(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0010540
(43) 공개일자 2022년01월25일

(51) 국제특허분류(Int. Cl.)
C12N 15/90 (2006.01) *A61K 31/7088* (2006.01)
A61K 48/00 (2006.01) *C12N 15/10* (2017.01)
C12N 15/113 (2010.01) *C12N 9/22* (2006.01)
C12N 9/78 (2006.01)
(52) CPC특허분류
C12N 15/907 (2013.01)
A61K 31/7088 (2013.01)
(21) 출원번호 10-2021-7041287
(22) 출원일자(국제) 2020년05월20일
심사청구일자 없음
(85) 번역문제출일자 2021년12월16일
(86) 국제출원번호 PCT/US2020/033807
(87) 국제공개번호 WO 2020/236936
국제공개일자 2020년11월26일
(30) 우선권주장
62/850,919 2019년05월21일 미국(US)

(71) 출원인
빔 테라퓨틱스, 인크.
미국 02139 매사추세츠주 케임브리지 랜즈다운 스트리트 26 세컨드 플로어
(72) 발명자
게르케 제이슨 마이클
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트 빔 테라퓨틱스, 인크. 내
페트로시안 나탈리
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트 빔 테라퓨틱스, 인크. 내
(74) 대리인
문두현

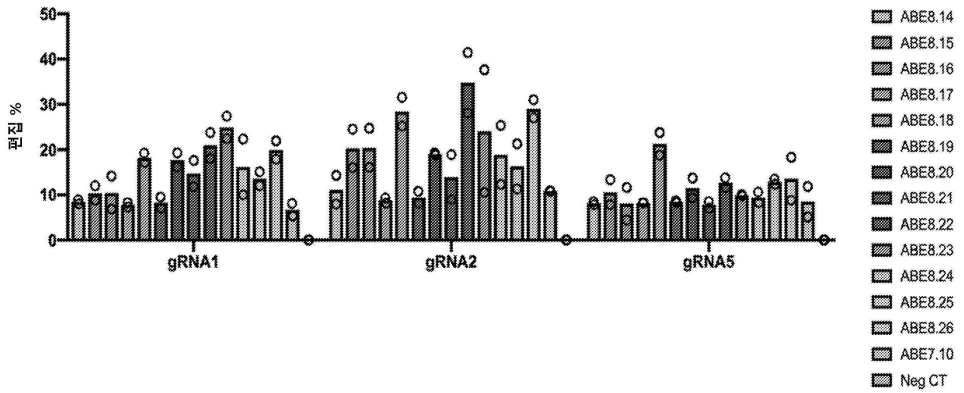
전체 청구항 수 : 총 188 항

(54) 발명의 명칭 **프로그래밍가능한 염기 편집기 시스템을 이용하여 단일염기다형성을 편집하는 방법**

(57) 요 약

레트 증후군 (RETT)과 연관된 돌연변이를 변경하기 위한 조성물 및 방법이 기재되어 있다. 본원에는 가이드 폴리뉴클레오타이드와 함께 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 핵염기 편집 도메인을 포함하는 염기 편집기 (예를 들어, ABE8)를 사용하는 조성물 및 방법이 제공된다. 또한 본원에는 표적 뉴클레오타이드 서열의 핵염기를 편집하기 위한 염기 편집기 시스템이 제공된다.

대 표 도



(52) CPC특허분류
A61K 48/00 (2013.01)
C12N 15/102 (2013.01)
C12N 15/113 (2013.01)
C12N 9/22 (2013.01)
C12N 9/78 (2013.01)
C12N 2310/20 (2017.05)

명 세 서

청구범위

청구항 1

대상체에서 메틸 CpG 결합 단백질 2 (*MECP2*) 유전자 또는 이의 조절 요소를 편집하는 방법으로서, 상기 방법은 이를 필요로 하는 대상체에게 (i) 아데노신 염기 편집기 또는 상기 아데노신 염기 편집기를 암호화하는 핵산 서열 및 (ii) 가이드 폴리뉴클레오타이드 또는 상기 가이드 폴리뉴클레오타이드를 암호화하는 핵산 서열을 투여하는 단계를 포함하고, 상기 아데노신 염기 편집기는 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 상기 아데노신 데아미나제 도메인은 TadA 참조 서열에 비해 아미노산 위치 82 또는 166 또는 이의 상응하는 위치에서 아미노산 치환을 포함하고, 상기 가이드 폴리뉴클레오타이드는 아데노신 염기 편집기가 레트 증후군 (RETT)과 연관된 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에서 A에서 G로의 핵염기 변경을 수행하도록 지시하고; 상기 A에서 G로의 핵염기 변경은 RETT와 연관된 SNP에 있으며, 이는 *MECP2* 유전자에 의해 암호화되는 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 또는 R306C 아미노산 돌연변이를 초래하는 방법.

청구항 2

제1항에 있어서, 상기 TadA 참조 서열이 아미노산 서열

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

를 포함하는, 방법

청구항 3

제1항 또는 제2항에 있어서, 상기 A에서 G로의 핵염기 변경이 RETT와 연관된 상기 SNP를 야생형 핵염기로 변화시키는, 방법.

청구항 4

제1항 또는 제2항에 있어서, 상기 A에서 G로의 핵염기 변경이 RETT와 연관된 상기 SNP를 비야생형 핵염기로 변화시켜 RETT의 하나 이상의 개선된 증상을 초래하는, 방법.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP에서 상기 A에서 G로의 핵염기 변경이 상기 메틸 CpG 결합 단백질 2 (MECP2) 폴리펩타이드에서 시스테인을 아르기닌으로 또는 정지 코돈을 아르기닌으로 변화시키는, 방법.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP가 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래하는, 방법.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 가이드 폴리뉴클레오타이드가 RETT와 연관된 상기 SNP를 포함하는 상기 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는, 방법.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 상기 아데노신 염기 편집기가 RETT와 연관된 상기 SNP를 포함하는 상기 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성하는, 방법.

청구항 9

제8항에 있어서, 상기 가이드 폴리뉴클레오타이드가 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함하는, 방법.

청구항 10

(i) 아데노신 염기 편집기 또는 상기 아데노신 염기 편집기를 암호화하는 핵산 서열 및 (ii) 가이드 폴리뉴클레오타이드 또는 상기 가이드 폴리뉴클레오타이드를 암호화하는 핵산 서열을 포함하는 염기 편집기 시스템으로서, 상기 아데노신 염기 편집기는 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 상기 아데노신 데아미나제 도메인은 TadA 참조 서열에 비해 아미노산 위치 82 또는 166 또는 이의 상응하는 위치에서 아미노산 치환을 포함하고, 상기 가이드 폴리뉴클레오타이드는 아데노신 염기 편집기가 레트 증후군(RETT)과 연관된 SNP를 포함하는 메틸 CpG 결합 단백질 2 (*MECP2*) 유전자 또는 이의 조절 요소에서 A에서 G로의 핵염기 변경을 수행하도록 지시하고; 상기 A에서 G로의 핵염기 변경은 RETT와 연관된 SNP에 있으며, 이는 *MECP2* 유전자에 의해 암호화되는 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 또는 R306C 아미노산 돌연변이를 초래하는 염기 편집기 시스템.

청구항 11

제10항에 있어서, 상기 A에서 G로의 핵염기 변경이 RETT와 연관된 상기 SNP를 야생형 핵염기로 변화시키는, 염기 편집기 시스템.

청구항 12

제10항에 있어서, 상기 A에서 G로의 핵염기 변경이 RETT와 연관된 상기 SNP를 비야생형 핵염기로 변화시켜 RETT의 하나 이상의 개선된 증상을 초래하는, 염기 편집기 시스템.

청구항 13

제10항 내지 제12항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP가 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래하는, 염기 편집기 시스템.

청구항 14

제10항 내지 제13항 중 어느 한 항에 있어서, 상기 가이드 폴리뉴클레오타이드가 RETT와 연관된 상기 SNP를 포함하는 상기 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는, 염기 편집기 시스템.

청구항 15

제10항 내지 제14항 중 어느 한 항에 있어서, 상기 아데노신 염기 편집기가 RETT와 연관된 상기 SNP를 포함하는 상기 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성하는, 염기 편집기 시스템.

청구항 16

제15항에 있어서, 상기 가이드 폴리뉴클레오타이드가 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함하는, 염기 편집기 시스템.

청구항 17

레트 증후군 (RETT)과 연관된 단일염기다형성 (SNP)를 포함하는 *MECP2* 폴리뉴클레오타이드를 편집하는 방법으로서, 상기 방법은 상기 *MECP2* 폴리뉴클레오타이드를 하나 이상의 가이드 폴리뉴클레오타이드와의 복합체를 이루고 있는 아데노신 데아미나제 염기 편집기 8 (ABE8)과 접촉시키는 단계를 포함하고, 상기 ABE8은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 상기 가이드 폴리뉴클레오타이드 중 하나 이상은 RETT와 연관된 *MECP2* 폴리뉴클레오타이드에서 SNP의 A·T에서 G·C로의 변경을 수행

하도록 상기 염기 편집기를 표적화하고, 상기 변경은 R133C 또는 R306C 중 하나 또는 둘 다인 방법.

청구항 18

제17항에 있어서, 상기 접촉이 세포, 진핵 세포, 포유류 세포, 또는 인간 세포에서 이루어지는, 방법.

청구항 19

제17항 또는 제18항에 있어서, 상기 세포가 생체내 또는 생체외에 있는, 방법.

청구항 20

제17항 내지 제19항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP에서 상기 A·T에서 G·C로의 변경이 상기 메틸 CpG 결합 단백질 2 (Mecp2) 폴리펩타이드에서 시스테인을 아르기닌으로 또는 정지 코돈을 아르기닌으로 변화시키는, 방법.

청구항 21

제17항 내지 제20항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP가 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래하는, 방법.

청구항 22

제17항 내지 제21항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 (*Streptococcus thermophilus*) *1* Cas9 (St1Cas9), 스텝토코커스 카니스 (*Steptococcus canis*) Cas9 (ScCas9)로부터 선택된 Cas9 또는 이의 변이체인, 방법.

청구항 23

제17항 내지 제22항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변경된 프로토스페이서-인접 모티프 (PAM)에 결합하는 변형된 SpCas9를 포함하는, 방법.

청구항 24

제23항에 있어서, 상기 변형된 SpCas9가 5'-NGT-3' 또는 5'-NGG-3'로부터 선택된 핵산 서열을 포함하는 PAM에 결합하는, 방법.

청구항 25

제23항에 있어서, 상기 변형된 SpCas9가 NGT PAM 변이체에 결합하는, 방법.

청구항 26

제24항 또는 제25항에 있어서, 상기 NGT PAM 변이체가 변형된 SpCas9의 하나 이상의 잔기 1335, 1337, 1135, 1136, 1218, 및/또는 1219에서 아미노산 치환, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 27

제23항 내지 제25항 중 어느 한 항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R, 및 L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 28

제23항 내지 제25항 중 어느 한 항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, 및 T1337, 및 L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, 및

T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 29

제17항 내지 제28항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 뉴클레아제 불활성 또는 닉카제 변이체인, 방법.

청구항 30

제29항에 있어서, 상기 닉카제 변이체가 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 31

제17항 내지 제30항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 데옥시리보핵산 (DNA)에서 아데노신을 탈아민화할 수 있는, 방법.

청구항 32

제17항 내지 제31항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 위치 82 및/또는 166에서 변경을 포함하는, 방법.

청구항 33

제32항에 있어서, 상기 아데노신 데아미나제 도메인이 아미노산 위치 82 및 166에서 변경을 포함하는, 방법.

청구항 34

제32항에 있어서, 상기 아데노신 데아미나제 도메인이 V82S 변경, T166R 변경, 또는 V82S 변경과 T166R 변경 둘 다로부터 선택된 변경을 포함하는, 방법.

청구항 35

제32항 내지 제34항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 하기의 변경 중 하나 이상을 추가로 포함하는, 방법: Y147T, Y147R, Q154S, Y123H, 및 Q154R.

청구항 36

제32항 내지 제35항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 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 그룹으로부터 선택된 변경을 포함하는, 방법

청구항 37

제17항 내지 제36항 중 어느 한 항에 있어서, 상기 ABE8이 아데노신 데아미나제 변이체 단량체를 포함하고, 상기 아데노신 데아미나제 단량체는 V82S 및 T166R 변경을 포함하는, 방법.

청구항 38

제17항 내지 제37항 중 어느 한 항에 있어서, 상기 ABE8이 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 방법.

청구항 39

제37항에 있어서, 상기 아데노신 데아미나제 변이체 단량체가 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및

Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함하는, 방법.

청구항 40

제17항에 있어서, 상기 ABE8이 TadA*8 도메인 및 야생형 TadA 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 방법.

청구항 41

제37항에 있어서, 상기 아데노신 데아미나제 단량체가 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 변경을 추가로 포함하는, 방법.

청구항 42

제17항에 있어서, 상기 ABE8 염기 편집기가 야생형 TadA 도메인, 및 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 그룹으로부터 선택된 변경의 조합을 포함하는 아데노신 데아미나제 변이체를 포함하는 이종이량체를 포함하는, 방법.

청구항 43

제17항 내지 제42항 중 어느 한 항에 있어서, 상기 가이드 폴리뉴클레오타이드가 AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함하는, 방법.

청구항 44

제17항에 있어서, 상기 아데노신 데아미나제가 TadA 데아미나제인, 방법.

청구항 45

제44항에 있어서, 상기 TadA 데아미나제가 TadA*8 변이체인, 방법.

청구항 46

제45항에 있어서, 상기 TadA*8 변이체가 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, 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, 및 TadA*8.24로 이루어진 그룹으로부터 선택되는, 방법.

청구항 47

제17항에 있어서, 상기 ABE8 염기 편집기가 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, 및 ABE8.24로 이루어진 그룹으로부터 선택되는, 방법.

청구항 48

제17항 내지 제47항 중 어느 한 항에 있어서, 상기 하나 이상의 가이드 RNA가 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 상기 crRNA는 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는, 방법.

청구항 49

제17항 내지 제48항 중 어느 한 항에 있어서, 상기 ABE8 염기 편집기가 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성하는, 방법.

청구항 50

제17항 내지 제49항 중 어느 한 항에 있어서, 상기 ABE8 염기 편집기가 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진, 방법:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

청구항 51

세포 또는 이의 전구체 내로

ABE8 염기 편집기, 또는 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드 (여기서, 상기 ABE8 염기 편집기는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함한다); 및

레트 증후군 (RETT)과 연관된 *MECP2* 폴리뉴클레오타이드에서 SNP의 A·T에서 G·C로의 변경을 수행하도록 염기 편집기를 표적화하는 하나 이상의 가이드 폴리뉴클레오타이드 (여기서, 상기 변경은 R133C 또는 R306C 중 하나 또는 둘 다이다)

를 도입하여 생성되는 세포.

청구항 52

제51항에 있어서, 상기 세포가 뉴런인, 세포.

청구항 53

제51항 또는 제52항에 있어서, 상기 뉴런이 MECP2 폴리펩타이드를 발현하는, 세포.

청구항 54

제51항 내지 제53항 중 어느 한 항에 있어서, 상기 세포가 RETT를 갖는 대상체로부터 유래되는, 세포.

청구항 55

제51항 내지 제54항 중 어느 한 항에 있어서, 상기 세포가 포유류 세포 또는 인간 세포인, 세포.

청구항 56

제51항 내지 제55항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP에서 상기 A·T에서 G·C로의 변경이 메틸 CpG 결합 단백질 2 (MECP2) 폴리펩타이드에서 시스테인을 아르기닌으로, 정지 코돈을 아르기닌으로 변화시키는, 세포.

청구항 57

제51항 내지 제56항 중 어느 한 항에 있어서, 상기 RETT와 연관된 상기 SNP가 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래하는, 세포.

청구항 58

제51항 내지 제57항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 스트렙토코커스 피오게네스 Cas9 (SpCas9), 스타필로코커스 아우레우스 Cas9 (SaCas9), 스트렙토코커스 써모필러스 *1* Cas9 (St1Cas9), 스텝토코커스 카니스 Cas9 (ScCas9)로부터 선택된 Cas9 또는 이의 변이체인, 세포.

청구항 59

제51항 내지 제58항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이

변경된 프로토스페이서-인접 모티프 (PAM)에 결합하는 변형된 SpCas9를 포함하는, 세포.

청구항 60

제59항에 있어서, 상기 변형된 SpCas9가 5'-NGT-3' 또는 5'-NGG-3'로부터 선택된 핵산 서열을 포함하는 PAM에 결합하는, 세포.

청구항 61

제60항에 있어서, 상기 변형된 SpCas9가 NGT PAM 변이체에 결합하는, 세포.

청구항 62

제61항에 있어서, 상기 NGT PAM 변이체가 변형된 SpCas9의 하나 이상의 잔기 1335, 1337, 1135, 1136, 1218, 및/또는 1219에서 아미노산 치환, 또는 이의 상응하는 아미노산 치환을 포함하는, 세포.

청구항 63

제59항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R, 및 L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함하는, 세포.

청구항 64

제59항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, 및 T1337, 및 L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함하는, 세포.

청구항 65

제51항 내지 제64항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 뉴클레아제 불활성 또는 닉카제 변이체인, 세포.

청구항 66

제65항에 있어서, 상기 닉카제 변이체가 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함하는, 세포.

청구항 67

제51항 내지 제66항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 데옥시리보핵산 (DNA)에서 아데노신을 탈아민화할 수 있는, 세포.

청구항 68

제51항 내지 제67항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

의 아미노산 위치 82 및/또는 166에서 변경을 포함하는, 세포.

청구항 69

제68항에 있어서, 상기 아데노신 데아미나제 도메인이 아미노산 위치 82 및 166에서 변경을 포함하는, 세포.

청구항 70

제69항에 있어서, 상기 아데노신 데아미나제 도메인이 V82S 변경, T166R 변경, 또는 V82S 변경과 T166R 변경 둘 다로부터 선택된 변경을 포함하는, 세포.

청구항 71

제68항 내지 제70항 중 어느 한 항에에 있어서, 상기 아데노신 데아미나제가 하기의 변경 중 하나 이상을 추가로 포함하는, 세포: Y147T, Y147R, Q154S, Y123H, 및 Q154R.

청구항 72

제68항 내지 제71항 중 어느 한 항에에 있어서, 상기 아데노신 데아미나제 도메인이 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 그룹으로부터 선택된 변경의 조합을 포함하는, 세포.

청구항 73

제68항 내지 제72항 중 어느 한 항에에 있어서, 상기 ABE8이 아데노신 데아미나제 변이체 단량체를 포함하고, 상기 아데노신 데아미나제 단량체는 V82S 및 T166R 변경을 포함하는, 세포.

청구항 74

제68항 내지 제72항 중 어느 한 항에에 있어서, 상기 ABE8이 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 세포.

청구항 75

제73항에 있어서, 상기 아데노신 데아미나제 변이체 단량체가 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함하는, 세포.

청구항 76

제51항에 있어서, 상기 ABE8이 TadA*7.10 도메인 및 TadA*8 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 세포.

청구항 77

제73항에 있어서, 상기 아데노신 데아미나제 변이체 단량체가 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함하는, 세포.

청구항 78

제51항에 있어서, 상기 ABE8 염기 편집기가 TadA*7.10 도메인, 및 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 그룹으로부터 선택된 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 이종이량체를 포함하는, 세포.

청구항 79

제51항에 있어서, 상기 가이드 폴리뉴클레오타이드가 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함하는, 세포.

청구항 80

제51항에 있어서, 상기 아데노신 데아미나제가 TadA 데아미나제인, 세포.

청구항 81

제80항에 있어서, 상기 TadA 데아미나제가 TadA*8 변이체인, 세포.

청구항 82

제81항에 있어서, 상기 TadA*8 변이체가 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, 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, 및 TadA*8.24로 이루어진 그룹으로부터 선택되는, 세포.

청구항 83

제51항에 있어서, 상기 ABE8 염기 편집기가 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, 및 ABE8.24로 이루어진 그룹으로부터 선택되는, 세포.

청구항 84

제51항 내지 제83항 중 어느 한 항에 있어서, 상기 하나 이상의 가이드 RNA가 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 상기 crRNA는 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는, 세포.

청구항 85

제51항 내지 제84항 중 어느 한 항에 있어서, 상기 염기 편집기가 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성하는, 방법.

청구항 86

제51항 내지 제85항 중 어느 한 항에 있어서, 상기 ABE8 염기 편집기가 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진, 세포:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

청구항 87

제51항 내지 제86항 중 어느 한 항에 있어서, 상기 gRNA가 하기 서열을 갖는 스캐폴드를 포함하는, 세포:

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGCACCGAGTCGGTGCTTTTTTT.

청구항 88

대상체에게 다음을 투여하는 단계를 포함하여, 상기 대상체에서 레트 증후군 (RETT)을 치료하는 방법:

ABE8 염기 편집기, 또는 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드 (여기서, 상기 ABE8 염기 편집기는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함한다); 및

RETT와 연관된 MECP2 폴리뉴클레오타이드에서 SNP의 A·T에서 G·C로의 변경을 수행하도록 상기 ABE8 염기 편집기를 표적화하는 하나 이상의 가이드 폴리뉴클레오타이드 (여기서, 상기 변경은 R133C 및/또는 R306C 중 하나 또는 둘 다이다).

청구항 89

제88항에 있어서, 상기 대상체가 포유동물 또는 인간인, 방법.

청구항 90

제88항 또는 제89항에 있어서, 상기 ABE8 염기 편집기, 또는 상기 ABE8 염기 편집기를 암호화하는 폴리뉴클레오타이드, 및 상기 하나 이상의 가이드 폴리뉴클레오타이드를 상기 대상체의 세포에 전달하는 단계를 포함하고, 임의로 상기 세포는 뉴런인, 방법.

청구항 91

제88항 내지 제90항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP에서 상기 A·T에서 G·C로의 변경이 메틸 CpG 결합 단백질 2 (MECP2) 폴리펩타이드에서 시스테인을 아르기닌으로 또는 정지 코돈을 아르기닌으로 변화시키는, 방법.

청구항 92

제88항 내지 제91항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP가 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래하는, 방법.

청구항 93

제88항 내지 제92항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 스트렙토코커스 피오게네스 Cas9 (SpCas9), 스타필로코커스 아우레우스 Cas9 (SaCas9), 스트렙토코커스 써모필러스 *1* Cas9 (St1Cas9), 스텝토코커스 카니스 Cas9 (ScCas9)로부터 선택된 Cas9 또는 이의 변이체인, 방법.

청구항 94

제88항 내지 제93항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변경된 프로토스페이서-인접 모티프 (PAM)에 결합하는 변형된 SpCas9를 포함하는, 방법.

청구항 95

제94항에 있어서, 상기 변형된 SpCas9가 5'-NGT-3' 또는 5'-NGG-3'로부터 선택된 핵산 서열을 포함하는 PAM에 결합하는, 방법.

청구항 96

제94항에 있어서, 상기 변형된 SpCas9가 NGT PAM 변이체에 결합하는, 방법.

청구항 97

제96항에 있어서, 상기 NGT PAM 변이체가 변형된 SpCas9의 하나 이상의 잔기 1335, 1337, 1135, 1136, 1218, 및/또는 1219에서 아미노산 치환, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 98

제94항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R, 및 L1111, D1135L, S1136R, G1218S, E1219V, D1332A, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 99

제94항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, 및 T1337, 및 L1111R, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 100

제88항 내지 제99항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이

뉴클레아제 불활성 또는 닉카제 변이체인, 방법.

청구항 101

제100항에 있어서, 상기 닉카제 변이체가 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 102

제81항 내지 제101항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 데옥시리보핵산 (DNA)에서 아데노신을 탈아민화할 수 있는, 방법.

청구항 103

제88항 내지 제102항에 있어서, 상기 아데노신 데아미나제 도메인이

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

의 아미노산 위치 82 및 /또는 166에서 변경을 포함하는, 방법.

청구항 104

제103항에 있어서, 상기 아데노신 데아미나제 도메인이 아미노산 위치 82 및 166에서 변경을 포함하는, 방법.

청구항 105

제103항에 있어서, 상기 아데노신 데아미나제 도메인이 V82S 변경, T166R 변경, 또는 V82S 변경과 T166R 변경 둘 다로부터 선택된 변경을 포함하는, 방법.

청구항 106

제103항 내지 제105항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 하기의 변경 중 하나 이상을 추가로 포함하는, 방법: Y147T, Y147R, Q154S, Y123H, 및 Q154R.

청구항 107

제103항 내지 제106항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 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 그룹으로부터 선택된 변경을 포함하는, 방법.

청구항 108

제88항 내지 제107항 중 어느 한 항에 있어서, 상기 ABE8이 아데노신 데아미나제 변이체 단량체를 포함하고, 상기 아데노신 데아미나제 단량체는 V82S 및 T166R 변경을 포함하는, 방법.

청구항 109

제88항 내지 제107항 중 어느 한 항에 있어서, 상기 ABE8이 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 방법.

청구항 110

제108항에 있어서, 상기 아데노신 데아미나제 변이체 단량체가 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함하는, 방법.

청구항 111

제88항에 있어서, 상기 ABE8이 TadA*7.10 도메인 및 TadA*8 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 방법.

청구항 112

제108항에 있어서, 상기 아데노신 데아미나제 변이체 단량체가 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함하는, 방법.

청구항 113

제88항에 있어서, 상기 ABE8 염기 편집기가 TadA7.10 도메인, 및 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 그룹으로부터 선택된 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 이종이량체를 포함하는, 방법.

청구항 114

제88항 내지 제113항 중 어느 한 항에 있어서, 상기 가이드 폴리뉴클레오타이드가 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 갖는, 방법.

청구항 115

제102항에 있어서, 상기 아데노신 데아미나제가 TadA 데아미나제인, 방법.

청구항 116

제115항에 있어서, 상기 TadA 데아미나제가 TadA*8 변이체인, 방법.

청구항 117

제116항에 있어서, 상기 TadA*8 변이체가 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, 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, 및 TadA*8.24로 이루어진 그룹으로부터 선택되는, 방법.

청구항 118

제88항에 있어서, 상기 ABE8 염기 편집기가 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, 및 ABE8.24로 이루어진 그룹으로부터 선택되는, 방법.

청구항 119

제88항 내지 제118항 중 어느 한 항에 있어서, 상기 하나 이상의 가이드 RNA가 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 상기 crRNA는 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는, 방법.

청구항 120

제88항 내지 제119항 중 어느 한 항에 있어서, 상기 염기 편집기가 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성하는, 방법.

청구항 121

대상체에서 레트 증후군 (RETT)을 치료하는 방법으로서, 상기 방법은:

이를 필요로 하는 대상체에게 (i) 아데노신 염기 편집기 또는 상기 아데노신 염기 편집기를 암호화하는 핵산 서열 및 (ii) 가이드 폴리뉴클레오타이드 또는 상기 가이드 폴리뉴클레오타이드를 암호화하는 핵산 서열을 투여하여 상기 대상체에서 RETT를 치료하는 단계를 포함하고, 상기 아데노신 염기 편집기는 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 상기 아데노신 데아미나제 도메인은 TadA 참조 서열에 비해 아미노산 위치 82 또는 166 또는 이의 상응하는 위치에서 아미노산 치환을 포함하고,

상기 가이드 폴리뉴클레오타이드는 상기 아데노신 염기 편집기가 상기 대상체에서 RETT와 연관된 SNP를 포함하는 메틸 CpG 결합 단백질 2 (*MECP2*) 유전자 또는 이의 조절 요소에서 A에서 G로의 핵염기 변경을 수행하도록 지시하고, RETT와 연관된 상기 SNP는 상기 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 또는 R306C 아미노산 돌연변이를 초래하는 방법.

청구항 122

제121항에 있어서, 상기 투여가 RETT와 연관된 적어도 하나의 증상을 개선하는, 방법.

청구항 123

제121항에 있어서, 상기 투여가 상기 아데노신 데아미나제에서 아미노산 치환이 없는 염기 편집기를 사용한 치료와 비교하여 RETT와 관련된 적어도 하나의 증상의 더 빠른 개선을 초래하는, 방법.

청구항 124

제121항 내지 제123항 중 어느 한 항에 있어서, 상기 A에서 G로의 핵염기 변경이 RETT와 연관된 상기 SNP를 야생형 핵염기로 변화시키는, 방법.

청구항 125

제121항 내지 제124항 중 어느 한 항에 있어서, 상기 A에서 G로의 핵염기 변경이 레트 증후군과 연관된 SNP를 비야생형 핵염기로 변화시켜 개선된 RETT 증상을 초래하는, 방법.

청구항 126

제121항 내지 제125항 중 어느 한 항에 있어서, 상기 가이드 폴리뉴클레오타이드가 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는, 방법.

청구항 127

제121항 내지 제126항 중 어느 한 항에 있어서, 상기 아데노신 염기 편집기가 RETT와 연관된 상기 SNP를 포함하는 상기 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성하는, 방법.

청구항 128

제127항에 있어서, 상기 가이드 폴리뉴클레오타이드가 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함하는, 방법.

청구항 129

제1항 내지 제9항 또는 제121항 내지 제128항, 또는 제10항 내지 제16항 중 어느 한 항에 있어서, 상기 가이드 폴리뉴클레오타이드가 상기 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 적어도 10개의 인접 뉴클레오타이드를 포함하는 핵산 서열을 포함하는, 방법 또는 염기 편집기.

청구항 130

제129항에 있어서, 상기 가이드 폴리뉴클레오타이드가 상기 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 15,

16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 또는 40개의 인접 뉴클레오타이드를 포함하는 핵산 서열을 포함하는, 방법.

청구항 131

제17항 내지 제50항 또는 제88항 내지 제120항, 또는 제51항 내지 제87항 중 어느 한 항에 있어서, 상기 가이드 폴리뉴클레오타이드가 상기 *MECP2* 폴리뉴클레오타이드에 상보적인 적어도 10개의 인접 뉴클레오타이드를 포함하는 핵산 서열을 포함하는, 방법 또는 세포.

청구항 132

제131항에 있어서, 상기 가이드 폴리뉴클레오타이드가 상기 *MECP2* 폴뉴클레오타이드에 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 또는 40개의 인접 뉴클레오타이드를 포함하는 핵산 서열을 포함하는, 방법.

청구항 133

제121항 내지 제128항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP가 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 및/또는 R306C 아미노산 돌연변이를 초래하는, 방법.

청구항 134

제133항에 있어서, RETT와 연관된 상기 SNP가 상기 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 아미노산 돌연변이를 초래하는, 방법.

청구항 135

제133항에 있어서, RETT와 연관된 상기 SNP가 상기 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R306C 아미노산 돌연변이를 초래하는, 방법.

청구항 136

제10항 내지 제16항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP에서 상기 A에서 G로의 핵염기 변경이 상기 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 아미노산 돌연변이를 초래하는, 염기 편집기 시스템.

청구항 137

제10항 내지 제16항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP에서 상기 A에서 G로의 핵염기 변경이 상기 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R306C 아미노산 돌연변이를 초래하는, 염기 편집기 시스템.

청구항 138

제17항 내지 제50항 또는 제88항 내지 제120항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP의 변경이 R133C와 R306C를 둘 다 포함하는, 방법.

청구항 139

제17항 내지 제50항 또는 제88항 내지 제120항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP의 변경이 R133C인, 방법.

청구항 140

제17항 내지 제50항 또는 제88항 내지 제120항 중 어느 한 항에 있어서, RETT와 연관된 상기 SNP의 변경이 R306C인, 방법.

청구항 141

제51항 내지 제87항 중 어느 한 항에 있어서, 레트 증후군 (RETT)과 연관된 상기 SNP의 변경이 R133C인, 세포.

청구항 142

제51항 내지 제87항 중 어느 한 항에 있어서, 레트 증후군 (RETT)과 연관된 상기 SNP의 변경이 R306C인, 세포.

청구항 143

MECP2 단백질을 암호화하는 *MECP2* 유전자에 완벽하게 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개의 인접 뉴클레오타이드를 포함하는 가이드 폴리뉴클레오타이드 또는 가이드 RNA.

청구항 144

제143항에 있어서, 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함하는, 가이드 폴리뉴클레오타이드 또는 가이드 RNA.

청구항 145

제143항 또는 제144항에 있어서, 스캐폴드 서열을 추가로 포함하고, 상기 스스캐폴드 서열은 임의로 다음과 같은, 가이드 폴리뉴클레오타이드 또는 가이드 RNA:

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC

CGAGTCGGTGCTTTTTTT.

청구항 146

아데노신 데아미나제 염기 편집기 8 (ABE8) 및 가이드 RNA를 포함하는 조성물로서, 상기 ABE8은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 상기 가이드 RNA는 레트 증후군과 연관된 *MECP2* 폴리뉴클레오타이드에서 SNP의 A·T에서 G·C로의 변경을 수행하도록 염기 편집기를 표적화하고, 상기 변경은 R133C 또는 R306C 중 하나 또는 둘 다인 조성물.

청구항 147

제146항에 있어서, RETT와 연관된 상기 SNP에서 상기 A·T에서 G·C로의 변경이 메틸 CpG 결합 단백질 2 (MECP2) 폴리펩타이드에서 시스테인을 아르기닌으로 또는 정지 코돈을 아르기닌으로 변화시키는, 조성물.

청구항 148

제146항 또는 제147항에 있어서, RETT와 연관된 상기 SNP가 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래하는, 조성물.

청구항 149

제146항 내지 제148항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 스트렙토코커스 피오게네스 Cas9 (SpCas9), 스타필로코커스 아우레우스 Cas9 (SaCas9), 스트렙토코커스 써모필러스 *1* Cas9 (St1Cas9), 스텝토코커스 카니스 Cas9 (ScCas9)로부터 선택된 Cas9 또는 이의 변이체인, 조성물.

청구항 150

제146항 내지 제149항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변경된 프로토스페이서-인접 모티프 (PAM)에 결합하는 변형된 SpCas9를 포함하는, 조성물.

청구항 151

제146항 내지 제149항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 뉴클레아제 불활성 또는 닉카제 변이체인, 조성물.

청구항 152

제151항에 있어서, 상기 닉카제 변이체가 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함하는, 조성물.

청구항 153

제146항 내지 제152항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 데옥시리보핵산 (DNA)에서 아데노신을 탈아민화할 수 있는, 조성물.

청구항 154

제146항 내지 제153항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 위치 82 및/또는 166에서 변경을 포함하는, 조성물.

청구항 155

제154항에 있어서, 상기 아데노신 데아미나제 도메인이 V82S 변경, T166R 변경, 또는 V82S 변경과 T166R 변경 둘 다로부터 선택된 변경을 포함하는, 조성물.

청구항 156

제154항 또는 제155항에 있어서, 상기 아데노신 데아미나제 도메인이 하기의 변경 중 하나 이상을 추가로 포함하는, 조성물: Y147T, Y147R, Q154S, Y123H, 및 Q154R.

청구항 157

제154항 내지 제156항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 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 그룹으로부터 선택된 변경을 포함하는, 조성물.

청구항 158

제146항 내지 제157항 중 어느 한 항에 있어서, 상기 ABE8이 아데노신 데아미나제 변이체 단량체를 포함하고, 상기 아데노신 데아미나제 단량체가 V82S 및 T166R 변경을 포함하는, 조성물.

청구항 159

제148항 내지 제158항 중 어느 한 항에 있어서, 상기 ABE8이 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 조성물.

청구항 160

제158항에 있어서, 상기 아데노신 데아미나제 변이체 단량체가 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함하는, 조성물.

청구항 161

제146항에 있어서, 상기 ABE8이 TadA*8 도메인 및 야생형 TadA 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 조성물.

청구항 162

제158항에 있어서, 상기 아데노신 데아미나제 변이체 단량체가 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함하는, 조성물.

청구항 163

제146항에 있어서, 상기 ABE8 염기 편집기가 야생형 TadA 도메인, 및 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 그룹으로부터 선택된 변경의 조합을 포함하는 아데노신 데아미나제 변이체를 포함하는 이종이량체를 포함하는, 조성물.

청구항 164

제146항 내지 제163항 중 어느 한 항에 있어서, 상기 가이드 RNA가 AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함하는, 조성물.

청구항 165

제146항에 있어서, 상기 아데노신 데아미나제가 TadA 데아미나제인, 조성물.

청구항 166

제165항에 있어서, 상기 TadA 데아미나제가 TadA*8 변이체인, 조성물.

청구항 167

제166항에 있어서, 상기 TadA*8 변이체가 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, 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, 및 TadA*8.24로 이루어진 그룹으로부터 선택되는, 조성물.

청구항 168

제146항에 있어서, 상기 ABE8 염기 편집기가 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, 및 ABE8.24로 이루어진 그룹으로부터 선택되는, 조성물.

청구항 169

제146항 내지 제168항 중 어느 한 항에 있어서, 상기 가이드 RNA가 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 상기 crRNA는 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는, 조성물.

청구항 170

제146항 내지 제169항 중 어느 한 항에 있어서, 상기 ABE8 염기 편집기가 RETT와 연관된 상기 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성하는, 조성물.

청구항 171

제146항 내지 제170항 중 어느 한 항에 있어서, 상기 ABE8 염기 편집기가 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진, 조성물:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

청구항 172

제146항 내지 제171항 중 어느 한 항에 있어서, 지질을 추가로 포함하고, 임의로 상기 지질이 양이온성 지질인, 조성물.

청구항 173

제146항 내지 제172항 중 어느 한 항에 있어서, 약제학적으로 허용되는 부형제 또는 희석제를 포함하는 약제학적 조성물인, 조성물.

청구항 174

제173항에 있어서, 레트 증후군의 치료를 위한, 약제학적 조성물.

청구항 175

제173항에 있어서, 상기 gRNA 및 상기 ABE8 염기 편집기가 함께 또는 별도로 제형화되는, 약제학적 조성물.

청구항 176

제172항 내지 제175항 중 어느 한 항에 있어서, 포유류 세포에서 발현에 적합한 벡터를 추가로 포함하고, 상기 벡터는 상기 ABE8 염기 편집기를 암호화하는 폴리뉴클레오타이드를 포함하는, 약제학적 조성물.

청구항 177

제176항에 있어서, 상기 벡터가 바이러스 벡터인, 약제학적 조성물.

청구항 178

제177항에 있어서, 상기 바이러스 벡터가 레트로바이러스 벡터, 아데노바이러스 벡터, 렌티바이러스 벡터, 헤르페스바이러스 벡터, 또는 아데노-연관된 바이러스 벡터 (AAV)인, 약제학적 조성물.

청구항 179

제172항 내지 제178항 중 어느 한 항에 있어서, 포유류 세포에서 발현에 적합한 리보핵입자를 추가로 포함하는, 약제학적 조성물.

청구항 180

레트 증후군을 치료하는 방법으로서, 상기 방법은 제172항 내지 제179항 중 어느 한 항의 약제학적 조성물을 이를 필요로 하는 대상체에게 투여하는 단계를 포함하는 방법.

청구항 181

대상체에서 레트 증후군의 치료에 있어서 제172항 내지 제179항 중 어느 한 항의 약제학적 조성물의 용도.

청구항 182

제180항 또는 제181항에 있어서, 상기 대상체가 포유동물 또는 인간인, 방법 또는 용도.

청구항 183

제51항 내지 제87항 중 어느 한 항의 세포를 포함하는 조성물.

청구항 184

제183항에 있어서, 약제학적으로 허용되는 담체 또는 희석제를 추가로 포함하는, 조성물.

청구항 185

(i) ABE8 염기 편집기를 암호화하는 핵산; 및 (ii) 제143항 내지 제145항 중 어느 한 항의 가이드 폴리뉴클레오타이드 또는 가이드 RNA를 포함하는 약제학적 조성물.

청구항 186

제185항에 있어서, 지질을 추가로 포함하는, 약제학적 조성물.

청구항 187

제186항에 있어서, 상기 지질이 양이온성 지질인, 약제학적 조성물.

청구항 188

제186항 또는 제187항에 있어서, 상기 염기 편집기를 암호하하는 핵산이 mRNA인, 약제학적 조성물.

발명의 설명

기 술 분 야

[0001] **관련 출원에 대한 상호 참조**

[0002] 본 국제 PCT 출원은 2019년 5월 21일에 출원된 가특허 출원 번호 62/850,919의 우선권 및 이익을 주장하며, 이의 전체 내용은 그 전문이 본원에 참조로 포함된다.

배 경 기 술

[0003] 레트 증후군 (RTT 또는 RETT)은 중추 신경계 (CNS)에서 염색질 및 전사 상태를 변형하는 암호화된 단백질의 능력을 손상시키거나 없애는 메틸-CpG-결합 단백질 2 (*Mecp2*) 유전자에서 이질적인 돌연변이 그룹에 의해 유발된다. 기능적 *Mecp2*를 전달하기 위한 유전자 요법 또는 내인성 *Mecp2* (*MECP2*) mRNA 전사체를 복구하기 위해 RNA 편집을 사용하는 것은 CNS 전체에 광범위하게 전달될 때 치료 개입에 대한 유망한 접근법이다. 그러나, 두 접근법은 치료 효능을 달성하기 위해 중요한 난제를 극복해야 한다. *Mecp2* 유전자 요법은 세포당을 기본으로 전달된 유전자의 투여량 또는 *Mecp2* 복제 증후군의 표현형을 모방하는 위험을 엄격하게 제어해야 한다. RNA 편집 플랫폼은 RTT 진단의 45% 초과를 차지하는 가장 만연한 *Mecp2* 돌연변이를 정밀하게 교정할 수 없으며 또한 효율적이고 가이드되지 않은 표적-이탈(off-target) 편집을 유도할 수도 없다.

[0004] 레트 증후군 (RTT)을 야기하는 *Mecp2*의 유전적 돌연변이는 매우 이종성이다. 결과적으로, 치료에 대해 선호되는 전략은 재조합 아데노-연관된 바이러스 (rAAV)에 의해 운반되는 야생형 *Mecp2*를 전달하는 것이었다. 이 전략은 각 개체에서 인과적 돌연변이에 불가지론적이기 때문에, 성공적인 유전자 요법 접근법은 RTT 환자 집단의 많은 부분에 치료 옵션을 제공할 것이다. 그러나, 현재까지 이 전략은 제한된 성공을 거두었다. RTT 환자는 거의 항상 이형접합 여성이므로, 무작위 X-염색체 불활성화로 인한 중추 신경계 (CNS) 내에서 특징적인 야생형 및 돌연변이 X-연결된 MeCP2 모자이크 발현을 초래한다. 따라서, 이미 야생형 MeCP2를 발현하는 뉴런에서 야생형 MeCP2의 rAAV 전달 및 발현은 MeCP2 복제 증후군의 표현형을 부분적으로 모방할 가능성이 있다. 이와 일관되게, RTT-모델 마우스의 CNS에서 높은 형질도입 효율은 야생형 마우스에서 발견되는 것보다 대략 2배 더 큰 MeCP2 발현을 초래하였다.

[0005] 따라서, 레트 증후군을 가진 환자를 치료하기 위한 신규한 조성물 및 방법이 필요하다.

[0006] **참조 인용**

[0007] 본 명세서에 언급된 모든 공보, 특허 및 특허 출원은 각각의 개별 공보, 특허 또는 특허 출원이 참조로 포함되는 것으로 구체적이고 개별적으로 표시된 것처럼 동일한 정도로 본원에 참조로 포함된다. 어떠한 다른 표시가 없는 한, 본 명세서에 언급된 공보, 특허 및 특허 출원은 이들의 전문이 본원에 참조로 포함된다.

발명의 내용

[0008] 본 개시의 요약

[0009] 아래에 기재된 바와 같이, 본 발명은 프로그래밍 가능한 핵염기 편집기를 사용하여 병원성 아미노산의 정밀 교정을 위한 조성물 및 방법을 특징으로 한다. 특히, 본 발명의 조성물 및 방법은 레트 증후군 (RTT 또는 RETT)의 치료에 유용하다. 따라서, 본 발명은 유해한 돌연변이 (예를 들어, R106W, R133C, T158M, R255*, R270*, R306C)를 교정하기 위해 내인성 *Mecp2* 유전자에서 단일염기다형성을 정밀하게 교정하기 위해 아데노신 (A) 염기 편집기 (ABE) (예를 들어, ABE8)를 사용하여 레트 증후군을 치료하기 위한 조성물 및 방법을 제공한다.

[0010] 하나의 양상에서, 대상체에서 메틸 CpG 결합 단백질 2 (*MECP2*) 유전자 또는 이의 조절 요소를 편집하는 방법이

제공되며, 여기서, 상기 방법은 이를 필요로 하는 대상체에 (i) 아데노신 염기 편집기 또는 상기 아데노신 염기 편집기를 암호화하는 핵산 서열 및 (ii) 가이드 폴리뉴클레오타이드 또는 상기 가이드 폴리뉴클레오타이드를 암호화하는 핵산 서열을 투여하는 단계를 포함하고, 여기서, 아데노신 염기 편집기는 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 아데노신 데아미나제 도메인은 TadA 참조 서열에 비해 아미노산 위치 82 또는 166 또는 이의 상응하는 위치에서 아미노산 치환을 포함하고, 가이드 폴리뉴클레오타이드는 아데노신 염기 편집기가 레트 증후군 (RETT)과 연관된 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에서 A에서 G로의 핵염기 변경을 수행하도록 지시하고; A에서 G로의 핵염기 변경은 RETT와 연관된 SNP에 있으며, 이는 *MECP2* 유전자에 의해 암호화되는 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 또는 R306C 아미노산 돌연변이를 초래한다. 하나의 구현예에서, TadA 참조 서열은 아미노산 서열

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0011] 를 포함한다. 하나의 구현예에서, A에서 G로의 핵염기 변경은 RETT와 연관된 SNP를 야생형 핵염기로 변화시킨다. 하나의 구현예에서, A에서 G로의 핵염기 변경은 RETT와 연관된 SNP를 RETT의 하나 이상의 개선된 증상을 초래하는 비야생형 핵염기로 변화시킨다. 하나의 구현예에서, RETT와 연관된 SNP에서 A에서 G로의 핵염기 변경은 메틸 CpG 결합 단백질 2 (MECP2) 폴리펩타이드에서 시스테인을 아르기닌으로 또는 정지 코돈을 아르기닌으로 변화시킨다. 하나의 구현예에서, RETT와 연관된 SNP는 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 RETT와 연관된 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함한다. 하나의 구현예에서, 아데노신 염기 편집기는 RETT와 연관된 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3' 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함한다.

[0012] 하나의 양상에서, 염기 편집기 시스템이 제공되고, 여기서, 염기 편집기 시스템은 (i) 아데노신 염기 편집기 또는 상기 아데노신 염기 편집기를 암호화하는 핵산 서열 및 (ii) 가이드 폴리뉴클레오타이드 또는 상기 가이드 폴리뉴클레오타이드를 암호화하는 핵산 서열을 포함하고, 여기서, 아데노신 염기 편집기는 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 아데노신 데아미나제 도메인은 TadA 참조 서열에 비해 아미노산 위치 82 또는 166 또는 이의 상응하는 위치에서 아미노산 치환을 포함하고, 가이드 폴리뉴클레오타이드는 아데노신 염기 편집기가 레트 증후군 (RETT)과 연관된 SNP를 포함하는 메틸 CpG 결합 단백질 2 (*MECP2*) 유전자 또는 이의 조절 요소에서 A에서 G로의 핵염기 변경을 수행하도록 지시하고; A에서 G로의 핵염기 변경은 RETT와 연관된 SNP에 있으며, 이는 *MECP2* 유전자에 의해 암호화되는 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 또는 R306C 아미노산 돌연변이를 초래한다. 하나의 구현예에서, A에서 G로의 핵염기 변경은 RETT와 연관된 SNP를 야생형 핵염기로 변화시킨다. 하나의 구현예에서, A에서 G로의 핵염기 변경은 RETT와 연관된 SNP를 비야생형 핵염기로 변화시켜 RETT의 하나 이상의 개선된 증상을 초래한다. 하나의 구현예에서, RETT와 연관된 SNP는 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 RETT와 연관된 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함한다. 하나의 구현예에서, 아데노신 염기 편집기는 RETT와 연관된 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함한다.

[0013] 또 다른 양상에서, 레트 증후군 (RETT)과 연관된 단일염기다형성 (SNP)를 포함하는 *MECP2* 폴리뉴클레오타이드를 편집하는 방법이 제공되고, 여기서, 상기 방법은 *MECP2* 폴리뉴클레오타이드를 하나 이상의 가이드 폴리뉴클레오타이드와의 복합체에서 아데노신 데아미나제 염기 편집기 8 (ABE8)과 접촉시키는 단계를 포함하고, 상기 ABE8은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 상기 가이드 폴리뉴클레오타이드 중 하나 이상은 상기 염기 편집기를 표적으로 하여 RETT와 연관된 *MECP2* 폴리뉴클레오타이드에서 SNP의 A·T에서 G·C로의 변경을 수행하고, 상기 변경은 R133C 또는 R306C 중 하나 또는 둘 다이다.

하나의 구현예에서, 접촉은 세포, 진핵 세포, 포유류 세포, 또는 인간 세포에서 이루어진다. 하나의 구현예에서, 세포는 생체내 또는 생체외에 있다. 하나의 구현예에서, RETT와 연관된 SNP에서 A·T에서 G·C로의 변경은 메틸 CpG 결합 단백질 2 (Mecp2) 폴리펩타이드에서 시스테인을 아르기닌으로 또는 정지 코돈을 아르기닌으로 변화시킨다. 하나의 구현예에서, RETT와 연관된 SNP는 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래한다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 (*Streptococcus thermophilus*) *1* Cas9 (St1Cas9), 스텝토코커스 카니스 (*Steptococcus canis*) Cas9 (ScCas9)로부터 선택된 Cas9 또는 이의 변이체이다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 변경된 프로토스페이서-인접 모티프 (PAM)에 결합하는 변형된 SpCas9를 포함한다. 하나의 구현예에서, 변형된 SpCas9는 5'-NGT-3' 또는 5'-NGG-3'로부터 선택된 핵산 서열을 포함하는 PAM에 결합한다. 하나의 구현예에서, 변형된 SpCas9는 NGT PAM 변이체에 결합한다. 하나의 구현예에서, NGT PAM 변이체는 변형된 SpCas9의 하나 이상의 잔기 1335, 1337, 1135, 1136, 1218, 및/또는 1219에서 아미노산 치환, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 변형된 SpCas9는 아미노산 치환 L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R, 및 L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 변형된 SpCas9는 아미노산 치환 D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, 및 T1337, 및 L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 뉴클레아제 불활성 또는 닉카제 변이체이다. 하나의 구현예에서, 닉카제 변이체는 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 데옥시리보핵산 (DNA)에서 아데노신을 탈아민화할 수 있다. 하나의 구현예에서, 아데노신 데아미나제 도메인은

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH

AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG

SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

의 아미노산 위치 82 및/또는 166에서 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 아미노산 위치 82 및 166에서 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 V82S 변경, T166R 변경, 또는 V82S 변경과 T166R 변경 둘 다로부터 선택된 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 하기의 변경 중 하나 이상을 추가로 포함한다: Y147T, Y147R, Q154S, Y123H, 및 Q154R. 하나의 구현예에서, 아데노신 데아미나제 도메인은 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 그룹으로부터 선택된 변경을 포함한다. 하나의 구현예에서, ABE8은 아데노신 데아미나제 변이체 단량체를 포함하고, 여기서, 아데노신 데아미나제 단량체는 V82S 및 T166R 변경을 포함한다. 하나의 구현예에서, ABE8은 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체 단량체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함한다. 하나의 구현예에서, ABE8은 TadA*8 도메인 및 야생형 TadA 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 단량체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 변경을 추가로 포함한다. 하나의 구현예에서, ABE8 염기 편집기는 야생형 TadA 도메인, 및 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 그룹으로부터 선택된 변경의 조합을 포함하는 아데노신 데아미나제 변이체를 포함하는 이종이량체를 포함한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-

GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함한다. 하나의 구현예에서, 아데노신 데아미나제는 TadA 데아미나제이다. 하나의 구현예에서, TadA 데아미나제는 TadA*8 변이체이다. 하나의 구현예에서, TadA*8 변이체는 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, 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, 및 TadA*8.24로 이루어진 그룹으로부터 선택된다. 하나의 구현예에서, ABE8 염기 편집기는 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, 및 ABE8.24로 이루어진 그룹으로부터 선택된다. 하나의 구현예에서, 하나 이상의 가이드 RNA는 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 상기 crRNA는 RETT와 연관된 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함한다. 하나의 구현예에서, ABE8 염기 편집기는 RETT와 연관된 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성한다. 하나의 구현예에서, ABE8 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0014]

[0015] 또 다른 측면에서, 세포 또는 이의 전구체 내로 (i) ABE8 염기 편집기, 또는 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드 (여기서, 상기 ABE8 염기 편집기는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함한다); 및 (ii) 레트 증후군 (RETT)과 연관된 *MECP2* 폴리뉴클레오타이드에서 SNP의 A·T에서 G·C로의 변경을 수행하도록 ABE8 염기 편집기를 표적화하는 하나 이상의 가이드 폴리뉴클레오타이드 (여기서, 상기 변경은 R133C 또는 R306C 중 하나 또는 둘 다이다)를 도입하여 생성된 세포가 제공된다. 하나의 구현예에서, 세포는 뉴런이다. 하나의 구현예에서, 뉴런은 MECP2 폴리펩타이드를 발현한다. 하나의 구현예에서, 세포는 RETT를 갖는 대상체로부터 유래된다. 하나의 구현예에서, 세포는 포유류 세포 또는 인간 세포이다. 하나의 구현예에서, RETT와 연관된 SNP에서 A·T에서 G·C로의 변경은 메틸 CpG 결합 단백질 2 (MECP2) 폴리펩타이드에서 시스테인을 아르기닌으로, 정지 코돈을 아르기닌으로 변화시킨다. 하나의 구현예에서, RETT와 연관된 SNP는 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래한다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 스트렙토코커스 피오게네스 Cas9 (SpCas9), 스타필로코커스 아우레우스 Cas9 (SaCas9), 스트렙토코커스 써모필러스 *1* Cas9 (St1Cas9), 스텝토코커스 카니스 Cas9 (ScCas9)로부터 선택된 Cas9 또는 이의 변이체이다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 변경된 프로토스페이서-인접 모티프 (PAM)에 결합하는 변형된 SpCas9를 포함한다. 하나의 구현예에서, 변형된 SpCas9는 5'-NGT-3' 또는 5'-NGG-3'로부터 선택된 핵산 서열을 포함하는 PAM에 결합한다. 하나의 구현예에서, 변형된 SpCas9는 NGT PAM 변이체에 결합한다. 하나의 구현예에서, NGT PAM 변이체는 변형된 SpCas9의 하나 이상의 잔기 1335, 1337, 1135, 1136, 1218, 및/또는 1219에서 아미노산 치환, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 변형된 SpCas9는 아미노산 치환 L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R, 및 L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 변형된 SpCas9는 아미노산 치환 D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, 및 T1337, 및 L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 뉴클레아제 불활성 또는 닉카제 변이체이다. 하나의 구현예에서, 닉카제 변이체는 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 데옥시리보핵산 (DNA)에서 아데노신을 탈아민화할 수 있다. 하나의 구현예에서, 아데노신 데아미나제 도메인은

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

의 아미노산 위치 82 및/또는 166에서 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제는 하기의 변경 중 하나 이상을 추가로 포함한다: Y147T, Y147R, Q154S, Y123H, 및 Q154R. 하나의 구현예에서, 아데노신 데아미나제 도메인은 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 그룹으로부터 선택된 변경의 조합을 포함한다. 하나의 구현예에서, ABE8은 아데노신 데아미나제 변이체 단량체를 포함하고, 여기서, 아데노신 데아미나제 단량체는 V82S 및 T166R 변경을 포함한다. 하나의 구현예에서, ABE8은 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체 단량체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함한다. 하나의 구현예에서, ABE8은 TadA*7.10 도메인 및 TadA*8 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체 단량체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함한다. 하나의 구현예에서, ABE8 염기 편집기는 TadA*7.10 도메인, 및 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 그룹으로부터 선택된 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 이종이량체를 포함한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함한다. 하나의 구현예에서, 아데노신 데아미나제는 TadA 데아미나제이다. 하나의 구현예에서, TadA 데아미나제는 TadA*8 변이체이다. 하나의 구현예에서, TadA*8 변이체는 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, 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, 및 TadA*8.24로 이루어진 그룹으로부터 선택된다. 하나의 구현예에서, ABE8 염기 편집기는 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, 및 ABE8.24로 이루어진 그룹으로부터 선택된다. 하나의 구현예에서, 하나 이상의 가이드 RNA는 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 여기서, crRNA는 RETT와 연관된 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함한다. 하나의 구현예에서, 염기 편집기는 RETT와 연관된 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성한다. 하나의 구현예에서, ABE8 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0016] 하나의 구현예에서, gRNA는 하기 서열을 갖는 스캐폴드를 포함한다:

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC
CGAGTCGGTGCTTTTTTT.

[0017] 또 다른 양상에서, 대상체에서 레트 증후군 (RETT)을 치료하는 방법이 제공되고, 여기서, 상기 방법은 상기 대상체에게 ABE8 염기 편집기, 또는 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드 (여기서, 상기 ABE8 염기 편집기는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함한다);

및 RETT와 연관된 MECP2 폴리뉴클레오타이드에서 SNP의 A·T에서 G·C로의 변경을 수행하도록 ABE8 염기 편집기를 표적화하는 하나 이상의 가이드 폴리뉴클레오타이드 (여기서, 상기 변경은 R133C 및/또는 R306C 중 하나 또는 둘 다이다)를 투여하는 단계를 포함한다. 하나의 구현예에서, 대상체는 포유동물 또는 인간이다. 하나의 구현예에서, 방법은 ABE8 염기 편집기, 또는 상기 ABE8 염기 편집기를 암호화하는 폴리뉴클레오타이드, 및 상기 하나 이상의 가이드 폴리뉴클레오타이드를 대상체의 세포에 전달하는 단계를 포함하고, 임의로 상기 세포는 뉴런이다. 방법의 하나의 구현예에서, RETT와 연관된 SNP에서 A·T에서 G·C로의 변경은 메틸 CpG 결합 단백질 2 (MECP2) 폴리펩타이드에서 시스테인을 아르기닌으로 또는 정지 코돈을 아르기닌으로 변화시킨다. 하나의 구현예에서, RETT와 연관된 SNP는 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래한다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 스트렙토코커스 피오게네스 Cas9 (SpCas9), 스타필로코커스 아우레우스 Cas9 (SaCas9), 스트렙토코커스 써모필러스 *1* Cas9 (St1Cas9), 스텝토코커스 카니스 Cas9 (ScCas9)로부터 선택된 Cas9 또는 이의 변이체이다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 변경된 프로토스페이서-인접 모티프 (PAM)에 결합하는 변형된 SpCas9를 포함한다. 하나의 구현예에서, 변형된 SpCas9는 5'-NGT-3' 또는 5'-NGG-3'로부터 선택된 핵산 서열을 포함하는 PAM에 결합한다. 하나의 구현예에서, 변형된 SpCas9는 NGT PAM 변이체에 결합한다. 하나의 구현예에서, NGT PAM 변이체는 변형된 SpCas9의 하나 이상의 잔기 1335, 1337, 1135, 1136, 1218, 및/또는 1219에서 아미노산 치환, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 변형된 SpCas9는 아미노산 치환 L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R, 및 L1111, D1135L, S1136R, G1218S, E1219V, D1332A, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 변형된 SpCas9는 아미노산 치환 D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, 및 T1337, 및 L1111R, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 뉴클레아제 불활성 또는 닉카제 변이체이다. 하나의 구현예에서, 닉카제 변이체는 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 데옥시리보핵산 (DNA)에서 아데노신을 탈아민화할 수 있다. 하나의 구현예에서, 아데노신 데아미나제 도메인은

[0018] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 위치 82 및/또는 166에서 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 아미노산 위치 82 및 166에서 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 V82S 변경, T166R 변경, 또는 V82S 변경과 T166R 변경 둘 다로부터 선택된 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 하기의 변경 중 하나 이상을 추가로 포함한다: Y147T, Y147R, Q154S, Y123H, 및 Q154R. 하나의 구현예에서, 아데노신 데아미나제 도메인은 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 그룹으로부터 선택된 변경을 포함한다. 하나의 구현예에서, ABE8은 아데노신 데아미나제 변이체 단량체를 포함하고, 여기서, 아데노신 데아미나제 단량체는 V82S 및 T166R 변경을 포함한다. 하나의 구현예에서, ABE8은 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체 단량체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함한다. 하나의 구현예에서, ABE8은 TadA*7.10 도메인 및 TadA*8 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체 단량체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함한다. 하나의 구현예에서, ABE8 염기 편집기는 TadA7.10 도메인, 및 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 그룹으로부터 선택된 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 이종이량체를 포함한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드

는 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 갖는다. 하나의 구현예에서, 아데노신 데아미나제는 TadA 데아미나제이다. 하나의 구현예에서, TadA 데아미나제는 TadA*8 변이체이다. 하나의 구현예에서, TadA*8 변이체는 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, 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, 및 TadA*8.24로 이루어진 그룹으로부터 선택된다. 하나의 구현예에서, ABE8 염기 편집기는 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, 및 ABE8.24로 이루어진 그룹으로부터 선택된다. 하나의 구현예에서, 하나 이상의 가이드 RNA는 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 여기서, crRNA는 RETT와 연관된 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함한다. 하나의 구현예에서, 염기 편집기는 RETT와 연관된 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성한다.

[0019] 또 다른 양상에서, 대상체에서 레트 증후군 (RETT)을 치료하는 방법이 제공되고, 여기서, 상기 방법은 이를 필요로 하는 대상체에게 (i) 아데노신 염기 편집기 또는 상기 아데노신 염기 편집기를 암호화하는 핵산 서열 및 (ii) 가이드 폴리뉴클레오타이드 또는 상기 가이드 폴리뉴클레오타이드를 암호화하는 핵산 서열을 투여하는 단계를 포함하고, 여기서, 아데노신 염기 편집기는 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 아데노신 데아미나제 도메인은 TadA 참조 서열에 비해 아미노산 위치 82 또는 166 또는 이의 상응하는 위치에서 아미노산 치환을 포함하고, 가이드 폴리뉴클레오타이드는 아데노신 염기 편집기가 대상체에서 RETT와 연관된 SNP를 포함하는 메틸 CpG 결합 단백질 2 (*MECP2*) 유전자 또는 이의 조절 요소에서 A에서 G로의 핵염기 변경을 수행하도록 지시하고, RETT와 연관된 SNP는 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 또는 R306C 아미노산 돌연변이를 초래한다. 방법의 하나의 구현예에서, 투여는 RETT와 연관된 적어도 하나의 증상을 개선한다. 방법의 하나의 구현예에서, 투여는 아데노신 데아미나제에서 아미노산 치환이 없는 염기 편집기를 사용한 치료와 비교하여 RETT와 관련된 적어도 하나의 증상의 더 빠른 개선을 초래한다. 방법의 하나의 구현예에서, A에서 G로의 핵염기 변경은 RETT와 연관된 SNP를 야생형 핵염기로 변화시킨다. 하나의 구현예에서, A에서 G로의 핵염기 변경은 레트 증후군과 연관된 SNP를 비야생형 핵염기로 변화시켜 개선된 RETT 증후군을 초래한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 RETT와 연관된 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함한다. 하나의 구현예에서, 아데노신 염기 편집기는 RETT와 연관된 SNP를 포함하는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함한다.

[0020] 편집 방법의 상기 설명된 양상 및 염기 편집기 시스템의 상기 설명된 양상의 구현예 및 이의 구현예에서, 가이드 폴리뉴클레오타이드는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 적어도 10개의 인접 뉴클레오타이드를 포함하는 핵산 서열을 포함한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 *MECP2* 유전자 또는 이의 조절 요소에 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 또는 40개의 인접 뉴클레오타이드를 포함하는 핵산 서열을 포함한다.

[0021] 편집 또는 치료 방법의 상기 설명된 양상 및 세포의 상기 설명된 양상의 구현예 및 이의 구현예에서, 가이드 폴리뉴클레오타이드는 *MECP2* 폴리뉴클레오타이드에 상보적인 적어도 10개의 인접 뉴클레오타이드를 포함하는 핵산 서열을 포함한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 *MECP2* 폴리뉴클레오타이드에 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 또는 40개의 인접 뉴클레오타이드를 포함하는 핵산 서열을 포함한다.

[0022] RETT를 치료하는 상기 설명된 방법의 구현예 및 이의 구현예에서, RETT와 연관된 SNP는 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 및/또는 R306C 아미노산 돌연변이를 초래한다. 하나의

구현예에서, RETT와 연관된 SNP는 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 아미노산 돌연변이를 초래한다. 하나의 구현예에서, RETT와 연관된 SNP는 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R306C 아미노산 돌연변이를 초래한다.

[0023] 상기 설명된 염기 편집기 시스템의 구현예 및 이의 구현예에서, RETT와 연관된 SNP에서 A에서 G로의 핵염기 변경은 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R133C 아미노산 돌연변이를 초래한다. 염기 편집기 시스템의 하나의 구현예에서, RETT와 연관된 SNP에서 A에서 G로의 핵염기 변경은 *MECP2* 유전자에 의해 암호화된 MECP2 폴리펩타이드 또는 이의 변이체에서 R306C 아미노산 돌연변이를 초래한다.

[0024] *MECP2* 폴리뉴클레오타이드를 편집하는 상기 설명된 방법 또는 레트 증후군 (RETT)을 치료하는 방법의 구현예에서 및 이의 구현예에서, RETT와 연관된 SNP의 변경은 R133C와 R306C 둘 다를 포함한다. 하나의 구현예에서, RETT와 연관된 SNP의 변경은 R133C이다. 하나의 구현예에서, RETT와 연관된 SNP의 변경은 R306C이다.

[0025] 상기 설명된 세포의 구현예 및 이의 구현예에서, 레트 증후군 (RETT)과 연관된 SNP의 변경은 R133C이다. 하나의 구현예에서, 레트 증후군 (RETT)과 연관된 SNP의 변경은 R306C이다.

[0026] 또 다른 양상에서, 가이드 폴리뉴클레오타이드 또는 가이드 RNA가 제공되고, 여기서, 가이드 폴리뉴클레오타이드 또는 가이드 RNA (gRNA)는 MECP2 단백질을 암호화하는 *MECP2* 유전자에 완벽하게 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개의 인접 뉴클레오타이드를 포함한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드 또는 가이드 RNA는 5'-AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3', 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드 또는 가이드 RNA는 스캐폴드 서열을 추가로 포함하고, 여기서, 스캐폴드 서열은 임의로 다음과 같다:

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC

[0027] CGAGTCGGTGCTTTTTTT.

[0028] 또 다른 양상에서, 아데노신 데아미나제 염기 편집기 8 (ABE8) 및 가이드 RNA를 포함하는 조성물이 제공되고, 여기서, 조성물에서, ABE8은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하고, 가이드 RNA는 레트 증후군과 연관된 *MECP2* 폴리뉴클레오타이드에서 SNP의 A·T에서 G·C로의 변경을 수행하도록 염기 편집기를 표적으로 하고, 변경은 R133C 또는 R306C 중 하나 또는 둘 다이다. 조성물의 하나의 구현예에서, RETT와 연관된 SNP에서 A·T에서 G·C로의 변경은 메틸 CpG 결합 단백질 2 (MECP2) 폴리펩타이드에서 시스테인을 아르기닌으로 또는 정지 코돈을 아르기닌으로 변화시킨다. 하나의 구현예에서, RETT와 연관된 SNP는 아미노산 위치 133 및/또는 306에서 아르기닌을 포함하는 MECP2 폴리펩타이드의 발현을 초래한다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 스트렙토코커스 피오게네스 Cas9 (SpCas9), 스타필로코커스 아우레우스 Cas9 (SaCas9), 스트렙토코커스 써모필러스 *1* Cas9 (St1Cas9), 스텝토코커스 카니스 Cas9 (ScCas9)로부터 선택된 Cas9, 또는 이들의 변이체이다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 변경된 프로토스페이서-인접 모티프 (PAM)에 결합하는 변형된 SpCas9를 포함한다. 하나의 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 뉴클레아제 불활성 또는 닉카제 변이체이다. 하나의 구현예에서, 닉카제 변이체는 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 데옥시리보핵산 (DNA)에서 아데노신을 탈아민화할 수 있다. 조성물의 하나의 구현예에서, 아데노신 데아미나제 도메인은

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH

AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG

[0029] SLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 위치 82 및/또는 166에서 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 V82S 변경, T166R 변경, 또는 V82S 변경과 T166R 변경 둘 다로부터 선택된 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 도메인은 하기의 변경 중 하나 이상을 추가로 포함한다: Y147T, Y147R, Q154S, Y123H, 및 Q154R. 하나의 구현예에서, 아데노신 데아미나제 도메인은 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 그룹으로

부터 선택된 변경을 포함한다. 하나의 구현예에서, ABE8은 아데노신 데아미나제 변이체 단량체를 포함하고, 여기서, 아데노신 데아미나제 단량체는 V82S 및 T166R 변경을 포함한다. 하나의 구현예에서, ABE8은 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체 단량체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함한다. 하나의 구현예에서, ABE8은 TadA*8 도메인 및 야생형 TadA 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체 단량체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R로 이루어진 그룹으로부터 선택된 하나 이상의 변경을 추가로 포함한다. 하나의 구현예에서, ABE8 염기 편집기는 야생형 TadA 도메인, 및 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 그룹으로부터 선택된 변경의 조합을 포함하는 아데노신 데아미나제 변이체를 포함하는 이종이량체를 포함한다. 하나의 구현예에서, 가이드 RNA는 AGAGCAAAAGGCUUUUCCCU-3', 5'-UAGAGCAAAAGGCUUUUCCC-3', 5'-UAGAGCAAAAGGCUUUUCCCU-3', 5'-UUUAGAGCAAAAGGCUUUUCCCU-3', 5'-UCUUGCACUUCUUGAUGGGG-3', 5'-CUUGCACUUCUUGAUGGGGAG-3' 또는 5'-GUCUUGCACUUCUUGAUGGGGAG-3'로부터 선택된 핵산 서열을 포함한다. 하나의 구현예에서, 아데노신 데아미나제는 TadA 데아미나제이다. 하나의 구현예에서, TadA 데아미나제는 TadA*8 변이체이다. 하나의 구현예에서, TadA*8 변이체는 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, 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, 및 TadA*8.24로 이루어진 그룹으로부터 선택된다. 하나의 구현예에서, ABE8 염기 편집기는 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, 및 ABE8.24로 이루어진 그룹으로부터 선택된다. 조성물의 하나의 구현예에서, 가이드 RNA는 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 여기서, crRNA는 RETT와 연관된 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함한다. 하나의 구현예에서, ABE8 염기 편집기는 RETT와 연관된 SNP를 포함하는 *MECP2* 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체를 형성한다. 하나의 구현예에서, ABE8 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAG
SLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0030] 하나의 구현예에서, 조성물은 지질을 추가로 포함하고, 임의로 지질이 양이온성 지질이다.

[0031] 조성물의 상기 설명된 양상의 구현예 및 이의 구현예에서, 조성물은 약제학적으로 허용되는 부형제 또는 희석제를 포함하는 약제학적 조성물이다. 하나의 구현예에서, 약제학적 조성물은 레트 증후군의 치료를 위한 것이다. 약제학적 조성물의 하나의 구현예에서, gRNA 및 ABE8 염기 편집기는 함께 또는 별도로 제형화된다. 하나의 구현예에서, 약제학적 조성물은 포유류 세포에서의 발현에 적합한 벡터를 추가로 포함하고, 여기서, 벡터는 ABE8 염기 편집기를 암호화하는 폴리뉴클레오타이드를 포함한다. 약제학적 조성물의 하나의 구현예에서, 벡터는 바이러스 벡터이다. 약제학적 조성물의 하나의 구현예에서, 바이러스 벡터는 레트로바이러스 벡터, 아데노바이러스 벡터, 렌티바이러스 벡터, 헤르페스바이러스 벡터, 또는 아데노-연관된 바이러스 벡터 (AAV)이다. 하나의 구현예에서, 약제학적 조성물은 포유류 세포에서의 발현에 적합한 리보핵입자를 추가로 포함한다.

[0032] 또 다른 양상에서, 레트 증후군을 치료하는 방법이 제공되고, 여기서, 상기 방법은 이를 필요로 하는 대상체에게 임의의 상기 설명된 양상 및 구현예에 기재된 바와 같은 약제학적 조성물을 투여하는 단계를 포함한다. 상기 방법의 하나의 구현예에서, 대상체는 포유동물 또는 인간이다.

[0033] 또 다른 양상에서, 대상체에서 레트 증후군의 치료에서 임의의 상기 설명된 양상 및 구현예에 기재된 바와 같은 약제학적 조성물의 용도가 제공된다. 용도의 하나의 구현예에서, 대상체는 포유동물 또는 인간이다.

[0034] 하나의 양상에서, 임의의 상기 설명된 양상 및 구현예에 기재된 바와 같은 세포를 포함하는 조성물이 제공된다. 하나의 구현예에서, 조성물은 약제학적으로 허용되는 담체 또는 희석제를 추가로 포함한다.

[0035] 또 다른 양상에서, (i) ABE8 염기 편집기를 암호화하는 핵산; 및 (ii) 상기 설명된 양상 및 구현예에 기재된 바와 같은 가이드 폴리뉴클레오타이드 또는 가이드 RNA를 포함하는 약제학적 조성물이 제공된다. 하나의 구현예에서, 약제학적 조성물은 지질을 추가로 포함한다. 하나의 구현예에서, 지질은 양이온성 지질이다. 약제학적 조성물의 하나의 구현예에서, 염기 편집기를 암호화하는 핵산은 mRNA이다.

[0036] 또 다른 양상에서, ABE8 염기 편집기가 제공되고, 여기서, ABE8 염기 편집기는 (i) 아미노산 치환 L1111R, D1135V, G1218R, E1219F, A1322R, R1335V, T1337R, 및 L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함하는 변형된 SpCas9; 및 (ii) TadA*8 아데노신 데아미나제를 포함한다.

[0037] 또 다른 양상에서, ABE8 염기 편집기가 제공되고, 여기서, ABE8 염기 편집기는 (i) 아미노산 치환 D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, 및 T1337, 및 L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, 및 T1337M 중 하나 이상, 또는 이의 상응하는 아미노산 치환을 포함하는 변형된 SpCas9; 및 (ii) TadA*8 데아미나제를 포함한다.

도면의 간단한 설명

[0038] 본원 개시내용의 특징은 첨부된 청구범위에서 구체적으로 설명된다. 본 발명의 특징 및 이점에 대한 더 나은 이해는 본원 개시내용의 원리가 활용되는 예시적인 구현예를 설명하는 다음의 상세한 설명 및 첨부 도면을 참조하여 얻어질 것이다:

도 1은 RTT 돌연변이에 기반한 운동 행동 평가를 묘사하는 그래프이다.

도 2는 MECP2 단백질의 각 도메인의 기능과 공통 RTT 돌연변이의 위치를 도시한다.

도 3은 gRNA1, gRNA2, 및 gRNA5를 사용하는 BE8 염기 편집기 변이체를 사용하여 R106W RTT 돌연변이의 정밀 교정의 백분율을 도시하는 그래프이다. 각 gRNA와 쌍을 이루는 각 염기 편집기에 대한 결과는 왼쪽에서 오른쪽으로 ABE8.14 (가장 왼쪽)에서 Neg. Ct (가장 오른쪽)로 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0039] 본 발명은 프로그래밍 가능한 핵염기 편집기 (예를 들어, ABE8)를 사용하여 RTT와 연관된 병원성 아미노산의 정밀 교정을 위한 조성물 방법을 특징으로 한다.

[0040] 본 발명은, 적어도 부분적으로, 아데노신 데아미나제 변이체 (본원에서 아데노신 염기 편집기 8 또는 "ABE8"로 칭함)를 특징으로 하는 염기 편집기가 내인성 *Mecp2* 유전자 (예를 들어, R106W, R133C, T158M, R255*, R270*, R306C)에서 단일염기다형성 (SNP)를 정밀하게 교정한다는 발견에 기반한다. 하나의 구현예에서, *Mecp2* 유전자에서의 SNP는 R133C이다. 하나의 구현예에서, *Mecp2* 유전자에서의 SNP는 R306C이다.

[0041] 본원에는 진행성 신경발달 장애 레트 증후군 (RTT 또는 RETT) 및 이의 증상과 인과적으로 관련된 메틸-CpG-결합 단백질 2 (*Mecp2*) 유전자에서 하나 이상의 돌연변이를 정밀하게 교정하기 위한 염기 편집 및 염기 편집 시스템을 제공하는 조성물 및 방법이 기재되어 있다. RTT는 주로 여성에게 영향을 미치는 X-연관된 우성 장애로, *Mecp2* 유전자에 돌연변이가 있는 영향을 받는 개체의 96%와 연관이 있으며, 겉보기에 정상적인 초기 발달에 이어 미세한 운동 능력과 효과적인 의사소통의 상실, 고정관념적 움직임, 및 실행증 (apraxia) 또는 보행의 완전한 부재를 동반한 퇴행이 특징이다 (도 1 참조). 고통받는 개체의 추가 임상 특징은 머리 성장 속도의 비정상적인 출생 후 감속, 주기적인 호흡, 위장 기능 장애, 간질, 및 척추측만증을 포함한다.

[0042] 가장 만연한 RTT-유발 돌연변이는 시티딘에서 티미딘으로의 (C->T) 전이 돌연변이이며, 이는 C·G에서 T·A 염기 쌍 치환을 초래한다. 상기 치환은 A·T에서 G·C로의 치환을 촉매하는 아데노신 염기 편집기 (ABE)를 사용하여 야생형, 비-병원성 게놈 서열로 다시 되돌릴 수 있다. 더 나아가, 매우 만연한 RTT-유발 돌연변이는 유전자 요법을 사용하여 발생할 수 있는 *Mecp2* 유전자 과발현을 유도할 위험 없이 ABE를 사용하여 야생형 서열로 복귀하기 위한 잠재적인 표적이다. 따라서, A·T에서 G·C로의 DNA 염기 편집은 *Mecp2* 유전자에서 가장 만연한

RTT-유발 돌연변이 중 하나 이상을 정밀하게 교정할 가능성이 있다.

[0043] 하기 기재 사항 및 실시예는 본원 개시내용의 구현예를 예시한다. 본원 개시내용은 본원에 기재된 특정 구현예로 제한되지 않고 이와 같이 다양할 수 있는 것으로 이해되어야만 한다. 당업자는 이의 범위에 포함되는 본원 개시내용의 다수의 변화 및 변형이 있음을 인지할 것이다.

[0044] 본원에 사용된 섹션 제목은 구성 목적만을 위한 것이고 기재된 주제를 제한하는 것으로 해석되어서는 안된다.

[0045] 본원 개시내용의 다양한 특성이 단일 구현예와 관련하여 기재될 수 있지만, 특성은 또한 별도로 또는 임의의 적합한 조합으로 제공될 수 있다. 역으로, 본원의 개시내용은 명확하게 하기 위해 별도의 구현예와 관련하여 본원에 기재될 수 있지만, 본원의 개시내용은 또한 단일 구현예로 수행될 수 있다.

[0046] **정의**

[0047] 다음 정의는 해당 기술 분야의 정의를 보완하고 현재 출원에 대한 것이며 관련되거나 관련되지 않은 사례, 예를 들어, 공동 소유의 특허 또는 출원에 귀속되어서는 안된다. 본원에 기재된 것과 유사하거나 또는 균등한 임의의 방법과 물질들은 본원 개시내용의 시험을 위한 실시에 사용할 수 있지만, 바람직한 물질 및 방법들을 본원에 기재되어 있다. 따라서, 본원에 사용된 용어는 단지 특정 구현예를 기재하기 위한 것이고 본 발명을 제한하는 것으로 의도되지 않는다.

[0048] 달리 정의되지 않는 경우, 본원 개시내용 및 이의 구현예에서 사용된 모든 기술적 및 과학적 용어는 관련 기술 분야의 숙련가에 의해 통상적으로 이해되는 의미를 갖는다. 하기의 참조문헌은 당업자에게 본원 개시내용에 사용된 많은 용어의 일반적인 정의를 제공한다: Singleton *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991).

[0049] 본원에서, 단수의 사용은 구체적으로 달리 언급되지 않는 경우 복수를 포함한다. 본 명세서에 사용된 바와 같이, 단수 형태 "a," "an," 및 "the"는 문맥에 달리 명백하게 지적되지 않는 경우 복수의 언급을 포함한다. 본원에서, "또는"의 사용은 달리 언급되지 않는 경우 "및/또는"을 의미하고 포괄적인 것으로 이해된다. 추가로, "포함한다 (include)", "포함한다 (includes)" 및 "포함되는 (included)"과 같은 다른 형태뿐만 아니라 "포함하는"이라는 용어의 사용은 제한적이지 않다.

[0050] 본 명세서 및 청구항(들)에 사용된 바와 같은, 용어 "포함하는" (및 포함하는의 임의의 형태, 예를 들어, "포함한다 (comprise)" 및 "포함한다 (comprises)"), "갖는 (having)" (및 갖는의 임의의 형태, 예를 들어, "갖는다 (have)" 및 "갖는다 (has)"), "포함하는 (including)" (및 포함하는의 임의의 형태, 예를 들어, "포함한다 (includes)" 및 "포함한다 (include)") 또는 "함유하는 (containing)" (및 함유하는의 임의의 형태, 예를 들어 "함유한다 (contains)" 및 "함유한다 (contain)")는 포괄적이거나 개방형 (open-ended)이고 추가의 언급되지 않은 요소 또는 방법 단계를 배제하지 않는다. 본 명세서에 논의된 임의의 구현예는 본원 개시내용의 임의의 방법 또는 조성물과 관련하여 구현될 수 있고 그 반대도 마찬가지인 것으로 고려된다. 추가로, 본원 개시내용의 조성물을 사용하여 본원 개시내용의 방법을 성취할 수 있다.

[0051] 용어 "약" 또는 "대략적으로"는 당업자에 의한 결정시 특정 값에 대해 허용 가능한 오차 범위 내를 의미하고, 이것은 상기 값이 어떻게 측정되거나 결정되는지, 즉, 측정 시스템의 한계치에 부분적으로 의존한다. 예를 들어, "약"은 당업계의 관행에 따라 하나 (1) 또는 하나 (1) 초과의 표준 편차 내를 의미할 수 있다. 대안적으로, "약"은 소정의 값의 최대 20%, 최대 10%, 최대 5%, 또는 최대 1%의 범위를 의미할 수 있다. 대안적으로, 특히, 생물학적 시스템 또는 공정과 관련하여, 용어는 한자릿수 (order of magnitude) 이내, 바람직하게는 5배 이내, 보다 바람직하게는 2배 이내를 의미할 수 있다. 특정 값이 본원 및 청구항에 기재되는 경우, 달리 언급되지 않는 한, 용어 "약"은 특정 값에 대해 허용되는 오차 범위 내에 있다고 가정해야 함을 의미한다.

[0052] 본원에 제공된 범위는 상기 범위 내 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 그룹으로부터 임의의 수, 수의 조합, 또는 서브-범위를 포함하는 것으로 이해된다.

[0053] 명세서에서 "일부 구현예", "구현예", "하나의 구현예" 또는 "다른 구현예"에 대한 언급은 구현예와 연계하여

기재된 특정 특성, 구조 또는 특징이 적어도 일부 구현예에 포함되지만 본원 개시내용의 모든 구현예에 필수적으로 포함되는 것은 아님을 의미한다.

[0054] "아데노신 데아미나제"란 아데닌 또는 아데노신의 가수분해 탈아민화를 촉매할 수 있는 폴리펩타이드 또는 이의 단편을 의미한다. 일부 구현예에서, 데아미나제 또는 데아미나제 도메인은 아데노신의 이노신으로 또는 데옥시아데노신의 데옥시이노신으로의 가수분해 탈아민화를 촉매하는 아데노신 데아미나제이다. 일부 구현예에서, 아데노신 데아미나제는 데옥시리보핵산 (DNA)에서 아데닌 또는 아데노신의 가수분해 탈아민화를 촉매한다. 본원에 제공된 아데노신 데아미나제 (예를 들어, 가공된 아데노신 데아미나제, 변화된 아데노신 데아미나제)는 임의의 유기체, 예를 들어, 세균으로부터 기원할 수 있다.

[0055] 일부 구현예에서, 데아미나제 또는 데아미나제 도메인은 인간, 침팬지, 고릴라, 몽키, 소, 개, 래트 또는 마우스와 같은 유기체 기원의 천연적으로 발생하는 데아미나제의 변이체이다. 일부 구현예에서, 데아미나제 또는 데아미나제 도메인은 자연에서 발생하지 않는다. 예를 들어, 일부 구현예에서, 데아미나제 또는 데아미나제 도메인은 천연적으로 발생하는 데아미나제와 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.

[0056] 일부 구현예에서, 아데노신 데아미나제는 TadA 데아미나제이다. 일부 구현예에서, TadA 데아미나제는 TadA 변이체이다. 일부 구현예에서, TadA 변이체는 TadA*8이다. 야생형 TadA(wt) 아데노신 데아미나제는 하기의 서열 (또한 TadA 참조 서열로 칭함)을 갖는다:

[0057] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD.

[0058] 일부 구현예에서, 아데노신 데아미나제는 하기의 서열에서의 변경을 포함한다:

[0059] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0060] (또한 TadA*7.10으로 칭함).

[0061] 일부 구현예에서, TadA*7.10은 적어도 하나의 변경을 포함한다. 일부 구현예에서, TadA*7.10은 아미노산 82 및/또는 166에서의 변경을 포함한다. 특정 구현예에서, 상기 언급된 서열의 변이체는 하기의 변경 중 하나 이상을 포함한다: Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R. 다른 구현예에서, TadA*7.10 서열의 변이체는 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; 또는 I76Y + V82S + Y123H + Y147R + Q154R로부터 선택된 변경의 조합을 포함한다.

[0062] 다른 구현예에서, TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 잔기 149, 150, 151, 152, 153, 154, 155, 156, 또는 157에서 시작하는 C 말단의 결실을 포함하는, 결실을 포함하는 데아미나제 변이체가 제공된다. 다른 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 하기의 변경 중 하나 이상을 포함하는 TadA (예를 들어, TadA*8) 단량체이다: Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R. 다른 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 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; 또는 I76Y + V82S + Y123H + Y147R + Q154R로부터 선택된 변경의 조합을 포함하는 TadA (예를 들어, TadA*8) 단량체이다.

[0063] 여전히 다른 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 2개의 아데노신 데아미나제 도메인 (예를 들어, TadA*8)을 포함하는 동종이량체이고

이의 각각은 하기의 변경 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R 중 하나 이상을 갖는다. 다른 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 2개의 아데노신 데아미나제 도메인 (예를 들어, TadA*8)을 포함하는 동종이량체이고, 이의 각각은 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; 또는 I76Y + V82S + Y123H + Y147R + Q154R로부터 선택된 변경의 조합을 갖는다.

[0064] 다른 구현예에서, 아데노신 데아미나제 변이체는 야생형 TadA 아데노신 데아미나제 도메인, 및 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 하기의 변경 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R 중 하나 이상을 포함하는 아데노신 데아미나제 변이체 도메인 (예를 들어, TadA*8)을 포함하는 이종이량체이다. 다른 구현예에서, 아데노신 데아미나제 변이체는 야생형 TadA 아데노신 데아미나제 도메인, 및 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 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 + Q154R의 그룹으로부터 선택된 변경의 조합을 포함하는 아데노신 데아미나제 변이체 도메인 (예를 들어, TadA*8)을 포함하는 이종이량체이다.

[0065] 다른 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10 도메인, 및 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 하기의 변경 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R 중 하나 이상의 포함하는 아데노신 데아미나제 변이체 도메인 (예를 들어, TadA*8)을 포함하는 이종이량체이다. 다른 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10 도메인, 및 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 하기의 변경의 조합을 포함하는 아데노신 데아미나제 변이체 도메인 (예를 들어, TadA*8)을 포함하는 이종이량체이다: 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.

[0066] 하나의 구현예에서, 아데노신 데아미나제는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 TadA*8이다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV

[0067] EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0068] 일부 구현예에서, TadA*8은 절단된다. 일부 구현예에서, 절단된 TadA*8은 전장 TadA*8에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 N-말단 아미노산 잔기가 결실된다. 일부 구현예에서, 절단된 TadA*8은 전장 TadA*8에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 C-말단 아미노산 잔기가 결실된다. 일부 구현예에서, 아데노신 데아미나제 변이체는 전장 TadA*8이다.

[0069] 특정 구현예에서, 아데노신 데아미나제 이종이량체는 TadA*8 도메인 및 다음 중 하나로부터 선택된 아데노신 데아미나제 도메인을 포함한다:

[0070] 스타필로코커스 아우레우스 (에스. 아우레우스 (*S. aureus*)) TadA:

MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAH
AEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGS

[0071] LMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[0072] 바실러스 서브틸리스 (*Bacillus subtilis*) (비. 서브틸리스 (*B. subtilis*)) TadA:

MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEML
VIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMN

[0073] LLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[0074] 살모넬라 타이피무리움 (*Salmonella typhimurium*) (에스. 티피무리움 (*S. typhimurium*)) TadA:

MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEG
WNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIG
RVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIK

[0075] ALKKADRAEGAGPAV

[0076] 쉐와넬라 푸트레파시엔스 (*Shewanella putrefaciens*) (에스. 푸트레파시엔스 (*S. putrefaciens*)) TadA:

MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEI
LCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGT

[0077] VVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[0078] 해모필러스 인플루엔자 (*Haemophilus influenzae*) F3031 (에이취. 인플루엔자 (*H. influenzae*)) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAH
AEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYK

[0079] TGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[0080] 콜로박터 크레슨투스 (*Caulobacter crescentus*) (씨. 크레슨투스 (*C. crescentus*)) TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAH
DPTAHAEIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADD

[0081] PKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

[0082] 게오박터 설푸레두센스 (*Geobacter sulfurreducens*) (지. 설푸레두센스 (*G. sulfurreducens*)) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSN
DPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDP
KGGAAGSLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALF

[0083] IDERKVPPEP

[0084] TadA*7.10

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV

[0085] EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0086] "아데노신 데아미나제 염기 편집기 8 (ABE8) 폴리펩타이드 또는 "ABE8"이란 하기의 참조 서열의 아미노산 위치 82 및/또는 166에서 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는, 본원에 정의된 바와 같은 염기 편집기를 의미한다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV

[0087] EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0088] 일부 구현예에서, ABE8은 참조 서열에 비해 본원에 기재된 바와 같은 추가의 변경을 포함한다.

[0089] "아데노신 데아미나제 염기 편집기 8 (ABE8) 폴리뉴클레오타이드"란 ABE8을 암호화하는 폴리뉴클레오타이드를 의미한다.

[0090] "투여하는"은 본원에 기재된 하나 이상의 조성물을 환자 또는 대상체에게 제공하는 것으로서 본원에 언급된다. 예를 들어 그리고 제한 없이, 조성물 투여, 예를 들어, 주사는 정맥내 (i.v.) 주사, 피하 (s.c.) 주사, 피내 (i.d.) 주사, 복막내 (i.p.) 주사, 또는 근육내 (i.m.) 주사에 의해 수행될 수 있다. 하나 이상의 상기 경로가 사용될 수 있다. 비경구 투여는 예를 들어, 볼러스 주사에 의해 또는 시간 경과에 따른 점진적 관류에 의한 것일 수 있다. 일부 구현예에서, 비경구 투여는 혈관내, 정맥내, 근육내, 동맥내, 척수강내, 종양내, 피내, 복강내, 경기관 (transtracheally), 피하 (subcutaneously), 피하 (subcuticularly), 관절내, 피막하 (subcapsularly), 지주막하 및 흉골내 주입 또는 주사를 포함한다. 대안적으로, 또는 동시에, 투여는 경구 경로에 의한 것일 수 있다.

[0091] "제제"란 임의의 소분자의 화학적 화합물, 항체, 핵산 분자 또는 폴리펩타이드 또는 이의 단편을 의미한다.

[0092] "변경"이란 본원에 기재된 것들과 같이 표준 당업계 공지된 방법에 의해 검출된 바와 같은 유전자 또는 폴리펩타이드의 발현 수준 또는 활성에서의 변화 (증가 또는 감소)를 의미한다. 본원에 사용된 바와 같은 변경은 발현 수준의 10% 변화, 바람직하게는 25% 변화, 보다 바람직하게는 40% 변화, 가장 바람직하게는 발현 수준의 50% 이상의 변화를 포함한다. "개선한다"란 질환의 발병 또는 진행의 감소, 억제, 감쇠, 약화, 정지 또는 안정화를 의미한다.

[0093] "유사체"란 동일하지 않지만 유사한 기능 또는 구조적 특성을 갖는 분자를 의미한다. 예를 들어, 폴리펩타이드 유사체는 상응하는 천연적으로 발생하는 폴리펩타이드의 생물학적 활성을 보유하면서 천연적으로 발생하는 폴리펩타이드에 비해 유사체의 기능을 증진시키는 특정 생화학적 변형을 갖는다. 상기 생화학적 변형은 예를 들어, 리간드 결합을 변경하지 않고 유사체의 프로테아제 내성, 막 투과성 및/또는 반감기를 증가시킬 수 있다. 유사체는 비천연 아미노산을 포함할 수 있다.

[0094] "메틸 CpG 결합 단백질 2 (Mecp2) 단백질"이란 NCBI 수탁 번호 NP_004983에 대해 적어도 약 95% 아미노산 서열 동일성을 갖는 폴리펩타이드 또는 이의 단편을 의미한다. 특정 구현예에서, Mecp2 단백질은 하기의 참조 서열에 비해 하나 이상의 변경을 포함한다. 특정 구현예에서, RTT와 연관된 Mecp2 단백질은 R106W, R168*, R133C, T158M, R255*, R270*, 및 R306C로부터 선택된 하나 이상의 돌연변이를 포함한다. 예시적인 Mecp2 아미노산 서열은 아래에 제공된다.

```
  1 mvagmlglre eksedqdlqg lkdkplkfkk vkkdkkeeke gkhepvqpsa hhsaepaeag
 61 kaetsegsgs apavpeasas pkqrrsiird rgpmyddptl pegwtrklkq rksgrsagky
121 dvylinpqgk afrskvelia yfekvgdtsl dpndfdftvt grgspsrreq kppkkpkspk
181 apgtgrgrgr pkgsgttrpk aatsegvqvk rvlekspgkl lvkmpfqtsp ggkaegggat
241 tstqvmvikr pgrkrkaead pqaipkkrgr kpgsvvaaaa aeakkkavke ssirsvqetv
301 lpikkrktre tvsievkevv kpllvstlge ksgkglktck spgrkskess pkgrsssass
361 ppkkehhhhh hhsespkapv pllpplpppp pepessedpt sppepqdlss svckeekmpr
421 ggslesdgcp kepaktqpav ataataaeky khrgegerkd ivsssmprpn reepvdsrtp
481 vtervs
```

[0095] "*Mecp2* 폴리뉴클레오타이드"란 Mecp2 단백질 또는 이의 단편을 암호화하는 핵산 분자를 의미한다. NCBI 수탁 번호 NM_004992에서 입수 가능한 예시적인 *Mecp2* 폴리뉴클레오타이드의 서열은 아래에 제공된다. 특정 구현예에서, *Mecp2* 폴리뉴클레오타이드는 하기의 참조 서열에 비해 하나 이상의 변경을 포함한다. 특정 구현예에서, RTT와 연관된 *Mecp2* 폴리뉴클레오타이드는 316C>T, 397C>T, 473C>T, 763C>T, 808C>T 및 916C>T로부터 선택된 하나 이상의 돌연변이를 포함한다.

[0096]

```
   1 ccggcgtcgg cggcgcgcgc gctccctcct ctcggagaga gggctgtggt aaaagccgtc
  61 cggaaaatgg ccgccgccgc cgccgccgcg ccgagcggag gaggaggagg aggcgaggag
 121 gagagactgc tccataaaaa tacagactca ccagttcctg ctttgatgtg acatgtgact
 181 ccccagaata caccttgctt ctgtagacca gctccaacag gattccatgg tagctgggat
 241 gttagggctc agggaagaaa agtcagaaga ccaggacctc cagggcctca aggacaaacc
 301 cctcaagttt aaaaaggtga agaaagataa gaaagaagag aaagagggca agcatgagcc
 361 cgtgcagcca tcagcccacc actctgctga gcccgcagag gcaggcaaag cagagacatc
 421 agaagggtca ggctccgccc cggctgtgcc ggaagcttct gcctccccca aacagcggcg

 481 ctccatcatc cgtgaccggg gacccatgta tgatgacccc accctgcctg aaggctggac
 541 acggaagctt aagcaaagga aatctggccg ctctgctggg aagtatgatg tgtatttgat
 601 caatccccag ggaaaagcct ttcgctctaa agtggagttg attgcgtact tcgaaaaggt
 661 aggcgacaca tccctggacc ctaatgattt tgacttcacg gtaactggga gagggagccc
 721 ctcccggcga gagcagaaac cacctaagaa gcccaaatct cccaaagctc caggaactgg
 781 cagaggccgg ggacgcccca aagggagcgg caccacgaga cccaaggcgg ccacgtcaga
 841 gggtgtgcag gtgaaaaggg tcctggagaa aagtcctggg aagctccttg tcaagatgcc
 901 ttttcaaact tcgccagggg gcaaggctga ggggggtggg gccaccacat ccacccaggt
 961 catggtgatc aaacgccccg gcaggaagcg aaaagctgag gccgaccctc aggccattcc
1021 caagaaacgg ggccgaaagc cggggagtgt ggtggcagcc gctgccgccg aggccaaaaa
1081 gaaagccgtg aaggagtctt ctatccgatc tgtgcaggag accgtactcc ccatcaagaa
1141 gcgcaagacc cgggagacgg tcagcatcga ggtcaaggaa gtggtgaagc ccctgctggt
1201 gtccaccctc ggtgagaaga gcgggaaagg actgaagacc tgtaagagcc ctgggcggaa
1261 aagcaaggag agcagcccca aggggcgcag cagcagcgcc tcctcacccc ccaagaagga
1321 gcaccaccac catcaccacc actcagagtc cccaaaggcc cccgtgccac tgctcccacc
1381 cctgccccca cctccacctg agcccgagag ctccgaggac cccaccagcc cccctgagcc
1441 ccaggacttg agcagcagcg tctgcaaaga ggagaagatg cccagaggag gctcactgga
1501 gagcgacggc tgccccaagg agccagctaa gactcagccc gcggttgcca ccgccgccac
1561 ggccgcagaa aagtacaaac accgagggga gggagagcgc aaagacattg tttcatcctc
1621 catgccaagg ccaaacagag aggagcctgt ggacagccgg acgcccgtga ccgagagagt
1681 tagctgactt tacacggagc ggattgcaaa gcaaaccaac aagaataaag gcagctgttg
1741 tctcttctcc ttatgggtag ggctctgaca aagcttcccg attaactgaa ataaaaaata
1801 ttttttttc tttcagtaaa cttagagttt cgtggcttca gggtgggagt agttggagca
1861 ttggggatgt ttttcttacc gacaagcaca gtcaggttga agacctaacc agggccagaa
1921 gtagctttgc acttttctaa actaggctcc ttcaacaagg cttgctgcag atactactga
1981 ccagacaagc tgttgaccag gcacctcccc tcccgcccaa acctttcccc catgtggtcg
2041 ttagagacag agcgacagag cagttgagag gacactcccg ttttcggtgc catcagtgcc
2101 ccgtctacag ctcccccagc tccccccacc tcccccactc ccaaccacgt tgggacaggg
2161 aggtgtgagg caggagagac agttggattc tttagagaag atggatatga ccagtggcta
2221 tggcctgtgc gatcccaccc gtggtggctc aagtctggcc ccacaccagc cccaatccaa
2281 aactggcaag gacgcttcac aggacaggaa agtggcacct gtctgctcca gctctggcat
2341 ggctaggagg ggggagtccc ttgaactact gggtgtagac tggcctgaac cacaggagag
2401 gatggcccag ggtgaggtgg catggtccat tctcaaggga cgtcctccaa cgggtggcgc
2461 tagaggccat ggaggcagta ggacaaggtg caggcaggct ggcctggggt caggccgggc
2521 agagcacagc ggggtgagag ggattcctaa tcactcagag cagtctgtga cttagtggac
2581 aggggagggg gcaaaggggg aggagaagaa aatgttcttc cagttacttt ccaattctcc
2641 tttagggaca gcttagaatt atttgcacta ttgagtcttc atgttcccac ttcaaaacaa
2701 acagatgctc tgagagcaaa ctggcttgaa ttggtgacat ttagtccctc aagccaccag
2761 atgtgacagt gttgagaact acctggattt gtatatatac ctgcgcttgt tttaaagtgg
2821 gctcagcaca tagggttccc acgaagctcc gaaactctaa gtgtttgctg caattttata
```

[0097]

2881 aggacttcct gattggtttc tcttctcccc ttccatttct gccttttgtt catttcatcc
2941 tttcacttct ttcccttcct ccgtcctcct ccttcctagt tcatcccttc tcttccaggc
3001 agccgcggtg cccaaccaca cttgtcggct ccagtcccca gaactctgcc tgccctttgt
3061 cctcctgctg ccagtaccag ccccaccctg ttttgagccc tgaggaggcc ttgggctctg
3121 ctgagtccga cctggcctgt ctgtgaagag caagagagca gcaaggtctt gctctcctag
3181 gtagccccct cttccctggt aagaaaaagc aaaaggcatt tcccaccctg aacaacgagc
3241 cttttcaccc ttctactcta gagaagtgga ctggaggagc tgggcccgat ttggtagttg
3301 aggaaagcac agaggcctcc tgtggcctgc cagtcatcga gtggcccaac aggggctcca
3361 tgccagccga ccttgacctc actcagaagt ccagagtcta gcgtagtgca gcagggcagt
3421 agcggtacca atgcagaact cccaagaccc gagctgggac cagtacctgg gtccccagcc
3481 cttcctctgc tccccctttt ccctcggagt tcttcttgaa tggcaatgtt ttgcttttgc
3541 tcgatgcaga caggggggca gaacaccaca catttcactg tctgtctggt ccatagctgt
3601 ggtgtagggg cttagaggca tgggcttgct gtgggttttt aattgatcag ttttcatgtg
3661 ggatcccatc tttttaacct ctgttcagga agtccttatc tagctgcata tcttcatcat
3721 attggtatat ccttttctgt gtttacagag atgtctctta tatctaaatc tgtccaactg
3781 agaagtacct tatcaaagta gcaaatgaga cagcagtctt atgcttccag aaacacccac
3841 aggcatgtcc catgtgagct gctgccatga actgtcaagt gtgtgttgtc ttgtgtattt
3901 cagttattgt ccctggcttc cttactatgg tgtaatcatg aaggagtgaa acatcataga
3961 aactgtctag cacttccttg ccagtcttta gtgatcagga accatagttg acagttccaa
4021 tcagtagctt aagaaaaaac cgtgtttgtc tcttctggaa tggttagaag tgagggagtt
4081 tgccccgttc tgtttgtaga gtctcatagt tggactttct agcatatatg tgtccatttc
4141 cttatgctgt aaaagcaagt cctgcaacca aactcccatc agcccaatcc ctgatccctg
4201 atcccttcca cctgctctgc tgatgacccc cccagcttca cttctgactc ttccccagga
4261 agggaagggg ggtcagaaga gagggtgagt cctccagaac tcttcctcca aggacagaag
4321 gctcctgccc ccatagtggc ctcgaactcc tggcactacc aaaggacact tatccacgag
4381 agcgcagcat ccgaccaggt tgtcactgag aagatgttta ttttggtcag ttgggttttt
4441 atgtattata cttagtcaaa tgtaatgtgg cttctggaat cattgtccag agctgcttcc
4501 ccgtcacctg ggcgtcatct ggtcctggta agaggagtgc gtggcccacc aggccccccc
4561 gtcacccatg acagttcatt cagggccgat ggggcagtcg tggttgggaa cacagcattt
4621 caagcgtcac tttatttcat tcgggcccca cctgcagctc cctcaaagag gcagttgccc
4681 agcctctttc ccttccagtt tattccagag ctgccagtgg ggcctgaggc tccttagggt
4741 tttctctcta tttccccctt tcttcctcat tccctcgtct ttcccaaagg catcacgagt
4801 cagtcgcctt tcagcaggca gccttggcgg tttatcgccc tggcaggcag gggccctgca
4861 gctctcatgc tgcccctgcc ttggggtcag gttgacagga ggttggaggg aaagccttaa
4921 gctgcaggat tctcaccagc tgtgtccggc ccagttttgg ggtgtgacct caatttcaat
4981 tttgtctgta cttgaacatt atgaagatgg gggcctcttt cagtgaattt gtgaacagca
5041 gaattgaccg acagctttcc agtacccatg gggctaggtc attaaggcca catccacagt
5101 ctcccccacc cttgttccag ttgttagtta ctacctcctc tcctgacaat actgtatgtc
5161 gtcgagctcc ccccaggtct acccctcccg gccctgcctg ctggtgggct tgtcatagcc
5221 agtgggattg ccggtcttga cagctcagtg agctggagat acttggtcac agccaggcgc

[0098]

5281 tagcacagct cccttctgtt gatgctgtat tcccatatca aaagacacag gggacaccca
5341 gaaacgccac atcccccaat ccatcagtgc caaactagcc aacggcccca gcttctcagc
5401 tcgctggatg gcggaagctg ctactcgtga gcgccagtgc gggtgcagac aatcttctgt
5461 tgggtggcat cattccaggc ccgaagcatg aacagtgcac ctgggacagg gagcagcccc
5521 aaattgtcac ctgcttctct gcccagcttt tcattgctgt gacagtgatg gcgaaagagg
5581 gtaataacca gacacaaact gccaagttgg gtggagaaag gagtttcttt agctgacaga
5641 atctctgaat tttaaatcac ttagtaagcg gctcaagccc aggagggagc agagggatac
5701 gagcggagtc ccctgcgcgg gaccatctgg aattggttta gcccaagtgg agcctgacag
5761 ccagaactct gtgtcccccg tctaaccaca gctccttttc cagagcattc cagtcaggct
5821 ctctgggctg actgggccag gggaggttac aggtaccagt tctttaagaa gatctttggg
5881 catatacatt tttagcctgt gtcattgccc caaatggatt cctgtttcaa gttcacacct
5941 gcagattcta ggacctgtgt cctagacttc agggagtcag ctgtttctag agttcctacc
6001 atggagtggg tctggaggac ctgcccggtg ggggggcaga gccctgctcc ctccgggtct
6061 tcctactctt ctctctgctc tgacgggatt tgttgattct ctccattttg gtgtctttct
6121 cttttagata ttgtatcaat ctttagaaaa ggcatagtct acttgttata aatcgttagg
6181 atactgcctc ccccagggtc taaaattaca tattagaggg gaaaagctga acactgaagt
6241 cagttctcaa caatttagaa ggaaaaccta gaaaacattt ggcagaaaat tacatttcga
6301 tgtttttgaa tgaatacgag caagctttta caacagtgct gatctaaaaa tacttagcac
6361 ttggcctgag atgcctggtg agcattacag gcaaggggaa tctggaggta gccgacctga
6421 ggacatggct tctgaacctg tcttttggga gtggtatgga aggtggagcg ttcaccagtg
6481 acctggaagg cccagcacca ccctccttcc cactcttctc atcttgacag agcctgcccc
6541 agcgctgacg tgtcaggaaa acacccaggg aactaggaag gcacttctgc ctgaggggca
6601 gcctgccttg cccactcctg ctctgctcgc ctcggatcag ctgagccttc tgagctggcc
6661 tctcactgcc tccccaaggc cccctgcctg ccctgtcagg aggcagaagg aagcaggtgt
6721 gagggcagtg caaggaggga gcacaacccc cagctcccgc tccgggctcc gacttgtgca
6781 caggcagagc ccagaccctg gaggaaatcc tacctttgaa ttcaagaaca tttggggaat
6841 ttggaaatct ctttgccccc aaacccccat tctgtcctac ctttaatcag gtcctgctca
6901 gcagtgagag cagatgaggt gaaaaggcca agaggtttgg ctcctgccca ctgatagccc
6961 ctctccccgc agtgtttgtg tgtcaagtgg caaagctgtt cttcctggtg accctgatta
7021 tatccagtaa cacatagact gtgcgcatag gcctgctttg tctcctctat cctgggcttt
7081 tgttttgctt tttagttttg cttttagttt ttctgtccct tttatttaac gcaccgacta
7141 gacacacaaa gcagttgaat ttttatatat atatctgtat attgcacaat tataaactca
7201 ttttgcttgt ggctccacac acacaaaaaa agacctgtta aaattatacc tgttgcttaa
7261 ttacaatatt tctgataacc atagcatagg acaagggaaa ataaaaaaag aaaaaaaaga
7321 aaaaaaaacg acaaatctgt ctgctggtca cttcttctgt ccaagcagat tcgtggtctt
7381 ttcctcgctt ctttcaaggg ctttcctgtg ccaggtgaag gaggctccag gcagcaccca
7441 ggttttgcac tcttgtttct cccgtgcttg tgaaagaggt cccaaggttc tgggtgcagg
7501 agcgctccct tgacctgctg aagtccggaa cgtagtcggc acagcctggt cgccttccac
7561 ctctgggagc tggagtccac tggggtggcc tgactccccc agtccccttc ccgtgacctg
7621 gtcagggtga gcccatgtgg agtcagcctc gcaggcctcc ctgccagtag ggtccgagtg

[0099]

```
 7681 tgtttcatcc ttcccactct gtcgagcctg ggggctggag cggagacggg aggcctggcc
 7741 tgtctcggaa cctgtgagct gcaccaggta gaacgccagg gaccccagaa tcatgtgcgt
 7801 cagtccaagg ggtcccctcc aggagtagtg aagactccag aaatgtccct ttcttctccc
 7861 ccatcctacg agtaattgca tttgcttttg taattcttaa tgagcaatat ctgctagaga
 7921 gtttagctgt aacagttctt tttgatcatc tttttttaat aattagaaac accaaaaaaa
 7981 tccagaaact tgttcttcca aagcagagag cattataatc accagggcca aaagcttccc
 8041 tccctgctgt cattgcttct tctgaggcct gaatccaaaa gaaaaacagc cataggccct
 8101 ttcagtggcc gggctacccg tgagcccttc ggaggaccag ggctggggca gcctctgggc
 8161 ccacatccgg ggccagctcc ggcgtgtgtt cagtgttagc agtgggtcat gatgctcttt
 8221 cccacccagc ctgggatagg ggcagaggag gcgaggaggc cgttgccgct gatgtttggc
 8281 cgtgaacagg tgggtgtctg cgtgcgtcca cgtgcgtgtt ttctgactga catgaaatcg
 8341 acgcccgagt tagcctcacc cggtgacctc tagccctgcc cggatggagc ggggcccacc
 8401 cggttcagtg tttctgggga gctggacagt ggagtgcaaa aggcttgcag aacttgaagc
 8461 ctgctccttc ccttgctacc acggcctcct ttccgtttga tttgtcactg cttcaatcaa
 8521 taacagccgc tccagagtca gtagtcaatg aatatatgac caaatatcac caggactgtt
 8581 actcaatgtg tgccgagccc ttgcccatgc tgggctcccg tgtatctgga cactgtaacg
 8641 tgtgctgtgt ttgctccccc tccccttcct tctttgccct ttacttgtct ttctggggtt
 8701 tttctgtttg ggtttggttt ggtttttatt tctccttttg tgttccaaac atgaggttct
 8761 ctctactggt cctcttaact gtggtgttga ggcttatatt tgtgtaattt ttggtgggtg
 8821 aaaggaattt tgctaagtaa atctcttctg tgtttgaact gaagtctgta ttgtaactat
 8881 gtttaaagta attgttccag agacaaatat ttctagacac tttttcttta caaacaaaag
 8941 cattcggagg gaggggggatg gtgactgaga tgagagggga gagctgaaca gatgacccct
 9001 gcccagatca gccagaagcc acccaaagca gtggagccca ggagtcccac tccaagccag
 9061 caagccgaat agctgatgtg ttgccacttt ccaagtcact gcaaaaccag gttttgttcc
 9121 gcccagtgga ttcttgtttt gcttcccctc cccccgagat tattaccacc atcccgtgct
 9181 tttaaggaaa ggcaagattg atgtttcctt gaggggagcc aggaggggat gtgtgtgtgc
 9241 agagctgaag agctggggag aatggggctg ggcccaccca agcaggaggc tgggacgctc
 9301 tgctgtgggc acaggtcagg ctaatgttgg cagatgcagc tcttcctgga caggccaggt
 9361 ggtgggcatt ctctctccaa ggtgtgcccc gtgggcatta ctgtttaaga cacttccgtc
 9421 acatcccacc ccatcctcca gggctcaaca ctgtgacatc tctattcccc accctccct
 9481 tcccagggca ataaaatgac catggagggg gcttgcactc tcttggctgt cacccgatcg
 9541 ccagcaaaac ttagatgtga gaaaacccct tcccattcca tggcgaaaac atctccttag
 9601 aaaagccatt accctcatta ggcatggttt tgggctccca aaacacctga cagcccctcc
 9661 ctcctctgag aggcggagag tgctgactgt agtgaccatt gcatgccggg tgcagcatct
 9721 ggaagagcta ggcagggtgt ctgccccctc ctgagttgaa gtcatgctcc cctgtgccag
 9781 cccagaggcc gagagctatg gacagcattg ccagtaacac aggccaccct gtgcagaagg
 9841 gagctggctc cagcctggaa acctgtctga ggttgggaga ggtgcacttg gggcacaggg
 9901 agaggccggg acacacttag ctggagatgt ctctaaaagc cctgtatcgt attcaccttc
 9961 agtttttgtg ttttgggaca attactttag aaaataagta ggtcgtttta aaaacaaaaa
10021 ttattgattg ctttttttgta gtgttcagaa aaaaggttct ttgtgtatag ccaaatgact
```

[0100]

```
10081 gaaagcactg atatatttaa aaacaaaagg caatttatta aggaaatttg taccatttca
10141 gtaaacctgt ctgaatgtac ctgtatacgt ttcaaaaaca ccccccccc actgaatccc
10201 tgtaacctat ttattatata aagagtttgc cttataaatt t
```

[0101]

[0102] "염기 편집기 (BE)" 또는 "핵염기 편집기 (NBE)"란 폴리뉴클레오타이드에 결합하고 핵염기 변형 활성을 갖는 제제를 의미한다. 다양한 구현예에서, 염기 편집기는 핵염기 변형 폴리펩타이드 (예를 들어, 데아미나제), 및 가이드 폴리뉴클레오타이드 (예를 들어, 가이드 RNA)와 연합된 핵산 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함한다. 다양한 구현예에서, 제제는 염기 편집 활성을 갖는 단백질 도메인, 즉, 핵산 분자 (예를 들어, DNA) 내 염기 (예를 들어, A, T, C, G, 또는 U)를 변형시킬 수 있는 도메인을 포함하는 생분자 복합체이다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 데아미나제 도메인에 융합되거나 연결된다. 하나의 구현예에서, 제제는 염기 편집 활성을 갖는 도메인을 포함하는 융합 단백질이다. 또 다른 구현예에서, 염기 편집 활성을 갖는 단백질 도메인은 가이드 RNA (예를 들어, 데아미나제에 융합된 가이드 RNA 및 RNA 결합 도메인 상의 RNA 결합 모티프를 통해)에 연결된다. 일부 구현예에서, 염기 편집기 활성을 갖는 도메

인은 핵산 분자 내 염기를 탈아민화할 수 있다. 일부 구현예에서, 염기 편집기는 DNA 분자 내 하나 이상의 염기를 탈아민화할 수 있다. 일부 구현예에서, 염기 편집기는 DNA 내 아데노신 (A)을 탈아민화할 수 있다. 일부 구현예에서, 염기 편집기는 아데노신 염기 편집기 (ABE)이다.

[0103] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 변이체 (예를 들어, TadA*8)를 환형의 퍼뮤턴트 Cas9 (예를 들어, spCAS9 또는 saCAS9) 및 양분된 핵 국소화 서열을 포함하는 스캐폴드에 클로닝함에 의해 생성된다 (예를 들어, ABE8). 환형의 퍼뮤턴트 Cas9는 당업계에 공지되어 있고, 예를 들어, 문헌 (참조: Oakes *et al.*, Cell 176, 254-267, 2019)에 기재되어 있다. 예시적인 환형의 퍼뮤턴트는 하기에 제시되어 있고, 여기서, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭한다.

[0104] CP5 (MSP "NGC=NGG와 같은 돌연변이 정규 Cas9를 갖는 팜 변이체", PID=단백질 상호작용 도메인 및 "D10A" 닉카제와 함께):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQK
GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL
DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPS**
KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENED
ILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI
LDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0105]

[0106] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 염기 편집기 8 (ABE8)이다. 일부 구현예에서, ABE8은 하기 **표 9**로부터의 염기 편집기로부터 선택된다. 일부 구현예에서, ABE8은 TadA로부터 유래된 아데노신 데아미나제 변이체를 함유한다. 일부 구현예에서, ABE8의 아데노신 데아미나제 변이체는 하기 **표 9**에 기재된 바와 같은 TadA*8 변이체이다. 일부 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10 변이체 (예를 들어, TadA*8)이고, 이는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R의 그룹으로부터 선택된 변경의 하나 이상을 포함한다. 다양한 구현예에서, ABE8은 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 그룹으로부터 선택된 변경의 조합과 함께 TadA*7.10 변이체 (예를 들어, TadA*8)를 포함한다. 일부 구현예에서, ABE8은 단량체 작제물이다. 일부 구현예에서, ABE8은 이종이량체 작제물이다. 일부 구현예에서, ABE8 염기 편집기는 하기의 서열을 포함한다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV

[0107] EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0108] 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 CRISPR 연합된 (예를 들어, Cas 또는 Cpf1) 효소이다. 일부 구현예에서, 염기 편집기는 데아미나제 도메인과 융합된 촉매적으로 데드 Cas9 (dCas9)이다. 일부 구현예에서, 염기 편집기는 데아미나제 도메인과 융합된 Cas9 닉카제 (nCas9)이다. 일부 구현예에서, 염기 편집기는 염기 절제 복구 (BER)의 저해제에 융합된다. 일부 구현예에서, 염기 절제 복구의 저해제는 우라실 DNA 글리코실라제 저해제 (UGI)이다. 일부 구현예에서, 염기 절제 복구의 저해제는 이노신 염기 절제 복구 저해제이다. 염기 편집기의 세부사항은 국제 PCT 출원 번호 PCT/2017/045381 (WO 2018/027078) 및 PCT/US2016/058344 (WO2017/070632)에 기재되어 있고, 이의 각각은 이의 전문이 본원에 참조로 포함된다. 또한 문헌 (참조: 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), and 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 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[0109] 예를 들어, 염기 편집 조성물에 사용된 바와 같은 아데닌 염기 편집기 ABE, 본원에 기재된 시스템 및 방법은 아래에 제공된 바와 같은 핵산 서열 (8877 염기 쌍)을 갖는다 (참조: Addgene, Watertown, MA.; 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 서열과 적어도 95% 이상의 동일성을 갖는 폴리뉴클레오타이드 서열이 또한 포함된다.

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACAT
GACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG
TTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCC
ATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGT
CAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGAAACGGACA
[0110] 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

[0111]

CATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAG
AAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACAT
ACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGA
AGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCC
GGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGG
CTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAA
GCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTA
AGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGA
GAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGA
ACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGAC
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

[0112]

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

[0113]

[0114] "염기 편집 활성"이란 폴리뉴클레오타이드 내에서 염기를 화학적으로 변경하는 작용을 함을 의미한다. 하나의 구현예에서, 제1 염기는 제2 염기로 전환된다. 하나의 구현예에서, 예를 들어, 염기 편집 활성은 예를 들어, A·T를 G·C로 전환시키는 아데노신 또는 아데닌 데아미나제 활성이다.

[0115] 일부 구현예에서, 염기 편집 활성은 편집 효율에 의해 평가된다. 염기 편집 효율은 임의의 적합한 수단에 의해, 예를 들어, 생거 서열분석 또는 차세대 서열분석에 의해 측정될 수 있다. 일부 구현예에서, 염기 편집 효율은 염기 편집기에 의해 수행된 핵염기 전환을 갖는 총 서열분석 판독의 백분율, 예를 들어, G·C 염기 쌍으로 전환된 표적 A·T 염기 쌍을 갖는 총 서열분석 판독의 백분율에 의해 측정된다. 일부 구현예에서, 염기 편집 효율은 염기 편집이 세포 집단에서 수행되는 경우, 염기 편집기에 의해 수행된 핵염기 전환을 갖는 총 세포의 백분율에 의해 측정된다.

[0116] 용어 "염기 편집기 시스템"은 표적 뉴클레오타이드 서열의 핵염기를 편집하기 위한 시스템을 언급한다. 다양한 구현예에서, 염기 편집기 (BE) 시스템은 (1) 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 (예를 들어, Cas9); (2) 상기 핵염기를 탈아민화하기 위한 데아미나제 도메인 (예를 들어, 아데노신 데아미

나제); 및 (3) 하나 이상의 가이드 폴리뉴클레오타이드 (예를 들어, 가이드 RNA)를 포함한다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이다. 일부 구현예에서, 염기 편집기는 아데닌 또는 아데노신 염기 편집기 (ABE)이다. 일부 구현예에서, 염기 편집기 시스템은 아데노신 데아미나제 염기 편집기 (ABE8)이다. 일부 구현예에서, ABE8은 단량체 작제물이다. 일부 구현예에서, 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 구현예에서, ABE8은 이종 작제물이다. 일부 구현예에서, 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.

[0117] 일부 구현예에서, 핵염기 편집기 시스템은 하나 초과의 염기 편집 성분을 포함할 수 있다. 예를 들어, 염기 편집기 시스템은 하나 초과의 아데노신 데아미나제를 포함할 수 있다. 일부 구현예에서, 단일 가이드 폴리뉴클레오타이드는 상이한 데아미나제를 표적 핵산 서열에 표적화하기 위해 사용될 수 있다. 일부 구현예에서, 단일 쌍의 가이드 폴리뉴클레오타이드는 상이한 데아미나제를 표적 핵산 서열에 표적화하기 위해 사용될 수 있다.

[0118] 염기 편집기 시스템의 데아미나제 도메인 및 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 성분은 서로 공유적으로 또는 비공유적으로, 또는 연합 및 이의 상호작용의 임의의 조합과 연합될 수 있다. 예를 들어, 일부 구현예에서, 데아미나제 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미나제 도메인에 융합되거나 연결될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미나제 도메인을 데아미나제 도메인과 비공유적으로 상호작용하거나 연합함에 의해 표적 뉴클레오타이드 서열에 표적화시킬 수 있다. 예를 들어, 일부 구현예에서, 데아미나제 도메인은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 일부인 추가의 이종성 부분 또는 도메인과 상호작용하거나, 이와 연합하거나 이와 복합체를 형성할 수 있는 추가의 이종성 부분 또는 도메인을 포함할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가의 이종성 부분은 단백질 도메인일 수 있다. 일부 구현예에서, 추가의 이종성 부분은 K 상동성 (KH) 도메인, MS2 코트 단백질 도메인, PP7 코트 단백질 도메인, SfMu Com 코트 단백질 도메인, 무균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인지 모티프일 수 있다.

[0119] 염기 편집기 시스템은 가이드 폴리뉴클레오타이드 성분을 추가로 포함할 수 있다. 염기 편집기 시스템의 성분들은 공유 결합, 비공유 상호작용, 또는 이의 연합과 상호작용의 임의의 조합을 통해 서로 연합될 수 있음을 인지해야 한다. 일부 구현예에서, 데아미나제 도메인은 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 예를 들어, 일부 구현예에서, 데아미나제 도메인은 가이드 폴리뉴클레오타이드의 일부 또는 분절 (예를 들어, 폴리뉴클레오타이드 모티프)과 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있는 추가의 이종성 부분 또는 도메인 (예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)을 포함할 수 있다. 일부 구현예에서, 추가의 이종성 부분 또는 도메인 (예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)은 데아미나제 도메인에 융합되거나 연결될 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가의 이종성 부분은 단백질 도메인일 수 있다. 일부 구현예에서, 추가의 이종성 부분은 K 상동성 (KH) 도메인, MS2 코트 단백질 도메인, PP7 코트 단백질 도메인, SfMu Com 코트 단백질 도메인, 무균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인지 모티프일 수 있다.

[0120] 일부 구현예에서, 염기 편집기 시스템은 염기 절제 복구 (BER) 성분의 저해제를 추가로 포함할 수 있다. 염기 편집기 시스템의 성분들은 공유 결합, 비공유 상호작용, 또는 이의 연합과 상호작용의 임의의 조합을 통해 서로 연합될 수 있음을 인지해야 한다. BER 성분의 저해제는 염기 절제 복구 (BER) 저해제를 포함할 수 있다. 일부 구현예에서, 염기 절제 복구 (BER)의 저해제는 우라실 DNA 글리코실라제 저해제 (UGI)일 수 있다. 일부 구현예에서, 염기 절제 복구의 저해제는 이노신 염기 절제 복구 (BER) 저해제일 수 있다. 일부 구현예에서, 염기 절제 복구 (BER)의 저해제는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 염기 절제 복구 (BER)의 저해제에 융합되거나 연결될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오티드 결합 도메인은 데아미나제 도메인 및 염기 절단 복구 (BER)의 저해제에 융합되거나 연결될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오티드 결합 도메인은 염기 절제 복구의 저해제를 염기 절제 복구 (BER)의 저해제와 비공유적으로 상호작용하거나 이와 연합함에 의해 표적 뉴클레오타이드 서열에 표적화시킬 수 있다. 예를 들어, 일부 구현예에서, 염기 절제 복구 (BER) 성분의 저해제는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 일부인 추가의 이종성 부분 또는 도메인과 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있는 추가의 이종성 부분 또는 도메인을 포함할 수 있다. 일부 구현예에서, 염기 절제 복구 (BER)의 저해제는 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 예를 들어, 일부 구현예에서, 염기 절제 복구 (BER)의 저해제는 가이드 폴리뉴클레오타이드의 일부 또는 분절 (예를 들어, 폴리뉴클레오타이드 모티프)과 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있는 추가의 이종성 부분 또는 도메인 (예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)을 포함할 수 있다. 일부 구현예에서, 가이드 폴리뉴클레오타이드의 추가의 이종성 부분 또는 도메인 (예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)은 염기 절제 복구 (BER)의 저해제에 융합되거나 연결될 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가의 이종성 부분은 단백질 도메인일 수 있다. 일부 구현예에서, 추가의 이종성 부분은 K 상동성 (KH) 도메인, MS2 코트 단백질 도메인, PP7 코트 단백질 도메인, SfMu Com 코트 단백질 도메인, 무균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인지 모티프일 수 있다.

[0121] 용어 "Cas9" 또는 "Cas9 도메인"은 Cas9 단백질, 또는 이의 단편을 포함하는 RNA 가이드된 뉴클레아제 (예를 들어, Cas9의 활성, 불활성, 또는 부분적 활성 DNA 절단 도메인, 및/또는 Cas9의 gRNA 결합 도메인을 포함하는 단백질)를 언급한다. Cas9 뉴클레아제는 또한 때로는 Casn1 뉴클레아제 또는 CRISPR (클러스터형 규칙적 간격을 둔 짧은 팔린드롬 반복체) 연합된 뉴클레아제로서 언급된다. CRISPR은 이동 유전학적 요소 (바이러스, 전이할 수 있는 요소 (transposable elements) 및 접합성 플라스미드)에 대한 보호를 제공하는 후천성 면역계이다. CRISPR 클러스터는 스페이서, 선행 이동 요소에 상보적인 서열 및 표적 공격 핵산을 포함한다. CRISPR 클러스터는 CRISPR RNA (crRNA)로 전사되고 프로세싱된다. II형 CRISPR 시스템에서, 전구-crRNA의 올바른 프로세싱은 트랜스-암호화된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 단백질을 요구한다. tracrRNA는 전구-crRNA의 리보뉴클레아제 3-원조 프로세싱에 대한 가이드로서 작용한다. 후속적으로, Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 환형 dsDNA 표적을 엔도핵산분해적으로 절단한다. crRNA에 상보적이지 않은 표적 가닥은 먼저 엔도핵산분해적으로 절단됨에 이어서 3'-5' 엑소핵산분해적으로 절단 제거한다. 실제로, DNA-결합 및 절단은 전형적으로 단백질 및 2개의 RNA를 필요로 한다. 그러나, 단일 가이드 RNA ("sgRNA", 또는 단순히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 양상을 단일 RNA 종으로 혼입하기 위해 가공될 수 있다. 예를 들어, 문헌 (Jinek M., *et al.*, *Science* 337:816-821(2012))을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다. Cas9는 자가 대 비-자가의 구분을 도와주기 위해 CRISPR 반복 서열 (PAM 또는 프로토스페이서 인접 모티프)에서 짧은 모티프를 인지한다. Cas9 뉴클레아제 서열 및 구조는 당업자에게 널리 공지되어 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); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., *et al.*, *Science* 337:816-821(2012)) 참조, 이의 각각의 전체 내용은 본원에 참조로 포함됨). Cas9 오톨로그는 에스. 피오게네스 (S. *Pyogenes*) 및 에스. 써모필러스 (*S. thermophilus*)를 포함하지만 이에 제한되지 않는 다양한 종

에 기재되어 있다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본원 개시내용을 기준으로 당업자에게 자명할 것이고, 상기 Cas9 뉴클레아제 및 서열은 이의 전문이 본원에 참조로 포함되는 문헌 (참조: 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 유전자좌로부터 Cas9 서열을 포함한다.

[0122] 예시적인 Cas9는 스트렙토코커스 피오게네스 Cas9 (spCas9)이고, 이의 아미노산 서열은 아래에 제공된다:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQL
FEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
IKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD

[0123]

[0124] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0125] 뉴클레아제-불활성화된 Cas9 단백질은 상호교환적으로 "dCas9" 단백질 (뉴클레아제-"데드" Cas9) 또는 촉매 불활성 Cas9로서 언급될 수 있다. 불활성 DNA 절단 도메인을 갖는 Cas9 단백질 (또는 이의 단편)을 생성하기 위한 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, 이의 각각의 전체 내용은 본원에 참조로 포함됨). 예를 들어, Cas9의 DNA 절단 도메인은 2개의 서브도메인인 HNH 뉴클레아제 서브도메인 및 RuvC1 서브도메인을 포함하는 것으로 공지되어 있다. HNH 서브도메인은 gRNA에 상보적인 가닥을 절단하는 반면 RuvC1 서브도메인은 비-상보적 가닥을 절단한다. 이들 서브도메인 내 돌연변이는 Cas9의 뉴클레아제 활성을 사일런싱시킬 수 있다. 예를 들어, 돌연변이 D10A 및 H840A는 에스. 피오게네스 Cas9의 뉴클레아제 활성을 완전히 불활성화시킨다 (참조: Jinek *et al., Science.* 337:816-821(2012); Qi *et al., Cell.* 28;152(5):1173-83 (2013)). 일부 구현예에서, Cas9 뉴클레아제는 불활성 (예를 들어, 불활성화된) DNA 절단 도메인을 갖고, 즉, Cas9는 "nCas9" 단백질 ("닉카제" Cas9에 대해)로서 언급되는 닉카제이다. 일부 구현예에서, Cas9의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 구현예에서, 단백질은 2개의 Cas9 도메인 중 하나를 포함한다: (1) Cas9의 gRNA 결합 도메인; 또는 (2) Cas9의 DNA 절단 도메인. 일부 구현예에서, Cas9 또는 이의 단편을 포함하는 단백질은 "Cas9 변이체"로서 언급된다. Cas9 변이체는 Cas9 또는 이의 단편과 상동성을 공유한다. 예를 들어, 상기 Cas9 변이체는 야생형 Cas9와 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, Cas9 변이체는 야생형 Cas9와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 아미노산 변화를 가질 수 있

다. 일부 구현예에서, Cas9 변이체는 Cas9의 단편 (예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함하여, 상기 단편은 야생형 Cas9의 상응하는 단편과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, 단편은 상응하는 야생형 Cas9의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 동일하거나, 이의 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%이다.

[0126] 일부 구현예에서, 단편은 적어도 100개 아미노산 길이이다. 일부 구현예에서, 단편은 적어도 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 또는 적어도 1300개 아미노산 길이이다.

[0127] 일부 구현예에서, 야생형 Cas9는 스트렙토코커스 피오게네스 기원의 Cas9에 상응한다 (NCBI 참조 서열: NC_017053.1, 다음과 같은 뉴클레오타이드 및 아미노산 서열).

[0128]
ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCAC
TGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCA
AAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAA
CGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTC
AAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAG
AAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAG
AAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCG
CTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATT
TAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTA
TTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAG
TAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTG
GGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAA
GATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAAT
TGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAG
ATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTAC
GATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTA

TAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCC
AAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTG
GTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCA
TCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAA
AAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTG
GCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAA
TTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTG
ATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTT
TATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGG
TGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAAT
TAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGAT
AGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTT
GGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATA
GGGGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAG
CTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGA
TAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTA
TGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGA
CAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTAT
TTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCG
TTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATG
AAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAA
TACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGG
ACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTC
ATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGA
TAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCA
AGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTT
GATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACA
AATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAG
TGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGT
GAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGAT
TAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTA
AAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAAT
ATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAAT
CGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCA

[0129]

AAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCC
AAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCC
AAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGG
AAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGA
AGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGA
CTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG
CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTA
TATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTT
TGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTG
TTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCA
ATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATG
CCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGA
GGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQL
FEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
IKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL

[0130]

YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD

[0131]

[0132] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0133] 일부 구현예에서, 야생형 Cas9는 하기의 뉴클레오타이드 및/또는 아미노산 서열에 상응하거나 이를 포함한다:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAAC
CGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTA
AAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAA
CGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAG
CAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAG
AGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAA
AAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAG
GTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATC
TAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTG
TTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTC
TAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCG
GTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAA
GATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAAT
TGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTG
ACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATA
TAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTC
AAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTT
GTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACA
TCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCA
AAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTG
GCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAA
TTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTG
ACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTG
TACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGG
AGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAAT
TGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGAT

[0134]

CGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCT
GGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATC
GGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAG
TTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGA
CAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTA
TGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGA
CAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCAT
ACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACA
TTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGG
ATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGA
AAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATG
TTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCC
TTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAG
TGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATG
CGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAA
CTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGC
ACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCA
AAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTT
AGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACT
CATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCC
GTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCT
AACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTT
AATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGA
GAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTT
TCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGA
CCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAG
TTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAG
CGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAA
GGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGT
TGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTC
CTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACT
TTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGA
GAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAA
[0135] CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGC

CGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAG
ACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTT
GGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGA
TTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0136]

[0137] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0138] 일부 구현예에서, 야생형 Cas9는 스트렙토코커스 피오게네스로부터의 Cas9에 상응한다 (NCBI 참조 서열: NC_002737.2 (하기와 같은 뉴클레오타이드 서열); 및 Uniprot 참조 서열: Q99ZW2 (하기와 같은 아미노산 서열)).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCAC
[0139] TGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCA

AAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAA
CGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTC
AAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAG
AAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAG
AAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCG
CTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATT
TAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTA
TTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAG
TAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTG
GGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAA
GATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAAT
TGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAG
ATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTAC
GATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTA
TAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCC
AAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTG
GTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCA
TCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAA
AAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTG
GCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAA
TTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTG
ATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTT
TATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGG
TGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAAT
TAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGAT
AGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTT
GGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATA
GGGAGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAG
CTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGA
TAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTA
TGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGA
CAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTAT
TTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATA
[0140] TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGT

ATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGA
AAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATG
TGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGT
TTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATC
GGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACG
CCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAA
CTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGC
ACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTA
AAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTA
CGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTT
GATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTC
GTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCT
AATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCT
AATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGC
GCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTC
TCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGA
TCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGG
TGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAA
AGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAA
AGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGC
TGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTT
TTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATT
GTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGC
GTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAA
CCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGC
TGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAG
ATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA
GGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
[0141] DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY

DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0142]

[0143] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0144] 일부 구현예에서, Cas9는 코리네박테리움 울세란스 (*Corynebacterium ulcerans*) (NCBI Refs: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아 (*Corynebacterium diphtheria*) (NCBI Refs: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라 (*Spiroplasma syrphidicola*) (NCBI Ref: NC_021284.1); 프레보텔라 인터메디아 (*Prevotella intermedia*) (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨스 (*Spiroplasma taiwanense*) (NCBI Ref: NC_021846.1); 스트렙토코커스 이니애 (*Streptococcus iniae*) (NCBI Ref: NC_021314.1); 벨리엘라 발티카 (*Belliella baltica*) (NCBI Ref: NC_018010.1); 사이크로플렉서스 토르쿠이스I (*Psychroflexus torquisI*) (NCBI Ref: NC_018721.1); 스트렙토코커스 써모필러스 (*Streptococcus thermophilus*) (NCBI Ref: YP_820832.1), 리스테리아 이노쿠아 (*Listeria innocua*) (NCBI Ref: NP_472073.1), 캄필로박터 제주니 (*Campylobacter jejuni*) (NCBI Ref: YP_002344900.1) 또는 나이세리아 메닌기티디스 (*Neisseria. meningitidis*) (NCBI Ref: YP_002342100.1)로부터 기원하는 Cas9를 언급하거나, 임의의 다른 유기체 기원의 Cas9를 언급한다.

[0145] 일부 구현예에서, Cas9는 나이세리아 메닌기티디스 (Nme)로부터 기원한다. 일부 구현예에서, Cas9는 Nme1, Nme2 또는 Nme3이다. 일부 구현예에서, Nme1, Nme2 또는 Nme3에 대한 PAM-상호작용 도메인은 각각 N_4GAT, N_4CC, 및 N_4CAAA이다 (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). 예시적인 나이세리아 메닌기티디스 Cas9 단백질, Nme1Cas9 (NCBI 참조: WP_002235162.1; II형 CRISPR RNA-가이드된 엔도뉴클레아제 Cas9)는 하기의 아미노산 서열을 갖는다:

```
   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
 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
```

[0146]

[0147] 또 다른 예시적인 나이세리아 메닌기티디스 Cas9 단백질, Nme2Cas9, (NCBI 참조: WP_002230835; II형 CRISPR RNA-가이드된 엔도뉴클레아제 Cas9)는 하기의 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
```

[0148]

[0149] 일부 구현예에서, dCas9는 Cas9 뉴클레아제 활성을 불활성화시키는 하나 이상의 돌연변이를 갖는 Cas9 아미노산 서열에 부분적으로 또는 전반적으로 상응하거나 포함한다. 예를 들어, 일부 구현예에서, dCas9 도메인은 또 다른 Cas9에서 D10A 및 H840A 돌연변이 또는 상응하는 돌연변이를 포함한다. 일부 구현예에서, dCas9는 dCas9 (D10A 및 H840A)의 아미노산 서열을 포함한다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0150]

[0151] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0152] 일부 구현예에서, Cas9 도메인은 D10A 돌연변이를 포함하고, 위치 840에서 잔기는 상기 제공된 아미노산 서열에서, 또는 본원에 제공된 임의의 아미노산 서열에서 상응하는 위치에 히스티딘을 유지한다.

[0153] 다른 구현예에서, D10A 및 H840A 이외의 돌연변이를 갖는 dCas9 변이체가 제공되고, 상기 변이체는 예를 들어, 뉴클레아제 불활성화된 Cas9 (dCas9)를 유도한다. 상기 돌연변이는 예를 들어 D10 및 H840에서 다른 아미노산 치환, 또는 Cas9의 뉴클레아제 도메인 내 다른 치환 (예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인에서의 치환)을 포함한다. 일부 구현예에서, dCas9의 변이체 또는 동족체가 제공되고, 이는 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, dCas9의 변이체가 제공되고, 약 5개 아미노산, 약 10개 아미노산, 약 15개 아미노산, 약 20개 아미노산, 약 25개 아미노산, 약 30개 아미노산, 약 40개 아미노산, 약 50개 아미노산, 약 75개 아미노산, 약 100개 이상의 아미노산만큼 더 짧거나 더 긴 아미노산 서열을 갖는다.

[0154] 일부 구현예에서, 본원에 제공된 바와 같은 Cas9 융합 단백질은 Cas9 단백질의 전장 아미노산 서열, 예를 들어, 본원에 제공된 Cas9 서열 중 하나를 포함한다. 그러나, 다른 구현예에서, 본원에 제공된 바와 같은 융합 단백질은 전장 Cas9 서열을 포함하지 않고 단지 하나 이상의 이의 단편을 포함한다. 적합한 Cas9 도메인 및 Cas9 단편의 예시적인 아미노산 서열이 본원에 제공되고, Cas9 도메인 및 단편의 추가의 적합한 서열은 당업자에게 자명할 것이다.

[0155] 변이체 및 이의 동족체를 포함하는, 추가의 Cas9 단백질 (예를 들어, 뉴클레아제 데드 Cas9 (dCas9), Cas9 닉카제 (nCas9), 또는 뉴클레아제 활성 Cas9)이 본원의 범위 내에 있는 것으로 인지해야 한다. 예시적인 Cas9 단백질은 제한 없이 하기에 제공된 것들을 포함한다. 일부 구현예에서, Cas9 단백질은 뉴클레아제 데드 Cas9 (dCas9)이다. 일부 구현예에서, Cas9 단백질은 Cas9 닉카제 (nCas9)이다. 일부 구현예에서, Cas9 단백질은 뉴클레아제 활성 Cas9이다.

[0156] 예시적인 촉매적 불활성 Cas9 (dCas9):

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF
EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD

[0157]

[0158] 예시적인 촉매적 Cas9 닉카제 (nCas9):

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF
EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD

[0159]

[0160] 예시적인 촉매적 활성 Cas9:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF
EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL
KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD.

[0161]

[0162] 일부 구현예에서, Cas9는 고세균 (archaea) (예를 들어, 나노고세균) 기원의 Cas9를 언급하고, 이것은 단세포 원핵 미생물의 도메인 및 킹덤을 구성한다. 일부 구현예에서, Cas9는 CasX 또는 CasY를 언급하고, 이는 예를 들어, 이의 전체 내용이 참조로 포함되는 문헌 (참조: Burstein *et al.*, "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21)에 기재되어 있다. 게놈-분리 균 유전체학을 사용하여, 생활 고세균 도메인에서 최초 보고된 Cas9를 포함하는, 다수의 CRISPR-Cas 시스템을 동정하였다. 이러한 다양한 Cas9 단백질은 활성 CRISPR-Cas 시스템의 일부로서 거의 연구되지 않은 나노고세균에서 발견되었다. 세균에서, 2개의 이전에 공지되지 않은 시스템인 CRISPR-CasX 및 CRISPR-CasY가 발견되었고, 이는 지금까지 발견된 가장 콤팩트한 시스템 중 하나이다. 일부 구현예에서, Cas9는 CasX, 또는 CasX의 변이체를 언급한다. 일부 구현예에서, Cas9는 CasY, 또는 CasY의 변이체를 언급한다. 다른 RNA-가이드된 DNA 결합 단백질이 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)로서 사용될 수 있고 본원 개시내용의 범위 내에 있는 것으로 인지되어야 한다.

[0163] 특정 구현예에서, 본원 개시내용의 방법에 유용한 napDNAbps는 당업계에 공지되고 예를 들어, 문헌 (참조: Oakes *et al.*, Cell 176, 254-267, 2019)에 기재된 환형의 퍼뮤턴트를 포함한다. 예시적인 환형의 퍼뮤턴트는 하기에 제시되어 있고, 여기서, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭한다.

[0164] **CP5** (MSP "NGC=NGG와 같이 돌연변이 정규 Cas9를 갖는 팜 변이체", PID=단백질 상호작용 도메인 및 "D10A" 닉카제와 함께):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKFLQK
GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL
DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEYKVPS
KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENED
ILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI
LDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL
ESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0165]

[0166] 염기 편집기에 혼입될 수 있는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 비제한적인 예는 CRISPR 단백질-유래된 도메인, 제한 뉴클레아제, 메가뉴클레아제, TAL 뉴클레아제 (TALEN), 및 아연 핑거 뉴클레아제 (ZFN)를 포함한다.

[0167] 일부 구현예에서, 본원에 제공된 임의의 융합 단백질의 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)은 CasX 또는 CasY 단백질일 수 있다. 일부 구현예에서, napDNAbp는 CasX 단백질이다. 일부 구현예에서, napDNAbp는 CasY 단백질이다. 일부 구현예에서, napDNAbp는 천연적으로 발생하는 CasX 또는 CasY 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, napDNAbp는 천연적으로 발생하는 CasX 또는 CasY 단백질이다. 일부 구현예에서, napDNAbp는 본원에 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 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종 기원의 Cas12b/C2c1, CasX 및 CasY가 또한 본원의 개시내용에 따라 사용될 수 있음을 인지해야 한다.

[0168] Cas12b/C2c1 (uniprot.org/uniprot/T0D7A2#2)

[0169] sp|T0D7A2|C2C1_ALIAG CRISPR-연합된 엔도-뉴클레아제 C2c1 OS = 알리사이클로바실러스 액시도테레스트리스 (*Alicyclobacillus acidoterrestris*) (균주 ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTA
EECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLAD
KDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYT
DSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQ
EHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNV
QRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPI
WTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDP
NEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRP
PYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVA
RKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLA
YLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAV
YESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKV

[0170]

SGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLI
LLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGI
RCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIH
ADLNAAQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYE
RERGKKRRKVFAQEKLSEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRI
EGYLVKQIRSRVPLQDSACENTGDI

[0171]

[0172] CasX (uniprot.org/uniprot/FONN87; uniprot.org/uniprot/FONH53)

[0173] >tr|FONN87|FONN87_SULIH CRISPR-연합된 Casx 단백질 OS = 설폴로부스 이슬란디쿠스 (*Sulfolobus islandicus*) (균주 HVE10/4) GN= SiH_0402 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKK
KKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPE
FYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLI
QNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLE
KRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLYFANRDLIMNLNSDDG
KVRDLKLISAYVNGELIRGEG

[0174]

[0175] >tr|FONH53|FONH53_SULIR CRISPR 연합된 단백질, Casx OS = 설폴로부스 이슬란디쿠스 (균주 REY15A) GN=SiRe_0771 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKK
KKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPE
FYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLI
QNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLE
KRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGK
VRDLKLISAYVNGELIRGEG

[0176]

[0177] 델타프로테오박테리아 CasX

MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNL
RMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQ
CGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLGKFGQRALDF
YSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKR
LESLRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLRL
KGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDHK
KREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAEDA
QSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSIG
SDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVIDL
TFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVAL
TFERREVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQ
RAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQGK
RTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITYADMDVMLVRLKKT
SDGWATTLNNKELKAEYQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLL
KKRFSHRPVQEQFVCLDCGHEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYK
RRLKEVWKPNA

[0178]

[0179] CasY (ncbi.nlm.nih.gov/protein/APG80656.1)

[0180] >APG80656.1 CRISPR-연합된 단백질 CasY (배양되지 않은 파르쿠박테리아 그룹 박테리움 (Parcubacteria group bacterium))

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGL
SNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLY
DELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASL
GERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKN
EGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRI
LAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGE
SDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKE
RLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTD
ALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDI
VSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEE
YIDLIELHKTALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSE
LRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEP
ESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQ
NKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSER
VFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVK
GLKLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYA
TLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGT
VRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADI
QASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI

[0181]

[0182] 용어 "Cas12" 또는 "Cas12 도메인"은 Cas12 단백질 또는 이의 단편 (예를 들어, Cas12의 활성, 불활성 또는 부분 활성 DNA 절단 도메인, 및/또는 Cas12의 gRNA 결합 도메인을 포함하는 단백질)을 포함하는 RNA 가이드된 뉴클레아제를 언급한다. Cas12는 부류 2, V형 CRISPR/Cas 시스템에 속한다. Cas12 뉴클레아제는 종종 CRISPR (클러스터링된 규칙적 간격을 둔 짧은 팔린드롬 반복체) 연합된 뉴클레아제로서 언급된다. 예시적인 바실러스 히사시 (*Bacillus hisashii*) Cas 12b (BhCas12b) Cas 12 도메인의 서열은 아래에 제공된다:

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEH
HEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEAN
QLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLA
EYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKV
EKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDEN
EPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQ
ATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGK
VDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGN
VGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLR
VMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRK
AREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKP
YKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEP
GEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFED
LSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSV
VTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNL
QKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIK
KGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISK
[0183] LTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK.

[0184] BhCas12b 아미노산 서열에 대해 적어도 85% 이상의 동일성을 갖는 아미노산 서열이 또한 본원 개시내용의 방법에 유용하다.

[0185] 용어 "보존성 아미노산 치환" 또는 "보존성 돌연변이"는 하나의 아미노산의 공통된 성질을 갖는 또 다른 아미노산으로의 대체를 언급한다. 개별 아미노산 간의 공통된 성질을 한정하는 기능적 방식은 상동성 유기체의 상응하는 단백질 간의 아미노산 변화의 정규화된 빈도를 분석하는 것이다 (참조: Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York (1979)). 상기 분석에 따라, 아미노산 그룹은 그룹 내 아미노산이 우선적으로 서로 교환하는 경우 및 따라서 전체 단백질 구조에 대한 이들의 영향에서 서로 가장 유사한 경우 한정될 수 있다 (참조: 상기 Schulz, G. E. and Schirmer, R. H.). 보존성 돌연변이의 비제한적인 예는 아미노산, 예를 들어, 라이신의 아르기닌으로의 아미노산 치환 및 그 반대의 아미노산 치환을 포함하여 양전하가 유지될 수 있고; 글루탐산의 아르파르트산으로의 아미노산 치환 및 그 반대의 아미노산 치환을 포함하여 음전하가 유지될 수 있고; 세린의 트레오닌으로의 아미노산 치환을 포함하여 유리 -OH가 유지될 수 있고; 글루타민의 아스파라긴으로의 아미노산 치환을 포함하여 유리 $-NH_2$가 유지될 수 있다.

[0186] 본원에서 상호교환적으로 사용된 용어 "암호화 서열" 또는 "단백질 암호화 서열"은 단백질을 암호화하는 폴리뉴클레오타이드 분절을 언급한다. 영역 또는 서열은 개시 코돈에 의해 5' 말단에 보다 근접하고 정지 코돈과 함께 3' 말단에 보다 근접하여 결합된다. 암호화 서열은 또한 개방 판독 프레임으로 언급될 수 있다.

[0187] 본원에 사용된 바와 같은 용어 "데아미나제" 또는 "데아미나제 도메인"은 탈아민화 반응을 촉매하는 단백질 또는 효소를 언급한다. 일부 구현예에서, 데아미나제는 아데닌의 하이포크산틴으로의 가수분해 탈아민화를 촉매하는 아데노신 데아미나제이다. 일부 구현예에서, 데아미나제는 아데노신 또는 아데닌 (A)의 이노신 (I)으로의 가수분해 탈아민화를 촉매하는 아데노신 데아마나제이다. 일부 구현예에서, 데아미나제 또는 데아미나제 도메인은 아데노신 또는 데옥시아데노신의 이노신 또는 데옥시이노신 각각으로의 가수분해 탈아민화를 촉매하는 아데노신 데아미나제이다. 일부 구현예에서, 아데노신 데아미나제는 데옥시리보핵산 (DNA)에서 아데노신의 가수분해 탈아민화를 촉매한다. 본원에 제공된 아데노신 데아미나제 (예를 들어, 가공된 아데노신 데아미나제, 변화된 아데노신 데아미나제)는 임의의 유기체, 예를 들어, 세균으로부터 기원할 수 있다. 일부 구현예에서, 아데노신 데아미나제는 세균, 예를 들어, 이. 콜리 (*E. coli*), 에스. 아우레우스 (*S. aureus*), 에스. 타이피. (*S. typhi*), 에스. 푸트레파시엔스 (*S. putrefaciens*), 에이치. 인플루엔자 (*H. influenzae*), 또는 씨. 크레슨투스 (*C. crescentus*)로부터 기원한다.

[0188] 일부 구현예에서, 아데노신 데아미나제는 TadA 데아미나제이다. 일부 구현예에서, TadA 데아미나제는 TadA*7.10 변이체이다. 일부 구현예에서, TadA*7.10 변이체는 TadA*8이다. 일부 구현예에서, 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 구현예에서, 데아미나제 또는 데아미나제 도메인은 인간, 침팬지, 고릴라, 몽키, 소, 개, 래트 또는 마우스와 같은 유기체 기원의 천연적으로 발생하는 데아미나제의 변이체이다. 일부 구현예에서, 데아미나제 또는 데아미나제 도메인은 천연적으로 발생하지 않는다. 예를 들어, 일부 구현예에서, 데아미나제 또는 데아미나제 도메인은 천연적으로 발생하는 데아미나제와 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. 예를 들어, 데아미나제 도메인은 국제 PCT 출원 번호 PCT/2017/045381 (WO 2018/027078) 및 PCT/US2016/058344 (WO 2017/070632)에 기재되어 있고, 이의 각각은 이의 전문이 본원에 참조로 포함된다. 또한 문헌 (참조: 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)), and 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 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[0189] "검출 가능한 표지"란 관심 대상의 분자에 연결된 경우 분광측정, 광화학적, 생화학적, 면역화학적 또는 화학적 수단을 통해 후자가 검출되도록 하는 조성물을 의미한다. 예를 들어, 유용한 표지는 방사성 동위원소, 자기 비드, 금속성 비드, 콜로이드성 입자, 형광성 염료, 전자-밀도 시약, 효소 (예를 들어, 효소-연결된 면역흡착 검정 (ELISA)에서 통상적으로 사용됨), 비오틴, 디곡시게닌, 또는 합텐을 포함한다.

[0190] "질환"이란 세포, 조직 또는 기관의 정상 기능을 손상시키거나 방해하는 임의의 병태 또는 장애를 의미한다. 질환의 예는 레트 증후군을 포함한다.

[0191] 본원에 사용된 바와 같은 용어 "유효량"은 생물학적 반응을 유발하기에 충분한 생물학적 활성제의 양을 언급한다. 일부 구현예에서, 효과량은 치료받지 않은 환자에 비해 질환의 증상을 개선하는데 필요한 양이다. 질환의 치료학적 치료를 위해 본원에 기재된 방법 및 용도를 수행하는데 사용되는 유효량의 활성제(들)는 투여 방식, 대상체의 연령, 체중 및 일반적인 건강에 따라 달라진다. 궁극적으로, 담당 의사 또는 수의사는 적당한 양 및 투여 용법을 결정할 것이다. 상기 양은 "유효"량으로서 언급된다. 하나의 구현예에서, 유효량은 세포 (예를 들어, 시험관내 세포 또는 생체내 세포)에서 관심 대상의 유전자에 변경을 도입하기에 충분한 본원에 기재된 염기 편집기 (예를 들어, 프로그래밍 가능한 DNA 결합 단백질, 핵염기 편집기 및 gRNA를 포함하는 융합 단백질)의 양이다. 하나의 구현예에서, 유효량은 치료학적 효과를 달성하기 위해 (예를 들어, 레트 증후군 또는 이의 증상 또는 병태를 감소시키거나 제어하기 위해) 요구되는 염기 편집기의 양이다. 상기 치료 효과는 대상체, 조직 또는 기관의 모든 세포에서 Mecp2를 변경하기에 충분할 필요는 없지만 대상체, 조직 또는 기관에 존재하는 세포의 약 1%, 5%, 10%, 25%, 50%, 75% 이상에서 Mecp2를 변경하는 데만 필요하다. 하나의 구현예에서, 유효량은 레트 증후군의 하나 이상의 증상을 개선하기에 충분하다.

[0192] 일부 구현예에서, 본원에 제공된 융합 단백질, 예를 들어, nCas9 도메인 및 데아미나제 도메인 (예를 들어, 아데노신 데아미나제)을 포함하는 핵염기 편집기의 유효량은 본원에 기재된 핵염기 편집기에 의해 특이적으로 결합되고 편집되는 표적 부위의 편집을 유도하기에 충분한 융합 단백질의 양을 언급한다. 당업자에 의해 인지되는 바와 같이, 제제, 예를 들어, 융합 단백질, 뉴클레아제, 하이브리드 단백질, 단백질 이량체, 단백질 (또는 단백질 이량체) 및 폴리뉴클레오타이드의 복합체, 또는 폴리뉴클레오타이드의 유효량은 다양한 인자, 예를 들어, 목적하는 생물학적 반응, 예를 들어, 특이적 대립유전자, 게놈, 또는 편집될 표적 부위, 표적화된 세포 또는 조직 및/또는 사용되는 제제에 따라 달라질 수 있다.

[0193] "단편"이란 폴리펩타이드 또는 핵산 분자 부분을 의미한다. 상기 부분은 바람직하게는, 참조 핵산 분자 또는 폴리펩타이드의 전체 길이의 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 또는 90%를 포함한다. 단편은 10, 20, 30, 40, 50, 60, 70, 80, 90, 또는 100, 200, 300, 400, 500, 600, 700, 800, 900, 또는 1000개 뉴클레오타이드 또는 아미노산을 포함할 수 있다.

[0194] "가이드 RNA" 또는 "gRNA"란 표적 서열에 특이적일 수 있고 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 단백질 (예를 들어, Cas9 또는 Cas12)과 복합체를 형성할 수 있는 폴리뉴클레오타이드를 의미한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 가이드 RNA (gRNA)이다. gRNA는 2개 이상의 RNA의 복합체로서 또는 단일 RNA 분자로서 존재할 수 있다. 단일 RNA 분자로서 존재하는 gRNA는 단일-가이드 RNA (sgRNA)로서 언급될 수 있지만, "gRNA"는 단일 분자로서 또는 2개 이상의 분자의 복합체로서 존재하는 가이드 RNA를 언급하는데 상호교환적으로 사용된다. 전형적으로, 단일 RNA 종으로서 존재하는 gRNA는 2개의 도메인을 포함한다: (1) 표적 핵산과 상동성을 공유하고 (예를 들어, Cas9 또는 Cas12 복합체의 표적으로의 결합을 지시하는) 도메인; 및 (2) Cas9 또는 Cas12 단백질에 결합하는 도메인. 일부 구현예에서, 도메인 (2)은 tracrRNA로서 공지된 서열에 상응하고, 스템-루프 구조를 포함한다. 예를 들어, 일부 구현예에서, 도메인 (2)은 이의 전체 내용이 참조로 본원에 포함된 문헌 (참조: Jinek *et al.*, Science 337:816-821(2012))에 제공된 바와 같은 tracrRNA와 동일하거나 상동성이다. gRNA의 다른 예 (예를 들어, 도메인 2를 포함하는 것들)는 "Switchable Cas9 Nucleases and Uses Thereof" 표제의 2013년 9월 6일자로 출원된 미국 가특허 출원 U.S.S.N 제61/874,682호 및 "Delivery System For Functional Nucleases" 표제의 2013년 9월 6일자로 출원된 미국 가특허 출원 U.S.S.N. 제61/874,746호에서 찾을 수 있고, 각각의 전체 내용은 이들의 전문이 본원에 참조로 포함된다. 일부 구현예에서, gRNA는 2개 이상의 도메인 (1) 및 (2)를 포함하고, "연장된 gRNA"로서 언급될 수 있다. 연장된 gRNA는 본원에 기재된 바와 같이, 2개 이상의 Cas9 또는 Cas12 단백질에 결합하고 2개 이상의 특유한 영역에서 표적 핵산에 결합할 것이다. gRNA는 표적 부위와 상보체를 형성하는 뉴클레오티드 서열을 포함하고, 이는 뉴클레아제/RNA 복합체의 상기 표적 부위로의 결합을 매개하여 뉴클레아제:RNA 복합체의 서열 특이성을 제공한다.

[0195] "이종이량체"란 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.

[0196] "하이드리드화"는 상보적 핵염기 간의 왓슨-크릭, 후그스틴 또는 역의 (reversed) 후그스틴 수소 결합일 수 있는 수소 결합을 의미한다. 예를 들어, 아데닌 및 티민은 수소 결합의 형성을 통해 쌍을 형성하는 상보적 핵염기이다.

[0197] "증가시킨다"란 적어도 10%, 25%, 50%, 75%, 또는 100%의 양성 변경을 의미한다.

[0198] 용어 "염기 복구의 저해제" 또는 "IBR"은 핵산 복구 효소, 예를 들어, 염기 절제 복구 효소의 활성을 저해할 수 있는 단백질을 언급한다. 일부 구현예에서, IBR은 이노신 염기 절제 복구 (BER)의 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 구현예에서, IBR은 Endo V 또는 hAAG의 저해제이다. 일부 구현예에서, IBR은 촉매 불활성 EndoV 또는 촉매 불활성 hAAG이다. 일부 구현예에서, 염기 복구 저해제는 EndoV 또는 hAAG의 저해제이다. 일부 구현예에서, 염기 복구 저해제는 촉매 불활성 EndoV 또는 촉매 불활성 hAAG이다.

[0199] 일부 구현예에서, 염기 복구 저해제는 우라실 글리코실라제 저해제 (UGI)이다. UGI는 우라실-DNA 글리코실라제 염기-절제 복구 효소를 저해할 수 있는 단백질을 언급한다. 일부 구현예에서, UGI 도메인은 야생형 UGI 또는 야생형 UGI의 단편을 포함한다. 일부 구현예에서, 본원에 제공된 UGI 단백질은 UGI 또는 UGI 단편과 상동성인 UGI 및 단백질의 단편을 포함한다. 일부 구현예에서, 염기 복구 저해제는 이노신 염기 절제 복구의 저해제이다. 일부 구현예에서, 염기 복구 저해제는 "촉매 불활성 이노신 특이적 뉴클레아제" 또는 "데드 이노신 특이적 뉴클레아제"이다. 임의의 특정 이론에 국한시키고자 하는 것 없이, 촉매 불활성 이노신 글리코실라제 (예를 들어, 알킬 아데닌 글리코실라제 (AAG))는 이노신에 결합할 수 있지만 무염기 부위를 생성할 수 없거나 이노신을 제거할 수 있어 새롭게 형성된 이노신 잔기를 DNA 손상/복구 기전으로부터 입체적으로 차단할 수 있다. 일부 구현예에서, 촉매 불활성 이노신 특이적 뉴클레아제는 핵산 내 이오신에 결합할 수 있지만 핵산을 절단하지 못한다. 비제한적 예시의 촉매 불활성 이노신 특이적 뉴클레아제는 예를 들어, 인간으로부터의 촉매 불활성 알킬 아데노신 글리코실라제 (AAG 뉴클레아제), 및 예를 들어, 이. 콜리로부터의 촉매 불활성 엔도뉴클레아제 V (EndoV 뉴클레아제)를 포함한다. 일부 구현예에서, 촉매 불활성 AAG 뉴클레아제는 또 다른 AAG 뉴클레아제 내 E125Q 돌연변이 또는 상응하는 돌연변이를 포함한다.

[0200] "인테인 (intein)"은 그 자체를 절개할 수 있고 단백질 스플라이싱으로서 공지된 공정에서 나머지 단편 (익스테인 (extein))을 펩타이드 결합으로 연결할 수 있는 단백질의 단편이다. 인테인은 또한 "단백질 인트론"으로서

언급된다. 단백질 자체를 절개하고 나머지 부분을 연결하는 인테인의 과정은 본원에서 "단백질 스플라이싱" 또는 "인테인-매개된 단백질 스플라이싱"으로 칭한다. 일부 구현예에서, 전구체 단백질의 인테인 (인테인-매개된 단백질 스플라이싱 전 단백질을 함유하는 인테인)은 2개의 유전자로부터 기원한다. 상기 인테인은 본원에서 스플릿 인테인 (예를 들어, 스플릿 인테인-N 및 스플릿 인테인-C)으로서 언급된다. 예를 들어, 시아노박테리아에서 DNA 폴리머라제 III의 촉매 서브유닛인 DnaE는 2개의 별도의 유전자 dnaE-n 및 dnaE-c에 의해 암호화된다. dnaE-n 유전자에 의해 암호화된 인테인은 본원에서 "인테인-N"으로서 언급될 수 있다. dnaE-c 유전자에 의해 암호화된 인테인은 본원에서 "인테인-C"로서 언급될 수 있다.

[0201] 다른 인테인 시스템이 또한 사용될 수 있다. 예를 들어, dnaE 인테인, Cfa-N (예를 들어, 스플릿 인테인-N) 및 Cfa-C (예를 들어, 스플릿 인테인-C) 인테인 쌍을 기반으로 하는 합성 인테인이 기재되어 있다 (예를 들어, 문헌 (참조: Stevens *et al.*, J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5), 본원에 참조로 포함됨). 본원 개시내용에 따라 사용될 수 있는 인테인 쌍의 비제한적인 예는 다음을 포함한다: Cfa DnaE 인테인, Ssp GyrB 인테인, Ssp DnaX 인테인, Ter DnaE3 인테인, Ter ThyX 인테인, Rma DnaB 인테인 및 Cne Prp8 인테인 (예를 들어, 본원에 참조로 포함된 미국 특허 제8,394,604호에 기재되어 있음).

[0202] 인테인의 예시적인 뉴클레오타이드 및 아미노산 서열이 제공된다.

[0203] **DnaE 인테인-N DNA:**

TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGT
GGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTCAGCCAG
TTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGTCTC
ATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAAT

[0204] CTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

[0205] **DnaE 인테인-N 단백질:**

CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDR

[0206] GEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

[0207] **DnaE 인테인-C DNA:**

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGA

[0208] TATTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT

[0209] **인테인-C:** MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

[0210] **Cfa-N DNA:**

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTGT
CGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACAGCCCA
TTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGCATC
ATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCCAATAGATGAGAT

[0211] ATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

[0212] **Cfa-N 단백질:**

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSI

[0213] IRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGATAAT
ATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACAACTTCC

[0214] **Cfa-C DNA:** TTCTCAAGAACGGTCTCGTAGCCAGCAAC

[0215] **Cfa-C 단백질:**

[0216] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

[0217] 인테인-N 및 인테인-C는 각각 스플릿 Cas9의 N-말단 부분 및 스플릿 Cas9의 C-말단 부분의 연결을 위해 스플릿 Cas9의 N-말단 부분 및 스플릿 Cas9의 C-말단 부분에 융합될 수 있다. 예를 들어, 일부 구현예에서, 인테인-N은 스플릿 Cas9의 N-말단 부분의 C-말단에 융합되어, 즉, N--[스플릿 Cas9의 N-말단 부분]-[인테인-N]--C의 구조를 형성한다. 일부 구현예에서, 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-매개된 단백질의 기전은 당업계에 공지되어 있고, 예를 들어, 본원에 참조로 포함된 문헌 (참조: Shah *et al.*, Chem Sci. 2014; 5(1):446-461)에 기재되어 있다. 인테인을 디자인하고 사용하기 위한 방법은 당업계에 공지되어 있고, 예를 들어, WO2014004336, WO2017132580, US20150344549, 및 US20180127780에 기재되어 있고, 이들 각각은 이들의 전문이 본원에 참조로 포함된다.

[0218] 용어 "단리된", "정제된" 또는 "생물학적으로 순수한"은 이의 고유 상태에서 발견된 바와 같이 정상적으로 여기에 수반되는 성분들로부터 다양한 정도로 유리된 물질을 언급한다. "단리물"은 본래의 공급원 또는 주변으로부터 분리 정도를 지칭한다. "정제한다"는 단리보다 높은 분리 정도를 지칭한다. "정제된" 또는 "생물학적으로 순수한" 단백질은 다른 물질이 상당히 제거되어 임의의 불순물이 단백질의 생물학적 성질에 실질적으로 영향을 미치지 않거나 다른 부작용을 유발하지 않는다. 즉, 본원에 기재된 핵산 또는 펩타이드는 재조합 DNA 기술에 의해 생성된 경우 세포 물질, 바이러스 물질 또는 배양 배지, 또는 화학적으로 합성된 경우 화학적 전구체 또는 다른 화학물질이 상당히 제거된 경우 정제된다. 순도 및 균질성은 전형적으로 분석 화학 기술, 예를 들어, 폴리아크릴아미드 겔 전기영동 또는 고성능 액체 크로마토그래피를 사용하여 결정된다. 용어 "정제된"은 핵산 또는 단백질이 전기영동 겔에서 필수적으로 하나의 밴드를 생성함을 지칭할 수 있다. 변형, 예를 들어, 인산화 또는 당화에 적용될 수 있는 단백질에 대해, 상이한 변형은 상이한 단리된 단백질을 생성할 수 있고 이들은 별도로 정제될 수 있다.

[0219] "단리된 폴리뉴클레오타이드"란 본원에 기재된 핵산 분자가 유래된 천연적으로 발생하는 유기체의 게놈에서 유전자를 플랭킹하는 유전자가 제거된 핵산 (예를 들어, DNA)을 의미한다. 따라서, 용어는, 예를 들어, 벡터로; 자가 복제 플라스미드 또는 바이러스; 또는 원핵생물 또는 진핵생물의 게놈 DNA로 혼입되거나, 다른 서열과 무관하게 별도의 분자 (예를 들어, PCR 또는 제한 엔도뉴클레아제 분해에 의해 생성된 cDNA 또는 게놈 또는 cDNA 단편)로서 존재하는 재조합 DNA를 포함한다. 추가로, 용어는 추가의 폴리펩타이드 서열을 암호화하는 하이브리드 유전자의 일부인 재조합 DNA 뿐만 아니라 DNA 분자로부터 전사된 RNA 분자를 포함한다.

[0220] "단리된 폴리펩타이드"란 천연적으로 여기에 수반되는 성분으로부터 분리된 본원에 개시된 폴리펩타이드를 의미한다. 전형적으로, 폴리펩타이드는 단백질 및 천연적으로 연합된 천연적으로 발생하는 유기 분자로부터 적어도 60 중량%로 제거된 경우 단리된다. 바람직하게는, 제제는 적어도 75 중량%, 보다 바람직하게 적어도 90 중량%, 및 가장 바람직하게 적어도 99 중량%의 본원에 기재된 폴리펩타이드이다. 본원 개시내용의 단리된 폴리펩타이드는 예를 들어, 상기 폴리펩타이드를 암호화하는 재조합 핵산의 발현에 의해; 또는 상기 단백질을 화학적으로 합성함에 의해 천연 공급원으로부터의 추출에 의해 수득될 수 있다. 순도는 임의의 적당한 방법, 예를 들어, 컬럼 크로마토그래피, 폴리아크릴아미드 겔 전기영동 또는 HPLC 분석에 의해 측정될 수 있다.

[0221] 본원에 사용된 바와 같은 용어 "링커"는 공유 링커 (예를 들어, 공유 결합), 비-공유 링커, 화학적 그룹, 또는 2개의 분자 또는 모이어티를 연결하는 분자, 예를 들어, 단백질 복합체 또는 리보뉴클레오 복합체의 2개의 성분, 또는 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 (예를 들어, dCas9) 및 데아미나제 도메인 (예를 들어, 아데노신 데아미나제)과 같은 융합 단백질의 2개의 도메인을 언급할 수 있다. 링커는 염기 편집기 시스템의 상이한 성분 또는 성분의 상이한 부분을 연결할 수 있다. 예를 들어, 일부 구현예에서, 링커는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 가이드 폴리뉴클레오타이드 결합 도메인, 및 데아미나제의 촉매 도메인을 연결할 수 있다. 일부 구현예에서, 링커는 CRISPR 폴리펩타이드와 데아미나제를 연결할 수 있다. 일부 구현예에서, 링커는 Cas9와 데아미나제를 연결할 수 있다. 일부 구현예에서, 링커는 dCas9와 데아미나제를 연결할 수 있다. 일부 구현예에서, 링커는 nCas9와 데아미나제를 연결할 수 있다. 일부 구현예에서, 링커는 가이드 폴리뉴클레오타이드와 데아미나제를 연결할 수 있다. 일부 구현예에서, 링커는 염기 편집기 시스템의 탈아민화 성분과 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 성분을 연결할 수 있다. 일부 구현예에서, 링커는 염기 편집기 시스템의 탈아민화 성분의 RNA-결합 부분과 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 성분을 연결할 수 있다. 일부 구현예에서,

링커는 염기 편집기 시스템의 탈아민화 성분의 RNA-결합 부분과 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 성분의 RNA-결합 부분을 연결할 수 있다. 링커는 2개의 그룹, 분자 또는 다른 모이어티 사이에 위치하거나 이에 의해 플랭킹되고, 공유 결합 또는 비-공유 상호작용을 통해 각각 하나에 연결됨에 따라 2개를 연결할 수 있다. 일부 구현예에서, 링커는 유기 분자, 그룹, 중합체, 또는 화학적 모이어티일 수 있다. 일부 구현예에서, 링커는 폴리뉴클레오타이드일 수 있다. 일부 구현예에서, 링커는 DNA 링커일 수 있다. 일부 구현예에서, 링커는 RNA 링커일 수 있다. 일부 구현예에서, 링커는 리간드에 결합할 수 있는 압타머를 포함할 수 있다. 일부 구현예에서, 리간드는 탄수화물, 펩타이드, 단백질 또는 핵산일 수 있다. 일부 구현예에서, 링커는 리보스위치로부터 유래할 수 있는 압타머를 포함할 수 있다. 압타머가 유래하는 리보스위치는 테오필린 리보스위치, 티아민 피로포스페이트 (TPP) 리보스위치, 아데노신 코발라민 (AdoCbl) 리보스위치, S-아데노실 메티오닌 (SAM) 리보스위치, SAH 리보스위치, 플라빈 모노뉴클레오타이드 (FMN) 리보스위치, 테트라하이드로폴레이트 리보스위치, 라이신 리보스위치, 글라이신 리보스위치, 퓨린 리보스위치, GlmS 리보스위치, 또는 프레-쿠에오신1 (PreQ1) 리보스위치로부터 선택될 수 있다. 일부 구현예에서, 링커는 폴리펩타이드 또는 단백질 도메인, 예를 들어, 폴리펩타이드 리간드에 결합된 압타머를 포함할 수 있다. 일부 구현예에서, 폴리펩타이드 리간드는 K 상동성 (KH) 도메인, MS2 코트 단백질 도메인, PP7 코트 단백질 도메인, SfMu Com 코트 단백질 도메인, 무균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인지 모티프일 수 있다. 일부 구현예에서, 폴리펩타이드 리간드는 염기 편집기 시스템 성분의 일부일 수 있다. 예를 들어, 핵염기 편집 성분은 데아미나제 도메인 및 RNA 인지 모티프를 포함할 수 있다.

[0222] 일부 구현예에서, 링커는 아미노산 또는 다수의 아미노산 (예를 들어, 펩타이드 또는 단백질)일 수 있다. 일부 구현예에서, 링커는 약 5-100개 아미노산 길이, 예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 또는 90-100개 아미노산 길이일 수 있다. 일부 구현예에서, 링커는 약 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 또는 450-500개 아미노산 길이일 수 있다. 더 길거나 짧은 링커가 또한 고려될 수 있다.

[0223] 일부 구현예에서, 링커는 Cas9 뉴클레아제 도메인을 포함하는 RNA-프로그래밍 가능한 뉴클레아제의 gRNA 결합 도메인과 핵산 편집 단백질 (예를 들어, 아데노신 데아미나제)의 촉매 도메인을 연결한다. 일부 구현예에서, 링커는 dCas9와 핵산 편집 단백질을 연결한다. 예를 들어, 링커는 2개의 그룹, 분자 또는 다른 모이어티 사이에 위치하거나 이에 의해 플랭킹되고 공유 결합을 통해 각각 하나에 연결됨에 따라 2개를 연결한다. 일부 구현예에서, 링커는 아미노산 또는 다수의 아미노산 (예를 들어, 펩타이드 또는 단백질)이다. 일부 구현예에서, 링커는 유기 분자, 그룹, 중합체, 또는 화학적 모이어티이다. 일부 구현예에서, 링커는 5-200개 아미노산 길이, 예를 들어, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 35, 45, 50, 55, 60, 60, 65, 70, 70, 75, 80, 85, 90, 90, 95, 100, 101, 102, 103, 104, 105, 110, 120, 130, 140, 150, 160, 175, 180, 190, 또는 200개 아미노산 길이이다. 더 길거나 더 짧은 링커가 또한 고려된다. 일부 구현예에서, 링커는 XTEN 링커로도 언급될 수 있는 아미노산 서열 SGSETPGTSESATPES를 포함한다. 일부 구현예에서, 링커는 아미노산 서열 SGGS를 포함한다. 일부 구현예에서, 링커는 $(SGGS)_n$, $(GGGS)_n$, $(GGGGS)_n$, $(G)_n$, $(EAAAK)_n$, $(GGS)_n$, SGSETPGTSESATPES, 또는 $(XP)_n$ 모티프, 또는 이들의 임의의 조합을 포함하고, 여기서, n은 독립적으로 1 내지 30의 정수이고, X는 임의의 아미노산이다. 일부 구현예에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15이다. 일부 구현예에서, 링커는 다수의 프롤린 잔기를 포함하고, 5-21, 5-14, 5-9, 5-7개 아미노산 길이, 예를 들어, PAPAP, PAPAPA, PAPAPAP, PAPAPAPA, $P(AP)_4$, $P(AP)_7$, $P(AP)_{10}$이다 상기 프롤린-풍부 링커는 또한 "강성" 링커로 칭한다.

[0224] 일부 구현예에서, 염기 편집기의 도메인은 SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGS, 또는 GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS의 아미노산 서열을 포함하는 링커를 통해 융합된다.

[0225] 일부 구현예에서, 염기 편집기의 도메인은 또한 XTEN 링커로서 언급될 수 있는 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다. 일부 구현예에서, 링커는 24개 아미노산 길이이다. 일부 구현예에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPES를 포함한다. 일부 구현예에서, 링커는 40개 아미노산 길이이다. 일부 구현예에서, 링커는 아미노산 서열SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS를 포함한다. 일부 구현예에서, 링커는 64개 아미노산 길이이다. 일부 구현예에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGS SGGS를 포함한다.

[0226] 일부 구현예에서, 링커는 92개 아미노산 길이이다. 일부 구현예에서, 링커는 아미노산 서열

PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGSEPATS를 포함한다.

[0227] 본원에 사용된 바와 같은 용어 "돌연변이"는 서열, 예를 들어, 핵산 또는 아미노산 서열 내 잔기의 또 다른 잔기로의 치환, 서열 내 하나 이상의 잔기의 결실 또는 삽입을 언급한다. 돌연변이는 전형적으로 본원에서 본래의 잔기에 이어서 서열 내 잔기의 위치를 표시하고 새롭게 치환된 잔기를 확인하여 기재된다. 본원에 제공된 아미노산 치환 (돌연변이)을 제조하기 위한 다양한 방법은 당업계에 널리 공지되어 있고 예를 들어, 문헌 (참조: Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))에 의해 제공된다. 일부 구현예에서, 현재 개시된 염기 편집기는 의도되지 않은 점 돌연변이와 같은, 상당수의 의도되지 않은 돌연변이를 생성하지 않고 핵산 (예를 들어, 대상체의 게놈 내 핵산) 내 점 돌연변이와 같은 "의도된 돌연변이"를 효율적으로 생성할 수 있다. 일부 구현예에서, 의도된 돌연변이는 의도된 돌연변이를 생성하도록 특이적으로 디자인된, 가이드 폴리뉴클레오타이드 (예를 들어, gRNA)에 결합하는 특이적 염기 편집기 (예를 들어, 아데노신 염기 편집기)에 의해 생성되는 돌연변이이다. 일반적으로, 서열 (예를 들어, 본원에 기재된 바와 같은 아미노산 서열)에서 만들어지거나 동정된 돌연변이는 참조 (또는 야생형) 서열, 즉, 돌연변이를 함유하지 않는 서열에 대한 넘버링된다. 당업자는 참조 서열에 비해 아미노산 및 핵산 서열에서 돌연변이의 위치를 결정하는 방법을 용이하게 이해할 것이다.

[0228] 용어 "비-보존성 돌연변이"는 상이한 그룹 간 아미노산 치환, 예를 들어, 라이신의 트립토판으로의 치환, 또는 페닐알라닌의 세린으로의 치환 등을 포함한다. 이 경우에, 비-보존성 아미노산 치환은 기능성 변이체의 생물학적 활성을 방해하지 않도록 또는 저해하지 않도록 하는 것이 바람직하다. 비-보존성 아미노산 치환은 기능성 변이체의 생물학적 활성을 증진시켜 기능성 변이체의 생물학적 활성이 야생형 단백질과 비교하여 증가될 수 있다.

[0229] 용어 "핵 국소화 서열", "핵 국소화 신호" 또는 "NLS"는 세포 핵으로의 단백질의 도입을 촉진시키는 아미노산 서열을 언급한다. 핵 국소화 서열은 당업계에 공지되어 있고, 예를 들어, 플랭크 등 (Plank *et al.*)의 2000년 11월 23일에 출원된 국제 PCT 출원 번호 PCT/EP2000/011690에 기재되고 2001년 5월 31일에 WO/2001/038547로 공개되어 있으며, 이의 내용이 예시적인 핵 국소화 서열의 이들의 기재를 위해 본원에 참조로 포함된다. 다른 구현예에서, NLS는 예를 들어, 문헌 (참조: Koblan et al., Nature Biotech. 2018 doi:10.1038/nbt.4172)에 기재된 최적화된 NLS이다. 일부 구현예에서, NLS는 하기의 아미노산 서열을 포함한다:

[0230] KRTADGSEFESPKKKRKV,

[0231] KRPAATKKAGQAKKKK,

[0232] KKTELQTTNAENKTKKL,

[0233] KRGINDRNFWRGENGRKTR,

[0234] RKSGKIAAIVVKRPRK,

[0235] PKKKRKV, 또는

[0236] MDSLLMNRRKFLYQFKNVRWAKGRRETYLC.

[0237] 본원에 사용된 바와 같은 용어 "핵산" 및 "핵산 분자"는 핵염기 및 산성 모이어티, 예를 들어, 뉴클레오사이드, 뉴클레오타이드 또는 뉴클레오타이드의 중합체를 포함하는 화합물을 언급한다. 전형적으로, 중합체 핵산, 예를 들어, 3개 이상의 뉴클레오타이드를 포함하는 핵산 분자는 선형 분자이고, 여기서, 인접 뉴클레오타이드는 포스포디에스테르 연결을 통해 서로 연결된다. 일부 구현예에서, "핵산"은 개별 핵산 잔기 (예를 들어. 뉴클레오타이드 및/또는 뉴클레오사이드)를 언급한다. 일부 구현예에서, "핵산"은 3개 이상의 개별 뉴클레오타이드 잔기를 포함하는 올리고뉴클레오타이드 쇄를 언급한다. 본원에 사용된 바와 같은 용어 "올리고뉴클레오타이드" 및 "폴리뉴클레오타이드"는 뉴클레오타이드 중합체 (예를 들어, 적어도 3개의 뉴클레오타이드 스트링)를 언급하기 위해 상호교환적으로 사용될 수 있다. 일부 구현예에서, "핵산"은 단일 및/또는 이중 가닥 DNA 뿐만 아니라 RNA를 포함한다. 핵산은 예를 들어, 게놈, 전사체, mRNA, tRNA, rRNA, siRNA, snRNA, 플라스미드, 코스미드, 염색체, 염색분체, 또는 다른 천연적으로 발생하는 핵산 분자와 관련하여 천연적으로 존재할 수 있다. 한편, 핵산 분자는 비천연적으로 발생하는 분자, 예를 들어, 재조합 DNA 또는 RNA, 인공 염색체, 가공된 게놈, 또는 이의 단편, 또는 합성 DNA, RNA, DNA/RNA 하이브리드일 수 있거나, 비-천연적으로 발생하는 뉴클레오타이드 또는 뉴클레오사이드를 포함한다. 추가로, 용어 "핵산", "DNA", "RNA", 및/또는 유사 용어는 핵산 유사체, 예를

들어, 포스포디에스테르 골격과는 다른 것을 갖는 유사체를 포함한다. 핵산은 천연 공급원으로부터 정제될 수 있고, 재조합 발현 시스템을 사용하여 생성되고, 임의로 정제되고, 화학적으로 합성될 수 있다. 경우에 따라, 예를 들어, 화학적으로 합성된 분자의 경우에, 핵산은 화학적으로 변형된 염기 또는 당 및 골격 변형을 갖는 유사체와 같은 뉴클레오사이드 유사체를 포함할 수 있다. 핵산 서열은 달리 지적되지 않는 경우 5'에서 3' 방향으로 제시된다. 일부 구현예에서, 핵산은 천연 뉴클레오사이드 (예를 들어, 아데노신, 티미딘, 구아노신, 시티딘, 우리딘, 데옥시아데노신, 데옥시티미딘, 데옥시구아노신, 및 데옥시시티딘); 뉴클레오사이드 유사체 (예를 들어, 2-아미노아데노신, 2-티오티미딘, 이노신, 피롤로-피리미딘, 3-메틸 아데노신, 5-메틸시티딘, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-요오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-시티딘, C5-메틸시티딘, 2-아미노아데노신, 7-데아자아데노신, 7-데아자구아노신, 8-옥소아데노신, 8-옥소구아노신, O(6)-메틸구아닌, 및 2-티오시티딘); 화학적으로 변형된 염기; 생물학적으로 변형된 염기 (예를 들어, 메틸화된 염기); 인터컬레이팅된 염기; 변형된 당 (예를 들어, 2'-플루오로리보스, 리보스, 2'-데옥시리보스, 아라비노스 및 헥소스); 및/또는 변형된 포스페이트 그룹 (예를 들어, 포스포로티오에이트 및 5'-*N*-포스포르아미디트 결합)이거나 이들을 포함한다.

[0238] 본원에 상호교환적으로 사용되는 용어 "핵염기", "질소성 염기" 또는 "염기"는 뉴클레오사이드를 형성하고 이는 차례로 뉴클레오타이드의 성분인 질소-함유 생물학적 화합물을 언급한다. 핵염기가 염기 쌍을 형성하고 하나가 또 다른 하나에 스택킹하는 능력은 직접적으로 리보핵산 (RNA)과 데옥시리보핵산 (DNA)와 같은 장쇄 나선 구조를 유도한다. 5개 핵염기 - 아데닌 (A), 시토신 (C), 구아닌 (G), 티민 (T), 및 우라실 (U)은 1차 또는 카노니칼로 불리운다. 아데닌 및 구아닌은 퓨린으로부터 유래하고, 시토신, 우라실, 및 티민은 피리미딘으로부터 유래한다. DNA 및 RNA는 또한 변형된 다른 (비-1차) 염기를 함유할 수 있다. 비제한적인 예시적 변형된 핵염기는 하이포크산틴, 크산틴, 7-메틸구아닌, 5,6-디하이드로우라실, 5-메틸시토신 (m5C), 및 5-하이드로메틸시토신을 포함할 수 있다. 하이포크산틴 및 크산틴은 돌연변이 유발제를 통해 생성될 수 있고 이들 둘 다는 탈아민화 (아민 그룹의 카보닐 그룹으로의 대체)를 통해서 생성될 수 있다. 하이포크산틴은 아데닌으로부터 변형될 수 있다. 크산틴은 구아닌으로부터 변형될 수 있다. 우라실은 시토신의 탈아민화로부터 비롯될 수 있다. "뉴클레오사이드"는 핵염기 및 5개 탄소 당 (리보스 또는 데옥시리보스)으로 이루어진다. 뉴클레오사이드의 예는 아데노신, 구아노신, 우리딘, 시티딘, 5-메틸우리딘 (m5U), 데옥시아데노신, 데옥시구아노신, 티미딘, 데옥시우리딘 및 데옥시시티딘을 포함한다. 변형된 핵염기를 갖는 뉴클레오사이드의 예는 이노신 (I), 크산토신 (X), 7-메틸구아노신 (m7G), 디하이드로우리딘 (D), 5-메틸시티딘 (m5C), 및 슈도우리딘 (Ψ)을 포함한다. "뉴클레오타이드"는 핵염기, 5개 탄소 당 (리보스 또는 데옥시리보스) 및 적어도 하나의 포스페이트 그룹으로 이루어진다.

[0239] 용어 "핵산 프로그래밍 가능한 DNA 결합 단백질" 또는 "napDNAbp"는 "폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인"과 상호교환적으로 사용되어 상기 napDNAbp를 특이적 핵산 서열에 가이드하는 가이드 핵산 (예를 들어, gRNA)와 같은 핵산 (예를 들어, DNA 또는 RNA)과 연합된 단백질을 언급할 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 RNA 결합 도메인이다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 Cas9 단백질이다. Cas9 단백질은 Cas9 단백질을, 가이드 RNA와 상보적인 특이적 DNA 서열로 가이드하는 가이드 RNA와 연합될 수 있다. 일부 구현예에서, napDNAbp는 Cas9 도메인, 예를 들어, 뉴클레아제 활성 Cas9, Cas9 닉카제 (nCas9), 또는 뉴클레아제 불활성 Cas9 (dCas9)이다. 핵산 프로그래밍 가능한 DNA 결합 단백질의 비제한적인 예는, 제한 없이, Cas9 (예를 들어, dCas9 및 nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 및 Cas12i를 포함한다. Cas 효소의 비제한적인 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (또한 Csn1 또는 Csx12로서 공지된), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, 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, II형 Cas 이펙터 단백질, V형 Cas 이펙터 단백질, VI형 Cas 이펙터 단백질, CARF, DinG, 이의 상동체, 또는 이의 변형된 또는 가공된 버전을 포함한다. 다른 핵산 프로그램 가능한 DNA 결합 단백질은 또한 본원 개시내용의 범위 내에 있지만, 이들은 구체적으로 본원 개시내용에 열거되지 않을 수 있다. 예를 들어, 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 참조하고, 이의 각각의 전체 내용은 본원에 참조로 포함된다.

[0240] 본원에 사용된 바와 같은 용어 "핵염기 편집 도메인" 또는 "핵염기 편집 단백질"은 RNA 또는 DNA에서 핵염기 변형, 예를 들어, 시토신 (또는 시티딘)의 우라실 (또는 우리딘) 또는 티민 (또는 티미딘)으로의 그리고 아데닌 (또는 아데노신)의 하이포크산틴 (또는 이노신)으로의 탈아민화, 및 비-주형 뉴클레오타이드 부가 및 삽입을 촉매할 수 있는 단백질 또는 효소를 언급한다. 일부 구현예에서, 핵염기 편집 도메인은 데아미나제 도메인 (예를 들어, 아데닌 데아미나제 또는 아데노신 데아미나제)이다. 일부 구현예에서, 핵염기 편집 도메인은 천연적으로 발생하는 핵염기 편집 도메인일 수 있다. 일부 구현예에서, 핵염기 편집 도메인은 천연적으로 발생하는 핵염기 편집 도메인으로부터 가공되거나 변화된 핵염기 편집 도메인일 수 있다. 핵염기 편집 도메인은 임의의 유기체, 예를 들어, 세균, 인간, 침팬지, 고릴라, 몽키, 소, 개, 래트, 또는 마우스로부터 기원할 수 있다.

[0241] 본원에 사용된 바와 같이, "제제를 수득하는"에서와 같이 "수득하는"은 합성, 구매, 또는 다르게 제제를 획득하는 것을 포함한다.

[0242] 본원에 사용된 바와 같은 "환자" 또는 "대상체"는 질환 또는 장애를 앓는 것으로 진단되거나, 질환 또는 장애를 갖거나 발병할 위험에 있거나 질환 또는 장애를 갖거나 발병한 것으로 의심되는 포유동물 대상체 또는 개체를 언급한다. 일부 구현예에서, 용어 "환자"는 질환 또는 장애가 발병할 평균 확률보다 높은 포유동물 대상체를 언급한다. 예시적인 환자는 인간, 비-인간 영장류, 고양이, 개, 돼지, 소, 고양이, 말, 낙타, 라마 (llama), 염소, 양, 설치류 (예를 들어, 마우스, 토끼, 래트 또는 기니아 피그) 및 본원에 기재된 치료요법이 이득이 될 수 있는 다른 포유류일 수 있다. 예시적인 인간 환자는 남성 및/또는 여성일 수 있다.

[0243] "이를 필요로 하는 환자" 또는 "이를 필요로 하는 대상체"는 예를 들어, 레트 증후군 (RTT)이지만 이에 제한되지 않는 질환 또는 장애로 진단되거나, 이의 위험에 처하거나, 이를 갖거나, 가질 것으로 미리 결정되거나, 또는 가질 것으로 의심되는 환자로서 본원에서 언급된다.

[0244] 용어 "병원성 돌연변이", "병원성 변이체", 질환 유발 돌연변이", "질환 유발 변이체", "해로운 돌연변이" 또는 "소인 돌연변이"는 특정 질환 또는 장애에 대한 개체의 민감성 또는 소인을 증가시키는 유전학적 변경 또는 돌연변이를 언급한다. 일부 구현예에서, 병원성 돌연변이는 유전자에 의해 암호화된 단백질에서 적어도 하나의 병원성 아미노산에 의해 치환된 적어도 하나의 야생형 아미노산을 포함한다.

[0245] 본원에 사용된 바와 같은 용어 "약제학적으로 허용되는 담체"는 신체의 하나의 부위 (예를 들어, 전달 부위)로부터 또 다른 부위 (예를 들어, 신체의 기관, 조직 또는 일부)로 화합물을 운반하거나 수송하는데 관여하는, 약제학적으로 허용되는 물질, 조성물 또는 비히클, 예를 들어, 액체 또는 고체 충전제, 희석제, 부형제, 제조 보조제 (예를 들어, 윤활제, 탈크 마그네슘, 칼슘 또는 아연 스테아레이트, 또는 스테아르산), 또는 용매 캡슐화 물질을 의미한다. 약제학적으로 허용되는 담체는 제형의 다른 성분과 상용성이고 대상체의 조직에 해롭지 않다 (예를 들어, 생리학적 상용성, 멸균 또는 생리학적 pH 등)는 의미에서 "허용되는"이다. "부형제," "담체," "약제학적으로 허용되는 담체," "비히클" 등과 같은 용어는 본원에서 상호교환적으로 사용된다.

[0246] 용어 "약제학적 조성물"은 약제학적 용도를 위해 제형화된 조성물을 의미한다.

[0247] 용어 "단백질", "펩타이드", "폴리펩타이드" 및 이들의 문법적 등가물은 본원에서 상호교환적으로 사용되고, 펩타이드 (아미드) 결합에 의해 함께 연결된 아미노산 잔기의 중합체를 언급한다. 상기 용어는 임의의 크기, 구조 또는 기능의 단백질, 펩타이드 또는 폴리펩타이드를 언급한다. 전형적으로, 단백질, 펩타이드 또는 폴리펩타이드는 적어도 3개의 아미노산 길이이다. 단백질, 펩타이드 또는 폴리펩타이드는 개별 단백질 또는 단백질 집합체를 언급할 수 있다. 단백질, 펩타이드 또는 폴리펩타이드 내 하나 이상의 아미노산은 예를 들어, 탄수화물 그룹, 하이드록실 그룹, 포스페이트 그룹, 파르네실 그룹, 이소파르네실 그룹, 지방산 그룹, 접합, 기능성화 또는 다른 변형 등을 위한 링커와 같은 화학적 실체의 부가에 의해 변형될 수 있다. 단백질, 펩타이드 또는 폴리펩타이드는 또한 단일 분자일 수 있거나 다중-분자 복합체일 수 있다. 단백질, 펩타이드, 또는 폴리펩타이드는 단지 천연적으로 발생하는 단백질 또는 펩타이드의 단편일 수 있다. 단백질, 펩타이드 또는 폴리펩타이드는 천연적으로 발생하거나, 재조합이거나 합성이거나 이의 임의의 조합일 수 있다. 본원에 사용된 바와 같은 용어 "융합 단백질"은 적어도 2개의 상이한 단백질로부터의 단백질 도메인을 포함하는 하이브리드 폴리펩타이드를 언급한다. 하나의 단백질은 융합 단백질의 아미노-말단 (N-말단) 부분에 또는 카복시 말단 (C-말단) 단백질에 위치함에 따라서 각각 아미노-말단 융합 단백질 또는 카복시 말단 융합 단백질을 형성할 수 있다. 단백질은 상이

한 도메인, 예를 들어, 핵산 결합 도메인 (예를 들어, 단백질의 결합을 표적 부위로 지시하는 Cas9의 gRNA 결합 도메인) 및 핵산 편집 단백질의 핵산 절단 도메인 또는 촉매 도메인을 포함할 수 있다. 일부 구현예에서, 단백질은 단백질성 부분, 예를 들어, 핵산 결합 도메인을 구성하는 아미노산 서열, 및 유기 화합물, 예를 들어, 핵산 절단 제제로서 작용할 수 있는 화합물을 포함한다. 일부 구현예에서, 단백질은 핵산, 예를 들어, RNA 또는 DNA와 복합체로 존재하거나 이와 연합되어 있다. 본원에 제공된 임의의 단백질은 당업계에 공지된 임의의 방법에 의해 생성될 수 있다. 예를 들어, 본원에 제공된 단백질은 특히 펩타이드 링커를 포함하는 융합 단백질에 적합한 재조합 단백질 발현 및 정제를 통해 생성될 수 있다. 재조합 단백질 발현 및 정제를 위한 방법은 널리 공지되어 있고, 문헌 (참조: Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))에 기재된 것들을 포함하고, 이의 전체 내용은 본원에서 참조로 포함된다.

[0248] 본원에 기재된 폴리펩타이드 및 단백질 (기능성 부분 및 이의 기능성 변이체 포함)은 하나 이상의 천연적으로 발생하는 아미노산 대신 합성 아미노산을 포함할 수 있다. 상기 합성 아미노산은 당업계에 공지되어 있고, 예를 들어, 아미노사이클로헥산 카복실산, 노르류신, α-아미노 n-데칸산, 호모세린, S-아세틸아미노메틸-시스테인, 트랜스-3- 및 트랜스-4-하이드록시프롤린, 4-아미노페닐알라닌, 4-니트로페닐알라닌, 4-클로로페닐알라닌, 4-카복시페닐알라닌, β-페닐세린 β-하이드록시페닐알라닌, 페닐글라이신, α-나프틸알라닌, 사이클로헥실알라닌, 사이클로헥실글라이신, 인돌린-2-카복실산, 1,2,3,4-테트라하이드로이소퀴놀린-3-카복실산, 아미노말론산, 아미노말론산 모노아미드, N'-벤질-N'-메틸-라이신, N',N'-디벤질-라이신, 6-하이드록시라이신, 오르니틴, α-아미노사이클로펜탄 카복실산, α-아미노사이클로헥산 카복실산, α-아미노사이클로헵탄 카복실산, α-(2-아미노-2-노르보르난)-카복실산, α,γ-디아미노부티르산, α,β-디아미노프로피온산, 호모페닐알라닌 및 α-3급-부틸글라이신을 포함한다. 폴리펩타이드 및 단백질은 폴리펩타이드 작제물의 하나 이상의 아미노산의 해독 후 변형과 관련될 수 있다. 해독 후 변형의 비제한적인 예는 인산화, 아세틸화 및 포밀화를 포함하는 아실화, 글리코실화 (N-연결된 및 O-연결된을 포함함), 아미드화, 하이드록실화, 메틸화 및 에틸화를 포함하는 알킬화, 유비퀴틸화, 피롤리돈 카복실산의 부가, 디설파이드 브릿지의 형성, 황화, 미리스토일화, 팔미토일화, 이소프레닐화, 파르네실화, 게라닐화, 글리피화 (glypiation), 지질화 및 요오드화를 포함한다.

[0249] 단백질 또는 핵산과 관련하여 본원에 사용된 바와 같은 용어 "재조합체"는 천연적으로 발생하지 않지만 인간 가공 생성물인 단백질 또는 핵산을 언급한다. 예를 들어, 일부 구현예에서, 재조합 단백질 또는 핵산 분자는 임의의 천연적으로 발생하는 서열과 비교하여 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 또는 적어도 7개 돌연변이를 포함하는 아미노산 또는 뉴클레오타이드 서열을 포함한다.

[0250] "감소시킨다"란 적어도 10%, 25%, 50%, 75%, 또는 100%의 음성 변경을 의미한다.

[0251] "참조"란 표준 또는 대조군 조건을 의미한다. 하나의 구현예에서, 참조는 야생형 또는 건강한 세포이다. 다른 구현예에서 및 제한 없이, 참조는 시험 조건에 적용되지 않거나, 위약 또는 정규 식염수, 배지, 완충액 및/또는 관심 대상의 폴리뉴클레오타이드를 함유하지 않는 대조군 벡터에 적용된 미처리된 세포이다.

[0252] "참조 서열"은 서열 비교용 기준으로서 사용되는 한정된 서열이다. 참조 서열은 특정 서열; 예를 들어, 전장 cDNA 또는 유전자 서열의 분절, 또는 완전한 cDNA 또는 유전자 서열의 서브세트 또는 전체일 수 있다. 폴리펩타이드에 대해, 참조 폴리펩타이드 서열의 길이는 일반적으로 적어도 약 16개 아미노산, 적어도 약 20개 아미노산, 적어도 약 25개 아미노산, 약 35개 아미노산, 약 50개 아미노산 또는 약 100개 아미노산일 것이다. 핵산에 대해, 참조 핵산 서열의 길이는 일반적으로 적어도 약 50개 뉴클레오타이드, 적어도 약 60개 뉴클레오타이드, 적어도 약 75개 뉴클레오타이드, 약 100개 뉴클레오타이드 또는 약 300개 뉴클레오타이드 또는 이에 대한 또는 이들 사이의 임의의 정수일 것이다. 일부 구현예에서, 참조 서열은 관심 대상의 단백질의 야생형 서열이다. 다른 구현예에서, 참조 서열은 야생형 단백질을 암호화하는 폴리뉴클레오타이드 서열이다.

[0253] 용어 "RNA-프로그래밍 가능한 뉴클레아제" 및 "RNA-가이드된 뉴클레아제"는 절단을 위한 표적이 아닌 하나 이상의 RNA(들)와 함께 (예를 들어, ~와 결합하거나 연합되어) 사용된다. 일부 구현예에서, RNA-프로그램 가능한 뉴클레아제는, RNA와 복합체로 있는 경우 뉴클레아제:RNA 복합체로서 언급될 수 있다. 전형적으로, 결합된 RNA(들)는 가이드 RNA (gRNA)로서 언급된다. gRNA는 2개 이상의 RNA 복합체로서 또는 단일 RNA 분자로서 존재할 수 있다. 단일 RNA 분자로서 존재하는 gRNA는 단일-가이드 RNA (sgRNA)로서 언급될 수 있지만 "gRNA"는 단일 분자로서 또는 2개 이상의 분자의 복합체로서 존재하는 가이드 RNA를 언급하기 위해 상호교환적으로 사용된다. 전형적으로, 단일 RNA 종으로서 존재하는 gRNA는 2개의 도메인을 포함한다: (1) 표적 핵산과 상동성을 공유하는 (예를 들어, Cas9 복합체의 표적으로의 결합을 지시하는) 도메인; 및 (2) Cas9 단백질에 결합하는 도메

인. 일부 구현예에서, 도메인 (2)은 tracrRNA로서 공지된 서열에 상응하고, 스템-루프 구조를 포함한다. 예를 들어, 일부 구현예에서, 도메인 (2)은 이의 전체 내용이 참조로 본원에 포함된 문헌 (참조: Jinek *et al.*, Science 337:816-821(2012))에 제공된 바와 같은 tracrRNA와 동일하거나 상동성이다. gRNA의 다른 예 (예를 들어, 도메인 2를 포함하는 것들)는 "Switchable Cas9 Nucleases and Uses Thereof" 표제의 2013년 9월 6일자로 출원된 미국 가특허 출원 U.S.S.N 제61/874,682호 및 "Delivery System For Functional Nucleases" 표제의 2013년 9월 6일자로 출원된 미국 가특허 출원 U.S.S.N. 제61/874,746호에서 찾을 수 있고, 각각의 전체 내용은 이들의 전문이 본원에 참조로 포함된다. 일부 구현예에서, gRNA는 2개 이상의 도메인 (1) 및 (2)를 포함하고, "연장된 gRNA"로서 언급될 수 있다. 예를 들어, 연장된 gRNA는 예를 들어, 본원에 기재된 바와 같이, 2개 이상의 Cas9 단백질에 결합하고 2개 이상의 특유한 영역에서 표적 핵산에 결합한다. gRNA는 표적 부위와 상보체를 형성하는 핵산 서열을 포함하고, 이는 뉴클레아제/RNA 복합체의 상기 표적 부위로의 결합을 매개하여 뉴클레아제:RNA 복합체의 서열 특이성을 제공한다.

[0254] 일부 구현예에서, RNA-프로그래밍 가능한 뉴클레아제는 (CRISPR-연합된 시스템) Cas9 엔도뉴클레아제, 예를 들어, 스트렙토코커스 피오게네스 (Streptococcus pyogenes)로부터의 Cas9 (Csn1)이다 (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).

[0255] RNA-프로그래밍 가능한 뉴클레아제 (예를 들어, Cas9)는 DNA 절단 부위를 표적화하기 위해 RNA:DNA 하이브리드화를 사용하기 때문에, 이들 단백질은 원칙적으로 가이드 RNA에 의해 특정된 임의의 서열에 표적화될 수 있다. 부위-특이적 절단을 위해 (예를 들어, 게놈을 변형시키기 위해) Cas9와 같은 RNA-프로그램 가능한 뉴클레아제를 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; 이들 각각의 전체 내용은 본원에 참조로 포함됨).

[0256] 용어 "단일염기다형성 (SNP)"는 게놈 내 특정 위치에 존재하는 단일 뉴클레오타이드에서의 변화이고, 여기서, 각각의 변화는 집단 내 일부 감지할 수 있는 정도 (예를 들어, > 1%)로 존재한다. 예를 들어, 인간 게놈 내 특정 염기 위치에서, C 뉴클레오타이드는 대부분의 개체에서 나타나지만 소수의 개체에서 나타날 수 있고, 상기 위치는 A에 의해 점유된다. 이것은 상기 특정 위치에서 SNP가 있음을 의미하고 2개의 가능한 뉴클레오타이드 변화, C 또는 A는 상기 위치에 대해 대립유전자인 것으로 일컬어진다. SNP는 질환에 대한 민감성에서의 차이를 뒷받침한다. 질병의 중증도 및 우리의 신체가 치료에 응답하는 방식은 또한 유전학적 변화를 나타낸다. SNP는 유전자의 암호화 영역, 유전자의 비-암호화 영역 내 또는 유전자 간 영역 (유전자 사이의 영역) 내에 존재할 수 있다. 일부 구현예에서, 암호화 서열 내 SNP는 필수적으로 유전학적 코드의 축퇴성으로 인해 생성된 단백질의 아미노산 서열을 변화시키지 않는다. 암호화 영역 내 SNP는 2개의 유형: 인접하거나 인접하지 않은 SNP이다. 유사한 SNP는 단백질 서열에 영향을 주지 않고, 인접하지 않은 SNP는 단백질의 아미노산 서열을 변화시킨다. 인접하지 않은 SNP는 2개 유형: 미스센스 및 넌센스이다. 단백질-암호화 영역 내에 있지 않은 SNP는 여전히 유전자 스플라이싱, 전사 인자 결합, 전령 RNA 분해 또는 비암호화 RNA의 서열에 영향을 미칠 수 있다. 상기 유형의 SNP에 의해 영향을 받은 유전자 발현은 eSNP (발현 SNP)로서 언급되고, 유전자로부터 업스트림 또는 다운스트림에 있을 수 있다. 단일 뉴클레오타이드 변이체 (SNV)는 빈도의 어떠한 제한 없이 단일 뉴클레오타이드에서의 변화이고 체세포에서 발생할 수 있다. 체세포 단일 뉴클레오타이드 변화 (예를 들어, 암과 연관됨)는 또한 단일-뉴클레오타이드 변경으로 불릴 수 있다.

[0257] "특이적으로 결합한다"란 핵산 분자, 폴리펩타이드, 또는 이의 복합체 (예를 들어, 핵산 프로그래밍 가능한 DNA 결합 단백질, 가이드 핵산), 본원 개시내용의 폴리펩타이드 및/또는 핵산 분자를 인지하고 결합하는 화합물 또는 분자를 의미하지만, 이는 샘플, 예를 들어, 생물학적 샘플에서 다른 분자를 실질적으로 인지하지 못하고 결합하지 않는다.

[0258] 본원에 기재된 방법에 유용한 핵산 분자는 본원 개시내용의 폴리펩타이드 또는 이의 단편을 암호화하는 임의의 핵산 분자를 포함한다. 상기 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 전형적으로 실질적

인 동일성을 나타낼 것이다. 내인성 서열과 "실질적인 동일성"을 갖는 폴리뉴클레오타이드는 전형적으로 이중 가닥 핵산 분자의 적어도 하나의 가닥과 하이브리드화할 수 있다. 본원에 기재된 방법에 유용한 핵산 분자는 본원 개시내용의 폴리펩타이드 또는 이의 단편을 암호화하는 임의의 핵산 분자를 포함한다. 상기 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 전형적으로 실질적인 동일성을 나타낼 것이다. 내인성 서열과 "실질적인 동일성"을 갖는 폴리뉴클레오타이드는 전형적으로 이중 가닥 핵산 분자의 적어도 하나의 가닥과 하이브리드화할 수 있다. "하이브리드화한다"란 다양한 엄중 조건하에서 상보적인 폴리뉴클레오타이드 서열 (예를 들어, 본원에 기재된 유전자) 또는 이의 일부 간에 쌍을 이루어 이중 가닥 분자를 형성함을 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).

[0259] 예를 들어, 엄중 염 농도는 통상적으로 약 750 mM 미만의 NaCl 및 75 mM의 삼나트륨 시트레이트, 바람직하게 약 500 mM 미만의 NaCl 및 50 mM 삼나트륨 시트레이트 및 보다 바람직하게 약 250 mM 미만의 NaCl 및 25 mM의 삼나트륨 시트레이트일 것이다. 낮은 엄중 하이브리드화는 유기 용매, 예를 들어, 포름아미드의 부재하에 수득될 수 있고, 높은 엄중 하이브리드화는 적어도 약 35% 포름아미드 및 보다 바람직하게 적어도 약 50% 포름아미드의 존재하에 수득될 수 있다. 엄중 온도 조건은 통상적으로 적어도 약 30℃, 보다 바람직하게 적어도 약 37℃, 및 가장 바람직하게 적어도 약 42℃의 온도를 포함할 것이다. 다양한 추가의 파라미터, 예를 들어, 하이브리드화 시간, 세제의 농도, 예를 들어, 나트륨 도데실 설페이트 (SDS), 및 캐리어 DNA의 내포 또는 배제는 당업자에게 널리 공지되어 있다. 다양한 수준의 엄중도는 필요한 만큼 이들 다양한 조건을 조합함에 의해 성취된다. 특정 구현예에서, 하이브리드화는 750 mM NaCl, 75 mM 삼나트륨 시트레이트 및 1% SDS에서 30℃에서 수행할 것이다. 보다 바람직한 구현예에서, 하이브리드화는 500 mM NaCl, 50 mM 삼나트륨 시트레이트, 1% SDS, 35% 포름아미드, 및 100 μg/ml의 변성된 연어 정자 DNA (ssDNA)에서 37℃에서 수행할 것이다. 가장 바람직한 구현예에서, 하이브리드화는 250 mM NaCl, 25 mM 삼나트륨 시트레이트, 1% SDS, 50% 포름아미드, 및 200 μg/ml ssDNA에서 42℃에서 수행할 것이다. 이들 조건에 대한 유용한 변화는 당업자에게 용이하게 자명할 것이다.

[0260] 대부분의 적용을 위해, 하이브리드화에 이어서 세척 단계는 또한 엄중도에서 다양할 것이다. 세척 엄중 조건은 염 농도 및 온도에 의해 한정될 수 있다. 상기와 같이, 세척 엄중도는 염 농도를 감소시킴에 의해 또는 온도를 증가시킴에 의해 증가될 수 있다. 예를 들어, 세척 단계를 위한 엄중 염 농도는 바람직하게 약 30 mM 미만의 NaCl 및 3 mM의 삼나트륨 시트레이트 및 가장 바람직하게 약 15 mM 미만의 NaCl 및 1.5 mM 삼나트륨 시트레이트일 것이다. 세척 단계를 위한 엄중 온도 조건은 통상적으로 적어도 약 25℃, 보다 바람직하게 적어도 약 42℃, 및 보다 더 바람직하게 적어도 약 68℃의 온도를 포함할 것이다. 바람직한 구현예에서, 세척 단계는 30 mM NaCl, 3 mM 삼나트륨 시트레이트 및 0.1% SDS에서 25℃에서 수행할 것이다. 보다 바람직한 구현예에서, 세척 단계는 15 mM NaCl, 1.5 mM 삼나트륨 시트레이트 및 0.1% SDS에서 42℃에서 수행할 것이다. 보다 바람직한 구현예에서, 세척 단계는 15 mM NaCl, 1.5 mM 삼나트륨 시트레이트 및 0.1% SDS에서 68℃에서 수행할 것이다. 이들 조건에 대한 추가의 변화는 당업자에게 용이하게 자명할 것이다. 하이브리드화 기술은 당업자에게 널리 공지되어 있고, 예를 들어, 문헌 (참조: 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); and Sambrook *et al.,* Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York)에 기재되어 있다.

[0261] "스플릿"이란 2개 이상의 단편으로 나누어짐을 의미한다.

[0262] "스플릿 Cas9 단백질" 또는 "스플릿 Cas9"는 2개의 별도의 뉴클레오타이드 서열에 의해 암호화된 N-말단 단편 및 C-말단 단편으로서 제공되는 Cas9 단백질을 언급한다. Cas9 단백질의 N-말단 부분 및 C-말단 부분에 상응하는 폴리펩타이드는 스플라이싱되어 "재구성된" Cas9 단백질을 형성할 수 있다. 특정 구현예에서, Cas9 단백질은 예를 들어, 이의 각각이 본원에 참조로 포함된, 문헌 (참조: Nishimasu *et al.,* Cell, Volume 156, Issue 5, pp. 935-949, 2014)에 기재되거나, 문헌 (참조: Jiang *et al.* (2016) Science 351: 867-871. PDB file: 5F9R)에 기재된 바와 같이, 단백질의 무질서 영역 내에서 2개의 단편으로 나누어진다. 일부 구현예에서, 단백질은 약 아미노산 A292-G364, F445-K483, 또는 E565-T637 사이의 SpCas9의 영역 내 임의의 C, T, A 또는 S에서, 또는 임의의 다른 Cas9, Cas9 변이체 (예를 들어, nCas9, dCas9), 또는 다른 napDNAbp 내 상응하는 위치에서 2개의 단편으로 나누어진다. 일부 구현예에서, 단백질은 SpCas9 T310, T313, A456, S469, 또는 C574에서 2개의 단편으로 나누어진다. 일부 구현예에서, 단백질을 2개의 단편으로 나누는 공정은 상기 단백질을 "스프릿팅"하는 것으로서 언급된다.

[0263] 다른 구현예에서, Cas9 단백질의 N-말단 부분은 아미노산 1-573 또는 1-637 에스. 피오게네스 Cas9 야생형 (SpCas9) (NCBI 참조 서열: NC_002737.2, 유니프롯 참조 서열: Q99ZW2)을 포함하고, Cas9 단백질의 C-말단 부분은 SpCas9 야생형의 아미노산 574-1368 또는 638-1368의 부분 또는 이의 상응하는 위치를 포함한다.

[0264] 스플릿 Cas9의 C-말단 부분은 스플릿 Cas9의 N-말단 부분과 연결되어 완전한 Cas9 단백질을 형성할 수 있다. 일부 구현예에서, Cas9 단백질의 C-말단 부분은 Cas9 단백질의 N-말단 부분이 종료되는 위치에서 개시한다. 이와 같이, 일부 구현예에서, 스플릿 Cas9의 C-말단 부분은 spCas9의 아미노산 (551-651)-1368 부분을 포함한다. "(551-651)-1368"은 아미노산 551-651 (포괄적) 사이의 아미노산에서 개시하고 아미노산 1368에서 종료됨을 의미한다. 예를 들어, 스플릿 Cas9의 C-말단 부분은 spCas9의 아미노산 551-1368, 552-1368, 553-1368, 554-1368, 555-1368, 556-1368, 557-1368, 558-1368, 559-1368, 560-1368, 561-1368, 562-1368, 563-1368, 564-1368, 565-1368, 566-1368, 567-1368, 568-1368, 569-1368, 570-1368, 571-1368, 572-1368, 573-1368, 574-1368, 575-1368, 576-1368, 577-1368, 578-1368, 579-1368, 580-1368, 581-1368, 582-1368, 583-1368, 584-1368, 585-1368, 586-1368, 587-1368, 588-1368, 589-1368, 590-1368, 591-1368, 592-1368, 593-1368, 594-1368, 595-1368, 596-1368, 597-1368, 598-1368, 599-1368, 600-1368, 601-1368, 602-1368, 603-1368, 604-1368, 605-1368, 606-1368, 607-1368, 608-1368, 609-1368, 610-1368, 611-1368, 612-1368, 613-1368, 614-1368, 615-1368, 616-1368, 617-1368, 618-1368, 619-1368, 620-1368, 621-1368, 622-1368, 623-1368, 624-1368, 625-1368, 626-1368, 627-1368, 628-1368, 629-1368, 630-1368, 631-1368, 632-1368, 633-1368, 634-1368, 635-1368, 636-1368, 637-1368, 638-1368, 639-1368, 640-1368, 641-1368, 642-1368, 643-1368, 644-1368, 645-1368, 646-1368, 647-1368, 648-1368, 649-1368, 650-1368, 또는 651-1368 중 어느 하나의 부분을 포함할 수 있다. 일부 구현예에서, 스플릿 Cas9 단백질의 C-말단 부분은 SpCas9의 아미노산 574-1368 또는 638-1368의 부분을 포함한다.

[0265] "대상체"란 인간 또는 비-인간 포유류, 예를 들어, 소, 말, 개, 양 또는 고양이를 포함하지만 이에 제한되지 않는 포유류를 의미한다. 대상체는 가축, 노동력을 생산하고 식품과 같은 상품을 제공하기 위해 길러진 가정용 동물, 소, 염소, 닭, 말, 돼지, 토끼 및 양을 포함하지만 이에 제한되지 않는다.

[0266] "실질적으로 동일한"이란 참조 아미노산 서열 (예를 들어, 본원에 기재된 아미노산 서열 중 어느 하나) 또는 핵산 서열 (예를 들어, 본원에 기재된 핵산 서열 중 어느 하나)과 적어도 50% 동일성 나타내는 폴리펩타이드 또는 핵산 분자를 의미한다. 바람직하게는, 상기 서열은 아미노산 수준 또는 핵산에서 비교를 위해 사용되는 서열과 적어도 60%, 보다 바람직하게는 80% 또는 85%, 및 보다 바람직하게는 90%, 95% 또는 심지어 99% 동일하다.

[0267] 서열 동일성은 전형적으로 서열 분석 소프트웨어를 사용하여 측정된다 (예를 들어, 기관 (the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705)의 서열 분석 소프트웨어 패키지, BLAST, BESTFIT, COBALT, EMBOSS Needle, GAP, 또는 PILEUP/PRETTYBOX 프로그램). 상기 소프트웨어는 상동성 정도를 다양한 치환, 결실 및/또는 다른 변형에 할당함에 의해 동일하거나 유사한 서열을 매칭시킨다. 보존적 치환은 전형적으로 하기의 그룹 내 치환을 포함한다: 글라이신, 알라닌; 발린, 이소류신, 류신; 아스파르트산, 글루탐산, 아스파라긴, 글루타민; 세린, 트레오닌; 라이신, 아르기닌; 및 페닐알라닌, 티로신. 동일성 정도를 결정하기 위한 예시적 접근법에서, BLAST 프로그램이 사용될 수 있고, 확률 스코어는 e^{-3}과 e^{-100} 사이이고 이는 밀접하게 관련된 서열을 지적한다. COBALT는 예를 들어, 하기의 파라미터와 함께 사용된다:

[0268] a) 정렬 파라미터: 갭 페널티-11, -1 및 종료-갭 페널티-5, -1,

[0269] b) CDD 파라미터: CDD 파라미터에 RPS BLAST를 사용; Blast E-값 0.003; 보존된 컬럼을 발견하고 CDD 파라미터에서 재계산, 및

[0270] c) 조회 클러스터링 파라미터: CDD 파라미터에 조회 클러스터를 사용; 워드 크기 4; 최대 클러스터 거리 0.8; 알파벳 규칙.

[0271] EMBOSS 바늘은 예를 들어, 하기의 파라미터와 함께 사용한다:

[0272] a) 매트릭스: BLOSUM62;

[0273] b) 갭 개방: 10;

[0274] c) 갭 연장: 0.5;

[0275] d) 아웃풋 포맷: 쌍;

[0276] e) 종료 갭 페널티: 거짓;

[0277] f) 종료 갭 개방: 10; 및

[0278] g) 종료 갭 연장: 0.5.

[0279] 용어 "표적 부위"는 핵염기 편집기에 의해 변형된 핵산 분자 내 서열을 언급한다. 하나의 구현예에서, 표적 부위는 데아미나제 또는 데아미나제 (예를 들어, 아데닌 데아미나제)를 포함하는 융합 단백질에 의해 탈아민화된다.

[0280] 본원에 사용된 바와 같은 용어 "치료한다", "치료하는", "치료" 등은 장애 및/또는 이와 연관된 증상(들)을 감소시키거나 개선시키거나 목적하는 약리학적 및/또는 병리학적 효과를 수득함을 언급한다. 배제하는 것은 아니지만, 장애 또는 병태를 치료하는 것은 장애, 병태 또는 이와 연관된 증상이 완전히 제거될 것을 요구하지 않음을 인지할 것이다. 일부 구현예에서, 효과는 치료학적이고, 즉, 제한 없이, 효과는 질환 및/또는 상기 질환에 기인할 수 있는 부작용 증상을 부분적으로 또는 완전하게 감소시키거나, 감쇠시키거나, 폐지시키거나, 약화시키거나, 완화시키거나, 감소시키거나, 이를 치유한다. 일부 구현예에서, 효과는 예방적이고, 즉, 효과는 질환 또는 병태의 발병 또는 재발로부터 보호하거나 이를 예방한다. 이를 위해, 본원에 개시된 방법은 본원에 기재된 바와 같이 치료학적 유효량의 조성물을 투여함을 포함한다.

[0281] "우라실 글리코실라제 저해제" 또는 "UGI"란 우라실-절제 복구 시스템을 저해하는 제제를 의미한다. 하나의 구현예에서, 제제는 숙주 우라실-DNA 글리코실라제에 결합하고 DNA로부터 우라실 잔기의 제거를 예방하는 단백질 또는 이의 단편이다. 하나의 구현예에서, UGI는 우라실-DNA 글리코실라제 염기-절제 복구 효소를 저해할 수 있는 단백질, 이의 단편 또는 도메인이다. 일부 구현예에서, UGI 도메인은 야생형 UGI 또는 이의 변형된 버전을 포함한다. 일부 구현예에서, UGI 도메인은 하기에 제시된 예시적인 아미노산 서열의 단편을 포함한다. 일부 구현예에서, UGI 단편은 하기에 제공된 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 구현예에서, UGI는 하기에 제시된 바와 같이 예시적인 UGI 아미노산 서열 또는 이의 단편과 상동성인 아미노산 서열을 포함한다. 일부 구현예에서, UGI, 또는 이의 일부는 하기에 제시된 바와 같이 야생형 UGI 또는 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:

[0282] >sp|P14739|UNGI_BPPB2 우라실-DNA 글리코실라제 저해제

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD

[0283] APEYKPWALVIQDSNGENKIKML.

[0284] 용어 "벡터"는 핵산 서열을 세포에 도입하여 형질전환된 세포를 유도하는 수단을 언급한다. 벡터는 플라스미드, 트랜스포존, 파아지, 바이러스, 리포좀, 및 에피좀을 포함한다. "발현 벡터"는 수용자 세포에서 발현될 뉴클레오타이드 서열을 포함하는 핵산 서열이다. 발현 벡터는 개시, 정지, 인핸서, 프로모터 및 분비 서열과 같은 도입된 서열의 발현을 촉진시키고/시키거나 용이하게 하기 위해 추가의 핵산 서열을 포함할 수 있다.

[0285] 본원에서 변수의 임의의 정의에 있는 화학 그룹 목록의 인용은 그 변수를 임의의 단일 그룹 또는 나열된 그룹의 조합으로서 정의하는 것을 포함한다. 본원에서 변수 또는 양상에 대한 구현예의 언급은 임의의 단일 구현예로서 또는 임의의 다른 구현예 또는 이의 부분과 조합하여 해당 구현예를 포함한다.

[0286] 본원에 제공되고 기재된 임의의 조성물 또는 방법은 본원에 제공되고 기재된 임의의 다른 조성물 및 방법 중 하나 이상과 조합될 수 있다.

[0287] DNA 편집은 유전학적 수준에서 병원성 돌연변이를 교정함에 의해 질환 상태를 변형시키는 실행 가능한 수단으로서 출현하였다. 최근까지, 모든 DNA 편집 플랫폼은 특정 게놈 부위에서 DNA 이중 가닥 절단 (DSB)을 유도하고, 반-확률적 방식의 생성물 결과를 결정하기 위한 내인성 DNA 복구 경로에 의존함에 의해 기능하여 복잡한 유전학적 생성물의 집단을 유도한다. 정확하지만, 사용자 정의된 복구 결과는 상동성 지시된 복구 (HDR) 경로를 통해 성취될 수 있고, 다수의 챌린지는 치료학적 관련 세포 유형에서 HDR을 사용한 고효율 복구를 차단하였다. 실제

로, 상기 경로는 경쟁의 오류 성향 비-상동성 말단 연결 경로에 비해 비효율적이다. 추가로, HDR은 세포 주기의 G1 및 S기에 엄격히 제한되어 유사분열 후 세포에서 DSB의 정확한 복구를 차단한다. 결과로서, 이들 집단에서 고효율과 함께 사용자 정의된 프로그래밍 가능한 방식으로 게놈 서열을 변경하기 어렵거나 불가능한 것으로 입증되었다.

[0288] **핵염기 편집기**

[0289] 본원에서는 폴리뉴클레오타이드의 표적 뉴클레오타이드 서열을 편집하거나, 변형시키거나 변경하기 위한 염기 편집기 또는 핵염기 편집기가 개시된다. 본원에서는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 핵염기 편집 도메인 (예를 들어, 아데노신 데아미나제)를 포함하는 핵염기 편집기 또는 염기 편집기가 기재된다. 결합된 가이드 폴리뉴클레오타이드 (예를 들어, gRNA)와 접합된 경우 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 표적 폴리뉴클레오타이드 서열과 특이적으로 결합 (즉, 결합된 가이드 핵산의 염기와 표적 폴리뉴클레오타이드 서열의 염기 간에 상보적 염기 쌍 형성을 통해)할 수 있고, 이로써 염기 편집기는 편집시키고자 하는 표적 핵산 서열에 위치한다. 일부 구현예에서, 표적 뉴클레오타이드 서열은 단일 가닥 DNA 또는 이중 가닥 DNA를 포함한다. 일부 구현예에서, 표적 폴리뉴클레오타이드 서열은 RNA를 포함한다. 일부 구현예에서, 표적 폴리뉴클레오타이드 서열은 DNA-RNA 하이브리드를 포함한다.

[0290] ***폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인***

[0291] 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 또한 RNA에 결합하는 핵산 프로그래밍 가능한 단백질을 포함할 수 있는 것으로 인지되어야 한다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 RNA로 가이드하는 핵산과 연합될 수 있다. 다른 핵산 프로그래밍 가능한 DNA 결합 단백질은 또한 본원 개시내용의 범위 내에 있지만, 이들은 구체적으로 본원 개시내용에 열거되어 있지 않다.

[0292] 염기 편집기의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 그 자체가 하나 이상의 도메인을 포함할 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 하나 이상의 뉴클레아제 도메인을 포함할 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 뉴클레아제 도메인은 엔도뉴클레아제 또는 엑소뉴클레아제를 포함할 수 있다. 본원에서 용어 "엑소뉴클레아제"는 자유 말단으로부터 핵산 (예를 들어, RNA 또는 DNA)을 분해할 수 있는 단백질 또는 폴리펩타이드를 언급하고, 용어 "엔도뉴클레아제"는 핵산 (예를 들어, DNA 또는 RNA)내 내부 영역을 촉매 (예를 들어, 절단)할 수 있는 단백질 또는 폴리펩타이드를 언급한다. 일부 구현예에서, 엔도뉴클레아제는 이중 가닥 핵산의 단일 가닥을 절단할 수 있다. 일부 구현예에서, 엔도뉴클레아제는 이중 가닥 핵산 분자의 양 가닥을 절단할 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 데옥시리보뉴클레아제일 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 리보뉴클레아제일 수 있다.

[0293] 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 뉴클레아제 도메인은 표적 폴리뉴클레오타이드의 0개, 1개 또는 2개 가닥을 절단할 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 닉카제 도메인을 포함할 수 있다. 본원에서, 용어 "닉카제"는 듀플렉스 핵산 분자 (예를 들어, DNA)에서 2개 가닥 중 하나의 가닥만을 절단할 수 있는 뉴클레아제 도메인을 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 언급한다. 일부 구현예에서, 닉카제는 하나 이상의 돌연변이를 활성 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인에 도입함에 의해 완전한 촉매 활성 (예를 들어, 천연) 형태의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인으로부터 유래할 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인이 Cas9로부터 유래된 닉카제 도메인을 포함하는 경우, Cas9-유래된 닉카제 도메인은 D10A 돌연변이 및 위치 840에 히스티딘을 포함할 수 있다. 상기 구현예에서, 잔기 H840은 촉매 활성을 보유하고, 이로써 핵산 듀플렉스의 단일 가닥을 절단할 수 있다. 또 다른 예에서, Cas9-유래된 닉카제 도메인은 H840A 돌연변이를 포함할 수 있고, 위치 10에서 아미노산 잔기는 D로 남아있다. 일부 구현예에서, 닉카제는 닉카제 활성을 위해 요구되지 않는 뉴클레아제 도메인의 전부 또는 일부를 제거함에 의해 완전한 촉매 활성 (예를 들어, 천연) 형태의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인으로부터 유래할 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인이 Cas9로부터 유래된 닉카제 도메인을 포함하는 경우, Cas9-유래된 닉카제 도메인은 RuvC 도메인 또는 HNH 도메인의 전부 또는 일부의 결실을 포함할 수 있다.

[0294] 예시적인 촉매적 활성 Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL

[0295] ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV

YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD.

[0296]

[0297] 닉카제 도메인을 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기는 따라서 특이적 폴리뉴클레오타이드 표적 서열에서 (예를 들어, 결합된 가이드 핵산의 상보적 서열에 의한 결정시) 단일 가닥 DNA 절단 (닉)을 생성할 수 있다. 일부 구현예에서, 닉카제 도메인 (예를 들어, Cas9-유래된 닉카제 도메인)을 포함하는 염기 편집기에 의해 절단되는 핵산 듀플렉스 표적 폴리뉴클레오타이드 서열의 가닥은 염기 편집기에 의해 편집되지 않는 가닥 (즉, 염기 편집기에 의해 절단되는 가닥은 편집될 염기를 포함하는 가닥의 반대편에 있다)이다. 다른 구현예에서, 닉카제 도메인 (예를 들어, Cas9-유래된 닉카제 도메인)을 포함하는 염기 편집기는 편집을 위해 표적화되는 DNA 분자의 가닥을 절단할 수 있다. 상기 구현예에서, 비-표적화된 가닥은 절단되지 않는다.

[0298] 또한 본원에는 촉매적 데드 (즉, 표적 폴리뉴클레오타이드 서열을 절단할 수 없는) 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기가 제공된다. 본원에서 용어 "촉매적 데드" 및 "뉴클레아제 데드"는 핵산의 가닥을 절단하지 못하는 무능력을 유도하는 하나 이상의 돌연변이 및/또는 결실을 갖는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 언급하기 위해 상호교환적으로 사용된다. 일부 구현예에서, 촉매적 데드 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 염기 편집기는 하나 이상의 뉴클레아제 도메인에서 특정 점 돌연변이의 결과로서 뉴클레아제 활성이 부재일 수 있다. 예를 들어, Cas9 도메인을 포함하는 염기 편집기의 경우에, Cas9는 D10A 돌연변이 및 H840A 돌연변이 둘 다를 포함할 수 있다. 상기 돌연변이는 뉴클레아제 도메인 둘 다를 불활성화시킴으로써 뉴클레아제 활성을 상실시킨다. 다른 구현예에서, 촉매적 데드 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 촉매 도메인 (예를 들어, RuvC1 및/또는 HNH 도메인)의 전부 또는 일부의 하나 이상의 결실을 포함할 수 있다. 추가의 구현예에서, 촉매적 데드 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 뉴클레아제 도메인의 전부 또는 일부의 결실뿐만 아니라 점 돌연변이 (예를 들어, D10A 또는 H840A)를 포함한다.

[0299] 또한 본원에는 이전에 기능성 버전의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인으로부터의 촉매적 데드 폴리뉴클레오타이드 프로그래밍 가능한 결합 도메인을 생성할 수 있는 돌연변이가 고려된다. 예를 들어, 촉매적 데드 Cas9 ("dCas9")의 경우에, D10A 및 H840A 이외의 다른 돌연변이를 갖는 변이체가 제공되고 이는 뉴클레아제 불활성화된 Cas9를 초래한다. 상기 돌연변이는 예를 들어, D10 및 H840에서 다른 아미노

산 치환, 또는 Cas9의 뉴클레아제 도메인 내 다른 치환 (예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인에서의 치환)을 포함한다. 추가의 적합한 뉴클레아제-불활성 dCas9 도메인은 본원 개시내용 및 당해 분야의 지식을 기준으로 당업자에게 자명할 수 있고, 본원 개시내용의 범위 내에 있다. 상기 추가의 예시적인 적합한 뉴클레아제-불활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A, 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하지만 이에 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, 이의 전체 내용은 본원에 참조로 포함됨).

[0300] 염기 편집기에 혼입될 수 있는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 비제한적인 예는 CRISPR 단백질-유래된 도메인, 제한 뉴클레아제, 메가뉴클레아제, TAL 뉴클레아제 (TALEN), 및 아연 핑거 뉴클레아제 (ZFN)를 포함한다. 일부 구현예에서, 염기 편집기는 결합된 가이드 핵산을 통해 핵산의 CRISPR (즉, 클러스터형 규칙적 간격을 둔 짧은 팔린드롬 반복체)-매개된 변형 동안에 핵산 서열에 결합할 수 있는 천연 또는 변형된 단백질 또는 이의 일부를 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함한다. 상기 단백질은 본원에서 "CRISPR 단백질"로서 언급된다. 따라서, 본원에서는 CRISPR 단백질의 전부 또는 일부를 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기 (즉, 또한 염기 편집기의 "CRISPR 단백질-유래된 도메인"으로서 언급되는 CRISPR 단백질의 전부 또는 일부를 도메인으로서 포함하는 염기 편집기)가 개시된다. 염기 편집기에 혼입된 CRISPR 단백질-유래된 도메인은 야생형 또는 천연 버전의 CRISPR 단백질과 비교하여 변형될 수 있다. 예를 들어, 하기된 바와 같이, CRISPR 단백질-유래된 도메인은 야생형 또는 천연 버전의 CRISPR 단백질과 비교하여 하나 이상의 돌연변이, 삽입, 결실, 재정렬 및/또는 재조합을 포함할 수 있다.

[0301] CRISPR은 이동 유전학적 요소 (바이러스, 전이할 수 있는 요소 (transposable elements) 및 접합성 플라스미드)에 대한 보호를 제공하는 후천성 면역계이다. CRISPR 클러스터는 스페이서, 선행 이동 요소에 상보적인 서열 및 표적 공격 핵산을 포함한다. CRISPR 클러스터는 CRISPR RNA (crRNA)로 전사되고 프로세싱된다. II형 CRISPR 시스템에서, 전구-crRNA의 올바른 프로세싱은 트랜스-암호화된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 단백질을 요구한다. tracrRNA는 전구-crRNA의 리보뉴클레아제 3-원조 프로세싱에 대한 가이드로서 작용한다. 후속적으로, Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 환형 dsDNA 표적을 엔도핵산분해적으로 절단한다. crRNA에 상보적이지 않은 표적 가닥은 먼저 엔도핵산분해적으로 절단됨에 이어서 3'-5' 엑소핵산분해적으로 절단 제거한다. 실제로, DNA-결합 및 절단은 전형적으로 단백질 및 2개의 RNA를 요구한다. 그러나, 단일 가이드 RNA ("sgRNA, 또는 단순히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 양상을 단일 RNA 종으로 혼입하기 위해 가공될 수 있다. 예를 들어, 문헌 (참조: Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. Science 337:816-821(2012))을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다. Cas9는 자가 대 비-자가의 구분을 도와주기 위해 CRISPR 반복 서열 (PAM 또는 프로토스페이서 인접 모티프)에서 짧은 모티프를 인지한다.

[0302] 일부 구현예에서, 본원에 기재된 방법은 가공된 Cas 단백질을 사용할 수 있다. 가이드 RNA (gRNA)는 Cas-결합을 위해 필요한 스캐폴드 서열 및 변형된 게놈 표적을 한정하는 사용자 정의된 ~20개 뉴클레오타이드 스페이서로 구성된 짧은 합성 RNA이다. 따라서, 당업자는 Cas 단백질의 게놈 표적을 변화시킬 수 있고, 특이성은 부분적으로 gRNA 표적화 서열이 나머지 게놈과 비교하여 게놈 표적에 대해 얼마나 특이적인지에 의해 결정된다.

[0303] 일부 구현예에서, gRNA 스캐폴드 서열은 다음과 같다: GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU.

[0304] 일부 구현예에서, 염기 편집기에 혼입된 CRISPR 단백질-유래된 도메인은 결합된 가이드 핵산과 접합되는 경우 표적 폴리뉴클레오타이드에 결합할 수 있는 엔도뉴클레아제 (예를 들어, 데옥시리보뉴클레아제 또는 리보뉴클레아제)이다. 일부 구현예에서, 염기 편집기에 혼입된 CRISPR 단백질-유래된 도메인은 결합된 가이드 핵산과 접합되는 경우 표적 폴리뉴클레오타이드에 결합할 수 있는 닉카제이다. 일부 구현예에서, 염기 편집기에 혼입된 CRISPR 단백질-유래된 도메인은 결합된 가이드 핵산과 접합되는 경우 표적 폴리뉴클레오타이드에 결합할 수 있는 촉매적 데드 도메인이다. 일부 구현예에서, 염기 편집기의 CRISPR 단백질 유래된 도메인에 의해 결합되는 표적 폴리뉴클레오타이드는 DNA이다. 일부 구현예에서, 염기 편집기의 CRISPR 단백질-유래된 도메인에 의해 결합되는 표적 폴리뉴클레오타이드는 RNA이다.

[0305] 본원에 사용될 수 있는 Cas 단백질은 부류 1 및 부류 2를 포함한다. Cas 단백질의 비제한적인 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas9 (또한 Csn1 또는

Csx12로서 공지된), 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, Cs0, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 및 Cas12i, CARF, DinG, 이의 상동체 또는 이의 변형된 버전을 포함한다. 2개의 기능성 엔도뉴클레아제 도메인을 갖는 Cas9와 같이 변형되지 않은 CRISPR 효소는 DNA 절단 활성을 가질 수 있다: RuvC 및 HNH. CRISPR 효소는 예를 들어, 표적 서열 내 및/또는 표적 서열의 상보체 내에 표적 서열에서 하나의 가닥 또는 가닥 둘 다의 절단을 지시할 수 있다. 예를 들어, CRISPR 효소는 표적 서열의 제1 또는 마지막 뉴클레오타이드로부터 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500개 이상의 염기 쌍 내에 하나의 가닥 또는 가닥 둘 다의 절단을 지시할 수 있다.

[0306] 상응하는 야생형 효소와 관련하여, 돌연변이된 CRISPR 효소가 표적 서열을 함유하는 표적 폴리뉴클레오타이드의 하나의 가닥 또는 가닥 둘 다를 절단하는 능력이 부재인 CRISPR 효소를 암호화하는 벡터가 사용될 수 있다. Cas9는 야생형 예시적인 Cas9 폴리펩타이드 (예를 들어, 에스. 피오게네스로부터의 Cas9)와 적어도 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성 및/또는 서열 상동성 또는 적어도 약 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성 및/또는 서열 상동성을 갖는 폴리펩타이드를 언급할 수 있다. Cas9는 야생형 예시적인 Cas9 폴리펩타이드 (예를 들어, 에스. 피오게네스로부터)와 최대 또는 최대 약 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성 및/또는 서열 상동성을 갖는 폴리펩타이드를 언급할 수 있다. Cas9는 결실, 삽입, 치환, 변이체, 돌연변이, 융합, 키메라 또는 이의 임의의 조합과 같은 아미노산 변화를 포함할 수 있는 야생형 또는 변형된 형태의 Cas9 단백질을 언급할 수 있다.

[0307] 일부 구현예에서, 염기 편집기의 CRISPR 단백질-유래된 도메인은 코리네박테리움 울세란스 (NCBI Refs: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아 (NCBI Refs: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라 (NCBI Ref: NC_021284.1); 프레보텔라 인터메디아 (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨스 (NCBI Ref: NC_021846.1); 스트렙토코커스 이니애 (NCBI Ref: NC_021314.1); 벨리엘라 발티카 (NCBI Ref: NC_018010.1); 사이크로플렉서스 토르쿠이스 (NCBI Ref: NC_018721.1); 스타필로코커스 써모필러스 (NCBI Ref: YP_820832.1); 리스테리아 이노쿠아 (NCBI Ref: NP_472073.1); 캄필로박터 제주니 (NCBI Ref: YP_002344900.1); 나이세리아 메닌기티디스 (NCBI Ref: YP_002342100.1), 스트렙토코커스 피오게네스, 또는 스타필로코커스 아우레우스로부터 기원하는 Cas9의 전부 또는 일부를 포함할 수 있다.

[0308] ***핵염기 편집기의 Cas9 도메인***

[0309] Cas9 뉴클레아제 서열 및 구조는 당업자에게 널리 공지되어 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); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., *et al.*, Science 337:816-821(2012)) 참조, 이의 전체 내용은 본원에 참조로 포함됨). Cas9 오톨로그는 에스. 피오게네스 및 에스. 써모필러스를 포함하지만 이에 제한되지 않는 다양한 종에 기재되어 있다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본원 개시내용을 기준으로 당업자에게 자명할 것이고, 상기 Cas9 뉴클레아제 및 서열은 이의 전문이 본원에 참조로 포함되는 문헌 (참조: 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.

[0310] 일부 양상에서, 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)은 Cas9 도메인이다. 비제한적으로, 예시적인 Cas9 도메인이 본원에 제공된다. Cas9 도메인은 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인 (dCas9), 또는 Cas9 닉카제 (nCas9)일 수 있다. 일부 구현예에서, Cas9 도메인은 뉴클레아제 활성 도메인이다. 예를 들어, Cas9 도메인은 듀플렉스 핵산의 가닥 둘 다 (예를 들어, 듀플렉스 DNA 분자의 가닥 둘 다)를 절단하는 Cas9 도메인일 수 있다. 일부 구현예에서, Cas9 도메인은 본원에 제시된 바와 같은 아미노산 서열 중 하나를 포함한다. 일부 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 임의의 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 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 임의의 하나와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,

35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 임의의 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 1200개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0311] 일부 구현예에서, Cas9의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 구현예에서, 단백질은 2개의 Cas9 도메인 중 하나를 포함한다: (1) Cas9의 gRNA 결합 도메인; 또는 (2) Cas9의 DNA 절단 도메인. 일부 구현예에서, Cas9 또는 이의 단편을 포함하는 단백질은 "Cas9 변이체"로서 언급된다. Cas9 변이체는 Cas9 또는 이의 단편과 상동성을 공유한다. 예를 들어, Cas9 변이체는 야생형 Cas9 서열과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, Cas9 변이체는 야생형 Cas9와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 아미노산 변화를 가질 수 있다.

[0312] 일부 구현예에서, Cas9 변이체는 Cas9의 단편 (예를 들어, gRNA 결합 도메인 또는 DNA 절단 도메인)을 포함하여, 상기 단편은 야생형 Cas9의 상응하는 단편과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, 단편은 상응하는 야생형 Cas9의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 동일하거나, 이의 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%이다. 일부 구현예에서, 단편은 적어도 100개 아미노산 길이이다. 일부 구현예에서, 상기 단편은 적어도 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 또는 적어도 1300개 아미노산 길이이다.

[0313] 일부 구현예에서, 본원에 제공된 바와 같은 Cas9 융합 단백질은 Cas9 단백질의 전장 아미노산 서열, 예를 들어, 본원에 제공된 Cas9 서열 중 하나를 포함한다. 그러나, 다른 구현예에서, 본원에 제공된 바와 같은 융합 단백질은 전장 Cas9 서열을 포함하지 않고 단지 하나 이상의 이의 단편을 포함한다. 적합한 Cas9 도메인 및 Cas9 단편의 예시적인 아미노산 서열이 본원에 제공되고, Cas9 도메인 및 단편의 추가의 적합한 서열은 당업자에게 자명할 것이다.

[0314] 일부 구현예에서, 야생형 Cas9는 스트렙토코커스 피오게네스 기원의 Cas9 (NCBI 참조 서열: NC_017053.1)에 상응한다. 예시적인 뉴클레오타이드 및 아미노산 서열은 다음과 같다:

[0315]

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCAC
TGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCA
AAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAA
CGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTC
AAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAG
AAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAG
AAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCG
CTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATT
TAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTA
TTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAG
TAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTG
GGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAA
GATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAAT
TGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAG
ATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTAC
GATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTA
TAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCC
AAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTG
GTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCA
TCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAA
AAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTG
GCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAA
TTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTG
ATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTT

TATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGG
TGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAAT
TAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGAT
AGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTT
GGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATA
GGGGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAG
CTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGA
TAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTA
TGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGA
CAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTAT
TTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCG
TTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATG
AAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAA
TACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGG
ACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTC
ATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGA
TAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCA
AGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTT
GATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACA
AATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAG
TGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGT
GAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGAT
TAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTA
AAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAAT
ATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAAT
CGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCA
AAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCC
AAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCC
AAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGG
AAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGA
AGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGA
CTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG
CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTA
[0316] TATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTT

TGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTG
TTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCA
ATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATG
CCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGA
GGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQL
FEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
IKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMER
SSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP
IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
GD

[0317]

[0318] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0319] 일부 구현예에서, 야생형 Cas9는 하기의 뉴클레오타이드 및/또는 아미노산 서열에 상응하거나 이를 포함한다:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAAC
[0320] CGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTA

AAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAA
CGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAG
CAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAG
AGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAA
AAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAG
GTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATC
TAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTG
TTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTC
TAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCG
GTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAA
GATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAAT
TGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTG
ACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATA
TAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTC
AAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTT
GTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACA
TCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCA
AAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTG
GCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAA
TTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTG
ACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTG
TACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGG
AGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAAT
TGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGAT
CGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCT
GGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATC
GGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAG
TTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGA
CAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTA
TGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGA
CAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCAT
ACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACA
[0321] TTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGG

ATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGA
AAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATG
TTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCC
TTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAG
TGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATG
CGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAA
CTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGC
ACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCA
AAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTT
AGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACT
CATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCC
GTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCT
AACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTT
AATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGA
GAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTT
TCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGA
CCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAG
TTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAG
CGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAA
GGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGT
TGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTC
CTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACT
TTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGA
GAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGC
CGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAG
ACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTT
GGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGA
TTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE

[0322]

DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0323]

[0324] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0325] 일부 구현예에서, 야생형 Cas9는 스트렙토코커스 피오게네스 (NCBI 참조 서열: NC_002737.2 (다음과 같은 뉴클레오타이드 서열); 및 유니프롯 참조 서열: Q99ZW2 (다음과 같은 아미노산 서열) 기원의 Cas9에 상응한다:

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCAC
TGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCA
AAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAA
CGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTC
AAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAG
AAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAG
AAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCG
CTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATT
TAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTA
TTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAG
TAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTG
GGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAA

[0326]

GATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAAT
TGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAG
ATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTAC
GATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTA
TAAAGAAATCTTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCC
AAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTG
GTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCA
TCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAA
AAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTG
GCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAA
TTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTG
ATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTT
TATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGG
TGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAAT
TAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGAT
AGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTT
GGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATA
GGGAGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAG
CTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGA
TAAGCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTA
TGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGA
CAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTAT
TTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATA
TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGT
ATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGA
AAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATG
TGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGT
TTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATC
GGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACG
CCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAA
CTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGC
ACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTA
AAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTA

[0327] CGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTT

GATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTC
GTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCT
AATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCT
AATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGC
GCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTC
TCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGA
TCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGG
TGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAA
AGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAA
AGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGC
TGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTT
TTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATT
GTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGC
GTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAA
CCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGC
TGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAG
ATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA
GGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV

[0328]

REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0329]

[0330] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0331] 일부 구현예에서, Cas9는 코리네박테리움 울세란스 (NCBI Refs: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아 (NCBI Refs: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라 (NCBI Ref: NC_021284.1); 프레보텔라 인터메디아 (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨스 (NCBI Ref: NC_021846.1); 스트렙토코커스 이니애 (NCBI Ref: NC_021314.1); 벨리엘라 발티카 (NCBI Ref: NC_018010.1); 사이크로플렉서스 토르쿠이스I (NCBI Ref: NC_018721.1); 스타필로코커스 써모필러스 (NCBI Ref: YP_820832.1), 리스테리아 이노쿠

아 (NCBI Ref: NP_472073.1), 캄필로박터 제주니 (NCBI Ref: YP_002344900.1) 또는 나이세리아 메닌기티디스 (NCBI Ref: YP_002342100.1)로부터 기원하는 Cas9를 언급하거나, 임의의 다른 기원의 Cas9를 언급한다.

[0332] 변이체 및 이의 동족체를 포함하는, 추가의 Cas9 단백질 (예를 들어, 뉴클레아제 데드 Cas9 (dCas9), Cas9 닉카제 (nCas9), 또는 뉴클레아제 활성 Cas9)이 본원의 범위 내에 있는 것으로 인지해야 한다. 예시적인 Cas9 단백질은 제한 없이 하기에 제공된 것들을 포함한다. 일부 구현예에서, Cas9 단백질은 뉴클레아제 활성 Cas9이다. 일부 구현예에서, Cas9 단백질은 뉴클레아제 데드 Cas9 (dCas9)이다. 일부 구현예에서, Cas9 단백질은 Cas9 닉카제 (nCas9)이다.

[0333] 일부 구현예에서, Cas9 도메인은 뉴클레아제-불활성 Cas9 도메인 (dCas9)이다. 예를 들어, dCas9 도메인은 듀플렉스 핵산 분자에 듀플렉스 핵산 분자의 어느 가닥도 절단하는 것 없이 (예를 들어, gRNA 분자를 통해) 결합할 수 있다. 뉴클레아제-불활성화된 Cas9 단백질은 상호교환적으로 "dCas9" 단백질 (뉴클레아제-"데드" Cas9에 대해) 또는 촉매 불활성 Cas9로서 언급될 수 있다. 불활성 DNA 절단 도메인을 갖는 Cas9 단백질 (또는 이의 단편)을 생성하기 위한 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, 이의 각각의 전체 내용은 본원에 참조로 포함됨). 예를 들어, Cas9의 DNA 절단 도메인은 2개의 서브도메인인 HNH 뉴클레아제 서브도메인 및 RuvC1 서브도메인을 포함하는 것으로 공지되어 있다. HNH 서브도메인은 gRNA에 상보적인 가닥을 절단하는 반면 RuvC1 서브도메인은 비-상보적 가닥을 절단한다. 이들 서브도메인 내 돌연변이는 Cas9의 뉴클레아제 활성을 사일런싱시킬 수 있다. 예를 들어, 돌연변이 D10A 및 H840A는 에스. 피오게네스 Cas9의 뉴클레아제 활성을 완전히 불활성화시킨다 (참조: Jinek *et al., Science.* 337:816-821(2012); Qi *et al., Cell.* 28;152(5):1173-83 (2013)).

[0334] 일부 구현예에서, dCas9는 Cas9 뉴클레아제 활성을 불활성화시키는 하나 이상의 돌연변이를 갖는 Cas9 아미노산 서열에 부분적으로 또는 전반적으로 상응하거나 포함한다. 일부 구현예에서, 뉴클레아제-불활성 dCas9 도메인은 본원에 제시된 아미노산 서열의 D10X 돌연변이 및 H840X 돌연변이, 또는 본원에 제공된 임의의 아미노산 서열에서 상응하는 돌연변이를 포함하고, 여기서, X는 임의의 아미노산 변화이다. 일부 구현예에서, 뉴클레아제-불활성 dCas9 도메인은 본원에 제시된 아미노산 서열의 D10A 돌연변이 및 H840A 돌연변이, 또는 본원에 제공된 임의의 아미노산 서열에서 상응하는 돌연변이를 포함한다. 일부 구현예에서, 뉴클레아제-불활성 Cas9 도메인은 클로닝 벡터 pPlatTET-gRNA2 (수탁 번호 BAV54124)에 제시된 아미노산 서열을 포함한다. 일부 구현예에서, dCas9는 dCas9 (D10A 및 H840A)의 아미노산 서열을 포함한다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0335]

[0336] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0337] 일부 구현예에서, 예시적인 촉매적 불활성 Cas9 (dCas9)의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0338]

[0339] (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, 이의 전체 내용은 본원에 참조로 포함됨).

[0340] 일부 구현예에서, 예시적인 촉매적 불활성 Cas9 (dCas9)의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0341]

[0342] 일부 구현예에서, Cas9 도메인은 D10A 돌연변이를 포함하고, 위치 840에서 잔기는 상기 제공된 아미노산 서열에서, 또는 본문에 제공된 임의의 아미노산 서열에서 상응하는 위치에 히스티딘을 유지한다. 다른 구현예에서, D10A 및 H840A 이외의 돌연변이를 갖는 dCas9 변이체가 제공되고, 상기 변이체는 예를 들어, 뉴클레아제 불활성화된 Cas9 (dCas9)를 유도한다. 상기 돌연변이는 예를 들어 D10 및 H840에서 다른 아미노산 치환, 또는 Cas9의 뉴클레아제 도메인 내 다른 치환 (예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인에서의 치환)을 포함한다. 일부 구현예에서, dCas9의 변이체 또는 동족체가 제공되고, 이는 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, dCas9의 변이체가 제공되고, 이는 약 5개 아미노산, 약 10개 아미노산, 약 15개 아미노산, 약 20개 아미노산, 약 25개 아미노산, 약 30개 아미노산, 약 40개 아미노산, 약 50개 아미노산, 약 75개 아미노산, 약 100개 이상의 아미노산만큼 더 짧거나 더 긴 아미노산 서열을 갖는다.

[0343] 추가의 적합한 뉴클레아제-불활성 dCas9 도메인은 본원 개시내용 및 당해 분야의 지식을 기준으로 당업자에게 자명할 것이고 본원 개시내용의 범위 내에 있다. 상기 추가의 예시적인 적합한 뉴클레아제-불활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A, 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하지만 이에 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, 이의 전체 내용은 본원에 참조로 포함됨).

[0344] 일부 구현예에서, Cas9 뉴클레아제는 불활성 (예를 들어, 불활성화된) DNA 절단 도메인을 갖고, 즉, Cas9는 "nCas9" 단백질 ("닉카제" Cas9에 대해)로서 언급되는 닉카제이다. 일부 구현예에서, Cas9 도메인은 Cas9 닉카제이다. Cas9 닉카제는 듀플렉스 핵산 분자 (예를 들어, 듀플렉스 DNA 분자)의 단 하나의 가닥을 절단할 수 있는 Cas9 단백질일 수 있다. 일부 구현예에서, Cas9 닉카제는 듀플렉스 핵산 분자의 표적 가닥을 절단하고, Cas9 닉카제가 Cas9에 결합된 gRNA (예를 들어, sgRNA)와 쌍을 형성하는 (이에 상보적인) 염기인 가닥을 절단함

을 의미한다. 일부 구현예에서, Cas9 닉카제는 D10A 돌연변이를 포함하고 위치 840에 히스티딘을 갖는다. 일부 구현예에서, Cas9 닉카제는 듀플렉스 핵산 분자의 비-표적, 비-염기-편집 가닥을 절단하고, 이는 Cas9 닉카제가 Cas9에 결합된 gRNA (예를 들어, sgRNA)와 쌍을 형성하는 염기가 아닌 가닥을 절단함을 의미한다. 일부 구현예에서, Cas9 닉카제는 H840A 돌연변이를 포함하고 10번 위치에 아스파르트산 잔기 또는 상응하는 돌연변이를 갖는다. 일부 구현예에서, Cas9 닉카제는 본원에 제공된 Cas9 닉카제 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 추가의 적합한 Cas9 뉴클레아제는 본원 개시내용 및 당해 분야의 지식에 기반하여 당업자에게 자명할 것이고 본원 개시내용의 범위 내에 있다.

[0345] 예시적인 촉매적 Cas9 닉카제 (nCas9)의 아미노산 서열은 다음과 같다:

[0346]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0347]

[0348] 일부 구현예에서, Cas9는 고세균 (예를 들어, 나노고세균) 기원의 Cas9를 언급하고, 이것은 단세포 원핵 미생물의 도메인 및 킹덤을 구성한다. 일부 구현예에서, 핵산 프로그래밍 가능한 DNA 결합 단백질은 CasX 또는 CasY 단백질일 수 있고, 이는 예를 들어, 이의 전체 내용이 참조로 포함되는 문헌 (참조: Burstein *et al.*, "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21)에 기재되어 있다. 게놈 분리 균유전체학을 사용하여, 생활 고세균 도메인에서 최초 보고된 Cas9를 포함하는, 다수의 CRISPR-Cas 시스템을 동정하였다. 상기 다양한 Cas9 단백질은 활성 CRISPR-Cas 시스템의 일부로서 거의 연구되지 않은 나노고세균에서 발견되었다. 세균에서, 2개의 이전에 공지되지 않은 시스템인 CRISPR-CasX 및 CRISPR-CasY가 발견되었고, 이는 지금까지 발견된 가장 콤팩트한 시스템 중 하나이다. 일부 구현예에서, 본원에 기재된 염기 편집기 시스템에서, Cas9는 CasX, 또는 CasX의 변이체에 의해 대체된다. 일부 구현예에서, 본원에 기재된 염기 편집기 시스템에서, Cas9는 CasY, 또는 CasY의 변이체에 의해 대체된다. 다른 RNA-가이드된 DNA 결합 단백질이 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)로서 사용될 수 있고 본원 개시내용의 범위 내에 있음을 인지해야 한다.

[0349] 일부 구현예에서, 본원에 제공된 임의의 융합 단백질의 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)은 CasX 또는 CasY 단백질일 수 있다. 일부 구현예에서, napDNAbp는 CasX 단백질이다. 일부 구현예에서, napDNAbp는 CasY 단백질이다. 일부 구현예에서, napDNAbp는 천연적으로 발생하는 CasX 또는 CasY 단백질과 적

어도 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 구현예에서, napDNAbp는 천연적으로 발생하는 CasX 또는 CasY 단백질이다. 일부 구현예에서, napDNAbp는 본원에 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 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종 기원의 CasX 및 CasY가 또한 본원 개시내용에 따라 사용될 수 있음을 인지해야 한다.

[0350] 예시적인 CasX ((uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53) tr|F0NN87|F0NN87_SULIHCRISPR-연합된 Casx 단백질 OS = 설폴로부스 이슬란디쿠스 (균주 HVE10/4) GN = SiH_0402 PE=4 SV=1) 아미노산 서열은 다음과 같다:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKK
KKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPE
FYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLI
QNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLE
KRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLYFANRDLIMNLNSDDG
KVRDLKLISAYVNGELIRGEG

[0351]

[0352] 예시적인 CasX (>tr|F0NH53|F0NH53_SULIR CRISPR 연합된 단백질, Casx OS = 설폴로부스 이슬란디쿠스 (균주 REY15A) GN=SiRe_0771 PE=4 SV=1) 아미노산 서열은 다음과 같다:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKK
KKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPE
FYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLI
QNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLE
KRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGK
VRDLKLISAYVNGELIRGEG

[0353]

[0354] 델타프로테오박테리아 CasX

MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAANNL
RMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTAGFACSQ
CGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLGKFGQRALDF
YSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIIIEHQKVVKGNQKR
LESLRELAGKENLEYPSVTLPPQPHTKEGVDfAYNEVIARVRMWVNLNLWQKLKLSRDDAKPLLR
LKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEKRNTILEGYNYLPNENDH
KKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERIDKKIAGLTSHIEREEARNAED
AQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYGDLRGNPFAVEAENRVVDISGFSI
GSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRFTDGTDIKKSGKWQGLLYGGGKAKVID
LTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETGLIKLANGRVIEKTIYNKKIGRDEPALFVA
LTFERREVVDPSNIKPVNLIGVARGENIPAVIALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEK
QRAIQAAKEVEQRRAGGYSRKFASKSRNLADDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQG
KRTFMTERQYTKMEDWLTAKLAYEGLTSKTYLSKTLAQYTSKTCSNCGFTITYADMDVMLVRLKK
TSDGWATTLNNKELKAEYQITYYNRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFL
LKKRFSHRPVQEQFVCLDCGHEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFY
KRRLKEVWKPNA

[0355]

[0356] 예시적인 CasY ((ncbi.nlm.nih.gov/protein/APG80656.1)>APG80656.1 CRISPR-연합된 단백질 CasY [배양되지 않은 파쿠박테리아 그룹 박테리움]) 아미노산 서열은 다음과 같다:

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGL
SNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLY
DELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASL
GERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKN
EGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRI
LAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGE
SDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKE
RLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTD
ALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDI
VSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEE
YIDLIELHKTALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSE
LRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEP
ESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQ
NKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSER
VFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVK
GLKLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYA
TLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGT
VRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADI
QASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI

[0357]

[0358] Cas9 뉴클레아제는 2개의 기능성 엔도뉴클레아제 도메인을 갖는다: RuvC 및 HNH. Cas9는 표적 DNA의 반대 가닥을 절단하기 위해 뉴클레아제 도메인을 위치시키는 표적 결합시 형태적 변화를 진행한다. Cas9-매개된 DNA 절단의 최종 결과는 표적 DNA (PAM 서열의 업스트림의 약 3 내지 4개 뉴클레오타이드) 내 이중 가닥 절단 (DSB)이다. 이어서, 생성된 DSB는 2개의 일반 복구 경로 중 하나에 의해 복구된다: (1) 효율적이지만 오류 성향 비-상동성 말단 연결 (NHEJ) 경로; 또는 (2) 덜 효율적이지만 고충실도 상동성 지시된 복구 (HDR) 경로.

[0359] 비-상동성 말단 연결 (NHEJ) 및/또는 상동성 지시된 복구 (HDR)의 "효율"은 임의의 편리한 방법에 의해 계산될 수 있다. 예를 들어, 효율은 성공적인 HDR의 백분율로 표현될 수 있다. 예를 들어, 서베이어 (surveyor) 뉴클레아제 검정을 사용하여 절단 생성물을 생성할 수 있고 기질에 대한 생성물의 비율을 사용하여 백분율을 계산할 수 있다. 예를 들어, 서베이어 뉴클레아제 효소를 사용하여 성공적인 HDR의 결과로서 새롭게 통합된 제한 서열을 포함하는 DNA를 직접 절단할 수 있다. 더 절단된 기질은 보다 큰 퍼센트의 HDR (보다 큰 HDR 효율)을 지적한다. 예시적인 예로서, HDR의 분율 (백분율)은 하기의 식을 사용하여 계산될 수 있다: [(절단 생성물)/(기질 + 절단 생성물)] (예를 들어, (b+c)/(a+b+c), 여기서, "a"는 DNA 기질의 밴드 강도이고 "b" 및 "c"는 절단 생성물이다).

[0360] 일부 구현예에서, 효율은 성공적인 NHEJ의 백분율로 표현될 수 있다. 예를 들어, T7 엔도뉴클레아제 I 검정을 사용하여 절단 생성물을 생성할 수 있고 기질에 대한 생성물의 비율을 사용하여 NHEJ의 백분율을 계산할 수 있다. T7 엔도뉴클레아제 I은 야생형 및 돌연변이체 DNA 가닥의 하이브리드화로부터 비롯된 미스매칭된 헤테로듀플렉스 DNA를 절단한다 (NHEJ는 본래의 절단 부위에서 소형 무작위 삽입 또는 결실 (indel)을 생성한다). 보다 많은 절단은 보다 큰 퍼센트의 NHEJ (보다 큰 NHEJ의 효율)를 지적한다. 예시적인 예로서, NHEJ의 분율 (백분율)은 하기의 식을 사용하여 계산될 수 있다: $(1-(1-(b+c)/(a+b+c))^{1/2})\times 100$, 여기서, "a"는 DNA 기질의 밴드 강도이고, "b" 및 "c"는 절단 생성물이다 (참조: Ran *et. al.*, Cell. 2013 Sep. 12; 154(6):1380-9; and Ran *et al.*, Nat Protoc. 2013 Nov.; 8(11): 2281-2308).

[0361] NHEJ 복구 경로는 가장 활성의 복구 기전이고, 이것은 흔히 DSB 부위에 소형 뉴클레오타이드 삽입 또는 결실 (indel)을 유발한다. NHEJ-매개된 DSB 복구의 무작위는 Cas9 및 gRNA 또는 가이드 폴리뉴클레오타이드를 발현하는 세포 집단이 다양한 어레이의 돌연변이를 유도할 수 있기 때문에 중요한 수행 관련성을 갖는다. 일부 구현예에서, NHEJ는 표적 DNA에 소형 삽입-결실을 유발하여 아미노산 결실, 삽입 또는 표적화된 유전자의 개방 판독 프레임 (ORF) 내 미성숙한 정지 코돈을 유도하는 프레임쉬프트 돌연변이를 초래한다. 이상적인 최종 결과는 표적화된 유전자 내 기능 상실 돌연변이이다.

[0362] NHEJ-매개된 DSB 복구는 흔히 유전자의 개방 판독 프레임을 붕괴시키고, 상동성 지시된 복구 (HDR)를 사용하여

단일 뉴클레오타이드 변화에서 형광단 또는 태그의 부가와 같은 대형 삽입에 이르는 특이적 뉴클레오타이드 변화를 생성할 수 있다.

[0363] 유전자 편집을 위해 HDR을 사용하기 위해, 목적하는 서열을 포함하는 DNA 복구 주형은 gRNA(들) 및 Cas9 또는 Cas9 닉카제와 함께 관심 대상의 세포 유형에 전달될 수 있다. 복구 주형은 목적하는 편집은 물론 표적의 바로 업스트림 및 다운스트림에 있는 추가의 상동성 서열 (좌측 및 우측 상동성 아암으로 칭함)을 포함할 수 있다. 각각의 상동성 아암의 길이는 도입되는 변화의 크기에 좌우될 수 있고, 보다 큰 삽입은 보다 긴 상동성 아암을 요구한다. 복구 주형은 단일 가닥 올리고뉴클레오타이드, 이중 가닥 올리고뉴클레오타이드 또는 이중 가닥 DNA 플라스미드일 수 있다. HDR의 효율은 Cas9, gRNA 및 외인성 복구 주형을 발현하는 세포에서도 일반적으로 낮다 (<10%의 변형된 대립유전자). HDR의 효율은 HDR이 세포 주기의 S 및 G2기 동안에 발생하기 때문에 세포를 동조(synchronizing)시킴으로써 증진될 수 있다. NHEJ에 관여하는 유전자를 화학적으로 또는 유전학적으로 억제하면 HDR 빈도를 증가시킬 수도 있다.

[0364] 일부 구현예에서, Cas9는 변형된 Cas9이다. 소정의 gRNA 표적화 서열은 부분 상동성이 존재하는 게놈 전반에 걸쳐 추가의 부위를 가질 수 있다. 이들 부위는 표적-이탈으로 불리우고 gRNA를 디자인하는 경우 고려될 필요가 있다. gRNA 디자인을 최적화시키는 것에 추가로, CRISPR 특이성은 또한 Cas9로의 변형을 통해 증가될 수 있다. Cas9는 2개의 뉴클레아제 도메인, RuvC 및 HNH의 조합 활성을 통해 이중 가닥 절단 (DSB)을 생성한다. SpCas9의 D10A 돌연변이체인 Cas9 닉카제는 하나의 뉴클레아제 도메인을 보유하고 DSB가 아닌 DNA 닉 (nick)을 생성한다. 닉카제 시스템은 또한 특이적 유전자 편집을 위해 HDR-매개된 유전자 편집과 조합될 수 있다.

[0365] 일부 구현예에서, Cas9는 변이체 Cas9 단백질이다. 변이체 Cas9 폴리펩타이드는 야생형 Cas9 단백질의 아미노산 서열과 비교하는 경우 하나의 아미노산이 상이한 (예를 들어, 결실, 삽입, 치환, 융합을 갖는) 아미노산 서열을 갖는다. 일부 경우에, 변이체 Cas9 폴리펩타이드는 Cas9 폴리펩타이드의 뉴클레아제 활성을 감소시키는 아미노산 변화 (예를 들어, 결실, 삽입 또는 치환)를 갖는다. 예를 들어, 일부 경우에, 변이체 Cas9 폴리펩타이드는 상응하는 야생형 Cas9 단백질의 뉴클레아제 활성의 50% 미만, 40% 미만, 30% 미만, 20% 미만, 10% 미만, 5% 미만, 또는 1% 미만을 갖는다. 일부 구현예에서, 변이체 Cas9 단백질은 실질적인 뉴클레아제 활성을 갖지 않는다. 대상 Cas9 단백질이 실질적인 뉴클레아제 활성을 갖지 않는 변이체 Cas9 단백질인 경우, 이것은 "dCas9"로서 언급될 수 있다.

[0366] 일부 구현예에서, 변이체 Cas9 단백질은 감소된 뉴클레아제 활성을 갖는다. 예를 들어, 변이체 Cas9 단백질은 야생형 Cas9 단백질, 예를 들어, 야생형 Cas9 단백질의 엔도뉴클레아제 활성의 약 20% 미만, 약 15% 미만, 약 10% 미만, 약 5% 미만, 약 1% 미만, 또는 약 0.1% 미만을 나타낸다.

[0367] 일부 구현예에서, 변이체 Cas9 단백질은 가이드 표적 서열의 상보적 가닥을 절단할 수 있지만 이중 가닥 가이드 표적 서열의 비-상보적 가닥을 절단하는 감소된 능력을 갖는다. 예를 들어, 변이체 Cas9 단백질은 RuvC 도메인의 기능을 감소시키는 돌연변이 (아미노산 치환)를 가질 수 있다. 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 D10A (아미노산 위치 10에서 아스파르테이트에서 알라닌으로)를 갖고 따라서 이중 가닥 가이드 표적 서열의 상보적 가닥을 절단할 수 있지만 이중 가닥 가이드 표적 서열의 비-상보성 가닥을 절단하는 감소된 능력 (따라서, 변이체 Cas9 단백질이 이중 가닥 표적 핵산을 절단하는 경우 이중 가닥 절단 (DSB) 대신 단일 가닥 절단 (SSB)을 유도한다)을 갖는다 (예를 들어, 문헌 (Jinek *et al.*, Science. 2012 Aug. 17; 337(6096):816-21) 참조).

[0368] 일부 구현예에서, 변이체 Cas9 단백질은 이중 가닥 가이드 표적 서열의 비-상보성 가닥을 절단할 수 있지만 가이드 표적 서열의 상보성 가닥을 절단하는 감소된 능력을 갖는다. 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 구현예에서, 변이체 Cas9 단백질은 H840A (아미노산 위치 840에서 히스티딘에서 알라닌으로) 돌연변이를 갖고 따라서 가이드 표적 서열의 비-상보성 가닥을 절단할 수 있지만 가이드 표적 서열의 상보성 가닥을 절단하는 감소된 능력 (따라서, 변이체 Cas9 단백질이 이중 가닥 가이드 표적 서열을 절단하는 경우 DSB 대신 SSB를 유도하는)을 갖는다. 상기 Cas9 단백질은 가이드 표적 서열 (예를 들어, 단일 가닥 가이드 표적 서열)을 절단하는 감소된 능력을 갖지만 가이드 표적 서열 (예를 들어, 단일 가닥 가이드 표적 서열)에 결합하는 능력을 보유한다.

[0369] 일부 구현예에서, 변이체 Cas9 단백질은 이중 가닥 표적 DNA의 상보성 및 비-상보성 가닥 둘 다를 절단하는 감소된 능력을 갖는다. 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 D10A 및 H840A 돌연변이 둘 다를 함유하여, 상기 폴리펩타이드는 이중 가닥 표적 DNA의 상보성 및 비-상보성 가닥 둘 다를 절단하는 감소된

능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다.

[0370] 또 다른 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 W476A 및 W1126A 돌연변이를 함유하여 폴리펩타이드는 표적 DNA를 절단하는 감소된 능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다.

[0371] 또 다른 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 P475A, W476A, N477A, D1125A, W1126A, 및 D1127A 돌연변이를 함유하여 폴리펩타이드는 표적 DNA를 절단하는 감소된 능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다.

[0372] 또 다른 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 H840A, W476A 및 W1126A 돌연변이를 함유하여 폴리펩타이드는 표적 DNA를 절단하는 감소된 능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다. 또 다른 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 H840A, D10A, W476A 및 W1126A 돌연변이를 함유하여 폴리펩타이드는 표적 DNA를 절단하는 감소된 능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다. 일부 구현예에서, 변이체 Cas9는 Cas9 HNH 도메인에서 위치 840에서 촉매 His 잔기 (A840H)를 복구하였다.

[0373] 또 다른 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 H840A, P475A, W476A, N477A, D1125A, W1126A, 및 D1127A 돌연변이를 함유하여 폴리펩타이드는 표적 DNA를 절단하는 감소된 능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다. 또 다른 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, 및 D1127A 돌연변이를 함유하여 폴리펩타이드는 표적 DNA를 절단하는 감소된 능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다. 일부 구현예에서, 변이체 Cas9 단백질이 W476A 및 W1126A 돌연변이를 함유하는 경우 또는 상기 변이체 Cas9 단백질이 P475A, W476A, N477A, D1125A, W1126A, 및 D1127A 돌연변이를 함유하는 경우, 변이체 Cas9 단백질은 효율적으로 PAM 서열에 결합하지 않는다. 따라서, 일부 상기 구현예에서, 상기 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 방법은 PAM 서열을 요구하지 않는다. 다른 말로, 일부 구현예에서, 상기 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 방법은 가이드 RNA를 포함할 수 있지만, 방법은 PAM 서열의 부재하에 수행될 수 있다 (그리고, 결합 특이성은 따라서 가이드 RNA의 표적화 분절에 의해 제공된다). 다른 잔기는 상기 효과를 성취하기 위해 돌연변이될 수 있다 (즉, 하나 또는 다른 뉴클레아제 부분을 불활성화시킬 수 있다). 비제한적인 예로서, 잔기 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, 및/또는 A987은 변경 (즉, 치환된)될 수 있다. 또한, 알라닌 치환과는 다른 돌연변이가 적합하다.

[0374] 일부 구현예에서, 감소된 촉매 활성 (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 갖는 경우)을 갖는 변이체 Cas9 단백질은 이것이 가이드 RNA와 상호작용하는 능력을 보유하는 한 부위 특이적 방식으로 (이것은 여전히 가이드 RNA에 의해 표적 DNA 서열에 가이드되기 때문에) 표적 DNA에 여전히 결합할 수 있다.

[0375] 일부 구현예에서, 변이체 Cas 단백질은 spCas9, spCas9-VRQR, spCas9-VRER, xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK, 또는 spCas9-LRVSQL일 수 있다.

[0376] 일부 구현예에서, 아미노산 치환체 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R (SpCas9-MQKFRAER)을 포함하고 변경된 PAM 5'-NGC-3'에 대해 특이성을 갖는 변형된 SpCas9가 사용되었다.

[0377] 에스. 피오게네스 Cas9에 대한 대안은 포유동물 세포에서 절단 활성을 나타내는 Cpf1 패밀리로부터 RNA-가이드된 엔도뉴클레아제를 포함할 수 있다. 프레보텔라 (*Prevotella*) 및 프란시셀라 1 (*Francisella 1*)로부터의 CRISPR(CRISPR/Cpf1)은 CRISPR/Cas9 시스템과 유사한 DNA-편집 기술이다. Cpf1은 부류 II CRISPR/Cas 시스템의 RNA-가이드된 엔도뉴클레아제이다. 이와 같이 획득된 면역 기전은 프레보텔라 (*Prevotella*) 및 프란시셀라

(*Francisella*) 세균에서 발견된다. Cpf1 유전자는 바이러스 DNA를 발견하고 절단하기 위해 가이드 RNA를 사용하는 엔도뉴클레아제를 암호화하는, CRISPR 유전자좌와 연합되어 있다. Cpf1은 Cas9 보다 소형이거나 보다 단순한 엔도뉴클레아제이고 CRISPR/Cas9 시스템 한계의 일부를 극복한다. Cas9 뉴클레아제와 달리, Cpf1-매개된 DNA 절단의 결과는 짧은 3' 오버행과 함께 이중 가닥 절단이다. Cpf1의 엇갈린 절단 패턴은 통상적인 제한 효소 클로닝과 유사하게, 방향성 유전자 전달 가능성을 열어 유전자 편집의 효율을 증가시킬 수 있다. 상기 기재된 Cas9 변이체 및 오톨로그 처럼, Cpf1은 또한 CRISPR에 의해 SpCas9가 선호하는 NGG PAM 부위가 없는 AT-풍부 영역 또는 AT-풍부 게놈에 표적화될 수 있는 부위의 수를 증대시킬 수 있다. Cpf1 유전자좌는 혼합된 알파/베타 도메인, RuvC-I에 이어서 나선 영역, RuvC-II 및 아연 핑거 유사 도메인을 포함한다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사한 RuvC-유사 엔도뉴클레아제 도메인을 갖는다.

[0378] 추가로, Cas9와 달리, Cpf1은 HNH 엔도뉴클레아제 도메인을 갖지 않고, Cpf1의 N-말단은 Cas9의 알파-나선 인지 돌출부를 갖지 않는다. Cpf1 CRISPR-Cas 도메인 구조는 Cpf1이 기능적으로 특유하고, 부류 2, V형 CRISPR 시스템으로서 분류됨을 보여준다. Cpf1 유전자좌는 II형 시스템보다 I형 및 III형에 보다 유사한 Cas1, Cas2 및 Cas4 단백질을 암호화한다. 기능적 Cpf1은 트랜스-활성화 CRISPR RNA (tracrRNA)를 필요로 하지 않고 따라서 CRISPR (crRNA)만이 요구된다. 이것은 게놈 편집에 이로운데, 이는 Cpf1이 Cas9보다 소형인 것 뿐만 아니라 이것은 보다 소형의 sgRNA 분자 (대략적으로 Cas9 만큼 많은 뉴클레오타이드의 절반)를 갖기 때문이다. Cpf1-crRNA 복합체는 Cas9에 의해 표적화된 G-풍부 PAM과는 대조적으로 프로토스페이서 인접 모티프 5'-YTN-3' 또는 5'-TTN-3'의 동정에 의해 표적 DNA 또는 RNA를 절단한다. PAM의 동정 후, Cpf1은 4 또는 5개 뉴클레오타이드 오버행을 갖는 점성 말단 유사 DNA 이중 가닥 절단을 도입한다.

[0379] 일부 구현예에서, Cas9는 변경된 PAM 서열에 대해 특이성을 갖는 Cas9이다. 일부 구현예에서, 추가 Cas9 변이체 및 PAM 서열은 문헌 (참조: Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020), 이의 전문이 본원에 참조로 포함됨)에 기재되어 있다. 일부 구현예에서, Cas9 변이체는 특이적 PAM 요건을 갖지 않는다. 일부 구현예에서, Cas9 변이체, 예를 들어, SpCas9 변이체는 NRNH PAM에 대해 특이성을 가지며, 여기서, R은 A 또는 G이고 H는 A, C 또는 T이다. 일부 구현예에서, SpCas9 변이체는 PAM 서열 AAA, TAA, CAA, GAA, TAT, GAT, 또는 CAC에 대해 특이성을 갖는다. 일부 구현예에서, SpCas9 변이체는 서열번호 1에 넘버링된 바와 같이 위치 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1218, 1219, 1221, 1249, 1256, 1264, 1290, 1318, 1317, 1320, 1321, 1323, 1332, 1333, 1335, 1337, 또는 1339에서 또는 이의 상응하는 위치에서 아미노산 치환을 포함한다. 일부 구현예에서, SpCas9 변이체는 서열번호 1에 넘버링된 바와 같이 위치 1114, 1135, 1218, 1219, 1221, 1249, 1320, 1321, 1323, 1332, 1333, 1335, 또는 1337에서 또는 이의 상응하는 위치에서 아미노산 치환을 포함한다. 일부 구현예에서, SpCas9 변이체는 서열번호 1에 넘버링된 바와 같이 위치 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1219, 1221, 1256, 1264, 1290, 1318, 1317, 1320, 1323, 1333에서 또는 이의 상응하는 위치에서 아미노산 치환을 포함한다. 일부 구현예에서, SpCas9 변이체는 서열번호 1에 넘버링된 바와 같이 위치 1114, 1131, 1135, 1150, 1156, 1180, 1191, 1218, 1219, 1221, 1227, 1249, 1253, 1286, 1293, 1320, 1321, 1332, 1335, 1339에서 또는 이의 상응하는 위치에서 아미노산 치환을 포함한다. 일부 구현예에서, SpCas9 변이체는 서열번호 1에 넘버링된 바와 같이 위치 1114, 1127, 1135, 1180, 1207, 1219, 1234, 1286, 1301, 1332, 1335, 1337, 1338, 1349에서 또는 이의 상응하는 위치에서 아미노산 치환을 포함한다. SpCas9 변이체의 예시적인 아미노산 치환 및 PAM 특이성은 **표 1A-1D**에 나타낸다.

[0380] [표 1A]

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

[0381]

[0382] [표 1B]

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

[0383]

[0384] [표 1C]

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	

[0385]

[0386] [표 1D]

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

[0387]

[0388] 일부 구현예에서, Cas9는 나이세리아 메닌기티디스 Cas9 (NmeCas9) 또는 이의 변이체이다. 일부 구현예에서, NmeCas9는 NNNNGAYW PAM에 대해 특이성을 가지고, 여기서, Y는 C 또는 T이고, W는 A 또는 T이다. 일부 구현예에서, NmeCas9는 NNNNGYTT PAM에 대해 특이성을 가지고, 여기서, Y는 C 또는 T이다. 일부 구현예에서, NmeCas9는 NNNNGTCT PAM에 대해 특이성을 갖는다. 일부 구현예에서, NmeCas9는 Nme1 Cas9이다. 일부 구현예에서, 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 대해 특이성을 갖는다. 일부 구현예에서, Nme1Cas9는 NNNNGATT PAM, NNNNCCTA PAM, NNNNCCTC PAM, NNNNCCTT PAM, 또는 NNNNCCTG PAM에 대해 특이성을 갖는다. 일부 구현예에서, NmeCas9는 CAA PAM, CAAA PAM, 또는 CCA PAM에 대해 특이성을 갖는다. 일부 구현예에서, NmeCas9는 Nme2 Cas9이다. 일부 구현예에서, NmeCas9는 NNNNCC (N4CC) PAM에 대해 특이성을 가지고, 여기서, N은 A, G, C, 또는 T 중 임의의 하나이다. 일부 구현예에서, NmeCas9는 NNNNCCGT PAM, NNNNCCGGPAM, NNNNCCCA PAM, NNNNCCCT PAM, NNNNCCCC PAM, NNNNCCAT PAM, NNNNCCAG PAM, NNNNCCAT PAM, 또는 NNNGATT PAM에 대해 특이성을 갖는다. 일부 구현예에서, NmeCas9는 Nme3Cas9이다. 일부 구현예에서, NmeCas9는 NNNNCAAA PAM, NNNNCC PAM, 또는 NNNNCNNN PAM에 대해 특이성을 갖는다. 추가 NmeCas9 특징 및 PAM 서열은 문헌 (참조: Edraki *et al.* Mol. Cell. (2019) 73(4): 714-726)에 기재되어 있으며 상기 문헌은 이의 전문이 본원에 참조로 포함된다.

[0389] Nme1Cas9의 예시적인 아미노산 서열이 아래에 제공된다:

[0390] II형 CRISPR RNA-가이드된 엔도뉴클레아제 Cas9 [나이세리아 메닌기티디스] WP_002235162.1

```
   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
 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
```

[0391]

[0392] Nme2Cas9의 예시적인 아미노산 서열이 아래에 제공된다:

[0393] II형 CRISPR RNA-가이드된 엔도뉴클레아제 Cas9 [나이세리아 메닌기티디스] WP_002230835.1

```
   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
```

[0394]

[0395] *핵염기 편집기의 Cas12 도메인*

[0396] 전형적으로, 미생물 CRISPR-Cas 시스템은 부류 1 및 부류 2 시스템으로 분류된다. 부류 1 시스템은 멀티서브유닛 이펙터 복합체를 갖고, 부류 2 시스템은 단일 단백질 이펙터를 갖는다. 예를 들어, Cas9 및 Cpf1은 부류 2 이펙터이지만 상이한 유형 (각각 II형 및 V형)이다. Cpf1에 추가로, 부류 2, 유형 V CRISPR-Cas 시스템은 또한 Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 및 Cas12i를 포함한다.

예를 들어, 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; and Yan et al., "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91)을 참조하고; 이의 각각의 전체 내용은 본원에 참조로 포함된다. 유형 V Cas 단백질은 RuvC (또는 RuvC-유사) 엔도뉴클레아제 도메인을 함유한다. 성숙한 CRISPR RNA (crRNA)의 생성은 일반적으로 tracrRNA-독립적이지만, Cas12b/C2c1은 예를 들어, crRNA의 생성을 위해 tracrRNA를 필요로 한다. Cas12b/C2c1은 DNA 절단을 위해 crRNA 및 tracrRNA 둘 다에 의존한다.

[0397] 본원 개시내용에 고려되는 핵산 프로그래밍 가능한 DNA 결합 단백질은 부류 2, 유형 V (Cas12 단백질)로서 분류된 Cas 단백질을 포함한다. Cas 부류 2, 유형 V 단백질의 비제한적인 예는 Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 및 Cas12i, 이의 상동체 또는 이의 변형된 버전을 포함한다. 본원에 사용된 바와 같은 Cas12 단백질은 또한 Cas12 뉴클레아제, Cas12 도메인, 또는 Cas12 단백질 도메인으로서 언급될 수 있다. 일부 구현예에서, 본원 개시내용의 Cas12 단백질은 데아미나제 도메인과 같은 내부적으로 융합된 단백질 도메인에 의해 중단된 아미노산 서열을 포함한다.

[0398] 일부 구현예에서, Cas12 도메인은 뉴클레아제 불활성 Cas12 도메인 또는 Cas12 닉카제이다. 일부 구현예에서, Cas12 도메인은 뉴클레아제 활성 도메인이다. 예를 들어, Cas12 도메인은 듀플렉스 핵산의 하나의 가닥 (예를 들어, 듀플렉스 DNA 분자)을 절단하는 Cas12 도메인일 수 있다. 일부 구현예에서, Cas12 도메인은 본원에 제시된 바와 같은 아미노산 서열 중 하나를 포함한다. 일부 구현예에서, Cas12 도메인은 본원에 제시된 아미노산 서열 중 임의의 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 구현예에서, Cas12 도메인은 본원에 제시된 아미노산 서열 중 임의의 하나와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 구현예에서, Cas12 도메인은 본원에 제시된 아미노산 서열 중 임의의 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 1200개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0399] 일부 구현예에서, Cas12의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 구현예에서, 단백질은 2개의 Cas12 도메인 중 하나를 포함한다: (1) Cas12의 gRNA 결합 도메인; 또는 (2) Cas12의 DNA 절단 도메인. 일부 구현예에서, Cas12 또는 이의 단편을 포함하는 단백질은 "Cas12 변이체"로서 언급된다. Cas12 변이체는 Cas12 또는 이의 단편과 상동성을 공유한다. 예를 들어, Cas12 변이체는 야생형 Cas12와 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, Cas12 변이체는 야생형 Cas12와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 아미노산 변화를 가질 수 있다. 일부 구현예에서, Cas12 변이체는 Cas12의 단편 (예를 들어, gRNA 결합 도메인 또는 DNA 절단 도메인)을 포함하여, 단편은 야생형 Cas12의 상응하는 단편과 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 96% 동일한, 적어도 약 97% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, 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%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%이다. 일부 구현예에서, 단편은 적어도 100개 아미노산 길이이다. 일부 구현예에서, 단편은 적어도 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 또는 적어도 1300개 아미노산 길이이다.

[0400] 일부 구현예에서, Cas12는 Cas12 뉴클레아제 활성을 변경하는 하나 이상의 돌연변이를 갖는 Cas12 아미노산 서열에 부분적으로 또는 전반적으로 상응하거나 포함한다. 상기 돌연변이는 예를 들어, Cas12의 RuvC 뉴클레아제

도메인 내 아미노산 치환을 포함한다. 일부 구현예에서, Cas12의 변이체 또는 동족체가 제공되고, 이는 야생형 Cas12와 적어도 약 70% 동일한, 적어도 약 80% 동일한, 적어도 약 90% 동일한, 적어도 약 95% 동일한, 적어도 약 98% 동일한, 적어도 약 99% 동일한, 적어도 약 99.5% 동일한, 또는 적어도 약 99.9% 동일하다. 일부 구현예에서, Cas12의 변이체가 제공되고, 이는 약 5개 아미노산, 약 10개 아미노산, 약 15개 아미노산, 약 20개 아미노산, 약 25개 아미노산, 약 30개 아미노산, 약 40개 아미노산, 약 50개 아미노산, 약 75개 아미노산, 약 100개 이상의 아미노산 만큼 더 짧거나 더 긴 아미노산 서열을 갖는다.

[0401] 일부 구현예에서, 본원에 제공된 바와 같은 Cas12 융합 단백질은 Cas12 단백질의 전장 아미노산 서열, 예를 들어, 본원에 제공된 Cas12 서열 중 하나를 포함한다. 그러나, 다른 구현예에서, 본원에 제공된 바와 같은 융합 단백질은 전장 Cas12 서열을 포함하지 않고 단지 하나 이상의 이의 단편을 포함한다. 적합한 Cas12 도메인의 예시적인 아미노산 서열이 본원에 제공되고, Cas12 도메인 및 단편의 추가의 적합한 서열은 당업자에게 자명할 것이다.

[0402] 일반적으로, 부류 2, 유형 V Cas 단백질은 단일 기능성 RuvC 엔도뉴클레아제 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 (참조: Strecker et al., Nature Communications, 2019, 10(1): Art. No.: 212). 하나의 구현예에서, 변이체 Cas12 폴리펩타이드는 야생형 Cas12 단백질의 아미노산 서열과 비교하는 경우 1, 2, 3, 4, 5개 이상의 아미노산이 상이한 (예를 들어, 결실, 삽입, 치환, 융합을 갖는) 아미노산 서열을 갖는다. 일부 경우에, 변이체 Cas12 폴리펩타이드는 Cas12 폴리펩타이드의 활성을 감소시키는 아미노산 변화 (예를 들어, 결실, 삽입 또는 치환)를 갖는다. 예를 들어, 일부 경우에, 변이체 Cas12는 상응하는 야생형 Cas12b 단백질의 닉카제 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 닉카제 활성을 갖지 않는다.

[0403] 일부 경우에, 변이체 Cas12b 단백질은 감소된 닉카제 활성을 갖는다. 예를 들어, 변이체 Cas12b 단백질은 야생형 Cas12b 단백질의 닉카제 활성의 약 20% 미만, 약 15% 미만, 약 10% 미만, 약 5% 미만, 약 1% 미만, 또는 약 0.1% 미만을 나타낸다.

[0404] 일부 구현예에서, Cas12 단백질은 포유동물 세포에서 활성을 나타내는 Cas12a/Cpf1 패밀리 기원의 RNA-가이드된 엔도뉴클레아제를 포함한다. 프레보텔라 (Prevotella) 및 프란시셀라 1 (Francisella 1)로부터의 CRISPR (CRISPR/Cpf1)은 CRISPR/Cas9 시스템과 유사한 DNA 편집 기술이다. Cpf1은 부류 II CRISPR/Cas 시스템의 RNA-가이드된 엔도뉴클레아제이다. 이와 같이 획득된 면역 기전은 프레보텔라 및 프란시셀라 세균에서 발견된다. Cpf1 유전자는 바이러스 DNA를 발견하고 절단하기 위해 가이드 RNA를 사용하는 엔도뉴클레아제를 암호화하는, CRISPR 유전자좌와 연합되어 있다. Cpf1은 Cas9보다 소형이거나 보다 단순한 엔도뉴클레아제이고 CRISPR/Cas9 시스템 한계의 일부를 극복한다. Cas9 뉴클레아제와 달리, Cpf1-매개된 DNA 절단의 결과는 짧은 3' 오버행과 함께 이중 가닥 절단이다. Cpf1의 엇갈린 절단 패턴은 통상적인 제한 효소 클로닝과 유사하게, 방향성 유전자 전달 가능성을 열어 유전자 편집의 효율을 증가시킬 수 있다. 상기 기재된 Cas9 변이체 및 오톨로그 처럼, Cpf1은 또한 CRISPR에 의해 SpCas9가 선호하는 NGG PAM 부위가 없는 AT-풍부 영역 또는 AT-풍부 게놈에 표적화될 수 있는 부위의 수를 증대시킬 수 있다. Cpf1 유전자좌는 혼합된 알파/베타 도메인, RuvC-I에 이어서 나선 영역, RuvC-II 및 아연 핑거 유사 도메인을 포함한다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사한 RuvC-유사 엔도뉴클레아제 도메인을 갖는다. 추가로, Cas9와 달리, Cpf1은 HNH 엔도뉴클레아제 도메인을 갖지 않고, Cpf1의 N-말단은 Cas9의 알파-나선 인지 돌출부를 갖지 않는다. Cpf1 CRISPR-Cas 도메인 구조는 Cpf1이 기능적으로 특유하고, 부류 2, V형 CRISPR 시스템으로서 분류됨을 보여준다. Cpf1 유전자좌는 II형 시스템보다 I형 및 III형과 보다 유사한 Cas1, Cas2 및 Cas4 단백질을 암호화한다. 기능적 Cpf1은 트랜스-활성화 CRISPR RNA (tracrRNA)를 요구하지 않고 따라서 CRISPR (crRNA)만이 요구된다. 이것은 게놈 편집에 이로운데 이는 Cpf1이 Cas9 보다 소형인 것 뿐만 아니라 이것은 보다 소형의 sgRNA 분자 (대략적으로 Cas9 만큼 많은 뉴클레오타이드의 절반)를 갖기 때문이다. Cpf1-crRNA 복합체는 Cas9에 의해 표적화된 G-풍부 PAM과는 대조적으로 프로토스페이서 인접 모티프 5'-YTN-3' 또는 5'-YTTN-3'의 동정에 의해 표적 DNA 또는 RNA를 표적화한다. PAM의 동정 후, Cpf1은 4 또는 5개 뉴클레오타이드 오버행을 갖는 점성 말단 유사 DNA 이중 가닥 절단을 도입한다.

[0405] 본원 개시내용의 일부 양상에서, 상응하는 야생형 효소와 관련하여, 돌연변이된 CRISPR 효소가 표적 서열을 함유하는 표적 폴리뉴클레오타이드의 하나의 가닥 또는 가닥 둘 다를 절단하는 능력이 부재인 CRISPR 효소를 암호화하는 벡터가 사용될 수 있다. Cas12는 야생형 예시적인 Cas12 폴리펩타이드 (예를 들어, 바실러스 히사시로부터의 Cas12)와 적어도 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%

서열 동일성 및/또는 서열 상동성 또는 적어도 약 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성 및/또는 서열 상동성을 갖는 폴리펩타이드를 언급할 수 있다. Cas12는 야생형 예시적인 Cas12 폴리펩타이드 (예를 들어, 바실러스 히사시 (BhCas12b), 바실러스 종 V3-13 (BvCas12b) 및 알리사이클로바실러스 액시디필러스 (Alicyclobacillus acidiphilus) (AaCas12b)로부터)와 최대 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성 및/또는 서열 상동성 또는 최대 약 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성 및/또는 서열 상동성을 갖는 폴리펩타이드를 언급할 수 있다. Cas12는 야생형 또는 결실, 삽입, 치환, 변이체, 돌연변이, 융합, 키메라 또는 이의 임의의 조합과 같은 아미노산 변화를 포함할 수 있는 변형된 형태의 Cas12 단백질을 언급할 수 있다.

[0406] 핵산 프로그래밍 가능한 DNA 결합 단백질 본원 개시내용의 일부 양상은 핵산 프로그램 가능한 DNA 결합 단백질로서 작용하고, 이를 사용하여 염기 편집기와 같은 단백질을 특정 핵산 (예를 들어, DNA 또는 RNA) 서열에 가이드할 수 있는, 도메인을 포함하는 융합 단백질을 제공한다. 특정 구현예에서, 융합 단백질은 핵산 프로그래밍 가능한 DNA 결합 단백질 도메인 및 데아미나제 도메인을 포함한다. 핵산 프로그래밍 가능한 DNA 결합 단백질의 비제한적인 예는 Cas9 (예를 들어, dCas9 및 nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 및 Cas12i를 포함한다. Cas 효소의 비제한적인 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (또한 Csn1 또는 Csx12로서 공지되어 있음), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, 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, II형 Cas 이펙터 단백질, V형 Cas 이펙터 단백질, VI형 Cas 이펙터 단백질, CARF, DinG, 이의 상동체, 또는 이의 변형된 또는 가공된 버전을 포함한다. 다른 핵산 프로그램 가능한 DNA 결합 단백질은 또한 본원 개시내용의 범위 내에 있지만, 이들은 구체적으로 본원 개시내용에 열거되지 않을 수 있다. 예를 들어, 문헌 (Makarova *et al.* "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" CRISPR 2018 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 참조하고, 이의 각각의 전체 내용은 본원에 참조로 포함된다.

[0407] Cas9와는 상이한 PAM 특이성을 갖는 핵산 프로그램 가능한 DNA-결합 단백질의 하나의 예는 프레보텔라 (*Prevotella*) 및 프란시셀라 (*Francisella* 1) (Cpf1)로부터 클러스터링된 규칙적 사이공간의 짧은 팔린드롬 반복체이다. Cas9와 유사하게, Cpf1은 또한 부류 2 CRISPR 이펙터이다. Cpf1이 Cas9와는 별개의 특성으로 강한 DNA 간섭을 매개하는 것으로 나타났다. Cpf1은 tracrRNA 부재의 단일의 RNA-가이드된 엔도뉴클레아제이고, 이것은 T-풍부 프로토스페이서-인접 모티프(TTN, TTTN, 또는 YTN)를 사용한다. 더욱이, Cpf1은 엇갈린 이중 가닥 절단을 통해 DNA를 절단한다. 16개 Cpf1-패밀리 단백질 중에서, 액시다미노코커스 (Acidaminococcus) 및 라크노스피라세아 (Lachnospiraceae)로부터의 2개의 효소는 인간 세포에서 효율적인 게놈-편집 활성을 갖는 것으로 나타난다. Cpf1 단백질은 당업계에 공지되어 있고, 예를 들어, 문헌 (참조: Yamano *et al.*, "Crystal structure of Cpf1 in complex with guide RNA and target DNA." Cell (165) 2016, p. 949-962; 이의 전체 내용은 본원에 참조로 포함됨)에 이전에 기재되어 있다.

[0408] 본 발명의 조성물 및 방법에 또한 유용한 것은 가이드 뉴클레오타이드 서열 프로그래밍 가능한 DNA-결합 단백질 도메인으로서 사용될 수 있는 뉴클레아제-불활성 Cpf1 (dCpf1) 변이체이다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사하지만 HNH 엔도뉴클레아제 도메인을 갖지 않는 RuvC-유사 엔도뉴클레아제를 갖고, Cpf1의 N-말단은 Cas9의 알파-나선 인지 돌출부를 갖지 않는다. 문헌 (참조: Zetsche *et al.*, *Cell*, 163, 759-771, 2015 (이는 본원에 참조로 포함됨))에서는 Cpf1의 RuvC-유사 도메인이 DNA 가닥 둘 다를 절단하는데 관여하고 RucC-유사 도메인의 불활성화가 Cpf1 뉴클레아제 활성을 불활성화시킴을 보여주었다. 예를 들어, 프란시셀라 노비시다 (*Francisella novicida*) Cpf1에서 D917A, E1006A, 또는 D1255A에 상응하는 돌연변이는 Cpf1 뉴클레아제 활성을 불활성화시킨다. 일부 구현예에서, 본원의 개시내용의 dCpf1은 D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, 또는 D917A/E1006A/D1255A에 상응하는 돌연변이를 포함한다. Cpf1의 RuvC 도메인을 불활성화시키는 임의의 돌연변이, 예를 들어, 치환 돌연변이, 결실 또는 삽입이 본원 개시내용에 따라 사용될 수 있는 것으로 이해되어야만 한다.

[0409] 일부 구현예에서, 본원에 제공된 임의의 융합 단백질의 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)은

Cpf1 단백질일 수 있다. 일부 구현예에서, Cpf1 단백질은 Cpf1 닉카제 (nCpf1)이다. 일부 구현예에서, Cpf1 단백질은 뉴클레아제 불활성 Cpf1 (dCpf1)이다. 일부 구현예에서, Cpf1, nCpf1 또는 dCpf1은 본원에 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 구현예에서, dCpf1은 본원에 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 포함하고, D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, 또는 D917A/E1006A/D1255A에 상응하는 돌연변이를 포함한다. 다른 세균 종 기원의 Cpf1이 또한 본원 개시내용에 따라 사용될 수 있음을 인지해야 한다.

[0410] 야생형 프란시셀라 노비시다 Cpf1 (D917, E1006, 및 D1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEI
LSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDA
KKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDI
PTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSL
DEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLF
KQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSK
IYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLET
IKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAE
DDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY
ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENK
GEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIE
DCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKIN
NIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF
KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP
TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVAD
[0411] VNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0412] 프란시셀라 노비시다 Cpf1 D917A (A917, E1006, 및 D1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEI
LSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDA
KKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDI
PTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSL
DEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLF
KQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSK
IYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLET
IKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAE
DDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY
ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENK
GEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIE
DCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKIN
NIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF
KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP
TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVAD
VNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0413]

[0414] 프란시셀라 노비시다 Cpf1 E1006A (D917, A1006, 및 D1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEI
LSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDA
KKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDI
PTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSL
DEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLF
KQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSK
IYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLET
IKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAE
DDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY
ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENK
GEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIE
DCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKIN
NIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF
KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP
TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVAD
VNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0415]

[0416] 프란시셀라 노비시다 Cpf1 D1255A (D917, E1006, 및 A1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEI
LSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDA
KKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDI
PTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSL
DEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLF
KQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSK
IYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLET
IKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAE
DDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY
ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENK
GEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIE
DCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKIN
NIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF
KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP
TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVAD
VNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0417]

[0418] 프란시셀라 노비시다 Cpf1 D917A/E1006A (A917, A1006, 및 D1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEI
LSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDA
KKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDI
PTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSL
DEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLF
KQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSK
IYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLET
IKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAE
DDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY
ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENK
GEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIE
DCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKIN
NIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF
KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP
TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVAD
VNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0419]

[0420] 프란시셀라 노비시다 Cpf1 D917A/D1255A (A917, E1006, 및 A1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEI
LSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDA
KKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDI
PTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSL
DEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLF
KQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSK
IYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLET
IKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAE
DDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY
ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENK
GEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIE
DCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKIN
NIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF
KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP
TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVAD
VNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0421]

[0422] 프란시셀라 노비시다 Cpf1 E1006A/D1255A (D917, A1006, 및 A1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEI
LSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDA
KKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDI
PTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSL
DEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLF
KQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSK
IYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLET
IKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAE
DDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY
ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENK
GEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIE
DCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKIN
NIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF
KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP
TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVAD
VNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0423]

[0424] 프란시셀라 노비시다 Cpf1 D917A/E1006A/D1255A (A917, A1006, 및 A1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEI
LSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDA
KKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDI
PTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSL
DEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLF
KQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSK
IYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLET
IKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAE
DDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNY
ITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENK
GEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIE
DCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA
KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKIN
NIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF
KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYP
TKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVAD
VNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0425]

[0426] 일부 구현예에서, 융합 단백질에 존재하는 Cas9 도메인의 하나는 PAM 서열이 요구되지 않는 가이드 뉴클레오타이드 서열-프로그래밍 가능한 DNA-결합 단백질 도메인으로 대체될 수 있다.

[0427] 일부 구현예에서, 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)은 미생물 CRISPR-Cas 시스템의 단일 이펙터이다. 미생물 CRISPR-Cas 시스템의 단일 이펙터는 제한 없이 Cas9, Cpf1, Cas12b/C2c1, 및 Cas12c/C2c3을 포함한다. 전형적으로, 미생물 CRISPR-Cas 시스템은 부류 1 및 부류 2 시스템으로 분류된다. 부류 1 시스템은 멀티서브유닛 이펙터 복합체를 갖고, 부류 2 시스템은 단일 단백질 이펙터를 갖는다. 예를 들어, Cas9 및 Cpf1은 부류 2 이펙터이다. Cas9 및 Cpf1에 추가로, 3개의 별개의 부류 2 CRISPR-Cas 시스템 (Cas12b/C2c1 및 Cas12c/C2c3)은 문헌 (참조: Shmakov *et al.,* "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", *Mol. Cell,* 2015 Nov. 5; 60(3): 385-397, 이의 전체 내용은 본원에 참조로 포함됨)에 기재되어 있다. 시스템의 2개의 이펙터, Cas12b/C2c1 및 Cas12c/C2c3은 Cpf1과 관련된 RuvC-유사 엔도뉴클레아제 도메인을 함유한다. 제3 시스템은 2개의 예측된 HEPN RNase 도메인을 갖는 이펙터를 포함한다. 성숙한 CRISPR RNA의 생성은 Cas12b/C2c1에 의한 CRISPR의 생성과 달리 tracrRNA 독립적이다. Cas12b/C2c1은 DNA 절단을 위해 CRISPR RNA 및 tracrRNA 둘 다에 의존한다.

[0428] 알리사이클로바실러스 액시도테라스트리스 Cas12b/C2c1 (AacC2c1)의 결정 구조는 키메라 단일 분자 가이드 RNA (sgRNA)와의 복합체로 보고되었다. 문헌 (예를 들어, Liu *et al.,* "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell,* 2017 Jan. 19; 65(2):310-322)을 참조하고 이의 전체 내용은 본원에 참조로 포함된다. 결정 구조는 3원 복합체로서 표적 DNA에 결합된 알리사이클로바실러스 액시도테레스트리스 (*Alicyclobacillus acidoterrestris*) C2c1에서도 보고되었다. 예를 들어, 문헌 (예를 들어, Yang *et al.*, "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell,* 2016 Dec. 15; 167(7):1814-1828)을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다. 표적 및 비-표적 DNA 가닥 둘 다와 함께 AacC2c1의 촉매 적격의 형태는 단일 RuvC 촉매 포켓 내에 독립적으로 위치하는 것으로 캡쳐되었고, Cas12b/C2c1-매개된 절단은 표적 DNA의 엇갈린 7개 뉴클레오타이드 절단을 유도한다. Cas12b/C2c1 3원 복합체와 이전에 동정된 Cas9 및 Cpf1 대응물 간의 구조적 비교는 CRISPR-Cas9 시스템에 의해 사용되는 기전의 다양성을 입증한다. 일부 구현예에서, 본원에 제공된 임의의 융합 단백질의 핵산 프로그램 가능한 DNA 결합 단백질 (napDNAbp)은 Cas12b/C2c1 또는 Cas12c/C2c3 단백질일 수 있다. 일부 구현예에서, napDNAbp는 Cas12b/C2c1 단백질이다. 일부 구현예에서, napDNAbp는 Cas12c/C2c3 단백질이다. 일부 구현예에서, napDNAbp는 천연적으로 발생하는 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 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, napDNAbp는 천연적으로 발생하는 Cas12b/C2c1 또는 Cas12c/C2c3 단백질이다. 일부 구현예에서, napDNAbp는 본원에 제공된 napDNAbp 서열 중 임의의 하나와 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종 기원의 Cas12b/C2c1 또는 Cas12c/C2c3이 또한 본원 개시내용에 따라 사용될 수 있음을 인지해야 한다.

[0429] Cas12b/C2c1 ((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR-연합된 엔도뉴클레아제 C2c1 OS = 알리사이클로바실러스 액시도테레스트리스 (균주 ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1) 아미노산 서열은 다음과 같다:

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTA
EECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLAD
KDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYT
DSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQ
EHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNV
QRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPI
WTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDP
NEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRP
PYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVA
RKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLA
YLRLLVRCGSEDVGRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAV
YESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKV
SGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLI
LLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGI
RCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIH
ADLNAAQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYE
RERGKKRRKVFAQEKLSEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRI
EGYLVKQIRSRVPLQDSACENTGDI

[0430]

[0431] BhCas12b (바실러스 히사시) NCBI 참조 서열: WP_095142515

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEH
HEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEAN
QLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLA
EYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKV
EKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDEN
EPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQ
ATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGK
VDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGN
VGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLR
VMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRK
AREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKP
YKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEP
GEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFED
LSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSV
VTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNL
QKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIK
KGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISK
LTNQYSISTIEDDSSKQSMKRPAATKKAGQAKKKK

[0432]

[0433] 일부 구현예에서, Cas12b는 BvCas12b V4로, BhCas12b의 변이체이며 BhCas12B에 대해 다음과 같은 변화를 포함한다: 893R, K846R, 및 E837G. BhCas12b(V4)는 다음과 같이 표현된다: 5' mRNA Cap---5' UTR---bhCas12b---STOP 서열 --- 3' UTR --- 120폴리A 테일

[0434] 5'UTR:

[0435] GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACC

[0436] 3' UTR (TriLink 표준 UTR)

GCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCT

[0437] GCACCCGTACCCCCGTGGTCTTTGAATAAAGTCTGA

[0438] bhCas12b (V4)의 핵산 서열

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAG
ATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGG
TGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTAC
GAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCT
GTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGG
TGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAA
GCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGG
AACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGG
AAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAG
CTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAA
AGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCA
TTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG
AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGC
TCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGT
ACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGAC
GAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGA
GGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATC
ACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAG
CAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAG
AAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGA
AAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAG
GGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGA
[0439] GGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACAC

TCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGC
GGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTC
CAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGA
CCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGC
CTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGT
GGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCG
TGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTG
CGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCT
GCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGAC
AAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTAC
AAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGG
CAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCC
TGAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACC
GAACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCT
GAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCT
ACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTC
GAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAA
GTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGG
GAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGT
AGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAG
ACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGA
AGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAG
AACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCA
GGTGGACGGCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGT
TCGGCGAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAA
ATCAAGAAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAG
CTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATG
TGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATC
AGCAAGCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAA
AAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG

[0440]

[0441] 일부 구현예에서, Cas12b는 BvCas12B이다. 일부 구현예에서, Cas12b는 아래에 제공된 예시적인 BvCas12b 서열에 넘버링된 바와 같이 아미노산 치환 S893R, K846R, 및 E837G를 포함한다.

[0442] BvCas12b (바실러스 종 V3-13) NCBI 참조 서열: WP_101661451.1

MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAYQAELIN
IIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGT
SNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQKDIEWL
PLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQLKEKTESYYKEHLTGGEEWIEKIRKF
EKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLPESASPEELWKVVAEQQNKMSEGFGD
PKVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTLPDAIEHPLWIRYESPGGT
NLNLFKLEEKQKKNYYVTLSKIIWPSEEKWIEKENIEIPLAPSIQFNRQIKLKQHVKGKQEISFS
DYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIGKA
LKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASVSIFEVVKEL
PKDQEQKLFYSINDTELFAIHKRSFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRL
ETKKTPDERKKAIHKLMEIVQSYDSWTASQKEVWEKELNLLTNMAAFNDEIWKESLVELHHRIEP
YVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRLLISWSKRSRTPGEANRIETDEPFGSSLL
QHIQNVKDDRLKQMANLIIMTALGFKYDKEEKDRYKRWKETYPACQIILFENLNRYLFNLDRSRR
ENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTL
KRLIEDGFINESELAYLKKGDIIPSQGGELFVTLSKRYKKDSDNNELTVIHADINAAQNLQKRFW
QQNSEVYRVPCQLARMGEDKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTT
FDLQDLDGFEDISKTIELAQEQQKKYLTMFRDPSGYFFNNETWRPQKEYWSIVNNIIKSCLKKKI
LSNKVEL.

[0443]

[0444] 일부 구현예에서, Cas12b는 BTCas12b.BTCas12b 바실러스 써모아밀로보란스 (*Bacillus thermoamylovorans*) NCBI 참조 서열: WP_041902512이다.

MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPKNPKKV
SKAEIQAELWDFVLKMQKCNSFTHEVDKDVVFNILRELYEELVPSSVEKKGEANQLSNKF
LYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLAE
YGLIPLFIPFTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEE
YEKVEKEHKTLEERIKEDIQAFKSLEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREII
QKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYAT
FCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTV
QLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGT
LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKFVNF
KPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLF
FPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFE
DITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGK
EVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQ
LNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERS
RFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL
QDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKLVTTHADINAAQNLQ
KRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWGNAGK
LKIKKGSSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFG
KLERILISKLTNQYSISTIEDDSSKQSM

[0445]

[0446] 일부 구현예에서, napDNAbp는 Cas12c를 언급한다. 일부 구현예에서, Cas12c 단백질은 Cas12c1 또는 Cas12c1의 변이체이다. 일부 구현예에서, Cas12 단백질은 Cas12c2 또는 Cas12c2의 변이체이다. 일부 구현예에서, Cas12 단백질은 올레이필러스 속 (Oleiphilus sp.) HI0009 (즉, OspCas12c) 기원의 Cas12c 단백질 또는 OspCas12c의 변이체이다. 이들 Cas12c 분자는 문헌 (참조: Yan et al., "Functionally Diverse Type V CRISPR-Cas

Systems," Science, 2019 Jan. 4; 363: 88-91; 이의 전체 내용은 본원에 참조로 포함됨)에 기재되어 있다. 일부 구현예에서, napDNAbp는 천연적으로 발생하는 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 구현예에서, napDNAbp는 천연적으로 발생하는 Cas12c1, Cas12c2, 또는 OspCas12c 단백질이다. 일부 구현예에서, napDNAbp는 본원에 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 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종 기원의 Cas12c1, Cas12c2, 또는 OspCas12c가 또한 본원의 개시내용에 따라 사용될 수 있음을 인지해야 한다.

[0447] Cas12c1

MQTKKTHLHLISAKASRKYRRTIACLSDTAKKDLERRKQSGAADPAQELSCLKTIKFKLEVPEGS
KLPSFDRISQIYNALETIEKGSLSYLLFALILSGFRIFPNSSAAKTFASSSCYKNDQFASQIKEI
FGEMVKNFIPSELESILKKGRRKNNKDWTEENIKRVLNSEFGRKNSEGSSALFDSFLSKFSQELF
RKFDSWNEVNKKYLEAAELLDSMLASYGPFDSVCKMIGDSDSRNSLPDKSTIAFTNNAEITVDIE
SSVMPYMAIAALLREYRQSKSKAAPVAYVQSHLTTTNGNGLSWFFKFGLDLIRKAPVSSKQSTSD
GSKSLQELFSVPDDKLDGLKFIKEACEALPEASLLCGEKGELLGYQDFRTSFAGHIDSWVANYVN
RLFELIELVNQLPESIKLPSILTQKNHNLVASLGLQEAEVSHSLELFEGLVKNVRQTLKKLAGID
ISSSPNEQDIKEFYAFSDVLNRLGSIRNQIENAVQTAKKDKIDLESAIEWKEWKKLKKLPKLNGL
GGGVPKQQELLDKALESVKQIRHYQRIDFERVIQWAVNEHCLETVPKFLVDAEKKKINKESSTDF
AAKENAVRFLLEGIGAAARGKTDSVSKAAYNWFVVNNFLAKKDLNRYFINCQGCIYKPPYSKRRS
LAFALRSDNKDTIEVVWEKFETFYKEISKEIEKFNIFSQEFQTFLHLENLRMKLLLRRIQKPIPA
EIAFFSLPQEYYDSLPPNVAFLALNQEITPSEYITQFNLYSSFLNGNLILLRRSRSYLRAKFSWV
GNSKLIYAAKEARLWKIPNAYWKSDEWKMILDSNVLVFDKAGNVLPAPTLKKVCEREGDLRLFYP
LLRQLPHDWCYRNPFVKSVGREKNVIEVNKEGEPKVASALPGSLFRLIGPAPFKSLLDDCFFNPL
DKDLRECMLIVDQEISQKVEAQKVEASLESCTYSIAVPIRYHLEEPKVSNQFENVLAIDQGEAGL
AYAVFSLKSIGEAETKPIAVGTIRIPSIRRLIHSVSTYRKKKQRLQNFKQNYDSTAFIMRENVTG
DVCAKIVGLMKEFNAFPVLEYDVKNLESGSRQLSAVYKAVNSHFLYFKEPGRDALRKQLWYGGDS
WTIDGIEIVTRERKEDGKEGVEKIVPLKVFPGRSVSARFTSKTCSCCGRNVFDWLFTEKKAKTNK
KFNVNSKGELTTADGVIQLFEADRSKGPKFYARRKERTPLTKPIAKGSYSLEEIERRVRTNLRRA
PKSKQSRDTSQSQYFCVYKDCALHFSGMQADENAAINIGRRFLTALRKNRRSDFPSNVKISDRLL
DN

[0448]

[0449] Cas12c2

MTKHSIPLHAFRNSGADARKWKGRIALLAKRGKETMRTLQFPLEMSEPEAAAINTTPFAVAYNAI
EGTGKGTLFDYWAKLHLAGFRFFPSGGAATIFRQQAVFEDASWNAAFCQQSGKDWPWLVPSKLYE
RFTKAPREVAKKDGSKKSIEFTQENVANESHVSLVGASITDKTPEDQKEFFLKMAGALAEKFDSW
KSANEDRIVAMKVIDEFLKSEGLHLPSLENIAVKCSVETKPDNATVAWHDAPMSGVQNLAIGVFA
TCASRIDNIYDLNGGKLSKLIQESATTPNVTALSWLFGKGLEYFRTTDIDTIMQDFNIPASAKES
IKPLVESAQAIPTMTVLGKKNYAPFRPNFGGKIDSWIANYASRLMLLNDILEQIEPGFELPQALL
DNETLMSGIDMTGDELKELIEAVYAWVDAAKQGLATLLGRGGNVDDAVQTFEQFSAMMDTLNGTL
NTISARYVRAVEMAGKDEARLEKLIECKFDIPKWCKSVPKLVGISGGLPKVEEEIKVMNAAFKDV
RARMFVRFEEIAAYVASKGAGMDVYDALEKRELEQIKKLKSAVPERAHIQAYRAVLHRIGRAVQN
CSEKTKQLFSSKVIEMGVFKNPSHLNNFIFNQKGAIYRSPFDRSRHAPYQLHADKLLKNDWLELL
AEISATLMASESTEQMEDALRLERTRLQLQLSGLPDWEYPASLAKPDIEVEIQTALKMQLAKDTV
TSDVLQRAFNLYSSVLSGLTFKLLRRSFSLKMRFSVADTTQLIYVPKVCDWAIPKQYLQAEGEIG
IAARVVTESSPAKMVTEVEMKEPKALGHFMQQAPHDWYFDASLGGTQVAGRIVEKGKEVGKERKL
VGYRMRGNSAYKTVLDKSLVGNTELSQCSMIIEIPYTQTVDADFRAQVQAGLPKVSINLPVKETI
TASNKDEQMLFDRFVAIDLGERGLGYAVFDAKTLELQESGHRPIKAITNLLNRTHHYEQRPNQRQ
KFQAKFNVNLSELRENTVGDVCHQINRICAYYNAFPVLEYMVPDRLDKQLKSVYESVTNRYIWSS
TDAHKSARVQFWLGGETWEHPYLKSAKDKKPLVLSPGRGASGKGTSQTCSCCGRNPFDLIKDMKP
RAKIAVVDGKAKLENSELKLFERNLESKDDMLARRHRNERAGMEQPLTPGNYTVDEIKALLRANL

[0450] RRAPKNRRTKDTTVSEYHCVFSDCGKTMHADENAAVNIGGKFIADIEK

[0451] OspCas12c

MTKLRHRQKKLTHDWAGSKKREVLGSNGKLQNPLLMPVKKGQVTEFRKAFSAYARATKGEMTDGR
KNMFTHSFEPFKTKPSLHQCELADKAYQSLHSYLPGSLAHFLLSAHALGFRIFSKSGEATAFQAS
SKIEAYESKLASELACVDLSIQNLTISTLFNALTTSVRGKGEETSADPLIARFYTLLTGKPLSRD
TQGPERDLAEVISRKIASSFGTWKEMTANPLQSLQFFEEELHALDANVSLSPAFDVLIKMNDLQG
DLKNRTIVFDPDAPVFEYNAEDPADIIIKLTARYAKEAVIKNQNVGNYVKNAITTTNANGLGWLL
NKGLSLLPVSTDDELLEFIGVERSHPSCHALIELIAQLEAPELFEKNVFSDTRSEVQGMIDSAVS
NHIARLSSSRNSLSMDSEELERLIKSFQIHTPHCSLFIGAQSLSQQLESLPEALQSGVNSADILL
GSTQYMLTNSLVEESIATYQRTLNRINYLSGVAGQINGAIKRKAIDGEKIHLPAAWSELISLPFI
GQPVIDVESDLAHLKNQYQTLSNEFDTLISALQKNFDLNFNKALLNRTQHFEAMCRSTKKNALSK
PEIVSYRDLLARLTSCLYRGSLVLRRAGIEVLKKHKIFESNSELREHVHERKHFVFVSPLDRKAK
KLLRLTDSRPDLLHVIDEILQHDNLENKDRESLWLVRSGYLLAGLPDQLSSSFINLPIITQKGDR
RLIDLIQYDQINRDAFVMLVTSAFKSNLSGLQYRANKQSFVVTRTLSPYLGSKLVYVPKDKDWLV
PSQMFEGRFADILQSDYMVWKDAGRLCVIDTAKHLSNIKKSVFSSEEVLAFLRELPHRTFIQTEV
RGLGVNVDGIAFNNGDIPSLKTFSNCVQVKVSRTNTSLVQTLNRWFEGGKVSPPSIQFERAYYKK
DDQIHEDAAKRKIRFQMPATELVHASDDAGWTPSYLLGIDPGEYGMGLSLVSINNGEVLDSGFIH
INSLINFASKKSNHQTKVVPRQQYKSPYANYLEQSKDSAAGDIAHILDRLIYKLNALPVFEALSG
NSQSAADQVWTKVLSFYTWGDNDAQNSIRKQHWFGASHWDIKGMLRQPPTEKKPKPYIAFPGSQV
SSYGNSQRCSCCGRNPIEQLREMAKDTSIKELKIRNSEIQLFDGTIKLFNPDPSTVIERRRHNLG
PSRIPVADRTFKNISPSSLEFKELITIVSRSIRHSPEFIAKKRGIGSEYFCAYSDCNSSLNSEAN

[0452] AAANVAQKFQKQLFFEL

[0453] 일부 구현예에서, napDNAbp는 Cas12g, Cas12h, 또는 Cas12i를 언급하고 이는, 예를 들어, 문헌 (참조: Yan et al., "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91; 각각의 전체 내용은 본원에 참조로 포함됨)에 기재되어 있다. 10 테라바이트 초과의 서열 데이터를 집계하여, Cas12g, Cas12h, 및 Cas12i를 포함하여 이전에 특성화된 부류 V 단백질과 약한 유사성을 나타내는 V형 Cas 단백질의 새로운 분류가 동정되었다. 일부 구현예에서, Cas12 단백질은 Cas12g 또는 Cas12g의 변이체이다. 일부 구현예에

서, Cas12 단백질은 Cas12h 또는 Cas12h의 변이체이다. 일부 구현예에서, Cas12 단백질은 Cas12i 또는 Cas12i의 변이체이다. 다른 RNA-가이드된 DNA 결합 단백질이 napDNAbp로서 사용될 수 있고, 본원 개시내용의 범위 내에 있음을 인지해야 한다. 일부 구현예에서, napDNAbp는 천연적으로 발생하는 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 구현예에서, napDNAbp는 천연적으로 발생하는 Cas12g, Cas12h, 또는 Cas12i 단백질이다. 일부 구현예에서, napDNAbp는 본원에 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 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종 기원의 Cas12g, Cas12h, 또는 Cas12i가 또한 본원의 개시내용에 따라 사용될 수 있음을 인지해야 한다. 일부 구현예에서, Cas12i는 Cas12i1 또는 Cas12i2이다.

[0454] Cas12g1

MAQASSTPAVSPRPRPRYREERTLVRKLLPRPGQSKQEFRENVKKLRKAFLQFNADVSGVCQWAI
QFRPRYGKPAEPTETFWKFFLEPETSLPPNDSRSPEFRRLQAFEAAAGINGAAALDDPAFTNELR
DSILAVASRPKTKEAQRLFSRLKDYQPAHRMILAKVAAEWIESRYRRAHQNWERNYEEWKKEKQE
WEQNHPELTPEIREAFNQIFQQLEVKEKRVRICPAARLLQNKDNCQYAGKNKHSVLCNQFNEFKK
NHLQGKAIKFFYKDAEKYLRCGLQSLKPNVQGPFREDWNKYLRYMNLKEETLRGKNGGRLPHCKN
LGQECEFNPHTALCKQYQQQLSSRPDLVQHDELYRKWRREYWREPRKPVFRYPSVKRHSIAKIFG
ENYFQADFKNSVVGLRLDSMPAGQYLEFAFAPWPRNYRPQPGETEISSVHLHFVGTRPRIGFRFR
VPHKRSRFDCTQEELDELRSRTFPRKAQDQKFLEAARKRLLETFPGNAEQELRLLAVDLGTDSAR
AAFFIGKTFQQAFPLKIVKIEKLYEQWPNQKQAGDRRDASSKQPRPGLSRDHVGRHLQKMRAQAS
EIAQKRQELTGTPAPETTTDQAAKKATLQPFDLRGLTVHTARMIRDWARLNARQIIQLAEENQVD
LIVLESLRGFRPPGYENLDQEKKRRVAFFAHGRIRRKVTEKAVERGMRVVTVPYLASSKVCAECR
KKQKDNKQWEKNKKRGLFKCEGCGSQAQVDENAARVLGRVFWGEIELPTAIP

[0455]

[0456] Cas12h1

MKVHEIPRSQLLKIKQYEGSFVEWYRDLQEDRKKFASLLFRWAAFGYAAREDDGATYISPSQALL
ERRLLLGDAEDVAIKFLDVLFKGGAPSSSCYSLFYEDFALRDKAKYSGAKREFIEGLATMPLDKI
IERIRQDEQLSKIPAEEWLILGAEYSPEEIWEQVAPRIVNVDRSLGKQLRERLGIKCRRPHDAGY
CKILMEVVARQLRSHNETYHEYLNQTHEMKTKVANNLTNEFDLVCEFAEVLEEKNYGLGWYVLWQ
GVKQALKEQKKPTKIQIAVDQLRQPKFAGLLTAKWRALKGAYDTWKLKKRLEKRKAFPYMPNWDN
DYQIPVGLTGLGVFTLEVKRTEVVVDLKEHGKLFCSHSHYFGDLTAEKHPSRYHLKFRHKLKLRK
RDSRVEPTIGPWIEAALREITIQKKPNGVFYLGLPYALSHGIDNFQIAKRFFSAAKPDKEVINGL
PSEMVVGAADLNLSNIVAPVKARIGKGLEGPLHALDYGYGELIDGPKILTPDGPRCGELISLKRD
IVEIKSAIKEFKACQREGLTMSEETTTWLSEVESPSDSPRCMIQSRIADTSRRLNSFKYQMNKEG
YQDLAEALRLLDAMDSYNSLLESYQRMHLSPGEQSPKEAKFDTKRASFRDLLRRRVAHTIVEYFD
DCDIVFFEDLDGPSDSDSRNNALVKLLSPRTLLLYIRQALEKRGIGMVEVAKDGTSQNNPISGHV
GWRNKQNKSEIYFYEDKELLVMDADEVGAMNILCRGLNHSVCPYSFVTKAPEKKNDEKKEGDYGK
RVKRFLKDRYGSSNVRFLVASMGFVTVTTKRPKDALVGKRLYYHGGELVTHDLHNRMKDEIKYLV
EKEVLARRVSLSDSTIKSYKSFAHV

[0457]

[0458] Cas12i1

MSNKEKNASETRKAYTTKMIPRSHDRMKLLGNFMDYLMDGTPIFFELWNQFGGGIDRDIISGTAN
KDKISDDLLLAVNWFKVMPINSKPQGVSPSNLANLFQQYSGSEPDIQAQEYFASNFDTEKHQWKD
MRVEYERLLAELQLSRSDMHHDLKLMYKEKCIGLSLSTAHYITSVMFGTGAKNNRQTKHQFYSKV
IQLLEESTQINSVEQLASIILKAGDCDSYRKLRIRCSRKGATPSILKIVQDYELGTNHDDEVNVP
SLIANLKEKLGRFEYECEWKCMEKIKAFLASKVGPYYLGSYSAMLENALSPIKGMTTKNCKFVLK
QIDAKNDIKYENEPFGKIVEGFFDSPYFESDTNVKWVLHPHHIGESNIKTLWEDLNAIHSKYEED
IASLSEDKKEKRIKVYQGDVCQTINTYCEEVGKEAKTPLVQLLRYLYSRKDDIAVDKIIDGITFL
SKKHKVEKQKINPVIQKYPSFNFGNNSKLLGKIISPKDKLKHNLKCNRNQVDNYIWIEIKVLNTK
TMRWEKHHYALSSTRFLEEVYYPATSENPPDALAARFRTKTNGYEGKPALSAEQIEQIRSAPVGL
RKVKKRQMRLEAARQQNLLPRYTWGKDFNINICKRGNNFEVTLATKVKKKKEKNYKVVLGYDANI
VRKNTYAAIEAHANGDGVIDYNDLPVKPIESGFVTVESQVRDKSYDQLSYNGVKLLYCKPHVESR
RSFLEKYRNGTMKDNRGNNIQIDFMKDFEAIADDETSLYYFNMKYCKLLQSSIRNHSSQAKEYRE
EIFELLRDGKLSVLKLSSLSNLSFVMFKVAKSLIGTYFGHLLKKPKNSKSDVKAPPITDEDKQKA
DPEMFALRLALEEKRLNKVKSKKEVIANKIVAKALELRDKYGPVLIKGENISDTTKKGKKSSTNS
FLMDWLARGVANKVKEMVMMHQGLEFVEVNPNFTSHQDPFVHKNPENTFRARYSRCTPSELTEKN
RKEILSFLSDKPSKRPTNAYYNEGAMAFLATYGLKKNDVLGVSLEKFKQIMANILHQRSEDQLLF
PSRGGMFYLATYKLDADATSVNWNGKQFWVCNADLVAAYNVGLVDIQKDFKKK

[0459]

[0460] Cas12i2

MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFFKMLQGLFGGITPEIVRFSTEQEKQQQDI
ALWCAVNWFRPVSQDSLTHTIASDNLVEKFEEYYGGTASDAIKQYFSASIGESYYWNDCRQQYYD
LCRELGVEVSDLTHDLEILCREKCLAVATESNQNNSIISVLFGTGEKEDRSVKLRITKKILEAIS
NLKEIPKNVAPIQEIILNVAKATKETFRQVYAGNLGAPSTLEKFIAKDGQKEFDLKKLQTDLKKV
IRGKSKERDWCCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKEIQ
SLNNLLVVKKLNDFFDSEFFSGEETYTICVHHLGGKDLSKLYKAWEDDPADPENAIVVLCDDLKN
NFKKEPIRNILRYIFTIRQECSAQDILAAAKYNQQLDRYKSQKANPSVLGNQGFTWTNAVILPEK
AQRNDRPNSLDLRIWLYLKLRHPDGRWKKHHIPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHL
PKLTDQTAIRVNKKHVKAAKTEARIRLAIQQGTLPVSNLKITEISATINSKGQVRIPVKFDVGRQ
KGTLQIGDRFCGYDQNQTASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQLS
YEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDAICKYQPRLYKFNKEYAYLLR
DIVRGKSLVELQQIRQEIFRFIEQDCGVTRLGSLSLSTLETVKAVKGIIYSYFSTALNASKNNPI
SDEQRKEFDPELFALLEKLELIRTRKKKQKVERIANSLIQTCLENNIKFIRGEGDLSTTNNATKK
KANSRSMDWLARGVFNKIRQLAPMHNITLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIG
DWVLRKLSQNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEELLKWRSDRKSNIPCWVLQNRLAEK
LGNKEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAANIALTVKGIGEQSSDE
ENPDGSRIKLQLTS

[0461]

[0462] 염기 편집기의 대표적인 핵산 및 단백질 서열은 다음과 같다:

[0463] P153에서 BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAG
ATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGG
TGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTAC
GAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCT
GTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGG
TGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAA
GCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGG
AACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCggaggct
ctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCA
AAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAAT
CGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCC
TTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTTT
GAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGT
TCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACC
ACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTT
CGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGG
ATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCTCCTGGGAAGAAGAGA
AGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAG
TACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAA
GTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATTCAGGCCC
TGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAGAAGGTCGAG
AAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGCTCTGGAACA
GTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGA
GCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAG
CCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGA
TTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGT
ACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCC

[0464] ACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAG

CAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGA
CAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTG
GACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGG
CAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAG
CCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTG
GGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCT
GAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGA
TCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTG
ATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAA
GCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAG
CCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCC
AGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCA
GCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACA
GCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTAC
AAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGT
GAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACA
TCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGC
GAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCT
GAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACG
ACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTG
AGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAG
ACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGG
GCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTG
ACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCT
GGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCA
GCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAG
AAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACG
GCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGC
GAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAA
GAAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCG
ACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTC
CCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAA
GCTGACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGC

[0465]

CGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCA*
*GATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TA
A

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEH
HEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEAN
QLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPGGSGGSSEVEFSHEYWMRHALTLAKR
ARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEP
CVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRM
PRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSWEEEKKKWEEDKKKDPLAKILGKLAEYG
LIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKE
YKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPS
EKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATF
TLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDI
VLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGR
IYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMS
IDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKARE
DNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKD
WVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEV
RRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSN
YNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTK
EKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKR
FWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGS
SKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTN
QYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0466]

[0467] K255에서 BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAG
ATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGG
TGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTAC
GAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCT
GTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGG
TGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAA
GCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGG
AACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGG

[0468]

AAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAG
CTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAA
AGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCA
TTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG
AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAggaggctctggaggaagcTCCGA
AGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATG
AACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAAT
AGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCT
TGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGT
GCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACG
GGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAACCACCGGGTAGAAATCAC
AGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTTTCGCATGCCCAGGCGGG
TCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTGGATCTGAAACACCTGGC
ACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGAGGACATCCAGGCTCTGAAGGCTCTGGAACA
GTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGTACCGGCTGA
GCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAG
CCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGA
TTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGT
ACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCC
ACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAG
CAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGA
CAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTG
GACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGG
CAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAG
CCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTG
GGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCT
GAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGA
TCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTG
ATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAA
GCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAG
CCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCC
AGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCA
GCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACA
GCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTAC

[0469]

AAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGT
GAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACA
TCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGC
GAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCT
GAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACG
ACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTG
AGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAG
ACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGG
GCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTG
ACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCT
GGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCA
GCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAG
AAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGG
CCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGG
GCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAG
GGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCT
GGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCA
GCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGC
GGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATT*
*ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEH
HEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEAN
QLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLA
EYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKV
EKEYKTLEERIKGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRA
IGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTS
ESATPESSGEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPS
EKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATF
TLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDI
VLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGR
IYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMS

[0470]

IDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKARE
DNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKD
WVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEV
RRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSN
YNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTK
EKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKR
FWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGS
SKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTN
QYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0471]

[0472] D306에서 BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAG
ATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGG
TGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTAC
GAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCT
GTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGG
TGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAA
GCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGG
AACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGG
AAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAG
CTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAA
AGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCA
TTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG
AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGC
TCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGT
ACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGAC
ggaggctctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGAC
TCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATC
GCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATC
ATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGCTGTACGT
CACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTAT
TCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGC
ATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCG
[0473] TTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCT

CTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGAGAACGAG
CCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGA
TTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCACCCTGAGT
ACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAGCAGCAGGCC
ACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAGAAGCGGCAG
CAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGAAAAAGCTGA
CAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTG
GACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGG
CAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGCGGAG
CCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGCGGCAACGTG
GGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTCCAAGTCTCT
GAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGACCGAGTGGA
TCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTG
ATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAA
GCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGAG
CCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTGCGGAAGGCC
AGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCTGCACTTCCA
GCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGACAAGAGAACA
GCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTAC
AAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGT
GAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACA
TCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCGAACCTGGC
GAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCTGAACGCCCT
GAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCTACTGCTACG
ACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTG
AGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAG
ACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGG
GCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCGTCGTG
ACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAGACTGACCCT
GGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGAAGTTCATCA
GCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAGAACCTGCAG
AAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGG
CCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGG
[0474] GCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAG

GGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCT
GGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCA
GCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGC
GGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATT*
*ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEH
HEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEAN
QLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLA
EYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKV
EKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDGG
SGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMN
HRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGENEPS
EKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATF
TLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDI
VLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGNVGR
IYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMS
IDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKARE
DNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKD
WVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEV
RRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSN
YNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTK
EKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKR
FWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGS
SKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTN
QYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0475]

[0476] D980에서 BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAG
ATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGG
TGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTAC

[0477] GAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCT

GTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGG

TGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAA

GCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGG

AACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGG

AAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAG

CTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAA

AGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCA

TTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG

AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGC

TCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGT

ACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGAC

GAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGA

GGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATC

ACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAG

CAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAG

AAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGA

AAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAG

GGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGA

GGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACAC

TCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGC

GGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTC

CAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGA

CCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGC

CTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGT

GGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCG

TGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTG

CGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCT

GCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGAC

AAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTAC

AAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGG

CAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCC

TGAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACC

GAACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCT

[0478] GAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCT

ACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTC
GAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAA
GTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGG
GAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGT
AGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAG
ACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGA
AGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAG
AACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCA
GGTGGACggaggctctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACG
CATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTC
AACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGC
GGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCGACGC
TGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGA
GTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCA
TCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGT
TGTGTCGTTTTTTCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACT
GACGGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCGG
CCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGG
GCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAGAAG
GGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCT
GGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCA
GCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGC
GGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATT*
*ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEH
HEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEAN
QLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLA
EYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKV
EKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDEN
EPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQ
ATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGK
VDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGN

VGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLR
VMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRK
AREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKP
YKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEP
GEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFED
LSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSV
VTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNL
QKRFWTRTHGFYKVYCKAYQVDGGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
RVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDG
SSGSETPGTSESATPESSGGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGS
SKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTN
QYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0479]

[0480]

[0481] K1019에서 BhCas12b GGSGGS-ABE8-Xten20

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAG
ATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTGAAGAAAGGCCTCTGGAAAACCCACGAGG
TGCTGAACCACGGAATCGCCTACTACATGAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTAC
GAGCACCACGAGCAGGACCCCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCT
GTGGGATTTCGTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGG
TGTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGAAGGGCGAA
GCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGAAAGGG
AACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACCTGAAGATTGCCGGCGATCCCTCCTGGG
AAGAAGAGAAGAAGAAGTGGGAAGAAGATAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAG
CTGGCTGAGTACGGACTGATCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAA
AGAAATCAAGTGGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCA
TTCAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGAATACGAG
AAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGACATCCAGGCTCTGAAGGC
TCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCTGCGGGACACCCTGAACACCAACGAGT
ACCGGCTGAGCAAGAGAGGCCTTAGAGGCTGGCGGGAAATCATCCAGAAATGGCTGAAAATGGAC
GAGAACGAGCCCTCCGAGAAGTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGA
GGCCGGCGATTACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATC
ACCCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGACGCCAAG
[0482] CAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGTCCGATTCGAGGAAAG

AAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGCAGCTGCACACCGAGAAGCTGAAGA
AAAAGCTGACAGTGCAGCTGGACCGGCTGATCTACCCTACAGAATCTGGCGGCTGGGAAGAGAAG
GGCAAAGTGGACATTGTGCTGCTGCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGA
GGAAAAGGGCAAGCACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACAC
TCGGCGGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGAAAGC
GGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAGAGTCCCCAGTGTC
CAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCAACTTCAAGCCCAAAGAACTGA
CCGAGTGGATCAAGGACAGCAAGGGCAAGAAACTGAAGTCCGGCATCGAGTCCCTGGAAATCGGC
CTGAGAGTGATGAGCATCGACCTGGGACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGT
GGATCAGAAGCCCGACATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCG
TGCACAGAGCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCTG
CGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGCGGAACGTGCT
GCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTCACCAAGTGGATCAGCAGAC
AAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGAGCTGATCCAGATCCGCGAGCTGATGTAC
AAGCCTTACAAGGACTGGGTCGCCTTCCTGAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGG
CAAAGAAGTGAAGCACTGGCGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCC
TGAAGAACATCGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACC
GAACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTGAATCACCT
GAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCATCATGCACGCCCTGGGCT
ACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGAACCCCGCCTGCCAGATCATCCTGTTC
GAGGATCTGAGCAACTACAACCCCTACGAGGAAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAA
GTGGTCCAGACGCGAGATCCCCAGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGG
GAGAAGTGGGCGCTCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGT
AGCGTCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAGGGCAG
ACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGACAAAGGCGGCGAGA
AGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACACACGCCGACATCAACGCCGCTCAG
AACCTGCAGAAGCGGTTCTGGACAAGAACCCACGGCTTCTACAAGGTGTACTGCAAGGCCTACCA
GGTGGACGGCCAGACCGTGTACATCCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGT
TCGGCGAGGGCTACTTCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGggaggc
tctggaggaagcTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGC
AAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAA
TCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCC
CTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTTATGATGCCGACGCTGTACGTCACGTT
TGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTG

[0483]

TTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATCATCCAGGCATGAAC
CACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTCGTTTTTT
TCGCATGCCCAGGCGGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCTCTTCTG
GATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGCCTGAAAATCAAGAAG
GGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAGACAGCTTCGACCT
GGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGACCCCAGCGGCAATGTGTTCCCCA
GCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGAAAGCTGGAACGCATCCTGATCAGCAAGCTG
ACCAACCAGTACTCCATCAGCACCATCGAGGACGACAGCAGCAAGCAGTCTATGAAAAGGCCGGC
GGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCC*TACCCATACGATGTTCCAGATT*
*ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC*TAA

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEH
HEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVEKKGEAN
QLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKLA
EYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYEKV
EKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDEN
EPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQ
ATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGK
VDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRDHLRRYPHKVESGN
VGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIKDSKGKKLKSGIESLEIGLR
VMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRK
AREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKP
YKDWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEP
GEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFED
LSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSV
VTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNL
QKRFWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKGGSG
GSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHR
VEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGLKIKKGS
SKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTN
QYSISTIEDDSSKQSMKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA

[0484]

[0485] 상기 서열에 대해, Kozak 서열은 굵게 표시하고 밑줄 치고; ---------- 는 N-말단 핵 국소화 신호 (NLS)를 표시하고; 소문자는 GGGSGGS 링커를 지칭하고; ━ ━ ━ ━ 는 ABE8을 암호화하는 서열을 표시하고, 변형되지 않은 서열은 BhCas12b를 암호화하고; 두줄 밑줄은 Xten20 링커를 지칭하고; 한줄 밑줄은 C-말단 NLS를 지칭하고; GGATCC 는 GS 링커를 지칭하고; 이탤릭 문자는 3x 헤마글루티닌 (HA) 태그의 암호화 서열을 나타낸다.

[0486] 일부 구현예에서, Cas9 도메인은 스타필로코커스 아우레우스로부터 기원하는 Cas9 도메인 (SaCas9)이다. 일부 구현예에서, SaCas9 도메인은 뉴클레아제 활성 SaCas9, 뉴클레아제 불활성 SaCas9 (SaCas9d), 또는 SaCas9 닉카제 (SaCas9n)이다. 일부 구현예에서, SaCas9는 본원에 제공된 임의의 아미노산 서열에서 N579A 돌연변이, 또는 상응하는 돌연변이를 포함한다.

[0487] 일부 구현예에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 비-카노니칼 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 구현예에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 NNGRRT 또는 NNGRRT PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 구현예에서, SaCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 E781X, N967X, 및 R1014X 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함하고, 여기서, X는 임의의 아미노산이다. 일부 구현예에서, SaCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 E781K, N967K, 및 R1014H 돌연변이, 또는 하나 이상의 상응하는 돌연변이 중 하나 이상을 포함한다. 일부 구현예에서, SaCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 E781K, N967K, 또는 R1014H 돌연변이,

또는 상응하는 돌연변이를 포함한다.

[0488] 일부 구현예에서, 변이체 Cas 단백질은 SpCas9, SpCas9-VRQR, SpCas9-VRER, xCas9 (sp), SaCas9, SaCas9-KKH, SpCas9-MQKSER, SpCas9-LRKIQK, 또는 SpCas9-LRVSQL일 수 있다.

[0489] 예시적인 SaCas9 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE
LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS
FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALND
LNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNL
KVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTG
THNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQ
SIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLI
EKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**N**SKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRY
ATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFK
EWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR
ELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ
YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKING
ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEV
KSKKHPQIIKKG

[0490]

[0491] 밑줄 치고 굵게 표시한 상기 잔기 N579는 돌연변이시켜 (예를 들어, A579로) SaCas9 닉카제를 생성할 수 있다.

[0492] 예시적인 SaCas9n 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE
LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS
FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALND
LNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNL
KVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTG
THNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQ
SIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLI
EKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE**A**SKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRY
ATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFK
EWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR
ELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ
YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKING
ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEV
KSKKHPQIIKKG

[0493]

[0494] N579로부터 돌연변이되어 SaCas9 닉카제를 생성할 수 있는 상기 잔기 A579는 밑줄 치고 굵게 표시한다.

[0495] 예시적인 SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE
LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQS
FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALND
LNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNL
KVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTG
THNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQ
SIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLI
EKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEE<u>A</u>SKKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRY
ATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFK
EWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR
<u>*K*</u>LINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQ
YGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFY<u>*K*</u>NDLIKING
ELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP<u>*H*</u>IIKTIASKTQSIKKYSTDILGNLYEV
KSKKHPQIIKKG.

[0496]

[0497] N579로부터 돌연변이되어 SaCas9 닉카제를 생성할 수 있는 상기 잔기 A579는 밑줄 치고 굵게 표시한다. E781, N967, 및 R1014로부터 돌연변이되어 SaKKH Cas9를 생성할 수 있는 상기 잔기 K781, K967, 및 H1014는 밑줄 치고 이탤릭으로 나타낸다.

[0498] 염기 편집기의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 그 자체가 하나 이상의 도메인을 포함할 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 하나 이상의 뉴클레아제 도메인을 포함할 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 뉴클레아제 도메인은 엔도뉴클레아제 또는 엑소뉴클레아제를 포함할 수 있다. 본원에서 용어 "엑소뉴클레아제"는 자유 말단으로부터 핵산 (예를 들어, RNA 또는 DNA)을 분해할 수 있는 단백질 또는 폴리펩타이드를 언급하고, 용어 "엔도뉴클레아제"는 핵산 (예를 들어, DNA 또는 RNA) 내 내부 영역을 촉매 (예를 들어, 절단)할 수 있는 단백질 또는 폴리펩타이드를 언급한다. 일부 구현예에서, 엔도뉴클레아제는 이중 가닥 핵산의 단일 가닥을 절단할 수 있다. 일부 구현예에서, 엔도뉴클레아제는 이중 가닥 핵산 분자의 양 가닥을 절단할 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 데옥시리보뉴클레아제일 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 리보뉴클레아제일 수 있다.

[0499] 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 뉴클레아제 도메인은 표적 폴리뉴클레오타이드의 0개, 1개 또는 2개 가닥을 절단할 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 닉카제 도메인을 포함할 수 있다. 본원에서, 용어 "닉카제"는 듀플렉스 핵산 분자 (예를 들어, DNA)에서 2개 가닥 중 하나의 가닥만을 절단할 수 있는 뉴클레아제 도메인을 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 언급한다. 일부 구현예에서, 닉카제는 하나 이상의 돌연변이를 활성 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인에 도입함에 의해 완전한 촉매 활성 (예를 들어, 천연) 형태의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인으로부터 유래할 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인이 Cas9로부터 유래된 닉카제 도메인을 포함하는 경우, Cas9-유래된 닉카제 도메인은 D10A 돌연변이 및 위치 840에 히스티딘을 포함할 수 있다. 상기 구현예에서, 잔기 H840은 촉매 활성을 보유하고, 이로써 핵산 듀플렉스의 단일 가닥을 절단할 수 있다. 또 다른 예에서, Cas9-유래된 닉카제 도메인은 H840A 돌연변이를 포함할 수 있고, 위치 10에서 아미노산 잔기는 D로 남아있다. 일부 구현예에서, 닉카제는 닉카제 활성을 위해 요구되지 않는 뉴클레아제 도메인의 전부 또는 일부를 제거함에 의해 완전한 촉매 활성 (예를 들어, 천연) 형태의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인으로부터 유래할 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인이 Cas9로부터 유래된 닉카

제 도메인을 포함하는 경우, Cas9-유래된 닉카제 도메인은 RuvC 도메인 또는 HNH 도메인의 전부 또는 일부의 결실을 포함할 수 있다.

[0500] 닉카제 도메인을 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기는 따라서 특이적 폴리뉴클레오타이드 표적 서열에서 (예를 들어, 결합된 가이드 핵산의 상보적 서열에 의한 결정시) 단일 가닥 DNA 절단 (닉)을 생성할 수 있다. 일부 구현예에서, 닉카제 도메인 (예를 들어, Cas9-유래된 닉카제 도메인)을 포함하는 염기 편집기에 의해 절단되는 핵산 듀플렉스 표적 폴리뉴클레오타이드 서열의 가닥은 염기 편집기에 의해 편집되지 않는 가닥 (즉, 염기 편집기에 의해 절단되는 가닥은 편집될 염기를 포함하는 가닥의 반대편에 있다)이다. 다른 구현예에서, 닉카제 도메인 (예를 들어, Cas9-유래된 닉카제 도메인)을 포함하는 염기 편집기는 편집을 위해 표적화되는 DNA 분자의 가닥을 절단할 수 있다. 상기 구현예에서, 비-표적화된 가닥은 절단되지 않는다.

[0501] 또한 본원에는 촉매적 데드 (즉, 표적 폴리뉴클레오타이드 서열을 절단할 수 없는) 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기가 제공된다. 본원에서 용어 "촉매적 데드" 및 "뉴클레아제 데드"는 핵산의 가닥을 절단하지 못하는 무능력을 유도하는 하나 이상의 돌연변이 및/또는 결실을 갖는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 언급하기 위해 상호교환적으로 사용된다. 일부 구현예에서, 촉매적 데드 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 염기 편집기는 하나 이상의 뉴클레아제 도메인에서 특정 점 돌연변이의 결과로서 뉴클레아제 활성이 부재일 수 있다. 예를 들어, Cas9 도메인을 포함하는 염기 편집기의 경우에, Cas9는 D10A 돌연변이 및 H840A 돌연변이 둘 다를 포함할 수 있다. 상기 돌연변이는 뉴클레아제 도메인 둘 다를 불활성화시킴으로써 뉴클레아제 활성을 상실시킨다. 다른 구현예에서, 촉매적 데드 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 촉매 도메인 (예를 들어, RuvC1 및/또는 HNH 도메인)의 전부 또는 일부의 하나 이상의 결실을 포함할 수 있다. 추가의 구현예에서, 촉매적 데드 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 뉴클레아제 도메인의 전부 또는 일부의 결실 뿐만 아니라 점 돌연변이 (예를 들어, D10A 또는 H840A)를 포함한다.

[0502] 또한 본원에는 이전에 기능성 버전의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인으로부터의 촉매적 데드 폴리뉴클레오타이드 프로그래밍 가능한 결합 도메인을 생성할 수 있는 돌연변이가 고려된다. 예를 들어, 촉매적 데드 Cas9 ("dCas9")의 경우에, D10A 및 H840A 이외의 다른 돌연변이를 갖는 변이체가 제공되고 이는 뉴클레아제 불활성화된 Cas9를 유도한다. 상기 돌연변이는 예를 들어, D10 및 H840에서 다른 아미노산 치환, 또는 Cas9의 뉴클레아제 도메인 내 다른 치환 (예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인에서의 치환)을 포함한다.

[0503] 추가의 적합한 뉴클레아제-불활성 dCas9 도메인은 본원 개시내용 및 당해 분야의 지식을 기준으로 당업자에게 자명할 수 있고, 본원 개시내용의 범위 내에 있다. 상기 추가의 예시적인 적합한 뉴클레아제-불활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A, 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하지만 이에 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)) 참조: 833-838, 이의 전체 내용은 본원에 참조로 포함됨). 일부 구현예에서, dCas9 도메인은 본원에 제공된 dCas9 도메인 중 임의의 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 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 임의의 하나와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 임의의 하나와 비교하여 적어도 10, 적어도 15, 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 적어도 90, 적어도 100, 적어도 150, 적어도 200, 적어도 250, 적어도 300, 적어도 350, 적어도 400, 적어도 500, 적어도 600, 적어도 700, 적어도 800, 적어도 900, 적어도 1000, 적어도 1100, 또는 적어도 1200개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0504] 염기 편집기에 혼입될 수 있는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 비제한적인 예는 CRISPR 단백질-유래된 도메인, 제한 뉴클레아제, 메가뉴클레아제, TAL 뉴클레아제 (TALEN), 및 아연 핑거 뉴클레아제 (ZFN)를 포함한다. 일부 구현예에서, 염기 편집기는 결합된 가이드 핵산을 통해 핵산의 CRISPR (즉, 클러스터형 규칙적 간격을 둔 짧은 팔린드롬 반복체)-매개된 변형 동안에 핵산 서열에 결합할 수 있는 천연 또는 변형된 단백질 또는 이의 일부를 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합

도메인을 포함한다. 상기 단백질은 본원에서 "CRISPR 단백질"로서 언급된다. 따라서, 본원에서는 CRISPR 단백질의 전부 또는 일부를 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기 (즉, 또한 염기 편집기의 "CRISPR 단백질-유래된 도메인"으로서 언급되는 CRISPR 단백질의 전부 또는 일부를 도메인으로서 포함하는 염기 편집기)가 개시된다. 염기 편집기에 혼입된 CRISPR 단백질-유래된 도메인은 야생형 또는 천연 버전의 CRISPR 단백질과 비교하여 변형될 수 있다. 예를 들어, 하기 기재된 바와 같이, CRISPR 단백질-유래된 도메인은 야생형 또는 천연 버전의 CRISPR 단백질과 비교하여 하나 이상의 돌연변이, 삽입, 결실, 재정렬 및/또는 재조합을 포함할 수 있다.

[0505] 일부 구현예에서, 염기 편집기에 혼입된 CRISPR 단백질-유래된 도메인은 결합된 가이드 핵산과 접합되는 경우 표적 폴리뉴클레오타이드에 결합할 수 있는 엔도뉴클레아제 (예를 들어, 데옥시리보뉴클레아제 또는 리보뉴클레아제)이다. 일부 구현예에서, 염기 편집기에 혼입된 CRISPR 단백질-유래된 도메인은 결합된 가이드 핵산과 접합되는 경우 표적 폴리뉴클레오타이드에 결합할 수 있는 닉카제이다. 일부 구현예에서, 염기 편집기에 혼입된 CRISPR 단백질-유래된 도메인은 결합된 가이드 핵산과 접합되는 경우 표적 폴리뉴클레오타이드에 결합할 수 있는 촉매적 데드 도메인이다. 일부 구현예에서, 염기 편집기의 CRISPR 단백질 유래된 도메인에 의해 결합되는 표적 폴리뉴클레오타이드는 DNA이다. 일부 구현예에서, 염기 편집기의 CRISPR 단백질-유래된 도메인에 의해 결합되는 표적 폴리뉴클레오타이드는 RNA이다.

[0506] 일부 구현예에서, 염기 편집기의 CRISPR 단백질-유래된 도메인은 코리네박테리움 울세란스 (NCBI Refs: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아 (NCBI Refs: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라 (NCBI Ref: NC_021284.1); 프레보텔라 인터메디아 (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨스 (NCBI Ref: NC_021846.1); 스트렙토코커스 이니애 (NCBI Ref: NC_021314.1); 벨리엘라 발티카 (NCBI Ref: NC_018010.1); 사이크로플렉서스 토르쿠이스 (NCBI Ref: NC_018721.1); 스타필로코커스 써모필러스 (NCBI Ref: YP_820832.1); 리스테리아 이노쿠아 (NCBI Ref: NP_472073.1); 캄필로박터 제주니 (NCBI Ref: YP_002344900.1); 나이세리아 메닌기티디스 (NCBI Ref: YP_002342100.1), 스트렙토코커스 피오게네스, 또는 스타필로코커스 아우레우스로부터 기원하는 Cas9의 전부 또는 일부를 포함할 수 있다.

[0507] 일부 구현예에서, 염기 편집기의 Cas9-유래된 도메인은 스타필로코커스 아우레스로부터 기원하는 Cas9 도메인 (SaCas9)이다. 일부 구현예에서, SaCas9 도메인은 뉴클레아제 활성 SaCas9, 뉴클레아제 불활성 SaCas9 (SaCas9d), 또는 SaCas9 닉카제 (SaCas9n)이다. 일부 구현예에서, SaCas9 도메인은 N579X 돌연변이를 포함한다. 일부 구현예에서, SaCas9 도메인은 N579A 돌연변이를 포함한다. 일부 구현예에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 비-카노니칼 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 구현예에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 NNGRRT PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 구현예에서, SaCas9 도메인은 E781X, N967X, 및 R1014X 돌연변이 중 하나 이상을 포함한다.

[0508] 염기 편집기는 고충실도 Cas9인 Cas9의 전부 또는 일부로부터 유래된 도메인을 포함할 수 있다. 일부 구현예에서, 염기 편집기의 고충실도 Cas9 도메인은 가공된 Cas9 도메인이고, 이는 상응하는 야생형 Cas9 도메인과 비교하여, Cas9 도메인과, DNA의 당-포스페이트 골격 간의 정전기 상호작용을 감소시키는 하나 이상의 돌연변이를 포함한다. DNA의 당-포스페이트 골격과 감소된 정전기 상호작용을 갖는 고충실도 Cas9 도메인은 적은 표적-이탈 효과를 가질 수 있다. 일부 구현예에서, Cas9 도메인 (예를 들어, 야생형 Cas9 도메인)은 Cas9 도메인과 DNA의 당-포스페이트 골격 간의 연합을 감소시키는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, Cas9 도메인은 Cas9 도메인과 DNA의 당-포스페이트 골격 간의 연합을 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% 이상 까지 감소시키는 하나 이상의 돌연변이를 포함한다.

[0509] 일부 구현예에서, 변형된 Cas9는 고충실도 Cas9 효소이다. 일부 구현예에서, 고충실도 Cas9 효소는 SpCas9 (K855A), eSpCas9(1.1), SpCas9-HF1, 또는 초정밀 Cas9 변이체 (HypaCas9)이다. 변형된 Cas9 eSpCas9(1.1)는 HNH/RuvC 그루브와 비표적 DNA 가닥 간의 상호작용을 약화시키는 알라닌 치환을 포함하여 가닥 분리 및 표적-이탈 부위에서의 절단을 방지한다. 유사하게, SpCas9-HF1은 Cas9와 DNA 포스페이트 골격의 상호작용을 방해하는 알라닌 치환을 통해 표적-이탈 편집을 낮춘다. HypaCas9는 Cas9 교정 및 표적 식별을 증가시키는 REC3 도메인의 돌연변이 (SpCas9 N692A/M694A/Q695A/H698A)를 포함한다. 3개의 고충실도 효소 모두는 야생형 Cas9보다 표적-이탈 편집이 덜 발생한다. 예시적인 고충실도 Cas9가 아래에 제공된다.

[0510] Cas9에 비해 고충실도 Cas9 도메인 돌연변이는 굵게 표시하고 밑줄 쳐서 나타낸다

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0511]

[0512] ***가이드 폴리뉴클레오타이드***

[0513] 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 가이드 RNA이다. 본원에 사용된 바와 같이, 용어 "가이드 RNA (gRNA)" 및 이의 문법적 등가물은 표적 DNA에 대해 특이적일 수 있고 Cas 단백질과 복합체를 형성할 수 있는 RNA를 언급할 수 있다. RNA/Cas 복합체는 Cas 단백질의 표적 DNA로의 "가이딩"을 도울 수 있다. Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 환형 dsDNA 표적을 엔도핵산분해적으로 절단한다. crRNA에 상보적이지 않은 표적 가닥은 먼저 엔도핵산분해적으로 절단됨에 이어서 3'-5' 엑소핵산분해적으로 절단 제거한다. 실제로, DNA-결합 및 절단은 전형적으로 단백질 및 2개의 RNA를 요구한다. 그러나, 단일 가이드 RNA ("sgRNA, 또는 단순히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 양상을 단일 RNA 종으로 혼입하기 위해 가공될 수 있다. 예를 들어, 문헌 (Jinek M. *et al.*, Science 337:816-821(2012))을 참조하고 이의 전체 내용은 본원에 참조로 포함된다. Cas9는 자가 대 비-자가의 구분을 도와주기 위해 CRISPR 반복 서열 (PAM 또는 프로토스페이서 인접 모티프)에서 짧은 모티프를 인지한다. Cas9 뉴클레아제 서열 및 구조는 당업자에게 널리 공지되어 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, 이의 각각의 전체 내용은 본원에 참조로 포함됨). Cas9 오톨로그는 에스. 피오게네스 및 에스. 써모필러스를 포함하지만 이에 제한되지 않는 다양한 종에 기재되어 있다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본원 개시내용을 기준으로 당업자에게 자명할 수 있고, 상기 Cas9 뉴클레아제 및 서열은 이의 전문이 본원에 참조로 포함되는 문헌 (참조: 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 구현예에서, Cas9 뉴클레아제는 불활성(예를 들어, 불활성화된) DNA 절단 도메인을 갖고, 즉, Cas9는 닉카제이다.

[0514] 일부 구현예에서, 가이드 폴리뉴클레오타이드는 적어도 하나의 단일 가이드 RNA ("sgRNA" 또는 "gRNA")이다.

일부 구현예에서, 가이드 폴리뉴클레오타이드는 적어도 하나의 tracrRNA이다. 일부 구현예에서, 가이드 폴리뉴클레오타이드는 폴리뉴클레오타이드 프로그래밍 가능한 DNA-결합 도메인 (예를 들어, Cas9 또는 Cpf1)을 표적 뉴클레오타이드 서열로 가이드하기 위해 PAM 서열을 요구하지 않는다.

[0515] 본원에 개시된 염기 편집기의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 (예를 들어, CRISPR-유래된 도메인)은 가이드 폴리뉴클레오타이드와 연합함에 의해 표적 뉴클레오타이드 서열을 인지할 수 있다. 가이드 폴리뉴클레오타이드 (예를 들어, gRNA)는 전형적으로 단일 가닥이고 폴리뉴클레오타이드의 표적 서열에 부위 특이적으로 결합 (즉, 상보적 염기 쌍 형성을 통해)하도록 프로그래밍되어 가이드 핵산과 접합된 염기 편집기를 표적 서열로 지시할 수 있다. 가이드 폴리뉴클레오타이드는 DNA일 수 있다. 가이드 폴리뉴클레오타이드는 RNA일 수 있다. 당업자에 의해 이해되는 바와 같이, 가이드 폴리뉴클레오티드 서열에서, 우라실 (U)은 서열에서 티민 (T)을 대체한다. 일부 구현예에서, 가이드 폴리뉴클레오타이드는 천연 뉴클레오타이드 (예를 들어, 아데노신)를 포함한다. 일부 구현예에서, 가이드 폴리뉴클레오타이드는 비-천연 (또는 비천연) 뉴클레오타이드 (예를 들어, 펩타이드 핵산 또는 뉴클레오타이드 유사체)를 포함한다. 일부 구현예에서, 가이드 핵산 서열의 표적화 영역은 적어도 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개 뉴클레오타이드 길이일 수 있다. 가이드 핵산의 표적화 영역은 10 내지 30개 뉴클레오타이드 길이, 또는 15 내지 25개 뉴클레오타이드 길이 또는 15 내지 20개 뉴클레오타이드 길이일 수 있다. 일부 구현예에서, 가이드 폴리뉴클레오타이드는 특히 5' 말단에서 1, 2, 3, 4개 등의 뉴클레오타이드에 의해 절단될 수 있다. 비제한적인 예로서, 20개 뉴클레오타이드 길이의 가이드 폴리뉴클레오타이드는 특히 5' 말단에서 1, 2, 3, 4개 등의 뉴클레오타이드에 의해 절단될 수 있다.

[0516] 일부 구현예에서, 가이드 폴리뉴클레오타이드는 2개 이상의 개별 폴리뉴클레오타이드를 포함하고, 이는 예를 들어, 상보성 염기 쌍 형성 (예를 들어, 이중 가이드 폴리뉴클레오타이드)을 통해 서로 상호작용할 수 있다. 예를 들어, 가이드 폴리뉴클레오타이드는 CRISPR RNA (crRNA) 및 트랜스-활성화 CRISPR RNA (tracrRNA)를 포함할 수 있다. 예를 들어, 가이드 폴리뉴클레오타이드는 하나 이상의 트랜스-활성화 CRISPR RNA (tracrRNA)를 포함할 수 있다.

[0517] II형 CRISPR 시스템에서, CRISPR 단백질 (예를 들어, Cas9)에 의한 핵산의 표적화는 전형적으로 표적 서열을 인지하는 서열을 포함하는 제1 RNA 분자 (crRNA)와 가이드 RNA-CRISPR 단백질 복합체를 안정화시키는 스캐폴드 영역을 형성하는 반복체 서열을 포함하는 제2 RNA 분자 (trRNA) 간에 상보적 염기 쌍 형성을 요구한다. 상기 이중 가이드 RNA 시스템은 가이드로 폴리뉴클레오타이드로서 사용되어 본원에 개시된 염기 편집기를 표적 폴리뉴클레오타이드 서열로 지시할 수 있다.

[0518] 일부 구현예에서, 본원에 제공된 염기 편집기는 단일 가이드 폴리뉴클레오타이드 (예를 들어, sgRNA)를 사용한다. 일부 구현예에서, 본원에 제공된 염기 편집기는 이중 가이드 폴리뉴클레오타이드 (예를 들어, 이중 gRNA)를 사용한다. 일부 구현예에서, 본원에 제공된 염기 편집기는 하나 이상의 가이드 폴리뉴클레오타이드 (예를 들어, 다중 gRNA)를 사용한다. 일부 구현예에서, 단일 가이드 폴리뉴클레오타이드는 본원에 기재된 상이한 염기 편집기에 대해 사용된다. 예를 들어, 단일 가이드 폴리뉴클레오타이드는 아데노신 염기 편집기를 위해 사용될 수 있다.

[0519] 다른 구현예에서, 가이드 폴리뉴클레오타이드는 단일 분자 (즉, 단일-분자 가이드 핵산)로 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. 본원에서, 용어 가이드 폴리뉴클레오타이드 서열은 염기 편집기와 상호작용할 수 있고 염기 편집기를 표적 폴리뉴클레오타이드 서열로 지시할 수 있는 임의의 단일, 이중 또는 다중-분자 핵산을 고려한다.

[0520] 전형적으로, 가이드 폴리뉴클레오타이드 (예를 들어, crRNA/trRNA 복합체 또는 gRNA)는 표적 폴리뉴클레오타이드 서열을 인지하고 이와 결합할 수 있는 서열을 포함하는 "폴리뉴클레오타이드-표적화 분절" 및 염기 편집기의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 성분 내 가이드 폴리뉴클레오타이드를 안정화시키는 "단백질-결합 분절"을 포함한다. 일부 구현예에서, 가이드 폴리뉴클레오타이드의 폴리뉴클레오타이드 표적화 분절은 DNA 폴리뉴클레오타이드를 인지하고 이와 결합하여 DNA 내 염기의 편집을 촉진시킨다. 다른 구현예에서, 가이드 폴리뉴클레오타이드의 폴리뉴클레오타이드 표적화 분절은 RNA 폴리뉴클레오타이드를 인지하고 이와 결합하여 RNA 내 염기의 편집을 촉진시킨다. 본원에서 "분절"은 분자의 섹션 또는 영역, 예를 들어, 가이드 폴리뉴클레오타이드 내 뉴클레오타이드의 연속 스트레치를 언급한다. 분절은 또한 분절이 하나 초과의 분자의 영역을 포함할 수 있도록 하는 복합체의 영역/섹션을 언급할 수 있다. 예를 들어, 가이드 폴리뉴클레오타이

드가 다중 핵산 분자를 포함하는 경우, 이의 단백질-결합 분절은 예를 들어, 상보체 영역을 따라 하이브리드화하는 다중 분리된 분자의 전부 또는 일부를 포함할 수 있다. 일부 구현예에서, 2개의 분리된 분자를 포함하는 DNA-표적화 RNA의 단백질-결합 분절은 (i) 100개 염기 쌍 길이인 제1 RNA 분자의 염기 쌍 40-75개; 및 (ii) 50개 염기 쌍 길이인 제2 RNA 분자의 염기 쌍 10-25개를 포함할 수 있다. 특정 문맥에서 구체적으로 달리 정의되지 않는 한, "분절"의 정의는 특정 수의 총 염기 쌍으로 제한되지 않고, 소정의 RNA 분자로부터의 임의의 특정 수의 염기 쌍으로 제한되지 않고, 복합체 내 특정 수의 분리된 분자로 제한되지 않고, 임의의 총 길이를 갖는 RNA 분자 영역을 포함할 수 있고 다른 분자와 상보성을 갖는 영역을 포함할 수 있다.

[0521] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 2개 이상의 RNA, 예를 들어, CRISPR RNA (crRNA) 및 트랜스활성화 crRNA (tracrRNA)를 포함할 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 때로는 단일 쇄 RNA, 또는 crRNA와 tracrRNA의 일부 (예를 들어, 기능성 부분)의 융합에 의해 형성된 단일 가이드 RNA (sgRNA)를 포함할 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 또한 crRNA 및 tracrRNA를 포함하는 이중 RNA일 수 있다. 추가로, crRNA는 표적 DNA와 하이브리드화할 수 있다.

[0522] 상기 논의된 바와 같이, 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 발현 생성물일 수 있다. 예를 들어, 가이드 RNA를 암호화하는 DNA는 가이드 RNA를 암호화하는 서열을 포함하는 벡터일 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 가이드 RNA 및 프로모터를 암호화하는 서열을 포함하는 단리된 가이드 RNA 또는 플라스미드 DNA로 세포를 형질감염시킴에 의해 세포에 전달될 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 또한 바이러스-매개된 유전자 전달을 사용하는 것과 같은 다른 방식으로 세포에 전달될 수 있다.

[0523] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 단리될 수 있다. 예를 들어, 가이드 RNA는 단리된 RNA 형태로 세포 또는 유기체에 형질감염될 수 있다. 가이드 RNA는 당업계에 공지된 임의의 시험관내 전사 시스템을 사용한 시험관내 전사에 의해 제조될 수 있다. 가이드 RNA는 가이드 RNA에 대한 암호화 서열을 포함하는 플라스미드 형태 보다는 단리된 RNA 형태로 세포에 전달될 수 있다.

[0524] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 3개의 영역을 포함할 수 있다: 염색체 서열에서 표적 부위에 상보적일 수 있는 5' 말단에서 제1 영역, 스템 루프 구조를 형성할 수 있는 제2 내부 영역 및 단일 가닥일 수 있는 제3의 3' 영역. 각각의 가이드 RNA의 제1 영역은 또한 각각의 가이드 RNA가 융합 단백질을 특정 표적 부위로 가이드하도록 상이할 수 있다. 추가로, 각각의 가이드 RNA의 제2 및 제3 영역은 모든 가이드 RNA에서 동일할 수 있다.

[0525] 가이드 RNA 또는 가이드 폴리뉴클레오타이드의 제1 영역은 가이드 RNA의 제1 영역이 표적 부위와 염기 쌍을 형성할 수 있도록 염색체 서열에서 표적 부위에서 서열에 상보적일 수 있다. 일부 구현예에서, 가이드 RNA의 제1 영역은 10개 뉴클레오타이드 내지 25개 뉴클레오타이드 이상 또는 약 10개 뉴클레오타이드 내지 25개 뉴클레오타이드 (즉, 10개 뉴클레오타이드 내지 뉴클레오타이드, 또는 약 10개 뉴클레오타이드 내지 약 25개 뉴클레오타이드; 또는 10개 뉴클레오타이드 내지 약 25개 뉴클레오타이드; 또는 약 10개 뉴클레오타이드 내지 25개 뉴클레오타이드) 이상을 포함할 수 있다. 예를 들어, 가이드 RNA의 제1 영역과 염색체 서열에서 표적 부위 간에 염기 쌍을 형성하는 영역은 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25개 이상 또는 약 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25개 이상의 뉴클레오타이드 길이일 수 있다. 일부 구현예에서, 가이드 RNA의 제1 영역은 19, 20, 또는 21개 또는 약 19, 20, 또는 21개 뉴클레오타이드 길이일 수 있다.

[0526] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 또한 2차 구조를 형성하는 제2 영역을 포함할 수 있다. 예를 들어, 가이드 RNA에 의해 형성된 2차 구조는 스템 (또는 헤어핀) 및 루프를 포함할 수 있다. 루프 및 스템의 길이는 다양할 수 있다. 예를 들어, 루프는 3 내지 10개 또는 약 3 내지 10개 뉴클레오타이드 길이의 범위일 수 있고 스템은 6 내지 20개 또는 약 6 내지 20개 염기 쌍 길이의 범위일 수 있다. 스템은 1 내지 10개 또는 약 10개 뉴클레오타이드의 하나 이상의 돌출부를 포함할 수 있다. 제2 영역의 전체 길이는 16 내지 60개 또는 약 16 내지 60개 뉴클레오타이드 길이의 범위일 수 있다. 예를 들어, 루프는 4개 또는 약 4개 뉴클레오타이드 길이일 수 있고 스템은 12개 또는 약 12개 염기 쌍일 수 있다.

[0527] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 또한 필수적으로 단일 가닥일 수 있는 3' 말단에서 제3 영역을 포함할 수 있다. 예를 들어, 제3 영역은 때로는 관심 대상의 세포에서 임의의 염색체 서열과 상보적이지 않고 때로는 가이드 RNA의 나머지와 상보적이지 않다. 추가로, 제3 영역의 길이는 다양할 수 있다. 제3 영역은 4개 또는 약 4개 초과의 뉴클레오타이드 길이일 수 있다. 예를 들어, 제3 영역의 길이는 5 내지 60개 또는 약 5 내지 60개 뉴클레오타이드 길이의 범위일 수 있다.

[0528] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 유전자 표적의 임의의 엑손 또는 인트론을 표적화할 수 있다. 일부 구현예에서, 가이드는 유전자의 엑손 1 또는 2를 표적화할 수 있고; 다른 구현예에서, 가이드는 유전자의 엑손 3 또는 4를 표적화할 수 있다. 조성물은 모두 동일한 엑손을 표적화하는 다중 가이드 RNA 또는 일부 구현예에서 상이한 엑손을 표적화할 수 있는 다중 가이드 RNA를 포함할 수 있다. 유전자의 엑손 및 인트론은 표적화될 수 있다.

[0529] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 약 20개 뉴클레오타이드의 핵산 서열을 표적화할 수 있다. 표적 핵산은 약 20개 미만의 뉴클레오타이드일 수 있다. 표적 핵산은 적어도 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30개 또는 적어도 약 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30개, 또는 1-100개 뉴클레오타이드 길이 사이일 수 있다. 표적 핵산은 최대 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50개 또는 최대 약 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50개 또는 1-100개 뉴클레오타이드 길이 사이일 수 있다. 표적 핵산 서열은 PAM의 첫번째 뉴클레오타이드의 5'에 바로 인접하게 20개 염기 또는 약 20개 염기일 수 있다. 가이드 RNA는 핵산 서열을 표적화할 수 있다. 표적 핵산은 적어도 1-10, 1-20, 1-30, 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, 또는 1-100개 또는 적어도 약 1-10, 1-20, 1-30, 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, 또는 1-100개 뉴클레오타이드일 수 있다.

[0530] 가이드 폴리뉴클레오타이드, 예를 들어, 가이드 RNA는 세포의 게놈에서 또 다른 핵산, 예를 들어, 표적 핵산 또는 프로토스페이서에 하이브리드화할 수 있는 핵산을 언급할 수 있다. 가이드 폴리뉴클레오타이드는 RNA일 수 있다. 가이드 폴리뉴클레오타이드는 DNA일 수 있다. 가이드 폴리뉴클레오타이드는 핵산 서열에 부위 특이적으로 결합하도록 프로그래밍될 수 있거나 디자인될 수 있다. 가이드 폴리뉴클레오타이드는 폴리뉴클레오타이드 쇄를 포함할 수 있고 단일 가이드 폴리뉴클레오타이드로 언급될 수 있다. 가이드 폴리뉴클레오타이드는 2개의 폴리뉴클레오타이드 쇄를 포함할 수 있고 이중 가이드 폴리뉴클레오타이드로 언급될 수 있다. 가이드 RNA는 RNA 분자로서 세포 또는 배아에 도입될 수 있다. 예를 들어, RNA 분자는 시험관내 전사될 수 있고/있거나 화학적으로 합성될 수 있다. RNA는 합성 DNA 분자, 예를 들어, gBlocks® 유전자 단편으로부터 전사될 수 있다. 이어서, 가이드 RNA는 RNA 분자로서 세포 또는 배아에 도입될 수 있다. 가이드 RNA는 또한 비-RNA 핵산 분자, 예를 들어, DNA 분자 형태로 세포 또는 배아에 도입될 수 있다. 예를 들어, 가이드 RNA를 암호화하는 DNA는 관심 대상의 세포 또는 배아에서 가이드 RNA의 발현을 위해 프로모터 제어 서열에 작동적으로 연결될 수 있다. RNA 암호화 서열은 RNA 폴리머라제 III (Pol III)에 인지되는 프로모터 서열에 작동적으로 연결될 수 있다. 가이드 RNA를 발현하도록 사용될 수 있는 플라스미드 벡터는 px330 벡터 및 px333 벡터를 포함하지만 이에 제한되지 않는다. 일부 구현예에서, 플라스미드 벡터 (예를 들어, px333 벡터)는 적어도 2개의 가이드 RNA-암호화 DNA 서열을 포함할 수 있다.

[0531] 가이드 폴리뉴클레오타이드, 예를 들어, 가이드 RNA를 선택하고, 디자인하고 입증하고, 서열을 표적화하기 위한 방법은 본원에 기재되고 당업자에게 공지되어 있다. 예를 들어, 핵염기 편집기 시스템에서 데아미나제 도메인 (예를 들어, AID 도메인)의 잠재적 기질 혼잡의 영향을 최소화하기 위해, 탈아민화를 위해 의도치 않게 표적화될 수 있는 잔기 (예를 들어, 표적 핵산 유전자좌 내 ssDNA 상에 잠재적으로 위치할 수 있는 표적-이탈 C 잔기)의 수는 최소화될 수 있다. 추가로, 소프트웨어 도구를 사용하여 표적 핵산 서열에 상응하는 gRNA를 최적화할 수 있고, 예를 들어, 게놈에 걸쳐 총 표적-이탈 활성을 최소화할 수 있다. 예를 들어, 에스. 피오게네스 Cas9를 사용한 각각의 가능한 표적화 도메인 선택을 위해, 모든 표적-이탈 서열 (이전의 선택된 PAM, 예를 들어, NAG 또는 NGG)은 최대 특정 수 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 미스매칭된 염기쌍을 함유하는 게놈에 걸쳐 동정될 수 있다. 표적 부위에 상보적인 gRNA의 제1 영역은 동정될 수 있고, 모든 제1 영역 (예를 들어, crRNA)은 이의 총 예측된 표적-이탈 스코어에 따라 등급화될 수 있고; 상위 등급의 표적화 도메인은 최대 표적-적중 및 최소 표적-이탈 활성을 가질 가능성이 있는 것들을 나타낸다. gRNA를 표적화하는 후보물은 당업계에 공지되고/되거나 본원에 제시된 바와 같은 방법을 사용함에 의해 기능적으로 평가될 수 있다.

[0532] 비제한적인 예로서, Cas9와 함께 사용하기 위한 가이드 RNA의 crRNA 내 표적 DNA 하이브리드화 서열은 DNA 서열 검색 알고리즘을 사용하여 동정될 수 있다. gRNA 디자인은 문헌 (참조: 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 기재된 바와 같이 공용 도구 cas-오핀더 (offinder)를 기반으로 하는 맞춤형 gRNA 디자인 소프트웨어를 사용하여 수행될 수 있다. 상기 소프트웨어는 게놈-와이드 표적-이탈 성향을 계산한 후 가이드를 스코어링한다. 전형적으로 완벽한 매칭에서 7개의 미스매칭 범위의 매칭은 17 내지 24개 길이 범위의 가이드를 위해 고려된다. 표적-이탈 부위가 계산적으로 결정되면, 합

계 스코어가 각각의 가이드에 대해 계산되고 웹-인터페이스를 사용하여 표 형식으로 요약된다. PAM 서열에 인접한 잠재적 표적 부위를 동정하는 것에 추가로, 소프트웨어는 또한 선택된 표적 부위로부터 1, 2, 3개 또는 3개 초과의 뉴클레오타이드가 상이한 모든 PAM 인접한 서열을 동정한다. 표적핵산 서열, 예를 들어, 표적 유전자에 대한 게놈 DNA 서열이 수득될 수 있고, 반복 요소들은 공개적으로 가용한 도구, 예를 들어, RepeatMasker 프로그램을 사용하여 스크리닝될 수 있다. RepeatMasker는 반복적인 요소들 및 낮은 복잡성의 영역에 대한 인풋 DNA 서열을 검색한다. 아웃풋은 소정의 탐색 서열에 존재하는 반복체의 상세한 주석이다.

[0533] 동정 후, 가이드 RNA, 예를 들어, crRNA의 제1 영역은 표적 부위까지의 이들의 거리, 이들의 직교성 및 관련 PAM 서열 (예를 들어, 에스. 피오게네스에 대해 NGG PAM 또는 에스. 아우레우스에 대한 NNGRRV PAM을 함유하는 인간 게놈 내 근접한 매칭의 동정을 기반으로 하는 5'G)과 함께 근접한 매칭에 대한 5' 뉴클레오타이드의 존재를 기준으로 하는 역가로 등급화된다. 본원에 사용된 바와 같이, 직교성은 표적 서열에 대한 최소 수의 미스매칭을 함유하는 인간 게놈 내 서열의 수를 언급한다. "고수준의 직교성" 또는 "양호한 직교성"은 예를 들어, 의도된 표적 이외에 인간 게놈 내 동일한 서열을 갖지 않거나 표적 서열 내 1개 또는 2개 미스매칭을 함유한 서열을 갖지 않는 20량체 표적화 도메인을 언급할 수 있다. 양호한 직교성을 갖는 표적화 도메인은 표적-이탈 DNA 절단을 최소화하기 위해 선택될 수 있다.

[0534] 일부 구현예에서, 리포터 시스템은 염기-편집 활성을 검출하고 후보 가이드 폴리뉴클레오타이드를 시험하기 위해 사용될 수 있다. 일부 구현예에서, 리포터 시스템은 염기 편집 활성이 리포터 유전자의 발현을 유도하는 리포터 유전자-기반 검정을 포함할 수 있다. 예를 들어, 리포터 시스템은 탈활성화된 개시 코돈, 예를 들어, 3'-TAC-5'에서 3'-CAC-5'로의 주형 가닥 상에 돌연변이를 포함하는 리포터 유전자를 포함할 수 있다. 표적 C의 성공적인 탈아민화 시, 상응하는 mRNA는 5'-GUG-3' 대신 5'-AUG-3'로서 전사될 것이고 이는 리포터 유전자의 해독을 가능하게 한다. 적합한 리포터 유전자는 당업자에게 자명할 것이다. 리포터 유전자의 비제한적인 예는 녹색 형광 단백질 (GFP), 적색 형광 단백질 (RFP), 루시퍼라제, 분비된 알칼린 포스파타제 (SEAP)를 암호화하는 유전자, 또는 이의 발현이 검출 가능하고 당업자에게 자명한 임의의 다른 유전자를 포함한다. 리포터 시스템은 많은 상이한 gRNA를 시험하기 위해, 예를 들어, 표적 DNA 서열과 관련하여 각각의 데아미나제가 표적화하는 뉴클레오타이드 잔기(들)가 어느 것인지를 결정하기 위해 사용될 수 있다. 비-주형 가닥 뉴클레오타이드 잔기를 표적화하는 sgRNA는 또한 특이적 염기 편집 단백질, 예를 들어, Cas9 데아미나제 융합 단백질의 표적-이탈 효과를 평가하기 위해 시험될 수 있다. 일부 구현예에서, 상기 gRNA는 돌연변이된 개시 코돈이 gRNA와 염기 쌍을 형성하지 않도록 디자인될 수 있다. 가이드 폴리뉴클레오타이드는 표준 뉴클레오타이드, 변형된 뉴클레오타이드 (예를 들어, 슈도우리딘), 뉴클레오타이드 이성체 및/또는 뉴클레오타이드 유사체를 포함할 수 있다. 일부 구현예에서, 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 형광성 염료), 검출 태그 (예를 들어, 비오틴, 디곡시게닌 등), 양자 도트 또는 골드 입자일 수 있다.

[0535] 가이드 폴리뉴클레오타이드는 화학적으로 및/또는 효소적으로 합성될 수 있다. 예를 들어, 가이드 RNA는 표준 포스포르아미드트-기반 고체-상 합성 방법을 사용하여 합성될 수 있다. 대안적으로, 가이드 RNA는 가이드 RNA를 암호화하는 DNA를 파아지 RNA 폴리머라제에 의해 인지되는 프로모터 제어 서열로 작동적으로 연결함에 의해 시험관내 합성될 수 있다. 적합한 파아지 프로모터 서열의 예는 T7, T3, SP6 프로모터 서열, 또는 이의 변이체를 포함한다. 가이드 RNA가 2개의 별도의 분자 (예를 들어, crRNA 및 tracrRNA)를 포함하는 구현예에서, crRNA는 화학적으로 합성되고 tracrRNA는 효소적으로 합성될 수 있다.

[0536] 일부 구현예에서, 염기 편집기 시스템은 다중 가이드 폴리뉴클레오타이드, 예를 들어, gRNA를 포함할 수 있다. 예를 들어, gRNA는 하나 이상의 표적 유전자좌 (예를 들어, 적어도 하나 (1) gRNA, 적어도 2개 gRNA, 적어도 5개 gRNA, 적어도 10개 gRNA, 적어도 20개 gRNA, 적어도 30개 gRNA, 또는 적어도 50개 gRNA)에 염기 편집기를 표적화할 수 있다. 일부 구현예에서, 다중 gRNA 서열은 반복적으로 정렬되고 바람직하게 직접 반복체에 의해 분리된다.

[0537] 가이드 RNA 또는 가이드 폴리뉴클레오타이드를 암호화하는 DNA 서열은 또한 벡터의 일부일 수 있다. 일부 구현예에서, 벡터는 추가의 발현 제어 서열 (예를 들어, 인핸서 서열, 코작 서열, 폴리아데닐화 서열, 전사 종결 서열 등), 선택 가능한 마커 서열 (예를 들어, GFP 또는 항생제 내성 유전자, 예를 들어, 푸로마이신), 복제 오리진 등을 포함한다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드를 암호화하는 DNA 분자는 또한 선형 또는 환형일 수 있다.

[0538] 일부 구현예에서, 염기 편집기 시스템의 하나 이상의 성분은 DNA 서열에 의해 암호화될 수 있다. 상기 DNA 서열은 별도로 또는 함께 발현 시스템, 예를 들어, 세포에 도입될 수 있다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 가이드 RNA를 암호화하는 DNA 서열은 세포에 도입될 수 있고, 각각의 DNA 서열은 별도의 분자 (예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 암호화 서열을 함유하는 하나의 벡터 및 가이드 RNA 암호화 서열을 함유하는 제2 벡터)의 일부일 수 있거나 둘 다는 동일한 분자 (예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 가이드 RNA 둘 다에 대한 암호화 (및 조절) 서열을 함유하는 하나의 벡터)의 일부일 수 있다.

[0539] 가이드 폴리뉴클레오타이드는 하나 이상의 변형을 포함하여 새롭거나 증진된 특성을 갖는 핵산을 제공할 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.

[0540] 일부 구현예에서, gRNA 또는 가이드 폴리뉴클레오타이드는 변형을 포함할 수 있다. 변형은 gRNA 또는 가이드 폴리뉴클레오타이드의 임의의 위치에서 가해질 수 있다. 하나 초과의 변형은 단일 gRNA 또는 가이드 폴리뉴클레오타이드에 가해질 수 있다. gRNA 또는 가이드 폴리뉴클레오타이드는 변형 후 품질 관리를 거칠 수 있다. 일부 구현예에서, 품질 제어는 PAGE, HPLC, MS, 또는 이의 임의의 조합을 포함할 수 있다.

[0541] gRNA 또는 가이드 폴리뉴클레오타이드의 변형은 치환, 삽입, 결실, 화학적 변형, 물리적 변형, 안정화, 정제, 또는 이들의 임의의 조합일 수 있다.

[0542] gRNA 또는 가이드 폴리뉴클레오타이드는 또한 5' 아데닐레이트, 5' 구아노신-트리포스페이트 캡, 5' N7-메틸구아노신-트리포스페이트 캡, 5' 트리포스페이트 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, 무염기성, 아크리딘, 아조벤젠, 비오틴, 비오틴 BB, 비오틴 TEG, 콜레스테릴 TEG, 데스티오비오틴 TEG, DNP TEG, DNP-X, DOTA, dT-비오틴, 이중 비오틴, PC 비오틴, 프소랄렌 C2, 프소랄렌 C6, TINA, 3'DABCYL, 블랙 홀 켄처 1, 블랙 홀 켄처 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, 리보뉴클레오타이드 유사체, 2'-O-메틸 리보뉴클레오사이드 유사체, 당 변형된 유사체, 워블 (wobble)/범용 염기, 형광성 염료 표지, 2'-플루오로 RNA, 2'-O-메틸 RNA, 메틸포스페이트, 포스포디에스테르 DNA, 포스포디에스테르 RNA, 포스포티오에이트 DNA, 포스포로티오에이트 RNA, UNA, 슈도우리딘-5'-트리포스페이트, 5'-메틸시티딘-5'-트리포스페이트 또는 이들의 임의의 조합에 의해 변형될 수 있다.

[0543] 일부 구현예에서, 변형은 영구적이다. 다른 구현예에서, 변형은 일과성이다. 일부 구현예에서, 다중 변형은 gRNA 또는 가이드 폴리뉴클레오타이드에 가해진다. gRNA 또는 가이드 폴리뉴클레오타이드 변형은 뉴클레오타이드의 생리화학적 성질, 예를 들어, 이들의 형태, 극성, 소수성, 화학적 반응성, 염기 쌍 형성 상호작용 또는 이들의 임의의 조합을 변경할 수 있다.

[0544] 변형은 또한 포스포로티오에이트 치환체일 수 있다. 일부 구현예에서, 천연 포스포티에스테르 결합은 세포 뉴클레아제에 의한 신속한 분해에 민감해질 수 있고; 포스포로티오에이트 (PS) 결합 치환체를 사용한 뉴클레오타드 상호 연결의 변형은 세포 분해에 의한 가수분해에 대해 보다 안정할 수 있다. 변형은 gRNA 또는 가이드 폴리뉴클레오타이드에서 안정성을 증가시킬 수 있다. 변형은 또한 생물학적 활성을 증진시킬 수 있다. 일부 구현예에서, 포스포로티오에이트 증진된 RNA gRNA는 RNase A, RNase T1, 소 혈청 뉴클레아제, 또는 이들의 임의의 조합물을 저해할 수 있다. 이들 성질은 PS-RNA gRNA가, 뉴클레아제에 대한 노출이 생체내 또는 시험관내에서 가능성이 높은 적용에서 사용될 수 있도록 한다. 예를 들어, 포스포로티오에이트 (PS) 결합은 엑소뉴클레아제 분해를 저해할 수 있는 gRNA의 5'- 또는 3'-말단에서 13 내지 5개 뉴클레오타이드 사이에 도입될 수 있다. 일부 구현예에서, 포스포로티오에이트 결합은 전체 gRNA에 걸쳐 부가되어 엔도뉴클레아제에 의한 공격을 감소시킬 수 있다.

[0545] ***프로토스페이서 인접 모티프***

[0546] 용어 "프로토스페이서 인접 모티프 (PAM)" 또는 PAM-유사 모티프는 CRISPR 세균 후천성 면역계에서 Cas9 뉴클레아제에 의해 표적화된 DNA 서열 직후 2-6 염기 쌍 DNA 서열을 언급한다. 일부 구현예에서, PAM은 5' PAM (즉, 프로토스페이서의 5' 말단의 업스트림에 위치함)일 수 있다. 다른 구현예에서, PAM은 3' PAM (즉, 프로토스페

이서의 5' 말단의 다운스트림에 위치함)일 수 있다. PAM 서열은 표적 결합을 위해 필수적이고, 정확한 서열은 Cas 단백질 유형에 의존한다. 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 이에 제한되지 않는다. Y는 피리미딘이고; N는 임의의 뉴클레오타이드 염기이고; W는 A 또는 T이다.

[0547] 본원에 제공된 염기 편집기는 카노니칼 또는 비-카노니칼 프로토스페이서 인접 모티프 (PAM) 서열을 포함하는 뉴클레오타이드 서열에 결합할 수 있는 CRISPR 단백질 유래된 도메인을 포함할 수 있다. PAM 부위는 표적 폴리뉴클레오타이드 서열에 인접한 뉴클레오타이드 서열이다. 본원 개시내용의 일부 양상은 상이한 PAM 특이성을 갖는 CRISPR 단백질의 전부 또는 일부를 포함하는 염기 편집기를 제공한다.

[0548] 예를 들어, 전형적으로 Cas9 단백질, 예를 들어, 에스. 피오게네스 기원의 Cas9 (spCas9)는 특정 핵산 영역에 결합하기 위해 카노니칼 NGG PAM 서열을 필요로 하고, 여기서, "NGG"에서 "N"은 아데닌 (A), 티민 (T), 구아닌 (G), 또는 시토신 (C)이고, G는 구아닌이다. PAM은 CRISPR 단백질-특이적일 수 있고 상이한 CRISPR 단백질 유래된 도메인을 포함하는 상이한 염기 편집기 간에 상이할 수 있다. PAM은 표적 서열의 5' 또는 3'일 수 있다. PAM은 표적 서열의 업스트림 또는 다운스트림일 수 있다. PAM은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 뉴클레오타이드 길이일 수 있다. 흔히, PAM은 2-6개 사이의 뉴클레오타이드 길이이다.

[0549] 일부 구현예에서, PAM은 "NRN" PAM이고, 여기서, "NRN"에서 "N"은 아데닌 (A), 티민 (T), 구아닌 (G), 또는 시토신 (C)이고, R은 아데닌 (A) 또는 구아닌 (G)이고; 또는 PAM은 "NYN" PAM이고, 여기서, NYN에서 "N"은 아데닌 (A), 티민 (T), 구아닌 (G), 또는 시토신 (C)이고, Y는 시티딘 (C) 또는 티민 (T)으로, 예를 들어, 문헌 (R.T. Walton et al., 2020, *Science,* 10.1126/science.aba8853 (2020))에 기재되어 있고, 이의 전체 내용은 본원에 참조로 포함된다.

[0550] 여러 PAM 변이체는 아래 **표 2**에 기재되어 있다.

[0551] [표 2] 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
SpyMacCas9	NAA
Cpf1	5′ (TTTV)

[0552]

[0553] 일부 구현예에서, PAM은 NGT이다. 일부 구현예에서, NGC PAM은 Cas9 변이체에 의해 인지된다. 일부 구현예에서, NGC PAM 변이체는 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R (총체적으로 "MQKFRAER"로 칭함)로부터 선택된 하나 이상의 아미노산 치환을 포함한다.

[0554] 일부 구현예에서, PAM은 NGT이다. 일부 구현예에서, NGT PAM은 Cas9 변이체에 의해 인지된다. 일부 구현예에서, NGT PAM 변이체는 하나 이상의 잔기 1335, 1337, 1135, 1136, 1218, 및/또는 1219에서 표적화된 돌연변이를 통해 생성된다. 일부 구현예에서, NGT PAM 변이체는 하나 이상의 잔기 1219, 1335, 1337, 1218에서 표적화

된 돌연변이를 통해 생성된다. 일부 구현예에서, NGT PAM 변이체는 하나 이상의 잔기 1135, 1136, 1218, 1219, 및 1335에서 표적화된 돌연변이를 통해 생성된다. 일부 구현예에서, NGT PAM 변이체는 아래 **표 3** 및 **표 4**에 제공된 표적화된 돌연변이 세트로부터 선택된다.

[0555] [표 3] 잔기 1219, 1335, 1337, 1218에서 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	

[0556]

[0557] [표 4] 잔기 1135, 1136, 1218, 1219, 및 1335에서 NGT PAM 변이체 돌연변이

	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			

[0558]

[0559] 일부 구현예에서, NGT PAM 변이체는 표 3 및 4에서 변이체 5, 7, 28, 31 또는 36으로부터 선택된다. 일부 구현예에서, 변이체는 개선된 NGT PAM 인지를 갖는다.

[0560] 일부 구현예에서, NGT PAM 변이체는 잔기 1219, 1335, 1337, 및/또는 1218에서 돌연변이를 갖는다. 일부 구현예에서, NGT PAM 변이체는 아래 표 5에 제공된 변이체로부터 개선된 인지에 대해 돌연변이로 선택된다.

[0561] [표 5] 잔기 1219, 1335, 1337, 및 1218에서 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

[0562]

[0563] 일부 구현예에서, NGT PAM에 대해 특이성을 갖는 염기 편집기는 아래 **표 6**에 제공된 바와 같이 생성될 수 있다.

[0564] [표 6] 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

[0565]

[0566] 일부 구현예에서, NGTN 변이체는 변이체 1이다. 일부 구현예에서, NGTN 변이체는 변이체 2이다. 일부 구현예에서, NGTN 변이체는 변이체 3이다. 일부 구현예에서, NGTN 변이체는 변이체 4이다. 일부 구현예에서, NGTN 변이체는 변이체 5이다. 일부 구현예에서, NGTN 변이체는 변이체 6이다.

[0567] 일부 구현예에서, Cas9 도메인은 스타필로코커스 피오게네스 (*Staphylococcus pyogenes*)로부터 기원하는 Cas9 도메인 (SpCas9)이다. 일부 구현예에서, SpCas9 도메인은 뉴클레아제 활성 SpCas9, 뉴클레아제 불활성 SpCas9 (SpCas9d), 또는 SpCas9 닉카제 (SpCas9n)이다. 일부 구현예에서, SpCas9는 본원에 제공된 임의의 아미노산 서열에서 D9X 돌연변이, 또는 상응하는 돌연변이를 포함하고, 여기서, X는 D를 제외한 임의의 아미노산이다. 일부 구현예에서, SpCas9는 본원에 제공된 임의의 아미노산 서열에서 D9A 돌연변이, 또는 상응하는 돌연변이를 포함한다. 일부 구현예에서, SpCas9 도메인, SpCas9d 도메인 또는 SpCas9n 도메인은 비-카노니칼 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 구현예에서, SpCas9 도메인, SpCas9d 도메인 또는 SpCas9n 도메인은 NGG, NGA 또는 NGCG PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135X, R1335X, 및 T1337X 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함하고, 여기서, X는 임의의 아미노산이다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135E, R1335Q, 및 T1337R 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함한다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135E, R1335Q 및 T1337R 돌연변이, 또는 상응하는 돌연변이를 포함한다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135X, R1335X, 및 T1337X 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함하고, 여기서, X는 임의의 아미노산이다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135V, R1335Q, 및 T1337R 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함한다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135V, R1335Q 및 T1337R 돌연변이, 또는 상응하는 돌연변이를 포함한다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135X, G1218X, R1335X 및 T1337X 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함하고, 여기서, X는 임의의 아미노산이다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135V, G1218R, R1335Q, 및 T1337R 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함한다. 일부 구현예에서, SpCas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D1135V, G1218R, R1335Q 및 T1337R 돌연변이, 또는 상응하는 돌연변이를 포함한다.

[0568] 일부 구현예에서, 본원에 제공된 임의의 융합 단백질의 Cas9 도메인은 본원에 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 구현예에서, 본원에 제공된 임의의 융합 단백질의 Cas9 도메인은 본원에 기재된 임의의 Cas9 폴리펩타이드의 아미노산 서열을 포함한다. 일부 구현예에서, 본원에 제공된 임의의 융합 단백질의 Cas9 도메인은 본원에 기재된 임의의 Cas9 폴리펩타이드의 아미노산 서열로 이루어진다.

[0569] 일부 예에서, 본원에 개시된 염기 편집기의 CRISPR 단백질-유래된 도메인에 의해 인지되는 PAM은 염기 편집기를 암호화하는 삽입체 (예를 들어, AAV 삽입체)에 대해 분리된 올리고뉴클레오타이드 상의 세포에 제공될 수 있다.

상기 구현예에서, 분리된 올리고뉴클레오타이드 상에 PAM 제공은 인접한 PAM이 표적 서열과 동일한 폴리뉴클레오타이드 상에 존재하지 않기 때문에 달리 절단될 수 없는 표적 서열의 절단을 가능하게 할 수 있다.

[0570] 하나의 구현예에서, 에스. 피오게네스 Cas9 (SpCas9)는 게놈 가공을 위한 CRISPR 엔도뉴클레아제로서 사용될 수 있다. 그러나, 기타의 것들이 사용될 수 있다. 일부 구현예에서, 상이한 엔도뉴클레아제는 특정 게놈 표적에 표적화하기 위해 사용될 수 있다. 일부 구현예에서, 비-NGG PAM 서열을 갖는 합성 SpCas9-유래된 변이체가 사용될 수 있다. 추가로, 다양한 종 기원의 다른 Cas9 오톨로그가 동정되었고 이들 "비-SpCas9"는 또한 본원 개시내용을 위해 유용할 수 있는 다양한 PAM 서열에 결합할 수 있다. 예를 들어, 상대적으로 대형 크기의 SpCas9 (대략적으로 4 kb 암호화 서열)는 세포에서 효율적으로 발현될 수 없는 SpCas9 cDNA를 갖는 플라스미드를 유도할 수 있다. 역으로, 스타필로코커스 아우레우스 Cas9 (SaCas9)에 대한 암호화 서열은 SpCas9보다 대략적으로 1 킬로베이스 짧아, 능히 이것이 세포에서 효율적으로 발현되도록 한다. SpCas9와 유사하게, SaCas9 엔도뉴클레아제는 시험관내 포유동물 세포에서 및 생체내 마우스에서 표적 유전자를 변형시킬 수 있다. 일부 구현예에서, Cas 단백질은 상이한 PAM 서열을 표적화할 수 있다. 일부 구현예에서, 표적 유전자는 예를 들어, Cas9 PAM, 5'-NGG에 인접해 있을 수 있다. 다른 구현예에서, 다른 Cas9 오톨로그는 상이한 PAM 요건을 가질 수 있다. 예를 들어, 에스. 써모필러스의 것들 (CRISPR1에 대해 5'-NNAGAA 및 CRISPR3에 대해 5'-NGGNG) 및 나이세리아 메닌기티디스 (*Neisseria meningiditis*)의 것들 (5'-NNNNGATT)과 같은 기타 PAM은 또한 표적 유전자에 인접해 있는 것으로 발견될 수 있다.

[0571] 일부 구현예에서, 에스. 피오게네스 시스템에 대해, 표적 유전자 서열은 5'-NGG PAM에 선행 (즉, 이에 대해 5')할 수 있고, 20-nt 가이드 RNA 서열은 반대 가닥과 염기 쌍을 형성하여 PAM에 인접한 Cas9 절단을 매개할 수 있다. 일부 구현예에서, 인접한 절단은 PAM의 업스트림의 3개 염기 쌍에 있을 수 있거나 약 3개 염기 쌍에 있을 수 있다. 일부 구현예에서, 인접한 절단은 PAM의 업스트림의 10개 염기 쌍에 있을 수 있거나 약 10개 염기 쌍에 있을 수 있다. 일부 구현예에서, 인접한 절단은 PAM의 업스트림의 0-20개 염기 쌍에 있을 수 있거나 약 0-20개 염기 쌍에 있을 수 있다. 예를 들어, 인접한 절단은 PAM의 업스트림의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개 염기 쌍 다음에 있을 수 있다. 인접한 절단은 또한 PAM의 다운스트림의 1 내지 30개 염기 쌍에 있을 수 있다. PAM 서열에 결합할 수 있는 예시적인 SpCas9 단백질의 서열은 다음과 같다:

[0572] 예시적인 PAM-결합 SpCas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0573]

[0574] 예시적인 PAM-결합 SpCas9n의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0575]

[0576] 예시적인 PAM-결합 SpEQR Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0577]

[0578] 상기 서열에서, D1135, R1335, 및 T1337로부터 돌연변이되어 SpEQR Cas9를 생성할 수 있는 잔기 E1135, Q1335, 및 R1337은 밑줄 치고 굵게 표시한다.

[0579] 예시적인 PAM-결합 SpVQR Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0580]

[0581] D1135, R1335, 및 T1337로부터 돌연변이되어 SpVQR Cas9를 생성할 수 있는 상기 잔기 V1135, Q1335, 및 R1337은 밑줄 치고 굵게 표시한다.

[0582] 예시적인 PAM-결합 SpVRER Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA**R**ELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STKEVLDATLIHQSITGLYETRIDLSQL
GGD.

[0583]

[0584] D1135, G1218, R1335, 및 T1337로부터 돌연변이되어 SpVRER Cas9를 생성할 수 있는 상기 잔기 V1135, R1218, E1335, 및 R1337은 밑줄 치고 굵게 표시한다.

[0585] 예시적인 PAM-결합 SpVRQR Cas9의 아미노산 서열은 다음과 같다: 예시적인 SpVRQR Cas9

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA**R**ELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQL
GGD.

[0586]

[0587] D1135, G1218, R1335, 및 T1337로부터 돌연변이되어 SpVRQR Cas9를 생성할 수 있는 상기 잔기 V1135, R1218, Q1335, 및 R1337은 밑줄 치고 굵게 표시한다.

[0588] 일부 구현예에서, 가공된 SpCas9 변이체는 3'H (비-G PAM)에 의해 플랭킹된 프로토스페이서 인접 모티프 (PAM) 서열을 인지할 수 있다 (**표 1A-1D** 참조). 일부 구현예에서, SpCas9 변이체는 NRNH PAM (여기서, R은 A 또는 G이고, H는 A, C 또는 T이다)을 인지한다. 일부 구현예에서, 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 전문이 본원에 참조로 포함됨).

[0589] 일부 구현예에서, Cas9 도메인은 재조합 Cas9 도메인이다. 일부 구현예에서, 재조합 Cas9 도메인은 SpyMacCas9 도메인이다. 일부 구현예에서, SpyMacCas9 도메인은 뉴클레아제 활성 SpyMacCas9, 뉴클레아제 불활성 SpyMacCas9 (SpyMacCas9d), 또는 SpyMacCas9 닉카제 (SpyMacCas9n)이다. 일부 구현예에서, SaCas9 도메인, SaCas9d 도메인 또는 SaCas9n 도메인은 비-카노니칼 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 구현예에서, SpyMacCas9 도메인, SpCas9d 도메인 또는 SpCas9n 도메인은 NAA PAM 서열을 갖는 핵산 서열에 결합할 수 있다.

[0590] 고유 5'-NAAN-3' PAM 특이성을 갖는, 스트렙토코커스 마카카에 (*Streptococcus macacae*)에서 Spy Cas9의 예시적인 Cas9 A 동족체의 서열은 당업계에 공지되어 있고 예를 들어, 문헌 (참조: Jakimo *et al.*, (www.biorxiv.org/content/biorxiv/early/2018/09/27/429654.full.pdf)에 기재되어 있고, 아래에 제공된다.

[0591] 예시적인 SpyMacCas9

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQL
FEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF
IKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN
IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQN
GGLFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQ
NPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHH
LDSDLSNDYLQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLL
GFTQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

[0592]

[0593] 일부 구현예에서, 변이체 Cas9 단백질은 H840A, P475A, W476A, N477A, D1125A, W1126A, 및 D1218A 돌연변이를 함유하여 폴리펩타이드는 표적 DNA 또는 RNA를 절단하는 감소된 능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다. 또 다른 비제한적인 예로서, 일부 구현예에서, 변이체 Cas9 단백질은 D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, 및 D1218A 돌연변이를 함유하여 폴리펩타이드는 표적 DNA를 절단하는 감소된 능력을 갖는다. 상기 Cas9 단백질은 표적 DNA (예를 들어, 단일 가닥 표적 DNA)를 절단하는 감소된 능력을 갖지만 표적 DNA (예를 들어, 단일 가닥 표적 DNA)에 결합하는 능력을 보유한다. 일부 구현예에서, 변이체 Cas9 단백질이 W476A 및 W1126A 돌연변이를 함유하는 경우 또는 상기 변이체 Cas9 단백질이 P475A, W476A, N477A, D1125A, W1126A, 및 D1218A 돌연변이를 함유하는 경우, 변이체 Cas9 단백질은 효율적으로 PAM 서열에 결합하지 않는다. 따라서, 일부 상기 경우에, 상기 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 방법은 PAM 서열을 요구하지 않는다. 다시 말해, 일부 구현예에서, 상기 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 방법은 가이드 RNA를 포함할 수 있지만 상기 방법은 PAM 서열의 부재하에 수행될 수 있다 (그리고, 결합 특이성은 따라서 가이드 RNA의 표적화 분절에 의해 제공된다). 다른 잔기는 상기 효과를 성취하기 위해 돌연변이될 수 있다 (즉, 하나 또는 다른 핵염기 부분을 불활성화시킬 수 있다). 비제한적인 예로서, 잔기 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, 및/또는 A987은 변경 (즉, 치환된)될 수 있다. 또한, 알라닌 치환과는 다른 돌연변이가 적합하다.

[0594] 일부 구현예에서, 염기 편집기의 CRISPR 단백질-유래된 도메인은 카노니칼 PAM 서열 (NGG)을 갖는 Cas9 단백질 전부 또는 일부를 포함할 수 있다. 다른 구현예에서, 염기 편집기의 Cas9-유래된 도메인은 비-카노니칼 PAM 서열을 사용할 수 있다. 상기 서열은 당업계에 기재되었고 당업자에게 자명할 것이다. 예를 들어, 비-카노니칼 PAM 서열에 결합하는 Cas9 도메인은 문헌 (참조: Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" Nature 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition" Nature Biotechnology 33, 1293-1298 (2015); 이의 각각의 전체 내용은 본원에 참조로 포함됨)에 기재되어 있다.

[0595] ***PAM 배타성이 감소된 Cas9 도메인***

[0596] 전형적으로, Cas9 단백질, 예를 들어, 에스. 피오게네스 기원의 Cas9 (spCas9)는 특정 핵산 영역에 결합하기 위해 카노니칼 NGG PAM 서열을 필요로 하고, 여기서, "NGG"에서 "N"은 아데노신 (A), 티미딘 (T), 또는 시토신 (C)이고, G는 구아노신이다. 이것은 게놈 내 목적하는 염기를 편집하는 능력을 제한할 수 있다. 일부 구현예에서, 본원에 제공된 염기 편집 융합 단백질은 정확한 위치에서, 예를 들어, PAM의 업스트림인 표적 염기를 포함하는 영역에 위치할 필요가 있을 수 있다. 예를 들어, 문헌 (참조: Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016)을 참조하고, 이의 전체 내용은 본원에 참조로 본원에 포함된다. 따라서, 일부 구현예에서, 본원에 제공된 임의의 융합 단백질은 카노니칼 (예를 들어, NGG) PAM 서열을 포함하지 않는 뉴클레오타이드 서열에 결합할 수 있는 Cas9 도메인을 포함할 수 있다. 비-카노니칼 PAM 서열에 결합하는 Cas9 도메인은 당업계에 기재되었고 당업자에게 자명할 것이다. 예를 들어, 비-카노니칼 PAM 서열에 결합하는 Cas9 도메인은 문헌 (참조: Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); 이의 각각의 내용은 본원에 참조로 포함됨)에 기재되어 있다.

[0597] ***고충실도 Cas9 도메인***

[0598] 본원 개시내용의 일부 양상은 고충실도 Cas9 도메인을 제공한다. 일부 구현예에서, 고충실도 Cas9 도메인은 가공된 Cas9 도메인이고, 이는 상응하는 야생형 Cas9 도메인과 비교하여, Cas9 도메인과, DNA의 당-포스페이트 골격 간의 정전기 상호작용을 감소시키는 하나 이상의 돌연변이를 포함한다. 임의의 특정 이론에 국한시키고자 하는 것 없이, DNA의 당-포스페이트 골격과 감소된 정전기 상호작용을 갖는 고충실도 Cas9 도메인은 적은 표적-이탈 효과를 가질 수 있다. 일부 구현예에서, Cas9 도메인 (예를 들어, 야생형 Cas9 도메인)은 Cas9 도메인과 DNA의 당-포스페이트 골격 간의 연합을 감소시키는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, Cas9 도메인은 Cas9 도메인과 DNA의 당-포스페이트 골격 간의 연합을 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%까지 감소시키는 하나 이상의 돌연변이를 포함한다.

[0599] 일부 구현예에서, 본원에 제공된 임의의 Cas9 융합 단백질은 본원에 제공된 임의의 아미노산 서열에서 N497X, R661X, Q695X 및/또는 Q926X 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함하고, 여기서, X는 임의의 아미노산이다. 일부 구현예에서, 본원에 제공된 임의의 Cas9 융합 단백질은 본원에 제공된 임의의 아미노산 서열에서 N497A, R661A, Q695A, 및/또는 Q926A 돌연변이, 또는 상응하는 돌연변이 중 하나 이상을 포함한다. 일부 구현예에서, Cas9 도메인은 본원에 제공된 임의의 아미노산 서열에서 D10A 돌연변이, 또는 상응하는 돌연변이를 포함한다. 고충실도를 갖는 Cas9 도메인은 당업계에 공지되어 있고 당업자에게 자명할 것이다. 예를 들어, 고충실도를 갖는 Cas9 도메인은 문헌 (참조: Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." *Nature* 529, 490-495 (2016); and Slaymaker, I.M., *et al.* "Rationally engineered Cas9 nucleases with improved specificity." *Science* 351, 84-88 (2015); 이의 전체 내용은 본원에 참조로 포함됨)에 기재되어 있다.

[0600] 일부 구현예에서, 변형된 Cas9는 고충실도 Cas9 효소이다. 일부 구현예에서, 고충실도 Cas9 효소는 SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, 또는 하이퍼 정확한 Cas9 변이체 (HypaCas9)이다. 변형된 Cas9 eSpCas9(1.1)는 HNH/RuvC 그루브와 비-표적 DNA 가닥 간의 상호작용을 약화시키는 알라닌 치환을 함유하여 가닥 분리 및 표적-이탈 부위에서의 절단을 방지한다. 유사하게, SpCas9-HF1은 DNA 포스페이트 골격과 Cas9의 상호작용을 붕괴시키는 알라닌 치환을 통한 표적-이탈 편집을 저하시킨다. HypaCas9는 Cas9 교정 및 표적 식별을 증가시키는 REC3 도메인에서 돌연변이 (SpCas9 N692A/M694A/Q695A/H698A)를 함유한다. 모든 3개의 고충실도 효소는 야생형 Cas9보다 적은 표적-이탈 편집을 생성한다.

[0601] 예시적인 고충실도 Cas9는 아래에 제공된다. Cas9에 비해 고충실도 Cas9 도메인 돌연변이는 굵게 표시하고 밑줄 쳐서 나타낸다.

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD.

[0602]

[0603] ***핵 국소화 서열 (NLS)을 포함하는 융합 단백질***

[0604] 일부 구현예에서, 본원에 제공된 융합 단백질은 하나 이상 (예를 들어, 2, 3, 4, 5개)의 핵 표적화 서열, 예를 들어, 핵 국소화 서열 (NLS)을 추가로 포함한다. 하나의 구현예에서, 이분된 NLS가 사용된다. 일부 구현예에서, NLS는 NLS를 포함하는 단백질의 세포 핵 (예를 들어, 핵 수송에 의해)으로의 혼입을 촉진시키는 아미노산 서열을 포함한다. 일부 구현예에서, 본원에 제공된 임의의 융합 단백질은 핵 국소화 서열 (NLS)을 추가로 포함한다. 일부 구현예에서, NLS는 융합 단백질의 N-말단에 융합된다. 일부 구현예에서, NLS는 융합 단백질의 C-말단에 융합된다. 일부 구현예에서, NLS는 Cas9 도메인의 N-말단에 융합된다. 일부 구현예에서, NLS는 nCas9 도메인 또는 dCas9 도메인의 C-말단에 융합된다. 일부 구현예에서, NLS는 데아미나제의 N-말단에 융합된다. 일부 구현예에서, NLS는 데아미나제의 C-말단에 융합된다. 일부 구현예에서, NLS는 하나 이상의 링커를 통해 융합 단백질에 융합된다. 일부 구현예에서, NLS는 링커 없이 융합 단백질에 융합된다. 일부 구현예에서, NLS는 본원에 제공되거나 참조된 NLS 서열 중 어느 하나의 아미노산 서열을 포함한다. 추가의 핵 국소화 서열은 당업계에 공지되어 있고 당업자에게 자명할 것이다. 예를 들어, NLS 서열은 문헌 (참조: Plank *et al.*, PCT/EP2000/011690)에 기재되어 있고, 이의 내용은 예시적인 핵 국소화 서열에 대한 이의 기재를 위해 본원에 참조로 포함된다. 일부 구현예에서, NLS는 하기의 아미노산 서열을 포함한다:

[0605] PKKKRKVEGADKRTADGSEFESPKKKRKV,

[0606] KRTADGSEFESPKKKRKV,

[0607] KRPAATKKAGQAKKKK,

[0608] KKTELQTTNAENKTKKL,

[0609] KRGINDRNFWRGENGRKTR,

[0610] RKSGKIAAIVVKRPRKPKKKRKV, 또는

[0611] MDSLLMNRRKFLYQFKNVRWAKGRRETYLC.

[0612] 일부 구현예에서, NLS는 링커에 존재하거나, NLS는 링커, 예를 들어, 본원에 기재된 링커에 의해 플랭킹된다. 일부 구현예에서, N-말단 또는 C-말단 NLS는 양분된 NLS이다. 이분된 NLS는 2개의 염기성 아미노산 클러스터를

포함하고, 이들은 상대적으로 짧은 스페이서 서열에 의해 분리되어 있다 (따라서 이분된- 2개 부분, 단일부분의 NLS가 아니다). 뉴클레오플라스민, KR[PAATKKAGQA]KKKK의 NLS는 흔한 이분된 신호의 원형이다: 기본 아미노산의 2개의 클러스터는 약 10개 아미노산의 스페이서에 의해 분리되어 있다. 예시적인 양분된 NLS의 서열은 다음과 같다: PKKKRKVEGADKRTADGSEFESPKKKRKV.

[0613] 일부 구현예에서, 본원에 기재된 융합 단백질은 링커 서열을 포함하지 않는다. 일부 구현예에서, 하나 이상의 도메인 또는 단백질 사이에 링커 서열이 존재한다. 일부 구현예에서, 아데노신 데아미나제 및 Cas9 도메인과 함께 예시적인 Cas9 융합 단백질의 일반 구조는 하기의 구조 중 임의의 하나를 포함하고, 여기서, NLS는 핵 국소화 서열 (예를 들어, 본원에 제공된 임의의 NLS)이고, NH_2는 융합 단백질의 N-말단이고, COOH는 융합 단백질의 C-말단이다:

[0614] NH2-NLS-[아데노신 데아미나제]-[Cas9 도메인]-COOH;

[0615] NH2-NLS-[Cas9 도메인]-[아데노신 데아미나제]-COOH;

[0616] NH2-[아데노신 데아미나제]-[Cas9 도메인]-NLS-COOH; 또는

[0617] NH2-[Cas9 도메인]-[아데노신 데아미나제]-NLS-COOH.

[0618] 본원 개시내용의 융합 단백질은 하나 이상의 추가의 특성을 포함할 수 있음을 인지해야 한다. 예를 들어, 일부 구현예에서, 융합 단백질은 융합 단백질의 가용화, 정제 또는 검출을 위해 유용한 서열 태그 뿐만 아니라 저해제, 세포질 국소화 서열, 배출 서열, 예를 들어, 핵 배출 서열 또는 다른 국소화 서열을 포함할 수 있다. 본원에 제공된 적합한 단백질 태그는 비오틴 카복실라제 캐리어 단백질 (BCCP) 태그, myc-태그, 칼모듈린 (calmodulin)-태그, FLAG-태그, 헤마글루티닌 (HA)-태그, 또한 히스티딘 태그 또는 His-태그로서 언급되는 폴리히스티딘 태그, 말토스 결합 단백질 (MBP)-태그, nus-태그, 글루타티온-S-트랜스퍼라제 (GST)-태그, 녹색 형광성 단백질 (GFP)-태그, 티오레독신-태그, S-태그, 소프트태그 (예를 들어, 소프트태그 1, 소프트태그 3), strep-태그, 비오틴 리가제 태그, FlAsH 태그, V5 태그 및 SBP-태그를 포함하지만 이에 제한되지 않는다. 추가의 적합한 서열은 당업자에게 자명할 것이다. 일부 구현예에서, 융합 단백질은 하나 이상의 His 태그를 포함한다.

[0619] 하나 이상의 핵 국소화 서열 (NLS)을 포함하는 CRISPR 효소를 암호화하는 벡터가 사용될 수 있다. 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개의 NLS가 사용될 수 있다. CRISPR 효소는 아미노-말단에서 또는 이의 근처에서 NLS, 카복시 말단에서 또는 이의 근처에서 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 초과 NLS 또는 이들의 임의의 조합 (예를 들어, 아미노-말단에서 하나 이상의 NLS 및 카복시 말단에서 하나 이상의 NLS)을 포함할 수 있다. 하나 초과의 NLS가 존재하는 경우, 각각은 서로 독립적으로 선택될 수 있어 단일 NLS는 하나 초과의 카피로 및/또는 하나 이상의 카피로 존재하는 하나 이상의 다른 NLS와 조합하여 존재할 수 있다.

[0620] 방법에 사용되는 CRISPR 효소는 약 6개 NLS를 포함할 수 있다. NLS는 NLS에 최근접 아미노산이 N- 또는 C-말단으로부터 폴리펩타이드 쇄를 따라 약 50개 아미노산 내에, 예를 들어, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 또는 50개 아미노산 내에 있는 경우 N- 또는 C-말단 근처에 고려된다.

[0621] ***베타성이 감소된 Cas9 도메인***

[0622] 전형적으로, Cas9 단백질, 예를 들어, 에스. 피오게네스 기원의 Cas9 (spCas9)는 특정 핵산 영역에 결합하기 위해 카노니칼 NGG PAM 서열을 필요로 하고, 여기서, "NGG"에서 "N"은 아데노신 (A), 티미딘 (T), 또는 시토신 (C)이고, G는 구아노신이다. 이것은 게놈 내 목적하는 염기를 편집하는 능력을 제한할 수 있다. 일부 구현예에서, 본원에 제공된 염기 편집 융합 단백질은 정확한 위치에서, 예를 들어, PAM의 업스트림인 표적 염기를 포함하는 영역에 위치할 필요가 있을 수 있다. 예를 들어, 문헌 (참조: Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016))을 참조하고, 이의 전체 내용은 본원에 참조로 본원에 포함된다. 따라서, 일부 구현예에서, 본원에 제공된 임의의 융합 단백질은 카노니칼 (예를 들어, NGG) PAM 서열을 포함하지 않는 뉴클레오타이드 서열에 결합할 수 있는 Cas9 도메인을 포함할 수 있다. 비-카노니칼 PAM 서열에 결합하는 Cas9 도메인은 당업계에 기재되었고 당업자에게 자명할 것이다. 예를 들어, 비-카노니칼 PAM 서열에 결합하는 Cas9 도메인은 문헌 (참조: Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus*

aureus CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); Nishimasu, H., *et al.*, "Engineered CRISPR-Cas9 nuclease with expanded targeting space" *Science*. 2018 Sep 21;361(6408):1259-1262, Chatterjee, P., *et al.*, Minimal PAM specificity of a highly similar SpCas9 ortholog" *Sci Adv*. 2018 Oct 24;4(10):eaau0766. doi: 10.1126/sciadv.aau0766; 이의 각각의 내용은 본원에 참조로 포함됨)에 기재되어 있다.

[0623] ***내부 삽입물을 갖는 융합 단백질***

[0624] 본원에는 핵산 프로그래밍 가능한 핵산 결합 단백질, 예를 들어, napDNAbp에 융합된 이종성 폴리펩타이드를 포함하는 융합 단백질이 제공된다. 이종성 폴리펩타이드는 본래의 또는 야생형 napDNAbp 폴리펩타이드 서열에서 발견되지 않는 폴리펩타이드일 수 있다. 이종성 폴리펩타이드는 napDNAbp의 C-말단에, napDNAbp의 N-말단에서 napDNAbp에 융합될 수 있거나, napDNAbp의 내부 위치에 삽입될 수 있다. 일부 구현예에서, 이종성 폴리펩타이드는 napDNAbp의 내부 위치에 삽입된다.

[0625] 일부 구현예에서, 이종성 폴리펩타이드는 데아미나제 또는 이의 기능성 단편이다. 예를 들어, 융합 단백질은 Cas9 또는 Cas12 (예를 들어, Cas12b/C2c1), 폴리펩타이드의 N-말단 단편 및 및 C-말단 단편에 의해 플랭킹된 데아미나제를 포함할 수 있다. 융합 단백질 내 데아미나제는 아데노신 데아미나제일 수 있다. 일부 구현예에서, 아데노신 데아미나제는 TadA (예를 들어, TadA7.10 또는 TadA*8)이다. 일부 구현예에서, TadA는 TadA*8이다. 본원에 기재된 바와 같은 TadA 서열 (예를 들어, TadA7.10 또는 TadA*8)은 상기 기재된 융합 단백질에 적합한 데아미나제이다.

[0626] 데아미나제는 환형의 퍼뮤턴트 데아미나제일 수 있다. 예를 들어, 데아미나제는 환형의 퍼뮤턴트 아데노신 데아미나제일 수 있다. 일부 구현예에서, 데아미나제는 TadA 참조 서열에 넘버링된 바와 같이 아미노산 잔기 116에서 환형으로 퍼뮤턴트화된 환형의 퍼뮤턴트 TadA이다. 일부 구현예에서, 데아미나제는 TadA 참조 서열에 넘버링된 바와 같이 아미노산 잔기 136에서 환형으로 치환된 환형의 퍼뮤턴트 TadA이다. 일부 구현예에서, 데아미나제는 TadA 참조 서열에 넘버링된 바와 같이 아미노산 잔기 65에서 환형으로 치환된 환형의 퍼뮤턴트 TadA이다.

[0627] 융합 단백질은 하나 초과의 데아미나제를 포함할 수 있다. 융합 단백질은 예를 들어 1, 2, 3, 4, 5개 이상의 데아미나제를 포함할 수 있다. 일부 구현예에서, 융합 단백질은 하나의 데아미나제를 포함한다. 일부 구현예에서, 융합 단백질은 2개의 데아미나제를 포함한다. 2개 이상의 데아미나제는 동종이량체일 수 있다. 2개 이상의 데아미나제는 이종이량체일 수 있다. 2개 이상의 데아미나제는 napDNAbp에 탠덤으로 삽입될 수 있다. 일부 구현예에서, 2개 이상의 데아미나제는 napDNAbp에 탠덤으로 있을 수 없다.

[0628] 일부 구현예에서, 융합 단백질 내 napDNAbp는 Cas9 폴리펩타이드 또는 이의 단편이다. Cas9 폴리펩타이드는 변이체 Cas9 폴리펩타이드일 수 있다. 일부 구현예에서, Cas9 폴리펩타이드는 Cas9 닉카제 (nCas9) 폴리펩타이드 또는 이의 단편이다. 일부 구현예에서, Cas9 폴리펩타이드는 뉴클레아제 데드 Cas9 (dCas9) 폴리펩타이드 또는 이의 단편이다. 융합 단백질 내 Cas9 폴리펩타이드는 전장 Cas9 폴리펩타이드일 수 있다. 일부 경우에, 융합 단백질 내 Cas9 폴리펩타이드는 전장 Cas9 폴리펩타이드일 수 없다. Cas9 폴리펩타이드는 예를 들어, 천연적으로 발생하는 Cas9 단백질에 비해 N-말단 또는 C-말단에서 절단될 수 있다. Cas9 폴리펩타이드는 환형으로 퍼뮤턴트화된 Cas9 단백질일 수 있다. Cas9 폴리펩타이드는 표적 폴리뉴클레오타이드 및 가이드 핵산 서열에 여전히 결합할 수 있는, Cas9 폴리펩타이드의 단편, 일부 또는 도메인일 수 있다.

[0629] 일부 구현예에서, Cas9 폴리펩타이드는 스트렙토코커스 피오게네스 Cas9 (SpCas9), 스타필로코커스 아우레우스 Cas9 (SaCas9), 스트렙토코커스 써모필러스 1 Cas9 (St1Cas9), 또는 이의 단편 또는 변이체이다.

[0630] 융합 단백질의 Cas9 폴리펩타이드는 천연적으로 발생하는 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.

[0631] 융합 단백질의 Cas9 폴리펩타이드는 아래에 제시된 Cas9 아미노산 서열 (아래에 "Cas9 참조 서열"로 불림)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:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDE
VAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ
TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN
FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY
EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI
SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK
AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDF
QFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAK
YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG
ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS
KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYN
KHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD

[0632]

[0633] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0634] Cas9 폴리펩티드의 N- 및 C-말단 단편에 의해 플랭킹된 이종성 촉매 도메인을 포함하는 융합 단백질은 또한 본 원에 기재된 방법에서 염기 편집에 유용하다. Cas9 및 하나 이상의 데아미나제 도메인, 예를 들어, 아데노신 데아미나제를 포함하거나, Cas9 서열에 의해 플랭킹된 아데노신 데아미나제 도메인을 포함하는 융합 단백질은 또한 표적 서열의 고도로 특이적이고 효율적인 염기 편집에 유용하다. 하나의 구현예에서, 키메라 Cas9 융합 단백질은 Cas9 폴리펩티드 내에 삽입된 이종성 촉매 도메인 (예를 들어, 아데노신 데아미나제)을 포함한다. 일부 구현예에서, 융합 단백질은 Cas9 내에 삽입된 아데노신 데아미나제 도메인 및 시티딘 데아미나제 도메인을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 Cas9 내에 융합되고 아데노신 데아미나제는 C-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas9 내에 융합되고 아데노신 데아미나제는 N-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas9 내에 융합되고 아데노신 데아미나제는 C-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas9 내에 융합되고 아데노신 데아미나제는 N-말단에 융합된다.

[0635] 다양한 구현예에서, 촉매 도메인은 아데노신 데아미나제 활성과 같은 DNA 변형 활성 (예를 들어, 데아미나제 활성)을 갖는다. 일부 구현예에서, 아데노신 데아미나제는 TadA (예를 들어, TadA7.10)이다. 일부 구현예에서, TadA는 TadA*8이다. 일부 구현예에서, TadA*8은 Cas9 내에 융합되고 아데노신 데아미나제는 C-말단에 융합된다. 일부 구현예에서, TadA*8은 Cas9 내에 융합되고 아데노신 데아미나제는 N-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas9 내에 융합되고 TadA*8은 C-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas9 내에 융합되고 TadA*8은 N-말단에 융합된다. TadA*8 및 아데노신 데아미나제 및 Cas9와의 융합 단백질의 예시적인 구조는 다음과 같이 제공된다:

[0636] NH_2-[Cas9(TadA*8)]-[아데노신 데아미나제]-COOH;

[0637] NH_2-[아데노신 데아미나제]-[Cas9(TadA*8)]-COOH;

[0638] NH_2-[Cas9(아데노신 데아미나제)]-[TadA*8]-COOH; 또는

[0639] NH_2-[TadA*8]-[Cas9(아데노신 데아미나제)]-COOH.

[0640] 일부 구현예에서, 상기 일반 구조에서 사용된 "-"는 임의의 링커의 존재를 지적한다.

[0641] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 적합한 위치에서 napDNAbp (예를 들어, Cas9 또는 Cas12 (예를 들어, Cas12b/C2c1))에 삽입될 수 있어서, napDNAbp는 표적 폴리뉴클레오타이드 및 가이드 핵산에 결합하는 이의 능력을 보유한다. 데아미나제 (예를 들어, 아데노신 데아미나제)는 데아미나제의 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. 데아미나제 (예를 들어, 아데노신 데아미나제)는 예를 들어 무질서 영역 또는 결정학적 연구에 의해 나타난 바와 같이 고온 인자 또는 B-인자를 포함하는 영역에서 napDNAbp에 삽입될 수 있다. 덜 정렬되거나, 무질서하거나, 구조화되지 않은 단백질 영역, 예를 들어 용매 노출 영역 및 루프는 구조 또는 기능을 손상시키지 않고 삽입을 위해 사용될 수 있다. 데아미나제 (예를 들어, 아데노신 데아미나제)는 가요성 루프 영역 또는 용매-노출된 영역에서 napDNAbp에 삽입될 수 있다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 Cas9 또는 Cas12b/C2c1 폴리펩타이드의 가요성 루프에 삽입된다.

[0642] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)의 삽입 위치는 Cas9 폴리펩타이드의 결정 구조의 B-인자 분석에 의해 결정된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 평균 초과의 B-인자 (예를 들어, 무질서 영역을 포함하는 총 단백질 또는 단백질 도메인과 비교하여 보다 높은 B 인자)를 포함하는 Cas9 폴리펩타이드의 영역에 삽입된다. B-인자 또는 온도 인자는 이들의 평균 위치로부터의 원자의 변동 (예를 들어, 결정 격자에서 온도-의존성 원자 진동 또는 정체 무질서의 결과로서)을 지적할 수 있다. 골격 원자에 대한 높은 B-인자 (예를 들어, 평균 초과의 B-인자)는 상대적으로 높은 국소 이동을 갖는 영역을 지적할 수 있다. 상기 영역은 구조 또는 기능을 손상시키지 않고 데아미나제를 삽입하기 위해 사용될 수 있다. 데아미나제 (예를 들어, 아데노신 데아미나제)는 총 단백질에 대해 평균 B-인자보다 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, 또는 200% 초과인 B-인자와 함께 Cα 원자를 갖는 잔기를 갖는 위치에 삽입될 수 있다. 데아미나제 (예를 들어, 아데노신 데아미나제)는 잔기를 포함하는 Cas9 단백질 도메인에 대한 평균 B-인자보다 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, 또는 200% 초과인 B-인자와 함께 Cα 원자를 갖는 잔기를 갖는 위치에 삽입될 수 있다. 평균 초과의 B-인자를 포함하는 Cas9 폴리펩타이드 위치는 예를 들어, 상기 Cas9 참조 서열에 넘버링된 바와 같이 잔기 768, 792, 1052, 1015, 1022, 1026, 1029, 1067, 1040, 1054, 1068, 1246, 1247, 및 1248을 포함할 수 있다. 평균 초과의 B-인자를 포함하는 Cas9 폴리펩타이드 영역은 예를 들어, 상기 Cas9 참조 서열에 넘버링된 바와 같이 잔기 792-872, 792-906, 및 2-791을 포함할 수 있다.

[0643] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 하기로 이루어진 그룹으로부터 선택된 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에서 napDNAbp에 삽입될 수 있다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, 및 1248. 일부 구현예에서, 이종성 폴리펩타이드는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 위치 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249, 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 구현예에서, 이종성 폴리펩타이드는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 위치 769-770, 792-793, 793-794, 1016-1017, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1053-1054, 1055-1056, 1068-1069, 1069-1070, 1248-1249, 또는 1249-1250, 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 구현예에서, 이종성 폴리펩타이드는 하기로 이루어진 그룹으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체한다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, 및 1248. 삽입 위치와 관련하여, 상기 Cas9 참조 서열에 대한 언급은 설명을 목적으로 하는 것으로 이해되어야만 한다. 본원에 논의된 바와 같은 삽입은 상기 Cas9 참조 서열의 Cas9 폴리펩타이드 서열로 제한되지 않지만, 변이체 Cas9 폴리펩타이드, 예를 들어, Cas9 닉카제 (nCas9), 뉴클레아제 데드 Cas9 (dCas9), 뉴클레아제 도메인이 부재인 Cas9 변이체, 절단된 Cas9, 또는 부분적 또는 완전한 HNH 도메인이 부재인 Cas9 도메인에서 상응하는 위치에 삽입을 포함한다.

[0644] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 하기로 이루어진 그룹으로부터 선택된 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에서 napDNAbp에 삽입될 수 있다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 768, 792, 1022, 1026, 1040, 1068, 및 1247. 일부 구현예에서, 이종성 폴리펩타이드는 상

기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1068-1069, 또는 1247-1248, 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 구현예에서, 이종성 폴리펩타이드는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 위치 769-770, 793-794, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1069-1070, 또는 1248-1249, 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 구현예에서, 이종성 폴리펩타이드는 하기로 이루어진 그룹으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체한다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 768, 792, 1022, 1026, 1040, 1068, 및 1247.

[0645] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 본원에 기재된 바와 같은 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드에서 상응하는 아미노산 잔기에서 napDNAbp에 삽입될 수 있다. 하나의 구현예에서, 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 하기로 이루어진 그룹으로부터 선택된 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드에서 상응하는 아미노산 잔기에서 napDNAbp에 삽입될 수 있다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1002, 1003, 1025, 1052-1056, 1242-1247, 1061-1077, 943-947, 686-691, 569-578, 530-539, 및 1060-1077. 데아미나제 (예를 들어, 아데노신 데아미나제)는 잔기의 N-말단 또는 C-말단에 삽입될 수 있거나 상기 잔기를 대체한다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 잔기의 C-말단에서 삽입된다.

[0646] 일부 구현예에서, 아데노신 데아미나제 (예를 들어, TadA)는 하기로 이루어진 그룹으로부터 선택된 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 11015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246. 일부 구현예에서, 아데노신 데아미나제 (예를 들어, TadA)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 잔기 792-872, 792-906, 또는 2-791의 위치, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기의 위치에 삽입된다. 일부 구현예에서, 아데노신 데아미나제는 하기로 이루어진 그룹으로부터 선택되는 아미노산의 N-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246. 일부 구현예에서, 아데노신 데아미나제는 하기로 이루어진 그룹으로부터 선택되는 아미노산의 C-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246. 일부 구현예에서, 아데노신 데아미나제는 하기로 이루어진 그룹으로부터 선택되는 아미노산, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체한다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246

[0647] 일부 구현예에서, CBE (예를 들어, APOBEC1)는 하기로 이루어진 그룹으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1016, 1023, 1029, 1040, 1069, 및 1247. 일부 구현예에서, ABE는 하기로 이루어진 그룹으로부터 선택되는 아미노산의 N-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1016, 1023, 1029, 1040, 1069, 및 1247. 일부 구현예에서, ABE는 하기로 이루어진 그룹으로부터 선택되는 아미노산의 C-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1016, 1023, 1029, 1040, 1069, 및 1247. 일부 구현예에서, ABE는 하기로 이루어진 그룹으로부터 선택되는 아미노산, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1016, 1023, 1029, 1040, 1069, 및 1247.

[0648] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 768, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 768의 N-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 768의 C-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 768, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0649] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 791에 삽입되거나, 아미노산 잔기 792에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아

미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 791의 N-말단에 삽입되거나, 아미노산 792의 N-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 791의 C-말단에 삽입되거나, 아미노산 792의 N-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 791을 대체하기 위해 삽입되거나, 아미노산 792를 대체하기 위해 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0650] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1016, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1016의 N-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1016의 C-말단에, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1016, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0651] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1022에 삽입되거나, 아미노산 잔기 1023에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1022의 N-말단에 삽입되거나, 아미노산 잔기 1023의 N-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1022의 C-말단에 삽입되거나, 아미노산 잔기 1023의 C-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1022를 대체하기 위해 삽입되거나, 아미노산 잔기 1023을 대체하기 위해 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0652] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1026에 삽입되거나, 아미노산 잔기 1029에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1026의 N-말단에 삽입되거나, 아미노산 잔기 1029의 N-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1026의 C-말단에 삽입되거나, 아미노산 잔기 1029의 C-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1026을 대체하기 위해 삽입되거나, 아미노산 잔기 1029를 대체하기 위해 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0653] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1040, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1040의 N-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1040의 C-말단, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1040, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0654] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1052에 삽입되거나, 아미노산 잔기 1054에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1052의 N-말단에 삽입되거나, 아미노산 잔기 1054의 N-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를

들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1052의 C-말단에 삽입되거나, 아미노산 잔기 1054의 C-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1052를 대체하기 위해 삽입되거나, 아미노산 잔기 1054를 대체하기 위해 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0655] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1067에 삽입되거나, 아미노산 잔기 1068에 삽입되거나, 아미노산 잔기 1069에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1067의 N-말단에 삽입되거나, 아미노산 잔기 1068의 N-말단에 삽입되거나, 아미노산 잔기 1069의 N-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1067의 C-말단에 삽입되거나, 아미노산 잔기 1068의 C-말단에 삽입되거나, 아미노산 잔기 1069의 C-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1067을 대체하기 위해 삽입되거나, 아미노산 잔기 1068을 대체하기 위해 삽입되거나, 아미노산 잔기 1069를 대체하기 위해 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0656] 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1246에 삽입되거나, 아미노산 잔기 1247에 삽입되거나, 아미노산 잔기 1248에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1246의 N-말단에 삽입되거나, 아미노산 잔기 1247의 N-말단에 삽입되거나, 아미노산 잔기 1248의 N-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1246의 C-말단에 삽입되거나, 아미노산 잔기 1247의 C-말단에 삽입되거나, 아미노산 잔기 1248의 C-말단에 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1246을 대체하기 위해 삽입되거나, 아미노산 잔기 1247을 대체하기 위해 삽입되거나, 아미노산 잔기 1248을 대체하기 위해 삽입되거나, 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[0657] 일부 구현예에서, 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 Cas9 폴리펩타이드의 가요성 루프에 삽입된다. 가요성 루프 부분은 상기 Cas9 참조 서열에 넘버링된 바와 같은 530-537, 569-570, 686-691, 943-947, 1002-1025, 1052-1077, 1232-1247, 또는 1298-1300, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기로 이루어진 그룹으로부터 선택될 수 있다. 가요성 루프 부분은 상기 Cas9 참조 서열에 넘버링된 바와 같은 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 또는 1248-1297, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기로 이루어진 그룹으로부터 선택될 수 있다.

[0658] 이종성 폴리펩타이드 (예를 들어, 아데닌 데아미나제)는 하기 아미노산 잔기에 상응하는 Cas9 폴리펩타이드 영역, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 삽입될 수 있다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1017-1069, 1242-1247, 1052-1056, 1060-1077, 1002-1003, 943-947, 530-537, 568-579, 686-691,1242-1247, 1298-1300, 1066-1077, 1052-1056, 또는 1060-1077.

[0659] 이종성 폴리펩타이드 (예를 들어, 아데닌 데아미나제)는 Cas9 폴리펩타이드의 결실 영역의 위치에 삽입될 수 있다. 결실 영역은 Cas9 폴리펩타이드의 N-말단 또는 C-말단 부분에 상응할 수 있다. 일부 구현예에서, 결실 영역은 상기 Cas9 참조 서열에 넘버링된 바와 같은 잔기 792-872, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 상응한다. 일부 구현예에서, 결실 영역은 상기 Cas9 참조 서열에 넘버링된 바와 같은 잔기 792-906, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 상응한다. 일부 구현예에서, 결실 영역은 상기 Cas9 참조 서열에 넘버링된 바와 같은 잔기 2-791, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 상응한다. 일부 구현예에서, 결실 영역은 상기 Cas9 참조 서열에 넘버링된 바와 같은 잔기 1017-1069, 또는 이의 상응하는 아미노산 잔기에 상응한다.

[0660] 예시적인 내부 융합 염기 편집기는 아래 **표 7A**에 제공된다:

[0661] [표 7A]: Cas9 단백질에서 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-환형 퍼뮤턴트 116 ins1067	ISLAY29
IBE031	TadA-환형 퍼뮤턴트 136 ins1248	ISLAY31
IBE032	TadA-환형 퍼뮤턴트 136 ins 1052	ISLAY32
IBE035	델타 792-872 TadA ins	ISLAY35
IBE036	델타 792-906 TadA ins	ISLAY36
IBE043	TadA-환형 퍼뮤턴트 65 ins1246	ISLAY43
IBE044	TadA ins C-말단 절단물 2 791	ISLAY44

[0662]

[0663] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 Cas9 폴리펩타이드의 구조적 또는 기능성 도메인 내 삽입될 수 있다. 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 Cas9 폴리펩타이드의 2개의 구조적 또는 기능성 도메인 사이에 삽입될 수 있다. 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 예를 들어, Cas9 폴리펩타이드로부터 도메인을 결실시킨 후 Cas9 폴리펩타이드의 구조적 또는 기능성 도메인의 위치에 삽입될 수 있다. Cas9 폴리펩타이드의 구조적 또는 기능성 도메인은 예를 들어, RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, 또는 HNH를 포함할 수 있다.

[0664] 일부 구현예에서, Cas9 폴리펩타이드는 하기로 이루어진 그룹으로부터 선택된 하나 이상의 도메인이 부재이다: RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, 또는 HNH 도메인. 일부 구현예에서, Cas9 폴리펩타이드는 뉴클레아제 도메인이 부재이다. 일부 구현예에서, Cas9 폴리펩타이드는 HNH 도메인이 부재이다. 일부 구현예에서, Cas9 폴리펩타이드는 HNH 도메인 부분이 부재여서 Cas9 폴리펩타이드는 감소되거나 폐지된 HNH 활성을 갖는다. 일부 구현예에서, Cas9 폴리펩타이드는 뉴클레아제 도메인의 결실을 포함하고, 데아미나제는 뉴클레아제 도메인을 대체하기 위해 삽입된다. 일부 구현예에서, HNH 도메인은 결실되고 데아미나제는 이의 위치에 삽입된다. 일부 구현예에서, RuvC 도메인의 하나 이상은 결실되고 데아미나제는 이의 위치에 삽입된다.

[0665] 이종성 폴리펩타이드를 포함하는 융합 단백질은 napDNAbp의 N-말단 및 C-말단 단편에 의해 플랭킹될 수 있다. 일부 구현예에서, 융합 단백질은 Cas9 폴리펩타이드의 N-말단 단편 및 C-말단 단편에 의해 플랭킹된 데아미나제를 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 구현예에서, N-말단 단편 및 C-말단 단편의 어느 것도 HNH 도메인을 포함하지 않는다.

[0666] 일부 구현예에서, N-말단 Cas9 단편의 C-말단은 융합 단백질이 표적 핵염기를 탈아민화하는 경우, 표적 핵염기에 인접해 있는 아미노산을 포함한다. 일부 구현예에서, C 말단 Cas9 단편의 N-말단은 융합 단백질이 표적 핵염기를 탈아민화하는 경우, 표적 핵염기에 인접해 있는 아미노산을 포함한다. 상이한 데아미나제의 삽입 위치는 표적 핵염기와, N 말단 Cas9 단편의 C-말단 또는 C 말단 Cas9 단편의 N-말단에서 아미노산 사이가 인접하도록 상이할 수 있다. 예를 들어, ABE의 삽입 위치는 하기로 이루어진 그룹으로부터 선택된 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 있을 수 있다: 상기 Cas9 참조 서열에 넘버링된 바와

같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246.

[0667] 융합 단백질의 N-말단 Cas9 단편 (즉, 융합 단백질에서 데아미나제를 플랭킹하는 N-말단 Cas9 단편)은 Cas9 폴리펩타이드의 N-말단을 포함할 수 있다. 융합 단백질의 N-말단 Cas9 단편은 적어도 약 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 또는 1300개 아미노산 길이를 포함할 수 있다. 융합 단백질의 N-말단 Cas9 단편은 하기의 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 상응하는 서열을 포함할 수 있다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918 또는 1-1100. N-말단 Cas9 단편은 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918 또는 1-1100, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기와 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% 서열 동일성을 포함하는 서열을 포함할 수 있다.

[0668] 융합 단백질의 C-말단 Cas9 단편 (즉, 융합 단백질에서 데아미나제를 플랭킹하는 C-말단 Cas9 단편)은 Cas9 폴리펩타이드의 C-말단을 포함할 수 있다. 융합 단백질의 C-말단 Cas9 단편은 적어도 약 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 또는 1300개 아미노산 길이를 포함할 수 있다. 융합 단백질의 C-말단 Cas9 단편은 하기의 아미노산 잔기, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기에 상응하는 서열을 포함할 수 있다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, 또는 56-1368. N-말단 Cas9 단편은 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, 또는 56-1368, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기와 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% 서열 동일성을 포함하는 서열을 포함할 수 있다.

[0669] 함께 취해진 N-말단 Cas9 단편 및 C-말단 Cas9 단편은 예를 들어, 상기 Cas9 참조 서열에 제시된 바와 같이 전장의 천연적으로 발생하는 Cas9 폴리펩타이드 서열에 상응할 수 없다.

[0670] 본원에 기재된 융합 단백질은 비-표적 부위 (예를 들어, 표적-이탈 부위)에서 감소된 탈아민화와 함께 표적화된 탈아민화, 예를 들어, 감소된 게놈 와이드 스퓨리어스 탈아민화를 수행할 수 있다. 본원에 기재된 융합 단백질은 비-표적 부위에서 감소된 바이스탠더 탈아민화와 함께 표적화된 탈아민화를 수행할 수 있다. 목적하지 않은 탈아민화 또는 표적-이탈 탈아민화는 예를 들어, Cas9 폴리펩타이드의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 말단 융합 단백질과 비교하여 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 또는 적어도 99%까지 감소될 수 있다. 목적하지 않은 탈아민화 또는 표적-이탈 탈아민화는 예를 들어, Cas9 폴리펩타이드의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 말단 융합 단백질과 비교하여 적어도 1배, 적어도 2배, 적어도 3배, 적어도 4배, 적어도 5배, 적어도 10배, 적어도 15배, 적어도 20배, 적어도 30배, 적어도 40배, 적어도 50배, 적어도 60배, 적어도 70배, 적어도 80배, 적어도 90배, 또는 적어도 100배까지 감소될 수 있다.

[0671] 일부 구현예에서, 융합 단백질의 데아미나제 (예를 들어, 아데노신 데아미나제)는 R-루프 범위 내 2개 이하의 핵염기를 탈아민화한다. 일부 구현예에서, 융합 단백질의 데아미나제는 R-루프 범위 내 3개 이하의 핵염기를 탈아민화한다. 일부 구현예에서, 융합 단백질의 데아미나제는 R-루프 범위 내 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이하의 핵염기를 탈아민화한다. R-루프는 DNA:RNA 하이브리드, DNA:DNA 또는 RNA:RNA 상보성 구조 및 단일 가닥의 DNA와 연합된 구조를 포함하는 3개 가닥의 핵산 구조이다. 본원에 사용된 바와 같은 R-루프는 표적 폴리뉴클레오타이드가 CRISPR 복합체 또는 염기 편집 복합체와 접촉되는 경우 형성될 수 있고, 여기서, 가이드 폴리뉴클레오타이드의 일부, 예를 들어. 가이드 RNA는 표적 폴리뉴클레오타이드의 일부, 예를 들어, 표적 DNA와 하이브리드화하고 이로 대체된다. 일부 구현예에서, R-루프는 스페이서 서열 및 표적 DNA 상보성 서열의 하이브리드화된 영역을 포함한다. R-루프 영역은 약 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 50개 핵염기 쌍의 길이일 수 있다. 일부 구현예에서, R-루프 영역은 약 20개 핵염기 쌍의 길이이다. 본원에 사용된 바와 같은 R-루프 영역은 가이드 폴리뉴클레오타이드와 하이브리드화하는 표적 DNA 가닥에 제한되지 않는 것으로 이해되어야만 한다. 예를 들어, R-루프 영역 내 표적 핵염기의 편집은 가이드 RNA에 상보적인 가닥을 포함하는 DNA 가닥에 대한 것일 수 있거나, 가이드 RNA에 상보적인 가닥의 반대 가닥인 DNA 가닥에 대한 것일 수 있다. 일부 구현예에서, R-루프의 영역 내 편집은 표적 DNA 서열에서 가이드 RNA

에 대한 비-상보적 가닥 (프로토스페이서 가닥) 상에 핵염기의 편집을 포함한다.

[0672] 본원에 기재된 융합 단백질은 카노니칼 염기 편집과는 상이한 편집 윈도우에서 표적 탈아민화를 수행할 수 있다. 일부 구현예에서, 표적 핵염기는 표적 폴리뉴클레오타이드 서열 내 PAM 서열의 약 1 내지 약 20개 염기 업스트림에 있다. 일부 구현예에서, 표적 핵염기는 표적 폴리뉴클레오타이드 서열 내 PAM 서열의 약 2 내지 약 12개 염기 업스트림에 있다. 일부 구현예에서, 표적 핵염기는 PAM 서열로부터 약 1 내지 9개 염기 쌍, 약 2 내지 10개 염기 쌍, 약 3 내지 11개 염기 쌍, 약 4 내지 12개 염기 쌍, 약 5 내지 13개 염기 쌍, 약 6 내지 14개 염기 쌍, 약 7 내지 15개 염기 쌍, 약 8 내지 16개 염기 쌍, 약 9 내지 17개 염기 쌍, 약 10 내지 18개 염기 쌍, 약 11 내지 19개 염기 쌍, 약 12 내지 20개 염기 쌍, 약 1 내지 7개 염기 쌍, 약 2 내지 8개 염기 쌍, 약 3 내지 9개 염기 쌍, 약 4 내지 10개 염기 쌍, 약 5 내지 11개 염기 쌍, 약 6 내지 12개 염기 쌍, 약 7 내지 13개 염기 쌍, 약 8 내지 14개 염기 쌍, 약 9 내지 15개 염기 쌍, 약 10 내지 16개 염기 쌍, 약 11 내지 17개 염기 쌍, 약 12 내지 18개 염기 쌍, 약 13 내지 19개 염기 쌍, 약 14 내지 20개 염기 쌍, 약 1 내지 5개 염기 쌍, 약 2 내지 6개 염기 쌍, 약 3 내지 7개 염기 쌍, 약 4 내지 8개 염기 쌍, 약 5 내지 9개 염기 쌍, 약 6 내지 10개 염기 쌍, 약 7 내지 11개 염기 쌍, 약 8 내지 12개 염기 쌍, 약 9 내지 13개 염기 쌍, 약 10 내지 14개 염기 쌍, 약 11 내지 15개 염기 쌍, 약 12 내지 16개 염기 쌍, 약 13 내지 17개 염기 쌍, 약 14 내지 18개 염기 쌍, 약 15 내지 19개 염기 쌍, 약 16 내지 20개 염기 쌍, 약 1 내지 3개 염기 쌍, 약 2 내지 4개 염기 쌍, 약 3 내지 5개 염기 쌍, 약 4 내지 6개 염기 쌍, 약 5 내지 7개 염기 쌍, 약 6 내지 8개 염기 쌍, 약 7 내지 9개 염기 쌍, 약 8 내지 10개 염기 쌍, 약 9 내지 11개 염기 쌍, 약 10 내지 12개 염기 쌍, 약 11 내지 13개 염기 쌍, 약 12 내지 14개 염기 쌍, 약 13 내지 15개 염기 쌍, 약 14 내지 16개 염기 쌍, 약 15 내지 17개 염기 쌍, 약 16 내지 18개 염기 쌍, 약 17 내지 19개 염기 쌍, 약 18 내지 20개 염기 쌍 떨어져 있거나 이의 업스트림에 있다. 일부 구현예에서, 표적 핵염기는 PAM 서열로부터 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 염기 쌍 떨어져 있거나 이의 업스트림에 있다. 일부 구현예에서, 표적 핵염기는 PAM 서열의 약 1, 2, 3, 4, 5, 6, 7, 8, 또는 9개 염기 쌍 업스트림에 있다. 일부 구현예에서, 표적 핵염기는 PAM 서열의 약 2, 3, 4, 또는 6개 염기 쌍 업스트림에 있다.

[0673] 융합 단백질은 하나 초과의 이종성 폴리펩타이드를 포함할 수 있다. 예를 들어, 융합 단백질은 하나 이상의 UGI 도메인 및/또는 하나 이상의 핵 국소화 신호를 추가로 포함할 수 있다. 2개 이상의 이종성 도메인은 탠덤으로 삽입될 수 있다. 2개 이상의 이종성 도메인은 이들이 탠덤으로 NapDNAbp에 있지 않도록 하는 위치에 삽입될 수 있다.

[0674] 융합 단백질은 데아미나제와 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 구현예에서, 융합 단백질은 N-말단 Cas9 단편과 데아미나제 사이에 링커를 포함한다. 일부 구현예에서, 융합 단백질은 C-말단 Cas9 단편과 데아미나제 사이에 링커를 포함한다. 일부 구현예에서, napDNAbp의 N-말단 및 C-말단 단편은 링커를 사용하여 데아미나제에 연결된다. 일부 구현예에서, N-말단 및 C-말단 단편은 링커 없이 데아미나제 도메인에 연결된다. 일부 구현예에서, 융합 단백질은 N-말단 Cas9 단편과 데아미나제 사이에 링커를 포함하지만, C-말단 Cas9 단편과 데아미나제 사이에 링커를 포함하지 않는다. 일부 구현예에서, 융합 단백질은 C-말단 Cas9 단편과 데아미나제 사이에 링커를 포함하지만, N-말단 Cas9 단편과 데아미나제 사이에 링커를 포함하지 않는다.

[0675] 일부 구현예에서, 융합 단백질 내의 napDNAbp는 Cas12 폴리펩티드, 예를 들어, Cas12b/C2c1, 또는 이의 단편이다. Cas12 폴리펩타이드는 변이체 Cas12 폴리펩타이드일 수 있다. 다른 구현예에서, Cas12 폴리펩타이드의 N- 또는 C-말단 단편은 핵산 프로그래밍 가능한 DNA 결합 도메인 또는 RuvC 도메인을 포함한다. 다른 구현예에서, 융합 단백질은 Cas12 폴리펩타이드와 촉매 도메인 사이에 링커를 포함한다. 다른 구현예에서, 링커의 아미노산 서열은 GGSGGS 또는 GSSGSETPGTSESATPESSG이다. 다른 구현예에서, 링커는 강성 링커이다. 상기 양상의 다른 구현예에서, 링커는 GGAGGCTCTGGAGGAAGC 또는 GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC에 의해 암호화되어 있다.

[0676] Cas12 폴리펩타이드의 N- 및 C-말단 단편에 의해 플랭킹된 이종성 촉매 도메인을 포함하는 융합 단백질은 또한 본원에 기재된 바와 같은 방법에서 염기 편집을 위해 유용하다. Cas12 및 하나 이상의 데아미나제 도메인, 예를 들어, 아데노신 데아미나제를 포함하거나 Cas12 서열에 의해 플랭킹된 아데노신 데아미나제를 포함하는 융합 단백질은 또한 표적 서열의 고도의 특이적 및 효율적 염기 편집을 위해 유용하다. 하나의 구현예에서, 키메라 Cas12 융합 단백질은 Cas12 폴리펩타이드 내에 삽입된 이종성 촉매 도메인 (예를 들어, 아데노신 데아미나제)을

포함한다. 일부 구현예에서, 융합 단백질은 Cas12 내에 삽입된 아데노신 데아미나제 도메인 및 아데노신 데아미나제 도메인을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 Cas12 내에서 융합되고 아데노신 데아미나제는 C-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas12 내에서 융합되고 아데노신 데아미나제는 N-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas12 내에서 융합되고 아데노신 데아미나제는 C-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas12 내에서 융합되고 아데노신 데아미나제는 N-말단에 융합된다. 아데노신 데아미나제 및 아데노신 데아미나제와 Cas12와의 융합 단백질의 예시적인 구조는 다음과 같이 제공된다:

[0677] NH_2-[Cas12(아데노신 데아미나제)]-[아데노신 데아미나제]-COOH;

[0678] NH_2-[아데노신 데아미나제]-[Cas12(아데노신 데아미나제)]-COOH;

[0679] NH_2-[Cas12(아데노신 데아미나제)]-[아데노신 데아미나제]-COOH; 또는

[0680] NH_2-[아데노신 데아미나제]-[Cas12(아데노신 데아미나제)]-COOH;

[0681] 일부 구현예에서, 상기 일반 구조에서 사용된 "-"는 임의의 링커의 존재를 지적한다.

[0682] 다양한 구현예에서, 촉매 도메인은 아데노신 데아미나제 활성과 같은 DNA 변형 활성 (예를 들어, 데아미나제 활성)을 갖는다. 일부 구현예에서, 아데노신 데아미나제는 TadA (예를 들어, TadA7.10)이다. 일부 구현예에서, TadA는 TadA*8이다. 일부 구현예에서, TadA*8은 Cas12 내에서 융합되고 아데노신 데아미나제는 C-말단에 융합된다. 일부 구현예에서, TadA*8은 Cas12 내에서 융합되고 아데노신 데아미나제는 N-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas12 내에서 융합되고 TadA*8은 C-말단에 융합된다. 일부 구현예에서, 아데노신 데아미나제는 Cas12 내에서 융합되고 TadA*8은 N-말단에 융합된다. TadA*8 및 시티딘 데아미나제 및 Cas12와의 융합 단백질의 예시적인 구조는 다음과 같이 제공된다:

[0683] N-[Cas12(TadA*8)]-[아데노신 데아미나제]-C;

[0684] N-[아데노신 데아미나제]-[Cas12(TadA*8)]-C;

[0685] N-[Cas12(아데노신 데아미나제)]-[TadA*8]-C; 또는

[0686] N-[TadA*8]-[Cas12(아데노신 데아미나제)]-C.

[0687] 일부 구현예에서, 상기 일반 구조에서 사용된 "-"는 임의의 링커의 존재를 지적한다.

[0688] 다른 구현예에서, 융합 단백질은 하나 이상의 촉매 도메인을 함유한다. 다른 구현예에서, 하나 이상의 촉매 도메인 중 적어도 하나는 Cas12 폴리펩타이드 내에 삽입되거나 Cas12 N-말단 또는 C-말단에 융합된다. 다른 구현예에서, 하나 이상의 촉매 도메인 중 적어도 하나는 Cas12 폴리펩타이드의 루프, 알파 나선 영역, 비구조화된 부분, 또는 용매 접근 가능한 부분 내에 삽입된다. 다른 구현예에서, Cas12 폴리펩타이드는 Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i이다. 다른 구현예에서, Cas12 폴리펩타이드는 바실러스 히사시 Cas12b, 바실러스 써모아밀로보란스 Cas12b, 바실러스 종 *V3-13* Cas12b, 또는 알리사이클로바실러스 액시디필러스 Cas12b와 적어도 약 85% 아미노산 서열 동일성을 갖는다. 다른 구현예에서, Cas12 폴리펩타이드는 바실러스 히사시 Cas12b, 바실러스 써모아밀로보란스 Cas12b, 바실러스 종 *V3-13* Cas12b, 또는 알리사이클로바실러스 액시디필러스 Cas12b와 적어도 약 90% 아미노산 서열 동일성을 갖는다. 다른 구현예에서, Cas12 폴리펩타이드는 바실러스 히사시 Cas12b, 바실러스 써모아밀로보란스 Cas12b, 바실러스 종 *V3-13* Cas12b, 또는 알리사이클로바실러스 액시디필러스 Cas12b와 적어도 약 95% 아미노산 서열 동일성을 갖는다. 다른 구현예에서, Cas12 폴리펩타이드는 바실러스 히사시 Cas12b, 바실러스 써모아밀로보란스 Cas12b, 바실러스 종 *V3-13* Cas12b, 또는 알리사이클로바실러스 액시디필러스 Cas12b의 단편을 포함하거나 필수적으로 이들로 이루어진다.

[0689] 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 위치 153-154, 255-256, 306-307, 980-981, 1019-1020, 534-535, 604-605 또는 344-345, 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h 또는 Cas12i의 상응하는 아미노산 잔기 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 P153과 S154 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 K255와 E256 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 D980과 G981 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 K1019와 L1020 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 F534와 P535 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 K604와 G605 사이에 삽입

된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 H344와 F345 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BvCas12b의 아미노산 위치 147과 148, 248과 249, 299와 300, 991과 992, 또는 1031과 1032 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i의 상응하는 아미노산 잔기 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BvCas12b의 아미노산 P147과 D148 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BvCas12b의 아미노산 G248과 G249 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BvCas12b의 아미노산 P299와 E300 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BvCas12b의 아미노산 G991과 E992 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BvCas12b의 아미노산 K1031과 M1032 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 AaCas12b의 아미노산 위치 157과 158, 258과 259, 310과 311, 1008과 1009, 또는 1044와 1045 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i의 상응하는 아미노산 잔기 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 AaCas12b의 아미노산 P157과 G158 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 AaCas12b의 아미노산 V258과 G259 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 AaCas12b의 아미노산 D310과 P311 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 AaCas12b의 아미노산 G1008과 E1009 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 AaCas12b의 아미노산 G1044와 K1045 사이에 삽입된다.

[0690] 다른 구현예에서, 융합 단백질은 핵 국소화 신호 (예를 들어, 양분된 핵 국소화 신호)를 포함한다. 다른 구현예에서, 핵 국소화 신호의 아미노산 서열은 MAPKKKRKVGIHGVPAA이다. 상기 양상의 다른 구현예에서, 핵 국소화 신호는 하기의 서열에 의해 암호화된다:

[0691] ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC. 다른 구현예에서, Cas12b 폴리펩타이드는 RuvC 도메인의 촉매 활성을 사일런싱시키는 돌연변이를 함유한다. 다른 구현예에서, Cas12b 폴리펩타이드는 D574A, D829A 및/또는 D952A 돌연변이를 함유한다. 다른 구현예에서, 융합 단백질은 태그 (예를 들어, 인플루엔자 헤마글루티닌 태그)를 추가로 함유한다.

[0692] 일부 구현예에서, 융합 단백질은 내부적으로 융합된 핵염기 편집 도메인 (예를 들어, 데아미나제 도메인, 예를 들어, 아데노신 데아미나제 도메인 전부 또는 일부)과 함께 napDNAbp 도메인 (예를 들어, Cas12-유래된 도메인)을 포함한다. 일부 구현예에서, napDNAbp는 Cas12b이다. 일부 구현예에서, 염기 편집기는 아래 **표 7B**에 제공된 유전자좌에 삽입된 내부적으로 융합된 TadA*8 도메인과 함께 BhCas12b 도메인을 포함한다.

[0693] [표 7B] Cas12b 단백질에서 삽입 유전자좌

BhCas12b	삽입 부위	aa 사이에 삽입된
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 삽입된**
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 삽입된**
위치 1	157	PG
위치 2	258	VG
위치 3	310	DP
위치 4	1008	GE
위치 5	1044	GK

[0694]

[0695] 비제한적인 예로, 아데노신 데아미나제 (예를 들어, ABE8.13)는 BhCas12b에 삽입하여 핵산 서열을 효과적으로 편집하는 융합 단백질 (예를 들어, ABE8.13-BhCas12b)을 생성할 수 있다.

[0696] 일부 구현예에서, 본원에 기재된 염기 편집 시스템은 Cas9 내에 삽입된 TadA를 갖는 ABE를 포함한다. Cas9 내에 TadA가 삽입된 관련 ABE의 서열이 제공된다.

[0697] 101 Cas9 TadAins 1015

[0698] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0699] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0700] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0701] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0702] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0703] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0704] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0705] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0706] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0707] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0708] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0709] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0710] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0711] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0712] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[0713] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[0714] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0715] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[0716] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0717] YPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHAL

[0718] TLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQG

[0719] GLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGS

[0720] LMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSST

[0721] DYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[0722] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[0723] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[0724] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0725] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[0726] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[0727] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0728] LGGD

[0729] 102 Cas9 TadAins 1022

[0730] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0731] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0732] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0733] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0734] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0735] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0736] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0737] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0738] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0739] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0740] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0741] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0742] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0743] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0744] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0745] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[0746] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[0747] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0748] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[0749] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0750] YPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESSGSEVEFSHE

[0751] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[0752] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[0753] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[0754] KKAQSSTDAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[0755] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[0756] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[0757] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0758] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[0759] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[0760] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0761] LGGD

[0762] 103 Cas9 TadAins 1029

[0763] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0764] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0765] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0766] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0767] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0768] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0769] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0770] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0771] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0772] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0773] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0774] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0775] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0776] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0777] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0778] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[0779] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[0780] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0781] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[0782] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0783] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGSSGSETPGTSESATPESSGS

[0784] EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH

[0785] DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV

[0786] VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMP

[0787] RQVFNAQKKAQSSTDGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[0788] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[0789] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[0790] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0791] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[0792] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[0793] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0794] LGGD

[0795] 103 Cas9 TadAins 1040

[0796] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0797] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0798] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0799] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0800] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0801] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0802] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0803] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0804] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0805] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0806] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0807] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0808] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0809] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0810] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0811] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[0812] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[0813] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0814] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[0815] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0816] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSGSSGSETPGT

[0817] SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI

[0818] GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA

[0819] GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC

[0820] AALLCYFFRMPRQVFNAQKKAQSSTDNIMNFFKTEITLANGEIRKRPLIE

[0821] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[0822] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[0823] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0824] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[0825] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[0826] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0827] LGGD

[0828] 105 Cas9 TadAins 1068

[0829] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0830] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0831] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0832] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0833] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0834] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0835] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0836] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0837] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0838] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0839] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0840] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0841] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0842] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0843] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0844] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[0845] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[0846] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0847] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[0848] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0849] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[0850] TLANGEIRKRPLIETNGEGSSGSETPGTSESATPESSGSEVEFSHEYWMR

[0851] HALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL

[0852] RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA

[0853] AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQ

[0854] SSTDTGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[0855] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[0856] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0857] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[0858] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[0859] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0860] LGGD

[0861] 106 Cas9 TadAins 1247

[0862] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0863] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0864] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0865] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0866] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0867] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0868] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0869] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0870] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0871] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0872] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0873] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0874] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0875] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0876] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0877] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[0878] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[0879] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0880] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[0881] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0882] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[0883] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[0884] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[0885] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[0886] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGGSS

[0887] GSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL

[0888] VLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTF

[0889] EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE

[0890] GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQKQLFVEQHKH

[0891] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[0892] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0893] LGGD

[0894] 107 Cas9 TadAins 1054

[0895] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0896] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0897] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0898] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0899] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0900] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0901] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0902] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0903] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0904] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0905] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0906] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0907] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0908] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0909] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0910] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[0911] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[0912] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0913] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[0914] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0915] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[0916] TLANGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE

[0917] VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID

[0918] ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN

[0919] HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIRKRPLIE

[0920] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[0921] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[0922] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0923] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[0924] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[0925] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0926] LGGD

[0927] 108 Cas9 TadAins 1026

[0928] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0929] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0930] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0931] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0932] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0933] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0934] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0935] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0936] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0937] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0938] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0939] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0940] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0941] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0942] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0943] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[0944] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[0945] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[0946] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[0947] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0948] YPKLESEFVYGDYKVYDVRKMIAKSEGSSGSETPGTSESATPESSGSEVE

[0949] FSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT

[0950] AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG

[0951] VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQV

[0952] FNAQKKAQSSTDQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[0953] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[0954] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[0955] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0956] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[0957] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[0958] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0959] LGGD

[0960] 109 Cas9 TadAins 768

[0961] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0962] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0963] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0964] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0965] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0966] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0967] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[0968] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[0969] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[0970] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[0971] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[0972] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0973] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0974] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[0975] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[0976] MGRHKPENIVIEMARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMR

[0977] HALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL

[0978] RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA

[0979] AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRTTQKGQKNSR

[0980] ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL

[0981] DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK

[0982] MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT

[0983] KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI

[0984] NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE

[0985] IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR

[0986] DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP

[0987] KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP

[0988] IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL

[0989] PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK

[0990] RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD

[0991] TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0992] 110.1 Cas9 TadAins 1250

[0993] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0994] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[0995] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[0996] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[0997] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[0998] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[0999] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1000] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1001] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1002] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1003] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1004] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1005] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1006] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1007] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1008] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1009] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1010] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1011] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1012] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1013] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1014] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1015] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1016] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1017] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG

[1018] SSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA

[1019] VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV

[1020] TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI

[1021] TEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSK

[1022] RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD

[1023] TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1024] 110.2 Cas9 TadAins 1250

[1025] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1026] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1027] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1028] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1029] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1030] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1031] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1032] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1033] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1034] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1035] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1036] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1037] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1038] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1039] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1040] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1041] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1042] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1043] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1044] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1045] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1046] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1047] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1048] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1049] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG

[1050] SSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP

[1051] VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT

[1052] LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR

[1053] VEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISE

[1054] FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK

[1055] YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1056] 110.3 Cas9 TadAins 1250

[1057] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1058] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1059] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1060] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1061] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1062] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1063] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1064] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1065] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1066] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1067] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1068] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1069] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1070] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1071] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1072] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1073] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1074] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1075] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1076] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1077] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1078] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1079] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1080] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1081] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG

[1082] SSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER

[1083] EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI

[1084] DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM

[1085] NHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQ

[1086] ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA

[1087] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1088] 110.4 Cas9 TadAins 1250

[1089] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1090] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1091] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1092] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1093] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1094] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1095] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1096] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1097] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1098] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1099] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1100] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1101] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1102] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1103] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1104] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1105] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1106] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1107] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1108] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1109] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1110] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1111] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1112] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1113] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG

[1114] SSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER

[1115] EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI

[1116] DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM

[1117] NHRVEITEGILADECAALLCYFFRMRREDNEQKQLFVEQHKHYLDEIIEQ

[1118] ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA

[1119] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1120] 110.5 Cas9 TadAins 1249

[1121] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1122] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1123] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1124] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1125] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1126] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1127] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1128] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1129] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1130] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1131] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1132] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1133] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1134] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1135] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1136] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1137] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1138] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1139] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1140] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1141] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1142] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1143] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1144] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1145] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSGS

[1146] SGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE

[1147] VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID

[1148] ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN

[1149] HRVEITEGILADECAALLCYFFRMRRPEDNEQKQLFVEQHKHYLDEIIEQ

[1150] ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA

[1151] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1152] 110.5 Cas9 TadAins 델타 59-66 1250

[1153] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1154] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1155] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1156] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1157] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1158] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1159] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1160] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1161] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1162] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1163] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1164] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1165] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1166] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1167] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1168] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1169] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1170] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1171] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1172] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1173] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1174] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1175] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1176] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1177] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG

[1178] SSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDERE

[1179] VPVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFE

[1180] PCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEG

[1181] ILADECAALLCYFFRMPRQVFNAQKKAQSSTDEDNEQKQLFVEQHKHYLD

[1182] EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN

[1183] LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG

[1184] D

[1185] 110.6 Cas9 TadAins 1251

[1186] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1187] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1188] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1189] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1190] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1191] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1192] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1193] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1194] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1195] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1196] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1197] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1198] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1199] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1200] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1201] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1202] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1203] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1204] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1205] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1206] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1207] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1208] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1209] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1210] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE

[1211] GSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDE

[1212] REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL

[1213] IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[1214] MNHRVEITEGILADECAALLCYFFRMRRDNEQKQLFVEQHKHYLDEIIEQ

[1215] ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA

[1216] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1217] 110.7 Cas9 TadAins 1252

[1218] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1219] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1220] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1221] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1222] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1223] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1224] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1225] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1226] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1227] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1228] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1229] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1230] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1231] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1232] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1233] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1234] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1235] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1236] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1237] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1238] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1239] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1240] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1241] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1242] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE

[1243] DGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARD

[1244] EREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR

[1245] LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

[1246] GMNHRVEITEGILADECAALLCYFFRMRRNEQKQLFVEQHKHYLDEIIEQ

[1247] ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA

[1248] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1249] 110.8 Cas9 TadAins 델타 59-66 C-절단물 1250

[1250] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1251] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1252] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1253] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1254] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1255] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1256] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1257] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1258] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1259] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1260] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1261] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1262] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1263] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1264] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1265] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1266] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1267] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1268] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1269] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1270] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1271] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1272] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1273] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1274] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG

[1275] SSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA

[1276] VLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC

[1277] AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE

[1278] CAALLCYFFRMPRQEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA

[1279] NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR

[1280] YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1281] 111.1 Cas9 TadAins 997

[1282] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1283] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1284] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1285] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1286] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1287] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1288] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1289] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1290] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1291] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1292] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1293] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1294] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1295] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1296] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1297] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1298] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1299] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1300] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1301] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALSHE

[1302] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[1303] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[1304] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[1305] KKAQSSTDGSSGSETPGTSESATPESSGIKKYPKLESEFVYGDYKVYDVR

[1306] KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET

[1307] GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL

[1308] IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM

[1309] ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE

[1310] LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI

[1311] IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG

[1312] APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1313] 111.2 Cas9 TadAins 997

[1314] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1315] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1316] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1317] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1318] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1319] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1320] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1321] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1322] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1323] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1324] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1325] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1326] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1327] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1328] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1329] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1330] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1331] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1332] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1333] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALSHE

[1334] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[1335] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[1336] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[1337] KKAQSSTDGSSGSSGSETPGTSESATPESSGGSSIKKYPKLESEFVYGDY

[1338] KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI

[1339] ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK

[1340] RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL

[1341] LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM

[1342] LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK

[1343] HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF

[1344] TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS

[1345] QLGGD

[1346] 112 델타 HNH TadA

[1347] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1348] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1349] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1350] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1351] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1352] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1353] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1354] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1355] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1356] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1357] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1358] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1359] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1360] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1361] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1362] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE

[1363] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[1364] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[1365] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[1366] KKAQSSTDGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND

[1367] KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL

[1368] IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK

[1369] TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK

[1370] TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA

[1371] KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK

[1372] LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG

[1373] SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH

[1374] RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL

[1375] IHQSITGLYETRIDLSQLGGD

[1376] 113 N-말단 단일 TadA 나선 절단물 165-말단

[1377] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG

[1378] LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG

[1379] RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR

[1380] MPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSV

[1381] GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR

[1382] TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH

[1383] PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG

[1384] HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL

[1385] SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL

[1386] SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP

[1387] LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG

[1388] ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL

[1389] GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT

[1390] RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE

[1391] YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE

[1392] DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL

[1393] EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL

[1394] INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ

[1395] GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE

[1396] NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL

[1397] QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK

[1398] SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF

[1399] IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDF

[1400] RKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVY

[1401] DVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN

[1402] GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS

[1403] DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI

[1404] TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS

[1405] AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL

[1406] DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT

[1407] NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG

[1408] GD

[1409] 114 N-말단 단일 TadA 나선 절단물 165-말단 델타 59-65

[1410] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRTAH

[1411] AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR

[1412] NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRSGGS

[1413] SGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITD

[1414] EYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT

[1415] RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV

[1416] DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD

[1417] LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE

[1418] NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD

[1419] DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK

[1420] RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY

[1421] KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL

[1422] RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI

[1423] TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE

[1424] LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE

[1425] CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL

[1426] TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

[1427] QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH

[1428] IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG

[1429] QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY

[1430] VDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE

[1431] EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE

[1432] TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY

[1433] KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA

[1434] KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV

[1435] WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK

[1436] KDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS

[1437] FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG

[1438] NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI

[1439] SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA

[1440] FKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1441] 115.1 Cas9 TadAins1004

[1442] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1443] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1444] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1445] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1446] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1447] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1448] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1449] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1450] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1451] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1452] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1453] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1454] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1455] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1456] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1457] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1458] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1459] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1460] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1461] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1462] YPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV

[1463] PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDA

[1464] TLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH

[1465] RVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIAKSEQ

[1466] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[1467] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[1468] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[1469] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[1470] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[1471] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[1472] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1473] 115.2 Cas9 TadAins1005

[1474] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1475] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1476] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1477] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1478] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1479] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1480] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1481] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1482] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1483] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1484] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1485] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1486] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1487] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1488] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1489] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1490] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1491] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1492] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1493] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1494] YPKLGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE

[1495] VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID

[1496] ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN

[1497] HRVEITEGILADECAALLCYFFRMPRQESEFVYGDYKVYDVRKMIAKSEQ

[1498] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[1499] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[1500] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[1501] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[1502] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[1503] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[1504] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1505] 115.3 Cas9 TadAins1006

[1506] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1507] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1508] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1509] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1510] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1511] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1512] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1513] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1514] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1515] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1516] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1517] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1518] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1519] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1520] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1521] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1522] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1523] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1524] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1525] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1526] YPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDER

[1527] EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI

[1528] DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM

[1529] NHRVEITEGILADECAALLCYFFRMPRQSEFVYGDYKVYDVRKMIAKSEQ

[1530] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[1531] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[1532] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[1533] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[1534] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[1535] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[1536] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1537] 115.4 Cas9 TadAins1007

[1538] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1539] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1540] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1541] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1542] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1543] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1544] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1545] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1546] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1547] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1548] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1549] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1550] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1551] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1552] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1553] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1554] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1555] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1556] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1557] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1558] YPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDE

[1559] REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL

[1560] IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[1561] MNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIAKSEQ

[1562] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[1563] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[1564] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[1565] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[1566] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[1567] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[1568] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1569] 116.1 Cas9 TadAins C-말단 절단물2 792

[1570] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1571] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1572] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1573] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1574] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1575] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1576] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1577] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1578] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1579] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1580] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1581] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1582] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1583] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1584] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1585] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGGSSGSETP

[1586] GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR

[1587] VIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM

[1588] CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD

[1589] ECAALLCYFFRMPRQSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE

[1590] LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK

[1591] KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI

[1592] TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE

[1593] INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ

[1594] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[1595] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[1596] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[1597] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[1598] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[1599] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[1600] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1601] 116.2 Cas9 TadAins C-말단 절단물2 791

[1602] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1603] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1604] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1605] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1606] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1607] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1608] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1609] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1610] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1611] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1612] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1613] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1614] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1615] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1616] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1617] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPG

[1618] TSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRV

[1619] IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC

[1620] AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE

[1621] CAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE

[1622] LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK

[1623] KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI

[1624] TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE

[1625] INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ

[1626] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[1627] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[1628] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[1629] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[1630] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[1631] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[1632] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1633] 116.3 Cas9 TadAins C-말단 절단물2 790

[1634] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1635] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1636] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1637] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1638] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1639] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1640] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1641] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1642] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1643] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1644] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1645] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1646] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1647] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1648] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1649] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEGSSGSETPGT

[1650] SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI

[1651] GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA

[1652] GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC

[1653] AALLCYFFRMPRQLGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE

[1654] LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK

[1655] KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI

[1656] TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE

[1657] INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ

[1658] EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG

[1659] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

[1660] PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN

[1661] PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA

[1662] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS

[1663] KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[1664] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1665] 117 Cas9 델타 1017-1069

[1666] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1667] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1668] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1669] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1670] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1671] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1672] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1673] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1674] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1675] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1676] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1677] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1678] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1679] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1680] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1681] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1682] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1683] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1684] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1685] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1686] YPKLESEFVYGDYKVYSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA

[1687] VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV

[1688] TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI

[1689] TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIVWDKGRDFATVR

[1690] KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF

[1691] DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA

[1692] KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVN

[1693] FLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD

[1694] ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK

[1695] RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1696] 118 Cas9 TadA-CP116ins 1067

[1697] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1698] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1699] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1700] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1701] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1702] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1703] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1704] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1705] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1706] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1707] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1708] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1709] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1710] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1711] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1712] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1713] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1714] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1715] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1716] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1717] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1718] TLANGEIRKRPLIETNMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[1719] KKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRAR

[1720] DEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY

[1721] RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY

[1722] PGGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[1723] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[1724] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[1725] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[1726] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[1727] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1728] LGGD

[1729] 119 Cas9 TadAins 701

[1730] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1731] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1732] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1733] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1734] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1735] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1736] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1737] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1738] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1739] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1740] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1741] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1742] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1743] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1744] SGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV

[1745] GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATL

[1746] YVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV

[1747] EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDLTFKEDIQKAQVS

[1748] GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMA

[1749] RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY

[1750] YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR

[1751] GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA

[1752] GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS

[1753] DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK

[1754] VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[1755] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[1756] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[1757] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[1758] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[1759] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[1760] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1761] LGGD

[1762] 120 Cas9 TadACP136ins 1248

[1763] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1764] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1765] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1766] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1767] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1768] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1769] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1770] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1771] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1772] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1773] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1774] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1775] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1776] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1777] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1778] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1779] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1780] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1781] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1782] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1783] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1784] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1785] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1786] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1787] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSMN

[1788] HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT

[1789] SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI

[1790] GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA

[1791] GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKH

[1792] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[1793] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1794] LGGD

[1795] 121 Cas9 TadACP136ins 1052

[1796] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1797] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1798] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1799] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1800] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1801] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1802] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1803] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1804] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1805] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1806] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1807] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1808] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1809] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1810] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1811] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1812] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1813] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1814] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1815] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1816] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1817] TLAMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS

[1818] ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL

[1819] NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEP

[1820] CVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGNGEIRKRPLIE

[1821] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[1822] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[1823] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[1824] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[1825] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[1826] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1827] LGGD

[1828] 122 Cas9 TadACP136ins 1041

[1829] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1830] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1831] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1832] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1833] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1834] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1835] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1836] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1837] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1838] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1839] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1840] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1841] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1842] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1843] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1844] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1845] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1846] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1847] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1848] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1849] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSMNHRVEITEG

[1850] ILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES

[1851] SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAI

[1852] GLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI

[1853] GRVVFGVRNAKTGAAGSLMDVLHYPGNIMNFFKTEITLANGEIRKRPLIE

[1854] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[1855] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[1856] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[1857] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[1858] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[1859] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1860] LGGD

[1861] 123 Cas9 TadACP139ins 1299

[1862] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1863] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1864] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1865] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1866] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1867] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1868] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1869] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1870] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1871] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1872] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1873] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1874] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1875] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1876] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1877] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1878] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1879] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1880] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1881] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1882] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[1883] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[1884] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[1885] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[1886] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE

[1887] DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRMN

[1888] HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT

[1889] SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI

[1890] GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA

[1891] GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGDKPIREQAENIIHLFT

[1892] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1893] LGGD

[1894] 124 Cas9 델타 792-872 TadAins

[1895] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1896] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1897] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1898] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1899] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1900] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1901] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1902] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1903] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1904] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1905] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1906] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1907] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1908] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1909] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1910] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE

[1911] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[1912] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[1913] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[1914] KKAQSSTDEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA

[1915] GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS

[1916] DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK

[1917] VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[1918] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[1919] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[1920] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[1921] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[1922] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[1923] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1924] LGGD

[1925] 125 Cas9 델타 792-906 TadAins

[1926] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1927] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1928] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1929] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1930] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1931] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1932] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1933] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1934] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1935] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1936] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1937] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1938] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1939] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1940] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1941] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE

[1942] YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE

[1943] IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA

[1944] KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ

[1945] KKAQSSTDGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK

[1946] LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI

[1947] KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT

[1948] EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT

[1949] EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK

[1950] VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL

[1951] PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS

[1952] PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR

[1953] DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI

[1954] HQSITGLYETRIDLSQLGGD

[1955] 126 TadA CP65ins 1003

[1956] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1957] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1958] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1959] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1960] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1961] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1962] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1963] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1964] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1965] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1966] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[1967] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[1968] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[1969] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[1970] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[1971] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[1972] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[1973] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[1974] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[1975] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[1976] YPKTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGR

[1977] VVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM

[1978] PRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHA

[1979] LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPLESEFVYGDYK

[1980] VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[1981] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[1982] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[1983] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[1984] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[1985] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[1986] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[1987] LGGD

[1988] 127 TadA CP65ins 1016

[1989] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[1990] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[1991] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[1992] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[1993] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[1994] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[1995] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[1996] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[1997] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[1998] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[1999] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2000] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2001] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2002] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2003] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2004] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2005] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2006] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2007] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2008] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2009] YPKLESEFVYGDYKVTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM

[2010] CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD

[2011] ECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSE

[2012] VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD

[2013] PYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[2014] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2015] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2016] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2017] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2018] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2019] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2020] LGGD

[2021] 128 TadA CP65ins 1022

[2022] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2023] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2024] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2025] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2026] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2027] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2028] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2029] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2030] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2031] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2032] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2033] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2034] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2035] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2036] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2037] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2038] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2039] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2040] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2041] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2042] YPKLESEFVYGDYKVYDVRKMITAHAEIMALRQGGLVMQNYRLIDATLYV

[2043] TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI

[2044] TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESAT

[2045] PESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN

[2046] RAIGLHDPAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[2047] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2048] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2049] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2050] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2051] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2052] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2053] LGGD

[2054] 129 TadA CP65ins 1029

[2055] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2056] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2057] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2058] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2059] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2060] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2061] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2062] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2063] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2064] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2065] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2066] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2067] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2068] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2069] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2070] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2071] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2072] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2073] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2074] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2075] YPKLESEFVYGDYKVYDVRKMIAKSEQEITAHAEIMALRQGGLVMQNYRL

[2076] IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[2077] MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETP

[2078] GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR

[2079] VIGEGWNRAIGLHDPGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE

[2080] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2081] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2082] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2083] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2084] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2085] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2086] LGGD

[2087] 130 TadA CP65ins 1041

[2088] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2089] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2090] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2091] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2092] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2093] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2094] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2095] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2096] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2097] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2098] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2099] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2100] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2101] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2102] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2103] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2104] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2105] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2106] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2107] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2108] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSTAHAEIMALR

[2109] QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAA

[2110] GSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQS

[2111] STDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV

[2112] PVGAVLVLNNRVIGEGWNRAIGLHDPNIMNFFKTEITLANGEIRKRPLIE

[2113] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2114] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2115] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2116] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2117] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2118] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2119] LGGD

[2120] 131 TadA CP65ins 1054

[2121] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2122] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2123] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2124] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2125] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2126] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2127] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2128] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2129] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2130] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2131] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2132] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2133] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2134] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2135] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2136] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2137] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2138] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2139] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2140] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2141] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[2142] TLANTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG

[2143] RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR

[2144] MPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRH

[2145] ALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPGEIRKRPLIE

[2146] TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR

[2147] NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL

[2148] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[2149] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH

[2150] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2151] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2152] LGGD

[2153] 132 TadA CP65ins 1246

[2154] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[2155] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

[2156] LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD

[2157] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

[2158] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

[2159] NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

[2160] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

[2161] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

[2162] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

[2163] YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

[2164] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

[2165] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[2166] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[2167] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

[2168] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

[2169] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

[2170] VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

[2171] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

[2172] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

[2173] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[2174] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

[2175] TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

[2176] QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

[2177] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

[2178] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGTAH

[2179] AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR

[2180] NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFN

[2181] AQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKR

[2182] ARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKH

[2183] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

[2184] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[2185] LGGD

[2186] 일부 구현예에서, 동정된 위치에서 Cas9 폴리펩타이드 내로 TadA 또는 이의 변이체를 삽입하기 위해 아데노신 데아미나제 염기 편집기가 생성되었다.

[2187] 비제한적이지만 예시적인 융합 단백질은 국제 PCT 출원 번호 PCT/US2020/016285 및 미국 가출원 번호 62/852,228 및 62/852,224에 기재되어 있고, 이의 내용은 이들의 전문이 본원에 참조로 포함된다.

[2188] ***핵염기 편집 도메인***

[2189] 본원에는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 핵염기 편집 도메인 (예를 들어, 데아미나제 도메인)을 포함하는 융합 단백질을 포함하는 염기 편집기가 기재되어 있다. 염기 편집기는 표적 서열을 인지할 수 있는 가이드 폴리뉴클레오타이드와 상호작용함으로써 표적 폴리뉴클레오타이드 서열에서 하나 이상의 염기를 편집하기 위해 프로그래밍될 수 있다. 표적 서열이 인지되면, 염기 편집기는 편집이 발생하는 폴리뉴클레오타이드 상에 부착되고 염기 편집기의 데아미나제 도메인 성분은 이어서 표적 염기를 편집할 수 있다.

[2190] 일부 구현예에서, 핵염기 편집 도메인은 데아미나제 도메인을 포함한다. 일부 구현예에서, 데아미나제 도메인은 아데닌 데아미나제 또는 아데노신 데아미나제일 수 있다. 일부 구현예에서, 용어 "아데닌 데아미나제" 및 "아데노신 데아미나제"는 상호교환적으로 사용될 수 있다. 핵염기 편집 단백질의 세부사항은 국제 PCT 출원 번호 PCT/2017/045381 (WO 2018/027078) 및 PCT/US2016/058344 (WO 2017/070632)에 기재되어 있고, 이의 각각은 이의 전문이 본원에 참조로 포함된다. 또한 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); and 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))을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[2191] ***A에서 G로의 편집***

[2192] 일부 구현예에서, 본원에 기재된 염기 편집기는 아데노신 데아미나제를 포함하는 데아미나제 도메인을 포함할 수 있다. 염기 편집기의 상기 아데노신 데아미나제 도메인은 아데닌 (A)을 탈아민화하여 구아닌 (G)의 염기 쌍 형성 성질을 나타내는 이노신 (I)을 형성함으로써 아데닌 (A) 핵염기의 구아닌 (G) 핵염기로의 편집을 촉진시킬 수 있다. 아데노신 데아미나제는 데옥시리보핵산 (DNA)에서 데옥시아데노신 잔기의 아데닌을 탈아민화 (즉, 아민 그룹을 제거하는)할 수 있다.

[2193] 일부 구현예에서, 본원에 제공된 핵염기 편집기는 하나 이상의 단백질 도메인을 함께 융합하여 융합 단백질을 생성시킴으로써 제조될 수 있다. 특정 구현예에서, 본원에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성 (예를 들어, 효율, 선택성 및 특이성)을 개선시키는 하나 이상의 특성을 포함한다. 예를 들어, 본원에 제공된 융합 단백질은 감소된 뉴클레아제 활성을 갖는 Cas9 도메인을 포함할 수 있다. 일부 구현예에서, 본원에 제공된 융합 단백질은 뉴클레아제 활성을 갖지 않는 Cas9 도메인 (dCas9), 또는 듀플렉스 DNA 분자의 하나의 가닥을 절단하는, Cas9 닉카제로서 언급되는 Cas9 도메인 (nCas9)을 가질 수 있다. 특정 이론에 국한시키고자 하는 것 없이, 촉매 잔기 (예를 들어, H840)의 존재는 표적화된 A의 반대편의 T를 함유하는 비편집된 (예를 들어, 비-탈아민화된) 가닥을 절단하는 Cas9의 활성을 유지한다. Cas9의 촉매 잔기의 돌연변이 (예를 들어, D10에서 A10으로)는 표적화된 A 잔기를 함유하는 편집된 가닥의 절단을 방지한다. 상기 Cas9 변이체는 gRNA-한정된 표적 서열을 기준으로 특정 위치에서 단일 가닥 DNA 절단(닉)을 생성하여 비-편집된 가닥의 복구를 유도하고 궁극적으로 비-편집된 가닥 상에 T에서 C로의 변화를 유도한다. 일부 구현예에서, A에서 G로의 염기 편집기는 이노신 염기 절제 복구의 저해제, 예를 들어, 우라실 글리코실라제 저해제 (UGI) 도메인 또는 촉매적 불활성 이노신 특이적 뉴클레아제를 추가로 포함한다. 임의의 특정 이론에 국한시키고자 하는 것 없이, UGI 도메인 또는 촉매적 불활성 이노신 특이적 뉴클레아제는 탈아민화된 아데노신 잔기 (예를 들어, 이노신)의 염기 절제 복구를 저해하거나 방지할 수 있고, 이는 염기 편집기의 활성 또는 효율을 개선시킬 수 있다.

[2194] 아데노신 데아미나제를 포함하는 염기 편집기는 DNA, RNA 및 DNA-RNA 하이브리드를 포함하는, 임의의 폴리뉴클레오타이드 상에 작용할 수 있다. 특정 구현예에서, 아데노신 데아미나제를 포함하는 염기 편집기는 RNA를 포

함하는 폴리뉴클레오타이드의 표적 A를 탈아민화할 수 있다. 예를 들어, 염기 편집기는 RNA 폴리뉴클레오타이드 및/또는 DNA-RNA 하이브리드 폴리뉴클레오타이드의 표적 A를 탈아민화할 수 있는 아데노신 데아미나제 도메인을 포함할 수 있다. 하나의 구현예에서, 염기 편집기에 혼입된 아데노신 데아미나제는 RNA (ADAR, 예를 들어, ADAR1 또는 ADAR2)에 작용하는 아데노신 데아미나제의 전부 또는 일부를 포함한다. 또 다른 구현예에서, 염기 편집기에 혼입된 아데노신 데아미나제는 tRNA (ADAT)에 작용하는 아데노신 데아미나제의 전부 또는 일부를 포함한다. 아데노신 데아미나제 도메인을 포함하는 염기 편집기는 또한 DNA 폴리뉴클레오타이드의 A 핵염기를 탈아민화할 수 있다. 하나의 구현예에서, 염기 편집기의 아데노신 데아미나제 도메인은 ADAT가 DNA 내 표적 A를 탈아민화하도록 하는 하나 이상의 돌연변이를 포함하는 ADAT의 전부 또는 일부를 포함한다. 예를 들어, 염기 편집기는 하기의 돌연변이 중 하나 이상을 포함하는 에스케리치아 콜리 (*Escherichia coli*) (EcTadA)로부터의 ADAT의 전부 또는 일부를 포함할 수 있다: 또 다른 아데노신 데아미나제에서 D108N, A106V, D147Y, E155V, L84F, H123Y, I156F, 또는 상응하는 돌연변이.

[2195] 아데노신 데아미나제는 임의의 적합한 유기체 (예를 들어, 이. 콜리)로부터 유래할 수 있다. 일부 구현예에서, 아데노신 데아미나제는 원핵생물 기원이다. 일부 구현예에서, 아데노신 데아미나제는 세균 기원이다. 일부 구현예에서, 아데노신 데아미나제는 에스케리치아 콜리, 스타필로코커스 아우레우스, 살모넬라 타이피, 쉐와넬라 푸트레파시엔스, 해모필러스 인플루엔자, 콜로박터 크레슨투스, 또는 바실러스 서브틸리스 기원이다. 일부 구현예에서, 아데노신 데아미나제는 이. 콜리 기원이다. 일부 구현예에서, 아데닌 데아미나제는 천연적으로 발생하는 아데노신 데아미나제이고, 이는 본원에 제공된 임의의 돌연변이에 상응하는 하나 이상의 돌연변이 (예를 들어, ecTadA 내 돌연변이)를 포함한다. 임의의 상동성 단백질 내 상응하는 잔기는 예를 들어, 상동성 잔기의 서열 정렬 및 결정에 의해 동정될 수 있다. 따라서, 본원에 기재된 임의의 돌연변이 (예를 들어, ecTadA에서 동정된 임의의 돌연변이)에 상응하는 임의의 천연적으로 발생하는 아데노신 데아미나제 (예를 들어, ecTadA와 상동성을 갖는)에서 돌연변이가 생성될 수 있다.

[2196] ***아데노신 데아미나제***

[2197] 일부 구현예에서, 본원에 기재된 융합 단백질은 아데노신 데아미나제를 포함하는 데아미나제 도메인을 포함할 수 있다. 염기 편집기의 상기 아데노신 데아미나제 도메인은 아데닌 (A)을 탈아민화하여 구아닌 (G)의 염기 쌍 형성 성질을 나타내는 이노신 (I)을 형성함으로써 아데닌 (A) 핵염기의 구아닌 (G) 핵염기로의 편집을 촉진시킬 수 있다. 아데노신 데아미나제는 데옥시리보핵산 (DNA)에서 데옥시아데노신 잔기의 아데닌을 탈아민화 (즉, 아민 그룹을 제거하는)할 수 있다.

[2198] 일부 구현예에서, 본원에 제공된 아데노신 데아미나제는 아데닌을 탈아민화할 수 있다. 일부 구현예에서, 본원에 제공된 아데노신 데아미나제는 DNA의 데옥시아데노신 잔기에서 아데닌을 탈아민화할 수 있다. 일부 구현예에서, 아데닌 데아미나제는 천연적으로 발생하는 아데노신 데아미나제이고, 이는 본원에 제공된 임의의 돌연변이에 상응하는 하나 이상의 돌연변이 (예를 들어, ecTadA 내 돌연변이)를 포함한다. 당업자는 예를 들어, 서열 정렬 및 상동성 잔기의 결정에 의해 임의의 상동성 단백질에서 상응하는 잔기를 동정할 수 있을 것이다. 따라서, 당업자는 임의의 천연적으로 발생하는 아데노신 데아미나제 (예를 들어, ecTadA와 상동성을 갖는)에서 돌연변이를 생성할 수 있을 것이고, 이는 본원에 기재된 임의의 돌연변이, 예를 들어, ecTadA에서 동정된 임의의 돌연변이에 상응한다. 일부 구현예에서, 아데노신 데아미나제는 원핵생물 기원이다. 일부 구현예에서, 아데노신 데아미나제는 세균 기원이다. 일부 구현예에서, 아데노신 데아미나제는 에스케리치아 콜리, 스타필로코커스 아우레우스, 살모넬라 타이피, 쉐와넬라 푸트레파시엔스, 해모필러스 인플루엔자, 콜로박터 크레슨투스, 또는 바실러스 서브틸리스 기원이다. 일부 구현예에서, 아데노신 데아미나제는 이. 콜리 기원이다.

[2199] 본원 개시내용은 효율 (>50-60%) 및 특이성이 증가된 아데노신 데아미나제 변이체를 제공한다. 특히, 본원에 기재된 아데노신 데아미나제 변이체는 폴리뉴클레오타이드 내 목적하는 염기를 편집할 가능성이 높고 변경시키고자 하지 않는 염기 (즉, "바이스탠더")를 편집할 가능성이 적다.

[2200] 일부 구현예에서, 아데노신 데아미나제는 TadA 데아미나제이다. 특정 구현예에서, TadA는 이의 전문이 본원에 참조로 포함된 PCT/US2017/045381 (WO2018/027078)에 기재된 TadA의 임의의 하나이다.

[2201] 일부 구현예에서, 본원 개시내용의 핵염기 편집기는 하기의 서열 중 변경을 포함하는 아데노신 데아미나제 변이체이다:

[2202] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (또한 TadA*7.10으로 칭함).

[2203] 특정 구현예에서, 융합 단백질은 단일 (예를 들어, 단량체로서 제공된) TadA*8 변이체를 포함한다. 일부 구현예에서, TadA*8은 Cas9 닉카제에 연결된다. 일부 구현예에서, 본원 개시내용의 융합 단백질은 TadA*8 변이체에 연결된 야생형 TadA (TadA(wt))의 이종이량체로서 포함한다. 다른 구현예에서, 본원 개시내용의 융합 단백질은 TadA*8 변이체에 연결된 TadA*7.10의 이종이량체로서 포함한다. 일부 구현예에서, 염기 편집기는 TadA*8 변이체 단량체를 포함하는 ABE8이다. 일부 구현예에서, 염기 편집기는 TadA*8 변이체 및 TadA(wt)의 이종이량체를 포함하는 ABE8이다. 일부 구현예에서, 염기 편집기는 TadA*8 변이체 및 TadA*7.10의 이종이량체를 포함하는 ABE8이다. 일부 구현예에서, 염기 편집기는 TadA*8 변이체의 이종이량체를 포함하는 ABE8이다. 일부 구현예에서, TadA*8 변이체는 **표 9**로부터 선택된다. 일부 구현예에서, ABE8은 **표 8, 9, 10** 또는 **11**로부터 선택된다. 관련 서열은 다음과 같다:

[2204] 야생형 TadA (TadA(wt)) 또는 "TadA 참조 서열"

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRV

[2205] EITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[2206] TadA*7.10:

MSEVEFSHEYW MRHALTLAKR ARDEREVPVG AVLVLNNRVI GEGWNRAIGL
HDPTAHAEIM ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT

[2207] GAAGSLMDVL HYPGMNHRVE ITEGILADEC AALLCYFFRM PRQVFNAQKK AQSSTD

[2208] 일부 구현예에서, 아데노신 데아미나제는 본원에 제공된 임의의 아데노신 데아미나제에 제시된 아미노산 서열 중 임의의 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. 본원에 제공된 아데노신 데아미나제는 하나 이상의 돌연변이 (예를 들어, 본원에 제공된 임의의 돌연변이)를 포함할 수 있는 것으로 인지되어야 한다. 본원 개시내용은 특정 퍼센트 동일성 + 본원에 기재된 임의의 돌연변이 또는 이의 조합을 갖는 임의의 데아미나제 도메인을 제공한다. 일부 구현예에서, 아데노신 데아미나제는 참조 서열 또는 본원에 제공된 임의의 아데노신 데아미나제와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 당업계에 공지되거나 본원에 기재된 아미노산 서열 중 임의의 하나와 비교하여 적어도 5, 적어도 10, 적어도 15, 적어도 20, 적어도 25, 적어도 30, 적어도 35, 적어도 40, 적어도 45, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 적어도 90, 적어도 100, 적어도 110, 적어도 120, 적어도 130, 적어도 140, 적어도 150, 적어도 160, 또는 적어도 170개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[2209] 일부 구현예에서, TadA 데아미나제는 전장 이. 콜리 TadA 데아미나제이다. 예를 들어, 특정 구현예에서, 아데노신 데아미나제는 하기 아미노산 서열을 포함한다:

MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDP
TAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD

[2210] VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD.

[2211] 그러나, 본원에 유용한 추가의 아데노신 데아미나제가 당업자에게 자명하고 본원 개시내용의 범위 내에 있는 것으로 인지되어야 한다. 예를 들어, 아데노신 데아미나제는 tRNA에 작용하는 아데노신 데아미나제 (ADAT)의 동족체일 수 있다. 제한 없이, 예시적인 AD AT 동족체의 아미노산 서열은 다음을 포함한다:

[2212] 스타필로코커스 아우레우스 TadA:

MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIER
AAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGS

[2213] LMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[2214] 바실러스 서브틸리스 TadA:

MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKA
LGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVS

[2215] GVLEEECGGMLSAFFRELRKKKKAARKNLSE

[2216] 살모넬라 타이피무리움 (에스. 티피무리움) TadA:

MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDP
TAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLID

[2217] VLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[2218] 쉐와넬라 푸트레파시엔스 (에스. 푸트레파시엔스) TadAA:

MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLE
NYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGV

[2219] LAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[2220] 해모필러스 인플루엔자 F3031 (에이취. 인플루엔자) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEII
ALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKM

[2221] NHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[2222] 콜로박터 크레슨투스 (씨. 크레슨투스) TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIA
AMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTC

[2223] HWRPEVTGGVLADESADLLRGFFRARRKAKI

[2224] 게오박터 설푸레두센스 (지. 설푸레두센스) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMI
AIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRL

[2225] NHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[2226] 이. 콜리 (*E. Coli*) TadA (ecTadA)의 구현예는 하기를 포함한다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV

[2227] EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[2228] 일부 구현예에서, 아데노신 데아미나제는 원핵생물 기원이다. 일부 구현예에서, 아데노신 데아미나제는 세균 기원이다. 일부 구현예에서, 아데노신 데아미나제는 에스케리치아 콜리, 스타필로코커스 아우레우스, 살모넬라 타이피, 쉐와넬라 푸트레파시엔스, 해모필러스 인플루엔자, 콜로박터 크레슨투스, 또는 바실러스 서브틸리스 기원이다. 일부 구현예에서, 아데노신 데아미나제는 이. 콜리 기원이다.

[2229] 하나의 구현예에서, 본원 개시내용의 융합 단백질은 Cas9 닉카제에 연결된 TadA*7.10에 연결된 야생형 TadA를 포함한다. 특정 구현예에서, 융합 단백질은 단일 TadA*7.10 도메인 (예를 들어, 단량체로서 제공된)을 포함한다. 다른 구현예에서, ABE7.10 편집기는 TadA*7.10 및 TadA(wt)를 포함하고 이들은 이종이량체를 형성할 수 있다.

[2230] 일부 구현예에서, 아데노신 데아미나제는 본원에 제공된 임의의 아데노신 데아미나제에 제시된 아미노산 서열 중 임의의 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. 본원에 제공된 아데노신 데아미나제는 하나 이상의 돌연변이 (예를 들어, 본원에 제공된 임의의 돌연변이)를 포함할 수 있는 것으로 인지되어야 한다. 본원 개시내용은 특정 퍼센트 동일성 + 본원에 기재된 임의의 돌연변이 또는 이의 조합을 갖는 임의의 데아미나제 도메인을 제공한다. 일부 구현예에서, 아데노신 데아미나

제는 참조 서열 또는 본원에 제공된 임의의 아데노신 데아미나제와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 당업계에 공지되거나 본원에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 5, 적어도 10, 적어도 15, 적어도 20, 적어도 25, 적어도 30, 적어도 35, 적어도 40, 적어도 45, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 적어도 90, 적어도 100, 적어도 110, 적어도 120, 적어도 130, 적어도 140, 적어도 150, 적어도 160, 또는 적어도 170개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[2231] 본원에 제공된 임의의 돌연변이 (예를 들어, TadA 참조 서열을 기준으로)가 다른 아데노신 데아미나제, 예를 들어, 이. 콜리 TadA (ecTadA), 에스. 아우레우스 (*S. aureus*) TadA (saTadA), 또는 다른 아데노신 데아미나제 (예를 들어, 세균 아데노신 데아미나제)로 도입될 수 있는 것으로 인지되어야 한다. 추가의 데아미나제가 본원에 제공된 바와 같이 돌연변이될 수 있는 상동성 아미노산 잔기를 동정하기 위해 유사하게 정렬될 수 있다는 것은 당업자에게 자명할 것이다. 따라서, TadA 참조 서열에서 동정된 임의의 돌연변이는 상동성 아미노산 잔기를 갖는 다른 아데노신 데아미나제 (예를 들어, ecTada)에 만들어질 수 있다. 또한 본원에 제공된 임의의 돌연변이는 개별적으로 또는 TadA 참조 서열 또는 또 다른 아데노신 데아미나제와 임의의 조합으로 만들어질 수 있는 것으로 인지되어야 한다.

[2232] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 D108X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 또 다른 아데노신 데아미나제에서 D108G, D108N, D108V, D108A, 또는 D108Y 돌연변이 또는 상응하는 돌연변이를 포함한다.

[2233] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 A106X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 A106V 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, 야생형 TadA 또는 ecTadA)에서 상응하는 돌연변이를 포함한다.

[2234] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 E155X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X의 존재는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 E155D, E155G 또는 E155V 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2235] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 D147X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X의 존재는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 D147Y 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2236] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 A106X, E155X 또는 D147X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 E155D, E155G, 또는 E155V 돌연변이를 포함한다. 일부 구현예에서, 아데노신 데아미나제는 D147Y를 포함한다.

[2237] 예를 들어, 아데노신 데아미나제는 TadA 참조 서열에서 D108N, A106V, E155V, 및/또는 D147Y 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함할 수 있다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 하기의 돌연변이 그룹 (돌연변이 그룹은 ";"에 의해 분리된다), 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 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, 본원에 제공된 상응하는 돌연변이의 임의의 조합은 아데노신 데아미나제 (예를 들어, ecTadA)에 만들어질 수 있음이 인지되어야 한다.

[2238] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 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 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이를 포함하고, 여기서, X의 존재는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 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 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이를 포함한다.

[2239] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8X, D108X, 및/또는 N127X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이를 포함하고, 여기서, X는 임의의 아미노산의 존재를 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, D108N 및/또는 N127S 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2240] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 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 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 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 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이를 포함한다.

[2241] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8X, D108X, N127X, D147X, R152X, 및 Q154X로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8X, M61X, M70X, D108X, N127X, Q154X, E155X 및 Q163X로 이루어진 그룹으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 7개 또는 8개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8X, D108X, N127X, E155X, 및 T166X로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다.

[2242] 일부 구현예에서, 아데노신 데아미나제는 H8X, A106X, 및 D108X로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개 돌연변이, 또는 또 다른 아데노신 데아미나제에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다. 일부 구현예에서, 아데노신 데아미나제는 H8X, R26X, L68X, D108X, N127X, D147X 및 E155X로 이루어진 그룹으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 7개 또는 8개 돌연변이, 또는 또 다른 아데노신 데아미나제에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8X, D108X, A109X, N127X, 및 E155X로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다.

[2243] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, D108N, N127S, D147Y, R152C, 및 Q154H로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아

미나제는 TadA 참조 서열에서 H8Y, M61I, M70V, D108N, N127S, Q154R, E155G 및 Q163H로 이루어진 그룹으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 7개 또는 8개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, D108N, N127S, E155V, 및 T166P로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, A106T, D108N, N127S, E155D, 및 K161Q로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, R26W, L68Q, D108N, N127S, D147Y, 및 E155V로 이루어진 그룹으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 7개 또는 8개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, D108N, A109T, N127S, 및 E155G로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함한다.

[2244] 본원에 제공된 임의의 돌연변이들 및 임의의 추가의 돌연변이들 (예를 들어, ecTadA 아미노산 서열을 기준으로)은 임의의 다른 아데노신 데아미나제에 도입될 수 있다. 또한 본원에 제공된 임의의 돌연변이는 개별적으로 또는 TadA 참조 서열 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)와 임의의 조합으로 만들어질 수 있다.

[2245] A에서 G로의 핵염기 편집 단백질의 세부사항은 문헌 (참조: 국제 PCT 출원 번호 PCT/2017/045381 (WO2018/027078) 및 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), 이의 전체 내용은 본원에 참조로 포함됨)에 기재되어 있다.

[2246] 일부 구현예에서, 아데노신 데아미나제는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이를 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 D108N, D108G 또는 D108V 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 A106V 또는 D108N 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 R107C 및 D108N 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, D108N, N127S, D147Y, 및 Q154H 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, D108N, N127S, D147Y, 및 E155V 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 D108N, D147Y 및 E155V 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, D108N 및 N127S 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 A106V, D108N, D147Y, 및 E155V 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이들을 포함한다.

[2247] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 S2X, H8X, I49X, L84X, H123X, N127X, I156X 및/또는 K160X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미나제에서 하나 이상의 상응하는 돌연변이들을 포함하고, 여기서, X의 존재는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 S2A, H8Y, I49F, L84F, H123Y, N127S, I156F 및/또는 K160S 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이들을 포함한다.

[2248] 일부 구현예에서, 아데노신 데아미나제는 L84X 돌연변이 아데노신 데아미나제를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 L84F 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2249] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 H123X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H123Y 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2250] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 I156X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 I156F 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2251] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 L84X, A106X, D108X, H123X, D147X, E155X, 및 I156X로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개, 5개, 6개 또는 7개의 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 S2X, I49X, A106X, D108X, D147X, 및 E155X로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8X, A106X, D108X, N127X, 및 K160X로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 지적한다.

[2252] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 L84F, A106V, D108N, H123Y, D147Y, E155V, 및 I156F로 이루어진 그룹으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 또는 7개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 S2A, I49F, A106V, D108N, D147Y, 및 E155V로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개, 5개 또는 6개 돌연변이를 포함한다.

[2253] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H8Y, A106T, D108N, N127S, 및 K160S로 이루어진 그룹으로부터 선택되는 1개, 2개, 3개, 4개 또는 5개 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이 또는 돌연변이들을 포함한다.

[2254] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 E25X, R26X, R107X, A142X 및/또는 A143X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이를 포함하고, 여기서, X의 존재는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 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 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이들을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에 상응하는 본원에 기재된 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이들을 포함한다.

[2255] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 E25X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 E25M, E25D, E25A, E25R, E25S 또는 E25Y 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2256] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 R26X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에

서 R26G, R26N, R26Q, R26C, R26L 또는 R26K 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2257] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 R107X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 R107P, R107K, R107A, R107N, R107W, R107H 또는 R107S 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2258] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 A142X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 A142N, A142D 또는 A142G 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2259] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 A143X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143S, A143Q 및/또는 A143R 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2260] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H36X, N37X, P48X, I49X, R51X, M70X, N72X, D77X, E134X, S146X, Q154X, K157X 및/또는 K161X 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이들을 포함하고, 여기서, X의 존재는 야생형 아데노신 데아미나제에서 상응하는 아미노산이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H36L, N37T, N37S, P48T, P48L, I49V, R51H, R51L, M70L, N72S, D77G, E134G, S146R, S146C, Q154H, K157N 및/또는 K161T 돌연변이 중 하나 이상, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 하나 이상의 상응하는 돌연변이들을 포함한다.

[2261] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 H36X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H36L 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2262] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 N37X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 N37T 또는 N37S 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2263] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 P48X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 P48T 또는 P48L 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2264] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 R51X 돌연변이, 또는 또 다른 아데노신 데아미나제에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 R51H 또는 R51L 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2265] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 S146X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 S146R 또는 S146C 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2266] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 K157X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 K157N 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2267] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 P48X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 P48S, P48T 또는 P48A 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2268] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 A142X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 A142N 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2269] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 W23X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 W23R 또는 W23L 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2270] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 R152X 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 R152P 또는 R52H 돌연변이, 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[2271] 하나의 구현예에서, 아데노신 데아미나제는 돌연변이 H36L, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F 및 K157N을 포함할 수 있다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에 비해 하기의 돌연변이 조합을 포함하고, 여기서, 조합의 각각의 돌연변이는 "_"에 의해 분리되고 돌연변이의 각각의 조합은 괄호 사이에 있다:

(A106V_D108N),
(R107C_D108N),
(H8Y_D108N_N127S_D147Y_Q154H),
(H8Y_D108N_N127S_D147Y_E155V),
(D108N_D147Y_E155V),
(H8Y_D108N_N127S),
(H8Y_D108N_N127S_D147Y_Q154H),
(A106V_D108N_D147Y_E155V),
(D108Q_D147Y_E155V),
(D108M_D147Y_E155V),

[2272]

(D108L_D147Y_E155V),
(D108K_D147Y_E155V),
(D108I_D147Y_E155V),
(D108F_D147Y_E155V),
(A106V_D108N_D147Y),
(A106V_D108M_D147Y_E155V),
(E59A_A106V_D108N_D147Y_E155V),
(E59A cat dead_A106V_D108N_D147Y_E155V),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D103A_D104N),
(G22P_D103A_D104N),
(D103A_D104N_S138A),
(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V
_I156F),
(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_
I156F),
(R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V
_I156F),
(R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),
(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V
_I156F),
(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(A106V_D108N_A142N_D147Y_E155V),
(R26G_A106V_D108N_A142N_D147Y_E155V),
(E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),
(R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),
(E25D_R26G_A106V_D108N_A142N_D147Y_E155V),
[2273] (A106V_R107K_D108N_A142N_D147Y_E155V),

(A106V_D108N_A142N_A143G_D147Y_E155V),
(A106V_D108N_A142N_A143L_D147Y_E155V),
(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),
(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),
(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),
(N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),
(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
(H36L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),
(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F),
(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),
(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),
(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F
_K157N), (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

[2274]

(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),
(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(P48S_A142N),
(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),
(P48T_I49V_A142N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N),
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_
I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F
_K161T),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V
_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V
_I156F _K157N),

[2275]

(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P
_E155V_I156F_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F
_K161T),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V
_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155V
_I156F _K157N).

[2276]

[2277] 특정 구현예에서, 본원에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성을 개선시키는 하나 이상의 특성을 포함한다. 예를 들어, 본원에 제공된 임의의 융합 단백질은 감소된 뉴클레아제 활성을 갖는 Cas9 도메인을 포함할 수 있다. 일부 구현예에서, 본원에 제공된 임의의 융합 단백질은 뉴클레아제 활성을 갖지 않는 Cas9 도메인 (dCas9), 또는 듀플렉스 DNA 분자의 하나의 가닥을 절단하는, Cas9 닉카제로서 언급되는 Cas9 도메인 (nCas9)을 가질 수 있다.

[2278] 일부 구현예에서, 아데노신 데아미나제는 TadA*7.10이다. 일부 구현예에서, TadA*7.10은 적어도 하나의 변경을 포함한다. 특정 구현예에서, TadA*7.10은 하기의 변경 중 하나 이상을 포함한다: Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및 Q154R. 변경 Y123H는 또한 본원에서 H123H로서 언급된다 (TadA*7.10에서 변경 H123Y는 Y123H (wt)로 복귀함). 다른 구현예에서, TadA*7.10은 하기의 그룹으로부터 선택된 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. 특정 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 잔기 149, 150, 151, 152, 153, 154, 155, 156, 및 157에서 시작하는 C 말단의 결실을 포함한다.

[2279] 다른 구현예에서, 본원 개시내용의 염기 편집기는 하기의 변경 중 하나 이상을 포함하는 아데노신 데아미나제 변이체 (예를 들어, TadA*8)를 포함하는 단량체이다: TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 TadA7.10 또는 TadA 참조 서열과 비교하여 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R. 다른 구현예에서, 아데노신 데아미나제 변이체 (TadA*8)는 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 하기의 그룹으로부터 선택된 변경의 조합을 포함하는 단량체이다: 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. 다른 구현예에서, 염기 편집기는 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 야생형 TadA 아데노신 데아미나제 및 하기의 변경 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R 중 하나 이상을 포함하는 아데노신 데아미나제 변이체 (예를 들어, TadA*8)를 포함하는 이종이량체이다. 다른 구현예에서, 염기 편집기는 TadA*7.10 도메인 및 하기의 그룹으로부터 선택된 변경의 조합을 포함하는 TadA*7.10, TadA 참조 서열, 또는 또 다른 TadA에서의 상응하는 돌연변이에 비해 아데노신 데아미나제 변이체 도메인 (예를 들어 TadA*8)을 포함하는 이종이량체이다: 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.

[2280] 하나의 구현예에서, 아데노신 데아미나제는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 TadA*8이다:

[2281]
```
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD
```

[2282] 일부 구현예에서, TadA*8은 절단된다. 일부 구현예에서, 절단된 TadA*8은 전장 TadA*8에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 N-말단 아미노산 잔기가 결실된다. 일부 구현예에서, 절단된 TadA*8은 전장 TadA*8에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 C-말단 아미노산 잔기가 결실된다. 일부 구현예에서, 아데노신 데아미나제 변이체는 전장 TadA*8이다.

[2283] 일부 구현예에서, 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.

[2284] 하나의 구현예에서, 본원 개시내용의 융합 단백질은 Cas9 닉카제에 연결된 본원에 기재된 아데노신 데아미나제 변이체 (예를 들어, TadA*8)에 연결된 야생형 TadA를 포함한다. 특정 구현예에서, 융합 단백질은 (예를 들어, 단량체로서 제공된) 단일 TadA*8 도메인을 포함한다. 다른 구현예에서, 염기 편집기는 TadA*8 및 TadA(wt)를 포함하고 이들은 이종이량체를 형성할 수 있다. 예시적인 서열은 다음과 같다:

[2285] TadA(wt) 또는 "TadA 참조 서열":

[2286]
```
MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD
```

[2287] TadA*7.10:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
[2288] EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[2289] TadA*8:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
[2290] EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[2291] 일부 구현예에서, 아데노신 데아미나제는 본원에 제공된 임의의 아데노신 데아미나제에 제시된 아미노산 서열 중 임의의 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. 본원에 제공된 아데노신 데아미나제는 하나 이상의 돌연변이 (예를 들어, 본원에 제공된 임의의 돌연변이)를 포함할 수 있는 것으로 인지되어야 한다. 본원의 개시내용은 특정 퍼센트 동일성 + 본원에 기재된 임의의 돌연변이 또는 이의 조합을 갖는 임의의 데아미나제 도메인을 제공한다. 일부 구현예에서, 아데노신 데아미나제는 참조 서열 또는 본원에 제공된 임의의 아데노신 데아미나제와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 구현예에서, 아데노신 데아미나제는 당업계에 공지되거나 본원에 기재된 아미노산 서열 중 임의의 하나와 비교하여 적어도 5, 적어도 10, 적어도 15, 적어도 20, 적어도 25, 적어도 30, 적어도 35, 적어도 40, 적어도 45, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 적어도 90, 적어도 100, 적어도 110, 적어도 120, 적어도 130, 적어도 140, 적어도 150, 적어도 160, 또는 적어도 170개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[2292] 특정 구현예에서, TadA*8은 굵게 표시한 하기의 임의의 위치에서 하나 이상의 돌연변이를 포함한다. 다른 구현예에서, TadA*8은 밑줄로 표시한 임의의 위치에서 하나 이상의 돌연변이를 포함한다:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG 50
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG 100
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR 150
[2293] MPRQVFNAQK KAQSSTD

[2294] 예를 들어, TadA*8은 단독으로 또는 TadA*7.10, TadA 참조 서열 또는 또 다른 TadA에서 상응하는 돌연변이와 상대적으로 하기 Y147T, Y147R, Q154S, Y123H, 및/또는 Q154R 중 임의의 하나 이상과 조합하여 아미노산 위치 82 및/또는 166 (예를 들어, V82S, T166R)에서 변경을 포함한다. 특정 구현예에서, 변경의 조합은 TadA*7.10, TadA 참조 서열 또는 또 다른 TadA에서 상응하는 돌연변이에 비해 하기의 그룹으로부터 선택된다: 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.

[2295] 일부 구현예에서, 아데노신 데아미나제는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 TadA*8이다:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG LHDPTAHAEI
MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCTFFR
[2296] MPRQVFNAQK KAQSSTD

[2297] 일부 구현예에서, TadA*8은 절단된다. 일부 구현예에서, 절단된 TadA*8은 전장 TadA*8에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 N-말단 아미노산 잔기가 결실된다. 일부 구

현예에서, 절단된 TadA*8은 전장 TadA*8에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 C-말단 아미노산 잔기가 결실된다. 일부 구현예에서, 아데노신 데아미나제 변이체는 전장 TadA*8이다.

[2298] 하나의 구현예에서, 본원 개시내용의 융합 단백질은 Cas9 닉카제에 연결된, 본원에 기재된 아데노신 데아미나제 변이체 (예를 들어, TadA*8)에 연결된 야생형 TadA를 포함한다. 특정 구현예에서, 융합 단백질은 단일 TadA*8 도메인 (예를 들어, 단량체로서 제공된)을 포함한다. 다른 구현예에서, 염기 편집기는 TadA*8 및 TadA(wt)을 포함하고 이들은 이종이량체를 형성할 수 있다.

[2299] *추가의 도메인*

[2300] 본원에 기재된 염기 편집기는 핵염기 편집, 폴리뉴클레오타이드의 핵염기의 변형 또는 변경의 촉진을 도와주는 임의의 도메인을 포함할 수 있다. 일부 구현예에서, 염기 편집기는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 (예를 들어, Cas9), 핵염기 편집 도메인 (예를 들어, 데아미나제 도메인) 및 하나 이상의 추가의 도메인을 포함한다. 일부 구현예에서, 추가의 도메인은 염기 편집기의 효소 또는 촉매 기능, 염기 편집기의 결합 기능을 촉진시킬 수 있거나 목적하는 염기 편집 결과를 방해할 수 있는 세포 기구 (예를 들어, 효소)의 저해제일 수 있다. 일부 구현예에서, 염기 편집기는 뉴클레아제, 닉카제, 리컴비나제, 데아미나제, 메틸트랜스퍼라제, 메틸라제, 아세틸라제, 아세틸트랜스퍼라제, 전사 활성화인자, 또는 전사 리프레서 도메인을 포함할 수 있다.

[2301] 일부 구현예에서, 염기 편집기는 우라실 글리코실라제 저해제 (UGI) 도메인을 포함한다. 일부 구현예에서, U의 존재에 대한 세포 DNA 복구 반응: G 헤테로듀플렉스 DNA는 세포에서 핵염기 편집 효율에서의 감소에 관여할 수 있다. 상기 구현예에서, 우라실 DNA 글리코실라제 (UDG)는 세포에서 DNA로부터 U의 제거를 촉매할 수 있고, 이는 염기 절제 복구 (BER)를 개시하여 대부분 U:G 쌍의 C:G 쌍으로의 복귀를 유도할 수 있다. 상기 구현예에서, BER은 단일 가닥에 결합하고, 편집된 염기를 차단하고, UGI를 저해하고, BER을 저해하고, 편집된 염기를 보호하고/하거나 비-편집된 가닥의 복구를 촉진시키는 하나 이상의 도메인을 포함하는 염기 편집기에서 저해될 수 있다. 따라서, 본원 개시내용은 UGI 도메인을 포함하는 염기 편집기 융합 단백질을 고려한다.

[2302] 일부 구현예에서, 염기 편집기는 도메인으로서 이중 가닥 절단 (DSB) 결합 단백질의 전부 또는 일부를 포함한다. 예를 들어, DSB 결합 단백질은 DSB의 말단에 결합할 수 있고 이들을 분해로부터 보호할 수 있는 박테리오파아지 Mu의 Gam 단백질을 포함할 수 있다. 문헌 (참조: 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))을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[2303] 추가로, 일부 구현예에서, Gam 단백질은 염기 편집기의 N 말단에 융합될 수 있다. 일부 구현예에서, Gam 단백질은 염기 편집기의 C-말단에 융합될 수 있다. 박테리오파아지 Mu의 Gam 단백질은 이중 가닥 절단 (DSB)의 말단에 결합하고 이들을 분해로부터 보호할 수 있다. 일부 구현예에서, DSB의 유리된 말단에 결합하는 Gam을 사용하여 염기 편집 과정 동안에 삽입-결실 형성을 감소시킬 수 있다. 일부 구현예에서, 174개 잔기 Gam 단백질은 염기 편집기의 N 말단에 융합된다. 문헌 (참조: 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))을 참조한다. 일부 구현예에서, 돌연변이 또는 돌연변이들은 야생형 도메인과 비교하여 염기 편집기 도메인의 길이를 변화시킬 수 있다. 예를 들어, 적어도 하나의 도메인에서 적어도 하나의 아미노산의 결실은 염기 편집기의 길이를 감소시킬 수 있다. 또 다른 경우에, 돌연변이 또는 돌연변이들은 야생형 도메인과 비교하여 도메인의 길이를 변화시키지 않는다. 예를 들어, 임의의 도메인에서 치환(들)은 염기 편집기의 길이를 변화시키지 않는다.

[2304] 일부 구현예에서, 염기 편집기는 도메인으로서 핵산 폴리머라제 (NAP) 전부 또는 일부를 포함할 수 있다. 예를 들어, 염기 편집기는 진핵 세포 NAP의 전부 또는 일부를 포함할 수 있다. 일부 구현예에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 DNA 폴리머라제이다. 일부 구현예에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 트랜스레젼 (translesion) 폴리머라제 활성을 갖는다. 일부 구현예에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 트랜스레젼 DNA 폴리머라제이다. 일부 구현예에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파, 또는 폴리머라제 에타이다. 일부 구현예에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 진핵 세포 폴리머라제 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, mu, 또는 nu 성분이다. 일부 구현예에서, 염기 편집기에 혼입된 NAP 또는 이의 일부는 핵산

폴리머라제 (예를 들어, 트랜스레젼 DNA 폴리머라제)와 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 99.5% 동일한 아미노산 서열을 포함한다.

[2305] **염기 편집기 시스템**

[2306] 본원에 제공된 염기 편집기 시스템의 사용은 하기의 단계를 포함한다: (a) 대상체의 폴리뉴클레오타이드 (예를 들어, 이중 또는 단일 가닥 DNA 또는 RNA)의 표적 뉴클레오타이드 서열을 핵염기 편집기 (예를 들어, 아데노신 염기 편집기) 및 가이드 핵산 (예를 들어, gRNA)을 포함하는 염기 편집기 시스템과 접촉시키는 단계로서, 상기 표적 뉴클레오타이드 서열이 표적화된 핵염기 쌍을 포함하는 단계; (b) 상기 표적 영역의 가닥 분리를 유도하는 단계; (c) 상기 표적 영역의 단일 가닥 내 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계; 및 (d) 상기 표적 영역의 하나 이하의 가닥을 절단하는 단계로서, 상기 제1 핵염기에 상보적인 제3 핵염기가 상기 제2 핵염기에 상보적인 제4 핵염기로 대체되는, 단계. 일부 구현예에서, 단계 (b)는 생략되는 것으로 인지되어야 한다. 일부 구현예에서, 상기 표적화된 핵염기 쌍은 하나 이상의 유전자에서 다수의 핵염기 쌍이다. 일부 구현예에서, 본원에 제공된 염기 편집기 시스템은 하나 이상의 유전자에서 다수의 핵염기 쌍의 멀티플렉스 편집을 수행할 수 있다. 일부 구현예에서, 다수의 핵염기 쌍은 동일한 유전자에 위치한다. 일부 구현예에서, 다수의 핵염기 쌍은 하나 이상의 유전자에 위치하고, 여기서, 적어도 하나의 유전자는 상이한 유전자좌에 위치한다.

[2307] 일부 구현예에서, 절단된 단일 가닥 (닉 가닥)은 가이드 핵산에 하이브리드화한다. 일부 구현예에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥에 반대편에 있다. 일부 구현예에서, 염기 편집기는 Cas9 도메인을 포함한다. 일부 구현예에서, 제1 염기는 아데닌이고 제2 염기는 G, C, A, 또는 T가 아니다. 일부 구현예에서, 제2 염기는 이노신이다.

[2308] 본원에 제공된 바와 같은 염기 편집 시스템은 촉매 결함 스트렙토코커스 피오게네스 Cas9, 아데노신 데아미나제, 및 이중 가닥 DNA 절단을 생성하지 않고 공여자 DNA 주형을 요구하지 않고 과량의 확률적 삽입 및 결실을 유도하지 않으면서 DNA에서 프로그래밍 가능한 단일 뉴클레오타이드 (A→G) 변화를 유도하는 염기 절제 복구의 저해제를 함유하는 융합 단백질을 사용하는 게놈 편집에 대한 신규 접근법을 제공한다.

[2309] 본원에서 염기 편집기 시스템을 사용하는 핵염기를 편집하기 위한 시스템, 조성물 및 방법이 제공된다. 일부 구현예에서, 염기 편집기 시스템은 (1) 핵염기를 편집하기 위해 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 핵염기 편집 도메인 (예를 들어, 데아미나제 도메인)을 포함하는 염기 편집기 (BE); 및 (2) 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인과 접합된 가이드 폴리뉴클레오타이드 (예를 들어, 가이드 RNA)를 포함한다. 일부 구현예에서, 염기 편집기 시스템은 아데노신 염기 편집기 (ABE)를 포함한다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 RNA 결합 도메인이다. 일부 구현예에서, 핵염기 편집 도메인은 데아미나제 도메인이다. 일부 구현예에서, 데아미나제 도메인은 아데닌 데아미나제 또는 아데노신 데아미나제일 수 있다.

[2310] 일부 구현예에서, 아데노신 염기 편집기는 DNA에서 아데닌을 탈아민화할 수 있다. 일부 구현예에서, ABE는 변화된 TadA 변이체를 포함한다.

[2311] 핵염기 편집 단백질의 세부사항은 국제 PCT 출원 번호 PCT/2017/045381 (WO2018/027078) 및 PCT/US2016/058344 (WO2017/070632)에 기재되어 있고, 이의 각각은 이의 전문이 본원에 참조로 포함된다. 또한 문헌 (참조: 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))을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[2312] 일부 구현예에서, 단일 가이드 폴리뉴클레오타이드는 데아미나제를 표적 핵산 서열에 표적화하기 위해 사용될 수 있다. 일부 구현예에서, 단일 쌍의 가이드 폴리뉴클레오타이드는 상이한 데아미나제를 표적 핵산 서열에 표적화하기 위해 사용될 수 있다.

[2313] 염기 편집기 시스템의 핵염기 성분 및 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 성분은 서로

공유적으로 또는 비공유적으로 연합될 수 있다. 예를 들어, 일부 구현예에서, 데아미나제 도메인은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미나제 도메인에 융합되거나 연결될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미나제 도메인을 데아미나제 도메인과 비공유적으로 상호작용하거나 연합함에 의해 표적 뉴클레오타이드 서열에 표적화시킬 수 있다. 예를 들어, 일부 구현예에서, 뉴클레아제 편집 성분, 예를 들어, 데아미나제 성분은 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 일부인 추가의 이종성 부분 또는 도메인과 상호작용하거나, 이와 연합하거나 이와 복합체를 형성할 수 있는 추가의 이종성 부분 또는 도메인을 포함할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가의 이종성 부분은 단백질 도메인일 수 있다. 일부 구현예에서, 추가의 이종성 부분은 K 상동성 (KH) 도메인, MS2 코트 단백질 도메인, PP7 코트 단백질 도메인, SfMu Com 코트 단백질 도메인, 무균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인지 모티프일 수 있다.

[2314] 염기 편집기 시스템은 가이드 폴리뉴클레오타이드 성분을 추가로 포함할 수 있다. 염기 편집기 시스템의 성분들은 공유 결합, 비공유 상호작용, 또는 이의 연합과 상호작용의 임의의 조합을 통해 서로 연합될 수 있음이 인지되어야 한다. 일부 구현예에서, 데아미나제 도메인은 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 예를 들어, 일부 구현예에서, 염기 편집기 시스템의 핵염기 편집 성분, 예를 들어, 데아미나제 성분은 가이드 폴리뉴클레오타이드의 일부 또는 분절 (예를 들어, 폴리뉴클레오타이드 모티프)과 상호작용하거나, 연합하거나 이와 복합체를 형성할 수 있는 추가의 이종성 부분 또는 도메인 (예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)을 포함할 수 있다. 일부 구현예에서, 추가의 이종성 부분 또는 도메인 (예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)은 데아미나제 도메인에 융합되거나 연결될 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가의 이종성 부분은 단백질 도메인일 수 있다. 일부 구현예에서, 추가의 이종성 부분은 K 상동성 (KH) 도메인, MS2 코트 단백질 도메인, PP7 코트 단백질 도메인, SfMu Com 코트 단백질 도메인, 무균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인지 모티프일 수 있다.

[2315] 일부 구현예에서, 염기 편집기 시스템은 염기 절제 복구 저해제 (BER)의 성분을 추가로 포함할 수 있다. 염기 편집기 시스템의 성분들은 공유 결합, 비공유 상호작용, 또는 이의 연합과 상호작용의 임의의 조합을 통해 서로 연합될 수 있음이 인지되어야 한다. BER 성분의 저해제는 염기 절제 복구 저해제를 포함할 수 있다. 일부 구현예에서, 염기 절제 복구의 저해제는 우라실 DNA 글리코실라제 저해제 (UGI)일 수 있다. 일부 구현예에서, 염기 절제 복구의 저해제는 이노신 염기 절제 복구 저해제일 수 있다. 일부 구현예에서, 염기 절제 복구의 저해제는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 염기 절제 복구의 저해제에 융합되거나 연결될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 데아미나제 도메인 및 염기 절제 복구의 저해제에 융합되거나 연결될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인은 염기 절제 복구의 저해제를 염기 절제 복구의 저해제와 비공유적으로 상호작용하거나 연합함에 의해 표적 뉴클레오타이드 서열에 표적화시킬 수 있다. 예를 들어, 일부 구현예에서, 염기 절제 복구 성분의 저해제는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인의 일부인 추가의 이종성 부분 또는 도메인과 상호작용하거나, 이와 연합하거나 이와 복합체를 형성할 수 있는 추가의 이종성 부분 또는 도메인을 포함할 수 있다. 일부 구현예에서, 염기 절제 복구의 저해제는 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 예를 들어, 일부 구현예에서, 염기 절제 복구의 저해제는 가이드 폴리뉴클레오타이드의 일부 또는 분절 (예를 들어,

폴리뉴클레오타이드 모티프)과 상호작용하거나, 연합하거나 이와 복합체를 형성할 수 있는 추가의 이종성 부분 또는 도메인 (예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)을 포함할 수 있다. 일부 구현예에서, 가이드 폴리뉴클레오타이드의 추가의 이종성 부분 또는 도메인 (예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오타이드 결합 도메인)은 염기 절제 복구의 저해제에 융합되거나 연결될 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드에 결합하거나 이와 상호작용하거나 이와 연합하거나 이와 복합체를 형성할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 가이드 폴리뉴클레오타이드에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리펩타이드 링커에 결합할 수 있다. 일부 구현예에서, 추가의 이종성 부분은 폴리뉴클레오타이드 링커에 결합할 수 있다. 추가의 이종성 부분은 단백질 도메인일 수 있다. 일부 구현예에서, 추가의 이종성 부분은 K 상동성 (KH) 도메인, MS2 코트 단백질 도메인, PP7 코트 단백질 도메인, SfMu Com 코트 단백질 도메인, 무균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인지 모티프일 수 있다.

[2316] 일부 구현예에서, 염기 편집기는 편집된 가닥의 염기 절제 복구 (BER)를 저해한다. 일부 구현예에서, 염기 편집기는 비-편집된 가닥을 보호하거나 이에 결합한다. 일부 구현예에서, 염기 편집기는 UGI 활성을 포함한다. 일부 구현예에서, 염기 편집기는 촉매적 불활성 이노신-특이적 뉴클레아제를 포함한다. 일부 구현예에서, 염기 편집기는 닉카제 활성을 포함한다. 일부 구현예에서, 염기 쌍의 의도된 편집은 PAM 부위의 업스트림에 있다. 일부 구현예에서, 염기 쌍의 의도된 편집은 PAM 부위의 업스트림의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드이다. 일부 구현예에서, 의도된 염기 쌍의 편집은 PAM 부위의 다운스트림에 있다. 일부 구현예에서 의도된 편집된 염기 쌍은 PAM 부위의 다운스트림의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드이다.

[2317] 일부 구현예에서, 방법은 카노니칼 (예를 들어, NGG) PAM 부위를 필요로 하지 않는다. 일부 구현예에서, 핵염기 편집기는 링커 또는 스페이서를 포함한다. 일부 구현예에서, 링커 또는 스페이서는 1-25개 아미노산 길이이다. 일부 구현예에서, 링커 또는 스페이서는 5-20개 아미노산 길이이다. 일부 구현예에서, 링커 또는 스페이서는 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 아미노산 길이이다.

[2318] 일부 구현예에서, 본원에 제공된 염기 편집 융합 단백질은 정확한 위치, 예를 들어, 표적 염기가 한정된 영역 (예를 들어, "탈아민화 윈도우") 내에 위치할 필요가 있다. 일부 구현예에서, 표적은 4개 염기 영역 내에 있을 수 있다. 일부 구현예에서, 상기 한정된 표적 영역은 PAM의 대략적으로 15개 염기 업스트림에 있을 수 있다. 문헌 (참조: 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))을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[2319] 일부 구현예에서, 표적 영역은 표적 윈도우를 포함하고, 여기서, 상기 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 구현예에서, 표적 윈도우는 1-10개 뉴클레오타이드를 포함한다. 일부 구현예에서, 표적 윈도우는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드 길이이다. 일부 구현예에서, 의도된 염기 쌍의 편집은 표적 윈도우 내에 있다. 일부 구현예에서, 표적 윈도우는 의도된 염기 쌍의 편집을 포함한다. 일부 구현예에서, 상기 방법은 본원에 제공된 임의의 염기 편집기를 사용하여 수행된다. 일부 구현예에서, 표적 윈도우는 탈아민화 윈도우이다. 탈아민화 윈도우는 염기 편집기가 표적 뉴클레오타이드에 작용하여 탈아민화하는 한정된 영역일 수 있다. 일부 구현예에서, 탈아민화 윈도우는 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 염기 영역 내에 있다. 일부 구현예에서, 탈아민화 윈도우는 PAM의 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개 염기 업스트림에 있다.

[2320] 본원 개시내용의 염기 편집기는 표적 폴리뉴클레오타이드 서열의 편집을 촉진시키는, 임의의 도메인, 특성 또는 아미노산 서열을 포함할 수 있다. 예를 들어, 일부 구현예에서, 염기 편집기는 핵 국소화 서열 (NLS)을 포함한다. 일부 구현예에서, 염기 편집기의 NLS는 데아미나제 도메인과 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 사이에 위치한다. 일부 구현예에서, 염기 편집기의 NLS는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 C-말단에 위치한다.

[2321] 본원에 개시된 바와 같이 염기 편집기에 존재할 수 있는 다른 예시적 특성은 융합 단백질의 가용화, 정제 또는 검출을 위해 유용한 서열 태그 뿐만 아니라 세포질 국소화 서열, 배출 서열, 예를 들어, 핵 배출 서열 또는 다

른 국소화 서열과 같은 국소화 서열이다. 본원에 제공된 적합한 단백질 태그는 비오틴 카복실라제 캐리어 단백질 (BCCP) 태그, myc-태그, 칼모듈린-태그, FLAG-태그, 헤마글루티닌 (HA)-태그, 또한 히스티딘 태그 또는 His-태그로서 언급되는 폴리히스티딘 태그, 말토스 결합 단백질 (MBP)-태그, nus-태그, 글루타티온-S-트랜스퍼라제 (GST)-태그, 녹색 형광성 단백질 (GFP)-태그, 티오레독신-태그, S-태그, 소프트태그 (예를 들어, 소프트태그 1, 소프트태그 3), strep-태그, 비오틴 리가제 태그, FlAsH 태그, V5 태그 및 SBP-태그를 포함하지만 이에 제한되지 않는다. 추가의 적합한 서열은 당업자에게 자명할 것이다. 일부 구현예에서, 융합 단백질은 하나 이상의 His 태그를 포함한다.

[2322] 융합 단백질에 포함될 수 있는 단백질 도메인의 비제한적인 예는 데아미나제 도메인 (예를 들어, 아데노신 데아미나제), 우라실 글리코실라제 저해제 (UGI) 도메인, 에피토프 태그 및 수용체 유전자 서열을 포함한다.

[2323] 에피토프 태그의 비제한적인 예는 히스티딘 (His) 태그, V5 태그, FLAG 태그, 인플루엔자 헤마글루티닌 (HA) 태그, Myc 태그, VSV-G 태그 및 티오레독신 (Trx) 태그를 포함한다. 리포터 유전자의 예는 글루타티온-5-트랜스퍼라제 (GST), 서양고추냉이 퍼옥시다제 (HRP), 클로람페니콜 아세틸트랜스퍼라제 (CAT) 베타-갈락토시다제, 베타-글루쿠로니다제, 루시퍼라제, 녹색 형광 단백질 (GFP), HcRed, DsRed, 시안 형광 단백질 (CFP), 황색 형광 단백질 (YFP), 및 청색 형광 단백질 (BFP)을 포함하는 자가형광 단백질을 포함하지만 이에 제한되지 않는다. 추가의 단백질 서열은 DNA 분자에 결합하거나 말토스 결합 단백질 (MBP), S-태그, Lex A DNA 결합 도메인 (DBD) 융합, GAL4 DNA 결합 도메인 융합, 및 헤르페스 심플렉스 바이러스 (HSV) BP16 단백질 융합을 포함하지만 이에 제한되지 않는 다른 세포 분자에 결합하는 아미노산 서열을 포함할 수 있다.

[2324] 일부 구현예에서, 아데노신 염기 편집기 (ABE)는 DNA에서 아데닌을 탈아민화할 수 있다. 일부 구현예에서, ABE는 BE3의 APOBEC1 성분을 천연 또는 가공된 이. 콜리 TadA, 인간 ADAR2, 마우스 ADA, 또는 인간 ADAT2로 대체함에 의해 생성된다. 일부 구현예에서, ABE는 변화된 TadA 변이체를 포함한다. 일부 구현예에서, ABE는 ABE 1.2 (TadA*-XTEN-nCas9-NLS)이다. 일부 구현예에서, TadA*는 A106V 및 D108N 돌연변이를 포함한다.

[2325] 일부 구현예에서, ABE는 제2 세대 ABE이다. 일부 구현예에서, ABE는 ABE2.1이고, 이는 TadA*에서 추가의 돌연변이 D147Y 및 E155V (TadA*2.1)를 포함한다. 일부 구현예에서, ABE는 촉매 불활성화된 버전의 인간 알킬 아데닌 DNA 글리코실라제 (E125Q 돌연변이를 갖는 AAG)에 융합된 ABE2.1인 ABE2.2이다. 일부 구현예에서, ABE는 촉매 불활성화된 버전의 이. 콜리 Endo V (D35A 돌연변이로 불활성화된)에 융합된 ABE2.1인 ABE2.3이다. 일부 구현예에서, ABE는 ABE2.6이고, 이는 ABE2.1에서의 링커 보다 2배 긴 링커 (32개 아미노산, $(SGGS)_2$-XTEN-$(SGGS)_2$)를 갖는 ABE2.6이다. 일부 구현예에서, ABE는 ABE2.7이고, 이는 추가의 야생형 TadA 단량체로 테더링된 ABE2.1이다. 일부 구현예에서, ABE는 ABE2.8이고, 이는 추가의 TadA *2.1 단량체로 테더링된 ABE2.1이다. 일부 구현예에서, ABE는 ABE2.9이고, 이는 변화된 TadA (TadA*2.1)의 ABE2.1의 N-말단으로의 직접적인 융합체이다. 일부 구현예에서, ABE는 ABE2.10이고, 이는 야생형 TadA의 ABE2.1의 N-말단으로의 직접적인 융합체이다. 일부 구현예에서, ABE는 ABE2.11이고, 이는 TadA* 단량체의 N-말단에서 불활성화 E59A 돌연변이를 갖는 ABE2.9이다. 일부 구현예에서, ABE는 ABE2.12이고, 이는 내부 TadA* 단량체에서 불활성화 E59A 돌연변이를 갖는 ABE2.9이다.

[2326] 일부 구현예에서, ABE는 제3 세대 ABE이다. 일부 구현예에서, ABE는 ABE3.1이고, 이는 3개의 추가의 TadA 돌연변이 (L84F, H123Y, 및 I156F)를 갖는 ABE2.3이다.

[2327] 일부 구현예에서, ABE는 제4 세대 ABE이다. 일부 구현예에서, ABE는 ABE4.3이고, 이는 추가의 TadA 돌연변이 A142N (TadA*4.3)을 갖는 ABE3.1이다.

[2328] 일부 구현예에서, ABE는 제5 세대 ABE이다. 일부 구현예에서, ABE는 ABE5.1이고, 이는 생존 클론 (H36L, R51L, S146C, 및 K157N)으로부터의 돌연변이의 컨센서스 세트를 ABE3.1로 가져옴으로써 생성된다. 일부 구현예에서, ABE는 ABE5.3이고, 이는 내부 변화된 TadA*에 융합된 야생형 이. 콜리 TadA를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE는 아래 **표 8**에 나타낸 바와 같이 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 구현예에서, ABE는 제6 세대 ABE이다. 일부 구현예에서, ABE는 아래 **표 8**에 나타낸 바와 같이 ABE6.1, ABE6.2, ABE6.3, ABE6.4, ABE6.5, 또는 ABE6.6이다. 일부 구현예에서, ABE는 제7 세대 ABE이다. 일부 구현예에서, ABE는 아래 **표 8**에 나타낸 바와 같이 ABE7.1, ABE7.2, ABE7.3, ABE7.4, ABE7.5, ABE7.6, ABE7.7, ABE7.8, ABE 7.9, 또는 ABE7.10이다.

[2329] [표 8] ABE의 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
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

[2330]

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
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

[2331]

[2332] 일부 구현예에서, 염기 편집기는 제8 세대 ABE(ABE8)이다. 일부 구현예에서, ABE8은 TadA*8 변이체를 함유한다. 일부 구현예에서, ABE8은 TadA*8 변이체 ("ABE8.x-m")를 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.1-m이고, 이는 Y147T 돌연변이를 갖는 TadA*7.10 (TadA*8.1)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.2-m이고, 이는 Y147R 돌연변이를 갖는 TadA*7.10 (TadA*8.2)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.3-m이고, 이는 Q154S 돌연변이를 갖는 TadA*7.10 (TadA*8.3)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.4-m이고, 이는 Y123H 돌연변이를 갖는 TadA*7.10 (TadA*8.4)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.5-m이고, 이는 V82S 돌연변이를 갖는 TadA*7.10 (TadA*8.5)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.6-m이고, 이는 T166R 돌연변이를 갖는 TadA*7.10 (TadA*8.6)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.7-m이고, 이는 Q154R 돌연변이를 갖는 TadA*7.10 (TadA*8.7)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.8-m이고, 이는 Y147R, Q154R 및 Y123H 돌연변이를 갖는 TadA*7.10 (TadA*8.8)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.9-m이고, 이는 Y147R, Q154R 및 I76Y 돌연변이를 갖는 TadA*7.10 (TadA*8.9)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.10-m이고, 이는 Y147R, Q154R 및 T166R 돌연변이를 갖는 TadA*7.10 (TadA*8.10)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.11-m이고, 이는 Y147T 및 Q154R 돌연변이를 갖는 TadA*7.10 (TadA*8.11)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.12-m이고, 이는 Y147T 및 Q154S 돌연변이를 갖는 TadA*7.10 (TadA*8.12)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.13-m이고, 이는 Y123H (H123Y로부터 복귀된 Y123H), Y147R, Q154R 및 I76Y 돌연변이를 갖는 TadA*7.10 (TadA*8.13)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.14-m이고, 이는 I76Y 및 V82S 돌연변이를 갖는 TadA*7.10 (TadA*8.14)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.15-m이고, 이는 V82S 및 Y147R 돌연변이를 갖는 TadA*7.10 (TadA*8.15)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.16-m이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), Y147R 돌연변이를 갖는 TadA*7.10 (TadA*8.16)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.17-m이고, 이는 V82S 및 Q154R 돌연변이를 갖는 TadA*7.10 (TadA*8.17)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.18-m이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), 및 Q154R 돌연변이를 갖는 TadA*7.10 (TadA*8.18)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.19-m이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), Y147R 및 Q154R 돌연변이를 갖는 TadA*7.10 (TadA*8.19)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.20-m이고, 이는 I76Y, V82S, Y123H (H123Y로부터 복귀된 Y123H), Y147R 및 Q154R 돌연변이를 갖는 TadA*7.10 (TadA*8.20)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.21-m이고, 이는 Y147R 및 Q154S 돌연변이를 갖는 TadA*7.10 (TadA*8.21)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.22-m이고, 이는 V82S 및 Q154S 돌연변이를 갖는 TadA*7.10 (TadA*8.22)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.23-m이고, 이는 V82S 및 Y123H (H123Y로부터 복귀된 Y123H) 돌연변이를 갖는 TadA*7.10 (TadA*8.23)을 함유하는 단량체성 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.24-m이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), 및 Y147T 돌연변이를 갖는 TadA*7.10 (TadA*8.24)을 함유하는 단량체성 작제물을 갖는다.

[2333] 일부 구현예에서, ABE8은 TadA*8 변이체에 융합된 야생형 이. 콜리 TadA ("ABE8.x-d")를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.1-d이고, 이는 Y147T 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.1)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.2-d이고, 이는 Y147R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.2)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.3-d이고, 이는 Q154S 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.3)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.4-d이고, 이는 Y123H 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.4)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.5-d이고, 이는 V82S 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.5)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.6-d이고, 이는 T166R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.6)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.7-d이고, 이는 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.7)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.8-d이고, 이는 Y147R, Q154R 및 Y123H 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.8)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.9-d이고, 이는 Y147R, Q154R 및 I76Y 돌연

변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.9)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.10-d이고, 이는 Y147R, Q154R 및 T166R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.10)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.11-d이고, 이는 Y147T 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.11)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.12-d이고, 이는 Y147T 및 Q154S 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.12)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.13-d이고, 이는 Y123H (H123Y로부터 복귀된 Y123H), Y147R, Q154R 및 I76Y 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.13)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.14-d이고, 이는 I76Y 및 V82S 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.14)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.15-d이고, 이는 V82S 및 Y147R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.15)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.16-d이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), 및 Y147R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.16)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.17-d이고, 이는 V82S 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.17)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.18-d이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.18)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.19-d이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), Y147 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.19)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.20-d이고, 이는 I76Y, V82S, Y123H (H123Y로부터 복귀된 Y123H), Y147R 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.20)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.21-d이고, 이는 Y147R 및 Q154S 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.21)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.22-d이고, 이는 V82S 및 Q154S 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.22)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.23-d이고, 이는 V82S 및 Y123H (H123Y로부터 복귀된 Y123H)돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.23)를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.24-d이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), 및 Y147T 돌연변이를 갖는 TadA*7.10에 융합된 야생형 이. 콜리 TadA (TadA*8.24)를 함유하는 이종이량체 작제물을 갖는다.

[2334] 일부 구현예에서, ABE8은 TadA*8 변이체에 융합된 TadA*7.10 ("ABE8.x-7")을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.1-7이고, 이는 Y147T 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.1)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.2-7이고, 이는 Y147R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.2)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.3-7이고, 이는 Q154S 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.3)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.4-7이고, 이는 Y123H 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.4)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.5-7이고, 이는 V82S 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.5)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.6-7이고, 이는 T166R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.6)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.7-7이고, 이는 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.7)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.8-7이고, 이는 Y147R, Q154R 및 Y123H 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.8)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.9-7이고, 이는 Y147R, Q154R 및 I76Y 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.9)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.10-7이고, 이는 Y147R, Q154R 및 T166R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.10)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.11-7이고, 이는 Y147T 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.11)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.12-7이고, 이는 Y147T 및 Q154S 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.12)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.13-7, 이는 Y123H (H123Y로부터 복귀된 Y123H), Y147R, Q154R 및 I76Y 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.13)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.14-7이고, 이는 I76Y 및 V82S 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.14)을 함유하는 이

종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.15-7이고, 이는 V82S 및 Y147R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.15)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.16-7이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H) 및 Y147R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.16)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.17-7이고, 이는 V82S 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.17)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.18-7이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H) 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.18)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.19-7이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H), Y147R 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.19)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.20-7이고, 이는 I76Y, V82S, Y123H (H123Y로부터 복귀된 Y123H), Y147R 및 Q154R 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.20)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.21-7이고, 이는 Y147R 및 Q154S 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.21)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.22-7이고, 이는 V82S 및 Q154S 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.22)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.23-7이고, 이는 V82S 및 Y123H (H123Y로부터 복귀된 Y123H) 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.23)을 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE8은 ABE8.24-7이고, 이는 V82S, Y123H (H123Y로부터 복귀된 Y123H) 및 Y147T 돌연변이를 갖는 TadA*7.10에 융합된 TadA*7.10 (TadA*8.24)을 함유하는 이종이량체 작제물을 갖는다.

[2335] 일부 구현예에서, ABE는 아래 **표 9**에 나타낸 바와 같이 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.

[2336] [표 9] 아데노신 데아미나제 염기 편집기 8 변이체

ABE8	아데노신 데아미나제	아데노신 데아미나제 기재사항
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
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)

[2337]

ABE8	아데노신 데아미나제	아데노신 데아미나제 기재사항
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)
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)

[2338]

[2339] 일부 구현예에서, 염기 편집기 (예를 들어, ABE8)는 아데노신 데아미나제 변이체 (예를 들어, TadA*8)를 환형의 퍼뮤턴트 Cas9 (예를 들어, CP5 또는 CP6) 및 양분된 핵 국소화 서열을 포함하는 스캐폴드에 클로닝함에 의해 생성된다. 일부 구현예에서, 염기 편집기 (예를 들어, ABE7.9, ABE7.10, 또는 ABE8)는 NGC PAM CP5 변이체 (에스. 피로게네스 (*S. pyrogenes*) Cas9 또는 spVRQR Cas9)이다. 일부 구현예에서, 염기 편집기 (예를 들어, ABE7.9, ABE7.10, 또는 ABE8)는 AGA PAM CP5 변이체 (에스. 피로게네스 Cas9 또는 spVRQR Cas9)이다. 일부 구현예에서, 염기 편집기 (예를 들어, ABE7.9, ABE7.10, 또는 ABE8)는 NGC PAM CP6 변이체 (에스. 피로게네스 Cas9 또는 spVRQR Cas9)이다. 일부 구현예에서, 염기 편집기 (예를 들어, ABE7.9, ABE7.10, 또는 ABE8)는 AGA PAM CP6 변이체 (에스. 피로게네스 Cas9 또는 spVRQR Cas9)이다.

[2340] 일부 구현예에서, ABE는 아래 **표 10**에 나타낸 바와 같은 유전자형을 갖는다.

[2341] [표 10] ABE의 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

[2342]

[2343] 아래 **표 11**에 나타낸 바와 같이, 40개 ABE8의 유전자형이 기재된다. ABE의 변화된 이. 콜리 TadA 부분에서 잔기 위치가 지적된다. ABE8에서 돌연변이 변화는 ABE7.10 돌연변이로부터 구분되는 경우 나타낸다. 일부 구현예에서, ABE는 아래 **표 11**에 나타낸 바와 같은 ABE들 중 하나의 유전자형을 갖는다.

[2344] [표 11] 변화된 TadA에서 잔기 동일성

	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								

[2345]

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				
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						

[2346]

[2347] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.1이다:

[2348] ABE8.1_Y147T_CP5_NGC PAM_단량체

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSESATPESSGGS***
SGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAK
FLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLI
HQSITGLYETRIDLSQLGGD***GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY
LALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE
ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[2349]

[2350] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭한다.

[2351] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.1이다:

[2352] pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_단량체

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSG*GSSGSETPGTSESATPESSGGS*
*SGGS*EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAK
FLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLI
HQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVITDEY
KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEM
AKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY
LALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE
ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQT
VKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[2353]

[2354] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭한다.

[2355] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.14이다:

[2356] NGC PAM CP5를 갖는 pNMG-357_ABE8.14

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDGGSSGGS*SGSETPGTSESATPESSGGSS*
*GGS*MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA
SAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS
NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR
LIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS
KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI
GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI
LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE
NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKS
DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[2357]

[2358] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭한다.

[2359] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.8-m이다:

[2360] ABE8.8-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGS*
*SGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD*EGADKRTADGSEFESPKKKRKV**

[2361]

[2362] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭하고, 두 줄 밑줄 친 서열은 돌연변이를 지적한다.

[2363] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.8-d이다:

[2364] ABE8.8-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGS*
*SGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMN
HRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL
YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ
KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ
KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKD
FQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALP
SKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[2365]

[2366] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭하고, 두 줄 밑줄 친 서열은 돌연변이를 지적한다.

[2367] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.13-m이다:

[2368] ABE8.13-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGS*
*SGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV*

[2369]

[2370] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭하고, 두 줄 밑줄 친 서열은 돌연변이를 지적한다.

[2371] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.13-d이다:

[2372] ABE8.13-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGS*
*SGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMN
HRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL
YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ
KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ
KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKD
FQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALP
SKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[2373]

[2374] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭하고, 두 줄 밑줄 친 서열은 돌연변이를 지적한다.

[2375] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.17-m이다:

[2376] ABE8.17-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGS*
*SGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGD*EGADKRTADGSEFESPKKKRKV**

[2377]

[2378] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭하고, 두 줄 밑줄 친 서열은 돌연변이를 지적한다.

[2379] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.17-d이다:

[2380] ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGS*
*SGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL
YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ
KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ
KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKD
FQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALP
SKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[2381]

[2382] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭하고, 두 줄 밑줄 친 서열은 돌연변이를 지적한다.

[2383] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.20-m이다:

[2384] ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGS*
*SGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV*

[2385]

[2386] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭하고, 두 줄 밑줄 친 서열은 돌연변이를 지적한다.

[2387] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열 또는 이의 단편을 포함하거나 이로 본질적으로 이루어진 ABE8.20-d이다:

[2388] ABE8.20-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGS*
*SGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMN
HRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESS*
*GGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS
NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL
YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE
ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ
KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ
KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKD
FQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALP
SKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[2389]

[2390] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 양분된 핵 국소화 서열을 지칭하고, 두 줄 밑줄 친 서열은 돌연변이를 지적한다.

[2391] 일부 구현예에서, 본원 개시내용의 ABE8은 하기의 서열로부터 선택된다:

[2392] 01. 모노ABE8.1_bpNLS + Y147T

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2393]

[2394] 02. 모노ABE8.1_bpNLS + Y147R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCRFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2395]

[2396] 03. 모노ABE8.1_bpNLS + Q154S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2397]

[2398] 04. 모노ABE8.1_bpNLS + Y123H

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
[2399] SQLGGDEGADKRTADGSEFESPKKKRKV

[2400] 05. 모노ABE8.1_bpNLS + V82S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2401]

[2402] 06. 모노ABE8.1_bpNLS + T166R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2403]

[2404] 07. 모노ABE8.1_bpNLS + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2405]

[2406] 08. 모노ABE8.1_bpNLS + Y147R_Q154R_Y123H

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2407]

[2408] 09. 모노ABE8.1_bpNLS + Y147R_Q154R_I76Y

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2409]

[2410] 10. 모노ABE8.1_bpNLS + Y147R_Q154R_T166R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCRFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2411]

[2412] 11. 모노ABE8.1_bpNLS + Y147T_Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCTFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2413]

[2414] 12. 모노ABE8.1_bpNLS + Y147T_Q154S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCTFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2415]

[2416] 13. 모노ABE8.1_bpNLS + H123Y123H_Y147R_Q154R_I76Y

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRV
EITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2417]

[2418] 14. 모노ABE8.1_bpNLS + V82S + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQ
GGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV
EITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGS
SGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI
KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYV
GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNS
RERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV
PQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDL
SQLGGDEGADKRTADGSEFESPKKKRKV

[2419]

[2420] 일부 구현예에서, 염기 편집기는 핵염기 편집 도메인 (예를 들어, 데아미나제 도메인의 전부 또는 일부)에 융합된 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 (예를 들어, Cas9 유래된 도메인)을 포함하는 융합 단백질이다. 특정 구현예에서, 본원에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성을 개선시키는 하나 이상의 특성을 포함한다. 예를 들어, 본원에 제공된 임의의 융합 단백질은 감소된 뉴클레아제 활성을 갖는 Cas9 도메인을 포함할 수 있다. 일부 구현예에서, 본원에 제공된 임의의 융합 단백질은 뉴클레아제 활성을 갖지 않는 Cas9 도메인 (dCas9), 또는 듀플렉스 DNA 분자의 하나의 가닥을 절단하는, Cas9 닉카제로서 언급되는 Cas9 도메인 (nCas9)을 가질 수 있다.

[2421] 일부 구현예에서, 염기 편집기는 우라실 글리코실라제 저해제 (UGI)의 전부 또는 일부를 포함하는 도메인을 추가로 포함한다. 일부 구현예에서, 염기 편집기는 우라실 DNA 글리코실라제 (UDG)와 같은 우라실 결합 단백질 (UBP)의 전부 또는 일부를 포함하는 도메인을 포함한다. 일부 구현예에서, 염기 편집기는 핵산 폴리머라제의 전부 또는 일부를 포함하는 도메인을 포함한다. 일부 구현예에서, 염기 편집기에 혼입된 핵산 폴리머라제 또는 이의 일부는 트랜스레젼 DNA 폴리머라제이다.

[2422] 일부 구현예에서, 염기 편집기의 도메인은 다중 도메인을 포함할 수 있다. 예를 들어, Cas9로부터 유래된 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기는 REC 돌출부 및 상기 REC 돌출부에 상응하는 NUC 돌출부 및 야생형 또는 천연 Cas9의 NUC 돌출부를 포함한다. 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 구현예에서, 염기 편집기의 하나 이상의 도메인은 도메인을 포함하는 폴리펩타이드의 야생형 버전에 비해 돌연변이 (예를 들어, 치환, 삽입, 결실)를 포함한다. 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인의 HNH 도메인은 H840A 치환을 포함할 수 있다. 또 다른 예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인의 RuvCI 도메인은 D10A 치환을 포함할 수 있다.

[2423] 본원에 기재된 염기 편집기의 상이한 도메인 (예를 들어, 인접한 도메인)은 하나 이상의 링커 도메인 (예를 들어, XTEN 링커 도메인)을 사용하거나 사용하지 않고 서로 연결될 수 있다. 일부 구현예에서, 링커 도메인은 결합 (예를 들어, 공유 결합), 화학적 그룹, 또는 2개의 분자 또는 모이어티를 연결하는 분자, 예를 들어, 융합 단백질의 2개의 도메인, 예를 들어, 제1 도메인 (예를 들어, Cas9-유래된 도메인) 및 제2 도메인 (예를 들어, 아데노신 데아미나제 도메인)일 수 있다. 일부 구현예에서, 링커는 공유 결합 (예를 들어, 탄소-탄소 결합, 디설파이드 결합, 탄소-헤테로원자 결합 등)이다. 특정 구현예에서, 링커는 아미드 연결의 탄소 질소 결합이다. 특정 구현예에서, 링커는 환식 또는 비환식, 치환된 또는 비치환된, 분지된 또는 비분지된 지방족 또는 헤테로지방족 링커이다. 특정 구현예에서, 링커는 중합체 (예를 들어, 폴리에틸렌, 폴리에틸렌 글리콜, 폴리아미드, 폴리에스테르 등)이다. 특정 구현예에서, 링커는 단량체, 이량체 또는 아미노알칸산의 중합체를 포함한다. 일부 구현예에서, 링커는 아미노알칸산(예를 들어, 글라이신, 에탄산, 알라닌, 베타-알라닌, 3-아미노프로판산, 4-아미노부탄산, 5-펜탄산 등)을 포함한다. 일부 구현예에서, 링커는 단량체, 이량체 또는 아미노헥산산 (Ahx)의 중합체를 포함한다. 특정 구현예에서, 링커는 카보사이클릭 모이어티 (예를 들어, 사이클로펜탄, 사이클로헥산)를 기반으로 한다. 다른 구현예에서, 링커는 폴리에틸렌 글리콜 모이어티 (PEG)를 포함한다. 특정 구현예에서, 링커는 아릴 또는 헤테로아릴 모이어티를 포함한다. 특정 구현예에서, 링커는 페닐 환을 기반으로 한다. 링커는 펩타이드로부터의 친핵체 (예를 들어, 티올, 아미노)의 링커로의 부착을 촉진시키기 위해 기능성화된 모이어티를 포함할 수 있다. 임의의 친전자체는 링커의 일부로서 사용될 수 있다. 예시적인 친전자체는 활성화된 에스테르, 활성화된 아미드, 마이클 수용체, 알킬 할라이드, 아릴 할라이드, 아실 할라이드 및 이소티오시아네이트를 포함하지만 이에 제한되지 않는다. 일부 구현예에서, 링커는 Cas9 뉴클레아제 도메인, 및 핵산 편집 단백질의 촉매 도메인을 포함하는 RNA-프로그래밍 가능한 뉴클레아제의 gRNA 결합 도메인을 연결한다. 일부 구현예에서, 링커는 dCas9와 제2 도메인 (예를 들어, UGI 등)을 연결한다.

[2424] 전형적으로, 링커는 2개의 그룹, 분자 또는 다른 모이어티 사이에 위치하거나 이에 의해 플랭킹되고 공유 결합을 통해 각각 하나에 연결됨에 따라 2개를 연결한다. 일부 구현예에서, 링커는 아미노산 또는 다수의 아미노산 (예를 들어, 펩타이드 또는 단백질)이다. 일부 구현예에서, 링커는 유기 분자, 그룹, 중합체, 또는 화학적 모이어티이다. 일부 구현예에서, 링커는 2-100개 아미노산 길이, 예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 또는 150-200개 아미노산 길이이다. 일부 구현예에서, 링커는 약 3 내지 약 104개 (예를 들어, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 또는 100개) 아미노산 길이이다. 더 길거나 더 짧은 링커가 또한 고려된다. 일부 구현예에서, 링커 도메인은 또한 XTEN 링커로서 언급될 수 있는 아미노산 서열 SGSETPGTSESATPES를 포함한다. 융합 단백질 도메인을 연결하기 위한 임의의 방법 (예를 들어, 매우 가요성 형태의 링커 $(SGGS)_n$, $(GGGS)_n$, $(GGGGS)_n$, 및 $(G)_n$으로부터 보다 강성 형태의 링커 $(EAAAK)_n$, $(GGS)_n$, SGSETPGTSESATPES (예를 들어, 문헌 (Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82) 참조; 이의 전체 내용은 본원에 참조로 포함됨) 또는 $(XP)_n$ 모티프 범위)을 사용하여 핵염기 편집기에 대한 활성을 위한 최적의 길이를 성취할 수 있다. 일부 구현예에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15이다. 일부 구현예에서, 링커는 $(GGS)_n$ 모티프를 포함하고, 여기서, n은 1, 3, 또는 7이다. 일부 구현예에서, 본원에 제공된 융합 단백질의 Cas9 도메인은 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다. 일부 구현예에서, 링커는 다수의 프롤린 잔기를 포함하고, 5-21, 5-14, 5-9, 5-7개 아미노산 길이, 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; 이의 전체 내용은 본원에 참조로 포함됨). 상기 프롤린-풍부 링커는 또한 "강성" 링커로 칭한다.

[2425] 본원에 기재된 융합 단백질은 핵산 편집 도메인을 포함한다. 일부 구현예에서, 데아미나제는 아데노신 데아미나제이다. 일부 구현예에서, 데아미나제는 척추동물 데아미나제이다. 일부 구현예에서, 데아미나제는 무척추동물 데아미나제이다. 일부 구현예에서, 데아미나제는 인간, 침팬지, 고릴라, 붕키, 소, 개, 래트, 또는 마우스 데아미나제이다. 일부 구현예에서, 데아미나제는 인간 데아미나제이다. 일부 구현예에서, 데아미나제는 래트 데아미나제이다.

[2426] *링커*

[2427] 특정 구현예에서, 링커는 본원에 기재된 임의의 펩타이드 또는 펩타이드 도메인을 연결하기 위해 사용될 수 있다. 링커는 공유 결합만큼 단순할 수 있거나 이것은 많은 원자 길이의 중합체 링커일 수 있다. 특정 구현예에서, 링커는 폴리펩타이드이거나 아미노산을 기반으로 한다. 다른 구현예에서, 링커는 펩타이드와 유사하지 않다. 특정 구현예에서, 링커는 공유 결합 (예를 들어, 탄소-탄소 결합, 디설파이드 결합, 탄소-헤테로원자 결합 등)이다. 특정 구현예에서, 링커는 아미드 연결의 탄소-질소 결합이다. 특정 구현예에서, 링커는 환식 또는 비환식, 치환된 또는 비치환된, 분지된 또는 비분지된 지방족 또는 헤테로지방족 링커이다. 특정 구현예에서, 링커는 중합체 (예를 들어, 폴리에틸렌, 폴리에틸렌 글리콜, 폴리아미드, 폴리에스테르 등)이다. 특정 구현예에서, 링커는 단량체, 이량체 또는 아미노알칸산의 중합체를 포함한다. 특정 구현예에서, 링커는 아미노알칸산 (예를 들어, 글라이신, 에탄산, 알라닌, 베타-알라닌, 3-아미노프로판산, 4-아미노부탄산, 5-펜탄산 등)을 포함한다. 특정 구현예에서, 링커는 단량체, 이량체 또는 아미노헥산산의 중합체 (Ahx)를 포함한다. 특정 구현예에서, 링커는 카보사이클릭 모이어티 (예를 들어, 사이클로펜탄, 사이클로헥산)를 기반으로 한다. 다른 구현예에서, 링커는 폴리에틸렌 글리콜 모이어티 (PEG)를 포함한다. 다른 구현예에서, 링커는 아미노산을 포함한다. 특정 구현예에서, 링커는 펩타이드를 포함한다. 특정 구현예에서, 링커는 아릴 또는 헤테로아릴 모이어티를 포함한다. 특정 구현예에서, 링커는 페닐 환을 기반으로 한다. 링커는 펩타이드 기원의 친핵체 (예를 들어, 티올, 아미노)의 링커로의 접착을 촉진시키는 기능성화된 모이어티를 포함할 수 있다. 임의의 친전자체는 링커의 일부로서 사용될 수 있다. 예시적인 친전자체는 활성화된 에스테르, 활성화된 아미드, 마이클 수용체, 알킬 할라이드, 아릴 할라이드, 아실 할라이드 및 이소티오시아네이트를 포함하지만 이에 제한되지 않는다.

[2428] 일부 구현예에서, 링커는 아미노산 또는 다수의 아미노산 (예를 들어, 펩타이드 또는 단백질)이다. 일부 구현예에서, 링커는 결합 (예를 들어, 공유 결합), 유기 분자, 그룹, 중합체 또는 화학적 모이어티이다. 일부 구현예에서, 링커는 약 3 내지 약 104개 (예를 들어, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 또는 100개) 아미노산 길이이다.

[2429] 일부 구현예에서, 아데노신 데아미나제 및 napDNAbp는 4, 16, 32, 또는 104개 아미노산 길이인 링커를 통해 융합된다. 일부 구현예에서, 링커는 약 3 내지 약 104개 아미노산 길이이다. 일부 구현예에서, 본원에 제공된 임의의 융합 단백질은 아데노신 데아미나제 및 Cas9 도메인을 포함하고 이들은 서로 링커를 통해 융합되어 있다. 데아미나제 도메인 (예를 들어, 가공된 ecTadA)와 Cas9 도메인 간에 다양한 링커 길이 및 가요성 (예를 들어, 매우 가요성 형태의 링커$(GGGS)_n$, $(GGGGS)_n$, 및 $(G)_n$으로부터 보다 강성 형태의 링커 $(EAAAK)_n$, $(SGGS)_n$, SGSETPGTSESATPES (예를 들어, 문헌 (Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82) 참조; 이의 전체 내용은 본원에 참조로 포함됨) 및 $(XP)_n$ 범위)을 사용하여 핵염기 편집기에 대한 활성을 위한 최적의 길이를 성취할 수 있다. 일부 구현예에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15이다. 일부 구현예에서, 링커는 $(GGS)_n$ 모티프를 포함하고, 여기서, n은 1, 3, 또는 7이다. 일부 구현예에서, 본원에 제공된 임의의 융합 단백질의 아데노신 데아미나제 및 Cas9 도메인은 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커 (예를 들어, XTEN 링커)를 통해 융합된다.

[2430] *가이드 RNA와의 Cas9 복합체*

[2431] 본원 개시내용의 일부 양상은 본원에 제공된 임의의 융합 단백질, 및 융합 단백질의 Cas9 도메인 (예를 들어, dCas9, 뉴클레아제 활성 Cas9, 또는 Cas9 닉카제)에 결합된 가이드 RNA (예를 들어, RTT 표적 가능한 돌연변이를 보유하는 Mecp2 대립유전자를 표적화하는 가이드)를 포함하는 복합체를 제공한다. 이러한 복합체는 리보핵단백질 (RNP)로도 칭한다. 융합 단백질 도메인을 연결하기 위한 임의의 방법 (예를 들어, 매우 가요성 형태의 링커 $(GGGS)_n$, $(GGGGS)_n$, 및 $(G)_n$으로부터 보다 강성 형태의 링커 $(EAAAK)_n$, $(SGGS)_n$, SGSETPGTSESATPES (예를 들어, 문헌 (Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82) 참조; 이의 전체 내용은 본원에 참조로 포함됨) 및 $(XP)_n$의 범위)을 사용하여 핵염기 편집기에 대한 활성을 위한 최적의 길이를 성취할 수 있다. 일부 구현예에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15이다. 일부 구현예에서, 링커는 $(GGS)_n$ 모티프를 포함하고, 여기서, n은 1, 3, 또는 7이다. 일부 구현예에서, 본원에 제공

된 융합 단백질의 Cas9 도메인은 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다.

[2432] 일부 구현예에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 15-100개 뉴클레오타이드 길이이고 표적 서열에 상보적인 적어도 10개 인접 뉴클레오타이드의 서열을 포함한다. 일부 구현예에서, 가이드 RNA는 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 또는 50개 뉴클레오타이드 길이이다. 일부 구현예에서, 가이드 RNA는 표적 서열과 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 또는 40개 인접 뉴클레오타이드의 서열을 포함한다. 일부 구현예에서, 표적 서열은 DNA 서열이다. 일부 구현예에서, 표적 서열은 세균, 효모, 진균류, 곤충, 식물 또는 동물의 게놈에서의 서열이다. 일부 구현예에서, 표적 서열은 인간의 게놈에서의 서열이다. 일부 구현예에서, 표적 서열의 3' 말단은 카노니칼 PAM 서열 (NGG)에 바로 인접해 있다. 일부 구현예에서, 표적 서열의 3' 말단은 비-카노니칼 PAM 서열 (예를 들어, 표 1 또는 5'-NAA-3'에 열거된 서열)에 바로 인접해 있다. 일부 구현예에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 RTT 표적 가능한 돌연변이를 보유하는 Mecp2 대립유전자의 서열에 상보적이다.

[2433] 본원 개시내용의 일부 양상은 본원에 제공된 융합 단백질 또는 복합체를 사용하는 방법을 제공한다. 예를 들어, 본원 개시내용의 일부 양상은 DNA 분자를 본원에 제공된 임의의 융합 단백질 및 적어도 하나의 가이드 RNA와 접촉시킴을 포함하는 방법을 제공하고, 여기서, 가이드 RNA는 약 15-100개 뉴클레오타이드 길이이고 표적 서열과 상보적인 적어도 10개 인접 뉴클레오타이드의 서열을 포함한다. 일부 구현예에서, 표적 서열의 3' 말단은 AGC, GAG, TTT, GTG, 또는 CAA 서열에 바로 인접해 있다. 일부 구현예에서, 표적 서열의 3' 말단은 NGA, NAA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, 또는 5' (TTTV) 서열에 바로 인접해 있다.

[2434] 각각의 서열에서 특정 위치 또는 잔기의 넘버링은 사용되는 특정 단백질 및 넘버링 기획에 의존하는 것으로 이해될 것이다. 넘버링은 예를 들어, 성숙한 단백질의 전구체 및 성숙한 단백질 자체에서 상이할 수 있고 종에 따른 서열에서의 차이는 넘버링에 영향을 미칠 수 있다. 당업자는 당업계에 널리 공지된 방법에 의해, 예를 들어, 서열 정렬 및 상동성 잔기의 결정에 의해 임의의 상동성 단백질 및 각각의 암호화 핵산에서 각각의 잔기를 동정할 수 있을 것이다.

[2435] 본원에 개시된 바와 같이, Cas9 도메인 및 아데노신 데아미나제 변이체 (예를 들어, TadA*8)를 포함하는 임의의 융합 단백질을 표적 부위, 예를 들어, 편집될 돌연변이를 포함하는 부위에 표적화하기 위해, 가이드 RNA, 예를 들어, sgRNA와 함께 융합 단백질을 전형적으로 동시 발현시킬 필요가 있다는 것은 당업자에게 자명할 것이다. 본원의 다른 곳에서 보다 상세하게 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 구현예에서, 가이드 RNA는 상기 가이드 서열이 표적 서열에 상보적인 서열을 포함하는 구조를 포함한다. 가이드 서열은 전형적으로 20개 뉴클레오타이드 길이다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특이적 게놈 표적 부위에 표적화하기 위해 적합한 가이드 RNA의 서열은 본원의 개시내용을 토대로 당업자에게 자명할 것이다. 상기 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오타이드의 업스트림 또는 다운스트림 50개 뉴클레오타이드 내의 핵산 서열에 상보적인 가이드 서열을 포함한다. 임의의 제공된 융합 단백질을 특이적 표적 서열에 표적화하기에 적합한 일부 예시적인 가이드 RNA 서열은 본원에 제공된다.

[2436] 본원에 개시된 염기 편집기의 도메인은 임의의 순서로 정렬될 수 있다. 예를 들어, 폴리뉴클레오타이드-프로그래밍 가능한 뉴클레오타이드-결합 도메인 및 데아미나제 도메인을 포함하는 융합 단백질을 포함하는 염기 편집기의 비제한적인 예는 다음과 같이 정렬될 수 있다:

[2437] NH_2-[핵염기 편집 도메인]-링커1-[예를 들어, Cas9 유래된 도메인]-COOH;

[2438] NH_2-[예를 들어, 아데노신 데아미나제]-링커1-[예를 들어, Cas9 유래된 도메인]-COOH;

[2439] NH_2-[예를 들어, 아데노신 데아미나제]-링커1-[예를 들어, Cas9 유래된 도메인]-링커2-[UGI]-COOH;

[2440] NH_2-[예를 들어, TadA*7.10]-링커1-[예를 들어, Cas9 유래된 도메인]-COOH;

[2441] NH_2-[예를 들어, 아데노신 데아미나제]-링커1-[예를 들어, Cas9 유래된 도메인]-COOH;

[2442] NH_2-[예를 들어, TadA*7.10]-링커1-[예를 들어, Cas9 유래된 도메인]-COOH;

[2443] NH_2-[예를 들어, TadA*7.10]-링커1-[예를 들어, Cas9 유래된 도메인]-링커2-[UGI]-COOH

[2444] NH_2-[예를 들어, 아데노신 데아미나제]-[예를 들어, Cas9 유래된 도메인]-COOH;

[2445] NH_2-[예를 들어, Cas9 유래된 도메인]-[예를 들어, 아데노신 데아미나제]-COOH;

[2446] NH_2-[예를 들어, 아데노신 데아미나제]-[예를 들어, Cas9 유래된 도메인]-[이노신 BER 저해제]-COOH;

[2447] NH_2-[예를 들어, 아데노신 데아미나제]-[이노신 BER 저해제]-[예를 들어, Cas9 유래된 도메인]-COOH;

[2448] NH_2-[이노신 BER 저해제]-[예를 들어, 아데노신 데아미나제]-[예를 들어, Cas9 유래된 도메인]-COOH;

[2449] NH_2-[예를 들어, Cas9 유래된 도메인]-[예를 들어, 아데노신 데아미나제]-[이노신 BER 저해제]-COOH;

[2450] NH_2-[예를 들어, Cas9 유래된 도메인]-[이노신 BER 저해제]-[예를 들어, 아데노신 데아미나제]-COOH; 또는

[2451] NH_2-[이노신 BER 저해제]-[예를 들어, Cas9 유래된 도메인]-[예를 들어, 아데노신 데아미나제]-COOH.

[2452] 추가로, 일부 구현예에서, Gam 단백질은 염기 편집기의 N 말단에 융합될 수 있다. 일부 구현예에서, Gam 단백질은 염기 편집기의 C 말단에 융합될 수 있다. 박테리오파아지 Mu의 Gam 단백질은 이중 가닥 절단 (DSB)의 말단에 결합하고 이들을 분해로부터 보호할 수 있다. 일부 구현예에서, DSB의 유리된 말단에 결합하는 Gam을 사용하여 염기 편집 과정 동안에 삽입-결실 형성을 감소시킬 수 있다. 일부 구현예에서, 174개 잔기 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))을 참조한다. 일부 구현예에서, 돌연변이 또는 돌연변이들은 야생형 도메인에 비해 염기 편집기 도메인의 길이를 변화시킬 수 있다. 예를 들어, 적어도 하나의 도메인에서 적어도 하나의 아미노산의 결실은 염기 편집기의 길이를 감소시킬 수 있다. 또 다른 경우에, 돌연변이 또는 돌연변이들은 야생형 도메인에 비해 도메인의 길이를 변화시키지 않는다. 예를 들어, 임의의 도메인에서 치환은 염기 편집기의 길이를 변화시키지 않는다.

[2453] 일부 구현예에서, 본원에 제공된 염기 편집 융합 단백질은 예를 들어, 표적 염기가 한정된 영역 (예를 들어, "탈아민화 윈도우") 내에 위치하는 정확한 위치에 위치할 필요가 있다. 일부 구현예에서, 표적은 4개 염기 영역 내에 있을 수 있다. 일부 구현예에서, 상기 한정된 표적 영역은 PAM의 대략적으로 15개 염기 업스트림에 있을 수 있다. 문헌 (Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); 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); and 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))을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[2454] 한정된 표적 영역은 탈아민화 윈도우일 수 있다. 탈아민화 윈도우는 염기 편집기가 작용하고 표적 뉴클레오타이드를 탈아민화시키는 한정된 영역일 수 있다. 일부 구현예에서, 탈아민화 윈도우는 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 염기 영역 내에 있다. 일부 구현예에서, 탈아민화 윈도우는 PAM의 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개 염기 업스트림에 있다.

[2455] 본원 개시내용의 염기 편집기는 표적 폴리뉴클레오타이드 서열의 편집을 촉진시키는, 임의의 도메인, 특성 또는 아미노산 서열을 포함할 수 있다. 예를 들어, 일부 구현예에서, 염기 편집기는 핵 국소화 서열 (NLS)을 포함한다. 일부 구현예에서, 염기 편집기의 NLS는 데아미나제 도메인과 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 사이에 위치한다. 일부 구현예에서, 염기 편집기의 NLS는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인의 C-말단에 위치한다.

[2456] 융합 단백질에 포함될 수 있는 단백질 도메인의 비제한적인 예는 데아미나제 도메인 (예를 들어, 아데노신 데아미나제), 우라실 글리코실라제 저해제 (UGI) 도메인, 에피토프 태그, 수용체 유전자 서열 및/또는 하기의 활성 중 하나 이상을 갖는 단백질 도메인을 포함한다: 메틸라제 활성, 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, RNA 절단 활성, 및 핵산 결합 활성. 추가의 도메인은 이종 기능성 도메인일 수 있다. 상기 이종 기능성 도메인은 DNA 메틸화, DNA 손상, DNA 복구, 표적 DNA와 연합된 표

적 폴리펩타이드 (예를 들어, 히스톤, DNA-결합 단백질 등)의 변형과 같은 기능 활성을 부여하여 예를 들어, 히스톤 메틸화, 히스톤 아세틸화, 히스톤 유비퀴틴화 등을 유도할 수 있다.

[2457] 부여된 다른 기능은 메틸트랜스퍼라제 활성, 데메틸라제 활성, 탈아민화 활성, 디스뮤타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 리컴비나제 활성, 폴리머라제 활성, 리가제 활성, 헬리카제 활성, 포토리아제 활성 또는 글리코실라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스파타제 활성, 유비퀴틴 리가제 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모일화 (SUMOylation) 활성, 탈수모일화 (deSUMOylation) 활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성, 리모델링 활성, 프로테아제 활성, 옥시도리덕타제 활성, 트랜스퍼라제 활성, 하이드롤라제 활성, 리아제 활성, 이소머라제 활성, 신타제 활성, 신테타제 활성, 및 탈미리스토일화 활성, 또는 이들의 임의의 조합을 포함할 수 있다.

[2458] 에피토프 태그의 비제한적인 예는 히스티딘 (His) 태그, V5 태그, FLAG 태그, 인플루엔자 헤마글루티닌 (HA) 태그, Myc 태그, VSV-G 태그 및 티오레독신 (Trx) 태그를 포함한다. 리포터의 예는 글루타티온-5-트랜스퍼라제 (GST), 서양고추냉이 퍼옥시다제 (HRP), 클로람페니콜 아세틸트랜스퍼라제 (CAT) 베타-갈락토시다제, 베타-글루쿠로니다제, 루시퍼라제, 녹색 형광 단백질 (GFP), HcRed, DsRed, 시안 형광 단백질 (CFP), 황색 형광 단백질 (YFP), 및 청색 형광 단백질 (BFP)을 포함하는 자가형광 단백질을 포함하지만 이에 제한되지 않는다. 추가의 단백질 서열은 DNA 분자에 결합하거나 말토스 결합 단백질 (MBP), S-태그, Lex A DNA 결합 도메인 (DBD) 융합, GAL4 DNA 결합 도메인 융합, 및 헤르페스 심플렉스 바이러스 (HSV) BP16 단백질 융합을 포함하지만 이에 제한되지 않는 다른 세포 분자에 결합하는 아미노산 서열을 포함할 수 있다.

[2459] ***가이드 RNA와의 Cas12 복합체***

[2460] 본원 개시내용의 일부 양상은 본원에 제공된 임의의 융합 단백질 및 가이드 RNA (예를 들어, 편집을 위해 표적 폴리뉴클레오타이드를 표적화하는 가이드)를 포함하는 복합체를 제공한다.

[2461] 일부 구현예에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 15-100개 뉴클레오타이드 길이이고 표적 서열에 상보적인 적어도 10개 인접 뉴클레오타이드의 서열을 포함한다. 일부 구현예에서, 가이드 RNA는 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 또는 50개 뉴클레오타이드 길이이다. 일부 구현예에서, 가이드 RNA는 표적 서열과 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 또는 40개 인접 뉴클레오타이드의 서열을 포함한다. 일부 구현예에서, 표적 서열은 DNA 서열이다. 일부 구현예에서, 표적 서열은 세균, 효모, 진균류, 곤충, 식물 또는 동물의 게놈에서의 서열이다. 일부 구현예에서, 표적 서열은 인간의 게놈에서의 서열이다. 일부 구현예에서, 표적 서열의 3' 말단은 카노니칼 PAM 서열에 바로 인접해 있다. 일부 구현예에서, 표적 서열의 3' 말단은 비-카노니칼 PAM 서열에 바로 인접해 있다.

[2462] 본원 개시내용의 일부 양상은 본원에 제공된 융합 단백질 또는 복합체를 사용하는 방법을 제공한다. 예를 들어, 본원 개시내용의 일부 양상은 DNA 분자를 본원에 제공된 임의의 융합 단백질 및 적어도 하나의 가이드 RNA와 접촉시키는 단계를 포함하는 방법을 제공하고, 여기서, 가이드 RNA는 약 15-100개 뉴클레오타이드 길이이고 표적 서열과 상보적인 적어도 10개 인접 뉴클레오타이드의 서열을 포함한다. 일부 구현예에서, 표적 서열의 3' 말단은 예를 들어, TTN, DTTN, GTTN, ATTN, ATTC, DTTNT, WTTN, HATY, TTTN, TTTV, TTTC, TG, RTR, 또는 YTN PAM 부위에 바로 인접해 있다.

[2463] 각각의 서열에서 특정 위치 또는 잔기의 넘버링은 사용되는 특정 단백질 및 넘버링 기획에 의존하는 것으로 이해될 것이다. 넘버링은 예를 들어, 성숙한 단백질의 전구체 및 성숙한 단백질 자체에서 상이할 수 있고, 종에 따른 서열에서의 차이는 넘버링에 영향을 미칠 수 있다. 당업자는 당업계에 널리 공지된 방법에 의해, 예를 들어, 서열 정렬 및 상동성 잔기의 결정에 의해 임의의 상동성 단백질 및 각각의 암호화 핵산에서 각각의 잔기를 동정할 수 있을 것이다.

[2464] 본원에 기재된 바와 같이 임의의 융합 단백질을 표적 부위, 예를 들어, 편집될 돌연변이를 포함하는 부위에 표적화하기 위해, 가이드 RNA와 함께 융합 단백질을 동시 발현시킬 필요가 있다는 것은 당업자에게 자명할 것이다. 본원의 다른 곳에서 보다 상세하게 설명된 바와 같이, 가이드 RNA는 전형적으로 Cas12 결합을 가능하게 하는 tracrRNA 프레임워크, 및 Cas12: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

구현예에서, 가이드 RNA는 상기 가이드 서열이 표적 서열에 상보적인 서열을 포함하는 구조를 포함한다. 가이드 서열은 전형적으로 20개 뉴클레오타이드 길이이다. Cas12:핵산 편집 효소/도메인 융합 단백질을 특이적 게놈 표적 부위에 표적화하기 위해 적합한 가이드 RNA의 서열은 본원의 개시내용을 토대로 당업자에게 자명할 것이다. 상기 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오타이드의 업스트림 또는 다운스트림 50개 뉴클레오타이드 내의 핵산 서열에 상보적인 가이드 서열을 포함한다. 임의의 제공된 융합 단백질을 특이적 표적 서열에 표적화하기에 적합한 일부 예시적인 가이드 RNA 서열은 본원에 제공된다.

[2465] 본원에 개시된 염기 편집기의 도메인은 데아미나제 도메인이 Cas12 단백질에 내재화된 이상 임의의 순서로 정렬될 수 있다. 예를 들어, Cas12 도메인 및 데아미나제 도메인을 포함하는 융합 단백질을 포함하는 염기 편집기의 비제한적인 예는 다음과 같이 정렬될 수 있다:

[2466] NH2-[Cas12 도메인]-링커1-[ABE8]-링커2-[Cas12 도메인]-COOH;

[2467] NH2-[Cas12 도메인]-링커1-[ABE8]-[Cas12 도메인]-COOH;

[2468] NH2-[Cas12 도메인]-[ABE8]-링커2-[Cas12 도메인]-COOH;

[2469] NH2-[Cas12 도메인]-[ABE8]-[Cas12 도메인]-COOH;

[2470] NH2-[Cas12 도메인]-링커1-[ABE8]-링커2-[Cas12 도메인]-[이노신 BER 저해제]-COOH;

[2471] NH2-[Cas12 도메인]-링커1-[ABE8]-[Cas12 도메인]-[이노신 BER 저해제]-COOH;

[2472] NH2-[Cas12 도메인]-[ABE8]-링커2-[Cas12 도메인]-[이노신 BER 저해제]-COOH;

[2473] NH2-[Cas12 도메인]-[ABE8]-[Cas12 도메인]-[이노신 BER 저해제]-COOH;

[2474] NH2-[이노신 BER 저해제]-[Cas12 도메인]-링커1-[ABE8]-링커2-[Cas12 도메인]-COOH;

[2475] NH2-[이노신 BER 저해제]-[Cas12 도메인]-링커1-[ABE8]-[Cas12 도메인]-COOH;

[2476] NH2-[이노신 BER 저해제]-[Cas12 도메인]-[ABE8]-링커2-[Cas12 도메인]-COOH;

[2477] NH2-[이노신 BER 저해제]NH2-[Cas12 도메인]-[ABE8]-[Cas12 도메인]-COOH;

[2478] 추가로, 일부 경우에, Gam 단백질은 염기 편집기의 N 말단에 융합될 수 있다. 일부 구현예에서, Gam 단백질은 염기 편집기의 C 말단에 융합될 수 있다. 박테리오파아지 Mu의 Gam 단백질은 이중 가닥 절단 (DSB)의 말단에 결합하고 이들을 분해로부터 보호할 수 있다. 일부 구현예에서, DSB의 유리된 말단에 결합하는 Gam을 사용하여 염기 편집 과정 동안에 삽입-결실 형성을 감소시킬 수 있다. 일부 구현예에서, 174개 잔기 Gam 단백질은 염기 편집기의 N 말단에 융합된다. 문헌 (참조: 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))을 참조한다. 일부 경우에, 돌연변이 또는 돌연변이들은 야생형 도메인에 비해 염기 편집기 도메인의 길이를 변화시킬 수 있다. 예를 들어, 적어도 하나의 도메인에서 적어도 하나의 아미노산의 결실은 염기 편집기의 길이를 감소시킬 수 있다. 또 다른 경우에, 돌연변이 또는 돌연변이들은 야생형 도메인에 비해 도메인의 길이를 변화시키지 않는다. 예를 들어, 임의의 도메인에서 치환(들)은 염기 편집기의 길이를 변화시키지 않는다.

[2479] 일부 구현예에서, 본원에 제공된 염기 편집 융합 단백질은 정확한 위치, 예를 들어, 표적 염기가 한정된 영역 (예를 들어, "탈아민화 윈도우") 내에 위치할 필요가 있다. 일부 경우에, 표적은 4-염기 영역 내에 있을 수 있다. 일부 경우에, 상기 한정된 표적 영역은 PAM의 대략적으로 15개 염기 업스트림에 있을 수 있다. 문헌 (참조: 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); and 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))을 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[2480] 한정된 표적 영역은 탈아민화 윈도우일 수 있다. 탈아민화 윈도우는 염기 편집기가 표적 뉴클레오타이드에 작용하여 탈아민화하는 한정된 영역일 수 있다. 일부 구현예에서, 탈아민화 윈도우는 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 염기 영역 내에 있다. 일부 구현예에서, 탈아민화 윈도우는 PAM의 5, 6, 7, 8, 9, 10, 11, 12, 13,

14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개 염기 업스트림에 있다.

[2481] 본원 개시내용의 염기 편집기는 표적 폴리뉴클레오타이드 서열의 편집을 촉진시키는, 임의의 도메인, 특성 또는 아미노산 서열을 포함할 수 있다. 예를 들어, 일부 구현예에서, 염기 편집기는 핵 국소화 서열 (NLS)을 포함한다. 일부 구현예에서, 염기 편집기의 NLS는 데아미나제 도메인과 napDNAbp 도메인 사이에 위치한다. 일부 구현예에서, 염기 편집기의 NLS는 napDNAbp 도메인의 C-말단에 위치한다.

[2482] 융합 단백질에 포함된 단백질 도메인은 이종 기능성 도메인일 수 있다. 융합 단백질에 포함될 수 있는 단백질 도메인의 비제한적인 예는 데아미나제 도메인 (예를 들어, 아데노신 데아미나제), 우라실 글리코실라제 저해제 (UGI) 도메인, 에피토프 태그 및 수용체 유전자 서열을 포함한다. 단백질 도메인은 예를 들어, 하기의 활성 중 하나 이상을 갖는 이종 기능성 도메인일 수 있다: 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 유전자 사일런싱 활성, 염색질 변형 활성, 후성적 변형 활성, 히스톤 변형 활성, RNA 절단 활성, 및 핵산 결합 활성. 상기 이종 기능성 도메인은 표적 DNA와 연합된 표적 폴리펩타이드 (예를 들어, 히스톤, DNA 결합 단백질 등)의 변형과 같은 기능 활성을 부여하여 예를 들어, 히스톤 메틸화, 히스톤 아세틸화, 히스톤 유비퀴틴화 등을 유도할 수 있다. 부여된 다른 기능 및/또는 활성은 트랜스포사제 활성, 인테그라제 활성, 리컴비나제 활성, 리가제 활성, 유비퀴틴 리가제 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모일화 (SUMOylation) 활성, 탈수모일화 (deSUMOylation) 활성, 또는 상기 임의의 조합을 포함할 수 있다.

[2483] 도메인은 에피토프 태그, 리포터 단백질, 기타 결합 도메인을 사용하여 검출되거나 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 (Trx) 태그를 포함한다. 리포터의 예는 글루타티온-5-트랜스퍼라제 (GST), 서양고추냉이 퍼옥시다제 (HRP), 클로람페니콜 아세틸트랜스퍼라제 (CAT) 베타-갈락토시다제, 베타-글루쿠로니다제, 루시퍼라제, 녹색 형광 단백질 (GFP), HcRed, DsRed, 시안 형광 단백질 (CFP), 황색 형광 단백질 (YFP), 및 청색 형광 단백질 (BFP)을 포함하는 자가형광 단백질을 포함하지만 이에 제한되지 않는다. 추가의 단백질 서열은 DNA 분자에 결합하거나 말토스 결합 단백질 (MBP), S-태그, Lex A DNA 결합 도메인 (DBD) 융합, GAL4 DNA 결합 도메인 융합, 및 헤르페스 심플렉스 바이러스 (HSV) BP16 단백질 융합을 포함하지만 이에 제한되지 않는 다른 세포 분자에 결합하는 아미노산 서열을 포함할 수 있다.

[2484] 일부 구현예에서, BhCas12b 가이드 폴리뉴클레오타이드는 하기의 서열을 갖는다:

[2485] BhCas12b sgRNA 스캐폴드 (밑줄 친) + 20nt 내지 23nt 가이드 서열 (N_n으로 지칭됨)

[2486] 5′GUUCUGTCUUUUGGUCAGGACAACCGUCUAGCUAUAAGUGCUGCAGGGUGUGAGAAACUCCUA
UUGCUGGACGAUGUCUCUUACGAGGCAUUAGCACNNNNNNNNNNNNNNNNNNNN-3′

[2487] 일부 구현예에서, BvCas12b 및 AaCas12b 가이드 폴리뉴클레오타이드는 하기의 서열을 갖는다:

[2488] BvCas12b sgRNA 스캐폴드 (밑줄 친) + 20nt 내지 23nt 가이드 서열 (N_n으로 지칭됨)

[2489] 5′GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGGA
GCACCUGAAAACAGGUGCUUGGCACNNNNNNNNNNNNNNNNNNNN-3′

[2490] AaCas12b sgRNA 스캐폴드 (밑줄 친) + 20nt 내지 23nt 가이드 서열 (N_n으로 지칭됨)

[2491] 5′GUCUAAAGGACAGAAUUUUUCAACGGGUGUGCCAAUGGCCACUUUCCAGGUGGCAAAGCCCGU
UGAACUUCUCAAAAAGAACGAUCUGAGAAGUGGCACNNNNNNNNNNNNNNNNNNNN-3′

[2492] ***아데노신 데아미나제 변이체 및 Cas9 도메인을 포함하는 융합 단백질을 사용하는 방법***

[2493] 본원 개시내용의 일부 양상은 본원에 제공된 융합 단백질 또는 복합체를 사용하는 방법을 제공한다. 예를 들어, 본원 개시내용의 일부 양상은 돌연변이 형태의 단백질을 암호화하는 DNA 분자를 본원에 제공된 임의의 융합 단백질 및 적어도 하나의 가이드 RNA와 접촉시키는 단계를 포함하는 방법을 제공하고, 여기서, 상기 가이드 RNA는 약 15-100개 뉴클레오타이드 길이이고 표적 서열과 상보적인 적어도 10개 인접 뉴클레오타이드의 서열을 포함한다. 일부 구현예에서, 표적 서열의 3' 말단은 카노니칼 PAM 서열 (NGG)에 바로 인접해 있다. 일부 구현예에서, 표적 서열의 3' 말단은 카노니칼 PAM 서열 (NGG)에 바로 인접해 있지 않다. 일부 구현예에서, 표적 서열의 3' 말단은 AGC, GAG, TTT, GTG, 또는 CAA 서열에 바로 인접해 있다. 일부 구현예에서, 표적 서열의 3' 말

단은 NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, 또는 5' (TTTV) 서열에 바로 인접해 있다.

[2494] 일부 구현예에서, 본원 개시내용의 융합 단백질은 관심 대상의 표적을 돌연변이시키기 위해 사용된다. 특히, 본원에 기재된 아데노신 데아미나제 핵염기 편집기 (예를 들어, ABE8)는 표적 서열 내에서 다중 돌연변이를 만들 수 있다. 이러한 돌연변이는 표적의 기능에 영향을 미칠 수 있다. 예를 들어, 아데노신 데아미나제 핵염기 편집기 (예를 들어, ABE8)를 사용하여 조절 영역을 표적화하는 경우, 조절 영역의 기능이 변경되고 다운스트림 단백질의 발현이 감소된다.

[2495] 각각의 서열에서 특정 위치 또는 잔기의 넘버링은 사용되는 특정 단백질 및 넘버링 기획에 의존하는 것으로 이해될 것이다. 넘버링은 예를 들어, 성숙한 단백질의 전구체 및 성숙한 단백질 자체에서 상이할 수 있고 종에 따른 서열에서의 차이는 넘버링에 영향을 미칠 수 있다. 당업자는 당업계에 널리 공지된 방법에 의해, 예를 들어, 서열 정렬 및 상동성 잔기의 결정에 의해 임의의 상동성 단백질 및 각각의 암호화 핵산에서 각각의 잔기를 동정할 수 있을 것이다.

[2496] 본원에 개시된 바와 같이 Cas9 도메인 및 아데노신 데아미나제 변이체 (예를 들어, ABE8)를 포함하는 임의의 융합 단백질을 표적 부위, 예를 들어, 편집될 돌연변이를 포함하는 부위에 표적화하기 위해, 가이드 RNA, 예를 들어, sgRNA와 함께 융합 단백질을 전형적으로 동시 발현시킬 필요가 있다는 것은 당업자에게 자명할 것이다. 본원의 다른 곳에서 보다 상세하게 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 구현예에서, 가이드 RNA는 상기 가이드 서열이 표적 서열에 상보적인 서열을 포함하는 구조를 포함한다. 가이드 서열은 전형적으로 20개 뉴클레오타이드 길이다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특이적 게놈 표적 부위에 표적화하기 위해 적합한 가이드 RNA의 서열은 본원의 개시내용을 토대로 당업자에게 자명할 것이다. 상기 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오타이드의 업스트림 또는 다운스트림 50개 뉴클레오타이드 내의 핵산 서열에 상보적인 가이드 서열을 포함한다. 임의의 제공된 융합 단백질을 특이적 표적 서열에 표적화하기에 적합한 일부 예시적인 가이드 RNA 서열은 본원에 제공된다.

[2497] ***염기 편집기 효율***

[2498] CRISPR-Cas9 뉴클레아제는 표적화된 게놈 편집을 매개하기 위해 광범위하게 사용되었다. 대부분의 게놈 편집 적용에서, Cas9는 가이드 폴리뉴클레오타이드 (예를 들어, 단일 가이드 RNA (sgRNA))와의 복합체를 형성하고, sgRNA 서열에 의해 특정된 표적 부위에서 이중 가닥 DNA 절단 (DSB)을 유도한다. 세포는 주로 비-상동성 말단-연결 (NHEJ) 복구 경로를 통해 상기 DSB에 반응하여 유전자를 붕괴시키는 프레임쉬프트 돌연변이를 유발할 수 있는 확률적 삽입 또는 결실 (삽입-결실)을 초래한다. DSB를 플랭킹하는 서열과 고도의 상동성을 갖는 공여자 DNA 주형의 존재하에, 유전자 교정은 상동성 지시된 복구 (HDR)로서 공지된 대안 경로를 통해 성취될 수 있다. 불행하게도, 대부분의 비-섭동 (non-perturbative) 조건하에서, HDR은 세포 상태 및 세포 유형에 의존하여 비효율적이고, 보다 큰 삽입-결실의 빈도에 의해 지배된다. 인간 질환과 연관된 공지된 유전학적 변화의 대부분이 점 돌연변이므로, 보다 효율적이고 명확하게 정확한 점 돌연변이를 제조할 수 있는 방법이 필요하다. 본원에 제공된 바와 같은 염기 편집 시스템은 이중 가닥 DNA 절단을 생성하지 않고 공여자 DNA 주형을 요구하지 않고 확률적 삽입 및 결실을 과도하게 유도하지 않으면서 게놈 편집을 위한 새로운 방법을 제공한다.

[2499] 본원에 제공된 염기 편집기는 상당한 비율의 삽입-결실을 생성하지 않으면서 돌연변이를 포함하는 단백질을 암호화하는 특정 뉴클레오타이드 염기를 변경할 수 있다. 본원에 사용된 바와 같은 "삽입-결실"은 핵산 내 뉴클레오타이드 염기의 삽입 또는 결실을 언급한다. 상기 삽입 또는 결실은 유전자의 암호화 영역 내 프레임 전환 돌연변이를 유도할 수 있다. 일부 구현예에서, 표적 뉴클레오타이드 서열 내 대다수의 삽입 또는 결실 (예를 들어, 삽입-결실)을 생성시키지 않으면서 핵산 내 특정 뉴클레오타이드를 효율적으로 변형시키는 (예를 들어, 돌연변이시키는) 염기 편집기를 생성하는 것이 바람직할 수 있다. 특정 구현예에서, 본원에 제공된 임의의 염기 편집기는 삽입-결실에 비해 보다 큰 비율의 의도된 변형 (예를 들어, 점 돌연변이 또는 탈아민화)을 생성할 수 있다.

[2500] 일부 구현예에서, 본원에 제공된 임의의 염기 편집기 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 삽입-결실 형성을 초래한다.

[2501] 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 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 삽입-결실 형성을 초래한다.

[2502] 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 표적 폴리뉴클레오타이드 서열에서 0.8% 미만의 삽입-결실 형성을 초래한다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 표적 폴리뉴클레오타이드 서열에서 최대 0.8%의 삽입-결실 형성을 초래한다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 표적 폴리뉴클레오타이드 서열에서 0.3% 미만의 삽입-결실 형성을 초래한다. 일부 구현예에서, 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 ABE7 염기 편집기 중 하나를 포함하는 염기 편집기 시스템과 비교하여 표적 폴리뉴클레오타이드 서열에서 보다 낮은 삽입-결실 형성을 초래한다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 ABE7.10을 포함하는 염기 편집기 시스템과 비교하여 표적 폴리뉴클레오타이드 서열에서 보다 낮은 삽입-결실 형성을 초래한다.

[2503] 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 ABE7 염기 편집기 중 하나를 포함하는 염기 편집기 시스템과 비교하여 삽입-결실 빈도를 감소시켰다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 ABE7 염기 편집기 중 하나를 포함하는 염기 편집기 시스템과 비교하여 삽입-결실 빈도를 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% 감소시킨다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 염기 편집기 시스템은 ABE7.10을 포함하는 염기 편집기 시스템과 비교하여 삽입-결실 빈도를 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% 감소시킨다.

[2504] 본원 개시내용은 증가된 효율 및 특이성을 갖는 아데노신 데아미나제 변이체 (예를 들어, ABE8 변이체)를 제공한다. 특히, 본원에 기재된 아데노신 데아미나제 변이체는 폴리뉴클레오타이드 내 목적하는 염기를 편집할 가능성이 높고 변경시키고자 하지 않은 염기 (예를 들어, "바이스탠더")를 편집할 가능성이 적다.

[2505] 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 바이스탠더 편집 또는 돌연변이를 감소시켰다. 일부 구현예에서, 의도되지 않은 편집 또는 돌연변이는 표적 뉴클레오타이드 서열의 표적 윈도우에서 의도되지 않거나 비-표적 위치에서 바이스탠더 돌연변이 또는 바이스탠더 편집, 예를 들어, 표적 염기의 염기 편집 (예를 들어, A 또는 C)이다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집기 시스템과 비교하여 바이스탠더 편집 또는 돌연변이를 감소시켰다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집기 시스템과 비교하여 바이스탠더 편집 또는 돌연변이를 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% 또는 적어도 99% 감소시켰다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집기 시스템과 비교하여 바이스탠더 편집 또는 돌연변이를 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.

[2506] 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 스퓨리어스 편집을 감소시켰다. 일부 구현예에서, 의도되지 않은 편집 또는 돌연변이는 게놈의 의도되지 않거나 비-표적 영역에서 스퓨리어스 돌연변이 또는 스퓨리어스 편집, 예를 들어, 표적 염기 (예를 들어, A 또는 C)의 비-특이적 편집 또는 가이드 독립적 편집이다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집기 시스템과 비교하여 스퓨리어스 편집을 감소시켰다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 시스템은 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집기 시스템과 비교하여 스퓨리어스 편집을 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% 또는 적어도 99% 감소시켰다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집 시스템은 ABE7 염기 편집기, 예를 들어, ABE7.10을 포함하는 염기 편집기 시스템과 비교하여 스퓨리어스 편집을 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.

[2507] 본원 개시내용의 일부 양상은 본원에 제공된 임의의 염기 편집기가 상당한 수의 비의도된 돌연변이, 예를 들어, 비의도된 점 돌연변이 (즉, 바이스탠더의 돌연변이)를 생성하지 않고 핵산 (예를 들어, 대상체의 게놈 내 핵산)에서 의도된 돌연변이, 예를 들어, 점 돌연변이를 효율적으로 생성할 수 있다. 일부 구현예에서, 본원에 제공된 임의의 염기 편집기는 적어도 0.01%의 의도된 돌연변이 (즉, 적어도 0.01% 염기 편집 효율)를 생성할 수 있다. 일부 구현예에서, 본원에 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.

[2508] 일부 구현예에서, 본원에 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 구현예에서, 염기 편집 효율은 세포 집단에서 편집된 핵염기의 백분율을 계산하여 측정될 수 있다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 세포 집단에서 편집된 핵염기에 의한 측정시 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.

[2509] 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 ABE7 염기 편집기와 비교하여 보다 높은 염기 편집 효율을 갖는다. 일부 구현예에서, 본원에 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.

[2510] 일부 구현예에서, 본원에 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.

[2511] 일부 구현예에서, 본원에 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 표적-적중 염기 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 세포 집단에서 편집된 표적 핵염기에 의한 측정시 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 표적-적중 염기 편집 효율을 갖는다.

[2512] 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 ABE7 염기 편집기와 비교하여 보다 높은 표적-적중 염기 편집 효율을 갖는다. 일부 구현예에서, 본원에 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 표적-적중 염기 편집 효율을 갖는다.

[2513] 일부 구현예에서, 본원에 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 표적-적중 염기 편집 효율을 갖는다.

[2514] 본원에 기재된 ABE8 염기 편집기 변이체는 플라스미드, 벡터, LNP 복합체 또는 mRNA를 통해 숙주 세포에 전달될 수 있다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 mRNA로서 숙주 세포에 전달된다. 일부 구현예에서, 핵산 기반 전달 시스템, 예를 들어, mRNA를 통해 전달된 ABE8 염기 편집기는 편집된 핵염기에 의한 측정시 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 표적-적중 염기 편집 효율을 갖는다. 일부 구현예에서, mRNA 시스템에 의해 전달되는 ABE8 염기 편집기는 플라스미드 또는 벡터 시스템에 의해 전달되는 ABE8 염기 편집기와 비교하여 보다 높은 염기 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우 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%의 표적-적중 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우 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 표적-적중 염기 편집 효율을 갖는다.

[2515] 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 임의의 염기 편집기 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 표적-이탈 편집을 초래한다.

[2516] 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우, 보다 낮은 가이드된 표적-이탈 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우, 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 표적-이탈 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우 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 표적-이탈 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우, 가이드 표적-이탈 편집 효율을 적어도 약 2.2배 감소시킨다.

[2517] 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우 보다 낮은 가이드-독립적 표적-이탈 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우, 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 표적-이탈 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우 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 표적-이탈 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 임의의 ABE8 염기 편집기 변이체는 플라스미드 또는 벡터 시스템에 의해 전달되는 경우와 비교하여 mRNA 시스템에 의해 전달되는 경우 가이드-독립적 표적-이탈 편집 효율 (예를 들어, 스퓨리어스 RNA 탈아민화)을 134.0배 감소시킨다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체는 게놈 전반에 걸쳐 가이드-독립적 돌연변이율을 증가시키지 않는다.

[2518] 본원 개시내용의 일부 양상은 본원에 제공된 임의의 염기 편집기가 상당한 수의 비의도된 돌연변이 (예를 들어, 스퓨리어스 표적-이탈 편집 또는 바이스탠더 편집)를 생성하지 않고 핵산 (예를 들어, 대상체의 게놈 내 핵산)에서 의도된 돌연변이, 예를 들어, 점 돌연변이를 효율적으로 생성할 수 있다는 인지를 기반으로 한다. 일부 구현예에서, 의도된 돌연변이는 표적 유전자 내 돌연변이를 변경하거나 교정하도록 특이적으로 디자인된, gRNA에 결합하는 특이적 염기 편집기에 의해 생성되는 돌연변이이다. 본원 개시내용의 일부 양상은 본원에 제공된 임의의 염기 편집기가 상당한 수의 비의도된 돌연변이를 생성하지 않고 핵산 (예를 들어, 대상체의 게놈 내 핵

산)에서 의도된 돌연변이를 효율적으로 생성할 수 있다는 인지를 기반으로 한다. 일부 구현예에서, 의도된 돌연변이는 의도된 돌연변이를 변경하거나 교정하도록 특이적으로 디자인된, gRNA에 결합하는 특이적 염기 편집기에 의해 생성되는 돌연변이이다. 일부 구현예에서, 의도된 돌연변이는 정지 코돈, 예를 들어, 유전자의 암호화 영역 내 미성숙한 정지 코돈을 생성하는 돌연변이이다. 일부 구현예에서, 의도된 돌연변이는 정지 코돈을 제거하는 돌연변이다. 일부 구현예에서, 의도된 돌연변이는 유전자의 스플라이싱을 변경하는 돌연변이이다. 일부 구현예에서, 의도된 돌연변이는 유전자의 조절 서열 (예를 들어, 유전자 프로모터 또는 유전자 리프레서)을 변경하는 돌연변이이다.

[2519] 일부 구현예에서, 본원에 제공된 염기 편집기는 1:1 초과인 의도된 점 돌연변이 대 삽입-결실 (즉, 비의도된 돌연변이)의 비율을 생성할 수 있다. 일부 구현예에서, 본원에 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 8.5:1, 적어도 9:1, 적어도 10:1, 적어도 11:1, 적어도 12:1, 적어도 13:1, 적어도 14:1, 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 삽입-결실의 비율을 생성시킬 수 있다. 본원에 기재된 바와 같은 염기 편집기의 특징은 본원에 제공된 임의의 융합 단백질 또는 융합 단백질을 사용하는 방법에 적용될 수 있는 것으로 인지되어야 한다.

[2520] 의도된 돌연변이 및 삽입-결실의 수는 예를 들어, 문헌 (참조: 국제 PCT 출원 번호 PCT/2017/045381 (WO2018/027078) 및 PCT/US2016/058344 (WO2017/070632); 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), 이의 전체 내용은 본원에 참조로 포함됨)에 기재된 바와 같은 임의의 적합한 방법을 사용하여 결정될 수 있다.

[2521] 일부 구현예에서, 삽입-결실 빈도를 계산하기 위해, 서열분석 판독은 삽입-결실이 발생할 수 있는 윈도우의 양 측면을 플랭킹하는 2개의 10-bp 서열과의 정확한 매칭에 대해 스캐닝한다. 정확한 매칭이 위치하지 않는 경우, 판독은 분석으로부터 배제된다. 상기 삽입-결실 윈도우의 길이가 정밀하게 참조 서열과 매칭하는 경우, 상기 판독은 삽입-결실을 함유하지 않는 것으로서 분류된다. 삽입-결실 윈도우가 참조 서열 보다 2개 이상의 염기로 길거나 짧은 경우, 서열 분석 판독은 각각 삽입 또는 결실로서 분류된다. 일부 구현예에서, 본원에 제공된 염기 편집기는 핵산 영역 내 삽입-결실의 형성을 제한할 수 있다. 일부 구현예에서, 상기 영역은 염기 편집기에 의해 표적화된 뉴클레오타이드에 있거나, 염기 편집기에 의해 표적화된 뉴클레오타이드의 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 뉴클레오타이드 내 영역에 있다.

[2522] 표적 뉴클레오타이드 영역에서 형성되는 삽입-결실의 수는 핵산 (예를 들어, 세포의 게놈 내 핵산)이 염기 편집기에 노출되는 시간의 양에 따를 수 있다. 일부 구현예에서, 삽입-결실의 수 또는 비율은 표적 뉴클레오타이드 서열 (예를 들어, 세포의 게놈 내 핵산)을 염기 편집기에 노출시키는 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 후 결정된다. 본원에 기재된 바와 같은 염기 편집기의 특징은 본원에 제공된 임의의 융합 단백질 또는 융합 단백질을 사용하는 방법에 적용될 수 있는 것으로 인지되어야 한다.

[2523] 일부 구현예에서, 본원에 제공된 염기 편집기는 핵산 영역 내 삽입-결실의 형성을 제한할 수 있다. 일부 구현예에서, 영역은 염기 편집기에 의해 표적화된 뉴클레오타이드에 있거나 염기 편집기에 의해 표적화된 뉴클레오타이드의 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 뉴클레오타이드 내 영역에 있다. 일부 구현예에서, 본원에 제공된 임의의 염기 편집기는 핵산 영역에서 삽입-결실의 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 형성되는 삽입-결실의 수는 핵산 (예를 들어, 세포의 게놈 내 핵산)이 염기 편집기에 노출되는 시간의 양에 따를 수 있다. 일부 구현예에서, 삽입-결실의 임의의 수 또는 비율은 핵산 (예를 들어, 세포의 게놈 내 핵산)을 염기 편집기에 노출시키는 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 후 결정된다.

[2524] ***멀리플렉스 편집***

[2525] 일부 구현예에서, 본원에 제공된 염기 편집기 시스템은 하나 이상의 유전자에서 다수의 핵염기 쌍의 멀티플렉스 편집을 수행할 수 있다. 일부 구현예에서, 다수의 핵염기 쌍은 동일한 유전자에 위치한다. 일부 구현예에서, 다수의 핵염기 쌍은 하나 이상의 유전자에 위치하고, 여기서, 적어도 하나의 유전자는 상이한 유전자좌에 위치한다. 일부 구현예에서, 멀티플렉스 편집은 하나 이상의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 하나 이상의 염기 편집기 시스템을 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 하나 이상의 염기 편집기 시스템을 단일 가이드 폴리뉴클레오타이드와 함께 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 하나 이상의 염기 편집기 시스템을 다수의 가이드 폴리뉴클레오타이드와 함께 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 하나 이상의 폴리뉴클레오타이드를 단일 염기 편집기 시스템과 함께 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드와 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오타이드의 혼합물을 포함할 수 있다. 본원에 기재된 바와 같은 임의의 염기 편집기를 사용한 멀티플렉스 편집의 특징은 본원에 제공된 임의의 염기 편집기를 사용한 방법의 임의의 조합에 적용될 수 있는 것으로 인지되어야 한다. 또한, 본원에 기재된 바와 같은 임의의 염기 편집기를 사용한 멀티플렉스 편집은 다수의 핵염기 쌍의 순차적 편집을 포함할 수 있는 것으로 인지되어야 한다.

[2526] 일부 구현예에서, 다수의 핵염기 쌍은 하나 이상의 유전자에 위치한다. 일부 구현예에서, 다수의 핵염기 쌍은 동일한 유전자에 위치한다. 일부 구현예에서, 하나 이상의 유전자에서 적어도 하나의 유전자는 상이한 유전자좌에서 위치한다. 일부 구현예에서, 편집은 적어도 하나의 단백질 암호화 영역에서 다수의 핵염기 쌍의 편집이다. 일부 구현예에서, 편집은 적어도 하나의 단백질 비-암호화 영역에서 다수의 핵염기 쌍의 편집이다. 일부 구현예에서, 편집은 적어도 하나의 단백질 암호화 영역 및 적어도 하나의 단백질 비-암호화 영역에서 다수의 핵염기 쌍의 편집이다.

[2527] 일부 구현예에서, 편집은 하나 이상의 가이드 폴리뉴클레오타이드와 연계한다. 일부 구현예에서, 염기 편집기 시스템은 하나 이상의 염기 편집기 시스템을 포함할 수 있다. 일부 구현예에서, 염기 편집기 시스템은 하나 이상의 염기 편집기 시스템을 단일 가이드 폴리뉴클레오타이드와 함께 포함할 수 있다. 일부 구현예에서, 염기 편집기 시스템은 하나 이상의 염기 편집기 시스템을 다수의 가이드 폴리뉴클레오타이드와 함께 포함할 수 있다. 일부 구현예에서, 편집은 단일 염기 편집기 시스템과 함께 하나 이상의 가이드 폴리뉴클레오타이드와 연계한다. 일부 구현예에서, 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드와 연계한다. 일부 구현예에서, 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오타이드와 연계한다. 일부 구현예에서, 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드와 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오타이드의 혼합물과 연계한다. 본원에 기재된 바와 같은 임의의 염기 편집기를 사용한 멀티플렉스 편집의 특징은 본원에 제공된 임의의 염기 편집기를 사용한 방법의 임의의 조합에 적용될 수 있는 것으로 인지되어야 한다. 또한 편집은 다수의 핵염기 쌍의 순차적 편집을 포함할 수 있는 것으로 인지되어야 한다.

[2528] 일부 구현예에서, 하나 이상의 유전자에서 다수의 핵염기 쌍의 멀티플렉스 편집을 수행할 수 있는 염기 편집기 시스템은 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함한다. 일부 구현예에서, 하나 이상의 유전자에서 다수의 핵염기 쌍의 멀티플렉스 편집을 수행할 수 있는 염기 편집기 시스템은 ABE7 염기 편집기 중 하나를 포함한다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 멀티플렉스 편집을 수행할 수 있는 염기 편집기 시스템은 ABE7 염기 편집기 중 하나를 포함하는 멀티플렉스 편집을 수행할 수 있는 염기 편집기 시스템과 비교하여 보다 높은 멀티플렉스 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 멀티플렉스 편집을 수행할 수 있는 염기 편집기 시스템은 ABE7 염기 편집기 중 하나를 포함하는 멀티플렉스 편집을 수행할 수 있는 염기 편집기 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%, 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 멀티플렉스 편집 효율을 갖는다. 일부 구현예에서, 본원에 기재된 ABE8 염기 편집기 변이체 중 하나를 포함하는 멀티플렉스 편집을 수행할 수 있는 염기 편집기 시스템은 ABE7 염기 편집기 중 하나를 포함하는 멀티플렉스 편집을 수행할 수 있는 염기 편집기 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 3.2, 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 멀티플렉스 편집 효율을 갖는다.

[2529] **핵산을 편집하기 위한 방법**

[2530] 본원 개시내용의 일부 양상은 핵산을 편집하기 위한 방법을 제공한다. 일부 구현예에서, 상기 방법은 단백질을 암호화하는 핵산 분자 (예를 들어, 이중-가닥 DNA 서열의 염기쌍)의 핵염기를 편집하기 위한 방법이다. 일부 구현예에서, 상기 방법은 하기의 단계를 포함한다: a) 핵산 (예를 들어, 이중-가닥 DNA 서열)의 표적 영역을 염기 편집기 및 가이드 핵산 (예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 상기 표적 영역의 단일 가닥 내 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계, 및 d) 상기 표적 영역의 하나 이하의 가닥을 절단하는 단계로서, 상기 제1 핵염기에 상보적인 제3 핵염기가 상기 제2 핵염기에 상보적인 제4 핵염기로 대체되는, 단계. 일부 구현예에서, 상기 방법은 핵산 내 20% 미만의 삽입-결실 형성을 초래한다. 일부 구현예에서, 단계 b는 생략된다는 것이 인지되어야 한다. 일부 구현예에서, 상기 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만, 또는 0.1% 미만의 삽입-결실 형성을 초래한다. 일부 구현예에서, 상기 방법은 상기 제2 핵염기를 상기 제4 핵염기에 상보적인 제5 핵염기로 대체하여 의도된 편집된 염기 쌍을 생성 (예를 들어, G·C에서 A·T로)하는 단계를 추가로 포함한다. 일부 구현예에서, 의도된 염기쌍의 적어도 5%가 편집된다. 일부 구현예에서, 의도된 염기쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 또는 50%가 편집된다.

[2531] 일부 구현예에서, 표적 뉴클레오타이드에서 의도된 생성물 대 비의도된 생성물의 비율은 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 구현예에서, 의도된 돌연변이 대 삽입-결실 형성의 비율은 1:1, 10:1, 50:1, 100:1, 500:1 초과, 또는 1000:1 이상보다 크다. 일부 구현예에서, 절단된 단일 가닥 (닉 가닥)은 가이드 핵산에 하이브리드화한다. 일부 구현예에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥에 반대편에 있다. 일부 구현예에서, 염기 편집기는 dCas9 도메인을 포함한다. 일부 구현예에서, 염기 편집기는 비-편집된 가닥을 보호하거나 이에 결합한다. 일부 구현예에서, 의도된 편집된 염기 쌍은 PAM 부위의 업스트림이다. 일부 구현예에서, 의도된 편집된 염기쌍은 PAM 부위의 업스트림의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드이다. 일부 구현예에서, 의도된 편집된 염기 쌍은 PAM 부위의 다운스트림이다. 일부 구현예에서 의도된 편집된 염기 쌍은 PAM 부위의 다운스트림의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드이다. 일부 구현예에서, 상기 방법은 카노니칼 (예를 들어, NGG) PAM 부위를 필요로하지 않는다. 일부 구현예에서, 핵염기 편집기는 링커를 포함한다. 일부 구현예에서, 링커는 1-25개 아미노산 길이이다. 일부 구현예에서, 링커는 5-20개 아미노산 길이이다. 일부 구현예에서, 링커는 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 아미노산 길이이다. 하나의 구현예에서, 링커는 32개 아미노산 길이이다. 또 다른 구현예에서, "긴 링커"는 적어도 약 60개 아미노산 길이이다. 다른 구현예에서, 링커는 약 3-100개 아미노산 길이이다. 일부 구현예에서, 상기 표적 영역은 표적 윈도우를 포함하고, 여기서, 상기 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 구현예에서, 표적 윈도우는 1-10개 뉴클레오타이드를 포함한다. 일부 구현예에서, 표적 윈도우는 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, 또는 1개 뉴클레오타이드 길이이다. 일부 구현예에서, 표적 윈도우는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드 길이이다. 일부 구현예에서, 의도된 편집된 염기 쌍은 표적 윈도우 내에 있다. 일부 구현예에서, 표적 윈도우는 의도된 편집된 염기 쌍을 포함한다. 일부 구현예에서, 상기 방법은 본원에 제공된 임의의

염기 편집기를 사용하여 수행된다.

[2532] 일부 구현예에서, 본원 개시내용은 뉴클레오타이드 (예를 들어, 단백질을 암호화하는 유전자에서 SNP)를 편집하기 위한 방법을 제공한다. 일부 구현예에서, 본원 개시내용은 이중-가닥 DNA 서열의 핵염기 쌍을 편집하기 위한 방법을 제공한다. 일부 구현예에서, 상기 방법은 a) 이중-가닥 DNA 서열의 표적 영역을 염기 편집기 및 가이드 핵산 (예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계로서, 상기 표적 영역이 표적 핵염기 쌍을 포함하는, 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 상기 표적 영역의 단일 가닥 내 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계, d) 상기 표적 영역의 하나 이하의 가닥을 절단하는 단계로서 상기 제1 핵염기에 상보적인 제3 핵염기가 상기 제2 핵염기에 상보적인 제4 핵염기로 대체되고, 상기 핵염기가 상기 제4 핵염기에 상보적인 제5 핵염기로 대체되어 의도된 편집된 염기쌍을 생성하고, 상기 의도된 편집된 염기쌍의 효율이 적어도 5%인, 단계를 포함한다. 일부 구현예에서, 단계 b는 생략되는 것으로 인지되어야 한다. 일부 구현예에서, 의도된 염기쌍의 적어도 5%가 편집된다. 일부 구현예에서, 의도된 염기쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 또는 50%가 편집된다. 일부 구현예에서, 상기 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만, 또는 0.1% 미만의 삽입-결실 형성을 유발한다. 일부 구현예에서, 표적 뉴클레오타이드에서 의도된 생성물 대 비의도된 생성물의 비율은 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 구현예에서, 의도된 돌연변이 대 삽입-결실 형성의 비율은 1:1, 10:1, 50:1, 100:1, 500:1 초과, 또는 1000:1 이상이다. 일부 구현예에서, 절단된 단일 가닥은 가이드 핵산에 하이브리드화한다. 일부 구현예에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥에 반대편에 있다. 일부 구현예에서, 의도된 편집된 염기 쌍은 PAM 부위의 업스트림이다. 일부 구현예에서, 의도된 편집된 염기쌍은 PAM 부위의 업스트림의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드이다. 일부 구현예에서, 의도된 편집된 염기쌍은 PAM 부위의 다운스트림이다. 일부 구현예에서 의도된 편집된 염기 쌍은 PAM 부위의 다운스트림의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드이다. 일부 구현예에서, 상기 방법은 카노니칼 (예를 들어, NGG) PAM 부위를 필요로하지 않는다. 일부 구현예에서, 링커는 1-25개 아미노산 길이이다. 일부 구현예에서, 링커는 5-20개 아미노산 길이이다. 일부 구현예에서, 링커는 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 아미노산 길이이다. 일부 구현예에서, 상기 표적 영역은 표적 윈도우를 포함하고, 여기서, 상기 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 구현예에서, 표적 윈도우는 1-10개 뉴클레오타이드를 포함한다. 일부 구현예에서, 표적 윈도우는 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, 또는 1개 뉴클레오타이드 길이이다. 일부 구현예에서, 표적 윈도우는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드 길이이다. 일부 구현예에서, 의도된 편집된 염기 쌍은 표적 윈도우 내에 존재한다. 일부 구현예에서, 표적 윈도우는 의도된 편집된 염기 쌍을 포함한다. 일부 구현예에서, 핵염기 편집기는 본원에 제공된 염기 편집기의 임의의 하나이다.

[2533] ***숙주 세포에서 융합 단백질의 발현***

[2534] 아데노신 데아미나제 변이체를 포함하는 본원 개시내용의 융합 단백질은 실제로 세균, 효모, 진균류, 곤충, 식물 및 동물 세포를 포함하지만 이에 제한되지 않는 관심 대상의 임의의 숙주 세포에서 당업자에게 공지된 통상적인 방법을 사용하여 발현될 수 있다. 예를 들어, 본원 개시내용의 아데노신 데아미나제를 암호화하는 DNA는 cDNA 서열을 기준으로 CDS의 업스트림 및 다운스트림에 대해 적합한 프라이머를 디자인함에 의해 클로닝될 수 있다. 클로닝된 DNA는 직접적으로 또는 경우에 따라 제한 효소를 사용한 분해 후, 또는 염기 편집 시스템의 하나 이상의 추가의 성분을 암호화하는 DNA로 연결된 적합한 링커 및/또는 핵 국소화 신호의 첨가 후 일 수 있다. 염기 편집 시스템은 숙주 세포에서 해독되어 복합체를 형성한다.

[2535] 본원에 기재된 단백질 도메인을 암호화하는 DNA는 DNA를 화학적으로 합성하여, 또는 PCR 방법 및 깁슨 어셈블리 방법을 사용하여 전장을 암호화하는 DNA를 작제함으로써 합성된 부분적 중첩 올리고DNA 단쇄를 연결하여 수득될 수 있다. 화학적 합성 또는 PCR 방법 또는 깁슨 어셈블리 방법의 조합에 의해 전장 DNA를 작제하는 이점은 사용될 코돈이 DNA가 도입되는 숙주에 따라 CDS 전장에서 디자인될 수 있다는 것이다. 이종성 DNA의 발현에서, 단백질 발현 수준은 이의 DNA 서열을 숙주 유기체에서 고도로 흔하게 사용되는 코돈으로 전환시킴에 의해 증가하는 것으로 예상된다. 사용될 숙주에서의 코돈 사용 빈도의 데이터로서, 예를 들어, Kazusa DNA 연구소의 홈페이지에 기재된 유전자 코드 사용 빈도 데이터베이스 (http://www.kazusa.or.jp/codon/index.html)가 사용될 수 있거나, 각각의 숙주에서 코돈 사용 빈도를 보여주는 문헌이 참조될 수 있다. 수득된 데이터 및 도입될 DNA 서열을 참조하여, DNA 서열을 위해 사용되는 것들 중에서 숙주에서 낮은 사용 빈도를 보여주는 코돈은 동일한 아미노산을 암호화하고 높은 사용 빈도를 보여주는 코돈으로 전환될 수 있다.

[2536] 핵산 서열-인지 모듈 및/또는 핵산 염기 전환 효소를 암호화하는 DNA를 함유하는 발현 벡터는 예를 들어, DNA를 적합한 발현 벡터 내 프로모터의 다운스트림에 연결함으로써 생성될 수 있다.

[2537] 발현 벡터가 에스케리치아 콜리-유래된 플라스미드 (예를 들어, pBR322, pBR325, pUC12, pUC13); 바실러스 서브틸리스-유래된 플라스미드 (예를 들어, pUB110, pTP5, pC194); 효모-유래된 플라스미드 (예를 들어, pSH19, pSH15); 곤충 세포 발현 플라스미드 (예를 들어, pFast-Bac); 동물 세포 발현 플라스미드 (예를 들어, pA1-11, pXT1, pRc/CMV, pRc/RSV, pcDNAI/Neo)인 경우, 박테리오파아지, 예를 들어, 람다.파아지 등; 곤충 바이러스 벡터, 예를 들어, 바큘로바이러스 등 (예를 들어, BmNPV, AcNPV); 동물 바이러스 벡터, 예를 들어, 레트로바이러스, 백시니아 바이러스, 아데노바이러스 등이 사용된다.

[2538] 일부 구현예에서, 유전자 발현을 위해 사용될 숙주에 적당한 임의의 프로모터가 사용될 수 있다. DSB를 사용하는 통상적인 방법에서, 숙주 세포의 생존율은 때로는 독성으로 인해 현저하게 감소하기 때문에, 유도성 프로모터를 사용함에 의한 유도의 개시에 의해 세포 수를 증가시키는 것이 요구될 수 있다. 그러나, 충분한 세포 증식은 또한 본원 개시내용의 핵산 변형 효소 복합체를 발현시킴에 의해 부여될 수 있기 때문에, 항시성 프로모터는 또한 제한 없이 사용될 수 있다.

[2539] 예를 들어, 숙주가 동물 세포인 경우, SRα 프로모터, SV40 프로모터, LTR 프로모터, CMV (시토메갈로바이러스) 프로모터, RSV (라우스 사코마 바이러스) 프로모터, MoMuLV (몰로니 마우스 백혈병 바이러스) LTR, HSV-TK (심플 헤르페스 바이러스 티미딘 키나제) 프로모터 등이 사용된다. 이들 중에서, CMV 프로모터, SRα 프로모터 등이 바람직하다.

[2540] 숙주가 에스케리치아 콜리인 경우, trp 프로모터, lac 프로모터, recA 프로모터, λP_L 프로모터, lpp 프로모터, T7 프로모터 등이 바람직하다.

[2541] 숙주가 바실러스 속인 경우, SPO1 프로모터, SPO2 프로모터, penP 프로모터 등이 바람직하다.

[2542] 숙주가 효모인 경우, Gal1/10 프로모터, PHO5 프로모터, PGK 프로모터, GAP 프로모터, ADH 프로모터 등이 바람직하다.

[2543] 숙주가 곤충 세포인 경우, 폴리헤드린 프로모터, P10 프로모터 등이 바람직하다.

[2544] 숙주가 식물 세포인 경우, CaMV35S 프로모터, CaMV19S 프로모터, NOS 프로모터 등이 바람직하다.

[2545] 상기 언급된 것들 외에 발현 벡터로서, 인핸서, 스플라이싱 신호, 종결인자, 폴리A 부가 신호, 선택 마커, 예를 들어, 약물 내성 유전자, 영양요구성 상보성 유전자 등, 복제 오리진 등을 함유하는 것이 사용될 수 있다.

[2546] 본원에 기재된 단백질 도메인을 암호화하는 RNA는 예를 들어, 주형으로서 상기 언급된 핵산 서열-인지 모듈 및/또는 핵산 염기 전환 효소를 암호화하는 DNA를 암호화하는 벡터를 사용함에 의해 공지된 시험관내 전사 시스템 자체에서 mRNA로의 전사에 의해 제조될 수 있다.

[2547] 본원 개시내용의 융합 단백질은 핵산 서열-인지 모듈 및/또는 핵산 염기 전환 효소를 암호화하는 DNA를 포함하는 발현 벡터를 숙주 세포에 도입하고 숙주 세포를 배양함으로써 세포내 발현될 수 있다.

[2548] 숙주로서, 에스케리치아 속, 바실러스 속, 효모, 곤충 세포, 곤충, 동물 세포 등이 사용된다.

[2549] 에스케리치아 속으로서, 에스케리치아 콜리 K12.cndot.DH1 (Proc. Natl. Acad. sci. USA, 60, 160 (1968)), 에스케리치아 콜리 JM103 (Nucleic Acids Research, 9, 309 (1981)), 에스케리치아 콜리 JA221 (Journal of Molecular Biology, 120, 517 (1978)), 에스케리치아 콜리 HB101 (Journal of Molecular Biology, 41, 459 (1969)), 에스케리치아 콜리 C600 (Genetics, 39, 440 (1954)) 등이 사용된다.

[2550] 바실러스 속으로서, 바실러스 서브틸리스 M1114 (Gene, 24, 255 (1983)), 바실러스 서브틸리스 207-21 (Journal of Biochemistry, 95, 87 (1984)) 등이 사용된다.

[2551] 효모로서, 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*) AH22, AH22R, NA87-11A, DKD-5D, 20B-12, 쉬조사카로마이세스 폼베 (*Schizosaccharomyces pombe*) NCYC1913, NCYC2036, 피키아 파스토리스 (*Pichia pastoris*) KM71 등이 사용된다.

[2552] 바이러스가 AcNPV인 경우 곤충 세포로서, 양배추 군벌레 유충-유래된 확립된 계통의 세포 (스포도프테라 프루기페르다 (*Spodoptera frugiperda*) 세포; Sf 세포), 트리코플루시아 니 (*Trichoplusia ni*)의 중간 장으로부터 유

래된 MG1 세포, 트리코플루시아 니의 란으로부터 유래된 하이 파이브 (High Five)™ 세포, 마메스트라 브라시카 (*Mamestra brassicae*)-유래된 세포, 에스티그메나 아크레아 (*Estigmena acrea*)-유래된 세포 등이 사용된다. 바이러스가 BmNPV인 경우, 봄빅스 모리 (*Bombyx mori*)-유래된 확립된 주의 세포 (봄빅스 모리 N 세포; BmN 세포) 등은 곤충 세포로서 사용된다. sf 세포로서, 예를 들어, Sf9 세포 (ATCC CRL1711), Sf21 세포 (상기 모두, 생체내, 13, 213-217 (1977)) 등이 사용된다.

[2553] 곤충으로서, 예를 들어, 봄믹스 모리, 드로소필라, 크리켓 등의 유충이 사용된다 (참조: Nature, 315, 592 (1985)).

[2554] 동물 세포로서, 세포주 예를 들어, 몽키 COS-7 세포, 몽키 베로 세포, 차이니즈 햄스터 난소 (CHO) 세포, dhfr 유전자-결핍 CHO 세포, 마우스 L 세포, 마우스 AtT-20 세포, 마우스 골수종 세포, 래트 GH3 세포, 인간 FL 세포 등, 만능 줄기 세포 예를 들어, iPS 세포, 인간 및 다른 포유류의 ES 세포 등, 및 다양한 조직으로부터 제조된 1차 배양된 세포가 사용된다. 추가로, 제브라피시 배아, 제노푸스 난모세포 등이 또한 사용될 수 있다.

[2555] 식물 세포로서, 현탁 배양된 세포, 캘러스, 원형질체, 다양한 식물로부터 제조된 잎 분절, 뿌리 분절 등 (예를 들어, 낟알, 예를 들어, 쌀, 밀, 옥수수 등, 제품 작물, 예를 들어, 토마토, 오이, 가지 등, 정원 식물, 예를 들어, 카네이션, 유스토마 루셀리아눔 (*Eustoma russellianum*) 등, 실험 식물, 예를 들어, 토바코, 아라비도프시스 탈리아나 (*arabidopsis thaliana*) 등)이 사용된다.

[2556] 상기 언급된 모든 숙주 세포는 반수체 (단배체) 또는 다배체 (예를 들어, 이배체, 삼배체, 사배체 등)일 수 있다. 통상적인 돌연변이 도입 방법에서, 돌연변이는 원칙적으로 헤테로 유전자 유형을 생성하기 위해 단지 하나의 상동성 염색체에 도입된다. 따라서, 목적하는 표현형은 우성 돌연변이가 존재하지 않는 경우 발현되지 않고 동형접합은 불편하게 노동 및 시간을 필요로 한다. 대조적으로, 본원 개시내용에 따라, 돌연변이는 게놈 내 상동성 염색체 상에 임의의 대립유전자에 도입될 수 있기 때문에, 목적하는 표현형은 열성 돌연변이의 경우에도 단일 생성으로 발현될 수 있고, 이는 통상적인 방법의 문제가 해결될 수 있으므로 극히 유용하다.

[2557] 발현 벡터는 숙주의 종류에 따라 공지된 방법 (예를 들어, 리소자임 방법, 컴피턴트 방법, PEG 방법, $CaCl_2$ 공침전 방법, 전기천공 방법, 미세주사 방법, 입자 총 방법, 지질감염 방법, 아그로박테리움 방법 등)에 의해 도입될 수 있다.

[2558] 에스케리치아 콜리는 예를 들어, 문헌 (참조: Proc. Natl. Acad. sci. USA, 69, 2110 (1972), Gene, 17, 107 (1982)) 등에 기재된 방법에 따라 형질전환될 수 있다.

[2559] 바실러스 속은 예를 들어, 문헌 (참조: Molecular & General Genetics, 168, 111 (1979)) 등에 기재된 방법에 따라 벡터에 도입될 수 있다.

[2560] 효모는 예를 들어, 문헌 (참조: Methods in Enzymology, 194, 182-187 (1991), Proc. Natl. Acad. sci. USA, 75, 1929 (1978)) 등에 기재된 방법에 따라 벡터에 도입될 수 있다.

[2561] 곤충 세포 및 곤충은 예를 들어, 문헌 (참조: Bio/Technology, 6, 47-55 (1988)) 등에 기재된 방법에 따라 벡터에 도입될 수 있다.

[2562] 동물 세포는 예를 들어, 문헌 (참조: Cell Engineering additional volume 8, New Cell Engineering Experiment Protocol, 263-267 (1995) (published by Shujunsha), and Virology, 52, 456 (1973))에 기재된 방법에 따라 벡터에 도입될 수 있다.

[2563] 벡터가 도입된 세포는 숙주의 종류에 따라 공지된 방법에 따라 배양될 수 있다.

[2564] 예를 들어, 에스케리치아 콜리 또는 바실러스 속이 배양된 경우, 액체 배지는 배양을 위해 사용될 배지로서 바람직할 수 있다. 배지는 바람직하게 형질전환체의 성장을 위해 필요한 탄소원, 질소원, 무기 물질 등을 함유한다. 탄소원의 예는 글루코스, 덱스트린, 가용성 전분, 슈크로스 등을 포함하고; 질소원의 예는 무기 또는 유기 물질, 예를 들어, 암모늄 염, 니트레이트 염, 옥수수 침지액, 펩톤, 카세인, 육류 추출물, 대두 케이크, 감자 추출물 등을 포함하고; 무기 물질의 예는 염화칼슘, 인산이수소나트륨, 염화마그네슘 등을 포함한다. 배지는 효모 추출물, 비타민, 성장 촉진 인자 등을 함유할 수 있다. 배지의 pH는 바람직하게 약 5 내지 약 8이다.

[2565] 에스케리치아 콜리를 배양하기 위한 배지로서, 예를 들어, 글루코스, 카사미노산 (Journal of Experiments in Molecular Genetics, 431-433, Cold Spring Harbor Laboratory, New York 1972)을 함유하는 M9 배지가 바람직할 수 있다. 필요한 경우, 예를 들어, 3β-인돌릴아크릴산과 같은 제제는 배지에 첨가하여 프로모터의 효율적

인 기능을 보장할 수 있다. 에스케리치아 콜리는 일반적으로 약 15 내지 약 43℃에서 배양된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[2566] 바실러스 속은 일반적으로 약 30℃ 내지 약 40℃에서 배양된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[2567] 효모를 배양하기 위한 배지의 예는 버크홀더 (Burkholder) 최소 배지 (Proc. Natl. Acad. sci. USA, 77, 4505 (1980)), 0.5% 카사미노산 (Proc. Natl. Acad. sci. USA, 81, 5330 (1984)) 등을 함유하는 SD 배지를 포함한다. 배지의 pH는 바람직하게 약 5 내지 약 8이다. 배양은 일반적으로 약 20℃ 내지 약 35℃에서 수행된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[2568] 곤충 세포 또는 곤충을 배양하기 위한 배지로서, 예를 들어, 적절하게 불활성화된 10% 소 혈청 등과 같은 첨가제를 함유하는 그레이스 (Grace) 곤충 배지 (Nature, 195, 788 (1962))가 사용된다. 배지의 pH는 바람직하게 약 6.2 내지 약 6.4이다. 배양은 일반적으로 약 27℃에서 수행된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[2569] 동물 세포를 배양하기 위한 배지로서, 예를 들어, 약 5 내지 약 20%의 태아 소 혈청을 함유하는 최소 필수 배지 (MEM) (Science, 122, 501 (1952)), 둘베코 변형 이글 배지 (DMEM) (Virology, 8, 396 (1959)), RPMI 1640 배지 (The Journal of the American Medical Association, 199, 519 (1967)), 199 배지 (Proceeding of the Society for the Biological Medicine, 73, 1 (1950)) 등이 사용된다. 배지의 pH는 바람직하게 약 6 내지 약 8이다. 배양은 일반적으로 약 30℃ 내지 약 40℃에서 수행된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[2570] 식물 세포를 배양하기 위한 배지로서, 예를 들어, MS 배지, LS 배지, B5 배지 등이 사용된다. 배지의 pH는 바람직하게 약 5 내지 약 8이다. 배양은 일반적으로 약 20℃ 내지 약 30℃에서 수행된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[2571] 고등 진핵 세포, 예를 들어, 동물 세포, 곤충 세포, 식물 세포 등이 숙주 세포로서 사용되는 경우, 본원 개시내용의 염기 편집 시스템을 암호화하는 DNA (예를 들어, 아데노신 데아미나제 변이체를 포함함)는 유도성 프로모터 (예를 들어, 메탈로티오네인 프로모터 (중금속 이온에 의해 유도된), 열 쇼크 단백질 프로모터 (열 쇼크에 의해 유도된), Tet-온/Tet-오프 시스템 프로모터 (테트라사이클린 또는 이의 유도체의 부가 또는 제거에 의해 유도됨), 스테로이드-반응성 프로모터 (스테로이드 호르몬 또는 이의 유도체에 의해 유도된) 등)의 조절하에 숙주 세포에 도입되고, 유도 물질은 적절한 단계에서 배지에 첨가되어 (또는 배지로부터 제거되어) 핵산-변형 효소 복합체의 발현을 유도하고, 배양은 소정의 기간동안 수행하여 염기 편집 및 돌연변이의 표적 유전자로의 도입을 수행하고, 염기 편집 시스템의 일과성 발현이 실현될 수 있다.

[2572] 원핵 세포, 예를 들어, 에스코리치아 콜리 등은 유도성 프로모터를 사용할 수 있다. 유도성 프로모터의 예는 lac 프로모터 (IPTG에 의해 유도됨), cspA 프로모터 (냉 쇼크에 의해 유도됨), araBAD 프로모터 (아라비노스에 의해 유도됨) 등을 포함하지만 이에 제한되지 않는다.

[2573] 대안적으로, 상기 언급된 유도성 프로모터는 또한 고등 진핵 세포, 예를 들어, 동물 세포, 곤충 세포, 식물 세포 등이 숙주 세포로서 사용되는 경우 벡터 제거 기전으로서 사용될 수 있다. 즉, 벡터에는 숙주 세포에서 기능하는 복제 오리진이 탑재되고, 단백질을 암호화하는 핵산의 발현의 복제에 필요한 단백질 (예를 들어, 동물 세포에 대해 SV40 및 대형 T 항원, oriP 및 EBNA-1 등)을 암호화하는 핵산은 상기 언급된 유도성 프로모터에 의해 조절된다. 결과로서, 벡터는 유도 물질이 제거되는 경우 유도 물질의 존재하에 자가 복제될 수 있고, 자가 복제는 가용하지 않고, 벡터는 천연적으로 세포 분열과 함께 감소한다 (자가 복제는 Tet-오프 시스템 벡터에서 테트라사이클린 및 독시사이클린의 첨가에 의해 가능하지 않다).

[2574] **염기 편집기를 사용하는 방법**

[2575] 질환 연관된 유전자 및 대립유전자 내 점 돌연변이의 교정은 치료제 및 기본 연구에서의 적용과 함께 유전자 교정을 위한 새로운 전략을 제공한다.

[2576] 본원 개시내용은 본원에 제공된 염기 편집기 시스템에 의해 교정될 수 있는 점 돌연변이와 연관되거나 이에 의해 유발되는 질환으로 진단된 대상체의 치료를 위한 방법을 제공한다. 예를 들어, 일부 구현예에서, 상기 질환, 예를 들어, 유전학적 돌연변이에 의해 유발된 질환을 갖는 대상체에게, 질환 연관된 유전자 내 점 돌연변이를 교정할 수 있는 유효량의 핵염기 편집기 (예를 들어, 아데노신 데아미나제 염기 편집기)를 투여하는 단계를 포함하는 방법이 제공된다. 본원 개시내용은 데아미나제 매개된 유전자 편집에 의해 교정될 수 있는 점 돌

연변이와 연관되거나 이에 의해 유발되는 질환의 치료를 위한 방법을 제공한다. 본원에 제공된 전략 및 융합 단백질로 치료될 수 있는 적합한 질환은 본원 개시내용을 토대로 당업자에게 자명할 것이다.

[2577] 본원에는 질환 또는 장애와 연관된 표적 뉴클레오타이드 서열에서 핵염기를 편집하기 위한 염기 편집기 또는 염기 편집기 시스템을 사용하는 방법이 제공된다. 일부 구현예에서, 염기 편집기 (예를 들어, 아데노신 데아미나제 및 Cas12 도메인을 포함함)의 활성은 점 돌연변이의 교정을 초래한다. 일부 구현예에서, 표적 DNA 서열은 질환 또는 장애와 연관된 G→A 점 돌연변이를 포함하고, 돌연변이체 A 염기의 탈아민화는 질환 또는 장애와 연관되지 않은 서열을 초래한다. 일부 구현예에서, 표적 DNA 서열은 질환 또는 장애와 연관된 T→C 점 돌연변이를 포함하고, 돌연변이체 C 염기의 탈아민화는 질환 또는 장애와 연관되지 않은 서열을 초래한다.

[2578] 일부 구현예에서, 표적 DNA 서열은 단백질을 암호화하고, 점 돌연변이는 코돈 내에 있고 야생형 코돈과 비교하여 돌연변이체 코돈에 의해 암호화된 아미노산에서의 변화를 초래한다. 일부 구현예에서, 돌연변이체 A의 탈아민화는 돌연변이체 코돈에 의해 암호화된 아미노산의 변화를 초래한다. 일부 구현예에서, 돌연변이체 A의 탈아민화는 야생형 아미노산을 암호화하는 코돈을 초래한다. 일부 구현예에서, 돌연변이체 C의 탈아민화는 돌연변이체 코돈에 의해 암호화된 아미노산의 변화를 초래한다. 일부 구현예에서, 돌연변이체 C의 탈아민화는 야생형 아미노산을 암호화하는 코돈을 초래한다. 일부 구현예에서, 대상체는 질환 또는 장애를 갖거나 질환 또는 장애로 진단되었다.

[2579] 일부 구현예에서, 본원에 제공된 아데노신 데아미나제는 DNA의 데옥시아데노신 잔기를 탈아민화시킬 수 있다. 본원 개시내용의 다른 양상은 아데노신 데아미나제 (본원에 기재된 바와 같이 DNA에서 데옥시아데노신을 탈아민화시키는 아데노신 데아미나제) 및 특정 뉴클레오타이드 서열에 결합할 수 있는 도메인 (예를 들어, Cas9 또는 Cpf1 단백질)을 포함하는 융합 단백질을 제공한다. 예를 들어, 아데노신은 전형적으로 시토신 잔기와 염기쌍을 형성하는 이노신 잔기로 전환될 수 있다. 상기 융합 단백질은 특히 핵산 서열의 표적화된 편집을 위해 유용하다. 상기 융합 단백질은, 시험관내 DNA의 표적화된 편집을 위해, 예를 들어, 돌연변이체 세포 또는 동물의 생성을 위해; 표적화된 돌연변이의 도입을 위해, 예를 들어, 동일하거나 또 다른 대상체로 후속적으로 재도입되는 대상체로부터 수득된 세포에서 생체외 세포에서 유전학적 결함의 교정을 위해; 그리고 생체내 표적화된 돌연변이의 도입을 위해 사용될 수 있고, 예를 들어, 유전자 결함의 교정 또는 질환 연관된 유전자에서 G에서 A로, 또는 T에서 C로의 돌연변이에서 질환 연관된 유전자 내 불활성화 돌연변이의 도입은 본원에 제공된 핵염기 편집기를 사용하여 처리될 수 있다. 본원 개시내용은 데아미나제 및 핵염기 편집기를 사용하는, 데아미나제, 융합 단백질, 핵산, 벡터, 세포, 조성물, 방법, 키트, 시스템 등을 제공한다.

[2580] ***Mecp2 유전자에서 뉴클레오티드를 표적화하기 위한 핵염기 편집기의 용도***

[2581] Mecp2 유전자에서 뉴클레오타이드를 표적화하는 핵염기 편집기의 적합성은 본원에 기재된 바와 같이 평가된다. 하나의 구현예에서, 관심 대상의 단일 세포에는 리포터 (예를 들어, GFP)를 암호화하는 소량의 벡터와 함께 본원에 기재된 핵염기 편집기를 암호화하는 핵산 분자 또는 분자들을 형질감염시키거나, 형질도입하거나, 또는 달리 변형시킨다. 이들 세포는 불멸화된 인간 세포주, 예를 들어, 293T, K562 또는 U20S 세포일 수 있다. 대안적으로, 1차 인간 세포가 사용될 수 있다. 세포는 또한 대상체 또는 개체로부터, 예를 들어, 조직 생검, 수술, 혈액, 혈장, 혈청 또는 다른 생물학적 유체로부터 수득될 수 있다. 상기 세포는 궁극적인 세포 표적과 관련될 수 있다.

[2582] 전달은 아래에 추가로 기재된 같이 바이러스 벡터를 사용하여 수행될 수 있다. 하나의 구현예에서, 형질감염은 지질 형질감염 (예를 들어, 리포펙타민 또는 푸겐)을 사용하여 또는 전기천공에 의해 수행될 수 있다. 형질감염 후, GFP의 발현은 형광성 현미경 또는 유동 세포측정에 의해 결정되어 일정하고 고수준의 형질감염을 확인할 수 있다. 이들 예비 형질감염은 상이한 핵염기 편집기를 포함하여 편집기의 어느 조합이 최대 활성을 부여하는지를 결정할 수 있다.

[2583] 핵염기 편집기의 활성은 본원에 기재된 바와 같이, 즉, 표적 유전자를 서열분석하여 표적 서열 내 변경을 검출함으로써 평가한다. 생거 서열분석을 위해, 정제된 PCR 앰플리콘은 플라스미드 골격에 클로닝하고, 형질전환시키고, 소량 분리 (miniprepped)하고, 단일 프라이머를 사용하여 서열분석하였다. 서열분석은 또한 차세대 서열분석 기술을 사용하여 수행될 수 있다. 차세대 서열분석을 사용하는 경우, 앰플리콘은 300-500 bp일 수 있고 의도된 절단 부위는 비대칭으로 위치한다. PCR 후, 차세대 서열분석 어댑터 및 바코드 (예를 들어, 일루미나 멀티플렉스 어댑터 및 인덱스)는 앰플리콘의 말단에, 예를 들어, 고속처리 서열분석 (예를 들어, 일루미나 MiSeq 상에서)에 사용하기 위해 부가될 수 있다.

[2584] 초기 시험에서 최대 수준의 표적 특이적 변경을 유도하는 융합 단백질은 추가의 평가를 위해 선택될 수 있다.

[2585] 특정 구현예에서, 핵염기 편집기를 사용하여 관심 대상의 폴리뉴클레오타이드를 표적화한다. 하나의 구현예에서, 본원에 기재된 바와 같은 핵염기 편집기는 핵산 서열, 예를 들어 RTT-연관된 돌연변이를 보유하는 *Mecp2* 폴리뉴클레오티드를 표적화하여 표적 유전자, 즉 *Mecp2*를 변경하는 데 사용되는 가이드 RNA와 함께 세포 (예를 들어, 뉴런)에 전달된다.

[2586] 일부 구현예에서, 염기 편집기는 가이드 RNA에 의해 표적화되어 관심 대상의서열에 하나 이상의 편집을 도입한다. 일부 구현예에서, *Mecp2* 유전자 내로 도입된 하나 이상의 변경은 하기 표 10에 제시된 바와 같다.

[2587] ***의도된 돌연변이 생성***

[2588] 일부 구현예에서, 본원에 제공된 방법의 목적은 유전자 편집을 통해 기능부전 유전자의 기능을 복구하는 것이다. 일부 구현예에서, 기능부전 유전자의 기능은 의도된 돌연변이를 도입함에 의해 복구된다. 일부 구현예에서, 본원에 제공된 방법을 사용하여 유전자 생성물의 정상 기능을 붕괴시킬 수 있다. 본원에 제공된 핵염기 편집 단백질은 예를 들어, 인간 세포 배양물에서 질환-연관된 돌연변이를 교정함에 의해 시험관내 유전자 편집-기반 인간 치료제에 대해 입증될 수 있다. 당업자라면 본원에 제공된 핵염기 편집 단백질, 예를 들어, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 (예를 들어, Cas9) 및 핵염기 편집 도메인 (예를 들어, 아데노신 데아미나제 도메인)을 포함하는 융합 단백질이 사용되어 임의의 단일 점 A에서 G로 또는 C에서 T 돌연변이를 교정할 수 있음을 이해할 것이다. 첫 번째 경우에, 돌연변이체 A의 I로의 탈아민화는 돌연변이를 교정하고, 후자의 경우 돌연변이체 T와 염기쌍을 이루는 A의 탈아민화에 이어 복제 라운드는 돌연변이를 교정한다.

[2589] 일부 구현예에서, 본원 개시내용은 의도되지 않은 점 돌연변이와 같이, 상당수의 의도되지 않은 돌연변이를 생성하는 것 없이 핵산 (예를 들어, 대상체의 게놈 내 핵산) 내 점 돌연변이와 같은 의도된 돌연변이를 효율적으로 생성하는 염기 편집기를 제공한다. 일부 구현예에서, 의도된 돌연변이는 의도된 돌연변이를 생성하도록 특이적으로 디자인된, 가이드 폴리뉴클레오타이드 (예를 들어, gRNA)에 결합하는 특이적 염기 편집기 (예를 들어, 아데노신 염기 편집기)에 의해 생성되는 돌연변이이다. 일부 구현예에서, 의도된 돌연변이는 질환 또는 장애, 예를 들어, 레트 증후군와 연관된 돌연변이이다. 일부 구현예에서, 의도된 돌연변이는 질환 또는 장애, 예를 들어, 레트 증후군와 연관된 아데닌 (A)에서 구아닌 (G)으로의 점 돌연변이 (예를 들어, SNP)이다. 일부 구현예에서, 의도된 돌연변이는 유전자의 암호화 영역 또는 비-암호화 영역 (예를 들어, 조절 영역 또는 요소) 내 아데닌 (A)에서 구아닌 (G)으로의 점 돌연변이이다. 일부 구현예에서, 의도된 돌연변이는 정지 코돈, 예를 들어, 유전자의 암호화 영역 내 미성숙한 정지 코돈을 생성하는 점 돌연변이이다. 일부 구현예에서, 의도된 돌연변이는 정지 코돈을 제거하는 돌연변이다.

[2590] 일부 구현예에서, 본원에 제공된 임의의 염기 편집기는 1:1 초과인 의도된 돌연변이 대 비의도된 돌연변이 (예를 들어, 의도된 점 돌연변이:비의도된 점 돌연변이)의 비율을 생성시킬 수 있다. 일부 구현예에서, 본원에 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 비의도된 돌연변이 (예를 들어, 의도된 점 돌연변이:비의도된 점 돌연변이)의 비율을 생성시킬 수 있다.

[2591] 염기 편집기 효율의 세부사항은 국제 PCT 출원 번호 PCT/2017/045381 (WO 2018/027078) 및 PCT/US2016/058344 (WO 2017/070632)에 기재되어 있고, 이의 각각은 이의 전문이 본원에 참조로 포함된다. 또한 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); and 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), 이의 전체 내용은 본원에 참조로 인용됨)을 참조한다.

[2592] 일부 구현예에서, 본원에 제공된 방법을 사용하여 하나 이상의 유전자에서 다수의 핵염기 쌍의 편집은 적어도 하나의 의도된 돌연변이의 형성을 유도한다. 일부 구현예에서, 상기 적어도 하나의 의도된 돌연변이의 상기 형

성은 질환-유발 돌연변이의 정밀 교정을 초래한다. 멀티플렉스 편집은 본원에 제공된 임의의 방법 또는 방법의 조합을 사용하여 성취될 수 있는 것으로 인지되어야 한다.

[2593] ***병원성 돌연변이의 정밀 교정***

[2594] 일부 구현예에서, 의도된 돌연변이는 병원성 돌연변이 또는 질환-유발 돌연변이의 정밀 교정이다. 병원성 돌연변이는 병원성 단일 염기 다형체 (SNP)일 수 있거나 SNP에 의해 유발될 수 있다. 예를 들어, 병원성 돌연변이는 유전자에 의해 암호화된 단백질의 아미노산 변화일 수 있다. 또 다른 예에서, 병원성 돌연변이는 유전자에서 병원성 SNP일 수 있다. 정밀 교정은 병원성 돌연변이를 야생형 상태로 되돌리는 것일 수 있다. 일부 구현예에서, 병원성 돌연변이는 질환 또는 장애와 연관된 G→A 점 돌연변이이고, 여기서 A에서 G로의 염기 편집기 (ABE)에 의한 돌연변이 A 염기의 탈아민화는 질환 또는 장애와 연관되지 않은 서열을 생성한다. 일부 구현예에서, 병원성 돌연변이는 C→T 점 돌연변이이다. C→T 점 돌연변이는 예를 들어, A에서 G로의 염기 편집기 (ABE)를 반대 가닥으로 표적화하고 병원성 T 핵염기의 상보체 A를 편집함으로써 교정될 수 있다. 염기 편집기는 병원성 SNP에 또는 병원성 SNP의 상보체를 표적으로 할 수 있다. 병원성 또는 질환-유발 돌연변이 및 기타 서열 변이에 대한 설명의 명명법은 문헌 (참조: den Dunnen, J.T. and Antonarakis, S.E., "Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations: A Discussion." Human Mutation 15:712 (2000), 이의 전체 내용은 본원에 참조로 포함됨)에 기재되어 있다.

[2595] 특정 구현예에서, 질환 또는 장애는 레트 증후군 (RTT)이다. 일부 구현예에서, 병원성 돌연변이는 *Mecp2* 유전자에 있다.

[2596] **전달 시스템**

[2597] ***핵염기 편집기 및 gRNA의 핵산-기반 전달***

[2598] 본원 개시내용에 따라 염기 편집 시스템을 암호화하는 핵산은 당업계에 공지된 방법에 의해 또는 본원에 기재된 바와 같이 시험관내 또는 생체내 대상체에게 투여될 수 있거나, 세포에 전달될 수 있다. 하나의 구현예에서, 핵염기 편집기는 예를 들어, 벡터 (예를 들어, 바이러스 또는 비-바이러스 벡터), 비-벡터-기반 방법 (예를 들어, 나출형 DNA, DNA 복합체, 지질 나노입자를 사용하여) 또는 이들의 조합에 의해 전달될 수 있다.

[2599] 핵염기 편집기를 암호화하는 핵산은 예를 들어, 형질감염 또는 전기천공에 의해 나출형 DNA 또는 RNA로서 세포 (예를 들어, 조혈 세포 또는 이들의 선조체, 조혈 줄기 세포 및/또는 유도된 만능 줄기 세포)에 직접 전달될 수 있거나, 표적 세포에 의한 취득을 촉진시키는 분자 (예를 들어, N-아세틸갈락토사민)에 접합될 수 있다. 핵산 벡터, 예를 들어, 본원에 기재된 벡터가 또한 사용될 수 있다.

[2600] 핵산 벡터는 본원에 기재된 융합 단백질의 도메인을 암호화하는 하나 이상의 서열을 포함할 수 있다. 벡터는 또한 단백질을 암호화하는 서열과 연합된 (예를 들어, 이에 삽입된 또는 융합된) 신호 펩타이드 (예를 들어, 핵 국소화, 핵소체 국소화 또는 미토콘드리아 국소화를 위해)를 암호화하는 서열을 포함할 수 있다. 하나의 예로서, 핵산 벡터는 하나 이상의 핵 국소화 서열 (예를 들어, SV40으로부터 핵 국소화 서열), 및 아데노신 데아미나제 변이체 (예를 들어, TadA*8)를 포함하는 Cas9 암호화 서열을 포함할 수 있다.

[2601] 핵산 벡터는 또한 임의의 적합한 수의 조절/제어 요소, 예를 들어, 프로모터, 인핸서, 인트론, 폴리아데닐화 신호, Kozak 컨센서스 서열, 또는 내부 리보솜 진입 부위 (IRES)를 포함할 수 있다. 이들 요소들은 당업계에 널리 공지되어 있다. 조혈 세포에 대해, 적합한 프로모터는 IFN베타 또는 CD45를 포함할 수 있다.

[2602] 본원 개시내용에 따른 핵산 벡터는 재조합 바이러스 벡터를 포함한다. 예시적인 바이러스 벡터는 본원에 제시된다. 당업계에 공지된 다른 바이러스 벡터가 또한 사용될 수 있다. 추가로, 바이러스 입자는 핵산 및/또는 펩타이드 형태로 염기 편집 시스템 성분을 전달하기 위해 사용될 수 있다. 예를 들어, "속빈" 바이러스 입자는 임의의 적합한 카고 (cargo)를 함유하도록 어셈블리될 수 있다. 바이러스 벡터 및 바이러스 입자는 또한 표적 조직 특이성을 변경하기 위해 표적화 리간드를 혼입하도록 가공될 수 있다.

[2603] 바이러스 벡터에 추가로, 비-바이러스 벡터는 본원의 개시내용에 따른 게놈 편집 시스템을 암호화하는 핵산을 전달하기 위해 사용될 수 있다. 비-바이러스 핵산 벡터의 하나의 중요한 카테고리는 유기 또는 무기일 수 있는 나노입자이다. 나노입자는 당업계에 널리 공지되어 있다. 임의의 적합한 나노입자 디자인을 사용하여 게놈 편집 시스템 성분 또는 상기 성분을 암호화하는 핵산을 전달할 수 있다. 예를 들어, 유기 (예를 들어, 지질 및/또는 중합체) 나노입자는 본원 개시내용의 특정 구현예에서 전달 비히클로서 사용하기 위해 적합할 수 있다. 나노입자 제형 및/또는 유전자 전달에 사용하기 위한 예시적인 지질은 **표 12** (아래)에 나타낸다.

[2604] [표 12]

유전자 전달을 위해 사용되는 지질

지질	약어	특성
1,2-디올레오일-sn-글리세로-3-포스파티딜콜린	DOPC	헬퍼
1,2-디올레오일-sn-글리세로-3-포스파티딜에탄올아민	DOPE	헬퍼
콜레스테롤		헬퍼
N-[1-(2,3-디올레일옥시)프로필]N,N,N-트리메틸암모늄 클로라이드	DOTMA	양이온
1,2-디올레오일옥시-3-트리메틸암모늄-프로판	DOTAP	양이온
디옥타데실아미도글라이실스퍼민	DOGS	양이온
N-(3-아미노프로필)-N,N-디메틸-2,3-비스(도데실옥시)-1-프로판아미늄 브로마이드	GAP-DLRIE	양이온
세틸트리메틸암모늄 브로마이드	CTAB	양이온
6-라우록시헥실 오르니티네이트	LHON	양이온
1-(2,3-디올레오일옥시프로필)-2,4,6-트리메틸피리디늄	2Oc	양이온
2,3-디올레오일옥시-N-[2(스퍼민카복스아미도-에틸]-N,N-디메틸-1-프로판아미늄 트리플루오로아세테이트	DOSPA	양이온
1,2-디올레일-3-트리메틸암모늄-프로판	DOPA	양이온
N-(2-하이드록시에틸)-N,N-디메틸-2,3-비스(테트라데실옥시)-1-프로판아미늄 브로마이드	MDRIE	양이온
디미리스토옥시프로필 디메틸 하이드록시에틸 암모늄 브로마이드	DMRI	양이온
3β-[N-(N',N'-디메틸아미노에탄)-카바모일]콜레스테롤	DC-Chol	양이온
비스-구아니듐-트렌-콜레스테롤	BGTC	양이온
1,3-디오데옥시-2-(6-카복시-스퍼밀)-프로필아미드	DOSPER	양이온
디메틸옥타데실암모늄 브로마이드	DDAB	양이온
디옥타데실아미도글리실스퍼미딘	DSL	양이온
rac-[(2,3-디옥타데실옥시프로필)(2-하이드록시에틸)]-디메틸암모늄 클로라이드	CLIP-1	양이온
rac-[2(2,3-디헥사데실옥시프로필-옥시메틸옥시)에틸]트리메틸암모늄 브로마이드	CLIP-6	양이온
에틸디미리스토일포스파티딜콜린	EDMPC	양이온
1,2-디스테아릴옥시-N,N-디메틸-3-아미노프로판	DSDMA	양이온
1,2-디미리스토일-트리메틸암모늄 프로판	DMTAP	양이온
O,O'-디미리스틸-N-라이실 아스파르테이트	DMKE	양이온
1,2-디스테아로일-sn-글리세로-3-에틸포스포콜린	DSEPC	양이온
N-팔미토일 D-에리트로-스핀고실 카바모일-스퍼민	CCS	양이온
N-t-부틸-NO-테트라데실-3-테트라데실아미노프로피온아미딘	디 C14-아미딘	양이온
옥타데세놀리옥시[에틸-2-헵타데세닐-3 하이드록시에틸] 이미다졸리늄 클로라이드	DOTIM	양이온
N1-콜레스테릴옥시카보닐-3,7-디아자노난-1,9-디아민	CDAN	양이온
2-(3-[비스(3-아미노-프로필)-아미노]프로필아미노)-N-디테트라데실카바모일메-에틸-아세트아미드	RPR209120	양이온
1,2-딜리놀레일옥시-3-디메틸아미노프로판	DLinDMA	양이온
2,2-딜리놀레일-4-디메틸아미노에틸-[1,3]-디옥솔란	DLin-KC2-DMA	양이온
딜리놀레일-메틸-4-디메틸아미노부티레이트	DLin-MC3-DMA	양이온

[2605]

[2606] **표 13**은 유전자 전달 및/또는 나노입자 제형에 사용하기 위한 예시적인 중합체를 열거한다.

[2607] [표 13]

유전자 전달을 위해 사용되는 중합체

중합체	약어
폴리(에틸렌)글리콜	PEG
폴리에틸렌이민	PEI
디티오비스(숙신이미딜프로피오네이트)	DSP
디메틸-3,3'-디티오비스프로피온이미데이트	DTBP
폴리(에틸렌 이민)비스카바메이트	PEIC
폴리(L-라이신)	PLL
히스티딘 변형된 PLL	
폴리(N-비닐피롤리돈)	PVP
폴리(프로필렌이민)	PPI
폴리(아미도아민)	PAMAM
폴리(아미노에틸렌이민)	SS-PAEI
트리에틸렌테트라민	TETA
폴리(β-아미노에스테르)	
폴리(4-하이드록시-L-프롤린 에스테르)	PHP
폴리(알릴아민)	
폴리(α-[4-아미노부틸]-L-글리콜산)	PAGA
폴리(D,L-락틱-코-글리콜산)	PLGA
폴리(N-에틸-4-비닐피리디늄 브로마이드)	
폴리(포스파젠)	PPZ
폴리(포스포에스테르)	PPE
폴리(포스포르아미데이트)	PPA
폴리(N-2-하이드록시프로필메타크릴아미드)	pHPMA
폴리(2-(디메틸아미노)에틸 메타크릴레이트)	pDMAEMA
폴리(2-아미노에틸 프로필렌 포스페이트)	PPE-EA
키토산	
갈락토실화된 키토산	
N-도다실화된 키토산	
히스톤	
콜라겐	
덱스트란-스퍼민	D-SPM

[2608]

[2609] **표 14**는 본원에 기재된 융합 단백질을 암호화하는 폴리뉴클레오타이드에 대한 전달 방법을 요약한다.

[2610] [표 14]

전달	벡터/모드	비-분열 세포로의 전달	발현의 지속기간	게놈 통합	전달된 분자 유형
물리적	(예를 들어, 전기천공, 입자 건, 인산칼슘 형질감염	예	일과성	아니오	핵산 및 단백질
바이러스	레트로바이러스	아니오	안정한	예	RNA
	렌티바이러스	예	안정한	변형과 함께 예/아니오	RNA
	아데노바이러스	예	일과성	아니오	DNA
	아데노-관련 바이러스(AAV)	예	안정한	아니오	DNA
	백시니아 바이러스	예	매우 일과성	아니오	DNA
	헤르페스 심플렉스 바이러스	예	안정한	아니오	DNA
비-바이러스	양이온 리포좀	예	일과성	전달된 것에 의존	핵산 및 단백질
	중합체 나노입자	예	일과성	전달된 것에 의존	핵산 및 단백질
생물학적 비-바이러스	약독화된 세균	예	일과성	아니오	핵산
전달 비히클	가공된 박테리오파아지	예	일과성	아니오	핵산
	포유동물 바이러스-유사 입자	예	일과성	아니오	핵산
	생물학적 리포좀: 적혈구 고스트 및 엑소좀	예	일과성	아니오	핵산

[2611]

[2612] 또 다른 양상에서, 게놈 편집 시스템 성분 또는 상기 성분, 예를 들어, 핵산 결합 단백질, 예를 들어, Cas9 또는 이의 변이체 및 관심 대상의 게놈 핵산 서열을 표적화하는 gRNA를 암호화하는 핵산의 전달은 리보핵산단백질(RNP)을 세포에 전달함으로써 성취될 수 있다. RNP는 핵산 결합 단백질, 예를 들어, 표적화 gRNA와 복합체 형태의 Cas9를 포함한다. RNP는 공지된 방법, 예를 들어, 문헌 (참조: Zuris, J.A. *et al*., 2015, *Nat. Biotechnology*, 33(1):73-80)에 보고된 바와 같이 전기천공, 핵감염 또는 양이온성 지질-매개된 방법을 사용하여 세포에 전달될 수 있다. RNP는 CRISPR 염기 편집 시스템에 사용하기 위해, 특히 1차 세포와 같이 형질감염시키기가 어려운 세포에 사용하기 위해 유리하다. 추가로, RNP는 또한 특히 진핵 세포 프로모터, 예를 들어, CRISPR 플라스미드에 사용될 수 있는 CMV 또는 EF1A가 잘 발현되지 않는 경우 세포에서 단백질 발현과 함께 일어날 수 있는 어려움을 완화시킬 수 있다. 유리하게, RNP의 사용은 외래 DNA의 세포로의 전달을 필요로 하지 않는다. 더욱이, 핵산 결합 단백질 및 gRNA 복합체를 포함하는 RNP는 시간 경과에 따라 분해되기 때문에, RNP의 사용은 표적-이탈 효과를 제한하기 위한 잠재력을 갖는다. 플라스미드 기반 기술의 것과 유사한 방식으로, RNP를 사용하여 결합 단백질 (예를 들어, Cas9 변이체)을 전달하고 상동성 지시된 복구 (HDR)를 지시할 수 있다.

[2613] 염기 편집기 암호화 핵산 분자 발현을 구동시키기 위해 사용되는 프로모터는 AAV ITR을 포함할 수 있다. 이것은 벡터 내 공간을 차지할 수 있는 추가의 프로모터 요소에 대한 필요성을 제거하기 때문에 유리할 수 있다. 확보된 추가의 공간을 사용하여 가이드 핵산 또는 선택가능한 마커와 같은 추가의 요소의 발현을 구동시킬 수 있다. ITR 활성은 상대적으로 약하여 선택된 뉴클레아제의 과발현으로 인한 잠재적 독성을 감소시키기 위해 사용될 수 있다.

[2614] 임의의 적합한 프로모터를 사용하여 염기 편집기 및 경우에 따라 가이드 핵산의 발현을 구동시킬 수 있다. 보편적 발현을 위해, 사용될 수 있는 프로모터는 CMV, CAG, CBh, PGK, SV40, 페리틴 중쇄 또는 경쇄 등을 포함한

다. 뇌 또는 다른 CNS 세포 발현을 위해, 적합한 프로모터는 다음을 포함할 수 있다: 모든 뉴런에 대해 시냅신 I (SynapsinI), 흥분성 뉴런에 대해 CaMKII알파, GABA성 (GABAergic) 뉴런에 대해 GAD67 또는 GAD65 또는 VGAT 등. 간 세포 발현을 위해, 적합한 프로모터는 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.

[2615] 일부 구현예에서, 본원 개시내용의 염기 편집기는 별도의 프로모터가 동일한 핵산 분자 내 염기 편집기 및 상용성 가이드 핵산의 발현을 구동시키도록 하기에 충분히 작은 크기를 갖는다. 예를 들어, 벡터 또는 바이러스 벡터는 염기 편집기를 암호화하는 핵산에 작동적으로 연결된 제1 프로모터 및 가이드 핵산에 작동적으로 연결된 제2 프로모터를 포함할 수 있다.

[2616] 가이드 핵산의 발현을 구동시키기 위해 사용되는 프로모터는 다음을 포함할 수 있다: Pol III 프로모터, 예를 들어, U6 또는 H1, gRNA 아데노 연관된 바이러스 (AAV)를 발현시키기 위해 Pol II 프로모터 및 인트론 카세트의 사용.

[2617] ***바이러스 벡터***

[2618] 따라서, 본원에 기재된 염기 편집기는 바이러스 벡터와 함께 전달될 수 있다. 일부 구현예에서, 본원에 개시된 염기 편집기는 바이러스 벡터에 함유된 핵산 상에 암호화될 수 있다. 일부 구현예에서, 염기 편집기 시스템의 하나 이상의 성분은 하나 이상의 바이러스 벡터 상에 암호화될 수 있다. 예를 들어, 염기 편집기 및 가이드 핵산은 단일 바이러스 벡터 상에 암호화될 수 있다. 다른 구현예에서, 염기 편집기 및 가이드 핵산은 상이한 바이러스 벡터 상에 암호화된다. 어느 경우에나, 염기 편집기 및 가이드 핵산은 각각 프로모터 및 터미네이터에 작동적으로 연결될 수 있다. 바이러스 벡터 상에 암호화된 성분의 조합은 선택된 바이러스 벡터의 카고 크기 제한에 의해 결정될 수 있다.

[2619] 염기 편집기의 전달을 위한 RNA 또는 DNA 바이러스 기반 시스템의 사용은 배양물 중에 또는 숙주에서 바이러스를 특이적 세포에 표적화시키고 바이러스 페이로드를 핵 또는 숙주 세포 게놈으로 트래픽킹하기 위해 고도로 변화된 공정을 이용한다. 바이러스 벡터는 배양물 중에 세포에 또는 환자 (생체내)에 직접 투여될 수 있거나, 이들은 시험관내에서 세포를 처리하기 위해 사용될 수 있고, 변형된 세포는 임의로 환자 (생체외)에게 투여될 수 있다. 통상적인 바이러스 기반 시스템은 유전자 전달을 위해 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관된 및 헤르페스 심플렉스 바이러스 벡터를 포함할 수 있다. 숙주 게놈 내 통합은 레트로바이러스, 렌티바이러스 및 아데노 연관된 바이러스 유전자 전달 방법으로 가능하고 이는 흔히 삽입된 전이유전자의 장기 발현을 초래한다. 추가로, 높은 형질도입 효율은 많은 상이한 세포 유형 및 표적 조직에서 관찰되었다.

[2620] 바이러스 벡터는 렌티바이러스 (예를 들어, HIV 및 FIV-기반 벡터), 아데노바이러스 (예를 들어, AD100), 레트로바이러스 (예를 들어, 몰로니 뮤린 백혈병 바이러스, MML-V), 헤르페스바이러스 벡터 (예를 들어, HSV-2), 및 아데노 연합된 바이러스 (AAV), 또는 다른 플라스미드 또는 바이러스 벡터 유형을 포함할 수 있고 이는 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)로부터 그리고 임상 시험, 및 렌티바이러스, AAV 및 아데노바이러스를 포함하는 임상 시험에 관한 공보로부터의 제형 및 용량을 사용한다. 예를 들어, AAV에 대해, 투여 경로, 제형 및 용량은 미국 특허 제8,454,972호 및 AAV를 포함하는 임상 시험에서와 같을 수 있다. 아데노바이러스에 대해, 투여 경로, 제형 및 용량은 미국 특허 제8,404,658호 및 아데노바이러스를 포함하는 임상 시험에서와 같을 수 있다. 플라스미드 전달에 대해, 투여 경로, 제형 및 용량은 미국 특허 제5,846,946호 및 플라스미드를 포함하는 임상 시험에서와 같을 수 있다. 용량은 평균 70 kg 개체 (예를 들어, 남성 성인 인간)를 기준으로 하거나 추정될 수 있고, 상이한 체중 및 종의 환자, 대상체, 포유류에 대해 조정될 수 있다. 투여 빈도는 연령, 성별, 일반적인 건강, 환자 또는 대상체의 기타 상태, 다루어지는 특정 상태 또는 증상을 포함한 일반적인 요인에 따라 의료 또는 수의사 (예를 들어, 의사, 수의사)의 범위 내이다. 바이러스 벡터는 관심 대상의 조직에 주사될 수 있다. 세포-유형 특이적 염기 편집을 위해, 염기 편집기 및 임의의 가이드 핵산의 발현은 세포-유형 특이적 프로모터에 의해 구동될 수 있다.

[2621] 레트로바이러스의 지향성은 외래 외피 단백질을 통합하여 표적 세포의 잠재적인 표적 집단을 확장함으로써 변경될 수 있다. 렌티바이러스 벡터는 비-분열 세포를 형질도입하거나 감염시킬 수 있고 일반적으로 높은 바이러스 역가를 생성할 수 있는 레트로바이러스 벡터이다. 레트로바이러스 유전자 전달 시스템의 선택은 따라서 표적 조직에 의존한다. 레트로바이러스 벡터는 최대 6-10 kb의 외래 서열에 대한 패키징 용량을 갖는 시스 작용 긴

말단 반복체로 구성된다. 최소 시스 작용 LTR은 벡터의 복제 및 패키징에 충분하며, 이어서 상기 벡터는 치료학적 유전자를 표적 세포에 통합하여 영구적인 전이유전자 발현을 제공하기 위해 사용된다. 광범위하게 사용되는 레트로바이러스 벡터는 뮤린 백혈병 바이러스 (MuLV), 기본 에이프 백혈병 바이러스 (GaLV: gibbon ape leukemia virus), 시미안 면역 결핍 바이러스 (SIV), 인간 면역 결핍 바이러스 (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).

[2622] 레트로바이러스 벡터, 특히 렌티바이러스 벡터는 표적 세포로의 효율적인 통합을 위해 소정의 길이 보다 작은 폴리뉴클레오타이드 서열을 필요로 할 수 있다. 예를 들어, 9 kb 초과 길이의 레트로바이러스 벡터는 보다 작은 크기의 것들과 비교하여 낮은 바이러스 역가를 유도할 수 있다. 일부 양상에서, 본원 개시내용의 염기 편집기는 레트로바이러스 벡터를 통한 표적 세포로의 효율적인 패키징 및 전달을 가능하게 하기 위해 충분한 크기를 갖는다. 일부 구현예에서, 염기 편집기는 가이드 핵산 및/또는 표적화할 수 있는 뉴클레아제 시스템의 기타 성분들과 함께 발견되는 경우에도 효율적인 팩킹 및 전달을 가능하게 하기 위한 크기를 갖는다.

[2623] 일과성 발현이 바람직한 적용에서 아데노바이러스 기반 시스템이 사용될 수 있다. 아데노바이러스 기반 벡터는 많은 세포 유형에서 매우 높은 효율로 형질도입할 수 있고 세포 분열을 요구하지 않는다. 상기 벡터를 사용하여 고역가 및 고수준의 발현이 수득되었다. 상기 벡터는 상대적으로 단순한 시스템에서 대량으로 생성될 수 있다. 아데노 연관된 바이러스 ("AAV") 벡터는 또한 예를 들어, 핵산 및 펩타이드의 시험관내 생성에서 그리고 생체내 및 생체외 유전자 치료요법 과정을 위해 표적 핵산을 세포에 형질도입하기 위해 사용될 수 있다 (예를 들어, 문헌 (West *et al.*, Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994)) 참조). 재조합 AAV 벡터의 작제는 문헌 (미국 특허 번호 5,173,414; 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 기재되어 있다.

[2624] AAV는 파르보바이러스 패밀리에 속하는 작은 단일 가닥의 DNA 의존성 바이러스이다. 4.7 kb 야생형 (wt) AAV 게놈은 각각 4개의 복제 단백질과 3개의 캡시드 단백질을 암호화하는 2개의 유전자로 구성되고, 145-bp 역위 말단 반복체 (ITR)에 의해 양측면 상에 플랭킹된다. 비리온은 동일한 개방 판독 프레임의로부터 이지만 상이한 스플라이싱 (Vp1) 및 대안적 해독 개시 부위 (각각 Vp2 및 Vp3)로부터 1:1:10의 비율로 생성되는 3개의 캡시드 단백질인 Vp1, Vp2, 및 Vp3으로 구성된다. Vp3은 비리온에서 가장 풍부한 서브유닛이고 바이러스의 지향성을 한정하는 세포 표면에서 수용체 인지에 관여한다. 바이러스 감염성에 기능하는 포스포리파제 도메인은 Vp1의 고유 N 말단에서 동정되었다.

[2625] wt AAV와 유사하게, 재조합 AAV (rAAV)는 벡터 전이유전자 카세트를 플랭킹하기 위해 *시스*-작용 145-bp ITR을 사용하고 외래 DNA의 패키징을 위해 최대 4.5 kb를 제공한다. 감염에 이어서, rAAV는 본원 개시내용의 융합 단백질을 발현할 수 있고 환형 헤드 투 테일 컨카테머로 에피좀적으로 존재함에 의해 숙주 게놈으로의 통합 없이 지속할 수 있다. 상기 시험관내 및 생체내 시스템을 사용한 다수의 rAAV 성공 사례가 있지만, 제한된 패키징 능력은 유전자의 암호화 서열의 길이가 wt AAV 게놈과 크기에 동일하거나 이 보다 큰 경우 AAV-매개된 유전자 전달의 용도를 제한하였다.

[2626] 바이러스 벡터는 상기 적용을 기준으로 선택될 수 있다. 예를 들어, 생체내 유전자 전달을 위해, AAV는 다른 바이러스 벡터 보다 유리할 수 있다. 일부 구현예에서, AAV는 낮은 독성을 가능하게 하고, 이는 면역 반응을 활성화시킬 수 있는 세포 입자의 초원심분리를 필요로 하지 않는 정제 방법으로 인한 것일 수 있다. 일부 구현예에서, AAV는 이것이 숙주 게놈에 통합하지 않기 때문에 삽입 돌연변이를 유발할 가능성을 낮게 한다. 아데노바이러스는 통상적으로 이들이 유도하는 강한 면역원성 반응 때문에 백신으로서 사용된다. 바이러스 벡터의 패키징 능력은 벡터에 패키징될 수 있는 염기 편집기의 크기를 제한할 수 있다.

[2627] AAV는 2개의 145개 염기 역위 말단 반복체 (ITR)를 포함하는 약 4.5 Kb 또는 4.75 Kb의 패키징 능력을 갖는다. 이것은 프로모터 뿐만 아니라 기재된 염기 편집기를 의미하고 전사 터미네이터는 단일 바이러스 벡터에 피팅될 수 있다. 4.5 또는 4.75 Kb보다 큰 작제물은 유의적으로 감소된 바이러스 생성을 유도할 수 있다. 예를 들어, SpCas9는 매우 크고, 유전자 자체는 4.1 Kb 초과이고, 이는 AAV에 팩킹되기 어렵게 한다. 따라서, 본원 개시내용의 구현예는 통상적인 염기 편집기 보다 길이가 보다 짧은 기재된 염기 편집기를 사용함을 포함한다. 일부 예에서, 염기 편집기는 4 kb 미만이다. 개시된 염기 편집기는 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4

kb, 3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, 또는 1.5 kb 미만일 수 있다. 일부 구현예에서, 개시된 염기 편집기는 4.5 kb 이하의 길이이다.

[2628] AAV는 AAV1, AAV2, AAV5 또는 이들의 임의의 조합일 수 있다. 당업자는 표적화될 세포에 관련하여 AAV의 유형을 선택할 수 있고; 예를 들어, 당업자는 뇌 또는 신경 세포를 표적화하기 위해 AAV 혈청형 1, 2, 5 또는 하이브리드 캡시드 AAV1, AAV2, AAV5 또는 이들의 임의의 조합을 선택할 수 있고; 당업자는 심장 조직을 표적화하기 위해 AAV4를 선택할 수 있다. AAV8은 간으로의 전달을 위해 유용하다. 이들 세포에 관한 특정 AAV 혈청형의 목록은 문헌 (참조: Grimm, D. et al, J. Virol. 82: 5887-5911 (2008))에서 찾을 수 있다.

[2629] 렌티바이러스는 유사분열 및 유사분열 후 세포 둘 다에서 이들의 유전자를 감염시키고 발현하는 능력을 갖는 복합 레트로바이러스이다. 가장 통상적으로 공지된 렌티바이러스는 인간 면역결핍 바이러스 (HIV)이고, 이는 광범위한 세포 유형을 표적화하기 위해 기타 바이러스의 외피 당단백질을 사용한다.

[2630] 렌티바이러스는 다음과 같이 제조될 수 있다. pCasES10 (렌티바이러스 전달 플라스미드 골격을 함유하는)을 클로닝 한 후, 낮은 계대 (p=5)에서 HEK293FT는 T-75 플라스크에 씨딩하고 10% 태아 소 혈청을 갖고 항생제가 없는 DMEM에서 형질감염시키기전 날에 50% 컨플루언스하도록 하였다. 20시간 후, 배지는 OptiMEM (무혈청) 배지로 갈아주고 형질감염은 4시간 후 수행하였다. 세포는 10 μg의 렌티바이러스 전달 플라스미드 (pCasES10) 및 하기의 패키징 플라스미드로 형질감염시킨다: 5 μg의 pMD2.G (VSV-g 슈도타입), 및 7.5 μg의 psPAX2 (gag/pol/rev/tat). 형질감염은 양이온성 지질 전달제 (50 μl의 리포펙타민 2000 및 100 μl의 플러스 시약 (Plus reagent))를 갖는 4 mL OptiMEM에서 수행할 수 있다. 6시간 후, 배지는 10% 태아 소 혈청을 갖는 무항생제 DMEM으로 갈아준다. 이들 방법은 세포 배양 동안에 혈청을 사용하지만 무혈청 방법이 바람직하다.

[2631] 렌티바이러스는 다음과 같이 정제될 수 있다. 바이러스 상등액은 48시간 후 수거한다. 상등액은 먼저 파쇄물을 제거하고 0.45 μm 낮은 단백질 결합 (PVDF) 필터를 통해 여과하였다. 이어서, 이들을 24,000 rpm에서 2시간 동안 원심분리에서 회전시킨다. 바이러스 펠렛은 4℃에서 밤새 50 μl의 DMEM에서 재현탁시킨다. 이어서, 이들을 분취하고 즉시 -80℃에서 동결시킨다.

[2632] 또 다른 구현예에서, 말 감염성 빈혈 바이러스 (EIAV)를 기반으로 하는 최소 비-영장류 렌티바이러스 벡터가 또한 고려된다. 또 다른 구현예에서, RetinoStat.RTM은 망막하 주사를 통해 전달되는 것으로 고려되는 혈관신생 억제 단백질 엔도스타틴 및 안지오스타틴을 발현하는 말 감염성 빈혈 바이러스 기반 렌티바이러스 유전자 치료 요법 벡터이다. 또 다른 구현예에서, 자가-불활성화 렌티바이러스 벡터의 사용이 고려된다.

[2633] 시스템의 임의의 RNA, 예를 들어, 가이드 RNA 또는 염기 편집기-암호화 mRNA는 RNA 형태로 전달될 수 있다. 염기 편집기-암호화 mRNA는 시험관내 전사를 사용하여 생성될 수 있다. 예를 들어, 뉴클레아제 mRNA는 하기의 요소들을 함유하는 PCR 카세트를 사용하여 합성될 수 있다: T7 프로모터, 선택적 kozak 서열 (GCCACC), 뉴클레아제 서열, 및 3' UTR, 예를 들어, 베타 글로빈-폴리A 테일로부터의 3' UTR. 카세트는 T7 폴리머라제에 의한 전사를 위해 사용될 수 있다. 가이드 폴리뉴클레오타이드 (예를 들어, gRNA)는 또한 T7 프로모터에 이어서 서열 "GG", 및 가이드 폴리뉴클레오타이드 서열을 함유하는 카세트로부터의 시험관내 전사를 사용하여 전사될 수 있다.

[2634] 발현을 증진시키고 가능한 독성을 감소시키기 위해, 염기 편집기-암호화 서열 및/또는 가이드 핵산은 하나 이상의 변형된 뉴클레오사이드를 포함하도록, 예를 들어, 슈도-U 또는 5-메틸-C를 사용하여 변형될 수 있다.

[2635] AAV 벡터의 작은 패키징 능력은 상기 크기를 초과하는 다수의 유전자의 전달 및/또는 큰 생리학적 조절 요소들의 사용을 어렵게 한다. 이들 과제는 예를 들어, 전달될 단백질(들)을 2개 이상의 단편으로 나눔으로써 해결될 수 있고, 여기서, 상기 N-말단 단편은 스플릿 인테인-N으로 융합되고, C-말단 단편은 스플릿 인테인-C에 융합된다. 이어서, 이들 단편은 2개 이상의 AAV 벡터로 패키징된다. 본원에 사용된 바와 같은, "인테인"은 플랭킹 N-말단 및 C-말단 익스테인 (예를 들어, 연결될 단편)을 연결하는 자가 스플라이싱 인트론 (예를 들어, 펩타이드)을 언급한다. 이종성 단백질 단편을 연결하기 위한 특정 인테인의 용도는 예를 들어, 문헌 (참조: Wood *et al.*, J. Biol. Chem. 289(21); 14512-9 (2014))에 기재되어 있다. 예를 들어, 단백질 단편을 분리하기 위해 융합되는 경우, 인테인 IntN 및 IntC는 서로 인지하고, 이들 자체를 스플라이스 제거하고 이들이 융합된 단백질 단편의 플랭킹 N- 및 C-말단 엑스테인을 동시 연결하여 2개의 단백질 단편으로부터 전장 단백질을 재구성한다. 다른 적합한 인테인은 당업자에게 자명할 것이다.

[2636] 본원 개시내용의 융합 단백질의 단편은 길이가 다양할 수 있다. 일부 구현예에서, 단백질 단편은 2개 아미노산

내지 약 1000개 아미노산 길이의 범위이다. 일부 구현예에서, 단백질 단편은 약 5개 아미노산 내지 약 500개 아미노산 길이의 범위이다. 일부 구현예에서, 단백질 단편은 약 20개 아미노산 내지 약 200개 아미노산 길이의 범위이다. 일부 구현예에서, 단백질 단편은 약 10개 아미노산 내지 약 100개 아미노산 길이의 범위이다. 다른 길이의 적합한 단백질 단편은 당업자에게 자명할 것이다.

[2637] 하나의 구현예에서, 이중 AAV 벡터는 대형 전이유전자 발현 카세트를 2개의 개별 반쪽 (5' 및 3' 말단, 또는 헤드 및 테일)으로 스플릿팅함에 의해 생성하고, 여기서, 카세트의 각각의 반쪽은 단일 AAV 벡터 (< 5 kb)에 패키징된다. 이어서, 전장 전이유전자 발현 카세트의 재어셈블리는 이중 AAV 벡터 둘 다에 이어서 다음과 같은 것들에 의한 동일한 세포의 동시 감염시 성취된다: (1) 5' 게놈과 3' 게놈 (이중 AAV 중첩 벡터) 간의 상동성 재조합 (HR); (2) 5' 게놈과 3' 게놈의 ITR-매개된 테일 투 헤드 컨카테머화 (이중 AAV *트랜*스-스플라이싱 벡터); 또는 (3) 이들 2개의 기전의 조합 (이중 AAV 하이브리드 벡터). 생체내 이중 AAV 벡터의 사용은 전장 단백질의 발현을 유도한다. 이중 AAV 벡터 플랫폼의 사용은 크기가 >4.7 kb인 전이유전자에 대해 효율적이고 실행 가능한 유전자 전달 전략을 제공한다.

[2638] ***인테인***

[2639] 일부 구현예에서, 뉴클레아제 (예를 들어, Cas9)의 일부 또는 단편은 인테인에 융합된다. 뉴클레아제는 인테인의 N-말단 또는 C-말단에 융합될 수 있다. 일부 구현예에서, 융합 단백질의 부분 또는 단편은 인테인에 융합되고 AAV 캡시드 단백질에 융합된다. 인테인, 뉴클레아제 및 캡시드 단백질은 함께 임의의 정렬 (예를 들어, 뉴클레아제-인테인-캡시드, 인테인-뉴클레아제-캡시드, 캡시드-인테인-뉴클레아제 등)로 융합될 수 있다. 일부 구현예에서, 인테인의 N-말단은 융합 단백질의 C-말단에 융합되고 인테인의 C-말단은 AAV 캡시드 단백질의 N-말단에 융합된다.

[2640] 인테인 (개재 단백질)은 다양한 유기체에서 발견되는 자동 처리 도메인으로, 단백질 스플라이싱으로 공지된 공정을 수행한다. 단백질 스플라이싱은 펩타이드 결합의 절단 및 형성 둘 다로 구성된 다단계 생화학적 반응이다. 단백질 스플라이싱의 내인성 기질은 인테인 함유 유기체에서 발견되는 단백질이지만, 인테인은 또한 실제로 임의의 폴리펩타이드 골격을 화학적으로 조작하기 위해 사용될 수 있다.

[2641] 단백질 스플라이싱에서, 인테인은 그 자체를 2개의 펩타이드 결합을 절단함에 의해 전구체 폴리펩타이드로부터 절제 제거됨으로써 새로운 펩타이드 결합의 형성을 통해 플랭킹 익스테인 (외부 단백질) 서열을 연결한다. 상기 재정렬은 해독 후 (또는 능히 해독과 동시에) 일어난다. 인테인-매개된 단백질 스플라이싱은 자발적으로 일어나고 단지 인테인 도메인의 폴딩을 요구한다.

[2642] 인테인의 약 5%는 스플릿 인테인이고, 이는 전사되어 2개의 별도의 폴리펩타이드인 N-인테인 및 C-인테인으로서 해독되고, 이의 각각은 하나의 익스테인에 융합된다. 해독 시, 인테인 단편은 자발적으로 및 비공유적으로 카노니칼 인테인 구조로 어셈블리하여 트랜스로 단백질 스플라이싱을 수행한다. 단백질 스플라이싱의 기전은 일련의 아실-전달 반응을 가능하게 하여 이는 인테인-익스테인 접합부에서 2개의 펩타이드 결합의 절단 및 N- 및 C-익스테인 간의 새로운 펩타이드 결합의 형성을 유도한다. 상기 공정은 N-익스테인과 인테인의 N-말단을 연결하는 펩타이드 결합의 활성화에 의해 개시된다. 실제로 모든 인테인은 C-말단 N-익스테인 잔기의 카보닐 탄소를 공격하는 이들의 N-말단에서 시스테인 또는 세린을 갖는다. 상기 N의 O/S 아실로의 전환은 통상적으로 발견되는 아스파르테이트와 함께 보존된 트레오닌 및 히스티딘 (TXXH 모티프로서 언급됨)에 의해 촉진되어, 선형 (티오)에스테르 중간체의 형성을 초래한다. 이어서, 상기 중간체는 시스테인, 세린 또는 트레오닌인 제1 C-익스테인 잔기 (+1)의 친핵성 공격에 의한 트랜스-(티오) 에스테르화에 적용한다. 수득한 분지된 (티오)에스테르 중간체는 고유 전환을 통해 분리된다: 인테인의 고도로 보존된 C-말단 아스파라긴의 폐환. 상기 과정은 히스티딘 (고도로 보존된 HNF 모티프에서 발견되는) 및 끝에서 두번째 히스티딘에 의해 촉진되고, 또한 아스파르테이트를 포함할 수 있다. 이러한 숙신이미드 형성 반응은 인테인을 반응 복합체로부터 절제하고 비-펩타이드 연결을 통해 부탁된 익스테인을 잔류시킨다. 상기 구조는 신속하게 인테인-독립적 양상으로 안정한 펩타이드 결합으로 재정렬한다.

[2643] 일부 구현예에서, 염기 편집기 (예를 들어, ABE, CBE)의 N-말단 단편은 스플릿 인테인-N에 융합되고, C-말단 단편은 스플릿 인테인-C에 융합된다. 이어서, 이들 단편은 2개 이상의 AAV 벡터로 패키징된다. 이종성 단백질 단편을 연결하기 위한 특정 인테인의 용도는 예를 들어, 문헌 (참조: Wood *et al*., J. Biol. Chem. 289(21); 14512-9 (2014))에 기재되어 있다. 예를 들어, 단백질 단편을 분리하기 위해 융합되는 경우, 인테인 IntN 및 IntC는 서로 인지하고, 이들 자체를 스플라이스 제거하고 이들이 융합된 단백질 단편의 플랭킹 N- 및 C-말단 엑스테인을 동시 연결하여 2개의 단백질 단편으로부터 전장 단백질을 재구성한다. 다른 적합한 인테인은 당업자

에게 자명할 것이다.

[2644] 일부 구현예에서, ABE는 SpCas9의 선택된 영역 내 Ala, Ser, Thr, 또는 Cys 잔기에서 N- 및 C-말단 단편으로 스플릿하였다. 이들 영역은 Cas9 결정 구조 분석에 의해 동정된 루프 영역에 상응한다. 각각의 단편의 N-말단은 인테인-N에 융합되고, 각각의 단편의 C-말단은 하기 서열 중 굵은 대문자로 지적된 아미노산 위치 S303, T310, T313, S355, A456, S460, A463, T466, S469, T472, T474, C574, S577, A589, 및 S590에서 인테인 C에 융합된다.

```
   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 nedilediv 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
```

[2645]

[2646] ***약제학적 조성물***

[2647] 본원 개시내용의 다른 양상은 본원에 기재된 임의의 염기 편집기, 융합 단백질 또는 융합 단백질-가이드 폴리뉴클레오타이드 복합체를 포함하는 약제학적 조성물에 관한 것이다. 일부 구현예에서, 약제학적 조성물은 약제학적으로 허용되는 담체를 추가로 포함한다. 일부 구현예에서, 약제학적 조성물은 추가의 제제 (예를 들어, 특이적 전달을 위해, 반감기를 증가시키기 위해 또는 다른 치료학적 화합물을 위해)를 포함한다.

[2648] 약제학적으로 허용되는 담체로서 작용할 수 있는 물질의 일부 비제한적인 예는 다음을 포함한다: (1) 당, 예를 들어, 락토스, 글루코스 및 슈크로스; (2) 전분, 예를 들어, 옥수수 전분 및 감자 전분; (3) 셀룰로스, 및 이의 유도체, 예를 들어, 나트륨 카복시메틸 셀룰로스, 메틸셀룰로스, 에틸 셀룰로스, 미세결정 셀룰로스 및 셀룰로스 아세테이트; (4) 분말 트라가칸트; (5) 맥아; (6) 젤라틴; (7) 윤활제, 예를 들어, 마그네슘 스테아레이트, 나트륨 라우릴 설페이트 및 탈크; (8) 부형제, 예를 들어, 코코아 버터 및 좌제 왁스; (9) 오일, 예를 들어, 땅콩유, 면화씨유, 잇꽃유, 참깨유, 올리브유, 옥수수유 및 대두유; (10) 글리콜, 예를 들어, 프로필렌 글리콜; (11) 폴리올, 예를 들어, 글리세린, 소르비톨, 만니톨 및 폴리에틸렌 글리콜 (PEG); (12) 에스테르, 예를 들어, 에틸 올레에이트 및 에틸 라우레이트; (13) 한천; (14) 완충제, 예를 들어, 수산화마그네슘 및 수산화알루미늄; (15) 알긴산; (16) 발열성 물질 제거수; (17) 등장성 식염수; (18) 링거 용액; (19) 에틸 알콜; (20) pH 완충액; (21) 폴리에스테르, 폴리카보네이트 및/또는 다중무수물; (22) 벌크제, 예를 들어, 폴리펩타이드 및 아미노산; (23) 혈청 알콜, 예를 들어, 에탄올; 및 (23) 약제학적 제형에 사용되는 다른 비독성 상용성 물질. 완충제, 습윤화제, 유화제, 희석제, 캡슐화제, 피부 침투 증진제, 착색제, 이형제, 코팅제, 감미제, 향제, 방향제, 보존제 및 항산화제는 또한 제형 중에 존재할 수 있다. 예를 들어, 담체는 식염수, 완충 식염수, 덱스트로스, 아르기닌, 슈크로스, 물, 글리세롤, 에탄올, 소르비톨, 덱스트란, 나트륨 카복시메틸 셀룰로스, 및 이들의 조합물을 포함할 수 있지만 이에 제한되지 않는다.

[2649] 약제학적 조성물은 약 5.0 내지 약 8.0 범위에서와 같이 생리학적 pH를 반영하는 소정의 수준에서 제형의 pH를 유지하기 위해 하나 이상의 pH 완충 화합물을 포함할 수 있다. 수성 액체 제형에 사용되는 pH 완충 화합물은 아미노산 또는 아미노산의 혼합물, 예를 들어, 히스티딘, 또는 히스티딘 및 글라이신과 같은 아미노산의 혼합물일 수 있다. 대안적으로, pH 완충 화합물은 일부 구현예에서 약 5.0 내지 약 8.0의 범위에서와 같이 소정의 수준에서 제형의 pH를 유지하고 칼슘 이온을 킬레이팅하지 않는 제제이다. 상기 pH 완충 화합물의 예시적인 예는 이미다졸 및 아세테이트 이온을 포함하지만 이에 제한되지 않는다. pH 완충 화합물은 소정의 수준에서 제형의 pH를 유지하기에 적합한 임의의 양으로 존재할 수 있다.

[2650] 약제학적 조성물은 또한 하나 이상의 삼투 조절제, 즉, 제형의 삼투 성질 (예를 들어, 등장성, 삼투압 및/또는 삼투 압력)을 수용자 개체의 혈류 및 혈액 세포에 허용되는 수준까지 조절하는 화합물을 함유할 수 있다. 삼투 조절제는 칼슘 이온을 킬레이팅하지 않는 제제일 수 있다. 삼투 조절제는 제형의 삼투 성질을 조절하는 당업자에게 공지되거나 가용한 임의의 화합물일 수 있다. 당업자는 경험적으로 본 발명의 제형에 사용하기 위해 소정의 삼투 조절제의 적합성을 결정할 수 있다. 삼투 조절제의 적합한 유형의 예시적인 예는 다음을 포함하지만 이에 제한되지 않는다: 염, 예를 들어, 염화나트륨 및 나트륨 아세테이트; 당, 예를 들어, 슈크로스, 덱스트로스, 및 만니톨; 아미노산, 예를 들어, 글라이신; 및 하나 이상의 이들 제제의 혼합물 및/또는 제제의 유형. 삼투 조절제(들)는 제형의 삼투 성질을 조절하기에 충분한 임의의 농도로 존재할 수 있다.

[2651] 일부 구현예에서, 약제학적 조성물은 대상체에게 전달하기 위해, 예를 들어, 유전자 편집을 위해 제형화된다. 일부 구현예에서, 본원에서 고려되는 약제학적 조성물의 투여는 주입, 투입 또는 비경구적으로를 포함하지만 이에 제한되지 않는 통상적인 기술을 사용하여 수행될 수 있다. 일부 구현예에서, 비경구 투여는 혈관내, 정맥내, 근육내, 동맥내, 척수강내, 종양내, 피내, 복강내, 기관으로, 피하로, 피내, 관절내, 캡슐내, 지주막하 및 흉골내 주입하거나 주사함을 포함한다. 일부 구현예에서, 본원에 기재된 약제학적 조성물을 투여하는 적합한 경로는 제한 없이 다음을 포함한다: 국소, 피하, 경피, 피내, 병변내, 관절내, 복강내, 방광내, 경점막, 잇몸, 치내, 와우내, 경고막, 기관내, 경막외, 척수강내, 근육내, 정맥내, 혈관내, 골내, 안주변, 종양내, 뇌내 및 뇌실내 투여.

[2652] 일부 구현예에서, 본원에 기재된 약제학적 조성물은 국소적으로 환부 (예를 들어, 종양 부위)에 투여된다. 일부 구현예에서, 본원에 기재된 약제학적 조성물은 주사에 의해, 카테터에 의해, 좌제에 의해, 또는 이식체에 의해 대상체에게 투여되고, 이식체는 다공성, 비-다공성, 또는 젤라틴성 물질이고, 이는 막, 예를 들어, 시알라스틱 막 또는 섬유를 포함한다.

[2653] 다른 구현예에서, 본원에 기재된 약제학적 조성물은 조절 방출 시스템으로 전달된다. 하나의 구현예에서, 펌프가 사용될 수 있다 (예를 들어, 문헌 (Langer, 1990, Science 249: 1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald *et al.*, 1980, Surgery 88:507; Saudek et al, 1989, N. Engl. J. Med. 321:574) 참조). 또 다른 실시양태에서, 중합체 물질이 사용될 수 있다. (예를 들어, 문헌 (Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61) 참조). 또한 문헌 (Levy *et al.*, 1985, Science 228: 190; During *et al.*, 1989, Ann. neurol. 25:351; Howard *et al.*, 1989, J. neurosurg. 71: 105)을 참조한다. 다른 조절 방출 시스템은 예를 들어, 상기 문헌 (Langer)에서 논의된다.

[2654] 일부 구현예에서, 약제학적 조성물은 대상체, 예를 들어, 인간에게 정맥내 또는 피하 투여를 위해 적합한 조성물로서 통상의 과정에 따라 제형화된다. 일부 구현예에서, 주사에 의한 투여용 약제학적 조성물은 가용화제로서 멸균 등장성 용도에서의 용제 및 주사 부위에서 통증을 완화하기 위한 리도카인과 같은 국소 마취제이다. 일반적으로, 성분은 별도로 또는 단일 투여 형태로 함께 혼합되어 공급되고, 예를 들어, 활성제의 양을 지적하는 앰푸울 또는 샤쉐와 같은 기밀하게 밀봉된 컨테이너에 동결건조된 분말 또는 무수 농축물로서 공급된다. 약제가 주입에 의해 투여되어야만 하는 경우, 멸균 약제학적 등급수 또는 식염수를 함유하는 주입병으로 분배될 수 있다. 약제학적 조성물이 주사에 의해 투여되는 경우, 멸균 주사용수 또는 식염수의 앰푸울이 제공되어 성분들은 투여 전 혼합될 수 있다.

[2655] 전신 투여를 위한 약제학적 조성물은 액체, 예를 들어, 멸균 식염수, 락테이트화된 링거 또는 행크 용액일 수 있다. 추가로, 약제학적 조성물은 고체 형태일 수 있고 사용 직전 재용해되거나 현탁될 수 있다. 동결건조된 형태가 또한 고려된다. 약제학적 조성물은 또한 비경구 투여에 적합한 리포좀 또는 미세결정과 같은 지질 입자 또는 소포 내에 함유될 수 있다. 입자는 조성물이 그 안에 함유되어 있는 한, 단층 또는 다층과 같은 임의의

적합한 구조일 수 있다. 화합물은 푸소겐성 지질 디올레오일포스파티딜에탄올아민 (DOPE), 저수준 (5-10 mol %)의 양이온성 지질을 함유하는 "안정화된 플라스미드-지질 입자" (SPLP)에 포집될 수 있고, 폴리에틸렌글리콜 (PEG) 코팅에 의해 안정화될 수 있다 (참조: Zhang Y. P. *et al.*, Gene Ther. 1999, 6: 1438-47). 양으로 하전된 지질, 예를 들어, N-[1-(2,3-디올레오일옥시)프로필]-N,N,N-트리메틸-암모늄메틸설페이트, 또는 "DOTAP"는 특히 상기 입자 및 소포를 위해 바람직하다. 상기 지질 입자의 제조는 널리 공지되어 있다. 예를 들어, 미국 특허 제4,880,635호; 제4,906,477호; 제4,911,928호; 제4,917,951호; 제4,920,016호; 및 제4,921,757호를 참조하고; 이의 각각은 본원에 참조로 포함된다.

[2656] 본원에 기재된 약제학적 조성물은 예를 들어, 유닛 용량으로서 투여되거나 패키징될 수 있다. 본원 개시내용의 약제학적 조성물을 참조로 사용되는 경우 용어 "유닛 용량"은 대상체에 대한 유닛 용량으로서 적합한 물리적으로 구분된 유닛을 언급하고, 각각의 유닛은 요구되는 희석제; 즉, 담체 또는 비히클과 연합된 목적하는 치료학적 효과를 생성하도록 계산된 소정량의 활성 물질을 함유한다.

[2657] 추가로, 약제학적 조성물은 (a) 동결건조된 형태로 본원 개시내용의 화합물을 함유하는 컨테이너 및 (b) 약제학적으로 허용되는 희석제 (예를 들어, 본원 개시내용의 동결건조된 화합물의 재구성 또는 희석을 위해 사용되는 멸균성)를 함유하는 제2 컨테이너를 포함하는 약제학적 키트로서 제공될 수 있다. 임의로 이러한 용기(들)는, 의약품 또는 생물학적 제품의 제조, 사용 또는 판매를 규제하는 정부 기관에 의해 규정된 형태의 통지와 관련되어 있으며, 당해 통지는 인간 투여를 위한 제조, 사용 또는 판매에 대한 기관의 승인을 반영한다.

[2658] 또 다른 양상에서, 레트 증후군 (RTT)의 치료에 유용한 물질을 함유하는 제품이 포함된다. 일부 구현예에서, 제품은 컨테이너 및 표지를 포함한다. 적합한 컨테이너는 예를 들어, 병, 바이알, 주사기 및 시험 튜브를 포함한다. 컨테이너들은 유리 또는 플라스틱과 같은 다양한 물질로부터 형성될 수 있다. 일부 구현예에서, 컨테이너는 본원에 기재된 질환을 치료하기 위해 효과적인 조성물을 유지하고 멸균 접근 포트를 가질 수 있다. 예를 들어, 컨테이너는 피하 주사 바늘에 의해 천공될 수 있는 스토퍼를 갖는 정맥내 용액 백 또는 바이알일 수 있다. 조성물 중에 활성제는 본원 개시내용의 화합물이다. 일부 구현예에서, 컨테이너 상에 또는 이와 연합된 표지는 조성물이 선택된 질환을 치료하기 위해 사용됨을 지적한다. 제품은 포스페이트-완충 식염수, 링거 용액, 또는 덱스트로스 용액과 같은 약제학적으로 허용되는 완충액을 포함하는 제2 컨테이너를 추가로 포함할 수 있다. 이것은 다른 완충제, 희석제, 필터, 바늘, 주사기 및 사용 설명서가 있는 팩키지 삽입물을 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[2659] 일부 구현예에서, 본원에 기재된 임의의 융합 단백질, gRNA, 및/또는 복합체는 약제학적 조성물의 일부로서 제공된다. 일부 구현예에서, 약제학적 조성물은 본원에 제공된 임의의 융합 단백질을 포함한다. 일부 구현예에서, 약제학적 조성물은 본원에 제공된 임의의 복합체를 포함한다. 일부 구현예에서, 약제학적 조성물은 gRNA와 양이온성 지질과 복합체를 형성하는 RNA-가이드된 뉴클레아제 (예를 들어, Cas9)를 포함하는 리보핵산단백질 복합체를 포함한다. 일부 구현예에서, 약제학적 조성물은 gRNA, 핵산 프로그래밍 가능한 DNA 결합 단백질, 양이온성 지질, 및 약제학적으로 허용되는 부형제를 포함한다. 약제학적 조성물은 임의로 하나 이상의 추가의 치료학적 활성 물질을 포함할 수 있다.

[2660] 일부 구현예에서, 본원에 제공된 조성물은 대상체에, 예를 들어, 인간 대상체에 투여되어 대상체 내 표적화된 게놈 변형을 수행한다. 일부 구현예에서, 세포는 대상체로부터 수득되고 본원에 제공된 임의의 약제학적 조성물과 접촉시킨다. 일부 구현예에서, 대상체로부터 제거되고 약제학적 조성물과 생체외 접촉된 세포는 임의로 목적하는 게놈 변형이 세포에서 수행되거나 검출된 후 대상체에 재도입한다. 뉴클레아제를 포함하는 약제학적 조성물을 전달하는 방법은 공지되어 있고 예를 들어, 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 기재되어 있고, 상기 특허의 모든 개시내용은 이들의 전문이 본원에 참조로 포함된다. 비록 본원에 제공된 약제학적 조성물의 기재가 원칙적으로 인간에게 투여하기에 적합한 약제학적 조성물에 관한 것이지만, 이러한 조성물이 일반적으로 예를 들어, 수의학적 용도를 위해 모든 종류의 동물 또는 유기체에 투여하기에 적합하다는 것은 당업자라면 이해할 것이다.

[2661] 각종 동물에게 투여하기에 적합한 조성물을 제공하기 위해서 사람에게 투여하는데 적합한 약제학적 조성물의 변형은 잘 이해되고, 통상의 수의학 약리학자는 존재하는 경우 단지 통상적인 실험으로 이러한 변형을 디자인 및/또는 수행할 수 있다. 약제학적 조성물의 투여가 고려되는 대상체는 인간 및/또는 다른 영장류; 소, 돼지, 말, 양, 고양이, 개, 마우스 및/또는 래트와 같은 포유류, 가정용 동물, 애완동물 및 상업 관련 포유동물; 및/또는 닭, 오리, 거위 및/또는 칠면조와 같은 상업 관련 조류를 포함한 조류를 포함하지만, 이들에 한정되는 것은 아

니다.

[2662] 본원에 기재된 약제학적 조성물의 제형은 약리학 분야에 공지되거나 이후에 개발된 임의의 방법에 의해 제조될 수 있다. 일반적으로, 이러한 제조 방법은 활성 성분(들)을 부형제 및/또는 하나 이상의 기타 보조 성분과 연관시키고, 이어서 필요에 따라 및/또는 경우에 따라, 생성물을 목적하는 단일 또는 다중 복용량 단위로 성형 및/또는 패키징하는 단계를 포함한다. 약제학적 제형은 약제학적으로 허용되는 부형제를 추가로 포함할 수 있으며, 이것은 본원에서 사용되는 바와 같이 임의의 및 모든 용매, 분산 매질, 희석제 또는 다른 액체 비히클, 분산물 또는 현탁액 보조제, 표면 활성제, 등장화제, 증점제 또는 유화제, 보존제, 고체 결합제, 윤활제 등을 포함하며, 목적하는 특정 투여 형태에 적합화된다. 문헌 (Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; 이의 전문은 본원에 참조로 인용됨))은 약제학적 조성물을 제형화하는데 사용되는 다양한 부형제 및 이의 제조를 위한 공지된 기술을 개시한다. 또한, 뉴클레아제를 포함하는 약제학적 조성물을 제조하기 위한 추가의 적합한 방법, 시약, 부형제 및 용매에 대해 이의 전문이 본원에 참조로 포함된 PCT 출원 PCT/US2010/055131 (공개번호 WO2011053982 A8, 2010년 11월 2일 출원됨)을 참조한다.

[2663] 임의의 통상적인 부형제 매질이, 예를 들면, 임의의 바람직하지 못한 생물학적 효과를 생성하거나 약제학적 조성물의 임의의 다른 성분(들)과 유해한 방식으로 상호작용함으로써 물질 또는 이의 유도체와 상용성이 아닌 경우를 제외하고는 이의 용도는 본원의 개시내용의 범위 내 있는 것으로 고려된다.

[2664] 상기 기재된 바와 같은 조성물은 유효량으로 투여될 수 있다. 유효량은 투여 방식, 치료받는 특정 병태 및 목적하는 결과에 의존한다. 이것은 또한 병태 단계, 대상체의 연령 및 신체 조건, 존재하는 경우 동시 치료요법의 특성, 및 의학적 수행자에게 널리 공지된 유사 인자에 의존할 수 있다. 치료학적 적용을 위해, 이것은 의학적으로 목적하는 결과를 성취하기에 충분한 양이다.

[2665] ***RETT 치료 방법***

[2666] 또한 본원에 기재된 염기 편집기 시스템 (예를 들어, ABE8 염기 편집기 및 gRNA)을 암호화하는 폴리뉴클레오타이드를 포함하는 치료학적 유효량의 약제학적 조성물을 대상체 (예를 들어, 인간과 같은 포유동물)에 투여하는 단계를 포함하여, 레트 증후군 (RTT 또는 RETT) 및/또는 RETT를 유발하는 *Mecp2*의 유전적 돌연변이를 치료하는 방법이 또한 제공된다. 일부 구현예에서, 염기 편집기는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함하는 융합 단백질이다. *Mecp2* 유전자의 돌연변이를 포함하는 핵산 서열의 A·T에서 G·C로의 변경을 수행하도록 염기 편집기를 표적화하는 하나 이상의 가이드 폴리뉴클레오타이드 및 염기 편집기로 대상체의 세포를 형질도입한다.

[2667] 본원의 방법은 본원에 기재된 유효량의 조성물을 대상체 (상기 치료를 필요로 하는 것으로 동정된 대상체 또는 상기 질환 위험이 있는 것으로 의심되고 상기 치료를 필요로 하는 대상체를 포함하는)에게 투여하는 단계를 포함한다. 상기 치료를 필요로 하는 대상체의 동정은 대상체 또는 건강 관리 전문의의 판단하에 있을 수 있고 주관적 (예를 들어, 견해) 또는 객관적 (예를 들어, 시험 또는 진단 방법에 의해 측정될 수 있는)일 수 있다.

[2668] 치료학적 방법은 일반적으로 예를 들어, 염기 편집기 및 이를 필요로 하는 대상체 (예를 들어, 인간 환자)의 *Mecp2* 유전자를 표적화하는 gRNA를 암호화하는 벡터를 포함하는, 치료학적 유효량의 약제학적 조성물의 투여하는 단계를 포함한다. 상기 치료는 RTT를 앓고 있거나, 이를 갖거나, 걸리거나, 또는 이의 위험에 처한 대상체, 특히 인간 대상체에게 적합하게 투여될 것이다. 본원의 조성물은 또한 RTT가 관련될 수 있는 임의의 다른 장애의 치료에 사용될 수 있다.

[2669] 하나의 구현예에서, 치료 진행 모니터링 방법이 제공된다. 상기 방법은 RTT와 연관된 장애 또는 이의 증상을 앓거나 이에 민감할 수 있는 대상체에서 진단학적 마커 (Marker) (예를 들어, RTT와 연관된 SNP)의 수준 또는 진단적 측정 (예를 들어, 스크리닝, 검정)을 결정하는 단계를 포함하고, 여기서, 상기 대상체는 질환 또는 이의 증상을 치료하기에 충분한 치료학적 양의 조성물을 투여 받는다. 상기 방법에서 결정된 마커의 수준은 대상체의 질환의 상태를 확립하기 위해 건강한 정상 대조군 또는 다른 질환에 걸린 환자에서 마커의 공지된 수준과 비교될 수 있다. 바람직한 구현예에서, 대상체 내 마커의 제2 수준은 제1 수준의 결정 이후의 시점에 결정되고, 2개의 수준은 질환의 과정 또는 치료요법의 효능을 모니터링하기 위해 비교된다. 특정 바람직한 구현예에서, 대상체에서 마커의 치료 전 수준은 본원 개시내용의 방법에 따른 치료를 개시하기 전에 결정되고; 이어서, 마커의 상기 치료 전 수준은 치료의 효능을 결정하기 위해 치료가 시작된 후 대상체의 마커 수준과 비교할 수 있다.

[2670] 일부 구현예에서, 세포는 대상체로부터 수득되고 본원에 제공된 바와 같은 약제학적 조성물과 접촉시킨다. 일

부 구현예에서, 대상체로부터 제거되고 약제학적 조성물과 생체외 접촉된 세포는, 임의로 목적하는 게놈 변형이 세포에서 수행되거나 검출된 후 대상체에 재도입한다. 뉴클레아제를 포함하는 약제학적 조성물을 전달하는 방법은 예를 들어, 미국 특허 번호 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; 및 7,163,824에 기재되어 있고, 상기 모든 특허의 개시내용은 이들의 전문이 본원에 참조로 포함된다. 비록 본원에 제공된 약제학적 조성물의 기재가 원칙적으로 인간에게 투여하기에 적합한 약제학적 조성물에 관한 것이지만, 이러한 조성물이 일반적으로 모든 종류의 동물 또는 유기체에, 예를 들어, 수의학적 용도로 투여하기에 적합하다는 것은 당업자라면 이해할 것이다.

[2671] ***키트***

[2672] 본원 개시내용의 다양한 양상은 염기 편집기 시스템을 포함하는 키트를 제공한다. 하나의 구현예에서, 키트는 핵염기 편집기 융합 단백질을 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 작제물을 포함한다. 하나의 구현예에서, 융합 단백질은 아데닌 데아미나제 및 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)을 포함한다. 일부 구현예에서, 키트는 관심 대상의 핵산 분자, 예를 들어, *Mecp2* RTT-연관된 돌연변이를 표적화할 수 있는 적어도 하나의 가이드 RNA를 포함한다. 일부 구현예에서, 키트는 적어도 하나의 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 작제물을 포함한다.

[2673] 일부 구현예에서, 키트는 하나 이상의 *Mecp2* RTT-연관된 돌연변이를 편집하기 위해 키트를 사용하기 위한 지침서를 제공한다. 지침서는 일반적으로 핵산 분자를 편집하기 위한 키트의 용도에 대한 정보를 포함할 것이다. 다른 구현예에서, 지침서는 하기 중 적어도 하나를 포함한다: 주의사항; 경고; 임상 연구; 및/또는 참조. 지침서는 컨테이너 (존재하는 경우) 상에 직접 인쇄될 수 있거나, 표지로서 컨테이너에 적용되거나 별도의 시트, 팜플렛, 카드 또는 폴더로서 컨테이너 내 또는 이와 함께 공급될 수 있다. 추가의 구현예에서, 키트는 적합한 작동 파라미터를 위해 표지 또는 별도의 삽입물 (패키지 삽입물)의 형태로 지침서를 포함할 수 있다. 여전히 또 다른 구현예에서, 키트는 검출, 계산 또는 정규화를 위해 표준물(들)로서 사용될 적당한 양성 및 음성 대조군 또는 대조군 샘플을 갖는 하나 이상의 컨테이너를 포함할 수 있다. 키트는 (멸균) 포스페이트-완충 식염수, 링거 용액, 또는 덱스트로스 용액과 같은 약제학적으로 허용되는 완충액을 포함하는 제2 컨테이너를 추가로 포함할 수 있다. 키트는 다른 완충제, 희석제, 필터, 바늘, 주사기 및 사용 설명서가 있는 패키지 삽입물을 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[2674] 특정 구현예에서, 키트는 레트 증후군을 가진 대상체의 치료에 유용하다.

[2675] 본원 개시내용의 수행은, 달리 지적되지 않는 한, 통상적인 분자 생물학 기술 (재조합 기술 포함), 미생물학, 세포 생물학, 생화학 및 면역학 기술을 사용하고, 이들 기술들은 당업자의 범위 내에 있다. 상기 기술은 문헌, 예를 들어, 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 자세히 설명되어 있다. 이들 기술은 본원 개시내용의 폴리뉴클레오타이드 및 폴리펩타이드의 생성에 적용될 수 있고, 예를 들어, 본원 개시내용을 제조하고 수행하는데 고려될 수 있다. 특정 구현예에 대해 특히 유용한 기술은 하기의 섹션에서 논의될 것이다.

[2676] **실시예**

[2677] 하기 실시예들은 본원 개시내용의 검정, 스크리닝 및 치료학적 방법을 만들고 사용하는 방법에 대한 완전한 개시내용 및 설명을 당업자에게 제공하기 위해 제시하는 것으로, 본원 개시내용 및 이의 구현예에서 설명된 것의 범위를 제한하도록 의도되지 않는다.

[2678] **실시예 1 세포의 Mecp2 RTT-연관된 돌연변이 교정을 위한 A·T에서 G·C로의 DNA 염기 편집**

[2679] 8개의 가장 만연한 RTT-유발 Mecp2 돌연변이 중 6개는 검증된 프로토스페이서 인접 모티프 (PAM) 서열 선호도를 갖는 Cas9 모이어티를 사용하는 A·T에서 G·C로의 DNA 염기 편집기 (ABE)를 사용하여 야생형 서열로의 복귀를 위해 표적화되었다 (**도 2**). 어떤 가이드 RNA (gRNA) 및 ABE-Cas9 플랫폼이 표적화된 Mecp2 돌연변이를 가장 효율적이고 정밀하게 교정할 수 있는지 결정하기 위해, R255X를 포함하는 RTT 표적 가능한 돌연변이를 보유하는 *Mecp2* 대립유전자 (표 15)를 렌티바이러스 형질도입에 의해 HEK293T 세포에 게놈 통합하였다. 주어진 돌연변이에 대한 gRNA 및 ABE-Cas9 편집기의 편집 효율을 측정하였다. ABE-Cas9 편집기 및 gRNA를 암호화하는 DNA의 형

질감염 5일 후, 세포를 용해시키고 miSeq 분석에 의해 원하는 부위에서 염기 편집에 대해 분석하였다.

[2680] [표 15] 6개 RETT 돌연변이

원인 돌연변이	표적 서열	PAM		Cas9 PAM-활성화 돌연변이	편집 위치
c.316C>T	CCATGTCCAGCCTTCAGGCA	GGG	SpCas9		3
(2.77%)	TCCATGTCCAGCCTTCAGGC	AGG	SpCas9		4
R106W	TTCCATGTCCAGCCTTCAGG	CAG	SpCas9	xCas9	5
	GCTTCCATGTCCAGCCTTCA	GGC	SpCas9	MQKSER/ICKSER/VRKSER; Nureki	7
	CTTCCATGTCCAGCCTTCAGG	CAGGGT	SaCas9		6
	AGCTTCCATGTCCAGCCTTCAGG	CAGGGT	SaCas9		8
	AGCTTCCATGTCCAGCCTTC	CTTA	Cpf1		8
	CTTAAGCTTCCATGTCCAGC	TTTG	Cpf1		12
c.397C>T	AGCAAAAGGCTTTTCCCTGG	GGA	SpCas9	VRQR	4
(4.56%)	GAGCAAAAGGCTTTTCCCTG	GGG	SpCas9		5
R133C	AGAGCAAAAGGCTTTTCCCT	GGG	SpCas9		6
	TAGAGCAAAAGGCTTTTCCC	TGG	SpCas9		7
	TAGAGCAAAAGGCTTTTCCCT	GGGGAT	SaCas9		7
	TTTAGAGCAAAAGGCTTTTCCCT	GGGGAT	SaCas9		9
c.473C>T	CCATGAAGTCAAAATCATTA	GGG	SpCas9		3
(8.81%)	ACCATGAAGTCAAAATCATT	AGG	SpCas9		4
T158M	TACCATGAAGTCAAAATCAT	TAG	SpCas9	xCas9; ScCas9	5
	TTACCATGAAGTCAAAATCAT	TAGGGT	SaCas9		6
	AGTTACCATGAAGTCAAAATCAT	TAGGGT	SaCas9		8
c.763C>T (6.68%) R255X	CTTTTCACTTCCTGCCGGGG	CGT	SpCas9	LRSVQL/LRKIQK/LRSVQK; Nureki	7
c.808C>T	TCAGCCCCGTTTCTTGGGAA	TGG	SpCas9		3
(5.8%)	GCTTTCAGCCCCGTTTCTTG	GGA	SpCas9	VRQR	7
R270X	CGGCTTTCAGCCCCGTTTCTT	GGGAAT	SaCas9		9
	CCCGGCTTTCAGCCCCGTTTCTT	GGGAAT	SaCas9		11
c.916C>T (5.17%)	GCACTTCTTGATGGGGAGTA	CGG	SpCas9		3
R306C	TCTTGCACTTCTTGATGGGG	AGT	SpCas9	LRSVQL/LRKIQK/LRSVQK	7
	CTTGCACTTCTTGATGGGGAG	TACGGT	SaCas9	KKH	6
	GTCTTGCACTTCTTGATGGGGAG	TACGGT	SaCas9	KKH	8

[2681]

[2682] 4개의 가장 심각한 돌연변이인 R255X, R106W, T158M, 및 R270X에 대한 특정 표적/스페이서 서열을 아래 표 16에 나타낸다:

[2683] [표 16] 레트 돌연변이 표적 및 PAM 서열

돌연변이	스페이서	PAM
R255X	CTTTTC**A**CTTCCTGCCGGGG	NGTT
R106W	AGCTTCC**A**TGTCCAGCCTTC	NGG
T158M	ACC**A**TG**AA**GTCAAAATCATT	NGG
R270X	GCTTTC**A**GCCCCGTTTCTTG	NGG

[2684]

[2685] 상기 표적/스페이서 서열은 표적 가능한 "a" 핵염기를 굵게 표시하고 밑줄 쳐서 나타낸다. R255X 돌연변이는 NGTT PAM 변이체를 사용하여 표적화될 수 있고 R106W, T158M 및 R270X 돌연변이는 NGG PAM을 사용하여 표적화될

수 있다.

[2686] **실시예 2 R106W 돌연변이 교정을 위한 염기 편집**

[2687] R106W는 레트 증후군에서 4개의 가장 심각한 돌연변이 중 하나이다. ABE는 표적화된 부위에서 A>G를 효율적으로 변환함으로써 R106W의 교정에 사용될 수 있다. 염기 편집은 실시예 1에서 논의된 바와 같이 RETT 돌연변이를 함유하는 MECP2의 단일 게놈 통합된 카피를 함유하는 MECP2-HEK293T 세포주에서 시험되었다. 5개 가이드 RNA (gRNA)는 ABE8 염기 편집기 변이체를 사용하여 RTT에서 R106W 돌연변이를 표적으로 하여 시험되었다 (**도 3**).

[2688] gRNA는 본원에서 제공되거나 당업자의 지식을 기반으로 결정되고 당업자에게 이해되는 바와 같이 질환-연관된 유전자에 대한 스캐폴드 서열 및 스페이서 서열 (표적 서열)을 포함한다 (예를 들어, 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), and 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) 참조).

[2689] R106W 돌연변이에 대한 5개 DNA 표적 서열은 아래에 제공된다:

DNA **표적 서열** 1: CCATGTCCAGCCTTCAGGCA**GGG**

DNA **표적 서열** 2: TCCATGTCCAGCCTTCAGGC**AGG**

DNA **표적 서열** 3: TTCCATGTCCAGCCTTCAGG**CAG**

DNA **표적 서열** 4: GCTTCCATGTCCAGCCTTCA**GGC**

DNA **표적 서열** 5: AGCTTCCATGTCCAGCCTTC**AGG**

[2690]

[2691] PAM 서열은 밑줄 쳐 있고 굵은 글꼴이다. 상기 표 15에 나타낸 바와 같이, 표적 서열 1, 2 및 5는 NGG PAM (즉, SpCas9)을 이용한다. 표적 서열 3은 xCas9를 필요로 한다 (본원에 참조로 포함된 문헌 (Liu *et al.* WO2017070633)에 기재된 바와 같음). 표적 서열 4는 Cas9 PAM 변이체를 필요로 한다 (본원에 참조로 포함된 문헌 (Joung *et al.* WO2019040650)에 기재된 바와 같음).

[2692] 상기 DNA 표적 서열 각각의 상보체에 하이브리드화하는 5개 가이드 RNA 서열은 다음과 같다:

gRNA1: 5′-gCCAUGUCCAGCCUUCAGGCA-3′

gRNA2: 5′-gUCCAUGUCCAGCCUUCAGGC-3′

gRNA3: 5′-gUUCCAUGUCCAGCCUUCAGG-3′

gRNA4: 5′-GCUUCCAUGUCCAGCCUUCA-3′

gRNA5: 5′-gAGCUUCCAUGUCCAGCCUUC-3′

[2693]

[2694] DNA 표적 서열과 gRNA 서열 둘 다에서 소문자 "g"는 폴리머라제가 전사를 개시해야 하는 서열의 불일치를 나타낸다. 일부 구현예에서, 가이드 폴리뉴클레오타이드는 특히 5' 말단에서 1, 2, 3, 4 등의 뉴클레오타이드에 의해 절단될 수 있다.

[2695] 각각의 gRNA에 대해, 스캐폴드 서열은 다음과 같다:

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC

[2696] CGAGTCGGTGCTTTTTTT. 당업자에 의해 이해되는 바와 같이, 가이드 폴리뉴클레오티드 서열 (gRNA)에서 우라실 (U)은 서열에서 티민 (T)을 대체한다.

[2697] 사용된 ABE 염기 편집기는 ABE8 이종이량체 변이체를 포함한다: ABE8.14, ABE8.15, ABE8.16, ABE8.17, ABE8.18, ABE8.19, ABE8.20, ABE8.21, ABE8.22, ABE8.23, ABE8.24, ABE8.25, 및 ABE8.26 (표 9 참조). 양성 대조군 염기 편집기 ABE7.10 및 음성 대조군 또한 비교를 위해 사용되었다.

[2698] 플라스미드 형질감염을 위해, HEK293T 세포를 형질감염시키고 Opti-MEM 배지 및 Lipofectamine 2000을 사용하여 250 ng의 gRNA 및 750 ng의 ABE8 변이체 염기 편집기 발현 플라스미드로 플레이팅하였다. NGG PAM 서열을 갖는 ABE8 염기 편집기 변이체가 사용되었다. 세포를 용해하고 형질감염 5일 후 서열분석을 위해 준비하였다 (형질감염 후 3일 째에 배지를 교체함).

[2699] 5개 gRNA 서열 중에서, gRNA1, gRNA2 및 gRNA5가 최상의 A에서 G로의 편집을 제공하였다 (**도 3**). ABE8 염기 편집기 변이체는 ABE7.10 대조군과 비교하여 염기 편집 활성이 증가하였다. **도 3**에 나타낸 바와 같이, gRNA2를 사용하여 달성된 A에서 G로의 유전자 편집의 양은 약 35%였다.

[2700] **실시예 3 R133C 돌연변이의 교정을 위한 염기 편집**

[2701] ABE는 표적 부위에서 A>G를 효율적으로 전환하여 R133C의 교정에 사용될 수 있다. (도 1 및 2). 염기 편집은 실시예 1에서 논의된 바와 같이 RETT 돌연변이를 함유하는 *MECP2*의 단일 게놈 통합된 카피를 함유하는 MECP2-HEK293T 세포주에서 시험된다. 4개 가이드 RNA (gRNA)는 ABE8 염기 편집기 변이체를 사용하여 RETT에서 R133C 돌연변이를 표적으로 하여 시험된다.

[2702] gRNA는 본원에서 제공되거나 당업자의 지식을 기반으로 결정되고 당업자에게 이해되는 바와 같이 질환-연관된 유전자에 대한 스캐폴드 서열 및 스페이서 서열 (표적 서열)을 포함한다 (예를 들어, 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), and 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) 참조).

[2703] R133C 돌연변이에 대한 4개 DNA 표적 서열은 아래에 제공된다:

[2704] **DNA 표적 서열** 1: AGAGCAAAAGGCTTTTCCCT

DNA 표적 서열 2: TAGAGCAAAAGGCTTTTCCC

DNA 표적 서열 3: AGAGCAAAAGGCTTTTCCCT

DNA 표적 서열 4: TTTAGAGCAAAAGGCTTTTCCCT

[2705] 표적 서열은 상기 표 15에 나타낸 바와 같이 PAM 서열을 이용한다. 표적 서열 1과 2는 SpCas9를 사용한다. 표적 서열 3과 4는 SaCas9를 사용한다.

[2706] 상기 DNA 표적 서열 각각의 상보체에 하이브리드화하는 4개 가이드 RNA 서열 (gRNA)은 다음과 같다:

[2707] gRNA1: 5′-gAGAGCAAAAGGCUUUUCCCU-3′

gRNA2: 5′-gUAGAGCAAAAGGCUUUUCCC-3′

gRNA3: 5′-gUAGAGCAAAAGGCUUUUCCCU-3′

gRNA4: 5′-gUUUAGAGCAAAAGGCUUUUCCCU-3′

[2708] DNA 표적 서열과 gRNA 서열 둘 다에서 소문자 "g"는 폴리머라제가 전사를 개시해야 하는 서열의 불일치를 나타낸다. 일부 구현예에서, 가이드 폴리뉴클레오타이드는 특히 5' 말단에서 1, 2, 3, 4 등의 뉴클레오타이드에 의해 절단될 수 있다.

[2709] 각각의 gRNA에 대해, 스캐폴드 서열은 다음과 같다:

[2710] GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC CGAGTCGGTGCTTTTTTT. 당업자에 의해 이해되는 바와 같이, 가이드 폴리뉴클레오티드 서열 (gRNA)에서 우라실 (U)은 서열에서 티민 (T)을 대체한다.

[2711] 사용을 위한 ABE 염기 편집기는 ABE8 이종이량체 변이체를 포함한다: ABE8.14, ABE8.15, ABE8.16, ABE8.17, ABE8.18, ABE8.19, ABE8.20, ABE8.21, ABE8.22, ABE8.23, ABE8.24, ABE8.25, 및 ABE8.26 (표 9 참조). 양성

대조군 염기 편집기 ABE7.10 및 음성 대조군 또한 비교를 위해 사용된다.

[2712] 플라스미드 형질감염을 위해, HEK293T 세포를 형질감염시키고 Opti-MEM 배지 및 Lipofectamine 2000을 사용하여 250 ng의 gRNA 및 750 ng의 ABE8 변이체 염기 편집기 발현 플라스미드로 플레이팅하였다. 세포를 용해하고 형질감염 5일 후 서열분석을 위해 준비하였다 (형질감염 후 3일 째에 배지를 교체함).

[2713] **실시예 4 R306C 돌연변이의 교정을 위한 염기 편집**

[2714] ABE는 표적 부위에서 A>G를 효율적으로 전환하여 R306C의 교정에 사용할 수 있다 (도 1 및 2). 염기 편집은 실시예 1에서 논의된 바와 같이 RETT 돌연변이를 함유하는 *MECP2*의 단일 게놈 통합된 카피를 함유하는 MECP2-HEK293T 세포주에서 시험된다. 3개 가이드 RNA (gRNA)는 ABE8 염기 편집기 변이체를 사용하여 RETT에서 R306C 돌연변이를 표적으로 하여 시험된다.

[2715] gRNA는 본원에서 제공되거나 당업자의 지식을 기반으로 결정되고 당업자에게 이해되는 바와 같이 질환-연관된 유전자에 대한 스캐폴드 서열 및 스페이서 서열 (표적 서열)을 포함한다 (예를 들어, 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), and 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) 참조).

[2716] R306C 돌연변이에 대한 3개 DNA 표적 서열은 아래에 제공된다:

DNA 표적 서열 1: TCTTGCACTTCTTGATGGGG

DNA 표적 서열 2: CTTGCACTTCTTGATGGGGAG

[2717] **DNA 표적 서열** 3: GTCTTGCACTTCTTGATGGGGAG

[2718] 표적 서열은 상기 표 15에 나타낸 바와 같이 PAM 서열을 이용한다. 표적 서열 1은 SpCas9를 사용한다. 표적 서열 2와 3은 SaCas9를 사용한다.

[2719] 상기 DNA 표적 서열 각각의 상보체에 하이브리드화하는 3개 가이드 RNA 서열 (gRNA)은 다음과 같다:

gRNA1: 5′-gUCUUGCACUUCUUGAUGGGG-3′

gRNA2: 5′-gCUUGCACUUCUUGAUGGGGAG-3′

[2720] gRNA3: 5′-gGUCUUGCACUUCUUGAUGGGGAG-3′

[2721] DNA 표적 서열과 gRNA 서열 둘 다에서 소문자 "g"는 폴리머라제가 전사를 개시해야 하는 서열의 불일치를 나타낸다. 일부 구현예에서, 가이드 폴리뉴클레오타이드는 특히 5' 말단에서 1, 2, 3, 4 등의 뉴클레오타이드에 의해 절단될 수 있다.

[2722] 각각의 gRNA에 대해, 스캐폴드 서열은 다음과 같다:

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCAC

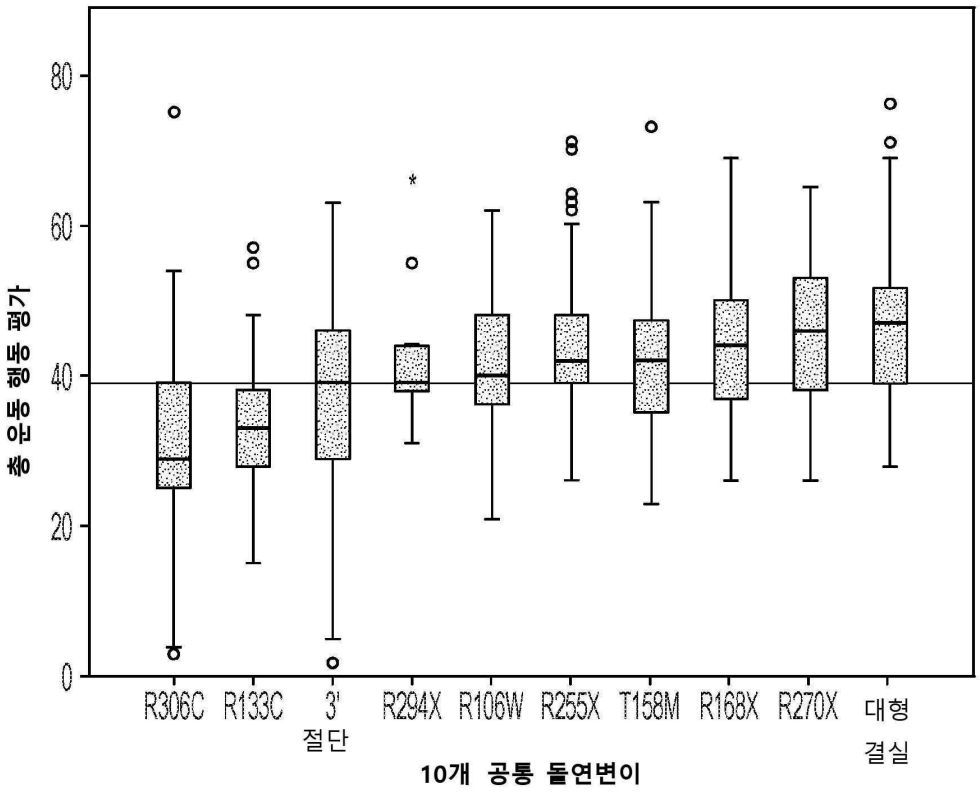
[2723] CGAGTCGGTGCTTTTTTT. 당업자에 의해 이해되는 바와 같이, 가이드 폴리뉴클레오티드 서열 (gRNA)에서 우라실 (U)은 서열에서 티민 (T)을 대체한다.

[2724] 사용을 위한 ABE 염기 편집기는 ABE8 이종이량체 변이체를 포함한다: ABE8.14, ABE8.15, ABE8.16, ABE8.17, ABE8.18, ABE8.19, ABE8.20, ABE8.21, ABE8.22, ABE8.23, ABE8.24, ABE8.25, 및 ABE8.26 (표 9 참조). 양성 대조군 염기 편집기 ABE7.10 및 음성 대조군 또한 비교를 위해 사용되었다.

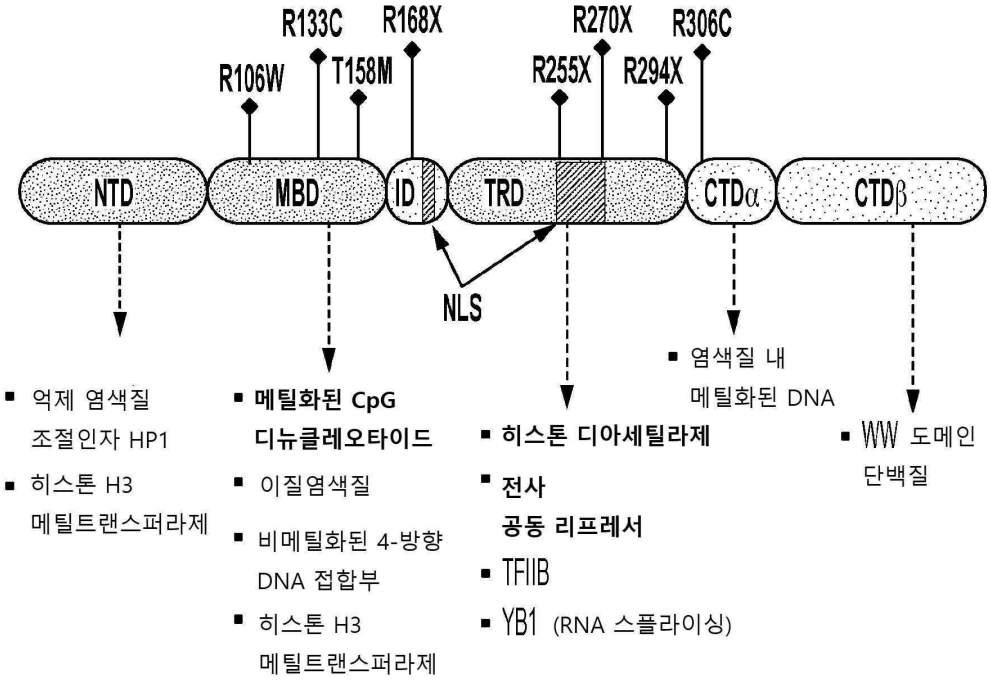
[2725] 플라스미드 형질감염을 위해, HEK293T 세포를 형질감염시키고 Opti-MEM 배지 및 Lipofectamine 2000을 사용하여 250 ng의 gRNA 및 750 ng의 ABE8 변이체 염기 편집기 발현 플라스미드로 플레이팅하였다. 세포를 용해하고 형질감염 5일 후 서열분석을 위해 준비하였다 (형질감염 후 3일 째에 배지를 교체함).

도면

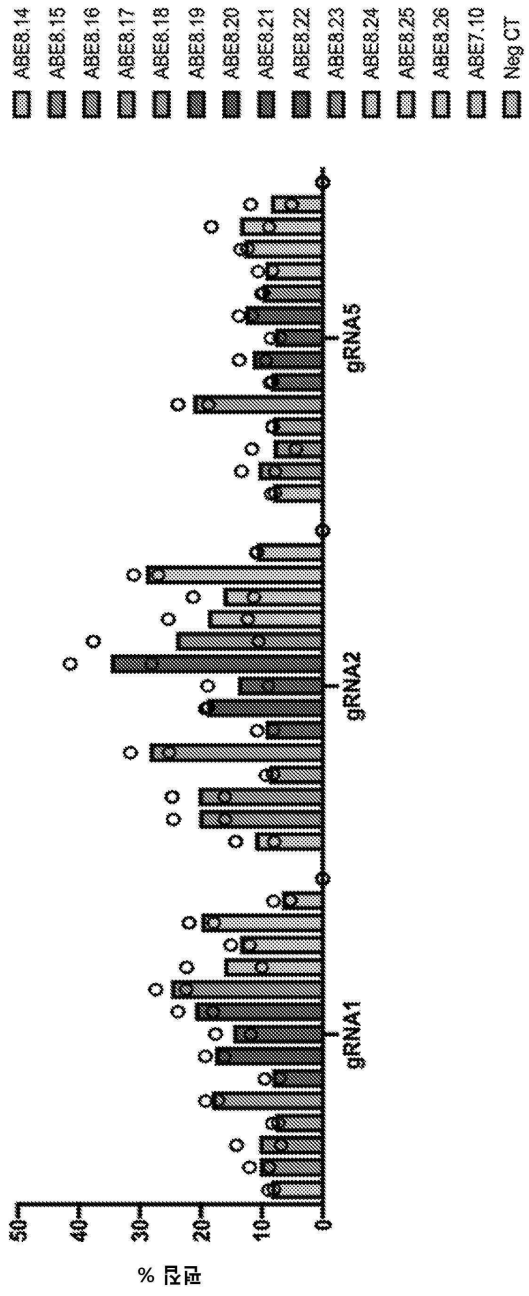
도면1


충동 행동 평가
0
20
40
60
80
R306C
R133C
3' 절단
R294X
R106W
R255X
T158M
R168X
R270X
대형 결실
10개 공통 돌연변이

도면2


R133C
R168X
R270X
R306C
R106W
T158M
R255X
R294X
NTD
MBD
ID
TRD
CTDα
CTDβ
NLS
▪ 억제 염색질 조절인자 HP1
▪ 히스톤 H3 메틸트랜스퍼라제
▪ 메틸화된 CpG 디뉴클레오타이드
▪ 이질염색질
▪ 비메틸화된 4-방향 DNA 접합부
▪ 히스톤 H3 메틸트랜스퍼라제
▪ 히스톤 디아세틸라제
▪ 전사 공동 리프레서
▪ TFIIB
▪ YB1 (RNA 스플라이싱)
▪ 염색질 내 메틸화된 DNA
▪ WW 도메인 단백질

도면3



서 열 목 록

SEQUENCE LISTING

<110> BEAM THERAPEUTICS INC.

<120> METHODS OF EDITING A SINGLE NUCLEOTIDE POLYMORPHISM USING PROGRAMMABLE BASE EDITOR SYSTEMS

<130> 180802.042401/PCT

<140> PCT/US2020/033807

<141> 2020-05-20

<150> 62/850,919

<151> 2019-05-21

<160> 256

<170> PatentIn version 3.5

<210> 1

<211> 167

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 1

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> 2

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 2

agagcaaaag gcuuuucccu 20

<210> 3

<211> 20

<

212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 3

uagagcaaaa ggcuuuuccc 20

<210> 4

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 4

uagagcaaaa ggcuuuuccc u 21

<210> 5

<211> 23

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 5

uuuagagcaa aaggcuuuuc ccu 23

<210> 6

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 6

ucuugcacuu cuugaugggg 20

<210> 7

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 7

cuugcacuuc uugaugggga g 21

<210> 8

<211> 23

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 8

gucuugcacu ucuugauggg gag 23

<210> 9

<211> 167

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 9

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> 10
<211> 83
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 10
gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60
ggcaccgagt cggtgctttt ttt 83
<210> 11

<211> 167

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

TadA sequence

<400> 11

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> 12

<211> 160

<212> PRT

<213> Staphylococcus aureus

<400> 12

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> 13

<211> 161

<212> PRT

<213> Bacillus subtilis

<400> 13

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> 14
<211> 183
<212> PRT
<213> Salmonella typhimurium
<400> 14
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> 15

<211> 164

<212> PRT

<213> Shewanella putrefaciens

<400> 15

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> 16

<211> 173

<212> PRT

<213> Haemophilus influenzae

<400> 16

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> 17
<211> 161
<212> PRT
<213> Caulobacter crescentus
<400> 17

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> 18

<211> 179

<212> PRT

<213> Geobacter sulfurreducens

<400> 18

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> 19

<211> 486

<212> PRT

<213> Homo sapiens

<400> 19

Met Val Ala Gly Met Leu Gly Leu Arg Glu Glu Lys Ser Glu Asp Gln
1 5 10 15
Asp Leu Gln Gly Leu Lys Asp Lys Pro Leu Lys Phe Lys Lys Val Lys
20 25 30
Lys Asp Lys Lys Glu Glu Lys Glu Gly Lys His Glu Pro Val Gln Pro
35 40 45
Ser Ala His His Ser Ala Glu Pro Ala Glu Ala Gly Lys Ala Glu Thr

50 55 60
Ser Glu Gly Ser Gly Ser Ala Pro Ala Val Pro Glu Ala Ser Ala Ser
65 70 75 80
Pro Lys Gln Arg Arg Ser Ile Ile Arg Asp Arg Gly Pro Met Tyr Asp
85 90 95
Asp Pro Thr Leu Pro Glu Gly Trp Thr Arg Lys Leu Lys Gln Arg Lys
100 105 110
Ser Gly Arg Ser Ala Gly Lys Tyr Asp Val Tyr Leu Ile Asn Pro Gln

115 120 125
Gly Lys Ala Phe Arg Ser Lys Val Glu Leu Ile Ala Tyr Phe Glu Lys
130 135 140
Val Gly Asp Thr Ser Leu Asp Pro Asn Asp Phe Asp Phe Thr Val Thr
145 150 155 160
Gly Arg Gly Ser Pro Ser Arg Arg Glu Gln Lys Pro Pro Lys Lys Pro
165 170 175
Lys Ser Pro Lys Ala Pro Gly Thr Gly Arg Gly Arg Gly Arg Pro Lys

180 185 190
Gly Ser Gly Thr Thr Arg Pro Lys Ala Ala Thr Ser Glu Gly Val Gln
195 200 205
Val Lys Arg Val Leu Glu Lys Ser Pro Gly Lys Leu Leu Val Lys Met
210 215 220
Pro Phe Gln Thr Ser Pro Gly Gly Lys Ala Glu Gly Gly Gly Ala Thr
225 230 235 240

Thr Ser Thr Gln Val Met Val Ile Lys Arg Pro Gly Arg Lys Arg Lys

245 250 255

Ala Glu Ala Asp Pro Gln Ala Ile Pro Lys Lys Arg Gly Arg Lys Pro

260 265 270

Gly Ser Val Val Ala Ala Ala Ala Ala Glu Ala Lys Lys Lys Ala Val

275 280 285

Lys Glu Ser Ser Ile Arg Ser Val Gln Glu Thr Val Leu Pro Ile Lys

290 295 300

Lys Arg Lys Thr Arg Glu Thr Val Ser Ile Glu Val Lys Glu Val Val

305 310 315 320

Lys Pro Leu Leu Val Ser Thr Leu Gly Glu Lys Ser Gly Lys Gly Leu

325 330 335

Lys Thr Cys Lys Ser Pro Gly Arg Lys Ser Lys Glu Ser Ser Pro Lys

340 345 350

Gly Arg Ser Ser Ser Ala Ser Ser Pro Pro Lys Lys Glu His His His

355 360 365

His His His His Ser Glu Ser Pro Lys Ala Pro Val Pro Leu Leu Pro

370 375 380

Pro Leu Pro Pro Pro Pro Pro Glu Pro Glu Ser Ser Glu Asp Pro Thr

385 390 395 400

Ser Pro Pro Glu Pro Gln Asp Leu Ser Ser Ser Val Cys Lys Glu Glu

405 410 415

Lys Met Pro Arg Gly Gly Ser Leu Glu Ser Asp Gly Cys Pro Lys Glu

420 425 430

Pro Ala Lys Thr Gln Pro Ala Val Ala Thr Ala Ala Thr Ala Ala Glu

435 440 445

Lys Tyr Lys His Arg Gly Glu Gly Glu Arg Lys Asp Ile Val Ser Ser

450 455 460

Ser Met Pro Arg Pro Asn Arg Glu Glu Pro Val Asp Ser Arg Thr Pro

465 470 475 480

Val Thr Glu Arg Val Ser

485

<210> 20

<211> 10241

<212> DNA

<213> Homo sapiens

<400> 20

ccggcgtcgg cggcgcgcgc gctccctcct ctcggagaga gggctgtggt aaaagccgtc 60

cggaaaatgg ccgccgccgc cgccgccgcg ccgagcggag gaggaggagg aggcgaggag 120

gagagactgc tccataaaaa tacagactca ccagttcctg ctttgatgtg acatgtgact 180

ccccagaata caccttgctt ctgtagacca gctccaacag gattccatgg tagctgggat 240

gttagggctc agggaagaaa agtcagaaga ccaggacctc cagggcctca aggacaaacc 300

cctcaagttt aaaaaggtga agaaagataa gaaagaagag aaagagggca agcatgagcc 360

cgtgcagcca tcagcccacc actctgctga gcccgcagag gcaggcaaag cagagacatc 420

agaagggtca ggctccgccc cggctgtgcc ggaagcttct gcctccccca aacagcggcg 480

ctccatcatc cgtgaccggg gacccatgta tgatgacccc accctgcctg aaggctggac 540

acggaagctt aagcaaagga aatctggccg ctctgctggg aagtatgatg tgtatttgat 600

caatccccag ggaaaagcct ttcgctctaa agtggagttg attgcgtact tcgaaaaggt 660

aggcgacaca tccctggacc ctaatgattt tgacttcacg gtaactggga gagggagccc 720

ctcccggcga gagcagaaac cacctaagaa gcccaaatct cccaaagctc caggaactgg 780

cagaggccgg ggacgcccca aagggagcgg caccacgaga cccaaggcgg ccacgtcaga 840

gggtgtgcag gtgaaaaggg tcctggagaa aagtcctggg aagctccttg tcaagatgcc 900

ttttcaaact tcgccagggg gcaaggctga ggggggtggg gccaccacat ccacccaggt 960

catggtgatc aaacgccccg gcaggaagcg aaaagctgag gccgaccctc aggccattcc 1020

caagaaacgg ggccgaaagc cggggagtgt ggtggcagcc gctgccgccg aggccaaaaa 1080

gaaagccgtg aaggagtctt ctatccgatc tgtgcaggag accgtactcc ccatcaagaa 1140

gcgcaagacc cgggagacgg tcagcatcga ggtcaaggaa gtggtgaagc ccctgctggt 1200

gtccaccctc ggtgagaaga gcgggaaagg actgaagacc tgtaagagcc ctgggcggaa 1260

aagcaaggag agcagcccca aggggcgcag cagcagcgcc tcctcacccc ccaagaagga 1320

gcaccaccac catcaccacc actcagagtc cccaaaggcc cccgtgccac tgctcccacc 1380

cctgccccca cctccacctg agcccgagag ctccgaggac cccaccagcc cccctgagcc 1440

ccaggacttg agcagcagcg tctgcaaaga ggagaagatg cccagaggag gctcactgga 1500

gagcgacggc tgccccaagg agccagctaa gactcagccc gcggttgcca ccgccgccac 1560

ggccgcagaa aagtacaaac accgaggggga gggagagcgc aaagacattg tttcatcctc 1620

catgccaagg ccaaacagag aggagcctgt ggacagccgg acgcccgtga ccgagagagt 1680

tagctgactt tacacggagc ggattgcaaa gcaaaccaac aagaataaag gcagctgttg 1740

tctcttctcc ttatgggtag ggctctgaca aagcttcccg attaactgaa ataaaaaata 1800

ttttttttte tttcagtaaa cttagagttt cgtggcttca gggtgggagt agttggagca 1860

ttggggatgt ttttcttacc gacaagcaca gtcaggttga agacctaacc agggccagaa 1920

gtagctttgc acttttctaa actaggctcc ttcaacaagg cttgctgcag atactactga 1980

ccagacaagc tgttgaccag gcacctcccc tcccgcccaa acctttcccc catgtggtcg 2040

ttagagacag agcgacagag cagttgagag gacactcccg ttttcggtgc catcagtgcc 2100

ccgtctacag ctccccccagc tccccccacc tcccccactc ccaaccacgt tgggacaggg 2160

aggtgtgagg caggagagac agttggattc tttagagaag atggatatga ccagtggcta 2220

tggcctgtgc gatcccaccc gtggtggctc aagtctggcc ccacaccagc cccaatccaa 2280

aactggcaag gacgcttcac aggacaggaa agtggcacct gtctgctcca gctctggcat 2340

ggctaggagg ggggagtccc ttgaactact gggtgtagac tggcctgaac cacaggagag 2400

gatggcccag ggtgaggtgg catggtccat tctcaaggga cgtcctccaa cgggtggcgc 2460

tagaggccat ggaggcagta ggacaaggtg caggcaggct ggcctggggt caggccgggc 2520

agagcacagc ggggtgagag ggattcctaa tcactcagag cagtctgtga cttagtggac 2580

aggggagggg gcaaaggggg aggagaagaa aatgttcttc cagttacttt ccaattctcc 2640

tttagggaca gcttagaatt atttgcacta ttgagtcttc atgttcccac ttcaaaacaa 2700

acagatgctc tgagagcaaa ctggcttgaa ttggtgacat ttagtccctc aagccaccag 2760

atgtgacagt gttgagaact acctggattt gtatatatac ctgcgcttgt tttaaagtgg 2820

gctcagcaca tagggttccc acgaagctcc gaaactctaa gtgtttgctg caattttata 2880

aggacttcct gattggtttc tcttctcccc ttccatttct gccttttgtt catttcatcc 2940

tttcacttct ttcccttcct ccgtcctcct ccttcctagt tcatcccttc tcttccaggc 3000

agccgcggtg cccaaccaca cttgtcggct ccagtcccca gaactctgcc tgccctttgt 3060

cctcctgctg ccagtaccag ccccaccctg ttttgagccc tgaggaggcc ttgggctctg 3120

ctgagtccga cctggcctgt ctgtgaagag caagagagca gcaaggtctt gctctcctag 3180

gtagcccccct cttccctggt aagaaaaagc aaaaggcatt tcccaccctg aacaacgagc 3240

cttttcaccc ttctactcta gagaagtgga ctggaggagc tgggcccgat ttggtagttg 3300

aggaaagcac agaggcctcc tgtggcctgc cagtcatcga gtggcccaac aggggctcca 3360

tgccagccga ccttgacctc actcagaagt ccagagtcta gcgtagtgca gcagggcagt 3420

agcggtacca atgcagaact cccaagaccc gagctgggac cagtacctgg gtccccagcc 3480
cttcctctgc tccccctttt ccctcggagt tcttcttgaa tggcaatgtt ttgcttttgc 3540
tcgatgcaga caggggggcca gaacaccaca catttcactg tctgtctggt ccatagctgt 3600
ggtgtagggg cttagaggca tgggcttgct gtgggttttt aattgatcag ttttcatgtg 3660
ggatcccatc tttttaacct ctgttcagga agtccttatc tagctgcata tcttcatcat 3720
attggtatat ccttttctgt gtttacagag atgtctctta tatctaaatc tgtccaactg 3780
agaagtacct tatcaaagta gcaaatgaga cagcagtctt atgcttccag aaacacccac 3840

aggcatgtcc catgtgagct gctgccatga actgtcaagt gtgtgttgtc ttgtgtattt 3900
cagttattgt ccctggcttc cttactatgg tgtaatcatg aaggagtgaa acatcataga 3960
aactgtctag cacttccttg ccagtcttta gtgatcagga accatagttg acagttccaa 4020
tcagtagctt aagaaaaaac cgtgtttgtc tcttctggaa tggttagaag tgagggagtt 4080
tgccccgttc tgtttgtaga gtctcatagt tggactttct agcatatatg tgtccatttc 4140
cttatgctgt aaaagcaagt cctgcaacca aactcccatc agcccaatcc ctgatccctg 4200
atcccttcca cctgctctgc tgatgacccc cccagcttca cttctgactc ttccccagga 4260

agggaagggg ggtcagaaga gagggtgagt cctccagaac tcttcctcca aggacagaag 4320
gctcctgccc ccatagtggc ctcgaactcc tggcactacc aaaggacact tatccacgag 4380
agcgcagcat ccgaccaggt tgtcactgag aagatgttta ttttggtcag ttgggttttt 4440
atgtattata cttagtcaaa tgtaatgtgg cttctggaat cattgtccag agctgcttcc 4500
ccgtcacctg ggcgtcatct ggtcctggta agaggagtgc gtggcccacc aggccccccc 4560
gtcacccatg acagttcatt cagggccgat ggggcagtcg tggttgggaa cacagcattt 4620
caagcgtcac tttatttcat tcgggcccca cctgcagctc cctcaaagag gcagttgccc 4680

agcctctttc ccttccagtt tattccagag ctgccagtgg ggcctgaggc tccttagggt 4740
tttctctcta tttccccctt tcttcctcat tccctcgtct ttcccaaagg catcacgagt 4800
cagtcgcctt tcagcaggca gccttggcgg tttatcgccc tggcaggcag gggccctgca 4860
gctctcatgc tgcccctgcc ttggggtcag gttgacagga ggttggaggg aaagccttaa 4920
gctgcaggat tctcaccagc tgtgtccggc ccagttttgg ggtgtgacct caatttcaat 4980
tttgtctgta cttgaacatt atgaagatgg gggcctcttt cagtgaattt gtgaacagca 5040
gaattgaccg acagctttcc agtacccatg gggctaggtc attaaggcca catccacagt 5100

ctcccccacc cttgttccag ttgttagtta ctacctcctc tcctgacaat actgtatgtc 5160

gtcgagctcc ccccaggtct acccctcccg gccctgcctg ctggtgggct tgtcatagcc 5220

agtgggattg ccggtcttga cagctcagtg agctggagat acttggtcac agccaggcgc 5280

tagcacagct cccttctgtt gatgctgtat tcccatatca aaagacacag gggacaccca 5340

gaaacgccac atcccccaat ccatcagtgc caaactagcc aacggcccca gcttctcagc 5400

tcgctggatg gcggaagctg ctactcgtga gcgccagtgc gggtgcagac aatcttctgt 5460

tgggtggcat cattccaggc ccgaagcatg aacagtgcac ctgggacagg gagcagcccc 5520

aaattgtcac ctgcttctct gcccagcttt tcattgctgt gacagtgatg gcgaaagagg 5580

gtaataacca gacacaaact gccaagttgg gtggagaaag gagtttcttt agctgacaga 5640

atctctgaat tttaaatcac ttagtaagcg gctcaagccc aggagggagc agagggatac 5700

gagcggagtc ccctgcgcgg gaccatctgg aattggttta gcccaagtgg agcctgacag 5760

ccagaactct gtgtcccccg tctaaccaca gctccttttc cagagcattc cagtcaggct 5820

ctctgggctg actgggccag gggaggttac aggtaccagt tctttaagaa gatctttggg 5880

catatacatt tttagcctgt gtcattgccc caaatggatt cctgtttcaa gttcacacct 5940

gcagattcta ggacctgtgt cctagacttc agggagtcag ctgtttctag agttcctacc 6000

atggagtggg tctggaggac ctgcccggtg gggggcaga gccctgctcc ctccgggtct 6060

tcctactctt ctctctgctc tgacgggatt tgttgattct ctccattttg gtgtctttct 6120

cttttagata ttgtatcaat ctttagaaaa ggcatagtct acttgttata aatcgttagg 6180

atactgcctc ccccagggtc taaaattaca tattagaggg gaaaagctga acactgaagt 6240

cagttctcaa caatttagaa ggaaaaccta gaaaacattt ggcagaaaat tacatttcga 6300

tgtttttgaa tgaatacgag caagctttta caacagtgct gatctaaaaa tacttagcac 6360

ttggcctgag atgcctggtg agcattacag gcaaggggaa tctggaggta gccgacctga 6420

ggacatggct tctgaacctg tcttttggga gtggtatgga aggtggagcg ttcaccagtg 6480

acctggaagg cccagcacca ccctccttcc cactcttctc atcttgacag agcctgcccc 6540

agcgctgacg tgtcaggaaa acacccaggg aactaggaag gcacttctgc ctgaggggca 6600

gcctgccttg cccactcctg ctctgctcgc ctcggatcag ctgagccttc tgagctggcc 6660

tctcactgcc tccccaaggc cccctgcctg ccctgtcagg aggcagaagg aagcaggtgt 6720

gagggcagtg caaggaggga gcacaacccc cagctcccgc tccgggctcc gacttgtgca 6780

caggcagagc ccagaccctg gaggaaatcc tacctttgaa ttcaagaaca tttggggaat 6840

ttggaaatct ctttgccccc aaaccccccat tctgtcctac ctttaatcag gtcctgctca 6900

gcagtgagag cagatgaggt gaaaaggcca agaggtttgg ctcctgccca ctgatagccc 6960

ctctccccgc agtgtttgtg tgtcaagtgg caaagctgtt cttcctggtg accctgatta 7020

tatccagtaa cacatagact gtgcgcatag gcctgctttg tctcctctat cctgggcttt 7080

tgttttgctt tttagttttg cttttagttt ttctgtccct tttatttaac gcaccgacta 7140

gacacacaaa gcagttgaat ttttatatat atatctgtat attgcacaat tataaactca 7200

ttttgcttgt ggctccacac acacaaaaaa agacctgtta aaattatacc tgttgcttaa 7260

ttacaatatt tctgataacc atagcatagg acaagggaaa ataaaaaaag aaaaaaaaga 7320

aaaaaaaacg acaaatctgt ctgctggtca cttcttctgt ccaagcagat tcgtggtctt 7380

ttcctcgctt ctttcaaggg ctttcctgtg ccaggtgaag gaggctccag gcagcaccca 7440

ggttttgcac tcttgtttct cccgtgcttg tgaaagaggt cccaaggttc tgggtgcagg 7500

agcgctccct tgacctgctg aagtccggaa cgtagtcggc acagcctggt cgccttccac 7560

ctctgggagc tggagtccac tggggtggcc tgactccccc agtccccttc ccgtgacctg 7620

gtcagggtga gcccatgtgg agtcagcctc gcaggcctcc ctgccagtag ggtccgagtg 7680

tgtttcatcc ttcccactct gtcgagcctg ggggctggag cggagacggg aggcctggcc 7740

tgtctcggaa cctgtgagct gcaccaggta gaacgccagg gaccccagaa tcatgtgcgt 7800

cagtccaagg ggtcccctcc aggagtagtg aagactccag aaatgtccct ttcttctccc 7860

ccatcctacg agtaattgca tttgcttttg taattcttaa tgagcaatat ctgctagaga 7920

gtttagctgt aacagttctt tttgatcatc tttttttaat aattagaaac accaaaaaaa 7980

tccagaaact tgttcttcca aagcagagag cattataatc accagggcca aaagcttccc 8040

tccctgctgt cattgcttct tctgaggcct gaatccaaaa gaaaaacagc cataggccct 8100

ttcagtggcc gggctacccg tgagcccttc ggaggaccag ggctggggca gcctctgggc 8160

ccacatccgg ggccagctcc ggcgtgtgtt cagtgttagc agtgggtcat gatgctcttt 8220

cccacccagc ctgggatagg ggcagaggag gcgaggaggc cgttgccgct gatgtttggc 8280

cgtgaacagg tgggtgtctg cgtgcgtcca cgtgcgtgtt ttctgactga catgaaatcg 8340

acgcccgagt tagcctcacc cggtgacctc tagccctgcc cggatggagc ggggcccacc 8400

cggttcagtg tttctgggga gctggacagt ggagtgcaaa aggcttgcag aacttgaagc 8460

ctgctccttc ccttgctacc acggcctcct ttccgtttga tttgtcactg cttcaatcaa 8520

taacagccgc tccagagtca gtagtcaatg aatatatgac caaatatcac caggactgtt 8580

actcaatgtg tgccgagccc ttgcccatgc tgggctcccg tgtatctgga cactgtaacg 8640

tgtgctgtgt ttgctcccct tccccttcct tctttgccct ttacttgtct ttctggggtt 8700

tttctgtttg ggtttggttt ggtttttatt tctccttttg tgttccaaac atgaggttct 8760

ctctactggt cctcttaact gtggtgttga ggcttatatt tgtgtaattt ttggtgggtg 8820

aaaggaattt tgctaagtaa atctcttctg tgtttgaact gaagtctgta ttgtaactat 8880

gtttaaagta attgttccag agacaaatat ttctagacac tttttcttta caaacaaaag 8940
cattcggagg gaggggggatg gtgactgaga tgagagggga gagctgaaca gatgacccct 9000
gcccagatca gccagaagcc acccaaagca gtggagccca ggagtcccac tccaagccag 9060
caagccgaat agctgatgtg ttgccacttt ccaagtcact gcaaaaccag gttttgttcc 9120
gcccagtgga ttcttgtttt gcttcccctc cccccgagat tattaccacc atcccgtgct 9180
tttaaggaaa ggcaagattg atgtttcctt gaggggagcc aggaggggat gtgtgtgtgc 9240
agagctgaag agctggggag aatggggctg ggcccaccca agcaggaggc tgggacgctc 9300

tgctgtgggc acaggtcagg ctaatgttgg cagatgcagc tcttcctgga caggccaggt 9360
ggtgggcatt ctctctccaa ggtgtgcccc gtgggcatta ctgtttaaga cacttccgtc 9420
acatcccacc ccatcctcca gggctcaaca ctgtgacatc tctattcccc accctcccct 9480
tcccagggca ataaaatgac catggagggg gcttgcactc tcttggctgt cacccgatcg 9540
ccagcaaaac ttagatgtga gaaaacccct tcccattcca tggcgaaaac atctccttag 9600
aaaagccatt accctcatta ggcatggttt tgggctccca aaacacctga cagcccctcc 9660
ctcctctgag aggcggagag tgctgactgt agtgaccatt gcatgccggg tgcagcatct 9720

ggaagagcta ggcagggtgt ctgccccctc ctgagttgaa gtcatgctcc cctgtgccag 9780
cccagaggcc gagagctatg gacagcattg ccagtaacac aggccaccct gtgcagaagg 9840
gagctggctc cagcctggaa acctgtctga ggttgggaga ggtgcacttg gggcacaggg 9900
agaggccggg acacacttag ctggagatgt ctctaaaagc cctgtatcgt attcaccttc 9960
agtttttgtg ttttgggaca attactttag aaaataagta ggtcgtttta aaaacaaaaa 10020
ttattgattg ctttttttgta gtgttcagaa aaaaggttct ttgtgtatag ccaaatgact 10080
gaaagcactg atatatttaa aaacaaaagg caatttatta aggaaatttg taccatttca 10140

gtaaacctgt ctgaatgtac ctgtatacgt ttcaaaaaca ccccccccccc actgaatccc 10200
tgtaacctat ttattatata aagagtttgc cttataaatt t 10241

<210> 21

<211> 1410

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 21

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> 22

<211

> 8811

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 22

atatgccaag tacgcccccc 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 aggccccccc 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 ccgccccccct 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> 23

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 23

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp 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> 24

<211> 4104

<212> DNA

<213> Streptococcus pyogenes

<400> 24

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> 25

<211> 4212

<212> DNA

<213> Streptococcus pyogenes

<400> 25

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> 26
<211> 1368
<212> PRT
<213> Streptococcus pyogenes
<400> 26

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> 27

<211> 4107

<212> DNA

<213> Streptococcus pyogenes

<400> 27

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> 28
<211> 1368
<212> PRT
<213> Streptococcus pyogenes
<400> 28
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> 29

<211> 1082

<212> PRT

<213> Neisseria meningitidis

<400> 29

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> 30

<211> 1082

<212> PRT

<213> Neisseria meningitidis

<400> 30

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> 31
<211> 1368
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 31
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> 32

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 32

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

```
Lys Val  Leu Ser Met Pro Gln  Val Asn Ile Val Lys  Lys Thr Glu
    1085                1090                1095
Val Gln  Thr Gly Gly Phe Ser  Lys Glu Ser Ile Leu  Pro Lys Arg

    1100                1105                1110
Asn Ser  Asp Lys Leu Ile Ala  Arg Lys Lys Asp Trp  Asp Pro Lys
    1115                1120                1125
Lys Tyr  Gly Gly Phe Asp Ser  Pro Thr Val Ala Tyr  Ser Val Leu
    1130                1135                1140
Val Val  Ala Lys Val Glu Lys  Gly Lys Ser Lys Lys  Leu Lys Ser
    1145                1150                1155
Val Lys  Glu Leu Leu Gly Ile  Thr Ile Met Glu Arg  Ser Ser Phe

    1160                1165                1170
Glu Lys  Asn Pro Ile Asp Phe  Leu Glu Ala Lys Gly  Tyr Lys Glu
    1175                1180                1185
Val Lys  Lys Asp Leu Ile Ile  Lys Leu Pro Lys Tyr  Ser Leu Phe
    1190                1195                1200
Glu Leu  Glu Asn Gly Arg Lys  Arg Met Leu Ala Ser  Ala Gly Glu
    1205                1210                1215
Leu Gln  Lys Gly Asn Glu Leu  Ala Leu Pro Ser Lys  Tyr Val Asn

    1220                1225                1230
Phe Leu  Tyr Leu Ala Ser His  Tyr Glu Lys Leu Lys  Gly Ser Pro
    1235                1240                1245
Glu Asp  Asn Glu Gln Lys Gln  Leu Phe Val Glu Gln  His Lys His
    1250                1255                1260
Tyr Leu  Asp Glu Ile Ile Glu  Gln Ile Ser Glu Phe  Ser Lys Arg
    1265                1270                1275
Val Ile  Leu Ala Asp Ala Asn  Leu Asp Lys Val Leu  Ser Ala Tyr

    1280                1285                1290
Asn Lys  His Arg Asp Lys Pro  Ile Arg Glu Gln Ala  Glu Asn Ile
    1295                1300                1305
Ile His  Leu Phe Thr Leu Thr  Asn Leu Gly Ala Pro  Ala Ala Phe
```

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 33

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 33

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 34

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 34

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 35

<211> 1129

<212> PRT

<213> Alicyclobacillus acidoterrestris

<400> 35

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> 36
<211> 345
<212> PRT
<213> Sulfolobus islandicus
<400> 36
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> 37

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 37

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> 38

<211> 986

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Deltaproteobacteria CasX sequence

<400> 38

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 Ala Tyr Asn Glu Val Ile Ala Arg Val Arg Met Trp Val
290 295 300

Asn Leu Asn Leu Trp Gln Lys Leu Lys Leu Ser Arg Asp Asp Ala Lys
305 310 315 320
Pro Leu Leu Arg Leu Lys Gly Phe Pro Ser Phe Pro Val Val Glu Arg
325 330 335
Arg Glu Asn Glu Val Asp Trp Trp Asn Thr Ile Asn Glu Val Lys Lys
340 345 350
Leu Ile Asp Ala Lys Arg Asp Met Gly Arg Val Phe Trp Ser Gly Val
355 360 365

Thr Ala Glu Lys Arg Asn Thr Ile Leu Glu Gly Tyr Asn Tyr Leu Pro
370 375 380
Asn Glu Asn Asp His Lys Lys Arg Glu Gly Ser Leu Glu Asn Pro Lys
385 390 395 400
Lys Pro Ala Lys Arg Gln Phe Gly Asp Leu Leu Leu Tyr Leu Glu Lys
405 410 415
Lys Tyr Ala Gly Asp Trp Gly Lys Val Phe Asp Glu Ala Trp Glu Arg
420 425 430

Ile Asp Lys Lys Ile Ala Gly Leu Thr Ser His Ile Glu Arg Glu Glu
435 440 445
Ala Arg Asn Ala Glu Asp Ala Gln Ser Lys Ala Val Leu Thr Asp Trp
450 455 460
Leu Arg Ala Lys Ala Ser Phe Val Leu Glu Arg Leu Lys Glu Met Asp
465 470 475 480
Glu Lys Glu Phe Tyr Ala Cys Glu Ile Gln Leu Gln Lys Trp Tyr Gly
485 490 495

Asp Leu Arg Gly Asn Pro Phe Ala Val Glu Ala Glu Asn Arg Val Val
500 505 510
Asp Ile Ser Gly Phe Ser Ile Gly Ser Asp Gly His Ser Ile Gln Tyr
515 520 525
Arg Asn Leu Leu Ala Trp Lys Tyr Leu Glu Asn Gly Lys Arg Glu Phe

530 535 540

Tyr Leu Leu Met Asn Tyr Gly Lys Lys Gly Arg Ile Arg Phe Thr Asp
545 550 555 560

Gly Thr Asp Ile Lys Lys Ser Gly Lys Trp Gln Gly Leu Leu Tyr Gly
565 570 575

Gly Gly Lys Ala Lys Val Ile Asp Leu Thr Phe Asp Pro Asp Asp Glu
580 585 590

Gln Leu Ile Ile Leu Pro Leu Ala Phe Gly Thr Arg Gln Gly Arg Glu
595 600 605

Phe Ile Trp Asn Asp Leu Leu Ser Leu Glu Thr Gly Leu Ile Lys Leu
610 615 620

Ala Asn Gly Arg Val Ile Glu Lys Thr Ile Tyr Asn Lys Lys Ile Gly
625 630 635 640

Arg Asp Glu Pro Ala Leu Phe Val Ala Leu Thr Phe Glu Arg Arg Glu
645 650 655

Val Val Asp Pro Ser Asn Ile Lys Pro Val Asn Leu Ile Gly Val Ala
660 665 670

Arg Gly Glu Asn Ile Pro Ala Val Ile Ala Leu Thr Asp Pro Glu Gly
675 680 685

Cys Pro Leu Pro Glu Phe Lys Asp Ser Ser Gly Gly Pro Thr Asp Ile
690 695 700

Leu Arg Ile Gly Glu Gly Tyr Lys Glu Lys Gln Arg Ala Ile Gln Ala
705 710 715 720

Ala Lys Glu Val Glu Gln Arg Arg Ala Gly Gly Tyr Ser Arg Lys Phe
725 730 735

Ala Ser Lys Ser Arg Asn Leu Ala Asp Asp Met Val Arg Asn Ser Ala
740 745 750

Arg Asp Leu Phe Tyr His Ala Val Thr His Asp Ala Val Leu Val Phe
755 760 765

Ala Asn Leu Ser Arg Gly Phe Gly Arg Gln Gly Lys Arg Thr Phe Met
770 775 780

Thr Glu Arg Gln Tyr Thr Lys Met Glu Asp Trp Leu Thr Ala Lys Leu
785 790 795 800
Ala Tyr Glu Gly Leu Thr Ser Lys Thr Tyr Leu Ser Lys Thr Leu Ala
805 810 815

Gln Tyr Thr Ser Lys Thr Cys Ser Asn Cys Gly Phe Thr Ile Thr Tyr
820 825 830
Ala Asp Met Asp Val Met Leu Val Arg Leu Lys Lys Thr Ser Asp Gly
835 840 845
Trp Ala Thr Thr Leu Asn Asn Lys Glu Leu Lys Ala Glu Tyr Gln Ile
850 855 860
Thr Tyr Tyr Asn Arg Tyr Lys Arg Gln Thr Val Glu Lys Glu Leu Ser
865 870 875 880

Ala Glu Leu Asp Arg Leu Ser Glu Glu Ser Gly Asn Asn Asp Ile Ser
885 890 895
Lys Trp Thr Lys Gly Arg Arg Asp Glu Ala Leu Phe Leu Leu Lys Lys
900 905 910
Arg Phe Ser His Arg Pro Val Gln Glu Gln Phe Val Cys Leu Asp Cys
915 920 925
Gly His Glu Val His Ala Ala Glu Gln Ala Ala Leu Asn Ile Ala Arg
930 935 940

Ser Trp Leu Phe Leu Asn Ser Asn Ser Thr Glu Phe Lys Ser Tyr Lys
945 950 955 960
Ser Gly Lys Gln Pro Phe Val Gly Ala Trp Gln Ala Phe Tyr Lys Arg
965 970 975
Arg Leu Lys Glu Val Trp Lys Pro Asn Ala
980 985

<210> 39
<211> 1210
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
uncultured Parcubacteria group bacterium sequence

<400> 39

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> 40

<211> 1140

<212> PRT

<213> Bacillus hisashii

<400> 40

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

1130 1135 1140
<210> 41
<211> 306
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<400> 41
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> 42

<211> 102

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 42

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> 43

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 43

atgatcaaga tagctacaag gaagtatctt ggcaaacaaa acgtttatga tattggagtc 60

gaaagagatc acaactttgc tctgaagaac ggattcatag cttctaat 108

<210> 44

<211> 36

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 44

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> 45

<211> 303

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 45

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> 46

<211> 101

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 46

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> 47

<211> 159

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 47

atgaagagga ctgccgatgg atcagagttt gaatctccca agaagaagag gaaagtaaag 60

ataatatctc gaaaaagtct tggtacccaa aatgtctatg atattggagt ggagaaagat 120

cacaacttcc ttctcaagaa cggtctcgta gccagcaac 159

<210> 48

<211> 53

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 48

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> 49

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 49

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
1 5 10 15

<210> 50

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 50

Ser Gly Gly Ser
1

<210> 51

<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> 51

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> 52

<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> 52

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> 53

<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> 53

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> 54
<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> 54
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
1 5 10 15
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
20 25 30

<210> 55
<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> 55

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> 56

<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> 56

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> 57

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<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

<220><221> MISC_FEATURE

<222> (1)..(60)

<223> This sequence may encompass 1-30 "Xaa Pro"

repeating units

<400> 57

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> 58

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 58

Pro Ala Pro Ala Pro

1 5

<210> 59

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 59

Pro Ala Pro Ala Pro Ala

1 5

<210> 60

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 60

Pro Ala Pro Ala Pro Ala Pro

1 5

<210> 61

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 61

Pro Ala Pro Ala Pro Ala Pro Ala

1 5

<210> 62

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 62

Pro Ala Pro Ala Pro Ala Pro Ala Pro
1 5

<210> 63

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 63

Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
1 5 10 15

<210> 64

<211> 21

<212>

PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 64

Pro Ala 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> 65

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 65

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> 66

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 66

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> 67

<211> 104

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 67

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> 68

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 68

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> 69

<211> 40

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 69

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> 70

<211> 64

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 70

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 Ser Gly Gly Ser
50 55 60

<210> 71

<211> 92

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 71

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> 72

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 72

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg

1 5 10 15

Lys Val

<210> 73

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 73

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 74

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 74

Lys Lys Thr Glu Leu Gln Thr Thr Asn Ala Glu Asn Lys Thr Lys Lys

1 5 10 15

Leu

<210> 75

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 75

Lys Arg Gly Ile Asn Asp Arg Asn Phe Trp Arg Gly Glu Asn Gly Arg
1 5 10 15
Lys Thr Arg

<210> 76

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 76

Arg Lys Ser Gly Lys Ile Ala Ala Ile Val Val Lys Arg Pro Arg Lys

1 5 10 15

<210> 77

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 77

Pro Lys Lys Lys Arg Lys Val
1 5

<210> 78

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 78

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> 79

<211> 84

<212> PRT

<213> Bacillus phage PBS2

<400> 79

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> 80

<211> 80

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 80

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

<210> 81

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 81

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> 82

<211> 987

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Deltaproteobacteria CasX sequence

<400> 82

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> 83
<211> 1300
<212> PRT
<213> Francisella novicida
<400> 83
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> 84
<211> 1300
<212> PRT
<213> Francisella novicida
<400> 84

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> 85

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 85

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> 86

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 86

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> 87
<211> 1300
<212> PRT
<213> Francisella novicida
<400> 87
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> 88

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 88

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> 89
<211> 1300
<212> PRT
<213> Francisella novicida

<400> 89
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> 90
<211> 1300
<212> PRT
<213> Francisella novicida
<400> 90

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> 91

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 91

gggaaataag agagaaaaga agagtaagaa gaaatataag agccacc 47

<210> 92

<211> 101

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 92

gctggagcct cggtggccat gcttcttgcc ccttgggcct ccccccagcc cctcctcccc 60

ttcctgcacc cgtacccccg tggtctttga ataaagtctg a 101

<210> 93

<211> 3420

<212> DNA

<213> Bacillus hisashii

<400> 93

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> 94
<211> 1112
<212> PRT
<213> Bacillus sp.
<400> 94

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> 95

<211> 1108

<212> PRT

<213> Bacillus thermoamylovorans

<400> 95

```
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> 96
<211> 1302
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
Cas12c1 sequence
<400> 96
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> 97

<211> 1218

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12c2 sequence

<400> 97

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> 98
<211> 1252
<212> PRT
<213> Oleiphilus sp.
<400> 98
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> 99

<211

> 767

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12g1 sequence

<400> 99

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> 100

<211> 870

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12h1 sequence

<400> 100

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> 101

<211> 1093

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12i1 sequence

<400> 101

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> 102

<211> 1054

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Cas12i2 sequence

<400> 102

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> 103

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 103

Gly Gly Ser Gly Gly Ser

1 5

<210> 104

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 104

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 cacccctgag 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> 105

<211> 1361

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 105

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 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> 106

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 106

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> 107

<211> 1361

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 107

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> 108

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 108

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> 109

<211> 1361

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 109

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> 110
<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 110

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> 111
<211> 1361
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 111

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> 112

<211> 4092

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 112

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 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> 113

<211> 1361

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 113

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> 114

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 114

Gly Gly Gly Ser Gly Gly Ser

1 5

<210> 115

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 115

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> 116

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 116

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> 117
<211> 1052
<212> PRT
<213> Staphylococcus aureus
<400> 117
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> 118

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 118

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

Ala 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 Ala 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 Ala 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 Ala 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> 119

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 119

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 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> 120

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 120

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> 121

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 121

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> 122

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 122

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 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> 123

<211> 1359

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 123

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> 124

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 124

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> 125

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 125

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> 126

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 126

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 127

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 127

Gly Gly Gly Ser

1

<210> 128

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 128

Gly Gly Gly Gly Ser

1 5

<210> 129

<211> 5

<212> PRT

<213> Artificial Sequence

<220

><223> Description of Artificial Sequence: Synthetic

peptide

<400> 129

Glu Ala Ala Ala Lys

1 5

<210> 130

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 130

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> 131

<211> 18

<212> DNA

<213> Artificial Sequence

<220

><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 131

ggaggctctg gaggaagc 18

<210> 132

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 132

ggctcttctg gatctgaaac acctggcaca agcgagagcg ccacccctga gagctctggc 60

<210> 133

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 133

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala

1 5 10 15

Ala

<210> 134

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 134

atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc c 51

<210> 135

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 135

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> 136
<211> 1554
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 136
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> 137

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 137

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> 138
<211> 1554
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 138

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> 139

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 139

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> 140

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 140

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> 141

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 141

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> 142
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 142
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> 143

<211> 1540

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 143

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> 144
<211> 1540
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 144

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> 145
<211> 1543
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 145

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> 146

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 146

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> 147

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 147

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> 148

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 148

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> 149

<211> 1551

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 149

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> 150
<211> 1546
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 150
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> 151

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 151

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> 152
<211> 1533
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 152
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> 153

<211> 1549

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 153

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> 154
<211> 1555
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 154
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> 155
<211> 1421
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 155

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> 156

<211> 1552

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 156

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> 157

<211> 1545

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 157

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> 158
<211> 1541
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 158
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> 159

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 159

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> 160

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 160

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> 161
<211> 1541
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 161
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> 162

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 162

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> 163

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 163

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> 164

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 164

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> 165

<211> 1484

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 165

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> 166
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 166

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> 167

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 167

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> 168

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 168

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> 169

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 169

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> 170

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 170

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> 171
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 171
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> 172

<211> 1454

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 172

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> 173
<211> 1420
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 173

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> 174
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 174
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> 175

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 175

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> 176

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 176

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> 177
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 177
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> 178

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 178

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> 179

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 179

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> 180

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 180

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> 181

<211> 178

<212> PRT

<213> Escherichia coli

<400> 181

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> 182

<211> 167

<212> PRT

<213> Escherichia coli

<400> 182

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> 183

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 183

Ser Gly Gly Ser Ser Gly Gly Ser

1 5

<210> 184

<211> 1609

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 184

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> 185

<211> 1807

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 185

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> 186

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 186

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> 187
<211> 1786
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 187
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> 188

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 188

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> 189

<211> 1786

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 189

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>

190

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 190

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> 191
<211> 1786
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 191
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> 192

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 192

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> 193

<211> 1786

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 193

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> 194
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 194

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> 195

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 195

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> 196
<211> 1588
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 196

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> 197

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 197

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> 198

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 198

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> 199
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 199
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> 200

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 200

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> 201
<211> 1588
<212> PRT

<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 201

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> 202
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 202
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> 203

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 203

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> 204

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 204

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> 205

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 205

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> 206

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 206

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> 207
<211> 1588
<212> PRT

<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 207
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> 208
<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> 208
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> 209
<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> 209
guucugtcuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60
cuauugcugg acgaugucuc uuacgaggca uuagcacnnn nnnnnnnnnn nnnnnnn 117
<210> 210
<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> 210
gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggagcaccug aaaacaggug cuuggcacnn nnnnnnnnnn nnnnnnnn 108
<210> 211
<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> 211
gucuaaagga cagaauuuuu caacgggugu gccaauggcc acuuuccagg uggcaaagcc 60
cguugaacuu cucaaaaaga acgaucugag aaguggcacn nnnnnnnnnn nnnnnnnnn 119
<210> 212

<211> 1368

<212> PRT

<213

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 212

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> 213
<211> 20
<212> DNA
<213> Unknown
<220><223> Description of Unknown:
target sequence

<400> 213

ccatgtccag ccttcaggca 20

<210> 214

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 214

tccatgtcca gccttcaggc 20

<210> 215

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 215

ttccatgtcc agccttcagg 20

<210> 216

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 216

gcttccatgt ccagccttca 20

<210> 217

<211> 21

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 217

cttccatgtc cagccttcag g 21

<210> 218

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 218

agcttccatg tccagccttc agg 23

<210> 219

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 219

agcttccatg tccagccttc 20

<210> 220

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 220

cttaagcttc catgtccagc 20

<210> 221

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 221

agcaaaaggc ttttccctgg 20

<210> 222

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 222

gagcaaaagg cttttccctg 20

<210> 223

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 223

agagcaaaag gcttttccct 20

<210> 224

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 224

tagagcaaaa ggcttttccc 20

<210> 225

<211> 21

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 225

tagagcaaaa ggcttttccc t 21

<210> 226

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 226

tttagagcaa aaggcttttc cct 23

<210> 227

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 227

ccatgaagtc aaaatcatta 20

<210> 228

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 228

accatgaagt caaaatcatt 20

<210> 229

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 229

taccatgaag tcaaaatcat 20

<210> 230

<211> 21

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 230

ttaccatgaa gtcaaaatca t 21

<210> 231

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 231

agttaccatg aagtcaaaat cat 23

<210> 232

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 232

cttttcactt cctgccgggg 20

<210> 233

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 233

tcagccccgt ttcttgggaa 20

<210> 234

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 234

gctttcagcc ccgtttcttg 20

<210> 235

<211> 21

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 235

cggctttcag ccccgtttct t 21

<210> 236

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 236

cccggctttc agccccgttt ctt 23

<210> 237

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 237

gcacttcttg atggggagta 20

<210> 238

<211> 20

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 238

tcttgcactt cttgatgggg 20

<210> 239

<211> 21

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 239

cttgcacttc ttgatgggga g 21

<210> 240

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 240

gtcttgcact tcttgatggg gag 23

<210> 241

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 241

ccatgtccag ccttcaggca ggg 23

<210> 242

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 242

tccatgtcca gccttcaggc agg 23

<210> 243

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 243

ttccatgtcc agccttcagg cag 23

<210> 244

<211> 23

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

target sequence

<400> 244

gcttccatgt ccagccttca ggc 23

<210> 245

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence

in view of a mismatch in the gRNA, as the particular

polymerase requires a "g" to initiate transcription

<400> 245

gccaugucca gccuucaggc a 21

<210> 246

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence
in view of a mismatch in the gRNA, as the particular
polymerase requires a "g" to initiate transcription

<400> 246

guccaugucc agccuucagg c 21

<210> 247

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence
in view of a mismatch in the gRNA, as the particular

polymerase requires a "g" to initiate transcription

<400> 247

guuccauguc cagccuucag g 21

<210> 248

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 248

gcuuccaugu ccagccuuca 20

<210> 249

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence
in view of a mismatch in the gRNA, as the particular
polymerase requires a "g" to initiate transcription

<400> 249

gagcuuccau guccagccuu c 21

<210> 250

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence in view of a mismatch in the gRNA, as the particular polymerase requires a "g" to initiate transcription

<400> 250

gagagcaaaa ggcuuuuccc u 21

<210> 251

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence in view of a mismatch in the gRNA, as the particular polymerase requires a "g" to initiate transcription

<400> 251

guagagcaaa aggcuuuucc c 21

<210> 252

<211> 22

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence in view of a mismatch in the gRNA, as the particular polymerase requires a "g" to initiate transcription

<400> 252

guagagcaaa aggcuuuucc cu 22

<210> 253

<211> 24

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence in view of a mismatch in the gRNA, as the particular polymerase requires a "g" to initiate transcription

<400> 253

guuuagagca aaaggcuuuu cccu 24

<210> 254

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence in view of a mismatch in the gRNA, as the particular

polymerase requires a "g" to initiate transcription

<400> 254

gucuugcacu ucuugauggg g 21

<210> 255

<211> 22

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence in view of a mismatch in the gRNA, as the particular polymerase requires a "g" to initiate transcription

<400> 255

gcuugcacuu cuugaugggg ag 22

<210> 256

<211> 24

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<220><223> 5' "G" in the gRNA sequence is added to the sequence in view of a mismatch in the gRNA, as the particular polymerase requires a "g" to initiate transcription

<400> 256

ggucuugcac uucuugaugg ggag 24