

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0127797

(43) 공개일자 2019년11월13일

(51) 국제특허분류(Int. Cl.)

C12N 9/78 (2006.01) C12N 15/10 (2017.01)

C12N 15/62 (2006.01) C12N 9/22 (2006.01)

C12N 9/24 (2006.01)

(52) CPC특허분류

C12N 9/78 (2013.01)

C12N 15/102 (2013.01)

(21) 출원번호 10-2019-7029551

(22) 출원일자(국제) 2018년03월09일

심사청구일자 없음

(85) 번역문제출일자 2019년10월08일

(86) 국제출원번호 PCT/US2018/021878

(87) 국제공개번호 WO 2018/165629

국제공개일자 2018년09월13일

(30) 우선권주장

62/470,175 2017년03월10일 미국(US)

(71) 출원인

프레지던트 앤드 펠로우즈 오브 하바드 칼리지

미합중국, 메사추세츠 02138, 캠프리지, 퀸시스트리트17

(72) 발명자

리우, 데이비드, 알.

미국 02420 메사추세츠주 렉싱턴 휘트먼 서클 3

코블란, 루크, 더블유.

미국 02246 메사추세츠주 브루클라인 먼마우스 스트리트 71

(74) 대리인

양영준, 김영

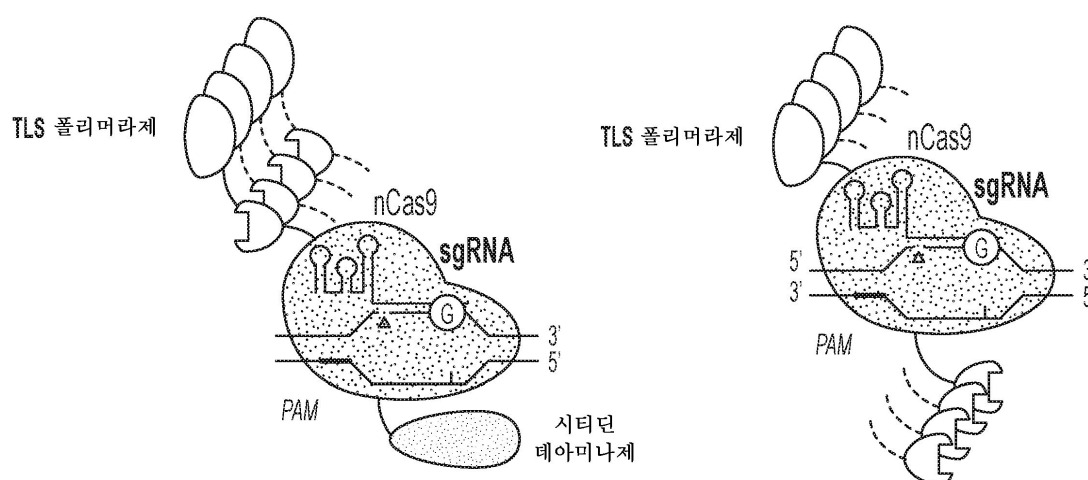
전체 청구항 수 : 총 181 항

(54) 발명의 명칭 시토신에서 구아닌으로의 염기 편집제

(57) 요약

본 개시내용의 일부 측면은 세포 또는 대상체의 게놈 내, 예를 들어 인간 게놈 내의 단일 부위의 편집을 포함한, 핵산의 표적화된 편집에 유용한 조성물, 전략, 시스템, 시약, 방법 및 키트를 제공한다. 일부 실시양태에서, 핵산(예를 들어, 게놈 DNA) 내 시토신(C)에서 구아닌(G)으로의 변화를 유도할 수 있는 융합 단백질이 제공된다. 일부 실시양태에서, 핵산 프로그램가능 DNA 결합 단백질(예를 들어, Cas9) 및 핵산 편집 단백질 또는 단백질 도메인, 예를 들어 데아미나제 도메인, 폴리머라제 도메인 및/또는 염기 절제 효소의 융합 단백질이 제공된다. 일부 실시양태에서, 표적화된 핵산 편집 방법이 제공된다. 일부 실시양태에서, 표적화된 핵산 편집 단백질, 예를 들어 핵산 프로그램가능 DNA 결합 단백질(예를 들어, Cas9) 및 핵산 편집 단백질 또는 도메인의 융합 단백질의 생성을 위한 시약 및 키트가 제공된다.

대표도



(52) CPC특허분류

C12N 15/62 (2013.01)

C12N 9/22 (2013.01)

C12N 9/2497 (2013.01)

C12Y 302/02027 (2013.01)

C12Y 305/04005 (2013.01)

C07K 2319/00 (2013.01)

C07K 2319/095 (2013.01)

명세서

청구범위

청구항 1

(i) 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp) 및 (ii) 시티딘 데아미나제 도메인 및 (iii) 우라실 결합 단백질 (UBP)을 포함하는 융합 단백질.

청구항 2

제1항에 있어서, 우라실 결합 단백질이 우라실 변형 효소인 융합 단백질.

청구항 3

제1항 또는 제2항에 있어서, 우라실 결합 단백질이 우라실 염기 절제 효소인 융합 단백질.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 우라실 결합 단백질이 우라실 DNA 글리코실라제 (UDG) 효소인 융합 단백질.

청구항 5

제4항에 있어서, 우라실 DNA 글리코실라제 (UDG) 효소가 인간, 마우스, 래트, 개, 원숭이, 소, 레서스 마카크, 침팬지, 고릴라, 바다 칠성장어로부터의 것 또는 우라실 DNA 글리코실라제 (UDG) 활성을 갖는 그의 임의의 변이체인 융합 단백질.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 우라실 결합 단백질이 서열식별번호: 49 (UdgX)의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함하는 UdgX인 융합 단백질.

청구항 7

제6항에 있어서, 우라실 결합 단백질이 서열식별번호: 49 (UdgX)의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 8

제1항 내지 제5항 중 어느 한 항에 있어서, 우라실 결합 단백질이 서열식별번호: 48 (UDG)의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함하는 UDG인 융합 단백질.

청구항 9

제8항에 있어서, 우라실 결합 단백질이 서열식별번호: 48 (UDG)의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 10

제1항 내지 제5항 중 어느 한 항에 있어서, 우라실 결합 단백질이 서열식별번호: 50 (UdgX*)의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함하는 UdgX*인 융합 단백질.

청구항 11

제10항에 있어서, 우라실 결합 단백질이 서열식별번호: 50 (UdgX*)의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 12

제1항 내지 제5항 중 어느 한 항에 있어서, 우라실 결합 단백질이 서열식별번호: 51 (UdgX_{On})의 아미노산 서열

에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함하는 UdgX_On인 융합 단백질.

청구항 13

제12항에 있어서, 우라실 결합 단백질이 서열식별번호: 51 (UdgX_On)의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 14

제1항 내지 제5항 중 어느 한 항에 있어서, 우라실 결합 단백질이 서열식별번호: 53 (SMUG1)의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함하는 SMUG1인 융합 단백질.

청구항 15

제14항에 있어서, 우라실 결합 단백질이 서열식별번호: 53 (SMUG1)의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서,

[시티딘 데아미나제]-[napDNAbp]-[UBP] 구조를 포함하며, 여기서 각 경우의 "]"-"는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서, 시티딘 데아미나제와 napDNAbp가 링커를 통해 융합된 것인 융합 단백질.

청구항 18

제17항에 있어서, 시티딘 데아미나제와 napDNAbp가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 19

제1항 내지 제18항 중 어느 한 항에 있어서, napDNAbp와 UBp가 링커를 통해 융합된 것인 융합 단백질.

청구항 20

제19항에 있어서, napDNAbp와 UBp가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 21

제1항 내지 제20항 중 어느 한 항에 있어서, (iv) 핵산 폴리머라제 도메인 (NAP)을 추가로 포함하는 융합 단백질.

청구항 22

제21항에 있어서, 핵산 폴리머라제 도메인이 진핵 핵산 폴리머라제 도메인인 융합 단백질.

청구항 23

제21항 또는 제22항에 있어서, 핵산 폴리머라제 도메인이 DNA 폴리머라제 도메인인 융합 단백질.

청구항 24

제21항 내지 제23항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 트랜스레전(translesion) 폴리머라제 활성을 갖는 것인 융합 단백질.

청구항 25

제21항 내지 제24항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 트랜스레전 DNA 폴리머라제인 융합 단백질.

청구항 26

제21항 내지 제25항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파 및 폴리머라제 에타로부터의 것인 융합 단백질.

청구항 27

제21항 내지 제25항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, 뮤 및 누로 이루어진 진핵 폴리머라제의 군으로부터 선택된 것인 융합 단백질.

청구항 28

제21항 내지 제27항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 29

제21항 내지 제28항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 30

제21항 내지 제29항 중 어느 한 항에 있어서,

[시티딘 데아미나제]-[napDNAbp]-[UBP]-[NAP];

[시티딘 데아미나제]-[napDNAbp]-[NAP]-[UBP];

[시티딘 데아미나제]-[NAP]-[napDNAbp]-[UBP]; 또는

[NAP]-[시티딘 데아미나제]-[napDNAbp]-[UBP]

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

청구항 31

제21항 내지 제30항 중 어느 한 항에 있어서, NAP와 UBPa 링커를 통해 융합된 것인 융합 단백질.

청구항 32

제31항에 있어서, NAP와 UBPa 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 33

제21항 내지 제32항 중 어느 한 항에 있어서, NAP와 napDNAbp가 링커를 통해 융합된 것인 융합 단백질.

청구항 34

제33항에 있어서, NAP와 napDNAbp가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 35

제21항 내지 제34항 중 어느 한 항에 있어서, NAP와 시티딘 데아미나제가 링커를 통해 융합된 것인 융합 단백질.

청구항 36

제35항에 있어서, NAP와 시티딘 데아미나제가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 37

제1항 내지 제36항 중 어느 한 항에 있어서, napDNAbp가 Cas9 도메인을 포함하는 것인 융합 단백질.

청구항 38

제37항에 있어서, Cas9 도메인이 서열식별번호: 4-26 중 어느 하나에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 39

제37항 또는 제38항에 있어서, Cas9 도메인이 서열식별번호: 4-26 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 40

제37항에 있어서, Cas9 도메인이 Cas9 니카제 (nCas9)인 융합 단백질.

청구항 41

제40항에 있어서, Cas9 니카제 (nCas9)가 서열식별번호: 10, 13, 16 또는 21 중 어느 하나에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 42

제40항 또는 제41항에 있어서, nCas9가 서열식별번호: 10, 13, 16 또는 21 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 43

제37항에 있어서, Cas9 도메인이 뉴클레아제 불활성 Cas9 (dCas9)인 융합 단백질.

청구항 44

제43항에 있어서, 뉴클레아제 불활성 Cas9 (dCas9)가 서열식별번호: 7, 8, 9 또는 22 중 어느 하나에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 45

제43항 또는 제44항에 있어서, dCas9가 서열식별번호: 7, 8, 9 또는 22 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 46

(i) 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp) 및 (ii) 시티딘 데아미나제 도메인 및 (iii) 핵산 폴리머라제 (NAP) 도메인을 포함하는 융합 단백질.

청구항 47

제46항에 있어서, 핵산 폴리머라제 도메인이 진핵 핵산 폴리머라제 도메인인 융합 단백질.

청구항 48

제46항 또는 제47항에 있어서, 핵산 폴리머라제 도메인이 DNA 폴리머라제 도메인인 융합 단백질.

청구항 49

제46항 내지 제48항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 트랜스레전 폴리머라제 활성을 갖는 것인 융합 단백질.

청구항 50

제46항 내지 제49항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 트랜스레전 DNA 폴리머라제인 융합 단백질.

청구항 51

제46항 내지 제50항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파 및 폴리머라제 에타로부터의 것인 융합 단백질.

청구항 52

제46항 내지 제51항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, 뮤 및 누로 이루어진 진핵 폴리머라제의 군으로부터 선택된 것인 융합 단백질.

청구항 53

제46항 내지 제52항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 54

제46항 내지 제53항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 55

제1항 내지 제54항 중 어느 한 항에 있어서, 시티딘 테아미나제 도메인이 아포지단백질 B mRNA-편집 복합체(APOBEC) 패밀리 테아미나제로부터의 테아미나제인 융합 단백질.

청구항 56

제55항에 있어서, APOBEC 패밀리 테아미나제가 APOBEC1 테아미나제, APOBEC2 테아미나제, APOBEC3A 테아미나제, APOBEC3B 테아미나제, APOBEC3C 테아미나제, APOBEC3D 테아미나제, APOBEC3F 테아미나제, APOBEC3G 테아미나제 및 APOBEC3H 테아미나제로 이루어진 군으로부터 선택된 것인 융합 단백질.

청구항 57

제1항 내지 제56항 중 어느 한 항에 있어서, 시티딘 테아미나제 도메인이 서열식별번호: 67-101 중 어느 하나의 아미노산 서열에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 58

제1항 내지 제57항 중 어느 한 항에 있어서, 시티딘 테아미나제 도메인이 서열식별번호: 67-101 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 59

제1항 내지 제57항 중 어느 한 항에 있어서, 시티딘 테아미나제 도메인이 서열식별번호: 93의 W90Y, R126E 및 R132E로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 테아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 래트 APOBEC1 (rAPOBEC1) 테아미나제인 융합 단백질.

청구항 60

제1항 내지 제57항 중 어느 한 항에 있어서, 시티딘 테아미나제 도메인이 서열식별번호: 91의 W90Y, Q126E 및

R132E로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 테아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 인간 APOBEC1 (hAPOBEC1) 테아미나제인 융합 단백질.

청구항 61

제1항 내지 제57항 중 어느 한 항에 있어서, 시티딘 테아미나제 도메인이 서열식별번호: 77의 W285Y, R320E 및 R326E로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 테아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 인간 APOBEC3G (hAPOBEC3G) 테아미나제인 융합 단백질.

청구항 62

제1항 내지 제54항 중 어느 한 항에 있어서, 시티딘 테아미나제 도메인이 활성화-유도된 테아미나제 (AID)인 융합 단백질.

청구항 63

제1항 내지 제54항 중 어느 한 항에 있어서, 시티딘 테아미나제 도메인이 페트로미존 마리누스(*Petromyzon marinus*)로부터의 시티딘 테아미나제 1 (pmCDA1)인 융합 단백질.

청구항 64

제46항 내지 제63항 중 어느 한 항에 있어서, napDNAbp가 Cas9 도메인을 포함하는 것인 융합 단백질.

청구항 65

제64항에 있어서, Cas9 도메인이 서열식별번호: 4-26 중 어느 하나에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 66

제64항 또는 제65항에 있어서, Cas9 도메인이 서열식별번호: 4-26 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 67

제64항에 있어서, Cas9 도메인이 Cas9 니카제 (nCas9)인 융합 단백질.

청구항 68

제67항에 있어서, Cas9 니카제 (nCas9)가 서열식별번호: 10, 13, 16 또는 21에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 69

제67항 또는 제68항에 있어서, nCas9가 서열식별번호: 10, 13, 16 또는 21의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 70

제64항에 있어서, Cas9 도메인이 뉴클레아제 불활성 Cas9 (dCas9)인 융합 단백질.

청구항 71

제70항에 있어서, 뉴클레아제 불활성 Cas9 (dCas9)가 서열식별번호: 7, 8, 9 또는 22에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 72

제70항 또는 제71항에 있어서, 뉴클레아제 불활성 Cas9 (dCas9)가 서열식별번호: 7, 8, 9 또는 22의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 73

제46항 내지 제72항 중 어느 한 항에 있어서,

[시티딘 테아미나제]-[napDNAbp]-[NAP] 구조를 포함하며, 여기서 각 경우의 "]"-["는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 74

제46항 내지 제73항 중 어느 한 항에 있어서, 시티딘 테아미나제와 napDNAbp가 링커를 통해 융합된 것인 융합 단백질.

청구항 75

제74항에 있어서, 시티딘 테아미나제와 napDNAbp가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 76

제46항 내지 제75항 중 어느 한 항에 있어서, napDNAbp와 NAP가 링커를 통해 융합된 것인 융합 단백질.

청구항 77

제76항에 있어서, napDNAbp와 NAP가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 78

(i) 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp) 및 (ii) 염기 절제 효소 (BEE)를 포함하는 융합 단백질.

청구항 79

제78항에 있어서, 염기 절제 효소가 시토신 (C) 또는 티민 (T) 염기 절제 효소인 융합 단백질.

청구항 80

제78항 또는 제79항에 있어서, 염기 절제 효소가 시토신 (C) 염기 절제 효소인 융합 단백질.

청구항 81

제79항 또는 제80항에 있어서, 염기 절제 효소가 서열식별번호: 48에 대해 적어도 85% 동일한 아미노산 서열을 포함하고, 서열식별번호: 48의 아미노산 잔기 156에서 A를 포함하는 것인 융합 단백질.

청구항 82

제80항 또는 제81항에 있어서, 염기 절제 효소가 서열식별번호: 65의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 83

제78항 또는 제79항에 있어서, 염기 절제 효소가 티민 (T) 염기 절제 효소인 융합 단백질.

청구항 84

제83항에 있어서, 염기 절제 효소가 서열식별번호: 48에 대해 적어도 85% 동일한 아미노산 서열을 포함하고, 서열식별번호: 48의 아미노산 잔기 213에서 D를 포함하는 것인 융합 단백질.

청구항 85

제84항 또는 제85항에 있어서, 염기 절제 효소가 서열식별번호: 66의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 86

제78항 내지 제85항 중 어느 한 항에 있어서,

[napDNAbp]-[BEE] 또는 [BEE]-[napDNAbp] 구조를 포함하며, 여기서 각 경우의 "]"-"는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 87

제78항 내지 제86항 중 어느 한 항에 있어서, BEE와 napDNAbp가 링커를 통해 융합된 것인 융합 단백질.

청구항 88

제86항 또는 제87항에 있어서, BEE와 napDNAbp가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 89

제78항 내지 제88항 중 어느 한 항에 있어서, (iii) 핵산 폴리머라제 (NAP) 도메인을 추가로 포함하는 융합 단백질.

청구항 90

제89항에 있어서, 핵산 폴리머라제 도메인이 진핵 핵산 폴리머라제 도메인인 융합 단백질.

청구항 91

제89항 또는 제90항에 있어서, 핵산 폴리머라제 도메인이 DNA 폴리머라제 도메인인 융합 단백질.

청구항 92

제89항 내지 제91항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 트랜스레전 폴리머라제 활성을 갖는 것인 융합 단백질.

청구항 93

제89항 내지 제92항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 트랜스레전 DNA 폴리머라제인 융합 단백질.

청구항 94

제89항 내지 제93항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파 및 폴리머라제 에타로부터의 것인 융합 단백질.

청구항 95

제89항 내지 제94항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, 뮤 및 누로 이루어진 진핵 폴리머라제의 군으로부터 선택된 것인 융합 단백질.

청구항 96

제89항 내지 제95항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 80%, 85%, 90%, 95%, 98% 또는 99% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 97

제89항 내지 제96항 중 어느 한 항에 있어서, 핵산 폴리머라제 도메인이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 98

제89항 내지 제97항 중 어느 한 항에 있어서,

[napDNAbp]-[BEE]-[NAP];

[napDNAbp]-[NAP]-[BEE];

[NAP]-[napDNAbp]-[BEE];

[NAP]-[BEE]-[napDNAbp];

[BEE]-[napDNAbp]-[NAP]; 또는

[BEE]-[NAP]-[napDNAbp]

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

청구항 99

제89항 내지 제98항 중 어느 한 항에 있어서, NAP와 BEE가 링커를 통해 융합된 것인 융합 단백질.

청구항 100

제99항에 있어서, NAP와 BEE가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 101

제89항 내지 제100항 중 어느 한 항에 있어서, BEE와 napDNAbp가 링커를 통해 융합된 것인 융합 단백질.

청구항 102

제101항에 있어서, BEE와 napDNAbp가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 103

제89항 내지 제102항 중 어느 한 항에 있어서, BEE와 napDNAbp가 링커를 통해 융합된 것인 융합 단백질.

청구항 104

제103항에 있어서, BEE와 napDNAbp가 서열식별번호: 102-109 중 어느 하나의 아미노산 서열을 포함하는 링커를 통해 융합된 것인 융합 단백질.

청구항 105

제78항 내지 제104항 중 어느 한 항에 있어서, napDNAbp가 Cas9 도메인을 포함하는 것인 융합 단백질.

청구항 106

제105항에 있어서, Cas9 도메인이 서열식별번호: 4-26 중 어느 하나에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 107

제105항 또는 제106항에 있어서, Cas9 도메인이 서열식별번호: 4-26 중 어느 하나의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 108

제105항에 있어서, Cas9 도메인이 Cas9 니카제 (nCas9)인 융합 단백질.

청구항 109

제108항에 있어서, Cas9 니카제 (nCas9)가 서열식별번호: 10, 13, 16 또는 21에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 110

제108항 또는 제109항에 있어서, nCas9가 서열식별번호: 10, 13, 16 또는 21의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 111

제105항에 있어서, Cas9 도메인이 뉴클레아제 불활성 Cas9 (dCas9)인 융합 단백질.

청구항 112

제111항에 있어서, 뉴클레아제 불활성 Cas9 (dCas9)가 서열식별번호: 7, 8, 9 또는 22에 대해 적어도 85% 동일한 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 113

제B34항 또는 제B35항에 있어서, dCas9가 서열식별번호: 7, 8, 9 또는 22의 아미노산 서열을 포함하는 것인 융합 단백질.

청구항 114

(i) 서열식별번호: 4-40 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제1 아미노산 서열;
(ii) 서열식별번호: 67-101 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제2 아미노산 서열; 및
(iii) 서열식별번호: 48-53 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제3 아미노산 서열을 포함하는 융합 단백질.

청구항 115

제114항에 있어서,

(iv) 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제4 아미노산 서열을 추가로 포함하는 융합 단백질.

청구항 116

제114항 또는 제115항에 있어서, 제1 아미노산 서열이 서열식별번호: 4-40 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 117

제114항 내지 제116항 중 어느 한 항에 있어서, 제2 아미노산 서열이 서열식별번호: 67-101 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 118

제114항 내지 제117항 중 어느 한 항에 있어서, 제3 아미노산 서열이 서열식별번호: 48-53 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 119

제115항 내지 제118항 중 어느 한 항에 있어서, 제4 아미노산 서열이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 120

제114항 내지 제119항 중 어느 한 항에 있어서,

NH_2 -[제2 아미노산 서열]-[제1 아미노산 서열]-[제3 아미노산 서열]-COOH;
 NH_2 -[제2 아미노산 서열]-[제3 아미노산 서열]-[제1 아미노산 서열]-COOH;
 NH_2 -[제1 아미노산 서열]-[제2 아미노산 서열]-[제3 아미노산 서열]-COOH;
 NH_2 -[제1 아미노산 서열]-[제3 아미노산 서열]-[제2 아미노산 서열]-COOH;
 NH_2 -[제3 아미노산 서열]-[제1 아미노산 서열]-[제2 아미노산 서열]-COOH;
 NH_2 -[제3 아미노산 서열]-[제2 아미노산 서열]-[제1 아미노산 서열]-COOH

구조를 포함하며, 여기서 각 경우의 "-"는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 121

제115항 내지 제119항 중 어느 한 항에 있어서,

NH_2 -[제4 아미노산 서열]-[제2 아미노산 서열]-[제1 아미노산 서열]-[제3 아미노산 서열]-COOH;
 NH_2 -[제2 아미노산 서열]-[제4 아미노산 서열]-[제1 아미노산 서열]-[제3 아미노산 서열]-COOH;
 NH_2 -[제2 아미노산 서열]-[제1 아미노산 서열]-[제4 아미노산 서열]-[제3 아미노산 서열]-COOH;
 NH_2 -[제2 아미노산 서열]-[제1 아미노산 서열]-[제3 아미노산 서열]-[제4 아미노산 서열]-COOH;
 NH_2 -[제4 아미노산 서열]-[제2 아미노산 서열]-[제3 아미노산 서열]-[제1 아미노산 서열]-COOH;
 NH_2 -[제2 아미노산 서열]-[제4 아미노산 서열]-[제3 아미노산 서열]-[제1 아미노산 서열]-COOH;
 NH_2 -[제2 아미노산 서열]-[제3 아미노산 서열]-[제4 아미노산 서열]-[제1 아미노산 서열]-COOH;
 NH_2 -[제2 아미노산 서열]-[제3 아미노산 서열]-[제1 아미노산 서열]-[제4 아미노산 서열]-COOH;
 NH_2 -[제4 아미노산 서열]-[제1 아미노산 서열]-[제2 아미노산 서열]-[제3 아미노산 서열]-COOH;
 NH_2 -[제1 아미노산 서열]-[제4 아미노산 서열]-[제2 아미노산 서열]-[제3 아미노산 서열]-COOH;
 NH_2 -[제1 아미노산 서열]-[제2 아미노산 서열]-[제4 아미노산 서열]-[제3 아미노산 서열]-COOH;
 NH_2 -[제1 아미노산 서열]-[제3 아미노산 서열]-[제4 아미노산 서열]-[제2 아미노산 서열]-COOH;
 NH_2 -[제4 아미노산 서열]-[제1 아미노산 서열]-[제3 아미노산 서열]-[제2 아미노산 서열]-COOH;
 NH_2 -[제1 아미노산 서열]-[제3 아미노산 서열]-[제4 아미노산 서열]-[제2 아미노산 서열]-COOH;
 NH_2 -[제1 아미노산 서열]-[제3 아미노산 서열]-[제2 아미노산 서열]-[제4 아미노산 서열]-COOH;
 NH_2 -[제4 아미노산 서열]-[제3 아미노산 서열]-[제1 아미노산 서열]-[제2 아미노산 서열]-COOH;
 NH_2 -[제3 아미노산 서열]-[제4 아미노산 서열]-[제1 아미노산 서열]-[제2 아미노산 서열]-COOH;
 NH_2 -[제3 아미노산 서열]-[제1 아미노산 서열]-[제4 아미노산 서열]-[제2 아미노산 서열]-COOH;
 NH_2 -[제3 아미노산 서열]-[제1 아미노산 서열]-[제2 아미노산 서열]-[제4 아미노산 서열]-COOH;
 NH_2 -[제4 아미노산 서열]-[제3 아미노산 서열]-[제2 아미노산 서열]-[제1 아미노산 서열]-COOH;

NH_2 -[제3 아미노산 서열]-[제4 아미노산 서열]-[제2 아미노산 서열]-[제1 아미노산 서열]-COOH;

NH_2 -[제3 아미노산 서열]-[제2 아미노산 서열]-[제4 아미노산 서열]-[제1 아미노산 서열]-COOH;

NH_2 -[제3 아미노산 서열]-[제2 아미노산 서열]-[제1 아미노산 서열]-[제4 아미노산 서열]-COOH

구조를 포함하며, 여기서 각 경우의 "-"는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 122

(i) 서열식별번호: 4-40 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제1 아미노산 서열;

(ii) 서열식별번호: 67-101 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제2 아미노산 서열; 및

(iii) 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제3 아미노산 서열을 포함하는 융합 단백질.

청구항 123

제122항에 있어서, 제1 아미노산 서열이 서열식별번호: 4-40 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 124

제122항 또는 제123항에 있어서, 제2 아미노산 서열이 서열식별번호: 67-101 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 125

제122항 내지 제124항 중 어느 한 항에 있어서, 제3 아미노산 서열이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 126

제122항 내지 제125항 중 어느 한 항에 있어서,

NH_2 -[제2 아미노산 서열]-[제1 아미노산 서열]-[제3 아미노산 서열]-COOH;

NH_2 -[제2 아미노산 서열]-[제3 아미노산 서열]-[제1 아미노산 서열]-COOH;

NH_2 -[제1 아미노산 서열]-[제2 아미노산 서열]-[제3 아미노산 서열]-COOH;

NH_2 -[제1 아미노산 서열]-[제3 아미노산 서열]-[제2 아미노산 서열]-COOH;

NH_2 -[제3 아미노산 서열]-[제1 아미노산 서열]-[제2 아미노산 서열]-COOH;

NH_2 -[제3 아미노산 서열]-[제2 아미노산 서열]-[제1 아미노산 서열]-COOH

구조를 포함하며, 여기서 각 경우의 "-"는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 127

(i) 서열식별번호: 4-40 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제1 아미노산 서열;

(ii) 서열식별번호: 65-66 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제2 아미노산 서열을 포함하는 융합 단백질.

청구항 128

제127항에 있어서,

(iii) 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 80% 동일한 제3 아미노산 서열을 추가로 포함하는 융합 단백질.

청구항 129

제127항 또는 제128항에 있어서, 제1 아미노산 서열이 서열식별번호: 4-40 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 130

제127항 내지 제129항 중 어느 한 항에 있어서, 제2 아미노산 서열이 서열식별번호: 65-66 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 131

제128항 내지 제130항 중 어느 한 항에 있어서, 제3 아미노산 서열이 서열식별번호: 54-64 중 어느 하나의 아미노산 서열에 대해 적어도 85%, 90%, 95%, 98% 또는 99% 동일한 것인 융합 단백질.

청구항 132

제127항 내지 제131항 중 어느 한 항에 있어서,

NH_2 -[제2 아미노산 서열]-[제1 아미노산 서열]-COOH; 또는

NH_2 -[제1 아미노산 서열]-[제2 아미노산 서열]-COOH

구조를 포함하며, 여기서 각 경우의 "-"는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 133

제128항 내지 제131항 중 어느 한 항에 있어서,

NH_2 -[제2 아미노산 서열]-[제1 아미노산 서열]-[제3 아미노산 서열]-COOH;

NH_2 -[제2 아미노산 서열]-[제3 아미노산 서열]-[제1 아미노산 서열]-COOH;

NH_2 -[제1 아미노산 서열]-[제2 아미노산 서열]-[제3 아미노산 서열]-COOH;

NH_2 -[제1 아미노산 서열]-[제3 아미노산 서열]-[제2 아미노산 서열]-COOH;

NH_2 -[제3 아미노산 서열]-[제1 아미노산 서열]-[제2 아미노산 서열]-COOH;

NH_2 -[제3 아미노산 서열]-[제2 아미노산 서열]-[제1 아미노산 서열]-COOH

구조를 포함하며, 여기서 각 경우의 "-"는 임의적인 링커를 포함하는 것인 융합 단백질.

청구항 134

핵산 분자 및 제1항 내지 제133항 중 어느 한 항의 융합 단백질을 포함하는 복합체.

청구항 135

제134항에 있어서, 핵산 분자가 RNA인 복합체.

청구항 136

제134항 또는 제135항에 있어서, 핵산 분자가 gRNA인 복합체.

청구항 137

제134항 내지 제136항 중 어느 한 항에 있어서, 핵산 분자가 sgRNA인 복합체.

청구항 138

제1항 내지 제133항 중 어느 한 항의 융합 단백질을 포함하는 제약 조성물.

청구항 139

제134항 내지 제137항 중 어느 한 항의 복합체를 포함하는 제약 조성물.

청구항 140

제138항 또는 139항에 있어서, 제약상 허용되는 부형제를 추가로 포함하는 제약 조성물.

청구항 141

핵산 분자를 제1항 내지 제133항 중 어느 한 항의 융합 단백질 또는 제134항 내지 제137항 중 어느 한 항의 복합체와 접촉시키는 것을 포함하는 방법.

청구항 142

제141항에 있어서, 핵산 분자가 DNA인 방법.

청구항 143

제141항 또는 제142항에 있어서, 핵산 분자가 게놈의 DNA인 방법.

청구항 144

a) 이중-가닥 DNA 서열의 표적 영역을, 핵염기 편집제 및 가이드 핵산을 포함하는 복합체와 접촉시키는 단계이며, 여기서 표적 영역은 표적 핵염기 쌍을 포함하는 것인 단계;
b) 상기 표적 영역의 가닥 분리를 유도하는 단계; 및
c) 표적 영역의 단일 가닥에서 시토신 또는 티민을 절제하는 단계를 포함하는, 이중-가닥 DNA 서열의 핵염기 쌍을 편집하는 방법.

청구항 145

제144항에 있어서, 단계 c가 표적 영역의 단일 가닥에서 시토신을 절제하는 것을 포함하는 것인 방법.

청구항 146

a) 이중-가닥 DNA 서열의 표적 영역을 핵산 프로그램가능 DNA 결합 단백질과 접촉시키는 단계이며, 여기서 표적 영역은 표적 핵염기 쌍을 포함하는 것인 단계;
b) 표적 영역의 상기 표적 핵염기 쌍의 제1 핵염기를 절제하는 단계; 및
c) 제1 핵염기를 제1 핵염기와 상이한 제2 핵염기로 대체하는 단계를 포함하는, 이중-가닥 DNA 서열의 핵염기 쌍을 편집하는 방법.

청구항 147

(a) 제1항 내지 제133항 중 어느 한 항의 융합 단백질을 코딩하는 핵산 서열; 및
(b) (a)의 발현을 구동하는 이중 프로모터를 포함하는 핵산 구축물을 포함하는 키트.

청구항 148

제1항 내지 제133항 중 어느 한 항의 융합 단백질을 코딩하는 폴리뉴클레오타이드.

청구항 149

제148항의 폴리뉴클레오티드를 포함하는 벡터.

청구항 150

제1항 내지 제133항 중 어느 한 항의 융합 단백질을 포함하는 세포.

청구항 151

제134항 내지 제137항 중 어느 한 항의 복합체를 포함하는 세포.

청구항 152

제1항 내지 제133항 중 어느 한 항의 융합 단백질을 코딩하는 핵산 분자를 포함하는 세포.

청구항 153

제150항 내지 제152항 중 어느 한 항에 있어서, 시험관내 세포인 세포.

청구항 154

제150항 내지 제152항 중 어느 한 항에 있어서, 생체내 세포인 세포.

청구항 155

질환 또는 장애를 갖거나 갖는 것으로 의심되는 대상체에게 제1항 내지 제133항 중 어느 한 항의 융합 단백질, 제134항 내지 제137항 중 어느 한 항의 복합체, 제148항의 폴리뉴클레오티드 또는 제149항의 벡터를 투여하는 것을 포함하는, 질환 또는 장애를 갖거나 갖는 것으로 의심되는 대상체를 치료하는 방법.

청구항 156

a. 이중-가닥 DNA 서열의 표적 영역을, 핵염기 편집제 및 가이드 핵산을 포함하는 복합체와 접촉시키는 단계이며, 여기서 표적 영역은 표적 핵염기 쌍을 포함하는 것인 단계;

b. 상기 표적 영역의 가닥 분리를 유도하는 단계;

c. 표적 영역의 단일 가닥에서 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계;

d. 상기 제2 핵염기를 이중-가닥 DNA 서열로부터 절제하여 무염기성 부위를 생성하는 단계; 및

e. 상기 표적 영역의 1개 이하의 가닥을 절단하는 단계

를 포함하며, 여기서 무염기성 부위 반대편 제3 핵염기는 제4 핵염기에 의해 대체되는 것인, 이중-가닥 DNA 서열의 핵염기 쌍을 편집하는 방법.

청구항 157

제156항에 있어서, 20% 미만 내지 1% 미만의 삽입결실 형성을 유발하는 방법.

청구항 158

제156항에 있어서, 제4 핵염기에 상보적인 제5 핵염기를 무염기성 부위 내로 삽입함으로써 의도된 편집된 염기 쌍을 생성하는 단계를 추가로 포함하는 방법.

청구항 159

제158항에 있어서, 의도된 편집된 염기 쌍을 생성하는 효율이 적어도 5%인 방법.

청구항 160

제159항에 있어서, 효율이 적어도 10% 내지 적어도 50%인 방법.

청구항 161

제158항에 있어서, 의도된 편집된 염기쌍 대 의도되지 않은 편집된 염기쌍의 비가 2:1 내지 10:1인 방법.

청구항 162

제158항에 있어서, 의도된 편집된 염기쌍 대 삼입결실 형성의 비가 2:1 내지 200:1인 방법.

청구항 163

제156항에 있어서, 절단된 가닥이 가이드 핵산에 혼성화되는 것인 방법.

청구항 164

제156항에 있어서, 절단된 단일 가닥이 제1 핵염기를 포함하는 가닥의 반대편 가닥인 방법.

청구항 165

제156항에 있어서, 제1 핵염기가 시토신인 방법.

청구항 166

제156항에 있어서, 제2 핵염기가 우라실인 방법.

청구항 167

제156항에 있어서, 핵염기 편집제가 우라실 결합 단백질을 포함하는 것인 방법.

청구항 168

제156항에 있어서, 핵염기 편집제가 니카제 활성을 포함하는 것인 방법.

청구항 169

제159항에 있어서, 의도된 편집된 염기 쌍이 PAM 부위의 상류에 있는 것인 방법.

청구항 170

제169항에 있어서, 의도된 편집된 염기 쌍이 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오티드 상류에 있는 것인 방법.

청구항 171

제158항에 있어서, 의도된 편집된 염기 쌍이 PAM 부위의 하류에 있는 것인 방법.

청구항 172

제171항에 있어서, 의도된 편집된 염기 쌍이 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오티드 하류에 있는 것인 방법.

청구항 173

제156항에 있어서, 핵염기 편집제가 링커를 포함하는 것인 방법.

청구항 174

제173항에 있어서, 링커가 1-25개 아미노산 길이인 방법.

청구항 175

제173항에 있어서, 링커가 5-20개 아미노산 길이인 방법.

청구항 176

제173항에 있어서, 링커가 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 아미노산 길이인 방법.

청구항 177

제156항에 있어서, 표적 영역이 표적 윈도우를 포함하고, 여기서 표적 윈도우가 표적 핵염기 쌍을 포함하는 것인 방법.

청구항 178

제177항에 있어서, 표적 윈도우가 1-10개 뉴클레오티드를 포함하는 것인 방법.

청구항 179

제177항에 있어서, 표적 윈도우가 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2 또는 1개 뉴클레오티드 길이인 방법.

청구항 180

제177항에 있어서, 표적 윈도우가 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오티드 길이인 방법.

청구항 181

제177항 내지 제180항 중 어느 한 항에 있어서, 표적 윈도우가 의도된 편집된 염기 쌍을 포함하는 것인 방법.

발명의 설명

배경 기술

[0001] 핵산 서열의 표적화된 편집, 예를 들어 표적화된 절단 또는 게놈 DNA 내로의 특정 변형의 표적화된 도입은 유전자 기능의 연구를 위한 고도로 유망한 접근법이고, 또한 인간 유전 질환에 대한 새로운 요법을 제공하는 잠재력을 갖는다. 많은 유전 질환은 원칙적으로 게놈 내 특정 위치에서의 특정 뉴클레오티드 변화 (예를 들어, 질환과 연관된 유전자의 특정 코돈에서의 C에서 G로 또는 G에서 C로의 변화)에 영향을 미침으로써 치료될 수 있기 때문에, 이러한 정확한 유전자 편집을 달성하기 위한 프로그램가능한 방식의 개발은 강력한 신규 연구 도구, 뿐만 아니라 유전자 편집-기반 치료제에 대한 잠재적 신규 접근법 둘 다를 나타낸다.

발명의 내용

[0002] 폴리뉴클레오티드 (예를 들어, DNA)를 변형시키는, 예를 들어 폴리뉴클레오티드 내에 시토신에서 구아닌으로의 돌연변이를 생성하는 조성물, 키트 및 방법이 본원에 제공된다. 본원에 보다 상세하게 기재된 바와 같이, 염기 편집 (예를 들어, C에서 G로의 편집)은 핵염기 (예를 들어, 시토신 (C))를 제거하여 핵산 서열 내에 무염기성 부위를 생성함으로써 달성되었다. 이어서, 무염기성 부위 반대편 핵염기 (예를 들어, 구아닌)가, 예를 들어 내인성 트랜스레전(translesion) 폴리머라제에 의해 상이한 핵염기 (예를 들어, 시토신)로 대체된다. 본원에 기재된 염기 편집 융합 단백질은 핵산 (예를 들어, 게놈 DNA) 내에 특정 돌연변이 (예를 들어, C에서 G로의 돌연변이)를 생성할 수 있으며, 이는, 예를 들어 핵산 돌연변이, 예를 들어 C에서 G로 또는 G에서 C로의 돌연변이를 수반하는 질환을 치료하는데 사용될 수 있다.

[0003] C에서 G로의 염기 편집제의 하나의 예는 핵산 프로그램가능 DNA 결합 단백질 (예를 들어, Cas9 도메인), 우라실 DNA 글리코실라제 (UDG) 도메인 및 시티딘 데아미나제를 함유하는 융합 단백질을 포함한다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, 이러한 염기 편집 융합 단백질은 특정 핵산 서열에 결합할 수 있고 (예를 들어, Cas9 도메인을 통해), UDG에 의해 핵산 서열 내의 시토신을 우라실로 탈아미노화한 다음 이를 핵산 분자로부터 절제할 수 있다. 이어서, 무염기성 부위 반대편 핵염기는, 예를 들어 내인성 트랜스레전 폴리머라제에 의해 또 다른 염기 (예를 들어, 시토신)로 대체될 수 있다. 전형적으로, 염기 복구 기구 (예를 들어, 세포에서)는, 무염기성 부위 반대편 핵염기를 다른 염기 (예를 들어, 아데닌, 구아닌 또는 티민)가 대체할 수 있긴 하지만, 무염기성 부위 반대편 핵염기를 시토신으로 대체한다. 추가로, 염기 편집제 내로 트랜스레전 폴리머라제를 혼입시키는 것은 무염기성 부위 반대편 시토신 혼입을 증가시킬 수 있는 것으로 밝혀졌다. 따라서, 염기 편집제는 염기 편집 효율을 개선시키기 위해 다양한 트랜스레전 폴리머라제가 혼입되도록 조작되었다. 무염기성 부위 반대편 C 통합에 대한 선호를 증가시키는 트랜스레전 폴리머라제는 C에서 G로의 핵염기 편집을 개선시킬 수 있다. 비-C 핵염기 (예를 들어, 아데닌, 구아닌 및 티민)를 우선적으로 통합시키는 다른 트랜스레전 폴

리머라제는 대안적 돌연변이 (예를 들어, C에서 A로의 돌연변이)를 생성하는데 사용될 수 있다는 것이 인지되어야 한다.

[0004] 또 다른 예로서, 염기 편집 융합 단백질은 핵산 프로그램가능 DNA 결합 단백질 (예를 들어, Cas9 도메인), 및 핵염기 (특히 시토신)를 제거하는 염기 절제 효소를 포함할 수 있다. 상기 기재된 바와 같이 UDG를 사용하여 시토신을 우리딘으로 탈아미노화하고 우리딘을 절제하기보다는, 염기 편집제는 핵염기, 예컨대 시토신 또는 티민을 먼저 탈아미노화하지 않으면서 이를 인식하고 제거하는 염기 절제 효소를 포함할 수 있다. 따라서, 염기 편집제 (예를 들어, C에서 G로의 염기 편집제)는, 핵산 프로그램가능 DNA 결합 단백질 (예를 들어, Cas9 도메인)을, 핵산 분자로부터 시토신 또는 티민을 제거하는 염기 절제 효소에 융합시킴으로써 조작되었다. 추가로, 상기 기재된 염기 편집제와 같이, 트랜스레전 폴리머라제를 이러한 염기 편집제 내로 혼입시켜, 염기 편집제의 염기 절제 효소에 의해 생성된 무염기성 부위 반대편 시토신 혼입을 증가시켰다. 염기 편집 전략을 약속하는 예시적인 염기 편집 단백질 및 개략적 표현은, 예를 들어 도 1-6, 33-36, 40 및 52에서 볼 수 있다.

[0005] 일부 실시양태에서, 본 개시내용은 염기 편집이 가능한 융합 단백질을 제공한다. 예시적인 염기 편집 융합 단백질은 하기를 포함한다. 일부 실시양태에서, 융합 단백질은 (i) 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp), (ii) 시티딘 데아미나제 도메인 및 (iii) 우라실 결합 단백질 (UBP)을 포함한다. 일부 실시양태에서, 융합 단백질은 (iv) 핵산 폴리머라제 도메인 (NAP)을 추가로 포함한다. 또 다른 예로서, 융합 단백질은 (i) 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp), (ii) 시티딘 데아미나제 도메인 및 (iii) 핵산 폴리머라제 (NAP) 도메인을 포함할 수 있다. 또 다른 예로서, 융합 단백질은 (i) 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp) 및 (ii) 염기 절제 효소 (BEE)를 포함할 수 있다. 일부 실시양태에서, 융합 단백질은 (iii) 핵산 폴리머라제 (NAP) 도메인을 추가로 포함한다. 염기 편집제 및 염기 편집제의 사용 방법은 하기에 추가로 상세하게 기재된다.

도면의 간단한 설명

[0006] 도 1은 C에서 T로 및 C에서 G로의 염기 편집을 예시하는 일반적인 개략도를 보여준다. 특정 DNA 폴리머라제 (예를 들어, 트랜스레전 폴리머라제)는 무염기성 부위 반대편 염기를 G로 대체하는 것으로 공지되어 있다. C에서 G로의 염기 편집을 달성하기 위한 하나의 전략은 무염기성 부위의 생성을 유도하고, 이어서 이러한 폴리머라제를 동원하거나 테더링하여 무염기성 부위 반대편 G를 C로 대체하는 것이다.

도 2는 C에서 G로의 편집을 위한 무염기성 부위 생성 및 염기-특이적 복구를 통한 염기 편집을 예시하는 일반적인 개략도를 보여준다.

도 3은 C에서 G로의 염기 편집을 위해 무염기성 부위가 형성되는, 도 1로부터의 도식 1을 예시하는 개략도를 보여준다. 무염기가 효율적으로 생성되는 경우, 이는 C에서 G로의 편집 경로를 통한 총 유동을 증가시킬 수 있다.

도 4는 무염기성 부위 형성의 증가가 사용되는, C에서 G로의 염기 편집을 위한 접근법 1을 예시하는 개략도를 보여준다. 예를 들어 UDG 도메인 및 트랜스레전 폴리머라제를 사용하여 무염기가 효율적으로 생성되는 경우, 이는 C에서 G로의 편집 경로를 통한 총 유동을 증가시킬 수 있다.

도 5는 염기 편집에 대한 UdgX의 효과를 예시하는 개략도를 보여준다. 최소 우라실 절제 활성으로 우라실에 단단히 결합하는 것으로 확인된 UDG의 오르토로그인 UdgX는 C에서 G로의 편집 양을 증가시킨다. 1.)에서 UdgX*는 시험관내 검정을 통해 우라실 결합 활성이 결여된 것으로 결정된 UDG의 변이체이다. 2.)에서 UdgX_On은 시험관내 검정을 통해 우라실 절제를 증가시키는 것으로 나타난 변이체이다. 3.)에서 UDG 직접 융합체는 우라실을 절제한다.

도 6은 우라실 글리코실라제 억제제 (UGI), Cas9 도메인 (예를 들어, nCas9) 및 시티딘 데아미나제를 함유하는 예시적인 C에서 T로의 염기 편집제 (예를 들어, BE3)를 예시하는 개략도 (좌측)를 보여준다. 우측은 우라실 DNA 글리코실라제 (UDG) (또는 그의 변이체), Cas9 도메인 (예를 들어, nCas9) 및 시티딘 데아미나제를 함유하는 C에서 G로의 염기 편집제를 예시하는 개략도이다.

도 7은 7종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 WT Hap1 세포 내 HEK2 부위에서의 총 편집 백분율을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로

그래프 제시된다.

도 8은 WT Hap1 세포에서 추가의 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_REV7; 및 SMUG1, 여기서 BE3 및 BE3_UdgX는 도 4로부터 반복됨)를 사용한 HEK2 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 9는 WT Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE3_REV7; BE2_UDG; BE3_UDG BE2_UdgX_On; BE3_UdgX_On; 및 SMUG1)를 사용한 HEK2 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 G에서 A, C 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 10은 7종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 WT Hap1 세포 내 RNF2 부위에서의 총 편집 백분율을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 11은 WT Hap1 세포에서 추가의 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_REV7; 및 SMUG1, 여기서 BE3 및 BE3_UdgX는 도 7로부터 반복됨)를 사용한 RNF2 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 12는 WT Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE3_REV7; BE2_UDG; BE3_UDG BE2_UdgX_On; BE3_UdgX_On; 및 SMUG1)를 사용한 RNF2 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 G에서 A, C 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 13은 7종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 WT Hap1 세포 내 FANCF 부위에서의 총 편집 백분율을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G 염기로의 편집은 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다.

도 14는 WT Hap1 세포에서 추가의 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_REV7; 및 SMUG1, 여기서 BE3 및 BE3_UdgX는 도 10으로부터 반복됨)를 사용한 FANCF 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다.

도 15는 WT Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE3_REV7; BE2_UDG; BE3_UDG BE2_UdgX_On; BE3_UdgX_On; 및 SMUG1)를 사용한 FANCF 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 C에서 A, G 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 16은 7종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 UDG^{-/-} Hap1 세포 내 HEK2 부위에서의 총 편집 백분율을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 17은 UDG^{-/-} Hap1 세포에서 추가의 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_REV7; 및 SMUG1, 여기서 BE3 및 BE3_UdgX는 도 13으로부터 반복됨)를 사용한 HEK2 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 18은 $UDG^{-/-}$ Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE3_REV7; BE2_UDG; BE3_UDG BE2_UdgX_On; BE3_UdgX_On; 및 SMUG1)를 사용한 HEK2 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 G에서 A, C 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 19는 7종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 $UDG^{-/-}$ Hap1 세포 내 RNF2 부위에서의 총 편집 백분율을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 20은 $UDG^{-/-}$ Hap1 세포에서 추가의 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_REV7; 및 SMUG1, 여기서 BE3 및 BE3_UdgX는 도 16으로부터 반복됨)를 사용한 RNF2 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 21은 $UDG^{-/-}$ Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE3_REV7; BE2_UDG; BE3_UDG BE2_UdgX_On; BE3_UdgX_On; 및 SMUG1)를 사용한 RNF2 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 G에서 A, C 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 22는 7종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 $UDG^{-/-}$ Hap1 세포 내 FANCF 부위에서의 총 편집 백분율을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G 염기로의 편집은 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다.

도 23은 $UDG^{-/-}$ Hap1 세포에서 추가의 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_REV7; 및 SMUG1, 여기서 BE3 및 BE3_UdgX는 도 19로부터 반복됨)를 사용한 FANCF 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다.

도 24는 $UDG^{-/-}$ Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE3_REV7; BE2_UDG; BE3_UDG BE2_UdgX_On; BE3_UdgX_On; 및 SMUG1)를 사용한 FANCF 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 C에서 A, G 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 25는 $REV1^{-/-}$ Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 HEK2 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 26은 $REV1^{-/-}$ Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 HEK2 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 G에서 A, C 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 27은 $REV1^{-/-}$ Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 RNF2 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 편집된 C를 함유하는 가닥의 반대편 DNA 가닥에 대해 서열분석을 수행함에 따라 도트형 막대 (G)가 채워진

막대 (C)로 이동한 것으로 그래프 제시된다.

도 28은 REV1^{-/-} Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 RNF2 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 G에서 A, C 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 29는 REV1^{-/-} Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 FANCF 부위에서의 총 편집 백분율을 보여준다. 상단 패널은 미가공 편집 값을 보여준다. 하단 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G로의 염기 편집은 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다.

도 30은 REV1^{-/-} Hap1 세포에서 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 FANCF 부위에서의 편집 특이성 비를 보여준다. 상단 패널은 편집의 총 백분율 및 C에서 A, G 또는 T로 이루어진 편집의 비를 보여준다. 하단 패널은 특이성 비 값의 그래프 표현이다.

도 31은 나타낸 C에서 G로의 염기 편집제를 사용한 HEK2, RNF2 및 FANCF 부위에서의 총 편집의 퍼센트에 대한 미가공 편집 값의 그래프 표현을 보여준다.

도 32는 HEK2, RNF2 및 FANCF 부위에서의 총 편집의 퍼센트에 대한 특이성 비의 그래프 표현을 보여준다.

도 33은 C에서 G로의 염기 편집을 위해 무염기성 부위 반대편에 C의 혼입을 증가시키는 접근법을 예시하는 개략도를 보여준다. 무염기성 부위 반대편 C 통합에 대한 선호가, 예를 들어 폴리머라제 (예를 들어, 트랜스레전 폴리머라제)를 사용하여 증가되는 경우, 총 C에서 G로의 염기 편집이 또한 증가될 것이다.

도 34는 C에서 G로의 염기 편집을 위해 무염기성 부위 반대편에 C의 혼입을 증가시키는 접근법을 예시하는 개략도를 보여준다. 무염기성 부위 반대편 C 통합에 대한 선호가, 예를 들어 트랜스레전 폴리머라제를 염기 편집제 내로 혼입시킴으로써 증가되는 경우, 총 C에서 G로의 염기 편집이 또한 증가될 수 있다.

도 35는 도 33 및 34의 C에서 G로의 염기 편집 접근법에서 사용될 수 있는 상이한 폴리머라제를 예시하는 개략도를 보여준다.

도 36은 우라실 글리코실라제 억제제 (UGI), Cas9 도메인 (예를 들어, nCas9) 및 시티딘 데아미나제를 함유하는 예시적인 C에서 T로의 염기 편집제 (예를 들어, BE3)를 예시하는 개략도 (좌측)를 보여준다. 우측은 트랜스레전 폴리머라제, Cas9 도메인 (예를 들어, nCas9) 및 시티딘 데아미나제를 함유하는 C에서 G로의 염기 편집제를 예시하는 개략도이다.

도 37은 REV1, Pol 카파, Pol 에타 및 Pol 이오타에 테더링된 염기 편집제를 사용한 WT 세포 내 HEK2 부위에서의 염기 편집을 보여준다. C에서 G로의 편집은 우측 패널 상의 그래프 표현에서 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다. Pol 카파 테더링은 C에서 G로의 편집 효율을 극적으로 증가시킨다. 미가공 편집 값은 좌측 패널에 제시된다.

도 38은 REV1, Pol 카파, Pol 에타 및 Pol 이오타에 테더링된 염기 편집제를 사용한 WT 세포 내 RNF2 부위에서의 염기 편집을 보여준다. C에서 G로의 편집은 우측 패널 상의 그래프 표현에서 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다. Pol 카파 테더링은 C에서 G로의 편집 효율을 극적으로 증가시킨다. 미가공 편집 값은 좌측 패널에 제시된다.

도 39는 REV1, Pol 카파, Pol 에타 및 Pol 이오타에 테더링된 염기 편집제를 사용한 WT 세포 내 FANCF 부위에서의 염기 편집을 보여준다. C에서 G로의 편집은 우측 패널 상의 그래프 표현에서 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다. Pol 카파 테더링은 C에서 G로의 편집 효율을 극적으로 증가시킨다. 미가공 편집 값은 좌측 패널에 제시된다.

도 40은 우라실 DNA 글리코실라제 (UDG), 트랜스레전 폴리머라제, Cas9 도메인 (예를 들어, nCas9) 및 시티딘 데아미나제를 함유하는 예시적인 C에서 G로의 염기 편집제를 예시하는 개략도 (좌측)를 보여준다. 우측은 트랜스레전 폴리머라제, Cas9 도메인 (예를 들어, nCas9) 및 염기 절제 효소 (예를 들어, C 또는 T 잔기를 절제할 수 있는 UDG 변이체)를 함유하는 C에서 G로의 염기 편집제를 예시하는 개략도이다.

도 41은 Pol 카파 또는 Pol 이오타 테더링된 구축물을 사용하여 HEK2, RNF2 및 FANCF 부위에서 도 40의 좌측 패널에 예시된 염기 편집제 (우라실 DNA 글리코실라제 (UDG), 트랜스레전 폴리머라제, Cas9 도메인 및 시티딘 데아미나제를 함유하는 염기 편집제)를 사용한 C에서 G로의 염기 편집을 보여준다. C에서 G로의 편집은 HEK2 및 RNF2의 경우 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로, FANCF의 경우 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다.

도 42는 도 40의 우측 패널에 제시된 Pol 카파, Pol 에타, Pol 이오타 및 REV1에 테더링된 염기 편집제 (트랜스레전 폴리머라제, Cas9 도메인, 및 T를 절제하는 염기 절제 효소 (UDG 147)를 함유하는 염기 편집제)를 사용한 WT 세포 내 HEK2 부위에서의 염기 편집을 보여준다. C에서 G로의 양은 HEK2 부위 내 특정 잔기에서 그래프로 예시된다. UDG 147은 T를 직접 제거하는 UDG 변이체이다.

도 43은 도 40의 우측 패널에 제시된 Pol 카파, Pol 에타, Pol 이오타 및 REV1에 테더링된 염기 편집제 (트랜스레전 폴리머라제, Cas9 도메인, 및 T를 절제하는 염기 절제 효소 (UDG 147)를 함유하는 염기 편집제)를 사용한 WT 세포 내 RNF2 부위에서의 염기 편집을 보여준다. C에서 G로의 양은 HEK2 부위 내 특정 잔기에서 그래프로 예시된다. UDG 147은 T를 직접 제거하는 UDG 변이체이다.

도 44는 도 40의 우측 패널에 제시된 Pol 카파, Pol 에타, Pol 이오타 및 REV1에 테더링된 염기 편집제 (트랜스레전 폴리머라제, Cas9 도메인, 및 T를 절제하는 염기 절제 효소 (UDG 147)를 함유하는 염기 편집제)를 사용한 WT 세포 내 FANCF 부위에서의 염기 편집을 보여준다. C에서 G로의 양은 HEK2 부위 내 특정 잔기에서 그래프로 예시된다. UDG 147은 T를 직접 제거하는 UDG 변이체이다.

도 45는 도 40의 우측 패널에 제시된 Pol 카파, Pol 에타, Pol 이오타 및 REV1에 테더링된 염기 편집제 (트랜스레전 폴리머라제, Cas9 도메인, 및 C를 절제하는 염기 절제 효소 (UDG 204)를 함유하는 염기 편집제)를 사용한 WT 세포 내 HEK2 부위에서의 염기 편집을 보여준다. C에서 G로의 양은 HEK2 부위 내 특정 잔기에서 그래프로 예시된다. UDG 204는 C를 직접 제거하는 UDG 변이체이다.

도 46은 도 40의 우측 패널에 제시된 Pol 카파, Pol 에타, Pol 이오타 및 REV1에 테더링된 염기 편집제 (트랜스레전 폴리머라제, Cas9 도메인, 및 C를 절제하는 염기 절제 효소 (UDG 204)를 함유하는 염기 편집제)를 사용한 WT 세포 내 RNF2 부위에서의 염기 편집을 보여준다. C에서 G로의 양은 HEK2 부위 내 특정 잔기에서 그래프로 예시된다. UDG 204는 C를 직접 제거하는 UDG 변이체이다.

도 47은 도 40의 우측 패널에 제시된 Pol 카파, Pol 에타, Pol 이오타 및 REV1에 테더링된 염기 편집제 (트랜스레전 폴리머라제, Cas9 도메인, 및 C를 절제하는 염기 절제 효소 (UDG 204)를 함유하는 염기 편집제)를 사용한 WT 세포 내 FANCF 부위에서의 염기 편집을 보여준다. C에서 G로의 양은 HEK2 부위 내 특정 잔기에서 그래프로 예시된다. UDG 204는 C를 직접 제거하는 UDG 변이체이다.

도 48은 MSH2가 DNA 내 우라실 (U)에서 시토신 (C)으로의 전환을 용이하게 할 수 있는, 염기 복구에 있어서 MSH2의 역할을 예시하는 개략도를 보여준다.

도 49는 6종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; 및 BE3_UDG)를 사용한 MSH2-/- 세포 내 HEK2 부위에서의 염기 편집을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G 염기로의 편집은 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 50은 6종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; 및 BE3_UDG)를 사용한 MSH2-/- 세포 내 RNF2 부위에서의 염기 편집을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G 염기로의 편집은 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 51은 6종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; 및 BE3_UNG)를 사용한 MSH2-/- 세포 내 FANCF 부위에서의 염기 편집을 보여준다. 미가공 편집 값은 좌측 패널에 제시된다. 우측 패널은 미가공 편집 값의 그래프 표현을 보여주며, 여기서 C에서 G 염기로의 편집은 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다.

도 52는 UDG (또는 UDG 변이체), Cas9 (예를 들어, nCas9) 도메인 및 시티딘 데아미나제를 함유하는 C에서 G로의 염기 편집제가 트랜스레전 폴리머라제에 의해 트랜스로 발현되는 염기 편집 접근법을 예시하는 개략도를 보여준다.

도 53은 다양한 폴리머라제 (Pol 카파, Pol 에타, Pol 이오타, REV1, Pol 베타 및 Pol 델타)에 의해 트랜스로 발현된 5종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; 및 BE3_UDG)를 사용한 HEK293 세포 내 HEK2 부위에서의 염기 편집을 보여준다. C에서 G로의 염기 편집은 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 54는 다양한 폴리머라제 (Pol 카파, Pol 에타, Pol 이오타, REV1, Pol 베타 및 Pol 델타)에 의해 트랜스로 발현된 5종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; 및 BE3_UDG)를 사용한 HEK293 세포 내 RNF2 부위에서의 염기 편집을 보여준다. C에서 G로의 염기 편집은 도트형 막대 (G)가 채워진 막대 (C)로 이동한 것으로 그래프 제시된다.

도 55는 다양한 폴리머라제 (Pol 카파, Pol 에타, Pol 이오타, REV1, Pol 베타 및 Pol 델타)에 의해 트랜스로 발현된 5종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; 및 BE3_UDG)를 사용한 HEK293 세포 내 FANCF 부위에서의 염기 편집을 보여준다. C에서 G로의 염기 편집은 채워진 막대 (C)가 도트형 막대 (G)로 이동한 것으로 그래프 제시된다.

발명을 실시하기 위한 구체적인 내용

- [0007] 본원 및 청구범위에 사용된 단수 형태는 문맥이 달리 명백하게 나타내지 않는 한 단수 및 복수 지시대상을 포함한다. 따라서, 예를 들어, "작용제"에 대한 언급은 단일 작용제 및 복수의 이러한 작용제를 포함한다.
- [0008] 본원에 사용된 용어 "테아미나제" 또는 "테아미나제 도메인"은 탈아미노화 반응을 촉매하는 단백질 또는 효소를 지칭한다. 일부 실시양태에서, 테아미나제 또는 테아미나제 도메인은 각각 시티딘 또는 데옥시시티딘의 우리딘 또는 데옥시우리딘으로의 가수분해적 탈아미노화를 촉매하는 시티딘 테아미나제이다. 일부 실시양태에서, 테아미나제 또는 테아미나제 도메인은 시토신의 우라실로의 가수분해적 탈아미노화를 촉매하는 시티딘 테아미나제 도메인이다. 일부 실시양태에서, 테아미나제 또는 테아미나제 도메인은 유기체, 예컨대 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트 또는 마우스로부터의 자연-발생 테아미나제이다. 일부 실시양태에서, 테아미나제 또는 테아미나제 도메인은 자연에서 발생하지 않는 유기체로부터의 자연-발생 테아미나제의 변이체이다. 예를 들어, 일부 실시양태에서, 테아미나제 또는 테아미나제 도메인은 유기체로부터의 자연-발생 테아미나제와 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.
- [0009] 용어 "염기 편집제 (BE)" 또는 "핵염기 편집제 (NBE)"는 핵산 서열 (예를 들어, DNA 또는 RNA) 내의 염기 (예를 들어, A, T, C, G 또는 U)에 변형을 가할 수 있는 폴리펩티드를 포함하는 작용제를 지칭한다. 일부 실시양태에서, 염기 편집제는 핵산 내의 염기를 탈아미노화 수 있다. 일부 실시양태에서, 염기 편집제는 DNA 분자 내의 염기를 탈아미노화할 수 있다. 일부 실시양태에서, 염기 편집제는 DNA 내의 시토신 (C)을 탈아미노화할 수 있다. 일부 실시양태에서, 염기 편집제는 DNA 분자 내의 염기를 절제할 수 있다. 일부 실시양태에서, 염기 편집제는 핵산 (예를 들어, DNA 또는 RNA) 분자 내의 아데닌, 구아닌, 시토신, 티민 또는 우라실을 절제할 수 있다. 일부 실시양태에서, 염기 편집제는 시티딘 테아미나제에 융합된 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)을 포함하는 단백질 (예를 들어, 융합 단백질)이다. 일부 실시양태에서, 염기 편집제는 우라실 결합 단백질 (UBP), 예컨대 우라실 DNA 글리코실라제 (UDG)에 융합된다. 일부 실시양태에서, 염기 편집제는 핵산 폴리머라제 (NAP) 도메인에 융합된다. 일부 실시양태에서, NAP 도메인은 트랜스레전 DNA 폴리머라제이다. 일부 실시양태에서, 염기 편집제는 napDNAbp, 시티딘 테아미나제 및 UBP (예를 들어, UDG)를 포함한다. 일부 실시양태에서, 염기 편집제는 napDNAbp, 시티딘 테아미나제 및 핵산 폴리머라제 (예를 들어, 트랜스레전 DNA 폴리머라제)를 포함한다. 일부 실시양태에서, 염기 편집제는 napDNAbp, 시티딘 테아미나제, UBP (예를 들어, UDG) 및 핵산 폴리머라제 (예를 들어, 트랜스레전 DNA 폴리머라제)를 포함한다.
- [0010] 일부 실시양태에서, 염기 편집제의 napDNAbp는 Cas9 도메인이다. 일부 실시양태에서, 염기 편집제는 시티딘 테아미나제에 융합된 Cas9 단백질을 포함한다. 일부 실시양태에서, 염기 편집제는 시티딘 테아미나제에 융합된 Cas9 니카제 (nCas9)를 포함한다. 일부 실시양태에서, Cas9 니카제는 D10A 돌연변이를 포함하고, 서열식별번호: 6의 잔기 840에서의 히스티딘, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 이는 Cas9가 핵산 듀플렉스의 단지 한 가닥만을 절단할 수 있게 한다. 일부 실시양태에서, 염기 편집제는 시티딘 테아미나제에 융합된 뉴클레아제-불활성 Cas9 (dCas9)를 포함한다. 일부 실시양태에서, dCas9 도메인은 서열식별번호: 6의 D10A 및 H840A 돌연변이, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26 중 어느 하나에서의 상응하는 돌연변이를 포함하며, 이는 Cas9 단

백질의 뉴클레아제 활성을 불활성화시킨다.

[0011] 본원에 사용된 용어 "링커"는 2개의 분자 또는 모이어티, 예를 들어 융합 단백질의 2개의 도메인, 예컨대, 예를 들어 뉴클레아제-불활성 Cas9 도메인 및 핵산-편집 도메인 (예를 들어, 시티딘 데아미나제)을 연결하는 결합 (예를 들어, 공유 결합), 화학적 기 또는 분자를 지칭한다. 일부 실시양태에서, 링커는 Cas9 뉴클레아제 도메인을 포함한 RNA-프로그램가능 뉴클레아제의 gRNA 결합 도메인, 및 핵산-편집 단백질의 촉매 도메인을 연결한다. 일부 실시양태에서, 링커는 dCas9 및 핵산 편집 단백질을 연결한다. 전형적으로, 링커는 2개의 기, 분자 또는 다른 모이어티 사이에 위치하거나 그에 플랭킹되고, 공유 결합을 통해 서로 연결되어, 2개를 연결한다. 일부 실시양태에서, 링커는 아미노산 또는 복수의 아미노산 (예를 들어, 펩티드 또는 단백질)이다. 일부 실시양태에서, 링커는 유기 분자, 기, 중합체 또는 화학적 모이어티이다. 일부 실시양태에서, 링커는 5-100개 아미노산 길이, 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, 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. 보다 긴 또는 보다 짧은 링커가 또한 고려된다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 102)를 포함하며, 이는 또한 XTEN 링커로 지칭될 수 있다. 일부 실시양태에서, 링커는 아미노산 서열 SGGS (서열식별번호: 103)를 포함한다. 일부 실시양태에서, 링커는 (SGGS)_n (서열식별번호: 103), (GGGS)_n (서열식별번호: 104), (GGGG)_n (서열식별번호: 105), (G)_n (서열식별번호: 121), (EAAAK)_n (서열식별번호: 106), (GGG)_n (서열식별번호: 122), SGSETPGTSESATPES (서열식별번호: 102), (XP)_n 모티프 (서열식별번호: 123), SGSSGGSETPGTSESATPESGGSSGGSS (서열식별번호: 108), GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSGSGS (서열식별번호: 109), SGSSGGSGGS (서열식별번호: 120), 또는 이들 중 임의의 것의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이고, X는 임의의 아미노산이다. 일부 실시양태에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15이다.

[0012] 본원에 사용된 용어 "돌연변이"는 서열, 예를 들어 핵산 또는 아미노산 서열 내의 잔기의 또 다른 잔기로의 치환, 또는 서열 내의 1개 이상의 잔기의 결실 또는 삽입을 지칭한다. 돌연변이는 전형적으로 원래 잔기에 이어서 서열 내의 그 잔기의 위치 및 새롭게 치환된 잔기의 정체를 확인함으로써 본원에 기재된다. 본원에 제공된 아미노산 치환 (돌연변이)을 제조하는 다양한 방법은 관련 기술분야에 널리 공지되어 있고, 예를 들어 문헌 [Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 제공되어 있다.

[0013] 본원에 사용된 용어 "우라실 결합 단백질" 또는 "UBP"는 우라실에 결합할 수 있는 단백질을 지칭한다. 일부 실시양태에서, 우라실 결합 단백질은 우라실 변형 효소이다. 일부 실시양태에서, 우라실 결합 단백질은 우라실 염기 절제 효소이다. 일부 실시양태에서, 우라실 결합 단백질은 우라실 DNA 글리코실라제 (UDG)이다. 일부 실시양태에서, 우라실 결합 단백질은 야생형 UDG (예를 들어, 인간 UDG)가 우라실에 결합하는 친화도의 적어도 1%, 2%, 3%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 또는 적어도 95%인 친화도로 우라실에 결합한다.

[0014] 본원에 사용된 용어 "염기 절제 효소" 또는 "BEE"는 핵산 분자 (예를 들어, DNA 또는 RNA)로부터 염기 (예를 들어, A, T, C, G 또는 U)를 제거할 수 있는 단백질을 지칭한다. 일부 실시양태에서, BEE는 DNA로부터 시토신을 제거할 수 있다. 일부 실시양태에서, BEE는 DNA로부터 티민을 제거할 수 있다. 예시적인 BEE는, 비제한적으로, 문헌 [Sang et al., "A Unique Uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, Vol. 43, No. 17 2015]에 기재된 바와 같은 UDG Tyr147Ala 및 UDG Asn204Asp를 포함하고; 상기 문헌의 전체 내용은 본원에 참조로 포함된다.

[0015] 용어 "핵산 폴리머라제" 또는 "NAP"는 뉴클레오티드 (예를 들어, 데옥시리보뉴클레오티드 및 리보뉴클레오티드)로부터 핵산 분자 (예를 들어, DNA 및 RNA)를 합성하는 효소를 지칭한다. 일부 실시양태에서, NAP는 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 트랜스레전 폴리머라제이다. 트랜스레전 폴리머라제는 돌연변이유발에서, 예를 들어 복제 포크를 재시작하거나 또는 DNA 손상의 존재로 인해 게놈에 남아있는 갭을 채우는 것에 의해 소정 역할을 한다. 예시적인 트랜스레전 폴리머라제는, 비제한적으로, Pol 베타, Pol 람다, Pol 에타, Pol 뮤, Pol 이오타, Pol 카파, Pol 알파, Pol 델타, Pol 감마 및 Pol 누를 포함한다.

[0016] 용어 "핵 국제화 서열" 또는 "NLS"는, 예를 들어 핵 수송에 의해 세포 핵 내로의 단백질 유입을 촉진하는 아미

노산 서열을 지칭한다. 핵 국제화 서열은 관련 기술분야에 공지되어 있으며, 통상의 기술자에게 분명할 것이다. 일부 실시양태에서, NLS는 단일부분 NLS이다. 일부 실시양태에서, NLS는 이부분 NLS이다. 이부분 NLS는 비교적 짧은 스페이스 서열 (예를 들어, 2-20개의 아미노산, 5-15개의 아미노산, 또는 8-12개의 아미노산)에 의해 분리된다. 예를 들어, NLS 서열은 2000년 11월 23일에 출원되고 2001년 5월 31일에 WO/2001/038547로 공개된 국제 PCT 출원 PCT/EP2000/011690 (Plank et al.); 및 문헌 [Kethar, K.M.V., et al., "Application of bioinformatics-coupled experimental analysis reveals a new transport-competent nuclear localization signal in the nucleoprotein of Influenza A virus strain" BMC Cell Biol, 2008, 9: 22]에 기재되어 있고; 각각의 문헌의 내용은 예시적인 핵 국제화 서열의 개시를 위해 본원에 참조로 포함된다. 일부 실시양태에서, NLS는 아미노산 서열 PKKKRKV (서열식별번호: 41), MDSLLMNRKFLYQFKNVRWAKGRRETYLC (서열식별번호: 42), KRTADGSEFESPKKKRKV (서열식별번호: 43), KRGINDRNFWRGNGRKRTR (서열식별번호: 44), KKTGGPIYRRVDGKWR (서열식별번호: 45), RRELILYDKKEIRRIWR (서열식별번호: 46), 또는 AVSRKRKA (서열식별번호: 47)를 포함한다.

[0017]

용어 "핵산 프로그램가능 DNA 결합 단백질" 또는 "napDNAbp"는 핵산 (예를 들어, DNA 또는 RNA), 예컨대 napDNAbp를 특이적 핵산 서열로 가이드하는 가이드 핵산과 회합되는 단백질을 지칭한다. 예를 들어, Cas9 단백질은 Cas9 단백질을 가이드 RNA에 상보적인 특정 DNA 서열로 가이드하는 가이드 RNA와 회합될 수 있다. 일부 실시양태에서, napDNAbp는 부류 2 미생물 CRISPR-Cas 이펙터이다. 일부 실시양태에서, napDNAbp는 Cas9 도메인, 예를 들어 뉴클레아제 활성 Cas9, Cas9 니카제 (nCas9) 또는 뉴클레아제 불활성 Cas9 (dCas9)이다. 핵산 프로그램가능 DNA 결합 단백질의 예는, 비제한적으로, Cas9 (예를 들어, dCas9 및 nCas9), CasX, CasY, Cpf1, C2c1, C2c2, C2c3 및 아르코노트를 포함한다. 그러나, 핵산 프로그램가능 DNA 결합 단백질은 또한 RNA에 결합하는 핵산 프로그램가능 단백질을 포함한다는 것이 인지되어야 한다. 예를 들어, napDNAbp는 napDNAbp를 RNA로 가이드하는 핵산과 회합될 수 있다. 다른 핵산 프로그램가능 DNA 결합 단백질은 이들이 본 개시내용에 구체적으로 열거되어 있지 않을 수 있더라도 또한 본 개시내용의 범주 내에 있다.

[0018]

용어 "Cas9" 또는 "Cas9 도메인"은 Cas9 단백질, 또는 그의 단편 (예를 들어, Cas9의 활성, 불활성, 또는 부분적 활성 DNA 절단 도메인, 및/또는 Cas9의 gRNA 결합 도메인을 포함하는 단백질)을 포함하는 RNA-가이드 뉴클레아제를 지칭한다. Cas9 뉴클레아제는 또한 때때로 casn1 뉴클레아제 또는 CRISPR (클러스터링된 규칙적 간격의 짧은 회문식 반복부)-연관 뉴클레아제로 지칭된다. CRISPR은 이동성 유전 요소 (바이러스, 전위가능 요소 및 접합 플라스미드)에 대한 보호를 제공하는 적응 면역계이다. CRISPR 클러스터는 스페이스, 선행 이동성 요소에 상보적인 서열, 및 표적 침입 핵산을 함유한다. CRISPR 클러스터는 전사되고, CRISPR RNA (crRNA)로 프로세싱된다. 유형 II CRISPR 시스템에서 프리-crRNA의 교정 프로세싱은 트랜스-코딩된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 단백질을 요구한다. tracrRNA는 프리-crRNA의 리보뉴클레아제 3-보조된 프로세싱을 위한 가이드로서의 역할을 한다. 후속적으로, Cas9/crRNA/tracrRNA는 스페이스에 상보적인 선행 또는 원형 dsDNA 표적을 핵산내부분해로 절단한다. crRNA에 상보적이지 않은 표적 가닥은 먼저 핵산내부분해로 절단된 다음, 핵산외부분해로 3'-5' 트리밍된다. 사실상, DNA-결합 및 절단은 전형적으로 단백질 및 둘 다의 RNA를 요구한다. 그러나, 단일 가이드 RNA ("sgRNA" 또는 간단히 "gRNA")는 crRNA 및 tracrRNA 둘 다의 측면을 단일 RNA 종 내로 혼입하도록 조작될 수 있다. 예를 들어, 문헌 [Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)]을 참조하며, 이의 전체 내용은 본원에 참조로 포함된다. Cas9는 자기 대 비-자기를 구별하는 것을 돕기 위한 CRISPR 반복 서열 (PAM 또는 프로토스페이스 인접 모티프)에서의 짧은 모티프를 인식한다. Cas9 뉴클레아제 서열 및 구조는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다 (예를 들어, 문헌 ["Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti et al., J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); 및 "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)] 참조, 상기 문헌 각각의 전체 내용은 본원에 참조로 포함됨). Cas9 오르토로그는 에스. 피오게네스(*S. pyogenes*) 및 에스. 씰모필루스(*S. thermophilus*)를 포함하나 이에 제한되지는 않는 다양한 종에서 설명되었다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 분명할 것이며, 이러한 Cas9 뉴클레아제 및 서열은 문헌 [Chylinski, Rhun, and

Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737] (이의 전체 내용은 본원에 참조로 포함됨)에 개시된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다. 일부 실시양태에서, Cas9 뉴클레아제는 불활성 (예를 들어, 불활성화된) DNA 절단 도메인을 가지며, 즉, Cas9는 니카제이다.

[0019] 뉴클레아제-불활성화된 Cas9 단백질은 "dCas9" 단백질 (뉴클레아제-"기능상실" Cas9에 대한 것)로 상호교환가능하게 지칭될 수 있다. 불활성 DNA 절단 도메인을 갖는 Cas9 단백질 (또는 그의 단편)을 생성하는 방법은 공지되어 있다 (예를 들어, 문헌 [Jinek et al., Science. 337:816-821(2012); Qi et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) Cell. 28;152(5):1173-83] 참조, 상기 문헌 각각의 전체 내용은 본원에 참조로 포함됨). 예를 들어, Cas9의 DNA 절단 도메인은 2종의 서브도메인인 HNH 뉴클레아제 서브도메인 및 RuvC1 서브도메인을 포함하는 것으로 공지되어 있다. HNH 서브도메인은 gRNA에 상보적인 가닥을 절단하고, 반면에 RuvC1 서브도메인은 비-상보적 가닥을 절단한다. 이들 서브도메인 내의 돌연변이는 Cas9의 뉴클레아제 활성을 침묵시킬 수 있다. 예를 들어, 돌연변이 D10A 및 H840A는 에스. 피오게네스 Cas9의 뉴클레아제 활성을 완전히 불활성화시킨다 (Jinek et al., Science. 337:816-821(2012); Qi et al., Cell. 28;152(5):1173-83 (2013)). 일부 실시양태에서, Cas9의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 실시양태에서, 단백질은 다음 2개의 Cas9 도메인 중 1개를 포함한다: (1) Cas9의 gRNA 결합 도메인; 또는 (2) Cas9의 DNA 절단 도메인. 일부 실시양태에서, Cas9 또는 그의 단편을 포함하는 단백질은 "Cas9 변이체"로 지칭된다. Cas9 변이체는 Cas9 또는 그의 단편에 대한 상동성을 공유한다. 예를 들어 Cas9 변이체는 야생형 Cas9에 대해 적어도 약 70% 동일하거나, 적어도 약 80% 동일하거나, 적어도 약 90% 동일하거나, 적어도 약 95% 동일하거나, 적어도 약 96% 동일하거나, 적어도 약 97% 동일하거나, 적어도 약 98% 동일하거나, 적어도 약 99% 동일하거나, 적어도 약 99.5% 동일하거나, 또는 적어도 약 99.9% 동일하다. 일부 실시양태에서, Cas9 변이체는 야생형 Cas9와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 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, 50개 또는 그 초과인 아미노산 변화를 가질 수 있다. 일부 실시양태에서, Cas9 변이체는 Cas9의 단편 (예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함하며, 이러한 단편은 야생형 Cas9의 상응하는 단편에 대해 적어도 약 70% 동일하거나, 적어도 약 80% 동일하거나, 적어도 약 90% 동일하거나, 적어도 약 95% 동일하거나, 적어도 약 96% 동일하거나, 적어도 약 97% 동일하거나, 적어도 약 98% 동일하거나, 적어도 약 99% 동일하거나, 적어도 약 99.5% 동일하거나, 또는 적어도 약 99.9% 동일하다. 일부 실시양태에서, 단편은 상응하는 야생형 Cas9의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5%이다.

[0020] 일부 실시양태에서, 단편은 적어도 100개 아미노산의 길이이다. 일부 실시양태에서, 단편은 적어도 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250 또는 1300개 아미노산의 길이이다. 일부 실시양태에서, 야생형 Cas9는 스트렙토코쿠스 피오게네스로부터의 Cas9 (NCBI 참조 서열: NC_017053.1, 서열식별번호: 1 (뉴클레오타이드); 서열식별번호: 4 (아미노산))에 상응한다.

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGG
GCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAA
ATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAG
TGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATAC
ACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCG
AAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAG
ACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTA
TCATGAGAAATATCCAACATCTATCATCTGCGAAAAAATTGGCAGATTCTACT
GATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTC
GTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAA
ACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATT
AACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCA
AGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAAATGGCTTG
TTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTT
TGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAGATACTTACGATGATGAT
TTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAG
CTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGA
AATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCAT
CAAGACTTGACTCTTTTAAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATA
AAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGG

[0021]

AGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGAT
GGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAA
CGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACCTGGGTGAGCTGCATG
CTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTAAAAGACAATCGTGAGAA
GATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTG
GCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATG
GAATTTGAAGAAGTTGTCGATAAAGGTGCTCAGCTCAATCATTTATTGAACGC
ATGACAACTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGT
TTGCTTTATGAGTATTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTA
CTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTG
TTGATTTACTCTTCAAAACAAATCGAAAAAGTAACCGTTAAGCAATTAAGAAG
ATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTCAGGAGTTGAAGA
TAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGAT
AAAGATTTTTTGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAA
CATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATG
CTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGG
TTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGC
AAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGC
AGCTGATCCATGATGATAGTTTTGACATTTAAAGAAGATATTCAAAAAGCACAGG
TGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCC
TGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAA
GTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAG
ACAATCAAAAAGGGCCAGAAAATTCGCGAGAGCGTATGAAACGAATCGAAGA
AGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATAC
TCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATG
TATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACA
TTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCG
TTCTGATAAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAA
AAAGATGAAAAACTATTGGAGACAACCTTCTAAACGCCAAGTTAATCACTCAACG
TAAGTTTGATAATTTAACGAAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAA
GCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGG
CACAAATTTTGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTAT
TCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTCCGAAAA
GATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATG

[0022]

CGTATCTAAATGCCGTCGTTGGAACGCTTTGATTAAGAAATATCCAAAACCTTGA
 ATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCT
 AAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATA
 TCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTTCGCAAC
 GCCCTCTAATCGAAACTAATGGGGAACTGGAGAAATTGTCTGGGATAAAGGGC
 GAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAA
 GAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAG
 AAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGG
 TGGTTTGTAGTCCAACGGTAGCTTATTCAGTCCTAGTGTTGCTAAGGTGGAA
 AAAGGGAAATCGAAGAAGTTAAAATCCGTAAAGAGTTACTAGGGATCACAATT
 ATGGAAAGAAGTTCCTTTGAAAAAATCCGATTGACTTTTTAGAAAGCTAAAGGAT
 ATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGA
 GTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGG
 AAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTATATTAGCTAGTCAT
 TATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTG
 GAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTCTA
 AGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTCTTAGTGCATATAACAA
 ACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTAC
 GTTGACGAATCTTGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATC
 GTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATC
 CATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA
 (SEQ ID NO: 1)

MDKKYSIGLDIGTNSVGWAVITDDYKVP SKKFKVLGNTDRHSIKKNLIGALLFGSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEG
 DLNPDNSDVKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRLENLIAQLPG
 EKRNLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYAD
 LFLAAKNLSDAILSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
 YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
 FDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFA
 WMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
 YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEER

[0023]

LKTYAHLFDDKVMKQLKRRRYTGWGRSLRKLINGIRDKQSGKTILDFLKSDGFANR
 NFMQLIHDDSLTFKEDIQKAQVSGQGHSLSHEQIANLAGSPAIIKGILQTVKIVDELVK
VMGHKPENVIEMARENQTTQKGQKNSRERMKRIEIGIKELGSQILKEHPVENTOLO
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKMKKNYWRQLLNAKLITQRFDNLTKAERGGELSELDKAGFIKRO
LVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKAT
AKYFFYSNIMNFEKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQ
VNIVKKTEVQTGGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
 VEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIILPKYSLFELE
 NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHK
 HYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPA
 AFKYFDTTIDRKRYTSTKEVLDTLHQSTGLYETRIDLSQLGGD (SEQ ID NO: 4)

[0024]

[0025]

(단일 밀줄: HNH 도메인; 이중 밀줄: RuvC 도메인)

[0026]

일부 실시양태에서, 야생형 Cas9는 서열식별번호: 2 (뉴클레오티드) 및/또는 서열식별번호: 5 (아미노산)에 상응하거나, 또는 이를 포함한다:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATCCGTTGGATGGG
 CTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGA
 ACACAGACCGTCATTTCGATTAATAAAGAATCTTATCGGTGCCCTCTATTCGATAG
 TGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATAC
 ACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTAGCAATGAGATGGCC
 AAAGTTGACGATTCTTTCTTTACCGTTTGAAGAGTCCTTCCTTGTCGAAGAGG
 ACAAGAAACATGAACGGCACCCCATCTTTGGAACATAGTAGATGAGGTGGCAT
 ATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAA
 CTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTT
 CCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACCTCGGATGTCGAC
 AAAGTTGATCCAGTTAGTACAAACCTATAATCAGTTGTTGAAGAGAACCCTA
 TAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCTCTCTAAATC
 CCGACGGCTAGAAAACCTGATCGCACAAATTACCGGAGAGAAGAAAAATGGGTT
 GTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTAAGTCGAAC
 TTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGAC
 GATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTGG
 CTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATAC
 TGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACAT

[0027]

CACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAAT
 ATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACG
 GCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGA
 TGGATGGGACGGAAGAGTTGCTTGTAAACTCAATCGCGAAGATCTACTGCGAA
 AGCAGCGGACTTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATT
 GCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCCTCAAAGACAATCGT
 GAAAAGATTGAGAAAATCCTAACCTTTTCGCATACCTTACTATGTGGGACCCCTGG
 CCCGAGGGAACCTCTCGGTTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTA
 CTCCATGGAATTTTGAGGAAGTTGTGCGATAAAGGTGCGTCAGCTCAATCGTTTCAT
 CGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAA
 GCACAGTTTACTTTACGAGTATTTACAGTGTACAATGAACTCACGAAAGTTAAG
 TATGTCAGTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAA
 GCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTG
 AAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTGAGATCTCCGGGG
 TAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAAT
 TAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATAT
 AGTGTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAA
 ACATACGCTCACCTGTTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCT
 ATACGGGCTGGGGACGATTGTGCGCGAAACTTATCAACGGGATAAGAGACAAGC
 AAAGTGGAATAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAA
 CTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAG
 GCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTG
 GTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGC
 TAGTTAAGGTCATGGGACGTCACAAACCGGAAAAACATTGTAATCGAGATGGCAC
 GCGAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCGAGAGCGGATGAAG
 AGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCT
 GTGGAAAATACCCAATTGCAGAACGAGAACTTTACCTCTATTACCTACAAAATG
 GAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGA
 CGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAA
 GTGCTTACACGCTCGGATAAGAACCGAGGGGAAAAGTGACAATGTTCCAAGCGAG
 GAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAACTG
 ATAACGCAAAGAAAGTTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCT
 GAACTTGACAAGGCCGATTATTAACGTCAGCTCGTGGAACCCGCCAAATC
 ACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAG

[0028]

AACGATAAGCTGATTTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTG
TCGGACTTCAGAAAGGATTTTCAATTCTATAAAAGTTAGGGAGATAAATAACTACC
ACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAA
ATACCCGAAGCTAGAAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTC
CGTAAGATGATCGCGAAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATA
CTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACG
GAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCG
TATGGGATAAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCC
AAGTCAACATAGTAAAGAAAAGTGAAGTGCAGACCGGAGGGTTTTCAAAGGAAT
CGATTCTTCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGG
ACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGT
AGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAACTGAAGTCAGTCAAAGAAT
TATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTT
CCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAGGATCTCATAATTAACTACC
AAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGC
CGGAGAGCTTCAAAAGGGGAACGAACCTCGCACTACCGTCTAAATACGTGAATTT
CCTGTATTTAGCGTCCCATACGAGAAGTTGAAAGGTTACCTGAAGATAACGAA
CAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGC
AAATTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGT
ATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAA
ATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTAT
TTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGAC
GCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGT
CACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACT
ACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACG
ATGACAAGGCTGCAGGA (SEQ ID NO: 2)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLEDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLYESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR

[0029]

TFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPOSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPOVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIILPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVLADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDTLIHQSTGLYETRIDLSQLGGD (SEQ ID NO: 5)

[0030]

[0031]

(단일 밀줄: HNH 도메인; 이중 밀줄: RuvC 도메인)

[0032]

일부 실시양태에서, 야생형 Cas9는 스트랩토코쿠스 피오게네스로부터의 Cas9 (NCBI 참조 서열: NC_002737.2, 서열식별번호: 3 (뉴클레오타이드); 및 유니프루트 참조 서열: Q99ZW2, 서열식별번호: 6 (아미노산))에 상응한다.

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGG
GCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAA
ATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAG
TGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATAC
ACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCG
AAAGTAGATGATAGTTTCTTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAG
ACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTA
TCATGAGAAATATCCAATCTATCATCTGCGAAAAAATTGGTAGATTCTACT
GATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTC
GTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAA
ACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATT
AACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCA
AGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTA

[0033]

TTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTT
 TGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGAT
 TTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAG
 CTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGA
 AATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCAT
 CAAGACTTGACTCTTTTAAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATA
 AAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGG
 AGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGAT
 GGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAA
 CGGACCTTTGACAACGGCTCTATTCCCATCAAATTCACCTGGGTGAGCTGCATG
 CTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTAAAAGACAATCGTGAGAA
 GATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTG
 GCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATG
 GAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTATTGAACGC
 ATGACAACTTTGATAAAAAATCTTCAAATGAAAAAGTACTACCAAAACATAGT
 TTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTA
 CTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTG
 TTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAGAAG
 ATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTCAGGAGTTGAAGA
 TAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGAT
 AAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAA
 CATTGACCTTATTGGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATG
 CTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGG
 TTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGC
 AAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGC
 AGCTGATCCATGATGATAGTTTTGACATTTAAAGAAGACATTCAAAAAGCACAAG
 TGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCC
 TGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAA
 GTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAAT
 CAGACAACCTCAAAAAGGGCCAGAAAATTCGCGAGAGCGTATGAAACGAATCGA
 AGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAA
 TACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGAC
 ATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATC
 ACATTGTTCCACAAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAAC

[0034]

GCGTTCTGATAAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGT
 CAAAAAGATGAAAAACTATTGGAGACAACCTTCTAAACGCCAAGTTAATCACTCA
 ACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACCTGAT
 AAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATG
 TGGCACAAATTTTGATAGTCGCATGAATACTAAATACGATGAAAAATGATAAAC
 TTATTCGAGAGGTTAAAGTGATTACCTTAAAAATCTAAATTAGTTTCTGACTTCCG
 AAAAGATTTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCAT
 GATGCGTATCTAAATGCCGTCGTTGGAAGTGCCTTTGATTAAAGAAATATCCAAAAC
 TTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCTGTAATGATT
 GCTAAGTCTGAGCAAGAAATAGGCAAGCAACCGCAAAATATTTCTTTTACTCTA
 ATATCATGAACCTTCTCAAACAGAAATTACACTTGCAAATGGAGAGATTTCGCAA
 ACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGG
 GCGAGATTTTGCCACAGTGCACAAAGTATTGTCCATGCCCCAAGTCAATATTGTC
 AAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAA
 AGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATAT
 GGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGG
 AAAAAGGGAAATCGAAGAAGTTAAAAATCCGTAAAGAGTTACTAGGGATCACAA
 TTATGGAAAGAAGTTTCCTTTGAAAAAATCCGATTGACTTTTTAGAAAGCTAAAGG
 ATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTT
 GAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAA
 GGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTC
 ATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTG
 TGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTC
 TAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGATATAAC
 AACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTT
 ACGTTGACGAATCTTGAGACTCCCGCTGCTTTTAAATATTTTGATACAACAATTG
 ATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCA
 ATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGAC
 TGA (SEQ ID NO: 3)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLEDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIKFRGHFLIEG
 DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLP

[0035]

GEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDLVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEKGILGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
PQVNIVKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKDKLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDTLIHQSTGLYETRIDLSQLGGD (SEQ ID NO: 6)

[0036]

[0037]

(단일 밀줄: HNH 도메인; 이중 밀줄: RuvC 도메인)

[0038]

일부 실시양태에서, Cas9는 하기로부터의 Cas9를 지칭한다: 코리네박테리움 울세란스(*Corynebacterium ulcerans*) (NCBI Ref: NC_015683.1, NC_017317.1); 코리네박테리움 디프테리아(*Corynebacterium diphtheria*) (NCBI Ref: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라(*Spiroplasma syrophidicola*) (NCBI Ref: NC_021284.1); 프레보텔라 인터메디아(*Prevotella intermedia*) (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨세(*Spiroplasma taiwanense*) (NCBI Ref: NC_021846.1); 스트렙토코쿠스 이니아에(*Streptococcus iniae*) (NCBI Ref: NC_021314.1); 벨리엘라 발티카(*Belliella baltica*) (NCBI Ref: NC_018010.1); 사이크로플렉수스 토르퀴시(*Psychroflexus torquis*) (NCBI Ref: NC_018721.1); 스트렙토코쿠스 써모필루스(*Streptococcus thermophilus*) (NCBI Ref: YP_820832.1); 리스테리아 인노쿠아(*Listeria innocua*) (NCBI Ref: NP_472073.1); 캄필로박터 제주니(*Campylobacter jejuni*) (NCBI Ref: YP_002344900.1); 또는 네이세리아 메닝기티디스(*Neisseria meningitidis*) (NCBI Ref: YP_002342100.1); 또는 임의의 다른 유기체로부터의 Cas9.

[0039]

일부 실시양태에서, dCas9는 Cas9 뉴클레아제 활성을 불활성화시키는 1개 이상의 돌연변이를 갖는 Cas9 아미노산 서열에 상응하거나, 또는 이를 부분적으로 또는 전체적으로 포함한다. 예를 들어, 일부 실시양태에서, dCas9 도메인은 서열식별번호: 6의 D10A 및 H840A 돌연변이 또는 또 다른 Cas9에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, dCas9는 서열식별번호: 7의 아미노산 서열을 포함한다.

[0040] dCas9 (D10A 및 H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
 DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
 GEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYA
 DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
 KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
 TFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFA
 WMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
 YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERL
 KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
 FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDDAIVPOSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGEIK
RQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFEKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
 AKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIHKLPKYSLFE
 LENGKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
 HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA
 PAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD (SEQ ID NO: 7)

[0041]

[0042] (단일 밀줄: HNH 도메인; 이중 밀줄: RuvC 도메인)

[0043]

일부 실시양태에서, Cas9 도메인은 D10A 돌연변이를 포함하고, 반면에 위치 840에서의 잔기는 서열식별번호: 6에 제공된 아미노산 서열에서 또는 또 다른 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에 제시된 Cas9에서의 상응하는 위치에서 히스티딘을 유지한다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, 촉매 잔기 H840의 존재는 표적화된 A 반대편에 T를 함유하는 비-편집 (예를 들어, 비-탈아미노화) 가닥을 절단하기 위한 Cas9의 활성을 유지한다. H840의 회복 (예를 들어, dCas9의 A840으로부터)은 A를 함유하는 표적 가닥의 절단을 유발하지 않는다. 이러한 Cas9 변이체는 gRNA-한정된 표적 서열에 기초하여 특정 위치에서 단일-가닥 DNA 파괴 (닉)를 생성할 수 있어, 비-편집 가닥의 복구를 유도함으로써, 궁극적으로 비-편집 가닥 상의 T에서 C로의 변화를 유발한다.

[0044]

다른 실시양태에서, 예를 들어 뉴클레아제 불활성화된 Cas9 (dCas9)를 유발하는, D10A 및 H840A 이외의 돌연변이를 갖는 dCas9 변이체가 제공된다. 이러한 돌연변이는, 예로서, D10 및 H840에서의 다른 아미노산 치환, 또는 Cas9의 뉴클레아제 도메인 내에서의 다른 치환 (예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인 내에서의 치환)을 포함한다. 일부 실시양태에서, 서열식별번호: 6, 7, 8, 9 또는 22에 대해 적어도 약 70% 동일하거나, 적어도 약 80% 동일하거나, 적어도 약 90% 동일하거나, 적어도 약 95% 동일하거나, 적어도 약 98% 동일하거나, 적어도 약 99% 동일하거나, 적어도 약 99.5% 동일하거나, 또는 적어도 약 99.9% 동일한 dCas9의 변이체 또는 상동체 (예를 들어, 서열식별번호: 6, 7, 8, 9 또는 22의 변이체)가 제공된다. 일부 실시양태에서, 약 5개 아미노산, 약 10개 아미노산, 약 15개 아미노산, 약 20개 아미노산, 약 25개 아미노산, 약 30개 아미노산, 약 40개 아미노산, 약 50개 아미노산, 약 75개 아미노산, 약 100개 아미노산 또는 그 초과 만큼 서열식별번호: 7, 8, 9 또는 22보다 더 짧거나 더 긴 아미노산 서열을 갖는 dCas9의 변이체 (예를 들어, 서열식별번호: 6, 7, 8, 9 또는 22의 변이체)가 제공된다.

[0045]

일부 실시양태에서, 본원에 제공된 바와 같은 Cas9 융합 단백질은 Cas9 단백질의 전장 아미노산 서열, 예를 들

어 본원에 제공된 Cas9 서열 중 하나를 포함한다. 그러나, 다른 실시양태에서, 본원에 제공된 바와 같은 융합 단백질은 전장 Cas9 서열을 포함하는 것이 아니라, 단지 그의 단편만을 포함한다. 예를 들어, 일부 실시양태에서, 본원에 제공된 Cas9 융합 단백질은 Cas9 단편을 포함하며, 여기서 단편은 crRNA 및 tracrRNA 또는 sgRNA에 결합하지만, 예를 들어 그것이 단지 뉴클레아제 도메인의 말단절단된 형태만을 포함하거나 또는 뉴클레아제 도메인을 전혀 포함하지 않는다는 점에서 기능적 뉴클레아제 도메인을 포함하지 않는다.

[0046] 적합한 Cas9 도메인 및 Cas9 단편의 예시적인 아미노산 서열은 본원에 제공되며, Cas9 도메인 및 단편의 추가의 적합한 서열은 관련 기술분야의 통상의 기술자에게 분명할 것이다.

[0047] 일부 실시양태에서, Cas9는 하기로부터의 Cas9를 지칭한다: 코리넨박테리움 울세란스 (NCBI Ref: NC_015683.1, NC_017317.1); 코리넨박테리움 디프테리아 (NCBI Ref: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라 (NCBI Ref: NC_021284.1); 프레보텔라 인터메디아 (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨세 (NCBI Ref: NC_021846.1); 스트렙토코쿠스 이니아에 (NCBI Ref: NC_021314.1); 헬리엘라 발티카 (NCBI Ref: NC_018010.1); 사이크로플렉스 토르퀴시 (NCBI Ref: NC_018721.1); 스트렙토코쿠스 쉐모필루스 (NCBI Ref: YP_820832.1); 리스테리아 인노쿠아 (NCBI Ref: NP_472073.1); 캄필로박터 제주니 (NCBI Ref: YP_002344900.1); 또는 네이세리아 메닌기티디스 (NCBI Ref: YP_002342100.1).

[0048] 추가의 Cas9 단백질 (예를 들어, 뉴클레아제 기능상실 Cas9 (dCas9), Cas9 니카제 (nCas9) 또는 뉴클레아제 활성 Cas9)은, 그의 변이체 및 상동체를 포함하여, 본 개시내용의 범주 내에 있다는 것이 인지되어야 한다. 예시적인 Cas9 단백질은, 비제한적으로, 하기 제공된 것을 포함한다. 일부 실시양태에서, Cas9 단백질은 뉴클레아제 기능상실 Cas9 (dCas9)이다. 일부 실시양태에서, dCas9는 아미노산 서열 (서열식별번호: 7, 8, 9 또는 22)을 포함한다. 일부 실시양태에서, Cas9 단백질은 Cas9 니카제 (nCas9)이다. 일부 실시양태에서, nCas9는 아미노산 서열 (서열식별번호: 10, 13, 16 또는 21)을 포함한다. 일부 실시양태에서, Cas9 단백질은 뉴클레아제 활성 Cas9이다. 일부 실시양태에서, 뉴클레아제 활성 Cas9는 아미노산 서열 (서열식별번호: 4, 5, 6, 11, 12, 14, 15, 16, 17, 18, 19, 20, 23, 24, 25 또는 26)을 포함한다.

[0049]

예시적인 촉매 불활성 Cas9 (dCas9):

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETA
 EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH
 PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGE
 KKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADL
 FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
 KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
 DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
 MTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVY
 NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
 VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERL
 KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
 FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDLVK
 VMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGSGILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK
 NRGKSDNPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGELSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
 REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
 ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
 PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
 AKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFE
 LENGKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
 HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA
 PAAFKYFDTTIDRKRYTSTKEVLDTLHQSIITGLYETRIDLSQLGGD (SEQ ID NO: 8)

[0050]

[0051]

예시적인 Cas9 니카제 (nCas9):

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETA
 EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH
 PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDL
 NPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGE
 KKNGLFGNLIASLGLTPNFKSNFLAEDAKLQLSKDTYDDDLNLLAQIGDQYADL
 FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
 KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
 DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAW
 MTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVY
 NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
 VEISGVEDRFNASLGTYHDLKKIHKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERL
 KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRN
 FMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIANLAGSPAICKGILQTVKVVDELVK
 VMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
 NRGKSDNPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
 REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
 ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
 PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
 AKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIILPKYSLFE
 LENGKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
 HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA
 PAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD (SEQ ID NO:

10)

[0052]

[0053]

예시적인 촉매 활성 Cas9:

DKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLLYLALAHMIKFRGHFLIEGDL
NPDNSDVKDLFIQLVQTYNQLFEEPNINASGVDAKAILSARLSKSRRLLENLIAQLPGE
KKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDQKSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLMS
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDTLHQSTGLYETRIDLSQLGGD (SEQ ID NO:

[0054]

11).

[0055]

본원에 사용된 용어 "Cas9 니카제"는 듀플렉스 핵산 분자 (예를 들어, 듀플렉스 DNA 분자)의 단지 1개의 가닥만을 절단할 수 있는 Cas9 단백질을 지칭한다. 일부 실시양태에서, Cas9 니카제는 D10A 돌연변이를 포함하고, 서열식별번호: 6의 위치 H840에서의 히스티딘, 또는 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26 중 어느 하나에서의 상응하는 돌연변이를 갖는다. 예를 들어, Cas9 니카제는 서열식별번호: 10, 13, 16 또는 21에 제시된 바와 같은 아미노산 서열을 포함할 수 있다. 이러한 Cas9 니카제는 활성 HNH 뉴클레아제 도메인을 갖고, DNA의 비-표적화된 가닥, 즉 gRNA에 의해 결합된 가닥을 절단할 수 있다. 추가로, 이러한 Cas9 니카제는 불활성 RuvC 뉴클레아제 도메인을 갖고, DNA의 표적화된 가닥, 즉 염기 편집이 요망되는 가닥을 절단할 수 없다.

[0056]

일부 실시양태에서, Cas9는 단세포형 원핵 미생물의 도메인 및 계를 구성하는 고세균 (예를 들어 나노고세균)으로부터의 Cas9를 지칭한다. 일부 실시양태에서, Cas9는 CasX 또는 CasY를 지칭하며, 이는 예를 들어 문헌 [Burstein et al., "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21] (이의 전체 내용은 본원에 참조로 포함됨)에 기재되었다. 게놈-분해 메타유전체학을 사용하여, 생명체의 고세균 도메인에서 최초 보고된 Cas9를 포함한, 수많은 CRISPR-Cas 시스템이 확인되었다. 이 분기 Cas9 단백질은 활성 CRISPR-Cas 시스템의 일부로서 거의 연구되지 않은 나노고세균에서 발견되었다. 박테리아에서, 2종의 이전에 공지되지 않은 시스템인 CRISPR-CasX 및 CRISPR-CasY가 발견되었으며, 이는 이제껏 발견된 시스템 중 가장 소형이다. 일부 실시양태에서, Cas9는 CasX 또는 CasX의 변이체를 지칭한다. 일부 실시양태에서, Cas9는 CasY 또는 CasY의 변이체를 지칭한다. 다른 RNA-가이드 DNA 결합 단백질은 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)로서 사용될 수 있고 본 개시내용의 범주 내에 있다는 것이 인지되어야 한다.

[0057]

일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)은 CasX 또는 CasY 단백질일 수 있다. 일부 실시양태에서, napDNAbp는 CasX 단백질이다. 일부 실시양태에서, CasX 단백질은 뉴클레아제 불활성 CasX 단백질 (dCasX), CasX 니카제 (CasXn), 또는 뉴클레아제

활성 CasX이다. 일부 실시양태에서, napDNAbp는 CasY 단백질이다. 일부 실시양태에서, CasY 단백질은 뉴클레아제 불활성 CasY 단백질 (dCasY), CasY 니카제 (CasYn), 또는 뉴클레아제 활성 CasY이다. 일부 실시양태에서, napDNAbp는 자연 발생 CasX 또는 CasY 단백질에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, napDNAbp는 자연 발생 CasX 또는 CasY 단백질이다. 일부 실시양태에서, napDNAbp는 서열식별번호: 27-29 중 어느 하나에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, napDNAbp는 서열식별번호: 27-29 중 어느 하나의 아미노산 서열을 포함한다. 다른 박테리아 종으로부터의 CasX 및 CasY는 또한 본 개시내용에 따라 사용될 수 있다는 것이 인지되어야 한다.

[0058] CasX (uniprot.org/uniprot/F0NN87; <http://www.uniprot.org/uniprot/F0NH53>)

[0059] >tr|F0NN87|F0NN87_SULIH CRISPR-연관 Casx 단백질 OS=솔폴로부스 이슬란디쿠스 (균주 HVE10/4) GN=SiH_0402 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFNFSQV
KECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAGWVLTRLGKAK
VSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFLGLWIARKVVSSVTNPVSV
VRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVL
ANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG
(SEQ ID NO: 27)

[0060]

[0061] >tr|F0NH53|F0NH53_SULIR CRISPR 연관 단백질, Casx OS=솔폴로부스 이슬란디쿠스 (균주 REY15A) GN=SiRe_0771 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFNFSQV
KECEEVSAPSFVKPEFYKFGGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKA
KVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFLGLWIARKVVSSVTNPVSV
VVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIV
LANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG
(SEQ ID NO: 28)

[0062]

[0063] CasY (ncbi.nlm.nih.gov/protein/APG80656.1)

[0064] >APG80656.1 CRISPR-연관 단백질 CasY [비배양 파르쿠파테리아 군 박테리움]

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDD
YVGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNAEV
RGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRE
RHKDQC�KLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTC
CLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGE
GFLGRLRENKITEKKAMMDITDAWRGQEQEEELEKRLRILAALTIKREPKFDNH
GGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKDLKKAKEMINRFGESDTKEEAV
VSSLESIEKIVPDDSDADDEKPDIPAIAYRRFLSDGRLTLNRFVQREDVQEALIKERLE
AEKKKKPKKRRKKSDAEDEKETIDFKELFPHLAKPLKLVNFYGDSKRELYKKYKN
AAIYTDALWKAVEKIYKSAFSSSLKNSFFDITDFDKDFFIKRLQKIFS VYRRFNTDKWK
PIVKN SFAPYCDIVSLAENEVLYKPKQSRSRKSAIDKNRVRLPSTENIAKAGIALARE
LSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLAVTETQLDISALDFVENGTVKD
FMKTRDGNLVLEGRFLEMFSQSIVFSELRLGLMSRKEFITRSAIQTMNGKQAE
LLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLKLKQMRYYPHYFGY
ELTRTGQGDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQ
FLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPF
TIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGL
KLDQRRGT FAMPSTKIARIESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKI
KKVYATLKKADVSEIDADKNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAE
MQVDETITTTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISKK
MRGNSCLFICPFCRANADADIQASQTIALRLRYVKEEKKVEDYFERFRKLKNIKVLGQ
MKKI (SEQ ID NO: 29)

[0065]

[0066] 본원에 사용된 용어 "유효량"은 목적하는 생물학적 반응을 도출해내기 위해 충분한 생물학적 활성제의 양을 지칭한다. 예를 들어, 일부 실시양태에서, 핵염기 편집제의 유효량은 핵염기 편집제에 의해 특이적으로 결합되는 표적 부위의 돌연변이를 유도하기에 충분한 핵염기 편집제의 양을 지칭할 수 있다. 일부 실시양태에서, 본원에 제공된 융합 단백질, 예를 들어 핵산 프로그램가능 DNA 결합 단백질 및 데아미나제 도메인 (예를 들어, 시티딘 데아미나제 도메인)을 포함하는 융합 단백질의 유효량은 융합 단백질에 의해 특이적으로 결합되고 편집되는 표적 부위의 편집을 유도하기에 충분한 융합 단백질의 양을 지칭할 수 있다. 통상의 기술자에 의해 인지될 바와 같이, 작용제, 예를 들어 융합 단백질, 핵염기 편집제, 데아미나제, 하이브리드 단백질, 단백질 이량체, 단백질 (또는 단백질 이량체) 및 폴리뉴클레오티드의 복합체, 또는 폴리뉴클레오티드의 유효량은, 예를 들어 목적하는 생물학적 반응, 예를 들어 편집될 특이적 대립유전자, 게놈 또는 표적 부위, 표적화될 세포 또는 조직, 및 사용될 작용제와 같은 다양한 인자에 좌우될 수 있다.

[0067]

본원에 사용된 용어 "핵산" 및 "핵산 분자"는 핵염기 및 산성 모이어티, 예를 들어 뉴클레오시드, 뉴클레오티드, 또는 뉴클레오티드의 중합체를 포함하는 화합물을 지칭한다. 전형적으로, 중합체 핵산, 예를 들어 3개 이상의 뉴클레오티드를 포함하는 핵산 분자는 선형 분자이며, 여기서 인접 뉴클레오티드는 포스포디에스테르 연결을 통해 서로 연결된다. 일부 실시양태에서, "핵산"은 개별 핵산 잔기 (예를 들어 뉴클레오티드 및/또는 뉴클레오시드)를 지칭한다. 일부 실시양태에서, "핵산"은 3개 이상의 개별 뉴클레오티드 잔기를 포함하는 올리고뉴클레오티드 쇄를 지칭한다. 본원에 사용된 용어 "올리고뉴클레오티드" 및 "폴리뉴클레오티드"는 뉴클레오티드의 중합체 (예를 들어, 일련의 적어도 3개의 뉴클레오티드)를 지칭하는 것으로 상호교환가능하게 사용될 수 있다. 일부 실시양태에서, "핵산"은 RNA 뿐만 아니라 단일 및/또는 이중-가닥 DNA를 포괄한다. 핵산은, 예를 들어 게놈, 전사체, mRNA, tRNA, rRNA, siRNA, snRNA, 플라스미드, 코스미드, 염색체, 크로마티드 또는 다른 자연 발생 핵산 분자와 관련하여 자연 발생될 수 있다. 다른 한편으로, 핵산 분자는 비-자연 발생 분자, 예를 들어 재조합 DNA 또는 RNA, 인공 염색체, 조작된 게놈, 또는 그의 단편, 또는 합성 DNA, RNA, DNA/RNA 하이브리드이거나, 또는 비-자연 발생 뉴클레오티드 또는 뉴클레오시드를 포함할 수 있다. 추가로, 용어 "핵산", "DNA", "RNA", 및/또는 유사한 용어는 핵산 유사체, 예를 들어 포스포디에스테르 백본 이외의 것을 갖는 유사체를 포함한다. 핵산은 천연 공급원으로부터 정제되고, 재조합 발현 시스템을 사용하여 생성되고, 임의로 정제되

고, 화학적으로 합성 등이 될 수 있다. 적절한 경우, 예를 들어 화학적으로 합성된 분자의 경우, 핵산은 뉴클레오타이드 유사체, 예컨대 화학적으로 변형된 염기 또는 당, 및 백본 변형을 갖는 유사체를 포함할 수 있다. 핵산 서열은 달리 나타내지 않는 한 5'에서 3'로의 방향으로 제시된다. 일부 실시양태에서, 핵산은 천연 뉴클레오타이드 (예를 들어 아데노신, 티미딘, 구아노신, 시티딘, 우리딘, 데옥시아데노신, 데옥시티미딘, 데옥시구아노신 및 데옥시시티딘); 뉴클레오타이드 유사체 (예를 들어, 2-아미노아데노신, 2-티오티미딘, 이노신, 피콜로-피리미딘, 3-메틸 아데노신, 5-메틸시티딘, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-아이오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-시티딘, C5-메틸시티딘, 2-아미노아데노신, 7-테아자아데노신, 7-테아자구아노신, 8-옥소아데노신, 8-옥소구아노신, O(6)-메틸구아닌 및 2-티오시티딘); 화학적으로 변형된 염기; 생물학적으로 변형된 염기 (예를 들어, 메틸화 염기); 삽입된 염기; 변형된 당 (예를 들어, 2'-플루오로리보스, 리보스, 2'-데옥시리보스, 아라비노스 및 핵소스); 및/또는 변형된 포스페이트 기 (예를 들어, 포스포로티오에이트 및 5'-N-포스포아미다이트 연결)이거나 또는 이를 포함한다.

[0068] 본원에 사용된 용어 "증식성 질환"은 세포 또는 세포 집단이 비정상적으로 상승된 증식률을 나타낸다는 점에서 세포 또는 조직 항상성이 교란된 임의의 질환을 지칭한다. 증식성 질환은 과다증식성 질환, 예컨대 전-신생물성 과다형성 상태 및 신생물성 질환을 포함한다. 신생물성 질환은 세포의 비정상적 증식을 특징으로 하며, 양성 및 악성 신생물 둘 다를 포함한다. 악성 신생물은 또한 암으로 지칭된다.

[0069] 용어 "단백질", "펩티드" 및 "폴리펩티드"는 본원에서 상호교환가능하게 사용되며, 펩티드 (아미드) 결합에 의해 함께 연결된 아미노산 잔기의 중합체를 지칭한다. 상기 용어는 임의의 크기, 구조 또는 기능의 단백질, 펩티드 또는 폴리펩티드를 지칭한다. 전형적으로, 단백질, 펩티드 또는 폴리펩티드는 적어도 3개의 아미노산 길이일 것이다. 단백질, 펩티드 또는 폴리펩티드는 개별 단백질 또는 단백질의 집합을 지칭할 수 있다. 단백질, 펩티드 또는 폴리펩티드에서의 아미노산 중 1개 이상은, 예를 들어 화학 물질, 예컨대 탄수화물 기, 히드록실 기, 포스페이트 기, 파르네실 기, 이소파르네실 기, 지방산 기, 접합용 링커의 첨가, 관능화, 또는 다른 변형 등에 의해 변형될 수 있다. 단백질, 펩티드 또는 폴리펩티드는 또한 단일 분자일 수 있거나 또는 다중-분자 복합체일 수 있다. 단백질, 펩티드 또는 폴리펩티드는 단지 자연 발생 단백질 또는 펩티드의 단편일 수 있다. 단백질, 펩티드 또는 폴리펩티드는 자연 발생, 재조합 또는 합성, 또는 그의 임의의 조합일 수 있다. 본원에 사용된 용어 "융합 단백질"은 적어도 2종의 상이한 단백질로부터의 단백질 도메인을 포함하는 하이브리드 폴리펩티드를 지칭한다. 1종의 단백질은 융합 단백질의 아미노-말단 (N-말단) 부분에 또는 카르복시-말단 (C-말단) 단백질에 위치하며, 따라서 각각 "아미노-말단 융합 단백질" 또는 "카르복시-말단 융합 단백질"을 형성할 수 있다. 단백질은 핵산 편집 단백질의 상이한 도메인, 예를 들어 핵산 결합 도메인 (예를 들어, 표적 부위에 대한 단백질의 결합을 지시하는 Cas9의 gRNA 결합 도메인) 및 핵산 절단 도메인 또는 촉매 도메인을 포함할 수 있다. 일부 실시양태에서, 단백질은 단백질성 부분, 예를 들어 핵산 결합 도메인을 구성하는 아미노산 서열, 및 유기 화합물, 예를 들어 핵산 절단제로서 작용할 수 있는 화합물을 포함한다. 일부 실시양태에서, 단백질은 핵산, 예를 들어 RNA와 복합체로 존재하거나 또는 그와 회합된다. 본원에 제공된 단백질 중 임의의 것은 관련 기술분야에 공지된 임의의 방법에 의해 생성될 수 있다. 예를 들어, 본원에 제공된 단백질은 펩티드 링커를 포함하는 융합 단백질에 특별히 적합한 재조합 단백질 발현 및 정제를 통해 생성될 수 있다. 재조합 단백질 발현 및 정제를 위한 방법은 널리 공지되어 있고 문헌 [Green and Sambrook, Molecular Cloning: A Laboratory Manual (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))] (이의 전체 내용은 본원에 참조로 포함됨)에 기재된 것을 포함한다.

[0070] 용어 "RNA-프로그래밍 가능 뉴클레아제" 및 "RNA-가이드 뉴클레아제"는 본원에서 상호교환가능하게 사용되며, 절단을 위한 표적이 아닌 1개 이상의 RNA(들)와 복합체를 형성하는 (예를 들어, 그와 결합하거나 회합하는) 뉴클레아제를 지칭한다. 일부 실시양태에서, RNA-프로그래밍 가능 뉴클레아제는, RNA와의 복합체로 존재하는 경우, 뉴클레아제:RNA 복합체로 지칭될 수 있다. 전형적으로, 결합된 RNA(들)는 가이드 RNA (gRNA)로 지칭된다. gRNA는 2개 이상의 RNA의 복합체로서, 또는 단일 RNA 분자로서 존재할 수 있다. 단일 RNA 분자로서 존재하는 gRNA는 단일-가이드 RNA (sgRNA)로 지칭될 수 있으며, "gRNA"는 단일 분자로서 또는 2개 이상의 분자의 복합체로서 존재하는 가이드 RNA를 지칭하는데 상호교환가능하게 사용된다. 전형적으로, 단일 RNA 종으로서 존재하는 gRNA는 다음 2개의 도메인을 포함한다: (1) 표적 핵산에 대한 상동성을 공유하는 (예를 들어, 그리고 표적에 대한 Cas9 복합체의 결합을 지시하는) 도메인; 및 (2) Cas9 단백질에 결합하는 도메인. 일부 실시양태에서, 도메인 (2)는 tracrRNA로 공지된 서열에 상응하며, 스템-루프 구조를 포함한다. 예를 들어, 일부 실시양태에서, 도메인 (2)는 문헌 [Jinek et al., Science 337:816-821(2012)] (이의 전체 내용은 본원에 참조로 포함됨)에 제공된 바와 같은 tracrRNA와 동일하거나 상동이다. gRNA의 다른 예 (예를 들어, 도메인 2를 포함하는 것)는 2013년 9월 6

일에 출원된 발명의 명칭 "스위칭가능한 Cas9 뉴클레아제 및 그의 용도"인 미국 특허 가출원 U.S.S.N. 61/874,682, 및 2013년 9월 6일에 출원된 발명의 명칭 "기능적 뉴클레아제를 위한 전달 시스템"인 미국 특허 가출원 U.S.S.N. 61/874,746에서 발견될 수 있으며, 각각의 전체 내용은 그 전문이 본원에 참조로 포함된다. 일부 실시양태에서, gRNA는 도메인 (1) 및 (2) 중 2개 이상을 포함하며, "연장된 gRNA"로 지칭될 수 있다. 예를 들어, 연장된 gRNA는, 예를 들어 본원에 기재된 바와 같이 2종 이상의 Cas9 단백질에 결합하고, 2개 이상의 별개의 영역에서 표적 핵산에 결합할 것이다. gRNA는 표적 부위를 보완하는 뉴클레오티드 서열을 포함하며, 이는 상기 표적 부위에 대한 뉴클레아제/RNA 복합체의 결합을 매개하여, 뉴클레아제:RNA 복합체의 서열 특이성을 제공한다. 일부 실시양태에서, RNA-프로그램가능 뉴클레아제는 (CRISPR-연관 시스템) Cas9 엔도뉴클레아제, 예를 들어 스트렙토코쿠스 피오게네스로부터의 Cas9 (Csn1)이다 (예를 들어, 문헌 ["Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); 및 "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012)] 참조, 상기 문헌 각각의 전체 내용은 본원에 참조로 포함됨).

[0071] RNA-프로그램가능 뉴클레아제 (예를 들어, Cas9)는 DNA 절단 부위를 표적화하기 위해 RNA:DNA 혼성화를 사용하기 때문에, 이들 단백질은 원칙적으로 가이드 RNA에 의해 특정된 임의의 서열에 표적화될 수 있다. 부위-특이적 절단 (예를 들어, 게놈을 변형시키기 위함)을 위해 RNA-프로그램가능 뉴클레아제, 예컨대 Cas9를 사용하는 방법은 관련 기술분야에 공지되어 있다 (예를 들어, 문헌 [Cong, L. et al., Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali, P. et al., RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang, W.Y. et al., Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek, M. et al., RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo, J.E. et al., Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic acids research (2013); Jiang, W. et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239 (2013)] 참조, 상기 문헌 각각의 전체 내용은 본원에 참조로 포함됨).

[0072] 본원에 사용된 용어 "대상체"는 개별 유기체, 예를 들어 개별 포유동물을 지칭한다. 일부 실시양태에서, 대상체는 인간이다. 일부 실시양태에서, 대상체는 비-인간 포유동물이다. 일부 실시양태에서, 대상체는 비-인간 영장류이다. 일부 실시양태에서, 대상체는 설치류이다. 일부 실시양태에서, 대상체는 양, 염소, 소, 고양이 또는 개이다. 일부 실시양태에서, 대상체는 척추동물, 양서류, 파충류, 어류, 곤충, 파리 또는 선충류이다. 일부 실시양태에서, 대상체는 연구용 동물이다. 일부 실시양태에서, 대상체는 유전자 조작되며, 예를 들어 유전자 조작된 비-인간 대상체이다. 대상체는 어느 한 성별일 수 있고 임의의 발달 단계에 있을 수 있다.

[0073] 용어 "표적 부위"는 염기 편집제, 예컨대 시티딘 데아미나제를 포함하는 융합 단백질 (예를 들어, 본원에 제공된 dCas9-시티딘 데아미나제 융합 단백질)에 의해 변형되는 핵산 분자 내의 서열을 지칭한다.

[0074] 용어 "치료", "치료하다" 및 "치료하는"은 본원에 기재된 바와 같은 질환 또는 장애, 또는 그의 1종 이상의 증상을 역전시키거나, 완화시키거나, 그의 발병을 지연시키거나 또는 그의 진행을 억제하는 것을 목표로 하는 임상 개입을 지칭한다. 본원에 사용된 용어 "치료", "치료하다" 및 "치료하는"은 본원에 기재된 바와 같은 질환 또는 장애, 또는 그의 1종 이상의 증상을 역전시키거나, 완화시키거나, 그의 발병을 지연시키거나 또는 그의 진행을 억제하는 것을 목표로 하는 임상 개입을 지칭한다. 일부 실시양태에서, 치료는 1종 이상의 증상이 발생하였고/거나 질환이 진단된 후에 투여될 수 있다. 다른 실시양태에서, 치료는 증상의 부재 하에서, 예를 들어 증상의 발병을 예방 또는 지연시키거나, 또는 질환의 발병 또는 진행을 억제하기 위해 투여될 수 있다. 예를 들어, 치료는 증상의 발병 전에 (예를 들어, 증상의 병력에 비추어 및/또는 유전적 또는 다른 감수성 인자에 비추어) 걸리기 쉬운 개체에게 투여될 수 있다. 치료는 또한, 예를 들어 그의 재발을 방지 또는 지연시키기 위해, 증상이 해소된 후에 계속될 수 있다.

[0075] 단백질 또는 핵산과 관련하여 본원에 사용된 용어 "재조합"은 자연에서 발생하는 것이 아니라 인간 조작의 생성물인 단백질 또는 핵산을 지칭한다. 예를 들어, 일부 실시양태에서, 재조합 단백질 또는 핵산 분자는 임의의

자연 발생 서열과 비교하여 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개 또는 적어도 7개의 돌연변이를 포함하는 아미노산 또는 뉴클레오타이드 서열을 포함한다.

[0076] 발명의 상세한 설명

[0077] 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)

[0078] 본 개시내용의 일부 측면은 단백질, 예컨대 염기 편집제를 특정 핵산 (예를 들어, DNA 또는 RNA) 서열에 가이드 하는데 사용될 수 있는 핵산 프로그램가능 DNA 결합 단백질을 제공한다. 핵산 프로그램가능 DNA 결합 단백질은, 비제한적으로, Cas9 (예를 들어, dCas9 및 nCas9), CasX, CasY, Cpf1, C2c1, C2c2, C2C3, 및 아르코 노트를 포함한다. Cas9와 상이한 PAM 특이성을 갖는 핵산 프로그램가능 DNA 결합 단백질의 한 예는 프레보텔라 및 프란시셀라 1로부터의 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (Cpf1)이다. Cas9와 유사하게, Cpf1 도 또한 부류 2 CRISPR 이펙터이다. Cpf1은 Cas9와는 다른 특색을 가지면서 강건한 DNA 간섭을 매개한다. Cpf1은 tracrRNA가 결합된 단일 RNA-가이드 엔도뉴클레아제이고, T-풍부 프로토스페이서-인접 모티프 (TTN, TTTN 또는 YTN)를 이용한다. 더욱이, Cpf1은 엇갈린 DNA 이중-가닥을 통해 DNA를 절단한다. 16종의 Cpf1-패밀리 단백질 중에서, 아시다미노코쿠스(*Acidaminococcus*) 및 라크노스피라세아에(*Lachnospiraceae*)로부터의 2종의 효소는 인간 세포에서의 효율적인 게놈-편집 활성을 갖는 것으로 밝혀져 있다. Cpf1 단백질은 관련 기술분야에 공지되어 있고, 이전에, 예를 들어 문헌 [Yamano et al., "Crystal structure of Cpf1 in complex with guide RNA and target DNA," Cell (165) 2016, p. 949-962] (이의 전체 내용은 본원에 참조로 포함됨)에 기재되었다.

[0079] 가이드 뉴클레오타이드 서열-프로그램가능 DNA 결합 단백질 도메인으로서 사용될 수 있는 뉴클레아제-불활성 Cpf1 (dCpf1) 변이체가 또한 본 발명의 조성물 및 방법에 유용하다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사한 RuvC-유사 엔도뉴클레아제 도메인을 갖지만, HNH 엔도뉴클레아제 도메인을 갖지 않고, Cpf1의 N-말단은 Cas9의 알파-나선형 인식 엽을 갖지 않는다. Cpf1의 RuvC-유사 도메인은 두 DNA 가닥의 절단을 담당하고 RuvC-유사 도메인의 불활성화는 Cpf1 뉴클레아제 활성을 불활성화시킨다는 것이 문헌 [Zetsche et al., Cell, 163, 759-771, 2015] (이는 본원에 참조로 포함됨)에 제시되었다. 예를 들어, 프란시셀라 노비시다(*Francisella novicida*) Cpf1 (서열식별번호: 30)에서의 D917A, E1006A 또는 D1255A에 상응하는 돌연변이는 Cpf1 뉴클레아제 활성을 불활성화시킨다. 일부 실시양태에서, 본 개시내용의 dCpf1은 서열식별번호: 30에서의 D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A 또는 D917A/E1006A/D1255A에 상응하는 돌연변이, 또는 또 다른 Cpf1에서의 상응하는 돌연변이(들)를 포함한다. Cpf1의 RuvC 도메인을 불활성화시키는 임의의 돌연변이, 예를 들어 치환 돌연변이, 결실 또는 삽입이 본 개시내용에 따라 사용될 수 있다는 것이 이해될 것이다.

[0080] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)은 Cpf1 단백질일 수 있다. 일부 실시양태에서, Cpf1 단백질은 Cpf1 니카제 (nCpf1)이다. 일부 실시양태에서, Cpf1 단백질은 뉴클레아제 불활성 Cpf1 (dCpf1)이다. 일부 실시양태에서, Cpf1, nCpf1 또는 dCpf1은 서열 식별번호: 30-37 중 어느 하나에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, dCpf1은 서열식별번호: 30-37 중 어느 하나에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함하고, 서열식별번호: 30에서의 D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A 및/또는 D917A/E1006A/D1255A에 상응하는 돌연변이, 또는 또 다른 Cpf1에서의 상응하는 돌연변이(들)를 포함한다. 일부 실시양태에서, dCpf1은 서열식별번호: 30-37 중 어느 하나의 아미노산 서열을 포함한다. 다른 박테리아 종으로부터의 Cpf1도 또한 본 개시내용에 따라 사용될 수 있다는 것이 인지되어야 한다.

[0081] 야생형 프란시셀라 노비시다 Cpf1 (서열식별번호: 30) (D917, E1006 및 D1255는 볼드체이고 밑줄표시되어 있다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK
 AKQIIDKYHQFFIEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQ
 ISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
 IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIHYRIVDDNLPKFLENKAKYESLKDKAP
 EAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNYYLNQSGITKFNTIIG
 GKRVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
 DSDVVTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
 SQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAL
 EFNKHRDIDKQCRFEELANFAAIPMIFDEIAQNKDNLAQISIKYQNGKKDLLQASA
 EDDVKAIKDLLDQTNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVP
 LYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVNM
 KKNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIR
 NHSTHTKNGSPQGYEKFENIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
 DEFYREVENQGYKLTFFENISESYIDSVVNQGLYLFQIYNKDFSAYSKGRPNLHTLY
 WKALFDERNLQDVVYKLNGEAEIFYRKQSIPKKITHPAKEAIANKNDNPKKESVFE
 YDLIKDKRFTEDKFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHL
 AYYTLVDGKGNIQKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLN
 YLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIYYVPAGFTSKICPVTGFVN
 QLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSR
 LINFRNSDKNHNWDTREYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLT
 SVLNILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGL
 KGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 30)

[0082]

[0083] 프란시셀라 노비시다 Cpf1 D917A (서열식별번호: 31) (A917, E1006 및 D1255는 볼드체이고 밑줄표시되어 있다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK
 AKQIIDKYHQFFIEILSSVCISEDLLQNYSDVYFKLKSSDDNLQKDFKSAKDTIKKQ
 ISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
 IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAP
 EAINYEQIKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNNYLNQSGITKFNTIIG
 GKRVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
 DSDVVTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
 SQQVFDDYSVIGTAVLEYITTQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAL
 EFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASA
 EDDVKAIKDLLDQTNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVP
 LYNKIRNYITQKPYSDKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVNM
 KKNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIR
 NHSTHTKNGSPQKGYEKFENIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
 DEFYREVENQGYKLTFFENISESYIDSVVNQGLYLFQIYNKDFSAYSIGRPNLHTLY
 WKALFDERNLQDVVYKLNGEAEFYRKQSIPKKITHPAKEAIANKNDNPKKESVFE
 YDLIKDKRFTEDKFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHL
 AYYTLVDGKGNIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLN
 YLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVN
 QLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSR
 LINFRNSDKNHNWDTREVVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLT
 SVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGL
 KGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 31)

[0084]

[0085]

프란시셀라 노비시다 Cpf1 E1006A (서열식별번호: 32) (D917, A1006 및 D1255는 볼드체이고 밑줄표시되어 있다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK
 AKQIIDKYHQFFIEILSSVCISEDLLQNYSDVYFKLKKSDDDLQKDFKSAKDTIKKQ
 ISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
 IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAP
 EAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIG
 GK FVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
 DSDVVTMQSFYEQIAAFKTVEEKSiketLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
 SQQVFDDYSVIGTAVLEYITQQAIPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAL
 EFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASA
 EDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVP
 LYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYVLGVMN
 KKNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIR
 NHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
 DEFYREVENQGYKLT FENISESYIDSVVNQGLYLFQIYNKDFSAYS KGRPNLHTLY
 WKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIAANKNDNPKKESVFE
 YDLIKDKRFTEDKFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHL
 AYYTLVDGKGNIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKL VIEYNAIVVF^ADLNFGFKRGRFKVEKQVYQKLEKMLIEKL
 NYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFV
 NQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGS
 RLINFRNSDKNHNWDTREVPYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAK
 LTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA^DANGAYHIG
 LKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 32)

[0086]

[0087] 프란시셀라 노비시다 Cpf1 D1255A (서열식별번호: 33) (D917, E1006 및 A1255는 볼드체이고 밑줄표시되어 있다)

MSIQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK
 AKQIIDKYHQFFIEILSSVCISEDLLQNYSDVYFKLKSSDDNLQKDFKSAKDTIKKQ
 ISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
 IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAP
 EAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNYYLNQSGITKFNTIIG
 GKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
 DSDVVTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
 SQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAL
 EFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASA
 EDDVKAIKDLLDQTNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVP
 LYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYLGVMN
 KKNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIR
 NHSTHTKNGSPQKGYEKFENIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
 DEFYREVENQGYKLTFFENISESYIDSVVNQGLYLFQIYNKDFSAYS KGRPNLHTLY
 WKALFDERNLQDVVYKLNGEAEIFYRKQSIPKKITHPAKEAIANKNDNPKKESVFE
 YDLIKDKRFTEDKFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHL
 AYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEIYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLN
 YLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIYYVPAGFTSKICPVTGFVN
 QLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSR
 LINFRNSDKNHNWDTREVVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLT
 SVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGL
 KGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 33)

[0088]

[0089] 프란시셀라 노비시다 Cpf1 D917A/E1006A (서열식별번호: 34) (A917, A1006 및 D1255는 볼드체이고 밑줄표시되어 있다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK
 AKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQ
 ISEYIKDSEKFNLFNQNLDIAKKQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
 IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAP
 EAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIG
 GKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
 DSDVVTMQSFYEQIAAFKTVEEKSiketLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
 SQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAL
 EFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASA
 EDDVKAIKDLLDQTNLLHLKLIKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVP
 LYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYLGVMN
 KKNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFPNPSDILRIR
 NHSTHTKNGSPQKGYEKFENIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
 DEFYREVENQGYKLTfENISESYIDSvVNQGKLYLFQIYNKDFSAYSKGRPNLHTLY
 WKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIAANKNDNPKKESVFE
 YDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHL
 AYYTLVDGKGNIQKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKL
 NYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIYYVPAGFTSKICPVTGFV
 NQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGS
 RLINFRNSDKNHNWDTREYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAK
 LTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIG
 LKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 34)

[0090]

[0091]

프란시셀라 노비시다 Cpf1 D917A/D1255A (서열식별번호: 35) (A917, E1006 및 A1255는 볼드체이고 밑줄표시되어 있다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK
 AKQIIDKYHQFFIEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQ
 ISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
 IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAP
 EAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNYYLNQSGITKFNTIIG
 GKFVNGENTKRKGINEYINLYSQQINDKTLKYYKMSVLFKQILSDTESKSFVIDKLED
 DSDVVTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
 SQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAL
 EFNKHRDIDKQCRFEELANFAAIMFDEIAQNKDNLAQISIKYQNQGKKDLLQASA
 EDDVKAIKDLLDQTNLLHLKLIKHFISQSEDKANILDKDEHFYLVFEECYFELANIVP
 LYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYLGVMN
 KKNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFNPSIEDILRIR
 NHSTHTKNGSPQKGYEKFENIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
 DEFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLY
 WKALFDERNLQDVVYKLNGEAEFYRKQSIPKKITHPAKEAIANKNDNPKKESVFE
 YDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHL
 AYYTLVDGKGNIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVIENAIIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLN
 YLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVN
 QLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGWTIASFGSR
 LINFRNSDKNHNWDTREVVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLT
 SVLNILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGL
 KGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 35)

[0092]

[0093]

프란시셀라 노비시다 Cpf1 E1006A/D1255A (서열식별번호: 36) (D917, A1006 및 A1255는 볼드체이고 밑줄표시되어 있다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK
 AKQIIDKYHQFFIEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQ
 ISEYIKDSEKFKNLFNQNLDIAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
 IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAP
 EAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLEVFIEIANFNYYLNQSGITKFNTIIG
 GKFTVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
 DSDVVTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
 SQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAL
 EFNKHRDIDKQCRFEELANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASA
 EDDVKAIKDLLDQTNLLHLKLIFHISQSEDKANILDKDEHFYLVFEECYFELANIVP
 LYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYLGVMN
 KKNKIFDDKAIKENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIR
 NHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYSI
 DEFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLY
 WKALFDERNLQDVVYKLNGEAEFYRKQSIPKKITHPAKEAIANKNDNPKKESVFE
 YDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHL
 AYYTLVDGKGNIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
 MKEGYLSQVVHEIAKLVEIYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKL
 NYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFV
 NQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGWTIASFGS
 RLINFRNSDKNHNWDTREVPYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAK
 LTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAANGAYHIG
 LKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 36)

[0094]

[0095]

프란시셀라 노비시다 Cpf1 D917A/E1006A/D1255A (서열식별번호: 37) (A917, A1006 및 A1255는 볼드체이고 밑줄 표시되어 있다)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK
AKQIIDKYHQFFIEILSSVCISEDLLQNYSDVYFKLKKSDDDLQKDFKSAKDTIKKQ
ISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEI
IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFLENKAKYESLKDKAP
EAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNYYLNQSGITKFNTIIG
GKFVNGENTKRKGINEYINLYSQQINDKTLKYYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDL
SQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAL
EFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNNQKKDLLQASA
EDDVKAIKDLLDQTNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVP
LYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMM
KKNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIR
NHSTHTKNGSPQKGYEKFENIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTENISESYIDSVVNQKLYLFQIYNKDFSAYSKGRPNLHTLY
WKALFDERNLQDVVYKLNGEAEFYRKQSIPKKITHPAKEAIAANKNDNPKKESVFE
YDLIKDKRFTEDKFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI¹ARGERHL
AAYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKE
MKEGYLSQVVHEIAKLVEIYNAIVVF²ADLNFGFKRGRFKVEKQVYQKLEKMLIEKL
NYLVFKDNEFDKTTGGVLRAYQLTAPFETFKMKMGKQTGIYYVPAGFTSKICPVTGFV
NQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGS
RLINFRNSDKNHNWDTREVPYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAK
LTSVLNTILQMRNSKTGTLDYLISPVADVNGNFFDSRQAPKNMPQDA³ANGAYHIG
LKGLMLLGRIKNNQEGKKNLVIKNEEYFEFVQNRNN (SEQ ID NO: 37)

[0096]

[0097]

일부 실시양태에서, 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)은 정규 (NGG) PAM 서열을 요구하지 않는 핵산 프로그램가능 DNA 결합 단백질이다. 일부 실시양태에서, napDNAbp는 아르코노트 단백질이다. 이러한 핵산 프로그램가능 DNA 결합 단백질의 한 예는 나트로노박테리움 그레고리 (*Natronobacterium gregoryi*)로부터의 아르코노트 단백질 (NgAgo)이다. NgAgo는 ssDNA-가이드 엔도뉴클레아제이다. NgAgo는 ~24개 뉴클레오타이드의 5' 인산화 ssDNA (gDNA)에 결합하여 그것을 그의 표적 부위로 가이드하고, gDNA 부위에서 DNA 이중-가닥 파괴를 만들 것이다. Cas9와 달리, NgAgo-gDNA 시스템은 프로토스페이서-인접 모티프 (PAM)를 요구하지 않는다. 뉴클레아제 불활성 NgAgo (dNgAgo)를 사용하는 것은 표적화될 수 있는 염기를 매우 확장시킬 수 있다. NgAgo의 특징화 및 용도는 문헌 [Gao et al., Nat Biotechnol., 2016 Jul;34(7):768-73. PubMed PMID: 27136078; Swarts et al., Nature. 507(7491) (2014):258-61; 및 Swarts et al., Nucleic Acids Res. 43(10) (2015):5120-9] (이들 각각은 본원에 참조로 포함됨)에 기재되었다. 나트로노박테리움 그레고리 아르코노트의 서열은 서열식별번호: 38에 제공된다.

[0098] 야생형 나트로노박테리움 그레고리 아르코노트 (서열식별번호: 38)

MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQD
 NGERRYITLWKNTTPKDVFTYDYATGSTYIFTNIDYEVKDGYNLTATYQTTVENAT
 AQEVGTTDEDETFAAGGEPLDHLLDDALNETPDDAETESDSGHVMTSFASRDQLPEW
 TLHTYTLTATDGAKTDEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGE
 TPLDLDCGVRVEADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLS
 KEPVLTCDLHEDLHERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRV
 KTTYRPRRGHIVWGLRDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADR
 RVVETRRQGHGDDAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEK
 AQAFARLDPVRLNGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDE
 TFSKGIVNPPESEFEVAVVLPQQAQWDTMADLLNQAGAPPTRSETVQYDAF
 SSPESISLNVAGAIDPSEVDAAFVVLPPDQEGFADLASPTETDELKALANMGIYSQ
 MAYFDRFRDAKIFYTRNVALGLLAAAGGVAFTEHAMPGDADMFIGIDVSRYPED
 GASGQINIAATATAVYKDGTLGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGE
 SPTHIVHRDGMNEDLDPATEFLNEQGVYDIVEIRKQPQTRLLAVSDVQYDTPVKS
 IAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQVYLLSQSH
 IQVHNSTARLPITTAYADQASTHATKGYLVQTGAFESNVGFL (SEQ ID NO: 38)

[0099]

[0100] 일부 실시양태에서, napDNAbp는 아르코노트 단백질의 원핵 상동체이다. 아르코노트 단백질의 원핵 상동체는 공지되어 있으며, 예를 들어 문헌 [Makarova K., et al., "Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements", Biol Direct. 2009 Aug 25;4:29. doi: 10.1186/1745-6150-4-29] (이의 전체 내용은 본원에 참조로 포함됨)에 기재되었다. 일부 실시양태에서, napDNAbp는 마리니토가 피에조필라(*Marinitoga piezophila*) 아르코노트 (MpAgo) 단백질이다. CRISPR-연관 마리니토가 피에조필라 아르코노트 (MpAgo) 단백질은 5'-인산화 가이드를 사용하여 단일-가닥 표적 서열을 절단한다. 5' 가이드는 모든 공지된 아르코노트에 의해 사용된다. MpAgo-RNA 복합체의 결정 구조는 5' 포스페이트 상호작용을 차단하는 잔기를 포함하는 가이드 가닥 결합 부위를 보여준다. 이 데이터는 5'-히드록실화 가이드에 대한 비정규 특이성을 갖는 아르코노트 하위부류의 진화를 시사한다. 예를 들어, 문헌 [Kaya et al., "A bacterial Argonaute with noncanonical guide RNA specificity", Proc Natl Acad Sci U S A. 2016 Apr 12;113(15):4057-62] (이의 전체 내용은 본원에 참조로 포함됨)을 참조한다. 다른 아르코노트 단백질이 사용될 수 있으며 본 개시내용의 범주 내에 있다는 것이 인지되어야 한다.

[0101]

일부 실시양태에서, 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)은 미생물 CRISPR-Cas 시스템의 단일 이펙터이다. 미생물 CRISPR-Cas 시스템의 단일 이펙터는, 비제한적으로, Cas9, Cpf1, C2c1, C2c2 및 C2c3을 포함한다. 전형적으로, 미생물 CRISPR-Cas 시스템은 부류 1 및 부류 2 시스템으로 나누어진다. 부류 1 시스템은 멀티서브유닛 이펙터 복합체를 갖고, 반면에 부류 2 시스템은 단일 단백질 이펙터를 갖는다. 예를 들어, Cas9 및 Cpf1은 부류 2 이펙터이다. Cas9 및 Cpf1 이외에, 3종의 특유한 부류 2 CRISPR-Cas 시스템 (C2c1, C2c2 및 C2c3)이 문헌 [Shmakov et al., "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", Mol. Cell, 2015 Nov 5; 60(3): 385-397] (이의 전체 내용은 본원에 참조로 포함됨)에 기재되었다. 상기 시스템 중 2종인 C2c1 및 C2c3의 이펙터는 Cpf1과 관련된 RuvC-유사 엔도뉴클레아제 도메인을 함유한다. 제3 시스템인 C2c2는 2개의 예측된 HEPN RNase 도메인을 갖는 이펙터를 함유한다. 성숙 CRISPR RNA의 생산은 C2c1에 의한 CRISPR RNA의 생산과 달리 tracrRNA-비의존적이다. C2c1은 DNA 절단을 위해 CRISPR RNA 및 tracrRNA 둘 다에 의존한다. 박테리아 C2c2는 그의 RNA-활성화된 단일-가닥 RNA 분해 활성과는 다른 CRISPR RNA 성숙을 위한 고유한 RNase 활성을 보유하는 것으로 밝혀졌다. 이들 RNase 기능은 서로 및 Cpf1의 CRISPR RNA-프로세싱 거동과 상이하다. 예를 들어, 문헌 [East-Seletsky, et al., "Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection", Nature, 2016 Oct 13;538(7624):270-273] (이의 전체 내용은 본원에 참조로 포함됨)을 참조한다. 렙토트리키아 샤히이(*Leptotrichia shahii*)의 C2c2의 시험관내 생화학적 분석은 C2c2가 단일 CRISPR RNA에 의해 가이드되고 상보적 프로토스페이서를 보유하는 ssRNA 표적을 절단하도록 프로그램화될 수 있다는 것을 제시하였다. 2개의 보존된 HEPN 도메인 내의 촉매 잔기는 절단을 매개한다. 촉매 잔기 내의 돌연변이는 촉매 불활성 RNA-결합 단백질을 생성한다. 예를 들어, 문헌

[Abudayyeh et al., "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector", Science, 2016 Aug 5; 353(6299)] (이의 전체 내용은 본원에 참조로 포함됨)을 참조한다.

[0102] 알리시클로바실루스 아시도테레스트리스(*Alicyclobacillus acidoterrestris*) C2c1 (AacC2c1)의 결정 구조는 키메라 단일-분자 가이드 RNA (sgRNA)와의 복합체로 보고되었다. 예를 들어, 문헌 [Liu et al., "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", Mol. Cell, 2017 Jan 19;65(2):310-322] (이의 전체 내용은 본원에 참조로 포함됨)을 참조한다. 결정 구조는 또한 3원 복합체로서 표적 DNA에 결합된 알리시클로바실루스 아시도테레스트리스 C2c1에서 보고되었다. 예를 들어, 문헌 [Yang et al., "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", Cell, 2016 Dec 15;167(7):1814-1828] (이의 전체 내용은 본원에 참조로 포함됨)을 참조한다. AacC2c1의 촉매 적격 입체형태는 표적 및 비-표적 DNA 가닥 둘 다와 함께, 단일 RuvC 촉매 포켓 내에 독립적으로 위치한 채로 포획되었으며, 이때 C2c1-매개 절단은 표적 DNA의 엇갈린 7개-뉴클레오타이드 파괴를 일으켰다. C2c1 3원 복합체와 이전에 확인된 Cas9 및 Cpf1 대응부 사이의 구조적 비교는 CRISPR-Cas9 시스템에 의해 사용된 메커니즘의 다양성을 입증한다.

[0103] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)은 C2c1, C2c2 또는 C2c3 단백질일 수 있다. 일부 실시양태에서, napDNAbp는 C2c1 단백질이다. 일부 실시양태에서, napDNAbp는 C2c2 단백질이다. 일부 실시양태에서, napDNAbp는 C2c3 단백질이다. 일부 실시양태에서, napDNAbp는 자연 발생 C2c1, C2c2 또는 C2c3 단백질에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, napDNAbp는 자연 발생 C2c1, C2c2 또는 C2c3 단백질이다. 일부 실시양태에서, napDNAbp는 서열식별번호: 39-40 중 어느 하나에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 박테리아 종으로부터의 C2c1, C2c2 또는 C2c3이 또한 본 개시내용에 따라 사용될 수 있다는 것이 인지되어야 한다.

[0104] C2c1 (uniprot.org/uniprot/T0D7A2#)

[0105] sp|T0D7A2|C2C1_ALIAG CRISPR-연관 엔도뉴클레아제 C2c1 OS=알리시클로바실루스 아시도테레스트리스 (균주 ATCC 49025 / DSM 3922 / CIP 106132 / NCIMB 13137 / GD3B) GN=c2c1 PE=1 SV=1

MAVKSIVKLRLLDDMPDIRAGLWKLHKEVNAVRYYTEWLSLLRQENLYRRSPNG
DGEQECDKTAECKAELLRLRARQVENGHRGPAGSDELLQLARQLYELLVPQAI
GAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEKEKA
ETRKSADRTADVLRLALDFGLKPLMRVYTDSEMSVVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQVRVGQYAKLVEQKNRFEQKNFVGQEHVLVHLVNQLQQDM
KEASPGLESKEQTAHYVTGRALRGSDKVFKEWGLAPDAPFDLYDAEIKNVQRRNT
RRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDAT
AHPWTRFDKLGGLNHQYTFLEFGERRHAIKHLKLVENGVAEVDVTPISM
SEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRDQLAHMHRRRGARD
VYLVNSVVRVQSQSEARGERRPPYAAVFRVLVDNHRFAVHFDFKLSDYLAHPDDGKL
GSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNKGRVPFFFPKGNLNLVAV
HERSQLKLPGETESKDLRAIREERQRTLRLQLRTQLAYLRLLVRCGEDVGRRRERSW
AKLIEQPVDAAHMTDPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVVRH
MGKQVRDWRKDVRSGERPKIRGYAKDVVGNSIEQIEYLERQYKFLKSWSFSGKVS
GQVIRAEKSGRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWWA
KYPPCQLILLEELSEYQFNDRPPSENNQLMQWSHRGVFQELINQAQVHDLVGTM
YAAFSSRFDARTGAPGIRCRRVPARCTQEHNPFPWWLNKFFVEHTLDACPLRAD
DLIPTGEGEIVSPFSAEEGDFHQIHADLNAAQNLQQLWSDFDISQIRLRCDWGEVD
GELVLIPRLTGKRTADSYSNKVFTNTGVTTYERERGGKRRKVFVFAQEKLSSEEAELL
VEADEAREKSVVLMRDPGSIINRGWNTRQKEFWSMVNQRIEGLYLVKQIRSRVPLQD
SACENTGDI (SEQ ID NO: 39)

[0106]

- [0107] C2c2 (uniprot.org/uniprot/P0DOC6)
- [0108] >sp|P0DOC6|C2C2_LEPSD CRISPR-연관 엔도리보뉴클레아제 C2c2 OS=렙토트리키아 샤히이 (균주 DSM 19757 / CCUG 47503 / CIP 107916 / JCM 16776 / LB37) GN=c2c2 PE=1 SV=1
- MGNLFGHKRWYEVDRDKKDFKIKRKVKVKRNYDGNKYILNINENNKEKIDNNKFIR
KYINYKKNDNLKEFTRKFHAGNLFKLKGKEGIIRIENNDDFLETEEVVLYIEAYGKS
EKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEIEIDIRDEYTNKTLNDCSILRI
IENDELETKKSIYEIFKNINMSLYKIIIEKIENETEKVFENRYEEHLREKLLKDDKIDVI
LTNFMEREKIKSNLEILGFVKFYLVNMGDKKKSKNKKMLVEKILNINVDLTVEDIAD
FVIKELEFWNITKRIEKVKKVNNEFLEKRRNRITYIKSYVLLDKHEKFKIERENKKDKI
VKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNFD
KKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKILK
RVKQYTLHEIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFASTNMELNK
IFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNITNNFIRKFTKI
GTNERNRILHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFCKDKKNII
TKINDIKISEENNNDIKYLPFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKE
LYKKLILEDLEENESKNIFLQELKKTGNGIDEIDENIENIENYKNAQISASKGNKKA
KYQKKVIECYIGYLRKNEYELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDKTI
VINDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNLRNEC
ITENWNLNLEEFIQMKKEIEKDFDDFKIQTKKEIFNNYEDIKNNILTEFKDDINGCDV
LEKKLEKIVFDDETKFEIDKKSNIQDEQRKLSNINKKDLKKKVDQYIKDKDQEIKS
KILCRIIFNSDFLKYYKKEIDNLIEDMESENENKFQEIYYPKERKNELIYKKNLFLNIG
NPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNYSKEYKE
KYIKKLKENDDDFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNLYLNKIESYLIDIN
WKLAIQMARFERDMHYIVNGLRELGIKLSGYNTGISRAYPKRNGSDGFYTTTAYYK
FFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHFYIVRNPFADYSIAEQIDRV
SNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVS
LELESYNSDYIKNLIIELLTKIENTNDTL (SEQ ID NO: 40)
- [0109]
- [0110] 핵염기 편집제의 Cas9 도메인
- [0111] 일부 측면에서, 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp)은 Cas9 도메인이다. 비제한적, 예시적인 Cas9 도메인이 본원에 제공된다. Cas9 도메인은 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 불활성 Cas9 도메인, 또는 Cas9 니카제일 수 있다. 일부 실시양태에서, Cas9 도메인은 뉴클레아제 활성 도메인이다. 예를 들어, Cas9 도메인은 듀플렉스 핵산의 양쪽 가닥 (예를 들어, 듀플렉스 DNA 분자의 양쪽 가닥)을 절단하는 Cas9 도메인일 수 있다. 일부 실시양태에서, Cas9 도메인은 서열식별번호: 4-29에 제시된 바와 같은 아미노산 서열 중 어느 하나를 포함한다. 일부 실시양태에서, Cas9 도메인은 본원에 제공된 임의의 Cas9 또는 서열식별번호: 4-29에 제시된 아미노산 서열 중 하나에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 본원에 제공된 임의의 Cas9 또는 서열식별번호: 4-29에 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, 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, 50개 또는 그 초과와 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 본원에 제공된 임의의 Cas9 또는 서열식별번호: 4-29에 제시된 아미노산 서열 중 어느 하나와 비교하여 적어도 10, 적어도 15, 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 적어도 90, 적어도 100, 적어도 150, 적어도 200, 적어도 250, 적어도 300, 적어도 350, 적어도 400, 적어도 500, 적어도 600, 적어도 700, 적어도 800, 적어도 900, 적어도 1000, 적어도 1100 또는 적어도 1200개의 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.
- [0112] 일부 실시양태에서, Cas9 도메인은 뉴클레아제-불활성 Cas9 도메인 (dCas9)이다. 예를 들어, dCas9 도메인은

듀플렉스 핵산 분자의 어느 하나의 가닥을 절단하지 않고 듀플렉스 핵산 분자에 (예를 들어, gRNA 분자를 통해) 결합할 수 있다. 일부 실시양태에서, 뉴클레아제-불활성 dCas9 도메인은 서열식별번호: 6에 제시된 아미노산 서열의 D10X 돌연변이 및 H840X 돌연변이, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 하나에서의 상응하는 돌연변이를 포함하고, 여기서 X는 임의의 아미노산 변화이다. 일부 실시양태에서, 뉴클레아제-불활성 dCas9 도메인은 서열식별번호: 6에 제시된 아미노산 서열의 D10A 돌연변이 및 H840A 돌연변이, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다. 한 예로서, 뉴클레아제-불활성 Cas9 도메인은 서열식별번호: 9에 제시된 아미노산 서열을 포함한다 (클로닝 벡터 pPlatTET-gRNA2, 수탁 번호 BAV54124).

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLI
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFL
VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDDTKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSVDKLFILVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLI
LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGP
LARGNSRFAMWTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLEYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
KSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTV
KVVDDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEELGKELGSQILKE
HPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDN
KVLTRSDKNRGKSDNVPSEEVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRRDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD
(SEQ ID NO: 9)

[0113]

[0114] (예를 들어 문헌 [Qi et al., "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." Cell. 2013; 152(5):1173-83]을 참조하며, 이의 전체 내용은 본원에 참조로 포함됨).

[0115] 추가의 적합한 뉴클레아제-불활성 dCas9 도메인은 본 개시내용 및 관련 기술분야의 지식에 기초하여 관련 기술 분야의 통상의 기술자에게 분명할 것이며, 본 개시내용의 범주 내에 있다. 이러한 추가의 예시적인 적합한 뉴클레아제-불활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A, 및 D10A/D839A/H840A/N863A 돌연변이체 도메인을 포함하나 이에 제한되지는 않는다 (예를 들어, 문헌 [Prashant et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature Biotechnology. 2013; 31(9): 833-838]을 참조하며, 이의 전체 내용은 본원에 참조로 포함됨). 일부 실시양태에서, dCas9 도메인은 본원에 제공된 dCas9 도메인 중 어느 하나에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 서열식별번호: 7, 8, 9 또는 22에 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 또는 그 초과와 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 서열식별번호: 7, 8, 9 또는 22에 제시된 아미노산 서열 중 어느 하나와 비교하여 적어도 10, 적어도 15, 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 60, 적어도 70, 적어도 80, 적어도 90, 적어도 100, 적어도 150, 적어도 200, 적어도 250, 적어도 300, 적어도 350, 적어도 400, 적어도 500, 적어도 600, 적어도 700, 적어도 800, 적어도 900, 적어도 1000, 적어도 1100 또는 적어도 1200개의 동일한 인접 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0116] 일부 실시양태에서, Cas9 도메인은 Cas9 니카제이다. Cas9 니카제는 듀플렉스 핵산 분자 (예를 들어, 듀플렉스 DNA 분자)의 단지 1개의 가닥만을 절단할 수 있는 Cas9 단백질일 수 있다. 일부 실시양태에서, Cas9 니카제는 듀플렉스 핵산 분자의 표적 가닥을 절단하며, 이는 Cas9 니카제가 Cas9에 결합된 gRNA (예를 들어, sgRNA)에 염기 쌍형성된 (그에 상보적인) 가닥을 절단함을 의미한다. 일부 실시양태에서, Cas9 니카제는 D10A 돌연변이를 포함하고, 서열식별번호: 6의 위치 840에서의 히스티딘, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26 중 어느 하나에서의 돌연변이를 갖는다. 예를 들어, Cas9 니카제는 서열식별번호: 10, 13, 16 또는 21에 제시된 바와 같은 아미노산 서열을 포함할 수 있다. 일부 실시양태에서, Cas9 니카제는 듀플렉스 핵산 분자의 비-표적, 비-염기-편집된 가닥을 절단하며, 이는 Cas9 니카제가 Cas9에 결합된 gRNA (예를 들어, sgRNA)에 염기 쌍형성되지 않은 가닥을 절단함을 의미한다. 일부 실시양태에서, Cas9 니카제는 H840A 돌연변이를 포함하고, 서열식별번호: 6의 위치 10에서의 아스파르트산 잔기, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26 중 어느 하나에서의 상응하는 돌연변이를 갖는다. 일부 실시양태에서, Cas9 니카제는 본원에 제공된 Cas9 니카제 중 어느 하나에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 추가의 적합한 Cas9 니카제는 본 개시내용 및 관련 기술분야의 지식에 기초하여 관련 기술분야의 통상의 기술자에게 분명할 것이며, 본 개시내용의 범주 내에 있다.

[0117] 감소된 PAM 배타성을 갖는 Cas9 도메인

[0118] 본 개시내용의 일부 측면은 상이한 PAM 특이성을 갖는 Cas9 도메인을 제공한다. 전형적으로, Cas9 단백질, 예컨대 에스. 피오게네스로부터의 Cas9 (spCas9)는 정규 NGG PAM 서열이 특정한 핵산 영역에 결합할 것을 요구하며, 여기서 "NGG" 내 "N"은 아데닌 (A), 티민 (T), 구아닌 (G) 또는 시토신 (C)이고, G는 구아닌이다. 이는 게놈 내에서 목적하는 염기를 편집하는 능력을 제한할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집 융합 단백질은, 예를 들어 표적 염기가 PAM의 대략 15개 염기 상류에 있는 4개의 염기 영역 (예를 들어, "탈아미노화 윈도우") 내인 정확한 위치에 놓일 필요가 있을 수 있다. 문헌 [Komor, A.C., et al., "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016)]을 참조하며, 이의 전체 내용은 본원에 참조로 포함된다. 일부 실시양태에서, 탈아미노화 윈도우는 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 염기 영역 내이다. 일부 실시양태에서, 탈아미노화 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 상류에 있다. 따라서, 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 정규 (예를 들어, NGG) PAM 서열을 함유하지 않는 뉴클레오타이드 서열에 결합할 수 있는 Cas9 도메인을 함유할 수 있다. 비-정규 PAM 서열에 결합하는 Cas9 도메인은 관련 기술분야에 기재되었으며, 통상의 기술자에게 분명할 것이다. 예를 들어, 비-정규 PAM 서열에 결합하는 Cas9 도메인은 문헌 [Kleinstiver, B. P., et al., "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" Nature 523, 481-485 (2015); 및 Kleinstiver, B. P., et al., "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition" Nature Biotechnology 33, 1293-1298 (2015)] (각각의 전체 내용은 본원에 참조로 포함됨)에 기재되었다.

[0119] 일부 실시양태에서, Cas9 도메인은 스태필로코쿠스 아우레우스로부터의 Cas9 도메인 (SaCas9)이다. 일부 실시양태에서, SaCas9 도메인은 뉴클레아제 활성 SaCas9, 뉴클레아제 불활성 SaCas9 (SaCas9d), 또는 SaCas9 니카제 (SaCas9n)이다. 일부 실시양태에서, SaCas9는 아미노산 서열 서열식별번호: 12를 포함한다. 일부 실시양태에서, SaCas9는 서열식별번호: 12의 N579X 돌연변이, 또는 서열식별번호: 13-14에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함하고, 여기서 X는 N을 제외한 임의의 아미노산이다. 일부 실시양태에서, SaCas9는 서열식별번호: 12의 N579A 돌연변이, 또는 서열식별번호: 13-14에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다.

[0120] 일부 실시양태에서, SaCas9 도메인, SaCas9d 도메인 또는 SaCas9n 도메인은 비-정규 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시양태에서, SaCas9 도메인, SaCas9d 도메인 또는 SaCas9n 도메인은 NNGRRT PAM 서열을 갖는 핵산 서열에 결합할 수 있고, 여기서 N = A, T, C 또는 G이고, R = A 또는 G이다. 일부 실시양태에서,

SaCas9 도메인은 서열식별번호: 12의 E781X, N967X 및 R1014X 돌연변이 중 1개 이상, 또는 서열식별번호: 13-14에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함하고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, SaCas9 도메인은 서열식별번호: 12의 E781K, N967K 및 R1014H 돌연변이 중 1개 이상, 또는 서열식별번호: 13-14에 제공된 아미노산 서열 중 임의의 것에서의 1개 이상의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SaCas9 도메인은 서열식별번호: 12의 E781K, N967K 또는 R1014H 돌연변이, 또는 서열식별번호: 13-14에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다.

[0121] 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 12-14 중 어느 하나에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 12-14 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 12-14 중 어느 하나의 아미노산 서열로 이루어진다.

[0122] 예시적인 SaCas9 서열

```
KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK
RGARRLKRHHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSA
ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGE
VRGSINRFKTSIDYVKEAKQLLKVKQAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPF
GWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLE
YYEKFQIENVFQKQKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI
TARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLS
LKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFI
QSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKE
NAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEVVDHIIIPRSVSFDNSFNKVL
VKQEEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAQKGRISKTKKEYLLEERDI
NRFSVQKDFINRLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKW
KFKKERNKGYKHAEDALIANADFIFKEWKLDKAKKVMENQMFEEKQAESMPEI
ETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKNRELINDTLYSTRKDDKGNTLIV
NNLNGLYDKDNDKLKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKY
YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYRF
DVYLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKKLKKISNQAEFIASFYNNND
LIKINGELYRVIGVNNDLLNRIEIVNMIDITYREYLENMNDKRPPIIKTIASKTQSIKKY
STDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 12)
```

[0123]

[0124] 밑줄표시되어 있고 볼드체인 서열식별번호: 12의 잔기 N579는 돌연변이되어 (예를 들어, A579로) SaCas9 니카제를 생성할 수 있다.

[0125] 예시적인 SaCas9n 서열

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK
 RGARRLKRRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA
 ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGE
 VRGSINRFKTSYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPF
 GWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLE
 YYEFQIENNVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI
 TARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLS
 LKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFI
 QSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKE
 NAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVL
 VKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAAGKGRISKTKKEYLLEERDI
 NRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKW
 KFKKERNKGYKHAEDALIIANADFIFKEWKLDKAKKVMENQMFEKQAESMPEI
 ETEQEYKEIFITPHQIKHIKDFKDYKYSRVDKKNRELINDTLYSTRKDDKGNTLIV
 NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDEKNPLYKY
 YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYRF
 DVYLDNGVYKFVTVKNLDVIKENYEVNSKCYEEAKKLKKISNQAEFIASFYNNND
 LIKINGELRVIGVNNDDLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKY
 STDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 13).

[0126]

[0127] 서열식별번호: 12의 N579로부터 돌연변이되어 SaCas9 니카제를 생성할 수 있는 서열식별번호: 13의 잔기 A579는 밑줄표시되어 있고 볼드체이다.

[0128] 예시적인 SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK
 RGARRLKRRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA
 ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGE
 VRGSINRFKTSYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPF
 GWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLE
 YYEFQIENNVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDI
 TARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLS
 LKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFI
 QSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKE
 NAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVL
 VKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAAGKGRISKTKKEYLLEERDI
 NRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKW
 KFKKERNKGYKHAEDALIIANADFIFKEWKLDKAKKVMENQMFEKQAESMPEI
 ETEQEYKEIFITPHQIKHIKDFKDYKYSRVDKKNRKLINDTLYSTRKDDKGNTLIV
 NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPTQYQKLKLIMEQYGDEKNPLYKY
 YEETGNYLTKYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKPYRF
 DVYLDNGVYKFVTVKNLDVIKENYEVNSKCYEEAKKLKKISNQAEFIASFYNND
 LIKINGELRVIGVNNDDLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKY
 STDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 14).

[0129]

[0130] 서열식별번호: 12의 N579로부터 돌연변이되어 SaCas9 니카제를 생성할 수 있는 서열식별번호: 14의 잔기 A579는 밑줄표시되어 있고 볼드체이다. 서열식별번호: 12의 E781, N967 및 R1014로부터 돌연변이되어 SaKKH Cas9를 생

상할 수 있는 서열식별번호: 14의 잔기 K781, K967 및 H1014는 밑줄표시되어 있고 이탤릭체이다.

[0131]

일부 실시양태에서, Cas9 도메인은 스트렙토코쿠스 피오게네스로부터의 Cas9 도메인 (SpCas9)이다. 일부 실시양태에서, SpCas9 도메인은 뉴클레아제 활성 SpCas9, 뉴클레아제 불활성 SpCas9 (SpCas9d), 또는 SpCas9 니카제 (SpCas9n)이다. 일부 실시양태에서, SpCas9는 아미노산 서열 서열식별번호: 15를 포함한다. 일부 실시양태에서, SpCas9는 서열식별번호: 15의 D9X 돌연변이, 또는 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함하고, 여기서 X는 D를 제외한 임의의 아미노산이다. 일부 실시양태에서, SpCas9는 서열식별번호: 15의 D9A 돌연변이, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인, SpCas9d 도메인 또는 SpCas9n 도메인은 비-정규 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시양태에서, SpCas9 도메인, SpCas9d 도메인 또는 SpCas9n 도메인은 NGG, NGA 또는 NGCG PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134X, R1334X 및 T1336X 돌연변이 중 1개 이상, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함하고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134E, R1334Q 및 T1336R 돌연변이, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134E, R1334Q 및 T1336R 돌연변이, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134X, R1334X 및 T1336X 돌연변이 중 1개 이상, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함하고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134V, R1334Q 및 T1336R 돌연변이 중 1개 이상, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134V, R1334Q 및 T1336R 돌연변이, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134X, G1217X, R1334X 및 T1336X 돌연변이 중 1개 이상, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함하고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134V, G1217R, R1334Q 및 T1336R 돌연변이 중 1개 이상, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, SpCas9 도메인은 서열식별번호: 15의 D1134V, G1217R, R1334Q 및 T1336R 돌연변이, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 어느 하나에서의 상응하는 돌연변이를 포함한다.

[0132]

일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 15-19 중 어느 하나에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 15-19 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것의 Cas9 도메인은 서열식별번호: 15-19 중 어느 하나의 아미노산 서열로 이루어진다.

[0133] 예시적인 SpCas9

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
 ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLEESFLV
 EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
 GHFLIEGDLNPDNSVDKLFQVLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
 NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA
 QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
 VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
 DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPYYVGPLA
 RGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHSL
 LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
 EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
 GFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
 VDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
 KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
 QFYKVRINNYHHAHDAYLNAVVGTAIIKKYKLESEFVYGDYKVYDVRKMIKSE
 QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
 KVLSPQVNVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
 SVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIKLP
 KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
 QLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVLSAYNKHRDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQSTGLYETRIDLSQLGGD (SEQ
 ID NO: 15)

[0134]

[0135] 예시적인 SpCas9n

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
 ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLEESFLV
 EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFR
 GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
 NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA
 QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
 VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
 DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPYYVGPLA
 RGNRSFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHSL
 LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
 EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
 GFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
 VDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
 KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
 QFYKVINNYHHAHDAYLNAVVGTAIIKYPKLESEFVYGDYKVYDVRKMIKSE
 QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
 KVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
 SVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIKLP
 KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
 QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD (SEQ
 ID NO: 16)

[0136]

[0137] 예시적인 SpEQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
 ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLV
 EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFR
 GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
 NLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLA
 QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
 VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
 DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLA
 RGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
 LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
 EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
 GFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
 VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEELGELGSQILKEHP
 VENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
 KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
 QFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSE
 QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
 KVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFESPTVAY
 SVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKDLIIKLP
 KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
 QLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRK**QYR**STKEVLDATLIHQSIITGLYETRIDLSQLGGD (SEQ
 ID NO: 17)

[0138]

[0139] 서열식별번호: 15의 D1134, R1334 및 T1336으로부터 돌연변이되어 SpEQR Cas9를 생성할 수 있는 서열식별번호: 17의 잔기 E1134, Q1334 및 R1336은 밑줄표시되어 있고 볼드체이다.

[0140] 예시적인 SpVQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG
 ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLV
 EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFR
 GHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
 NLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLA
 QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
 VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
 DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLA
 RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
 LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLTLFEDR
 EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
 GFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
 VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
 VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
 LTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
 KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
 QFYKREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSE
 QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
 KVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAY
 SVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIKLP
 KYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
 QLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
 LTNLGAPAAFKYFDTTIDRK**QYR**STKEVLDATLIHQSIITGLYETRIDLSQLGGD (SEQ
 ID NO: 18)

[0141]

[0142] 서열식별번호: 15의 D1134, R1334 및 T1336으로부터 돌연변이되어 SpVQR Cas9를 생성할 수 있는 서열식별번호: 18의 잔기 V1134, Q1334 및 R1336은 밑줄표시되어 있고 볼드체이다.

[0143] 예시적인 SpVRER Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIG
ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLEESFLV
EEDKKHERHPFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFR
GHFLIEGDLNPDNSDVKLFQVLQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA
QIGDQYADFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKNRE
DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEENEDILEDIVLTTLTFEDR
EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSD
GFANRNFQMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEELGILGSQILKEHP
VENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
QFYKVRINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIKLP
KYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
QLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKEYSTKEVLDATLIHQ SITGLYETRIDLSQLGGD (SEQ
ID NO: 19)

[0144]

[0145] 서열식별번호: 15의 D1134, G1217, R1334 및 T1336으로부터 돌연변이되어 SpVRER Cas9를 생성할 수 있는 서열식별번호: 19의 잔기 V1134, R1217, Q1334 및 R1336은 밑줄표시되어 있고 볼드체이다.

[0146]

고충실도 Cas9 도메인

[0147]

본 개시내용의 일부 측면은 본원에 제공된 핵염기 편집제의 고충실도 Cas9 도메인을 제공한다. 일부 실시양태에서, 고충실도 Cas9 도메인은 상응하는 야생형 Cas9 도메인과 비교하여, Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 정전기적 상호작용을 감소시키는 1개 이상의 돌연변이를 포함하는 조작된 Cas9 도메인이다. 어떠한 특정한 이론에 얽매는 것을 원하지는 않지만, DNA의 당-포스페이트 백본과의 감소된 정전기적 상호작용을 갖는 고충실도 Cas9 도메인은 보다 적은 오프-타겟 효과를 가질 수 있다. 일부 실시양태에서, Cas9 도메인 (예를 들어, 야생형 Cas9 도메인)은 Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 회합을 감소시키는 1개 이상의 돌연변이를 포함한다. 일부 실시양태에서, Cas9 도메인은 Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 회합을 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70% 또는 그 초과만큼 감소시키는 1개 이상의 돌연변이를 포함한다.

[0148]

일부 실시양태에서, 본원에 제공된 임의의 Cas9 융합 단백질은 서열식별번호: 6에 제공된 아미노산 서열의 N497X, R661X, Q695X 및/또는 Q926X 돌연변이 중 1개 이상, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이(들)를 포함하고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 본원에 제공된 임의의 Cas9 융합 단백질은 서열식별번호: 6에 제공된 아미노산 서열의 N497A, R661A, Q695A 및/또는 Q926A 돌연변이 중 1개 이상, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, Cas9 도메인은 서열식별번호: 6에 제공된 아미노산 서열의 D10A 돌연변이, 또는 본원에 제공된 임

의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이를 포함한다. 일부 실시양태에서, Cas9 도메인 (예를 들어, 본원에 제공된 임의의 융합 단백질의 것)은 서열식별번호: 20에 제시된 바와 같은 아미노산 서열을 포함한다. 일부 실시양태에서, Cas9 도메인은 서열식별번호: 20에 대해 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. 고충실도를 갖는 Cas9 도메인은 관련 기술분야에 공지되어 있으며, 통상의 기술자에게 분명할 것이다. 예를 들어, 고충실도를 갖는 Cas9 도메인은 문헌 [Kleinstiver, B.P., et al. "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." Nature 529, 490-495 (2016); 및 Slaymaker, I.M., et al. "Rationally engineered Cas9 nucleases with improved specificity." Science 351, 84-88 (2015)]에 기재되어 있고; 각각의 전체 내용은 본원에 참조로 포함된다.

[0149] 본원에 제공된 임의의 염기 편집제, 예를 들어 본원에 제공된 임의의 C에서 G로의 염기 편집제는 본원에 기재된 바와 같은 Cas9 도메인을 변형시켜 고충실도 염기 편집제, 예를 들어 고충실도의 C에서 G로의 염기 편집제를 생성함으로써 고충실도 염기 편집제로 전환될 수 있다는 것인 인지되어야 한다. 일부 실시양태에서, 고충실도 Cas9 도메인은 dCas9 도메인이다. 일부 실시양태에서, 고충실도 Cas9 도메인은 nCas9 도메인이다.

[0150] 서열식별번호: 6의 Cas9 대비 돌연변이가 볼드체 및 밑줄로 제시된 고충실도 Cas9 도메인

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIG
ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLV
EEDKKHERHPFIGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFR
GHFLIEGDLNPDNSDVKLFQVLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
NLIAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLKIKDKDFLDNEENEDILEDIVLTTLTFEDR
EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSD
GFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV
VDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGKELGSQILKEHP
VENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF
QFYKVINNYHHADAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSE
QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLP
KYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQK
QLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSTGLYETRIDLSQLGGD (SEQ
ID NO: 20)

[0151]

[0152] 본 개시내용은 또한 napDNAbp의 단편, 예컨대 본원에 제공된 napDNAbp 중 임의의 것의 말단절단물을 제공한다. 일부 실시양태에서, napDNAbp는 napDNAbp의 N-말단에 1개 이상의 아미노산이 부재하는 N-말단 절단물이다. 일부 실시양태에서, napDNAbp는 napDNAbp의 N-말단에 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 또는 50개 아미노산이 부재한다. 예를 들어, napDNAbp의 N-말단 절단물은 본원에 제공된 임의의 napDNAbp, 예컨대 서열식별번호: 4-40 중 어느 하나에 제공된 napDNAbp 중 어느 하나의 N-

말단 절단물일 수 있다. 일부 실시양태에서, napDNAbp는 napDNAbp의 C-말단에 1개 이상의 아미노산이 부재하는 C-말단 절단물이다. 일부 실시양태에서, napDNAbp는 napDNAbp의 C-말단에 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 50개 아미노산이 부재한다. 예를 들어, napDNAbp의 C-말단 절단물은 본원에 제공된 임의의 napDNAbp, 예컨대 서열식별번호: 4-40 중 어느 하나에 제공된 NAP 중 어느 하나의 C-말단 절단물일 수 있다.

[0153] 일부 실시양태에서, 본원에 제공된 임의의 napDNAbp는 본원에 제공된 임의의 napDNAbp, 예컨대 서열식별번호: 4-40에 제공된 napDNAbp 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, 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, 50개 또는 그 초과 아미노산 변화를 갖는다.

[0154] 우라실 결합 단백질 (UBP)

[0155] 우라실 결합 단백질 또는 UBP는 우라실에 결합할 수 있는 단백질을 지칭한다. 일부 실시양태에서, 우라실 결합 단백질은 우라실 변형 효소이다. 일부 실시양태에서, 우라실 결합 단백질은 우라실 염기 절제 효소이다. 일부 실시양태에서, 우라실 결합 단백질은 우라실 DNA 글리코실라제 (UDG)이다. 일부 실시양태에서, 우라실 결합 단백질은 야생형 UDG (예를 들어, 인간 UDG)가 우라실에 결합하는 친화도의 적어도 1%, 2%, 3%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 또는 적어도 95%인 친화도로 우라실에 결합한다. 일부 실시양태에서, 우라실 결합 단백질은 우라실에 결합하는 야생형 우라실 결합 단백질, 예컨대 야생형 UDG (예를 들어, 인간 UDG)와 비교하여 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, 50개 또는 그 초과 아미노산 변화를 가질 수 있다.

[0156] 일부 실시양태에서, UBP는 우라실 변형 효소이다. 일부 실시양태에서, UBP는 우라실 염기 절제 효소이다. 일부 실시양태에서, UBP는 우라실 DNA 글리코실라제이다. 일부 실시양태에서, UBP는 본원에 제공된 임의의 우라실 결합 단백질이다. 예를 들어, UBP는 UDG, UdgX, UdgX*, UdgX_On 또는 SMUG1일 수 있다. 일부 실시양태에서, UBP는 우라실 결합 단백질, 우라실 염기 절제 효소 또는 우라실 DNA 글리코실라제 (UDG) 효소에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, UBP는 본원에 제공된 임의의 우라실 결합 단백질, 예를 들어 하기 제공된 UBP 및 UBP 변이체 중 임의의 것에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, UBP는 서열식별번호: 48-53 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, UBP는 서열식별번호: 48-53 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 우라실 결합 단백질은 본원에 제공된 임의의 UBP, 예컨대 서열식별번호: 48-53 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, 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, 50개 또는 그 초과 아미노산 변화를 갖는다.

[0157] 본 개시내용은 또한 UBP의 단편, 예컨대 본원에 제공된 임의의 UBP의 말단절단물을 제공한다. 일부 실시양태에서, UBP는 UBP의 N-말단에 1개 이상의 아미노산이 부재하는 N-말단 절단물이다. 일부 실시양태에서, UBP는 UBP의 N-말단에 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 50개 아미노산이 부재한다. 예를 들어, UBP의 N-말단 절단물은 본원에 제공된 임의의 UBP, 예컨대 서열식별번호: 48-53 중 어느 하나에 제공된 UBP 중 어느 하나의 N-말단 절단물일 수 있다. 일부 실시양태에서, UBP는 UBP의 C-말단에 1개 이상의 아미노산이 부재하는 C-말단 절단물이다. 일부 실시양태에서, UBP는 UBP의 C-말단에 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 50개 아미노산이 부재한다. 예를 들어, UBP의 C-말단 절단물은 본원에 제공된 임의의 UBP, 예컨대 서열식별번호: 48-53 중 어느 하나에 제공된 UBP 중 어느 하나의 C-말단 절단물일 수 있다.

[0158] 다른 UBP는 관련 기술분야의 통상의 기술자에게 분명할 것이고 본 개시내용의 범주 내에 있다는 것이 인지되어야 한다. 예를 들어, UBP는 이전에 문헌 [Sang et al., "A Unique Uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, Vol. 43, No. 17 2015]에 기재되었고; 이의 전체 내용은 본원에 참조로 포함된다.

UDG

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEE
PGTPSSPLSAEQLDRIQRNKAALLRLAARNVPVGFGEWKKHLSGEFGKPYFIKL
MGFVAEERKHHTVYPPPHQVFTWTQMCDIKDVKVVLGQDPYHGPNQAHGLCFVS
QRPVPPPPSLENIYKELSTDIEDFVHPGHGDLGWAKQGVLLNVLTVRAHQANSH
KERGWEQFTDAVVSWLNQNSNGLVFLWGSYAQKKGSAIDRKRHHVLQTAHPSPL
SVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL (SEQ ID NO: 48)

UdgX

MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGECPG
DKEDLAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKT
PSRTEVVACRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDV
PGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRLVAADVVP (SEQ ID NO: 49)

UdgX* (R107S)

MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGECPG
DKEDLAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKT
PSRTEVVACRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDV
PGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRLVAADVVP (SEQ ID NO: 50)

UdgX_On (H109S)

MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGECPG
DKEDLAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKT
PSRTEVVACRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDV
PGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRLVAADVVP (SEQ ID NO: 51)

Rev7

MTTLTRQDLNFGQVVADVLCEFLVAVHLILYVREVYPVGIFQKRKKYNVPVQMSC
HPELNQYIQDTLHCVKPLLEKNDVEKVVVVILDKEHRPVEKFVFEITQPPLLSISSDSL
LSHVEQLLRALFILKISVCDVLDHNPFGCTFTVLVHTREAATRNMEKIQVIKDFPWIL
ADEQDVHMHDPRLIPLKMTSDILKMQLYVEERAHKGS (SEQ ID NO: 52)

Smug1

MPQAFLGSIHEPAGALMEPQPCPGSLAESFLEEELRLNAELSQLQFSEPVGIIYNPVE
YAWEPHRNYVTRYCQGPKVFLFLGMNPGPFQMAQTGVPPFGEVSMVRDWLGIVGPV
LTPPQEHKRPVGLGLECPQSEVSGARFWGFFRNLCGQPEVFFHCFVHNLCPLLFLAP
SGRNLTAPALPAKQREQLLGICDAALCRQVQLLGVRLVVGVGRLAEQRARRALAGL
MPEVQVEGLLHPSRNPQANKGWEAVAKERLNLGLLPLLLK (SEQ ID NO: 53)

핵산 폴리머라제 (NAP)

핵산 폴리머라제 또는 NAP는 뉴클레오타이드 (예를 들어, 데옥시리보뉴클레오타이드 및 리보뉴클레오타이드)로부터 핵산 분자 (예를 들어, DNA 및 RNA)를 합성하는 효소를 지칭한다. 일부 실시양태에서, NAP는 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 트랜스레전 폴리머라제이다. 트랜스레전 폴리머라제는 돌연변이 유발에서, 예를 들어 복제 포크를 재시작하거나 또는 DNA 손상의 존재로 인해 게놈에 남아있는 갭을 채우는 것에 의해 소정 역할을 한다. 예시적인 트랜스레전 폴리머라제는, 비제한적으로, Pol 베타, Pol 람다, Pol 에타,

Pol 뮤, Pol 이오타, Pol 카파, Pol 알파, Pol 델타, Pol 감마 및 Pol 누를 포함한다.

[0163] 일부 실시양태에서, NAP는 진핵 핵산 폴리머라제이다. 일부 실시양태에서, NAP는 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 트랜스레전 폴리머라제 활성을 갖는다. 일부 실시양태에서, NAP는 트랜스레전 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파 또는 폴리머라제 에타이다. 일부 실시양태에서, NAP는 진핵 폴리머라제 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, 뮤 또는 누이다. 일부 실시양태에서, NAP는 자연 발생 핵산 폴리머라제 (예를 들어, 트랜스레전 DNA 폴리머라제)에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, NAP는 본원, 예를 들어 하기에 제공된 핵산 폴리머라제 중 임의의 것에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 예를 들어, NAP는 서열식별번호: 54-64 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함할 수 있다. 일부 실시양태에서, NAP는 서열식별번호: 54-64 중 어느 하나의 아미노산 서열을 포함한다. 다른 NAP는 관련 기술분야의 통상의 기술자에게 분명할 것이고 본 개시내용의 범주 내에 있다는 것이 인지되어야 한다. 일부 실시양태에서, NAP는 서열식별번호: 54-64 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 핵산 폴리머라제는 본원에 제공된 임의의 NAP, 예컨대 서열식별번호: 54-64 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, 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, 50개 또는 그 초과와 아미노산 변화를 갖는다.

[0164] 본 개시내용은 또한 NAP의 단편, 예컨대 본원에 제공된 임의의 NAP의 말단절단물을 제공한다. 일부 실시양태에서, NAP는 NAP의 N-말단에 1개 이상의 아미노산이 부재하는 N-말단 절단물이다. 일부 실시양태에서, NAP는 NAP의 N-말단에 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 50개 아미노산이 부재한다. 예를 들어, NAP의 N-말단 절단물은 본원에 제공된 임의의 NAP, 예컨대 서열식별번호: 54-64 중 어느 하나에 제공된 NAP 중 어느 하나의 N-말단 절단물일 수 있다. 일부 실시양태에서, NAP는 NAP의 C-말단에 1개 이상의 아미노산이 부재하는 C-말단 절단물이다. 일부 실시양태에서, NAP는 NAP의 C-말단에 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 50개 아미노산이 부재한다. 예를 들어, NAP의 C-말단 절단물은 본원에 제공된 임의의 NAP, 예컨대 서열식별번호: 54-64 중 어느 하나에 제공된 NAP 중 어느 하나의 C-말단 절단물일 수 있다.

[0165] Pol 베타

MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVIAKYPHKIKSG
AEAKKLPGVGTGKIAEKIDEFLATGKLRKLEKIRQDDTSSINFLTRVSGIGPSAARKFV
DEGIKTLEDLRKNEDKLNHHQRIGLKLYFGDFEKRIPREMLQMMDIVLNEVKKVDSE
YIATVCGSFRRGAESSGDMVLLTHPSFTSESTKQPKLLHQVVEQLQKVHFITDTLSK
GETKFMGVCQLPSKNDEKEYPHRRIDIRLIPKDQYYCGVLYFTGSDIFNKNMRAHAL
EKGFTINEYTIRPLGVTGVAGEPLPVDSEKDIFDYIQWKYREPKDRSE (SEQ ID NO:

54)

[0166]

[0167] PoI 람다

MDPRGILKAFPKRQKIHADASSKVLAKIPRREEGEEAEWLSSLRAHVVRTGIGRAR
 AELFEKQIVQHGGQLCPAQGPVTHIVVDEGMDYERALLRLPQLPPGAQLVKSA
 WLSLCLQERRLDVAGFSIFPSRYLDHPQPSKAEQDASIPPGTHEALLQTALSPPPPP
 TRPVSPQKAKEAPNTQAQPISSDDEASDGEETQVSAADLEALISGHYPTSLEGDCEPS
 PPAVLDKWVCAQPSSQKATNHNHHITEKLEVLAKAYSVQGDKWRALGYAKAINA
 LKSFHKPVTSYQEACSIPIGIGKRM AEKIIIEILESGHLRKLDHISESVPVLELFSNIWGAG
 TKTAQMWYQQGFRSLEDIRSQASLTQQQAIGLKHYSDFLERMPREEATEIEQTVQKA
 AQAFNSGLLCVACGSYRRGKATCGDVLITHPDGRSHRGIFSRLDLSLRQEGFLTD
 DLVSQEENGQQQKYLGVCRPLPGPGRRHRRLDIIVVPYSEFACALLYFTGSAHFNRSM
 RALAKTKGMSLSEHALSTAVVRNTHGCKVGPGRVLPPTTEKDVFRLLGLPYREPAE
 RDW (SEQ ID NO: 55)

[0168]

[0169] PoI 에타

MATGQDRVVALVDMDCFFVQVEQRQNPFLRNKPCAVVQYKSWKGGGIIAVSYEA
 RAFGVTRSMWADDAKKLCPDLLLAQVRESRGKANLTKYREASVEVMEIMSRFAVIE
 RASIDEAYVDLTSAVQERLQKLQGQPISADLLPSTYIEGLPQGPTTAETVQKEGMRK
 QGLFQWLDLQIDNLTSPDLQLTVGAVIVEEMRAAIERETGFQCSAGISHNKVLAKL
 ACGLNKNRQTLVSHGSVPQLFSQMPIRKIRSLGGKLGASVIEILGIEYMGELTQFTES
 QLQSHFGEKNGSWLYAMCRGIEHDPVKPRQLPKTIGCSKNFPGKTALATREQVQWW
 LLQLAQELEERLT KD RNDNDRVATQLVVSIRVQGDKRLSSLRRCCALTRYDAHKMS
 HDAFTVIKNCNTSGIQTEWSPPLTMLFLCATKFSASAPSSSTDITSFLSSDPSSLPKVPV
 TSSEAKTQGS GPAVTATKATTSLSEFFQKAAERQKVKEASLSSLTAPTQAPMSNSPS
 KPSPFPQTSQSTGTEPFFKQKSLLLKQKQLNNSVSSPQQNPWSNCKALPNSLPTEYP
 GCVPVCEGVSKLEESSKATPAEMDLAHSQSMHASSASKSVLEVTQKATPNPSLLA
 AEDQVPCEKCGSLVPVWDMPEHMDYHFALELQKSFLQPHSSNPQVVS AVSHQGKR
 NPKSPLACTNKRPRPEGMQTLESFFKPLTH (SEQ ID NO: 56)

[0170]

[0171] PoI 뮤

MLPKRRRRARVGSPSGDAASSTPPSTRFPGVAIYLV EPRMGRSRRRAFLTGLARSKGFR
 VLDACSSEATHVMEETSAAEAVSWQERRMAAAPPGCTPPALLDISWLTESLGAGQ
 PVPVECRHRLEVAGPRKGPLSPAWMPAYACQRPTPLTHHNTGLSEALEILAEAAAGFE
 GSEGRLLTFCRAASVLKALPSPVTTLSQLQGLPHFGEHSSRVVQELLEHGVCEEVER
 VRRSERYQTMKLFTQIFGVGKTA DRWYREGLRTLDDLREQPQKLTQQQKAGLQH
 HQDLSTPVLRSVDALQQVVEEAVGQALPGATVTLTG GFRRGKLQGHVDVDFLITHP
 KEGQEAGLLPRVMCRLQDQGLILYHQHQHSCCESPTRLAQQSHMDAFERSFCIFRLP
 QPPGAAVGGSTRPCPSWKAVRVLDLVVAPVSQFPFALLGWTGSKLFQRELRRFSRKE
 KGLWLNSHGLFDPEQKTFQAAASEEDIFRHLGLEYPPEQRNA (SEQ ID NO: 57)

[0172]

[0173]

Pol 이오타

MEKLGVEPEEEGGDDDEEDAEAWAMELADVGAASSQGVHDQVLPTPNASSRVI
VHVDLDCFYAQVEMISNPELKDKPLGVQQKYLVVTCNYEARLGVKKLMNVRDA
KEKCPQLVLVNGEDLTRYREMSYKVTELEEFSPVVERLGFDENFVDLTEMVEKRL
QQLQSDELSAVTVSGHVVYNNQSNLLDVLHIRLLVGSQIAAEMREAMYNQLGLTGC
AGVASNKLLAKLVSGVFKPNQQTIVLLPESCQHLSLNIKEIPGIGYKTAKCLEALG
INSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSGPPQSFSEEDSFKKCSSEV
EAKNKIEELLASLLNRVCQDGRKPHTVRLIIRRYSEKHYGRESRQCPIPSHVIQKLGT
GNYDVMTPMVDILMKLFRNMVNVKMPFHLTLLSVCFCNLKALNTAKKGLIDYYLM
PSLSTTSRSGKHSFKMKDTHMEDFPKDKETNRDFLPSGRIESTRTRESPLDTTNFSKE
KDINEFPLCSLPEGVDQEVFKQLPVDIQEEILSGKSREKFQKGKSVSCPLHASRGVLSF
FSKKQMOPINPRDHLSSSKQVSSVSPCEPGTSGFNSSSSSYMSSQKDYSYYLDNRL
KDERISQGPKEPQGFHTNSNPAVSAFHSFPNLQSEQLFSRNHTTDSHKQTVATDSHE
GLTENREPDVDEKITFSPDIDPQVFYELPEAVQKELLAEWKRAGSDFHIGHK (SEQ
ID NO: 58)

[0174]

[0175]

Pol 카과

MDSTKEKCDSYKDDLLRMGLNDNKAGMEGLDKEKINKIIMEATKGSRFYGNELK
KEKQVNQRIENMMQQAQITSQQLRKAQLQVDRFAMELEQSRNLSNTIVHIDMDAF
YAAVEMRDNPCLKDKPIAVGSMMLSTSNYHARRFGVRAAMPGFIAKRLCPQLIIVP
PNFDKYRAVSKEVKEILADYDPNFMAMSLDEAYLNITKHLEERQNWPEDKRRYFIK
MGSSVENDNPGKEVNKLSEHERSISPLLFEESPSDVQPPGDPFQVNFEEQNNPQILQN
SVVFGTSAQEVVKEIRFRIEQKTTLTASAGIAPNTMLAKVCSDKNKPNGQYQILPNRQ
AVMDFIKDLPIRKVSGIGKVTEKMLKALGIITCTELYQQRALLSLLFSETSWHYFLHIS
LGLGSTHLTRDGERKSMSVERTFSEINKAEEQYSLCQELCSELAQDLQKERLKGRTV
TIKKNVNFEVKTRASTVSSVSTAEEIFAIAKELLKTEIDADFPPLRLRLMGVRISSE
PNEEDRKHQQRSIIGFLQAGNQALSATECTLEKTDKDKFVKPLEMSHKKSFFDKKRS
ERKWSHQDTFKCEAVNKQSFQTSQPFQVLKKKMNNENLEISENSDDCQILTCPVCFRA
QGCISLEALNKHVDECLDGPSISENFKMFSCSHVSATKVNKKENVPASSLCEKQDYE
AHPKIKEISSVDCIALVDTIDNSSKAESIDALSNKHSKEECSSLPSKSFNIEHCHQNSSS
TVSLENEDVGSFRQEYRQPYLCEVKTGQALVCPVCNVEQKTSDLTLFNVHVDVCLN
KSFQELRKDKFNPVNQPKESSRSTGSSSGVQKAVTRTRKRPGLMTKYSTSKKIKPNNP
KHTLDIFFK (SEQ ID NO: 59)

[0176]

[0177]

Pol 알파

MAPVHGDDCEIGASALSDSGSFVSSRARREKSKKGRQEALERLKKAKAGEKYKYE
VEDFTGVYEEVDEEQYSKLVQARQDDDWIVDDDGIGYVEDGREIFDDDLDDALDA
DEKGDGDKARNKDKRNVKKLAVTKPNNIKSMFIACAGKKTADKAVDLSKDGLLD
ILQDLNTETPQITPPPMILKKRSIGASPNPFSVHTATAVPSGKIASPVSRKEPPLTPV
PLKRAEFAGDDVQVESTEEEQESGAMEFEDGDFDEPMEVEEVDLEPMAAKAWDKE
SEPAEEVKQEADSGKGTVSYLGSFLPDVSCWDIDQEGDSSFSVQEVQVDSSHLPLVK
GADEEQVFHFYWLDAYEDQYNQPGVVFLFGKVVWIESAETHVSCCVMVKNERTLYF
LPREMKIDLNTGKETGTPISMKDVYEEFDEKIATKYKIMKFKSKPVEKNYAFEIPDVP
EKSEYLEVKYSAEMPQLPQDLKGETFSHVFGTNTSSLELFLMNRKIKGPCWLEVKSP
QLLNQPVSWCKVEAMALKPDLVNVIKDVSPPLVVMFAFSMKTMQNAKNHQNEIIA
MAALVHHSFALDKAAPKPPFQSHFCVVSFKPKDCIFPYAFKEVIEKKNVKEVAATER
TLLGFFLAKVHKIDPDIVGHNIYGFELEVLLQRINVCKAPHWSKIGRLKRSNMPKLG
GRSGFGERNATCGRMICDVEISAKELIRCKSYHLSLVQQILKTERVVIPMENIQNMY
SESSQLLYLLEHTWKDAKFIQIMCELNVLPALQITNIAGNIMSRTLMGGRSERNEF
LLLHAFYENNYIVPDKQIFRKPQQKLGDDEEIDGDTNKYKKGRKKAAYAGGLVLD
PKVGFYDKFILLDFNSLYPSIIEFNICFTTVQRVASEAQKVTEDEGEQEQIPELPDPSL
EMGILPREIRKLVERRKQVKQLMKQQDLNPDLLQYDIRQKALKLTANSMYGCLGFS
YSRFYAKPLAALVTYKGREILMHTKEMVQKMNLEVIYGDTSIMINTNSTNLEEVFK
LGNKVKSEVNKLYKLEIDIDGVFKSLLLKKKKYAALVVEPTSDGNYVTKQELKG
LDIVRRDWCDLAKDTGNFVIGQILSDQSRDTIVENIQKRLIEIGENVLNGSVVPSQFEI
NKALTKDPQDYPDKKSLPHVHVALWINSQGGRKVKAGDTVSYVICQDGSNLTAQR
AYAPEQLQKQDNLTIDTQYYLAQQIHPVVARICEPIDGIDAVLIATWLGLDPTQFRVH
HYHKDEENDALLGGPAQLTDEEKYRDCERFKCPCPTCGTENIYDNVFDGSGTDMEP
SLYRCSNIDCKASPLTFTVQLSNKLIMDIRRFIKKYDGLWICEEPTCRNRTRHLPLQF
SRTGPLCPACMKATLQPEYSDKSLYTQLCFYRYIFDAECALEKLTDDHEKDKLKKQF
FTPKVLQDYRKLKNTAEQFLSRSGYSEVNLSKLFAGCAVKS (SEQ ID NO: 60)

[0178]

[0179]

PoI 델타

MDGKRRPGPGVPPKRARGGLWDDDDAPRPSQFEEDLALMEEMEAHRLQEQUEE
EELQSVLEGVADGQVPPSAIDPRWLRPTPALDPQTEPLIFQQLEIDHYVGPAQPVPG
GPPPSHGSVPVLRAFGVTDEGFSVCCHIHGFAPYFYTPAPPFGPEHMGDLQRELNL
AISRDSRGGRELTGPAVLAVELCSRESMFGYHGHGSPFLRITVALPRLVAPARRLLE
QGIRVAGLGTSPFAPYEANVDFEIRFMVDTDIVGCNWLELPAGKYALRLKEKATQCQ
LEADVLSWDVVSHPPEGPWQRIAPLRVLSFDIECAGRKGIFPEPERDPVIQCSLGLRW
GEPEPFLRLALTLRPCAPILGAKVQSYEKEEDLLQAWSTFIRIMDPDVITGYNIQNFDL
PYLISRAQTLKVQTFPFLGRVAGLCSNIRDSSFQSKQTGRRDTKVVSVMVGRVQMDM
LQVLLREYKLSYTLNAVSFHFLGEQKEDVQHSIITDLQNGNDQTRRRRLAVYCLKDA
YLPLRLLERLMVLVNAVEMARVTGVPLSYLLSRGQQVKVVSQLLRQAMHEGLLMP
VVKSEGGEDYTGATVIEPLKGYDVPATLDFSSLYPSIMMAHNLCYTTLLRPGTAQ
KLGLTEDQFIRTPTGDEFVKTSVRKGLLPQILENLLSARKRAKAEKAKETDPLRRQVL
DGRQLALKVSANSVYGFTGAQVGKLPCEISQSVTGFRQMIEKTKQLVESKYTVEN
GYSTSAAKVVYGDTSVMCRFGVSSVAEAMALGREAADWVSGHFPSPIRLEFEKVYF
PYLLISKRYAGLLFSSRPDAHMRMDCKGLEAVRRDNCPLVANLVTASLRLLIDRD
PEGAVAHQAQDVISDLLCNRIDISQLVITKELTRAASDYAGKQAHVELAERMKRDPG
SAPSLGDRVPYVIISAAKGVAAYMKSEDPLFVLEHSLPIDTQYYLEQQLAKPLLRIFEP
ILGEGRAEAVLLRGDHTRCTVLTGKVGGLLAFKRRNCCIGCRTVLSHQGAVCEF
CQPRESELYQKEVSHLNALEERFSRLWTQCQRCQGSLEDVICTSRDCPIFYMRKKV
RKDLEDQEQLLRRFGPPGPEAW (SEQ ID NO: 61)

[0180]

[0181] Pol 감마

MSRLLWRKVAGATVGPVPAPGRWVSSSVASDPSDGQRRRQQQQQQQQQQQ
 QPQQPVLSSEGGQLRHNPLDIQMLSRGLHEQIFGQGGEMPGEAAVRRSVEHLQKH
 GLWGQPAVPLPDVELRLPLYGDNLDQHFRLLAQKQSLPYLEAANLLLQAQLPPKPP
 AWAWAEGWTRYGPEGEAVPVAIPEERALVFDVEVCLAEGTCPTLAVASPSAWYSW
 CSQRLVEERYSWTSQSPADLPLEVPTGASSPTQRDWQEQLVVGHNVSFDRAHIRE
 QYLIQGSRMFLDTMSMHMAISGLSSFQSRSLWIAAKQGKHKVQPPTKQGQKSQRKA
 RRGPAISSWDWLDISSVNSLAEVHRLYVGGPPLEKEPRELFVKGTMKDIRENFQDLM
 QYCAQDVWATHEVFQQQLPLFLERCPHPVTLAGMLEMGVSYLPVNQNWERYLAE
 QGTYEELQREMKKSLMDLANDACQLLSGERYKEDPWLWDLEWDLQEFKQKKAKK
 VKKEPATASKLPIEGAGAPGDPMDQEDLGPCSEEEEFQQDVMARACLQKLKGTTEL
 LPKRPQHLPHPGWYRKLCPRLDDPAWTPGPSLLSLQMRVTPKLMALTWDGFPLHY
 SERHGWGYLVPGRRDNLAKLPTGTTLESAGVVCYPYRAIESLYRKHCLEQGKQQLMP
 QEAGLAEFLLTDNSAIWQTVEELDYLEVEAEAKMENLRAAVPGQPLALTARGGPK
 DTQPSYHHGNPNYNDVDIPGCWFFKLPHKDGNSCNVGSFPAKDFLPKMEDGTLQAG
 PGGASGPRALEINKMISFWRNAHKRISSQMVVWLPRSALPRAVIRHPDYDEEGLYGA
 ILPQVVTAGTITRRAVEPTWLTASNARPDRVGSELKAMVQAPPGYTLVGADVDSQE
 LWIAAVLGDAHFAGMHGCTAFGWMTLQGRKSRGTDLHSKTATTVGISREHAKIFNY
 GRIYGAGQPFAERLLMQFNHRLTQQEAAEKAQQMYAATKGLRWYRLSDEGEWL
 RELNLPVDRTEGGWISLQDLRKVQRETARKSQWKKWEVVAERAWKGGTESEMFN
 KLESIAATSDIPRTPVLGCCISRALEPSAVQEEFMTSRVNWVQSSAVDYLHMLVAM
 KWLFEFAIDGRFCISIHDEVRYLVREEDRYRAALALQITNLLTRCMFAYKLGLNDLP
 QSVAFFSAVDIDRCLRKEVTMDCKTPSNPTGMERRYGIPQGEALDIYQIELTKGSLE
 KRSQPGP (SEQ ID NO: 62)

[0182]

[0183] Pol 뉴

MENYEALVGFDLCNTPLSSVAQKIMSAMHSGDLVDSKTWKGSTETMEVINKSSVKY
 SVQLEDKRTQSPEKKDLKSLRSQTSRGSAKLSPQSFSVRLTDQLSADQKQKSISSLTL
 SCLIPQYNQEASVLQKKGHKRKHFLMENINNENKGSINLKRKHITYNNLSEKTSKQ
 MALEEDTDDAEGYLNNGNSGALKKHFCDIRHLDDWAKSQLIEMLKQAAALVITVM
 YTDGSTQLGADQTPVSSVRGIVVLVKRQAEGGHGCPDAPACGPVLEGFVSDDPCIYI
 QIEHSAIWDQEQAHHQFARNVLFQTMKCKCPVICFNAKDFVRIVLQFFGNDGSKW
 HVADFIGLDPRIAAWLIDPSDATPSFEDLVEKYCEKSITVKVNSTYGNSSRNIVNQNV
 RENLKTLYRLTMDLCSKLDYGLWQLFRTLELPLIPILAVMESHAIQVNKEEMKTS
 ALLGARLKELEQEAHFVAGERFLITSNNQLREILFGKLKLHLLSQRNSLPRTGLQKYP
 STSEAVLNALRDLHPLPKIILEYRQVHKIKSTFVDGLLACMKKGSISSTWNQTGTVTG
 RLSAKHPNIQGISKHPIQITTPKNFKGKEDKILTISPRAMFVSSKGHTFLAADFSQIELRI
 LTHLSGDPELLKLQESERDDVFSTLTSQWKDVPVEQVTHADREQTKKVYAVVYG
 AGKERLAACLGVPQIEAAQFLESFLQYKKIKDFARAAIAQCHQTGCVVSIMGRRRP
 LPRIHAHDQQLRAQAERQAVNFVVQGSAADLCKLAMIHVFTAVAASHTLTARLVAQ
 IHDELLFEVEDPQIECAALVRRRTMESLEQVQALELQLQVPLKVSLSAGRSWGHLVP
 LQEAWGPPPGPCRTESPSNSLAAPGSPASTQPPPLHFSPSFCL (SEQ ID NO: 63)

[0184]

[0185] Rev1

MRRGGWRKRAENDGWETWGGYMAAKVQKLEEQFRSDAAMQKDGTSSTIFSGVAI
YVNGYTDPSAEELRKLMLHGGQYHVYYSRSKTTHIATNLPNAKIKELKGEKVIRP
EWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPLPGPSNIAKQLNNRVN
HIVKKIETENEVKVNGMNSWNEEDENNDFSVDLEQTSPGRKQNGIPHRGTAIFN
GHTPSSNGALKTQDCLVPMVNSVASRLSPAFSQEEDKAEKSSTDFRDCTLQQLQQST
RNTDALRNPHTNSFSLSPHLSNTKINGAHHSTVQGPSSTKSTSSVSTFSKAAPSVPSK
PSDCNFISNFYSHSRLHHISMWKCELTEFVNTLQQRQNGIFPGREKLKKMKTGRSALV
VTDTGDMSVLNSPRHQSCIMHVDMDCCFFVSVGIRNRPDLKGKPVAVTSNRGTGRAP
LRPGANPQLEWQYYQNKILKGKAADIPDSSLWENPDSAQANGIDSVLSRAEIASCSY
EARQLGIKNGMFFGHAKQLCPNLQAVPYDFHAYKEVAQTLYETLASYTHNIEAVSC
DEALVDITEILAETKLTPDEFANAVRMEIKDQTKCAASVGIGSNILLARMATRKAKPD
GQYHLKPEEVDDFIRGQLVTNLPVGVGHSMESKLASLGIKTCGDLQYMTMAKLQKEF
GPKTGQMLYRFRGLDDRVRTEKERKSVSAEINYGIRFTQPKEAEAFLLSLSEEIQR
RLEATGMKGKRLTLKIMVRKPGAPVETAKFGGHHGICDNIARTVTLDQATDNAKIIGK
AMLNMFHTMKLNISDMRGVGIHVNLVPTNLNPSTCPSRPSVQSSHFPSSGSYSVRDV
FQVQKAKKSTEEHKEVFRAAVDLEISSASRTCTFLPPFAHLPTSPDTNKAESSGKW
NGLHTPVSQVSRNLNLSIEVSPSQLDQSVLEALPPDLREQVEQVCAVQQAESHGDKK
KEPVNGCNTGILPQPVGTVLLQIPEPQESNSDAGINLIALPAFSQVDPEVFALPAELQ
RELKAAVDQRQRQGENSTHQQSASASVPKNPLHLKAAVKEKKRNKKKTIGSPKR
IQSPLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLSSL
QSDPAGCVRPPAPNLAGAVEFNDVKTLLEWITTISDPMEEDILQVVKYCTDLIEEKD
LEKLDLVIKYMKRLMQQSVESVWNMAFDLFDNVQVVLQQTYGSTLKVT (SEQ ID
NO: 64)

[0186]

[0187] 염기 절제 효소 (BEE)

[0188] 염기 절제 효소 또는 BEE는 핵산 분자 (예를 들어, DNA 또는 RNA)로부터 염기 (예를 들어, A, T, C, G 또는 U)를 제거할 수 있는 단백질을 지칭한다. 일부 실시양태에서, BEE는 DNA로부터 시토신을 제거할 수 있다. 일부 실시양태에서, BEE는 DNA로부터 티민을 제거할 수 있다. 예시적인 BEE는, 비제한적으로, 문헌 [Sang et al., "A Unique Uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, Vol. 43, No. 17 2015]에 기재된 바와 같은 UDG Tyr147Ala 및 UDG Asn204Asp를 포함하고; 이의 전체 내용은 본원에 참조로 포함된다.

[0189]

일부 실시양태에서, 염기 절제 효소 (BEE)는 시토신, 티민, 아데닌, 구아닌 또는 우라실 염기 절제 효소이다. 일부 실시양태에서, 염기 절제 효소 (BEE)는 시토신 염기 절제 효소이다. 일부 실시양태에서, BEE는 티민 염기 절제 효소이다. 일부 실시양태에서, 염기 절제 효소는 자연 발생 BEE에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 염기 절제 효소는 본원에 제공된 임의의 BEE, 예를 들어 하기 UDG (Tyr147Ala) 또는 UDG (Asn204Asp)에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 염기 절제 효소는 서열식별번호: 65-66 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 염기 절제 효소는 서열식별번호: 65-66 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 염기 절제 효소는 본원에 제공된 임의의 BEE, 예컨대 서열식별번호: 65-66 중 어느 하나와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 또는 그 초과와 아미노산 변화를 갖는다.

[0190]

본 개시내용은 또한 BEE의 단편, 예컨대 본원에 제공된 임의의 BEE의 말단절단물을 제공한다. 일부 실시양태에서, BEE는 BEE의 N-말단에 1개 이상의 아미노산이 부재하는 N-말단 절단물이다. 일부 실시양태에서, BEE는 BEE의 N-말단에 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 또는 50 개 아미노산이 부재한다. 예를 들어, BEE의 N-말단 절단물은 본원에 제공된 임의의 BEE, 예컨대 서열식별번호: 65-66 중 어느 하나에 제공된 BEE 중 어느 하나의 N-말단 절단물일 수 있다. 일부 실시양태에서, BEE는 BEE의 C-말단에 1개 이상의 아미노산이 부재하는 C-말단 절단물이다. 일부 실시양태에서, BEE는 BEE의 C-말단에 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 50개 아미노산이 부재한다. 예를 들어, BEE의 C-말단 절단물은 본원에 제공된 임의의 BEE, 예컨대 서열식별번호: 65-66 중 어느 하나에 제공된 BEE 중 어느 하나의 C-말단 절단물일 수 있다.

[0191] 다른 BEE는 관련 기술분야의 통상의 기술자에게 분명할 것이고 본 개시내용의 범주 내에 있다는 것이 인지되어야 한다. 예를 들어, BEE는 이전에 문헌 [Sang et al., "A Unique Uracil-DNA binding protein of the uracil DNA glycosylase superfamily," Nucleic Acids Research, Vol. 43, No. 17 2015]에 기재되었고; 이의 전체 내용은 본원에 참조로 포함된다.

[0192] UDG (Tyr147Ala) - 돌연변이된 잔기는 볼드체 및 밑줄로 표시된다.

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEE
PGTPSSPLSAEQLDRIQRNKAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKL
MGFVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQDPAHGPNQAHGLCFSV
QRPVPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLNVLTVRAHQANSH
KERGWEQFTDAVVSWLNQNSNGLVFLWGSYAQKKGSAIDRKRHHVLQTAHPSPL
SVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL (SEQ ID NO: 65)

[0193]

[0194] UDG (Asn204Asp) - 돌연변이된 잔기는 볼드체 및 밑줄로 표시된다.

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEE
PGTPSSPLSAEQLDRIQRNKAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKL
MGFVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQDPYHGPNQAHGLCFSV
QRPVPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLNDAVLTVRAHQANSH
KERGWEQFTDAVVSWLNQNSNGLVFLWGSYAQKKGSAIDRKRHHVLQTAHPSPL
SVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL (SEQ ID NO: 66)

[0195]

[0196] 테아미나제 도메인

[0197] 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질 또는 염기 편집체는 시티딘 테아미나제 도메인을 포함한다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 C에서 U로의 염기 변화를 촉매할 수 있다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 아포지단백질 B mRNA-편집 복합체 (APOBEC) 패밀리의 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC1 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC2 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3A 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3B 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3C 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3D 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3E 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3F 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3G 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC3H 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 APOBEC4 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 활성화-유도된 테아미나제 (AID)이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 척추동물 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 무척추동물 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 인간, 침팬지, 고릴라, 원숭이, 소, 개, 래트 또는 마우스 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 인간 테아미나제이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 래트 테아미나제, 예를 들어 rAPOBEC1이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 페트로미존 마린우스(*Petromyzon marinus*) 시티딘 테아미나제1 (pmCDA1)이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 인간 APOBEC3G (서열식별번호: 77)이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 인간 APOBEC3G의 단편 (서열식별번호:

100)이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 D316R_D317R 돌연변이를 포함하는 인간 APOBEC3G 변이체 (서열식별번호: 99)이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 서열식별번호: 77에서의 D316R_D317R 돌연변이에 상응하는 돌연변이를 포함하는, 인간 APOBEC3G의 단편 (서열식별번호: 101)이다.

[0198] 일부 실시양태에서, 시티딘 테아미나제 도메인은 자연-발생 시티딘 테아미나제에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일하다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 본원에 제공된 시티딘 테아미나제 중 임의의 것에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일하다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 서열식별번호: 67-101 중 어느 하나의 테아미나제 도메인에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일하다. 일부 실시양태에서, 핵산 편집 도메인은 서열식별번호: 67-101 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 본원에 제공된 임의의 시티딘 테아미나제 도메인, 예컨대 서열식별번호: 67-101 중 어느 하나와 비교하여 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50개 또는 그 초과 아미노산 변화를 갖는다.

[0199] 본 개시내용은 또한 시티딘 테아미나제 도메인의 단편, 예컨대 본원에 제공된 임의의 시티딘 테아미나제 도메인의 말단절단물을 제공한다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 시티딘 테아미나제 도메인의 N-말단에 1개 이상의 아미노산이 부재하는 N-말단 절단물이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 시티딘 테아미나제 도메인의 N-말단에 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 또는 50개 아미노산이 부재한다. 예를 들어, 시티딘 테아미나제 도메인의 N-말단 절단물은 본원에 제공된 임의의 시티딘 테아미나제 도메인, 예컨대 서열식별번호: 67-101 중 어느 하나에 제공된 시티딘 테아미나제 도메인 중 어느 하나의 N-말단 절단물일 수 있다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 시티딘 테아미나제 도메인의 C-말단에 1개 이상의 아미노산이 부재하는 C-말단 절단물이다. 일부 실시양태에서, 시티딘 테아미나제 도메인은 시티딘 테아미나제 도메인의 C-말단에 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 또는 50개 아미노산이 부재한다. 예를 들어, 시티딘 테아미나제 도메인의 C-말단 절단물은 본원에 제공된 임의의 시티딘 테아미나제 도메인, 예컨대 서열식별번호: 67-101 중 어느 하나에 제공된 시티딘 테아미나제 도메인 중 어느 하나의 C-말단 절단물일 수 있다.

[0200] 일부 예시적인 시티딘 테아미나제 도메인은, 비제한적으로, 하기 제공된 것들을 포함한다. 일부 실시양태에서, 각각의 서열의 활성 도메인, 예를 들어 국제화 신호가 없는 도메인 (핵 방출 신호, 세포질 국제화 신호가 없는 핵국제화 서열)이 사용될 수있음이 이해되어야 한다.

[0201] 인간 AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNKC
HVELLFLRYISDWLDLPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTAR
LYFCEDRKAEPGLRRLHRAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGLHEN
SVRLSRQLRRILLPLYEVDDLRLDAFRTLGL (SEQ ID NO: 67)

[0202]

(밑줄: 핵 국제화 서열; 이중 밑줄: 핵 유출 신호)

[0204] 마우스 AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSCSLDFGHLRNKSGC
HVELLFLRYISDWLDLPGRCYRVTWFTSWSPCYDCARHVAEFLRWPNLSLRIFTAR
LYFCEDRKAEPGLRRLHRAGVQIGIMTFKDYFCWNTFVENRERTFKAWGLHEN
SVRLTRQLRRILLPLYEVDDLRLDAFRMLGF (SEQ ID NO: 68)

[0205]

(밑줄: 핵 국제화 서열; 이중 밑줄: 핵 유출 신호)

[0206]

- [0207] 개 AID:
- MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAAR
LYFCEDRKAEPGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHEN
SVRLSRQLRRILLPLYEYVDDLRDAFRTLGL (SEQ ID NO: 69)
- [0208]
- [0209] (밀줄: 핵 국제화 서열; 이중 밀줄: 핵 유출 신호)
- [0210] 소 AID:
- MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTAR
LYFCDKERKAEPGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHE
NSVRLSRQLRRILLPLYEYVDDLRDAFRTLGL (SEQ ID NO: 70)
- [0211]
- [0212] (밀줄: 핵 국제화 서열; 이중 밀줄: 핵 유출 신호)
- [0213] 래트 AID
- MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRS
LLMKQRKFLYHFKNVRWAKGRHETYL CYVVKRRDSATSFSLDFGYLRNKSCHVE
LLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTG
WGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSRLSRLLRILLPL
YEYVDDLRDAFRTLGL (SEQ ID NO: 71)
- [0214]
- [0215] (밀줄: 핵국제화 서열; 이중 밀줄: 핵유출 신호)
- [0216] 마우스 APOBEC-3:
- MGPFCGLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPV
SLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEAEQIVRFL
ATHHNLSDIFSSRLYNVQDPETQQNLCLVQEGAQVAAMDLYEFKKCWKKFVDN
GRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEG
RRMDPLSEEEFYQFYNQVRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGG
QHAELFLDKIRSMELSQVITCYLTWSPCPNCAWQLAAFKRDRPDILHIYTSRLYFHW
KRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRRL
RRIKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 72)
- [0217]
- [0218] (이텔릭체: 핵산 편집 도메인)
- [0219] 래트 APOBEC-3:
- MGPFCGLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVS
LHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEAEQVLRFLA
THHNLSDIFSSRLYNIRDPENQQNLCLVQEGAQVAAMDLYEFKKCWKKFVDNNGG
RRFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRR
VHLLSEEEFYQFYNQVRVKHLCYYHGVPYLCYQLEQFNGQAPLKGCLLSEKGGKH
AEILFLDKIRSMELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDILHIYTSRLYFHWKR
PFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLLHR
IKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 73)
- [0220]
- [0221] (이텔릭체: 핵산 편집 도메인)

- [0222] 레서스 마카크 APOBEC-3G:
- MVEPMDPRTFVSNFNRRPILSGLNTVWLCCEVKT KDPSGPPLDAKIFQGKVYSKAKY
HPEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVA
 RLYYFWKPDYQQALRILCQKRGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRN
 NLPKHYTLLQATLGELLRLHMDPGTFTSNFNKPPWVSGQHETYLCKYVERLHNDT
 WVPLNQHRGFLRNQAPNIHGFPGKGRHAELCFDLIPFWKLDGQQYRVTCFTSWSPCF
 CAQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGRLALHRDGAKIAMMNYSEFEYC
 WDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 74)
- [0223]
- [0224] (이텔릭체: 핵산 편집 도메인; 밑줄: 세포질 국제화 신호)
- [0225] 침팬지 APOBEC-3G:
- MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEYVKTGKPSRPPLDAKIFRGQ
VYSKLKYHPEMRFFHWFSKWRKLHRDQEYEVWYISWSPCTKCTRDVATFLAEDPKV
 TLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRE
 LFEPWNNLPKYYILLHIMLGEILRHSMPTFTSNFNELWVRGRHETYLCEVERL
 HNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFDLVIPFWKLDLHQDYRVTCFTS
 WSPCFSCAQEMAKFISNNKHVSLCIFAARIYDDQGRQCQGLRTLAKAGAKISIMTYSE
 FKHCWDTFVDHQGCPFPQWDGLEHSQALSGRLRAILQNQGN (SEQ ID NO: 75)
- [0226]
- [0227] (이텔릭체: 핵산 편집 도메인; 밑줄: 세포질 국제화 신호)
- [0228] 그린 원숭이 APOBEC-3G:
- MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEYVKT KDPSGPPLDANIFQKG
LYPEAKDHPKMFHWFHWRQLHRDQEYEVWYVSWSPCTRCANSVATFLAEDPKV
 TLTIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQG
 KPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNKPPWVSGQRETYLCKYVE
 RSHNDTWVLLNQHRGFLRNQAPDRHGFPGKGRHAELCFDLIPFWKLDLDDQQYRVTCFT
 SWSPCFSCAQKMAKFISNNKHVSLCIFAARIYDDQGRQCQGLRTLHRDGAKIAVMNY
 SEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 76)
- [0229]
- [0230] (이텔릭체: 핵산 편집 도메인; 밑줄: 세포질 국제화 신호)
- [0231] 인간 APOBEC-3G:
- MKPHFRNTVERMYRDTFSYNFYNNRPILSRRNTVWLCYEYVKTGKPSRPPLDAKIFRGQ
VYSELKYHPEMRFFHWFSKWRKLHRDQEYEVWYISWSPCTKCTRDMATFLAEDPKV
 TLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRE
 LFEPWNNLPKYYILLHIMLGEILRHSMPTFTFNFNNEPWWVRGRHETYLCEVERM
 HNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFDLVIPFWKLDLDDQQYRVTCFTS
 WSPCFSCAQEMAKFISNNKHVSLCIFTARIYDDQGRQCQGLRTLAEGAKISIMTYSE
 FKHCWDTFVDHQGCPFPQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 77)
- [0232]
- [0233] (이텔릭체: 핵산 편집 도메인; 밑줄: 세포질 국제화 신호)

- [0234] 인간 APOBEC-3F:
- MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDKIFRGQ
VYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWVSWTPCPDCVAKLAEFLAEHPNVTL
TISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPW
YKFDDNYAFLHRTLKEILRNPMYEPHIFYFHFKNLRKAYGRNESWLCFTMEVVK
HHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNIEVTWYTSWSPCECA
GEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCW
ENFVYNDDEPFKPKWGLKYNFLFLDSKLQEILE (SEQ ID NO: 78)
- [0235] (이텔릭체: 핵산 편집 도메인)
- [0236] 인간 APOBEC-3B:
- MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFR
GQVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWVSWTPCPDCVAKLAEFLSEHPN
VTLTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQF
MPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRQRQTYLCYEVERLD
NGTWVLMQDQMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFIWS
PCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTY
DEFEYCWDTFVYRQGCQPFQPDWGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 79)
- [0238] (이텔릭체: 핵산 편집 도메인)
- [0239] 래트 APOBEC-3:
- MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTIFYFHFKNVRYAWGRKN
NFLCYEVNGMDALPVPLRQGVFRKQGGHIAELCFIYWFHDKVLRVLSPMEEFKVT
WYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVH
VAAMDLPFCKKCWNKFVDNDGQPFPRPWRMLRINFSFYDCKLQEIFSRMNLLREDVF
YLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLE
KMRSMELSQVRITCYLTWSPCPNCARQLAFAFKDHPDLILRIYTSRLYFYWRKKFQK
GLCTLWRSIGIHVDVMDLPQFADCWTNFNPNQRPFRPWNELEKNSWRIQRRRLRIKE
SWGL (SEQ ID NO: 80)
- [0241] 소 APOBEC-3B:
- DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVL
FKQQFGNQPRVPAPYYRRKTYLCYQLKQRNDLTLDRCGFRNKKQRHAEIRFIDKINS
LDLNPSQSYKIICYITWSPCPNCANELVNFIITRNHLKLEIFASRLYFHWIKSFKMGLQ
DLQNAISVAVMTHTEFEDCWEQFVDNQSRPFQPDWKLEQYSASIRRLRILTAPI
(SEQ ID NO: 81)
- [0243]

- [0244] 침팬지 APOBEC-3B:
- MNPQIRNPMEMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFR
GQMYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWVSWTPCPDCVAKLAKFLAEH
PNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQP
FMPWYKFDDNYAFLHRTLKEIIRHLMDDPTFTFNFNNDPLVLRHQTLYLCYEVERLD
NGTWVLMQHMGLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFWIS
WSPCFSWGCGAQVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIM
TYDEFEYCWDTFVYRQGCQFQPDWGLEEHSQALSGRLRAILQVRASSLCMVPHRPPP
PPQSPGPCPLCSEPPLGSLPTGRPAPSLPFLLTASFSPPPASLPPLPSLSLSPGHLVPV
SFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG (SEQ ID NO: 82)
- [0245]
- [0246] 인간 APOBEC-3C:
- MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVF
RNQVDSETHCHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSN
VNLTIFTARLYYFQYPCYQEGRLSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKP
WKGLKTNFRLLKRRLRESLQ (SEQ ID NO: 83)
- [0247]
- [0248] (이텔릭체: 핵산 편집 도메인)
- [0249] 고릴라 APOBEC3C
- MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVF
RNQVDSETHCHAERCFLSWFCDDILSPNTNYQVTWYTSWSPCECAGEVAEFLARHSN
VNLTIFTARLYYFQDTDYQEGRLSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFK
PWKGLKYNFRFLKRRLQEILE (SEQ ID NO: 84)
- [0250]
- [0251] (이텔릭체: 핵산 편집 도메인)
- [0252] 인간 APOBEC-3A:
- MEASPASGRHLMDDPHFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLH
NQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFWISWSPCFSWGCGAGEVRAFLQ
ENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGC
PFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 85)
- [0253]
- [0254] (이텔릭체: 핵산 편집 도메인)
- [0255] 레서스 마카크 APOBEC-3A:
- MDGSPASRPRHLMDDPNTFTFNFNNDLSVRGRHQTLYLCYEVERLDNGTWVPMDERR
GFLCNKAKNVPCGDYGGHVELRFLCEVPSWQLDPAQTYRVTFWISWSPCFRRGCAGQ
VRVFLQENKHVRLRIFAARIYDYDPLYQEALRTLRLDAGAQVSIMTYEEFKHCWDTF
VDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 86)
- [0256]
- [0257] (이텔릭체: 핵산 편집 도메인)
- [0258] 소 APOBEC-3A:
- MDEYTFTENFNNGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCH
AELYFLGKIHSWNLDNRQHRYLTCFISWSPCYDCAQKLTTFLKENHHISLHILASRIYTH
NRFGCHQSGLCCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQAL
CTELQAILKTQQN (SEQ ID NO: 87)
- [0259]

- [0260] (이텔릭체: 핵산 편집 도메인)
- [0261] 인간 APOBEC-3H:
- MALLTAETFERLQFNNKRRLRRPYYPKALLCYQLTPQNGSTPTRGYFENKKKCHAEI
CFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWC
KPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRA
IKRRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 88)
- [0262]
- [0263] (이텔릭체: 핵산 편집 도메인)
- [0264] 레서스 마카크 APOBEC-3H:
- MALLTAKTFSLQFNNKRRVNKPYYPKALLCYQLTPQNGSTPTRGHLKNKKKDHAE
IRFINKIKSMGLDETQCYQVTCYLTWSPCSCAGELVDFIKAHRHLNLRIFASRLYYH
WRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNS
QAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR (SEQ ID NO: 89)
- [0265]
- [0266] 인간 APOBEC-3D:
- MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFR
GPVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWVFSWNPCLPCVV
KVTKFLAEHPNVTLTISAARLYYYRDRDRWRVLLRLHKAGARVKIMDYEDFAYCW
ENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHKNNLLKACG
RNESWLCFTMEVTKHHSVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNY
EVTWYTSWSPCECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGAS
VKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRRLREILQ (SEQ ID NO:
90)
- [0267]
- [0268] (이텔릭체: 핵산 편집 도메인)
- [0269] 인간 APOBEC-1:
- MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACCLYEIKWGMSRKiWRSSGKN
TTNHVEVNFIEKFTSERDFHPSMSCSITWFLSWSPCWECQAIREFLSRHPGVTLVIIYV
ARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQY
PPLWMMLYALELHCILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLI
HPSVAWR (SEQ ID NO: 91)
- [0270]
- [0271] 마우스 APOBEC-1:
- MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQN
TSNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIA
RLYHHTDQNRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPSPNEAYWPRYPHL
WVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK
(SEQ ID NO: 92)
- [0272]
- [0273] 래트 APOBEC-1:
- MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT
NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR
LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYPSPNEAHWPYPHLLW
VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHLLWATGLK (SEQ
ID NO: 93)
- [0274]

- [0275] 인간 APOBEC-2:
- MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRN
VEYSSGRNKTFLCYVVEAQGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPA
LRYNVTWYVSSSPCAACADRIKTLSTKNLRLLLVGRLFMWEEPEIQAALKKLKE
AGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK
(SEQ ID NO: 94)
- [0276]
- [0277] 마우스 APOBEC-2:
- MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFR
NVEYSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAAEEAFFNTILPAFDP
ALKYNVTWYVSSSPCAACADRILKTLSTKNLRLLLVSRLFMWEEPEVQAALKKLK
EAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK
(SEQ ID NO: 95)
- [0278]
- [0279] 래트 APOBEC-2:
- MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFR
NVEYSSGRNKTFLCYVVEAQSKGGQVQATQGYLEDEHAGAHAAEEAFFNTILPAFDP
ALKYNVTWYVSSSPCAACADRILKTLSTKNLRLLLVSRLFMWEEPEVQAALKKLK
EAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK
(SEQ ID NO: 96)
- [0280]
- [0281] 소 APOBEC-2:
- MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRN
VEYSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAAEEAFFNSIMPTFDPA
LRYMVTWYVSSSPCAACADRIVKTLNKTKNLRLLLVGRLFMWEEPEIQAALRKLKE
AGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK (SEQ
ID NO: 97)
- [0282]
- [0283] 페트로미존 마리누스 CDA1 (pmCDA1)
- MTDAEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFWGYAVNK
PQSGTERGIHAEIFSIRKVEEYLRDNPQQFTINWYSSWSPCADCAEKILEWYNQELRG
NGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHQ
LNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 98)
- [0284]
- [0285] 인간 APOBEC3G D316R_D317R
- MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSELKYHPEMRFFHWFWSKWRKLHRDQEYEV TWYISWSPCTKCTRDMATFLAEDP
KVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYS
QRELFEPWNNLPKYIILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEV
ERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFDVIPFWKLDLDQDYRV
TCFTSWSPCFSCAQEMAKFISKKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISI
MTYSEFKHCWDTFVDHQGCPFPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO:
99)
- [0286]

[0287] 인간 APOBEC3G 쇠 A

MDPPTFTFNFNNEPWVRGRHETLYLCYEVRMHNDTWVLLNQRRGFLCNQAPHKHG
FLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKKNHVS LCI
FTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLD
EHSQDLSGRLRAILQ (SEQ ID NO: 100)

[0288]

[0289] 인간 APOBEC3G 쇠 A D120R_D121R

MDPPTFTFNFNNEPWVRGRHETLYLCYEVRMHNDTWVLLNQRRGFLCNQAPHKHG
FLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKKNHVS LCI
FTARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDE
HSQDLSGRLRAILQ (SEQ ID NO: 101)

[0290]

[0291] 염기 편집제의 편집 윈도우를 조정하는 데아미나제 도메인

[0292] 본 개시내용의 일부 측면은, 예를 들어 데아미나제 도메인에서의 점 돌연변이를 생성함으로써 본원에 제공된 융합 단백질 중 임의의 것의 데아미나제 도메인 촉매적 활성을 조정하는 것이 융합 단백질 (예를 들어, 염기 편집제)의 프로세싱성에 영향을 미친다는 인식에 기초한다. 예를 들어, 염기 편집 융합 단백질 내의 데아미나제 도메인의 촉매적 활성을 감소시키지만 제거하지는 않는 돌연변이는 데아미나제 도메인이 표적 잔기에 인접한 잔기의 탈아미노화를 촉매함으로써 탈아미노화 윈도우를 좁힐 가능성을 적게 할 수 있다. 탈아미노화 윈도우를 좁히는 능력은 특정 표적 잔기의 인접한 잔기의 원하지 않는 탈아미노화를 방지할 수 있으며, 이는 오프-타겟 효과를 감소시키거나 방지할 수 있다.

[0293]

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

[0294]

일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 93)의 H121X, H122X, R126X, R126X, R118X, W90X, W90X 및 R132X로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함하고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 rAPOBEC1 (서열식별번호: 93)의 H121R, H122R, R126A, R126E, R118A, W90A, W90Y 및 R132E로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다.

[0295]

일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 77)의 D316X, D317X, R320X, R320X, R313X, W285X, W285X, R326X로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함하고, 여기서 X는 임의의 아미노산이다. 일부 실시양태에서, 본원에 제공된 융합 단백질 중 임의의 것은 hAPOBEC3G (서열식별번호: 77)의 D316R, D317R, R320A, R320E, R313A, W285A, W285Y, R326E로 이루어진 군으로부터 선택된 1개 이상의 돌연변이, 또는 또 다른 APOBEC 데아미나제에서의 1개 이상의 상응하는 돌연변이를 포함하는 APOBEC 데아미나제를 포함한다.

[0297]

[0298]

- 89 -

- [0299] 본 개시내용의 일부 측면은 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp), 시티딘 데아미나제, 및 우라실 결합 단백질 (UBP)을 포함하는 융합 단백질을 제공한다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질은 염기 편집제이다. 일부 실시양태에서, UBP는 우라실 변형 효소이다. 일부 실시양태에서, UBP는 우라실 염기 절제 효소이다. 일부 실시양태에서, UBP는 우라실 DNA 글리코실라제이다. 일부 실시양태에서, UBP는 본원에 제공된 임의의 우라실 결합 단백질이다. 예를 들어, UBP는 UDG, UdgX, UdgX*, UdgX_On 또는 SMUG1일 수 있다. 일부 실시양태에서, UBP는 우라실 결합 단백질, 우라실 염기 절제 효소 또는 우라실 DNA 글리코실라제 (UDG) 효소에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, UBP는 본원에 제공된 임의의 우라실 결합 단백질에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 예를 들어, UBP는 서열식별번호: 48-53 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함할 수 있다. 일부 실시양태에서, UBP는 서열식별번호: 48-53 중 어느 하나의 아미노산 서열을 포함한다.
- [0300] 일부 실시양태에서, napDNAbp는 Cas9 도메인, Cpf1 도메인, CasX 도메인, CasY 도메인, C2c1 도메인, C2c2 도메인, C2c3 도메인 또는 아르코노트 도메인이다. 일부 실시양태에서, napDNAbp는 본원에 제공된 임의의 napDNAbp이다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질의 napDNAbp는 Cas9 도메인이다. Cas9 도메인은 본원에 제공된 Cas9 도메인 또는 Cas9 단백질 (예를 들어, dCas9 또는 nCas9) 중 임의의 것일 수 있다. 일부 실시양태에서, 본원에 제공된 임의의 Cas9 도메인 또는 Cas9 단백질 (예를 들어, dCas9 또는 nCas9)은 본원에 제공된 임의의 시티딘 데아미나제와 융합될 수 있다. 일부 실시양태에서, 융합 단백질은 하기 구조를 포함한다:
- [0301] NH_2 -[시티딘 데아미나제]-[napDNAbp]-[UBP]-COOH;
- [0302] NH_2 -[시티딘 데아미나제]-[UBP]-[napDNAbp]-COOH;
- [0303] NH_2 -[UBP]-[시티딘 데아미나제]-[napDNAbp]-COOH;
- [0304] NH_2 -[UBP]-[napDNAbp]-[시티딘 데아미나제]-COOH;
- [0305] NH_2 -[napDNAbp]-[UBP]-[시티딘 데아미나제]-COOH; 또는
- [0306] NH_2 -[napDNAbp]-[시티딘 데아미나제]-[UBP]-COOH.
- [0307] 일부 실시양태에서, 시티딘 데아미나제, napDNAbp (예를 들어, Cas9 도메인) 및 UBP를 포함하는 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시양태에서, 시티딘 데아미나제 도메인과 napDNAbp 사이에 링커가 존재한다. 일부 실시양태에서, 시티딘 데아미나제 도메인과 UBP 사이에 링커가 존재한다. 일부 실시양태에서, napDNAbp와 UBP 사이에 링커가 존재한다. 일부 실시양태에서, 상기 일반적 아키텍처에 사용된 "-"는 임의적인 링커 서열의 존재를 나타낸다. 일부 실시양태에서, 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 UBP, 및/또는 napDNAbp와 UBP는 본원에 제공된 임의의 링커를 통해 융합된다. 예를 들어, 일부 실시양태에서 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 UBP, 및/또는 napDNAbp와 UBP는 하기 "링커" 표제의 섹션에 제공된 링커 중 임의의 것을 통해 융합된다. 일부 실시양태에서, 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 UBP, 및/또는 napDNAbp와 UBP는 1 내지 200개 아미노산을 포함하는 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 UBP, 및/또는 napDNAbp와 UBP는 1 내지 5, 1 내지 10, 1 내지 20, 1 내지 30, 1 내지 40, 1 내지 50, 1 내지 60, 1 내지 80, 1 내지 100, 1 내지 150, 1 내지 200, 5 내지 10, 5 내지 20, 5 내지 30, 5 내지 40, 5 내지 60, 5 내지 80, 5 내지 100, 5 내지 150, 5 내지 200, 10 내지 20, 10 내지 30, 10 내지 40, 10 내지 50, 10 내지 60, 10 내지 80, 10 내지 100, 10 내지 150, 10 내지 200, 20 내지 30, 20 내지 40, 20 내지 50, 20 내지 60, 20 내지 80, 20 내지 100, 20 내지 150, 20 내지 200, 30 내지 40, 30 내지 50, 30 내지 60, 30 내지 80, 30 내지 100, 30 내지 150, 30 내지 200, 40 내지 50, 40 내지 60, 40 내지 80, 40 내지 100, 40 내지 150, 40 내지 200, 50 내지 60, 50 내지 80, 50 내지 100, 50 내지 150, 50 내지 200, 60 내지 80, 60 내지 100, 60 내지 150, 60 내지 200, 80 내지 100, 80 내지 150, 80 내지 200, 100 내지 150, 100 내지 200, 또는 150 내지 200개 아미노산 길이를 포함하는 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 UBP, 및/또는 napDNAbp와 UBP는 4, 16, 24, 32, 91 또는 104개 아미노산 길이를 포함하는 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 UBP, 및/또는 napDNAbp와 UBP는 SGSETPGTSESATPES (서열식

별번호: 102), SGG (서열식별번호: 103), SGGSSGSETPGTSESATPESGG (서열식별번호: 107), SGGSSGSSGSETPGTSESATPESGGSSGG (서열식별번호: 108), GGGGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGGSGS (서열식별번호: 109), 또는 SGGSSGSSGG (서열식별번호: 120)의 아미노산 서열을 포함하는 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 테아미나제와 napDNAbp, 시티딘 테아미나제와 UBP, 및/또는 napDNAbp와 UBP는 XTEN 링커로도 지칭될 수 있는, 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 102)를 포함하는 링커를 통해 융합된다.

[0308] 뉴클레아제 프로그램가능 DNA 결합 단백질 (napDNAbp), 시티딘 테아미나제 및 핵산 폴리머라제 (NAP) 도메인을 포함하는 융합 단백질

[0309] 본 개시내용의 일부 측면은 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp), 시티딘 테아미나제 및 핵산 폴리머라제 (NAP) 도메인을 포함하는 융합 단백질을 제공한다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질은 염기 편집제이다. 일부 실시양태에서, NAP는 진핵 핵산 폴리머라제이다. 일부 실시양태에서, NAP는 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 트랜스레전 폴리머라제 활성을 갖는다. 일부 실시양태에서, NAP는 트랜스레전 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파 또는 폴리머라제 에타이다. 일부 실시양태에서, NAP는 진핵 폴리머라제 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, 뮤 또는 누이다. 일부 실시양태에서, NAP는 핵산 폴리머라제 (예를 들어, 트랜스레전 DNA 폴리머라제)에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, NAP는 본원에 제공된 임의의 핵산 폴리머라제에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 예를 들어, NAP는 서열식별번호: 54-64 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함할 수 있다. 일부 실시양태에서, NAP는 서열식별번호: 54-64 중 어느 하나의 아미노산 서열을 포함한다.

[0310] 일부 실시양태에서, napDNAbp는 Cas9 도메인, Cpf1 도메인, CasX 도메인, CasY 도메인, C2c1 도메인, C2c2 도메인, C2c3 도메인 또는 아르코노트 도메인이다. 일부 실시양태에서, napDNAbp는 본원에 제공된 임의의 napDNAbp이다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질의 napDNAbp는 Cas9 도메인이다. Cas9 도메인은 본원에 제공된 Cas9 도메인 또는 Cas9 단백질 (예를 들어, dCas9 또는 nCas9) 중 임의의 것일 수 있다. 일부 실시양태에서, 본원에 제공된 임의의 Cas9 도메인 또는 Cas9 단백질 (예를 들어, dCas9 또는 nCas9)은 본원에 제공된 임의의 시티딘 테아미나제와 융합될 수 있다. 일부 실시양태에서, 융합 단백질은 하기 구조를 포함한다:

[0311] NH_2 -[시티딘 테아미나제]-[napDNAbp]-[NAP]-COOH;

[0312] NH_2 -[시티딘 테아미나제]-[NAP]-[napDNAbp]-COOH;

[0313] NH_2 -[NAP]-[시티딘 테아미나제]-[napDNAbp]-COOH;

[0314] NH_2 -[NAP]-[napDNAbp]-[시티딘 테아미나제]-COOH;

[0315] NH_2 -[napDNAbp]-[NAP]-[시티딘 테아미나제]-COOH; 또는

[0316] NH_2 -[napDNAbp]-[시티딘 테아미나제]-[NAP]-COOH.

[0317] 일부 실시양태에서, 시티딘 테아미나제, napDNAbp (예를 들어, Cas9 도메인) 및 NAP를 포함하는 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시양태에서, 시티딘 테아미나제 도메인과 napDNAbp 사이에 링커가 존재한다. 일부 실시양태에서, 시티딘 테아미나제 도메인과 NAP 사이에 링커가 존재한다. 일부 실시양태에서, napDNAbp와 NAP 사이에 링커가 존재한다. 일부 실시양태에서, 상기 일반적 아키텍처에 사용된 "-"는 임의적인 링커 서열의 존재를 나타낸다. 일부 실시양태에서, 시티딘 테아미나제와 napDNAbp, 시티딘 테아미나제와 NAP, 및/또는 napDNAbp와 NAP는 본원에 제공된 임의의 링커를 통해 융합된다. 예를 들어, 일부 실시양태에서 시티딘 테아미나제와 napDNAbp, 시티딘 테아미나제와 NAP, 및/또는 napDNAbp와 NAP는 하기 "링커" 표제의 섹션에 제공된 임의의 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 테아미나제와 napDNAbp, 시티딘 테아미나제와 NAP, 및/또는 napDNAbp와 NAP는 1 내지 200개 아미노산을 포함하는 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 테아미나제와 napDNAbp, 시티딘 테아미나제와 NAP, 및/또는 napDNAbp와 NAP는 1 내지 5, 1 내지 10, 1

내지 20, 1 내지 30, 1 내지 40, 1 내지 50, 1 내지 60, 1 내지 80, 1 내지 100, 1 내지 150, 1 내지 200, 5 내지 10, 5 내지 20, 5 내지 30, 5 내지 40, 5 내지 60, 5 내지 80, 5 내지 100, 5 내지 150, 5 내지 200, 10 내지 20, 10 내지 30, 10 내지 40, 10 내지 50, 10 내지 60, 10 내지 80, 10 내지 100, 10 내지 150, 10 내지 200, 20 내지 30, 20 내지 40, 20 내지 50, 20 내지 60, 20 내지 80, 20 내지 100, 20 내지 150, 20 내지 200, 30 내지 40, 30 내지 50, 30 내지 60, 30 내지 80, 30 내지 100, 30 내지 150, 30 내지 200, 40 내지 50, 40 내지 60, 40 내지 80, 40 내지 100, 40 내지 150, 40 내지 200, 50 내지 60 50 내지 80, 50 내지 100, 50 내지 150, 50 내지 200, 60 내지 80, 60 내지 100, 60 내지 150, 60 내지 200, 80 내지 100, 80 내지 150, 80 내지 200, 100 내지 150, 100 내지 200, 또는 150 내지 200개 아미노산 길이를 포함하는 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 NAP, 및/또는 napDNAbp와 NAP는 4, 16, 32 또는 104개 아미노산 길이를 포함하는 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 NAP, 및/또는 napDNAbp와 NAP는 SGSETPGTSESATPES (서열식별번호: 102), SGGSS (서열식별번호: 103), SGGSSGSETPGTSESATPES (서열식별번호: 107), SGGSSGSSGSETPGTSESATPES (서열식별번호: 108), GGGSSGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGS (서열식별번호: 109), 또는 SGGSSGSSGSS (서열식별번호: 120)의 아미노산 서열을 포함하는 링커를 통해 융합된다. 일부 실시양태에서, 시티딘 데아미나제와 napDNAbp, 시티딘 데아미나제와 NAP, 및/또는 napDNAbp와 NAP는 XTEN 링커로도 치환될 수 있는, 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 102)를 포함하는 링커를 통해 융합된다.

[0318] 뉴클레아제 프로그램가능 DNA 결합 단백질 (napDNAbp), 시티딘 데아미나제, 우라실 결합 단백질 (UBP) 및 핵산 폴리머라제 (NAP) 도메인을 포함하는 융합 단백질

[0319] 본 개시내용의 일부 측면은 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp), 시티딘 데아미나제, 우라실 결합 단백질 (UBP) 및 핵산 폴리머라제 (NAP) 도메인을 포함하는 융합 단백질을 제공한다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질은 염기 편집제이다. 일부 실시양태에서, NAP는 진핵 핵산 폴리머라제이다. 일부 실시양태에서, NAP는 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 트랜스레전 폴리머라제 활성을 갖는다. 일부 실시양태에서, NAP는 트랜스레전 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파 또는 폴리머라제 에타이다. 일부 실시양태에서, NAP는 진핵 폴리머라제 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, 뮤 또는 누이다. 일부 실시양태에서, NAP는 핵산 폴리머라제 (예를 들어, 트랜스레전 DNA 폴리머라제)에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, NAP는 본원에 제공된 임의의 핵산 폴리머라제에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 예를 들어, NAP는 서열식별번호: 54-64 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함할 수 있다. 일부 실시양태에서, NAP는 서열식별번호: 54-64 중 어느 하나의 아미노산 서열을 포함한다.

[0320] 일부 실시양태에서, UBP는 우라실 변형 효소이다. 일부 실시양태에서, UBP는 우라실 염기 절제 효소이다. 일부 실시양태에서, UBP는 우라실 DNA 글리코실라제이다. 일부 실시양태에서, UBP는 본원에 제공된 임의의 우라실 결합 단백질이다. 예를 들어, UBP는 UDG, UdgX, UdgX*, UdgX_{On} 또는 SMUG1일 수 있다. 일부 실시양태에서, UBP는 우라실 결합 단백질, 우라실 염기 절제 효소 또는 우라실 DNA 글리코실라제 (UDG) 효소에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, UBP는 본원에 제공된 임의의 우라실 결합 단백질에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 예를 들어, UBP는 서열식별번호: 48-53 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함할 수 있다. 일부 실시양태에서, UBP는 서열식별번호: 48-53 중 어느 하나의 아미노산 서열을 포함한다.

[0321] 일부 실시양태에서, napDNAbp는 Cas9 도메인, Cpf1 도메인, CasX 도메인, CasY 도메인, C2c1 도메인, C2c2 도메인, C2c3 도메인 또는 아르코노트 도메인이다. 일부 실시양태에서, napDNAbp는 본원에 제공된 임의의 napDNAbp이다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질의 napDNAbp는 Cas9 도메인이다. Cas9 도메인은 본원에 제공된 Cas9 도메인 또는 Cas9 단백질 (예를 들어, dCas9 또는 nCas9) 중 임의의 것일 수 있다. 일부 실시양태에서, 본원에 제공된 임의의 Cas9 도메인 또는 Cas9 단백질 (예를 들어, dCas9 또는 nCas9)은 본원에 제공된 임의의 시티딘 데아미나제와 융합될 수 있다. 일부 실시양태에서, 융합 단백질은 하기 구조를 포함한다:

- [0322] NH_2 -[NAP]-[시티딘 데아미나제]-[napDNAbp]-[UBP]-COOH;
- [0323] NH_2 -[시티딘 데아미나제]-[NAP]-[napDNAbp]-[UBP]-COOH;
- [0324] NH_2 -[시티딘 데아미나제]-[napDNAbp]-[NAP]-[UBP]-COOH;
- [0325] NH_2 -[시티딘 데아미나제]-[napDNAbp]-[UBP]-[NAP]-COOH;
- [0326] NH_2 -[NAP]-[시티딘 데아미나제]-[UBP]-[napDNAbp]-COOH;
- [0327] NH_2 -[시티딘 데아미나제]-[NAP]-[UBP]-[napDNAbp]-COOH;
- [0328] NH_2 -[시티딘 데아미나제]-[UBP]-[NAP]-[napDNAbp]-COOH;
- [0329] NH_2 -[시티딘 데아미나제]-[UBP]-[napDNAbp]-[NAP]-COOH;
- [0330] NH_2 -[NAP]-[UBP]-[시티딘 데아미나제]-[napDNAbp]-COOH;
- [0331] NH_2 -[UBP]-[NAP]-[시티딘 데아미나제]-[napDNAbp]-COOH;
- [0332] NH_2 -[UBP]-[시티딘 데아미나제]-[NAP]-[napDNAbp]-COOH;
- [0333] NH_2 -[UBP]-[시티딘 데아미나제]-[napDNAbp]-[NAP]-COOH;
- [0334] NH_2 -[NAP]-[UBP]-[napDNAbp]-[시티딘 데아미나제]-COOH;
- [0335] NH_2 -[UBP]-[NAP]-[napDNAbp]-[시티딘 데아미나제]-COOH;
- [0336] NH_2 -[UBP]-[napDNAbp]-[NAP]-[시티딘 데아미나제]-COOH;
- [0337] NH_2 -[UBP]-[napDNAbp]-[시티딘 데아미나제]-[NAP]-COOH;
- [0338] NH_2 -[NAP]-[napDNAbp]-[UBP]-[시티딘 데아미나제]-COOH;
- [0339] NH_2 -[napDNAbp]-[NAP]-[UBP]-[시티딘 데아미나제]-COOH;
- [0340] NH_2 -[napDNAbp]-[UBP]-[NAP]-[시티딘 데아미나제]-COOH;
- [0341] NH_2 -[napDNAbp]-[UBP]-[시티딘 데아미나제]-[NAP]-COOH;
- [0342] NH_2 -[NAP]-[napDNAbp]-[시티딘 데아미나제]-[UBP]-COOH;
- [0343] NH_2 -[napDNAbp]-[NAP]-[시티딘 데아미나제]-[UBP]-COOH;
- [0344] NH_2 -[napDNAbp]-[시티딘 데아미나제]-[NAP]-[UBP]-COOH; 또는
- [0345] NH_2 -[napDNAbp]-[시티딘 데아미나제]-[UBP]-[NAP]-COOH.
- [0346] 일부 실시양태에서, 시티딘 데아미나제, napDNAbp (예를 들어, Cas9 도메인), UBP 및 NAP를 포함하는 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시양태에서, 시티딘 데아미나제 도메인과 napDNAbp, NAP 및/또는 UBP 사이에 링커가 존재한다. 일부 실시양태에서, napDNAbp와 시티딘 데아미나제 도메인, NAP 및/또는 UBP 사이에 링커가 존재한다. 일부 실시양태에서, NAP와 시티딘 데아미나제, napDNAbp 및/또는 UBP 사이에 링커가 존재한다. 일부 실시양태에서, UBP와 시티딘 데아미나제, napDNAbp 및 NAP 사이에 링커가 존재한다. 일부 실시양태에서, 상기 일반적 아키텍처에 사용된 "-"는 임의적인 링커 서열의 존재를 나타낸다. 일부 실시양태에서, 링커는 본원, 예를 들어 "링커" 표제의 섹션에 제공된 링커 중 임의의 것이다. 일부 실시양태에서, 링커는 1 내지 200개 아미노산을 포함한다. 일부 실시양태에서, 링커는 1 내지 5, 1 내지 10, 1 내지 20, 1 내지 30, 1 내지 40, 1 내지 50, 1 내지 60, 1 내지 80, 1 내지 100, 1 내지 150, 1 내지 200, 5 내지 10, 5 내지 20, 5 내지 30, 5 내지 40, 5 내지 60, 5 내지 80, 5 내지 100, 5 내지 150, 5 내지 200, 10 내지 20, 10 내지 30,

10 내지 40, 10 내지 50, 10 내지 60, 10 내지 80, 10 내지 100, 10 내지 150, 10 내지 200, 20 내지 30, 20 내지 40, 20 내지 50, 20 내지 60, 20 내지 80, 20 내지 100, 20 내지 150, 20 내지 200, 30 내지 40, 30 내지 50, 30 내지 60, 30 내지 80, 30 내지 100, 30 내지 150, 30 내지 200, 40 내지 50, 40 내지 60, 40 내지 80, 40 내지 100, 40 내지 150, 40 내지 200, 50 내지 60 50 내지 80, 50 내지 100, 50 내지 150, 50 내지 200, 60 내지 80, 60 내지 100, 60 내지 150, 60 내지 200, 80 내지 100, 80 내지 150, 80 내지 200, 100 내지 150, 100 내지 200, 또는 150 내지 200개 아미노산 길이를 포함한다. 일부 실시양태에서, 링커는 4, 16, 32 또는 104개 아미노산 길이를 포함한다. 일부 실시양태에서, 링커는 SGSETPGTSESATPES (서열식별번호: 102), SGG (서열식별번호: 103), SGGSSGSETPGTSESATPESSGGS (서열식별번호: 107), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (서열식별번호: 108), GGGSGSPGPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGS (서열식별번호: 109), 또는 SGGSGGSGGS (서열식별번호: 120)의 아미노산 서열을 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 102)를 포함하며, 이는 또한 XTEN 링커로 지칭될 수 있다.

[0347] 뉴클레아제 프로그램가능 DNA 결합 단백질 (napDNAbp) 및 염기 절제 효소 (BEE)를 포함하는 융합 단백질

[0348] 본 개시내용의 일부 측면은 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp) 및 염기 절제 효소를 포함하는 융합 단백질을 제공한다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질은 염기 편집제이다. 일부 실시양태에서, 염기 절제 효소 (BEE)는 시토신, 티민, 아데닌, 구아닌 또는 우라실 염기 절제 효소이다. 일부 실시양태에서, 염기 절제 효소 (BEE)는 시토신 염기 절제 효소이다. 일부 실시양태에서, BEE는 티민 염기 절제 효소이다. 일부 실시양태에서, 염기 절제 효소는 자연 발생 BEE에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 염기 절제 효소는 서열식별번호: 65-66 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, 염기 절제 효소는 서열식별번호: 65-66 중 어느 하나의 아미노산 서열을 포함한다.

[0349] 일부 실시양태에서, napDNAbp는 Cas9 도메인, Cpf1 도메인, CasX 도메인, CasY 도메인, C2c1 도메인, C2c2 도메인, C2c3 도메인 또는 아르코노트 도메인이다. 일부 실시양태에서, napDNAbp는 본원에 제공된 임의의 napDNAbp이다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질의 napDNAbp는 Cas9 도메인이다. Cas9 도메인은 본원에 제공된 Cas9 도메인 또는 Cas9 단백질 (예를 들어, dCas9 또는 nCas9) 중 임의의 것일 수 있다. 일부 실시양태에서, 본원에 제공된 임의의 Cas9 도메인 또는 Cas9 단백질 (예를 들어, dCas9 또는 nCas9)은 본원에 제공된 임의의 시티딘 데아미나제와 융합될 수 있다. 일부 실시양태에서, 융합 단백질은 하기 구조를 포함한다:

[0350] NH_2 -[BEE]-[napDNAbp]-COOH; 또는

[0351] NH_2 -[napDNAbp]-[BEE]-COOH.

[0352] 일부 실시양태에서, 융합 단백질은 핵산 폴리머라제 (NAP)를 추가로 포함한다. 일부 실시양태에서, NAP는 진핵 핵산 폴리머라제이다. 일부 실시양태에서, NAP는 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 트랜스레전 폴리머라제 활성을 갖는다. 일부 실시양태에서, NAP는 트랜스레전 DNA 폴리머라제이다. 일부 실시양태에서, NAP는 Rev7, Rev1 복합체, 폴리머라제 이오타, 폴리머라제 카파 또는 폴리머라제 에타이다. 일부 실시양태에서, NAP는 진핵 폴리머라제 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, 뮤 또는 누이다. 일부 실시양태에서, NAP는 핵산 폴리머라제 (예를 들어, 트랜스레전 DNA 폴리머라제)에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시양태에서, NAP는 본원에 제공된 임의의 핵산 폴리머라제에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함한다. 예를 들어, NAP는 서열식별번호: 54-64 중 어느 하나에 대해 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 동일한 아미노산 서열을 포함할 수 있다. 일부 실시양태에서, NAP는 서열식별번호: 54-64 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, 융합 단백질은 하기 구조를 포함한다:

[0353] NH_2 -[BEE]-[napDNAbp]-[NAP]-COOH;

[0354] NH_2 -[BEE]-[NAP]-[napDNAbp]-COOH;

[0355] NH_2 -[NAP]-[BEE]-[napDNAbp]-COOH;

[0356] NH_2 -[NAP]-[napDNAbp]-[BEE]-COOH;

[0357] NH_2 -[napDNAbp]-[NAP]-[BEE]-COOH; 또는

[0358] NH_2 -[napDNAbp]-[BEE]-[NAP]-COOH.

[0359] 일부 실시양태에서, napDNAbp (예를 들어, Cas9 도메인) 및 BEE를 포함하는 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시양태에서, napDNAbp (예를 들어, Cas9 도메인), BEE 및 NAP를 포함하는 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시양태에서, napDNAbp와 BEE 사이에 링커가 존재한다. 일부 실시양태에서, BEE와 NAP 및/또는 napDNAbp 사이에 링커가 존재한다. 일부 실시양태에서, NAP와 BEE 및/또는 napDNAbp 사이에 링커가 존재한다. 일부 실시양태에서, napDNAbp와 BEE 및/또는 NAP 사이에 링커가 존재한다. 일부 실시양태에서, 상기 일반적 아키텍처에 사용된 "-"는 임의적인 링커 서열의 존재를 나타낸다. 일부 실시양태에서, 링커는 본원, 예를 들어 "링커" 표제의 섹션에 제공된 링커 중 임의의 것이다. 일부 실시양태에서, 링커는 1 내지 200개 아미노산을 포함한다. 일부 실시양태에서, 링커는 1 내지 5, 1 내지 10, 1 내지 20, 1 내지 30, 1 내지 40, 1 내지 50, 1 내지 60, 1 내지 80, 1 내지 100, 1 내지 150, 1 내지 200, 5 내지 10, 5 내지 20, 5 내지 30, 5 내지 40, 5 내지 60, 5 내지 80, 5 내지 100, 5 내지 150, 5 내지 200, 10 내지 20, 10 내지 30, 10 내지 40, 10 내지 50, 10 내지 60, 10 내지 80, 10 내지 100, 10 내지 150, 10 내지 200, 20 내지 30, 20 내지 40, 20 내지 50, 20 내지 60, 20 내지 80, 20 내지 100, 20 내지 150, 20 내지 200, 30 내지 40, 30 내지 50, 30 내지 60, 30 내지 80, 30 내지 100, 30 내지 150, 30 내지 200, 40 내지 50, 40 내지 60, 40 내지 80, 40 내지 100, 40 내지 150, 40 내지 200, 50 내지 60, 50 내지 80, 50 내지 100, 50 내지 150, 50 내지 200, 60 내지 80, 60 내지 100, 60 내지 150, 60 내지 200, 80 내지 100, 80 내지 150, 80 내지 200, 100 내지 150, 100 내지 200, 또는 150 내지 200개 아미노산 길이를 포함한다. 일부 실시양태에서, 링커는 4, 16, 32 또는 104개 아미노산 길이를 포함한다. 일부 실시양태에서, 링커는 SGSETPGTSESATPES (서열식별번호: 102), SGGG (서열식별번호: 103), SGGSSGSETPGTSESATPESGGG (서열식별번호: 107), SGGSSGGSSGSETPGTSESATPESGGSSGGG (서열식별번호: 108), GGGSGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGG (서열식별번호: 109), 또는 SGGSGGSGG (서열식별번호: 120)의 아미노산 서열을 포함한다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 102)를 포함하며, 이는 또한 XTEN 링커로 지칭될 수 있다.

[0360] 핵 국제화 서열 (NLS)을 포함하는 융합 단백질

[0361] 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질은 1개 이상의 핵 표적화 서열, 예를 들어 핵 국제화 서열 (NLS)을 추가로 포함한다. 일부 실시양태에서, NLS는 NLS를 포함하는 단백질의 세포 핵 내로의 유입 (예를 들어, 핵 수송에 의한 것)을 용이하게 하는 아미노산 서열을 포함한다. 일부 실시양태에서, 본원에 제공된 임의의 융합 단백질은 핵 국제화 서열 (NLS)을 추가로 포함한다. 일부 실시양태에서, NLS는 융합 단백질의 N-말단에 융합된다. 일부 실시양태에서, NLS는 융합 단백질의 C-말단에 융합된다. 일부 실시양태에서, NLS는 napDNAbp의 N-말단에 융합된다. 일부 실시양태에서, NLS는 napDNAbp의 C-말단에 융합된다. 일부 실시양태에서, NLS는 NAP의 N-말단에 융합된다. 일부 실시양태에서, NLS는 NAP의 C-말단에 융합된다. 일부 실시양태에서, NLS는 시티딘 데아미나제의 N-말단에 융합된다. 일부 실시양태에서, NLS는 시티딘 데아미나제의 C-말단에 융합된다. 일부 실시양태에서, NLS는 UBP의 N-말단에 융합된다. 일부 실시양태에서, NLS는 UBP의 C-말단에 융합된다. 일부 실시양태에서, NLS는 BEE의 N-말단에 융합된다. 일부 실시양태에서, NLS는 BEE의 C-말단에 융합된다. 일부 실시양태에서, NLS는 1종 이상의 링커를 통해 융합 단백질에 융합된다. 일부 실시양태에서, NLS는 링커 없이 융합 단백질에 융합된다. 일부 실시양태에서, NLS는 본원에 제공되거나 언급된 NLS 서열 중 어느 하나의 아미노산 서열을 포함한다. 일부 실시양태에서, NLS는 서열식별번호: 41 또는 서열식별번호: 42에 제시된 바와 같은 아미노산 서열을 포함한다. 핵 국제화 서열은 관련 기술분야에 공지되어 있으며, 통상의 기술자에게 분명한 것이다. 예를 들어, NLS 서열은 PCT/EP2000/011690 (Plank et al.)에 기재되어 있고, 이의 내용은 예시적인 핵 국제화 서열의 개시를 위해 본원에 참조로 포함된다. 일부 실시양태에서, NLS는 아미노산 서열 PKKKRKV (서열식별번호: 41), MDLLMNRKFLYQFKNVRWAKGRRETYLC (서열식별번호: 42), KRTADGSEFSPKKRKV (서열식별번호: 43), KRGINDRNFWRGNGRKTR (서열식별번호: 44), KKTGGPIYRRVDGKWRR (서열식별번호: 45), RRELILYDKKEIRRIWR (서열식별번호: 46), 또는 AVSRKRKA (서열식별번호: 47)를 포함한다.

[0362]

링커

[0363]

특정 실시양태에서, 링커는 본원에 기재된 단백질 또는 단백질 도메인 중 임의의 것을 연결하는데 사용될 수 있다. 링커는 공유 결합만큼 단순할 수 있거나 또는 다수의 원자 길이의 중합체 링커일 수 있다. 특정 실시양태에서, 링커는 폴리펩티드이거나 또는 아미노산을 기재로 한다. 다른 실시양태에서, 링커는 펩티드-유사가 아니다. 특정 실시양태에서, 링커는 공유 결합 (예를 들어, 탄소-탄소 결합, 디설피드 결합, 탄소-헤테로원자 결합 등)이다. 특정 실시양태에서, 링커는 아미드 연결의 탄소-질소 결합이다. 특정 실시양태에서, 링커는 시클릭 또는 비-시클릭, 치환 또는 비치환, 분지형 또는 비분지형 지방족 또는 헤테로지방족 링커이다. 특정 실시양태에서, 링커는 중합체 (예를 들어, 폴리에틸렌, 폴리에틸렌 글리콜, 폴리아미드, 폴리에스테르 등)이다. 특정 실시양태에서, 링커는 아미노알칸산의 단량체, 이량체 또는 중합체를 포함한다. 특정 실시양태에서, 링커는 아미노알칸산 (예를 들어, 글리신, 에탄산, 알라닌, 베타-알라닌, 3-아미노프로판산, 4-아미노부탄산, 5-펜탄산 등)을 포함한다. 특정 실시양태에서, 링커는 아미노헥산산 (Ahx)의 단량체, 이량체 또는 중합체를 포함한다. 특정 실시양태에서, 링커는 카르보시클릭 모이어티 (예를 들어, 시클로펜탄, 시클로헥산)를 기재로 한다. 다른 실시양태에서, 링커는 폴리에틸렌 글리콜 모이어티 (PEG)를 포함한다. 다른 실시양태에서, 링커는 아미노산을 포함한다. 특정 실시양태에서, 링커는 펩티드를 포함한다. 특정 실시양태에서, 링커는 아릴 또는 헤테로아릴 모이어티를 포함한다. 특정 실시양태에서, 링커는 페닐 고리를 기재로 한다. 링커는 펩티드에서 링커로 친핵체 (예를 들어, 티올, 아미노)의 부착을 용이하게 하는 관능화된 모이어티를 포함할 수 있다. 임의의 친전자체는 링커의 일부로서 사용될 수 있다. 예시적인 친전자체는 활성화된 에스테르, 활성화된 아미드, 마이클 수용자, 알킬 할라이드, 아릴 할라이드, 아실 할라이드 및 이소티오시아네이트를 포함하나 이에 제한되지는 않는다.

[0364]

일부 실시양태에서, 링커는 아미노산 또는 복수의 아미노산 (예를 들어, 펩티드 또는 단백질)이다. 일부 실시양태에서, 링커는 결합 (예를 들어, 공유 결합), 유기 분자, 기, 중합체 또는 화학적 모이어티이다. 일부 실시양태에서, 링커는 5-100개 아미노산의 길이, 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, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140, 140-150 또는 150-200개 아미노산의 길이이다. 보다 긴 또는 보다 짧은 링커도 또한 고려된다. 일부 실시양태에서, 링커는 아미노산 서열 SGSETPGTSESATPES (서열식별번호: 102)를 포함하며, 이는 또한 XTEN 링커로 지칭될 수 있다. 일부 실시양태에서, 링커는 아미노산 서열 SGG (서열식별번호: 103)를 포함한다. 일부 실시양태에서, 링커는 (SGG)_n (서열식별번호: 103), (GGG)_n (서열식별번호: 104), (GGGG)_n (서열식별번호: 105), (G)_n (서열식별번호: 121), (EAAAK)_n (서열식별번호: 106), (GGG)_n (서열식별번호: 122), SGSETPGTSESATPES (서열식별번호: 102), SGGSGGSGG (서열식별번호: 120), 또는 (XP)_n 모티프 (서열식별번호: 123), 또는 이들 중 임의의 것의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이고, X는 임의의 아미노산이다. 일부 실시양태에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15이다. 일부 실시양태에서, 링커는 SGSETPGTSESATPES (서열식별번호: 102) 및 SGG (서열식별번호: 103)를 포함한다. 일부 실시양태에서, 링커는 SGGSSGSETPGTSESATPES (서열식별번호: 107)를 포함한다. 일부 실시양태에서, 링커는 SGGSSGSGSETPGTSESATPES (서열식별번호: 108)를 포함한다. 일부 실시양태에서, 링커는 GGGSGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGG (서열식별번호: 109)를 포함한다. 일부 실시양태에서, 링커는 SGGSGGSGG (서열식별번호: 120)를 포함한다.

[0365]

가이드 핵산과의 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp) 복합체

[0366]

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

[0367]

일부 실시양태에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 15-100개 뉴클레오타이드의 길이이고, 표적 서열에 상보적인 적어도 10개의 인접 뉴클레오타이드의 서열을 포함한다. 일부 실시양태에서, 가이드 RNA는 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 또는 50개 뉴클레오타이드의 길이이다. 일부 실시양태에서, 가이드 RNA는 표적 서열에 상보적인 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 또는 40개의 인접 뉴클레오타이드의 서열을 포함한다. 일부 실시양태에서, 표적 서열은 DNA

서열이다. 일부 실시양태에서, 표적 서열은 RNA 서열이다. 일부 실시양태에서, 표적 서열은 포유동물의 게놈에서의 서열이다. 일부 실시양태에서, 표적 서열은 인간의 게놈에서의 서열이다. 일부 실시양태에서, 표적 서열의 3' 말단은 정규 PAM 서열 (NGG)에 바로 인접한다. 일부 실시양태에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 질환 또는 장애와 연관된 서열에 상보적이다. 일부 실시양태에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 본원에 제공된 질환 또는 장애 중 임의의 것과 연관된 유전자에서의 돌연변이를 갖는 질환 또는 장애와 연관된 서열에 상보적이다. 일부 실시양태에서, 가이드 핵산 (예를 들어, 가이드 RNA)은 본원에 제공된 바와 같은 질환 또는 장애와 연관된 유전자 중 임의의 것에 상보적이다.

[0368] 융합 단백질의 사용 방법

[0369] 본 개시내용의 일부 측면은 본원에 제공된 융합 단백질 (예를 들어, 염기 편집제), 또는 본원에 제공된 가이드 핵산 (예를 들어, gRNA) 및 융합 단백질 (예를 들어, 염기 편집제)을 포함하는 복합체 중 임의의 것을 사용하는 방법을 제공한다. 예를 들어, 본 개시내용의 일부 측면은 DNA 또는 RNA 분자를 본원에 제공된 융합 단백질 또는 염기 편집제 중 임의의 것 및 적어도 1종의 가이드 핵산 (예를 들어, 가이드 RNA)과 접촉시키는 것을 포함하며, 여기서 가이드 핵산 (예를 들어, 가이드 RNA)은 약 15-100개 뉴클레오타이드의 길이이고, 표적 서열에 상보적인 적어도 10개의 인접 뉴클레오타이드의 서열을 포함하는 것인 방법을 제공한다. 일부 실시양태에서, 표적 서열의 3' 말단은 정규 spCas9 PAM 서열 (NGG)에 바로 인접한다. 일부 실시양태에서, 표적 서열의 3' 말단은 spCas9 정규 PAM 서열 (NGG)에 바로 인접하지 않는다. 일부 실시양태에서, 표적 서열의 3' 말단은 AGC, GAG, TTT, GTG 또는 CAA 서열에 바로 인접한다.

[0370] 일부 실시양태에서, 표적 DNA 서열은 질환 또는 장애와 연관된 서열을 포함한다. 일부 실시양태에서, 표적 DNA 서열은 질환 또는 장애와 연관된 점 돌연변이를 포함한다. 일부 실시양태에서, 융합 단백질 (예를 들어, napDNAbp, 시티딘 데아미나제 및 우라실 결합 단백질 UBP를 포함하는 것) 또는 복합체의 활성은 점 돌연변이의 교정을 유발한다. 일부 실시양태에서, 표적 DNA 서열은 질환 또는 장애와 연관된 G에서 C로 또는 C에서 G로의 점 돌연변이를 포함하고, 돌연변이체 C 염기의 탈아미노화 및/또는 절제는 질환 또는 장애와 연관되지 않은 서열을 생성한다. 일부 실시양태에서, 표적 DNA 서열은 단백질을 코딩하고, 점 돌연변이는 코돈 내에 존재하고, 야생형 코돈과 비교하여 돌연변이체 코돈에 의해 코딩되는 아미노산의 변화를 유발한다. 일부 실시양태에서, 돌연변이체 C의 탈아미노화는 돌연변이체 코돈에 의해 코딩되는 아미노산의 변화를 유발한다. 일부 실시양태에서, 돌연변이체 C의 탈아미노화는 야생형 아미노산을 코딩하는 코돈을 생성한다. 일부 실시양태에서, 접촉은 대상체에서 생체내 접촉이다. 일부 실시양태에서, 대상체는 질환 또는 장애를 갖거나 갖는 것으로 진단되었다. 일부 실시양태에서, 질환 또는 장애는 22q13.3 결실 증후군; 2-메틸-3-히드록시부티르산노증; 3 메틸크로토닐-CoA 카르복실라제 1 결핍; 3-메틸크로토닐 CoA 카르복실라제 2 결핍; 3-메틸글루타콘산노증 유형 2; 3-메틸글루타콘산노증 유형 3; 3-메틸글루타콘산노증 유형 V; 3-옥소-5 알파-스테로이드 델타 4-데히드로게나제 결핍; 46,XY 성별 반전, 유형 1; 46,XY 참남녀혼합증, SRY-관련; 4-히드록시페닐피루베이트 디옥시게나제 결핍; 비정상적 안면 형상; 비정상적 글리코실화 (CDG IIa); 연골무발생증 유형 2; 완전색맹 2; 완전색맹 5; 완전색맹 6; 완전색맹 7; 후천성 헤모글로빈 H 질환; 침두합지증 유형 I; 호르몬 저항성 동반 또는 비동반 선단이골증 1; 선단이골증 2, 호르몬 저항성 동반 또는 비동반; 얼굴말단뼈발생이상, 신시내티 유형; ACTH 저항성; 급성 뉴런병 증성 고셔병; 아담스-올리버 증후군; 아담스-올리버 증후군 2; 아담스-올리버 증후군 4; 아담스-올리버 증후군 6; 아데닌 포스포리보실트랜스퍼라제 결핍; 아데닐로숙시네이트 리아제 결핍; 청소년 신경석증; 부신백질이영양증; 성인 접합부 수포성 표피박리증; 성인 신경 세로이드 리포푸신증; ADULT 증후군; 연령-관련 황반 변성 14; 연령-관련 황반 변성 3; 에카르디 구티에레스 증후군 5; 에카르디-구티에레스 증후군 6; 알렉산더병; 알파 지중해빈혈; 알파-B 크리스탈린병증; 알포트 증후군, 상염색체 열성; 알포트 증후군, X-연관 열성; 소아기 교대성 편마비 2; 알츠하이머병; 알츠하이머병, 유형 1; 알츠하이머병, 유형 3; 불완전 사기질형성증, 저성숙 유형, IIA3; 불완전 사기질형성증, 유형 1E; 아미쉬 치사 소두증; AML - 급성 골수성 백혈병; 아밀로이드생성 트랜스티레틴 아밀로이드증; 근위축성 측삭 경화증 16, 소아; 근위축성 측삭 경화증 6, 상염색체 열성; 근위축성 측삭 경화증 유형 1; 근위축성 측삭 경화증 유형 10; 근위축성 측삭 경화증 유형 2; 근위축성 측삭 경화증 유형 9; 안테르센 타일 증후군; 빈혈, 적혈구생성이상 선천성, 유형 IV; 빈혈, 비구상 용혈성, G6PD 결핍에 기인; 빈혈, 철적혈모구성, 피리독신-불응성, 상염색체 열성; 안젤만 증후군; 혈관병증, 유전성, 신병증, 동맥류 및 근육 경련 동반; 면역 결핍 동반 무한성 외배엽 이형성증; 손발톱결여증; 생식기 기형 및 스테로이드생성 장애 동반 엔틀리-빅슬러 증후군; 생식기 기형 및 스테로이드생성 장애 비동반 엔틀리-빅슬러 증후군; 재생불량성 빈혈; 아포지단백질 a-i 결핍; 아르기나제 결핍; 부정맥유발성 우심실 심근병증; 부정맥유발성 우심실 심근병증, 유형 11; 부정맥유발성 우심실 심근병증, 유형 9; 영아기 동맥 석회화; 동맥 비틀림 증후군; 선천성 다발성 관절만곡증 원위 유형 1; 관절만곡증 신장 기능장애 담즙정체 증후군; 관절만곡증, 원위, 유형 5d; 아트 증후군; 아스파

르틸글루코사민노, 핀란드 유형; 질식성 가슴 이영양증 2; 비타민 E 결핍 동반 운동실조; 운동실조-모세혈관확장증 증후군; 운동실조-모세혈관확장증-유사 장애; 아텔로스테오제네시스 유형 1; 심방 세동; 심방 세동, 가족성, 10; 심방 중격 결손 4; 유전성 불보름 위축증(Aтроphia bulborum hereditaria); ATR-X 증후군; 비정형 용혈성-요독성 증후군 1; 청각 신경병증, 상염색체 열성, 1; 컷바퀴관절염 증후군 1; 자가면역 질환, 다기관, 영아-발병; 자가면역 림프증식성 증후군, 유형 1A; 자가면역 림프증식성 증후군, 유형 V; 상염색체 우성 야간전두엽 간질; 미토콘드리아 DNA 결실 동반 상염색체 우성 진행성 외안근마비 2; 미토콘드리아 DNA 결실 동반 상염색체 우성 진행성 외안근마비 3; 미토콘드리아 DNA 결실 동반 상염색체 우성 진행성 외안근마비 4; 상염색체 열성 선천성 어린선 1; 상염색체 열성 선천성 어린선 5; 상염색체 열성 저인산혈증 비타민 D 불응성 구루병; 악센펠트-리거 기형; 악센펠트-리거 증후군 유형 1; 악센펠트-리거 증후군 유형 3; 바레이처-윈터 증후군 1; 바르데-비들 증후군; 바르데-비들 증후군 10; 바르데-비들 증후군 12; 바르데-비들 증후군 2; 바르데-비들 증후군 3; 바르데-비들 증후군 4; 바르데-비들 증후군 9; 바터 증후군 산전 유형 2; 바터 증후군, 유형 4b; 기저 신경절 질환, 비오틴-반응성; 베커 근육 이영양증; 양성 가족성 신생아 발작 1; 양성 가족성 신생아-영아 발작; 양성 재발성 간내 담즙정체 2; 버나드-술리에 증후군, 유형 B; 베타 지중해빈혈; 비에티 결정질 각막막 이영양증; 담즙산 합성 결손, 선천성, 2; 비오틴다제 결핍; 출혈 장애, 혈소판-유형, 19; 혈액형--루터안 억제제; 블롬 증후군; 보슬리-살리-알로레이니 증후군; 보우셔 노이하우저 증후군; 단지증 유형 B2; 유방암; 유방-난소암, 가족성 1; 유방-난소암, 가족성 2; 기관지확장증; 브라운-비알레토-반 라에레 증후군; 브라운-비알레토-반 라에레 증후군 2; 수포성 어린선형 홍피증; 버킷 림프종; 굴지 이형성증; Cap 근병증 2; 탄수화물-결핍 당단백질 증후군 유형 I; 탄수화물-결핍 당단백질 증후군 유형 II; 결장 암증; 췌장 암증; 심장 부정맥; 심근병증, 영아 치사성, 시토크롬 C 옥시다제 결핍에 기인 3; 심장얼굴피부 증후군; 심장얼굴피부 증후군 2; 심근병증; 심근병증, 제한성; 카니 복합증, 유형 1; 카르니틴 팔미토일트랜스퍼라제 I 결핍; 백내장 1; 백내장, 선천성, 감각신경성 난청, 다운 증후군-유사 안면 외관, 저신장 및 정신 지체 동반; 카테콜아민성 다형성 심실성 빈맥; 중심핵병; 중앙 성조숙증; 소뇌 운동실조 및 저생식선자극 생식선기능저하증; 진행성 외안근마비 동반 소뇌 운동실조 영아; 소뇌 운동실조, 난청, 및 기면증; 뇌 아밀로이드 혈관병증, APP-관련; 피질하 경색 및 백질뇌병증을 동반한 상염색체 우성 뇌동맥병증; 뇌 해면 기형 1; 뇌성 마비, 경직성 사지마비, 1; 뇌-늑골-하악 증후군; 세로이드 리포푸신증 뉴런 1; 세로이드 리포푸신증 뉴런 10; 세로이드 리포푸신증 뉴런 6; 세로이드 리포푸신증 뉴런 7; 세로이드 리포푸신증 뉴런 8; 세로이드 리포푸신증, 뉴런, 13; 세로이드 리포푸신증, 뉴런, 2; 체디악-히가시 증후군; 차르 증후군; 샤르코-마리-투스병; 샤르코-마리-투스병 유형 1B; 샤르코-마리-투스병 유형 2B; 샤르코-마리-투스병 유형 2D; 샤르코-마리-투스병 유형 2I; 샤르코-마리-투스병 유형 2K; 샤르코-마리-투스병, 축삭, 성대 부전마비 동반, 상염색체 열성; 샤르코-마리-투스병, 탈수초성, 유형 1C; 샤르코-마리-투스병, 우성 중간체 E; 샤르코-마리-투스병, 유형 2; 샤르코-마리-투스병, 유형 2A2; 샤르코-마리-투스병, 유형 4C; 샤르코-마리-투스병, 유형 4G; 샤르코-마리-투스병, 유형 IA; 샤르코-마리-투스병, 유형 IE; 샤르코-마리-투스병, 유형 IF; 샤르코-마리-투스병, X-연관 열성, 유형 5; CHARGE 연합증; CHILD 증후군; 콜레스타놀 축적 질환; 콜레스테롤 모노옥시게나제 (측쇄 절단) 결핍; 점상 연골이형성증 1, X-연관 열성; CHOPS 증후군; 염색체 9q 결실 증후군; 만성 육아종성 질환, X-연관; 섬모 이상운동증, 원발성, 14; 섬모 이상운동증, 원발성, 19; 섬모 이상운동증, 원발성, 3; 섬모 이상운동증, 원발성, 7; 쇄골두개 이골증; 코케인 증후군 유형 A; 코린-로우리 증후군; 코헨 증후군; 골 질환; 결장직장암, 유전성, 비폴립증, 유형 1; 육아종 동반 복합 세포성 및 체액성 면역 결손; 복합 산화성 인산화 결핍 24; 복합 산화성 인산화 결핍 9; 공통 가변성 면역결핍 7; 보체 성분 9 결핍; 추체-간체 이영양증 10; 추체-간체 이영양증 11; 추체-간체 이영양증 3; 추체-간체 이영양증 5; 추체-간체 이영양증 6; 선천성 부신 저형성증, X-연관; 선천성 무거핵구성 혈소판감소증; 선천성 무홍채증; 선천성 양측 정관 결손증; 선천성 백내장, 청각 상실, 및 신경변성; 선천성 구축성 거미가락증; 선천성 폴레이트 흡수 결함; 선천성 글리코실화 장애 유형 1K; 선천성 글리코실화 장애 유형 1M; 선천성 글리코실화 장애 유형 1t; 선천성 글리코실화 장애 유형 1u; 선천성 글리코실화 장애 유형 2C; 선천성 전신성 지방이영양증 유형 1; 선천성 전신성 지방이영양증 유형 2; 선천성 심장 결손, 다중 유형, 1, X-연관; 선천성 락타제 결핍; 선천성 긴 QT 증후군; 뇌 및 눈 기형 동반 선천성 근육 이영양증-디스트로글리칸병증, 유형 A2; 뇌 및 눈 기형 동반 선천성 근육 이영양증-디스트로글리칸병증, 유형 A7; 정신 지체 동반 선천성 근육 이영양증-디스트로글리칸병증, 유형 B1; 정신 지체 동반 선천성 근육 이영양증-디스트로글리칸병증, 유형 B2; 섬유 유형 불균등 동반 선천성 근병증; 선천성 근긴장증, 상염색체 우성 형태; 선천성 근긴장증, 상염색체 열성 형태; 선천성 비진행성 야맹증, 상염색체 우성 3; 선천성 비진행성 야맹증, 유형 1A; 선천성 비진행성 야맹증, 유형 1F; 코프로포르피린증; 각막 이영양증, 푸크스 내피, 8; 각막 상피 이영양증; 각막 취약성 구형각막, 청색 공막 및 관절 파운동성; 코넬리아 드 랑귀 증후군 1; 코넬리아 드 랑귀 증후군 4; 피질 이형성증, 복합체, 다른 뇌 기형 동반 3; 코르티손 리덕타제 결핍 1; 코우텐 증후군 2; 두개외배엽 이형성증 1; 두개안면 난청 증후군; 두개골관절병증; 두개골유합증; 두개골유합증 3; 두개

골유합증 및 치아 기형; 크레아틴 결핍, X-연관; 크리글러 나자르 증후군, 유형 1; 크루존 증후군; 잠복안구 증후군; 잠복고환, 일측성 또는 양측성; 쿠싱 지혈유합증; 베어 및 스티븐슨 뇌회상 피부 증후군; 시스타티오닌뇨증; 낭성 섬유증; 시스틴축적증, 안구 비신병증성; 시토크롬-c 옥시다제 결핍; 다논병; 난청, 상염색체 우성 12; 난청, 상염색체 우성 20; 난청, 상염색체 열성 1A; 난청, 상염색체 열성 63; 난청, 상염색체 열성 8; 난청, 상염색체 열성 9; 아세틸-CoA 아세틸트랜스퍼라제 결핍; 알파-만노시다제 결핍; 페록시다제 결핍; 글리세롤 키나제 결핍; 구아니디노아세테이트 메틸트랜스퍼라제 결핍; 히드록시메틸글루타릴-CoA 리아제 결핍; 아이오다이드 퍼옥시다제 결핍; 말로닐-CoA 데카르복실라제 결핍; UDP글루코스-헥소스-1-포스페이트 우리딜트랜스퍼라제 결핍; 말 및 언어 발달 지연; 델타 지중해빈혈; 텐트병 1; 드뷔쿠아 증후군; 데스모스테롤로시스 (Desmosterolosis); DFNA 2 비증후군성 청각 상실; 당뇨병 유형 2; 당뇨병, 인슐린-의존성, 20; 수족지신장뇌 (Digitorenocerebral) 증후군; 확장성 심근병증 1FF; 확장성 심근병증 1G; 확장성 심근병증 1S; 확장성 심근병증 1X; 확장성 심근병증 3B; 시토크롬 p450 옥시도리덕타제 결핍에 기인한 스테로이드생성 장애; 원위 유전성 모터 뉴런병증 유형 2B; 첩모중생증-림프부종 증후군; 드래쉬 증후군; 뒤시엔느 근육 이영양증; 선천성 이상각화증 상염색체 우성; 선천성 이상각화증 X-연관; 선천성 이상각화증, 상염색체 우성, 2; 선천성 이상각화증, 상염색체 열성, 5; 이상긴장증 1; 이상긴장증 27; 이상긴장증 5, 도파-반응성 유형; 이상긴장증, 도파-반응성, 고페닐알라닌혈증 동반 또는 비동반, 상염색체 열성; 조기 영아 간질성 뇌병증 13; 조기 영아 간질성 뇌병증 2; 조기 영아 간질성 뇌병증 8; 조기 영아 간질성 뇌병증 9; 조기 근간대성 뇌병증; 외배엽 이형성증-합지증 증후군 1; 가락결손증, 외배엽 이형성증, 및 구순열/구개 증후군 3; 엘러스-단로스 증후군, 전형적 유형; 엘러스-단로스 증후군, 히드록시리신-결핍; 엘러스-단로스 증후군, 근육구축 유형; 엘러스-단로스 증후군, 유형 4; 아이스펠트 유형 선천성 근육 이영양증; 타원적혈구증 3; 자궁내막 암증; 종말판 아세틸콜린에스테라제 결핍; 비대 전정 수도관 증후군; 엔테로키나제 결핍; 단순 수포성 표피박리증, 쾨브너 유형; 간질, 야간 전두엽, 유형 3; 간질, 진행성 근간대성 1A (운베리히트 및 룬트보르크); 간질, 진행성 근간대성 2b; 간질성 뇌병증, 조기 영아, 1; 간질성 뇌병증, 조기 영아, 24; 간질성 뇌병증, 조기 영아, 28; 간질성 뇌병증, 조기 영아, 31; 골단 연골종 이형성증, 미우라 유형; 간혈성 운동실조 유형 1; 간혈성 운동실조, 유형 6; 삼화성 통증 증후군, 가족성, 3; 적혈구증가증, 가족성, 2; 적혈구증가증, 가족성, 3; 운동실조 동반 홍반각피증; 삼출성 유리체망막병증 1; 삼출성 유리체망막병증 5; 파브리병; 파브리병, 심장 변이형; 인자 v 및 인자 viii, 복합 결핍, 2; 두드러기 및 난청 동반 가족성 아밀로이드 신병증; 가족성 유방암; 가족성 한랭 두드러기; 가족성 열성 발작 8; 가족성 편마비성 편두통 유형 3; 가족성 비후성 심근병증 1; 가족성 비후성 심근병증 10; 가족성 비후성 심근병증 11; 가족성 비후성 심근병증 20; 가족성 비후성 심근병증 23; 가족성 비후성 심근병증 4; 가족성 비후성 심근병증 6; 가족성 형성부전성, 사구체낭성 신장; 가족성 영아 근무력증; 가족성 소아 통풍; 가족성 지중해열; 연관 골수성 악성종양 동반 가족성 혈소판 장애; 가족성 공뇌증; 가족성 만발성 피부 포르피린증; 가족성 내장 아밀로이드증, 오스터택 유형; 판코니 빈혈, 상보군 C; 판코니 빈혈, 상보군 F; 판코니 빈혈, 상보군 G; 판코니 빈혈, 상보군 J; 판코니 빈혈, 상보군 T; 파버 지방육아종증; 태아성 헤모글로빈 정량적 형질 유전자좌 1; 태아성 헤모글로빈 정량적 형질 유전자좌 6; 섬유성연골이형성증; 정신 지체 동반 또는 비동반 언어 장애 동반 초점성 간질; 초점성 분절성 사구체경화증 6; 망막종증와 저형성증 및 초로기 백내장 증후군; 전두비부 이형성증 1; 전두비부 이형성증 2; 전두측두엽 치매; 프록토스-비포스파타제 결핍; 푸마라제 결핍; 갈락토실세라미드 베타-갈락토시다제 결핍; 담낭 질환 4; 감스트로프-볼파르트 증후군; 강글리오시드 시알리다제 결핍; 강글리오시드증 GM1 유형 3; 가드너 증후군; 적혈구형성이상 동반 GATA-1-관련 혈소판감소증; 고셔병; 고셔병 유형 3C; 고셔병, 주산기 치사성; 고셔병, 유형 1; 열성 발작 동반 전신성 간질 플러스, 유형 1; 열성 발작 동반 전신성 간질 플러스, 유형 2; 열성 발작 동반 전신성 간질 플러스, 유형 9; 게르스트만-스트라우슬러-샤잉크 증후군; 글라즈만 혈소판무력증; 녹내장 1, 개방각, F; 녹내장, 선천성; 전반적 발달 지연; 글루코코르티코이드 결핍 4; 글루타르 산뇨증, 유형 1; 글리코겐 축적 질환 IIIa; 글리코겐 축적 질환 IV, 선천성 신경근육; 글리코겐 축적 질환 IXb; 심장 글리코겐 축적 질환, 치사성 선천성; 글리코겐 축적 질환, 유형 II; 글리코겐 축적 질환, 유형 IV; 글리코겐 축적 질환, 유형 V; 글리코겐 축적 질환, 유형 VI; 글리코실포스파티딜이노시톨 결핍; 회색 혈소판 증후군; 그리셀리 증후군 유형 2; 성장 및 정신 지체, 하악안면 이골증, 소두증, 및 구개열; 면역결핍 동반 성장 호르몬 불감성; 혈색소증 유형 1; 혈색소증 유형 3; 헥소키나제 결핍에 기인한 용혈성 빈혈; 용혈성 빈혈, 비구상적혈구, 글루코스 포스페이트 이소머라제 결핍에 기인; 혈철증, 전신, 무세룰로플라스민혈증에 기인; 헨레캄 림프관 확장증-림프부종 증후군; 유전성 장병성 말단피부염; 유전성 혈관부종 유형 1; 유전성 유방 및 난소암 증후군; 유전성 암-소인 증후군; 유전성 미만성 위암; 구상체 동반 유전성 미만성 백질뇌병증; 유전성 인자 II 결핍 질환; 유전성 인자 IX 결핍 질환; 유전성 인자 VIII 결핍 질환; 유전성 인자 XI 결핍 질환; 유전성 프록토스노증; 유전성 평활근증증 및 신세포암; 유전성 림프부종 유형 I; 유전성 신경통 근위축; 유전성 비폴립증 결장직장암 유형 5; 유전성 비폴립증 결장직장 신생물; 유전성 체장염; 유전성 부신경절종-크롬친화세포종 증후군; 유전성

열성변형적혈구증가증; 유전성 감각 신경병증 유형 1D; 유전성 철적혈모구성 빈혈; 착위, 내장, X-연관; 이소증; 허쉬쉬스프룽병 신경질신경모세포종; 조직구성 수질성 망상증; 완전전뇌증 11; 완전전뇌증 2; 완전전뇌증 3; 완전전뇌증 4; MTHFR 결핍에 기인한 호모시스테인혈증; CBS 결핍에 기인한 호모시스틴뇨증; 후롤러 증후군; 갑상선 휘르틀레 세포 암증; 허친슨-길포드 증후군; 고칼슘뇨, 소아기, 자기-제한성; 고콜레스테롤혈증; 파도놀람증 3; 유전성 파도놀람증; 고펜티딘혈증 백내장 증후군; 고지단백혈증, 유형 I; 고지단백혈증, 유형 ID; 고리신혈증; 고오르니틴혈증-고암모니아혈증-호모시트룰린뇨증 증후군; 고프로인술린혈증; 양안과다격리증, 증증, 얼굴중간 융기, 근시, 정신 지체 및 골 취약성 동반; 비대성 심근병증; 저칼슘혈증, 상염색체 우성 1; 저칼슘혈증, 상염색체 우성 1, 바터 증후군 동반; 저연골형성증; 철 과부하 동반 저색소성 소구성 빈혈; 간 내 글리코겐 신테타제 결핍 동반 저혈당증; 후각결여 동반 또는 비동반 저생식선자극 생식선기능저하증 13; 발한저하성 X-연관 외배엽 이형성증; 저칼륨혈증 주기성 마비 1; 저마그네슘혈증 1, 장; 저마그네슘혈증 5, 신장, 안구 침범 동반; 저마그네슘혈증, 발작, 및 정신 지체; 저수초형성 백질이영양증 7; 저수초형성 백질이영양증 8, 치아 부족증 및/또는 저생식선자극 생식선기능저하증 동반 또는 비동반; 저단백혈증, 과다이하; 갑상선기능저하증, 선천성, 비갑상선종성, 1; 갑상선기능저하증, 선천성, 비갑상선종성, 5; 갑상선기능저하증, 선천성, 비갑상선종성, 6; 티갑소증 6; 티갑소증-림프부종-모세혈관확장증 증후군; I 세포 질환; 심상성 어린선; 특발성 기저 신경절 석회화 5; 면역결핍 12; 면역결핍 23; 면역결핍 24; 면역결핍 30; 면역결핍 31a; 면역결핍 31C; 과다 IgM 동반 면역결핍 유형 1; 봉입체 근병증 2; 영아 소뇌-망막 변성; 영아 GM1 강글리오시드증; 영아 저포스파타제증; 영아 안진, X-연관; 인슐린-저항성 당뇨병 및 흑색 극세포종; 지적 장애; 중간체 메이플 시럽뇨 질환 유형 2; 침습성 폐렴구균성 질환, 재발성 고립성, 2; 홍채-각막-섬유주 이발생증; 뇌 내 철 축적; 잭슨-바이스 증후군; 야콥-크로이츠펠트병; 주버트 증후군 23; 소아 GM1강글리오시드증; 소아 폴립증 증후군; 카부키 메이크업 증후군; 칼만 증후군 3; 칼만 증후군 4; 칼만 증후군 5; 칼만 증후군 6; 원추각막 1; 켈슈터 증후군; 쿠겔베르그-벨란더병; 라포르병; 랑거 전완부 이형성증 증후군; 라론-유형 고립성 소마토포로핀 결손; 라슨 증후군, 우성 유형; 임신 모체 급성 지방간 동반 LCHAD 결핍; 레베르 선천성 흑암시 13; 레베르 선천성 흑암시 4; 레베르 선천성 흑암시 9; 리병; 레오파드 증후군; 레오파드 증후군 1; 레오파드 증후군 2; 요정증 증후군; 레리 웨일 연골 뼈형성이상증; 레쉬-니한 증후군; 백질이영양증, 저수초형성, 6; 운동실조 동반 백질뇌병증; 뇌간 및 척수 침범 및 락테이트 상승 동반 백질뇌병증; 소멸 백질 동반 백질뇌병증; 라이디히 세포 무발생증; 리-프라우메니 증후군 1; 지대 근육 이영양증; 지대 근육 이영양증, 유형 1B; 지대 근육 이영양증, 유형 1C; 지대 근육 이영양증, 유형 1E; 지대 근육 이영양증, 유형 2A; 지대 근육 이영양증, 유형 2B; 지대 근육 이영양증, 유형 2E; 지대 근육 이영양증, 유형 2F; 지대 근육 이영양증, 유형 2L; 지대 근육 이영양증-디스트로글리칸병증, 유형 C1; 지대 근육 이영양증-디스트로글리칸병증, 유형 C14; 지대 근육 이영양증-디스트로글리칸병증, 유형 C2; 지대 근육 이영양증-디스트로글리칸병증, 유형 C7; 활택뇌증 1; 긴 QT 증후군 1; 긴 QT 증후군 13; 긴 QT 증후군 15; 긴 QT 증후군 2; 긴 QT 증후군 9; 긴 QT 증후군, LQT1 하위유형; 장쇄 3-히드록시아실-CoA 데히드로게나제 결핍; 로웨 증후군; 황체형성 호르몬 저항성, 여성; 림프증식성 증후군 1; 림프증식성 증후군 1, X-연관; 린치 증후군 I; 린치 증후군 II; 거대혈소판감소증, 가족성, 버나드-술리에 유형; 원뿔세포 침범 동반 황반 이영양증; 마지드 증후군; 식도 악성 종양; 전립선 악성 종양; 하악말단 이골증; 메이플 시럽뇨 질환; 메이플 시럽뇨 질환 유형 1A; 메이플 시럽뇨 질환 유형 2; 마르팡 증후군; 마리 오나 유전성 티갑소증 1; 청소년 성숙기-발병 당뇨병, 유형 2; 청소년 성숙기-발병 당뇨병, 유형 3; 중간쇄 아실-조효소 A 데히드로게나제 결핍; 마이어-골린 증후군 5; 멜닉-프레이저 증후군; MEN2 표현형: 비분류; MEN2 표현형: 미지; 멘케스 킨키-헤어 증후군; 자연 폐경기, 양적 형질 유전자좌(Menopause, natural, age at, quantitative trait locus) 3; 정신 지체 30, X-연관; 정신 지체 및 뇌교 및 소뇌 저형성증 동반; 정신 지체, 상염색체 우성 13; 정신 지체, 상염색체 우성 16; 정신 지체, 상염색체 우성 29; 정신 지체, 상염색체 우성 38; 정신 지체, 상염색체 우성 7; 정신 지체, 상염색체 열성 34; 정신 지체, 상염색체 열성 49; 정신 지체, 상동증적 운동, 간질, 및/또는 뇌 기형; 정신 지체, 증후군성, 클라에스-엔센 유형, X-연관; 정신 지체, X-연관, 증후군성 13; 정신 지체, X-연관, 증후군성 32; 정신 지체, X-연관, 증후군성, 레이몬드 유형; 정신 지체, X-연관, 증후군성, wu 유형; 정신 지체-저긴장 얼굴 증후군 X-연관, 1; 메로신 결핍 선천성 근육 이영양증; 이염성 백질이영양증; 골간단 연골종이형성증, 슈미트 유형; 메틸코발라민 결핍, cblg 유형; 메틸말론산뇨증, mut(0) 유형; 소두증 및 맥락망막병증, 상염색체 열성, 2; 맥락망막병증, 림프부종 또는 정신 지체 동반 또는 비동반 소두증; 소구성 빈혈; 왜소음경증; 소안구증 증후군성 3; 소안구증 증후군성 5; 소안구증, 고립성 3; 소안구증, 고립성 6; 소안구증, 고립성, 결손증 동반 7; 당뇨병 미세혈관 합병증 7; 경도 비-PKU 고펜틸알라닌혈증; 미토콘드리아 복합체 I 결핍; 미토콘드리아 복합체 II 결핍; 미토콘드리아 복합체 III 결핍; 미토콘드리아 DNA 고갈 증후군 13 (뇌근병증 유형); 미토콘드리아 DNA 고갈 증후군 2; 미토콘드리아 DNA 고갈 증후군 9 (메틸말론산뇨증 동반 뇌근병증); 미토콘드리아 단백 에노일-CoA 히드라타제 1 결핍; 미토콘드리아 삼관능성 단백결 결핍; 미요시 근육 이영양증 1; 미요시 근육 이영양증 3; 모어-트라네브자에르그

(Mohr-Tranebjaerg) 증후군; 모자이크 반문상의 이수성 증후군; 모와트-윌슨 증후군; 점액지질증 III 감마; 뮤코폴리사카라이드증 유형 VI; 뮤코폴리사카라이드증, MPS-II; 뮤코폴리사카라이드증, MPS-III-B; 뮤코폴리사카라이드증, MPS-I-S; 뮤코폴리사카라이드증, MPS-IV-A; 뮤코폴리사카라이드증, MPS-IV-B; 뮌케 증후군; 물리브리 난장이 증후군; 다발성 선천성 기형; 다발성 내분비 신생물, 유형 1; 다발성 내분비 신생물, 유형 2; 다발성 내분비 신생물, 유형 2a; 다중 골단 이형성증 1; 다중 골단 이형성증 5; 다중 외골종 유형 2; 다중 이상편 증후군 에스코바르 유형; 다중 슬파타제 결핍; 절단 각피증; 근무력증, 지대, 가족성; 근무력 증후군, 선천성, 9, 아세틸콜린 수용체 결핍 근무력 증후군 연관, 선천성, 9, 아세틸콜린 수용체 결핍 연관; 근무력 증후군, 선천성, 시냅스전 및 시냅스후 결손; 근무력 증후군, 선천성, 세관 응집체 동반 2; 근무력 증후군, 느린-채널 선천성; 근간대성 간질 근병증 감각 운동실조; 근간대성경련, 가족성 피질; 근원섬유성 근병증 1; 근파동 1; 체위성 근육 위축 동반 근병증, X-연관; 근병증, 액틴, 선천성, 파랑의 얇은 근필라멘트; 근병증, 중심핵성; 근병증, 원위, 1; 근병증, 고립성 미토콘드리아, 상염색체 우성; 근병증, 신체 감소, X-연관, 조기 발병, 중증; 선천성 근긴장증; 손발톱 장애, 비증후군성 선천성, 8; 소안구증 4; 기면증 7; 네이티브 아메리칸 근병증; 나바호 신경간병증; 네발린 근병증 3; 신생아 저긴장증; 신생아 인슐린-의존성 당뇨병; 시트린 결핍에 의해 유발된 신생아 간내 담즙정체; 난소 신생물; 신석증/골다공증, 저인산혈증, 2; 신결석증 16; 신결석증 18; 신증후군, 유형 10; 노이-락소바 증후군 1; 뇌 철 축적을 동반한 신경변성 5; 신경뇌하수체 요붕증; 니콜레이테스-바레이저 증후군; 니만-픽병 유형 C1; 니만-픽병, 유형 A; 니만-픽병, 유형 B; 니만-픽병, 유형 c1, 소아 형태; 노나카 근병증; 비-케톤성 고글리신혈증; 누난 증후군 1; 누난 증후군 5; 누난 증후군 7; 누난 증후군 8; 제공되지 않음; 명시되지 않음; 눈피부 백색증 유형 3; 안인두 근육 이영양증; 옵시모디스플라시아(Opsismodysplasia); 시신경 위축 9; 시신경 위축 및 백내장, 상염색체 우성; 시신경 저형성증 및 중추 신경계 이상; 구강-안면-수족 지 증후군; 오르니틴 아미노트랜스퍼라제 결핍; 오르니틴 카르바모일트랜스퍼라제 결핍; 구강안면열 11; 구안지 증후군 6; 오르트산뇨증; 골형성 부전증 유형 12; 골형성 부전증 유형 13; 골형성 부전증 유형 III; 정상 공막을 동반한 골형성 부전증, 우성 형태; 골형성 부전증, 열성 주산기 치사성; 골화석증 상염색체 우성 유형 1; 골화석증 상염색체 열성 7; 귀-구개-수족지 증후군, 유형 I; 피부부비후골막증 증후군; 팔리스터-홀 증후군; 파피용-르페브르 증후군; 부신경절종 1; 부신경절종 4; 부갑상선 암증; 두정공 2; 파킨슨병 1; 파킨슨병 7; 파킨슨병 9; 발작성 야간 혈색소뇨 1; 부분 하이포크산틴-구아닌 포스포리보실트랜스퍼라제 결핍; 피부 박리 증후군, 말단 유형; 펠거-휴에트 이상; 펠리제우스-메르츠바허병; 펜드레드 증후군; 영구 신생아 당뇨병; 피옥시슴 생물발생 장애 6B; 피옥시슴 생물발생 장애 9B; 포이즈-예거스 증후군; 파이퍼 증후군; 페닐케톤뇨; 크롬친화세포종; 포스포글리세레이트 키나제 1 결핍; 포스포리보실피로포스페이트 신테타제 과다활성; 감광성 모발유황이영양증; 피어슨 증후군; 색소성 담창구 변성; 피트-홉킨스 증후군; 피트-홉킨스-유사 증후군 2; 뇌하수체 의존성 고코르티솔증; 뇌하수체 호르몬 결핍, 복합 1; 뇌하수체 호르몬 결핍, 복합 4; 뇌하수체 호르몬 결핍, 복합 5; 혈소판-유형 출혈 장애 16; 다응집성 적혈구 증후군(Polyagglutinable erythrocyte syndrome); 결절성 다발동맥염; 다낭성 신장 질환, 영아 유형; 폴리글루코산체 근병증 2; 다소뇌회증, 양측성 전두두정; 다발신경병증, 청각 상실, 운동실조, 색소성 망막염, 및 백내장; 교뇌소뇌 저형성증, 유형 1B; 교뇌소뇌 저형성증, 유형 1c; 교뇌소뇌 저형성증, 유형 9; 포레티-볼츠하우저 증후군; 축전성 다지증 2; 조기 염색분체 분리 형질; 조기 난소 부전 5; 조기 난소 부전 7; 조기 난소 부전 9; 원발성 상염색체 열성 소두증 1; 원발성 상염색체 열성 소두증 2; 원발성 상염색체 열성 소두증 5; 원발성 상염색체 열성 소두증 6; 원발성 심모 이상운동증; 원발성 확장성 심근병증; 원발성 가족성 비후성 심근병증; 원발성 고옥살산뇨증, 유형 I; 원발성 고옥살산뇨증, 유형 III; 원발성 국제화 피부 아밀로이드증 1; 원발성 개방각 녹내장 소아 발병 1; 원발성 폐고혈압; 원발성 폐고혈압 4; 프림로스 증후군; 진행성 골화성 근염; 진행성 경화성 회백질위축증; 증식성 혈관병증 및 무뇌수두증-수두증 증후군; 프로페르딘 결핍, X-연관; 프로피온산혈증; 슈도-후틀러 다발이영양증; 제1형 가성저알도스테론증 상염색체 우성; 가성저알도스테론증 유형 2B; 가성저알도스테론증, 유형 2; 가성부갑상선기능저하증 유형 1A; 탄성섬유 가성황색증; 다중 응고 인자 결핍 동반 탄성섬유 가성황색증-유사 장애; 유전성 출혈성 모세혈관확장증 관련 폐동맥 고혈압; 폐 섬유증 및/또는 골수 부전, 텔로미어-관련, 2; 농축이글증; 피리독신-의존성 간질; 피루베이트 데히드로게나제 E1-알파 결핍; 요골 무형성-혈소판감소 증후군; 레인 증후군; RAS병증; 열성 이영양성 수포성 표피박리증; 라이펜스타인 증후군; 신장 카르니틴 수송 결손; 신세포 암증, 유두상, 1; 신장 이형성증; 신장 저요산혈증 2; 신세포관 산증, 원위, 용혈성 빈혈 동반; 망막 원추 이영양증 3A; 색소성 망막염; 색소성 망막염 10; 색소성 망막염 11; 색소성 망막염 14; 색소성 망막염 2; 색소성 망막염 25; 색소성 망막염 33; 색소성 망막염 35; 색소성 망막염 4; 색소성 망막염 43; 색소성 망막염 50; 색소성 망막염 56; 색소성 망막염 73; 색소성 망막염 74; 망막모세포종; 레트 장애; 레트 증후군, 선천성 변이형; 레트 증후군, 자펠라 변이형; 횡문근양 종양 소인 증후군 2; 근형부 점상 연골이형성증 유형 1; 리엔호프 증후군; 로버츠-SC 해표지증 증후군; 로비노우 증후군; RRM2B-관련 미토콘드리아 질환; 루빈스타인-테이비 증후군; 사에트레-코첸 증후군; 견갑비골 근병증, X-연관 우

성; 선들러병, 유형 1; 선들러병, 유형 3; 슈나이더 결정질 각막 이영양증; 시클 증후군 1; 발작; 선택적 치아 무발생증 1; 시니어-로켄 증후군 8; 감각 실조 신경병증, 구음장애, 및 안근부전마비; SeSAME 증후군; ADA 결핍에 기인한 중증 복합 면역결핍; 소두증, 성장 지연, 및 이온화 방사선에 대한 감수성 동반 중증 복합 면역결핍; 중증 선천성 호중구감소증; 중증 선천성 호중구감소증 4, 상염색체 열성; 영아기에서의 중증 근간대성 간질; 중증 X-연관 근세관성 근병증; 짧은 QT 증후군; 짧은 QT 증후군 2; 비특이적 골격 이상을 동반한 저신장; 저신장, 이도 폐쇄증, 하악 저형성증, 골격 이상; 저신장, 특발성, 상염색체; 저신장, 특발성, X-연관; 다지증 동반 또는 비동반 짧은-늑골 흉부 이형성증 13; 다지증 동반 짧은-늑골 흉부 이형성증 14; 다지증 동반 또는 비동반 짧은-늑골 흉부 이형성증 3; 쉬프린젠 증후군; 쉬프린젠-골드버그 증후군; 슈와크만 증후군; 시알산 축적 질환, 중증 영아 유형; 시알산증, 유형 II; 동기능 부전 증후군 2, 상염색체 우성; B-세포 면역결핍, 주기열 및 발달 지체 동반 철적혈구구성 빈혈; 시토스테롤혈증; 쇼그렌-라르손 증후군; 스미스-렘리-오피츠 증후군; 소르스비안저 이영양증; 소토스 증후군 1; 소토스 증후군 2; 경직성 운동실조 샤를부아-사기네이 유형; 경직성 하반신마비 11, 상염색체 열성; 경직성 하반신마비 30, 상염색체 열성; 경직성 하반신마비 4, 상염색체 우성; 경직성 하반신마비 54, 상염색체 열성; 경직성 하반신마비 6; 경직성 하반신마비 7; 경직성 하반신마비 8; 정자발생 부전 8; 구상적혈구증 유형 4; 스펡고지질 활성화제 단백질 1 결핍; 스펡고미엘린/콜레스테롤 지질증; 척수성 근육 위축, 하지 우세형 2, 상염색체 우성; 척수성 근육 위축, 유형 II; 척수소뇌성 운동실조 14; 척수소뇌성 운동실조 21; 척수소뇌성 운동실조 35; 척수소뇌성 운동실조 38; 척수소뇌성 운동실조, 상염색체 열성 12; 척추늑골 이골증 2; 관절 이완 동반 척추골단골간단 이형성증; 척추골단골간단 이형성증, 파키스탄 유형; 선천성 척추골단 이형성증; 추체-간체 이영양증 동반 척추골간단 이형성증; 두경부 편평 세포 암종; 스타르가르트병 1; 스타르가르트병 3; 스틸 증후군; 스틸클러 증후군 유형 1; 스틸프 피부 증후군; 자상-연관 혈관병증, 영아-발병; 아 급성 뉴런병증성 고셔병; 숙시닐-CoA 아세토아세테이트 트랜스퍼라제 결핍; 슈퍼옥시드 디스뮤타제, 상승된 세포외; 판막위 대동맥 협착; 지질유합증-단지증 증후군; 합지증 유형 9; 탄지에르병; 족근 수근 융합 증후군 (Tarsal carpal coalition syndrome); 테이-삭스병; 테이-삭스병, B1 변이형; T-세포 전립선구성 백혈병; 탬플-바레이처 증후군; 탬타미 축전성 단지증 증후군; 팔로 사징증; 흉부 대동맥류 및 대동맥 박리; 혈소판감소증 2; 혈소판감소증, X-연관; 혈소판감소증, X-연관, 간혈성; 활성화 단백질 C 저항성에 기인한 혈전성향증; 혈전성향증, 유전성, 단백질 C 결핍에 기인, 상염색체 우성; 혈전성향증, 유전성, 단백질 C 결핍에 기인, 상염색체 열성; 갑상선암, 비수질성, 4; 갑상선 호르몬생성이상 1; 갑상선종독성 주기성 마비; 티에즈 증후군; 치아 무발생증, 선택적, 3; 치아 무발생증, 선택적, X-연관, 1; 일과성 신생아 당뇨병 1; 일과성 신생아 당뇨병 2; 트리처 콜린스 증후군 2; 모발비지질 이형성증 유형 I; 어린선 동반 트리글리세리드 축적 질환; 트리오스포스페이트 이소머라제 결핍; 삼지골 무지; 결절성 경화증 1; 결절성 경화증 2; 결절성 경화증 증후군; 티로시나제-음성 눈 피부 백색증; 티로시나제-양성 눈 피부 백색증; 티로신혈증 유형 2; 울리히 선천성 근육 이영양증; 비분류; 운베리히트-룬트보르크 증후군; 업쇼-술만 증후군; 우리던 5-프라임 모노포스페이트 히드롤라제 결핍, 용혈성 빈혈에 기인; 어서 증후군, 유형 1D; 어서 증후군, 유형 1F; 어서 증후군, 유형 2A; 반더우드 증후군; 혼합 포르피린증; 대두증 및 뇌실확장증 동반 VATER 연합증; 심실 중격 결손 3; 비타민 D-의존성 구루병, 유형 1; 비타민 D-의존성 구루병, 유형 2; 비타민 k-의존성 응고 인자, 복합 결핍, 1; 난황상 이영양증; 폰 히켈-린다우 증후군; 폰 빌레브란트병, 유형 2b; 바르덴부르크 증후군 유형 1; 바르덴부르크 증후군 유형 2E, 신경계 침범 비동반; 바르덴부르크 증후군 유형 4A; 바르덴부르크 증후군 유형 4B; 바르덴부르크 증후군 유형 4C; 마커-바르부르크 선천성 근육 이영양증; 바르부르크 마이크로 증후군 3; 사마귀, 저감마글로불린혈증, 감염, 및 골수배출 장애(myelokathexis); 베르드니히-호프만병; 베르너 증후군; 비아커 증후군; 비데만-스타이너 증후군; 윈체스터 증후군; 울프람 증후군 2; 제로시토시스(Xerocytosis); 색소성 건피증, D군; 색소성 건피증, 군 G; X-연관 무감마글로불린혈증; X-연관 유전성 운동 및 감각 신경병증; 스테릴-술파타제 결핍 동반 X-연관 어린선; X-연관 정신 지체 41; X-연관 정신 지체 90; X-연관 뇌실주위 이소증; 짐메르만-라반트 증후군; 또는 짐메르만-라반트 증후군 2이다.

[0371] 일부 실시양태에서, 표적 DNA 서열은 질환 또는 장애와 연관된 서열을 포함한다. 일부 실시양태에서, 표적 DNA 서열은 질환 또는 장애와 연관된 점 돌연변이를 포함한다. 일부 실시양태에서, 질환 또는 장애와 연관된 점 돌연변이는 질환 또는 장애와 연관된 유전자에서의 것이다. 일부 실시양태에서, 질환 또는 장애와 관련된 유전자는 AARS2, AASS, ABCA1, ABCA4, ABCB11, ABCB6, ABCC6, ABCC8, ABCD1, ABCG8, ABHD12, ABHD5, ACADM, ACAT1, ACE, ACO2, ACTA1, ACTB, ACTG1, ACTN2, ACVR1, ACVRL1, ADA, ADAMTS13, ADAR, ADGRG1, ADSL, AFF4, AGA, AGBL1, AGL, AGPAT2, AGRN, AGXT, AIPL1, AKR1D1, ALAD, ALAS2, ALDH3A2, ALDH7A1, ALDOB, ALG1, ALPL, ALS2, ALX3, ALX4, AMPD2, AMT, ANKS6, ANO5, APC, APOA1, APOE, APP, APRT, AQP2, AR, ARHGEF9, ARID2, ARL6, ARSA, ARSB, ARSE, ARX, ASAH1, ASB10, ASPM, ATF6, ATL1, ATM, ATP13A2, ATP1A3, ATP6V1B2, ATP7A, ATR,

ATRX, AVP, B2M, B3GALT6, BAAT, BARD1, BBS10, BBS12, BBS2, BBS4, BBS9, BCKDHA, BCKDHB, BCS1L, BEST1, BHLHA9, BICD2, BLM, BMP1, BMP4, BMPR2, BRAF, BRCA1, BRCA2, BRIP1, BTB, BTK, C10orf2, C1GALT1C1, C5orf42, C9, CA1, CACNA1S, CALM2, CANT1, CAPN3, CASK, CASQ2, CASR, CAV3, CBS, CCBE1, CCDC39, CD40LG, CDC6, CDC73, CDH1, CDH23, CDKL5, CDKN2A, CDON, CECR1, CENPJ, CEP120, CEP83, CFP, CFTR, CHAT, CHCHD10, CHD7, CHRNA1, CHRN2, CHRNG, CHST14, CHSY1, CLCN1, CLCN2, CLCN5, CLCNKA, CLDN16, CLDN19, CLIC2, CLN6, CLN8, CNGA3, CNM2, CNTNAP2, COA5, COL11A1, COL1A1, COL1A2, COL27A1, COL2A1, COL3A1, COL4A1, COL4A5, COL5A1, COL5A2, COL6A1, COL6A3, COL7A1, COLQ, COMP, CP, CPOX, CPT1A, CPT2, CR2, CRADD, CREBBP, CRH, CRX, CRYAB, CSF1R, CSTB, CTH, CTLA4, CTNS, CTPS1, CTSC, CTSD, CTSF, CTSK, CUL3, CXCR4, CYBB, CYP1B1, CYP27A1, CYP27B1, CYP4F22, CYP4V2, CYP7B1, DARS2, DBT, DCLRE1C, DCX, DDHD2, DES, DGUOK, DHCR24, DHCR7, DKC1, DLG3, DLL4, DMD, DMP1, DNAH11, DNAH5, DNAJB6, DNAJC19, DNM1, DNM2, DNMT1, DOCK6, DOK7, DOLK, DPAGT1, DPM2, DSC2, DSP, DYNC1H1, DYNC2H1, DYRK1A, DYSF, ECEL1, ECHS1, EDA, EDN3, EEF1A2, EFHC1, EFTUD2, EGLN1, EHMT1, EIF2B5, ELN, ELOVL4, ELOVL5, EMP2, ENPP1, EOGT, ERCC2, ERCC8, ESCO2, ETFDH, EXOSC3, EXOSC8, EXT2, EYA1, EYS, F12, F2, F5, F8, F9, FAM20C, FANCA, FANCF, FANCG, FAS, FBLN5, FBN1, FBN2, FBP1, FBXL4, FCGR3B, FGF8, FGFR1, FGFR2, FGFR3, FH, FHL1, FKTN, FLCN, FLG, FLNA, FLNB, FLT4, FLVCR2, FOXC1, FOXE1, FOXG1, FOXL2, FRAS1, FRMD7, FTL, FUS, G6PC3, G6PD, GAA, GABRA1, GABRG2, GAD1, GALT, GALNS, GALT, GAMT, GARS, GATA1, GATA6, GBA, GBA2, GBE1, GCDH, GCH1, GCK, GDAP1, GDI1, GFAP, GGCX, GHR, GJA8, GJB1, GJB2, GK, GLB1, GLI3, GLRA1, GMPBB, GNAI3, GNAS, GNAT1, GNE, GNPTAB, GNPTG, GPI, GPIHBP1, GPT2, GRIA3, GRIN2A, GRIN2B, GRIP1, GRN, GSC, GUCY2D, GYG1, GYS2, H6PD, HADHB, HBB, HBD, HBG1, HBG2, HCN1, HCN4, HESX1, HEXA, HFE, HFM1, HGSNAT, HINT1, HK1, HMGCL, HNF1A, HNF1B, HOGA1, HOXA1, HPD, HPGD, HPRT1, HR, HSD17B10, HSPB1, IDS, IDUA, IFT122, IFT80, IGHMBP2, IKBKG, IL11RA, IL12RB1, IMPDH1, IMPG2, INF2, ING1, INPPL1, INSL3, INSR, IRF6, IRX5, ISPD, ITGA2B, ITGB3, ITK, JAGN1, KCNA1, KCNH1, KCNH2, KCNJ1, KCNJ10, KCNJ11, KCNJ18, KCNJ2, KCNJ5, KCNK3, KCNQ1, KCNQ2, KCNQ4, KDM5C, KIAA0196, KIAA0586, KIF11, KIF1A, KIF2A, KISS1, KISS1R, KLF1, KMT2A, KMT2D, KRAS, KRIT1, KRT1, KRT5, KRT6A, LAMA1, LAMA2, LAMB2, LAMB3, LAMP2, LBR, LCT, LDLR, LIPA, LITAF, LMBR1, LMNA, LPIN2, LPL, LRIT3, LRP5, LRRC6, LRTOMT, LYST, LYZ, MAD1L1, MAF, MALT1, MAN2B1, MAPK1, MASTL, MATN3, MC2R, MCCC1, MCCC2, MCFD2, MCM8, MCOLN1, MCPH1, MECP2, MEF2C, MEFV, MEN1, MESP2, MET, MFN2, MFSB8, MGAT2, MITF, MKKS, MLH1, MLYCD, MMACHC, MMP14, MOG, MPL, MPV17, MPZ, MRE11A, MRPL3, MSH2, MSH6, MSR1, MSX1, MT-ATP6, MTHFR, MTM1, MT-ND1, MTR, MUSK, MUT, MYBPC3, MYC, MYH7, MYL2, MYL3, MYO1E, MYOC, NAGA, NAGLU, NARS2, NBEAL2, NBN, NDP, NDUFA1, NDUFA13, NDUFAF3, NDUFS8, NEFL, NEU1, NEXN, NFIX, NHEJ1, NHLRC1, NIPA1, NIPBL, NKX2-5, NLRP3, NMNAT1, NNT, NOBOX, NOG, NOL3, NOTCH3, NPC1, NPR2, NROB1, NR3C2, NR5A1, NRXN1, NSD1, NSDHL, NT5C3A, NYX, OAT, OCA2, OCRL, OFD1, OPA3, OPCML, OSMR, OTC, OTOF, OTX2, OXCT1, PAFAH1B1, PAH, PAK3, PALB2, PANK2, PAPSS2, PARK7, PAX2, PAX3, PAX6, PAX9, PCCA, PCCB, PCDH15, PCDH19, PCYT1A, PDE4D, PDE6A, PDE6B, PDE6C, PDE6H, PDGFB, PDHA1, PET100, PEX10, PEX7, PGK1, PGM1, PGM3, PHGDH, PHKB, PHOX2B, PIEZO1, PIGM, PITPNM3, PITX2, PKHD1, PKP2, PLA2G6, PLK4, PLOD1, PLP1, PMM2, PMP22, PMS2, PNPLA6, POLG, POLG2, POLR1A, POLR1D, POLR3A, POLR3B, POMT1, POMT2, POR, POU1F1, PPOX, PPT1, PRKACG, PRKAG2, PRKAR1A, PRKCG, PRNP, PROC, PROK2, PROKR2, PRPF31, PRPS1, PRSS56, PSAP, PSEN1, PTEN, PTPN11, PURA, PVRL4, PYGL, PYGM, RAB18, RAB27A, RAB7A, RAD21, RAD51C, RAF1, RAG2, RAX, RAX2, RB1, RBM8A, RDH12, RET, RHO, RIT1, RNF216, ROGDI, RP2, RPGR, RPS6KA3, RRM2B, RSPO4, RUNX1, RUNX2, RYR1, RYR2, SACS, SAMHD1, SBDS, SCN11A, SCN1A, SCN2A, SCN5A, SCN8A, SCNN1B, SDHAF1, SDHB, SDHD, SEMA4A, SEPNI1, SERPINF1, SERPING1, SETBP1, SGCB, SGCD, SH2D1A, SH3TC2, SHANK3, SHH, SHOX, SIGMAR1, SIX3, SKI, SLC11A2, SLC17A5, SLC19A3, SLC1A3, SLC22A5, SLC25A13, SLC25A15, SLC25A19, SLC25A22, SLC25A38, SLC25A4, SLC26A4, SLC2A10, SLC2A9, SLC33A1, SLC35C1, SLC39A4, SLC46A1, SLC4A1, SLC52A2, SLC52A3, SLC5A5, SLC6A5, SLC6A8, SLC9A3R1, SMAD2, SMAD4, SMARCA2, SMARCA4, SMN1, SMPD1, SNCA, SNRNP200, SNRNPB, SOD1, SOD3, SOX9, SPAST, SPATA5, SPG11, SPG7, SPTB, SRD5A2, SRY, STAC3, STAR, STAT1, STAT3, STAT5B, STK11, STS, STX1B, STXBP1, SUCLG1, SUMF1, TARDBP, TAZ, TBC1D24, TBX1, TBX20, TCF12, TCF4, TECTA, TERC, TERT, TFAP2B, TFR2, TGFB3, TGFB1, TGFB2, TGIF1, TGM1, TGM5, TGM6, THRA, THRB, TIMM8A, TK2, TMEM173, TMEM240, TMEM98, TMPRSS15, TMPRSS3, TMPRSS6, TNFRSF11A, TNNT1, TOR1A, TP53, TP63, TPI1, TPM1, TPM2, TPM3, TPO, TPP1, TRIM37, TRNT1, TRPM6, TRPS1, TSC1, TSC2, TSHR, TSPAN12, TTPA, TTR, TUBB4A, TULP1, TYMP, TYR, TYRP1, UBE2T, UBE3A, UBIAD1, UMOD, UMPS, UROD, USH2A, USP8, VDR, VHL, VPS13B, VPS33B, VWF, WAS, WDR19, WDR45, WDR62, WDR72, WFS1, WNK4, WNT5A, WRN, WT1, WWOX, ZBTB20, ZC4H2, ZDHHC9, ZEB2, ZFP57, ZIC3, 또는 ZNF469로 이루어진 군으로부터 선택된다.

- [0372] 일부 실시양태는 본원에 제공된 DNA 편집 융합 단백질을 사용하는 방법을 제공한다. 일부 실시양태에서, 융합 단백질은 표적 핵염기, 예를 들어 C 잔기를 탈아미노화함으로써 점 돌연변이를 핵산 내로 도입하는데 사용된다. 일부 실시양태에서, 융합 단백질을 사용하여 표적 C를 U로 탈아미노화한 다음, 이를 제거하여 이전에 C 잔기에 의해 점유된 무염기성 부위를 생성한다. 일부 실시양태에서, 표적 핵염기의 탈아미노화는 유전적 결함의 교정, 예를 들어 유전자 산물에서의 기능 상실을 초래하는 점 돌연변이의 교정을 유발한다. 일부 실시양태에서, 본원에 제공된 방법은 탈활성화 점 돌연변이를 질환 또는 장애와 연관된 유전자 산물을 코딩하는 유전자 또는 대립 유전자 내로 도입하는데 사용된다. 예를 들어, 일부 실시양태에서, DNA 편집 융합 단백질을 사용하여 탈활성화 점 돌연변이를 종양유전자 내로 도입하는 (예를 들어, 증식성 질환의 치료에서) 방법이 본원에 제공된다. 탈활성화 돌연변이는 일부 실시양태에서, 코딩 서열에서의 조기 정지 코돈을 생성할 수 있으며, 이는 말단절단된 유전자 산물, 예를 들어 전장 단백질의 기능이 결여된 말단절단된 단백질의 발현을 유발한다.
- [0373] 일부 실시양태에서, 본원에 제공된 방법의 목적은 게놈 편집을 통해 기능이상 유전자의 기능을 회복시키는 것이다. 본원에 제공된 핵염기 편집 단백질은, 예를 들어 인간 세포 배양물에서 질환-연관 돌연변이를 교정함으로써, 시험관내에서 유전자 편집-기반 인간 치료제에 대해 검증될 수 있다. 본원에 제공된 핵염기 편집 단백질, 예를 들어 핵산 프로그램가능 DNA 결합 단백질 (예를 들어, Cas9), 시티딘 데아미나제 및 우라실 결합 단백질을 포함하는 융합 단백질은 임의의 단일 점 C에서 G로 또는 G에서 C로의 돌연변이를 교정하는데 사용될 수 있다는 것이 통상의 기술자에 의해 이해될 것이다. 전자의 경우에, 돌연변이체 C에서 U로의 탈아미노화 및 후속하여 U의 절제는 돌연변이를 교정하고, 후자의 경우에, C에서 U로의 탈아미노화 및 후속하여 돌연변이체 G와 염기-쌍 형성되는 U의 절제, 이어서 복제의 라운드는 돌연변이를 교정한다.
- [0374] 질환-연관 유전자 및 대립유전자에서의 점 돌연변이의 성공적인 교정은 치료제 및 기초 연구에서의 적용과 함께 유전자 교정을 위한 새로운 전략을 열고 있다. 핵산 프로그램가능 DNA 결합 단백질 (napDNAbp), 시티딘 데아미나제 및 우라실 결합 단백질을 포함하는 개시된 융합 단백질과 같은 부위-특이적 단일-염기 변형 시스템은 또한 "역" 유전자 요법에서의 적용을 가지며, 여기서 특정 유전자 기능은 고의로 억제되거나 폐지된다. 이들 경우에, 단백질 내 불활성화 돌연변이 또는 단백질의 기능을 억제하는 돌연변이를 유도하는 부위-특이적 돌연변이 잔기는 시험관내, 생체외 또는 생체내에서 단백질 기능을 폐지하거나 억제하는데 사용될 수 있다.
- [0375] 본 개시내용은 본원에 제공된 DNA 편집 융합 단백질에 의해 교정될 수 있는 점 돌연변이와 연관되거나 그에 의해 유발되는 질환을 갖는 것으로 진단된 대상체의 치료 방법을 제공한다. 예를 들어, 일부 실시양태에서, 이러한 질환, 예를 들어 상기 기재된 바와 같은 점 돌연변이와 연관된 암을 갖는 대상체에게, 점 돌연변이 (예를 들어, C에서 G로 또는 G에서 C로의 점 돌연변이)를 교정하거나 탈활성화 돌연변이를 질환-연관 유전자 내로 도입하는 염기 편집제 융합 단백질의 유효량을 투여하는 것을 포함하는 방법이 제공된다. 일부 실시양태에서, 질환은 증식성 질환이다. 일부 실시양태에서, 질환은 유전 질환이다. 일부 실시양태에서, 질환은 신생물성 질환이다. 일부 실시양태에서, 질환은 대사 질환이다. 일부 실시양태에서, 질환은 리소솜 축적 질환이다. 점 돌연변이를 교정하거나 탈활성화 돌연변이를 질환-연관 유전자 내로 도입함으로써 치료될 수 있는 다른 질환은 관련 기술분야의 통상의 기술자에게 공지되어 있을 것이며, 본 개시내용은 이에 관하여 제한되지 않는다.
- [0376] 본 개시내용은 병원성 G에서 C로 또는 C에서 G로의 돌연변이를 포함하는 유전자의 목록을 제공한다. 이러한 병원성 G에서 C로 또는 C에서 G로의 돌연변이는 본원에 제공된 방법 및 조성물을 사용하여, 예를 들어 C를 G로 및/또는 G를 C로 돌연변이시켜 유전자 기능을 회복시킴으로써 교정될 수 있다.
- [0377] 일부 실시양태에서, 융합 단백질은 정규 PAM을 인식하고 따라서 플랭킹 서열에서 각각 정규 PAM, 예를 들어 NGG에 의해 병원성 G에서 C로 또는 C에서 G로의 돌연변이를 교정할 수 있다. 예를 들어, 정규 PAM을 인식하는 Cas9 단백질은 서열식별번호: 6에 의해 제공된 바와 같은 스트렙토코쿠스 피오게네스 Cas9의 아미노산 서열에 대해, 또는 서열식별번호: 6의 RuvC 및 HNH 도메인을 포함하는 그의 단편에 대해 적어도 80%, 85%, 90%, 95%, 97%, 98% 또는 99% 동일한 아미노산 서열을 포함한다.
- [0378] napDNAbp (예를 들어, Cas9 도메인)를 포함하는, 본원에 제공된 융합 단백질 중 임의의 것을 표적 부위, 예를 들어 편집될 점 돌연변이를 포함하는 부위에 표적화하기 위해, 전형적으로 융합 단백질을 가이드 RNA, 예를 들어 sgRNA와 함께 공-발현시킬 필요가 있다는 것은 관련 기술분야의 통상의 기술자에게 분명할 것이다. 본원의 다른 곳에서 보다 상세히 설명된 바와 같이, 가이드 RNA는 전형적으로 Cas9 결합을 허용하는 tracrRNA 프레임워크, 및 가이드 서열을 포함하며, 이는 Cas9:핵산 편집 효소/도메인 융합 단백질에 서열 특이성을 부여한다. 일부 실시양태에서, 가이드 RNA는 구조 5'-[가이드 서열]-guuuuagagcuagaaauagcaaguuuuuuuuagggcuaguccguuucaacuugaaaaaguggcaccgagucggugcuuuuu-3' (서열식별번호:

119)를 포함하며, 여기서 가이드 서열은 표적 서열에 상보적인 서열을 포함한다. 일부 실시양태에서, 가이드 서열은 표적 핵산에 상보적인 핵산 서열을 포함한다. 가이드 서열은 전형적으로 20개 뉴클레오타이드의 길이이다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특이적 게놈 표적 부위에 표적화하는데 적합한 가이드 RNA의 서열은 본 개시내용에 기초하여 관련 기술분야의 통상의 기술자에게 분명할 것이다. 이러한 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오타이드의 50개 뉴클레오타이드 상류 또는 하류 내의 핵 서열에 상보적인 가이드 서열을 포함한다.

[0379] 염기 편집제 효율

[0380] 본 개시내용의 일부 측면은 본원에 제공된 염기 편집제 중 임의의 것이 유의한 비율의 삽입결실을 생성하지 않으면서 특이적 뉴클레오타이드 염기를 변형시킬 수 있다는 인식에 기초한다. 본원에 사용된 "삽입결실"은 핵산 내의 뉴클레오타이드 염기의 삽입 또는 결실을 지칭한다. 이러한 삽입 또는 결실은 유전자의 코딩 영역 내의 프레임 시프트 돌연변이를 초래할 수 있다. 일부 실시양태에서, 핵산에서 다수의 삽입 또는 결실 (즉, 삽입결실)을 생성하지 않으면서, 핵산 내의 특이적 뉴클레오타이드를 효율적으로 변형시키는 (예를 들어 돌연변이시키거나 탈아미노화하는) 염기 편집제를 생성하는 것이 바람직하다. 특정 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 보다 큰 비율의 의도된 변형 (예를 들어, 점 돌연변이 또는 탈아미노화) 대 삽입결실을 생성할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제는 1:1 초과인 의도된 점 돌연변이 대 삽입결실의 비를 생성할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제는 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5:1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 200:1, 적어도 300:1, 적어도 400:1, 적어도 500:1, 적어도 600:1, 적어도 700:1, 적어도 800:1, 적어도 900:1 또는 적어도 1000:1 또는 그 초과인 의도된 점 돌연변이 대 삽입결실의 비를 생성할 수 있다. 의도된 돌연변이 및 삽입결실의 수는 임의의 적합한 방법, 예를 들어 하기 실시예에 사용된 방법을 사용하여 결정될 수 있다. 일부 실시양태에서, 삽입결실 빈도를 계산하기 위해, 서열분석 판독물을 삽입결실이 발생할 수 있는 윈도우의 양쪽 측면에 플랭킹되어 있는 2개의 10-bp 서열에 대한 정확한 매치에 대해 스캐닝한다. 정확한 매치가 위치하지 않는 경우에, 판독물을 분석으로부터 제외한다. 이 삽입결실 윈도우의 길이가 참조 서열에 정확하게 매치되는 경우에, 판독물을 삽입결실을 함유하지 않는 것으로서 분류한다. 삽입결실 윈도우가 참조 서열보다 2개 이상의 염기 더 길거나 더 짧은 경우에, 서열분석 판독물을 각각 삽입 또는 결실로서 분류한다.

[0381] 일부 실시양태에서, 본원에 제공된 염기 편집제는 핵산의 영역에서의 삽입결실의 형성을 제한할 수 있다. 일부 실시양태에서, 영역은 염기 편집제에 의해 표적화되는 뉴클레오타이드, 또는 염기 편집제에 의해 표적화되는 뉴클레오타이드의 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 뉴클레오타이드 내의 영역에 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 핵산의 영역에서의 삽입결실의 형성을 1% 미만, 1.5% 미만, 2% 미만, 2.5% 미만, 3% 미만, 3.5% 미만, 4% 미만, 4.5% 미만, 5% 미만, 6% 미만, 7% 미만, 8% 미만, 9% 미만, 10% 미만, 12% 미만, 15% 미만 또는 20% 미만으로 제한할 수 있다. 핵산 영역에서 형성된 삽입결실의 수는 핵산 (예를 들어, 세포의 게놈 내의 핵산)이 염기 편집제에 노출되는 시간의 양에 의존할 수 있다. 일부 실시양태에서, 삽입결실의 수 또는 비율은 염기 편집제에 핵산 (예를 들어, 세포의 게놈 내의 핵산)을 적어도 1시간, 적어도 2시간, 적어도 6시간, 적어도 12시간, 적어도 24시간, 적어도 36시간, 적어도 48시간, 적어도 3일, 적어도 4일, 적어도 5일, 적어도 7일, 적어도 10일 또는 적어도 14일 노출시킨 후에 결정된다.

[0382] 본 개시내용의 일부 측면은 본원에 제공된 염기 편집제 중 임의의 것이 유의한 수의 비의도된 돌연변이, 예컨대 비의도된 점 돌연변이를 생성하지 않으면서 핵산 (예를 들어 대상체의 게놈 내의 핵산)에서 의도된 돌연변이, 예컨대 점 돌연변이를 효율적으로 생성할 수 있다는 인식에 기초한다. 일부 실시양태에서, 의도된 돌연변이는 의도된 돌연변이를 생성하도록 특이적으로 설계된 gRNA에 결합된 특이적 염기 편집제에 의해 생성된 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 질환 또는 장애와 연관된 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 질환 또는 장애와 연관된 시토신 (C)에서 구아닌 (G)으로의 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 질환 또는 장애와 연관된 구아닌 (G)에서 시토신 (C)으로의 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 유전자의 코딩 영역 내의 시토신 (C)에서 구아닌 (G)으로의 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 유전자의 코딩 영역 내의 구아닌 (G)에서 시토신 (C)으로의 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 정지 코돈, 예를 들어 유전자의 코딩 영역 내의 조기 정지 코돈을 생성하는 점 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 정지 코돈을 제거하는 돌연변이이다. 일부 실시양태에서, 의도된 돌연변이는 유전자의 스플라이싱을 변경시키는 돌연변이이다. 일부 실시양태에서,

의도된 돌연변이는 유전자의 조절 서열 (예를 들어, 유전자 프로모터 또는 유전자 리프레서)을 변경시키는 돌연변이이다. 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 1:1 초과인 의도된 돌연변이 대 비의도된 돌연변이 (예를 들어, 의도된 점 돌연변이:비의도된 점 돌연변이)의 비를 생성할 수 있다. 일부 실시양태에서, 본원에 제공된 염기 편집제 중 임의의 것은 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5:1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 150:1, 적어도 200:1, 적어도 250:1, 적어도 500:1 또는 적어도 1000:1 또는 그 초과인 의도된 돌연변이 대 비의도된 돌연변이 (예를 들어, 의도된 점 돌연변이:비의도된 점 돌연변이)의 비를 생성할 수 있다. 본원에서 "염기 편집제 효율" 섹션에 기재된 염기 편집제의 특징은 본원에 제공된 융합 단백질 또는 융합 단백질을 사용하는 방법 중 임의의 것에 적용될 수 있다는 것이 인지되어야 한다.

[0383] 핵산을 편집하는 방법

[0384] 본 개시내용의 일부 측면은 핵산을 편집하는 방법을 제공한다. 일부 실시양태에서, 방법은 핵산의 핵염기 (예를 들어, 이중-가닥 DNA 서열의 염기 쌍)를 편집하는 방법이다. 일부 실시양태에서, 방법은 하기 단계를 포함한다: a) 핵산의 표적 영역 (예를 들어, 이중-가닥 DNA 서열)을 염기 편집제 (예를 들어, 시티딘 데아미나제 및 우라실 결합 단백질에 융합된 Cas9 도메인) 및 가이드 핵산 (예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계이며, 여기서 표적 영역은 표적화된 핵염기 쌍을 포함하는 것인 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 표적 영역의 단일 가닥에서의 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계, d) 제2 핵염기를 절제하여 무염기성 부위를 생성하는 단계, 및 e) 제1 핵염기 염기에 상보적인 제3 핵염기를 시토신 (C)인 제4 핵염기로 대체하는 단계. 일부 실시양태에서, 방법은 핵산에서 20% 미만의 삽입결실 형성을 유발한다. 일부 실시양태에서, 단계 b는 생략된다는 것이 인지되어야 한다. 일부 실시양태에서, 제1 핵염기는 시토신 (C)이다. 일부 실시양태에서, 제2 핵염기는 탈아미노화된 시토신 또는 우라실이다. 일부 실시양태에서, 제3 핵염기는 구아닌 (G)이다. 일부 실시양태에서, 제4 핵염기는 시토신 (C)이다. 일부 실시양태에서, 제5 핵염기는 단계 (d)에서 생성된 무염기성 부위 내로 라이게이션된다. 일부 실시양태에서, 제5 핵염기는 구아닌 (G)이다. 일부 실시양태에서, 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만 또는 0.1% 미만의 삽입결실 형성을 유발한다. 일부 실시양태에서, 의도된 염기 쌍의 적어도 5%가 편집된다. 일부 실시양태에서, 의도된 염기 쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45% 또는 50%가 편집된다.

[0385] 일부 실시양태에서, 표적 뉴클레오타이드에서의 의도된 생성물 대 비의도된 생성물의 비는 적어도 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1 또는 200:1 또는 그 초과이다. 일부 실시양태에서, 의도된 점 돌연변이 대 삽입결실 형성의 비는 1:1, 10:1, 50:1, 100:1, 500:1 또는 1000:1 또는 그 초과보다 더 크다. 일부 실시양태에서, 절단된 단일 가닥 (니킹된 가닥)은 가이드 핵산에 혼성화된다. 일부 실시양태에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥의 반대편이다. 일부 실시양태에서, 염기 편집제는 Cas9 도메인을 포함한다. 일부 실시양태에서, 염기 편집제는 니카제 활성을 포함한다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 상류에 있다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오타이드 상류에 있다. 일부 실시양태에서, 의도된 편집된 염기쌍은 PAM 부위의 하류에 있다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오타이드 하류에 있다. 일부 실시양태에서, 방법은 정규 (예를 들어, NGG) PAM 부위를 요구하지 않는다. 일부 실시양태에서, 핵염기 편집제는 링커를 포함한다. 일부 실시양태에서, 링커는 1 내지 25개 아미노산의 길이이다. 일부 실시양태에서, 링커는 5 내지 20개 아미노산의 길이이다. 일부 실시양태에서, 링커는 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 아미노산의 길이이다. 일부 실시양태에서, 표적 영역은 표적 윈도우를 포함하며, 여기서 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 실시양태에서, 표적 윈도우는 1-10개의 뉴클레오타이드를 포함한다. 일부 실시양태에서, 표적 윈도우는 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2 또는 1개 뉴클레오타이드의 길이이다. 일부 실시양태에서, 표적 윈도우는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오타이드의 길이이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 표적 윈도우 내에 있다. 일부 실시양태에서, 표적 윈도우는 의도된 편집된 염기 쌍을 포함한다. 일부 실시양태에서, 방법은 본원에 제공된 염기 편집제 중 임의의 것을 사용하여 수행된다. 일부 실시양태에서, 표적 윈도우는 탈아미노화 윈도우이다.

[0386] 일부 실시양태에서, 본 개시내용은 뉴클레오티드를 편집하는 방법을 제공한다. 일부 실시양태에서, 본 개시내용은 이중-가닥 DNA 서열의 핵염기 쌍을 편집하는 방법을 제공한다. 일부 실시양태에서, 방법은 a) 이중-가닥 DNA 서열의 표적 영역을 염기 편집제 및 가이드 핵산 (예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계이며, 여기서 표적 영역은 표적 핵염기 쌍을 포함하는 것인 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 표적 영역의 단일 가닥에서의 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계, d) 제2 핵염기를 절제하여 무염기성 부위를 생성하는 단계, 및 e) 제1 핵염기 염기에 상보적인 제3 핵염기를 시토신 (C)인 제4 핵염기로 대체하여 의도된 편집된 염기 쌍을 생성하는 단계를 포함하며, 여기서 의도된 편집된 염기 쌍을 생성하는 효율은 적어도 5%이다. 일부 실시양태에서, 단계 b는 생략된다는 것이 인지되어야 한다. 일부 실시양태에서, 의도된 염기 쌍의 적어도 5%가 편집된다. 일부 실시양태에서, 의도된 염기 쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45% 또는 50%가 편집된다. 일부 실시양태에서, 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만 또는 0.1% 미만의 삽입결실 형성을 유발한다. 일부 실시양태에서, 표적 뉴클레오티드에서의 의도된 생성물 대 비의도된 생성물의 비는 적어도 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1 또는 200:1 또는 그 초과보다 더 크다. 일부 실시양태에서, 의도된 점 돌연변이 대 삽입결실 형성의 비는 1:1, 10:1, 50:1, 100:1, 500:1 또는 1000:1 또는 그 초과보다 더 크다. 일부 실시양태에서, 절단된 단일 가닥은 가이드 핵산에 혼성화된다. 일부 실시양태에서, 핵염기 편집제는 니카제 활성을 포함한다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 상류에 있다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오티드 상류에 있다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 하류에 있다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 PAM 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오티드 하류에 있다. 일부 실시양태에서, 방법은 정규 (예를 들어, NGG) PAM 부위를 요구하지 않는다. 일부 실시양태에서, 핵염기 편집제는 링커를 포함한다. 일부 실시양태에서, 링커는 1 내지 25개 아미노산의 길이이다. 일부 실시양태에서, 링커는 5 내지 20개 아미노산의 길이이다. 일부 실시양태에서, 링커는 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 아미노산의 길이이다. 일부 실시양태에서, 표적 영역은 표적 윈도우를 포함하며, 여기서 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 실시양태에서, 표적 윈도우는 1-10개의 뉴클레오티드를 포함한다. 일부 실시양태에서, 표적 윈도우는 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2 또는 1개 뉴클레오티드의 길이이다. 일부 실시양태에서, 표적 윈도우는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오티드의 길이이다. 일부 실시양태에서, 의도된 편집된 염기 쌍은 표적 윈도우 내에서 발생한다. 일부 실시양태에서, 표적 윈도우는 의도된 편집된 염기 쌍을 포함한다. 일부 실시양태에서, 핵염기 편집제는 본원에 제공된 염기 편집제 중 어느 하나이다.

[0387] 제약 조성물

[0388] 본 개시내용의 다른 측면은 본원에 기재된 염기 편집제, 융합 단백질 또는 융합 단백질-gRNA 복합체 중 어느 것을 포함하는 제약 조성물에 관한 것이다. 본원에 사용된 용어 "제약 조성물"은 제약 용도를 위해 제제화된 조성물을 지칭한다. 일부 실시양태에서, 제약 조성물은 제약상 허용되는 담체를 추가로 포함한다. 일부 실시양태에서, 제약 조성물은 추가의 작용제 (예를 들어 특정 전달, 반감기 증가 또는 다른 치료 화합물을 위해)를 포함한다.

[0389] 본원에 사용된 용어 "제약상 허용되는 담체"는 신체의 한 부위 (예를 들어, 전달 부위)로부터 또 다른 부위 (예를 들어, 신체의 기관, 조직 또는 부분)로 화합물을 운반하거나 수송하는 데 수반되는 제약상 허용되는 물질, 조성물 또는 비히클, 예컨대 액체 또는 고체 충전제, 희석제, 부형제, 제조 보조제 (예를 들어, 윤활제, 활석, 스테아르산마그네슘, 스테아르산칼슘 또는 스테아르산아연, 또는 스테아르산), 또는 용매 캡슐화 물질을 의미한다. 제약상 허용되는 담체는 제제의 다른 성분과 상용성이고 대상체의 조직에 유해하지 않다는 (예를 들어, 생리학상 상용성, 멸균, 생리적 pH 등) 의미에서 "허용된다". 제약상 허용되는 담체로서의 역할을 할 수 있는 물질의 일부 예는 하기를 포함한다: (1) 당, 예컨대 락토스, 글루코스 및 수크로스; (2) 전분, 예컨대 옥수수 전분 및 감자 전분; (3) 셀룰로스 및 그의 유도체, 예컨대 소듐 카르복시메틸 셀룰로스, 메틸셀룰로스, 에틸 셀룰로스, 미세결정질 셀룰로스 및 셀룰로스 아세테이트; (4) 분말화 트라가칸트; (5) 맥아; (6) 젤라틴; (7) 윤활제, 예컨대 스테아르산마그네슘, 소듐 라우릴 황페이트 및 활석; (8) 부형제, 예컨대 코코아 버터 및 좌제 왁스; (9) 오일, 예컨대 땅콩 오일, 목화씨 오일, 홍화 오일, 참깨 오일, 올리브 오일, 옥수수 오일 및 대두 오일; (10) 글리콜, 예컨대 프로필렌 글리콜; (11) 폴리올, 예컨대 글리세린, 소르비톨, 만니톨 및 폴리에틸렌 글리콜 (PEG); (12) 에스테르, 예컨대 에틸 올레에이트 및 에틸 라우레이트; (13) 한천; (14) 완충제, 예컨대 수

산화마그네슘 및 수산화알루미늄; (15) 알긴산; (16) 발열원 무함유 물; (17) 등장성 염수; (18) 링거액; (19) 에틸 알콜; (20) pH 완충 용액; (21) 폴리에스테르, 폴리카르보네이트 및/또는 폴리무수물; (22) 벌킹제, 예컨대 폴리펩티드 및 아미노산 (23) 혈청 성분, 예컨대 혈청 알부민, HDL 및 LDL; (22) C2-C12 알콜, 예컨대 에탄올; 및 (23) 제약 제제에 사용된 다른 비-독성 상용성 물질. 습윤제, 착색제, 방출제, 코팅제, 감미제, 향미제, 펄프제, 보존제 및 항산화제가 또한 제제에 존재할 수 있다. "부형제", "담체", "제약상 허용되는 담체" 등과 같은 용어는 본원에서 상호교환가능하게 사용된다.

[0390] 일부 실시양태에서, 제약 조성물은 대상체에의 전달을 위해, 예를 들어 유전자 편집을 위해 제제화된다. 본원에 기재된 제약 조성물을 투여하는 적합한 경로는 비제한적으로 하기를 포함한다: 국소, 피하, 경피, 피내, 병변내, 관절내, 복강내, 방광내, 경점막, 치은, 치아내, 안구내, 경고실, 기관내, 경막외, 척수강내, 근육내, 정맥내, 혈관내, 골내, 안구주위, 종양내, 뇌내 및 뇌실내 투여.

[0391] 일부 실시양태에서, 본원에 기재된 제약 조성물은 이환 부위 (예를 들어, 종양 부위)에 국부로 투여된다. 일부 실시양태에서, 본원에 기재된 제약 조성물은 주사에 의해, 카테터에 의해, 좌제에 의해 또는 이식물에 의해 대상체에게 투여되며, 이식물은 막, 예컨대 실라스틱 막, 또는 섬유를 포함한, 다공성, 비-다공성 또는 젤라틴성 물질이다.

[0392] 다른 실시양태에서, 본원에 기재된 제약 조성물은 제어 방출 시스템으로 전달된다. 한 실시양태에서, 펌프가 사용될 수 있다 (예를 들어, 문헌 [Langer, 1990, Science 249:1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al., 1989, N. Engl. J. Med. 321:574] 참조). 또 다른 실시양태에서, 중합체 물질이 사용될 수 있다. (예를 들어, 문헌 [Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61] 참조. 또한, 문헌 [Levy et al., 1985, Science 228:190; During et al., 1989, Ann. Neurol. 25:351; Howard et al., 1989, J. Neurosurg. 71:105] 참조.). 다른 제어 방출 시스템은, 예를 들어 [Langer, 상기 문헌]에서 논의된다.

[0393] 일부 실시양태에서, 제약 조성물은 대상체, 예를 들어 인간에의 정맥내 또는 피하 투여에 적합화된 조성물로서 상용 절차에 따라 제제화된다. 일부 실시양태에서, 주사에 의한 투여를 위한 제약 조성물은 멸균 등장성 수성 완충제 중 용액이다. 필요한 경우, 제약은 또한 가용화제 및 주사 부위에서 통증을 완화시키기 위한 국부 마취제, 예컨대 리그노카인을 포함할 수 있다. 일반적으로, 성분은 개별적으로 또는 함께 혼합되어 단위 투여 형태로, 예를 들어 활성제의 양을 표시하는 기밀 용기, 예컨대 앰플 또는 사체 내의 건조 동결건조 분말 또는 무수 농축물로서 공급된다. 제약이 주입에 의해 투여될 경우에, 이는 멸균 제약 등급 물 또는 염수를 함유하는 주입 병으로 분배될 수 있다. 제약 조성물이 주사에 의해 투여되는 경우에, 성분이 투여 전에 혼합될 수 있도록 멸균 주사용수 또는 염수의 앰플이 제공될 수 있다.

[0394] 전신 투여를 위한 제약 조성물은 액체, 예를 들어 멸균 염수, 락테이트화된 링거액 또는 탕크 용액일 수 있다. 추가로, 제약 조성물은 고체 형태일 수 있고, 사용 직전 재-용해되거나 현탁될 수 있다. 동결건조된 형태가 또한 고려된다.

[0395] 제약 조성물은 지질 입자 또는 소포, 예컨대 리포솜 또는 미세결정 내에 함유될 수 있으며, 이는 또한 비경구 투여에 적합하다. 입자는 조성물이 그 안에 함유되는 한, 입자의 적합한 구조, 예컨대 단층 또는 복수층일 수 있다. 화합물은 융합생성 지질 디올레오일포스파티딜에탄올아민 (DOPE), 낮은 수준 (5-10 몰%)의 양이온성 지질을 함유하는 "안정화된 플라스미드-지질 입자" (SPLP)에 포획될 수 있고, 폴리에틸렌글리콜 (PEG) 코팅에 의해 안정화될 수 있다 (Zhang Y. P. et al., Gene Ther. 1999, 6:1438-47). 양으로 하전된 지질, 예컨대 N-[1-(2,3-디올레오일옥시)프로필]-N,N,N-트리메틸-암모늄메틸설페이트 또는 "DOTAP"는 이러한 입자 및 소포에 특히 바람직하다. 이러한 지질 입자의 제조는 널리 공지되어 있다. 예를 들어, 미국 특허 번호 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016 및 4,921,757 (이들 각각은 본원에 참조로 포함됨)을 참조한다.

[0396] 본원에 기재된 제약 조성물은, 예를 들어 단위 용량으로 투여되거나 포장될 수 있다. 본 개시내용의 제약 조성물에 관하여 사용되는 경우 용어 "단위 용량"은 대상체에 대한 단일 투여량으로서 적합한 물리적 이산 단위를 지칭하며, 각각의 단위는 요구되는 희석제; 즉, 담체 또는 비히클과 함께 목적하는 치료 효과를 생성하도록 계산된 미리 결정된 양의 활성 물질을 함유한다.

[0397] 추가로, 제약 조성물은 (a) 동결건조 형태의 본 발명의 화합물 (예를 들어, 융합 단백질 또는 염기 편집제)을

함유하는 용기 및 (b) 주사용 제약상 허용되는 희석제 (예를 들어, 멸균수)를 함유하는 제2 용기를 포함하는 제약 키트로서 제공될 수 있다. 제약상 허용되는 희석제는 본 발명의 동결건조 화합물의 재구성 또는 희석을 위해 사용될 수 있다. 인간 투여를 위한 제조, 사용 또는 판매의 기관에 의한 승인을 반영하는, 제약 또는 생물학적 제품의 제조, 사용 또는 판매를 관할하는 정부 기관에 의해 규정된 형태의 통지서가 이러한 용기(들)에 임의로 회합될 수 있다.

[0398] 또 다른 측면에서, 상기 기재된 질환의 치료에 유용한 물질을 함유하는 제조 물품이 포함된다. 일부 실시양태에서, 제조 물품은 용기 및 라벨을 포함한다. 적합한 용기는, 예를 들어 병, 바이알, 시린지 및 시험 튜브를 포함한다. 용기는 다양한 물질, 예컨대 유리 또는 플라스틱으로 형성될 수 있다. 일부 실시양태에서, 용기는 본원에 기재된 질환을 치료하는데 효과적인 조성물을 보유하며, 멸균 접근 포트를 가질 수 있다. 예를 들어, 용기는 피하 주사 바늘에 의해 뚫릴 수 있는 마개를 갖는 정맥내 용액 백 또는 바이알일 수 있다. 조성물 중 활성제는 본 발명의 화합물이다. 일부 실시양태에서, 용기 상의 또는 그와 회합된 라벨은 조성물이 선택 질환을 치료하는데 사용된다는 것을 표시한다. 제조 물품은 제약상 허용되는 완충제, 예컨대 포스페이트 완충염수, 링거액 또는 텍스트로스 용액을 함유하는 제2 용기를 추가로 포함할 수 있다. 이는 다른 완충제, 희석제, 필터, 바늘, 시린지, 및 사용에 대한 지침서를 갖는 패키지 삽입물을 포함한, 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[0399] 키트, 벡터, 세포

[0400] 본 개시내용의 일부 측면은 (a) 본원에 제공된 바와 같은 임의의 융합 단백질을 코딩하는 뉴클레오타이드 서열; 및 (b) (a)의 서열의 발현을 구도하는 이중 프로모터를 포함하는 핵산 구축물을 포함하는 키트를 제공한다. 일부 실시양태에서, 키트는 가이드 RNA 백본을 코딩하는 발현 구축물을 추가로 포함하며, 여기서 구축물은 표적 서열과 동일하거나 그에 상보적인 핵산 서열의 가이드 RNA 백본 내로의 클로닝을 허용하도록 위치된 클로닝 부위를 포함한다.

[0401] 본 개시내용의 일부 측면은 본원에 제공된 바와 같은 융합 단백질의 napDNAbp (예를 들어, Cas9 단백질)를 코딩하는 폴리뉴클레오타이드를 제공한다. 본 개시내용의 일부 측면은 이러한 폴리뉴클레오타이드를 포함하는 벡터를 제공한다. 일부 실시양태에서, 벡터는 폴리뉴클레오타이드의 발현을 구도하는 이중 프로모터를 포함한다.

[0402] 본 개시내용의 일부 측면은 본원에 제공된 임의의 융합 단백질, 본원에 제공된 임의의 융합 단백질을 코딩하는 핵산 분자, 본원에 제공된 임의의 융합 단백질 및 gRNA를 포함하는 복합체, 및/또는 본원에 제공된 임의의 벡터를 포함하는 세포를 제공한다.

[0403] 상기 리포터 시스템의 예시적인 실시양태의 설명은 단지 예시 목적으로 제공되며, 제한적인 것으로 의도되지 않는다. 추가적인 리포터 시스템, 예를 들어 상기 상세히 기재된 예시적인 시스템의 변형이 또한 본 개시내용에 포함된다.

[0404] 실시예

[0405] 무염기성 부위 생성 및 조작된 특이적 복구를 통한 시토신 (C)에서 구아닌 (G)으로의 염기 편집제

[0406] HEK2, RNF2 및 FANCF 부위에 대한 서열분석 데이터가 하기에 제시된다. 제시된 데이터는 윈도우에서 가장 많이 편집된 C에 대한 염기 편집 값을 나타낸다. 이는 HEK2의 경우 C6, RNF2의 경우 C6, 및 FANCF의 경우 C6이다. 염기 편집 전 및 후 3개의 상이한 부위에 대한 서열은 하기와 같다: HEK2: GAACACAAAGCATAGACTGC (서열식별번호: 110) (서열분석은 다음을 관독함: CTTGTGTTTCGTATCTGACG (서열식별번호: 111)); RNF2: GTCATCTTAGTCATTACCTG (서열식별번호: 112) (서열분석은 다음을 관독함: CAGTAGAATCAGTAATGGAC (서열식별번호: 113)); 및 FANCF: GGAATCCCTTCTGCAGCACC (서열식별번호: 114) (서열분석은 상기와 같은 것을 관독함). HEK2 및 RNF2 둘 다에 대해, 비-표적 가닥을 서열분석하였다 (이 가닥은 표적 C에 상보적인 G를 함유함). FANCF에 대해, 표적 가닥을 서열분석하였다 (이 가닥은 표적 C를 함유함). C에서 T로의 염기 편집 (예를 들어, C에서 T로의 염기 편집제인 BE3을 사용함) 및 C에서 G로의 염기 편집에 대한 개략도가 도 1 및 2에 제시된다. 특정 DNA 폴리머라제는 무염기성 부위 반대편 염기를 G로 대체하는 것으로 공지되어 있다. C에서 G로의 염기 편집을 달성하기 위한 하나의 전략은 무염기성 부위의 생성을 유도한 다음 상기 폴리머라제를 동원하거나 테더링하여 무염기성 부위 반대편 G를 C로 대체하는 것이다. 이는 C 및 T가 절제되고 폴리머라제의 미리 결정된 염기 선호에 기초하여 모든 폴리머라제에 의해 복구될 수 있는 경우, 모든 편집제에 대한 접근을 제공할 수 있다.

[0407] 상이한 융합 구축물이 하기에 요약되고 표 1에 제시된다. UdgX는 최소 우라실-절제 활성으로 우라실에 단단히 결합하는 것으로 공지된 UDG의 이소형이다. UdgX*는 UdgX의 돌연변이된 버전이며 (Sang et al. NAR, 2015),

문헌 [Sang et al.]에서 시험관내 검정에 의해 우라실 절제 활성이 결여된 것으로 관찰되었다. UdgX_On은 UdgX의 또 다른 돌연변이된 버전이며 (Sang et al. NAR, 2015), 문헌 [Sang et al.]에서 보고된 동일한 시험관내 검정에서 증가된 우라실 절제 활성을 갖는 것으로 관찰되었다. UDG는 무염기성 부위를 생성하기 위해 DNA로부터 우라실의 절제를 담당하는 효소이다. Rev7은 무염기성 부위 반대편에 C를 혼입시키는 것으로 공지된 Rev1/Rev3/Rev7 복합체의 성분이다. Rev1은 상기 언급된 복합체의 효소 성분이다. 폴리머라제 알파, 베타, 감마, 델타, 엡실론, 감마, 에타, 이오타, 카파, 람다, 뮤 및 누는 무염기성 부위 반대편 염기 혼입에 대해 상이한 선호를 갖는 진핵 폴리머라제이다.

표 1: 구축물 참조 설명

구축물	정의
BE3	공개된 염기 편집 구축물
BE3_UdgX	우라실 결합 단백질인 UdgX로 대체된 UGI
BE3_UdgX*	우라실에 대한 감소된 결합 친화도를 갖는 UdgX 이소형으로 대체된 UGI
BE3_REV7	C-통합 트랜스레전 합성 기구의 성분으로 대체된 UGI
BE2_UDG	UGI가 우라실 데글리코실라제로 대체된 dCas9 기반 구축물 (니킹 부재)
BE3_UDG	우라실 데글리코실라제 (BE3)로 대체된 UGI
BE2_UdgX_On	UGI가 우라실 절제를 증가시키는 활성화 돌연변이를 갖는 UdgX로 대체된 dCas9 구축물
BE3_UdgX_On	우라실 절제를 증가시키는 활성화 돌연변이를 갖는 UdgX로 대체된 UGI
SMUG1	ssDNA 우라실 데글리코실라제인 SMUG1로 대체된 UGI

실시예에 사용된 구축물:

BE3_전장 - 이는 시티딘 데아미나제, nCas9 및 우라실 글리코실라제 억제제 (UGI) 도메인을 포함하는 C에서 T로의 염기 편집제 구축물이다.

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLLYEINWGGRHSIWRHTSQNT
 NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR
 LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPRYPHLW
 VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSET
 PGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLI
 GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSSFHRLSEESFL
 VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSKADLRLIYLALAHMIK
 FRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRR
 LENLIAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDXYDDDLN
 LAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
 ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
 REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGP
 LARGNSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPK
 HSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE

DYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL
KSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTV
KVVDDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGKELGSQILKE
HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN
KVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR
KDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGDSG
GSTNLSDIIEKETGKQLVIQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENVMML
TSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKRKRV (SEQ ID NO: 115)

[0413]

[0414]

BE3_UGI 부재 - 이 구축물은 UGI 도메인이 결여된 상기 BE3 구축물이다.

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWGGRHSIWRHTSQNT
NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR
LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNEAHWPYRPHLW
VRLYVLELYCIIILGPLCLNLRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSET
PGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLI
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFL
VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLN
LAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGP
LARGNSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLF
EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL

[0415]

KSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTV
KVVDDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGKELGSQILKE
HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN
KVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR
KDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKEGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEEN
EQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIH
LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGD
(SEQ ID NO: 116)

[0416]

[0417]

Cas9 니카제 서열 - BE3에서 사용된다.

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEISNEMAKVDDSFHRLSESLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHSLLEYEFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDDELVK
VMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM

[0418]

PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
AKVEGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEENEQKQLFVEQ
HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGD (SEQ ID NO:
21)

[0419]

[0420] dCas9 서열 - BE2에서 사용된다.

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDFKFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
GEKKNGLFGNLIASLGLTPNFKSNFDAEDAKLQLSKDTYDDDLNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDQKSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVK
VMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD (SEQ ID NO:
22)

[0421]

[0422] BE3_UGI를 UDG, UdgX 변이체, 폴리머라제로 대체함 - 하기 구축물에서, NLS 서열은 밑줄표시에 의해 확인되고, 링커는 이탤릭체로 확인된다. 하기 서열에 나타난 "[UGI]"는 UDG, UDG 변이체 (예를 들어, UDG, UdgX* (R107S), 및 UdgX_On (H109S)), Rev7, 및 Smug1이 삽입된 위치를 확인한다 (BE3의 UGI가 아닌). 하기 서열에 나타난 "[폴리머라제]"는 폴리머라제 (예를 들어, Pol 베타, Pol 람다, Pol 에타, Pol 뮤, Pol 이오타, Pol 카파, Pol 알파, Pol 델타, Pol 감마 및 Pol 누), 및 Rev1이 삽입된 위치를 확인한다.

MSSETGPPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWGGRHSIWRHTSQNT
 NKHVEVNFIKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR
 LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHILW
 VRLYVLELYCIILGLPCLNLRKQQLTFFTIALQSCHYQRLPPHILWATGLKSGSET
 PGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI
 GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFL
 VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIK
 FRGHFLIEGDLNPDNSVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSR
 LENLIAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDYYDDLDNL
 LAQIGDQYADFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
 ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN
 REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPYYVGP
 LARGNSRFAMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPK
 HSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
 DYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLF
 EDREMIEERLKYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKTILDFL
 KSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTV
 KVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEELGKELGSQILKE
 HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDQDHIVPQSFLKDDSIDN
 KVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
 ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFR
 KDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMI
 AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
 TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLQNEKLYLYYLQNGRDMYV
 DQELDINRLSDYDQDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK
 NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS

[0423]

RMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVG
 GTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEIT
 LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
 SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELL
 GITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQK
 GNELALPSKYVNFLYLASHYEKLKGPEDNEQKQLFVEQHKHYLDEIEQISEFSKRV
 ILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
 TKEVLDATLIHQSTGLYETRIDLSQLGGDSGGG(SAQ ID NO:

[0424]

117)[UGI]SGGSGGSGGS(SAQ ID NO: 120)[폴리머라제]PKKKRKV (SAQ ID NO: 41)

[0425]

N-말단 UDQ (삽입물 UDQ (Tyr147Ala) 또는 UDQ (Asn204Asp)) + C-말단에서의 Cas9 니카제 및 폴리머라제 - 하
 기 구축물에서, NLS 서열은 밑줄표시에 의해 확인되고, 링커는 이탤릭체로 확인된다. 하기 서열에 나타낸
 "[UDQ변이체]"는 UDQ Tyr147Ala 및 UDQ Asn204Asp가 삽입된 위치를 확인한다. 하기 서열에 나타낸 "[폴리머라
 제]"는 폴리머라제 (예를 들어, Pol 베타, Pol 람다, Pol 에타, Pol 뮤, Pol 이오타, Pol 카파, Pol 알파, Pol
 델타, Pol 감마 및 Pol 누), 및 Rev1이 삽입된 위치를 확인한다.

[UDG변이체]SETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG
NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVD
DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR
LIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKA
ILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD
TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE
HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI
LTFRIPIYYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASAQSFIERMTNFDK
NLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNR
KVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI
LEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD
KQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSHEHIANLAG
SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLRSDKNRGSNDNPSEEVVKKMKNYWRQLLNAKLITQRKFDN

[0426]

LTKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKSKLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDY
KVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDWDP
KKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELLGITIMERSSEKPNIDFLEAK
GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
EKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHARDK
PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRI
DLSQLGGDSGGS(SEQ ID NO: 118)[폴리머라제]PKKKRKV(SEQ ID NO: 41)

[0427]

[0428]

실시예 1: C에서 G로의 접근법 1 - 무염기성 부위 형성 증가

[0429]

무염기성 부위가 보다 효율적으로 생성되는 경우, C에서 G로의 염기 편집 경로를 통한 총 유동이 증가될 것으로 예상된다. 이러한 접근법에서 사용된 염기 편집제의 개략적 표현이 도 3 및 4에 제시된다. 최소 우라실 절제 활성으로 우라실에 단단히 결합하는 것으로 확인된 UDG의 오르토로그인 UdgX를 사용하는 것은 C에서 G로의 편집 양을 증가시킨다. 어떠한 특정한 이론에 얽매이는 것을 원하지는 않지만, U에 대한 UdgX 근사-공유 결합은 트랜스레전 폴리머라제-유형 복구를 착수시키는 손상을 모방한다. 추가로, UdgX는 낮은 수준의 촉매 활성을 가지며, 이는 단단한 결합과 조합되어 U를 절제하고 무염기성 부위 형성을 유도한다. 무염기성 부위 형성은 오프-타겟 생성물을 가능하게 하고, 이러한 손상의 우선적인 생성은 더 많은 생성물을 유도한다. 이는 상이한 실험 및 염기 편집제를 통해 지지되며, 이는 도 5 및 6에 예시된다.

[0430]

7종의 염기 편집제 (BE3; BE3_UdgX; BE3_UdgX*; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 WT 세포 내 HEK2, RNF2 및 FANCF 부위에서의 C에서 G로의 염기 편집의 결과가 도 7 내지 15에 제시된다. 이들 도면은 표준 C에서 T로의 편집으로부터의 서열 교란에 대해 높은, 중간 및 낮은 용인을 갖는 3개의 대표적인 부위에서의 가장 많이 편집된 위치 (C6)에서 C에서 G로의 편집에 대한 결과를 보여준다.

[0431]

다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 UDG-/- 세포 내 HEK2, RNF2 및 FANCF 부위에서의 C에서 G로의 염기 편집의 결과가 도 16 내지 24에 제시된다.

[0432]

다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 REV1-/- 세포 내 HEK2, RNF2 및 FANCF 부위에서의 C에서 G로의 염기 편집의 결과가 도 25 내지 30에 제시된다.

[0433] 다양한 C에서 G로의 염기 편집제 (BE3; BE3_UdgX; BE2_UNG; BE3_UNG; BE2UdgX_On; BE3UdgX_On; 및 SMUG1)를 사용한 3종의 각 세포 유형 (WT, UDG-/- 및 REV1-/- 세포) 내 HEK2, RNF2 및 FANCF 부위에서의 C에서 G로의 염기 편집의 결과가 도 31 및 32에 요약된다.

[0434] 실시예 2: C에서 G로의 접근법 2 - 무염기성 부위 반대편 C 혼입 증가

[0435] 무염기성 부위 반대편 C 통합에 대한 선호 증가는 총 C에서 G로의 염기 편집의 증가를 유도할 것이다. 이러한 접근법 및 이러한 접근법에서 사용된 염기 편집제에 대한 개략도가 도 33 및 34에 예시된다. C에서 G로의 염기 편집에 대한 이러한 접근법에서 사용될 수 있는 다양한 폴리머라제가 도 35에 제시된다. 간략하게 무염기성 부위 생성은 C에서 비-T로의 생성물 형성을 유도한다. Rev1은 dC 트랜스퍼라제 활성을 갖는다. 이러한 경로를 제거하거나 또는 무염기성 손상이 복구되는 방법을 변경하는 것은 새로운 염기 편집제를 유도하여야 한다. 이러한 경로가 단독으로 이러한 생성물의 형성을 담당한다면 Rev1-/- 녹아웃 세포주는 C에서 G로의 편집이 결여될 것이다. 다양한 폴리머라제의 융합은 증가된 C에서 G로의 염기 편집을 유도하는 무염기성 부위 반대편 복구를 위한 폴리머라제 선호에 기초하여 반대 가닥의 복구를 유도할 것이다. 예시적인 염기 편집제가 도 36에 예시된다.

[0436] 다양한 염기 편집제 (BE3; BE3_UdgX; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 WT 세포 내 HEK2, RNF2 및 FANCF 부위에서의 C에서 G로의 염기 편집의 결과가 도 37 내지 39에 제시된다.

[0437] 인간 폴리머라제 η , ι , κ 및 REV1에 의한 무염기성 부위 및 G 반대편 1-염기 혼입에 대한 정상-상태 동역학적 파라미터가 표 2에 제공된다. 문헌 [Choi et al. J mol Bio. 2010]을 참조한다.

[0438] 표 2: 폴리머라제 η , ι , κ 및 REV1에 대한 정상-상태 동역학적 파라미터

폴리머라제	주형	dNTP	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)	dNTP 선택성 비 ^a	상대 효율 ^b
η	AP 부위	A	40 ± 6	0.12 ± 0.004	3.0	0.95	0.065
		T	290 ± 50	0.92 ± 0.05	3.2	1	0.070
		G	8.5 ± 1.0	0.005 ± 0.0001	0.59	0.19	0.013
		C	210 ± 20	0.14 ± 0.01	0.67	0.21	0.015
	G	C	2.6 ± 0.1	0.12 ± 0.005	46	1	1
ι	AP 부위	A	210 ± 40	0.54 ± 0.04	2.6	0.45	1.4
		T	130 ± 20	0.74 ± 0.02	5.7	1	3.0
		G	120 ± 10	0.47 ± 0.01	3.9	0.69	2.1
		C	570 ± 140	0.77 ± 0.05	1.4	0.24	0.74
	G	C	300 ± 30	0.57 ± 0.02	1.9	1	1
κ	AP 부위	A	1600 ± 200	0.077 ± 0.005	0.048	0.77	0.00065
		T	2300 ± 700	0.017 ± 0.002	0.0074	0.12	0.00010
		G	400 ± 70	0.0032 ± 0.0002	0.008	0.13	0.00011
		C	780 ± 220	0.049 ± 0.005	0.063	1	0.00085
	G	C	3.8 ± 0.5	0.28 ± 0.01	74	1	1
REV1	AP 부위	A	140 ± 50	0.000025 ± 0.000002	0.00018	0.0031	0.00019
		T	190 ± 30	0.000072 ± 0.000003	0.00038	0.0067	0.00040
		G	190 ± 50	0.000031 ± 0.000003	0.00016	0.0029	0.00017
		C	210 ± 30	0.012 ± 0.001	0.057	1	0.061
	G	C	12.8 ± 50	0.012 ± 0.0003	0.94	1	1

[0439]

[0440] ^adNTP 선택성 비는 각각의 dNTP 혼입에 대한 k_{cat}/K_m 을 AP 부위 반대편 dNTP 혼입에 대한 가장 높은 k_{cat}/K_m 으로 나눔으로써 계산됨.

[0441] ^b상대 효율은 AP 부위 반대편 각각의 dNTP 혼입에 대한 k_{cat}/K_m 을 G 반대편 dCTP 혼입에 대한 k_{cat}/K_m 으로 나눔으로써 계산됨.

[0442] 인간 폴리머라제 α 및 δ /PCNA에 의한 무염기성 부위 및 G 반대편 1-염기 혼입에 대한 정상-상태 동역학적 파라미터가 표 3에 제공된다.

[0443] 표 3: 폴리머라제 α 및 δ /PCNA에 대한 정상-상태 동역학적 파라미터

폴리머라제	주형	dNTP	K_m (μ M)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)	dNTP 선택성 비 ^a	상대 효율 ^b
α	AP 부위	A	570 ± 100	0.0083 ± 0.0001	0.015	1	0.0010
		T	250 ± 60	0.00046 ± 0.00003	0.0018	0.12	0.00012
		G	550 ± 120	0.00024 ± 0.00002	0.0004	0.027	0.00003
		C	980 ± 50	0.00047 ± 0.00001	0.0005	0.033	0.00003
	G	C	0.42 ± 0.09	0.0064 ± 0.0003	15		1
δ /PCNA	AP 부위	A	25 ± 6	0.0067 ± 0.0004	0.27	1	0.012
		T	62 ± 16	0.0060 ± 0.0004	0.097	0.36	0.0044
		G	110 ± 20	0.010 ± 0.001	0.091	0.34	0.0041
		C	880 ± 160	0.0069 ± 0.0006	0.0078	0.029	0.0004
	G	C	0.27 ± 0.05	0.0059 ± 0.0002	22		1

[0444]

[0445] ^adNTP 선택성 비는 각각의 dNTP 혼입에 대한 k_{cat}/K_m 을 AP 부위 반대편 dNTP 혼입에 대한 가장 높은 k_{cat}/K_m 으로 나눔으로써 계산됨.

[0446] ^b상대 효율은 AP 부위 반대편 각각의 dNTP 혼입에 대한 k_{cat}/K_m 을 G 반대편 dCTP 혼입에 대한 k_{cat}/K_m 으로 나눔으로써 계산됨.

[0447] 표 4 - 염기 편집 접근법 2에 사용될 수 있는 폴리머라제.

폴리머라제	크기 (아미노산)
패밀리 X	
베타	335
람다	575
뮤	494
패밀리 B	
알파	1462
델타	1107
엡실론	2286
패밀리 Y	
에타	713
이오타	740
카파	870
Rev1	1251
제타 (Rev3/Rev7)	3130

[0448]

[0449] 실시예 3: C에서 G로의 접근법 3 - 무염기성 부위 형성 및 C 혼입 둘 다 증가

[0450] 증가된 C에서 G로의 염기 편집을 위한 무염기성 부위 형성 및 C 혼입 둘 다를 증가시키기 위한 염기 편집제의 개략도가 도 40에 예시된다. 폴리머라제 테더링된 구축물, 특히 Pol 카파의 첨가는 C에서 G로의 염기 편집을 증가시킨다. Pol 카파 또는 Pol 이오타 테더링된 구축물을 사용한 HEK2, RNF2 및 FANCF 부위에서의 염기 편집의 결과가 도 41에 제시된다. WT 세포에서 추가의 폴리머라제 테더링된 구축물을 사용한 HEK2, RNF2 및 FANCF 부위 내 시토신 잔기에서의 염기 편집의 결과가 도 42 내지 47에 제시된다. UDG 147은 T를 직접 제거하고 C에서 G로의 염기 편집을 증가시키는 효소이고 (도 42 내지 44), 반면 UDG 204는 C를 직접 제거하고 C에서 G로의 염기 편집을 증가시키는 효소이다 (도 45 내지 47).

[0451] 실시예 4: C에서 G로의 접근법 4 - C에서 G로의 유동을 증가시키기 위한 대안적 복구 경로 제거

[0452] C에서 G로의 편집을 개선시키는 하나의 방법은 대안적 복구 경로를 제거하거나 하향조정하는 것이다. 한 예로서, 복구 경로 단백질 MSH2^{-/-}를 제거하는 것은 C에서 G로의 염기 편집의 증가를 유도할 수 있으며, 이는 도 48에 제시된다. 다양한 염기 편집제 (BE3; BE3_UdgX; BE2_UdgX-On; BE3_UdgX-On; BE2_UDG; 및 BE3_UDG)를 사용

한 MSH2^{-/-} 세포 내 HEK2, RNF2 및 FANCF 부위에서의 C에서 G로의 염기 편집의 결과가 도 49 내지 51에 제시된다.

[0453] 실시예 5: C에서 G로의 접근법 5 - 성분의 트랜스 발현

[0454] 함께 기능하는 염기 편집제 성분을 확인하기 위한 하나의 접근법은 그러한 성분을 함께 세포에서 트랜스로 발현 시키는 것이다. C에서 G로의 돌연변이를 유도하는 염기 편집제 성분 (예를 들어, 폴리머라제, 우라실 결합 단백질, 염기 절제 효소, 시티딘 데아미나제 및/또는 핵산 프로그램가능 DNA 결합 단백질)이 확인되면, 이들은 염기 편집제를 생성하도록 테더링될 수 있다. APOBEC-Cas9 니카제에 융합된 발현된 UDG 및 UdgX 변이체 및 동시에 트랜스로 과다발현된 TLS 폴리머라제는 RNF2 부위에서 C에서 G로의 편집을 유도한다. 성분의 트랜스 발현을 예시하는 개략도가 도 52에 제시된다.

[0455] 다양한 폴리머라제 (Pol 카파, Pol 에타, Pol 이오타, REV1, Pol 베타 및 Pol 텔타)에 의해 트랜스로 발현된 5 종의 상이한 염기 편집제 (BE3; BE3_UdgX; BE2_UdgX_On; BE3_UdgX_On; BE2_UDG; 및 BE3_UDG)를 사용한 HEK293 세포 내 HEK2, RNF2 및 FANCF에서의 염기 편집의 결과가 도 53 내지 55에 제시된다.

[0456] 참고문헌

1. Chan, K., Resnick, M. A., Gordenin, D. A. The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis. *DNA Repair* 12, 878-889 (2013).

[0457]

2. Choi, J. Y., Lim, S., Kim, E. J., Jo, A., and Guengerich F. P. Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases alpha, delta, eta, iota, kappa, and Rev1. *Journal of Molecular Biology* 404, 34-44 (2010).
3. Dianov, G. L. and Hubsher U. Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Research*, 1-8 (2013).
4. Fortini, P., Pasucci, B., Sobol, R. W., Wilson, S. H., and Dogliotti, E. Different DNA polymerases are involved in the Short- and lon-patch base excision repair in mammalian cells. *Biochemistry* 37, 3575-3580 (1998).
5. Jiricny, J. The multifaceted mismatch-repair system. *Nature Rev. Molecular Cell Biology* 7, 335-346 (2006).
6. Katafuchi A. and Nohmi T. DNA polymerases involved in the incorporation of oxidized nucleotides into DNA: their efficiency and template base preference. *Mutation Research* 703, 24-31 (2010).
7. Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A., and Krokan, E. H. Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO* 15, 3442-3447 (1996).
8. Krokan, H. E. and Bjoras, M. Base Excision Repair, *Cold Spring Harbor Perspectives in Biology*, 1-22 (2013).
9. Kunkel, T. A. and Erie, D. A. Eukaryotic mismatch repair in relation to RNA replication. *Annual Reviews Genetics* 49, 291-313 (2015).
10. Li, G. M. Mechanisms and functions of DNA mismatch repair. *Cell Research* 18, 85-98 (2008).
11. Lin, W., Xin, H., Wu, X., Yuan, F., and Wang, Z. The human REV1 gene codes for a DNA template-dependent dCMP transferase. *Nucleic Acids Research* 27, 4468-4475 (1999).
12. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavil, B., Alseth, I., Krokan, H. E., and Tainer, J. A. Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* 80, 869-878 (1995).
13. Prasad, R., Poltoratsky, V., Hou, E. W., and Wilson, S. H. Rev1 is a base excision repair enzyme with 5' deoxyribose phosphate lyase activity. *Nucleic Acid Research*, 1-10 (2016).
14. Robertson, A. B., Klungland, A., Rognes, T., and Leiros, I. Base excision repair: the long and the short of it. *Cell Molecular Life Sciences* 66, 981-993 (2009).
15. Sale, J. E., Lehmann, A. R., and Woodgate, R. Y-Family DNA polymerases and their role in tolerance of cellular DNA damage. *Nature Rev. Molecular Cell Biology* 13, 141-152 (2012).
16. Sang, P. B., Srinath, T., Patil, A. G., Woo, E. J., and Varshney, U. A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Research*, 1-12 (2015).
17. Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* 373, 487-493 (1995).
18. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* 384, 87-92 (1996).
19. Weill, J. C. and Reynaud C. A. DNA polymerases in adaptive immunity. *Nature Rev. Immunology* 8, 302-312 (2008).
20. Yasui, A. Alternative excision repair pathways. *Cold Spring Harbor Perspectives in Biology*, 1-8 (2013).

[0458]

[0459]

실시예 6: - Cas9 변이체 서열

[0460]

본 개시내용은 Cas9 변이체, 예를 들어 1개 이상의 돌연변이 (예를 들어, dCas9 또는 Cas9 니카제를 생성하기 위한 것)를 포함할 수 있는 1종 이상의 유기체로부터의 Cas9 단백질을 제공한다. 일부 실시양태에서, 하기 표에 의해 확인되는, Cas9 단백질의 아미노산 잔기 중 1개 이상이 돌연변이될 수 있다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 D10 및/또는 H840 잔기, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이가 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 D10 잔기, 또는 본원에 제공된 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이는 D를 제외한 임의의 아미노산 잔기로 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 D10 잔기, 또는 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이는 A로 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 H840 잔기, 또는 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 잔기는 H이다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 H840 잔기, 또는 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된

아미노산 서열 중 임의의 것에서의 상응하는 돌연변이는 H를 제외한 임의의 아미노산 잔기로 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 H840 잔기, 또는 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 돌연변이는 A로 돌연변이된다. 일부 실시양태에서, 서열식별번호: 6에 제공된 아미노산 서열의 D10 잔기, 또는 임의의 Cas9, 예컨대 서열식별번호: 4-26에 제공된 아미노산 서열 중 임의의 것에서의 상응하는 잔기는 D이다.

[0461] 다양한 종으로부터의 Cas9 서열을 정렬시켜, 서열식별번호: 6의 D10 및 H840의 상응하는 상동 아미노산 잔기가 다른 Cas9 단백질에서 확인되어 상동 아미노산 잔기의 상응하는 돌연변이를 갖는 Cas9 변이체의 생성을 허용할 수 있는지 여부를 결정하였다. 정렬은 NCBI 제약-기반 다중 정렬 도구 (COBALT (st-va.ncbi.nlm.nih.gov/tools/cobalt에서 접근가능함))를 사용하여 하기 파라미터로 수행하였다. 정렬 파라미터: 갭 페널티 -11,-1; 엔드-갭 페널티 -5,-1. CDD 파라미터: RPS BLAST 사용; 블라스트 E-값 0.003; 보존된 칼럼 발견 및 리컴퓨팅. 쿼리 클러스터링 파라미터: 쿼리 클러스터 사용; 워드 크기 4; 최대 클러스터 거리 0.8; 알파벳 규칙적.

[0462] 4개의 Cas9 서열의 예시적인 정렬이 하기에 제공된다. 정렬에서의 Cas9 서열은 다음과 같다: 서열 1 (S1): 서열식별번호: 23 | WP_010922251 | gi 499224711 | 유형 II CRISPR RNA-가이드 엔도뉴클레아제 Cas9 [스트렙토코쿠스 피오케네스]; 서열 2 (S2): 서열식별번호: 24 | WP_039695303 | gi 746743737 | 유형 II CRISPR RNA-가이드 엔도뉴클레아제 Cas9 [스트렙토코쿠스 갈롤리티쿠스]; 서열 3 (S3): 서열식별번호: 25 | WP_045635197 | gi 782887988 | 유형 II CRISPR RNA-가이드 엔도뉴클레아제 Cas9 [스트렙토코쿠스 미티스]; 서열 4 (S4): 서열식별번호: 26 | 5AXW_A | gi 924443546 | 스타필로코쿠스 아우레우스 Cas9. HNH 도메인 (볼드체 및 밑줄표시) 및 RuvC 도메인 (박스표시)이 각각의 4개의 서열에 대해 확인된다. S1에서의 아미노산 잔기 10 및 840 및 정렬된 서열에서의 상동 아미노산은 하기 각각의 아미노산 잔기에서 별표로 확인된다.

```

S1 1  --MDKK--YSIGLD*IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI--GALLFDSG--ETAETRLKRTARRRYT 73
S2 1  --MTKKNYSIGLD*IGTNSVGWAVITDDYKVPAKMKVLGNTDKKYIKKNLI--GALLFDSG--ETAETRLKRTARRRYT 74
S3 1  --M-KKGYSIGLD*IGTNSVGFAVITDDYKVPSKMKVLGNTDKRFIKKNLI--GALLFDEG--TTAEARLKRTARRRYT 73
S4 1  GSHMKRNYILGLD*IGITSVGYGII--DYET-----RDVIDAGVRLFKEANVENNEGRRSKRGARRLKR 61

S1 74  RRKNRICYLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSSTDKADLRL 153
S2 75  RRKNRLRYLQEIFANEIAKVDESFFQRLDESFLTDDDKTFDSHPIFGNKAEDDAYHQFPTIYHLRKLADSSEKADLRL 154
S3 74  RRKNRLRYLQEIFSEEMSKVDSFFHRLDDSLIPEDKRESKYPIFATLTEEKEYHKQFPTIYHLRKLADSSEKTDLRL 153
S4 62  RRRHRIQRVKKLL-----FDYNLLTD-----HSELGINPYEARVKGLSQKLSEEE 107

S1 154  IYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEK 233
S2 155  VYLALAHMIKFRGHFLIEGELNAENTDVQKIFADFGVYVNRFTDDSHLSEITVDVASILTEKISKSRRLENLIKYYPTK 234
S3 154  IYLALAHMIKYRGHFLYEAFFDIKNNDIQKIFNEFISIDNTFEGSSLSGQNAQVEAIFTDKISKAKRERVLKLPFDEK 233
S4 108  FSAALLHLAKRRG-----VHNVEVEEDT----- 131

S1 234  KNLFGNLIALSLGLTPNFKNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAILLSDILRVNTEIT 313
S2 235  KNTLFGNLIALALGLQPNFKNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLYDAILLSGILTVDDNST 314
S3 234  STGLFSEFLKLIIVGNQADFKKHFDLEDKAPLQFSKDTYEDLENLLGQIGDDFTDLFVSAKKLYDAILLSGILTVTDPST 313
S4 132  ----GNELS-----TKEQISRN----- 144

S1 314  KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM--DGTEELLV 391
S2 315  KAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIKDKNKNGYAGYIENGVKQDEFYKYLKNILSKIKIDGSDYFLD 394
S3 314  KAPLSASMIERYENHQNDLAALKQFIKNNLPEKYDEVFSDQSKDGYAGYIDGKTTQETFYKYIKNLLSKF--EGTDYFLD 391
S4 145  ----SKALEEKYVAELQ-----LRLKKDG----- 165

S1 392  KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAMWTRKSEE 471
S2 395  KIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKEKQDRIEKILTFRIPYYVGPLVRKDSRFWAWEYRSDE 474
S3 392  KIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEYYPFLKDNKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDE 471
S4 166  --EVRSINRFKTS-----YVKEAKQLLVQKAYHQLDQSFIDTYIDLLETRTRYEGP--GEGSPFGW-----K 227

S1 472  TITPWNFEFEEVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL 551
S2 475  KITPWNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPHKSHVYETYAVYNELTKIKYVNEQGKE--SFFDSNMKQEIFDH 553
S3 472  AIRPWNFEFEEIVDKASSAEDFINKMNTYDLYLPEEKVLPHKSLLEYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQ 551

```

[0463]

S4 228 DIKEW-----YEMLMGHCYFPEELRSVKYAYNADLYNALNDLNLVITRDENEK---LEYEKFQIIEN 289

S1 552 LFTNRRKVTVKQLKEDYFKKIECFDSVEISGVEDR---FNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTTLTFED 628

S2 554 VFENRRKVTKEKLLNLYNKEFPEYRIKDLIGLDKENKSFNASLGTYHDLKKIL-DKAFLLDDKVNEEVIEDIKTLTLTFED 632

S3 552 LFKENRRKVTKEKDIHYLHN-VDGYDGIELKGIEKQ---FNASLSTYHDLKIIKDKFMDDAKNAEILNIVHTLTIFED 627

S4 290 VFQKQKKPTLQIAKEILVNEEDIKGYRVTSVGKPEF---TNLKVYHDKIDITARKEII---ENAEELLDQIAKILTIYQS 363

S1 629 REMIEERLKYAHLFDDKVMKQLKR-RRYTGWGRLSRKLINGIRDKQSGKTIIDFLKSDGFANRNFMQLIHDDSLTFKED 707

S2 633 KDMIHERLKQYSIDFTANQLKKLER-RHYTGWGRLSYKLINGIRNKENKNTILDYLIDDGSANRNFMQLINDTLFPKQI 711

S3 628 REMIKQRLAQYDSLFDEKVIKALTR-RHYTGWGRLSAKLINGICDKQTGNTILDYLIDDGKINRNFMQLINDGLSFKEI 706

S4 364 SEDIQEELTNLNLSELTQEIEQISNLKGYTGTHNLSLKAINLILDE-----LWHTNDNQIAIFNRLKLV----- 428

S1 708 IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT-----QKGGKNSRERM 781

S2 712 IQKSQVVGVDDEIAVVHDLPGSPAIKKGILQSVKIVDELVKVMG-GNPDNIVIEMARENQTT-----NRGRSQSQORI 784

S3 707 IQKAQVIGTKDDVKQVQELSGSPAIKKGILQSIKIVDELVKVMG-HAPESIVIEMARENQTT-----ARGKKNQQRY 779

S4 429 -KKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG-LPNDIIIELAIREKNSKDAQKMINEMQKRNRRQTN 505

S1 782 KRIEEGIKELSGOIL-----KEHPVENTQLONEKLYLYLQNGRDMYVDQELDINRLSD----YVDH*IVPQSFLKDD 850

S2 785 KKLQNSLKEKLSNILEEKPSYIEDKVENSHLQNDQLFLYYIQNGKDMYTGDELDIDHLS-----YDIDH*IIPQAFIKDD 860

S3 780 KRIEDSLKILASGL---DSNILEKENTDNNQLQNDRLFLYYIQNGKDMYTGDELDINQLSS----YDIDH*IIPQAFIKDD 852

S4 506 ERIEELIRTGK-----ENAKYLEIKLHDMQEGKCLYSLEAIPLEDLNNPFNYEVDH*IIPRSVSDFN 570

S1 851 SIDNKVLTSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDN-LTKAERGGL-SELD-----KAGFIKRQLV 922

S2 861 SIDNRVLTSSAKNRGKSDDVPSLDIVRAKAEWVRLYKSGLSIKRKFND-LTKAERGGL-TEAD-----KAGFIKRQLV 932

S3 853 SIDNRVLTSSKDNRGKSDNVPSEEVQKKRAFQQLLDSKLISERKFNN-LTKAERGGL-DEPD-----KVGFIKRQLV 924

S4 571 SFNNKVLVQEEASKGKGNRTTPQYLSSSDSKISYETFKKHILNAKKGGRISKTKEFYLLLEERDINRFSVQKDFINRNLV 650

S1 923 ETRQITKHVAQILDSRMTKYDENDKLIREVKVITLKSILVSDFRKDFQFYKVRINNYHHHADAYLNAVVGTAIIKKYP 1002

S2 933 ETRQITKHVAQILDARFNTHEHDENDKVIDVKVITLKSILVSDFRKDFQFYKVRINNYHHHADAYLNAVVGTAIIKKYP 1012

S3 925 ETRQITKHVAQILDARNTVEVNEKDKKNTVKIITLKSILVSNFRKEFRLYKVRINNYHHHADAYLNAVVAKAILKKYP 1004

S4 651 DTRYATRGMLNLLRSYFRVN-----NLDVKVSINGGFTSFLRRKWKFKKERNKGYKHAEDALIIA----- 712

S1 1003 KLESEFVYGDYKVYDVRKMIASEQ--EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG--- 1077

S2 1013 KLASEFVYGEYKDYDIRKFIITNSSD----KATAKYFFYSNLMNFFKTKVKYADGTVFERPIIETNAD-GEIAWNKQ--- 1083

S3 1005 KLEPEFVYGEYQYDLKRYISRSKDPKEVEKATEKYFFYSNLMNFFKEEVHYADGTIVKREINIEYKSDTGEIAWNKE--- 1081

S4 713 --NADFIKKEWKLKDAKVMENQM-----FEEKQAESMPETEYQEYKEIFITPHQIK 764

S1 1078 -----RDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKD---WDPKKYGGFDSPTVAYSVLVAVK 1149

S2 1084 -----IDFEKVRKVLSPQVNIKKTEVQTGGFSKESILPKGSDKLIIPKTKKYYWDTKKYGGFDSPTVAYSVVFVADV 1158

S3 1082 -----KDFAIKKKVLSPQVNIKKTEVQTGGFSKESILPKGNSDKLIIPKTKILDDTTYGGFDSPIAYSILLIADI 1156

S4 765 HIKDFKDYKSHRVDDKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL---KKLIN-KSP---EKLMLYHH 835

S1 1150 EKGSKKLKSVKELLGITIMERSSFEKNPI-DFLEAKG----YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG 1223

S2 1159 EKGAKKLKTVKELVGISIMERSFFEENPV-EFLENKG----YHNIREDKLIIKLPKYSLFEFEGRRRLASASELQKG 1232

S3 1157 EKGAKKLKTVKTLVGITIMEKAAFEENPI-TFLENKG----YHNVKENILCLPKYSLFELENGRRLLASAKELQKG 1230

S4 836 DPQTYQKLK-----LIMEQYGDENPLYKYEEETGNLYTKYSKKNNGPVIKKIKYYGNKLNALDITDDYPNSRNKV 907

S1 1224 NELALPSKYVNFYLYASHYEKLKSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKH----- 1297

S2 1233 NEMVLPGYLVLELLYHAHRADNF-----NSTEYLYNVSEHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSM----- 1301

S3 1231 NEIVLPVYLTTLTYHKNVHL-----DEPGHLEYIQKHRNEFKDLLNLVSEFSQKYVLADANLEKIKSLYADN----- 1299

S4 908 VKLSLKPYRFD-VYLDNGVYKFV-----TVKNLDVIK--KENYYEVNSKAYEEAKKLKISNQAEFIASFYNNDLIKING 979

S1 1298 RDKPIREQAENIIHLFTLNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI-----GLYETRI----DLSQL 1365

S2 1302 DNFSIEEISNSFINLLTALGAPADFNFLGKIPRKRYTSTKECLNATLIHQSI-----GLYETRI----DLSKL 1369

S3 1300 EQADIIEILANSFINLLTALGAPAAFKFGKIDRKRYTSTVEILNATLIHQSI-----GLYETWI----DLSKL 1367

S4 980 ELYRVIGVNNDLNRIEVMIMIDITYR-EYLENMNDKRPPRIIKTIASKT---QSIKKYSTDILGNLYEVKSKKHPQIIKK 1055

S1 1366 GGD 1368 (SEQ ID NO: 23)

S2 1370 GEE 1372 (SEQ ID NO: 24)

S3 1368 GED 1370 (SEQ ID NO: 25)

S4 1056 G-- 1056 (SEQ ID NO: 26)

[0464]

[0465]

[0466]

정렬은 참조 Cas9 아미노산 서열 또는 아미노산 잔기에 상동인 아미노산 서열 및 아미노산 잔기가, 관련 기술분야에 공지된 정렬 프로그램 및 알고리즘을 사용하여 참조 서열 또는 참조 잔기와 정렬되는 아미노산 서열 또는 잔기를 확인함으로써, 상이한 종으로부터의 Cas9 서열을 포함하나 이에 제한되지 않는 Cas9 서열 변이체에 걸쳐

확인될 수 있다는 것을 입증한다. 본 개시내용은 서열식별번호: 23-26에서 별표에 의해 확인되는 아미노산 잔기 중 1개 이상 (예를 들어, 각각 S1, S2, S3 및 S4)이 본원에 기재된 바와 같이 돌연변이된 Cas9 변이체를 제공한다. 서열식별번호: 23-26에서 별표에 의해 확인되는 잔기에 상응하는 서열식별번호: 6의 Cas9에서의 잔기 D10 및 H840은 본원에서 "상동" 또는 "상응하는" 잔기로 지칭된다. 이러한 상동 잔기는, 예를 들어 상기 기재된 바와 같은 서열 정렬에 의해, 및 참조 서열 또는 잔기와 정렬되는 서열 또는 잔기를 확인함으로써 확인될 수 있다. 유사하게, 본원에서 서열식별번호: 6에서 확인되는 돌연변이, 예를 들어 서열식별번호: 6에서의 잔기 10 및 840의 돌연변이에 상응하는 Cas9 서열에서의 돌연변이는 본원에서 "상동" 또는 "상응하는" 돌연변이로 지칭된다. 예를 들어, 상기 4개의 정렬된 서열에 대해 서열식별번호: 6 또는 S1 (서열식별번호: 23)에서의 D10A 돌연변이에 상응하는 돌연변이는 S2에 대해 D11A, S3에 대해 D10A, 및 S4에 대해 D13A이고; 서열식별번호: 6 또는 S1 (서열식별번호: 23)에서의 H840A에 대한 상응하는 돌연변이는 S2에 대해 H850A, S3에 대해 H842A, 및 S4에 대해 H560A이다.

[0467] 추가로, 상이한 종으로부터의 여러 Cas9 서열을 상기 약술된 동일한 알고리즘 및 정렬 파라미터를 사용하여 정렬시켰다. 상이한 종으로부터의 여러 Cas9 서열 ('632 공보의 서열식별번호: 11-260)을 상기 약술된 동일한 알고리즘 및 정렬 파라미터를 사용하여 정렬시켰으며, 이는 예를 들어 2017년 4월 27일에 공개된 발명의 명칭 "핵염기 편집제 및 그의 용도"의 특허 공개 번호 W02017/070632 ("'632 공보") (이는 본원에 참조로 포함됨)에 제시되어 있다. 다른 Cas9 단백질의 잔기에 상동인 아미노산 잔기는 상응하는 돌연변이를 다른 Cas9 단백질 내로 혼입시키는데 사용될 수 있는 이러한 방법을 사용하여 확인할 수 있다. 서열식별번호: 6의 잔기 10 및 840에 상동인 아미노산 잔기를 상기 약술된 것과 동일한 방식으로 확인하였다. 정렬은 본원에 제공되며, 참조로 포함된다. HNH 도메인 (볼드체 및 밑줄표시) 및 RuvC 도메인 (박스표시)이 각각의 4개의 서열 (서열식별번호: 23-26)에 대해 확인된다. 서열식별번호: 6에서의 아미노산 잔기 10 및 840에 상응하는 단일 잔기는 정렬에서의 서열식별번호: 23에서 박스표시되어 있으며, 이는 정렬된 서열에서의 상응하는 아미노산 잔기를 확인시켜 준다.

[0468] 등가물 및 범주, 참고문헌에 의한 포함

[0469] 관련 기술분야의 통상의 기술자는 상용 실험에 지나지 않는 것을 사용하여, 본원에 기재된 본 발명의 구체적인 실시양태에 대한 많은 등가물을 인식하거나 확인할 수 있을 것이다. 본 발명의 범주는 상기 설명에 제한되도록 의도된 것은 아니며, 오히려 첨부된 청구범위에 제시된 바와 같다.

[0470] 청구범위에서 단수표현은 달리 나타내지 않는 한 또는 달리 문맥으로부터 명백하지 않는 한 하나 또는 하나 초과를 의미할 수 있다. 군의 하나 이상의 구성원들 사이에 "또는"을 포함하는 청구범위 또는 설명은 달리 나타내지 않는 한 또는 달리 문맥으로부터 명백하지 않는 한 군 구성원 중 하나, 하나 초과 또는 모두가 주어진 생성물 또는 공정에 존재하거나, 그에 사용되거나, 또는 그와 달리 관련되어 있는 경우에 충족된 것으로 간주된다. 본 발명은 군의 정확하게 하나의 구성원이 주어진 생성물 또는 공정에 존재하거나, 그에 사용되거나, 또는 그와 달리 관련되어 있는 실시양태를 포함한다. 본 발명은 또한 군 구성원의 하나 초과 또는 모두가 주어진 생성물 또는 공정에 존재하거나, 그에 사용되거나, 또는 그와 달리 관련되어 있는 실시양태를 포함한다.

[0471] 추가로, 본 발명은 하나 이상의 청구항 또는 명세서의 관련 부분으로부터의 하나 이상의 제한, 요소, 절, 서술적 용어 등이 또 다른 청구항 내에 도입되는 모든 변동, 조합 및 순열을 포괄하는 것으로 이해될 것이다. 예를 들어, 또 다른 청구항에 대해 종속항인 임의의 청구항은 동일 기반 청구항에 대해 종속항인 임의의 다른 청구항에서 발견되는 하나 이상의 제한을 포함하도록 변형될 수 있다. 추가로, 청구항이 조성물을 인용하는 경우에, 달리 나타내지 않거나 또는 모순 또는 불일치가 발생할 것임이 관련 기술분야의 통상의 기술자에게 분명하지 않는 한, 본원에 개시된 임의의 목적을 위해 상기 조성물을 사용하는 방법이 포함되고, 본원에 개시된 임의의 제조 방법 또는 관련 기술분야에 공지된 다른 방법에 따라 상기 조성물을 제조하는 방법이 포함되는 것으로 이해될 것이다.

[0472] 요소들이, 예를 들어 마쿠쉬 군 포맷의 목록으로서 제시되는 경우에, 요소의 각각의 하위군이 또한 개시되고, 임의의 요소(들)가 이러한 군으로부터 제거될 수 있는 것으로 이해될 것이다. 또한 용어 "포괄하는"은 개방된 것을 의도하고 추가의 요소 또는 단계의 포함을 허용한다는 것에 주의한다. 일반적으로, 본 발명 또는 본 발명의 측면이 특정한 요소, 특색, 단계 등을 포함하는 것으로 언급된 경우, 본 발명 또는 본 발명의 측면의 특정 실시양태가 그러한 요소, 특색, 단계 등으로 이루어지거나 본질적으로 이루어지는 것으로 이해되어야 한다. 단 순성의 목적을 위해, 그 실시양태는 본원에 구체적으로 제시되지 않았다. 따라서, 하나 이상의 요소, 특색, 단계 등을 포함하는 본 발명의 각각의 실시양태에 대해, 본 발명은 또한 그 요소, 특색, 단계 등으로 이루어지거

나 본질적으로 이루어지는 실시양태를 제공한다.

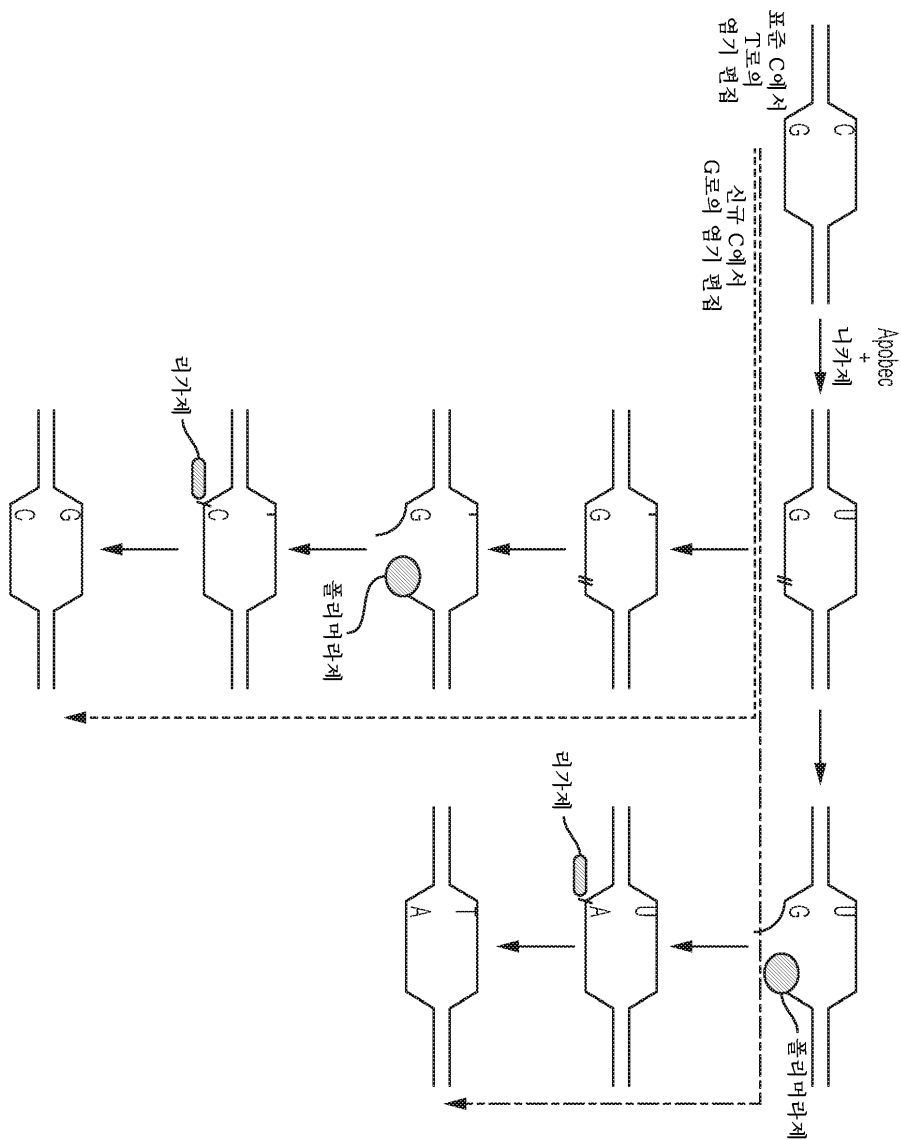
[0473] 범위가 주어진 경우, 종점들이 포함된다. 추가로, 달리 나타내지 않는 한 또는 달리 문맥 및/또는 관련 기술분야의 통상의 기술자의 이해로부터 명백하지 않는 한, 범위로서 표현된 값은 문맥이 달리 명백하게 지시하지 않는 한 본 발명의 다른 실시양태에서 언급된 범위 내에 있는 임의의 특정 값을 상기 범위의 하한치의 단위의 10분의 1로 취할 수 있는 것으로 이해될 것이다. 달리 나타내지 않는 한 또는 달리 문맥 및/또는 관련 기술분야의 통상의 기술자의 이해로부터 명백하지 않는 한, 범위로서 표현된 값은 주어진 범위 내에 있는 임의의 하위범위를 취할 수 있고, 여기서 하위범위의 종점은 상기 범위의 하한치의 단위의 10분의 1과 동일한 정도의 정확도로 표현되는 것으로 또한 이해될 것이다.

[0474] 추가로, 본 발명의 임의의 특정한 실시양태는 어느 하나 이상의 청구항으로부터 명백하게 배제될 수 있는 것으로 이해될 것이다. 범위가 주어지는 경우에, 범위 내에 있는 임의의 값은 어느 하나 이상의 청구항으로부터 명백하게 배제될 수 있다. 본 발명의 조성물 및/또는 방법의 임의의 실시양태, 요소, 특색, 적용 또는 측면은 어느 하나 이상의 청구항으로부터 배제될 수 있다. 간결성의 목적을 위해, 하나 이상의 요소, 특색, 목적 또는 측면이 배제된 모든 실시양태는 본원에 명백하게 제시되지 않는다.

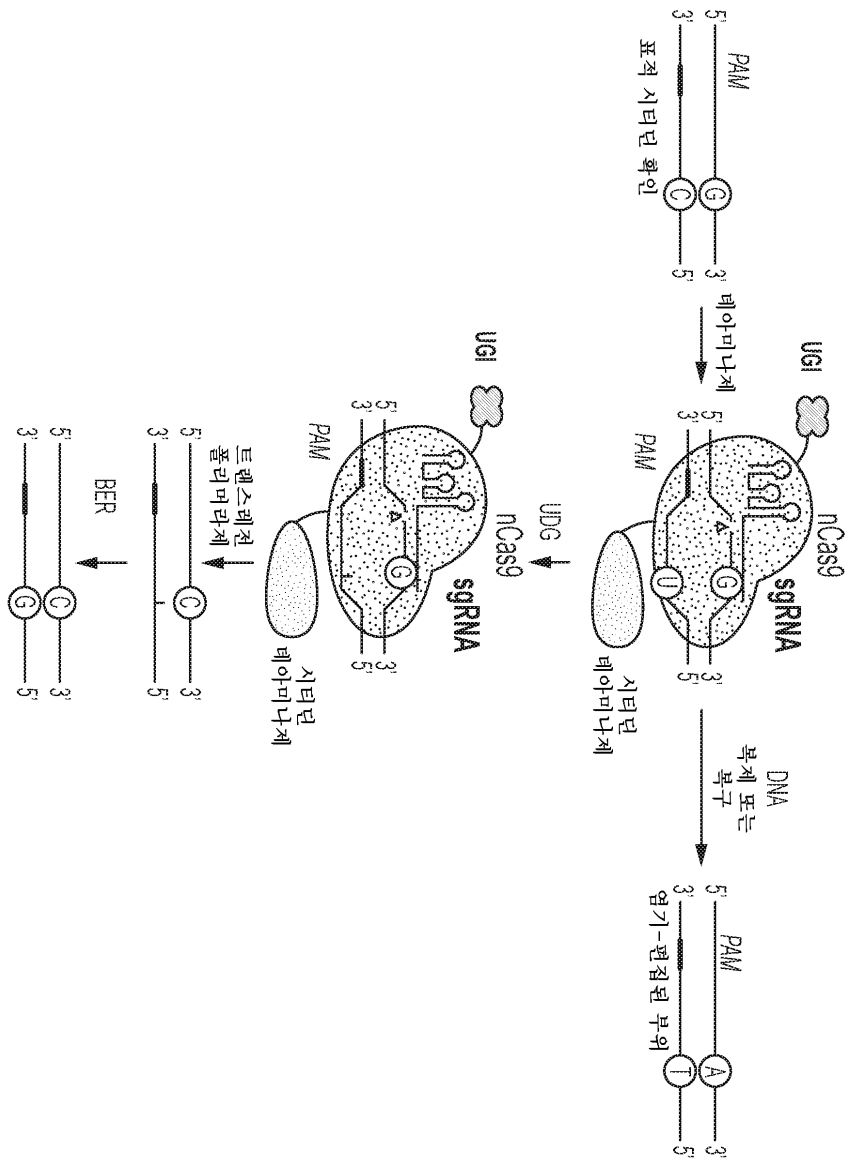
[0475] 상기 열거된 항목을 포함한, 본원에 언급된 모든 공개물, 특허 및 서열 데이터베이스 목록은 마치 각각의 개별 공개물 또는 특허가 특이적으로 및 개별적으로 참조로 포함되는 것으로 나타내어진 바와 같이 그 전문이 본원에 참조로 포함된다. 상충되는 경우에, 본원의 임의의 정의를 포함하여 본 출원이 우선할 것이다.

도면

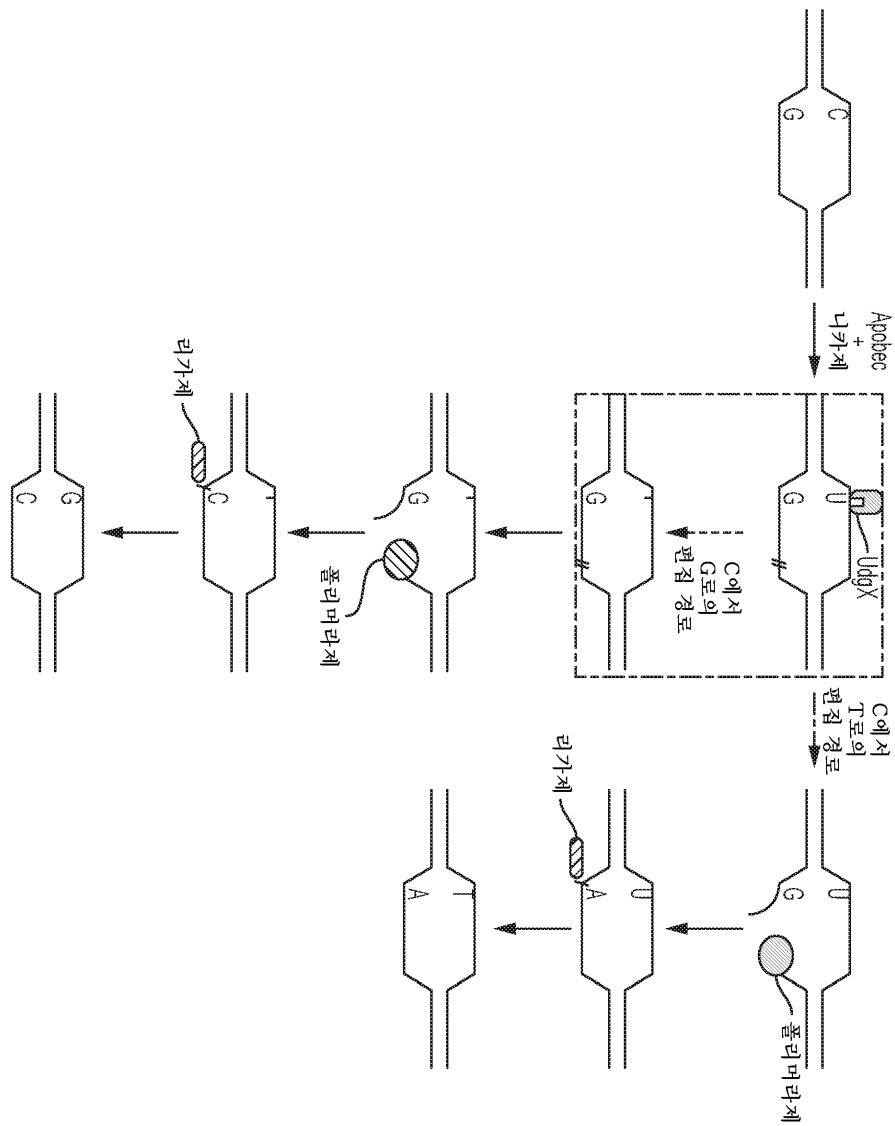
도면1



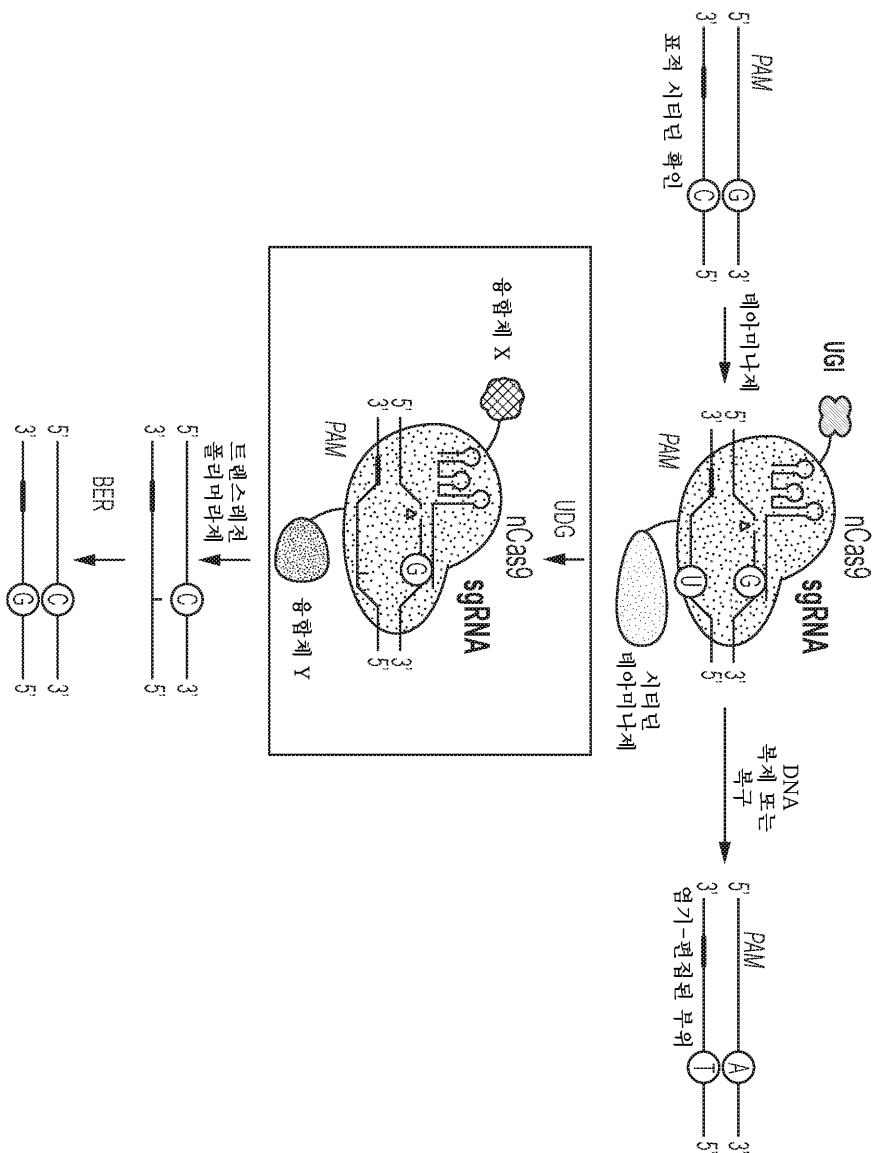
도면2



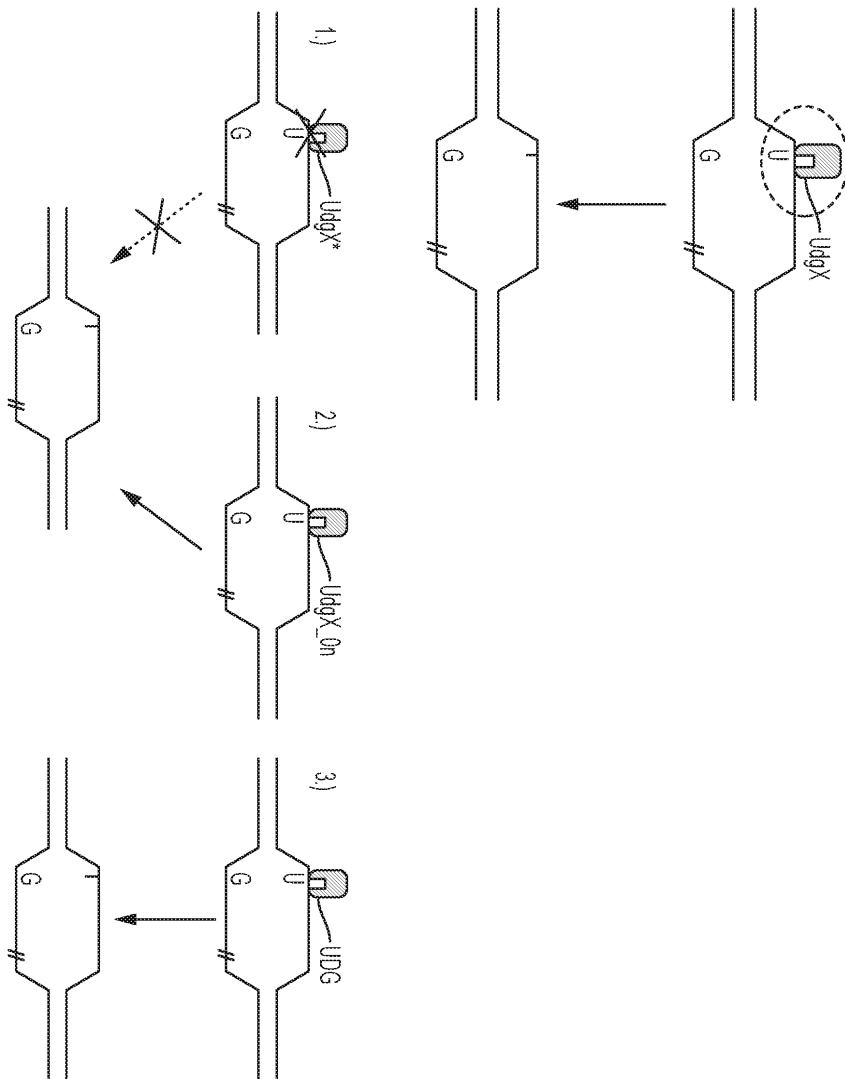
도면3



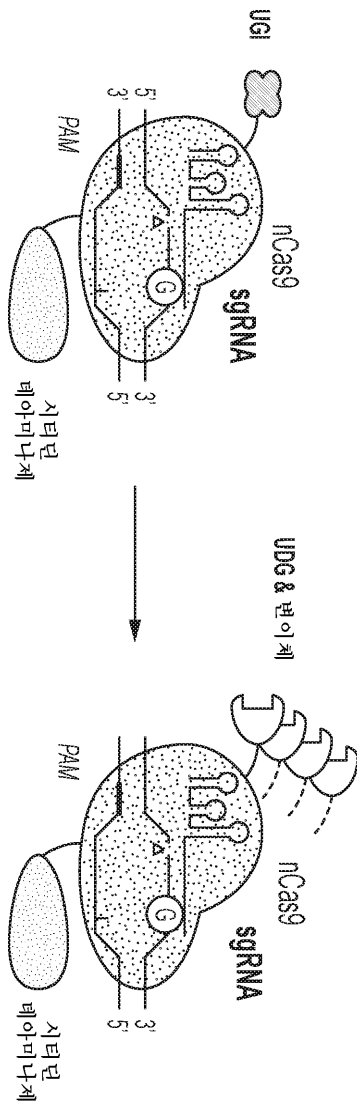
도면4



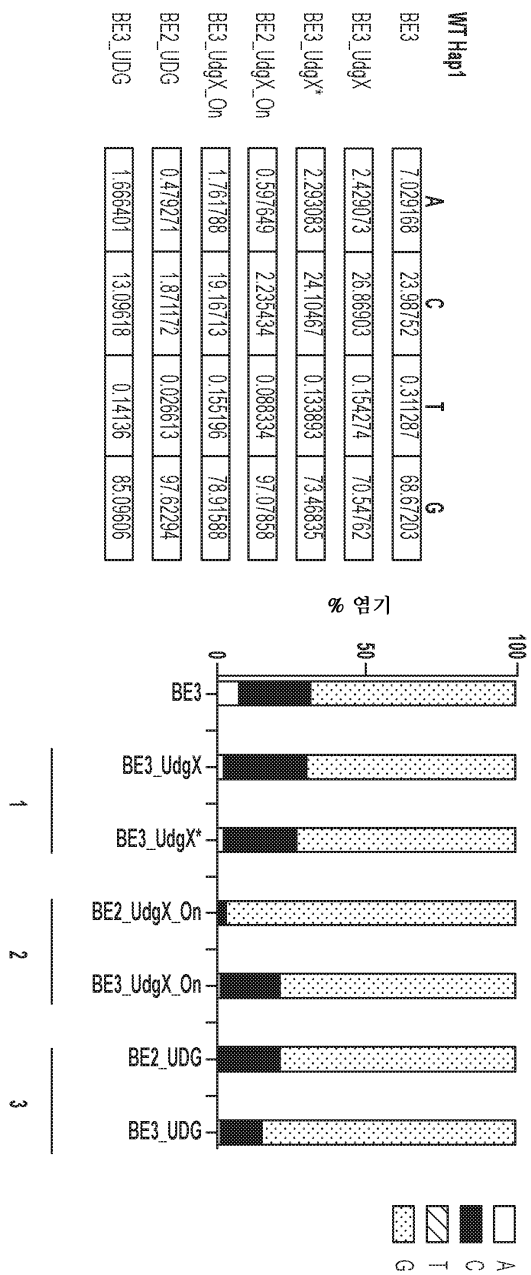
도면5



도면6



도면7

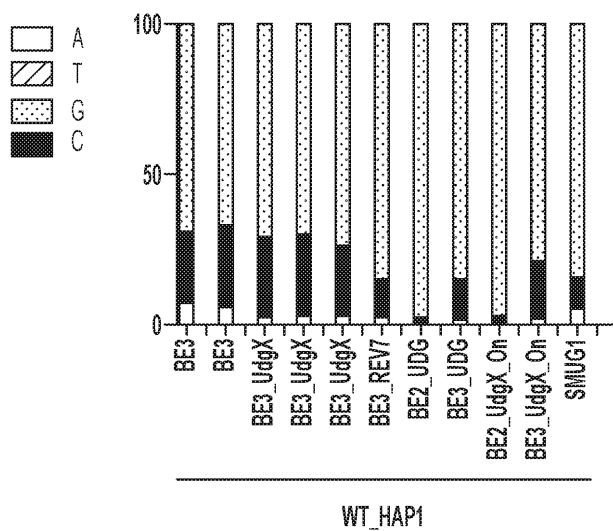


도면8

HEK2

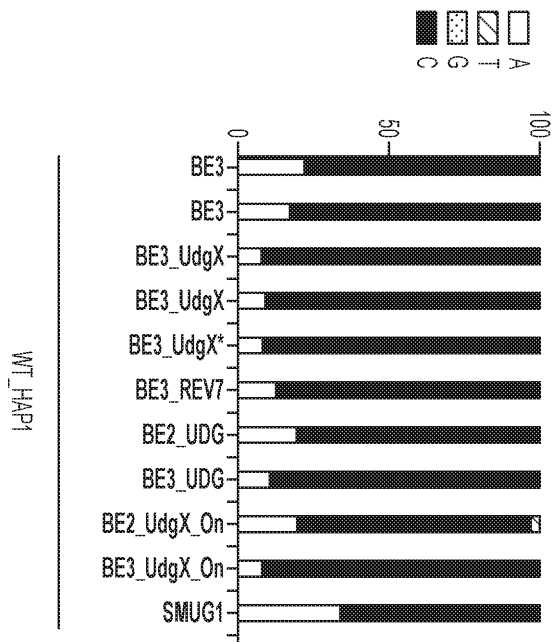
WT Hap1

	A	C	T	G
BE3	7.029168	23.98752	0.311287	68.67203
BE3	5.872296	27.17249	0.27076	66.68446
BE3_UdgX	2.429073	26.86903	0.154274	70.54762
BE3_UdgX	2.908524	26.48944	0.438746	70.1633
BE3_UdgX*	2.293083	24.10467	0.133893	73.46835
BE3_REV7	2.010573	12.94713	0.013356	85.02894
BE2_UDG	0.479271	1.871172	0.026613	97.62294
BE3_UDG	1.666401	13.09618	0.14136	85.09606
BE2_UdgX_On	0.597649	2.235434	0.088334	97.07858
BE3_UdgX_On	1.761788	19.16713	0.155196	78.91588
SMUG1	5.542508	10.38046	0.234852	83.84218

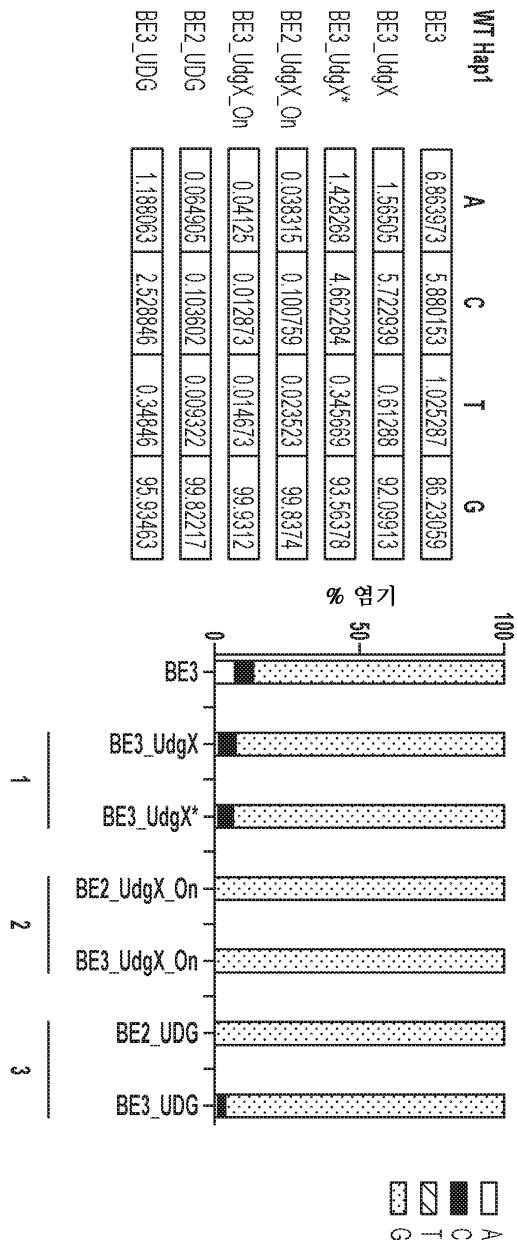


도면9

HEK2				
WT Hap1				
	총 편차	A	C	T
BE3	31.3279752	22.43735	76.56901	0.99364
BE3	33.315546	17.62629	81.56099	0.812714
BE3_UdgX	29.4523771	8.24746	91.22873	0.523809
BE3_UdgX	29.8367097	9.748139	88.78137	1.47049
BE3_UdgX*	26.5316462	8.642822	90.85252	0.504655
BE3_REV7	14.97105859	13.42973	86.48106	0.089209
BE2_UDG	2.37705617	20.16239	78.71804	1.119573
BE3_UDG	14.9039407	11.18094	87.87059	0.948472
BE2_UdgX_On	2.921416827	20.45751	76.51882	3.023668
BE3_UdgX_On	21.08411823	8.355998	90.90792	0.73608
SMUG1	16.15782057	34.30233	64.24419	1.453468

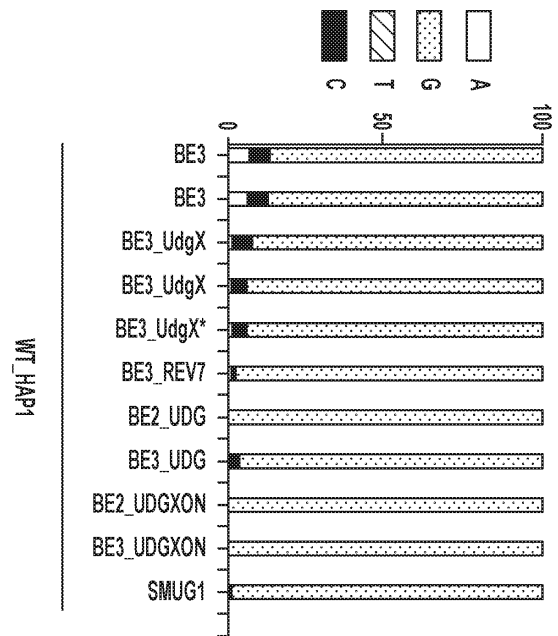


도면10



도면11

RNF2				
WT Hap1				
	A	C	T	G
BE3	6.863973	5.880153	1.025287	86.23059
BE3	6.501816	5.566838	0.980042	86.9613
BE3_UdgX	1.56505	5.722939	0.61288	92.09913
BE3_UdgX	1.512982	4.334513	0.317906	93.8346
BE3_UdgX	1.428268	4.662284	0.345669	93.56378
BE3_REV7	0.971986	1.605961	0.14826	97.2738
BE2_UDG	0.064905	0.103602	0.009322	99.82217
BE3_UDG	1.188063	2.528846	0.34846	95.93463
BE2_UdgX_On	0.038315	0.100759	0.023523	99.8374
BE3_UdgX_On	0.04125	0.012873	0.014673	99.9312
SMUG1	0.799327	0.799327	0.12621	98.27514

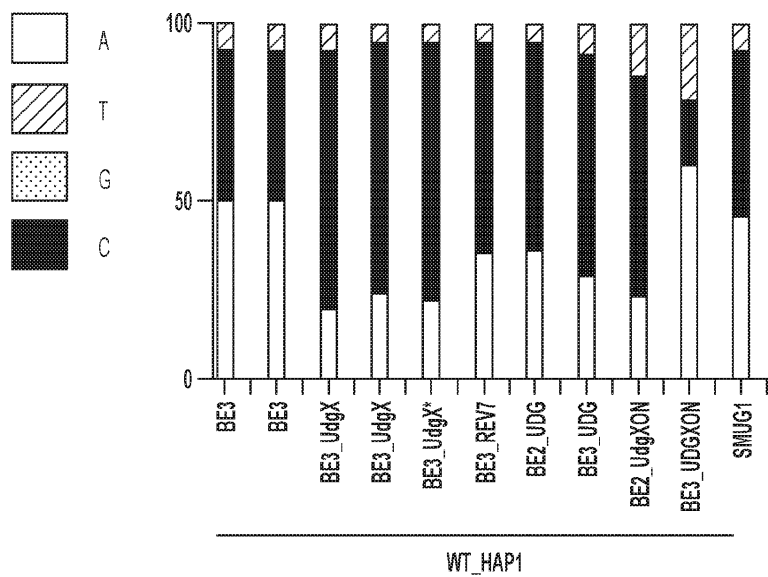


도면12

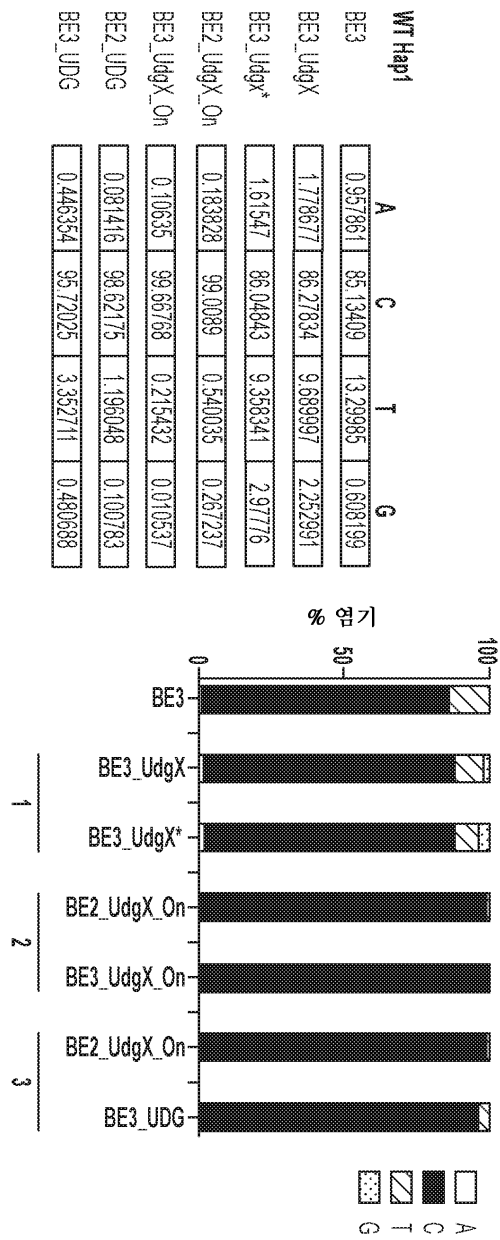
RNF2

WT Hap1

총 편집	A	C	T
BE3	13.769413	49.84942	42.70446
BE3	13.038696	49.86554	42.61805
BE3_UdgX	7.9008694	19.80858	72.43429
BE3_UdgX	6.1654013	24.53988	70.30383
BE3_UdgX	6.4362206	22.1911	72.43823
BE3_REV7	2.7262062	35.65341	58.90827
BE2_UDG	0.177828896	36.49852	58.25915
BE3_UDG	4.0653692	29.22399	62.20458
BE2_UdgX_On	0.162597289	23.5645	61.96868
BE3_UdgX_On	0.068796831	59.95976	18.71227
SMUG1	1.724863273	46.34146	46.34146



도면13

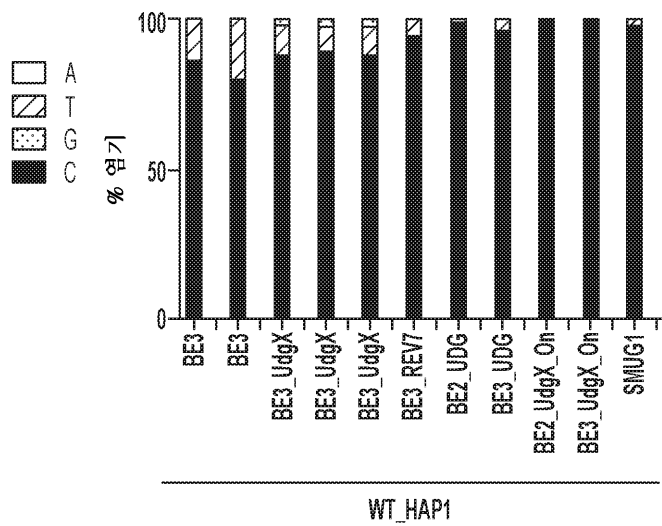


도면14

FANCF

WT Hap1

	A	C	T	G
BE3	0.957861	85.13409	13.29985	0.608199
BE3	1.318066	78.56728	19.15355	0.961108
BE3_UdgX	1.778677	86.27834	9.689997	2.252991
BE3_UdgX	1.764677	87.32044	8.332098	2.582781
BE3_UdgX	1.61547	86.04843	9.689997	2.97776
BE3_REV7	0.514632	93.70418	5.31733	0.463864
BE2_UDG	0.081416	98.62175	1.196048	0.100783
BE3_UDG	0.446354	95.72025	3.352711	0.480688
BE2_UdgXON	0.183828	99.0089	0.540035	0.267237
BE3_UDGXON	0.10635	99.66768	0.215432	0.010537
SMUG1	0.330974	97.17443	2.400242	0.094351

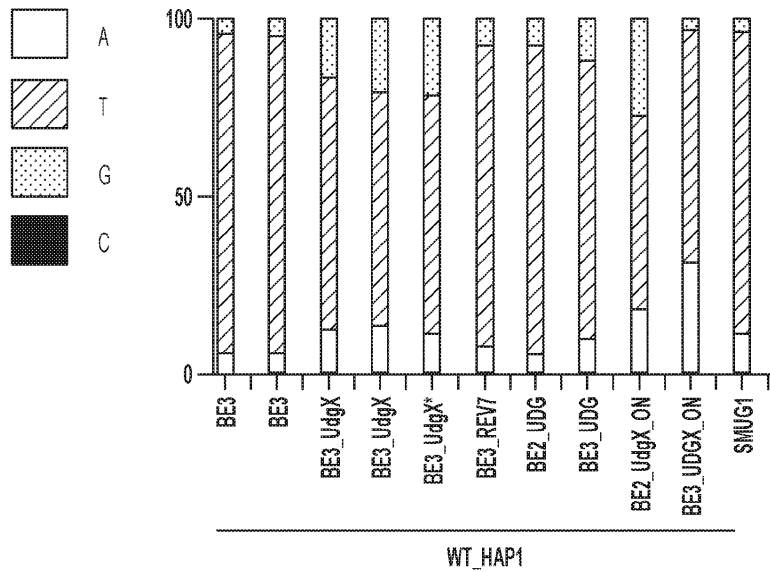


도면15

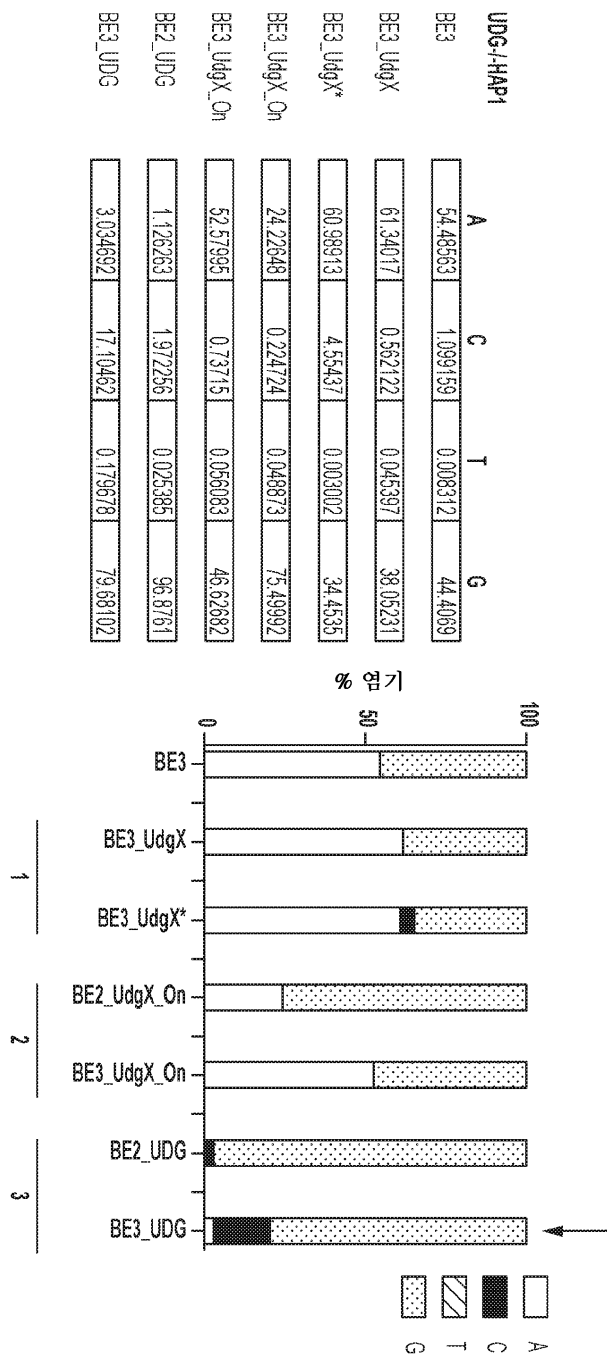
FANCF

WT Hap1

	총 편집	A	T	G
BE3	14.8659096	6.443336	89.46543	4.091234
BE3	21.4327257	6.149792	89.36591	4.4843
BE3_UdgX	13.721665	12.96254	70.61823	16.41922
BE3_UdgX	12.679556	13.9175	65.71285	20.36965
BE3_UdgX	13.951571	11.57913	67.07733	21.34355
BE3_REV7	6.2958255	8.174178	84.45803	7.367795
BE2_UDG	1.37824733	5.907229	86.78036	7.31241
BE2_UDG	4.279753	10.42942	78.33889	11.23169
BE2_UdgXON	0.991099728	18.54785	54.48845	26.9637
BE3_UDGXON	0.332319917	32.00235	64.82678	3.170875
SMUG1	2.825567162	11.71353	84.94728	3.339192

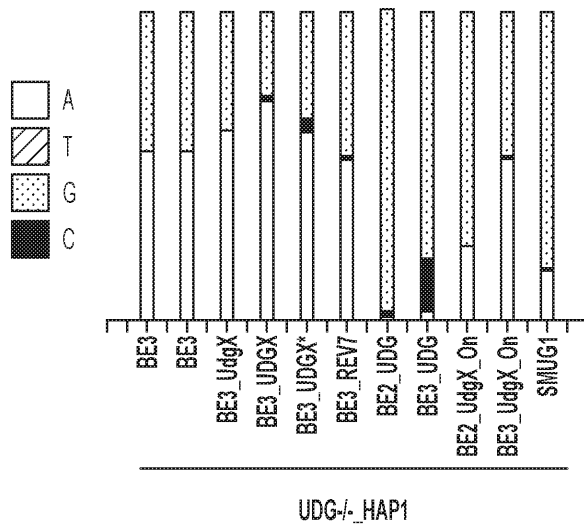


도면16



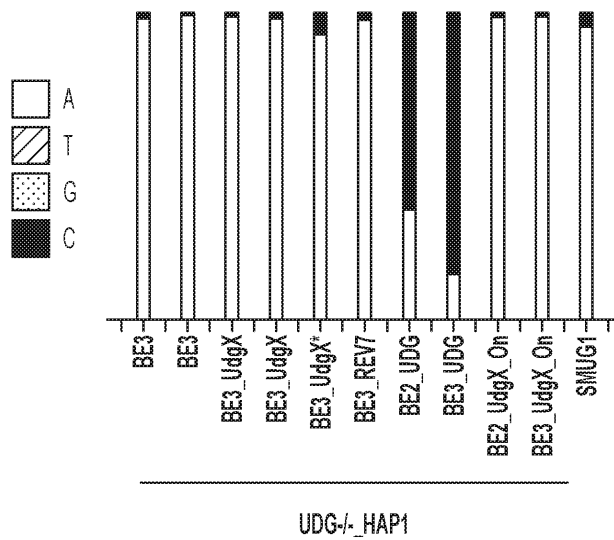
도면17

UDG-/- HAP1	A	C	T	G
BE3	54.48563	1.099159	0.008312	44.4069
BE3	54.88124	0.280479	0.037441	44.80084
BE3_UdgX	61.34017	0.562122	0.045397	38.05231
BE3_UDGX	71.34399	1.200882	0.015775	27.43935
BE3_UDGX	60.98913	4.55437	0.003002	34.4535
BE3_REV7	52.13184	1.191155	0.028973	46.64804
BE2_UDG	1.126263	1.972256	0.025385	96.8761
BE3_UDG	3.034692	17.10462	0.179678	79.68102
BE2_UdgX_On	24.22648	0.224724	0.048873	75.49992
BE3_UdgX_On	52.57995	0.73715	0.056083	46.62682
SMUG1	16.2789	0.700013	0.052671	82.96842

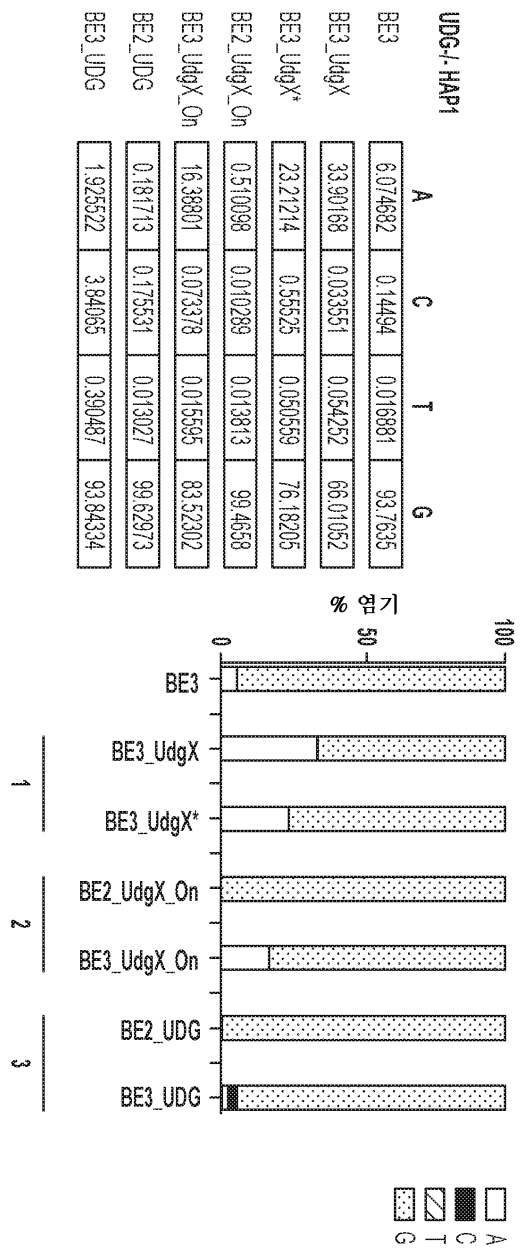


도면18

UDG-/- HAP1	총 편집	A	C	T
BE3	55.59310115	98.0079	1.97715	0.014952
BE3	55.19916029	99.42405	0.508121	0.06783
BE3_UdgX	61.94768881	99.0193	0.907414	0.073283
BE3_UDGX	72.56064713	98.32325	1.655005	0.021741
BE3_UDGX	65.54650222	93.04712	6.948304	0.00458
BE3_REV7	53.35196778	97.71306	2.232636	0.054305
BE2_UDG	3.12390396	36.05306	63.13434	0.812604
BE3_UDG	20.3189896	14.93525	84.18047	0.884284
BE2_UdgX_On	24.50008069	98.88328	0.917237	0.19948
BE3_UdgX_On	53.37318037	98.5138	1.381124	0.105076
SMUG1	17.03158156	95.58066	4.110088	0.309256

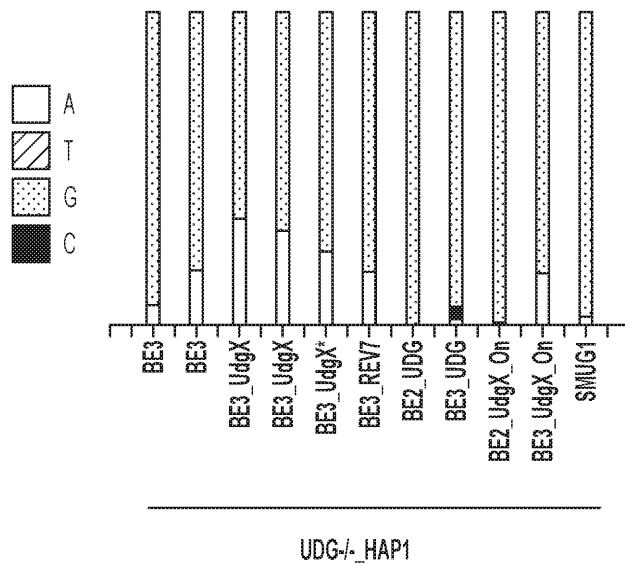


도면19



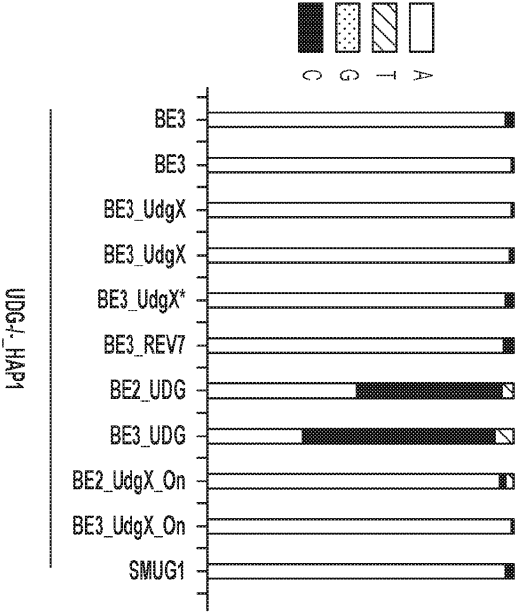
도면20

UDG-/- HAP1	A	C	T	G
BE3	6.074682	0.14494	0.016881	93.7635
BE3	17.20021	0.014772	0.073862	82.71116
BE3_UdgX	33.90168	0.033551	0.054252	66.01052
BE3_UDGX	29.92735	0.262275	0.056048	69.75433
BE3_UDGX	23.21214	0.55525	0.050559	76.18205
BE3_REV7	16.81834	0.510068	0.045985	82.62561
BE2_UDG	0.181713	0.175531	0.013027	99.62973
BE3_UDG	1.925522	3.84065	0.390487	93.84334
BE2_UdgX_On	0.510098	0.010289	0.013813	99.4658
BE3_UdgX_On	16.38801	0.073378	0.015595	83.52302
SMUG1	2.500624	0.048749	0.021911	97.42872

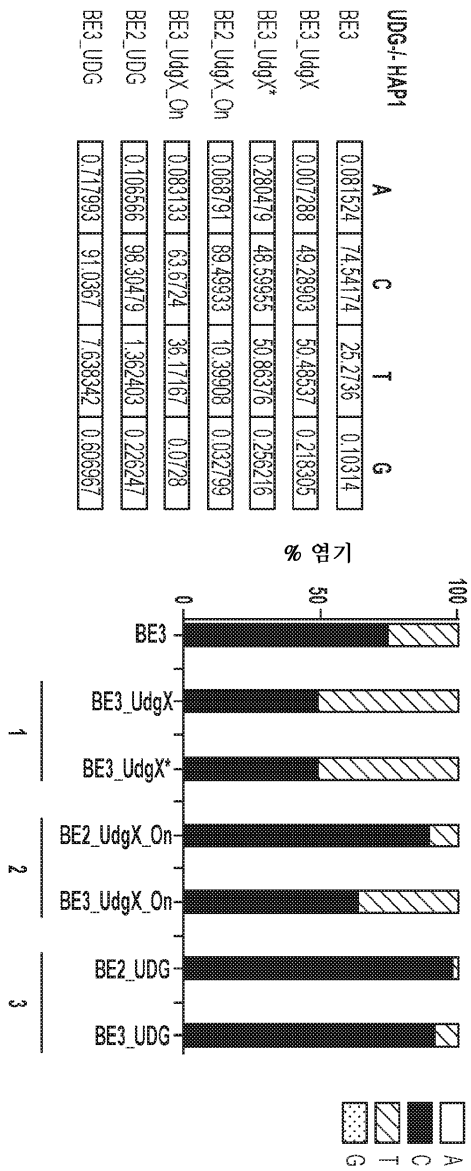


도면21

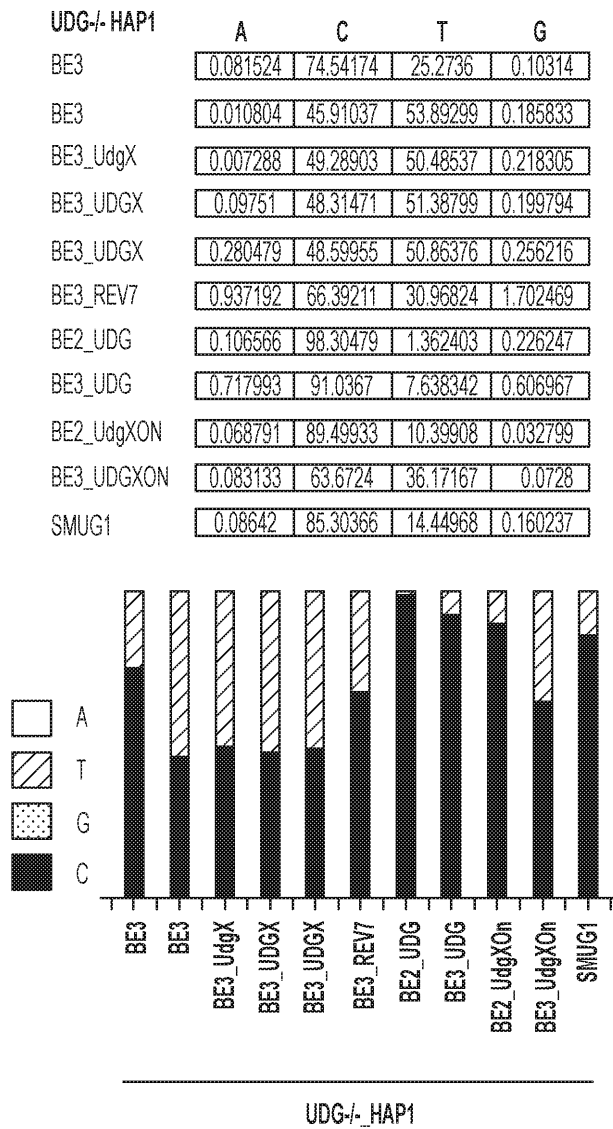
UDG ⁺ -HAP1				
	총 편집	A	C	T
BE3	6.23650279	97.40526	2.324062	0.270674
BE3	17.28884399	99.48733	0.085444	0.427222
BE3_UdgX	33.98948321	99.74168	0.098709	0.159615
BE3_UDGX	30.24567257	98.94754	0.867148	0.185308
BE3_UDGX	23.81794891	97.4565	2.331223	0.212274
BE3_REV7	17.374393	96.79958	2.935747	0.26467
BE2_UDG	0.37027144	49.07573	47.40609	3.518187
BE3_UDG	6.1566565	31.27544	62.38205	6.342507
BE2_UdgX_On	0.53420081	95.48813	1.926121	2.585752
BE3_UdgX_On	16.47696452	99.46002	0.445334	0.094646
SMUG1	2.57128436	97.25195	1.895891	0.85216



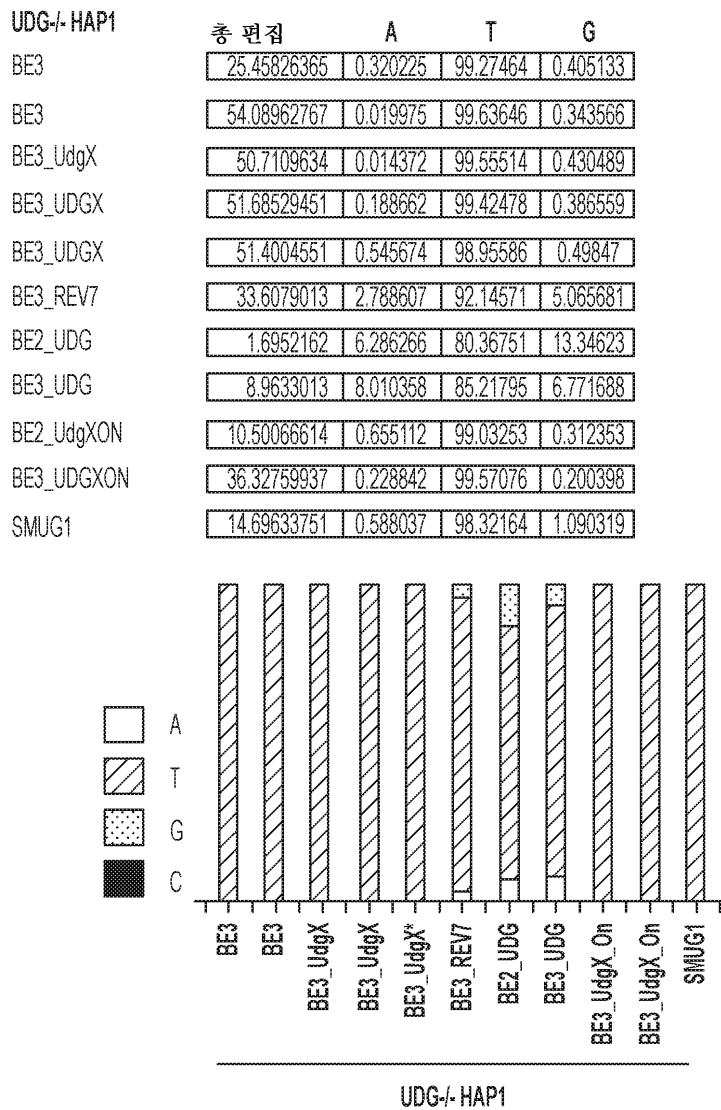
도면22



도면23



도면24



도면25

REV1-/-_HAP1				
BE3	A	C	T	G
	27.90154	38.82184	0.21661	33.06002

BE3_UdgX	5.394911	62.86946	0.220717	31.51492

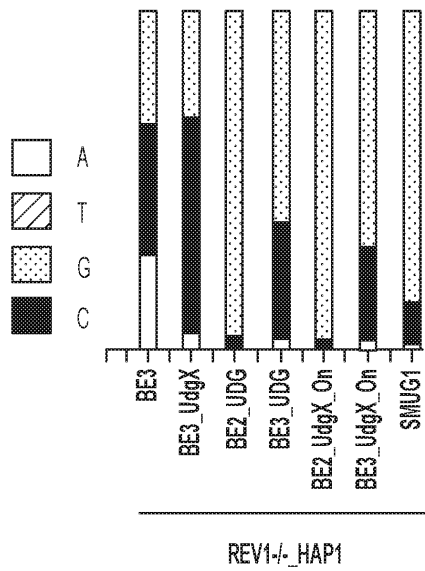
BE2_UDG	0.583164	3.171128	0.018236	96.22747

BE3_UDG	3.612539	33.5042	0.285178	62.59809

BE2_UdgX_On	0.787453	2.146845	0.072979	96.99272

BE3_UdgX_On	2.845842	27.20856	0.126609	69.81899

SMUG1	2.136539	12.36191	0.104799	85.39675



도면26

REV1-/-_HAP1	총 편집	A	C	T
BE3	66.9399898	41.68142	57.99499	0.323588

BE3_UdgX	68.4850877	7.877497	91.80022	0.322284

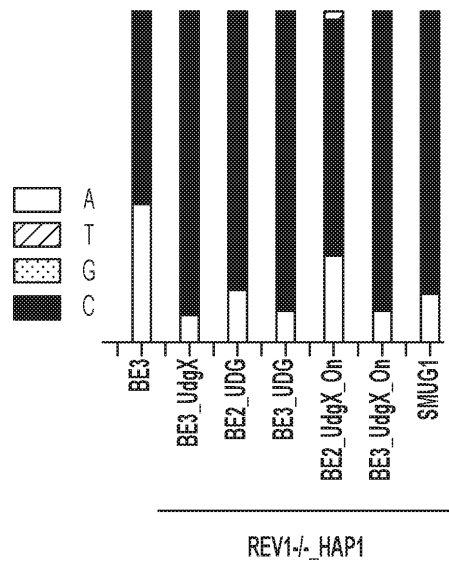
BE2_UDG	3.77252838	15.45818	84.05843	0.483394

BE3_UDG	37.4019174	9.6587	89.57883	0.76247

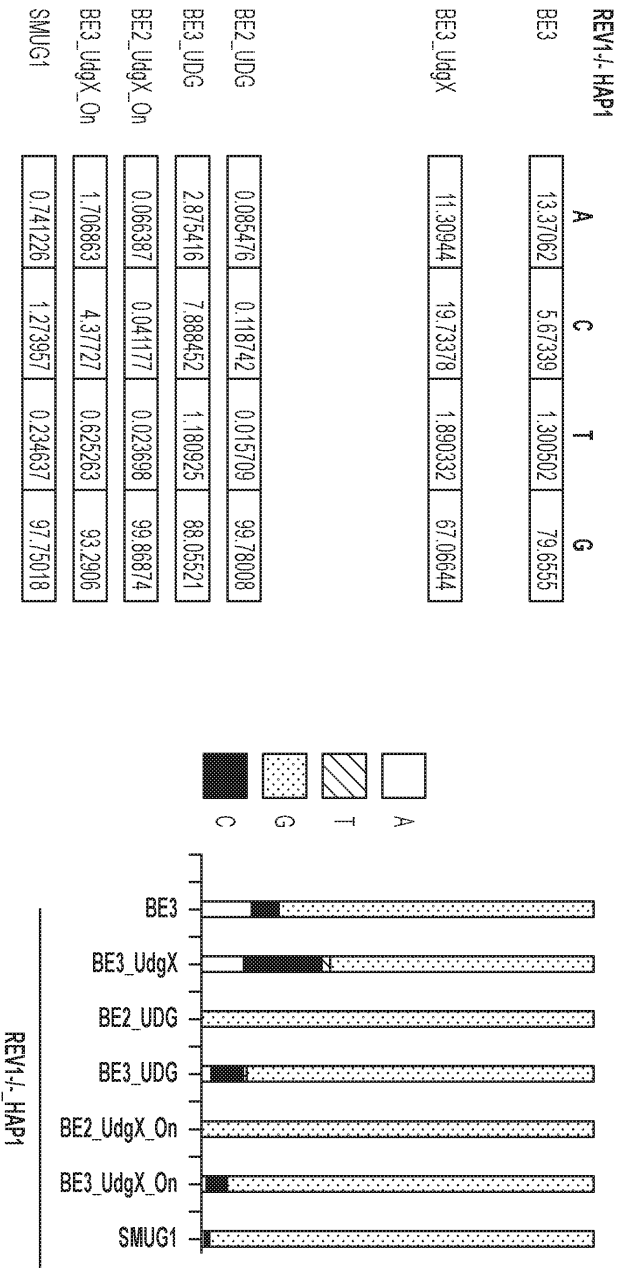
BE2_UdgX_On	3.007276279	26.18491	71.38834	2.426748

BE3_UdgX_On	30.18101476	9.429247	90.151126	0.419498

SMUG1	14.60324913	14.63057	84.65179	0.717639

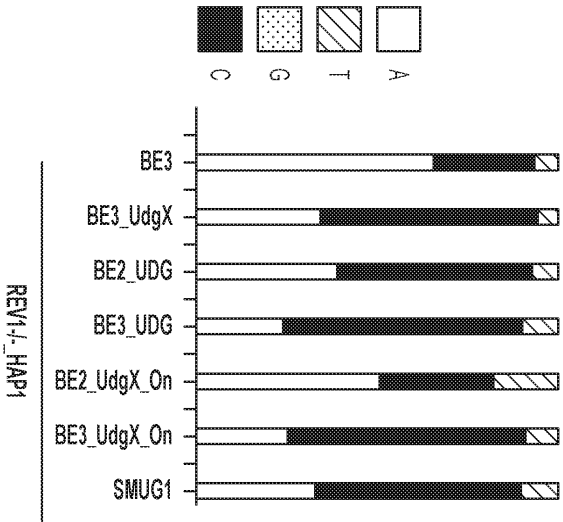


도면27



도면28

REV1-/- HAP1	총 편집			
	A	C	T	
BE3	20,344,512	65,721,0161	27,886,5868	6,392,9712
BE3_UdgX	32,933,552	34,340,1769	59,919,9868	5,739,6338
BE2_UDG	0,219,2751	38,865,538	53,991,6084	7,142,8357
BE3_UDG	11,944,793	24,072,5478	66,040,9268	9,886,2545
BE2_UdgX_On	0,131,260725	50,576,1844	31,370,0384	18,053,7772
BE3_UdgX_On	6,709,395472	25,439,874	65,240,8983	9,319,21428
SMUG1	2,249,82028	32,946,027	56,624,8126	10,429,1604



도면29

REV1-/- HAP1

	A	C	T	G
BE3	1.850796	59.33426	38.02208	0.792859

BE3_UdgX	3.486185	65.86237	22.91265	7.738802

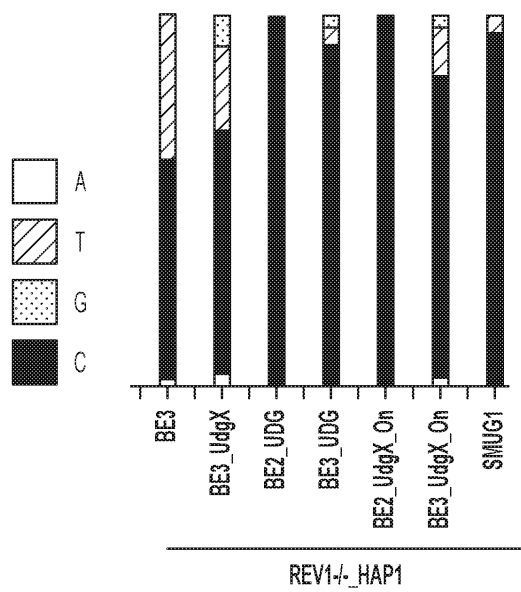
BE2_UDG	0.222439	98.32175	1.1797	0.276117

BE3_UDG	0.840939	91.43678	4.79438	2.927905

BE2_UdgXON	0.208095	98.36564	1.119386	0.306881

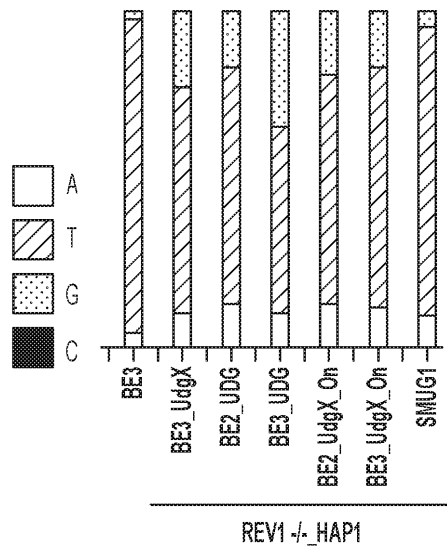
BE3_UDGXON	2.108873	82.15436	12.85805	2.878713

SMUG1	0.428121	95.46036	3.901712	0.20981

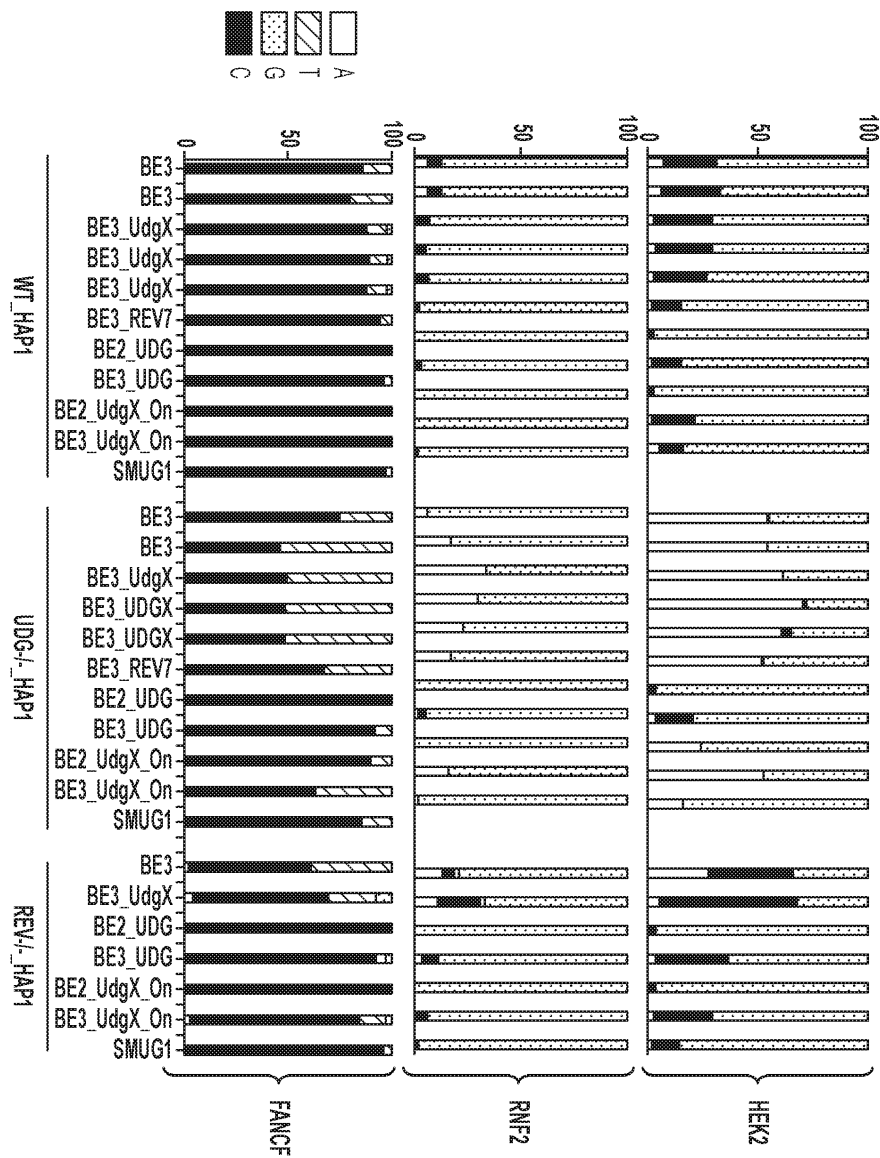


도면30

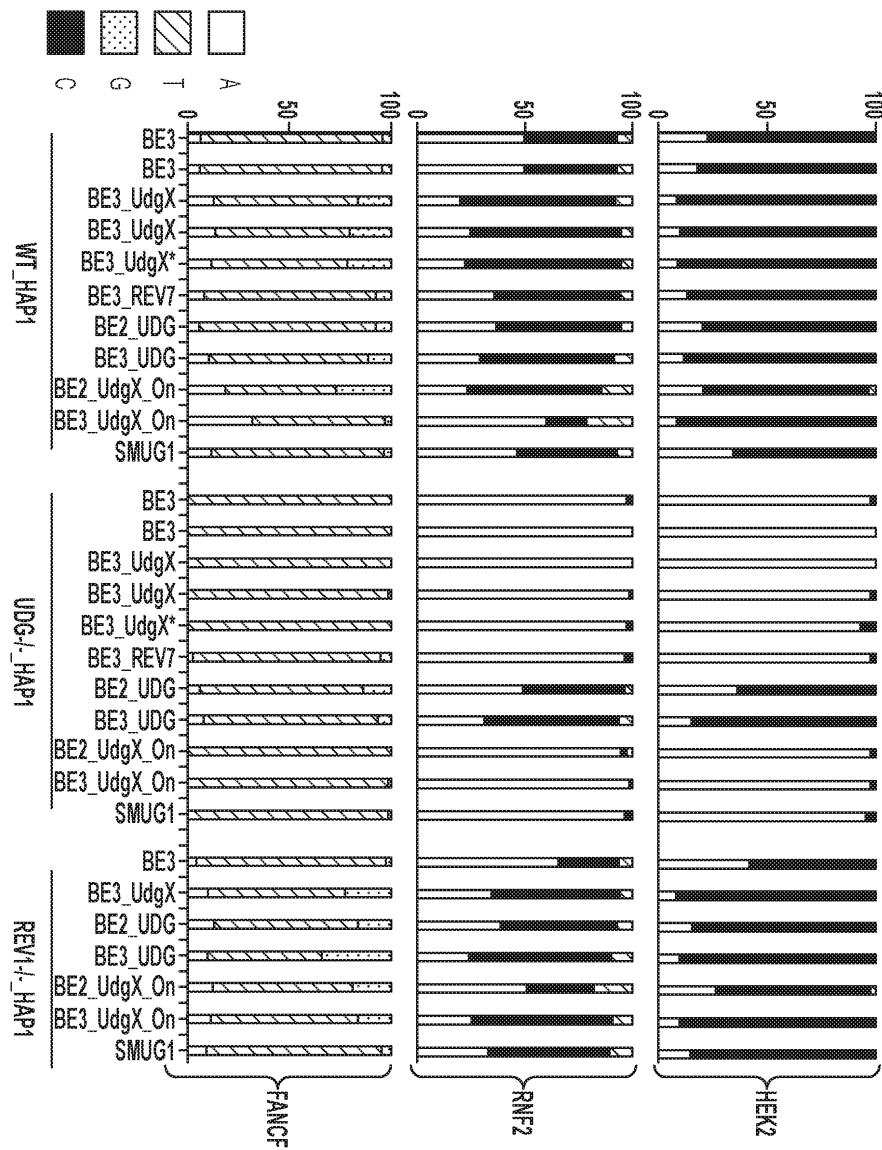
REV1 Δ -HAP1	총 편집	A	T	G
BE3	40.665735	4.55124197	93.4990601	1.94969795
BE3_UdgX	34.137637	10.212145	67.1184417	22.6694132
BE2_UDG	1.6782564	13.2541786	70.2931924	16.4526291
BE3_UDG	8.5632244	9.82035926	55.9880225	34.1916183
BE2_UdgXON	1.634361336	12.732475	68.490701	18.776824
BE3_UDGXON	17.84563906	11.8173022	72.051513	16.1311847
SMUG1	4.539642813	9.43071161	85.9475655	4.62172285



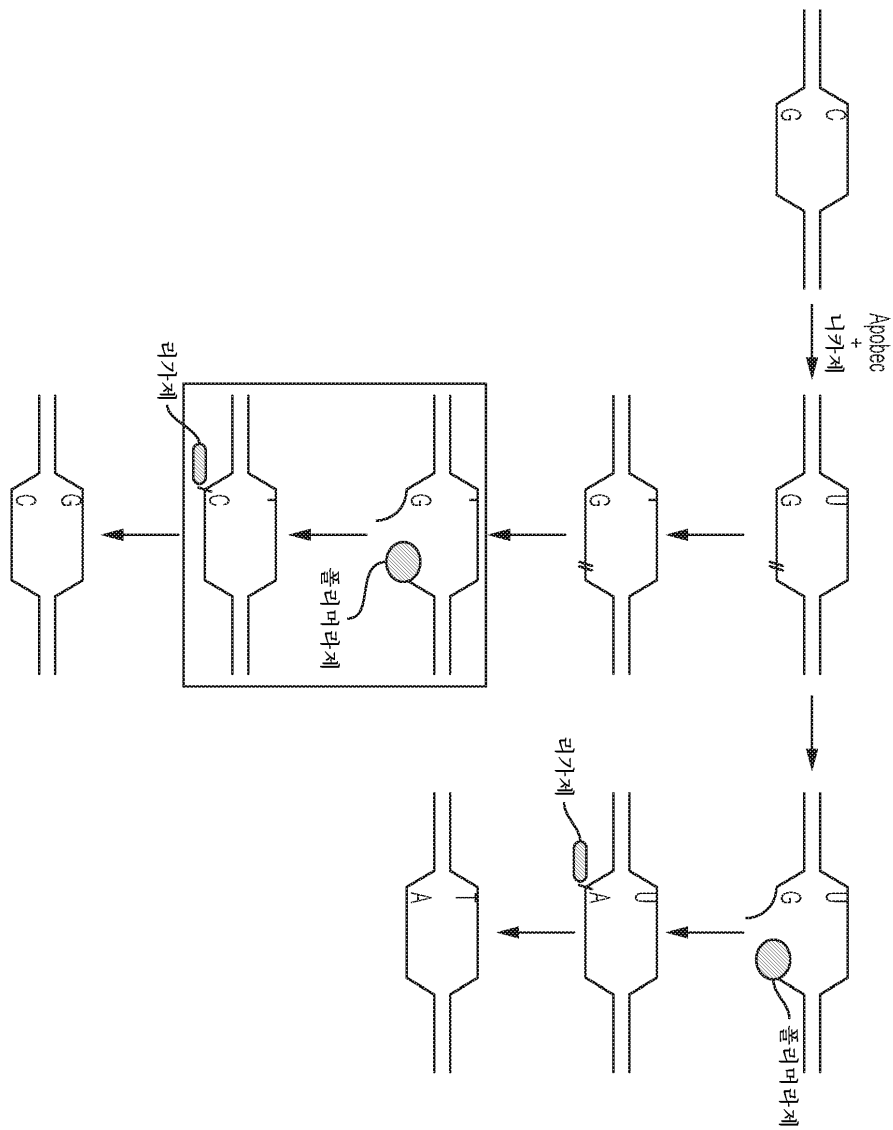
도면31



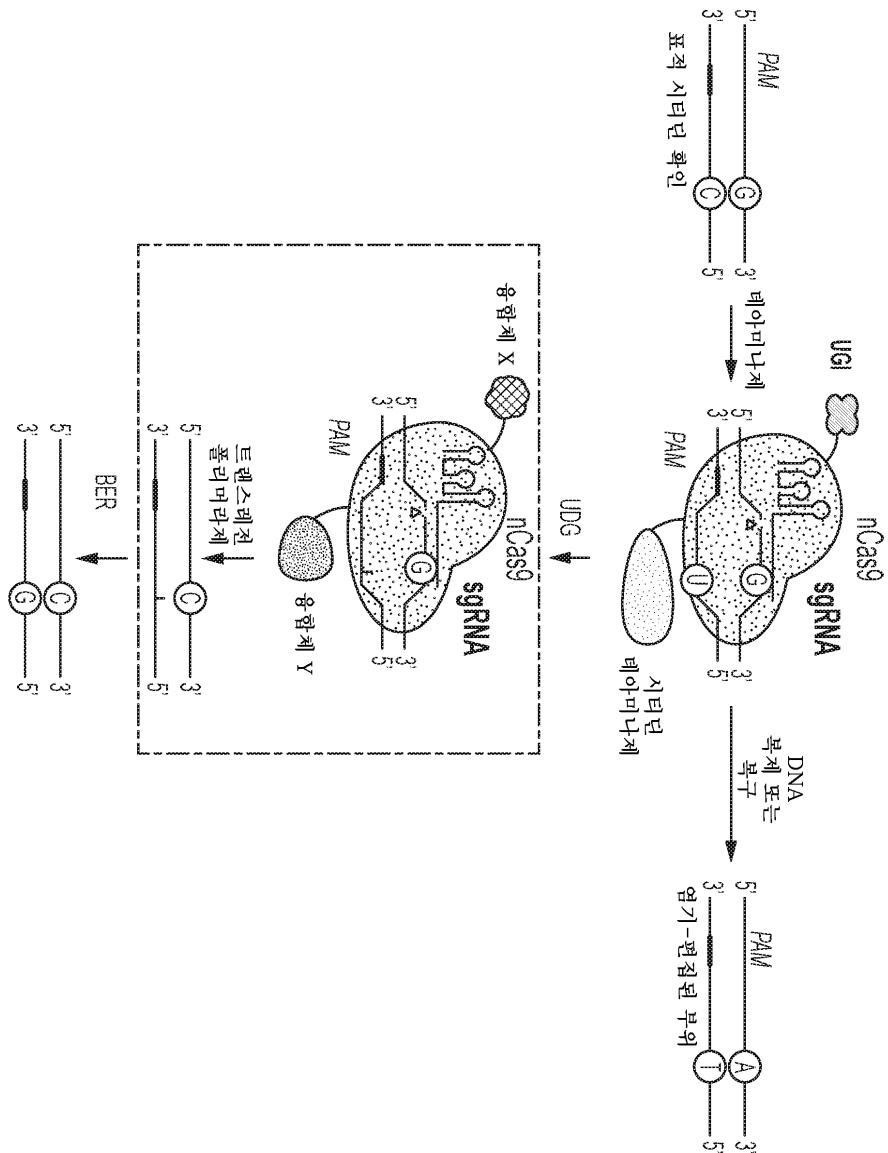
도면32



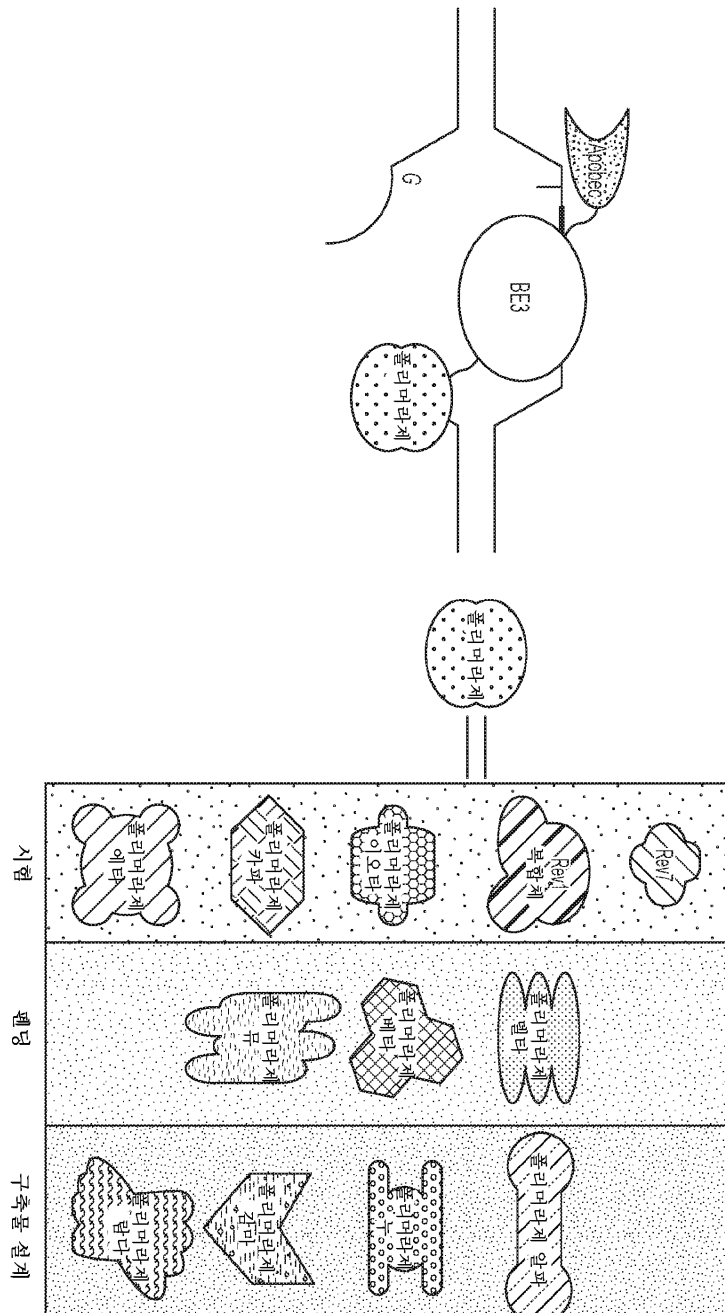
도면33



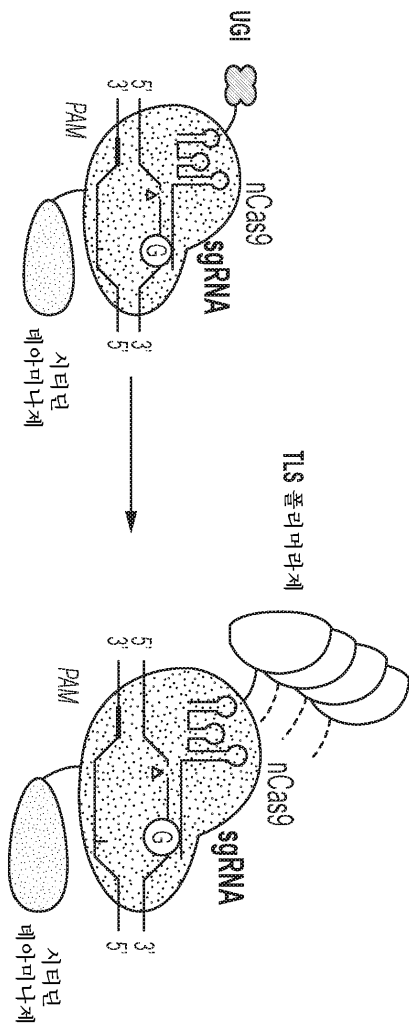
도면34



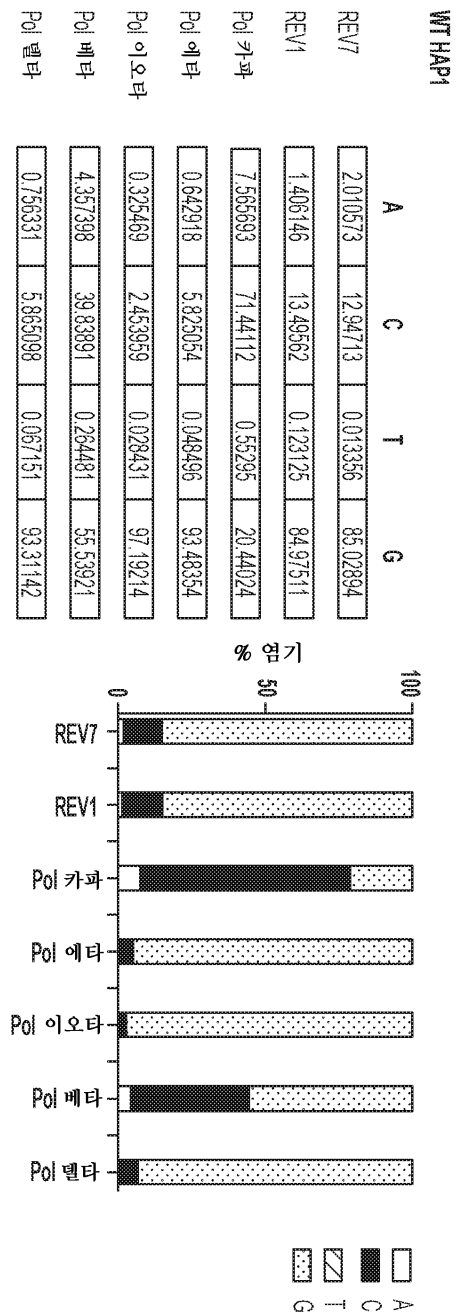
도면35



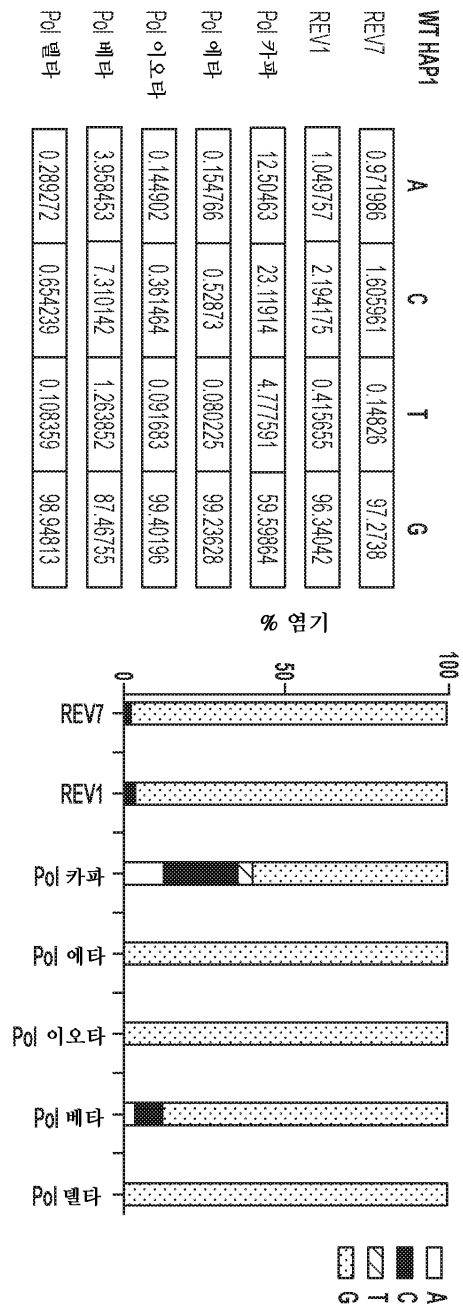
도면36



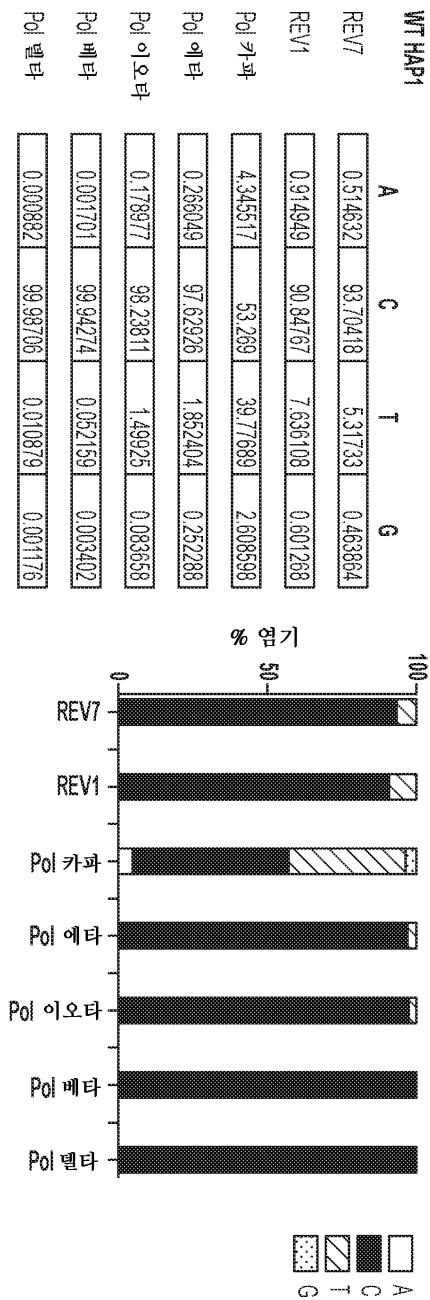
도면37



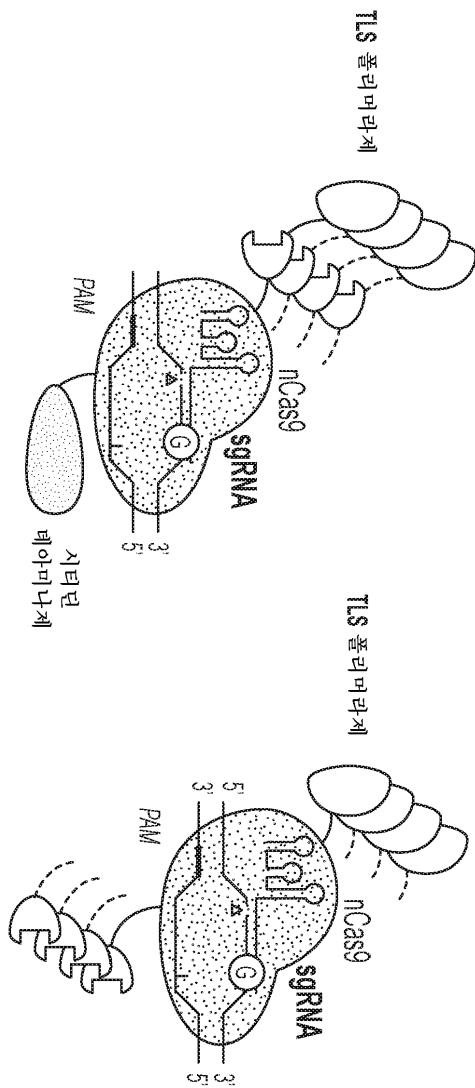
도면38



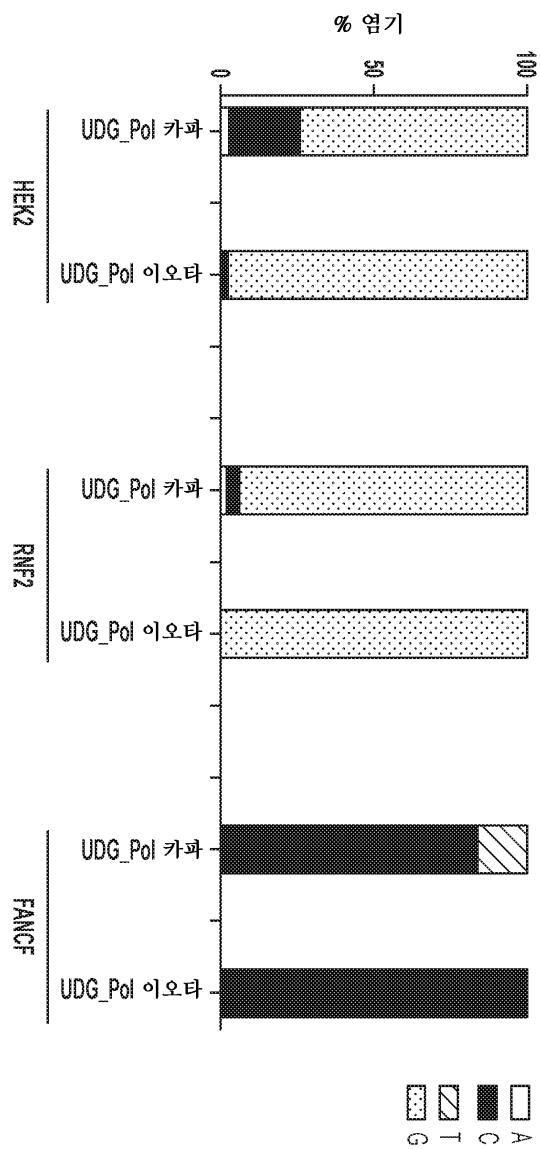
도면39



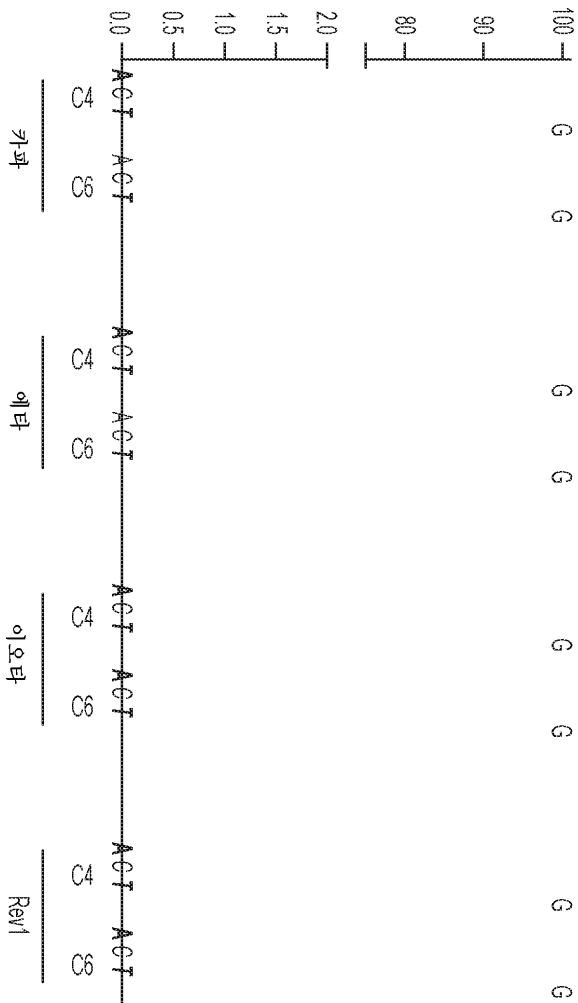
도면40



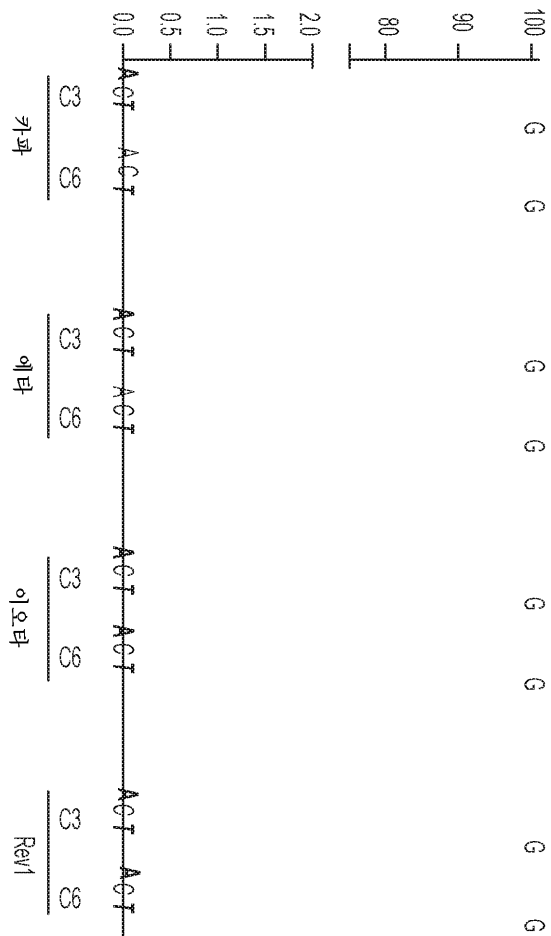
도면41



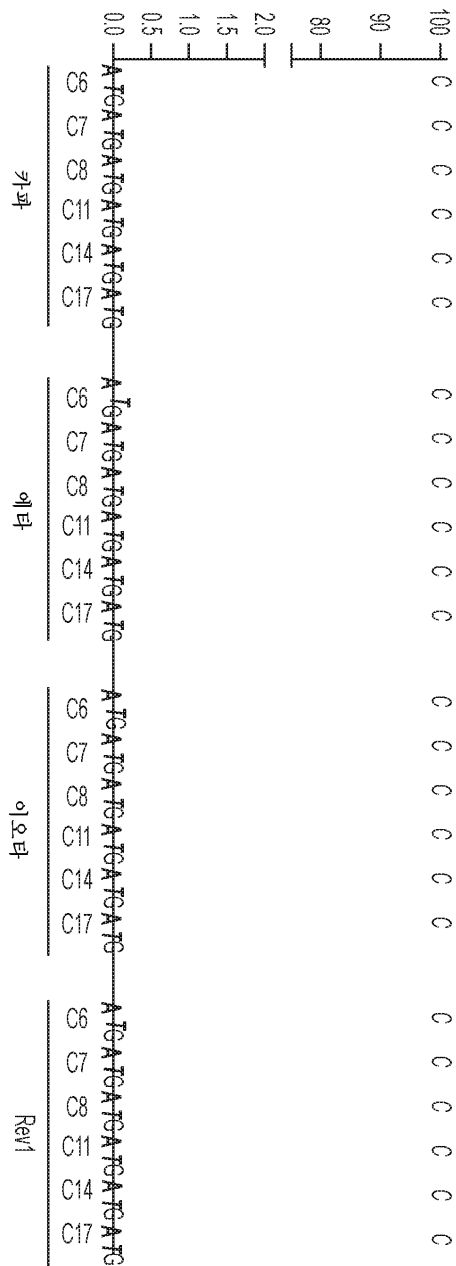
도면42



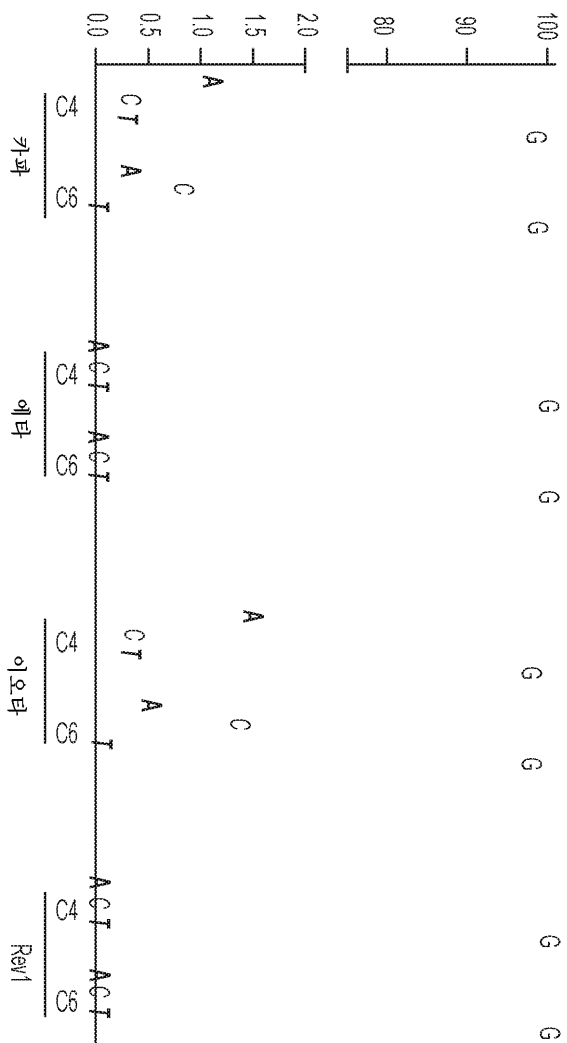
도면43



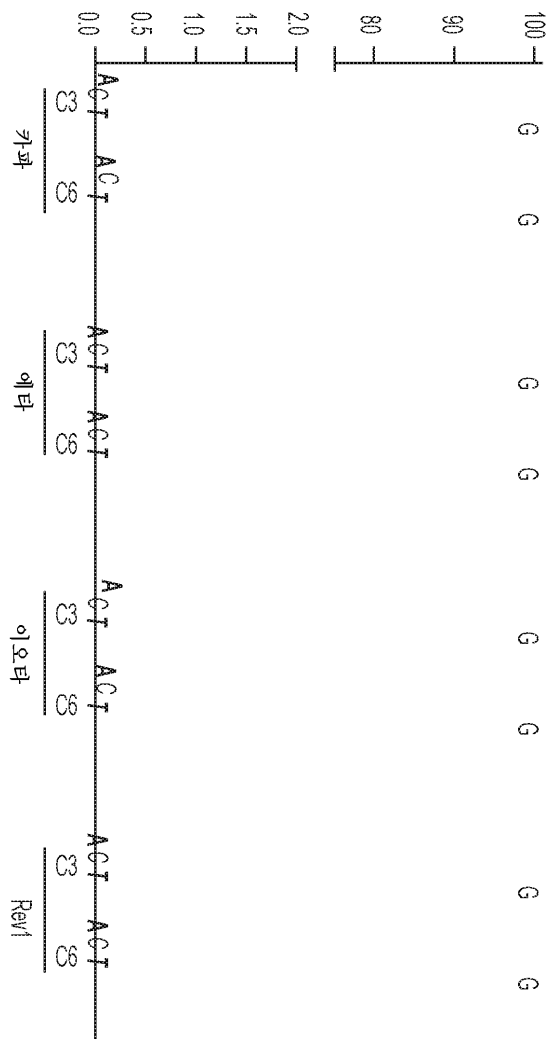
도면44



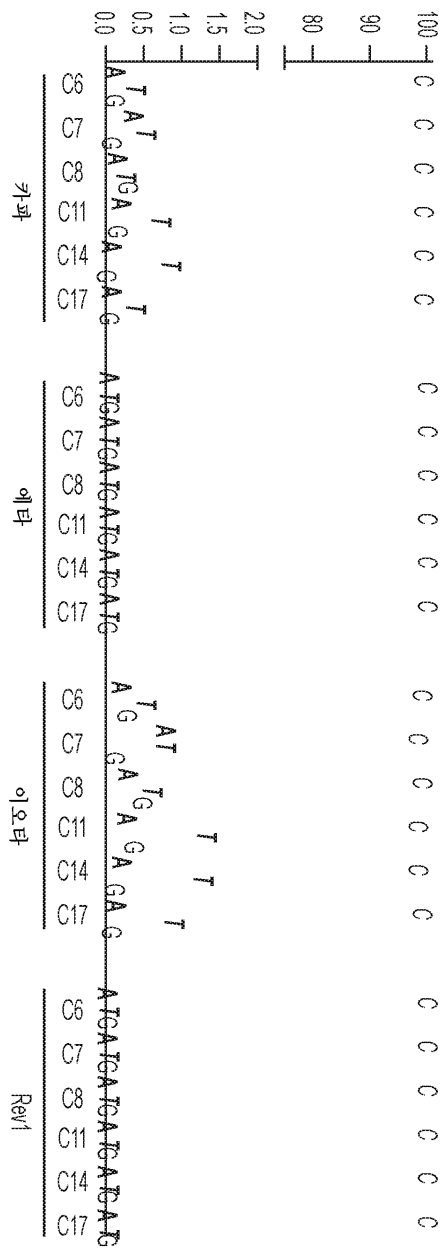
도면45



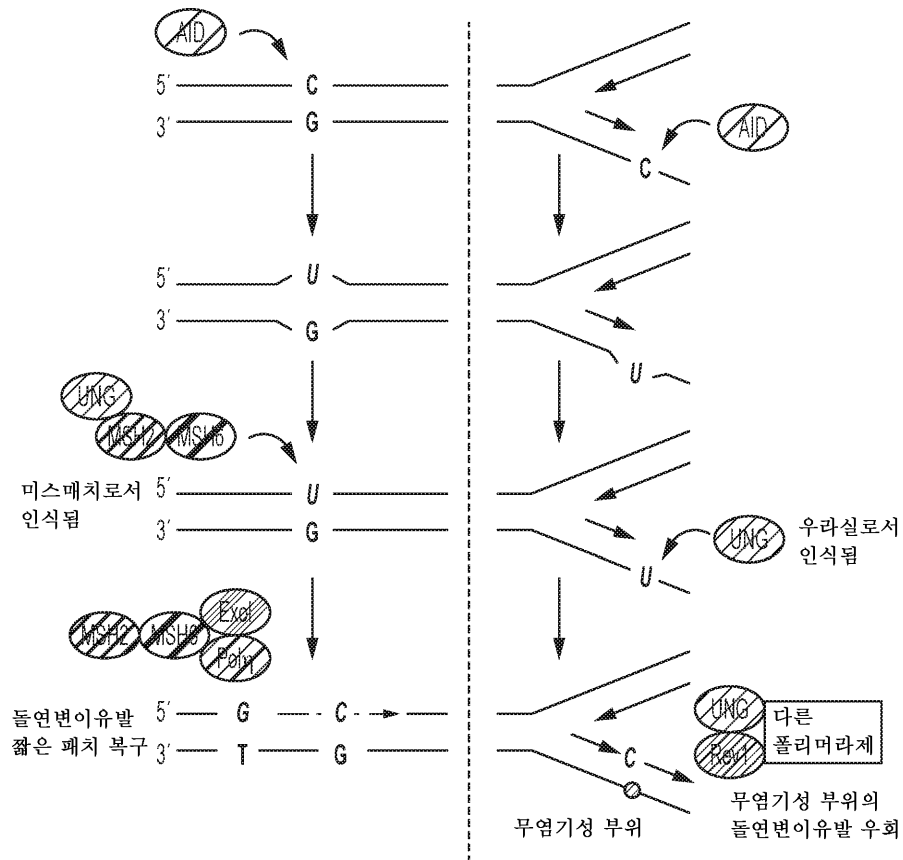
도면46



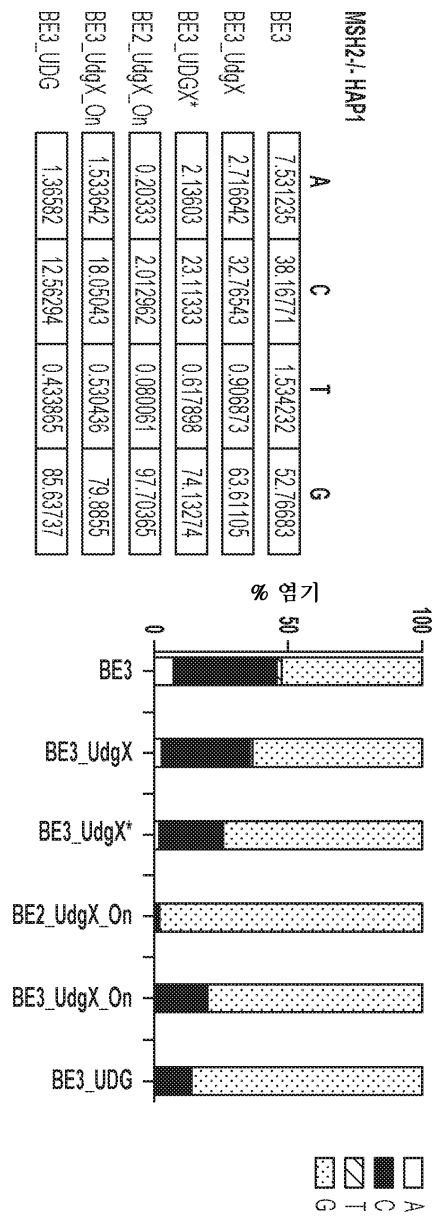
도면47



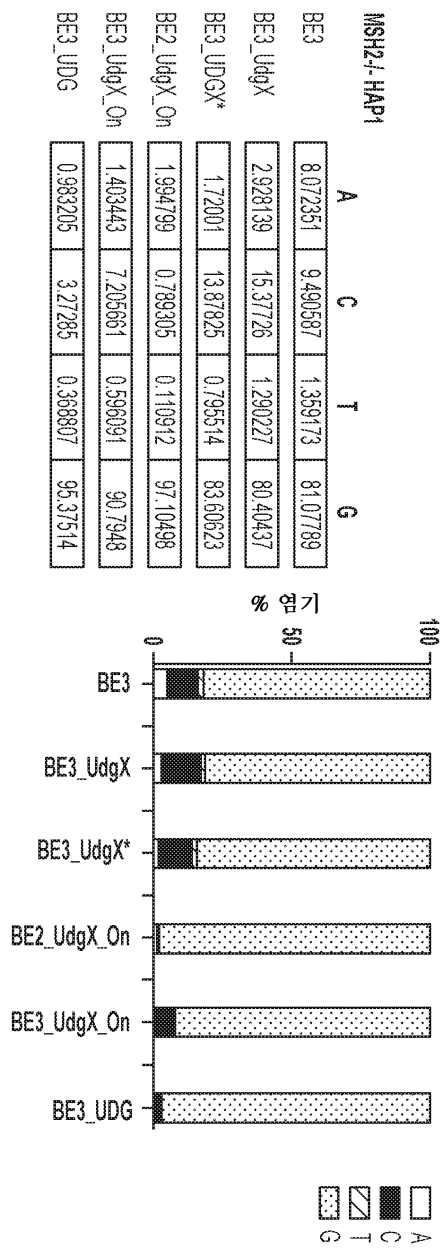
도면48



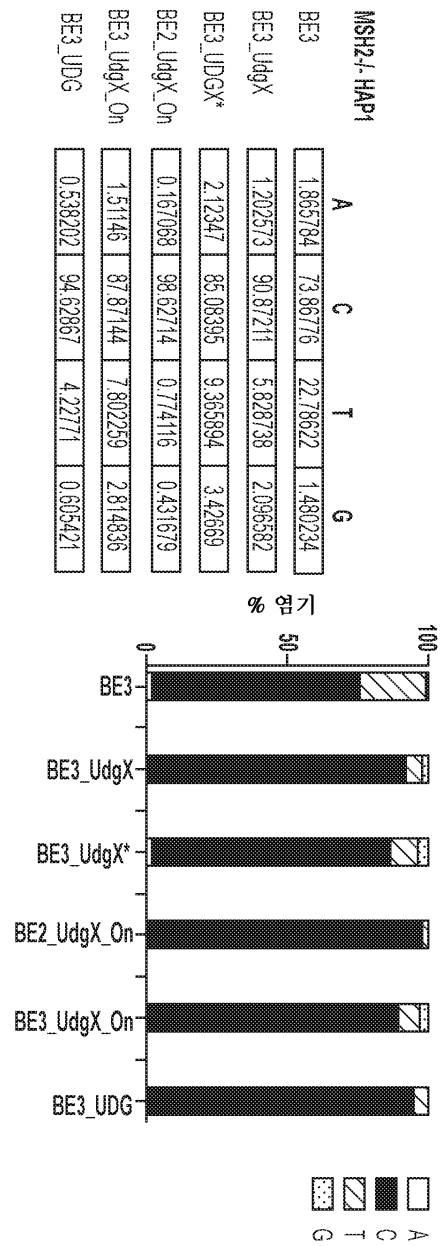
도면49



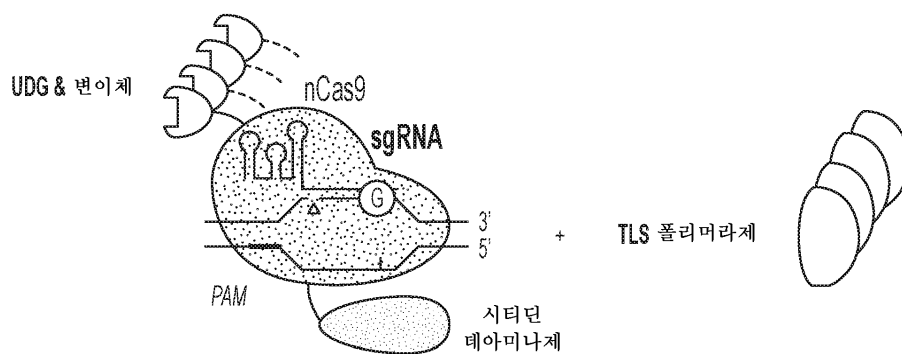
도면50



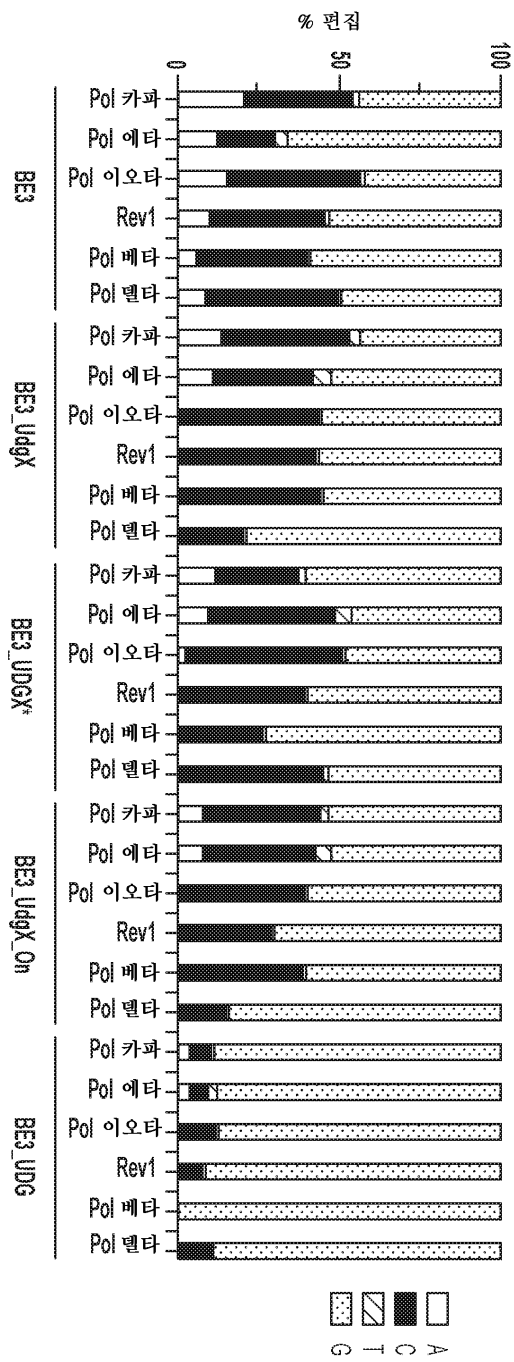
도면51



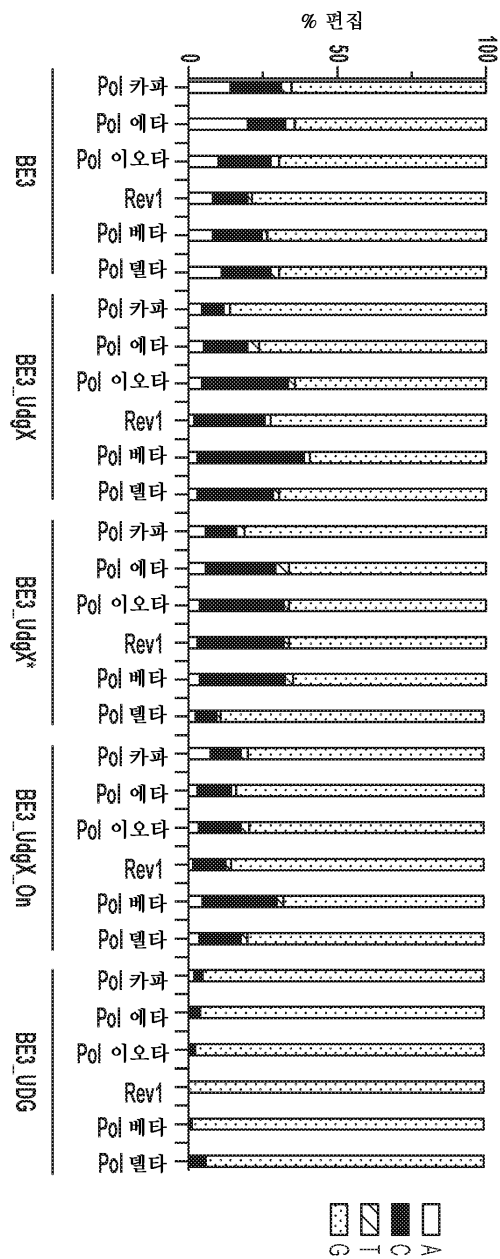
도면52



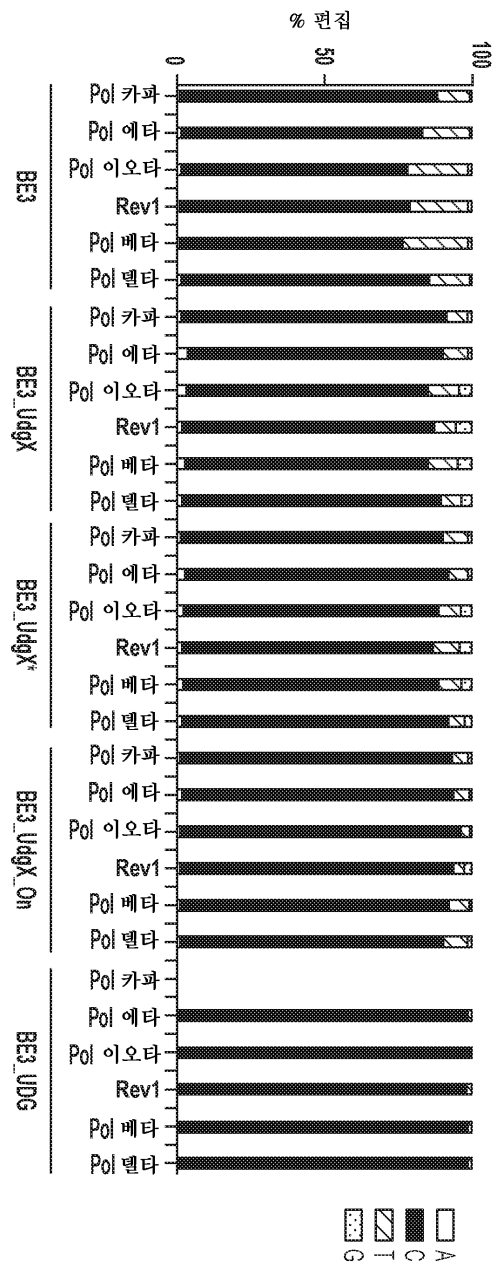
도면53



도면54



도면55



서열 목록

SEQUENCE LISTING

<110> President and Fellows of Harvard College

<120> CYTOSINE TO GUANINE BASE EDITOR

<130> H0824.70253W000

<140> Not Yet Assigned

<141> 2018-03-09

<150> US 62/470,175

<151> 2017-03-10

<160> 123

<170> PatentIn version 3.5

<210> 1

<211> 4104

<212> DNA

<213> Streptococcus pyogenes

<400> 1

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg	60
atcactgatg attataaggt tccgtctaaa aagttcaagg ttcctgggaaa tacagaccgc	120
cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa	180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt	240
tatctacagg agatTTTTTc aaatgagatg gcgaaagtag atgatatgtt ctttcacga	300
cttgaagagt ctttttttgt ggaagaagac aagaagcatg aacgtcatcc ttttttggaa	360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa	420
aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcgt	480
atgattaagt ttcgtgggtc ttttttgatt gagggagatt taaatcctga taatagtgt	540
giggacaaac tatttatcca gttggtacaa atctacaatc aattatttga agaaaaccct	600
attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga	660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttgggaat	720
ctcattgctt tgtcattggg attgaccctt aattttaaat caaattttga tttggcagaa	780
gatgctaaat tacagcttcc aaaagatact tacgatgatg atttagataa tttattggcg	840
caaattggag atcaatatgc tgatttgttt ttggcagcta agaattttatc agatgctatt	900
ttactttcag atatcctaag agtaaatagt gaaataacta aggcctccct atcagcttca	960
atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagtctga	1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca	1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta	1140
gaaaaaatgg atgggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc	1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcgt	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 actacaaaaa catagtttgc tttatgagta ttttacggtt	1560
tataacgaat tgacaaaggt caaatatgtt actgaggga tgcgaaaacc agcatttctt	1620
tcaggatgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc	1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt	1740
tcaggagtgt aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt	1800
attaaagata aagatTTTTT ggataatgaa gaaaatgaag atatcttaga ggatattgtt	1860
ttacattga cctatttga agataggggg atgattgagg aaagacttaa aacatatgct	1920
cacctctttg atgataaggt gatgaacag cttaaacgtc gccgttatac tggttgggga	1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta	2040
gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat	2100
agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttta	2160
catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggat tttacagact	2220
glaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcgttatt	2280
gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg	2340
aaacgaatcg aagaaggat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt	2400
gaaaatactc aattgcaaaa tgaagctc tatctctatt atctacaaaa tggaagagac	2460
atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt	2520
gttcacaaa 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
aaatacgtg aaaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa	2880
ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac	2940
catcatgcc atgatgcgta tctaaatgcc gtcgttggaa ctgctttgat taagaaatat	3000
ccaaaacttg aatcgaggtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg	3060
attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaat	3120
atcatgaact tcttcaaaa agaaattaca cttgcaaatg gagagattcg caaacgcct	3180
ctaatacgaaa ctaatgggga aactggagaa attgtctggg ataaaggcg agattttgcc	3240

acagtgcgca aagtattgtc catgccccaa gtcaatatgt tcaagaaaac agaagtacag 3300

acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct 3360

cgtaaaaaag actgggatcc aaaaaatat ggtgggtttg atagtccaac ggtagcttat 3420

tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa 3480

gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaatcc gattgacttt 3540

ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat 3600

agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa 3660

aaaggaaatg agctggctct gccaaagcaa tatgtgaatt ttttatattt agctagtcac 3720

tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag 3780

cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt 3840

ttagcagatg ccaatttaga taaagtctct agtgcataa 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 taggaggtag ctga 4104

<210> 2

<211> 4212

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<400> 2

atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc 60

ataaccgatg aatacaaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt 120

cattcgatta aaaagaatct tatcggtgcc ctctatttcg atagtggcga aacggcagag 180

gcgactcgcc tgaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt 240

tacttacaag aaatttttag caatgagatg gccaaagtgt acgattcttt ctttcaccgt 300

ttggaagagt ctttcttgt cgaagaggac aagaaacatg aacggcacc catcttttga 360

aacatagtag atgaggtggc atatcatgaa aagtaccaa cgatttatca cctcagaaaa 420

aagctagttg actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat 480

atgataaagt tccgtgggca ctttctcatt gaggtgatc 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 atcagtafc ggacttattt ttggctgccca aaaaccttag cgatgcaatc	900
ctctatctg acatactgag agttaatact gagattacca aggcgcctgt atccgcttca	960
atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc ctagtcctgt	1020
cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca	1080
ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta	1140
gagaagatgg atgggacgga agagttgctt gtaaaactca atcggaaga tctactgca	1200
aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat	1260
gctatactta gaaggcagga ggatttttat ccgttctca aagacaatcg tgaaaagatt	1320
gagaaaatcc taacctttcg cataccttac tatgtgggac ccctggcccg agggaactct	1380
cggttcgcat ggatgacaag aaagtcgaa gaaacgatta ctccatggaa ttttgaggaa	1440
gttgtcgata aaggcgcgc agctcaatcg ttcacgaga ggatgaccaa ctttgacaag	1500
aatttaccga acgaaaaagt attgcctaag cacagtttac ttacgagta tttcacagt	1560
tacaatgaac tcacgaaagt taagtatgic actgagggca tgcgtaaacc cgcctttcta	1620
agcggagaac agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca	1680
gttaagcaat tgaagagga ctactttaag aaaattgaat gcttcgattc tgcgagatc	1740
tccgggtag aagatcgatt taatgcgtca ctgggtacgt atcatgacct cctaaagata	1800
attaaagata aggacttctt ggataacgaa gagaatgaag atatcttaga agatatagt	1860
ttgactctta cctctttga agatcgggaa atgattgagg aaagactaaa aacatcgc	1920
cacctgttcg acgataaggt tatgaacag ttaaagaggc gtcgctatac gggctgggga	1980
cgattgtcgc ggaacttat caacgggata agagacaagc aaagtggtaa aactattctc	2040
gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac	2100
tctttaacct tcaaagagga tatacaaaag gcacaggttt ccggacaagg ggactcattg	2160
cacgaacata ttgcgaatc tctgtgttcg ccagccatca aaaaggcat actccagaca	2220
gtcaaagtag tggatgagct agttaaggic atgggacgtc acaaacgga aaacattgta	2280
atcgagatgg cagcgaaaa tcaaacgact cagaaggggc aaaaaacag tcgagagcgg	2340
atgaagagaa tagaagaggg tattaagaa ctgggcagcc agatcttaaa ggagcatcct	2400
gtggaaaata ccaattgca 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 aaagaaagt t cgataactta 2700
actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccggtttat taaacgtcag 2760
ctcgtggaaa cccgccaat cacaagcat gttgcacaga tactagattc ccgaatgaat 2820
acgaaatagc acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca 2880
aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940
taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000
taccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060

atgatcgga aaagcgaaca ggagataggc aaggctacag ccaaatactt cttttattct 3120
aacattatga atttctttaa gacggaaatc actctggcaa acggagagat acgcaaacga 3180
cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc 3240
gcgacggtga gaaaagtttt gtccatgcc caagtcaaca tagtaagaa aactgaggtg 3300
cagaccggag ggttttcaaa ggaatcgatt ctccaaaaa ggaatagtga taagctcatc 3360
gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgtagccc tacagttgcc 3420
tattctgtcc tagtagtggc aaaagttgag aagggaata ccaagaaact gaagtcagtc 3480

aaagaattat tggggataac gattatggag cgctcgtctt ttgaaaagaa ccccatcgac 3540
ttccttgagg cgaaaggta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600
tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660
caaaaggga acgaactgc actaccgtct aaatacgtga atttctgta tttagcgtcc 3720
cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact tttgttgag 3780
cagcacaac attatctcga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840
atcctagctg atgccaatct ggacaaagta ttaagcgcac acaacaagca cagggataaa 3900

cccatacgtg agcaggcgga aaatattatc catttgitta ctcttacaa ctcggcgct 3960
ccagccgat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020
gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcgata 4080
gattgtcac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140
tacaagacc atgacggta ttataaagat catgacatcg attacaagga tgacgatgac 4200
aaggctgcag ga 4212

<210> 3

<211> 4107

<212> DNA

<213> Streptococcus pyogenes

<400> 3

```

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg      60
atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc      120
cacagtatca aaaaaaatct tataggggct cttttatttg acagtgagaa gacagcggaa      180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt      240
tatctacagg agatTTTTTc aaatgagatg gcgaaagtag atgatagttt ctttcatcga      300
cttgaagagt cttttttggt ggaagaagac aagaagcatg aacgtcatcc ttttttga      360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa      420

aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcatt      480
atgattaagt ttctgtgtca ttttttgatt gagggagatt taaatcctga taatagtgat      540
gtggacaaac tatttatcca gttgttacia acctacaatc aattatttga agaaaacctt      600
attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga      660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttggaat      720
ctcattgctt tgcattggg tttgacctt aattttaaat caaattttga tttggcagaa      780
gatgctaaat tacagcttcc aaaagatact tacgatgatg atttagataa tttattggcg      840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaattttatc agatgctatt      900
ttactttcag atatcctaag agtaataact gaaataacta aggctccctc atcagcttca      960
atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagtctga      1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca      1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta      1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc      1200
aagcaacgga ctttgacaa cggtctctatt ccccatcaaa ttcacttggg tgagctgcat      1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt      1320
gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt      1380
cgttttgcat ggatgactcg gaagctgaa gaaacaatta ccccatggaa ttttgaagaa      1440
gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa      1500
aatcttccaa atgaaaaagt actacaaaaa catagtgtgc tttatgagta ttttacggtt      1560
tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt      1620

```


tcaggatgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc	1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt	1740
tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt	1800
attaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt	1860
ttaacattga ccttatttga agataggagg atgattgagg aaagacttaa aacatatgct	1920
cacctctttg atgataaggt gatgaacag cttaaacgtc gccgtttatac tggttgggga	1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta	2040
gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat	2100
agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttaa	2160
catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat ttacagact	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 aaagtttctt taaagacgat tcaatagaca ataaggtctt aacgcgttct	2580
gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa	2640
aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta	2700
acgaaagctg aacgtggagg tttagtgtaa cttgataaag ctggttttat caaacgcca	2760
ttggttgaaa ctgccaat cactaagcat gtggcacaaa ttttgatag tcgcatgaat	2820
actaaatagc atgaaaatga taaacttatt cgagaggtta aagtattac 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 agaaatagc aaagcaaccg caaaatattt cttttactct	3120
aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc	3180
cctctaatac aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt	3240
gccacagtgc gcaaagtatt gtccatgcc caagtcaata ttgtcaagaa aacagaagta	3300
cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt	3360

gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420
tattcagtcc tagtgggtgc taagggtgaa aaagggaat 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 attattttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840
attttagcag atgccaatit agataaagtt cttagtgcac ataacaaaca tagagacaaa 3900
ccaatacgtg aacaagcaga aaatattatt cattatttta cgttgacgaa tcttggagct 3960
cccgtgcttt ttaaataatt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
gaagtttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080
gatttgagtc agctaggagg tgactga 4107

<210> 4

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 4

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> 5
 <211> 1368
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic polypeptide
 <400> 5
 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> 6

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 6

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240
 Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255
 Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270

 Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285
 Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300
 Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320
 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335

 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380
 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400

 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460

 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725	730	735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly		
740	745	750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln		
755	760	765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile		
770	775	780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro		
785	790	795
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
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
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg		
965	970	975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val			
980	985	990	
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe			
995	1000	1005	
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala			
1010	1015	1020	
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe			
1025	1030	1035	
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala			
1040	1045	1050	
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu			
1055	1060	1065	
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val			
1070	1075	1080	
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr			
1085	1090	1095	
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys			
1100	1105	1110	
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro			
1115	1120	1125	
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val			
1130	1135	1140	
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys			
1145	1150	1155	
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser			
1160	1165	1170	
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys			
1175	1180	1185	
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu			
1190	1195	1200	
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly			

1205	1210	1215
Glu Leu Gln Lys Gly Asn Glu	Leu Ala Leu Pro Ser	Lys Tyr Val
1220	1225	1230
Asn Phe Leu Tyr Leu Ala Ser	His Tyr Glu Lys Leu	Lys Gly Ser
1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys	Gln Leu Phe Val Glu	Gln His Lys
1250	1255	1260
His Tyr Leu Asp Glu Ile Ile	Glu Gln Ile Ser Glu	Phe Ser Lys
1265	1270	1275
Arg Val Ile Leu Ala Asp Ala	Asn Leu Asp Lys Val	Leu Ser Ala
1280	1285	1290
Tyr Asn Lys His Arg Asp Lys	Pro Ile Arg Glu Gln	Ala Glu Asn
1295	1300	1305
Ile Ile His Leu Phe Thr Leu	Thr Asn Leu Gly Ala	Pro Ala Ala
1310	1315	1320
Phe Lys Tyr Phe Asp Thr Thr	Ile Asp Arg Lys Arg	Tyr Thr Ser
1325	1330	1335
Thr Lys Glu Val Leu Asp Ala	Thr Leu Ile His Gln	Ser Ile Thr
1340	1345	1350
Gly Leu Tyr Glu Thr Arg Ile	Asp Leu Ser Gln Leu	Gly Gly Asp
1355	1360	1365
<210> 7		
<211> 1368		
<212> PRT		
<213> Artificial Sequence		
<220><223> Synthetic polypeptide		
<400> 7		
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val		
1 5 10 15		
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe		

20

25

30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45
 Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 50 55 60
 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
 65 70 75 80
 Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
 85 90 95
 Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110
 His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125
 His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140
 Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160
 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175
 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 180 185 190
 Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205
 Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220
 Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240
 Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255
 Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270
 Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285
 Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300
 Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320
 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335
 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380
 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780
 Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

 785 790 795 800
 Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
 805 810 815
 Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
 820 825 830
 Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
 835 840 845
 Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

 850 855 860
 Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 865 870 875 880
 Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 885 890 895
 Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

 915 920 925
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960
 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

 980 985 990
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020

Lys Ser	Glu Gln Glu Ile Gly	Lys Ala Thr Ala Lys	Tyr Phe Phe
1025	1030	1035	
Tyr Ser	Asn Ile Met Asn Phe	Phe Lys Thr Glu Ile	Thr Leu Ala
1040	1045	1050	
Asn Gly	Glu Ile Arg Lys Arg	Pro Leu Ile Glu Thr	Asn Gly Glu
1055	1060	1065	
Thr Gly	Glu Ile Val Trp Asp	Lys Gly Arg Asp Phe	Ala Thr Val
1070	1075	1080	
Arg Lys	Val Leu Ser Met Pro	Gln Val Asn Ile Val	Lys Lys Thr
1085	1090	1095	
Glu Val	Gln Thr Gly Gly Phe	Ser Lys Glu Ser Ile	Leu Pro Lys
1100	1105	1110	
Arg Asn	Ser Asp Lys Leu Ile	Ala Arg Lys Lys Asp	Trp Asp Pro
1115	1120	1125	
Lys Lys	Tyr Gly Gly Phe Asp	Ser Pro Thr Val Ala	Tyr Ser Val
1130	1135	1140	
Leu Val	Val Ala Lys Val Glu	Lys Gly Lys Ser Lys	Lys Leu Lys
1145	1150	1155	
Ser Val	Lys Glu Leu Leu Gly	Ile Thr Ile Met Glu	Arg Ser Ser
1160	1165	1170	
Phe Glu	Lys Asn Pro Ile Asp	Phe Leu Glu Ala Lys	Gly Tyr Lys
1175	1180	1185	
Glu Val	Lys Lys Asp Leu Ile	Ile Lys Leu Pro Lys	Tyr Ser Leu
1190	1195	1200	
Phe Glu	Leu Glu Asn Gly Arg	Lys Arg Met Leu Ala	Ser Ala Gly
1205	1210	1215	
Glu Leu	Gln Lys Gly Asn Glu	Leu Ala Leu Pro Ser	Lys Tyr Val
1220	1225	1230	
Asn Phe	Leu Tyr Leu Ala Ser	His Tyr Glu Lys Leu	Lys Gly Ser
1235	1240	1245	
Pro Glu	Asp Asn Glu Gln Lys	Gln Leu Phe Val Glu	Gln His Lys

1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 8

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 8

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> 9

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 9

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> 10

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 10

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> 11
<211> 1367
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic polypeptide
<400> 11
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> 12

<211> 1052

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 12

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> 13

<211> 1052

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 13

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> 14
 <211> 1052
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic polypeptide
 <400> 14
 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
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
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
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> 15

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 15

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> 16

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 16

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> 17

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 17

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
 1 5 10 15
 Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
 20 25 30
 Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
 35 40 45
 Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
 50 55 60
 Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
 65 70 75 80
 Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
 85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350
 Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
 355 360 365

 Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
 370 375 380
 Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
 385 390 395 400
 Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
 405 410 415
 Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
 420 425 430

 Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
 435 440 445
 Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
 450 455 460
 Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
 465 470 475 480
 Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
 485 490 495

 Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
 500 505 510
 Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
 515 520 525
 Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
 530 535 540
 Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
 545 550 555 560

 Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
 565 570 575
 Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
 580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
 595 600 605
 Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
 610 615 620

 Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
 625 630 635 640
 Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
 645 650 655
 Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
 660 665 670
 Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
 675 680 685

 Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
 690 695 700
 Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
 705 710 715 720
 Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
 725 730 735
 Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
 740 745 750

 His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
 755 760 765
 Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
 770 775 780
 Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
 785 790 795 800
 Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
 805 810 815

 Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
 820 825 830
 Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp

835	840	845	
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly			
850	855	860	
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn			
865	870	875	880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe			
885	890	895	
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys			
900	905	910	
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys			
915	920	925	
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu			
930	935	940	
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys			
945	950	955	960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu			
965	970	975	
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val			
980	985	990	
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val			
995	1000	1005	
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys			
1010	1015	1020	
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr			
1025	1030	1035	
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn			
1040	1045	1050	
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr			
1055	1060	1065	
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg			
1070	1075	1080	

Lys Val	Leu Ser Met Pro Gln	Val Asn Ile Val Lys	Lys Thr Glu
1085	1090	1095	
Val Gln	Thr Gly Gly Phe Ser	Lys Glu Ser Ile Leu	Pro Lys Arg
1100	1105	1110	
Asn Ser	Asp Lys Leu Ile Ala	Arg Lys Lys Asp Trp	Asp Pro Lys
1115	1120	1125	
Lys Tyr	Gly Gly Phe Glu Ser	Pro Thr Val Ala Tyr	Ser Val Leu
1130	1135	1140	
Val Val	Ala Lys Val Glu Lys	Gly Lys Ser Lys Lys	Leu Lys Ser
1145	1150	1155	
Val Lys	Glu Leu Leu Gly Ile	Thr Ile Met Glu Arg	Ser Ser Phe
1160	1165	1170	
Glu Lys	Asn Pro Ile Asp Phe	Leu Glu Ala Lys Gly	Tyr Lys Glu
1175	1180	1185	
Val Lys	Lys Asp Leu Ile Ile	Lys Leu Pro Lys Tyr	Ser Leu Phe
1190	1195	1200	
Glu Leu	Glu Asn Gly Arg Lys	Arg Met Leu Ala Ser	Ala Gly Glu
1205	1210	1215	
Leu Gln	Lys Gly Asn Glu Leu	Ala Leu Pro Ser Lys	Tyr Val Asn
1220	1225	1230	
Phe Leu	Tyr Leu Ala Ser His	Tyr Glu Lys Leu Lys	Gly Ser Pro
1235	1240	1245	
Glu Asp	Asn Glu Gln Lys Gln	Leu Phe Val Glu Gln	His Lys His
1250	1255	1260	
Tyr Leu	Asp Glu Ile Ile Glu	Gln Ile Ser Glu Phe	Ser Lys Arg
1265	1270	1275	
Val Ile	Leu Ala Asp Ala Asn	Leu Asp Lys Val Leu	Ser Ala Tyr
1280	1285	1290	
Asn Lys	His Arg Asp Lys Pro	Ile Arg Glu Gln Ala	Glu Asn Ile
1295	1300	1305	
Ile His	Leu Phe Thr Leu Thr	Asn Leu Gly Ala Pro	Ala Ala Phe

1310 1315 1320
 Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser Thr
 1325 1330 1335
 Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
 1340 1345 1350
 Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
 1355 1360 1365

<210> 18

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 18

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
 1 5 10 15
 Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
 20 25 30
 Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
 35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
 50 55 60
 Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
 65 70 75 80
 Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
 85 90 95
 Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
 100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
 115 120 125
 Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
 130 135 140
 Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160
 Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
 165 170 175

 Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
 180 185 190
 Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
 195 200 205
 Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
 210 215 220
 Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
 225 230 235 240

 Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
 245 250 255
 Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
 260 265 270
 Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
 275 280 285
 Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
 290 295 300

 Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
 305 310 315 320
 Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
 325 330 335
 Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
 340 345 350
 Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
 355 360 365

 Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
 370 375 380
 Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
 385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655
 Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
 660 665 670
 Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
 675 680 685

 Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
 690 695 700
 Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
 705 710 715 720
 Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
 725 730 735
 Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
 740 745 750

 His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
 755 760 765
 Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
 770 775 780
 Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
 785 790 795 800
 Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
 805 810 815

 Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
 820 825 830
 Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
 835 840 845
 Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
 850 855 860
 Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
 865 870 875 880

 Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
 885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val Leu

1130	1135	1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser		
1145	1150	1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe		
1160	1165	1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu		
1175	1180	1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe		
1190	1195	1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu		
1205	1210	1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn		
1220	1225	1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro		
1235	1240	1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His		
1250	1255	1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg		
1265	1270	1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr		
1280	1285	1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile		
1295	1300	1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe		
1310	1315	1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser Thr		
1325	1330	1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly		
1340	1345	1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp		
1355	1360	1365

<210> 19

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 19

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met

145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460
 Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
 465 470 475 480
 Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
 485 490 495

 Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
 500 505 510
 Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
 515 520 525
 Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
 530 535 540
 Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
 545 550 555 560

 Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
 565 570 575
 Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
 580 585 590
 Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
 595 600 605
 Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
 610 615 620

 Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
 625 630 635 640
 Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
 645 650 655
 Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
 660 665 670
 Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
 675 680 685

 Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
 690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
 705 710 715 720
 Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
 725 730 735
 Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
 740 745 750

 His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
 755 760 765
 Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
 770 775 780
 Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
 785 790 795 800
 Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
 805 810 815

 Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
 820 825 830
 Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
 835 840 845
 Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
 850 855 860
 Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
 865 870 875 880

 Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
 885 890 895
 Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
 900 905 910
 Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
 915 920 925
 His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
 930 935 940

 Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945	950	955	960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu			
	965	970	975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val			
	980	985	990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val			
	995	1000	1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys			
	1010	1015	1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr			
	1025	1030	1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn			
	1040	1045	1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr			
	1055	1060	1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg			
	1070	1075	1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu			
	1085	1090	1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg			
	1100	1105	1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys			
	1115	1120	1125
Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val Leu			
	1130	1135	1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser			
	1145	1150	1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe			
	1160	1165	1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu			
	1175	1180	1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
 1190 1195 1200
 Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Arg Glu
 1205 1210 1215
 Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
 1220 1225 1230
 Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
 1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
 1250 1255 1260
 Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
 1265 1270 1275
 Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
 1280 1285 1290
 Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
 1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
 1310 1315 1320
 Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Glu Tyr Arg Ser Thr
 1325 1330 1335
 Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
 1340 1345 1350
 Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
 1355 1360 1365

<210> 20

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 20

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> 21

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 21

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> 22

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 22

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> 23
 <211> 1368
 <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 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>	24		
<211>	1372		
<212>	PRT		

<213> Streptococcus gallolyticus

<400> 24

```

Met Thr Lys Lys Asn Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser
1           5           10           15
Val Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys
           20           25           30
Met Lys Val Leu Gly Asn Thr Asp Lys Lys Tyr Ile Lys Lys Asn Leu
           35           40           45
Leu Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg
           50           55           60

Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Leu
65           70           75           80
Arg Tyr Leu Gln Glu Ile Phe Ala Asn Glu Ile Ala Lys Val Asp Glu
           85           90           95
Ser Phe Phe Gln Arg Leu Asp Glu Ser Phe Leu Thr Asp Asp Asp Lys
           100          105          110
Thr Phe Asp Ser His Pro Ile Phe Gly Asn Lys Ala Glu Glu Asp Ala
           115          120          125

Tyr His Gln Lys Phe Pro Thr Ile Tyr His Leu Arg Lys His Leu Ala
           130          135          140
Asp Ser Ser Glu Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala
145          150          155          160
His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Glu Leu Asn
           165          170          175
Ala Glu Asn Thr Asp Val Gln Lys Ile Phe Ala Asp Phe Val Gly Val
           180          185          190

Tyr Asn Arg Thr Phe Asp Asp Ser His Leu Ser Glu Ile Thr Val Asp
           195          200          205
Val Ala Ser Ile Leu Thr Glu Lys Ile Ser Lys Ser Arg Arg Leu Glu
           210          215          220
Asn Leu Ile Lys Tyr Tyr Pro Thr Glu Lys Lys Asn Thr Leu Phe Gly

```

225 230 235 240
 Asn Leu Ile Ala Leu Ala Leu Gly Leu Gln Pro Asn Phe Lys Thr Asn
 245 250 255

 Phe Lys Leu Ser Glu Asp Ala Lys Leu Gln Phe Ser Lys Asp Thr Tyr
 260 265 270
 Glu Glu Asp Leu Glu Glu Leu Leu Gly Lys Ile Gly Asp Asp Tyr Ala
 275 280 285
 Asp Leu Phe Thr Ser Ala Lys Asn Leu Tyr Asp Ala Ile Leu Leu Ser
 290 295 300
 Gly Ile Leu Thr Val Asp Asp Asn Ser Thr Lys Ala Pro Leu Ser Ala
 305 310 315 320

 Ser Met Ile Lys Arg Tyr Val Glu His His Glu Asp Leu Glu Lys Leu
 325 330 335
 Lys Glu Phe Ile Lys Ala Asn Lys Ser Glu Leu Tyr His Asp Ile Phe
 340 345 350
 Lys Asp Lys Asn Lys Asn Gly Tyr Ala Gly Tyr Ile Glu Asn Gly Val
 355 360 365
 Lys Gln Asp Glu Phe Tyr Lys Tyr Leu Lys Asn Ile Leu Ser Lys Ile
 370 375 380

 Lys Ile Asp Gly Ser Asp Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp
 385 390 395 400
 Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln
 405 410 415
 Ile His Leu Gln Glu Met His Ala Ile Leu Arg Arg Gln Gly Asp Tyr
 420 425 430
 Tyr Pro Phe Leu Lys Glu Lys Gln Asp Arg Ile Glu Lys Ile Leu Thr
 435 440 445

 Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Val Arg Lys Asp Ser Arg
 450 455 460
 Phe Ala Trp Ala Glu Tyr Arg Ser Asp Glu Lys Ile Thr Pro Trp Asn
 465 470 475 480

Phe Asp Lys Val Ile Asp Lys Glu Lys Ser Ala Glu Lys Phe Ile Thr
 485 490 495
 Arg Met Thr Leu Asn Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro
 500 505 510

 Lys His Ser His Val Tyr Glu Thr Tyr Ala Val Tyr Asn Glu Leu Thr
 515 520 525
 Lys Ile Lys Tyr Val Asn Glu Gln Gly Lys Glu Ser Phe Phe Asp Ser
 530 535 540
 Asn Met Lys Gln Glu Ile Phe Asp His Val Phe Lys Glu Asn Arg Lys
 545 550 555 560
 Val Thr Lys Glu Lys Leu Leu Asn Tyr Leu Asn Lys Glu Phe Pro Glu
 565 570 575

 Tyr Arg Ile Lys Asp Leu Ile Gly Leu Asp Lys Glu Asn Lys Ser Phe
 580 585 590
 Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Lys Lys Ile Leu Asp Lys
 595 600 605
 Ala Phe Leu Asp Asp Lys Val Asn Glu Glu Val Ile Glu Asp Ile Ile
 610 615 620
 Lys Thr Leu Thr Leu Phe Glu Asp Lys Asp Met Ile His Glu Arg Leu
 625 630 635 640

 Gln Lys Tyr Ser Asp Ile Phe Thr Ala Asn Gln Leu Lys Lys Leu Glu
 645 650 655
 Arg Arg His Tyr Thr Gly Trp Gly Arg Leu Ser Tyr Lys Leu Ile Asn
 660 665 670
 Gly Ile Arg Asn Lys Glu Asn Asn Lys Thr Ile Leu Asp Tyr Leu Ile
 675 680 685
 Asp Asp Gly Ser Ala Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp
 690 695 700

 Thr Leu Pro Phe Lys Gln Ile Ile Gln Lys Ser Gln Val Val Gly Asp
 705 710 715 720
 Val Asp Asp Ile Glu Ala Val Val His Asp Leu Pro Gly Ser Pro Ala

725	730	735
Ile Lys Lys Gly Ile Leu Gln Ser Val Lys Ile Val Asp Glu Leu Val		
740	745	750
Lys Val Met Gly Gly Asn Pro Asp Asn Ile Val Ile Glu Met Ala Arg		
755	760	765
Glu Asn Gln Thr Thr Asn Arg Gly Arg Ser Gln Ser Gln Gln Arg Leu		
770	775	780
Lys Lys Leu Gln Asn Ser Leu Lys Glu Leu Gly Ser Asn Ile Leu Asn		
785	790	795
Glu Glu Lys Pro Ser Tyr Ile Glu Asp Lys Val Glu Asn Ser His Leu		
805	810	815
Gln Asn Asp Gln Leu Phe Leu Tyr Tyr Ile Gln Asn Gly Lys Asp Met		
820	825	830
Tyr Thr Gly Asp Glu Leu Asp Ile Asp His Leu Ser Asp Tyr Asp Ile		
835	840	845
Asp His Ile Ile Pro Gln Ala Phe Ile Lys Asp Asp Ser Ile Asp Asn		
850	855	860
Arg Val Leu Thr Ser Ser Ala Lys Asn Arg Gly Lys Ser Asp Asp Val		
865	870	875
Pro Ser Leu Asp Ile Val Arg Ala Arg Lys Ala Glu Trp Val Arg Leu		
885	890	895
Tyr Lys Ser Gly Leu Ile Ser Lys Arg Lys Phe Asp Asn Leu Thr Lys		
900	905	910
Ala Glu Arg Gly Gly Leu Thr Glu Ala Asp Lys Ala Gly Phe Ile Lys		
915	920	925
Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile		
930	935	940
Leu Asp Ala Arg Phe Asn Thr Glu His Asp Glu Asn Asp Lys Val Ile		
945	950	955
Arg Asp Val Lys Val Ile Thr Leu Lys Ser Asn Leu Val Ser Gln Phe		
965	970	975

Arg Lys Asp Phe Glu Phe Tyr Lys Val Arg Glu Ile Asn Asp Tyr His			
980	985	990	
His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Leu			
995	1000	1005	
Lys Lys Tyr Pro Lys Leu Ala Ser Glu Phe Val Tyr Gly Glu Tyr			
1010	1015	1020	
Lys Lys Tyr Asp Ile Arg Lys Phe Ile Thr Asn Ser Ser Asp Lys			
1025	1030	1035	
Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Leu Met Asn Phe Phe			
1040	1045	1050	
Lys Thr Lys Val Lys Tyr Ala Asp Gly Thr Val Phe Glu Arg Pro			
1055	1060	1065	
Ile Ile Glu Thr Asn Ala Asp Gly Glu Ile Ala Trp Asn Lys Gln			
1070	1075	1080	
Ile Asp Phe Glu Lys Val Arg Lys Val Leu Ser Tyr Pro Gln Val			
1085	1090	1095	
Asn Ile Val Lys Lys Val Glu Thr Gln Thr Gly Gly Phe Ser Lys			
1100	1105	1110	
Glu Ser Ile Leu Pro Lys Gly Asp Ser Asp Lys Leu Ile Pro Arg			
1115	1120	1125	
Lys Thr Lys Lys Val Tyr Trp Asp Thr Lys Lys Tyr Gly Gly Phe			
1130	1135	1140	
Asp Ser Pro Thr Val Ala Tyr Ser Val Phe Val Val Ala Asp Val			
1145	1150	1155	
Glu Lys Gly Lys Ala Lys Lys Leu Lys Thr Val Lys Glu Leu Val			
1160	1165	1170	
Gly Ile Ser Ile Met Glu Arg Ser Phe Phe Glu Glu Asn Pro Val			
1175	1180	1185	
Glu Phe Leu Glu Asn Lys Gly Tyr His Asn Ile Arg Glu Asp Lys			
1190	1195	1200	
Leu Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Phe Glu Gly Gly			

1205 1210 1215
 Arg Arg Arg Leu Leu Ala Ser Ala Ser Glu Leu Gln Lys Gly Asn
 1220 1225 1230
 Glu Met Val Leu Pro Gly Tyr Leu Val Glu Leu Leu Tyr His Ala
 1235 1240 1245
 His Arg Ala Asp Asn Phe Asn Ser Thr Glu Tyr Leu Asn Tyr Val
 1250 1255 1260

Ser Glu His Lys Lys Glu Phe Glu Lys Val Leu Ser Cys Val Glu
 1265 1270 1275
 Asp Phe Ala Asn Leu Tyr Val Asp Val Glu Lys Asn Leu Ser Lys
 1280 1285 1290
 Ile Arg Ala Val Ala Asp Ser Met Asp Asn Phe Ser Ile Glu Glu
 1295 1300 1305
 Ile Ser Asn Ser Phe Ile Asn Leu Leu Thr Leu Thr Ala Leu Gly
 1310 1315 1320

Ala Pro Ala Asp Phe Asn Phe Leu Gly Glu Lys Ile Pro Arg Lys
 1325 1330 1335
 Arg Tyr Thr Ser Thr Lys Glu Cys Leu Asn Ala Thr Leu Ile His
 1340 1345 1350
 Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Lys
 1355 1360 1365

Leu Gly Glu Glu
 1370

<210> 25

<211> 1370

<212> PRT

<213> Streptococcus gordonii

<400> 25

Met Lys Lys Gly Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15
 Gly Phe Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Met
 20 25 30

Lys Val Leu Gly Asn Thr Asp Lys Arg Phe Ile Lys Lys Asn Leu Ile
 35 40 45
 Gly Ala Leu Leu Phe Asp Glu Gly Thr Thr Ala Glu Ala Arg Arg Leu
 50 55 60

 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Leu Arg
 65 70 75 80
 Tyr Leu Gln Glu Ile Phe Ser Glu Glu Met Ser Lys Val Asp Ser Ser
 85 90 95
 Phe Phe His Arg Leu Asp Asp Ser Phe Leu Ile Pro Glu Asp Lys Arg
 100 105 110
 Glu Ser Lys Tyr Pro Ile Phe Ala Thr Leu Thr Glu Glu Lys Glu Tyr
 115 120 125

 His Lys Gln Phe Pro Thr Ile Tyr His Leu Arg Lys Gln Leu Ala Asp
 130 135 140
 Ser Lys Glu Lys Thr Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160
 Met Ile Lys Tyr Arg Gly His Phe Leu Tyr Glu Glu Ala Phe Asp Ile
 165 170 175
 Lys Asn Asn Asp Ile Gln Lys Ile Phe Asn Glu Phe Ile Ser Ile Tyr
 180 185 190

 Asp Asn Thr Phe Glu Gly Ser Ser Leu Ser Gly Gln Asn Ala Gln Val
 195 200 205
 Glu Ala Ile Phe Thr Asp Lys Ile Ser Lys Ser Ala Lys Arg Glu Arg
 210 215 220
 Val Leu Lys Leu Phe Pro Asp Glu Lys Ser Thr Gly Leu Phe Ser Glu
 225 230 235 240
 Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Lys Lys His Phe
 245 250 255

 Asp Leu Glu Asp Lys Ala Pro Leu Gln Phe Ser Lys Asp Thr Tyr Asp
 260 265 270
 Glu Asp Leu Glu Asn Leu Leu Gly Gln Ile Gly Asp Asp Phe Thr Asp

275 280 285
 Leu Phe Val Ser Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly
 290 295 300
 Ile Leu Thr Val Thr Asp Pro Ser Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320

 Met Ile Glu Arg Tyr Glu Asn His Gln Asn Asp Leu Ala Ala Leu Lys
 325 330 335
 Gln Phe Ile Lys Asn Asn Leu Pro Glu Lys Tyr Asp Glu Val Phe Ser
 340 345 350
 Asp Gln Ser Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Thr
 355 360 365
 Gln Glu Thr Phe Tyr Lys Tyr Ile Lys Asn Leu Leu Ser Lys Phe Glu
 370 375 380

 Gly Thr Asp Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp Phe Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415
 Gln Glu Met Asn Ala Ile Leu Arg Arg Gln Gly Glu Tyr Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Lys Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445

 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Arg Asp Phe Ala Trp
 450 455 460
 Leu Thr Arg Asn Ser Asp Glu Ala Ile Arg Pro Trp Asn Phe Glu Glu
 465 470 475 480
 Ile Val Asp Lys Ala Ser Ser Ala Glu Asp Phe Ile Asn Lys Met Thr
 485 490 495
 Asn Tyr Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser
 500 505 510

 Leu Leu Tyr Glu Thr Phe Ala Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Phe Ile Ala Glu Gly Leu Arg Asp Tyr Gln Phe Leu Asp Ser Gly Gln
 530 535 540
 Lys Lys Gln Ile Val Asn Gln Leu Phe Lys Glu Asn Arg Lys Val Thr
 545 550 555 560
 Glu Lys Asp Ile Ile His Tyr Leu His Asn Val Asp Gly Tyr Asp Gly
 565 570 575

 Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ala Ser Leu Ser Thr
 580 585 590
 Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Glu Phe Met Asp Asp
 595 600 605
 Ala Lys Asn Glu Ala Ile Leu Glu Asn Ile Val His Thr Leu Thr Ile
 610 615 620
 Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ala Gln Tyr Asp Ser
 625 630 635 640

 Leu Phe Asp Glu Lys Val Ile Lys Ala Leu Thr Arg Arg His Tyr Thr
 645 650 655
 Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Cys Asp Lys
 660 665 670
 Gln Thr Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Lys Ile
 675 680 685
 Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Gly Leu Ser Phe Lys
 690 695 700

 Glu Ile Ile Gln Lys Ala Gln Val Ile Gly Lys Thr Asp Asp Val Lys
 705 710 715 720
 Gln Val Val Gln Glu Leu Ser Gly Ser Pro Ala Ile Lys Lys Gly Ile
 725 730 735
 Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met Gly His
 740 745 750
 Ala Pro Glu Ser Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr
 755 760 765

 Ala Arg Gly Lys Lys Asn Ser Gln Gln Arg Tyr Lys Arg Ile Glu Asp

770	775	780	
Ser Leu Lys Ile Leu Ala	Ser Gly Leu Asp Ser Asn Ile Leu Lys Glu		
785	790	795	800
Asn Pro Thr Asp Asn Asn Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr			
	805	810	815
Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly Glu Ala Leu Asp Ile			
	820	825	830
Asn Gln Leu Ser Ser Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe			
	835	840	845
Ile Lys Asp Asp Ser Leu Asp Asn Arg Val Leu Thr Ser Ser Lys Asp			
	850	855	860
Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Ile Glu Val Val Gln Lys			
865	870	875	880
Arg Lys Ala Phe Trp Gln Gln Leu Leu Asp Ser Lys Leu Ile Ser Glu			
	885	890	895
Arg Lys Phe Asn Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Asp Glu			
	900	905	910
Arg Asp Lys Val Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln			
	915	920	925
Ile Thr Lys His Val Ala Gln Ile Leu Asp Ala Arg Tyr Asn Thr Glu			
	930	935	940
Val Asn Glu Lys Asp Lys Lys Asn Arg Thr Val Lys Ile Ile Thr Leu			
945	950	955	960
Lys Ser Asn Leu Val Ser Asn Phe Arg Lys Glu Phe Arg Leu Tyr Lys			
	965	970	975
Val Arg Glu Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn			
	980	985	990
Ala Val Val Ala Lys Ala Ile Leu Lys Lys Tyr Pro Lys Leu Glu Pro			
	995	1000	1005
Glu Phe Val Tyr Gly Glu Tyr Gln Lys Tyr Asp Leu Lys Arg Tyr			
1010	1015	1020	

Ile Ser Arg Ser Lys Asp Pro	Lys Glu Val Glu Lys	Ala Thr Glu
1025	1030	1035
Lys Tyr Phe Phe Tyr Ser Asn	Leu Leu Asn Phe Phe	Lys Glu Glu
1040	1045	1050
Val His Tyr Ala Asp Gly Thr	Ile Val Lys Arg Glu	Asn Ile Glu
1055	1060	1065
Tyr Ser Lys Asp Thr Gly Glu	Ile Ala Trp Asn Lys	Glu Lys Asp
1070	1075	1080
Phe Ala Ile Ile Lys Lys Val	Leu Ser Leu Pro Gln	Val Asn Ile
1085	1090	1095
Val Lys Lys Arg Glu Val Gln	Thr Gly Gly Phe Ser	Lys Glu Ser
1100	1105	1110
Ile Leu Pro Lys Gly Asn Ser	Asp Lys Leu Ile Pro	Arg Lys Thr
1115	1120	1125
Lys Asp Ile Leu Leu Asp Thr	Thr Lys Tyr Gly Gly	Phe Asp Ser
1130	1135	1140
Pro Val Ile Ala Tyr Ser Ile	Leu Leu Ile Ala Asp	Ile Glu Lys
1145	1150	1155
Gly Lys Ala Lys Lys Leu Lys	Thr Val Lys Thr Leu	Val Gly Ile
1160	1165	1170
Thr Ile Met Glu Lys Ala Ala	Phe Glu Glu Asn Pro	Ile Thr Phe
1175	1180	1185
Leu Glu Asn Lys Gly Tyr His	Asn Val Arg Lys Glu	Asn Ile Leu
1190	1195	1200
Cys Leu Pro Lys Tyr Ser Leu	Phe Glu Leu Glu Asn	Gly Arg Arg
1205	1210	1215
Arg Leu Leu Ala Ser Ala Lys	Glu Leu Gln Lys Gly	Asn Glu Ile
1220	1225	1230
Val Leu Pro Val Tyr Leu Thr	Thr Leu Leu Tyr His	Ser Lys Asn
1235	1240	1245
Val His Lys Leu Asp Glu Pro	Gly His Leu Glu Tyr	Ile Gln Lys

1250 1255 1260
 His Arg Asn Glu Phe Lys Asp Leu Leu Asn Leu Val Ser Glu Phe
 1265 1270 1275
 Ser Gln Lys Tyr Val Leu Ala Asp Ala Asn Leu Glu Lys Ile Lys
 1280 1285 1290
 Ser Leu Tyr Ala Asp Asn Glu Gln Ala Asp Ile Glu Ile Leu Ala
 1295 1300 1305
 Asn Ser Phe Ile Asn Leu Leu Thr Phe Thr Ala Leu Gly Ala Pro
 1310 1315 1320

 Ala Ala Phe Lys Phe Phe Gly Lys Asp Ile Asp Arg Lys Arg Tyr
 1325 1330 1335
 Thr Thr Val Ser Glu Ile Leu Asn Ala Thr Leu Ile His Gln Ser
 1340 1345 1350
 Ile Thr Gly Leu Tyr Glu Thr Trp Ile Asp Leu Ser Lys Leu Gly
 1355 1360 1365
 Glu Asp
 1370
 <210> 26
 <211> 1056
 <212> PRT
 <213> Staphylococcus aureus
 <400> 26
 Gly Ser His Met Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile

 1 5 10 15
 Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile
 20 25 30
 Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu
 35 40 45
 Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg
 50 55 60
 His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu

65					70				75				80			
Thr	Asp	His	Ser	Glu	Leu	Ser	Gly	Ile	Asn	Pro	Tyr	Glu	Ala	Arg	Val	
				85				90				95				
Lys	Gly	Leu	Ser	Gln	Lys	Leu	Ser	Glu	Glu	Glu	Phe	Ser	Ala	Ala	Leu	
				100				105				110				
Leu	His	Leu	Ala	Lys	Arg	Arg	Gly	Val	His	Asn	Val	Asn	Glu	Val	Glu	
				115				120				125				
Glu	Asp	Thr	Gly	Asn	Glu	Leu	Ser	Thr	Lys	Glu	Gln	Ile	Ser	Arg	Asn	
				130				135				140				
Ser	Lys	Ala	Leu	Glu	Glu	Lys	Tyr	Val	Ala	Glu	Leu	Gln	Leu	Glu	Arg	
145				150				155				160				
Leu	Lys	Lys	Asp	Gly	Glu	Val	Arg	Gly	Ser	Ile	Asn	Arg	Phe	Lys	Thr	
				165				170				175				
Ser	Asp	Tyr	Val	Lys	Glu	Ala	Lys	Gln	Leu	Leu	Lys	Val	Gln	Lys	Ala	
				180				185				190				
Tyr	His	Gln	Leu	Asp	Gln	Ser	Phe	Ile	Asp	Thr	Tyr	Ile	Asp	Leu	Leu	
				195				200				205				
Glu	Thr	Arg	Arg	Thr	Tyr	Tyr	Glu	Gly	Pro	Gly	Glu	Gly	Ser	Pro	Phe	
210				215				220								
Gly	Trp	Lys	Asp	Ile	Lys	Glu	Trp	Tyr	Glu	Met	Leu	Met	Gly	His	Cys	
225				230				235				240				
Thr	Tyr	Phe	Pro	Glu	Glu	Leu	Arg	Ser	Val	Lys	Tyr	Ala	Tyr	Asn	Ala	
				245				250				255				
Asp	Leu	Tyr	Asn	Ala	Leu	Asn	Asp	Leu	Asn	Asn	Leu	Val	Ile	Thr	Arg	
				260				265				270				
Asp	Glu	Asn	Glu	Lys	Leu	Glu	Tyr	Tyr	Glu	Lys	Phe	Gln	Ile	Ile	Glu	
				275				280				285				
Asn	Val	Phe	Lys	Gln	Lys	Lys	Lys	Pro	Thr	Leu	Lys	Gln	Ile	Ala	Lys	
290				295				300								
Glu	Ile	Leu	Val	Asn	Glu	Glu	Asp	Ile	Lys	Gly	Tyr	Arg	Val	Thr	Ser	
305				310				315				320				

Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys

325 330 335

Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp

340 345 350

Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln

355 360 365

Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu

370 375 380

Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu

385 390 395 400

Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn

405 410 415

Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp

420 425 430

Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile

435 440 445

Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile

450 455 460

Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu

465 470 475 480

Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu

485 490 495

Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile

500 505 510

Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys

515 520 525

Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile

530 535 540

Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His

545 550 555 560

Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val

565 570 575
 Leu Val Lys Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe

 580 585 590
 Gln Tyr Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys
 595 600 605
 Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr
 610 615 620
 Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val
 625 630 635 640
 Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr

 645 650 655
 Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu
 660 665 670
 Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg
 675 680 685
 Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His
 690 695 700
 Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu

 705 710 715 720
 Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe
 725 730 735
 Glu Glu Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu
 740 745 750
 Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp
 755 760 765
 Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg

 770 775 780
 Glu Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly
 785 790 795 800
 Asn Thr Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn
 805 810 815

Asp Lys Leu Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met
 820 825 830
 Tyr His His Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu
 835 840 845
 Gln Tyr Gly Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr
 850 855 860
 Gly Asn Tyr Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile
 865 870 875 880
 Lys Lys Ile Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile
 885 890 895
 Thr Asp Asp Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu
 900 905 910
 Lys Pro Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe
 915 920 925
 Val Thr Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu
 930 935 940
 Val Asn Ser Lys Ala Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser
 945 950 955 960
 Asn Gln Ala Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys
 965 970 975
 Ile Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu
 980 985 990
 Asn Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu
 995 1000 1005
 Glu Asn Met Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile
 1010 1015 1020
 Ala Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu
 1025 1030 1035
 Gly Asn Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile
 1040 1045 1050
 Lys Lys Gly

1055

<210> 27

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 27

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> 28

<211> 345

<212> PRT

<

213> Sulfolobus islandicus

<400> 28

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>	29																	
<211>	1210																	
<212>	PRT																	
<213>	uncultured Parcubacteria group bacterium																	
<400>	29																	
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 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> 30
 <211> 1300
 <212> PRT
 <213> Francisella novicida
 <400> 30
 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> 31

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 31

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> 32

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 32

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> 33

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 33

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
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
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
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> 34

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 34

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> 35

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 35

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> 36

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 36

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> 37

<211> 1300

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 37

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> 38

<211> 887

<212> PRT

<213> *Natronobacterium gregoryi*

<400> 38

Met Thr Val Ile Asp Leu Asp Ser Thr Thr Thr Ala Asp Glu Leu Thr

1 5 10 15

Ser Gly His Thr Tyr Asp Ile Ser Val Thr Leu Thr Gly Val Tyr Asp

20 25 30

Asn Thr Asp Glu Gln His Pro Arg Met Ser Leu Ala Phe Glu Gln Asp

35 40 45

Asn Gly Glu Arg Arg Tyr Ile Thr Leu Trp Lys Asn Thr Thr Pro Lys

50 55 60

Asp Val Phe Thr Tyr Asp Tyr Ala Thr Gly Ser Thr Tyr Ile Phe Thr

65 70 75 80

Asn Ile Asp Tyr Glu Val Lys Asp Gly Tyr Glu Asn Leu Thr Ala Thr

85 90 95

Tyr Gln Thr Thr Val Glu Asn Ala Thr Ala Gln Glu Val Gly Thr Thr
100 105 110

Asp Glu Asp Glu Thr Phe Ala Gly Gly Glu Pro Leu Asp His His Leu
115 120 125

Asp Asp Ala Leu Asn Glu Thr Pro Asp Asp Ala Glu Thr Glu Ser Asp
130 135 140

Ser Gly His Val Met Thr Ser Phe Ala Ser Arg Asp Gln Leu Pro Glu
145 150 155 160

Trp Thr Leu His Thr Tyr Thr Leu Thr Ala Thr Asp Gly Ala Lys Thr
165 170 175

Asp Thr Glu Tyr Ala Arg Arg Thr Leu Ala Tyr Thr Val Arg Gln Glu
180 185 190

Leu Tyr Thr Asp His Asp Ala Ala Pro Val Ala Thr Asp Gly Leu Met
195 200 205

Leu Leu Thr Pro Glu Pro Leu Gly Glu Thr Pro Leu Asp Leu Asp Cys
210 215 220

Gly Val Arg Val Glu Ala Asp Glu Thr Arg Thr Leu Asp Tyr Thr Thr
225 230 235 240

Ala Lys Asp Arg Leu Leu Ala Arg Glu Leu Val Glu Glu Gly Leu Lys
245 250 255

Arg Ser Leu Trp Asp Asp Tyr Leu Val Arg Gly Ile Asp Glu Val Leu
260 265 270

Ser Lys Glu Pro Val Leu Thr Cys Asp Glu Phe Asp Leu His Glu Arg
275 280 285

Tyr Asp Leu Ser Val Glu Val Gly His Ser Gly Arg Ala Tyr Leu His
290 295 300

Ile Asn Phe Arg His Arg Phe Val Pro Lys Leu Thr Leu Ala Asp Ile
305 310 315 320

Asp Asp Asp Asn Ile Tyr Pro Gly Leu Arg Val Lys Thr Thr Tyr Arg
325 330 335

Pro Arg Arg Gly His Ile Val Trp Gly Leu Arg Asp Glu Cys Ala Thr

340 345 350
 Asp Ser Leu Asn Thr Leu Gly Asn Gln Ser Val Val Ala Tyr His Arg
 355 360 365

 Asn Asn Gln Thr Pro Ile Asn Thr Asp Leu Leu Asp Ala Ile Glu Ala
 370 375 380
 Ala Asp Arg Arg Val Val Glu Thr Arg Arg Gln Gly His Gly Asp Asp
 385 390 395 400
 Ala Val Ser Phe Pro Gln Glu Leu Leu Ala Val Glu Pro Asn Thr His
 405 410 415
 Gln Ile Lys Gln Phe Ala Ser Asp Gly Phe His Gln Gln Ala Arg Ser
 420 425 430

 Lys Thr Arg Leu Ser Ala Ser Arg Cys Ser Glu Lys Ala Gln Ala Phe
 435 440 445
 Ala Glu Arg Leu Asp Pro Val Arg Leu Asn Gly Ser Thr Val Glu Phe
 450 455 460
 Ser Ser Glu Phe Phe Thr Gly Asn Asn Glu Gln Gln Leu Arg Leu Leu
 465 470 475 480
 Tyr Glu Asn Gly Glu Ser Val Leu Thr Phe Arg Asp Gly Ala Arg Gly
 485 490 495

 Ala His Pro Asp Glu Thr Phe Ser Lys Gly Ile Val Asn Pro Pro Glu
 500 505 510
 Ser Phe Glu Val Ala Val Val Leu Pro Glu Gln Gln Ala Asp Thr Cys
 515 520 525
 Lys Ala Gln Trp Asp Thr Met Ala Asp Leu Leu Asn Gln Ala Gly Ala
 530 535 540
 Pro Pro Thr Arg Ser Glu Thr Val Gln Tyr Asp Ala Phe Ser Ser Pro
 545 550 555 560

 Glu Ser Ile Ser Leu Asn Val Ala Gly Ala Ile Asp Pro Ser Glu Val
 565 570 575
 Asp Ala Ala Phe Val Val Leu Pro Pro Asp Gln Glu Gly Phe Ala Asp
 580 585 590

Leu Ala Ser Pro Thr Glu Thr Tyr Asp Glu Leu Lys Lys Ala Leu Ala
 595 600 605
 Asn Met Gly Ile Tyr Ser Gln Met Ala Tyr Phe Asp Arg Phe Arg Asp
 610 615 620

 Ala Lys Ile Phe Tyr Thr Arg Asn Val Ala Leu Gly Leu Leu Ala Ala
 625 630 635 640
 Ala Gly Gly Val Ala Phe Thr Thr Glu His Ala Met Pro Gly Asp Ala
 645 650 655
 Asp Met Phe Ile Gly Ile Asp Val Ser Arg Ser Tyr Pro Glu Asp Gly
 660 665 670
 Ala Ser Gly Gln Ile Asn Ile Ala Ala Thr Ala Thr Ala Val Tyr Lys
 675 680 685

 Asp Gly Thr Ile Leu Gly His Ser Ser Thr Arg Pro Gln Leu Gly Glu
 690 695 700
 Lys Leu Gln Ser Thr Asp Val Arg Asp Ile Met Lys Asn Ala Ile Leu
 705 710 715 720
 Gly Tyr Gln Gln Val Thr Gly Glu Ser Pro Thr His Ile Val Ile His
 725 730 735
 Arg Asp Gly Phe Met Asn Glu Asp Leu Asp Pro Ala Thr Glu Phe Leu
 740 745 750

 Asn Glu Gln Gly Val Glu Tyr Asp Ile Val Glu Ile Arg Lys Gln Pro
 755 760 765
 Gln Thr Arg Leu Leu Ala Val Ser Asp Val Gln Tyr Asp Thr Pro Val
 770 775 780
 Lys Ser Ile Ala Ala Ile Asn Gln Asn Glu Pro Arg Ala Thr Val Ala
 785 790 795 800
 Thr Phe Gly Ala Pro Glu Tyr Leu Ala Thr Arg Asp Gly Gly Gly Leu
 805 810 815

 Pro Arg Pro Ile Gln Ile Glu Arg Val Ala Gly Glu Thr Asp Ile Glu
 820 825 830
 Thr Leu Thr Arg Gln Val Tyr Leu Leu Ser Gln Ser His Ile Gln Val

835 840 845
His Asn Ser Thr Ala Arg Leu Pro Ile Thr Thr Ala Tyr Ala Asp Gln
850 855 860
Ala Ser Thr His Ala Thr Lys Gly Tyr Leu Val Gln Thr Gly Ala Phe
865 870 875 880

Glu Ser Asn Val Gly Phe Leu

885

<210> 39

<211> 1129

<212> PRT

<213> Alicyclobacillus acidoterrestris

<400> 39

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> 40

<211> 1389

<212> PRT

<213> *Leptotrichia shahii*

<400> 40

Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys

1 5 10 15

Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp

20 25 30

Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys

35 40 45

Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn

50 55 60

Asp Asn Ile Leu Lys Glu Phe Thr Arg Lys Phe His Ala Gly Asn Ile

65 70 75 80

Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn

85 90 95

Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr

100 105 110

Gly Lys Ser Glu Lys Leu Lys Ala Leu Gly Ile Thr Lys Lys Lys Ile

115 120 125

Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile

130 135 140

Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg

145 150 155 160

Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg

165 170 175

Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile

180 185 190

Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile

195 200 205

Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His

210 215 220
 Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr
 225 230 235 240
 Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu

 245 250 255
 Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser
 260 265 270
 Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp
 275 280 285
 Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe
 290 295 300
 Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu

 305 310 315 320
 Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
 325 330 335
 Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
 340 345 350
 Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
 355 360 365
 Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile

 370 375 380
 Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
 385 390 395 400
 Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys
 405 410 415
 Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
 420 425 430
 Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys

 435 440 445
 Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
 450 455 460

Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
 465 470 475 480
 Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp
 485 490 495
 Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
 500 505 510
 Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu
 515 520 525
 Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile
 530 535 540
 Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys
 545 550 555 560
 Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp
 565 570 575
 Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile
 580 585 590
 Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg
 595 600 605
 Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile
 610 615 620
 Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu
 625 630 635 640
 Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp
 645 650 655
 Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser
 660 665 670
 Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro
 675 680 685
 Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn
 690 695 700
 Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu

705 710 715 720
 Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu
 725 730 735
 Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
 740 745 750
 Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys

 755 760 765
 Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
 770 775 780
 Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
 785 790 795 800
 Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
 805 810 815
 Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile

 820 825 830
 Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser
 835 840 845
 Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val
 850 855 860
 Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu
 865 870 875 880
 Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp

 885 890 895
 Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys
 900 905 910
 Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn
 915 920 925
 Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp
 930 935 940
 Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile

 945 950 955 960

Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu
 965 970 975
 Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys
 980 985 990
 Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser
 995 1000 1005
 Lys Ile Leu Cys Arg Ile Ile Phe Asn Ser Asp Phe Leu Lys Lys
 1010 1015 1020
 Tyr Lys Lys Glu Ile Asp Asn Leu Ile Glu Asp Met Glu Ser Glu
 1025 1030 1035
 Asn Glu Asn Lys Phe Gln Glu Ile Tyr Tyr Pro Lys Glu Arg Lys
 1040 1045 1050
 Asn Glu Leu Tyr Ile Tyr Lys Lys Asn Leu Phe Leu Asn Ile Gly
 1055 1060 1065
 Asn Pro Asn Phe Asp Lys Ile Tyr Gly Leu Ile Ser Asn Asp Ile
 1070 1075 1080
 Lys Met Ala Asp Ala Lys Phe Leu Phe Asn Ile Asp Gly Lys Asn
 1085 1090 1095
 Ile Arg Lys Asn Lys Ile Ser Glu Ile Asp Ala Ile Leu Lys Asn
 1100 1105 1110
 Leu Asn Asp Lys Leu Asn Gly Tyr Ser Lys Glu Tyr Lys Glu Lys
 1115 1120 1125
 Tyr Ile Lys Lys Leu Lys Glu Asn Asp Asp Phe Phe Ala Lys Asn
 1130 1135 1140
 Ile Gln Asn Lys Asn Tyr Lys Ser Phe Glu Lys Asp Tyr Asn Arg
 1145 1150 1155
 Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu Val Glu Phe Asn Tyr
 1160 1165 1170
 Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu
 1175 1180 1185
 Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val

1190	1195	1200
Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn		
1205	1210	1215
Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly		
1220	1225	1230
Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser		
1235	1240	1245
Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu		

1250	1255	1260
Ser Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg		
1265	1270	1275
Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp		
1280	1285	1290
Tyr Ser Ile Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser		
1295	1300	1305
Tyr Ser Thr Arg Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu		

1310	1315	1320
Val Phe Lys Lys Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys		
1325	1330	1335
Lys Phe Lys Leu Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met		
1340	1345	1350
Lys Pro Lys Lys Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser		
1355	1360	1365
Asp Tyr Ile Lys Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu		

1370	1375	1380
Asn Thr Asn Asp Thr Leu		

1385

<210> 41

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 41

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 42

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 42

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> 43

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 43

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg

1 5 10 15

Lys Val

<210> 44

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 44

Lys Arg Gly Ile Asn Asp Arg Asn Phe Trp Arg Gly Glu Asn Gly Arg

1 5 10 15

Lys Thr Arg

<210> 45

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 45

Lys Lys Thr Gly Gly Pro Ile Tyr Arg Arg Val Asp Gly Lys Trp Arg

1 5 10 15

Arg

<210> 46

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 46

Arg Arg Glu Leu Ile Leu Tyr Asp Lys Glu Glu Ile Arg Arg Ile Trp

1 5 10 15

Arg

<210> 47

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 47

Ala Val Ser Arg Lys Arg Lys Ala

1 5

<210> 48

<211> 313

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 48

Met Ile Gly Gln Lys Thr Leu Tyr Ser Phe Phe Ser Pro Ser Pro Ala
1 5 10 15
Arg Lys Arg His Ala Pro Ser Pro Glu Pro Ala Val Gln Gly Thr Gly
20 25 30
Val Ala Gly Val Pro Glu Glu Ser Gly Asp Ala Ala Ala Ile Pro Ala
35 40 45
Lys Lys Ala Pro Ala Gly Gln Glu Glu Pro Gly Thr Pro Pro Ser Ser
50 55 60
Pro Leu Ser Ala Glu Gln Leu Asp Arg Ile Gln Arg Asn Lys Ala Ala
65 70 75 80
Ala Leu Leu Arg Leu Ala Ala Arg Asn Val Pro Val Gly Phe Gly Glu
85 90 95
Ser Trp Lys Lys His Leu Ser Gly Glu Phe Gly Lys Pro Tyr Phe Ile
100 105 110
Lys Leu Met Gly Phe Val Ala Glu Glu Arg Lys His Tyr Thr Val Tyr
115 120 125
Pro Pro Pro His Gln Val Phe Thr Trp Thr Gln Met Cys Asp Ile Lys
130 135 140
Asp Val Lys Val Val Ile Leu Gly Gln Asp Pro Tyr His Gly Pro Asn
145 150 155 160
Gln Ala His Gly Leu Cys Phe Ser Val Gln Arg Pro Val Pro Pro Pro
165 170 175
Pro Ser Leu Glu Asn Ile Tyr Lys Glu Leu Ser Thr Asp Ile Glu Asp
180 185 190
Phe Val His Pro Gly His Gly Asp Leu Ser Gly Trp Ala Lys Gln Gly
195 200 205
Val Leu Leu Leu Asn Ala Val Leu Thr Val Arg Ala His Gln Ala Asn
210 215 220
Ser His Lys Glu Arg Gly Trp Glu Gln Phe Thr Asp Ala Val Val Ser
225 230 235 240
Trp Leu Asn Gln Asn Ser Asn Gly Leu Val Phe Leu Leu Trp Gly Ser

245 250 255
Tyr Ala Gln Lys Lys Gly Ser Ala Ile Asp Arg Lys Arg His His Val
260 265 270
Leu Gln Thr Ala His Pro Ser Pro Leu Ser Val Tyr Arg Gly Phe Phe

275 280 285
Gly Cys Arg His Phe Ser Lys Thr Asn Glu Leu Leu Gln Lys Ser Gly
290 295 300
Lys Lys Pro Ile Asp Trp Lys Glu Leu

305 310

<210> 49

<211> 209

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 49

Met Ala Gly Ala Gln Asp Phe Val Pro His Thr Ala Asp Leu Ala Glu

1 5 10 15

Leu Ala Ala Ala Ala Gly Glu Cys Arg Gly Cys Gly Leu Tyr Arg Asp

20 25 30

Ala Thr Gln Ala Val Phe Gly Ala Gly Gly Arg Ser Ala Arg Ile Met

35 40 45

Met Ile Gly Glu Gln Pro Gly Asp Lys Glu Asp Leu Ala Gly Leu Pro

50 55 60

Phe Val Gly Pro Ala Gly Arg Leu Leu Asp Arg Ala Leu Glu Ala Ala

65 70 75 80

Asp Ile Asp Arg Asp Ala Leu Tyr Val Thr Asn Ala Val Lys His Phe

85 90 95

Lys Phe Thr Arg Ala Ala Gly Gly Lys Arg Arg Ile His Lys Thr Pro

100 105 110

Ser Arg Thr Glu Val Val Ala Cys Arg Pro Trp Leu Ile Ala Glu Met

115 120 125

Thr Ser Val Glu Pro Asp Val Val Val Leu Leu Gly Ala Thr Ala Ala

130 135 140

Lys Ala Leu Leu Gly Asn Asp Phe Arg Val Thr Gln His Arg Gly Glu

145 150 155 160

Val Leu His Val Asp Asp Val Pro Gly Asp Pro Ala Leu Val Ala Thr

 165 170 175

Val His Pro Ser Ser Leu Leu Arg Gly Pro Lys Glu Glu Arg Glu Ser

 180 185 190

Ala Phe Ala Gly Leu Val Asp Asp Leu Arg Val Ala Ala Asp Val Arg

 195 200 205

Pro

<210> 50

<211> 209

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 50

Met Ala Gly Ala Gln Asp Phe Val Pro His Thr Ala Asp Leu Ala Glu

1 5 10 15

Leu Ala Ala Ala Ala Gly Glu Cys Arg Gly Cys Gly Leu Tyr Arg Asp

 20 25 30

Ala Thr Gln Ala Val Phe Gly Ala Gly Gly Arg Ser Ala Arg Ile Met

 35 40 45

Met Ile Gly Glu Gln Pro Gly Asp Lys Glu Asp Leu Ala Gly Leu Pro

50 55 60

Phe Val Gly Pro Ala Gly Arg Leu Leu Asp Arg Ala Leu Glu Ala Ala

65 70 75 80

Asp Ile Asp Arg Asp Ala Leu Tyr Val Thr Asn Ala Val Lys His Phe

 85 90 95

Lys Phe Thr Arg Ala Ala Gly Gly Lys Arg Ser Ile His Lys Thr Pro

 100 105 110

Ser Arg Thr Glu Val Val Ala Cys Arg Pro Trp Leu Ile Ala Glu Met
 115 120 125
 Thr Ser Val Glu Pro Asp Val Val Val Leu Leu Gly Ala Thr Ala Ala
 130 135 140
 Lys Ala Leu Leu Gly Asn Asp Phe Arg Val Thr Gln His Arg Gly Glu
 145 150 155 160
 Val Leu His Val Asp Asp Val Pro Gly Asp Pro Ala Leu Val Ala Thr
 165 170 175

Val His Pro Ser Ser Leu Leu Arg Gly Pro Lys Glu Glu Arg Glu Ser
 180 185 190
 Ala Phe Ala Gly Leu Val Asp Asp Leu Arg Val Ala Ala Asp Val Arg
 195 200 205

Pro

<210> 51

<211> 209

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 51

Met Ala Gly Ala Gln Asp Phe Val Pro His Thr Ala Asp Leu Ala Glu
 1 5 10 15

Leu Ala Ala Ala Ala Gly Glu Cys Arg Gly Cys Gly Leu Tyr Arg Asp
 20 25 30

Ala Thr Gln Ala Val Phe Gly Ala Gly Gly Arg Ser Ala Arg Ile Met
 35 40 45

Met Ile Gly Glu Gln Pro Gly Asp Lys Glu Asp Leu Ala Gly Leu Pro
 50 55 60

Phe Val Gly Pro Ala Gly Arg Leu Leu Asp Arg Ala Leu Glu Ala Ala
 65 70 75 80

Asp Ile Asp Arg Asp Ala Leu Tyr Val Thr Asn Ala Val Lys His Phe
 85 90 95

Lys Phe Thr Arg Ala Ala Gly Gly Lys Arg Arg Ile Ser Lys Thr Pro
 100 105 110
 Ser Arg Thr Glu Val Val Ala Cys Arg Pro Trp Leu Ile Ala Glu Met
 115 120 125
 Thr Ser Val Glu Pro Asp Val Val Val Leu Leu Gly Ala Thr Ala Ala
 130 135 140

Lys Ala Leu Leu Gly Asn Asp Phe Arg Val Thr Gln His Arg Gly Glu
 145 150 155 160
 Val Leu His Val Asp Asp Val Pro Gly Asp Pro Ala Leu Val Ala Thr
 165 170 175
 Val His Pro Ser Ser Leu Leu Arg Gly Pro Lys Glu Glu Arg Glu Ser
 180 185 190
 Ala Phe Ala Gly Leu Val Asp Asp Leu Arg Val Ala Ala Asp Val Arg
 195 200 205

Pro

<210> 52

<211> 211

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 52

Met Thr Thr Leu Thr Arg Gln Asp Leu Asn Phe Gly Gln Val Val Ala
 1 5 10 15
 Asp Val Leu Cys Glu Phe Leu Glu Val Ala Val His Leu Ile Leu Tyr
 20 25 30
 Val Arg Glu Val Tyr Pro Val Gly Ile Phe Gln Lys Arg Lys Lys Tyr
 35 40 45

Asn Val Pro Val Gln Met Ser Cys His Pro Glu Leu Asn Gln Tyr Ile
 50 55 60
 Gln Asp Thr Leu His Cys Val Lys Pro Leu Leu Glu Lys Asn Asp Val
 65 70 75 80

Glu Lys Val Val Val Val Ile Leu Asp Lys Glu His Arg Pro Val Glu
 85 90 95
 Lys Phe Val Phe Glu Ile Thr Gln Pro Pro Leu Leu Ser Ile Ser Ser
 100 105 110

 Asp Ser Leu Leu Ser His Val Glu Gln Leu Leu Arg Ala Phe Ile Leu
 115 120 125
 Lys Ile Ser Val Cys Asp Ala Val Leu Asp His Asn Pro Pro Gly Cys
 130 135 140
 Thr Phe Thr Val Leu Val His Thr Arg Glu Ala Ala Thr Arg Asn Met
 145 150 155 160
 Glu Lys Ile Gln Val Ile Lys Asp Phe Pro Trp Ile Leu Ala Asp Glu
 165 170 175

 Gln Asp Val His Met His Asp Pro Arg Leu Ile Pro Leu Lys Thr Met
 180 185 190
 Thr Ser Asp Ile Leu Lys Met Gln Leu Tyr Val Glu Glu Arg Ala His
 195 200 205
 Lys Gly Ser
 210
 <210> 53
 <211> 270
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic polypeptide
 <400> 53
 Met Pro Gln Ala Phe Leu Leu Gly Ser Ile His Glu Pro Ala Gly Ala
 1 5 10 15

 Leu Met Glu Pro Gln Pro Cys Pro Gly Ser Leu Ala Glu Ser Phe Leu
 20 25 30
 Glu Glu Glu Leu Arg Leu Asn Ala Glu Leu Ser Gln Leu Gln Phe Ser
 35 40 45
 Glu Pro Val Gly Ile Ile Tyr Asn Pro Val Glu Tyr Ala Trp Glu Pro
 50 55 60

His Arg Asn Tyr Val Thr Arg Tyr Cys Gln Gly Pro Lys Glu Val Leu
65 70 75 80

Phe Leu Gly Met Asn Pro Gly Pro Phe Gly Met Ala Gln Thr Gly Val
85 90 95

Pro Phe Gly Glu Val Ser Met Val Arg Asp Trp Leu Gly Ile Val Gly
100 105 110

Pro Val Leu Thr Pro Pro Gln Glu His Pro Lys Arg Pro Val Leu Gly
115 120 125

Leu Glu Cys Pro Gln Ser Glu Val Ser Gly Ala Arg Phe Trp Gly Phe
130 135 140

Phe Arg Asn Leu Cys Gly Gln Pro Glu Val Phe Phe His His Cys Phe
145 150 155 160

Val His Asn Leu Cys Pro Leu Leu Phe Leu Ala Pro Ser Gly Arg Asn
165 170 175

Leu Thr Pro Ala Glu Leu Pro Ala Lys Gln Arg Glu Gln Leu Leu Gly
180 185 190

Ile Cys Asp Ala Ala Leu Cys Arg Gln Val Gln Leu Leu Gly Val Arg
195 200 205

Leu Val Val Gly Val Gly Arg Leu Ala Glu Gln Arg Ala Arg Arg Ala
210 215 220

Leu Ala Gly Leu Met Pro Glu Val Gln Val Glu Gly Leu Leu His Pro
225 230 235 240

Ser Pro Arg Asn Pro Gln Ala Asn Lys Gly Trp Glu Ala Val Ala Lys
245 250 255

Glu Arg Leu Asn Glu Leu Gly Leu Leu Pro Leu Leu Leu Lys
260 265 270

<210> 54

<211> 335

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 54

Met Ser Lys Arg Lys Ala Pro Gln Glu Thr Leu Asn Gly Gly Ile Thr
1 5 10 15

Asp Met Leu Thr Glu Leu Ala Asn Phe Glu Lys Asn Val Ser Gln Ala
20 25 30

Ile His Lys Tyr Asn Ala Tyr Arg Lys Ala Ala Ser Val Ile Ala Lys
35 40 45

Tyr Pro His Lys Ile Lys Ser Gly Ala Glu Ala Lys Lys Leu Pro Gly
50 55 60

Val Gly Thr Lys Ile Ala Glu Lys Ile Asp Glu Phe Leu Ala Thr Gly
65 70 75 80

Lys Leu Arg Lys Leu Glu Lys Ile Arg Gln Asp Asp Thr Ser Ser Ser
85 90 95

Ile Asn Phe Leu Thr Arg Val Ser Gly Ile Gly Pro Ser Ala Ala Arg
100 105 110

Lys Phe Val Asp Glu Gly Ile Lys Thr Leu Glu Asp Leu Arg Lys Asn
115 120 125

Glu Asp Lys Leu Asn His His Gln Arg Ile Gly Leu Lys Tyr Phe Gly
130 135 140

Asp Phe Glu Lys Arg Ile Pro Arg Glu Glu Met Leu Gln Met Gln Asp
145 150 155 160

Ile Val Leu Asn Glu Val Lys Lys Val Asp Ser Glu Tyr Ile Ala Thr
165 170 175

Val Cys Gly Ser Phe Arg Arg Gly Ala Glu Ser Ser Gly Asp Met Asp
180 185 190

Val Leu Leu Thr His Pro Ser Phe Thr Ser Glu Ser Thr Lys Gln Pro
195 200 205

Lys Leu Leu His Gln Val Val Glu Gln Leu Gln Lys Val His Phe Ile
210 215 220

Thr Asp Thr Leu Ser Lys Gly Glu Thr Lys Phe Met Gly Val Cys Gln
225 230 235 240

Leu Pro Ser Lys Asn Asp Glu Lys Glu Tyr Pro His Arg Arg Ile Asp
 245 250 255
 Ile Arg Leu Ile Pro Lys Asp Gln Tyr Tyr Cys Gly Val Leu Tyr Phe
 260 265 270
 Thr Gly Ser Asp Ile Phe Asn Lys Asn Met Arg Ala His Ala Leu Glu
 275 280 285
 Lys Gly Phe Thr Ile Asn Glu Tyr Thr Ile Arg Pro Leu Gly Val Thr
 290 295 300

 Gly Val Ala Gly Glu Pro Leu Pro Val Asp Ser Glu Lys Asp Ile Phe
 305 310 315 320
 Asp Tyr Ile Gln Trp Lys Tyr Arg Glu Pro Lys Asp Arg Ser Glu
 325 330 335

 <210> 55
 <211> 575
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic polypeptide
 <400> 55
 Met Asp Pro Arg Gly Ile Leu Lys Ala Phe Pro Lys Arg Gln Lys Ile
 1 5 10 15
 His Ala Asp Ala Ser Ser Lys Val Leu Ala Lys Ile Pro Arg Arg Glu
 20 25 30
 Glu Gly Glu Glu Ala Glu Glu Trp Leu Ser Ser Leu Arg Ala His Val
 35 40 45
 Val Arg Thr Gly Ile Gly Arg Ala Arg Ala Glu Leu Phe Glu Lys Gln
 50 55 60
 Ile Val Gln His Gly Gly Gln Leu Cys Pro Ala Gln Gly Pro Gly Val
 65 70 75 80
 Thr His Ile Val Val Asp Glu Gly Met Asp Tyr Glu Arg Ala Leu Arg
 85 90 95
 Leu Leu Arg Leu Pro Gln Leu Pro Pro Gly Ala Gln Leu Val Lys Ser
 100 105 110

Ala Trp Leu Ser Leu Cys Leu Gln Glu Arg Arg Leu Val Asp Val Ala
115 120 125
Gly Phe Ser Ile Phe Ile Pro Ser Arg Tyr Leu Asp His Pro Gln Pro
130 135 140
Ser Lys Ala Glu Gln Asp Ala Ser Ile Pro Pro Gly Thr His Glu Ala
145 150 155 160
Leu Leu Gln Thr Ala Leu Ser Pro Pro Pro Pro Thr Arg Pro Val
165 170 175
Ser Pro Pro Gln Lys Ala Lys Glu Ala Pro Asn Thr Gln Ala Gln Pro
180 185 190
Ile Ser Asp Asp Glu Ala Ser Asp Gly Glu Glu Thr Gln Val Ser Ala
195 200 205
Ala Asp Leu Glu Ala Leu Ile Ser Gly His Tyr Pro Thr Ser Leu Glu
210 215 220
Gly Asp Cys Glu Pro Ser Pro Ala Pro Ala Val Leu Asp Lys Trp Val
225 230 235 240
Cys Ala Gln Pro Ser Ser Gln Lys Ala Thr Asn His Asn Leu His Ile
245 250 255
Thr Glu Lys Leu Glu Val Leu Ala Lys Ala Tyr Ser Val Gln Gly Asp
260 265 270
Lys Trp Arg Ala Leu Gly Tyr Ala Lys Ala Ile Asn Ala Leu Lys Ser
275 280 285
Phe His Lys Pro Val Thr Ser Tyr Gln Glu Ala Cys Ser Ile Pro Gly
290 295 300
Ile Gly Lys Arg Met Ala Glu Lys Ile Ile Glu Ile Leu Glu Ser Gly
305 310 315 320
His Leu Arg Lys Leu Asp His Ile Ser Glu Ser Val Pro Val Leu Glu
325 330 335
Leu Phe Ser Asn Ile Trp Gly Ala Gly Thr Lys Thr Ala Gln Met Trp
340 345 350
Tyr Gln Gln Gly Phe Arg Ser Leu Glu Asp Ile Arg Ser Gln Ala Ser

355 360 365
 Leu Thr Thr Gln Gln Ala Ile Gly Leu Lys His Tyr Ser Asp Phe Leu
 370 375 380
 Glu Arg Met Pro Arg Glu Glu Ala Thr Glu Ile Glu Gln Thr Val Gln
 385 390 395 400
 Lys Ala Ala Gln Ala Phe Asn Ser Gly Leu Leu Cys Val Ala Cys Gly

405 410 415
 Ser Tyr Arg Arg Gly Lys Ala Thr Cys Gly Asp Val Asp Val Leu Ile
 420 425 430
 Thr His Pro Asp Gly Arg Ser His Arg Gly Ile Phe Ser Arg Leu Leu
 435 440 445
 Asp Ser Leu Arg Gln Glu Gly Phe Leu Thr Asp Asp Leu Val Ser Gln
 450 455 460
 Glu Glu Asn Gly Gln Gln Gln Lys Tyr Leu Gly Val Cys Arg Leu Pro

465 470 475 480
 Gly Pro Gly Arg Arg His Arg Arg Leu Asp Ile Ile Val Val Pro Tyr
 485 490 495
 Ser Glu Phe Ala Cys Ala Leu Leu Tyr Phe Thr Gly Ser Ala His Phe
 500 505 510
 Asn Arg Ser Met Arg Ala Leu Ala Lys Thr Lys Gly Met Ser Leu Ser
 515 520 525
 Glu His Ala Leu Ser Thr Ala Val Val Arg Asn Thr His Gly Cys Lys

530 535 540
 Val Gly Pro Gly Arg Val Leu Pro Thr Pro Thr Glu Lys Asp Val Phe
 545 550 555 560
 Arg Leu Leu Gly Leu Pro Tyr Arg Glu Pro Ala Glu Arg Asp Trp
 565 570 575

<210> 56

<211> 713

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 56

Met Ala Thr Gly Gln Asp Arg Val Val Ala Leu Val Asp Met Asp Cys
1 5 10 15

Phe Phe Val Gln Val Glu Gln Arg Gln Asn Pro His Leu Arg Asn Lys
20 25 30

Pro Cys Ala Val Val Gln Tyr Lys Ser Trp Lys Gly Gly Gly Ile Ile
35 40 45

Ala Val Ser Tyr Glu Ala Arg Ala Phe Gly Val Thr Arg Ser Met Trp
50 55 60

Ala Asp Asp Ala Lys Lys Leu Cys Pro Asp Leu Leu Leu Ala Gln Val
65 70 75 80

Arg Glu Ser Arg Gly Lys Ala Asn Leu Thr Lys Tyr Arg Glu Ala Ser
85 90 95

Val Glu Val Met Glu Ile Met Ser Arg Phe Ala Val Ile Glu Arg Ala
100 105 110

Ser Ile Asp Glu Ala Tyr Val Asp Leu Thr Ser Ala Val Gln Glu Arg
115 120 125

Leu Gln Lys Leu Gln Gly Gln Pro Ile Ser Ala Asp Leu Leu Pro Ser
130 135 140

Thr Tyr Ile Glu Gly Leu Pro Gln Gly Pro Thr Thr Ala Glu Glu Thr
145 150 155 160

Val Gln Lys Glu Gly Met Arg Lys Gln Gly Leu Phe Gln Trp Leu Asp
165 170 175

Ser Leu Gln Ile Asp Asn Leu Thr Ser Pro Asp Leu Gln Leu Thr Val
180 185 190

Gly Ala Val Ile Val Glu Glu Met Arg Ala Ala Ile Glu Arg Glu Thr
195 200 205

Gly Phe Gln Cys Ser Ala Gly Ile Ser His Asn Lys Val Leu Ala Lys
210 215 220

Leu Ala Cys Gly Leu Asn Lys Pro Asn Arg Gln Thr Leu Val Ser His
225 230 235 240

Gly Ser Val Pro Gln Leu Phe Ser Gln Met Pro Ile Arg Lys Ile Arg
 245 250 255
 Ser Leu Gly Gly Lys Leu Gly Ala Ser Val Ile Glu Ile Leu Gly Ile
 260 265 270

 Glu Tyr Met Gly Glu Leu Thr Gln Phe Thr Glu Ser Gln Leu Gln Ser
 275 280 285
 His Phe Gly Glu Lys Asn Gly Ser Trp Leu Tyr Ala Met Cys Arg Gly
 290 295 300
 Ile Glu His Asp Pro Val Lys Pro Arg Gln Leu Pro Lys Thr Ile Gly
 305 310 315 320
 Cys Ser Lys Asn Phe Pro Gly Lys Thr Ala Leu Ala Thr Arg Glu Gln
 325 330 335

 Val Gln Trp Trp Leu Leu Gln Leu Ala Gln Glu Leu Glu Glu Arg Leu
 340 345 350
 Thr Lys Asp Arg Asn Asp Asn Asp Arg Val Ala Thr Gln Leu Val Val
 355 360 365
 Ser Ile Arg Val Gln Gly Asp Lys Arg Leu Ser Ser Leu Arg Arg Cys
 370 375 380
 Cys Ala Leu Thr Arg Tyr Asp Ala His Lys Met Ser His Asp Ala Phe
 385 390 395 400

 Thr Val Ile Lys Asn Cys Asn Thr Ser Gly Ile Gln Thr Glu Trp Ser
 405 410 415
 Pro Pro Leu Thr Met Leu Phe Leu Cys Ala Thr Lys Phe Ser Ala Ser
 420 425 430
 Ala Pro Ser Ser Ser Thr Asp Ile Thr Ser Phe Leu Ser Ser Asp Pro
 435 440 445
 Ser Ser Leu Pro Lys Val Pro Val Thr Ser Ser Glu Ala Lys Thr Gln
 450 455 460

 Gly Ser Gly Pro Ala Val Thr Ala Thr Lys Lys Ala Thr Thr Ser Leu
 465 470 475 480
 Glu Ser Phe Phe Gln Lys Ala Ala Glu Arg Gln Lys Val Lys Glu Ala

485 490 495
 Ser Leu Ser Ser Leu Thr Ala Pro Thr Gln Ala Pro Met Ser Asn Ser
 500 505 510
 Pro Ser Lys Pro Ser Leu Pro Phe Gln Thr Ser Gln Ser Thr Gly Thr
 515 520 525

 Glu Pro Phe Phe Lys Gln Lys Ser Leu Leu Leu Lys Gln Lys Gln Leu
 530 535 540
 Asn Asn Ser Ser Val Ser Ser Pro Gln Gln Asn Pro Trp Ser Asn Cys
 545 550 555 560
 Lys Ala Leu Pro Asn Ser Leu Pro Thr Glu Tyr Pro Gly Cys Val Pro
 565 570 575
 Val Cys Glu Gly Val Ser Lys Leu Glu Glu Ser Ser Lys Ala Thr Pro
 580 585 590

 Ala Glu Met Asp Leu Ala His Asn Ser Gln Ser Met His Ala Ser Ser
 595 600 605
 Ala Ser Lys Ser Val Leu Glu Val Thr Gln Lys Ala Thr Pro Asn Pro
 610 615 620
 Ser Leu Leu Ala Ala Glu Asp Gln Val Pro Cys Glu Lys Cys Gly Ser
 625 630 635 640
 Leu Val Pro Val Trp Asp Met Pro Glu His Met Asp Tyr His Phe Ala
 645 650 655

 Leu Glu Leu Gln Lys Ser Phe Leu Gln Pro His Ser Ser Asn Pro Gln
 660 665 670
 Val Val Ser Ala Val Ser His Gln Gly Lys Arg Asn Pro Lys Ser Pro
 675 680 685
 Leu Ala Cys Thr Asn Lys Arg Pro Arg Pro Glu Gly Met Gln Thr Leu
 690 695 700
 Glu Ser Phe Phe Lys Pro Leu Thr His
 705 710
 <210> 57
 <211> 494

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 57

Met Leu Pro Lys Arg Arg Arg Ala Arg Val Gly Ser Pro Ser Gly Asp

1 5 10 15

Ala Ala Ser Ser Thr Pro Pro Ser Thr Arg Phe Pro Gly Val Ala Ile

20 25 30

Tyr Leu Val Glu Pro Arg Met Gly Arg Ser Arg Arg Ala Phe Leu Thr

35 40 45

Gly Leu Ala Arg Ser Lys Gly Phe Arg Val Leu Asp Ala Cys Ser Ser

50 55 60

Glu Ala Thr His Val Val Met Glu Glu Thr Ser Ala Glu Glu Ala Val

65 70 75 80

Ser Trp Gln Glu Arg Arg Met Ala Ala Ala Pro Pro Gly Cys Thr Pro

85 90 95

Pro Ala Leu Leu Asp Ile Ser Trp Leu Thr Glu Ser Leu Gly Ala Gly

100 105 110

Gln Pro Val Pro Val Glu Cys Arg His Arg Leu Glu Val Ala Gly Pro

115 120 125

Arg Lys Gly Pro Leu Ser Pro Ala Trp Met Pro Ala Tyr Ala Cys Gln

130 135 140

Arg Pro Thr Pro Leu Thr His His Asn Thr Gly Leu Ser Glu Ala Leu

145 150 155 160

Glu Ile Leu Ala Glu Ala Ala Gly Phe Glu Gly Ser Glu Gly Arg Leu

165 170 175

Leu Thr Phe Cys Arg Ala Ala Ser Val Leu Lys Ala Leu Pro Ser Pro

180 185 190

Val Thr Thr Leu Ser Gln Leu Gln Gly Leu Pro His Phe Gly Glu His

195 200 205

Ser Ser Arg Val Val Gln Glu Leu Leu Glu His Gly Val Cys Glu Glu

Glu Gln Lys Thr Phe Phe Gln Ala Ala Ser Glu Glu Asp Ile Phe Arg
465 470 475 480

His Leu Gly Leu Glu Tyr Leu Pro Pro Glu Gln Arg Asn Ala
485 490

<210> 58

<211> 740

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 58

Met Glu Lys Leu Gly Val Glu Pro Glu Glu Gly Gly Gly Asp Asp

1 5 10 15

Asp Glu Glu Asp Ala Glu Ala Trp Ala Met Glu Leu Ala Asp Val Gly

20 25 30

Ala Ala Ala Ser Ser Gln Gly Val His Asp Gln Val Leu Pro Thr Pro

35 40 45

Asn Ala Ser Ser Arg Val Ile Val His Val Asp Leu Asp Cys Phe Tyr

50 55 60

Ala Gln Val Glu Met Ile Ser Asn Pro Glu Leu Lys Asp Lys Pro Leu

65 70 75 80

Gly Val Gln Gln Lys Tyr Leu Val Val Thr Cys Asn Tyr Glu Ala Arg

85 90 95

Lys Leu Gly Val Lys Lys Leu Met Asn Val Arg Asp Ala Lys Glu Lys

100 105 110

Cys Pro Gln Leu Val Leu Val Asn Gly Glu Asp Leu Thr Arg Tyr Arg

115 120 125

Glu Met Ser Tyr Lys Val Thr Glu Leu Leu Glu Glu Phe Ser Pro Val

130 135 140

Val Glu Arg Leu Gly Phe Asp Glu Asn Phe Val Asp Leu Thr Glu Met

145 150 155 160

Val Glu Lys Arg Leu Gln Gln Leu Gln Ser Asp Glu Leu Ser Ala Val

165 170 175

Thr Val Ser Gly His Val Tyr Asn Asn Gln Ser Ile Asn Leu Leu Asp

180 185 190
Val Leu His Ile Arg Leu Leu Val Gly Ser Gln Ile Ala Ala Glu Met
195 200 205
Arg Glu Ala Met Tyr Asn Gln Leu Gly Leu Thr Gly Cys Ala Gly Val
210 215 220
Ala Ser Asn Lys Leu Leu Ala Lys Leu Val Ser Gly Val Phe Lys Pro
225 230 235 240
Asn Gln Gln Thr Val Leu Leu Pro Glu Ser Cys Gln His Leu Ile His

245 250 255
Ser Leu Asn His Ile Lys Glu Ile Pro Gly Ile Gly Tyr Lys Thr Ala
260 265 270
Lys Cys Leu Glu Ala Leu Gly Ile Asn Ser Val Arg Asp Leu Gln Thr
275 280 285
Phe Ser Pro Lys Ile Leu Glu Lys Glu Leu Gly Ile Ser Val Ala Gln
290 295 300
Arg Ile Gln Lys Leu Ser Phe Gly Glu Asp Asn Ser Pro Val Ile Leu

305 310 315 320
Ser Gly Pro Pro Gln Ser Phe Ser Glu Glu Asp Ser Phe Lys Lys Cys
325 330 335
Ser Ser Glu Val Glu Ala Lys Asn Lys Ile Glu Glu Leu Leu Ala Ser
340 345 350
Leu Leu Asn Arg Val Cys Gln Asp Gly Arg Lys Pro His Thr Val Arg
355 360 365
Leu Ile Ile Arg Arg Tyr Ser Ser Glu Lys His Tyr Gly Arg Glu Ser

370 375 380
Arg Gln Cys Pro Ile Pro Ser His Val Ile Gln Lys Leu Gly Thr Gly
385 390 395 400
Asn Tyr Asp Val Met Thr Pro Met Val Asp Ile Leu Met Lys Leu Phe
405 410 415
Arg Asn Met Val Asn Val Lys Met Pro Phe His Leu Thr Leu Leu Ser

420 425 430
Val Cys Phe Cys Asn Leu Lys Ala Leu Asn Thr Ala Lys Lys Gly Leu

435 440 445
Ile Asp Tyr Tyr Leu Met Pro Ser Leu Ser Thr Thr Ser Arg Ser Gly

450 455 460
Lys His Ser Phe Lys Met Lys Asp Thr His Met Glu Asp Phe Pro Lys

465 470 475 480
Asp Lys Glu Thr Asn Arg Asp Phe Leu Pro Ser Gly Arg Ile Glu Ser

485 490 495
Thr Arg Thr Arg Glu Ser Pro Leu Asp Thr Thr Asn Phe Ser Lys Glu

500 505 510
Lys Asp Ile Asn Glu Phe Pro Leu Cys Ser Leu Pro Glu Gly Val Asp

515 520 525
Gln Glu Val Phe Lys Gln Leu Pro Val Asp Ile Gln Glu Glu Ile Leu

530 535 540
Ser Gly Lys Ser Arg Glu Lys Phe Gln Gly Lys Gly Ser Val Ser Cys

545 550 555 560
Pro Leu His Ala Ser Arg Gly Val Leu Ser Phe Phe Ser Lys Lys Gln

565 570 575
Met Gln Asp Ile Pro Ile Asn Pro Arg Asp His Leu Ser Ser Ser Lys

580 585 590
Gln Val Ser Ser Val Ser Pro Cys Glu Pro Gly Thr Ser Gly Phe Asn

595 600 605
Ser Ser Ser Ser Ser Tyr Met Ser Ser Gln Lys Asp Tyr Ser Tyr Tyr

610 615 620
Leu Asp Asn Arg Leu Lys Asp Glu Arg Ile Ser Gln Gly Pro Lys Glu

625 630 635 640
Pro Gln Gly Phe His Phe Thr Asn Ser Asn Pro Ala Val Ser Ala Phe

645 650 655
His Ser Phe Pro Asn Leu Gln Ser Glu Gln Leu Phe Ser Arg Asn His

660 665 670

Thr Thr Asp Ser His Lys Gln Thr Val Ala Thr Asp Ser His Glu Gly
675 680 685

Leu Thr Glu Asn Arg Glu Pro Asp Ser Val Asp Glu Lys Ile Thr Phe

690 695 700

Pro Ser Asp Ile Asp Pro Gln Val Phe Tyr Glu Leu Pro Glu Ala Val

705 710 715 720

Gln Lys Glu Leu Leu Ala Glu Trp Lys Arg Ala Gly Ser Asp Phe His

725 730 735

Ile Gly His Lys

740

<210> 59

<211> 870

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 59

Met Asp Ser Thr Lys Glu Lys Cys Asp Ser Tyr Lys Asp Asp Leu Leu

1 5 10 15

Leu Arg Met Gly Leu Asn Asp Asn Lys Ala Gly Met Glu Gly Leu Asp

20 25 30

Lys Glu Lys Ile Asn Lys Ile Ile Met Glu Ala Thr Lys Gly Ser Arg

35 40 45

Phe Tyr Gly Asn Glu Leu Lys Lys Glu Lys Gln Val Asn Gln Arg Ile

50 55 60

Glu Asn Met Met Gln Gln Lys Ala Gln Ile Thr Ser Gln Gln Leu Arg

65 70 75 80

Lys Ala Gln Leu Gln Val Asp Arg Phe Ala Met Glu Leu Glu Gln Ser

85 90 95

Arg Asn Leu Ser Asn Thr Ile Val His Ile Asp Met Asp Ala Phe Tyr

100 105 110

Ala Ala Val Glu Met Arg Asp Asn Pro Glu Leu Lys Asp Lys Pro Ile

115 120 125

Ala Val Gly Ser Met Ser Met Leu Ser Thr Ser Asn Tyr His Ala Arg

130 135 140

Arg Phe Gly Val Arg Ala Ala Met Pro Gly Phe Ile Ala Lys Arg Leu

145 150 155 160

Cys Pro Gln Leu Ile Ile Val Pro Pro Asn Phe Asp Lys Tyr Arg Ala

165 170 175

Val Ser Lys Glu Val Lys Glu Ile Leu Ala Asp Tyr Asp Pro Asn Phe

180 185 190

Met Ala Met Ser Leu Asp Glu Ala Tyr Leu Asn Ile Thr Lys His Leu

195 200 205

Glu Glu Arg Gln Asn Trp Pro Glu Asp Lys Arg Arg Tyr Phe Ile Lys

210 215 220

Met Gly Ser Ser Val Glu Asn Asp Asn Pro Gly Lys Glu Val Asn Lys

225 230 235 240

Leu Ser Glu His Glu Arg Ser Ile Ser Pro Leu Leu Phe Glu Glu Ser

245 250 255

Pro Ser Asp Val Gln Pro Pro Gly Asp Pro Phe Gln Val Asn Phe Glu

260 265 270

Glu Gln Asn Asn Pro Gln Ile Leu Gln Asn Ser Val Val Phe Gly Thr

275 280 285

Ser Ala Gln Glu Val Val Lys Glu Ile Arg Phe Arg Ile Glu Gln Lys

290 295 300

Thr Thr Leu Thr Ala Ser Ala Gly Ile Ala Pro Asn Thr Met Leu Ala

305 310 315 320

Lys Val Cys Ser Asp Lys Asn Lys Pro Asn Gly Gln Tyr Gln Ile Leu

325 330 335

Pro Asn Arg Gln Ala Val Met Asp Phe Ile Lys Asp Leu Pro Ile Arg

340 345 350

Lys Val Ser Gly Ile Gly Lys Val Thr Glu Lys Met Leu Lys Ala Leu

355 360 365

Gly Ile Ile Thr Cys Thr Glu Leu Tyr Gln Gln Arg Ala Leu Leu Ser

370 375 380
 Leu Leu Phe Ser Glu Thr Ser Trp His Tyr Phe Leu His Ile Ser Leu

 385 390 395 400
 Gly Leu Gly Ser Thr His Leu Thr Arg Asp Gly Glu Arg Lys Ser Met
 405 410 415
 Ser Val Glu Arg Thr Phe Ser Glu Ile Asn Lys Ala Glu Glu Gln Tyr
 420 425 430
 Ser Leu Cys Gln Glu Leu Cys Ser Glu Leu Ala Gln Asp Leu Gln Lys
 435 440 445
 Glu Arg Leu Lys Gly Arg Thr Val Thr Ile Lys Leu Lys Asn Val Asn

 450 455 460
 Phe Glu Val Lys Thr Arg Ala Ser Thr Val Ser Ser Val Val Ser Thr
 465 470 475 480
 Ala Glu Glu Ile Phe Ala Ile Ala Lys Glu Leu Leu Lys Thr Glu Ile
 485 490 495
 Asp Ala Asp Phe Pro His Pro Leu Arg Leu Arg Leu Met Gly Val Arg
 500 505 510
 Ile Ser Ser Phe Pro Asn Glu Glu Asp Arg Lys His Gln Gln Arg Ser

 515 520 525
 Ile Ile Gly Phe Leu Gln Ala Gly Asn Gln Ala Leu Ser Ala Thr Glu
 530 535 540
 Cys Thr Leu Glu Lys Thr Asp Lys Asp Lys Phe Val Lys Pro Leu Glu
 545 550 555 560
 Met Ser His Lys Lys Ser Phe Phe Asp Lys Lys Arg Ser Glu Arg Lys
 565 570 575
 Trp Ser His Gln Asp Thr Phe Lys Cys Glu Ala Val Asn Lys Gln Ser

 580 585 590
 Phe Gln Thr Ser Gln Pro Phe Gln Val Leu Lys Lys Lys Met Asn Glu
 595 600 605
 Asn Leu Glu Ile Ser Glu Asn Ser Asp Asp Cys Gln Ile Leu Thr Cys
 610 615 620

Pro Val Cys Phe Arg Ala Gln Gly Cys Ile Ser Leu Glu Ala Leu Asn
625 630 635 640
Lys His Val Asp Glu Cys Leu Asp Gly Pro Ser Ile Ser Glu Asn Phe
645 650 655
Lys Met Phe Ser Cys Ser His Val Ser Ala Thr Lys Val Asn Lys Lys
660 665 670
Glu Asn Val Pro Ala Ser Ser Leu Cys Glu Lys Gln Asp Tyr Glu Ala
675 680 685
His Pro Lys Ile Lys Glu Ile Ser Ser Val Asp Cys Ile Ala Leu Val
690 695 700
Asp Thr Ile Asp Asn Ser Ser Lys Ala Glu Ser Ile Asp Ala Leu Ser
705 710 715 720
Asn Lys His Ser Lys Glu Glu Cys Ser Ser Leu Pro Ser Lys Ser Phe
725 730 735
Asn Ile Glu His Cys His Gln Asn Ser Ser Ser Thr Val Ser Leu Glu
740 745 750
Asn Glu Asp Val Gly Ser Phe Arg Gln Glu Tyr Arg Gln Pro Tyr Leu
755 760 765
Cys Glu Val Lys Thr Gly Gln Ala Leu Val Cys Pro Val Cys Asn Val
770 775 780
Glu Gln Lys Thr Ser Asp Leu Thr Leu Phe Asn Val His Val Asp Val
785 790 795 800
Cys Leu Asn Lys Ser Phe Ile Gln Glu Leu Arg Lys Asp Lys Phe Asn
805 810 815
Pro Val Asn Gln Pro Lys Glu Ser Ser Arg Ser Thr Gly Ser Ser Ser
820 825 830
Gly Val Gln Lys Ala Val Thr Arg Thr Lys Arg Pro Gly Leu Met Thr
835 840 845
Lys Tyr Ser Thr Ser Lys Lys Ile Lys Pro Asn Asn Pro Lys His Thr
850 855 860
Leu Asp Ile Phe Phe Lys

865	870															
<210>	60															
<211>	1468															
<212>	PRT															
<213>	Artificial Sequence															
<220><223>	Synthetic polypeptide															
<400>	60															
Met	Ala	Pro	Val	His	Gly	Asp	Asp	Cys	Glu	Ile	Gly	Ala	Ser	Ala	Leu	
1				5					10					15		
Ser	Asp	Ser	Gly	Ser	Phe	Val	Ser	Ser	Arg	Ala	Arg	Arg	Glu	Lys	Lys	
				20					25					30		
Ser	Lys	Lys	Gly	Arg	Gln	Glu	Ala	Leu	Glu	Arg	Leu	Lys	Lys	Ala	Lys	
				35					40					45		
Ala	Gly	Glu	Lys	Tyr	Lys	Tyr	Glu	Val	Glu	Asp	Phe	Thr	Gly	Val	Tyr	
				50					55					60		
Glu	Glu	Val	Asp	Glu	Glu	Gln	Tyr	Ser	Lys	Leu	Val	Gln	Ala	Arg	Gln	
65					70					75					80	
Asp	Asp	Asp	Trp	Ile	Val	Asp	Asp	Asp	Gly	Ile	Gly	Tyr	Val	Glu	Asp	
					85					90					95	
Gly	Arg	Glu	Ile	Phe	Asp	Asp	Asp	Leu	Glu	Asp	Asp	Ala	Leu	Asp	Ala	
				100					105					110		
Asp	Glu	Lys	Gly	Lys	Asp	Gly	Lys	Ala	Arg	Asn	Lys	Asp	Lys	Arg	Asn	
				115					120					125		
Val	Lys	Lys	Leu	Ala	Val	Thr	Lys	Pro	Asn	Asn	Ile	Lys	Ser	Met	Phe	
				130					135					140		
Ile	Ala	Cys	Ala	Gly	Lys	Lys	Thr	Ala	Asp	Lys	Ala	Val	Asp	Leu	Ser	
145					150					155					160	
Lys	Asp	Gly	Leu	Leu	Gly	Asp	Ile	Leu	Gln	Asp	Leu	Asn	Thr	Glu	Thr	
				165					170					175		
Pro	Gln	Ile	Thr	Pro	Pro	Pro	Val	Met	Ile	Leu	Lys	Lys	Lys	Arg	Ser	
				180					185					190		
Ile	Gly	Ala	Ser	Pro	Asn	Pro	Phe	Ser	Val	His	Thr	Ala	Thr	Ala	Val	

195 200 205
Pro Ser Gly Lys Ile Ala Ser Pro Val Ser Arg Lys Glu Pro Pro Leu

210 215 220
Thr Pro Val Pro Leu Lys Arg Ala Glu Phe Ala Gly Asp Asp Val Gln
225 230 235 240
Val Glu Ser Thr Glu Glu Glu Gln Glu Ser Gly Ala Met Glu Phe Glu
245 250 255
Asp Gly Asp Phe Asp Glu Pro Met Glu Val Glu Glu Val Asp Leu Glu
260 265 270
Pro Met Ala Ala Lys Ala Trp Asp Lys Glu Ser Glu Pro Ala Glu Glu

275 280 285
Val Lys Gln Glu Ala Asp Ser Gly Lys Gly Thr Val Ser Tyr Leu Gly
290 295 300
Ser Phe Leu Pro Asp Val Ser Cys Trp Asp Ile Asp Gln Glu Gly Asp
305 310 315 320
Ser Ser Phe Ser Val Gln Glu Val Gln Val Asp Ser Ser His Leu Pro
325 330 335
Leu Val Lys Gly Ala Asp Glu Glu Gln Val Phe His Phe Tyr Trp Leu

340 345 350
Asp Ala Tyr Glu Asp Gln Tyr Asn Gln Pro Gly Val Val Phe Leu Phe
355 360 365
Gly Lys Val Trp Ile Glu Ser Ala Glu Thr His Val Ser Cys Cys Val
370 375 380
Met Val Lys Asn Ile Glu Arg Thr Leu Tyr Phe Leu Pro Arg Glu Met
385 390 395 400
Lys Ile Asp Leu Asn Thr Gly Lys Glu Thr Gly Thr Pro Ile Ser Met

405 410 415
Lys Asp Val Tyr Glu Glu Phe Asp Glu Lys Ile Ala Thr Lys Tyr Lys
420 425 430
Ile Met Lys Phe Lys Ser Lys Pro Val Glu Lys Asn Tyr Ala Phe Glu
435 440 445

Ile Pro Asp Val Pro Glu Lys Ser Glu Tyr Leu Glu Val Lys Tyr Ser
450 455 460

Ala Glu Met Pro Gln Leu Pro Gln Asp Leu Lys Gly Glu Thr Phe Ser

465 470 475 480

His Val Phe Gly Thr Asn Thr Ser Ser Leu Glu Leu Phe Leu Met Asn
485 490 495

Arg Lys Ile Lys Gly Pro Cys Trp Leu Glu Val Lys Ser Pro Gln Leu
500 505 510

Leu Asn Gln Pro Val Ser Trp Cys Lys Val Glu Ala Met Ala Leu Lys
515 520 525

Pro Asp Leu Val Asn Val Ile Lys Asp Val Ser Pro Pro Pro Leu Val

530 535 540

Val Met Ala Phe Ser Met Lys Thr Met Gln Asn Ala Lys Asn His Gln
545 550 555 560

Asn Glu Ile Ile Ala Met Ala Ala Leu Val His His Ser Phe Ala Leu
565 570 575

Asp Lys Ala Ala Pro Lys Pro Pro Phe Gln Ser His Phe Cys Val Val
580 585 590

Ser Lys Pro Lys Asp Cys Ile Phe Pro Tyr Ala Phe Lys Glu Val Ile

595 600 605

Glu Lys Lys Asn Val Lys Val Glu Val Ala Ala Thr Glu Arg Thr Leu
610 615 620

Leu Gly Phe Phe Leu Ala Lys Val His Lys Ile Asp Pro Asp Ile Ile
625 630 635 640

Val Gly His Asn Ile Tyr Gly Phe Glu Leu Glu Val Leu Leu Gln Arg
645 650 655

Ile Asn Val Cys Lys Ala Pro His Trp Ser Lys Ile Gly Arg Leu Lys

660 665 670

Arg Ser Asn Met Pro Lys Leu Gly Gly Arg Ser Gly Phe Gly Glu Arg
675 680 685

Asn Ala Thr Cys Gly Arg Met Ile Cys Asp Val Glu Ile Ser Ala Lys

690 695 700
Glu Leu Ile Arg Cys Lys Ser Tyr His Leu Ser Glu Leu Val Gln Gln
705 710 715 720
Ile Leu Lys Thr Glu Arg Val Val Ile Pro Met Glu Asn Ile Gln Asn

725 730 735
Met Tyr Ser Glu Ser Ser Gln Leu Leu Tyr Leu Leu Glu His Thr Trp
740 745 750
Lys Asp Ala Lys Phe Ile Leu Gln Ile Met Cys Glu Leu Asn Val Leu
755 760 765
Pro Leu Ala Leu Gln Ile Thr Asn Ile Ala Gly Asn Ile Met Ser Arg
770 775 780
Thr Leu Met Gly Gly Arg Ser Glu Arg Asn Glu Phe Leu Leu Leu His

785 790 795 800
Ala Phe Tyr Glu Asn Asn Tyr Ile Val Pro Asp Lys Gln Ile Phe Arg
805 810 815
Lys Pro Gln Gln Lys Leu Gly Asp Glu Asp Glu Glu Ile Asp Gly Asp
820 825 830
Thr Asn Lys Tyr Lys Lys Gly Arg Lys Lys Ala Ala Tyr Ala Gly Gly
835 840 845
Leu Val Leu Asp Pro Lys Val Gly Phe Tyr Asp Lys Phe Ile Leu Leu

850 855 860
Leu Asp Phe Asn Ser Leu Tyr Pro Ser Ile Ile Gln Glu Phe Asn Ile
865 870 875 880
Cys Phe Thr Thr Val Gln Arg Val Ala Ser Glu Ala Gln Lys Val Thr
885 890 895
Glu Asp Gly Glu Gln Glu Gln Ile Pro Glu Leu Pro Asp Pro Ser Leu
900 905 910
Glu Met Gly Ile Leu Pro Arg Glu Ile Arg Lys Leu Val Glu Arg Arg

915 920 925
Lys Gln Val Lys Gln Leu Met Lys Gln Gln Asp Leu Asn Pro Asp Leu
930 935 940

Ile Leu Gln Tyr Asp Ile Arg Gln Lys Ala Leu Lys Leu Thr Ala Asn
945 950 955 960
Ser Met Tyr Gly Cys Leu Gly Phe Ser Tyr Ser Arg Phe Tyr Ala Lys
965 970 975
Pro Leu Ala Ala Leu Val Thr Tyr Lys Gly Arg Glu Ile Leu Met His
980 985 990
Thr Lys Glu Met Val Gln Lys Met Asn Leu Glu Val Ile Tyr Gly Asp
995 1000 1005
Thr Asp Ser Ile Met Ile Asn Thr Asn Ser Thr Asn Leu Glu Glu
1010 1015 1020
Val Phe Lys Leu Gly Asn Lys Val Lys Ser Glu Val Asn Lys Leu
1025 1030 1035
Tyr Lys Leu Leu Glu Ile Asp Ile Asp Gly Val Phe Lys Ser Leu
1040 1045 1050
Leu Leu Leu Lys Lys Lys Lys Tyr Ala Ala Leu Val Val Glu Pro
1055 1060 1065
Thr Ser Asp Gly Asn Tyr Val Thr Lys Gln Glu Leu Lys Gly Leu
1070 1075 1080
Asp Ile Val Arg Arg Asp Trp Cys Asp Leu Ala Lys Asp Thr Gly
1085 1090 1095
Asn Phe Val Ile Gly Gln Ile Leu Ser Asp Gln Ser Arg Asp Thr
1100 1105 1110
Ile Val Glu Asn Ile Gln Lys Arg Leu Ile Glu Ile Gly Glu Asn
1115 1120 1125
Val Leu Asn Gly Ser Val Pro Val Ser Gln Phe Glu Ile Asn Lys
1130 1135 1140
Ala Leu Thr Lys Asp Pro Gln Asp Tyr Pro Asp Lys Lys Ser Leu
1145 1150 1155
Pro His Val His Val Ala Leu Trp Ile Asn Ser Gln Gly Gly Arg
1160 1165 1170
Lys Val Lys Ala Gly Asp Thr Val Ser Tyr Val Ile Cys Gln Asp

1175	1180	1185
Gly Ser Asn Leu Thr Ala Ser	Gln Arg Ala Tyr Ala	Pro Glu Gln
1190	1195	1200
Leu Gln Lys Gln Asp Asn Leu	Thr Ile Asp Thr Gln	Tyr Tyr Leu
1205	1210	1215
Ala Gln Gln Ile His Pro Val	Val Ala Arg Ile Cys	Glu Pro Ile
1220	1225	1230
Asp Gly Ile Asp Ala Val Leu	Ile Ala Thr Trp Leu	Gly Leu Asp
1235	1240	1245
Pro Thr Gln Phe Arg Val His	His Tyr His Lys Asp	Glu Glu Asn
1250	1255	1260
Asp Ala Leu Leu Gly Gly Pro	Ala Gln Leu Thr Asp	Glu Glu Lys
1265	1270	1275
Tyr Arg Asp Cys Glu Arg Phe	Lys Cys Pro Cys Pro	Thr Cys Gly
1280	1285	1290
Thr Glu Asn Ile Tyr Asp Asn	Val Phe Asp Gly Ser	Gly Thr Asp
1295	1300	1305
Met Glu Pro Ser Leu Tyr Arg	Cys Ser Asn Ile Asp	Cys Lys Ala
1310	1315	1320
Ser Pro Leu Thr Phe Thr Val	Gln Leu Ser Asn Lys	Leu Ile Met
1325	1330	1335
Asp Ile Arg Arg Phe Ile Lys	Lys Tyr Tyr Asp Gly	Trp Leu Ile
1340	1345	1350
Cys Glu Glu Pro Thr Cys Arg	Asn Arg Thr Arg His	Leu Pro Leu
1355	1360	1365
Gln Phe Ser Arg Thr Gly Pro	Leu Cys Pro Ala Cys	Met Lys Ala
1370	1375	1380
Thr Leu Gln Pro Glu Tyr Ser	Asp Lys Ser Leu Tyr	Thr Gln Leu
1385	1390	1395
Cys Phe Tyr Arg Tyr Ile Phe	Asp Ala Glu Cys Ala	Leu Glu Lys
1400	1405	1410

Leu Thr Thr Asp His Glu Lys Asp Lys Leu Lys Lys Gln Phe Phe
 1415 1420 1425
 Thr Pro Lys Val Leu Gln Asp Tyr Arg Lys Leu Lys Asn Thr Ala
 1430 1435 1440
 Glu Gln Phe Leu Ser Arg Ser Gly Tyr Ser Glu Val Asn Leu Ser
 1445 1450 1455
 Lys Leu Phe Ala Gly Cys Ala Val Lys Ser
 1460 1465

<210> 61

<211> 1107

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 61

Met Asp Gly Lys Arg Arg Pro Gly Pro Gly Pro Gly Val Pro Pro Lys
 1 5 10 15
 Arg Ala Arg Gly Gly Leu Trp Asp Asp Asp Ala Pro Arg Pro Ser
 20 25 30
 Gln Phe Glu Glu Asp Leu Ala Leu Met Glu Glu Met Glu Ala Glu His
 35 40 45

Arg Leu Gln Glu Gln Glu Glu Glu Leu Gln Ser Val Leu Glu Gly
 50 55 60
 Val Ala Asp Gly Gln Val Pro Pro Ser Ala Ile Asp Pro Arg Trp Leu
 65 70 75 80
 Arg Pro Thr Pro Pro Ala Leu Asp Pro Gln Thr Glu Pro Leu Ile Phe
 85 90 95
 Gln Gln Leu Glu Ile Asp His Tyr Val Gly Pro Ala Gln Pro Val Pro
 100 105 110

Gly Gly Pro Pro Pro Ser His Gly Ser Val Pro Val Leu Arg Ala Phe
 115 120 125
 Gly Val Thr Asp Glu Gly Phe Ser Val Cys Cys His Ile His Gly Phe
 130 135 140

Ala Pro Tyr Phe Tyr Thr Pro Ala Pro Pro Gly Phe Gly Pro Glu His
145 150 155 160
Met Gly Asp Leu Gln Arg Glu Leu Asn Leu Ala Ile Ser Arg Asp Ser
165 170 175

Arg Gly Gly Arg Glu Leu Thr Gly Pro Ala Val Leu Ala Val Glu Leu
180 185 190
Cys Ser Arg Glu Ser Met Phe Gly Tyr His Gly His Gly Pro Ser Pro
195 200 205
Phe Leu Arg Ile Thr Val Ala Leu Pro Arg Leu Val Ala Pro Ala Arg
210 215 220
Arg Leu Leu Glu Gln Gly Ile Arg Val Ala Gly Leu Gly Thr Pro Ser
225 230 235 240

Phe Ala Pro Tyr Glu Ala Asn Val Asp Phe Glu Ile Arg Phe Met Val
245 250 255
Asp Thr Asp Ile Val Gly Cys Asn Trp Leu Glu Leu Pro Ala Gly Lys
260 265 270
Tyr Ala Leu Arg Leu Lys Glu Lys Ala Thr Gln Cys Gln Leu Glu Ala
275 280 285
Asp Val Leu Trp Ser Asp Val Val Ser His Pro Pro Glu Gly Pro Trp
290 295 300

Gln Arg Ile Ala Pro Leu Arg Val Leu Ser Phe Asp Ile Glu Cys Ala
305 310 315 320
Gly Arg Lys Gly Ile Phe Pro Glu Pro Glu Arg Asp Pro Val Ile Gln
325 330 335
Ile Cys Ser Leu Gly Leu Arg Trp Gly Glu Pro Glu Pro Phe Leu Arg
340 345 350
Leu Ala Leu Thr Leu Arg Pro Cys Ala Pro Ile Leu Gly Ala Lys Val
355 360 365

Gln Ser Tyr Glu Lys Glu Glu Asp Leu Leu Gln Ala Trp Ser Thr Phe
370 375 380
Ile Arg Ile Met Asp Pro Asp Val Ile Thr Gly Tyr Asn Ile Gln Asn

385 390 395 400
 Phe Asp Leu Pro Tyr Leu Ile Ser Arg Ala Gln Thr Leu Lys Val Gln
 405 410 415
 Thr Phe Pro Phe Leu Gly Arg Val Ala Gly Leu Cys Ser Asn Ile Arg
 420 425 430

 Asp Ser Ser Phe Gln Ser Lys Gln Thr Gly Arg Arg Asp Thr Lys Val
 435 440 445
 Val Ser Met Val Gly Arg Val Gln Met Asp Met Leu Gln Val Leu Leu
 450 455 460
 Arg Glu Tyr Lys Leu Arg Ser Tyr Thr Leu Asn Ala Val Ser Phe His
 465 470 475 480
 Phe Leu Gly Glu Gln Lys Glu Asp Val Gln His Ser Ile Ile Thr Asp
 485 490 495

 Leu Gln Asn Gly Asn Asp Gln Thr Arg Arg Arg Leu Ala Val Tyr Cys
 500 505 510
 Leu Lys Asp Ala Tyr Leu Pro Leu Arg Leu Leu Glu Arg Leu Met Val
 515 520 525
 Leu Val Asn Ala Val Glu Met Ala Arg Val Thr Gly Val Pro Leu Ser
 530 535 540
 Tyr Leu Leu Ser Arg Gly Gln Gln Val Lys Val Val Ser Gln Leu Leu
 545 550 555 560

 Arg Gln Ala Met His Glu Gly Leu Leu Met Pro Val Val Lys Ser Glu
 565 570 575
 Gly Gly Glu Asp Tyr Thr Gly Ala Thr Val Ile Glu Pro Leu Lys Gly
 580 585 590
 Tyr Tyr Asp Val Pro Ile Ala Thr Leu Asp Phe Ser Ser Leu Tyr Pro
 595 600 605
 Ser Ile Met Met Ala His Asn Leu Cys Tyr Thr Thr Leu Leu Arg Pro
 610 615 620

 Gly Thr Ala Gln Lys Leu Gly Leu Thr Glu Asp Gln Phe Ile Arg Thr
 625 630 635 640

Pro Thr Gly Asp Glu Phe Val Lys Thr Ser Val Arg Lys Gly Leu Leu
645 650 655

Pro Gln Ile Leu Glu Asn Leu Leu Ser Ala Arg Lys Arg Ala Lys Ala
660 665 670

Glu Leu Ala Lys Glu Thr Asp Pro Leu Arg Arg Gln Val Leu Asp Gly
675 680 685

Arg Gln Leu Ala Leu Lys Val Ser Ala Asn Ser Val Tyr Gly Phe Thr
690 695 700

Gly Ala Gln Val Gly Lys Leu Pro Cys Leu Glu Ile Ser Gln Ser Val
705 710 715 720

Thr Gly Phe Gly Arg Gln Met Ile Glu Lys Thr Lys Gln Leu Val Glu
725 730 735

Ser Lys Tyr Thr Val Glu Asn Gly Tyr Ser Thr Ser Ala Lys Val Val
740 745 750

Tyr Gly Asp Thr Asp Ser Val Met Cys Arg Phe Gly Val Ser Ser Val
755 760 765

Ala Glu Ala Met Ala Leu Gly Arg Glu Ala Ala Asp Trp Val Ser Gly
770 775 780

His Phe Pro Ser Pro Ile Arg Leu Glu Phe Glu Lys Val Tyr Phe Pro
785 790 795 800

Tyr Leu Leu Ile Ser Lys Lys Arg Tyr Ala Gly Leu Leu Phe Ser Ser
805 810 815

Arg Pro Asp Ala His Asp Arg Met Asp Cys Lys Gly Leu Glu Ala Val
820 825 830

Arg Arg Asp Asn Cys Pro Leu Val Ala Asn Leu Val Thr Ala Ser Leu
835 840 845

Arg Arg Leu Leu Ile Asp Arg Asp Pro Glu Gly Ala Val Ala His Ala
850 855 860

Gln Asp Val Ile Ser Asp Leu Leu Cys Asn Arg Ile Asp Ile Ser Gln
865 870 875 880

Leu Val Ile Thr Lys Glu Leu Thr Arg Ala Ala Ser Asp Tyr Ala Gly

885 890 895
 Lys Gln Ala His Val Glu Leu Ala Glu Arg Met Arg Lys Arg Asp Pro
 900 905 910
 Gly Ser Ala Pro Ser Leu Gly Asp Arg Val Pro Tyr Val Ile Ile Ser
 915 920 925
 Ala Ala Lys Gly Val Ala Ala Tyr Met Lys Ser Glu Asp Pro Leu Phe
 930 935 940

 Val Leu Glu His Ser Leu Pro Ile Asp Thr Gln Tyr Tyr Leu Glu Gln
 945 950 955 960
 Gln Leu Ala Lys Pro Leu Leu Arg Ile Phe Glu Pro Ile Leu Gly Glu
 965 970 975
 Gly Arg Ala Glu Ala Val Leu Leu Arg Gly Asp His Thr Arg Cys Lys
 980 985 990
 Thr Val Leu Thr Gly Lys Val Gly Gly Leu Leu Ala Phe Ala Lys Arg
 995 1000 1005

 Arg Asn Cys Cys Ile Gly Cys Arg Thr Val Leu Ser His Gln Gly
 1010 1015 1020
 Ala Val Cys Glu Phe Cys Gln Pro Arg Glu Ser Glu Leu Tyr Gln
 1025 1030 1035
 Lys Glu Val Ser His Leu Asn Ala Leu Glu Glu Arg Phe Ser Arg
 1040 1045 1050
 Leu Trp Thr Gln Cys Gln Arg Cys Gln Gly Ser Leu His Glu Asp
 1055 1060 1065

 Val Ile Cys Thr Ser Arg Asp Cys Pro Ile Phe Tyr Met Arg Lys
 1070 1075 1080
 Lys Val Arg Lys Asp Leu Glu Asp Gln Glu Gln Leu Leu Arg Arg
 1085 1090 1095
 Phe Gly Pro Pro Gly Pro Glu Ala Trp
 1100 1105

 <210> 62
 <211> 1239
 <212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 62

Met Ser Arg Leu Leu Trp Arg Lys Val Ala Gly Ala Thr Val Gly Pro

1 5 10 15
Gly Pro Val Pro Ala Pro Gly Arg Trp Val Ser Ser Ser Val Pro Ala

20 25 30
Ser Asp Pro Ser Asp Gly Gln Arg Arg Arg Gln Gln Gln Gln Gln Gln

35 40 45
Gln Gln Gln Gln Gln Gln Gln Pro Gln Gln Pro Gln Val Leu Ser Ser

50 55 60
Glu Gly Gly Gln Leu Arg His Asn Pro Leu Asp Ile Gln Met Leu Ser

65 70 75 80
Arg Gly Leu His Glu Gln Ile Phe Gly Gln Gly Gly Glu Met Pro Gly

85 90 95
Glu Ala Ala Val Arg Arg Ser Val Glu His Leu Gln Lys His Gly Leu

100 105 110
Trp Gly Gln Pro Ala Val Pro Leu Pro Asp Val Glu Leu Arg Leu Pro

115 120 125
Pro Leu Tyr Gly Asp Asn Leu Asp Gln His Phe Arg Leu Leu Ala Gln

130 135 140
Lys Gln Ser Leu Pro Tyr Leu Glu Ala Ala Asn Leu Leu Leu Gln Ala

145 150 155 160
Gln Leu Pro Pro Lys Pro Pro Ala Trp Ala Trp Ala Glu Gly Trp Thr

165 170 175
Arg Tyr Gly Pro Glu Gly Glu Ala Val Pro Val Ala Ile Pro Glu Glu

180 185 190
Arg Ala Leu Val Phe Asp Val Glu Val Cys Leu Ala Glu Gly Thr Cys

195 200 205
Pro Thr Leu Ala Val Ala Ile Ser Pro Ser Ala Trp Tyr Ser Trp Cys

210 215 220

Ser Gln Arg Leu Val Glu Glu Arg Tyr Ser Trp Thr Ser Gln Leu Ser
 225 230 235 240
 Pro Ala Asp Leu Ile Pro Leu Glu Val Pro Thr Gly Ala Ser Ser Pro
 245 250 255
 Thr Gln Arg Asp Trp Gln Glu Gln Leu Val Val Gly His Asn Val Ser
 260 265 270
 Phe Asp Arg Ala His Ile Arg Glu Gln Tyr Leu Ile Gln Gly Ser Arg
 275 280 285
 Met Arg Phe Leu Asp Thr Met Ser Met His Met Ala Ile Ser Gly Leu
 290 295 300
 Ser Ser Phe Gln Arg Ser Leu Trp Ile Ala Ala Lys Gln Gly Lys His
 305 310 315 320
 Lys Val Gln Pro Pro Thr Lys Gln Gly Gln Lys Ser Gln Arg Lys Ala
 325 330 335
 Arg Arg Gly Pro Ala Ile Ser Ser Trp Asp Trp Leu Asp Ile Ser Ser
 340 345 350
 Val Asn Ser Leu Ala Glu Val His Arg Leu Tyr Val Gly Gly Pro Pro
 355 360 365
 Leu Glu Lys Glu Pro Arg Glu Leu Phe Val Lys Gly Thr Met Lys Asp
 370 375 380
 Ile Arg Glu Asn Phe Gln Asp Leu Met Gln Tyr Cys Ala Gln Asp Val
 385 390 395 400
 Trp Ala Thr His Glu Val Phe Gln Gln Gln Leu Pro Leu Phe Leu Glu
 405 410 415
 Arg Cys Pro His Pro Val Thr Leu Ala Gly Met Leu Glu Met Gly Val
 420 425 430
 Ser Tyr Leu Pro Val Asn Gln Asn Trp Glu Arg Tyr Leu Ala Glu Ala
 435 440 445
 Gln Gly Thr Tyr Glu Glu Leu Gln Arg Glu Met Lys Lys Ser Leu Met
 450 455 460
 Asp Leu Ala Asn Asp Ala Cys Gln Leu Leu Ser Gly Glu Arg Tyr Lys

465 470 475 480
 Glu Asp Pro Trp Leu Trp Asp Leu Glu Trp Asp Leu Gln Glu Phe Lys
 485 490 495
 Gln Lys Lys Ala Lys Lys Val Lys Lys Glu Pro Ala Thr Ala Ser Lys
 500 505 510
 Leu Pro Ile Glu Gly Ala Gly Ala Pro Gly Asp Pro Met Asp Gln Glu

 515 520 525
 Asp Leu Gly Pro Cys Ser Glu Glu Glu Glu Phe Gln Gln Asp Val Met
 530 535 540
 Ala Arg Ala Cys Leu Gln Lys Leu Lys Gly Thr Thr Glu Leu Leu Pro
 545 550 555 560
 Lys Arg Pro Gln His Leu Pro Gly His Pro Gly Trp Tyr Arg Lys Leu
 565 570 575
 Cys Pro Arg Leu Asp Asp Pro Ala Trp Thr Pro Gly Pro Ser Leu Leu

 580 585 590
 Ser Leu Gln Met Arg Val Thr Pro Lys Leu Met Ala Leu Thr Trp Asp
 595 600 605
 Gly Phe Pro Leu His Tyr Ser Glu Arg His Gly Trp Gly Tyr Leu Val
 610 615 620
 Pro Gly Arg Arg Asp Asn Leu Ala Lys Leu Pro Thr Gly Thr Thr Leu
 625 630 635 640
 Glu Ser Ala Gly Val Val Cys Pro Tyr Arg Ala Ile Glu Ser Leu Tyr

 645 650 655
 Arg Lys His Cys Leu Glu Gln Gly Lys Gln Gln Leu Met Pro Gln Glu
 660 665 670
 Ala Gly Leu Ala Glu Glu Phe Leu Leu Thr Asp Asn Ser Ala Ile Trp
 675 680 685
 Gln Thr Val Glu Glu Leu Asp Tyr Leu Glu Val Glu Ala Glu Ala Lys
 690 695 700
 Met Glu Asn Leu Arg Ala Ala Val Pro Gly Gln Pro Leu Ala Leu Thr

 705 710 715 720

Ala Arg Gly Gly Pro Lys Asp Thr Gln Pro Ser Tyr His His Gly Asn
725 730 735

Gly Pro Tyr Asn Asp Val Asp Ile Pro Gly Cys Trp Phe Phe Lys Leu
740 745 750

Pro His Lys Asp Gly Asn Ser Cys Asn Val Gly Ser Pro Phe Ala Lys
755 760 765

Asp Phe Leu Pro Lys Met Glu Asp Gly Thr Leu Gln Ala Gly Pro Gly
770 775 780

Gly Ala Ser Gly Pro Arg Ala Leu Glu Ile Asn Lys Met Ile Ser Phe
785 790 795 800

Trp Arg Asn Ala His Lys Arg Ile Ser Ser Gln Met Val Val Trp Leu
805 810 815

Pro Arg Ser Ala Leu Pro Arg Ala Val Ile Arg His Pro Asp Tyr Asp
820 825 830

Glu Glu Gly Leu Tyr Gly Ala Ile Leu Pro Gln Val Val Thr Ala Gly
835 840 845

Thr Ile Thr Arg Arg Ala Val Glu Pro Thr Trp Leu Thr Ala Ser Asn
850 855 860

Ala Arg Pro Asp Arg Val Gly Ser Glu Leu Lys Ala Met Val Gln Ala
865 870 875 880

Pro Pro Gly Tyr Thr Leu Val Gly Ala Asp Val Asp Ser Gln Glu Leu
885 890 895

Trp Ile Ala Ala Val Leu Gly Asp Ala His Phe Ala Gly Met His Gly
900 905 910

Cys Thr Ala Phe Gly Trp Met Thr Leu Gln Gly Arg Lys Ser Arg Gly
915 920 925

Thr Asp Leu His Ser Lys Thr Ala Thr Thr Val Gly Ile Ser Arg Glu
930 935 940

His Ala Lys Ile Phe Asn Tyr Gly Arg Ile Tyr Gly Ala Gly Gln Pro
945 950 955 960

Phe Ala Glu Arg Leu Leu Met Gln Phe Asn His Arg Leu Thr Gln Gln

965	970	975
Glu Ala Ala Glu Lys Ala Gln Gln Met Tyr Ala Ala Thr Lys Gly Leu		
980	985	990
Arg Trp Tyr Arg Leu Ser Asp Glu Gly Glu Trp Leu Val Arg Glu Leu		
995	1000	1005
Asn Leu Pro Val Asp Arg Thr Glu Gly Gly Trp Ile Ser Leu Gln		
1010	1015	1020
Asp Leu Arg Lys Val Gln Arg Glu Thr Ala Arg Lys Ser Gln Trp		
1025	1030	1035
Lys Lys Trp Glu Val Val Ala Glu Arg Ala Trp Lys Gly Gly Thr		
1040	1045	1050
Glu Ser Glu Met Phe Asn Lys Leu Glu Ser Ile Ala Thr Ser Asp		
1055	1060	1065
Ile Pro Arg Thr Pro Val Leu Gly Cys Cys Ile Ser Arg Ala Leu		
1070	1075	1080
Glu Pro Ser Ala Val Gln Glu Glu Phe Met Thr Ser Arg Val Asn		
1085	1090	1095
Trp Val Val Gln Ser Ser Ala Val Asp Tyr Leu His Leu Met Leu		
1100	1105	1110
Val Ala Met Lys Trp Leu Phe Glu Glu Phe Ala Ile Asp Gly Arg		
1115	1120	1125
Phe Cys Ile Ser Ile His Asp Glu Val Arg Tyr Leu Val Arg Glu		
1130	1135	1140
Glu Asp Arg Tyr Arg Ala Ala Leu Ala Leu Gln Ile Thr Asn Leu		
1145	1150	1155
Leu Thr Arg Cys Met Phe Ala Tyr Lys Leu Gly Leu Asn Asp Leu		
1160	1165	1170
Pro Gln Ser Val Ala Phe Phe Ser Ala Val Asp Ile Asp Arg Cys		
1175	1180	1185
Leu Arg Lys Glu Val Thr Met Asp Cys Lys Thr Pro Ser Asn Pro		
1190	1195	1200

Thr Gly Met Glu Arg Arg Tyr Gly Ile Pro Gln Gly Glu Ala Leu

1205

1210

1215

Asp Ile Tyr Gln Ile Ile Glu Leu Thr Lys Gly Ser Leu Glu Lys

1220

1225

1230

Arg Ser Gln Pro Gly Pro

1235

<210> 63

<211> 900

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 63

Met Glu Asn Tyr Glu Ala Leu Val Gly Phe Asp Leu Cys Asn Thr Pro

1 5 10 15

Leu Ser Ser Val Ala Gln Lys Ile Met Ser Ala Met His Ser Gly Asp

20

25

30

Leu Val Asp Ser Lys Thr Trp Gly Lys Ser Thr Glu Thr Met Glu Val

35

40

45

Ile Asn Lys Ser Ser Val Lys Tyr Ser Val Gln Leu Glu Asp Arg Lys

50

55

60

Thr Gln Ser Pro Glu Lys Lys Asp Leu Lys Ser Leu Arg Ser Gln Thr

65

70

75

80

Ser Arg Gly Ser Ala Lys Leu Ser Pro Gln Ser Phe Ser Val Arg Leu

85

90

95

Thr Asp Gln Leu Ser Ala Asp Gln Lys Gln Lys Ser Ile Ser Ser Leu

100

105

110

Thr Leu Ser Ser Cys Leu Ile Pro Gln Tyr Asn Gln Glu Ala Ser Val

115

120

125

Leu Gln Lys Lys Gly His Lys Arg Lys His Phe Leu Met Glu Asn Ile

130

135

140

Asn Asn Glu Asn Lys Gly Ser Ile Asn Leu Lys Arg Lys His Ile Thr

145 150 155 160
 Tyr Asn Asn Leu Ser Glu Lys Thr Ser Lys Gln Met Ala Leu Glu Glu
 165 170 175
 Asp Thr Asp Asp Ala Glu Gly Tyr Leu Asn Ser Gly Asn Ser Gly Ala
 180 185 190
 Leu Lys Lys His Phe Cys Asp Ile Arg His Leu Asp Asp Trp Ala Lys
 195 200 205
 Ser Gln Leu Ile Glu Met Leu Lys Gln Ala Ala Ala Leu Val Ile Thr

 210 215 220
 Val Met Tyr Thr Asp Gly Ser Thr Gln Leu Gly Ala Asp Gln Thr Pro
 225 230 235 240
 Val Ser Ser Val Arg Gly Ile Val Val Leu Val Lys Arg Gln Ala Glu
 245 250 255
 Gly Gly His Gly Cys Pro Asp Ala Pro Ala Cys Gly Pro Val Leu Glu
 260 265 270
 Gly Phe Val Ser Asp Asp Pro Cys Ile Tyr Ile Gln Ile Glu His Ser

 275 280 285
 Ala Ile Trp Asp Gln Glu Gln Glu Ala His Gln Gln Phe Ala Arg Asn
 290 295 300
 Val Leu Phe Gln Thr Met Lys Cys Lys Cys Pro Val Ile Cys Phe Asn
 305 310 315 320
 Ala Lys Asp Phe Val Arg Ile Val Leu Gln Phe Phe Gly Asn Asp Gly
 325 330 335
 Ser Trp Lys His Val Ala Asp Phe Ile Gly Leu Asp Pro Arg Ile Ala

 340 345 350
 Ala Trp Leu Ile Asp Pro Ser Asp Ala Thr Pro Ser Phe Glu Asp Leu
 355 360 365
 Val Glu Lys Tyr Cys Glu Lys Ser Ile Thr Val Lys Val Asn Ser Thr
 370 375 380
 Tyr Gly Asn Ser Ser Arg Asn Ile Val Asn Gln Asn Val Arg Glu Asn
 385 390 395 400

Leu Lys Thr Leu Tyr Arg Leu Thr Met Asp Leu Cys Ser Lys Leu Lys

405 410 415

Asp Tyr Gly Leu Trp Gln Leu Phe Arg Thr Leu Glu Leu Pro Leu Ile

420 425 430

Pro Ile Leu Ala Val Met Glu Ser His Ala Ile Gln Val Asn Lys Glu

435 440 445

Glu Met Glu Lys Thr Ser Ala Leu Leu Gly Ala Arg Leu Lys Glu Leu

450 455 460

Glu Gln Glu Ala His Phe Val Ala Gly Glu Arg Phe Leu Ile Thr Ser

465 470 475 480

Asn Asn Gln Leu Arg Glu Ile Leu Phe Gly Lys Leu Lys Leu His Leu

485 490 495

Leu Ser Gln Arg Asn Ser Leu Pro Arg Thr Gly Leu Gln Lys Tyr Pro

500 505 510

Ser Thr Ser Glu Ala Val Leu Asn Ala Leu Arg Asp Leu His Pro Leu

515 520 525

Pro Lys Ile Ile Leu Glu Tyr Arg Gln Val His Lys Ile Lys Ser Thr

530 535 540

Phe Val Asp Gly Leu Leu Ala Cys Met Lys Lys Gly Ser Ile Ser Ser

545 550 555 560

Thr Trp Asn Gln Thr Gly Thr Val Thr Gly Arg Leu Ser Ala Lys His

565 570 575

Pro Asn Ile Gln Gly Ile Ser Lys His Pro Ile Gln Ile Thr Thr Pro

580 585 590

Lys Asn Phe Lys Gly Lys Glu Asp Lys Ile Leu Thr Ile Ser Pro Arg

595 600 605

Ala Met Phe Val Ser Ser Lys Gly His Thr Phe Leu Ala Ala Asp Phe

610 615 620

Ser Gln Ile Glu Leu Arg Ile Leu Thr His Leu Ser Gly Asp Pro Glu

625 630 635 640

Leu Leu Lys Leu Phe Gln Glu Ser Glu Arg Asp Asp Val Phe Ser Thr

645 650 655
 Leu Thr Ser Gln Trp Lys Asp Val Pro Val Glu Gln Val Thr His Ala

 660 665 670
 Asp Arg Glu Gln Thr Lys Lys Val Val Tyr Ala Val Val Tyr Gly Ala
 675 680 685
 Gly Lys Glu Arg Leu Ala Ala Cys Leu Gly Val Pro Ile Gln Glu Ala
 690 695 700
 Ala Gln Phe Leu Glu Ser Phe Leu Gln Lys Tyr Lys Lys Ile Lys Asp
 705 710 715 720
 Phe Ala Arg Ala Ala Ile Ala Gln Cys His Gln Thr Gly Cys Val Val

 725 730 735
 Ser Ile Met Gly Arg Arg Arg Pro Leu Pro Arg Ile His Ala His Asp
 740 745 750
 Gln Gln Leu Arg Ala Gln Ala Glu Arg Gln Ala Val Asn Phe Val Val
 755 760 765
 Gln Gly Ser Ala Ala Asp Leu Cys Lys Leu Ala Met Ile His Val Phe
 770 775 780
 Thr Ala Val Ala Ala Ser His Thr Leu Thr Ala Arg Leu Val Ala Gln

 785 790 795 800
 Ile His Asp Glu Leu Leu Phe Glu Val Glu Asp Pro Gln Ile Pro Glu
 805 810 815
 Cys Ala Ala Leu Val Arg Arg Thr Met Glu Ser Leu Glu Gln Val Gln
 820 825 830
 Ala Leu Glu Leu Gln Leu Gln Val Pro Leu Lys Val Ser Leu Ser Ala
 835 840 845
 Gly Arg Ser Trp Gly His Leu Val Pro Leu Gln Glu Ala Trp Gly Pro

 850 855 860
 Pro Pro Gly Pro Cys Arg Thr Glu Ser Pro Ser Asn Ser Leu Ala Ala
 865 870 875 880
 Pro Gly Ser Pro Ala Ser Thr Gln Pro Pro Pro Leu His Phe Ser Pro
 885 890 895

Ser Phe Cys Leu

900

<210> 64

<211> 1251

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 64

Met Arg Arg Gly Gly Trp Arg Lys Arg Ala Glu Asn Asp Gly Trp Glu

1 5 10 15
Thr Trp Gly Gly Tyr Met Ala Ala Lys Val Gln Lys Leu Glu Glu Gln

20 25 30
Phe Arg Ser Asp Ala Ala Met Gln Lys Asp Gly Thr Ser Ser Thr Ile

35 40 45
Phe Ser Gly Val Ala Ile Tyr Val Asn Gly Tyr Thr Asp Pro Ser Ala

50 55 60
Glu Glu Leu Arg Lys Leu Met Met Leu His Gly Gly Gln Tyr His Val

65 70 75 80
Tyr Tyr Ser Arg Ser Lys Thr Thr His Ile Ile Ala Thr Asn Leu Pro

85 90 95
Asn Ala Lys Ile Lys Glu Leu Lys Gly Glu Lys Val Ile Arg Pro Glu

100 105 110
Trp Ile Val Glu Ser Ile Lys Ala Gly Arg Leu Leu Ser Tyr Ile Pro

115 120 125
Tyr Gln Leu Tyr Thr Lys Gln Ser Ser Val Gln Lys Gly Leu Ser Phe

130 135 140
Asn Pro Val Cys Arg Pro Glu Asp Pro Leu Pro Gly Pro Ser Asn Ile

145 150 155 160
Ala Lys Gln Leu Asn Asn Arg Val Asn His Ile Val Lys Lys Ile Glu

165 170 175
Thr Glu Asn Glu Val Lys Val Asn Gly Met Asn Ser Trp Asn Glu Glu

180 185 190

Asp Glu Asn Asn Asp Phe Ser Phe Val Asp Leu Glu Gln Thr Ser Pro

195 200 205

Gly Arg Lys Gln Asn Gly Ile Pro His Pro Arg Gly Ser Thr Ala Ile

210 215 220

Phe Asn Gly His Thr Pro Ser Ser Asn Gly Ala Leu Lys Thr Gln Asp

225 230 235 240

Cys Leu Val Pro Met Val Asn Ser Val Ala Ser Arg Leu Ser Pro Ala

245 250 255

Phe Ser Gln Glu Glu Asp Lys Ala Glu Lys Ser Ser Thr Asp Phe Arg

260 265 270

Asp Cys Thr Leu Gln Gln Leu Gln Gln Ser Thr Arg Asn Thr Asp Ala

275 280 285

Leu Arg Asn Pro His Arg Thr Asn Ser Phe Ser Leu Ser Pro Leu His

290 295 300

Ser Asn Thr Lys Ile Asn Gly Ala His His Ser Thr Val Gln Gly Pro

305 310 315 320

Ser Ser Thr Lys Ser Thr Ser Ser Val Ser Thr Phe Ser Lys Ala Ala

325 330 335

Pro Ser Val Pro Ser Lys Pro Ser Asp Cys Asn Phe Ile Ser Asn Phe

340 345 350

Tyr Ser His Ser Arg Leu His His Ile Ser Met Trp Lys Cys Glu Leu

355 360 365

Thr Glu Phe Val Asn Thr Leu Gln Arg Gln Ser Asn Gly Ile Phe Pro

370 375 380

Gly Arg Glu Lys Leu Lys Lys Met Lys Thr Gly Arg Ser Ala Leu Val

385 390 395 400

Val Thr Asp Thr Gly Asp Met Ser Val Leu Asn Ser Pro Arg His Gln

405 410 415

Ser Cys Ile Met His Val Asp Met Asp Cys Phe Phe Val Ser Val Gly

420 425 430

Ile Arg Asn Arg Pro Asp Leu Lys Gly Lys Pro Val Ala Val Thr Ser

435 440 445
Asn Arg Gly Thr Gly Arg Ala Pro Leu Arg Pro Gly Ala Asn Pro Gln

450 455 460
Leu Glu Trp Gln Tyr Tyr Gln Asn Lys Ile Leu Lys Gly Lys Ala Ala

465 470 475 480
Asp Ile Pro Asp Ser Ser Leu Trp Glu Asn Pro Asp Ser Ala Gln Ala

485 490 495
Asn Gly Ile Asp Ser Val Leu Ser Arg Ala Glu Ile Ala Ser Cys Ser

500 505 510
Tyr Glu Ala Arg Gln Leu Gly Ile Lys Asn Gly Met Phe Phe Gly His

515 520 525
Ala Lys Gln Leu Cys Pro Asn Leu Gln Ala Val Pro Tyr Asp Phe His

530 535 540
Ala Tyr Lys Glu Val Ala Gln Thr Leu Tyr Glu Thr Leu Ala Ser Tyr

545 550 555 560
Thr His Asn Ile Glu Ala Val Ser Cys Asp Glu Ala Leu Val Asp Ile

565 570 575
Thr Glu Ile Leu Ala Glu Thr Lys Leu Thr Pro Asp Glu Phe Ala Asn

580 585 590
Ala Val Arg Met Glu Ile Lys Asp Gln Thr Lys Cys Ala Ala Ser Val

595 600 605
Gly Ile Gly Ser Asn Ile Leu Leu Ala Arg Met Ala Thr Arg Lys Ala

610 615 620
Lys Pro Asp Gly Gln Tyr His Leu Lys Pro Glu Glu Val Asp Asp Phe

625 630 635 640
Ile Arg Gly Gln Leu Val Thr Asn Leu Pro Gly Val Gly His Ser Met

645 650 655
Glu Ser Lys Leu Ala Ser Leu Gly Ile Lys Thr Cys Gly Asp Leu Gln

660 665 670
Tyr Met Thr Met Ala Lys Leu Gln Lys Glu Phe Gly Pro Lys Thr Gly

675 680 685

Gln Met Leu Tyr Arg Phe Cys Arg Gly Leu Asp Asp Arg Pro Val Arg
690 695 700

Thr Glu Lys Glu Arg Lys Ser Val Ser Ala Glu Ile Asn Tyr Gly Ile

705 710 715 720

Arg Phe Thr Gln Pro Lys Glu Ala Glu Ala Phe Leu Leu Ser Leu Ser
725 730 735

Glu Glu Ile Gln Arg Arg Leu Glu Ala Thr Gly Met Lys Gly Lys Arg
740 745 750

Leu Thr Leu Lys Ile Met Val Arg Lys Pro Gly Ala Pro Val Glu Thr
755 760 765

Ala Lys Phe Gly Gly His Gly Ile Cys Asp Asn Ile Ala Arg Thr Val

770 775 780

Thr Leu Asp Gln Ala Thr Asp Asn Ala Lys Ile Ile Gly Lys Ala Met
785 790 795 800

Leu Asn Met Phe His Thr Met Lys Leu Asn Ile Ser Asp Met Arg Gly
805 810 815

Val Gly Ile His Val Asn Gln Leu Val Pro Thr Asn Leu Asn Pro Ser
820 825 830

Thr Cys Pro Ser Arg Pro Ser Val Gln Ser Ser His Phe Pro Ser Gly

835 840 845

Ser Tyr Ser Val Arg Asp Val Phe Gln Val Gln Lys Ala Lys Lys Ser
850 855 860

Thr Glu Glu Glu His Lys Glu Val Phe Arg Ala Ala Val Asp Leu Glu
865 870 875 880

Ile Ser Ser Ala Ser Arg Thr Cys Thr Phe Leu Pro Pro Phe Pro Ala
885 890 895

His Leu Pro Thr Ser Pro Asp Thr Asn Lys Ala Glu Ser Ser Gly Lys

900 905 910

Trp Asn Gly Leu His Thr Pro Val Ser Val Gln Ser Arg Leu Asn Leu
915 920 925

Ser Ile Glu Val Pro Ser Pro Ser Gln Leu Asp Gln Ser Val Leu Glu

930	935	940	
Ala Leu Pro Pro Asp Leu Arg Glu Gln Val Glu Gln Val Cys Ala Val			
945	950	955	960
Gln Gln Ala Glu Ser His Gly Asp Lys Lys Lys Glu Pro Val Asn Gly			
	965	970	975
Cys Asn Thr Gly Ile Leu Pro Gln Pro Val Gly Thr Val Leu Leu Gln			
	980	985	990
Ile Pro Glu Pro Gln Glu Ser Asn Ser Asp Ala Gly Ile Asn Leu Ile			
	995	1000	1005
Ala Leu Pro Ala Phe Ser Gln Val Asp Pro Glu Val Phe Ala Ala			
1010	1015	1020	
Leu Pro Ala Glu Leu Gln Arg Glu Leu Lys Ala Ala Tyr Asp Gln			
	1025	1030	1035
Arg Gln Arg Gln Gly Glu Asn Ser Thr His Gln Gln Ser Ala Ser			
1040	1045	1050	
Ala Ser Val Pro Lys Asn Pro Leu Leu His Leu Lys Ala Ala Val			
1055	1060	1065	
Lys Glu Lys Lys Arg Asn Lys Lys Lys Lys Thr Ile Gly Ser Pro			
1070	1075	1080	
Lys Arg Ile Gln Ser Pro Leu Asn Asn Lys Leu Leu Asn Ser Pro			
	1085	1090	1095
Ala Lys Thr Leu Pro Gly Ala Cys Gly Ser Pro Gln Lys Leu Ile			
1100	1105	1110	
Asp Gly Phe Leu Lys His Glu Gly Pro Pro Ala Glu Lys Pro Leu			
1115	1120	1125	
Glu Glu Leu Ser Ala Ser Thr Ser Gly Val Pro Gly Leu Ser Ser			
1130	1135	1140	
Leu Gln Ser Asp Pro Ala Gly Cys Val Arg Pro Pro Ala Pro Asn			
	1145	1150	1155
Leu Ala Gly Ala Val Glu Phe Asn Asp Val Lys Thr Leu Leu Arg			
1160	1165	1170	

Glu Trp Ile Thr Thr Ile Ser Asp Pro Met Glu Glu Asp Ile Leu
1175 1180 1185

Gln Val Val Lys Tyr Cys Thr Asp Leu Ile Glu Glu Lys Asp Leu
1190 1195 1200

Glu Lys Leu Asp Leu Val Ile Lys Tyr Met Lys Arg Leu Met Gln

1205 1210 1215

Gln Ser Val Glu Ser Val Trp Asn Met Ala Phe Asp Phe Ile Leu
1220 1225 1230

Asp Asn Val Gln Val Val Leu Gln Gln Thr Tyr Gly Ser Thr Leu
1235 1240 1245

Lys Val Thr
1250

<210> 65

<211> 313

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 65

Met Ile Gly Gln Lys Thr Leu Tyr Ser Phe Phe Ser Pro Ser Pro Ala

1 5 10 15
Arg Lys Arg His Ala Pro Ser Pro Glu Pro Ala Val Gln Gly Thr Gly

20 25 30
Val Ala Gly Val Pro Glu Glu Ser Gly Asp Ala Ala Ala Ile Pro Ala

35 40 45
Lys Lys Ala Pro Ala Gly Gln Glu Glu Pro Gly Thr Pro Pro Ser Ser

50 55 60
Pro Leu Ser Ala Glu Gln Leu Asp Arg Ile Gln Arg Asn Lys Ala Ala

65 70 75 80
Ala Leu Leu Arg Leu Ala Ala Arg Asn Val Pro Val Gly Phe Gly Glu

85 90 95
Ser Trp Lys Lys His Leu Ser Gly Glu Phe Gly Lys Pro Tyr Phe Ile

100 105 110

Lys Leu Met Gly Phe Val Ala Glu Glu Arg Lys His Tyr Thr Val Tyr
 115 120 125
 Pro Pro Pro His Gln Val Phe Thr Trp Thr Gln Met Cys Asp Ile Lys
 130 135 140
 Asp Val Lys Val Val Ile Leu Gly Gln Asp Pro Ala His Gly Pro Asn
 145 150 155 160
 Gln Ala His Gly Leu Cys Phe Ser Val Gln Arg Pro Val Pro Pro Pro
 165 170 175
 Pro Ser Leu Glu Asn Ile Tyr Lys Glu Leu Ser Thr Asp Ile Glu Asp
 180 185 190
 Phe Val His Pro Gly His Gly Asp Leu Ser Gly Trp Ala Lys Gln Gly
 195 200 205
 Val Leu Leu Leu Asn Ala Val Leu Thr Val Arg Ala His Gln Ala Asn
 210 215 220
 Ser His Lys Glu Arg Gly Trp Glu Gln Phe Thr Asp Ala Val Val Ser
 225 230 235 240
 Trp Leu Asn Gln Asn Ser Asn Gly Leu Val Phe Leu Leu Trp Gly Ser
 245 250 255
 Tyr Ala Gln Lys Lys Gly Ser Ala Ile Asp Arg Lys Arg His His Val
 260 265 270
 Leu Gln Thr Ala His Pro Ser Pro Leu Ser Val Tyr Arg Gly Phe Phe
 275 280 285
 Gly Cys Arg His Phe Ser Lys Thr Asn Glu Leu Leu Gln Lys Ser Gly
 290 295 300
 Lys Lys Pro Ile Asp Trp Lys Glu Leu
 305 310
 <210> 66
 <211> 313
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic polypeptide

<400> 66

Met Ile Gly Gln Lys Thr Leu Tyr Ser Phe Phe Ser Pro Ser Pro Ala
1 5 10 15

Arg Lys Arg His Ala Pro Ser Pro Glu Pro Ala Val Gln Gly Thr Gly
20 25 30

Val Ala Gly Val Pro Glu Glu Ser Gly Asp Ala Ala Ala Ile Pro Ala
35 40 45

Lys Lys Ala Pro Ala Gly Gln Glu Glu Pro Gly Thr Pro Pro Ser Ser
50 55 60

Pro Leu Ser Ala Glu Gln Leu Asp Arg Ile Gln Arg Asn Lys Ala Ala
65 70 75 80

Ala Leu Leu Arg Leu Ala Ala Arg Asn Val Pro Val Gly Phe Gly Glu
85 90 95

Ser Trp Lys Lys His Leu Ser Gly Glu Phe Gly Lys Pro Tyr Phe Ile
100 105 110

Lys Leu Met Gly Phe Val Ala Glu Glu Arg Lys His Tyr Thr Val Tyr
115 120 125

Pro Pro Pro His Gln Val Phe Thr Trp Thr Gln Met Cys Asp Ile Lys
130 135 140

Asp Val Lys Val Val Ile Leu Gly Gln Asp Pro Tyr His Gly Pro Asn
145 150 155 160

Gln Ala His Gly Leu Cys Phe Ser Val Gln Arg Pro Val Pro Pro Pro
165 170 175

Pro Ser Leu Glu Asn Ile Tyr Lys Glu Leu Ser Thr Asp Ile Glu Asp
180 185 190

Phe Val His Pro Gly His Gly Asp Leu Ser Gly Trp Ala Lys Gln Gly
195 200 205

Val Leu Leu Leu Asp Ala Val Leu Thr Val Arg Ala His Gln Ala Asn
210 215 220

Ser His Lys Glu Arg Gly Trp Glu Gln Phe Thr Asp Ala Val Val Ser
225 230 235 240

Trp Leu Asn Gln Asn Ser Asn Gly Leu Val Phe Leu Leu Trp Gly Ser

245										250										255																			
Tyr	Ala	Gln	Lys	Lys	Gly	Ser	Ala	Ile	Asp	Arg	Lys	Arg	His	His	Val																								
260										265										270																			
Leu	Gln	Thr	Ala	His	Pro	Ser	Pro	Leu	Ser	Val	Tyr	Arg	Gly	Phe	Phe																								
275										280										285																			
Gly	Cys	Arg	His	Phe	Ser	Lys	Thr	Asn	Glu	Leu	Leu	Gln	Lys	Ser	Gly																								
290										295										300																			
Lys	Lys	Pro	Ile	Asp	Trp	Lys	Glu	Leu																															
305										310																													
<210>										67																													
<211>										198																													
<212>										PRT																													
<213>										Homo sapiens																													
<400>										67																													
Met	Asp	Ser	Leu	Leu	Met	Asn	Arg	Arg	Lys	Phe	Leu	Tyr	Gln	Phe	Lys																								
1										5										10										15									
Asn	Val	Arg	Trp	Ala	Lys	Gly	Arg	Arg	Glu	Thr	Tyr	Leu	Cys	Tyr	Val																								
20										25										30																			
Val	Lys	Arg	Arg	Asp	Ser	Ala	Thr	Ser	Phe	Ser	Leu	Asp	Phe	Gly	Tyr																								
35										40										45																			
Leu	Arg	Asn	Lys	Asn	Gly	Cys	His	Val	Glu	Leu	Leu	Phe	Leu	Arg	Tyr																								
50										55										60																			
Ile	Ser	Asp	Trp	Asp	Leu	Asp	Pro	Gly	Arg	Cys	Tyr	Arg	Val	Thr	Trp																								
65										70										75										80									
Phe	Thr	Ser	Trp	Ser	Pro	Cys	Tyr	Asp	Cys	Ala	Arg	His	Val	Ala	Asp																								
85										90										95																			
Phe	Leu	Arg	Gly	Asn	Pro	Asn	Leu	Ser	Leu	Arg	Ile	Phe	Thr	Ala	Arg																								
100										105										110																			
Leu	Tyr	Phe	Cys	Glu	Asp	Arg	Lys	Ala	Glu	Pro	Glu	Gly	Leu	Arg	Arg																								
115										120										125																			
Leu	His	Arg	Ala	Gly	Val	Gln	Ile	Ala	Ile	Met	Thr	Phe	Lys	Asp	Tyr																								

130 135 140
Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys
145 150 155 160
Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
165 170 175
Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190

Phe Arg Thr Leu Gly Leu

195
<210> 68
<211> 198
<212> PRT
<213> Mus musculus
<400> 68

Met Asp Ser Leu Leu Met Lys Gln Lys Lys Phe Leu Tyr His Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Cys Ser Leu Asp Phe Gly His
35 40 45

Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Glu
85 90 95
Phe Leu Arg Trp Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125
Leu His Arg Ala Gly Val Gln Ile Gly Ile Met Thr Phe Lys Asp Tyr
130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Arg Thr Phe Lys
 145 150 155 160
 Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Thr Arg Gln Leu
 165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
 180 185 190

Phe Arg Met Leu Gly Phe
 195

<210> 69

<211> 197

<212> PRT

<213> Canis lupus familiaris

<400> 69

Met Asp Ser Leu Leu Met Lys Gln Arg Lys Phe Leu Tyr His Phe Lys
 1 5 10 15
 Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
 20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly His
 35 40 45

Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
 50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
 65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
 85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Ala Ala Arg
 100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
 115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr
 130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Lys Thr Phe Lys

145 150 155 160
 Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
 165 170 175
 Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
 180 185 190
 Phe Arg Thr Leu Gly
 195
 <210> 70
 <211> 198
 <212> PRT
 <213> Bos taurus
 <400> 70
 Met Asp Ser Leu Leu Lys Lys Gln Arg Gln Phe Leu Tyr Gln Phe Lys
 1 5 10 15

 Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
 20 25 30
 Val Lys Arg Arg Asp Ser Pro Thr Ser Phe Ser Leu Asp Phe Gly His
 35 40 45
 Leu Arg Asn Lys Ala Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
 50 55 60
 Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
 65 70 75 80

 Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
 85 90 95
 Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
 100 105 110
 Leu Tyr Phe Cys Asp Lys Glu Arg Lys Ala Glu Pro Glu Gly Leu Arg
 115 120 125
 Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp
 130 135 140

 Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe

145 150 155 160
 Lys Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln
 165 170 175
 Leu Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp
 180 185 190
 Ala Phe Arg Thr Leu Gly
 195
 <210> 71
 <211> 238
 <212> PRT
 <213> Rattus norvegicus

 <400> 71
 Met Ala Val Gly Ser Lys Pro Lys Ala Ala Leu Val Gly Pro His Trp
 1 5 10 15
 Glu Arg Glu Arg Ile Trp Cys Phe Leu Cys Ser Thr Gly Leu Gly Thr
 20 25 30
 Gln Gln Thr Gly Gln Thr Ser Arg Trp Leu Arg Pro Ala Ala Thr Gln
 35 40 45
 Asp Pro Val Ser Pro Pro Arg Ser Leu Leu Met Lys Gln Arg Lys Phe
 50 55 60

 Leu Tyr His Phe Lys Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr
 65 70 75 80
 Tyr Leu Cys Tyr Val Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser
 85 90 95
 Leu Asp Phe Gly Tyr Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu
 100 105 110
 Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys
 115 120 125

 Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala
 130 135 140
 Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg
 145 150 155 160

Ile Phe Thr Ala Arg Leu Thr Gly Trp Gly Ala Leu Pro Ala Gly Leu
165 170 175
Met Ser Pro Ala Arg Pro Ser Asp Tyr Phe Tyr Cys Trp Asn Thr Phe
180 185 190
Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu
195 200 205
Asn Ser Val Arg Leu Ser Arg Arg Leu Arg Arg Ile Leu Leu Pro Leu
210 215 220
Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu Gly
225 230 235
<210> 72
<211> 429
<212> PRT
<213> Mus musculus
<400> 72
Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro
1 5 10 15
Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn
20 25 30
Leu Gly Tyr Ala Lys Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val
35 40 45
Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val Phe
50 55 60
Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe
65 70 75 80
His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile
85 90 95
Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile
100 105 110
Val Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe Ser
115 120 125

Ser Arg Leu Tyr Asn Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys

130 135 140

Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu

145 150 155 160

Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe

165 170 175

Arg Pro Trp Lys Arg Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys

180 185 190

Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser

195 200 205

Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr

210 215 220

Arg Phe Cys Val Glu Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu

225 230 235 240

Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr

245 250 255

His Arg Met Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly

260 265 270

Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His

275 280 285

Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln

290 295 300

Val Thr Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala

305 310 315 320

Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His

325 330 335

Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys

340 345 350

Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp

355 360 365

Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg

370 375 380
Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln

385 390 395 400
Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val

405 410 415
Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser

420 425

<210> 73

<211> 429

<212> PRT

<213> Rattus norvegicus

<400> 73

Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro

1 5 10 15

Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn

20 25 30

Leu Arg Tyr Ala Ile Asp Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val

35 40 45

Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val Phe

50 55 60

Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe

65 70 75 80

His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile

85 90 95

Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Val

100 105 110

Leu Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe Ser

115 120 125

Ser Arg Leu Tyr Asn Ile Arg Asp Pro Glu Asn Gln Gln Asn Leu Cys

130 135 140

Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu

145 150 155 160
 Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe
 165 170 175
 Arg Pro Trp Lys Lys Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys
 180 185 190
 Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser
 195 200 205

 Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr
 210 215 220
 Arg Phe Cys Val Glu Arg Arg Arg Val His Leu Leu Ser Glu Glu Glu
 225 230 235 240
 Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr
 245 250 255
 His Gly Val Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly
 260 265 270

 Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His
 275 280 285
 Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln
 290 295 300
 Val Ile Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
 305 310 315 320
 Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His
 325 330 335

 Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys
 340 345 350
 Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp
 355 360 365
 Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg
 370 375 380
 Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln
 385 390 395 400

Arg Arg Leu His Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val

405 410 415

Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser

420 425

<210> 74

<211> 370

<212> PRT

<213> Macaca mulatta

<400> 74

Met Val Glu Pro Met Asp Pro Arg Thr Phe Val Ser Asn Phe Asn Asn

1 5 10 15

Arg Pro Ile Leu Ser Gly Leu Asn Thr Val Trp Leu Cys Cys Glu Val

20 25 30

Lys Thr Lys Asp Pro Ser Gly Pro Pro Leu Asp Ala Lys Ile Phe Gln

35 40 45

Gly Lys Val Tyr Ser Lys Ala Lys Tyr His Pro Glu Met Arg Phe Leu

50 55 60

Arg Trp Phe His Lys Trp Arg Gln Leu His His Asp Gln Glu Tyr Lys

65 70 75 80

Val Thr Trp Tyr Val Ser Trp Ser Pro Cys Thr Arg Cys Ala Asn Ser

85 90 95

Val Ala Thr Phe Leu Ala Lys Asp Pro Lys Val Thr Leu Thr Ile Phe

100 105 110

Val Ala Arg Leu Tyr Tyr Phe Trp Lys Pro Asp Tyr Gln Gln Ala Leu

115 120 125

Arg Ile Leu Cys Gln Lys Arg Gly Gly Pro His Ala Thr Met Lys Ile

130 135 140

Met Asn Tyr Asn Glu Phe Gln Asp Cys Trp Asn Lys Phe Val Asp Gly

145 150 155 160

Arg Gly Lys Pro Phe Lys Pro Arg Asn Asn Leu Pro Lys His Tyr Thr

165 170 175

Leu Leu Gln Ala Thr Leu Gly Glu Leu Leu Arg His Leu Met Asp Pro

180 185 190
Gly Thr Phe Thr Ser Asn Phe Asn Asn Lys Pro Trp Val Ser Gly Gln

195 200 205
His Glu Thr Tyr Leu Cys Tyr Lys Val Glu Arg Leu His Asn Asp Thr

210 215 220
Trp Val Pro Leu Asn Gln His Arg Gly Phe Leu Arg Asn Gln Ala Pro

225 230 235 240
Asn Ile His Gly Phe Pro Lys Gly Arg His Ala Glu Leu Cys Phe Leu

245 250 255
Asp Leu Ile Pro Phe Trp Lys Leu Asp Gly Gln Gln Tyr Arg Val Thr

260 265 270
Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala

275 280 285
Lys Phe Ile Ser Asn Asn Glu His Val Ser Leu Cys Ile Phe Ala Ala

290 295 300
Arg Ile Tyr Asp Asp Gln Gly Arg Tyr Gln Glu Gly Leu Arg Ala Leu

305 310 315 320
His Arg Asp Gly Ala Lys Ile Ala Met Met Asn Tyr Ser Glu Phe Glu

325 330 335
Tyr Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe Gln Pro

340 345 350
Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg

355 360 365
Ala Ile

370

<210> 75

<211> 384

<212> PRT

<213> Pan troglodytes

<400> 75

Met Lys Pro His Phe Arg Asn Pro Val Glu Arg Met Tyr Gln Asp Thr

1 5 10 15

Phe Ser Asp Asn Phe Tyr Asn Arg Pro Ile Leu Ser His Arg Asn Thr
20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Lys Leu Lys Tyr
50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
85 90 95

Cys Thr Lys Cys Thr Arg Asp Val Ala Thr Phe Leu Ala Glu Asp Pro
100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
130 135 140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145 150 155 160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
165 170 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
180 185 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Ser Asn Phe Asn Asn
195 200 205

Glu Leu Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
210 215 220

Glu Arg Leu His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
245 250 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp

260 265 270
 Leu His Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
 275 280 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Asn Asn Lys His
 290 295 300
 Val Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg
 305 310 315 320
 Cys Gln Glu Gly Leu Arg Thr Leu Ala Lys Ala Gly Ala Lys Ile Ser
 325 330 335
 Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
 340 345 350

His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser
 355 360 365
 Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
 370 375 380

<210> 76

<211> 377

<212> PRT

<213> Chlorocebus sabaeus

<400> 76

Met Asn Pro Gln Ile Arg Asn Met Val Glu Gln Met Glu Pro Asp Ile
 1 5 10 15
 Phe Val Tyr Tyr Phe Asn Asn Arg Pro Ile Leu Ser Gly Arg Asn Thr

20 25 30
 Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Asp Pro Ser Gly Pro Pro
 35 40 45
 Leu Asp Ala Asn Ile Phe Gln Gly Lys Leu Tyr Pro Glu Ala Lys Asp
 50 55 60
 His Pro Glu Met Lys Phe Leu His Trp Phe Arg Lys Trp Arg Gln Leu
 65 70 75 80
 His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Val Ser Trp Ser Pro

85	90	95	
Cys Thr Arg Cys Ala Asn Ser Val Ala Thr Phe Leu Ala Glu Asp Pro			
100	105	110	
Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Lys			
115	120	125	
Pro Asp Tyr Gln Gln Ala Leu Arg Ile Leu Cys Gln Glu Arg Gly Gly			
130	135	140	
Pro His Ala Thr Met Lys Ile Met Asn Tyr Asn Glu Phe Gln His Cys			
145	150	155	160
Trp Asn Glu Phe Val Asp Gly Gln Gly Lys Pro Phe Lys Pro Arg Lys			
165	170	175	
Asn Leu Pro Lys His Tyr Thr Leu Leu His Ala Thr Leu Gly Glu Leu			
180	185	190	
Leu Arg His Val Met Asp Pro Gly Thr Phe Thr Ser Asn Phe Asn Asn			
195	200	205	
Lys Pro Trp Val Ser Gly Gln Arg Glu Thr Tyr Leu Cys Tyr Lys Val			
210	215	220	
Glu Arg Ser His Asn Asp Thr Trp Val Leu Leu Asn Gln His Arg Gly			
225	230	235	240
Phe Leu Arg Asn Gln Ala Pro Asp Arg His Gly Phe Pro Lys Gly Arg			
245	250	255	
His Ala Glu Leu Cys Phe Leu Asp Leu Ile Pro Phe Trp Lys Leu Asp			
260	265	270	
Asp Gln Gln Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe			
275	280	285	
Ser Cys Ala Gln Lys Met Ala Lys Phe Ile Ser Asn Asn Lys His Val			
290	295	300	
Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys			
305	310	315	320
Gln Glu Gly Leu Arg Thr Leu His Arg Asp Gly Ala Lys Ile Ala Val			
325	330	335	

Met Asn Tyr Ser Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Asp Arg

340 345 350

Gln Gly Arg Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln

355 360 365

Ala Leu Ser Gly Arg Leu Arg Ala Ile

370 375

<210> 77

<211> 384

<212> PRT

<213> Homo sapiens

<400> 77

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr

20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro

35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr

50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu

65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro

85 90 95

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro

100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp

115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly

130 135 140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys

145 150 155 160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
 165 170 175
 Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
 180 185 190
 Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn
 195 200 205
 Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
 210 215 220
 Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
 225 230 235 240
 Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
 245 250 255
 His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
 260 265 270
 Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
 275 280 285
 Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His
 290 295 300
 Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg
 305 310 315 320
 Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser
 325 330 335
 Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
 340 345 350
 His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser
 355 360 365
 Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
 370 375 380
 <210> 78
 <211> 373
 <212> PRT
 <213> Homo sapiens

<400> 78

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Arg
35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Gln Pro Glu His
50 55 60

His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu Pro
65 70 75 80

Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro Cys
85 90 95

Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ala Glu His Pro Asn
100 105 110

Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg
115 120 125

Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg Val
130 135 140

Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe Val
145 150 155 160

Tyr Ser Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn
165 170 175

Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met
180 185 190

Glu Ala Met Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Arg
195 200 205

Lys Ala Tyr Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val
210 215 220

Val Lys His His Ser Pro Val Ser Trp Lys Arg Gly Val Phe Arg Asn
225 230 235 240

Gln Val Asp Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser
245 250 255

Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr
260 265 270

Trp Tyr Thr Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala
275 280 285

Glu Phe Leu Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala
290 295 300

Arg Leu Tyr Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser
305 310 315 320

Leu Ser Gln Glu Gly Ala Ser Val Glu Ile Met Gly Tyr Lys Asp Phe
325 330 335

Lys Tyr Cys Trp Glu Asn Phe Val Tyr Asn Asp Asp Glu Pro Phe Lys
340 345 350

Pro Trp Lys Gly Leu Lys Tyr Asn Phe Leu Phe Leu Asp Ser Lys Leu
355 360 365

Gln Glu Ile Leu Glu
370

<210> 79

<211> 382

<212> PRT

<213> Homo sapiens

<400> 79

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr
1 5 10 15

Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu
35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Val Tyr Phe Lys Pro Gln
50 55 60

Tyr His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu

65						70						75						80
Pro	Ala	Tyr	Lys	Cys	Phe	Gln	Ile	Thr	Trp	Phe	Val	Ser	Trp	Thr	Pro			
			85						90						95			
Cys	Pro	Asp	Cys	Val	Ala	Lys	Leu	Ala	Glu	Phe	Leu	Ser	Glu	His	Pro			
			100						105						110			
Asn	Val	Thr	Leu	Thr	Ile	Ser	Ala	Ala	Arg	Leu	Tyr	Tyr	Tyr	Trp	Glu			
			115						120						125			
Arg	Asp	Tyr	Arg	Arg	Ala	Leu	Cys	Arg	Leu	Ser	Gln	Ala	Gly	Ala	Arg			
			130						135						140			
Val	Thr	Ile	Met	Asp	Tyr	Glu	Glu	Phe	Ala	Tyr	Cys	Trp	Glu	Asn	Phe			
			145						150						155			
Val	Tyr	Asn	Glu	Gly	Gln	Gln	Phe	Met	Pro	Trp	Tyr	Lys	Phe	Asp	Glu			
			165						170						175			
Asn	Tyr	Ala	Phe	Leu	His	Arg	Thr	Leu	Lys	Glu	Ile	Leu	Arg	Tyr	Leu			
			180						185						190			
Met	Asp	Pro	Asp	Thr	Phe	Thr	Phe	Asn	Phe	Asn	Asn	Asp	Pro	Leu	Val			
			195						200						205			
Leu	Arg	Arg	Arg	Gln	Thr	Tyr	Leu	Cys	Tyr	Glu	Val	Glu	Arg	Leu	Asp			
			210						215						220			
Asn	Gly	Thr	Trp	Val	Leu	Met	Asp	Gln	His	Met	Gly	Phe	Leu	Cys	Asn			
			225						230						235			
Glu	Ala	Lys	Asn	Leu	Leu	Cys	Gly	Phe	Tyr	Gly	Arg	His	Ala	Glu	Leu			
			245						250						255			
Arg	Phe	Leu	Asp	Leu	Val	Pro	Ser	Leu	Gln	Leu	Asp	Pro	Ala	Gln	Ile			
			260						265						270			
Tyr	Arg	Val	Thr	Trp	Phe	Ile	Ser	Trp	Ser	Pro	Cys	Phe	Ser	Trp	Gly			
			275						280						285			
Cys	Ala	Gly	Glu	Val	Arg	Ala	Phe	Leu	Gln	Glu	Asn	Thr	His	Val	Arg			
			290						295						300			
Leu	Arg	Ile	Phe	Ala	Ala	Arg	Ile	Tyr	Asp	Tyr	Asp	Pro	Leu	Tyr	Lys			
			305						310						315			
															320			

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met
 325 330 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln
 340 345 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala
 355 360 365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
 370 375 380

<210> 80

<211> 396

<212> PRT

<213> Rattus norvegicus

<400> 80

Met Gln Pro Gln Gly Leu Gly Pro Asn Ala Gly Met Gly Pro Val Cys

1 5 10 15
 Leu Gly Cys Ser His Arg Arg Pro Tyr Ser Pro Ile Arg Asn Pro Leu
 20 25 30

Lys Lys Leu Tyr Gln Gln Thr Phe Tyr Phe His Phe Lys Asn Val Arg
 35 40 45

Tyr Ala Trp Gly Arg Lys Asn Asn Phe Leu Cys Tyr Glu Val Asn Gly
 50 55 60

Met Asp Cys Ala Leu Pro Val Pro Leu Arg Gln Gly Val Phe Arg Lys

65 70 75 80
 Gln Gly His Ile His Ala Glu Leu Cys Phe Ile Tyr Trp Phe His Asp
 85 90 95

Lys Val Leu Arg Val Leu Ser Pro Met Glu Glu Phe Lys Val Thr Trp
 100 105 110

Tyr Met Ser Trp Ser Pro Cys Ser Lys Cys Ala Glu Gln Val Ala Arg
 115 120 125

Phe Leu Ala Ala His Arg Asn Leu Ser Leu Ala Ile Phe Ser Ser Arg

130

135

140

Leu Tyr Tyr Tyr Leu Arg Asn Pro Asn Tyr Gln Gln Lys Leu Cys Arg
 145 150 155 160
 Leu Ile Gln Glu Gly Val His Val Ala Ala Met Asp Leu Pro Glu Phe
 165 170 175
 Lys Lys Cys Trp Asn Lys Phe Val Asp Asn Asp Gly Gln Pro Phe Arg
 180 185 190
 Pro Trp Met Arg Leu Arg Ile Asn Phe Ser Phe Tyr Asp Cys Lys Leu
 195 200 205
 Gln Glu Ile Phe Ser Arg Met Asn Leu Leu Arg Glu Asp Val Phe Tyr
 210 215 220
 Leu Gln Phe Asn Asn Ser His Arg Val Lys Pro Val Gln Asn Arg Tyr
 225 230 235 240
 Tyr Arg Arg Lys Ser Tyr Leu Cys Tyr Gln Leu Glu Arg Ala Asn Gly
 245 250 255
 Gln Glu Pro Leu Lys Gly Tyr Leu Leu Tyr Lys Lys Gly Glu Gln His
 260 265 270
 Val Glu Ile Leu Phe Leu Glu Lys Met Arg Ser Met Glu Leu Ser Gln
 275 280 285
 Val Arg Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
 290 295 300
 Arg Gln Leu Ala Ala Phe Lys Lys Asp His Pro Asp Leu Ile Leu Arg
 305 310 315 320
 Ile Tyr Thr Ser Arg Leu Tyr Phe Tyr Trp Arg Lys Lys Phe Gln Lys
 325 330 335
 Gly Leu Cys Thr Leu Trp Arg Ser Gly Ile His Val Asp Val Met Asp
 340 345 350
 Leu Pro Gln Phe Ala Asp Cys Trp Thr Asn Phe Val Asn Pro Gln Arg
 355 360 365
 Pro Phe Arg Pro Trp Asn Glu Leu Glu Lys Asn Ser Trp Arg Ile Gln
 370 375 380
 Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu

385	390										395									
<210>	81																			
<211>	227																			
<212>	PRT																			
<213>	Bos taurus																			
<400>	81																			
Asp	Gly	Trp	Glu	Val	Ala	Phe	Arg	Ser	Gly	Thr	Val	Leu	Lys	Ala	Gly					
1				5					10					15						
Val	Leu	Gly	Val	Ser	Met	Thr	Glu	Gly	Trp	Ala	Gly	Ser	Gly	His	Pro					
			20					25					30							
Gly	Gln	Gly	Ala	Cys	Val	Trp	Thr	Pro	Gly	Thr	Arg	Asn	Thr	Met	Asn					
		35					40					45								
Leu	Leu	Arg	Glu	Val	Leu	Phe	Lys	Gln	Gln	Phe	Gly	Asn	Gln	Pro	Arg					
	50					55					60									
Val	Pro	Ala	Pro	Tyr	Tyr	Arg	Arg	Lys	Thr	Tyr	Leu	Cys	Tyr	Gln	Leu					
65					70				75					80						
Lys	Gln	Arg	Asn	Asp	Leu	Thr	Leu	Asp	Arg	Gly	Cys	Phe	Arg	Asn	Lys					
			85						90					95						
Lys	Gln	Arg	His	Ala	Glu	Ile	Arg	Phe	Ile	Asp	Lys	Ile	Asn	Ser	Leu					
		100						105					110							
Asp	Leu	Asn	Pro	Ser	Gln	Ser	Tyr	Lys	Ile	Ile	Cys	Tyr	Ile	Thr	Trp					
	115						120					125								
Ser	Pro	Cys	Pro	Asn	Cys	Ala	Asn	Glu	Leu	Val	Asn	Phe	Ile	Thr	Arg					
	130					135					140									
Asn	Asn	His	Leu	Lys	Leu	Glu	Ile	Phe	Ala	Ser	Arg	Leu	Tyr	Phe	His					
145					150				155						160					
Trp	Ile	Lys	Ser	Phe	Lys	Met	Gly	Leu	Gln	Asp	Leu	Gln	Asn	Ala	Gly					
			165					170						175						
Ile	Ser	Val	Ala	Val	Met	Thr	His	Thr	Glu	Phe	Glu	Asp	Cys	Trp	Glu					
		180						185					190							
Gln	Phe	Val	Asp	Asn	Gln	Ser	Arg	Pro	Phe	Gln	Pro	Trp	Asp	Lys	Leu					
	195						200					205								

Glu Gln Tyr Ser Ala Ser Ile Arg Arg Arg Leu Gln Arg Ile Leu Thr

210 215 220

Ala Pro Ile

225

<210> 82

<211> 490

<212> PRT

<213> Pan troglodytes

<400> 82

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Trp Met Tyr Gln Arg Thr

1 5 10 15

Phe Tyr Tyr Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr

20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Arg Arg Gly His Ser Asn Leu

35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Met Tyr Ser Gln Pro Glu

50 55 60

His His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu

65 70 75 80

Ser Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro

85 90 95

Cys Pro Asp Cys Val Ala Lys Leu Ala Lys Phe Leu Ala Glu His Pro

100 105 110

Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu

115 120 125

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg

130 135 140

Val Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp

165 170 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Ile Arg His Leu

180 185 190
 Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val

 195 200 205
 Leu Arg Arg His Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp
 210 215 220
 Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn
 225 230 235 240
 Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu
 245 250 255
 Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile

 260 265 270
 Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly
 275 280 285
 Cys Ala Gly Gln Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg
 290 295 300
 Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys
 305 310 315 320
 Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met

 325 330 335
 Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln
 340 345 350
 Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala
 355 360 365
 Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Val Arg Ala Ser Ser Leu
 370 375 380
 Cys Met Val Pro His Arg Pro Pro Pro Pro Pro Gln Ser Pro Gly Pro

 385 390 395 400
 Cys Leu Pro Leu Cys Ser Glu Pro Pro Leu Gly Ser Leu Leu Pro Thr
 405 410 415
 Gly Arg Pro Ala Pro Ser Leu Pro Phe Leu Leu Thr Ala Ser Phe Ser
 420 425 430

Phe Pro Pro Pro Ala Ser Leu Pro Pro Leu Pro Ser Leu Ser Leu Ser
 435 440 445
 Pro Gly His Leu Pro Val Pro Ser Phe His Ser Leu Thr Ser Cys Ser
 450 455 460
 Ile Gln Pro Pro Cys Ser Ser Arg Ile Arg Glu Thr Glu Gly Trp Ala
 465 470 475 480
 Ser Val Ser Lys Glu Gly Arg Asp Leu Gly
 485 490
 <210> 83
 <211> 190
 <212> PRT
 <213> Homo sapiens
 <400> 83
 Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr
 1 5 10 15
 Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu
 20 25 30
 Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val
 35 40 45
 Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His
 50 55 60
 Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu
 65 70 75 80
 Ser Pro Asn Thr Lys Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
 85 90 95
 Cys Pro Asp Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser
 100 105 110
 Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Tyr
 115 120 125
 Pro Cys Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala
 130 135 140

Val Glu Ile Met Asp Tyr Glu Asp Phe Lys Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Asp Asn Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Thr

165 170 175

Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ser Leu Gln

180 185 190

<210> 84

<211> 190

<212> PRT

<213> Gorilla gorilla

<400> 84

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr

1 5 10 15

Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu

20 25 30

Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val

35 40 45

Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His

50 55 60

Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu

65 70 75 80

Ser Pro Asn Thr Asn Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro

85 90 95

Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser

100 105 110

Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Asp

115 120 125

Thr Asp Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala

130 135 140

Val Lys Ile Met Asp Tyr Lys Asp Phe Lys Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Asp Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Tyr
165 170 175
Asn Phe Arg Phe Leu Lys Arg Arg Leu Gln Glu Ile Leu Glu
180 185 190
<210> 85
<211> 199
<212> PRT
<213> Homo sapiens
<400> 85
Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His
1 5 10 15
Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Thr Tyr
20 25 30
Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met
35 40 45
Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys
50 55 60
Gly Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro
65 70 75 80
Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile
85 90 95
Ser Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala
100 105 110
Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg
115 120 125
Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg
130 135 140
Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His
145 150 155 160
Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp
165 170 175
Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala

180 185 190
Ile Leu Gln Asn Gln Gly Asn
195

<210> 86
<211> 202
<212> PRT
<213> Macaca mulatta
<400> 86

Met Asp Gly Ser Pro Ala Ser Arg Pro Arg His Leu Met Asp Pro Asn
1 5 10 15
Thr Phe Thr Phe Asn Phe Asn Asn Asp Leu Ser Val Arg Gly Arg His
20 25 30
Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Trp
35 40 45
Val Pro Met Asp Glu Arg Arg Gly Phe Leu Cys Asn Lys Ala Lys Asn

50 55 60
Val Pro Cys Gly Asp Tyr Gly Cys His Val Glu Leu Arg Phe Leu Cys
65 70 75 80
Glu Val Pro Ser Trp Gln Leu Asp Pro Ala Gln Thr Tyr Arg Val Thr
85 90 95
Trp Phe Ile Ser Trp Ser Pro Cys Phe Arg Arg Gly Cys Ala Gly Gln
100 105 110
Val Arg Val Phe Leu Gln Glu Asn Lys His Val Arg Leu Arg Ile Phe

115 120 125
Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Gln Glu Ala Leu Arg
130 135 140
Thr Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Glu Glu
145 150 155 160
Phe Lys His Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe
165 170 175
Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg

180 185 190
 Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
 195 200
 <210> 87
 <211> 185
 <212> PRT
 <213> Bos taurus
 <400> 87
 Met Asp Glu Tyr Thr Phe Thr Glu Asn Phe Asn Asn Gln Gly Trp Pro
 1 5 10 15
 Ser Lys Thr Tyr Leu Cys Tyr Glu Met Glu Arg Leu Asp Gly Asp Ala
 20 25 30
 Thr Ile Pro Leu Asp Glu Tyr Lys Gly Phe Val Arg Asn Lys Gly Leu
 35 40 45
 Asp Gln Pro Glu Lys Pro Cys His Ala Glu Leu Tyr Phe Leu Gly Lys
 50 55 60
 Ile His Ser Trp Asn Leu Asp Arg Asn Gln His Tyr Arg Leu Thr Cys
 65 70 75 80
 Phe Ile Ser Trp Ser Pro Cys Tyr Asp Cys Ala Gln Lys Leu Thr Thr
 85 90 95
 Phe Leu Lys Glu Asn His His Ile Ser Leu His Ile Leu Ala Ser Arg
 100 105 110
 Ile Tyr Thr His Asn Arg Phe Gly Cys His Gln Ser Gly Leu Cys Glu
 115 120 125
 Leu Gln Ala Ala Gly Ala Arg Ile Thr Ile Met Thr Phe Glu Asp Phe
 130 135 140
 Lys His Cys Trp Glu Thr Phe Val Asp His Lys Gly Lys Pro Phe Gln
 145 150 155 160
 Pro Trp Glu Gly Leu Asn Val Lys Ser Gln Ala Leu Cys Thr Glu Leu
 165 170 175
 Gln Ala Ile Leu Lys Thr Gln Gln Asn
 180 185

<210> 88

<211> 200

<212> PRT

<213> Homo sapiens

<400> 88

Met Ala Leu Leu Thr Ala Glu Thr Phe Arg Leu Gln Phe Asn Asn Lys

1 5 10 15

Arg Arg Leu Arg Arg Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr

20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly Tyr Phe Glu

35 40 45

Asn Lys Lys Lys Cys His Ala Glu Ile Cys Phe Ile Asn Glu Ile Lys

50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu

65 70 75 80

Thr Trp Ser Pro Cys Ser Ser Cys Ala Trp Glu Leu Val Asp Phe Ile

85 90 95

Lys Ala His Asp His Leu Asn Leu Gly Ile Phe Ala Ser Arg Leu Tyr

100 105 110

Tyr His Trp Cys Lys Pro Gln Gln Lys Gly Leu Arg Leu Leu Cys Gly

115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Phe Pro Lys Phe Ala Asp Cys

130 135 140

Trp Glu Asn Phe Val Asp His Glu Lys Pro Leu Ser Phe Asn Pro Tyr

145 150 155 160

Lys Met Leu Glu Glu Leu Asp Lys Asn Ser Arg Ala Ile Lys Arg Arg

165 170 175

Leu Glu Arg Ile Lys Ile Pro Gly Val Arg Ala Gln Gly Arg Tyr Met

180 185 190

Asp Ile Leu Cys Asp Ala Glu Val

195 200

<210> 89

<211> 210

<212> PRT

<213> Macaca mulatta

<400> 89

Met Ala Leu Leu Thr Ala Lys Thr Phe Ser Leu Gln Phe Asn Asn Lys

1 5 10 15

Arg Arg Val Asn Lys Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr

20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly His Leu Lys

35 40 45

Asn Lys Lys Lys Asp His Ala Glu Ile Arg Phe Ile Asn Lys Ile Lys

50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu

65 70 75 80

Thr Trp Ser Pro Cys Pro Ser Cys Ala Gly Glu Leu Val Asp Phe Ile

85 90 95

Lys Ala His Arg His Leu Asn Leu Arg Ile Phe Ala Ser Arg Leu Tyr

100 105 110

Tyr His Trp Arg Pro Asn Tyr Gln Glu Gly Leu Leu Leu Leu Cys Gly

115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Leu Pro Glu Phe Thr Asp Cys

130 135 140

Trp Glu Asn Phe Val Asp His Lys Glu Pro Pro Ser Phe Asn Pro Ser

145 150 155 160

Glu Lys Leu Glu Glu Leu Asp Lys Asn Ser Gln Ala Ile Lys Arg Arg

165 170 175

Leu Glu Arg Ile Lys Ser Arg Ser Val Asp Val Leu Glu Asn Gly Leu

180 185 190

Arg Ser Leu Gln Leu Gly Pro Val Thr Pro Ser Ser Ser Ile Arg Asn

195 200 205

Ser Arg

210

<210> 90

<211> 386

<212>

> PRT

<213> Homo sapiens

<400> 90

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr

20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu

35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Pro Val Leu Pro Lys Arg Gln

50 55 60

Ser Asn His Arg Gln Glu Val Tyr Phe Arg Phe Glu Asn His Ala Glu

65 70 75 80

Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Arg Leu Pro Ala Asn Arg

85 90 95

Arg Phe Gln Ile Thr Trp Phe Val Ser Trp Asn Pro Cys Leu Pro Cys

100 105 110

Val Val Lys Val Thr Lys Phe Leu Ala Glu His Pro Asn Val Thr Leu

115 120 125

Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Arg Asp Arg Asp Trp Arg

130 135 140

Trp Val Leu Leu Arg Leu His Lys Ala Gly Ala Arg Val Lys Ile Met

145 150 155 160

Asp Tyr Glu Asp Phe Ala Tyr Cys Trp Glu Asn Phe Val Cys Asn Glu

165 170 175

Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn Tyr Ala Ser

180 185 190

Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met Glu Ala Met

195 200 205

Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Leu Lys Ala Cys
 210 215 220
 Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val Thr Lys His
 225 230 235 240
 His Ser Ala Val Phe Arg Lys Arg Gly Val Phe Arg Asn Gln Val Asp
 245 250 255

 Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys
 260 265 270
 Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr Trp Tyr Thr
 275 280 285
 Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu
 290 295 300
 Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Cys
 305 310 315 320

 Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Cys Ser Leu Ser Gln
 325 330 335
 Glu Gly Ala Ser Val Lys Ile Met Gly Tyr Lys Asp Phe Val Ser Cys
 340 345 350
 Trp Lys Asn Phe Val Tyr Ser Asp Asp Glu Pro Phe Lys Pro Trp Lys
 355 360 365
 Gly Leu Gln Thr Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ile
 370 375 380

 Leu Gln
 385
 <210> 91
 <211> 236
 <212> PRT
 <213> Homo sapiens
 <400> 91
 Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg
 1 5 10 15
 Arg Ile Glu Pro Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu

20 25 30
 Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg
 35 40 45
 Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val

 50 55 60
 Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met
 65 70 75 80
 Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
 85 90 95
 Ser Gln Ala Ile Arg Glu Phe Leu Ser Arg His Pro Gly Val Thr Leu
 100 105 110
 Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg

 115 120 125
 Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
 130 135 140
 Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
 145 150 155 160
 Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met
 165 170 175
 Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys

 180 185 190
 Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu
 195 200 205
 His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu
 210 215 220
 Ala Thr Gly Leu Ile His Pro Ser Val Ala Trp Arg
 225 230 235
 <210> 92
 <211> 229
 <212> PRT
 <213> Mus musculus
 <400> 92

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15
 Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
 20 25 30
 Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
 35 40 45
 Ser Val Trp Arg His Thr Ser Gln Asn Thr Ser Asn His Val Glu Val
 50 55 60
 Asn Phe Leu Glu Lys Phe Thr Thr Glu Arg Tyr Phe Arg Pro Asn Thr

65 70 75 80
 Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
 85 90 95
 Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg His Pro Tyr Val Thr Leu
 100 105 110
 Phe Ile Tyr Ile Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg
 115 120 125
 Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140
 Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro
 145 150 155 160
 Pro Ser Asn Glu Ala Tyr Trp Pro Arg Tyr Pro His Leu Trp Val Lys
 165 170 175
 Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
 180 185 190
 Leu Lys Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205
 Thr Leu Gln Thr Cys His Tyr Gln Arg Ile Pro Pro His Leu Leu Trp
 210 215 220
 Ala Thr Gly Leu Lys

225

<210> 93

<211> 229

<212> PRT

<213> Rattus norvegicus

<400> 93

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220
Ala Thr Gly Leu Lys
225
<210> 94
<211> 224
<212> PRT
<213> Homo sapiens
<400> 94
Met Ala Gln Lys Glu Glu Ala Ala Val Ala Thr Glu Ala Ala Ser Gln
1 5 10 15
Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu
20 25 30
Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45
Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60
Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly
65 70 75 80
Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala
85 90 95
Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110
Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125
Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140
Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160
Glu Ile Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175
Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 95
<211> 224
<212> PRT
<213> Mus musculus

<400> 95

Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln
1 5 10 15

Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
20 25 30

Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro
35 40 45

Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Val Gln Ser Lys Gly
65 70 75 80

Gly Gln Ala Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly
85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110

Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125

Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140

Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160

Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val

180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile

195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 96

<211> 224

<212> PRT

<213> Rattus norvegicus

<400> 96

Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln

1 5 10 15

Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu

20 25 30

Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro

35 40 45

Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly

50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly

65 70 75 80

Gly Gln Val Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly

85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp

100 105 110

Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys

115 120 125

Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn

130 135 140

Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro

145 150 155 160

Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu

165 170 175
 Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Leu Trp Gln Asn Phe Val
 180 185 190
 Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
 195 200 205
 Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220
 <210> 97
 <211> 224
 <212> PRT
 <213> Bos taurus
 <400> 97

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Glu Pro Ala Ser Gln
 1 5 10 15
 Asn Gly Glu Glu Val Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
 20 25 30
 Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
 35 40 45

Ala His Tyr Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
 50 55 60
 Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly
 65 70 75 80
 Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Thr
 85 90 95
 Asn His Ala Glu Glu Ala Phe Phe Asn Ser Ile Met Pro Thr Phe Asp
 100 105 110

Pro Ala Leu Arg Tyr Met Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
 115 120 125
 Ala Ala Cys Ala Asp Arg Ile Val Lys Thr Leu Asn Lys Thr Lys Asn
 130 135 140
 Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
 145 150 155 160

Glu Ile Gln Ala Ala Leu Arg Lys Leu Lys Glu Ala Gly Cys Arg Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 98

<211> 208

<212> PRT

<213> Petromyzon marinus

<400> 98

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr

1 5 10 15
Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
20 25 30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
35 40 45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
50 55 60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg

65 70 75 80
Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
85 90 95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
100 105 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
115 120 125

Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn

130 135 140

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
 145 150 155 160
 Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
 165 170 175
 Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
 180 185 190
 Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val

195 200 205

<210> 99

<211> 384

<212> PRT

<213> Homo sapiens

<400> 99

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
 1 5 10 15
 Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
 20 25 30
 Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
 35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr
 50 55 60
 His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
 65 70 75 80
 His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
 85 90 95
 Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro
 100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
 115 120 125
 Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
 130 135 140
 Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys

145 150 155 160
 Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
 165 170 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
 180 185 190
 Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn
 195 200 205
 Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
 210 215 220
 Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
 225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
 245 250 255
 His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
 260 265 270
 Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
 275 280 285
 Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His
 290 295 300

Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg
 305 310 315 320
 Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser
 325 330 335
 Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
 340 345 350
 His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser
 355 360 365

Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
 370 375 380

<210> 100

<211> 184

<212> PRT

<213> Homo sapiens

<400> 100

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val

1 5 10 15

Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His

20 25 30

Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn

35 40 45

Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu

50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp

65 70 75 80

Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala

85 90 95

Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys

100 105 110

Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly

115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr

130 135 140

Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys

145 150 155 160

Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser

165 170 175

Gly Arg Leu Arg Ala Ile Leu Gln

180

<210> 101

<211> 184

<212> PRT

<213> Homo sapiens

<400> 101

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val
1 5 10 15
Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His
20 25 30
Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn
35 40 45
Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu
50 55 60
Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp
65 70 75 80
Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala
85 90 95
Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys
100 105 110
Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln Glu Gly
115 120 125
Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr
130 135 140
Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys
145 150 155 160
Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser
165 170 175
Gly Arg Leu Arg Ala Ile Leu Gln
180
<210> 102
<211> 16
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic polypeptide
<400> 102
Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
1 5 10 15

<210> 103

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (5)..(120)

<223> May be absent

<400> 103

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> 104

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (5)..(120)

<223> May be absent

<400> 104

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> 105

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (6)..(150)

<223> May be absent

<400> 105

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> 106

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (6)..(150)

<223> May be absent

<400> 106

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> 107

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 107

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> 108

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 108

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> 109

<211> 104

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 109

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> 110

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<400> 110

gaacacaaag catagactgc

20

<210> 111

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<400> 111

cttgtgtttc gtatctgacg 20

<210> 112

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<400> 112

gtcatcttag tcattacctg 20

<210> 113

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<400> 113

cagtagaatc agtaatggac 20

<210> 114

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<400> 114

ggaatccctt ctgcagcacc 20

<210> 115

<211> 1710

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 115

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45
 Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
 50 55 60
 Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

 65 70 75 80
 Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
 85 90 95
 Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
 100 105 110
 Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
 115 120 125
 Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

 130 135 140
 Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
 145 150 155 160
 Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
 165 170 175
 Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
 180 185 190
 Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

 195 200 205
 Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
 210 215 220
 Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
 225 230 235 240
 Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
 245 250 255
 Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro

 260 265 270
 Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
 275 280 285

Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
 290 295 300
 Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
 305 310 315 320
 Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
 325 330 335
 Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
 340 345 350
 Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
 355 360 365
 Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
 370 375 380
 Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
 385 390 395 400
 Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
 405 410 415
 Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
 420 425 430
 Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
 435 440 445
 Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
 450 455 460
 Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
 465 470 475 480
 Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
 485 490 495
 Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
 500 505 510
 Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
 515 520 525
 Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile

530 535 540
 Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
 545 550 555 560
 Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
 565 570 575
 Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys

 580 585 590
 Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
 595 600 605
 Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
 610 615 620
 Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
 625 630 635 640
 Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His

 645 650 655
 Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
 660 665 670
 Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
 675 680 685
 Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
 690 695 700
 Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

 705 710 715 720
 Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
 725 730 735
 Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
 740 745 750
 Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
 755 760 765
 Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu

 770 775 780

Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
 785 790 795 800
 Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
 805 810 815
 Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
 820 825 830
 Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
 835 840 845
 Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
 850 855 860
 Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
 865 870 875 880
 Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
 885 890 895
 Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
 900 905 910
 Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
 915 920 925
 Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
 930 935 940
 Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
 945 950 955 960
 Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
 965 970 975
 Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
 980 985 990
 Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
 995 1000 1005
 Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
 1010 1015 1020
 Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

1025	1030	1035
Ile Leu Lys Glu His Pro Val	Glu Asn Thr Gln Leu	Gln Asn Glu
1040	1045	1050
Lys Leu Tyr Leu Tyr Tyr Leu	Gln Asn Gly Arg Asp	Met Tyr Val
1055	1060	1065
Asp Gln Glu Leu Asp Ile Asn	Arg Leu Ser Asp Tyr	Asp Val Asp
1070	1075	1080
His Ile Val Pro Gln Ser Phe	Leu Lys Asp Asp Ser	Ile Asp Asn
1085	1090	1095
Lys Val Leu Thr Arg Ser Asp	Lys Asn Arg Gly Lys	Ser Asp Asn
1100	1105	1110
Val Pro Ser Glu Glu Val Val	Lys Lys Met Lys Asn	Tyr Trp Arg
1115	1120	1125
Gln Leu Leu Asn Ala Lys Leu	Ile Thr Gln Arg Lys	Phe Asp Asn
1130	1135	1140
Leu Thr Lys Ala Glu Arg Gly	Gly Leu Ser Glu Leu	Asp Lys Ala
1145	1150	1155
Gly Phe Ile Lys Arg Gln Leu	Val Glu Thr Arg Gln	Ile Thr Lys
1160	1165	1170
His Val Ala Gln Ile Leu Asp	Ser Arg Met Asn Thr	Lys Tyr Asp
1175	1180	1185
Glu Asn Asp Lys Leu Ile Arg	Glu Val Lys Val Ile	Thr Leu Lys
1190	1195	1200
Ser Lys Leu Val Ser Asp Phe	Arg Lys Asp Phe Gln	Phe Tyr Lys
1205	1210	1215
Val Arg Glu Ile Asn Asn Tyr	His His Ala His Asp	Ala Tyr Leu
1220	1225	1230
Asn Ala Val Val Gly Thr Ala	Leu Ile Lys Lys Tyr	Pro Lys Leu
1235	1240	1245
Glu Ser Glu Phe Val Tyr Gly	Asp Tyr Lys Val Tyr	Asp Val Arg
1250	1255	1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265

1270

1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280

1285

1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295

1300

1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310

1315

1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325

1330

1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340

1345

1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355

1360

1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1370

1375

1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385

1390

1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400

1405

1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415

1420

1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430

1435

1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445

1450

1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460

1465

1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475

1480

1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1490 1495 1500
 Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1505 1510 1515
 Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1520 1525 1530
 Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545
 Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1550 1555 1560
 Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575
 Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1580 1585 1590
 Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1595 1600 1605
 Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile

1610 1615 1620
 Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu

1625 1630 1635
 Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu

1640 1645 1650
 Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu

1655 1660 1665
 Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp

1670 1675 1680
 Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met

1685 1690 1695
 Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1700 1705 1710

<210> 116

<211> 1612

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 116

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210					215					220									
Ala	Thr	Gly	Leu	Lys	Ser	Gly	Ser	Glu	Thr	Pro	Gly	Thr	Ser	Glu	Ser				
225					230					235					240				
Ala	Thr	Pro	Glu	Ser	Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Ala	Ile	Gly				
245					250					255									
Thr	Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val	Pro				
260					265					270									
Ser	Lys	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys				
275					280					285									
Lys	Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu				
290					295					300									
Ala	Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys				
305					310					315					320				
Asn	Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys				
325					330					335									
Val	Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu				
340					345					350									
Glu	Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp				
355					360					365									
Glu	Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys				
370					375					380									
Lys	Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu				
385					390					395					400				
Ala	Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly				
405					410					415									
Asp	Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu				
420					425					430									
Val	Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser				
435					440					445									
Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg				
450					455					460									

Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480

Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510

Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525

Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540

Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560

Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575

Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590

Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620

Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640

Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655

Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670

Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685

Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700

Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720
 Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
 725 730 735

 Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
 740 745 750
 Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
 755 760 765
 Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
 770 775 780
 Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
 785 790 795 800

 Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
 805 810 815
 Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
 820 825 830
 Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
 835 840 845
 Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
 850 855 860

 Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
 865 870 875 880
 Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
 885 890 895
 Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
 900 905 910
 Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
 915 920 925

 Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
 930 935 940
 Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
 945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080

His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190	1195	1200
Ser Lys Leu Val Ser Asp Phe	Arg Lys Asp Phe Gln	Phe Tyr Lys
1205	1210	1215
Val Arg Glu Ile Asn Asn Tyr	His His Ala His Asp	Ala Tyr Leu
1220	1225	1230
Asn Ala Val Val Gly Thr Ala	Leu Ile Lys Lys Tyr	Pro Lys Leu
1235	1240	1245
Glu Ser Glu Phe Val Tyr Gly	Asp Tyr Lys Val Tyr	Asp Val Arg
1250	1255	1260
Lys Met Ile Ala Lys Ser Glu	Gln Glu Ile Gly Lys	Ala Thr Ala
1265	1270	1275
Lys Tyr Phe Phe Tyr Ser Asn	Ile Met Asn Phe Phe	Lys Thr Glu
1280	1285	1290
Ile Thr Leu Ala Asn Gly Glu	Ile Arg Lys Arg Pro	Leu Ile Glu
1295	1300	1305
Thr Asn Gly Glu Thr Gly Glu	Ile Val Trp Asp Lys	Gly Arg Asp
1310	1315	1320
Phe Ala Thr Val Arg Lys Val	Leu Ser Met Pro Gln	Val Asn Ile
1325	1330	1335
Val Lys Lys Thr Glu Val Gln	Thr Gly Gly Phe Ser	Lys Glu Ser
1340	1345	1350
Ile Leu Pro Lys Arg Asn Ser	Asp Lys Leu Ile Ala	Arg Lys Lys
1355	1360	1365
Asp Trp Asp Pro Lys Lys Tyr	Gly Gly Phe Asp Ser	Pro Thr Val
1370	1375	1380
Ala Tyr Ser Val Leu Val Val	Ala Lys Val Glu Lys	Gly Lys Ser
1385	1390	1395
Lys Lys Leu Lys Ser Val Lys	Glu Leu Leu Gly Ile	Thr Ile Met
1400	1405	1410
Glu Arg Ser Ser Phe Glu Lys	Asn Pro Ile Asp Phe	Leu Glu Ala
1415	1420	1425

Lys Gly Tyr Lys Glu Val Lys	Lys Asp Leu Ile Ile	Lys Leu Pro
1430	1435	1440
Lys Tyr Ser Leu Phe Glu Leu	Glu Asn Gly Arg Lys	Arg Met Leu
1445	1450	1455
Ala Ser Ala Gly Glu Leu Gln	Lys Gly Asn Glu Leu	Ala Leu Pro
1460	1465	1470
Ser Lys Tyr Val Asn Phe Leu	Tyr Leu Ala Ser His	Tyr Glu Lys
1475	1480	1485
Leu Lys Gly Ser Pro Glu Asp	Asn Glu Gln Lys Gln	Leu Phe Val
1490	1495	1500
Glu Gln His Lys His Tyr Leu	Asp Glu Ile Ile Glu	Gln Ile Ser
1505	1510	1515
Glu Phe Ser Lys Arg Val Ile	Leu Ala Asp Ala Asn	Leu Asp Lys
1520	1525	1530
Val Leu Ser Ala Tyr Asn Lys	His Arg Asp Lys Pro	Ile Arg Glu
1535	1540	1545
Gln Ala Glu Asn Ile Ile His	Leu Phe Thr Leu Thr	Asn Leu Gly
1550	1555	1560
Ala Pro Ala Ala Phe Lys Tyr	Phe Asp Thr Thr Ile	Asp Arg Lys
1565	1570	1575
Arg Tyr Thr Ser Thr Lys Glu	Val Leu Asp Ala Thr	Leu Ile His
1580	1585	1590
Gln Ser Ile Thr Gly Leu Tyr	Glu Thr Arg Ile Asp	Leu Ser Gln
1595	1600	1605
Leu Gly Gly Asp		
1610		
<210> 117		
<211> 1929		
<212> PRT		
<213> Artificial Sequence		
<220><223> Synthetic polypeptide		
<400> 117		

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly

245 250 255
 Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
 260 265 270
 Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
 275 280 285

 Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
 290 295 300
 Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
 305 310 315 320
 Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
 325 330 335
 Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
 340 345 350

 Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
 355 360 365
 Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
 370 375 380
 Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
 385 390 395 400
 Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
 405 410 415

 Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
 420 425 430
 Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
 435 440 445
 Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
 450 455 460
 Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
 465 470 475 480

 Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
 485 490 495

Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
 500 505 510
 Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
 515 520 525
 Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
 530 535 540

 Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
 545 550 555 560
 Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
 565 570 575
 Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
 580 585 590
 Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
 595 600 605

 Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
 610 615 620
 Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
 625 630 635 640
 Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
 645 650 655
 Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
 660 665 670

 Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
 675 680 685
 Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
 690 695 700
 Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
 705 710 715 720
 Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
 725 730 735

 Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu

740	745	750	
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu			
755	760	765	
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu			
770	775	780	
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn			
785	790	795	800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile			
805	810	815	
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn			
820	825	830	
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys			
835	840	845	
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val			
850	855	860	
Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu			
865	870	875	880
Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys			
885	890	895	
Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn			
900	905	910	
Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys			
915	920	925	
Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp			
930	935	940	
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln			
945	950	955	960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala			
965	970	975	
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val			
980	985	990	

Lys Val Met Gly Arg His Lys Pro	Glu Asn Ile Val Ile	Glu Met Ala
995	1000	1005
Arg Glu Asn Gln Thr Thr Gln	Lys Gly Gln Lys Asn	Ser Arg Glu
1010	1015	1020
Arg Met Lys Arg Ile Glu Glu	Gly Ile Lys Glu Leu	Gly Ser Gln
1025	1030	1035
Ile Leu Lys Glu His Pro Val	Glu Asn Thr Gln Leu	Gln Asn Glu
1040	1045	1050
Lys Leu Tyr Leu Tyr Tyr Leu	Gln Asn Gly Arg Asp	Met Tyr Val
1055	1060	1065
Asp Gln Glu Leu Asp Ile Asn	Arg Leu Ser Asp Tyr	Asp Val Asp
1070	1075	1080
His Ile Val Pro Gln Ser Phe	Leu Lys Asp Asp Ser	Ile Asp Asn
1085	1090	1095
Lys Val Leu Thr Arg Ser Asp	Lys Asn Arg Gly Lys	Ser Asp Asn
1100	1105	1110
Val Pro Ser Glu Glu Val Val	Lys Lys Met Lys Asn	Tyr Trp Arg
1115	1120	1125
Gln Leu Leu Asn Ala Lys Leu	Ile Thr Gln Arg Lys	Phe Asp Asn
1130	1135	1140
Leu Thr Lys Ala Glu Arg Gly	Gly Leu Ser Glu Leu	Asp Lys Ala
1145	1150	1155
Gly Phe Ile Lys Arg Gln Leu	Val Glu Thr Arg Gln	Ile Thr Lys
1160	1165	1170
His Val Ala Gln Ile Leu Asp	Ser Arg Met Asn Thr	Lys Tyr Asp
1175	1180	1185
Glu Asn Asp Lys Leu Ile Arg	Glu Val Lys Val Ile	Thr Leu Lys
1190	1195	1200
Ser Lys Leu Val Ser Asp Phe	Arg Lys Asp Phe Gln	Phe Tyr Lys
1205	1210	1215
Val Arg Glu Ile Asn Asn Tyr	His His Ala His Asp	Ala Tyr Leu

1220	1225	1230
Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu		
1235	1240	1245
Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg		
1250	1255	1260
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala		
1265	1270	1275
Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu		
1280	1285	1290
Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu		
1295	1300	1305
Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp		
1310	1315	1320
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile		
1325	1330	1335
Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser		
1340	1345	1350
Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Gln Asn Glu Lys Leu		
1355	1360	1365
Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln		
1370	1375	1380
Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile		
1385	1390	1395
Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val		
1400	1405	1410
Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro		
1415	1420	1425
Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu		
1430	1435	1440
Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr		
1445	1450	1455

Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
1460 1465 1470

Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val
1475 1480 1485

Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn
1490 1495 1500

Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
1505 1510 1515

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
1520 1525 1530

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala
1535 1540 1545

Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser
1550 1555 1560

Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met
1565 1570 1575

Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr
1580 1585 1590

Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr
1595 1600 1605

Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn
1610 1615 1620

Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala
1625 1630 1635

Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys
1640 1645 1650

Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu
1655 1660 1665

Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp
1670 1675 1680

Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr

1685	1690	1695
Ser Val Leu Val Val Ala Lys	Val Glu Lys Gly Lys	Ser Lys Lys
1700	1705	1710
Leu Lys Ser Val Lys Glu Leu	Leu Gly Ile Thr Ile	Met Glu Arg
1715	1720	1725
Ser Ser Phe Glu Lys Asn Pro	Ile Asp Phe Leu Glu	Ala Lys Gly
1730	1735	1740
Tyr Lys Glu Val Lys Lys Asp	Leu Ile Ile Lys Leu	Pro Lys Tyr
1745	1750	1755
Ser Leu Phe Glu Leu Glu Asn	Gly Arg Lys Arg Met	Leu Ala Ser
1760	1765	1770
Ala Gly Glu Leu Gln Lys Gly	Asn Glu Leu Ala Leu	Pro Ser Lys
1775	1780	1785
Tyr Val Asn Phe Leu Tyr Leu	Ala Ser His Tyr Glu	Lys Leu Lys
1790	1795	1800
Gly Ser Pro Glu Asp Asn Glu	Gln Lys Gln Leu Phe	Val Glu Gln
1805	1810	1815
His Lys His Tyr Leu Asp Glu	Ile Ile Glu Gln Ile	Ser Glu Phe
1820	1825	1830
Ser Lys Arg Val Ile Leu Ala	Asp Ala Asn Leu Asp	Lys Val Leu
1835	1840	1845
Ser Ala Tyr Asn Lys His Arg	Asp Lys Pro Ile Arg	Glu Gln Ala
1850	1855	1860
Glu Asn Ile Ile His Leu Phe	Thr Leu Thr Asn Leu	Gly Ala Pro
1865	1870	1875
Ala Ala Phe Lys Tyr Phe Asp	Thr Thr Ile Asp Arg	Lys Arg Tyr
1880	1885	1890
Thr Ser Thr Lys Glu Val Leu	Asp Ala Thr Leu Ile	His Gln Ser
1895	1900	1905
Ile Thr Gly Leu Tyr Glu Thr	Arg Ile Asp Leu Ser	Gln Leu Gly
1910	1915	1920

Gly Asp Ser Gly Gly Ser

1925

<210> 118

<211> 1385

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 118

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Asp Lys

1 5 10 15

Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala

20 25 30

Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu

35 40 45

Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu

50 55 60

Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr

65 70 75 80

Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln

85 90 95

Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His

100 105 110

Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg

115 120 125

His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys

130 135 140

Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp

145 150 155 160

Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys

165 170 175

Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser

180 185 190

Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu
 195 200 205

Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile
 210 215 220

Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala
 225 230 235 240

Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala
 245 250 255

Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala
 260 265 270

Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu
 275 280 285

Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu
 290 295 300

Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg
 305 310 315 320

Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys
 325 330 335

Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val
 340 345 350

Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser
 355 360 365

Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu
 370 375 380

Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu
 385 390 395 400

Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg
 405 410 415

Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu
 420 425 430

His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp

435 440 445
 Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr
 450 455 460

 Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg
 465 470 475 480
 Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp
 485 490 495
 Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp
 500 505 510
 Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr
 515 520 525

 Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr
 530 535 540
 Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala
 545 550 555 560
 Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln
 565 570 575
 Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu
 580 585 590

 Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His
 595 600 605
 Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu
 610 615 620
 Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu
 625 630 635 640
 Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe
 645 650 655

 Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp
 660 665 670
 Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser
 675 680 685

Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg
690 695 700

Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp
705 710 715 720

Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His
725 730 735

Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln
740 745 750

Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys
755 760 765

Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln
770 775 780

Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly
785 790 795 800

Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn
805 810 815

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly
820 825 830

Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp
835 840 845

Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser
850 855 860

Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser
865 870 875 880

Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp
885 890 895

Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
900 905 910

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
915 920 925

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val

930	935	940	
Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp			
945	950	955	960
Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val			
	965	970	975
Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn			
	980	985	990
Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr			
	995	1000	1005
Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr			
	1010	1015	1020
Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser			
	1025	1030	1035
Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser			
	1040	1045	1050
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly			
	1055	1060	1065
Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly			
	1070	1075	1080
Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys			
	1085	1090	1095
Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val			
	1100	1105	1110
Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn			
	1115	1120	1125
Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys			
	1130	1135	1140
Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val			
	1145	1150	1155
Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val			
	1160	1165	1170

Lys Glu	Leu Leu Gly Ile Thr	Ile Met Glu Arg Ser	Ser Phe Glu
1175	1180	1185	
Lys Asn	Pro Ile Asp Phe Leu	Glu Ala Lys Gly Tyr	Lys Glu Val
1190	1195	1200	
Lys Lys	Asp Leu Ile Ile Lys	Leu Pro Lys Tyr Ser	Leu Phe Glu
1205	1210	1215	
Leu Glu	Asn Gly Arg Lys Arg	Met Leu Ala Ser Ala	Gly Glu Leu
1220	1225	1230	
Gln Lys	Gly Asn Glu Leu Ala	Leu Pro Ser Lys Tyr	Val Asn Phe
1235	1240	1245	
Leu Tyr	Leu Ala Ser His Tyr	Glu Lys Leu Lys Gly	Ser Pro Glu
1250	1255	1260	
Asp Asn	Glu Gln Lys Gln Leu	Phe Val Glu Gln His	Lys His Tyr
1265	1270	1275	
Leu Asp	Glu Ile Ile Glu Gln	Ile Ser Glu Phe Ser	Lys Arg Val
1280	1285	1290	
Ile Leu	Ala Asp Ala Asn Leu	Asp Lys Val Leu Ser	Ala Tyr Asn
1295	1300	1305	
Lys His	Arg Asp Lys Pro Ile	Arg Glu Gln Ala Glu	Asn Ile Ile
1310	1315	1320	
His Leu	Phe Thr Leu Thr Asn	Leu Gly Ala Pro Ala	Ala Phe Lys
1325	1330	1335	
Tyr Phe	Asp Thr Thr Ile Asp	Arg Lys Arg Tyr Thr	Ser Thr Lys
1340	1345	1350	
Glu Val	Leu Asp Ala Thr Leu	Ile His Gln Ser Ile	Thr Gly Leu
1355	1360	1365	
Tyr Glu	Thr Arg Ile Asp Leu	Ser Gln Leu Gly Gly	Asp Ser Gly
1370	1375	1380	
Gly Ser			
1385			

<210> 119

<211> 82

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<400> 119

guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu ccguuaucaa cuugaaaaag 60

uggcaccgag ucggugcuuu uu 82

<210> 120

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<400> 120

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser

1 5 10

<210> 121

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (2)..(30)

<223> May be absent

<400> 121

Gly 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> 122

<211> 90

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (4)..(90)

<223> May be absent

<400> 122

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> 123

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic polypeptide

<220><221> misc_feature

<222> (1)..(1)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (3)..(60)

<223> May be absent

<400> 123

Xaa Pro 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 Xaa Pro

20 25 30

Xaa Pro 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