(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0126680
(43) 공개일자 2021년10월20일

(51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) *A61K 31/7105* (2006.01)
C07K 14/81 (2006.01) *C12N 15/90* (2006.01)
C12N 5/071 (2010.01) *C12N 9/22* (2006.01)
C12N 9/78 (2006.01)
(52) CPC특허분류
C12N 15/102 (2013.01)
A61K 31/7105 (2013.01)
(21) 출원번호 10-2021-7029279
(22) 출원일자(국제) 2020년02월13일
심사청구일자 없음
(85) 번역문제출일자 2021년09월10일
(86) 국제출원번호 PCT/US2020/018195
(87) 국제공개번호 WO 2020/168135
국제공개일자 2020년08월20일
(30) 우선권주장
62/805,238 2019년02월13일 미국(US)
(뒷면에 계속)

(71) 출원인
빔 테라퓨틱스, 인크.
미국 02139 매사추세츠주 케임브리지 랜즈다운 스트리트 26 세컨드 플로어
(72) 발명자
가우델리 니콜
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트
페커 마이클
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트
(뒷면에 계속)
(74) 대리인
문두현

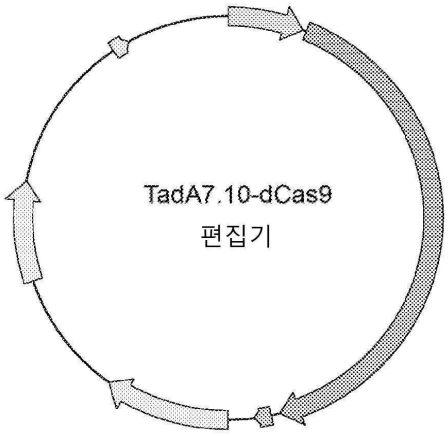
전체 청구항 수 : 총 177 항

(54) 발명의 명칭 **알파-1 항트립신 결핍증을 치료하기 위한 조성물 및 방법**

(57) 요 약

본 발명은 알파-1 항트립신 (A1AT) 결핍증과 연관된 해로운 돌연변이(deleterious mutations)를 편집하기 위한 조성물 및 방법을 특징으로 한다. 특정 구현예에서, 본 발명은 전례 없는 수준의 효율을 갖는 아데노신 데아미나제 염기 편집기 ABE8을 사용하여, A1AT 폴리뉴클레오타이드에서의 돌연변이를 교정하기 위한 방법을 제공한다.

대 표 도 - 도1a



(52) CPC특허분류
C07K 14/8125 (2013.01)
C12N 15/90 (2013.01)
C12N 5/067 (2013.01)
C12N 9/22 (2013.01)
C12N 9/78 (2013.01)
C12Y 305/04004 (2013.01)
C07K 2319/80 (2013.01)
C12N 2310/20 (2017.05)
C12N 2510/00 (2013.01)

(72) 발명자
제츠체 베른트
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트
슬레이메이커 이안
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트
유 이
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트
본 데이비드 에이.
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트
이 승주
미국 매사추세츠주 02139 캠브리지 세컨드 플로어 26 랜즈다운 스트리트

(30) 우선권주장
62/805,271 2019년02월13일 미국(US)
62/852,224 2019년05월23일 미국(US)
62/852,228 2019년05월23일 미국(US)
62/931,722 2019년11월06일 미국(US)
62/941,569 2019년11월27일 미국(US)
62/966,526 2020년01월27일 미국(US)

명 세 서

청구범위

청구항 1

알파-1 항트립신 결핍증과 연관된 단일 뉴클레오타이드 다형성(SNP)을 포함하는 알파-1 항트립신 폴리뉴클레오타이드를 편집하는 방법으로서, 상기 방법이 상기 폴리뉴클레오타이드를, 하나 이상의 가이드 RNA와, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인, 및

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD의 아미노산 위치 82 또는 166에서 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 적어도 하나의 염기 편집기 도메인을 포함하는 염기 편집기와 접촉시키는 단계를 포함하고, 여기서, 상기 가이드 RNA는 상기 염기 편집기를 표적화하여 알파-1 항트립신 결핍증과 연관된 SNP의 변경을 초래하는, 방법.

청구항 2

제1항에 있어서, 상기 아데노신 데아미나제 변이체가 아미노산 위치 82 및 166에서의 변경을 포함하는, 방법.

청구항 3

제1항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S 변경을 포함하는, 방법.

청구항 4

제1항에 있어서, 상기 아데노신 데아미나제 변이체가 T166R 변경을 포함하는, 방법.

청구항 5

제1항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S 및 T166R 변경을 포함하는, 방법.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체는 하기의 변경 중 하나 이상을 추가로 포함하는, 방법: Y147T, Y147R, Q154S, Y123H, 및 Q154R.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 하기의 변경을 포함하는, 방법: 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.

청구항 8

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 149, 150, 151, 152, 153, 154, 155, 156, 및 157로 이루어진 그룹으로부터 선택되는 잔기에서 시작하는 C말단의 결실을 포함하는, 방법.

청구항 9

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 염기 편집기 도메인이 아데노신 데아미나제 변이체 단량체를 포함하고, 상기 아데노신 데아미나제 단량체가 V82S 및 T166R 변경을 포함하는, 방법.

청구항 10

제1항에 있어서, 상기 염기 편집기 도메인이 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 방법.

청구항 11

제10항에 있어서, 상기 아데노신 데아미나제 변이체가 추가로 Y147T, Y147R, Q154S, Y123H, V82S, T166R 및 Q154R로 이루어진 그룹으로부터 선택되는 변경을 포함하는, 방법.

청구항 12

제1항에 있어서, 상기 염기 편집기 도메인이 TadA7.10 도메인 및 아데노신 데아미나제 변이체 도메인을 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 방법.

청구항 13

제12항에 있어서, 상기 아데노신 데아미나제 변이체가 추가로 Y147T, Y147R, Q154S, Y123H, V82S, T166R 및 Q154R로 이루어진 그룹으로부터 선택되는 변경을 포함하는, 방법.

청구항 14

제1항에 있어서, 상기 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 선택되는 변경을 함유하는 아데노신 데아미나제 변이체를 포함하는, 방법.

청구항 15

제1항에 있어서, 상기 아데노신 데아미나제 변이체가 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8 또는 이의 단편인, 방법:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 전장 ABE8에 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 결실된 절단된 ABE8인, 방법.

청구항 17

제1항 내지 제15항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 전장 ABE8에 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 결실된 절단된 ABE8인, 방법.

청구항 18

알파-1 항트립신 결핍증과 연관된 단일 뉴클레오타이드 다형성(SNP)을 포함하는 알파-1 항트립신 폴리뉴클레오타이드를 편집하는 방법으로서, 상기 방법이 알파-1 항트립신 폴리뉴클레오타이드를 하나 이상의 가이드 RNA와, 하기의 서열을 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인:

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFM
QPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYE
KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITG
LYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVI**
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA

KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
YDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*, (여기서, 상기 굵게 표시된 서열은 Cas9 기원의 서열을 지적하고, 상기 이탤릭 서열은 링커 서열을 지칭하고, 상기 밑줄 친 서열은 이분된 (bipartite) 핵 국소화 서열을 지칭함), 및

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 위치 82 및/또는 166에서 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 적어도 하나의 염기 편집기 도메인을 포함하는 융합 단백질과 접촉시킴을 포함하는, 방법.

청구항 19

제18항에 있어서, 상기 아데노신 데아미나제 변이체가 아미노산 위치 82 및 166에서의 변경을 포함하는, 방법.

청구항 20

제18항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S 변경을 포함하는, 방법.

청구항 21

제18항에 있어서, 상기 아데노신 데아미나제 변이체가 T166R 변경을 포함하는, 방법.

청구항 22

제18항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S 및 T166R 변경을 포함하는, 방법.

청구항 23

제18항 내지 제22항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 추가로 하기의 변경 중 하나 이상을 포함하는, 방법: Y147T, Y147R, Q154S, Y123H, 및 Q154R.

청구항 24

제18항에 있어서, 상기 아데노신 데아미나제 변이체가 하기의 변경을 포함하는, 방법: 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.

청구항 25

제1항 내지 제24항 중 어느 한 항에 있어서, 상기 세포가 생체내 또는 생체외인, 방법.

청구항 26

제1항 내지 제24항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 SNP에서 A·T에서 G·C로의 변경이 아미노산 위치 342에서 글루탐산을 라이신으로 변화시키는, 방법.

청구항 27

제1항 내지 제24항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 상기 SNP가 아미노산 위치 342에서 라이신을 갖는 알파-1 항트립신 폴리펩타이드의 발현을 유도하는, 방법.

청구항 28

제1항 내지 제24항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 상기 SNP가 글루탐산을 라이신으로 치환하는, 방법.

청구항 29

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 Y147R + Q154R +Y123H를 포함하는, 방법.

청구항 30

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 Y147R + Q154R + I76Y를 포함하는, 방법.

청구항 31

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 Y147R + Q154R + T166R을 포함하는, 방법.

청구항 32

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 Y147T + Q154R을 포함하는, 방법.

청구항 33

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 Y147T + Q154S를 포함하는, 방법.

청구항 34

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 Y147R + Q154S를 포함하는, 방법.

청구항 35

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S + Q154S를 포함하는, 방법.

청구항 36

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S + Y147R을 포함하는, 방법.

청구항 37

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S + Q154R을 포함하는, 방법.

청구항 38

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S + Y123H를 포함하는, 방법.

청구항 39

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 I76Y + V82S를 포함하는, 방법.

청구항 40

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S + Y123H + Y147T를 포함하는, 방법.

청구항 41

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S + Y123H + Y147R을 포함하는, 방법.

청구항 42

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S + Y123H + Q154R을 포함하는, 방법.

청구항 43

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 Y123H+ Y147R + Q154R + I76Y를 포함하는, 방법.

청구항 44

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S + Y123H + Y147R + Q154R을 포함하는, 방법.

청구항 45

제24항에 있어서, 상기 아데노신 데아미나제 변이체가 I76Y + V82S + Y123H + Y147R + Q154R을 포함하는, 방법.

청구항 46

제1항 내지 제45항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변형된 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 (*Streptococcus thermophilus*) 1 Cas9 (St1Cas9), 변형된 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 또는 이의 변이체인, 방법.

청구항 47

제46항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변경된 프로토스페이서-인접 모티프 (PAM) 특이성 또는 비-G PAM에 대해 특이성을 갖는 SpCas9의 변이체를 포함하는, 방법.

청구항 48

제47항에 있어서, 상기 변경된 PAM이 상기 핵산 서열 5'-NGC-3'에 대해 특이성을 갖는, 방법.

청구항 49

제47항 또는 제48항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 50

제1항 내지 제49항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 뉴

클레아제 불활성 또는 닉카제 변이체인, 방법.

청구항 51

제50항에 있어서, 상기 닉카제 변이체가 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 52

제1항 내지 제51항 중 어느 한 항에 있어서, 상기 염기 편집기가 추가로 아연 핑거 도메인을 포함하는, 방법.

청구항 53

제1항 내지 제52항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 데옥시리보핵산 (DNA)에서 아데닌을 탈아민화시킬 수 있는, 방법.

청구항 54

제1항 내지 제53항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제가 천연적으로 발생하지 않는 변형된 아데노신 데아미나제인, 방법.

청구항 55

제1항 내지 제54항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제가 TadA 데아미나제인, 방법.

청구항 56

제55항에 있어서, 상기 TadA 데아미나제가 TadA*7.10인, 방법.

청구항 57

제1항 내지 제56항 중 어느 한 항에 있어서, 상기 하나 이상의 가이드 RNA가 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 여기서, 상기 crRNA가 알파-1 항트립신 결핍증과 연관된 SNP를 포함하는 알파-1 항트립신 핵산 서열에 상보적인 핵산 서열을 포함하는, 방법.

청구항 58

제1항 내지 제57항 중 어느 한 항에 있어서, 상기 염기 편집기가 알파-1 항트립신 결핍증과 연관된 SNP를 포함하는 알파-1 항트립신 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체 형태로 있는, 방법.

청구항 59

제1항 내지 제58항 중 어느 한 항의 염기 편집기 및 하기로 이루어진 그룹으로부터 선택되는 핵산 서열을 포함하는 가이드 RNA를 포함하는 염기 편집 시스템:

5′-ACCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-CCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-CAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-AUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-UCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-CGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′

청구항 60

제59항에 있어서, 상기 gRNA가 추가로 핵산 서열 5'-GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'를 포함하는, 염기 편집 시스템.

청구항 61

세포 또는 이의 선조체로 하기를 도입함에 의해 생성된 세포:

상기 세포에 대한 염기 편집기, 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드 (여기서, 상기 염기 편집기는 제1항 내지 제60항 중 어느 한 항에 기재된 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 아데노신 데아미나제 도메인을 포함한다), 및

알파-1 항트립신 결핍증과 연관된 SNP의 A·T에서 G·C로의 변경을 초래하기 위해 염기 편집기를 표적화하는 하나 이상의 가이드 폴리뉴클레오타이드.

청구항 62

제61항에 있어서, 상기 생성된 세포가 간세포 또는 이의 선조체인, 세포.

청구항 63

제61항 또는 제62항에 있어서, 상기 세포가 알파-1 항트립신 결핍증을 갖는 대상체로부터 기원하는, 세포.

청구항 64

제61항 내지 제63항 중 어느 한 항에 있어서, 상기 세포가 포유동물 세포 또는 인간 세포인, 세포.

청구항 65

제61항 내지 제64항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 SNP에서 A·T에서 G·C로의 변경이 상기 알파-1 항트립신 폴리펩타이드에서 글루탐산을 라이신으로 변화시키는, 세포.

청구항 66

제61항 내지 제65항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 상기 SNP가 아미노산 위치 42에서 라이신을 갖는 알파-1 항트립신 폴리펩타이드의 발현을 유도하는, 세포.

청구항 67

제61항 내지 제65항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 상기 SNP가 글루탐산을 라이신으로 치환하는, 세포.

청구항 68

제61항 내지 제67항 중 어느 한 항에 있어서, 상기 세포가 알파-1 항트립신 결핍증과 연관된 SNP의 A·T에서 G·C로의 변경에 대해 선택되는, 세포.

청구항 69

제61항 내지 제68항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변형된 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 (*Streptococcus thermophilus*) 1 Cas9 (St1Cas9), 변형된 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 또는 이의 변이체인, 세포.

청구항 70

제69항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변경된 프로토스페이서-인접 모티프 (PAM) 특이성 또는 비-G PAM에 대해 특이성을 갖는 SpCas9의 변이체를 포함하는, 세포.

청구항 71

제70항에 있어서, 상기 변경된 PAM이 상기 핵산 서열 5'-NGC-3'에 대해 특이성을 갖는, 세포.

청구항 72

제70항 또는 제71항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R, 또는 이의 상응하는 아미노산 치환을 포함하는, 세포.

청구항 73

제61항 내지 제72항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 뉴클레아제 불활성 또는 닉카제 변이체인, 세포.

청구항 74

제73항에 있어서, 상기 닉카제 변이체가 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함하는, 세포.

청구항 75

제71항 내지 제74항 중 어느 한 항에 있어서, 상기 염기 편집기가 추가로 아연 핑거 도메인을 포함하는, 세포.

청구항 76

제71항 내지 제75항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 도메인이 데옥시리보핵산 (DNA)에서 아데닌을 탈아민화시킬 수 있는, 세포.

청구항 77

제71항 내지 제76항 중 어느 한 항에 있어서, 상기 하나 이상의 가이드 폴리뉴클레오타이드가 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 여기서, 상기 crRNA가 알파-1 항트립신 결핍증과 연관된 SNP를 포함하는 알파-1 항트립신 핵산 서열에 상보적인 핵산 서열을 포함하는, 세포.

청구항 78

제71항 내지 제77항 중 어느 한 항에 있어서, 상기 염기 편집기 및 상기 하나 이상의 가이드 폴리뉴클레오타이드가 세포 내 복합체를 형성하는, 세포.

청구항 79

제78항에 있어서, 상기 염기 편집기가 알파-1 항트립신 결핍증과 연관된 SNP를 포함하는 알파-1 항트립신 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체 형태로 있는, 세포.

청구항 80

대상체에서 알파-1 항트립신 결핍증을 치료하기 위한 방법으로서, 상기 방법이 제71항 내지 제79항 중 어느 한 항의 세포를 상기 대상체에게 투여함을 포함하는, 방법.

청구항 81

제80항에 있어서, 상기 세포가 상기 대상체에 자가인, 방법.

청구항 82

제80항에 있어서, 상기 세포가 상기 대상체에 동종이계인, 방법.

청구항 83

제71항 내지 제82항 중 어느 한 항의 세포로부터 증식되거나 확장된 단리된 세포 또는 세포 집단.

청구항 84

간세포 또는 이의 선조체를 생성하는 방법으로서, 상기 방법이 하기의 단계를 포함하는, 방법:

(a) 알파-1 항트립신 결핍증과 연관된 SNP를 포함하는 간세포 선조체에,

염기 편집기, 또는 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드(여기서, 상기 염기 편집기는 제1항 내지 제60항 중 어느 한 항에 기재된 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 아데노신 데아미나제 변이체 도메인을 포함한다), 및

알파-1 항트립신 결핍증과 연관된 SNP의 A·T에서 G·C로의 변경을 초래하기 위해 염기 편집기를 표적화하는 하나 이상의 가이드 폴리뉴클레오타이드를 도입하는 단계; 및

(b) 상기 간세포 선조체를 간세포로 분화시키는 단계.

청구항 85

간세포를 생성하는 방법으로서, 상기 방법이 하기의 단계를 포함하는, 방법:

(a) 알파-1 항트립신 결핍증과 연관된 SNP를 포함하는 간세포에,

염기 편집기, 또는 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드(여기서, 상기 염기 편집기는 제1항 내지 제60항 중 어느 한 항에 기재된 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 아데노신 데아미나제 변이체 도메인을 포함한다), 및

알파-1 항트립신 결핍증과 연관된 SNP의 A·T에서 G·C로의 변경을 초래하기 위해 상기 염기 편집기를 표적화하는 하나 이상의 가이드 폴리뉴클레오타이드를 도입하는 단계.

청구항 86

제85항에 있어서, 상기 간세포 또는 간세포 선조체가 포유 동물 세포 또는 인간 세포인, 방법.

청구항 87

제85항 또는 제86항에 있어서, 알파-1 항트립신 결핍증과 연관된 SNP에서 상기 A·T에서 G·C로의 변경이 상기 알파-1 항트립신 폴리펩타이드에서 글루탐산을 라이신으로 변화시키는, 방법.

청구항 88

제85항 내지 제87항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 상기 SNP가 아미노산 위치 342에서 라이신을 갖는 알파-1 항트립신 결핍증 폴리펩타이드의 발현을 유도하는, 방법.

청구항 89

제85항 내지 제88항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 상기 SNP가 글루탐산을 라이신으로 치환하는, 방법.

청구항 90

제85항 내지 제89항 중 어느 한 항에 있어서, 상기 세포가 알파-1 항트립신 결핍증과 연관된 상기 SNP의 A·T에서 G·C로의 변경에 대해 선택되는, 방법.

청구항 91

제85항 내지 제90항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변형된 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 (*Streptococcus thermophilus*) 1 Cas9 (St1Cas9), 변형된 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 또는 이의 변이체인, 방법.

청구항 92

제85항 내지 제91항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변경된 프로토스페이서-인접 모티프 (PAM) 특이성을 갖는 변형된 SpCas9를 포함하는, 방법.

청구항 93

제92항에 있어서, 상기 변경된 PAM이 상기 핵산 서열 5'-NGC-3'에 대해 특이성을 갖는, 방법.

청구항 94

제92항 또는 제93항에 있어서, 상기 변형된 SpCas9가 아미노산 치환 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 95

제85항 내지 제94항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 뉴클레아제 불활성 또는 닉카제 변이체인, 방법.

청구항 96

제95항에 있어서, 상기 닉카제 변이체가 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 97

제85항 내지 제96항 중 어느 한 항에 있어서, 상기 염기 편집기가 추가로 아연 핑거 도메인을 포함하는, 방법.

청구항 98

제85항 내지 제97항 중 어느 한 항에 있어서, 상기 하나 이상의 가이드 RNA가 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 여기서, 상기 crRNA가 알파-1 항트립신 결핍증과 연관된 SNP를 포함하는 알파-1 항트립신 핵산 서열에 상보적인 핵산 서열을 포함하는, 방법.

청구항 99

제85항 내지 제98항 중 어느 한 항에 있어서, 상기 염기 편집기 및 상기 하나 이상의 가이드 폴리뉴클레오타이드가 세포 내 복합체를 형성하는, 방법.

청구항 100

제99항에 있어서, 상기 염기 편집기가 알파-1 항트립신 결핍증과 연관된 상기 SNP를 포함하는 알파-1 항트립신 핵산 서열에 상보적인 핵산 서열을 포함하는 단일 가이드 RNA (sgRNA)와 복합체 형태로 있는, 방법.

청구항 101

제1항 내지 제58항 중 어느 한 항에 있어서, 상기 Cas9가 StCas9 또는 SaCas9인, 방법.

청구항 102

제1항 내지 제58항 중 어느 한 항에 있어서, 상기 Cas9이 변형된 SaCas9인, 방법.

청구항 103

제102항에 있어서, 상기 변형된 SaCas9가 아미노산 치환 E782K, N968K, 및 R1015H, 또는 이의 상응하는 아미노산 치환을 포함하는, 방법.

청구항 104

제102항에 있어서, 상기 변형된 SaCas9이 하기 아미노산 서열을 포함하는, 방법.

KRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
IQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQ
KAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNS
ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
NRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
PRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKT
KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSF
LRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIE
TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNG
LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIE
VNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG.

청구항 105

대상체에서 알파-1 항트립신 결핍증 (A1AD)을 치료하기 위한 방법으로서, 상기 방법이 하기의 단계를 포함하는, 방법: 상기 대상체에게 Cas9 또는 Cas12 폴리펩타이드 내 삽입된 아데노신 데아미나제 변이체를 포함하는 융합 단백질, 또는 상기 융합 단백질을 암호화하는 폴리뉴클레오타이드; 및 상기 융합 단백질을 표적화하여 A1AD와 연관된 단일 뉴클레오타이드 다형성(SNP)의 A·T에서 G·C로의 변경을 초래하는 하나 이상의 가이드 폴리뉴클레오타이드를 투여하여 상기 대상체에서 A1AD를 치료하는 단계.

청구항 106

대상체에서 알파-1 항트립신 결핍증 (A1AD)을 치료하는 방법으로서, 상기 방법이 하기의 단계를 포함하는, 방법:

상기 대상체에게 아데노신 염기 편집기 8 (ABE8), 또는 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드 (여기서, ABE8은 Cas9 또는 Cas12 폴리펩타이드 내 삽입된 아데노신 데아미나제 변이체를 포함한다); 및 ABE8을 표적화하여 A1AD와 연관된 SNP의 A·T에서 G·C로의 변경을 초래하는 하나 이상의 가이드 폴리뉴클레오타이드를 투여하여 상기 대상체에서 A1AD를 치료하는 단계.

청구항 107

제105항 또는 제106항에 있어서, 상기 ABE8이 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-d로부터 선택되는, 방법.

청구항 108

제105항 내지 제107항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 서열을 포함하고; 상기 아미노산 서열이 적어도 하나의 변경을 포함하는, 방법.

청구항 109

제108항에 있어서, 상기 아데노신 데아미나제 변이체가 아미노산 위치 82 및/또는 166에서의 변경을 포함하는, 방법.

청구항 110

제108항 또는 제109항에 있어서, 상기 적어도 하나의 변경이 하기를 포함하는, 방법: V82S, T166R, Y147T, Y147R, Q154S, Y123H, 및/또는 Q154R.

청구항 111

제105항 내지 제110항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 하기의 변경의 조합 중 하나를 포함하는, 방법: 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.

청구항 112

제105항 내지 제111항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, 또는 TadA*8.24인, 방법.

청구항 113

제105항 내지 제112항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 149, 150, 151, 152, 153, 154, 155, 156, 및 157로 이루어진 그룹으로부터 선택되는 잔기에서 시작하는 C말단의 결실을 포함하는, 방법.

청구항 114

제105항 내지 제113항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 TadA*8 아데노신 데아미나제 변이체 도메인을 포함하는 아데노신 데아미나제 단량체인, 방법.

청구항 115

제105항 내지 제113항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 야생형 아데노신 데아미나제 도메인 및 TadA*8 아데노신 데아미나제 변이체 도메인을 포함하는 아데노신 데아미나제 이종이량체인, 방법.

청구항 116

제105항 내지 제113항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 TadA 도메인 및 TadA*8 아데노신 데아미나제 변이체 도메인을 포함하는 아데노신 데아미나제 이종이량체인, 방법.

청구항 117

제105항 내지 제116항 중 어느 한 항에 있어서, A1AD와 연관된 상기 SNP에서 상기 A·T에서 G·C로의 변경이 아미노산 위치 342에서 글루탐산을 라이신으로 변화시키는, 방법.

청구항 118

제105항 내지 제116항 중 어느 한 항에 있어서, A1AD와 연관된 상기 SNP가 아미노산 위치 342에서 라이신을 갖

는 알파-1 항트립신 폴리펩타이드의 발현을 유도하는, 방법.

청구항 119

제105항 내지 제116항 중 어느 한 항에 있어서, 알파-1 항트립신 결핍증과 연관된 상기 SNP가 글루탐산을 라이신으로 치환하는, 방법.

청구항 120

제105항 내지 제119항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 Cas9 또는 Cas12 폴리펩타이드의 가요성 루프, 알파 나선 영역, 비구조적 부분 또는 용매 접근 가능한 부분 내에 삽입되는, 방법.

청구항 121

제105항 내지 제119항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 Cas9 또는 Cas12 폴리펩타이드의 N-말단 단편 및 C-말단 단편에 의해 플랭킹된, 방법.

청구항 122

제121항에 있어서, 상기 융합 단백질 또는 ABE8이 구조 NH_2-[Cas9 또는 Cas12 폴리펩타이드의 N-말단 단편]-[아데노신 데아미나제 변이체]-[Cas9 또는 Cas12 폴리펩타이드의 C-말단 단편]-COOH를 포함하고, 여기서 "]-["는 각각의 경우 임의의 링커인, 방법.

청구항 123

제121항 또는 제122항에 있어서, N-말단 단편의 C-말단 또는 C 말단 단편의 N-말단이 Cas9 또는 Cas12 폴리펩타이드의 가요성 루프의 일부를 포함하는, 방법.

청구항 124

제120항에 있어서, 상기 가요성 루프가, 상기 아데노신 데아미나제 변이체가 표적 핵염기를 탈아민화시키는 경우, 표적 핵염기에 근접한 아미노산을 포함하는, 방법.

청구항 125

제105항 내지 제124항 중 어느 한 항에 있어서, 상기 대상체에게 가이드 핵산 서열을 투여하여 A1AD와 연관된 SNP 표적 핵염기의 탈아민화를 수행함을 추가로 포함하는, 방법.

청구항 126

제125항에 있어서, 상기 SNP 표적 핵염기의 탈아민화가 표적 핵염기를 야생형 핵염기로 또는 비-야생형 핵염기로 대체하고, 상기 표적 핵염기의 탈아민화가 A1AD의 증상을 개선시키는, 방법.

청구항 127

제126항에 있어서, A1AD와 연관된 상기 SNP의 탈아민화가 글루탄산을 라이신으로 치환시키는, 방법.

청구항 128

제105항 내지 제127항 중 어느 한 항에 있어서, 상기 표적 핵염기가 표적 폴리뉴클레오타이드 서열에서 PAM 서열로부터 1 내지 20개 핵염기로 이격되어 있는, 방법.

청구항 129

제128항에 있어서, 상기 표적 핵염기가 PAM 서열의 2 내지 12개 핵염기 업스트림에 있는, 방법.

청구항 130

제121항 내지 제129항 중 어느 한 항에 있어서, 상기 Cas9 또는 Cas12 폴리펩타이드의 N-말단 단편 또는 C-말단 단편이 표적 폴리뉴클레오타이드 서열에 결합하는, 방법.

청구항 131

제130항에 있어서,

상기 N-말단 단편 또는 상기 C-말단 단편이 RuvC 도메인을 포함하거나;

상기 N-말단 단편 또는 상기 C-말단 단편이 HNH 도메인을 포함하거나;

상기 N-말단 단편 및 상기 C-말단 단편의 어느 것도 HNH 도메인을 포함하지 않거나;

상기 N-말단 단편 및 상기 C-말단 단편의 어느 것도 RuvC 도메인을 포함하지 않는, 방법.

청구항 132

제115항 내지 제131항 중 어느 한 항에 있어서, 상기 Cas9 또는 Cas12 폴리펩타이드가 하나 이상의 구조적 도메인에서 부분적 또는 완전한 결실을 포함하고, 상기 데아미나제가 Cas9 또는 Cas12 폴리펩타이드의 부분적 또는 완전한 결실 위치에 삽입되는, 방법.

청구항 133

제132항에 있어서,

상기 결실이 RuvC 도메인 내에 있거나;

상기 결실이 HNH 도메인 내에 있는, 방법.

청구항 134

제105항 내지 제133항 중 어느 한 항에 있어서, 상기 융합 단백질 또는 ABE8이 Cas9 폴리펩타이드를 포함하는, 방법.

청구항 135

제134항에 있어서, 상기 Cas9 폴리펩타이드가 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 1 (*Streptococcus thermophilus 1*) Cas9 (St1Cas9), 또는 이의 변이체인, 방법.

청구항 136

제134항 또는 제135항에 있어서, 상기 Cas9 폴리펩타이드가 하기의 아미노산 서열 (Cas9 참조 서열) 또는 이의 상응하는 영역을 포함하는, 방법:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

(한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인; (Cas9 참조 서열).

청구항 137

제136항에 있어서,

상기 Cas9 폴리펩타이드가 Cas9 폴리펩타이드 참조 서열에 넘버링된 바와 같은 아미노산 1017-1069 또는 이의 상응하는 아미노산의 결실을 포함하거나;

상기 Cas9 폴리펩타이드가 Cas9 폴리뉴클레오타이드 참조 서열에 넘버링된 바와 같은 아미노산 792-872 또는 이의 상응하는 아미노산의 결실을 포함하거나;

상기 Cas9 폴리펩타이드가 Cas9 폴리뉴클레오타이드 참조 서열에 넘버링된 바와 같은 아미노산 792-906 또는 이의 상응하는 아미노산의 결실을 포함하는, 방법.

청구항 138

제134항 내지 제137항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 상기 Cas9 폴리펩타이드의 가요성 루프 내에 삽입되는, 방법.

청구항 139

제138항에 있어서, 상기 가요성 루프가 Cas9 참조 서열에서 넘버링된 바와 같은 위치 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300에서 또는 이의 상응하는 아미노산 위치에서 아미노산 잔기로 이루어진 그룹으로부터 선택된 영역을 포함하는, 방법.

청구항 140

제136항에 있어서, 상기 데아미나제 변이체가 Cas9 참조 서열에서 넘버링된 바와 같은 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 삽입되는, 방법.

청구항 141

제136항에 있어서, 상기 데아미나제 변이체가 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입되는, 방법.

청구항 142

제136항에 있어서, 상기 데아미나제 변이체가 Cas9 참조 서열에서 넘버링된 바와 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입되는, 방법.

청구항 143

제136항에 있어서, 아데노신 데아미나제 변이체가 표 13A에서 동정된 유전자좌에서 Cas9 폴리펩타이드 내에 삽입된, 방법.

청구항 144

제136항에 있어서, 상기 N-말단 단편이 Cas9 참조 서열의 아미노산 잔기 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 및/또는 1248-1297 또는 이의 상응하는 잔기를 포함하는, 방법.

청구항 145

제136항에 있어서, 상기 C-말단 단편이 Cas9 참조 서열의 아미노산 잔기 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 및/또는 538-568 또는 이의 상응하는 잔기를 포함하는, 방법.

청구항 146

제134항 내지 제145항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩타이드가 변형된 Cas9이고, 변경된 PAM 또는 비-G PAM에 대해 특이성을 갖는, 방법.

청구항 147

제134항 내지 제146항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩타이드가 닉카제이거나 상기 Cas9 폴리펩타이드가 불활성 뉴클레아제인, 방법.

청구항 148

제134항 내지 제145항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩타이드가 변형된 SpCas9 폴리펩타이드인, 방법.

청구항 149

제148항에 있어서, 상기 변형된 SpCas9 폴리펩타이드가 아미노산 치환 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R (SpCas9-MQKFRAER)을 포함하고, 변경된 PAM 5'-NGC-3'에 대해 특이성을 갖는, 방법.

청구항 150

제105항 내지 제133항 중 어느 한 항에 있어서, 상기 융합 단백질 또는 ABE8이 Cas12 폴리펩타이드를 포함하는, 방법.

청구항 151

제150항에 있어서, 상기 아데노신 데아미나제 변이체가 Cas12 폴리펩타이드로 삽입되는, 방법.

청구항 152

제150항 또는 제151항에 있어서, 상기 Cas12 폴리펩타이드가 Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i인, 방법.

청구항 153

제151항 또는 제152항에 있어서, 상기 아데노신 데아미나제 변이체가:

아미노산 위치: a) BhCas12b의 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;

b) BvCas12b의 147 및 148, 248 및 249, 299 및 300, 991 및 992, 또는 1031 및 103 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i의 상응하는 아미노산 잔기; 또는

c) AaCas12b의 157 및 158, 258 및 259, 310 및 311, 1008 및 1009, 또는 1044 및 1045, 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i의 상응하는 아미노산 잔기 사이에 삽입되는, 방법.

청구항 154

제151항에 있어서, 아데노신 데아미나제 변이체가 표 13B에서 동정된 유전자좌에서 Cas12 폴리펩타이드 내에 삽입된, 방법.

청구항 155

제154항에 있어서, 상기 Cas12 폴리펩타이드가 Cas12b인, 방법.

청구항 156

제154항에 있어서, 상기 Cas12 폴리펩타이드가 BhCas12b 도메인, BvCas12b 도메인, 또는 AACas12b 도메인을 포함하는, 방법.

청구항 157

제105항 내지 제156항 중 어느 한 항에 있어서, 상기 가이드 RNA가 CRISPR RNA (crRNA) 및 트랜스-활성화 crRNA (tracrRNA)를 포함하는, 방법.

청구항 158

제105항 내지 제157항 중 어느 한 항에 있어서, 상기 대상체가 포유류 또는 인간인, 방법.

청구항 159

제59항 또는 제60항의 염기 편집 시스템, 및 약제학적으로 허용되는 담체, 비히클 또는 부형제를 포함하는 약제학적 조성물.

청구항 160

제61항 내지 제79항 중 어느 한 항의 세포, 또는 약제학적으로 허용되는 담체, 비히클 또는 부형제를 포함하는 약제학적 조성물.

청구항 161

제59항 또는 제60항의 염기 편집 시스템을 포함하는 키트.

청구항 162

제61항 내지 제79항 중 어느 한 항의 세포를 포함하는 키트.

청구항 163

제161항 또는 제162항에 있어서, 사용 지침서와 함께 패키지 삽입물을 추가로 포함하는, 키트.

청구항 164

폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인, 및
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 위치 82 또는 166에서 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 적어도 하나의 염기 편집기 도메인 및 가이드 RNA를 포함하는 염기 편집기 시스템으로서, 상기 가이드 RNA가 상기 염기 편집기를 표적화하여 알파-1 항트립신 결핍증과 연관된 SNP의 변경을 초래하는, 염기 편집기 시스템.

청구항 165

제164항에 있어서, 상기 아데노신 데아미나제 변이체가 V82S 변경 및/또는 T166R 변경을 포함하는, 염기 편집기 시스템.

청구항 166

제165항에 있어서, 상기 아데노신 데아미나제 변이체가 추가로 하기의 변경 중 하나 이상을 포함하는, 염기 편집기 시스템: Y147T, Y147R, Q154S, Y123H, 및 Q154R.

청구항 167

제165항 또는 제166항에 있어서, 상기 염기 편집기 도메인이 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함하는, 염기 편집기 시스템.

청구항 168

제164항 내지 제167항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 전장 TadA8에 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 결실된 절단된 TadA8인, 염기 편집기 시스템.

청구항 169

제164항 내지 제167항 중 어느 한 항에 있어서, 상기 아데노신 데아미나제 변이체가 전장 TadA8에 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 결실된 절단된 TadA8인, 염기 편집기 시스템.

청구항 170

제164항 내지 제69항 중 어느 한 항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변형된 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 1 (*Streptococcus thermophilus 1*) Cas9 (St1Cas9), 변형된 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 또는 이의 변이체인, 염기 편집기 시스템.

청구항 171

제170항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 변경된 프로토스페이서-인접 모티프 (PAM) 특이성 또는 비-G PAM에 대해 특이성을 갖는 SpCas9의 변이체인, 염기 편집기 시스템.

청구항 172

제170항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 뉴클레아제 불활성 Cas9인, 염기 편집기 시스템.

청구항 173

제170항에 있어서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이 Cas9 닉카제인, 염기 편집기 시스템.

청구항 174

하나 이상의 가이드 RNA와, 하기의 서열을 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인:

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFM
QPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYE
KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITG
LYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVI**
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
YDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*,

(여기서, 상기 굵게 표시된 서열은 Cas9 기원의 서열을 지적하고, 상기 이탤릭 서열은 링커 서열을 지칭하고, 상기 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭함), 및

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

의 아미노산 위치 82 및/또는 166에서 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 적어도 하나의 염기 편집기 도메인을 포함하는 융합 단백질을 포함하는 염기 편집기 시스템.

청구항 175

제164항 내지 제174항 중 어느 한 항의 염기 편집기 시스템을 포함하는 세포.

청구항 176

제175항에 있어서, 상기 세포가 인간 세포 또는 포유류 세포인, 세포.

청구항 177

제175항에 있어서, 상기 세포가 생체외, 생체내 또는 시험관내인, 세포.

발명의 설명

기 술 분 야

[0001] **관련 출원에 대한 상호 참조**

[0002] 본 출원은 PCT 국제 특허 출원으로서 2019년 2월 13일에 출원된 미국 가출원 제62/805,238호; 2019년 2월 13일에 출원된 제62/805,271호; 2019년 5월 23일에 출원된 제62/852,224호; 2019년 5월 23일에 출원된 제62/852,228호; 2019년 11월 6일에 출원된 제62/931,722호, 2019년 11월 27일에 출원된 제62/941,569호; 및 2020년 1월 27일에 출원된 제62/966,526호에 대한 우선권의 이득을 주장하고, 이의 모든 내용은 이들의 전문이 본원에 참조로 포함된다.

[0003] **참조 인용**

[0004] 본 명세서에 언급된 모든 공보, 특허 및 특허 출원은 각각의 개별 공보, 특허 또는 특허 출원이 구체적으로 및 개별적으로 참조로 인용되는 것과 동일한 정도로 참조로 본원에 인용된다. 달리 지적되지 않는 경우, 본 명세서에 언급된 공보, 특허 및 특허 출원은 이들의 전문이 본원에 참조로 포함된다.

배 경 기 술

[0005] 건강한 개체에서, 알파-1 항트립신 (A1AT)은 간 내 간세포에 의해 생성되고 전신 순환계로 분비되고, 여기서 이것은 프로테아제 저해제로서 기능한다. 이것은 특히 호중구 엘라스타제의 양호한 저해제여서 폐와 같은 조직 및 기관을 엘라스틴 분해로부터 보호한다. 알파-1 항트립신 결핍증 (A1AD)을 갖는 환자에서, A1AT를 암호화하는 유전자에서의 변형은 단백질 생성을 감소시킨다. 결과적으로, 폐 내 엘라스틴은 호중구 엘라스타제에 의해 보다 신속하게 분해되고 기간 경과에 따라 폐 탄력성을 상실시키고 만성 폐쇄성 폐 질환 (COPD)을 발병시킨다.

[0006] 대부분의 통상의 병원성 A1AT 변이체는 구아닌의 아데닌으로의 돌연변이이고 이는 아미노산 342에서 글루타메이트의 라이신으로의 치환을 유도한다. 상기 치환은 단백질이 간세포 내에서 잘못 폴딩되어 중합되도록 하고 궁극적으로 독성 응집물은 간 손상 및 간경변을 유도할 수 있다. 간 독성은 유전자 녹아웃 (CRISPR/ZFN/TALEN) 또는 유전자 녹다운 (siRNA)에 의해 해결될 수 있지만, 어느 접근법도 폐 병리를 해소하지 못한다. 폐 병리는 단백질 대체 치료요법으로 해소될 수 있지만 상기 치료요법은 또한 간 독성을 해소하는 데 실패한다. 유전자 치료요법은 또한 A1AT 유전자 결함을 해소하기 위해 부적절하다. A1AD를 갖는 환자의 간은 이미 내인성 A1AT에 의해 유발된 중증의 질환 부담하에 있기 때문에, 간에서 A1AT를 증가시키는 유전자 치료요법은 비생산적이다.

[0007] 따라서, A1AD를 갖는 환자를 치료하여 폐 병리 및 간 독성 둘 다를 해소하는 방법이 요구된다.

발명의 내용

[0008] **개요**

[0009] 하기된 바와 같이, 본 발명은 알파-1 항트립신 결핍증 (A1AD)와 연관된 해로운 돌연변이(deleterious mutations)를 편집하기 위한 조성물 및 방법을 특징으로 한다. 특정 구현예에서, 본 발명은 전례 없는 수준 (예를 들어, >60-70%)의 효율 및 특이성을 갖는, "ABE8"로 호칭되는 변형된 아데노신 데아미나제를 사용하여 A1AD를 치료하기 위한 방법을 제공하여 A1AD와 연관된 돌연변이를 교정한다.

[0010] 하나의 양상에서, 본 발명은 알파-1 항트립신 결핍증과 연관된 단일 뉴클레오타이드 다형성(SNP)을 함유하는 알파-1 항트립신 폴리뉴클레오타이드를 편집하는 방법을 제공하고, 상기 방법은 상기 폴리뉴클레오타이드를, 하나 이상의 가이드 RNA와, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인, 및

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0011] 의 아미노산 위치 82 또는 166에서 변경을 함유하는 아데노신 데아미나제 변이체인 적어도 하나의 염기 편집기 도메인을 함유하는 염기 편집기와 접촉시킴을 포함하고, 여기서, 상기 가이드 RNA는 염기 편집기를 표적화하여 알파-1 항트립신 결핍증과 연관된 SNP의 변경을 초래한다.

[0012] 또 다른 양상에서, 본 발명은 알파-1 항트립신 결핍증과 연관된 단일 뉴클레오타이드 다형성(SNP)을 함유하는 알파-1 항트립신 폴리뉴클레오타이드를 편집하는 방법을 제공하고, 상기 방법은 알파-1 항트립신 폴리뉴클레오타이드를 하나 이상의 가이드 RNA와, 하기의 서열을 함유하는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합

도메인:

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYR
STKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAI
GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV
LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*,

[0013] (여기서, 상기 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고 밑줄 친 서열은 이분된(bipartite) 핵 국소화 서열을 지칭함), 및

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0014] 의 아미노산 위치 82 또는 166에서 변경을 함유하는 아데노신 데아미나제 변이체를 함유하는 적어도 하나의 염기 편집기 도메인을 포함하는 융합 단백질과 접촉시킴을 포함한다.

[0015] 또 다른 양상에서, 본 발명은 임의의 이전의 양상의 융합 단백질 및 하기로부터 선택되는 핵산 서열을 함유하는 가이드 RNA를 함유하는 염기 편집 시스템을 제공한다:

5′-ACCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC
AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU
GGCACCGAGU CGGUGCUUUU-3′
5′-CCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-CAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-AUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-UCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′
5′-CGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU
AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3′

[0016]

[0017] 또 다른 양상에서, 세포는 세포 또는 이의 선조체에:

[0018] 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인 및 임의의 이전의 양상에 기재된 아데노신 데아미나제 도메인을 함유하는 염기 편집기, 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드; 및 상기 염기 편집기를 표적화하여 알파-1 항트립신 결핍증과 연관된 SNP의 A·T에서 G·C로의 변경을 초래하는 하나 이상의 가이드 폴리뉴클레오타이드를 도입함에 의해 생성된다. 하나의 구현예에서, 생성된 세포는 간세포 또는 이의 선조체이다. 또 다른 구현예에서, 상기 세포는 알파-1 항트립신 결핍증을 갖는 대상체로부터 기원한다. 또 다른 구현예에서, 세포는 포유동물 세포 또는 인간 세포이다.

[0019] 상기 양상의 다양한 구현예에서, gRNA는 핵산 서열 5'-GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3을 추가로 함유한다.

[0020] 또 다른 양상에서, 본 발명은 대상체에서 알파-1 항트립신 결핍증을 치료하는 방법을 제공하고, 상기 방법은 임의의 이전의 양상의 세포를 상기 대상체에게 투여하는 단계를 포함한다. 하나의 양상에서, 상기 세포는 상기 대상체에 자가이거나 동종이계이다.

[0021] 또 다른 양상에서, 본 발명은 단리된 세포 또는 상기된 양상 및 구현예의 세포로부터 증식되거나 확장된 세포 집단을 제공한다.

[0022] 또 다른 양상에서, 본 발명은 간세포를 생성하는 방법을 제공하고, 상기 방법은: (a) 알파-1 항트립신 결핍증과 연관된 SNP를 함유하는 간세포에 상기된 양상 및 구현예의 임의의 하나에 기재된 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 아데노신 데아미나제 변이체 도메인을 포함하는 염기 편집기 또는 상기 염기 편집기를 암호화하는 뉴클레오타이드; 및 상기 염기 편집기를 표적화하여 알파-1 항트립신 결핍증과 연관된 SNP의 A·T에서 G·C로의 변경을 초래하는 하나 이상의 가이드 폴리뉴클레오타이드를 도입함을 포함한다.

[0023] 다양한 구현예에서, 간세포는 포유동물 세포 또는 인간 세포이다.

[0024] 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 아미노산 위치 82 및 166에서의 변경을 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 V82S 변경을 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 T166R 변경을 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 V82S 및 T166R 변경을 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 하기의 변경 중 하나 이상을 추가로 포함한다: Y147T, Y147R, Q154S, Y123H, 및 Q154R. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 하기의 변경을 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. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 Y147R+ Q154R+Y123H를 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 Y147R+Q154R+I76Y를 포함한다. 상기된 양상의 구현예에서, 아

데노신 데아미나제 변이체는 Y147R+Q154R+T166R을 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 Q147T+Q154R을 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 Q147T+Q154S를 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 Y147R+Q154S를 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 V82S+Q154S를 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 V82S+Y147R을 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 V82S+Q154R을 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 V82S+Y123H를 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 I76Y+V82S를 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 V82S+Y123H+Y147T를 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 V82S+Y123H+Y147R을 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 V82S+Y123H+Q154R을 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 Y123H+Y147R+Q154R+I76Y를 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 V82S+Y123H+Y147R+Q154R을 포함한다. 상기된 양상의 구현예에서, 아데노신 데아미나제 변이체는 I76Y+V82S+Y123H+Y147R+ Q154R을 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 149, 150, 151, 152, 153, 154, 155, 156, 및 157로 이루어진 그룹으로부터 선택되는 잔기에서 시작하는 C말단의 결실을 포함한다. 상기 양상의 다른 구현예에서, 염기 편집기는 V82S 및 T166R을 함유하는 단일 아데노신 데아미나제 변이체를 포함한다. 상기 양상의 다른 구현예에서, 상기 염기 편집기 도메인은 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, Q154R로 이루어진 그룹으로부터 선택되는 변경을 추가로 포함한다. 상기 양상의 다른 구현예에서, 상기 염기 편집기 도메인은 TadA7.10 도메인 및 아데노신 데아미나제 변이체를 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 Y147T, Y147R, Q154S, Y123H, V82S, T166R, Q154R로 이루어진 그룹으로부터 선택되는 변경을 추가로 포함한다. 상기 양상의 다른 구현예에서, 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 선택되는 변경을 함유하는 아데노신 데아미나제 변이체를 포함한다. 상기 양상의 다른 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8 또는 이의 단편이다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0025] 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 전장 ABE8에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, 또는 20개 N-말단 아미노산 잔기가 결실된 절단된 ABE8을 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 변이체는 전장 ABE8에 비해 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, 또는 20개 C-말단 아미노산 잔기가 결실된 절단된 ABE8이다.

[0026] 상기 양상의 다른 구현예에서, 알파-1 항트립신 결핍증과 연관된 SNP에서 A·T에서 G·C로의 변경은 알파-1 항트립신 폴리펩타이드에서 글루탐산을 라이신으로 변화시킨다. 상기 양상의 다른 구현예에서, 알파-1 항트립신 결핍증과 연관된 SNP는 아미노산 위치 342에서 라이신을 갖는 알파-1 항트립신 폴리펩타이드의 발현을 유도한다. 상기 양상의 다른 구현예에서, 알파-1 항트립신 결핍증과 연관된 SNP는 글루탐산을 라이신으로 치환시킨다. 상기 양상의 다른 구현예에서, 세포는 알파-1 항트립신 결핍증과 연관된 SNP의 A·T에서 G·C로의 변경에 대해 선택된다. 상기 양상의 다른 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 변형된 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 1 (*Streptococcus thermophilus 1*) Cas9 (St1Cas9), 변형된 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 또는 이의 변이체이다. 상기 양상의 다른 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 변경된 프로토스페이서-인접 모티프 (PAM) 특이성 또는 비-G PAM에 대해 특이성을 갖는 SpCas9의 변이체를 포함한다. 상기 양상의 다른 구현예에서, 변경된 PAM은 핵산 서열 5'-NGC-3'에 대해 특이성을 갖는다. 상기 양상의 다른 구현예에서, 변경된 SpCas9는 아미노산 치환 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R, 또는 이의 상응하는 아미노산 치환을 포함한다. 상기 양상의 다른 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 뉴클레아제 불활성 또는 닉카제 변이

체이다. 상기 양상의 다른 구현예에서, 닉카제 변이체는 아미노산 치환 D10A 또는 이의 상응하는 아미노산 치환을 포함한다. 상기 양상의 다른 구현예에서, 염기 편집기는 아연 핑거 도메인을 추가로 포함한다. 상기 양상의 다른 구현예에서, 아데노신 데아미나제 도메인은 데옥시리보핵산 (DNA)에서 아데닌을 탈아민화시킬 수 있다. 상기 양상의 다른 구현예에서, 하나 이상의 가이드 RNA는 CRISPR RNA (crRNA) 및 트랜스-암호화된 소형 RNA (tracrRNA)를 포함하고, 여기서, 상기 crRNA는 알파-1 항트립신 결핍증과 연관된 SNP를 함유하는 알파-1 항트립신 핵산 서열에 상보적인 핵산 서열을 포함한다. 상기 양상의 다른 구현예에서, 염기 편집기 및 상기 하나 이상의 가이드 폴리뉴클레오타이드는 세포 내 복합체를 형성한다. 상기 양상의 다른 구현예에서, 염기 편집기는 알파-1 항트립신 결핍증과 연관된 SNP를 함유하는 알파-1 항트립신 핵산 서열에 상보적인 핵산 서열을 함유하는 단일 가이드 RNA (sgRNA)와 복합체 형태로 있다.

[0027] 또 다른 양상에서, 대상체에서 알파-1 항트립신 결핍증 (A1AD)을 치료하기 위한 방법이 제공되고, 여기서, 상기 방법은 상기 대상체에게 Cas9 또는 Cas12 폴리펩타이드 내 삽입된 아데노신 데아미나제 변이체를 포함하는 융합 단백질, 또는 상기 융합 단백질을 암호화하는 폴리뉴클레오타이드; 및 상기 융합 단백질을 표적화하여 A1AD와 연관된 단일 뉴클레오타이드 다형성(SNP)의 A·T에서 G·C로의 변경을 초래하는 하나 이상의 가이드 폴리뉴클레오타이드를 투여하여 상기 대상체에서 A1AD를 치료함을 포함한다.

[0028] 또 다른 양상에서, 대상체에서 알파-1 항트립신 결핍증 (A1AD)을 치료하기 위한 방법이 제공되고, 여기서, 상기 방법은 상기 대상체에게 아데노신 염기 편집기, Cas9 또는 Cas12 폴리펩타이드 내 삽입된 아데노신 데아미나제 변이체를 포함하는 ABE8 또는 상기 염기 편집기를 암호화하는 폴리뉴클레오타이드; 및 ABE8을 표적화하여 A1AD와 연관된 SNP의 A·T에서 G·C로의 변경을 초래하는 하나 이상의 가이드 폴리뉴클레오타이드를 투여하여 상기 대상체에서 A1AD를 치료함을 포함한다.

[0029] 상기된 방법의 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-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-d로부터 선택된다. 상기된 방법의 구현예에서, 아데노신 데아미나제 변이체는 하기의 아미노산 서열을 포함하고:

[0030] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD; 여기서, 상기 아미노산 서열은 적어도 하나의 변경을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체는 아미노산 위치 82 및/또는 166에서의 변경을 포함한다. 하나의 구현예에서, 적어도 하나의 변경은 다음을 포함한다: V82S, T166R, Y147T, Y147R, Q154S, Y123H, 및/또는 Q154R.

[0031] 상기된 방법의 구현예에서, 아데노신 데아미나제 변이체는 하기 변경의 조합 중 하나를 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. 상기된 방법의 구현예에서, 아데노신 데아미나제 변이체는 TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, 또는 TadA*8.24이다. 하나의 구현예에서, 아데노신 데아미나제 변이체는 149, 150, 151, 152, 153, 154, 155, 156, 및 157로 이루어진 그룹으로부터 선택되는 잔기에서 시작하는 C말단의 결실을 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체는 TadA*8 아데노신 데아미나제 변이체 도메인을 포함하는 아데노신 데아미나제 단량체이다. 하나의 구현예에서, 아데노신 데아미나제 변이체는 야생형 아데노신 데아미나제 도메인 및 TadA*8 아데노신 데아미나제 변이체 도메인을 포함하는 아데노신 데아미나제 이종이량체이다. 하나의 구현예에서, 아데노신 데아미나제 변이체는 TadA 도메인 및 TadA*8 아데노신 데아미나제 변이체 도메인을 포함하는 아데노신 데아미나제 이종이량체이다.

[0032] 상기된 방법의 구현예에서, A1AD와 연관된 SNP에서 A·T에서 G·C로의 변경은 아미노산 위치 342에서 글루탐산을 라이신으로 변화시킨다. 상기된 방법의 구현예에서, A1AD와 연관된 SNP는 아미노산 위치 342에서 라이신을

갖는 알파-1 항트립신 폴리펩타이드의 발현을 유도한다. 상기된 방법의 구현예에서, 알파-1 항트립신 결핍증과 연관된 SNP는 글루탐산을 라이신으로 치환시킨다.

[0033] 상기된 방법의 구현예에서, 아데노신 데아미나제 변이체는 Cas9 또는 Cas12 폴리펩타이드의 가요성 루프, 알파 나선 영역, 비구조적 부분 또는 용매 접근 가능한 부분 내 삽입된다. 상기된 방법의 구현예에서, 아데노신 데아미나제 변이체는 Cas9 또는 Cas12 폴리펩타이드의 N-말단 단편 및 C-말단 단편에 의해 플랭킹된다.

[0034] 상기된 방법의 구현예에서, 융합 단백질 또는 ABE8은 구조 NH_2-[Cas9 또는 Cas12 폴리펩타이드의 N-말단 단편]-[아데노신 데아미나제 변이체]-[Cas9 또는 Cas12 폴리펩타이드의 C-말단 단편]-COOH를 포함하고, 여기서 "]-["는 각각의 경우 임의의 링커이다. 하나의 구현예에서, N-말단 단편의 C-말단 또는 C 말단 단편의 N-말단은 Cas9 또는 Cas12 폴리펩타이드의 가요성 루프의 일부를 포함한다. 하나의 구현예에서, 가요성 루프는 아데노신 데아미나제 변이체가 표적 핵염기를 탈아민화시키는 경우 표적 핵염기에 근접한 아미노산을 포함한다.

[0035] 상기된 방법의 구현예에서, 상기 방법은 대상체에게 가이드 핵산 서열을 투여하여 A1AD와 연관된 SNP 표적 핵염기의 탈아민화를 수행하는 단계를 추가로 포함한다. 상기된 방법의 구현예에서, SNP 표적 핵염기의 탈아민화는 표적 핵염기를 야생형 핵염기로 또는 비-야생형 핵염기로 대체하고, 상기 표적 핵염기의 탈아민화는 A1AD의 증상을 개선시킨다. 상기된 방법의 구현예에서, A1AD와 연관된 SNP의 탈아민화는 글루탄산을 라이신으로 치환시킨다.

[0036] 상기된 방법의 구현예에서, 표적 핵염기는 표적 폴리뉴클레오타이드 서열에서 PAM 서열로부터 1 내지 20개 핵염기로 이격되어 있다. 하나의 구현예에서, 표적 핵염기는 PAM 서열의 2 내지 12개 핵염기 업스트림에 있다. 상기된 방법의 구현예에서, Cas9 또는 Cas12 폴리펩타이드의 N-말단 단편 또는 C-말단 단편은 표적 폴리뉴클레오타이드 서열에 결합한다. 특정 구현예에서, N-말단 단편 또는 C-말단 단편은 RuvC 도메인을 포함하거나; N-말단 단편 또는 C-말단 단편은 HNH 도메인을 포함하거나; N-말단 단편 및 C-말단 단편의 어느 것도 HNH 도메인을 포함하지 않거나; N-말단 단편 및 C-말단 단편의 어느 것도 RuvC 도메인을 포함하지 않는다. 하나의 구현예에서, Cas9 또는 Cas12 폴리펩타이드는 하나 이상의 구조적 도메인에서 부분적 또는 완전한 결실을 포함하고, 여기서, 데아미나제는 Cas9 또는 Cas12 폴리펩타이드의 부분적 또는 완전한 결실 위치에 삽입된다. 특정 구현예에서, 결실은 RuvC 도메인에 내에 있고; 결실은 HNH 도메인에 내에 있고; 또는 결실은 RuvC 도메인 및 C-말단 도메인을 브릿징한다.

[0037] 상기된 방법의 구현예에서, 융합 단백질 또는 ABE8은 Cas9 폴리펩타이드를 포함한다. 하나의 구현예에서, Cas9 폴리펩타이드는 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 1 (*Streptococcus thermophilus 1*) Cas9 (St1Cas9), 또는 이의 변이체이다. 하나의 구현예에서, Cas9 폴리펩타이드는 하기의 아미노산 서열 (Cas9 참조 서열)을 포함한다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF
VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0038] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인; (Cas9 참조 서열), 또는 이의 상응하는 영역). 특정 구현예에서, Cas9 폴리펩타이드는 Cas9 폴리펩타이드 참조 서열에서 넘버링된 바와 같은 아미노산 1017-1069 또는 이의 상응하는 아미노산의 결실을 포함하거나; Cas9 폴리펩타이드는 Cas9 폴리펩타이드 참조 서열에서 넘버링된 바와 같은 아미노산 792-872 또는 이의 상응하는 아미노산의 결실을 포함하거나; Cas9 폴리펩타이드는 Cas9 폴리펩타이드 참조 서열에서 넘버링된 바와 같은 아미노산 792-906 또는 이의 상응하는 아미노산의 결실을 포함한다. 상기된 방법의 하나의 구현예에서, 아데노신 데아미나제 변이체는 Cas9 폴리펩타이드의 가요성 루프 내에 삽입된다.

[0039] 하나의 구현예에서, 가요성 루프는 Cas9 참조 서열에서 넘버링된 바와 같은 위치 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300에서 또는 이의 상응하는 아미노산 위치에서 아미노산 잔기로 이루어진 그룹으로부터 선택된 영역을 포함한다.

[0040] 상기된 방법의 구현예에서, 데아미나제 변이체는 Cas9 참조 서열에서 넘버링된 바와 같은 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 위치 사이에 삽입된다. 상기된 방법의 구현예에서, 데아미나제 변이체는 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 상기된 방법의 구현예에서, 데아미나제 변이체는 Cas9 참조 서열에서 넘버링된 바와 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 상기된 방법의 구현예에서, 아데노신 데아미나제 변이체는 표 13A에서 동정된 유전자좌에서 Cas9 폴리펩타이드 내에 삽입된다. 하나의 구현예에서, N-말단 단편은 Cas9 참조 서열의 아미노산 잔기 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 및/또는 1248-1297 또는 이의 상응하는 잔기를 포함한다. 하나의 구현예에서, C-말단 단편은 Cas9 참조 서열의 아미노산 잔기 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 및/또는 538-568 또는 이의 상응하는 잔기를 포함한다.

[0041] 상기된 방법의 구현예에서, Cas9 폴리펩타이드는 변형된 Cas9이고, 변경된 PAM 또는 비-G PAM에 대해 특이성을

갖는다. 상기된 방법의 구현예에서, Cas9 폴리펩타이드는 닉카제이거나, Cas9 폴리펩타이드는 뉴클레아제 불활성이다. 상기된 방법의 구현예에서, Cas9 폴리펩타이드는 변형된 SpCas9 폴리펩타이드이다. 하나의 구현예에서, 변형된 SpCas9 폴리펩타이드는 아미노산 치환 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R (SpCas9-MQKFRAER)을 포함하고 변경된 PAM 5'-NGC-3'에 대해 특이성을 갖는다.

[0042] 상기된 방법의 또 다른 구현예에서, 융합 단백질 또는 ABE8은 Cas12 폴리펩타이드를 포함한다. 하나의 구현예에서, 아데노신 데아미나제 변이체는 Cas12 폴리펩타이드에 삽입된다. 하나의 구현예에서, Cas12 폴리펩타이드는 Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i이다. 하나의 구현예에서, 아데노신 데아미나제 변이체는 아미노산 위치: a) BhCas12b의 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; b) BvCas12b의 147 및 148, 248 및 249, 299 및 300, 991 및 992, 또는 1031 및 103 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i의 상응하는 아미노산 잔기; 또는 c) AaCas12b의 157 및 158, 258 및 259, 310 및 311, 1008 및 1009, 또는 1044 및 1045, 또는 Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, 또는 Cas12i의 상응하는 아미노산 잔기 사이에 삽입된다. 하나의 구현예에서, 아데노신 데아미나제 변이체는 표 13B에서 동정된 유전자좌에서 Cas12 폴리펩타이드 내에 삽입된다. 하나의 구현예에서, Cas12 폴리펩타이드는 Cas12b이다. 하나의 구현예에서, Cas12 폴리펩타이드는 BhCas12b 도메인, BvCas12b 도메인, 또는 AACas12b 도메인을 포함한다.

[0043] 상기된 방법의 구현예에서, 가이드 RNA는 CRISPR RNA (crRNA) 및 트랜스-활성화 crRNA (tracrRNA)를 포함한다. 상기된 방법의 구현예에서, 대상체는 포유류 또는 인간이다.

[0044] 또 다른 양상에서, 상기된 방법, 양상 및 구현예의 임의의 하나의 염기 편집 시스템, 및 약제학적으로 허용되는 담체, 비히클 또는 부형제를 포함하는 약제학적 조성물이 제공된다.

[0045] 하나의 양상에서, 상기된 양상 및 구현예의 세포, 및 약제학적으로 허용되는 담체, 비히클 또는 부형제를 포함하는 약제학적 조성물이 제공된다.

[0046] 또 다른 양상에서, 상기된 방법, 양상 및 구현예의 임의의 하나의 염기 편집 시스템을 포함하는 키트가 제공된다.

[0047] 또 다른 양상에서, 상기된 양상 및 구현예의 임의의 하나의 세포를 포함하는 키트가 제공된다. 키트의 구현예에서, 상기 키트는 사용 지침서와 함께 패키지 삽입물을 추가로 포함한다.

[0048] 하나의 양상에서, 본원에 제공된 것은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인, 및

[0049] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 위치 82 또는 166에서 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 적어도 하나의 염기 편집기 도메인을 포함하는 염기 편집기이다.

[0050] 하나의 양상에서, 염기 편집기 시스템은 상기된 이의 염기 편집기 및 가이드 RNA를 포함하고, 상기 가이드 RNA는 상기 염기 편집기를 표적화하여 알파-1 항트립신 결핍증과 연관된 SNP의 변경을 초래한다. 일부 구현예에서, 아데노신 데아미나제 변이체는 V82S 변경 및/또는 T166R 변경을 포함한다. 일부 구현예에서, 아데노신 데아미나제 변이체는 하기의 변경 중 하나 이상을 추가로 포함한다: Y147T, Y147R, Q154S, Y123H, 및 Q154R. 일부 구현예에서, 염기 편집기 도메인은 야생형 아데노신 데아미나제 도메인 및 아데노신 데아미나제 변이체를 포함하는 아데노신 데아미나제 이종이량체를 포함한다. 일부 구현예에서, 아데노신 데아미나제 변이체는 전장 TadA8에 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 결실된 절단된 TadA8이다. 일부 구현예에서, 아데노신 데아미나제 변이체는 전장 TadA8에 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 결실된 절단된 TadA8이다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 변형된 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 1 (*Streptococcus thermophilus 1*) Cas9 (St1Cas9), 변형된 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 또는 이의 변이체이다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 변경된 프로토스페이서-인접 모티프 (PAM) 특이성 또는 비-G PAM에 대해 특이성을 갖는

SpCas9의 변이체이다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 뉴클레아제 불활성 Cas9이다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인은 Cas9 닉카제이다.

[0051] 하나의 양상에서, 본원에서는 하나 이상의 가이드 RNA와, 하기의 서열을 포함하는 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인을 포함하는 융합 단백질을 포함하는 염기 편집기 시스템이 제공된다:

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFM
QPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYE
KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITG
LYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVI**
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAA
KNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD
YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK
YDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*, 여기서, 상기 굵게 표시된 서열은 Cas9 기원의 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열, 및

[0052] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD 의 아미노산 위치 82 및/또는 166에서 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는 적어도 하나의 염기 편집기 도메인을 지칭한다.

[0053]

[0054] 하나의 양상에서, 상기된 염기 편집기 시스템 중 어느 하나를 포함하는 세포가 제공된다. 일부 구현예에서, 세포는 인간 세포 또는 포유동물 세포이다. 일부 구현예에서, 상기 세포는 생체외, 생체내 또는 시험관내이다.

[0055] 본 발명은 알파-1 항트립신 결핍증 (A1AD)과 연관된 돌연변이를 편집하기 위한 조성물 및 방법을 제공한다. 본 발명에 의해 정의된 조성물 및 제품은 단리되거나 다르게는 하기 제공된 실시예와 연계하여 제조된다. 본 발명의 다른 특성 및 이점은 상세한 설명 및 청구항으로부터 자명하다.

[0056] **정의**

[0057] 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 물질들을 본 발명의 시험을 위한 실시에서 사용할 수 있지만, 바람직한 물질 및 방법들을 본원에 기재한다. 따라서, 본원에 사용된 용어는 단지 특정 구현예를 기재하기 위한 것이고 본 발명을 제한하는 것으로 의도되지 않는다.

[0058] 달리 정의되지 않는 경우, 본원에서 사용된 모든 기술적 및 과학적 용어는 본 발명이 속하는 기술 분야의 통상의 기술자에 의해 통상적으로 이해되는 의미를 갖는다. 하기의 참조문헌은 당업자에게 본 발명에 사용되는 많은 용어의 일반 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).

[0059] 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, "포함하는"이라는 용어 뿐만 아니라 "포함한다 (include)", "포함한다 (includes)" 및 "포함되는"과 같은 다른 형태의 사용은 제한적이지 않다.

[0060] 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.

[0061] 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 1 또는 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, 상기 용어는 예를 들어, 값의 5배 이내 또는 2배 이내와 같은 한 자릿수 (order of magnitude) 이내를 의미할 수 있다. 특정 값이 본원 및 청구항에 기재되는 경우, 달리 언급되지 않는다면, 용어 "약"은 특정 값에 대해 허용되는 오차 범위 내를 의미하는 것으로 추정되어야만 한다.

[0062] 본원에 제공된 범위는 제1 및 마지막 값을 포함하는 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.

[0063] 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.

[0064] "아데노신 데아미나제"는 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, 진화된 아데노신 데아미나제)는 임의의 유기체, 예를 들어, 세균으로부터 기원할 수 있다.

[0065] 일부 구현예에서, 아데노신 데아미나제는 하기의 서열에서의 변경을 포함한다:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
MPRQVFNAQK KAQSSTD

[0066]

[0067] (또한 TadA*7.10으로 호칭됨).

[0068] 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 하기 변경의 하나 이상을 포함한다: Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R. 변경 Y123H는 또한 본원에서 H123H로서 언급된다 (TadA*7.10에서 변경 H123Y는 Y123H (wt)로 복귀함). 다른 구현예에서, TadA*7.10 서열의 변이체는 하기의 그룹으로부터 선택되는 변경의 조합을 포함한다: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R.

[0069] 다른 구현예에서, 본 발명은 잔기 149, 150, 151, 152, 153, 154, 155, 156, 또는 157에서 개시하는 C 말단의 결실을 포함하는 결실, 예를 들어, TadA*8을 포함하는 아데노신 데아미나제 변이체를 제공한다. 다른 구현예에서, 아데노신 데아미나제 변이체는 하기의 변경의 하나 이상을 포함하는 TadA (예를 들어, TadA*8) 단량체이다: Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R. 다른 구현예에서, 아데노신 데아미나제 변이체는 하기의 그룹으로부터 선택되는 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.

[0070] 여전히 다른 구현예에서, 아데노신 데아미나제 변이체는 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 2개의 아데노신 데아미나제 도메인 (예를 들어, TadA*8)을 포함하는 동종이량체이고, 이의 각각은 하기의 그룹으로부터 선택되는 변경의 조합을 갖는다: 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.

[0071] 다른 구현예에서, 아데노신 데아미나제 변이체는 야생형 TadA 아데노신 데아미나제 도메인 및 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 하기의 그룹으로부터 선택된 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.

[0072] 다른 구현예에서, 아데노신 데아미나제 변이체는 TadA*7.10 도메인 및 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 TadA*7.10 도메인, 및 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; 또는 I76Y + V82S + Y123H + Y147R + Q154R.

[0073] 하나의 구현예에서, 아데노신 데아미나제는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 TadA*8 또는 이의 단편이다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

[0074] GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0075] 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 결실이다. 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 결실이다. 일부 구현예에서, 아데노신 데아미나제 변이체는 전장 TadA*8이다.

[0076] 특정 구현예에서, 아데노신 데아미나제 이종이량체는 TadA*8 도메인, 및 하기 중 하나로부터 선택되는 아데노신 데아미나제 도메인을 포함한다:

[0077] 스타필로코커스 아우레우스 (*Staphylococcus aureus*) (에스. 아우레우스 (*S. aureus*)) TadA:

MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIA
IERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQS

[0078] NFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN

[0079] 바실러스 서브틸리스 (*Bacillus subtilis*) (비. 서브틸리스 (*B. subtilis*)) TadA:

MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEA
CKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNH

[0080] QAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[0081] 살모넬라 티피무리움 (*Salmonella typhimurium*) (에스. 티피무리움 (*S. typhimurium*)) TadA:

MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGA

[0082] AGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV

[0083] 쉐와넬라 푸트레파시엔스 (*Shewanella putrefaciens*) (에스. 푸트레파시엔스 (*S. putrefaciens*)) TadA:

MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGK
KLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQV

[0084] EVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[0085] 해모필러스 인플루엔자 (*Haemophilus influenzae*) F3031 (에이취. 인플루엔자 (*H. influenzae*)) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHA
EIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHF

[0086] FDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK

[0087] 콜로박터 크레슨투스 (*Caulobacter crescentus*) (씨. 크레슨투스 (*C. crescentus*)) TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHA
EIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKF

[0088] FAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

[0089] 게오박터 설푸레두센스 (*Geobacter sulfurreducens*) (지. 설푸레두센스 (*G. sulfurreducens*)) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHA
EMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDL

[0090] SADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP

[0091] TadA*7.10

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

[0092] GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0093] "아데노신 데아미나제 염기 편집기 8 (ABE8) 폴리펩타이드"는 하기의 참조 서열의 아미노산 위치 82 및/또는 166에서 변경을 포함하는 아데노신 데아미나제 변이체를 포함하는, 본원에 정의되고/되거나 기재된 바와 같은 염기 편집기 (BE)를 의미한다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT

[0094] GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0095] 일부 구현예에서, ABE8은 참조 서열에 비해 추가의 변경을 포함한다.

[0096] "아데노신 데아미나제 염기 편집기 8 (ABE8) 폴리뉴클레오타이드"는 ABE8 폴리펩타이드를 암호화하는 폴리뉴클레오타이드 (폴리뉴클레오타이드 서열)를 의미한다.

[0097] "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. 대안적으로, 또는 동시에, 투여는 경구 경로에 의한 것일 수 있다.

[0098] "제제"란 임의의 소분자의 화학적 화합물, 항체, 핵산 분자 또는 폴리펩타이드 또는 이의 단편을 의미한다.

[0099] "알파-1 항트립신 (A1AT) 단백질"은 UniProt 승인 번호 P01009와 적어도 약 95% 아미노산 서열 동일성을 갖는 폴리펩타이드 또는 이의 단편을 의미한다. 특정 구현예에서, A1AT 단백질은 하기의 참조 서열에 비해 하나 이상의 변경을 포함한다. 하나의 특정 구현예에서, A1AD와 연관된 A1AT 단백질은 E342K 돌연변이를 포함한다. 예시적인 A1AT 아미노산 서열은 하기에 제공된다.

[0100] >sp|P01009|A1AT_인간 알파-1-항트립신 OS=호모 사피엔스 OX=9606 GN=SERPINA1 PE=1 SV=3

MPSSVSWGILLLAGLCCLVPVSLAEDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLY
RQLAHQSNSTNIFFSPVSIATAFAMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELL
RTLNQPDSQLQLTTGNGLFLSEGLKLVDKFLEDVKKLYHSEAFTVNFGDTEEAKKQINDYVE
KGTQGKIVDLVKELDRDTVFALVNYIFFKGKWERPFEVKDTEEEDFHVDQVTTVKVPMMKRL
GMFNIQHCKKLSSWVLLMKYLGNATAIFFLPDEGKLQHLENELTHDIITKFLENEDRRSASL
HLPKLSITGTYDLKSVLGQLGITKVFSNGADLSGVTEEAPLKLSKAVHKAVLTIDEKGTEAA

[0101] GAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK

[0102] 상기 A1AT 단백질 서열에서, 제1의 24개 아미노산은 신호 펩타이드 (밑줄 친)를 구성한다. A1AD (즉, E342K)에 돌연변이된 서열의 위치 342는 아미노산 "1"로서 신호 서열 이후 세팅 아미노산 잔기 "E"를 기준으로 결정된다.

[0103] "변경"이란 본원에 기재된 것들과 같이 표준 당업계 방법에 의해 검출된 바와 같은 유전자 또는 폴리펩타이드의 구조, 발현 수준 또는 활성에서의 변화 (예를 들어, 증가 또는 감소)를 의미한다. 본원에 사용된 바와 같은 변경은 폴리뉴클레오타이드 또는 폴리펩타이드 서열에서의 변화 또는 발현 수준에서의 변화, 예를 들어, 25% 변화, 40% 변화, 50% 변화 이상을 포함한다.

[0104] "개선한다"는 질환의 발병 또는 진행의 감소, 억제, 감쇠, 약화, 정지 또는 안정화를 의미한다.

[0105] "유사체"란 동일하지 않지만 유사한 기능 또는 구조적 특징을 갖는 분자를 의미한다. 예를 들어, 폴리뉴클레오타이드 또는 폴리펩타이드 유사체는 상응하는 천연적으로 발생하는 폴리뉴클레오타이드 또는 폴리펩타이드의 생물학적 활성을 보유하면서 천연적으로 발생하는 폴리뉴클레오타이드 또는 폴리펩타이드에 비해 유사체의 기능을 증진시키는 특정 변형을 갖는다. 상기 변형은 예를 들어, 리간드 결합을 변경하지 않고 DNA에 대한 유사체의 친화성, 효율, 특이성, 프로테아제 또는 뉴클레아제 내성, 막 투과성 및/또는 반감기를 증가시킬 수 있다. 유사체는 비천연 뉴클레오타이드 또는 아미노산을 포함할 수 있다.

[0106] "염기 편집기 (BE)" 또는 "핵염기 편집기 (NBE)"는 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, 상기 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, 상기 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 RNA와, RNA 결합 도메인 상의 RNA 결합 모티프를 통해)에 연결된다. 일부 구현예에서, 염기 편집기 활성을 갖는 도메인은 핵산 분자 내 염기를 탈아민화시킬 수 있다. 일부 구현예에서, 염기 편집기는 DNA 분자 내 하나 이상의 염기를 탈아민화시킬 수 있다. 일부 구현예에서, 염기 편집기는 DNA 내 아데노신 (A)을 탈아민화시킬 수 있다. 일부 구현예에서, 염기 편집기는 아데노신 염기 편집기 (ABE)이다.

[0107] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 변이체 (예를 들어, TadA*8)를 환형의 퍼뮤턴트 Cas9 (예를 들어, spCAS9 또는 saCAS9) 및 이분된 핵 국소화 서열을 포함하는 스캐폴드에 클로닝함에 의해 생성된다. 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 이분된 핵 국소화 서열을 지칭한다.

[0108] CP5 (MSP "NGC=NGG와 같은 돌연변이 정규 Cas9를 갖는 팜 변이체", PID=단백질 상호작용 도메인 및 "D10A" 닉카제와 함께):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYR
STKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAI
GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV
LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0109]

[0110] 일부 구현예에서, ABE8은 하기 **표 6-9, 13,** 또는 **14**로부터의 염기 편집기로부터 선택된다. 일부 구현예에서, ABE8은 TadA로부터 유래된 아데노신 데아미나제 변이체를 함유한다. 일부 구현예에서, ABE8의 아데노신 데아미나제 변이체는 하기 **표 7, 9, 13** 또는 **14**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 선택되는 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 + Q154R의 그룹으로부터 선택되는 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, ABE8은 서열을 포함한다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0111]

[0112] 일부 구현예에서, 폴리뉴클레오타이드 프로그램 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. 이들 염기 편집기의 세부사항은 국제 PCT 출원 PCT/2017/045381 (WO 2018/027078) 및 PCT/US2016/058344 (WO2017/070632)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); 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.

[0113] 예를 들어, 염기 편집 조성물에 사용된 바와 같은 아데닌 염기 편집기 (ABE), 본원에 기재된 시스템 및 방법은 하기에 제공된 바와 같은 핵산 서열 (8877 염기쌍) (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.)을 갖는다. ABE 핵산 서열과 적어도 95% 이상의 동일성을 갖는 폴리뉴클레오타이드 서열이 또한 포함된다.

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC
ATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGAT
GCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC
CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAAC
TCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTT
AGTGAACCGTCAGATCCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC
ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGA
GTTTAGCCACGAGTATTGGATGAGGCACGCACTGACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAG
TCCCCGTGGGCGCCGTGCTGGTGCACAACAATAGAGTGATCGGAGAGGGATGGAACAGGCCAATCGGC
CGCCACGACCCTACCGCACACGCAGAGATCATGGCACTGAGGCAGGGAGGCCTGGTCATGCAGAATTA
CCGCCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGCGTGATGTGCGCAGGAGCAATGATCC
ACAGCAGGATCGGAAGAGTGGTGTTCGGAGCACGGGACGCCAAGACCGGCGCAGCAGGCTCCCTGATG
GATGTGCTGCACCACCCCGGCATGAACCACCGGGTGGAGATCACAGAGGGAATCCTGGCAGACGAGTG
CGCCGCCCTGCTGAGCGATTTCTTTAGAATGCGGAGACAGGAGATCAAGGCCCAGAAGAAGGCACAGA
GCTCCACCGACTCTGGAGGATCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGGCACAAGCGAGTCC
GCCACACCAGAGAGCTCCGGCGGCTCCTCCGGAGGATCCTCTGAGGTGGAGTTTTCCCACGAGTACTG
[0114] GATGAGACATGCCCTGACCCTGGCCAAGAGGGCACGCGATGAGAGGGAGGTGCCTGTGGGAGCCGTGC

TGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAACAGCC
CATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCAC
CCTGTACGTGACATTCGAGCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCG
TGGTGTTTGGCGTGAGGAACGCAAAAACCGGCGCCGCAGGCTCCCTGATGGACGTGCTGCACTACCCC
GGCATGAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCAGATGAATGTGCCGCCCTGCTGTGCTA
TTTCTTTCGGATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGGCCCAGAGCTCCACCGACTCCGGAG
GATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCGAGAGCGCAACACCTGAAAGCAGC
GGGGGCAGCAGCGGGGGGTCAGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGG
CTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACC
GGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACC
CGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGAT
CTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGG
AAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAG
AAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCT
GATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACC
CCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACG
GCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCC
TGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTG
AGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCT
GTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGA
TCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTG
CTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAA
CGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCC
TGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAG
CGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCG
GCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCA
TCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAG
GAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGA
GCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACG
AGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCC
TTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGT
GAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGG
AAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTC
CTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAG
AGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGA
AGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAG

[0115]

TCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGAT
CCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCC
TGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAG
GTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAG
AGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCA
TCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAG
CTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCT
GTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGG
TGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAG
ATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGAC
CAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAA
CCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT
GACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGA
TTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCG
TCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAG
GTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTT
CTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGC
GGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACC
GTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTT
CAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACC
CTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAA
AAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAG
CTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCA
TCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGC
GAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCA
CTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGC
ACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAAT
CTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATAT
CATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCG
ACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGC
CTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACTCTGGCGGCTCAAAAAGAACCGCCGA
CGGCAGCGAATTCGAGCCCAAGAAGAAGAGGAAAGTCTAACCGGTCATCATCACCATCACCATTGAGT
TTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCG
TGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCG
CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTG
GGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCT
GGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGA

[0116]

AATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGC
CTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT
CGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAA
GGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGC
AAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAG
CATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT
TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCT
TTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTC
GTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAA
CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGA
TTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACT
AGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC
TTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACACTCAGTGGAACGAAAAC
TCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAA
ATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA
GTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTC
ACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA
CTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAAT
AGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTC
ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTA
GCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCA
GCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAAC
CAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATA
CCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCA
AGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATC
TTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAA
GGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGT
TATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCAC
ATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGA
CTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAG
GTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAA
GAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATT
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC
CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTC

[0117]

AATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATT

[0118] TACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC

[0119] "염기 편집 활성"은 폴리뉴클레오타이드 내에서 염기를 화학적으로 변경하는 작용을 함을 의미한다. 하나의 구현예에서, 제1 염기는 제2 염기로 전환된다. 하나의 구현예에서, 염기 편집 활성은 아데노신 또는 아데닌 데아미나제 활성, 예를 들어, A·T를 G·C로 전환시키는 활성이다. 염기 편집 활성은 또한 아데노신 또는 아데닌 데아미나제 활성, 예를 들어, A·T를 G·C로 전환시키는 활성 및 시티딘 데아미나제 활성, 예를 들어, 표적 C·G를 to T·A로 전환시키는 활성을 포함할 수 있다. 일부 구현예에서, 염기 편집 활성은 편집 효율에 의해 평가된다. 염기 편집 효율은 임의의 적합한 수단, 예를 들어, 생거 서열 분석 또는 차세대 서열 분석에 의해 측정될 수 있다. 일부 구현예에서, 염기 편집 효율은 염기 편집기에 의해 수행된 핵염기 전환을 갖는 총 서열분석 판독의 퍼센트, 예를 들어, G.C 염기쌍으로 전환된 표적 A.T 염기 쌍을 갖는 총 서열분석 판독의 퍼센트에 의해 측정된다. 일부 구현예에서, 염기 편집 효율은 염기 편집이 세포 집단에서 수행된 경우, 염기 편집기에 의해

수행된 핵염기 전환을 갖는 총 세포의 퍼센트에 의해 측정된다.

[0120] 용어 "염기 편집기 시스템"은 표적 뉴클레오타이드 서열의 핵염기를 편집하기 위한 시스템을 언급한다. 다양한 구현예에서, 염기 편집기 시스템은 (1) 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 (예를 들어, Cas9); (2) 상기 핵염기를 탈아민화하기 위한 데아미나제 도메인 (예를 들어. 아데노신 데아미나제); 및 (3) 하나 이상의 가이드 폴리뉴클레오타이드 (예를 들어, 가이드 RNA)를 포함한다. 일부 구현예에서, 상기 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 ABE8이다.

[0121] 일부 구현예에서, 염기 편집기 시스템은 하나 초과의 염기 편집 성분을 포함할 수 있다. 예를 들어, 염기 편집기 시스템은 하나 초과의 데아미나제를 포함할 수 있다. 일부 구현예에서, 염기 편집기 시스템은 하나 이상의 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.

[0122] 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.

[0123] 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 있다. 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, 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.

[0124] 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 있다. BER 성분의 저해제는 BER 저해제를 포함할 수 있다. 일부 구현예에서, BER의 저해제는 우라실 DNA 글리코실라제 저해제 (UGI)일 수 있다. 일부 구현예에서, BER의 저해제는 이노신 BER 저해제일 수 있다. 일부 구현예에서, BER의 저해제는 폴리뉴클레오타이드 프로그램 가능한 뉴클레오타이드 결합 도메인에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 BER의 저해제에 융합되거나 연결될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 데아미나제 도메인 및 BER의 저해제에 융합되거나 연결될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 BER의 저해제를 BER의 저해제와 비공유적으로 상호작용하거나 연합함에 의해 표적 뉴클레오타이드 서열에 표적화시킬 수 있다. 예를 들어, 일부 구현예에서, 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.

[0125] 일부 구현예에서, BER의 저해제는 가이드 폴리뉴클레오타이드에 의해 표적 뉴클레오타이드 서열에 표적화될 수 있다. 예를 들어, 일부 구현예에서, BER의 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 BER의 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.

[0126] 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 요구한다. 그러나, 단일 가이드 RNA ("sgRNA", 또는 단순히 "gNRA")는 crRNA 및 tracrRNA 둘 다의 양상을 단일 RNA 종으로 혼입하기 위해 가공될 수 있다. 예를 들어, 문헌 (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. Cas9 뉴클레아제 서열 및 구조는 당업자에게 널리 공지되어 있다 (참조: 예를 들어, "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), 이의 각각의 전체 내용은 본원에 참조로 포함된다). Cas9 오톨로그는 에스. 피오게네스 (*S. Pyogenes*) 및 에스. 써모필러스 (*S.*

thermophilus)를 포함하지만 이에 제한되지 않는 다양한 종에 기재되었다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본원 개시내용을 기준으로 당업자에게 자명할 것이고, 상기 Cas9 뉴클레아제 및 서열은 이의 전문이 본원에 참조로 인용되는 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 기재된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다.

[0127] 예시적인 Cas9는 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (spCas9)이고, 이의 아미노산 서열은 하기에 제공된다:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0128]

[0129] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0130] 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 (참조: 예를 들어, 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, 이의 각각의 전문의 내용은 본원에 참조로 포함된다). 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 (*S. Pyogenes*) 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 있다.

[0131] 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, 상기 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.

[0132] 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.

[0133] 일부 구현예에서, 야생형 Cas9는 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) 기원의 Cas9(NCBI 참조 서열: NC_017053.1, 다음과 같은 뉴클레오타이드 및 아미노산 서열)에 상응한다.

[0134]
ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
GTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
CTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTC
TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATC
CAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGA
TGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
AGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTG
ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGA
TACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGT
TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGT
GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGA
CTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT

TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT
TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA
TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAA
GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT
GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
AACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTA
TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAG
CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAA
GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
TTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCA
CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGA
TTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTT
GATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGA
AAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAG
GTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAA
AATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATT
AGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAG
ACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAAC
GTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAA
GTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTG
GCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGA
GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCT
ATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTT
GGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAA
AGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAA
[0135] AATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGA

GAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAA
AGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGA
AAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGAC
AAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAAC
GGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAAT
CCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATT
GACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAA
ATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTAC
AAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCAT
TATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCA
TAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAG
CAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGT
GAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTT
TAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATG
CCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA
GGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG

[0136]

EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0137]

[0138] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0139] 일부 구현예에서, 야생형 Cas9는 하기의 뉴클레오타이드 및/또는 아미노산 서열에 상응하거나 이를 포함한다:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCAT
AACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATT
CGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACT
CGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACA
AGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGT
CCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTC
AACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTG
GGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATC
CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGA
TGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCAC
AATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTG
ACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGA
CACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTAT
TTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACT
GAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGA
CTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCT
TTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTC
TACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACT
CAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAA
TCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAA
GACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCT
GGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAT
GGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACC

[0140] AACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTA

TTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCG
CCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAA
GTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGA
GATCTCCGGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGA
TAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTG
TTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCA
CCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGAT
TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTT
CTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAAC
CTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATA
TTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTG
GATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACG
CGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAG
AGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG
CAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA
ACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGA
AGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGAC
AATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGC
GAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
AACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCAT
GTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCG
GGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAAT
TCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTC
GTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTA
CAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAG
CCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAAC
GGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGA
TAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAA
AGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGT
GATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCC
TACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGA
AGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACC

[0141] AAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGC

TTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCC
CATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCA
GCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCC
TAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATA
CGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGC
ATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAG
ACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAG
CTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGA
CGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0142]

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

[0144] 일부 구현예에서, 야생형 Cas9는 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*)로부터의 Cas9 (NCBI 참조 서열: NC_002737.2 (하기와 같은 뉴클레오타이드 서열); 및 Uniprot 참조 서열에 상응한다: Q99ZW2 (하기와 같은 아미노산 서열).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
GTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
CTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTC
TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATC
CAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGA
TGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
AGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTG
ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGA
TACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGT
TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACT
GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGA
CTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT
TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA
TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAA
GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT
GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
AACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTA
TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAG
CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAA
GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
[0145] TTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCA

CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATA
TTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTT
GATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACG
TGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAG
AAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTG
CAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGA
ATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTA
AAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGAT
AACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGC
CAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTG
AACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCAT
GTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCG
AGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAAT
TCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTC
GTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTA
TAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCG
CAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAAT
GGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGA
TAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCA
AGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCG
GACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCC
AACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAA
AATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCG
ATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACC
TAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAAT
TACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGT
CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCA
GCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTT
TAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATA
CGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
[0146] TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAG

ATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAG
CTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0147]

[0148] (서열번호 1. 한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0149] 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.

[0150] 일부 구현예에서, Cas9는 나이세리아 메니기티디스 (*Neisseria menigitidis*) Cas9 (NmeCas9) 또는 이의 변이체이다. 일부 구현예에서, NmeCas9는 NNNNGAYW PAM에 대해 특이성을 갖고, 여기서, Y는 C 또는 T이고, W는 A 또는 T이다. 일부 구현예에서, NmeCas9는 NNNNGYTT PAM에 대해 특이성을 갖고, 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 갖고, 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. 일부 구현예에서, Nme1, Nme2 또는 Nme3에 대해 PAM-상호작용 도메인은 각각 N_4GAT, N_4CC, 및 N_4CAAA이다. 추가의 NmeCas9 특성 및 PAM 서열은 문헌 (참조: Edraki et al., A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for *In Vivo* Genome Editing, *Mol. Cell.* (2019) 73(4): 714-726)에 기재되어 있고, 이는 본원에 이의 전문이 참조로 포함된다.

[0151] 예시적인 나이세리아 메닌기티디스 (*Neisseria meningitidis*) 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
```

[0152]

[0153] 또 다른 예시적인 나이세리아 메닌기티디스 (*Neisseria meningitidis*) 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
1081 vr
```

[0154]

[0155] 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 Cas9에 D10A 및 H840A 돌연변이 또는 상응하는 돌연변이를 포함한다. 일부 구현예에서, dCas9는 dCas9 (D10A 및 H840A)의 아미노산 서열을 포함한다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0156]

[0157] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0158] 일부 구현예에서, Cas9 도메인은 D10A 돌연변이를 포함하고, 위치 840에서 잔기는 상기 제공된 아미노산 서열에서, 또는 본문에 제공된 임의의 아미노산 서열에서 상응하는 위치에 히스티딘을 유지한다.

[0159] 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 아미노산 만큼 더 짧거나 더 긴 아미노산 서열을 갖는다.

[0160] 일부 구현예에서, 본원에 제공된 바와 같은 Cas9 융합 단백질은 Cas9 단백질의 전장 아미노산 서열, 예를 들어, 본원에 제공된 Cas9 서열 중 하나를 포함한다. 다른 구현예에서, 그러나, 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 자명하다.

[0161] 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.

[0162] 예시적인 촉매적 불활성 Cas9 (dCas9)의 아미노산 서열은 다음과 같다:

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATR
LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDE
VAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ
LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0163]

[0164] 예시적인 촉매적 Cas9 닉카제 (nCas9)의 아미노산 서열은 다음과 같다:

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATR
LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDE
VAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ
LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0165]

[0166] 예시적인 촉매적 활성 Cas9의 아미노산 서열은 다음과 같다:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATR
LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDE
VAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ
LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
KSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD.

[0167]

[0168] 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 인용되는 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 컴팩트한 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 범위내에 있는 것으로 인지되어야 한다.

[0169] 일부 구현예에서, Cas9는 변경된 PAM 서열에 대한 특이성을 갖는 Cas9 변이체이다. 일부 구현예에서, 추가의 Cas9 변이체 및 PAM 서열은 문헌 (참조: sequences are described in Miller et al., Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat Biotechnol* (2020). doi.org/10.1038/s41587-020-0412-8)에 기재되어 있고, 이의 전문은 본원에 참조로 포함된다. 일부 구현예에서, Cas9 변이체는 어떠한 특정 PAM 요건을 갖지 않는다. 일부 구현예에서, Cas9 변이체, 예를 들어, SpCas9 변이체는 NRNH PAM에 대해 특이성을 갖고, 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 하기 참조 서열에 상대적으로 넘버링된 바와 같은 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.

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0170]

[0171] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0172] 일부 구현예에서, SpCas9 변이체는 상기 참조 서열에 상대적으로 넘버링된 바와 같은 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 상기 참조 서열에 상대적으로 넘버링된 바와 같은 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 상기 참조 서열에 상대적으로 넘버링된 바와 같은 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 상기 참조 서열에 상대적으로 넘버링된 바와 같은 위치 1114, 1127, 1135, 1180, 1207, 1219, 1234, 1286, 1301, 1332, 1335, 1337, 1338 또는 1349에서 아미노산 치환을 포함한다. SpCas9 변이체의 예시적인 아미노산 치환 및 PAM 특이성은 표 A-D 및 표 8에 나타낸다.

[0173] [표 A]

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

[0174]

[0175] [표 B]

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

[0176]

[0177] [표 C]

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	

[0178]

[0179] [표 D]

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

[0180]

[0181] 특정 구현예에서, 본 발명의 방법에 유용한 napDNAbps는 당업계에 공지되고 예를 들어, 문헌 (참조: Oakes *et al.*, Cell 176, 254-267, 2019)에 기재된 환형 퍼뮤턴트를 포함한다. 예시적인 환형의 퍼뮤턴트는 하기에 제시되어 있고, 여기서, 굵게 표시된 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭한다.

[0182] **CP5** (MSP "NGC=NGG와 같이 돌연변이 정규 Cas9를 갖는 팜 변이체", PID=단백질 상호작용 도메인 및 "D10A" 닉카제와 함께):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYR
STKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAI
GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV
LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0183]

[0184] 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.

[0185] 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.

[0186] Cas12b/C2c1 (uniprot.org/uniprot/T0D7A2#2)

[0187] sp|T0D7A2|C2C1_ALIAG CRISPR-연합된 엔도-뉴클레아제 C2c1 OS = 알리사이클로바실러스 액시도-테레스트리스 (*Alicyclobacillus acido-terrestris*) (균주 ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECD
KTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKF
LSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFG
LKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQ
KNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLA
PDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREV
DDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRG
ARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSE
GLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLL
KLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAAN
HMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPK
IRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKE
DRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLM
QWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPW
WLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDF
DISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQE
KLSEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSR
VPLQDSACENTGDI

[0188]

[0189] CasX (uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53)

[0190] >tr|F0NN87|F0NN87_SULIH CRISPR-연합된 Casx 단백질 OS = 설폴로부스 아슬란디쿠스 (*Sulfolobus islandicus*) (균주 HVE10/4)GN= SiH_0402 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGK
AKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAP
SFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTP
TRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTIN
GGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLY
FANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0191]

[0192] >tr|F0NH53|F0NH53_SULIR CRISPR 연합된 단백질, Casx OS = 설폴로부스 아이슬란디쿠스 (*Sulfolobus islandicus*) (균주 REY15A) GN=SiRe_0771 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGK
AKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAP
SFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTP
TRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTIN
GGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYF
ANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0193]

[0194] 델타프로테오박테리아 CasX

MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAA
NNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTA
GFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLG
KFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIII
EHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDAYNEVIARVRMWVNLNLWQ
KLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAEK
RNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERID
KKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWYG
DLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIRF
TDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLETG
LIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVIA
LTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNLA
DDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGLT
SKTYLSKTLAQYTSKTCSNCGFTITYADMDVMLVRLKKTSDGWATTLNNKELKAEYQITYYN
RYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCGH
EVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0195]

[0196] CasY(ncbi.nlm.nih.gov/protein/APG80656.1)

[0197] >APG80656.1 CRISPR-연합된 단백질 CasY [배양되지 않은 파르쿠박테리아 그룹 박테리움 (Parcubacteria group bacterium)]

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGL
YGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTL
KGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKN
AKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFN
KLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELKKAMMDITDAW
RGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLK
GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSD
GRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKL
VPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQK
IFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTEN
IAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVE
NGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAEL
LYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELT
RTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHR
PKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQ
RYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTK
IARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDAD
KNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLID
AIKDFMRPPIFDENDTPFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIAL
LRYVKEEKKVEDYFERFRKLKN IKVLGQMKKI

[0198]

[0199] 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 그

룹은 그룹 내 아미노산이 우선적으로 서로 교환하는 경우 및 따라서 전체 단백질 구조에 대한 이들의 영향에서 서로 가장 유사한 경우 한정될 수 있다 (참조: Schulz, G. E. and Schirmer, R. H., supra). 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 유리된-OH가 유지될 수 있고; 글루타민의 아스파라긴으로의 아미노산 치환을 포함하여 유리된 $-NH_2$가 유지될 수 있다.

[0200] 본원에서 상호교환적으로 사용된 바와 같은 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.

[0201] 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 데옥시아데노신이 각각 이노신으로 또는 데옥시이노신으로의 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, 진화된 아데노신 데아미나제)는 임의의 유기체, 예를 들어, 세균으로부터 기원할 수 있다. 일부 구현예에서, 아데노신 데아미나제는 세균, 예를 들어, 에스케리치아 콜리 (*Escherichia. coli*), 스타필로코커스 아우레우스 (*Staphylococcus. aureus*), 살모넬라 티피무리움 (*Salmonella. typhimurium*), 슈와넬라 푸트레파시엔스 (*Shewanella. putrefaciens*), 해모필러스 인플루엔자 (*Haemophilus influenzae*), 또는 콜로박터 크레슨투스 (*Caulobacter. crescentus*)로부터 기원한다.

[0202] 일부 구현예에서, 아데노신 데아미나제는 TadA 데아미나제이다. 일부 구현예에서, TadA 데아미나제는 TadA 변이체이다. 일부 구현예에서, TadA 변이체는 TadA*8이다. 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 PCT/2017/045381 (WO 2018/027078) 및 PCT/US2016/058344 (WO2017/070632)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); 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.

[0203] "검출한다"는 검출될 분석물의 존재, 부재 또는 양을 동정하는 것을 언급한다. 하나의 구현예에서, 폴리뉴클레오타이드 또는 폴리펩타이드에서 서열 변경이 검출된다. 또 다른 구현예에서, 삽입-결실(indel)의 존재가 검출된다.

[0204] "검출 가능한 표지"는 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, ELISA에서 통상적으로 사용되는 바와 같이), 비오틴, 디곡시게닌, 또는 합텐을 포함한다.

[0205] "질환"은 세포, 조직 또는 기관의 정상 기능을 손상시키거나 방해하는 임의의 병태 또는 장애를 의미한다. 하나의 구현예에서, 질환은 A1AD이다.

[0206] 본원에 사용된 바와 같은 용어 "유효량"은 목적하는 생물학적 반응을 유발하기에 충분한 생물학적 활성제의 양을 언급한다. 특정 구현예에서, 유효량은 세포에서 A1AT 돌연변이를 변경하여 치료학적 효과를 성취하기에 충분한 염기 편집기 시스템 (예를 들어, 프로그래밍 가능한 DNA 결합 단백질, 핵염기 편집기 및 gRNA를 포함하는 융합 단백질)의 양이다. 상기 치료학적 효과는 조직 또는 기관의 모든 세포에서 A1AD를 변경하기에 충분할 필요가 없지만 대상체, 조직 또는 기관에 존재하는 세포의 약 1%, 5%, 10%, 25%, 50%, 75% 이상에서만 A1AD를 변경하기에 충분하다. 하나의 구현예에서, 유효량은 A1AD의 하나 이상의 증상을 개선하기에 충분하다. 질환의 치료학적 치료를 위해 본 발명을 수행하기 위해 사용되는 활성제(들)의 유효량은 투여 방식, 대상체의 연령, 체중 및 일반 건강에 따라 다양하다. 궁극적으로, 담당 의사 또는 수의사는 적당한 양 및 투여 용법을 결정한다. 상기 양은 "유효"량으로서 언급된다. 하나의 구현예에서, 유효량은 세포 (예를 들어, 시험관내 세포 또는 생체내 세포)에서 관심 대상의 유전자에 변경을 도입하기에 충분한 본 발명의 염기 편집기 (예를 들어, 프로그래밍 가능한 DNA 결합 단백질, 핵염기 편집기 및 gRNA를 포함하는 융합 단백질)의 양이다. 하나의 구현예에서, 유효량은 치료학적 효과를 성취하기 위해 (예를 들어, 질환 또는 이의 증상 또는 병태를 감소시키거나 제어하기 위해) 요구되는 염기 편집기의 양이다.

[0207] "단편"은 폴리펩타이드 또는 핵산 분자 부분을 의미한다. 상기 부분은 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.

[0208] "가이드 RNA" 또는 "gRNA"는 표적 서열에 특이적일 수 있고 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 단백질 (예를 들어, Cas9 또는 Cpf1)과 복합체를 형성할 수 있는 폴리뉴클레오타이드를 의미한다. 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 가이드 RNA (gRNA)이다. gRNA는 2개 이상의 RNA의 복합체 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 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 인용된 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 단백질에 결합하고 2개 이상의 특유한 영역에서 표적 핵산에 결합한다. gRNA는 표적 부위와 상보체를 형성하는 뉴클레오티드 서열을 포함하고, 이는 뉴클레아제/RNA 복합체의 상기 표적 부위로의 결합을 매개하여 뉴클레아제:RNA 복합체의 서열 특이성을 제공한다. 당업자에 의해 인지되는 바와 같이, RNA 폴리뉴클레오타이드 서열, 예를 들어, gRNA 서열은 DNA 폴리뉴클레오타이드 서열에 포함되는 핵염기 티민 (T) 대신 핵염기 우라실(U), 피리미딘 유도체를 포함한다. RNA에서, 우라실 염기는 아데닌과 쌍을 형성하고 DNA 전사 동안에 티민을 대체한다.

[0209] "하이브리드화"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.

[0210] 용어 "염기 복구의 저해제" 또는 "IBR"은 핵산 복구 효소, 예를 들어, 염기 절제 복구 (BER) 효소의 활성을 저해할 수 있는 단백질을 언급한다. 일부 구현예에서, IBR은 이노신 염기 절제 복구의 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 Endo V 또는 hAAG의 저해제이다. 일부 구현예에서, 염기 복구 저해제는 촉매 불활성 EndoV 또는 촉매 불활성 hAAG이다.

[0211] 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.

[0212] "증가시킨다"는 적어도 10%, 25%, 50%, 75%, 또는 100%의 양성 변경을 의미한다.

[0213] "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 공정은 본원에서 "단백질 스플라이싱" 또는 "인테인-매개된 단백질 스플라이싱"으로 호칭된다. 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 서브뉴닛인 DnaE는 2개의 별도의 유전자 dnaE-n 및 dnaE-c에 의해 암호화되어 있다. 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.

[0214] 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 보고되었다 (예를 들어, 문헌 (참조: Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5, 본원에 참조로 인용됨). 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, 문헌 (본원에 참조로 인용된 미국 특허 제8,394,604호)).

[0215] 인테인의 예시적인 뉴클레오타이드 및 아미노산 서열이 제공된다.

[0216] **DnaE 인테인-N DNA:**

TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGAT
TGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTC
AGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGAT
GGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCC

[0217] TATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT

[0218] **DnaE 인테인-N 단백질:**

CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDR

[0219] GEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

[0220] **DnaE 인테인-C DNA:**

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGA

[0221] TATTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT

[0222] **인테인-C:** MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN

[0223] **Cfa-N DNA:**

TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGAT
TGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACAC
AGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGAT
GGAAGCATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCC

[0224] AATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTGCCA

[0225] **Cfa-N 단백질:**

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED

[0226] GSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP

[0227] **Cfa-C DNA:**

ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGAT
AATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCACA

[0228] ACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC

[0229] **Cfa-C 단백질:**

[0230] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN

[0231] 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 인용된 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.

[0232] 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 단리 보다 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, 본 발명의 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.

[0233] "단리된 폴리뉴클레오타이드"는 본 발명의 핵산 분자가 유래된 천연적으로 발생하는 유기체의 게놈에서 유전자를 플랭킹하는 유전자가 제거된 핵산 (예를 들어, DNA)을 의미한다. 상기 용어는 따라서, 예를 들어, 벡터로; 자가 복제 플라스미드 또는 바이러스; 또는 원핵세포 또는 진핵세포의 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.

[0234] "단리된 폴리펩타이드"는 천연적으로 여기에 수반되는 성분으로부터 분리된 본 발명의 폴리펩타이드를 의미한다. 전형적으로, 폴리펩타이드는 단백질 및 천연적으로 연합된 천연적으로 발생하는 유기 분자로부터 적어도 60 중량%로 제거된 경우 단리된다. 바람직하게, 상기 제제는 적어도 75 중량%, 보다 바람직하게 적어도 90 중량%, 및 가장 바람직하게 적어도 99 중량%의 본 발명의 폴리펩타이드이다. 본 발명의 단리된 폴리펩타이

드는 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.

[0235] 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 도메인과 같은, 융합 단백질의 2개의 도메인 ((예를 들어, PCT/US19/44935 에 기재된 바와 같은, 아데노신 데아미나제, 또는 아데노신 데아미나제 및 시티딘 데아미나제)을 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.

[0236] 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. 보다 길거나 짧은 링커가 또한 고려될 수 있다.

[0237] 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, 상기 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. 보다 길고 또는 보다 짧은 링커가 또한 고려된다.

[0238] 일부 구현예에서, 핵염기 편집기의 도메인은 SGGSSGSETPGTSESATPESSGGS, SGGSSGGSSGSETPGTSESATPESSGGSSGGS,

또는 GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS의 아미노산 서열을 포함하는 링커를 통해 융합된다. 일부 구현예에서, 핵염기 편집기의 도메인은 또한 XTEN 링커로서 언급될 수 있는 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다. 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 이들 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.

[0239] 일부 구현예에서, 링커는 24개 아미노산 길이이다. 일부 구현예에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPES를 포함한다. 일부 구현예에서, 링커는 40개 아미노산 길이이다. 일부 구현예에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS를 포함한다. 일부 구현예에서, 링커는 64개 아미노산 길이이다. 일부 구현예에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGS SGGS를 포함한다. 일부 구현예에서, 링커는 92개 아미노산 길이이다. 일부 구현예에서, 링커는 아미노산 서열 PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS를 포함한다.

[0240] "마커"는 질환 또는 장애와 연관된 발현 수준 또는 활성이 변경된 임의의 단백질 또는 폴리뉴클레오타이드를 의미한다.

[0241] 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 기재된 염기 편집기는 의도되지 않은 점 돌연변이와 같은, 상당수의 의도되지 않은 돌연변이 없이 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.

[0242] 일반적으로, 서열 (예를 들어, 본원에 기재된 바와 같은 아미노산 서열)에서 만들어지거나 동정된 돌연변이는 참조 (또는 야생형) 서열, 즉, 돌연변이를 함유하지 않는 서열에 상대적으로 넘버링된다. 당업자는 참조 서열에 상대적으로 아미노산 및 핵산 서열에서 돌연변이의 위치를 결정하는 법을 용이하게 이해할 것이다.

[0243] 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 상기 기능성 변이체의 생물학적 활성이 야생형 단백질과 비교하여 증가될 수 있다.

[0244] 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 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 KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRK, PKKKRKV, 또는 MDSLLMNRRKFLYQFKNVRWAKGRRETYLC의 아미노산 서열을 포함한다.

[0245] 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. 뉴클레오타

이드 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.

[0246] 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 또는 가이드 폴리뉴클레오타이드 (예를 들어, gRNA)와 같은 핵산 (예를 들어, DNA 또는 RNA)과 연합된 단백질을 언급할 수 있다. 일부 구현예에서, 상기 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이다. 일부 구현예에서, 상기 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 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, 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 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.

[0247] 본원에 상호교환적으로 사용되는 용어 "핵염기", "질소성 염기" 또는 "염기"는 뉴클레오타이드의 성분인 뉴클레오사이드를 형성하는 질소-함유 생물학적 화합물을 언급한다. 핵염기가 염기쌍을 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 카노니칼

로 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.

[0248] 본원에 사용된 바와 같은 용어 "핵염기 편집 도메인" 또는 "핵염기 편집 단백질"은 RNA 또는 DNA에서 핵염기 변형, 예를 들어, 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 하나 초과의 데아미나제 도메인 (예를 들어, PCT/US19/44935에 기재된 바와 같이 아데노신 데아미나제, 또는 아데노신 데아미나제 및 시티딘 또는 시토신 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 진화된 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.

[0249] 본원에 사용된 바와 같이, "제제를 수득하는"에서와 같이 "수득하는"은 합성, 구매, 생성, 제조 또는 다르게 제제를 획득하는 것을 포함한다.

[0250] 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 확률 보다 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 남성이고/이거나 여성일 수 있다.

[0251] "이를 필요로 하는 환자" 또는 "이를 필요로 하는 대상체"는 본원에서 질환 또는 장애를 앓는 것으로 진단되거나, 질환 또는 장애를 갖거나, 질환 또는 장애에 민감하거나 갖는 것으로 미리 결정되거나 가질 것으로 의심되는 환자 또는 개체로서 언급된다.

[0252] 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.

[0253] 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 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 상이한 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.

[0254] 본원에 기재된 폴리펩타이드 및 단백질 (기능성 부분 및 이의 기능성 변이체를 포함하는)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-연결된), 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.

[0255] 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.

[0256] "감소시킨다"는 적어도 10%, 25%, 50%, 75%, 또는 100%의 음성 변경을 의미한다.

[0257] "참조"는 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 비처리된 세포이다.

[0258] "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 아미노산이다. 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 정수이다. 일부 구현예에서, 참조 서열은 관심 대상의 단백질의 야생형 서열이다. 다른 구현예에서, 참조 서열은 야생형 단백질을 암호화하는 폴리뉴클레오타이드 서열이다.

[0259] 용어 "RNA-프로그래밍 가능한 뉴클레아제" 및 "RNA-가이드된 뉴클레아제"는 절단을 위한 표적이 아닌 하나 이상의 RNA(들)와 함께 (예를 들어, 와 결합하거나 연합된) 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 (예

를 들어, Cas9 복합체의 표적으로의 결합을 지시하는) 도메인; 및 (2) Cas9 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 인용된 문헌 (참조: Jinek et ah, 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.

[0260] 일부 구현예에서, RNA-프로그래밍 가능한 뉴클레아제는 (CRISPR-연합된 시스템) Cas9 엔도뉴클레아제, 예를 들어, 스트렙토코커스 피오게네스 (Streptococcus pyogenes)로부터의 Cas9 (Csnl)이다 (참조: 예를 들어, "Complete genome sequence of an Ml strain of Streptococcus pyogenes." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C, Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011)).

[0261] 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 (참조: 예를 들어, Cong, L. et ah, Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali, P. et ah, RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang, W.Y. et ah, Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek, M. et ah, RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo, J.E. et ah, Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic acids research (2013); Jiang, W. et ah RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239 (2013); 이들 각각의 전체 내용은 본원에 참조로 포함된다).

[0262] 용어 "단일 뉴클레오타이드 다형성(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 영향받은 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 변화는 또한 단일 뉴클레오타이드 변경으로 호칭될 수 있다.

[0263] "특이적으로 결합한다"는 핵산 분자, 폴리펩타이드, 또는 이의 복합체 (예를 들어, 핵산 프로그래밍 가능한 DNA 결합 단백질, 가이드 핵산), 본 발명의 폴리펩타이드 및/또는 핵산 분자를 인지하고 결합하는 화합물 또는 분자

를 의미하지만, 이는 샘플, 예를 들어, 생물학적 샘플에서 다른 분자를 실질적으로 인지하지 못하고 결합하지 않는다.

[0264] 본 발명의 방법에 유용한 핵산 분자는 본 발명의 폴리펩타이드 또는 이의 단편을 암호화하는 임의의 핵산 분자를 포함한다. 상기 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 전형적으로 실질적인 동일성을 나타낸다. 내인성 서열과 "실질적인 동일성"을 갖는 폴리뉴클레오타이드는 전형적으로 이중 가닥 핵산 분자의 적어도 하나의 가닥과 하이브리드화할 수 있다. 본 발명의 방법에 유용한 핵산 분자는 본 발명의 폴리펩타이드 또는 이의 단편을 암호화하는 임의의 핵산 분자를 포함한다. 상기 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 전형적으로 실질적인 동일성을 나타낸다. 내인성 서열과 "실질적인 동일성"을 갖는 폴리뉴클레오타이드는 전형적으로 이중 가닥 핵산 분자의 적어도 하나의 가닥과 하이브리드화할 수 있다. "하이브리드화한다"는 다양한 엄중 조건하에서 상보적인 폴리뉴클레오타이드 서열 (예를 들어, 본원에 기재된 유전자) 또는 이의 일부 간에 쌍을 이루어 이중 가닥 분자를 형성함을 의미한다. (참조: 예를 들어, Wahl, G. M. and S. L. BERger (1987) Methods Enzymol. 152:399; Kimmel, A. R. (1987) Methods Enzymol. 152:507).

[0265] 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 시트레이트이다. 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 포함한다. 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. 하나의 구현예에서, 하이브리드화는 750 mM NaCl, 75 mM 삼나트륨 시트레이트 및 1% SDS에서 30℃에서 수행한다. 또 다른 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 수행한다. 또 다른 구현예에서, 하이브리드화는 250 mM NaCl, 25 mM 삼나트륨 시트레이트, 1% SDS, 50% 포름아미드, 및 200 μg/ml ssDNA에서 42℃에서 수행한다. 이들 조건에 대한 유용한 변화는 당업자에게 용이하게 자명하다.

[0266] 대부분의 적용을 위해, 하이브리드화에 이어서 세척 단계는 또한 엄중도에서 다양하다. 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 시트레이트이다. 세척 단계를 위한 엄중 온도 조건은 통상적으로 적어도 약 25℃, 보다 바람직하게 적어도 약 42℃, 및 보다 더 바람직하게 적어도 약 68℃의 온도를 포함한다. 하나의 구현예에서, 세척 단계는 30 mM NaCl, 3 mM 삼나트륨 시트레이트 및 0.1% SDS에서 25℃에서 수행한다. 보다 바람직한 구현예에서, 세척 단계는 15 mM NaCl, 1.5 mM 삼나트륨 시트레이트 및 0.1% SDS에서 42℃에서 수행한다. 보다 바람직한 구현예에서, 세척 단계는 15 mM NaCl, 1.5 mM 삼나트륨 시트레이트 및 0.1% SDS에서 68℃에서 수행한다. 이들 조건에 대한 추가의 변화는 당업자에게 용이하게 자명하다. 하이브리드화 기술은 당업자에게 널리 공지되어 있고, 예를 들아, 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.

[0267] "스플릿"은 2개 이상의 단편으로 나누어짐을 의미한다.

[0268] "스플릿 Cas9 단백질" 또는 "스플릿 Cas9"은 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 인용된, 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 위

치에서 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.

[0269] 다른 구현예에서, Cas9 단백질의 N-말단 부분은 아미노산 1-573 또는 1-637 에스. 피오게네스 (S. pyogenes) Cas9 야생형(SpCas9) (NCBI 참조 서열: NC_002737.2, 유니프롯 참조 서열: Q99ZW2), 또는 이의 상응하는 위치/돌연변이를 포함하고, Cas9 단백질의 C-말단 부분은 SpCas9 야생형의 아미노산 574-1368 또는 638-1368의 부분을 포함한다.

[0270] 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 651-1368의 어느 하나의 부분을 포함할 수 있다. 일부 구현예에서, 스플릿 Cas9 단백질의 C-말단 부분은 SpCas9의 아미노산 574-1368 또는 638-1368 부분을 포함한다.

[0271] "*서핀1A* 폴리뉴클레오타이드"는 A1AT 단백질 또는 이의 단편을 암호화하는 핵산 분자를 의미한다. 예시적인 서핀1A 폴리뉴클레오타이드의 서열은 NCBI 승인 번호. NM_000295에서 가용하고, 하기에 제공된다:

```
 1 acaatgactc ctttcggtaa gtgcagtgga agctgtacac tgcccaggca aagcgtccgg
61 gcagcgtagg cgggcgactc agatcccagc cagtggactt agcccctgtt tgctcctccg
121 ataactgggg tgaccttggt taatattcac cagcagcctc ccccgttgcc cctctggatc
181 cactgcttaa atacggacga ggacagggcc ctgtctcctc agcttcaggc accaccactg
241 acctgggaca gtgaatcgac aatgccgtct tctgtctcgt ggggcatcct cctgctggca
301 ggcctgtgct gcctggtccc tgtctccctg gctgaggatc cccagggaga tgctgcccag
361 aagacagata catcccacca tgatcaggat cacccaacct tcaacaagat caccccccaac
421 ctggctgagt tcgccttcag cctataccgc cagctggcac accagtccaa cagcaccaat
481 atcttcttct ccccagtgag catcgctaca gcctttgcaa tgctctccct ggggaccaag
```

[0272]

```
541 gctgacactc acgatgaaat cctggagggc ctgaatttca acctcacgga gattccggag
601 gctcagatcc atgaaggctt ccaggaactc ctccgtaccc tcaaccagcc agacagccag
661 ctccagctga ccaccggcaa tggcctgttc ctcagcgagg gcctgaagct agtggataag
721 tttttggagg atgttaaaaa gttgtaccac tcagaagcct tcactgtcaa cttcggggac
781 accgaagagg ccaagaaaca gatcaacgat tacgtggaga agggtactca agggaaaatt
841 gtggatttgg tcaaggagct tgacagagac acagtttttg ctctggtgaa ttacatcttc
901 tttaaaggca aatgggagag acccctttgaa gtcaaggaca ccgaggaaga ggacttccac
961 gtggaccagg tgaccaccgt gaaggtgcct atgatgaagc gtttaggcat gtttaacatc
1021 cagcactgta agaagctgtc cagctgggtg ctgctgatga aatacctggg aatgccacc
1081 gccatcttct tcctgcctga tgaggggaaa ctacagcacc tggaaaatga ctcacccac
1141 gatatcatca ccaagttcct ggaaaatgaa gacagaaggt ctgccagctt catttaccc
1201 aaactgtcca ttactggaac ctatgatctg aagagcgtcc tgggtcaact ggcatcact
1261 aaggtcttca gcaatggggc tgacctctcc ggggtcacag aggaggcacc ctgaagctc
1321 tccaaggccg tgcataaggc tgtgctgacc atcgacgaga aagggactga gctgctggg
1381 gccatgtttt tagaggccat acccatgtct atccccccgg aggtcaagtt aacaaaccc
1441 tttgtcttct taatgattga acaaaatacc aagtctcccc tcttcatggg aaagtggtg
1501 aatcccaccc aaaaataact gcctctcgct cctcaacccc tcccctccat cctggcccc
1561 ctccctggat gacattaaag aagggttgag ctggtccctg cctgcatgtg ctgtaaatc
1621 cctcccatgt tttctctgag tctccctttg cctgctgagg ctgtatgtgg ctccaggta
1681 acagtgctgt cttcgggccc cctgaactgt gttcatggag catctggctg gtaggcaca
1741 tgctgggctt gaatccaggg gggactgaat cctcagctta cggacctggg ccatctgtt
1801 tctggagggc tccagtcttc cttgtcctgt cttggagtcc ccaagaagga tcacagggg
1861 aggaaccaga taccagccat gaccccaggc tccaccaagc atcttcatgt cccctgctc
1921 atcccccact ccccccccacc cagagttgct catcctgcca gggctggctg gcccacccc
1981 aaggctgccc tcctgggggc cccagaactg cctgatcgtg ccgtggccca ttttgtggc
2041 atctgcagca acacaagaga gaggacaatg tcctcctctt gacccgctgt acctaacca
2101 gactcgggcc ctgcacctct caggcacttc tggaaaatga ctgaggcaga tcttcctga
2161 agcccattct ccatggggca acaaggacac ctattctgtc cttgtccttc atcgctgcc
2221 ccagaaagcc tcacatatct ccgtttagaa tcaggtccct tctccccaga gaagaggag
2281 ggtctctgct ttgttttctc tatctcctcc tcagacttga ccaggcccag aggccccag
2341 aagaccatta ccctatatcc cttctcctcc ctagtcacat ggccataggc tgctgatgg
2401 ctcaggaagg ccattgcaag gactcctcag ctatgggaga ggaagcacat acccattga
2461 cccccgcaac ccctcccttt cctcctctga gtcccgactg gggccacatg agcctgact
2521 tctttgtgcc tgttgctgtc cctgcagtct tcagagggcc accgcagctc agtgccacg
2581 gcaggaggct gttcctgaat agcccctgtg gtaagggcca ggagagtcct ccatcctcc
2641 aaggccctgc taaaggacac agcagccagg aagtcccctg ggcccctagc gaaggacag
2701 cctgctccct ccgtctctac caggaatggc cttgtcctat ggaaggcact ccccatccc
2761 aaactaatct aggaatcact gtctaaccac tcactgtcat gaatgtgtac taaaggatg
2821 aggttgagtc ataccaaata gtgatttcga tagttcaaaa tggtgaaatt gcaattcta
2881 catgattcag tctaatcaat ggataccgac tgtttcccac acaagtctcc gttctctta
2941 agcttactca ctgacagcct ttcactctcc acaaatacat taaagatatg ccatcacca
```

[0273]

```
3001 agccccctag gatgacacca gacctgagag tctgaagacc tggatccaag tctgacttt
3061 tccccctgac agctgtgtga ccttcgtgaa gtcgccaaac ctctctgagc ccagtcatt
3121 gctagtaaga cctgcctttg agttggtatg atgttcaagt tagataacaa atgtttata
3181 cccattagaa cagagaataa atagaactac atttcttgca
```

[0274]

[0275] PAM 서열은 강조되고, 아데닌 염기 편집 후 올바른 서열을 나타낸다.

[0276] "대상체"는 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 포함하나 이에 국한되지 않는다.

[0277] "실질적으로 동일한"은 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. 하나의 구현예에서, 상기 서열은 비교를 위해 사용되는 아미노산 수준 또는 핵산에서

적어도 60%, 80% 또는 85%, 90%, 95% 또는 심지어 99% 동일하다.

[0278] 서열 동일성은 전형적으로 서열 분석 소프트웨어를 사용하여 측정된다 (참조: 예를 들어, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). 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} 내지 e^{-100}이고 이는 밀접하게 관련된 서열을 지적한다.

[0279] COBALT는 예를 들어, 하기의 파라미터와 함께 사용된다:

[0280] a) 정렬 파라미터: 갭 페널티-11,-1 및 종료-갭 페널티-5,-1,

[0281] b) CDD 파라미터: CDD 파라미터에 RPS BLAST를 사용; Blast E-값 0.003; 보존된 컬럼을 발견하고 CDD 파라미터에서 재계산, 및

[0282] c) 조회 클러스터링 파라미터: CDD 파라미터에 조회 클러스터를 사용; 워드 크기 4; 최대 클러스터 거리 0.8; 알파벳 규칙.

[0283] EMBOSS 바늘은 예를 들어, 하기의 파라미터와 함께 사용한다:

[0284] a) 매트릭스: BLOSUM62;

[0285] b) 갭 개방: 10;

[0286] c) 갭 연장: 0.5;

[0287] d) 아웃풋 포맷: 쌍;

[0288] e) 종료 갭 페널티: 거짓;

[0289] f) 종료 갭 개방: 10; 및

[0290] g) 종료 갭 연장: 0.5.

[0291] 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.

[0292] 본원에 사용된 바와 같은 용어 "치료한다", "치료하는", "치료" 등은 장애 및/또는 이와 연관된 증상을 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, 상기 효과는 치료학적이고, 제한 없이, 상기 효과는 질환 및/또는 상기 질환에 기인할 수 있는 부작용 증상을 부분적으로 또는 완전하게 감소시키거나, 감쇠시키거나, 폐지시키거나, 약화시키거나, 완화시키거나, 감소시키거나, 이를 치유한다. 일부 구현예에서, 상기 효과는 예방적이고, 즉, 상기 효과는 질환 또는 병태의 발병 또는 재발로부터 보호하거나 이를 예방한다. 이를 위해, 본원에 기재된 방법은 본원에 기재된 바와 같이 치료학적 유효량의 조성물을 투여함을 포함한다. 하나의 구현예에서, 질환은 알파-1 항트립신 결핍증(A1AD)이다.

[0293] "우라실 글리코실라제 저해제" 또는 "UGI"는 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:

[0294] >sp|P14739|UNGI_BPPB2 우라실-DNA 글리코실라제 저해제

[0295] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD APEYKPWALVIQDSNGENKIKML.

[0296] 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.

[0297] 본원에 제공된 임의의 조성물 또는 방법은 본원에 제공된 임의의 다른 조성물 및 방법의 하나 이상과 조합될 수 있다.

[0298] 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 필수적이 아님을 의미한다.

[0299] 본원에서 변수의 임의의 정의에서 화학적 그룹의 목록의 언급은 단일 그룹 또는 열거된 그룹의 조합으로서 해당 변수의 정의를 포함한다. 본원의 변수 또는 양상에 대한 구현예의 언급은 임의의 단일 구현예 또는 임의의 다른 구현예 또는 이의 일부와 조합된 구현예를 포함한다.

[0300] 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 가닥을 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. 실제로, 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.

도면의 간단한 설명

[0301] **도 1a - 1c**는 플라스미드를 도시한다. **도 1a**는 TadA7.10-dCas9 염기 편집기를 암호화하는 발현 벡터이다. **도 1b**는 클로람페니콜 내성 (CamR) 및 스트렙토마이신 내성 (SpectR)을 부여하는 단백질을 암호화하는 핵산 분자를 포함하는 플라스미드이다. 상기 플라스미드는 또한 2개의 점 돌연변이에 의해 불능화된 가나마이신 내성 유전자를 포함한다. **도 1c**는 클로람페니콜 내성 (CamR) 및 스트렙토마이신 내성 (SpectR)을 부여하는 단백질을 암호화하는 핵산 분자를 포함하는 플라스미드이다. 상기 플라스미드는 또한 2개의 점 돌연변이에 의해 불능화된 가나마이신 내성 유전자를 포함한다.

도 2는 결손 가나마이신 내성 유전자를 포함하는, **도 1a-c**에 도시된 발현 벡터가 형질도입된 세균 콜로니의 이미지이다. 벡터는 오류 성향 PCR을 사용하여 생성된 ABE7.10 변이체를 함유하였다. 이들 "진화된" ABE7.10 변이체를 발현하는 세균 세포는 증가하는 농도의 가나마이신을 사용한 가나마이신 내성에 대해 선택하였다. 아데노신 데아미나제 활성을 갖는 ABE7.10 변이체를 발현하는 세균은 가나마이신 내성 유전자에 도입된 돌연변이를 교정하여 가나마이신 내성을 복구할 수 있다. 가나마이신 내성 세포는 추가의 분석에 대해 선택하였다.

도 3은 표 6에 열거된 선택된 ABE8의 효과 및 특이성을 정량하는 그래프이다. 편집은 HEK293T 세포에서 알파-1 항-트립신 유전자좌에서 검정하였다.

도 4a 및 4b는 ABE8의 편집 효율 및 특이성을 도해한다. **도 4a 및 4b**는 표 6에 기재된 선택된 ABE8의 염기 편집 및 특이성을 정량하는 그래프이다. 단일의 변이체 TadA 데아미나제 도메인 또는 야생형 TadA 데아미나제를

평가하였다.

도 5는 바이스탠더(bystander) A에 대해 온-표적(on-target) 아데닌 (A) 염기를 편집하기 위한 ABE8의 효과를 보여주는 그래프를 제공한다. 특히, ABE8은 효과적인 TadA 데아미나제, ABE7.10과 비교하여 A1AD 부위의 편집 (즉, A·T에서 G·C로의 전환)에서 5배 증가를 유도한다.

도 6a-6d는 염기 편집기 가공을 통해 1차 PiZ 섬유아세포에서 핵염기 교정의 개선된 비율을 생성하는 것과 관련된 핵산 서열, 표 및 막대 그래프를 제공한다. **도 6a**는 A1AD와 연관된 PiZ 돌연변이를 암호화하는 표적 부위 DNA 서열을 보여준다. 이러한 서열은 20개 뉴클레오타이드 프로토스페이서 및 비-카노니칼 spCas9 NGC PAM을 포함한다. **도 6b**는 PiZ 돌연변이를 교정하기 위해 사용되는 다양한 편집기의 TadA 데아미나제 및 Cas9 PAM 변이체 구성성분 둘 다를 기재하는 표를 제공한다. **도 6c 및 6d**는 네온 전기천공 시스템을 사용한 염기 편집 시약으로 형질감염된 환자-유래된 PiZZ 섬유아세포 (GM11423 코리엘 바이오리파지토리)에서 관찰된 편집율을 도시하는 막대-그래프를 제공한다. 각각의 처리는 70,000개 섬유아세포, 염기 편집기를 암호화하는 100 ng의 mRNA 및 50 ng의 알파-1 교정 gRNA를 함유하는 10 μl의 전기천공 완충액으로 이루어졌다. 회수 48h 후, 세포를 용해시키고 관심 대상의 유전자좌는 표적화된 앰플리콘 서열 분석에 의해 조사하였다. 상기 데이터는 2개의 독립적 실험으로부터 수득하였다. 이들 데이터 및 결과는 NGC PAM 인지의 최적화 (변이체 1-3, **도 6b 및 6c**) 및 ABE8/9 돌연변이의 혼입을 통한 TadA 데아미나제의 최적화 (변이체 4-9, **도 6b-6d**) 둘 다로부터 효율을 편집하는데 있어서의 개선을 입증한다.

도 7a-7d는 NSG-PiZ 유전자전이 마우스에서 지질 나노입자 (LNP)-매개된 전달 및 염기 편집에 의해 생성된 혈청 A1AT에서의 증가와 관련된 핵산 서열, 표 및 그래프를 제공한다. **도 7a**는 20개 뉴클레오타이드 프로토스페이서 및 비-카노니칼 spCas9 NGC PAM을 포함하는 표적 부위 DNA 서열을 보여준다. **도 7b**는 PiZ 돌연변이를 교정하기 위해 사용되는 다양한 편집기의 TadA 데아미나제 및 Cas9 PAM 변이체 구성성분 둘 다를 기재하는 표를 제공한다. **도 7c**는 염기 편집기를 암호화하는 1:1 중량비의 gRNA와 mRNA를 함유하는 1.5 mg/kg의 LNP를 사용한 치료 7일 후 NSG-PiZ 유전자전이 마우스 모델로부터 전체 간 gDNA에서 관찰된 편집률을 도시하는 그래프를 제공한다. 시판되는 NSG-PiZ 마우스는 부분 간절제술 후 인간 간 세포를 위해 안정한 배경을 제공하는, 면역결핍 NOD-SCID 감마 (NSG) 배경 상에 돌연변이체 인간 SERPINA1 (Glu342Lys 돌연변이)를 발현한다. (The Jackson Laboratory, Mount Desert Island, ME). 상기 결과는 ngcABEvar9가 조기 버전의 변이체 8 보다 높은 편집률을 산출함을 입증하였다. **도 7d**는 MSD 샌드위치 면역검정에 의한 측정시 전처리 샘플에 상대적으로 혈청 알파-1 안티트립신에서의 증가와 상호관련됨을 보여주는 그래프를 제공한다. 이들 결과를 기준으로, ABE8 시약을 사용한 염기 편집은 알파-1 안티트립신의 결핍증 및 이의 잠재적 폐 후유증을 해소할 수 있다.

도 8은 NRNN PAM 공간 내 모든 가능한 PAM을 접근하기 위해 Cas9 변이체를 도시하는 표이다. 이들의 PAM에서 3개 이하의 정의된 뉴클레오타이드의 인지를 요구하는 Cas9 변이체 만이 열거된다. 비-G PAM 변이체는 SpCas9-NRRH, SpCas9-NRTH, 및 SpCas9-NRCH를 포함하고 (참조: Miller, S.M., *et al*. Continuous evolution of SpCas9 variants compatible with non-G PAMs, *Nat. Biotechnol*. (2020), (//doi.org/10.1038/s41587-020-0412-8), 이의 내용은 이들의 전문이 본원에 참조로 포함된다.

발명을 실시하기 위한 구체적인 내용

[0302] 하기된 바와 같이, 본 발명은 알파-1 항트립신 결핍증 (A1AD)와 연관된 돌연변이를 변경하기 위한 조성물 및 방법을 특징으로 한다. 일부 구현예에서, 편집은 해로운 돌연변이를 교정하여 상기 편집된 폴리뉴클레오타이드가 야생형 참조 폴리뉴클레오타이드 서열과 구분될 수 없도록 한다. 또 다른 구현예에서, 상기 편집은 해로운 돌연변이를 변경하여 편집된 폴리뉴클레오타이드가 유순한 돌연변이를 포함하도록 한다.

[0303] 본 발명은 적어도 부분적으로 아데노신 데아미나제 변이체를 포함하는 염기 편집기가 효율적으로 및 정확하게 A1AD와 연관된 해로운 돌연변이를 효율적으로 및 정확하게 편집할 수 있다는 발견을 기초로 한다.

[0304] **알파-1 항트립신 결핍증(A1AD)**

[0305] 알파-1 항트립신 (A1A)은 염색체 14 상에 SERPINA1 유전자에 의해 암호화된 프로테아제 저해제이다. 상기 당단백질은 주로 간에서 합성되고, 혈액으로 분비되며 건강한 성인에서 혈청 농도는 1.5-3.0 g/L (20-52 μmol/L)이다. 그것은 폐 간질과 폐포 내막 유체로 확산되어 호중구 엘라스타제를 불활성화하여 프로테아제 매개 손상으로부터 폐 조직을 보호한다. 알파-1 항트립신 결핍증(A1AD)은 상염색체 공동우성 양상으로 유전된다. 100개 초과의 SERPINA1 유전자의 유전학적 변이체가 기재되었지만 모두가 질환과 연관되어 있지 않다. 이들 변이체의 알파벳 지정은 겔 전기영동상에서 이들의 이동 속도를 기초로 한다. 가장 통상의 변이체는 M (중간 이동성) 대

립유전자 (PiM)이고, 2개의 가장 빈번한 결핍증 대립유전자는 PiS 및 PiZ이다 (후자는 가장 느린 이동 속도를 갖는다). 측정 가능한 혈청 단백질을 생성하지 않는 여러 돌연변이가 기재되었고; 이들은 "널 (null)" 대립유전자로서 언급된다. 가장 통상적인 유전자형은 MM이고, 이는 알파-1 항트립신의 정상 혈청 수준을 생성한다. 중증의 결핍증을 갖는 대부분의 사람들은 Z 대립유전자 (ZZ)에 대해 동형접합성이다. 미국에서 60,000명 초과의 환자들은 중증의 ZZ 표현형을 갖는다. Z 단백질은 간세포의 소포체에서 이의 생성 동안에 잘못 폴딩되어 중합하고; 이들 비정상적인 중합체는 간에 포집되어 알파-1 항트립신의 혈청 수준을 크게 감소시킨다. 결핍 또는 불안정한 A1AT 생성은 A1AD에 걸린 환자에서 간 및/또는 폐 병리를 유발한다. 알파-1 항트립신 결핍증을 갖는 환자에 나타낸 간 질환은 간세포에서 비정상적인 알파-1 항트립신 단백질의 축적, 및 오토파지, 소포체 스트레스 반응 및 아폽토시스를 포함하는 결과적인 세포 반응에 의해 유발된다. 알파-1 항트립신의 감소된 순환계 수준은 폐에서 증가된 호중구 엘라스타제 활성을 유도하고; 프로테아제와 항프로테아제의 상기 불균형은 상기 병태와 연관된 폐 질환을 유도한다.

[0306] 알파-1 항트립신 결핍증("A1AD")은 백인에서 가장 통상적이고 이것은 대부분 흔히 폐 및 간에 영향을 미친다. 폐에서, 가장 통상적인 증상은 폐 기저부에서 가장 두드러진 조기-발병 (30대 및 40대의 환자)의 범세엽성 폐기종이다. 그러나, 기관지확장증과 같이 확산성 또는 상엽 (upper lobe) 폐기종이 발생할 수 있다. 가장 빈번하게 기재된 증상은 호흡 곤란, 쌕쌕거림 및 기침을 포함한다. 영향을 받은 개체의 폐 기능 시험은 COPD와 일치하는 결과를 보여주지만, 기관지확장제 반응이 관찰될 수 있고 천식으로서 오진될 수 있다. ZZ 유전자형에 의해 유발되는 간 질환은 다양한 방식으로 나타난다. 영향을 받은 영아는 신생아기에 담즙정체성 황달, 때로는 담즙색체 변 (연한 색 또는 점토색) 및 간비대가 나타날 수 있다. 혈중에 접합된 빌리루빈, 트랜스아미나제 및 감마-글루타밀 트랜스퍼라제 수준은 상승된다. 보다 높은 연령의 어린이 및 성인에서 간 질환은 트랜스아미나제 수치가 우발적으로 증가하거나 정맥류 출혈 또는 복수를 포함하여 확립된 간경변증의 징후와 함께 나타날 수 있다. 알파-1 항트립신 결핍증은 또한 환자가 간세포 암종에 걸리게 한다. 동형접합성 ZZ 유전자형은 간 질환이 발병하는데 필요하지만, 이형접합성 Z 돌연변이는 예를 들어, C형 간염 감염 및 낭성 섬유증 간 질환에서 보다 중증의 간 질환에 대한 보다 큰 위험을 부여함에 의해 다른 질환에 대한 유전학적 변형제로서 작용할 수 있다.

[0307] A1AD의 2개의 가장 통상적인 임상적 변이체는 E264V (PiS) 및 E342K (PiZ) 대립유전자이다. 임상적 단일 뉴클레오타이드 변이체 E342K (PiZ)는 불안정하고/하거나 불활성 A1AT 단백질을 유도하고, 결과로서 간 및 폐 독성을 유발한다. 유전은 상염색체 공동우성이다. 절반 초과의 A1AD 환자는 적어도 하나의 카피의 돌연변이 E342K를 함유한다.

병원성 돌연변이의 보정

	유전자	병원성 돌연변이	염기 편집기	gRNA 표적화 서열	PAM
1.	SERPINA1	E342K	ABE	GACAAGAAAGGGACUGAAGC	NGC
2.	SERPINA1	E342K	ABE	AUCGACAAGAAAGGGACUGA	NGC
3.	SERPINC1	R48C (R79C)	ABE	ACACACCGGUUGGUGGCCUC	NGG

[0308]

[0309] 일부 구현예에서, 질환 또는 장애는 알파-1 항트립신 결핍증 (A1AD)이다. 일부 구현예에서, 병원성 돌연변이는 유전자 *SERPINA1*에 있다. 일부 구현예에서, *SERPINA1*의 돌연변이는 E342K (PiZ 대립유전자)이다. 일부 구현예에서, 위치 7에서 A는 G로 편집되어 PiZ 대립유전자를 야생형 대립유전자로 복구시킨다.

[0310] **핵염기 편집기**

[0311] 본원에서는 폴리뉴클레오타이드의 표적 뉴클레오타이드 서열을 편집하거나, 변형시키거나 변경하기 위한 염기 편집기 또는 핵염기 편집기가 기재된다. 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.

[0312] ***폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인***

[0313] 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.

[0314] 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, 상기 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.

[0315] 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.

[0316] 예시적인 촉매적 활성 Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD.

[0317]

[0318] 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.

[0319] 또한 본원에서는 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.

[0320] 또한 본원에서는 이전에 기능성 버전의 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인으로

부터의 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. 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 (참조: 예를 들어, Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, 이의 전체 내용은 본원에 참조로 포함된다).

[0321] 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 기재된다. 염기 편집기에 도입된 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.

[0322] 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 "gNRA")는 crRNA 및 tracrRNA 둘 다의 양상을 단일 RNA 종으로 혼입하기 위해 가공될 수 있다. 예를 들어, 문헌 (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.

[0323] 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.

[0324] 일부 구현예에서, gRNA 스캐폴드 서열은 다음과 같다: GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU.

[0325] 일부 구현예에서, 염기 편집기에 도입된 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 도입된 CRISPR 단백질-유래된 도메인은 결합된 가이드 핵산과 접합되는 경우 표적 폴리뉴클레오타이드에 결합할 수 있는 닉카제이다. 일부 구현예에서, 염기 편집기에 도입된 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 결

합되는 표적 폴리뉴클레오타이드는 RNA이다.

[0326] 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 염기쌍 내에 하나의 가닥 또는 가닥 둘 다의 절단을 지시할 수 있다.

[0327] 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 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 야생형 또는 결실, 삽입, 치환, 변이체, 돌연변이, 융합, 키메라 또는 이의 임의의 조합과 같은 아미노산 변화를 포함할 수 있는 변형된 형태의 Cas9 단백질을 언급할 수 있다.

[0328] 일부 구현예에서, 염기 편집기의 CRISPR 단백질 유래된 도메인은 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 니오쿠아 (*Listeria innocua*) (NCBI Ref: NP_472073.1), 캄필로박터 제주니 (*Campylobacter jejuni*) (NCBI Ref: YP_002344900.1); 나이세리아 메닌기티디스 (*Neisseria. meningitidis*) (NCBI Ref: YP_002342100.1), 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*), 또는 스타필로코커스 아우레우스 (*Staphylococcus aureus*)로부터 기원하는 Cas9의 전부 또는 일부를 포함할 수 있다.

[0329] ***핵염기 편집기의 Cas9 도메인***

[0330] Cas9 뉴클레아제 서열 및 구조는 당업자에게 널리 공지되어 있다 (참조: 예를 들어, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C, Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), 이의 각각의 전체 내용은 본원에 참조로 인용됨). Cas9 오톨로그는 에스. 피오게네스 (*S. Pyogenes*) 및 에스. 써모필러스 (*S. thermophilus*)를 포함하지만 이에 제한되지 않는 다양한 종에 기재되었다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본원 개시내용을 기준으로 당업자에게 자명할 것이고, 상기 Cas9 뉴클레아제 및 서열은 이의 전문이 본원에 참조로 인용되는 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 기재된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다.

[0331] 일부 구현예에서, 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 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 어느 하나와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 어느 하나와 비교하여 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 350개, 적어도 400개, 적어도 500개, 적어도 600개, 적어도 700개, 적어도 800개, 적어도 900개, 적어도 1000개, 적어도 1100개, 또는 적어도 1200개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0332] 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. 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, 상기 단편은 상응하는 야생형 Cas9의 아미노산 길이와 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.

[0333] 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 자명하다.

[0334] Cas9 단백질은 Cas9 단백질을, 가이드 RNA와 상보성을 갖는 특이적 DNA 서열로 가이드하는 가이드 RNA와 연합될 수 있다. 일부 구현예에서, 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 Cas9 도메인, 예를 들어, 뉴클레아제 활성 Cas9, Cas9 닉카제 (nCas9), 또는 뉴클레아제 불활성 Cas9 (dCas9)이다. 핵산 프로그래밍 가능한 DNA 결합 단백질의 예는 제한 없이 Cas9 (예를 들어, dCas9 및 nCas9), CasX, CasY, Cpf1, Cas12b/C2C1, 및 Cas12c/C2C3을 포함한다.

[0335] 일부 구현예에서, 야생형 Cas9는 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) 기원의 Cas9(NCBI 참조 서열: NC_017053.1, 다음과 같은 뉴클레오타이드 및 아미노산 서열)에 상응한다.

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
GTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
CTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT

[0336]

GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTC
TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATC
CAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGA
TGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
AGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTG
ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGA
TACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGT
TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGT
GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGA
CTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT
TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA
TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAA
GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT
GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
AACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTA
TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAG
CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAA
GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
TTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCA
CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGA
TTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTT
GATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGA
AAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAG
GTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAA
AATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATT

[0337]

AGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAG
ACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAAC
GTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAA
GTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTG
GCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGA
GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCT
ATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTT
GGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAA
AGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAA
AATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGA
GAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAA
AGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGA
AAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGAC
AAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAAC
GGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAAT
CCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATT
GACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAA
ATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTAC
AAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCAT
TATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCA
TAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAG
CAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGT
GAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTT
TAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATG
CCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA
GGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF

[0338]

YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0339]

[0340] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0341] 일부 구현예에서, 야생형 Cas9는 하기의 뉴클레오타이드 및/또는 아미노산 서열에 상응하거나 이를 포함한다:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCAT
AACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATT
CGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACT
CGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACA
AGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGT
CCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTC
AACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTG
GGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATC
CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGA
TGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCAC
AATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTG
ACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGA

[0342] CACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTAT

TTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACT
GAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGA
CTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCT
TTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTC
TACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACT
CAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAA
TCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAA
GACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCT
GGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAT
GGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACC
AACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTA
TTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCG
CCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAA
GTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGA
GATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGA
TAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTG
TTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCA
CCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGAT
TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTT
CTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAAC
CTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATA
TTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTG
GATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACG
CGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAG
AGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG
CAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA
ACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGA
AGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGAC
AATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGC
GAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
AACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCAT
GTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCG
GGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAAT
[0343] TCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTC

GTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTA
CAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAG
CCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAAC
GGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGA
TAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAA
AGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGT
GATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCC
TACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGA
AGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACC
AAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGC
TTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCC
CATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCA
GCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCC
TAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATA
CGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGC
ATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAG
ACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAG
CTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGA
CGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
[0344]

QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD
[0345]

[0346] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0347] 일부 구현예에서, 야생형 Cas9는 스트렙토코커스 피오게네스 (*Streptococcus*

[0348] *pyogenes*) (NCBI 참조 서열: NC_002737.2 (다음과 같은 뉴클레오타이드 서열); 및 유니프롯 참조 서열: Q99ZW2 (다음과 같은 아미노산 서열)로부터의 Cas9에 상응한다:

[0349]
ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
GTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
CTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTC
TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATC
CAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGA
TGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
AGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTG
ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGA
TACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGT
TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACT
GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGA
CTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT
TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA

TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAA
GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT
GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
AACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTA
TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAG
CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAA
GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
TTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCA
CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATA
TTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTT
GATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACG
TGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAG
AAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTG
CAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGA
ATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTA
AAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGAT
AACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGC
CAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTG
AACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCAT
GTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCG
AGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAAT
TCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTC
GTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTA
TAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCG
CAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAAT
GGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGA
TAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCA
AGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCG

[0350]

GACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCC
AACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAA
AATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCG
ATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACC
TAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAAT
TACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGT
CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCA
GCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTT
TAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATA
CGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAG
ATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAG
CTAGGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS

[0351]

HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0352]

[0353] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[0354] 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.

[0355] 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.

[0356] 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. 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. 하나의 예로서, 뉴클레아제-불활성 Cas9 도메인은 클로닝 벡터 pPlatTET-gRNA2 (승인 번호 BAV54124)에 제시된 아미노산 서열을 포함한다.

[0357] 예시적인 촉매적 불활성 Cas9 (dCas9)의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0358]

[0359] (참조: 예를 들어, Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell*. 2013; 152(5):1173-83, 이의 전체 내용은 본원에 참조로 포함된다).

[0360] 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 제한되

지 않는다 (참조: 예를 들어, Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, 이의 전체 내용은 본원에 참조로 포함된다).

[0361] 일부 구현예에서, Cas9 뉴클레아제는 불활성(예를 들어, 불활성화된) DNA 절단 도메인을 갖고, 즉, Cas9는 "nCas9" 단백질 ("닉카제" Cas9에 대해)로서 언급되는 닉카제이다. 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 (참조: 예를 들어, 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, 이의 각각의 전문의 내용은 본원에 참조로 포함된다). 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 (*S. Pyogenes*) Cas9의 뉴클레아제 활성을 완전히 불활성화시킨다 (참조: Jinek *et al., Science.* 337:816-821(2012); Qi *et al., Cell.* 28;152(5):1173-83 (2013)).

[0362] 일부 구현예에서, dCas9 도메인은 본원에 제공된 dCas9 도메인 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 어느 하나와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 구현예에서, Cas9 도메인은 본원에 제시된 아미노산 서열 중 어느 하나와 비교하여 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 350개, 적어도 400개, 적어도 500개, 적어도 600개, 적어도 700개, 적어도 800개, 적어도 900개, 적어도 1000개, 적어도 1100개, 또는 적어도 1200개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0363] 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 Cas9에 D10A 및 H840A 돌연변이 또는 상응하는 돌연변이를 포함한다.

[0364] 일부 구현예에서, dCas9는 dCas9 (D10A 및 H840A)의 아미노산 서열을 포함한다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0365]

[0366] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인).

[0367] 일부 구현예에서, Cas9 도메인은 D10A 돌연변이를 포함하고, 위치 840에서 잔기는 상기 제공된 아미노산 서열에서, 또는 본문에 제공된 임의의 아미노산 서열에서 상응하는 위치에 히스티딘을 유지한다.

[0368] 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 아미노산 만큼 더 짧거나 더 긴 아미노산 서열을 갖는다.

[0369] 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 (에 상보적인) 염기인 가닥을 절단함을 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, 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 지식을 기준으로 당업자에게 자명할 것이고 본원 개시내용의 범위 내에 있다.

[0370] 예시적인 촉매적 Cas9 닉카제 (nCas9)의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0371]

[0372] 일부 구현예에서, Cas9는 고세균 (archaea) (예를 들어, 나노고세균) 기원의 Cas9를 언급하고, 이것은 단세포 원핵 미생물의 도메인 및 킹덤을 구성한다. 일부 구현예에서, 프로그래밍 가능한 뉴클레오타이드 결합 단백질은 CasX 또는 CasY 단백질일 수 있고, 이는 예를 들어, 이의 전체 내용이 참조로 인용되는 문헌 (참조: 예를 들어, 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 컴팩트한 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 범위내에 있는 것으로 인지되어야 한다.

[0373] 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, 프로그래밍 가능한 뉴클레오타이드 결합 단백질은 천연적으로 발생하는 CasX 또는 CasY 단백질이다. 일부 구현예에서, 프로그래밍 가능한 뉴클레오타이드 결합 단백질은 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 본원의 개시내용에 따라 사용될 수 있음을 인지해야 한다.

[0374] 예시적인 CasX ((uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53) tr|F0NN87|F0NN87_SULIHCRISPR-연합된 Casx 단백질 OS = 설폴로부스 아이슬란디쿠스 (*Sulfolobus islandicus*) (균주 HVE10/4) GN = SiH_0402 PE=4 SV=1) 아미노산 서열은 다음과 같다:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGK
AKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAP
SFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTP
TRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVRIYTISDAVGQNPTTIN
GGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTG SKRLEDLLY
FANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

[0375]

[0376] 예시적인 CasX (>tr|F0NH53|F0NH53_SULIR CRISPR 연합된 단백질, Casx OS = 설폴로부스 아이슬란디쿠스 (*Sulfolobus islandicus*) (균주 REY15A) GN=SiRe_0771 PE=4 SV=1) 아미노산 서열은 다음과 같다:

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGK
AKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAP
SFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTP
TRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSVVSIYTISDAVGQNPTTIN
GGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYEYLTGSKRLEDLLYF
ANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG.

[0377]

[0378] 델타프로테오박테리아 CasX

MEKRINKIRKKLSADNATKPVSRSGPMKTLLVRVMTDDLKKRLEKRRKKPEVMPQVISNNAA
NNLRMLLDDYTKMKEAILQVYWQEFKDDHVGLMCKFAQPASKKIDQNKLKPEMDEKGNLTTA
GFACSQCGQPLFVYKLEQVSEKGKAYTNYFGRCNVAEHEKLILLAQLKPVKDSDEAVTYSLG
KFGQRALDFYSIHVTKESTHPVKPLAQIAGNRYASGPVGKALSDACMGTIASFLSKYQDIII
EHQKVVKGNQKRLESLRELAGKENLEYPSVTLPPQPHTKEGVDfAYNEVIARVRMWVNLNLW
QKLKLSRDDAKPLLRLKGFPSFPVVERRENEVDWWNTINEVKKLIDAKRDMGRVFWSGVTAE
KRNTILEGYNYLPNENDHKKREGSLENPKKPAKRQFGDLLLYLEKKYAGDWGKVFDEAWERI
DKKIAGLTSHIEREEARNAEDAQSKAVLTDWLRAKASFVLERLKEMDEKEFYACEIQLQKWY
GDLRGNPFAVEAENRVVDISGFSIGSDGHSIQYRNLLAWKYLENGKREFYLLMNYGKKGRIR
FTDGTDIKKSGKWQGLLYGGGKAKVIDLTFDPDDEQLIILPLAFGTRQGREFIWNDLLSLET
GLIKLANGRVIEKTIYNKKIGRDEPALFVALTFERREVVDPSNIKPVNLIGVARGENIPAVI
ALTDPEGCPLPEFKDSSGGPTDILRIGEGYKEKQRAIQAAKEVEQRRAGGYSRKFASKSRNL
ADDMVRNSARDLFYHAVTHDAVLVFANLSRGFGRQGKRTFMTERQYTKMEDWLTAKLAYEGL
TSKTYLSKTLAQYTSKTCSNCGFTITYADMDVMLVRLKKTSDGWATTLNNKELKAEYQITYY
NRYKRQTVEKELSAELDRLSEESGNNDISKWTKGRRDEALFLLKKRFSHRPVQEQFVCLDCG
HEVHAAEQAALNIARSWLFLNSNSTEFKSYKSGKQPFVGAWQAFYKRRLKEVWKPNA

[0379]

[0380] 예시적인 CasY ((ncbi.nlm.nih.gov/protein/APG80656.1)>APG80656.1 CRISPR-연합된 단백질 CasY [배양되지 않은 파쿠박테리아 그룹 박테리움]) 아미노산 서열은 다음과 같다:

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGL
YGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTL
KGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKN
AKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNRNRGEVLFN
KLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELKKAMMDITDAW
RGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLK
GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSD
GRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKL
VPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQK
IFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTEN
IAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVE
NGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAEL
LYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELT
RTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHR
PKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFTIFPEKSAEEEGQ
RYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTK
IARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDAD
KNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLID
AIKDFMRPPIFDENDTPFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIAL
LRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI.

[0381]

[0382] 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 절단이다. 수득한 DSB는 이어서 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.

[0383] 비-상동성 말단 연결 (NHEJ) 및/또는 상동성 지시된 복구 (HDR)의 "효율"은 임의의 간편한 방법에 의해 계산될 수 있다. 예를 들어, 일부 구현예에서, 효율은 성공적인 HDR의 퍼센트로 표현될 수 있다. 예를 들어, 서베이어 (surveyor) 뉴클레아제 검정을 사용하여 절단 생성물을 생성할 수 있고 기질에 대한 생성물의 비율을 사용하여 퍼센트를 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 (퍼센트)은 하기의 수학식을 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).

[0384] 일부 구현예에서, 효율은 성공적인 NHEJ의 퍼센트로 표현될 수 있다. 예를 들어, T7 엔도뉴클레아제 I 검정을 사용하여 절단 생성물을 생성할 수 있고 기질에 대한 생성물의 비율을 사용하여 NHEJ의 퍼센트를 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 (퍼센트)은 하기의 수학식을 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).

[0385] 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. 대부분의 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 유도한다. 이상적인 최종 결과

는 표적화된 유전자 내 기능 상실 돌연변이이다.

[0386] 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.

[0387] 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 호칭됨)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)시킴에 의해 증진될 수 있다. 화학적으로 또는 유전학적으로 NHEJ에 관여하는 유전자의 억제는 또한 HDR 빈도를 증가시킬 수 있다.

[0388] 일부 구현예에서, Cas9는 변형된 Cas9이다. 소정의 gRNA 표적화 서열은 부분 상동성이 존재하는 게놈 전반에 걸쳐 추가의 부위를 가질 수 있다. 이들 부위는 오프-표적으로 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.

[0389] 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.

[0390] 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.

[0391] 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)을 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)를 유도한다)을 갖는다 (참조: 예를 들어, Jinek *et al.*, Science. 2012 Aug. 17; 337(6096):816-21).

[0392] 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.

[0393] 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.

[0394] 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.

[0395] 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.

[0396] 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.

[0397] 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, 상기 방법은 PAM 서열을 요구하지 않는다. 다른 말로, 일부 구현예에서, 상기 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 상기 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 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.

[0398] 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.

[0399] 일부 구현예에서, 변이체 Cas 단백질은 spCas9, spCas9-VRQR, spCas9-VRER, xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK, 또는 spCas9-LRVSQL일 수 있다.

[0400] 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 SpCas9이 사용되었다.

[0401] 에스. 피오게네스 (*S. Pyogenes*) Cas9에 대한 대안은 포유동물 세포에서 절단 활성을 나타내는 Cpf1 패밀리로부

터 RNA-가이드된 엔도뉴클레아제를 포함할 수 있다. 프레보텔라(*Prevotella*) 및 프란시셀라 (*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 다르게, 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. 상기된 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, 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 시스템으로부터 기원하는 것 보다 I형 및 III형과 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 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'의 동정에 의해 표적 DNA 또는 RNA를 절단한다. PAM의 동정 후, Cpf1은 4 또는 5개 뉴클레오타이드 오버행의 점성 말단 유사 DNA 이중 가닥 절단을 도입한다.

[0402] *핵염기 편집기의 Cas12 도메인*

[0403] 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; 이의 각각의 전체 내용은 본원에 참조로 포함된다)을 참조한다. 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.

[0404] 본 발명에 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, 본 발명의 Cas12 단백질은 데아미나제 도메인과 같은 내부적으로 융합된 단백질 도메인에 의해 중단된 아미노산 서열을 포함한다.

[0405] 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)를 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 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, Cas12 도메인은 본원에 제시된 아미노산 서열 중 어느 하나와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 이상의 돌연변이를 갖는 아

미노산 서열을 포함한다. 일부 구현예에서, Cas12 도메인은 본원에 제시된 아미노산 서열 중 어느 하나와 비교하여 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 350개, 적어도 400개, 적어도 500개, 적어도 600개, 적어도 700개, 적어도 800개, 적어도 900개, 적어도 1000개, 적어도 1100개, 또는 적어도 1200개 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0406] 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, 상기 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, 상기 단편은 상응하는 야생형 Cas12의 아미노산 길이와 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% 동일하다. 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.

[0407] 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, 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.

[0408] 일부 구현예에서, 본원에 제공된 바와 같은 Cas12 융합 단백질은 Cas12 단백질의 전장 아미노산 서열, 예를 들어, 본원에 제공된 Cas12 서열 중 하나를 포함한다. 다른 구현예에서, 그러나, 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 자명하다.

[0409] 일반적으로, 부류 2, 유형 V Cas 단백질은 단일 기능성 RuvC 엔도뉴클레아제 도메인을 갖는다 (참조: 예를 들어, Chen et al., "CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity," Science 360:436-439 (2018)). 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.

[0410] 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.

[0411] 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 (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. 상기된 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 다르게, 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 시스템 보다 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.

[0412] 본 발명의 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 히사시 (Bacillus hisashii)로부터의 Cas12)와 적어도 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 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.

[0413] ***핵산 프로그래밍 가능한 DNA 결합 단백질***

[0414] 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 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, Csx15, 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 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.

[0415] 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. Cas9과 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 포함된다)에 이전에 기재되었다.

[0416] 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 포함된다))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.

[0417] 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 기재된 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 기재된 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.

[0418] 야생형 프란시셀라 노비시다 (*Francisella novicida*) Cpf1 (D917, E1006, 및 D1255는 굵게 표시하고 밑줄친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
KNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKE
TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTL
YWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0419]

[0420] 프란시셀라 노비시다 (*F*rancisella novicida) Cpf1 D917A (A917, E1006, 및 D1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
KNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKE
TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTL
YWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0421]

[0422] 프란시셀라 노비시다 (*F*rancisella novicida) Cpf1 E1006A (D917, A1006, 및 D1255는 굵게 표시하고 밑줄

친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
KNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKE
TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTL
YWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0423]

[0424] 프란시셀라 노비시다 (*F*rancisella novicida) Cpf1 D1255A (D917, E1006, 및 A1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
KNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKE
TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTL
YWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0425]

[0426] 프란시셀라 노비시다 (*F*rancisella novicida) Cpf1 D917A/E1006A (A917, A1006, 및 D1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
KNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKE
TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTL
YWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0427]

[0428] 프란시셀라 노비시다 (*F*rancisella novicida) Cpf1 D917A/D1255A (A917, E1006, 및 A1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
KNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKE
TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTL
YWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0429]

[0430] 프란시셀라 노비시다 (*F*rancisella novicida) Cpf1 E1006A/D1255A (D917, A1006, 및 A1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
KNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKE
TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTL
YWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0431]

[0432] 프란시셀라 노비시다 (*F*rancisella novicida) Cpf1 D917A/E1006A/D1255A(A917, A1006, 및 A1255는 굵게 표시하고 밑줄 친다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFI
EEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFN
QNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENR
KNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDY
KTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQ
INDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKE
TLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPS
KKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAI
LFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFY
NPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTL
YWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLI
KDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI
AKLVIEYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRA
YQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKIC
YNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLL
KDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNF
FDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0433]

[0434] 일부 구현예에서, 융합 단백질에 존재하는 Cas9 도메인의 하나는 PAM 서열이 요구되지 않는 가이드 뉴클레오타이드 서열-프로그래밍 가능한 DNA 결합 단백질 도메인으로 대체될 수 있다.

[0435] 일부 구현예에서, Cas9 도메인은 스타필로코커스 아우레우스 (*Staphylococcus aureus*)로부터 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.

[0436] 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.

[0437] 예시적인 SaCas9 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
IQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQ
KAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNS
ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
NRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
PRSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKT
KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSF
LRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIE
TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNG
LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIE
VNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0438]

[0439] 밑줄 치고 굵게 표시한 상기 잔기 N579는 돌연변이시켜 (예를 들어, A579로) SaCas9 닉카제를 생성한다.

[0440] 예시적인 SaCas9n 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
IQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQ
KAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNS
ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
NRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
PRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKT
KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSF
LRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIE
TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNG
LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIE
VNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

[0441]

[0442] N579로부터 돌연변이되어 SaCas9 닉카제를 생성할 수 있는 상기 잔기 A579는 밑줄치고 굵게 표시한다.

[0443] 예시적인 SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
IQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQ
KAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNS
ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
NRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
PRSVSFDNSFNNKVLVKQEE<u>A</u>SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKT
KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSF
LRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIE
TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR<u>K</u>LINDTLYSTRKDDKGNTLIVNNLNG
LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFY<u>K</u>NDLIKINGELYRVIGVNNDLLNRIE
VNMIDITYREYLENMNDKRPP<u>H</u>IIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG.

[0444]

[0445] N579로부터 돌연변이되어 SaCas9 닉카제를 생성할 수 있는 상기 잔기 A579는 밑줄치고 굵게 표시한다. E781, N967, 및 R1014로부터 돌연변이되어 SaKKH Cas9를 생성할 수 있는 상기 잔기 K781, K967, 및 H1014는 밑줄치고 이탤릭으로 나타낸다.

[0446] 일부 구현예에서, napDNAbp는 환형 퍼뮤턴트이다. 하기의 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭한다.

[0447] CP5 (MSP "NGC" PID 및 "D10A" 닉카제와 함께):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYR
STKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAI
GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV
LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0448]

[0449] 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 포함된다)에 기재되었다. 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.

[0450] 알리사이클로바실러스 액시도테라스트리스 (*Alicyclobaccillus acidoterrastris*) 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. 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.

[0451] 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 어느 하나와 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 세균 종 기원의 Cas12b/C2c1 또는 Cas12c/C2c3이 또한 본원 개시내용에 따라 사용될 수 있음을 인지해야 한다.

[0452] Cas12b/C2c1 ((uniprot.org/uniprot/T0D7A2#2) sp|T0D7A2|C2C1_ALIAG CRISPR-연합된 엔도뉴클레아제 C2c1 OS = 알리사이클로바실러스 액시도테레스트리스 (*Alicyclobacillus acido-terrestris*) (균주 ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1) 아미노산 서열은 다음과 같다:

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECD
KTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKF
LSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFG
LKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQ
KNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLA
PDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREV
DDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRG
ARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSE
GLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQLL
KLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQPVDAAN
HMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPK
IRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKE
DRLKKLADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLM
QWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPW
WLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDF
DISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQE
KLSEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSR
VPLQDSACENTGDI.

[0453]

[0454] BhCas12b (바실러스 히사시 (Bacillus hisashii)) NCBI 참조 서열: WP_095142515

MAPKKKRKVGIHGVPAAATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAI
YEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDEVFNILRELYEELVPSSVE
KKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDP
LAKILGKLAEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQALERFLSWES
WNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLR
GWREIIQKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPY
LYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKL
TVQLDRLIYPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGT
LGGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKVVNFKP
KELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLFFPIK
GTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITERE
KRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLEVEIGKEVKHWRKS
LSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKED
RLKKMANTIIMHALGYCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSR
REIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGR
LTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCK
AYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGSSKQSSSELVDS
DILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISKLTNQYSISTIE
DDSSKQSMKRPAATKKAGQAKKKK

[0455]

[0456] 일부 구현예에서, Cas12b는 BvCas12B이다. 일부 구현예에서, Cas12b는 하기에 제공된 예시적인 BvCas12b 아미노산 서열에 넘버링된 바와 같이 아미노산 치환 S893R, K846R, 및 E837G를 포함한다.

[0457] BvCas12b (바실러스 종. V3-13) NCBI 참조 서열: WP_101661451.1

MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKEAYQAE
LINIIRNQQRNNGSSEEHGSDQEILALLRQLYELIIPSSIGESGDANQLGNKFLYPLVDPNS
QSGKGTSNAGRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFT
NIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQLKEKTESYYKEHLT
GGEEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKLPESASPEELWK
VVAEQQNKMSEGFGDPKVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTL
PDAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWIEKENIEIPLAPSI
QFNRQIKLKQHVKGKQEISFSDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFF
NLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLL
EGMRVMSIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKRSFLLNLPGEVVT
KNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTPDERKKAIHKLMEIVQSYDSWTASQKE
VWEKELNLLTNMAAFNDEIWKESLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNID
ELEDTRRLLISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFK
YDKEEKDRYKRWKETYPACQIILFENLNRYLFNLDRSRRENSRLMKWAHRSIPRTVSMQGEM
FGLQVGDVRSEYSSRFHAKTGAPGIRCHALTEEDLKAGSNTLKRLIEDGFINESELAYLKKG
DIIPSQGGELFVTLSKRYKKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMG
EDKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTTFDLQDLDGFEDISK
TIELAQEQQKKYLTMFRDPSGYFFNNETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL

[0458]

[0459] ***가이드 폴리뉴클레오타이드***

[0460] 하나의 구현예에서, 가이드 폴리뉴클레오타이드는 가이드 RNA이다. RNA/Cas 복합체는 Cas 단백질의 표적 DNA로

의 "가이딩"을 도와줄 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 "gNRA")는 crRNA 및 tracrRNA 둘 다의 양상을 단일 RNA 종으로 혼입하기 위해 가공될 수 있다. 문헌 (예를 들어, 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 (참조: 예를 들어, "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); and "Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M.*et al*, Science 337:816-821(2012), 이의 각각의 전체 내용은 본원에 참조로 포함된다). Cas9 오톨로그는 에스. 피오게네스 (*S. Pyogenes*) 및 에스. 써모필러스 (*S. thermophilus*)를 포함하지만 이에 제한되지 않는 다양한 종에 기재되었다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본원 개시내용을 기준으로 당업자에게 자명할 수 있고, 상기 Cas9 뉴클레아제 및 서열은 이의 전문이 본원에 참조로 인용되는 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 기재된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다. 일부 구현예에서, Cas9 뉴클레아제는 불활성(예를 들어, 불활성화된) DNA 절단 도메인을 갖고, 즉, Cas9는 닉카제이다.

[0461] 일부 구현예에서, 가이드 폴리뉴클레오타이드는 적어도 하나의 단일 가이드 RNA ("sgRNA" 또는 "gNRA")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.

[0462] 본원에 기재된 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. 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.

[0463] 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.

[0464] 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 기재된 염기 편집기를 표적 폴리뉴클레오타이드 서열로 지시할 수 있다.

[0465] 일부 구현예에서, 본원에 제공된 염기 편집기는 단일 가이드 폴리뉴클레오타이드 (예를 들어, gRNA)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 예를 들어, PCT/US19/44935에 기재

된 바와 같이 아데노신 염기 편집기, 또는 아데노신 염기 편집기 및 시티딘 염기 편집기를 위해 사용될 수 있다.

[0466] 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.

[0467] 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 폴리뉴클레오타이드가 다중 핵산 분자를 포함하는 경우, 이의 단백질-결합 분절은 예를 들어, 상보체 영역을 따라 하이브리드화하는 다중 분리된 분자의 전부 또는 일부를 포함할 수 있다. 2개의 분리된 분자를 포함하는 DNA-표적화 RNA의 단백질-결합 분절은 (i) 100개 염기쌍 길이인 제1 RNA 분자의 염기쌍 40-75개; 및 (ii) 50개 염기쌍 길이인 제2 RNA 분자의 염기쌍 10-25개를 포함할 수 있다. 특정 문맥에서 구체적으로 달리 정의되지 않는 경우 "분절"의 정의는 특정 수의 총 염기쌍으로 제한되지 않고, 소정의 RNA 분자로부터의 임의의 특정 수의 염기쌍으로 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.

[0468] 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 단일쇄 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.

[0469] 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.

[0470] 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.

[0471] 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.

[0472] 가이드 RNA 또는 가이드 폴리뉴클레오타이드의 제1 영역은 가이드 RNA의 제1 영역이 표적 부위와 염기쌍을 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. 때로는, 가이드 RNA의 제1 영역은 19, 20, 또는 21개 또는 약 19, 20, 또는 21개 뉴클레오타이드 길이일 수 있다.

[0473] 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 염기쌍 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 염기쌍일 수 있다.

[0474] 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 범위일 수 있다.

[0475] 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.

[0476] 가이드 RNA 또는 가이드 폴리뉴클레오타이드는 20개 또는 약 20개 뉴클레오타이드의 핵산 서열을 표적화할 수 있다. 표적 핵산은 20개 미만 또는 약 20개 미만의 뉴클레오타이드일 수 있다. 표적 핵산은 적어도 또는 적어도 약 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30개 길이 또는 1 내지 100개 뉴클레오타이드 길이 사이일 수 있다. 표적 핵산은 최대 또는 최대 약 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50개 길이 또는 1 내지 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 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.

[0477] 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 부위의 서열에 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. 가이드 RNA는 이어서 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.

[0478] 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 오프-표적 C 잔기)의 수는 최소화될 수 있다. 추가로, 소프트웨어 도구를 사용하여 표적 핵산 서열에 상응하는 gRNA를 최적화

할 수 있고, 예를 들어, 게놈에 걸쳐 총 오프-표적 활성을 최소화할 수 있다. 예를 들어, 에스. 피오게네스 Cas9를 사용한 각각의 가능한 표적화 도메인 선택을 위해, 모든 오프-표적 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 오프-표적 스코어에 따라 등급화될 수 있고; 상위 등급의 표적화 도메인은 최대 온-표적 및 최소 오프-표적 활성을 가질 가능성이 있는 것들을 나타낸다. gRNA를 표적화하는 후보물은 당업계에 공지되고/되거나 본원에 제시된 바와 같은 방법을 사용함에 의해 기능적으로 평가될 수 있다.

[0479] 비제한적인 예로서, Cas9과 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 오프-표적 성향을 계산한 후 가이드를 스코어링한다. 전형적으로 완벽한 매칭에서 7개의 미스매칭 범위의 매칭은 17 내지 24개 길이 범위의 가이드를 위해 고려된다. 오프-표적 부위가 계산적으로 결정되면, 합계 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.

[0480] 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 같은 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 오프-표적 DNA 절단을 최소화하기 위해 선택될 수 있다.

[0481] 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 탈아민화시, 상응하는 mRNA는 5'-GUG-3' 대신 5'-AUG-3'로서 전사될 것이고 이는 리포터 유전자의 해독을 가능하게 한다. 적합한 리포터 유전자는 당업자에게 자명하다. 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, 표적 DNA 서열과 관련하여 각각의 데아미나제가 표적화하는 잔기(들)이 어느 것인지를 결정하기 위해 사용될 수 있다. 비-주형 가닥을 표적화하는 sgRNA는 또한 특이적 염기 편집 단백질, 예를 들어, Cas9 데아미나제 융합 단백질의 오프-표적 효과를 평가하기 위해 시험될 수 있다. 일부 구현예에서, 상기 gRNA는 돌연변이된 개시 코돈이 gRNA와 염기쌍을 형성하지 않도록 디자인될 수 있다. 가이드 폴리뉴클레오타이드는 표준 리보뉴클레오타이드, 변형된 리보뉴클레오타이드 (예를 들어, 슈도우리딘), 리보뉴클레오타이드 이성체 및/또는 리보뉴클레오타이드 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 적합한 형광성 염료), 검출 태그 (예를 들어, 비오틴, 디곡시게닌 등), 양자 도트 또는 골드 입자일 수 있다.

[0482] 가이드 폴리뉴클레오타이드는 화학적으로 합성되거나, 효소적으로 합성되거나 이의 조합일 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 작

동적으로 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.

[0483] 일부 구현예에서, 염기 편집기 시스템은 다중 가이드 폴리뉴클레오타이드, 예를 들어, gRNA를 포함할 수 있다. 예를 들어, gRNA는 염기 편집기 시스템에 포함되는 하나 이상의 표적 유전자좌 (예를 들어, 적어도 1개 gRNA, 적어도 2개 gRNA, 적어도 5개 gRNA, 적어도 10개 gRNA, 적어도 20개 gRNA, 적어도 30개 gRNA, 적어도 50개 gRNA)에 표적화될 수 있다. 다중 gRNA 서열은 반복적으로 정렬되고 바람직하게 직접 반복체에 의해 분리된다.

[0484] 가이드 RNA 또는 가이드 폴리뉴클레오타이드를 암호화하는 DNA 서열은 또한 벡터의 일부일 수 있다. 추가로, 벡터는 추가의 발현 제어 서열 (예를 들어, 인핸서 서열, 코작 서열, 폴리아데닐화 서열, 전사 종결 서열 등), 선택가능한 마커 서열 (예를 들어, GFP 또는 항생제 내성 유전자, 예를 들어, 푸로마이신), 복제 오리진, 등을 포함할 수 있다. 가이드 RNA를 암호화하는 DNA 분자는 또한 선형일 수 있다. 가이드 RNA 또는 가이드 폴리뉴클레오타이드를 암호화하는 DNA 분자는 또한 환형일 수 있다.

[0485] 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.

[0486] 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.

[0487] 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 관리는 PAGE, HPLC, MS, 또는 이의 임의의 조합을 포함할 수 있다.

[0488] gRNA 또는 가이드 폴리뉴클레오타이드의 변형은 치환, 삽입, 결실, 화학적 변형, 물리적 변형, 안정화, 정제, 또는 이들의 임의의 조합일 수 있다.

[0489] 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.

[0490] 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, 염기쌍 형성 상호작용 또는 이들의 임의의 조합을 변경할 수 있다.

[0491] PAM 서열은 당업계에 공지된 임의의 PAM 서열일 수 있다. 적합한 PAM 서열은 NGG, NGA, NGC, NGN, NGT, 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; N은 임의의 뉴클레오타이드 염기이고; W는 A 또는 T이다.

[0492] 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 이의 임의의 조합물을 저해할 수 있다. 이들 성질은 PS-RNA gRNA가, 뉴클레아제에 대한 노출이 생체내 또는 시험관내에서 가능성이 높은 적용에서 사용될 수 있게 한다. 예를 들어, 포스포로티오에이트 (PS) 결합은 엑소뉴클레아제 분해를 저해할 수 있는 gRNA의 5'- 또는 ''-말단에서 마지막 3 내지 5개 뉴클레오타이드 사이에 도입될 수 있다. 일부 구현예에서, 포스포로티오에이트 결합은 전체 gRNA에 걸쳐 부가되어 엔도뉴클레아제에 의한 공격을 감소시킬 수 있다.

[0493] ***프로토스페이서 인접 모티프***

[0494] 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 위치한)일 수 있다. 다른 구현예에서, PAM은 3' PAM (즉, 프로토스페이서의 5' 말단의 다운스트림에 위치한)일 수 있다.

[0495] PAM 서열은 표적 결합을 위해 필수적이고, 정확한 서열은 Cas 단백질 유형에 의존한다.

[0496] 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. 예를 들어, Cas9 단백질, 예를 들어, 에스. 피오게네스 (*S. Pyogenes*) 기원의 Cas9 (spCas9)는 전형적으로 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 뉴클레오타이드 길이이다. 여러 PAM 변이체는 하기 표 1에 기재되어 있다.

[0497] [표 1]

Cas9 단백질 및 상응하는 PAM 서열

변이체	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
spCas9-MQKFRAER	NGC
Cpf1	5' (TTTV)
SpyMac	5' -NAA-3'

[0498]

[0499] 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 호칭됨)로부터 선택된 하나 이상의 아미노산 치환을 포함한다.

[0500] 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 표적화된 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 하기 표 2 및 3에 제공된 표적화된 돌연변이 세트로부터 선택된다.

[0501] [표 2]

잔기 1219, 1335, 1337, 1218 에서 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	

[0502]

[0503] [표 3]

잔기 1135, 1136, 1218, 1219, 및 1335 에서 NGT PAM 변이체 돌연변이

변이체	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

[0504]

[0505] 일부 구현예에서, NGT PAM 변이체는 표 2 및 3에서 변이체 5, 7, 28, 31 또는 36으로부터 선택된다. 일부 구현예에서, 변이체는 개선된 NGT PAM 인지를 갖는다.

[0506] 일부 구현예에서, NGT PAM 변이체는 잔기 1219, 1335, 1337, 및/또는 1218에서 돌연변이를 갖는다. 일부 구현예에서, NGT PAM 변이체는 하기 표 4에 제공된 변이체로부터 개선된 인지에 대해 돌연변이로 선택된다.

[0507] [표 4]]

잔기 1219, 1335, 1337 및 1218 에서 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

[0508]

[0509] 일부 구현예에서, NGT PAM에 대해 특이성을 갖는 염기 편집기는 하기 **표 5A**에 제공된 바와 같이 생성될 수 있다.

[0510] [표 5A]

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

[0511]

[0512] 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.

[0513] 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 D10X 돌연변이, 또는 상응하는 돌연변이를 포함하고, 여기서, X는 D를 제외한 임의의 아미노산이다. 일부 구현예에서, SpCas9는 본원에 제공된 임의의 아미노산 서열에서 D10A 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.

[0514] 일부 구현예에서, Cas9는 변경된 PAM 서열에 대한 특이성을 갖는 Cas9 변이체이다. 일부 구현예에서, 추가의 Cas9 변이체 및 PAM 서열은 문헌 (참조: Miller et al., Continuous evolution of SpCas9 variants compatible with non-G PAMs. Nat Biotechnol (2020))에 기재되어 있다. https://doi.org/10.1038/s41587-020-0412-8, 이의 전체 내용은 본원에 참조로 포함된다. 일부 구현예에서, Cas9 변이체는 특이적 PAM 요건을 갖지 않는다. 일부 구현예에서, Cas9 변이체, 예를 들어, SpCas9 변이체는 NRNH PAM에 대해 특이성을 갖고, 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 같은 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 같은 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 같은 위치 1114, 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 같은 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 하기 표 5B, 5C, 5D, 및 5E에 나타낸다.

[0515] [표 5B]

SpCas9/PAM	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

[0516]

[0517] [표 5C]

SpCas9/PAM	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

[0518]

[0519] [표 5D]

SpCas9/PAM	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	

[0520]

[0521] [표 5E]

SpCas9/PAM	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

[0522]

[0523] 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.

[0524] 일부 예에서, 본원에 기재된 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.

[0525] 구현예에서, 에스. 피오게네스 (*S. Pyogenes*)) 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 4kb 암호화 서열)는 세포에서 효율적으로 발현될 수 없는 SpCas9 cDNA를 갖는 플라스미드를 유도할 수 있다. 역으로, 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9)에 대한 암호화 서열은 SpCas9 보다 대략적으로 1킬로베이스 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. 써모필러스 (*S. Thermophilus*)의 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.

[0526] 일부 구현예에서, 에스. 피오게네스 (*S. Pyogenes*) 시스템에 대해, 표적 유전자 서열은 5'-NGG PAM에 선행 (즉, 이에 대해 5')할 수 있고, 20-nt 가이드 RNA 서열은 반대 가닥과 염기쌍을 형성하여 PAM에 인접한 Cas9 절단을 매개할 수 있다. 일부 구현예에서, 인접한 절단은 PAM의 업스트림에 있을 수 있거나 업스트림의 약 3개 염기쌍에 있을 수 있다. 일부 구현예에서, 인접한 절단은 PAM의 업스트림에 있을 수 있거나 업스트림의 약 10개 염기쌍에 있을 수 있다. 일부 구현예에서, 인접한 절단은 PAM의 업스트림에 있을 수 있거나 업스트림의 약 0-20개 염기쌍에 있을 수 있다. 예를 들어, 인접한 절단은 PAM의 업스트림에 있을 수 있거나 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 염기쌍 다음에 있을 수 있다. 인접한 절단은 또한 PAM의 다운스트림의 1 내지 30개 염기쌍에 있을 수 있다. PAM 서열에 결합할 수 있는 예시적인 SpCas9 단백질의 서열은 다음과 같다:

[0527] 예시적인 PAM-결합 SpCas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0528]

[0529] 예시적인 PAM-결합 SpCas9n의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0530]

[0531] 예시적인 PAM-결합 SpEQR Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESVLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0532]

[0533] 상기 서열에서, D1134, R1335 및 T1336로부터 돌연변이되어 SpEQR Cas9를 생성할 수 있는 잔기 E1134, Q1334, 및 R1336은 밑줄치고 굵게 표시한다.

[0534] 예시적인 PAM-결합 SpVQR Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0535]

[0536] 상기 서열에서, D1134, R1335 및 T1336로부터 돌연변이되어 SpVQR Cas9를 생성할 수 있는 잔기 V1134, Q1334, 및 R1336은 밑줄치고 굵게 표시한다.

[0537] 예시적인 PAM-결합 SpVRER Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA**R**ELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STKEVLDATLIHQSITGLYETRIDLSQ
LGGD.

[0538]

[0539] 상기 서열에서, D1134, G1217, R1335 및 T1336으로부터 돌연변이되어 SpVRER Cas9를 생성할 수 있는 잔기 V1134, R1217, Q1334, 및 R1336은 밑줄치고 굵게 표시한다.

[0540] 일부 구현예에서, 가공된 SpCas9 변이체는 3′ H (비-G PAM)에 의해 플랭킹된 프로토스페이서 인접 모티프 (PAM) 서열을 인지할 수 있다 (도 8a-8e를 참조한다). 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. 이들 변이체는 예를 들어, 문헌 (참조: Miller, S.M., *et al*. Continuous evolution of SpCas9 variants compatible with non-G PAMs, *Nat. Biotechnol*. (2020), (//doi.org/10.1038/s41587-020-0412-8), 이의 내용은 이들의 전문이 본원에 참조로 인용됨)에 기재된 바와 같이 파아지-원조 비-연속 진화 (PANCE)를 통해 진화하였다.

[0541] 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.

[0542] 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, 하기에 제공된다.

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSL
HEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQ
KPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDI
GDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEII
SFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQ
KQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

[0543]

[0544] 일부 경우에, 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, 상기 방법은 PAM 서열을 요구하지 않는다. 다시 말해, 일부 구현예에서, 상기 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 상기 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 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.

[0545] 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 보고되었고 당업자에게 자명하다. 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 포함된다)에 기재되었다.

[0546] ***PAM 독점성이 감소된 Cas9 도메인***

[0547] 전형적으로, Cas9 단백질, 예를 들어, 에스. 피오게네스 (*S. Pyogenes*) 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 (Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016), 이의 전체 내용은 참조로 본원에 인용된다)을 참조한다. 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 자명하다. 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 포함된다)에 기재되었다.

[0548] ***고충실도 Cas9 도메인***

[0549] 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 오프-표적 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.

[0550] 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 자명하다. 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 인용됨)에 기재되어 있다.

[0551] 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 오프-표적 부위에서의 절단을 방지한다. 유사하게, SpCas9-HF1은 DNA 포스페이트 골격과 Cas9의 상호작용을 붕괴시키는 알라닌 치환을 통한 오프-표적 편집을 저하시킨다. HypaCas9는 Cas9 교정 및 표적 식별을 증가시키는 REC3 도메인에서 돌연변이 (SpCas9 N692A/M694A/Q695A/H698A)를 함유한다. 모든 3개의 고충실도 효소는 야생형 Cas9 보다 적은 오프-표적 편집을 생성한다.

[0552] 예시적인 고충실도 Cas9는 하기에 제공된다.

[0553] Cas9에 상대적인 고충실도 Cas9 도메인 돌연변이는 굵게 표시하고 밑줄 쳐서 나타낸다.

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATR
LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDE
VAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ
LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT**A**
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFL
KSDGFANRNFM**A**LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

[0554]

[0555] ***핵 국소화 서열 (NLS)을 포함하는 융합 단백질***

[0556] 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 자명하다. 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 아미노산 서열 PKKKRKVEGADKRTADGSEFESPKKKRKV, KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRKPKKKRKV, 또는 MDSLLMNRRKFLYQFKNVRWAKGRRETYLC를 포함한다.

[0557] 일부 구현예에서, NLS는 링커에 존재하거나, NLS는 링커, 예를 들어, 본원에 기재된 링커에 의해 플랭킹된다. 일부 구현예에서, N-말단 또는 C-말단 NLS는 이분된 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 이분된 NLS의 서열은 다음과 같다:

[0558] PKKKRKVEGADKRTADGSEFESPKKKRKV

[0559] 일부 구현예에서, 본 발명의 융합 단백질은 링커 서열을 포함하지 않는다. 일부 구현예에서, 하나 이상의 도메인 또는 단백질 간에 링커 서열이 존재한다.

[0560] 본원 개시내용의 융합 단백질은 하나 이상의 추가의 특성을 포함할 수 있는 것으로 인지되어야 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.

[0561] 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, 10 또는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 초과의 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.

[0562] 상기 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.

[0563] ***핵염기 편집 도메인***

[0564] 본원에서는 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인 및 핵염기 편집 도메인 (예를 들어, 데아미나제 도메인)을 포함하는 융합 단백질을 포함하는 염기 편집기가 기재된다. 염기 편집기는 표적 서열을 인지할 수 있는 가이드 폴리뉴클레오타이드와 상호작용함에 의해 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.

[0565] 일부 구현예에서, 핵염기 편집 도메인은 데아미나제 도메인을 포함한다. 특히 본원에 기재된 바와 같이, 데아미나제 도메인은 아데노신 데아미나제를 포함한다. 일부 구현예에서, 용어 "아데닌 데아미나제" 및 "아데노신 데아미나제"는 상호교환적으로 사용될 수 있다. 핵염기 편집 단백질의 세부사항은 문헌 (참조: 국제 PCT 출원 PCT/2017/045381 (WO2018/027078) 및 PCT/US2016/058344 (WO2017/070632)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); 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.

[0566] ***A에서 G로의 편집***

[0567] 일부 구현예에서, 본원에 기재된 염기 편집기는 아데노신 데아미나제를 포함하는 데아미나제 도메인을 포함할 수 있다. 염기 편집기의 상기 아데노신 데아미나제 도메인은 아데닌 (A)를 탈아민화시켜 구아닌 (G)의 염기쌍 형성 성질을 나타내는 이노신 (I)을 형성함에 의해 아데닌 (A) 핵염기의 구아닌 (G) 핵염기로의 편집을 촉진시킬 수 있다. 아데노신 데아미나제는 데옥시리보핵산 (DNA)에서 데옥시아데노신 잔기의 아데닌을 탈아민화 (즉 아민 그룹을 제거하는)시킬 수 있다.

[0568] 일부 구현예에서, 본원에 제공된 핵염기 편집기는 하나 이상의 단백질 도메인을 함께 융합함에 의해 융합 단백질을 생성시킴에 의해 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 뉴크레아제 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 잔기의 존재 (예를 들어, H840)는 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, A-에서-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.

[0569] 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 포함하는 폴리뉴클레오타이드의 표적 A를 탈아민화시킬 수 있다. 예를 들어, 염기 편집기는 RNA 폴리뉴클레오타이드 및/또는 DNA-RNA 하이브리드 폴리뉴클레오타이드의 표적 A를 탈아민화시킬 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 탈아민화시킬 수 있다. 하나의 구현예에서, 염기 편집기의 아데노신 데아미나제 도메인은 ADAT가 DNA 내 표적 A를 탈아민화시키도록 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, I157F, 또는 상응하는 돌연변이.

[0570] 아데노신 데아미나제는 임의의 적합한 유기체 (예를 들어, 이. 콜리)로부터 유래할 수 있다. 일부 구현예에서, 아데닌 데아미나제는 천연적으로 발생하는 아데노신 데아미나제이고, 이는 본원에 제공된 임의의 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, 임의의 천연적으로 발생하는 아데노신 데아미나제 (예를 들어, ecTadA와 상동성을 갖는)에서 돌연변이를 생성할 수 있고, 이는 본원에 기재된 임의의 돌연변이 (예를 들어, ecTadA에서 동정된 임의의 돌연변이)에 상응한다.

[0571] ***아데노신 데아미나제***

[0572] 일부 구현예에서, 본원에 기재된 염기 편집기는 아데노신 데아미나제를 포함하는 데아미나제 도메인을 포함할 수 있다. 염기 편집기의 상기 아데노신 데아미나제 도메인은 아데닌 (A)를 탈아민화시켜 구아닌 (G)의 염기쌍 형성 성질을 나타내는 이노신 (I)을 형성함에 의해 아데닌 (A) 핵염기의 구아닌 (G) 핵염기로의 편집을 촉진시킬 수 있다. 아데노신 데아미나제는 데옥시리보핵산 (DNA)에서 데옥시아데노신 잔기의 아데닌을 탈아민화 (즉 아민 그룹을 제거하는)시킬 수 있다.

[0573] 일부 구현예에서, 본원에 제공된 아데노신 데아미나제는 아데닌을 탈아민화시킬 수 있다. 일부 구현예에서, 본원에 제공된 아데노신 데아미나제는 DNA의 데옥시아데노신 잔기에서 아데닌을 탈아민화시킬 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 있고, 이는 본원에 기재된 임의의 돌연변이, 예를 들어, ecTadA에서 동정된 임의의 돌연변이에 상응한다. 일부 구현예에서, 아데노신 데아미나제는 원핵세포 기원이다. 일부 구현예에서, 아데노신 데아미나제는 세균 기원이다. 일부 구현예에서, 아데노신 데아미나제는 에스케리치아 콜리 (*Escherichia*

coli), 스타필로코커스 아우레우스 (*Staphylococcus aureus*), 살모넬라 타이피 (*Salmonella typhi*), 쉐와넬라 푸트레파시엔스 (*Shewanella putrefaciens*), 해모필러스 인플루엔자 (*Haemophilus influenzae*), 콜로박터 크레슨투스 (*Caulobacter crescentus*), 또는 바실러스 서브틸리스 (*Bacillus subtilis*)로부터 기원한다. 일부 구현예에서, 아데노신 데아미나제는 이. 콜리 기원이다.

[0574] 본 발명은 효율 (>50-60%) 및 특이성을 증가시키는 아데노신 데아미나제 변이체를 제공한다. 특히, 본원에 기재된 아데노신 데아미나제 변이체는 폴리뉴클레오타이드 내 목적하는 염기를 편집할 가능성이 높고 변경시키고자 하지 않은 염기 (즉, "바이스탠더")를 편집할 가능성이 적다.

[0575] 특정 구현예에서, TadA는 이의 전문이 참조로 본원에 인용된 PCT/US2017/045381 (WO2018/027078)에 기재된 TadA의 임의의 하나이다.

[0576] 일부 구현예에서, 본 발명의 핵염기 편집기는 하기의 서열 중 변경을 포함하는 아데노신 데아미나제 변이체이다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0577] (또한 TadA*7.10으로서 호칭됨).

[0578] 특정 구현예에서, 융합 단백질은 단일(예를 들어, 단량체로서 제공된) TadA*8 변이체를 포함한다. 일부 구현예에서, TadA*8은 Cas9 닉카제에 연결된다. 일부 구현예에서, 본 발명의 융합 단백질은 TadA*8 변이체에 연결된 야생형 TadA (TadA(wt))의 이종이량체로서 포함한다. 다른 구현예에서, 본 발명의 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 7로부터 선택된다. 일부 구현예에서, ABE8은 표 7로부터 선택된다. 관련 서열은 다음과 같다:

[0579] 야생형 TadA (TadA(wt)) 또는 "TadA 참조 서열"

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0580] (서열번호 2)

[0581] TadA*7.10:

MSEVEFSHEYW MRHALTLAKR ARDEREVPVG AVLVLNNRVI GEGWNRAIGL
HDPTAHAEIM ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR
VVFGVRNAKT GAAGSLMDVL HYPGMNHRVE ITEGILADEC AALLCYFFRM
PRQVFNAQKK AQSSTD

[0582]

[0583] 일부 구현예에서, 아데노신 데아미나제는 본원에 제공된 임의의 아데노신 데아미나제에 제시된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 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 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 아미

노산 서열을 포함한다.

[0584] 일부 구현예에서, TadA 데아미나제는 전장 이. 콜리 TadA 데아미나제이다. 예를 들어, 특정 구현예에서, 아데노신 데아미나제는 하기 아미노산 서열을 포함한다:

[0585]
```
MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGA
AGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD.
```

[0586] 그러나 본원에 유용한 추가의 아데노신 데아미나제가 당업자에게 자명하고 본원 개시내용의 범위 내에 있는 것으로 이해되어야만 한다. 예를 들어, 아데노신 데아미나제는 tRNA에 작용하는 아데노신 데아미나제 (ADAT)의 동족체일 수 있다. 제한 없이, 예시적인 AD AT 동족체의 아미노산 서열은 다음을 포함한다:

[0587] 스타필로코커스 아우레우스 (*Staphylococcus aureus*) TadA:

[0588]
```
MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIA
IERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGS
LMNLLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN
```

[0589] 바실러스 서브틸리스 (*Bacillus subtilis*) TadA:

[0590]
```
MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEA
CKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNH
QAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE
```

[0591] 살모넬라 티피무리움 (*Salmonella typhimurium*) (에스. 티피무리움 (*S. typhimurium*)) TadA:

[0592]
```
MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGA
AGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV
```

[0593] 쉐와넬라 푸트레파시엔스 (*Shewanella putrefaciens*) (에스. 푸트레파시엔스 (*S. putrefaciens*)) TadA:

[0594]
```
MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGK
KLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQV
EVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE
```

[0595] 해모필러스 인플루엔자 (*Haemophilus influenzae*) F3031 (에이취. 인플루엔자 (*H. influenzae*)) TadA:

[0596]
```
MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHA
EIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHF
FDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK
```

[0597] 콜로박터 크레슨투스 (*Caulobacter crescentus*) (씨. 크레슨투스 (*C. crescentus*)) TadA:

[0598]
```
MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHA
EIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKF
FAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI
```

[0599] 게오박터 설푸레두센스 (*Geobacter sulfurreducens*) (지. 설푸레두센스 (*G. sulfurreducens*)) TadA:

[0600]
```
MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHA
EMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDL
SADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALF IDERKVPPEP
```

[0601] 이. 콜리 (*E. Coli*) TadA (ecTadA)의 구현예는 하기를 포함한다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA

LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

[0602] GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0603] 일부 구현예에서, 아데노신 데아미나제는 원핵세포 기원이다. 일부 구현예에서, 아데노신 데아미나제는 세균 기원이다. 일부 구현예에서, 아데노신 데아미나제는 에스케리치아 콜리 (*Escherichia coli*), 스타필로코커스 아우레우스 (*Staphylococcus aureus*), 살모넬라 타이피 (*Salmonella typhi*), 쉐와넬라 푸트레파시엔스 (*Shewanella putrefaciens*), 해모필러스 인플루엔자 (*Haemophilus influenzae*), 콜로박터 크레슨투스 (*Caulobacter crescentus*), 또는 바실러스 서브틸리스 (*Bacillus subtilis*)로부터 기원한다. 일부 구현예에서, 아데노신 데아미나제는 이. 콜리 기원이다.

[0604] 하나의 구현예에서, 본 발명의 융합 단백질은 Cas9 닉카제에 연결된 TadA7.10에 연결된 야생형 TadA를 포함한다. 특정 구현예에서, 융합 단백질은 단일 TadA7.10 도메인 (예를 들어, 단량체로서 제공된)을 포함한다. 다른 구현예에서, ABE7.10 편집기는 TadA7.10 및 TadA(wt)을 포함하고 이들은 이종이량체를 형성할 수 있다.

[0605] 일부 구현예에서, 아데노신 데아미나제는 본원에 제공된 임의의 아데노신 데아미나제에 제시된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 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 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.

[0606] 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 자명하다. 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 ecTadA 참조 서열 또는 또 다른 아데노신 데아미나제와 임의의 조합으로 만들어질 수 있는 것으로 인지되어야 한다.

[0607] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 또 다른 아데노신 데아미나제에서 D108G, D108N, D108V, D108A, 또는 D108Y 돌연변이 또는 상응하는 돌연변이를 포함한다.

[0608] 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 다른 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.

[0609] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열 내 E155X 돌연변이, 또는 또 다른 아데노신 데아미나

제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함하고, 여기서, X의 존재는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 다른 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.

[0610] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 D147Y 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[0611] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 E155D, E155G, 또는 E155V 돌연변이를 포함한다. 일부 구현예에서, 아데노신 데아미나제는 D147Y를 포함한다.

[0612] 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 D 147Y; 및 D108N, A106V, E155V, 및 D147Y. 그러나, 본원에 제공된 상응하는 돌연변이의 임의의 조합은 아데노신 데아미나제 (예를 들어, ecTadA)에 만들어질 수 있음이 인지되어야 한다.

[0613] 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 다른 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.

[0614] 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.

[0615] 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 다른 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.

[0616] 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 데

아미나제에서 상응하는 아미노산 이외의 다른 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 아미노산이외의 다른 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 아미노산이외의 다른 임의의 아미노산의 존재를 지적한다.

[0617] 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 다른 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 아미노산이외의 다른 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 아미노산이외의 다른 임의의 아미노산의 존재를 지적한다.

[0618] 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 데아미나제는 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.

[0619] 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.

[0620] 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 인용됨)에 기재되어 있다.

[0621] 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 아

데노신 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.

[0622] 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 아미노산이외의 다른 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.

[0623] 일부 구현예에서, 아데노신 데아미나제는 L84X 돌연변이 아데노신 데아미나제를 포함하고, 여기서, X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 L84F 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[0624] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H123Y 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[0625] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 I156F 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[0626] 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 다른 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 다른 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 아미노산이외의 다른 임의의 아미노산의 존재를 지적한다.

[0627] 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.

[0628] 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.

[0629] 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 아미노산이외의 다른 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.

[0630] 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 다른 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.

[0631] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 R26G, R26N, R26Q, R26C, R26L 또는 R26K 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[0632] 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 다른 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.

[0633] 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 다른 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.

[0634] 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 다른 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.

[0635] 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H36X, N37X, 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 다른 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.

[0636] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 H36L 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[0637] 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 다른 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.

[0638] 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 다른 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.

[0639] 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 다른 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.

[0640] 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 다른 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.

[0641] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 K157N 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[0642] 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 다른 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.

[0643] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 A142N 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연변이를 포함한다.

[0644] 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 다른 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.

[0645] 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 다른 임의의 아미노산을 지적한다. 일부 구현예에서, 아데노신 데아미나제는 TadA 참조 서열에서 R152P 또는 R52H 돌연변이 또는 또 다른 아데노신 데아미나제 (예를 들어, ecTadA)에서 상응하는 돌연

변이를 포함한다.

[0646] 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 상대적으로 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),
(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),
(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),

[0647] (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),
(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),

[0648]

(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),
(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),

[0649]

(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_E155
V
_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152
P _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_E155
V
_I156F _K157N).

[0650]

[0651] 특정 구현예에서, 본원에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성을 개선시키는 하나 이상의 특성을 포함한다. 예를 들어, 본원에 제공된 임의의 융합 단백질은 감소된 뉴크레아제 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.

[0652] 일부 구현예에서, 아데노신 데아미나제는 TadA*7.10이다. 일부 구현예에서, TadA*7.10은 적어도 하나의 변경을 포함한다. 특정 구현예에서, TadA*7.10은 TadA*7.10으로의 하기의 변경 또는 추가의 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 선택되는 변경의 조합을 포함한다: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; 및 I76Y + V82S + Y123H + Y147R + Q154R. 특정 구현예에서, 아데노신 데아미나제 변이체는 잔기 149, 150, 151, 152, 153, 154, 155, 156, 및 157에서 시작하는 C 말단의 결실을 포함한다.

[0653] 다른 구현예에서, 본 발명의 염기 편집기는 하기의 변경 중 하나 이상을 포함하는 아데노신 데아미나제 변이체 (예를 들어, TadA*8)를 포함하는 단량체이다: TadA7.10 또는 TadA 참조 서열과 비교하여 Y147T, Y147R, Q154S, Y123H, V82S, T166R, 및/또는 Q154R. 다른 구현예에서, 아데노신 데아미나제 변이체 (TadA*8)는 하기의 그룹으로부터 선택되는 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 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 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.

[0654] 하나의 구현예에서, 아데노신 데아미나제는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 TadA*8 또는 이의 단편이다:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

[0655] GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD

[0656] 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 결실이다. 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 결실이다. 일부 구현예에서, 아데노신 데아미나제 변이체는 전장 TadA*8이다.

[0657] 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, TadA*8.24이다.

[0658] 하나의 구현예에서, 본 발명의 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)를 포함하고 이들은 이종이량체를 형성할 수 있다. 예시적인 서열은 다음과 같다:

[0659] TadA(wt), "TadA 참조 서열":

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP

[0660] GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0661] TadA*7.10:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

[0662] GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0663] TadA*8:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

[0664] GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD.

[0665] 일부 구현예에서, 아데노신 데아미나제는 본원에 제공된 임의의 아데노신 데아미나제에 제시된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 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 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.

[0666] 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
MPRQVFNAQK KAQSSTD

[0667]

[0668] 예를 들어, TadA*8은 단독으로 또는 하기 Y147T, Y147R, Q154S, Y123H, 및/또는 Q154R의 임의의 하나 이상과 조합하여 아미노산 위치 82 및/또는 166 (예를 들어, V82S, T166R)에서 변경을 포함한다. 특정 구현예에서, 변경의 조합은 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.

[0669] 일부 구현예에서, 아데노신 데아미나제는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 TadA*8 또는 이의 단편이다:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCTFFR
MPRQVFNAQK KAQSSTD

[0670]

[0671] 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 결실이다. 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 결실이다. 일부 구현예에서, 아데노신 데아미나제 변이체는 전장 TadA*8이다.

[0672] 하나의 구현예에서, 본 발명의 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)를 포함하고 이들은 이종이량체를 형성할 수 있다.

[0673] ***추가의 도메인***

[0674] 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.

[0675] 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, 본원의 개시내용은 UGI 도메인을 포함하는 염기 편집기 융합 단백질을 고려한다.

[0676] 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 박

테리오파아지 Mu의 Gam 단백질을 포함할 수 있다. 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 참조하고, 이의 전체 내용은 본원에 참조로 포함된다.

[0677] 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 공정 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 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.

[0678] 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 도입된 NAP 또는 이의 일부는 DNA 폴리머라제이다. 일부 구현예에서, 염기 편집기에 도입된 NAP 또는 이의 일부는 트랜스레젼 (translesion) 폴리머라제 활성을 갖는다. 일부 구현예에서, 염기 편집기에 도입된 NAP 또는 이의 일부는 트랜스레젼 DNA 폴리머라제이다. 일부 구현예에서, 염기 편집기에 도입된 NAP 또는 이의 일부는 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파, 또는 폴리머라제 에타이다. 일부 구현예에서, 염기 편집기에 도입되는 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 도입되는 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.

[0679] **염기 편집기 시스템**

[0680] 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.

[0681] 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.

[0682] 본원에 제공된 바와 같은 염기 편집 시스템은 촉매 결함 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) 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 (C→T 또는 A→G) 변화를 유도하는 염기 절제 복구의 저해제를 함유하는 융합 단백질을 사용하는 게놈 편집에 대한 신규 접근법을 제공한다.

[0683] 본원에서는 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, 상기 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인은 폴리뉴클레오타이드 프로그래밍 가능한 DNA 결합 도메인이다. 일부 구현예에서, 상기 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 데아미나제이다. 일부 구현예에서, 아데노신 염기 편집기는 DNA에서 아데닌을 탈아민화할 수 있다. 일부 구현예에서, ABE는 진화된 TadA 변이체를 포함한다.

[0684] 핵염기 편집 단백질의 세부사항은 국제 PCT 출원 PCT/2017/045381 (WO2018/027078) 및 PCT/US2016/058344 (WO2017/070632)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); 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.

[0685] 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.

[0686] 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.

[0687] 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 이종

성 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.

[0688] 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.

[0689] 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.

[0690] 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, 염기쌍의 의도된 편집은 PAM 부위의 업스트림에 있다. 일부 구현예에서, 염기쌍의 의도된 편집은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 뉴클레오타이드 업스트림에 있다. 일부 구현예에서, 의도된 염기쌍의 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.

[0691] 일부 구현예에서, 상기 방법은 카노니칼 (예를 들어, NGG) PAM 부위를 필요로하지 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.

[0692] 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 (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.

[0693] 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, 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 염기쌍의 편집은 표적 윈도우 내에 있다. 일부 구현예에서, 표적 윈도우는 의도된 염기쌍의 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 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.

[0694] 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.

[0695] 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 다른 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.

[0696] 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.

[0697] 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.

[0698] 일부 구현예에서, 아데노신 염기 편집기 (ABE)는 DNA에서 아데닌을 탈아민화할 수 있다. 일부 구현예에서, ABE는 BE3의 APOBEC1 성분을 천연 또는 가공된 이. 콜리 (*E. coli*) TadA, 인간 ADAR2, 마우스 ADA, 또는 인간 ADAT2로 대체함에 의해 생성된다. 일부 구현예에서, ABE는 진화된 TadA 변이체를 포함한다. 일부 구현예에서, ABE는 ABE 1.2 (TadA*-XTEN-nCas9-NLS)이다. 일부 구현예에서, TadA*는 A106V 및 D108N 돌연변이를 포함한다.

[0699] 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 구

현예에서, 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 진화된 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.

[0700] 일부 구현예에서, ABE는 제3 세대 ABE이다. 일부 구현예에서, ABE는 ABE3.1이고, 이는 3개의 추가의 TadA 돌연변이 (L84F, H123Y, 및 I157F)를 갖는 ABE2.3이다.

[0701] 일부 구현예에서, ABE는 제4 세대 ABE이다. 일부 구현예에서, ABE는 ABE4.3이고, 이는 추가의 TadA 돌연변이 A142N (TadA*4.3)을 갖는 ABE3.1이다.

[0702] 일부 구현예에서, ABE는 제5 세대 ABE이다. 일부 구현예에서, ABE는 ABE5.1이고, 이는 생존 클론 (H36L, R51L, S146C, 및 K157N)으로부터의 돌연변이의 컨센서스 세트의 ABE3.1에 수입함에 의해 생성된다. 일부 구현예에서, ABE는 ABE5.3이고, 이는 내부 진화된 TadA*에 융합된 야생형 이. 콜리 TadA를 함유하는 이종이량체 작제물을 갖는다. 일부 구현예에서, ABE는 하기 표 6에 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 하기 표 6에 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 하기 표 6에 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.

[0703] [표 6] ABE의 유전자형

[0704]

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	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

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
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	I	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	I	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	I	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	I	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	I	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

[0705]

[0706] 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*8.18)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.

[0707] 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 돌연변이를 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.

[0708] 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 이종이량체 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.

[0709] 일부 구현예에서, ABE는 하기 표 7에 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.

[0710] [표 7]

염기 편집기 - ABE8

ABE8 명칭	아데노신 데아미나제	아데노신 데아미나제 기재
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)
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)

[0711]

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)

[0712]

[0713] 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 이분된 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 (*S. pyrogenes*) Cas9 또는 spVRQR Cas9)이다. 일부 구현예에서, 염기 편집기 (예를 들어, ABE7.9, ABE7.10, 또는 ABE8)는 NGC PAM CP6 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 CP6 변이체 (에스. 피로게네스 (*S. pyrogenes*) Cas9 또는 spVRQR Cas9)이다.

[0714] 일부 구현예에서, ABE는 하기 **표 8**에 나타낸 바와 같은 유전자형을 갖는다.

[0715] [표 8] ABE의 유전자형

[0716]

	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

[0717] 하기 **표 9**에 나타낸 바와 같이, 40개 ABE8의 유전자형이 기재된다. ABE의 진화된 이. 콜리 TadA 부분에서 잔기 위치가 지적된다. ABE8에서 돌연변이 변화는 ABE7.10 돌연변이로부터 구분되는 경우 나타낸다. 일부 구현예에서, ABE는 **표 9**에 나타낸 바와 같은 ABE들 중 하나의 유전자형을 갖는다.

[0718] [표 9] 진화된 TadA에서 잔기 동일성

[0719]

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	**R**	**L**	**A**	**L**	**I**	**V**	**F**	**V**	**N**	**Y**	**C**	**Y**	**P**	**Q**	**V**	**F**	**N**	**T**
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				

ABE8.23-m						S				H							
ABE8.24-m						S				H		T					
ABE8.1-d												T					
ABE8.2-d												R					
ABE8.3-d														S			
ABE8.4-d										H							
ABE8.5-d						S											
ABE8.6-d																	R
ABE8.7-d														R			
ABE8.8-d										H		R		R			
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					

[0720]

[0721] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.1 또는 이의 단편이다:

[0722] ABE8.1_Y147T_CP5_NGC PAM_단량체

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSES***
ATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF

[0723]

KYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGS*
*GGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED
IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN
AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV*

[0724]

[0725] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭한다.

[0726] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.1 또는 이의 단편이다:

[0727] pNMG-B335 ABE8.1_Y147T_CP5_NGC PAM_단량체

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSG*GSSGSETPGTSES*
*ATPESSGGSSGGS***EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK**
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF
KYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGS*
*GGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE**
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED
IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN
AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEOEGADKRTADGSEFESPKKKRKV*

[0728]

[0729] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭한다.

[0730] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.14 또는 이의 단편이다:

[0731] NGC PAM CP5를 갖는 pNMG-357_ABE8.14

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDGGSSGGS*SGSETPGTSESA*
*TPESSGGSSGGS*MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHYPGMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD*SGGSSGG*
*SSGSETPGTSESATPESSGGSSGGS***EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE**
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFMQPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPRAFKYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGS*
*GGSGGSGGSGGSGGM***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI**
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF
LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS
PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG
SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN
KVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEF
ESPKKKRKV*

[0732]

[0733] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭한다.

[0734] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.8-m 또는 이의 단편이다:

[0735] ABE8.8-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0736]

[0737] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭하고, 두줄 밑줄 친 서열은 돌연변이를 지적한다.

[0738] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.8-d 또는 이의 단편이다:

[0739] ABE8.8-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGG*
*SSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT**
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS
KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
EVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0740]

[0741] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭하고, 두줄 밑줄 친 서열은 돌연변이를 지적한다.

[0742] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.13-m 또는 이의 단편이다:

[0743] ABE8.13-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0744]

[0745] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭하고, 두줄 밑줄 친 서열은 돌연변이를 지적한다.

[0746] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.13-d 또는 이의 단편이다:

[0747] ABE8.13-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
LHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGG*
*SSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT**
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS
KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
EVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0748]

[0749] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭하고, 두줄 밑줄 친 서열은 돌연변이를 지적한다.

[0750] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.17-m 또는 이의 단편이다:

[0751] ABE8.17-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0752]

[0753] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭하고, 두줄 밑줄 친 서열은 돌연변이를 지적한다.

[0754] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.17-d 또는 이의 단편이다:

[0755] ABE8.17-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
LHDPTAHAEIMALRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTD*SGGSSGG*
*SSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT**
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS
KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
EVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0756]

[0757] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭하고, 두줄 밑줄 친 서열은 돌연변이를 지적한다.

[0758] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.20-m 또는 이의 단편이다:

[0759] ABE8.20-m

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYETRIDLSQLGGD*EGADKRTADGSEFESPKKKRKV**

[0760]

[0761] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭하고, 두줄 밑줄 친 서열은 돌연변이를 지적한다.

[0762] 일부 구현예에서, 염기 편집기는 아데노신 데아미나제 활성을 갖는, 하기의 서열을 포함하거나 이들로 필수적으로 이루어진 ABE8.20-d 또는 이의 단편이다:

[0763] ABE8.20-d

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSES*
*ATPESSGGSSGGS*SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
LHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD*SGGSSGG*
*SSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNT**
DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN
LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQY
ADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI
PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMK
RIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG
GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG
KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQV
NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS
KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
EVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV*

[0764]

[0765] 상기 서열에서, 일반 텍스트는 아데노신 데아미나제 서열을 지칭하고, 굵은 서열은 Cas9로부터 유래된 서열을 지적하고, 이탤릭 서열은 링커 서열을 지칭하고, 밑줄 친 서열은 이분된 핵 국소화 서열을 지칭하고, 두줄 밑줄 친 서열은 돌연변이를 지적한다.

[0766] 일부 구현예에서, 본 발명의 ABE8은 하기의 서열로부터 선택된다:

[0767] 01. 모노ABE8.1_bpNLS + Y147T

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0768]

[0769] 02. 모노ABE8.1_bpNLS + Y147R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCRFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
[0770] TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0771] 03. 모노ABE8.1_bpNLS + Q154S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0772]

[0773] 04. 모노ABE8.1_bpNLS + Y123H

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0774]

[0775] 05. 모노ABE8.1_bpNLS + V82S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0776]

[0777] 06. 모노ABE8.1_bpNLS + T166R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSRDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0778]

[0779] 07. 모노ABE8.1_bpNLS + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0780]

[0781] 08. 모노ABE8.1_bpNLS + Y147R_Q154R_Y123H

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI

[0782] TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0783] 09. 모노ABE8.1_bpNLS + Y147R_Q154R_I76Y

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0784]

[0785] 10. 모노ABE8.1_bpNLS + Y147R_Q154R_T166R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSRDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0786]

[0787] 11. 모노ABE8.1_bpNLS + Y147T_Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0788]

[0789] 12. 모노ABE8.1_bpNLS + Y147T_Q154S

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRSVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0790]

[0791] 13. 모노ABE8.1_bpNLS + H123Y123H_Y147R_Q154R_I76Y

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI
TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0792]

[0793] 14. 모노ABE8.1_bpNLS + V82S + Q154R

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSES
ATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSD
AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDK
GASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK
AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI
RDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ
ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGG
FSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKV
LSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSI

[0794] TGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRKV

[0795] 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 뉴크레아제 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.

[0796] 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 도입된 핵산 폴리머라제 또는 이의 일부는 트랜스레젼 DNA 폴리머라제이다.

[0797] 일부 구현예에서, 염기 편집기의 도메인은 다중 도메인을 포함할 수 있다. 예를 들어, Cas9로부터 유래된 폴리뉴클레오타이드 프로그래밍 가능한 뉴클레오타이드 결합 도메인을 포함하는 염기 편집기는 야생형 또는 천연 Cas9의 REC 엽 및 NUC 엽에 상응하는 REC 엽(lobe) 및 NUC 엽을 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 버전과 비교하여 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.

[0798] 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.

[0799] 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 (참조: 예를 들어, 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; 이의 전체 내용은 본원에 참조로 인용됨) 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 (참조: 예를 들어, 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; 이의 전체 내용은 본원에 참조로 인용됨). 상기 프롤린-풍부 링커는 또한 "강성" 링커로 호칭된다.

[0800] 본 발명의 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, 몽키, 소, 개, 래트, 또는 마우스 데아미나제이다. 일부 구현예에서, 데아미나제는 인간 데아미나제이다. 일부 구현예에서, 데아미나제는 래트 데

아미나제이다.

[0801] *링커*

[0802] 특정 구현예에서, 링커는 본 발명의 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.

[0803] 일부 구현예에서, 상기 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.

[0804] 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(참조: 예를 들어, 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; 이의 전체 내용은 본원에 참조로 인용됨) 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.

[0805] *가이드 RNA와의 Cas9 복합체*

[0806] 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, A 돌연변이를 표적화하는 가이드)를 포함하는 복합체를 제공한다. 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 (참조: 예를 들어, 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; 이의 전체 내용은 본원에 참조로 인용됨) 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 구현

예에서, 본원에 제공된 융합 단백질의 Cas9 도메인은 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다.

[0807] 일부 구현예에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 15 내지 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 관심 대상의 유전자(예를 들어, 질환 또는 장애와 연관된 유전자) 내 서열에 상보적이다. 일부 구현예에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 알파-1 항트립신 결핍증 (A1AD)과 연관된 서열에 상보적이다.

[0808] 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, 상기 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, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, 또는 5' (TTTV) 서열에 바로 인접해 있다.

[0809] 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 단백질에서 및 각각의 암호화 핵산에서의 각각의 잔기를 동정할 수 있을 것이다.

[0810] 본원에 개시된 바와 같이 임의의 융합 단백질을 표적 부위, 예를 들어, 편집될 돌연변이를 포함하는 부위에 표적화하기 위해, 가이드 RNA와 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 tracrRNA은 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 표적화하기 위해 적합한 일부 예시적인 가이드 RNA 서열은 본원에 제공된다.

[0811] *가이드 RNA와의 Cas12 복합체*

[0812] 본원 개시내용의 일부 양상은 본원에 제공된 임의의 융합 단백질 및 가이드 RNA (예를 들어, 편집을 위해 표적 폴리뉴클레오타이드를 표적화하는 가이드)를 포함하는 복합체를 제공한다.

[0813] 일부 구현예에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 15 내지 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.

[0814] 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, 상기 가이드 RNA는 약 15-100개 뉴클레오타이드 길이이고 표적 서열과 상보적인 적어도 10개 인접 뉴클레오타이드의 서열을 포함한다. 일부 구현예에서, 표적 서열의 3' 말단은예를 들어, TTN, DTTN, GTTN, ATTN, ATTC, DTTNT, WTTN, HATY, TTTN, TTTV, TTTC, TG, RTR, 또는 YTN PAM 부위에 바로 인접해 있다.

[0815] 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 단백질에서 및 각각의 암호화 핵산에서의 각각의 잔기를 동정할 수 있을 것이다.

[0816] 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 자명하다. 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 tracrRNA은 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. 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 표적화하기 위해 적합한 일부 예시적인 가이드 RNA 서열은 본원에 제공된다.

[0817] 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:

[0818] NH2-[Cas12 도메인]-링커1-[ABE8]-링커2-[Cas12 도메인]-COOH;

[0819] NH2-[Cas12 도메인]-링커1-[ABE8]-[Cas12 도메인]-COOH;

[0820] NH2-[Cas12 도메인]-[ABE8]-링커2-[Cas12 도메인]-COOH;

[0821] NH2-[Cas12 도메인]-[ABE8]-[Cas12 도메인]-COOH;

[0822] NH2-[Cas12 도메인]-링커1-[ABE8]-링커2-[Cas12 도메인]-[이노신 BER 저해제]-COOH;

[0823] NH2-[Cas12 도메인]-링커1-[ABE8]-[Cas12 도메인]-[이노신 BER 저해제]-COOH;

[0824] NH2-[Cas12 도메인]-[ABE8]-링커2-[Cas12 도메인]-[이노신 BER 저해제]-COOH;

[0825] NH2-[Cas12 도메인]-[ABE8]-[Cas12 도메인]-[이노신 BER 저해제]-COOH;

[0826] NH2-[이노신 BER 저해제]-[Cas12 도메인]-링커1-[ABE8]-링커2-[Cas12 도메인]-COOH;

[0827] NH2-[이노신 BER 저해제]-[Cas12 도메인]-링커1-[ABE8]-[Cas12 도메인]-COOH;

[0828] NH2-[이노신 BER 저해제]-[Cas12 도메인]-[ABE8]-링커2-[Cas12 도메인]-COOH;

[0829] NH2-[이노신 BER 저해제]NH2-[Cas12 도메인]-[ABE8]-[Cas12 도메인]-COOH;

[0830] 추가로, 일부 경우에, Gam 단백질은 염기 편집기의 N 말단에 융합될 수 있다. 일부 구현예에서, Gam 단백질은 염기 편집기의 C말단에 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 공정 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 경우에, 돌연변이 또는 돌연변이들은 야생형 도메인

과 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.

[0831] 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(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.

[0832] 한정된 표적 영역은 탈아민화 윈도우일 수 있다. 탈아민화 윈도우는 염기 편집기가 표적 뉴클레오타이드에 작용하여 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.

[0833] 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)를 포함한다. 일부 구현예에서, 염기 편집기의 NLS는 데아미나제 도메인과 napDNAbp 도메인 사이에 위치한다. 일부 구현예에서, 염기 편집기의 NLS는 napDNAbp 도메인의 C-말단에 위치한다.

[0834] 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.

[0835] 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.

[0836] 일부 구현예에서, BhCas12b 가이드 폴리뉴클레오타이드는 하기의 서열을 갖는다:

[0837] BhCas12b sgRNA 스캐폴드(밑줄 친) + 20nt 내지 23nt 가이드 서열 (N_n으로 지칭됨)

5'GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAG
AAACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAGCACNNNNNNNNNNN

[0838] NNNNNNNNN-3'

[0839] 일부 구현예에서, BvCas12b 및 AaCas12b 가이드 폴리뉴클레오타이드는 하기의 서열을 갖는다:

[0840] BvCas12b sgRNA 스캐폴드(밑줄 친) + 20nt 내지 23nt 가이드 서열(N_n으로 지칭됨)

5'GACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAAGTAATTAAAAATTACC
CACCACAGGAGCACCTGAAAACAGGTGCTTGGCACNNNNNNNNNNNNNNNNNN
NN-3'

[0841]

[0842] AaCas12b sgRNA 스캐폴드(밑줄 친) + 20nt 내지 23nt 가이드 서열(N_n으로 지칭됨)

5'GTCTAAAGGACAGAATTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGG
CAAAGCCCGTTGAACTTCTCAAAAAGAACGATCTGAGAAGTGGCACNNNNNNNN
NNNNNNNNNNNN-3'

[0843]

[0844] ***아데노신 데아미나제 변이체 및 Cas9 도메인을 포함하는 융합 단백질을 사용하는 방법***

[0845] 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 접촉시킴을 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' 말단은 NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, 또는 5' (TTTV) 서열에 바로 인접해 있다.

[0846] 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 단백질에서 및 각각의 암호화 핵산에서의 각각의 잔기를 동정할 수 있을 것이다.

[0847] 본원에 기재된 바와 같이 Cas9 도메인 및 아데노신 데아미나제 변이체 (예를 들어, ABE8)를 포함하는 임의의 융합 단백질을 표적 부위에, 예를 들어, 편집될 돌연변이를 포함하는 부위에 표적화시키기 위해, 가이드 RNA, 예를 들어, sgRNA와 함께 융합 단백질을 동시 발현시킬 필요가 있다는 것은 당업자에게 자명하다. 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 길이이다. 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 표적화하기 위해 적합한 일부 예시적인 가이드 RNA 서열이 본원에 제공된다.

[0848] ***염기 편집기 효율***

[0849] 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 돌연변이임으로, 보다 효율적이고 명확하게 정확한 점 돌연변이를 제조할 수 있는 방법이 요구된다. 본원에

제공된 바와 같은 염기 편집 시스템은 이중 가닥 DNA 절단을 생성하지 않고 공여자 DNA 주형을 요구하지 않고 과도한 확률적 삽입 및 결실을 유도하지 않으면서 게놈 편집을 제공하는 새로운 방법을 제공한다.

[0850] 본 발명의 융합 단백질은 유의적 비율의 삽입-결실을 생성하는 것 없이 돌연변이를 포함하는 단백질을 암호화하는 특정 뉴클레오타이드 염기를 유리하게 변형시킨다. 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, 핵산 내 대다수의 삽입 또는 결실 (예를 들어, 삽입-결실)을 생성시키지 않고 핵산 내 특정 뉴클레오타이드를 효율적으로 변형 (예를 들어, 돌연변이)시키는 염기 편집기를 생성하는 것이 바람직할 수 있다. 특정 구현예에서, 본원에 제공된 임의의 염기 편집기는 삽입-결실에 비해 보다 큰 비율의 의도된 변형 (예를 들어, 돌연변이)을 생성할 수 있다.

[0851] 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.

[0852] 본원 개시내용의 일부 양상은 본원에 제공된 임의의 염기 편집기가 상당한 수의 비의도된 돌연변이, 예를 들어, 비의도된 점 돌연변이를 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.

[0853] 일부 구현예에서, 본원에 제공된 염기 편집기는 1:1 초과인 의도된 돌연변이 대 삽입-결실의 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 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 200:1, 적어도 300:1, 적어도 400:1, 적어도 500:1, 적어도 600:1, 적어도 700:1, 적어도 800:1, 적어도 900:1, 또는 적어도 1000:1 이상인 의도된 돌연변이 대 삽입-결실의 비율을 생성시킬 수 있다.

[0854] 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 인용됨)에 기재된 바와 같은 임의의 적합한 방법을 사용하여 결정될 수 있다.

[0855] 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 정확하게 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.

[0856] 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, 적

어도 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.

[0857] 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. 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 1 시간, 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.

[0858] 본원 개시내용의 일부 양상은 본원에 제공된 임의의 염기 편집기가 상당한 수의 비의도된 돌연변이를 생성하지 않고 핵산 (예를 들어, 대상체의 게놈 내 핵산)에서 의도된 돌연변이를 효율적으로 생성할 수 있다는 인지를 기반으로 한다. 일부 구현예에서, 의도된 돌연변이는 구체적으로 HBG 돌연변이를 변경하거나 교정하도록 특이적으로 디자인된, gRNA에 결합하는 특이적 염기 편집기에 의해 생성되는 돌연변이이다.

[0859] 일부 구현예에서, 본원에 제공된 임의의 염기 편집기는 1:1 초과인 의도된 돌연변이 대 비의도된 돌연변이 (예를 들어, 의도된 돌연변이:비의도된 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 돌연변이의 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.

[0860] ***멀티플렉스 편집***

[0861] 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 요구하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 요구하는 적어도 하나의 가이드 폴리뉴클레오타이드를 포함할 수 있다. 일부 구현예에서, 멀티플렉스 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 요구하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드와 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 요구하는 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.

[0862] 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 유전자좌에 위치한다.

[0863] 일부 구현예에서, 편집은 적어도 하나의 단백질 암호화 영역에서 다수의 핵염기 쌍의 편집이다. 일부 구현예에

서, 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.

[0864] 일부 구현예에서, 상기 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 요구하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드와 연계한다. 일부 구현예에서, 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 요구하는 적어도 하나의 가이드 폴리뉴클레오타이드와 연계한다. 일부 구현예에서, 편집은 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 요구하지 않는 적어도 하나의 가이드 폴리뉴클레오타이드와 표적 폴리뉴클레오타이드 서열로의 표적 결합을 위해 PAM 서열을 요구하는 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.

[0865] **핵산을 편집하기 위한 방법**

[0866] 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.

[0867] 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 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 뉴클레오

타이드 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.

[0868] 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 제2 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 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, 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.

[0869] ***숙주 세포에서 융합 단백질의 발현***

[0870] 아데노신 데아미나제 변이체를 포함하는 본 발명의 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, 본 발명의 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.

[0871] 본원에 기재된 단백질 도메인을 암호화하는 DNA는 DNA를 화학적으로 합성함에 의해 또는 PCR 방법 및 깁슨 어셈블리 방법을 사용하여 전장을 암호화하는 DNA를 작제함에 의해 합성된 부분적 중첩 올리고 DNA 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.

[0872] 핵산 서열-인지 모듈 및/또는 핵산 염기 전환 효소를 암호화하는 DNA를 함유하는 발현 벡터는 예를 들어, DNA를 적합한 발현 벡터 내 프로모터의 다운스트림에 연결함에 의해 생성될 수 있다.

[0873] 발현 벡터로서, 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); 박테리오파아지, 예를 들어, 람다.파아지 등; 곤충 바이러스 벡터, 예를 들어, 바쿨로바이러스 등(예를 들어, BmNPV, AcNPV); 동물 바이러스 벡터, 예를 들어, 레트로바이러스, 백시니아 바이러스, 아데노바이러스 등이 사용된다.

[0874] 프로모터로서, 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 본 발명의 핵산 변형 효소 복합체를 발현시킴에 의해 부여될 수 있기 때문에, 항시성 프로모터는 또한 제한 없이 사용될 수 있다.

[0875] 예를 들어, 숙주가 동물 세포인 경우, SR.알파. 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, SR.알파. 프로모터 등이 바람직할 수 있다.

[0876] 숙주가 에스케리치아 콜리인 경우, trp 프로모터, lac 프로모터, recA 프로모터, 람다.P.sub.L 프로모터, lpp 프로모터, T7 프로모터 등이 바람직할 수 있다.

[0877] 숙주가 바실러스 속인 경우, SPO1 프로모터, SPO2 프로모터, penP 프로모터 등이 바람직할 수 있다.

[0878] 숙주가 효모인 경우, Gal1/10 프로모터, PHO5 프로모터, PGK 프로모터, GAP 프로모터, ADH 프로모터 등이 바람직할 수 있다.

[0879] 숙주가 곤충 세포인 경우, 폴리헤드린 프로모터, P10 프로모터 등이 바람직할 수 있다.

[0880] 숙주가 식물 세포인 경우, CaMV35S 프로모터, CaMV19S 프로모터, NOS 프로모터 등이 바람직할 수 있다.

[0881] 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.

[0882] 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.

[0883] 본 발명의 융합 단백질은 핵산 서열-인지 모듈 및/또는 핵산 염기 전환 효소를 숙주 세포에 도입하고 상기 숙주 세포를 배양함에 의해 세포내 발현될 수 있다.

[0884] 숙주로서, 에스케리치아 속, 바실러스 속, 효모, 곤충 세포, 곤충, 동물 세포 등이 사용된다.

[0885] 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.

[0886] 바실러스 속으로서, 바실러스 서브틸리스 M1114 [Gene, 24, 255 (1983)], 바실러스 서브틸리스 207-21 [Journal of Biochemistry, 95, 87 (1984)] 등이 사용된다.

[0887] 효모로서, 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*) AH22, AH22R.sup.-, NA87-11A, DKD-5D, 20B-12, 쉬조사카로마이세스 폼베 (*Schizosaccharomyces pombe*) NCYC1913, NCYC2036, 피키아 파스토리스 (*Pichia pastoris*) KM71 등이 사용된다.

[0888] 바이러스가 AcNPV인 경우 곤충 세포로서, 양배추 군벌레 유충-유래된 확립된 계통의 세포 (스포도프테라 프루기

페르다 (*Spodoptera frugiperda*) 세포; Sf 세포), 트리코플루시아 니 (*Trichoplusia ni*)의 중간 장으로부터 유래된 MG1 세포, 하이 파이브 (High Five).TM. 트리코플루시아 니의 란으로부터 유래된 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 세포, 예를 들어, Sf9 세포 (ATCC CRL1711), Sf21 세포 [상기 모두, 생체내, 13, 213-217 (1977)] 등이 사용된다.

[0889] 곤충으로서, 예를 들어, 봄믹스 모리, 드로소필라, 크리켓 등의 유충이 사용된다 [참조: Nature, 315, 592 (1985)].

[0890] 동물 세포로서, 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.

[0891] 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.

[0892] 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, 본 발명에 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.

[0893] 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.

[0894] 에스케리치아 콜리는 예를 들어, 문헌 (참조: Proc. Natl. Acad. Sci. USA, 69, 2110 (1972), Gene, 17, 107 (1982) 등)에 기재된 방법에 따라 형질전환될 수 있다.

[0895] 바실러스 속은 예를 들어, 문헌 (참조: Molecular & General Genetics, 168, 111 (1979) 등)에 기재된 방법에 따라 벡터에 도입될 수 있다.

[0896] 효모에는 예를 들어, 문헌 (참조: Methods in Enzymology, 194, 182-187 (1991), Proc. Natl. Acad. Sci. USA, 75, 1929 (1978) 등)에 기재된 방법에 따라 벡터가 도입될 수 있다.

[0897] 곤충 세포 및 곤충에는 예를 들어, 문헌 (참조: Bio/Technology, 6, 47-55 (1988) 등)에 기재된 방법에 따라 벡터가 도입될 수 있다.

[0898] 동물 세포에는 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 벡터가 도입될 수 있다.

[0899] 벡터가 도입된 세포는 숙주의 종류에 따라 공지된 방법에 따라 배양될 수 있다.

[0900] 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 - 약 8이다.

[0901] 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 바람직

할 수 있다. 필요한 경우, 예를 들어, 3.베타.-인돌릴아크릴산과 같은 제제는 배지에 첨가하여 프로모터의 효율적인 기능을 보장할 수 있다. 에스케리치아 콜리는 일반적으로 약 15-약 43℃에서 배양된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[0902] 바실러스 속은 일반적으로 역 30 - 약 40℃에서 배양된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[0903] 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 - 약 8이다. 배양은 일반적으로 약 20℃-약 35℃에서 수행된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[0904] 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.

[0905] 동물 세포를 배양하기 위한 배지로서, 예를 들어, 약 5- 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 - 약 8이다. 배양은 일반적으로 약 30℃-약 40℃에서 수행된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[0906] 식물 세포를 배양하기 위한 배지로서, 예를 들어, MS 배지, LS 배지, B5 배지 등이 사용된다. 배지의 pH는 바람직하게 약 5 - 약 8이다. 배양은 일반적으로 약 20℃-약 30℃에서 수행된다. 필요한 경우, 통기 및 교반이 수행될 수 있다.

[0907] 고등 진핵 세포, 예를 들어, 동물 세포, 곤충 세포, 식물 세포 등이 숙주 세포로서 사용되는 경우, 본 발명의 염기 편집 시스템을 암호화하는 DNA (예를 들어, 아데노신 데아미나제 변이체를 포함하는)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 유도된), 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 기간 동안 수행하여 염기 편집 및 돌연변이의 표적 유전자로의 도입을 수행하고, 염기 편집 시스템의 일과성 발현이 실현될 수 있다.

[0908] 원핵 세포, 예를 들어, 에스코리치아 콜리 등은 유도성 프로모터를 사용할 수 있다. 유도성 프로모터의 예는 lac 프로모터 (IPTG에 의해 유도된), cspA 프로모터 (냉 쇼크에 의해 유도된), araBAD 프로모터 (아라비노스에 의해 유도된) 등을 포함하지만 이에 제한되지 않는다.

[0909] 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).

[0910] **전달 시스템**

[0911] ***핵염기 편집기 및 gRNA의 핵산-기반 전달***

[0912] 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, 누출된 DNA, DNA 복합체, 지질 나노입자를 사용하여) 또는 이들의 조합에 의해 전달될 수 있다. 하나의 구현예에서, 핵염기 편집기는 선택적으로 세포 (예를 들어, 간세포, 배아 줄기 세포, 유도된 만능 줄기 세포 (iPSC), 오가노이드)에 전달된다. 다른 구현예에서, 핵염기 편집기를 암호화하는 핵산은 간 (hepatocyte) (간

(liver)) 세포 또는 이들의 선조체, 및/또는 알파1 항-트립신 (A1AT) 유전자에서 돌연변이를 포함하는 유도된 만능 줄기 세포에 전달된다. 상기 세포를 사용하여 알파1 항-트립신 유전자 편집의 기능적 효과를 검정할 수 있다. 하나의 구현예에서, 변경된 알파1 항-트립신 유전자의 효과는 간세포에서 조사된다.

[0913] 핵염기 편집기를 암호화하는 핵산은 예를 들어, 형질감염 또는 전기천공에 의해 누출된 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.

[0914] 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, ABE8)를 포함하는 Cas9 암호화 서열을 포함할 수 있다.

[0915] 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.

[0916] 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.

[0917] 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 10 (하기)에 나타낸다.

[0918] [표 10]

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	20c	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-N0-테트라데실-3-테트라데실아미노프로피온아미딘	디C14-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-디헥트라데실카바모일메-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

[0919]

[0920] 표 11은 유전자 전달 및/또는 나노입자 제형에 사용하기 위한 예시적인 중합체를 열거한다.

[0921] [표 11]

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

[0922]

[0923] 표 12는 본원에 기재된 융합 단백질을 암호화하는 폴리뉴클레오타이드에 대한 전달 방법을 요약한다.

[0924] [표 12]

전달	벡터/모드	비-분열 세포로의 전달	발현의 지속기간	게놈 통합	전달된 분자 유형
물리적	(예를 들어, 전기천공, 입자건, 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 의존하여	핵산 및 단백질
	중합체 나노입자	예	일과성	전달된 것에 의존하여	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

[0925]

[0926] 또 다른 양상에서, 게놈 편집 시스템 성분 또는 상기 성분, 예를 들어, 핵산 결합 단백질, 예를 들어, Cas9 또

는 이의 변이체 및 관심 대상의 게놈 핵산 서열을 표적화하는 gRNA를 암호화하는 핵산의 전달은 리보핵산단백질(RNP)을 세포에 전달함에 의해 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 오프-표적 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.

[0927] 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 선택 가능한 마커와 같은 추가의 요소의 발현을 구동시킬 수 있다. ITR 활성은 상대적으로 약하여 선택된 뉴클레아제의 과발현으로 인한 잠재적 독성을 감소시키기 위해 사용될 수 있다.

[0928] 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 포함한다. 뇌 또는 다른 CNS 세포 발현을 위해, 적합한 프로모터는 다음을 포함할 수 있다: 모든 뉴런에 대해 시냅신I (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 세포에 대해, 적합한 프로모터는 IFN베타 또는 CD45를 포함할 수 있다. 골아세포를 위해 적합한 프로모터는 OG-2를 포함할 수 있다.

[0929] 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.

[0930] 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.

[0931] ***바이러스 벡터***

[0932] 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 (cargo) 크기 제한에 의해 결정될 수 있다.

[0933] 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 진화된 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 유도한다. 추가로, 높은 형질도입 효율은 많은 상이한 세포 유형 및 표적 조직에서 관찰되었다.

[0934] 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)으로부터 및 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 (예: 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. 일부 양상에서, 본원의 개시내용은 예를 들어, 렌티바이러스 벡터 또는 재조합 아데노-연관된 바이러스 벡터와 같은 바이러스 벡터를 사용하여 알파1 항-트립신 돌연변이를 표적화하는 핵염기 편집기의 바이러스 전달에 관한 것이다.

[0935] 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, 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 (참조: 예를 들어, 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).

[0936] 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.

[0937] 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 (참조: 예를 들어, 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)). 재조합 AAV 벡터의 작제는 문헌 (참조: 미국 특허 제5,173,414호; 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); and Samulski *et al.*, J. Virol. 63:03822-3828 (1989))을 포함하는 다수의 공보에 기재되어 있다.

[0938] AAV는 파르보바이러스 패밀리에 속하는 작은 단일 가닥의 DNA 의존성 바이러스이다. 4.7 kb 야생형 (wt) AAV 게놈은 각각 4개의 복제 단백질 3개의 캡시드 단백질을 암호화하는 2개의 유전자로 구성되고 145-bp 역위 말단 반복체 (ITR)에 의해 양측면 상에 플랭킹된다. 비리온은 동일한 개방 판독 프레임으로부터 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 한정하

는 세포 표면에서 수용체 인지에 관여한다. 바이러스 감염성에 기능하는 포스포리파제 도메인은 Vp1의 고유 N 말단에서 동정되었다.

[0939] 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 본 발명의 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 wtAAV 게놈과 크기에 동일하거나 이 보다 큰 경우 AAV-매개된 유전자 전달의 용도를 제한하였다.

[0940] 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.

[0941] 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. 기재된 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.

[0942] 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 이의 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.

[0943] 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.

[0944] 렌티바이러스는 다음과 같이 제조될 수 있다. pCasES10 (렌티바이러스 전달 플라스미드 골격을 함유하는)을 클로닝한 후, 낮은 계대 (p=5)에서 HEK293FT는 T-75 플라스크에 씨딩하고 10% 태아 소 혈청을 갖고 항생제가 없는 DMEM에서 형질감염시키기 전 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 ul의 플러스 시약 (Plus reagent))을 갖는 4mL 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.

[0945] 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.

[0946] 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.

[0947] 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.

[0948] 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.

[0949] 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 익스테인을 동시 연결하여 2개의 단백질 단편으로부터 전장 단백질을 재구성한다. 다른 적합한 인테인은 당업자에게 자명하다.

[0950] 본 발명의 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 자명하다.

[0951] 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, 상기 카세트의 각각의 반쪽은 단일 AAV 벡터 (<5kb)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' 및 3' 게놈 (이중 AAV 중첩 벡터) 간의 상동성 재조합 (HR); (2) 5' 및 3' 게놈의 ITR-매개된 테일 투 헤드 컨카테머화 (이중 AAV *트랜스*-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.

[0952] ***인테인***

[0953] 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.

[0954] 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.

[0955] 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.

[0956] 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 공정은 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.

[0957] 일부 구현예에서, 염기 편집기 (예를 들어, ABE, CBE)의 N-말단 단편은 스플릿 인테인-N에 융합되고, C-말단 단편은 스플릿 인테인-C에 융합된다. 이들 단편은 이어서 2개 이상의 AAV 벡터에 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 익스테인을 동시 연결하여 2개의 단백질 단편으로부터 전장 단백질을 재구성한다. 다른 적합한 인테인은 당업자에게 자명하다.

[0958] 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 ll**S**dilrvn**T** ei**T**kaplsas mikrydehhq dltllkalvr qqlpekykei ffdq**S**kngya
 361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
 421 ailrrqedfy pflkdnreki ekiltfripy yvgpl**A**rgn**S** rf**A**wm**T**rk**S**e e**T**i**T**pwnfee
 481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
 541 sgeqkkaivd llfktnrkvt vkqlkedyfk kie**C**fd**S**vei sgvedrfn**AS** lgtyhdllki
 601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfddkvmkq lkrrrytgwg
 661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgdsl
 721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
 781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
 841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
 901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrmn tkydendkli revkvitlks
 961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd
```

[0959]

[0960] ***돌연변이를 표적화하기 위한 핵염기 편집기의 용도***

[0961] 돌연변이를 표적화하는 핵염기 편집기의 적합성은 본원에 기재된 바와 같이 평가한다. 하나의 구현예에서, 관

심 대상의 단일 세포에는 리포터 (예를 들어, GFP)를 암호화하는 소량의 벡터와 함께 염기 편집 시스템을 형질도입하였다. 이들 세포는 불멸화된 인간 세포주, 예를 들어, 293T, K562 또는 U2OS을 포함하는, 당업계에 공지된 임의의 세포주일 수 있다. 대안적으로, 1차 세포 (예를 들어, 인간)가 사용될 수 있다. 상기 세포는 궁극적인 세포 표적과 관련될 수 있다.

[0962] 전달은 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.

[0963] 핵염기 편집기의 활성은 본원에 기재된 바와 같이, 즉 세포의 게놈을 서열분석하여 표적 서열 내 변경을 검출함에 의해 평가한다. 생거 서열분석을 위해, 정제된 PCR 앰플리콘은 플라스미드 골격에 클로닝하고, 형질전환시키고, 소량분리하고 단일 프라이머를 사용하여 서열분석하였다. 서열분석은 또한 차세대 서열분석 기술을 사용하여 수행될 수 있다. 차세대 서열분석을 사용하는 경우, 앰플리콘은 300-500bp일 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.

[0964] 초기 시험에서 최대 수준의 표적 특이적 변경을 유도하는 융합 단백질은 추가의 평가를 위해 선택될 수 있다.

[0965] 특정 구현예에서, 핵염기 편집기를 사용하여 관심 대상의 폴리뉴클레오타이드를 표적화한다. 하나의 구현예에서, 본 발명의 핵염기 편집기는 세포의 게놈 내 관심 대상의 돌연변이를 표적화하기 위해 사용되는 가이드 RNA와 함께 세포 (예를 들어, 조혈 세포 또는 이들의 선조체, 조혈 줄기 세포 및/또는 유도된 만능 줄기 세포)에 전달되어 돌연변이를 변경한다. 일부 구현예에서, 염기 편집기는 가이드 RNA에 의해 표적화되어 관심 대상의 유전자 서열에 하나 이상의 편집을 도입한다.

[0966] 시스템은 하나 이상의 상이한 벡터를 포함할 수 있다. 하나의 양상에서, 염기 편집기는 목적하는 세포 유형, 바람직하게 진핵 세포, 바람직하게 포유동물 세포 또는 인간 세포에서 발현을 위해 최적화된 코돈이다.

[0967] 일반적으로, 코돈 최적화는 본래의 서열의 적어도 하나의 코돈 (예를 들어, 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 또는 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 이상의 코돈)을 본래의 아미노산 서열을 유지하면서 숙주 세포의 유전자에 보다 흔하게 또는 가장 흔하게 사용되는 코돈으로 대체함에 의해 관심 대상의 숙주 세포에서 증진된 발현을 위해 핵산 서열을 변형시키는 공정을 언급한다. 다양한 종은 특정 아미노산의 특정 코돈에 대해 특정 편향을 나타낸다. 코돈 편향(유기체 간의 코돈 용법에서의 차이)은 흔히 전령 RNA(mRNA)의 해독 효율과 상관관계가 있고, 이는 이어서 무엇 보다 해독되는 코돈의 성질 및 특정 운반 RNA(tRNA) 분자의 가용성에 의존하는 것으로 사료된다. 세포에서 선택된 tRNA의 우세는 일반적으로 펩타이드 합성에서 가장 흔하게 사용되는 코돈을 반영한다. 따라서, 유전자는 코돈 최적화를 기반으로 소정의 유기체에서 최적의 유전자 발현을 위해 조정될 수 있다. 코돈 용법 표는 예를 들어, 웹사이트 (www.kazusa.orjp/codon/ (visited Jul. 9, 2002))에 가용한 "코돈 용법 데이터베이스"에서 용이하게 가용하고, 이들 표는 다수의 방식으로 적용될 수 있다. 문헌 (Nakamura, Y., *et al.* "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000))을 참조한다. 특정 숙주 세포에서의 발현을 위해 특정 서열을 코돈 최적화하기 위한 컴퓨터 알고리즘 또한 가용하고, 예를 들어, 유전자 포르지 (Gene Forge) (Aptagen; Jacobus, Pa.) 또한 가용하다. 일부 구현예에서, 가공된 뉴클레아제를 암호화하는 서열에서 하나 이상의 코돈 (예를 들어, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 이상 또는 모든 코돈)은 특정 아미노산에 대해 가장 자주 사용되는 코돈에 상응한다.

[0968] 패키징 세포는 전형적으로 숙주 세포를 감염시킬 수 있는 바이러스 입자를 형성하기 위해 사용된다. 상기 세포는 아데노바이러스를 패키징하는 293개 세포, 및 레트로바이러스를 패키징하는 psi.2 세포 또는 PA317 세포를 포함한다. 유전자 치료요법에 사용되는 바이러스 벡터는 일반적으로 핵산 벡터를 바이러스 입자에 패키징하는 세포주를 생성함에 의해 생성된다. 벡터는 전형적으로 패키징 및 숙주로의 후속 통합을 위해 요구되는 최소 바이러스 서열을 포함하고, 기타 바이러스 서열은 발현된 폴리뉴클레오타이드(들)에 대한 발현 카세트에 의해 대체된다. 상실된 바이러스 기능은 전형적으로 패키징 세포주에 의해 트랜스로 공급된다. 예를 들어, 유전자 치료요법에 사용되는 AAV 벡터는 전형적으로 패키징 및 숙주 게놈으로의 통합을 위해 요구되는 AAV 게놈으로부터의 ITR 서열만을 갖는다. 바이러스 DNA는 다른 AAV 유전자, 즉, rep 및 cap를 암호화하지만 ITR 서열이 부재인 헬퍼 플라스미드를 포함하는 세포주에 패키징될 수 있다. 세포주는 또한 헬퍼로서 아데노바이러스로 감염될 수

있다. 헬퍼 바이러스는 AAV 벡터의 복제 및 헬퍼 플라스미드로부터의 AAV 유전자의 발현을 촉진시킬 수 있다. 헬퍼 플라스미드는 일부 경우에 ITR 서열의 부재로 인해 상당량으로 패키징되지 않는다. 아데노바이러스에 의한 오염은 예를 들어, 아데노바이러스가 AAV 보다 더 민감한 열처리에 의해 감소될 수 있다.

[0969] **약제학적 조성물**

[0970] 본원 개시내용의 다른 양상은 본원에 기재된 임의의 염기 편집기, 융합 단백질 또는 융합 단백질-가이드 폴리뉴클레오타이드 복합체를 포함하는 약제학적 조성물에 관한 것이다. 본원에 사용된 바와 같은 용어 "약제학적 조성물"은 약제학적 용도를 위해 제형화된 조성물을 언급한다. 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.

[0971] 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.

[0972] 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) 겔라틴; (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. 습윤화제, 착색제, 방출제, 코팅제, 감미제, 향미제, 방향제, 보존제 및 항산화제는 또한 제형 중에 존재할 수 있다. "부형제", "담체", "약제학적으로 허용되는 담체", "비히클" 등과 같은 용어는 본원에서 상호교환적으로 사용된다.

[0973] 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 바람직하게 약 5.0 내지 약 8.0의 범위에서와 같이 소정의 수준에서 제형의 pH를 유지하고 칼슘 이온을 킬레이팅하지 않는 제제이다. 상기 pH 완충 화합물의 예시적 예는 이미다졸 및 아세테이트 이온을 포함하지만 이에 제한되지 않는다. pH 완충 화합물은 소정의 수준에서 제형의 pH를 유지하기에 적합한 임의의 양으로 존재할 수 있다.

[0974] 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.

[0975] 일부 구현예에서, 약제학적 조성물은 대상체에게 전달하기 위해, 예를 들어, 유전자 편집을 위해 제형화된다. 본원에 기재된 약제학적 조성물을 투여하는 적합한 경로는 제한 없이 다음을 포함한다: 국소, 피하, 경피, 피내, 병변내, 관절내, 복막내, 혈관내, 경점막, 잇몸, 치아내, 와우내, 경고막, 기관내, 경막외, 척수강내, 근육내, 정맥내, 혈관내, 골내, 안주변, 종양내, 뇌내 및 뇌실내 투여.

[0976] 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.

[0977] 다른 구현예에서, 본원에 기재된 약제학적 조성물은 조절 방출 시스템으로 전달된다. 하나의 구현예에서, 펌프가 사용될 수 있다 (참조: 예를 들어, 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). 또 다른 구현예에서, 중합체 물질을 사용할 수 있다 (참조: 예를 들어, 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.). 또한 문헌 (Levy *et al.*, 1985, Science 228: 190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard et ah, 1989, J. Neurosurg. 71: 105.)을 참조한다. 다른 조절 방출 시스템은 예를 들어, 상기 문헌 (Langer)에서 논의된다.

[0978] 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.

[0979] 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 투여를 위해 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 ah, Gene Ther. 1999, 6: 1438-47). 양전하 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.

[0980] 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.

[0981] 추가로, 약제학적 조성물은 (a) 동결건조된 형태로 본 발명의 화합물을 함유하는 컨테이너 및 (b) 약제학적으로 허용되는 희석제 (예를 들어, 본 발명의 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.

[0982] 또 다른 양상에서, 상기된 질환의 치료를 위해 유용한 물질을 함유하는 제품이 포함된다. 일부 구현예에서, 제품은 컨테이너 및 표지를 포함한다. 적합한 컨테이너는 예를 들어, 병, 바이알, 주사기 및 시험 튜브를 포함한다. 상기 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 본 발명의 화합물이다. 일부 구현예에서, 컨테이너 상에 또는 이와 연합된 표지는 조성물이 선택된 질환을 치료하기 위해 사용됨을 지적한다. 제품은 포스페이트-완충 식염수, 링거 용액, 또

는 덱스트로스 용액과 같은 약제학적으로 허용되는 완충액을 포함하는 제2 컨테이너를 추가로 포함할 수 있다. 이것은 다른 완충제, 희석제, 필터, 바늘, 주사기 및 사용 설명서가 있는 패키지 삽입물을 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[0983] 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.

[0984] 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 임으로 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 문헌의 모든 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 수의학적 용도로 투여하기에 적합하다는 것은 당업자라면 이해할 것이다.

[0985] 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 아니다.

[0986] 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 목적하는 특정 투여 형태에 적합한, 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 포함한다. 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 기재한다. 또한, 문헌 (PCT 출원 PCT/US2010/055131 (공개번호 WO2011/053982 A8, 2010년 11월 출원됨)은 뉴클레아제를 포함하는 약제학적 조성물을 제조하기 위한 추가의 적합한 방법, 시약, 부형제 및 용매에 대해 본원에 참조로 포함된다.

[0987] 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 범위내 있는 것으로 고려된다.

[0988] 상기된 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 개업 의사에게 널리 공지된 유사 인자에 의존할 수 있다. 치료학적 적용을 위해, 이것은 의학적으로 목적하는 결과를 성취하기에 충분한 양이다.

[0989] 일부 구현예에서, 본원 개시내용에 따른 조성물은 임의의 다양한 질환, 장애 및/또는 병태의 치료를 위해 사용될 수 있다.

[0990] ***키트***

[0991] 본원 개시내용의 다양한 양상은 염기 편집기 시스템을 포함하는 키트를 제공한다. 하나의 구현예에서, 상기 키트는 핵염기 편집기 융합 단백질을 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 작제물을 포함한다. 융합 단백질은 데아미나제 (예를 들어, 아데노신 데아미나제) 및 핵산 프로그래밍 가능한 DNA 결합 단백질 (napDNAbp)을 포함한다. 일부 구현예에서, 상기 키트는 관심 대상의 핵산 분자를 표적화할 수 있는 적어도 하나의 가이드 RNA를 포함한다. 일부 구현예에서, 상기 키트는 적어도 하나의 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 작제물을 포함한다. 하나의 구현예에서, 상기 키트는 데아미나제, 및 알파-1 항트립신 폴리뉴클레오타이드를 표적화할 수 있는 가이드 RNA를 포함하는 핵염기 편집기 융합 단백질을 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 작제물을 포함한다.

[0992] 키트는 일부 구현예에서, 하나 이상의 돌연변이를 편집하기 위해 키트를 사용하기 위한 지침서를 제공한다. 상기 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. 이것은 다른 완충제, 희석제, 필터, 바늘, 주사기 및 사용 설명서가 있는 패키지 삽입물을 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[0993] ***내부 삽입물을 갖는 융합 단백질***

[0994] 본원에서는 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.

[0995] 일부 구현예에서, 이종성 폴리펩타이드는 데아미나제 또는 이의 기능성 단편이다. 예를 들어, 융합 단백질은 Cas9 또는 Cas12 (예를 들어, Cas12b/C2c1), 폴리페타이드의 N-말단 단편 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 상기된 융합 단백질을 위해 적합한 데아미나제이다.

[0996] 데아미나제는 환형 퍼뮤턴트 데아미나제일 수 있다. 예를 들어, 데아미나제는 환형 퍼뮤턴트 아데노신 데아미나제일 수 있다. 일부 구현예에서, 데아미나제는 TadA 참조 서열에서 넘버링된 바와 같이 아미노산 잔기 116에서 환형으로 퍼뮤턴트화된 환형 퍼뮤턴트 TadA이다. 일부 구현예에서, 데아미나제는 TadA 참조 서열에서 넘버링된 바와 같이 아미노산 잔기 136에서 환형으로 치환된 환형 퍼뮤턴트 TadA이다. 일부 구현예에서, 데아미나제는 TadA 참조 서열에서 넘버링된 바와 같이 아미노산 잔기 65에서 환형으로 치환된 환형 퍼뮤턴트 TadA이다.

[0997] 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. 융합 단백질에서 2개 이상의 데아미나제는 예를 들어, PCT/US19/44935에 기재된 바와 같이 아데노신 데아미나제, 시티딘 데아미나제 또는 이의 조합일 수 있다. 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.

[0998] 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 상대적으로 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.

[0999] 일부 구현예에서, Cas9 폴리펩타이드는 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 스타필로코커스 아우레우스 (*Staphylococcus aureus*) Cas9 (SaCas9), 스트렙토코커스 써모필러스 1 (*Streptococcus thermophilus 1*) Cas9 (St1Cas9), 또는 이의 단편 또는 변이체이다.

[1000] 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.

[1001] 융합 단백질의 Cas9 폴리펩타이드는 하기에 제시된 Cas9 아미노산 서열 (하기에서 "Cas9 참조 서열"로 호칭됨)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.

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1002]

[1003] (한줄 밑줄: HNH 도메인; 두줄 밑줄: RuvC 도메인)

[1004] 일부 구현예에서, 융합 단백질 내 napDNAbp는 Cas12 폴리펩타이드, 예를 들어, Cas12b/C2c1 또는 이의 단편이다. Cas12 폴리펩타이드는 변이체 Cas12 폴리펩타이드일 수 있다.

[1005] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 적합한 위치에서 napDNAbp (예를 들어, Cas9 또는 Cas12 (예를 들어, Cas12b/C2c1))에 삽입될 수 있어, 상기 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.

[1006] 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 국소적 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-인자 보다 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-인자 보다 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.

[1007] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 하기로 이루어진 그룹으로부터 선택된 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, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기. 일부 구현예에서, 이종성 폴리펩타이드는 상기 Cas9 참조 서열에 넘버링된 바와 같은 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 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, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기. 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.

[1008] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 하기로 이루어진 그룹으로부터 선택된 아미노산 잔기에서 napDNAbp에 삽입될 수 있다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 768, 792, 1022, 1026, 1040, 1068, 및 1247, 또는 또 다른 Cas9 폴리펩타이드에서 상응하는 아미노산 잔기. 일부 구현예에서, 이종성 폴리펩타이드는 상기 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 잔기를 대체한다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 768, 792, 1022, 1026, 1040, 1068, 및 1247, 또는 또 다른 Cas9 폴리펩타이드에서 상응하는 아미노산 잔기.

[1009] 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 본원에 기재된 바와 같은 아미노산 잔기에서 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 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, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기. 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-말단에 삽입된다.

[1010] 일부 구현예에서, 아데노신 데아미나제 (예를 들어, TadA)는 하기로 이루어진 그룹으로부터 선택된 아미노산 잔기에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기. 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-말단에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기. 일부 구현예에서, 아데노신 데아미나제는 하기로 이루어진 그룹으로부터 선택되는 아미노산의 C-말단에 삽입된다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기. 일부 구현예에서, 아데노신 데아미나제는 하기로 이루어진 그룹으로부터 선택되는 아미노산을 대체한다: 상기 Cas9 참조 서열에 넘버링된 바와 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기.

[1011] 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 768, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 구현예에서, 데아미나제 (예를 들어, 아데노신 데아미나제)는 상기 Cas9 참조 서열에 넘버링된 바와 같은 아미노산 잔기 768의 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 768, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[1012] 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 아미노산 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, 또는 아미노산 792의 N-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 791을 대체하기 위해 삽입되거나, 아미노산 792를 대체하기 위해 삽입되거나 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[1013] 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.

[1014] 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.

[1015] 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.

[1016] 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-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 1040, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기를 대체하기 위해 삽입된다.

[1017] 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.

[1018] 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.

[1019] 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 다

른 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.

[1020] 일부 구현예에서, 이종성 폴리펩타이드 (예를 들어, 데아미나제)는 Cas9 폴리펩타이드의 가요성 루프에 삽입된다. 가요성 루프 부분은 상기 Cas9 참조 서열에 넘버링된 바와 같이 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 하기로 이루어진 그룹으로부터 선택될 수 있다: 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 잔기.

[1021] 이종성 폴리펩타이드 (예를 들어, 아데닌 데아미나제)는 아미노산 잔기에 상응하는 Cas9 폴리펩타이드 영역에 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, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기.

[1022] 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.

[1023] 예시적인 내부 융합 염기 편집기는 하기 **표 13A**에 제공되고, 또한 PCT/US20/16285에 기재되어 있다.

[1024] [표 13A]

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 136ins 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 2791	ISLAY44

[1025]

[1026] 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.

[1027] 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.

[1028] 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.

[1029] 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.

[1030] 일부 구현예에서, N-말단 Cas9 단편의 C-말단은 융합 단백질이 표적 핵염기를 탈아민화시키는 경우, 표적 핵염기에 인접해 있는 아미노산을 포함한다. 일부 구현예에서, C-말단 Cas9 단편의 N-말단은 융합 단백질이 표적 핵염기를 탈아민화시키는 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 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, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기.

[1031] 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 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, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기. N-말단 Cas9 단편은 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, 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.

[1032] 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 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, 또는 또 다른 Cas9 폴리펩타이드 내 상응하는 아미노산 잔기. 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.

[1033] 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.

[1034] 본원에 기재된 융합 단백질은 비-표적 부위 (예를 들어, 오프-표적 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 않는 탈아민화 또는 오프-표적 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 오프-표적 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.

[1035] 일부 구현예에서, 융합 단백질의 데아미나제 (예를 들어, 아데노신 데아미나제)는 R-루프 범위 내 2개 이하의 핵염기를 탈아민화시킨다. 일부 구현예에서, 융합 단백질의 데아미나제는 R-루프 범위 내 3개 이하의 핵염기를 탈아민화시킨다. 일부 구현예에서, 융합 단백질의 데아미나제는 R-루프 범위 내 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이하의 핵염기를 탈아민화시킨다. 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 길이를 가질 수 있다. 일부 구현예에서, R-루프 영역은 약 20 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 RNA에 대한 비-상보적 가닥 (프로토스페이서 가닥) 상에 핵염기의 편집을 포함한다.

[1036] 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 염기쌍, 약 2 내지 10개 염기쌍, 약 3 내지 11개 염기쌍, 약 4 내지 12개 염기쌍, 약 5 내지 13개 염기쌍, 약 6 내지 14개 염기쌍, 약 7 내지 15개 염기쌍, 약 8 내지 16개 염기쌍, 약 9 내지 17개 염기쌍, 약 10 내지 18개 염기쌍, 약 11 내지 19개 염기쌍, 약 12 내지 20개 염기쌍, 약 1 내지 7개 염기쌍, 약 2 내지 8개 염기쌍, 약 3 내지 9개 염기쌍, 약 4 내지 10개 염기쌍, 약 5 내지 11개 염기쌍, 약 6 내지 12개 염기쌍, 약 7 내지 13개 염기쌍, 약 8 내지 14개 염기쌍, 약 9 내지 15개 염기쌍, 약 10 내지 16개 염기쌍, 약 11 내지 17개 염기쌍, 약 12 내지 18개 염기쌍, 약 13 내지 19개 염기쌍, 약 14 내지 20개 염기쌍, 약 1 내지 5개 염기쌍, 약 2 내지 6개 염기쌍, 약 3 내지 7개 염기쌍, 약 4 내지 8개 염기쌍, 약 5 내지 9개 염기쌍, 약 6 내지 10개 염기쌍, 약 7 내지 11개 염기쌍, 약 8 내지 12개 염기쌍, 약 9 내지 13개 염기쌍, 약 10 내지 14개 염기쌍, 약 11 내지 15개 염기쌍, 약 12 내지 16개 염기쌍, 약 13 내지 17개 염기쌍, 약 14 내지 18개 염기쌍, 약 15 내지 19개 염기쌍, 약 16 내지 20개 염기쌍, 약 1 내지 3개 염기쌍, 약 2 내지 4개 염기쌍, 약 3 내지 5개 염기쌍, 약 4 내지 6개 염기쌍, 약 5 내지 7개 염기쌍, 약 6 내지 8개 염기쌍, 약 7 내지 9개 염기쌍, 약 8 내지 10개 염기쌍, 약 9 내지 11개 염기쌍, 약 10 내지 12개 염기쌍, 약 11 내지 13개 염기쌍, 약 12 내지 14개 염기쌍, 약 13 내지 15개 염기쌍, 약 14 내지 16개 염기쌍, 약 15 내지 17개 염기쌍, 약 16 내지 18개 염기쌍, 약 17 내지 19개 염기쌍, 약 18 내지 20개 염기쌍 떨어져 있거나 이의 업스트림에 있다. 일부 구현예에서, 표적 핵염기는 PAM 서열로부터 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 염기쌍 떨어져 있거나 이의 업스트림에 있다. 일부 구현예에서, 표적 핵염기는 PAM 서열의 약 1, 2, 3, 4, 5, 6, 7, 8, 또는 9개 염기쌍 업스트림에 있다. 일부 구현예에서, 표적 핵염기는 PAM 서열의 약 2, 3, 4, 또는 6개 염기쌍 업스트림에 있다.

[1037] 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 탠덤

으로 삽입될 수 있다. 2개 이상의 이종성 도메인은 이들이 탠덤으로 NapDNAbp에 있지 않도록 하는 위치에 삽입될 수 있다.

[1038] 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.

[1039] 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.

[1040] Cas9 또는 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. Cas9 또는 Cas12 및 하나 이상의 데아미나제 도메인, 예를 들어, 아데노신 데아미나제를 포함하거나 Cas9 또는 Cas12 서열에 의해 플랭킹된 아데노신 데아미나제를 포함하는 융합 단백질은 또한 표적 서열의 고도의 특이적 및 효율적 염기 편집을 위해 유용하다. 하나의 구현예에서, 키메라 Cas9 또는 Cas12 융합 단백질은 Cas12 폴리펩타이드 내 삽입된 이종성 촉매 도메인을 함유한다.

[1041] 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. 다른 구현예에서, 융합 단백질은 하나 이상의 촉매 도메인을 함유한다. 다른 구현예에서, 하나 이상의 촉매 도메인의 적어도 하나는 Cas12 폴리펩타이드 내 삽입되거나 Cas12 N-말단 또는 C-말단에 융합된다. 다른 구현예에서, 하나 이상의 촉매 도메인의 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 (*Bacillus hisashii*) Cas12b, 바실러스 써모아밀로보란스 (*Bacillus thermoamylovorans*) Cas12b, 바실러스 종 (*Bacillus sp.*) *V3-13* Cas12b, 또는 알리사이클로바실러스 액시디필러스 (*Alicyclobacillus acidiphilus*) Cas12b와 적어도 약 85% 아미노산 서열 동일성을 갖는다. 다른 구현예에서, Cas12 폴리펩타이드는 바실러스 히사시 (*Bacillus hisashii*) Cas12b, 바실러스 써모아밀로보란스 (*Bacillus thermoamylovorans*) Cas12b, 바실러스 종 (*Bacillus sp.*) *V3-13* Cas12b, 또는 알리사이클로바실러스 액시디필러스 (*Alicyclobacillus acidiphilus*) Cas12b와 적어도 약 90% 아미노산 서열 동일성을 갖는다. 다른 구현예에서, Cas12 폴리펩타이드는 바실러스 히사시 (*Bacillus hisashii*) Cas12b, 바실러스 써모아밀로보란스 (*Bacillus thermoamylovorans*) Cas12b, 바실러스 종 (*Bacillus sp.*) *V3-13* Cas12b, 또는 알리사이클로바실러스 액시디필러스 (*Alicyclobacillus acidiphilus*) Cas12b와 적어도 약 95% 아미노산 서열 동일성을 갖는다. 다른 구현예에서, Cas12 폴리펩타이드는 바실러스 히사시 (*Bacillus hisashii*) Cas12b, 바실러스 써모아밀로보란스 (*Bacillus thermoamylovorans*) Cas12b, 바실러스 종 (*Bacillus sp.*) *V3-13* Cas12b, 또는 알리사이클로바실러스 액시디필러스 (*Alicyclobacillus acidiphilus*) Cas12b의 단편을 포함하거나 필수적으로 이들로 이루어진다.

[1042] 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 K255과 E256 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 D980과 G981 사이에 삽입된다. 다른 구현예에서, 촉매 도메인은 BhCas12b의 아미노산 K1019과 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 P299과 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.

[1043] 다른 구현예에서, 융합 단백질은 핵 국소화 신호(예를 들어, 이분된 핵 국소화 신호)를 포함한다. 다른 구현예에서, 핵 국소화 신호의 아미노산 서열은 MAPKKKRKVGIHGVPAA이다. 상기 양상의 다른 구현예에서, 핵 국소화 신호는 하기의 서열에 의해 암호화된다:

[1044] 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.

[1045] 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 하기 **표 13B**에 제공된 유전자좌에 삽입된 내부적으로 융합된 TadA*8 도메인과 함께 BhCas12b 도메인을 포함한다.

[1046] [표 13B]

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

[1047]

[1048] 비제한적으로 예를 들면, 아데노신 데아미나제 (예를 들어, ABE8.13)는 BhCas12b에 삽입하여 핵산 서열을 효과적으로 편집하는 융합 단백질 (*예를 들어*, ABE8.13-BhCas12b)을 생성할 수 있다.

[1049] 비제한적이지만 예시적인 융합 단백질은 미국 가출원 제62/852,228호 및 제62/852,224호에 기재되어 있고, 이의 내용은 이들의 전문이 본원에 참조로 포함된다.

[1050] 본 발명의 수행은 달리 지적되지 않는 경우 통상적인 분자 생물학 기술(재조합 기술을 포함하는), 미생물학, 세포 생물학, 생화학 및 면역학 기술을 사용하고, 이들 기술들은 당업자의 범위 내에 있다. 상기 기술은 문헌(참조: 예를 들어, "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")에서 완전하게 설명된다(Coligan, 1991). 이들 기술은 본 발명의 폴리뉴클레오타이드 및 폴리펩타이드의 생성에 적용될 수 있고, 예를 들어, 본 발명을 제조하고 수행하는데 고려될 수 있다. 특정 구현예에 대해 특히 유용한 기술은 하기의 섹션에서 논의된다.

[1051] 하기 실시예들은 본 발명의 검정, 스크리닝 및 치료학적 방법을 제조하고 사용하는 방법에 대한 완전한 개시내용 및 기재를 사용하여 당업자에게 제공하기 위해 제시하는 것이지, 본 발명자들이 본 발명으로 간주하는 범위를 한정하려는 것은 아니다.

[1052] **실시예**

[1053] ***실시예 1: 증가된 편집 효율을 갖는 진화된 아데노신 염기 편집기***

[1054] Tad7.10-dCas9 융합 단백질을 포함하는 염기 편집 시스템은 대략 10 내지 20% 효율로 표적 폴리뉴클레오타이드를 편집할 수 있지만 보다 높은 효율을 요구하는 용도에 대해서는 이들의 사용이 제한될 수 있다. 증가된 효율 및 특이성을 갖는 아데닌 염기 편집기를 동정하기 위해, 아데노신 데아미나제 TadA 7.10을 포함하는 작제물은 오류 성향 PCR로 돌연변이유발 시키고 이어서 핵산 프로그래밍 가능한 DNA 결합 단백질인 dCas9를 암호화하는 핵산 서열에 인접한 발현 벡터에 클로닝하였다 (**도 1a**). 돌연변이유발된 TadA-dCas9 염기 편집기는 이들 실시예에서 ABE8 (아데닌 데아미나제 변이체)로 언급된다. 아데노신 변이체를 포함하는 발현 벡터는 클로람페니콜 내성 (CamR) 및 스트렙토마이신 내성 (SpectR)을 암호화하고 2점 돌연변이 (진화 라운드 7 전략)에 의해 비기능성이 된 가나마이신 내성 유전자를 갖는 선택 플라스미드와 함께 컴피턴트 세균 세포에 동시 형질감염시켰다 (**도 1b**). 세포는 가나마이신 내성의 복구에 대해 선택되었고 이는 아데노신 데아미나제 활성에 대해 판독된다. 아데노신 변이체를 포함하는 발현 벡터는 클로람페니콜 내성 (CamR) 및 스트렙토마이신 내성 (SpectR)을 암호화하고 3점 돌연변이 (진화 라운드 8 전략)에 의해 비기능성이 된 가나마이신 내성 유전자를 갖는 선택 플라스미드와 함께 컴피턴트 세균 세포에 동시 형질감염시켰다 (**도 1c**).

[1055] 불활성화된 가나마이신 내성 유전자 핵산 서열은 하기에 제공된다:

ccggaattgccagctggggcgccctctggtaaggttgggaagccctgcaaagtaaactggat
ggctttcttgccgccaaggatctgatggcgcaggggatcaagatctgatcaagagacaggat
gaggatcct**ttcgcATGATCGAATAAGAT**GGATTGCACGCAGGTTCTCCGGCC**GCTTAGGTG**
GAGCGCCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTT
CCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGA
ATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCA
GCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGG
GCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAA
TGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGC
ATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGA
GCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCG
AGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGC
TTTTCTGGA*TTCATTAACTGTGGCCGGCT*GGGTGTGGCGGACCGCTATCAGGACATAGCGTT
GGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTT
ACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTC
TAA

[1056]

[1057] 상기 서열에서, 소문자는 가나마이신 내성 프로모터 영역을 나타내고, 굵게 표시된 서열은 표적화된 불활성화 부분(Q4* 및 W15*)을 지적하고, 이탤릭 서열은 가나마이신 내성 유전자 (D208N)의 표적화된 불활성 부위를 지칭

하고, 밑줄 친 서열은 PAM 서열을 지칭한다.

[1058] 다시, 세포는 증가하는 가나마이신 농도와 함께 일련의 아가로스 플레이트 상에 분주하였다. 효율적인 염기 편집 활성을 갖는 아데노신 데아미나제 변이체는 가나마이신 내성 유전자에 존재하는 돌연변이를 교정할 수 있었고 추가의 분석을 위해 선택되었다. 세균 세포에서 효율적인 염기 편집을 보여주는 아데노신 데아미나제 변이체 염기 편집기는 **표 14**에 기재된다. 선택된 아데노신 데아미나제 변이체를 포함하는 염기 편집기를 암호화하는 포유동물 발현 벡터를 제조하였다.

[1059] [표 14]

신규 아데닌 염기 편집기 ABE8

플라스미드 ID	기재 내용	기능
280	ABE8.1	단량체_TadA*7.10 + Y147T
281	ABE8.2	단량체_TadA*7.10 + Y147R
282	ABE8.3	단량체_TadA*7.10 + Q154S
283	ABE8.4	단량체_TadA*7.10 + Y123H
284	ABE8.5	단량체_TadA*7.10 + V82S
285	ABE8.6	단량체_TadA*7.10 + T166R
286	ABE8.7	단량체_TadA*7.10 + Q154R
287	ABE8.8	단량체_ Y147R_Q154R_Y123H
288	ABE8.9	단량체_ Y147R_Q154R_I76Y
289	ABE8.10	단량체_ Y147R_Q154R_T166R
290	ABE8.11	단량체_ Y147T_Q154R
291	ABE8.12	단량체_ Y147T_Q154S
292	ABE8.13	단량체_H123Y123H_ Y147R_Q154R_I76Y
293	ABE8.14	이종이량체_TadA*7.10 + Y147T
294	ABE8.15	이종이량체_TadA*7.10 + Y147R
295	ABE8.16	이종이량체_TadA*7.10 + Q154S
296	ABE8.17	이종이량체_TadA*7.10 + Y123H
297	ABE8.18	이종이량체_TadA*7.10 + V82S
298	ABE8.19	이종이량체_TadA*7.10 + T166R
299	ABE8.20	이종이량체_TadA*7.10 + Q154R
300	ABE8.21	이종이량체_ Y147R_Q154R_Y123H
301	ABE8.22	이종이량체_ Y147R_Q154R_I76Y
302	ABE8.23	이종이량체_ Y147R_Q154R_T166R
303	ABE8.24	이종이량체_ Y147T_Q154R
304	ABE8.25	이종이량체_ Y147T_Q154S

[1060]

[1061] ***실시예 2: ABE8을 사용한 알파-1 항-트립신 돌연변이의 교정***

[1062] 선택된 ABE8 작제물은 E342K 돌연변이를 포함하는 A1AT를 발현하는 HEK293 세포 (HEK293T-E342K)에서 염기 편집 활성에 대해 시험하였다. 하나의 접근법에서, HEK293T-E342K 세포는 HEK293 세포에 대해 최적화된 고효율 저독성 DNA 형질감염 시약, 250 ng의 gRNA 플라스미드 및 750 ng의 ABE8 플라스미드 [0407]을 사용한 3 μl:1 μg 비율로 Mirus TransIT293으로 일과성으로 형질감염시켰다. HEK293T-E342K는 2.5 ug Var-3 ABE mRNA 및 1000 ng gRNA 191 길이 20nt를 사용한 네온 전기천공으로 형질감염시켰다. spCas9 염기 편집기에 대해 sgRNA로서 제공된 gRNA 골격은 다음과 같다:

[1063] 5'- GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'

[1064] 기재된 방법에 유용한 gRNA는 하기를 포함한다:

5'-ACCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'

5'-CCAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'

5'-CAUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'

5'-AUCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'

5'-UCGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'

5'-CGACAAGAAAGGGACUGA GUUUUAGAGC UAGAAAUAGC AAGUUAAAAU AAGGCUAGUC CGUUAUCAAC UUGAAAAAGU GGCACCGAGU CGGUGCUUUU-3'

[1065]

[1066] 플라스미드 감염을 위해 4일 및 RNA 전기천공을 위해 2일 후, 게놈 DNA는 0.05% SDS, 25 μg/ml 프로테이나제 K, 10 mM Tris pH 8.0의 단순 용해 완충액으로 추출함에 이어서 85℃에서 열불활성화하였다. 게놈 부위는 PCR 증폭시키고 MiSeq 상에서 서열분석하였다. 결과는 각각의 위치에서 염기 빈도에 대해 및 퍼센트 삽입-결실에 대해 이전에 기재된 바와 같이 분석하였다. 삽입-결실 계산을 위한 세부사항은 국제 PCT 출원 번호 PCT/2017/045381 및 PCT/US2016/058344에 기재되고, 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.

[1067] 선택된 ABE8의 편집 활성(표 15 참조)은 상이한 제조업자 (AxoLabs, Germany and Synthego, Menlo Park, CA)가 제조한 19 또는 20-뉴클레오타이드 가이드 RNA를 사용하여 HEK293T-E342K에서 검정하였다. **도 3, 4a, 및 4b**에 나타낸 바와 같이, ABE8은 대조군 편집기 (AVT686)에 비해 현저한 효율 및 특이성을 보여주었다.

[1068] [표 15]

아데노신 염기 편집기

플라스미드 ID	기재 내용	기능
MSP392	pUTR-trilink-monoTadA-ABE7.10-MQKFRAER	A1AD
MSP393	pUTR-trilink-ABE7.10(Y146T)-MQKFRAER	A1AD
MSP394	pUTR-trilink-ABE7.10(Y146R)-MQKFRAER	A1AD
MSP395	pUTR-trilink-ABE7.10(V81S)-MQKFRAER	A1AD
MSP396	pUTR-trilink-ABE7.10(Q153R)-MQKFRAER	A1AD
MSP397	pUTR-trilink-ABE7.10(Q153S)-MQKFRAER	A1AD
MSP414	pUTR-trilink-ABE7.10(V82S, Y147T)-MQKFRAER	A1AD
MSP415	pUTR-trilink-ABE7.10(V82S, Y147R)-MQKFRAER	A1AD
MSP416	pUTR-trilink-ABE7.10(V82S, Q154R)-MQKFRAER	A1AD
MSP417	pUTR-trilink-ABE7.10(V82S, Q154S)-MQKFRAER	A1AD
MSP418	pUTR-trilink-monoTadA-ABE7.10(V82S, Y147T)-MQKFRAER	A1AD
MSP419	pUTR-trilink-monoTadA-ABE7.10(V82S, Y147R)-MQKFRAER	A1AD
MSP420	pUTR-trilink-monoTadA-ABE7.10(V82S, Q154R)-MQKFRAER	A1AD
MSP421	pUTR-trilink-monoTadA-ABE7.10(V82S, Q154S)-MQKFRAER	A1AD
MSP422	pUTR-trilink-monoTadA-ABE7.10(V82S)-MQKFRAER	A1AD
MSP432	pUTR-trilink-monoTadA-ABE7.10(V82S, H123H, Y147T)-MQKFRAER	A1AD
MSP433	pUTR-trilink-monoTadA-ABE7.10(V82S, H123H)-MQKFRAER	A1AD

[1069]

[1070] 추가로, ABE8은 바이스탠더 A 보다 온-표적 아데닌(A) 염기에 대해 정확한 편집을 제공하고(**도 5**) A1AD 표적 부위에서 고도로 효율적인 치료학적 관련 편집을 가능하게 한다. 특히, ABE8은 **도 5**에 관찰된 바와 같이 A1AD 부위에서 편집 (즉, A·T에서 G·C로의 전환)에서 5배 증가를 유도한다. ABE8을 사용한 E342K의 염기 편집을 통한 정확한 돌연변이 교정은 예를 들어, 순환 AAT 수준을 11 μM 초과로 복구시키고 A1AD에 걸린 대상체에서 폐 및 간 기능 둘 다를 개선시킨다.

[1071] **도 6a-6d**는 계속되는 염기 편집기 가공을 통해 1차 PiZ 섬유아세포에서 핵염기 교정의 개선된 비율을 생성하는 것과 관련된 데이터 및 결과를 보여준다. **도 7a-7d**는 NSG-PiZ 유전자전이 마우스에서 지질 나노입자 (LNP)-매개된 전달 및 염기 편집에 의해 생성된 혈청 A1AT에서의 증가와 관련된 데이터 및 결과를 제공한다.

[1072] **다른 구현예**

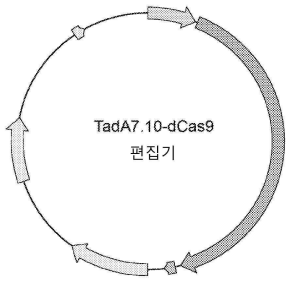
[1073] 이전의 기재로부터, 변화 및 변형이 본원에 기재된 발명에 가해져 이를 다양한 용법 및 조건에 적응하도록 수행될 수 있음은 자명할 것이다. 상기 구현예는 또한 하기의 청구범위 내에 있다.

[1074] 본원에서 변수의 임의의 정의에서 요소들 목록의 언급은 임의의 단일 요소 또는 열거된 요소의 조합(또는 서브조합)으로서 상기 변수의 정의를 포함한다. 본원의 구현예의 언급은 임의의 단일 구현예 또는 임의의 다른 구현예 또는 이의 일부와 조합된 구현예를 포함한다.

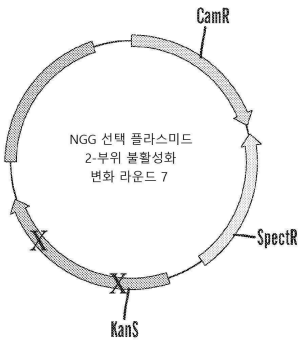
[1075] 본 명세서에 언급된 모든 공보, 특허 및 특허 출원은 본원에서 각각의 개별 공보, 특허 또는 특허 출원이 구체적으로 및 개별적으로 참조로 인용되는 것과 동일한 정도로 참조로 본원에 인용된다. 달리 지적되지 않는 경우, 본 명세서에 언급된 공보, 특허 및 특허 출원은 이들의 전문이 본원에 참조로 포함된다.

도면

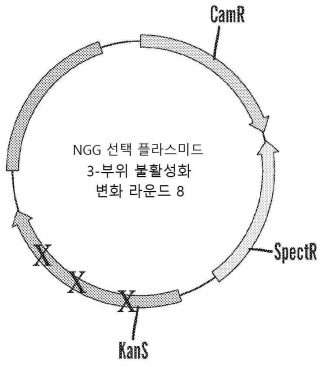
도면1a



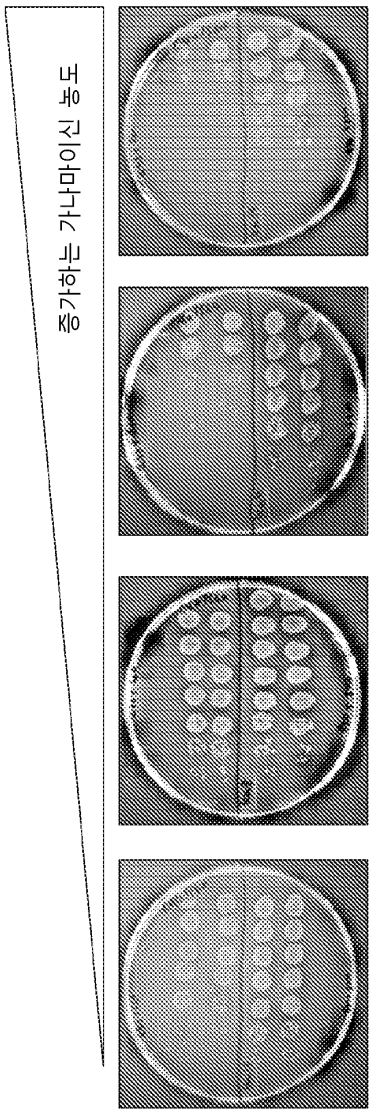
도면1b



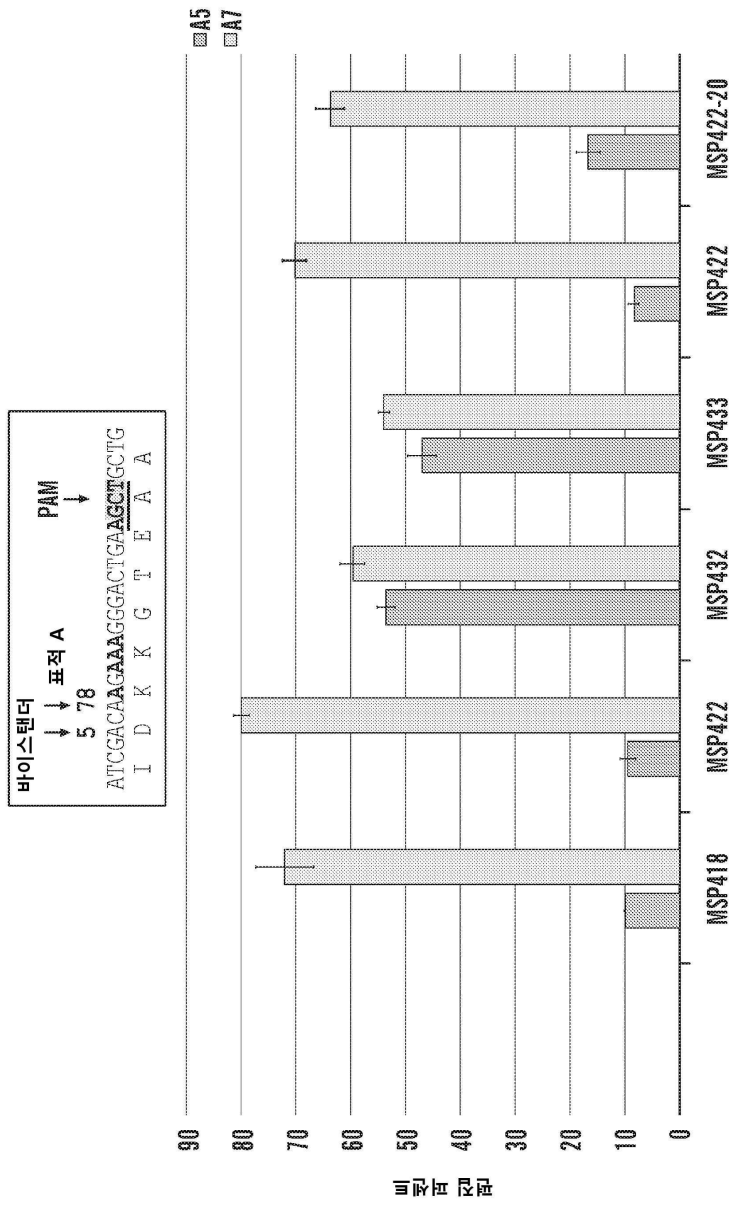
도면1c



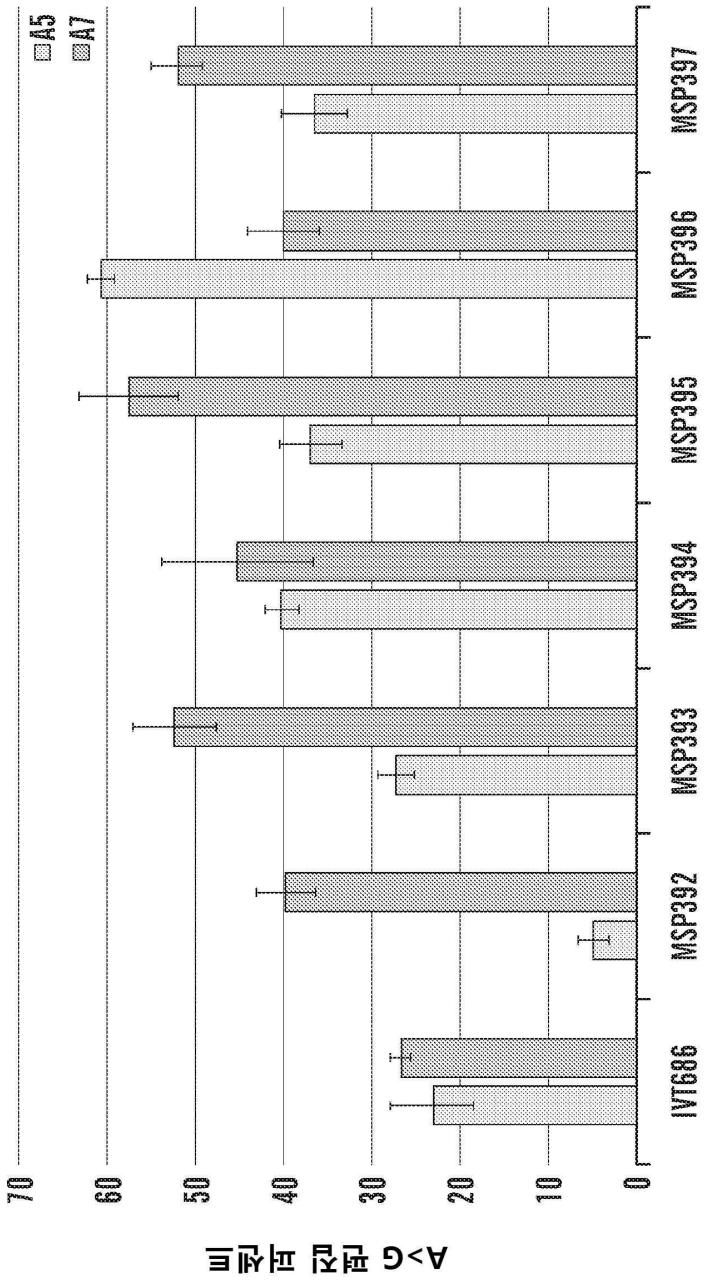
도면2



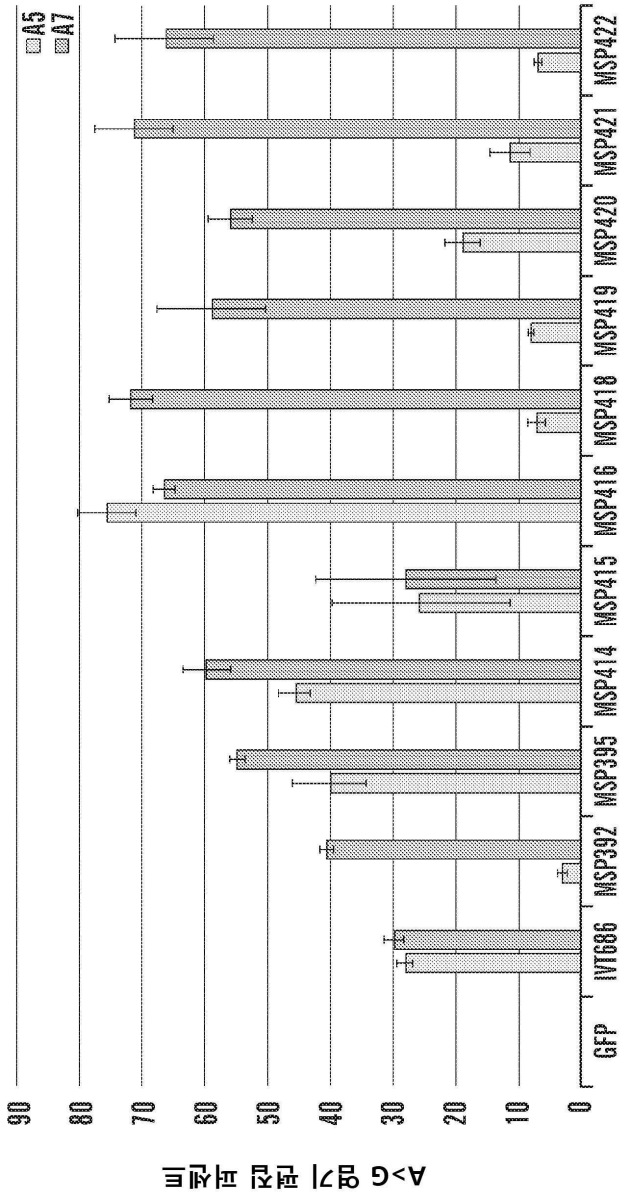
도면3


A5
A7
편집 퍼센트
90
80
70
60
50
40
30
20
10
0
MSP418
MSP422
MSP432
MSP433
MSP422
MSP422-20
바이스탠더
PAM
표적 A
5 78
ATCGACAAGAAAGGGACTGAAGCTGCTG
I D K K G T E A A

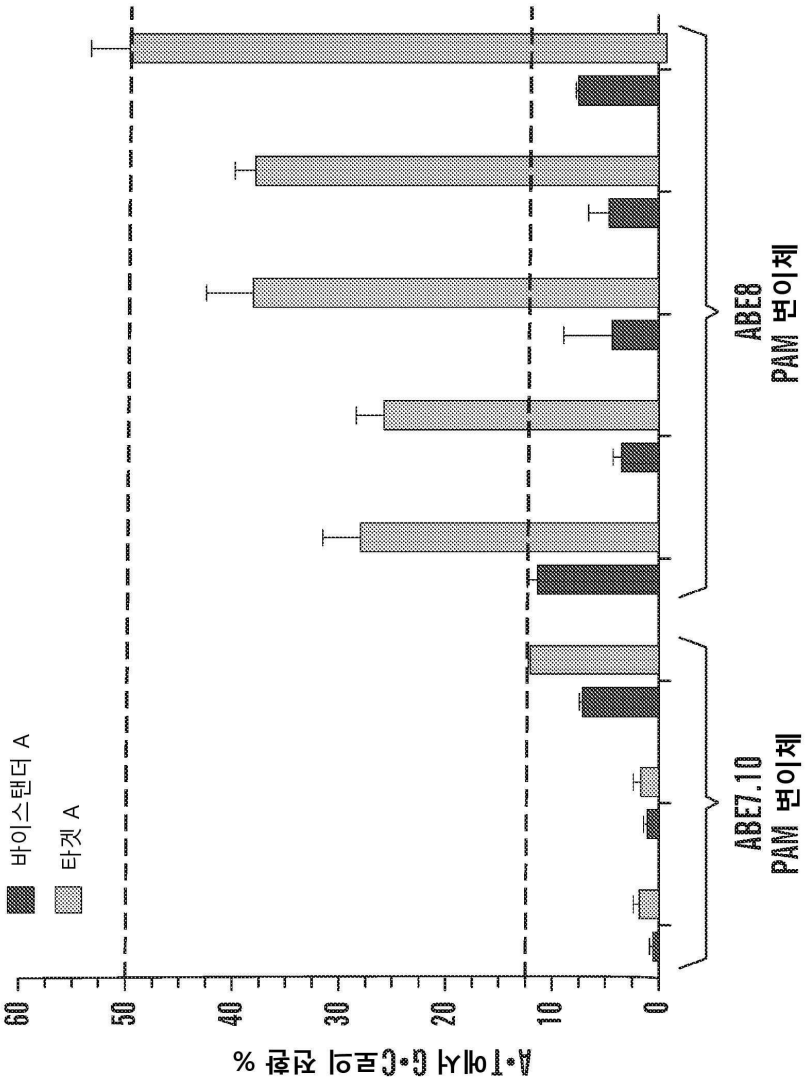
도면4a


A5
A7
MSP397
MSP396
MSP395
MSP394
MSP393
MSP392
IVT686
0
10
20
30
40
50
60
70
A>G 편집 퍼센트

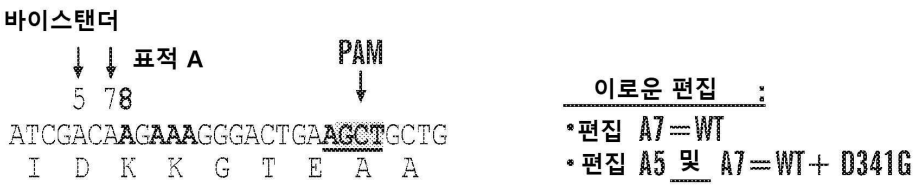
도면4b


A5
A7
GFP
WT686
MSP392
MSP395
MSP414
MSP415
MSP416
MSP418
MSP419
MSP420
MSP421
MSP422
0
10
20
30
40
50
60
70
80
90
A>G 염기 편집 퍼센트

도면5


바이스탠더 A
타겟 A
ABE8
PAM 변이체
ABE7.10
PAM 변이체
A·T에서 C·G로의 전환 %
0
10
20
30
40
50
60

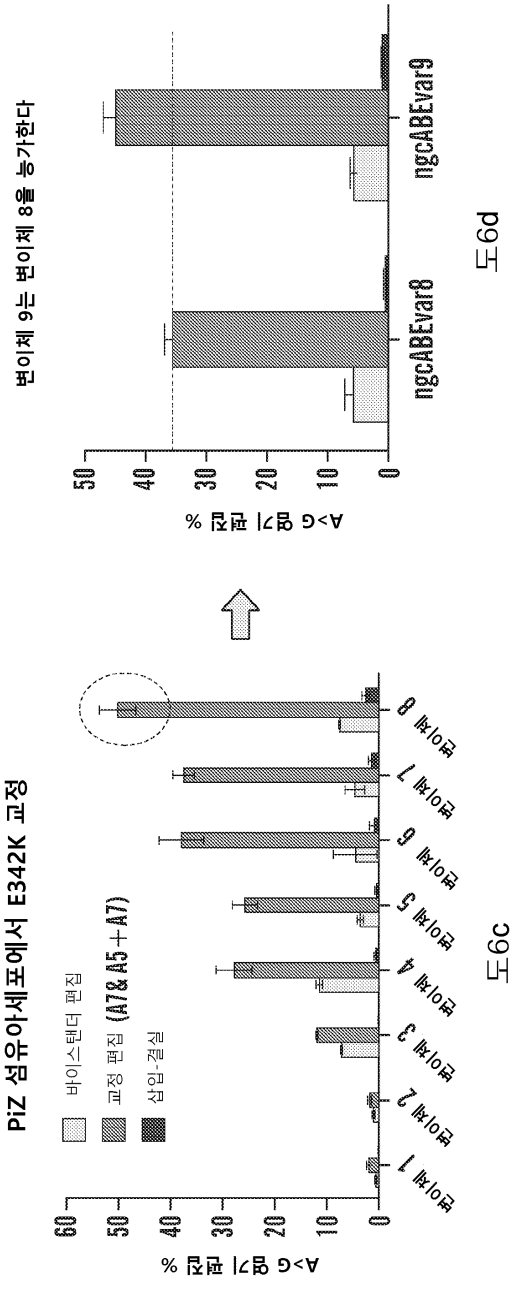
도면6a


바이스탠더
표적 A
PAM
5 78
ATCGACAAGAAAGGGACTGAAGCTGCTG
I D K K G T E A A
이로운 편집 :
•편집 A7=WT
•편집 A5 및 A7=WT+ D341G

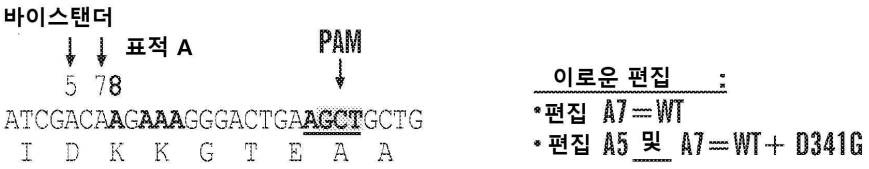
도면6b

코드명	TadA	TadA 변이체	PAM 변이체	날짜
ngcABEvar1	이중	TadA_7.10	MQKSER	Jul-18
ngcABEvar2	이중	TadA_7.10	nureki	Jul-18
ngcABEvar3	이중	TadA_ 7.10	MQKFRAER	Sep-18
ngcABEvar4	이중	TadA_7.10+V82S	MQKFRAER	Dec-18
ngcABEvar5	모노	TadA_7.10+V82S	MQKFRAER	Jan-19
ngcABEvar6	모노	TadA_7.10+I76Y+V82S	MQKFRAER	Jun-19
ngcABEvar7	모노	TadA_7.10+V82T	MQKFRAER	Jun-19
ngcABEvar8	모노	TadA_7.10+I76Y+V82T	MQKFRAER	Jul-19
ngcABEvar9	모노	TadA_7.10+I76Y+V82T +Y147T+Q154S	MQKFRAER	Oct-19

도면6cd



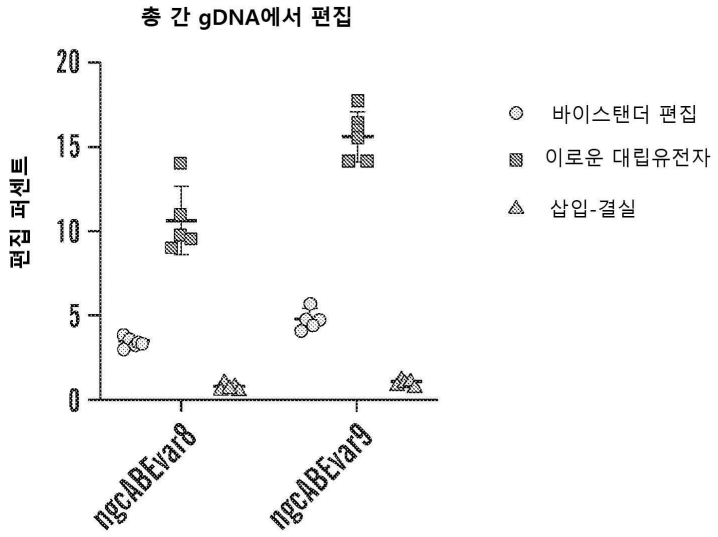
도면7a



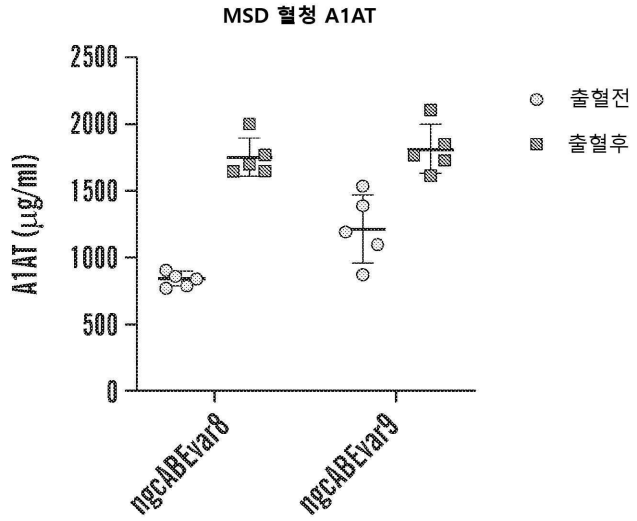
도면7b

코드명	TadA	TadA 변이체	PAM 변이체	날짜
ngcABEvar1	이중	TadA_7.10	MQKSER	Jul-18
ngcABEvar2	이중	TadA_7.10	nureki	Jul-18
ngcABEvar3	이중	TadA_ 7.10	MQKFRAER	Sep-18
ngcABEvar4	이중	TadA_7.10+V82S	MQKFRAER	Dec-18
ngcABEvar5	모노	TadA_7.10+V82S	MQKFRAER	Jan-19
ngcABEvar6	모노	TadA_7.10+I76Y+V82S	MQKFRAER	Jun-19
ngcABEvar7	모노	TadA_7.10+V82T	MQKFRAER	Jun-19
ngcABEvar8	모노	TadA_7.10+I76Y+V82T	MQKFRAER	Jul-19
ngcABEvar9	모노	TadA_7.10+I76Y+V82T +Y147T+Q154S	MQKFRAER	Oct-19

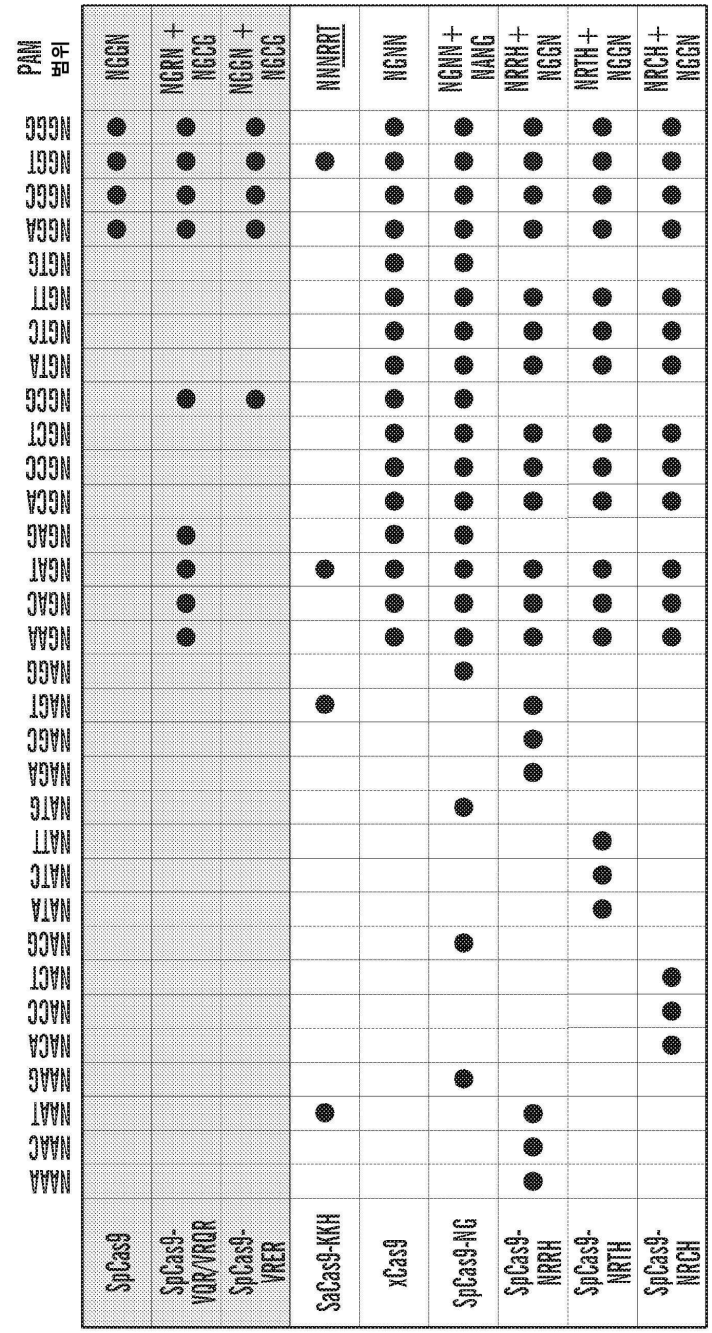
도면7c



도면7d


MSD 혈청 A1AT
2500
2000
1500
1000
500
0
A1AT (μg/ml)
ngcABEvar8
ngcABEvar9
출혈전
출혈후

도면8



서 열 목 록

SEQUENCE LISTING

<110> BEAM THERAPEUTICS INC.

<120> COMPOSITIONS AND METHODS FOR TREATING ALPHA-1 ANTITRYPSIN DEFICIENCY

<130> 52885-788.602

<140> PCT/US2020/018195

<141> 2020-02-13

<150> 62/966,526

<151> 2020-01-27

<150> 62/941,569

<151> 2019-11-27

<150> 62/931,722

<151> 2019-11-06

<150> 62/852,228

<151> 2019-05-23

<150> 62/852,224

<151> 2019-05-23

<150> 62/805,271

<151> 2019-02-13

<150> 62/805,238

<151> 2019-02-13

<160> 148

<170> PatentIn version 3.5

<210> 1

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 1

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> 2

<211> 167

<212> PRT

<213> Unknown

<220

><223> Description of Unknown:

TadA sequence

<400> 2

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> 3
<211> 167
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 3
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> 4

<211> 1410

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 4

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> 5

<211> 103

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 5

accaucgaca agaaagggac ugaguuuuag agcuagaaau agcaaguuaa aauaaggcua 60

guccguuauc aacuugaaaa aguggcaccg agucggugcu uuu 103

<210> 6

<211> 102

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 6

ccaucgacaa gaaagggacu gaguuuuaga gcuagaaaua gcaaguuaaa auaaggcuag 60

uccguuauca acuugaaaaa guggcaccga gucggugcuu uu 102

<210> 7

<211> 101

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 7

caucgacaag aaagggacug aguuuuagag cuagaaauag caaguuaaaa uaaggcuagu 60

ccguuaucaa cuugaaaaag uggcaccgag ucggugcuuu u 101

<210> 8

<211> 100

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 8

aucgacaaga aagggacuga guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 9

<211> 99

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 9

ucgacaagaa agggacugag uuuuagagcu agaaauagca aguuaaaaua aggcuagucc 60

guuaucaacu ugaaaaagug gcaccgaguc ggugcuuuu 99

<210> 10

<211> 98

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 10

cgacaagaaa gggacugagu uuuagagcua gaaauagcaa guuaaaauaa ggcuaguccg 60

uuaucaacuu gaaaaagugg caccgagucg gugcuuuu 98

<210> 11

<211> 80

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 11

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

<210> 12

<211> 167

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 12

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Thr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp
165

<210> 13

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 13

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> 14

<211> 160

<212> PRT

<213> Staphylococcus aureus

<400> 14

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> 15
<211> 161
<212> PRT
<213> Bacillus subtilis
<400> 15

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> 16

<211> 183

<212> PRT

<213> Salmonella typhimurium

<400> 16

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> 17
<211> 164
<212> PRT
<213> Shewanella putrefaciens
<400> 17
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> 18

<211> 173

<212> PRT

<213> Haemophilus influenzae

<400> 18

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> 19

<211> 161

<212> PRT

<213> Caulobacter crescentus

<400> 19

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> 20

<211> 179

<212> PRT

<213> Geobacter sulfurreducens

<400> 20

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> 21
<211> 418
<212> PRT
<213> Homo sapiens
<400> 21
Met Pro Ser Ser Val Ser Trp Gly Ile Leu Leu Leu Ala Gly Leu Cys
1 5 10 15
Cys Leu Val Pro Val Ser Leu Ala Glu Asp Pro Gln Gly Asp Ala Ala

20 25 30

Gln Lys Thr Asp Thr Ser His His Asp Gln Asp His Pro Thr Phe Asn

35 40 45

Lys Ile Thr Pro Asn Leu Ala Glu Phe Ala Phe Ser Leu Tyr Arg Gln

50 55 60

Leu Ala His Gln Ser Asn Ser Thr Asn Ile Phe Phe Ser Pro Val Ser

65 70 75 80

Ile Ala Thr Ala Phe Ala Met Leu Ser Leu Gly Thr Lys Ala Asp Thr

85 90 95

His Asp Glu Ile Leu Glu Gly Leu Asn Phe Asn Leu Thr Glu Ile Pro

100 105 110

Glu Ala Gln Ile His Glu Gly Phe Gln Glu Leu Leu Arg Thr Leu Asn

115 120 125

Gln Pro Asp Ser Gln Leu Gln Leu Thr Thr Gly Asn Gly Leu Phe Leu

130 135 140

Ser Glu Gly Leu Lys Leu Val Asp Lys Phe Leu Glu Asp Val Lys Lys

145 150 155 160

Leu Tyr His Ser Glu Ala Phe Thr Val Asn Phe Gly Asp Thr Glu Glu

165 170 175

Ala Lys Lys Gln Ile Asn Asp Tyr Val Glu Lys Gly Thr Gln Gly Lys

180 185 190

Ile Val Asp Leu Val Lys Glu Leu Asp Arg Asp Thr Val Phe Ala Leu

195 200 205

Val Asn Tyr Ile Phe Phe Lys Gly Lys Trp Glu Arg Pro Phe Glu Val

210 215 220

Lys Asp Thr Glu Glu Glu Asp Phe His Val Asp Gln Val Thr Thr Val

225 230 235 240

Lys Val Pro Met Met Lys Arg Leu Gly Met Phe Asn Ile Gln His Cys

245 250 255

Lys Lys Leu Ser Ser Trp Val Leu Leu Met Lys Tyr Leu Gly Asn Ala

260 265 270

Thr Ala Ile Phe Phe Leu Pro Asp Glu Gly Lys Leu Gln His Leu Glu

275 280 285

Asn Glu Leu Thr His Asp Ile Ile Thr Lys Phe Leu Glu Asn Glu Asp

290 295 300

Arg Arg Ser Ala Ser Leu His Leu Pro Lys Leu Ser Ile Thr Gly Thr

305 310 315 320

Tyr Asp Leu Lys Ser Val Leu Gly Gln Leu Gly Ile Thr Lys Val Phe

325 330 335

Ser Asn Gly Ala Asp Leu Ser Gly Val Thr Glu Glu Ala Pro Leu Lys

340 345 350

Leu Ser Lys Ala Val His Lys Ala Val Leu Thr Ile Asp Glu Lys Gly

355 360 365

Thr Glu Ala Ala Gly Ala Met Phe Leu Glu Ala Ile Pro Met Ser Ile

370 375 380

Pro Pro Glu Val Lys Phe Asn Lys Pro Phe Val Phe Leu Met Ile Glu

385 390 395 400

Gln Asn Thr Lys Ser Pro Leu Phe Met Gly Lys Val Val Asn Pro Thr

405 410 415

Gln Lys

<210> 22

<211> 8811

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 22

atatgccaag tacgcccccct attgacgtca atgacggtaa atggcccgcc tggcattatg 60

cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg 120

ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact 180

cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa 240

atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta 300

ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct 360

agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat gaaacggaca 420

gccgacggaa gcgagttcga gtcaccaaag aagaagcgga aagtctctga agtcgagttt 480

agccacgagt attggatgag gcacgcactg accctggcaa agcgagcatg ggatgaaaga 540

gaagtccccg tgggcgccgt gctggtgcac aacaatagag tgatcggaga gggatggaac 600

aggccaatcg gccgccacga ccctaccgca cacgcagaga tcatggcact gaggcaggga 660

ggcctggtca tgcagaatta ccgcctgatc gatgccaccc tgtatgtgac actggagcca 720

tgcgtgatgt gcgcaggagc aatgatccac agcaggatcg gaagagtggt gttcggagca 780

cgggacgcca agaccggcgc agcaggctcc ctgatggatg tgctgcacca ccccggcatg 840

aaccaccggg tggagatcac agagggaatc ctggcagacg agtgcgccgc cctgctgagc 900

gatttcttta gaatgcggag acaggagatc aaggcccaga agaaggcaca gagctccacc 960

gactctggag gatctagcgg aggatcctct ggaagcgaga caccaggcac aagcgagtcc 1020

gccacaccag agagctccgg cggctcctcc ggaggatcct ctgaggtgga gttttcccac 1080

gagtactgga tgagacatgc cctgaccctg gccaagaggg cacgcgatga gagggaggtg 1140

cctgtgggag ccgtgctggt gctgaacaat agagtgatcg gcgagggctg gaacagagcc 1200

atcggcctgc acgacccaac agcccatgcc gaaattatgg ccctgagaca gggcggcctg 1260

gtcatgcaga actacagact gattgacgcc accctgtacg tgacattcga gccttgcgtg 1320

atgtgcgccg gcgccatgat ccactctagg atcggccgcg tggtgtttgg cgtgaggaac 1380

gcaaaaaccg gcgccgcagg ctccctgatg gacgtgctgc actaccccgg catgaatcac 1440

cgcgtcgaaa ttaccgaggg aatcctggca gatgaatgtg ccgccctgct gtgctatttc 1500

tttcggatgc ctagacaggt gttcaatgct cagaagaagg cccagagctc caccgactcc 1560

ggaggatcta gcggaggctc ctctggctct gagacacctg gcacaagcga gagcgcaaca 1620

cctgaaagca gcgggggcag cagcgggggg tcagacaaga agtacagcat cggcctggcc 1680

atcggcacca actctgtggg ctgggccgtg atcaccgacg agtacaaggt gcccagcaag 1740

aaattcaagg tgctgggcaa caccgaccgg cacagcatca agaagaacct gatcggagcc 1800

ctgctgttcg acagcggcga aacagccgag gccacccggc tgaagagaac cgccagaaga 1860

agatacacca gacggaagaa ccggatctgc tatctgcaag agatcttcag caacgagatg 1920

gccaaggtgg acgacagctt cttccacaga ctggaagagt ccttcctggt ggaagaggat 1980

aagaagcacg agcggcaccc catcttcggc aacatcgtgg acgaggtggc ctaccacgag 2040

aagtacccca ccatctacca cctgagaaag aaactggtgg acagcaccga caaggccgac 2100

ctgcggctga tctatctggc cctggcccac atgatcaagt tccggggcca cttcctgatc 2160

gagggcgacc tgaaccccga caacagcgac gtggacaagc tgttcatcca gctggtgcag 2220

acctacaacc agctgttcga ggaaaacccc atcaacgcca gcggcgtgga cgccaaggcc 2280

atcctgtctg ccagactgag caagagcaga cggctggaaa atctgatcgc ccagctgccc 2340

ggcgagaaga agaatggcct gttcggaaac ctgattgccc tgagcctggg cctgaccccc 2400

aacttcaaga gcaacttcga cctggccgag gatgccaaac tgcagctgag caaggacacc 2460

tacgacgacg acctggacaa cctgctggcc cagatcggcg accagtacgc cgacctgttt 2520

ctggccgcca agaacctgtc cgacgccatc ctgctgagcg acatcctgag agtgaacacc 2580

gagatcacca aggccccct gagcgcctct atgatcaaga gatacgacga gcaccaccag 2640

gacctgaccc tgctgaaagc tctcgtgcgg cagcagctgc ctgagaagta caaagagatt 2700

ttcttcgacc agagcaagaa cggctacgcc ggctacattg acggcggagc cagccaggaa 2760

gagttctaca agttcatcaa gcccatcctg gaaaagatgg acggcaccga ggaactgctc 2820

gtgaagctga acagagagga cctgctgcgg aagcagcgga ccttcgacaa cggcagcatc 2880

ccccaccaga tccacctggg agagctgcac gccattctgc ggcggcagga agatttttac 2940

ccattcctga aggacaaccg ggaaaagatc gagaagatcc tgaccttccg catcccctac 3000

tacgtgggcc ctctggccag gggaaacagc agattcgcct ggatgaccag aaagagcgag 3060

gaaaccatca ccccctggaa cttcgaggaa gtggtggaca agggcgcttc cgcccagagc 3120

ttcatcgagc ggatgaccaa cttcgataag aacctgccca acgagaaggt gctgcccaag 3180

cacagcctgc tgtacgagta cttcaccgtg tataacgagc tgaccaaagt gaaatacgtg 3240

accgagggaa tgagaaagcc cgccttcctg agcggcgagc agaaaaaggc catcgtggac 3300

ctgctgttca agaccaaccg gaaagtgacc gtgaagcagc tgaaagagga ctacttcaag 3360

aaaatcgagt gcttcgactc cgtggaaatc tccggcgtgg aagatcggtt caacgcctcc 3420

ctgggcacat accacgatct gctgaaaatt atcaaggaca aggacttcct ggacaatgag 3480

gaaaacgagg acattctgga agatatcgtg ctgaccctga cactgtttga ggacagagag 3540

atgatcgagg aacggctgaa aacctatgcc cacctgttcg acgacaaagt gatgaagcag 3600

ctgaagcggc ggagatacac cggctggggc aggctgagcc ggaagctgat caacggcatc 3660

cgggacaagc agtccggcaa gacaatcctg gatttcctga agtccgacgg cttcgccaac 3720

agaaacttca tgcagctgat ccacgacgac agcctgacct ttaaagagga catccagaaa 3780

gcccaggtgt ccggccaggg cgatagcctg cacgagcaca ttgccaatct ggccggcagc 3840

cccgccatta agaagggcat cctgcagaca gtgaaggtgg tggacgagct cgtgaaagtg 3900

atgggccggc acaagcccga gaacatcgtg atcgaaatgg ccagagagaa ccagaccacc 3960

cagaagggac agaagaacag ccgcgagaga atgaagcgga tcgaagaggg catcaaagag 4020

ctgggcagcc agatcctgaa agaacacccc gtggaaaaca cccagctgca gaacgagaag 4080
ctgtacctgt actacctgca gaatgggcgg gatatgtacg tggaccagga actggacatc 4140
aaccggctgt ccgactacga tgtggaccat atcgtgcctc agagctttct gaaggacgac 4200
tccatcgaca acaaggtgct gaccagaagc gacaagaacc ggggcaagag cgacaacgtg 4260
ccctccgaag aggtcgtgaa gaagatgaag aactactggc ggcagctgct gaacgccaag 4320
ctgattaccc agagaaagtt cgacaatctg accaaggccg agagaggcgg cctgagcgaa 4380
ctggataagg ccggcttcat caagagacag ctggtggaaa cccggcagat cacaaagcac 4440

gtggcacaga tcctggactc ccggatgaac actaagtacg acgagaatga caagctgatc 4500
cgggaagtga aagtgatcac cctgaagtcc aagctggtgt ccgatttccg gaaggatttc 4560
cagttttaca aagtgcgcga gatcaacaac taccaccacg cccacgacgc ctacctgaac 4620
gccgtcgtgg gaaccgccct gatcaaaaag taccctaagc tggaaagcga gttcgtgtac 4680
ggcgactaca aggtgtacga cgtgcggaag atgatcgcca agagcgagca ggaaatcggc 4740
aaggctaccg ccaagtactt cttctacagc aacatcatga acttttcaa gaccgagatt 4800
accctggcca acggcgagat ccggaagcgg cctctgatcg agacaaacgg cgaaaccggg 4860

gagatcgtgt gggataaggg ccgggatttt gccaccgtgc ggaaagtgct gagcatgccc 4920
caagtgaata tcgtgaaaaa gaccgaggtg cagacaggcg gcttcagcaa agagtctatc 4980
ctgcccaaga ggaacagcga taagctgatc gccagaaaga aggactggga ccctaagaag 5040
tacggcggct tcgacagccc caccgtggcc tattctgtgc tggtggtggc caaagtggaa 5100
aagggcaagt ccaagaaact gaagagtgtg aaagagctgc tggggatcac catcatggaa 5160
agaagcagct tcgagaagaa tcccatcgac tttctggaag ccaagggcta caaagaagtg 5220
aaaaaggacc tgatcatcaa gctgcctaag tactccctgt tcgagctgga aaacggccgg 5280

aagagaatgc tggcctctgc cggcgaactg cagaagggaa acgaactggc cctgccctcc 5340
aaatatgtga acttcctgta cctggccagc cactatgaga agctgaaggg ctcccccgag 5400
gataatgagc agaaacagct gtttgtggaa cagcacaagc actacctgga cgagatcatc 5460
gagcagatca gcgagttctc caagagagtg atcctggccg acgctaatct ggacaaagtg 5520
ctgtccgcct acaacaagca ccgggataag cccatcagag agcaggccga gaatatcatc 5580
cacctgttta ccctgaccaa tctgggagcc cctgccgcct tcaagtactt tgacaccacc 5640
atcgaccgga agaggtacac cagcaccaaa gaggtgctgg acgccaccct gatccaccag 5700

agcatcaccg gcctgtacga gacacggatc gacctgtctc agctgggagg tgactctggc 5760

ggctcaaaaa gaaccgccga cggcagcgaa ttcgagccca agaagaagag gaaagtctaa 5820

ccggtcatca tcaccatcac cattgagttt aaacccgctg atcagcctcg actgtgcctt 5880

ctagttgcca gccatctgtt gtttgcccct cccccgtgcc ttccttgacc ctggaaggtg 5940

ccactcccac tgtcctttcc taataaaatg aggaaattgc atcgcattgt ctgagtaggt 6000

gtcattctat tctgggggggt ggggtggggc aggacagcaa gggggaggat tgggaagaca 6060

atagcaggca tgctggggat gcggtgggct ctatggcttc tgaggcggaa agaaccagct 6120

ggggctcgat accgtcgacc tctagctaga gcttggcgta atcatggtca tagctgtttc 6180

ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat acgagccgga agcataaagt 6240

gtaaagccta gggtgcctaa tgagtgagct aactcacatt aattgcgttg cgctcactgc 6300

ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc caacgcgcgg 6360

ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc gctcactgac tcgctgcgct 6420

cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa ggcggtaata cggttatcca 6480

cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga 6540

accgtaaaaa ggccgcgttg ctggcgtttt tccataggct ccgcccccct gacgagcatc 6600

acaaaaatcg acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg 6660

cgtttccccc tggaagctcc ctcgtgcgct ctcctgttcc gaccctgccg cttaccggat 6720

acctgtccgc ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt 6780

atctcagttc ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa ccccccgttc 6840

agcccgaccg ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacacg 6900

acttatcgcc actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg 6960

gtgctacaga gttcttgaag tggtggccta actacggcta cactagaaga acagtatttg 7020

gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg 7080

gcaaacaaac caccgctggt agcggtggtt tttttgtttg caagcagcag attacgcgca 7140

gaaaaaaagg atctcaagaa gatcctttga tcttttctac ggggtctgac actcagtgga 7200

acgaaaactc acgttaaggg attttggtca tgagattatc aaaaaggatc ttcacctaga 7260

tccttttaaa ttaaaaatga agttttaaat caatctaaag tatatatgag taaacttggt 7320

ctgacagtta ccaatgctta atcagtgagg cacctatctc agcgatctgt ctatttcgtt 7380

catccatagt tgcctgactc cccgtcgtgt agataactac gatacgggag ggcttaccat 7440

ctggccccag tgctgcaatg ataccgcgag acccacgctc accggctcca gatttatcag 7500

caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact ttatccgcct 7560

ccatccagtc tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt 7620

tgcgcaacgt tgttgccatt gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg 7680

cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatccccc atgttgtgca 7740

aaaaagcggt tagctccttc ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt 7800

tatcactcat ggttatggca gcactgcata attctcttac tgtcatgcca tccgtaagat 7860

gcttttctgt gactggtgag tactcaacca agtcattctg agaatagtgt atgcggcgac 7920

cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa 7980

aagtgctcat cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt 8040

tgagatccag ttcgatgtaa cccactcgtg cacccaactg atcttcagca tcttttactt 8100

tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa 8160

gggcgacacg gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt 8220

atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa 8280

taggggttcc gcgcacattt ccccgaaaag tgccacctga cgtcgacgga tcgggagatc 8340

gatctcccga tcccctaggg tcgactctca gtacaatctg ctctgatgcc gcatagttaa 8400

gccagtatct gctccctgct tgtgtgttgg aggtcgctga gtagtgcgcg agcaaaattt 8460

aagctacaac aaggcaaggc ttgaccgaca attgcatgaa gaatctgctt agggttaggc 8520

gttttgcgct gcttcgcgat gtacgggcca gatatacgcg ttgacattga ttattgacta 8580

gttattaata gtaatcaatt acggggtcat tagttcatag cccatatatg gagttccgcg 8640

ttacataact tacggtaaat ggcccgcctg gctgaccgcc caacgacccc cgcccattga 8700

cgtcaataat gacgtatgtt cccatagtaa cgccaatagg gactttccat tgacgtcaat 8760

gggtggagta tttacggtaa actgcccact tggcagtaca tcaagtgtat c 8811

<210> 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> 1082

<212> PRT

<213> Neisseria meningitidis

<400> 28

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> 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 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> 30

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 30

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> 31
<211> 1367
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 31

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> 32
<211> 1367
<212> PRT
<213> Streptococcus pyogenes
<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 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> 33

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 33

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> 34

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 34

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> 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> 306

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

DnaE Intein-N sequence

<400> 40

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> 41

<211> 102

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

DnaE Intein-N sequence

<400> 41

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> 42

<211> 108

<212> DNA

<213> Unknown

<220><223> Description of Unknown:

DnaE Intein-C sequence

<400> 42

atgatcaaga tagctacaag gaagtatctt ggcaaacaaa acgtttatga tattggagtc 60

gaaagagatc acaactttgc tctgaagaac ggattcatag cttctaat 108

<210> 43

<211> 36

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Intein-C sequence

<400> 43

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> 44

<211> 303

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 44

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> 45

<211> 101

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 45

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> 46

<211> 159

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 46

atgaagagga ctgccgatgg atcagagttt gaatctccca agaagaagag gaaagtaaag 60

ataatatctc gaaaaagtct tggtacccaa aatgtctatg atattggagt ggagaaagat 120

cacaacttcc ttctcaagaa cggtctcgta gccagcaac 159

<210> 47

<211> 53

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 47

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> 48

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 48

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> 49

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 49

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> 50

<211> 104

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 50

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> 51

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 51

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1 5 10 15

<210> 52

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 52

Ser Gly Gly Ser

1

<210> 53

<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> 53

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> 54
<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> 54
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> 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 "Gly Gly Gly Gly Ser" repeating units

<400> 55

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> 56
<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> 56

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
1 5 10 15
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
20 25 30

<210> 57
<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> 57

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> 58
<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> 58
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> 59

<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> 59

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

1 5 10 15

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

20 25 30

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

35 40 45

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

50 55 60

<210> 60

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 60

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> 61

<211> 40

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 61

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> 62

<211> 64

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 62

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> 63

<211> 92

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 63

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> 64

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 64

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg

1 5 10 15

Lys Val

<210> 65

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 65

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 66

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 66

Lys Lys Thr Glu Leu Gln Thr Thr Asn Ala Glu Asn Lys Thr Lys Lys

1 5 10 15

Leu

<210> 67

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 67

Lys Arg Gly Ile Asn Asp Arg Asn Phe Trp Arg Gly Glu Asn Gly Arg

1 5 10 15

Lys Thr Arg

<210> 68

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 68

Arg Lys Ser Gly Lys Ile Ala Ala Ile Val Val Lys Arg Pro Arg Lys

1 5 10 15

<210> 69

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 69

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 70

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 70

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> 71

<211> 3184

<212> DNA

<213> Homo sapiens

<400> 71

acaatgactc ctttcggtaa gtgcagtgga agctgtacac tgcccaggca aagcgtccgg 60

gcagcgtagg cgggcgactc agatcccagc cagtggactt agcccctgtt tgctcctccg 120

ataactgggg tgaccttggt taatattcac cagcagcctc ccccgttgcc cctctggatc 180

cactgcttaa atacggacga ggacagggcc ctgtctcctc agcttcaggc accaccactg 240

acctgggaca gtgaatcgac aatgccgtct tctgtctcgt ggggcatcct cctgctggca 300

ggcctgtgct gcctggtccc tgtctccctg gctgaggatc cccagggaga tgctgcccag 360

aagacagata catcccacca tgatcaggat cacccaacct tcaacaagat caccccccaac 420

ctggctgagt tcgccttcag cctataccgc cagctggcac accagtccaa cagcaccaat 480

atcttcttct ccccagtgag catcgctaca gcctttgcaa tgctctccct ggggaccaag 540

gctgacactc acgatgaaat cctggagggc ctgaatttca acctcacgga gattccggag 600

gctcagatcc atgaaggctt ccaggaactc ctccgtaccc tcaaccagcc agacagccag 660

ctccagctga ccaccggcaa tggcctgttc ctcagcgagg gcctgaagct agtggataag 720

tttttggagg atgttaaaaa gttgtaccac tcagaagcct tcactgtcaa cttcggggac 780

accgaagagg ccaagaaaca gatcaacgat tacgtggaga agggtactca agggaaaatt 840

gtggatttgg tcaaggagct tgacagagac acagtttttg ctctggtgaa ttacatcttc 900

tttaaaggca aatgggagag acccctttgaa gtcaaggaca ccgaggaaga ggacttccac 960

gtggaccagg tgaccaccgt gaaggtgcct atgatgaagc gtttaggcat gtttaacatc 1020

cagcactgta agaagctgtc cagctgggtg ctgctgatga aatacctggg aatgccaccg 1080

ccatcttctt cctgcctgat gaggggaaac tacagcacct ggaaaatgac tcacccacga 1140

tatcatcacc aagttcctgg aaaatgaaga cagaaggtct gccagcttca tttacccaaa 1200

ctgtccatta ctggaaccta tgatctgaag agcgtcctgg gtcaactggc atcactaagg 1260

tcttcagcaa tggggctgac ctctccgggg tcacagagga ggcaccctga agctctccaa 1320

ggccgtgcat aaggctgtgc tgaccatcga cgagaaaggg actgagctgc tggggccatg 1380

tttttagagg ccatacccat gtctatcccc cccgaggtca agttaacaaa ccctttgtct 1440

tcttaatgat tgaacaaaat accaagtctc ccctcttcat gggaaagtgg tgaatcccac 1500

ccaaaaataa ctgcctctcg ctcctcaacc cctcccctcc atcctggccc cctccctgga 1560

tgacattaaa gaagggttga gctggtccct gcctgcatgt gctgtaaatc cctcccatgt 1620

tttctctgag tctccctttg cctgctgagg ctgtatgtgg ctccaggtaa cagtgctgtc 1680

ttcgggcccc ctgaactgtg ttcatggagc atctggctgg taggcacatg ctgggcttga 1740

atccaggggg gactgaatcc tcagcttacg gacctgggcc atctgtttct ggagggctcc 1800

agtcttcctt gtcctgtctt ggagtcccca agaaggatca caggggagga accagatacc 1860

agccatgacc ccaggctcca ccaagcatct tcatgtcccc tgctcatccc ccactccccc 1920

ccacccagag ttgctcatcc tgccagggct ggctggccca ccccaaggct gccctcctgg 1980

gggccccaga actgcctgat cgtgccgtgg cccattttgt ggcatctgca gcaacacaag 2040

agagaggaca atgtcctcct cttgacccgc tgtacctaac cagactcggg ccctgcacct 2100

ctcaggcact tctggaaaat gactgaggca gatcttcctg aagcccattc tccatggggc 2160

aacaaggaca cctattctgt ccttgtcctt catcgctgcc ccagaaagcc tcacatatct 2220

ccgtttagaa tcaggtccct tctccccaga gaagaggagg gtctctgctt tgttttctct 2280

atctcctcct cagacttgac caggcccaga ggccccagaa gaccattacc ctatatccct 2340

tctcctccct agtcacatgg ccataggctg ctgatggctc aggaaggcca ttgcaaggac 2400

tcctcagcta tgggagagga agcacatacc cattgacccc cgcaacccct ccctttcctc 2460

ctctgagtcc cgactggggc cacatgagcc tgacttcttt gtgcctgttg ctgtccctgc 2520

agtcttcaga gggccaccgc agctcagtgc cacggcagga ggctgttcct gaatagcccc 2580

tgtggtaagg gccaggagag tcctccatcc tccaaggccc tgctaaagga cacagcagcc 2640

aggaagtccc ctgggcccct agcgaaggac agcctgctcc ctccgtctct accaggaatg 2700

gccttgtcct atggaaggca ctccccatcc caaactaatc taggaatcac tgtctaacca 2760

ctcactgtca tgaatgtgta ctaaaggatg aggttgagtc ataccaaata gtgatttcga 2820

tagttcaaaa tggtgaaatt gcaattctac atgattcagt ctaatcaatg gataccgact 2880

gtttcccaca caagtctccg ttctcttaag cttactcact gacagccttt cactctccac 2940

aaatacatta aagatatgcc atcaccaagc cccctaggat gacaccagac ctgagagtct 3000

gaagacctgg atccaagtct gacttttccc cctgacagct gtgtgacctt cgtgaagtcg 3060

ccaaacctct ctgagcccag tcattgctag taagacctgc ctttgagttg gtatgatgtt 3120

caagttagat aacaaatgtt tatacccatt agaacagaga ataaatagaa ctacatttct 3180

tgca 3184

<210> 72

<211> 84

<212> PRT

<213> Bacillus phage PBS2

<400> 72

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> 73

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 73

gacaagaaag ggacugaagc 20

<210> 74

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 74

aucgacaaga aagggacuga 20

<210> 75

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 75

acacaccggu ugguggccuc 20

<210> 76

<211> 987

<212> PRT

<213> Unknown

<220><223> Description of Unknown: Deltaproteobacteria CasX sequence

<400> 76

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> 77
<211> 1300
<212> PRT
<213> Francisella novicida
<400> 77
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> 78

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 78

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> 79

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 79

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> 80
<211> 1300
<212> PRT
<213> Francisella novicida
<400> 80

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> 81

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 81

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> 82

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 82

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> 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 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> 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 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> 85

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 85

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> 86

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 86

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> 87

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 87

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> 88

<211> 1140

<212> PRT

<213> Bacillus hisashii

<400> 88

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
85 90 95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
100 105 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
115 120 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
130 135 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145 150 155 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
165 170 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
180 185 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
195 200 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
210 215 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

225 230 235 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

245 250 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys

260 265 270

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg

275 280 285

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu

290 295 300

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

305 310 315 320

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys

325 330 335

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr

340 345 350

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro

355 360 365

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

370 375 380

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn

385 390 395 400

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn

405 410 415

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys

420 425 430

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

435 440 445

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg

450 455 460

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His

465 470 475 480

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr

485 490 495

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

500 505 510

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met

515 520 525

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys

530 535 540

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu

545 550 555 560

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

565 570 575

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly

580 585 590

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys

595 600 605

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu

610 615 620

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr

625 630 635 640

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu

645 650 655

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe

660 665 670

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

675 680 685

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

690 695 700

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

705 710 715 720

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
725 730 735
Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

740 745 750
Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
755 760 765
Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
770 775 780
Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
785 790 795 800
Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

805 810 815
Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
820 825 830
Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
835 840 845
Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
850 855 860
Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu

865 870 875 880
Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
885 890 895
Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr
900 905 910
Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
915 920 925
Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu

930 935 940
Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
945 950 955 960
Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln

965 970 975

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys

980 985 990

Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp

995 1000 1005

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile

1010 1015 1020

Leu Lys Asp Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys

1025 1030 1035

Ile Lys Lys Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp

1040 1045 1050

Ser Asp Ile Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys

1055 1060 1065

Gly Glu Lys Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe

1070 1075 1080

Pro Ser Asp Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu

1085 1090 1095

Glu Arg Ile Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser

1100 1105 1110

Thr Ile Glu Asp Asp Ser Ser Lys Gln Ser Met Lys Arg Pro Ala

1115 1120 1125

Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Lys

1130 1135 1140

<210> 89

<211> 1112

<212> PRT

<213> Bacillus sp.

<400> 89

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> 90

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 90

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Val Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Glu Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 91

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 91

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> 92

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 92

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> 93

<211> 1359

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 93

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> 94
<211> 1367
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 94

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Ala

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Ala Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Ala Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Ala Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 95

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 95

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> 96

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 96

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> 97

<211> 178

<212> PRT

<213> Escherichia coli

<400> 97

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> 98

<211> 167

<212> PRT

<213> Escherichia coli

<400> 98

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> 99
<211> 8
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic peptide
<400> 99
Ser Gly Gly Ser Ser Gly Gly Ser
1 5
<210> 100
<211> 1609

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 100

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> 101

<211> 1807

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 101

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> 102
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 102
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> 103

<211> 1786

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 103

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> 104

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 104

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> 105
<211> 1786
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 105

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> 106

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 106

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> 107
<211> 1786
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 107
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>

108

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 108

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> 109
<211> 1786
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 109
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> 110

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 110

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> 111

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 111

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> 112
<211> 1588
<212> PRT

<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 112

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> 113
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 113
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> 114

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 114

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> 115

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 115

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> 116

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 116

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> 117

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 117

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> 118
<211> 1588
<212> PRT

<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 118
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> 119
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 119

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> 120

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 120

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> 121

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 121

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> 122
<211> 1588
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 122
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> 123

<211> 1588

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 123

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> 124
<211> 4
<212> PRT

<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic peptide
<400> 124
Ser Gly Gly Ser
1

<210> 125
<211> 4
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic peptide
<400> 125
Gly Gly Gly Ser
1

<210> 126
<211> 5
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 126

Gly Gly Gly Gly Ser

1 5

<210> 127

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 127

Glu Ala Ala Ala Lys

1 5

<210> 128

<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> 128

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> 129

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 129

Pro Ala Pro Ala Pro

1 5

<210> 130

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 130

Pro Ala Pro Ala Pro Ala

1 5

<210>

131

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 131

Pro Ala Pro Ala Pro Ala Pro

1 5

<210> 132

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 132

Pro Ala Pro Ala Pro Ala Pro Ala

1 5

<210> 133

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 133

Pro Ala Pro Ala Pro Ala Pro Ala Pro

1 5

<210> 134

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 134

Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

1 5 10 15

<210> 135

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 135

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> 136

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<220><221> modified_base

<222> (98)..(117)

<223> a, c, t, g, unknown or other

<400> 136

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatgtctc ttacgaggca ttagcacnnn nnnnnnnnnn nnnnnnn 117

<210> 137

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<220><221> modified_base

<222> (89)..(108)

<223> a, c, t, g, unknown or other

<400> 137

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcacnn nnnnnnnnnn nnnnnnnn 108

<210> 138

<211> 119

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<220><221> modified_base

<222> (100)..(119)

<223> a, c, t, g, unknown or other

<400> 138

gtctaaagga cagaattttt caacgggtgt gccaatggcc actttccagg tggcaaagcc 60

cgttgaactt ctcaaaaaga acgatctgag aagtggcacn nnnnnnnnnn nnnnnnnnn 119

<210> 139

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 139

Gly Gly Ser Gly Gly Ser
1 5

<210> 140

<211> 20

<212> PRT

<213
> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 140

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> 141

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 141

ggaggctctg gaggaagc 18

<210> 142

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 142

ggctcttctg gatctgaaac acctggcaca agcgagagcg ccacccctga gagctctggc 60

<210> 143

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 143

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala

<210> 144

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 144

atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc c 51

<210> 145

<211> 933

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 145

ccggaattgc cagctggggc gccctctggt aaggttggga agccctgcaa agtaaactgg 60

atggctttct tgccgccaag gatctgatgg cgcaggggat caagatctga tcaagagaca 120

ggatgaggat cctttcgcat gatcgaataa gatggattgc acgcaggttc tccggccgct 180

taggtggagc gcctattcgg ctatgactgg gcacaacaga caatcggctg ctctgatgcc 240

gccgtgttcc ggctgtcagc gcaggggcgc ccggttcttt ttgtcaagac cgacctgtcc 300

ggtgccctga atgaactgca ggacgaggca gcgcggctat cgtggctggc cacgacgggc 360

gttccttgcg cagctgtgct cgacgttgtc actgaagcgg gaagggactg gctgctattg 420

ggcgaagtgc cggggcagga tctcctgtca tctcaccttg ctcctgccga gaaagtatcc 480

atcatggctg atgcaatgcg gcggctgcat acgcttgatc cggctacctg cccattcgac 540

caccaagcga aacatcgcat cgagcgagca cgtactcgga tggaagccgg tcttgtcgat 600
caggatgatc tggacgaaga gcatcagggg ctcgcgccag ccgaactgtt cgccaggctc 660
aaggcgcgca tgcccgacgg cgaggatctc gtcgtgaccc atggcgatgc ctgcttgccg 720
aatatcatgg tggaaaatgg ccgcttttct ggattcatta actgtggccg gctgggtgtg 780
gcggaccgct atcaggacat agcgttggct acccgtgata ttgctgaaga gcttggcggc 840
gaatgggctg accgcttcct cgtgctttac ggtatcgccg ctcccgattc gcagcgcatc 900
gccttctatc gccttcttga cgagttcttc taa 933

<210> 146
<211> 1052
<212> PRT
<213> Staphylococcus aureus
<400> 146

Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val Gly
1 5 10 15
Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val
20 25 30
Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser
35 40 45
Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln
50 55 60
Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser
65 70 75 80
Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser
85 90 95
Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala
100 105 110
Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly
115 120 125
Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu
130 135 140

Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp
145 150 155 160
Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr Val
165 170 175
Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu

180 185 190
Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg
195 200 205
Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp
210 215 220
Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
225 230 235 240
Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn

245 250 255
Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
260 265 270
Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
275 280 285
Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
290 295 300
Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro

305 310 315 320
Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
325 330 335
Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys
340 345 350
Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
355 360 365
Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn

370 375 380
Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn

385 390 395 400

Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile

405 410 415

Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln

420 425 430

Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val

435 440 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile

450 455 460

Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu

465 470 475 480

Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg

485 490 495

Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly

500 505 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met

515 520 525

Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp

530 535 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg

545 550 555 560

Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln

565 570 575

Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser

580 585 590

Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu

595 600 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr

610 615 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe

625 630 635 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met
645 650 655
Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val
660 665 670
Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
675 680 685
Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala

690 695 700
Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705 710 715 720
Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
725 730 735
Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
740 745 750
Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr

755 760 765
Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg 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> 147

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 147

atcgacaaga aagggactga agctgctg 28

<210> 148

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 148

Ile Asp Lys Lys Gly Thr Glu Ala Ala
1 5