(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0121113
(43) 공개일자 2021년10월07일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) *C12N 15/62* (2006.01)
C12N 9/80 (2006.01)
(52) CPC특허분류
C12N 9/22 (2013.01)
C12N 15/62 (2013.01)
(21) 출원번호 10-2021-7026724
(22) 출원일자(국제) 2020년01월31일
심사청구일자 없음
(85) 번역문제출일자 2021년08월23일
(86) 국제출원번호 PCT/US2020/016285
(87) 국제공개번호 WO 2020/160514
국제공개일자 2020년08월06일
(30) 우선권주장
62/799,702 2019년01월31일 미국(US)

(71) 출원인
빔 테라퓨틱스, 인크.
미국 02139 매사추세츠주 케임브리지 랜즈다운 스트리트 26 세컨드 플로어
(72) 발명자
슬레이메이커 이안
미국 매사추세츠주 02139 캠브리지 랜즈다운 스트리트 26 세컨드 플로어
제크 제이슨 마이클
미국 매사추세츠주 02139 캠브리지 랜즈다운 스트리트 26 세컨드 플로어
(뒷면에 계속)
(74) 대리인
문두현

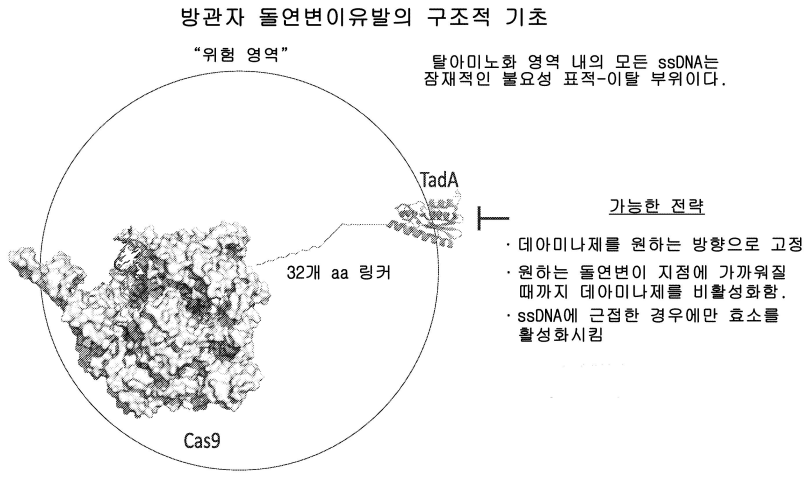
전체 청구항 수 : 총 160 항

(54) 발명의 명칭 **비-표적 탈아미노화가 감소된 핵염기 편집기 및 핵염기 편집기를 특성규명하기 위한 분석**

(57) 요 약

본 발명은 비-표적 탈아미노화가 감소된 염기 편집기, 상기 염기 편집기를 이용하는 방법과, 예를 들어, 프로그래밍된, 표적-적중 탈아미노화와 비교하여, 염기 편집기를 감소된 비-표적 탈아미노화를 갖는 것으로 특성규명하기 위한 분석을 특징으로 한다.

대 표 도



(52) CPC특허분류
C12N 9/80 (2013.01)
C12Y 305/04005 (2013.01)
C07K 2319/00 (2013.01)
C12N 2310/20 (2017.05)
C12N 2320/11 (2013.01)
(72) 발명자
고델리 니콜
미국 매사추세츠주 02139 캠브리지 랜즈다운 스트리트 26 세컨드 플로어
유 이
미국 매사추세츠주 02139 캠브리지 랜즈다운 스트리트 26 세컨드 플로어

명 세 서

청구범위

청구항 1

Cas9 폴리펩티드의 가요성 루프 내에 삽입된 데아미나제를 포함하는 융합 단백질로서, 상기 융합 단백질이 다음 구조를 포함하는 것인, 융합 단백질:

NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH,

여기서, "]-["의 각 경우는 선택적 링커이다.

청구항 2

Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된(flanked) 데아미나제를 포함하는 융합 단백질로서, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단은 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함하는, 융합 단백질.

청구항 3

제1항 또는 제2항에 있어서, 상기 융합 단백질의 데아미나제가 표적 폴리뉴클레오티드 서열에서 표적 핵염기를 탈아미노화시키는, 융합 단백질.

청구항 4

제3항에 있어서, 상기 융합 단백질이 상기 표적 핵염기를 탈아미노화시키는 경우, 상기 가요성 루프가 상기 표적 핵염기에 근접한 아미노산을 포함하는, 융합 단백질.

청구항 5

제4항에 있어서, 상기 가요성 루프가 상기 Cas9 폴리펩티드의 알파-나선 구조의 일부를 포함하는, 융합 단백질.

청구항 6

제4항 또는 제5항에 있어서, 상기 표적 핵염기가 서열번호: 1의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 단부 말단 융합 단백질과 비교하여 더 낮은 표적-이탈 탈아미노화로 탈아미노화되는, 융합 단백질.

청구항 7

제4항 또는 제5항에 있어서, 상기 표적 핵염기가 상기 표적 폴리뉴클레오티드 서열 내의 프로토스페이서 인접 모티프(PAM) 서열로부터 1 내지 20개의 핵염기 떨어져 있는, 융합 단백질.

청구항 8

제7항에 있어서, 상기 표적 핵염기가 상기 PAM 서열의 상류에 있는 2개 내지 12개 핵염기인, 융합 단백질.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 가요성 루프가 서열번호: 1에서 번호매겨진 것과 같은 위치 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300에서의 아미노산 잔기로 이루어진 군으로부터 선택된 영역, 또는 이들의 상응하는 영역을 포함하는, 융합 단백질.

청구항 10

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249, 또는 이들의 상응하는 아미노산 위치 사이에 삽입되는, 융합 단백질.

청구항 11

제10항에 있어서, 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248, 또는 이들의 상응하는 아미노산 위치 사이에 삽입되는, 융합 단백질.

청구항 12

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 서열번호: 1에서 번호매겨진 것과 같은 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입되는, 융합 단백질.

청구항 13

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 N-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 상기 Cas9 폴리펩티드의 아미노산 잔기 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 및/또는 1248-1297, 또는 이들의 상응하는 잔기를 포함하는, 융합 단백질.

청구항 14

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 C-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 상기 Cas9 폴리펩티드의 아미노산 잔기 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 및/또는 538-568, 또는 이들의 상응하는 잔기를 포함하는, 융합 단백질.

청구항 15

제3항 내지 제8항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드의 상기 N 말단 단편 또는 상기 C 말단 단편이 상기 표적 폴리뉴클레오티드 서열에 결합하는, 융합 단백질.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드의 상기 N-말단 단편 또는 상기 C-말단 단편이 DNA 결합 도메인을 포함하는, 융합 단백질.

청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 RuvC 도메인을 포함하는, 융합 단백질.

청구항 18

제1항 내지 제17항 중 어느 한 항에 있어서, 상기 N-말단 단편 또는 상기 C 말단 단편이 HNH 도메인을 포함하는, 융합 단백질.

청구항 19

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 N-말단 단편 및 상기 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는, 융합 단백질.

청구항 20

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 N-말단 단편 및 상기 C-말단 단편 중 어느 것도 RuvC 도메인을 포함하지 않는, 융합 단백질.

청구항 21

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 하나 이상의 구조적 도메인에서 부분적 또는 완전한 결실을 포함하는, 융합 단백질.

청구항 22

제21항에 있어서, 상기 데아미나제가 상기 Cas9 폴리펩티드의 부분적 또는 완전한 결실 위치에 삽입되는 것인, 융합 단백질.

청구항 23

제21항 또는 제22항에 있어서, 상기 결실이 RuvC 도메인 내에 있는, 융합 단백질.

청구항 24

제21항 또는 제22항에 있어서, 상기 결실이 HNH 도메인 내에 있는, 융합 단백질.

청구항 25

제21항 또는 제22항에 있어서, 상기 결실이 RuvC 도메인과 C-말단 도메인, L-I 도메인과 HNH 도메인, 또는 RuvC 도메인과 L-I 도메인을 연결하는, 융합 단백질.

청구항 26

제21항 또는 제22항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017-1069 또는 이들의 상응하는 아미노산의 결실을 포함하는, 융합 단백질.

청구항 27

제21항 또는 제22항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-872 또는 이들의 상응하는 아미노산의 결실을 포함하는, 융합 단백질.

청구항 28

제21항 또는 제22항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-906 또는 이들의 상응하는 아미노산의 결실을 포함하는, 융합 단백질.

청구항 29

Cas9 폴리펩티드 내에 삽입된 데아미나제를 포함하는 융합 단백질로서, 상기 융합 단백질은 다음 구조를 포함하는 것인, 융합 단백질:

NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH,

여기서, "]-["의 각 경우는 선택적 링커이고,

상기 Cas9 폴리펩티드는 HNH 도메인의 완전한 결실을 포함하고,

상기 데아미나제는 상기 결실 위치에 삽입된다.

청구항 30

제29항에 있어서, 상기 N 말단 단편의 C 말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 791인, 융합 단백질.

청구항 31

제30항에 있어서, 상기 C 말단 단편의 N 말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 907인, 융합 단백질.

청구항 32

제30항에 있어서, 상기 C 말단 단편의 N 말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 873인, 융합 단백질.

청구항 33

Cas9 폴리펩티드 내에 삽입된 데아미나제를 포함하는 융합 단백질로서, 상기 융합 단백질은 다음 구조를 포함하

는 것인, 융합 단백질:

NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH,

여기서, "]-["의 각 경우는 선택적 링커이고,

상기 Cas9는 RuvC 도메인의 완전한 결실을 포함하고,

상기 데아미나제는 상기 결실 위치에 삽입된다.

청구항 34

제1항 내지 제33항 중 어느 한 항에 있어서, 상기 데아미나제가 시티딘 데아미나제 또는 아데노신 데아미나제인, 융합 단백질

청구항 35

제34항에 있어서, 상기 시티딘 데아미나제가 APOBEC 시티딘 데아미나제, 활성화 유도 시티딘 데아미나제(AID), 또는 CDA인, 융합 단백질

청구항 36

제35항에 있어서, 상기 APOBEC 데아미나제가 APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3E, APOBEC3F, APOBEC3G, APOBEC3H, 또는 APOBEC4인, 융합 단백질

청구항 37

제36항에 있어서, 상기 APOBEC 데아미나제가 rAPOBEC1인, 융합 단백질.

청구항 38

제34항 내지 제37항 중 어느 한 항에 있어서, UGI 도메인을 추가로 포함하는, 융합 단백질.

청구항 39

제34항에 있어서, 상기 아데노신 데아미나제가 TadA 데아미나제인, 융합 단백질.

청구항 40

제39항에 있어서, 상기 TadA 데아미나제가 변형된 TadA인, 융합 단백질.

청구항 41

제40항에 있어서, 상기 TadA 데아미나제가 TadA 7.10인, 융합 단백질.

청구항 42

제41항에 있어서, 상기 아데노신 데아미나제가 TadA 이량체인, 융합 단백질.

청구항 43

제42항에 있어서, 상기 TadA 이량체가 TadA 7.10 및 야생형 TadA를 포함하는, 융합 단백질.

청구항 44

제3항 내지 제43항 중 어느 한 항에 있어서, 상기 선택적 링커가 (SGGS)n, (GGGS)n, (GGGGS)n, (G)n, (EAAAAK)n, (GGS)n, SGSETPGTSESATPES, 또는 (XP)n 모티프, 또는 이들의 조합을 포함하고, 상기 n은 독립적으로 1 내지 30 사이의 정수인,융합 단백질.

청구항 45

제1항 내지 제43항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드의 N 말단 단편이 링커 없이 데아미나제에 융합되는, 융합 단백질.

청구항 46

제1항 내지 제43항 중 어느 한 항에 있어서, 상기 Cas9의 상기 C 말단 단편이 링커 없이 데아미나제에 융합되는, 융합 단백질.

청구항 47

제1항 내지 제46항 중 어느 한 항에 있어서, 추가의 촉매 도메인을 추가로 포함하는, 융합 단백질.

청구항 48

제47항에 있어서, 상기 추가의 촉매 도메인이 제2 데아미나제인, 융합 단백질

청구항 49

제48항에 있어서, 상기 제2 데아미나제가 상기 융합 단백질의 상기 N 말단 또는 상기 C 말단에 융합된 것인, 융합 단백질.

청구항 50

제48항 또는 제49항에 있어서, 상기 데아미나제가 시티딘 데아미나제 또는 아데노신 데아미나제인, 융합 단백질.

청구항 51

제1항 내지 제50항 중 어느 한 항에 있어서, 핵 위치결정 신호(nuclear localization signal)를 추가로 포함하는, 융합 단백질.

청구항 52

제51항에 있어서, 상기 핵 위치결정 신호가 이분 핵 위치결정 신호인, 융합 단백질.

청구항 53

제1항 내지 제52항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 스트렙토코커스 피오게네스(*Streptococcus pyogenes*) Cas9(SpCas9), 스타필로코커스 아우레우스(Staphylococcus aureus) Cas9(SaCas9), 스트렙토코커스 써모필러스(*Streptococcus thermophilus*) 1 Cas9(St1Cas9), 또는 이들의 변이체인, 융합 단백질.

청구항 54

제1항 내지 제53항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 변형된 Cas9이고 변경된 PAM에 대한 특이성을 갖는, 융합 단백질.

청구항 55

제1항 내지 제54항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 닉카아제인, 융합 단백질.

청구항 56

제1항 내지 제54항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 뉴클레아제 비활성인, 융합 단백질.

청구항 57

제3항 내지 제56항 중 어느 한 항에 있어서, 가이드 핵산 서열과 복합체화하여 상기 표적 핵염기의 탈아미노화를 수행하는 것인, 융합 단백질.

청구항 58

제57항에 있어서, 상기 표적 폴리뉴클레오티드와 추가로 복합체화하는, 융합 단백질.

청구항 59

제1항 내지 제58항 중 어느 한 항의 융합 단백질을 코딩하는 폴리뉴클레오티드.

청구항 60

제59항의 폴리뉴클레오티드를 포함하는 발현 벡터.

청구항 61

제60항에 있어서, 상기 발현 벡터가 포유류 발현 벡터인, 발현 벡터.

청구항 62

제61항에 있어서, 상기 벡터가 아데노-관련 바이러스(AAV), 레트로바이러스 벡터, 아데노바이러스 벡터, 렌티바이러스 벡터, 센다이 바이러스 벡터, 및 헤르페스바이러스 벡터로 이루어진 군으로부터 선택된 바이러스 벡터인, 발현 벡터.

청구항 63

제60항 내지 제62항 중 어느 한 항에 있어서, 상기 벡터가 프로모터를 포함하는 것인, 발현 벡터.

청구항 64

제1항 내지 제58항 중 어느 한 항의 융합 단백질, 제59항의 폴리뉴클레오티드, 또는 제60항 내지 제63항 중 어느 한 항의 벡터를 포함하는 세포.

청구항 65

제64항에 있어서, 상기 세포가 박테리아 세포, 식물 세포, 곤충 세포, 인간 세포 또는 포유류 세포인, 세포.

청구항 66

제1항 내지 제58항 중 어느 한 항의 융합 단백질, 제59항의 폴리뉴클레오티드, 또는 제60항 내지 제63항 중 어느 한 항의 벡터를 포함하는 키트.

청구항 67

염기 편집 방법으로서,

폴리뉴클레오티드 서열을 제1항 내지 제58항 중 어느 한 항의 융합 단백질과 접촉시키는 단계를 포함하며, 상기 융합 단백질의 데아미나제는 상기 폴리뉴클레오티드 내의 핵염기를 탈아미노화시킴으로써, 상기 폴리뉴클레오티드 서열을 편집하는 것인, 방법.

청구항 68

제67항에 있어서, 상기 표적 폴리뉴클레오티드 서열을 가이드 핵산 서열과 접촉시켜 상기 표적 핵염기의 탈아미노화를 수행하는 단계를 추가로 포함하는, 방법.

청구항 69

표적 폴리뉴클레오티드 서열에서 표적 핵염기를 편집하는 방법으로서, 상기 방법은 상기 표적 폴리뉴클레오티드 서열을 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하는 융합 단백질과 접촉시키는 단계를 포함하며, 상기 융합 단백질의 데아미나제는 상기 표적 폴리뉴클레오티드 서열 중의 상기 표적 핵염기를 탈아미노화시키고, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단은 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함하는, 방법.

청구항 70

표적 폴리뉴클레오티드 서열에서 표적 핵염기를 편집하는 방법으로서, 상기 방법은 상기 표적 폴리뉴클레오티드 서열을 Cas9 폴리펩티드의 가요성 루프 내에 삽입된 데아미나제를 포함하는 융합 단백질과 접촉시키는 단계를 포함하며, 상기 융합 단백질은 NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH 구조를 포함하고, 여기서, "]-["의 각 경우는 선택적 링커이고, 상기 융합 단백질의 데아미나제는 상기 표적 폴리뉴클레오

티드 서열에서 상기 표적 핵염기를 탈아미노화시키는, 방법.

청구항 71

제69항 또는 제70항에 있어서, 상기 표적 폴리뉴클레오티드 서열을 가이드 핵산 서열과 접촉시켜 상기 표적 핵염기의 탈아미노화를 수행하는 단계를 추가로 포함하는, 방법.

청구항 72

제71항에 있어서, 상기 가이드 핵산 서열이 상기 표적 폴리뉴클레오티드 서열 중의 프로토스페이서 서열에 상보적인 스페이서 서열을 포함하며, 그로 인해 R-루프를 형성하는, 방법.

청구항 73

제69항 내지 제72항 중 어느 한 항에 있어서, 상기 표적 뉴클레오티드가 서열번호: 1의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 단부 말단 방법과 비교하여, 더 낮은 표적-이탈 탈아미노화로 탈아미노화되는, 방법.

청구항 74

제72항에 있어서, 상기 융합 단백질의 데아미나제가 R-루프의 영역 내에서 2개 이하의 핵염기를 탈아미노화시키는, 방법.

청구항 75

제69항 내지 제74항 중 어느 한 항에 있어서, 상기 표적 핵염기가 상기 표적 폴리뉴클레오티드 서열 중의 PAM 서열로부터 떨어져 있는 1개 내지 20개 핵염기인, 방법.

청구항 76

제75항에 있어서, 상기 표적 핵염기가 PAM 서열의 상류에 있는 2개 내지 12개 핵염기인, 방법.

청구항 77

제69항 내지 제76항 중 어느 한 항에 있어서, 상기 융합 단백질의 데아미나제가 상기 표적 핵염기를 탈아미노화시킬 때, 상기 가요성 루프가 상기 표적 핵염기에 근접한 아미노산을 포함하는, 방법.

청구항 78

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 가요성 루프가 서열번호: 1에서 번호매겨진 것과 같은 위치 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300에서의 아미노산 잔기로 이루어진 군으로부터 선택된 영역, 또는 이의 상응하는 영역을 포함하는, 방법.

청구항 79

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249 또는 이들의 상응하는 아미노산 위치 사이에 삽입되는, 방법.

청구항 80

제79항에 있어서, 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248, 또는 이들의 상응하는 아미노산 위치 사이에 삽입되는, 방법.

청구항 81

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070 또는 1247-1248 또는 이들의 상응하는 아미

노산 위치 사이에 삽입되는, 방법.

청구항 82

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 N-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 상기 Cas9 폴리펩티드의 아미노산 잔기 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 및/또는 1248-1297, 또는 이들의 상응하는 잔기를 포함하는, 방법.

청구항 83

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 C-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 상기 Cas9 폴리펩티드의 아미노산 잔기 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 및/또는 538-568, 또는 이들의 상응하는 잔기를 포함하는, 방법.

청구항 84

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드의 N 말단 단편 또는 C 말단 단편이 상기 표적 폴리뉴클레오티드 서열에 결합하는, 방법.

청구항 85

제69항 내지 제84항 중 어느 한 항에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 RuvC 도메인을 포함하는, 방법.

청구항 86

제69항 내지 제85항 중 어느 한 항에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 HNH 도메인을 포함하는, 방법.

청구항 87

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 N-말단 단편 및 상기 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는, 방법.

청구항 88

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 N-말단 단편 및 C-말단 단편 중 어느 것도 RuvC 도메인을 포함하지 않는, 방법.

청구항 89

제69항 내지 제77항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 하나 이상의 구조적 도메인에서 부분적 또는 완전한 결실을 포함하는, 방법.

청구항 90

제89항에 있어서, 상기 데아미나제가 상기 Cas9 폴리펩티드의 부분적 또는 완전한 결실 위치에 삽입되는, 방법.

청구항 91

제89항 또는 제90항에 있어서, 상기 결실이 RuvC 도메인 내에 있는, 방법.

청구항 92

제89항 또는 제90항에 있어서, 상기 결실이 HNH 도메인 내에 있는, 방법.

청구항 93

제89항 또는 제90항에 있어서, 상기 결실이 RuvC 도메인과 C-말단 도메인, L-I 도메인과 HNH 도메인, 또는 RuvC 도메인과 L-I 도메인을 연결하는, 방법.

청구항 94

제90항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017-1069 또는 이들의 상응하는 아미노산의 결실을 포함하는, 방법.

청구항 95

제90항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-872 또는 이들의 상응하는 아미노산의 결실을 포함하는, 방법.

청구항 96

제90항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-906 또는 이들의 상응하는 아미노산의 결실을 포함하는, 방법.

청구항 97

제69항 내지 제96항 중 어느 한 항에 있어서, 상기 데아미나제가 시티딘 데아미나제인, 방법.

청구항 98

제69항 내지 제96항 중 어느 한 항에 있어서, 상기 데아미나제가 아데노신 데아미나제인, 방법.

청구항 99

제69항 내지 제98항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 변형된 Cas9이고, 변경된 프로토스페이서-인접 모티프(PAM)에 대한 특이성을 갖는, 방법.

청구항 100

제69항 내지 제99항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 닉카아제인, 방법.

청구항 101

제69항 내지 제99항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 뉴클레아제 비활성인, 방법.

청구항 102

제67항 내지 제101항 중 어느 한 항에 있어서, 상기 접촉이 세포에서 수행되는 것인, 방법.

청구항 103

제102항에 있어서, 상기 세포가 포유류 세포 또는 인간 세포인, 방법.

청구항 104

제103항에 있어서, 상기 세포가 만능 세포인, 방법.

청구항 105

제102항 내지 제104항 중 어느 한 항에 있어서, 상기 세포가 생체내(*in vivo*) 또는 생체외에(*ex vivo*) 있는, 방법.

청구항 106

제67항 내지 제101항 중 어느 한 항에 있어서, 상기 접촉이 세포의 집단에서 수행되는, 방법.

청구항 107

제1-6항에 있어서, 상기 세포의 집단이 포유류 세포 또는 인간 세포인, 방법.

청구항 108

대상체에서 유전적 상태를 치료하는 방법으로서, 상기 방법은 상기 대상체에게 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하는 융합 단백질 또는 상기 융합 단배질을 코딩하는 폴리뉴클레오티드, 및 가이드 핵산 서열 또는 상기 가이드 핵산 서열을 코딩하는 폴리뉴클레오티드를 투여하는 단계를 포함하며, 상기 가이드 핵산 서열은 상기 융합 단백질이 상기 대상체에서 표적 폴리뉴클레오티드 서열 중의 표적 핵염기를 탈아미노화시키도록 안내하고, 그로 인해 상기 유전적 상태를 치료하는, 방법.

청구항 109

대상체에서 유전적 상태를 치료하는 방법으로서, 상기 방법은 상기 대상체에게 Cas9 폴리펩티드의 가요성 루프 내에 삽입된 데아미나제를 포함하는 융합 단백질을 투여하는 단계를 포함하며, 상기 융합 단백질은 NH_2-[Cas9의 N-말단]-[데아미나제]-[Cas9의 C-말단 단편]-COOH 구조를 포함하며, 여기서, "]-["의 각 경우는 선택적 링커이고, 상기 융합 단백질의 데아미나제는 상기 대상체에서 상기 표적 뉴클레오티드 서열 중의 상기 표적 핵염기를 탈아미노화시킴으로써, 상기 유전적 상태를 치료하는, 방법.

청구항 110

제108항 또는 제109항에 있어서, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단이 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함하는, 방법.

청구항 111

제108항 내지 제110항 중 어느 한 항에 있어서, 상기 대상체에게 가이드 핵산 서열을 투여하여 상기 표적 핵염기의 탈아미노화를 수행하는 단계를 추가로 포함하는, 방법.

청구항 112

제108항 내지 제111항 중 어느 한 항에 있어서, 상기 표적 핵염기가 상기 유전적 상태와 관련된 돌연변이를 포함하는, 방법.

청구항 113

제112항에 있어서, 상기 표적 핵염기의 탈아미노화가 상기 표적 핵염기를 야생형 핵염기로 대체하는, 방법.

청구항 114

제112항에 있어서, 상기 표적 핵염기의 탈아미노화가 상기 표적 핵염기를 비-야생형 핵염기로 대체하고, 상기 표적 핵염기의 탈아미노화가 상기 유전적 상태의 증상을 개선하는, 방법.

청구항 115

제108항 내지 제111항 중 어느 한 항에 있어서, 상기 표적 폴리뉴클레오티드 서열이 상기 표적 핵염기 이외의 핵염기에서 상기 유전적 상태와 관련된 돌연변이를 포함하는, 방법.

청구항 116

제115항에 있어서, 상기 표적 핵염기의 탈아미노화가 상기 유전적 상태의 증상을 개선하는 것인, 방법.

청구항 117

제108항 내지 제116항 중 어느 한 항에 있어서, 상기 표적 핵염기가 상기 표적 폴리뉴클레오티드 서열의 PAM 서열로부터 떨어져 있는 1개 내지 20개 핵염기인, 방법.

청구항 118

제117항에 있어서, 상기 표적 핵염기가 PAM 서열의 상류에 있는 2개 내지 12개 핵염기인, 방법.

청구항 119

제108항 내지 제118항 중 어느 한 항에 있어서, 상기 융합 단백질의 데아미나제가 상기 표적 핵염기를 탈아미노화시킬 때, 상기 가요성 루프가 상기 표적 핵염기에 근접한 아미노산을 포함하는, 방법.

청구항 120

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 가요성 루프가 서열번호: 1에서 번호매겨진 것과 같은 위치 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300에서의 아미노산 잔기로 이루어진 군으로부터 선택된 영역, 또는 이들의 상응하는 영역을 포함하는, 방법.

청구항 121

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249 또는 이들의 상응하는 아미노산 위치 사이에 삽입되는, 방법.

청구항 122

제121항에 있어서, 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248, 또는 이들의 상응하는 아미노산 위치 사이에 삽입되는, 방법.

청구항 123

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070 또는 1247-1248 또는 이들의 상응하는 아미노산 위치 사이에 삽입되는, 방법.

청구항 124

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 N-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 상기 Cas9 폴리펩티드의 아미노산 잔기 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 및/또는 1248-1297, 또는 이들의 상응하는 잔기를 포함하는, 방법.

청구항 125

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 C-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 상기 Cas9 폴리펩티드의 아미노산 잔기 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 및/또는 538-568, 또는 이들의 상응하는 잔기를 포함하는, 방법.

청구항 126

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드의 상기 N 말단 단편 또는 상기 C 말단 단편이 상기 표적 폴리뉴클레오티드 서열에 결합하는, 방법.

청구항 127

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 RuvC 도메인을 포함하는, 방법.

청구항 128

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 HNH 도메인을 포함하는, 방법.

청구항 129

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 N-말단 단편 및 상기 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는, 방법.

청구항 130

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 N-말단 단편 및 상기 C-말단 단편 중 어느 것도 RuvC 도메

인을 포함하지 않는, 방법.

청구항 131

제108항 내지 제119항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 하나 이상의 구조적 도메인에서 부분적 또는 완전한 결실을 포함하는, 방법.

청구항 132

제131항에 있어서, 상기 데아미나제가 상기 Cas9 폴리펩티드의 부분적 또는 완전한 결실 위치에 삽입되는, 방법.

청구항 133

제131항 또는 제132항에 있어서, 상기 결실이 RuvC 도메인 내에 있는, 방법.

청구항 134

제131항 또는 제132항에 있어서, 상기 결실이 HNH 도메인 내에 있는, 방법.

청구항 135

제131항 또는 제132항에 있어서, 상기 결실이 RuvC 도메인과 C-말단 도메인, L-I 도메인과 HNH 도메인, 또는 RuvC 도메인과 L-I 도메인을 연결하는, 방법.

청구항 136

제131항 또는 제132항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017-1069 또는 이들의 상응하는 아미노산의 결실을 포함하는, 방법.

청구항 137

제131항 또는 제132항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-872 또는 이들의 상응하는 아미노산의 결실을 포함하는, 방법.

청구항 138

제131항 또는 제132항에 있어서, 상기 Cas9 폴리펩티드가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-906 또는 이의 상응하는 아미노산의 결실을 포함하는, 방법.

청구항 139

제108항 내지 제138항 중 어느 한 항에 있어서, 상기 데아미나제가 시티딘 데아미나제인, 방법.

청구항 140

제108항 내지 제138항 중 어느 한 항에 있어서, 상기 데아미나제가 아데노신 데아미나제인, 방법.

청구항 141

제108항 내지 제140항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 변형된 Cas9이고, 변경된 PAM에 대한 특이성을 갖는, 방법.

청구항 142

제108항 내지 제141항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 닉카아제인, 방법.

청구항 143

제108항 내지 제142항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 뉴클레아제 비활성인, 방법.

청구항 144

제108항 내지 제143항 중 어느 한 항에 있어서, 상기 대상체가 포유동물인, 방법.

청구항 145

제108항 내지 제144항 중 어느 한 항에 있어서, 상기 대상체가 인간인, 방법.

청구항 146

복수의 융합 단백질을 포함하는 최적화된 염기 편집을 위한 단백질 라이브러리로서, 상기 복수의 융합 단백질 각각은 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하고, 상기 융합 단백질 각각의 N-말단 단편은 상기 나머지 복수의 융합 단백질의 N-말단 단편과 상이하거나, 상기 융합 단백질 각각의 C-말단 단편은 상기 나머지 복수의 융합 단백질의 C-말단 단편과 상이하며, 상기 융합 단백질 각각의 데아미나제는 표적 폴리뉴클레오티드 서열 중의 프로토스페이서 인접 모티프(PAM) 서열에 근접한 표적 핵염기를 탈아미노화시키며, 상기 N 말단 단편 또는 상기 C 말단 단편은 상기 표적 폴리뉴클레오티드 서열에 결합하는, 단백질 라이브러리.

청구항 147

제146항에 있어서, 상기 PAM 서열로부터 1 내지 20개의 핵염기 떨어져 있는 각각의 핵염기에 대해, 상기 복수의 융합 단백질 중 적어도 하나가 상기 핵염기를 탈아미노화시키는, 단백질 라이브러리.

청구항 148

제147항에 있어서, 상기 복수의 융합 단백질 각각의 상기 Cas9 폴리펩티드의 N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단이 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함하는, 단백질 라이브러리.

청구항 149

제146항 내지 제148항 중 어느 한 항에 있어서, 상기 복수의 융합 단백질 중 적어도 하나가 서열번호: 1의 N 말단 또는 C 말단에 융합된 상기 데아미나제를 포함하는 단부 말단 융합 단백질과 비교하여 더 낮은 표적-이탈 탈아미노화로 상기 표적 핵염기를 탈아미노화시키는, 단백질 라이브러리.

청구항 150

제146항 내지 제149항 중 어느 한 항에 있어서, 상기 복수의 융합 단백질 중 적어도 하나가 상기 PAM 서열의 상류에 있는 표적 핵염기 2 내지 12개 핵염기를 탈아미노화시키는, 단백질 라이브러리.

청구항 151

제146항 내지 제150항 중 어느 한 항에 있어서, 상기 융합 단백질이 상기 표적 핵염기를 탈아미노화시킬 때, 상기 복수의 융합 단백질의 N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단이 상기 표적 핵염기에 근접한 아미노산을 포함하는, 단백질 라이브러리.

청구항 152

제146항 내지 제150항 중 어느 한 항에 있어서, 상기 융합 단백질 중의 적어도 하나의 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249 또는 이들의 상응하는 아미노산 위치 사이에 있는, 단백질 라이브러리.

청구항 153

제152항에 있어서, 상기 융합 단백질 중의 적어도 하나의 상기 데아미나제가 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248, 또는 이들의 상응하는 아미노산 위치 사이에 있는, 단백질 라이브러리.

청구항 154

제146항 내지 제150항 중 어느 한 항에 있어서, 상기 융합 단백질 중의 적어도 하나의 상기 데아미나제가 서열

번호: 1에서 번호매겨진 것과 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070 또는 1247-1248 또는 이들의 상응하는 아미노산 위치 사이에 있는, 단백질 라이브러리.

청구항 155

제146항 내지 제154항 중 어느 한 항에 있어서, 데아미나제가 아데노신 데아미나제인 단백질 라이브러리.

청구항 156

제146항 내지 제154항 중 어느 한 항에 있어서, 상기 데아미나제가 시티딘 데아미나제인, 단백질 라이브러리.

청구항 157

제146항 내지 제156항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 스트렙토코커스 피오게네스(*Streptococcus pyogenes*) Cas9(SpCas9), 스타필로코커스 아우레우스(Staphylococcus aureus) Cas9(SaCas9), 스트렙토코커스 써모필러스(*Streptococcus thermophilus*) 1 Cas9(St1Cas9), 또는 이들의 변이체인, 단백질 라이브러리.

청구항 158

제146항 내지 제157항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 변형된 Cas9이고 변경된 프로토스페이서-인접 모티프(PAM)에 대한 특이성을 갖는 것인, 단백질 라이브러리.

청구항 159

제146항 내지 제157항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 닉카아제인, 단백질 라이브러리.

청구항 160

제146항 내지 제157항 중 어느 한 항에 있어서, 상기 Cas9 폴리펩티드가 뉴클레아제 비활성인, 단백질 라이브러리.

발명의 설명

기 술 분 야

[0001] **상호 참조**

[0002] 본 출원은 2019년 1월 31일에 출원된 미국 가특허 출원 제62/799,702호의 이익을 주장하는 PCT 국제출원이며, 상기 출원의 전문은 참조로 본 명세서에 통합된다.

배 경 기 술

[0003] **발명의 배경**

[0004] 핵염기 편집기로 지칭되는, CRISPR-Cas 단백질의 정확한 표적화와 결합된 데아미나제는 표적 폴리뉴클레오티드에 특정 점 돌연변이를 도입하는 활성을 갖는다. 핵염기 편집기는 이중-가닥 DNA 절단을 도입하지 않고 염기 변화를 유도하며, 표적 A·T를 G·C로 전환하는 아데노신 염기 편집기와 표적 C·G를 T·A로 전환하는 시티딘 염기 편집기를 포함한다. 그러나, 세포에 핵염기 편집기를 도입하면 게놈-전체의(genome-wide) 불요성(spurious) 탈아미노화, 방관자 돌연변이, 및 표적 근위 편집을 포함하여, 원하지 않는 염기 편집기-관련 편집을 생성할 가능성이 있다. 불요성 탈아미노화 이벤트는 게놈 전체에서 발생할 수 있는데, 가이드 RNA에 의한 CRISPR-Cas 도메인의 프로그래밍을 통한 표적화된 염기 편집과 독립적으로 작용하는 염기 편집기 탈아미노화 도메인에 의해 촉매된다. 이론에 구속됨이 없이, 게놈-전체의 불요성 탈아미노화 이벤트는, 예를 들어 "DNA 호흡"으로 인해 단일 가닥 DNA 기질이 형성되는 곳에서, 또는 DNA 복제 분기점에서 발생할 가능성이 있다. 표적 근위 편집은 표적화된 서열 외부에서 발생하지만, 표적 영역의 ~200bp 상류 또는 하류 내에 있는 염기 편집 이벤트이다. 방관자 돌연변이는 원하는 표적 핵염기가 아닌 표적-적중 Cas9/sgRNA 가이드된 염기 편집 윈도우 내에서 발생하는 돌연변이이다. 방관자 돌연변이는 침묵 돌연변이(아미노산 변화 없음) 또는 비동의(non-synonymous) 돌연변이(아미노산 변화 있음) 중 어느 하나를 초래할 수 있다. 따라서, 비-표적 탈아미노화가 감소된 염기 편집기가 요구되

고 있다.

발명의 내용

[0005] **발명의 요약**

[0006] 이하에 기술된 바와 같이, 본 발명은, 예를 들어, 프로그래밍된, 표적-적중 탈아미노화와 비교하여, 비-표적(non-target) 탈아미노화가 감소된 핵염기 편집기 조성물과, 핵염기 편집기를 그러한 것으로 특성규명하기 위한 방법 및 분석을 특징으로 한다.

[0007] 본 발명에 의해 정의된 조성물 및 물품은 아래에 제공된 실시예와 관련하여 단리되거나 달리 제조되었다. 본 발명의 다른 특징 및 이점은 상세한 설명, 및 청구범위로부터 명백할 것이다.

[0008] 한 양상에서, Cas9 폴리펩티드의 가요성 루프 내에 삽입된 데아미나제를 포함하는 융합 단백질이 본 명세서에 제공되며, 융합 단백질은 다음 구조를 포함한다:

[0009] NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH, 여기서 "]-["의 각 경우는 선택적 링커이다.

[0010] 한 양상에서, Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하는 융합 단백질이 본 명세서에 제공되며, 상기 N-말단 단편의 C-말단 또는 상기 C-말단 단편의 N-말단은 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함한다.

[0011] 일부 실시형태에서, 융합 단백질의 데아미나제는 표적 폴리뉴클레오티드 서열에서 표적 핵염기를 탈아미노화시킨다. 일부 실시형태에서, 가요성 루프는 융합 단백질이 표적 핵염기를 탈아미노화시킬 때 표적 핵염기에 근접한 아미노산을 포함한다. 일부 실시형태에서, 가요성 루프(flexible loop)는 Cas9 폴리펩티드의 알파-나선 구조의 일부를 포함한다. 일부 실시형태에서, 표적 핵염기는 서열번호: 1의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 단부 말단(end terminus) 융합 단백질과 비교하여 더 낮은 표적-이탈(off-target) 탈아미노화로 탈아미노화된다.

[0012] 일부 실시형태에서, 표적 핵염기는 표적 폴리뉴클레오티드 서열 중의 프로토스페이서 인접 모티프(PAM) 서열로부터 1개 내지 20개 핵염기 떨어져 있다. 일부 실시형태에서, 표적 핵염기는 PAM 서열의 상류에 있는 2 내지 12개 핵염기이다. 일부 실시형태에서, 가요성 루프는 서열번호: 1에서 번호매겨진 것과 같은 위치 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300에서의 아미노산 잔기로 이루어진 군으로부터 선택된 영역, 또는 이의 상응하는 영역을 포함한다. 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-10521, 1040-10541, 1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249 또는 이의 상응하는 아미노산 위치 사이에 삽입된다.

[0013] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다.

[0014] 일부 실시형태에서, N-말단 단편은 서열번호: 1에서 번호매겨진 것과 같은 Cas의 폴리펩티드의 아미노산 잔기 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 및/또는 1248-1297, 또는 이의 상응하는 잔기를 포함한다. 일부 실시형태에서, C-말단 단편은 서열번호: 1에서 번호매겨진 것과 같은 Cas의 폴리펩티드의 아미노산 잔기 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 및/또는 538-568, 또는 이의 상응하는 잔기를 포함한다. 일부 실시형태에서, Cas9 폴리펩티드의 N 말단 단편 또는 C 말단 단편은 표적 폴리뉴클레오티드 서열에 결합한다.

[0015] 일부 실시형태에서, Cas9 폴리펩티드의 N-말단 단편 또는 C-말단 단편은 DNA 결합 도메인을 포함한다. 일부 실시형태에서, N-말단 단편 또는 C-말단 단편은 RuvC 도메인을 포함한다. 일부 실시형태에서, N-말단 단편 또는 C 말단 단편은 HNH 도메인을 포함한다. 일부 실시형태에서, N-말단 단편 및 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는다. 일부 실시형태에서, N-말단 단편 및 C-말단 단편 중 어느 것도 RuvC 도메인을 포함하지 않는다. 일부 실시형태에서, Cas9 폴리펩티드는 하나 이상의 구조적 도메인에 부분적 또는 완전한 결실을 포함한

다. 일부 실시형태에서, 데아미나제는 Cas9 폴리펩티드의 부분적 또는 완전한 결실 위치에 삽입된다.

[0016] 일부 실시형태에서, 결실은 RuvC 도메인 내에 있다. 일부 실시형태에서, 결실은 HNH 도메인 내에 있다. 일부 실시형태에서, 결실은 RuvC 도메인과 C-말단 도메인, L-I 도메인과 HNH 도메인, 또는 RuvC 도메인과 L-I 도메인을 연결한다. 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017-1069 또는 이의 상응하는 아미노산의 결실을 포함한다. 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-872 또는 이의 상응하는 아미노산의 결실을 포함한다. 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-906 또는 이의 상응하는 아미노산의 결실을 포함한다.

[0017] 한 양상에서, Cas9 폴리펩티드 내에 삽입된 데아미나제를 포함하는 융합 단백질이 본 명세서에 제공되며, 상기 융합 단백질은 다음 구조를 포함한다:

[0018] NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH, 여기서 "]-["의 각 경우는 선택적 링커이고, 상기 Cas9 폴리펩티드는 HNH 도메인의 완전한 결실을 포함하고, 상기 데아미나제는 상기 결실 위치에 삽입된다.

[0019] 일부 실시형태에서, N 말단 단편의 C 말단 아미노산은 서열번호: 1에서 번호매겨진 것과 같은 아미노산 791이다. 일부 실시형태에서, C 말단 단편의 N 말단 아미노산은 서열번호: 1에서 번호매겨진 것과 같은 아미노산 907이다. 일부 실시형태에서, C 말단 단편의 N 말단 아미노산은 서열번호: 1에서 번호매겨진 것과 같은 아미노산 873이다.

[0020] 한 양상에서, Cas9 폴리펩티드 내에 삽입된 데아미나제를 포함하는 융합 단백질이 본 명세서에 제공되며, 상기 융합 단백질은 다음 구조를 포함한다:

[0021] NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH, 여기서 "]-["의 각 경우는 선택적 링커이고, 상기 Cas9는 RuvC 도메인의 완전한 결실을 포함하고, 상기 데아미나제는 상기 결실 위치에 삽입된다.

[0022] 일부 실시형태에서, 데아미나제는 시티딘 데아미나제 또는 아데노신 데아미나제이다. 일부 실시형태에서, 시티딘 데아미나제는 APOBEC 시티딘 데아미나제, 활성화 유도 시티딘 데아미나제(AID), 또는 CDA이다. 일부 실시형태에서, APOBEC 데아미나제는 APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3E, APOBEC3F, APOBEC3G, APOBEC3H, 또는 APOBEC4이다. 일부 실시형태에서, APOBEC 데아미나제는 rAPOBEC1이다. 일부 실시형태에서, 상기 양상 중 어느 하나의 융합 단백질은 UGI 도메인을 추가로 포함한다.

[0023] 일부 실시형태에서, 아데노신 데아미나제는 TadA 데아미나제이다. 일부 실시형태에서, TadA 데아미나제는 변형된 TadA이다. 일부 실시형태에서, TadA 데아미나제는 TadA 7.10이다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 이량체이다. 일부 실시형태에서, TadA 이량체는 TadA 7.10 및 야생형 TadA를 포함한다. 일부 실시형태에서, 선택적 링커는 (SGGS)n, (GGGS)n, (GGGGS)n, (G)n, (EAAAAK)n, (GGS)n, SGSETPGTSESATPES, 또는 (XP)n 모티프, 또는 이들의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이다.

[0024] 일부 실시형태에서, Cas9 폴리펩티드의 N 말단 단편은 링커 없이 데아미나제에 융합된다. 일부 실시형태에서, Cas9의 C 말단 단편은 링커 없이 데아미나제에 융합된다. 일부 실시형태에서, 상기 양상 중 어느 하나의 융합 단백질은 추가의 촉매 도메인을 추가로 포함한다.

[0025] 일부 실시형태에서, 추가 촉매 도메인은 제2 데아미나제이다. 일부 실시형태에서, 제2 데아미나제는 융합 단백질의 N 말단 또는 C 말단에 융합된다. 일부 실시형태에서, 데아미나제는 시티딘 데아미나제 또는 아데노신 데아미나제이다. 일부 실시형태에서, 상기 양상 중 어느 하나의 융합 단백질은 핵 위치결정 신호(nuclear localization signal)를 추가로 포함한다. 일부 실시형태에서, 핵 위치결정 신호는 이분(bipartite) 핵 위치결정 신호이다. 일부 실시형태에서, Cas9 폴리펩티드는 스트렙토코커스 피오게네스(*Streptococcus pyogenes*) Cas9(SpCas9), 스타필로코커스 아우레우스(*Staphylococcus aureus*) Cas9(SaCas9), 스트렙토코커스 써모필러스(*Streptococcus thermophilus*) 1 Cas9(St1Cas9), 또는 이의 변이체이다. 일부 실시형태에서, Cas9 폴리펩티드는 변형된 Cas9이고, 변경된 PAM에 대한 특이성을 갖는다. 일부 실시형태에서, Cas9 폴리펩티드는 닉카아제(nickase)이다. 일부 실시형태에서, Cas9 폴리펩티드는 뉴클레아제 비활성(inactive)이다. 일부 실시형태에서, 상기 양상 중 어느 하나의 융합 단백질은 가이드 핵산 서열과 복합체화하여 표적 핵염기의 탈아미노화를 수행한다. 일부 실시형태에서, 융합 단백질은 표적 폴리뉴클레오티드와 추가로 복합체화한다.

[0026] 본 명세서에서는 상기 양상 중 어느 하나의 융합 단백질을 코딩하는 폴리뉴클레오티드를 제공한다.

[0027] 본 명세서에서는 상술한 폴리뉴클레오티드를 포함하는 발현 벡터를 제공한다.

[0028] 일부 실시형태에서, 발현 벡터는 포유동물 발현 벡터이다. 일부 실시형태에서, 벡터는 아데노-관련 바이러스(AAV), 레트로바이러스 벡터, 아데노바이러스 벡터, 렌티바이러스 벡터, 센다이 바이러스 벡터, 및 헤르페스바이러스 벡터로 이루어진 군으로부터 선택된 바이러스 벡터이다. 일부 실시형태에서, 벡터는 프로모터를 포함한다.

[0029] 본 명세서에서는 상기 양상 중 어느 하나의 융합 단백질, 상술한 폴리뉴클레오티드, 또는 상술한 벡터를 포함하는 세포를 제공한다.

[0030] 일부 실시형태에서, 세포는 박테리아 세포, 식물 세포, 곤충 세포, 인간 세포, 또는 포유류 세포이다.

[0031] 본 명세서에서는 상기 양상 중 어느 하나의 융합 단백질, 상술한 폴리뉴클레오티드, 또는 상술한 벡터를 포함하는 키트를 제공한다.

[0032] 폴리뉴클레오티드 서열을 상기 양상 중 어느 하나의 융합 단백질과 접촉시키는 단계를 포함하는 염기 편집 방법이 본 명세서에 제공되며, 상기 융합 단백질의 데아미나제는 폴리뉴클레오티드 중의 핵염기를 탈아미노시키며, 그로 인해 폴리뉴클레오티드 서열을 편집한다.

[0033] 일부 실시형태에서, 방법은 표적 폴리뉴클레오티드 서열을 가이드 핵산 서열과 접촉시켜 표적 핵염기의 탈아미노화를 수행하는 단계를 추가로 포함한다.

[0034] 한 양상에서, 본 명세서에서는 표적 폴리뉴클레오티드 서열에서 표적 핵염기를 편집하는 방법을 제공하고, 이 방법은 표적 폴리뉴클레오티드 서열을 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하는 융합 단백질과 접촉시키는 단계를 포함하며, 상기 융합 단백질의 데아미나제는 표적 폴리뉴클레오티드 서열에서 표적 핵염기를 탈아미노화시키며, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단이 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함한다.

[0035] 표적 폴리뉴클레오티드 서열에서 표적 핵염기를 편집하는 방법이 본 명세서에 제공되며, 이 방법은 표적 폴리뉴클레오티드 서열을 Cas9 폴리펩티드의 가요성 루프 내에 삽입된 데아미나제를 포함하는 융합 단백질과 접촉시키는 단계를 포함하며, 상기 융합 단백질은 NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH 구조를 포함하며, 여기서, "]-["의 각 경우는 선택적 링커이고, 상기 융합 단백질의 데아미나제는 상기 표적 폴리뉴클레오티드 서열에서 상기 표적 핵염기를 탈아미노화시킨다.

[0036] 일부 실시형태에서, 방법은 표적 폴리뉴클레오티드 서열을 가이드 핵산 서열과 접촉시켜 표적 핵염기의 탈아미노화를 수행하는 단계를 추가로 포함한다. 일부 실시형태에서, 가이드 핵산 서열은 표적 폴리뉴클레오티드 서열의 프로토스페이서 서열에 상보적인 스페이서 서열을 포함하며, 그로 인해 R-루프를 형성한다. 일부 실시형태에서, 표적 핵염기는 서열번호: 1의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 단부 말단 방법과 비교하여 더 낮은 표적-이탈 탈아미노화로 탈아미노화된다. 일부 실시형태에서, 융합 단백질의 데아미나제는 단백질은 R 루프 영역 내에서 2개 이하의 핵염기를 탈아미노화시킨다. 일부 실시형태에서, 표적 핵염기는 표적 폴리뉴클레오티드 서열 중의 PAM 서열로부터 1개 내지 20개 핵염기 떨어져 있다. 일부 실시형태에서, 표적 핵염기는 PAM 서열의 상류에 있는 2 내지 12개 핵염기이다.

[0037] 일부 실시형태에서, 가요성 루프는 융합 단백질의 데아미나제가 표적 핵염기를 탈아미노화시킬 때 표적 핵염기에 근접한 아미노산을 포함한다. 일부 실시형태에서, 가요성 루프는 서열번호: 1에서 번호매겨진 것과 같은 위치 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300에서의 아미노산 잔기로 이루어진 군으로부터 선택된 영역, 또는 이의 상응하는 영역을 포함한다. 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-10521, 1040-10541, 1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다.

[0038] 일부 실시형태에서, N-말단 단편은 서열번호: 1에서 번호매겨진 것과 같은 Cas9의 폴리펩티드의 아미노산 잔기 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 및/또는 1248-1297, 또는 이의 상응하는 잔기를 포함한다. 일부 실시형태에서, C-말단 단편은 서열번호: 1에서 번호매겨진 것과 같은 Cas9의 폴리펩티드의 아미노산 잔기 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 및/또는 538-568, 또는 이의 상응하는 잔기를 포함한다. 일부 실시형태에서, Cas9 폴리펩티드의 N 말단 단편 또는 C 말단 단편은 표적 폴리뉴클레오티드 서열에 결합한다. 일부 실시형태에서, N-말단 단편 또는 C-말단 단편은 RuvC 도메인을 포함한다. 일부 실시형태에서, N-말단 단편 또는 C-말단 단편은 HNH 도메인을 포함한다. 일부 실시형태에서, N-말단 단편 및 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는다. 일부 실시형태에서, N-말단 단편 및 C-말단 단편 중 어느 것도 RuvC 도메인을 포함하지 않는다.

[0039] 일부 실시형태에서, Cas9 폴리펩티드는 하나 이상의 구조적 도메인에 부분적 또는 완전한 결실을 포함한다. 일부 실시형태에서, 데아미나제는 Cas9 폴리펩티드의 부분적 또는 완전한 결실 위치에 삽입된다. 일부 실시형태에서, 결실은 RuvC 도메인 내에 있다. 일부 실시형태에서, 결실은 HNH 도메인 내에 있다. 일부 실시형태에서, 결실은 RuvC 도메인과 C-말단 도메인, L-I 도메인과 HNH 도메인, 또는 RuvC 도메인과 L-I 도메인을 연결한다. 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017-1069 또는 이의 상응하는 아미노산의 결실을 포함한다.

[0040] 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-872 또는 이의 상응하는 아미노산의 결실을 포함한다. 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 792-906 또는 이의 상응하는 아미노산의 결실을 포함한다. 일부 실시형태에서, 데아미나제는 시티딘 데아미나제이다. 일부 실시형태에서, 데아미나제는 아데노신 데아미나제이다. 일부 실시형태에서, Cas9 폴리펩티드는 변형된 Cas9이고, 변경된 프로토스페이서-인접 모티프(PAM)에 대한 특이성을 갖는다. 일부 실시형태에서, Cas9 폴리펩티드는 닉카아제이다. 일부 실시형태에서, Cas9 폴리펩티드는 뉴클레아제 비활성이다.

[0041] 일부 실시형태에서, 접촉은 세포에서 수행된다. 일부 실시형태에서, 세포는 포유류 세포 또는 인간 세포이다. 일부 실시형태에서, 세포는 만능 세포(pluripotent cell)이다. 일부 실시형태에서, 세포는 생체내(*in vivo*) 또는 생체외에(*ex vivo*) 있다. 일부 실시형태에서, 접촉은 세포 집단에서 수행된다. 일부 실시형태에서, 세포 집단은 포유류 세포 또는 인간 세포이다.

[0042] 한 양상에서, 대상체에서 유전적 상태를 치료하는 방법이 본 명세서에 제공되며, 이 방법은 대상체에게 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하는 융합 단백질 또는 상기 융합 단백질을 코딩하는 폴리뉴클레오티드, 및 가이드 핵산 서열 또는 상기 가이드 핵산 서열을 코딩하는 폴리뉴클레오티드를 투여하는 단계를 포함하고, 상기 가이드 핵산 서열은 상기 융합 단백질이 상기 대상체의 표적 폴리뉴클레오티드 서열에서 표적 핵염기를 탈아미노화시키도록 유도하며, 그로 인해 상기 유전적 질환을 치료한다.

[0043] 대상체의 유전적 상태를 치료하는 방법이 본 명세서에 제공되며, 이 방법은 상기 대상체에 Cas9 폴리펩티드의 가요성 루프 내에 삽입된 데아미나제를 포함하는 융합 단백질을 투여하는 단계를 포함하고, 상기 융합 단백질은 NH_2-[Cas9의 N-말단 단편]-[데아미나제]-[Cas9의 C-말단 단편]-COOH 구조를 포함하며, "]-["의 각 경우는 선택적 링커이고, 상기 융합 단백질의 데아미나제는 상기 대상체의 폴리뉴클레오티드 서열에서 상기 표적 핵염기를 탈아미노화시키며, 그로 인해 상기 유전적 상태를 치료한다.

[0044] 일부 실시형태에서, N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단은 Cas9 폴리펩티드의 가요성 루프의 일부를 포함한다. 일부 실시형태에서, 방법은 대상체에게 가이드 핵산 서열을 투여하여 표적 핵염기의 탈아미노화를 수행하는 단계를 추가로 포함한다. 일부 실시형태에서, 표적 핵염기는 유전적 상태와 관련된 돌연변이를 포함한다. 일부 실시형태에서, 표적 핵염기의 탈아미노화는 표적 핵염기를 야생형 핵염기로 대체한다. 일부 실시형태에서, 표적 핵염기의 탈아미노화는 표적 핵염기를 비야생형 핵염기로 대체하고, 상기 표적 핵염기의 탈아미노화는 상기 유전적 상태의 증상을 개선한다.

[0045] 일부 실시형태에서, 표적 폴리뉴클레오티드 서열은 표적 핵염기 이외의 핵염기에서의 유전적 상태와 관련된 돌연변이를 포함한다. 일부 실시형태에서, 표적 핵염기의 탈아미노화는 유전적 상태의 증상을 개선한다. 일부 실시형태에서, 표적 핵염기는 표적 폴리뉴클레오티드 서열 중의 PAM 서열로부터 1 내지 20개의 핵염기 떨어져 있다. 일부 실시형태에서, 표적 핵염기는 PAM 서열의 상류에 있는 2 내지 12개 핵염기이다. 일부 실시형태에서, 가요성 루프는 융합 단백질의 데아미나제가 표적 핵염기를 탈아미노화시킬 때 표적 핵염기에 근접한 아미노산을 포함한다.

[0046] 일부 실시형태에서, 가요성 루프는 서열번호: 1에서 번호매겨진 것과 같은 위치 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300에서의 아미노산 잔기로 이루어진 군으로부터 선택된 영역, 또는 이의 상응하는 영역을 포함한다.

[0047] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-10521, 1040-10541, 1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다.

[0048] 일부 실시형태에서, N-말단 단편은 서열번호: 1에서 번호매겨진 것과 같은 Cas의 폴리펩티드의 아미노산 잔기 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 및/또는 1248-1297, 또는 이의 상응하는 잔기를 포함한다. 일부 실시형태에서, C-말단 단편은 서열번호: 1에서 번호매겨진 것과 같은 Cas의 폴리펩티드의 아미노산 잔기 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 및/또는 538-568, 또는 이의 상응하는 잔기를 포함한다. 일부 실시형태에서, Cas9 폴리펩티드의 N 말단 단편 또는 C 말단 단편은 표적 폴리뉴클레오티드 서열에 결합한다. 일부 실시형태에서, N-말단 단편 또는 C-말단 단편은 RuvC 도메인을 포함한다. 일부 실시형태에서, N-말단 단편 또는 C-말단 단편은 HNH 도메인을 포함한다.

[0049] 일부 실시형태에서, N-말단 단편 및 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는다. 일부 실시형태에서, N-말단 단편 및 C-말단 단편 중 어느 것도 RuvC 도메인을 포함하지 않는다. 일부 실시형태에서, Cas9 폴리펩티드는 하나 이상의 구조적 도메인에 부분적 또는 완전한 결실을 포함한다. 일부 실시형태에서, 데아미나제는 Cas9 폴리펩티드의 부분적 또는 완전한 결실 위치에 삽입된다. 일부 실시형태에서, 결실은 RuvC 도메인 내에 있다. 일부 실시형태에서, 결실은 HNH 도메인 내에 있다. 일부 실시형태에서, 결실은 RuvC 도메인과 C-말단 도메인, L-I 도메인과 HNH 도메인, 또는 RuvC 도메인과 L-I 도메인을 연결한다. 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017-1069 또는 이의 상응하는 아미노산의 결실을 포함한다.

[0050] 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1로 번호매겨진 것과 같은 아미노산 792-872 또는 이의 상응하는 아미노산의 결실을 포함한다. 일부 실시형태에서, Cas9 폴리펩티드는 서열번호: 1로 번호매겨진 것과 같은 아미노산 792-906 또는 이의 상응하는 아미노산의 결실을 포함한다. 일부 실시형태에서, 데아미나제는 시티딘 데아미나제이다. 일부 실시형태에서, 데아미나제는 아데노신 데아미나제이다. 일부 실시형태에서, Cas9 폴리펩티드는 변형된 Cas9이고, 변경된 PAM에 대한 특이성을 갖는다. 일부 실시형태에서, Cas9 폴리펩티드는 닉카아제이다. 일부 실시형태에서, Cas9 폴리펩티드는 뉴클레아제 비활성이다. 일부 실시형태에서, 대상체는 포유동물이다. 일부 실시형태에서, 대상은 인간이다.

[0051] 복수의 융합 단백질을 포함하는 최적화된 염기 편집을 위한 단백질 라이브러리가 본 명세서에 제공되며, 상기 복수의 융합 단백질 각각은 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하고, 상기 N-융합 단백질 각각의 말단 단편은 상기 나머지 복수의 융합 단백질의 N-말단 단편과 상이하거나, 또는 상기 융합 단백질 각각의 C-말단 단편은 상기 나머지 복수의 융합 단백질의 C-말단 단편과 상이하며, 상기 융합 단백질 각각의 데아미나제는 표적 폴리뉴클레오티드 서열 내의 프로토스페이서 인접 모티프(PAM) 서열에 근접한 표적 핵염기를 탈아미노화시키며, 상기 N 말단 단편 또는 C 말단 단편은 상기 표적 폴리뉴클레오티드 서열에 결합한다.

[0052] 일부 실시형태에서, PAM 서열로부터 1개 내지 20개 핵염기 떨어져 있는 각 핵염기에 대해, 복수의 융합 단백질 중 적어도 하나는 핵염기를 탈아미노화시킨다. 일부 실시형태에서, 복수의 융합 단백질 각각의 Cas9 폴리펩티드의 N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단은 Cas9 폴리펩티드의 가요성 루프의 일부를 포함한다. 일부 실시형태에서, 복수의 융합 단백질 중 적어도 하나는 서열번호: 1의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 단부 말단 융합 단백질과 비교하여 더 낮은 표적-이탈 탈아미노화로 표적 핵염기를 탈아미노화시킨다. 일부 실시형태에서, 복수의 융합 단백질 중 적어도 하나는 PAM 서열의 상류에 있는 표적 핵염기 2 내지 12개 핵염기를 탈아미노화시킨다. 일부 실시형태에서, 복수의 융합 단백질의 N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단은 융합 단백질이 표적 핵염기를 탈아미노화시킬 때 표적 핵염기에 근접한 아미노산을 포함한다.

[0053] 일부 실시형태에서, 융합 단백질 중의 적어도 하나의 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미

노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249 또는 이의 상응하는 아미노산 위치 사이에 있다. 일부 실시형태에서, 융합 단백질 중의 적어도 하나의 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1040-1041, 1068-1069, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 있다. 일부 실시형태에서, 융합 단백질 중의 적어도 하나의 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 1016-1017, 1023-1024, 1029-1030, 1040-1041, 1069-1070 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 있다. 일부 실시형태에서, 데아미나제는 아데노신 데아미나제이다. 일부 실시형태에서, 데아미나제는 시티딘 데아미나제이다.

[0054] 일부 실시형태에서, Cas9 폴리펩티드는 스트렙토코커스 피오게네스 Cas9(SpCas9), 스타필로코커스 아우레우스 Cas9(SaCas9), 스트렙토코커스 써모필러스 1 Cas9(St1Cas9), 또는 이의 변이체이다. 일부 실시형태에서, Cas9 폴리펩티드는 변형된 Cas9이고, 변경된 프로토스페이서-인접 모티프(PAM)에 대한 특이성을 갖는다. 일부 실시형태에서, Cas9 폴리펩티드는 닉카아제이다. 일부 실시형태에서, Cas9 폴리펩티드는 뉴클레아제 비활성이다.

[0055] **정의**

[0056] 달리 정의되지 않는 한, 본 명세서에서 사용된 모든 기술 및 과학 용어는 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 의미를 갖는다. 다음 참고 문헌은 본 발명에 사용된 많은 용어에 대한 일반적인 정의를 제공한다: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); 및 Hale & Marham, The Harper Collins Dictionary of Biology (1991). 본 명세서에 사용된, 다음 용어는, 달리 명시되지 않는 한 아래에 부여된 의미를 갖는다.

[0057] 달리 정의되지 않는 한, 본 명세서에서 사용된 모든 기술 및 과학 용어는 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 의미를 갖는다. 다음 참고 문헌은 본 발명에 사용된 많은 용어에 대한 일반적인 정의를 제공한다: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). 본 명세서에 사용된, 다음 용어는, 달리 명시되지 않는 한 아래에 부여된 의미를 갖는다.

[0058] "아데노신 데아미나제"는 아데닌 또는 아데노신의 가수분해 탈아미노화를 촉매할 수 있는 폴리펩티드 또는 이의 단편을 의미한다. 일부 실시형태에서, 데아미나제 또는 데아미나제 도메인은 아데노신의 이노신으로의 또는 데옥시아데노신의 데옥시이노신으로의 가수분해 탈아미노화를 각각 촉매할 수 있는 아데노신 데아미나제이다. 일부 실시형태에서, 아데노신 데아미나제는 데옥시리보핵산(DNA)에서 아데닌 또는 아데노신의 가수분해 탈아미노화를 촉매한다. 본 명세서에서 제공되는 아데노신 데아미나제(예를 들어, 조작된 아데노신 데아미나제, 진화된(evolved) 아데노신 데아미나제)는 임의의 유기체, 예컨대 박테리아로부터 유래될 수 있다. 일부 실시형태에서, 데아미나제 또는 데아미나제 도메인은 유기체로부터의 자연적으로 발생하는 데아미나제의 변이체이다. 일부 실시양태에서, 데아미나제 또는 데아미나제 도메인은 자연적으로 발생하지 않는다. 예를 들어, 일부 실시형태에서, 데아미나제 또는 데아미나제 도메인은 자연적으로 발생하는 데아미나제에 대해 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%, 적어도 99.5% 동일하다. 일부 실시형태에서, 아데노신 데아미나제는 박테리아, 예컨대 E. 콜라이(*E. coli*), S. 아우레우스(*S. aureus*), S. 티피(*S. typhi*), S. 퓨트레파시엔스(*S. putrefaciens*), H. 인플루엔자(*H. influenzae*), 또는 C. 크레센투스(*C. crescentus*)로부터 유래된다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 데아미나제이다. 일부 실시형태에서, TadA 데아미나제는 E. 콜라이(*E. coli*) TadA(ecTadA) 데아미나제 또는 이의 단편이다.

[0059] 예를 들어, 절단된(truncated) ecTadA는 전체 길이 ecTadA에 비해 하나 이상의 N-말단 아미노산이 결여될 수 있다. 일부 실시형태에서, 절단된 ecTadA는 전체 길이 ecTadA에 비해 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 6개, 17개, 18개, 19개, 또는 20개 N-말단 아미노산 잔기가 결여될 수 있다. 일부 실시형태에서, 절단된 ecTadA는 전체 길이 ecTadA에 비해 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 6개, 17개, 18개, 19개, 또는 20개 C-말단 아미노산 잔기가 결여될 수 있다. 일부 실시형태에서, ecTadA 데아미나제는 N-말단 메티오닌을 포함하지 않는다. 일부 실시형태

에서, TadA 데아미나제는 N-말단 절단된 TadA이다. 특정 실시형태에서, TadA는, 그 전문이 본 명세서에 참조로 통합되는, PCT/US2017/045381에 기재된 TadA 중 임의의 하나이다.

[0060] 특정 실시형태에서, 아데노신 데아미나제는, "TadA 참조 서열"로 지칭되는, 다음 아미노산 서열을 포함한다:

[0061] MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPT AHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKT

[0062] GAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD .

[0063] 일부 실시형태에서, TadA 데아미나제는 전체 길이 E. 콜라이 TadA 데아미나제이다. 예를 들어, 특정 실시형태에서, 아데노신 데아미나제는 다음 아미노산 서열을 포함한다:

[0064] MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEG WNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIG RVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEI KAQKKAQSSTD.

[0065] 그러나, 본 출원에 유용한 추가의 아데노신 데아미나제는 당업자에게 명백할 것이며, 본 개시의 범위 내에 있다는 것을 이해해야 한다. 예를 들어, 아데노신 데아미나제는 tRNA(AD AT)에 작용하는 아데노신 데아미나제의 상동체(homolog)일 수 있다. 예시적인 AD AT 상동체는, 제한됨이 없이, 하기를 포함한다:

[0066] 스타필로코커스 아우레우스(*Staphylococcus aureus*) TadA:

[0067] MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAH AEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCS GS LMNLLQQS NFNHRAIVDKG VLKE AC S TLLTTFFKNLRANKKS TN

[0068] 바실러스 서브틸리스(*Bacillus subtilis*) TadA:

[0069] MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEML VIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGC S GTLMN LLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE

[0070] 살모넬라 티피무리움(*Salmonella typhimurium*)(*S. typhimurium*) TadA:

[0071] MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEG WNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIG RVVFGARDAKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIK ALKKADRAEGAGPAV

[0072] 슈와넬라 퓨트레파시엔스(*Shewanella putrefaciens*)(*S. putrefaciens*) TadA:

[0073] MDE YWMQVAMQM AEKAEAAGE VPVGA VLVKDGQQIATGYNLS IS QHDPT AHAEI

[0074] LCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGT VVNLLQHPAFNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE

[0075] 해모필루스 인플루엔자(*Haemophilus influenzae*) F3031(*H. influenzae*) TadA:

[0076] MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQS DPT AH AEIIALRNG AKNIQN YRLLNS TLY VTLEPCTMC AG AILHS RIKRLVFG AS D YK TGAIGSRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSD K

[0077] 카울로박터 크레센투스(*Caulobacter crescentus*)(*C. crescentus*) TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAH
DPTAHAEIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADD
[0078] PKGGAVVHGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI

[0079] 지오박테 설푸레두센스(*Geobacter sulfurreducens*) (*G. sulfurreducens*) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSN
DPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDP
KGGAAGSLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALF
[0080] IDERKVPPEP

[0081] TadA7.10

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
[0082] GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0083] TadA7.10 또는 TadA7.10 변이체를 포함하는 예시적인 서열은 다음을 포함한다:

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLY
VTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVE
[0084] ITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0085] TadA7.10 CP65

[0086] TAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF

GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQ
VFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTL
[0087] AKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP

[0088] TadA7.10 CP83

YRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS
ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL
[0089] NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN

[0090] TadA7.10 CP136

MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETP
GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR
VIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
[0091] CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

[0092] TadA7.10 C-절단체(truncate)

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLY
VTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVE
[0093] ITEGILADECAALLCYFFRMPRQVFN

[0094] TadA7.10 C-절단체 2

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLY
VTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVE

[0095] ITEGILADECAALLCYFFRMPRQ

[0096] TadA7.10 델타59-66+C-절단체

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD

[0097] ECAALLCYFFRMPRQVFN

[0098] TadA7.10 델타 59-66

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD

[0099] ECAALLCYFFRMPRQVFNAQKKAQSSTD

[0100] "작용제(agent)"는 임의의 소분자 화학적 화합물, 항체, 핵산 분자, 또는 폴리펩티드, 또는 이의 단편을 의미한다.

[0101] "돌연변이를 변경하다"에 의해

[0102] "변경(alteration)"은 본 명세서에 기재된 것과 같은 표준 기술 공지 방법에 의해 검출되는 바와 같은 유전자 또는 폴리펩티드의 구조, 발현 수준 또는 활성의 변화를 의미한다. 본 명세서에 사용된 변경(예를 들어, 증가 또는 감소)은 발현 수준에서의 10% 변화, 25% 변화, 40% 변화, 및 발현 수준에서의 50% 또는 그 초과의 변화를 포함한다.

[0103] "유사체(analog)"는 동일하지 않지만 유사한 기능적 또는 구조적 특징을 갖는 분자를 의미한다. 예를 들어, 폴리뉴클레오티드 유사체는 상응하는 자연적으로 발생하는 폴리뉴클레오티드의 생물학적 활성을 유지하면서 자연적으로 발생하는 폴리뉴클레오티드에 비해 유사체의 기능을 향상시키는 특정 변형을 갖는다. 이러한 변형은 DNA에 대한 폴리뉴클레오티드의 친화도, 반감기, 및/또는 뉴클레아제 내성을 증가시킬 수 있다. 유사체는 비천연 뉴클레오티드 또는 아미노산을 포함할 수 있다.

[0104] 본 개시에서, "포함하다(comprises)", "포함하는(comprising)", "함유하는(containing)" 및 "가지는(having)" 등은 미국 특허법에서 이들에 부여된 의미를 가질 수 있고, "포함하다(includes)", "포함하는(including)" 등을 의미할 수 있고; "본질적으로 구성되는(consisting essentially of)" 또는 "본질적으로 구성되다(consists essentially)"는 마찬가지로 미국 특허법에 부여된 의미를 가지며, 이 용어는 인용된 것 이상의 존재로 인해 인용된 것의 기본 또는 신규 특성이 변경되지 않는 한, 인용된 것 이상의 존재를 허용하지만 선행 기술의 실시형태는 제외한다.

[0105] "염기 편집기(BE)" 또는 "핵염기 편집기(NBE)"는 폴리뉴클레오티드에 결합하고 핵염기 변형 활성을 갖는 작용제를 의미한다. 한 실시형태에서, 작용제는 염기 편집 활성을 갖는 도메인, 즉 핵산 분자(예를 들어, DNA) 내의 염기(예를 들어, A, T, C, G, 또는 U)를 변형할 수 있는 도메인을 포함하는 융합 단백질이다. 일부 실시형태에서, 염기 편집 활성을 갖는 도메인은 핵산 분자 내의 염기를 탈아미노화할 수 있다. 일부 실시형태에서, 염기 편집기는 DNA 분자 내의 염기를 탈아미노화할 수 있다. 일부 실시형태에서, 염기 편집기는 DNA 내의 시토신(C) 또는 아데노신을 탈아미노화할 수 있다. 일부 실시형태에서, 염기 편집기는 시티딘 염기 편집기(CBE)이다. 일부 실시형태에서, 염기 편집기는 아데노신 염기 편집기(ABE)이다. 일부 실시형태에서, 염기 편집기는 아데노신 염기 편집기(ABE) 및 시티딘 염기 편집기(CBE)이다. 일부 실시형태에서, 염기 편집기는 아데노신 데아미나제에 융합된 뉴클레아제-비활성 Cas9(dCas9)이다. 일부 실시형태에서, Cas9는 원형 치환체(circular permutant) Cas9(예를 들어, spCas9 또는 saCas9)이다. 원형 치환 Cas9는 당업계에 공지되어 있으며, 예를 들어, 문헌[Oakes et al., Cell 176, 254-267, 2019]에 설명되어 있다. 일부 실시형태에서, 염기 편집기는 염기 절제 복구의 억제제,

예를 들어, UGI 도메인에 융합된다. 일부 실시형태에서, 융합 단백질은 데아미나제에 융합된 Cas9 닉카아제 및 UGI 도메인과 같은 염기 절제 복구의 억제제를 포함한다. 다른 실시예에서, 염기 편집기는 무염기성(abasic) 염기 편집기이다.

[0106] 일부 실시형태에서, 아데노신 데아미나제는 TadA로부터 진화된다. 일부 실시형태에서, 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인은 CRISPR 관련(예를 들어, Cas 또는 Cpf1) 효소이다. 일부 실시형태에서, 염기 편집기는 데아미나제 도메인에 융합된 촉매적으로 멸실된 Cas9(dCas9)이다. 일부 실시형태에서, 염기 편집기는 데아미나제 도메인에 융합된 Cas9 닉카아제(nCas9)이다. 일부 실시형태에서, 데아미나제 도메인은 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인의 N-말단 또는 C-말단 단편이다. 일부 실시형태에서, 데아미나제는 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인의 N-말단 및 C-말단 단편에 의해 측접된다. 일부 실시형태에서, 데아미나제 도메인은 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인의 부위에 삽입된다. 일부 실시형태에서, 염기 편집기는 염기 절제 복구(BER의) 억제제에 융합된다. 일부 실시형태에서, 염기 절제 복구의 억제제는 우라실 DNA 글리코실라제 억제제(UGI)이다. 일부 실시형태에서, 염기 절제 복구의 억제제는 이노신 염기 절제 복구 억제제이다. 염기 편집기에 대한 자세한 사항은 PCT 국제출원 번호 PCT/2017/045381(WO2018/027078)호 및 PCT/US2016/058344(WO2017/070632)호에 기재되어 있으며, 이들 각각은 그 전문이 본 명세서에 참조로 통합된다. 또한 이들의 전체 내용이 본 명세서에 참조로 통합되는, 하기 문헌 참조: Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A뷿 to G뷗 in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017); 및 Rees, H.A., *et al.*, "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1.

[0107] 일부 실시형태에서, 데아미나제 도메인은 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인의 영역 내로 삽입된다. 일부 실시형태에서, 삽입 부위는 napDNAbp의 구조적 분석에 의해 결정된다. 일부 실시형태에서, 삽입 부위는 가요성 루프이다. 일부 실시형태에서, 데아미나제 도메인은 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인 내의 부위 내로 삽입되고, 상기 부위는 1029, 1026, 1054, 1022, 1015, 1068, 1247, 1040, 1248, 및 768로 이루어진 아미노산 위치의 그룹으로부터의 적어도 하나에서 선택된다. 일부 실시형태에서, 데아미나제 도메인은 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인의 도메인을 대체하여 삽입된다. 일부 실시형태에서, 도메인은 RuvC, Rec1, Rec2, 및 HNH로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 데아미나제 도메인은 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인 내의 소정 범위의 아미노산 잔기 대신 삽입되고, 여기서 상기 범위의 아미노산 잔기는 서열번호: 1로 번호매겨진 것과 같은 Cas9의 잔기 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 및 1298-1300 또는 이의 상응하는 위치로 이루어진 그룹에서 선택된다. Cas9 아미노산 서열을 비교함으로써 상이한 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인에서 상동성 영역을 확인하는 방법은 당업자에게 명백할 것이다. 일부 실시형태에서, 염기 편집기는 폴리뉴클레오티드 프로그래밍가능한 DNA 결합 도메인의 1개 부위 이상에 삽입된 1개 이상의 데아미나제 도메인을 포함하며, 여기서 상기 부위는 위에 기재되어 있다.

[0108] 일부 실시형태에서, 염기 편집기는 아데노신 데아미나제 변이체(예를 들어, TadA*7.10)를 원형 치환체(circular permutant) Cas9(예를 들어, spCAS9) 및 이분 핵 위치결정 서열을 포함하는 스캐폴드로 클로닝함으로써 생성된다. 원형 치환체 Cas9는 당업계에 공지되어 있으며, 예를 들어 문헌[Oakes *et al.*, Cell 176, 254-267, 2019]에 기술되어 있다. 예시적인 원형 치환체 서열은 하기에 제시되어 있으며, 여기서 굵은 글씨체 서열은 Cas9로부터 유래된 서열을 나타내고, 이탤릭체 서열은 링커 서열을 나타내고, 밑줄친 서열은 이분된 핵 위치결정 서열을 나타낸다.

[0109] **CP5** (MSP "NGC=NGG와 같은 정규 Cas9 돌연변이가 있는 Pam 변이체" PID=단백질 상호작용 도메인 및 "D10A" 닉카아제를 지님):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM

[0110] **LASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE**

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

[0111]

[0112] 핵염기 구성요소(components) 및 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 구성요소는 공유적으로 또는 비공유적으로 서로 연관될 수 있다. 예를 들어, 일부 실시형태에서, 데아미나제 도메인은 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 도메인에 의해 표적 뉴클레오티드 서열에 표적화될 수 있다. 일부 실시형태에서, 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 도메인은 데아미나제 도메인에 융합되거나 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 도메인은 데아미나제 도메인과 비공유적으로 상호작용하거나 연관됨으로써 데아미나제 도메인을 표적 뉴클레오티드 서열에 대해 표적화할 수 있다. 예를 들어, 일부 실시형태에서, 핵염기 편집 구성요소, 예를 들어, 데아미나제 구성요소는 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 도메인의 일부인 추가의 이종성(heterologous) 부분 또는 도메인과 상호작용할 수 있거나, 이와 연관될 수 있거나, 또는 이와 복합체화할 수 있는 추가의 이종성 부분 또는 도메인을 포함할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리펩티드에 결합하거나, 이와 상호작용하거나, 이와 연관되거나, 또는 이와 복합체화할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리뉴클레오티드에 결합하거나, 이와 상호작용하거나, 이와 연관하거나, 또는 이와 복합체화할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 가이드 폴리뉴클레오티드에 결합할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리펩티드 링커에 결합할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리뉴클레오티드 링커에 결합할 수 있다. 추가의 이종 부분은 단백질 도메인일 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인식 모티프일 수 있다.

[0113] 염기 편집기 시스템은 가이드 폴리뉴클레오티드 구성요소를 추가로 포함할 수 있다. 염기 편집기 시스템의 구성요소는 공유 결합, 비공유 상호작용, 또는 이들의 연관 및 상호작용의 임의의 조합을 통해 서로 연관될 수 있음을 이해해야 한다. 일부 실시형태에서, 데아미나제 도메인은 가이드 폴리뉴클레오티드에 의해 표적 뉴클레오티드 서열로 표적화될 수 있다. 예를 들어, 일부 실시형태에서, 핵염기 편집기 시스템의 핵염기 편집 구성요소, 예를 들어, 데아미나제 구성요소는 가이드 폴리뉴클레오티드의 일부 또는 세그먼트(예를 들어, 폴리뉴클레오티드 모티프)와 상호작용할 수 있거나, 이와 연관하거나, 또는 이와 복합체화할 수 있는 추가의 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오티드 결합 도메인)을 포함할 수 있다. 일부 실시형태에서, 추가의 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오티드 결합 도메인)은 데아미나제 도메인에 융합되거나 연결될 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리펩티드에 결합하거나, 이와 상호작용하거나, 이와 연관되거나, 또는 이와 폴리펩티드와 복합체화할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리뉴클레오티드에 결합하거나, 이와 상호작용하거나, 이와 연관되거나, 또는 이와 복합체화할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 가이드 폴리뉴클레오티드에 결

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

[0114] 일부 실시형태에서, 염기 편집기 시스템은 염기 절제 복구(BER) 구성요소의 억제제를 추가로 포함할 수 있다. 염기 편집기 시스템의 구성요소는 공유 결합, 비공유 상호작용, 또는 이들의 연관 및 상호작용의 임의의 조합을 통해 서로 연관될 수 있음을 이해해야 한다. BER 구성요소의 억제제는 염기 절제 복구 억제제를 포함할 수 있다. 일부 실시형태에서, 염기 절제 복구의 억제제는 우라실 DNA 글리코실라제 억제제(UGI)일 수 있다. 일부 실시형태에서, 염기 절제 복구의 억제제는 이노신 염기 절제 복구 억제제일 수 있다. 일부 실시형태에서, 염기 절제 복구의 억제제는 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 도메인에 의해 표적 뉴클레오티드 서열에 표적화될 수 있다. 일부 실시형태에서, 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 도메인은 염기 절제 복구의 억제제에 융합되거나 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 도메인은 데아미나제 도메인 및 염기 절제 복구의 억제제에 융합되거나 연결될 수 있다. 일부 실시형태에서, 폴리뉴클레오티드 프로그래밍가능한 뉴클레오티드 결합 도메인은 염기 절제 복구 억제제와 비공유적으로 상호작용하거나 연관됨으로써 염기 절제 복구 억제제를 표적 뉴클레오티드 서열에 대해 표적화할 수 있다. 예를 들어, 일부 실시형태에서, 염기 절제 복구 구성요소의 억제제는 폴리뉴클레오티드 프로그래밍가능한 결합 도메인의 일부인 추가의 이종성 부분 또는 도메인과 상호작용할 수 있거나, 이와 연관될 수 있거나, 또는 이와 복합체화할 수 있는 추가의 이종성 부분 또는 도메인을 포함할 수 있다. 일부 실시형태에서, 염기 절제 복구의 억제제는 가이드 폴리뉴클레오티드에 의해 표적 뉴클레오티드 서열에 표적화될 수 있다. 예를 들어, 일부 실시형태에서, 염기 절제 복구의 억제제는 가이드 폴리뉴클레오티드의 일부 또는 세그먼트(예를 들어, 폴리뉴클레오티드 모티프)와 상호작용할 수 있거나, 이와 연관될 수 있거나, 또는 이와 복합체화할 수 있는 추가의 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오티드 결합 도메인)을 포함할 수 있다. 일부 실시형태에서, 가이드 폴리뉴클레오티드의 추가의 이종성 부분 또는 도메인(예를 들어, RNA 또는 DNA 결합 단백질과 같은 폴리뉴클레오티드 결합 도메인)은 염기 절제 복구의 억제제에 융합되거나 연결될 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리뉴클레오티드에 결합하거나, 이와 상호작용하거나, 이와 연관하거나, 또는 이와 복합체화할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 가이드 폴리뉴클레오티드에 결합할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리펩티드 링커에 결합할 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 폴리뉴클레오티드 링커에 결합할 수 있다. 추가적인 이종 부분은 단백질 도메인일 수 있다. 일부 실시형태에서, 추가의 이종성 부분은 K 상동성(KH) 도메인, MS2 외피 단백질 도메인, PP7 외피 단백질 도메인, SfMu Com 외피 단백질 도메인, 멸균 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인식 모티프일 수 있다.

[0115] "염기 편집 활성"은 폴리뉴클레오티드 내의 염기를 화학적으로 변경시키는 작용을 지칭한다. 한 실시형태에서, 제1 염기는 제2 염기로 전환된다. 한 실시형태에서, 염기 편집 활성은, 예를 들어, 표적 C·G를 T·A로 전환하는, 시티딘 데아미나제 활성이다. 또 다른 실시형태에서, 염기 편집 활성은, 예를 들어, A·T를 G·C로 전환하는, 아데노신 데아미나제 활성이다.

[0116] 용어 "Cas9" 또는 "Cas9 도메인"은 Cas9 단백질, 또는 이의 단편(예를 들어, Cas9의 활성, 비활성 또는 부분 활성 DNA 절단 도메인 및/또는 Cas9의 gRNA 결합 도메인을 포함하는 단백질)을 포함하는 RNA 가이드된 뉴클레아제를 지칭한다. Cas9 뉴클레아제는 또한 때때로 Casn1 뉴클레아제 또는 CRISPR(일정한 간격을 두고 규칙적으로 분포하는 짧은 회문 반복부) 관련 뉴클레아제로도 지칭된다. CRISPR은 이동성 유전 요소(바이러스, 전이가능한 요소 및 접합 플라스미드)에 대한 보호를 제공하는 적응 면역 시스템이다. CRISPR 클러스터는 스페이서, 선행 이동 요소에 상보적인 서열, 및 표적 침입 핵산을 포함한다. CRISPR 클러스터는 전사되어 CRISPR RNA(crRNA)로 가공된다. 유형 II CRISPR 시스템에서 pre-crRNA의 올바른 가공에는 트랜스-코딩된 작은 RNA(tracrRNA), 내인성 리보뉴클레아제3(rnc) 및 Cas9 단백질을 필요로 한다. tracrRNA는 pre-crRNA의 리보뉴클레아제3-보조(aided) 가공을 위한 가이드 역할을 한다. 후속적으로, Cas9/crRNA/tracrRNA는 스페이서에 상보적인 선형 또는 원형 dsDNA 표적을 핵산내부분해방식으로(endonucleolytically) 절단한다. crRNA에 상보적이지 않은 표적 가닥은 먼저 내핵분해적으로 절단된 다음 3'-5' 핵산외부분해방식으로(exonucleolytically) 트리밍된다. 자연에서, DNA-결합 및 절단에는 일반적으로 단백질과 두 RNA가 모두 필요하다. 그러나, 단일 가이드 RNA("sgRNA" 또는 간단히 "gNRA")는 crRNA 및 tracrRNA 둘 다의 양상을 단일 RNA 종에 통합하도록 조작될 수 있다. 예를 들어, 이의 전체 내용이 본 명세서에 참조로 통합되는, 문헌[Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A.,

Charpentier E. *Science* 337:816-821(2012)] 참조. Cas9는 CRISPR 반복 서열에서 짧은 모티프(PAM 또는 프로토스페이서 인접 모티프)를 인식하여 자가 대 비자기를 구별하는 것을 돕는다. Cas9 뉴클레아제 서열 및 구조는 당업자에게 잘 알려져 있다(예를 들어, 이들 각각의 전체 내용이 본 명세서에 참조로 통합되는, 다음 문헌 참조: "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)). Cas9 오르토로그는 S. 피오게네스(*S. pyogenes*) 및 S. 써모필루스(*S. thermophilus*)를 포함하지만, 이로만 제한되는 것은 아닌, 다양한 종에서 설명되었다. 추가의 적합한 Cas9 뉴클레아제 및 서열은 본 개시에 기초하여 당업자에게 명백할 것이며, 이러한 Cas9 뉴클레아제 및 서열은, 그 전체 내용이 본 명세서에 참조로 통합되는, 문헌 [Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737]에 개시된 유기체 및 유전자좌로부터의 Cas9 서열을 포함한다.

[0117] 뉴클레아제-비활성화된 Cas9 단백질은 "dCas9" 단백질(뉴클레아제-"멸실된" Cas9의 경우) 또는 촉매적으로 비활성인 Cas9로 상호교환적으로 지칭될 수 있다. 비활성 DNA 절단 도메인을 갖는 Cas9 단백질(또는 이의 단편)을 생성하는 방법은 공지되어 있다(예를 들어, 각각의 전체 내용이 본 명세서에 참고로 포함되는, 다음 문헌 참조: 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). 예를 들어, Cas9의 DNA 절단 도메인은 2개의 서브도메인, HNH 뉴클레아제 서브도메인 및 RuvC1 서브도메인을 포함하는 것으로 알려져 있다. HNH 서브 도메인은 gRNA에 상보적인 가닥을 절단하는 반면, RuvC1 서브 도메인은 비-상보적인 가닥을 절단한다. 이러한 서브도메인 내의 돌연변이는 Cas9의 뉴클레아제 활성을 침묵시킬 수 있다. 예를 들어, 돌연변이 D10A 및 H840A는 S. 피오게네스(*S. pyogenes*) Cas9의 뉴클레아제 활성을 완전히 비활성화시킨다(Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, *Cell*. 28;152(5):1173-83 (2013)). 일부 실시형태에서, Cas9 뉴클레아제는 비활성(예를 들어, 비활성화된) DNA 절단 도메인을 갖고, 즉 Cas9는 "nCas9" 단백질("닉카아제" Cas9의 경우)로 지칭되는 닉카아제이다. 일부 실시형태에서, Cas9의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 실시형태에서, 단백질은 다음 2개의 Cas9 도메인 중 하나를 포함한다: (1) Cas9의 gRNA 결합 도메인; 또는 (2) Cas9의 DNA 절단 도메인. 일부 실시형태에서, Cas9 또는 이의 단편을 포함하는 단백질은 "Cas9 변이체"로 지칭된다. Cas9 변이체는 Cas9, 또는 이의 단편에 대한 상동성을 공유한다. 예를 들어, Cas9 변이체는 야생형 Cas9와 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97%, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5%, 또는 적어도 약 99.9% 동일하다. 일부 실시형태에서, Cas9 변이체는 야생형 Cas9과 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개, 또는 그 이상의 아미노산 변화를 가질 수 있다. 일부 실시형태에서, Cas9 변이체는 Cas9의 단편(예를 들어, gRNA 결합 도메인 또는 DNA-절단 도메인)을 포함하여, 단편이 야생형 Cas9의 상응하는 단편에 대해 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 96% 동일, 적어도 약 97% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일, 또는 적어도 약 99.9% 동일하다. 일부 실시형태에서, 단편은 상응하는 야생형 Cas9의 아미노산 길이의 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70 이상%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95% 동일, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%이다.

[0118] 일부 실시형태에서, 단편은 적어도 100개 아미노산 길이이다. 일부 실시형태에서, 단편은 적어도 100개, 150개, 200개, 250개, 300개, 350개, 400개, 450개, 500개, 550개, 600개, 650개, 700개, 750개, 800개, 850개, 900개, 950개, 1000개, 1050개, 1100개, 1150개, 1200개, 1250개, 또는 적어도 1300개의 아미노산 길이이다. 일부 실시형태에서, 야생형 Cas9는 스트렙토코커스 피오게네스(*Streptococcus pyogenes*)로부터의 Cas9에 상응한다(NCBI 참조 서열: NC_017053.1, 뉴클레오티드 및 아미노산 서열은 아래와 같다).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
GTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT
CTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTC
TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG
GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATC
CAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGA
TGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC

[0119] AGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTG

ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGA
TACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGT
TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGT
GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGA
CTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT
TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT
TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT
AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA
TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAA
GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT
GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
AACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTA
TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAG
CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAA
GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
TTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCA
CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGA
TTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTT
GATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGA
AAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAG
GTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAA
AATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATT
AGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAG
ACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAAC
GTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAA
GTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTG

[0120] GCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGA

GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCT
ATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTT
GGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAA
AGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAA
AATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGA
GAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAA
AGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGA
AAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGAC
AAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAAC
GGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAAT
CCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATT
GACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAA
ATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTAC
AAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCAT
TATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCA
TAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAG
CAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGT
GAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTT
TAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATG
CCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA
GGAGGTGACTGA

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV

DELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL

[0121] LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV

[0122] GGD

[0123] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인)

[0124] 일부 실시형태에서, 야생형 Cas9는 아래 뉴클레오티드 및/또는 아미노산 서열에 상응하거나, 또는 이를 포함한다:

[0125]

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCAT
AACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATT
CGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACT
CGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACA
AGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGT
CCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTC
AACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTG
GGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATC
CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGA
TGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCAC
AATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTG
ACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGA
CACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTAT
TTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACT
GAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGA
CTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCT
TTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTC

TACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACT
CAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAA
TCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAA
GACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCT
GGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAT
GGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACC
AACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTA
TTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCG
CCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAA
GTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGA
GATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGA
TAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTG
TTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCA
CCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGAT
TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTT
CTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAAC
CTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATA
TTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTG
GATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACG
CGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAG
AGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG
CAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA
ACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGA
AGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGAC
AATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGC
GAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
AACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCAT
GTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCG
GGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAAT
TCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTC
GTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTA
CAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAG
CCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAAC
[0126] GGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGA

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

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN

[0127]

GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0128] LGGD

[0129] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인)

[0130] 일부 실시형태에서, 야생형 Cas9는 스트렙토코커스 피오게네스(*Streptococcus pyogenes*)로부터의 Cas9(NCBI 참조 서열: NC_002737.2(아래와 같은 뉴클레오티드 서열); 및 Uniprot 참조 서열: Q99ZW2(아래와 같은 아미노산 서열))에 상응한다.

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

[0131] GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT

GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT
GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA
AACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTA
TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAG
CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAA
GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA
AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT
TTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCA
CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT
TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT
TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATA
TTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTT
GATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACG
TGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAG
AAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTG
CAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGA
ATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTA
AAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGAT
AACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGC
CAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTG
AACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCAT
GTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCG
AGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAAT
TCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTC
GTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTA
TAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCG
CAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAAT
GGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGA
TAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCA
AGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCG
GACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCC
[0132] AACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAA

AATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCG
ATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACC
TAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAAT
TACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGT
CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCA
GCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTT
TAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATA
CGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGC
TTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAG
ATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAG
CTAGGAGGTGACTGA

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

[0133]

REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0134] (서열번호: 1. 단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인)

[0135] 일부 실시형태에서, Cas9는, 코리네박테리움 울세란스(*Corynebacterium ulcerans*)(NCBI 참조: NC_0156831, NC_017317.1); 코리네박테리움 디프테리아(*Corynebacterium diphtheria*)(NCBI 참조: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라(*Spiroplasma syrphidicola*)(NCBI 참조: NC_021284.1); 프레보텔라 인터메디아(*Prevotella intermedia*)(NCBI 참조: NC_017861.1); 스피로플라스마 타이와넨세(*Spiroplasma taiwanense*)(NCBI 참조:NC_021846.1); 스트렙토코커스 이니애(*Streptococcus iniae*)(NCBI 참조: NC_021314.1); 벨리엘라 발티카(*Belliella baltica*)(NCBI 참조: NC_018010.1); 사이크로플렉수스 토르퀴스(*Psychroflexus torquis*)I(NCBI 참조: NC_018721.1); 스트렙토코커스 써모필러스(*Streptococcus thermophilus*)(NCBI 참조: YP_820832.1), 리스테리아 인노쿠아(*Listeria innocua*)(NCBI 참조: NP_472073.1), 캄필로박터 제주니(*Campylobacter jejuni*)(NCBI 참조: YP_002344900.1) 또는 나이세리아 메닌지티디스(*Neisseria meningitidis*)(NCBI 참조: YP_002342100.1)으로부터의 Cas9 또는 임의의 다른 유기체로부터의 Cas9을 지칭한다.

[0136] 일부 실시형태에서, dCas9는 Cas9 뉴클레아제 활성을 비활성화시키는 하나 이상의 돌연변이를 갖는 Cas9 아미노산 서열에 상응하거나, 또는 부분적으로 또는 전체적으로 이를 포함한다. 예를 들어, 일부 실시형태에서, dCas9

도메인은 D10A 및 H840A 돌연변이 또는 또 다른 Cas9에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, dCas9는 dCas9(D10A 및 H840A)의 아미노산 서열을 포함한다:

[0137]
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN

[0138]
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0139] (단일 밑줄: HNH 도메인; 이중 밑줄: RuvC 도메인).

[0140] 일부 실시형태에서, Cas9 도메인은 D10A 돌연변이를 포함하는 반면, 위치 840의 잔기는 상기 제공된 아미노산 서열에서, 또는 본 명세서에 제공된 임의의 아미노산 서열 중의 상응하는 위치에서 히스티딘으로 유지된다.

[0141] 다른 실시형태에서, D10A 및 H840A 이외의 돌연변이를 갖는 dCas9 변이체가 제공되며, 이는, 예를 들어, 뉴클레아제 비활성화된 Cas9(dCas9)를 초래한다. 예시의 일환으로, 이러한 돌연변이는 D10 및 H840에서의 다른 아미노산 치환, 또는 Cas9의 뉴클레아제 도메인 내의 다른 치환(예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인에서의 치환)을 포함한다. 일부 실시형태에서, dCas9의 변이체 또는 상동체(homologues)는 적어도 약 70% 동일, 적어도 약 80% 동일, 적어도 약 90% 동일, 적어도 약 95% 동일, 적어도 약 98% 동일, 적어도 약 99% 동일, 적어도 약 99.5% 동일, 또는 적어도 약 99.9% 동일한 것으로 제공된다. 일부 실시형태에서, 약 5개 아미노산, 약 10개 아미노산, 약 15개 아미노산, 약 20개 아미노산, 약 25개 아미노산, 약 30 아미노산, 약 40 아미노산, 약 50개 아미노산, 약 75개 아미노산, 약 100개 아미노산 또는 그 이상 만큼 더 짧거나 더 긴, 아미노산 서열을 갖는 dCas9의 변이체가 제공된다.

[0142] 일부 실시형태에서, 본 명세서에 제공된 바와 같은 Cas9 융합 단백질은 Cas9 단백질의 전체 길이 아미노산 서열, 예를 들어, 본 명세서에 제공된 Cas9 서열 중 하나를 포함한다. 그러나, 다른 실시형태에서, 본 명세서에 제공된 바와 같은 융합 단백질은 전체 길이 Cas9 서열을 포함하지 않고, 이의 하나 이상의 단편만을 포함한다. 적합한 Cas9 도메인 및 Cas9 단편의 예시적인 아미노산 서열이 본 명세서에 제공되며, Cas9 도메인 및 단편의 추가의 적합한 서열은 당업자에게 명백할 것이다.

[0143] 일부 실시형태에서, Cas9은 하기로부터의 Cas9을 지칭한다: 코리네박테리움 울세란스(*Corynebacterium ulcerans*)(NCBI 참조: NC_0156831, NC_017317.1); 코리네박테리움 디프테리아(*Corynebacterium diphtheria*)(NCBI 참조: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라(*Spiroplasma syrphidicola*)(NCBI 참조: NC_021284.1); 프레보텔라 인터메디아(*Prevotella intermedia*)(NCBI 참조: NC_017861.1); 스피로플라스마 타이와넨세(*Spiroplasma taiwanense*)(NCBI 참조:NC_021846.1); 스트렙토코커스

이니애(*Streptococcus iniae*)(NCBI 참조: NC_021314.1); 벨리엘라 발티카(*Belliella baltica*)(NCBI 참조: NC_018010.1); 사이크로플렉수스 토르퀴스(*Psychroflexus torquis*)I(NCBI 참조: NC_018721.1); 스트렙토코커스 써모필러스(*Streptococcus thermophilus*)(NCBI 참조: YP_820832.1), 리스테리아 인노쿠아(*Listeria innocua*)(NCBI 참조: NP_472073.1), 캄필로박터 제주니(*Campylobacter jejuni*)(NCBI 참조: YP_002344900.1); 또는 나이세리아 메닌지티디스(*Neisseria meningitidis*)(NCBI 참조: YP_002342100.1)

[0144] 추가 Cas9 단백질(예를 들어, 뉴클레아제 멸실된 Cas9(dCas9), Cas9 닉카아제(nCas9), 또는 뉴클레아제 활성 Cas9)이, 이들의 변이체 및 상동체를 포함하여, 본 개시의 범위 내에 있음을 이해해야 한다. 예시적인 Cas9 단백질은, 제한됨이 없이, 아래에 제공된 것들을 포함한다. 일부 실시형태에서, Cas9 단백질은 뉴클레아제 멸실된 Cas9(dCas9)이다. 일부 실시형태에서, Cas9 단백질은 Cas9 닉카아제(nCas9)이다. 일부 실시형태에서, Cas9 단백질은 뉴클레아제 활성 Cas9이다.

[0145] 예시적인 촉매적으로 비활성인 Cas9(dCas9):

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
[0146] LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ

HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
[0147] PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0148] 예시적인 촉매적으로 Cas9 닉카아제(nCas9):

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA

[0149] PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0150] 예시적인 촉매적으로 활성인 Cas9:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL

[0151] NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE

KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
[0152] PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD.

[0153] 일부 실시형태에서, Cas9는 단일-세포 원핵 미생물의 도메인 및 킹덤을 구성하는 고세균(예를 들어, 나노고세균(nanoarchaea)으로부터의 Cas9를 지칭한다. 일부 실시형태에서, Cas9은 CasX 또는 CasY를 지칭하며, 이는, 예를 들어, 이의 전체 내용이 본 명세서에 참조로 통합되는, 문헌[Burstein *et al.*, "New CRISPR-Cas systems from uncultivated microbes" Cell Res. 2017 Feb 21 doi: 10.1038/cr.2017.21]에 기재된 것이다. 게놈-해체 메타유전체학(genome-resolved metagenomics)을 이용하여, 생명의 고세균 도메인에서 처음으로 보고된 Cas9를 포함하는, 다수의 CRISPR-Cas 시스템이 확인되었다. 이 분기된 Cas9 단백질은 활성 CRISPR-Cas 시스템의 일부로 거의 연구되지 않은 나노고세균에서 발견되었다. 박테리아에서, 이전에 알려지지 않은 2가지 시스템인, CRISPR-CasX 및 CRISPR-CasY가 발견되었으며, 이들은 지금까지 발견된 가장 컴팩트한 시스템 중 하나이다. 일부 실시형태에서, Cas9는 CasX, 또는 CasX의 변이체를 지칭한다. 일부 실시형태에서, Cas9는 CasY, 또는 CasY의 변이체를 지칭한다. 다른 RNA-가이드된 DNA 결합 단백질이 핵산 프로그래밍가능한 DNA 결합 단백질(napDNAbp)로 사용될 수 있으며, 본 개시의 범위 내에 있음을 이해해야 한다.

[0154] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 핵산 프로그래밍가능한 DNA 결합 단백질(napDNAbp)은 CasX 또는 CasY 단백질일 수 있다. 일부 실시형태에서, napDNAbp는 CasX 단백질이다. 일부 실시형태에서, napDNAbp는 CasY 단백질이다. 일부 실시형태에서, napDNAbp는 자연적으로-발생하는 Cas12b/C2c1 또는 CasX 또는 CasY 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, napDNAbp는 자연적으로-발생하는 CasX 또는 CasY 단백질이다. 일부 실시형태에서, napDNAbp는 본 명세서에 제공된 임의의 CasX 또는 CasY 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 박테리아 종으로부터의 CasX 및 CasY도 본 개시에 따라 사용될 수 있음을 이해해야 한다.

[0155] CasX (uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53)

[0156] >tr|F0NN87|F0NN87_SULIH CRISPR-associated Casx protein OS = *Sulfolobus islandicus* (strain HVE10/4) GN = SiH_0402 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQV
KECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAK
VSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSV
VRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVL
[0157] ANYIYEYLTG SKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0158] >tr|F0NH53|F0NH53_SULIR CRISPR associated protein, Casx OS = *Sulfolobus islandicus* (strain REY15A) GN=SiRe_0771 PE=4 SV=1

MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQV
KECEEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKA
KVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVS
VVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIV
[0159] LANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG

[0160] CasY (ncbi.nlm.nih.gov/protein/APG80656.1)

[0161] >APG80656.1 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium]

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDD
YVGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEV
RGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRE
RHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTC
CLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLG
EGFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNH
WGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEA
VVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERL
EAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYK
NAAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKW
KPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALA
RELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGTV
KDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAE
LLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFG
YELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQT
QFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQ
PFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVK
GLKLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQ
KIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFCGACKKLWRA
EMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISK
KMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKN IKVLG
[0162] QMKKI

[0163] "시티딘 데아미나제"는 아미노 기를 카보닐 기로 전환시키는 탈아미노 반응을 촉매할 수 있는 폴리펩티드 또는 이의 단편을 의미한다. 한 실시형태에서, 시티딘 데아미나제는 시토신을 우라실로 또는 5-메틸시토신을 티민으로 전환시킨다. 페트로미존 마리누스(Petromyzon marinus)에서 유래된, PmCDA1(페트로미존 마리누스 시토신 데아미나제 1, "PmCDA1"), 포유동물(예를 들어, 인간, 돼지, 소, 말, 원숭이 등)로부터 유래된, AID(활성화-유도된 시티딘 데아미나제; AICDA), 및 APOBEC은 예시적인 시티딘 데아미나제이다.

[0164] PmCDA1의 염기 서열 및 아미노산 서열 및 인간 AID의 CDS의 염기 서열 및 아미노산 서열은 본 명세서에서 아래에 제시되어 있다:

[0165] >tr|A5H718|A5H718_PETMA Cytosine deaminase OS=페트로미존 마리누스(Petromyzon marinus) OX=7757 PE=2

SV=1

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSG
TERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLK
IWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKT

[0166] LKRAEKRRSELSIMIQVKILHTTKSPAV

[0167] >EF094822.1 Petromyzon marinus isolate PmCDA.21 cytosine deaminase mRNA, complete cds

TGACACGACACAGCCGTGTATATGAGGAAGGGTAGCTGGATGGGGGGGGGGGGAATACGTTCAGAGAGGA
CATTAGCGAGCGTCTTGTTGGTGGCCTTGAGTCTAGACACCTGCAGACATGACCGACGCTGAGTACGTGA
GAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAAATCCGTGTCGCA
TAGATGCTACGTTCTCTTTGAATTAAAACGACGGGGTGAACGTAGAGCGTGTTTTTGGGGCTATGCTGTG
AATAAACCACAGAGCGGGACAGAACGTGGAATTCACGCCGAAATCTTTAGCATTAGAAAAGTCGAAGAAT
ACCTGCGCGACAACCCCGGACAATTCACGATAAATTGGTACTCATCCTGGAGTCCTTGTGCAGATTGCGC
TGAAAAGATCTTAGAATGGTATAACCAGGAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGC
AAACTCTATTACGAGAAAAATGCGAGGAATCAAATTGGGCTGTGGAACCTCAGAGATAACGGGGTTGGGT
TGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAAAATATTCATCCAATCGTCGCACAATCAATT
GAATGAGAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACGGAGCGAGTTGTCCATTATG

[0168] ATTCAGGTAAAAATACTCCACACCACTAAGAGTCCTGCTGTTTAAGAGGCTATGCGGATGGTTTTC

[0169] >tr|Q6QJ80|Q6QJ80_HUMAN Activation-induced cytidine deaminase OS=Homo sapiens OX=9606 GN=AICDA PE=2 SV=1

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELL
FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK

[0170] AEPEGLRRLHRAGVQIAIMTFKAPV

[0171] >NG_011588.1:5001-15681 Homo sapiens activation induced cytidine deaminase (AICDA), RefSeqGene (LRG_17) on chromosome 12

AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGGAGGCAAGAAGACACTCTG
GACACCACTATGGACAGGTAAAGAGGCAGTCTTCTCGTGGGTGATTGCACTGGCCTTCCTCTCAGAGCAA
ATCTGAGTAATGAGACTGGTAGCTATCCCTTTCTCTCATGTAACTGTCTGACTGATAAGATCAGCTTGAT
CAATATGCATATATATTTTTTGATCTGTCTCCTTTTCTTCTATTCAGATCTTATACGCTGTCAGCCCAAT
TCTTTCTGTTTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAAGTAGTGCGTACAATGTA
CTGATTCGTCCTGAGATTTGTACCATGGTTGAAACTAATTTATGGTAATAATATTAACATAGCAAATCTT
TAGAGACTCAAATCATGAAAAGGTAATAGCAGTACTGTACTAAAAACGGTAGTGCTAATTTTCGTAATAA
TTTTGTAAATATTCAACAGTAAAACAACTTGAAGACACACTTTCCTAGGGAGGCGTTACTGAAATAATTT
AGCTATAGTAAGAAAATTTGTAATTTTAGAAATGCCAAGCATTCTAAATTAATTGCTTGAAAGTCACTAT
GATTGTGTCCATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAATGTTAAAGGCCCAATTGTTAGG
CAGTTAATGGCACTTTTACTATTAACTAATCTTTCCATTTGTTCAGACGTAGCTTAACTTACCTCTTAGG

[0172] TGTGAATTTGGTTAAGGTCCTCATAATGTCTTTATGTGCAGTTTTTGATAGGTTATTGTCATAGAACTTA

TTCTATTCCTACATTTATGATTACTATGGATGTATGAGAATAACACCTAATCCTTATACTTTACCTCAAT
TTAACTCCTTTATAAAGAACTTACATTACAGAATAAAGATTTTTTAAAAATATATTTTTTTGTAGAGACA
GGGTCTTAGCCCAGCCGAGGCTGGTCTCTAAGTCCTGGCCCAAGCGATCCTCCTGCCTGGGCCTCCTAAA
GTGCTGGAATTATAGACATGAGCCATCACATCCAATATACAGAATAAAGATTTTTAATGGAGGATTTAAT
GTTCTTCAGAAAATTTTCTTGAGGTCAGACAATGTCAAATGTCTCCTCAGTTTACACTGAGATTTTGAAA
ACAAGTCTGAGCTATAGGTCCTTGTGAAGGGTCCATTGGAAATACTTGTTCAAAGTAAAATGGAAAGCAA
AGGTAAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAAAGGAGAAAAGATGAAATTCAACAGGACAGAA
GGGAAATATATTATCATTAAGGAGGACAGTATCTGTAGAGCTCATTAGTGATGGCAAAATGACTTGGTCA
GGATTATTTTTAACCCGCTTGTTTCTGGTTTGCACGGCTGGGGATGCAGCTAGGGTTCTGCCTCAGGGAG
CACAGCTGTCCAGAGCAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCTTCCTACTCAG
GACAGAAATGACGAGAACAGGGAGCTGGAAACAGGCCCCTAACCAGAGAAGGGAAGTAATGGATCAACAA
AGTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTGACTGGTAACATGTGACAGAAACAGTGTA
GGCTTATTGTATTTTCATGTAGAGTAGGACCCAAAAATCCACCCAAAGTCCTTTATCTATGCCACATCCT
TCTTATCTATACTTCCAGGACACTTTTTCTTCCTTATGATAAGGCTCTCTCTCTCTCCACACACACACAC
ACACACACACACACACACACACACACACACACACAAACACACACCCCGCCAACCAAGGTGCATGTAAAAAGA
TGTAGATTCCTCTGCCTTTCTCATCTACACAGCCCAGGAGGGTAAGTTAATATAAGAGGGATTTATTGGT
AAGAGATGATGCTTAATCTGTTTAACACTGGGCCTCAAAGAGAGAATTTCTTTTCTTCTGTACTTATTAA
GCACCTATTATGTGTTGAGCTTATATATACAAAGGGTTATTATATGCTAATATAGTAATAGTAATGGTGG
TTGGTACTATGGTAATTACCATAAAAATTATTATCCTTTTAAAATAAAGCTAATTATTATTGGATCTTTT
TTAGTATTCATTTTATGTTTTTTATGTTTTTGATTTTTTAAAAGACAATCTCACCCTGTTACCCAGGCTG
GAGTGCAGTGGTGCAATCATAGCTTTCTGCAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCCTTGG
CCTCCCAAAGTGTTGGGATACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTAAAATATG
CATTTTAAATTTTAAAATAATATGGCTAATTTTTACCTTATGTAATGTGTATACTGGCAATAAATCTAGT
TTGCTGCCTAAAGTTTAAAGTGCTTTCCAGTAAGCTTCATGTACGTGAGGGGAGACATTTAAAGTGAAAC
AGACAGCCAGGTGTGGTGGCTCACGCCTGTAATCCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTT
GAGCCCTGGAGTTCAAGACCAGCCTGAGCAACATGGCAAAACGCTGTTTCTATAACAAAAATTAGCCGGG
CATGGTGGCATGTGCCTGTGGTCCCAGCTACTAGGGGGCTGAGGCAGGAGAATCGTTGGAGCCCAGGAGG
TCAAGGCTGCACTGAGCAGTGCTTGCGCCACTGCACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCA
AAAAAATAAGAAGAAAAATTAAAAATAAATGGAAACAACTACAAAGAGCTGTTGTCCTAGATGAGCTACT
TAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTCACCTGCACTACATTATTAAAAT
ATCAATTCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTCATAGTACCTTTATTCACAAAACC
CCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAAATAAACAAAATGTGCTATATCCATGCAA
TGGAATACCACCCTGCAGTACAAAGAAGCTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAAG
AGTCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTCTAGAAAATGAAAGTAACTTATAGTTAC
AGAAAGCAAATCAGGGCAGGCATAGAGGCTCACACCTGTAATCCCAGCACTTTGAGAGGCCACGTGGGAA
GATTGCTAGAACTCAGGAGTTCAAGACCAGCCTGGGCAACACAGTGAAACTCCATTCTCCACAAAAATGG
GAAAAAAAGAAAGCAAATCAGTGGTTGTCCTGTGGGGAGGGGAAGGACTGCAAAGAGGGAAGAAGCTCTG
GTGGGGTGAGGGTGGTGATTCAGGTTCTGTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAA
AATATTCGTAGAATTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTATACCTCAATGTAAGAA
AAAATAATGTGTAAGAAAACTTTCAATTCTCTTGCCAGCAAACGTTATTCAAATTCCTGAGCCCTTTACT
[0173] TCGCAAATTCTCTGCACTTCTGCCCCGTACCATTAGGTGACAGCACTAGCTCCACAAATTGGATAAATGC

ATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTCATATCAACCATGCTGTACAGCT
TGTGTTGCTGTCTGCAGCTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTTGGGTTACCAGAGTATTT
CCACAAATGCTATTCAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAGG
AGGAGAAATCAGTCATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTTAAAAAAGC
AGCAGAGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAAATATGTGCCTGTAGAAAGACT
AGAGGAATAAACACAAGAATCTTAACAGTCATTGTCATTAGACACTAAGTCTAATTATTATTATTAGACA
CTATGATATTTGAGATTTAAAAAATCTTTAATATTTTAAAATTTAGAGCTCTTCTATTTTTCCATAGTAT
TCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTT
TGGTCTTGTTGCCCATGCTGGAGTGGAATGGCATGACCATAGCTCACTGCAACCTCCACCTCCTGGGTTC
AAGCAAAGCTGTCGCCTCAGCCTCCCGGGTAGATGGGATTACAGGCGCCCACCACCACACTCGGCTAATG
TTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAGG
ATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGATGTAGGCCACTGCGCCCGGCCAAGTATTGC
TCTTATACATTAAAAAACAGGTGTGAGCCACTGCGCCCAGCCAGGTATTGCTCTTATACATTAAAAAATA
GGCCGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAAGCCAAGGCGGGCAGAACACCCGAGGT
CAGGAGTCCAAGGCCAGCCTGGCCAAGATGGTGAAACCCCGTCTCTATTAAAAATACAAACATTACCTGG
GCATGATGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCCGCGGAGCCTGGCA
GATCTGCCTGAGCCTGGGAGGTTGAGGCTACAGTAAGCCAAGATCATGCCAGTATACTTCAGCCTGGGCG
ACAAAGTGAGACCGTAACAAAAAAAAAAAAAATTTAAAAAAAGAAATTTAGATCAAGATCCAACTGTAAAA
AGTGGCCTAAACACCACATTAAAGAGTTTGGAGTTTATTCTGCAGGCAGAAGAGAACCATCAGGGGGTCT
TCAGCATGGGAATGGCATGGTGCACCTGGTTTTTGTGAGATCATGGTGGTGACAGTGTGGGGAATGTTAT
TTTGGAGGGACTGGAGGCAGACAGACCGGTTAAAAGGCCAGCACAACAGATAAGGAGGAAGAAGATGAGG
GCTTGGACCGAAGCAGAGAAGAGCAAACAGGGAAGGTACAAATTCAAGAAATATTGGGGGGTTTGAATCA
ACACATTTAGATGATTAATTAAATATGAGGACTGAGGAATAAGAAATGAGTCAAGGATGGTTCCAGGCTG
CTAGGCTGCTTACCTGAGGTGGCAAAGTCGGGAGGAGTGGCAGTTTAGGACAGGGGGCAGTTGAGGAATA
TTGTTTTGATCATTTTGAGTTTGAGGTACAAGTTGGACACTTAGGTAAAGACTGGAGGGGAAATCTGAAT
ATACAATTATGGGACTGAGGAACAAGTTTATTTTATTTTTTGTTTCGTTTTCTTGTTGAAGAACAAATTT
AATTGTAATCCCAAGTCATCAGCATCTAGAAGACAGTGGCAGGAGGTGACTGTCTTGTGGGTAAGGGTTT
GGGGTCCTTGATGAGTATCTCTCAATTGGCCTTAAATATAAGCAGGAAAAGGAGTTTATGATGGATTCCA
GGCTCAGCAGGGCTCAGGAGGGCTCAGGCAGCCAGCAGAGGAAGTCAGAGCATCTTCTTTGGTTTAGCCC
AAGTAATGACTTCCTTAAAAAGCTGAAGGAAAATCCAGAGTGACCAGATTATAAACTGTACTCTTGCATT
TTCTCTCCCTCCTCTCACCCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTC
CGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACATCCT
TTTCACTGGACTTTGGTTATCTTCGCAATAAGGTATCAATTAAAGTCGGCTTTGCAAGCAGTTTAATGGT
CAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTTTTTTGGCATTTGTGTCTCT
ATCACATTCCTCAAATCCTTTTTTTTATTTCTTTTTCCATGTCCATGCACCCATATTAGACATGGCCCAA
AATATGTGATTTAATTCCTCCCCAGTAATGCTGGGCACCCTAATACCACTCCTTCCTTCAGTGCCAAGAA
CAACTGCTCCCAAACTGTTTACCAGCTTTCCTCAGCATCTGAATTGCCTTTGAGATTAATTAAGCTAAAA
GCATTTTTATATGGGAGAATATTATCAGCTTGTCCAAGCAAAAATTTTAAATGTGAAAAACAAATTGTGT
CTTAAGCATTTTTGAAAATTAAGGAAGAAGAATTTGGGAAAAAATTAACGGTGGCTCAATTCTGTCTTCC
AAATGATTTCTTTTCCCTCCTACTCACATGGGTCGTAGGCCAGTGAATACATTCAACATGGTGATCCCCA

[0174] GAAAACTCAGAGAAGCCTCGGCTGATGATTAATTAAATTGATCTTTCGGCTACCCGAGAGAATTACATTT

CCAAGAGACTTCTTCACCAAAATCCAGATGGGTTTACATAAACTTCTGCCCACGGGTATCTCCTCTCTCC
TAACACGCTGTGACGTCTGGGCTTGGTGGAATCTCAGGGAAGCATCCGTGGGGTGGAAGGTCATCGTCTG
GCTCGTTGTTTGATGGTTATATTACCATGCAATTTTCTTTGCCTACATTTGTATTGAATACATCCCAATC
TCCTTCCTATTCGGTGACATGACACATTCTATTTCAGAAGGCTTTGATTTTATCAAGCACTTTCATTTAC
TTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTGTCTGCTTTACCAAAATCTATTTCCCCT
TTTCAGATCCTCCCAAATGGTCCTCATAAACTGTCCTGCCTCCACCTAGTGGTCCAGGTATATTTCCACA
ATGTTACATCAACAGGCACTTCTAGCCATTTTCCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACACA
AATTAAATCTTCGGTGAGGTAGTGTGATGCTGCTTCCTCCCAACTCAGCGCACTTCGTCTTCCTCATTCC
ACAAAAACCCATAGCCTTCCTTCACTCTGCAGGACTAGTGCTGCCAAGGGTTCAGCTCTACCTACTGGTG
TGCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAACACAAGGACAATAGCTGCAAGCATCCCCAAAGATC
ATTGCAGGAGACAATGACTAAGGCTACCAGAGCCGCAATAAAAGTCAGTGAATTTTAGCGTGGTCCTCTC
TGTCTCTCCAGAACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCC
TGGCCGCTGCTACCGCGTCACCTGGTTCACCTCCTGGAGCCCCTGCTACGACTGTGCCCGACATGTGGCC
GACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCCTCTACTTCTGTGAGGACC
GCAAGGCTGAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCCATCATGACCTTCAA
AGGTGCGAAAGGGCCTTCCGCGCAGGCGCAGTGCAGCAGCCCGCATTCGGGATTGCGATGCGGAATGAAT
GAGTTAGTGGGGAAGCTCGAGGGGAAGAAGTGGGCGGGGATTCTGGTTCACCTCTGGAGCCGAAATTAAA
GATTAGAAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAGGCCCCGAGGAAATGAGAAAATGGGGCCAGG
GTTGCTTCTTTCCCCTCGATTTGGAACCTGAACTGTCTTCTACCCCCATATCCCCGCCTTTTTTTCCTTT
TTTTTTTTTTGAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCACGAAAGAACTTTCAAAGCC
TGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGTAAGGGGCT
TCCTCGCTTTTTAAATTTTCTTTCTTTCTCTACAGTCTTTTTTGGAGTTTCGTATATTTCTTATATTTTC
TTATTGTTCAATCACTCTCAGTTTTCATCTGATGAAAACTTTATTTCTCCTCCACATCAGCTTTTTCTTC
TGCTGTTTCACCATTCAGAGCCCTCTGCTAAGGTTCCTTTTCCCTCCCTTTTCTTTCTTTTGTTGTTTCA
CATCTTTAAATTTCTGTCTCTCCCCAGGGTTGCGTTTCCTTCCTGGTCAGAATTCTTTTCTCCTTTTTTT
TTTTTTTTTTTTTTTTTTTTAAACAAACAAACAAAAAACCCAAAAAAACTCTTTCCCAATTTACTTTCTT
CCAACATGTTACAAAGCCATCCACTCAGTTTAGAAGACTCTCCGGCCCCACCGACCCCCAACCTCGTTTT
GAAGCCATTCACTCAATTTGCTTCTCTCTTTCTCTACAGCCCCTGTATGAGGTTGATGACTTACGAGACG
CATTTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAG
ACAGTGGATAAAAAACAGTCCTTCAAGTCTTCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTT
ACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAAAATAGAGAAGGAACACAGGTC
TGGCCAGGGACGTGCTGCAATTGGTGCAGTTTTGAATGCAACATTGTCCCCTACTGGGAATAACAGAACT
GCAGGACCTGGGAGCATCCTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAAGGT
AGATCCTAAAAAGCATGGTGAGAGGATCAAATGTTTTTATATCAACATCCTTTATTATTTGATTCATTTG
AGTTAACAGTGGTGTTAGTGATAGATTTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAA
CTCTTCCATCAGGCCATGATCTATAGGACCTCCTAATGAGAGTATCTGGGTGATTGTGACCCCAAACCAT
CTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTTAATCAGCAGAAGCATGTTTTTATGTTTGT
ACAAAAGAAGATTGTTATGGGTGGGGATGGAGGTATAGACCATGCATGGTCACCTTCAAGCTACTTTAAT
AAAGGATCTTAAAATGGGCAGGAGGACTGTGAACAAGACACCCTAATAATGGGTTGATGTCTGAAGTAGC
AAATCTTCTGGAAACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCACAAACTTCACATATC

[0175] ATAATTAGCAAACAATTGGAAGGAAGTTGCTTGAATGTTGGGGAGAGGAAAATCTATTGGCTCTCGTGGG

TCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGATGCTTCTCCCA
AAGGTATATTAACTATATAAGAGAGTTGTGACAAAACAGAATGATAAAGCTGCGAACCGTGGCACACGCT
CATAGTTCTAGCTGCTTGGGAGGTTGAGGAGGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGG
GCAACATAACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAGAAAGAGAGAGGGCCGGGCGTGGTG
GCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGCCGGGCGGATCACCTGTGGTCAGGAGTTTGAGA
CCAGCCTGGCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAAATTAGCCAGGCGTGGTAGCAGG
CACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGTTGCA
GTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGACAAGAGCAAGACTCTGTCTCAGAAAAAAAAA
AAAAAAAGAGAGAGAGAGAAAGAGAACAATATTTGGGAGAGAAGGATGGGGAAGCATTGCAAGGAAAT
TGTGCTTTATCCAACAAAATGTAAGGAGCCAATAAGGGATCCCTATTTGTCTCTTTTGGTGTCTATTTGT
CCCTAACAACTGTCTTTGACAGTGAGAAAAATATTCAGAATAACCATATCCCTGTGCCGTTATTACCTAG
CAACCCTTGCAATGAAGATGAGCAGATCCACAGGAAAACTTGAATGCACAACTGTCTTATTTTAATCTTA
TTGTACATAAGTTTGTAAAAGAGTTAAAAATTGTTACTTCATGTATTCATTTATATTTTATATTATTTTG
CGTCTAATGATTTTTTATTAACATGATTTCCTTTTCTGATATATTGAAATGGAGTCTCAAAGCTTCATAA
ATTTATAACTTTAGAAATGATTCTAATAACAACGTATGTAATTGTAACATTGCAGTAATGGTGCTACGAA
GCCATTTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAATTTGCTTCTGGCTCACTTTCAATCAGTTA
AATAAATGATAAATAATTTTGGAAGCTGTGAAGATAAAATACCAAATAAAATAATATAAAAGTGATTTAT

[0176] ATGAAGTTAAAATAAAAAATCAGTATGATGGAATAAACTTG

[0177] 아포리포단백질 B mRNA 편집 효소, 촉매성 폴리펩티드-유사(APOBEC)는 진화적으로 보존된 시티딘 데아미나제 계열이다. 이 패밀리의 구성원은 C-에서-U 편집 효소이다. APOBEC 유사 단백질의 N-말단 도메인은 촉매 도메인이고 C-말단 도메인은 유사촉매(pseudocatalytic) 도메인이다. 보다 구체적으로, 촉매 도메인은 아연 의존성 시티딘 데아미나제 도메인이며 시티딘 탈아미노화에 중요하다. APOBEC 패밀리 구성원은 APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D("APOBEC3E"가 현재 이것을 지칭함), APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4 및 활성화 유도(시티딘) 데아미나제를 포함한다. SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, VRER-BE3, YE1-BE3, EE-BE3, YE2-BE3, 및 YEE-BE3을 포함하나 이로만 제한되는 것은 아닌 다수의 변형된 시티딘 데아미나제가 상업적으로 이용가능하며, Addgene(플라스미드 85169, 85170, 85171, 85172, 85173, 85174, 85175, 85176, 85177)에서 입수가능하다.

[0178] 본 개시의 양상에 따라 Cas9에 융합될 수 있는 다른 예시적인 데아미나제는 아래에 제공된다. 일부 실시형태에서, 각 서열의 활성 도메인, 예를 들어, 위치결정 신호(핵 위치결정 서열, 핵 외수송 신호 없이, 세포질 위치결정 신호)가 없는 도메인이 사용될 수 있음을 이해해야 한다.

[0179] 인간 AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPE
GLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEV

[0180] DDLRDAFRTLGL (밑줄: 핵 위치결정 서열; 이중 밑줄: 핵 외수송 신호)

[0181] 마우스 AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPE
GLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEV

[0182] DDLRDAFRMLGF (밑줄: 핵 위치결정 서열; 이중 밑줄: 핵 외수송 신호)

[0183] 개 AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPE
GLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEV

[0184] DDLRDAFRTLGL (밑줄: 핵 위치결정 서열; 이중 밑줄: 핵 외수송 신호)

[0185] 소 AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFL
RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEP
EGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYE

[0186] VDDLRDAFRTLGL (밑줄: 핵 위치결정 서열; 이중 밑줄: 핵 외수송 신호)

[0187] 랫트 AID

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQR
KFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYISDWDLD

[0188] PGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYF

[0189] YCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLGL (밑줄: 핵 위치결정 서열; 이중 밑줄: 핵 외수송 신호)

[0190] 마우스 APOBEC-3

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKD
NI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATHHNLSLDIFSSRLYNVQD
PETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPV
PSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNG
QAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTS
RLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKE

[0191] SWGLQDLVNDFGNLQLGPPMS (이탤릭체: 핵산 편집 도메인)

[0192] 랫트 APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNK
DNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLATHHNLSLDIFSSRLYNIR
DPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIP
VPSSSSSTLSNICLTKGLPETRFCVERRRVHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFN
GQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYT
SRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIK

[0193] ESWGLQDLVNDFGNLQLGPPMS (이탤릭체: 핵산 편집 도메인)

[0194] 레수스 마카크(*Rhesus macaque*) APOBEC-3 G:

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYHPEMR
FLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDY
QQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELL
RHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPKG
RHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISNNEHVSLCIFAARIYDDQ

[0195] GRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 위치결정 신호)

[0196] 침팬지 APOBEC-3 G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKY*HPEM*
*RFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALR
SLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTS

[0197] NFNNELWVRGRHETYLCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLD*

*LHQDYRVTCFTSWSPCFSC*AQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTY

[0198] SEFKHCWDTFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 위치결정 신호)

[0199] 녹색 원숭이 APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKD*HPEM*
*KFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALR
ILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTS
NFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLD*
*DQQYRVTCFTSWSPCFSC*AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYS
[0200] EFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 위치결정 신호)

[0201] 인간 APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKY*HPEM*
*RFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALR
SLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTF
NFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLD*
*LDQDYRVTCFTSWSPCFSC*AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTY
[0202] SEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

[0203] (이탤릭체: 핵산 편집 도메인; 밑줄: 세포질 위치결정 신호)

[0204] 인간 APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEH*HAEM*
*CFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCR
LSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHF
KNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETH*CHAERCFLSWFCDDILSPNTNYEVT*
*WYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCW
[0205] ENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE (이탤릭체: 핵산 편집 도메인)

[0206] 인간 APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQY*HAE*
*MCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALC
[0207] RLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNN

DPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQI*
*YRVTWFISWSPCFSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDE
[0208] FEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

[0209] (이탤릭체: 핵산 편집 도메인)

[0210] 랫트 APOBEC-3B:

MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCA
LPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNL
SLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFY
DCKLQEIFSRMNLLREDVFYLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQH
VEILFLEKMRSMELSQVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTL
[0211] WRSGIHVDVMDLPQFADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL

[0212] 소 APOBEC-3B:

DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAP
YYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPSQSYKIICYITWSPCPNCANE
LVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPW
[0213] DKLEQYSASIRRRLQRILTAPI

[0214] 침팬지 APOBEC-3B:

MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAE
MCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALC
RLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNN
DPLVLRRHQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQI
YRVTWFISWSPCFSWGCAGQVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDE
FEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEP
PLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSFHSLTSCSIQPPCSSRIRET
[0215] EGWASVSKEGRDLG

[0216] 인간 APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETH*CHAE*
*RCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLR
[0217] SLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ

[0218] 고릴라 APOBEC3C

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETH*CHAE*
*RCFLSWECDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLYYFQDTDYQEGLR
[0219] SLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRFLKRRLQEILE

[0220] (이탤릭체: 핵산 편집 도메인)

[0221] 인간 APOBEC-3A:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYG
R*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLY
[0222] KEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

[0223] (이탤릭체: 핵산 편집 도메인)

[0224] 레수스 마카크(*Rhesus macaque*) APOBEC-3A:

MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCG
DYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVRVFLQENKHVRLRIFAARIYDYD
[0225] PLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN

[0226] (이탤릭체: 핵산 편집 도메인)

[0227] 소 APOBEC-3 A:

MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*HAELYFLGKIHSW*
*NLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARI
[0228] TIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN

[0229] (이탤릭체: 핵산 편집 도메인)

[0230] 인간 APOBEC-3H:

MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEICFINEIKSMGL*
*DETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVM
[0231] GFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV

[0232] (이탤릭체: 핵산 편집 도메인)

[0233] 레수스 마카크(*Rhesus macaque*) APOBEC-3H:

MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGL
DETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVM
GLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSS
[0234] SIRNSR

[0235] 인간 APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHR
QEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPC*LPCVVKVTKFLAEHPNVTLTISAARLY
YYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNP
MEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFC*
*DDILSPNTNYEVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGAS
[0236] VKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

[0237] (이탤릭체: 핵산 편집 도메인)

[0238] 인간 APOBEC-1:

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIK
KFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVN
SGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLT
[0239] FFRLHLQNCHYQTIPPHILLATGLIHPSVAWR

[0240] 마우스 APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLE
KFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLIS
SGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLT
[0241] FFTITLQTCHYQRIPPHLLWATGLK

[0242] 랫트 APOBEC-1 :

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS
SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT
[0243] FFTIALQSCHYQRLPPHILWATGLK

[0244] 인간 APOBEC-2:

MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKT
FLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRII
KTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQP
[0245] WEDIQENFLYYEEKLADILK

[0246] 마우스 APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT
FLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRIL
KTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEP
[0247] WEDIQENFLYYEEKLADILK

[0248] 랫트 APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT
FLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRIL
KTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEP
[0249] WEDIQENFLYYEEKLADILK

[0250] 소 APOBEC-2:

MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKT
FLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIV
KTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEP
[0251] WEDIQENFLYYEEKLADILK

[0252] 페트로미존 마리누스(*Petromyzon marinus*) CDA1 (pmCDA1)

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE
IFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQI
GLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSFMIQVKILHTTK

[0253] SPAV

[0254] 인간 APOBEC3G D316R D317R

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEM
RFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALR
SLCQKRDGPRATMKFNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHFMLGEILRHSMDPPTFTFN
FNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDL
DQDYRVTC
FTSWSPCFSCAQEMAKFISKKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFV

[0255] DHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

[0256] 인간 APOBEC3G 사슬 A

MDPPTFTFNFNNEPWWGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGA

[0257] KISF TYSEFKHCWDTFVDHQGCPFQPWDGLD EHSQDLSGRLRAILQ

[0258] 인간 APOBEC3G 사슬 A D120R D121R

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLD
VIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAG

[0259] AKISFMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ

[0260] 용어 "데아미나제" 또는 "데아미나제 도메인"은 탈아미노화 반응을 촉매하는 단백질 또는 이의 단편을 지칭한다.

[0261] "검출하다"는 검출할 분석물의 존재, 부재 또는 양을 확인하는 것을 의미한다. 한 실시형태에서, 폴리뉴클레오티드 또는 폴리펩티드의 서열 변경이 검출된다. 또 다른 실시형태에서, 인델(indels)의 존재가 검출된다.

[0262] "검출가능한 표지"는 관심 분자에 연결된 경우, 분광, 광화학, 생화학, 면역화학, 또는 화학적 수단을 통해, 상기 분자를 검출가능하게 하는 조성물을 의미한다. 예를 들어, 유용한 표지로는 방사성 동위 원소, 자기 비드, 금속 비드, 콜로이드 입자, 형광 염료, 전자 밀도 시약, 효소(예를 들어, ELISA에서 일반적으로 사용됨), 비오틴, 디곡시게닌, 또는 합텐을 포함한다.

[0263] "단편"은 폴리펩티드 또는 핵산 분자의 일부를 의미한다. 이러한 일부는 참조 핵산 분자 또는 폴리펩티드의 전체 길이의 적어도 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 또는 90%를 함유한다. 단편은 10개, 20개, 30개, 40개, 50개, 60개, 70개, 80개, 90개, 또는 100개, 200개, 300개, 400개, 500개, 600개, 700개, 800개, 900개, 또는 1000개의 뉴클레오티드 또는 아미노산을 함유할 수 있다.

[0264] "혼성화(hybridization)"는 상보적인 핵염기들 사이의, 왓슨-크릭, 후그스틴 또는 역(reversed) 후그스틴 수소 결합일 수 있는, 수소 결합을 의미한다. 예를 들어, 아데닌과 티민은 수소 결합 형성을 통해 쌍을 이루는 상보적인 핵염기이다.

[0265] 용어 "염기 복구의 억제제" 또는 "IBR"은 핵산 복구 효소, 예를 들어, 염기 절제 복구 효소의 활성을 억제할 수 있는 단백질을 지칭한다. 일부 실시형태에서, IBR은 이노신 염기 절제 복구의 억제제이다. 예시적인 염기 복구의 억제제는 APE1, Endo III, Endo IV, Endo V, Endo VIII, Fpg, hOGG1, hNEIL1, T7 Endo1, T4PDG, UDG, hSMUG1, 및 hAAG의 억제제를 포함한다. 일부 실시형태에서, IBR은 Endo V 또는 hAAG의 억제제이다. 일부 실시형태에서, IBR은 촉매적으로 비활성인 EndoV 또는 촉매적으로 비활성인 hAAG이다.

[0266] 용어 "단리된(isolated)", "정제된" 또는 "생물학적으로 순수한"은 이의 천연 상태에서 발견되는 것과 같이, 일반적으로 동반되는 구성요소(components)가 다양한 정도로 없는 물질을 지칭한다. "단리하다"는 원래의 소스 또는 주변과의 분리 정도를 나타낸다. "정제하다"는 단리보다 더 높은 분리 정도를 나타낸다. "정제된" 또는 "생물학적으로 순수한" 단백질은 불순물이 단백질의 생물학적 특성에 실질적으로 영향을 미치거나 다른 부작용을 일으키지 않도록 다른 물질이 충분히 포함되어 있지 않다. 즉, 본 발명의 핵산 또는 펩티드는 재조합 DNA 기술

에 의해 생산될 때 세포 물질, 바이러스 물질 또는 배양 배지가 실질적으로 없거나, 화학적으로 합성될 때 화학적 전구체 또는 기타 화학 물질이 없는 경우 정제된 것이다. 순도 및 균질성은 일반적으로 분석 화학 기술(예를 들어, 폴리아크릴아미드 겔 전기영동 또는 고성능 액체 크로마토그래피)을 사용하여 결정된다. 용어 "정제된"은 핵산 또는 단백질이 전기영동 겔에서 본질적으로 하나의 밴드를 생성함을 나타낼 수 있다. 예를 들어, 인산화 또는 글리코실화와 같은 변경을 겪게 될 수 있는 단백질의 경우, 다른 변형은 다른 분리된 단백질을 생성할 수 있으며, 이는 별도로 정제될 수 있다.

[0267] "단리된 폴리뉴클레오티드"는, 본 발명의 핵산 분자가 유래된, 유기체의 자연적으로-발생하는 게놈에서, 해당 유전자에 측접한 유전자들이 없는 핵산(예를 들어, DNA)을 의미한다. 따라서, 이 용어는, 예를 들어, 벡터 내로; 자율 복제 플라스미드 또는 바이러스 내로; 또는 원핵 생물 또는 진핵 생물의 게놈 DNA 내로 통합되거나; 또는 다른 서열과 무관하게 별도의 분자(예를 들어, PCR 또는 제한 엔도뉴클레아제 분해에 의해 생성된 cDNA, 또는 게놈 또는 cDNA 단편)로 존재하는 재조합 DNA를 포함한다. 또한, 이 용어는 DNA 분자로부터 전사된 RNA 분자뿐만 아니라, 추가 폴리펩티드 서열을 코딩하는 하이브리드 유전자의 일부인 재조합 DNA를 포함한다.

[0268] "단리된 폴리펩티드"는 자연적으로 동반되는 구성요소로부터 분리된 본 발명의 폴리펩티드를 의미한다. 전형적으로, 폴리펩티드는 단백질 및 이와 자연적으로 결합한 자연적으로 발생하는 유기 분자가 중량을 기준으로 적어도 60% 없을 때 단리된 것이다. 바람직하게는, 제제(preparation)에는 본 발명의 폴리펩티드가 중량 기준으로 적어도 75%, 더 바람직하게는 적어도 90%, 및 가장 바람직하게는 적어도 99% 존재한다. 본 발명의 단리된 폴리펩티드는, 예를 들어, 천연 공급원으로부터 추출에 의해, 이러한 폴리펩티드를 코딩하는 재조합 핵산의 발현에 의해; 또는 단백질을 화학적으로 합성함으로써 수 득될 수 있다. 순도는 임의의 적절한 방법, 예를 들어, 컬럼 크로마토그래피, 폴리아크릴 아미드 겔 전기 영동, 또는 HPLC 분석에 의해 측정할 수 있다.

[0269] 본원에 사용된, 용어 "링커"는 결합(예를 들어, 공유 결합), 화학 그룹, 또는 2개의 분자 또는 모이어티, 예를 들어, 융합 단백질의 2개의 도메인을 연결하는 분자를 지칭한다. 일부 실시형태에서, 링커는, Cas9 도메인을 포함하는, RNA-프로그래밍가능한 뉴클레아제의 gRNA 결합 도메인, 및 핵산 편집 단백질(예를 들어, 시티딘 및/또는 아데노신 데아미나제)의 촉매 도메인을 결합시킨다. 일부 실시형태에서, 링커는 dCas9 및 핵산 편집 단백질을 결합시킨다. 전형적으로, 링커는 두 기, 분자, 또는 다른 모이어티 사이에 위치하거나, 측접하며, 공유 결합을 통해 각각에 연결되며, 그리하여 두 그룹을 연결한다. 일부 실시형태에서, 링커는 아미노산 또는 복수의 아미노산(예를 들어, 펩티드 또는 단백질)이다. 일부 실시형태에서, 링커는 유기 분자, 기, 중합체, 또는 화학적 모이어티이다. 일부 실시형태에서, 링커는 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 25개, 35개, 45개, 50개, 55개, 60개, 60개, 65개, 70개, 70개, 75개, 80개, 85개, 90개, 90개, 95개, 100개, 101개, 102개, 103개, 104개, 105개, 110개, 120개, 130개, 140개, 150개, 160개, 175개, 180개, 190개, 또는 200개의 아미노산 길이일 수 있다. 더 길거나 더 짧은 링커도 고려된다. 일부 실시형태에서, 링커는, XTEN 링커로도 지칭될 수 있는, 아미노산 서열 SGSETPGTSESATPES를 포함한다. 일부 실시형태에서, 링커는 아미노산 서열 SGGS를 포함한다. 일부 실시형태에서, 링커는 $(SGGS)_n$, $(GGGS)_n$, $(GGGGS)_n$, $(G)_n$, $(EAAAK)_n$, $(GGS)_n$, SGSETPGTSESATPES, 또는 $(XP)_n$ 모티프, 또는 이들 중 임의의 것의 조합을 포함하며, 여기서 n은 독립적으로 1 내지 30의 정수이고, X는 임의의 아미노산이다. 일부 실시형태에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15이다.

[0270] 일부 실시형태에서, 핵염기 편집기의 도메인은 SGGSSGGSSGSETPGTSESATPESSGGSSGGS, 또는 GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS 의 아미노산 서열을 포함하는 링커를 통해 융합된다. 일부 실시형태에서, 핵염기 편집기의 도메인은 XTEN 링커로도 지칭될 수 있는 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다. 일부 실시형태에서, 링커는 길이가 24개 아미노산이다. 일부 실시형태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPES를 포함한다. 일부 실시형태에서, 링커는 길이가 40개 아미노산이다. 일부 실시형태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS를 포함한다. 일부 실시형태에서, 링커는 길이가 64개 아미노산이다. 일부 실시형태에서, 링커는 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS 를 포함한다. 일부 실시형태에서, 링커는 길이가 92개 아미노산이다. 일부 실시형태에서, 링커는 아미노산 서열

PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP
GTSTEPSEGSAPGTSESATPESGPGSEPATS 를 포함한다.

[0271] 본 명세서에 사용된 용어 "돌연변이"는 서열, 예를 들어, 핵산 또는 아미노산 서열 내의 잔기의 또 다른 잔기로의 치환, 또는 서열 내의 하나 이상의 잔기의 결실 또는 삽입을 지칭한다. 돌연변이는 전형적으로 원래 잔기를 확인하고 이어서 서열 내 잔기의 위치 및 새로 치환된 잔기의 정체를 확인함으로써 본 명세서에 기술된다. 본 명세서에 제공된 아미노산 치환(돌연변이)을 만드는 다양한 방법은 당업계에 잘 알려져 있으며, 예를 들어 문헌(Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4^{th} ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))에 제공되어 있다.

[0272] 본 명세서에 사용된, 용어 "핵산" 및 "핵산 분자"는 핵염기 및 산성 모이어티를 포함하는 화합물, 예를 들어, 뉴클레오시드, 뉴클레오티드, 또는 뉴클레오티드 중합체를 지칭한다. 전형적으로, 중합체성 핵산, 예를 들어, 3개 이상의 뉴클레오티드를 포함하는 핵산 분자는, 인접한 뉴클레오티드가 포스포디에스터 연결을 통해 서로 연결되는, 선형 분자이다. 일부 실시형태에서, "핵산"은, 개별 핵산 잔기(예를 들어, 뉴클레오티드 및/또는 뉴클레오시드)를 지칭한다. 일부 실시형태에서, "핵산"은 3개 이상의 개별 뉴클레오티드 잔기를 포함하는 올리고뉴클레오티드 사슬을 지칭한다. 본 명세서에 사용된, 용어 "올리고뉴클레오티드" 및 "폴리뉴클레오티드"는 뉴클레오티드의 중합체(예를 들어, 적어도 3개의 뉴클레오티드의 스트링)를 지칭하기 위해 상호교환적으로 사용될 수 있다. 일부 실시형태에서, "핵산"은 단일 및/또는 이중 가닥 DNA뿐만 아니라, RNA를 포함한다. 핵산은, 예를 들어 게놈, 전사체, mRNA, tRNA, rRNA, siRNA, snRNA, 플라스미드, 코스미드, 염색체, 염색분체(chromatid), 또는 기타 자연적으로 발생하는 핵산 분자의 맥락에서 자연적으로 발생할 수 있다. 반면에, 핵산 분자는 비-자연적으로 발생하는 분자일 수 있는데, 예를 들어, 재조합 DNA 또는 RNA, 인공 염색체, 조작된 게놈, 또는 이의 단편 또는 합성 DNA, RNA, DNA/RNA 하이브리드이거나, 또는 비-자연적으로 발생하는 뉴클레오티드 또는 뉴클레오시드를 포함한다. 또한, 용어 "핵산", "DNA", "RNA", 및/또는 유사한 용어는 핵산 유사체, 예를 들어, 포스포디에스터 백본 이외의 것을 갖는 유사체를 포함한다. 핵산은 천연 공급원으로부터 정제할 수 있고, 재조합 발현 시스템을 사용하여 생산하고 임의로 정제하거나, 화학적 합성 등으로 얻을 수 있다. 적절한 경우, 예를 들어, 화학적으로 합성된 분자의 경우, 핵산은 화학적으로 변형된 염기 또는 당, 및 골격 변경을 갖는 유사체와 같은 뉴클레오시드 유사체를 포함할 수 있다. 핵산 서열은 달리 명시되지 않는 한, 5'에서 3' 방향으로 제시된다. 일부 실시형태에서, 핵산은 천연 뉴클레오시드(예를 들어, 아데노신, 티미딘, 구아노신, 시티딘, 우리딘, 데옥시아데노신, 데옥시티미딘, 데옥시구아노신 및 데옥시시티딘); 뉴클레오시드 유사체(예를 들어, 2-아미노아데노신, 2-티오티미딘, 이노신, 피롤로-피리미딘, 3-메틸아데노신, 5-메틸시티딘, 2-아미노아데노신, C5-브로모리 딘, C5-플루오로우리딘, C5-아이오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-시티딘, C5-메틸시티딘, 2-아미노아데노신, 7-데아자아데노신, 7-데아자구아노신, 8-옥소아데노신, 8-옥소구아노신, O(6)-메틸구아닌, 및 2-티오시티딘); 화학적으로 변형된 염기; 생물학적으로 변형된 염기(예를 들어, 메틸화된 염기); 삽입된(intercalated) 염기; 변형된 당(2'-, 예를 들어, 플루오로리보스, 리보스, 2'-데옥시리보스, 아라비노스, 및 헥소스); 및/또는 개질된 포스페이트 그룹(예를 들어, 포스포로티오에이트 및 5'-N-포스포라미다이트 연결)이거나 이를 포함한다.

[0273] 용어 "핵 위치결정 서열(nuclear localization sequence)", "핵 위치결정 신호" 또는 "NLS"는 단백질의 세포 핵으로의 도입을 촉진하는 아미노산 서열을 지칭한다. 핵 위치결정 서열은 당업계에 공지되어 있으며, 예를 들어, 2001년 5월 31일에 WO/2001/038547로 공개된, Plank 등이 2000년 11월 23일에 출원한, PCT 국제 출원 PCT/EP2000/011690에 기재되어 있으며, 이들의 내용은 예시적인 핵 위치결정 서열의 개시를 위해 본 명세서에 참조로 통합된다. 다른 실시형태에서, NLS는, 예를 들어, 문헌[Koblan *et al.*, Nature Biotech. 2018 doi:10.1038/nbt.4172]에 기술된 최적화된 NLS이다. 일부 실시형태에서, NLS는 아미노산 서열
KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK,
KKTELQTTNAENKTKKL, KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRK,
PKKKRKV, 또는 MDSLLMNRRKFLYQFKNVRWAKGRRETYLC 를 포함한다.

[0274] 본 개시는 핵산 프로그래밍가능한 핵산(예를 들어, DNA 또는 RNA) 결합 단백질을 제공한다. 핵산 프로그래밍가능한 핵산 결합 단백질은, 예를 들어, "핵산 프로그램가능한 DNA 결합 단백질" 또는 "napDNAbp"일 수 있다. "핵산 프로그래밍가능한 DNA 결합 단백질" 또는 "napDNAbp"라는 용어는 napDNAbp를 특정 핵산 서열로 가이드하는 가이드 핵산과 같은 핵산(예를 들어, DNA 또는 RNA)과 연관되는 단백질을 지칭한다. 예를 들어, Cas9 단백질은 Cas9 단백질을 가이드 RNA에 상보적인 특정 DNA 서열로 가이드하는 가이드 RNA와 연관될 수 있다. 일부 실시형태에서, napDNAbp는 Cas9 도메인, 예를 들어, 뉴클레아제 활성 Cas9, Cas9 닉카아제(nCas9), 또는 뉴클레아제

비활성 Cas9(dCas9)이다. 핵산 프로그래밍가능한 DNA 결합 단백질의 예는, 제한됨이 없이, Cas9(예를 들어, dCas9 및 nCas9), CasX, CasY, Cpf1, Cas12b/C2c1, 및 Cas12c/C2c3을 포함한다. 다른 핵산 프로그래밍가능한 DNA 결합 단백질도 본 개시의 범위 내에 있지만, 이들은 본 개시에서 구체적으로 나열되지 않을 수 있다.

[0275] 본 명세서에 사용된, "작용제를 수득하는"에서와 같이 "수득하는"은 작용제를 합성, 구매, 단리, 또는 달리 획득하는 것을 포함한다.

[0276] 용어 "RNA-프로그래밍가능한 뉴클레아제" 및 "RNA-가이드 뉴클레아제"는 절단을 위한 표적이 아닌 하나 이상의 RNA(들)와 함께 사용(예를 들어, 결합하거나 연관)된다. 일부 실시형태에서, RNA-프로그래밍가능한 뉴클레아제는, RNA와의 복합체화할 때, 뉴클레아제:RNA 복합체로 지칭될 수 있다. 전형적으로, 결합된 RNA(들)는 가이드 RNA(gRNA)로 지칭된다. gRNA는 2개 이상의 RNA의 복합체 또는 단일 RNA 분자로 존재할 수 있다. 단일 RNA 분자로 존재하는 gRNA는 단일 가이드 RNA(sgRNA)로 지칭될 수 있지만, "gRNA"는 단일 분자 또는 둘 이상의 분자의 복합체로 존재하는 가이드 RNA를 지칭하기 위해 상호교환적으로 사용된다. 전형적으로, 단일 RNA 종으로 존재하는 gRNA는 다음 2가지 도메인을 포함한다: (1) 표적 핵산에 대한 상동성을 공유하는 도메인(예를 들어, 표적에 대한 Cas9 복합체의 결합을 유도하는 것); 및 (2) Cas9 단백질에 결합하는 도메인. 일부 실시형태에서, 도메인 (2)는 tracrRNA로 공지된 서열에 상응하고, 스템-루프 구조를 포함한다. 예를 들어, 일부 실시형태에서, 도메인 (2)는 문헌[Jinek et al, Science 337:816-821(2012)]에 제공된 바와 같이 tracrRNA와 동일하거나 상동이며, 상기 문헌의 전체 내용은 본 명세서에 참조로 통합된다. gRNA의 다른 예(예를 들어, 도메인 2를 포함하는 것)는, "전환가능한 Cas9 뉴클레아제 및 이의 용도"라는 제목으로 2013년 9월 6일자로 출원된 미국 가특허 출원 U.S.S.N. 61/874,682호 및 "기능적 뉴클레아제를 위한 전달 시스템"이라는 제목으로 2013년 9월 6일에 출원된 미국 가특허 출원 U.S.S.N. 61/874,746호에서 찾을 수 있으며, 각각의 전체 내용은 그 전문이 본 명세서에 참조로 통합된다. 일부 실시형태에서, gRNA는 도메인 (1) 및 (2) 중 2개 이상을 포함하고, "확장된(extended) gRNA"로 지칭될 수 있다. 예를 들어, 확장된 gRNA는, 예를 들어, 본 명세서에 기재된 바와 같이 2개 이상의 Cas9 단백질에 결합하고, 2개 이상의 별개의 영역에서 표적 핵산에 결합할 것이다. gRNA는 뉴클레아제:RNA 복합체의 서열 특이성을 제공하는 표적 부위에 대한 뉴클레아제/RNA 복합체의 결합을 매개하는 표적 부위를 보완하는 뉴클레오티드 서열을 포함한다. 일부 실시형태에서, RNA-프로그래밍가능한 뉴클레아제는 (CRISPR-관련 시스템) Cas9 엔도뉴클레아제, 예를 들어, 다음 문헌 참조: "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., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma CM., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011)).

[0277] 단백질 또는 핵산과 관련하여 본 명세서에서 사용된 것과 같은 용어 "재조합"은 자연에서 발생하지 않지만, 인간 조작의 생성물인 단백질 또는 핵산을 의미한다. 예를 들어, 일부 실시형태에서, 재조합 단백질 또는 핵산 분자는 자연적으로 발생하는 서열과 비교하여 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 또는 적어도 7개의 돌연변이를 포함하는 아미노산 또는 뉴클레오티드 서열을 포함한다.

[0278] "감소하다"는 적어도 10%, 25%, 50%, 75%, 또는 100%의 음성적인 변경을 의미한다.

[0279] "참조"는 표준 또는 대조 조건을 의미한다.

[0280] "참조 서열"은 서열 비교를 위한 기준으로 사용되는 정의된 서열이다. 참조 서열은 특정 서열의 서브세트 또는 전체일 수 있으며; 예를 들어, 전체-길이 cDNA 또는 유전자 서열의 세그먼트, 또는 완전한 cDNA 또는 유전자 서열일 수 있다. 폴리펩티드의 경우, 참조 폴리펩티드 서열의 길이는 일반적으로 적어도 약 16개 아미노산, 적어도 약 20개 아미노산, 더 적어도 약 25개 아미노산, 훨씬 더 바람직하게는 약 35개 아미노산, 약 50개 아미노산, 또는 약 100개의 아미노산일 것이다. 핵산의 경우, 참조 핵산 서열의 길이는 일반적으로 적어도 약 50개의 뉴클레오티드, 적어도 약 60개의 뉴클레오티드, 적어도 약 75개의 뉴클레오티드, 및 약 100개 뉴클레오티드 또는 약 300개 뉴클레오티드 또는 상기 수치 부근 또는 그 사이의 임의의 정수일 것이다.

[0281] "특이적으로 결합한다"는 핵산 분자, 폴리펩티드, 또는 이의 복합체(예를 들어, 핵산 프로그래밍가능한 DNA 결합 단백질 및 가이드 핵산), 화합물, 또는 분자가 본 발명의 폴리펩티드 및/또는 핵산 분자를 인식하고 이에 결합하지만, 샘플, 예를 들어, 생물학적 샘플 내의 다른 분자를 실질적으로 인식하고 이에 결합하지 않는다는 것을 의미한다.

[0282] 본 발명의 방법에 유용한 핵산 분자는 본 발명의 폴리펩티드 또는 이의 단편을 코딩하는 임의의 핵산 분자를 포함한다. 이러한 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 일반적으로 실질적인 동일성을 나타낸다. 내인성 서열에 대해 "실질적인 동일성"을 갖는 폴리뉴클레오티드는 전형적으로 이중 가닥 핵산 분자의 적어도 하나의 가닥과 혼성화할 수 있다. 본 발명의 방법에 유용한 핵산 분자는 본 발명의 폴리펩티드 또는 이의 단편을 코딩하는 임의의 핵산 분자를 포함한다. 이러한 핵산 분자는 내인성 핵산 서열과 100% 동일할 필요는 없지만, 일반적으로 실질적인 동일성을 나타낸다. 내인성 서열에 대해 "실질적인 동일성"을 갖는 폴리뉴클레오티드는 전형적으로 이중-가닥 핵산 분자의 적어도 하나의 가닥과 혼성화할 수 있다. "혼성화하다"는 다양한 엄격도(stringency) 조건 하에서 상보적 폴리뉴클레오티드 서열(예를 들어, 본 명세서에 기재된 유전자) 또는 이의 일부 사이에 이중-가닥 분자를 형성하는 쌍을 의미한다. (예를 들어, 문헌[Wahl, G. M. and S. L. Berger (1987) Methods Enzymol. 152:399; Kimmel, A. R. (1987) Methods Enzymol. 152:507] 참조).

[0283] 예를 들어, 엄격한 염 농도는 일반적으로 NaCl 약 750 mM 및 트리소듐 시트레이트 75 mM 이하, 바람직하게는 NaCl 약 500 mM 및 트리소듐 시트레이트 50 mM 이하, 보다 바람직하게는 NaCl 약 250 mM 및 트리소듐 시트레이트 25 mM 이하일 것이다. 낮은 엄격도 혼성화는 유기 용매, 예를 들어, 포름아미드의 부재시 얻을 수 있는 반면, 높은 엄격도 혼성화는 적어도 약 35% 포름아미드, 더 바람직하게는 적어도 약 50% 포름아미드의 존재시에 얻을 수 있다. 엄격한 온도 조건은 일반적으로 적어도 약 30℃, 더 바람직하게는 적어도 약 37℃, 가장 바람직하게는 적어도 약 42℃의 온도를 포함한다. 혼성화 시간, 세제, 예를 들어, 소듐 도데실 설페이트(SDS) 농도와 같은 다양한 추가 파라미터, 및 담체 DNA의 포함 또는 배제가, 당업자에게 잘 알려져 있다. 필요에 따라 이러한 다양한 조건을 결합하여 다양한 수준의 엄격도가 달성된다. 한 실시형태에서, 혼성화는 750 mM NaCl, 75 mM 트리소듐 시트레이트, 및 1% SDS 중에서 30℃에서 일어날 것이다. 또 다른 실시형태에서, 혼성화는 500 mM NaCl, 50 mM 트리소듐 시트레이트, 1% SDS, 35% 포름아미드, 및 100 100 mu.g/㎖ 변성된 연어 정자 DNA(ssDNA) 중에서 37℃에서 일어날 것이다. 또 다른 실시형태에서, 혼성화는 250 mM NaCl, 25 mM 트리소듐 시트레이트, 1% SDS, 50% 포름아미드, 및 200 ㎍/㎖ ssDNA 중에서 42℃에서 일어날 것이다. 이러한 조건에 대한 유용한 변경은 당업자에게 쉽게 명백할 것이다.

[0284] 대부분의 적용에서, 혼성화에 뒤이은 세척 단계는 또한 엄격도에 있어서 다양할 것이다. 세척 엄격도 조건은 염분 농도와 온도로 정의할 수 있다. 상기한 것과 같이, 염분 농도를 낮추거나 온도를 높여 세척 엄격도를 높일 수 있다. 예를 들어, 세척 단계에 대한 엄격한 염 농도는 바람직하게는 NaCl 약 30 mM 및 트리소듐 시트레이트 3 mM 이하일 것이고, 가장 바람직하게는 NaCl 약 15 mM 및 트리소듐 시트레이트 1.5 mM 이하일 수 있다. 세척 단계에 대한 엄격한 온도 조건은 일반적으로 적어도 약 25℃, 더 바람직하게는 적어도 약 42℃, 더욱더 바람직하게는 적어도 약 68℃의 온도를 포함할 것이다. 일 실시형태에서, 세척 단계는 30 mM NaCl, 3 mM 트리소듐 시트레이트, 및 0.1% SDS 중에서 25℃에서 일어날 것이다. 더 바람직한 실시형태에서, 세척 단계는 15 mM NaCl, 1.5 mM 트리소듐 시트레이트, 및 0.1% SDS 중에서 42 C에서 일어날 것이다. 더 바람직한 실시형태에서, 세척 단계는 15 mM NaCl, 1.5 mM 트리소듐 시트레이트, 및 0.1% SDS에서 68℃에서 일어날 것이다. 이러한 조건에 대한 추가 변형은 당업자에게 자명할 것이다. 혼성화 기술은 당업자에게 잘 알려져 있으며, 예를 들어, 다음 문헌에 기재되어 있다: Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel *et al*. (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); 및 Sambrook *et al*., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

[0285] "실질적으로 동일한"은 참조 아미노산 서열(예를 들어, 본 명세서에 기재된 아미노산 서열 중 임의의 하나) 또는 핵산 서열(예를 들어, 본 명세서에 기재된 핵산 서열 중 어느 하나)에 대해 적어도 50% 동일성을 나타내는 폴리펩티드 또는 핵산 분자를 지칭한다. 한 실시형태에서, 이러한 서열은 비교에 사용된 서열과 아미노산 수준 또는 핵산에서 적어도 60%, 80% 또는 85%, 90%, 95% 또는 심지어 99% 동일하다.

[0286] 서열 동일성은 전형적으로 서열 분석 소프트웨어(예를 들어, 53705, 위스콘신, 메디슨, 유니버시티 애브뉴 1710 소재, 위스콘신 대학교 생명 공학 센터, 유전학 컴퓨터 그룹의 시퀀싱 소프트웨어 패키지, BLAST, BESTFIT, GAP, 또는 PILEUP/PRETTYBOX 프로그램)를 이용하여 측정된다. 이러한 소프트웨어는 다양한 치환, 결실, 및/또는 기타 변형에 대해 상동성의 정도를 할당하여 동일하거나 유사한 서열을 일치시킨다. 보존적 치환은 일반적으로 다음 그룹 내의 치환을 포함한다: 글라이신, 알라닌; 발린, 이소류신, 류신; 아스파르트산, 글루탐산, 아스파라긴, 글루타민; 세린, 트레오닌; 라이신, 아르기닌; 및 페닐알라닌, 티로신. 동일성의 정도를 결정하는 예시적인 접근법에서, 밀접하게 관련된 서열을 나타내는 e^{-3}과 e^{-100} 사이의 확률 점수와 함께, BLAST 프로그램이 사용될

수 있다. 예를 들어, COBALT는 다음 파라미터와 함께 사용된다.

[0287] "대상체"는 인간 또는 비-인간 포유동물, 예컨대 소, 말, 개, 양 또는 고양이를 포함하나, 이에 제한되지 않는, 포유동물을 의미한다. 대상체는 가축, 소, 염소, 닭, 말, 돼지, 토끼 및 양을 포함하되 이에 제한되지 않는, 노동력을 생산하고 상품, 예컨대 식품을 제공하기 위해 사육된 길들여진 동물이 포함된다.

[0288] 용어 "표적 부위"는 핵염기 편집기에 의해 변형된 핵산 분자 내의 서열을 의미한다. 한 실시형태에서, 표적 부위는 데아미나제 또는 데아미나제(예를 들어, 시티딘 또는 아데닌 데아미나제)를 포함하는 융합 단백질에 의해 탈아미노화된다.

[0289] RNA 프로그래밍가능한 뉴클레아제(예를 들어, Cas9)는 RNA:DNA 혼성화를 이용하여 DNA 절단 부위를 표적으로 하기 때문에 이러한 단백질은 원칙적으로 가이드 RNA에 의해 지정된 임의의 서열을 표적으로 할 수 있다. (예를 들어, 게놈을 변형하기 위해) 부위-특이적 절단을 위해 Cas9와 같은 RNA-프로그래밍가능한 뉴클레아제를 사용하는 방법은 당업계에 공지되어 있다(예를 들어, 각각의 전체 내용이 본 명세서에 참조로 통합되는 다음 문헌 참조: Cong, L. et al, Multiplex genome engineering using CRISPR/Cas systems. Cong, L. et ah, Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali, P. et ah, RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang, W.Y. et ah, Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek, M. et ah, RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo, J.E. et ah, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic acids research (2013); Jiang, W. et ah RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239 (2013)).

[0290] 본 명세서에 제공된 범위는 해당 범위 내의 모든 값에 대한 약칭으로 이해된다. 예를 들어, 1 내지 50의 범위는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 또는 50으로 이루어진 그룹으로부터의 임의의 숫자, 숫자의 조합, 또는 하위-범위를 포함하는 것으로 이해된다.

[0291] 구체적으로 언급되거나 문맥상 명백하지 않는 한, 본 명세서에 사용된 용어 "또는"은 포괄적인 것으로 이해된다. 구체적으로 언급되거나 문맥상 명백하지 않은 한, 본 명세서에 사용된 용어 "a", "an" 및 "the"는 단수 또는 복수로 이해된다.

[0292] 구체적으로 언급되거나 문맥상 명백하지 않는 한, 본 명세서에 사용된 용어 "약"은, 예를 들어, 평균의 2 표준편차 내에서 당업계의 정상적인 허용 오차 범위 내로 이해된다. 약은 명시된 값의 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05% 또는 0.01% 이내로 이해될 수 있다. 문맥상 명백하지 않는 한, 본 명세서에 제공된 모든 수치 값은 약이라는 용어에 의해 변형된다.

[0293] 본 명세서에서 변수의 임의의 정의에서의 화학 그룹 목록에 대한 인용은 해당 변수를 임의의 단일 그룹 또는 나열된 그룹의 조합으로 정의하는 것을 포함한다. 본 명세서에서 변수 또는 양상에 대한 실시형태의 인용은 임의의 단일 실시형태로서 또는 임의의 다른 실시형태 또는 이의 부분과 조합하여 해당 실시형태를 포함한다.

[0294] 본 명세서에 제공된 임의의 조성물 또는 방법은 본 명세서에 제공된 임의의 다른 조성물 및 방법 중 하나 이상과 조합될 수 있다.

도면의 간단한 설명

[0295] 도 1은 아데노신 핵염기 편집기의 모델을 도시하고, 방관자 돌연변이유발을 위한 구조적 기초를 부분적으로 제공한다.

도 2는 염기 편집을 위한 표적 DNA의 위치 예측을 나타내는 모델이다.

도 3은 도 2에서 표적 DNA의 위치에 근접한 데아미나제 도메인의 위치를 나타내는 모델을 도시한다.

도 4는 하나 이상의 데아미나제 도메인이 Cas9에 삽입될 수 있는 확인된 영역을 나타내는 아데노신 핵염기 편집기의 모델이다. 데아미나제 도메인이 단일 가닥 DNA(자홍색)를 표적으로 할 수 있는 위치에 근접한 루프(노란색). 관심 영역은 A, B, C, D, E, F, G 및 H로 표시된 영역을 포함한다.

도 5는 도 4의 모델의 확대도이며, 영역 B, C, D, E 및 F의 잔기를 보여준다.

도 6은 도 4의 모델의 확대도이며, 영역 F, G 및 H의 잔기를 보여준다.

도 7은 도 4의 모델의 확대도이며, 영역 A, B, C, D 및 E의 잔기를 보여준다.

도 8은 고-처리량 시험관내 탈아미노화 분석을 도시한다. 프로브의 불요성 탈아미노화는 염기 편집기용 기질을 함유하는 표적-적중 프로브를 지니는 핵염기 편집기를 함유하는 반응과 표적-이탈 탈아미노화를 검출하기 위한 프로브를 함유하는 염기 편집기 부재시의 반응과 비교하여 표적-적중 탈아미노화와 구별될 수 있다.

도 9는 표적-이탈 탈아미노화에 대한 형광 분석 결과를 나타내는 그래프이다.

도 10은 아데노신 염기 편집기(ABE) 대 트랜스형(*in trans*) TadA가 있는 ABE 시스템의 비교를 나타내는 그래프이다.

도 11은 불요성 표적-이탈 염기 편집을 위한 잠재적 기질을 도시한다.

도 12는 시스형(*in cis*) 및 트랜스형 데아미나제의 활성을 평가하기 위한 분석을 도시한다.

도 13은 시스형-트랜스형 분석에서 rAPOBEC1의 활성을 나타내는 그래프이다.

도 14는 시스형-트랜스형 분석에서 TadA-TadA7.10의 활성을 나타내는 그래프이다.

도 15는 염기 편집기 컨텍스트(트랜스형 ABE에서)에서 TadA-TadA7.10에 대해 더 낮은 트랜스형 활성이 관찰되었음을 도시한다.

도 16은 GFP 발현에 대한 용량-반응 결과를 나타낸 그래프이다. 빈 벡터로의 pmaxGFP 플라스미드의 적정은 감소된 GFP 발현 수준을 초래했다.

도 17은 아데노신 핵염기 편집기 ABE의 시스형 및 트랜스형 활성에 대한 용량-반응을 나타내는 그래프이다.

도 18은 시티딘 핵염기 편집기 BE4의 시스형 및 트랜스형 활성에 대한 용량-반응을 나타내는 그래프이다.

도 19는 감소된 불요성 탈아미노화에 대한 데아미나제의 스크리닝 결과를 나타내는 그래프이다. 데아미나제 ppAPOBEC-2(10), mAPOBEC-2(8), mAPOBEC-3(12), 및 mfAPOBEC-4(22)는 높은 시스형/트랜스형 활성을 나타냈다.

도 20a-q는 조사된 편집기의 염기 편집 활성을 나타낸다. 도 20a는 XTEN 링커에 의해 Cas9에 융합된 TadA를 지니는, ABE7.10의 개략도이다. 도 20b-q는 표적 서열 GAACACAAAGCATAGACTGC(HEK2) 및 GGACAGCTTTTCCTAGACAG(T39)를 지니는 R-루프의 범위 내에서 표적화 가닥에 대한 A에서 G로의 탈아미노화 백분율로 예시적인 내부 융합 염기 편집기의 염기 편집 활성을 보여준다. 도 20b, ABE7.10의 편집 활성. 도 20c, ISLAY008의 편집 활성. 도 20d, ISLAY003의 편집 활성. 도 20e, ISLAY002의 편집 활성. 도 20F, ISLAY007 편집 활성. 도 20g, ISLAY001의 편집 활성. 도 20h, ISLAY005의 편집 활성. 도 20i, ISLAY006의 편집 활성. 도 20j, ISLAY004 편집 활성. 도 20k, ISLAY021의 편집 활성. 도 20l, ISLAY031의 편집 활성. 도 20m, ISLAY020의 편집 활성. 도 20n, ISLAY036의 편집 활성. 도 20o, ISLAY035의 편집 활성. 도 20p, ISLAY028의 편집 활성. 도 20q, ISLAY009 편집 활성.

도 21a-b는 예시적인 염기 편집기의 개략도를 보여준다. 도 21a는 예시적인 염기 편집기 ABE7.10 및 예시적인 염기 편집기(IBE002, IBE004, IBE005, IBE006, IBE008, IBE009, 및 IBE020)의 개략도를 도시한다. 도 21b는 데아미나제 삽입 위치를 보여주는, 상기 염기 편집기의 공간적 삽화를 보여준다.

도 22a-d는 29개의 상이한 게놈 표적에서 ABE7.10과 비교한 예시적인 내부 융합 염기 편집기의 염기 편집 효율을 도시한다. 도 22a는 편집 효율이 최상의 위치에서 ABE7.10 편집으로 정규화된 것을 보여준다. 도 22b는 ABE7.10과 요약 및 비교한 IBE의 최대 편집 효율을 보여준다. 도 22c는 각 염기 편집기에 대한 피크 편집 위치의 가우스 평활화 표현(Gaussian smoothened representation)을 보여준다. 도 22d는 29개의 테스트 대상에서 정규화된 편집의 히트맵을 보여준다.

도 23은 각 부위에서 ABE7.10에 대해 정규화된 29개의 상이한 표적으로서 트랜스 편집 분석에 의해 측정된 불요성 탈아미노화를 도시한다.

도 24a-f는 6개의 게놈 유전자좌에서 A-염기 편집자의 편집 백분율을 보여준다: HEK4(도 24a), FANCF(도 24b), HEK-3(도 24c), HEK2-YY(도 24d), EMX1 (도 24e), HEK2(도 24f). X-축: 1은 PAM에서 가장 멀고 20은 PAM 근위인 핵염기 위치(PAM은 위치 21-23). Y축: Illumina 시퀀싱으로 측정한 A에서 G로의 편집 백분율.

도 25a-f는 6개의 게놈 유전자좌에서 C-염기 편집기의 편집 백분율: HEK4(도 25a), FANCF(도 25b), HEK-3(도

25c), HEK2-YY(도 25d), EMX1(도 25e), HEK2(도 25f). X-축: 1은 PAM에서 가장 멀고 20은 PAM 근위인 핵염기 위치(PAM은 위치 21-23). Y축: Illumina 시퀀싱으로 측정한 A에서 G로의 편집 백분율.

발명을 실시하기 위한 구체적인 내용

[0296] **발명의 상세한 설명**

[0297] 이하에 기술된 바와 같이, 본 발명은 비-표적 탈아미노화가 감소된 염기 편집기, 상기 염기 편집기를 이용하는 방법과, 예를 들어, 프로그래밍된 표적-적중 탈아미노화와 비교하여, 염기 편집기가 감소된 비-표적 탈아미노화를 갖는 것으로 특성규명하기 위한 분석을 특징으로 한다.

[0298] *아데노신 데아미나제*

[0299] 일부 실시형태에서, 본 발명의 핵염기 편집기는 아데노신 데아미나제 도메인을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 아데노신 데아미나제는 아데닌을 탈아미노화할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 아데노신 데아미나제는 DNA의 데옥시아데노신 잔기에서 아데닌을 탈아미노화할 수 있다. 아데노신 데아미나제는 임의의 적합한 유기체(예를 들어, E. 콜라이)로부터 유래될 수 있다. 일부 실시형태에서, 아데닌 데아미나제는 본 명세서에 제공된 돌연변이(예를 들어, ecTadA의 돌연변이) 중 임의의 것에 상응하는 하나 이상의 돌연변이를 포함하는 자연적으로 발생하는 아데노신 데아미나제이다. 당업자는, 예를 들어, 서열 정렬 및 상동성 잔기의 결정에 의해 임의의 상동성 단백질에서 상응하는 잔기를 확인할 수 있을 것이다. 따라서, 당업자는 본 명세서에 기재된 임의의 돌연변이, 예를 들어, ecTadA에서 확인된 임의의 돌연변이에 상응하는 임의의 자연적으로 발생하는 아데노신 데아미나제(예를 들어, ecTadA에 대한 상동성을 가짐)에서 돌연변이를 생성할 수 있을 것이다. 일부 실시형태에서, 아데노신 데아미나제는 원핵생물로부터 유래된다. 일부 실시형태에서, 아데노신 데아미나제는 박테리아로부터의 것이다. 일부 실시형태에서, 아데노신 데아미나제는 박테리아에서 유래된 것이다. 일부 실시형태에서, 아데노신 데아미나제는 에스케리치아 콜라이(*Escherichia coli*), 스타필로코커스 아우레우스(*Staphylococcus aureus*), 살모넬라 티피(*Salmonella typhi*), 슈와넬라 퓨트레파시엔스(*Shewanella putrefaciens*), 해모필루스 인플루엔자(*Haemophilus influenzae*), 카울로박터 크레센투스(*Caulobacter crescentus*), 또는 바실러스 서브틸리스(*Bacillus subtilis*)에서 유래된 것이다. 일부 실시형태에서, 아데노신 데아미나제는 E. 콜라이에서 유래된 것이다.

[0300] 한 실시형태에서, 본 발명의 융합 단백질은 Cas9 닉카아제에 연결된 TadA7.10에 연결된 야생형 TadA를 포함한다. 특정 실시형태에서, 융합 단백질은 단일 TadA7.10 도메인(예를 들어, 단량체로 제공됨)을 포함한다. 다른 실시형태에서, ABE7.10 편집기는 이종이량체를 형성할 수 있는 TadA7.10 및 TadA(wt)를 포함한다. 관련 서열은 다음과 같다:

[0301] TadA(wt):

[0302] SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIM
ALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVL
HHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD

[0303] TadA7.10:

[0304] SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

[0305] 일부 실시형태에서, 아데노신 데아미나제는 본 명세서에 제공된 아데노신 데아미나제 중 어느 하나에 제시된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 본 명세서에 제공된 아데노신 데아미나제는 하나 이상의 돌연변이(예를 들어, 본 명세서에 제공된 임의의 돌연변이)를 포함할 수 있음을 이해해야 한다. 본 개시는 특정 백분율 동일성 및 본 명세서에 기재된 임의의 돌연변이 또는 이들의 조합을 갖는 임의의 데아미나제 도메인을 제공한다. 일부 실시형태에서, 아데노신 데아미나제는 참조 서열, 또는 본 명세서에 제공된 아데노신 데아미나제와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개, 또는 그 이상의 돌연변이를 갖는 아미노

산 서열을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 당업계에 공지되거나 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 5개, 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 35개, 적어도 40개, 적어도 45개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 110개, 적어도 120개, 적어도 130개, 적어도 140개, 적어도 150개, 적어도 160개, 또는 적어도 170개의 동일한 연속 아미노산 잔기를 포함한다.

[0306] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 D108X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 D108G, D108N, D108V, D108A, 또는 D108Y 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다. 그러나, 본 명세서에 제공된 것과 같이 돌연변이될 수 있는 상동성 아미노산 잔기를 확인하기 위해 추가의 데아미나제가 유사하게 정렬될 수 있음을 인식해야 한다.

[0307] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A106X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A106V 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0308] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 E155X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X의 존재는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 E155D, E155G, 또는 E155V 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0309] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 D147X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X의 존재는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 D147Y 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0310] 본 명세서에 제공된 임의의 돌연변이(예를 들어, TadA 참조 서열의 ecTadA 아미노산 서열에 기초한 것)는 다른 아데노신 데아미나제, 예컨대, E. 콜라이 TadA(ecTadA), S. 아우레우스 TadA(saTadA), 또는 다른 아데노신 데아미나제(예를 들어, 박테리아 아데노신 데아미나제)에 도입될 수 있음을 인식해야 한다. ecTadA에서 돌연변이된 잔기에 얼마나 상동성이 있는지는 숙련된 기술자에게 명백할 것이다. 따라서, ecTadA에서 식별된 임의의 돌연변이는 상동성 아미노산 잔기를 갖는 다른 아데노신 데아미나제에서 만들어질 수 있다. 또한, 본 명세서에 제공된 임의의 돌연변이는 ecTadA 또는 또 다른 아데노신 데아미나제에서 개별적으로 또는 임의의 조합으로 만들어질 수 있음을 인식해야 한다. 예를 들어, 아데노신 데아미나제는 TadA 참조 서열에서의 D108N, A106V, E155V, 및/또는 D147Y 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 함유할 수 있다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 다음의 돌연변이 그룹(돌연변이 그룹은 ";"로 분리됨), 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다: D108N 및 A106V; D108N 및 E155V; D108N 및 D147Y; A106V 및 E155V; A106V 및 D147Y; E155V 및 D147Y; D108N, A106V, 및 E55V; D108N, A106V, 및 D147Y; D108N, E55V, 및 D147Y; A106V, E55V, 및 D 147Y; 및 D108N, A106V, E55V, 및 D147Y. 그러나, 본 명세서에 제공된 상응하는 돌연변이의 임의의 조합이 아데노신 데아미나제(예를 들어, ecTadA)에서 만들어질 수 있음을 인식해야 한다.

[0311] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 H8X, T17X, L18X, W23X, L34X, W45X, R51X, A56X, E59X, E85X, M94X, I95X, V102X, F104X, A106X, R107X, D108X, K110X, M118X, N127X, A138X, F149X, M151X, R153X, Q154X, I156X, 및/또는 K157X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함하며, 여기서 X의 존재는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 H8Y, T17S, L18E, W23L, L34S, W45L, R51H, A56E, 또는 A56S, E59G, E85K, 또는 E85G, M94L, 1951, V102A, F104L, A106V, R107C, 또는 R107H, 또는 R107P, D108G, 또는 D108N, 또는 D108V, 또는 D108A, 또는 D108Y, K1101, M118K, N127S, A138V, F149Y, M151V, R153C, Q154L, I156D, 및/또는 K157R 돌연변이, 또는 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함한다.

[0312] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 H8X, D108X, 및/또는 N127X 돌연

변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함하며, 여기서 X는 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 H8Y, D108N, 및/또는 N127S 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함한다.

[0313] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 H8X, R26X, M61X, L68X, M70X, A106X, D108X, A109X, N127X, D147X, R152X, Q154X, E155X, K161X, Q163X, 및/또는 T166X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 H8Y, R26W, M61I, L68Q, M70V, A106T, D108N, A109T, N127S, D147Y, R152C, Q154H 또는 Q154R, E155G 또는 E155V 또는 E155D, K161Q, Q163H, 및/또는 T166P 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함한다.

[0314] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8X, D108X, N127X, D147X, R152X, 및 Q154X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8X, M61X, M70X, D108X, N127X, Q154X, E155X, 및 Q163X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 7개, 또는 8개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8X, D108X, N127X, E155X, 및 T166X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에 상응하는 아미노산 이외의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 H8X, A106X, D108X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 또는 6개의 돌연변이, 또 다른 아데노신 데아미나제에서의 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 H8X, R126X, L68X, D108X, N127X, D147X, 및 E155X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 7개, 또는 8개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8X, D108X, A109X, N127X, 및 E155X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다.

[0315] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, D108N, N127S, D147Y, R152C, 및 Q154H로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, M61I, M70V, D108N, N127S, Q154R, E155G, 및 Q163H로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 7개, 또는 8개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, D108N, N127S, E155V, 및 T166P로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, A106T, D108N, N127S, E155D, 및 K161Q로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, R126W, L68Q, D108N, N127S, D147Y, 및 E155V로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 7개, 또는 8개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, D108N, A109T, N127S, 및 E155G로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함한다.

[0316] 일부 실시형태에서, 아데노신 데아미나제는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이

를 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 D108N, D108G, 또는 D108V 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A106V 및 D108N 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R107C 및 D108N 돌연변이, 또는 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, D108N, N127S, D147Y, 및 Q154H 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, R24W, D108N, N127S, D147Y, 및 E155V 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 D108N, D147Y, 및 E155V 돌연변이, 또는 다른 아데노신 데아미나제에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, D108N, 및 S127S 돌연변이, 또는 다른 아데노신 데아미나제에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A106V, D108N, D147Y, 및 E155V 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이들을 포함한다.

[0317] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 S2X, H8X, I49X, L84X, H123X, N127X, I156X, 및/또는 K160X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함하며, 여기서 X의 존재는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 S2A, H8Y, I49F, L84F, H123Y, N127S, I156F, 및/또는 K160S 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함한다.

[0318] 일부 실시형태에서, 아데노신 데아미나제는 L84X 돌연변이 아데노신 데아미나제를 포함하고, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 L84F 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0319] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H123X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H123Y 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0320] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 I157X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 I157F 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0321] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 L84X, A106X, D108X, H123X, D147X, E155X, 및 I156X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 또는 7개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 S2X, I49X, A106X, D108X, D147X, 및 E155X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 또는 6개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산의 존재를 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8X, A106X, D108X, N127X, 및 K160X로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 돌연변이 또는 돌연변이들을 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 아미노산의 존재를 나타낸다.

[0322] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 L84F, A106V, D108N, H123Y, D147Y, E155V, 및 I156F로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 6개, 또는 7개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 S2A, I49F, A106V, D108N, D147Y, 및 E155V로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 5개, 또는 6개의 돌연변이를 포함한다.

[0323] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H8Y, A106T, D108N, N127S, 및 K160S로 이루어진 군으로부터 선택된 1개, 2개, 3개, 4개, 또는 5개의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이 또는 돌연변이들을 포함한다.

[0324] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 E25X, R26X, R107X, A142X, 및/또는 A143X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함하며, X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 모든 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 E25M, E25D, E25A, E25R, E25V, E25S, E25Y, R26G, R26N, R26Q, R26C, R26L, R26K, R107P, R07K, R107A, R107N, R107W, R107H, R107S, A142N, A142D, A142G, A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, 및/또는 A143R 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에 상응하는 본 명세서에 기재된 하나 이상의 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함한다.

[0325] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 E25X 돌연변이, 또는 또 다른 아데노신 데아미나제(예를 들어, ecTadA)에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 E25M, E25D, E25A, E25R, E25V, E25S, 또는 E25Y 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0326] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R26X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R26G, R26N, R26Q, R26C, R26L, 또는 R26K 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0327] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R107X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R107P, R07K, R107A, R107N, R107W, R107H, 또는 R107S 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0328] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A142X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A142N, A142D, A142G 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0329] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A143X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, 및/또는 A143R 돌연변이, 또는 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0330] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 H36X, N37X, P48X, I49X, R51X, M70X, N72X, D77X, E134X, S146X, Q154X, K157X, 및/또는 K161X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함하며, 여기서 X의 존재는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 하나 이상의 H36L, N37T, N37S, P48T, P48L, I49V, R51H, R51L, M70L, N72S, D77G, E134G, S146R, S146C, Q154H, K157N, 및/또는 K161T 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 하나 이상의 상응하는 돌연변이들을 포함한다.

[0331] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H36X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 H36L 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0332] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 N37X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 N37T, 또는 N37S 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0333] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 P48X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 P48T, 또는 P48L 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0334] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R51X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R51H, 또는 R51L 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0335] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 S146X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 S146R, 또는 S146C 돌연변이, 또는 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0336] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 K157X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 K157N 돌연변이, 또는 또 다른 아데노신 데아미나제)에서의 상응하는 돌연변이를 포함한다.

[0337] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 P48X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 P48S, P48T, 또는 P48A 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0338] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A142X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 A142N 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0339] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 W23X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 W23R, 또는 W23L 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함한다.

[0340] 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R152X 돌연변이, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이를 포함하며, 여기서 X는 야생형 아데노신 데아미나제에서의 상응하는 아미노산 이외의 임의의 아미노산을 나타낸다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에서의 R152P, 또는 R52H 돌연변이, 또는 또 다른 아데노신 데아미나제의 상응하는 돌연변이를 포함한다.

[0341] 한 실시형태에서, 아데노신 데아미나제는 돌연변이 H36L, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F, 및 K157N을 포함할 수 있다. 일부 실시형태에서, 아데노신 데아미나제는 TadA 참조 서열에 비해 다음과 같은 돌연변이 조합을 포함하며, 조합의 각 돌연변이는 "_"로 분리되고 돌연변이의 각 조합은 괄호 내에 있다:

(R107C_D108N),
(H8Y_D108N_S 127S_D 147Y_Q154H), (H8Y_R24W_D108N_N127S_D147Y_E155V),
(D108N_D147Y_E155V), (H8Y_D108N_S 127S), (H8Y_D108N_N127S_D147Y_Q154H),
(A106V D108N D147Y E155V) (D108Q D147Y E155V) (D108M_D147Y_E155V),
(D108L_D147Y_E155V), (D108K_D147Y_E155V), (D108I_D147Y_E155V),
(D108F_D147Y_E155V), (A106V_D108N_D147Y), (A106V_D108M_D147Y_E155V),
(E59A_A106V_D108N_D147Y_E155V), (E59A cat
dead_A106V_D108N_D147Y_E155V),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F), (D103A_D014N),
(G22P_D 103 A_D 104N), (G22P_D 103 A_D 104N_S 138 A) , (D 103 A_D 104N_S
138A),
(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I15
6F),
(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I15
6F), (R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I15
6F), (R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),
(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I15
6F),
(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),
(A106V_D108N_A142N_D147Y_E155V),
(R26G_A106V_D108N_A142N_D147Y_E155V),
[0342] (E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),

(R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),
(E25D_R26G_A106V_D108N_A142N_D147Y_E155V),
(A106V_R107K_D108N_A142N_D147Y_E155V),
(A106V_D108N_A142N_A143G_D147Y_E155V),
(A106V_D108N_A142N_A143L_D147Y_E155V),
(H36L_R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F _K157N),
(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),
(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),
(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),
(N72S_L84F_A106V_D108N_H123Y_S 146R_D147Y_E155V_I156F),
(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
57N),
(H36L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_S 146R_D147Y_E155V_I156F_K161T),
(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),
(H36L_G67V_L84F_A106V_D108N_H123Y_S 146T_D147Y_E155V_I156F),
(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),
(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),
(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S 146C_D147Y_E155V_I156F
_K157N), (N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F_K157N_K161T),
(L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F_K157N_K160E_K161T),
(L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F_K157N_K160E), (R74Q

[0343] L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),
(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F), (P48S_A142N),
(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),
(P48T_I49V_A142N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F
_K157N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S
146C_A142N_D147Y_E155V_I156F
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F
_K157N),
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S
146C_D147Y_E155V_ I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S
146C_D147Y_E155V_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S
146C_A142N_D147Y_E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F
_K157N),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S 146R_D147Y_E155V_I156F
_K161T),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S
146C_D147Y_R152H_E155V_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S 146C_D147Y_R152P_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S
146C_D147Y_R152P_E155V _I156F _K157N),

[0344]

(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S 146C_D147Y_E155
V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S
146C_D147Y_R152P _E155V_I156F _K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S 146R_D147Y_E155V_I156F
_K161T),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S
146C_D147Y_R152P_E155V _I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S 146C_D147Y_R152P_E155
V_I156F _K157N).

[0345]

[0346] *시티딘 데아미나제*

[0347] 한 실시형태에서, 본 발명의 융합 단백질은 시티딘 데아미나제를 포함한다. 일부 실시형태에서, 본 명세서에 제공된 시티딘 데아미나제는 시토신 또는 5-메틸시토신을 우라실 또는 티민으로 탈아미노화할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 시토신 데아미나제는 DNA에서 시토신을 탈아미노화할 수 있다. 시티딘 데아미나

제는 임의의 적합한 유기체로부터 유래될 수 있다. 일부 실시형태에서, 시티딘 데아미나제는 본 명세서에 제공된 임의의 돌연변이에 상응하는 하나 이상의 돌연변이를 포함하는 자연적으로 발생하는 시티딘 데아미나제이다. 당업자는, 예를 들어, 서열 정렬 및 상동성 잔기의 결정에 의해, 임의의 상동성 단백질에서 상응하는 잔기를 확인할 수 있을 것이다. 따라서, 당업자는 본 명세서에 기재된 임의의 돌연변이에 상응하는 임의의 자연적으로 발생하는 시티딘 데아미나제에서 돌연변이를 생성할 수 있을 것이다. 일부 실시형태에서, 시티딘 데아미나제는 원핵 생물로부터 유래된다. 일부 실시형태에서, 시티딘 데아미나제는 박테리아로부터 유래된다. 일부 실시형태에서, 시티딘 데아미나제는 포유동물(예를 들어, 인간)로부터 유래되는 것이다.

[0348] 일부 실시형태에서, 시티딘 데아미나제는 본 명세서에 제시된 시티딘 데아미나제 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 본 명세서에 제공된 시티딘 데아미나제는 하나 이상의 돌연변이(예를 들어, 본 명세서에 제공된 임의의 돌연변이)를 포함할 수 있음을 이해해야 한다. 본 개시는 특정 백분율 동일성에 추가하여 본 명세서에 기재된 임의의 돌연변이 또는 이의 조합을 갖는 임의의 데아미나제 도메인을 제공한다. 일부 실시형태에서, 시티딘 데아미나제는 참조 서열과 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개, 또는 그 이상의 돌연변이를 갖는 아미노산 서열, 또는 본 명세서에 제공된 시티딘 데아미나제 중 임의의 것을 포함한다. 일부 실시형태에서, 시티딘 데아미나제는 당업계에 공지되거나 본 명세서에 기재된 아미노산 서열 중 어느 하나와 비교하여 적어도 5개, 적어도 10개, 적어도 15개, 적어도 20개, 적어도 25개, 적어도 30개, 적어도 35개, 적어도 40개, 적어도 45개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 110개, 적어도 120개, 적어도 130개, 적어도 140개, 적어도 150개, 적어도 160개, 또는 적어도 170개의 동일한 연속 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0349] 본 발명의 융합 단백질은 핵산 편집 도메인을 포함한다. 일부 실시형태에서, 핵산 편집 도메인은 C에서 U로의 염기 변화를 촉매할 수 있다. 일부 실시형태에서, 핵산 편집 도메인은 데아미나제 도메인이다. 일부 실시형태에서, 데아미나제는 시티딘 데아미나제 및 아데노신 데아미나제이다. 일부 실시형태에서, 데아미나제는 아포리포단백질 B mRNA-편집 복합체(APOBEC) 패밀리 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC1 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC2 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3A 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3B 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3C 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3D 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3E 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3F 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3G 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC3H 데아미나제이다. 일부 실시형태에서, 데아미나제는 APOBEC4 데아미나제이다. 일부 실시형태에서, 데아미나제는 활성화-유도된 데아미나제(AID)이다. 일부 실시형태에서, 데아미나제는 척추동물 데아미나제이다. 일부 실시형태에서, 데아미나제는 무척추동물 데아미나제이다. 일부 실시형태에서, 데아미나제는 인간, 침팬지, 고릴라, 원숭이, 소, 개, 랫트, 또는 마우스 데아미나제이다. 일부 실시형태에서, 데아미나제는 인간 데아미나제이다. 일부 실시형태에서, 데아미나제는 랫트 데아미나제, 예를 들어, rAPOBEC1이다. 일부 실시형태에서, 데아미나제는 칠성장어(페트로미존 마리누스(Petromyzon marinus)) 시티딘 데아미나제1(pmCDA1)이다. 일부 실시형태에서, 데아미나제는 인간 APOBEC3G이다. 일부 실시형태에서, 데아미나제는 인간 APOBEC3G의 단편이다. 일부 실시형태에서, 데아미나제는 D316R D317R 돌연변이를 포함하는 인간 APOBEC3G 변이체이다. 일부 실시형태에서, 데아미나제는 인간 APOBEC3G의 단편이고 D316R D317R 돌연변이에 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, 핵산 편집 도메인은 본 명세서에 기재된 임의의 데아미나제의 데아미나제 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.

[0350] 특정 실시형태에서, 본 명세서에 제공된 융합 단백질은 융합 단백질의 염기 편집 활성을 개선하는 하나 이상의 특징을 포함한다. 예를 들어, 본 명세서에 제공된 임의의 융합 단백질은 감소된 뉴클레아제 활성을 갖는 Cas9 도메인을 포함할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 뉴클레아제 활성을 갖지 않는 Cas9 도메인(dCas9), 또는 Cas9 닉카아제(nCas9)로 지칭되는 이중나선화된(duplexed) DNA 분자의 한 가닥을 절단하는 Cas9 도메인을 가질 수 있다.

[0351] *핵염기 편집기의 Cas9 도메인*

[0352] 일부 양상에서, 핵산 프로그래밍가능한 DNA 결합 단백질(napDNAbp)은 Cas9 도메인이다. Cas9 도메인의 비제한적인 예시가 본 명세서에 제공된다. Cas9 도메인은 뉴클레아제 활성 Cas9 도메인, 뉴클레아제 비활성 Cas9 도메인, 또는 Cas9 닉카아제일 수 있다. 일부 실시형태에서, Cas9 도메인은 뉴클레아제 활성 도메인이다. 예를 들어, Cas9 도메인은 이중나선화된 핵산의 두 가닥(예를 들어, 이중나선화된 DNA 분자의 두 가닥)을 절단하는 Cas9 도메인일 수 있다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 기재된 아미노산 서열 중 어느 하나를 포함한다. 일부 실시형태에서, Cas9 도메인은, 본 명세서에 제시된 아미노산 서열 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 제시된 아미노산 서열 중 어느 하나와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개, 또는 그 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 제시된 아미노산 서열 중 어느 하나와 비교하여 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 350개, 적어도 400개, 적어도 500개, 적어도 600개, 적어도 700개, 적어도 800개, 900개, 적어도 1000개, 적어도 1100개, 또는 적어도 1200개의 동일한 연속(contiguous) 아미노산 잔기를 포함한다.

[0353] 일부 실시형태에서, Cas9 도메인은 뉴클레아제-비활성 Cas9 도메인(dCas9)이다. 예를 들어, dCas9 도메인은 이중나선화된 핵산 분자의 어느 한 가닥을 절단하지 않고 이중나선화된 핵산 분자(예를 들어, gRNA 분자를 통해)에 결합할 수 있다. 일부 실시형태에서, 뉴클레아제-비활성 dCas9 도메인은 본 명세서에 제시된 아미노산 서열의 D10X 돌연변이 및 H840X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산 변화이다. 일부 실시형태에서, 뉴클레아제-비활성 dCas9 도메인은 본 명세서에 제시된 아미노산 서열의 D10A 돌연변이 및 H840A 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일 예로서, 뉴클레아제-비활성 Cas9 도메인은 클로닝 벡터 pPlatTET-gRNA2(수탁 번호 BAV54124)에 제시된 아미노산 서열을 포함한다.

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

[0354] (예를 들어, 이의 전체 내용이 본

명세서에 참조로 통합되는, 문헌[Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell*. 2013; 152(5):1173-83] 참조).

[0355] 추가의 적합한 뉴클레아제-비활성 dCas9 도메인은 본 개시 및 당해 분야의 지식에 기초하여 당업자에게 명백할 것이며, 본 개시의 범위 내에 있다. 이러한 추가의 예시적인 적합한 뉴클레아제-비활성 Cas9 도메인은 D10A/H840A, D10A/D839A/H840A, 및 D10A/D839A/H840A/N863A 돌연변이 도메인을 포함하지만, 이로만 제한되는 것은 아니다(예를 들어, 그 전문이 본 명세서에 참조로 통합되는, 문헌[Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838] 참조). 일부 실시형태에서, dCas9 도메인은 본 명세서에 제시된 dCas9 도메인 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일하다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 제시된 아미노산 서열 중 어느 하나와 비교하여 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 21개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개 또는 그 이상의 돌연변이를 갖는 아미노산 서열을 포함한다. 일부 실시형태에서, Cas9 도메인은 본 명세서에 제시된 아미노산 서열 중 어느 하나와 비교하여 적어도 10개, 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개, 적어도 50개, 적어도 60개, 적어도 70개, 적어도 80개, 적어도 90개, 적어도 100개, 적어도 150개, 적어도 200개, 적어도 250개, 적어도 300개, 적어도 350개, 적어도 400개, 적어도 500개, 적어도 600개, 적어도 700개, 적어도 800개, 900개, 적어도 1000개, 적어도 1100개, 또는 적어도 1200개의 동일한 연속 아미노산 잔기를 갖는 아미노산 서열을 포함한다.

[0356] 일부 실시형태에서, Cas9 도메인은 Cas9 닉카아제이다. Cas9 닉카아제는 이중나선화된 핵산 분자(예를 들어, 이중나선화된 DNA 분자)의 한 가닥만 절단할 수 있는 Cas9 단백질일 수 있다. 일부 실시형태에서, Cas9 닉카아제는 이중나선화된 핵산 분자의 표적 가닥을 절단하는데, 이는 Cas9 닉카아제가 Cas9에 결합된 gRNA(예를 들어, sgRNA)에 대해 (상보적인) 염기 쌍을 이루는 가닥을 절단함을 의미한다. 일부 실시형태에서, Cas9 닉카아제는 D10A 돌연변이를 포함하고 위치 840에 히스티딘을 갖는다. 일부 실시형태에서, Cas9 닉카아제는 이중나선화된 핵산 분자의 비-표적, 비-염기-편집 가닥을 절단하며, 이는 Cas9 닉카아제가 Cas9에 결합된 gRNA(예를 들어, sgRNA)에 염기 쌍이 아닌 가닥을 절단함을 의미한다. 일부 실시형태에서, Cas9 닉카아제는 H840A 돌연변이를 포함하고 위치 10에 아스파르트산 잔기 또는 상응하는 돌연변이를 갖는다. 일부 실시형태에서, Cas9 닉카아제는 본 명세서에 제공된 Cas9 닉카아제 중 어느 하나와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 적어도 99.5% 동일하다. 추가의 적합한 Cas9 닉카아제는 본 개시 및 해당 분야의 지식에 기초하여 당업자에게 명백할 것이며, 본 개시의 범위 내에 있다.

[0357] *배타성(exclusivity)이 감소된 Cas9 도메인*

[0358] 전형적으로, S. 피오게네스(*S. pyogenes*)로부터의 Cas9(spCas9)와 같은, Cas9 단백질은 특정 핵산 영역에 결합하기 위해 표준 NGG PAM 서열을 필요로 하며, 여기서 "NGG"의 "N"은 아데노신(A), 티미딘(T), 또는 시토신(C)이고, G는 구아노신이다. 이것은 게놈 내에서 원하는 염기를 편집하는 활성을 제한할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집 융합 단백질은 정확한 위치, 예를 들어, PAM의 상류에 있는 표적 염기를 포함하는 영역에 배치될 필요가 있을 수 있다. 예를 들어, 이의 전체 내용이 본 명세서에 참조로 통합되는, 문헌[Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016)] 참조. 따라서, 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 정규(예를 들어, NGG) PAM 서열을 함유하지 않는 뉴클레오티드 서열에 결합할 수 있는 Cas9 도메인을 함유할 수 있다. 비-정규 PAM 서열에 결합하는 Cas9 도메인은 당업계에 설명되어 있으며 숙련된 기술자에게 명백할 것이다. 예를 들어, 비-정규 PAM 서열에 결합하는 Cas9 도메인은, 각각의 전체 내용이 본 명세서에 참조로 통합되는, 다음 문헌에 설명되어 있다: 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). 여러 PAM 변형이 아래 표 1에 설명되어 있다:

[0359] **표 1. Cas9 단백질 및 해당 PAM 서열**

변이체	**PAM**
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN

[0360]

saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
Cpf1	5' (TTTV)

[0361]

[0362] 일부 실시형태에서, Cas9 도메인은 스타필로코커스 아우레우스(*Staphylococcus aureus*)(SaCas9)의 Cas9 도메인이다. 일부 실시형태에서, SaCas9 도메인은 뉴클레아제 활성 SaCas9, 뉴클레아제 비활성 SaCas9(SaCas9d), 또는 SaCas9 닉카아제(SaCas9n)이다. 일부 실시형태에서, SaCas9는 N579A 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다.

[0363] 일부 실시형태에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 비-정규 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 NNGRRT PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SaCas9 도메인은 하나 이상의 E781X, N967X, 및 R1014X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시형태에서, SaCas9 도메인은 하나 이상의 E781K, N967K, 및 R1014H 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 하나 이상의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, SaCas9 도메인은 E781K, N967K, 또는 R1014H 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다.

[0364] **예시적인 saCas9 서열**

[0365] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR

GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKT
SDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEW
YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIEN
VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENA
ELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE
LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK
KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIK
LHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKK
GNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFI
NRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIF
ITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDK
DNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTK
YSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVY
KFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYR
VIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLY
EVKSKKHPQIIKKG

[0366]

[0367] 밑줄이 그어져 있고 굵은 글씨체로 기재된, 상기 잔기 N579는 SaCas9 닉카아제를 생성하기 위해 돌연변이(예를 들어, A579로)될 수 있다.

[0368] **예시적인 SaCas9n 서열**

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR
GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKT
SDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEW
YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIEN
VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENA
ELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE
LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK
KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIK
LHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKK
GNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFI
NRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIF

[0369]

ITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDK
DNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTK
YSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVY
KFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYR
VIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLY
EVKSKKHPQIIKKG

[0370]

[0371] SaCas9 닉카아제를 생성하기 위해 N579로부터 돌연변이될 수 있는, 상기 잔기 A579는 밑줄이 그어져 있고 굵은 글씨체로 기재되어 있다.

[0372] 예시적인 SaKKH Cas9

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR
RRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR
GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKT
SDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEW
YEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIEN
VFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENA
ELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDE
LWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIK
KYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIK
LHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKK
GNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFI
NRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG
YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIF
ITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDK
DNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTK
YSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVY
KFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRV
IGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYE
VKSKKHPQIIKKG.

[0373]

[0374] SaCas9 닉카아제를 생성하기 위해 N579로부터 돌연변이될 수 있는, 상기 잔기 A579는 밑줄이 그어져 있고 굵은 글씨체로 기재되어 있다. SaKKH Cas9을 생성하기 위해 E781, N967, 및 R1014로부터 돌연변이될 수 있는, 상기 잔기 K781, K967, 및 H1014는 밑줄이 그어져 있고 이탤릭체로 기재되어 있다.

[0375] 일부 실시형태에서, Cas9 도메인은 스트렙토코커스 피오게네스(*Streptococcus pyogenes*)로부터의 Cas9(SpCas9) 도메인이다. 일부 실시형태에서, SpCas9 도메인은 뉴클레아제 활성 SpCas9, 뉴클레아제 비활성 SpCas9(SpCas9d), 또는 SpCas9 닉카아제(SpCas9n)이다. 일부 실시형태에서, SpCas9는 본 명세서에 제공된 임의의 아미노산 서열에서의 D9X 돌연변이, 또는 상응하는 돌연변이를 포함하며, 여기서 X는 D를 제외한 임의의 아미노산이다. 일부 실시형태에서, SpCas9는 D9A 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인, SpCas9d 도메인, 또는 SpCas9n 도메인은 비-정규 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SpCas9 도메인, SpCas9d 도메인, 또는 SpCas9n 도메인은 NGG, NGA, 또는 NGCG PAM 서열을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1134X, R1334X, 및 T1336X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1134E, R1334Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1134E, R1334Q, 및 T1337R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1134X, R1334X, 및 T1336X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1134V, R1334Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 아미노산 서열 중 임의의 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1134V, R1334Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1134X, G1217X, R1334X, 및 T1336X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1134V, G1217R, R1334Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1134V, G1217R, R1334Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다.

[0376] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 Cas9 폴리펩티드와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%,

적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일하다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 임의의 Cas9 폴리펩티드의 아미노산 서열을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 임의의 Cas9 폴리펩티드의 아미노산 서열로 구성된다.

[0377] 예시적인 SpCas9

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
[0378] PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0379] 예시적인 SpCas9n

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
[0380] EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH

PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0381]

[0382] 예시적인 SpEQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

[0383]

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**E**SPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

[0384]

[0385] 위의 서열에서, D1134, R1334, 및 T1336에서 돌연변이되어 SpEQR Cas9를 생성 할 수 있는, 상기 잔기 E1134, Q1334, 및 R1336은 밑줄이 그어져 있으며 굵은 글씨체로 표시되어 있다.

[0386] 예시적인 SpVQR Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
[0387] NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK

RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
[0388] PAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD

[0389] D1134, R1334, 및 T1336로부터 돌연변이되어 SpVQR Cas9를 생성할 수 있는, 상기 잔기 V1134, Q1334, 및 R1336은 밑줄이 그어져 있으며 굵은 글씨체로 표시되어 있다.

[0390] 예시적인 SpVRER Cas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASA**R**ELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ

[0391]

HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRK**E**Y**R**STKEVLDATLIHQSITGLYETRIDLSQLGGD.

[0392]

[0393] D1134, G1217, R1334, 및 T1336에서 돌연변이되어 SpVRER Cas9를 생성할 수 있는, 상기 잔기 V1134, R1217, E1334, 및 R1336은 밑줄이 그어져 있고 굵은 글씨체로 표시되어 있다. Cas9 뉴클레아제는 다음 2가지 기능적 엔도뉴클레아제 도메인을 가지고 있다: RuvC 및 HNH. Cas9는 표적 결합시 표적 DNA의 반대 가닥을 절단하도록 뉴클레아제 도메인을 위치시키는 두 번째 입체형태 변화를 겪는다. Cas9-매개 DNA 절단의 최종 결과는 표적 DNA 내의 이중 가닥 절단(DSB)이다(PAM 서열의 상류에서 약 3-4개 뉴클레오티드) 생성된 DSB는 다음 2가지 일반적인 복구 경로 중 하나에 의해 복구된다: (1) 효율적이지만 오류-빈번한 비상동성 말단 결합(NHEJ) 경로; 또는 (2) 덜 효율적이지만 고-충실도 상동성 직접 복구(HDR) 경로.

[0394] 비상동성 말단 결합(NHEJ) 및/또는 상동성 직접 복구(HDR)의 "효율"은 임의의 편리한 방법으로 계산할 수 있다. 예를 들어, 일부 경우에, 성공적인 HDR의 백분율의 관점에서 효율을 나타낼 수 있다. 예를 들어, 조사자(surveyor) 뉴클레아제 분석을 사용하여 절단 산물을 생성할 수 있으며 생성물 대 기질의 비율을 이용하여 백분율을 계산할 수 있다. 예를 들어, 성공적인 HDR의 결과로 새롭게 통합된 제한 서열을 포함하는 DNA를 직접 절단하는 조사자 뉴클레아제 효소를 사용할 수 있다. 더 많은 절단된 기질은 더 높은 HDR 백분율을 나타낸다(HDR의 효율이 더 높음). 예시적인 일예로서, HDR의 비율(fraction)(백분율)은 다음 방정식을 사용하여 계산할 수 있다: [(절단 생성물)/(기질 + 절단 생성물)](예를 들어, (b + c)/(a + b + c), 여기서 "a"는 DNA 기질의 밴드 강도이고 "b" 및 "c"는 절단 생성물이다).

[0395] 일부 경우에, 효율은 성공적인 NHEJ의 백분율의 관점에서 나타낼 수 있다. 예를 들어, T7 엔도뉴클레아제 I 분석을 사용하여 절단 생성물을 생성할 수 있으며 생성물 대 기질의 비율을 사용하여 NHEJ 백분율을 계산할 수 있다. T7 엔도뉴클레아제 I은 야생형 및 돌연변이 DNA 가닥의 혼성화로 인해 발생하는 불일치된 이종이중나선(heteroduplex) DNA를 절단한다(NHEJ는 본래의 절단 부위에서 작은 무작위 삽입 또는 결실(인델)을 생성한다). 더 많은 절단은 더 높은 NHEJ 백분율(NHEJ의 더 높은 효율)을 나타낸다. 예시적인 예로서, NHEJ의 비율(백분율)은 다음 방정식을 사용하여 계산할 수 있다: $(1-(1-(b+c)/(a+b+c))^{1/2}) \times 100$, 여기서 "a"는 DNA 기질의 밴드 강도이고, "b" 및 "c"는 절단 생성물이다(Ran *et al.*, 2013 Sep 12; 154(6):1380-9; 및 Ran *et al.*, Nat

Protoc 2013 Nov; 8(11): 2281-2308.

[0396] NHEJ 복구 경로는 가장 활동적인 복구 메커니즘이며, DSB 부위에 작은 뉴클레오티드 삽입 또는 결실(인델)을 자주 발생시킨다. NHEJ-매개 DSB 복구의 무작위성은, Cas9 및 gRNA 또는 가이드 폴리뉴클레오티드를 발현하는 세포 집단이 다양한 돌연변이 배열(arrays)을 초래할 수 있기 때문에, 중요한 실제적인 영향(implications)을 미친다. 대부분의 경우에서, NHEJ는 표적 DNA에서 작은 인델을 생성하여 아미노산 결실, 삽입, 또는 프레임 이동 돌연변이를 일으켜 표적 유전자의 오픈 리딩 프레임(ORF) 내에서 조기 정지 코돈을 야기한다. 이상적인 최종 결과는 표적 유전자 내의 기능소실(loss-of-function) 돌연변이이다.

[0397] NHEJ-매개 DSB 복구는 종종 유전자의 오픈 리딩 프레임을 파괴하지만, 상동성 직접 복구(HDR)는 단일 뉴클레오티드 변화에서 형광단 또는 태그 추가와 같은 거대 삽입에 이르는 특정 뉴클레오티드 변화를 생성하는 데 사용할 수 있다.

[0398] 유전자 편집을 위해 HDR을 활용하기 위해, 원하는 서열을 포함하는 DNA 복구 주형을 gRNA(들) 및 Cas9 또는 Cas9 닉카아제와 함께 관심 세포 유형 내로 전달할 수 있다. 복구 주형은 원하는 편집물(edit)뿐만 아니라 표적의 상류 및 하류에 바로 추가 상동 서열(왼쪽 & 오른쪽 상동성 암이라고 함)을 포함할 수 있다. 각 상동성 암의 길이는 도입되는 변경의 크기에 따라 달라질 수 있으며, 더 큰 삽입에는 더 긴 상동성 암을 필요로 한다. 복구 주형은 단일-가닥 올리고뉴클레오티드, 이중-가닥 올리고뉴클레오티드, 또는 이중-가닥 DNA 플라스미드일 수 있다. HDR의 효율은 일반적으로 Cas9, gRNA 및 외인성(exogenous) 복구 주형을 발현하는 세포에서 조차 낮다(변형된 대립유전자의 < 10%). HDR은 세포 주기의 S 및 G2 단계에서 발생하기 때문에, 세포를 동기화하여 HDR의 효율을 향상시킬 수 있다. NHEJ에 관여하는 유전자를 화학적으로 또는 유전적으로 억제하면 HDR 빈도를 또한 증가시킬 수 있다.

[0399] 일부 경우에, Cas9는 변형된 Cas9이다. 주어진 gRNA 표적화 서열은 부분 상동성이 존재하는 게놈 전체에 추가 부위를 가질 수 있다. 이러한 부위를 표적-이탈 부위(off-targets)라고 하며, gRNA를 설계할 때 고려해야 한다. gRNA 디자인을 최적화하는 것 외에도, Cas9에 대한 변경을 통해 CRISPR 특이성을 높일 수 있다. Cas9는 두개의 뉴클레아제 도메인, RuvC 및 HNH의 조합된 활성을 통해 이중 가닥 절단(DSB)을 생성한다. SpCas9의 D10A 돌연변이체인, Cas9 닉카아제는 하나의 뉴클레아제 도메인을 보유하고, DSB보다 DNA 닉을 생성한다. 닉카아제 시스템은 특정 유전자 편집을 위해 HDR-매개 유전자 편집과 조합될 수도 있다.

[0400] 일부 경우에, Cas9는 변이체 Cas9 단백질이다. 변이체 Cas9 폴리펩티드는 야생형 Cas9 단백질의 아미노산 서열과 비교할 때, 하나의 아미노산이 다른(예를 들어, 결실, 삽입, 치환, 융합을 갖는) 아미노산 서열을 갖는다. 일부 경우에, 변이체 Cas9 폴리펩티드는 Cas9 폴리펩티드의 뉴클레아제 활성을 감소시키는 아미노산 변화(예를 들어, 결실, 삽입 또는 치환)를 갖는다. 예를 들어, 일부 경우에, 변이체 Cas9 폴리펩티드는 상응하는 야생형 Cas9 단백질 활성의 50% 미만, 40% 미만, 30% 미만, 20% 미만, 10% 미만, 5% 미만, 또는 1% 미만의 뉴클레아제를 가진다. 일부 경우에, 변이체 Cas9 단백질은 실질적인 뉴클레아제 활성이 없다. 대상 Cas9 단백질이 실질적인 뉴클레아제 활성을 갖지 않는 변이체 Cas9 단백질인 경우, "dCas9"로 지칭될 수 있다.

[0401] 일부 경우에, 변이체 Cas9 단백질은 감소된 뉴클레아제 활성을 갖는다. 예를 들어, 변이체 Cas9 단백질은 야생형 Cas9 단백질, 예를 들어, 야생형 Cas9 단백질의 엔도뉴클레아제 활성의 약 20% 미만, 약 15% 미만, 약 10% 미만, 약 5% 미만, 약 1% 미만, 또는 약 0.1% 미만을 나타낸다.

[0402] 일부 경우에, 변이체 Cas9 단백질은 가이드 표적 서열의 상보적 가닥을 절단할 수 있지만 이중 가닥 가이드 표적 서열의 비-상보적 가닥을 절단하는 활성이 감소된다. 예를 들어, 변이체 Cas9 단백질은 RuvC 도메인의 기능을 감소시키는 돌연변이(아미노산 치환)를 가질 수 있다. 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 D10A(아미노산 위치 10에 아스파르테이트에서 알라닌)를 갖고, 그리하여 이중 가닥 가이드 표적 서열의 상보적 가닥을 절단할 수 있지만, 이중 가닥 가이드 표적 서열의 비-상보적 가닥을 절단하는 활성이 감소된다(따라서 변이체 Cas9 단백질은 이중 가닥 표적 핵산을 절단할 때, 이중 가닥 절단(DSB) 대신 단일 가닥 절단(SSB)을 초래한다)(예를 들어, 문헌[Jinek *et al.*, Science 2012 Aug 17; 337(6096):816-21] 참조).

[0403] 일부 경우에, 변이체 Cas9 단백질은 이중 가닥 가이드 표적 서열의 비-상보적 가닥을 절단할 수 있지만, 가이드 표적 서열의 상보적 가닥을 절단하는 활성이 감소된다. 예를 들어, 변이체 Cas9 단백질은 HNH 도메인(RuvC/HNH/RuvC 도메인 모티프)의 기능을 감소시키는 돌연변이(아미노산 치환)를 가질 수 있다. 비제한적인 예로서, 일부 경우에, 변이체 Cas9 단백질은 H840A(아미노산 위치 840에서 히스티딘에서 알라닌으로) 돌연변이를 갖고, 따라서 가이드 표적 서열의 비-상보성 가닥을 절단할 수 있지만, 가이드 표적 서열의 상보적 가닥을 절단

하는 활성이 감소된다(따라서 변이체 Cas9 단백질은 이중 가닥 가이드 표적 서열을 절단할 때 DSB 대신 SSB를 초래한다). 이러한 Cas9 단백질은 가이드 표적 서열(예를 들어, 단일 가닥 가이드 표적 서열)을 절단하는 감소된 활성을 갖지만, 가이드 표적 서열(예를 들어, 단일 가닥 가이드 표적 서열)에 결합하는 활성을 보유한다.

[0404] 일부 경우에, 변이체 Cas9 단백질은 이중 가닥 표적 DNA의 상보적 및 비-상보적 가닥 둘 다를 절단하는 활성이 감소된다. 비제한적인 예로서, 일부 실시형태에서, 변이체 Cas9 단백질은 D10A 및 H840A 돌연변이 둘 다를 보유하여 폴리펩티드가 이중 가닥 표적 DNA의 상보적 및 비-상보적 가닥 둘 다를 절단하는 활성이 감소된다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다.

[0405] 또 다른 비제한적 예로서, 일부 경우에, 변이체 Cas9 단백질은 W476A 및 W1126A 돌연변이를 보유하여 폴리펩티드가 표적 DNA를 절단하는 활성이 감소된다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다.

[0406] 또 다른 비제한적 예로서, 일부 경우에, 폴리펩티드가 표적 DNA를 절단하는 활성이 감소되도록 변이체 Cas9 단백질은 P475A, W476A, N477A, D1125A, W1126A, 및 D1127A 돌연변이를 보유한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다.

[0407] 또 다른 비제한적 예로서, 일부 경우에, 변이체 Cas9 단백질은 폴리펩티드가 표적 DNA를 절단하는 활성이 감소되도록 H840A, W476A, 및 W1126A, 돌연변이를 내포한다. 이러한 Cas9 단백질은 표적 DN (예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다. 또 다른 비제한적 예로서, 일부 경우에, 변이체 Cas9 단백질은 폴리펩티드가 표적 DNA를 절단하는 활성이 감소되도록 H840A, D10A, W476A, 및 W1126A, 돌연변이를 내포한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다. 일부 실시형태에서, 변이체 Cas9는 Cas9 HNH 도메인의 위치 840에 회복된 촉매 His 잔기(A840H)를 갖는다.

[0408] 또 다른 비제한적 예로서, 일부 경우에, 변이체 Cas9 단백질은 폴리펩티드가 표적 DNA를 절단하는 활성이 감소되도록 H840A, P475A, W476A, N477A, D1125A, W1126A, 및 D1127A 돌연변이를 내포한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다. 또 다른 비제한적 예로서, 일부 경우에, 변이체 Cas9 단백질은 폴리펩티드가 표적 DNA를 절단하는 활성이 감소되도록 D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, 및 D1127A 돌연변이를 내포한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다. 일부 경우에, 변이체 Cas9 단백질이 W476A 및 W1126A 돌연변이를 내포하거나, 또는 변이체 Cas9 단백질이 P475A, W476A, N477A, D1125A, W1126A, 및 D1127A 돌연변이를 내포할 때, 변이체 Cas9 단백질은 PAM 서열에 효율적으로 결합하지 않는다. 따라서, 이러한 일부 경우에, 이러한 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 이 방법은 PAM 서열을 필요로 하지 않는다. 달리 말해서, 일부 경우에, 이러한 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 이 방법은 가이드 RNA를 포함할 수 있지만, 방법은 PAM 서열의 부재하에 수행될 수 있다(결합의 특이성은 그리하여 가이드 RNA의 표적화 세그먼트에 의해 제공된다). 상기 효과를 달성(즉, 하나 또는 다른 뉴클레아제 부분을 비활성화)하기 위해, 다른 잔기를 돌연변이시킬 수 있다. 비제한적인 예로서, 잔기 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, 및/또는 A987은 변경(즉, 치환)될 수 있다. 또한, 알라닌 치환 이외의 돌연변이가 적합하다.

[0409] 일부 실시형태에서, 감소된 촉매 활성을 갖는 변이체 Cas9 단백질(예를 들어, Cas9 단백질이 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, 및/또는 A987 돌연변이, 예를 들어, D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, 및/또는 D986A를 갖는 경우)은, 가이드 RNA와 상호작용할 수 있는 활성을 보유하는 한, (가이드 RNA에 의해 표적 DNA 서열로 여전히 가이드되기 때문에) 여전히 부위-특이적 방식으로 표적 DNA에 결합할 수 있다.

[0410] 일부 실시형태에서, 변이체 Cas 단백질은 spCas9, spCas9-VRQR, spCas9-VRER, xCas9(sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK, 또는 spCas9-LRVSQL일 수 있다.

[0411] S. 피오게네스(*S. pyogenes*) Cas9의 대안은 포유류 세포에서 절단 활성을 나타내는 Cpf1 패밀리의 RNA-가이드된

엔도뉴클레아제를 포함할 수 있다. 프레보텔라(*Prevotella*) 및 프란시셀라(*Francisella*) 1(CRISPR/Cpf1)의 CRISPR은 CRISPR/Cas9 시스템과 유사한 DNA 편집 기술이다. Cpf1은 클래스 II CRISPR/Cas 시스템의 RNA-가이드 엔도 뉴클레아제이다. 이 후천성(acquired) 면역 메커니즘은 프레보텔라(*Prevotella*) 및 프란시셀라(*Francisella*) 박테리아에서 발견된다. Cpf1 유전자는 CRISPR 유전자좌와 연관되어, 가이드 RNA를 이용하여 바이러스 DNA를 찾고 절단하는 엔도뉴클레아제를 코딩한다. Cpf1은 Cas9보다 작고 단순한 엔도뉴클레아제이기 때문에, CRISPR/Cas9 시스템의 일부 제한을 극복할 수 있다. Cas9 뉴클레아제와 달리, Cpf1-매개 DNA 절단의 결과는 짧은 3' 오버행(overhang)이 있는 이중 가닥 파손이다. Cpf1의 엇갈린 절단 패턴은 전형적인 제한 효소 클로닝과 유사한, 방향성 유전자 전달 가능성을 열어 두어, 유전자 편집의 효율성을 증가시킬 수 있다. 위에서 설명한 Cas9 변이체 및 오르쏘로그와 마찬가지로, Cpf1은, CRISPR에 의해 표적화될 수 있는 부위의 수를 SpCas9가 선호하는 NGG PAM 부위가 없는 AT-풍부 영역 또는 AT-풍부 게놈으로 확장할 수도 있다. Cpf1 유전자좌는 혼합된 알파/베타 도메인, RuvC-I 다음에 나선형 영역, RuvC-II 및 징크 핑거-유사 도메인을 포함한다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사한 RuvC-유사 엔도뉴클레아제 도메인을 갖는다. 또한, Cpf1에는 HNH 엔도뉴클레아제 도메인이 없고, Cpf1의 N-말단에는 Cas9의 알파-나선 인식 로브가 없다. Cpf1 CRISPR-Cas 도메인 아키텍처는 Cpf1이 기능적으로 독특함을 보여주며, 이는 클래스 2, 타입 V CRISPR 시스템으로 분류되는 것임을 보여준다. Cpf1 유전자좌는 유형 II 시스템에서보다 유형 I 및 III와 더 유사한 Cas1, Cas2 및 Cas4 단백질을 코딩한다. 기능성 Cpf1은 트랜스-활성화 CRISPR RNA(tracrRNA)를 필요로 하지 않으며, 그리하여, CRISPR(crRNA)만 필요로 한다. 이는 Cpf1이 Cas9보다 작을 뿐만 아니라, 더 적은 sgRNA 분자(Cas9의 뉴클레오티드 수의 대략 절반)를 가지기 때문에, 게놈 편집에 도움이 된다. Cpf1-crRNA 복합체는 Cas9에 의해 표적화된 G-풍부 PAM과 대조적으로 모티프 5'-YTN-3'에 인접한 프로토스페이서의 식별에 의해 표적 DNA 또는 RNA를 절단한다. PAM의 식별 후, Cpf1은 4개 또는 5개의 뉴클레오티드 오버행의 점착-말단-유사(sticky-단부-like) DNA 이중 가닥 절단을 도입한다.

[0412] *2개의 napDNAbp, 데아미나제 도메인을 포함하는 융합 단백질*

[0413] 본 개시의 일부 양상은 닉카아제 활성을 갖는 napDNAbp 도메인(예를 들어, nCas 도메인) 및 촉매적으로 불활성인 napDNAbp(예를 들어, dCas 도메인) 및 핵염기 편집기(예를 들어, 아데노신 데아미나제 도메인, 시티딘 데아미나제 도메인)를 포함하는 융합 단백질을 제공하며, 여기서 적어도 napDNAbp 도메인은 링커에 의해 연결된다. Cas 도메인은 본 명세서에 제공된 Cas 도메인 또는 Cas 단백질(예를 들어, dCas9 및 nCas9) 중 임의의 것일 수 있음을 이해해야 한다. 일부 실시형태에서, Cas 도메인, DNA 결합 단백질 도메인, 또는 Cas 단백질 중 임의의 것은, 제한없이, Cas9(예를 들어, dCas9 및 nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 및 Cas12i를 포함한다. Cas9와 다른 PAM 특이성을 갖는 프로그래밍가능한 폴리뉴클레오티드 결합 단백질의 한 예는 프레보텔라(*Prevotella*) 및 프란시셀라(*Francisella*)1(Cpf1)의 클러스터링된 규칙적으로 간격을 두고 반복되는 짧은 회문 반복부이다. Cas9와 유사하게 Cpf1도 클래스 2 CRISPR 이펙터이다. 예를 들어, 제한 없이, 일부 실시형태에서, 융합 단백질은 다음 구조를 포함하며, 여기서 데아미나제는 아데노신 데아미나제 또는 시티딘 데아미나제이다:

[0414] NH_2-[데아미나제]-[nCas 도메인]-[dCas 도메인]-COOH;

[0415] NH_2-[데아미나제]-[dCas 도메인]-[nCas 도메인]-COOH;

[0416] NH_2-[nCas 도메인]-[dCas 도메인]-[데아미나제]-COOH;

[0417] NH_2-[dCas 도메인]-[nCas 도메인]-[데아미나제]-COOH;

[0418] NH_2-[nCas 도메인]-[데아미나제]-[dCas 도메인]-COOH;

[0419] NH_2-[dCas 도메인]-[데아미나제]-[nCas 도메인]-COOH;

[0420] 일부 실시형태에서, 상기 일반 아키텍처에서 사용된 "-"는 선택적 링커의 존재를 나타낸다. 일부 실시형태에서, 데아미나제 및 napDNAbp(예를 들어, Cas 도메인)는 링커 서열에 의해 연결되지 않고 직접 융합된다. 일부 실시형태에서, 링커는 데아미나제 도메인과 napDNAbp 사이에 존재한다. 일부 실시형태에서, 데아미나제 또는 다른 핵염기 편집기는 dCas에 직접 융합되고, 링커는 dCas 및 nCas9를 연결한다. 일부 실시형태에서, 데아미나제 및 napDNAbp는 본 명세서에 제공된 선택적 링커를 통해 융합된다. 예를 들어, 일부 실시형태에서, 데아미나제 및 napDNAbp는 "링커"라는 제목의 섹션에서 하기 제공된 선택적 링커를 통해 융합된다. 일부 실시형태에서, dCas 도메인 및 데아미나제는 바로 인접하고 nCas 도메인은 링커를 통해 이들 도메인(5' 또는 3')에 연결된다.

[0421] **내부 삽입이 있는 융합 단백질**

[0422] 본 개시는 핵산 프로그래밍가능한 핵산 결합 단백질, 예를 들어, napDNAbp에 융합된 이종성 폴리펩티드를 포함하는 융합 단백질을 제공한다. 이종성 폴리펩티드는 천연 또는 야생형 napDNAbp 폴리펩티드 서열에서 발견되지 않는 폴리펩티드일 수 있다. 이종성 폴리펩티드는 napDNAbp의 C-말단 단부(terminal end), napDNAbp의 N-말단 단부에서 napDNAbp에 융합되거나, 또는 napDNAbp의 내부 위치에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩티드는 napDNAbp의 내부 위치에 삽입된다.

[0423] 일부 실시형태에서, 이종성 폴리펩티드는 데아미나제 또는 이의 기능적 단편이다. 예를 들어, 융합 단백질은 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함할 수 있다. 융합 단백질 중의 데아미나제는 시티딘 데아미나제일 수 있다. 융합 단백질 중의 데아미나제는 아데노신 데아미나제일 수 있다.

[0424] 데아미나제는 원형 치환체 데아미나제일 수 있다. 예를 들어, 데아미나제는 원형 치환체 아데노신 데아미나제 또는 원형 치환체 시티딘 데아미나제일 수 있다. 일부 실시형태에서, 데아미나제는 TadA 참조 서열에서 번호매겨진 것과 같은 아미노산 잔기 116에서 원형 치환된, 원형 치환체 TadA이다. 일부 실시형태에서, 데아미나제는 TadA 참조 서열에서 번호매겨진 것과 같은 아미노산 잔기 136에서 원형 치환된, 원형 치환체 TadA이다. 일부 실시형태에서, 데아미나제는 TadA 참조 서열에서 번호매겨진 것과 같은 아미노산 잔기 65에서 원형 치환된, 원형 치환체 TadA이다.

[0425] 융합 단백질은 하나 이상의 데아미나제를 포함할 수 있다. 융합 단백질은, 예를 들어, 1개, 2개, 3개, 4개, 5개 또는 그 이상의 데아미나제를 포함할 수 있다. 일부 실시형태에서, 융합 단백질은 하나의 데아미나제를 포함한다. 일부 실시형태에서, 융합 단백질은 2개의 데아미나제를 포함한다. 융합 단백질에서 2개 이상의 데아미나제는 아데노신 데아미나제, 시티딘 데아미나제, 또는 이들의 조합일 수 있다. 2개 이상의 데아미나제는 동종이량체일 수 있다. 2개 이상의 데아미나제는 이종이량체일 수 있다. 2개 이상의 데아미나제는 napDNAbp에 나란히(in tandem) 삽입될 수 있다. 일부 실시형태에서, 2개 이상의 데아미나제는 napDNAbp에서 나란히 있지 않을 수 있다.

[0426] 일부 실시형태에서, 융합 단백질 중의 napDNAbp는 Cas9 폴리펩티드 또는 이의 단편이다. Cas9 폴리펩티드는 변이체 Cas9 폴리펩티드일 수 있다. 일부 실시형태에서, Cas9 폴리펩티드는 Cas9 닉카아제(nCas9) 폴리펩티드 또는 이의 단편이다. 일부 실시형태에서, Cas9 폴리펩티드는 뉴클레아제 멸실된 Cas9(dCas9) 폴리펩티드 또는 이의 단편이다. 융합 단백질의 Cas9 폴리펩티드는 전체 길이 Cas9 폴리펩티드일 수 있다. 일부 경우에, 융합 단백질의 Cas9 폴리펩티드는 전체 길이 Cas9 폴리펩티드가 아닐 수 있다. Cas9 폴리펩티드는, 예를 들어, 자연적으로 발생하는 Cas9 단백질에 비해 N-말단 또는 C-단부 말단에서 절단될 수 있다. Cas9 폴리펩티드는 원형으로 치환된 Cas9 단백질일 수 있다. Cas9 폴리펩티드는 표적 폴리뉴클레오티드 및 가이드 핵산 서열에 여전히 결합할 수 있는, Cas9 폴리펩티드의 단편, 부분 또는 도메인일 수 있다.

[0427] 일부 실시형태에서, Cas9 폴리펩티드는 스트렙토코커스 피오게네스(*Streptococcus pyogenes*) Cas9(SpCas9), 스타필로코커스 아우레우스(*Staphylococcus aureus*) Cas9(SaCas9), 스트렙토코커스 써모필러스(*Streptococcus thermophilus*) 1 Cas9(St1Cas9), 또는 이의 단편 또는 변이체이다.

[0428] 융합 단백질의 Cas9 폴리펩티드는 자연적으로 발생하는 Cas9 폴리펩티드와 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함할 수 있다.

[0429] 융합 단백질의 Cas9 폴리펩티드는 서열번호: 1로 제시된 Cas9 아미노산 서열과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95% 이상, 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함할 수 있다.

[0430] 이종성 폴리펩티드(예를 들어, 데아미나제)는, 예를 들어, napDNAbp가 표적 폴리뉴클레오티드 및 가이드 핵산에 결합하는 활성을 유지하도록 적절한 위치에서 napDNAbp(예를 들어, Cas9) 내에 삽입될 수 있다. 데아미나제는 데아미나제의 기능(예를 들어, 염기 편집 활성) 또는 napDNAbp(예를 들어, 표적 핵산 및 가이드 핵산에 결합하는 활성)를 손상시키지 않고 napDNAbp 내로 삽입될 수 있다. 데아미나제는, 예를 들어, 비정돈된(disordered) 영역 또는 결정학적 연구에 의해 나타난 바와 같이 고온 인자 또는 B-인자를 포함하는 영역에서 napDNAbp 내에 삽입될 수 있다. 덜 정돈되거나, 비정돈되거나, 또는 구조화되지 않은 단백질 영역, 예를 들어, 용매에 노출된 영역 및 루프가 구조 또는 기능을 손상시키지 않고 삽입에 사용될 수 있다. 데아미나제는 가요성 루프 영역 또는 용매 노출된 영역 내의 napDNAbp에 삽입될 수 있다. 일부 실시형태에서, 데아미나제는 Cas9 폴리펩티드 중의

가요성 루프 내에 삽입된다.

[0431] 일부 실시형태에서, 데아미나제의 삽입 위치는 Cas9 폴리펩티드의 결정 구조의 B-인자 분석에 의해 결정된다. 일부 실시형태에서, 데아미나제는 평균에 비해 더 많은 B-인자(예를 들어, 전체 단백질 또는 비정돈된 영역을 포함하는 단백질 도메인에 비해 더 높은 B 인자)를 포함하는 Cas9 폴리펩티드의 영역 내에 삽입된다. B-인자 또는 온도 인자는 이들의 평균 위치에서 (예를 들어, 온도-의존성 원자 진동 또는 결정 격자의 정적 비정돈의 결과로) 원자의 변동을 나타낼 수 있다. 백본 원자에 대한 높은 B-인자(예를 들어, 평균에 비해 더 많은 B-인자)는 상대적으로 높은 국부 이동성을 갖는 영역을 나타낼 수 있다. 이러한 영역은 구조나 기능을 손상시키지 않으면서 데아미나제를 삽입하는 데 사용될 수 있다. 데아미나제는 총 단백질에 대한 평균 B-인자에 비해 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, 또는 200%를 초과하여 더 많은 B-인자가 있는 Cα 원자를 갖는 잔기가 있는 위치에 삽입될 수 있다. 데아미나제는 잔기를 포함하는 Cas9 단백질 도메인에 대한 평균 B-인자에 비해 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, 또는 200%를 초과하여 더 많은 B-인자가 있는 Cα 원자를 가진 잔기가 있는 위치에 삽입될 수 있다. 평균에 비해 더 많은 B-인자를 포함하는 Cas9 폴리펩티드 위치는, 예를 들어, 서열번호: 1에서 번호매겨진 것과 같은 768, 792, 1052, 1015, 1022, 1026, 1029, 1067, 1040, 1054, 1068, 1246, 1247, 1248번 잔기를 포함할 수 있다. 평균보다 높은 B-인자를 포함하는 Cas9 폴리펩티드 영역은, 예를 들어, 서열번호: 1에서 번호매겨진 것과 같은 잔기 792-872, 792-906, 및 2-791을 포함할 수 있다.

[0432] 이종성 폴리펩티드(예를 들어, 데아미나제)는 서열번호: 1에서 번호매겨진 것과 같은 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, 및 1248로 이루어진 군에서 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에서 napDNAbp 내에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, 또는 1248-1249 또는 이의 상응하는 아미노산 위치 사이에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 769-770, 792-793, 793-794, 1016-1017, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1053-1054, 1055-1056, 1068-1069, 1069-1070, 1248-1249, 또는 1249-1250 또는 이의 상응하는 아미노산 위치 사이에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, 및 1248로 이루어진 군에서 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체한다. 삽입 위치와 관련하여 서열번호: 1에 대한 참조는 예시를 위한 것임을 이해해야 한다. 본 명세서에 논의된 바와 같은 삽입은 서열번호: 1의 Cas9 폴리펩티드 서열에 제한되지 않지만, 변이체 Cas9 폴리펩티드, 예를 들어, Cas9 닉카아제(nCas9), 뉴클레아제 멸실된 Cas9(dCas9), 뉴클레아제 도메인이 결여된 Cas9 변이체, 절단된 Cas9, 또는 HNH 도메인이 일부 또는 완전히 결여된 Cas9 도메인에서의 상응하는 위치들에서의 삽입을 포함한다.

[0433] 이종성 폴리펩티드(예를 들어, 데아미나제)는 서열번호: 1에서 번호매겨진 것과 같은 768, 792, 1022, 1026, 1040, 1068, 및 1247로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에서 napDNAbp 내에 삽입될 수 있다. 일부 실시형태에서, 이종성 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 768-769, 792-793, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1068-1069, 또는 1247-1248 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 실시형태에서, 이종성 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 아미노산 위치 769-770, 793-794, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1069-1070, 또는 1248-1249 또는 이의 상응하는 아미노산 위치 사이에 삽입된다. 일부 실시형태에서, 이종성 폴리펩티드는 서열번호: 1에서 번호매겨진 것과 같은 768, 792, 1022, 1026, 1040, 1068, 및 1247로 이루어진 군으로부터 선택된 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체한다.

[0434] 이종성 폴리펩티드(예를 들어, 데아미나제)는 도 4, 도 5, 도 6 또는 도 7에 나타낸 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드의 상응하는 아미노산 잔기에서 napDNAbp 내에 삽입될 수 있다. 이종성 폴리펩티드(예를 들어, 데아미나제)는 서열번호: 1에서 번호매겨진 것과 같은 1002, 1003, 1025, 1052-1056, 1242-1247, 1061-1077, 943-947, 686-691, 569-578, 530-539, 및 1060-1077로 이루어진 군으로부터 선택된 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드의 상응하는 아미노산 잔기에서 napDNAbp 내에 삽입될 수 있다. 데아미나제는 상기 잔기의 N-말단 또는 C-말단에 삽입되거나, 또는 잔기를 대체할 수 있다. 일부 실시형태에서, 데아미나제는 상기 잔기의 C-말단에 삽입된다.

[0435] 일부 실시형태에서, ABE(예를 들어, TadA)는 서열번호: 1에서 번호매겨진 것과 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246으로 이루어진 군으로부터 선택된 아미노산 잔기, 또는 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE(예를 들어, TadA)는 서열번호: 1에서 번호매겨진 것과 같은 잔기 792-872, 792-906, 또는 2-791, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하여 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246으로 이루어진 군으로부터 선택된 아미노산, 또는 또 다른 Cas9 폴리펩티드의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246로 이루어진 군으로부터 선택된 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246으로 이루어진 군으로부터 선택된 아미노산, 또는 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0436] 일부 실시형태에서, CBE(예를 들어, APOBEC1)는 서열번호: 1에서 번호매겨진 것과 같은 1016, 1023, 1029, 1040, 1069, 및 1247로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1016, 1023, 1029, 1040, 1069, 및 1247로 이루어진 군으로부터 선택된 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1016, 1023, 1029, 1040, 1069, 및 1247로 이루어진 군으로부터 선택된 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1016, 1023, 1029, 1040, 1069, 및 1247로 이루어진 군으로부터 선택된 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0437] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 768번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 768번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 768번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 768번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0438] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 791번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 791번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 791번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 791번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0439] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 792번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 792번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 792번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 792번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0440] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1016번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1016번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1016번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1016번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도

록 삽입된다.

[0441] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1022번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1022번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1022번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1022번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0442] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1023번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1023번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1023번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1023번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0443] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1026번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1026번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1026번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1026번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0444] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1029번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1029번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1029번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1029번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0445] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1040번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1040번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1040번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1040번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0446] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1052번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1052번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1052번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1052번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0447] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1054번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1054번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1054번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번

호매겨진 것과 같은 1054번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0448] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1067번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1067번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1067번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1067번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0449] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1068번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1068번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1068번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1068번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0450] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1069번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1069번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1069번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1069번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0451] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1246번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1246번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1246번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1246번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0452] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1247번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1247번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1247번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1247번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0453] 일부 실시형태에서, 데아미나제는 서열번호: 1에서 번호매겨진 것과 같은 1248번 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1248번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 N-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1248번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기의 C-말단에 삽입된다. 일부 실시형태에서, ABE는 서열번호: 1에서 번호매겨진 것과 같은 1248번 아미노산, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기를 대체하도록 삽입된다.

[0454] 일부 실시형태에서, 이종성 폴리펩티드(예를 들어, 데아미나제)는 Cas9 폴리펩티드의 가요성 루프에 삽입된다. 가요성 루프 부분은 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-570, 686-691, 943-947, 1002-1025, 1052-1077, 1232-1247, 또는 1298-1300으로 이루어진 군에서, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에서 선택될 수 있다. 가요성 루프 부분은 서열번호: 1에서 번호매겨진 것과 같은 1-529, 538-

568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 또는 1248-1297로 이루어진 군에서, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에서 선택될 수 있다.

[0455] 이종성 폴리펩티드(예를 들어, 데아미나제)는 서열번호: 1에서 번호매겨진 것과 같은 1017-1069, 1242-1247, 1052-1056, 1060-1077, 1002-1003, 943-947, 530-537, 568-579, 686-691, 1242-1247, 1298-1300, 1066-1077, 1052-1056, 또는 1060-1077 아미노산 잔기, 또는 또는 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 상응하는 Cas9 폴리펩티드 영역 내로 삽입될 수 있다.

[0456] 이종성 폴리펩티드(예를 들어, 데아미나제)는 Cas9 폴리펩티드의 결실된 영역을 대체하여 삽입될 수 있다. 결실된 영역은 Cas9 폴리펩티드의 N-말단 또는 C-말단 부분에 상응할 수 있다. 일부 실시형태에서, 결실된 영역은 서열번호: 1에서 번호매겨진 것과 같은 잔기 792-872, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 상응한다. 일부 실시형태에서, 결실된 영역은 서열번호: 1에서 번호매겨진 것과 같은 잔기 792-906, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 상응한다. 일부 실시형태에서, 결실된 영역은 서열번호: 1에서 번호매겨진 것과 같은 잔기 2-791, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 상응한다. 일부 실시형태에서, 결실된 영역은 서열번호: 1에서 번호매겨진 것과 같은 잔기 1017-1069, 또는 이의 상응하는 아미노산 잔기에 상응한다.

[0457] 이종성 폴리펩티드(예를 들어, 데아미나제)는 Cas9 폴리펩티드의 구조적 또는 기능적 도메인 내에 삽입될 수 있다. 이종성 폴리펩티드(예를 들어, 데아미나제)는 Cas9 폴리펩티드의 2개의 구조적 또는 기능적 도메인 사이에 삽입될 수 있다. 이종성 폴리펩티드(예를 들어, 데아미나제)는, 예를 들어, Cas9 폴리펩티드로부터 도메인을 결실시킨 후 Cas9 폴리펩티드의 구조적 또는 기능적 도메인을 대체하여 삽입될 수 있다. Cas9 폴리펩티드의 구조적 또는 기능적 도메인은, 예를 들어, RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, 또는 HNH를 포함할 수 있다.

[0458] 일부 실시형태에서, Cas9 폴리펩티드는 RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, 또는 HNH 도메인으로 이루어진 군으로부터 선택된 하나 이상의 도메인이 결여된다. 일부 실시형태에서, Cas9 폴리펩티드는 뉴클레아제 도메인이 결여된다. 일부 실시형태에서, Cas9 폴리펩티드는 HNH 도메인이 결여된다. 일부 실시형태에서, Cas9 폴리펩티드는 Cas9 폴리펩티드는 감소되거나 폐기된 HNH 활성을 갖도록 HNH 도메인의 일부가 결여된다.

[0459] 일부 실시형태에서, Cas9 폴리펩티드는 뉴클레아제 도메인의 결실을 포함하고, 데아미나제는 뉴클레아제 도메인을 대체하도록 삽입된다. 일부 실시형태에서, HNH 도메인이 결실되고 데아미나제가 그 자리에 삽입된다. 일부 실시형태에서, 하나 이상의 RuvC 도메인이 결실되고 데아미나제가 그 자리에 삽입된다.

[0460] 이종성 폴리펩티드를 포함하는 융합 단백질은 napDNAbp의 N-말단 및 C-말단 단편에 의해 측접될 수 있다. 일부 실시형태에서, 융합 단백질은 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함한다. N 말단 단편 또는 C 말단 단편은 표적 폴리뉴클레오티드 서열에 결합할 수 있다. N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단은 Cas9 폴리펩티드의 가요성 루프의 일부를 포함할 수 있다. N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단은 Cas9 폴리펩티드의 알파-나선 구조의 일부를 포함할 수 있다. N-말단 단편 또는 C-말단 단편은 DNA 결합 도메인을 포함할 수 있다. N-말단 단편 또는 C-말단 단편은 RuvC 도메인을 포함할 수 있다. N-말단 단편 또는 C-말단 단편은 HNH 도메인을 포함할 수 있다. 일부 실시형태에서, N-말단 단편 및 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는다.

[0461] 일부 실시형태에서, N 말단 Cas9 단편의 C-말단은 융합 단백질이 표적 핵염기를 탈아미노화시킬 때 표적 핵염기에 근접한 아미노산을 포함한다. 일부 실시형태에서, C 말단 Cas9 단편의 N-말단은 융합 단백질이 표적 핵염기를 탈아미노화시킬 때 표적 핵염기에 근접한 아미노산을 포함한다. 상이한 데아미나제의 삽입 위치는 표적 핵염기와 N 말단 Cas9 단편의 C-말단 또는 C 말단 Cas9 단편의 N-말단의 아미노산 사이에 근접성을 갖도록 하기 위해 상이할 수 있다. 예를 들어, ABE의 삽입 위치는 서열번호: 1에서 번호매겨진 것과 같은 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, 및 1246으로 이루어진 군으로부터 선택된 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 있을 수 있다. CBE의 적합한 삽입 위치는 서열번호: 1에서 번호매겨진 것과 같은 1016, 1023, 1029, 1040, 1069, 및 1247로 이루어진 군으로부터 선택되는 아미노산 잔기, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기일 수 있다. 특정 실시형태에서, ABE의 삽입은 상기 열거된 아미노산 잔기 중 어느 하나의 N 말단 또는 C 말단에 삽입될 수 있다. 일부 실시형태에서, ABE의 삽입은 상기 열거된 아미노산 잔기 중 임의의 하나를 대체하기 위해 삽입될 수 있다.

[0462] 융합 단백질의 N-말단 Cas9 단편(즉, 융합 단백질에서 데아미나제에 측접하는 N-말단 Cas9 단편)은 Cas9 폴리펩

티드의 N-말단을 포함할 수 있다. 융합 단백질의 N-말단 Cas9 단편은 적어도 약 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 또는 1300개 길이의 아미노산을 포함할 수 있다. 융합 단백질의 N-말단 Cas9 단편은 다음 아미노산 잔기에 상응하는 서열을 포함할 수 있다: 서열번호: 1에서 번호매겨진 것과 같은 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, 또는 1-1100, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기. N-말단 Cas9 단편은 서열번호: 1에서 번호매겨진 것과 같은 아미노산 잔기 1-56, 1-95, 1-200, 1-300, 1-400, 1 -500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, 또는 1-1100, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%의 서열 동일성을 갖는(comprising) 서열을 포함할 수 있다.

[0463] 융합 단백질의 C-말단 Cas9 단편(즉, 융합 단백질에서 데아미나제를 플랭킹하는 C-말단 Cas9 단편)은 Cas9 폴리펩티드의 C-말단을 포함할 수 있다. 융합 단백질의 C-말단 Cas9 단편은 적어도 약 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 또는 1300개 길이의 아미노산을 포함할 수 있다. 융합 단백질의 C-말단 Cas9 단편은 다음 아미노산 잔기에 상응하는 서열을 포함할 수 있다: 서열번호: 1에서 번호매겨진 것과 같은 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, 또는 56-1368, 또는 또 다른 Cas9 폴리펩티드의 상응하는 아미노산 잔기. N-말단 Cas9 단편은 서열번호: 1에서 번호매겨진 것과 같은 아미노산 잔기 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, 또는 56-1368, 또는 또 다른 Cas9 폴리펩티드에서의 상응하는 아미노산 잔기에 대해 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5%의 서열 동일성을 갖는(comprising) 서열을 포함할 수 있다. 함께 취해진 융합 단백질의 N-말단 Cas9 단편 및 C-말단 Cas9 단편은, 예를 들어, 서열번호: 1에 제시된 바와 같이, 전체 길이의 자연적으로 발생하는 Cas9 폴리펩티드 서열에 상응하지 않을 수 있다.

[0464] 본 명세서에 기술된 융합 단백질은, 감소된 게놈 전체의 불요성 탈아미노화와 같은, 비-표적 부위(예를 들어, 표적-이탈 부위)에서의 감소된 탈아미노화로 표적화된 탈아미노화를 수행할 있다. 본 명세서에 기술된 융합 단백질은 비-표적 부위에서 방관자 탈아미노화를 감소시키면서 표적화된 탈아미노화를 수행할 수 있다. 원하지 않는 탈아미노화 또는 표적-이탈 탈아미노화는, 예를 들어, Cas9 폴리펩티드의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 단부 말단 융합 단백질과 비교하여 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 또는 적어도 99% 만큼 감소될 수 있다. 원하지 않는 탈아미노화 또는 표적-이탈 탈아미노화는, 예를 들어, Cas9 폴리펩티드의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 단부 말단 융합 단백질과 비교하여, 적어도 1배, 적어도 2배, 적어도 3배, 적어도 4배, 적어도 5배, 적어도 10배, 적어도 15배, 적어도 20배, 적어도 30배, 적어도 40배, 적어도 50배, 적어도 60배, 적어도 70배, 적어도 80배, 적어도 90배, 또는 적어도 100배 만큼 감소될 수 있다.

[0465] 일부 실시형태에서, 융합 단백질의 데아미나제는 R-루프 범위 내에서 2개 이하의 핵염기를 탈아미노화시킨다. 일부 실시형태에서, 융합 단백질의 데아미나제는 R-루프 범위 내에서 3개 이하의 핵염기를 탈아미노화시킨다. 일부 실시형태에서, 융합 단백질의 데아미나제는 R-루프 범위 내에서 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 이하의 핵염기를 탈아미노화시킨다. R 루프는 DNA:RNA 하이브리드, DNA:DNA 또는 RNA:RNA 상보적 구조를 포함하는 3-가닥 핵산 구조이며 단일-가닥 DNA와 연관된다. 본 명세서에서 사용된, R-루프는 표적 폴리뉴클레오티드가 CRISPR 복합체 또는 염기 편집 복합체와 접촉될 때 형성될 수 있으며, 여기서 가이드 폴리뉴클레오티드, 예를 들어, 가이드 RNA의 일부는 표적 폴리뉴클레오티드, 예를 들어, 표적 DNA의 일부와 혼성화하고 이를 대체한다. 일부 실시형태에서, R-루프는 스페이서 서열 및 표적 DNA 상보적 서열의 혼성화된 영역을 포함한다. R-루프 영역은 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 또는 50개 길이의 핵염기 쌍일 수 있다. 일부 실시형태에서, R-루프 영역은 길이가 약 20개의 핵염기 쌍이다. 본 명세서에 사용된, R-루프 영역은 가이드 폴리뉴클레오티드와 혼성화하는 표적 DNA 가닥으로 제한되지 않는다는 것을 이해해야 한다. 예를 들어, R-루프 영역 내의 표적 핵염기의 편집은 가이드 RNA에 대한 상보적 가닥을 포함하는 DNA 가닥에 대한 것일 수 있거나, 가이드 RNA에 상보적인 가닥의 반대 가닥인 DNA 가닥에 대한 것일 수 있다. 일부 실시형태에서, R-루프 영역에서의 편집은 비-상보적 가닥(프로토스페이서 가닥) 상의 핵염기를 표적 DNA 서열의 가이드 RNA로 편집하는 것을 포함한다.

[0466] 본 명세서에 기술된 융합 단백질은 정규 염기 편집과는 상이한 편집 윈도우에서 표적 탈아미노화를 수행할 수 있다. 일부 실시형태에서, 표적 핵염기는 표적 폴리뉴클레오티드 서열에서 PAM 서열의 상류에 있는 약 1 내지

약 20개 염기이다. 일부 실시형태에서, 표적 핵염기는 표적 폴리뉴클레오티드 서열에서 PAM 서열의 상류에 있는 약 2개 내지 약 12개 염기이다. 일부 실시형태에서, 표적 핵염기는 PAM 서열로부터 약 1개 내지 9개 염기쌍, 약 2개 내지 10개 염기쌍, 약 3개 내지 11개 염기쌍, 약 4개 내지 12개 염기쌍, 약 5개 내지 13개 염기쌍, 약 6개 내지 14개 염기쌍, 약 7개 내지 15개의 염기쌍, 약 8개 내지 16개의 염기쌍, 약 9개 내지 17개의 염기쌍, 약 10개 내지 18개의 염기쌍, 약 11개 내지 19개의 염기쌍, 약 12개 내지 20개의 염기쌍, 약 1개 내지 7개의 염기쌍, 약 2개 내지 8개 염기쌍, 약 3개 내지 9개 염기쌍, 약 4개 내지 10개 염기쌍, 약 5개 내지 11개 염기쌍, 약 6개 내지 12개 염기쌍, 약 7개 내지 13개 염기쌍, 약 8개 내지 14개 염기쌍, 9개 내지 15개 염기쌍, 약 10개 내지 16개 염기쌍, 약 11개 내지 17개 염기쌍, 약 12개 내지 18개 염기쌍, 약 13개 내지 19개 염기쌍, 약 14개 내지 20개 염기쌍, 약 1개 내지 5개 염기쌍, 약 2개 내지 6개 염기쌍, 약 3개 내지 7개 염기쌍, 약 4개 내지 8개 염기쌍, 약 5개 내지 9개 염기쌍, 약 6개 내지 10개 염기쌍, 약 7개 내지 11개 염기쌍, 약 8개 내지 12개 염기쌍, 약 9개 내지 13개 염기쌍, 약 10 내지 14개 염기쌍, 약 11 내지 15개 염기쌍, 약 12개 내지 16개 염기쌍, 약 13개 내지 17개 염기쌍, 약 14개 내지 18개 염기쌍, 약 15개 내지 19개 염기쌍, 약 16개 내지 20개 염기쌍, 약 1개 내지 3개 염기쌍, 약 2개 내지 4개 염기 약 3개 내지 5개의 염기쌍, 약 4개 내지 6개 염기쌍, 약 5개 내지 7개의 염기쌍, 약 6개 내지 8개의 염기쌍, 약 7개 내지 9개의 염기쌍, 약 8개 내지 10개의 염기쌍, 약 9개 내지 11개의 염기쌍 , 약 10개 내지 12개의 염기쌍, 약 11개 내지 13개의 염기쌍, 약 12개 내지 14개의 염기쌍, 약 13개 내지 15개의 염기쌍, 약 14개 내지 16개의 염기쌍, 약 15개 내지 17개의 염기쌍, 약 16개 내지 18개의 염기쌍, 약 17개 내지 19개 염기쌍, 약 18개 내지 20개 염기쌍 만큼 떨어져 있거나 이의 상류에 존재한다. 일부 실시형태에서, 표적 핵염기는 PAM 서열로부터 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개, 또는 그 이상 더 많은 염기쌍 만큼 떨어져 있거나, 상류에 있다. 일부 실시형태에서, 표적 핵염기는 PAM 서열의 상류에 있는 약 1, 2, 3, 4, 5, 6, 7, 8 또는 9개 염기쌍이다. 일부 실시형태에서, 표적 핵염기는 PAM 서열의 상류에 있는 약 2, 3, 4 또는 6개 염기쌍이다.

[0467] 따라서, 정규 염기 편집기, 예를 들어, BE4와 비교하여 대안적인 바람직한 염기 편집 윈도우를 허용하는 염기 편집을 최적화하기 위해 융합 단백질 라이브러리 및 이를 사용하는 방법이 또한 본 명세서에 제공된다. 일부 실시형태에서, 본 개시는 복수의 융합 단백질을 포함하는 최적화된 염기 편집을 위한 단백질 라이브러리를 제공하며, 복수의 융합 단백질 각각은 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하고, 상기 융합 단백질 각각의 N-말단 단편은 상기 나머지 복수의 융합 단백질의 N-말단 단편과 상이하거나 상기 융합 단백질 각각의 C-말단 단편은 상기 나머지 복수의 융합 단백질의 C-말단과 상이하고, 융합 단백질 각각의 데아미나제는 표적 폴리뉴클레오티드 서열에서 프로토스페이서 인접 모티프(PAM) 서열에 근접한 표적 핵염기를 탈아미노화시키며, 상기 N 말단 단편 또는 상기 C 말단 단편은 표적 폴리뉴클레오티드 서열에 결합한다. 일부 실시형태에서, CRISPR R-루프 내의 각 핵염기에 대해, 복수의 융합 단백질 중 적어도 하나는 핵염기를 탈아미노화시킨다. 일부 실시형태에서, PAM 서열로부터 1 내지 20개 염기쌍 떨어져 있는 표적 폴리뉴클레오티드 중의 각 핵염기에 대해, 상기 복수의 융합 단백질 중 적어도 하나는 핵염기를 탈아미노화시킨다. 일부 실시형태에서, 최적화된 염기 편집을 가능케하는 융합 단백질 라이브러리를 포함하는 키트가 본 명세서에 제공된다.

[0468] 융합 단백질은 하나 이상의 이종성 폴리펩티드를 포함할 수 있다. 예를 들어, 융합 단백질은 하나 이상의 UGI 도메인 및/또는 하나 이상의 핵 위치결정 신호를 추가로 포함할 수 있다. 2개 이상의 이종성 도메인은 나란히 삽입될 수 있다. 2개 이상의 이종성 도메인은 NapDNAbp에서 나란히 있지 않게 되는 위치에 삽입될 수 있다.

[0469] 융합 단백질은 데아미나제와 napDNAbp 폴리펩티드 사이에 링커를 포함할 수 있다. 링커는 펩티드 또는 비펩티드 링커일 수 있다. 예를 들어, 링커는 XTEN, (GGGS)n, (GGGGS)n, (G)n, (EAAAAK)n, (GGS)n, SGSETPGTSESATPES일 수 있다. 일부 실시형태에서, 융합 단백질은 N-말단 Cas9 단편과 데아미나제 사이에 링커를 포함한다. 일부 실시형태에서, 융합 단백질은 C-말단 Cas9 단편과 데아미나제 사이에 링커를 포함한다. 일부 실시형태에서, napDNAbp의 N-말단 및 C-말단 단편은 링커로 데아미나제에 연결된다. 일부 실시형태에서, N-말단 및 C-말단 단편은 링커 없이 데아미나제 도메인에 연결된다. 일부 실시형태에서, 융합 단백질은 N-말단 Cas9 단편과 데아미나제 사이에 링커를 포함하지만, C-말단 Cas9 단편과 데아미나제 사이에 링커를 포함하지 않는다. 일부 실시형태에서, 융합 단백질은 C-말단 Cas9 단편과 데아미나제 사이에 링커를 포함하지만, N-말단 Cas9 단편과 데아미나제 사이에 링커를 포함하지 않는다.

[0470] 예시적인 TadA 또는 TadA7.10 서열은 아래에 제시되어 있다:

SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGR
VVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTD

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLY
VTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVE
ITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD

TAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVF
GVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQ
VFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP

YRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLH
YPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS

[0471] ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL

NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN

MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETP
GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR
VIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLY
VTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVE
ITEGILADECAALLCYFFRMPRQVFN

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLY
VTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVE
ITEGILADECAALLCYFFRMPRQ

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLCYFFRMPRQVFN

GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVG
AVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
[0472] ECAALLCYFFRMPRQVFNAQKKAQSSTD

[0473] 101 Cas9 TadAins 1015

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
[0474] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVGSSGSETPGTSESATPESSGSEVEFSHEYWMRHAL
TLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQG
GLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGS
LMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSST
DYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0475] LGGD

[0476] 102 Cas9 TadAins 1022

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIGSSGSETPGTSESATPESSGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
[0477] GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0478] LGGD

[0479] 103 Cas9 TadAins 1029

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGSSGSETPGTSESATPESSGS
EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRV
[0480] VFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMP
RQVFNAQKKAQSSTDGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0481] LGGD

[0482] 103 Cas9 TadAins 1040

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
[0483] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC
AALLCYFFRMPRQVFNAQKKAQSSTDNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0484] LGGD

[0485] 105 Cas9 TadAins 1068

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
[0486] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGEGSSGSETPGTSESATPESSGSEVEFSHEYWMR
HALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL
RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQ
SSTDTGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0487] LGGD

[0488] 106 Cas9 TadAins 1247

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
[0489] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGGSS
GSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVL
VLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTF
EPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITE
GILADECAALLCYFFRMPRQVFNAQKKAQSSTDSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0490] LGGD

[0491] 107 Cas9 TadAins 1054

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
[0492] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0493] LGGD

[0494] 108 Cas9 TadAins 1026

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
[0495] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEGSSGSETPGTSESATPESSGSEVE
FSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQV
FNAQKKAQSSTDQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0496] LGGD

[0497] 109 Cas9 TadAins 768

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMR
HALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL
RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREI
NNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQE
IGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP
KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL
PSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK
[0498] RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD
[0499] TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0500] 110.1 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA

[0501] VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV
TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI
TEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISEFSK
RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFD

[0502] TTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0503] 110.2 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
[0504] KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDAT
LYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHR
VEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
[0505] YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0506] 110.3 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
[0507] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
NHRVEITEGILADECAALLCYFFRMPREDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
[0508] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0509] 110.4 Cas9 TadAins 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
[0510] SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
NHRVEITEGILADECAALLCYFFRMRREDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
[0511] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0512] 110.5 Cas9 TadAins 1249

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
[0513] SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSGS
SGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCYFFRMRRPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
[0514] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0515] 110.5 Cas9 TadAins 델타 59-66 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
[0516] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSSGSETPGTSESATPESGSSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFE
PCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEG
ILADECAALLCYFFRMPRQVFNAQKKAQSSTDEDNEQKQLFVEQHKHYLD
EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN
LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
[0517] D

[0518] 110.6 Cas9 TadAins 1251

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
[0519] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
GSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMRRDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
[0520] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0521] 110.7 Cas9 TadAins 1252

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
[0522] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI

LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI

FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK

NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV

MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP

VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD

SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI

REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV

QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE

KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK

YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE

DGSSGSSGSETPGTSESATPESGSSSGSEVEFSHEYWMRHALTLAKRARD

EREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR

LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

GMNHRVEITEGILADECAALLCYFFRMRRNEQKQLFVEQHKHYLDEIIEQ

ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA

[0523] AFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0524] 110.8 Cas9 TadAins 델타 59-66 C-절단체 1250

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

[0525] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPG
SSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA
VLVLNNRVIGEGWNRAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC
AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMPRQEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0526]

[0527] 111.1 Cas9 TadAins 997

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGSSGSETPGTSESATPESSGIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM
ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE

[0528] LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI

IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG

[0529] APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0530] 111.2 Cas9 TadAins 997

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGSSGSSGSETPGTSESATPESSGGSSIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI
[0531] ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK

RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF
TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLS
[0532] QLGGD

[0533] 112 델타 HNH TadA

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND
KLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
[0534] IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKK
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL
[0535] IHQSITGLYETRIDLSQLGGD

[0536] 113 N-말단 단일 TadA 나선 절단된 165-단부

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG
RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR
MPRSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG
HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQL
SKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP
LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMT
RKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKL
INGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARE
NQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
[0537] QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK

SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDF
RKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVY
DVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS
AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL
DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG
[0538] GD

[0539] 114 N-말단 단일 TadA 나선 절단된 165-단부 델타 59-65

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRSGGS
SGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITD
EYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGD
LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD
DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK
RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
[0540] TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEH
IANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG
QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY
VDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE
EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA
KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK
KDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS
FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI
SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA
[0541] FKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0542] 115.1 Cas9 TadAins1004

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ

[0543] LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD

SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDA
TLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNH
RVEITEGILADECAALLCYFFRMPRQLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF

[0544] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0545] 115.2 Cas9 TadAins1005

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
[0546] NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD

LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLID
ATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCYFFRMPRQESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
[0547] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0548] 115.3 Cas9 TadAins1006

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
[0549] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR

KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLEGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLI
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGM
NHRVEITEGILADECAALLCYFFRMPRQSEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
[0550] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0551] 115.4 Cas9 TadAins1007

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
[0552] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDE
REVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRL
IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
[0553] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0554] 116.1 Cas9 TadAins C-말단 절단체2 792

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
[0555] LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR

LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGGSSGSETP
GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR
VIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLCYFFRMPRQSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0556]

[0557] 116.2 Cas9 TadAins C-말단 절단체2 791

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPG
TSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRV
IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMC
AGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADE
CAALLCYFFRMPRQGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD
PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
[0558] LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
[0559] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0560] 116.3 Cas9 TadAins C-말단 절단체2 790

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADEC
AALLCYFFRMPRQLGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE
INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
[0561] RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWD

PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN
PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS
KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
[0562] DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0563] 117 Cas9 델타 1017-1069

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYSSGSEVEFSHEYWMRHALTLAKRARDEREVPVGA
VLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYV

[0564] TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI

TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVN
FLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK

[0565] RYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0566] 118 Cas9 TadA-CP116ins 1067

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK

[0567] YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI

TLANGEIRKRPLIETNMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRAR
DEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY
RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0568] LGGD

[0569] 119 Cas9 TadAins 701

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPV
GAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATL
[0570] YVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV

EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDLTFKEDIQKAQVS
GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMA
RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY
YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0571] LGGD

[0572] 120 Cas9 TadACP136ins 1248

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI

[0573] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSMN
HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0574]

[0575] 121 Cas9 TadACP136ins 1052

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
[0576] KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY

YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLAMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGS
ETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVL
NNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEP
CVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGNGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0577] LGGD

[0578] 122 Cas9 TadACP136ins 1041

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
[0579] INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP

NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSMNHRVEITEG
ILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPES
SGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAI
GLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRI
GRVVFGVRNAKTGAAGSLMDVLHYPGNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[0580]

[0581] 123 Cas9 TadACP139ins 1299

[0582] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA

LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPE
DNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRMN
HRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA
GAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGDKPIREQAENIIHLFT
[0583] LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0584] LGGD

[0585] 124 Cas9 델타 792-872 TadAins

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML

[0586] ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0587] LGGD

[0588] 125 Cas9 델타 792-906 TadAins

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNA
KTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ
KKAQSSTDGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALI
KKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT
EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK
[0589] VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL
PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS
PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
[0590] HQSITGLYETRIDLSQLGGD

[0591] 126 TadA CP65ins 1003

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGR
VVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRM
PRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHA
[0592] LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0593] LGGD

[0594] 127 TadA CP65ins 1016

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
[0595] REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM
CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSE
VEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
PYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0596] LGGD

[0597] 128 TadA CP65ins 1022

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
[0598] MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMITAHAEIMALRQGGLVMQNYRLIDATLYV
TFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEI
TEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESAT
PESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWN
RAIGLHDPAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0599] LGGD

[0600] 129 TadA CP65ins 1029

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
[0601] LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEITAHAEIMALRQGGLVMQNYRL
IDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPG
MNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETP
GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNR
VIGEGWNRAIGLHDPGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0602] LGGD

[0603] 130 TadA CP65ins 1041

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
[0604] FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSTAHAEIMALR
QGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAA
GSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQS
STDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREV
PVGAVLVLNNRVIGEGWNRAIGLHDPNIMNFFKTEITLANGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0605] LGGD

[0606] 131 TadA CP65ins 1054

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
[0607] LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP

INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG
RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFR
MPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRH
ALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPGEIRKRPLIE
TNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR
NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT
LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
[0608] LGGD

[0609] 132 TadA CP65ins 1246

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP
NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLR
KQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD
LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQ
LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDD
SIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVE
KGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVR
NAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFN
AQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKR
ARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPSPEDNEQKQLFVEQHKH

[0610] YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT

LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ

[0611] LGGD

[0612] *프로토스페이서 인접 모티프*

[0613] 용어 "프로토스페이서 인접 모티프(PAM)" 또는 PAM-유사 모티프는 CRISPR 박테리아 적응성 면역계에서 Cas9 뉴클레아제에 의해 표적화된 DNA 서열 바로 뒤의 2-6개 염기 쌍 DNA 서열을 지칭한다. 일부 실시형태에서, PAM은 5' PAM(즉, 프로토스페이서의 5' 단부의 상류에 위치하는 것)일 수 있다. 다른 실시형태에서, PAM은 3' PAM(즉, 프로토스페이서의 5' 단부의 하류에 위치하는 것)일 수 있다.

[0614] PAM 서열은 표적 결합에 필수적이지만, 정확한 서열은 Cas 단백질의 유형에 따라 다르다.

[0615] 본 명세서에서 제공되는 염기 편집기는 정규 또는 비-정규 프로토스페이서 인접 모티프(PAM) 서열을 함유하는

뉴클레오티드 서열에 결합할 수 있는 CRISPR 단백질-유래 도메인을 포함할 수 있다. PAM 부위는 표적 폴리뉴클레오티드 서열에 근접한 뉴클레오티드 서열이다. 본 개시의 일부 양상은 상이한 PAM 특이성을 갖는 CRISPR 단백질의 전부 또는 일부를 포함하는 염기 편집기를 제공한다. 예를 들어, S. 피오게네스의 Cas9(spCas9)와 같은, Cas9 단백질은 일반적으로 특정 핵산 영역에 결합하기 위해 정규 NGG PAM 서열을 필요로 하며, 여기서 "NGG"의 "N"은 아데닌(A), 티민(T), 구아닌(G) 또는 시토신(C)이고, G는 구아닌이다. PAM은 CRISPR 단백질 특이적일 수 있으며, 다른 CRISPR 단백질 유래 도메인을 포함하는 다른 염기 편집기 간에 다를 수 있다. PAM은 표적 서열의 5' 또는 3'일 수 있다. PAM은 표적 서열의 상류 또는 하류에 있을 수 있다. PAM은 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 또는 그 이상의 뉴클레오티드 길이일 수 있다. 종종, PAM은 길이가 2 내지 6개 뉴클레오티드 사이이다. 몇몇 PAM 변형이 아래 표 1에 기재되어 있다.

[0616] 일부 실시형태에서, SpCas9는 PAM 핵산 서열 5'-NGC-3' 또는 5'-NGG-3'에 대한 특이성을 갖는다. 상기 양상의 다양한 실시형태에서, SpCas9는 표 1에 열거된 Cas9 또는 Cas9 변이체이다. 상기 양상의 다양한 실시형태에서, 변형된 SpCas9는 spCas9-MQKFRAER이다. 일부 실시형태에서, 변이체 Cas 단백질은 spCas9, spCas9-VRQR, spCas9-VRER, xCas9(sp), saCas9, saCas9-KKH, SpCas9-MQKFRAER, spCas9-MQKSER, spCas9-LRKIQK, 또는 spCas9-LRVSQL일 수 있다. 한 특정 실시형태에서, 아미노산 치환 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, 및 T1337R(SpCas9-MQKFRAER)을 포함하고, 변경된 PAM 5'-NGC-3'에 대한 특이성을 갖는 변형된 SpCas9가 사용된다.

[0617] 일부 실시형태에서, PAM은 NGT이다. 일부 실시형태에서, NGT PAM은 변이체이다. 일부 실시형태에서, NGT PAM 변이체는 하나 이상의 잔기 1335, 1337, 1135, 1136, 1218 및/또는 1219에서 표적화된 돌연변이를 통해 생성된다. 일부 실시형태에서, NGT PAM 변이체는 하나 이상의 잔기 1219, 1335, 1337, 1218에서 표적화된 돌연변이를 통해 생성된다. 일부 실시형태에서, NGT PAM 변이체는 하나 이상의 잔기 1135, 1136, 1218, 1219, 및 1335에서 표적화된 돌연변이를 통해 생성된다. 일부 실시형태에서, NGT PAM 변이체는 아래 표 2 및 3에 제공된 표적화된 돌연변이 세트에서 선택된다.

[0618] **표 2: 잔기 1219, 1335, 1337, 1218에서 NGT PAM 변이체 돌연변이**

[0619]

변이체	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R
9	L	L	T	
10	L	L	R	
11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

[0620] **표 3: 잔기 1135, 1136, 1218, 1219, 및 1335에서 NGT PAM 변이체 돌연변이**

[0621]

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

[0622]

[0623] 일부 실시형태에서, NGT PAM 변이체는 표 2 및 3의 변이체 5, 7, 28, 31, 또는 36에서 선택된다. 일부 실시형태에서, 변이체는 개선된 NGT PAM 인식을 나타낸다.

[0624] 일부 실시형태에서, NGT PAM 변이체는 잔기 1219, 1335, 1337, 및/또는 1218에서 돌연변이를 갖는다. 일부 실시형태에서, NGT PAM 변이체는 아래 표 4에 제공된 변이체로부터의 개선된 인식을 위한 돌연변이로 선택된다.

[0625] **표 4: 잔기 1219, 1335, 1337, 및 1218에서 NGT PAM 변이체 돌연변이**

변이체	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

[0626]

[0627] 일부 실시형태에서, NGT PAM은 아래 표 5에 제공된 변이체에서 선택된다.

[0628] **표 5. NGT PAM 변이체**

	NGTN 변이체	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
변이체 1	LRKIQK	L	R	K	I	-	Q	K
변이체 2	LRSVQK	L	R	S	V	-	Q	K
변이체 3	LRSVQL	L	R	S	V	-	Q	L
변이체 4	LRKIRQK	L	R	K	I	R	Q	K
변이체 5	LRSVRQK	L	R	S	V	R	Q	K
변이체 6	LRSVRQL	L	R	S	V	R	Q	L

[0629]

[0630] 일부 실시형태에서, Cas9 도메인은 스트렙토코커스 피오게네스(*Streptococcus pyogenes*)로부터의 Cas9(SpCas9) 도메인이다. 일부 실시형태에서, SpCas9 도메인은 뉴클레아제 활성 SpCas9, 뉴클레아제 비활성 SpCas9(SpCas9d), 또는 SpCas9 닉카아제(SpCas9n)이다. 일부 실시형태에서, SpCas9는 D9X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함하며, 본 명세서에 제공된 시티딘 데아미나제 또는 아데노신 데아미나제 중 임의의 것과 융합될 수 있다.

[0631] 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1135X, R1335X, 및 T1336X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1135E, R1335Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1135E, R1335Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1135X, R1335X, 및 T1336X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1135V, R1335Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 아미노산 서열 중 임의의 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1135V, R1335Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이들을 포함한다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1135X, G1217X, R1335X, 및 T1336X 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함하며, 여기서 X는 임의의 아미노산이다. 일부 실시형태에서, SpCas9 도메인은 하나 이상의 D1135V, G1217R, R1335Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, SpCas9 도메인은 D1135V, G1217R, R1335Q, 및 T1336R 돌연변이, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이들을 포함한다.

[0632] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 Cas9 폴리펩티드와 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일하다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 임의의 Cas9 폴리펩티드의 아미노산 서열을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 Cas9 도메인은 본 명세서에 기재된 임의의 Cas9 폴리펩티드의 아미노산 서열로 구성된다.

[0633] 일부 예에서, 본 명세서에 개시된 염기 편집기의 CRISPR 단백질-유래 도메인에 의해 인식되는 PAM은 염기 편집기를 코딩하는 삽입물(예를 들어, AAV 삽입물)에 대한 별도의 올리고뉴클레오티드 상의 세포에 제공될 수 있다. 이러한 실시형태에서, 별도의 올리고뉴클레오티드에 PAM을 제공하면, 인접한 PAM이 표적 서열과 동일한 폴리뉴클레오티드에 존재하지 않기 때문에, 그렇지 않으면 절단될 수 없는 표적 서열의 절단을 가능하게 할 수 있다.

[0634] 일 실시형태에서, S. 피오게네스(*S. pyogenes*) Cas9(SpCas9)를 게놈 조작을 위한 CRISPR 엔도뉴클레아제로 사용할 수 있다. 그러나, 다른 것을 사용할 수 있다. 일부 실시형태에서, 특정 게놈 표적을 표적으로 하기 위해 상이한 엔도뉴클레아제가 사용될 수 있다. 일부 실시형태에서, 비-NGG PAM 서열을 갖는 합성 SpCas9- 유래 변이체를 사용할 수 있다. 추가로, 다양한 종으로부터의 다른 Cas9 오르쏘로그가 확인되었으며, 이러한 "비-SpCas9"는 본 개시에 또한 유용할 수 있는 다양한 PAM 서열에 결합할 수 있다. 예를 들어, 비교적 큰 크기의 SpCas9(약 4 킬로베이스(kb) 코딩 서열)는 세포에서 효율적으로 발현될 수 없는 SpCas9 cDNA를 운반하는 플라스미드를 생성할 수 있다. 반대로 스타필로코커스 아우레우스(*Staphylococcus aureus*) Cas9(SaCas9)에 대한 코딩 서열은 SpCas9보다 약 1 킬로베이스 더 짧아 세포에서 효율적으로 발현될 수 있다. SpCas9와 유사하게, SaCas9 엔도뉴클레아제는 시험관내 포유동물 세포와 생체내 마우스에서 표적 유전자를 변형할 수 있다. 일부 실시형태에서,

Cas 단백질은 상이한 PAM 서열을 표적으로 삼을 수 있다. 일부 실시형태에서, 표적 유전자는, 예를 들어, Cas9 PAM, 5'-NGG에 인접할 수 있다. 다른 실시형태에서, 다른 Cas9 오르쏘로그는 다른 PAM 요구사항을 가질 수 있다. 예를 들어, S. 써모필러스(*S. thermophilus*)(CRISPR1의 경우 5'-NNAGAA, CRISPR3의 경우 5'-NGGNG) 및 나이세리아 메닌지티디스(*Neisseria meningiditis*)(5'-NNNNGATT)과 같은 다른 PAM도 표적 유전자 옆에서 발견될 수 있다.

[0635] 일부 실시형태에서, S. 피오게네스(*S. pyogenes*) 시스템의 경우, 표적 유전자 서열은 5'-NGG PAM에 선행(즉, 이에 대해 5'임)할 수 있고, 20-nt 가이드 RNA 서열은 반대 가닥과 염기 쌍을 이루어 PAM에 인접한 Cas9 절단을 매개할 수 있다. 일부 실시형태에서, 인접한 절단은 PAM의 상류에 있는 약 3개의 염기 쌍이거나 일 수 있다. 일부 실시형태에서, 인접한 절단은 PAM의 상류에 있는 약 10개의 염기 쌍이거나 일 수 있다. 일부 실시형태에서, 인접한 절단은 PAM의 상류에 약 0 내지 20개의 염기 쌍이거나 일 수 있다. 예를 들어, 인접한 절단은 PAM의 상류의 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 또는 30개 염기 쌍 다음에 있을 수 있다. 인접한 절단은 1 내지 30개의 염기 쌍만큼 PAM의 하류에 있을 수도 있다. PAM 서열에 결합할 수 있는 예시적인 SpCas9 단백질의 서열은 다음과 같다:

[0636] 예시적인 PAM-결합 SpCas9의 아미노산 서열은 다음과 같다:

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

[0637]

[0638] 예시적인 PAM-결합 SpCas9n 아미노산 서열은 다음과 같다:

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

[0639]

[0640] 예시적인 PAM-결합 SpEQR Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK

[0641]

VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQ
LGGD.

[0642] 이 서열에서, SpEQR Cas9를 생성

하기 위해 D1135, R1335 및 T1337에서 돌연변이될 수 있는 잔기 E1135, Q1335 및 R1337은 밑줄이 그어져 있고 굵은 글씨체로 표시되어 있다.

[0643] 예시적인 PAM-결합 SpVQR Cas9의 아미노산 서열은 다음과 같다:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
[0644] GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS

DKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLDATLIHQSITGLYETRIDLSQ
[0645] LGGD. 이 서열에서, SpVQR Cas9를 생성하기 위해 D1135, R1335 및 T1336에서 돌연변이될 수 있는 잔기 V1135, Q1335, 및 R1336은 밑줄이 그어져 있고 굵을 글씨체로 표시되어 있다.

[0646] 예시적인 PAM-결합 SpVRER Cas9의 아미노산 서열은 다음과 같다:

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

[0647]

[0648] 일부 실시형태에서, Cas9 도메인은 재조합 Cas9 도메인이다. 일부 실시형태에서, 재조합 Cas9 도메인은 SpyMacCas9 도메인이다. 일부 실시형태에서, SpyMacCas9 도메인은 뉴클레아제 활성 SpyMacCas9, 뉴클레아제 비활성 SpyMacCas9(SpyMacCas9d), 또는 SpyMacCas9 닉카아제(SpyMacCas9n)이다. 일부 실시형태에서, SaCas9 도메인, SaCas9d 도메인, 또는 SaCas9n 도메인은 비-정규 PAM을 갖는 핵산 서열에 결합할 수 있다. 일부 실시형태에서, SpyMacCas9 도메인, SpCas9d 도메인, 또는 SpCas9n 도메인은 NAA PAM 서열을 갖는 핵산 서열에 결합할 수 있다.

[0649] 예시적인 SpyMacCas9

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSP
LEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRD
RGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDL
SNDYLQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGF
[0650] TQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGED.

[0651] 일부 경우에, 변이체 Cas9 단백질은 폴리펩티드가 표적 DNA를 절단하는 활성이 감소되도록 H840A, P475A, W476A, N477A, D1125A, W1126A, 및 D1218A 돌연변이를 내포한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다. 또 다른 비제한적 예로서, 일부 경우에, 변이체 Cas9 단백질은 폴리펩티드가 표적 DNA를 절단하는 활성이 감소되도록 D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, 및 D1218A 돌연변이를 내포한다. 이러한 Cas9 단백질은 표적 DNA(예를 들어, 단일 가닥 표적 DNA)를 절단하는 활성이 감소되지만 표적 DNA(예를 들어, 단일 가닥 표적 DNA)에 결합하는 활성을 보유한다. 일부 경우에, 변이체 Cas9 단백질이 W476A 및 W1126A 돌연변이를 내포하거나 변이체 Cas9 단백질이 P475A, W476A, N477A, D1125A, W1126A, 및 D1218A 돌연변이를 내포할 때, 변이체 Cas9 단백질은 PAM 서열에 효율적으로 결합하지 않는다. 따라서, 일부 그러한 경우에서, 이러한 변이체 Cas9 단백질이 결합 방법에 사용될 때, 이 방법은 PAM 서열을 필요로 하지 않는다. 달리 말해서, 일부 경우에, 이러한 변이체 Cas9 단백질이 결합 방법에 사용되는 경우, 이 방법은 가이드 RNA를 포함할 수 있지만, 방법은 PAM 서열의 부재하에 수행될 수 있다(결합의 특이성은 그리하여 가이드 RNA의 표적화 세그먼트에 의해 제공된다). 상기 효과를 달성(즉, 하나 또는 다른 뉴클레아제 부분을 비활성화)하는 데 다른 잔기가 돌연변이될 수 있다. 비제한적인 예로서, 잔기 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, 및/또는 A987은 변경(즉, 치환)될 수 있다. 또한, 알라닌 치환 이외의 돌연변이가 적합하다.

[0652] 일부 실시형태에서, 염기 편집기의 CRISPR 단백질-유래 도메인은 정규 PAM 서열(NGG)을 갖는 Cas9 단백질의 전부 또는 일부를 포함할 수 있다. 다른 실시형태에서, 염기 편집기의 Cas9 유래 도메인은 비-정규 PAM 서열을 사용할 수 있다. 이러한 서열은 당업계에 기술되어 있으며 숙련된 기술자에게 명백할 것이다. 예를 들어, 비-정규 PAM 서열에 결합하는 Cas9 도메인은, 각각의 전체 내용이 본 명세서에 참조로 통합되는, 다음 문헌에 기재되어 있다: 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).

[0653] *고 충실도 Cas9 도메인*

[0654] 본 개시의 일부 양상은 고 충실도 Cas9 도메인을 제공한다. 일부 실시형태에서, 고 충실도 Cas9 도메인은, 상응

하는 야생형 Cas9 도메인에 비해, Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 정전기 상호작용을 감소시키는 하나 이상의 돌연변이를 포함하는 조작된 Cas9 도메인이다. 특정 이론에 구속되는 것을 원치않으면서, DNA의 당-포스페이트 백본과의 정전기적 상호작용을 감소시킨 고 충실도 Cas9 도메인은 표적-이탈 효과가 더 적을 수 있다. 일부 실시형태에서, Cas9 도메인(예를 들어, 야생형 Cas9 도메인)은 Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 연관성(association)을 감소시키는 하나 이상의 돌연변이를 포함한다. 일부 실시형태에서, Cas9 도메인은 Cas9 도메인과 DNA의 당-포스페이트 백본 사이의 연관성을 적어도 1%, 적어도 2%, 적어도 3%, 적어도 4%, 적어도 5%, 적어도 10%, 적어도 15%, 적어도 20%, 적어도 25%, 적어도 30%, 적어도 35%, 적어도 40%, 적어도 45%, 적어도 50%, 적어도 55%, 적어도 60%, 적어도 65%, 또는 적어도 70% 감소시키는 하나 이상의 돌연변이를 포함한다.

[0655] 일부 실시형태에서, 본 명세서에 제공된 임의의 Cas9 융합 단백질은 N497X, R661X, Q695X, 및/또는 Q926X 돌연변이 중 하나 이상, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함하며. 여기서 X는 임의의 아미노산이다. 일부 실시형태에서, 본 명세서에 제공된 임의의 Cas9 융합 단백질은 N497A, R661A, Q695A, 및/또는 Q926A 돌연변이 중 하나 이상, 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 일부 실시형태에서, Cas9 도메인은 D10A 돌연변이 또는 본 명세서에 제공된 임의의 아미노산 서열에서의 상응하는 돌연변이를 포함한다. 고 충실도를 갖는 Cas9 도메인은 당업계에 공지되어 있으며 숙련된 기술자에게 명백할 것이다. 예를 들어, 고 충실도를 갖는 Cas9 도메인은, 각각의 전체 내용이 본 명세서에 참조로 통합되는, 다음 문헌에 기재되어 있다: Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detec표 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).

[0656] **Cas9에 비해 고 충실도 Cas9 도메인 돌연변이는 굵은 글씨체로 표시되고 밑줄이 그어져 있다**

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFLKSDGFANRN
FM**A**LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK

[0657] NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK

RQLVETR**A**ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA

[0658] PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0659] *핵산 프로그래밍가능한 DNA 결합 단백질*

[0660] 본 개시의 일부 양상은 염기 편집기와 같은 단백질을, 특정 핵산(예를 들어, DNA 또는 RNA) 서열로 가이드하는 데 사용될 수 있는, 핵산 프로그래밍가능한 DNA 결합 단백질로서 역할하는 도메인을 포함하는 융합 단백질을 제

공한다. 특정 실시형태에서, 융합 단백질은 핵산 프로그래밍가능한 DNA 결합 단백질 도메인 및 데아미나제 도메인을 포함한다. DNA 결합 단백질은, 제한됨이 없이, Cas9(예를 들어, dCas9 및 nCas9), CasX, CasY, Cpf1, Cas12b/C2c1, 및 Cas12c/C2c3을 포함한다. Cas9와 다른 PAM 특이성을 갖는 프로그래밍가능한 폴리뉴클레오티드 결합 단백질의 한 예는 프레보텔라(*Prevotella*) 및 프란시셀라(*Francisella*)로부터의 일정한 간격을 두고 규칙적으로 분포하는 짧은 회문 반복부 1(Cpf1)이다. Cas9와 유사하게, Cpf1은 클래스 2 CRISPR 이펙터이기도 하다. Cpf1은 Cas9와는 다른 특징으로 강력한(roboust) DNA 간섭을 매개하는 것으로 밝혀졌다. Cpf1은 tracrRNA가 결여된 단일 RNA-가이드 엔도뉴클레아제이며, T가 풍부한 프로토스페이서 인접 모티프(TTN, TTTN, 또는 YTN)를 사용한다. 또한, Cpf1은 엇갈린 DNA 이중-가닥 파손을 통해 DNA를 절단한다. 16개의 Cpf1 계열 단백질 중 애시드아미노코쿠스(*Acidaminococcus*)와 라크노스피라세애(*Lachnospiraceae*)의 두 효소가 인간 세포에서 효율적인 게놈 편집 활성을 갖는 것으로 밝혀졌다. Cpf1 단백질은 당업계에 공지되어 있으며, 예를 들어, 이의 전체 내용이 본 명세서에 참조로 통합되는, 문헌[Yamano *et al*., "Crystal structure of Cpf1 in complex with guide RNA and target DNA." *Cell* (165) 2016, p. 949-962]에 이미 설명되었다.

[0661] 본 조성물 및 방법에서 가이드 뉴클레오티드 서열-프로그래밍가능한 DNA-결합 단백질 도메인으로 사용될 수 있는 뉴클레아제-비활성 Cpf1(dCpf1) 변이체가 또한 유용하다. Cpf1 단백질은 Cas9의 RuvC 도메인과 유사하지만 HNH 엔도뉴클레아제 도메인을 갖지 않는 RuvC-유사 엔도뉴클레아제 도메인을 가지며, Cpf1의 N-말단에는 Cas9의 알파-나선 인식 로브가 없다. 문헌[Zetsche *et al*., *Cell*, 163, 759-771, 2015](본 명세서에 참조로 통합됨)에서 Cpf1의 RuvC-유사 도메인이 양쪽 DNA 가닥을 절단하는 것을 담당하고 RuvC-유사 도메인의 비활성화는 Cpf1 뉴클레아제 활성을 비활성화한다 것을 밝혀냈다. 예를 들어, 프란시엘라 노비시다(*Francisella novicida*) Cpf1의 D917A, E1006A, 또는 D1255A에 상응하는 돌연변이는 Cpf1 뉴클레아제 활성을 비활성화한다. 일부 실시형태에서, 본 개시의 dCpf1은 D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, 또는 D917A/E1006A/D1255A에 상응하는 돌연변이를 포함한다. 임의의 돌연변이, 예를 들어, Cpf1의 RuvC 도메인을 비활성화하는 치환 돌연변이, 결실, 또는 삽입이 본 개시에 따라 사용될 수 있음을 이해해야 한다.

[0662] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 핵산 프로그래밍가능한 뉴클레오티드 결합 단백질은 Cpf1 단백질일 수 있다. 일부 실시형태에서, Cpf1 단백질은 Cpf1 닉카아제(nCpf1)이다. 일부 실시형태에서, Cpf1 단백질은 뉴클레아제 비활성 Cpf1(dCpf1)이다. 일부 실시형태에서, Cpf1, nCpf1, 또는 dCpf1은 본 명세서에 개시된 Cpf1 서열과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, dCpf1은 본 명세서에 개시된 Cpf1 서열과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함하며, D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, 또는 D917A/E1006A/D1255A에 상응하는 돌연변이를 포함한다. 다른 박테리아 종으로부터의 Cpf1도 본 개시에 따라 사용될 수 있음을 이해해야 한다.

[0663] 야생형 프란시엘라 노비시다(*Francisella novicida*) Cpf1(D917, E1006, 및 D1255는 굵은 글씨체로 기재되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
[0664] DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI

TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRI
KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0665]

[0666] 프란시엘라 노비시다(*Francisella novicid*a) Cpf1 D917A(A917, E1006, 및 D1255는 굵은 글씨체로 기재되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV

[0667]

SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRI
KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0668]

[0669] 프란시엘라 노비시다(*Francisella novicid*a) Cpf1 E1006A(D917, A1006, 및 D1255는 굵은 글씨체로 기재되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRI
[0670] KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0671] 프란시엘라 노비시다(*Francisella novicid*a) Cpf1 D1255A(D917, E1006, 및 A1255는 굵은 글씨체로 기재되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
[0672] KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT

TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRI
[0673] KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0674] 프란시엘라 노비시다(*Francisella novicid*a) Cpf1 D917A/E1006A(A917, A1006, 및 D1255는 굵은 글씨체로 기재되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN

[0675]

GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRI

[0676] KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0677] 프란시엘라 노비시다(*Francisella novicid*a) Cpf1 D917A/D1255A(A917, E1006, 및 A1255는 굵은 글씨체로 기재되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV

[0678] SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN

HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRI
[0679] KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0680] 프란시엘라 노비시다(*Francisella novicida*) Cpf1 E1006A/D1255A(D917, A1006, 및 A1255는 굵은 글씨체로 기재되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRI
[0681] KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0682] 프란시엘라 노비시다(*Francisella novicida*) Cpf1 D917A/E1006A/D1255A(A917, A1006, 및 A1255는 굵은 글씨체로 기재되고 밑줄이 그어져 있음)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
[0683] QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE

KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVFADLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAANGAYHIGLKGLMLLGRI
KNNQEGKKLNLVIKNEEYFEFVQNRNN

[0684]

[0685] 일부 실시형태에서, 융합 단백질에 존재하는 Cas9 도메인 중 하나는 PAM 서열에 대한 요구사항(requirements)이 없는 가이드 뉴클레오티드 서열-프로그래밍가능한 DNA-결합 단백질 도메인으로 대체될 수 있다.

[0686] 일부 실시형태에서, 핵산 프로그래밍가능한 DNA 결합 단백질(napDNAbp)은 미생물 CRISPR-Cas 시스템의 단일 이펙터이다. 미생물 CRISPR-Cas 시스템의 단일 이펙터는, 제한됨이 없이, Cas9, Cpf1, Cas12b/C2c1, 및 Cas12c/C2c3를 포함한다. 전형적으로, 미생물 CRISPR-Cas 시스템은 클래스 1 및 클래스 2 시스템으로 구분된다. 클래스 1 시스템은 다중서브유닛 이펙터 복합체를 갖는 반면, 클래스 2 시스템은 단일 단백질 이펙터를 갖는다. 예를 들어, Cas9 및 Cpf1은 클래스 2 이펙터이다. Cas9 및 Cpf1 외에도 3개의 별개의 클래스 2 CRISPR-Cas 시스템(Cas12b/C2c1 및 Cas12c/C2c3)이, 이의 전체 내용이 본 명세서에 참조로 통합되는, 문헌[Shmakov *et al.*, "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", *Mol. Cell*, 2015 Nov 5; 60(3): 385-397]에 기재되어 있다. 시스템의 두 이펙터인, Cas12b/C2c1 및 Cas12c/C2c3은, Cpf1과 관련된 RuvC-유사 엔도뉴클레아제 도메인을 함유한다. 세 번째 시스템은, 2개의 예측된 HEPN RNase 도메인이 있는 이펙터를 포함한다. 성숙한 CRISPR RNA의 생산은, Cas12b/C2c1에 의한 CRISPR RNA의 생산과 달리, tracrRNA에 독립적이다. Cas12b/C2c1은 DNA 절단을 위해 CRISPR RNA와 tracrRNA 둘 다에 의존적이다.

[0687] 알리사이클로바실러스 애시도테레스트리스(*Alicyclobaccillus acidoterrastris*) Cas12b/C2c1(AacC2c1)의 결정 구조는 키메라 단일 분자 가이드 RNA(sgRNA)와의 복합체인 것으로 밝혀졌다. 예를 들어, 이의 전체 내용이 본 명세서에 참조로 통합되는, 문헌[Liu *et al.*, "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", Mol. Cell, 2017 Jan 19; 65(2):310-322] 참조. 결정 구조는 3원(ternary) 복합체로서 표적 DNA에 결합된 알리사이클로바실러스 애시도테레스트리스(*Alicyclobacillus acidoterrestris*) C2c1에서도 보고되었다. 예를 들어, 이의 전체 내용이 본 명세서에 참조로 통합되는, 문헌[Yang *et al.*, "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell*, 2016 Dec 15; 167(7):1814-1828] 참조. 표적 가닥 및 비-표적 DNA 가닥 둘 다를 가진, AacC2c1의 촉매적으로 적격인 입체형태(conformations)는 단일 RuvC 촉매 포켓 내에 독립적으로 포획되어, Cas12b/C2c1-매개 절단으로 표적 DNA의 엇갈린(staggered) 7개-뉴클레오티드 파손(break)을 초래한다. Cas12b/C2c1 3원 복합체와 이전에 확인된 Cas9 및 Cpf1 대응물 간의 구조적 비교는 CRISPR-Cas9 시스템에서 사용되는 메커니즘의 다양성을 입증한다.

[0688] 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질의 핵산 프로그래밍가능한 DNA 결합 단백질(napDNAbp)은 Cas12b/C2c1, 또는 Cas12c/C2c3 단백질일 수 있다. 일부 실시형태에서, napDNAbp는 Cas12b/C2c1 단백질이다. 일부 실시형태에서, napDNAbp는 Cas12c/C2c3 단백질이다. 일부 실시형태에서, napDNAbp는 자연적으

로-발생하는 Cas12b/C2c1 또는 Cas12c/C2c3 단백질과 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 일부 실시형태에서, napDNAbp는 자연적으로-발생하는 Cas12b/C2c1 또는 Cas12c/C2c3 단백질이다. 일부 실시형태에서, napDNAbp는 본 명세서에 제공된 napDNAbp 서열 중 어느 하나와 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일한 아미노산 서열을 포함한다. 다른 박테리아 종으로부터의 Cas12b/C2c1 또는 Cas12c/C2c3도 본 개시에 따라 사용될 수 있음을 이해해야 한다.

[0689] Cas12b/C2c1 (uniprot.org/uniprot/T0D7A2#2)

[0690] sp|T0D7A2|C2C1_ALIAG CRISPR-associated endo-nuclease C2c1 OS = *Alicyclobacillus acido- terrestris* (strain ATCC 49025 / DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1

MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNG
DGEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAI
GAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKA
ETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDM
KEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNT
RRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDAT
AHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISM
SEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARD
VYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKL
GSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAV
HERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSW
AKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRH
MGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVS
GQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVA
KYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTM
YAAFSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRAD
DLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGEVD
GELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELL
VEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSRVPLQD
SACENTGDI

[0691]

[0692] *핵 위치결정 서열(NLS)을 포함하는 융합 단백질*

[0693] 일부 실시형태에서, 본 명세서에 제공된 융합 단백질은 하나 이상(예를 들어, 2개, 3개, 4개, 5개) 핵 표적화 서열, 예를 들어, 핵 위치결정 서열(NLS)을 추가로 포함한다. 한 실시형태에서, 이분 NLS가 사용된다. 일부 실시형태에서, NLS는 NLS를 포함하는 단백질의 세포핵 내로 (예를 들어, 핵 수송에 의해) 도입을 촉진하는 아미노산 서열을 포함한다. 일부 실시형태에서, 본 명세서에 제공된 임의의 융합 단백질은 핵 위치결정 서열(NLS)을 추가로 포함한다. 일부 실시형태에서, NLS는 융합 단백질의 N-말단에 융합된다. 일부 실시형태에서, NLS는 융합 단백질의 C-말단에 융합된다. 일부 실시형태에서, NLS는 Cas9 도메인의 N-말단에 융합된다. 일부 실시형태에서, NLS는 nCas9 도메인 또는 dCas9 도메인의 C-말단에 융합된다. 일부 실시형태에서, NLS는 데아미나제의 N-말단에 융합된다. 일부 실시형태에서, NLS는 데아미나제의 C-말단에 융합된다. 일부 실시형태에서, NLS는 하나 이상의 링커를 통해 융합 단백질에 융합된다. 일부 실시형태에서, NLS는 링커없이 융합 단백질에 융합된다. 일부 실시형태에서, NLS는 본 명세서에서 제공되거나 참조된 NLS 서열 중 어느 하나의 아미노산 서열을 포함한다. 추가의 핵 위치결정 서열은 당업계에 공지되어 있고 숙련된 기술자에게 명백할 것이다. 예를 들어, NLS 서열은 Plank 등의 PCT/EP2000/011690에 기재되어 있으며, 그 내용은 예시적인 핵 위치결정 서열의 개시를 위해 본 명세서에 참조로 통합된다. 일부 실시형태에서, NLS는 아미노산 서열

PKKKRKVEGADKRTADGSEFES PKKKRKV,
KRTADGSEFESPKKKRKV, KRPAATKKAGQAKKKK, KKTELQTTNAENKTKKL,
KRGINDRNFWRGENGRKTR, RKSGKIAAIVVKRPRKPKKKRKV, 또는
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC
를 포함한다.

[0694] 일부 실시형태에서, NLS는 링커 내에 존재하거나 NLS는 링커, 예를 들어, 본 명세서에 기재된 링커에 의해 측접된다. 일부 실시형태에서, N-말단 또는 C-말단 NLS는 이분 NLS이다. 이분 NLS는 비교적 짧은 스페이서 서열로 분리된 2개의 염기성(basic) 아미노산 클러스터를 포함한다(따라서 이분 - 2개 부분인 반면, 단분(monopartite) NLS는 그렇지 않다). 뉴클레오플라스민의 NLS인, KR[PAATKKAGQA]KKKK는 편재성(ubiquitous) 이분 신호의 원형(prototype)이며: 약 10개 아미노산의 스페이서로 분리된, 2개의 염기성 아미노산 클러스터이다. 예시적인 이분 NLS의 서열은 다음과 같다: **PKKKRKVEGADKRTADGSEFES PKKKRKV** .

[0695] 일부 실시형태에서, 본 발명의 융합 단백질은 링커 서열을 포함하지 않는다. 일부 실시형태에서, 하나 이상의 도메인 또는 단백질 사이에 링커 서열이 존재한다.

[0696] 본 개시의 융합 단백질은 하나 이상의 추가 특징을 포함할 수 있음을 이해해야 한다. 예를 들어, 일부 실시형태에서, 융합 단백질은 억제제, 세포질 위치결정 서열, 핵 외수송 서열(nuclear export sequences)과 같은 수송 서열, 또는 다른 위치결정 서열뿐만 아니라, 융합 단백질의 가용화, 정제, 또는 검출에 유용한 서열 태그를 포함할 수 있다. 본 명세서에 제공된 적합한 단백질 태그는, 이로만 제한되는 것은 아니지만, 비오틴 카복실라제 운반 단백질(BCCP; biotin carboxylase carrier protein) 태그, myc-태그, 칼모둘린-태그, FLAG-태그, 헤마글루티닌(HA)-태그, 히스티딘 태그 또는 His-태그로도 지칭되는 폴리히스티딘 태그, 말토스 결합 단백질(MBP)-태그, nus-태그, 글루타티온-S-트랜스퍼라제(GST)-태그, 녹색 형광 단백질(GFP)-태그, 티오레독신-태그, S-태그, Softags(예를 들어, Softag 1, Softag 3), 연쇄상 태그, 비오틴 리가아제 태그, FlAsH 태그, V5 태그, 및 SBP 태그를 포함한다. 추가의 적합한 서열은 해당 기술분야의 당업자에게 명백할 것이다. 일부 실시형태에서, 융합 단백질은 하나 이상의 His 태그를 포함한다.

[0697] *링커*

[0698] 특정 실시형태에서, 링커는 본 발명의 펩티드 또는 펩티드 도메인 중 임의의 것을 연결하는 데 사용될 수 있다. 링커는 공유 결합처럼 간단할 수도 있고, 원자들이 많은 길이의 중합체 링커일 수도 있다. 특정 실시형태에서, 링커는 폴리펩티드이거나 아미노산을 기반으로 한다. 다른 실시형태에서, 링커는 펩티드와 유사하지 않다. 특정 실시형태에서, 링커는 공유 결합(예를 들어, 탄소-탄소 결합, 이황화 결합, 탄소-헤테로원자 결합 등)이다. 특정 실시형태에서, 링커는 아미드 연결의 탄소-질소 결합이다. 특정 실시형태에서, 링커는 환형 또는 비환형, 치환 또는 비치환, 분지 또는 비분지 지방족 또는 헤테로지방족 링커이다. 특정 실시형태에서, 링커는 중합체성이다(예를 들어, 폴리에틸렌, 폴리에틸렌 글리콜, 폴리아미드, 폴리에스테르 등). 특정 실시형태에서, 링커는 아미노알칸산의 단량체, 이량체 또는 중합체를 포함한다. 특정 실시형태에서, 링커는 아미노알칸산(예를 들어, 글리신, 에탄올산, 알라닌, 베타-알라닌, 3-아미노프로판산, 4-아미노부탄산, 5-펜탄산 등)을 포함한다. 특정 실시형태에서, 링커는 아미노헥산산(Ahx)의 단량체, 이량체 또는 중합체를 포함한다. 특정 실시형태에서, 링커는 카보시클릭 모이어티(예를 들어, 시클로펜탄, 시클로헥산)를 기반으로 한다. 다른 실시형태에서, 링커는 폴리에틸렌 글리콜 모이어티(PEG)를 포함한다. 다른 실시형태에서, 링커는 아미노산을 포함한다. 특정 실시형태에서, 링커는 펩티드를 포함한다. 특정 실시형태에서, 링커는 아릴 또는 헤테로아릴 모이어티를 포함한다. 특정 실시형태에서, 링커는 페닐 고리를 기반으로 한다. 링커는 펩티드로부터 링커로의 친핵체(예를 들어, 티올, 아미노)의 부착을 용이하게 하는 기능화된 모이어티를 포함할 수 있다. 모든 친전 자체는 링커의 일부로 사용될 수 있다. 예시적인 친전자체는 활성화된 에스테르, 활성화된 아미드, 마이클(Michael) 수용체, 알킬 할라이드, 아릴 할라이드, 아실 할라이드 및 아이소티오시아네이트를 포함하지만, 이로만 제한되는 것은 아니다.

[0699] 일부 실시형태에서, 링커는 아미노산 또는 복수의 아미노산(예를 들어, 펩티드 또는 단백질)이다. 일부 실시형태에서, 링커는 결합(예를 들어, 공유결합), 유기 분자, 기, 중합체, 또는 화학적 모이어티이다. 일부 실시형태에서, 링커는 약 3 내지 104개(예를 들어, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 50개, 55개, 60개, 65개, 70개, 75개, 80개, 85개, 90개, 95개, 또는 100개) 아미노산 길이이다.

[0700] *가이드 RNA를 지니는 Cas9 복합체*

[0701] 본 개시의 일부 양상은 본 명세서에 제공된 임의의 융합 단백질, 및 가이드 RNA를 포함하는 복합체를 제공한다. 핵염기 편집기의 활성을 위한 최적의 길이를 달성하기 위해, 융합 단백질 도메인을 연결하는 임의의 방법이 사용될 수 있다(예를 들어, $(SGGS)_n$, $(GGGS)_n$, $(GGGGS)_n$, 및 $(G)_n$ 형태의 매우 유연한 링커로부터, $(EAAAK)_n$, $(SGGS)_n$, SGSETPGTSESATPES(예를 들어, 전체 내용이 본 명세서에 참조로 통합되는, 문헌[Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82] 참조), 및 $(XP)_n$ 모티프 형태의 더 강직한 링커에 이르는 범위의 것). 일부 실시형태에서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15이다. 일부 실시형태에서, 링커는 $(GGS)_n$ 모티프를 포함하며, 여기서 n은 1, 3, 또는 7이다. 일부 실시형태에서, 본 명세서에 제공된 융합 단백질의 Cas9 도메인은 아미노산 서열 SGSETPGTSESATPES를 포함하는 링커를 통해 융합된다.

[0702] 일부 실시형태에서, 가이드 핵산(예를 들어, 가이드 RNA)은 15개 내지 100개 뉴클레오티드 길이이고, 표적 서열에 상보적인 적어도 10개의 연속 뉴클레오티드 서열을 포함한다. 일부 실시형태에서, 가이드 RNA는 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 40개, 41개, 42개, 43개, 44개, 45개, 46개, 47개, 48개, 49개, 또는 50개의 뉴클레오티드 길이이다. 일부 실시형태에서, 가이드 RNA는 표적 서열에 상보적인 15개, 16개, 17개, 18개, 19개, 20개, 21개, 22개, 23개, 24개, 25개, 26개, 27개, 28개, 29개, 30개, 31개, 32개, 33개, 34개, 35개, 36개, 37개, 38개, 39개, 또는 40개의 연속 뉴클레오티드의 서열을 포함한다. 일부 실시형태에서, 표적 서열은 DNA 서열이다. 일부 실시형태에서, 표적 서열은 박테리아, 효모, 진균, 곤충, 식물, 또는 동물의 게놈 내의 서열이다. 일부 실시형태에서, 표적 서열은 인간 게놈의 서열이다. 일부 실시형태에서, 표적 서열의 3' 말단은 정규 PAM 서열(NGG)에 바로 인접해 있다. 일부 실시형태에서, 표적 서열의 3' 말단은 비-정규 PAM 서열(예를 들어, 표 1에 열거된 서열)에 바로 인접해 있다.

[0703] 본 개시의 일부 양상은 본 명세서에 제공된 융합 단백질, 또는 복합체를 이용하는 방법을 제공한다. 예를 들어, 본 개시의 일부 양상은 DNA 분자를 본 명세서에 제공된 임의의 융합 단백질 및 적어도 하나의 가이드 RNA와 접촉시키는 단계를 포함하는 방법을 제공하며, 여기서 가이드 RNA는 약 15개 내지 100개 뉴클레오티드 길이이고 표적 서열에 상보적인 적어도 10개의 연속 뉴클레오티드 서열을 포함한다. 일부 실시형태에서, 표적 서열의 3' 말단은 AGC, GAG, TTT, GTG, 또는 CAA 서열에 바로 인접해 있다. 일부 실시형태에서, 표적 서열의 3' 말단은 NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, 또는 5'(TTTV) 서열에 바로 인접해 있다.

[0704] 각각의 서열에서 특정 위치 또는 잔기의 번호 매기기는 사용된 특정 단백질 및 번호 매기기 체계에 의존한다는 것을 이해할 것이다. 번호 매기기는, 예를 들어, 성숙 단백질의 전구체와 성숙 단백질 자체에서 다를 수 있으며, 종에서 종마다의 서열의 차이가 번호 매기기에 영향을 미칠 수 있다. 당업자는, 예를 들어, 서열 정렬 및 상동성 잔기의 결정에 의해 당업계에 잘 알려진 방법에 의해 임의의 상동성 단백질 및 각각의 코딩 핵산에서 각각의 잔기를 확인할 수 있을 것이다.

[0705] 본 명세서에 개시된 임의의 융합 단백질을 표적 부위, 예를 들어, 편집될 돌연변이를 포함하는 부위로 표적화하기 위해, 전형적으로 가이드 RNA와 함께 융합 단백질을 공동-발현시키는 것이 필요하다는 것은 당업자에게 명백할 것이다. 본 명세서의 다른 곳에서 더 자세히 설명된 바와 같이, 가이드 RNA는 전형적으로 Cas9 결합을 허용하는 tracrRNA 프레임워크 및 Cas9:핵산 편집 효소/도메인 융합 단백질에 서열 특이성을 부여하는 가이드 서열을 포함한다. 대안적으로, 가이드 RNA 및 tracrRNA는 2개의 핵산 분자로서, 별개로 제공될 수 있다. 일부 실시형태에서, 가이드 RNA는 구조를 포함하고, 여기서 가이드 서열은 표적 서열에 상보적인 서열을 포함한다. 가이드 서열은 전형적으로 20개 뉴클레오티드 길이이다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특정 게놈 표적 부위로 표적화하기 위한 적합한 가이드 RNA의 서열은 본 개시에 기초하여 당업자에게 명백할 것이다. 이러한 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오티드의 상류 또는 하류에 있는 50개 뉴클레오티드 내의 핵산 서열에 상보적인 가이드 서열을 포함한다. 제공된 융합 단백질 중 임의의 것을 특정 표적 서열로 표적화하기에 적합한 일부 예시적인 가이드 RNA 서열이 본 명세서에서 제공된다.

[0706] *시티딘 데아미나제, 아데노신 데아미나제 및 Cas9 도메인을 포함하는 융합 단백질을 이용하는 방법*

[0707] 본 개시의 일부 양상은 본 명세서에 제공된 융합 단백질, 또는 복합체를 이용하는 방법을 제공한다. 예를 들어, 본 개시의 일부 양상은 돌연변이를 코딩하는 DNA 분자를 본 명세서에 제공된 임의의 융합 단백질, 및 적어도 하나의 가이드 RNA와 접촉시키는 단계를 포함하는 방법을 제공하며, 여기서 가이드 RNA는 약 15개 내지 100개 뉴

클레오티드 길이이고 표적 서열에 상보적인 적어도 10개의 연속 뉴클레오티드의 서열을 포함한다. 일부 실시형태에서, 표적 서열의 3' 말단은 정규 PAM 서열(NGG)에 바로 인접해 있다. 일부 실시형태에서, 표적 서열의 3' 말단은 정규 PAM 서열(NGG)에 바로 인접해 있지 않다. 일부 실시형태에서, 표적 서열의 3' 말단은 AGC, GAG, TTT, GTG, 또는 CAA 서열에 바로 인접해 있다. 일부 실시형태에서, 표적 서열의 3' 말단은 NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, 또는 5' (TTTV) 서열에 바로 인접해 있다.

[0708] 각각의 서열에서 특정 위치 또는 잔기의 번호 매기기는 사용된 특정 단백질 및 번호 매기기 체계에 의존한다는 것을 이해할 것이다. 번호 매기기는, 예를 들어, 성숙 단백질의 전구체와 성숙 단백질 자체에서 다를 수 있으며, 종에서 종마다의 서열의 차이가 번호 매기기에 영향을 미칠 수 있다. 당업자는, 예를 들어, 서열 정렬 및 상동성 잔기의 결정에 의해 당업계에 잘 알려진 방법에 의해 임의의 상동성 단백질 및 각각의 코딩 핵산에서 각각의 잔기를 확인할 수 있을 것이다.

[0709] 본 명세서에 개시된 것과 같은, Cas9 도메인 및 시티딘 데아미나제, 또는 아데노신 데아미나제를 포함하는 임의의 융합 단백질을 표적 부위, 예를 들어, 편집될 돌연변이를 포함하는 부위로 표적화하기 위해, 전형적으로 가이드 RNA, 예를 들어, sgRNA와 함께 융합 단백질을 공동-발현시키는 것이 필요하다는 것은 당업자에게 명백할 것이다. 본 명세서의 다른 곳에서 더 자세히 설명된 바와 같이, 가이드 RNA는 전형적으로 Cas9 결합을 허용하는 tracrRNA 프레임워크 및, Cas9:핵산 편집 효소/도메인 융합 단백질에 서열 특이성을 부여하는, 가이드 서열을 포함한다. 대안적으로, 가이드 RNA 및 tracrRNA는, 2개의 핵산 분자로서, 별개로 제공될 수 있다. 일부 실시형태에서, 가이드 RNA는 소정 구조를 포함하고, 여기서 가이드 서열은 표적 서열에 상보적인 서열을 포함한다. 가이드 서열은 전형적으로 20개 뉴클레오티드 길이이다. Cas9:핵산 편집 효소/도메인 융합 단백질을 특정 게놈 표적 부위로 표적화하기 위한 적합한 가이드 RNA의 서열은 본 개시에 기초하여 당업자에게 명백할 것이다. 이러한 적합한 가이드 RNA 서열은 전형적으로 편집될 표적 뉴클레오티드의 상류 또는 하류에 있는 50개 뉴클레오티드 이내의 핵산 서열에 상보적인 가이드 서열을 포함한다. 제공된 융합 단백질 중 임의의 것을 특정 표적 서열로 표적화하기에 적합한 일부 예시적인 가이드 RNA 서열이 본 명세서에서 제공된다.

[0710] *염기 편집기 효율*

[0711] 본 발명의 융합 단백질은 유리하게는, 상당한 비율의 인델을 생성함이 없이, 돌연변이를 포함하는 특정 뉴클레오티드 염기를 변형시킨다. 본 명세서에 사용된 용어 "인델(들)"은 핵산 내의 뉴클레오티드 염기의 삽입 또는 결실을 지칭한다. 이러한 삽입 또는 결실은 유전자의 코딩 영역 내에서 프레임 이동 돌연변이를 유발할 수 있다. 일부 실시형태에서, 핵산에서 다수의 삽입 또는 결실(즉, 인델)을 생성함이 없이, 핵산 내의 특정 뉴클레오티드를 효율적으로 변형(예를 들어, 돌연변이화)하는 염기 편집기를 생성하는 것이 바람직하다. 특정 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 인델 대비(versus) 더 큰 비율의 의도된 변형(예를 들어, 돌연변이)을 생성할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 1:1을 초과하는 의도된 점 돌연변이 대 인델의 비율을 생성할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5:1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 200:1, 적어도 300:1, 적어도 400:1, 적어도 500:1, 적어도 600:1, 적어도 700:1, 적어도 800:1, 적어도 900:1, 또는 적어도 1000:1, 또는 그 이상의 의도된 점 돌연변이 대 인델의 비율을 생성할 수 있다. 의도된 돌연변이 및 인델의 수는 임의의 적절한 방법을 DL용하여 결정될 수 있다.

[0712] 일부 실시형태에서, 본 명세서에 제공된 염기 편집기는 핵산 영역에서 인델의 형성을 제한할 수 있다. 일부 실시형태에서, 영역은 염기 편집기에 의해 표적화된 뉴클레오티드 또는 염기 편집기에 의해 표적화된 뉴클레오티드의 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 또는 10개 뉴클레오티드 내의 영역에 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 핵산 영역에서 인델의 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 인델의 수는 핵산(예를 들어, 세포의 게놈 내 핵산)이 염기 편집기에 노출된 시간에 따라 달라질 수 있다. 일부 실시형태에서, 임의의 수 또는 비율의 인델은 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 동안 염기 편집기에 핵산(예를 들어, 세포 게놈 내의 핵산)을 노출시킨 후 결정된다.

[0713] 본 개시의 일부 양상은 본 명세서에 제공된 임의의 염기 편집기가 상당한 수의 의도하지 않은 돌연변이를 생성

하지 않으면서 핵산(예를 들어, 대상체의 게놈 내의 핵산)에서 의도된 돌연변이를 효율적으로 생성할 수 있다는 인식에 기초한다. 일부 실시형태에서, 의도된 돌연변이는 돌연변이를 변경하거나 수정하도록 특별히 설계된 gRNA에 결합된 특정 염기 편집기에 의해 생성되는 돌연변이이다. 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 1:1보다 큰 의도된 돌연변이 대 의도하지 않은 돌연변이(예를 들어, 의도된 돌연변이:의도하지 않은 돌연변이)의 비율을 생성할 수 있다. 일부 실시형태에서, 본 명세서에 제공된 임의의 염기 편집기는 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5 : 1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 10:1, 적어도 12:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1 , 적어도 50:1, 적어도 100:1, 적어도 150:1, 적어도 200:1, 적어도 250:1, 적어도 500:1, 또는 적어도 1000:1 또는 그 이상인 의도된 돌연변이 대 의도하지 않은 돌연변이의 비율을 생성 할 수 있다. 본 명세서에 "*염기 편집기 효율*" 섹션에 기술된 염기 편집기의 특징은 임의의 융합 단백질, 또는 본 명세서에 제공된 융합 단백질을 이용하는 방법에 적용될 수 있음을 이해해야 한다.

[0714] *핵산 편집 방법*

[0715] 본 개시의 일부 양상은 핵산 편집 방법을 제공한다. 일부 실시형태에서, 방법은 관심 폴리펩티드(예를 들어, 질병 유전자의 발현 생성물)를 코딩하는 핵산 분자의 핵염기를 편집하는 방법이다. 일부 실시형태에서, 이 방법은 다음 단계를 포함한다: a) 핵산의 표적 영역(예를 들어, 이중-가닥 DNA 서열)을 염기 편집기 및 가이드 핵산(예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계, b) 상기 표적 영역의 가닥 분리를 유도하는 단계, c) 표적 영역의 단일 가닥에서 상기 표적 핵염기 쌍의 제1 핵염기를 제2 핵염기로 전환시키는 단계, 및 d) nCas9을 이용하여 상기 표적 영역의 하나 이하의 가닥을 절단하는 단계로서, 여기서 상기 제1 핵염기에 상보적인 제3 핵염기는 상기 제2 핵염기에 상보적인 제4 핵염기로 대체되는 단계. 일부 실시형태에서, 이 방법은 핵산에서 20% 미만의 인델 형성을 초래한다. 일부 실시형예에서, 단계 b가 생략된다는 것을 이해해야 한다. 일부 실시형태에서, 이 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2% 미만, 또는 0.1% 미만의 인델 형성을 초래한다. 일부 실시형태에서, 이 방법은 제2 핵염기를 제4 핵염기에 상보적인 제5 핵염기로 대체하며, 그로 인해 의도된 편집된 염기 쌍(예를 들어, G·C에서 A·T로)을 생성하는 단계를 추가로 포함한다. 일부 실시형태에서, 의도된 염기 쌍의 적어도 5%가 편집된다. 일부 실시형태에서, 의도된 염기 쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 또는 50%가 편집된다.

[0716] 일부 실시형태에서, 표적 뉴클레오티드에서 의도된 생성물 대 의도하지 않은 생성물의 비율은 적어도 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 또는 200:1이거나, 또는 이보다 더 크다. 일부 실시형태에서, 의도된 돌연변이 대 인델 형성의 비율은 1:1, 10:1, 50:1, 100:1, 500:1, 또는 1000:1을 초과하거나, 또는 이보다 더 크다. 일부 실시형태에서, 절단된 단일 가닥(니킹 가닥)은 가이드 핵산에 혼성화된다. 일부 실시형태에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥과 반대이다. 일부 실시형태에서, 염기 편집기는 dCas9 도메인을 포함한다. 일부 실시형태에서, 염기 편집기는 편집되지 않은 가닥을 보호하거나 이에 결합한다. 일부 실시형태에서, 의도 된 편집된 염기 쌍은 PAM 부위의 상류에 있다. 일부 실시형태에서, 의도된 편집된 염기 쌍은 PAM 부위의 상류에 있는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 또는 20개 뉴클레오티드이다. 일부 실시형태에서, 의도된 편집된 염기 쌍은 PAM 부위의 하류에 있다. 일부 실시형태에서, 의도된 편집된 염기 쌍은 PAM 부위의 뉴클레오티드 하류에 있는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 또는 20개 뉴클레오티드이다. 일부 실시형태에서, 방법은 정규(예를 들어, NGG) PAM 부위를 필요로 하지 않는다. 일부 실시형태에서, 핵염기 편집기는 링커를 포함한다. 일부 실시형태에서, 링커는 1 내지 25개의 아미노산 길이이다. 일부 실시형태에서, 링커는 5 내지 20개 아미노산 길이이다. 일부 실시형태에서, 링커는 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 또는 20개 아미노산 길이이다. 한 실시형태에서, 링커는 32개 아미노산 길이이다. 또 다른 실시형태에서, "긴 링커"는 적어도 약 60개 아미노산 길이이다. 다른 실시형태에서, 링커는 약 3 내지 100개의 아미노산 길이이다. 일부 실시형태에서, 표적 영역은 표적 윈도우를 포함하고, 여기서 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 실시형태에서, 표적 윈도우는 1 내지 10개의 뉴클레오티드를 포함한다. 일부 실시형태에서, 표적 윈도우는 1 내지 9개, 1 내지 8개, 1 내지 7개, 1 내지 6개, 1 내지 5개, 1 내지 4개, 1 내지 3개, 1 내지 2개, 또는 1개 뉴클레오티드 길이이다. 일부 실시형태에서, 표적 윈도우는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 또는 20개 뉴클레오티드 길이이다. 일부 실시형태에서, 의도된 편집 된 염기 쌍은 표적 윈도우 내에 있다. 일부 실시형태에서, 표적 윈도우는 의도된 편집된 염기 쌍을 포함한다. 일부

실시형태에서, 방법은 본 명세서에 제공된 임의의 염기 편집기를 이용하여 수행된다.

[0717] 일부 실시형태에서, 본 개시는 뉴클레오티드(예를 들어, SNP)를 편집하는 방법을 제공한다. 일부 실시형태에서, 본 개시는 이중-가닥 DNA 서열의 핵염기 쌍을 편집하는 방법을 제공한다. 일부 실시형태에서, 이 방법은 a) 이중-가닥 DNA 서열의 표적 영역을 염기 편집기 및 가이드 핵산(예를 들어, gRNA)을 포함하는 복합체와 접촉시키는 단계로서, 여기서 표적 영역은 표적 핵염기 쌍을 포함하는 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 표적 영역의 하나 이하의 가닥을 절단하는 단계,를 포함하며, 여기서 상기 제1 핵염기에 상보적인 제3 핵염기는 상기 제2 핵염기에 상보적인 제4 핵염기로 대체되고, 상기 제2 핵염기는 상기 제4 핵염기에 상보적인 제5 염기로 대체되며, 그로 인해 의도된 편집된 핵염기 쌍을 생성하며, 여기서 상기 의도된 편집된 염기 쌍을 생성하는 효율은 적어도 5%이다. 일부 실시형태에서, 단계 b는 생략된다는 것을 이해해야 한다. 일부 실시형태에서, 의도된 염기 쌍의 적어도 5%가 편집된다. 일부 실시형태에서, 의도된 염기 쌍의 적어도 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 또는 50%가 편집된다. 일부 실시형태에서, 이 방법은 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2%, 또는 0.1% 미만의 인델 형성을 초래한다. 일부 실시형태에서, 표적 뉴클레오티드에서 의도된 생성물 대 의도하지 않은 생성물의 비율은 적어도 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 또는 200:1이거나, 또는 이보다 더 크다. 일부 실시형태에서, 의도된 돌연변이 대 인델 형성의 비율은 1:1, 10:1, 50:1, 100:1, 500:1, 또는 1000:1을 초과하거나, 또는 이보다 더 크다. 일부 실시형태에서, 절단된 단일 가닥은 가이드 핵산에 혼성화된다. 일부 실시형태에서, 절단된 단일 가닥은 제1 핵염기를 포함하는 가닥과 반대이다. 일부 실시형태에서, 의도된 편집된 염기 쌍은 PAM 부위의 상류에 있다. 일부 실시형태에서, 의도된 편집된 염기 쌍은 PAM 부위의 상류에 있는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 또는 20개 뉴클레오티드이다. 일부 실시형태에서, 의도된 편집된 염기 쌍은 PAM 부위의 하류에 있다. 일부 실시형태에서, 의도된 편집된 염기 쌍은 PAM 부위의 하류에 있는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 또는 20개 뉴클레오티드이다. 일부 실시형태에서, 방법은 정규(예를 들어, NGG) PAM 부위를 필요로 하지 않는다. 일부 실시형태에서, 링커는 1 내지 25개의 아미노산 길이이다. 일부 실시형태에서, 링커는 5 내지 20개 아미노산 길이이다. 일부 실시형태에서, 링커는 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 또는 20개 아미노산 길이이다. 일부 실시형태에서, 표적 영역은 표적 윈도우를 포함하고, 여기서 표적 윈도우는 표적 핵염기 쌍을 포함한다. 일부 실시형태에서, 표적 윈도우는 1 내지 10개의 뉴클레오티드를 포함한다. 일부 실시형태에서, 표적 윈도우는 길이가 1 내지 9개, 1 내지 8개, 1 내지 7개, 1 내지 6개, 1 내지 5개, 1 내지 4개, 1 내지 3개, 1 내지 2개, 또는 1개 뉴클레오티드이다. 일부 실시형태에서, 표적 윈도우는 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 또는 20개 뉴클레오티드 길이이다. 일부 실시형태에서, 의도된 편집된 염기 쌍은 표적 윈도우 내에서 발생한다. 일부 실시형태에서, 표적 윈도우는 의도된 편집된 염기 쌍을 포함한다. 일부 실시형태에서, 핵염기 편집기는 본 명세서에 제공된 염기 편집기 중 임의의 하나이다.

[0718] ***다중 편집***

[0719] 일부 실시형태에서, 본 명세서에 제공된 염기 편집기 시스템은 하나 이상의 유전자에서 복수의 핵염기 쌍을 다중 편집할 수 있다. 일부 실시형태에서, 복수의 핵염기 쌍은 동일한 유전자에 위치한다. 일부 실시형태에서, 복수의 핵염기 쌍은 하나 이상의 유전자에 위치하며, 여기서 적어도 하나의 유전자는 상이한 유전자좌에 위치한다. 일부 실시형태에서, 다중 편집은 하나 이상의 가이드 폴리뉴클레오티드를 포함할 수 있다. 일부 실시형태에서, 다중 편집은 하나 이상의 염기 편집기 시스템을 포함할 수 있다. 일부 실시형태에서, 다중 편집은 단일 가이드 폴리뉴클레오티드를 갖는 하나 이상의 염기 편집기 시스템을 포함할 수 있다. 일부 실시형태에서, 다중 편집은 복수의 가이드 폴리뉴클레오티드를 갖는 하나 이상의 염기 편집기 시스템을 포함할 수 있다. 일부 실시형태에서, 다중 편집은 단일 염기 편집기 시스템을 갖는 하나 이상의 가이드 폴리뉴클레오티드를 포함할 수 있다. 일부 실시형태에서, 다중 편집은 표적 폴리뉴클레오티드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오티드를 포함할 수 있다. 일부 실시형태에서, 다중 편집은 표적 폴리뉴클레오티드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오티드를 포함할 수 있다. 일부 실시형태에서, 다중 편집은 표적 폴리뉴클레오티드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오티드 및 표적 폴리뉴클레오티드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오티드의 혼합을 포함할 수 있다. 본 명세서에 기술된 임의의 염기 편집기를 사용하는 다중 편집의 특성은 본 명세서에 제공된 임의의 염기 편집기를 사용하는 방법의 임의의 조합에 적용될 수 있음을 이해해야 한다. 본 명세서에 기재

된 바와 같은 임의의 염기 편집기를 사용한 다중 편집은 복수의 핵염기 쌍의 순차적 편집을 포함할 수 있음을 또한 이해해야 한다.

[0720] 일부 실시형태에서, 복수의 핵염기 쌍은 하나 이상의 유전자에 존재한다. 일부 실시형태에서, 복수의 핵염기 쌍은 동일한 유전자에 존재한다. 일부 실시형태에서, 하나 이상의 유전자 중 하나 이상의 유전자는 상이한 유전자좌에 위치한다.

[0721] 일부 실시형태에서, 편집은 적어도 하나의 단백질 코딩 영역에서 복수의 핵염기 쌍을 편집하는 것이다. 일부 실시형태에서, 편집은 적어도 하나의 단백질 비-코딩 영역에서 복수의 핵염기 쌍을 편집하는 것이다. 일부 실시형태에서, 편집은 적어도 하나의 단백질 코딩 영역 및 적어도 하나의 단백질 비-코딩 영역에서 복수의 핵염기 쌍을 편집하는 것이다.

[0722] 일부 실시형태에서, 편집은 하나 이상의 가이드 폴리뉴클레오티드와 연계된다. 일부 실시형태에서, 염기 편집기 시스템은 하나 이상의 염기 편집기 시스템을 포함할 수 있다. 일부 실시형태에서, 염기 편집기 시스템은 단일 가이드 폴리뉴클레오티드와 관련하여 하나 이상의 염기 편집기 시스템을 포함할 수 있다. 일부 실시형태에서, 염기 편집기 시스템은 복수의 가이드 폴리뉴클레오티드와 관련하여 하나 이상의 염기 편집기 시스템을 포함할 수 있다. 일부 실시형태에서, 편집은 단일 염기 편집기 시스템을 갖는 하나 이상의 가이드 폴리뉴클레오티드와 관련하여 이루어진다. 일부 실시형태에서, 편집은 표적 폴리뉴클레오티드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오티드와 연계된다. 일부 실시형태에서, 편집은 표적 폴리뉴클레오티드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오티드와 연계된다. 일부 실시형태에서, 편집은 표적 폴리뉴클레오티드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하지 않는 적어도 하나의 가이드 폴리뉴클레오티드 및 표적 폴리뉴클레오티드 서열에 대한 결합을 표적화하기 위해 PAM 서열을 필요로 하는 적어도 하나의 가이드 폴리뉴클레오티드의 혼합(mix)과 연계된다. 본 명세서에 기재된 임의의 염기 편집기를 사용하는 다중 편집의 특성은 본 명세서에 제공된 임의의 염기 편집기를 사용하는 방법의 임의의 조합에 적용될 수 있음을 이해해야 한다. 편집은 복수의 핵염기 쌍의 순차적 편집을 포함할 수 있음을 또한 이해해야 한다.

[0723] ***숙주 세포에서 융합 단백질의 발현***

[0724] 본 발명의 융합 단백질은 당업자에게 공지된 일상적인 방법을 사용하여 박테리아, 효모, 진균, 곤충, 식물 및 동물 세포를 포함하지만, 이로만 제한되는 것은 아닌, 사실상 임의의 관심 숙주 세포에서 발현될 수 있다. 예를 들어, 본 발명의 융합 단백질을 코딩하는 DNA는 cDNA 서열에 기초하여 CDS의 상류 및 하류에 적합한 프라이머를 설계함으로써 클로닝될 수 있다. 클로닝된 DNA는 직접, 또는 원하는 경우 제한 효소로의 소화 후, 또는 적절한 링커 및/또는 핵 위치결정 신호의 첨가 후 염기 편집 시스템의 하나 이상의 추가 구성요소를 코딩하는 DNA와 결찰될 수 있다. 염기 편집 시스템은 숙주 세포에서 번역되어 복합체화한다.

[0725] 핵염기 변형 활성을 갖는 하나 이상의 도메인을 코딩하는 하나 이상의 폴리뉴클레오티드(예를 들어, 아데노신 데아미나제, 시티딘 데아미나제, DNA 글리코실라제)를, napDNAbp를 코딩하는 폴리뉴클레오티드에 작동가능하게 연결하여 본 발명의 융합 단백질을 코딩하는 폴리뉴클레오티드를 제조함으로써 융합 단백질을 생성한다. 일부 실시형태에서, napDNAbp를 코딩하는 폴리뉴클레오티드, 및 핵염기 변형 활성을 갖는 도메인을 코딩하는 DNA는 각각 결합 도메인 또는 이의 결합 파트너를 코딩하는 DNA와 융합될 수 있거나, 또는 두 DNA 모두 분리(separation) 인테인을 코딩하는 DNA와 융합될 수 있으며, 이로써 핵산 서열-인식 변환 모듈과 핵산 염기 변환 효소가 숙주 세포에서 번역되어 복합체화한다. 이러한 경우, 링커 및/또는 핵 위치결정 신호는 원하는 경우 DNA 중 하나 또는 둘 다의 적절한 위치에 연결될 수 있다.

[0726] 본 명세서에 기재된 단백질 도메인을 코딩하는 DNA는 DNA를 화학적으로 합성함으로써, 또는 이의 전체 길이를 코딩하는 DNA를 구축하기 위해 PCR 방법 및 깁슨(Gibson) 조립 방법을 이용함으로써 합성된 부분적으로 중첩되는 짧은 올리고DNA 사슬을 연결함으로써 얻을 수 있다. 전체-길이의 DNA를 화학적 합성 또는 PCR 방법이나 깁슨 조립 방법의 조합으로 구축하는 것의 장점은 DNA가 도입되는 숙주에 따라 사용할 코돈을 CDS 전체-길이로 설계할 수 있다는 점이다. 이종성 DNA의 발현에 있어서, 그 DNA 서열을 숙주 유기체에서 많이 사용되는 코돈으로 변환함으로써 단백질 발현 수준이 증가할 것으로 기대된다. 사용될 숙주에 있어서의 코돈 사용 빈도의 데이터로서는, 예를 들면, Kazusa DNA Research Institute의 홈페이지에 공개된 유전자 코드 사용 빈도 데이터베이스(http://www.kazusa.or.jp/codon/index.html)를 이용하거나, 또는 각 숙주의 코돈 사용 빈도를 보여주는 문서를 참조할 수 있다. 획득한 데이터와 도입하고자 하는 DNA 서열을 참조하면, DNA 서열에 사용된 코돈들 중에서 해당 숙주에서 사용 빈도가 낮은 코돈은 동일한 아미노산을 코딩하고 사용 빈도가 높은 코돈으로 전환될 수 있

다.

[0727] 핵산 서열 인식 모듈 및/또는 핵산 염기 전환 효소를 코딩하는 DNA를 함유하는 발현 벡터는, 예를 들어, 적합한 발현 벡터에서 프로모터의 하류에 DNA를 연결함으로써 생성될 수 있다.

[0728] 발현 벡터로서, 에스케리치아 콜라이(*Escherichia coli*)-유래 플라스미드(예를 들어, pBR322, pBR325, pUC12, pUC13); 바실러스 서브틸리스(*Bacillus subtilis*) 유래-플라스미드(예를 들어, pUB110, pTP5, pC194); 효모-유래 플라스미드(예를 들어, pSH19, pSH15); 곤충 세포 발현 플라스미드(예를 들어, pFast-Bac); 동물 세포 발현 플라스미드(예를 들어, pA1-11, pXT1, pRc/CMV, pRc/RSV, pcDNAI/Neo); 람다 파지 등과 같은, 박테리오파지; 바큘로바이러스 등과 같은, 곤충 바이러스 벡터(예를 들어, BmNPV, AcNPV); 레트로바이러스, 백시니아 바이러스, 아데노바이러스 등과 같은, 바이러스 벡터 등이 사용된다.

[0729] 프로모터로서, 유전자 발현에 사용되는 숙주에 적합한 임의의 프로모터를 사용할 수 있다. DSB를 이용하는 종래의 방법에서는, 때때로 독성으로 인해 숙주 세포의 생존율이 현저하게 감소하는 경우가 있기 때문에, 유도성 프로모터를 이용하여 유도 개시까지 세포 수를 늘리는 것이 바람직할 수 있다. 그러나, 본 발명의 핵산-변형 효소 복합체를 발현시켜도 충분한 세포 증식이 가능하기 때문에, 구성 프로모터도 제한없이 사용할 수 있다.

[0730] 예를 들어, 숙주가 동물 세포인 경우, SR.알파. 프로모터, SV40 프로모터, LTR 프로모터, CMV(사이토메갈로 바이러스) 프로모터, RSV(Rous 육종 바이러스) 프로모터, MoMuLV(몰로니 마우스 백혈병 바이러스) LTR, HSV-TK(단순 헤르페스 바이러스 티미딘 키나제) 프로모터 등을 사용한다. 이들 중 CMV 프로모터, SR.알파 프로모터 등이 바람직하다. 한 실시형태에서, 프로모터는 CMV 프로모터 또는 SR 알파 프로모터이다. 숙주 세포가 에스케리치아 콜라이(*Escherichia coli*)인 경우, 다음 프모모터 중 어느 하나를 사용할 수 있다: trp 프로모터, lac 프로모터, recA 프로모터, 람다.P_L 프로모터, lpp 프로모터, T7 프로모터 등. 숙주가 바실러스(*Bacillus*) 속인 경우, 다음 프모모터 중 어느 하나를 사용할 수 있다: SPO1 프로모터, SPO2 프로모터, penP 프로모터 등. 숙주가 효모인 경우, 다음 프모모터 중 어느 하나를 사용할 수 있다: Gal1/10 프로모터, PHO5 프로모터, PGK 프로모터, GAP 프로모터, ADH 프로모터 등. 숙주가 곤충 세포인 경우, 다음 프모모터 중 어느 하나를 사용할 수 있다: 폴리헤드린 프로모터, P10 프로모터 등. 숙주가 식물 세포인 경우, 다음 프모모터 중 어느 하나를 사용할 수 있다: CaMV35S 프로모터, CaMV19S 프로모터, NOS 프로모터 등.

[0731] 일부 실시형태에서, 발현 벡터는 인핸서, 스플라이싱 신호, 종결자, 폴리A 부가 신호, 선택 마커, 예컨대 약물 내성 유전자, 영양요구성 상보 유전자 등, 복제 기원 등을 필요에 따라 함유할 수 있다.

[0732] 본 명세서에 기술된 단백질 도메인을 코딩하는 RNA는, 예를 들어, 상기 언급 된 핵산 서열-인식 모듈 및/또는 핵산 염기 전환 효소를 코딩하는 DNA를 코딩하는 벡터를 주형으로 사용하여 그 자체로 공지된 시험관 전사 시스템에서 mRNA로의 전사에 의해 제조될 수 있다.

[0733] 본 발명의 융합 단백질은 융합 단백질을 코딩하는 발현 벡터를 숙주 세포에 도입하고 숙주 세포를 배양함으로써 발현될 수 있다. 본 발명에 유용한 숙주 세포는 박테리아 세포, 효모, 곤충 세포, 포유류 세포 등을 포함한다.

[0734] 에스케리치아(*Escherichia*) 속에는 에스케리치아 콜라이(*Escherichia coli*) K12.cndot.DH1(Proc. Natl. Acad. Sci. USA, 60, 160 (1968)], 에스케리치아 콜라이(*Escherichia coli*) JM103(Nucleic Acids Research, 9, 309 (1981)], 에스케리치아 콜라이(*Escherichia coli*) JA221(Journal of Molecular Biology, 120, 517 (1978)], 에스케리치아 콜라이(*Escherichia coli*) HB101(Journal of Molecular Biology, 41, 459 (1969)], 에스케리치아 콜라이(*Escherichia coli*) C600(Genetics, 39, 440 (1954)] 등을 포함한다.

[0735] 바실러스(*Bacillus*) 속은 바실러스 서브틸리스(*Bacillus subtilis*) M1114(Gene, 24, 255 (1983)], 바실러스 서브틸리스(*Bacillus subtilis*) 207-21(Journal of Biochemistry, 95, 87 (1984)] 등을 포함한다.

[0736] 본 발명의 융합 단백질을 발현하는 데 유용한 효모는 사카로미세스 세레비시애(사카로미세스 세레비시애(Saccharomyces cerevisiae)) AH22, AH22R.sup.-, NA87-11A, DKD-5D, 20B-12, 스키조사카로미세스 폼페(*Schizosaccharomyces pombe*) NCYC1913, NCYC2036, 피치아 파스토리스(*Pichia pastoris*) KM71 등을 포함한다.

[0737] 융합 단백질은, 예를 들어, AcNPV와 같은, 바이러스 벡터를 사용하여 곤충 세포에서 발현된다. 곤충 숙주 세포는 다음 세포주 중 어느 하나를 포함한다: 도둑나방(cabbage armyworm) 유충-유래 확립 계통(스포돕테라 프루기퍼드(*Spodoptera frugiperd*) 세포; SF 세포), 트리초플루시아니(*Trichoplusiani*)의 중간 내장(mid-intestine)에서 유래된 MG1 세포, 트리초플루시아니(*Trichoplusiani*)의 알에서 유래된 High Five.TM. 세포, 마메스트라 브래스시카에(*Mamestra brassicae*)-유래 세포, 에스티그메나 아크레아(*Estigmena acrea*)-유래 세포

등이 사용된다. 바이러스가 BmNPV인 경우, 봄빅스 모리(*Bombyx mori*) 유래 확립된 계통(봄빅스 모리(*Bombyx mori*) N 세포; BmN 세포) 등의 세포가 곤충 세포로 사용된다. Sf 세포로서, 예를 들어, Sf9 세포(ATCC CRL1711), Sf21 세포(상기 모두, In Vivo, 13, 213-217 (1977)] 등.

[0738] 곤충으로서, 예를 들어, 봄빅스 모리(*Bombyx mori*), 드로소필라(*Drosophila*), 크리켓 등의 유충이 융합 단백질을 발현하는데 사용된다(Nature, 315, 592 (1985)).

[0739] 포유류 세포주를 사용하여 융합 단백질을 발현할 수 있다. 이러한 세포주는 원숭이 COS-7 세포, 원숭이 Vero 세포, 차이니즈 햄스터 난소(CHO) 세포, dhfr 유전자-결핍 CHO 세포, 마우스 L 세포, 마우스 AtT-20 세포, 마우스 골수종 세포, 랫트 GH3 세포, 인간 FL 세포 등을 포함한다. 인간 및 기타 포유류의 iPS 세포, ES 세포 등과 같은 만능 줄기 세포, 및 다양한 조직으로부터 제조된 1차 배양 세포가 사용된다. 또한, 제브라피쉬 배아, 제노푸스(*Xenopus*) 난모세포 등도 사용할 수 있다.

[0740] 식물 세포는 숙련된 기술자에게 잘 알려진 방법을 사용하여 배양 상태로 유지될 수 있다. 식물 세포 배양은 다양한 식물(예를 들어, 쌀, 밀, 옥수수 등과 같은 곡물, 토마토, 오이, 가지, 카네이션, 유스토마 러셀리아늄(*Eustoma russellianum*), 담배, 아라비돕시스 탈리아나(*Arabidopsis thaliana*)와 같은 산물 작물)으로부터 제조된, 배양 세포, 캘러스, 원형질체, 잎 세그먼트, 뿌리 세그먼트 등을 현탁시키는 단계를 포함한다.

[0741] 상기 언급된 모든 숙주 세포는 반수체(monoploid) 또는 배수체(예를 들어, 2 배체, 3 배체, 4 배체 등)일 수 있다. 기존의 돌연변이 도입 방법에서는, 원칙적으로 하나의 상동 염색체에만 돌연변이를 도입하여 이종 유전자 유형을 생성한다. 따라서, 우성 돌연변이가 발생하지 않는 한 원하는 표현형이 발현되지 않으며, 동형접합(homozygousness)은 불편하게 노동력과 시간을 필요로 한다. 이에 반해, 본 발명에 따르면, 게놈 내 상동 염색체상의 모든 대립 유전자에 돌연변이를 도입할 수 있기 때문에, 심지어 열성 돌연변이의 경우에도 원하는 표현형을 단일 세대로 발현시킬 수 있으며, 이는 기존 방법의 문제점을 해결할 수 있기 때문에 매우 유용하다.

[0742] 본 발명의 융합 단백질을 코딩하는 발현 벡터는 (예를 들어, 리소자임 방법, 적격(competent) 방법) PEG 방법, $CaCl_2$ 공침 방법, 전기천공 방법, 미세주입 방법, 입자 총 방법, 리포펙션 방법, 아그로박테리움(*Agrobacterium*) 방법 등을 사용하는) 임의의 형질감염 방법을 사용하여 숙주 세포에 도입된다. 형질감염 방법은 형질감염될 숙주 세포를 기준으로 선택된다.

[0743] 에스케리치아 콜라이(*Escherichia coli*)는, 예를 들어, 문헌[Proc. Natl. Acad. Sci. USA, 69, 2110 (1972), Gene, 17, 107 (1982)] 등에 기재된 방법에 따라 형질전환된다. 바실러스(*Bacillus*) 속은, 예를 들어, 문헌[Molecular & General Genetics, 168, 111 (1979)] 등에 기재되어 있다. 효모 세포는, 예를 들어, 문헌[Methods in Enzymology, 194, 182-187 (1991), Proc. Natl. Acad. Sci. USA, 75, 1929 (1978)] 등에 기재된 방법에 따라 벡터 내로 도입될 수 있다. 곤충 세포는, 예를 들어, 문헌[Bio/Technology, 6, 47-55 (1988)] 등에 기재된 방법에 따라 벡터 내로 도입될 수 있다. 포유류 세포는, 예를 들어, 문헌[Cell Engineering additional volume 8, New Cell Engineering Experiment Protocol, 263-267 (1995) (Shujunsha에 의해 출간됨)], 및 문헌[Virology, 52, 456 (1973)]에 기재된 방법에 따라 벡터 내로 도입될 수 있다.

[0744] 본 발명의 발현 벡터를 포함하는 세포는, 숙주에 따라 달라지는, 공지된 방법에 따라 배양된다. 예를 들어, 에스케리치아 콜라이(*Escherichia coli*) 또는 바실러스(*Bacillus*) 속을 배양할 때는, 배양에 사용하는 배지로는 액체 배지가 바람직할 수 있다. 배지는 바람직하게는 형질전환체의 성장에 필요한 탄소원, 질소원, 무기 물질 등을 포함한다. 탄소원의 예는 포도당, 덱스트린, 가용성 전분, 수크로스 등을 포함하고; 질소원의 예로는 암모늄염, 질산염, 옥수수 침지액, 펩톤, 카제인, 육류 추출물, 대두 케이크, 감자 추출물 등과 같은 무기 또는 유기 물질; 무기 물질의 예로는 염화칼슘, 인산이수소 나트륨, 염화 마그네슘 등을 포함한다. 배지는 효모 추출물, 비타민, 성장 촉진 인자 등을 포함할 수 있다. 배지의 pH는 바람직하게는 약 5 내지 약 8이다.

[0745] 에스케리치아 콜라이(*Escherichia coli*) 배양용 배지로는, 예를 들어, 포도당, 카사미노산(casamino acid)을 함유하는 M9 배지[Journal of Experiments in Molecular Genetics, 431-433, Cold Spring Harbor Laboratory, New York 1972]가 바람직할 수 있다. 필요한 경우, 예를 들어 3.베타.-인돌일아르릴산과 같은 작용제를 배지에 첨가하여 프로모터의 효율적인 기능을 보장할 수 있다. 에스케리치아 콜라이(*Escherichia coli*)는 일반적으로 약 15 내지 약 43℃에서 배양한다. 필요한 경우, 에어레이션 및 교반을 수행할 수 있다.

[0746] 바실러스(*Bacillus*) 속은 일반적으로 약 30 내지 약 40℃에서 배양한다. 필요한 경우, 에어레이션 및 교반을 수행한다.

[0747] 효모 배양을 위한 배지의 예는 버크홀더(*Burkholder*) 최소 배지(Proc. Natl. Acad. Sci. USA, 77, 4505 (1980)], 0.5% 카사미노산을 함유하는 SD 배지(Proc. Natl. Acad. Sci. USA, 81, 5330 (1984)] 등을 포함한다. 배지의 pH는 바람직하게는 약 5 내지 약 8이다. 배양은 일반적으로 약 20℃ 내지 약 35℃에서 수행된다. 필요한 경우, 폭기 및 교반을 수행할 수 있다.

[0748] 곤충 세포 또는 곤충을 배양하기 위한 배지로는, 예를 들어, 필요한 경우, 불활성화된 10% 소 혈청 등의 첨가제 등을 함유한 그레이스(Grace)의 곤충 배지(Nature, 195, 788 (1962)] 등이 사용된다. 배지의 pH는 바람직하게는 약 6.2 내지 약 6.4이다. 배양은 일반적으로 약 27℃에서 수행된다. 필요한 경우, 에어레이션 및 교반을 수행할 수 있다.

[0749] 동물 세포 배양을 위한 배지로, 예를 들어, 약 5 내지 약 20%의 소 태아 혈청을 함유하는 최소 필수 배지(MEM)(Science, 122, 501 (1952)], 둘베코 변형 이글 배지(DMEM)(Virology, 8, 396 (1959)], RPMI 1640 배지(The Journal of the American Medical Association, 199, 519 (1967)], 199 배지(Proceeding of the Society for the Biological Medicine, 73, 1 (1950)] 등이 사용된다. 배지의 pH는 바람직하게는 약 6 내지 약 8이다. 배양은 일반적으로 약 30℃ 내지 약 40℃에서 수행된다. 필요한 경우, 에어레이션 및 교반을 수행할 수 있다.

[0750] 식물 세포 배양용 배지로는, 예를 들어, MS 배지, LS 배지, B5 배지 등이 사용된다. 배지의 pH는 바람직하게는 약 5 내지 약 8이다. 배양은 일반적으로 약 20 ℃ 내지 약 30 ℃에서 수행된다. 필요한 경우, 에어레이션 및 교반을 수행할 수 있다.

[0751] 동물 세포, 곤충 세포, 식물 세포 등과 같은 고등 진핵 세포를 숙주 세포로 사용하는 경우, 본 발명의 염기 편집 시스템을 코딩하는 DNA는 유도성 프로모터(예를 들어, 메탈로티오네인 프로모터(중금속 이온에 의해 유도됨), 열 충격 단백질 프로모터(열 충격에 의해 유도됨), Tet-ON/Tet-OFF 시스템 프로모터(테트라사이클린 또는 이의 유도체의 첨가 또는 제거에 의해 유도됨), 스테로이드-반응성 프로모터(스테로이드 호르몬 또는 이의 유도체에 의해 유도됨) 등)의 조절 하에 숙주 세포에 도입되고, 유도 물질을 적절한 단계에서 배지에 첨가(또는 배지로부터 제거)하여 핵산-변형 효소 복합체의 발현을 유도하고, 배양을 주어진 기간 동안 수행하여 염기 편집을 진행하고, 표적 유전자에 돌연변이를 도입하면 염기 편집 시스템의 일시적인 발현을 구현할 수 있다.

[0752] 에스케리치아 콜라이(*Escherichia coli*) 등과 같은 원핵 세포는 유도성 프로모터를 이용할 수 있다. 유도성 프로모터의 예는, 이로만 제한되는 것은 아니지만, lac 프로모터(IPTG에 의해 유도됨), cspA 프로모터(콜드 쇼크에 의해 유도됨), araBAD 프로모터(아라비노스에 의해 유도됨) 등을 포함한다.

[0753] 대안적으로, 상술한 유도성 프로모터는 동물 세포, 곤충 세포, 식물 세포 등과 같은 고등 진핵 세포가 숙주 세포로 사용되는 경우 벡터 제거 메커니즘으로 사용될 수도 있다. 즉, 벡터에는 숙주 세포에서 기능하는 복제 기점과 복제에 필요한 단백질(예를 들어, 동물 세포의 경우, SV40 on 및 Large T 항원, oriP 및 EBNA-1 등)을 코딩하는 핵산이 탑재되고, 단백질을 코딩하는 핵산의 발현은 상기 언급된 유도성 프로모터에 의해 조절된다. 그 결과 벡터는 유도 물질 존재 하에서 자율적으로 복제가능하게 되나, 유도 물질이 제거되면, 자율적 복제가 불가능하게 되며, 벡터는 세포분열에 따라 자연적으로 유실되게 된다(자율 복제는 Tet-OFF 시스템 벡터에서 테트라사이클린과 독시사이클린의 첨가에 의해 불가능해짐).

[0754] **전달 시스템**

[0755] 핵염기 편집기 및 gRNA의 핵산-기반 전달

[0756] 본 개시에 따른 염기 편집 시스템(예를 들어, 다중-이펙터 핵염기 편집기)를 코딩하는 핵산은 대상체에게 투여되거나 당업계에 공지된 방법에 의해 또는 본 명세서에 기재된 바와 같이 시험관내(*in vitro*) 또는 생체내에서(*in vivo*) 세포로 전달될 수 있다. 예를 들어, 다중-이펙터 핵염기 편집기는, 예를 들어, 벡터(예를 들어, 바이러스 또는 비-바이러스 벡터), 비-벡터 기반 방법(예를 들어, 네이키드 DNA, DNA 복합체, 지질 나노입자를 이용), 또는 이들의 조합에 의해 전달될 수 있다.

[0757] 핵염기 편집기 또는 다중-이펙터 핵염기 편집기를 코딩하는 핵산은, 예를 들어, 형질감염 또는 또는 전기천공을 통해, 네이키드 DNA 또는 RNA로서 세포(예를 들어, 조혈 세포 또는 이들의 전구체, 조혈 줄기 세포, 및/또는 유도 만능 줄기 세포)에 직접 전달될 수 있거나, 또는 표적 세포에 의한 흡수를 촉진하는 분자(예를 들어, N-아세틸갈락토사민)에 접합될 수 있다. 본 명세서에 기술된 벡터와 같은, 핵산 벡터가 또한 사용될 수 있다.

[0758] 핵산 벡터는 본 명세서에 기재된 융합 단백질의 도메인을 코딩하는 하나 이상의 서열을 포함할 수 있다. 벡터는 또한 단백질을 코딩하는 서열과 연관된(예를 들어, 삽입되거나 융합된), 신호 펩티드(예를 들어, 핵 위치결정,

핵소체(nucleolar) 위치결정, 또는 미토콘드리아 위치결정을 위한 것)를 코딩하는 단일 서열을 포함할 수 있다. 일 예로서, 핵산 벡터는 하나 이상의 핵 위치결정 서열(예를 들어, SV40로부터의 핵 위치결정 서열), 및 데아미나제(아데노신 데아미나제 및/또는 시티딘 데아미나제)를 포함하는 Cas9 코딩 서열을 포함할 수 있다.

[0759] 핵산 벡터는 또한 임의의 적합한 수의 조절/제어 요소, 예를 들어, 프로모터, 인핸서, 인트론, 폴리아데닐화 신호, 코작(Kozak) 컨센서스 서열, 또는 내부 리보솜 진입 부위(IRES)를 포함할 수 있다. 이러한 요소는 당업계에 잘 알려져 있다. 조혈 세포의 경우, 적합한 프로모터는 IFN 베타 또는 CD45를 포함할 수 있다.

[0760] 본 개시에 따른 핵산 벡터는 재조합 바이러스 벡터를 포함한다. 예시적인 바이러스 벡터는 본 명세서에 설명되어 있다. 당업계에 알려진 다른 바이러스 벡터를 또한 사용할 수 있다. 또한, 바이러스 입자를 사용하여 핵산 및/또는 펩티드 형태로 염기 편집 시스템 구성요소를 전달할 수 있다. 예를 들어, "빈(empty)" 바이러스 입자는 임의의 적합한 화물을 함유하도록 조립될 수 있다. 바이러스 벡터 및 바이러스 입자는 표적 조직 특이성을 변경하기 위해 표적화 리간드를 통합하도록 조작될 수도 있다.

[0761] 바이러스 벡터 이외에, 비-바이러스 벡터가 본 개시에 따른 게놈 편집 시스템을 코딩하는 핵산을 전달하는 데 이용될 수 있다. 비-바이러스 핵산 벡터의 중요한 범주 중 하나는, 유기성 또는 무기성일 수 있는, 나노 입자이다. 나노 입자는 당업계에 잘 알려져 있다. 게놈 편집 시스템 구성요소 또는 이러한 구성요소를 코딩하는 핵산을 전달하기 위해 임의의 적합한 나노 입자 디자인을 이용할 수 있다. 예를 들어, 유기(예를 들어, 지질 및/또는 중합체) 나노 입자는 본 개시의 특정 실시형태에서 전달 비히클로서 사용하기에 적합할 수 있다. 나노 입자 제형 및/또는 유전자 전달에 사용하기 위한 예시적인 지질은 **표 6**(아래)에 제시되어 있다.

[0762] **표 6**

[0763] 유전자 전달에 사용되는 지질

지질	약어	특징
1,2-다이올레오일-sn-글리세로-3-포스파티딜콜린	DOPC	헬퍼
1,2-다이올레오일-sn-글리세로-3-포스파티딜에탄올아민	DOPE	헬퍼
콜레스테롤		헬퍼
N-[1-(2,3-다이올레일옥시)프로필]N,N,N-트리메틸암모늄 클로라이드	DOTMA	양이온성
1,2-다이올레오일옥시-3-트리메틸암모늄-프로판	DOTAP	양이온성
다이옥타데실아미도글리실스퍼민	DOGS	양이온성
N-(3-아미노프로필)-N,N-다이메틸-2,3-비스(도데실옥시)-1-프로판아미니움 브로마이드	GAP-DLRIE	양이온성
세틸트리메틸암모늄 브로마이드	CTAB	양이온성
6-라우르옥시헥실 오르니티네이트	LHON	양이온성
1-(2,3-다이올레오일옥시프로필)-2,4,6-트리메틸피리디니움	2Oc	양이온성
2,3-다이올레일옥시-N-[2(스퍼민카복사미도-에틸]-N,N-다이메틸-1-프로판아미니움 트리플루오로아세테이트	DOSPA	양이온성
1,2-다이올레일-3-트리메틸암모늄-프로판	DOPA	양이온성
N-(2-하이드록시에틸)-N,N-다이메틸-2,3-비스(테트라데실옥시)-1-프로판아미니움 브로마이드	MDRIE	양이온성
다이미리스토옥시프로필 다이메틸 하이드록시에틸 암모늄브로마이드	DMRI	양이온성
3β-[N-(N',N'-다이메틸아미노에탄)-카바모일]콜레스테롤	DC-Chol	양이온성
비스-구아니디움-트렌-콜레스테롤	BGTC	양이온성
1,3-다이오데옥시-2-(6-카복시-스퍼밀)-프로필아미드	DOSPER	양이온성
다이메틸옥타데실암모니움 브로마이드	DDAB	양이온성
다이옥타데실아미도글리실스퍼미딘	DSL	양이온성
rac-[(2,3-다이옥타데실옥시프로필)(2-하이드록시에틸)]-다이메틸암모니움 클로라이드	CLIP-1	양이온성
rac-[2(2,3-다이헥사데실옥시프로필-옥시메틸옥시)에틸]트리메틸암모니움 브로마이드	CLIP-6	양이온성
에틸다이미리스토일포스파티딜콜린	EDMPC	양이온성
1,2-다이스테아릴옥시-N,N-다이메틸-3-아미노프로판	DSDMA	양이온성
1,2-다이미리스토일-트리메틸암모늄 프로판	DMTAP	양이온성
O,O'-다이미리스틸-N-일실 아스파테이트	DMKE	양이온성
1,2-다이스테아로일-sn-글리세로-3-에틸포스포콜린	DSEPC	양이온성

[0764]

유전자 전달에 사용되는 지질

지질	약어	특징
N-팔미토일 D-에리쓰로-스핑고실 카바모일-스퍼민	CCS	양이온성
N-t-부틸-N0-테트라데실-3-테트라데실아미노프로피온아미딘	다이 C14-아미딘	양이온성
옥타데세노일옥시[에틸-2-헵타데세닐-3 하이드록시에틸] 이미다졸리니움 클로라이드	DOTIM	양이온성
N1 -콜레스테릴옥시카보닐-3,7-디아자노난-1,9-다이아민	CDAN	양이온성
2-(3-[비스(3-아미노-프로필)-아미노]프로필아미노)-N-다이테트라데실카바모일메틸-아세트아미드	RPR209120	양이온성
1,2-다이리놀레일옥시-3-다이메틸아미노프로판	DLinDMA	양이온성
2,2-다이리놀레일-4-다이메틸아미노에틸-[1,3]-다이옥솔란	DLin-KC2-DMA	양이온성
다이리놀레일-메틸-4-다이메틸아미노부티레이트	DLin-MC3-DMA	양이온성

[0765]

[0766] 표 7은 유전자 전달 및/또는 나노 입자 제형에 사용하기 위한 예시적인 중합체를 열거한다.

[0767] 표 7

유전자 전달에 사용되는 중합체

중합체	약어
폴리(에틸렌)글리콜	PEG
폴리에틸렌이민	PEI
다이티오비스 (석신이미딜프로피오네이트)	DSP
다이메틸-3,3'-다이티오비스프로피온이미데이트	DTBP
폴리(에틸렌 이민)비스카바메이트	PEIC
폴리(L-라이신)	PLL
히스티딘 개질된 PLL	
폴리(N-비닐피롤리돈)	PVP
폴리(프로필렌이민)	PPI
폴리(아미도아민)	PAMAM
폴리(아미도에틸렌이민)	SS-PAEI
트리에틸렌테트라민	TETA
폴리(β-아미노에스터)	

[0768]

유전자 전달에 사용되는 중합체

중합체	약어
폴리(4-하이드록시-L-프롤린 에스터)	PHP
폴리(알릴아민)	
폴리(α-[4-아미노부틸]-L-글리콜산)	PAGA
폴리(D,L-락틱-co-글리콜산)	PLGA
폴리(N-에틸-4-비닐피리디니움 브로마이드)	
폴리(포스파젠)	PPZ
폴리(포스포에스터)	PPE
폴리(포스포르아미데이트)	PPA
폴리(N-2-하이드록시프로필메타크릴아미드)	pHPMA
폴리 (2-(다이메틸아미노)에틸 메타크릴레이트)	pDMAEMA
폴리(2-아미노에틸 프로필렌 포스페이트)	PPE-EA
키토산	
갈락토실화된 키토산	
N-도다실화된(Dodacylated) 키토산	
히스톤	
콜라겐	
덱스트란-스퍼민	D-SPM

[0769]

[0770] 표 8은 본 명세서에 기술된 융합 단백질을 코딩하는 폴리뉴클레오티드에 대한 전달 방법을 요약한다.

[0771] 표 8

전달	벡터/방식	비-분열 세포로의 전달	발현의 지속	게놈 통합	전달되는 분자의 유형
물리적	(예를 들어, 전기천공, 입자 총, 칼슘 포스페이트 형질감염	네	일시적	아니오	핵산 및 단백질
바이러스	레트로바이러스	아니오	안정적	네	RNA
	렌티바이러스	네	안정적	변형에 따라 네/아니오	RNA
	아데노바이러스	네	일시적	아니오	DNA
	아데노-관련 바이러스 (AAV)	네	안정적	아니오	DNA
	백시니아 바이러스	네	매우 일시적	아니오	DNA
	헤르페스 심플렉스 바이러스	네	안정적	아니오	DNA
비-바이러스	양이온성 리포좀	네	일시적	전달되는 것에 따라 다름	핵산 및 단백질
	중합체성 나노입자	네	일시적	전달되는 것에 따라 다름	핵산 및 단백질
생물학적 비-바이러스 전달 비히클	약화된 박테리아	네	일시적	아니오	핵산
	조작된 박테리오파지	네	일시적	아니오	핵산
	포유류 바이러스-유사 입자	네	일시적	아니오	핵산
	생물학적 리포좀: 파열 적혈구 및 엑소좀	네	일시적	아니오	핵산

[0772]

[0773]

[0774] 또 다른 양상에서, 게놈 편집 시스템 구성요소 또는 이러한 구성요소를 코딩하는 핵산, 예를 들어, Cas9 또는 이의 변이체와 같은 핵산 결합 단백질, 및 관심있는 게놈 핵산 서열을 표적화하는 gRNA의 전달은 리보핵단백질(RNP)을 세포에 전달함으로써 달성된다. RNP는 표적화 gRNA와 복합체로 핵산 결합 단백질, 예를 들어, Cas9를 포함한다. RNP는 전기천공, 뉴클레오펙션, 또는 예를 들어, 문헌[Zuris, J.A. *et al.*, 2015, *Nat. Biotechnology*, 33(1):73-80]에 보고된 것과 같은, 양이온성 지질-매개 방법과 같은, 공지된 방법을 사용하여 세포에 전달될 수 있다. RNP는 CRISPR 염기 편집 시스템, 특히 1차 세포와 같이 형질감염이 어려운 세포에 사용하는 데 유리하다. 또한, RNP는 특히 CRISPR 플라스미드에 사용될 수 있는, CMV 또는 EF1A와 같은, 진핵 프로모터가 잘 발현되지 않을 때, 세포에서 단백질 발현으로 발생할 수 있는 어려움을 완화할 수 있다. 유리하게는, RNP의 사용은 외래 DNA의 세포로의 전달을 필요로 하지 않는다. 더욱이, 핵산 결합 단백질과 gRNA 복합체를 포함하는 RNP는 시간이 지남에 따라 분해되기 때문에, RNP의 사용은 표적-이탈 효과를 제한할 가능성이 있다. 플라스미드 기반 기술과 유사한 방식으로, RNP를 사용하여 결합 단백질(예를 들어, Cas9 변이체)을 전달하고 상동성 직접 복구(HDR)를 유도할 수 있다.

[0775] 핵산 분자 발현을 코딩하는 염기 편집기를 구동하는 데 사용되는 프로모터는 AAV ITR을 포함할 수 있다. 이는, 벡터에서 공간을 차지할 수 있는, 추가 프로모터 요소의 필요성을 제거하는 데 유리할 수 있다. 비워진 추가 공간은 가이드 핵산 또는 선택가능한 마커와 같은 추가 요소의 발현을 유도하는 데 사용될 수 있다. ITR 활성은 상대적으로 약하기 때문에, 선택된 뉴클레아제의 과발현으로 인한 잠재적 독성을 감소시키는 데 사용할 수 있다.

[0776] 임의의 적절한 프로모터를 사용하여 염기 편집기, 및 필요한 경우 가이드 핵산의 발현을 유도할 수 있다. 편재성(ubiquitous) 발현의 경우, 사용할 수 있는 프로모터는 CMV, CAG, CBh, PGK, SV40, 페리틴 중쇄 또는 경쇄 등을 포함한다. 뇌 또는 기타 CNS 세포 발현의 경우, 적합한 프로모터는 다음을 포함할 수 있다: 모든 뉴런에 대한 시냅신(Synapsin)I, 흥분성 뉴런의 경우 CaMKII알파 프로모터, GABA성 뉴런의 경우 GAD67 또는 GAD65 또는 VGAT 등. 간 세포 발현의 경우, 적합한 프로모터는 알부민 프로모터를 포함한다. 폐 세포 발현의 경우, 적합한 프로모터는 SP-B 프로모터를 포함할 수 있다. 내피 세포의 경우, 적합한 프로모터는 ICAM 프로모터를 포함할 수 있다. 조혈 세포의 경우, 적합한 프로모터는 IFN 베타 또는 CD45 프로모터를 포함할 수 있다. 조골 세포의 경우, 적합한 프로모터는 OG-2를 포함할 수 있다.

[0777] 일부 실시형태에서, 본 개시의 염기 편집기는 별개의 프로모터가 동일한 핵산 분자 내에서 염기 편집기 및 호환가능한 가이드 핵산의 발현을 유도하도록 하기에 충분히 작은 크기이다. 예를 들어, 벡터 또는 바이러스 벡터는 염기 편집기를 코딩하는 핵산에 작동가능하게 연결된 제1 프로모터 및 가이드 핵산에 작동가능하게 연결된 제2 프로모터를 포함할 수 있다.

[0778] 가이드 핵산의 발현을 유도하는 데 사용되는 프로모터는 다음을 포함할 수 있다: U6 또는 H1과 같은 Pol III 프로모터, Pol II 프로모터 및 인트론 카세트를 사용하여 gRNA 아데노-관련 바이러스(AAV)를 발현.

[0779] *바이러스 벡터*

[0780] 따라서, 본 명세서에 기재된 염기 편집기는 바이러스 벡터와 함께 전달될 수 있다. 일부 실시형태에서, 본 명세서에 개시된 염기 편집기는 바이러스 벡터에 함유된 핵산 상에서 코딩될 수 있다. 일부 실시형태에서, 염기 편집기 시스템의 하나 이상의 구성요소는 하나 이상의 바이러스 벡터 상에서 코딩될 수 있다. 예를 들어, 염기 편집기 및 가이드 핵산은 단일 바이러스 벡터에 코딩될 수 있다. 다른 실시형태에서, 염기 편집기 및 가이드 핵산은 상이한 바이러스 벡터 상에서 코딩된다. 어느 경우든, 염기 편집기 및 가이드 핵산은 각각 프로모터 및 터미네이터에 작동가능하게 연결될 수 있다. 바이러스 벡터에 코딩된 구성요소의 조합은 선택한 바이러스 벡터의 화물 크기 제한에 의해 결정될 수 있다.

[0781] 염기 편집기의 전달을 위한 RNA 또는 DNA 바이러스 기반 시스템의 사용은 바이러스를 배양 또는 숙주의 특정 세포로 표적화하고 바이러스 페이로드를 핵 또는 숙주 세포 게놈으로 트래피킹하는 고도로 진화된 프로세스를 활용한다. 바이러스 벡터는 배양된 세포로, 환자(생체내에서)에 직접 투여될 수 있거나, 이들은 시험관내에서 세포 치료에 사용될 수 있으며, 변형된 세포는 선택적으로 환자에게 투여될 수 있다(생체외에서). 기존의 바이러스 기반 시스템은 유전자 전달을 위한 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-관련 및 단순 헤르페스 바이러스 벡터를 포함할 수 있다. 숙주 게놈 내의 통합은 레트로바이러스, 렌티바이러스, 및 아데노-관련 바이러스 유전자 전달 방법으로 가능하며, 종종 삽입된 도입유전자의 장기 발현을 초래한다. 또한, 많은 상이한 세포 유형 및 표적 조직에서 높은 형질도입 효율이 관찰되었다.

[0782] 바이러스 벡터는 렌티바이러스(예를 들어, HIV 및 FIV 기반 벡터), 아데노바이러스(예를 들어, AD100), 레트로바이러스(예를 들어, 말로니 뮤린 백혈병 바이러스, MML-V), 헤르페스바이러스 벡터(예를 들어, HSV-2) 및 및 아데노 관련 바이러스(AAV), 또는 기타 플라스미드 또는 바이러스 벡터 유형, 특히, 예를 들어, 미국 특허 제8,454,972호(아데노바이러스에 대한 제형, 용량), 미국 특허 제8,404,658호(AAV에 대한 제형, 용량) 및 미국 특허 제5,846,946호(DNA 플라스미드에 대한 제형, 용량)로부터 그리고 렌티바이러스, AAV 및 아데노바이러스와 관련하는 임상 시험에 관한 임상 시험 및 간행물로부터의 제형 및 용량을 이용하는 것을 포함할 수 있다. 예를 들어, AAV의 경우, 투여 경로, 제형 및 용량은 미국 특허 제8,454,972호에 기재된 것과 같을 수 있고 아데노바이러스와 관련하는 임상 시험에서와 같을 수 있다. 플라스미드 전달의 경우, 투여 경로, 제형 및 용량은 미국 특허 제5,846,946호에서와 같을 수 있고 플라스미드에 관련하는 임상 연구에서와 같을 수 있다. 용량은 평균 70kg 개체(예를 들어, 성인 남성)을 기준으로 하거나 추정할 수 있으며, 환자, 대상체, 다양한 체중 및 종의 포유동물에 대해 조정할 수 있다. 투여 빈도는 환자 또는 대상체의 연령, 성별, 일반적인 건강, 기타 상태, 다루어지는 특정 상태 또는 증상을 포함한 일반적인 요인에 따라 의료 또는 수의 전문가(예를 들어, 의사, 수의사)의 영역 내에 있다. 바이러스 벡터는 관심 조직 내로 주입될 수 있다. 세포-유형 특이적 염기 편집의 경우, 염기 편집기 및 선택적 가이드 핵산의 발현은 세포-유형 특이적 프로모터에 의해 구동될 수 있다.

[0783] 레트로바이러스의 향성은 외래 외피 단백질을 통합함으로써 변경될 수 있으며, 표적 세포의 잠재적인 표적 집단을 확장시킬 수 있다. 렌티바이러스 벡터는 비 분열 세포를 형질도입하거나 감염시킬 수 있고, 일반적으로 높은 바이러스 역가를 생성할 수 있는 레트로바이러스 벡터이다. 따라서 레트로바이러스 유전자 전달 시스템의 선택은 표적 조직에 따라 달라진다. 레트로바이러스 벡터는 최대 6 내지 10 kb의 외래 서열에 대한 패키징 용량을

가진 시스-작용 긴 말단 반복부로 구성된다. 최소 시스-작용 LTR은 벡터의 복제 및 패키징에 충분하며, 이는 영구적인 도입유전자 발현을 제공하기 위해 치료 유전자를 표적 세포에 통합하는 데 사용된다. 널리 사용되는 레트로바이러스 벡터에는 뮤린 백혈병 바이러스(MuLV), 긴팔 원숭이 백혈병 바이러스(GaLV), 시미안 면역결핍 바이러스(SIV), 인간 면역 결핍 바이러스(HIV), 및 이들의 조합에 기반한 것들이 포함된다(예를 들어, 다음 문헌 참조: Buchscher *et al.*, J. Virol. 66:2731-2739 (1992); Johann *et al.*, J. Virol. 66:1635-1640 (1992); Sommnerfelt *et al.*, Virol. 176:58-59 (1990); Wilson *et al.*, J. Virol. 63:2374-2378 (1989); Miller *et al.*, J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

[0784] 레트로바이러스 벡터, 특히 렌티바이러스 벡터는, 표적 세포로의 효율적인 통합을 위해 주어진 길이보다 작은 폴리뉴클레오티드 서열을 필요로 할 수 있다. 예를 들어, 길이가 9kb보다 큰 레트로바이러스 벡터는 더 작은 크기에 비해 더 낮은 바이러스 역가를 초래할 수 있다. 일부 실시형태에서, 본 개시의 염기 편집기는 레트로바이러스 벡터를 통해 표적 세포로의 효율적인 패키징 및 전달을 가능하게하기에 충분한 크기이다. 일부 실시형태에서, 염기 편집기는 가이드 핵산 및/또는 표적화가능한 뉴클레아제 시스템의 다른 구성요소와 함께 발현되는 경우에도 효율적인 패킹 및 전달을 허용할 수 있는 크기이다.

[0785] 일시적인 발현이 선호되는 응용에서, 아데노바이러스 기반 시스템이 사용될 수 있다. 아데노바이러스 기반 벡터는 많은 세포 유형에서 매우 높은 형질도입 효율이 가능하며 세포 분열을 필요로 하지 않는다. 이러한 벡터로 높은 역가 및 발현 수준이 획득되었다. 이 벡터는 비교적 간단한 시스템에서 대량으로 생성될 수 있다. 아데노-관련 바이러스("AAV") 벡터는 또한, 예를 들어, 핵산 및 펩티드의 시험관내(*in vitro*) 생산에서, 그리고 생체내(*in vivo*) 및 생체외(*ex vivo*) 유전자 치료 절차에서 표적 핵산으로 세포를 형질도입하는 데 사용될 수 있다(예를 들어, 다음 문헌 참조: West *et al.*, Virology 160:38-47 (1987); 미국 특허 제4,797,368호; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994)). 재조합 AAV 벡터의 구축은, 다음 문헌을 포함하는, 다수의 간행물에 기술되어 있다: 미국 특허 제5,173,414호; Tratschin *et al.*, Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, *et al.*, Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); 및 Samulski *et al.*, J. Virol. 63:03822-3828 (1989).

[0786] AAV는 파보바이러스 패밀리에 속하는 작은 단일-가닥 DNA 의존성 바이러스이다. 4.7kb 야생형(wt) AAV 게놈은 각각 4개의 복제 단백질과 3개의 캡시드 단백질을 코딩하는 2개의 유전자로 구성되며, 각 측부 상에 145-bp 반전 말단 반복부(ITR)가 측접하고 있다. 비리온은 동일한 오픈 리딩 프레임에서, 그러나 차등 스플라이싱(Vp1) 및 대체 번역 개시 부위(각각, Vp2 및 Vp3)에서 1:1:10 비율로 생성되는, 3가지 캡시드 단백질인 Vp1, Vp2 및 Vp3로 구성된다. Vp3는 비리온에서 가장 풍부한 서브유닛이며 바이러스의 향성(tropism)을 정의하는 세포 표면에서의 수용체 인식에 참여한다. 바이러스 감염성에서 기여하는, 포스포리파제 도메인이 Vp1의 고유한 N 말단에서 확인되었다.

[0787] 야생형(wt) AAV와 유사하게, 재조합 AAV(rAAV)는 시스(*cis*)-작용 145-bp ITR을 이용하여 벡터 도입유전자(transgene) 카세트를 측접시켜, 외래 DNA의 패키징을 위해 최대 4.5kb를 제공한다. 감염에 뒤이어, rAAV는 본 발명의 융합 단백질을 발현할 수 있고 원형 머리-꼬리(head-to-tail) 콘카티머(concatemers)에 에피솜으로 존재함으로써 숙주 게놈으로 통합되지 않고 지속될 수 있다. 시험관내(*in vitro*) 및 생체내에서(*in vivo*), 이 시스템을 이용하는 rAAV 성공에 관한 다수의 예가 존재하지만, 제한된 패키징 용량은, 유전자의 코딩 서열의 길이가 야생형(wt) AAV 게놈 크기와 같거나 클 때, AAV-매개 유전자 전달의 사용을 제한시켰다.

[0788] 바이러스 벡터는 애플리케이션에 따라 선택될 수 있다. 예를 들어, 생체내(*in vivo*) 유전자 전달의 경우, AAV가 다른 바이러스 벡터보다 유리할 수 있다. 일부 실시형태에서, AAV는 낮은 독성을 허용하는데, 이는 면역 반응을 활성화할 수 있는 세포 입자의 초원심분리를 필요로 하지 않는 정제 방법 때문일 수 있다. 일부 실시형태에서, AAV는 숙주 게놈에 통합되지 않기 때문에 삽입 돌연변이 유발을 유발할 가능성이 낮다. 아데노바이러스는 그들이 유도하는 강한 면역원성 반응 때문에 일반적으로 백신으로 사용된다. 바이러스 벡터의 패키징 용량은 벡터로 패키징할 수 있는 염기 편집기의 크기를 제한할 수 있다.

[0789] AAV의 패키징 용량은 약 4.5Kb 또는 4.75Kb이며, 2개의 145개 염기의 역전 말단 반복부(ITR; inverted terminal repeats)을 포함한다. 이는 개시된 염기 편집기를 비롯하여 프로모터 및 전사 종결자가 단일 바이러스 벡터에 들어갈 수 있다는 것을 의미한다. 4.5 또는 4.75Kb보다 큰 구축물은 바이러스 생산을 크게 감소시킬 수 있다. 예를 들어, SpCas9는 상당히 크고, 유전자 자체가 4.1Kb 이상이므로, AAV 내로 패킹하기가 어렵다. 따라서, 본 개시의 실시형태는 종래의 염기 편집기보다 길이가 더 짧은 개시된 염기 편집기를 이용하는 것을 포함한다. 일부 예에서, 염기 편집기는 4kb 미만이다. 개시된 염기 편집기는 4.5kb, 4.4kb, 4.3kb, 4.2kb, 4.1kb,

4kb, 3.9kb, 3.8kb, 3.7kb, 3.6kb, 3.5kb, 3.4kb, 3.3kb, 3.2kb, 3.1kb, 3kb, 2.9kb, 2.8kb, 2.7kb, 2.6kb, 2.5kb, 2kb, 또는 1.5kb 미만일 수 있다. 일부 실시형태에서, 개시된 염기 편집기는 길이가 4.5kb 또는 그 미만이다.

[0790] AAV는 AAV1, AAV2, AAV5 또는 이들의 임의의 조합일 수 있다. 표적화할 세포와 관련하여 AAV의 유형을 선택할 수 있다. 예를 들어, 뇌 또는 신경 세포를 표적화하기 위해 AAV 혈청형 1, 2, 5 또는 하이브리드 캡시드 AAV1, AAV2, AAV5 또는 이들의 임의 조합을 선택할 수 있으며; 심장 조직을 표적으로 하기 위해 AAV4를 선택할 수 있다. AAV8은 간으로의 전달에 유용하다. 이들 세포에 대한 특정 AAV 혈청형의 표목록(tabulation)은 문헌[Grimm, D. *et al.*, J. Virol. 82: 5887-5911 (2008)]에서 찾을 수 있다.

[0791] 렌티바이러스는 유사 분열 및 유사 분열 후 세포 둘 다에서 자신의 유전자를 감염시키고 발현시키는 능력을 갖는 복합(complex) 레트로바이러스이다. 가장 일반적으로 알려진 렌티바이러스는 인간 면역 결핍 바이러스(HIV)로, 이것은 다른 바이러스의 외피 당단백질을 사용하여 광범위한 세포 유형을 표적으로 삼는다.

[0792] 렌티바이러스는 다음과 같이 제조할 수 있다. (렌티바이러스 전달 플라스미드 백본을 함유하는) pCasES10을 클로닝한 후, 낮은 계대(p = 5)에서 HEK293FT를 형질감염 전날 50% 컨플루언시(confluence)에서 항생제없이 10% 소 태아 혈청과 함께 DMEM이 있는 T-75 플라스크에 시딩한다. 20시간 후, 배지를 OptiMEM(무 혈청) 배지로 변경하고 4시간 후에 형질감염을 수행했다. 세포를 10 ㎍의 렌티바이러스 전달 플라스미드(pCasES10) 및 다음 패키징 플라스미드로 형질감염시킨다: 5 ㎍의 pMD2.G(VSV-g 유사형) 및 7.5 ㎍의 psPAX2(gag/pol/rev/tat). 형질감염은 양이온 성 지질 전달 작용제(50 ㎕ Lipofectamine 2000 및 100 ㎕ Plus 시약)를 사용하여 4 ㎖ OptiMEM에서 수행할 수 있다. 6시간 후, 배지를 10% 소 태아 혈청을 함유한 항생제가 없는 DMEM으로 교체한다. 이러한 방법은 세포 배양 중에 혈청을 사용하지만, 무-혈청 방법이 선호된다.

[0793] 렌티바이러스는 다음과 같이 정제할 수 있다. 바이러스성 상청액을 48시간 후 수확한다. 상청액을 먼저 이물질을 제거하고 0.45 ㎛ 저 단백질 결합(PVDF) 필터를 통해 여과시킨다. 그런 다음 이것을 24,000 rpm에서 2시간 동안 초원심분리기에서 회전시킨다. 바이러스 펠렛을 4℃에서 밤새 50 ㎕의 DMEM에 재현탁시킨다. 그런 다음, 분취한 후 즉시 -80℃에서 냉동한다.

[0794] 또 다른 실시형태에서, 말 감염성 빈혈 바이러스(EIAV)에 기초한 최소 비-영장류 렌티바이러스 벡터가 또한 고려된다. 또 다른 실시형태에서, RetinoStat.RTM, 혈관생성억제성(angiostatic) 단백질 엔도스타틴 및 안지오스타틴을 발현하는 말 감염성 빈혈 바이러스-기반 렌티바이러스 유전자 치료 벡터가 망막하 주사를 통해 전달되는 것으로 고려된다. 또 다른 실시형태에서, 자가-비활성화 렌티바이러스 벡터의 사용이 고려된다.

[0795] 시스템의 임의의 가이드 RNA, 예를 들어, 가이드 RNA 또는 염기 편집기-코딩 mRNA는 RNA 형태로 전달될 수 있다. 염기 편집기-코딩 mRNA는 시험관내(*in vitro*) 전사로 생성될 수 있다. 예를 들어, 뉴클레아제 mRNA는 다음 요소를 함유하는 PCR 카세트를 이용하여 합성할 수 있다: T7 프로모터, 선택적 코작 서열(GCCACC), 뉴클레아제 서열, 및 베타 글로빈-폴리A 테일로부터의 3' UTR과 같은 3' UTR. 카세트는 T7 중합효소에 의해 전사될 수 있다. 가이드 폴리뉴클레오티드(예를 들어, gRNA)는 또한 T7 프로모터, 이어서 서열 "GG", 및 가이드 폴리뉴클레오티드 서열을 함유하는 카세트로부터의 시험관내 전사를 이용하여 전사될 수 있다.

[0796] 발현을 향상시키고 있을 수 있는 독성을 감소시키기 위해, 염기 편집기-코딩 서열 및/또는 가이드 핵산은, 하나 이상의 변경된 뉴클레오시드, 예를 들어, 슈도(pseudo)-U 또는 5-메틸-C를 포함하도록 변형될 수 있다.

[0797] AAV 벡터의 작은 패키징 용량은 이 크기를 초과하는 많은 유전자의 전달 및/또는 큰 생리적 조절 요소의 사용을 어렵게 만든다. 이러한 문제는, 예를 들어, 전달될 단백질(들)을 2개 이상의 단편으로 분할하여 해결할 수 있으며, 여기서 N-말단 단편은 분할된 인테인-N에 융합되고 C-말단 단편은 분할된 인테인-C에 융합된다. 그런 다음 이러한 단편은 2개 이상의 AAV 벡터로 패키징된다. 한 실시형태태에서, 인테인은 AAV 캡시드 단백질에 이식된 다중-이펙터 염기 편집기 단백질의 단편 또는 일부를 결합하는 데 사용된다. 본 명세서에 사용된 "인테인"은 측접 N-말단 및 C-말단 엑스테인(예를 들어, 결합될 단편)을 결찰하는 자가-스플라이싱 단백질 인트론(예를 들어, 펩티드)을 지칭한다. 이종성 단백질 단편을 연결하기 위한 특정 인테인의 사용은, 예를 들어, 문헌[Wood *et al.*, J. Biol. Chem. 289(21); 14512-9 (2014)]에 기술되어 있다. 예를 들어, 분리된 단백질 단편에 융합될 때, 인테인 IntN 및 IntC는 서로를 인식하고, 스스로 스플라이싱하며, 이들이 융합된 단백질 단편의 측접 N- 및 C-말단 엑스테인을 동시에 결찰시켜, 그로 인해 두 단백질 단편으로부터의 전체 길이 단백질을 재구성한다. 다른 적합한 인테인은 당업자에게 명백할 것이다.

[0798] 본 발명의 융합 단백질의 단편은 길이가 다양할 수 있다. 일부 실시형태에서, 단백질 단편은 길이가 2개 아미노

산 내지 약 1000개 아미노산 범위이다. 일부 실시형태에서, 단백질 단편은 길이가 약 5개 아미노산 내지 약 500개 아미노산 범위이다. 일부 실시형태에서, 단백질 단편은 길이가 약 20개 아미노산 내지 약 200개 아미노산 범위이다. 일부 실시형태에서, 단백질 단편은 길이가 약 10개 아미노산 내지 약 100개 아미노산 범위이다. 다른 길이의 적합한 단백질 단편은 당업자에게 명백할 것이다.

[0799] 한 실시형태에서, 이중 AAV 벡터는 거대 도입유전자 발현 카세트를 2개의 분리된 절반(5' 및 3' 말단, 또는 머리와 꼬리)으로 분할함으로써 생성되며, 카세트의 각 절반은 (< 5kb의) 단일 AAV 벡터에 패키징된다. 전체-길이 도입유전자 발현 카세트의 재조립은 두 개의 이중 AAV 벡터에 의해 동일한 세포를 공동-감염시, 후속하여 하기에 의해 달성된다: (1) 5'와 3' 게놈 사이의 상동 재조합(HR)(이중 AAV 중첩 벡터); (2) 5' 및 3' 게놈의 ITR-매개 꼬리-머리 콘카티머형성(이중 AAV 트랜스(*trans*)-스플라이싱 벡터); 또는 (3) 이들 두 메커니즘의 조합(이중 AAV 하이브리드 벡터). 생체내에서(*in vivo*) 이중 AAV 벡터의 사용은 전체-길이 단백질을 초래한다. 이중 AAV 벡터 플랫폼의 사용은 크기가 >4.7kb인 도입유전자에 대한 효율적이고 실행가능한 유전자 전달 전략을 제시한다.

[0800] ***인테인***

[0801] 일부 실시형태에서, 뉴클레아제(예를 들어, Cas9)의 일부 또는 단편은 인테인에 융합된다. 뉴클레아제는 인테인의 N-말단 또는 C-말단에 융합될 수 있다. 일부 실시형태에서, 융합 단백질의 일부 또는 단편은 인테인에 융합되고 AAV 캡시드 단백질에 융합된다. 인테인, 뉴클레아제 및 캡시드 단백질은 임의의 배열(예를 들어, 뉴클레아제-인테인-캡시드, 인테인-뉴클레아제-캡시드, 캡시드-인테인-뉴클레아제 등)으로 함께 융합될 수 있다. 일부 실시형태에서, 인테인의 N-말단은 융합 단백질의 C-말단에 융합되고 인테인의 C-말단은 AAV 캡시드 단백질의 N-말단에 융합된다.

[0802] 인테인(개재 단백질)은 다양한 유기체에서 발견되는 자동-가공(auto-processing) 도메인이며, 단백질 스플라이싱으로 알려진 과정을 수행한다. 단백질 스플라이싱은 펩티드 결합의 절단과 형성 둘 다로 구성된 다단계 생화학 반응이다. 단백질 스플라이싱의 내인성 기질은 인테인-함유 유기체에서 발견되는 단백질이지만, 인테인은 사실상 모든 폴리펩티드 백본을 화학적으로 조작하는 데 사용할 수도 있다.

[0803] 단백질 스플라이싱에서, 인테인은 2개의 펩티드 결합을 절단함으로써 전구체 폴리펩티드에서 스스로를 잘라내어, 그로 인해 새로운 펩티드 결합의 형성을 통해 측접 엑스테인(외부 단백질) 서열을 결찰시킨다. 이 재배열은 번역후(또는 추정컨대 공동-번역으로) 일어난다. 인테인-매개 단백질 스플라이싱은 자발적으로 일어나며, 인테인 도메인의 폴딩만을 필요로 한다.

[0804] 약 5%의 인테인은 분할 인테인으로, 각각 하나의 엑스테인에 융합되는, 2개의 개별 폴리펩티드인 N-인테인과 C-인테인으로 전사되고 번역된다. 번역시, 인테인 단편은 자발적이고 비공유적으로 정규 인테인 구조로 조립되어 트랜스로 단백질 스플라이싱을 수행한다. 단백질 스플라이싱의 메커니즘은 일련의 아실 전달 반응을 수반하여 인테인-엑스테인 접합부에서 2개의 펩티드 결합에 대한 절단 및 N- 및 C-엑스테인 사이에 새로운 펩티드 결합의 형성을 초래한다. 이 과정은 N-엑스테인과 인테인의 N-말단을 연결하는 펩티드 결합의 활성화에 의해 시작된다. 사실상 모든 인테인은 이들의 N-말단에 C-말단 N-엑스테인 잔기의 카보닐 탄소를 공격하는 시스테인 또는 세린을 가지고 있다. 이 N에서 O/S 아실-이동은, 일반적으로 발견되는 아스파테이트와 함께, 보존된 트레오닌 및 히스티딘(TXXH 모티프로 지칭됨)에 의해 촉진되어, 선형(티오) 에스테르 중간체의 형성이 초래된다. 다음으로, 이 중간체는 시스테인, 세린, 또는 트레오닌인, 첫 번째 C-엑스테인 잔기(+1)의 친핵성 공격에 의해 트랜스-(티오)에스테르화를 겪게 된다. 생성된 분지형(티오) 에스테르 중간체는 고유한 변경을 통해 분해된다: 인테인의 고도로 보존된 C-말단 아스파라긴의 고리화. 이 과정은 히스티딘(고도로 보존된 HNF 모티프에서 발견됨)과 끝에서 두번째(penultimate) 히스티딘에 의해 촉진되며 아스파테이트를 포함할 수도 있다. 이 숙신이 미드 형성 반응은 반응성 복합체에서 인테인을 제거하고 비-펩티드 결합을 통해 부착된 엑스테인을 남긴다. 이 구조체는 인테인-비의존적 방식으로 안정적인 펩티드 결합으로 빠르게 재배열된다.

[0805] 일부 실시형태에서, 염기 편집기의 N-말단 단편(예를 들어, ABE, CBE)은 분할 인테인-N에 융합되고 C-말단 단편은 분할 인테인-C에 융합된다. 이러한 단편은 2개 이상의 AAV 벡터로 패키징된다. 이종성 단백질 단편을 연결하기 위한 특정 인테인의 사용은, 예를 들어, 문헌[Wood *et al.*, J. Biol. Chem. 289(21); 14512-9 (2014)]에 기술되어 있다. 예를 들어, 분리된 단백질 단편에 융합될 때, 인테인 IntN 및 IntC는 서로를 인식하고, 스스로 스플라이싱하며, 이들이 융합되는 단백질 단편의 측접 N- 및 C-말단 엑스테인을 동시에 결찰시키며, 그로 인해 두 단백질 단편로부터 전체 길이 단백질을 재구성한다. 다른 적합한 인테인은 당업자에게 자명할 것이다.

[0806] 일부 실시형태에서, ABE는 SpCas9의 선택된 영역 내의 Ala, Ser, Thr, 또는 Cys 잔기에서 N- 및 C-말단 단편으로 분할되었다. 이들 영역은 Cas9 결정 구조 분석에 의해 확인된 루프 영역에 해당한다. 각 단편의 N-말단은 인테인-N에 융합되고 각 단편의 C-말단은, 아래 서열에서 굵은 글씨 대문자로 표시된, 아미노산 위치 S303, T310, T313, S355, A456, S460, A463, T466, S469, T472, T474, C574, S577, A589, 및 S590에서 인테인 C에 융합된다.

1 mdkkysigld igtnsvgwav itdeykvpsk kfkvlgntdr hsikknliga llfdsgetae
61 atrlkrtarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnsd
181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknglfgn
241 lialslgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
301 ll**S**dilrvn**T** ei**T**kaplsas mikrydehhq dltllkalvr qqlpekykei ffdq**S**kngya
361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
421 ailrrqedfy pflkdnreki ekiltfripy yvgpl**A**rgn**S** rf**A**wm**T**rk**S**e e**T**i**T**pwnfee
481 vvdkgasaqs fiermtnfdk nlpnekvlpk hslllyeyftv ynеltkvkyv tegmrkpafl
541 sgeqkkaivd llfktnrkvt vkqlkedyfk kie**C**fd**S**vei sgvedrfn**AS** lgtyhdllki
601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfddkvmkq lkrrrytgwg
661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgdsl
721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrmn tkydendkli revkvitlks
961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd

[0807]

[0808] ***돌연변이를 표적화하기 위한 핵염기 편집기의 사용***

[0809] 하나 이상의 돌연변이를 표적으로 하는 핵염기 편집기 또는 다중-이펙터 핵염기 편집기의 적합성은 본 명세서에 기재된 바와 같이 평가된다. 한 실시형태에서, 관심있는 단일 세포는 리포터(예를 들어, GFP)를 코딩하는 소량의 벡터와 함께 염기 편집 시스템으로 형질도입된다. 이들 세포는 293T, K562 또는 U20S와 같은 불멸화된 인간 세포주를 포함하여 당업계에 공지된 임의의 세포주일 수 있다. 대안적으로, 일차 세포(예를 들어, 인간)가 사용될 수 있다. 이러한 세포는 최종 세포 표적과 관련이 있을 수 있다.

[0810] 전달은 바이러스 벡터를 사용하여 수행할 수 있다. 한 실시형태에서, 형질감염은 지질 형질 감염(예를 들어, 리포펙타민(Lipofectamine) 또는 퓨젠(Fugene))을 사용하거나 전기천공에 의해 수행될 수 있다. 형질감염 후, GFP의 발현은 형광 현미경 또는 유세포분석 중 어느 하나에 의해 결정되어 일관되고 높은 수준의 형질감염을 확인할 수 있다. 이러한 예비 형질감염은 가장 큰 활성을 제공하는 편집기 조합을 결정하기 위해 다른 핵염기 편집기를 포함할 수 있다.

[0811] 핵염기 편집기의 활성은 본 명세서에 기재된 바와 같이, 즉, 표적 서열의 변경을 검출하기 위해 세포의 게놈을 시퀀싱함으로써 평가된다. 생어(Sanger) 시퀀싱의 경우, 정제된 PCR 앰플리콘은 플라스미드 백본으로 복제되고,형질전환되고, 미니프렙화되며(miniprepped), 단일 프라이머로 시퀀싱된다. 시퀀싱은 또한 차세대 시퀀싱 기술을 사용하여 수행될 수 있다. 차세대 시퀀싱을 사용할 때, 앰플리콘은 의도된 절단 부위가 비대칭으로 배치된 300 내지 500 bp일 수 있다. PCR 후, 예를 들어, (예를 들어, Illumina MiSeq 상에서) 고 처리량 시퀀싱에 사용하기 위한, 차세대 시퀀싱 어댑터 및 바코드(예를 들어, Illumina 멀티플렉스 어댑터 및 인덱스)를 앰플리콘

의 말단에 추가할 수 있다.

[0812] 초기 시험에서 가장 높은 수준의 표적 특이적 변경을 유도하는 융합 단백질은 추가 평가를 위해 선택될 수 있다.

[0813] 특정 실시형태에서, 핵염기 편집기 또는 다중-이펙터 염기 편집기는 관심있는 폴리뉴클레오티드를 표적화하기 위해 사용된다. 한 실시형태에서, 본 발명의 핵염기 편집기 또는 다중-이펙터 염기 편집기는 세포의 게놈 내에서 관심있는 돌연변이를 표적으로 삼는 데 사용되는 가이드 RNA와 관련하여 세포(예를 들어, 조혈 세포 또는 이들의 전구 세포, 조혈 줄기 세포 및/또는 유도 만능 줄기 세포)에 전달되어, 그로 인해 돌연변이를 변경시킨다. 일부 실시형태에서, 염기 편집기는 관심있는 유전자의 서열에 하나 이상의 편집을 도입하기 위해 가이드 RNA에 의해 표적화된다.

[0814] 한 실시형태에서, 핵염기 편집기 또는 다중-이펙터 핵염기 편집기는, 이로만 제한되는 것은 아니지만, 스플라이스 부위, 인핸서, 및 전사 조절 요소를 포함하는, 조절 서열을 표적화하는 데 사용된다. 조절 요소에 의해 제어되는 유전자의 발현에 대한 변경의 효과는 당업계에 공지된 임의의 방법을 사용하여 분석된다.

[0815] 다른 실시형태에서, 본 발명의 핵염기 편집기 또는 다중-이펙터 핵염기 편집기는 상보성 결정 영역(CDR)을 코딩하는 폴리뉴클레오티드를 표적화하여 발현된 CDR에서 변경을 생성하는 데 사용된다. 그런 다음 CDR 기능에 대한 이러한 변경의 효과는, 예를 들어, 항원에 대한 CDR의 특이적 결합을 측정함으로써 분석된다.

[0816] 또 다른 실시형태에서, 본 발명의 다중-이펙터 핵염기 편집기는 유기체의 게놈 내에서 관심있는 폴리뉴클레오티드를 표적으로 하는 데 사용된다. 한 실시형태에서, 본 발명의 다중-이펙터 핵염기 편집기는 세포의 게놈 내에서 다양한 서열을 타일링하는 데 사용되는 가이드 RNA의 라이브러리와 관련하여 세포에 전달되어, 그로 인해 게놈 전체에 걸쳐 서열을 전체적으로(systematically) 변경시킨다.

[0817] 시스템은 하나 이상의 상이한 벡터를 포함할 수 있다. 일 양상에서, 염기 편집기는 원하는 세포 유형, 우선적으로 진핵 세포, 바람직하게는 포유류 세포 또는 인간 세포에서의 발현을 위해 코돈 최적화된다.

[0818] 일반적으로, 코돈 최적화는 천연 서열의 적어도 하나의 코돈(예를 들어, 약 1개, 2개, 3개, 4개, 5개, 10개, 15개, 20개, 25개, 50개, 또는 그 이상의 코돈)을 천연 아미노산 서열을 유지하면서 숙주 세포의 유전자에서 더 자주 또는 가장 빈번하게 사용되는 코돈으로 대체함으로써 관심 숙주 세포에서 발현을 향상시키기 위해 핵산 서열을 변경시키는 과정을 의미한다. 다양한 종은 특별한 아미노산의 특정 코돈에 대해 특별한 편향을 나타낸다. 코돈 편향(유기체 간 코돈 사용의 차이)은 종종 메신저 RNA(mRNA)의 번역 효율성과 관련이 있으며, 이는 결과적으로, 다른 무엇보다도, 번역되는 코돈의 특성 및 특정 전사 RNA(tRNA) 분자의 가용성에 의존하는 것으로 여겨진다. 세포에서 선택된 tRNA의 우세는 일반적으로 펩티드 합성에서 가장 빈번하게 사용되는 코돈의 반영이다. 따라서, 유전자는 코돈 최적화를 기반으로 주어진 유기체에서 최적의 유전자 발현을 위해 조정될 수 있다. 코돈 사용 테이블은, 예를 들어, www.kazusa.orjp/codon/(2002년 7월 9일 방문)에서 입수가능한 "코돈 사용 데이터베이스"에서 쉽게 사용할 수 있으며, 이러한 테이블은 여러 가지 방법으로 조정할 수 있다. 문헌[Nakamura, Y., *et al*. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000)] 참조. Gene Forge(Aptagen; Jacobus, Pa.)와 같은, 특정 숙주 세포에서 발현을 위해 특정 서열을 코돈 최적화하기 위한 컴퓨터 알고리즘도 또한 이용가능하다. 일부 실시형태에서, 조작된 뉴클레아제를 코딩하는 서열 내의 하나 이상의 코돈(예를 들어, 1개, 2개, 3개, 4개, 5개, 10개, 15개, 20개, 25개, 50개 또는 그 이상, 또는 모든 코돈)은 특정 아미노산에 대한 가장 빈번하게 사용되는 코돈에 상응한다.

[0819] 패키징 세포는 전형적으로 숙주 세포를 감염시킬 수 있는 바이러스 입자를 형성하는 데 사용된다. 이러한 세포는 아데노바이러스를 포장하는, 293 세포와 레트로바이러스를 포장하는, psi.2 세포 또는 PA317 세포를 포함한다. 유전자 치료에 사용되는 바이러스 벡터는 일반적으로 핵산 벡터를 바이러스 입자로 패키징하는 세포주를 생성하여 생성된다. 벡터는 전형적으로 패키징 및 숙주로의 후속 통합에 필요한 최소 바이러스 서열을 포함하고, 다른 바이러스 서열은 발현될 폴리뉴클레오티드(들)에 대한 발현 카세트로 대체된다. 소실된 바이러스 기능은 일반적으로 패키징 세포주에 의해 트랜스(*trans*)로 공급된다. 예를 들어, 유전자 치료에 사용되는 AAV 벡터는 일반적으로 패키징 및 숙주 게놈으로의 통합에 필요한 AAV 게놈으로부터의 ITR 서열만 보유한다. 바이러스 DNA는 다른 AAV 유전자, 즉 rep 및 cap을 코딩하지만, ITR 서열이 없는 헬퍼 플라스미드를 함유하는, 세포주에 패키징될 수 있다. 세포주는 헬퍼로서 아데노바이러스로 감염될 수도 있다. 헬퍼 바이러스는 헬퍼 플라스미드로부터 AAV 벡터의 복제 및 AAV 유전자의 발현을 촉진할 수 있다. 일부 경우에 있어서 헬퍼 플라스미드는 ITR 서열의 결여로 인해 상당한 양으로 패키징되지 않는다. 예를 들어, AAV보다 더 민감한 아데노바이러스에 대한 열처

리에 의해, 아데노바이러스로의 오염을 감소시킬 수 있다.

[0820] ***다중-이펙터 핵염기 편집기의 응용***

[0821] 다중-이펙터 핵염기 편집기를 이용하여 관심 폴리뉴클레오티드를 표적화하여 단백질 발현을 변형시키는 변경을 생성할 수 있다. 한 실시형태에서, 다중-이펙터 핵염기 편집기를 이용하여, 이로만 제한되는 것은 아니지만, 스플라이스 부위, 인핸서 및 전사 조절 요소를 포함하는, 비-코딩 또는 조절 서열을 변경시킨다. 그런 다음 조절 요소에 의해 제어되는 유전자의 발현에 대한 변경의 효과는 당업계에 공지된 임의의 방법을 사용하여 분석된다. 특정 실시형태에서, 다중-이펙터 핵염기 편집기는 조절 서열을 실질적으로 변경하며, 그로 인해 유전자 발현을 조절하는 이의 활성을 폐기시킬 수 있다. 유리하게는, 이것은, 다른 RNA 프로그래밍가능한 뉴클레아제와 달리, 게놈 표적 서열에서 이중-가닥 파손을 생성하지 않고 수행될 수 있다.

[0822] 다중-이펙터 핵염기 편집기를 이용하여 관심 폴리뉴클레오티드를 표적화하여 단백질 활성을 변형시키는 변경을 생성할 수 있다. 예를 들어, 돌연변이유발의 맥락에서, 다중-이펙터 핵염기 편집기는 오류-빈번 PCR 및 기타 중합효소-기반 방법에 비해 많은 이점을 갖는다. 본 발명의 다중-이펙터 핵염기 편집기는 표적 영역의 다수의 염기에서 변경을 생성하기 때문에, 이러한 돌연변이는, 코돈의 단일 뉴클레오티드 변화가 여전히 동일한 아미노산을 코딩할 수 있다는 점(예를 들어, 코돈 축퇴)을 감안하면 단백질 수준에서 발현될 가능성이 더 낮은, 오류-빈번 PCR에 의해 도입된 돌연변이에 비해 단백질 수준에서 발현될 가능성이 더 높다. 폴리뉴클레오티드 전체에 걸쳐 무작위 변경을 유도하는, 오류-빈번 PCR과 달리, 본 발명의 다중-이펙터 핵염기 편집기는 관심 단백질의 작거나 정의된 영역 내에서 특정 아미노산을 표적화하는 데 사용될 수 있다.

[0823] 다른 실시형태에서, 본 발명의 다중-이펙터 핵염기 편집기는 유기체의 게놈 내에서 관심 폴리뉴클레오티드를 표적으로 하는 데 사용된다. 한 실시형태에서, 유기체는 미생물균총(microbiome)의 미생물이다(예를 들어, 박테리오데테스(*Bacteriodetes*), 베루코마이크로비아(*Verrucomicrobia*), 피르미큐테스(*Firmicutes*); 감마프로테오박테리아(*Gammaproteobacteria*), 알파프로테오박테리아(*Alphaproteobacteria*), 박테리오데테스(*Bacteriodetes*), 클로스트리디아(*Clostridia*), 에리시페로트리치아(*Erysipelotrichia*), 바실리(*Bacilli*); 엔테로박테리아레스(*Enterobacteriales*), 박테리오달레스(*Bacteriodales*), 베루코마이크로비아레스(*Verrucomicrobiales*), 클로스트리디아레스(*Clostridiales*), 에리시오페로트리차레스(*Erysiopelotrichales*), 락토바실라레스(*Lactobacillales*); 엔테로박테리아세애(*Enterobacteriaceae*), 박테로이다세애(*Bacteroidaceae*), 에리시오페로트리차세애(*Erysiopelotrichaceae*), 프레보텔라세애(*Prevotellaceae*), 코로오박테리아세애(*Coriobacteriaceae*), 및 알카리제나세애(*Alcaligenaceae*); 에스키리치아(*Escherichia*), 박테로이데스(*Bacteroides*), 알리스티페스(*Alistipes*), 악커만시아(*Akkermansia*), 클로스트리디움(*Clostridium*), 락토바실러스(*Lactobacillus*)). 다른 실시형태에서, 유기체는 농업적으로 중요한 동물(예를 들어, 소, 양, 염소, 말, 닭, 칠면조) 또는 식물(예를 들어, 대두, 밀, 옥수수, 벼, 담배, 사과, 포도, 복숭아, 자두, 체리)이다. 한 실시형태에서, 본 발명의 다중-이펙터 핵염기 편집기는 세포의 게놈 내에서 다양한 서열을 타일링하는 데 사용되는 가이드 RNA의 라이브러리와 관련하여 세포에 전달되어, 그로 인해 게놈 전체에 걸쳐 서열을 전체적으로 변경시킨다.

[0824] 돌연변이는 구조-기능 분석을 용이하게 하거나 단백질의 내인성 활성을 변경하기 위해 임의의 다양한 단백질에서 생성될 수 있다. 돌연변이는, 예를 들어, 효소(예를 들어, 키나제, 포스파타제, 카복실라제, 포스포디에스테라제) 또는 효소 기질, 수용체 또는 이의 리간드, 및 항체 및 이의 항원에서, 생성될 수 있다. 한 실시형태에서, 다중-이펙터 핵염기 편집기는 효소의 활성 부위, 수용체의 리간드 결합 부위, 또는 항체의 상보성 결정 영역(CDR)을 코딩하는 핵산 분자를 표적으로 한다. 효소의 경우, 활성 부위에서 돌연변이를 유도하면 효소의 활성이 증가, 감소, 또는 폐기될 수 있다. 효소에 대한 돌연변이의 효과는 당 업계에 공지되고/되거나 숙련된 기술자에게 자명할 임의의 다수의 분석을 포함하는 효소 활성 분석으로 특성규명된다. 수용체의 경우, 리간드 결합 부위에서 만들어진 돌연변이는 리간드에 대한 수용체 친화성을 증가, 감소 또는 폐기시킬 수 있다. 이러한 돌연변이의 효과는 당업계에 공지되고/되거나 숙련된 기술자에게 명백할 수 있는 다수의 분석 중 임의의 것을 포함하는, 수용체/리간드 결합 분석으로 분석된다. 항체 CDR의 경우, CDR 내에서 만들어진 돌연변이는 항원에 대한 결합을 증가, 감소 또는 폐기시킬 수 있다. 달리, CDR 내에서 만들어진 돌연변이는 항원에 대한 항체의 특이성을 변경시킬 수 있다. 그런 다음, CDR 기능에 대한 이러한 변경의 효과는, 예를 들어, CDR의 이의 항원에 대한 특이적 결합을 측정함으로써 또는 임의의 다른 유형의 면역분석으로 분석된다.

[0825] **약제학적 조성물**

[0826] 본 개시의 다른 양상은 본 명세서에 기재된 임의의 염기 편집기, 융합 단백질, 또는 융합 단백질-가이드 폴리뉴

클레오티드 복합체를 포함하는 약제학적 조성물에 관한 것이다. 일부 실시형태에서, 약제학적 조성물은 약제학적으로 허용가능한 담체를 추가로 포함한다. 일부 실시형태에서, 약제학적 조성물은 (예를 들어, 특이적 전달, 반감기 증가, 또는 다른 치료 화합물을 위한) 추가 작용제를 포함한다.

[0827] 적합한 약제학적으로 허용가능한 담체는 일반적으로 약제학적 조성물을 대상체에게 투여하는 것을 돕거나, 약제학적 조성물을 전달가능한 제제로 가공하는 것을 돕거나, 투여 전에 약제학적 조성물을 저장하는 것을 돕는 불활성 물질을 포함한다. 약제학적으로 허용가능한 담체는 제형의 형태, 일관성, 점도, pH, 약동학, 용해도를 안정화, 최적화 또는 달리 변경할 수 있는 작용제를 포함할 수 있다.

[0828] 약제학적으로 허용가능한 담체로 작용할 수 있는 물질의 일부 비제한적인 예는 다음을 포함한다: (1) 락토스, 글루코스 및 수크로스와 같은, 당; (2) 옥수수 전분 및 감자 전분과 같은, 전분; (3) 셀룰로오스 및 그 유도체, 예컨대 나트륨 카복시 메틸셀룰로오스, 메틸셀룰로오스, 에틸셀룰로오스, 미정질 셀룰로오스 및 셀룰로오스 아세테이트; (4) 분말화된 트라가칸트; (5) 맥아; (6) 젤라틴; (7) 마그네슘 스테아레이트, 나트륨 라우릴설페이트 및 탈크와 같은, 윤활제; (8) 코코아 버터 및 좌약 왁스와 같은, 부형제; (9) 땅콩유, 면실유, 홍화유, 참기름, 올리브유, 옥수수 유 및 대두유와 같은, 오일; (10) 프로필렌 글리콜과 같은, 글리콜; (11) 글리세린, 소르비톨, 만니톨 및 폴리에틸렌 글리콜(PEG)과 같은, 폴리올; (12) 에틸올레에이트 및 에틸라우레이트와 같은, 에스터; (13) 한천; (14) 수산화 마그네슘 및 수산화 알루미늄과 같은, 완충제; (15) 알긴산; (16) 발열원이 없는 물; (17) 등장 식염수; (18) 링거 용액; (19) 에틸 알코올; (20) pH 완충 용액; (21) 폴리에스터, 폴리카보네이트 및/또는 폴리무수물; (22) 폴리펩티드 및 아미노산과 같은 증량제(bulking agents) (23) 에탄올과 같은, 혈청 알코올; 및 (23) 약제학적 제형에 사용되는 기타 무독성 양립성 물질. 완충제, 습윤제, 유화제, 희석제, 캡슐화제, 피부 침투 향상제, 착색제, 방출제(release agents), 코팅제, 감미제, 향미제, 방향제, 보존제(preservative) 및 항산화제가 또한 제형에 존재할 수 있다. 예를 들어, 담체는, 이로만 제한되는 것은 아니지만, 식염수, 완충 식염수, 덱스트로스, 아르기닌, 수크로스, 물, 글리세롤, 에탄올, 소르비톨, 덱스트란, 소듐 카르복시메틸 셀룰로스 및 이들의 조합을 포함할 수 있다.

[0829] 약제학적 조성물은 약 5.0 내지 약 8.0 범위와 같은, 생리학적 pH를 반영하는 미리결정된 수준으로 제형의 pH를 유지하기 위해 하나 이상의 pH 완충 화합물을 포함할 수 있다. 수성 액체 제형에 사용되는 pH 완충 화합물은 아미노산 또는 아미노산 혼합물, 예컨대, 히스티딘 또는 히스티딘 및 글라이신과 같은, 아미노산 혼합물일 수 있다. 대안적으로, pH 완충 화합물은 바람직하게는, 제형의 pH를 미리결정된 수준, 예컨대, 약 5.0 내지 약 8.0 범위로 유지하고, 칼슘 이온을 킬레이팅하지 않는 작용제이다. 이러한 pH 완충 화합물의 예시적인 예는, 이로만 제한되는 것은 아니지만, 이미다졸 및 아세테이트 이온을 포함한다. pH 완충 화합물은 제형의 pH를 미리결정된 수준으로 유지하기에 적합한 임의의 양으로 존재할 수 있다.

[0830] 약제학적 조성물은 또한 하나 이상의 삼투 조절제, 즉 제형의 삼투 특성(예를 들어, 긴장성, 삼투질농도(osmolality), 및/또는 삼투압)을 수혈받는 개체(recipient individuals)의 혈류 및 혈액 세포에 허용되는 수준으로 조절하는 화합물을 함유할 수 있다. 삼투 조절제는 칼슘 이온을 킬레이팅하지 않는 작용제일 수 있다. 삼투 조절제는 제형의 삼투 특성을 조절하는 당업자에게 공지되거나 입수가능한 임의의 화합물일 수 있다. 당업자는 본 발명의 제형에 사용하기 위한 주어진 삼투 조절제의 적합성을 경험적으로 결정할 수 있다. 적절한 유형의 삼투 조절제의 예시적인 예는, 이로만 제한되는 것은 아니지만, 염화나트륨 및 아세트산 나트륨과 같은, 염; 수크로스, 덱스트로스 및 만니톨과 같은, 당; 글라이신과 같은 아미노산; 및 하나 이상의 이들 작용제 및/또는 여러 유형의 혼합물을 포함한다. 삼투 조절제(들)는 제형의 삼투 특성을 조절하기에 충분한 임의의 농도로 존재할 수 있다.

[0831] 일부 실시형태에서, 약제학적 조성물은, 예를 들어, 유전자 편집을 위해 대상체에 전달하기 위해 제제화된다. 일부 실시형태에서, 본 명세서에서 고려되는 약제학적 조성물의 투여는 주입, 수혈 또는 비경구를 포함하나 이에 제한되지 않는 통상적인 기술을 사용하여 수행될 수 있다. 일부 실시형태에서, 비경구 투여는 혈관내, 정맥내, 근육내, 동맥내, 척추강내, 종양내, 피내, 복강내, 경기관, 피하, 표피하, 관절내, 피막하, 지주막하 및 흉골내 주입 또는 주사를 포함한다. 일부 실시형태태에서, 본 명세서에 기재된 약제학적 조성물의 적절한 투여 경로는, 제한됨이 없이, 다음을 포함한다: 국소, 피하, 경피, 피내, 병변내, 관절내, 복강내, 방광내, 경점막, 치은, 치내, 달팽이관내, 경막, 기관내, 경막내, 척추강내, 근육내, 정맥내, 혈관내, 골내(intraosseus), 안구주위(periocular), 종양내, 뇌내, 및 뇌실내(intracerebroventricular) 투여.

[0832] 일부 실시형태에서, 본 명세서에 기재된 약제학적 조성물은 질병 부위(예를 들어, 종양 부위)에 국소 투여된다. 일부 실시형태에서, 본 명세서에 기재된 약제학적 조성물은 주사, 카테터, 좌약 또는 임플란트에 의해 대상체에

게 투여되며, 임플란트는 막, 예컨대 시알성(sialastic) 막, 또는 섬유를 포함하는, 다공성, 비다공성, 또는 젤라틴성 물질로 되어 있다.

[0833] 다른 실시형태에서, 본 명세서에 기재된 약제학적 조성물은 제어 방출 시스템으로 전달된다. 일 실시형태에서, 펌프가 사용될 수 있다(예를 들어, 문헌[Langer, 1990, Science 249: 1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald *et al.*, 1980, Surgery 88:507; Saudek *et al.*, 1989, N. Engl. J. Med. 321:574] 참조). 다른 실시형태에서, 중합체 재료가 사용될 수 있다. (예를 들어, 다음 문헌 참조: Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds, Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol Sci Rev Macromol Chem 23:61). 또한, 다음 문헌 참조: Levy *et al.*, 1985, Science 228: 190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard *et al.*, 1989, J. Neurosurg. 71: 105.) 다른 제어 방출 시스템은, 예를 들어, Langer(*전게서*)에 논의되어 있다.

[0834] 일부 실시형태에서, 약제학적 조성물은 대상체, 예를 들어, 인간에게 정맥내 또는 피하 투여에 적합한 조성물로서 일상적인 절차에 따라 제제화된다. 일부 실시형태에서, 주사에 의한 투여용 약제학적 조성물은 가용화제로서 멸균 등장성 사용시의 용액 및 주사 부위의 통증을 완화하기 위한 리그노카인과 같은 국소 마취제이다. 일반적으로, 성분은 활성제의 양을 표시하는 앰플 또는 샤쉐(sachette)와 같은 밀폐된 용기에 건조 동결건조 분말 또는 물이 없는 농축물과 같이 단위 투여 형태로 별도로 공급되거나 함께 혼합된다. 약제가 주입으로 투여되는 경우, 멸균된 약제 등급의 물 또는 식염수가 들어 있는 주입병(infusion bottle)으로 조제할 수 있다. 약제학적 조성물이 주사로 투여되는 경우, 주사용 멸균수 또는 식염수의 앰플이 제공되어 투여 전에 성분이 혼합될 수 있다.

[0835] 전신 투여용 약제학적 조성물은 액체, 예를 들어, 멸균 식염수, 젖산 링거 또는 행크 용액일 수 있다. 또한, 약제학적 조성물은 고체 형태일 수 있고 사용 직전에 재용해되거나 현탁될 수 있다. 동결건조된 형태도 고려된다. 약제학적 조성물은 비경구 투여에도 적합한 리포솜 또는 미세 결정과 같은 지질 입자 또는 비히클 내에 포함될 수 있다. 입자는 조성물이 내부에 함유되어 있는 한 단층라멜라(unilamellar) 또는 복층라멜라(plurilamellar)와 같은 임의의 적합한 구조일 수 있다. 화합물은 융합성 지질 다이올레오일포스파티딜에탄올아민(DOPE), 낮은 수준(5 내지 10 몰%)의 양이온성 지질을 포함하는 "안정화된 플라스미드-지질 입자"(SPLP)에 포획될 수 있으며, 폴리에틸렌 글리콜(PEG) 코팅에 의해 안정화될 수 있다(Zhang Y. P. *et al.*, Gene Ther 1999, 6:1438-47). N-[1-(2,3-다이올레오일옥시)프로필]-N,N,N-트리메틸-암모늄메틸설페이트, 또는 "DOTAP"과 같이 양으로 하전된 지질이 이러한 입자 및 비히클에 특히 바람직하다. 이러한 지질 입자의 제조는 잘 알려져 있다. 예를 들어, 각각이 본 명세서에 참조로 통합되는 다음 문헌 참조: 미국 특허 제4,880,635호; 제4,906,477호; 제4,911,928호; 제4,917,951호; 제4,920,016호; 및 제4,921,757호.

[0836] 본 명세서에 기재된 약제학적 조성물은, 예를 들어, 단위 용량(unit dose)으로 투여되거나 패키징될 수 있다. 본 개시의 약제학적 조성물과 관련하여 사용될 때, 용어 "단위 용량"은 대상체에 대한 단일 용량으로 적합한 물리적으로 분리된 단위를 지칭하며, 각 단위는 필요한 희석제; 즉, 담체, 또는 비히클과 관련하여 원하는 치료 효과를 생성하도록 계산된 미리결정된 양의 활성 물질을 함유한다.

[0837] 또한, 약제학적 조성물은 (a) 동결건조된 형태로 본 발명의 화합물을 함유하는 용기 및 (b) (예를 들어, 본 발명의 동결건조 화합물의 재구성 또는 희석을 위해 사용되는 멸균된) 약제학적으로 허용가능한 희석제를 함유하는 제2 용기를 포함하는 제약 키트로서 제공될 수 있다. 임의로, 이러한 용기와 관련하여 의약품 또는 생물학적 제품의 제조, 사용 또는 판매를 규제하는 정부기관이 규정한 형식의 통지가 있을 수 있으며, 이 통지는 인간 투여를 위한 제조, 사용 또는 판매에 관한 정부기관의 승인을 반영한다.

[0838] 또 다른 양상에서, 상술한 질병의 치료에 유용한 물질을 함유하는 제조 물품이 포함된다. 일부 실시형태에서, 제조 물품은 용기 및 라벨을 포함한다. 적합한 용기는, 예를 들어, 병, 바이알, 주사기 및 시험관을 포함한다. 용기는 유리 또는 플라스틱과 같은 다양한 재료로 형성될 수 있다. 일부 실시형태에서, 용기는 본 명세서에 기재된 질병을 치료하는 데 효과적이며, 멸균 접근 포트를 가질 수 있는 조성물을 보유한다. 예를 들어, 용기는 정맥내 용액 백 또는 피하 주사 바늘로 뚫을 수 있는 마개가 있는 바이알일 수 있다. 조성물의 활성제는 본 발명의 화합물이다. 일부 실시형태에서, 용기 상의 또는 용기와 관련된 라벨은 조성물이 선택된 질환을 치료하기 위해 사용됨을 나타낸다. 제조 물품은 포스페이트 완충 식염수, 링거 용액 또는 덱스트로스 용액과 같은 약제학적으로 허용가능한 완충액을 포함하는 제2 용기를 추가로 포함할 수 있다. 다른 완충제, 희석제, 필터, 바늘, 주사기 및 사용 설명서가 있는 패키지 삽입물을 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추

가로 포함할 수 있다.

[0839] 일부 실시형태에서, 본 명세서에 기재된 임의의 융합 단백질, gRNA 및/또는 복합체는 약제학적 조성물의 일부로서 제공된다. 일부 실시형태에서, 약제학적 조성물은 본 명세서에 제공된 임의의 융합 단백질을 포함한다. 일부 실시형태에서, 약제학적 조성물은 본 명세서에 제공된 임의의 복합체를 포함한다. 일부 실시형태에서, 약제학적 조성물은 gRNA 및 양이온성 지질과 복합체화하는 RNA-가이드 뉴클레아제(예를 들어, Cas9)를 포함하는 리보핵단백질 복합체를 포함한다. 일부 실시형태에서, 약제학적 조성물은 gRNA, 핵산 프로그래밍가능한 DNA 결합 단백질, 양이온성 지질 및 약제학적으로 허용가능한 부형제를 포함한다. 약제학적 조성물은 임의로, 하나 이상의 추가 치료 활성 물질을 포함할 수 있다.

[0840] 일부 실시형태에서, 본 명세서에 제공된 조성물은 대상체 내에서 표적화된 게놈 변경을 수행하기 위해 대상체, 예를 들어, 인간 대상체에게 투여된다. 일부 실시형태에서, 세포는 대상체로부터 수득되고 본 명세서에 제공된 임의의 약제학적 조성물과 접촉된다. 일부 실시형태에서, 대상체로부터 제거되고 생체외에서 약제 학적 조성물과 접촉된 세포는 임의로 원하는 게놈 변형이 세포에서 수행되거나 검출된 후 대상체에게 재도입된다. 뉴클레아제를 포함하는 약제학적 조성물을 전달하는 방법은 공지되어 있으며, 예를 들어, 이의 개시내용은 그 전문이 본 명세서에 참조로 통합되는, 미국 특허 번호 제6,453,242호; 제6,503,717호; 제6,534,261호; 제6,599,692호; 제6,607,882호; 제6,689,558호; 제6,824,978호; 제6,933,113호; 제6,979,539호; 제7,013,219호; 및 제7,163,824호에 기술되어 있다. 본 명세서에 제공된 약제학적 조성물의 설명은 주로 인간에게 투여하기에 적합한 약제학적 조성물에 관한 것이지만, 이러한 조성물은 일반적으로 모든 종류의 동물 또는 유기체에 투여하기 위해, 예를 들어, 수의학적 사용을 위해 적합하다는 것을 당업자는 이해할 것이다.

[0841] 다양한 동물에 투여하기에 적합한 조성물을 만들기 위해 인간에게 투여하기에 적합한 약제학적 조성물의 변형은 잘 이해되어 있으며, 통상의 숙련된 수의학 약리학자는, 필요하다면, 단지 통상적인 실험으로 이러한 변경을 설계 및/또는 수행할 수 있다. 약제학적 조성물의 투여가 고려되는 대상체는, 이로만 제한되는 것은 아니지만, 인간 및/또는 비인간 영장류, 포유동물, 가축, 애완 동물, 및 상업적으로 관련된 포유동물, 예컨대 소, 돼지, 말, 양, 고양이, 개, 마우스 및/또는 랫트; 및/또는 닭, 오리, 거위 및/또는 칠면조와 같은 상업적으로 관련된 새를 포함한 새를 포함한다.

[0842] 본 명세서에 기재된 약제학적 조성물의 제형은 공지된 또는 이후 약리학 분야에서 개발되는 임의의 방법에 의해 제조될 수 있다. 일반적으로, 이러한 제조 방법은 활성 성분(들)을 부형제 및/또는 하나 이상의 다른 보조 성분과 결합시키는 단계를 포함하고, 그런 다음 필요 및/또는 바람직하다면, 제품을 원하는 단일 또는 다중-용량 단위로 성형 및/또는 포장하는 단계를 포함한다. 약제학적 제형은 약제 학적으로 허용되는 부형제를 추가로 포함할 수 있으며, 이는 본 명세서에 사용된, 임의의 및 모든 용매, 분산 매질, 희석제 또는 기타 액체 비히클, 분산 또는 현탁 보조제, 표면 활성제, 등장화제, 증점제 또는 유화제, 보존제, 고체 결합제, 윤활제 및 원하는 특정 투여 형태에 적합한, 기타의 것을 포함한다. 그 전문이 본 명세서에 참고로 포함되는, 문헌[Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006]은 약제학적 조성물의 제형화에 사용되는 다양한 부형제 및 그의 제조를 위한 공지된 기술을 개시한다. 뉴클레아제를 포함하는 약제학적 조성물을 제조하기 위한 추가의 적합한 방법, 시약, 부형제 및 용매에 대해서는 또한 본 명세서에 그 전문이 참조로 통합되는, PCT 출원 PCT/US2010/055131(공개 번호 WO2011/053982A8, 2010년 11월 2일에 출원됨)을 참조.

[0843] 임의의 통상적인 부형제 매질(medium)은 임의의 바람직하지 않은 생물학적 효과를 생성하거나 또는 달리 약제학적 조성물의 임의의 다른 성분(들)과 유해한 방식으로 상호작용함으로써 물질 또는 그 유도체와 양립할 수 없는 경우를 제외하고, 그 사용이 본 개시의 범위 내에 있는 것으로 고려된다.

[0844] 위에서 기술한 것과 같은, 조성물은 유효량으로 투여될 수 있다. 유효량은 투여 방식, 치료할 특정 상태 및 원하는 결과에 따라 달라진다. 또한 상태의 단계, 대상체의 연령 및 신체 상태, 동시 요법의 특성(있는 경우) 및 의료 전문가에게 잘 알려진 유사한 요인에 따라 달라질 수 있다. 치료적 적용의 경우, 이것은 의학적으로 바람직한 결과를 달성하기에 충분한 양이다.

[0845] 일부 실시형태에서, 본 개시에 따른 조성물은 임의의 다양한 질환, 장애, 및/또는 상태의 치료에 사용될 수 있다.

[0846] ***키트, 벡터, 세포***

[0847] 본 개시의 다양한 양상은 염기 편집기 시스템을 포함하는 키트를 제공한다. 한 실시형태에서, 키트는 핵염기 편

집기 융합 단백질을 코딩하는 뉴클레오티드 서열을 포함하는 핵산 구축물을 포함한다. 융합 단백질은 하나 이상의 데아미나제 도메인(예를 들어, 시티딘 데아미나제 및/또는 아데닌 데아미나제) 및 핵산 프로그래밍가능한 DNA 결합 단백질(napDNAbp)을 포함한다. 일부 실시형태에서, 키트는 관심있는 핵산 분자를 표적화할 수 있는 적어도 하나의 가이드 RNA를 포함한다. 일부 실시형태에서, 키트는 적어도 하나의 가이드 RNA를 코딩하는 뉴클레오티드 서열을 포함하는 핵산 구축물을 포함한다. 일부 실시형태에서, 키트는 (a) 본 명세서에 제공된 바와 같은 아데노신 데아미나제 및/또는 시티딘 데아미나제에 융합된 Cas9 도메인; 및 (b) (a)의 서열의 발현을 유도하는 이종성 프로모터를 코딩하는 뉴클레오티드 서열을 포함하는, 핵산 구축물을 포함한다.

[0848] 키트는, 일부 실시형태에서, 하나 이상의 돌연변이를 편집하기 위해 키트를 사용하기 위한 지침서(instructions)를 제공한다. 지침서에는 일반적으로 핵산 분자를 편집하기 위한 키트 사용에 대한 정보가 포함된다. 다른 실시형태에서, 지침서는 다음 중 적어도 하나를 포함한다: 예방 조치; 경고; 임상 연구; 및/또는 참조. 지침서는 용기(존재시)에 직접 인쇄되거나, 용기에 부착된 라벨로, 또는 용기 내에 또는 이와 함께 제공된 별도의 시트, 팜플렛, 카드 또는 폴더로 인쇄될 수 있다. 추가 실시형태에서, 키트는 적절한 작동 파라미터를 위한 라벨 또는 별도의 삽입물(패키지 삽입물) 형태의 지침서를 포함할 수 있다. 또 다른 실시형태에서, 키트는 검출, 보정 또는 정규화를 위한 표준(들)으로 사용되는 적절한 양성 및 음성 대조군 또는 대조군 샘플이 있는 하나 이상의 용기를 포함할 수 있다. 키트는 (멸균) 포스페이트-완충 식염수, 링거 용액 또는 덱스트로스 용액과 같은 약제학적으로 허용가능한 완충액을 포함하는 제2 용기를 추가로 포함할 수 있다. 다른 완충제, 희석제, 필터, 바늘, 주사기 및 사용 지침서가 있는 패키지 삽입물을 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[0849] 본 개시의 일부 양상은 본 명세서에 제공된 임의의 핵염기 편집기 또는 다중-이펙터 핵염기 편집기 또는 융합 단백질을 포함하는 세포를 제공한다. 일부 실시형태에서, 세포는 본 명세서에 제공된 임의의 뉴클레오티드 또는 벡터를 포함한다.

[0850] 본 발명의 실시는, 달리 지시되지 않는 한, 숙련된 기술자가 잘 이해하고 있는 범위 내의, 분자 생물학(재조합 기술 포함), 미생물학, 세포 생물학, 생화학 및 면역학의 통상적인 기술을 채택한다. 이러한 기술은, 다음과 같은 문헌에 자세히 설명되어 있다: "Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase Chain Reaction", (Mullis, 1994); "Current Protocols in Immunology" (Coligan, 1991). 이러한 기술들은 본 발명의 폴리뉴클레오티드 및 폴리펩티드의 생산에 적용할 수 있으며, 그 자체로 본 발명을 제조하고 실시하는 데 고려될 수 있다. 특정 실시형태에 대한 특히 유용한 기술은 이하의 섹션에서 논의될 것이다.

[0851] 이하의 실시예는 본 발명의 분석, 스크리닝, 및 치료학적 방법을 구현하고 사용하는 방법에 대한 완전한 개시 및 설명을 당업자에게 제공하기 위해 제시되며, 본 발명자들이 자신의 발명으로 간주하는 범위를 제한하려는 의도는 아니다.

[0852] **실시예**

[0853] **실시예 1: 비-표적 탈아미노화가 감소된 핵염기 편집기의 구축**

[0854] 핵염기 편집기(예를 들어, 링커에 의해 연결된 CRISPR-Cas 단백질과 데아미나제의 융합 단백질)를 사용하여 표적 폴리뉴클레오티드에 특정 점 돌연변이를 도입할 수 있다. 그러나, 핵염기 편집기는 의도하지 않은 게놈-전체의 불요성 탈아미노화, 방관자 돌연변이, 및 표적 근위 편집의 가능성을 가지고 있다. 이론에 구속됨이 없이, 염기 편집기에서 링커를 줄이거나 제거하면 의도하지 않은 탈아미노화 이벤트의 가능성이 감소하고/하거나 원하는 표적 탈아미노화를 촉진할 수 있다(도 1). 이것은 부분적으로 핵염기 편집기의 데아미나제 도메인에 대한 유효 활성 반경을 감소시키기 때문일 수 있다. DNA에 결합된 Cas9의 구조는 X-선 결정학에 의해 결정되었지만, 염기 편집이 일어나는 DNA 부분에 대한 구조적 정보는 존재하지 않는다. Cas9의 모델링은 염기 편집이 일어나는 DNA가 Cas9에 근접한 2개의 위치에 있을 수 있음을 예측한다(도 2). 이러한 예측에 기초하여, 이들 위치 중 하나 이상에 데아미나제 또는 이의 단편을 위치시키는 것은 원하지 않는 탈아미노화 이벤트를 감소시키면서 표적-적중 염기 편집을 촉진할 가능성이 있다(도 3). TadA 데아미나제 또는 이의 단편의 삽입이 가능한 아데노신 염기 편집기(예를 들어, TadA에 융합된 Cas9) 내의 몇몇 영역이 확인되었다(도 4-7). 그에 따라, 확인된 위치에서 Cas9 폴리펩티드 내로 TadA 또는 이의 변이체를 삽입하여 아데노신 데아미나제 염기 편집기를 생성시켰다.

[0855] **실시예 2: 표적-적중 및 표적-이탈 탈아미노화를 측정하기 위한 고-처리량 시험관내 분석**

[0856] 핵염기 편집기를 평가하고 불요성 탈아미노화를 포함하는 비-표적 탈아미노화 대 표적 탈아미노화를 측정하여 후보 구축물을 특성규명하기 위한 시험관내 분석을 개발했다. FRET-기반 버전의 분석은 검출을 위해 형광 리포터를 사용하지만, 이 분석은 겔-기반 판독에 적합할 수 있다(도 8). 시험관내 탈아미노화 분석을 위한 프로브는 탈아미노화를 위한 기질, 특히 핵염기 편집기를 위한 기질을 포함한다(도 8). 탈아미노화될 수 있는 뉴클레오티드를 함유하는 것 외에도, 프로브는 PAM 서열, 표적 특이적 서열 등을 포함하거나, 심지어 무작위 서열을 포함할 수 있다. 프로브 세트를 이용한 탈아미노화 반응은 병렬로 수행할 수 있다(예를 들어, 고-처리량 형식). 기질의 탈아미노화(C→U 또는 A→I)는 기질을 탈아미노화 특이적 엔도뉴클레아제(각각, USER/엔도뉴클레아제V)에 의해 절단가능하게 만든다(도 8). 기질의 절단은 퀜처(quencher) 분자로부터 형광 리포터를 분리하여 형광 신호를 생성시킨다(도 8). 에 대한 높은 표적-적중 대 표적-이탈 형광 비율은 염기 편집기가 효과적임을 나타낸다. 당해 기술분야에 공지된 임의의 상호작용하는 형광단 및 퀜처 쌍 또는 FRET 공여자-수용체 쌍을 사용할 수 있다. 특정 실시형태에서, 형광단은 FAM, TET, HEX, TAMRA, JOE, 또는 ROX 중 하나 이상이다. 다양한 실시형태에서, 퀜처는 5' Iowa Black® RQ(5IabRQ)를 포함하는 댑씰(dabcyl), 댑실(dabsyl), 블랙 홀 퀜처(Black Hole Quencher) 염료 중 하나 이상이다. 일반적으로, 퀜칭 염료는 여기 매칭 퀜칭 염료이다. 형광단-퀜처 쌍과 이들의 선택은, 예를 들어, 문헌[Marras, Selection of Fluorophore and Quencher Pairs for Fluorescent Nucleic Acid Hybridization Probes in Methods in Molecular Biology: Fluorescent Energy Transfer Nucleic Acid Probes: Designs and Protocols. Edited by: V.V. Didenko ⓒ Humana Press Inc., Totowa, NJ]에 설명되어 있다.

[0857] 이 분석에 대한 시연으로서, 아데노신 염기 편집기를 SpCas9(탈아미노화 도메인 없음)의 존재하 또는 단백질 없이 발생하는 탈아미노화에 대해 아데노신 염기 편집기의 표적-적중 탈아미노화를 비교함으로써 표적-이탈 탈아미노화를 생성시키는 가능성에 대해 분석했다(도 9). 아데노신 염기 편집기 반응은 SpCas9에 비해 높은 형광 신호를 생성했고 단백질 반응은 없었으며, 이는 ABE가 표적-적중 염기 편집에 효과적임을 나타낸다. 또 다른 예에서, 아데노신 염기 편집기를 SpCas9가 TadA와 트랜스로 존재하는 트랜스형 아데노신 염기 편집(ABE-TadA)과 비교했다(도 10). ABE는 ABE-TadA, SpCas9에 비해 증가된 형광을 생성시켰고, 단백질 반응은 없었으며, 표적-적중 염기 편집에 효과적이었다. 단일-가닥 구조 및 분지형 구조를 포함하는, 불요성 표적-이탈 염기 편집에 대한 잠재적 기질을 이 분석으로 테스트할 수 있고, 이는 게놈 내의 다른 구조(예를 들어, DNA "호흡", 복제 분기점, 전사 활성 DNA 등)를 반영할 수 있다. (도 11).

[0858] **실시예 3: 시스형 및 트랜스형 데아미나제의 활성을 평가하기 위한 분석.**

[0859] 시스형(CRISPR-Cas에 공유결합된 탈아미노화 도메인) 및 트랜스형(트랜스형으로 제공된 탈아미노화 도메인을 지니는 CRISPR-Cas 단백질)에서 데아미나제의 활성을 구별하기 위한 분석을 개발했다(도 12). 시스형에서 발생하는 탈아미노화는 표적화된 염기 편집에 의한 탈아미노화를 나타내는 반면, 트랜스형에서의 탈아미노화는 불요성 탈아미노화를 나타낸다. 시스형 대 트랜스형 활성의 높은 비율은 불요성 탈아미노화가 감소된 데아미나제가 염기 편집기로서 효과적임을 나타낸다.

[0860] 랫트 APOBEC1을 시스형-트랜스형 분석으로 테스트했다. 간단히 말해서, HEK293T 세포를 염기 편집기 BE4(rAPOBEC1-nCas9-UGI-UGI), rAPOBEC1 및 nCas9, nCas9 및 가이드 RNA, 또는 rAPOBEC1 및 가이드 RNA를 발현하는 구축물로 형질감염시켰다. 게놈 DNA를 세포로부터 분리하고 4개의 게놈 표적 부위에 대한 시퀀싱을 수득했다. 모든 부위에서, rAPOBEC1은 트랜스형 데아미나제 활성과 비교하여 시스형 데아미나제 활성이 더 높게 나타났을 뿐만 아니라, 표적화된 염기 편집을 위한 구성요소들 중 적어도 하나가 결핍된 다른 대조군 반응에 비해서도 높게 나타났다(도 13). 유사하게, TadA7.10은 또한 트랜스형 데아미나제 활성 및 기타 탈아미노화 이벤트에 비해 시스형 데아미나제 활성이 더 높은 것으로 나타났다(도 14). 가이드와 분리된 트랜스에서 아데노신 염기 편집기의 효과를 이해하기 위해, SaCas9-ABE 및 SaCas9 가이드를 SpCas9-ABE 및 SaCas9 가이드, 그리고 입체 장애 ABE 변이체 및 SaCas9 가이드와 함께 테스트했다(도 15). 이러한 맥락에서, SpCas9-ABE는 염기 편집기 맥락에서 TadA-TadA7.10에 대한 트랜스형 활성이 더 낮은 것으로 나타났다. ABE 및 입체 장애 ABE 변이체에 대한 시스형/트랜스형 활성의 비율을 SaCas9 가이드 분석으로부터의 트랜스형 측정 및 ABE 및 입체 장애 ABE 변이체의 활성을 이용하여 추정했다. ABE 및 입체 장애 ABE 변이체에 대한 추정 비율은 상대적으로 높았다. 시스형 및 트랜스형 활성에 대한 용량 반응 연구는 높은 시스형 대 트랜스형 활성이 용량에 의해 조절될 수 있는지 여부(예를 들어, 시스형 활성이 용량 증가에 따라 트랜스형 활성보다 더 빠르게 증가하는지 여부)를 결정하기 위해 수행했다. 테스트한 조건 하에서, 시스형 대 트랜스형 활성의 용량 반응은 관찰되지 않았다(도 16-18).

[0861] 시스형-트랜스형 분석을 이용하여 아래 표 9에 나열된 다양한 데아미나제의 감소된 불요성 탈아미노화를 평가했다.

[0862] 표 9. 시스형-트랜스형 분석을 이용하여 스크리닝된 데아미나제

1	rAPOBEC-1	9	hAPOBEC-2	17	hAPOBEC-3F	25	btAID
2	mAPOBEC-1	10	ppAPOBEC-2	18	hAPOBEC-3G	26	mAID
3	maAPOBEC-1	11	btAPOBEC-2	19	hAPOBEC-4	27	pmCDA-1
4	hAPOBEC-1	12	mAPOBEC-3	20	mAPOBEC-4	28	pmCDA-2
5	ppAPOBEC-1	13	hAPOBEC-3A	21	rAPOBEC-4	29	pmCDA-5
6	ocAPOBEC1	14	hAPOBEC-3B	22	mfAPOBEC-4	30	yCD
7	mdAPOBEC-1	15	hAPOBEC-3C	23	hAID	31	rAPOBEC-1- 델타 177-186
8	mAPOBEC-2	16	hAPOBEC-3D	24	clAID	32	rAPOBEC-1- 델타 202-213

[0863]

[0864] 흥미롭게도, ppAPOBEC-2, mAPOBEC-2, mAPOBEC-3, 및 mfAPOBEC-4를 포함하는, 몇몇 데아미나제는 높은 시스형/트랜스형 활성을 나타냈다.

[0865] rAPOBEC-1 라투스 노르베지쿠스(*Rattus norvegicus*)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT
NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR
LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW
[0866] VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

[0867] mAPOBEC-1 무스 무스쿨루스(*Mus musculus*)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQN
TSN
HVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARL
Y
HHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVK
[0868] LYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK

[0869] maAPOBEC-1 메소크리세투스 아우라투스(*Mesocricetus auratus*)

MSSETGPVVVDPTLRRRIEPHEFDAFFDQGELRKETCLLYEIRWGGRHNIWRHTGQN
TSRHVEINFIEKFTSERYFYPSTRCSIVWFLSWSPCGECSKAITEFLSGHPNVTLFIYAA
RLY
HHTDQRNRQGLRDLISRGVTIRIMTEQEYCYCWRNFVNYPPSNEVYWPRYPNLWMR
[0870] LYALELYCIHLGLPPCLKIKRRHQYPLTFFRLNLQSCHYQRIPPHILWATGFI

[0871] hAPOBEC-1 호모 사피엔스(*Homo sapiens*)

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKN
TTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYV
ARLF
WHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLW
MMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSV
[0872] AWR

[0873] ppAPOBEC-1 폰고 피그마에우스(*Pongo pygmaeus*)

[0874] MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKN TTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYV ARLF WHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLW MMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSV TWR

[0875] ocAPOBEC1 오릭토라구스 쿠니쿨루스(*Oryctolagus cuniculus*)

[0876] MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELRKEACLLYEIKWGASSKTWRSSGKN TTNHVEVNFLEKLTSEGRLGPSTCCSITWFLSWSPCWECSMAIREFLSQHPGVTLIIFV ARLF QHMDRRNRQGLKDLVTSGVTVRVMSVSEYCYCWENFVNYPPGKAAQWPRYPPRW MLMYALELYCIILGLPPCLKISRRHQKQLTFFSLTPQYCHYKMIPPYILLATGLLQPSV PWR

[0877] mdAPOBEC-1 모노델피스 도메스티카(*Monodelphis domestica*)

[0878] MNSKTGPSVGDATLRRRIKPWEFVAFFNPQELRKETCLLYEIKWGNQNIWRHSNQN TSQHAEINFMEKFTAERHFNSSVRCSITWFLSWSPCWECSKAIRKFLDHYPNVTLAIFI

[0879] SRLYWHMDQQHRQGLKELVHSGVTIQIMSYSEYHYCWRNFVDYPQGEEDYWPKYP YLWIMLYVLELHCIILGLPPCLKISGSHSNQLALFSLDLQDCHYQKIPYNVLVATGLV QPFVTWR

[0880] mAPOBEC-2 무스 무스쿨루스(*Mus musculus*)

[0881] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFR NVEYSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDP ALKYNVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKL KEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

[0882] hAPOBEC-2 호모 사피엔스(*Homo sapiens*)

[0883] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRN VE YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALR YNVTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAG CKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK

[0884] ppAPOBEC-2 폰고 피그마에우스(*Pongo pygmaeus*)

[0885] MAQKEEAAAATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRN VE YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALR YNVTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEELEIQDALKKLKEAG CKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK

[0886] btAPOBEC-2 보스 타우루스(*Bos Taurus*)

[0887] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRN VE YSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALR YMVTWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEA GCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK

[0888] mAPOBEC-3 무스 무스쿨루스(*Mus musculus*)

[0889] MQPQRLGPRAGMGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCY
EVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSW

[0890] SPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLY
EFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYISVPSSSSSTLSNICL
TKGLPETRFWVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNG
QAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRD
RPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPF
WPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS

[0891] hAPOBEC-3A 호모 사피엔스(*Homo sapiens*)

[0892] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLH
NQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAF
LQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQ
GCPFQPWDGLDEHSQALSGRLRAILQNQGN

[0893] hAPOBEC-3B 호모 사피엔스(*Homo sapiens*)

[0894] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFR
GQVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHP
NVTLTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQ
FMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERL
DNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFI
SWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSI
MTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN

[0895] hAPOBEC-3C 호모 사피엔스(*Homo sapiens*)

[0896] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVF
RNQVDSETHCHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARH
SNVNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPF
KPWKGLKTNFRLLKRRLRESLQ

[0897] hAPOBEC-3D 호모 사피엔스(*Homo sapiens*)

[0898] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFR
GPVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPC
VVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAY
CWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKA
CGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPN

[0899] TNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLS
QEGASVKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ

[0900] hAPOBEC-3F 호모 사피엔스(*Homo sapiens*)

[0901] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQ
VYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNV
TLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMP
WYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEV
VKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPC
PECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFK
YCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE

[0902] hAPOBEC-3G 호모 사피엔스(*Homo sapiens*)

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQ
VYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDP
KVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYS
QRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEV
ERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRV
TCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISI

[0903] MTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN

[0904] hAPOBEC-4 호모 사피엔스(*Homo sapiens*)

MEPIYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTT
F
PQTKHLTFYELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHII
L
YSNNSPCNEANHCCISKMYNFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLA
SL
WPRVVLSPISGGIWHSVLHSFISGVSGSHVFQPILTGRALADRHNAYEINAITGVKPYF
T
DVLLQTKRNPNTKAQEALESYPLNNAFPGQFFQMPSGQLQPNLPPDLRAPVVFVLVP
LRDLPPMHMGQNPNKPRNIVRHLNMPQMSFQETKDLGRLPTGRSVEIVEITEQFASS

[0905] KEADEKKKKKGKK

[0906] mAPOBEC-4 무스 무스쿨루스(*Mus musculus*)

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHEN

[0907] SVRLTRQLRRILLPLYEVDDLRDAFRMLGF

[0908] rAPOBEC-4 라투스 노르베지쿠스(*Rattus norvegicus*)

MEPLYEEYLTHSGTIVKPYYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWS
TYP
QTKHLTFYELRSSSGNLIQKGLASNCTGSHTHPESMLFERDGYLDSLIFHDSNIRHIIL
Y
SNNSPCDEANHCCISKMYNFLMNYPEVTLSVFFSQLYHTENQFPTSAWNREALRGLA
SLWPQVTLSAISGGIWQSILETFVSGISEGLTAVRPFTAGRTLTDRYNAYEINCITEVK
PYFT
DALHSWQKENQDQKVWAASENQPLHNTTPAQWQPDMSQDCRTPAVFMLVPYRDL
PPIHVNPSPQKPRTVVRHLNTLQLSASKVKALRKSPSGRPVKKEEARKGSTRSQEAN

[0909] ETNKSKWKKQTLFIKSNICHLLEREQKKIGILSSWSV

[0910] mfAPOBEC-4 마카카 파스시쿨라리스(*Macaca fascicularis*)

MEPTYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGT
TY
PQTKHLTFYELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLDSAIYNNDSIRHII
L
YCNNSPCNEANHCCISKVYNFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLA
SL
WPRVVLSPISGGIWHSVLHSFVSGVSGSHVFQPILTGRALTDRYNAYEINAITGVKPFF
T
DVLLHTKRNPNTKAQMALESYPLNNAFPGQSFQMTSGIPPDLRAPVVFVLLPLRDLP
PMHMGQDPNKPRNIIRHLNMPQMSFQETKDLERLPTRRSVETVEITERFASSKQAEE
[0911] KTKKKKGKK

[0912] hAID 호모 사피엔스(*Homo sapiens*)

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
[0913] HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR

LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN
[0914] SVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[0915] clAID 카니스 루푸스 파밀리아리스(*Canis lupus familiaris*)

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHEN
[0916] SVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[0917] btAID 보스 타우루스(*Bos Taurus*)

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTAR
LYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHE
[0918] NSVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[0919] mAID 무스 무스쿨루스(*Mus musculus*)

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN
[0920] SVRLSRQLRRILLPLYEVDDLRDAFRTLGL

[0921] pmCDA-1 페트로미존 마리누스(Petromyzon marinus)

MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYII
NNPNVCHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEE
QGHTLTMHFSRIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEAS
RRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGIPLHLFTLQTPLLSGRVVWWR
[0922] V

[0923] pmCDA-2 페트로미존 마리누스(*Petromyzon marinus*)

MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGH
AVNYNKQGTSIHAEVLLLSAVRAALLRRRRCEDGEEATRGCTLHCYSTYSPCRDCV
EYIQEFGASTGVRVVIHCCRLYELDVNRRRSEAEGVLRSLSRLGRDFRLMGPRDAIA
[0924] LLLGGRLANTADGESGASGNAWVTETNVVEPLVDMTGFGDEDLHAQVQRNKQIRE

AYANYASAVSLMLGELHVDPDKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGPEI

[0925] GRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLIVVMRKHNYQGP

[0926] pmCDA-5 페트로미존 마리누스(*Petromyzon marinus*)

MAGDENVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYII

NNPNVCHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPWLKNLLEE

QGHTLMMHFSRIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEAS

[0927] RRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGMPLHLFT

[0928] yCD 사카로미세스 세레비시애(*Saccharomyces cerevisiae*)

MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKDGSVLGRGHNMRF

QKGSATLHGEISTLENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVVGEN

[0929] VNFKSKGEKYLQTRGHEVVVVDDERCKKIMKQFIDERPQDWFEDIGE

[0930] rAPOBEC-1 (델타 177-186)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT

NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR

LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW

[0931] VRGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK

[0932] rAPOBEC-1 (델타 202-213)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT

NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR

LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW

[0933] VRLYVLELYCIILGLPPCLNILRRKQPQHYQRLPPHILWATGLK

[0934] **실시예 4**: CBE 및 ABE 내부 융합의 구축

[0935] CBE 및 ABE 내부 융합 구축물을 SpCas9 또는 D10A 돌연변이가 있는 SpCas9 닉카아제 내의 높은 b-인자 위치에 데아미나제를 클로닝하여 생성시켰다. 일부 경우에, Cas9의 구조적 또는 기능적 도메인을 부분적으로 결실 또는 결실시키고 TadA 도메인(IBE020)으로 대체했다. 동일한 방식으로 CBE를 삽입하고 C-말단 단부 상에서 우라실 DNA 글리코실라제 억제제(UGI) 도메인으로 변형시켰다.

[0936] 예시적인 내부 융합 염기 편집기가 아래 표 10에 제공되어 있다:

[0937] 표 10:

BE ID	변형	다른 ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09
IBE020	delta HNH TadA 792	ISLAY20
IBE021	N-말단 융합 단일 TadA 나선 절단된 165-단부	ISLAY21
IBE029	TadA-원형 치환체 116 ins1067	ISLAY29
IBE031	TadA- 원형 치환체 136 ins1248	ISLAY31
IBE032	TadA- 원형 치환체 136ins 1052	ISLAY32
IBE035	delta 792-872 TadA ins	ISLAY35
IBE036	delta 792-906 TadA ins	ISLAY36
IBE043	TadA-원형 치환체 65 ins1246	ISLAY43
IBE044	TadA ins C-말단 절단체 2 791	ISLAY44
HR001	GGS-rAPOBEC1-XTEN-ins-부위 1_Y1016-D10A-UGIx2	pHRB-043
HR002	GGS-rAPOBEC1-XTEN ins-부위 2_A1023-D10A-UGIx2	pHRB-044
HR003	GGS-rAPOBEC1-XTEN ins-부위 3 _E1029-D10A-UGIx2	pHRB-045
	GGS-rAPOBEC1-XTEN ins-부위 4_N1040-D10A-UGIx2	pHRB-046
HR004	GGS-rAPOBEC1-XTEN ins-부위 5 -T1069-D10A-UGIx2	pHRB-047
HR005	GGS-rAPOBEC1-XTEN ins-부위 6-G1247-D10A-UGIx2	pHRB-048

[0938]

[0939] 구축물의 서열은 아래에 제공되어 있다.

[0940] Cas9 TadAins 1015

[0941] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC

CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGGGTTCTAGCGGCAGCGAGACTCCCGGGACCT
CAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTA
CTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGT
GCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAA
TAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCG
[0942] ACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTC

ACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGAC
GAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGG
ACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATAT
TGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGT
CTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACTACGACGTGCGGAAGAT
GATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTA
CAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGAT
CCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGG
ATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGA
ATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC
TGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCT
AAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG
CCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTG
GGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTG
GAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAA
GTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGG
CGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCT
GTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCA
GAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCA
GATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGT
GCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAA
TATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTAC
TTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGAC
GCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTG
[0943] TCTCAGCTGGGAGGTGAC

[0944] Cas9 TadAins 1015

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
[0945] KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVGSSGSETPGTSESAT

PESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLH
DPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRN
AKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQ
SSTDYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
[0946] RIDLSQLGGD

[0947] Cas9 TadAins 1022

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
[0948] ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG

AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGGTTCTAGCG
GCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCG
AAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGA
GGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATC
GCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCAC
ATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGAC
TTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGC
TATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACG
GGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGG
GTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTAC
TTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTA
CTGACGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCT
ACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGA
TCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGG
ATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGA
ATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC
TGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCT
AAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG
CCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTG
GGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTG
GAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAA
GTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGG
CGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCT
[0949] GTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCA

GAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCA
GATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGT
GCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAA
TATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTAC
TTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGAC
GCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTG
[0950] TCTCAGCTGGGAGGTGAC

[0951] Cas9 TadAins 1022

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIGSSGSET
PGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEG
WNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIG
RVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVF
NAQKKAQSSTDAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
ETRIDLSQLGGD

[0952]

[0953] IBE003_Cas9: Cas9 TadAins 1029

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
[0954] GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA

AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
[0955] AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG

GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCA
CACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACA
CGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGC
AGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGG
ACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCT
TGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCT
TGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCG
GTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATT
ACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAAT
GTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCA
GAAAAAAGCACAATCCTCTACTGACGGCAAGGCTACCGCCAAGTACTTCTTCTAC
AGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATC
CGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGAT
AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAAT
ATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTG
CCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA
GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCC
AAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGG
GATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGA
AGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTA
CTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGA
ACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTA
CCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAA
ACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGAT
CAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCT
GTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATAT
CATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTT
GACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCC
ACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
AGCTGGGAGGTGAC

[0956]

[0957] Cas9 TadAins 1029

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF

[0958]

DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNN
RVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFF
RMPRQVFNAQKKAQSSTDGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
ETRIDLSQLGGD

[0959]

[0960] IBE004_Cas9: Cas9 TadAins 1040

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
[0961] GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT

CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCGGTTCTAGCG
GCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCG
AAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGA
GGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATC
GCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCAC
ATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGAC
TTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGC
TATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACG
GGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGG
GTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTAC
TTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTA
CTGACAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGAT
CCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGG
ATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGA
ATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC
[0962] TGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCT

AAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG
CCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTG
GGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTG
GAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAA
GTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGG
CGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCT
GTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCA
GAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCA
GATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGT
GCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAA
TATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTAC
TTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGAC
GCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTG
[0963] TCTCAGCTGGGAGGTGAC

[0964] Cas9 TadAins 1040

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDERE
VPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVT
FEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILAD
ECAALLCYFFRMPRQVFNAQKKAQSSTDNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGD

[0965]

[0966] IBE005_Cas9: Cas9 TadAins 1068

[0967] ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG

GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
[0968] CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT

TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACGGCGAAGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGA
GTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGG
ATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCC
GTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGG
GCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAG
GGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGT
TTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGT
TGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTG
CTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCG
GACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTA
ACGCCCAGAAAAAAGCACAATCCTCTACTGACACCGGGGAGATCGTGTGGGATA
AGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATA
TCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGC
CCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAG
AAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCA
AAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGG
ATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAA
GCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTAC
TCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAA
CTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACC
TGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA
GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGT
CCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCA
TCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGA
CACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCAC
CCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCA
[0969] GCTGGGAGGTGAC

[0970] Cas9 TadAins 1608

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
[0971] GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA

DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGEGSSGSETPGTSESATPESSGS
EVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDTG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGD

[0972]

[0973] IBE006_Cas9: Cas9 TadAins 1247

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG

[0974]

GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATT
TTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGA
CCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACA
GCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCT
TCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGG
GCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGC
[0975] TGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA

AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACT
ATGAGAAGCTGAAGGGCGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGT
CCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGAT
GAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGT
GGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGC
AATCGGACTCCACGACCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGG
AGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTT
GAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTG
TATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCT
GCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGA
CGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAAC
GCCCAGAAAAAAGCACAATCCTCTACTGACTCCCCCGAGGATAATGAGCAGAAA
CAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATC
AGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTG
TCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATC
ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTG
ACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCA
CCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCA
[0976] GCTGGGAGGTGAC

[0977] Cas9 TadAins 1247

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGGSSGSETPGTSE
SATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAI
GLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFG
VRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQK
[0978] KAQSSTDSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR

DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
[0979] RIDLSQLGGD

[0980] IBE007_Cas9: Cas9 TadAins 1054

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG

[0981] CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC

[0982]
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGTTCTAGCGGCAGCGAGACTCC
CGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCC
CATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAA
CGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAA
GGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATG
GCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGC
TGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCG
CATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCA
CTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAA
GGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCA
GGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGCGAGATCC
GGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGAT
AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAAT
ATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTG
CCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA
GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCC
AAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGG
GATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGA
AGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTA
CTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGA
ACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTA
CCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAA
ACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGAT
CAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCT
GTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATAT
CATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTT
GACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCC
ACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
AGCTGGGAGGTGAC

[0983] Cas9 TadAins 1054

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGSSGSETPGTSESATPESSGSEVEFSHEYWMRH
ALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
[0984] ETRIDLSQLGGD

[0985] IBE008_Cas9: Cas9 TadAins 1026

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
[0986] CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA

GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGGGTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAA
GTTCTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGAC
TCTCGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGT
GCTCAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGA
CCCCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCA
[0987] GAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATG

TGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCA
ACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCA
TGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGC
TGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGC
ACAATCCTCTACTGACCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTAC
AGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATC
CGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGAT
AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAAT
ATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTG
CCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA
GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCC
AAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGG
GATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGA
AGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTA
CTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGA
ACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTA
CCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAA
ACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGAT
CAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCT
GTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATAT
CATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTT
GACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCC
ACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
[0988] AGCTGGGAGGTGAC

[0989] Cas9 TadAins 1026

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEGSS
GSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVI
GEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIH
[0990] SRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMP

RQVFNAQKKAQSSTDQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
[0991] ETRIDLSQLGGD

[0992] IBE009_Cas9: Cas9 TadAins 768

[0993]

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG

TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGGGTTCTAGCGGCAGCGA
GACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGA
GTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGA
GATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATC
GGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAA
ATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATG
CGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCA
CTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCA
GGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACCACCGGGTAGAAATC
ACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCA
TGCCCAGGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGG
ATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTG
GAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGG
CGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGAT
GTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAG
GTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGA
AGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCT
GATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAG
CGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGAT
CACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGA
GAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGT
GTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTAC
CACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAA
AAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGAC
GTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAA
GTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCC
AACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGA
GATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCAT
GCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAA
AGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGA
CTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTG
CTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAA
AGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCAT
CGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAA
GCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGC
CTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGT
GAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGA
TAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGAT
CATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCT
GGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCA
GGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCC
[0994] TTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAG

GTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGG
[0995] ATCGACCTGTCTCAGCTGGGAGGTGAC

[0996] Cas9 TadAins 768

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALT
LAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN
YRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMN
HRVEITEGILADECAALLCYFFRMPRTTQKGQKNSRERMKRIEEGIKELGSQILKEHP
VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKV
LTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA
KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE
QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
[0997] FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0998] IBE020_delt: 델타 HNH TadA 792

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
[0999] GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC

AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGC
AAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAA
CAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCAC
TGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTA
TCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCG
GGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCA
AGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACC
ACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGT
GTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATC
CTCTACTGACGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACA
GCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCG
GATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGAT
CACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAA
GTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTC
GTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTAC
[1000] GGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGA

AATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC
AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAG
ACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACC
GTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG
CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAG
CTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGC
CCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCC
AAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAG
CAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGT
GAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAA
CGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACT
GGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCAC
AAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG
ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGG
GATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACC
AATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGA
GGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCA
[1001] CCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[1002] 델타 HNH TadA 792

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWM
RHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1003]

[1004] IBE021_N-te: N-말단 융합 단일 TadA 나선 절단된 165-단부

ATGTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCG
CAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCA
ACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCA
CTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATT
ATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGC
GGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCC
AAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAAC
CACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTG
TGTTACTTTTTTCGCATGCCCAGGTCTGGTGGTTCTTCTGGTGGTTCTAGCGGCAG
CGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTGGTTCTTCT
GGTGGTTCTGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTG
GGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTG
CTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTG
TTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAG
AAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAA
CGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCT
GGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGA
CGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACT
GGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCA
CATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAA
CAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTC
GAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCC
AGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAG
AAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCC
AACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAG
GACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTAC
GCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACA
TCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCA
AGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGC
AGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCT
ACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCA
AGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACA
GAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACC
AGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACC
CATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCC
CCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCA
GAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAG
GGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTG
CCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCC
TTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAAC
CGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGC
TTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCA
CATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGG
[1005] AAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACA

GAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAG
TGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGG
AAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTC
CTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGAC
AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGAT
AGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGC
ATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCAC
AAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAA
GGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAG
AGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGA
ACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACC
AGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTC
AGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACA
AGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATG
AAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTC
GACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGG
CTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACA
GATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCG
GGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGA
TTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCC
TACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAA
AGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCC
AAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAAC
ATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAG
CGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGC
CGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTG
AAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAG
AGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTA
CGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTG
GAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCAC
CATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAA
GGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCT
GTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCA
GAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCC
AGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTG
TTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAG
TTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCT
ACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACC
TGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCAC
CATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGAT
CCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGG
[1006] AGGTGAC

[1007] IBE021_N-te: N-말단 융합 단일 TadA 나선 절단된 165-단부

[1008] MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT

GAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRSGGSSGGSSGSETPG
TSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS
IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLAL
AHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD
LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL
TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNE
KVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV
KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIK
KGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL
GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVY
DVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE
QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD

[1009]

IBE029_ISLA: TadA-CP116ins 1067

[1010]

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA

[1011]

GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGG
ACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAA
CGCCCAGAAAAAAGCACAATCCTCTACTGACGGTTCTAGCGGCAGCGAGACTCC
[1012] CGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCC

CATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAA
CGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAA
GGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAAATCATG
GCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGC
TGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCG
CATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCA
CTGATGGACGTGCTGCATTACCCAGGCGGCGAAACCGGGGAGATCGTGTGGGAT
AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAAT
ATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTG
CCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAA
GAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCC
AAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGG
GATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGA
AGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTA
CTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGA
ACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTA
CCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAA
ACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGAT
CAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCT
GTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATAT
CATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTT
GACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCC
ACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTC
[1013] AGCTGGGAGGTGAC

[1014] IBE029_ISLA : TadA-CP116ins 1067

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
[1015] GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNMNHRVEITEGILADECAALLCY

FFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY
RLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
[1016] RIDLSQLGGD

[1017] IBE031_ISLA: TadACP136ins 1248

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
[1018] ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG

AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATT
TTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGA
CCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACA
GCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCT
TCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGG
GCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGC
TGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACT
ATGAGAAGCTGAAGGGCTCCATGAACCACCGGGTAGAAATCACAGAAGGCATAT
TGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGT
CTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGTTCTAGCGGCAGCGA
GACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGA
GTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGA
GATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATC
GGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCACATGCGGAA
ATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGACTTATCGATG
CGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAGCTATGATTCA
CTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGACGGGTGCCGCA
GGTTCACTGATGGACGTGCTGCATTACCCAGGCCCCGAGGATAATGAGCAGAAA

[1019]

CAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATC
AGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTG
TCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATC
ATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTG
ACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCA
CCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCA

[1020]

GCTGGGAGGTGAC

[1021] IBE031_ISLA : TadACP136ins 1248

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSH
EYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL
RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHYPGPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGD

[1022]

[1023] IBE032_ISLA: TadACP136ins 1052

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC

[1024]

AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
[1025] GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC

TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCATGAACCACCGGGTAGAAATCACAG
AAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCC
CAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATCCTCTACTGACGGTTCTAG
CGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTC
CGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGCAAAG
AGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAAT
CGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCACTGCA
CATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTATCGA
CTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCGGGAG
CTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCAAGAC
GGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCAACGGCGAGAT
CCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGG
ATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGA
ATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCC
TGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCT
AAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG
CCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTG
GGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTG
GAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAA
GTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGG
CGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCT
GTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCA
GAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCA
GATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGT
GCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAA
TATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTAC
TTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGAC
GCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTG
[1026] TCTCAGCTGGGAGGTGAC

[1027] IBE032_ISLA : TadACP136ins 1052

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
[1028] GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA

DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLAMNHRVEITEGILADECAALLCYFFRMPRQVFNA
QKKAQSSTDGSSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFE
PCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGNGEIRKRPLIETNGET
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
[1029] ETRIDLSQLGGD

[1030] IBE035_ISLA: 델타 792-872 TadAins

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
[1031] TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG

GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGC
AAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAA
CAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCAC
TGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTA
TCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCG
GGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCA
AGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACC
ACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGT
GTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATC
CTCTACTGACGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCT
GAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAG
AGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGA
AACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACAC
TAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAA
GTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAG
ATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACC
GCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTAC
AAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAA
GGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAG
ATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGC
GAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAA
[1032] GTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGC

GGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCC
AGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTG
GCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTG
AAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAG
AAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGAC
CTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAG
AGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCC
TCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCT
CCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACC
TGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCG
ACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCA
TCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAG
CCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAG
CACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTA
[1033] CGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[1034] IBE035_ISLA : 델타 792-872 TadAins

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWM
RHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDEEVVKKMKNYW
RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP
KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
[1035] VLDATLIHQSITGLYETRIDLSQLGGD

[1036] IBE036_ISLA : 델타 792-906 TadAins

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
[1037] GCTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCTCGC

AAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCTCAA
CAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCCCAC
TGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAATTA
TCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGCGCG
GGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACGCCA
AGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGAACC
ACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGTTGT
GTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCCCAGAAAAAAGCACAATC
CTCTACTGACGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCT
GGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGAT
GAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCAC
CCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTG
CGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTG
GGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGC
GACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAAT
CGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAG
ACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACA
AACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTG
CGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAG
ACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTG
ATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCC
ACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAG
AAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAG
CTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAA
AAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGG
CCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGC
CCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTG
AAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAG
CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATC
CTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGAT
AAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAAT
CTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGT
ACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCG
[1038] GCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGAC

[1039] IBE043_ISLA: 델타 792-906 TadAins

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
[1040] AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT

VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSEVEFSHEYWM
RHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHY
PGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMP
QVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAP
[1041] AAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1042] IBE043_ISLA: TadA CP65ins 1246

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
[1043] GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT

CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATT
TTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGA
CCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACA
GCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCT
TCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGG
GCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGC
TGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACT
ATGAGAAGCTGAAGGGCACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAG
GGCTTGTGATGCAGAATTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGA
ACCTTGCGTAATGTGCGCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTA
[1044] TTCGGTGTTCGCAACGCCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGC

ATTACCCAGGCATGAACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACG
AATGTGCGGCGCTGTTGTGTTACTTTTTTCGCATGCCCAGGCAGGTCTTTAACGCC
CAGAAAAAAGCACAATCCTCTACTGACGGTTCTAGCGGCAGCGAGACTCCCGGG
ACCTCAGAGTCCGCCACACCCGAAAGTTCTGGTTCCGAAGTCGAGTTTTCCCATG
AGTACTGGATGAGACACGCATTGACTCTCGCAAAGAGGGCTCGAGATGAACGCG
AGGTGCCCGTGGGGGCAGTACTCGTGCTCAACAATCGCGTAATCGGCGAAGGTT
GGAATAGGGCAATCGGACTCCACGACCCCTCCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA
GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGT
CCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCA
TCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGA
CACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCAC
CCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCA
[1045] GCTGGGAGGTGAC

[1046] IBE043_ISLA : TadA CP65ins 1246

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGTAHAEIMALRQ
GGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTDGSSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNR
AIGLHDPSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR
DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
[1047] RIDLSQLGGD

[1048] IBE044_ISLA: TadAins C-말단 절단체2 791

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG

[1049] GTTCTAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAAAGTT

CTGGTTCCGAAGTCGAGTTTTCCCATGAGTACTGGATGAGACACGCATTGACTCT
CGCAAAGAGGGCTCGAGATGAACGCGAGGTGCCCGTGGGGGCAGTACTCGTGCT
CAACAATCGCGTAATCGGCGAAGGTTGGAATAGGGCAATCGGACTCCACGACCC
CACTGCACATGCGGAAATCATGGCCCTTCGACAGGGAGGGCTTGTGATGCAGAA
TTATCGACTTATCGATGCGACGCTGTACGTCACGTTTGAACCTTGCGTAATGTGC
GCGGGAGCTATGATTCACTCCCGCATTGGACGAGTTGTATTCGGTGTTCGCAACG
CCAAGACGGGTGCCGCAGGTTCACTGATGGACGTGCTGCATTACCCAGGCATGA
ACCACCGGGTAGAAATCACAGAAGGCATATTGGCGGACGAATGTGCGGCGCTGT
TGTGTTACTTTTTTCGCATGCCCAGGCAGGGCAGCCAGATCCTGAAAGAACACCC
CGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAA
TGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTA
CGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAAC
AAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTC
CGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCA
AGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCC
TGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGC
AGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACG
ACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGC
TGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAA
CTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATC
AAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTAC
GACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGC
CAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTG
GCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGG
GGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAG
CATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAG
CAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAA
GGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTC
TGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGT
GAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCC
CATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCAT
CAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCT
GGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATA
TGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAG
GATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAG
ATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAAT
CTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAG
CAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCG
CCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAG
AGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACAC

[1050] GGATCGACCTGTCTCAGCTGGGAGGTGAC

[1051] IBE044_ISLA : TadAins C-말단 절단체2 791

[1052] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE

RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSSGSETPGTSES
ATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIG
LHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQGSQILKEH
PVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNK
VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA
KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIK
LPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE
QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[1053]

[1054] pHRB-043_GGS-rAP: rAPOBEC1-XTEN-ins-부위1_Y1016-D10A-UGIx2

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT

[1055]

ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGGAGGCTCTGGAGGAAGCAGCTCTGAGA
CAGGACCTGTGGCCGTGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGAGT
TCGAGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACG
AGATCAACTGGGGCGGCAGACACTCTATCTGGCGGCACACAAGCCAGAACACCA
ACAAGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGCGGTACTTCT
GCCCCAACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGA
GTGCAGCAGAGCCATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTC
[1056] ATCTATATCGCCCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGGCCTGC

GCGACCTGATCAGCAGCGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCG
GCTACTGCTGGCGGAACTTCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGC
CTAGATATCCTCACCTGTGGGTCCGACTGTACGTGCTGGAACTGTACTGCATCAT
CCTGGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGAC
CTTCTTCACAATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATC
CTGTGGGCCACCGGACTTAAGGGCTCTTCTGGATCTGAAACACCTGGCACAAGTG
AGAGCGCCACCCCTGAGAGCTCTGGCGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATT
TTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGA
CCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACA
GCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCT
TCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGG
GCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGC
TGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACT
ATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGG
AACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCA
AGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
AGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTA
CCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGA
CCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCA
GAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGA
CTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGA
GAAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGA
AGAGGTGGAAGAAGTGATCGGCAACAAGCCCGAGTCCGACATCCTGGTGCACAC
CGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTGACGCCCC
TGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGAT
CAAGATGCTGAGCGGCGGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGAT
ATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCAAGAGTCTATTCTCATG
CTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTG
GTCCATACAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCG
ACGCTCCCGAGTATAAGCCATGGGCACTTGTCATTCAGGACTCCAATGGGGAAA
ACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAAG
CGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAA
CCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGT
GCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCC
TGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCA
TTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA
GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTAT
GGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTA
GAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCA
[1057] CAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCT

AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG
CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGG
TCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC
CACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA
AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGAC
AGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCT
CCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAG
TTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACG
CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCT
ATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTA
GATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACC
GCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGG
AAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATT
AATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACG
TTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCA
TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC
AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCG
CCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
AAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTG
CACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAA
AACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGAT
CTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTT
AAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGC
AAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTG
CTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTT
GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
[1058] TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC

TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGAC
GTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGG
ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG
CGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC
CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAG
GCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGAT
[1059] CCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1060] pHRB-043_GGS-rAP : rAPOBEC1-XTEN-ins-부위1_Y1016-D10A-UGIx2

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYGGSGGSSSETGPV
AVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNF
IEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPR
NRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLEL
YCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSGSETPGTSES
ATPESSGDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
ETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES
DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGST
NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD

[1061] APEYKPWALVIQDSNGENKIKMLPKKKRKVEGADKRTADGSEFESPKKKRKV*

[1062] pHRB-044_GGS-rAP: rAPOBEC1-XTEN ins-부위2_A1023-D10A-UGIx2

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG

[1063] GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG

AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCGGAGGCT
CTGGAGGAAGCAGCTCTGAGACAGGACCTGTGGCCGTGGATCCCACACTGCGGA
GAAGAATTGAGCCCCACGAGTTCGAGGTGTTCTTCGACCCCAGAGAGCTGCGGA
AAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACACTCTATCTGGC
GGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGT
TTACGACCGAGCGGTACTTCTGCCCCAACACCAGATGCAGCATCACCTGGTTTCT
GAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGCCATCACCGAGTTTCTGTCCAGA
TATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGATC
CTAGAAATAGACAGGGCCTGCGCGACCTGATCAGCAGCGGAGTGACAATCCAGA
TCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACTTCGTGAACTACAGCC
CCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCGACTGTACGT
GCTGGAACTGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGA
AGAAAGCAGCCTCAGCTGACCTTCTTCACAATCGCCCTGCAGAGCTGCCACTACC
AGAGACTGCCTCCACACATCCTGTGGGCCACCGGACTTAAGGGCTCTTCTGGATC
TGAAACACCTGGCACAAGTGAGAGCGCCACCCCTGAGAGCTCTGGCAAGAGCGA
GCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAA
CTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTG
ATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTT
GCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACC
GAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGC
GATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTC
GACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGC
AAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA
AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAA
GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTG
GAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAA
CGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTAT
GAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAA
CAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAG
AGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAG
CACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACC
CTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACC
GGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGA
GCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACT
[1064] CTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAGA

AAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAG
AGGTGGAAGAAGTGATCGGCAACAAGCCCGAGTCCGACATCCTGGTGCACACCG
CCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTGACGCCCCTG
AGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCA
AGATGCTGAGCGGCGGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATAT
TATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCAAGAGTCTATTCTCATGCTC
CCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTC
CATACAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGAC
GCTCCCGAGTATAAGCCATGGGCACTTGTCATTCAGGACTCCAATGGGGAAAAC
AAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAAGCG
CACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACC
GGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGC
CTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTG
GAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATT
GTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG
CTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA
ATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAA
TGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGG
GAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG
GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTC
GTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCC
ACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAA
GGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCC
CCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTG
GCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAG
CAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT
TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTG
CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGC
AAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGC
GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGC
TCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGT
ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGA
AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA
ATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGT
TGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCAT
[1065] TCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAA

AAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATC
CGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAG
TGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC
CACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAA
AACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGC
ACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAA
ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTC
TCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCC
GCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATC
TCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCA
AAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGC
TTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTG
ACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCAT
AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACG
TCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGA
CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGC
GGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCC
AAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG
GACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGG
CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATC

[1066] CGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1067] pHRB-044_GGS-rAP : rAPOBEC1-XTEN ins-부위2_A1023-D10A-UGIx2

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
[1068] DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG

FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAGGSGG
SSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT
NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR
LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW
VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSG
SETPGTSESATPESSGKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG
ETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF
LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYL
ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNK
HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGL
YETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKP
ESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGS
TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
[1069] DAPEYKPWALVIQDSNGENKIKMLPKKKRKVEGADKRTADGSEFESPKKKRKV*

[1070] pHRB-045_GGS-rAP: rAPOBEC1-XTEN ins-부위3_E1029-D10A-UGIx2

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA

[1071] GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT

CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGAGGCTCTGGAGGAAGCAGCTCTGAGACAGGACCTGTGGCCG
TGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGAGTTCGAGGTGTTCTTCG
ACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAACTGGGGCG
GCAGACACTCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCACGTGGAAG
TGAACTTTATCGAGAAGTTTACGACCGAGCGGTACTTCTGCCCCAACACCAGATG
CAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGCCATC
ACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCT
GTACCACCACGCCGATCCTAGAAATAGACAGGGCCTGCGCGACCTGATCAGCAG
CGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAA
CTTCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTG
TGGGTCCGACTGTACGTGCTGGAACTGTACTGCATCATCCTGGGCCTGCCTCCAT
GCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACCTTCTTCACAATCGCCCT
GCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCCACCGGACTT
AAGGGCTCTTCTGGATCTGAAACACCTGGCACAAGTGAGAGCGCCACCCCTGAG
AGCTCTGGCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACT
[1072] TTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGA

TCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTG
CCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCG
AGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCG
ATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCG
ACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCA
AGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAA
GAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAG
AAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGG
AAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAAC
GAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATG
AGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAAC
AGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGA
GAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGC
ACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCC
TGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCG
GAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAG
CATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACTCT
GGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAGAAA
GAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAGAG
GTGGAAGAAGTGATCGGCAACAAGCCCGAGTCCGACATCCTGGTGCACACCGCC
TACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTGACGCCCCTGAG
TACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCAAG
ATGCTGAGCGGCGGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATATTA
TTGAGAAAGAAACCGGGAAACAGCTCGTGATTCAAGAGTCTATTCTCATGCTCCC
GGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTCCA
TACAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGACGCT
CCCGAGTATAAGCCATGGGCACTTGTCATTCAGGACTCCAATGGGGAAAACAAA
ATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAAGCGCAC
CGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACCGGT
CATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTT
CTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAA
GGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGG
AGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTT
CTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCT
TGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATT
CCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGA
GTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAA
ACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTT
CGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACA
GAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCT
GACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG
ACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
[1073] CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGG

CGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCC
AAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCG
GTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGC
AGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTT
CTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGC
GCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
AACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCG
CAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCT
CAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGG
ATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTA
TATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTAT
CTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGA
TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGC
GAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAA
GGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA
TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTT
GTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATT
CAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAA
AAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCC
GTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGT
GTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCC
ACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAA
ACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCA
CCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAA
CAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGA
ATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCT
CATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCC
GCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATC
TCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCA
AAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGC
TTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTG
ACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCAT
AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACG
TCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGA
CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGC
GGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCC
AAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG
GACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGG
CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATC
[1074] CGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1075] pHRB-045_GGS-rAP : rAPOBEC1-XTEN ins-부위3_E1029-D10A-UGIx2

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GGSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRH
TSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTL
FIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPR
YPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGL
KGSSGSETPGTSESATPESSGGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
ETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES
DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGST
NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
[1076] APEYKPWALVIQDSNGENKIKMLPKKKRKVEGADKRTADGSEFESPKKKRKV*

[1077] pHRB-046_GGS-rAP: rAPOBEC1-XTEN ins-부위4_N1040-D10A-UGIx2

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
[1078] AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC

GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
[1079] ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT

TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACGGAGGCT
CTGGAGGAAGCAGCTCTGAGACAGGACCTGTGGCCGTGGATCCCACACTGCGGA
GAAGAATTGAGCCCCACGAGTTCGAGGTGTTCTTCGACCCCAGAGAGCTGCGGA
AAGAGACATGCCTGCTGTACGAGATCAACTGGGGCGGCAGACACTCTATCTGGC
GGCACACAAGCCAGAACACCAACAAGCACGTGGAAGTGAACTTTATCGAGAAGT
TTACGACCGAGCGGTACTTCTGCCCCAACACCAGATGCAGCATCACCTGGTTTCT
GAGCTGGTCCCCTTGCGGCGAGTGCAGCAGAGCCATCACCGAGTTTCTGTCCAGA
TATCCCCACGTGACCCTGTTCATCTATATCGCCCGGCTGTACCACCACGCCGATC
CTAGAAATAGACAGGGCCTGCGCGACCTGATCAGCAGCGGAGTGACAATCCAGA
TCATGACCGAGCAAGAGAGCGGCTACTGCTGGCGGAACTTCGTGAACTACAGCC
CCAGCAACGAAGCCCACTGGCCTAGATATCCTCACCTGTGGGTCCGACTGTACGT
GCTGGAACTGTACTGCATCATCCTGGGCCTGCCTCCATGCCTGAACATCCTGAGA
AGAAAGCAGCCTCAGCTGACCTTCTTCACAATCGCCCTGCAGAGCTGCCACTACC
AGAGACTGCCTCCACACATCCTGTGGGCCACCGGACTTAAGGGCTCTTCTGGATC
TGAAACACCTGGCACAAGTGAGAGCGCCACCCCTGAGAGCTCTGGCATCATGAA
CTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTG
ATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTT
GCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACC
GAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGC
GATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTC
GACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGC
AAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAA
AGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAA
GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTG
GAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAA
CGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTAT
GAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAA
CAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAG
AGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAG
CACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACC
CTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACC
GGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGA
GCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGACT
CTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGAGA
AAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGAAG
AGGTGGAAGAAGTGATCGGCAACAAGCCCGAGTCCGACATCCTGGTGCACACCG
CCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTGACGCCCCTG
AGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGATCA
AGATGCTGAGCGGCGGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGATAT
TATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCAAGAGTCTATTCTCATGCTC
CCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTGGTC
CATACAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCGAC
GCTCCCGAGTATAAGCCATGGGCACTTGTCATTCAGGACTCCAATGGGGAAAAC
AAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAAGCG
[1080] CACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAACC

GGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGC
CTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTG
GAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATT
GTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG
CTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGA
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA
ATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAA
TGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGG
GAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG
GTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTC
GTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCC
ACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAA
GGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCC
CCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTG
GCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAG
CAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT
TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTG
CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGC
AAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGC
GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGC
TCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGT
ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGA
AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA
ATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGT
TGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCAT
TCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAA
AAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATC
CGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAG
TGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC
CACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAA
AACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGC
ACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAA
ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTC
TCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCC
[1081] GCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATC

TCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCA
AAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGC
TTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTG
ACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCAT
AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCT
GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACG
TCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGA
CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGC
GGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCC
AAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG
GACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGG
CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATC
[1082] CGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1083] pHRB-046_GGS-rAP : rAPOBEC1-XTEN ins-부위4_N1040-D10A-UGIx2

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNGGSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEI
NWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAI
TEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVN
YSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQ
RLPPHILWATGLKGSSGSETPGTSESATPESSGIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
[1084] c SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH

RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
ETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES
DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGST
NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
[1085] APEYKPWALVIQDSNGENKIKMLPKKKRKVEGADKRTADGSEFESPKKKRKV*

[1086] pHRB-047_GGS-rAP: rAPOBEC1-XTEN ins-부위5-T1069-D10A-UGIx2

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT
ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
[1087] ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG

TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACGGCGAAACCGGAGGCTCTGGAGGAAGCAGCTCTGAGACAG
GACCTGTGGCCGTGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGAGTTCG
AGGTGTTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGA
TCAACTGGGGCGGCAGACACTCTATCTGGCGGCACACAAGCCAGAACACCAACA
AGCACGTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGCGGTACTTCTGCCC
CAACACCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGC
AGCAGAGCCATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCATCT
ATATCGCCCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGGCCTGCGCG
ACCTGATCAGCAGCGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCGGCT
ACTGCTGGCGGAACTTCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGCCTA
GATATCCTCACCTGTGGGTCCGACTGTACGTGCTGGAACTGTACTGCATCATCCT
GGGCCTGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACCTTC
TTCACAATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGT
GGGCCACCGGACTTAAGGGCTCTTCTGGATCTGAAACACCTGGCACAAGTGAGA
GCGCCACCCCTGAGAGCTCTGGCGGGGAGATCGTGTGGGATAAGGGCCGGGATT
TTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGA
CCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACA
GCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCT
TCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGG
GCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGC
TGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACT
[1088] ATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGG

AACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCA
AGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
AGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTA
CCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGA
CCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCA
GAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTGA
CTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCGA
GAAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTGA
AGAGGTGGAAGAAGTGATCGGCAACAAGCCCGAGTCCGACATCCTGGTGCACAC
CGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTGACGCCCC
TGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGAT
CAAGATGCTGAGCGGCGGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGAT
ATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCAAGAGTCTATTCTCATG
CTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTG
GTCCATACAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCG
ACGCTCCCGAGTATAAGCCATGGGCACTTGTCATTCAGGACTCCAATGGGGAAA
ACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAAG
CGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAA
CCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGT
GCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCC
TGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCA
TTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA
GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTAT
GGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTA
GAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCA
CAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG
CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGG
TCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC
CACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA
AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGAC
AGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCT
CCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAG
TTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACG
CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
[1089] TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCT

ATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTA
GATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACC
GCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGG
AAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATT
AATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACG
TTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCA
TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC
AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCG
CCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
AAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTG
CACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAA
AACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGAT
CTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTT
AAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGC
AAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTG
CTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTT
GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC
TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGAC
GTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGG
ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG
CGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC
CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAG
GCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGAT
[1090] CCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1091] pHRB-047_GGS-rAP : rAPOBEC1-XTEN ins-부위5-T1069-D10A-UGIx2

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
[1092] AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT

VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGGSGGSSSETGPVAVDPTL
RRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTT
ERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGL
RDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGL
PPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSGSETPGTSESATPESSG
GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL
EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA
SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
ETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES
DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGST
NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
[1093] APEYKPWALVIQDSNGENKIKMLPKKKRKVEGADKRTADGSEFESPKKKRKV*

[1094] pHRB-048_GGS-rAP: rAPOBEC1-XTEN ins-부위6-G1247-D10A-UGIx2

ATGGACAAGAAGTACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGG
GCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC
AACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGAC
AGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATA
CACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGAT
GGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGA
AGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGT
GGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGA
CAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATC
AAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAA
AACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTG
AGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAA
GAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTC
AAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACC
TACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGAC
CTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGA
[1095] GAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGAT

ACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGC
TGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCG
GCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCA
TCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCC
TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG
AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
TCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
AAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
ACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATT
TTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGA
CCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACA
GCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCT
[1096] TCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGG

GCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGG
AAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACA
AAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGC
TGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGA
AACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACT
ATGAGAAGCTGAAGGGCGGAGGCTCTGGAGGAAGCAGCTCTGAGACAGGACCT
GTGGCCGTGGATCCCACACTGCGGAGAAGAATTGAGCCCCACGAGTTCGAGGTG
TTCTTCGACCCCAGAGAGCTGCGGAAAGAGACATGCCTGCTGTACGAGATCAAC
TGGGGCGGCAGACACTCTATCTGGCGGCACACAAGCCAGAACACCAACAAGCAC
GTGGAAGTGAACTTTATCGAGAAGTTTACGACCGAGCGGTACTTCTGCCCCAACA
CCAGATGCAGCATCACCTGGTTTCTGAGCTGGTCCCCTTGCGGCGAGTGCAGCAG
AGCCATCACCGAGTTTCTGTCCAGATATCCCCACGTGACCCTGTTCATCTATATC
GCCCGGCTGTACCACCACGCCGATCCTAGAAATAGACAGGGCCTGCGCGACCTG
ATCAGCAGCGGAGTGACAATCCAGATCATGACCGAGCAAGAGAGCGGCTACTGC
TGGCGGAACTTCGTGAACTACAGCCCCAGCAACGAAGCCCACTGGCCTAGATAT
CCTCACCTGTGGGTCCGACTGTACGTGCTGGAACTGTACTGCATCATCCTGGGCC
TGCCTCCATGCCTGAACATCCTGAGAAGAAAGCAGCCTCAGCTGACCTTCTTCAC
AATCGCCCTGCAGAGCTGCCACTACCAGAGACTGCCTCCACACATCCTGTGGGCC
ACCGGACTTAAGGGCTCTTCTGGATCTGAAACACCTGGCACAAGTGAGAGCGCC
ACCCCTGAGAGCTCTGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTG
GAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCC
AAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAAC
AAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTT
ACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCG
ACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACC
AGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGTG
ACTCTGGTGGAAGCGGAGGATCTGGCGGCAGCACCAATCTGAGCGACATCATCG
AGAAAGAGACAGGCAAGCAGCTGGTCATCCAAGAGTCCATCCTGATGCTGCCTG
AAGAGGTGGAAGAAGTGATCGGCAACAAGCCCGAGTCCGACATCCTGGTGCACA
CCGCCTACGATGAGAGCACCGACGAGAACGTGATGCTGCTGACCTCTGACGCCC
CTGAGTACAAGCCTTGGGCTCTCGTGATCCAGGACAGCAACGGCGAGAACAAGA
TCAAGATGCTGAGCGGCGGCTCTGGTGGCTCTGGCGGATCTACAAACCTGTCCGA
TATTATTGAGAAAGAAACCGGGAAACAGCTCGTGATTCAAGAGTCTATTCTCATG
CTCCCGGAAGAAGTCGAGGAAGTCATTGGAAACAAGCCTGAGAGCGATATTCTG
GTCCATACAGCCTACGACGAGTCTACCGATGAGAATGTCATGCTCCTCACCAGCG
ACGCTCCCGAGTATAAGCCATGGGCACTTGTCATTCAGGACTCCAATGGGGAAA
ACAAAATCAAAATGCTCCCAAAGAAAAAACGCAAGGTGGAGGGAGCTGATAAG
CGCACCGCCGATGGTTCCGAGTTCGAAAGCCCCAAGAAGAAGAGGAAAGTCTAA
CCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGT
GCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCC
TGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCA
TTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA
GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTAT
GGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTA
GAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCA
[1097] CAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCT

AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG
CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGG
TCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC
CACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA
AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGAC
AGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCT
CCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAG
TTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGG
CAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGCAGCAGATTACG
CGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCT
ATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTA
GATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACC
GCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGG
AAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATT
AATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACG
TTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCA
TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC
AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCG
CCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA
AAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTG
CACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAA
AACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGAT
CTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTT
AAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGC
AAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTG
CTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTT
GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
[1098] TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC

TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGAC
GTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGG
ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG
CGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC
CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAG
GCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGAT
[1099] CCGCTAGAGATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACC

[1100] pHRB-048_GGS-rAP : rAPOBEC1-XTEN ins-부위6-G1247-D10A-UGIx2

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRF
AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEL
VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS
DKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG
FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY
KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI
GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGGGSGGSSSETG
PVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEV
NFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHAD
PRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKGSSGSETPGTSE
SATPESSGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY
ETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPES
DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGST
NLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
[1101] APEYKPWALVIQDSNGENKIKMLPKKKRKVEGADKRTADGSEFESPKKKRKV*

[1102] **실시예 5**: ABE 내부 융합 염기 편집기

[1103] 세포에서 염기 편집을 평가하기 위해 HEK293T 세포를 Lipofectamine 2000(Life Technologies)을 사용하여 100 ng의 sgRNA 코딩 플라스미드 및 염기 편집기 코딩 플라스미드로 공동-형질감염시켰다. 4일 후 게놈 DNA를 분리하고 PCR을 통해 표적 게놈 영역을 증폭시켰다. PCR 생성물 라이브러리를 생성하기 위해 시퀀싱 어댑터를 추가했다. 염기 편집 영역을 포함하는 준비된 PCR 라이브러리를 Illumina MiSeq를 통해 시퀀싱했다.

[1104] IBE003, IBE008, IBE007, IBE002, IBE001, IBE005, IBE006, IBE004, IBE021, IBE031, IBE020, IBE036, IBE035, IBE028, 및 IBE009를 사용하여 2개의 다른 표적 서열, HEK2(GAACACAAAGCATAGACTGC) 및 T39(GGACAGCTTTCCTAGACAG)에서의 편집 백분율을 측정했다(**도 20c-q**).

[1105] **실시예 6**: 내부 융합 염기 편집기 효율.

[1106] HEK293T를 Lipofectamine 2000(Life Technologies)을 사용하여 100 ng의 sgRNA-코딩 플라스미드와 염기 편집기 코딩 플라스미드로 공동-형질감염시켰다. 4일 후, 게놈 DNA를 분리하고 PCR을 통해 표적 게놈 영역을 증폭시켰다. PCR 생성물 라이브러리를 생성하기 위해 시퀀싱 어댑터를 추가했다. 염기 편집 영역을 포함하는 준비된 PCR 라이브러리를 Illumina MiSeq를 통해 시퀀싱했다. 이 실험에 사용된 표적 부위는 아래 표 11에 나타나 있다.

[1107] 표 11:

명칭	서열
표적 1	GGCCTCCGTATCACTCTCTGACTGGGGT
표적 2	GTAACTGAACCCCTGCAATCAATGGGAT
표적 3	GCCACAGACTTTTCCATTTGCAGGAGT
표적 4	GCGAAAGGCTCGCGGCGAAGGAAGGAAT
표적 5	GCCTGGCAGATGAGAACCAGGAGGAAT
표적 6	GTATTACTATTATTATCTGAGATGGGGT
표적 7	GCCACAGTGGGAGGGGACATGGGGAAT
표적 8	GCCCTGATCTGCACTGAACAGAGGGGT
표적 9	GCCTCAAGTCTGGTTATTTTAGGGGGAT
표적 10	GGTCGACCCTTGGTATCCATGGGGGAT
표적 11	GAAAGAGACAGAGAAGGGGCAGGGGGT
표적 12	GAGTGGGAACTTTCTGATGCCATGGAAT
표적 13	GTGGGACTGATCCCTTAATGTGTGGGGT
표적 14	GCCCAGCTCCAGCCTCTGATGAGGGGT
표적 15	GAAGGCTTTACTGTATTACAGAAGGGGT
표적 16	GGAGCCAGAGACCAGTGGGCAGGGGGT
표적 17	GCTTTCCTTAGCTGTAAAAGAAAGGGAT
표적 18	GAGAAGAAACCAGGGAACAGGTAGGAGT
표적 19	GGCCTCCGTATCACTCTCTGACTGGGGT
표적 20	GATGTGTCTACTGTTACTTACAAGGAAT
표적 21	GACCAGGTCAGCAAACATGTTTGGAAT
표적 22	GCACCCAGGGGTTCTGCAGAGCAGGGAT
표적 23	GCCCAGCAATTCACTGTGAAGAGGGAT
표적 24	GACCAAAACGAGGGACATTTAGGGGAT
표적 25	GCTCCTCTCACCCTTATGACTCAGGGAT
표적 26	GACTCAGCGCCCCTGCCGGGCCTGGGAT
표적 27	GGTCGTAGCCAGTCCGAACCCCGGAGT
표적 28	GCATTCCACTCCGTCCGCCTCCGGAGT
표적 29	GGGTACCTGAGTGGGGTGCATTTGGGGT

[1108]

[1109] 시퀀싱 판독물을 원래 표적 서열에 맞춰 정렬시키고 편집 백분율을 측정했다. 도 22a를 참조하면, 29개의 게놈 표적에서 ABE7.10과 비교하여 내부 융합물의 효율을 조사했다. 편집 효율을 최상의 위치(위치 14, 이 그래프에서 20은 PAM에서 가장 멀고 1은 PAM에 가장 가까움)에서 ABE7.10 편집에 대해 정규화시켰다. 모든 부위에 걸쳐 내부 염기 편집기의 최대 편집 효율은 ABE7.10의 최대 편집 효율에 대해 정규화된다(도 22b). ABE7.10에 대해 정규화된 최대 편집 효율을 기반으로 한 유효 염기 편집 윈도우는 ABE7.10과 비교하여 내부 융합 A 염기 편집기에서 변경된 최대 편집 윈도우를 나타낸다(도 22c, d).

[1110] **실시예 7**: 내부 융합 염기 편집기의 불요성 탈아미노화 평가

[1111] 가이드가 있는 염기 편집기를 SaCas9 및 다양한 유전자좌를 표적으로 하는 가이드와 함께 형질감염시켰다. SaCas9는 불요성 염기 편집에 의해 탈아미노화되는 ssDNA를 생성시킨다(불요성 = 가이드-표적화된 것이 아님). SaCas9 표적 유전자좌를 시퀀싱하여 불요성 탈아미노화를 측정했다. 트랜스-편집을 각 부위에서 ABE7.10 트랜스 편집에 대해 정규화시켰다. 29개의 다른 IBE 표적에 걸친 불요성 탈아미노화를 비교를 위한 각 부위에서 ABE7.10 트랜스-편집 부위에 대해 정규화된 트랜스-편집 분석으로 측정했다. 전체 트랜스-편집을 해당 부위에서 ABE7.10 트랜스-편집에 대해 정규화하기 전에 부위별로 합산했다. 테스트된 내부 염기 편집기(IBE002, IBE004, IBE005, IBE006, IBE008, IBE009, IBE020)는 ABE7.10과 비교하여 감소된 평균 불요성 탈아미노화를 나타냈다.

[1112] **실시예 8**: 내부 융합 A 염기 편집기의 염기 편집 평가

[1113] ABE 내부 융합물을 이용한 염기 편집을 고-처리량 시퀀싱을 이용하여 실시예 6에 기술된 바와 같이 HEK293T 세포를 사용하여 평가했다. 이 분석에서, 가이드가 6개의 상이한 부위 HEK4, GGCACTGCGGCTGGAGGTGG(도 24a); FANCF, GTAGGGCCTTCGCGCACCTCA(도 24b); HEK-3, GGCCCAGACTGAGCACGTGA(도 24c); HEK2-YY,

GGAAACCTTGAATAAGAATGGA(도 24d); EMX1, GAGTCCGAGCAGAAGAAGAA(도 24e), 및 HEK2, GAACACAAAGCATAGACTGC(도 24f)를 표적으로 하도록 설계했다.

[1114] 10,000개 내지 20,000개 HEK293T 세포를 웰당 시딩했다. 75 ng의 sgRNA 및 175 ng의 염기 편집기 또는 Cas9 플라스미드를 1㎕의 Lipofectamine 2000으로 형질감염시켰다. 형질감염 4일 후, 게놈 DNA를 분리하고 표적 부위를 PCR 증폭하고 Illumina MiSeq를 통해 시퀀싱했다. 편집 백분율은 표시된 위치에서 A에서 G로의 돌연변이가 있는 40,000개의 Illumina 시퀀싱 판독물의 백분율로 계산했다. 내부 융합 아데노신 염기 편집기는 ABE7.10과 비교하여 상이한 최대 편집 윈도우와 감소된 표적-이탈 편집을 나타낸다. (도 24a-f).

[1115] **실시예 9**: 내부 융합 C 염기 편집기의 염기 편집에 대한 평가.

[1116] CBE 내부 융합을 이용한 염기 편집을 고-처리량 시퀀싱을 이용하여 실시예 6에 기술한 바와 같이 HEK293T 세포를 사용하여 평가했다. 이 분석에서 다음 CBE 염기 편집기를 사용했다: , BE4, HR001, HR002, HR003, HR004, HR005. 이 분석에서, 가이드를 6개의 상이한 부위 HEK4, GGCACTGCGGCTGGAGGTGG(도 25a); FANCF, GTAGGGCCTTCGCGCACCTCA(도 25b); HEK-3, GGCCCAGACTGAGCACGTGA(도 25c); HEK2-YY, GGAAACCTTGAATAAGAATGGA(도 25d); EMX1, GAGTCCGAGCAGAAGAAGAA(도 25e) 및 HEK2, GAACACAAAGCATAGACTGC(도 25f)를 표적으로 하도록 설계했다. 편집 백분율을 표시된 위치에서 C에서 T로의 돌연변이가 있는 40,000개의 Illumina 시퀀싱 판독물의 백분율로 계산했다. 내부 융합 시티딘 염기 편집기는 ABE7.10과 비교하여 상이한 최대 편집 윈도우와 감소된 표적-이탈 편집을 나타낸다. (도 24a-f).

[1117] 염기 편집기 시스템, 방법, 조성물 및 용도를 포함하는 아래에 번호매겨진 추가 실시형태가 본 명세서에서 구상된다:

[1118] 1. Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하는 융합 단백질로서, 상기 융합 단백질의 데아미나제는 표적 폴리뉴클레오티드 서열 중의 표적 핵염기를 탈아미노화시키며, 상기 N 말단 단편 또는 상기 C 말단 단편은 상기 표적 폴리뉴클레오티드 서열에 결합하고, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단은 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함하는, 융합 단백질.

[1119] 2. 실시형태 1에 있어서, 상기 표적 핵염기가 서열번호: 1의 N 말단 또는 C 말단에 융합된 상기 데아미나제를 포함하는 단부 말단(end terminus) 융합 단백질과 비교하여 더 낮은 표적-이탈 탈아미노화로 탈아미노화되는, 융합 단백질.

[1120] 3. 실시형태 1 또는 2에 있어서, 상기 표적 핵염기가 표적 폴리뉴클레오티드 서열 중의 프로토스페이서 인접 모티프(PAM) 서열로부터 1개 내지 20개 핵염기 떨어져 있는, 융합 단백질.

[1121] 4. 실시형태 3에 있어서, 상기 표적 핵염기가 PAM 서열의 상류에 있는 2개 내지 12개 핵염기인, 융합 단백질.

[1122] 5. 실시형태 1 내지 4 중 어느 한 실시형태에 있어서, 상기 융합 단백질이 상기 표적 핵염기를 탈아미노화시킬 때, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단이 상기 표적 핵염기에 근접한 아미노산을 포함하는, 융합 단백질 .

[1123] 6. 실시형태 1 내지 4 중 어느 한 실시형태에 있어서, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단이 상기 Cas9 폴리펩티드의 알파-나선 구조의 일부를 포함하는, 융합 단백질.

[1124] 7. 실시형태 1 내지 4 중 어느 한 실시형태에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 DNA 결합 도메인을 포함하는, 융합 단백질.

[1125] 8. 실시형태 1 내지 4 중 어느 한 실시형태에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 RuvC 도메인을 포함하는, 융합 단백질.

[1126] 9. 실시형태 1 내지 4 중 어느 한 실시형태에 있어서, 상기 N-말단 단편 및 상기 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는, 융합 단백질.

[1127] 10. 실시형태 1 내지 4 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드의 가요성 루프가 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치 또는 이들의 상응하는 위치에서의 아미노산을 포함하는, 융합 단백질.

[1128] 11. 실시형태 10에 있어서, 상기 N-말단 단편이 상기 Cas9 폴리펩티드의 N-말단에서 시작하고 서열번호: 1에서 번호매겨진 것과 같은 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 또는 1248-1297

사이의 위치 또는 이들의 상응하는 위치에서 종결되는 연속(contiguous) 서열인, 융합 단백질.

[1129] 12. 실시형태 10에 있어서, 상기 C-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 1301-1368, 1248-1297, 1078-1231, 1026-1051, 948-1001, 692-942, 580-685, 또는 538-568 사이의 위치에서 시작하고 상기 Cas9 폴리펩티드의 C-말단에서 종결되는 연속 서열인, 융합 단백질.

[1130] 13. 실시형태 10에 있어서, 상기 N-말단 단편의 상기 C-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1016, 1023, 1029, 1040, 1069, 또는 1247 또는 이의 상응하는 아미노산인, 융합 단백질.

[1131] 14. 실시형태 10에 있어서, 상기 C-말단 단편의 상기 N-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017, 1024, 1030, 1041, 1070, 1248 또는 이의 상응하는 아미노산인, 융합 단백질.

[1132] 15. 실시형태 11 내지 14 중 어느 한 실시형태에 있어서, 상기 데아미나제가 시티딘 데아미나제인, 융합 단백질.

[1133] 16. 실시형태 10에 있어서, 상기 N-말단 단편이 상기 Cas9 폴리펩티드의 N-말단에서 시작하고 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-10 -579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 또는 이들의 상응하는 위치에서 종결되는 연속 서열인, 융합 단백질.

[1134] 17. 실시형태 10에 있어서, 상기 C-말단은 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치에서 시작하고 상기 Cas9 폴리펩티드의 C-말단에서 종결되는 연속 서열인, 융합 단백질.

[1135] 18. 실시형태 10에 있어서, 상기 N-말단 단편의 C-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1022, 1029, 1040, 1068, 1069, 1247, 1054, 1026, 768, 791, 792, 1248, 1052, 또는 1246 또는 이의 상응하는 아미노산인, 융합 단백질.

[1136] 19. 실시형태 10에 있어서, 상기 C-말단 단편의 N-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1023, 1030, 1041, 1069, 1070, 1248, 1055, 1026, 769, 792, 793, 873, 907, 1249, 1053, 또는 1247, 또는 이의 상응하는 아미노산인, 융합 단백질.

[1137] 20. 실시형태 16 내지 19 중 어느 한 실시형태에 있어서, 상기 데아미나제가 아데노신 데아미나제인, 융합 단백질.

[1138] 21. 실시형태 1 내지 20 중 어느 한 실시형태에 있어서, 추가의 촉매 도메인을 추가로 포함하는, 융합 단백질.

[1139] 22. 실시형태 21에 있어서, 상기 추가의 촉매 도메인이 시티딘 데아미나제 또는 아데노신 데아미나제인, 융합 단백질.

[1140] 23. 실시형태 1 내지 22 중 어느 한 실시형태에 있어서, 상기 N-말단 단편과 상기 데아미나제 사이에 링커를 추가로 포함하는, 융합 단백질.

[1141] 24. 실시형태 1 내지 22 중 어느 한 실시형태에 있어서, 상기 C-말단 단편과 상기 데아미나제 사이에 링커를 추가로 포함하는, 융합 단백질.

[1142] 25. 실시형태 1 내지 22 중 어느 한 실시형태에 있어서, 핵 위치결정 신호를 추가로 포함하는, 융합 단백질.

[1143] 26. 실시형태 25에 있어서, 상기 핵 위치결정 신호가 이분 핵 위치결정 신호인, 융합 단백질.

[1144] 27. 실시형태 1 내지 26 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 스트렙토코커스 피오게네스(*Streptococcus pyogenes*) Cas9(SpCas9), 스타필로코커스 아우레우스(Staphylococcus aureus) Cas9(SaCas9), 스트렙토코커스 써모필러스(*Streptococcus thermophilus*) 1 Cas9(St1Cas9), 또는 이들의 변이체인, 융합 단백질.

[1145] 28. 실시형태 1 내지 27 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 변형된 Cas9이고, 변경된 PAM에 대한 특이성을 갖는, 융합 단백질.

[1146] 29. 실시형태 1 내지 28 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 닉카아제인, 융합 단백질.

[1147] 30. 실시형태 1 내지 28 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 뉴클레아제 비활성인, 융합 단백질.

[1148] 31. 실시형태 1 내지 30 중 어느 한 실시형태에 있어서, 가이드 핵산 서열과 복합체화하여 상기 표적 핵염기의 탈아미노화를 수행하는, 융합 단백질.

[1149] 32. 복수의 융합 단백질을 포함하는 최적화된 염기 편집을 위한 단백질 라이브러리로서, 상기 복수의 융합 단백질 각각은 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하고, 상기 복수의 융합 단백질의 각각의 N-말단 단편은 상기 나머지 복수의 융합 단백질의 N-말단 단편과 상이하거나, 상기 융합 단백질 각각의 C-말단 단편은 상기 나머지 복수의 융합 단백질의 C-말단 단편과 상이하며, 상기 융합 단백질 각각의 데아미나제는 표적 폴리뉴클레오티드 서열 중의 프로토스페이서 인접 모티프(PAM) 서열에 근접한 상기 표적 핵염기를 탈아미노화시키며, 상기 N 말단 단편 또는 상기 C 말단 단편은 상기 표적 폴리뉴클레오티드 서열에 결합하는, 단백질 라이브러리.

[1150] 33. 실시형태 32에 있어서, 상기 PAM 서열로부터 1개 내지 20개 핵염기 떨어져 있는 각 핵염기에 대해, 상기 복수의 융합 단백질 중 적어도 하나가 상기 핵염기를 탈아미노화시키는, 단백질 라이브러리.

[1151] 34. 실시형태 32에 있어서, 상기 복수의 융합 단백질 각각의 상기 Cas9 폴리펩티드의 N 말단 단편의 C-말단 또는 C 말단 단편의 N-말단이 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함하는, 단백질 라이브러리.

[1152] 35. 실시형태 32 내지 34 중 어느 한 실시형태에 있어서, 상기 복수의 융합 단백질 중 적어도 하나가 서열번호: 1의 N 말단 또는 C 말단에 융합된 상기 데아미나제를 포함하는 단부 말단 융합 단백질과 비교하여 더 낮은 표적-이탈 탈아미노화로 상기 표적 핵염기를 탈아미노화시키는, 단백질 라이브러리.

[1153] 36. 실시형태 32 내지 35 중 어느 한 실시형태에 있어서, 상기 복수의 융합 단백질 중 적어도 하나가 PAM 서열의 상류에 있는 표적 핵염기 2 내지 12개 핵염기를 탈아미노화시키는, 단백질 라이브러리.

[1154] 37. 실시형태 32 내지 36 중 어느 한 실시형태에 있어서, 상기 융합 단백질이 상기 표적 핵염기를 탈아미노화시킬 때, 상기 복수의 융합 단백질의 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단이 상기 표적 핵염기에 근접한 아미노산을 포함하는, 단백질 라이브러리.

[1155] 38. 실시형태 34 내지 36 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드의 가요성 루프가 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치 또는 이들의 상응하는 위치에서의 아미노산을 포함하는, 단백질 라이브러리.

[1156] 39. 실시형태 38에 있어서, 상기 복수의 융합 단백질이 상기 Cas9 폴리펩티드의 N-말단에서 시작하는 N-말단 단편을 포함하고 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치 또는 이들의 상응하는 위치에서 종결되는 연속 서열인 융합 단백질을 포함하는, 단백질 라이브러리.

[1157] 40. 실시형태 38에 있어서, 상기 복수의 융합 단백질이 상기 Cas9 폴리펩티드의 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치 또는 이들의 상응하는 위치에서 시작하는 C-말단 단편을 포함하고 상기 Cas9 폴리펩티드의 C-말단에서 종결되는 연속 서열인 융합 단백질을 포함하는, 단백질 라이브러리.

[1158] 41. 실시형태 38 내지 40 중 어느 한 실시형태에 있어서, 상기 N-말단 단편의 C-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1022, 1029, 1040, 1068, 1069, 1247, 1054, 1026, 768, 791, 792, 1248, 1052, 또는 1246 또는 이의 상응하는 아미노산인, 단백질 라이브러리.

[1159] 42. 실시형태 38 내지 40 중 어느 한 실시형태에 있어서, 상기 C-말단 단편의 N-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1023, 1030, 1041, 1069, 1070, 1248, 1055, 1026, 769, 792, 793, 873, 907, 1249, 1053, 또는 1247 또는 이의 상응하는 아미노산, 단백질 라이브러리.

[1160] 43. 실시형태 32 내지 42 중 어느 한 실시형태에 있어서, 상기 데아미나제가 아데노신 데아미나제인, 단백질 라이브러리.

[1161] 44. 실시형태 32 내지 42 중 어느 한 실시형태에 있어서, 상기 데아미나제가 시티딘 데아미나제인, 단백질 라이브러리.

[1162] 45. 실시형태 32 내지 44 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 스트렙토코커스 피오게네스(*Streptococcus pyogenes*) Cas9(SpCas9), 스타필로코커스 아우레우스(Staphylococcus aureus) Cas9(SaCas9), 스트렙토코커스 써모필루스(*Streptococcus thermophilus*) 1 Cas9(St1Cas9), 또는 이들의 변이체인, 단백질 라

이브러리.

[1163] 46. 실시형태 32 내지 45 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 변형된 Cas9이고 변경된 프로토스페이서-인접 모티프(PAM)에 대한 특이성을 갖는, 단백질 라이브러리.

[1164] 47. 실시형태 32 내지 46 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 닉카아제인, 단백질 라이브러리.

[1165] 48. 실시형태 32 내지 46 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 뉴클레아제 비활성인, 단백질 라이브러리.

[1166] 49. 실시형태 1 내지 31 중 어느 한 실시형태의 융합 단백질을 포함하는 세포.

[1167] 50. 실시형태 49에 있어서, 상기 세포가 포유류 세포 또는 인간 세포인, 세포.

[1168] 51. 표적 폴리뉴클레오티드 서열에서 표적 핵염기를 편집하는 방법으로서, 상기 방법은 상기 표적 폴리뉴클레오티드를 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하는 융합 단백질과 접촉시키는 단계를 포함하고, 상기 융합 단백질의 데아미나제는 상기 표적 폴리뉴클레오티드 서열에서 상기 표적 핵염기를 탈아미노화시키며, 상기 N 말단 단편 또는 상기 C 말단 단편은 상기 표적 폴리뉴클레오티드 서열에 결합하고, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단은 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함하는, 방법.

[1169] 52. 실시형태 51에 있어서, 상기 표적 폴리뉴클레오티드 서열을 가이드 핵산 서열과 접촉시켜 상기 표적 핵염기의 탈아미노화를 수행하는 단계를 추가로 포함하는, 방법.

[1170] 53. 실시형태 52에 있어서, 상기 가이드 핵산 서열이 상기 표적 폴리뉴클레오티드 서열의 프로토스페이서 서열에 상보적인 스페이서 서열을 포함하며, 그로 인해 R-루프를 형성하는, 방법.

[1171] 54. 실시형태 53에 있어서, 상기 표적 핵염기가 서열번호: 1의 N 말단 또는 C 말단에 융합된 데아미나제를 포함하는 단부 말단 방법과 비교하여 더 낮은 표적-이탈 탈아미노화로 탈아미노화되는, 방법.

[1172] 55. 실시형태 54에 있어서, 상기 융합 단백질의 데아미나제가 상기 R-루프의 범위 내에서 2개 이하의 핵염기를 탈아미노화시키는, 방법.

[1173] 56. 실시형태 51 내지 55 중 어느 한 실시형태에 있어서, 상기 표적 핵염기가 상기 표적 폴리뉴클레오티드 서열 내의 PAM 서열로부터 1개 내지 20개 핵염기 떨어져 있는, 방법.

[1174] 57. 실시형태 55에 있어서, 상기 표적 핵염기가 상기 PAM 서열의 상류에 있는 2개 내지 12개 핵염기인, 방법.

[1175] 58. 실시형태 51 내지 57 중 어느 한 실시형태에 있어서, 상기 융합 단백질의 데아미나제가 상기 표적 핵염기를 탈아미노화시킬 때, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단이 상기 표적 핵염기에 근접한 아미노산을 포함하는, 방법.

[1176] 59. 실시형태 51 내지 57 중 어느 한 실시형태에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 RuvC 도메인을 포함하는, 방법.

[1177] 60. 실시형태 51 내지 57 중 어느 한 실시형태에 있어서, 상기 N-말단 단편 및 상기 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는, 방법.

[1178] 61. 실시형태 51 내지 57 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드의 가요성 루프가 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치 또는 이들의 상응하는 위치에서의 아미노산을 포함하는, 방법

[1179] 62. 실시형태 61에 있어서, 상기 N-말단 단편이 상기 Cas9 폴리펩티드의 N-말단에서 시작하고 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치 또는 이들의 상응하는 위치에서 종결되는 연속 서열인, 방법.

[1180] 63. 실시형태 61에 있어서, 상기 C-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치 또는 이들의 상응하는 위치에서 시작하고 상기 Cas9 폴리펩티드의 C-말단에서 종결되는 연속 서열인, 방법.

[1181] 64. 실시형태 61에 있어서, 상기 N-말단 단편의 C-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미

노산 1016, 1023, 1029, 1040, 1069 또는 1247 또는 이의 상응하는 아미노산인, 방법.

[1182] 65. 실시형태 61에 있어서, 상기 C-말단 단편의 N-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017, 1024, 1030, 1041, 1070, 1248 또는 이의 상응하는 아미노산인, 방법.

[1183] 66. 실시형태 62 내지 65 중 어느 한 실시형태에 있어서, 상기 데아미나제가 시티딘 데아미나제인, 방법.

[1184] 67. 실시형태 61에 있어서, 상기 N-말단 단편이 상기 Cas9 폴리펩티드의 N-말단에서 시작하고 서열번호: 1에서 번호매겨진 것과 같은 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 또는 1248-1297 사이의 위치 또는 이들의 상응하는 위치에서 종결되는 연속 서열인, 방법.

[1185] 68. 실시형태 61에 있어서, 상기 C-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, 또는 1248-1248 사이의 위치 또는 이들의 상응하는 위치에서 시작하고 상기 Cas9 폴리펩티드의 C-말단에서 종결되는 연속 서열인, 방법.

[1186] 69. 실시형태 61에 있어서, 상기 N-말단 단편의 C-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1022, 1029, 1040, 1068, 1069, 1247, 1054, 1026, 768, 791, 792, 1248, 1052, 또는 1246 또는 이의 상응하는 아미노산인, 방법.

[1187] 70. 실시형태 61에 있어서, 상기 C-말단 단편의 N-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1023, 1030, 1041, 1069, 1070, 1248, 1055, 1026, 769, 792, 793, 873, 907, 1249, 1053, 또는 1247, 또는 이의 상응하는 아미노산인, 방법.

[1188] 71. 실시형태 67 내지 70 중 어느 한 실시형태에 있어서, 상기 데아미나제가 아데노신 데아미나제인, 방법.

[1189] 72. 실시형태 51 내지 71 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 변형된 Cas9이고 변경된 프로토스페이서-인접 모티프(PAM)에 대한 특이성을 갖는, 방법.

[1190] 73. 실시형태 51 내지 72 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 닉카아제인, 방법.

[1191] 74. 실시형태 51 내지 72 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 뉴클레아제 비활성인, 방법.

[1192] 75. 실시형태 51 내지 74 중 어느 한 실시형태에 있어서, 상기 접촉이 세포에서 수행되는, 방법.

[1193] 76. 실시형태 75에 있어서, 상기 세포가 포유류 세포 또는 인간 세포인, 방법.

[1194] 77. 실시형태 76에 있어서, 상기 세포가 만능 세포인, 방법.

[1195] 78. 실시형태 75 내지 77 중 어느 한 실시형태에 있어서, 상기 세포가 생체내(*in vivo*) 또는 생체외에(*ex vivo*) 있는 것인, 방법.

[1196] 79. 실시형태 51 내지 74 중 어느 한 실시형태에 있어서, 상기 접촉이 세포 집단에서 수행되는, 방법.

[1197] 80. 실시형태 79에 있어서, 상기 세포 집단이 포유류 세포 또는 인간 세포인, 방법.

[1198] 81. 대상체에서 유전적 상태를 치료하는 방법으로서, 상기 대상체에 Cas9 폴리펩티드의 N-말단 단편 및 C-말단 단편이 측접된 데아미나제를 포함하는 융합 단백질 또는 상기 융합 단백질을 코딩하는 폴리뉴클레오티드, 및 가이드 핵산 서열 또는 상기 가이드 핵산 서열을 코딩하는 폴리뉴클레오티드를 투여하는 단계를 포함하며, 상기 가이드 핵산 서열은 상기 융합 단백질이 상기 대상체에서 표적 폴리뉴클레오티드 서열 중의 표적 핵염기를 탈아미노화하도록 유도하고, 상기 N 말단 단편 또는 상기 C 말단 단편은 상기 표적 폴리뉴클레오티드 서열에 결합하며, 그로 인해 상기 유전적 상태를 치료하는, 방법.

[1199] 82. 실시형태 81에 있어서, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단이 상기 Cas9 폴리펩티드의 가요성 루프의 일부를 포함하는, 방법.

[1200] 83. 실시형태 81 또는 82에 있어서, 상기 대상체에게 가이드 핵산 서열을 투여하여 상기 표적 핵염기의 탈아미노화를 수행하는 단계를 추가로 포함하는, 방법.

[1201] 84. 실시형태 81 내지 83 중 어느 한 실시형태에 있어서, 상기 표적 핵염기가 상기 유전적 상태와 관련된 돌연변이를 포함하는, 방법.

[1202] 85. 실시형태 84에 있어서, 상기 표적 핵염기의 탈아미노화가 상기 표적 핵염기를 야생형 핵염기로 대체하는, 방법.

[1203] 86. 실시형태 84에 있어서, 상기 표적 핵염기의 탈아미노화가 상기 표적 핵염기를 비-야생형 핵염기로 대체하고, 상기 표적 핵염기의 탈아미노화가 상기 유전적 상태의 증상을 개선하는, 방법.

[1204] 87. 실시형태 81 내지 83 중 어느 한 실시형태에 있어서, 상기 표적 폴리뉴클레오티드 서열이 상기 표적 핵염기 이외의 핵염기에서의 상기 유전적 상태와 관련된 돌연변이를 포함하는, 방법.

[1205] 88. 실시형태 87에 있어서, 상기 표적 핵염기의 탈아미노화가 상기 유전적 상태의 증상을 개선하는, 방법.

[1206] 89. 실시형태 81 내지 88 중 어느 한 실시형태에 있어서, 상기 표적 핵염기가 상기 표적 폴리뉴클레오티드 서열 내의 PAM 서열로부터 1 내지 20개의 핵염기 떨어져 있는, 방법.

[1207] 90. 실시형태 89에 있어서, 상기 표적 핵염기가 상기 PAM 서열의 상류에 있는 2개 내지 12개 핵염기인, 방법.

[1208] 91. 실시형태 81 내지 90 중 어느 한 실시형태에 있어서, 상기 융합 단백질의 데아미나제가 상기 표적 핵염기를 탈아미노화시킬 때, 상기 N 말단 단편의 C-말단 또는 상기 C 말단 단편의 N-말단이 상기 표적 핵염기에 근접한 아미노산을 포함하는, 방법.

[1209] 92. 실시형태 81 내지 90 중 어느 한 실시형태에 있어서, 상기 N-말단 단편 또는 상기 C-말단 단편이 RuvC 도메인을 포함하는, 방법.

[1210] 93. 실시형태 81 내지 90 중 어느 한 실시형태에 있어서, 상기 N-말단 단편 및 상기 C-말단 단편 중 어느 것도 HNH 도메인을 포함하지 않는. 방법.

[1211] 94. 실시형태 81 내지 90 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드의 가요성 루프가 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치 또는 이들의 상응하는 위치에서의 아미노산을 포함하는, 방법.

[1212] 95. 실시형태 94에 있어서, 상기 N-말단 단편이 Cas9 폴리펩티드의 N-말단에서 시작하고 서열번호: 1에서 번호매겨진 것과 같은 V530-P537, F569-E579, D686-R691, Y943-D947, L1052-E1077, P1002-S1025, Y1232-G1247, 또는 R1298-K1300 위치에서 종결되는 연속 서열인, 방법.

[1213] 96. 실시형태 94에 있어서, 상기 C-말단 단편이 서열번호: 1에서 번호매겨진 것과 같은 530-537, 569-579, 686-691, 768-793, 943-947, 1002-1040, 1052-1077, 1232-1248, 또는 1298-1300 사이의 위치에서 시작하고 상기 Cas9 폴리펩티드의 C-말단에서 종결되는 연속 서열인, 방법.

[1214] 97. 실시형태 94에 있어서, 상기 N-말단 단편의 C-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1016, 1023, 1029, 1040, 1069, 1022, 1029, 1040, 1068, 1069, 1247, 1054, 1026, 768, 791, 792, 1246, 1247, 1248, 또는 1052, 또는 이의 상응하는 아미노산인, 방법.

[1215] 98. 실시형태 94에 있어서, 상기 C-말단 단편의 N-말단 아미노산이 서열번호: 1에서 번호매겨진 것과 같은 아미노산 1017, 1023, 1024, 1030, 1041, 1069, 1070, 1247, 1248, 1249, 1055, 1026, 769, 792, 793, 873, 907, 또는 1053, 또는 이의 상응하는 아미노산인, 방법.

[1216] 99. 실시형태 81 내지 98 중 어느 한 실시형태에 있어서, 상기 데아미나제가 시티딘 데아미나제인, 방법.

[1217] 100. 실시형태 81 내지 98 중 어느 한 실시형태에 있어서, 상기 데아미나제가 아데노신 데아미나제인, 방법.

[1218] 101. 실시형태 81 내지 100 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 변형된 Cas9이고, 변경된 PAM에 대한 특이성을 갖는, 방법.

[1219] 102. 실시형태 81 내지 101 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 닉카아제인, 방법.

[1220] 103. 실시형태 81 내지 101 중 어느 한 실시형태에 있어서, 상기 Cas9 폴리펩티드가 뉴클레아제 비활성인, 방법.

[1221] 104. 실시형태 81 내지 103 중 어느 한 실시형태에 있어서, 상기 대상체가 포유동물인, 방법.

[1222] 105. 실시형태 81 내지 104 중 어느 한 실시형태에 있어서, 상기 대상체가 인간인, 방법.

[1223] **기타 실시형태**

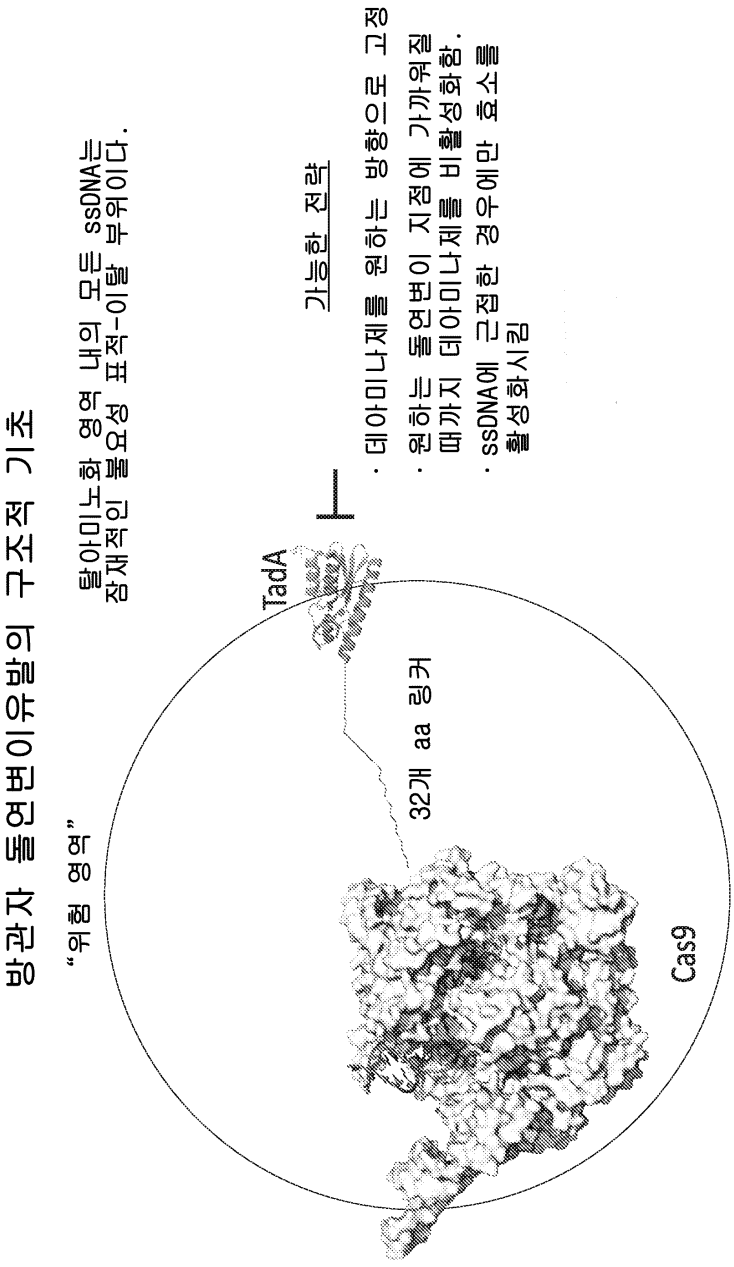
[1224] 전술한 설명으로부터, 다양한 사용 및 조건에 채택하기 위해 본 명세서에 설명된 발명에 대한 변경 및 수정이 이루어질 수 있음이 명백할 것이다. 이러한 실시 형태는 또한 다음의 청구범위 내에 있다.

[1225] 본 명세서에서 변수의 정의에서 요소의 목록을 언급하는 것은 열거된 요소의 임의의 단일 요소 또는 조합(또는 하위조합)으로서 해당 변수의 정의를 포함한다. 본 명세서의 실시형태에 대한 언급은 임의의 단일 실시형태로서 또는 임의의 다른 실시형태 또는 이의 일부와 조합된 실시형태를 포함한다.

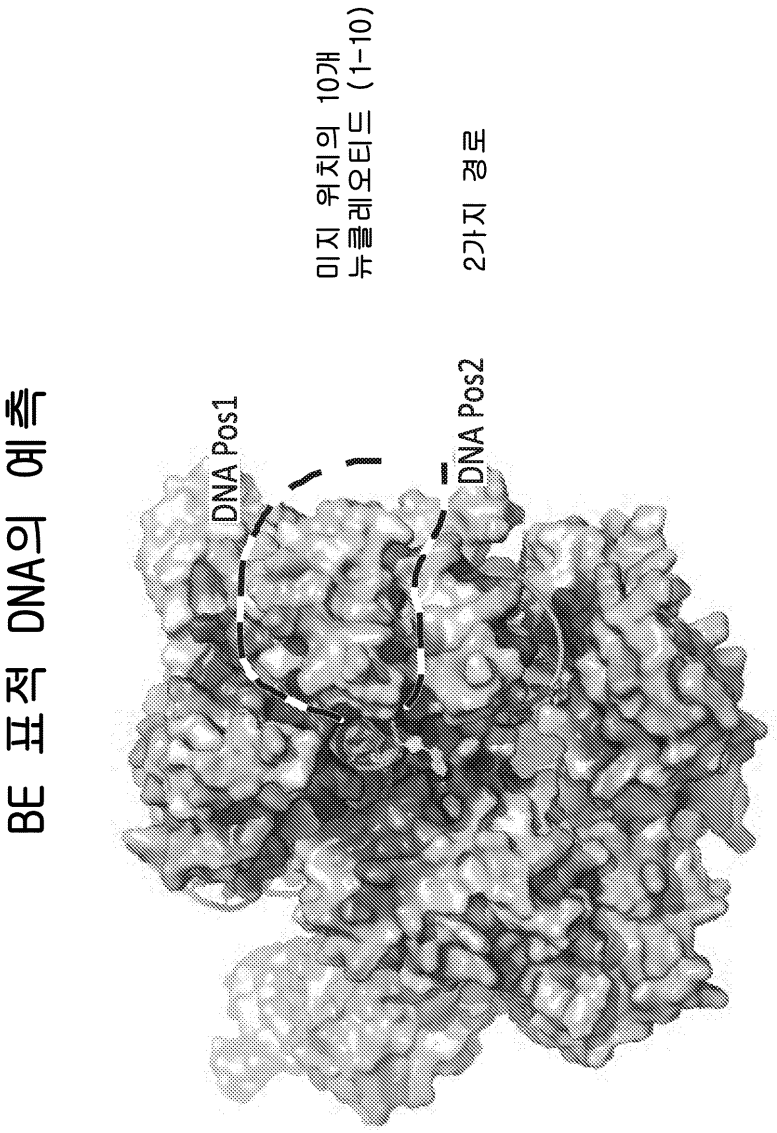
[1226] 본 명세서에서 언급된 모든 특허 및 간행물은 각각의 독립된 특허 및 간행물이 참조로 통합되도록 구체적으로 또 개별적으로 표시되는 것과 동일한 정도로 본 명세서에 참조로 통합된다.

도면

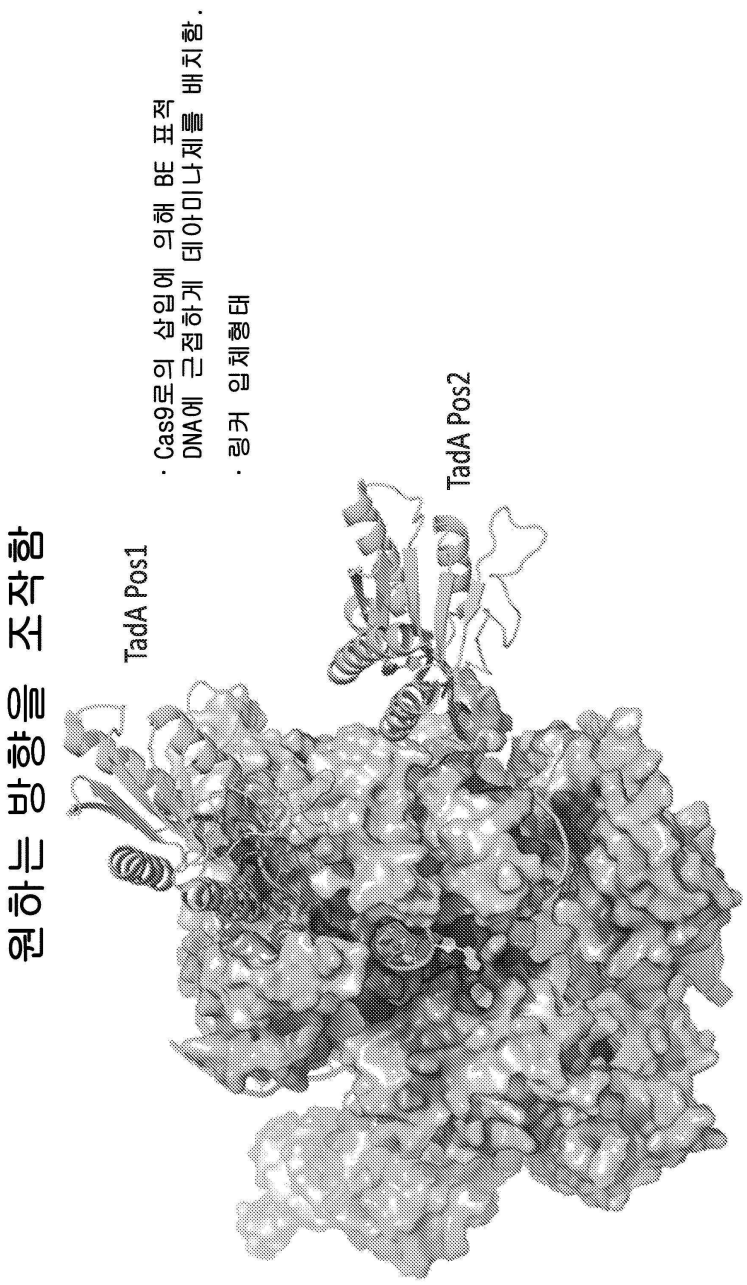
도면1



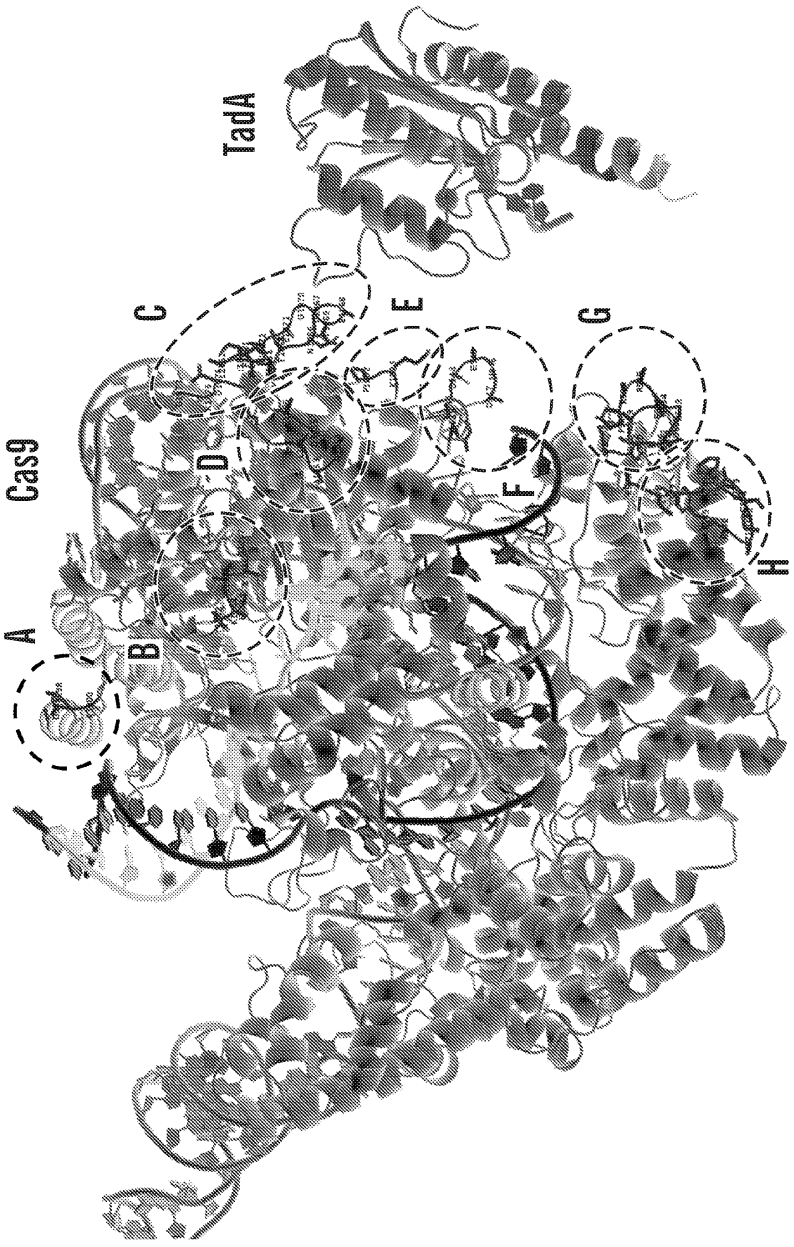
도면2



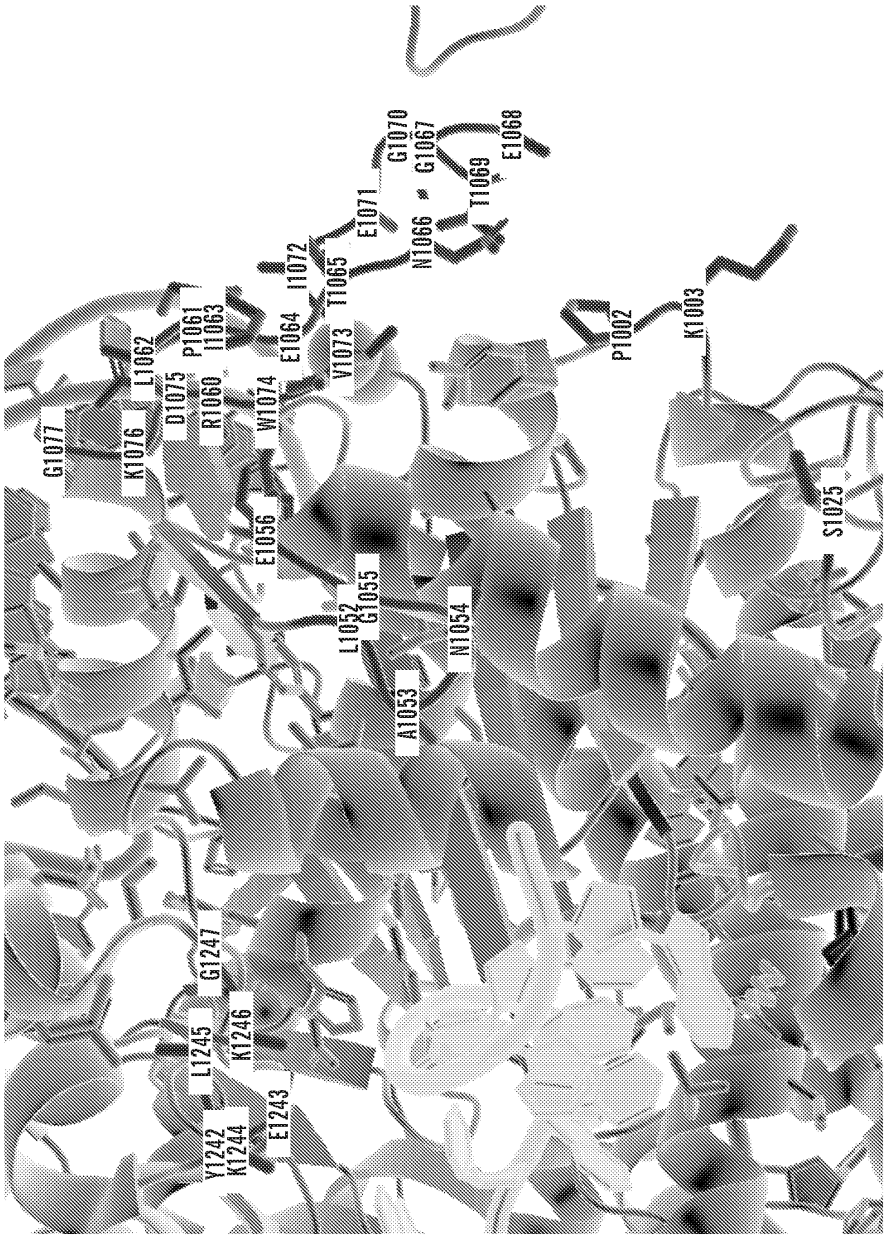
도면3


원하는 방향을 조작함
TadA Pos1
TadA Pos2
· Cas9로의 삽입에 의해 BE 표적
DNA에 근접하게 데아미나제를 배치함.
· 링커 입체형태

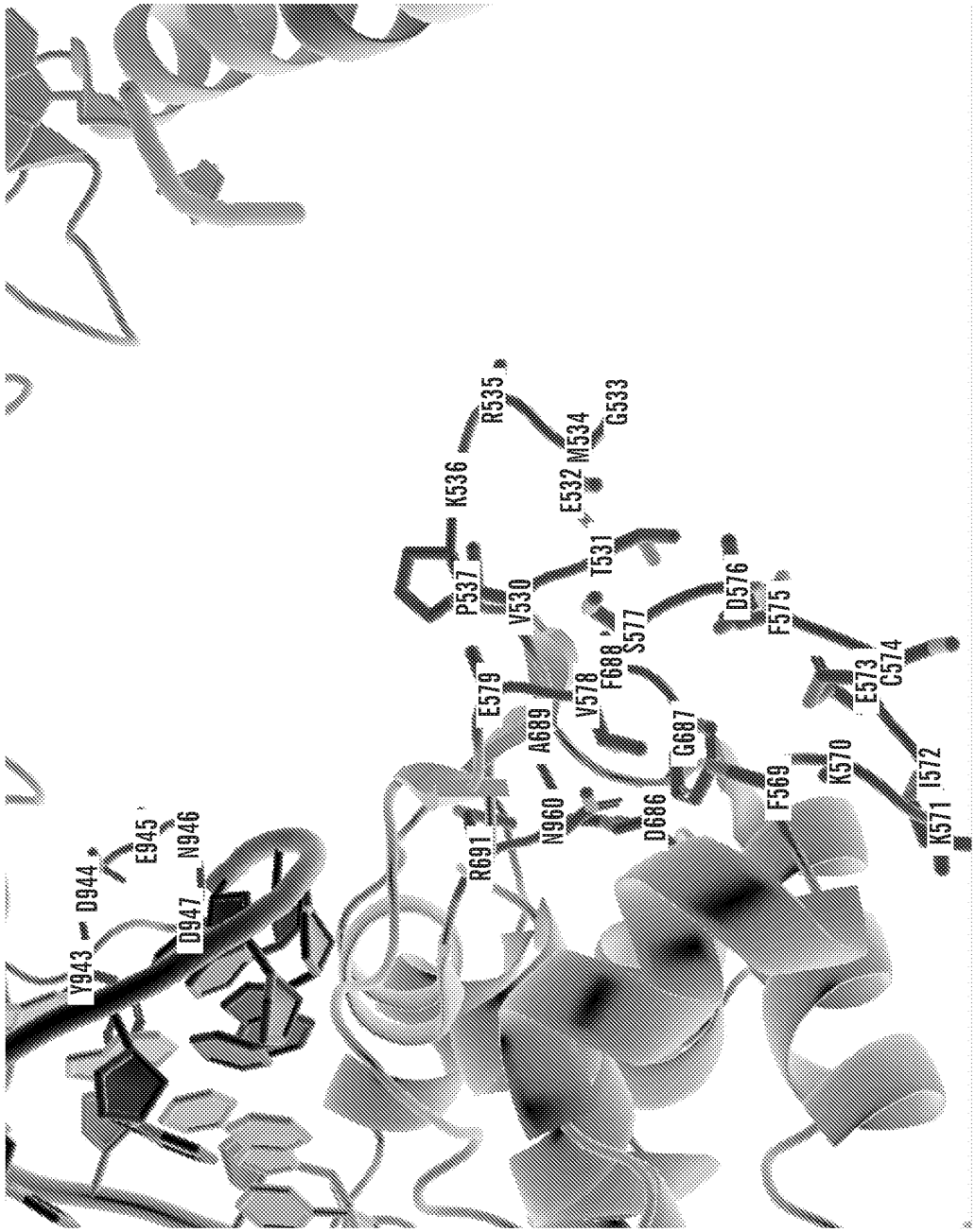
도면4



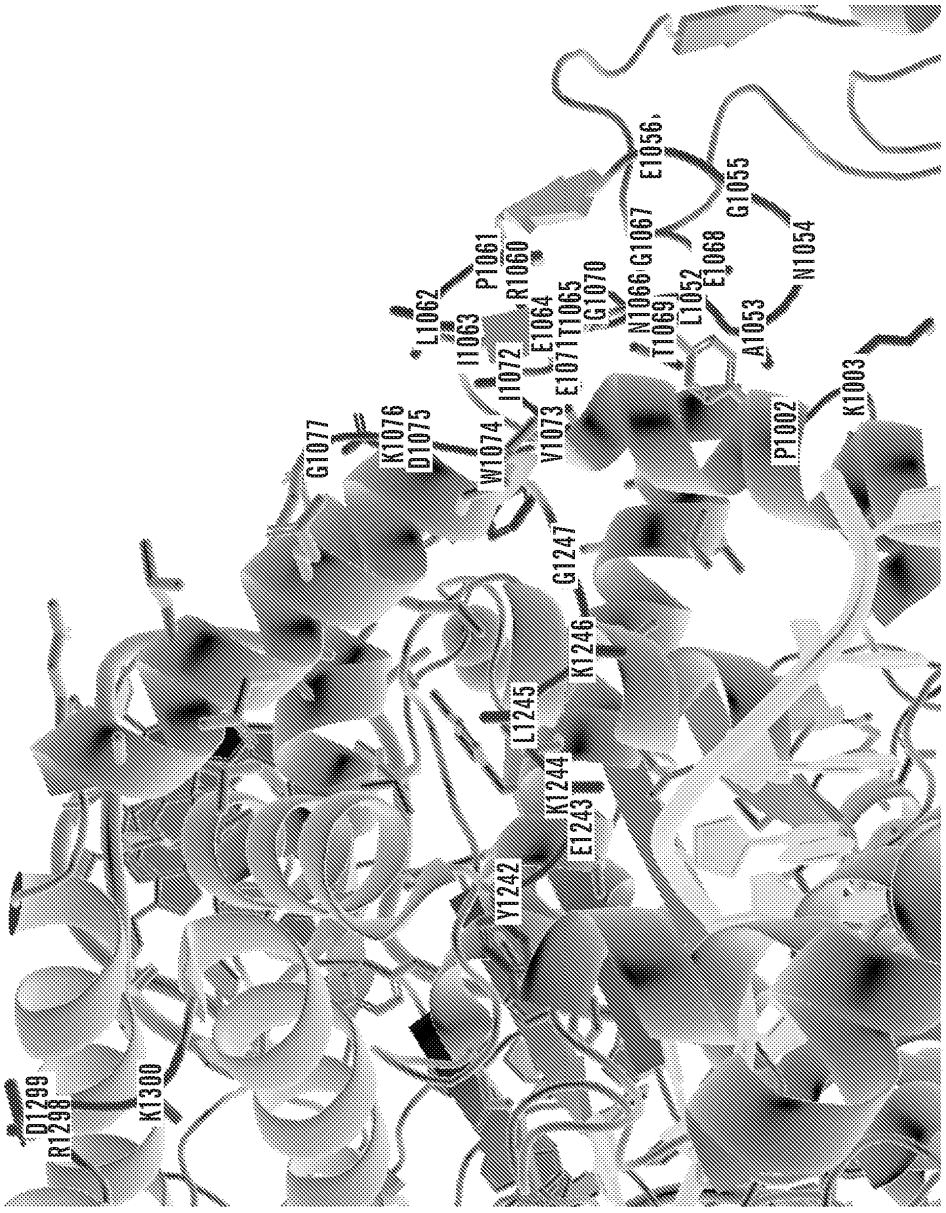
도면5


G1070
G1067
E1068
T1069
E1071
N1066
I1072
T1065
P1061
I1063
L1062
E1064
V1073
P1002
K1003
D1075
R1060
W1074
G1077
K1076
E1056
S1025
G1055
L1052
N1054
A1053
G1247
L1245
K1246
E1243
Y1242
K1244

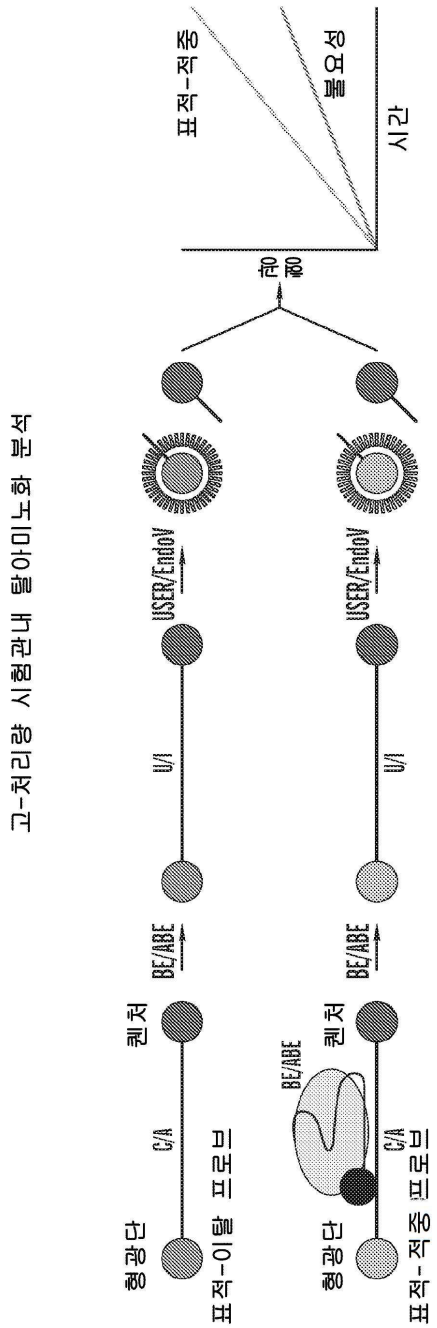
도면6


V943
D944
E945
N946
D947
R691
N960
A689
E579
V578
F688
S577
D686
G687
D576
F575
F569
K570
E573
C574
K571
I572
P537
V530
T531
K536
R535
E532
M534
G533

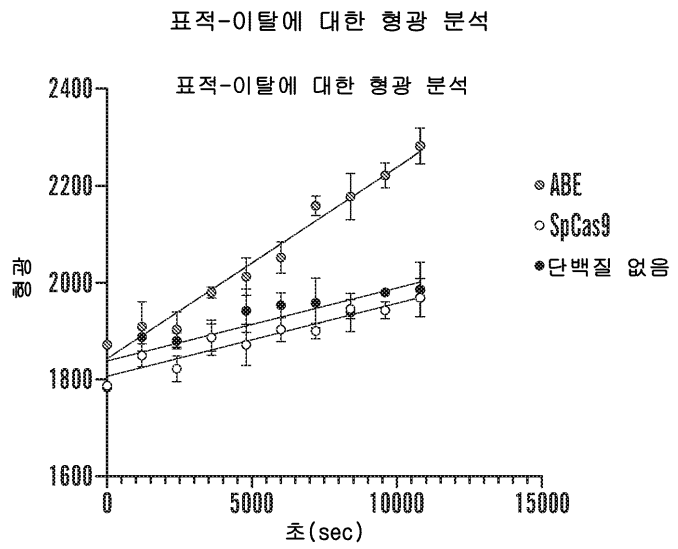
도면7


D1299
R1298
K1300
G1077
K1076
D1075
W1074
I1072
E1064
E1071T1065
V1073
L1062
I1063
P1061
R1060
G1070
N1066
G1067
E1056
T1069
L1052
E1068
G1055
N1054
A1053
P1002
K1003
G1247
K1246
L1245
K1244
E1243
Y1242
A, B, C, D, E

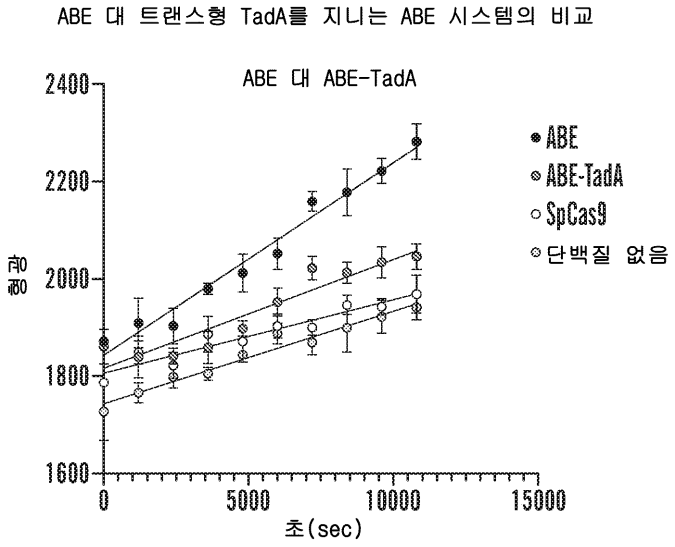
도면8


고-처리량 시험관내 탈아미노화 분석
표적-이탈 프로브
형광단
C/A
퀜처
BE/ABE
U/I
USER/EndoV
표적-적중 프로브
형광단
C/A
퀜처
BE/ABE
U/I
USER/EndoV
BE/ABE
결합
표적-적중
불요성
시간

도면9

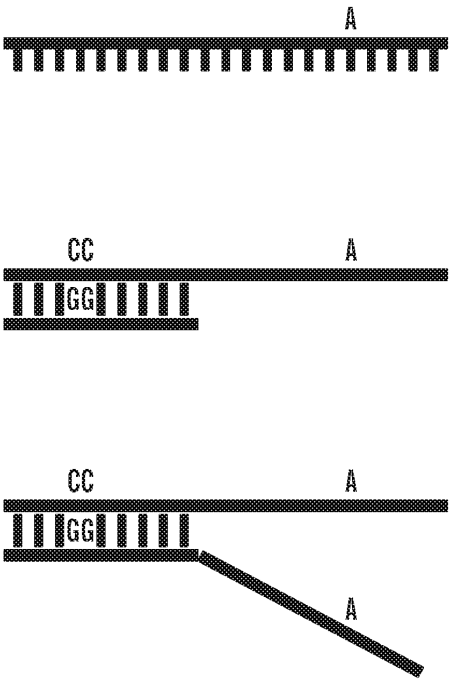

표적-이탈에 대한 형광 분석
표적-이탈에 대한 형광 분석
2400
2200
2000
1800
1600
0
5000
10000
15000
초(sec)
ABE
SpCas9
단백질 없음

도면10


ABE 대 트랜스형 TadA를 지니는 ABE 시스템의 비교
ABE 대 ABE-TadA
2400
2200
2000
1800
1600
0
5000
10000
15000
초(sec)
ABE
ABE-TadA
SpCas9
단백질 없음

도면11

불요성 표적-이탈 염기 편집에 대한 잠재적 기질을 테스트할 수 있음.

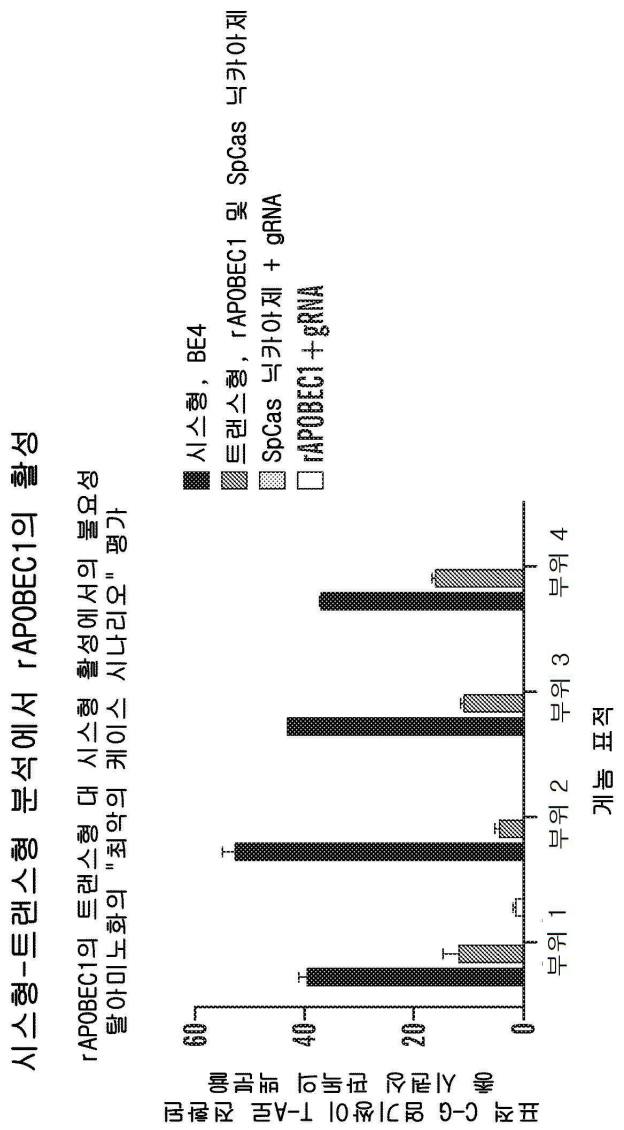


상기 분석에서 양호한 "불요성 탈아미노화" 기질에 대해 테스트함

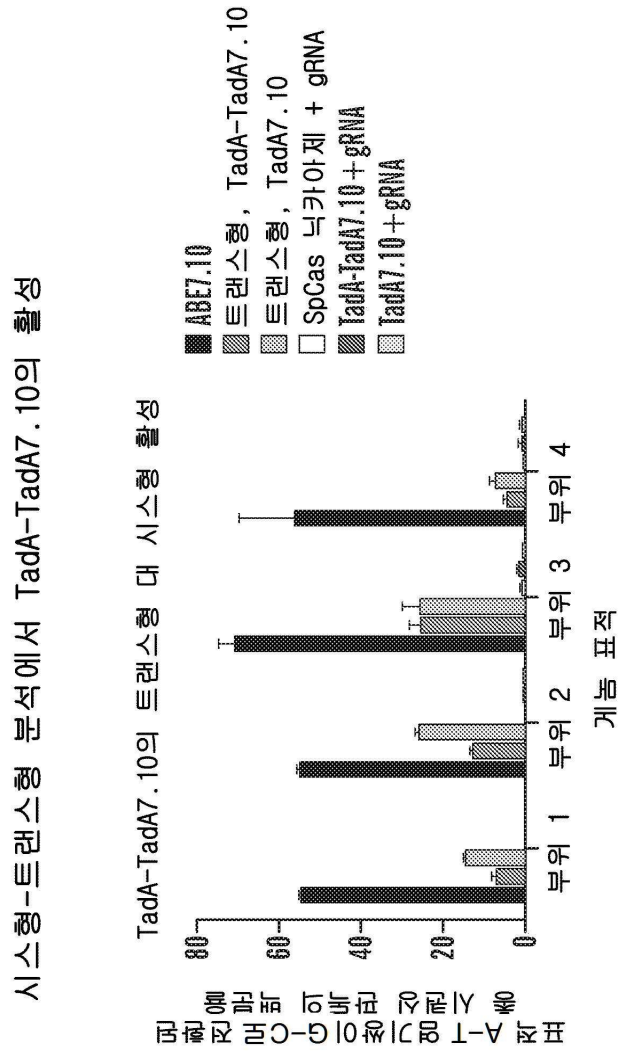
도면12

시스형-트랜스형 분석에서 데아미나제의 활성을 평가하기 위한 분석
실험 설계: 시스형(공유, 염기 편집기 컨텍스트) 대
트랜스형(불요성 탈아미노화 모방)에서 다양한 데아미나제를
이용한 염기 편집 평가
C에서 T로의 염기 편집
시스형
C에서 T로의 염기 편집
트랜스형
+
데아미나제
nCas9-UGI-UGI
데아미나제-nCas9-USI-UGI
ABE의 경우, UGI를 부착하지 않았음.

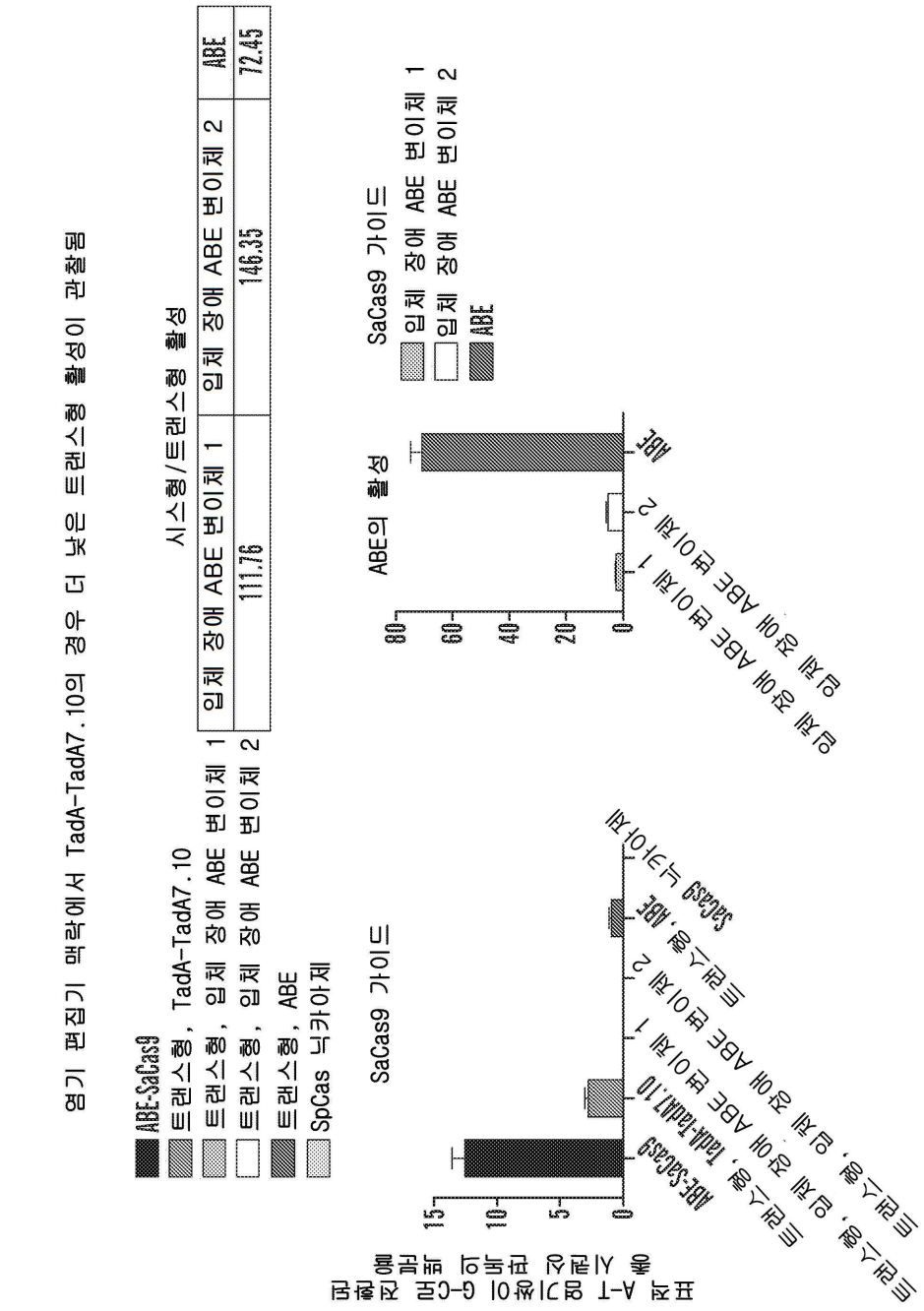
도면13


시스형-트랜스형 분석에서 rAPOBEC1의 활성
rAPOBEC1의 트랜스형 대 시스형 활성에서의 불요성
탈아미노화의 "최악의 케이스 시나리오" 평가
시스형, BE4
트랜스형, rAPOBEC1 및 SpCas 닉카아제
SpCas 닉카아제 + gRNA
rAPOBEC1+gRNA
표적 C-G 염기쌍의 T-A로 전환된
총 시퀀싱 판독의 백분율
60
40
20
0
부위 1
부위 2
부위 3
부위 4
게놈 표적

도면14


시스형-트랜스형 분석에서 TadA-TadA7.10의 활성
TadA-TadA7.10의 트랜스형 대 시스형 활성
ABE7.10
트랜스형, TadA-TadA7.10
트랜스형, TadA7.10
SpCas 닉카아제 + gRNA
TadA-TadA7.10+gRNA
TadA7.10+gRNA
표적 A-T 염기쌍의 G-C로 전환된
총 시퀀싱 판독의 백분율
0
20
40
60
80
부위 1
부위 2
부위 3
부위 4
게놈 표적

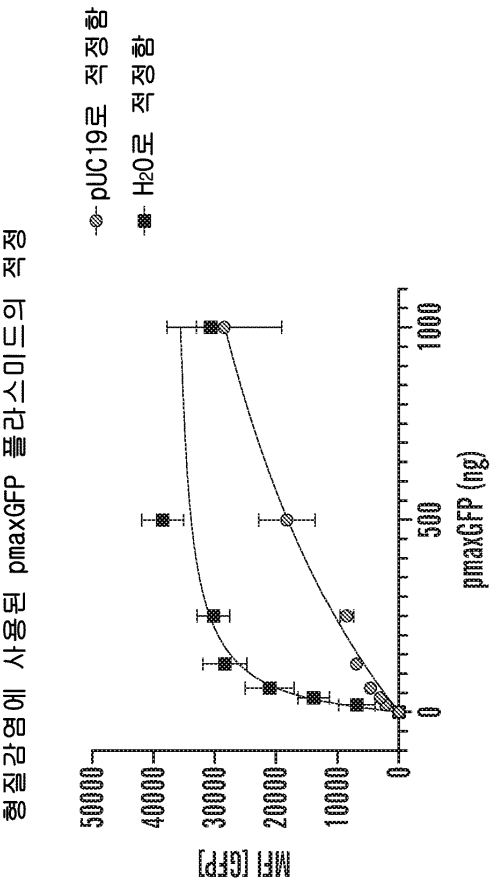
도면15


염기 편집기 맥락에서 TadA-TadA7.10의 경우 더 낮은 트랜스형 활성이 관찰됨
ABE-SaCas9
트랜스형, TadA-TadA7.10
트랜스형, 입체 장애 ABE 변이체 1
트랜스형, 입체 장애 ABE 변이체 2
트랜스형, ABE
SpCas 닉카아제
시스형/트랜스형 활성
입체 장애 ABE 변이체 1
입체 장애 ABE 변이체 2
ABE
111.76
146.35
72.45
SaCas9 가이드
ABE의 활성
입체 장애 ABE 변이체 1
입체 장애 ABE 변이체 2
ABE
표적 A-T 염기쌍의 G-C로 전환된 총 시퀀싱 판독의 백분율
SaCas9 가이드
ABE-SaCas9
TadA-TadA7.10
트랜스형, 입체 장애 ABE 변이체 1
트랜스형, 입체 장애 ABE 변이체 2
트랜스형, ABE
SpCas9 닉카아제

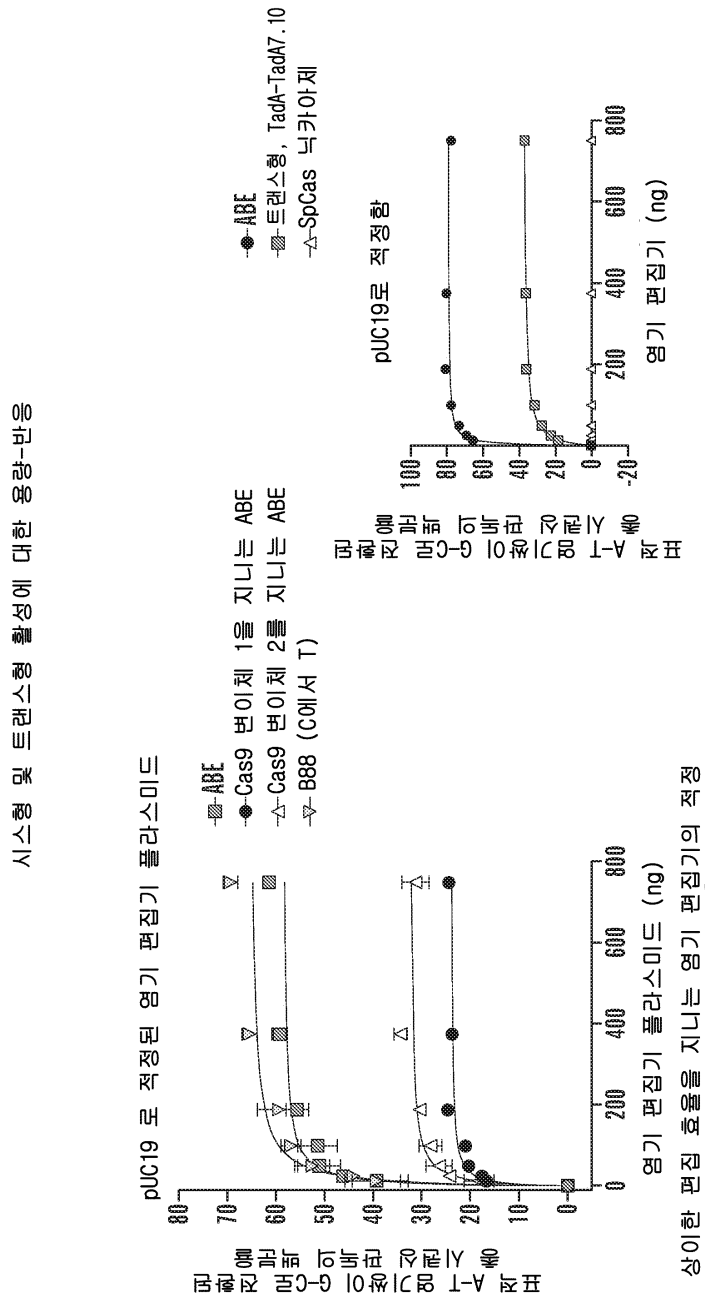
도면16

시스형 및 트랜스형 활성에 대한 용량-반응

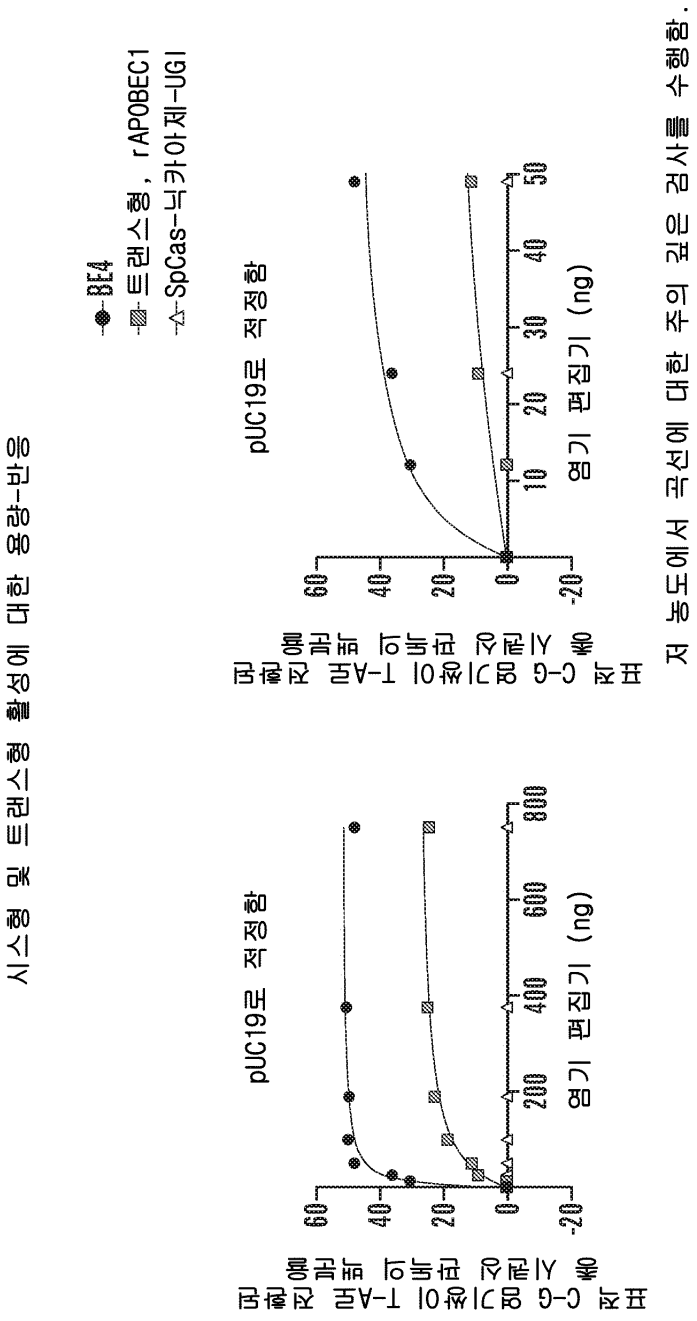
빈 벡터로의 pmaxGFP 플라스미드의 적정은 GFP의 발현 수준 감소를 초래하였음.



도면17

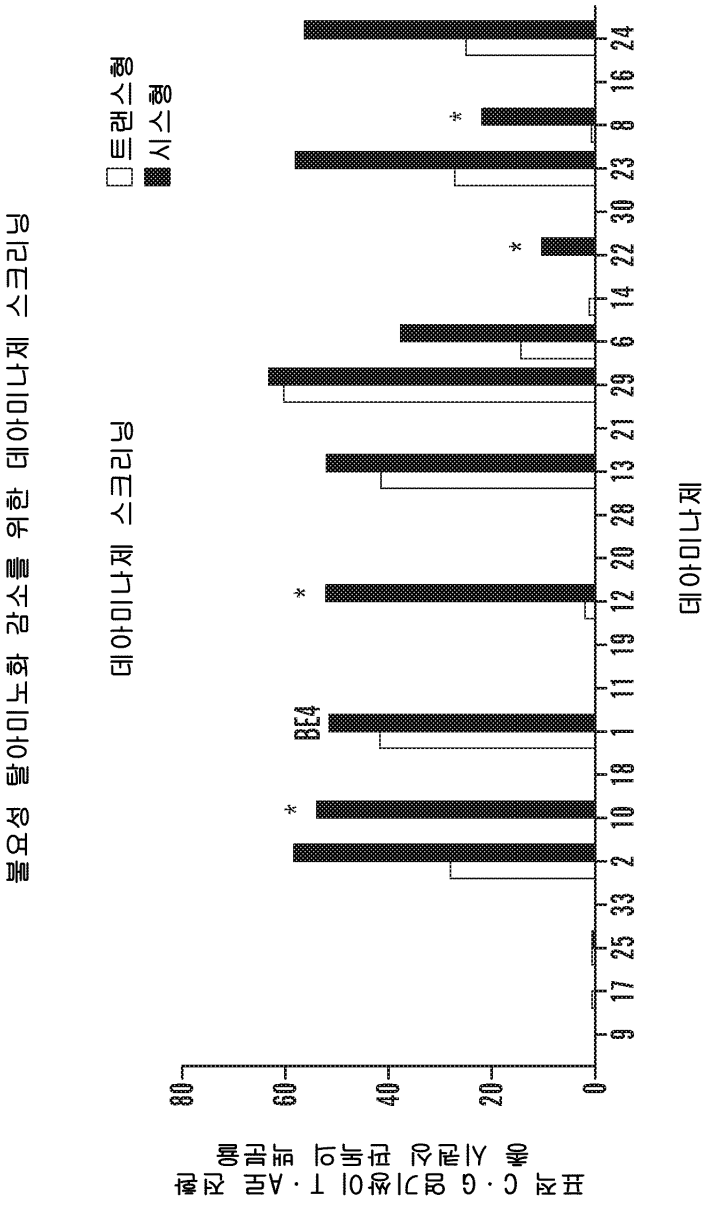

시스템 및 트랜스형 활성에 대한 용량-반응
pUC19 로 적정된 염기 편집기 플라스미드
ABE
Cas9 변이체 1을 지니는 ABE
Cas9 변이체 2를 지니는 ABE
B88 (C에서 T)
표적 A-T 염기쌍의 G-C로 전환된 백분율
염기 편집기 플라스미드 (ng)
ABE
트랜스형, TadA-TadA7.10
SpCas 닉카아제
pUC19로 적정함
표적 A-T 염기쌍의 G-C로 전환된 백분율
염기 편집기 (ng)
상이한 편집 효율을 지니는 염기 편집기의 적정

도면18

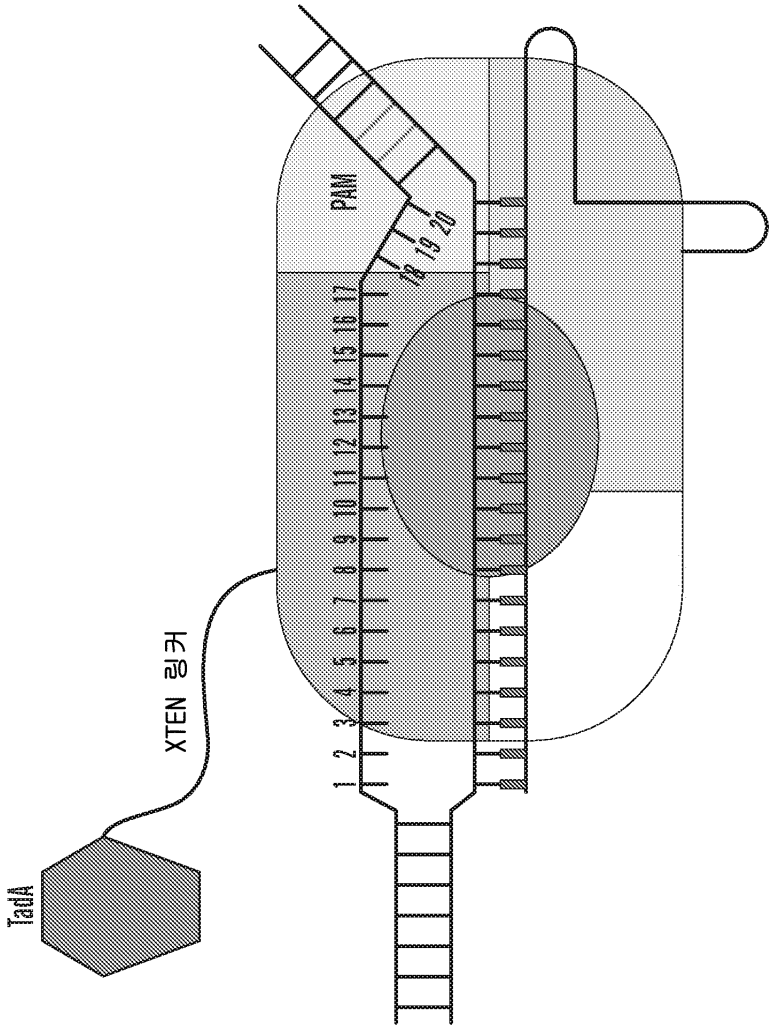


저 농도에서 곡선에 대한 주의 깊은 검사를 수행함.

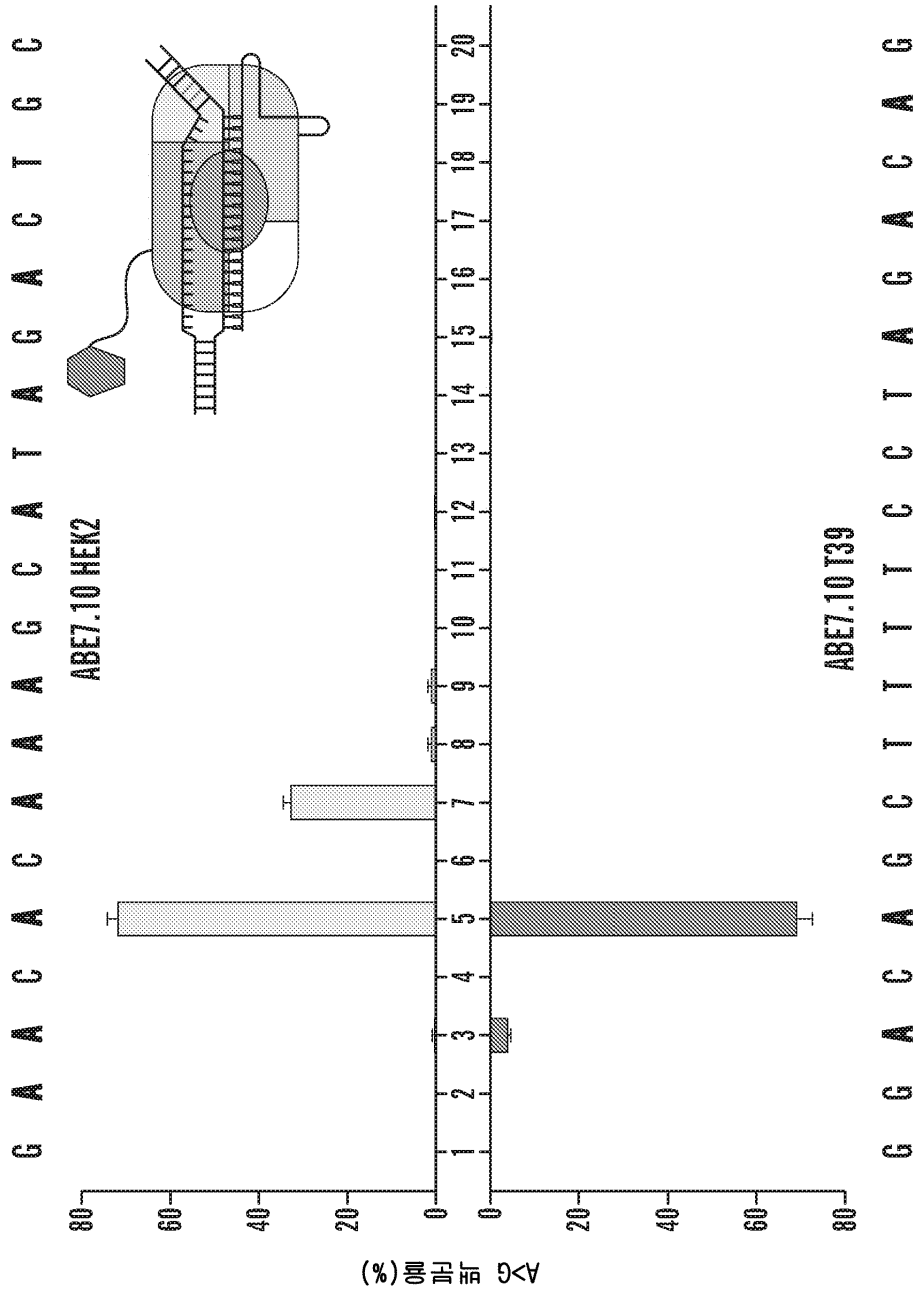
도면19


불요성 탈아미노화 감소를 위한 데아미나제 스크리닝
□트랜스형
■시스형
데아미나제 스크리닝
BE4
*
표적 C·G 염기쌍의 T·A로 전환
총 시퀀싱 판독의 백분율
0
20
40
60
80
9 17 25 33 2 10 18 1 11 19 12 20 28 13 21 29 6 14 22 30 23 8 16 24
데아미나제

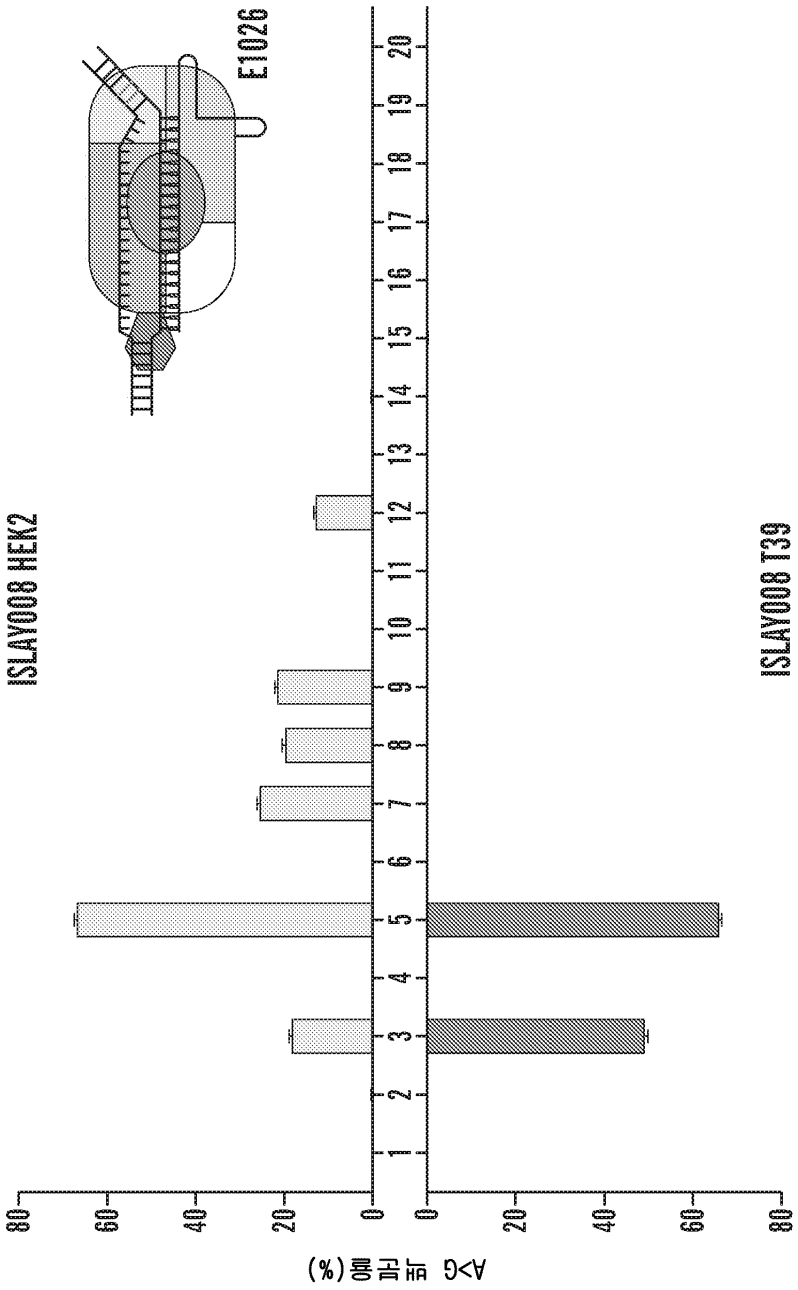
도면20a



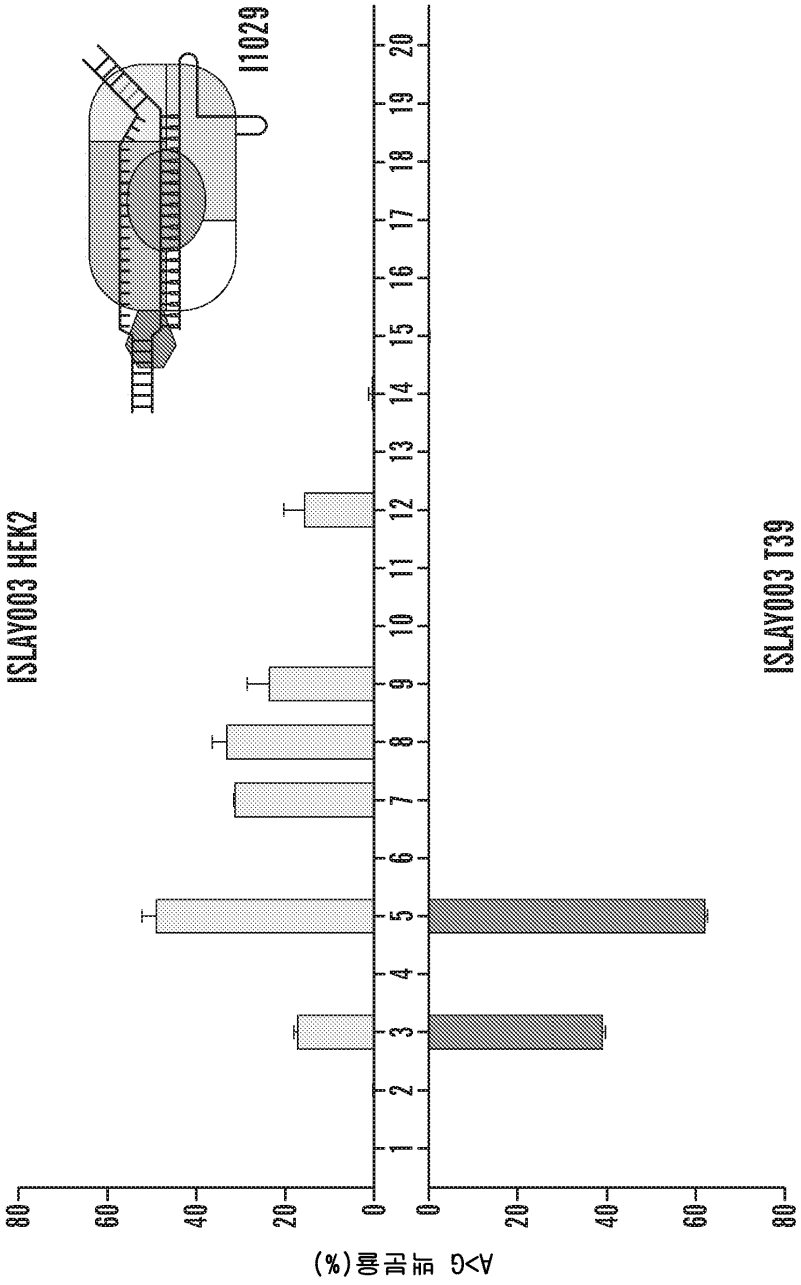
도면20b



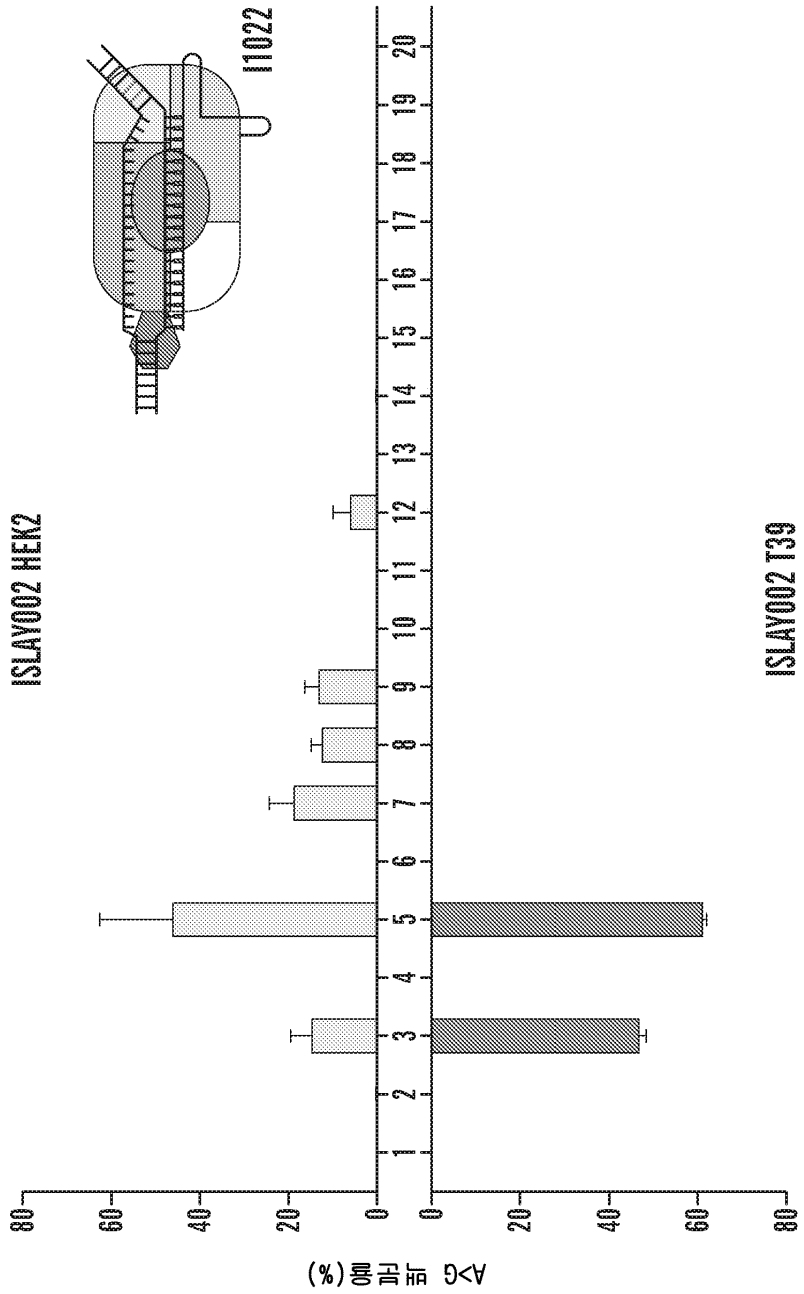
도면20c


E1026
ISLAY008 HEK2
ISLAY008 T39
A>G 변환율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

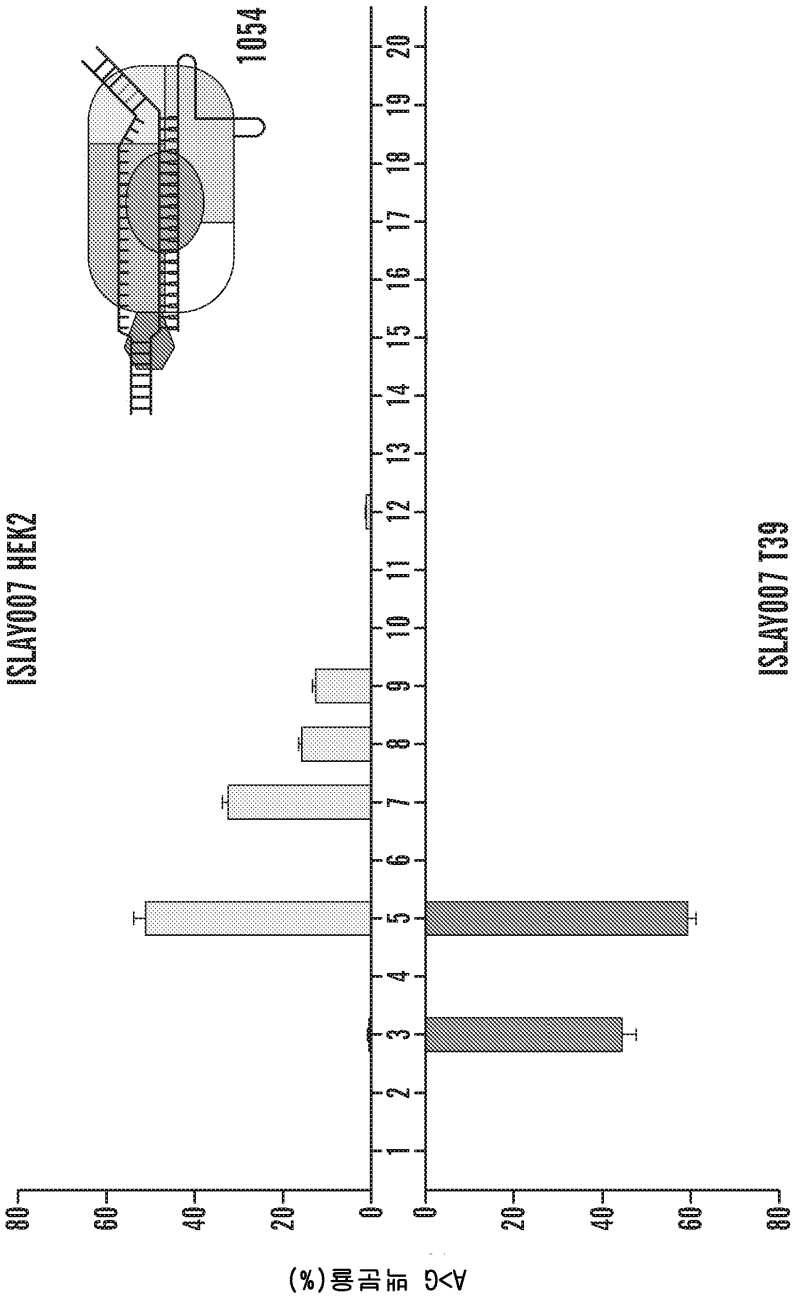
도면20d


I1029
ISLAY003 HEK2
ISLAY003 T39
A>G 변환율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

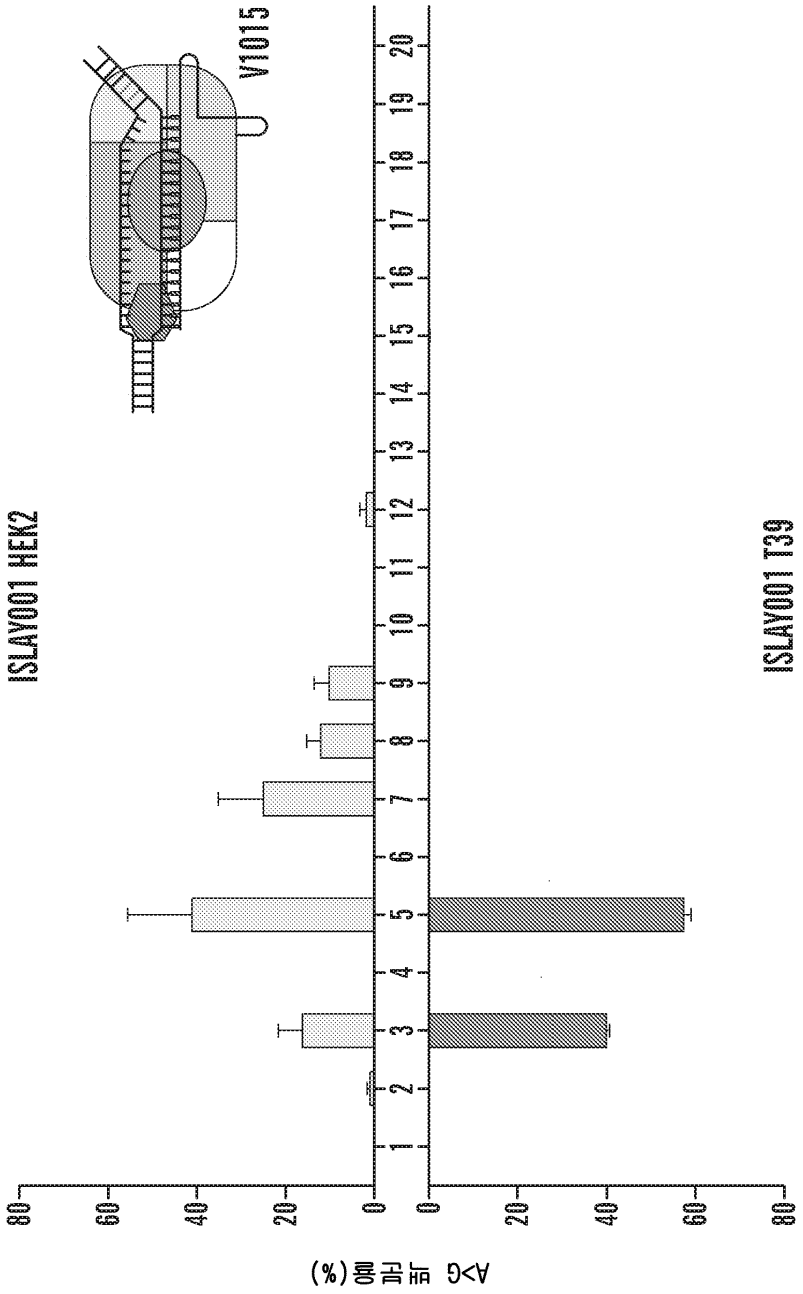
도면20e


I1022
ISLAY002 HEK2
ISLAY002 T39
A>G 편집율(%)
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

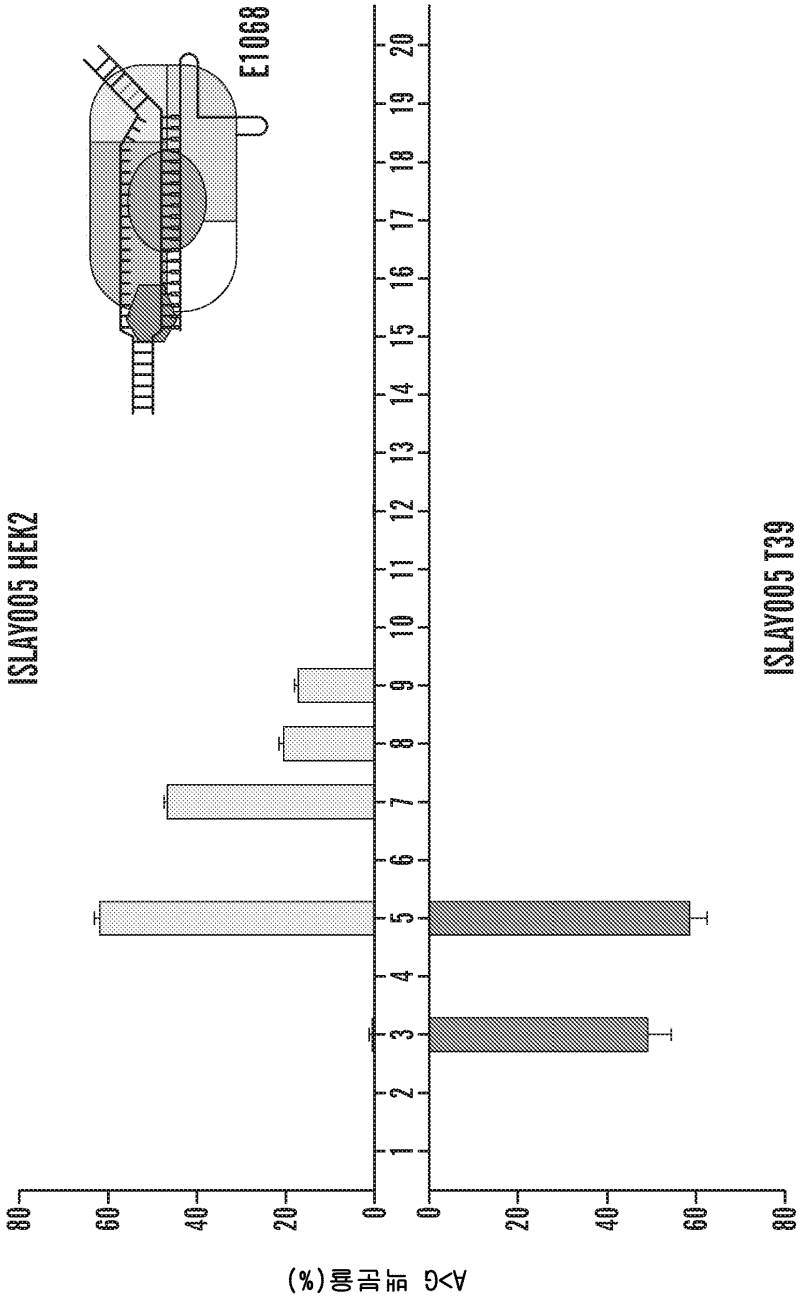
도면20f


1054
ISLAY007 HEK2
ISLAY007 T39
A>G 편집율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

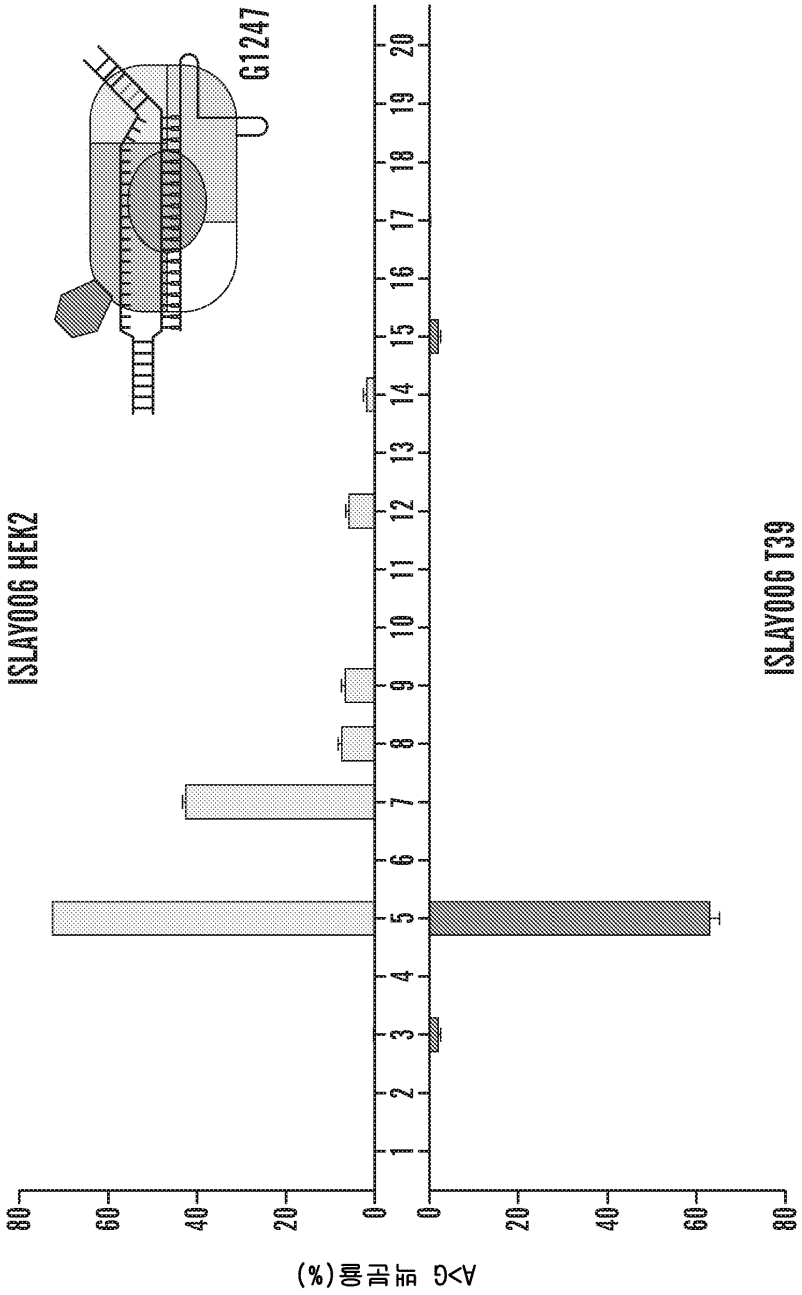
도면20g


V1015
ISLAY001 HEK2
ISLAY001 T39
A>G 변환율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

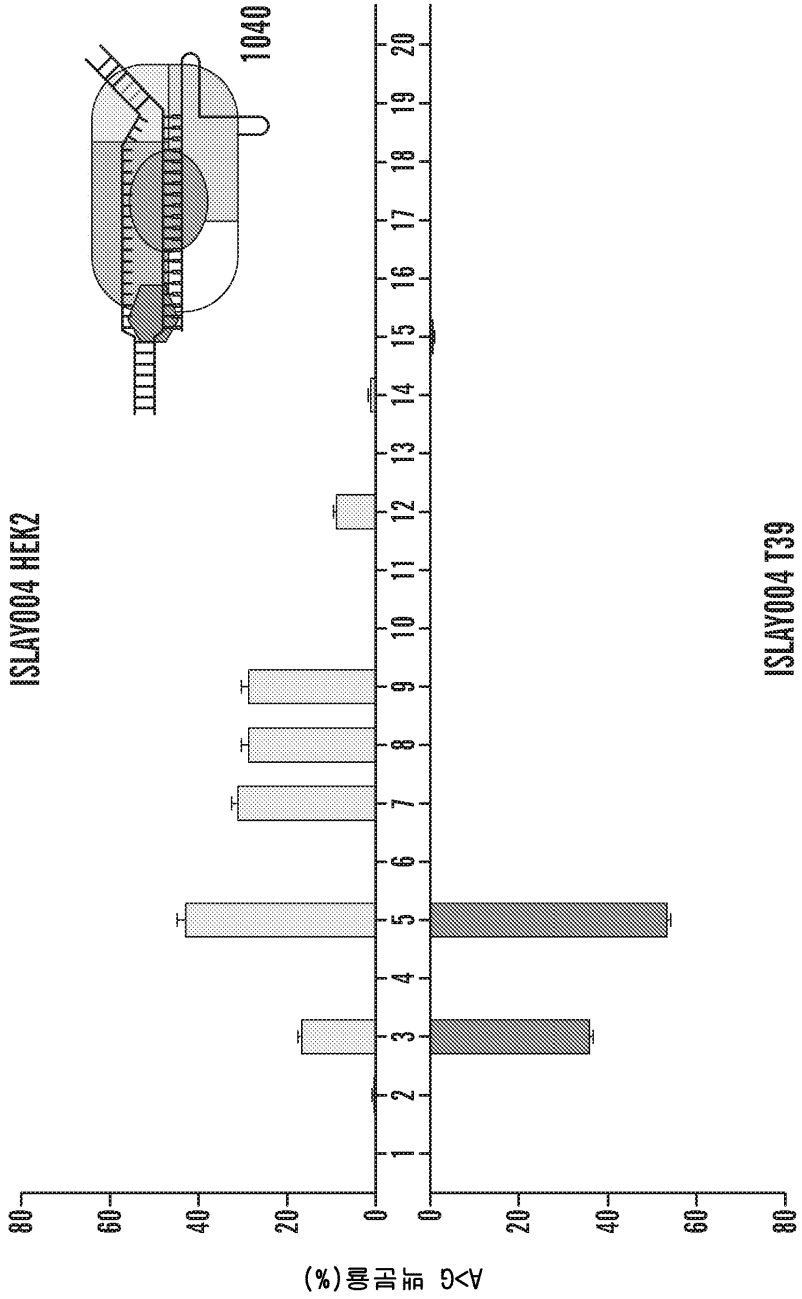
도면20h


E1068
ISLAY005 HEK2
ISLAY005 T39
A>G 편집율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

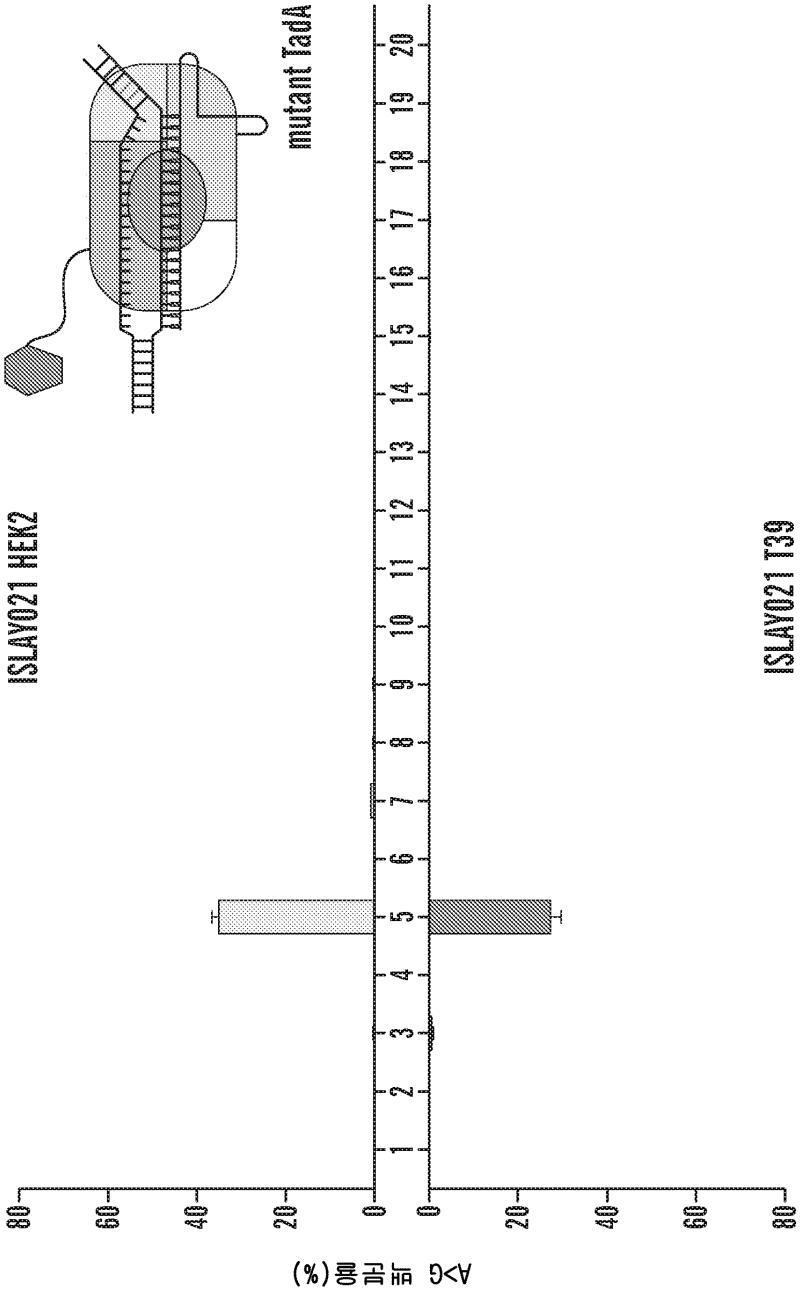
도면20i


G1247
ISLAY006 HEK2
ISLAY006 T39
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
80
60
40
20
0
0
20
40
60
80
A>G 빈도율(%)

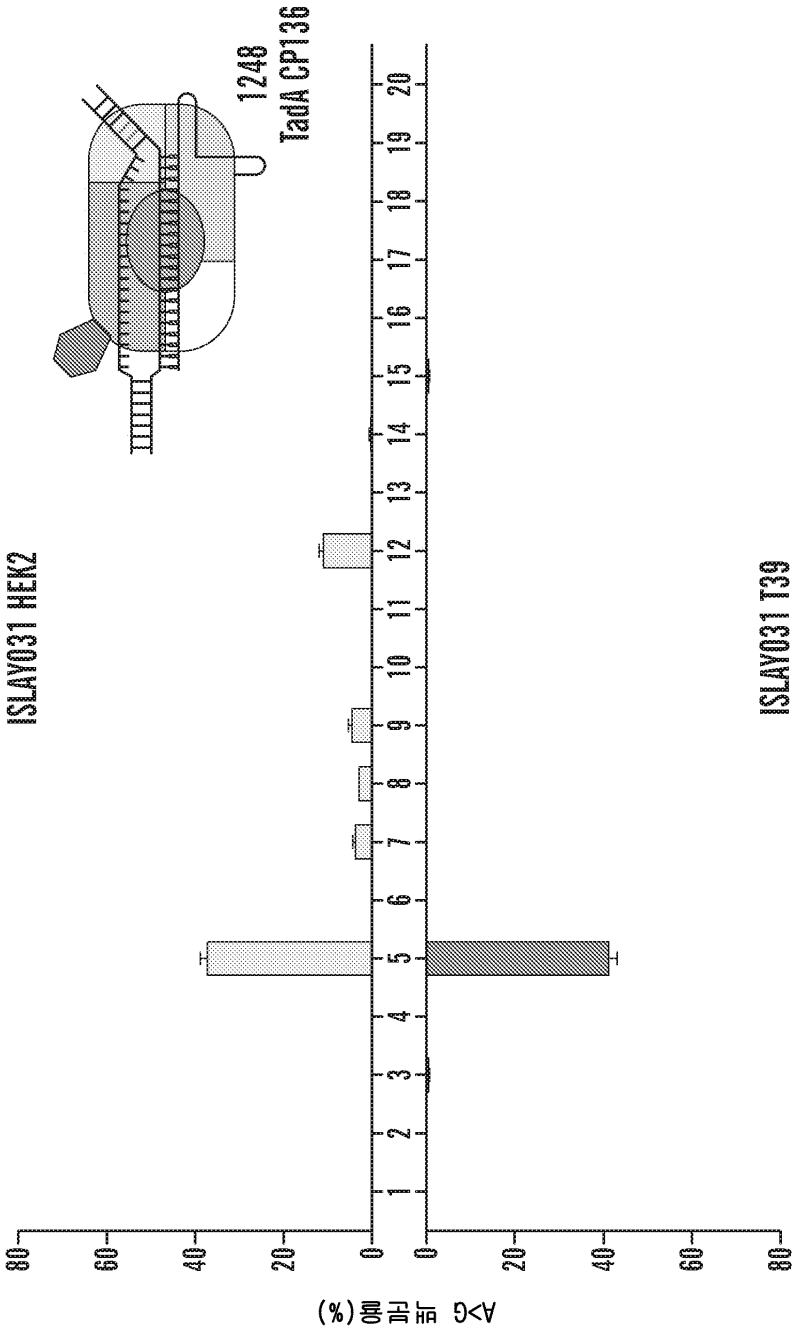
도면20j


1040
ISLAY004 HEK2
ISLAY004 T39
A>G 변환율(%)
80
60
40
20
0
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

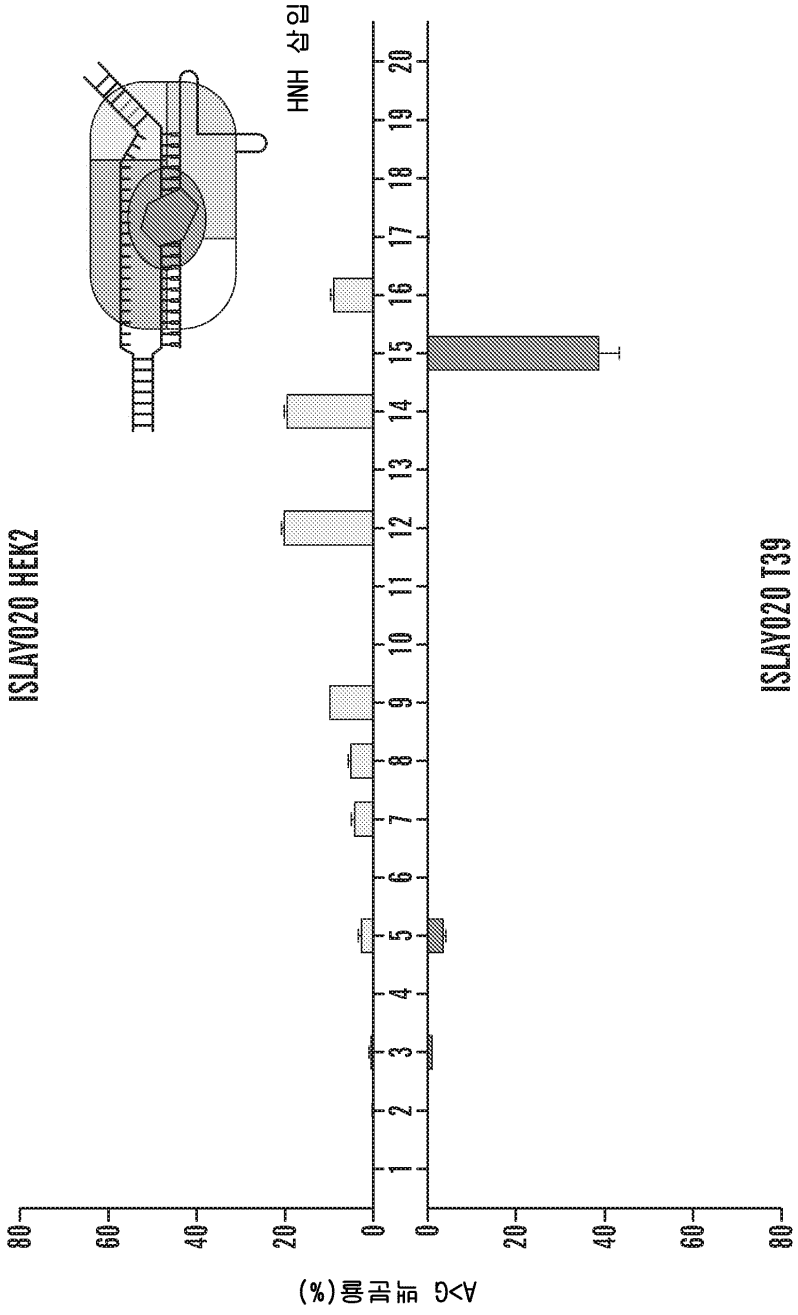
도면20k


mutant TadA
ISLAY021 HEK2
ISLAY021 T39
A>G 빈도율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

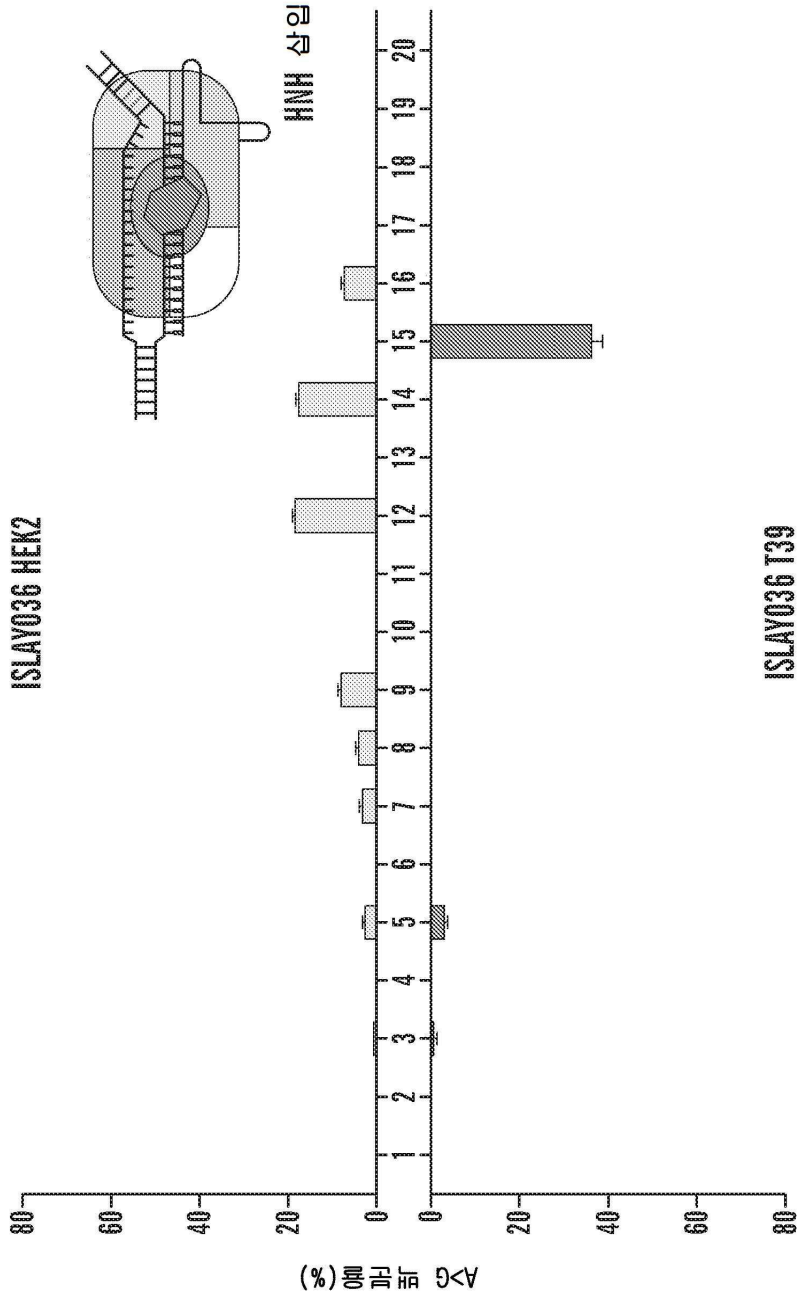
도면201


1248
TadA CP136
ISLAY031 HEK2
ISLAY031 T39
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
0 20 40 60 80
0 20 40 60 80
A>G 변환율(%)

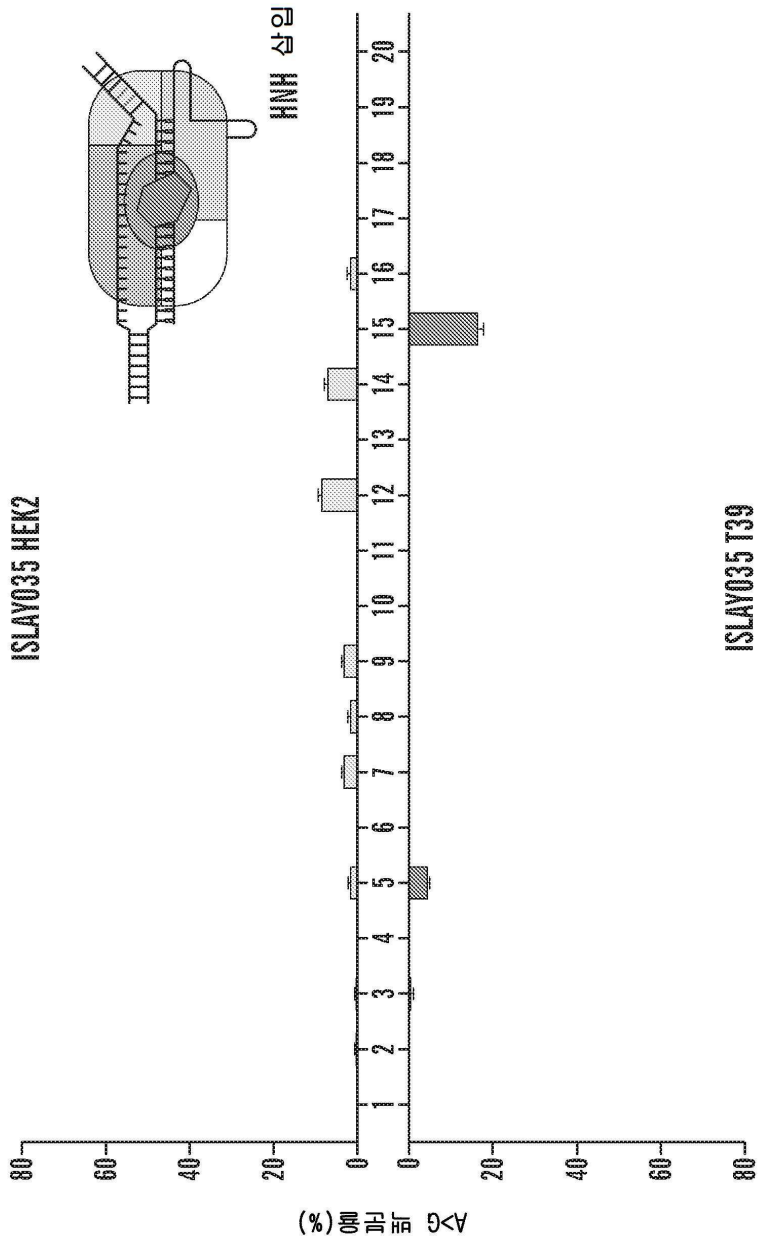
도면20m


HNH 삽입
ISLAY020 HEK2
ISLAY020 T39
A>G 변환율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

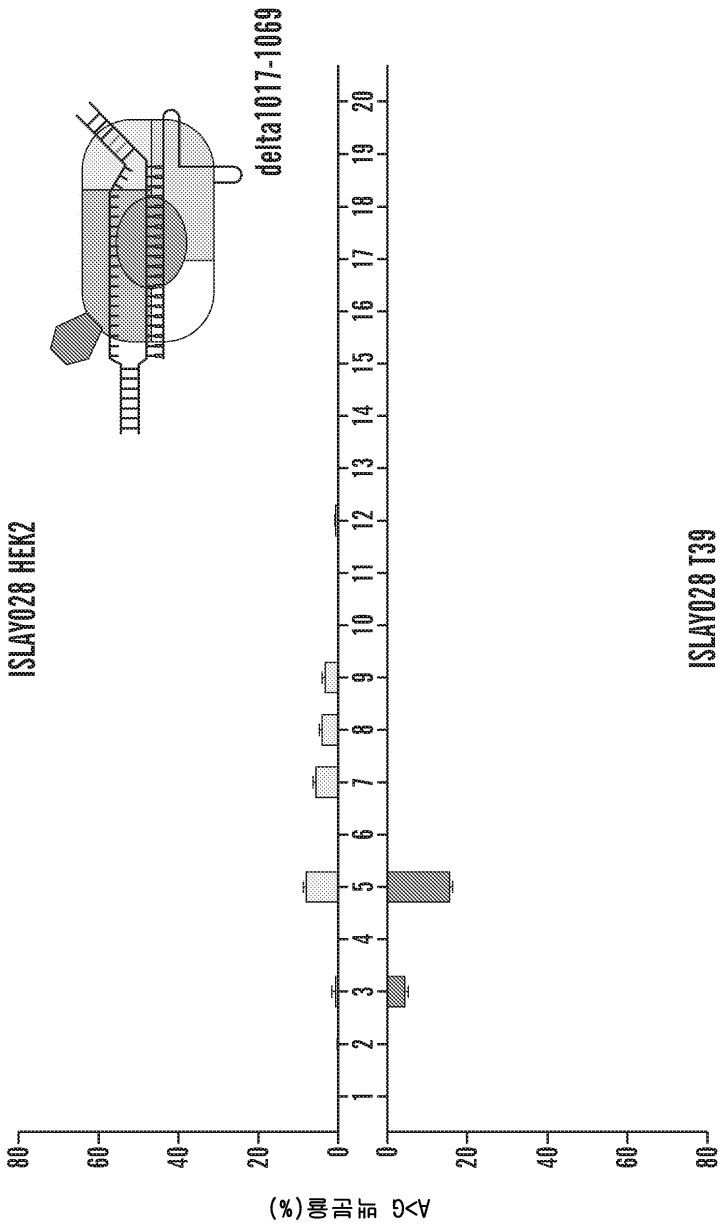
도면20n


ISLAY036 HEK2
ISLAY036 T39
HMW 삽입
A>G 변환율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

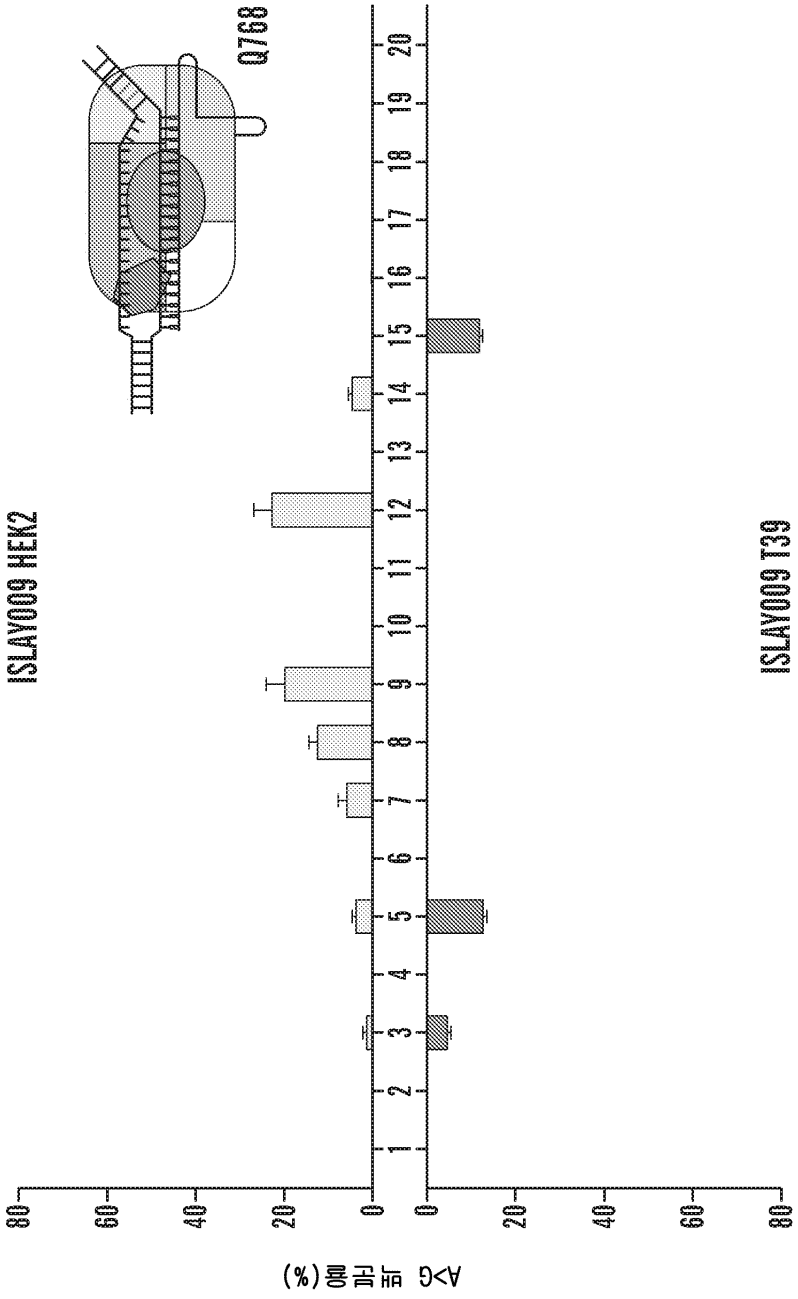
도면20o


HNH 삽입
ISLAY035 HEK2
ISLAY035 T39
A>G 빈도율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

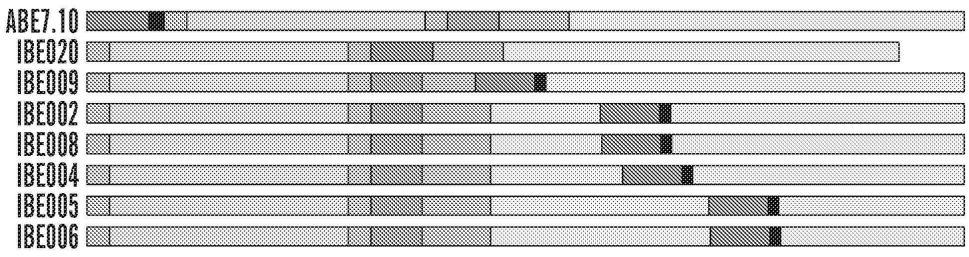
도면20p


delta1017-1069
ISLAY028 HEK2
ISLAY028 T39
A>G 변환율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

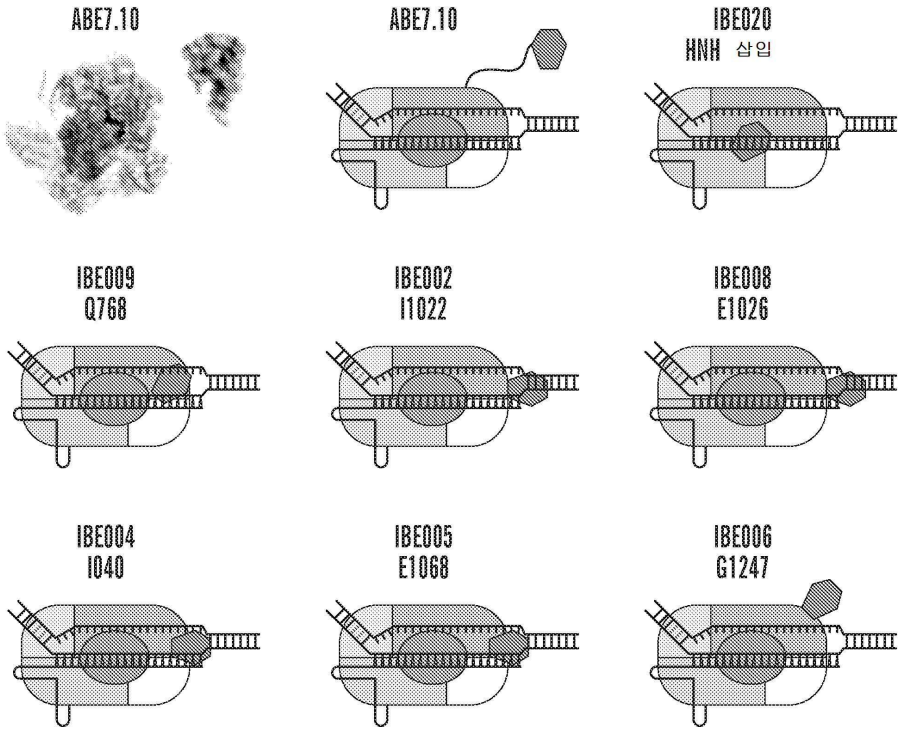
도면20q


Q768
ISLAY009 HEK2
ISLAY009 T39
A>G 편집율(%)
0
20
40
60
80
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

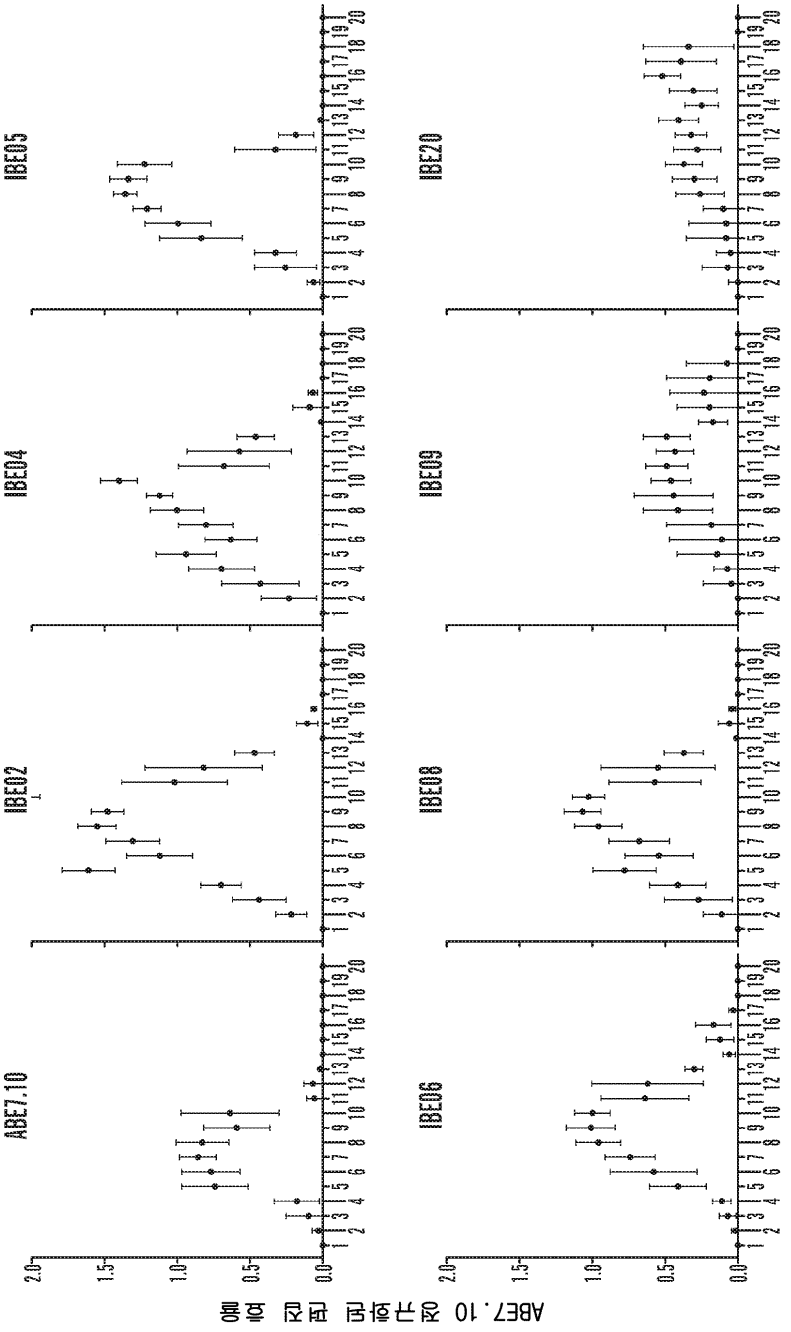
도면21a


ABE7.10
IBE020
IBE009
IBE002
IBE008
IBE004
IBE005
IBE006

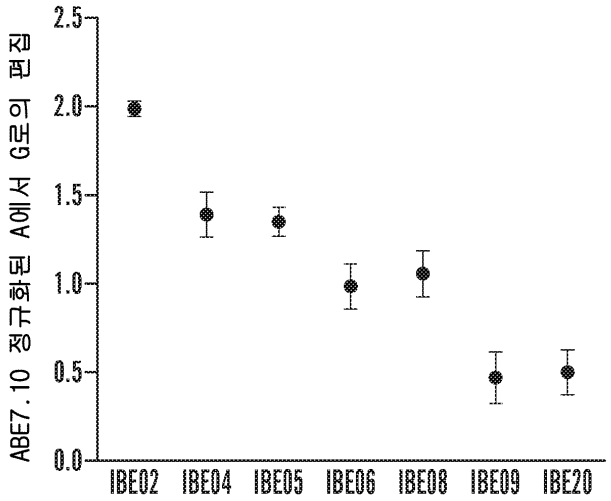
도면21b


ABE7.10
ABE7.10
IBE020
HNH 삽입
IBE009
Q768
IBE002
I1022
IBE008
E1026
IBE004
I040
IBE005
E1068
IBE006
G1247

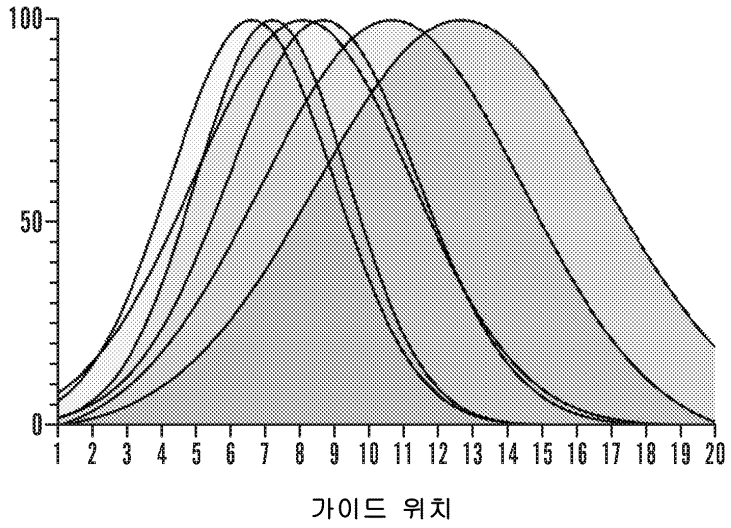
도면22a


ABE7.10
IBE02
IBE04
IBE05
IBE06
IBE08
IBE09
IBE20
ABE7.10 정규화된 편집 효율

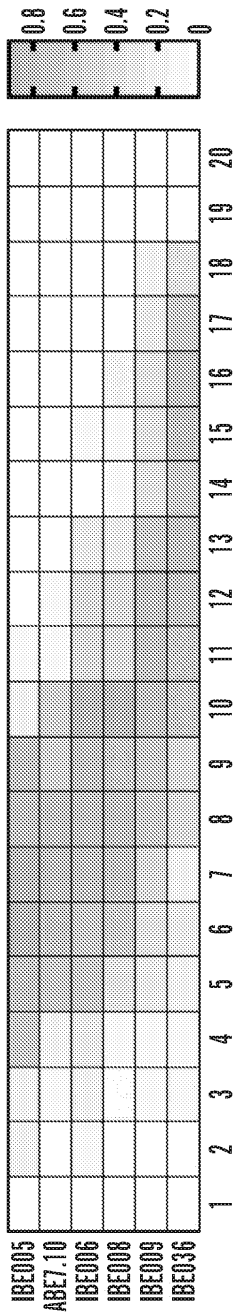
도면22b


ABE7.10 정규화된 A에서 G로의 편집
2.5
2.0
1.5
1.0
0.5
0.0
IBE02
IBE04
IBE05
IBE06
IBE08
IBE09
IBE20

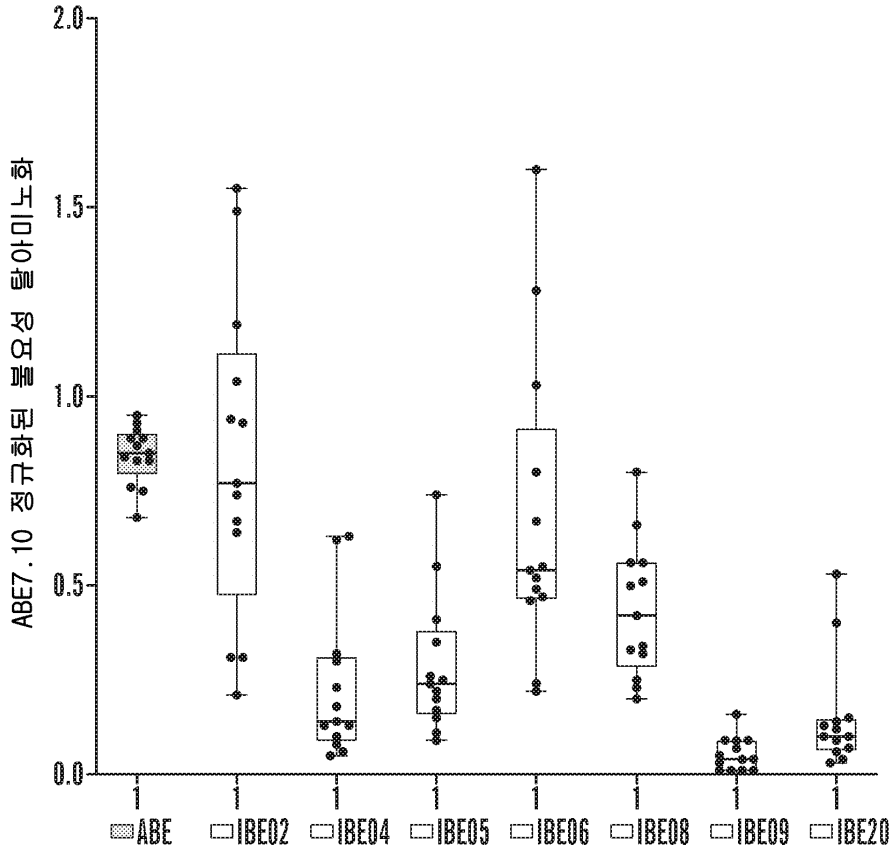
도면22c


100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
가이드 위치

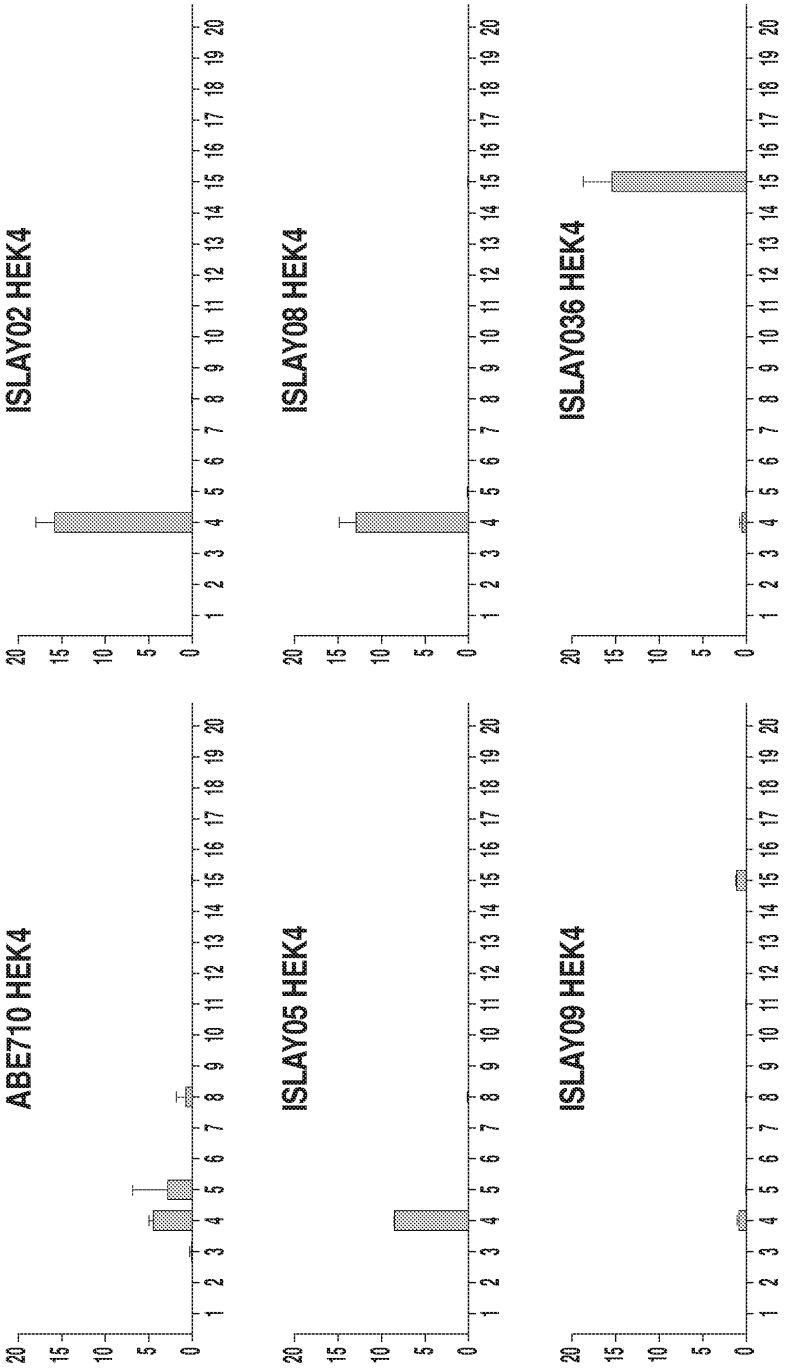
도면22d


0.8
0.6
0.4
0.2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
IBED05
ABE7.10
IBED06
IBED08
IBED09
IBED36

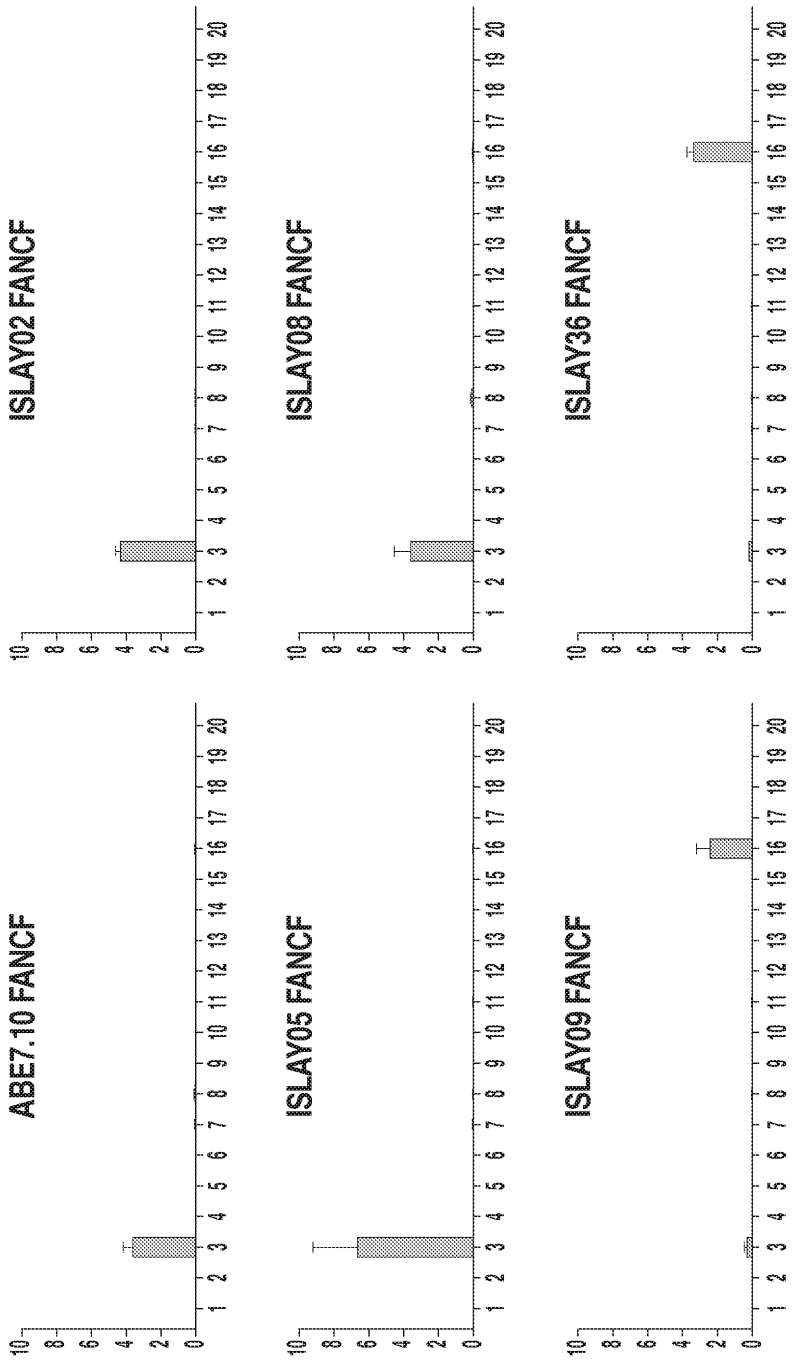
도면23


ABE7.10 정규화된 불요성 탈아미노화
2.0
1.5
1.0
0.5
0.0
ABE
IBE02
IBE04
IBE05
IBE06
IBE08
IBE09
IBE20

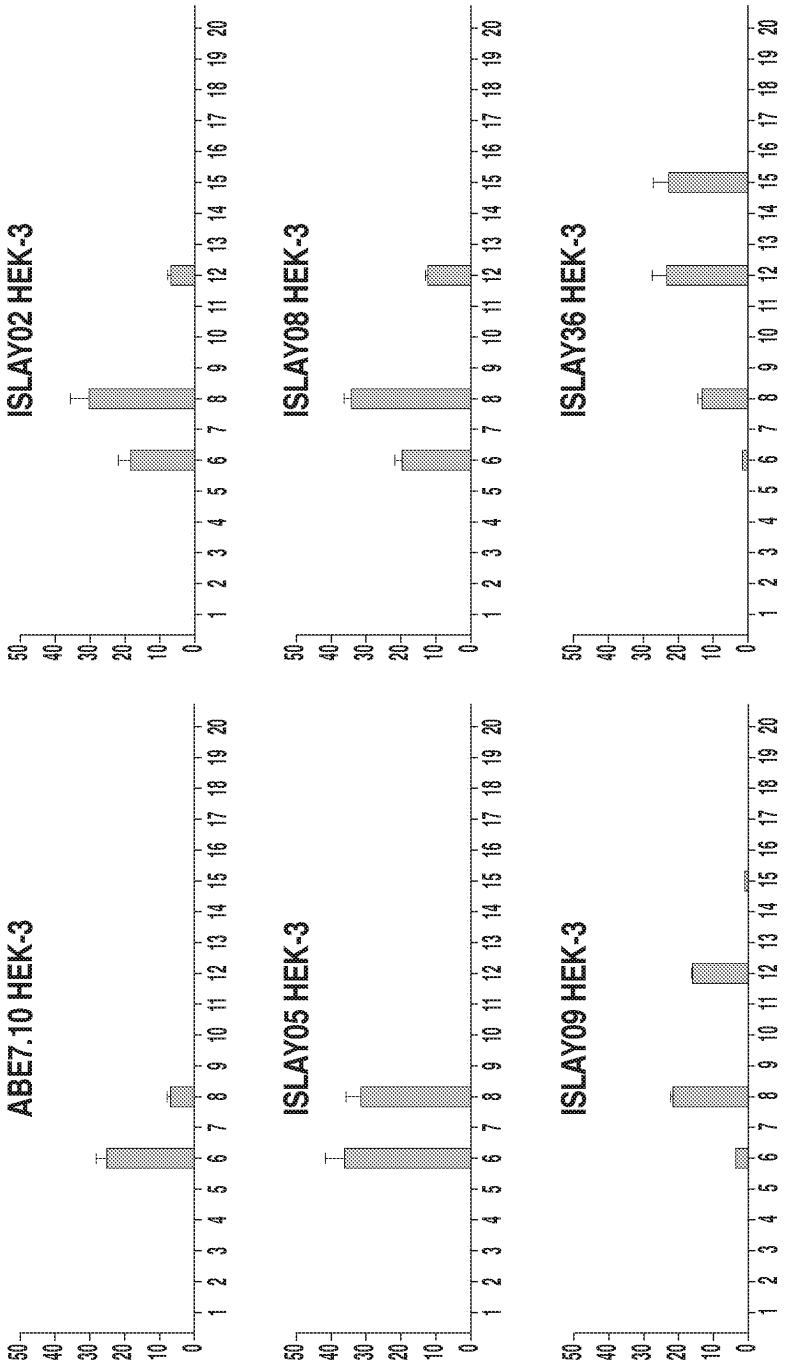
도면24a


ISLAY02 HEK4
ISLAY08 HEK4
ISLAY036 HEK4
ABE710 HEK4
ISLAY05 HEK4
ISLAY09 HEK4

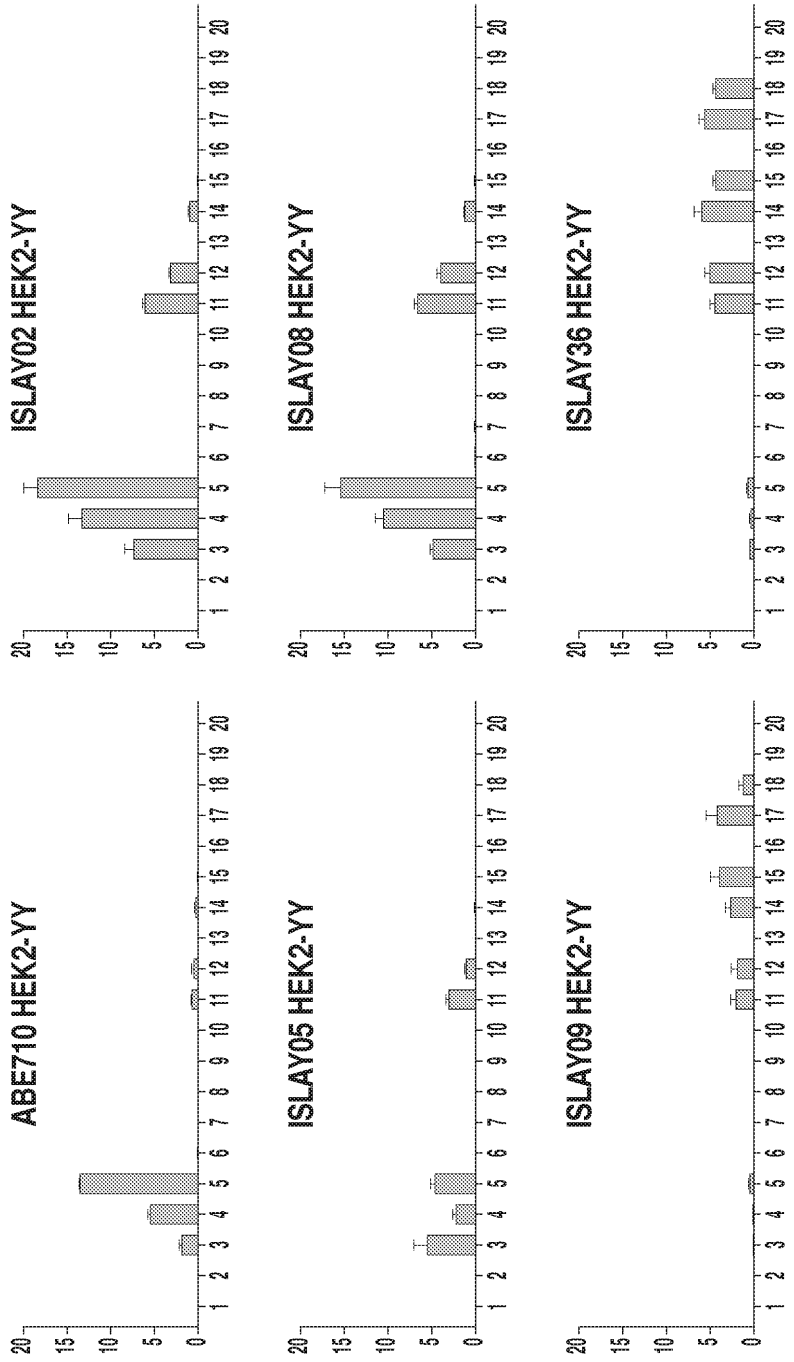
도면24b


ISLAY02 FANCF
ISLAY08 FANCF
ISLAY36 FANCF
ABE7.10 FANCF
ISLAY05 FANCF
ISLAY09 FANCF

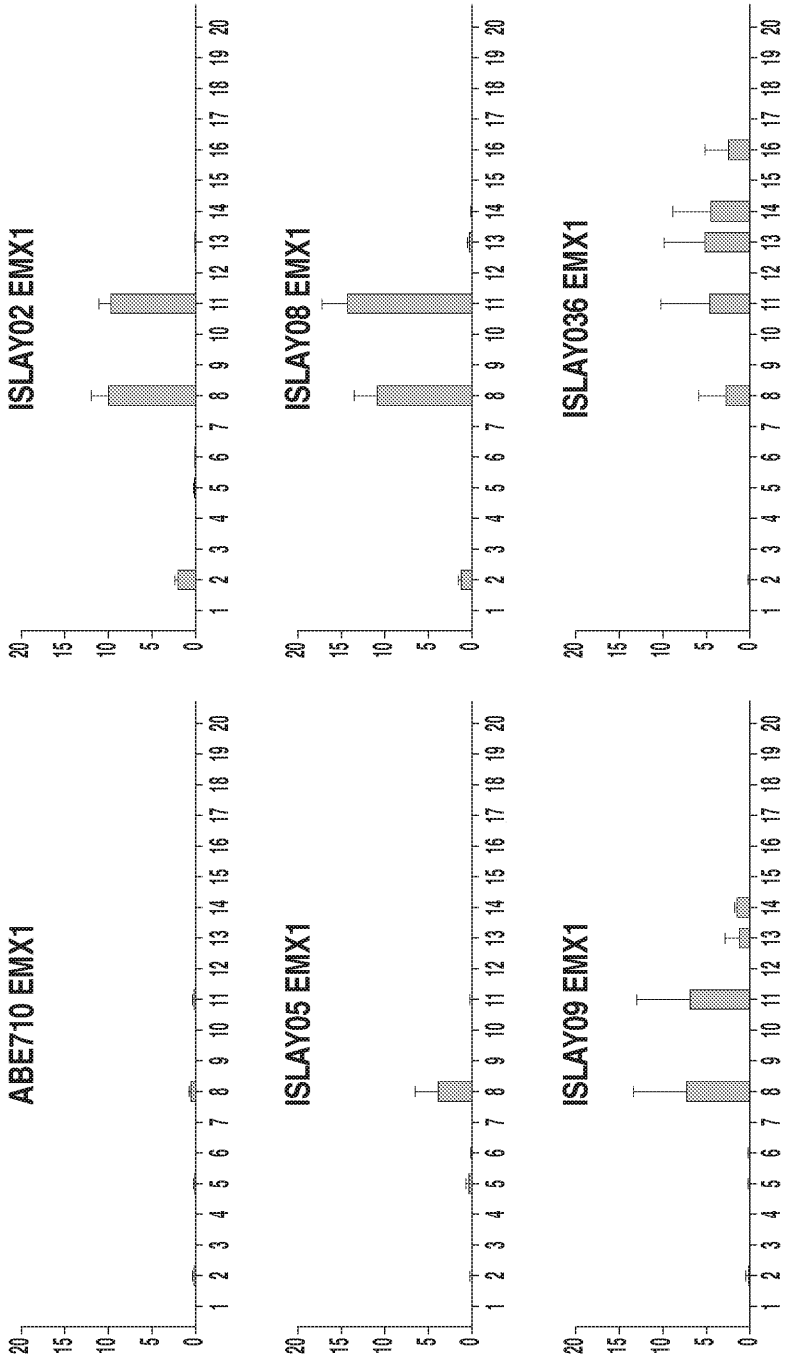
도면24c


ISLAY02 HEK-3
ISLAY08 HEK-3
ISLAY36 HEK-3
ABE7.10 HEK-3
ISLAY05 HEK-3
ISLAY09 HEK-3

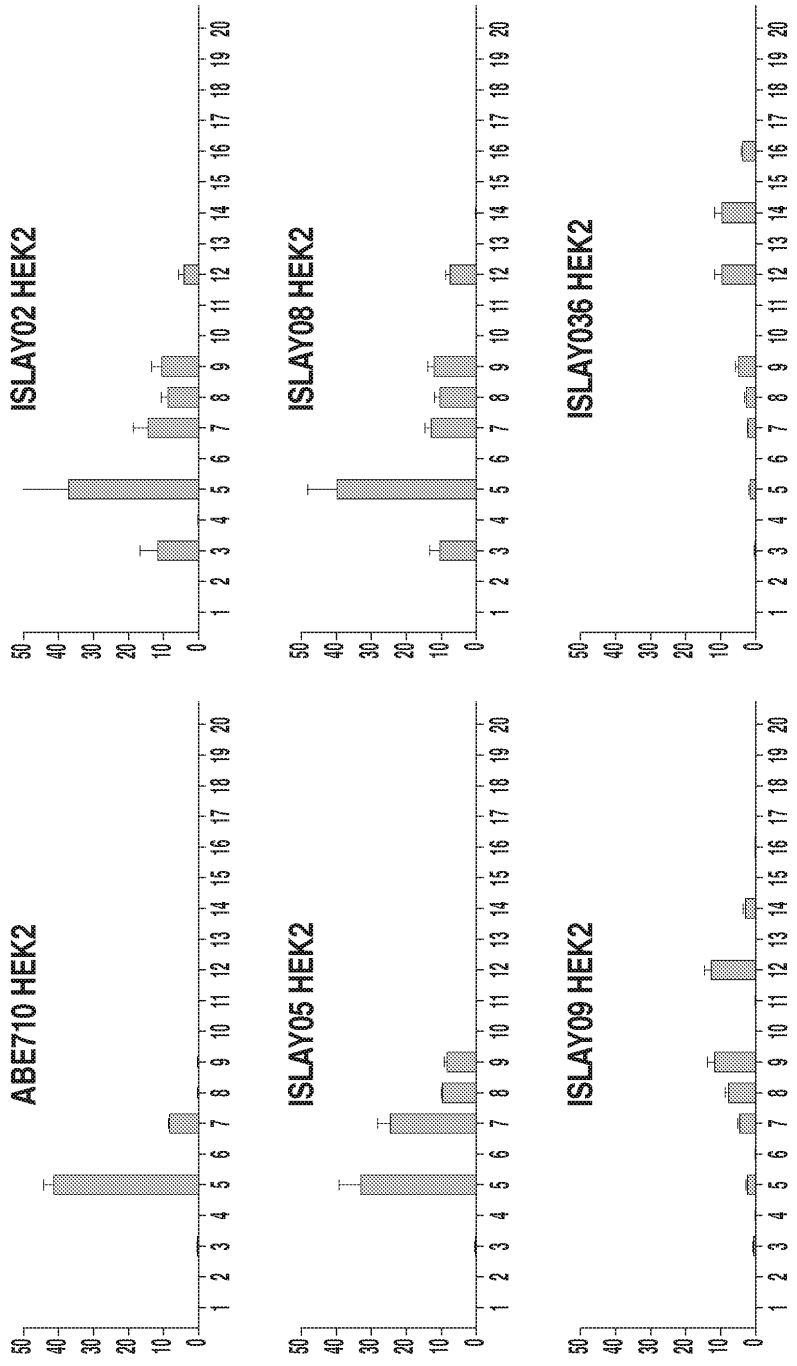
도면24d


ISLAY02 HEK2-YY
ISLAY08 HEK2-YY
ISLAY36 HEK2-YY
ABE710 HEK2-YY
ISLAY05 HEK2-YY
ISLAY09 HEK2-YY

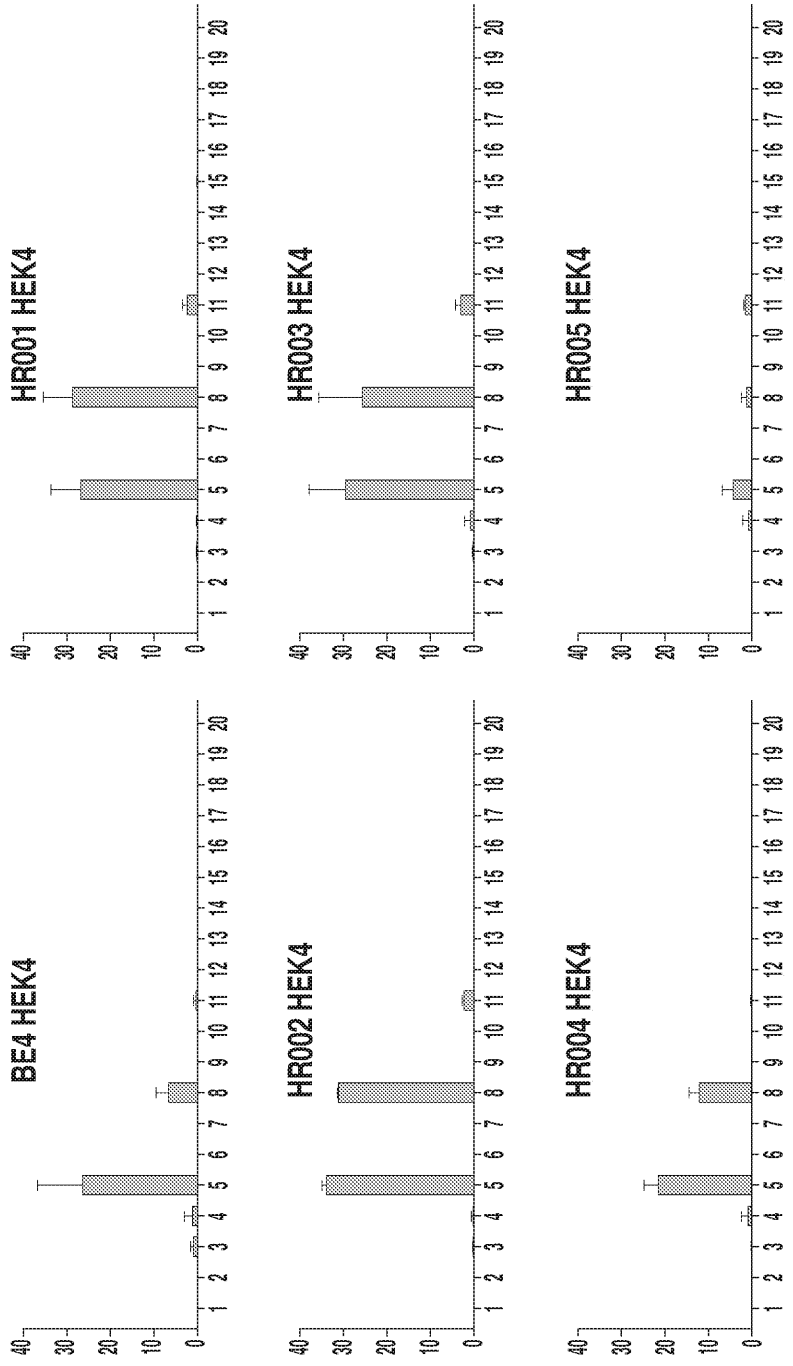
도면24e


ISLAY02 EMX1
ISLAY08 EMX1
ISLAY036 EMX1
ABE710 EMX1
ISLAY05 EMX1
ISLAY09 EMX1

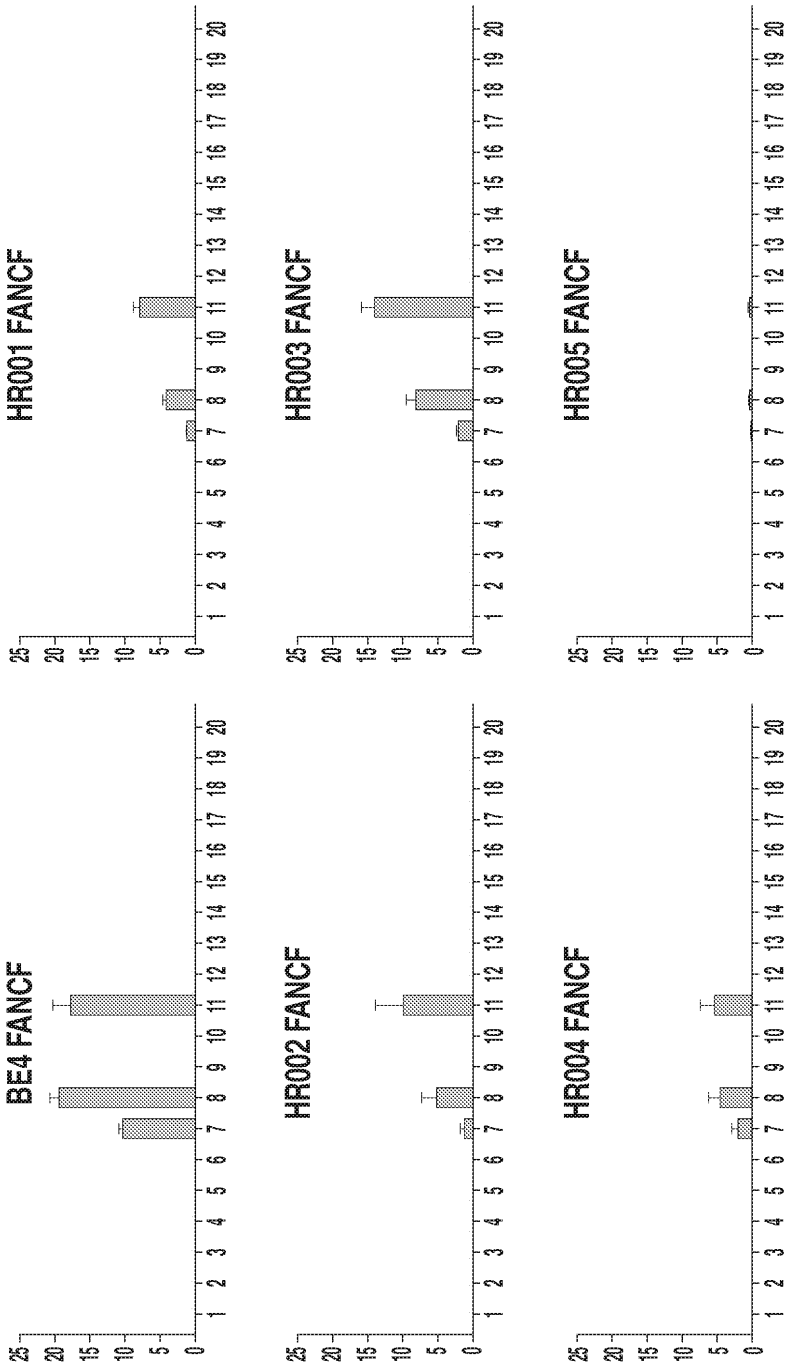
도면24f


ISLAY02 HEK2
ISLAY08 HEK2
ISLAY036 HEK2
ABE710 HEK2
ISLAY05 HEK2
ISLAY09 HEK2

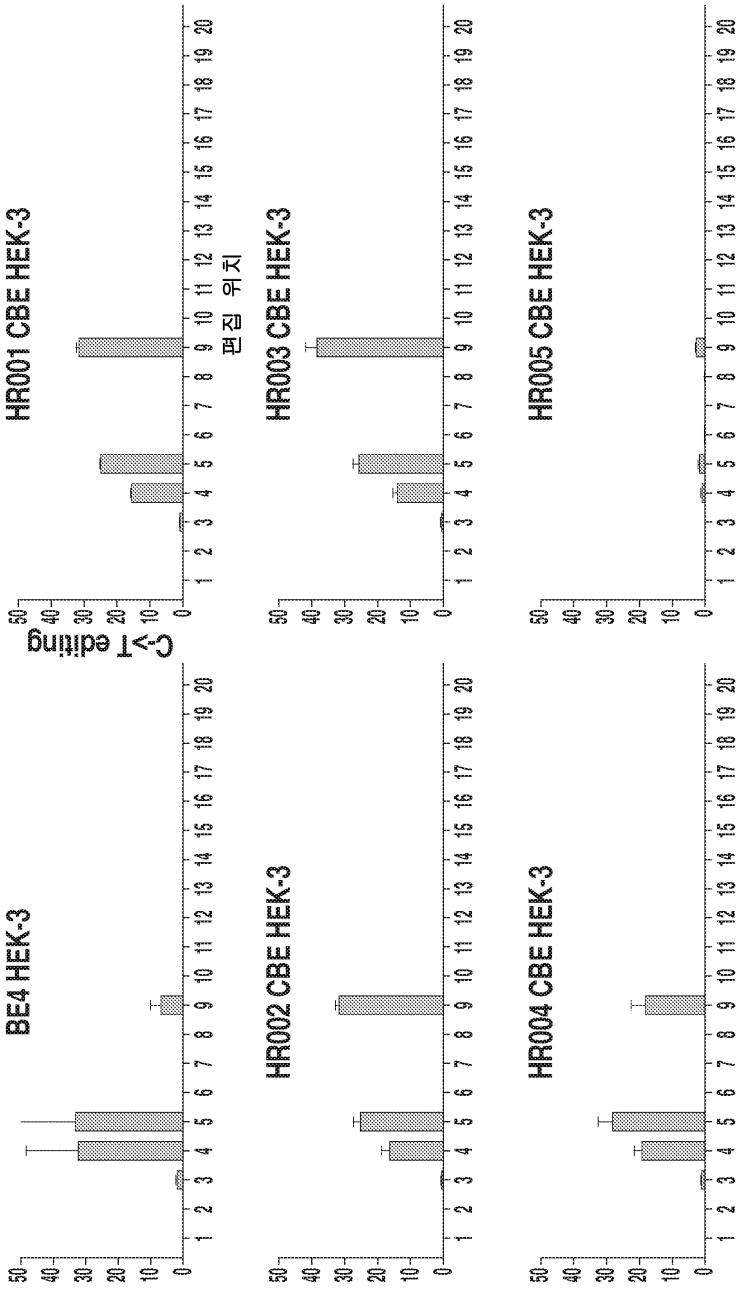
도면25a


HR001 HEK4
HR003 HEK4
HR005 HEK4
BE4 HEK4
HR002 HEK4
HR004 HEK4

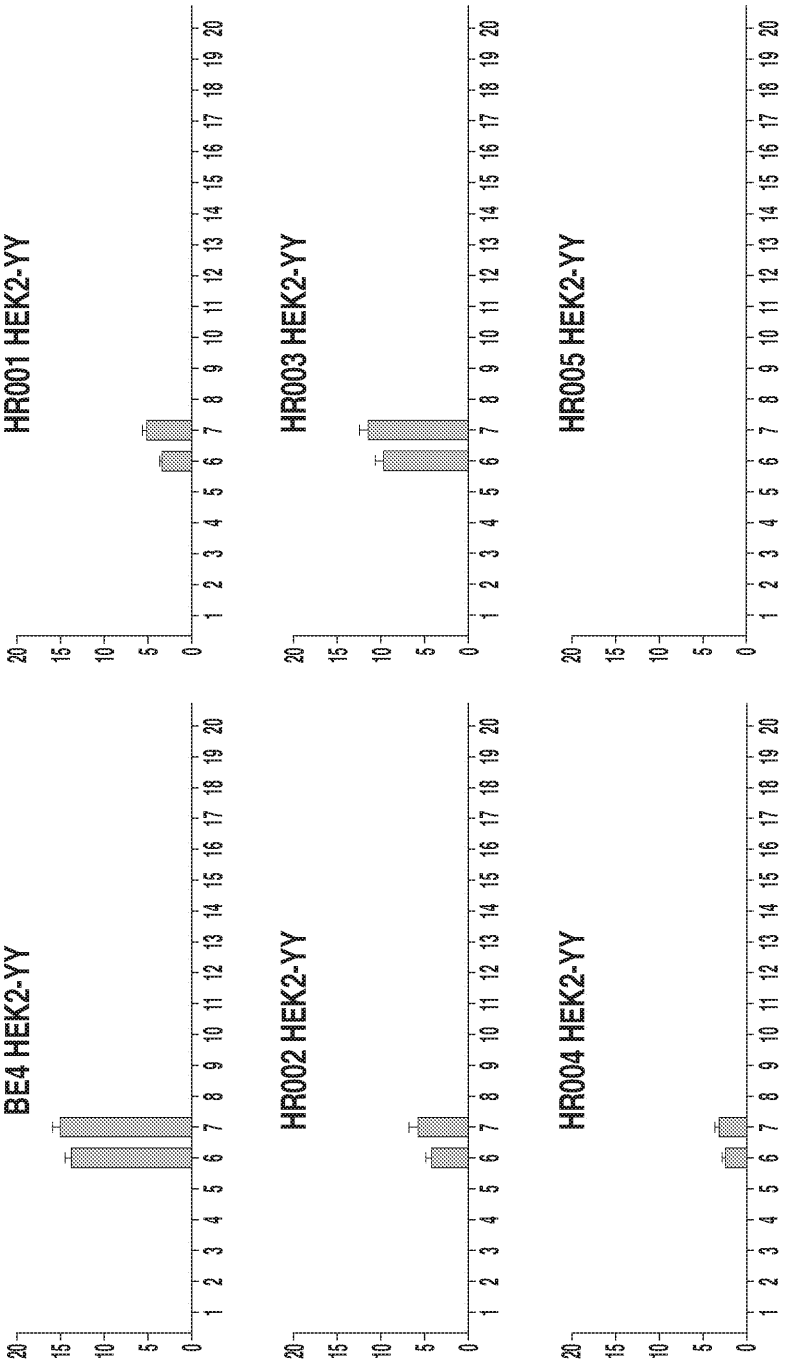
도면25b


HR001 FANCF
HR003 FANCF
HR005 FANCF
BE4 FANCF
HR002 FANCF
HR004 FANCF

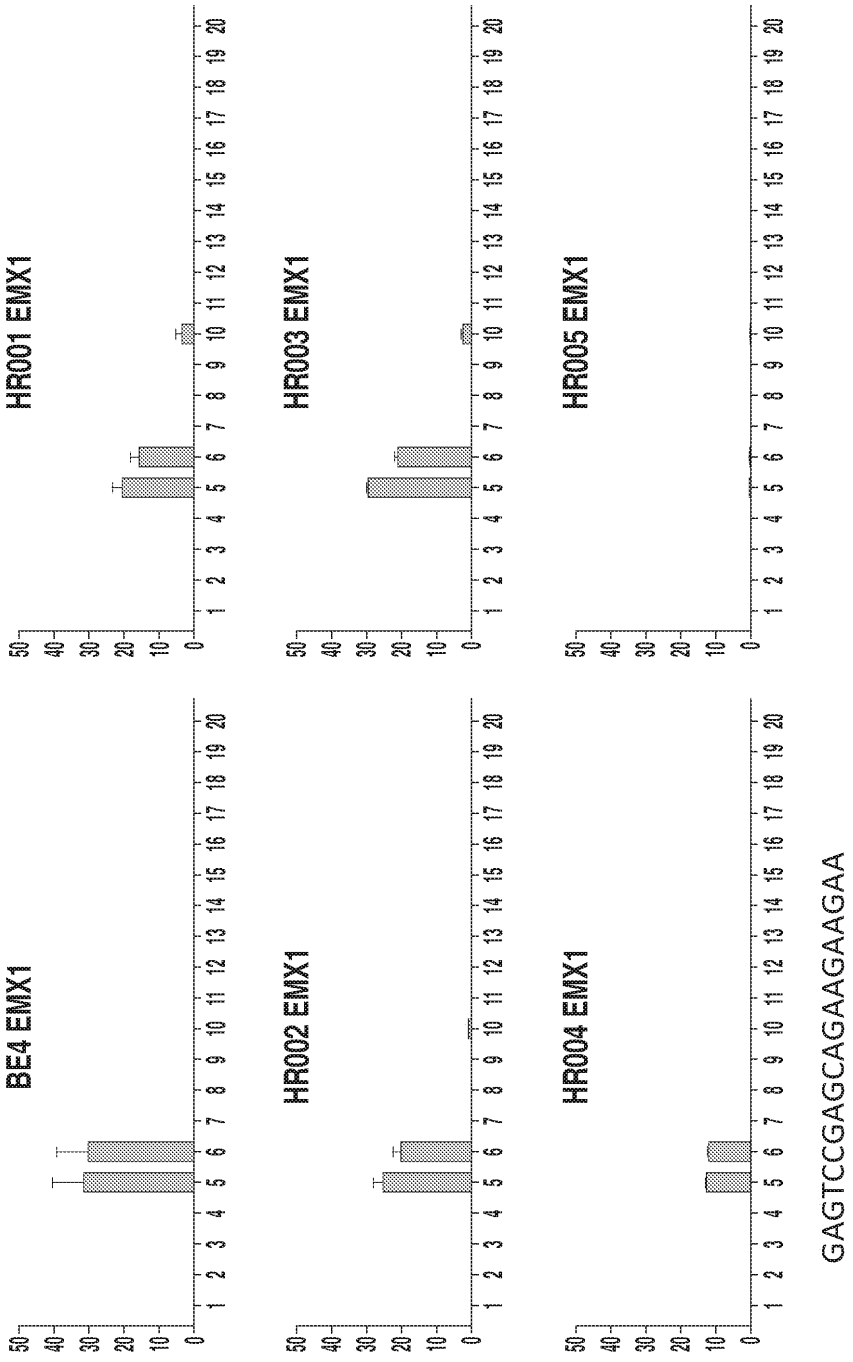
도면25c


C->T editing
편집 위치
BE4 HEK-3
HR001 CBE HEK-3
HR002 CBE HEK-3
HR003 CBE HEK-3
HR004 CBE HEK-3
HR005 CBE HEK-3

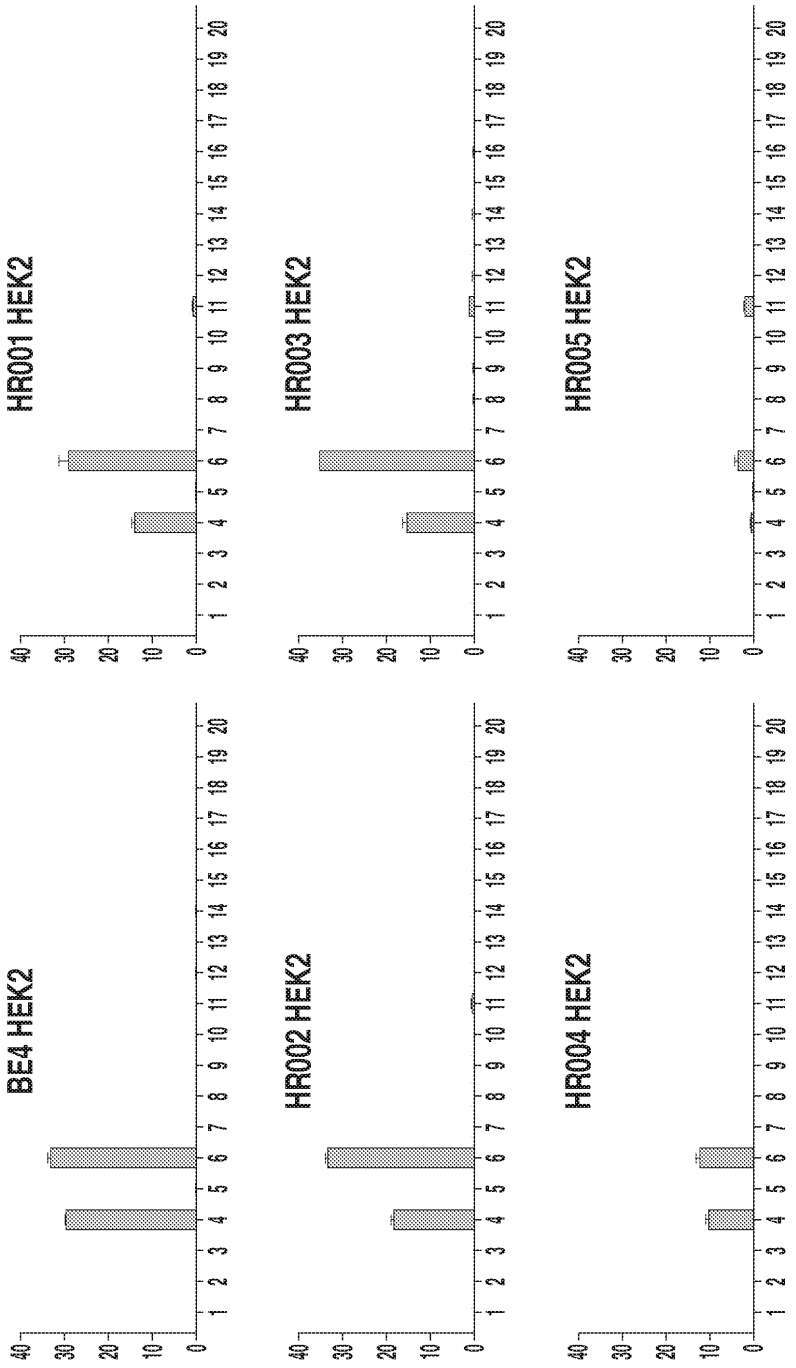
도면25d


HR001 HEK2-YY
HR003 HEK2-YY
HR005 HEK2-YY
BE4 HEK2-YY
HR002 HEK2-YY
HR004 HEK2-YY

도면25e


BE4 EMX1
HR001 EMX1
HR002 EMX1
HR003 EMX1
HR004 EMX1
HR005 EMX1
GAGTCCGAGCAGAAGAAGAA

도면25f



서 열 목 록

SEQUENCE LISTING

<110> BEAM THERAPEUTICS INC.

<120> NUCLEOBASE EDITORS HAVING REDUCED NON-TARGET DEAMINATION AND ASSAYS FOR CHARACTERIZING NUCLEOBASE EDITORS

<130> 52885-785.601

<140> PCT/US2020/016285

<141> 2020-01-31

<150> 62/799,702

<151> 2019-01-31

<160> 261

<170> PatentIn version 3.5

<210> 1

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 1

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

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

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

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

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

530 535 540

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

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

625 630 635 640

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

645 650 655

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

660 665 670

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

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

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

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

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

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

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

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

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

1115 1120 1125

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

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365
<210> 2
<211> 120
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<220><221> MISC_FEATURE

<222> (1)..(120)

<223> This sequence may encompass 1-30 "Ser Gly Gly Ser"
repeating units

<400> 2

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> 3

<211> 120

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<220><221> MISC_FEATURE

<222> (1)..(120)

<223> This sequence may encompass 1-30 "Gly Gly Gly Ser"

repeating units

<400> 3

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> 4

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<220><221> MISC_FEATURE

<222> (1)..(150)

<223> This sequence may encompass 1-30 "Gly Gly Gly Gly Ser"
repeating units

<400> 4

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> 5

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (1)..(30)

<223> This sequence may encompass 1-30 residues

<400> 5

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> 6

<211> 150

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<220><221> MISC_FEATURE

<222> (1)..(150)

<223> This sequence may encompass 1-30 "Glu Ala Ala Ala Lys" repeating units

<400> 6

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> 7

<211> 90

<212> PRT

<213> Artificial Sequence

<

220><223> Description of Artificial Sequence: Synthetic

polypeptide

<220><221> MISC_FEATURE

<222> (1)..(90)

<223> This sequence may encompass 1-30 "Gly Gly Ser"

repeating units

<400> 7

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> 8

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 8

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1 5 10 15

<210> 9

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<220><221> MOD_RES

<222> (1)..(1)

<223> Any amino acid

<220><221> MOD_RES

<222> (3)..(3)

<223> Any amino acid

<220><221> MOD_RES

<222> (5)..(5)

<223> Any amino acid

<220><221> MOD_RES

<222> (7)..(7)

<223> Any amino acid

<220><221> MOD_RES

<222> (9)..(9)

<223> Any amino acid

<220><221> MOD_RES

<222> (11)..(11)

<223> Any amino acid

<220><221> MOD_RES

<222> (13)..(13)

<223> Any amino acid

<220><221> MOD_RES

<222> (15)..(15)

<223> Any amino acid

<220><221> MOD_RES

<222> (17)..(17)

<223> Any amino acid

<220><221> MOD_RES

<222> (19)..(19)

<223> Any amino acid

<220><221> MOD_RES

<222> (21)..(21)

<223> Any amino acid

<220><221> MOD_RES

<222> (23)..(23)

<223> Any amino acid

<220><221> MOD_RES

<222> (25)..(25)

<223> Any amino acid

<220><221> MOD_RES

<222> (27)..(27)

<223> Any amino acid

<220><221> MOD_RES

<222> (29)..(29)

<223> Any amino acid

<220><221> MOD_RES

<222> (31)..(31)

<223> Any amino acid

<220><221> MOD_RES

<222> (33)..(33)

<223> Any amino acid

<220><221> MOD_RES

<222> (35)..(35)

<223> Any amino acid

<220><221> MOD_RES

<222> (37)..(37)

<223> Any amino acid

<220><221> MOD_RES

<222> (39)..(39)

<223> Any amino acid

<220><221

> MOD_RES

<222> (41)..(41)

<223> Any amino acid

<220><221> MOD_RES

<222> (43)..(43)

<223> Any amino acid

<220><221> MOD_RES

<222> (45)..(45)

<223> Any amino acid

<220><221> MOD_RES

<222> (47)..(47)

<223> Any amino acid

<220><221> MOD_RES

<222> (49)..(49)

<223> Any amino acid

<220><221> MOD_RES

<222> (51)..(51)

<223> Any amino acid

<220><221> MOD_RES

<222> (53)..(53)

<223> Any amino acid

<220><221> MOD_RES

<222> (55)..(55)

<223> Any amino acid

<220><221> MOD_RES

<222> (57)..(57)

<223> Any amino acid

<220><221> MOD_RES

<222> (59)..(59)

<223> Any amino acid

<220><221> MISC_FEATURE

<222> (1)..(60)

<223> This sequence may encompass 1-30 "Xaa Pro"

repeating units

<400> 9

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

1 5 10 15

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

20 25 30

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

35 40 45

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro

50 55 60

<210> 10

<211> 167

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Adenosine deaminase sequence

<400> 10

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro

35 40 45

Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

65 70 75 80

Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp
165

<210> 11

<211> 178

<212> PRT

<213> Escherichia coli

<400> 11

Met Arg Arg Ala Phe Ile Thr Gly Val Phe Phe Leu Ser Glu Val Glu
1 5 10 15

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg
20 25 30

Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val His Asn
35 40 45

Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile Gly Arg His Asp
50 55 60

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
65 70 75 80

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Leu Glu
85 90 95

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
100 105 110

Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala Ala Gly Ser Leu
115 120 125

Met Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr
130 135 140

Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Ser Asp Phe Phe
145 150 155 160

Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys Ala Gln Ser Ser
165 170 175

Thr Asp

<210> 12

<211> 160

<212> PRT

<213> Staphylococcus aureus

<400> 12

Met Gly Ser His Met Thr Asn Asp Ile Tyr Phe Met Thr Leu Ala Ile
1 5 10 15

Glu Glu Ala Lys Lys Ala Ala Gln Leu Gly Glu Val Pro Ile Gly Ala
20 25 30

Ile Ile Thr Lys Asp Asp Glu Val Ile Ala Arg Ala His Asn Leu Arg
35 40 45

Glu Thr Leu Gln Gln Pro Thr Ala His Ala Glu His Ile Ala Ile Glu
50 55 60

Arg Ala Ala Lys Val Leu Gly Ser Trp Arg Leu Glu Gly Cys Thr Leu
65 70 75 80

Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Thr Ile Val Met
85 90 95

Ser Arg Ile Pro Arg Val Val Tyr Gly Ala Asp Asp Pro Lys Gly Gly
100 105 110

Cys Ser Gly Ser Leu Met Asn Leu Leu Gln Gln Ser Asn Phe Asn His
115 120 125

Arg Ala Ile Val Asp Lys Gly Val Leu Lys Glu Ala Cys Ser Thr Leu
130 135 140

Leu Thr Thr Phe Phe Lys Asn Leu Arg Ala Asn Lys Lys Ser Thr Asn
145 150 155 160

<210> 13

<211> 161

<212> PRT

<213> Bacillus subtilis

<400> 13

Met Thr Gln Asp Glu Leu Tyr Met Lys Glu Ala Ile Lys Glu Ala Lys

1 5 10 15

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

Glu Pro Cys Pro Met Cys Ala Gly Ala Val Val Leu Ser Arg Val Glu
85 90 95
Lys Val Val Phe Gly Ala Phe Asp Pro Lys Gly Gly Cys Ser Gly Thr
100 105 110
Leu Met Asn Leu Leu Gln Glu Glu Arg Phe Asn His Gln Ala Glu Val
115 120 125
Val Ser Gly Val Leu Glu Glu Glu Cys Gly Gly Met Leu Ser Ala Phe
130 135 140

Phe Arg Glu Leu Arg Lys Lys Lys Lys Ala Ala Arg Lys Asn Leu Ser
145 150 155 160
Glu

<210> 14
<211> 183
<212> PRT
<213> Salmonella typhimurium
<400> 14
Met Pro Pro Ala Phe Ile Thr Gly Val Thr Ser Leu Ser Asp Val Glu
1 5 10 15
Leu Asp His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg
20 25 30
Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val His Asn

35 40 45

His Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile Gly Arg His Asp

50 55 60

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val

65 70 75 80

Leu Gln Asn Tyr Arg Leu Leu Asp Thr Thr Leu Tyr Val Thr Leu Glu

85 90 95

Pro Cys Val Met Cys Ala Gly Ala Met Val His Ser Arg Ile Gly Arg

100 105 110

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

115 120 125

Ile Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Ile

130 135 140

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

145 150 155 160

Arg Met Arg Arg Gln Glu Ile Lys Ala Leu Lys Lys Ala Asp Arg Ala

165 170 175

Glu Gly Ala Gly Pro Ala Val

180

<210> 15

<211> 164

<212> PRT

<213> Shewanella putrefaciens

<400> 15

Met Asp Glu Tyr Trp Met Gln Val Ala Met Gln Met Ala Glu Lys Ala

1 5 10 15

Glu Ala Ala Gly Glu Val Pro Val Gly Ala Val Leu Val Lys Asp Gly

20 25 30

Gln Gln Ile Ala Thr Gly Tyr Asn Leu Ser Ile Ser Gln His Asp Pro

35 40 45

Thr Ala His Ala Glu Ile Leu Cys Leu Arg Ser Ala Gly Lys Lys Leu

50 55 60

Glu Asn Tyr Arg Leu Leu Asp Ala Thr Leu Tyr Ile Thr Leu Glu Pro
65 70 75 80
Cys Ala Met Cys Ala Gly Ala Met Val His Ser Arg Ile Ala Arg Val
85 90 95
Val Tyr Gly Ala Arg Asp Glu Lys Thr Gly Ala Ala Gly Thr Val Val

100 105 110
Asn Leu Leu Gln His Pro Ala Phe Asn His Gln Val Glu Val Thr Ser
115 120 125
Gly Val Leu Ala Glu Ala Cys Ser Ala Gln Leu Ser Arg Phe Phe Lys
130 135 140
Arg Arg Arg Asp Glu Lys Lys Ala Leu Lys Leu Ala Gln Arg Ala Gln
145 150 155 160
Gln Gly Ile Glu

<210> 16

<211> 173
<212> PRT
<213> Haemophilus influenzae
<400> 16
Met Asp Ala Ala Lys Val Arg Ser Glu Phe Asp Glu Lys Met Met Arg
1 5 10 15
Tyr Ala Leu Glu Leu Ala Asp Lys Ala Glu Ala Leu Gly Glu Ile Pro
20 25 30
Val Gly Ala Val Leu Val Asp Asp Ala Arg Asn Ile Ile Gly Glu Gly
35 40 45
Trp Asn Leu Ser Ile Val Gln Ser Asp Pro Thr Ala His Ala Glu Ile

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

100 105 110

Tyr Lys Thr Gly Ala Ile Gly Ser Arg Phe His Phe Phe Asp Asp Tyr

115 120 125

Lys Met Asn His Thr Leu Glu Ile Thr Ser Gly Val Leu Ala Glu Glu

130 135 140

Cys Ser Gln Lys Leu Ser Thr Phe Phe Gln Lys Arg Arg Glu Glu Lys

145 150 155 160

Lys Ile Glu Lys Ala Leu Leu Lys Ser Leu Ser Asp Lys

165 170

<210> 17

<211> 161

<212> PRT

<213> Caulobacter crescentus

<400> 17

Met Arg Thr Asp Glu Ser Glu Asp Gln Asp His Arg Met Met Arg Leu

1 5 10 15

Ala Leu Asp Ala Ala Arg Ala Ala Ala Glu Ala Gly Glu Thr Pro Val

20 25 30

Gly Ala Val Ile Leu Asp Pro Ser Thr Gly Glu Val Ile Ala Thr Ala

35 40 45

Gly Asn Gly Pro Ile Ala Ala His Asp Pro Thr Ala His Ala Glu Ile

50 55 60

Ala Ala Met Arg Ala Ala Ala Ala Lys Leu Gly Asn Tyr Arg Leu Thr

65 70 75 80

Asp Leu Thr Leu Val Val Thr Leu Glu Pro Cys Ala Met Cys Ala Gly

85 90 95

Ala Ile Ser His Ala Arg Ile Gly Arg Val Val Phe Gly Ala Asp Asp

100 105 110

Pro Lys Gly Gly Ala Val Val His Gly Pro Lys Phe Phe Ala Gln Pro

115 120 125

Thr Cys His Trp Arg Pro Glu Val Thr Gly Gly Val Leu Ala Asp Glu

130 135 140

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

145 150 155 160

Ile

<210> 18

<211> 179

<212> PRT

<213> Geobacter sulfurreducens

<400> 18

Met Ser Ser Leu Lys Lys Thr Pro Ile Arg Asp Asp Ala Tyr Trp Met

1 5 10 15

Gly Lys Ala Ile Arg Glu Ala Ala Lys Ala Ala Ala Arg Asp Glu Val

20 25 30

Pro Ile Gly Ala Val Ile Val Arg Asp Gly Ala Val Ile Gly Arg Gly

35 40 45

His Asn Leu Arg Glu Gly Ser Asn Asp Pro Ser Ala His Ala Glu Met

50 55 60

Ile Ala Ile Arg Gln Ala Ala Arg Arg Ser Ala Asn Trp Arg Leu Thr

65 70 75 80

Gly Ala Thr Leu Tyr Val Thr Leu Glu Pro Cys Leu Met Cys Met Gly

85 90 95

Ala Ile Ile Leu Ala Arg Leu Glu Arg Val Val Phe Gly Cys Tyr Asp

100 105 110

Pro Lys Gly Gly Ala Ala Gly Ser Leu Tyr Asp Leu Ser Ala Asp Pro

115 120 125

Arg Leu Asn His Gln Val Arg Leu Ser Pro Gly Val Cys Gln Glu Glu

130 135 140

Cys Gly Thr Met Leu Ser Asp Phe Phe Arg Asp Leu Arg Arg Arg Lys

145 150 155 160

Lys Ala Lys Ala Thr Pro Ala Leu Phe Ile Asp Glu Arg Lys Val Pro

165 170 175

Pro Glu Pro

<210> 19

<211> 167

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 19

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp
165

<210> 20

<211> 186

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 20

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1 5 10 15

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
20 25 30

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro
35 40 45

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp
50 55 60

Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met
65 70 75 80

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp
85 90 95

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala
100 105 110

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala
115 120 125

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly
130 135 140

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys
145 150 155 160

Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn
165 170 175

Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp
180 185

<210> 21

<211> 186

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 21

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met
1 5 10 15

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro
20 25 30

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
35 40 45

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
50 55 60

Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
65 70 75 80

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg
85 90 95

Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr
100 105 110

Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr
115 120 125

Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met
130 135 140

Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val
145 150 155 160

Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly
165 170 175

Trp Asn Arg Ala Ile Gly Leu His Asp Pro
180 185

<210> 22

<211> 186

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 22

Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val

1 5 10 15

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

20 25 30

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val

35 40 45

Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile

50 55 60

Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro

65 70 75 80

Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly

85 90 95

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu

100 105 110

Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His

115 120 125

Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val

130 135 140

Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn

145 150 155 160

Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala

165 170 175

Leu Arg Gln Gly Gly Leu Val Met Gln Asn

180 185

<210> 23

<211> 186

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 23

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys
1 5 10 15

Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn
20 25 30

Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu
35 40 45

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu
50 55 60

Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala
65 70 75 80

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val
85 90 95

Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu
100 105 110

His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly
115 120 125

Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr
130 135 140

Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile
145 150 155 160

Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly
165 170 175

Ser Leu Met Asp Val Leu His Tyr Pro Gly
180 185

<210> 24

<211> 176

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 24

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1 5 10 15

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
20 25 30

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro
35 40 45

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp
50 55 60

Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met
65 70 75 80

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp
85 90 95

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala
100 105 110

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala
115 120 125

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly
130 135 140

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys
145 150 155 160

Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn
165 170 175

<210> 25

<211> 173

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 25

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

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

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala
115 120 125
Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly
130 135 140
Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys
145 150 155 160
Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln
165 170

<210> 26

<211> 168
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 26
Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1 5 10 15
Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg

20 25 30

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

35 40 45

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp

50 55 60

Asn Arg Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val

65 70 75 80

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

85 90 95

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg

100 105 110

Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu

115 120 125

Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr

130 135 140

Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe

145 150 155 160

Arg Met Pro Arg Gln Val Phe Asn

165

<210> 27

<211> 178

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 27

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

1 5 10 15

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg

20 25 30

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

35 40 45

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp

50 55 60

Asn Arg Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
65 70 75 80

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu
85 90 95

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
100 105 110

Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu

115 120 125

Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr
130 135 140

Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe
145 150 155 160

Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser
165 170 175

Thr Asp

<210> 28

<211> 1410

<212

> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 28

Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met
1 5 10 15

Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys
20 25 30

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
35 40 45

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

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
115 120 125
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg
130 135 140
Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr
145 150 155 160
Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
165 170 175

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

Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala
245 250 255
Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp
260 265 270
Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
275 280 285
Thr Asn Leu Gly Ala Pro Arg Ala Phe Lys Tyr Phe Asp Thr Thr Ile

290 295 300

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

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

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

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

Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn
545 550 555 560

Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg
565 570 575

Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly
580 585 590

Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly
595 600 605

Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys
610 615 620

Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu
625 630 635 640

Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn
645 650 655

Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu
660 665 670

Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu
675 680 685

His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu
690 695 700

Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr
705 710 715 720

Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe
725 730 735

Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val
740 745 750

Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn
755 760 765

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

Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys

785 790 795 800

Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu

805 810 815

Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu

820 825 830

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

835 840 845

Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro

850 855 860

Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr

865 870 875 880

Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg

885 890 895

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

900 905 910

Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp

915 920 925

Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val

930 935 940

Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys

945 950 955 960

Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile

965 970 975

Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met

980 985 990

Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val

995 1000 1005

Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu

1010 1015 1020

Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys

1025 1030 1035

Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn
1040 1045 1050
Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp
1055 1060 1065

Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu
1070 1075 1080
His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
1085 1090 1095
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
1100 1105 1110
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn
1115 1120 1125

Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys
1130 1135 1140
Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
1145 1150 1155
Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr
1160 1165 1170
Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu
1175 1180 1185

Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val
1190 1195 1200
Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu
1205 1210 1215
Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser
1220 1225 1230
Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu
1235 1240 1245

Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys
1250 1255 1260
Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile

1265 1270 1275

Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala

1280 1285 1290

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

1295 1300 1305

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

1310 1315 1320

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

1325 1330 1335

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

1340 1345 1350

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

1355 1360 1365

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

1370 1375 1380

Ala Lys Ser Glu Gln Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly

1385 1390 1395

Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

1400 1405 1410

<210> 29

<211> 4104

<212> DNA

<213> Streptococcus pyogenes

<400> 29

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300

cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540
gtggacaaac tatttatcca gttggtacaa atctacaatc aattatttga agaaaaccct 600
attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttgggaat 720
ctcattgctt tgtcattggg attgacccct aattttaaat caaattttga tttggcagaa 780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840
caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatagt gaaataacta aggctcccct atcagcttca 960
atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380
cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440
gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500
aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560
tataacgaat tgacaaaggt caaatatgtt actgagggaa tgcgaaaacc agcatttctt 1620
tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt 1800
attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860
ttaacattga ccttatttga agataggggg atgattgagg aaagacttaa aacatatgct 1920
cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040
gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100
agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttta 2160

catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggtat tttacagact 2220

gtaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcgttatt 2280

gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg 2340

aaacgaatcg aagaaggtat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt 2400

gaaaatactc aattgcaaaa tgaaaagctc tatctctatt atctacaaaa tggaagagac 2460

atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt 2520

gttccacaaa gtttcattaa agacgattca atagacaata aggtactaac gcgttctgat 2580

aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac 2640

tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg 2700

aaagctgaac gtggaggttt gagtgaactt gataaagctg gttttatcaa acgccaattg 2760

gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtcg catgaatact 2820

aaatacgatg aaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa 2880

ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac 2940

catcatgccc atgatgcgta tctaaatgcc gtcgttggaa ctgctttgat taagaaatat 3000

ccaaaacttg aatcggagtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg 3060

attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaat 3120

atcatgaact tcttcaaaac agaaattaca cttgcaaatg gagagattcg caaacgccct 3180

ctaatcgaaa ctaatgggga aactggagaa attgtctggg ataaagggcg agattttgcc 3240

acagtgcgca aagtattgtc catgccccaa gtcaatattg tcaagaaaac agaagtacag 3300

acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct 3360

cgtaaaaaag actgggatcc aaaaaaatat ggtggttttg atagtccaac ggtagcttat 3420

tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa 3480

gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaaatcc gattgacttt 3540

ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat 3600

agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa 3660

aaaggaaatg agctggctct gccaagcaaa tatgtgaatt ttttatattt agctagtcat 3720

tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag 3780

cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt 3840

ttagcagatg ccaatttaga taaagttctt agtgcatata acaaacatag agacaaacca 3900

atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc 3960

gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc tacaaaagaa 4020

gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat 4080

ttgagtcagc taggaggtga ctga 4104

<210> 30

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 30

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> 31

<211> 4212

<212> DNA

<213> Streptococcus pyogenes

<400> 31

atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc 60

ataaccgatg aatacaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt 120

cattcgatta aaaagaatct tatcggtgcc ctcctattcg atagtggcga aacggcagag 180
gcgactcgcc tgaaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt 240
tacttacaag aaatttttag caatgagatg gccaaagttg acgattcttt ctttcaccgt 300
ttggaagagt ccttccttgt cgaagaggac aagaaacatg aacggcaccc catctttgga 360
aacatagtag atgaggtggc atatcatgaa aagtacccaa cgatttatca cctcagaaaa 420
aagctagttg actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat 480
atgataaagt tccgtgggca ctttctcatt gagggtgatc taaatccgga caactcggat 540

gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct 600
ataaatgcaa gtggcgtgga tgcgaaggct attcttagcg cccgcctctc taaatcccga 660
cggctagaaa acctgatcgc acaattaccc ggagagaaga aaaatgggtt gttcggtaac 720
cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa 780
gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca 840
caaattggag atcagtatgc ggacttattt ttggctgcca aaaaccttag cgatgcaatc 900
ctcctatctg acatactgag agttaatact gagattacca aggcgccgtt atccgcttca 960

atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc cctagtccgt 1020
cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca 1080
ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta 1140
gagaagatgg atgggacgga agagttgctt gtaaaactca atcgcgaaga tctactgcga 1200
aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat 1260
gctatactta gaaggcagga ggatttttat ccgttcctca aagacaatcg tgaaaagatt 1320
gagaaaatcc taacctttcg cataccttac tatgtgggac ccctggcccg agggaactct 1380

cggttcgcat ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa 1440
gttgtcgata aaggtgcgtc agctcaatcg ttcatcgaga ggatgaccaa ctttgacaag 1500
aatttaccga acgaaaaagt attgcctaag cacagtttac tttacgagta tttcacagtg 1560
tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta 1620
agcggagaac agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca 1680
gttaagcaat tgaaagagga ctactttaag aaaattgaat gcttcgattc tgtcgagatc 1740
tccggggtag aagatcgatt taatgcgtca cttggtacgt atcatgacct cctaaagata 1800

attaaagata aggacttcct ggataacgaa gagaatgaag atatcttaga agatatagtg 1860

ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgct 1920

cacctgttcg acgataaggt tatgaaacag ttaaagaggc gtcgctatac gggctgggga 1980

cgattgtcgc ggaaacttat caacgggata agagacaagc aaagtggtaa aactattctc 2040

gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac 2100

tctttaacct tcaaagagga tatacaaaag gcacaggttt ccggacaagg ggactcattg 2160

cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaagggcat actccagaca 2220

gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaaccgga aaacattgta 2280

atcgagatgg cacgcgaaaa tcaaacgact cagaaggggc aaaaaaacag tcgagagcgg 2340

atgaagagaa tagaagaggg tattaaagaa ctgggcagcc agatcttaaa ggagcatcct 2400

gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg 2460

gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatcac 2520

attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgct tacacgctcg 2580

gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag 2640

aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagtt cgataactta 2700

actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccggatttat taaacgtcag 2760

ctcgtggaaa cccgccaaat cacaaagcat gttgcacaga tactagattc ccgaatgaat 2820

acgaaatacg acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca 2880

aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940

taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000

tacccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060

atgatcgcga aaagcgaaca ggagataggc aaggctacag ccaaatactt cttttattct 3120

aacattatga atttctttaa gacggaaatc actctggcaa acggagagat acgcaaacga 3180

cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc 3240

gcgacggtga gaaaagtttt gtccatgccc caagtcaaca tagtaaagaa aactgaggtg 3300

cagaccggag ggttttcaaa ggaatcgatt cttccaaaaa ggaatagtga taagctcatc 3360

gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc 3420

tattctgtcc tagtagtggc aaaagttgag aagggaaaat ccaagaaact gaagtcagtc 3480

aaagaattat tggggataac gattatggag cgctcgtctt ttgaaaagaa ccccatcgac 3540

ttccttgagg cgaaaggtta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600

tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660

caaaagggga acgaactcgc actaccgtct aaatacgtga atttcctgta tttagcgtcc 3720

cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact ttttgttgag 3780

cagcacaaac attatctcga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840

atcctagctg atgccaatct ggacaaagta ttaagcgcat acaacaagca cagggataaa 3900

cccatacgtg agcaggcgga aaatattatc catttgttta ctcttaccaa cctcggcgct 3960

ccagccgcat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020

gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080

gatttgtcac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140

tacaaagacc atgacggtga ttataaagat catgacatcg attacaagga tgacgatgac 4200

aaggctgcag ga 4212

<210> 32

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 32

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> 33

<211> 4107

<212> DNA

<213> Streptococcus pyogenes

<400> 33

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300

cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600

attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720

ctcattgctt tgtcattggg tttgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatact gaaataacta aggctcccct atcagcttca 960

atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560

tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100

agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta 2160

catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact 2220

gtaaaagttg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt 2280

attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt 2340

atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct 2400

gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460

gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac 2520

attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtctt aacgcgttct 2580

gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa 2640

aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700

acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgccaa 2760

ttggttgaaa ctcgccaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat 2820

actaaatacg atgaaaatga taaacttatt cgagaggtta aagtgattac cttaaaatct 2880

aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940

taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000

tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060

atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120

aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180

cctctaatcg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240

gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300

cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt 3360

gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420

tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt 3480

aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac 3540

tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600

tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660

caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttata tttagctagt 3720

cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780

cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840

attttagcag atgccaattt agataaagtt cttagtgcat ataacaaaca tagagacaaa 3900

ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct 3960

cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020

gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080

gatttgagtc agctaggagg tgactga 4107

<210> 34

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 34

Met Asp Lys Lys Tyr Ser Ile Gly Leu 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> 35

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 35

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> 36

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 36

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> 37

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 37

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> 38

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 38

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> 39

<211> 345

<212> PRT

<213> Sulfolobus islandicus

<400> 39

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> 40
<211> 1210
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
uncultured Parcubacteria group bacterium
<400> 40

Met Ser Lys Arg His Pro Arg Ile Ser Gly Val Lys Gly Tyr Arg Leu

1 5 10 15

His Ala Gln Arg Leu Glu Tyr Thr Gly Lys Ser Gly Ala Met Arg Thr

20 25 30

Ile Lys Tyr Pro Leu Tyr Ser Ser Pro Ser Gly Gly Arg Thr Val Pro

35 40 45

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

50 55 60

Leu Ser Asn Phe Asp Asp Leu Tyr Asn Ala Glu Lys Arg Asn Glu Glu

65 70 75 80

Lys Val Tyr Ser Val Leu Asp Phe Trp Tyr Asp Cys Val Gln Tyr Gly

85 90 95

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

100 105 110

Val Arg Gly Gly Ser Tyr Glu Leu Thr Lys Thr Leu Lys Gly Ser His

115 120 125

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

130 135 140

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

145 150 155 160

Ile Ile Asp Cys Phe Lys Ala Glu Tyr Arg Glu Arg His Lys Asp Gln

165 170 175

Cys Asn Lys Leu Ala Asp Asp Ile Lys Asn Ala Lys Lys Asp Ala Gly

180 185 190

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

195 200 205

Ile Ser Glu Gln Ser Glu Asn Asp Lys Pro Ser Phe Thr Asn Pro Leu

210 215 220

Asn Leu Thr Cys Cys Leu Leu Pro Phe Asp Thr Val Asn Asn Asn Arg

225 230 235 240

Asn Arg Gly Glu Val Leu Phe Asn Lys Leu Lys Glu Tyr Ala Gln Lys

245 250 255

Leu Asp Lys Asn Glu Gly Ser Leu Glu Met Trp Glu Tyr Ile Gly Ile

260 265 270

Gly Asn Ser Gly Thr Ala Phe Ser Asn Phe Leu Gly Glu Gly Phe Leu

275 280 285

Gly Arg Leu Arg Glu Asn Lys Ile Thr Glu Leu Lys Lys Ala Met Met

290 295 300

Asp Ile Thr Asp Ala Trp Arg Gly Gln Glu Gln Glu Glu Glu Leu Glu

305 310 315 320

Lys Arg Leu Arg Ile Leu Ala Ala Leu Thr Ile Lys Leu Arg Glu Pro

325 330 335

Lys Phe Asp Asn His Trp Gly Gly Tyr Arg Ser Asp Ile Asn Gly Lys

340 345 350

Leu Ser Ser Trp Leu Gln Asn Tyr Ile Asn Gln Thr Val Lys Ile Lys

355 360 365

Glu Asp Leu Lys Gly His Lys Lys Asp Leu Lys Lys Ala Lys Glu Met

370 375 380

Ile Asn Arg Phe Gly Glu Ser Asp Thr Lys Glu Glu Ala Val Val Ser

385 390 395 400

Ser Leu Leu Glu Ser Ile Glu Lys Ile Val Pro Asp Asp Ser Ala Asp

405 410 415

Asp Glu Lys Pro Asp Ile Pro Ala Ile Ala Ile Tyr Arg Arg Phe Leu

420 425 430

Ser Asp Gly Arg Leu Thr Leu Asn Arg Phe Val Gln Arg Glu Asp Val

435 440 445

Gln Glu Ala Leu Ile Lys Glu Arg Leu Glu Ala Glu Lys Lys Lys Lys

450 455 460

Pro Lys Lys Arg Lys Lys Lys Ser Asp Ala Glu Asp Glu Lys Glu Thr

465 470 475 480

Ile Asp Phe Lys Glu Leu Phe Pro His Leu Ala Lys Pro Leu Lys Leu

485 490 495

Val Pro Asn Phe Tyr Gly Asp Ser Lys Arg Glu Leu Tyr Lys Lys Tyr
500 505 510
Lys Asn Ala Ala Ile Tyr Thr Asp Ala Leu Trp Lys Ala Val Glu Lys

515 520 525
Ile Tyr Lys Ser Ala Phe Ser Ser Ser Leu Lys Asn Ser Phe Phe Asp
530 535 540
Thr Asp Phe Asp Lys Asp Phe Phe Ile Lys Arg Leu Gln Lys Ile Phe
545 550 555 560
Ser Val Tyr Arg Arg Phe Asn Thr Asp Lys Trp Lys Pro Ile Val Lys
565 570 575
Asn Ser Phe Ala Pro Tyr Cys Asp Ile Val Ser Leu Ala Glu Asn Glu

580 585 590
Val Leu Tyr Lys Pro Lys Gln Ser Arg Ser Arg Lys Ser Ala Ala Ile
595 600 605
Asp Lys Asn Arg Val Arg Leu Pro Ser Thr Glu Asn Ile Ala Lys Ala
610 615 620
Gly Ile Ala Leu Ala Arg Glu Leu Ser Val Ala Gly Phe Asp Trp Lys
625 630 635 640
Asp Leu Leu Lys Lys Glu Glu His Glu Glu Tyr Ile Asp Leu Ile Glu

645 650 655
Leu His Lys Thr Ala Leu Ala Leu Leu Leu Ala Val Thr Glu Thr Gln
660 665 670
Leu Asp Ile Ser Ala Leu Asp Phe Val Glu Asn Gly Thr Val Lys Asp
675 680 685
Phe Met Lys Thr Arg Asp Gly Asn Leu Val Leu Glu Gly Arg Phe Leu
690 695 700
Glu Met Phe Ser Gln Ser Ile Val Phe Ser Glu Leu Arg Gly Leu Ala

705 710 715 720
Gly Leu Met Ser Arg Lys Glu Phe Ile Thr Arg Ser Ala Ile Gln Thr
725 730 735
Met Asn Gly Lys Gln Ala Glu Leu Leu Tyr Ile Pro His Glu Phe Gln

740 745 750

Ser Ala Lys Ile Thr Thr Pro Lys Glu Met Ser Arg Ala Phe Leu Asp

755 760 765

Leu Ala Pro Ala Glu Phe Ala Thr Ser Leu Glu Pro Glu Ser Leu Ser

770 775 780

Glu Lys Ser Leu Leu Lys Leu Lys Gln Met Arg Tyr Tyr Pro His Tyr

785 790 795 800

Phe Gly Tyr Glu Leu Thr Arg Thr Gly Gln Gly Ile Asp Gly Gly Val

805 810 815

Ala Glu Asn Ala Leu Arg Leu Glu Lys Ser Pro Val Lys Lys Arg Glu

820 825 830

Ile Lys Cys Lys Gln Tyr Lys Thr Leu Gly Arg Gly Gln Asn Lys Ile

835 840 845

Val Leu Tyr Val Arg Ser Ser Tyr Tyr Gln Thr Gln Phe Leu Glu Trp

850 855 860

Phe Leu His Arg Pro Lys Asn Val Gln Thr Asp Val Ala Val Ser Gly

865 870 875 880

Ser Phe Leu Ile Asp Glu Lys Lys Val Lys Thr Arg Trp Asn Tyr Asp

885 890 895

Ala Leu Thr Val Ala Leu Glu Pro Val Ser Gly Ser Glu Arg Val Phe

900 905 910

Val Ser Gln Pro Phe Thr Ile Phe Pro Glu Lys Ser Ala Glu Glu Glu

915 920 925

Gly Gln Arg Tyr Leu Gly Ile Asp Ile Gly Glu Tyr Gly Ile Ala Tyr

930 935 940

Thr Ala Leu Glu Ile Thr Gly Asp Ser Ala Lys Ile Leu Asp Gln Asn

945 950 955 960

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

965 970 975

Leu Lys Leu Asp Gln Arg Arg Gly Thr Phe Ala Met Pro Ser Thr Lys

980 985 990

Ile Ala Arg Ile Arg Glu Ser Leu Val His Ser Leu Arg Asn Arg Ile
995 1000 1005

His His Leu Ala Leu Lys His Lys Ala Lys Ile Val Tyr Glu Leu
1010 1015 1020

Glu Val Ser Arg Phe Glu Glu Gly Lys Gln Lys Ile Lys Lys Val

1025 1030 1035

Tyr Ala Thr Leu Lys Lys Ala Asp Val Tyr Ser Glu Ile Asp Ala
1040 1045 1050

Asp Lys Asn Leu Gln Thr Thr Val Trp Gly Lys Leu Ala Val Ala
1055 1060 1065

Ser Glu Ile Ser Ala Ser Tyr Thr Ser Gln Phe Cys Gly Ala Cys
1070 1075 1080

Lys Lys Leu Trp Arg Ala Glu Met Gln Val Asp Glu Thr Ile Thr

1085 1090 1095

Thr Gln Glu Leu Ile Gly Thr Val Arg Val Ile Lys Gly Gly Thr
1100 1105 1110

Leu Ile Asp Ala Ile Lys Asp Phe Met Arg Pro Pro Ile Phe Asp
1115 1120 1125

Glu Asn Asp Thr Pro Phe Pro Lys Tyr Arg Asp Phe Cys Asp Lys
1130 1135 1140

His His Ile Ser Lys Lys Met Arg Gly Asn Ser Cys Leu Phe Ile

1145 1150 1155

Cys Pro Phe Cys Arg Ala Asn Ala Asp Ala Asp Ile Gln Ala Ser
1160 1165 1170

Gln Thr Ile Ala Leu Leu Arg Tyr Val Lys Glu Glu Lys Lys Val
1175 1180 1185

Glu Asp Tyr Phe Glu Arg Phe Arg Lys Leu Lys Asn Ile Lys Val
1190 1195 1200

Leu Gly Gln Met Lys Lys Ile
1205 1210

<210> 41

<211> 208

<212> PRT

<213> Petromyzon marinus

<400> 41

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr
1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
20 25 30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
35 40 45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
50 55 60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg
65 70 75 80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
85 90 95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
100 105 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
115 120 125

Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
130 135 140

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145 150 155 160

Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175

Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
180 185 190

Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205

<210> 42

<211> 766

<212> DNA

<213> Petromyzon marinus

<400> 42

tgacacgaca cagccgtgta tatgaggaag ggtagctgga tgggggggggg gggaatacgt 60

tcagagagga cattagcgag cgtcttgttg gtggccttga gtctagacac ctgcagacat 120

gaccgacgct gagtacgtga gaatccatga gaagttggac atctacacgt ttaagaaaca 180

gtttttcaac aacaaaaaat ccgtgtcgca tagatgctac gttctctttg aattaaaacg 240

acggggtgaa cgtagagcgt gtttttgggg ctatgctgtg aataaaccac agagcgggac 300

agaacgtgga attcacgccg aaatctttag cattagaaaa gtcgaagaat acctgcgcga 360

caaccccgga caattcacga taaattggta ctcatcctgg agtccttgtg cagattgcgc 420

tgaaaagatc ttagaatggt ataaccagga gctgcggggg aacggccaca ctttgaaaat 480

ctgggcttgc aaactctatt acgagaaaaa tgcgaggaat caaattgggc tgtggaacct 540

cagagataac ggggttgggt tgaatgtaat ggtaagtgaa cactaccaat gttgcaggaa 600

aatattcatc caatcgtcgc acaatcaatt gaatgagaat agatggcttg agaagacttt 660

gaagcgagct gaaaaacgac ggagcgagtt gtccattatg attcaggtaa aaatactcca 720

caccactaag agtcctgctg tttaagaggc tatgcggatg gttttc 766

<210> 43

<211> 145

<212> PRT

<213> Homo sapiens

<400> 43

Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45

Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Ala Pro

130 135 140

Val

145

<210> 44

<211> 10681

<212> DNA

<213> Homo sapiens

<400> 44

agagaaccat cattaattga agtgagattt ttctggcctg agacttgcag ggaggcaaga 60

agacactctg gacaccacta tggacaggta aagaggcagt cttctcgtgg gtgattgcac 120

tggccttcct ctcagagcaa atctgagtaa tgagactggt agctatccct ttctctcatg 180

taactgtctg actgataaga tcagcttgat caatatgcat atatattttt tgatctgtct 240

ccttttcttc tattcagatc ttatacgctg tcagcccaat tctttctgtt tcagacttct 300

cttgatttcc ctctttttca tgtggcaaaa gaagtagtgc gtacaatgta ctgattcgtc 360

ctgagatttg taccatggtt gaaactaatt tatggtaata atattaacat agcaaatctt 420

tagagactca aatcatgaaa aggtaatagc agtactgtac taaaaacggt agtgctaatt 480

ttcgtaataa ttttgtaaat attcaacagt aaaacaactt gaagacacac tttcctaggg 540

aggcgttact gaaataattt agctatagta agaaaatttg taattttaga aatgccaagc 600

attctaaatt aattgcttga aagtcactat gattgtgtcc attataagga gacaaattca 660

ttcaagcaag ttatttaatg ttaaaggccc aattgttagg cagttaatgg cacttttact 720

attaactaat ctttccattt gttcagacgt agcttaactt acctcttagg tgtgaatttg 780

gttaaggtcc tcataatgtc tttatgtgca gtttttgata ggttattgtc atagaactta 840

ttctattcct acatttatga ttactatgga tgtatgagaa taacacctaa tccttatact 900

ttacctcaat ttaactcctt tataaagaac ttacattaca gaataaagat tttttaaaaa 960

tatatttttt tgtagagaca gggtcttagc ccagccgagg ctggtctcta agtcctggcc 1020

caagcgatcc tcctgcctgg gcctcctaaa gtgctggaat tatagacatg agccatcaca 1080

tccaatatac agaataaaga tttttaatgg aggatttaat gttcttcaga aaattttctt 1140

gaggtcagac aatgtcaaat gtctcctcag tttacactga gattttgaaa acaagtctga 1200

gctataggtc cttgtgaagg gtccattgga aatacttgtt caaagtaaaa tggaaagcaa 1260

aggtaaaatc agcagttgaa attcagagaa agacagaaaa ggagaaaaga tgaaattcaa 1320

caggacagaa gggaaatata ttatcattaa ggaggacagt atctgtagag ctcattagtg 1380

atggcaaaat gacttggtca ggattatttt taacccgctt gtttctggtt tgcacggctg 1440

gggatgcagc tagggttctg cctcagggag cacagctgtc cagagcagct gtcagcctgc 1500

aagcctgaaa cactccctcg gtaaagtcct tcctactcag gacagaaatg acgagaacag 1560

ggagctggaa acaggcccct aaccagagaa gggaagtaat ggatcaacaa agttaactag 1620

caggtcagga tcacgcaatt catttcactc tgactggtaa catgtgacag aaacagtgta 1680

ggcttattgt attttcatgt agagtaggac ccaaaaatcc acccaaagtc ctttatctat 1740

gccacatcct tcttatctat acttccagga cacttttttct tccttatgat aaggctctct 1800

ctctctccac acacacacac acacacacac acacacacac acacacacac acaaacacac 1860

accccgccaa ccaaggtgca tgtaaaaaga tgtagattcc tctgcctttc tcatctacac 1920

agcccaggag ggtaagttaa tataagaggg atttattggt aagagatgat gcttaatctg 1980

tttaacactg ggcctcaaag agagaatttc ttttcttctg tacttattaa gcacctatta 2040

tgtgttgagc ttatatatac aaagggttat tatatgctaa tatagtaata gtaatggtgg 2100

ttggtactat ggtaattacc ataaaaatta ttatcctttt aaaataaagc taattattat 2160

tggatctttt ttagtattca ttttatgttt tttatgtttt tgatttttta aaagacaatc 2220

tcaccctgtt acccaggctg gagtgcagtg gtgcaatcat agctttctgc agtcttgaac 2280

tcctgggctc aagcaatcct cctgccttgg cctcccaaag tgttgggata cagtcatgag 2340

ccactgcatc tggcctagga tccatttaga ttaaaatatg cattttaaat tttaaaataa 2400

tatggctaat ttttacctta tgtaatgtgt atactggcaa taaatctagt ttgctgccta 2460

aagtttaaag tgctttccag taagcttcat gtacgtgagg ggagacattt aaagtgaaac 2520

agacagccag gtgtggtggc tcacgcctgt aatcccagca ctctgggagg ctgaggtggg 2580

tggatcgctt gagccctgga gttcaagacc agcctgagca acatggcaaa acgctgtttc 2640

tataacaaaa attagccggg catggtggca tgtgcctgtg gtcccagcta ctaggggct 2700

gaggcaggag aatcgttgga gcccaggagg tcaaggctgc actgagcagt gcttgcgcca 2760

ctgcactcca gcctgggtga caggaccaga ccttgcctca aaaaaataag aagaaaaatt 2820

aaaaataaat ggaaacaact acaaagagct gttgtcctag atgagctact tagttaggct 2880

gatattttgg tatttaactt ttaaagtcag ggtctgtcac ctgcactaca ttattaaaat 2940

atcaattctc aatgtatatc cacacaaaga ctggtacgtg aatgttcata gtacctttat 3000

tcacaaaacc ccaaagtaga gactatccaa atatccatca acaagtgaac aaataaacaa 3060

aatgtgctat atccatgcaa tggaatacca ccctgcagta caaagaagct acttggggat 3120

gaatcccaaa gtcatgacgc taaatgaaag agtcagacat gaaggaggag ataatgtatg 3180

ccatacgaaa ttctagaaaa tgaaagtaac ttatagttac agaaagcaaa tcagggcagg 3240

catagaggct cacacctgta atcccagcac tttgagaggc cacgtgggaa gattgctaga 3300

actcaggagt tcaagaccag cctgggcaac acagtgaaac tccattctcc acaaaaatgg 3360

gaaaaaaaga aagcaaatca gtggttgtcc tgtggggagg ggaaggactg caaagaggga 3420

agaagctctg gtggggtgag ggtggtgatt caggttctgt atcctgactg tggtagcagt 3480

ttggggtgtt tacatccaaa aatattcgta gaattatgca tcttaaatgg gtggagttta 3540

ctgtatgtaa attatacctc aatgtaagaa aaaataatgt gtaagaaaac tttcaattct 3600

cttgccagca aacgttattc aaattcctga gccctttact tcgcaaattc tctgcacttc 3660

tgccccgtac cattaggtga cagcactagc tccacaaatt ggataaatgc atttctggaa 3720

aagactaggg acaaaatcca ggcatcactt gtgctttcat atcaaccatg ctgtacagct 3780

tgtgttgctg tctgcagctg caatggggac tcttgatttc tttaaggaaa cttgggttac 3840

cagagtattt ccacaaatgc tattcaaatt agtgcttatg atatgcaaga cactgtgcta 3900

ggagccagaa aacaaagagg aggagaaatc agtcattatg tgggaacaac atagcaagat 3960

atttagatca ttttgactag ttaaaaaagc agcagagtac aaaatcacac atgcaatcag 4020

tataatccaa atcatgtaaa tatgtgcctg tagaaagact agaggaataa acacaagaat 4080

cttaacagtc attgtcatta gacactaagt ctaattatta ttattagaca ctatgatatt 4140

tgagatttaa aaaatcttta atattttaaa atttagagct cttctatttt tccatagtat 4200

tcaagtttga caatgatcaa gtattactct ttctttttt ttttttttt tttttttga 4260

gatggagttt tggtcttgtt gcccatgctg gagtggaatg gcatgaccat agctcactgc 4320

aacctccacc tcctgggttc aagcaaagct gtcgcctcag cctcccgggt agatgggatt 4380

acaggcgccc accaccacac tcggctaatg tttgtatttt tagtagagat ggggtttcac 4440

catgttggcc aggctggtct caaactcctg acctcagagg atccacctgc ctcagcctcc 4500

caaagtgctg ggattacaga tgtaggccac tgcgcccggc caagtattgc tcttatacat 4560

taaaaaacag gtgtgagcca ctgcgcccag ccaggtattg ctcttataca ttaaaaaata 4620

ggccggtgca gtggctcacg cctgtaatcc cagcactttg ggaagccaag gcgggcagaa 4680

cacccgaggt caggagtcca aggccagcct ggccaagatg gtgaaacccc gtctctatta 4740

aaaatacaaa cattacctgg gcatgatggt gggcgcctgt aatcccagct actcaggagg 4800
ctgaggcagg aggatccgcg gagcctggca gatctgcctg agcctgggag gttgaggcta 4860
cagtaagcca agatcatgcc agtatacttc agcctgggcg acaaagtgag accgtaacaa 4920
aaaaaaaaaa atttaaaaaa agaaatttag atcaagatcc aactgtaaaa agtggcctaa 4980
acaccacatt aaagagtttg gagtttattc tgcaggcaga agagaaccat caggggggtct 5040
tcagcatggg aatggcatgg tgcacctggt ttttgtgaga tcatggtggt gacagtgtgg 5100
ggaatgttat tttggaggga ctggaggcag acagaccggt taaaaggcca gcacaacaga 5160

taaggaggaa gaagatgagg gcttggaccg aagcagagaa gagcaaacag ggaaggtaca 5220
aattcaagaa atattggggg gtttgaatca acacatttag atgattaatt aaatatgagg 5280
actgaggaat aagaaatgag tcaaggatgg ttccaggctg ctaggctgct tacctgaggt 5340
ggcaaagtcg ggaggagtgg cagtttagga caggggggcag ttgaggaata ttgttttgat 5400
cattttgagt ttgaggtaca agttggacac ttaggtaaag actggagggg aaatctgaat 5460
atacaattat gggactgagg aacaagttta ttttatttttt tgtttcgttt tcttgttgaa 5520
gaacaaattt aattgtaatc ccaagtcatc agcatctaga agacagtggc aggaggtgac 5580

tgtcttgtgg gtaagggttt ggggtccttg atgagtatct ctcaattggc cttaaatata 5640
agcaggaaaa ggagtttatg atggattcca ggctcagcag ggctcaggag ggctcaggca 5700
gccagcagag gaagtcagag catcttcttt ggtttagccc aagtaatgac ttccttaaaa 5760
agctgaagga aaatccagag tgaccagatt ataaactgta ctcttgcatt ttctctccct 5820
cctctcaccc acagcctctt gatgaaccgg aggaagtttc tttaccaatt caaaaatgtc 5880
cgctgggcta agggtcggcg tgagacctac ctgtgctacg tagtgaagag gcgtgacagt 5940
gctacatcct tttcactgga ctttggttat cttcgcaata aggtatcaat taaagtcggc 6000

tttgcaagca gtttaatggt caactgtgag tgcttttaga gccacctgct gatggtatta 6060
cttccatcct tttttggcat ttgtgtctct atcacattcc tcaaatcctt ttttttattt 6120
ctttttccat gtccatgcac ccatattaga catggcccaa aatatgtgat ttaattcctc 6180
cccagtaatg ctgggcaccc taataccact ccttccttca gtgccaagaa caactgctcc 6240
caaactgttt accagctttc ctcagcatct gaattgcctt tgagattaat taagctaaaa 6300
gcatttttat atgggagaat attatcagct tgtccaagca aaaattttaa atgtgaaaaa 6360
caaattgtgt cttaagcatt tttgaaaatt aaggaagaag aatttgggaa aaaattaacg 6420

gtggctcaat tctgtcttcc aaatgatttc ttttccctcc tactcacatg ggtcgtaggc 6480

cagtgaatac attcaacatg gtgatcccca gaaaactcag agaagcctcg gctgatgatt 6540

aattaaattg atctttcggc tacccgagag aattacattt ccaagagact tcttcaccaa 6600

aatccagatg ggtttacata aacttctgcc cacgggtatc tcctctctcc taacacgctg 6660

tgacgtctgg gcttggtgga atctcaggga agcatccgtg gggtggaagg tcatcgtctg 6720

gctcgttgtt tgatggttat attaccatgc aattttcttt gcctacattt gtattgaata 6780

catcccaatc tccttcctat tcggtgacat gacacattct atttcagaag gctttgattt 6840

tatcaagcac tttcatttac ttctcatggc agtgcctatt acttctctta caatacccat 6900

ctgtctgctt taccaaaatc tatttcccct tttcagatcc tcccaaatgg tcctcataaa 6960

ctgtcctgcc tccacctagt ggtccaggta tatttccaca atgttacatc aacaggcact 7020

tctagccatt ttccttctca aaaggtgcaa aaagcaactt cataaacaca aattaaatct 7080

tcggtgaggt agtgtgatgc tgcttcctcc caactcagcg cacttcgtct tcctcattcc 7140

acaaaaaccc atagccttcc ttcactctgc aggactagtg ctgccaaggg ttcagctcta 7200

cctactggtg tgctcttttg agcaagttgc ttagcctctc tgtaacacaa ggacaatagc 7260

tgcaagcatc cccaaagatc attgcaggag acaatgacta aggctaccag agccgcaata 7320

aaagtcagtg aattttagcg tggtcctctc tgtctctcca gaacggctgc cacgtggaat 7380

tgctcttcct ccgctacatc tcggactggg acctagaccc tggccgctgc taccgcgtca 7440

cctggttcac ctcctggagc ccctgctacg actgtgcccg acatgtggcc gactttctgc 7500

gagggaaccc caacctcagt ctgaggatct tcaccgcgcg cctctacttc tgtgaggacc 7560

gcaaggctga gcccgagggg ctgcggcggc tgcaccgcgc cggggtgcaa atagccatca 7620

tgaccttcaa aggtgcgaaa gggccttccg cgcaggcgca gtgcagcagc ccgcattcgg 7680

gattgcgatg cggaatgaat gagttagtgg ggaagctcga ggggaagaag tgggcgggga 7740

ttctggttca cctctggagc cgaaattaaa gattagaagc agagaaaaga gtgaatggct 7800

cagagacaag gccccgagga aatgagaaaa tggggccagg gttgcttctt tcccctcgat 7860

ttggaacctg aactgtcttc tacccccata tccccgcctt tttttccttt ttttttttt 7920

gaagattatt tttactgctg gaatactttt gtagaaaacc acgaaagaac tttcaaagcc 7980

tgggaagggc tgcatgaaaa ttcagttcgt ctctccagac agcttcggcg catccttttg 8040

gtaaggggct tcctcgcttt ttaaattttc tttctttctc tacagtcttt tttggagttt 8100

cgtatatttc ttatattttc ttattgttca atcactctca gttttcatct gatgaaaact 8160

ttatttctcc tccacatcag ctttttcttc tgctgtttca ccattcagag ccctctgcta 8220

aggttccttt tccctccctt ttctttcttt tgttgtttca catctttaaa tttctgtctc 8280

tccccagggt tgcgtttcct tcctggtcag aattcttttc tccttttttt ttttttttt 8340

ttttttttt aaacaaacaa acaaaaaacc caaaaaaact ctttcccaat ttactttctt 8400

ccaacatgtt acaaagccat ccactcagtt tagaagactc tccggcccca ccgaccccca 8460

acctcgtttt gaagccattc actcaatttg cttctctctt tctctacagc ccctgtatga 8520

ggttgatgac ttacgagacg catttcgtac tttgggactt tgatagcaac ttccaggaat 8580

gtcacacacg atgaaatatc tctgctgaag acagtggata aaaaacagtc cttcaagtct 8640

tctctgtttt tattcttcaa ctctcacttt cttagagttt acagaaaaaa tatttatata 8700

cgactcttta aaaagatcta tgtcttgaaa atagagaagg aacacaggtc tggccaggga 8760

cgtgctgcaa ttggtgcagt tttgaatgca acattgtccc ctactgggaa taacagaact 8820

gcaggacctg ggagcatcct aaagtgtcaa cgtttttcta tgacttttag gtaggatgag 8880

agcagaaggt agatcctaaa aagcatggtg agaggatcaa atgtttttat atcaacatcc 8940

tttattattt gattcatttg agttaacagt ggtgttagtg atagattttt ctattctttt 9000

cccttgacgt ttactttcaa gtaacacaaa ctcttccatc aggccatgat ctataggacc 9060

tcctaatgag agtatctggg tgattgtgac cccaaaccat ctctccaaag cattaatatc 9120

caatcatgcg ctgtatgttt taatcagcag aagcatgttt ttatgtttgt acaaaagaag 9180

attgttatgg gtggggatgg aggtatagac catgcatggt caccttcaag ctactttaat 9240

aaaggatctt aaaatgggca ggaggactgt gaacaagaca ccctaataat gggttgatgt 9300

ctgaagtagc aaatcttctg gaaacgcaaa ctcttttaag gaagtcccta atttagaaac 9360

acccacaaac ttcacatatc ataattagca aacaattgga aggaagttgc ttgaatgttg 9420

gggagaggaa aatctattgg ctctcgtggg tctcttcatc tcagaaatgc caatcaggtc 9480

aaggtttgct acattttgta tgtgtgtgat gcttctccca aaggtatatt aactatataa 9540

gagagttgtg acaaaacaga atgataaagc tgcgaaccgt ggcacacgct catagttcta 9600

gctgcttggg aggttgagga gggaggatgg cttgaacaca ggtgttcaag gccagcctgg 9660

gcaacataac aagatcctgt ctctcaaaaa aaaaaaaaaa aaaaagaaag agagagggcc 9720

gggcgtggtg gctcacgcct gtaatcccag cactttggga ggccgagccg ggcggatcac 9780

ctgtggtcag gagtttgaga ccagcctggc caacatggca aaaccccgtc tgtactcaaa 9840

atgcaaaaat tagccaggcg tggtagcagg cacctgtaat cccagctact tgggaggctg 9900

aggcaggaga atcgcttgaa cccaggaggt ggaggttgca gtaagctgag atcgtgccgt 9960

tgcactccag cctgggcgac aagagcaaga ctctgtctca gaaaaaaaaa aaaaaaagag 10020

agagagagag aaagagaaca atatttggga gagaaggatg gggaagcatt gcaaggaaat 10080

tgtgctttat ccaacaaaat gtaaggagcc aataagggat ccctatttgt ctcttttggt 10140

gtctatttgt ccctaacaac tgtctttgac agtgagaaaa atattcagaa taaccatatc 10200

cctgtgccgt tattacctag caacccttgc aatgaagatg agcagatcca caggaaaact 10260
tgaatgcaca actgtcttat tttaatctta ttgtacataa gtttgtaaaa gagttaaaaa 10320
ttgttacttc atgtattcat ttatatttta tattattttg cgtctaatga ttttttatta 10380
acatgatttc cttttctgat atattgaaat ggagtctcaa agcttcataa atttataact 10440
ttagaaatga ttctaataac aacgtatgta attgtaacat tgcagtaatg gtgctacgaa 10500
gccatttctc ttgattttta gtaaactttt atgacagcaa atttgcttct ggctcacttt 10560
caatcagtta aataaatgat aaataatttt ggaagctgtg aagataaaat accaaataaa 10620

ataatataaa agtgatttat atgaagttaa aataaaaaat cagtatgatg gaataaactt 10680
g 10681

<210> 45
<211> 198
<212> PRT
<213> Homo sapiens
<400> 45

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> 46
<211> 198
<212> PRT
<213> Mus musculus
<400> 46
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> 47

<211> 198

<212> PRT

<213> Canis lupus

<400> 47

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

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly His

35 40 45

Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Ala Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Lys Thr Phe Lys
145 150 155 160
Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
165 170 175
Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190
Phe Arg Thr Leu Gly Leu
195

<210> 48
<211> 199
<212> PRT
<213> Bos taurus
<400> 48
Met Asp Ser Leu Leu Lys Lys Gln Arg Gln Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Pro Thr Ser Phe Ser Leu Asp Phe Gly His
35 40 45
Leu Arg Asn Lys Ala Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

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

115 120 125
Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp
130 135 140

Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe
145 150 155 160
Lys Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln
165 170 175
Leu Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp

180 185 190
Ala Phe Arg Thr Leu Gly Leu
195

<210> 49
<211> 239
<212> PRT
<213> Rattus norvegicus
<400> 49
Met Ala Val Gly Ser Lys Pro Lys Ala Ala Leu Val Gly Pro His Trp
1 5 10 15
Glu Arg Glu Arg Ile Trp Cys Phe Leu Cys Ser Thr Gly Leu Gly Thr
20 25 30
Gln Gln Thr Gly Gln Thr Ser Arg Trp Leu Arg Pro Ala Ala Thr Gln

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

100 105 110
Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys
115 120 125
Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala
130 135 140
Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg

145 150 155 160

Ile Phe Thr Ala Arg Leu Thr Gly Trp Gly Ala Leu Pro Ala Gly Leu

165 170 175

Met Ser Pro Ala Arg Pro Ser Asp Tyr Phe Tyr Cys Trp Asn Thr Phe

180 185 190

Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu

195 200 205

Asn Ser Val Arg Leu Ser Arg Arg Leu Arg Arg Ile Leu Leu Pro Leu

210 215 220

Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu Gly Leu

225 230 235

<210> 50

<211> 429

<212> PRT

<213> Mus musculus

<400> 50

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> 51

<211> 430
<212> PRT
<213> Rattus norvegicus
<400> 51
Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro
1 5 10 15
Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn
20 25 30
Arg Leu Arg Tyr Ala Ile Asp Arg Lys Asp Thr Phe Leu Cys Tyr Glu
35 40 45
Val Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val

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

115 120 125
Ser Ser Arg Leu Tyr Asn Ile Arg Asp Pro Glu Asn Gln Gln Asn Leu
130 135 140
Cys Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr

145 150 155 160

Glu Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg

165 170 175

Phe Arg Pro Trp Lys Lys Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser

180 185 190

Lys Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser

195 200 205

Ser Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu

210 215 220

Thr Arg Phe Cys Val Glu Arg Arg Arg Val His Leu Leu Ser Glu Glu

225 230 235 240

Glu Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr

245 250 255

Tyr His Gly Val Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn

260 265 270

Gly Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln

275 280 285

His Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser

290 295 300

Gln Val Ile Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys

305 310 315 320

Ala Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu

325 330 335

His Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln

340 345 350

Lys Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met

355 360 365

Asp Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys

370 375 380

Arg Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr

385 390 395 400

Gln Arg Arg Leu His Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu
405 410 415
Val Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser
420 425 430

<210> 52
<211> 370
<212> PRT
<213> Macaca mulatta
<400> 52

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> 53
<211> 384
<212> PRT
<213> Pan sp.
<400> 53

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> 54

<211> 377

<212> PRT

<213> Chlorocebus sabaeus

<400> 54

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> 55

<211> 384

<212> PRT

<213

> Homo sapiens

<400> 55

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> 56
<211> 373
<212> PRT

<213> Homo sapiens

<400> 56

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

> 57

<211> 382

<212> PRT

<213> Homo sapiens

<400> 57

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr

1 5 10 15

Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr

20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu

35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Val Tyr Phe Lys Pro Gln

50 55 60

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

115 120 125
Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg
130 135 140
Val Thr Ile Met Asp Tyr Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe
145 150 155 160
Val Tyr Asn Glu Gly Gln Gln Phe Met Pro Trp Tyr Lys Phe Asp Glu
165 170 175
Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Tyr Leu

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

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

305 310 315 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met

325 330 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln

340 345 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala

355 360 365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn

370 375 380

<210> 58

<211> 395

<212> PRT

<213> Rattus norvegicus

<400> 58

Met Gln Pro Gln Gly Leu Gly Pro Asn Ala Gly Met Gly Pro Val Cys

1 5 10 15

Leu Gly Cys Ser His Arg Arg Pro Tyr Ser Pro Ile Arg Asn Pro Leu

20 25 30

Lys Lys Leu Tyr Gln Gln Thr Phe Tyr Phe His Phe Lys Asn Val Arg

35 40 45

Tyr Ala Trp Gly Arg Lys Asn Asn Phe Leu Cys Tyr Glu Val Asn Gly

50 55 60

Met Asp Cys Ala Leu Pro Val Pro Leu Arg Gln Gly Val Phe Arg Lys

65 70 75 80

Gln Gly His Ile His Ala Glu Leu Cys Phe Ile Tyr Trp Phe His Asp

85 90 95

Lys Val Leu Arg Val Leu Ser Pro Met Glu Glu Phe Lys Val Thr Trp

100 105 110

Tyr Met Ser Trp Ser Pro Cys Ser Lys Cys Ala Glu Gln Val Ala Arg

115 120 125

Phe Leu Ala Ala His Arg Asn Leu Ser Leu Ala Ile Phe Ser Ser Arg

130 135 140

Leu Tyr Tyr Tyr Leu Arg Asn Pro Asn Tyr Gln Gln Lys Leu Cys Arg
145 150 155 160
Leu Ile Gln Glu Gly Val His Val Ala Ala Met Asp Leu Pro Glu Phe
165 170 175

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

Tyr Arg Arg Lys Ser Tyr Leu Cys Tyr Gln Leu Glu Arg Ala Asn Gly
245 250 255
Gln Glu Pro Leu Lys Gly Tyr Leu Leu Tyr Lys Lys Gly Glu Gln His
260 265 270
Val Glu Ile Leu Phe Leu Glu Lys Met Arg Ser Met Glu Leu Ser Gln
275 280 285
Val Arg Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
290 295 300

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

Phe Arg Pro Trp Asn Glu Leu Glu Lys Asn Ser Trp Arg Ile Gln Arg
370 375 380
Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu

385 390 395

<210> 59

<211> 226

<212> PRT

<213> Bos taurus

<400> 59

Asp Gly Trp Glu Val Ala Phe Arg Ser Gly Thr Val Leu Lys Ala Gly
1 5 10 15
Val Leu Gly Val Ser Met Thr Glu Gly Trp Ala Gly Ser Gly His Pro
20 25 30

Gly Gln Gly Ala Cys Val Trp Thr Pro Gly Thr Arg Asn Thr Met Asn
35 40 45
Leu Leu Arg Glu Val Leu Phe Lys Gln Gln Phe Gly Asn Gln Pro Arg
50 55 60
Val Pro Ala Pro Tyr Tyr Arg Arg Lys Thr Tyr Leu Cys Tyr Gln Leu
65 70 75 80
Lys Gln Arg Asn Asp Leu Thr Leu Asp Arg Gly Cys Phe Arg Asn Lys
85 90 95

Lys Gln Arg His Ala Glu Arg Phe Ile Asp Lys Ile Asn Ser Leu Asp
100 105 110
Leu Asn Pro Ser Gln Ser Tyr Lys Ile Ile Cys Tyr Ile Thr Trp Ser
115 120 125
Pro Cys Pro Asn Cys Ala Asn Glu Leu Val Asn Phe Ile Thr Arg Asn
130 135 140
Asn His Leu Lys Leu Glu Ile Phe Ala Ser Arg Leu Tyr Phe His Trp
145 150 155 160

Ile Lys Ser Phe Lys Met Gly Leu Gln Asp Leu Gln Asn Ala Gly Ile
165 170 175
Ser Val Ala Val Met Thr His Thr Glu Phe Glu Asp Cys Trp Glu Gln
180 185 190
Phe Val Asp Asn Gln Ser Arg Pro Phe Gln Pro Trp Asp Lys Leu Glu
195 200 205

Gln Tyr Ser Ala Ser Ile Arg Arg Arg Leu Gln Arg Ile Leu Thr Ala

210 215 220

Pro Ile

225

<210> 60

<211> 490

<212> PRT

<213> Pan sp.

<400> 60

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> 61

<211> 190

<212> PRT

<213> Homo sapiens

<400> 61

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> 62

<211> 190

<212> PRT

<213> Gorilla sp.

<400> 62

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr
1 5 10 15

Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu
20 25 30

Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val
35 40 45

Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His
50 55 60

Cys His Ala Glu Arg Cys Phe Leu Ser Trp Glu Cys Asp Asp Ile Leu
65 70 75 80

Ser Pro Asn Thr Asn Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
85 90 95

Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser
100 105 110

Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Asp
115 120 125

Thr Asp Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala
130 135 140

Val Lys Ile Met Asp Tyr Lys Asp Phe Lys Tyr Cys Trp Glu Asn Phe
145 150 155 160

Val Tyr Asn Asp Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Tyr
165 170 175
Asn Phe Arg Phe Leu Lys Arg Arg Leu Gln Glu Ile Leu Glu
180 185 190

<210> 63
<211> 199
<212> PRT
<213> Homo sapiens
<400> 63
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> 64

<211> 202

<212> PRT

<213> Macaca mulatta

<400> 64

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> 65

<211> 185

<212> PRT

<213> Bos taurus

<400> 65

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> 66

<211> 200

<212> PRT

<213> Homo sapiens

<400> 66

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> 67

<211> 210

<212> PRT

<213> Macaca mulatta

<400> 67

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> 68

<211> 386

<212> PRT

<213> Homo sapiens

<400> 68

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> 69

<211> 236

<212> PRT

<213> Homo sapiens

<400> 69

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> 70

<211> 229

<212> PRT

<213> Mus musculus

<400> 70

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> 71

<211> 229

<212> PRT

<213> Rattus norvegicus

<400> 71

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> 72

<211> 224

<212> PRT

<213> Homo sapiens

<400> 72

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> 73

<211> 224

<212> PRT

<213> Mus musculus

<400> 73

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> 74
<211> 224
<212> PRT
<213> Rattus norvegicus
<400> 74

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> 75

<211> 224

<212> PRT

<213> Bos taurus

<400> 75

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Ala Glu Pro Ala Ser Gln

1 5 10 15

Asn Gly Glu Glu Val Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu

20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro

35 40 45

Ala His Tyr Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly

50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly

65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Thr

85 90 95

Asn His Ala Glu Glu Ala Phe Phe Asn Ser Ile Met Pro Thr Phe Asp

100 105 110

Pro Ala Leu Arg Tyr Met Val Thr Trp Tyr Val Ser Ser Ser Pro Cys

115 120 125

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

130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160

Glu Ile Gln Ala Ala Leu Arg Lys Leu Lys Glu Ala Gly Cys Arg Leu
165 170 175
Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
180 185 190
Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205
Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 76
<211> 208
<212> PRT
<213> Petromyzon marinus
<400> 76
Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr
1 5 10 15
Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
20 25 30
Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
35 40 45
Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly

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

115 120 125

Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
130 135 140
Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145 150 155 160
Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175
Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser

180 185 190
Phe Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205
<210> 77
<211> 381
<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 Phe Asn Tyr Asp Glu Phe Gln His Cys Trp
145 150 155 160

Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn Asn
165 170 175

Leu Pro Lys Tyr Tyr Ile Leu Leu His Phe Met Leu Gly Glu Ile Leu
180 185 190

Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu
195 200 205

Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu
210 215 220

Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe
225 230 235 240

Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His
245 250 255

Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu
260 265 270

Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe
275 280 285

Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Lys His Val Ser
290 295 300

Leu Cys Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln
305 310 315 320

Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Thr
325 330 335

Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly
340 345 350

Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu
355 360 365

Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
370 375 380

<210> 78

<211> 182

<212> PRT

<213> Homo sapiens

<400> 78

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Trp
1 5 10 15

Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His Asn
20 25 30

Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn Gln
35 40 45

Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu Cys
50 55 60

Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp Tyr
65 70 75 80

Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln
85 90 95

Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys Ile
100 105 110

Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly Leu
115 120 125

Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Thr Tyr Ser Glu
130 135 140

Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe
145 150 155 160

Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser Gly Arg
165 170 175

Leu Arg Ala Ile Leu Gln
180

<210> 79

<211> 184

<212> PRT

<213> Homo sapiens

<400> 79

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val
1 5 10 15

Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His
20 25 30

Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn
35 40 45

Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu
50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp
65 70 75 80

Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala
85 90 95

Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys
100 105 110

Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln Glu Gly
115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Met Thr Tyr
130 135 140

Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys
145 150 155 160

Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser
165 170 175

Gly Arg Leu Arg Ala Ile Leu Gln
180

<210> 80

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 80

Ser Gly Gly Ser

1

<210> 81

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 81

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> 82

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 82

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> 83

<211> 104

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 83

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> 84

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 84

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser

20

<210> 85

<211> 40

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 85

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser

20 25 30

Ser Gly Gly Ser Ser Gly Gly Ser

35 40

<210> 86

<211> 64

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 86

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser

20 25 30

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

35 40 45

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser

50 55 60

<210> 87

<211> 92

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 87

Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser

1 5 10 15

Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser

20 25 30

Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu

35 40 45

Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser
50 55 60

Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr
65 70 75 80
Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser
85 90

<210> 88
<211> 18
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic peptide
<400> 88
Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg
1 5 10 15

Lys Val

<210> 89
<211> 16
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic peptide
<400> 89
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15
<210> 90
<211> 17
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic peptide
<400> 90

Lys Lys Thr Glu Leu Gln Thr Thr Asn Ala Glu Asn Lys Thr Lys Lys

1 5 10 15

Leu

<210> 91

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 91

Lys Arg Gly Ile Asn Asp Arg Asn Phe Trp Arg Gly Glu Asn Gly Arg

1 5 10 15

Lys Thr Arg

<210> 92

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 92

Arg Lys Ser Gly Lys Ile Ala Ala Ile Val Val Lys Arg Pro Arg Lys

1 5 10 15

<210> 93

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 93

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 94

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 94

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> 95

<211> 20

<212> DNA

<213> Homo sapiens

<400> 95

gaacacaaag catagactgc 20

<210> 96

<211> 20

<212> DNA

<213> Homo sapiens

<400> 96

ggacagcttt tcctagacag 20

<210> 97

<211> 166

<212> PRT

<213> Unknown

<220><223> Description of Unknown:
TadA sequence

<400> 97

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

1 5 10 15

Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val

20 25 30

Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile

35 40 45

Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr

65 70 75 80

Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser

85 90 95

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

100 105 110

Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg

115 120 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu

130 135 140

Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys

145 150 155 160

Ala Gln Ser Ser Thr Asp

165

<210> 98

<211> 166

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 98

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

1 5 10 15

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val

20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile

35 40 45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65 70 75 80
Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
85 90 95
Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala
100 105 110
Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg
115 120 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
130 135 140
Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys
145 150 155 160
Ala Gln Ser Ser Thr Asp
165

<210> 99
<211> 1052
<212> PRT
<213> Staphylococcus aureus
<400> 99
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> 100

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 100

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> 101

<211> 1052

<212> PRT

<213> Staphylococcus aureus

<400> 101

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

1 5 10 15

Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val

20 25 30

Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser

35 40 45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln

50 55 60

Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser

65 70 75 80

Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser

85 90 95

Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala

100 105 110

Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly

115 120 125

Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu
130 135 140
Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp
145 150 155 160
Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr Val
165 170 175
Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu
180 185 190

Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg
195 200 205
Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp
210 215 220
Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
225 230 235 240
Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn
245 250 255

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
260 265 270
Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
275 280 285
Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
290 295 300
Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro
305 310 315 320

Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
325 330 335
Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys
340 345 350
Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
355 360 365

Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn
370 375 380

Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn
385 390 395 400
Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile
405 410 415
Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln
420 425 430
Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val
435 440 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile
450 455 460
Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu
465 470 475 480
Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg
485 490 495
Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly
500 505 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met
515 520 525
Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp
530 535 540
Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg
545 550 555 560
Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln
565 570 575

Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser
580 585 590
Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu
595 600 605
Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr

610 615 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe

625 630 635 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met

645 650 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val

660 665 670

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

675 680 685

Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala

690 695 700

Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu

705 710 715 720

Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln

725 730 735

Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile

740 745 750

Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr

755 760 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Lys Leu Ile Asn

770 775 780

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

785 790 795 800

Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys

805 810 815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp

820 825 830

Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp

835 840 845

Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu

850 855 860

Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865 870 875 880
Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
885 890 895

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
900 905 910
Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
915 920 925
Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
930 935 940
Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
945 950 955 960

Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Ile Asn Gly Glu
965 970 975
Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu
980 985 990
Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn
995 1000 1005
Asp Lys Arg Pro Pro His Ile Ile Lys Thr Ile Ala Ser Lys Thr
1010 1015 1020

Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr
1025 1030 1035
Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 102
<211> 1367
<212> PRT
<213> Streptococcus pyogenes
<400> 102
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> 103

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 103

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> 104

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 104

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> 105

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 105

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> 106
<211> 4
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
peptide
<400> 106
Gly Gly Gly Ser

1

<210> 107

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 107

Gly Gly Gly Gly Ser
1 5

<210> 108

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 108

Glu Ala Ala Ala Lys
1 5

<210> 109

<211> 166

<212> PRT

<213> Unknown

<220><223> Description of Unknown: TadA sequence

<400> 109

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1 5 10 15

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val
20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
35 40 45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65 70 75 80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
85 90 95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala
100 105 110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg
115 120 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
130 135 140

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys
145 150 155 160

Ala Gln Ser Ser Thr Asp
165

<210> 110

<211> 186

<212> PRT

<213> Unknown

<220><223>

Description of Unknown:

TadA sequence

<400> 110

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1 5 10 15

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
20 25 30

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro
35 40 45

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp

50 55 60

Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met
65 70 75 80

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp
85 90 95
Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala
100 105 110
Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala

115 120 125
Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly
130 135 140
Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys
145 150 155 160
Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn
165 170 175
Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp
180 185

<210> 111
<211> 186
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
TadA sequence
<400> 111
Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met
1 5 10 15
Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro
20 25 30
Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
35 40 45

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
50 55 60
Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
65 70 75 80
Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg

85 90 95

Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr

100 105 110

Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr

115 120 125

Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met

130 135 140

Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val

145 150 155 160

Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly

165 170 175

Trp Asn Arg Ala Ile Gly Leu His Asp Pro

180 185

<210> 112

<211> 186

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

TadA sequence

<400> 112

Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val

1 5 10 15

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

20 25 30

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val

35 40 45

Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile

50 55 60

Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro

65 70 75 80

Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly

85 90 95

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu

100 105 110

Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His

115 120 125

Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val

130 135 140

Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn

145 150 155 160

Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala

165 170 175

Leu Arg Gln Gly Gly Leu Val Met Gln Asn

180 185

<210> 113

<211> 186

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

TadA sequence

<400> 113

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys

1 5 10 15

Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn

20 25 30

Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu

35 40 45

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu

50 55 60

Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala

65 70 75 80

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val

85 90 95

Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu
100 105 110
His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly
115 120 125
Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr
130 135 140
Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile
145 150 155 160

Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly
165 170 175
Ser Leu Met Asp Val Leu His Tyr Pro Gly
180 185

<210> 114
<211> 176
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
TadA sequence
<400> 114
Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1 5 10 15
Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg

20 25 30
His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro
35 40 45
Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp
50 55 60
Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met
65 70 75 80
Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

85 90 95
Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala

100 105 110

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala

115 120 125

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly

130 135 140

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys

145 150 155 160

Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn

165 170 175

<210> 115

<211> 173

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

TadA sequence

<400> 115

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

1 5 10 15

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg

20 25 30

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

35 40 45

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp

50 55 60

Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met

65 70 75 80

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

85 90 95

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala

100 105 110

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala

115 120 125

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly
130 135 140
Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys

145 150 155 160
Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln
165 170

<210> 116
<211> 168
<212> PRT
<213> Unknown
<220><223> Description of Unknown:
TadA sequence
<400> 116
Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1 5 10 15
Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
20 25 30

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro
35 40 45
Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp
50 55 60
Asn Arg Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
65 70 75 80
Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu
85 90 95

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
100 105 110
Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu
115 120 125
Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr
130 135 140
Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe

145 150 155 160

Arg Met Pro Arg Gln Val Phe Asn
165

<210> 117

<211> 178

<212> PRT

<213> Unknown

<220><223> Description of Unknown:
TadA sequence

<400> 117

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1 5 10 15

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
20 25 30

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro
35 40 45

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp
50 55 60

Asn Arg Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
65 70 75 80

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu
85 90 95

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
100 105 110

Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu
115 120 125

Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr
130 135 140

Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe
145 150 155 160

Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser

165 170 175

Thr Asp

<210> 118

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 118

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

165 170 175

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

225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

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

530 535 540

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

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

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

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Gly Ser Ser Gly Ser Glu Thr Pro

1010 1015 1020

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val

1025 1030 1035

Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala

1040 1045 1050

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu

1055 1060 1065

Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile

1070 1075 1080

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

1085 1090 1095

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr

1100 1105 1110

Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met

1115 1120 1125

Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala

1130 1135 1140

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro

1145 1150 1155

Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp

1160 1165 1170

Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln

1175 1180 1185

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Tyr Asp

1190 1195 1200

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

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

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

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

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

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

1445 1450 1455

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

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 119

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 119

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

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

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

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

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

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

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

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

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

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

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

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

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

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

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

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

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

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

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

```
        755                 760                 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005
```

```
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Gly
    1010                1015                1020
Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser Glu Ser  Ala Thr Pro
    1025                1030                1035
Glu Ser  Ser Gly Ser Glu Val  Glu Phe Ser His Glu  Tyr Trp Met
    1040                1045                1050

Arg His  Ala Leu Thr Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu
    1055                1060                1065
Val Pro  Val Gly Ala Val Leu  Val Leu Asn Asn Arg  Val Ile Gly
    1070                1075                1080
Glu Gly  Trp Asn Arg Ala Ile  Gly Leu His Asp Pro  Thr Ala His
    1085                1090                1095
Ala Glu  Ile Met Ala Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn
    1100                1105                1110

Tyr Arg  Leu Ile Asp Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys
    1115                1120                1125
Val Met  Cys Ala Gly Ala Met  Ile His Ser Arg Ile  Gly Arg Val
    1130                1135                1140
Val Phe  Gly Val Arg Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu
    1145                1150                1155
Met Asp  Val Leu His Tyr Pro  Gly Met Asn His Arg  Val Glu Ile
    1160                1165                1170

Thr Glu  Gly Ile Leu Ala Asp  Glu Cys Ala Ala Leu  Leu Cys Tyr
    1175                1180                1185
Phe Phe  Arg Met Pro Arg Gln  Val Phe Asn Ala Gln  Lys Lys Ala
    1190                1195                1200
Gln Ser  Ser Thr Asp Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1205                1210                1215
Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

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

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

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

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

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

1445 1450 1455

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

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550

<210> 120
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 120
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

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

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

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

530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

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

610 615 620

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

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

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

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Ser Ser Gly Ser Glu Thr Pro Gly

1025 1030 1035

Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu

1040 1045 1050

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys

1055 1060 1065

Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val

1070 1075 1080

Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly

1085 1090 1095

Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

1100 1105 1110

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

1115 1120 1125

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile

1130 1135 1140

His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys

1145 1150 1155

Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly

1160 1165 1170

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu

1175 1180 1185

Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val

1190 1195 1200

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

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

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

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

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

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

1445 1450 1455

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

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 121

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 121

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

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

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

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

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

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

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

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

530 535 540

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

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

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

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

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035

Tyr Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

1040 1045 1050

Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu

1055 1060 1065

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

1070 1075 1080

Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg

1085 1090 1095

Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro

1100 1105 1110

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val

1115 1120 1125

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe

1130 1135 1140

Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile

1145 1150 1155

Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala

1160 1165 1170

```
Gly Ser  Leu Met Asp Val Leu  His Tyr Pro Gly Met  Asn His Arg
    1175                 1180                 1185
Val Glu  Ile Thr Glu Gly Ile  Leu Ala Asp Glu Cys  Ala Ala Leu
    1190                 1195                 1200
Leu Cys  Tyr Phe Phe Arg Met  Pro Arg Gln Val Phe  Asn Ala Gln
    1205                 1210                 1215

Lys Lys  Ala Gln Ser Ser Thr  Asp Asn Ile Met Asn  Phe Phe Lys
    1220                 1225                 1230
Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                 1240                 1245
Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                 1255                 1260
Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                 1270                 1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                 1285                 1290
Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                 1300                 1305
Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                 1315                 1320
Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                 1330                 1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                 1345                 1350
Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                 1360                 1365
Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                 1375                 1380
Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                 1390                 1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
```

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

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

1445 1450 1455

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

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 122

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 122

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr

1070 1075 1080

Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp

1085 1090 1095

Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg

1100 1105 1110

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile

1115 1120 1125

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala

1130 1135 1140

His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln

1145 1150 1155

Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro

1160 1165 1170

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg

1175 1180 1185

Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser

1190 1195 1200

Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu

1205 1210 1215

Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys

1220 1225 1230

Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys

1235 1240 1245

Ala Gln Ser Ser Thr Asp Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

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

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

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

1445 1450 1455

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

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550

<210> 123

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 123

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

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

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

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

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

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

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

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

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

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

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

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

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

1115 1120 1125

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

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Gly

1235 1240 1245

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

1250 1255 1260

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met

1265 1270 1275

Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu

1280 1285 1290

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly

1295 1300 1305

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His

1310 1315 1320

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn

1325 1330 1335

Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys
1340 1345 1350
Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
1355 1360 1365
Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu
1370 1375 1380

Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile
1385 1390 1395
Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr
1400 1405 1410
Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala
1415 1420 1425
Gln Ser Ser Thr Asp Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550
<210> 124
<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 124

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

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

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

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

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

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

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

435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

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

530 535 540

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

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

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

690 695 700

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

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

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

885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

Asn Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala

1055 1060 1065

Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr

1070 1075 1080

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

1085 1090 1095

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val

1100 1105 1110

Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr

1115 1120 1125

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

1130 1135 1140

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

1145 1150 1155

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly

1160 1165 1170

Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly

1175 1180 1185

Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val

1190 1195 1200

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu

1205 1210 1215

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

1220 1225 1230

Lys Ala Gln Ser Ser Thr Asp Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

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

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

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

1445 1450 1455

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

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 125

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 125

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

1 5 10 15

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

20 25 30

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

35 40 45

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

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

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

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

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

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

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

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

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

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

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

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

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

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

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

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

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

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

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

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

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

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

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu

1025 1030 1035

Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His

1040 1045 1050

Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg

1055 1060 1065

Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn

1070 1075 1080

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp

1085 1090 1095

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu

1100 1105 1110

Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr

1115 1120 1125

Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg

1130 1135 1140

Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala

1145 1150 1155

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

1160 1165 1170

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala

1175 1180 1185

Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala

1190 1195 1200

Gln Lys Lys Ala Gln Ser Ser Thr Asp Gln Glu Ile Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

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

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305

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

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 126

<211> 1540

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 126

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

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

530 535 540

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

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
770 775 780
Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
785 790 795 800
His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

805 810 815
Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp
820 825 830
Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met
835 840 845
Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp
850 855 860
Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala

865 870 875 880

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala

885 890 895

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly

900 905 910

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys

915 920 925

Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Thr Thr Gln Lys

930 935 940

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

945 950 955 960

Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr

965 970 975

Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg

980 985 990

Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr

995 1000 1005

Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser

1010 1015 1020

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

1025 1030 1035

Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

1040 1045 1050

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

1055 1060 1065

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

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile

1115 1120 1125

Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln

1130 1135 1140

Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp

1145 1150 1155

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

1160 1165 1170

Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr

1175 1180 1185

Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys

1190 1195 1200

Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe

1205 1210 1215

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

1220 1225 1230

Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys

1235 1240 1245

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

1250 1255 1260

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser

1265 1270 1275

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

1280 1285 1290

Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser

1295 1300 1305

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys

1310 1315 1320

Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile

1325 1330 1335

Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe

1340 1345 1350

Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile

1355 1360 1365

Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys

1370 1375 1380

Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu

1385 1390 1395

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His

1400 1405 1410

Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln

1415 1420 1425

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

1430 1435 1440

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

1445 1450 1455

Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro

1460 1465 1470

Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr

1475 1480 1485

Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile

1490 1495 1500

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr

1505 1510 1515

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp

1520 1525 1530

Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 127

<211> 1540

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 127

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

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

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

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

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

225 230 235 240

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

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

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

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

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

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

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

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

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala

1250 1255 1260

Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr

1265 1270 1275

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

1280 1285 1290

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val

1295 1300 1305

Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr

1310 1315 1320

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

1325 1330 1335

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

1340 1345 1350

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly

1355 1360 1365

Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly

1370 1375 1380

Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val

1385 1390 1395

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu

1400 1405 1410

Cys Tyr Phe Phe Arg Met Pro Arg Glu Asp Asn Glu Gln Lys Gln

1415 1420 1425

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

1430 1435 1440

Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn
1445 1450 1455
Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro
1460 1465 1470
Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr
1475 1480 1485

Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
1490 1495 1500
Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr
1505 1510 1515
Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1520 1525 1530
Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 128
<211> 1543
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 128
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

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

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

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

565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

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

595 600 605

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

610 615 620

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

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

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

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

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

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

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

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

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

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

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

1115 1120 1125

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

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser

1250 1255 1260

Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser

1265 1270 1275

His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala

1280 1285 1290

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn

1295 1300 1305

Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His

1310 1315 1320

Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly

1325 1330 1335

Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val

1340 1345 1350

Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser

1355 1360 1365

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

1370 1375 1380

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn

1385 1390 1395

His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala

1400 1405 1410

Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Glu Asp Asn Glu

1415 1420 1425

Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu

1430 1435 1440

Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala

1445 1450 1455

Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg

1460 1465 1470

Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe

1475 1480 1485

Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

1490 1495 1500

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

1505 1510 1515

Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr

1520 1525 1530

Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 129

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 129

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

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

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

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

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

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

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

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser

1250 1255 1260

Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser Glu Val

1265 1270 1275

Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala

1280 1285 1290

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu

1295 1300 1305

Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile

1310 1315 1320

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

1325 1330 1335

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr

1340 1345 1350

Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met

1355 1360 1365

Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala

1370 1375 1380

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro

1385 1390 1395

Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp
1400 1405 1410

Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Glu
1415 1420 1425

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1430 1435 1440

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1445 1450 1455

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1460 1465 1470

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1475 1480 1485

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1490 1495 1500

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1505 1510 1515

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1520 1525 1530

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540 1545

<210> 130

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 130

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

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

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

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

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

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

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

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

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

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

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

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

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

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

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

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

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

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

1115 1120 1125

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

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser

1250 1255 1260

Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser Glu Val
1265 1270 1275
Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala
1280 1285 1290
Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu
1295 1300 1305
Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
1310 1315 1320

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
1325 1330 1335
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr
1340 1345 1350
Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met
1355 1360 1365
Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala
1370 1375 1380

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro
1385 1390 1395
Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp
1400 1405 1410
Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Arg Arg Glu
1415 1420 1425
Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1430 1435 1440

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1445 1450 1455
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1460 1465 1470
Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1475 1480 1485
His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1490 1495 1500

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1505 1510 1515
Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1520 1525 1530
Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540 1545

<210> 131
<211> 1546
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 131
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

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

130 135 140

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

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

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

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

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

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

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

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

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

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

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

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

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

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

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

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

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

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

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

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu

1250 1255 1260

Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser Glu Val Glu

1265 1270 1275

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys

1280 1285 1290

Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val

1295 1300 1305

Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly

1310 1315 1320

Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

1325 1330 1335

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu

1340 1345 1350

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile
1355 1360 1365
His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys
1370 1375 1380

Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly
1385 1390 1395
Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu
1400 1405 1410
Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Arg Arg Pro Glu
1415 1420 1425
Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1430 1435 1440

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1445 1450 1455
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1460 1465 1470
Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1475 1480 1485
His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1490 1495 1500

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1505 1510 1515
Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1520 1525 1530
Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540 1545

<210> 132
<211> 1551
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 132

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser

1250 1255 1260

Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Gly Ser Glu Val Glu

1265 1270 1275

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys

1280 1285 1290

Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val

1295 1300 1305

Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala His Ala

1310 1315 1320

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr

1325 1330 1335

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val

1340 1345 1350

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val

1355 1360 1365

Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met

1370 1375 1380

Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr

1385 1390 1395

Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe

1400 1405 1410

Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln

1415 1420 1425

Ser Ser Thr Asp Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1430 1435 1440

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1445 1450 1455
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1460 1465 1470
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1475 1480 1485
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1490 1495 1500

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1505 1510 1515
Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1520 1525 1530
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1535 1540 1545
Gly Gly Asp
1550

<210> 133
<211> 1546
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 133
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

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

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

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

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

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

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

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

1055 1060 1065

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

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

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

1115 1120 1125

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

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

1250 1255 1260

Ser Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser Glu

1265 1270 1275

Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu

1280 1285 1290

Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val
1295 1300 1305
Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
1310 1315 1320

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu
1325 1330 1335
Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala
1340 1345 1350
Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala
1355 1360 1365
Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn
1370 1375 1380

Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr
1385 1390 1395
Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
1400 1405 1410
Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Arg Arg
1415 1420 1425
Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1430 1435 1440

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1445 1450 1455
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1460 1465 1470
Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1475 1480 1485
His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1490 1495 1500

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1505 1510 1515
Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1520 1525 1530

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540 1545

<210> 134

<211> 1546

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 134

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

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

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

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

1115 1120 1125

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

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly
1250 1255 1260

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Ser Ser Ser Gly Ser
1265 1270 1275
Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1280 1285 1290
Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
1295 1300 1305
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg
1310 1315 1320

Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala
1325 1330 1335
Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp
1340 1345 1350
Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly
1355 1360 1365
Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg
1370 1375 1380

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

1385 1390 1395

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

1400 1405 1410

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Arg

1415 1420 1425

Arg Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1430 1435 1440

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

1445 1450 1455

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

1460 1465 1470

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1475 1480 1485

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1490 1495 1500

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1505 1510 1515

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu

1520 1525 1530

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540 1545

<210> 135

<211> 1533

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 135

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

1 5 10 15

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

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

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

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

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

770 775 780

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

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala

1250 1255 1260

Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr
1265 1270 1275
Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu
1280 1285 1290
Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val
1295 1300 1305
Ile Gly Glu Gly Trp Asn Arg Ala His Ala Glu Ile Met Ala Leu
1310 1315 1320

Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala
1325 1330 1335
Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala
1340 1345 1350
Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn
1355 1360 1365
Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr
1370 1375 1380

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
1385 1390 1395
Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg
1400 1405 1410
Gln Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1415 1420 1425
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1430 1435 1440

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1445 1450 1455
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1460 1465 1470
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1475 1480 1485

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1490 1495 1500

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1505 1510 1515
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1520 1525 1530

<210> 136
<211> 1549
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 136
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

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

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

385 390 395 400

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

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

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

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

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

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

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

Val Gly Thr Ala Leu Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

995 1000 1005

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

1010 1015 1020

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg

1025 1030 1035

Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala

1040 1045 1050

Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

1055 1060 1065

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly

1070 1075 1080

Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg

1085 1090 1095

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

1100 1105 1110

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

1115 1120 1125

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro
1130 1135 1140
Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp
1145 1150 1155

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr
1160 1165 1170
Pro Glu Ser Ser Gly Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu
1175 1180 1185
Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile
1190 1195 1200
Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe
1205 1210 1215

Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu
1220 1225 1230
Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly
1235 1240 1245
Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr
1250 1255 1260
Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys
1265 1270 1275

Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro
1280 1285 1290
Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp
1295 1300 1305
Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser
1310 1315 1320
Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu
1325 1330 1335

Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser
1340 1345 1350
Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr

1355 1360 1365

Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser

1370 1375 1380

Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala

1385 1390 1395

Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr

1400 1405 1410

Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly

1415 1420 1425

Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His

1430 1435 1440

Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser

1445 1450 1455

Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser

1460 1465 1470

Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu

1475 1480 1485

Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala

1490 1495 1500

Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr

1505 1510 1515

Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile

1520 1525 1530

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly

1535 1540 1545

Asp

<210> 137

<211> 1555

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 137

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

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

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

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

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
995 1000 1005

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
1010 1015 1020

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg
1025 1030 1035

Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala
1040 1045 1050

Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp
1055 1060 1065

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly
1070 1075 1080

Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg
1085 1090 1095

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
1100 1105 1110

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu
1115 1120 1125

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro
1130 1135 1140

Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp
1145 1150 1155

Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu
1160 1165 1170

Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Ile Lys Lys Tyr
1175 1180 1185

Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr
1190 1195 1200

Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys

1205 1210 1215

Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe
1220 1225 1230
Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro
1235 1240 1245
Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys
1250 1255 1260
Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln
1265 1270 1275

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser
1280 1285 1290
Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala
1295 1300 1305
Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
1310 1315 1320
Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys
1325 1330 1335

Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile
1340 1345 1350
Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe
1355 1360 1365
Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile
1370 1375 1380
Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys
1385 1390 1395

Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu
1400 1405 1410
Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His
1415 1420 1425
Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln
1430 1435 1440

Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu
1445 1450 1455

Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn
1460 1465 1470
Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro
1475 1480 1485
Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr
1490 1495 1500
Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
1505 1510 1515

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr
1520 1525 1530
Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1535 1540 1545
Leu Ser Gln Leu Gly Gly Asp
1550 1555
<210> 138
<211> 1421
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 138
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

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

530 535 540

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

545 550 555 560

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

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu
785 790 795 800
Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

805 810 815

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile

820 825 830

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His

835 840 845

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr

850 855 860

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

865 870 875 880

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly

885 890 895

Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu

900 905 910

His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

915 920 925

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg

930 935 940

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Gly

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe

1010 1015 1020

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His

1025 1030 1035

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

1040 1045 1050

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1055 1060 1065

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1070 1075 1080

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1085 1090 1095

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1100 1105 1110

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1115 1120 1125

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1130 1135 1140

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1145 1150 1155

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1160 1165 1170

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1175 1180 1185

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1190 1195 1200

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1205 1210 1215

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1220 1225 1230

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1235 1240 1245

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1250 1255 1260

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1265 1270 1275

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1280 1285 1290

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1295 1300 1305

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

1310 1315 1320

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1325 1330 1335

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1340 1345 1350

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1355 1360 1365

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1370 1375 1380

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1385 1390 1395

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1400 1405 1410

Asp Leu Ser Gln Leu Gly Gly Asp

1415 1420

<210> 139

<211> 1552

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 139

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110
Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Ser Gly Gly Ser Ser Gly Gly
145 150 155 160
Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
165 170 175
Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly
180 185 190
Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu
195 200 205

Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg
210 215 220
His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly
225 230 235 240
Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr
245 250 255
Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn
260 265 270

Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser
275 280 285
Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly

290 295 300

Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr
305 310 315 320

His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg
325 330 335

Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe
340 345 350

Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu
355 360 365

Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro
370 375 380

Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu
385 390 395 400

Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu
405 410 415

Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu
420 425 430

Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu
435 440 445

Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala
450 455 460

Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu
465 470 475 480

Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile
485 490 495

Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His
500 505 510

His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro
515 520 525

Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala
530 535 540

Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile
545 550 555 560
Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys
565 570 575
Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly
580 585 590

Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg
595 600 605
Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile
610 615 620
Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala
625 630 635 640
Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr
645 650 655

Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala
660 665 670
Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn
675 680 685
Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val
690 695 700
Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys
705 710 715 720

Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu
725 730 735
Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr
740 745 750
Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu
755 760 765
Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile
770 775 780

Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu

785 790 795 800

Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile

805 810 815

Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met

820 825 830

Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg

835 840 845

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

850 855 860

Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu

865 870 875 880

Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln

885 890 895

Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala

900 905 910

Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val

915 920 925

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

930 935 940

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

945 950 955 960

Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly

965 970 975

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

980 985 990

Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

995 1000 1005

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1010 1015 1020

His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1025 1030 1035

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1040 1045 1050
Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1055 1060 1065
Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1070 1075 1080
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1085 1090 1095

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
1100 1105 1110
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
1115 1120 1125
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
1130 1135 1140
Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
1145 1150 1155

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
1160 1165 1170
Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
1175 1180 1185
Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
1190 1195 1200
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
1205 1210 1215

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
1220 1225 1230
Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu
1235 1240 1245
Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
1250 1255 1260
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1265 1270 1275

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1280 1285 1290
Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
1295 1300 1305
Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1310 1315 1320
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1325 1330 1335

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1340 1345 1350
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1355 1360 1365
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1370 1375 1380
Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1385 1390 1395

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
1400 1405 1410
Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1415 1420 1425
Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1430 1435 1440
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1445 1450 1455

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1460 1465 1470
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1475 1480 1485
Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1490 1495 1500

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1505 1510 1515

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1520 1525 1530
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1535 1540 1545
Leu Gly Gly Asp
1550

<210> 140
<211> 1545
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 140

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

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

65 70 75 80
Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val
85 90 95
Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp
100 105 110
Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly
115 120 125

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met

130 135 140

Pro Arg Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro
145 150 155 160
Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly
165 170 175
Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser
180 185 190
Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys

195 200 205

Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu
210 215 220
Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg
225 230 235 240
Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile
245 250 255
Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp

260 265 270

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

325 330 335

His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn
340 345 350
Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr
355 360 365
Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp

370 375 380

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

385 390 395 400

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

405 410 415

Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn

420 425 430

Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr

435 440 445

Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala

450 455 460

Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser

465 470 475 480

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

485 490 495

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

500 505 510

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

515 520 525

Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala

530 535 540

Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met

545 550 555 560

Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu

565 570 575

Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His

580 585 590

Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro

595 600 605

Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg

610 615 620

Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala
625 630 635 640
Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu

645 650 655
Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met
660 665 670
Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His
675 680 685
Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val
690 695 700
Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu

705 710 715 720
Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val
725 730 735
Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe
740 745 750
Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu
755 760 765
Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu

770 775 780
Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu
785 790 795 800
Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr
805 810 815
Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg
820 825 830
Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg

835 840 845
Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly
850 855 860
Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr

865 870 875 880

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

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His

965 970 975

Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr

980 985 990

Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn

995 1000 1005

Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe

1010 1015 1020

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

1025 1030 1035

Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val

1040 1045 1050

Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu

1055 1060 1065

Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly

1070 1075 1080

Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu

1085 1090 1095

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp

1100 1105 1110

Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg

1115 1120 1125

Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe

1130 1135 1140

Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr

1145 1150 1155

His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala

1160 1165 1170

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

1175 1180 1185

Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu

1190 1195 1200

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

1205 1210 1215

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

1220 1225 1230

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

1235 1240 1245

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

1250 1255 1260

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

1265 1270 1275

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

1280 1285 1290

Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr

1295 1300 1305

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

1310 1315 1320

Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys

1325 1330 1335

Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys

1340 1345 1350

Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys

1355 1360 1365

Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu

1370 1375 1380

Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln

1385 1390 1395

Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu

1400 1405 1410

Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp

1415 1420 1425

Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu

1430 1435 1440

Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile

1445 1450 1455

Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys

1460 1465 1470

His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His

1475 1480 1485

Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr

1490 1495 1500

Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu

1505 1510 1515

Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr

1520 1525 1530

Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540 1545

<210> 141

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 141

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

1010 1015 1020

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1025 1030 1035

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly

1040 1045 1050

Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn

1055 1060 1065

Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met

1070 1075 1080

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile

1085 1090 1095

Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala

1100 1105 1110

Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val

1115 1120 1125

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu

1130 1135 1140

His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile

1145 1150 1155

Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met

1160 1165 1170

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

1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

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

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 142

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 142

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

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

325 330 335

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

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

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

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

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

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

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

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

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

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

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

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

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

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

1010 1015 1020

Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala

1025 1030 1035

Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val

1040 1045 1050

Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp

1055 1060 1065

Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile
1070 1075 1080
Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu
1085 1090 1095

Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys
1100 1105 1110
Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly
1115 1120 1125
Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val
1130 1135 1140
Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly
1145 1150 1155

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg
1160 1165 1170
Met Pro Arg Gln Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val
1175 1180 1185
Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly
1190 1195 1200
Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe
1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg
1220 1225 1230
Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
1235 1240 1245
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro
1250 1255 1260
Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe
1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1280 1285 1290
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

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

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 143

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 143

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

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

530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

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

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1010 1015 1020

Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His

1025 1030 1035

Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

1040 1045 1050

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly

1055 1060 1065

Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu

1070 1075 1080

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg

1085 1090 1095

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

1100 1105 1110

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

1115 1120 1125

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp

1130 1135 1140

Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu

1145 1150 1155

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe

1160 1165 1170

Arg Met Pro Arg Gln Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

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

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 144

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 144

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

1 5 10 15

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

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

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

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

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

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

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

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

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Gly Ser
995 1000 1005

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
1010 1015 1020

Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
1025 1030 1035
His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val
1040 1045 1050

Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu
1055 1060 1065
Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
1070 1075 1080
Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr
1085 1090 1095
Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val
1100 1105 1110

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val
1115 1120 1125
Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
1130 1135 1140
Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr
1145 1150 1155
Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe
1160 1165 1170

Phe Arg Met Pro Arg Gln Glu Phe Val Tyr Gly Asp Tyr Lys Val
1175 1180 1185
Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly
1190 1195 1200
Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe
1205 1210 1215
Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg
1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
1235 1240 1245
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

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

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr
1490 1495 1500
Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1505 1510 1515
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 145
<211> 1541
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 145
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

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

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Gly Ser Ser Gly Ser Glu Thr Pro

785 790 795 800

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu

805 810 815

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg

820 825 830

Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn

835 840 845

Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp

850 855 860

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val

865 870 875 880

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu
885 890 895
Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
900 905 910
Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu
915 920 925
Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr

930 935 940
Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe
945 950 955 960
Arg Met Pro Arg Gln Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn
965 970 975
Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly
980 985 990
Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp

995 1000 1005
Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp
1010 1015 1020
Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
1025 1030 1035
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
1040 1045 1050
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg

1055 1060 1065
Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
1070 1075 1080
Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
1085 1090 1095
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
1100 1105 1110
Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val

1115 1120 1125

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe

1130 1135 1140

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His

1145 1150 1155

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

1160 1165 1170

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

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

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 146

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 146

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

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

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

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

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

225 230 235 240

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

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

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

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

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

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

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

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

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Ser Gly Ser Glu Thr Pro Gly

785 790 795 800

Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe

805 810 815

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala

820 825 830

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn

835 840 845

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro

850 855 860

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

865 870 875 880

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro

885 890 895

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val

900 905 910

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met

915 920 925

Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu

930 935 940

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg

945 950 955 960

Met Pro Arg Gln Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn

965 970 975

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly
980 985 990

Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp
995 1000 1005

Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp
1010 1015 1020

Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
1025 1030 1035

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
1040 1045 1050

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg
1055 1060 1065

Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu
1070 1075 1080

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
1085 1090 1095

Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
1100 1105 1110

Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val
1115 1120 1125

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe
1130 1135 1140

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His
1145 1150 1155

Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys
1160 1165 1170

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val
1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly
1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

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

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala
1445 1450 1455
Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys
1460 1465 1470
Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr
1490 1495 1500
Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1505 1510 1515
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1520 1525 1530
Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 147
<211> 1541
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 147
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

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

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

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

565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

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

595 600 605

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

610 615 620

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

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

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

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr

785 790 795 800

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser

805 810 815

His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg

820 825 830

Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg

835 840 845

Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr

850 855 860

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln

865 870 875 880

Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys

885 890 895

Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val

900 905 910

Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp

915 920 925

Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly

930 935 940

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met

945 950 955 960

Pro Arg Gln Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn

965 970 975

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly

980 985 990

Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp

995 1000 1005

Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp

1010 1015 1020

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

1025 1030 1035

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

1040 1045 1050

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg

1055 1060 1065

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

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

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

1115 1120 1125

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe

1130 1135 1140

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His

1145 1150 1155

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

1160 1165 1170

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

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

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 148

<211> 1484

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 148

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

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

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

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

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

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

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

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

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

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

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

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Ser Ser Gly Ser Glu Val Glu
1010 1015 1020

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys
1025 1030 1035
Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val
1040 1045 1050
Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly
1055 1060 1065
Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
1070 1075 1080

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
1085 1090 1095
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile
1100 1105 1110
His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys
1115 1120 1125
Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly
1130 1135 1140

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu
1145 1150 1155
Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val

1160 1165 1170

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Glu Ile

1175 1180 1185

Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu

1190 1195 1200

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

1205 1210 1215

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

1220 1225 1230

Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly

1235 1240 1245

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

1250 1255 1260

Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu

1265 1270 1275

Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn

1280 1285 1290

Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys

1295 1300 1305

Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu

1310 1315 1320

Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys

1325 1330 1335

Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr

1340 1345 1350

Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn

1355 1360 1365

Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp

1370 1375 1380

Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu

1385 1390 1395

Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His
1400 1405 1410

Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
1415 1420 1425

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe
1430 1435 1440

Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val
1445 1450 1455

Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu
1460 1465 1470

Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1475 1480

<210> 149

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 149

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

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

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

```
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035
Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050
Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Met Asn
    1055                1060                1065
His Arg  Val Glu Ile Thr Glu  Gly Ile Leu Ala Asp  Glu Cys Ala
    1070                1075                1080

Ala Leu  Leu Cys Tyr Phe Phe  Arg Met Pro Arg Gln  Val Phe Asn
```

1085 1090 1095

Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser

1100 1105 1110

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

1115 1120 1125

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1130 1135 1140

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly

1145 1150 1155

Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn

1160 1165 1170

Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met

1175 1180 1185

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile

1190 1195 1200

Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala

1205 1210 1215

Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val

1220 1225 1230

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu

1235 1240 1245

His Tyr Pro Gly Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550

<210> 150

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 150

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Gly Ser Ser

690 695 700

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
705 710 715 720
Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
725 730 735
Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
740 745 750
Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

755 760 765

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
770 775 780
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
785 790 795 800
Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
805 810 815
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

820 825 830

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
835 840 845
Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
850 855 860
Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
865 870 875 880
Lys Ala Gln Ser Ser Thr Asp Leu Thr Phe Lys Glu Asp Ile Gln Lys

885 890 895

Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn
900 905 910
Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys
915 920 925
Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn

930 935 940

Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln

945 950 955 960

Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu

965 970 975

Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu

980 985 990

Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met

995 1000 1005

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp

1010 1015 1020

Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile

1025 1030 1035

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser

1040 1045 1050

Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr

1055 1060 1065

Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

1070 1075 1080

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

1085 1090 1095

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile

1100 1105 1110

Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

1115 1120 1125

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr

1130 1135 1140

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe

1145 1150 1155

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala

1160 1165 1170

Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro

1175 1180 1185

Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp

1190 1195 1200

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 151

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 151

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu

1250 1255 1260

Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val

1265 1270 1275

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser

1280 1285 1290

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1295 1300 1305

Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His

1310 1315 1320

Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

1325 1330 1335

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly

1340 1345 1350

Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu

1355 1360 1365

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg

1370 1375 1380

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

1385 1390 1395

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

1400 1405 1410

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp

1415 1420 1425

Val Leu His Tyr Pro Gly Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 152
<211> 1554
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 152
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu

1055 1060 1065

Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val

1070 1075 1080

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser

1085 1090 1095

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1100 1105 1110

Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His

1115 1120 1125

Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

1130 1135 1140

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly

1145 1150 1155

Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu

1160 1165 1170

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg

1175 1180 1185

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

1190 1195 1200

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

1205 1210 1215

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp

1220 1225 1230

Val Leu His Tyr Pro Gly Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 153

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 153

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

```
    450                 455                 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700
```

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala

1040 1045 1050

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg

1055 1060 1065

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly

1070 1075 1080

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

1085 1090 1095

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met

1100 1105 1110

Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu

1115 1120 1125

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly

1130 1135 1140

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His

1145 1150 1155

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn

1160 1165 1170

Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys

1175 1180 1185

Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
1190 1195 1200
Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu
1205 1210 1215
Met Asp Val Leu His Tyr Pro Gly Asn Ile Met Asn Phe Phe Lys
1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1250 1255 1260
Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 154

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 154

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Met Asn His Arg Val Glu Ile Thr Glu Gly

1295 1300 1305

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg

1310 1315 1320

Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser

1325 1330 1335

Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

1340 1345 1350

Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu

1355 1360 1365

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp

1370 1375 1380

Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg

1385 1390 1395

Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro

1400 1405 1410

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val

1415 1420 1425

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe

1430 1435 1440

Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile

1445 1450 1455

Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala

1460 1465 1470

Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545
Ser Gln Leu Gly Gly Asp
1550

<210> 155
<211> 1454
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 155

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu

785 790 795 800

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

805 810 815

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile

820 825 830

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His

835 840 845

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr

850 855 860

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

865 870 875 880

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly
885 890 895
Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu
900 905 910

His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu
915 920 925
Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg
930 935 940
Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Glu Glu
945 950 955 960
Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys
965 970 975

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly
980 985 990
Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val
995 1000 1005
Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser
1010 1015 1020
Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu
1025 1030 1035

Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg
1040 1045 1050
Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His
1055 1060 1065
His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu
1070 1075 1080
Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp
1085 1090 1095

Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln
1100 1105 1110
Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile

1115 1120 1125

Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile

1130 1135 1140

Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile

1145 1150 1155

Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu

1160 1165 1170

Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr

1175 1180 1185

Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp

1190 1195 1200

Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly

1205 1210 1215

Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala

1220 1225 1230

Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu

1235 1240 1245

Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn

1250 1255 1260

Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys

1265 1270 1275

Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu

1280 1285 1290

Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys

1295 1300 1305

Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr

1310 1315 1320

Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn

1325 1330 1335

Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp

1340 1345 1350

Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu
1355 1360 1365
Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His
1370 1375 1380
Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
1385 1390 1395

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe
1400 1405 1410
Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val
1415 1420 1425
Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu
1430 1435 1440
Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1445 1450

<210> 156
<211> 1420

<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 156
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu

785 790 795 800

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu

805 810 815

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile

820 825 830

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His
835 840 845

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr
850 855 860

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met
865 870 875 880

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly
885 890 895

Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu
900 905 910

His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu
915 920 925

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg
930 935 940

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Leu
945 950 955 960

Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr
965 970 975

Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
980 985 990

Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile
995 1000 1005

Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln
1010 1015 1020

Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp
1025 1030 1035

Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr
1040 1045 1050

Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr
1055 1060 1065

Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys

1070 1075 1080

Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe

1085 1090 1095

Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro

1100 1105 1110

Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys

1115 1120 1125

Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln

1130 1135 1140

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser

1145 1150 1155

Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala

1160 1165 1170

Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser

1175 1180 1185

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys

1190 1195 1200

Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile

1205 1210 1215

Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe

1220 1225 1230

Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile

1235 1240 1245

Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys

1250 1255 1260

Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu

1265 1270 1275

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His

1280 1285 1290

Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln

1295 1300 1305

Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu
1310 1315 1320
Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn
1325 1330 1335
Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro
1340 1345 1350
Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr
1355 1360 1365

Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
1370 1375 1380
Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr
1385 1390 1395
Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1400 1405 1410
Leu Ser Gln Leu Gly Gly Asp
1415 1420

<210> 157
<211> 1554
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 157
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Thr Ala His Ala Glu

995 1000 1005

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg

1010 1015 1020

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

1025 1030 1035

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

1040 1045 1050

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp

1055 1060 1065

Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu

1070 1075 1080

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe

1085 1090 1095

Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser

1100 1105 1110

Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu

1115 1120 1125

Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His

1130 1135 1140

Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg

1145 1150 1155

Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn

1160 1165 1170

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp

1175 1180 1185

Pro Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp

1190 1195 1200

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 158

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 158

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Thr Ala His Ala Glu Ile Met Ala

1010 1015 1020

Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp

1025 1030 1035

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly

1040 1045 1050

Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg

1055 1060 1065

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His

1070 1075 1080

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu

1085 1090 1095

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro

1100 1105 1110

Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp

1115 1120 1125

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr

1130 1135 1140

Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp
1145 1150 1155
Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
1160 1165 1170
Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile
1175 1180 1185

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Tyr Asp
1190 1195 1200
Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala
1205 1210 1215
Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys
1220 1225 1230
Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1250 1255 1260
Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275
Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 159

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 159

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Thr

1010 1015 1020

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

1025 1030 1035

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

1040 1045 1050

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly

1055 1060 1065

Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly

1070 1075 1080

Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val

1085 1090 1095

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu

1100 1105 1110

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

1115 1120 1125

Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro

1130 1135 1140

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val

1145 1150 1155

Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala

1160 1165 1170

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu

1175 1180 1185

Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile

1190 1195 1200

Gly Leu His Asp Pro Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 160

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 160

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Thr Ala His Ala Glu Ile Met Ala Leu
1025 1030 1035

Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala
1040 1045 1050

Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala
1055 1060 1065

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn

1070 1075 1080

Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr

1085 1090 1095

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala

1100 1105 1110

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg

1115 1120 1125

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly

1130 1135 1140

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro

1145 1150 1155

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met

1160 1165 1170

Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu

1175 1180 1185

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly

1190 1195 1200

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Gly Lys Ala

1205 1210 1215

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 161

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 161

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly

1040 1045 1050

Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val

1055 1060 1065

Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser

1070 1075 1080

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

1085 1090 1095

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn

1100 1105 1110

His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala

1115 1120 1125

Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn

1130 1135 1140

Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser

1145 1150 1155

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

1160 1165 1170

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1175 1180 1185

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly

1190 1195 1200

Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn

1205 1210 1215

Arg Ala Ile Gly Leu His Asp Pro Asn Ile Met Asn Phe Phe Lys

1220 1225 1230

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1235 1240 1245

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1250 1255 1260

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys

1280 1285 1290

Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg

1295 1300 1305

Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr

1340 1345 1350

Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu

1355 1360 1365

Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr

1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 162

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 162

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu

1055 1060 1065

Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr

1070 1075 1080

Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg

1085 1090 1095

Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala

1100 1105 1110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His

1115 1120 1125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala

1130 1135 1140

Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala

1145 1150 1155

Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu

1160 1165 1170

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser

1175 1180 1185

Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

1190 1195 1200

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

1205 1210 1215

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg

1220 1225 1230

Ala Ile Gly Leu His Asp Pro Gly Glu Ile Arg Lys Arg Pro Leu
1235 1240 1245
Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
1250 1255 1260
Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val
1265 1270 1275

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1280 1285 1290
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1295 1300 1305
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1310 1315 1320
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1325 1330 1335

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1340 1345 1350
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1355 1360 1365
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1370 1375 1380
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 163

<211> 1554

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 163

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Thr

1235 1240 1245

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

1250 1255 1260

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu

1265 1270 1275

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly

1280 1285 1290

Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly

1295 1300 1305

Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val

1310 1315 1320

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu

1325 1330 1335

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys

1340 1345 1350

Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro

1355 1360 1365

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val

1370 1375 1380

Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala

1385 1390 1395

Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu

1400 1405 1410

Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile

1415 1420 1425

Gly Leu His Asp Pro Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu

1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln

1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu

1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile

1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp

1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu

1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu

1535 1540 1545

Ser Gln Leu Gly Gly Asp

1550

<210> 164

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 164

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Glu Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 165

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 165

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 166
<211> 1368
<212> PRT
<213> Streptococcus pyogenes

<400> 166
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Arg

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Glu Tyr Arg Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 167

<211> 1359

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 167

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu

705 710 715 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Ile Gln Thr Val Gly Gln Asn Gly Gly Leu Phe Asp Asp Asn Pro

1100 1105 1110

Lys Ser Pro Leu Glu Val Thr Pro Ser Lys Leu Val Pro Leu Lys

1115 1120 1125

Lys Glu Leu Asn Pro Lys Lys Tyr Gly Gly Tyr Gln Lys Pro Thr

1130 1135 1140

Thr Ala Tyr Pro Val Leu Leu Ile Thr Asp Thr Lys Gln Leu Ile

1145 1150 1155

Pro Ile Ser Val Met Asn Lys Lys Gln Phe Glu Gln Asn Pro Val

1160 1165 1170

Lys Phe Leu Arg Asp Arg Gly Tyr Gln Gln Val Gly Lys Asn Asp

1175 1180 1185

Phe Ile Lys Leu Pro Lys Tyr Thr Leu Val Asp Ile Gly Asp Gly

1190 1195 1200

Ile Lys Arg Leu Trp Ala Ser Ser Lys Glu Ile His Lys Gly Asn

1205 1210 1215

Gln Leu Val Val Ser Lys Lys Ser Gln Ile Leu Leu Tyr His Ala

1220 1225 1230

His His Leu Asp Ser Asp Leu Ser Asn Asp Tyr Leu Gln Asn His

1235 1240 1245

Asn Gln Gln Phe Asp Val Leu Phe Asn Glu Ile Ile Ser Phe Ser

1250 1255 1260

Lys Lys Cys Lys Leu Gly Lys Glu His Ile Gln Lys Ile Glu Asn

1265 1270 1275

Val Tyr Ser Asn Lys Lys Asn Ser Ala Ser Ile Glu Glu Leu Ala

1280 1285 1290

Glu Ser Phe Ile Lys Leu Leu Gly Phe Thr Gln Leu Gly Ala Thr

1295 1300 1305

Ser Pro Phe Asn Phe Leu Gly Val Lys Leu Asn Gln Lys Gln Tyr

1310 1315 1320

Lys Gly Lys Lys Asp Tyr Ile Leu Pro Cys Thr Glu Gly Thr Leu

1325 1330 1335

Ile Arg Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Val Asp Leu

1340 1345 1350

Ser Lys Ile Gly Glu Asp

1355

<210> 168

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 168

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> 169

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 169

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> 170

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 170

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> 171

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 171

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> 172

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 172

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys

20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys

35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu

50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser

65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys

85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr

100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr

145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys

210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys

275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met

340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys

355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln

370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr

385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala

405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn

420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala

435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn

450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys

485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp

515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His

530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His

545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val

565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp

725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His

900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu

915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile

930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn

965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile

980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu

995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys

1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp

1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr

1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp

1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu

1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly

1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val

1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245

Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly

1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu

1280 1285 1290

Phe Val Gln Asn Arg Asn Asn

1295 1300

<210> 173

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 173

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> 174

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 174

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> 175

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 175

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> 176

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 176

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> 177

<211> 1129

<212> PRT

<213> Alicyclobacillus acidoterrestris

<400> 177

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> 178

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 178

Pro Lys Lys Lys Arg Lys Val Glu Gly Ala Asp Lys Arg Thr Ala Asp

1 5 10 15

Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val

20 25

<210> 179

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 179

Arg Lys Ser Gly Lys Ile Ala Ala Ile Val Val Lys Arg Pro Arg Lys

1 5 10 15

Pro Lys Lys Lys Arg Lys Val

20

<210> 180

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<220><221> MISC_FEATURE

<222> (1)..(21)

<223> This sequence may encompass 1, 3 or 7 "Gly Gly Ser"

repeating units

<400> 180

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

1 5 10 15

Gly Ser Gly Gly Ser

20

<210> 181

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 181

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> 182

<211> 229

<212> PRT

<213> Mesocricetus auratus

<400> 182

Met Ser Ser Glu Thr Gly Pro Val Val Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Asp Ala Phe Phe Asp Gln Gly Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Arg Trp Gly Gly Arg His

35 40 45

Asn Ile Trp Arg His Thr Gly Gln Asn Thr Ser Arg His Val Glu Ile

50 55 60

Asn Phe Ile Glu Lys Phe Thr Ser Glu Arg Tyr Phe Tyr Pro Ser Thr

65 70 75 80

Arg Cys Ser Ile Val Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Lys Ala Ile Thr Glu Phe Leu Ser Gly His Pro Asn Val Thr Leu

100 105 110

Phe Ile Tyr Ala Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Arg Gly Val Thr Ile Arg Ile Met

130 135 140

Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro

145 150 155 160

Pro Ser Asn Glu Val Tyr Trp Pro Arg Tyr Pro Asn Leu Trp Met Arg

165 170 175

Leu Tyr Ala Leu Glu Leu Tyr Cys Ile His Leu Gly Leu Pro Pro Cys

180 185 190

Leu Lys Ile Lys Arg Arg His Gln Tyr Pro Leu Thr Phe Phe Arg Leu

195 200 205

Asn Leu Gln Ser Cys His Tyr Gln Arg Ile Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Phe Ile

225

<210> 183

<211> 236

<212> PRT

<213> Pongo pygmaeus

<400> 183

Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Ser Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg

35 40 45

Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
50 55 60
Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Arg Phe His Ser Ser Ile
65 70 75 80
Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
85 90 95
Ser Gln Ala Ile Arg Glu Phe Leu Ser Gln His Pro Gly Val Thr Leu
100 105 110

Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160
Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met
165 170 175

Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys
180 185 190
Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Ala Phe Phe Arg Leu
195 200 205
His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu
210 215 220
Ala Thr Gly Leu Ile His Pro Ser Val Thr Trp Arg
225 230 235
<210> 184
<
211> 236
<212> PRT
<213> Oryctolagus cuniculus
<400> 184
Met Ala Ser Glu Lys Gly Pro Ser Asn Lys Asp Tyr Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro Trp Glu Phe Glu Val Phe Phe Asp Pro Gln Glu Leu

20 25 30

Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Ala Ser Ser

35 40 45

Lys Thr Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val

50 55 60

Asn Phe Leu Glu Lys Leu Thr Ser Glu Gly Arg Leu Gly Pro Ser Thr

65 70 75 80

Cys Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys

85 90 95

Ser Met Ala Ile Arg Glu Phe Leu Ser Gln His Pro Gly Val Thr Leu

100 105 110

Ile Ile Phe Val Ala Arg Leu Phe Gln His Met Asp Arg Arg Asn Arg

115 120 125

Gln Gly Leu Lys Asp Leu Val Thr Ser Gly Val Thr Val Arg Val Met

130 135 140

Ser Val Ser Glu Tyr Cys Tyr Cys Trp Glu Asn Phe Val Asn Tyr Pro

145 150 155 160

Pro Gly Lys Ala Ala Gln Trp Pro Arg Tyr Pro Pro Arg Trp Met Leu

165 170 175

Met Tyr Ala Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Lys Ile Ser Arg Arg His Gln Lys Gln Leu Thr Phe Phe Ser Leu

195 200 205

Thr Pro Gln Tyr Cys His Tyr Lys Met Ile Pro Pro Tyr Ile Leu Leu

210 215 220

Ala Thr Gly Leu Leu Gln Pro Ser Val Pro Trp Arg

225 230 235

<210> 185

<211> 235

<212> PRT

<213> Monodelphis domestica

<400> 185

Met Asn Ser Lys Thr Gly Pro Ser Val Gly Asp Ala Thr Leu Arg Arg

1 5 10 15

Arg Ile Lys Pro Trp Glu Phe Val Ala Phe Phe Asn Pro Gln Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys Trp Gly Asn Gln Asn

35 40 45

Ile Trp Arg His Ser Asn Gln Asn Thr Ser Gln His Ala Glu Ile Asn

50 55 60

Phe Met Glu Lys Phe Thr Ala Glu Arg His Phe Asn Ser Ser Val Arg

65 70 75 80

Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser

85 90 95

Lys Ala Ile Arg Lys Phe Leu Asp His Tyr Pro Asn Val Thr Leu Ala

100 105 110

Ile Phe Ile Ser Arg Leu Tyr Trp His Met Asp Gln Gln His Arg Gln

115 120 125

Gly Leu Lys Glu Leu Val His Ser Gly Val Thr Ile Gln Ile Met Ser

130 135 140

Tyr Ser Glu Tyr His Tyr Cys Trp Arg Asn Phe Val Asp Tyr Pro Gln

145 150 155 160

Gly Glu Glu Asp Tyr Trp Pro Lys Tyr Pro Tyr Leu Trp Ile Met Leu

165 170 175

Tyr Val Leu Glu Leu His Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu

180 185 190

Lys Ile Ser Gly Ser His Ser Asn Gln Leu Ala Leu Phe Ser Leu Asp

195 200 205

Leu Gln Asp Cys His Tyr Gln Lys Ile Pro Tyr Asn Val Leu Val Ala

210 215 220

Thr Gly Leu Val Gln Pro Phe Val Thr Trp Arg

225 230 235

<210> 186

<211> 224

<212> PRT

<213> Pongo pygmaeus

<400> 186

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Thr Glu Ala Ala Ser Gln
1 5 10 15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu
20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly
65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala
85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110

Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125

Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Leu
145 150 155 160

Glu Ile Gln Asp Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 187

<211> 440

<212> PRT

<213> Mus musculus

<400> 187

Met Gln Pro Gln Arg Leu Gly Pro Arg Ala Gly Met Gly Pro Phe Cys
1 5 10 15

Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro Ile Arg Asn Leu Ile
20 25 30

Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn Leu Gly Tyr Ala Lys
35 40 45

Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val Thr Arg Lys Asp Cys
50 55 60

Asp Ser Pro Val Ser Leu His His Gly Val Phe Lys Asn Lys Asp Asn
65 70 75 80

Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe His Asp Lys Val Leu
85 90 95

Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile Thr Trp Tyr Met Ser
100 105 110

Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile Val Arg Phe Leu Ala
115 120 125

Thr His His Asn Leu Ser Leu Asp Ile Phe Ser Ser Arg Leu Tyr Asn
130 135 140

Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys Arg Leu Val Gln Glu
145 150 155 160

Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu Phe Lys Lys Cys Trp
165 170 175

Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe Arg Pro Trp Lys Arg
180 185 190

Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys Leu Gln Glu Ile Leu
195 200 205

Arg Pro Cys Tyr Ile Ser Val Pro Ser Ser Ser Ser Ser Thr Leu Ser
210 215 220

Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr Arg Phe Trp Val Glu
225 230 235 240

Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu Phe Tyr Ser Gln Phe
245 250 255

Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr His Arg Met Lys Pro
260 265 270

Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly Gln Ala Pro Leu Lys
275 280 285

Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His Ala Glu Ile Leu Phe
290 295 300

Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln Val Thr Ile Thr Cys
305 310 315 320

Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala Trp Gln Leu Ala Ala
325 330 335

Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His Ile Tyr Thr Ser Arg
340 345 350

Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys Gly Leu Cys Ser Leu
355 360 365

Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp Leu Pro Gln Phe Thr
370 375 380

Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg Pro Phe Trp Pro Trp
385 390 395 400

Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln Arg Arg Leu Arg Arg
405 410 415

Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val Asn Asp Phe Gly Asn
420 425 430

Leu Gln Leu Gly Pro Pro Met Ser

435 440

<210> 188

<211> 367

<212> PRT

<213> Homo sapiens

<400> 188

Met Glu Pro Ile Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val
1 5 10 15

Lys Pro Tyr Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro
20 25 30

Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe
35 40 45

Cys Gln Ile Phe Gly Phe Pro Tyr Gly Thr Thr Phe Pro Gln Thr Lys
50 55 60

His Leu Thr Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln
65 70 75 80

Lys Gly His Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser
85 90 95

Met Leu Phe Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn
100 105 110

Asp Ser Ile Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asn
115 120 125

Glu Ala Asn His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Ile Thr
130 135 140

Tyr Pro Gly Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr
145 150 155 160

Glu Met Asp Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser
165 170 175

Leu Ala Ser Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly
180 185 190

Ile Trp His Ser Val Leu His Ser Phe Ile Ser Gly Val Ser Gly Ser
195 200 205

His Val Phe Gln Pro Ile Leu Thr Gly Arg Ala Leu Ala Asp Arg His
210 215 220
Asn Ala Tyr Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Tyr Phe Thr
225 230 235 240

Asp Val Leu Leu Gln Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Glu
245 250 255
Ala Leu Glu Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Phe Phe
260 265 270
Gln Met Pro Ser Gly Gln Leu Gln Pro Asn Leu Pro Pro Asp Leu Arg
275 280 285
Ala Pro Val Val Phe Val Leu Val Pro Leu Arg Asp Leu Pro Pro Met
290 295 300

His Met Gly Gln Asn Pro Asn Lys Pro Arg Asn Ile Val Arg His Leu
305 310 315 320
Asn Met Pro Gln Met Ser Phe Gln Glu Thr Lys Asp Leu Gly Arg Leu
325 330 335
Pro Thr Gly Arg Ser Val Glu Ile Val Glu Ile Thr Glu Gln Phe Ala
340 345 350
Ser Ser Lys Glu Ala Asp Glu Lys Lys Lys Lys Lys Gly Lys Lys
355 360 365

<210> 189
<211> 388
<212> PRT
<213> Rattus norvegicus
<400> 189
Met Glu Pro Leu Tyr Glu Glu Tyr Leu Thr His Ser Gly Thr Ile Val
1 5 10 15
Lys Pro Tyr Tyr Trp Leu Ser Val Ser Leu Asn Cys Thr Asn Cys Pro
20 25 30
Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Pro Tyr Thr Glu Phe
35 40 45

His Gln Thr Phe Gly Phe Pro Trp Ser Thr Tyr Pro Gln Thr Lys His

50 55 60

Leu Thr Phe Tyr Glu Leu Arg Ser Ser Ser Gly Asn Leu Ile Gln Lys
65 70 75 80

Gly Leu Ala Ser Asn Cys Thr Gly Ser His Thr His Pro Glu Ser Met
85 90 95

Leu Phe Glu Arg Asp Gly Tyr Leu Asp Ser Leu Ile Phe His Asp Ser
100 105 110

Asn Ile Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asp Glu

115 120 125

Ala Asn His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Met Asn Tyr
130 135 140

Pro Glu Val Thr Leu Ser Val Phe Phe Ser Gln Leu Tyr His Thr Glu
145 150 155 160

Asn Gln Phe Pro Thr Ser Ala Trp Asn Arg Glu Ala Leu Arg Gly Leu
165 170 175

Ala Ser Leu Trp Pro Gln Val Thr Leu Ser Ala Ile Ser Gly Gly Ile

180 185 190

Trp Gln Ser Ile Leu Glu Thr Phe Val Ser Gly Ile Ser Glu Gly Leu
195 200 205

Thr Ala Val Arg Pro Phe Thr Ala Gly Arg Thr Leu Thr Asp Arg Tyr
210 215 220

Asn Ala Tyr Glu Ile Asn Cys Ile Thr Glu Val Lys Pro Tyr Phe Thr
225 230 235 240

Asp Ala Leu His Ser Trp Gln Lys Glu Asn Gln Asp Gln Lys Val Trp

245 250 255

Ala Ala Ser Glu Asn Gln Pro Leu His Asn Thr Thr Pro Ala Gln Trp
260 265 270

Gln Pro Asp Met Ser Gln Asp Cys Arg Thr Pro Ala Val Phe Met Leu
275 280 285

Val Pro Tyr Arg Asp Leu Pro Pro Ile His Val Asn Pro Ser Pro Gln

290 295 300

Lys Pro Arg Thr Val Val Arg His Leu Asn Thr Leu Gln Leu Ser Ala

305 310 315 320

Ser Lys Val Lys Ala Leu Arg Lys Ser Pro Ser Gly Arg Pro Val Lys

325 330 335

Lys Glu Glu Ala Arg Lys Gly Ser Thr Arg Ser Gln Glu Ala Asn Glu

340 345 350

Thr Asn Lys Ser Lys Trp Lys Lys Gln Thr Leu Phe Ile Lys Ser Asn

355 360 365

Ile Cys His Leu Leu Glu Arg Glu Gln Lys Lys Ile Gly Ile Leu Ser

370 375 380

Ser Trp Ser Val

385

<210> 190

<211> 363

<212> PRT

<213> Macaca fascicularis

<400> 190

Met Glu Pro Thr Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val

1 5 10 15

Lys Pro Tyr Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro

20 25 30

Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe

35 40 45

Cys Gln Ile Phe Gly Phe Pro Tyr Gly Thr Thr Tyr Pro Gln Thr Lys

50 55 60

His Leu Thr Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln

65 70 75 80

Lys Gly His Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser

85 90 95

Met Leu Phe Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn

100 105 110

Asp Ser Ile Arg His Ile Ile Leu Tyr Cys Asn Asn Ser Pro Cys Asn
115 120 125
Glu Ala Asn His Cys Cys Ile Ser Lys Val Tyr Asn Phe Leu Ile Thr
130 135 140
Tyr Pro Gly Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr
145 150 155 160
Glu Met Asp Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser
165 170 175

Leu Ala Ser Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly
180 185 190
Ile Trp His Ser Val Leu His Ser Phe Val Ser Gly Val Ser Gly Ser
195 200 205
His Val Phe Gln Pro Ile Leu Thr Gly Arg Ala Leu Thr Asp Arg Tyr
210 215 220
Asn Ala Tyr Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Phe Phe Thr
225 230 235 240

Asp Val Leu Leu His Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Met
245 250 255
Ala Leu Glu Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Ser Phe
260 265 270
Gln Met Thr Ser Gly Ile Pro Pro Asp Leu Arg Ala Pro Val Val Phe
275 280 285
Val Leu Leu Pro Leu Arg Asp Leu Pro Pro Met His Met Gly Gln Asp
290 295 300

Pro Asn Lys Pro Arg Asn Ile Ile Arg His Leu Asn Met Pro Gln Met
305 310 315 320
Ser Phe Gln Glu Thr Lys Asp Leu Glu Arg Leu Pro Thr Arg Arg Ser
325 330 335
Val Glu Thr Val Glu Ile Thr Glu Arg Phe Ala Ser Ser Lys Gln Ala
340 345 350

Glu Glu Lys Thr Lys Lys Lys Lys Gly Lys Lys

355 360

<210> 191

<211> 22

<212> DNA

<213> Homo sapiens

<400> 191

ggaaaccttg aataagaatg ga 22

<210> 192

<211> 20

<212> DNA

<213> Homo sapiens

<400> 192

gagtccgagc agaagaagaa 20

<210> 193

<211> 20

<212> DNA

<213> Homo sapiens

<400> 193

ggcccagact gagcacgtga 20

<210> 194

<211> 198

<212> PRT

<213> Mus musculus

<400> 194

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> 195

<211> 224

<212> PRT

<213> Petromyzon marinus

<400> 195

Met Ala Gly Tyr Glu Cys Val Arg Val Ser Glu Lys Leu Asp Phe Asp

1 5 10 15

Thr Phe Glu Phe Gln Phe Glu Asn Leu His Tyr Ala Thr Glu Arg His

20 25 30

Arg Thr Tyr Val Ile Phe Asp Val Lys Pro Gln Ser Ala Gly Gly Arg

35 40 45

Ser Arg Arg Leu Trp Gly Tyr Ile Ile Asn Asn Pro Asn Val Cys His

50 55 60

Ala Glu Leu Ile Leu Met Ser Met Ile Asp Arg His Leu Glu Ser Asn
65 70 75 80
Pro Gly Val Tyr Ala Met Thr Trp Tyr Met Ser Trp Ser Pro Cys Ala
85 90 95
Asn Cys Ser Ser Lys Leu Asn Pro Trp Leu Lys Asn Leu Leu Glu Glu
100 105 110
Gln Gly His Thr Leu Thr Met His Phe Ser Arg Ile Tyr Asp Arg Asp

115 120 125
Arg Glu Gly Asp His Arg Gly Leu Arg Gly Leu Lys His Val Ser Asn
130 135 140
Ser Phe Arg Met Gly Val Val Gly Arg Ala Glu Val Lys Glu Cys Leu
145 150 155 160
Ala Glu Tyr Val Glu Ala Ser Arg Arg Thr Leu Thr Trp Leu Asp Thr
165 170 175
Thr Glu Ser Met Ala Ala Lys Met Arg Arg Lys Leu Phe Cys Ile Leu

180 185 190
Val Arg Cys Ala Gly Met Arg Glu Ser Gly Ile Pro Leu His Leu Phe
195 200 205
Thr Leu Gln Thr Pro Leu Leu Ser Gly Arg Val Val Trp Trp Arg Val
210 215 220

<210> 196
<211> 331
<212> PRT
<213> Petromyzon marinus
<400> 196

Met Glu Leu Arg Glu Val Val Asp Cys Ala Leu Ala Ser Cys Val Arg
1 5 10 15

His Glu Pro Leu Ser Arg Val Ala Phe Leu Arg Cys Phe Ala Ala Pro
20 25 30
Ser Gln Lys Pro Arg Gly Thr Val Ile Leu Phe Tyr Val Glu Gly Ala
35 40 45
Gly Arg Gly Val Thr Gly Gly His Ala Val Asn Tyr Asn Lys Gln Gly

50 55 60

Thr Ser Ile His Ala Glu Val Leu Leu Leu Ser Ala Val Arg Ala Ala
65 70 75 80

Leu Leu Arg Arg Arg Arg Cys Glu Asp Gly Glu Glu Ala Thr Arg Gly
85 90 95

Cys Thr Leu His Cys Tyr Ser Thr Tyr Ser Pro Cys Arg Asp Cys Val
100 105 110

Glu Tyr Ile Gln Glu Phe Gly Ala Ser Thr Gly Val Arg Val Val Ile
115 120 125

His Cys Cys Arg Leu Tyr Glu Leu Asp Val Asn Arg Arg Arg Ser Glu
130 135 140

Ala Glu Gly Val Leu Arg Ser Leu Ser Arg Leu Gly Arg Asp Phe Arg
145 150 155 160

Leu Met Gly Pro Arg Asp Ala Ile Ala Leu Leu Leu Gly Gly Arg Leu
165 170 175

Ala Asn Thr Ala Asp Gly Glu Ser Gly Ala Ser Gly Asn Ala Trp Val
180 185 190

Thr Glu Thr Asn Val Val Glu Pro Leu Val Asp Met Thr Gly Phe Gly
195 200 205

Asp Glu Asp Leu His Ala Gln Val Gln Arg Asn Lys Gln Ile Arg Glu
210 215 220

Ala Tyr Ala Asn Tyr Ala Ser Ala Val Ser Leu Met Leu Gly Glu Leu
225 230 235 240

His Val Asp Pro Asp Lys Phe Pro Phe Leu Ala Glu Phe Leu Ala Gln
245 250 255

Thr Ser Val Glu Pro Ser Gly Thr Pro Arg Glu Thr Arg Gly Arg Pro
260 265 270

Arg Gly Ala Ser Ser Arg Gly Pro Glu Ile Gly Arg Gln Arg Pro Ala
275 280 285

Asp Phe Glu Arg Ala Leu Gly Ala Tyr Gly Leu Phe Leu His Pro Arg
290 295 300

Ile Val Ser Arg Glu Ala Asp Arg Glu Glu Ile Lys Arg Asp Leu Ile
305 310 315 320
Val Val Met Arg Lys His Asn Tyr Gln Gly Pro
325 330

<210> 197
<211> 209

<212> PRT
<213> Petromyzon marinus
<400> 197
Met Ala Gly Asp Glu Asn Val Arg Val Ser Glu Lys Leu Asp Phe Asp
1 5 10 15
Thr Phe Glu Phe Gln Phe Glu Asn Leu His Tyr Ala Thr Glu Arg His
20 25 30
Arg Thr Tyr Val Ile Phe Asp Val Lys Pro Gln Ser Ala Gly Gly Arg
35 40 45
Ser Arg Arg Leu Trp Gly Tyr Ile Ile Asn Asn Pro Asn Val Cys His

50 55 60
Ala Glu Leu Ile Leu Met Ser Met Ile Asp Arg His Leu Glu Ser Asn
65 70 75 80
Pro Gly Val Tyr Ala Met Thr Trp Tyr Met Ser Trp Ser Pro Cys Ala
85 90 95
Asn Cys Ser Ser Lys Leu Asn Pro Trp Leu Lys Asn Leu Leu Glu Glu
100 105 110
Gln Gly His Thr Leu Met Met His Phe Ser Arg Ile Tyr Asp Arg Asp

115 120 125
Arg Glu Gly Asp His Arg Gly Leu Arg Gly Leu Lys His Val Ser Asn
130 135 140
Ser Phe Arg Met Gly Val Val Gly Arg Ala Glu Val Lys Glu Cys Leu
145 150 155 160
Ala Glu Tyr Val Glu Ala Ser Arg Arg Thr Leu Thr Trp Leu Asp Thr
165 170 175

Thr Glu Ser Met Ala Ala Lys Met Arg Arg Lys Leu Phe Cys Ile Leu

180 185 190

Val Arg Cys Ala Gly Met Arg Glu Ser Gly Met Pro Leu His Leu Phe

195 200 205

Thr

<210> 198

<211> 158

<212> PRT

<213> Saccharomyces cerevisiae

<400> 198

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

Ile Ala Tyr Glu Glu Ala Ala Leu Gly Tyr Lys Glu Gly Gly Val Pro

20 25 30

Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg

35 40 45

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val

100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe

130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu

145 150 155

<210> 199

<211> 218

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 199

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu
180 185 190

Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro

195 200 205

Pro His Ile Leu Trp Ala Thr Gly Leu Lys

210 215

<210> 200

<211> 218

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 200

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln His Tyr Gln Arg Leu Pro

195 200 205

Pro His Ile Leu Trp Ala Thr Gly Leu Lys

210 215

<210> 201

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 201

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgggttc tagcggcagc 3060

gagactcccg ggacctcaga gtccgccaca cccgaaagtt ctggttccga agtcgagttt 3120

tcccatgagt actggatgag acacgcattg actctcgcaa agagggctcg agatgaacgc 3180

gaggtgcccg tgggggcagt actcgtgctc aacaatcgcg taatcggcga aggttggaat 3240

agggcaatcg gactccacga ccccactgca catgcggaaa tcatggccct tcgacaggga 3300

gggcttgtga tgcagaatta tcgacttatc gatgcgacgc tgtacgtcac gtttgaacct 3360

tgcgtaatgt gcgcgggagc tatgattcac tcccgcattg gacgagttgt attcggtgtt 3420

cgcaacgcca agacgggtgc cgcaggttca ctgatggacg tgctgcatta cccaggcatg 3480

aaccaccggg tagaaatcac agaaggcata ttggcggacg aatgtgcggc gctgttgtgt 3540

tacttttttc gcatgcccag gcaggtcttt aacgcccaga aaaaagcaca atcctctact 3600

gactacgacg tgcggaagat gatcgccaag agcgagcagg aaatcggcaa ggctaccgcc 3660

aagtacttct tctacagcaa catcatgaac tttttcaaga ccgagattac cctggccaac 3720

ggcgagatcc ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 202

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 202

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcggtt ctagcggcag cgagactccc gggacctcag agtccgccac acccgaaagt 3120

tctggttccg aagtcgagtt ttcccatgag tactggatga gacacgcatt gactctcgca 3180

aagagggctc gagatgaacg cgaggtgccc gtgggggcag tactcgtgct caacaatcgc 3240

gtaatcggcg aaggttggaa tagggcaatc ggactccacg accccactgc acatgcggaa 3300

atcatggccc ttcgacaggg agggcttgtg atgcagaatt atcgacttat cgatgcgacg 3360

ctgtacgtca cgtttgaacc ttgcgtaatg tgcgcgggag ctatgattca ctcccgcatt 3420
ggacgagttg tattcggtgt tcgcaacgcc aagacgggtg ccgcaggttc actgatggac 3480
gtgctgcatt acccaggcat gaaccaccgg gtagaaatca cagaaggcat attggcggac 3540
gaatgtgcgg cgctgttgtg ttacttttt cgcatgccca ggcaggtctt taacgcccag 3600
aaaaaagcac aatcctctac tgacgccaag agcgagcagg aaatcggcaa ggctaccgcc 3660
aagtacttct tctacagcaa catcatgaac ttttttcaaga ccgagattac cctggccaac 3720
ggcgagatcc ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840
gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900
aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960
gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020
aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080
gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140
atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260
ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320
aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380
gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440
aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500
ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560
aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 203

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 203

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220
gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280
atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400
gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460
gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580
gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640
aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820
actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880
aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940
taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000
taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060
atgatcgcca agagcgagca ggaaatcggt tctagcggca gcgagactcc cgggacctca 3120
gagtccgcca cacccgaaag ttctggttcc gaagtcgagt tttcccatga gtactggatg 3180

agacacgcat tgactctcgc aaagagggct cgagatgaac gcgaggtgcc cgtgggggca 3240
gtactcgtgc tcaacaatcg cgtaatcggc gaaggttgga atagggcaat cggactccac 3300
gaccccactg cacatgcgga aatcatggcc cttcgacagg gagggcttgt gatgcagaat 3360
tatcgactta tcgatgcgac gctgtacgtc acgtttgaac cttgcgtaat gtgcgcggga 3420
gctatgattc actcccgcat tggacgagtt gtattcggtg ttcgcaacgc caagacgggt 3480
gccgcaggtt cactgatgga cgtgctgcat tacccaggca tgaaccaccg ggtagaaatc 3540
acagaaggca tattggcgga cgaatgtgcg gcgctgttgt gttacttttt tcgcatgccc 3600

aggcaggtct ttaacgccca gaaaaaagca caatcctcta ctgacggcaa ggctaccgcc 3660

aagtacttct tctacagcaa catcatgaac ttttttcaaga ccgagattac cctggccaac 3720

ggcgagatcc ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 204

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 204

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660
cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960
atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140
gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200
aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380
agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440
gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500
aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560
tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620
agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680
gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800
atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860
ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

ggttctagcg gcagcgagac tcccgggacc tcagagtccg ccacacccga aagttctggt 3180

tccgaagtcg agttttccca tgagtactgg atgagacacg cattgactct cgcaaagagg 3240

gctcgagatg aacgcgaggt gcccgtgggg gcagtactcg tgctcaacaa tcgcgtaatc 3300

ggcgaaggtt ggaatagggc aatcggactc cacgacccca ctgcacatgc ggaaatcatg 3360

gcccttcgac agggagggct tgtgatgcag aattatcgac ttatcgatgc gacgctgtac 3420

gtcacgtttg aaccttgcgt aatgtgcgcg ggagctatga ttcactcccg cattggacga 3480

gttgtattcg gtgttcgcaa cgccaagacg ggtgccgcag gttcactgat ggacgtgctg 3540

cattacccag gcatgaacca ccgggtagaa atcacagaag gcatattggc ggacgaatgt 3600

gcggcgctgt tgtgttactt ttttcgcatg cccaggcagg tctttaacgc ccagaaaaaa 3660

gcacaatcct ctactgacaa catcatgaac ttttttcaaga ccgagattac cctggccaac 3720

ggcgagatcc ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 205

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 205

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

aacatcatga acttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg 3180

cctctgatcg agacaaacgg cgaaggttct agcggcagcg agactcccgg gacctcagag 3240

tccgccacac ccgaaagttc tggttccgaa gtcgagtttt cccatgagta ctggatgaga 3300

cacgcattga ctctcgcaaa gagggctcga gatgaacgcg aggtgcccgt gggggcagta 3360

ctcgtgctca acaatcgcgt aatcggcgaa ggttggaata gggcaatcgg actccacgac 3420

cccactgcac atgcggaaat catggcccct cgacagggag ggcttgtgat gcagaattat 3480

cgacttatcg atgcgacgct gtacgtcacg tttgaacctt gcgtaatgtg cgcgggagct 3540

atgattcact cccgcattgg acgagttgta ttcggtgttc gcaacgccaa gacgggtgcc 3600

gcaggttcac tgatggacgt gctgcattac ccaggcatga accaccgggt agaaatcaca 3660

gaaggcatat tggcggacga atgtgcggcg ctgttgtgtt actttttcg catgcccagg 3720

caggtcttta acgcccagaa aaaagcacaa tcctctactg acaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560
aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620
ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 206
<211> 4662
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polynucleotide
<400> 206

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60
atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240
tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420
aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480
atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660
cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900
ctgctgagcg acatcctgag agtgaacacc gagatcacca aggccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140
gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

aacatcatga acttttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg 3180

cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt 3240

gccaccgtgc ggaaagtgct gagcatgccc caagtgaata tcgtgaaaaa gaccgaggtg 3300

cagacaggcg gcttcagcaa agagtctatc ctgcccaaga ggaacagcga taagctgatc 3360

gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc 3420

tattctgtgc tggtggtggc caaagtggaa aagggcaagt ccaagaaact gaagagtgtg 3480

aaagagctgc tggggatcac catcatggaa agaagcagct tcgagaagaa tcccatcgac 3540

tttctggaag ccaagggcta caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag 3600

tactccctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg 3660

cagaagggaa acgaactggc cctgccctcc aaatatgtga acttcctgta cctggccagc 3720

cactatgaga agctgaaggg cggttctagc ggcagcgaga ctcccgggac ctcagagtcc 3780

gccacacccg aaagttctgg ttccgaagtc gagttttccc atgagtactg gatgagacac 3840

gcattgactc tcgcaaagag ggctcgagat gaacgcgagg tgcccgtggg ggcagtactc 3900

gtgctcaaca atcgcgtaat cggcgaaggt tggaataggg caatcggact ccacgacccc 3960

actgcacatg cggaaatcat ggcccttcga cagggagggc ttgtgatgca gaattatcga 4020

cttatcgatg cgacgctgta cgtcacgttt gaaccttgcg taatgtgcgc gggagctatg 4080

attcactccc gcattggacg agttgtattc ggtgttcgca acgccaagac gggtgccgca 4140

ggttcactga tggacgtgct gcattaccca ggcatgaacc accgggtaga aatcacagaa 4200

ggcatattgg cggacgaatg tgcggcgctg ttgtgttact tttttcgcat gcccaggcag 4260

gtctttaacg cccagaaaaa agcacaatcc tctactgact cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 207

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 207

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680
gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800
atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860
ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220
gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280
atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340
atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400
gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580
gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640
aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760
ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820
actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940
taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000
taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060
atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120
aacatcatga acttttttcaa gaccgagatt accctggcca acggttctag cggcagcgag 3180
actcccggga cctcagagtc cgccacaccc gaaagttctg gttccgaagt cgagttttcc 3240
catgagtact ggatgagaca cgcattgact ctcgcaaaga gggctcgaga tgaacgcgag 3300

gtgcccgtgg gggcagtact cgtgctcaac aatcgcgtaa tcggcgaagg ttggaatagg 3360

gcaatcggac tccacgaccc cactgcacat gcggaaatca tggcccttcg acagggaggg 3420

cttgtgatgc agaattatcg acttatcgat gcgacgctgt acgtcacgtt tgaaccttgc 3480

gtaatgtgcg cgggagctat gattcactcc cgcattggac gagttgtatt cggtgttcgc 3540

aacgccaaga cgggtgccgc aggttcactg atggacgtgc tgcattaccc aggcatgaac 3600

caccgggtag aaatcacaga aggcatattg gcggacgaat gtgcggcgct gttgtgttac 3660

ttttttcgca tgcccaggca ggtctttaac gcccagaaaa aagcacaatc ctctactgac 3720

ggcgagatcc ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 208

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 208

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgaggg ttctagcggc agcgagactc ccgggacctc agagtccgcc 3120

acacccgaaa gttctggttc cgaagtcgag ttttcccatg agtactggat gagacacgca 3180

ttgactctcg caaagagggc tcgagatgaa cgcgaggtgc ccgtgggggc agtactcgtg 3240

ctcaacaatc gcgtaatcgg cgaaggttgg aatagggcaa tcggactcca cgaccccact 3300

gcacatgcgg aaatcatggc ccttcgacag ggagggcttg tgatgcagaa ttatcgactt 3360

atcgatgcga cgctgtacgt cacgtttgaa ccttgcgtaa tgtgcgcggg agctatgatt 3420

cactcccgca ttggacgagt tgtattcggt gttcgcaacg ccaagacggg tgccgcaggt 3480

tcactgatgg acgtgctgca ttacccaggc atgaaccacc gggtagaaat cacagaaggc 3540

atattggcgg acgaatgtgc ggcgctgttg tgttactttt ttcgcatgcc caggcaggtc 3600

tttaacgccc agaaaaaagc acaatcctct actgaccagg aaatcggcaa ggctaccgcc 3660

aagtacttct tctacagcaa catcatgaac tttttcaaga ccgagattac cctggccaac 3720

ggcgagatcc ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 209

<211> 4620

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 209

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagggttct agcggcagcg agactcccgg gacctcagag 2340

tccgccacac ccgaaagttc tggttccgaa gtcgagtttt cccatgagta ctggatgaga 2400

cacgcattga ctctcgcaaa gagggctcga gatgaacgcg aggtgcccgt gggggcagta 2460

ctcgtgctca acaatcgcgt aatcggcgaa ggttggaata gggcaatcgg actccacgac 2520

cccactgcac atgcggaaat catggccctt cgacagggag ggcttgtgat gcagaattat 2580

cgacttatcg atgcgacgct gtacgtcacg tttgaacctt gcgtaatgtg cgcgggagct 2640

atgattcact cccgcattgg acgagttgta ttcggtgttc gcaacgccaa gacgggtgcc 2700

gcaggttcac tgatggacgt gctgcattac ccaggcatga accaccgggt agaaatcaca 2760

gaaggcatat tggcggacga atgtgcggcg ctgttgtgtt actttttcg catgcccagg 2820

accacccaga agggacagaa gaacagccgc gagagaatga agcggatcga agagggcatc 2880

aaagagctgg gcagccagat cctgaaagaa caccccgtgg aaaacaccca gctgcagaac 2940

gagaagctgt acctgtacta cctgcagaat gggcgggata tgtacgtgga ccaggaactg 3000

gacatcaacc ggctgtccga ctacgatgtg gaccatatcg tgcctcagag ctttctgaag 3060

gacgactcca tcgacaacaa ggtgctgacc agaagcgaca agaaccgggg caagagcgac 3120

aacgtgccct ccgaagaggt cgtgaagaag atgaagaact actggcggca gctgctgaac 3180

gccaagctga ttacccagag aaagttcgac aatctgacca aggccgagag aggcggcctg 3240

agcgaactgg ataaggccgg cttcatcaag agacagctgg tggaaacccg gcagatcaca 3300

aagcacgtgg cacagatcct ggactcccgg atgaacacta agtacgacga gaatgacaag 3360

ctgatccggg aagtgaaagt gatcaccctg aagtccaagc tggtgtccga tttccggaag 3420

gatttccagt tttacaaagt gcgcgagatc aacaactacc accacgccca cgacgcctac 3480

ctgaacgccg tcgtgggaac cgccctgatc aaaaagtacc ctaagctgga aagcgagttc 3540

gtgtacggcg actacaaggt gtacgacgtg cggaagatga tcgccaagag cgagcaggaa 3600

atcggcaagg ctaccgccaa gtacttcttc tacagcaaca tcatgaactt tttcaagacc 3660

gagattaccc tggccaacgg cgagatccgg aagcggcctc tgatcgagac aaacggcgaa 3720

accggggaga tcgtgtggga taagggccgg gattttgcca ccgtgcggaa agtgctgagc 3780

atgccccaag tgaatatcgt gaaaaagacc gaggtgcaga caggcggctt cagcaaagag 3840

tctatcctgc ccaagaggaa cagcgataag ctgatcgcca gaaagaagga ctgggaccct 3900

aagaagtacg gcggcttcga cagccccacc gtggcctatt ctgtgctggt ggtggccaaa 3960

gtggaaaagg gcaagtccaa gaaactgaag agtgtgaaag agctgctggg gatcaccatc 4020

atggaaagaa gcagcttcga gaagaatccc atcgactttc tggaagccaa gggctacaaa 4080

gaagtgaaaa aggacctgat catcaagctg cctaagtact ccctgttcga gctggaaaac 4140

ggccggaaga gaatgctggc ctctgccggc gaactgcaga agggaaacga actggccctg 4200

ccctccaaat atgtgaactt cctgtacctg gccagccact atgagaagct gaagggctcc 4260
cccgaggata atgagcagaa acagctgttt gtggaacagc acaagcacta cctggacgag 4320
atcatcgagc agatcagcga gttctccaag agagtgatcc tggccgacgc taatctggac 4380
aaagtgctgt ccgcctacaa caagcaccgg gataagccca tcagagagca ggccgagaat 4440
atcatccacc tgtttaccct gaccaatctg ggagcccctg ccgccttcaa gtactttgac 4500
accaccatcg accggaagag gtacaccagc accaaagagg tgctggacgc caccctgatc 4560
caccagagca tcaccggcct gtacgagaca cggatcgacc tgtctcagct gggaggtgac 4620

<210> 210
<211> 21
<212> DNA
<213> Homo sapiens
<400> 210
gtagggcctt cgcgcacctc a 21
<210> 211
<211> 4263
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<400> 211
atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60
atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240
tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420
aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480
atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggctccg aagtcgagtt ttcccatgag 2400

tactggatga gacacgcatt gactctcgca aagagggctc gagatgaacg cgaggtgccc 2460

gtgggggcag tactcgtgct caacaatcgc gtaatcggcg aaggttggaa tagggcaatc 2520

ggactccacg accccactgc acatgcggaa atcatggccc ttcgacaggg agggcttgtg 2580

atgcagaatt atcgacttat cgatgcgacg ctgtacgtca cgtttgaacc ttgcgtaatg 2640

tgcgcgggag ctatgattca ctcccgcatt ggacgagttg tattcggtgt tcgcaacgcc 2700

aagacgggtg ccgcaggttc actgatggac gtgctgcatt acccaggcat gaaccaccgg 2760

gtagaaatca cagaaggcat attggcggac gaatgtgcgg cgctgttgtg ttacttttt 2820

cgcatgccca ggcaggtctt taacgcccag aaaaaagcac aatcctctac tgacggcggc 2880

ctgagcgaac tggataaggc cggcttcatc aagagacagc tggtggaaac ccggcagatc 2940

acaaagcacg tggcacagat cctggactcc cggatgaaca ctaagtacga cgagaatgac 3000

aagctgatcc gggaagtgaa agtgatcacc ctgaagtcca agctggtgtc cgatttccgg 3060

aaggatttcc agttttacaa agtgcgcgag atcaacaact accaccacgc ccacgacgcc 3120

tacctgaacg ccgtcgtggg aaccgccctg atcaaaaagt accctaagct ggaaagcgag 3180

ttcgtgtacg gcgactacaa ggtgtacgac gtgcggaaga tgatcgccaa gagcgagcag 3240

gaaatcggca aggctaccgc caagtacttc ttctacagca acatcatgaa ctttttcaag 3300

accgagatta ccctggccaa cggcgagatc cggaagcggc ctctgatcga gacaaacggc 3360

gaaaccgggg agatcgtgtg ggataagggc cgggattttg ccaccgtgcg gaaagtgctg 3420

agcatgcccc aagtgaatat cgtgaaaaag accgaggtgc agacaggcgg cttcagcaaa 3480

gagtctatcc tgcccaagag gaacagcgat aagctgatcg ccagaaagaa ggactgggac 3540

cctaagaagt acggcggctt cgacagcccc accgtggcct attctgtgct ggtggtggcc 3600

aaagtggaaa agggcaagtc caagaaactg aagagtgtga aagagctgct ggggatcacc 3660

atcatggaaa gaagcagctt cgagaagaat cccatcgact ttctggaagc caagggctac 3720

aaagaagtga aaaaggacct gatcatcaag ctgcctaagt actccctgtt cgagctggaa 3780

aacggccgga agagaatgct ggcctctgcc ggcgaactgc agaagggaaa cgaactggcc 3840

ctgccctcca aatatgtgaa cttcctgtac ctggccagcc actatgagaa gctgaagggc 3900

tcccccgagg ataatgagca gaaacagctg tttgtggaac agcacaagca ctacctggac 3960

gagatcatcg agcagatcag cgagttctcc aagagagtga tcctggccga cgctaatctg 4020

gacaaagtgc tgtccgccta caacaagcac cgggataagc ccatcagaga gcaggccgag 4080

aatatcatcc acctgtttac cctgaccaat ctgggagccc ctgccgcctt caagtacttt 4140

gacaccacca tcgaccggaa gaggtacacc agcaccaaag aggtgctgga cgccaccctg 4200

atccaccaga gcatcaccgg cctgtacgag acacggatcg acctgtctca gctgggaggt 4260

gac 4263

<210> 212

<211> 4656

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 212

atgtccgaag tcgagttttc ccatgagtac tggatgagac acgcattgac tctcgcaaag 60

agggctcgag atgaacgcga ggtgcccgtg ggggcagtac tcgtgctcaa caatcgcgta 120

atcggcgaag gttggaatag ggcaatcgga ctccacgacc ccactgcaca tgcggaaatc 180

atggcccttc gacagggagg gcttgtgatg cagaattatc gacttatcga tgcgacgctg 240

tacgtcacgt ttgaaccttg cgtaatgtgc gcgggagcta tgattcactc ccgcattgga 300

cgagttgtat tcggtgttcg caacgccaag acgggtgccg caggttcact gatggacgtg 360

ctgcattacc caggcatgaa ccaccgggta gaaatcacag aaggcatatt ggcggacgaa 420

tgtgcggcgc tgttgtgtta ctttttttcgc atgcccaggt ctggtggttc ttctggtggt 480

tctagcggca gcgagactcc cgggacctca gagtccgcca cacccgaaag ttctggtggt 540

tcttctggtg gttctgacaa gaagtacagc atcggcctgg ccatcggcac caactctgtg 600

ggctgggccg tgatcaccga cgagtacaag gtgcccagca agaaattcaa ggtgctgggc 660

aacaccgacc ggcacagcat caagaagaac ctgatcggag ccctgctgtt cgacagcggc 720

gaaacagccg aggccacccg gctgaagaga accgccagaa gaagatacac cagacggaag 780

aaccggatct gctatctgca agagatcttc agcaacgaga tggccaaggt ggacgacagc 840

ttcttccaca gactggaaga gtccttcctg gtggaagagg ataagaagca cgagcggcac 900

cccatcttcg gcaacatcgt ggacgaggtg gcctaccacg agaagtaccc caccatctac 960

cacctgagaa agaaactggt ggacagcacc gacaaggccg acctgcggct gatctatctg 1020

gccctggccc acatgatcaa gttccggggc cacttcctga tcgagggcga cctgaacccc 1080

gacaacagcg acgtggacaa gctgttcatc cagctggtgc agacctacaa ccagctgttc 1140

gaggaaaacc ccatcaacgc cagcggcgtg gacgccaagg ccatcctgtc tgccagactg 1200

agcaagagca gacggctgga aaatctgatc gcccagctgc ccggcgagaa gaagaatggc 1260

ctgttcggaa acctgattgc cctgagcctg ggcctgaccc ccaacttcaa gagcaacttc 1320

gacctggccg aggatgccaa actgcagctg agcaaggaca cctacgacga cgacctggac 1380

aacctgctgg cccagatcgg cgaccagtac gccgacctgt ttctggccgc caagaacctg 1440

tccgacgcca tcctgctgag cgacatcctg agagtgaaca ccgagatcac caaggccccc 1500

ctgagcgcct ctatgatcaa gagatacgac gagcaccacc aggacctgac cctgctgaaa 1560

gctctcgtgc ggcagcagct gcctgagaag tacaaagaga ttttcttcga ccagagcaag 1620

aacggctacg ccggctacat tgacggcgga gccagccagg aagagttcta caagttcatc 1680

aagcccatcc tggaaaagat ggacggcacc gaggaactgc tcgtgaagct gaacagagag 1740

gacctgctgc ggaagcagcg gaccttcgac aacggcagca tcccccacca gatccacctg 1800

ggagagctgc acgccattct gcggcggcag gaagattttt acccattcct gaaggacaac 1860

cgggaaaaga tcgagaagat cctgaccttc cgcatcccct actacgtggg ccctctggcc 1920

aggggaaaca gcagattcgc ctggatgacc agaaagagcg aggaaaccat cacccctgg 1980

aacttcgagg aagtggtgga caagggcgct tccgcccaga gcttcatcga gcggatgacc 2040

aacttcgata agaacctgcc caacgagaag gtgctgccca agcacagcct gctgtacgag 2100

tacttcaccg tgtataacga gctgaccaaa gtgaaatacg tgaccgaggg aatgagaaag 2160

cccgccttcc tgagcggcga gcagaaaaag gccatcgtgg acctgctgtt caagaccaac 2220

cggaaagtga ccgtgaagca gctgaaagag gactacttca agaaaatcga gtgcttcgac 2280

tccgtggaaa tctccggcgt ggaagatcgg ttcaacgcct ccctgggcac ataccacgat 2340

ctgctgaaaa ttatcaagga caaggacttc ctggacaatg aggaaaacga ggacattctg 2400

gaagatatcg tgctgaccct gacactgttt gaggacagag agatgatcga ggaacggctg 2460

aaaacctatg cccacctgtt cgacgacaaa gtgatgaagc agctgaagcg gcggagatac 2520

accggctggg gcaggctgag ccggaagctg atcaacggca tccgggacaa gcagtccggc 2580

aagacaatcc tggatttcct gaagtccgac ggcttcgcca acagaaactt catgcagctg 2640

atccacgacg acagcctgac ctttaaagag gacatccaga aagcccaggt gtccggccag 2700

ggcgatagcc tgcacgagca cattgccaat ctggccggca gccccgccat taagaagggc 2760

atcctgcaga cagtgaaggt ggtggacgag ctcgtgaaag tgatgggccg gcacaagccc 2820

gagaacatcg tgatcgaaat ggccagagag aaccagacca cccagaaggg acagaagaac 2880

agccgcgaga gaatgaagcg gatcgaagag ggcatcaaag agctgggcag ccagatcctg 2940

aaagaacacc ccgtggaaaa cacccagctg cagaacgaga agctgtacct gtactacctg 3000

cagaatgggc gggatatgta cgtggaccag gaactggaca tcaaccggct gtccgactac 3060

gatgtggacc atatcgtgcc tcagagcttt ctgaaggacg actccatcga caacaaggtg 3120

ctgaccagaa gcgacaagaa ccggggcaag agcgacaacg tgccctccga agaggtcgtg 3180

aagaagatga agaactactg gcggcagctg ctgaacgcca agctgattac ccagagaaag 3240

ttcgacaatc tgaccaaggc cgagagaggc ggcctgagcg aactggataa ggccggcttc 3300

atcaagagac agctggtgga aacccggcag atcacaaagc acgtggcaca gatcctggac 3360

tcccggatga acactaagta cgacgagaat gacaagctga tccgggaagt gaaagtgatc 3420

accctgaagt ccaagctggt gtccgatttc cggaaggatt tccagtttta caaagtgcgc 3480

gagatcaaca actaccacca cgcccacgac gcctacctga acgccgtcgt gggaaccgcc 3540

ctgatcaaaa agtaccctaa gctggaaagc gagttcgtgt acggcgacta caaggtgtac 3600

gacgtgcgga agatgatcgc caagagcgag caggaaatcg gcaaggctac cgccaagtac 3660

ttcttctaca gcaacatcat gaacttttтc aagaccgaga ttaccctggc caacggcgag 3720

atccggaagc ggcctctgat cgagacaaac ggcgaaaccg gggagatcgt gtgggataag 3780

ggccgggatt ttgccaccgt gcggaaagtg ctgagcatgc cccaagtgaa tatcgtgaaa 3840

aagaccgagg tgcagacagg cggcttcagc aaagagtcta tcctgcccaa gaggaacagc 3900

gataagctga tcgccagaaa gaaggactgg gaccctaaga agtacggcgg cttcgacagc 3960

cccaccgtgg cctattctgt gctggtggtg gccaaagtgg aaaagggcaa gtccaagaaa 4020

ctgaagagtg tgaaagagct gctggggatc accatcatgg aaagaagcag cttcgagaag 4080

aatcccatcg actttctgga agccaagggc tacaaagaag tgaaaaagga cctgatcatc 4140

aagctgccta agtactccct gttcgagctg gaaaacggcc ggaagagaat gctggcctct 4200

gccggcgaac tgcagaaggg aaacgaactg gccctgccct ccaaatatgt gaacttcctg 4260

tacctggcca gccactatga gaagctgaag ggctcccccg aggataatga gcagaaacag 4320

ctgtttgtgg aacagcacaa gcactacctg gacgagatca tcgagcagat cagcgagttc 4380

tccaagagag tgatcctggc cgacgctaat ctggacaaag tgctgtccgc ctacaacaag 4440

caccgggata agcccatcag agagcaggcc gagaatatca tccacctgtt taccctgacc 4500

aatctgggag cccctgccgc cttcaagtac tttgacacca ccatcgaccg gaagaggtac 4560

accagcacca aagaggtgct ggacgccacc ctgatccacc agagcatcac cggcctgtac 4620

gagacacgga tcgacctgtc tcagctggga ggtgac 4656

<210> 213

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 213

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220
gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340
atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400
gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460
gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580
gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640
aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760
ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820
actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880
aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940
taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000
taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060
atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

aacatcatga acttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg 3180
cctctgatcg agacaaacat gaaccaccgg gtagaaatca cagaaggcat attggcggac 3240
gaatgtgcgg cgctgttgtg ttacttttt cgcatgccca ggcaggtctt taacgcccag 3300
aaaaaagcac aatcctctac tgacggttct agcggcagcg agactcccgg gacctcagag 3360
tccgccacac ccgaaagttc tggttccgaa gtcgagtttt cccatgagta ctggatgaga 3420
cacgcattga ctctcgcaaa gagggctcga gatgaacgcg aggtgcccgt gggggcagta 3480
ctcgtgctca acaatcgcgt aatcggcgaa ggttggaata gggcaatcgg actccacgac 3540

cccactgcac atgcggaaat catggccctt cgacagggag ggcttgtgat gcagaattat 3600

cgacttatcg atgcgacgct gtacgtcacg tttgaacctt gcgtaatgtg cgcgggagct 3660

atgattcact cccgcattgg acgagttgta ttcggtgttc gcaacgccaa gacgggtgcc 3720

gcaggttcac tgatggacgt gctgcattac ccaggcggcg aaaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 214

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 214

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480
atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660
cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900
ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960
atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140
gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200
aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320
gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380
agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440
gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500
aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560
tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620
agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800
atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860
ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

aacatcatga acttttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg 3180

cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt 3240

gccaccgtgc ggaaagtgct gagcatgccc caagtgaata tcgtgaaaaa gaccgaggtg 3300

cagacaggcg gcttcagcaa agagtctatc ctgcccaaga ggaacagcga taagctgatc 3360

gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc 3420

tattctgtgc tggtggtggc caaagtggaa aagggcaagt ccaagaaact gaagagtgtg 3480

aaagagctgc tggggatcac catcatggaa agaagcagct tcgagaagaa tcccatcgac 3540

tttctggaag ccaagggcta caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag 3600

tactccctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg 3660

cagaagggaa acgaactggc cctgccctcc aaatatgtga acttcctgta cctggccagc 3720

cactatgaga agctgaaggg ctccatgaac caccgggtag aaatcacaga aggcatattg 3780

gcggacgaat gtgcggcgct gttgtgttac ttttttcgca tgcccaggca ggtctttaac 3840

gcccagaaaa aagcacaatc ctctactgac ggttctagcg gcagcgagac tcccgggacc 3900

tcagagtccg ccacacccga aagttctggt tccgaagtcg agttttccca tgagtactgg 3960

atgagacacg cattgactct cgcaaagagg gctcgagatg aacgcgaggt gcccgtgggg 4020

gcagtactcg tgctcaacaa tcgcgtaatc ggcgaaggtt ggaatagggc aatcggactc 4080

cacgacccca ctgcacatgc ggaaatcatg gcccttcgac agggagggct tgtgatgcag 4140

aattatcgac ttatcgatgc gacgctgtac gtcacgtttg aaccttgcgt aatgtgcgcg 4200

ggagctatga ttcactcccg cattggacga gttgtattcg gtgttcgcaa cgccaagacg 4260

ggtgccgcag gttcactgat ggacgtgctg cattacccag gccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 215

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 215

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

aacatcatga acttttttcaa gaccgagatt accctggcca tgaaccaccg ggtagaaatc 3180

acagaaggca tattggcgga cgaatgtgcg gcgctgttgt gttacttttt tcgcatgccc 3240

aggcaggtct ttaacgccca gaaaaaagca caatcctcta ctgacggttc tagcggcagc 3300

gagactcccg ggacctcaga gtccgccaca cccgaaagtt ctggttccga agtcgagttt 3360

tcccatgagt actggatgag acacgcattg actctcgcaa agagggctcg agatgaacgc 3420

gaggtgcccg tgggggcagt actcgtgctc aacaatcgcg taatcggcga aggttggaat 3480

agggcaatcg gactccacga ccccactgca catgcggaaa tcatggccct tcgacaggga 3540

gggcttgtga tgcagaatta tcgacttatc gatgcgacgc tgtacgtcac gtttgaacct 3600

tgcgtaatgt gcgcgggagc tatgattcac tcccgcattg gacgagttgt attcggtgtt 3660

cgcaacgcca agacgggtgc cgcaggttca ctgatggacg tgctgcatta cccaggcaac 3720

ggcgagatcc ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg 3780

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3840

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3900

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3960

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 4020

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 4080

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 4140

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 4200

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 4260

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500
ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560
aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620
ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 216
<211> 4362
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide
<400> 216

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240
tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420
aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660
cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960
atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggctccg aagtcgagtt ttcccatgag 2400

tactggatga gacacgcatt gactctcgca aagagggctc gagatgaacg cgaggtgccc 2460

gtgggggcag tactcgtgct caacaatcgc gtaatcggcg aaggttggaa tagggcaatc 2520

ggactccacg accccactgc acatgcggaa atcatggccc ttcgacaggg agggcttgtg 2580

atgcagaatt atcgacttat cgatgcgacg ctgtacgtca cgtttgaacc ttgcgtaatg 2640

tgcgcgggag ctatgattca ctcccgcatt ggacgagttg tattcggtgt tcgcaacgcc 2700

aagacgggtg ccgcaggttc actgatggac gtgctgcatt acccaggcat gaaccaccgg 2760

gtagaaatca cagaaggcat attggcggac gaatgtgcgg cgctgttgtg ttacttttt 2820

cgcatgccca ggcaggtctt taacgcccag aaaaaagcac aatcctctac tgacgaagag 2880

gtcgtgaaga agatgaagaa ctactggcgg cagctgctga acgccaagct gattacccag 2940

agaaagttcg acaatctgac caaggccgag agaggcggcc tgagcgaact ggataaggcc 3000

ggcttcatca agagacagct ggtggaaacc cggcagatca caaagcacgt ggcacagatc 3060

ctggactccc ggatgaacac taagtacgac gagaatgaca agctgatccg ggaagtgaaa 3120

gtgatcaccc tgaagtccaa gctggtgtcc gatttccgga aggatttcca gttttacaaa 3180

gtgcgcgaga tcaacaacta ccaccacgcc cacgacgcct acctgaacgc cgtcgtggga 3240

accgccctga tcaaaaagta ccctaagctg gaaagcgagt tcgtgtacgg cgactacaag 3300

gtgtacgacg tgcggaagat gatcgccaag agcgagcagg aaatcggcaa ggctaccgcc 3360

aagtacttct tctacagcaa catcatgaac tttttcaaga ccgagattac cctggccaac 3420

ggcgagatcc ggaagcggcc tctgatcgag acaaacggcg aaaccgggga gatcgtgtgg 3480

gataagggcc gggattttgc caccgtgcgg aaagtgctga gcatgcccca agtgaatatc 3540

gtgaaaaaga ccgaggtgca gacaggcggc ttcagcaaag agtctatcct gcccaagagg 3600

aacagcgata agctgatcgc cagaaagaag gactgggacc ctaagaagta cggcggcttc 3660

gacagcccca ccgtggccta ttctgtgctg gtggtggcca aagtggaaaa gggcaagtcc 3720

aagaaactga agagtgtgaa agagctgctg gggatcacca tcatggaaag aagcagcttc 3780

gagaagaatc ccatcgactt tctggaagcc aagggctaca aagaagtgaa aaaggacctg 3840

atcatcaagc tgcctaagta ctccctgttc gagctggaaa acggccggaa gagaatgctg 3900

gcctctgccg gcgaactgca gaagggaaac gaactggccc tgccctccaa atatgtgaac 3960

ttcctgtacc tggccagcca ctatgagaag ctgaagggct cccccgagga taatgagcag 4020

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4080

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4140

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4200

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4260

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4320

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4362

<210> 217

<211> 4260

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 217

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220
gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340
atgaagcgga tcgaagaggg catcaaagag ctgggctccg aagtcgagtt ttcccatgag 2400
tactggatga gacacgcatt gactctcgca aagagggctc gagatgaacg cgaggtgccc 2460
gtgggggcag tactcgtgct caacaatcgc gtaatcggcg aaggttggaa tagggcaatc 2520
ggactccacg accccactgc acatgcggaa atcatggccc ttcgacaggg agggcttgtg 2580
atgcagaatt atcgacttat cgatgcgacg ctgtacgtca cgtttgaacc ttgcgtaatg 2640
tgcgcgggag ctatgattca ctcccgcatt ggacgagttg tattcggtgt tcgcaacgcc 2700

aagacgggtg ccgcaggttc actgatggac gtgctgcatt acccaggcat gaaccaccgg 2760
gtagaaatca cagaaggcat attggcggac gaatgtgcgg cgctgttgtg ttacttttt 2820
cgcatgccca ggcaggtctt taacgcccag aaaaaagcac aatcctctac tgacggcctg 2880
agcgaactgg ataaggccgg cttcatcaag agacagctgg tggaaacccg gcagatcaca 2940
aagcacgtgg cacagatcct ggactcccgg atgaacacta agtacgacga gaatgacaag 3000
ctgatccggg aagtgaaagt gatcaccctg aagtccaagc tggtgtccga tttccggaag 3060
gatttccagt tttacaaagt gcgcgagatc aacaactacc accacgccca cgacgcctac 3120

ctgaacgccg tcgtgggaac cgccctgatc aaaaagtacc ctaagctgga aagcgagttc 3180
gtgtacggcg actacaaggt gtacgacgtg cggaagatga tcgccaagag cgagcaggaa 3240
atcggcaagg ctaccgccaa gtacttcttc tacagcaaca tcatgaactt tttcaagacc 3300
gagattaccc tggccaacgg cgagatccgg aagcggcctc tgatcgagac aaacggcgaa 3360
accggggaga tcgtgtggga taagggccgg gattttgcca ccgtgcggaa agtgctgagc 3420
atgccccaag tgaatatcgt gaaaaagacc gaggtgcaga caggcggctt cagcaaagag 3480
tctatcctgc ccaagaggaa cagcgataag ctgatcgcca gaaagaagga ctgggaccct 3540

aagaagtacg gcggcttcga cagccccacc gtggcctatt ctgtgctggt ggtggccaaa 3600

gtggaaaagg gcaagtccaa gaaactgaag agtgtgaaag agctgctggg gatcaccatc 3660

atggaaagaa gcagcttcga gaagaatccc atcgactttc tggaagccaa gggctacaaa 3720

gaagtgaaaa aggacctgat catcaagctg cctaagtact ccctgttcga gctggaaaac 3780

ggccggaaga gaatgctggc ctctgccggc gaactgcaga agggaaacga actggccctg 3840

ccctccaaat atgtgaactt cctgtacctg gccagccact atgagaagct gaagggctcc 3900

cccgaggata atgagcagaa acagctgttt gtggaacagc acaagcacta cctggacgag 3960

atcatcgagc agatcagcga gttctccaag agagtgatcc tggccgacgc taatctggac 4020

aaagtgctgt ccgcctacaa caagcaccgg gataagccca tcagagagca ggccgagaat 4080

atcatccacc tgtttaccct gaccaatctg ggagcccctg ccgccttcaa gtactttgac 4140

accaccatcg accggaagag gtacaccagc accaaagagg tgctggacgc caccctgatc 4200

caccagagca tcaccggcct gtacgagaca cggatcgacc tgtctcagct gggaggtgac 4260

<210> 218

<211> 4662

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 218

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

aacatcatga acttttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg 3180

cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt 3240

gccaccgtgc ggaaagtgct gagcatgccc caagtgaata tcgtgaaaaa gaccgaggtg 3300

cagacaggcg gcttcagcaa agagtctatc ctgcccaaga ggaacagcga taagctgatc 3360

gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc 3420

tattctgtgc tggtggtggc caaagtggaa aagggcaagt ccaagaaact gaagagtgtg 3480

aaagagctgc tggggatcac catcatggaa agaagcagct tcgagaagaa tcccatcgac 3540

tttctggaag ccaagggcta caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag 3600

tactccctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg 3660

cagaagggaa acgaactggc cctgccctcc aaatatgtga acttcctgta cctggccagc 3720

cactatgaga agctgaaggg cactgcacat gcggaaatca tggcccttcg acagggaggg 3780

cttgtgatgc agaattatcg acttatcgat gcgacgctgt acgtcacgtt tgaaccttgc 3840

gtaatgtgcg cgggagctat gattcactcc cgcattggac gagttgtatt cggtgttcgc 3900

aacgccaaga cgggtgccgc aggttcactg atggacgtgc tgcattaccc aggcatgaac 3960

caccgggtag aaatcacaga aggcatattg gcggacgaat gtgcggcgct gttgtgttac 4020

ttttttcgca tgcccaggca ggtctttaac gcccagaaaa aagcacaatc ctctactgac 4080

ggttctagcg gcagcgagac tcccgggacc tcagagtccg ccacacccga aagttctggt 4140

tccgaagtcg agttttccca tgagtactgg atgagacacg cattgactct cgcaaagagg 4200

gctcgagatg aacgcgaggt gcccgtgggg gcagtactcg tgctcaacaa tcgcgtaatc 4260

ggcgaaggtt ggaatagggc aatcggactc cacgacccct cccccgagga taatgagcag 4320

aaacagctgt ttgtggaaca gcacaagcac tacctggacg agatcatcga gcagatcagc 4380

gagttctcca agagagtgat cctggccgac gctaatctgg acaaagtgct gtccgcctac 4440

aacaagcacc gggataagcc catcagagag caggccgaga atatcatcca cctgtttacc 4500

ctgaccaatc tgggagcccc tgccgccttc aagtactttg acaccaccat cgaccggaag 4560

aggtacacca gcaccaaaga ggtgctggac gccaccctga tccaccagag catcaccggc 4620

ctgtacgaga cacggatcga cctgtctcag ctgggaggtg ac 4662

<210> 219

<211> 4623

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 219

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggttcta gcggcagcga gactcccggg 2400

acctcagagt ccgccacacc cgaaagttct ggttccgaag tcgagttttc ccatgagtac 2460

tggatgagac acgcattgac tctcgcaaag agggctcgag atgaacgcga ggtgcccgtg 2520

ggggcagtac tcgtgctcaa caatcgcgta atcggcgaag gttggaatag ggcaatcgga 2580

ctccacgacc ccactgcaca tgcggaaatc atggcccttc gacagggagg gcttgtgatg 2640

cagaattatc gacttatcga tgcgacgctg tacgtcacgt ttgaaccttg cgtaatgtgc 2700

gcgggagcta tgattcactc ccgcattgga cgagttgtat tcggtgttcg caacgccaag 2760

acgggtgccg caggttcact gatggacgtg ctgcattacc caggcatgaa ccaccgggta 2820

gaaatcacag aaggcatatt ggcggacgaa tgtgcggcgc tgttgtgtta ctttttcgc 2880

atgcccaggc agggcagcca gatcctgaaa gaacaccccg tggaaaacac ccagctgcag 2940

aacgagaagc tgtacctgta ctacctgcag aatgggcggg atatgtacgt ggaccaggaa 3000

ctggacatca accggctgtc cgactacgat gtggaccata tcgtgcctca gagctttctg 3060

aaggacgact ccatcgacaa caaggtgctg accagaagcg acaagaaccg gggcaagagc 3120

gacaacgtgc cctccgaaga ggtcgtgaag aagatgaaga actactggcg gcagctgctg 3180

aacgccaagc tgattaccca gagaaagttc gacaatctga ccaaggccga gaggcggc 3240

ctgagcgaac tggataaggc cggcttcatc aagagacagc tggtggaaac ccggcagatc 3300

acaaagcacg tggcacagat cctggactcc cggatgaaca ctaagtacga cgagaatgac 3360

aagctgatcc gggaagtgaa agtgatcacc ctgaagtcca agctggtgtc cgatttccgg 3420

aaggatttcc agttttacaa agtgcgcgag atcaacaact accaccacgc ccacgacgcc 3480

tacctgaacg ccgtcgtggg aaccgccctg atcaaaaagt accctaagct ggaaagcgag 3540

ttcgtgtacg gcgactacaa ggtgtacgac gtgcggaaga tgatcgccaa gagcgagcag 3600

gaaatcggca aggctaccgc caagtacttc ttctacagca acatcatgaa ctttttcaag 3660

accgagatta ccctggccaa cggcgagatc cggaagcggc ctctgatcga gacaaacggc 3720

gaaaccgggg agatcgtgtg ggataagggc cgggattttg ccaccgtgcg gaaagtgctg 3780

agcatgcccc aagtgaatat cgtgaaaaag accgaggtgc agacaggcgg cttcagcaaa 3840

gagtctatcc tgcccaagag gaacagcgat aagctgatcg ccagaaagaa ggactgggac 3900

cctaagaagt acggcggctt cgacagcccc accgtggcct attctgtgct ggtggtggcc 3960

aaagtggaaa agggcaagtc caagaaactg aagagtgtga aagagctgct ggggatcacc 4020

atcatggaaa gaagcagctt cgagaagaat cccatcgact ttctggaagc caagggctac 4080

aaagaagtga aaaaggacct gatcatcaag ctgcctaagt actccctgtt cgagctggaa 4140

aacggccgga agagaatgct ggcctctgcc ggcgaactgc agaagggaaa cgaactggcc 4200

ctgccctcca aatatgtgaa cttcctgtac ctggccagcc actatgagaa gctgaagggc 4260

tcccccgagg ataatgagca gaaacagctg tttgtggaac agcacaagca ctacctggac 4320

gagatcatcg agcagatcag cgagttctcc aagagagtga tcctggccga cgctaatctg 4380

gacaaagtgc tgtccgccta caacaagcac cgggataagc ccatcagaga gcaggccgag 4440

aatatcatcc acctgtttac cctgaccaat ctgggagccc ctgccgcctt caagtacttt 4500

gacaccacca tcgaccggaa gaggtacacc agcaccaaag aggtgctgga cgccaccctg 4560

atccaccaga gcatcaccgg cctgtacgag acacggatcg acctgtctca gctgggaggt 4620

gac 4623

<210> 220

<211> 1541

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 220

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Ser Gly Ser Glu Thr Pro Gly

785 790 795 800

Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe

805 810 815

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala

820 825 830

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn

835 840 845

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro

850 855 860

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

865 870 875 880

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro

885 890 895

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val

900 905 910

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met

915 920 925

Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu

930 935 940

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg

945 950 955 960

Met Pro Arg Gln Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn

965 970 975

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly

980 985 990

Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp

995 1000 1005

Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp

1010 1015 1020

Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

1025 1030 1035

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

1040 1045 1050

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg

1055 1060 1065

Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu

1070 1075 1080

Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg

1085 1090 1095

Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn

1100 1105 1110

Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val

1115 1120 1125

Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe

1130 1135 1140

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His

1145 1150 1155

Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys

1160 1165 1170

Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val

1175 1180 1185

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly

1190 1195 1200

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe

1205 1210 1215

Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg

1220 1225 1230

Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp

1235 1240 1245

Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro

1250 1255 1260

Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe

1265 1270 1275

Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1280 1285 1290

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp

1295 1300 1305

Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu

1310 1315 1320

Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly

1325 1330 1335

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp

1340 1345 1350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile

1355 1360 1365

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg

1370 1375 1380

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu

1385 1390 1395

Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser

1400 1405 1410

His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys

1415 1420 1425

Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile

1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala

1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys

1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu

1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr

1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala

1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp

1535 1540

<210> 221

<211> 8913

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 221

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660
cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900
ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140
gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200
aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320
gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440
gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500
aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560
tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620
agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680
gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860
ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacgg aggctctgga 3060

ggaagcagct ctgagacagg acctgtggcc gtggatccca cactgcggag aagaattgag 3120

ccccacgagt tcgaggtgtt cttcgacccc agagagctgc ggaaagagac atgcctgctg 3180

tacgagatca actggggcgg cagacactct atctggcggc acacaagcca gaacaccaac 3240

aagcacgtgg aagtgaactt tatcgagaag tttacgaccg agcggtactt ctgccccaac 3300

accagatgca gcatcacctg gtttctgagc tggtcccctt gcggcgagtg cagcagagcc 3360

atcaccgagt ttctgtccag atatccccac gtgaccctgt tcatctatat cgcccggctg 3420

taccaccacg ccgatcctag aaatagacag ggcctgcgcg acctgatcag cagcggagtg 3480

acaatccaga tcatgaccga gcaagagagc ggctactgct ggcggaactt cgtgaactac 3540

agccccagca acgaagccca ctggcctaga tatcctcacc tgtgggtccg actgtacgtg 3600

ctggaactgt actgcatcat cctgggcctg cctccatgcc tgaacatcct gagaagaaag 3660

cagcctcagc tgaccttctt cacaatcgcc ctgcagagct gccactacca gagactgcct 3720

ccacacatcc tgtgggccac cggacttaag ggctcttctg gatctgaaac acctggcaca 3780

agtgagagcg ccacccctga gagctctggc gacgtgcgga agatgatcgc caagagcgag 3840

caggaaatcg gcaaggctac cgccaagtac ttcttctaca gcaacatcat gaacttttc 3900

aagaccgaga ttaccctggc caacggcgag atccggaagc ggcctctgat cgagacaaac 3960

ggcgaaaccg gggagatcgt gtgggataag ggccgggatt ttgccaccgt gcggaaagtg 4020

ctgagcatgc cccaagtgaa tatcgtgaaa aagaccgagg tgcagacagg cggcttcagc 4080

aaagagtcta tcctgcccaa gaggaacagc gataagctga tcgccagaaa gaaggactgg 4140

gaccctaaga agtacggcgg cttcgacagc cccaccgtgg cctattctgt gctggtggtg 4200

gccaaagtgg aaaagggcaa gtccaagaaa ctgaagagtg tgaaagagct gctggggatc 4260

accatcatgg aaagaagcag cttcgagaag aatcccatcg actttctgga agccaagggc 4320

tacaaagaag tgaaaaagga cctgatcatc aagctgccta agtactccct gttcgagctg 4380

gaaaacggcc ggaagagaat gctggcctct gccggcgaac tgcagaaggg aaacgaactg 4440

gccctgccct ccaaatatgt gaacttcctg tacctggcca gccactatga gaagctgaag 4500

ggctcccccg aggataatga gcagaaacag ctgtttgtgg aacagcacaa gcactacctg 4560

gacgagatca tcgagcagat cagcgagttc tccaagagag tgatcctggc cgacgctaat 4620

ctggacaaag tgctgtccgc ctacaacaag caccgggata agcccatcag agagcaggcc 4680

gagaatatca tccacctgtt taccctgacc aatctgggag cccctgccgc cttcaagtac 4740

tttgacacca ccatcgaccg gaagaggtac accagcacca aagaggtgct ggacgccacc 4800

ctgatccacc agagcatcac cggcctgtac gagacacgga tcgacctgtc tcagctggga 4860

ggtgactctg gtggaagcgg aggatctggc ggcagcacca atctgagcga catcatcgag 4920

aaagagacag gcaagcagct ggtcatccaa gagtccatcc tgatgctgcc tgaagaggtg 4980

gaagaagtga tcggcaacaa gcccgagtcc gacatcctgg tgcacaccgc ctacgatgag 5040

agcaccgacg agaacgtgat gctgctgacc tctgacgccc ctgagtacaa gccttgggct 5100

ctcgtgatcc aggacagcaa cggcgagaac aagatcaaga tgctgagcgg cggctctggt 5160

ggctctggcg gatctacaaa cctgtccgat attattgaga aagaaaccgg gaaacagctc 5220

gtgattcaag agtctattct catgctcccg gaagaagtcg aggaagtcat tggaaacaag 5280

cctgagagcg atattctggt ccatacagcc tacgacgagt ctaccgatga gaatgtcatg 5340

ctcctcacca gcgacgctcc cgagtataag ccatgggcac ttgtcattca ggactccaat 5400

ggggaaaaca aaatcaaaat gctcccaaag aaaaaacgca aggtggaggg agctgataag 5460

cgcaccgccg atggttccga gttcgaaagc cccaagaaga agaggaaagt ctaaccggtc 5520

atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 5580

gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 5640

ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 5700

ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 5760

ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 5820

cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 5880

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 5940

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 6000

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 6060
gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 6120
ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 6180
caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 6240
aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 6300
atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 6360
cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 6420

ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 6480
gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg 6540
accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 6600
cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 6660
cagagttctt gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct 6720
gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6780
aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 6840

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6900
actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6960
taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 7020
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 7080
tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 7140
ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 7200
accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 7260

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 7320
acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 7380
tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 7440
cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 7500
tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 7560
ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 7620
gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 7680

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 7740

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 7800

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaagga ataagggcga 7860

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 7920

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 7980

ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 8040

ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 8100

atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 8160

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 8220

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 8280

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 8340

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 8400

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 8460

agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 8520

cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 8580

tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 8640

tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 8700

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 8760

caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 8820

aggtctatat aagcagagct ggtttagtga accgtcagat ccgctagaga tccgcggccg 8880

ctaatacgac tcactatagg gagagccgcc acc 8913

<210> 222

<211> 1837

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 222

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 Gly Gly Ser Gly Gly Ser Ser
1010 1015 1020

Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg Arg
1025 1030 1035
Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
1040 1045 1050
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg
1055 1060 1065

His Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val
1070 1075 1080
Glu Val Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys
1085 1090 1095
Pro Asn Thr Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro
1100 1105 1110
Cys Gly Glu Cys Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr
1115 1120 1125

Pro His Val Thr Leu Phe Ile Tyr Ile Ala Arg Leu Tyr His His
1130 1135 1140
Ala Asp Pro Arg Asn Arg Gln Gly Leu Arg Asp Leu Ile Ser Ser
1145 1150 1155
Gly Val Thr Ile Gln Ile Met Thr Glu Gln Glu Ser Gly Tyr Cys
1160 1165 1170
Trp Arg Asn Phe Val Asn Tyr Ser Pro Ser Asn Glu Ala His Trp
1175 1180 1185

Pro Arg Tyr Pro His Leu Trp Val Arg Leu Tyr Val Leu Glu Leu
1190 1195 1200
Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg
1205 1210 1215
Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile Ala Leu Gln Ser
1220 1225 1230
Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp Ala Thr Gly
1235 1240 1245

Leu Lys Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

1250 1255 1260

Ala Thr Pro Glu Ser Ser Gly Asp Val Arg Lys Met Ile Ala Lys

1265 1270 1275

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1280 1285 1290

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1295 1300 1305

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1310 1315 1320

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1325 1330 1335

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1340 1345 1350

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1355 1360 1365

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1370 1375 1380

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1385 1390 1395

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1400 1405 1410

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1415 1420 1425

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1430 1435 1440

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1445 1450 1455

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1460 1465 1470

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1475 1480 1485

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1490 1495 1500
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1505 1510 1515
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1520 1525 1530
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1535 1540 1545

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1550 1555 1560
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1565 1570 1575
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1580 1585 1590
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1595 1600 1605

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser
1610 1615 1620
Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile
1625 1630 1635
Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile
1640 1645 1650
Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro
1655 1660 1665

Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp
1670 1675 1680
Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro
1685 1690 1695
Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys
1700 1705 1710
Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu

1715 1720 1725

Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln
1730 1735 1740
Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly
1745 1750 1755
Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu
1760 1765 1770
Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu
1775 1780 1785

Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn
1790 1795 1800
Lys Ile Lys Met Leu Pro Lys Lys Lys Arg Lys Val Glu Gly Ala
1805 1810 1815
Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys
1820 1825 1830
Lys Arg Lys Val
1835

<210> 223
<211> 8913
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 223
atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60
atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240
tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgccg gaggctctgg aggaagcagc tctgagacag gacctgtggc cgtggatccc 3120

acactgcgga gaagaattga gccccacgag ttcgaggtgt tcttcgaccc cagagagctg 3180

cggaaagaga catgcctgct gtacgagatc aactggggcg gcagacactc tatctggcgg 3240

cacacaagcc agaacaccaa caagcacgtg gaagtgaact ttatcgagaa gtttacgacc 3300

gagcggtact tctgccccaa caccagatgc agcatcacct ggtttctgag ctggtcccct 3360

tgcggcgagt gcagcagagc catcaccgag tttctgtcca gatatcccca cgtgaccctg 3420

ttcatctata tcgcccggct gtaccaccac gccgatccta gaaatagaca gggcctgcgc 3480

gacctgatca gcagcggagt gacaatccag atcatgaccg agcaagagag cggctactgc 3540

tggcggaact tcgtgaacta cagccccagc aacgaagccc actggcctag atatcctcac 3600

ctgtgggtcc gactgtacgt gctggaactg tactgcatca tcctgggcct gcctccatgc 3660

ctgaacatcc tgagaagaaa gcagcctcag ctgaccttct tcacaatcgc cctgcagagc 3720

tgccactacc agagactgcc tccacacatc ctgtgggcca ccggacttaa gggctcttct 3780

ggatctgaaa cacctggcac aagtgagagc gccacccctg agagctctgg caagagcgag 3840

caggaaatcg gcaaggctac cgccaagtac ttcttctaca gcaacatcat gaacttttтc 3900

aagaccgaga ttaccctggc caacggcgag atccggaagc ggcctctgat cgagacaaac 3960

ggcgaaaccg gggagatcgt gtgggataag ggccgggatt ttgccaccgt gcggaaagtg 4020

ctgagcatgc cccaagtgaa tatcgtgaaa aagaccgagg tgcagacagg cggcttcagc 4080

aaagagtcta tcctgcccaa gaggaacagc gataagctga tcgccagaaa gaaggactgg 4140

gaccctaaga agtacggcgg cttcgacagc cccaccgtgg cctattctgt gctggtggtg 4200

gccaaagtgg aaaagggcaa gtccaagaaa ctgaagagtg tgaaagagct gctggggatc 4260
accatcatgg aaagaagcag cttcgagaag aatcccatcg actttctgga agccaagggc 4320
tacaaagaag tgaaaaagga cctgatcatc aagctgccta agtactccct gttcgagctg 4380
gaaaacggcc ggaagagaat gctggcctct gccggcgaac tgcagaaggg aaacgaactg 4440
gccctgccct ccaaatatgt gaacttcctg tacctggcca gccactatga gaagctgaag 4500
ggctcccccg aggataatga gcagaaacag ctgtttgtgg aacagcacaa gcactacctg 4560
gacgagatca tcgagcagat cagcgagttc tccaagagag tgatcctggc cgacgctaat 4620

ctggacaaag tgctgtccgc ctacaacaag caccgggata agcccatcag agagcaggcc 4680
gagaatatca tccacctgtt taccctgacc aatctgggag cccctgccgc cttcaagtac 4740
tttgacacca ccatcgaccg gaagaggtac accagcacca aagaggtgct ggacgccacc 4800
ctgatccacc agagcatcac cggcctgtac gagacacgga tcgacctgtc tcagctggga 4860
ggtgactctg gtggaagcgg aggatctggc ggcagcacca atctgagcga catcatcgag 4920
aaagagacag gcaagcagct ggtcatccaa gagtccatcc tgatgctgcc tgaagaggtg 4980
gaagaagtga tcggcaacaa gcccgagtcc gacatcctgg tgcacaccgc ctacgatgag 5040

agcaccgacg agaacgtgat gctgctgacc tctgacgccc ctgagtacaa gccttgggct 5100
ctcgtgatcc aggacagcaa cggcgagaac aagatcaaga tgctgagcgg cggctctggt 5160
ggctctggcg gatctacaaa cctgtccgat attattgaga aagaaaccgg gaaacagctc 5220
gtgattcaag agtctattct catgctcccg gaagaagtcg aggaagtcat tggaaacaag 5280
cctgagagcg atattctggt ccatacagcc tacgacgagt ctaccgatga gaatgtcatg 5340
ctcctcacca gcgacgctcc cgagtataag ccatgggcac ttgtcattca ggactccaat 5400
ggggaaaaca aaatcaaaat gctcccaaag aaaaaacgca aggtggaggg agctgataag 5460

cgcaccgccg atggttccga gttcgaaagc cccaagaaga agaggaaagt ctaaccggtc 5520
atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 5580
gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 5640
ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 5700
ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 5760
ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 5820
cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 5880

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 5940

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 6000

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 6060

gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 6120

ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 6180

caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 6240

aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 6300

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 6360

cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 6420

ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 6480

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccccc gttcagcccg 6540

accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 6600

cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 6660

cagagttctt gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct 6720

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6780

aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 6840

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6900

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6960

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 7020

gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 7080

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 7140

ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 7200

accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 7260

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 7320

acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 7380

tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 7440

cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 7500

tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 7560

ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 7620

gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 7680

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 7740

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 7800

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaagga ataagggcga 7860

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 7920

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 7980

ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 8040

ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 8100

atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 8160

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 8220

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 8280

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 8340

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 8400

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 8460

agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 8520

cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 8580

tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 8640

tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 8700

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 8760

caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 8820

aggtctatat aagcagagct ggtttagtga accgtcagat ccgctagaga tccgcggccg 8880

ctaatacgac tcactatagg gagagccgcc acc 8913

<210> 224

<211> 1837

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 224

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

Gly Gly Ser Gly Gly Ser Ser Ser Glu Thr Gly Pro Val Ala Val
1025 1030 1035
Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu Phe Glu Val
1040 1045 1050

Phe Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu Tyr
1055 1060 1065
Glu Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg His Thr Ser
1070 1075 1080
Gln Asn Thr Asn Lys His Val Glu Val Asn Phe Ile Glu Lys Phe
1085 1090 1095
Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr
1100 1105 1110

Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser Arg Ala Ile
1115 1120 1125
Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu Phe Ile Tyr
1130 1135 1140
Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg Gln Gly
1145 1150 1155
Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met Thr
1160 1165 1170

Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
1175 1180 1185
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val
1190 1195 1200
Arg Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro
1205 1210 1215
Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe
1220 1225 1230

Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro
1235 1240 1245
His Ile Leu Trp Ala Thr Gly Leu Lys Gly Ser Ser Gly Ser Glu

1250 1255 1260

Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Lys

1265 1270 1275

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1280 1285 1290

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1295 1300 1305

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1310 1315 1320

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1325 1330 1335

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1340 1345 1350

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1355 1360 1365

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1370 1375 1380

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1385 1390 1395

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1400 1405 1410

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1415 1420 1425

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1430 1435 1440

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1445 1450 1455

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1460 1465 1470

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1475 1480 1485

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1490 1495 1500
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1505 1510 1515
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1520 1525 1530

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1535 1540 1545
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1550 1555 1560
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1565 1570 1575
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1580 1585 1590

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1595 1600 1605
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser
1610 1615 1620
Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile
1625 1630 1635
Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile
1640 1645 1650

Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro
1655 1660 1665
Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp
1670 1675 1680
Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro
1685 1690 1695
Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys
1700 1705 1710

Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu

1715 1720 1725

Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln

1730 1735 1740

Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly

1745 1750 1755

Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu

1760 1765 1770

Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu

1775 1780 1785

Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn

1790 1795 1800

Lys Ile Lys Met Leu Pro Lys Lys Lys Arg Lys Val Glu Gly Ala

1805 1810 1815

Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys

1820 1825 1830

Lys Arg Lys Val

1835

<210> 225

<211> 8913

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 225

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460
gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580
gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640
aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760
ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880
aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940
taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000
taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060
atgatcgcca agagcgagca ggaaatcgga ggctctggag gaagcagctc tgagacagga 3120
cctgtggccg tggatcccac actgcggaga agaattgagc cccacgagtt cgaggtgttc 3180
ttcgacccca gagagctgcg gaaagagaca tgcctgctgt acgagatcaa ctggggcggc 3240

agacactcta tctggcggca cacaagccag aacaccaaca agcacgtgga agtgaacttt 3300
atcgagaagt ttacgaccga gcggtacttc tgccccaaca ccagatgcag catcacctgg 3360
tttctgagct ggtccccttg cggcgagtgc agcagagcca tcaccgagtt tctgtccaga 3420
tatccccacg tgaccctgtt catctatatc gcccggctgt accaccacgc cgatcctaga 3480
aatagacagg gcctgcgcga cctgatcagc agcggagtga caatccagat catgaccgag 3540
caagagagcg gctactgctg gcggaacttc gtgaactaca gccccagcaa cgaagcccac 3600
tggcctagat atcctcacct gtgggtccga ctgtacgtgc tggaactgta ctgcatcatc 3660

ctgggcctgc ctccatgcct gaacatcctg agaagaaagc agcctcagct gaccttcttc 3720
acaatcgccc tgcagagctg ccactaccag agactgcctc cacacatcct gtgggccacc 3780
ggacttaagg gctcttctgg atctgaaaca cctggcacaa gtgagagcgc caccctgag 3840
agctctggcg gcaaggctac cgccaagtac ttcttctaca gcaacatcat gaacttttc 3900
aagaccgaga ttaccctggc caacggcgag atccggaagc ggcctctgat cgagacaaac 3960
ggcgaaaccg gggagatcgt gtgggataag ggccgggatt ttgccaccgt gcggaaagtg 4020
ctgagcatgc cccaagtgaa tatcgtgaaa aagaccgagg tgcagacagg cggcttcagc 4080

aaagagtcta tcctgcccaa gaggaacagc gataagctga tcgccagaaa gaaggactgg 4140

gaccctaaga agtacggcgg cttcgacagc cccaccgtgg cctattctgt gctggtggtg 4200

gccaaagtgg aaaagggcaa gtccaagaaa ctgaagagtg tgaaagagct gctggggatc 4260

accatcatgg aaagaagcag cttcgagaag aatcccatcg actttctgga agccaagggc 4320

tacaaagaag tgaaaaagga cctgatcatc aagctgccta agtactccct gttcgagctg 4380

gaaaacggcc ggaagagaat gctggcctct gccggcgaac tgcagaaggg aaacgaactg 4440

gccctgccct ccaaatatgt gaacttcctg tacctggcca gccactatga gaagctgaag 4500

ggctcccccg aggataatga gcagaaacag ctgtttgtgg aacagcacaa gcactacctg 4560

gacgagatca tcgagcagat cagcgagttc tccaagagag tgatcctggc cgacgctaat 4620

ctggacaaag tgctgtccgc ctacaacaag caccgggata agcccatcag agagcaggcc 4680

gagaatatca tccacctgtt taccctgacc aatctgggag cccctgccgc cttcaagtac 4740

tttgacacca ccatcgaccg gaagaggtac accagcacca aagaggtgct ggacgccacc 4800

ctgatccacc agagcatcac cggcctgtac gagacacgga tcgacctgtc tcagctggga 4860

ggtgactctg gtggaagcgg aggatctggc ggcagcacca atctgagcga catcatcgag 4920

aaagagacag gcaagcagct ggtcatccaa gagtccatcc tgatgctgcc tgaagaggtg 4980

gaagaagtga tcggcaacaa gcccgagtcc gacatcctgg tgcacaccgc ctacgatgag 5040

agcaccgacg agaacgtgat gctgctgacc tctgacgccc ctgagtacaa gccttgggct 5100

ctcgtgatcc aggacagcaa cggcgagaac aagatcaaga tgctgagcgg cggctctggt 5160

ggctctggcg gatctacaaa cctgtccgat attattgaga aagaaaccgg gaaacagctc 5220

gtgattcaag agtctattct catgctcccg gaagaagtcg aggaagtcat tggaaacaag 5280

cctgagagcg atattctggt ccatacagcc tacgacgagt ctaccgatga gaatgtcatg 5340

ctcctcacca gcgacgctcc cgagtataag ccatgggcac ttgtcattca ggactccaat 5400

ggggaaaaca aaatcaaaat gctcccaaag aaaaaacgca aggtggaggg agctgataag 5460

cgcaccgccg atggttccga gttcgaaagc cccaagaaga agaggaaagt ctaaccggtc 5520

atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 5580

gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 5640

ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 5700

ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 5760

ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 5820

cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 5880

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 5940

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 6000

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 6060

gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 6120

ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 6180

caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 6240

aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 6300

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 6360

cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 6420

ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 6480

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg 6540

accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 6600

cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 6660

cagagttctt gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct 6720

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6780

aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 6840

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6900

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6960

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 7020

gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 7080

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 7140

ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 7200

accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 7260

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 7320

acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 7380

tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 7440

cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 7500

tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 7560

ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 7620

gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 7680

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 7740

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 7800

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 7860

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 7920
gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 7980
ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 8040
ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 8100
atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 8160
caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 8220
cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 8280

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 8340
aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 8400
taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 8460
agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 8520
cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 8580
tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 8640
tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 8700

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 8760
caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 8820
aggtctatat aagcagagct ggtttagtga accgtcagat ccgctagaga tccgcggccg 8880
ctaatacgac tcactatagg gagagccgcc acc 8913

<210> 226
<211> 1837
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 226

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 Gly Ser Gly Gly Ser Ser Ser Glu

1025 1030 1035

Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg Arg Ile Glu

1040 1045 1050

Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu Arg Lys

1055 1060 1065

Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser

1070 1075 1080

Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

1085 1090 1095

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn

1100 1105 1110

Thr Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly

1115 1120 1125

Glu Cys Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His

1130 1135 1140

Val Thr Leu Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp

1145 1150 1155

Pro Arg Asn Arg Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val

1160 1165 1170

Thr Ile Gln Ile Met Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg

1175 1180 1185

Asn Phe Val Asn Tyr Ser Pro Ser Asn Glu Ala His Trp Pro Arg

1190 1195 1200

Tyr Pro His Leu Trp Val Arg Leu Tyr Val Leu Glu Leu Tyr Cys

1205 1210 1215

Ile Ile Leu Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys

1220 1225 1230

Gln Pro Gln Leu Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His

1235 1240 1245

Tyr Gln Arg Leu Pro Pro His Ile Leu Trp Ala Thr Gly Leu Lys

1250 1255 1260

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr

1265 1270 1275

Pro Glu Ser Ser Gly Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1280 1285 1290

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1295 1300 1305

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1310 1315 1320

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1325 1330 1335

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1340 1345 1350

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1355 1360 1365

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1370 1375 1380

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1385 1390 1395

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1400 1405 1410

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1415 1420 1425

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1430 1435 1440

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1445 1450 1455

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1460 1465 1470

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1475 1480 1485

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1490 1495 1500

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1505 1510 1515

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1520 1525 1530

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1535 1540 1545

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1550 1555 1560

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1565 1570 1575

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1580 1585 1590

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1595 1600 1605

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser

1610 1615 1620

Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile

1625 1630 1635

Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile

1640 1645 1650

Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro

1655 1660 1665

Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp

1670 1675 1680

Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro

1685 1690 1695

Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys

1700 1705 1710

Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu

1715 1720 1725

Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln

1730 1735 1740

Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly

1745 1750 1755

Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu

1760 1765 1770

Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu

1775 1780 1785

Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn

1790 1795 1800

Lys Ile Lys Met Leu Pro Lys Lys Lys Arg Lys Val Glu Gly Ala

1805 1810 1815

Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys

1820 1825 1830

Lys Arg Lys Val

1835

<210> 227

<211> 8913

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 227

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660
cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900
ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140
gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200
aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320
gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440
gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500
aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560
tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620
agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680
gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860
ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100
agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

aacggaggct ctggaggaag cagctctgag acaggacctg tggccgtgga tcccacactg 3180

cggagaagaa ttgagcccca cgagttcgag gtgttcttcg accccagaga gctgcggaaa 3240

gagacatgcc tgctgtacga gatcaactgg ggcggcagac actctatctg gcggcacaca 3300

agccagaaca ccaacaagca cgtggaagtg aactttatcg agaagtttac gaccgagcgg 3360

tacttctgcc ccaacaccag atgcagcatc acctggtttc tgagctggtc cccttgcggc 3420

gagtgcagca gagccatcac cgagtttctg tccagatatc cccacgtgac cctgttcatc 3480

tatatcgccc ggctgtacca ccacgccgat cctagaaata gacagggcct gcgcgacctg 3540

atcagcagcg gagtgacaat ccagatcatg accgagcaag agagcggcta ctgctggcgg 3600

aacttcgtga actacagccc cagcaacgaa gcccactggc ctagatatcc tcacctgtgg 3660

gtccgactgt acgtgctgga actgtactgc atcatcctgg gcctgcctcc atgcctgaac 3720

atcctgagaa gaaagcagcc tcagctgacc ttcttcacaa tcgccctgca gagctgccac 3780

taccagagac tgcctccaca catcctgtgg gccaccggac ttaagggctc ttctggatct 3840

gaaacacctg gcacaagtga gagcgccacc cctgagagct ctggcatcat gaacttttc 3900

aagaccgaga ttaccctggc caacggcgag atccggaagc ggcctctgat cgagacaaac 3960

ggcgaaaccg gggagatcgt gtgggataag ggccgggatt ttgccaccgt gcggaaagtg 4020

ctgagcatgc cccaagtgaa tatcgtgaaa aagaccgagg tgcagacagg cggcttcagc 4080

aaagagtcta tcctgcccaa gaggaacagc gataagctga tcgccagaaa gaaggactgg 4140

gaccctaaga agtacggcgg cttcgacagc cccaccgtgg cctattctgt gctggtggtg 4200

gccaaagtgg aaaagggcaa gtccaagaaa ctgaagagtg tgaaagagct gctggggatc 4260

accatcatgg aaagaagcag cttcgagaag aatcccatcg actttctgga agccaagggc 4320

tacaaagaag tgaaaaagga cctgatcatc aagctgccta agtactccct gttcgagctg 4380

gaaaacggcc ggaagagaat gctggcctct gccggcgaac tgcagaaggg aaacgaactg 4440

gccctgccct ccaaatatgt gaacttcctg tacctggcca gccactatga gaagctgaag 4500

ggctcccccg aggataatga gcagaaacag ctgtttgtgg aacagcacaa gcactacctg 4560

gacgagatca tcgagcagat cagcgagttc tccaagagag tgatcctggc cgacgctaat 4620

ctggacaaag tgctgtccgc ctacaacaag caccgggata agcccatcag agagcaggcc 4680

gagaatatca tccacctgtt taccctgacc aatctgggag cccctgccgc cttcaagtac 4740

tttgacacca ccatcgaccg gaagaggtac accagcacca aagaggtgct ggacgccacc 4800

ctgatccacc agagcatcac cggcctgtac gagacacgga tcgacctgtc tcagctggga 4860

ggtgactctg gtggaagcgg aggatctggc ggcagcacca atctgagcga catcatcgag 4920

aaagagacag gcaagcagct ggtcatccaa gagtccatcc tgatgctgcc tgaagaggtg 4980

gaagaagtga tcggcaacaa gcccgagtcc gacatcctgg tgcacaccgc ctacgatgag 5040

agcaccgacg agaacgtgat gctgctgacc tctgacgccc ctgagtacaa gccttgggct 5100

ctcgtgatcc aggacagcaa cggcgagaac aagatcaaga tgctgagcgg cggctctggt 5160

ggctctggcg gatctacaaa cctgtccgat attattgaga aagaaaccgg gaaacagctc 5220

gtgattcaag agtctattct catgctcccg gaagaagtcg aggaagtcat tggaaacaag 5280

cctgagagcg atattctggt ccatacagcc tacgacgagt ctaccgatga gaatgtcatg 5340

ctcctcacca gcgacgctcc cgagtataag ccatgggcac ttgtcattca ggactccaat 5400

ggggaaaaca aaatcaaaat gctcccaaag aaaaaacgca aggtggaggg agctgataag 5460

cgcaccgccg atggttccga gttcgaaagc cccaagaaga agaggaaagt ctaaccggtc 5520

atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 5580

gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 5640

ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 5700

ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 5760

ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 5820

cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 5880

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 5940

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 6000

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 6060
gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 6120
ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 6180
caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 6240
aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 6300
atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 6360
cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 6420

ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 6480
gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg 6540
accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 6600
cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 6660
cagagttctt gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct 6720
gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6780
aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 6840

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6900
actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6960
taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 7020
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 7080
tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 7140
ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 7200
accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 7260

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 7320
acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 7380
tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 7440
cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 7500
tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 7560
ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 7620
gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 7680

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 7740

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 7800

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaagggа ataagggcga 7860

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 7920

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 7980

ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 8040

ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 8100

atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 8160

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 8220

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 8280

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 8340

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 8400

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 8460

agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 8520

cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 8580

tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 8640

tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 8700

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 8760

caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 8820

aggtctatat aagcagagct ggtttagtga accgtcagat ccgctagaga tccgcggccg 8880

ctaatacgac tcactatagg gagagccgcc acc 8913

<210> 228

<211> 1837

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 228

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 Gly Gly Ser Gly Gly Ser Ser Ser Glu Thr Gly Pro
1040 1045 1050
Val Ala Val Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu
1055 1060 1065

Phe Glu Val Phe Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys
1070 1075 1080
Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg
1085 1090 1095
His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val Asn Phe Ile
1100 1105 1110
Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys
1115 1120 1125

Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser
1130 1135 1140
Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
1145 1150 1155
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn
1160 1165 1170
Arg Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln
1175 1180 1185

Ile Met Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val
1190 1195 1200
Asn Tyr Ser Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His
1205 1210 1215
Leu Trp Val Arg Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu
1220 1225 1230
Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln
1235 1240 1245

Leu Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg

1250 1255 1260

Leu Pro Pro His Ile Leu Trp Ala Thr Gly Leu Lys Gly Ser Ser

1265 1270 1275

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1280 1285 1290

Ser Gly Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1295 1300 1305

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1310 1315 1320

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1325 1330 1335

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1340 1345 1350

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1355 1360 1365

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1370 1375 1380

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1385 1390 1395

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1400 1405 1410

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1415 1420 1425

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1430 1435 1440

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1445 1450 1455

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1460 1465 1470

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1475 1480 1485

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1490 1495 1500
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1505 1510 1515
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1520 1525 1530
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1535 1540 1545

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1550 1555 1560
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1565 1570 1575
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1580 1585 1590
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1595 1600 1605

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser
1610 1615 1620
Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile
1625 1630 1635
Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile
1640 1645 1650
Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro
1655 1660 1665

Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp
1670 1675 1680
Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro
1685 1690 1695
Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys
1700 1705 1710
Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu

1715 1720 1725

Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln
1730 1735 1740
Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly
1745 1750 1755
Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu
1760 1765 1770
Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu
1775 1780 1785

Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn
1790 1795 1800
Lys Ile Lys Met Leu Pro Lys Lys Lys Arg Lys Val Glu Gly Ala
1805 1810 1815
Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys
1820 1825 1830
Lys Arg Lys Val
1835

<210> 229
<211> 8913
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 229
atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60
atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240
tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520

atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580

gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640

aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700

accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760

ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880

aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940

taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000

taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060

atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120

aacatcatga acttttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg 3180

cctctgatcg agacaaacgg cgaaaccgga ggctctggag gaagcagctc tgagacagga 3240

cctgtggccg tggatcccac actgcggaga agaattgagc cccacgagtt cgaggtgttc 3300

ttcgacccca gagagctgcg gaaagagaca tgcctgctgt acgagatcaa ctggggcggc 3360

agacactcta tctggcggca cacaagccag aacaccaaca agcacgtgga agtgaacttt 3420

atcgagaagt ttacgaccga gcggtacttc tgccccaaca ccagatgcag catcacctgg 3480

tttctgagct ggtccccttg cggcgagtgc agcagagcca tcaccgagtt tctgtccaga 3540

tatccccacg tgaccctgtt catctatatc gcccggctgt accaccacgc cgatcctaga 3600

aatagacagg gcctgcgcga cctgatcagc agcggagtga caatccagat catgaccgag 3660

caagagagcg gctactgctg gcggaacttc gtgaactaca gccccagcaa cgaagcccac 3720

tggcctagat atcctcacct gtgggtccga ctgtacgtgc tggaactgta ctgcatcatc 3780

ctgggcctgc ctccatgcct gaacatcctg agaagaaagc agcctcagct gaccttcttc 3840

acaatcgccc tgcagagctg ccactaccag agactgcctc cacacatcct gtgggccacc 3900

ggacttaagg gctcttctgg atctgaaaca cctggcacaa gtgagagcgc caccctgag 3960

agctctggcg gggagatcgt gtgggataag ggccgggatt ttgccaccgt gcggaaagtg 4020

ctgagcatgc cccaagtgaa tatcgtgaaa aagaccgagg tgcagacagg cggcttcagc 4080

aaagagtcta tcctgcccaa gaggaacagc gataagctga tcgccagaaa gaaggactgg 4140

gaccctaaga agtacggcgg cttcgacagc cccaccgtgg cctattctgt gctggtggtg 4200

gccaaagtgg aaaagggcaa gtccaagaaa ctgaagagtg tgaaagagct gctggggatc 4260
accatcatgg aaagaagcag cttcgagaag aatcccatcg actttctgga agccaagggc 4320
tacaaagaag tgaaaaagga cctgatcatc aagctgccta agtactccct gttcgagctg 4380
gaaaacggcc ggaagagaat gctggcctct gccggcgaac tgcagaaggg aaacgaactg 4440
gccctgccct ccaaatatgt gaacttcctg tacctggcca gccactatga gaagctgaag 4500
ggctcccccg aggataatga gcagaaacag ctgtttgtgg aacagcacaa gcactacctg 4560
gacgagatca tcgagcagat cagcgagttc tccaagagag tgatcctggc cgacgctaat 4620

ctggacaaag tgctgtccgc ctacaacaag caccgggata agcccatcag agagcaggcc 4680
gagaatatca tccacctgtt taccctgacc aatctgggag cccctgccgc cttcaagtac 4740
tttgacacca ccatcgaccg gaagaggtac accagcacca aagaggtgct ggacgccacc 4800
ctgatccacc agagcatcac cggcctgtac gagacacgga tcgacctgtc tcagctggga 4860
ggtgactctg gtggaagcgg aggatctggc ggcagcacca atctgagcga catcatcgag 4920
aaagagacag gcaagcagct ggtcatccaa gagtccatcc tgatgctgcc tgaagaggtg 4980
gaagaagtga tcggcaacaa gcccgagtcc gacatcctgg tgcacaccgc ctacgatgag 5040

agcaccgacg agaacgtgat gctgctgacc tctgacgccc ctgagtacaa gccttgggct 5100
ctcgtgatcc aggacagcaa cggcgagaac aagatcaaga tgctgagcgg cggctctggt 5160
ggctctggcg gatctacaaa cctgtccgat attattgaga aagaaaccgg gaaacagctc 5220
gtgattcaag agtctattct catgctcccg gaagaagtcg aggaagtcat tggaaacaag 5280
cctgagagcg atattctggt ccatacagcc tacgacgagt ctaccgatga gaatgtcatg 5340
ctcctcacca gcgacgctcc cgagtataag ccatgggcac ttgtcattca ggactccaat 5400
ggggaaaaca aaatcaaaat gctcccaaag aaaaaacgca aggtggaggg agctgataag 5460

cgcaccgccg atggttccga gttcgaaagc cccaagaaga agaggaaagt ctaaccggtc 5520
atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 5580
gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 5640
ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 5700
ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 5760
ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 5820
cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 5880

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 5940

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 6000

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 6060

gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 6120

ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 6180

caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 6240

aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 6300

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 6360

cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 6420

ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 6480

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccccc gttcagcccg 6540

accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 6600

cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 6660

cagagttctt gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct 6720

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6780

aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 6840

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6900

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6960

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 7020

gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 7080

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 7140

ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 7200

accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 7260

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 7320

acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 7380

tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 7440

cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 7500

tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 7560

ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 7620

gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 7680

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 7740

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 7800

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaagga ataagggcga 7860

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 7920

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 7980

ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 8040

ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 8100

atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 8160

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 8220

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 8280

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 8340

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 8400

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 8460

agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 8520

cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 8580

tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 8640

tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 8700

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 8760

caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 8820

aggtctatat aagcagagct ggtttagtga accgtcagat ccgctagaga tccgcggccg 8880

ctaatacgac tcactatagg gagagccgcc acc 8913

<210> 230

<211> 1837

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 230

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 Gly Ser Gly Gly Ser Ser Ser Glu Thr Gly Pro Val Ala
1070 1075 1080
Val Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu Phe Glu
1085 1090 1095
Val Phe Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu
1100 1105 1110

Tyr Glu Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg His Thr
1115 1120 1125
Ser Gln Asn Thr Asn Lys His Val Glu Val Asn Phe Ile Glu Lys
1130 1135 1140
Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile
1145 1150 1155
Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser Arg Ala
1160 1165 1170

Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu Phe Ile
1175 1180 1185
Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg Gln
1190 1195 1200
Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
1205 1210 1215
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr
1220 1225 1230

Ser Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp
1235 1240 1245
Val Arg Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu

1250 1255 1260

Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr

1265 1270 1275

Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro

1280 1285 1290

Pro His Ile Leu Trp Ala Thr Gly Leu Lys Gly Ser Ser Gly Ser

1295 1300 1305

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

1310 1315 1320

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1325 1330 1335

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1340 1345 1350

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1355 1360 1365

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1370 1375 1380

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1385 1390 1395

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1400 1405 1410

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1415 1420 1425

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1430 1435 1440

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1445 1450 1455

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1460 1465 1470

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1475 1480 1485

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1490 1495 1500
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1505 1510 1515
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1520 1525 1530

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1535 1540 1545
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1550 1555 1560
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1565 1570 1575
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1580 1585 1590

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1595 1600 1605
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser
1610 1615 1620
Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile
1625 1630 1635
Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile
1640 1645 1650

Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro
1655 1660 1665
Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp
1670 1675 1680
Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro
1685 1690 1695
Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys
1700 1705 1710

Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu

1715 1720 1725

Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln

1730 1735 1740

Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly

1745 1750 1755

Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu

1760 1765 1770

Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu

1775 1780 1785

Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn

1790 1795 1800

Lys Ile Lys Met Leu Pro Lys Lys Lys Arg Lys Val Glu Gly Ala

1805 1810 1815

Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys

1820 1825 1830

Lys Arg Lys Val

1835

<210> 231

<211> 8913

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 231

atggacaaga agtacagcat cggcctggcc atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg agtacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aaactggtgg acagcaccga caaggccgac ctgcggctga tctatctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc 600

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacacc gagatcacca aggcccccct gagcgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgcttc cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc 1740

tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg cgatagcctg 2160

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca 2220

gtgaaggtgg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg 2280

atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag ccgcgagaga 2340

atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc 2400

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg 2460
gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat 2520
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc 2580
gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag 2640
aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagtt cgacaatctg 2700
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag 2760
ctggtggaaa cccggcagat cacaaagcac gtggcacaga tcctggactc ccggatgaac 2820

actaagtacg acgagaatga caagctgatc cgggaagtga aagtgatcac cctgaagtcc 2880
aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac 2940
taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag 3000
taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag 3060
atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc 3120
aacatcatga acttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg 3180
cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt 3240

gccaccgtgc ggaaagtgct gagcatgccc caagtgaata tcgtgaaaaa gaccgaggtg 3300
cagacaggcg gcttcagcaa agagtctatc ctgcccaaga ggaacagcga taagctgatc 3360
gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc 3420
tattctgtgc tggtggtggc caaagtggaa aagggcaagt ccaagaaact gaagagtgtg 3480
aaagagctgc tggggatcac catcatggaa agaagcagct tcgagaagaa tcccatcgac 3540
tttctggaag ccaagggcta caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag 3600
tactccctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg 3660

cagaagggaa acgaactggc cctgccctcc aaatatgtga acttcctgta cctggccagc 3720
cactatgaga agctgaaggg cggaggctct ggaggaagca gctctgagac aggacctgtg 3780
gccgtggatc ccacactgcg gagaagaatt gagccccacg agttcgaggt gttcttcgac 3840
cccagagagc tgcggaaaga gacatgcctg ctgtacgaga tcaactgggg cggcagacac 3900
tctatctggc ggcacacaag ccagaacacc aacaagcacg tggaagtgaa ctttatcgag 3960
aagtttacga ccgagcggta cttctgcccc aacaccagat gcagcatcac ctggtttctg 4020
agctggtccc cttgcggcga gtgcagcaga gccatcaccg agtttctgtc cagatatccc 4080

cacgtgaccc tgttcatcta tatcgcccgg ctgtaccacc acgccgatcc tagaaataga 4140

cagggcctgc gcgacctgat cagcagcgga gtgacaatcc agatcatgac cgagcaagag 4200

agcggctact gctggcggaa cttcgtgaac tacagcccca gcaacgaagc ccactggcct 4260

agatatcctc acctgtgggt ccgactgtac gtgctggaac tgtactgcat catcctgggc 4320

ctgcctccat gcctgaacat cctgagaaga aagcagcctc agctgacctt cttcacaatc 4380

gccctgcaga gctgccacta ccagagactg cctccacaca tcctgtgggc caccggactt 4440

aagggctctt ctggatctga aacacctggc acaagtgaga gcgccacccc tgagagctct 4500

ggctcccccg aggataatga gcagaaacag ctgtttgtgg aacagcacaa gcactacctg 4560

gacgagatca tcgagcagat cagcgagttc tccaagagag tgatcctggc cgacgctaat 4620

ctggacaaag tgctgtccgc ctacaacaag caccgggata agcccatcag agagcaggcc 4680

gagaatatca tccacctgtt taccctgacc aatctgggag cccctgccgc cttcaagtac 4740

tttgacacca ccatcgaccg gaagaggtac accagcacca aagaggtgct ggacgccacc 4800

ctgatccacc agagcatcac cggcctgtac gagacacgga tcgacctgtc tcagctggga 4860

ggtgactctg gtggaagcgg aggatctggc ggcagcacca atctgagcga catcatcgag 4920

aaagagacag gcaagcagct ggtcatccaa gagtccatcc tgatgctgcc tgaagaggtg 4980

gaagaagtga tcggcaacaa gcccgagtcc gacatcctgg tgcacaccgc ctacgatgag 5040

agcaccgacg agaacgtgat gctgctgacc tctgacgccc ctgagtacaa gccttgggct 5100

ctcgtgatcc aggacagcaa cggcgagaac aagatcaaga tgctgagcgg cggctctggt 5160

ggctctggcg gatctacaaa cctgtccgat attattgaga aagaaaccgg gaaacagctc 5220

gtgattcaag agtctattct catgctcccg gaagaagtcg aggaagtcat tggaaacaag 5280

cctgagagcg atattctggt ccatacagcc tacgacgagt ctaccgatga gaatgtcatg 5340

ctcctcacca gcgacgctcc cgagtataag ccatgggcac ttgtcattca ggactccaat 5400

ggggaaaaca aaatcaaaat gctcccaaag aaaaaacgca aggtggaggg agctgataag 5460

cgcaccgccg atggttccga gttcgaaagc cccaagaaga agaggaaagt ctaaccggtc 5520

atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 5580

gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 5640

ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 5700

ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 5760

ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 5820

cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 5880

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 5940

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 6000

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 6060

gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 6120

ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 6180

caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 6240

aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 6300

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 6360

cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 6420

ccgcctttct cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 6480

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg 6540

accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 6600

cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 6660

cagagttctt gaagtggtgg cctaactacg gctacactag aagaacagta tttggtatct 6720

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6780

aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 6840

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6900

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6960

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 7020

gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 7080

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 7140

ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 7200

accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 7260

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 7320

acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 7380

tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 7440

cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 7500

tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 7560

ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 7620

gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 7680

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 7740

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 7800

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 7860

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 7920

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 7980

ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 8040

ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 8100

atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 8160

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 8220

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 8280

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 8340

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 8400

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 8460

agtatttacg gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc 8520

cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgcccag tacatgacct 8580

tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt accatggtga 8640

tgcggttttg gcagtacatc aatgggcgtg gatagcggtt tgactcacgg ggatttccaa 8700

gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa cgggactttc 8760

caaaatgtcg taacaactcc gccccattga cgcaaatggg cggtaggcgt gtacggtggg 8820

aggtctatat aagcagagct ggtttagtga accgtcagat ccgctagaga tccgcggccg 8880

ctaatacgac tcactatagg gagagccgcc acc 8913

<210> 232

<211> 1837

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 232

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Gly

1235 1240 1245

Gly Ser Gly Gly Ser Ser Ser Glu Thr Gly Pro Val Ala Val Asp

1250 1255 1260

Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu Phe Glu Val Phe

1265 1270 1275

Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu Tyr Glu

1280 1285 1290

Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg His Thr Ser Gln

1295 1300 1305

Asn Thr Asn Lys His Val Glu Val Asn Phe Ile Glu Lys Phe Thr

1310 1315 1320

Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr Trp

1325 1330 1335

Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser Arg Ala Ile Thr

1340 1345 1350

Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu Phe Ile Tyr Ile

1355 1360 1365

Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg Gln Gly Leu

1370 1375 1380

Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met Thr Glu

1385 1390 1395

Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser Pro

1400 1405 1410

Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

1415 1420 1425

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro

1430 1435 1440

Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe

1445 1450 1455

Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His

1460 1465 1470

Ile Leu Trp Ala Thr Gly Leu Lys Gly Ser Ser Gly Ser Glu Thr

1475 1480 1485

Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Pro

1490 1495 1500

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1505 1510 1515

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1520 1525 1530

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1535 1540 1545

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1550 1555 1560

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1565 1570 1575

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1580 1585 1590

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1595 1600 1605

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser

1610 1615 1620

Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile

1625 1630 1635

Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile

1640 1645 1650

Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro

1655 1660 1665

Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp

1670 1675 1680

Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro

1685 1690 1695

Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys

1700 1705 1710

Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu

1715 1720 1725

Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln

1730 1735 1740

Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly

1745 1750 1755

Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu

1760 1765 1770

Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu

1775 1780 1785

Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn

1790 1795 1800

Lys Ile Lys Met Leu Pro Lys Lys Lys Arg Lys Val Glu Gly Ala

1805 1810 1815

Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys

1820 1825 1830

Lys Arg Lys Val

1835

<210> 233

<211> 28

<212> DNA

<213> Homo sapiens

<400> 233

ggcctccgta tcactctctg actggggt 28

<210> 234

<211> 28

<212> DNA

<213> Homo sapiens

<400> 234

gtaactgaac ccctgcaatc aatgggat 28

<210> 235

<211> 27

<212> DNA

<213> Homo sapiens

<400> 235

gccacagact tttccatttg caggagt 27

<210> 236

<211> 28

<212> DNA

<213> Homo sapiens

<400> 236

gcgaaaggct cgcggcgaag gaaggaat 28

<210> 237

<211> 27

<212> DNA

<213> Homo sapiens

<400> 237

gcctggcaga tgagaaccag gaggaat 27

<210> 238

<211> 28

<212> DNA

<213> Homo sapiens

<400> 238

gtattactat tattatctga gatggggt 28

<210> 239

<211> 27

<212> DNA

<213> Homo sapiens

<400> 239

gccacagtgg gaggggacat ggggaat 27

<210> 240

<211> 27

<212> DNA

<213> Homo sapiens

<400> 240

gccctgatct gcactgaaca gaggggt 27

<210> 241

<211> 28

<212> DNA

<213> Homo sapiens

<400> 241

gcctcaagtc tggttatttt agggggat 28

<210> 242

<211> 27

<212> DNA

<213> Homo sapiens

<400> 242

ggtcgaccct tggtatccat gggggat 27

<210> 243

<211> 27

<212> DNA

<213> Homo sapiens

<400> 243

gaaagagaca gagaaggggc agggggt 27

<210> 244

<211> 28

<212> DNA

<213> Homo sapiens

<400> 244

gagtgggaac tttctgatgc catggaat 28

<210> 245

<211> 28

<212> DNA

<213> Homo sapiens

<400> 245

gtgggactga tcccttaatg tgtggggt 28

<210> 246

<211> 27

<212> DNA

<213> Homo sapiens

<400> 246

gcccagctcc agcctctgat gaggggt 27

<210> 247

<211> 28

<212> DNA

<213> Homo sapiens

<400> 247

gaaggcttta ctgtattaca gaaggggt 28

<210> 248

<211> 27

<212> DNA

<213> Homo sapiens

<400> 248

ggagccagag accagtgggc agggggt 27

<210> 249

<211> 28

<212> DNA

<213> Homo sapiens

<400> 249

gctttcctta gctgtaaaag aaagggat 28

<210> 250

<211> 28

<212> DNA

<213> Homo sapiens

<400> 250

gagaagaaac cagggaacag gtaggagt 28

<210> 251

<211> 28

<212> DNA

<213> Homo sapiens

<400> 251

gatgtgtcta ctgttactta caaggaat 28

<210> 252

<211> 27

<212> DNA

<213> Homo sapiens

<400> 252

gaccaggtca gcaaacatgt ttggaat 27

<210> 253

<211> 28

<212> DNA

<213> Homo sapiens

<400> 253

gcacccaggg gttctgcaga gcagggat 28

<210> 254

<211> 27

<212> DNA

<213> Homo sapiens

<400> 254

gcccagcaat tcactgtgaa gagggat 27

<210> 255

<211> 27

<212> DNA

<213> Homo sapiens

<400> 255

gaccaaaacg agggacattt aggggat 27

<210> 256

<211> 28

<212> DNA

<213> Homo sapiens

<400> 256

gctcctctca cccttatgac tcagggat 28

<210> 257

<211> 28

<212> DNA

<213> Homo sapiens

<400> 257

gactcagcgc ccctgccggg cctgggat 28

<210> 258

<211> 27

<212> DNA

<213> Homo sapiens

<400> 258

ggtcgtagcc agtccgaacc ccggagt 27

<210> 259

<211> 27

<212> DNA

<213> Homo sapiens

<400> 259

gcattccact ccgtccgcct ccggagt 27

<210> 260

<211> 28

<212> DNA

<213> Homo sapiens

<400> 260

gggtacctga gtggggtgca tttggggt 28

<210> 261

<211> 20

<212> DNA

<213> Homo sapiens

<400> 261

ggcactgcgg ctggaggtgg 20