



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년11월10일
(11) 등록번호 10-2465067
(24) 등록일자 2022년11월04일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) C07K 14/195 (2006.01)
C12N 15/10 (2017.01) C12N 15/113 (2010.01)
(52) CPC특허분류
C12N 9/22 (2013.01)
C07K 14/195 (2013.01)
(21) 출원번호 10-2020-7026571
(22) 출원일자(국제) 2019년02월15일
심사청구일자 2020년09월15일
(85) 번역문제출일자 2020년09월15일
(65) 공개번호 10-2020-0121342
(43) 공개일자 2020년10월23일
(86) 국제출원번호 PCT/US2019/018335
(87) 국제공개번호 WO 2019/161290
국제공개일자 2019년08월22일
(30) 우선권주장
62/631,304 2018년02월15일 미국(US)
62/720,525 2018년08월21일 미국(US)
(56) 선행기술조사문헌
US20170314015 A1
KR1020170061697 A
비특허1
비특허2

(73) 특허권자
시그마-알드리치 컴퍼니., 엘엘씨
미국 미주리 63103 세인트 루이스 스프러스 스트리트 3050
(72) 발명자
시백 티모시
미국 미주리 63103 세인트 루이스 스프러스 스트리트 3050, 씨/오 시그마-알드리치 컴퍼니, 엘엘씨
첸 푸치앙
미국 미주리 63103 세인트 루이스 스프러스 스트리트 3050, 씨/오 시그마-알드리치 컴퍼니, 엘엘씨
데이비스 그레고리
미국 미주리 63103 세인트 루이스 스프러스 스트리트 3050, 씨/오 시그마-알드리치 컴퍼니, 엘엘씨
(74) 대리인
강명구, 이경민

전체 청구항 수 : 총 44 항

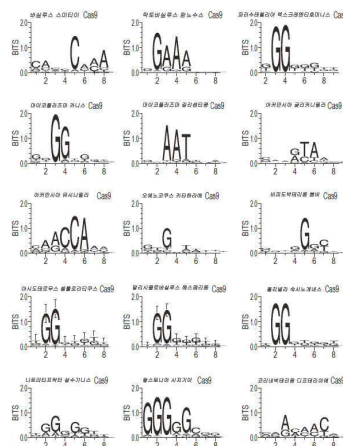
심사관 : 한지혜

(54) 발명의 명칭 **진핵 게놈 변형을 위한 조작된 CAS9 시스템**

(57) 요약

표적 DNA 결합을 위해 대체 프로토스페이서 인접 모티프들을 이용하는 조작된 Cas9 시스템, 상기 조작된 Cas9 시스템을 인코딩하는 핵산, 및 진핵 세포에서 표적 염색체 서열들을 변형하기 위해 상기 조작된 Cas9 시스템을 사용하는 방법.

대표도 - 도1



(52) CPC특허분류

C12N 15/102 (2013.01)

C12N 15/113 (2013.01)

C07K 2319/00 (2013.01)

C12N 2310/20 (2017.05)

명세서

청구범위

청구항 1

시험관내(*in vitro*)에서 진핵 세포에서 염색체 서열을 변형시키기 위한 시스템으로서, 이때 상기 시스템은 (a) 조작된(engineered) *바실루스 스미티이* (*Bacillus smithii*) Cas9 단백질; 그리고 (b) 조작된 가이드 RNA, 이때 조작된 가이드 RNA는 조작된 Cas9 단백질과 복합체를 형성하도록 설계되며, 그리고 상기 조작된 가이드 RNA는 염색체 서열에서 표적 서열과 혼성화되도록 설계된 5' 가이드 서열을 포함하고, 이때 상기 표적 서열은 프로토스페이서 인접 모티프 (PAM)에 대해 5'에 있으며, PAM은 5'-NNNNCAAA-3' 서열을 포함하며, 이때 N은 A, C, G 또는 T인, 시스템.

청구항 2

청구항 1에 있어서, 상기 조작된 Cas9 단백질은 그 야생형 대응물에 비해 적어도 하나의 변형을 포함하는, 시스템.

청구항 3

청구항 2에 있어서, 적어도 하나의 변형은 적어도 하나의 이중 도메인의 부가를 포함하는, 시스템.

청구항 4

청구항 3에 있어서, 적어도 하나의 이중(heterologous) 도메인은 핵 국제화 신호, 세포-침투 도메인, 마커 도메인, 염색질 조절 모티프, 후성 변형 도메인, 전사 조절 도메인, RNA 압타머 결합 도메인, 또는 이의 조합을 포함하는, 시스템.

청구항 5

청구항 2에 있어서, 상기 적어도 하나의 변형은 하나 이상의 아미노산의 치환, 하나 이상의 아미노산의 삽입, 하나 이상의 아미노산의 결실, 또는 이들의 조합을 포함하는, 시스템.

청구항 6

청구항 5에 있어서, 상기 적어도 하나의 변형은, RuvC 도메인, HNH 도메인, REC 도메인, PAM 상호작용 도메인, 또는 이의 조합 내부에 존재하는, 시스템.

청구항 7

청구항 1에 있어서, 상기 조작된 Cas9 단백질은 뉴클레아제이고 이중 가닥 서열의 두 가닥 모두를 절단하거나, 니카아제이고 이중 가닥 서열 중 한 가닥을 절단하거나, 뉴클레아제 또는 니카아제 활성을 가지지 않는, 시스템.

청구항 8

청구항 1에 있어서, 조작된 가이드 RNA는 단일 분자인, 시스템.

청구항 9

청구항 1에 있어서, 상기 조작된 가이드 RNA 서열은 진핵 세포에서 조작된 가이드 RNA 내부에서 염기 페어링을 촉진하거나, 조작된 가이드 RNA 내부에서 염기 페어링을 최소화하거나, 조작된 가이드 RNA의 안정성을 증가시키거나, 조작된 가이드 RNA의 전사를 촉진하거나, 또는 이의 조합인, 시스템.

청구항 10

청구항 1에 있어서, 상기 시스템은 적어도 하나의 공여체 폴리뉴클레오티드를 더 포함하며, 이때 상기 적어도 하나의 조작된 가이드 RNA는 적어도 하나의 조작된 Cas9 단백질과 공여체 폴리뉴클레오티드를 염색체 서열에서

표적 부위로 안내할 수 있고, 이로써 상기 공여체 폴리뉴클레오티드가 염색체 서열로의 통합이 일어나는, 시스템.

청구항 11

청구항 7에 있어서, 상기 조작된 Cas9 단백질은 서열 번호: 2, 119 또는 120에 적어도 90% 서열 동일성을 가지는 아미노산 서열을 가지는, 시스템.

청구항 12

청구항 7에 있어서, 상기 조작된 Cas9 단백질은 서열 번호: 2, 119 또는 120에서 제시된 아미노산 서열을 가지는, 시스템.

청구항 13

청구항 7의 시스템을 인코딩하는 복수의 핵산으로서, 이러한 복수의 핵산은 상기 조작된 Cas9 단백질을 인코딩하는 적어도 하나의 핵산, 및 상기 조작된 가이드 RNA를 인코딩하는 적어도 하나의 핵산을 포함하는, 복수의 핵산.

청구항 14

청구항 13에 있어서, 상기 조작된 Cas9 단백질을 인코딩하는 적어도 하나의 핵산은 RNA인, 복수의 핵산.

청구항 15

청구항 13에 있어서, 상기 조작된 Cas9 단백질을 인코딩하는 적어도 하나의 핵산은 DNA인, 복수의 핵산.

청구항 16

청구항 13에 있어서, 상기 조작된 Cas9 단백질을 인코딩하는 적어도 하나의 핵산은 진핵 세포에서의 발현에 코돈 최적화되는, 복수의 핵산.

청구항 17

청구항 16에 있어서, 진핵 세포는 인간 세포, 비-인간 포유동물 세포, 비-포유동물 척추동물 세포, 무척추동물 세포, 식물 세포, 또는 단세포 진핵 유기체인, 복수의 핵산.

청구항 18

청구항 13에 있어서, 상기 조작된 가이드 RNA를 인코딩하는 적어도 하나의 핵산은 DNA인, 복수의 핵산.

청구항 19

청구항 13에 있어서, 상기 조작된 Cas9 단백질을 인코딩하는 적어도 하나의 핵산은 *시험관내* RNA 합성 또는 박테리아 세포에서 단백질 발현을 위해 파지 프로모터 서열에 작동가능하게 연결되고, 상기 조작된 가이드 RNA를 인코딩하는 적어도 하나의 핵산은 *시험관내* RNA 합성을 위해 파지 프로모터 서열에 작동가능하게 연결되는, 복수의 핵산.

청구항 20

청구항 13에 있어서, 상기 조작된 Cas9 단백질을 인코딩하는 적어도 하나의 핵산은 진핵 세포에서 발현을 위해 진핵 프로모터 서열에 작동가능하게 연결되고, 상기 조작된 가이드 RNA를 인코딩하는 적어도 하나의 핵산은 진핵 세포에서 발현을 위해 진핵 프로모터 서열에 작동가능하게 연결되는, 복수의 핵산.

청구항 21

청구항 13의 복수의 핵산을 포함하는 적어도 하나의 벡터.

청구항 22

청구항 21에 있어서, 플라스미드 벡터, 바이러스 벡터, 또는 자가-복제 RNA 레플리콘인, 적어도 하나의 벡터.

청구항 23

청구항 1 내지 12 중 어느 한 항에 따른 조작된 Cas9 단백질 및 조작된 가이드 RNA를 포함하는 적어도 하나의 시스템을 포함하는 시험관내(*in vitro*) 진행 세포.

청구항 24

청구항 23에 있어서, 인간 세포, 비-인간 포유동물 세포, 식물 세포, 비-포유동물 척추동물 세포, 무척추동물 세포, 또는 단세포 진행 유기체인, 시험관내(*in vitro*) 진행 세포.

청구항 25

청구항 1 내지 12 중 어느 한 항에 따른 조작된 Cas9 단백질 및 조작된 가이드 RNA를 포함하는 적어도 하나의 시스템을 진행 세포에 도입하는 것을 포함하는, 진행 세포에서 염색체 서열을 변형시키는 시험관내(*in vitro*) 방법에 있어서, 이때 상기 적어도 하나의 조작된 가이드 RNA는 상기 적어도 하나의 조작된 Cas9 단백질을 염색체 서열 내 표적 부위로 유도하여, 염색체 서열의 변형이 일어나는, 방법.

청구항 26

청구항 25에 있어서, 상기 시스템은 적어도 하나의 공여체 폴리뉴클레오티드를 더 포함하며, 이때 상기 적어도 하나의 조작된 가이드 RNA는 적어도 하나의 조작된 Cas9 단백질과 공여체 폴리뉴클레오티드를 염색체 서열에서 표적 부위로 안내할 수 있고, 이로써 상기 공여체 폴리뉴클레오티드가 염색체 서열로의 통합이 일어나는, 방법.

청구항 27

청구항 25에 있어서, 상기 변형은 적어도 하나의 뉴클레오티드의 치환, 적어도 하나의 뉴클레오티드의 결실, 적어도 하나의 뉴클레오티드의 삽입, 적어도 하나의 뉴클레오티드의 전환, 적어도 하나의 뉴클레오티드의 변형, 적어도 관련된 히스티딘 단백질의 변형, 또는 이의 조합을 포함하는, 방법.

청구항 28

청구항 25에 있어서, 상기 조작된 Cas9 단백질은 뉴클레아제 또는 니카아제 활성을 가지고, 상기 변형은 적어도 하나의 삽입결실(indel)을 포함하는, 방법.

청구항 29

청구항 28에 있어서, 상기 변형은 염색체 서열의 비활성화를 포함하는, 방법.

청구항 30

청구항 26에 있어서, 상기 조작된 Cas9 단백질은 뉴클레아제 또는 니카아제 활성을 가지며, 상기 적어도 하나의 공여체 폴리뉴클레오티드는 세포 내부에 도입되고, 상기 변형은 염색체 서열의 적어도 하나의 뉴클레오티드의 변화를 포함하는, 방법.

청구항 31

청구항 30에 있어서, 상기 적어도 하나의 공여체 폴리뉴클레오티드는 염색체 서열 내 표적 부위 근방의 서열에 비해 적어도 하나의 뉴클레오티드 변화를 가지는 공여체 서열을 포함하는, 방법.

청구항 32

청구항 31에 있어서, 적어도 하나의 공여체 폴리뉴클레오티드는 외인성 서열에 해당하는 공여체 서열을 포함하는, 방법.

청구항 33

청구항 31에 있어서, 상기 공여체 서열은 염색체 서열의 표적 부위의 상류 및 하류에 위치한 서열들과 실질적 서열 동일성을 가지는 서열에 연결되는, 방법.

청구항 34

청구항 31에 있어서, 상기 공여체 서열은 적어도 하나의 조작된 Cas9 단백질에 의해 생성된 오버행과 양립가능한 짧은 오버행에 연결되는, 방법.

청구항 35

청구항 25에 있어서, 진핵 세포는 인간 세포, 비-인간 포유동물 세포, 식물 세포, 비-포유동물 척추동물 세포, 무척추동물 세포, 또는 단세포 진핵 유기체인, 방법.

청구항 36

적어도 하나의 염색질 조절 모티프에 연결된 Cas9 단백질을 포함하는 융합 단백질로서, 상기 Cas9 단백질은 *바실러스 스미티이* (*Bacillus smithii*) Cas9 단백질인, 융합 단백질.

청구항 37

청구항 36에 있어서, 상기 적어도 하나의 염색질 조절 모티프는 고 이동성 그룹 (HMG) 박스 (HMGB) DNA 결합 도메인, HMG 뉴클레오솜-결합 (HMGN) 단백질, 히스티딘 H1 변이체의 중심 구상 도메인, 염색질 재형성 복합체 단백질의 DNA 결합 도메인, 또는 이의 조합인, 융합 단백질.

청구항 38

청구항 36 또는 37에 있어서, 적어도 하나의 염색질 조절 모티프는 HMGB1 박스 A 도메인, HMGN1 단백질, HMGN2 단백질, HMGN3a 단백질, HMGN3b 단백질, 히스티딘 H1 중심 구상 도메인, 모방 스위치 (ISWI) 단백질 DNA 결합 도메인, 크로모도메인-헬리케이스-DNA 단백질 1 (CHD1) DNA 결합 도메인, 또는 이의 조합인, 융합 단백질.

청구항 39

청구항 36에 있어서, 적어도 하나의 염색질 조절 모티프는 화학 결합에 의해 직접적으로, 링커에 의해 간접적으로, 또는 이의 조합으로 Cas9 단백질에 연결되는, 융합 단백질.

청구항 40

청구항 36에 있어서, 상기 적어도 하나의 염색질 조절 모티프는 N-말단, C-말단, 내부 위치, 또는 이의 조합에서 Cas9 단백질에 연결되는, 융합 단백질.

청구항 41

청구항 36에 있어서, 적어도 하나의 핵 국제화 신호를 추가로 포함하는, 융합 단백질.

청구항 42

청구항 36에 있어서, 적어도 하나의 세포-침투 도메인, 적어도 하나의 마커 도메인, 또는 이의 조합을 추가로 포함하는, 융합 단백질.

청구항 43

청구항 36에 있어서, 상기 융합 단백질은 서열 번호:119 또는 120에 대해 적어도 90% 서열 동일성을 가지는 아미노산 서열을 가지는, 융합 단백질.

청구항 44

청구항 36 내지 43 중 어느 한 항에 있어서, 상기 융합 단백질은 서열 번호:119 또는 120에 제시된 아미노산 서열을 가지는, 융합 단백질.

청구항 45

삭제

청구항 46

삭제

발명의 설명

기술 분야

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2018년 2월 15일 출원된 미국 가특허 출원 제 62/631,304, 및 2018년 8월 21일 출원된 미국 가특허 출원 제 62/720,525의 이익을 주장하며, 이 문헌들 각각은 본원에 그 전문이 참고문헌으로 포함된다.

[0003] 서열 목록

[0004] 본원은 ASCII 형식으로 전자적으로 제출된 서열 목록을 포함하며 이는 본원에 그 전체가 참고로 포함된다. 2019년 2월 12일 생성된 상기 ASCII 사본 파일명은 P18_023PCT_SL.txt이며 370,305 바이트 크기이다.

[0005] 분야

[0006] 본 발명은 조작된 Cas9 시스템, 이러한 시스템을 인코딩하는 핵산, 및 게놈 변형을 위해 이러한 시스템을 사용하는 방법에 관한 것이다.

배경 기술

[0007] 배경

[0008] 게놈 편집 툴로서 최근 박테리아 클래스 2 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR) 및 CRISPR-관련 (Cas) CRISPR/Cas 시스템의 개발은 진핵생물 게놈 변형을 위한 부위 특이적 엔도뉴클레아제를 조합에 전례없는 용이성과 단순성을 제공했다. 그러나 각 CRISPR/Cas 시스템은 표적 DNA 결합을 위해 특정 프로토스페이서 인접 모티프 (PAM)를 필요로 하기 때문에 각 시스템은 특정 게놈 부위로 제한된다. 현재 가장 널리 채택되는 *스트렙토코쿠스 피오게네스* (*Streptococcus pyogenes*) Cas9 (SpyCas9)는 표적화를 위해 종종 발생하는 PAM (5'-NGG-3')를 사용하지만, 이는 여전히 이러한 모티프가 없는 많은 게놈 부위들로부터 배제되는데, 왜냐하면 진핵생물 게놈, 특히 포유동물과 식물의 게놈은 DNA 서열이 매우 복잡하고 이질적이기 때문이다. 더욱이, 상동 재조합 (HDR) 또는 염기 편집자, 가령, dCas9/시티딘 탈아미노효소 및 dCas9/아데노신 탈아미노효소를 사용하는 정밀한 유전자 편집은, 최적의 편집 결과를 구현하기 위해 단일 염기쌍 분해능에서조차도 정확한 DNA 결합 위치를 자주 필요로 한다. 그러므로, 게놈 커버리지 밀도를 높이기 위해 표적화에 새로운 PAM들을 사용하는 새로운 CRISPR/Cas 시스템을 개발할 필요성이 존재한다.

발명의 내용

[0009] 요약

[0010] 본 발명의 다양한 양상들 중 조작된 Cas9 단백질 및 조작된 가이드 RNA를 포함하는 조작된 Cas9 시스템이 포함되는데, 이 때 각 조작된 가이드 RNA는 조작된 Cas9 단백질과 복합체를 형성하도록 설계되고 이러한 조작된 가이드 RNA는 이중 가닥 서열에서 표적 서열과 혼성화하도록 설계된 5' 가이드 서열을 포함하고, 이 때 표적 서열은 프로토스페이서 인접 모티프 (PAM)에 대해 5'에 있으며 PAM은 표 A에 나열된 서열을 가진다.

[0011] 본 발명의 또 다른 양상은 상기 조작된 Cas9 시스템을 인코딩하는 복수의 핵산 및 복수의 상기 핵산을 포함하는 적어도 하나의 벡터를 포함한다.

[0012] 추가 측면은 적어도 하나의 조작된 Cas9 시스템 및/또는 상기 조작된 Cas9 시스템을 인코딩하는 적어도 하나의 핵산을 포함하는 진핵 세포를 포함한다.

[0013] 본 발명의 또 다른 측면은 진핵 세포에서 염색체 서열을 변형하는 방법을 포함한다. 상기 방법은 조작된 Cas9 단백질 및 조작된 가이드 RNA, 그리고 선택적으로 적어도 하나의 공여체 폴리뉴클레오타이드를 포함하는 적어도 하나의 조작된 Cas9 시스템 및/또는 상기 조작된 Cas9 시스템을 인코딩하는 적어도 하나의 핵산을 진핵 세포에 도입하는 것을 포함하며, 여기서 적어도 하나의 조작된 가이드 RNA는 적어도 하나의 조작된 Cas9 단백질을 염색체 서열의 표적 부위로 유도하여 염색체 서열의 변형이 발생하게 한다.

[0014] 본 발명의 다른 양상들 및 특징들은 이하에서 상세히 설명한다.

도면의 간단한 설명

[0015] **도면의 간단한 설명**

도 1은 Cas9 오솔로그에 의한 시험관내 표적 DNA 절단에 필요한 프로토스페이서 인접 모티프 (PAM)의 웹로그 분석을 보여준다. 가로축 숫자는 PAM 서열에서 뉴클레오타이드의 위치를 나타낸다.

도 2A는 McaCas9, McaCas9-HN1HB1 융합 (즉, 아미노 말단의 HMGN1 및 카르복실 말단의 HMGB1 상자 A) 및 McaCas9-HN1H1G 융합 (즉, 아미노 말단에서의 HMGN1 및 카르복실 말단에서 히스톤 H1 중심 구상 모티프)의 절단 효율을 (삽입결실의 퍼센트로서) 나타낸다. 각 유전자좌의 표적 부위는 표 6에 제공된다. 오차 막대는 평균 $\pm$ SD를 보여준다 ($n = 3$ 생물학적 복제본).

도 2B는 PexCas9, PexCas9-HN1HB1 융합 (즉, 아미노 말단의 HMGN1 및 카르복실 말단의 HMGB1 상자 A) 및 PexCas9-HN1H1G 융합 (즉, 아미노 말단에서의 HMGN1 및 카르복실 말단에서 히스톤 H1 중심 구상 모티프)의 절단 효율을 (삽입결실의 퍼센트로서) 나타낸다. 각 유전자좌의 표적 부위는 표 6에 제공된다. 오차 막대는 평균 $\pm$ SD를 보여준다 ($n = 3$ 생물학적 복제본).

도 2C는 BsmCas9, BsmCas9-HN1HB1 융합 (즉, 아미노 말단의 HMGN1 및 카르복실 말단의 HMGB1 상자 A) 및 BsmCas9-HN1H1G 융합 (즉, 아미노 말단에서의 HMGN1 및 카르복실 말단에서 히스톤 H1 중심 구상 모티프)의 절단 효율을 (삽입결실의 퍼센트로서) 나타낸다. 각 유전자좌의 표적 부위는 표 6에 제공된다. 오차 막대는 평균 $\pm$ SD를 보여준다 ($n = 3$ 생물학적 복제본).

도 2D는 LrhCas9, LrhCas9-HN1HB1 융합 (즉, 아미노 말단의 HMGN1 및 카르복실 말단의 HMGB1 상자 A) 및 LrhCas9-HN1H1G 융합 (즉, 아미노 말단에서의 HMGN1 및 카르복실 말단에서 히스톤 H1 중심 구상 모티프)의 절단 효율을 (삽입결실의 퍼센트로서) 나타낸다. 각 유전자좌의 표적 부위는 표 6에 제공된다. 오차 막대는 평균 $\pm$ SD를 보여준다 ($n = 3$ 생물학적 복제본).

도 3은 대조 Cas9 및 Cas9-CMM 융합 뉴클레아제의 오프-타겟 활성을 (삽입결실의 퍼센트로서) 보여준다. 오차 막대는 평균 $\pm$ SD를 보여준다 ($n = 3$ 생물학적 복제본).

발명을 실시하기 위한 구체적인 내용

[0016] **상세한 설명**

[0017] 본 발명은 표적 DNA 결합을 위해 대체 PAM을 사용하여 게놈 커버리지 밀도를 증가시키는 오솔로그 Cas9 시스템을 제공한다. 예를 들어, 이러한 대체 PAM 중 일부는 A 및/또는 T 잔기를 포함하고 다른 대체 PAMs는 GC가 풍부하다. 따라서 이러한 대체 PAM을 이용하는 조작된 Cas9 시스템은 이전에는 접근 할 수 없었던 게놈 유전자좌의 표적화된 게놈 편집 또는 게놈 변형을 가능하게 한다.

[0018] **(I) 조작된 Cas9 시스템**

[0019] 본 발명의 한 측면은 조작된 Cas9 단백질 및 조작된 가이드 RNA를 포함하는 조작된 Cas9 시스템을 제공하며, 여기서 각각의 조작된 가이드 RNA는 특정한 조작된 Cas9 단백질과 복합체를 형성하도록 설계된다. 각각의 조작된 가이드 RNA는 이중 가닥 서열에서 표적 서열과 혼성화하도록 설계된 5' 가이드 서열을 포함하고, 이 때 표적 서열은 프로토스페이서 인접 모티프 (PAM)에 대해 5'에 있으며 PAM은 표 A에 나열된 서열을 가진다. 이러한 조작된 Cas9 시스템은 자연 발생하지 않는다.

[0020] **(a) 조작된 Cas9 단백질**

[0021] 조작된 Cas9 단백질은 야생형 대응물에 비해 적어도 하나의 아미노산 치환, 삽입 또는 결실을 포함한다. Cas9 단백질은 다양한 박테리아에 존재하는 II형 CRISPR 시스템의 단일 이펙터 단백질이다. 본원에 개시된 조작된 Cas9 단백질은 아카리오클로리스 종 (*Acaryochloris* sp.), 아세토할로비움 종 (*Acetohalobium* sp.), 아시다미노코쿠스 종 (*Acidaminococcus* sp.), 아시디티오바실루스 종 (*Acidithiobacillus* sp.), 아시도테르무스 종 (*Acidothermus* sp.), 아커만시아 종 (*Akkermansia* sp.), 알리시클로바실루스 종 (*Alicyclobacillus* sp.), 알로크로마티움 종 (*Allochromatium* sp.), 암모니펙스 종 (*Ammonifex* sp.), 아나바에나 종 (*Anabaena* sp.), 아르트로스피라 종 (*Arthrosira* sp.), 바실루스 종 (*Bacillus* sp.), 비피도박테리움 종 (*Bifidobacterium* sp.), 부르크홀데리알레스 종 (*Burkholderiales* sp.), 칼디셀룰로시럽터 종 (*Caldicelulosiruptor* sp.), 캄필로박터 종 (*Campylobacter* sp.), 칸디다투스 종 (*Candidatus* sp.), 클로스트리디움 종 (*Clostridium* sp.), 코리네박테리움 종 (*Corynebacterium* sp.), 크로코스파에라 종 (*Crocospaera* sp.), 시아노테세 종 (*Cyanothecae* sp.), 엑시구오박테리움 종 (*Exiguobacterium* sp.), 피네골디아 종 (*Finegoldia* sp.), 프란시셀라 종 (*Francisella* sp.),

크테도노박터 종 (*Ktedonobacter* sp.), 라크노스피라세아에 종 (*Lachnospiraceae* sp.), 락토바실루스 종 (*Lactobacillus* sp.), 린그비아 종 (*Lyngbya* sp.), 마리노박터 종 (*Marinobacter* sp.), 메타노할로비움 종 (*Methanohalobium* sp.), 마이크로실라 종 (*Microscilla* sp.), 마이크로콜레우스 종 (*Microcoleus* sp.), 마이크로시스 티스 종 (*Microcystis* sp.), 마이코플라즈마 종 (*Mycoplasma* sp.), 나트라나에로비우스 종 (*Natranaerobius* sp.), 네이세리아 종 (*Neisseria* sp.), 니트라티프락터 종 (*Nitratifractor* sp.), 니트로소코쿠스 종 (*Nitrosococcus* sp.), 노카르디옵시스 종 (*Nocardiosis* sp.), 노둘라리아 종 (*Nodularia* sp.), 노스톡 종 (*Nostoc* sp.), 오에노코쿠스 종 (*Oenococcus* sp.), 오실라토리아 종 (*Oscillatoria* sp.), 파라수테렐라 종 (*Parasutterella* sp.), 펠로토마쿨룸 종 (*Pelotomaculum* sp.), 페트로로가 종 (*Petrotoga* sp.), 폴라로모나스 종 (*Polaromonas* sp.), 프레보텔라 종 (*Prevotella* sp.), 수도알테로모나스 종 (*Pseudoalteromonas* sp.), 랄스 토니아 종 (*Ralstonia* sp.), 스탕필로코쿠스 종 (*Staphylococcus* sp.), 스트렙토코쿠스 종 (*Streptococcus* sp.), 스트렙토마이세스 종 (*Streptomyces* sp.), 스트렙토스포란지움 종 (*Streptosporangium* sp.), 시네코코쿠스 종 (*Synechococcus* sp.), 테르모시포 종 (*Thermosipho* sp.), 베루코미크로비아 종 (*Verrucomicrobia* sp.), 및 올리넬라 종 (*Wolinella* sp.)에서 유래할 수 있다.

[0022] 특정 구체예들에서, 본원에 개시된 조작된 Cas9 단백질은 아시도테르무스 종 (*Acidothermus* sp.), 아커만시아 종 (*Akkermansia* sp.), 알리시클로바실루스 종 (*Alicyclobacillus* sp.), 바실루스 종 (*Bacillus* sp.), 비피도 박테리움 종 (*Bifidobacterium* sp.), 부르크홀데리아레스 종 (*Burkholderiales* sp.), 코리네박테리움 종 (*Corynebacterium* sp.), 락토바실루스 종 (*Lactobacillus* sp.), 마이코플라즈마 종 (*Mycoplasma* sp.), 니트라 티프락터 종 (*Nitratifractor* sp.), 오에노코쿠스 종 (*Oenococcus* sp.), 파라수테렐라 종 (*Parasutterella* sp.), 랄스토니아 종 (*Ralstonia* sp.), 또는 올리넬라 종 (*Wolinella* sp.)에서 유래할 수 있다.

[0023] 특정 구체예들에서, 본원에 개시된 조작된 Cas9 단백질은 is from 아시도테르무스 셀룰로리티쿠스 (*Acidothermus cellulolyticus*) (Ace), 아커만시아 글리카니필라 (*Akkermansia glycaniphila*) (Ag1), 아커만시아 뮤시니필라 (*Akkermansia muciniphila*) (Amu), 알리시클로바실루스 헤스페리둠 (*Alicyclobacillus hesperidum*) (Ahe), 바실루스 스미티이 (*Bacillus smithii*) (Bsm), 비피도박테리움 봄비 (*Bifidobacterium bombi*) (Bbo), 코리네박테리움 디프테리아 (*Corynebacterium diphtheria*) (Cdi), 락토바실루스 람노수스 (*Lactobacillus rhamnosus*) (Lrh), 마이코플라즈마 카니스 (*Mycoplasma canis*) (Mca), 마이코플라즈마 갈리셉 티쿰 (*Mycoplasma gallisepticum*) (Mga), 니트라티프락터 살수기니스 (*Nitratifractor saluginis*) (Nsa), 오 에노코쿠스 키타하라에 (*Oenococcus kitaharae*) (Ok1), 파라수테렐리아 엑스크레멘티호미니스 (*Parasutterella excrementihominis*) (Pex), 랄스토니아 시지गी (*Ralstonia syzygii*) (Rsy), 또는 올리넬라 속시노게네스 (*Wolinella succinogenes*) (Wsu)에서 유래한다.

[0024] 야생형 Cas9 단백질은 두 개의 뉴클레아제 도메인, 즉, RuvC 및 HNH 도메인을 포함하며, 이들 각각은 이중 가닥 서열의 한 가닥을 절단한다. Cas9 단백질은 또한 가이드 RNA (예컨대, REC1, REC2) 또는 RNA/DNA 이중이중체 (예컨대, REC3)와 상호작용하는 REC 도메인들, 및 프로토스페이서-인접 모티프 (PAM) (즉, PAM-상호작용 도메인)와 상호작용하는 도메인을 포함한다.

[0025] Cas9 단백질은 하나 이상의 변형 (즉, 하나 이상의 아미노산의 치환, 하나 이상의 아미노산의 결실, 하나 이상의 아미노산의 삽입)을 포함하도록 조작되어 Cas9 단백질은 변경된 활성, 특이성 및/또는 안정성을 가질 수 있다.

[0026] 예를 들어, Cas9 단백질은 뉴클레아제 도메인 중 하나 또는 둘 모두를 비활성화하기 위한 하나 이상의 돌연변이 및/또는 결실에 의해 조작 될 수 있다. 하나의 뉴클레아제 도메인의 비활성화는 이중 가닥 서열의 한 가닥을 절단하는 Cas9 단백질을 생성한다 (즉, Cas9 니카아제). RuvC 도메인은 D10A, D8A, E762A 및/또는 D986A와 같은 돌연변이에 의해 비활성화 될 수 있으며, HNH 도메인은 H840A, H559A, N854A, N856A 및/또는 N863A와 같은 돌연변이에 의해 비활성화 될 수 있다 (스트렙토코쿠스 피오케네스 Cas9, SpyCas9의 넘버링 시스템 참고). 두 뉴클레아제 도메인의 비활성화는 절단 활성이 없는 Cas9 단백질을 생성한다 (즉, 촉매적으로 비활성이거나 죽은 Cas9).

[0027] Cas9 단백질은 또한 하나 이상의 아미노산 치환, 결실 및/또는 삽입에 의해 조작되어 개선된 표적화 특이성, 개선된 충실도, 변경된 PAM 특이성, 감소된 오프-타겟 효과 및/또는 증가된 안정성을 가질 수 있다. 표적화 특이성을 개선하고, 충실도를 개선하고, 및/또는 오프-타겟 효과를 감소시키는 하나 이상의 돌연변이의 비-제한적인 예는 N497A, R661A, Q695A, K810A, K848A, K855A, Q926A, K1003A, R1060A 및/또는 D1135E를 포함한다 (SpyCas9의 넘버링 체계 참고).

[0028] (i) 이중 도메인

[0029] Cas9 단백질은 하나 이상의 이중 도메인을 포함하도록 조작 될 수 있다, 즉, Cas9는 하나 이상의 이중 도메인에 융합된다. 둘 이상의 이중 도메인이 Cas9와 융합되는 상황에서, 둘 이상의 이중 도메인은 동일하거나 다를 수 있다. 하나 이상의 이중 도메인은 N 말단 단부, C 말단 단부, 내부 위치 또는 이들의 조합에 융합 될 수 있다. 융합은 화학적 결합을 통한 직접 일 수 있으며, 또는 연결은 하나 이상의 링커를 통한 간접 일 수 있다. 다양한 구체예에서, 이중 도메인은 핵 국재화 신호, 세포-침투 도메인, 마커 도메인, 염색질 파괴 도메인, 후성 변형 도메인 (예를 들어, 시티딘 탈아미노효소 도메인, 히스톤 아세틸트랜스퍼라제 도메인 등), 전사 조절 도메인, RNA 압타머 결합 도메인, 또는 비-Cas9 뉴클레아제 도메인일 수 있다.

[0030] 일부 구체예에서 하나 이상의 이중 도메인은 핵 국재화 신호 (NLS) 일 수 있다. 핵 국재화 신호의 비-제한적인 예는 PKKKRKV (서열 번호: 78), PKKKRRV (서열 번호: 79), KRPAATKKAGQAKKKK (서열 번호: 80), YGRKKRRQRRR (서열 번호: 81), RKKRRQRRR (서열 번호: 82), PAAKRVKLD (서열 번호: 83), RQRRNELKRSP (서열 번호: 84), VSRKRPRP (서열 번호: 85), PPKKARED (서열 번호: 86), PQPKKKPL (서열 번호: 87), SALIKKKKKMAP (서열 번호: 88), PKQKKRK (서열 번호: 89), RKLKKIKKL (서열 번호: 90), REKKFLKRR (서열 번호: 91), KRKGDEVDGVDEVAKKSKK (서열 번호: 92), RKCLQAGMNLEARKTKK (서열 번호: 93), NQSSNFGPMKGGNFGGRSSGPGYGGGQYFAKPRNQGGY (서열 번호: 94) 및 RMRIZFKNGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (서열 번호: 95)를 포함한다.

[0031] 다른 구체예에서, 하나 이상의 이중 도메인은 세포-침투 도메인 일 수 있다. 적합한 세포-침투 도메인의 예는, 제한없이, GRKKRRQRRPPQPKKKRKV (서열 번호: 96), PLSSIFSRIGDPPKKKKRKV (서열 번호: 97), GALFLGLGAAGSTMGAPKKKKRKV (서열 번호: 98), GALFLGLGAAGSQWGAQPKWKKRKV (서열 번호: 99), KETQWGAWSQPKKKRKV (서열 번호: 100), YARAAARQARA (서열 번호: 101), THRLPRRRRRR (서열 번호: 102), GGRRARRRRRR (서열 번호: 103), RRQRRTSKLMKR (서열 번호: 104), GWTLNSAGYLLGKINLKALAALAKKIL (서열 번호: 105), KALAWEAKLAKALAKALAKHLAKALAKALKEA (서열 번호: 106) 및 RQIKIWFQNRRMKWKK (서열 번호: 107)을 포함한다.

[0032] 대안적 구체예에서, 하나 이상의 이중 도메인은 마커 도메인 일 수 있다. 마커 도메인은 형광 단백질 및 정제 또는 에피토프 태그를 포함한다. 적합한 형광 단백질에는, 제한 없이, 녹색 형광 단백질 (예컨대, GFP, eGFP, GFP-2, tagGFP, turboGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), 황색 형광 단백질 (예컨대, YFP, EYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), 청색 형광 단백질 (예컨대, BFP, EBFP, EBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), 시안 형광 단백질 (예컨대, ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), 적색 형광 단백질 (예컨대, mKate, mKate2, mPlum, DsRed 단량체, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), 오렌지색 형광 단백질 (예컨대, mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato), 및 이의 조합이 포함된다. 마커 도메인은 하나 이상의 형광 단백질 (예컨대, Suntag)의 일련 반복을 포함 할 수 있다. 적합한 정제 또는 에피토프 태그의 비-제한적 예들에는 6xHis (서열 번호: 134), FLAG[®], HA, GST, Myc, SAM, 등이 포함된다. CRISPR 복합체의 검출 또는 농축을 용이하게 하는 이중 융합의 비-제한적 예들은 스트렙타비딘 (Kiprianov 외, 인간 Antibodies, 1995, 6(3):93-101.), 아비딘 (Airenne 외, Biomolecular Engineering, 1999, 16(1-4):87-92), 아비딘의 단량체 형태 (Laitinen 외, Journal of Biological Chemistry, 2003, 278(6):4010-4014), 재조합 생산 동안 비오틴화를 용이하게 하는 펩티드 태그 (Cull 외, Methods in Enzymology, 2000, 326:430-440)를 포함한다.

[0033] 또한 다른 구체예들에서, 하나 이상의 이중 도메인은 염색질 조절 모티프 (CMM) 일 수 있다. CMM들의 비-제한적 예들은 고 이동성 그룹 (HMG) 단백질들 (예컨대, HMGB1, HMGB2, HMGB3, HMGN1, HMGN2, HMGN3a, HMGN3b, HMGN4, 및 HMGN5 단백질), 히스톤 H1 변이체들 (예컨대, 히스톤 H1.0, H1.1, H1.2, H1.3, H1.4, H1.5, H1.6, H1.7, H1.8, H1.9, 및 H.1.10)의 중심 구상 도메인, 또는 염색질 재형성 복합체의 DNA 결합 도메인 (예컨대, SWI/SNF (스위치/수크로스 비-발효성, SWItch/Sucrose Non-Fermentable), ISWI (모방 스위치, Imitation SWItch), CHD (크로모도메인-헬리카제-DNA 결합, Chromo도메인-헬리카제-DNA binding), Mi-2/NuRD (뉴클레오솜 재형성 및 탈아세틸화효소, Nucleosome Remodeling and 탈아세틸화효소), INO80, SWR1, 및 RSC 복합체에서 유래한 뉴클레오솜 상호작용 펩티드를 포함한다. 다른 구체예들에서, CMM들은 또한 토포아이스머라제, 헬리카제, 또는 바이러스 단백질에서 유래할 수 있다. CMM의 공급원은 변화할 수 있으며 변화할 것이다. CMM들은 인간, 동물 (즉, 척추동물 및 무척추동물), 식물, 조류, 또는 효모에서 유래할 수 있다. 특정 CMM의 비 제한적인 예

가 아래 표에 나열되어 있다. 당업자는 다른 종들의 상동체 및/또는 그 내부의 관련 융합 모티프를 용이하게 확인할 수 있다.

단백질	접근 번호	융합 모티프
인간 HMGN1	P05114	전장
인간 HMGN2	P05204	전장
인간 HMGN3a	Q15651	전장
인간 HMGN3b	Q15651-2	전장
인간 HMGN4	O00479	전장
인간 HMGN5	P82970	뉴클레오솜 결합 모티프
인간 HMGB1	P09429	박스 A
인간 히스티딘 H1.0	P07305	구상 모티프
인간 히스티딘 H1.2	P16403	구상 모티프
인간 CHD1	O14646	DNA 결합 모티프
효모 CHD1	P32657	DNA 결합 모티프
효모 ISWI	P38144	DNA 결합 모티프
인간 TOP1	P11387	DNA 결합 모티프
인간 헤르페스바이러스 8 LANA	J9QSF0	뉴클레오솜 결합 모티프
인간 CMV IE1	P13202	염색질 테더링 모티프

<i>M. 레프라이에</i> DNA 헬리카제	P40832	HhH 결합 모티프
--------------------------	--------	------------

또 다른 구체예에서, 하나 이상의 이중 도메인은 후성 변형 도메인 일 수 있다. 적합한 후성 변형 도메인의 비-제한적 예들은 DNA 탈아미노화 (예컨대, 시티딘 탈아미노효소, 아데노신 탈아미노효소, 구아닌 탈아미노효소), DNA 메틸전이효소 활성 (예컨대, 시토신 메틸전이효소), DNA 탈메틸효소 활성, DNA 아미노화, DNA 산화 활성, DNA 헬리카제 활성, 히스티딘 아세틸전이효소 (HAT) 활성 (예컨대, E1A 결합 단백질 p300에서 유래한 HAT 도메인), 히스티딘 탈아세틸화효소 활성, 히스티딘 메틸전이효소 활성, 히스티딘 탈메틸효소 활성, 히스티딘 키나제 활성, 히스티딘 포스포타제 활성, 히스티딘 유비퀴틴 리가제 활성, 히스티딘 탈유비퀴틴화 활성, 히스티딘 아데닐화 활성, 히스티딘 탈아데닐화 활성, 히스티딘 SUMO일화 활성, 히스티딘 데SUMOy일화 활성, 히스티딘 리보실화 활성, 히스티딘 탈리보실화 (deribosylation) 활성, 히스티딘 미리스토일화 활성, 히스티딘 탈미리스토일화 활성, 히스티딘 시트룰린화 활성, 히스티딘 알킬화 활성, 히스티딘 탈알킬화 활성, 또는 히스티딘 산화 활성을

가지진 것을 포함한다. 특정 구체예들에서, 후성 변형 도메인은 시티딘 탈아미노효소 활성화, 아데노신 탈아미노효소 활성화, 히스톤 아세틸전이효소 활성화, 또는 DNA 메틸전이효소 활성을 포함할 수 있다.

[0036] 다른 구체예에서, 하나 이상의 이중 도메인은 전사 조절 도메인 (즉, 전사 활성화 도메인 또는 전사 억제 도메인) 일 수 있다. 적합한 전사 활성화 도메인은 제한없이 단순 헤르페스 바이러스 VP16 도메인, VP64 (즉, VP16의 4개의 일렬 복제본), VP160 (즉, VP16의 10개의 일렬 복제본), NF- κ B p65 활성화 도메인 (p65), 엠스타인-바 바이러스 R 트랜스활성화인자 (Rta) 도메인, VPR (즉, VP64+p65+Rta), p300-의존성 전사 활성화 도메인, p53 활성화 도메인 1 및 2, 열 충격 인자 1 (HSF1) 활성화 도메인, Smad4 활성화 도메인 (SAD), cAMP 반응 요소 결합 단백질 (CREB) 활성화 도메인, E2A 활성화 도메인, 활성화된 T 세포의 핵 인자 (NFAT) 활성화 도메인, 또는 이들의 조합을 포함한다. 적합한 전사 억제자 도메인의 비 제한적인 예는 크루펠-관련 박스 (KRAB) 억제자 도메인, Mxi 억제자 도메인, 유도성 cAMP 조기 억제자 (ICER) 도메인, YY1 글리신 풍부 억제자 도메인, Sp1-유사 억제자, E(spl) 억제자를 포함하며, $\hat{I}\hat{E}$ B 억제자, Sin3 억제자, 메틸-CpG 결합 단백질 2 (MeCP2) 억제자, 또는 이들의 조합을 포함한다. 전사 활성화 또는 전사 억제인자 도메인은 Cas9 단백질에 유전적으로 융합되거나 비공유 단백질-단백질, 단백질-RNA, 또는 단백질-DNA 상호작용에 의해 결합될 수 있다.

[0037] 추가 구체예에서, 하나 이상의 이중 도메인은 RNA 압타머 결합 도메인 일 수 있다 (Koner mann 외, Nature, 2015, 517(7536):583-588; Zalatan 외, Cell, 2015, 160(1-2):339-50). 적합한 RNA 압타머 단백질 도메인의 예들에는 MS2 외피 단백질 (MCP), PP7 박테리오파지 외피 단백질 (PCP), Mu 박테리오파지 Com 단백질, 람다 박테리오파지 N22 단백질, 스템-루프 결합 단백질 (SLBP), 취약 X 정신 지체 증후군-관련 단백질 1 (FXR1), 박테리오파지로부터 유래한 단백질, 가령, AP205, BZ13, f1, f2, fd, fr, ID2, JP34/GA, JP501, JP34, JP500, KU1, M11, M12, MX1, NL95, PP7, ϕ Cb5, ϕ Cb8r, ϕ Cb12r, ϕ Cb23r, Q β , R17, SP- β , TW18, TW19, 및 VK, 이의 단편, 또는 이의 유도체가 포함된다.

[0038] 또 다른 구체예에서, 하나 이상의 이중 도메인은 비-Cas9 뉴클레아제 도메인 일 수 있다. 적합한 뉴클레아제 도메인은 임의의 엔도뉴클레아제 또는 엑소뉴클레아제로부터 얻을 수 있다. 뉴클레아제 도메인이 유도될 수 있는 엔도뉴클레아제의 비-제한적 예들은, 제한 엔도뉴클레아제 및 호밍 엔도뉴클레아제를 포함하나 이에 제한되는 것은 아니다. 일부 구체예들에서, 뉴클레아제 도메인은 유형 II-S 제한 엔도뉴클레아제로부터 유래될 수 있다. 유형 II-S 엔도뉴클레아제는 인식/결합 부위로부터 전형적으로 여러 개 염기쌍 떨어져있는 부위들의 DNA를 절단하고, 그리하여 분리가능한 결합 및 절단 도메인을 가진다. 이들 효소들은 일반적으로 엇갈린 위치들에서 DNA 각 가닥을 절단하기 위해 일시적으로 결합하여 이량체를 형성하는 단량체이다. 적합한 유형 II-S 엔도뉴클레아제의 비-제한적 예들에는 BfiI, BpmI, BsaI, BsgI, BsmBI, BsmI, BspMI, FokI, MboII, 및 SapI가 포함된다. 일부 구체예들에서, 뉴클레아제 도메인은 FokI 뉴클레아제 도메인 또는 이의 유도체일 수 있다. 유형 II-S 뉴클레아제 도메인은 2개의 상이한 뉴클레아제 도메인의 이량체화를 용이하게 하도록 변형될 수 있다. 예를 들면, FokI의 절단 도메인은 특정 아미노산 잔기들을 돌연변이시킴으로써 변형될 수 있다. 비-제한적 예로서, FokI 뉴클레아제 도메인의 위치 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, 및 538의 아미노산 잔기들은 변형을 위한 표적이다. 특정 구체예들에서, FokI 뉴클레아제 도메인은 Q486E, I499L, 및/또는 N496D 돌연변이들을 포함하는 제 1 FokI 절단-도메인, 및 E490K, I538K, 및/또는 H537R 돌연변이들을 포함하는 제 2 FokI 절단-도메인을 포함할 수 있다.

[0039] 하나 이상의 이중 도메인은 하나 이상의 화학적 결합 (예를 들어, 공유 결합)을 통해 Cas9 단백질에 직접 연결될 수 있거나, 하나 이상의 이중 도메인이 하나 이상의 링커를 통해 Cas9 단백질에 간접적으로 연결될 수 있다.

[0040] 링커는 최소한 하나의 공유 결합에 의해 하나 이상의 다른 화학적 작용기를 연결시키는 화학적 작용기이다. 적합한 링커들에는 아미노산, 펩티드, 뉴클레오타이드, 핵산, 유기 링커 분자들 (예컨대, 말레이미드 유도체, N-에톡시벤질이미다졸, 바이페닐-3,4',5-트라이카르복시산, p-아미노벤질옥시카르보닐, 등), 다이설파이드 링커, 및 폴리머 링커 (예컨대, PEG)가 포함된다. 링커는 알킬렌, 알켄일렌, 알킨일렌, 알킬, 알켄일, 알킨일, 알콕시, 아릴, 헤테로아릴, 아랄킬, 아랄켄일, 아랄킨일 등을 비롯한 (그러나 이에 제한되는 것은 아님) 하나 이상의 스페이싱 작용기들을 포함할 수 있다. 링커는 중성일 수 있거나, 양 또는 음 전하를 지닐 수 있다. 추가적으로, 링커는 절단가능할 수 있어서, 링커를 또 다른 화학적 작용기에 연결시키는 링커의 공유 결합이, pH, 온도, 염 농도, 빛, 촉매 또는 효소를 포함한 특정 조건하에서 파열되거나 절단될 수 있다. 일부 구체예들에서, 링커는 펩티드 링커일 수 있다. 펩티드 링커는 가요성 아미노산 링커 (예컨대, 소형, 비-극성 또는 극성 아미노산을 포함) 일 수 있다. 가요성 링커들의 비-제한적 예들에는 LEGGGS (서열 번호:108), TGSG (서열 번호:109),

GGSGGGSG (서열 번호:110), (GGGS)₁₋₄ (서열 번호:111), 및 (Gly)₆₋₈ (서열 번호:112)이 포함된다. 대안적으로, 펩티드 링커는 단단한 아미노산 링커일 수 있다. 이러한 링커들에는 (EAAAK)₁₋₄ (서열 번호:113), A(EAAAK)₂₋₅A (서열 번호:114), PAPAP (서열 번호:115), 및 (AP)₆₋₈ (서열 번호:116)이 포함된다. 적합한 링커들의 또 다른 예는 해당 분야에 잘 공지되어 있으며 링커들을 설계하는 프로그램들은 용이하게 이용가능하다 (Crasto 외., Protein Eng., 2000, 13(5):309-312).

[0041] 일부 구체예에서, 조작된 Cas9 단백질은 무세포 시스템, 박테리아 세포 또는 진핵 세포에서 재조합적으로 생산될 수 있고 표준 정제 수단을 사용하여 정제될 수 있다. 다른 구체예들에서, 조작된 Cas9 단백질은 조작된 Cas9 단백질을 인코딩하는 핵산으로부터 관심 진핵 세포의 *생체내에서* 생산된다 (아래 섹션 (II) 참고).

[0042] 조작된 Cas9 단백질이 뉴클레아제 또는 니카아제 활성을 포함하는 구체예에서, 조작된 Cas9 단백질은 적어도 하나의 핵 국제화 신호, 세포-침투 도메인 및/또는 마커 도메인, 뿐만 아니라 적어도 하나의 염색질 파괴 도메인을 추가로 포함 할 수 있다. 조작된 Cas9 단백질이 후성 변형 도메인에 연결되는 구체예에서, 조작된 Cas9 단백질은 적어도 하나의 핵 국제화 신호, 세포-침투 도메인 및/또는 마커 도메인, 뿐만 아니라 적어도 하나의 염색질 파괴 도메인을 추가로 포함 할 수 있다. 또한, 조작된 Cas9 단백질이 전사 조절 도메인에 연결되는 구체예에서, 조작된 Cas9 단백질은 적어도 하나의 핵 국제화 신호, 세포-침투 도메인 및/또는 마커 도메인, 뿐만 아니라 적어도 하나의 염색질 파괴 도메인 및/또는 적어도 하나의 RNA 압타머 결합 도메인을 추가로 포함 할 수 있다.

[0043] (ii) 특이적 조작된 Cas9 단백질

[0044] 특정 구체예들에서, 조작된 Cas9 단백질은 *바실루스 스미티이* (*Bacillus smithii*), *락토바실루스 람노수스* (*Lactobacillus rhamnosus*), *파라수테렐라 엑스크레멘티호미니스* (*Parasutterella excrementihominis*), *마이코플라즈마 카니스* (*Mycoplasma canis*), *마이코플라즈마 갈리셉티쿰* (*Mycoplasma gallisepticum*), *아커만시아 글리카니필라* (*Akkermansia glycaniphila*), *아커만시아 뮤시니필라* (*Akkermansia muciniphila*), *오에노코쿠스 키타하라에* (*Oenococcus kitaharae*), *비피도박테리움 봄비* (*Bifidobacterium bombi*), *아시도테르무스 셀룰로리티쿠스* (*Acidothermus cellulolyticus*), *알리시클로바실루스 헤스페리둠* (*Alicyclobacillus hesperidum*), *올리넬라 숙시노게네스* (*Wolinella succinogenes*), *니트라티프락터 살수기니스* (*Nitratifractor salsuginis*), *랄스토니아 시지기이* (*Ralstonia syzygii*), 또는 *코리네박테리움 디프테리아* (*Corynebacterium diphtheria*)에서 유래하며 적어도 하나의 NLS에 연결된다. 일부 예들에서, 조작된 Cas9 단백질은 서열 번호: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 또는 30에 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95% 또는 적어도 약 99% 서열 동일성을 가질 수 있다. 특정 구체예들에서, 조작된 Cas9 단백질은 서열 번호: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 또는 30에 적어도 약 95% 서열 동일성을 가질 수 있다. 다른 구체예들에서, 조작된 Cas9 단백질은 서열 번호: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 또는 30의 아미노산 서열을 가진다.

[0045] 다른 구체예들에서, 조작된 Cas9 단백질은 적어도 하나의 염색질 조절 모티프 (CMM)에 연결된, *바실루스 스미티이* (*Bacillus smithii*), *락토바실루스 람노수스* (*Lactobacillus rhamnosus*), *파라수테렐라 엑스크레멘티호미니스* (*Parasutterella excrementihominis*), *마이코플라즈마 카니스* (*Mycoplasma canis*), *마이코플라즈마 갈리셉티쿰* (*Mycoplasma gallisepticum*), *아커만시아 글리카니필라* (*Akkermansia glycaniphila*), *아커만시아 뮤시니필라* (*Akkermansia muciniphila*), *오에노코쿠스 키타하라에* (*Oenococcus kitaharae*), *비피도박테리움 봄비* (*Bifidobacterium bombi*), *아시도테르무스 셀룰로리티쿠스* (*Acidothermus cellulolyticus*), *알리시클로바실루스 헤스페리둠* (*Alicyclobacillus hesperidum*), *올리넬라 숙시노게네스* (*Wolinella succinogenes*), *니트라티프락터 살수기니스* (*Nitratifractor salsuginis*), *랄스토니아 시지기이* (*Ralstonia syzygii*), 또는 *코리네박테리움 디프테리아* (*Corynebacterium diphtheria*) Cas9 단백질 일 수 있다. Cas9 단백질과 CMM 사이의 연결은 직접 또는 링커를 통해 이루어질 수 있다. Cas9-CMM 융합 단백질은 적어도 하나의 NLS를 추가로 포함할 수 있다. 특정 구체예에서, Cas9-CMM 융합 단백질은 서열 번호:117, 118, 119, 1200, 121, 122, 123, 또는 124에 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95%, 또는 적어도 약 99% 서열 동일성을 가질 수 있다. 특정 구체예에서, Cas9-CMM 융합 단백질은 서열 번호: 117, 118, 119, 120, 121, 122, 123, 또는 124에 적어도 약 95% 서열 동일성을 가질 수 있다. 특정 구체예에서 Cas9-CMM 융합 단백질은 서열 번호: 117, 118, 119, 120, 121, 122, 123 또는 124의 아미노산 서열을 가진다.

[0046] (b) 조작된 가이드 RNA

- [0047] 조작된 가이드 RNA는 특정 조작된 Cas9 단백질과 복합체를 형성하도록 설계된다. 가이드 RNA는 (i) 표적 서열과 혼성화하는, 5' 단부에 가이드 서열을 내포하는 CRISPR RNA (crRNA) 및 (ii) Cas9 단백질을 동원하는 트랜스작용 crRNA (tracrRNA) 서열을 포함한다. 각 가이드 RNA의 crRNA 가이드 서열은 상이하다 (즉, 서열 특이적이다). tracrRNA 서열은 일반적으로 특정 박테리아 종들의 Cas9 단백질과 복합체를 형성하도록 설계된 가이드 RNA에서 동일하다.
- [0048] crRNA 가이드 서열은 이중 가닥 서열에서 표적 서열 (즉, 프로토스페이서)과 혼성화하도록 설계된다. 일반적으로, crRNA와 표적 서열 간의 상보성은 최소한 80%, 최소한 85%, 최소한 90%, 최소한 95%, 또는 최소한 99%이다. 특정 구체예들에서, 상보성은 완전하다 (즉, 100%). 다양한 구체예들에서, crRNA 가이드 서열의 길이는 약 15개 뉴클레오티드 내지 약 25개 뉴클레오티드 범위일 수 있다. 예를 들면, crRNA 가이드 서열은 약 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개 뉴클레오티드 길이일 수 있다. 특정 구체예들에서, crRNA는 약 19, 20, 또는 21개 뉴클레오티드 길이이다. 한 구체예에서, crRNA 가이드 서열은 20개 뉴클레오티드의 길이를 가진다.
- [0049] 가이드 RNA는 Cas9 단백질과 상호작용하는 하나 이상의 스템 루프 구조를 형성하는 반복 서열 및 단일 가닥으로 남아있는 3' 서열을 포함한다. 각 루프와 스템의 길이는 다를 수 있습니다. 예를 들면, 루프는 약 3 내지 약 10개 뉴클레오티드 길이 범위일 수 있으며, 스템은 약 6개 내지 약 20개 염기쌍 길이 범위일 수 있다. 스템은 하나 이상의 1 내지 약 10개 뉴클레오티드 팽창부를 포함할 수 있다. 단일 가닥 3' 영역의 길이는 다를 수 있다. 조작된 가이드 RNA에서 tracrRNA 서열은 일반적으로 관심 박테리아 종들에서의 야생형 tracrRNA의 코딩 서열에 기초한다. 야생형 서열은 2차 구조 형성을 촉진, 2차 구조 안정성을 증가, 진행 세포에서 발현을 촉진하도록 변형될 수 있다. 예를 들면, 하나 이상의 뉴클레오티드 변화가 가이드 RNA 코딩 서열에 도입될 수 있다 (아래 실시예 3 참고). tracrRNA 서열은 약 50개 뉴클레오티드 내지 약 300개 뉴클레오티드 길이 범위일 수 있다. 다양한 구체예들에서, tracrRNA는 약 50 내지 약 90개 뉴클레오티드, 약 90 내지 약 110개 뉴클레오티드, 약 110 내지 약 130개 뉴클레오티드, 약 130 내지 약 150개 뉴클레오티드, 약 150 내지 약 170개 뉴클레오티드, 약 170 내지 약 200개 뉴클레오티드, 약 200 내지 약 250개 뉴클레오티드, 또는 약 250 내지 약 300개 뉴클레오티드 길이 범위일 수 있다.
- [0050] 일반적으로, 조작된 가이드 RNA는 단일 분자 (즉, 단일 가이드 RNA 또는 sgRNA)이며, 이 때 crRNA 서열은 tracrRNA 서열에 연결된다. 그러나 일부 구체예들에서, 조작된 가이드 RNA는 2개의 별도 분자들일 수 있다. 제 1 분자는 제 2 분자의 5' 단부와 염기쌍을 이룰 수 있는 3' 서열 (약 6 내지 약 20개 뉴클레오티드를 포함)을 함유하는 crRNA를 포함하고, 이 때 제 2 분자는 제 1 분자의 3' 단부와 염기쌍을 이룰 수 있는 5' 서열 (약 6 내지 약 20개 뉴클레오티드를 포함)을 함유하는 tracrRNA를 포함한다.
- [0051] 일부 구체예들에서, 조작된 가이드 RNA의 tracrRNA 서열은 하나 이상의 압타머 서열들을 포함하도록 변형될 수 있다 (Konermann 외., Nature, 2015, 517(7536):583-588; Zalatan 외., Cell, 2015, 160(1-2):339-50). 적합한 압타머 서열들은 MCP, PCP, Com, SLBP, FXR1, AP205, BZ13, f1, f2, fd, fr, ID2, JP34/GA, JP501, JP34, JP500, KU1, M11, M12, MX1, NL95, PP7, ϕ Cb5, ϕ Cb8r, ϕ Cb12r, ϕ Cb23r, Q β , R17, SP- β , TW18, TW19, VK, 등의 단편, 또는 이들의 유도체에서 선택된 압타머 단백질들에 결합하는 서열들을 포함한다. 당업자는 압타머 길이가 달라질 수 있음을 이해한다.
- [0052] 다른 구체예들에서, 가이드 RNA는 최소한 하나의 탐지가능한 표지를 추가로 포함할 수 있다. 탐지가능한 표지는 형광단 (예컨대, FAM, TMR, Cy3, Cy5, 텍사스 레드, 오레곤 그린, 알렉사 플루오르, 할로 태그, 또는 적합한 형광 염료), 탐지 태그 (예컨대, 비오틴, 디곡시제닌, 등), 양자 점, 또는 금 입자들일 수 있다.
- [0053] 가이드 RNA는 표준 리보뉴클레오티드 및/또는 변형된 리보뉴클레오티드를 포함할 수 있다. 일부 구체예에서, 가이드 RNA는 표준 또는 변형된 데옥시리보뉴클레오티드를 포함할 수 있다. 가이드 RNA가 효소적으로 합성되는 (즉, 생체내 또는 시험관내) 구체예에서, 가이드 RNA는 일반적으로 표준 리보뉴클레오티드를 포함한다. 가이드 RNA가 화학적으로 합성되는 구체예들에서, 가이드 RNA는 표준 또는 변형된 리보뉴클레오티드 및/또는 데옥시리보뉴클레오티드를 포함할 수 있다. 변형된 리보뉴클레오티드 및/또는 데옥시리보뉴클레오티드는 염기 변형 (예컨대, 슈도우리딘, 2-티오우리딘, N6-메틸아데노신 등) 및/또는 당 변형 (예컨대, 2'-O-메틸, 2'-플루오로, 2'-아미노, 잠금 핵산 (LNA) 등)을 포함한다. 가이드 RNA의 골격은 또한 포스포로티오에이트 링키지, 보라노포스페이트 링키지 또는 펩티드 핵산을 포함하도록 변형될 수 있다.
- [0054] 특정 구체예들에서, 조작된 가이드 RNA는 서열 번호: 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 또는 45에 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95% 또는 적어도 약

99% 서열 동일성을 가질 수 있다. 일부 구체예들에서, 조작된 Cas9 가이드 RNA는 서열 번호: 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 또는 45의 서열을 가진다.

(c) PAM 서열

상기 설명한 조작된 Cas9 시스템은 새로운 PAM 서열의 상류에 위치한 이중 가닥 DNA의 특정 서열을 표적한다. 조작된 Cas9 시스템에 의해 선호되는 PAM 서열들을 축퇴성 PAM의 라이브러리를 사용하여 시험관내 식별하였으며 (실시에 1 및 도 1 참고), 게놈 편집 실험들 이후 시퀀싱하여 확인하였다 (실시에 2 참고). 본원에 개시된 조작된 Cas9 시스템 각각에 대한 PAM은 아래 표 A에 제시된다.

표 A. PAM 서열	
조작된 Cas9 시스템	PAM (5'-3')*
<i>바실루스 스미티이</i> (<i>Bacillus smithii</i>) Cas9 (BsmCas9)	NNNNCAAA
<i>락토바실루스 람노수스</i> (<i>Lactobacillus rhamnosus</i>) Cas9 (LrhCas9)	NGAAA
<i>파라수테렐라 엑스크레멘티호미니스</i> (<i>Parasutterella excrementihominis</i>) Cas9 (PexCas9)	NGG
<i>마이코플라즈마 카니스</i> (<i>Mycoplasma canis</i>) Cas9 (McaCas9)	NNGG
<i>마이코플라즈마 칼리셉티쿰</i> (<i>Mycoplasma gallisepticum</i>) Cas9 (MgaCas9)	NNAAT
<i>아커만시아 글리카니필라</i> (<i>Akkermansia glycaniphila</i>) Cas9 (AglCas9)	NNNRTA
<i>아커만시아 뮤시니필라</i> (<i>Akkermansia muciniphila</i>) Cas9 (AmuCas9)	MMACCA
<i>오에노코쿠스 키타하라에</i> (<i>Oenococcus kitaharae</i>) Cas9 (OkiCas9)	NNG
<i>비피도박테리움 볼비</i> (<i>Bifidobacterium bombi</i>) Cas9 (BboCas9)	NNNNGRY
<i>아시도테르무스 셀룰로리티쿠스</i> (<i>Acidothermus cellulolyticus</i>) Cas9 (AceCas9)	NGG
<i>알리시클로바실루스 헤스페리둠</i> (<i>Alicyclobacillus</i>)	NGG

<i>hesperidum</i>) Cas9 (AheCas9)	
울리넬라 속시노게네스 (<i>Wolinella succinogenes</i>) Cas9 (WsuCas9)	NGG
니트라티프락터 살수기니스 (<i>Nitratifactor salsuginis</i>) Cas9 (NsaCas9)	NRGNK
랄스토니아 시지기이 (<i>Ralstonia syzygii</i>) Cas9 (RsyCas9)	GGGRG
코리넨박테리움 디프테리아 (<i>Corynebacterium diphtheria</i>) Cas9 (CdiCas9)	NNAMMMC

* K는 G 또는 T이고; M은 A 또는 C이고; R은 A 또는 G이고; Y는 C 또는 T이고; 그리고 N은 A, C, G, 또는 T이다.

[0058] (II) 핵산

[0059] 본 발명의 또 다른 양상은 섹션 (I)에 상기 기재된 조작된 Cas9 시스템을 인코딩하는 핵산을 제공한다. 상기 시스템은 단일 핵산 또는 복수의 핵산들에 의해 인코딩 될 수 있다. 핵산은 DNA 또는 RNA, 선형 또는 원형, 단일-가닥 또는 이중-가닥 일 수 있다. RNA 또는 DNA는 관심 진핵 세포에서 단백질로의 효율적인 번역을 위해 코돈 최적화될 수 있다. 코돈 최적화 프로그램은 프리웨어로서 또는 상업적 공급업체들로부터 이용가능하다.

[0060] 일부 구체예들에서, 핵산은 서열 번호: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 또는 30의 아미노산 서열에 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95% 또는 적어도 약 99% 서열 동일성을 가지는 단백질을 인코딩한다. 특정 구체예들에서, 조작된 Cas9 단백질을 인코딩하는 핵산은 서열 번호:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 또는 29의 DNA 서열에 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95% 또는 적어도 약 99% 서열 동일성을 가질 수 있다. 특정 구체예에서, 조작된 Cas9 단백질을 인코딩하는 DNA는 서열 번호: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 또는 29의 DNA 서열을 가진다. 추가 구체예에서, 상기 핵산은 서열 번호: 117, 118, 119, 120, 121, 122, 123, 또는 124의 아미노산 서열에 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95%, 또는 적어도 약 99% 서열 동일성을 가지는 단백질을 인코딩한다.

[0061] 일부 구체예에서, 조작된 Cas9 단백질을 인코딩하는 핵산은 RNA일 수 있다. RNA는 *시험관내*에서 효소적으로 합성될 수 있다. 이를 위하여, 조작된 Cas9 단백질을 인코딩하는 DNA는 *시험관내* RNA 합성을 위해 파지 RNA 중합 효소에 의해 인식되는 프로모터 서열에 작동적으로 연결될 수 있다. 예를 들면, 프로모터 서열은 T7, T3, 또는 SP6 프로모터 서열 또는 T7, T3, 또는 SP6 프로모터 서열의 변이 일 수 있다. 조작된 단백질을 인코딩하는 DNA는 아래 상세히 설명된 벡터의 일부 일 수 있다. 이러한 구체예들에서, *시험관내*-전사된 RNA는 정제, 캡핑, 및 /또는 폴리아데닐화 될 수 있다. 다른 구체예들에서, 조작된 Cas9 단백질을 인코딩하는 RNA는 자가-복제 RNA의 일부 일 수 있다 (Yoshioka 외., Cell Stem Cell, 2013, 13: 246-254). 자가-복제 RNA는 비감염성, 자가-복제 Venezuelan 말 뇌염 (VEE) 바이러스 RNA 레플리콘으로부터 유래될 수 있는데, 이는 제한된 세포 분열수를 위하여 자가-복제할 수 있는 양성-센스, 단일-가닥 RNA로서, 관심 단백질들을 코딩화하기 위해 변형될 수 있다 (Yoshioka 외., Cell Stem Cell, 2013, 13: 246-254).

[0062] 다른 구체예에서, 조작된 Cas9 단백질을 인코딩하는 핵산은 DNA 일 수 있다. DNA 코딩 서열은 관심 세포에서 발현을 위해 최소한 하나의 프로모터 제어 서열에 작동적으로 연결될 수 있다. 특정 구체예들에서, DNA 코딩 서열은 박테리아 (*예컨대*, *대장균*) 세포 또는 진핵생물 (*예컨대*, 효모, 곤충, 또는 포유동물) 세포에서 조작된 Cas9 단백질의 발현을 위해 프로모터 서열에 작동적으로 연결될 수 있다. 적합한 박테리아 프로모터에는, 제한 없이, T7 프로모터, *lac* 오페론 프로모터, *trp* 프로모터, *tac* 프로모터 (이들은 *trp*과 *lac* 프로모터의 하이브리드임), 전술한 것들 중 어느 하나의 변이형, 및 전술한 것들의 임의의 조합이 포함된다. 적합한 진핵생물 프로모터의 비-제한적 예들에는 항시성의, 조절된 또는 세포- 또는 조직-특이적 프로모터가 포함된다. 적합한 진핵생물 항시적 프로모터 제어 서열에는, 거대세포바이러스 즉시 초기 프로모터 (CMV), 원숭이 바이러스 (SV40) 프로모터, 아데노바이러스 주요 후기 프로모터, 라우스 육종 바이러스 (RSV) 프로모터, 마우스 쥘 중앙 바이러스

(MMTV) 프로모터, 포스포글리세레이트 키나제 (PGK) 프로모터, 연장 인자 (ED1)-알파 프로모터, 유비퀴틴 프로모터, 액틴 프로모터, 튜블린 프로모터, 면역글로불린 프로모터, 이의 단편, 또는 전술한 것들의 임의의 조합이 포함되나, 이에 제한되는 것은 아니다. 적합한 진핵생물 조절된 프로모터 제어 서열의 예들에는 제한 없이 열충격, 금속, 스테로이드, 항생제, 또는 알콜에 의해 조절되는 것들이 포함된다. 조직-특이적 프로모터의 비-제한적 예들에는 B29 프로모터, CD14 프로모터, CD43 프로모터, CD45 프로모터, CD68 프로모터, 데스민 프로모터, 엘라스타제-1 프로모터, 엔도글린 프로모터, 피브로넥틴 프로모터, Flt-1 프로모터, GFAP 프로모터, GPIIb 프로모터, ICAM-2 프로모터, INF- α 프로모터, Mb 프로모터, NphsI 프로모터, OG-2 프로모터, SP-B 프로모터, SYN1 프로모터, 및 WASP 프로모터가 포함된다. 프로모터 서열은 야생형 일 수 있거나 보다 효율적인 또는 효과적인 발현을 위해 변형될 수 있다. 일부 구체예들에서, DNA 코딩 서열은 또한 폴리아데닐화 신호 (*예컨대*, SV40 폴리A 신호, 소 성장 호르몬 (BGH) 폴리A 신호, 등) 및/또는 최소한 하나의 전사 종결 서열에 연결될 수 있다. 어떤 상황에서는 조작된 Cas9 단백질이 박테리아 또는 진핵 세포에서 정제 될 수 있다.

[0063] 또 다른 구체예에서, 조작된 가이드 RNA는 DNA에 의해 인코딩 될 수 있다. 일부 예들에서, 조작된 가이드 RNA를 인코딩하는 DNA는 시험관내 RNA 합성을 위해 파지 RNA 중합효소에 의해 인식되는 프로모터 서열에 작동적으로 연결될 수 있다. 예를 들면, 프로모터 서열은 T7, T3, 또는 SP6 프로모터 서열 또는 T7, T3, 또는 SP6 프로모터 서열의 변이 일 수 있다. 다른 예에서, 조작된 가이드 RNA를 인코딩하는 DNA는 관심 진핵 세포에서의 발현을 위해 RNA 중합효소 III (Pol III)에 의해 인식되는 프로모터 서열에 작동가능하게 연결될 수 있다. 적합한 Pol III 프로모터의 예들에는, 포유동물 U6, U3, H1, 및 7SL RNA 프로모터가 포함되나, 이에 제한되는 것은 아니다.

[0064] 다양한 구체예에서, 조작된 Cas9 단백질을 인코딩하는 핵산은 벡터에 존재할 수 있다. 일부 구체예에서, 벡터는 조작된 가이드 RNA를 인코딩하는 핵산을 추가로 포함 할 수 있다. 적합한 벡터들에는 플라스미드 벡터, 바이러스 벡터, 및 자가-복제 RNA가 포함된다 (Yoshioka *외.*, Cell Stem Cell, 2013, 13: 246-254). 일부 구체예들에서, 복합체 또는 융합 단백질을 인코딩하는 핵산은 플라스미드 벡터에 존재할 수 있다. 적합한 플라스미드 벡터의 비-제한적 예들에는 pUC, pBR322, pET, pBluescript, 및 이의 변이체들이 포함된다. 다른 구체예들에서, 복합체 또는 융합 단백질을 인코딩하는 핵산은 바이러스 벡터의 일부 일 수 있다 (*예컨대*, 렌티바이러스 벡터, 아데노-연관 바이러스 벡터, 아데노바이러스 벡터, 등). 플라스미드 또는 바이러스 벡터는 또 다른 발현 제어 서열 (*예컨대*, 인핸서 서열, 코작 서열, 폴리아데닐화 서열, 전사 종결 서열, 등.), 선별 마커 서열 (*예컨대*, 항생제 내성 유전자), 복제 원점, 등을 포함할 수 있다. 벡터 및 이의 용도에 관한 추가 정보는 "Current Protocols in Molecular Biology" Ausubel *외.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001에서 찾을 수 있다.

[0065] (III) 진핵 세포

[0066] 본 발명의 또 다른 양상은 상기 섹션 (I)에 설명된 적어도 하나의 조작된 Cas9 시스템 및/또는 상기 섹션 (II)에 설명된 조작된 Cas9 단백질 및/또는 조작된 가이드 RNA를 인코딩하는 적어도 하나의 핵산을 포함하는 진핵 세포를 포함한다.

[0067] 진핵 세포는 인간 세포, 비-인간 포유동물 세포, 비-포유동물 척추동물 세포, 무척추동물 세포, 식물 세포, 또는 단세포 진핵 유기체일 수 있다. 적합한 진핵 세포들의 예들은 상기 섹션 (IV)(c)에 설명된다. 진핵 세포는 *시험관내*, *세포외*, 또는 *세포내* 일 수 있다.

[0068] (IV) 염색체 서열 변형 방법

[0069] 본 발명의 추가 양상은 진핵 세포에서 염색체 서열을 변형하는 방법을 포함한다. 일반적으로, 이 방법은 섹션 (I)에서 상술한 적어도 하나의 조작된 Cas9 시스템 및/또는 섹션 (II)에서 상술한 상기 조작된 Cas9 시스템을 인코딩하는 적어도 하나의 핵산을 관심 진핵 세포에 도입하는 것을 포함한다.

[0070] 조작된 Cas9 단백질이 뉴클레아제 또는 니카아제 활성을 포함하는 구체예에서, 염색체 서열 변형은 적어도 하나의 뉴클레오티드의 치환, 적어도 하나의 뉴클레오티드의 결실, 적어도 하나의 뉴클레오티드의 삽입을 포함 할 수 있다. 일부 구체예들에서, 상기 방법은 뉴클레아제 활성을 포함하는 하나의 조작된 Cas9 시스템 또는 니카아제 활성을 포함하고 공여체 폴리뉴클레오티드가 없는 2개의 조작된 Cas9 시스템을 진핵 세포에 도입하는 것을 포함하고, 이로써 조작된 Cas9 시스템 또는 시스템들은 염색체 서열의 표적 부위에 이중-가닥 절단을 도입하고 세포 DNA 복구 과정에 의한 이중-가닥 절단의 복구는 적어도 하나의 뉴클레오티드 변화 (즉, 삽입결실)를 도입

함으로써, 염색체 서열을 비활성화 (즉, 유전자 녹아웃) 시킨다. 다른 구체예들에서, 상기 방법은 뉴클레아제 활성을 포함하는 하나의 조작된 Cas9 시스템 또는 니카아제 활성, 그리고 공여체 폴리뉴클레오티드를 포함하는 2개의 조작된 Cas9 시스템을 진핵 세포에 도입하는 것을 포함하고, 이로써 조작된 Cas9 시스템 또는 시스템들은 염색체 서열의 표적 부위에 이중-가닥 절단을 도입하고 세포 DNA 복구 과정에 의한 이중-가닥 절단의 복구는 공여체 폴리뉴클레오티드 서열을 염색체 서열의 표적 부위 내부에 삽입 또는 교환하는 결과를 가져온다.

[0071] 조작된 Cas9 단백질이 후성 변형 활성 또는 전사 조절 활성을 포함하는 구체예들에서, 염색체 서열 변형은 염색체 서열의 표적 부위에서 또는 근방에서 적어도 하나의 뉴클레오티드의 전환, 표적 부위에서 또는 근방에서 적어도 하나의 뉴클레오티드의 변형, 표적 부위에서 또는 근방에서 적어도 하나의 히스티딘의 변형, 및/또는 표적 부위에서 또는 근방에서 전사의 변화를 포함할 수 있다.

[0072] (a) 세포로의 도입

[0073] 상기 언급한 바와 같이, 상기 방법은 적어도 하나의 조작된 Cas9 시스템 및/또는 상기 시스템을 인코딩하는 핵산 (및 선택적인 공여체 폴리뉴클레오티드)을 진핵 세포에 도입하는 것을 포함한다. 적어도 하나의 시스템 및/또는 핵산/공여체 폴리뉴클레오티드는 다양한 수단에 의해 관심 세포 내부에 도입될 수 있다.

[0074] 일부 구체예들에서, 상기 세포는 적절한 분자들 (즉, 단백질, DNA, 및/또는 RNA)로 형질감염 될 수 있다. 적합한 형질감염 방법들에는 뉴클레오펙션 (또는 전기전공), 칼슘 포스페이트-매개된 형질감염, 양이온 폴리머 형질감염 (예컨대, DEAE-덱스트란 또는 폴리에틸렌이민), 바이러스 형질도입, 비로솜 형질감염, 비리온 형질감염, 리포솜 형질감염, 양이온 리포솜 형질감염, 면역리포솜 형질감염, 비리포솜 지질 형질감염, 덴드리머 형질감염, 열 충격 형질감염, 자기주입법, 리포펙션, 유전자 총 전달, 임팔레క్ష션, 초음파천공법, 광학 형질감염, 및 특정 상표 제제 (proprietary agent)-개선된 핵산의 흡수가 포함된다. 형질감염 방법들은 해당 분야에 잘 공지되어 있다 (예컨대, "Current Protocols in Molecular Biology" Ausubel 외., John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001 참조). 다른 구체예들에서, 상기 분자들은 미세주입에 의하여 세포 내부에 도입될 수 있다. 예를 들면, 상기 분자들은 관심 세포들의 세포질 또는 핵 내부에 주사될 수 있다. 세포 내부에 도입되는 각 분자의 양은 변화할 수 있으나, 해당 분야의 숙련된 기술자는 적절한 양을 결정하는 수단을 잘 알고 있다.

[0075] 다양한 분자들은 동시에 또는 순차적으로 세포 내부에 도입될 수 있다. 예를 들어, 조작된 Cas9 시스템 (또는 이의 인코딩 핵산)과 공여체 폴리뉴클레오티드는 동시에 도입 될 수 있다. 대안적으로, 하나를 먼저 도입한 다음 다른 하나를 차후에 세포 내에 도입할 수 있다.

[0076] 일반적으로, 세포는 세포 성장 및/또는 유지에 적절한 조건하에서 유지된다. 적합한 세포 배양 조건들은 해당 분야에 널리 공지이며, 예를 들면, Santiago 외., Proc. Natl. Acad. Sci. USA, 2008, 105:5809-5814; Moehle 외., Proc. Natl. Acad. Sci. USA, 2007, 104:3055-3060; Urnov 외., Nature, 2005, 435:646-651; 및 Lombardo 외., Nat. Biotechnol., 2007, 25:1298-1306에 기재되어 있다. 해당 분야의 숙련된 기술자들은 세포의 배양 방법들이 해당 분야에 공지이며 세포 유형에 따라 달라질 것임을 이해하고 있을 것이다. 모든 사례에서, 특정 세포 유형에 대한 가장 우수한 기술들을 결정하기 위해 관례적인 최적화가 사용될 수 있다.

[0077] (b) 선택적 공여체 폴리뉴클레오티드

[0078] 조작된 Cas9 단백질이 뉴클레아제 또는 니카아제 활성을 포함하는 구체예들에서, 상기 방법은 최소한 하나의 공여체 폴리뉴클레오티드를 세포에 도입하는 단계를 추가로 포함할 수 있다. 공여체 폴리뉴클레오티드는 단일-가닥 또는 이중-가닥, 선형 또는 원형, 및/또는 RNA 또는 DNA 일 수 있다. 일부 구체예들에서, 공여체 폴리뉴클레오티드는 벡터, 예컨대, 플라스미드 벡터일 수 있다.

[0079] 공여체 폴리뉴클레오티드는 최소한 하나의 공여체 서열을 포함한다. 일부 양상들에서, 공여체 폴리뉴클레오티드의 공여체 서열은 내인성 또는 고유 염색체 서열의 변형된 형태일 수 있다. 예를 들면, 공여체 서열은 조작된 Cas9 시스템에 의해 표적되는 서열의 또는 서열 근방의 염색체 서열의 일부분과 실질적으로 동일할 수 있지만, 이는 최소한 하나의 뉴클레오티드 변화를 포함한다. 그러므로, 고유 서열과 통합 또는 교환시, 표적된 염색체 위치의 서열은 최소한 하나의 뉴클레오티드 변화를 포함한다. 예를 들면, 이러한 변화는 하나 이상의 뉴클레오티드 삽입, 하나 이상의 뉴클레오티드 결실, 하나 이상의 뉴클레오티드 치환, 또는 이의 조합일 수 있다. 변형된 서열의 "유전자 교정" 통합의 결과로서, 세포는 표적된 염색체 서열로부터 변형된 유전자 생성물을 생성할 수 있다.

- [0080] 다른 양상들에서, 공여체 폴리뉴클레오타이드의 공여체 서열은 외인성 서열일 수 있다. 본 명세서에서 사용되는, "외인성" 서열은 세포에 대해 고유적이지 않은 서열, 또는 그 고유 위치가 세포의 게놈에서 상이한 위치에 존재하는 서열을 지칭한다. 예를 들면, 외인성 서열은 단백질 코딩 서열을 포함할 수 있는데, 이는 외인성 프로모터 제어 서열에 작동적으로 연결될 수 있으므로, 게놈에 통합시, 세포는 통합된 서열에 의해 코드되는 단백질을 발현할 수 있다. 대안적으로, 외인성 서열은 염색체 서열에 통합될 수 있으며, 그리하여 그 발현은 내인성 프로모터 제어 서열에 의해 조절된다. 다른 반복들에서, 외인성 서열은 전사 제어 서열, 또 다른 발현 제어 서열, RNA 코딩 서열, 동일 수 있다. 상기 언급한 바와 같이, 외인성 서열의 염색체 서열로의 통합은 "녹인"이라 명명한다.
- [0081] 해당 분야의 숙련된 기술자가 이해할 수 있는 바와 같이, 공여체 서열의 길이는 달라질 수 있으며 달라질 것이다. 예를 들면, 공여체 서열은 수개의 뉴클레오타이드에서 수백개의 뉴클레오타이드까지의 길이에서 수십만개 뉴클레오타이드까지 변화할 수 있다.
- [0082] 전형적으로, 공여체 폴리뉴클레오타이드의 공여체 서열은 상류 서열 및 하류 서열에 의해 연결되고, 이는 각각 조작된 Cas9 시스템에 의해 표적되는 서열의 상류 및 하류에 위치한 서열들과 실질적 서열 동일성을 가진다. 이들 서열 유사성으로 인해, 공여체 폴리뉴클레오타이드의 상류 및 하류 서열들은 공여체 폴리뉴클레오타이드와 표적되는 염색체 서열 간의 상동 재조합을 가능하게 하여 공여체 서열이 염색체 서열 내에 통합 (또는 염색체 서열과 교환) 될 수 있다.
- [0083] 본 명세서에서 사용되는 상류 서열은, 조작된 Cas9 시스템에 의해 표적되는 서열의 상류 염색체 서열과 실질적인 서열 동일성을 공유하는 핵산 서열을 지칭한다. 유사하게, 하류 서열은 조작된 Cas9 시스템에 의해 표적되는 서열의 하류 염색체 서열과 실질적인 서열 동일성을 공유하는 핵산 서열을 지칭한다. 본 명세서에서 사용되는, 어구 "실질적 서열 동일성"은 최소한 약 75% 서열 동일성을 가지는 서열들을 지칭한다. 그러므로, 공여체 폴리뉴클레오타이드 내 상류 및 하류 서열들은 표적 서열에 대한 상류 또는 하류 서열과 약 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 가질 수 있다. 한 예시적 구체예에서, 공여체 폴리뉴클레오타이드 내 상류 및 하류 서열들은 조작된 Cas9 시스템에 의해 표적되는 서열에 대한 상류 또는 하류 염색체 서열과 약 95% 또는 100% 서열 동일성을 가질 수 있다.
- [0084] 일부 구체예들에서, 상류 서열은 조작된 Cas9 시스템에 의해 표적되는 서열의 바로 상류에 위치한 염색체 서열과 실질적 서열 동일성을 공유한다. 다른 구체예들에서, 상류 서열은 표적 서열로부터 약 백 (100)개 뉴클레오타이드 상류 이내에 위치한 염색체 서열과 실질적 서열 동일성을 공유한다. 그러므로, 예를 들면, 상류 서열은 표적 서열로부터 약 1 내지 약 20, 약 21 내지 약 40, 약 41 내지 약 60, 약 61 내지 약 80, 또는 약 81 내지 약 100 뉴클레오타이드 상류에 위치한 염색체 서열과 실질적 서열 동일성을 공유할 수 있다. 일부 구체예들에서, 하류 서열은 조작된 Cas9 시스템에 의해 표적되는 서열의 바로 하류에 위치한 염색체 서열과 실질적 서열 동일성을 공유한다. 다른 구체예들에서, 하류 서열은 표적 서열로부터 약 백 (100)개 뉴클레오타이드 하류 이내에 위치한 염색체 서열과 실질적 서열 동일성을 공유한다. 그러므로, 예를 들면, 하류 서열은 표적 서열로부터 약 1 내지 약 20, 약 21 내지 약 40, 약 41 내지 약 60, 약 61 내지 약 80, 또는 약 81 내지 약 100 뉴클레오타이드 하류에 위치한 염색체 서열과 실질적 서열 동일성을 공유할 수 있다.
- [0085] 각 상류 또는 하류 서열은 약 20개 뉴클레오타이드 내지 약 5000개 뉴클레오타이드 길이 범위일 수 있다. 일부 구체예들에서, 상류 및 하류 서열들은 약 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, 또는 5000개 뉴클레오타이드를 포함할 수 있다. 특정 구체예들에서, 상류 및 하류 서열들은 약 50 내지 약 1500개 뉴클레오타이드 길이 범위일 수 있다.
- [0086] **(c) 세포 유형**
- [0087] 다양한 진핵 세포들이 본 발명에 개시된 방법에서 사용하기에 적합하다. 예를 들면, 세포는 인간 세포, 비-인간 포유동물 세포, 비-포유동물 척추동물 세포, 무척추동물 세포, 곤충 세포, 식물 세포, 효모 세포, 또는 단세포 진핵생물 기관일 수 있다. 일부 구체예들에서, 세포는 하나의 세포 배아 일 수도 있다. 예를 들면, 비-인간 포유동물 배아는 쥐, 햄스터, 설치류, 토끼, 고양이, 개, 양, 돼지, 소, 말, 및 영장류 배아를 포함한다. 또 다른 구체예들에서, 세포는 줄기 세포, 가령, 배아 줄기 세포, ES-유사 줄기 세포, 태아 줄기 세포, 성체 줄기 세포, 동일 수 있다. 한 구체예에서, 줄기 세포는 인간 배아 줄기 세포가 아니다. 더욱이, 줄기 세포들은 그 전문이 본 명세서에 포함되는 W02003/046141, 또는 Chung 외. (Cell 줄기 세포, 2008, 2:113-117)에 개시된

기술들에 의해 제조되는 것들을 포함할 수 있다. 세포는 시험관내 (즉, 배양물 내), 생체의 (즉, 유기체로부터 분리된 조직 내) 또는 생체내 (즉, 유기체내) 존재할 수 있다. 예시적 구체예들에서, 세포는 포유동물 세포 또는 포유동물 세포주이다. 특정 구체예들에서, 세포는 인간 세포 또는 인간 세포주이다.

[0088] 적합한 포유동물 세포 또는 세포주들의 비-제한적 예들에는 인간 배아 신장 세포 (HEK293, HEK293T); 인간 자궁 경부 암종 세포 (HELA); 인간 폐 세포 (W138); 인간 간 세포 (Hep G2); 인간 U2-OS 골육종 세포, 인간 A549 세포, 인간 A-431 세포, 및 인간 K562 세포; 중국 햄스터 난소 (CHO) 세포, 새끼 햄스터 신장 (BHK) 세포; 생쥐 골수종 NS0 세포, 생쥐 배아 섬유모세포 3T3 세포 (NIH3T3), 생쥐 B 림프종 A20 세포; 생쥐 흑색종 B16 세포; 생쥐 근육모세포 C2C12 세포; 생쥐 골수종 SP2/0 세포; 생쥐 배아 중간엽 C3H-10T1/2 세포; 생쥐 암종 CT26 세포, 생쥐 전립선 DuCuP 세포; 생쥐 유방 EMT6 세포; 생쥐 간암 Hepa1c1c7 세포; 생쥐 골수종 J5582 세포; 생쥐 상피 MTD-1A 세포; 생쥐 심근 MyEnd 세포; 생쥐 신장 RenCa 세포; 생쥐 이자 RIN-5F 세포; 생쥐 흑색종 X64 세포; 생쥐 림프종 YAC-1 세포; 쥐 교모세포종 9L 세포; 쥐 B 림프종 RBL 세포; 쥐 신경모세포종 B35 세포; 쥐 간암 세포 (HTC); 버팔로 쥐 간 BRL 3A 세포; 개 신장 세포 (MDCK); 개 유선 (CMT) 세포; 쥐 골육종 D17 세포; 쥐 단핵구/대식세포 DH82 세포; 원숭이 신장 SV-40 형질전환된 섬유모세포 (COS7) 세포; 원숭이 신장 CVI-76 세포; 아프리카 녹색 원숭이 신장 (VERO-76) 세포가 포함된다. 포유동물 세포주의 보다 많은 목록은 미국 표준 균주 카탈로그 (ATCC, Manassas, VA)에서 찾을 수 있다.

[0089] (V) 응용

[0090] 본 명세서에 개시된 조성물 및 방법들은 다양한 치료, 진단, 산업 및 연구 분야에서 사용될 수 있다. 일부 구체예들에서, 본 발명은 유전자 기능을 모형화 및/또는 연구하기 위해, 관심의 유전적 또는 후성 조건들을 연구하기 위해, 또는 다양한 질병 또는 장애들에 관여하는 생화학적 경로들을 연구하기 위해 세포, 동물, 또는 식물에서 임의의 관심 염색체 서열을 변형하기 위해 사용될 수 있다. 예를 들면, 질병 또는 장애들을 모형화하는 유전자삽입 유기체가 생성될 수 있으며, 여기서 질병 또는 장애와 연관된 하나 이상의 핵산 서열들의 발현은 변화되어 있다. 질병 모델은 유기체에 대한 돌연변이들의 효과를 연구하기 위해, 질병의 발병 및/또는 진행을 연구하기 위해, 질병에 대한 제약학적 활성 화합물의 효과를 연구하기 위해, 및/또는 잠재적인 유전자 치료 전략의 효능을 평가하기 위해 사용될 수 있다.

[0091] 다른 구체예들에서, 상기 조성물 및 방법들은 효율적인 그리고 비용 효과적인 기능적 게놈 선별을 실시하기 위해 사용될 수 있으며, 특정 생물학적 과정에 관여하는 유전자들의 기능 그리고 유전자 발현에서 임의의 변형이 생물학적 과정에 어떻게 영향을 줄 수 있는지를 연구하기 위해, 또는 세포 표현형과 관련된 게놈 좌위의 포화 또는 딥 스캐닝 돌연변이생성을 실시하기 위해 사용될 수 있다. 예를 들면, 유전자 발현, 약물 내성, 및 질병의 역전에 필요한 기능적 요소들에 관한 중요한 최소 특징들 및 별개의 취약성들을 결정하기 위하여 포화 또는 딥 스캐닝 돌연변이생성이 사용될 수 있다.

[0092] 추가 구체예들에서, 본 명세서에 개시된 조성물 및 방법들은 질병 또는 장애의 존재를 확인하기 위한 진단 테스트에 및/또는 치료 옵션 결정에 사용하기 위해 사용될 수 있다. 적합한 진단 테스트들의 예는 암 세포에서 특정 돌연변이들 (예컨대, EGFR, HER2, 등에서 특정 돌연변이)의 탐지, 특정 질병들과 관련된 특정 돌연변이들의 탐지 (예컨대, 트라이뉴클레오티드 반복, 낫 적혈구병과 연관된 β -글로빈의 돌연변이들, 특정 SNP, 등), 간염의 탐지, 바이러스 (예컨대, Zika)의 탐지, 등을 포함한다.

[0093] 추가 구체예들에서, 본 명세서에 개시된 조성물 및 방법들은 특정 질병 또는 장애와 연관된 유전자 돌연변이들을 교정, 가령, 예컨대, 낫 적혈구병 또는 지중해빈혈과 연관된 글로빈 유전자 돌연변이들을 교정, 중증 복합성 면역 결핍증 (SCID)과 연관된 아데노신 탈아미노효소 유전자에서 돌연변이들을 교정, 헌팅턴 병의 질병-유발 유전자인 HTT의 발현을 감소, 또는 망막 색소변성의 치료를 위해 로돕신 유전자에서의 돌연변이들을 교정함에 사용될 수 있다. 이러한 변형들은 생체의 세포에서 이루어질 수 있다.

[0094] 또 다른 구체예들에서, 본 명세서에 개시된 조성물 및 방법들은 환경적 스트레스에 대한 내성이 증가된 또는 개선된 형질들을 가지는 작물들을 생성하기 위해 사용될 수 있다. 본 발명은 또한 개선된 형질을 가진 농장 동물 또는 생산 동물을 생성하기 위해 사용될 수도 있다. 예를 들면, 돼지는 생의학 모델, 특히, 재생 의학 또는 이종이식에서의 모델로서 매력적이게 하는 많은 특징들을 가진다.

[0095] 정의

[0096] 달리 정의되지 않는 한, 본 명세서에서 사용되는 모든 기술 및 과학 용어는 본 발명이 속하는 해당 분야의 기술자에 의해 일반적으로 이해되는 의미를 가진다. 다음 참고문헌들은 본 발명에서 사용되는 많은 용어들의 일반

적인 정의를 숙련된 기술자에게 제공한다: Singleton 외., 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 외. (eds.), Springer Verlag (1991); 및 Hale & Marham, The Harper Collins Dictionary of Biology (1991). 본 출원에서 사용되는, 하기 용어들은 달리 특정한 언급이 없는 한 하기 의미들을 가진다.

- [0097] 본 발명 또는 바람직한 구체예(들)의 구성요소들을 소개할 때, 관사 "하나 (a, an)", "그것" 및 "상기"는 하나 이상의 구성요소들이 존재함을 의미하는 것이다. 용어 "포함하는", "비롯한" 및 "가지는"은 포함적인 의미이며 나열된 구성요소들 이외에 추가 구성요소들이 존재할 수 있음을 의미한다.
- [0098] 수치값과 관련하여 사용될 때 용어 "약" x는, 예를 들면 $x \pm 5\%$ 를 의미한다.
- [0099] 본 명세서에서 사용되는 용어 "상보적인" 또는 "상보성"은 특정 수소 결합을 통한 염기 페어링에 의한 이중-가닥 핵산들의 결합을 지칭한다. 염기 페어링은 표준 왓슨-크릭 염기 페어링일 수 있다 (예컨대, 5'-A G T C-3'은 상보적 서열 3'-T C A G-5'과 쌍을 이룬다). 염기 페어링은 또한 후그스틴 또는 역 후그스틴 수소 결합일 수 있다. 상보성은 일반적으로 이중나선 영역에 대해 측정되므로, 예를 들면, 오버행은 제외된다. 이중나선 영역의 두 개 가닥들 간의 상보성은 부분적일 수 있으며 염기들 중 일부만 (예컨대, 70%) 상보적인 경우 백분율로 표현된다 (예컨대, 70%). 상보적이지 않은 염기들은 "미스매치"된다. 상보성은 또한 이중나선 영역의 모든 염기들이 상보적인 경우 완전할 수도 있다 (즉, 100%).
- [0100] 본 명세서에서 사용되는 용어 "CRISPR/Cas 시스템" 또는 "Cas9 시스템"은 Cas9 단백질 (즉, 뉴클레아제, 니카아제, 또는 촉매적 사멸 단백질) 및 가이드 RNA를 포함하는 복합체를 지칭한다.
- [0101] 본 명세서에서 사용되는 용어 "내인성 서열"은 세포에 대해 고유한 염색체 서열을 지칭한다.
- [0102] 본 명세서에서 사용되는 용어 "외인성"은 세포에 대해 고유하지 않은 서열 또는 세포의 게놈에서 그 고유 위치가 상이한 염색체 위치에 존재하는 염색체 서열을 지칭한다.
- [0103] 본 명세서에서 사용되는 용어 "유전자"는 유전자 생성물을 인코딩하는 DNA 영역 (엑손과 인트론 포함), 뿐만 아니라 조절 서열들이 코딩 및/또는 전사된 서열들에 인접하는지와 관계없이 유전자 생성물의 생성을 조절하는 모든 DNA 영역을 지칭한다. 따라서, 유전자는 프로모터 서열, 터미네이터, 번역 조절 서열, 가령, 리보솜 결합 부위 및 내부 리보솜 진입 부위, 인핸서, 슬라이서, 절연체, 경계 요소, 복제 원점, 기질 부착 부위 및 좌위 제어 영역을 포함하나 이에 제한되는 것은 아니다.
- [0104] 용어 "이중"은 관심 세포에 대해 내인성이 아닌 또는 고유하지 않은 엔터티를 지칭한다. 예를 들면, 이중 단백질은 외인성 출처, 가령, 외인적으로 도입된 핵산 서열로부터 유래한 또는 본래 유래하였던 단백질을 지칭한다. 일부 사례들에서, 이중 단백질은 관심 세포에 의하여 일반적으로 생성되지 않는다.
- [0105] 용어 "니카아제"는 이중-가닥 핵산 서열 중 한 가닥을 절단하는 (즉, 이중-가닥 서열을 절단시키는) 효소를 지칭한다. 예를 들면, 이중 가닥 절단 활성을 가지는 뉴클레아제는 니카아제로서 기능하여 이중-가닥 서열 중 한 가닥만을 절단하도록 하기 위한 돌연변이 및/또는 결실에 의해 변형될 수 있다.
- [0106] 본 명세서에서 사용되는 용어 "뉴클레아제"는 이중-가닥 핵산 서열 두 가닥 모두를 절단하는 효소를 지칭한다.
- [0107] 용어 "핵산" 및 "폴리뉴클레오티드"는 선형 또는 원형 입체형태의, 그리고 단일- 또는 이중-가닥 형태의 데옥시 리보뉴클레오티드 또는 리보뉴클레오티드 폴리머를 지칭한다. 본 출원의 목적에서, 이들 용어들은 폴리머의 길이에 대한 제한으로 해석되어서는 안된다. 이 용어들은 자연 뉴클레오티드, 뿐만 아니라 염기, 당 및/또는 포스페이트 모이어티 (예컨대, 포스포로티오에이트 골격)에서 변형되어 있는 뉴클레오티드의 공지된 유사체를 포함할 수 있다. 일반적으로, 특정 뉴클레오티드의 유사체는 동일한 염기-쌍 특이성을 가진다; 즉, A의 유사체는 T와 염기-쌍을 이루게 될 것이다.
- [0108] 용어 "뉴클레오티드"는 데옥시리보뉴클레오티드 또는 리보뉴클레오티드를 지칭한다. 뉴클레오티드는 표준 뉴클레오티드 (즉, 아데노신, 구아노신, 시티딘, 티미딘, 및 우라딘), 뉴클레오티드 이성질체, 또는 뉴클레오티드 유사체일 수 있다. 뉴클레오티드 유사체는 변형된 퓨린 또는 피리미딘 염기 또는 변형된 리보오스 모이어티를 가지는 뉴클레오티드를 지칭한다. 뉴클레오티드 유사체는 자연 발생 뉴클레오티드 (예컨대, 이노신, 수도우리딘, 등) 또는 비-자연 발생 뉴클레오티드 일 수 있다. 뉴클레오티드의 당 또는 염기 모이어티에 대한 변형의 비-제한적 예들에는 아세틸 그룹, 아미노 그룹, 카르복실 그룹, 카르복시메틸 그룹, 하이드록실 그룹, 메틸 그룹, 포스포릴 그룹, 및 싸이올 그룹의 부가 (또는 제거), 뿐만 아니라 염기의 탄소 및 질소 원자들의 다른 원자

들로의 치환 (예컨대, 7-테아자 퓨린)이 포함된다. 뉴클레오타이드 유사체에는 또한 다이테옥시 뉴클레오타이드, 2'-O-메틸 뉴클레오타이드, 잠금 핵산 (LNA), 펩티드 핵산 (PNA), 및 모르포린이 포함된다.

[0109] 용어 "폴리펩티드" 및 "단백질"은 호환적으로 사용되어 아미노산 잔기들의 폴리머를 지칭한다.

[0110] 용어 "표적 서열", "표적 염색체 서열" 및 "표적 부위"는 호환적으로 사용되어 조작된 Cas9 시스템이 표적되는 염색체 DNA 내 특정 서열, 및 조작된 Cas9 시스템이 DNA 또는 DNA와 관련된 단백질(들)을 변형시키는 부위를 지칭한다.

[0111] 핵산 및 아미노산 서열 동일성을 결정하는 기술들은 해당 분야에 공지되어 있다. 전형적으로, 이러한 기술들은 유전자에 대한 mRNA의 뉴클레오타이드 서열을 결정하는 단계 및/또는 이에 의해 인코딩되는 아미노산 서열을 결정하는 단계, 및 이들 서열들을 제 2 뉴클레오타이드 또는 아미노산 서열과 비교하는 단계를 포함한다. 또한 게놈 서열들을 결정하고 이러한 방식으로 비교할 수 있다. 일반적으로, 동일성은 두 개의 폴리뉴클레오타이드 또는 폴리펩티드 서열 각각의 정확한 뉴클레오타이드-대-뉴클레오타이드 또는 아미노산-대-아미노산 상응성을 지칭한다. 둘 이상의 서열들 (폴리뉴클레오타이드 또는 아미노산)은 그 퍼센트 동일성을 결정함으로써 비교될 수 있다. 2개 서열들, 핵산 또는 아미노산 서열들의 동일성 백분율은, 2개의 정렬된 서열들 간의 일치 정합수(exact matches)를 더 짧은 서열들의 길이로 나누고 100을 곱한 것이다. 핵산 서열들에 대한 대략적인 정렬은 Smith 및 Waterman의 국소적 상동성 알고리즘, *Advances in Applied Mathematics* 2:482-489 (1981)에 의해 제공된다. 이 알고리즘은 미국 Washington, D.C.의 National Biomedical Research Foundation이 펴낸 Dayhoff, *Atlas of Atlas of Protein Sequences and Structure*, M. O. Dayhoff ed., 5 suppl. 3:353-358에 의해 개발되고, Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986)에 의해 정규화된 점수 행렬을 사용함으로써 아미노산 서열들에 적용될 수 있다. 서열의 동일성 백분율을 결정하기 위한 이러한 알고리즘의 예시적인 실행은 "BestFit" 실용신안 출원에서 Genetics Computer Group (Madison, Wis.)에 의해 제공된다. 서열들 간의 동일성 또는 유사성 백분율을 계산하는 그 외 적합한 프로그램들은 일반적으로 해당 기술분야에 공지이며, 예를 들어, 또다른 정렬 프로그램은 기본 매개변수와 함께 사용되는 BLAST가 있다. 예를 들어, BLASTN 및 BLASTP는 다음 기본 매개변수를 사용하여 이용될 수 있다: 유전 부호=표준; 필터=없음; 가닥=모두; 컷오프=60; 기대값=10; 매트릭스=BLOSUM62; 디스크립션=50 서열들; 분류=HIGH SCORE; 데이터베이스=비다중(non-redundant), GenBank+EMBL+DDJB+PDB+ GenBank CDS translations+Swiss protein+Spupdate+PIR. 이들 프로그램들의 상세내용은 GenBank 웹사이트에서 찾을 수 있다.

[0112] 본 발명의 범위에서 벗어나지 않고 상기 세포 및 방법들에서 다양한 변화가 이루어질 수 있으므로, 상기 상세한 설명 및 하기 제공되는 실시예에 포함된 모든 주제들은 설명적으로 해석되어야 하며 제한적 의미로 해석되어서는 안된다.

[0113] 실시예

[0114] 다음 실시예들은 본 발명의 특정 양상들을 설명한다.

[0115] 실시예 1: Cas9 오솔로그에 의한 표적 DNA 절단을 위한 PAM 요건들의 결정

[0116] 바실루스 스미티이 (*Bacillus smithii*), 락토바실루스 람노수스 (*Lactobacillus rhamnosus*), 파라수테렐라 엑스크레멘티호미니스 (*Parasutterella excrementihominis*), 마이코플라즈마 카니스 (*Mycoplasma canis*), 마이코플라즈마 갈리셉티쿰 (*Mycoplasma gallisepticum*), 아커만시아 글리카니필라 (*Akkermansia glycaniphila*), 아커만시아 뮤시니필라 (*Akkermansia muciniphila*), 오에노코쿠스 키타하라에 (*Oenococcus kitaharae*), 비피도박테리움 봄비 (*Bifidobacterium bombi*), 아시도테르무스 셀룰로리티쿠스 (*Acidothermus cellulolyticus*), 알리시클로바실루스 헤스페리둠 (*Alicyclobacillus hesperidum*), 올리넬라 숙시노게네스 (*Wolinella succinogenes*), 니트라티프락터 살수기니스 (*Nitratifractor saluginis*), 칼스토니아 시지기이 (*Ralstonia syzygii*), 및 코리네박테리움 디프테리아 (*Corynebacterium diphtheria*)에서 유래한 Cas9 오솔로그들은 인간 세포에서의 발현을 위하여 코돈 최적화되었으며 C 말단에서 SV40 대형 T 항원 핵 국재화 (NLS)로 태그되었다 (서열 번호:1-30; 이하 표 6 참고). 각 오솔로그의 발현은 인간 거대세포바이러스 (CMV) 즉각적 초기 인핸서 및 프로모터에 의해 유도되었다. CRISPR RNA (crRNA) 및 추정 트랜스-활성화 crRNA (tracrRNA)는 함께 결합되어 단일 가이드 RNA (sgRNA)를 형성하였다 (서열 번호: 31-45; 아래 표 6 참고). 각 sgRNA의 발현은 인간 U6 프로모터에 의해 유도되었다. 시험관내 전사된 sgRNA는 시험관내 소화를 위한 보충제로서 T7 프로모터 태그된 PCR 주형으로부터 제조되었다.

[0117] 인간 K562 세포들을 Cas9 인코딩 플라스미드 및 sgRNA 발현 플라스미드를 사용하여 뉴클레오펙션에 의해 형질감

염시켰다. 각 형질감염은 2백만개의 세포, 5 µg의 Cas9 인코딩 플라스미드 DNA 및 3 µg의 sgRNA 발현 플라스미드 DNA로 구성되었다. 세포를 형질감염 약 24시간 후에 수확하고, 얼음처럼 차가운 PBS 완충액으로 세척하고, 4 ° C 냉장실에서 30분 동안 지속적으로 교반하면서 150 µL의 용해 용액 (20mM HEPES, pH 7.5; 100mM KCl; 5mM MgCl₂, 1mM DTT, 5 % 글리세롤, 0.1% Triton X-100, 1x 프로테아제 억제제)으로 용해시켰다. 상청액은 4 ° C에서 2분 동안 16,000 xg에서 원심 분리하여 잔류 세포 파편을 제거하여 준비하고 플라스미드 DNA PAM 라이브러리의 시험관내 분해를 위한 Cas9 RNP의 공급원으로 사용했습니다. 이 라이브러리는 4⁸ 개의 축퇴성 PAM들을 함유하였으며, 각각 다음 구성의 프로토스페이서 바로 앞에 있다: 5'-GTACAAACGGCAGAAGCTGGNNNNNNN-3' (서열 번호: 46). 각 시험관내 분해는 10 µL의 세포 용해물 상청액, 2 µL의 5x 분해 완충액 (100mM HEPES, pH 7.5; 500mM KCl; 25mM MgCl₂; 5mM DTT; 25 % 글리세롤), 800ng의 PAM 라이브러리 DNA, 그리고 20 µL 반응 부피에서 20 pmol의 시험관내 전사된 sgRNA 보충제로 구성되었다. 반응을 37 ° C에서 30 분간 유지 한 후 PCR 정제 키트로 정제하였다. Illumina NextSeq 시퀀싱 라이브러리는 분해된 제품에서 준비되고 딥 시퀀싱을 거쳤습니다. Weblogo 프로그램을 사용하여 심층 시퀀싱 데이터를 분석하여 각 Cas9 오솔로그에 대한 PAM 요건들을 추론했다.

[0118] 결과를 도 1에 요약한다. 결과는 시험관내 표적 DNA 절단을 위해 PAM을 함유하는 A 및/또는 T를 사용하는 여러 Cas9 오솔로그를 밝혀냈다. 이러한 Cas9 오솔로그는 AT 풍부한 게놈 부위를 표적으로 하는 수단을 제공 할 수 있다. 결과는 또한 GC가 풍부한 게놈 부위를 표적하기에 적합한 PAM을 사용하는 여러 Cas9 오솔로그를 밝혀냈다. 이러한 Cas9 오솔로그는 GC가 풍부한 게놈 부위에서 SpyCas9에 대한 대체 표적화 계획을 제공하여 표적화 분리능과 특이성을 높일 수 있다.

[0119] **실시예 2: 바실루스 스미티이 (*Bacillus smithii*) Cas9 (BsmCas9) 및 락토바실루스 람노수스 (*Lactobacillus rhamnosus*) Cas9 (LrhCas9)를 사용한 게놈 변형**

[0120] 도 1 및 (상기) 표 A에 도시된 바와 같이, 소형 BsmCas9 (1095 aa) (서열 번호: 2) 및 LrhCas9 (서열 번호: 4)는 표적 DNA 결합을 위해 각각 5'-NNNNCAAA-3' PAM 및 5'-NGAAA-3' PAM을 사용한다. 이러한 새로운 PAM 이용은 AT 풍부 게놈 부위들을 표적하기 위한 수단들을 제공한다. 유전자 편집을 설명하기 위해, 인간 K562 세포들 (1x10⁶)을 5 µg의 Cas9 인코딩 플라스미드 DNA 및 3 µg의 sgRNA 발현 플라스미드 DNA로 뉴클레오펙션하였다. 표적 게놈 부위에는 인간 티로신-단백질 포스파타제 비-수용체 2형 (PTN2) 유전자좌, 인간 빈 스피라클 호메오박스 1 (*EMX1*) 유전자좌, 인간 세포예정사 1 리간드 1 (*PDIL1*) 유전자좌, 인간 AAVS1 세이프 하버 유전자좌, 인간 사이토크롬 p450 산화환원효소 (*POR*) 유전자좌 및 인간 핵 수용체 서브패밀리 1 그룹 I 구성원 3 (*CAR*) 유전자좌가 포함된다. 형질감염 3일 후 DNA 추출 용액 (QuickExtract제)을 사용하여 게놈 DNA를 준비하고 표적된 게놈 영역들을 각각 PCR 증폭시켰다 (JumpStart Taq제 ReadyMix제). PCR 프라이머들을 표 1에 열거한다.

표 1. PCR 프라이머.			
유전자좌	정방향 프라이머 (5'-3')	역방향 프라이머 (5'-3')	크기 (bp)
<i>PTN2</i>	CTGTTTCCTGGGTTCCAATAACAAG AC (서열 번호: 47)	ACAAGGGCTCAAGTGGAGTG (서열 번호: 48)	290
<i>EMX1</i>	ATGGGAGCAGCTGGTCAGAG (서열 번호: 49)	CAGCCCATTGCTTGTCCCT (서열 번호: 50)	507
<i>PD1L1</i>	CTCGCCATTCCAGCCACTCAAAC (서열 번호: 51)	GGTTAAGTCGGGTTTCCTTGCAAG (서열 번호: 52)	341
<i>AAVS1</i>	TTCGGGTCACCTCTCACTCC (서열 번호: 53)	GGCTCCATCGTAAGCAAACC (서열 번호: 54)	469
<i>POR</i>	CTCCCCTGCTTCTTGTCGTAT (서열 번호: 55)	ACAGGTCGTGGACACTCACA (서열 번호: 56)	380
<i>CAR</i>	GGATCAAGTCAAGGGCATGT (서열 번호: 57)	ATGTAGCTGGACAGGCTTGG (서열 번호: 58)	347

[0121]

[0122]

증폭은 다음 조건을 사용하여 수행되었다: 초기 변성을 위해 2분 동안 98 ° C; 34주기는 15초 동안 98 ° C, 30초 동안 62 ° C, 및 45초 동안 72 ° C; 1주기는 5분 동안 72 ° C; 그리고 4 ° C에서 유지. PCR 생성물들을 Cel-1 뉴클레아제로 분해하고 10% 아크릴아미드 겔에서 분리하였다. 표적된 돌연변이 비율을 ImageJ를 사용하여 측정하고 퍼센트 삽입 및/또는 결실 (% 삽입결실)로 표현하였다. 결과를 표 2에 요약한다. 이러한 결과는 두 Cas9 오솔로그가 5'-NNNNCAAA-3' PAM (BsmCas9) 또는 5'-NGAAA-3' PAM (LrhCas9)을 사용하여 인간 세포에서 내인성 게놈 부위를 편집 할 수 있음을 보여준다.

표 2. 인간 K562 세포들에서 BsmCas9 및 LrhCas9 들 이용한 유전자 편집.				
Cas9	유전자좌/표적 #	프로토스페이서 서열 (5'-3')	PAM (5'-3')*	삼입결실 (%)
BsmCas9	<i>PTN2</i>	CTCATACATGGCTATAATAGAA (서열 번호: 59)	GGAG <u>CAAA</u>	11.9
	<i>EMX1</i> /표적 1	GAAGGTGTGGTTCCAGAACCGG (서열 번호: 60)	AGGAC <u>AAA</u>	6.8
	<i>EMX1</i> /표적 2	TGGTTCCAGAACCGGAGGACAA (서열 번호: 61)	AGTAC <u>AAA</u>	10.9
	<i>EMX1</i> /표적 3	CCCAGGTGAAGGTGTGGTTCCA (서열 번호: 62)	GAAC <u>CGGA</u>	0
	<i>EMX1</i> /표적 4	AGAACCGGAGGACAAAGTACAA (서열 번호: 63)	ACGGC <u>AGA</u>	0
LrhCas9	<i>PD1L1</i>	CCTCTGGCACATCCTCCAAA (서열 번호: 64)	TG <u>AAA</u>	38.9
	<i>AAVS1</i>	CTAGGGACAGGATTGGTGAC (서열 번호: 65)	AG <u>AAA</u>	32.7
	<i>POR</i> /표적 1	GCTCGTACTGGCGAATGCT	GG <u>AAA</u>	26.7

[0123]

		(서열 번호: 66)		
	POR/표적 2	GCTGAAGAGCTACGAGAACC (서열 번호: 67)	<u>AGAAG</u>	0
	POR/표적 3	CATGGGGGAGATGGGCCGGC (서열 번호: 68)	<u>TGAAG</u>	0
	CAR/표적 1	AGAGACTCTCTAGAAGGGAC (서열 번호: 69)	<u>AGAAA</u>	31.7
	CAR/표적 2	GTGAGAGTCTCCTCCCAATG (서열 번호: 70)	<u>GGAAA</u>	27.0
	CAR/표적 3	GGGAGGAGACTCTCACCTGA (서열 번호: 71)	<u>AGAAA</u>	0
* PAM의 결정인자 뉴클레오티드는 밑줄되어 있다.				

[0124] 실시예 3: 염색질 조절 모티프와의 융합에 의한 파라수테렐라 엑스크레멘티호미니스 (*Parasutterella excrementihominis*) Cas9 (PexCas9)

[0125] 파라수테렐라 엑스크레멘티호미니스 (*Parasutterella excrementihominis*) Cas9 (PexCas9-NLS) (서열 번호:6)는 TGSG 링커 (서열 번호:109)를 사용하여 N 말단에서 인간 HMG1 펩티드 (서열 번호:72)와, 그리고 LEGGS 링커 (서열 번호:108)를 사용하여 C 말단에서 인간 HMGB1 박스 A 펩티드 (PexCas9-HN1HB1 융합; 서열 번호:117) 또는 인간 히스티딘 H1 중심 구상 도메인 펩티드 (PexCas9-HN1H1G; 서열 번호:118)와의 융합에 의해 변형되었다.

표 3. 염색질 조절 모티프		
명칭	펩티드 서열	서열 번호:
HMGN1 (HN1)	MPKRKVSSAEGAAKEPKRRSARLSAKPPAKVEAKPKKAA AKDKSSDKKVQTKGKRGAKGKQAEVANQETKEDLPAEN GETKTEESPASDEAGEKEAKSD	72
인간 HMGB1 박스 A (HB1)	MGKGDPPKPRGKMSSYAFFVQTCREEHKKKHPDASVNF SEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKT YIPPKGE	73
인간 히스티딘 H1 중심 구상 도메인 (H1G)	STDHPKYS DMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVG ENADSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP	74

[0126]

[0127]

인간 K562 세포들 (1×10^6)을 물 당량 (각각 5 및 5.4 μg)의 PexCas9-NLS, PexCas9-HN1HB1 융합, 또는 PexCas9-HN1H1G 융합을 인코딩하는 플라스미드 DNA 및 인간 시토크롬 p450 산화환원효소 (*POR*) 유전자좌의 게놈 부위를 표적하는 3 μg 의 sgRNA 플라스미드로 형질감염시켰다. 형질감염 3일 후 DNA 추출 용액 (QuickExtract™)을 사용하여 게놈 DNA를 준비하고 표적된 게놈 영역을 정방향 프라이머 5'-CTCCCTGCTTCTGTCTGAT-3' (서열 번호:55) 및 역방향 프라이머 5'-ACAGGTCGTGGACACTCACA-3' (서열 번호:56)를 사용하여 PCR 증폭하였다. 증폭은 다음 조건을 사용하여 수행되었다: 초기 변성을 위해 2분 동안 1주기는 98 ° C; 34주기는 15초 동안 98 ° C, 30초 동안 62 ° C, 및 45초 동안 72 ° C; 1주기는 5분 동안 72 ° C; 그리고 4 ° C에서 유지. PCR 생성물들을 Cel-1 뉴클레아제로 분해하고 10% 아크릴아미드 겔에서 분리하였다. 표적된 돌연변이 비율을 ImageJ를 사용하여 측정하고 퍼센트 삽입 및/또는 결실 (% 삽입결실)로 표현하였다. 결과를 표 4에 요약한다. 결과는 적어도 하나의 염색질 조절 모티프와 Cas9 융합이 인간 세포의 내인성 표적에 대한 유전자 편집 효율을 향상시킨다는 것을 보여준다.

표 4 인간 K562 세포에서 PexCas9 및 PexCas9 융합 단백질을 사용한 유전자 편집.				
Cas9 뉴클레아제	유전자좌	프로토스페이서 (5'-3')	PAM (5'-3')*	삽입결실 (%)
PexCas9-NLS	<i>POR</i>	TGTACATGGGGGAGATGGGC (서열 번호: 75)	<u>CGG</u>	22.9
PexCas9-HN1HB1 융합				36.8
PexCas9HN1H1G 융합				43.7

[0128]

* PAM의 결정인자 뉴클레오티드는 밑줄되어 있다.

[0129] 실시예 4. sgRNA 변형에 의한 *마이코플라즈마 카니스 (Mycoplasma canis)* Cas9 (McaCas9) 시스템의 개선

[0130] McaCas9의 야생형 crRNA 코딩 서열은 반복 영역에 4개의 연속적인 티미딘 잔기를 함유하고, 4개의 티미딘 잔기 중 3개는 crRNA와 tracrRNA가 함께 결합되어 sgRNA를 형성 할 때 추정 tracrRNA 서열에서 3개의 아데노신 잔기와 쌍을 이룰 것으로 예측된다. 인간 RNA 중합효소 (Pol) III는 코딩 RNA 가닥에 4개 이상의 연속적인 티미딘 잔기를 전사 종결 신호로 사용하는 것으로 알려져 있다. 인간 세포에서 McaCas9 sgRNA의 초기 전사 종결을 방지하기 위해 T에서 C 돌연변이 및 상응하는 A에서 G 돌연변이를 sgRNA 스캐폴드에 도입하여 다음과 같은 서열의 변형된 sgRNA 스캐폴드를 형성했다: 5'-GUUUCUAGUGUUGUACAAUUAUUGGGUGAAAACCCAAAUUUGUACAUCUAGAUCAAGGCGCUUAAUUGCUGCCGUAUUGCGUAGCUUUCAGUUUUUU-3' (서열 번호: 76), 여기서 돌연변이된 뉴클레오티드는 밑줄되어 있다. 이 변형은 또한 sgRNA 스캐폴드 열역학적 안정성을 증가시킬 것으로 예상된다.

[0131] N 말단에 HMGN1 펩티드를 그리고 C 말단에 히스티딘 H1 구상 도메인 펩티드를 함유하는, McaCas9 융합 단백질을 인코딩하는 플라스미드 DNA 5.5 µg, 및 대조 sgRNA 스캐폴드 또는 변형된 sgRNA 스캐폴드를 인코딩하는 sgRNA 플라스미드 DNA 3 µg으로 인간 K562 세포들 (1x 10⁶)을 형질감염시켰다. 형질감염 3일 후 DNA 추출 용액 (QuickExtract™)을 사용하여 게놈 DNA를 준비하고 표적된 게놈 영역을 정방향 프라이머 5'-CTCCCTGCTTCTGTGCTAT-3' (서열 번호: 55) 및 역방향 프라이머 5'-ACAGGTCGTGGACACTCACA -3' (서열 번호: 56)를 사용하여 PCR 증폭하였다. 증폭은 다음 조건을 사용하여 수행되었다: 초기 변성을 위해 2분 동안 1주기는 98 ° C; 34주기는 15초 동안 98 ° C, 30초 동안 62 ° C, 및 45초 동안 72 ° C; 1주기는 5분 동안 72 ° C; 그리고 4 ° C에서 유지. PCR 생성물들을 Cel-1 뉴클레아제로 분해하고 10% 아크릴아미드 겔에서 분리하였다. 표적된 돌연변이 비율을 ImageJ를 사용하여 측정하고 퍼센트 삽입 및/또는 결실 (% 삽입결실)로 표현하였다. 결과를 표 5에 요약한다. 결과는 포유동물 세포에서 Cas9 오솔로그의 활성이 sgRNA 스캐폴드를 변형하여 향상 될 수 있음을 보여준다.

표 5. 대조 sgRNA 스캐폴드 또는 변형된 sgRNA 스캐폴드와 함께 McaCas9를 사용한 유전자 편집.				
sgRNA 스캐폴드	유전자좌	프로토스페이서 (5'-3')	PAM (5'-3')*	삽입결실 (%)
대조 sgRNA 스캐폴드	POR	ATAGATGCGGCCAAGGTGTACA (서열 번호: 77)	TGGG	25.5
변형된 sgRNA 스캐폴드				36.2

[0132] * PAM의 결정인자 뉴클레오티드는 밑줄되어 있다.

[0133] 실시예 5. 염색질 조절 모티프와의 융합에 의한 McaCas9, BsmCas9, PexCas9, 및 LrhCas9 활성의 개선

[0134] McaCas9-HN1HB1 (서열 번호: 123), McaCas9-HN1H1G (서열 번호: 124), BsmCas9-HN1HB1 (서열 번호: 119), Bsm-HN1H1G (서열 번호: 120), Lrh-HN1HB1 (서열 번호: 121), LrhCas9-HN1H1G (서열 번호: 122)를 생성하기 위해, McaCas9-NLS, BsmCas9-NLS, 및 LrhCas9-NLS 단백질들을 아미노 말단에서 HMGN1 (HN1)과 그리고 카르복실 말단에서 HMGB1 박스 A (HB1) 또는 히스티딘 H1 중심 구상 모티프 (H1G)와 연결시켜 또 다른 Cas9-CMM 융합 단백질들을 준비하였다. 이들 융합들 그리고 실시예 3에 상기 기재된 PexCas9-CMM 융합들의 뉴클레아제 활성을 본질적으로 실시예 2 및 3에 상기 기재된 바와 같이 상응하는 조작된 Cas9 단백질의 활성과 비교하였다. 표 6은 각 Cas9 뉴클레아제에 대한 특정 유전자좌에서의 표적 부위 (즉, 프로토스페이서 + PAM, 이는 볼드체로 나타내며 결정인자 뉴클레오티드는 밑줄되어 있다)를 나타낸다.

표 6. 유전자 편집 표적 부위			
Cas9	유전자좌	표적 부위 (5'-3')	서열 번호
Mca	<i>POR1</i>	ATAGATGCGGCCAAGGTGTACAT <u>G</u> GG	125
	<i>POR2</i>	CTACGAGAACCAGAAGCCGTGAGT <u>G</u> G	126
Bsm	<i>PTN2</i>	CTCATACATGGCTATAATAGAAGGAG <u>G</u> CAA	127
	<i>EMX1</i>	GAAGGTGTGGTTCAGAACCGGAGGAC <u>A</u> AA	128
	<i>EMX2</i>	TGGTTCAGAACCGGAGGACAAAGTAC <u>A</u> AA	129
Pex	<i>POR1</i>	TGTACATGGGGGAGATGGGCG <u>G</u>	130
	<i>AAVS1</i>	GGGGCCACTAGGGACAGGATT <u>G</u> G	131
Lrh	<i>POR1</i>	AGCTCGTACTGGCGAATGCTG <u>G</u> AAA	132
	<i>PD1L1</i>	CCTCTGGCACATCCTCCAAATG <u>A</u> AA	133

[0135]

[0136]

각 조건하에서 삽입결실의 퍼센트가 도 2A-D에 플롯되어 있다. 두 HN1HB1 및 HN1H1G 조합들 모두 적어도 하나의 부위에서 4 개의 Cas9 오솔로그를 크게 향상시켰다. 배수 변화 크기에 기초하였을 때, CMM 융합 변형은 McaCas9에서 최대 향상을 제공하였으며, 테스트한 2개 부위에서 5배 이상만큼 그 활성을 증가시켰다 (도 2A). CMM 융합은 PexCas9에서 2배 이상의 향상을 제공하였다 (도 2B). BsmCas9 활성은 한 부위에서 3배 이상 향상되었으나, 2번째 부위에서는 20%만이 증가되었고 3번째 부위에서는 영향이 없었다 (도 2C). 그러나 세 가지 BsmCas9 뉴클레아제들 모두 매우 효율적이었음 (>35% 삽입결실)을 주목하여야 한다. LrhCas9는 융합 변형없이도 테스트한 2개 부위들에서 매우 효율적이었다 (22% 및 33% 삽입결실) (도 2D). 그러나, HN1H1G 조합 또한 두 부위들 모두에서 유의한 향상을 제공하였는데, 활성 증가가 705 및 28%였다. 이들 결과는 CMM 융합 전략이 유전자 편집 효율을 향상시킨다는 것을 보여준다.

[0137]

실시예 6. Cas9-CMM 융합들의 오프-타겟 효과

[0138]

Cas9-CMM 융합들의 오프-타겟 활성을 평가하기 위하여, Surveyor 뉴클레아제 분석을 사용하여 각 표적 부위에 대해 1-5위에 랭크된 잠재적인 오프-타겟 부위들을 분석하였다. 상기 실시예 5에 기재된 Cas9 및 Cas9-CMM 융합 데이터 이외에도, 스트렙토코쿠스 피오게네스 (*Streptococcus pyogenes*) Cas9 (SpyCas9), SpyCas9-CMM 융합, 스트렙토코쿠스 파스테우리아누스 (*Streptococcus pasteurianus*) Cas9 (SpaCas9), Spa-CMM 융합, 캄필로박터 제주니 (*Campylobacter jejuni*) Cas9 (CjeCas9), 및 CjeCas9-SMM 융합들로부터 얻은 데이터 또한 분석하였다. 분석된 총 64개의 잠재적 오프-타겟 부위로부터, 11개의 부위에서 오프-타겟 절단이 탐지되었으며, 이는 테스트한 총 21개의 가이드 서열 중 9개의 가이드 서열로 인한 것이었다. 11개의 오프-타겟 부위들에서, 대조 Cas9 및 융합 뉴클레아제들은 동시적이었으며, 대조 SpyCas9에서 오프-타겟 절단이 전혀 탐지되지 않았던 POR Spy 1-OT1 부위는 예외였다. 종합하면, 융합 뉴클레아제 및 대조 Cas9 사이에는 유의한 차이가 없었다 (도 3). 예를 들면, 11개 오프-타겟 부위들 모두에 걸쳐, HN1H1G 융합 조합은 평균 $8.0 \pm 6.0\%$ 삽입결실을 보였으며 대조 Cas9는 평균 $7.5 \pm 5.1\%$ 삽입결실을 보였다. 유사하게, HN1HB1 융합 조합과 관련된 10개 오프-타겟 부위들에 걸쳐, 융합 조합과 대조 Cas9 사이에 유의한 차이는 없었다 ($6.9 \pm 5.7\%$ vs. $6.5 \pm 5.4\%$ 삽입결실). 종합하면, 이들 결과는 HN1HB1 및 HN1H1G 융합 조합에 의한 온-타겟 활성 향상이 일반적으로 오프-타겟 활성의 증가를 가져오는 것은 아님을 보여준다.

[0139]

조작된 Cas9 시스템

[0140]

표 7은 조작된 Cas9/NLS 단백질의 인간 코돈 최적화된 DNA 및 단백질 서열 (서열 번호: 1-30, 여기서 NLS 서열

은 밑줄 표시됨) 및 조작된 sgRNA의 DNA 서열 (서열 번호: 31-45; 5' 단부에 있는 N 잔기는 프로그래밍 가능한 표적 서열을 나타냄)을 제시한다. 또한 Cas9-CMM 융합들에서도 제시된다 (서열 번호: 117-124).

표 7. 조작된 Cas9 시스템

BsmCas9/NLS DNA 서열 (서열 번호: 1)

ATGAACTACAAGATGGGCTCGACATCGGAATCGCCTCTGTTGGATGGGCCGTGATCAACCTGGAC
CTGAAGAGAATCGAGGACCTCGGCGTGCGGATCTTCGACAAGGCTGAGCATCCTCAGAACGGCGAG
TCTCTGGCCCTGCCTAGAAGAATTGCCAGAAGCGCCAGACGGCGGCTGCGGAGAAGAAAGCACAGA
CTGGAACGGGATCAGACGGCTGCTGGTGTCCGAGAACGTGCTGACCAAAGAAGAGATGAACCTGCTG
TTCAAGCAGAAAAAGCAGATCGACGTGTGGCAGCTGAGAGTGGACGCCCTGGAAAGAAAGCTGAAC
AACGACGAGCTGGCCAGAGTGTCTGCTGCACCTGGCCAAGAGAAGAGGCTTCAAGAGCAACAGAAAG
AGCGAGCGGAACAGCAAAGAGAGCAGCGAGTTCCCTGAAGAACATCGAAGAGAACCAGAGCATTCTG
GCCCAGTACAGATCCGTGGGCGAGATGATCGTGAAGGACAGCAAGTTCGCCTACCACAAGCGGAAC
AAGCTGGACAGCTACAGCAACATGATCGCCAGGGACGATCTGGAAAGAGAGATCAAGCTGATCTTC
GAGAAGCAGCGCGAGTTCAACAACCCCGTGTGCACCGAGAGACTGGAAGAGAAGTACCTGAACATC
TGGTCCAGCCAGCGGCCCTTTCGCCTCCAAAGAGGACATCGAGAAAAAAGTGGGCTTCTGCACCTTC
GAGCCCAAAGAGAAAAAGAGCCCTAAGGCCACCTACACCTTCCAGAGCTTCATCGTGTGGGAGCAC
ATCAACAAGCTGCGGCTGGTGTCTCCCGACGAGACAAGAGCCCTGACCGAGATCGAGCGGAATCTG
CTGTATAAGCAGGCCCTTCAGCAAGAACAAGATGACCTACTACGACATCCGGAAGCTGCTGAACCTG
AGCGACGACATCCACTTCAAGGGCTGCTGTACGACCCCAAGAGCAGCCTGAAGCAGATTGAGAAC
ATCCGGTTTCTGGAAGTGGACTCTTACCACAAGATCCGGAAGTGCATCGAGAATGTGTACGGCAAG
GACGGCATCCGCATGTTCAACGAGACAGACATCGACACCTTCGGCTACGCCCTGACCATCTTCAAG
GACGACGAGGATATCGTGGCCTACCTGCAGAACGAGTACATCACCAAGAACGGCAAGCGGGTGTC
AATCTGGCCAACAAAGGTGTACGACAAGTCCCTGATCGACGAACTGCTGAATCTGTCTTCTCCAA
TTCCGCCCACTGAGCATGAAGGCCATCCGGAACATCCTGCCTTACATGGAACAGGGCGAAATCTAC
AGCAAGGCCCTGCGAACTGGCCGGCTACAACCTCACAGGCCCAAGAAGAAAGAGAAGGCCCTGCTG
CTGCCGTGTGATCCCCAATATCGCCAATCCTGTGGTCATGCGGGCCCTGACACAGAGCAGAAAGGTG
GTCAACGCCCATCATCAAGAAATACGGATCCCCCGTGTCCATCCACATCGAGCTGGCTAGGGATCTG
AGCCACAGCTTCGACGAGCGGAAGAAGATCCAGAAGGACCAGACCGAGAACCAGCAAGAAGAACGAA
ACCGCCATCAAGCAGCTGATCGAGTACGAGCTGACTAAGAACCCACCGGCCTGGACATCGTGAAG
TTCAAACCTTTGGAGCGAGCAGCAAGGCAGATGCATGTACTCCCTGAAGCCTATTGAGCTGGAAAGA
CTGCTGGAACCCGGCTACGTGGAAGTGGACCACATTCTGCCCTACAGCAGAAGCCTGGACGACAGC
TACGCCAACAAAGTGCTGGTCTGACAAAAGAGAACCAGCGAAAAGGGCAATCACACCCCTGTGGAA
TATCTCGGCCTGGGCTCTGAGCGGTGGAAGAAATTCGAGAAGTTCGTGCTGGCTAACAAAGCAGTTC
TCTAAGAAGAAGAAGCAGAACCTGCTCCGGCTGAGATACGAGGAAACCGAGGAAAAAGAGTTCAAA
GAGCGGAACCTGAACGACACCCGGTACATCTCCAAGTTCTTCGCCAACTTCATCAAAGAGCATCTG
AAGTTCGCCGACGGCGACGGCGGCCAGAAAGTGTACACAATCAACGGCAAGATCACCGCTCACCTG
AGAAGCAGATGGGACTTCAACAAGAACCAGGAAGAGAGCGACCTGCACCACGCTGTGGATGCTGTG
ATTGTGGCCTGTGCCACACAGGGCATGATCAAGAAGATTACCGAGTTCTACAAGGCCCGCAGCAG
AACAAAGAGTCCGCCAAGAAAAAAGAACCATCTTTCCCAGCCTTGGCCTCACTTCGCCGATGAG
CTGAAGGCTCGGCTGAGCAAGTTCCCTCAAGAGTCCATCGAGGCCCTTCGCTCTGGGCAACTACGAC
AGAAAGAAGCTGGAATCCCTGCGGCCTGTGTTGCTGTCCAGAATGCCCAAGAGATCCGTGACAGGC
GCTGCCCCACCAAGAGACACTGAGAAGATGCGTGGGCATCGACGAGCAGTCTGGCAAGATTACAGACC

[0141]

GCCGTGAAAAACAAAGCTGAGCGACATCAAGCTGGATAAGGACGGACACTTCCCCATGTACCAGAAA
GAGTCTGACCCCAAGACCTACGAGGCCATCAGACAGAGGCTGCTCGAACACAACAACGACCCTAAG
AAGGCCTTTCAAGAGCCACTGTACAAGCCCAAAAAGAATGGCGAGCCCGGACCAGTGATCCGGACC
GTGAAGATCATCGACACAAAGAACAAGGTGGTGCACCTGGACGGCAGCAAGACAGTGGCCTACAAC
TCCAACATCGTGGCGACCGACGTGTTTCGAGAAGGATGGCAAGTACTACTGCGTGCCCGTGTACACT
ATGGATATCATGAAGGGCACCCCTGCCTAACAAGGCCATCGAAGCCAACAAGCCCTACTCCGAGTGG
AAAGAGATGACCGAAGAGTACACGTTCCAGTTCAGTCTGTTCCTCCCAACGACCTCGTGCGCATCGTG
CTGCCAAGAGAGAAAAACCATCAAGACCAGCACCAACGAGGAAATCATCATTAAAGGACATCTTTGCC
TACTACAAGACCATCGACAGCGCCACAGGCGGCTGGAACTGATCTCCACGATCGGAACCTTCAGC
CTGAGAGGCGTGGGCTCTAAGACACTGAAGCGCTTTGAGAAGTATCAGGTGGACGTGCTGGGCAAC
ATCCACAAAAGTGAAGGGCGAGAAGAGAGTCCGCCCTGGCCGCTCTCTACCAACCAGAAAAAGGGAAAG
ACCGTGGACAGCCTGCAGAGCGTGTCCGATCCCAAGAAGAAGAGGAAGGTG

BsmCas9/NLS 단백질 서열 (서열 번호: 2)

MNYKMGDLIGIASVGVAVINLDLKRIEDLGVRIFDKAEHPQNGESLALPRRIARSARRRLRRRKHR
LERIRLLVSENVLTKEEMNLLFKQKKQIDVWQLRVDALERKLNDELARVLLHLAKRRGFKSNRK
SERNSKESSEFLKNIEENQSIQAQYRSVGEMIVKDSKFAYHKNKLDSSYNMIARDLEREIKLIF
EKQREFNNPVCTERLEEKYLNWSSQRPFAKEDIEKKVGFCTFEPKEKRAPKATYTFQSFIVWEH
INKLRLVSPDETRALTEIERNLLYKQAFSKNKMITYDIRKLLNLSDDIHFKGLLYDPKSSLKQIEN
IRFLELDSYHKIRKCIENVYKDGIRMFNETDIDTFGYALTIFKDEEDIVAYLQNEYITKNGKRV
NLANKVYDKSLIDELLNLSFSKFAHLSMKAIRNILPYMEQGEIYSKACELAGYNFTGPKKKEKALL
LPVIPNIANPVVMRALTQSRKVNAIIKKYGPSVSIHIELARDLSHSFDERKKIQKDQTENRKKNE
TAIKQLIEYELTKNPTGLDIVKFKLWSEQQGRCMYSLKPIELERLLEPGYVEVDHILPYSRLDDS
YANKVLVLTKENREKGNHTPVEYLGSGSERWKKFEKFLANKQFSKKKKQNLLRLRYEETEEKEFK
ERNLNDTRYISKFFANFIKEHLKFADGGGQKVYTINGKITAHLRSRWDFNKNREESDLHHAVDV
IVACATQGMIKKITEFYKAREQNKESAKKKEPIFPQWPWPHFADELKARLSKFPQESIEAFALGNYD
RKKLESRLRPVFSRMPKRSVTGAHQETLRRCVGIDEQSGKIQTAVKTKLSDIKLDKDGHFPMYQK
ESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEFPGPVIRTVKIIDTKNKVVHLDGSKTVAYN
SNIVRTDVFEKDGKYYCVPVYTMIMKGTLPNKAIEANKPYSEWKEMTEEYTFQFSLFPNDLVRI
LPREKTIKTSTNEEIIKIDIFAYYKTIDSATGGLELISHDRNFSLRGVGSKTLKRFEKYQVDVLGN
IHKVKGEKRVGLAAPTNOKKGKTVDLSQSVSDPKKKRKV

LrhCas9/NLS DNA 서열 (서열 번호: 3)

ATGACCAAGCTGAACCAGCCTTACGGCATCGGCCTGGACATCGGCAGCAATAGCATCGGCTTTGCC
GTGGTGGACGCCAACAGCCATCTGCTGAGACTGAAGGGCGAGACAGCCATCGGCGCCAGACTGTTT
AGAGAGGGACAGAGCGCCGCTGACAGACGGGGAAGCAGAACCACAAGAAGGCGGCTGTCCAGAACC
AGATGGCGGCTGAGCTTCTGCGGGATTTCTTCGCCCCCTCACATCACCAAGATCGACCCCGACTTC
TTTCTGCGGCAAAAATACTCCGAGATCAGCCCCAAGGACAAGGACAGGTTTAAGTACGAGAAGCGG
CTGTTCAACGACCCGACCGACCGGAGTTCTACGAGGACTACCCAGCATGTACCACCTGAGACTG
CACCTGATGACCCACACACACAAGGCCGATCCTCGGGAAATCTTCTGCGCCATCCACCACATCCTG
AAGTCCAGAGGGCACTTTCTGACACCCGGCGCTGCCAAGGACTTCAACACCGACAAAGTGGACCTT

GAGGACATCTTCCCCGCTCTGACAGAGGCTTACGCCAGGTGTACCCCGATCTGGAAGTGACCTTC
GATCTGGCCAAGGCCGACGACTTCAAGGCCAAGCTGCTGGACGAACAGGCCACACCTAGCGACACA
CAGAAAGCCCTGGTCAACCTGCTGCTGTCTAGCGACGGCGAGAAAGAAATCGTGAAGAAGCGGAAG
CAGGTCTGACCGAGTTGCGCCAAGGCCATCACCGGCCCTGAAAACAAAGTTCAATCTGGCCCTGGGC
ACCGAGGTGGACGAAGCTGATGCTTCCAAGTGGCAGTTACAGCATGGGCCAGCTGGACGACAAGTGG
TCCAACATCGAAAACAGCATGACCGACAGGGCACCAGAAATCTTCGAGCAGATCCAAGAGCTGTAC
CGGGCCAGACTGCTGAACGGAATTGTGCCTGCCGGCATGAGCCTGTCTCAGGCCAAAGTGGCCGAT
TACGGCCAGCACAAGAGGACCTGGAAGTGTTCAGACCTACCTGAAGAAGCTGAACGACACAGAG
CTGGCCAAGACCATCAGGGGCCCTGTACGATCGGTACATCAACGGCGACGACGCCAAGCCTTTCTTG
CGCGAGGATTTTGTGAAGGCCCTGACCAAAGAAGTGACAGCTCACCCCAACGAGGTGTCCGAACAG
CTGCTGAACAGGATGGGCCAAGCCAACTTCATGCTGAAGCAGCGGACCAAGGCCAACGGCGCCATT
CCTATTACAGCTGCAGCAGAGAGAGCTGGACCAGATCATTGCCAACAGAGCAAGTACTACGACTGG
CTGGCCGCTCCTAATCCTGTGGAAGCCACAGATGGAAGATGCCCTACCAGCTGGATGAGCTGCTC
AACTTTACATCCCCCTACTACGTGGGCCCTCTGATCACCCCTAAACAGCAGGCCGAGAGCGGCGAG
AATGTGTTCGCTTGGATGGTCCGAAAGGACCCAGCGGCAACATCACCCCTTACAACCTTCGACGAG
AAGGTGGACAGAGAGGCCAGCGCCAACACCTTCATCCAGAGAATGAAGACCACCGACACATACCTG
ATCGGCGAGGACGTGCTGCCTAAGCAGAGCCTGCTGTACCAGAAATACGAGGTGCTGAACGAGCTG
AACACGTGCGGATCAACAACGAGTGCCTGGGCACAGACCAGAAGCAGAGACTGATCAGAGAGGTG
TTCGAGCGGCACAGCAGCGTGACCATCAACAGGTGGCCGACAATCTGGTGGCCACGGCGATTTT
GCCAGACGGCCTGAGATTAGAGGACTGGCCGATGAGAAGCGGTTCTTGAGCAGCCTGAGCACCTAC
CACCAGCTGAAAGAGATCCTGCACGAGGCCATCGACGACCCCAACAACTGCTGGATATCGAGAAC
ATCATCACCTGGTCCACCGTGTTCGAGGACCACACCATCTTCGAGACAAAGCTGGCCGAGATCGAG
TGGCTGGACCCCAAGAAGATCAACGAGCTGTCTGGCATCAGATACAGAGGCTGGGGCCAGTTCTCC
CGGAAGCTGCTCGATGGACTGAAGCTTGGCAATGGCCACACCGTGATTCAAGAAGTGTGCTGAGC
AACCACAACCTGATGCAGATCCTGGCCGACGAGACACTGAAAGAAACCATGACAGAGCTGAATCAG
GACAAAGTGAAAACCGACGACATCGAGGATGTGATCAACGACGCCTACACAAGCCCCAGCAACAAA
AAGGCCCTCAGACAGGTGCTGAGAGTGGTCGAGGATATCAAGCACGCCGCCAACGGACAGGACCCT
AGCTGGCTGTTTATCGAAACCGCCGATGGAACAGGCACCGCCGGCAAGAGAACACAGAGCCGGCAG
AACAGATCCAGACCGTGTACGCCAACGCCGCTCAAGAGCTGATCGATTCTGCCGTGCCGGGCGAG
CTGGAAGATAAGATTGCTGACAAGGCCAGCTTCACCGACCGGCTGGTGTGTACTTTATGCAAGGC
GGCAGAGACATCTACACAGGCGCCCTCTGAACATCGACCAGCTGAGCCACTACGATATCGACCAC
ATTCTGCCCCAGAGCCTGATCAAGGACGACAGCCTGGACAACCGGGTGCTCGTGAACGCCACCATC
AACCGCGAGAAGAACAATGTGTTTGCCAGCACACTGTTCCGCCGAAAGATGAAGGCCACCTGGCGG
AAATGGCACGAAGCCGGACTGATCTCTGGCAGAAAGCTGCCGAATCTGATGCTGCGGCCCGACGAG
ATCGACAAGTTTGCCAAGGGCTTCGTGGCCCGGCAGCTGGTTGAGACAAGACAGATCATCAAGCTG
ACAGAGCAGATTGCCGCGCTCAGTACCCCAACACCAAGATTATTGCCGTGAAGGCCGGACTGTCC
CATCAGCTGAGAGAGGAAGTGGACTTCCCAAGAACCGGGACGTGAACCACTACCACCACGCCTTC
GATGCCCTTTCTGGCCGCTAGAATCGGCACCTACCTGCTGAAGAGATACCCCAAGCTGGCCCCATTC
TTCACCTACGGCGAGTTTGCTAAGGTGGACGTCAAGAAGTTCCCGGAGTTCAACTTCATCGGAGCC
CTGACACACGCCAAGAAGAACATTATCGCCAAGGACACCGGCAGATCGTGTGGGACAAAGAGCGG
GACATCAGAGAAGTGGACCGCATCTACAACCTCAAGCGGATGCTGATCACACACGAGGTGTACTTC
GAGACTGCCGACCTGTTCAAGCAGACCATCTACGCCGCTAAGGACAGCAAAGAGAGAGGCGGACG
AAGCAGCTGATCCCTAAGAAGCAGGGCTACCCCACTCAGGTGTACGGCGGCTACACACAAGAGAGC
GGCTCTTACAACGCCCTCGTCAGAGTGGCCGAGGCCGATACAACAGCCTACCAAGTGATCAAGATC
AGCGCCCAGAACGCCAGCAAGATCGCCTCCGCCAACCTGAAAAGCCGCGAGAAAGGCCAAACAGCTC

CTGAATGAGATCGTCTGTAAGCAGCTGGCTAAGCGGCGGAAGAACTGGAAGCCTAGCGCCAATAGC
TTCAAGATCGTGATCCCCAGATTTCGGCATGGGCACCCCTGTTCCAGAACGCTAAGTACGGCCTGTTT
ATGGTCAACAGCGACACCTACTACCGGAACCTACCAAGAACTCTGGCTGAGCCGGGAAAACAGAAA
CTGCTGAAAAAGCTGTTTCTCCATCAAATACGAGAAAACCCAGATGAACCACGACGCCCTGCAGGTC
TACAAGGCCATTATCGACAGGTGGAAAAGTTCTTCAAGCTGTACGACATCAACCAGTTCCGCGCC
AAGCTGAGCGGACGCCATCGAGAGATTTGAGAAGCTGCCCATCAATACCGACGGCAACAAGATCGGC
AAGACCGAGACTCTGAGACAGATCCTGATCGGACTGCAGGCCAATGGCACCCGGTCCAACGTGAAG
AACCTGGGCATCAAGACCGATCTGGGCCTGCTGCAAGTCGGCAGCGGAATCAAGCTGGACAAGGAT
ACCCAGATCGTGTATCAGAGCCCCCTCCGGCCTGTTTAAAGCGGAGAATCCCCTGGCTGACCTGCC
AAGAAGAAGAGGAAGGTG

LrhCas9/NLS 단백질 서열 (서열 번호: 4)

MTKLNQPYGIGLDIGSNSIGFAVVDANSHLLRLKGETAIGARLFREGQSAADRRGSRTTRRLSRT
RWRLSFLRDFPAPHITKIDPDFFLRQKYSEISPKDKDRFKYEKRLFNDRDTEAFYEDYPSMYHLRL
HLMTHTHKADPREIFLAITHHILKSRGHFLTPGAAKDFNTDKVDLEDIFPALTEAYAQVYPDLELTF
DLAKADDFKAKLLDEQATPSDTQKALVNLLLSSDGEKEIVKKRKQVLTEFAKAITGLKTKFNLALG
TEVDEADASNWQFSMGQLDDKWSNIETSMTDQGTEIFEQIQELYRARLLNGIVFAGMSLSQAKVAD
YGQHKEDLELFKTYLKKLNDHELAKTIRGLYDRYINGDDAKPFLREDFVKALTKEVTAHPNEVSEQ
LLNRMGQANFMLKQRTKANGAIPQLQORELDQIIANQSKYYDWLAAPNPVEAHRWKMPYQLDELL
NFHIPYYVGPLITPKQQAESGENVFAMVVRKDPGSGNITPYNFDEKVDREASANTFIQRMKTTDTYL
IGEDVLPKQSLLYQKYEVNLNELNNVRINNECLGTDQKQRLIREVFERHSSVTIKQVADNLVAHGDF
ARRPEIRGLADEKRFLSSLSTYHQLKEILHEAIDDPKLLDIENIITWSTVFEDHTIFETKLAIEIE
WLDPKKINELSGIRYRGWGQFSRKLDDGLKLGNHGTIVIQELMSNHNLMQILADETLKETMTELNQ
DKLKTDDIEDVINDAYTSPSNKKALRQVLRVVEDIKHAANGQDPSWLFETADGTGTAGKRTQSRQ
KQIQTVYANAAQELIDSAVRGELEDKIADKASFTDRLVLYFMQGGRIYTGAPLNIDQLSHYDIDH
ILPQSLIKDDSLDNRVLVNATINREKNNVFASLTFAGKMKATWRKWHEAGLISGRKLRNLMRLPDE
IDKFAGGFVARQLVETRQIIKLTEQIAAAQYPNTKIIAVKAGLSHQLREELDFPKNRDYNHYHHAF
DAFLAARIGTYLLKRYPKLAPFFTYGEFAKVDVKKFRENFFIGALTHAKKNIIAKDTGEIVWDER
DIRELDRIYNFKRMLITHEVYFETADLFKQTIYAAKDSKERGGSKQLIPKKQGYPTQVYGGYTQES
GSYNALVRVAEADTTAYQVIKISAQNASKIASANLKSREKQKLLNEIVVKQLAKRRKNWKPSANS
FKIVIPRFGMGTFLFQNAKYGLFMVNSDTYYRNYQELWLSRENQKLLKLLFSIKYEKTQMNHDALQV
YKAIIDQVEKFFKLYDINQFRAKLSDAIERFEKLPINTDGNKIGKTETLRQILIGLQANGTRSNVK
NLGIKTDLGLLQVGSIGKLDKDTQIVYQSPSGLFKRRIPLADLPKKRRKV

PexCas9/NLS DNA 서열 (서열 번호: 5)

ATGGGCAAGACCCACATCATCGGCGTTGGCCTGGATCTCGGCGGCACATACACAGGCACCTTCATC
ACCAGCCATCCTAGCGACGAAGCCGAGCACAGAGATCACAGCAGCGCCTTCACCGTGGTCAACAGC
GAGAACTGAGCTTCAGCAGCAAGAGCAGAACAGCCGTGCGGCACAGAGTGCGGAGCTACAAGGGC
TTCGACCTGCCTAGAAAGCTGCTGCTTCTGGTGGCCGAGTATCAGCTGCTGCAGAAGAAGCAGACA
CTGGCCCTGAGGAAAGAGAGAACTGAGAATCGCCCTGAGCGGCTACCTGAAGAGAAGAGGCTAC
GCCAGAACCAGGCGGAGACAGATACAAGCGTGTGGAATCTCTGGACCCAGCGTGTTCAGCAGC

GCTCCCAGCTTCACCAATTTCTTCAACGACAGCGAGCCCCCTGAACATCCAGTGGGAAGCCATTGCC
 AACTCTCCCGAGACAACAAAGGCCCTGAACAAAGAGCTGAGCGGCCAGAAAGAGGCCGACTTCAAG
 AAGTACATCAAGACCAGCTTTCCCGAGTACAGCGCCAAAGAGATTCTGGCCAACCTACGTGGAAGGC
 AGACGGGCCATTCTGGACGCCAGCAAGTATATCGCCAACCTGCAGAGCCTGGGCCACAAGCACAGA
 AGCAAGTACCTGAGCGACATTCTGCAGGACATGAAGCGGGACAGCCGGATCACCAGACTGAGCGAA
 GCCTTTGGCAGCACCGACAACCTGTGGCGGATCATCGGCCAACATCAGCAATCTGCAAGAACGGGCC
 GTGCGGTGGTACTTCAACGATGCCAAGTTCGAGCAGGGCCAAGAGCAGCTGGATGCCGTGAAGCTG
 AAGAATGTGCTCGTGCGGGCCCTGAAGTATCTGCGGAGTGACGACAAAGAGTGGAGCGCCTCTCAG
 AAGCAGATCATCCAGTCTCTGGAACAGAGCGGGCAGCTGCTGGATGTGCTGGCTGGACTCGACCCC
 GACAGAACAAATCCCTCCATACGAGGACCAGAACACAGACGGCCTCCTGAGGATCAGACCCTGTAT
 CTGAACCCCCAAGGCTCTGAGCAGCGAGTACGGCGAGAAGTGGAAAGTCTTGGGCCAACAAAGTTTGCC
 GGCCTTACCCTCTGCTGACCGAGGATCTGACCGAGATCCTGAAGAACACCGACAGAAAGTCCCGG
 ATCAAGATCAGATCCGATGTGCTGCCCCGACAGCGACTACAGACTGGCCTACATCCTGCAGAGAGCC
 TTCGATCGGTCTATCGCCCTGGACGAGTGCAGCATCAGAAGAACCGCCGAGGACTTCGAGAACGGC
 GTGGTCATCAAGAACGAGAAACTGGAAGATGTGCTGAGCGGACACCAGCTGGAAGAGTTTCTGGAA
 TTTGCCAATCGGTACTACCAAGAGACAGCCAAGGCCAAGAACGGCCTGTGGTTCCCAGAGAACGCC
 CTGCTGGAAAGAGCCGATCTGCACCCCTCCTATGAAGAACAAGATTCTGAACGTGATCGTCGGACAG
 GCCCTGGGAGTGTCTCTGCTGAGGGCACCGATTTCATCGAGGAAATTTGGAACAGCAAAGTGAAA
 GGCCGGTCCACCGTGCGGAGCATCTGTAACGCCATCGAGAATGAGAGAAAGACCTACGGACCCCTAC
 TTCAGCGAGGACTACAAGTTCTGTGAAAACGGCCCTGAAAGAGGGCAAAACCGAGAAAGAGCTGTCC
 AAGAAATTCCGCCGCCGTGATCAAGGTGCTGAAGATGGTGTCTGAGGTGGTGCCTTTATCGGAAAA
 GAGCTGCGGCTGTCTGACGAGGCCAGAGCAAGTTCGACAATCTGTACTCTCTGGCCCAGCTGTAC
 AACCTGATCGAGACAGAGCGGAACGGCTTCAGCAAGGTGTCACTGGCTGCCACCTGGAAAATGCC
 TGGCGGATGACCATGACAGATGGATCCGCCAGTGTGTAGACTGCCTGCCGATTGTGTGCGGCC
 TTCGACGGCTTTATCCGGAAGGCCATCGACCGGAACCTTTGGGAAGTCGCCAAGCGGATTGCCGAG
 GAAGTGAAGAAGTCCGTGCACTTCACCAACGGCACCGTGAAGATCCCTGTGGCCATCGAGGCCAAC
 AGCTTCAACTTTACCGCCAGCCTGACCGACCTGAAGTACATTCACTCAAAAGAACAGAAGCTCAAG
 AAGAAGTTGGAGGACATCCAGCGGAACGAAGAGAATCAAGAGAAGCGGTGGCTGAGCAAAGAGGAA
 CGGATCAGAGCCGACAGCCACGGCATCTGTGCCTATACTGGCAGACCCCTGGATGACGTGGGCGAG
 ATCGATCACATCATCCCCAGAAGCCTGACACTGAAGAAAAGCGAGAGCATCTACAACCTCCGAAGTG
 AACCTGATCTTCGTGTCTGCCAGGGCAATCAAGAAAAGAGAACAACATCTACCTGCTGAGCAAC
 CTCGCCAAGAACTACCTGGCCGCCGTGTTTGGCACAAGCGACCTGAGCCAGATCACCAACGAGATC
 GAGAGCACCGTGTCTGACGCTGAAAGCTGCTGGCAGACTGGGCTACTTCGATCTGCTGAGCGAAAA
 GAGCGGGCCTGCGCCAGACATGCCCTGTTTCTGAATAGCGACTCCGAGGCCAGACGCGCCGTGATT
 GATGTTCTTGGCTCTCGGAGAAAGGCCAGCGTGAACGGAACCCAGGCTTGGTTTGTGCGGTCCATC
 TTCTCCAAAGTGGCGGAGGCACTGGCCGCTTGGACACAAGAAACAGGCAACGAGCTGATCTTTGAC
 GCCATCAGCGTGCCAGCCGCCGATAGCTCTGAGATGAGGAAGAGATTCGCCGAGTACCGGCCCTGAG
 TTCAGAAAGCCCAAAGTGACAGCCTGTGGCCTCTCACAGCATCGACGCCATGTGCATCTATCTGGCC
 GCCTGCAGCGACCCCTTCAAGACCAAGAGAATGGGCTCTCAGCTGGCCATCTACGAGCCCATCAAC
 TTCGATAACCTGTTACCCGGCAGCTGTCAAGTGATCCAGAACACCCCTCGGAACCTCTCCGACAAG
 ACCAATATCGCTAACAGCCCCATCTTCAAAGAGACAATCTACGCCGAGCGGTTCTTGACATCATC
 GTGTCCAGAGGCGAGATTTTCATCGGCTACCCAGCAACATGCCCTTCGAGGAAAAGCCCAACCG
 ATCAGCATCGGCGGCAAGGACCCTTTCAGCATCCTGTCTGTGCTGGGCGCCTACCTGGATAAGGCC
 CCTAGCAGCGAGAAAAGAAAAGCTCACCATCTACCGGTCGTCAAGAACAAGCCTTCGAGCTGTTT
 TCCAAGGTGGCCGGCAGCAAGTTTACCGCCGAAGAAGATAAGGCCGCCAAGATCCTGGAAGCCCTG

CACCTTCGTGACCGTGAAACAGGATGTGGCCGCCACCGTGTCCGATCTGATCAAGAGCAAGAAAGAA
CTGAGCAAGGATAGCATCGAGAACCTGGCCAAGCAGAAGGGCTGCCTGAAGAAGGTGGAATACTCC
AGCAAAGAGTTCAAGTTCAAGGGCAGCCTGATCATCCCTGCCGCCGTGGAATGGGGAAAAGTGCTG
TGGAACGTGTTCAAAGAAAAACACGGCCGAAGAACTGAAGGACGAGAACGCTCTGAGGAAGGCCCTG
GAAGCTGCCTGGCCTAGCTCTTTTCGGCACCAGAAACCTGCACTCTAAGGCCAAGCGGGTGTTCAGC
CTGCCTGTGGTGGCTACACAATCTGGCGCCGTGCGGATCAGACGCAAGACAGCCTTCGGCGACTTC
GTGTACCAGAGCCAGGACACAAACAACCTGTACAGCAGCTTCCCCGTGAAGAACGGCAAGCTGGAT
TGGAGCAGCCCTATCATTACCCCGCTCTGCAGAACCGGAACCTGACCGCCTACGGCTACAGATTCT
GTGGACCACGACAGATCCATCAGCATGAGCGAGTTCAGAGAGGTGTACAACAAGGACGACCTGATG
CGGATCGAGCTGGCCAGGGAACAAGCAGCAGACGCTACCTGAGAGTGGAAATGCCCGGCGAGAAA
TTCCTCGCTTGGTTTGGCGAGAACAGCATCAGCCTGGGCTCCAGCTTCAAGTTCTCTGTGTCCGAG
GTGTTTCGACAACAAAATCTACACCGAGAACGCCGAGTTTACCAAGTTCTCTGCCTAAGCCTAGAGAG
GACAACAAGCACAACGGGACCATCTTTTTTGAACCTCGTGGGCCCCAGAGTGATCTTCAACTACATC
GTTGGCGGAGCCGCCAGCAGCCTGAAAGAAATCTTTAGCGAGGCCGGCAAAGAGCGGAGCCCCAAG
AAGAAGAGGAAGGTG

PexCas9/NLS 단백질 서열 (서열 번호: 6)

MGKTHIIGVGLDLGGTYTGTFITSHPSDEAEHRDHSSAFTVNVNSEKLSFSSKSRTAVRHRVRSYKG
FDLRRRLLLLVAEYQLLQKKQTLAPEERENLRIALSGYLKRRGYARTEAETDTSVLES LDPSVFSS
APSFNTNFNDSEPLNIQWEAIANS PETTKALNKELSGQKEADFKKYIKTSFPEYSAKEILANYVEG
RRAILDASKYIANLQSLGHKHSKYLS DILQDMKRDSRITLSEAFGSTDNLWRIIGNISNLQERA
VRWYFNDAKFEQGGQELDAVKLKNVLVRLKYLRSDDKEWSASQKQIIQSLEQSGDVLVDLAGLDP
DRTIPPYEDQNNRRPPEDQTLYLNPKALSSEYGEKWKSWANKFAGAYPLLTEDLTEILKNTDRKSR
IKIRSDVLPDSYRLAYILQRAFDRSIALDECSIRRTAEDFENGVVIKNEKLEDVLSGHQLEEFLE
FANRYYQETAKAKNGLWFPENALLERADLHPPMKNKILNVIVGQALGVSPAEGTDFIEEIWN SKVK
GRSTVRSICNAIENERKTYGPYFSEDYKFVK TALKEGKTEKELSKKFAAVIKVLKMOVSEVVPFIGK
ELRLSDEAQSKFDNLYSLAQLYNLIETERNGFSKVSLAAHLENAWRMTMTDGS AQCCRLPADCVRP
FDGFIRKAIDRNSWEVAKRIAEEVKKSVDFTNGTVKIPVAIEANSFNFTASLTDLKYIQLKEQKLK
KKLEDIQRNEENQEKRWLSKEERIRADSHGICAYTGRPLDDVGEIDHII PRSLTLKKSESIYNSEV
NLIFVSAQGNQEKKNNIYLLSNLAKNYLA AVFGTSDLSQITNEIESTVLQLKAAGRLGYFDLLSEK
ERACARHALFLNSDSEARRAVIDVLGSRRKASVNGTQAWFVRSIFS KVRQALAAWTQETGNELIFD
AISVPAADSSEMRRKFAEYRPEFRKPKVQPVASHSIDAMCIYLAACSDPFKTKRMGSQLAIYEPIN
FDNLFTGSCQVIQNTPRNFSDKTNIANSP IFKETIYAERFLDIIVSRGEIFIGYPSNMPFEKPNR
ISIGGKDPFSILSVLGAYLDKAPSSEKEKLT IYRVVKNKAFELFSKVAGSKFTAEDKAAKILEAL
HFVTVKQDVAATVSDLIKSKKELSKDSIENLAKQKGCLKKVEYSSKEFKFKGSLIIPAAVEWGKVL
WNVFKENTAEELKDENALRKALEAAWPSSFGTRNLHSAKRVFSLP VVATQSGAVRIRRKTAFGDF
VYQSQDTNNLYSSFPVKNGKLDWSSPIIHPALQNRNLTA YGYRFVDHRSISMSEFREVYNKDDL
RIELAQQGTSSRRYL RVEMPGEKFLAWFGENSISLGSSFKFSVSEVFDNKIYTENAEFTKFLPKPRE
DNKHNGTIFFELVGPRVIFNYIVGGAASSLKEIFSEAGKERSPKKKRKV

McaCas9/NLS DNA 서열 (서열 번호: 7)

ATGGAAAAGAAAGCGGAAAGTCACCCTGGGCTTCGACCTGGGAATCGCCTCTGTTGGATGGGCCATC
 GTGGACAGCGAGACAAACCAGGTGTACAAGCTGGGCAGCAGACTGTTTCGACGCCCCGTGACACCAAC
 CTGGAAAAGAAAGAACCCAGCGGGGACCAGAAAGGCTGCTGCGGAGAGAAAGTACCGGAACCAGAAA
 TTCTACAACCTGGTCAAGCGGACCGAGGTGTTGCGCCTGTCTAGCAGAGAGGCCATCGAGAACAGA
 TTCAGAGAGCTGAGCATCAAGTACCCCAACATCATCGAGCTGAAAACAAAGGCCCTGAGCCAAGAA
 GTGTGCCCCGACGAGATTGCTGCGGATTCTGCACGACTACCTGAAGAACCAGGGGCTACTTCTACGAC
 GAGAAAGAGACAAAAGAGGACTTCGACCAGCAGACCGTGGAAATCCATGCCTAGCTACAAGCTGAAC
 GAGTTCTACAAGAAGTACGGCTACTTCAAAGGCGCCCTGTCTCAGCCTACCGAGAGCGAGATGAAG
 GACAACAAGGACCTGAAAGAGGCATTCTTCTCGACTTCTCCAACAAGAGTGGCTGAAAGAGATC
 AACTACTTCTTCAACGTGCAGAAGAACATCCTGAGCGAGACATTCATCGAAGAGTTCAAGAAGATT
 TTCAGCTTCACCCGGGACATCAGCAAAGGCCAGGCAGCGACAATATGCCCTCTCCTTACGGCATC
 TTCGGCGAGTTTCGGCGACAATGGCCAAGGCGGCAGATACGAGCACATCTGGGACAAGAACATCGGC
 AAGTGCAGCATCTTACCAACGAGCAGAGAGCCCCTAAGTACCTGCCTAGCGCTCTGATCTTCAAC
 TTCCTGAACGAGCTGGCCAACATCAGACTGTACAGCACCAGACAAGAAGAATATCCAGCCTCTGTGG
 AAGCTGAGCAGCATCGATAAGCTGAATATCCTGCTGAACCTGTTCAACCTGCCTATCAGCGAGAAG
 AAGAAAAAGCTGACCAGCACCAACATCAACGACATCGTGAAGAAAGAGTCCATCAAGAGCATCATG
 CTGAGCGTCGAGGACATCGACATGATCAAGGATGAGTGGGCCGGCAAAGAACCCAACGTGTACGGC
 GTTGGACTGAGCGGCCTGAACATCGAGGAAAGCGCCAAAGAGAACAAGTTCAAGTTCCAAGACCTG
 AAGATCCTGAACGTCCTGATCAATCTGCTGGACAACGTGGGCATCAAGTTCGAGTTCAGGACCGC
 AGCGACATCATCAAGAACCTGGAAGTGTGGATAACCTGTACCTGTTCCTGATCTACCAGAAAGAG
 AGCAACAACAAGAGACGCTCCATCGACCTGTTTATCGCCAAAGAACAGTCCCTGAATATCGAGAAC
 CTGAAGCTCAAGCTCAAAGAGTTCCTGCTCGGAGCCGGCAACGAGTTCGAGAACCACAACAGCAAG
 ACCCACAGCCTGTCCAAGAAGGCCATTGACGCCATCCTGCCTAAGCTGCTCGACAACAACGAAGGC
 TGGAAATCTGGAAGCCATCAAGAATTACGACGAGGAAATCAAGAGCCAGATCGAGGACAACCTCCAGC
 CTGATGGCCAAGCAGGATAAGAAGTACCTGAACGACAACCTTCTCAAGGATGCCATTCTGCCGCCA
 AACGTGAAAGTGACCTTCCAGCAGGCCATCCTCATCTTCAACAAGATCATCCAGAAGTTCAGCAAG
 GATTCGAGATCGACAAGGTGCTGATCGAACTGGCCAGAGAGATGACCCAGGACCAAGAGAACGAC
 GCCCTGAAGGGAATCGCTAAGGCCCAGAAGTCCAAGAAAAGCCTGGTGGAAAGAGAGACTCGAAGCC
 AACACATCGACAAGAGCGTGTTCACCGATAAGTACGAGAAGCTTATCTACAAGATTTTCTGTGG
 ATCAGCCAGGACTTTAAGGACCCCTACACCGCGCCCAAGATCAGCGCCAATGAGATCGTGGATAAC
 AAGGTGGAAATCGACCACATCATCCCTTACAGCCTGTGCTTCGACGACAGCAGCGCCAACAAAGTG
 CTGGTGACACAAGCAGAGCAATCAAGAGAAGTCTAACAGCCTGCCGTACGAGTACATCAAGCAGGGC
 CACTCCGGCTGGAAGTGGGACGAGTTACCAAAATACGTGAAGCGGGTGTTCGTGAACAACGTGGAC
 TCTATCCTGAGCAAGAAAGAGCGCCTGAAGAAGTCCGAGAATCTGCTGACCACCAGCTACGACGGC
 TATGAGAAGCTGGGCTTCTTGCCAGAAACCTGAATGACACCAGATACGCCACCATCCTGTTCCGG
 GACCAGCTGAACAATTACGCCGAGCACCACCTGATCGATAACAAGAAAATGTTCAAAGTGATCGCC
 ATGAACGGGGCCGTGACCAGCTTCATCCGGAAGAACATGAGCTACGACAACAAGCTGCGGCTGAAG
 GACAGAAGCGACTTCAGCCACCACGCCCTACGACGCCGCCATCATTGCCCTGTTTCAGCAACAAGACC
 AAGACGCTGTACAACCTGATTGACCCACGCTGAACGGCATCATCAGCAAGAGAAGCGAAGGCTAT
 TGGGTCATCGAGGATCGGTACACAGGCGAGATCAAAGAGCTTAAGAAAAGAGGATTGGACCTCTATC
 AAGAACAATGTGACGGCCCCGAAGATCGCCAAAGAAATCGAGGAATATCTGATCGACCTGGACGAT
 GAGGTGTTCTTCAGCCGGAAGAACTAAGCGCAAGACCAACCGGCAGCTGTACAATGAGACAATCTAC
 GGAATCGCCGCCAAGACCGACGAGGACGGCATCACCACCTACTACAAGAAAGAAAAGTTCTCCATC
 CTGGACGACAAGGACATCTACCTGCGGCTGCTGAGAGAACGCGAGAAGTTCGTGATCAACCAGAGC
 AACCCCGAAGTGATCGACCAGATTATCGAGATCATCGAGAGCTACGGGAAAGAAAACAACATCCCC

AGCCGCGACGAGGCCATCAATATCAAGTACACGAAGAACAAGATTAACTACAACCTCTACCTCAAG
CAGTACATGCGGAGCCTGACCAAGAGCCTGGACCAGTTTCAGCGAGGGCTTCATCAATCAGATGATC
GCCAACAAGACGTTTCGTGCTGTATAACCCACCAAGAACACAACGCGGAAGATCAAGTTCCTGCGG
CTCGTGAACGATGTGAAGATCAACGATATTTCGCAAGAATCAAGTGATCAACAAGTTTAACGGGAAG
AACACGAGCCCAAGGCCTTCTACGAGAATATCAACAGCCTGGGCGCCATCGTGTTCAAGTCCTCC
GCCAACAACCTTCAAGACCCTGTCCATCAACACCCAGATCGCCATCTTCGGAGACAAGAACTGGGAT
ATCGAGGATTTCAAGACCTACAACATGGAAAAAATCGAGAAGTACAAAGAGATATACGGCATCGAC
AAAACCTACAACCTCCACAGCTTTATCTTCCCGGCACAATCCTGCTCGATAAGCAGAACAAAGAG
TTCTACTACATCAGCAGCATCCAGACCGTGAACGACCAAATTGAGCTGAAGTTTCTGAACAAGATC
GAGTTTAAGAACGACGACAACACCTCCGGGGCCAACAAGCCTCCTCGGAGACTGAGATTCCGCATT
AAGTCCATCATGAACAACCTACGAGCAGGTCGACATCAGCCCCCTTCGGCATCAACAAGAAGATATTC
GAGCCCAAGAAGAAGAGGAAGGTG

McaCas9/NLS 단백질 서열 (서열 번호:8)

MEKKRKVTLGFDLGIASVWGAIVDSETNQVYKLGSRFLDAPDTNLERRTQRGTRRLRRRRKYRNQK
FYNLVKRTEVFGLSREAIENRFRELSIKYPNIIELKTKALSQEVCPDEIAWILHDYLNKRGYFYD
EKETKEDFDQQTVESMPYSYKLNIFYKKYGYFKGALSQPTSEMKDNKDLKEAFFFDFFSNKEWLKEI
NYFFNVQKNILSETFIEEFKKIFSFTDISKGPSSDNMPSPYGFGEFGDNGQGGRYEHIWDKNIG
KCSIFTNEQRAPKYLPSALIFNFLNELANIRLYSTDKNKIQLWLKSSIDKLNILLNLFNLPISSEK
KKKLTSTNINDIVKESIKSIMLSVEDIDMIKDEWAGKEPNVYGVGLSGLNIEESAKENKFKFQDL
KILNVLINLLDNVGIKFEFKDRSDIKNLELLDNLFLYLYQKESNNKDSSIDLFIANKNSLNIE
LKLKLKEFLLGAGNEFENHNSKTHSLSKKAIIDAILPKLLDNNEGWNLEAIKNYDEEIKSQIEDNSS
LMAKQDKKYLNDNPLKDAILPPNVKVTQQAAILFNKIIQKFSKDFEIDKVVIELAREMTQDQEND
ALKGIAKAQKSKSLVEERLEANNIDKSVFNDKYEKLIYKIFLWISQDFKDPYTGAKISANEIVDN
KVEIDHIIPYSLCFDDSSANKVLVHKQSNQEKSNLSPYEYIKQGHSGWNWDEFTKYVVRVFNVD
SILSKKERLKKSENLLTTSYDGYEKLGLFARNLNDTRYATILFRDQLNNYAEHHLIDNKKMFKVIA
MNGAVTSFIRKNMSYDNKLRLLKDRSDFSHHAYDAIIALFSNKTLYNLIDPSLNGIISKRSSEGY
WVIEDRYTGEIKELKKEDWTSIKNNVQARKIAKEIEEYLIDLDDEVFFSRKTKRKTNRQLYNETIY
GIAAKTDEDGITNYYKKEKFSILDDKDIYLRLLREREKFINQSNPEVIDQIIIEIESYKKNIP
SRDEAINIKYTKNKINYNLYLKQYMRSLTKSLDQFSEGFINQMIANKTFVLYNPTKNTRKIKFLR
LVNDVKINDIRKNQVINKFNKGKNEPKAFYENINSLGAIVFKSSANNFKTLSINTQIAIFGDKNWD
IEDFKTYNMEKIEKYKEIYGIDKTYNFHSFIFPGTILLDKQNKEFYIISIQTVNDQIELKFLNKI
EFKNDNTSGANKPPRRRLRFGIKSIMNNYEQVDISPFGINKKIPEPKKKRKVPKKKKRKV

MgaCas9/NLS DNA 서열 (서열 번호:9)

ATGAACAACAGCATCAAGAGCAAGCCCGAAGTGACCATCGGCCTGGATCTCGGCGTTGGCTCTGTT
GGATGGGCGCATCGTGGACAACGAGACAAACATCATCCACCACCTGGGCAGCAGACTGTTTCAGCCAG
GCCAAGACAGCTGAGGACAGGCGGTCTTTTCAGAGGCGTCCGGAGACTGATCCGGCGGAGAAAGTAC
AAGCTGAAGAGATTTCGTGAACCTGATCTGGAAGTACAACAGCTACTTCGGCTTCAAGAACAAAGAG
GACATCCTGAACAACCTACCAAGAGCAGCAGAACTGCACAACACCGTGCTGAACCTGAAGCTCGAA
GCCCTGAACGCCAAGATCGACCCCAAGGCTCTGAGCTGGATTCTGCACGACTACCTGAAGAACCGG

[0142]

GGCCACTTCTACGAGGACAACCGGGACTTCAACGTGTACCCACAGAGGAACTGGCCAACTACTTC
GACGAGTTCGGCTACTACAAGGGCATCATCGACAGCAAGAACGACGACGATGATAAGCTGGAAGAG
GGCCTGACCAAGTACAAGTTCAGCAACCAGCACTGGCTGGAAGAAGTGAAGAAGGTCCTGAGCAAC
CAGACCGGCCTGCCTGAGAAGTTCAAAGAGGAATACGAGAGCCTGTTTCAGCTACGTGCGGAACTAC
TCTGAAGGCCCTGGCAGCATCAACAGCGTGTCCCCATACGGCATCTATCACCTGGACGAGAAAGAG
GGCAAAGTCGTCCAGAAGTATAACAACATCTGGGACAAGACCATCGGGAAAGTGCAGCATCTTCCCC
GACGAGTACAGAGCCCCCTAAGAACAGCCCTATCGCCATGATCTTCAACGAGATCAACGAGCTGAGC
ACCATCCGGTCCTACAGCATCTACCTGACCGGCTGGTTCATCAATCAAGAGTTCAAGAAGGCCTAC
CTGAACAAGCTGCTGGACCTGCTGATCAAGACCAACAGCGAGAAGCCCATCGACGCCCCGGCAGTTT
AAGAAGCTGCGGGAAGAGACAATCGCCGAGAGCATCGGCAAGAAACCCTGAAGGACGTGGAAAGC
GAGGAAAAGCTGGAAAAGGACGACCACAAGTGAAGCTGAAGGGCCTGAAGCTGAACACCAACGGC
AAGATCCAGTACAACGACCTGTCTAGCCTGGCCAAGTTCGTGCACAACTGAAGCAGCACCTGAAA
CTGGACTTTCTGCTGGAAGATCAGTACACCCCTCTGGACAAGATCAACTTCCTGCAGAGCCTGTAC
GTGTACCTGGGCAAGCACCTGAGATACAGCAACAGAGTGGACAGCGCCAACTGAAAGAGTTTCAGC
GACAGTCCCGGCTGTTTCGAGAGAGTGCTGCAAGAGCAGAAGGACGGCCTGTTCAAGCTGTTTGAG
CAGACCGACAAGGACGACGAGAAGATCCTGACACAGACCCACAGCCTGTCCACCAAGGCTATGCTG
CTGGCCATCACCAAGATGACCAACCTGGACAATGACGAGGATAACCAGAAGAAACGACAAAAGGC
TGGAACCTTCGAGGCCATCAAGAACCTCGACCAGAAGTTCATCGACATCACCAAGACGAACAACAAC
CTGAGCCTGAAGCAGGACAAGCGCTACCTGGATGACCAGTTCATCAACGACGCCATTCTGAGCCCT
GGCGTGAAGAGAATCCTGCGCGAGGCCACCAAGGTGTTCAACGCCATCCTCAAGCAGTTCTCCGAA
GAGTACGACGTGACCAAGGTGGTTCATCGAGCTGGCCAGAGAGCTGAGCGAAGAGAAAGAACTGGAA
AACACCAAGAACTACAAGAAGCTTATCAAGAAGAACGGCGATAAGATCAGCGAGGGACTGAAAGCC
CTGGGGATCGCCGAGGATAAGATCGAAGAGATCCTGAAGTCTCCCAACCAAGTCTTACAAAGTGCTG
CTGTGGCTGCAGCAGGACCACATCGATCCCTACAGCCAGAAAGAGATCGCCTTCGACGATATCCTG
ACCAAAACCGAAAAGACCGAGATCGACCACATCATTCCTTACTCCATCAGCTTCGACGACAGCAGC
AGCAACAACTGCTGGTGCTGGCCGAGTCCAATCAGGCCAAGTCCAACCAGACACCTTACGAGTTT
ATCAACTCCGGCAAGGCCGAGATCACCTGGGAAGTGTACGAGGCCTACTGCCACAAGTTCAAAAAC
GGCGACTCCAGCCTGCTGGACAGCACCCAGAGAAGCAAGAAATTCGCCAAGATGATGAAGACCGAC
ACCAGCTCTAAGTACGACATCGGCTTTCTGGCCCCGAACCTGAACGACACCAGATACGCCACCATC
GTGTTCCGGGACGCTCTGAAGGACTACGCCAACAACCACCTGGTGAAGATAAGCCCATGTTCAAG
GTCGTGTGCATCAACGGCGGCGTGACCAGCTTCCGTGCGGAAGAACTTTGACCCCAAGTCTTGGTAC
GCCAAGAAGGACAGAGACAAGAACATTCACCACGCCGTGGACGCCAGCATCATCTCCATCTTCAGC
AACGAGACTAAGACCCTGTTCAACCAGCTGACAAAGTTCGCCGACTACAAGCTGTTCAAGAATACC
GACGGCTCTTGAAGAAGATCGATCCTAAGACAGGCGTGGTGTGAGAAAGTGACCGACGAGAATTGG
AAGCAGATCCGCGTGCACAACAGGTGTCCGAGATCGCCAAAGTGATCGACAAGTACATCCAGGAC
AGCAACATCGAGCGGAAGGCCAGATACAGCCGGAAGATCGAGAACAAGACCAATATCAGCCTGTTT
AACGACACCGTGTAATCCGCCAAGAAAGTGGGCTACGAGGATCAGATCAAGCGCAAGAACCTGAAA
ACCTTGACATCCACGAGAGCGCCGAGGAAAACAAGAACAGCAAAGTGAAAAAGCAGTTTCGTGTAC
CGGAAGCTCGTGAACGTGTCCCTGCTGAACAATGACAAGCTGGCCGACCTGTTTCGCCGAGAAAGAA
GATATTCTGATGTACCGGGCCAATCCGTGGGTTCATCAACCTGGCCGAGCAGATTTTCAACGAGTAC
ACCGAGAACAAAAAGATCAAGAGCCAGAACGTGTTTCGAGAAGTACATGCTGGATCTGACCAAGAG
TTCCCCGAGAAGTTTAGCGAGGCCTTCGTGAAGTCCATGATCAGAAACAAGACCGCCATCATCTAC
AACGTGAGAAAGGATGTGGTGCACCGGATCAAGCGGCTGAAGATTCTGAGCAGCGAGCTGAAAGAA
AACAAAGTGGTCCAACGTGATCATCCGCTCCAAGAACGAGAGCGGCACCAAGCTGAGCTACCAGGAC
ACCATCAACTCTATCGCCCTGATGATCATGCGGAGCATCGACCCAACCGCCAAAAAACAGTACATC

AGGGTGCCCCCTGAACACCCCTGAATCTGCACCTGGGCGACCAGGACTTCGACCTGCACAATATCGAC
GCCTATCTGAAGAAGCCTAAGTTCTGTCAAGTACCTGAAGGCCAATGAGATCGGCGACGAGTATAAG
CCTTGCGCGCTGCTGACAAGCGGCACACTGCTGATCCACAAGAAAGACAAGAACTCATGTACATC
AGCAGCTTCCAGAACCTCAACGACCTCATCGAGATCAAGAATCTGATCGAGACAGAGTACAAAGAA
AACGTGGACTCAGACCCCAAGAAGAAGAAAAAGGCCAGCCAGATCCTGAGAAGCCTGAGCGTGATC
CTGAACGATTACATCCTGCTGGATGCCAAGTATAACTTCGACATCCTGGGCCTGTCTAAGAACAAG
ATTGACGAGATCCTCAACAGCAAGCTGGACCTCGACAAGATTGCCAAGCCCAAGAAGAAGAGGAAG
GTG

MgaCas9/NLS 단백질 서열 (서열 번호:10)

MNNSIKSKPEVTIGLDLGVGSVGVWAIVDNETNIIHHLGSRLFSQAKTAEDRRSFRGVRRRLIRRRKY
KLKRFVNLIWKYNSYFGFNKEDILNNYQEQQLHNTVLNLKLEALNAKIDPKALSWILHDYLNKR
GHFYEDNRDFNVYPTTELANYFDEFYGYKGIIDSKNDDDDKLEEGLTQYKFSNQHWLEEVKKVLSN
QTGLPEKFKEEYESLFSYVRNYSEGPGSINSVSPYGIYHLDEKEGKVQKYNNIWDKTIGKCSIFP
DEYRAPKNSPIAMIFNEINELSTIRSYSIYLTGWFINQEFKKAYLNKLLDLLIKTNSEKPIDARQF
KKLREETIAESIGKETLKDVSEEEKLEKDDHKWKLKGLKLNLTNGKIQYNDLSSLAKFVHKLKQHLK
LDLLEDQYTPLDKINFLQSLYVYLGKHLRYSNRVDSANLKEFSDSSRLFERVLQEQQDGLFKLFE
QTDKDEKILTQTHSLSTKAMLLAITRMTNLDNEDNQKNNDKGNFEAIKNFDQKFIDITKTNNN
LSLKQDKRYLDDQFINDAILSPGVKRILREATKVFNAILKQFSEEDVTKVVIELARELSEEKELE
NTKNYKKLIKNGDKISEGLKALGIAEDKIEEILKSPTKSYKVLLWLQDHDIPYSQKEIAFDDIL
TKTEKTEIDHILPYSISFDDSSSNKLLVLAESNQAKSNQTPYEFINSKAEITWEVYEAYCHKFKN
GDSLLDSTQRSKKFAKMMKTDTSKYDIGFLARNLNDTRYATIVERDALKDYANNHLVEDKPMFK
VVCINGGVTSFLRKNFDPKSWYAKKDRDKNIHHAVDASIIISIFSNETKTLFNQLTKFADYKLFKNT
DGSWKKIDPKTGTVSEVTDENWKQIRVRNQVSEIAKVIDKYIQDSNIERKARYSRKIENKTNISLF
NDTVYSARKVGYEDQIKRKNLKTLDIHESAEENKNSKVKKQFVYRKLNVNLSLLNNDKLADLFAEKE
DILMYRANPWVINLAEQIFNEYTENKKIKSQNVFEKYMLDLTKFPEKFSFAFVKSMIRNKTAIIY
NVEKDVVHRIKRLKILSSELKENKWSNVIIIRSKNESGTLKSYQDTINSIALMIMRSIDPTAKKQYI
RVPLNTLNLHLGDQDFDLHNIDAYLKKPKFVKYKANEIGDEYKPWRVLTSGTLLIHKKDKKLMYI
SSFQNLNDLIEIKNLIETEKENVDSDPKKKKKASQILRSLSVILNDYILLDAKYNFDILGLSKNK
IDEILNSKLDLDKIAKPKKKRKV

AglCas9/NLS DNA 서열 (서열 번호:11)

ATGCAGAACATCACCTTCAGCTTCGACGTGGGCTACGCCTCTATCGGATGGGCTGTTGTTTCAGGCC
CCTGCTCAGCCAGAGCAGGACCCTGGAATAGTGGCCTGTGGCACCGTGCTGTTCCCTAGCGACGAT
TGCCAGGCCCTTCAGCGGAGAACTACCGCGGCTGCGGAGGAACATCCGGTCCAGAAGAGTGCGG
ATCGAGCGGATCGGAAAGCTGCTGGTTACGGCCGGAATCCTGACACCTGAGGAAAAGGCCACACCT
GGACACCCCGCTCCATTCTTTTTTGGCAGCCCAGGCTTGGCAGGGCATCAGACAACTGTCTCCACTG
GAAGTGTGGCACATCCTGCGTTGGTACGCCACACAACAGAGGCTACGACAACAATGCCGCTGGGCC
ACCGTGTCCACCAAAGAGGATACCGAGAAAGTCAACAACGCCCCGACCTGATGCAGAAAGTTTGGC
GCCGAGACAATGTGCGCCACACTGTGCCATGCTATGGAACCTGGACATGGACGTGCCCGATGCCGCC
ATGACAGTGTCTACACCAGCCTACAGAACCCCTGAACAGCGCCTTTCTAGAGATGTGGTGCAGAGA

GAGGTGCTGGACATCCTGAGACACAGCGCCAGCCACATCAAAGAGCTGACCCCTGAGATCATCCGG
 CTGATCGCCAGCAAGAGGATCTGAGCACAGAGCAGAGAAGCGAGCTGGCCGCCAAGGGAATTAGA
 CTGGCCAGAAGATACCGGGGCAGCCTGCTGTTTGGACAGCTGCTGCCCAGATTGACAACCGGATC
 ATCAGACGGTGCCCCATCATCTGGGGCCACACATTTGAGCAGGCCAAGACCAGCGGCATGAGCGAG
 AAAGAAGCTCAGGCCCTGGCTGACAAGGTGGCCAAAGTGCCTACAGCCGACTGTCCCGAGTTCTAC
 GCCTACAGATTGCCCCGCATCCTGAACAATCTGAGAGCCAACGGACTGCCCTGCCTGTGGAAGTT
 CGCTGTGAACTGATGCAGGCCGCCAGAGCCGAGGGAAAACTGACAGCCGCCAAGATCAAGAAAGAA
 ATCATGAGGCTGATGGGCGACGTCGAGAGCAACATCGACCACTACTTCCATCTGCACCCCGACAGC
 GAGGAAGCCCTGATTCTCGATCCCGCTATGGAATGCCTGCACCGGACCGGACTGTACGATGCCCTC
 AGCTCTGTCTGTCGGAAGAGTGGCCCTGACCAGACTGCGGAGAGGCCAAAATCTGTACCCCTGCCTAC
 CTGCGGGACATGATGCTGAGACACGGCGAGGATACCCAGGCTCTGGATCTGGCCATTGCCAAGCAG
 CAGGGAAGAAAGGCCCTCGGCCCTAGAAAGAACGACACAGATGCCAGCGCCGACGCCAGCATTGCA
 TGGCAAGATAAGCCCCCTGGCTCCTAAGACAGCCTCTGGCAGAGCCCCCTTATGCCAGACCAGTTCTG
 AGACAGGCCGTGGACGAGATCATGAATGGCGAGGACCCTACCAGGCCAGCTCTGGATGAACAGCAT
 CCGGACGGCGAGGACAAGCCTTCTCACGGCTGTCTGTATGGCCTGCTGGACCCTGCCAGCAAAGAG
 AACGAGTACCTGAACAAGCTTCCCTGGACGCCCTGACAAACAATCACCTCGTGCGGCACCGGATG
 CTGATCCTGGATAGACTGACCCAGGACCTCGTCAGAGAGTTCGCTGACGGCGATCCCAGCAGAGTG
 GAACGGTTCTGTATCGAAGTGGGCAGAGAGCTGAGCGCCTTCTCTGGCATGACCAGCAAGCAGATC
 CAGTCCGAGCTGAACGAGCGGATGAAGCACTTCAAGAGCGCCGTGGCCTATCTGGCCAAACACGCC
 CCTGATATGGCCACATCTGCCGGCTGATCCGGAAGTGCAGAATCGCTATGGACATGAACCTGGCAG
 TGCCCTTTACCCG6CCAGACCTACATGCCCTACGACCTGCCCTAAGCTGGAACGCGAGCACATCGTG
 CCTACGCCAACAGAAAGACAGATGCCCTGTCTGCCCTGGTGCTGACATGGCTGGCCGTGAACAAG
 ATGAAGGGCAAGAGAACCCGCTACCACTTTATCAAAGAGTGCAGGGGCCAGAGCGTGCCCGGCAGA
 AATCAGAATATCGTGTCGTGAAAGCAGTACGAGACATTCGTGGAAAAGCTGGACACCAAGGGCCAC
 GCCGACGACGCCAAGAGAAAAAGACCCGGAAGAACTGATGATGGTGGACAGACTGAGCAGCCAG
 GGAACAAACGGCGAGTCTGAGCTGGATTTACCGAGGGCATGATGACCCAGAGCAGCCACCTGATG
 AAGATCGCCGCTAGAGGCGTGCGGAAGAACTTTCTCACGCCACCGTGACATGATCCCTGGCGCT
 ATTACTGGCACTGTGCGCAAGGCTTGGAAGGTGGCAGGATGCCTGGCCGGCATTTGTCTGAAGCC
 GTCGATCCCGTGACACACAGAATCCAGGACAAAGAGACACTGCGGCGGCTGACCCATCTGCATCAT
 GCATGGATGCCTGCGTGCTGGGACTGATCCCTCACCTGATTCCAGAGCACAGATCCGGCCTGCTG
 AGAAAAGCTCTGGCCGCTAGAAGGCTGCCCGAGAATGTTCCGGCAAGAGGTGGAAAGCGCCGTGTCC
 AAGCGGTACTACACCATCACAAAAGAGAGCAAACCTGGAACCTGCGGGATCTGCCACACACTGAAG
 AACTCTATCGCCGCCAAGCTGAGCGAGGGCAGAGTGGTGCAACACATCCCTGCCGATATGAGCGGA
 GCCAAGCTGGAAGAGACAATCTGGGGAATTGCCCTGGCCAGCACATCGACGACAATAGCGAGGTG
 GTCATCCGGCAGAAGTCCCTGAGCATCGGCAAGGACGGCAACAGAATCAGAACCAGAAAGACCGAC
 AAGCAGGGCAACCCCATCACCGAGAAGGCCTCTAAGCTCGTGGGCATCAAGCCTACCGGCACCAGC
 AAACCTGCAGCCCATCAGAGGCGTGATCATCATCAAGGACAACCTCGCCATTGCTCTGGACCCCGTG
 CCAACCATGATTCCCCACCACAACGTGTACAAGCGGCTGGAAGAACTGCGGAAGCTGAACCACGGT
 AGACATGTGCGGCTGCTGAAAAAGGGCATGCTGATCAGGCTGAGCCACCAGAAAGTCCGGCGACAAG
 AACGGCATGTGGAAAGTGCGGAGCATCCAGGACCAGGGCTCCTCTGGCCTGAAAGTGAATCTGCAG
 AGGCCCTACTACGCCGGCAAGATCGAGGACACCAGAACCAGAGAATTGGAAGAACGTGTCCATCAAG
 GCCCTGCTGAGCCAAGGCATGGAAATCTGCCAACCACCTACTGCGGCACCACACCTCCCAAGAAG
 AAGAGGAAGGTG

Ag1Cas9/NLS 단백질 서열 (서열 번호:12)

MQNITFSFDVGYASIGWAVVQAPAQPEQDPGIVACGTVLFPSDDCQAFQRRNYRRLRRNIRSRRVR
IERIGKLLVQAGILTPEEKATPGHPAPFFFLAAQAWQGIRQLSPLEVWHILRWYAHNNGYDNNAAWA
TVSTKEDTEKVNNARHLMQKFGAETMCATLCHAMELDMVDVDAAMTVSTPAYRTLNSAFPRDQVQR
EVLDIRHSASHIKELTPEIIRLIAQQEDLSTEQRSELAAGIRLARRYRGSLLFGQLLPRFDNRI
IRRCPIIWAHTFEQAKTSGMSEKEAQALADKVAKVPTADCEFYAYRFARILNNLRANGLPLPVEV
RCELMQAAARAEGKLTAARIKKEIMRLMGDVESNIDHYFHLHPDSEALILDPAMECLHRTGLYDAL
SSVVRVALTRLRRGKICTPAYLRDMMLRHGEDTQALDLAIKQQGRKAPRPRKNDTASADASIA
WQDKPLAPKTASGRAPYARPVLRQAVDEIMNGEDPTRPALDEQHPDGEDKPSHGCLYGLLDPASKE
NEYLNKPLPLDALTNHNLVRHRLILDRLTQDLVREFADGDPSSRVERFCIEVGRELSAFSGMTSKQI
QSELNERMKHFSAVAYLAKHAPDMATSAGLIRKCRIAMDMNWQCPFTGQTYMPYDLPKLEREHIV
PYANRKTDAISALVLTWLAVNKMKGKRTAYQFIKECEGQSVPRNQIVSVKQYETFVEKLDTKGH
ADDAKRRKTRKKLMMVDRSSQGTNGESELDFTEGMMTQSSHLMKIAARGVRKNFPHATVDMIPGA
ITGTVRKAWKVAGCLAGICPEAVDPVTHRIQDKETLRLRLTHLHHALDACVLGLIPLHPIEHRSGLL
RKALAARRLPENVRQEVESAVSKRYTITKESKLELRDLPTTLKNSIAAKLSEGRVVQHIPADMSG
AKLEETIWGIAPGQHIDDNSEVIRQKSLSIGKDGNRIRTRKTDKQGNPITEKASKLVGIKPTGTS
KLQPIRGVIIKDNFAIALDPVPTMIPHNNVYKRLEELRKLNHGRHVRLLLKKGMLIRLSHQKSGDK
NGMWKVRSIQDQSSGLKVNLRPPYYAGKIEDTRTENWKNVSIKALLSQGMEILPTTYCGTTPPKK
RKRV

AmuCas9/NLS DNA 서열 (서열 번호:13)

ATGAGCAGAAGCCTGACCTTCAGCTTCGACATCGGCTACGCCTCTATCGGCTGGGCCGTGATTGCC
TCTGCCAGCCACGATGATGCCGATCCTAGCGTGTGTGGCTGTGGCACCGTGCTGTTCCCAAGGAT
GATTGCCAGGCCTTCAAGCGGAGAGAGTACCGGCGGCTGCGGAGAAACATCCGGTCCAGAAGAGTG
CGGATCGAGCGGATTGGTAGACTGCTGGTGCAGGCCAGATCATCACCCCTGAGATGAAGGAAACC
AGCGGACACCCCGCTCCATTCTACCTGGCATCTGAGGCCCTGAAGGGCCACAGAACACTGGCCCT
ATTGAACTGTGGCATGTGCTGCGTTGGTACGCCACAACAGAGGCTACGACAACAACGCCAGCTGG
TCCAACAGCCTGTCTGAGGATGGTGGCAACGGCGAGGATACCGAGAGAGTGAAACACGCCAGGAC
CTGATGGACAAGCACGGCACAGCTACAATGGCCGAGACAATCTGCAGAGAGCTGAAGCTGGAAGAG
GGCAAAGCCGACGCTCCTATGGAAGTGTCTACCCCTGCCTACAAGAACCTGAACACCGCCTTTCCA
CGGCTGATCGTGGAAAAAGAGTGGGAGAAATCCTGGAACTGAGCGCCCTCTGATCCCTGGACTG
ACAGCCGAGATCATCGAGCTGATCGCCAGCATCACCTCTGACCACTGAACAGAGAGGCGTGCTG
CTCCAGCACGGCATTAAAGCTGGCCAGAAGATACAGAGGCAGCCTGCTGTTCCGCCAGCTGATCCCT
AGATTCGACAACAGGATCATCAGCAGATGCCCCGTGACATGGGCCCAAGTGTATGAGGCCGAGCTG
AAGAAGGGCAACAGCGAGCAGTCTGCCAGAGAGAGAGCCGAGAAGCTGAGCAAGGTGCCACCGCC
AATTGTCCCGAGTTCTACGAGTACCGGATGGCCAGAATCCTGTGCAACATCAGAGCCGACGGCGAG
CCTCTGAGCGCCGAGATTAGACGCGAGCTGATGAACAGGCCAGACAAGAGGGAAAGCTGACCAAG
GCCAGCCTGGAAAAGGCCATCTCTAGCCGGCTGGGCAAGAAACCGAGACAACGTGTCCAACCTAC
TTCACACTGCACCCCGACAGCGAGGAAGCCCTGTATCTGAATCCTGCCGTGGAAGTGCTGCAGAGA
AGCGGCATCGGCCAGATTCTGAGCCCCAGCGTGTACAGAATCGCCGCCAACAGACTGCGGAGAGGC
AAGAGCGTGACCCCTAACTACCTGCTGAATCTGCTGAAGTCCAGAGGCGAGTCTGGCGAGGCCCTG
GAAAAAAGATCGAGAAAGAGTCCAAGAAGAAAGAGGCCGACTACGCCGACACACCCCTGAAGCCT

AAGTACGCCACAGGCAGAGCCCCTTACGCCAGAACCCTGCTGAAGAAAGTGGTGGAAAGAGATCCTG
 GATGGCGAGGACCCTACCAGACCTGCTAGAGGGCGAAGCTCACCCTGACGGCGAACTGAAAGCCCAC
 GATGGCTGCCTGTACTGCCTGCTGGATACCGACAGCAGCGTGAACCAGCACCAGAAAGAGCGGAGA
 CTGGACACCATGACCAACAACCACCTCGTGCGGCACCGGATGCTGATCCTGGACAGACTCCTGAAG
 GATCTGATCCAGGACTTCGCGGACGGCCAGAAGGACAGAATCAGCAGAGTGTGCGTGGAAAGTCGGC
 AAAGAGCTGACCACCTTCAGCGCTATGGACAGCAAGAAGATCCAGCGGGAAGTGAACCTGCGGCAG
 AAGTCTCATACCGACGCCGTGAACAGACTGAAGAGAAAGCTTCCAGGCAAGGCCCTGAGCGCCAAC
 CTGATCAGAAAGTGCAGAATCGCAATGGACATGAACTGGACATGCCCTTCACCGGCGCCACATAT
 GCGGATCACGAGCTGGAAAATCTGGAACCTGGAACACATCGTGCCCCACAGCTTCAGACAGAGCAAT
 GCCCTGTCTAGCCTGGTGTGACATGGCCTGGCGTGAACAGGATGAAGGGACAGAGAACCCTGCTAC
 GACTTCGTGGAAACAAGAGCAAGAGAACCCCGTGCCTGACAAGCCCAACCTGCACATCTGCAGCCTG
 AACAACTATCGCGAGCTGGTGGAAAAGCTGGACGACAAGAAGGGACACGAGGACGACAGACGGCGG
 AAGAAGAAAAGAAAGGCCCTGCTGATGGTCCGAGGCCGTGTCTCACAACACCAGAGCCAGAACCAC
 GAGGCCATGAAAGAAATCGGCATGACCGAGGGCATGATGACCCAGAGCAGCCACCTGATGAAGCTG
 GCCTGCAAGAGCATCAAGACCAGCCTGCCTGACGCTCACATCGACATGATTCCAGGCGCCGTGACT
 GCCGAAGTTCGCAAGCCTGGGATGTGTTCCGGCGTGTTCAAAGAACTGTGCCCCGAAGCCGCCGAT
 CCTGACTCTGGCAAGATCCTGAAAGAGAACCTGCGGAGCCTGACTCATCTGCATCACGCCCTGGAT
 GCCTGTGTGCTGGGACTGATCCCCTACATCATCCCCGCTCACCACAATGGCCTGCTGAGAAGAGTC
 CTGGCCATGCGCAGAATCCCCGAGAACTGATCCCTCAAGTGCGGCCCGTGGCCAACCAGAGACAC
 TACGTGCTGAACGACGACGGCCGGATGATGCTGAGGGATCTGAGTGCCAGCCTGAAAGAAAACATC
 CGCGAGCAGCTGATGGAACAGCAGTGTATCCAGCAGTGCCTGCTGATATGGGCGGAGCAGTGTCT
 AAAGAAACAATGCAGCGGGTGTGAGCGTGGACGGCTCTGGCGAAGATGCTATGGTGTCCCTGTCT
 AAGAAGAAGGACGGCAAGAAAGAGAGAAGTCAAGTCAAGGCCCTCAAGCTCGTGGGAGTGTTCCT
 GAGGGCCCCAGCAAGCTGAAAGCTCTGAAGGCCGCCATCGAGATCGACGGCAATTATGGCGTGGCA
 CTGGACCCCCAAGCCTGTGGTCATCAGACACATCAAGGTGTTCAAGAGGATCATGGCCCTCAAAGAG
 CAGAACGGCGGCAAGCCAGTGCGCATCCTGAAAAAGGGCATGCTGATTCACCTGACCAGCAGCAAG
 GACCTAAGCACGCTGCGCTTTGGAGAATCGAGAGCATCCAGGACAGCAAAGGCGGCGTGAAACTG
 GACCTGCAGAGGGCTCATTGCGCCGTGCCTAAGAACAAGACCCACGAGTGCAATTGGAGAGAGGTG
 GACCTGATCTCCCTGCTGAAAAAGTACCAGATGAAGCGCTACCCACCAGCTACACCGGCACACCT
 AGACCCAAGAAGAAGAGGAAGGTG

AmuCas9/NLS 단백질 서열 (서열 번호:14)

MSRSITFSFDIGYASIGWAVIASASHDDADPSVCGCGTVLFPKDDCQAFKRREYRRLRRNIRSRV
 RIERIGRLLVQAQIITPEMKETSGHPAPFYLAASEALKGHRTLAPIELWHVLRWYAHNRGYDNNASW
 SNSLSEDDGNGEDTERVKHAQDLMDKHGTATMAETICRELKLEEGKADAPMEVSTPAYKNLNTAFP
 RLIVEKEVRRILELSAPLIPGLTAEIIEILIAQHHLPTTEQRGVLLQHGIKLARRYRGSLLFGQLIP
 RFDNRIISRCFVTWAQVYEAELKKGNSEQSARERAELSKVFTANCPEFYEYRMARILCNIRADGE
 PLSAEIRRELMNQARQEGKLTAKSLEKAISSRLGKETETNVSNYFTLHPDSEEALYLNPAVEVLQR
 SGIGQILSPSVYRIAANRLRRGKSVTPNYLLNLLKSRGESGEALEKKIEKESKKKEADYADTPLKP
 KYATGRAPYARTVLKKVVEEILDGEDPTRPARGEAHPDGELKAHDGCLYCLLDTDSSVNQHQKERR
 LDTMTNNHLVRHRMLILDRLKDLIQDFADGQKDRISRVCEVVGKELTTFSAMDSKKIQRELTLRQ
 KSHDAVNRLKRKLFGKALSANLIRKCRIAMDMNWTCPFTGATYGDHELENLELEHIVPHSFQSN
 ALSSIVLTWPGVNRMKGQRTGYDFVEQEENPVPDKPNLHICSLNNYRELVEKLDDKKGHEDDRR

[0143]

KKRRKALLMVRGLSHKHQSQNHEAMKEIGMTEGMMTQSSHLMKLACKSIKTSLPDAHIDMIPGAVT
AEVRKAWDVFGVFKELCPEAADPDSGKILKENLRSLSLTHLHHALDACVLGLIPYIIIPAHNGLLRV
LAMRRIPEKLIPOVRPVANQRHYVLNDDGRMMLRDLASLKENIREQLMEQQRVIQHVPAADMGGALL
KETMQRVLSVDGSGEDAMVSLSKKKDGKKEKNQVKASKLVGVFPEGPSKLKALKAAIEIDGNYGVA
LDPKPVVIRHIKVFKRIMALKEQNGGKPVRIILKKGMLIHLTSSKDPKHAGVWRIESIQDSKGGVKL
DLQRAHCAVPKNKTHECNWREVDLISLLKKYQMKRYPTSYTGTPRPKKKRKV

서열 번호:15. OkiCas9/NLS DNA

ATGGCCAGAGATTACAGCGTCGGCCTGGATATCGGCACCTCTTCTGTTGGATGGGCCGCCATCGAC
AACAAGTACACCTGATCCGGGCCAAGAGCAAGAACCTGATTGGCGTGC GGCTGTTTCGATAGCGCC
GTGACCGCCGAGAAGAGAAGAGGCTACAGAACCACCAGACGGCGCTGAGCAGACGGCATTGGAGA
CTGAGACTGCTGAACGACATCTTCGCCGGACCTCTGACCGATTTCGGCGACGAGAATTTCTGGCC
AGACTGAAGTACAGCTGGGTTCACCCTCAAGACCAGAGCAATCAGGCCCACTTTGCCGCCGGACTG
CTGTTTCGACAGCAAAGAGCAGGACAAGGACTTCTACCGGAAGTACCCCAACCATCTATCACCTGAGA
CTGGCCCTGATGAACGACGACCAAGAAGCAGACCTGAGAGAGGTGTACCTGGCCATCCACCACCTG
GTCAAGTACAGAGGCCACTTCTGATCGAGGGCGACGTGAAAGCCGACAGCGCCTTTGATGTGCAC
ACCTTCGCCCGACGCCATCCAGAGATACGCCGAGAGCAACAACCTCCGACGAGAACCTGCTGGGCAAG
ATCGACGAGAAGAAGCTGAGCGCTGCCCTGACCGATAAGCACGGCAGCAAAAGCCAGAGAGCCGAG
ACAGCCGAAACCGCCTTCGACATCTGGACCTGCAGTCCAAGAAGCAGATCCAGGCCATCCTGAAG
TCCGTCGTGGGCAACCAGGCCAATCTGATGGCCATTTTGGCCTGGACAGCAGCGCCATCAGCAAG
GACGAGCAGAAGAATACTAAGTTTCAGCTTCGACGACGCCGACATCGATGAGAAGATCGCCGATTCT
GAGGCCCTGCTGAGCGATACCGAGTTTCAGATTCTGTGCGATCTGAAGGCCGCCCTTTGACGGCCTG
ACACTGAAAATGCTGCTGGGCGACGACAAGACCGTGTCCGCTGCTATGGTTTCGACGGTTCAACGAG
CACCAGAAGGACTGGGAGTACATCAAGAGCCACATCCGGAACGCCAAGAACGCCGGCAATGGCCTG
TACGAGAAGTCTAAGAAGTTTCGACGGCATCAACGCCGCTATCTGGCTCTGCAGTCCGACAACGAG
GACGACAGAAAGAAGGCCAAGAAGATTTTCCAGGACGAGATCAGCTCCGCCGACATTCCCGATGAT
GTGAAGGCCGATTTCTGAAGAAGATTGACGACGATCAGTTCTGCTATCCAGCGGACCAAGAAC
AACGGCACAAATCCCTCACCAGCTGCACCGGAACGAGCTGGAAACAGATCATCGAGAAGCAGGGGATC
TACTACCCATTCTGAAGGACACCTACCAAGAGAACAGCCACGAGCTGAACAAAATCACAGCCCTG
ATCAACTTCAGGGTGCCCTACTACGTGGGCCCTCTGGTGGAAAGAGGAACAGAAAATCGCCGACGAC
GGCAAGAACATCCCCGATCCTACCAACCACTGGATGGTCCGAAAGTCCAACGACACCATCACACCC
TGGAACCTGAGCCAGGTGGTGCACCTGGATAAGAGCGGCAGAAAGATTCATCGAGCGGCTGACCGGC
ACCGATACCTATCTGATCGGAGAGCCACACTGCCCAAGAACAGCCTGCTGTACCAGAAATTCGAC
GTGCTGCAAGAACTGAACAACATCCGCGTGTCCGGCAGACGGCTGGACATTAGAGCCAAGCAGGAT
GCCTTCGAGCACCTGTTCAAGGTGCAGAAAACCGTGTCTGCTACCAATCTGAAGGACTTCTGGTG
CAAGCCGGCTACATCAGCGAGGACACCCAGATTGAAGGACTCGCCGACGTGAACGGAAAGAACTTC
ACAACGCCCTGACCACCTACAACCTACCTGGTGTCTGTGCTGGGCCGCGAGTTCTGGGAAAACCCC
AGCAACGAGGAAGTGTGGAAGAGATTACCGAGCTGCAGACCGTGTTCGAGGACAAGAAGGTGCTG
CGGAGACAGCTGGATCAGCTGGACGGAAGTGAAGCAGACCAACAGAGAGAAGCTTTCCCGGAAGCAC
TACACCGGCTGGGGCAGAAATCAGCAAGAAGCTGCTGACCACCAAGATCGTGCAGAACGCCGACAAG
ATCGATAACCAGACCTTCGATGTGCCCGGATGAACCAGAGCATCATCGACACCTGTACAACACC
AAGATGAACCTGATGGAAATCATCAACAATGCCGAGGATGACTTCGGCGTCAGAGCCTGGATCGAC
AAGCAGAACACCACCGATGGCGACGAGCAGGACGTGTACAGCCTGATCGATGAACTGGCTGGCCCC

AAAGAGATCAAGCGGGGCATCGTGCAGTCCTTTAGAATCCTGGACGACATCACCAAGGCCGTGGGC
TACGCCCCCTAAACGGGTGTACCTCGAATTTGCCAGAAAAGACCCAAGAGAGCCACCTGACCAACAGC
CGGAAGAACCAGCTGAGCACCCCTGCTGAAGAATGCCGGCCTGTCTGAGCTGGTCACACAGGTGTCC
CAGTATGATGCCCGCCGCTCTGCAGAACGACCGGCTGTATCTTTACTTCCTGCAGCAAGGCAAGGAC
ATGTACTCCGGCGAGAAGCTGAATCTGGACAACCTGAGCAACTACGACATCGACCACATCATCCCT
CAGGCTTACACCAAGGACAACAGCCTGGACAACAGAGTGTGGTGTCCAATATCACCAACCGGCGG
AAGTCCGACAGCAGCAACTATCTGCCCCGCTCTGATCGATAAGATGCGGCCCTTTTGGAGCGTGCTG
AGCAAGCAGGGGCTGCTGTCTAAGCACAAAGTTCCGCCAACCTGACCAGAACCAGAGACTTCGACGAT
ATGGAAAAAGAGCGGTTTATCGCCCGCAGCCTGGTGGAAACCCGGCAGATCATTAGAAGACGTGGCC
AGCCTGATTGACAGCCACTTCGGCGGAGAGACAAAAGCCGTGGCCATTAGAAGCAGCCTGACAGCC
GACATGCGGAGATACGTGGACATCCCCAAGAACCGGGACATCAACGACTACCACCACGCCCTTCGAT
GCCCTGCTGTTTAGCACAGTGGGCCAGTACACCGAGAACAGCGGCCCTGATGAAGAAGGGCCAGCTG
TCCGATTCTGCCGGCAACCAGTACAATCGGTACATCAAAGAGTGGATTACGCGCGCCAGGCTGAAC
GCACAGTCCAGAGAGTGAACCCCTTCGGCTTTGTCTGGGCTCCATGAGAAAATGCTGCCCCCTGGC
AAGCTGAACCCCGAGACAGGGGAGATACCCCGAGAGGAAAACGCCGACTGGTCTATCGCCGACCTG
GACTACCTGCACAAAGTGATGAATTTCCGGAAGATCACCGTGACCAGGCGGCTGAAGGATCAGAAA
GGACAGCTGTACGACGAGAGCAGATACCCCTCCGTGCTGCACGACGCCAAGTCTAAGGCCAGCATC
AACTTTGACAAGCACAAGCCCGTGGACCTGTACGGCGGCTTTAGCTCTGCCAAGCCTGCCTATGCC
GCACTGATCAAGTTCAAGAACAAGTTCCGGCTGGTCAACGTGCTGCGGCAGTGGACCTACAGCGAC
AAGAACTCCGAGGACTATATCCTTGAGCAGATCAGAGGCAAGTACCCTAAGGCCGAGATGGTGTCTG
TCTCACATCCCTTACGGCCAGCTGGTCAAGAAAGATGGCGCCCTGGTCACCATCTCTAGCGCCACA
GAGCTGCACAACTTTGAGCAGCTGTGGCTGCCTCTGGCCGACTACAAGCTGATCAACACACTGCTT
AAGACCAAAGAGGACAACCTCGTCGATATCCTGCACAACCGGCTGGATCTCCCCGAGATGACAATC
GAGAGCGCCTTCTACAAAGCCTTCGACTCCATCCTGAGCTTCGCCTTCAACAGATACGCCCTGCAC
CAGAACGCCCTCGTGAACTGCAGGCCACAGGGACGATTTCAATGCCCTGAACACGAGGATAAG
CAGCAGACCCTGGAAAGGATTCTGGACGCTCTGCATGCCTCTCCAGCCAGCAGCGACCTGAAGAAA
ATCAACCTGTCCAGCGGCTTCGGCCGGCTGTTTTCCCTAGCCACTTTACCCTGGCCGACACCGAC
GAGTTCATCTTCCAGAGCGTGACCGGCTGTTCAGCACCAGAAAACAGTGGCTCAGCTGTATCAA
GAGACAAAGCCCAAGAAGAAGAGGAAGTG

Ok1Cas9/NLS 단백질 서열 (서열 번호:16)

MARDYSVGLDIGTSSVGVAAIDNKYHLIRAKSKNLIGVRLFDSAVTAEKRRGYRTRRRRLSRRHWR
LRLNDIFAGPLTDFGDENFLARLKYSWVHPQDQSNQAHFAAGLLFDSKEQDKDFYRKYPTIYHLR
LALMNDQKHDLEVVYLAIHHLVKYRGHFLIEGDVKADSAFDVHTFADAIQRYAESNNSDENLLGK
IDEKKLSAALTDKHGSKSQRAETAETAFFDILDLSKKQIQAILKSVVGNQANLMAIFGLDSSAISK
DEQKNYKFSFDDADIDEKIADSEALLSDTEFEFLCDLKAFFDGLTLKMLLGDDKTVSAAMVRRFNE
HQKDWEYIKSHIRNAKNAGNGLYEKSCKFDGINAAYLALQSDNEDDRKKAKKIFQDEISSADIPDD
VKADFLKKIDDDQFLPIQRTKNNGTIPHQLHRNELEQIIIEKQGIYYPFLKDTYQENSHELKITAL
INFRVPYVVGPLVEEEQKIADDGKNIPDPTNHWMVRKSNDTITPWNLSQVVDLDKSGRRFIERLTG
TDTYLIGEPTLPKNSLLYQKFDVLQELNNIRVSGRRLDIRAKQDAFEHLFKVQKTVSATNLKDFLV
QAGYISEDTOIEGLADVNGKNFNALTTYNYLVSVLGREFVENPSNEELLEIEITELQTVFEDKKVL
RRQLDQLDGLSDHNREKLSRKHYTGWGRISKLLTTKIVQNADKIDNQTDFVPRMNQSIIDTLYNT
KMNLMEIINNAEDDFGVRAWIDKQNTTDGDEQDVYSLIDELAGPKEIKRGIVQSFRLDDITKAVG

YAPKRVYLEFARKTQESHLTNSRKNQLSTLLKNAGLSELVTQVSQYDAAALQNDRLYLFLQQGKD
MYSGEKLNLDNLSNYDIDHII PQAYTKDNSLDNRVLVSNITNRRKSDSSNYLPALIDKMRPFWSVL
SKQGLLSKHKFANLTRTRDFDDMEKERFIARSLVETRQIIKNVASLIDSHFGGETKAVAIRSSLTA
DMRRYVDIPKNRDINDYHHAFDALLFSTVGQYTENSGLMKKGQLSDSAGNQYNRYIKEWIHAARLN
AQSQRVNPFGFVVGSMRNAAPGKLN PETGEITPEENADWSIADLDYLHKVMNFRKITVTRRLKDQK
GQLYDESRYPVSLHDAKSKASINFDKHKPVDLYGGFSSAKPAYAALIKFKNKFRLVNVLRQWTSYD
KNSEDIYLEQIRGKYPKAEMVLSHIPYGQLVKKDGALVTISSATELHNFEQLWLPLADYKLINTLL
KTKEDNLVDILHNRLDLPENTIESAFYKAFDSILSFAFNRYALHQNALVKLQAHRRDDFNALNYEDK
QQTLEIRILDALHASPASSDLKKINLSSGFGRLFSFSHFTLADTDEFIFQSVTGLFSTQKTVAQLYQ
ETKPKKKRKV

BboCas9/NLS DNA 서열 (서열 번호:17)

ATGAGCCAGCACCCGGCGGTATAGAATCGGCATCGACGTGGGCGCTGAATAGCGTTGGACTGGCCGCC
GTGGAAATCGACGCCAACACGACAATCCTCTGGACGAGATCCCCATCAGCATCCTGAATGCCAG
AGCGTGATCCACGATGGCGGAGTGGACCCTGATGAGGCCAAGTCTGCTACAAGCAGACGGGCTTCT
GCTGGCGTGGCCAGAAGAACAAGACGGCTGCACAAGAGCAAGCGGCAGAGACTGGCCAAGCTGGAC
GAGGTGCTGAATGAGCTGGGCTACCCCGTGGAGATGAGAGCCAGTTTCCAGCCGCGAGCAACCC
TATATCGCTTGGCAAGTGGCGGCCAACTGGCCGAGACATTTCATCCCCGACGTGGAAACCCGGAAG
CGGATGATCTCTATCGCCATCCGGCACATTGCCCGGCATAGAGGATGGCGGAATCCCTACTCTTCT
GTGGCCGACGCCGAGCGGATGAGCCATACACCTTCTCCATTTCATGGTGGAAATACGCCAAGAAGCTG
GACTTCGAGATCAACGACAGACGGACCAACGGCTTCTATCACAGCCCTTGGCAGAGCGTGGACGAG
GAAGGCAAGAGACTGAGCAAGAGCGAGCTGGAAAAGCAGCCCAAGATCGAGGACTGGAACGACAAC
CCCATCAACGGCAAGACAATCGCCCGAGCTGGTTCGTCTCTTGGAAACCCAGACCAAGATCAGA
CGGGATCTGACACACGGCCTGCAGACCGAGAGCACCCCTGAATATCCAGACAGAGAAGCTGCACCAG
AGCGACTACATCCACGAACCTGGAAACCATCTTCGAGCGGCAGCACGTGGACCAGACAACCCGAAGAA
CAGCTGCTGGAAAGCCACCTTCCACACCAAGAATCCTAAGGCCGTGGGAGCCGCCGCTAAGCTCGTT
GGAAAAGATGCCCTGGACAGCCGGTACTACAGAGCCAGCAGAGCCACACCAGCCTTCGAAGAGTAC
AGAGTGATGGCCGCCATCGACACCCCTGCGGATTAGAGAGCACGGCACCGAGAGACAGCTGACCACC
GACGAGAGAAGAAAGCTGTTTCGACTTCATCAAGGGGCTGCCAGCAAAAAGACCAAGAACGAGCCC
AGCATCAGCTCCCTGACCTGGGGAGATGTGGCCGATTTTCTGGGCATCCAGCGGATCGATCTGAGA
GGCCTGGGCTCTCTGAAAGACGGCGAACCTGTGTCTGCCAAGCAGCCTCCTGTGATCGAGACAAAC
GACATCATGCAGAAGGCCCTGATCCAATCGCTGCCTGGTGGTCACAGGCCAACACCAAGAACGG
GACAGATTGCTCGAGTTTCATGAGCAACGCTGGCGCCATCAAGGACACCTCCGACGAAGTGCAGAAC
ATTGACGCGGAGATCAGCCAGCTGCTCGAAGAAGTACCGGCTCTGAGCTGGAATCCCTGGATAAG
ATCACCCCTGACCTCTGGCAGAGCCGCCTACAGCTCTCAGACCCCTGAGAAACATCACCAACTATATG
TACGAGACAGGCTGCGACCTGACCACAGCCAGACAAGAGCTGTACCACGTGGGCAAGAATTGGGCC
CCTCCTGCTCCTCCTATCTACGAGCACACAGGCAACCCCGAGCGTGGACAGAACCTTCAGCATCATC
CACAGATGGCTGTGCAACATGCGGGACAGTACGGCGAGCCCGAGACAGTGAATATCGAGTACGTC
CGCGACGGCTTCAGCAGCACATCTACACAGCTGGCCGAGCAGCGGAGCGGGATAGAAGATACGCC
GACAACTGAAGATGCTGAGCAACTACGAGGGCGCCAGCAGCAGATCAGATGTGCGGAGAATCAAG
GCCCTGCAGAGACAGAACTGCCAGTGCATCTACTGCGGCCGGACCATCACCTTCGAGACATGCCAG
ATGGACCATGTGCTGCCCGGAAAGGCCCTGGATCCGATAGCAAGTTCGAGAACCTGGTGGCCACA
TGCGGCGAGTGCAACAAGTCCAAGAGCGATACCCTGTACATGAAGTGGGCCAAGACATACCCCAAT

ACCAACCTGCAGGACGTGCTGAGAAGAATCCAAGAGTGGTCCAAGGACGGCTGGATGACCGACAAA
AGATGGCGGCAGTACAAAGAGGCCCTGATCCTGAGACTGGAAGCTACCGAGAAGCAAGAGCCCCTG
GACAATCGGAGCATGGAAAGCGTGTCTACATGGCCAGAGAGCTGCGGAACCGGATCTACGGCTTT
TACGGCTGGCAGCAGCAGGACGACGCCCTGAAACAAGGCAGACAGAGGGTGTTCGTGTCCAGCGGC
AGTATGACAGCCGCTGCCAGAAGGACCCCTTTTCGAGTCCCCACTGATTAAGGGCGCCGATGAGGAA
ACCTACGAGAGCAGCCTGCCTTGGCTGGATGGCATGAAGGGCAAGACCAGACTGGATCGGAGACAC
CATGCCGTGGACGCCAGCATCATTGCCATGATGAGGCCCCAGATCGTGAAGATCCTGACAGAGGCC
CAAGAGATCAGAAGCGAGCAGCAGCAGACAAGTACCGGAAGGGCCAGACACCTGACTACGTGTGCAAG
CGGCGGGACTACTGGCGGAATTGGAGAGGCACCCCTGACACCAGAGATGAGGAAGTGTTCAACTAC
TGGGCTGGGGAGCAGCTGAGAACCCTGACCGATCTGGTGTCCAGAAGATGGCCGACGACGAAATC
CCCGTGATCTACCCCAACCAGACTGAGACTCGGCAATGGCAGCGCCCAAGGATACCGTGGTGTCC
ATGATGACCCGGAAGTGGGCGACGAGCTGAGCATCACCGCCATCAACAAAGCCGAAAGCGGAGCC
CTGTACACAGCCCTGACCAGAGACAGCGACTTCGACTGGAAAACCGGCCTGAGCGCCAATCCTAAC
CGGCGGATCAGAGTGACAGATAAGTGGTTCGAGGCCGACGATAACCATCAAGTTTCTGGAACCTGCC
GTGGAAGTGGTGTGAAGAACAACACCAGAGCCAGAATCGACCCGAGGCTCTGGATAAGGTGCAC
AGCAGACTGTACGTGCCCGTCAGAGGCGGAATCGCCGAAGCCGGAATAGCATTACCACGTGCGG
TTCTACAAGATCCCCAAGCTGAACAGCAAGGGCAAGCAGACCGGCAGCATCTACGCCATGCTGAGA
GTGCTGACCATCGACCTGGCCATGAACCAGTACGACAAAGAGACAGGCAAGAAGCAGGACCTGTTT
ACCCTGCCACTGCCTGAAAGCAGCCTGAGCAGAAGATTCAGCGAGCCCCAACTGCGGCAGGCTCTG
ATCGATGGCAGCCGAATATCTCGGATGGGCCGTCTGGACGATGAGCTTGAGATCCCCGCCTTC
GCCAACGCCAGAATCACAGAGGAACAGGCCATTACCGCAGCTTCACCGACAGACTGCTGCACAGC
TTTCCCGGCACACACACAAGTTCAGATTCCGCCGGCTTCTCCCGAACACCGAGATCGCCATTAGACCT
GTGCAGCTGGCCTCTGAGGGCCTGATCGAAACCGATGAGAACCGGAAGAGACAGCAGCTGCGGCTG
ACCCAGCCTAACACCGAGTACAGCAACAGCATCAAGAACGTGCTGAAGTCCGGCCTGCACCTGAAA
GTGAACACCCTGTTCAGACAGGCATCCTGGTCACCAGGGCCAATAGCCAGGGAAAGCAGAGCATC
CGTTTCAGCACAGTGAAGAGCCCAAGAAGAAGAGGAAGGTG

BboCas9/NLS 단백질 서열 (서열 번호:18)

MSQHRRYRIGIDVGLNSVGLAAVEIDANHDNPLDEIPISILNAQSVIHDGGVDPDEAKSATSRRAS
AGVARRTRRLHKSKRQRLAKLDEVLNELGYPVEDESQFPAGSNPYIAWQVRAKLAETFI PDVETRK
RMISIAIRHIAHRGWRNPYSSVADAERMSHTPSPFMVEYAKKLDFEINDRRTNGFYHSPWQSVDE
EGKRLSKSELEKQPKIEDWNDNPINGKTIAQLVSSLEPQTKIRRDLTGHLQTESTLNIQTEKLHQ
SDYIHELETIFERQHVDQTTQEQLLEATFHTKNPKAVGAAAKLVGKDALDSRYRASRATPAFEY
RVMAAIDTLRIREHGTERQLTTDERRKLFDFIKGLPSKKTNEPSISSLTWGDVADFLGIQRIDLR
GLGSLKDGEFVSAKQPPVIETNDIMQKAPDPIAAMWSQANTKERDRFVEFMSNAGAIKDTSDVERN
IDAEISQLEELTGSELES�DKITLTSGRAAYSSQTLRNITNYMYETGCDLTARQELYHVGNWA
FPAPPIYEHTGNPSVDRTFSIIHRWLCNMMDQYGEPE TVNIEYVRDGFSSSTSTQLAEQRERDRRYA
DNLKMLSNYEGASSRSDVRRIKALQRQNCQCIYCGRTITFETCQMDHVLPRKGPSSDSKFENLVAT
CGECNKS KSDTLYMNWAKTYPNTNLQDVLRRIQEWSKDGWMTDKRWRQYKEALILRLEATEKQEPL
DNRSMESVSYMARELRNRIYGFYGWHQDDALKQGRQRFVSSGSMTAAARRTPFESPLIKGADEE
TYESSLPWLDGMKGKTRLDRRHHAVDASIIAMMRPQIVKILTEAQEIRSEQHDKYRKQTPDYVCK
RRDYWRNWRGTPDTRDEEVFNWAGEQLRTLTDLVSQKMADEIPVIYPTRLRLNGSAHKDTVVS
MMTRKVGDELSITAINKAESGALYTALTRDSDFDWKTGLSANPNRRIRVHDKWFEADDTIKFLEPA

[0144]

VEVVLKNNTRARIDFEALDKVHSTLYVFPVRGGIAEAGNSIHVRFYKIPKLNSKKGQTGSIYAMLR
VLTIDLAMNQYDKETGKKQDLFTLPLPESSLSRRFSEPKLRQALIDGTAEYLGWAVVDDELEIPAF
ANARITEEQAINGSFTDRLLHSFPGTHKFRFAGFSRNTEIAIRPVQLASEGLIETDENRKRQQLRL
TQPNTEYSNSIKNVLKSGHLKVNLTFTGTGILVTRANSQKGQSIRFSTVEEPPKKRKRV

AceCas9/NLS DNA 서열 (서열 번호:19)

ATGGGCGGATCTGAAGTGGGAACCGTGCTGTGACTTGGAGACTGGGAGTCGATGTGGGCGAGAGA
TCCATTGGACTGGCCGCCGTGTCTACGAAGAGGACAAGCCCAAAGAAATCCTGGCTGCTGTGTCC
TGGATTACGATGGCGGAGTGGGCGACGAAAGAAGCGGAGCTAGTAGACTGGCCCTGAGAGGCATG
GCCAGAAGGGCTAGACGGCTGCGGAGATTCCGTAGAGCCAGACTGCGCGACCTGGACATGCTGCTG
TCTGAACTCGGATGGACCCCTCTGCCTGACAAGAACGTGTCACTGTGGATGCCTGGCTGGCCAGA
AAGAGACTGGCCGAGGAATACGTGGTGGACGAGACAGAGAGAAGAAGGCTGCTGGGCTACGCCGTG
TCTCACATGGCTAGACATAGAGGCTGGCGGAACCCCTGGACCACCATCAAGGACCTGAAGAACCTG
CCTCAGCCTAGCGACAGCTGGGAGAGAACCAGAGAAAGCCTGGAAGCCCGGTACTCCGTGTCTCTG
GAACCTGGCACAGTTGGACAGTGGGSCCGGATACCTGCTGCAGAGAGCCCTGGCATCAGACTGAAC
CCTACACAGCAGAGCGCCGGAAGAAGGGCCGAACCTGTCTAATGCCACCGCCTTCGAGACAAGACTG
CGGCAAGAGGATGTGCTGTGGGAGCTGAGATGTATCGCCGACGTTCAAGGGCCTGCCTGAGGACGTG
GTGTCCAATGTGATCGACGCCGTGTTCTGCCAGAAAAGACCTAGCGTGCCCGCCGAGAGAATCGGC
AGAGATCCTCTCGATCCAGCCAGCTGAGAGCCAGCAGAGCCTGCCCTGGAATTTCAAGAGTACCGG
ATCGTGGCCGCTGTGGCCAACCTGAGAATCAGAGATGGCAGCGGCAGCAGACCCCTGAGTCTGGAA
GAAAGAAACGCCGTGATCGAGGCCCTGCTGGCCAGACAGAAAAGAGCCTCACTTGGAGCGACATT
GCCCTGGAAATCCTGAAGCTGCCCAACGAGAGCGACCTGACCTCTGTGCTGAAAGAGGATGGCCCA
AGCAGCCTGGCCTACTCTCAGTTCCGCCCTTTTCGATGAGACAAGCGCCCGGATCGCCGAGTTTATC
GCCAAGAACAGACGGAAGATCCCCACATTCCGCCAGTGGTGGCAAGAGCAGGATCGGACCAGTAGA
AGCGATCTGGTGGCTGCCCTGGCCGACAATTCTATTGCCCGCGAGGAAGAACAAGAGCTGCTGGTG
CATCTGCCCGACGCCGAACCTGAAGCTCTGGAAGGACTGGCTCTGCCCTCTGGCAGAGTGGCCTAT
AGCAGACTGACACTGAGCGGCCTGACCAGAGTGATGAGAGATGATGGCGTGGACGTGCACAACGCC
CGCAAGACATGCTTCGGAGTGGACGACAATTGGCGGCCCTCCACTGCCTGCTCTGCATGAAGCTACA
GGACACCCCGTGGTGGATAGAAACCTGGCTATCCTGCGGAAAGTTCTTGAGCAGCGCCACCATGAGA
TGGGGCCCTCCACAGTCTATCGTGGTGGAACTTGCCAGAGGCGCCAGCGAGAGCAGAGAAAGGCAG
GCCGAAGAAGAAGCCGCTCGGAGAGCCACAGAAAGGCCAACGACAGAATTAGAGCCGAACCTCAGA
GCCTCCGGCCTGAGCGATCCTTCTCCTGCCGATCTTGTTAGAGCCCGCCTGCTGGAAGTGTACGAC
TGCCACTGTATGTAAGTGTGGCGCCCTATCTCCTGGGAGAACAGCGAGCTGGATCACATCGTGCCC
AGAACAGATGGCGGATCCAACAGACACGAGAACCTGGCCATTACATGCGGCGCCTGCAACAAAGAA
AAAGGCAGAAGGCCCTTCGCCAGCTGGGCGGAGACAAGCAATAGAGTGACGCTGCGGGACGTGATC
GACCGGGTGCAGAAGCTGAAGTACAGCGGCAACATGTACTGGACCCGGGACGAGTTCAGCCGGTAC
AAGAAAAGCGTGGTGGCCCGGCTGAAGCGGAGAACCTCTGATCCTGAAGTGATCCAGAGCATCGAG
AGCACCGGCTATGCTGCCGTGGCTCTGAGAGATAGACTGCTGAGCTACGGCGAGAAGAATGGCGTG
GCACAGGTGGCCGTTTTTAGAGGCGGAGTGACAGCCGAGGCCAGAAGATGGCTGGACATCTCCATC
GAGCGGCTGTTCACTAGAGTGGCCATCTTCGCCAGAGCACCTCCACCAAGAGGCTGGATAGAAGG
CACCACGCGCTGGATGCTGTGGTGTGACAACACTGACACCCGGCGTGGCCAAGACTGGCTGAT
GCTAGAAGCAGAAGAGTGTCCGCCGAGTCTGGCGCAGACCAAGCGACGTGAACAGACACAGCACC
GAGGAACCTCAGAGCCCCGCTACAGACAGTGGAAAGAGAGCTGTTCTGCCCTGGGCGACCTGCTG

ATTTCTACCGCCGCCAGAGATTCTATCGCCGTGGCTGCTCCTCTGAGACTGAGGCCAACAGGCGCA
CTGCACGAGGAAACCTTGAGAGCCTTTAGCGAGCACACAGTGGGAGCCGCTTGGAAGGGCGCTGAG
CTGAGAGAATCGTGGAACCCGAAGTGTACGCCGCCTTCCTGGCACTTACAGATCCTGGCGGCAGA
TTCCTGAAGGTGTCCCTTAGCGAAGATGTGCTGCCCTGCCGACGAGAACAGGCACATTGTGCTGAGC
GACAGAGTGCTGGGCCCCAGAGACAGAGTGAAACTGTTCCCCGACGACCGGGGCGAGCATCAGAGTC
AGAGGTGGCGCAGCCTATATCGCCAGCTTTCACCACGCCAGAGTGTTTACAGATGGGGAAGCAGCCAC
TCTCCTAGCTTCGCCCTGCTGAGAGTCTCTCTGGCTGATCTGGCTGTGGCTGGCCTGCTTAGAGAT
GGGGTCGACGTGTTTACAGCCGAGCTGCCACCTTGGACTCCCGCTTGAGATATGCCTCTATCGCC
CTGGTCAAGGCCGTGGAAAGCGGCGACGCTAAGCAAGTTGGATGGCTGGTGGCTGGCGACGAACTG
GATTTTGGACCTGAGGGCGTGACAACCGCTGCCGGCGATCTGAGCATGTTCTGAAGTACTTTCCC
GAGCGGCACTGGGTCTGTACCGGCTTCGAAGATGACAAGAGGATCAACCTGAAGCCTGCCTTCCTG
TCTGCCGAACAGGCTGAGGTGCTGAGGACTGAGAGAAGCGACAGACCCGACACACTGACAGAGGCC
GGCGAAATTCTGGCCAGTTCTTCCCTAGATGTTGGCGGGCCACAGTGGCTAAGGTGCTGTGCCAT
CCTGGCTGACCGTGATCAGAAGAAGAGCCCTGGGACAGCTAGGTGGCGGAGAGGACATCTGCCT
TATTCATGGCGGCCTTGAGCGCCGATCCTTGAGTGCGGGAACACCTCCCAAGAAGAAGAGGAAG
GTG

AceCas9/NLS 단백질 서열 (서열 번호: 20)

MGGSEVGTVPVTVRLGVDVGERSIGLAAVSYEEDKPKEILAAVSWIHDGGVGDERSGASRLALRGM
ARRARRLRRFRARRLRDLMLLSELGWTFPLDKNVSPVDLARKRLAEEYVVDETERRRLLGYAV
SHMARHRGRWNPWTTIKDLKNLPQPSDSWERTRESLEARYSVSLEPGTVGQWAGYLLQRAPGIRLN
PTQQSAGRRRAELSNATAFETRLRQEDVLWELRCIADVQGLPEDVVSVIDAVFCQKRSPVPAERIG
RDPLDPSQLRASRACLEFQYRIVAANLRIRDGSGSRPLSLEERNAVIEALLAQTERSILTWSDI
ALEILKLPNESDLTSVPEEDGPSSLAYSQFAPFDETSARIAEFIAKNRRKIPTFAQWWQEQDRTSR
SDLVAALADNSIAGEEEQELLVHLPLDAELEALEGLALPSGRVAYSRLTSLGLTRVMRDDGVDVHNA
RKTCTGVDDNWRPPLPALHEATGHPVVDRLAILRKFLSSATMRWGPPQSIVVELARGASESRERQ
AEEEEARRAHRKANDRIRAE LRASGLSDPSPADLVRARLLELYDCHCMYCGAPISWENSELDHIVP
RTDGGSNRHENLAITCGACNKEKGRRFFASWAETSNRVQLRDVIDRVQKLKYSNGMYWTRDEFSRY
KKSVMARLKRRTSDPEVIQSIESTGYAAVALRDRLLSYGEKNGVAQVAVFRGGVTAEARRWLDISI
ERLFSRVAIFAQSTSTKRLDRRHHAVIDAVVLTTLTPGVAKTLADARSRRVSAEFWRRPSDVNRHST
EEPQSPAYRQWKESCSGLGDLLISTAARDSIAVAAPLRLRPTGALHEETLRAFSEHTVGAAWKGAE
LRRIVEPEVYAAFLALTDPPGRFLKVSPSEDVLPADENRHIVLSDRVLGPRDRVKLFPPDRGSIRV
RGGAAYIASFHHARVFRWGSSHSPSFALLRVSLADLAVAGLLRDGVDVFTAE LPPWTPAWRYASIA
LVKAVESGDAKQVGLVPGDELDFGPEGVTTAAGDLSMFLKYPPERHVVVTGFEDDKRINLKPAFL
SAEQAEVLRTERS DRPDTL TEAGEILAQFFPRCWRATVAKVLCHPGLTVIRRTALGQPRWRRGHLP
YSWRPWSADPWSGGTFFKKKKRKV

AheCas9/NLS DNA 서열 (서열 번호: 21)

ATGGCCTATAGACTGGGCTCGACATCGGCATCACATCTGTTGGATGGGCCGTCGTGGCCCTGGAA
AAGGATGAGTCTGGACTGAAGCCCGTGCGCATCCAGGATCTGGGCGTCAGAATCTTCGACAAGGCC
GAGGATAGCAAGACCGGCGCTTCTCTGGCTCTGCCCAGAAGAGAAGCCAGAAGCGCCAGAAGAAGA

ACCCGGCGGAGAAGGCACAGACTGTGGCGCGTGAAAAGACTGCTGGAACAGCACGGCATCCTGAGC
 ATGGAACAGATCGAGGCCCTGTACGCCAGAGAACAGCAGCCCTGATGTGTATGCCCTGAGAGTG
 GCCGGCCTGGACAGATGTCTGATCGCCGAAGAGATCGCCCGGGTGCTGATTACATTGCCACAGA
 AGAGGCTTCCAGAGCAACAGAAAGAGCGAGATCAAGGACAGCGACGCCGGCAAGCTGCTGAAGGCC
 GTGCAAGAGAACGAGAACCTGATGCAGAGCAAGGGCTACAGAACCGTGGCCGAGATGCTGGTGTCT
 GAGGCCACAAAAGACAGACGCCGAGGGAAAGCTGGTGCACGGCAAGAAGCACGGCTACGTACGAAC
 GTGCGGAACAAGGCCGGCGAGTACAGACACACAGTGTCCAGACAGGCCATCGTGGACGAAGTGCGG
 AAGATTTTCGCCGCTCAGAGAGCCCTGGGCAACGACGTGATGAGCGAGGAACCTGGAAGATAGCTAC
 CTGAAGATCCTGTGCAGCCAGCGGAACCTTCGATGATGGCCCTGGCGGCGATTCTCCTTATGGACAC
 GGAAGCGTTAGCCCCGACGGCGTCAGACAGAGCATCTACGAGAGAATGGTCCGGAAGCTGCACCTTC
 GAGACAGGCCGAGAAGAGAGCCCTAGAAGCAGCTACAGCTTCGAGCGGTTTCAGCTGCTGACCAAG
 GTGGTCAACCTGCGGATCTACCGGCAGCAAGAGGATGGCGGCAGATACCCTTGTGAAGTACCCAG
 ACCGAGCGGGCCAGAGTGATCGATTGTGCCTACGAGCAGACCAAGATCACCTACGGAAAGCTGAGA
 AAGCTGCTGGACATGAAGGACACCGAGAGCTTTGCCGGCCTGACCTACGGCCTGAACAGAAGCAGA
 AACAAGACCGAGGACACCGTGTTCTGTGGAATGAAGTTCTACCAAGAGTCCGCAAGGCCCTGCAG
 AGAGCCGGGGTTTTTCATTACAGGACCTGAGCATCGAGACACTGGACCAGATCGGCTGGATTCTGAGC
 GTGTGGAAGTCCGACGACAACCGGCGGAAGAAGCTGTCTACACTGGGCCTGAGCGACAACGTGATC
 GAAGAAGTGTGCCCCCTGAACGGCTCCAAGTTTGGCCACCTGAGCCTGAAGGCCATCAGAAAGATC
 CTGCCCTTTCCTGGAAGATGGGTACAGCTACGACGTGGCCTGTGAAGTGGCCGGCTATCAGTTTCAG
 GGCAAGACAGAGTACGTGAAGCAGCGGCTGCTGCCCTCCACTTGGAGAAGGCCAAGTGACAAACCCC
 GTTGTGCGCAGAGCACTGAGCCAGGCCATCAAGGTTGTGAACGCCGTGATCAGAAAGCACGGCAGC
 CCAGAGAGCATCCACATCGAACTGGCCAGAGAGCTGAGCAAGAACCTGGACGAGCGGAGAAAGATC
 GAGAAGGCCCCAGAAAGAAAATCAGAAGAACAACGAGCAGATTAAGGACGAGATCCGCGAGATCCTG
 GGATCCGCCCATGTGACCGGAAGAGACATCGTGAAGTACAAGCTGTTCAAACAGCAACAAGAGTTT
 TGCATGTACAGCGGCGAGAAGCTGGACGTGACCAGACTGTTTGAGCCTGGCTATGCCGAGGTGGAC
 CACATCATCCCTTACGGCATCAGCTTCGACGACTCCTACGACAACAAGGTGCTGGTTAAGACCGAG
 CAGAACC GGCGAGAAGGGCAATAGAACCCTCTGGAATACCTGCGGGACAAGCCTGAGCAGAAGGCC
 AAGTTTATCGCCCTGGTGGAAATCTATCCCTCTGAGCCAGAAAAAGAAAAACCACTCCTGATGGAC
 AAGCGGGCCATCGACCTGGAACAAGAGGGCTTCAGAGAGCGGAACCTGAGCGATACCCGGTACATC
 ACACGGGCCCTGATGAACCACATCCAGGCTTGGCTGCTGTTTCGACGAGACAGCCAGCACCAGATCC
 AAGAGGGTCCGTGTGTGTGAATGGCGCCGTGACCGCCTACATGAGAGCTAGATGGGGCTGACAAAG
 GATAGAGATGCCGGCGATAAGCACACGCGCTGATGCTGTGGTGGTGGCCTGTATCGGAGACAGC
 CTGATCCAGAGAGTGACCAATACGACAAGTTCAAGCGGAACGCCCTGGCCGACCGGAACAGATAT
 GTGCAGCAGGTTTCCAAGAGCGAGGGCATCACCCAGTACGTGGACAAAAGAAACCGGCGAGGTGTTT
 ACCTGGGAGTCCCTTCGATGAGCGGAAGTTCTTGCCTAACGAGCCCTGGAACCTTGGCCATTCTTC
 AGGGATGAGCTGCTGGCCAGACTGAGCGACGACCCCTCCAAGAACATCAGAGCCATCGGCCTGCTG
 ACCTACAGCGAGACTGAGCAGATCGATCCCATCTTCGTGTCCAGAATGCCACCAGAAAAGTGACC
 GGCGCAGCCCAAAAGAGACAATCAGATCCCCACGGATCGTGAAGGTGGACGATAACAAGGGCACCC
 GAGATCCAGGTGGTGGTGTCTAAGGTGGCCCTGACCGAGCTGAAGCTGACCAAGACGGCGAAATC
 AAGGATTACTTCAGGCCCCGAGGACGACCCAGACTGTACAACACCCTGAGAGAACGGCTGGTGCAG
 TTCGGCGGAGATGCCAAGGCCGCTTCAAGAACCCTGTACAAGATCAGCAAGGACGGCTCTGTG
 CGGACCCCTGTGCGGAAAGTGAAGATTCAAGAGAAGCTGACACTGGGCGTGCCAGTGCATGGCGGA
 AGAGGAATTGCCGAGAATGGCGGCATGGTCCGAATCGACGTGTTCCGCCAAAGGCGGCAAGTACTAC
 TTCGTGCCCATCTACGTGGCCGACGTGCTGAAGAGAGAGCTGCCCAACAGACTGGCCACCGCTCAC
 AAGCCTTACAGCGAATGGCGCGTGGTGGACGACAGCTACCAGTTCAAGTTCTCTCTGTACCCCAAC

GATGCCGTGATGATCAAGCCCAGCAGAGAGGTGGACATCACCTACAAGGACCGGAAAGAGCCCGTC
GGCTGCCGGATCATGTACTTTGTGTCCGCCAATATCGCCAGCGCCTCCATCAGCCTGAGAACCCAC
GATAACTCCGGCGAGCTGGAAGGACTGGGCATCCAAGGACTGGAAGTGTTCGAGAAATACGTCGTG
GGCCCTCTGGGCGACACACACCCTGTGTACAAAGAACGGCGGATGCCCTTCAGAGTGAACGGAAG
ATGAACCCCAAGAAGAAGAGGAAGGTG

AheCas9/NLS 단백질 서열 (서열 번호: 22)

MAYRLGLDIGITSVGWAVVALEKDESGLPVRIQDLGVRIFDKAEDSKTGASLALPRREARSARRR
TRRRRHRLWRVKRLLEQHGILSMEQIEALYAQRTSSPDVYALRVAGLDRCLIAEEIARVLIHIAHR
RGFQSNRKSEIKDSDAGKLLKAVQENENLMQSKGYRTVAEMLVSEATKTDAGKLVHGKKHGYVSN
VRNKAGEYRHTVSRQAIVDEVKIFAAQRALGNDVMSEELDSYLKILCSQRNFDDGPGGDSPLYGH
GSVSPDGVRSIYERMVGSCTFETGEKRAPRSSYSFERFQLLTQVNLRIYRQOEDGGRYPCELTQ
TERARVIDCAYEQTKITYGKLRKLLDMKDTESFAGLTGYNRSRNTKEDTVFVEMKPHYHEVRKALQ
RAGVFIQDLSETLDQIGWILSVWKSDDNRRKKLSTLGLSDNVIEELLPLNGSKFGLSLKAIKRI
LPFLEDGYSYDVACELAGYQFQGKTEYVKQRLPLPGEGETNPVVRALSAIKVVNAVIRKHGS
PESIHIELARELSKNLDERRKIEKAQKENQKNNEQIKDEIREILGSAHVTGRDIVKYKLFKQQQEF
CMYSGEKLQVTRLFEPGYAEVDHIIIPYGISFDDSYDNKVLVKTEQNRQKGNRTPLEYLRDKPEQKA
KFIALVESIPLSQKKKNHLLMDKRAIDLEQEGFRERNLSDTRYITRALMNHQAWLLFDETASTRS
KRVVCVNGAVTAYMRARWGLTKDRDAGDKHHAADAVVACIGDSLQIRVTKYDKFKRNALADNRNY
VQVSKSEGITQYVDKETGEVFTWESFDERKFLPNEPLEPWPFFRDELLARLSDPSKNIRAIPLL
TYSETEQIDPIFVSRMPTRKVTGAHKETIRSPRIVKVDNKGTEIQVVVSKVALTELKLTGDGEI
KDYFRPEDDPRLYNTLRERLVQFGGDAKAAPKEPVYKISKDGSVRTPVRKVKIQEKLTGVPVHGG
RGIAENGGMVRIDVFAKGGKYFVPIYVADVLKRELPNRLATAHKPYSEWRVDDSYQFKFSLYPN
DAVMIKPSREVDITYKDRKEPVGCRIMYFVSANIASASISLRTHDNSGELEGLGIQGLEVFKEYV
GPLGDTHPVYKERRMPFRVERKMNPKKKRKV

WsuCas9/NLS DNA 서열 (서열 번호: 23)

ATGCTGGTGTCCTTATCTCTGTGGATCTCGGCGGCAAGAATACCGGCTTCTTCAGCTTCACCGAC
AGCCTGGACAATAGCCAGAGCGGCACCGTGATCTACGACGAGAGCTTCGTGCTGAGCCAAGTGGGC
AGAAGAAGCAAGCGGCACAGCAAGCGGAACAACCTGAGAAACAAGCTGGTCAAGCGGCTGTTCTCTG
CTGATCCTGCAAGAGCACACCGGCTGAGCATCGACGTTCTGCCCCGATGAGATCCGGGGCCTGTTC
ACAAGAGAGGCTACACCTACGCCGCTTCGAGCTGGACGAGAAGAAGAAGGACGCCCTGGAAAGC
GATACCTGAAAGAGTTCCTGAGCGAGAAGCTGCAGTCCATCGACAGAGACAGCGACGTGGAAGAT
TTCCTGAACAGATCGCCAGCAACGCCGAGAGCTTTAAGGACTACAAGAAAGGCTTCGAGGCCGTG
TTCGCCAGCGCCACACACAGCCCCAACAAGAAGCTGGAAGTGAAGGACGAGCTGAAGTCCGAGTAC
GGCGAGAACGCCAAAGAAGTGTGCGCGGCTGAGAGTGACCAAAGAGATCCTGGACGAGTTCGAC
AAGCAAGAGAACCAGGGCAACCTGCCTCGGGCCAAGTACTTTGAGGAACTGGGCGAGTATATCGCC
ACCAACGAGAAAGTCAAGAGCTTCTTCGACAGCAACAGCCTGAAGCTGACCGACATGACCAAGCTG
ATCGGCAACATCAGCAACTACAGCTGAAAGAGCTGCGGCGGTACTTCAACGACAAAGAGATGGAA
AAGGGCGACATCTGGATTCCCAACAAGCTGCACAAGATCACCGAGAGATTTGTGCGGAGCTGGCAC
CCCAAGAACGACGCCGATAGACAGAGAAGGGCCGAGCTGATGAAGGACCTGAAGTCCAAAGAAATC

ATGGAAGTCTGACCAACACCGAGCCTGTGATGACAATCCCTCCTTACGACGACATGAACAACAGA
GGCGCCGTGAAGTGTGACACCCTGCGGCTGAATGAGGAATACCTGGACAAACATCTGCCCAACTGG
CGGGATATCGCCAAGAGACTGAACCACGGCAAGTTCAACGACGACCTGGCCGACTCTACCGTGAAG
GGCTACAGCGAGGATAGCACCCCTGCTGCACAGACTGCTGGACACCTCTAAAGAGATCGACATCTAC
GAGCTGCGGGGCAAGAAGCCCCAACGAGCTGCTGGTTAAGACACTGGGCCAGAGCGACGCCAACAGA
CTGTATGGCTTCGCCCCAGAACTACTATGAGCTGATCCGGCAGAAAGTGCAGCGCTGGCATTGCGGTG
CCCGTGAAGAACAAGGATGACTCCCTGAACCTGGAAGATAACTCCAACATGCTGAAGCGGTGCAAC
CACAATCCTCCACACAAGAAGAATCAGATCCACAACCTGGTGGCCGGCATCCTGGGAGTGAACCTG
GATGAGGCCAAGTTCGCGGAGTTCGAGAAAGAGCTTTGGAGCGCCAAAGTGGGCAACAAGAACTG
AGCGCCTACTGCAAGAACATCGAGGAAGTGAAGAACCCACGGCAACACCTTCAAGATCGATATA
GAGGAAGTGGCAAGAAGGACCCCGCCGAGCTGTCCAAAGAGGAAAAGGCCAAGCTGAGACTGACC
GACGACGTGATCCTGAATGAGTGGTCCAGAAAGATCGCCAACTTCTTTGACATCGACGACAAGCAC
CGGCAGCGGTTCAACAACCTGTTGAGCATGGCCAGCTGCACACAGTGATCGACACACCCAGAAGC
GGCTTCAGCTCTACCTGCAAAAGATGCACCGCCGAGAACAGGTTGAGAAGCGAGACAGCCTTCTAC
AACGACGAGACAGGGGAGTTCACAAGAAGGGCCACAGCCACCTGTCAGAGACTGCCCGCTGATACC
CAGAGGCCCTTTCAGCGGAAAGATCGAGCGGTACATCGACAAGCTGGGATACGAGCTGGCCAAGATC
AAGGCTAAAGAACTGGAAGGCATGGAAGCTAAAGAAATCAAGGTGCCCATCATCCTGGAACAGAAC
GCCTTCGAGTACGAGGAAAGCCTGCGGAAGTCTAAGACCGGATCCAACGACAGAGTGATCAACTCC
AAGAAAGACCGGACGGAAGAAAGTGGCCAAGGCCAAGAGAACGCCGAGGACAGGCTGAAGGAC
AAGGACAAGCGGATCAAGGCCTTCAGCAGCGGCATCTGCCCTTACTGCGGAGATACCATCGGAGAT
GACGGCGAGATCGACACATCCTGCCTAGAAGCCACACACTGAAAATCTACGGGACCGTGTTCAC
CCCGAGGGCAATCTGATCTACGTGCACAGAAAGTGCAACCAGGCCAAAGCCGACAGCATCTACAAG
CTGAGCGATATCAAGGCCGGCGTGTGAGCCAGTGGATTGAAGAACAGGTGGCCAACATTAAGGGG
TACAAGACCTTCAGCGTGTGTCCGCCGAACAGCAGAAGGCCCTTAGATACGCCCTGTTCCCTCCAG
AACGACAACGAGGCCCTACAAAAAGGTGGTGGACTGGCTGCGGACCGACAGTCTGCTAGAGTGAAC
GGCACACAGAAGTACCTGGCCAAAAAGATCCAAGAGAAGCTCACCAAGATGCTGCCTAACAAAGCAC
CTGAGCTTCGAGTTTCATCCTGGCCGATGCCACCGAGGTGTCAGAGCTGAGAAGGCAGTACGCCAGA
CAGAACCTCTGCTGGCTAAGGCCGAGAAGCAGGCCCTTCTTCTCACGCCATTGATGCCGTGATG
GCCTTCGTGGCCAGATACCAGAAGGTGTTCAAGGACGGCACCCCTCCTAACGCCGATGAGGTGGCA
AACTGGCTATGCTGGACAGCTGGAACCCCGCCTCTAATGAGCCTCTGACAAAGGGCCTGTCCACG
AACCAGAAAATCGAGAAGATGATCAAGAGCGGGCAGTACGGCCAGAAAAACATGAGAGAGGTGTTT
GGCAAGTCCATCTTCGGAGAGAATGCCATCGGCCGAGAGATACAAGCCCATCGTGGTTCAAGAAGGC
GGCTACTACATCGGCTACCCCGCCACAGTGAAAAAGGGCTACGAACTGAAGAACTGCAAGGTGGTC
ACCAGCAAGAACGATATTGCCAAGCTGGAAAAGATCATCAAGAACCAGGACCTGATCTCTCTGAAA
GAGAATCAGTACATCAAAATCTTCTCCATCAACAAGCAGACCATCAGCGAGCTGAGCAACCGCTAC
TTCAACATGAATTACAAGAACCTGGTCGAGCGGGACAAAGAAATTTGTGGGACTGCTTGAGTTTATC
GTCGAGAACTGCCGGTACTACACCAAGAAAGTGGACGTGAAGTTGCCCCCTAAGTACATCCACGAG
ACAAAGTACCCCTTCTACGATGACTGGCGGAGATTGACGAGGCCCTGGCGGTATCTGCAAGAAAAC
CAGAACAAAGACCGCTCCAAGGACCGCTTCGTGATCGATAAGAGCAGCCTGAACGAGTACTACCAG
CCAGACAAGAATGAGTACAAGCTGGACGTGGACACCCAGCCTATCTGGGACGACTTCTGCCGGTGG
TACTTCTTGACAGATACAAGACCGCCAACGACAAGAAGTCCATCCGCATCAAGGCCCGCAAGACA
TTCTCCCTGCTGGCTGAGTCTGGCGTGCAGGGCAAAGTGTTCGGGGCAAGAGAAAGATCCCTACC
GGCTACGCTATCAGGCCCTGCCTATGGACAACAACGTGATCGCTGGCGATTACGCCAACATTCTG
CTGGAAGCCAAACAGCAAGACCCTGAGCCTGGTGCCCTAAGAGCGGCATCAGCATTGAGAAGCAGCTG
GACAAAAAGCTCGACGTATCAAAAAGACCGACGTGCGCGGCCCTGGCAATCGACAACAACCTCTTC

[0145]

TTCAACGCCGACTTCGACACACACGGCATCCGGCTGATCGTGAAAAACACCAGCGTGAAAGTGGA
AACTTCCCCATCAGCGCCATCGATAAGTCCGCCAAGCGGATGATCTTCAGAGCCCTGTTTGAGAAA
GAGAAGGGGAAGCGCAAGAAAAAGACCACCATCAGCTTCAAAGAAAGCGGCCCTGTGCAGGACTAC
CTCAAGGTGTTCTGAAAAAGATCGTGAAGATCCAGCTGAGAACCGACGGCTCCATCTCCAACATC
GTCGTGCGGAAGAATGCCGCCGATTTACCCTGAGCTTTAGAAAGCGAGCACATCCAGAAACTGCTG
AAGCCCAAGAAGAAGAGGAAGGTG

WsuCas9/NLS 단백질 서열 (서열 번호: 24)

MLVSPISVDLGGKNTGFFSFTDSDLNSQSGTVIYDES FVLSQVGRSRKHSKRNNLRNKLVKRFL
LILQEHGSLSDVLPDEIRGLFNKRGYTYAGFELDEKKKDALES DTLKEFLSEKLQSIDRSDVED
FLNQIASNAESFKDYKKGF EAVFASATHSPNKKLELKDELKSEYGENAKELLAGLRVTKEILDEFD
KQENQGNLPRAKYFEELGEYIATNEKVKSFFDSNSLKLTDMTKLIGNISNYQLKELRRYFNDKEME
KGDWIPNKLHKITERFVRSWHPKNDADRQRAELMKDLKSKEIMELLTTTEPVM TIPPYDDMNR
GAVKCQTLRLNEEYLDKHLPNWRDIAKRLNHGKFNDLADSTVKGYSEDSTLLHRLLDTSKEIDIY
ELRGKPNELLVKTLGQSDANRLYGFAQNYELIRQKVRAGIWPVKNKDDSLNLEDNSNMLKRCN
HNPPHKKNQIHNLVAGILGVKLDEAKFAEF EKELWSAKVGNKKLSAYCKNIEELRKTHGNTFKIDI
EELRKKDPAELSKEEKAKLRLTDDVILNEWSQKIANFFDIDDKHRQRFNNLFSMAQLHTVIDTPRS
GFSSTCKRCTAENRFRSETAFYNDETGEFHKKATATCQRLPADTQRPFSGKIERYIDKLG YELAKI
KAKELEGMEAKEIKVPIILEQNAFEYEESLRKSKTGSNDRVIN SKKDRDGKKLAKAKENAEDRLKD
KDKRIKAFSSGICPYCGDTIGDDGEIDHILPRSHTLKIYGT VFNPEGNLIYVHQKCNQAKADSIYK
LSDIKAGVSAQWIEEQVANIKGYKTFSVLSAEQQKAFRYALFLQNDNEAYKKVVDWLRTDQSARVN
GTQKYLAKKIQEKLTKMLPNKHLSEFILADATEVSELRRQYARQNPLLAKEKQAPSSHAIDAVM
AFVARYQKVPKDGTPPNADEVAKLAMLD SWNPASNEPLTKGLSTNQKIEKMIKSGDYGQKNMREVF
GKSIFGENAIGERYKPIVVQEGGYIIGYPATVKKG YELKNCKVVT SKNDIAKLEKIIKNQDLISLK
ENQYIKIFSINKQTISELNRYFNMNYKNLVERDKEIVGLLEFIV ENCRYT KKVVDVKFAPKYIHE
TKYPFYDDWRRFDEAWRYLQENQNTSSKDRFVIDKSS LNEYYPDKNEYKLDVDTQPIWDDFCRW
YFLDRYKTANDKKSIRIKARKTFSLLAESGVQGVFR AKRKIPTGYAYQALPMDNNVIAGDYANIL
LEANSKTLSLVPKSGISIEKQLDKKLDVIKKT DVRGLAIDNNSFFNADFDTHGIRLIVENTSVKVG
NFPIS AIDKS AKRMIFRALFEKEKGKRKKTTISFKESGPVQDYLKVFLKKIVKIQLR TDGSI SNI
VVRKNAADFTLSFRSEHIQKLLKPKKKRKV

NsaCas9/NLS DNA 서열 (서열 번호: 25)

ATGAAGAAGATCCTGGGCGTCGACCTGGGCATCACCAGCTTTGGATACGCCATCCTGCAAGAGACA
GGCAAGGACCTGTACAGATGCCTGGACAACAGCGTGGTCATGCGGAACAACCCCTACGACGAGAAG
TCTGGCGAGAGCAGCCAGAGCATCCGCAGCACCCAGAAATCCATGCGGCGGGCTGATCGAGAAGCGG
AAGAAACGGATCAGATGCGTGGCCCAGACAATGGAACGCTACGGCATCCTGGACTACTCCGAGACA
ATGAAGATCAACGACCCCAAGAACAACCCGATCAAGAACAGATGGCAGCTGAGAGCCGTGGACGCC
TGGAAAAGACCTCTGAGCCCTCAAGAGCTGTTGCCATCTTTGCCACATGGCCAAGCACC GGGGC
TACAACTCTATCGCCACCGAGGACCTGATCTACGAGCTGGA ACTGGA ACTCGGCCTGAACGACCT
GAGAAAGAGTCCGAGAAGAAGGCCGACGAGCGGAGACAGGTGTACAACGCCCTGAGACACCTGGAA
GAACTGCGGAAGAAGTACGGCGGCGAGACAATCGCCCAGACCATCCACAGAGCTGTGGAAGCCGGC

NsaCas9/NLS 단백질 서열 (서열 번호: 26)

MKKILGVDLGITSFGYAILQETGKDLYRCLDNSVVMRNNPYDEKSGESSQSIRSTQKSMRRLIEKR
KKRIRCVAQTMERYGILDYSETMKINDPKNNPIKNRWQLRAVDWKRPLSPQELFAIFAHMAKHRG
YKSIATEDLIYELELELGLNDPEKESEKKADERRQVYNALRHLEELRKKYGGETIAQTIHRAVEAG
DLRSYRNHDDYEKMIRREDIEEEEIEKVLLRQAEALGALGLPEEQVSELIDELKACITDQEMPTIDES
LFGKCTFYKDELAAPAYSYSYLDYRLYKKLADLNIDGYEVTQEDREKVIEWVEKKIAOGKNNLKKIT
HKDLRKILGLAPEQKIFGVEDERIVKKGKEPRTFVPPFFFLADIAKFKELFASTQKHPDALQIFREL
AEILQRSKTPQEALDRLRALMAGKGIDTDDRELLELFKNKRSGTRELSHRYILEALPLFLEGYDEK
EVQRILGFDDREDYSRYPKSLRHLHLREGNLFKEENPINNHAVKSLASWALGLIADLSWRYGPF
EIILETTRDALPEKIRKEIDKAMREREKALDKIIGKYKKEFPSIDKRLARKIQLWERQKGLDLYSG
KVINLSQLLDGSADIEHIVPQSLGGLSTDYNTIVTLKSVNAAGKGNRLPGDWLAGNPDYRERIGMLS
EKGLIDWKKRKNLLAQSLDEIYTENTHSGKIRATSYLEALVAQVLKRYPPFPDPFLRKNIGVIRMI
PGKVTSTKTRSLGKSKSRETNFHHAEDALILSTLTRGWQNLRLHRLRDNYGKSEAEKELWKKYM
PHIEGLTLADYIDEAFRRFMSKGEESLFYRDMFDTIRSISYWDKKPLSASSHKETVYSSRHEVPT
LRKNILEAFDSLNVIKDRHKLTTTEEFMKRYDKEIRQKLWLHRIGNTNDESYRAVEERATQIAQILT
RYQLMDAQNDKEIDEKFQALKELITSPIEVTGKLLRKMRFVYDKLNAMQIDRGLVETDKNMLGIH
ISKGPNEKLIFRRMDVNNAHELQKERSGILCYLNEMLFIFNKKGLIHYGCLRSYLEKGQGSKYIAL
FNPRFPANPKAQPSKFTSDSKIKQVGIGSATGIIKAHLLDGHVRSYEVFGTLPEGSIEWFKEESG
YGRVEDDPHHPKKKKRV

RsyCas9/NLS DNA 서열 (서열 번호: 27)

ATGGCCGAGAAAGCAGCACAGATGGGGACTCGACATCGGCACCAATTCTATCGGCTGGGCCGTGATC
GCCCTGATCGAAGGCAGACCTGCTGGACTGGTGGCTACCGGCAGCAGAATCTTTAGCGACGGCAGA
AACCCCAAGGACGGCAGCTCTCTGGCCGTCGAGAGAAGAGGACCTCGGCAGATGCGGCGGAGAAGA
GACAGATATCTCCGGCGGAGGGACAGATTATGCAGGCCCTGATCAACGTGGGCCTGATGCCTGGG
GATGCCGCCGCTAGAAAAGCCCTGGTCACCGAGAATCCCTACGTGCTGAGACAGAGAGGCCTGGAC
CAAGCTCTGACCCCTGCCTGAATTTGGCAGAGCCCTGTTCCACCTGAACCAGCGGAGAGGCTTCCAG
AGCAACAGAAAAGACCGATCGGGCCACCGCCAAAGAAAGCGGCAAAGTGAAGAACGCCATTGCCGCC
TTCAGAGCCGGCATGGGCAATGCCAGAACAGTGGGAGAAGCCCTGGCCAGACGACTGGAAGATGGC
AGACCAGTGCGGGCCAGAATGGTCCGACAGGGCAAAGATGAGCACTACGAGCTGTATATCGCCAGA
GAGTGGATCGCCCAAGAGTTCCGATGCCCTGTGGGCCAGCCAGCAGAGATTTATGCTGAGGTGCTG
GCCGACGCCGCCAGAGATAGACTGAGAGCCATCCTGCTGTTCCAGCGGAAGCTGCTGCCTGTGCCT
GTGGGCAAGTGCTTCCTGGAACCTAACCAGCCTAGAGTGGCCGCTGCTCTGCCTAGCGCTCAGAGA
TTCAGACTGATGCAAGAGCTGAACCACCTGAGAGTGATGACCCTGGCCGACAAGAGAGAGAGGGCCC
CTGAGCTTCCAAGAGAGAAAACGATCTGCTGGCCAGCTGGTGGCCAGACCTAAGTGCGGCTTCGAC
ATGCTGCGGAAGATCGTGTTCCGGCGCCAACAAGAGGCCCTACAGATTCACCATCGAGAGCGAGCGG
CGGAAAGAACTGAAGGGCTGTGATACAGCCGCCAAGCTGGCCAAAGTGAATGCCCTGGGAAGTAGA
TGGCAGGCTCTGTCCCTGGACGAGCAGGATAGACTCGTGTGCCTGCTGCTGGACGGCGAGAATGAT
GCTGTGCTGGCTGATGCCCTGCGGGAACACTATGGACTGACAGACGCCCAGATCGACACACTGCTG
GGCCTGTCTTTTGAGGACGGCCACATGAGACTGGGGAGAAGCGCTCTGCTGAGAGTCTGGATGCC
CTGGAATCCGGAAGAGATGAGCAGGGACTGCCCTGTCTTACGATAAGGCTGTTGTGGCTGCCGGC
TATCCAGCTCACACAGCCGATCTGGAACAGGCGAGAGAGATGCACTGCCCTACTACGGCGAGCTG

CTGTGGCGGTATACACAGGATGCCCTACCGCCAAGAACGACGCCGAGAGAAAGTTTCGGCAAGATC
GCCAATCCTACCGTGACATCGGCCTGAATCAGCTGAGAAAAGCTTGTCATGCCCTGATCCAGAGA
TACGGCAAGCCCGCTCAGATCGTGGTGGAACTGGCCAGAAATCTGAAGGCTGGCCTGGAAGAGAAA
GAGCGGATCAAGAAAACAGCAGACCGCCAACCTGGAACGGAACGAGAGAATCCGGCAGAAGCTGCAG
GACGCTGGCGTGCCCGACAACAGAGAAAACCGGCTGCGGATGCGGCTGTTTCGAGGAACCTCGGACAA
GGCAATGGACTGGGCACCCCTTGCTACTCCGGCAGACAGATCAGCCTGCAGAGACTGTTTCAGC
AACGACGTGCAGGTGCAGCACATCCTGCCTTTCAGCAAGACCCCTGGATGACAGCTTCGCCAACAAG
GTGCTCGCCAGCACGACGCCAACAGATACAAGGGCAACAGAGGCCCTTTCGAGGCCTTCGGAGCC
AACAGAGATGGCTACGCCTGGGACGACATTAGAGCCAGAGCAGCCGTGCTGCCCGGAACAAGAGA
AACAGATTTGCCGAGACAGCCATGCAGGACTGGCTGCACAACGAGACTGACTTTCTGGCTCGGCAG
CTGACCGATACCGCCTACCTTAGCAGAGTGGCCAGGCAGTACCTGACCGCCATCTGCAGCAAGGAC
GACGTGTACGTTAGCCCCGGCAGACTGACTGCCATGCTGAGAGCTAAGTGGGGCCTGAACAGAGTG
CTGGATGGCGTGATGGAAGAACAGGGCAGACCGCCGTGAAGAACCAGGATGATCAGACACCAC
GCCATCGACGCCGTGGTTATTGGCGCCACAGATAGAGCCATGCTGCAACAGGTGGCCACACTGGCC
GCTAGAGCTAGAGAACAGGACGCCGAAAGGCTGATCGGCGACATGCCTACGCCTTGGCCTAATTTTC
CTTGAGGACGTGCGGGCTGCCGTGGCCAGATGTGTGGTTTCTCAAGCCCGACACAGGACAGAA
GGCGGCTGCATAACGATACAGCCTACGGCATTGTGGCCGGACCATTCGAGGATGGCAGATACAGA
GTGCGGCACCGGGTGTCCTGTTTCGATCTGAAACCTGGCGACCTGAGCAACGTCCGCTGTGATGCT
CCTCTGCAAGCCGAGCTGGAACCCATCTTCGAGCAGGACGATGCCAGGGCCAGAGAAGTGGCTCTT
ACAGCCCTGGCTGAGCGGTACAGACAGCGGAAAGTGTGGCTGGAAGAACTGATGAGCGTGTGCT
ATCAGACCCAGAGGCGAGGACGGAAGACCTGCCAGATAGCGCTCCTTACAAGGCCCTACAAGGGC
GACTCCCACTACTGCTATGAGCTGTTTCATCAATGAGCGCGGCAGATGGGATGGCGAGCTGATCTCT
ACCTTCCGGGCCAATCAGGCCGCTTACCGCGCGTTTCAGAAATGACCCAGCCAGGTTCAGAAGATAC
ACCGCTGGCGGTAGACCCCTGCTGATGAGACTGTGTATCAACGACTATATCGCCGTGGGCACAGCC
GCCGAGAGGACCATCTTTAGAGTGGTCAAGATGAGCGAGAACAAGATCACTCTGGCCGAGCACTTC
GAAGGCGGAACCTGAAACAGAGGGATGCCGACAAGGACGATCCCTTCAAGTATCTGACAAAGAGC
CCTGGCGCTCTGCGCGATCTGGGAGCTAGAAGAATCTTCGTGGACCTGATCGGCCGCGTGTGAGC
CCAGGCATTAAGGGCGATCCCAAGAAGAAGAGGAAGGTG

RsyCas9/NLS 단백질 서열 (서열 번호: 28)

MAEKQHRWGLDIGTNSIGWAVIALIEGRPAGLVATGSRIFSDGRNPKDGSSLAVERRGPRQMRRRR
DRYLRRRDRFMQALINVGLMPGDAAARKALVTENPYVLRQRLDQALTLPEFGRALFHLNQRRGFQ
SNRKTDRATAKESGKVKNAIAAFRAGMGNARTVGEALARRLEDGRPVRRARMVGGKDEHYELYIAR
EWIAQEPDALWASQQRFAEVLADAARDRLRAILLFQRKLLPVPVGKCFLEPNQPRVAAALPSAQR
FRLMQELNHLRVMTLADKRERPLSFQERNDLAQLVARPKCGFDMRLKIVFGANKEAYRFTIESER
RKELKGCDTAAKLAKVNALGTRWQALSLEQDRLVCLLLDGENDAVLADALREHYGLTDAQIDTLL
GLSFEDGHMRLGRSALLRVLDALESGRDEQGLPLSYDKAVVAAGYPAHTADLENGERDALPYYGEL
LWRYTQDAPTAKNDAERKFGKIANPTVHIGLNQLRKLVNALIQRYGKPAQIVVELARNLKAGLEEK
ERIKKQQTANLERNERIRQKLQDAGVPDNRENRLMRLEELGQGNLGTPCIYSGRQISLQRLFS
NDVQVDHILPFSKTLDDSFANKVLAQHDANRYKGNRGPFEAFGANRDGYAWDDIRARA AVLPRNKR
NRFAETAMQDWLHNETDFLARQLTDAYLSRVARQYLTAICSKDDVYVSPGRLTAMLRKWLNRV
LDGMVEEQGRPAVKNRDDHRHHAIDAVVIGATDRAMLQQVATLAARAREQDAERLIGDMPTPWFNF
LEDVRAAVARCVVSHKPDHGPEGGLHNDTAYGIVAGPFEDGRYRVHRVSLFDLKPGLSNVRCD

[0146]

PLQAELEPIFEQDDARAREVALTALAERYRQRKVLWLEELMSVLPIRPRGEDGKTLPD SAPYKAYKG
DSNYCYELFINERGRWDGELISTFRANQAAYRRFRNDPARFRRYTAGGRPLLMRLCINDYIAVGTA
AERTIFRVVKMSSENKITLAEHFEGGTLKQRDADKDDPFKYLTKSPGALRDLGARRIFVDLIGRVLD
PGIKGDPKKRKV

CdiCas9/NLS DNA 서열 (서열 번호: 29)

ATGAAGTACCACGTGGGCATCGACGTGGGCACCTTTTCTGTTGGACTGGCCGCCATCGAAGTGGAC
GATGCCGGAATGCCTATCAAGACCCTGAGCCTGGTGTCCACATCCACGATTCTGGACTGGACCCC
GACGAGATCAAGAGCGCCGTTACAAGACTGGCCAGCAGCGGAATCGCCAGAAGAACCAGACGGCTG
TACCGGCGGAAGAGAAGAAGGCTGCAGCAGCTGGACAAGTTCATCCAGAGACAAGGCTGGCCCGTG
ATCGAGCTGGAAGATTACAGCGACCCTCTGTACCCCTGGAAAAGTGCGGGCTGAACTGGCTGCCAGC
TATATCGCCGATGAGAAAAGAGCGGGGCGAGAAGCTGTCTGTGGCCCTGAGACACATTGCCAGACAC
AGAGGATGGCGGAACCCCTACGCCAAGGTGTCTCTCTGTATCTGCCTGACGGCCCTAGCGACGCC
TTCAAGGCCATCAGAGAGGAAATCAAGAGAGCCAGCGGCCAGCCTGTGCCTGAAACAGCTACAGTG
GGCCAGATGGTCACCCCTGTGTGAACTGGGCACCCCTGAAAGTTGAGAGGCGAAGGCGGAGTGCTGTCT
GCCAGACTCCAGCAGAGCGATTACGCCAGAGAGATCCAAGAGATTGCGCGGATGCAAGAGATCGGC
CAAGAGCTGTACAGAAAGATCATCGATGTGGTGTTCGCCGCCGAGTCTCCTAAGGGATCTGCCTCT
AGCAGAGTGGGCAAGACCCCTCTGCAGCCCGGCAAGAATAGAGCCCTGAAAGCCTCCGATGCCTTC
CAGAGATACCGGATCGCCGCTCTGATCGGCAACCTGAGAGTTAGAGTGGACGGCGAGAAGAGGATT
CTGAGCGTGGAAAGAGAAAAACCTGGTGTTCGACCACCTGGTCAATCTGACCCCTAAGAAAGAACCC
GAGTGGGTCAACAATCGCCGAGATCCTGGGAATCGACAGAGGCCAGCTGATCGGAACCGCCACCATG
ACAGATGATGGCGAAAGAGCCGGCGCTCGGCCTCCTACACATGACACCAATCGGAGCATCGTGAAC
AGCAGAATCGCCCCCTCTGGTGGACTGGTGGAAAACCGCCTCTGCTCTGGAACAGCACGCTATGGTC
AAGGCCCTGTCCAATGCCGAGGTGGACGACTTCGATTCTCCTGAGGGCGCCAAAGTGCAGGCCTTC
TTTGGCGACCTGGACGACGATGTGCACGCCAAGCTGGATAGCCTGCATCTGCCTGTTGGCAGAGCC
GCCTACAGCGAGGATACACTTGTGCGGCTGACCAGACGGATGCTGAGTGATGGCGTGGACCTGTAT
ACCGCCAGACTGCAAGAGTTTGGCATCGAGCCTAGCTGGACCCCTCCAACACCTAGAATCGGAGAG
CCCGTGGGAAACCCCGCTGTGGACAGAGTGCTGAAAACCGTGTCCAGATGGCTGGAAAGCGCCACC
AAAACATGGGGCGCTCCCGAGAGAGTGATCATCGAACACGTGCGCGAGGGCTTCGTGACCGAGAAA
AGGGCCAGAGAAATGGATGGCGACATGCGGAGAAGGGCCGCCAGAAATGCCAAGCTGTTCCAAGAA
ATGCAAGAAAAGCTGAACGTGCAGGGCAAGCCCTCCAGAGCCGACCTTTGGAGATACCAGAGCGTG
CAGAGACAGAACTGCCAGTGCGCCTACTGTGGCAGCCCTATCACCTTCAGCAACAGCGAGATGGAC
CACATCGTGCTAGAGCCGGCCAGGGATCCACCAACACCAGAGAAAATCTGGTGGCCGTGTGCCAC
AGATGCAACCAGAGCAAGGGCAACACCCCATTCGCCATCTGGGCCAAGAACACCTCTATCGAGGGC
GTGTCCGTGAAAGAAGCCGTGGAAAGAACCAGGCACTGGGTCAACCGATACCGGCATGAGAAGCAC
GACTTCAAGAAATTACCAAGGCCGTGGTGGAAACGGTTCCAGAGGGCCACAATGGACGAGGAAATT
GACGCCCCGAGCATGGAAAGCGTGGCTGGATGGCCAATGAGCTGAGAAAGTAGAGTGGCCAGCAC
TTCGCCAGCCACGGCACAAACAGTCAGAGTGACAGAGGCGAGCCTGACCGCCGAAGCTCGTAGAGCC
TCTGGAATCAGCGCAAGCTGAAGTTCTTTGACGGCGTGGGCAAGAGCAGACTGGACAGAAGGCAC
CACGCCATTGATGCCGCCGTGATCGCCTTACCAGCGACTATGTGGCCGAAACACTGGCCGTGCGG
AGCAACCTCAAAACAGAGCCAGGCTCACAGACAAGAGGCTCCTCAGTGGCGCGAGTTCACAGGCAAA
GATGCCGAACACAGAGCCGCTTGGAGAGTGTGGTGCCAGAAGATGGAAAACTGAGCGCCCTGCTG
ACCGAGGACCTGAGAGATGATAGAGTGGTGGTCATGAGCAACGTGCGCCTGAGACTCGGAAATGGC

AGCGCCCAAAAGAGACAATCGGAAAGCTGAGCAAAGTGAAGCTGTCCAGCCAGCTGAGCGTGTCC
GACATCGATAAGGCCAGCTCTGAGGCCCTTTGGTGCGCCCTGACAAGAGAACCTGGCTTCGACCCC
AAAGAGGGGACTGCCTGCCAATCCTGAGCGGCACATCAGAGTGAATGGCACCCATGTGTACGCCGGC
GACAACATCGGCCCTGTTTCCAGTGTCTGCCGGATCTATCGCTCTGAGAGGCGGATATGCCGAGCTG
GGCAGCTCTTTCCATCACGCCAGGGTGTACAAGATCACAAGCGGCAAGAAACCCGCCTTTGCCATG
CTGAGAGTGTATACCATCGACCTGCTGCCTTACCGGAACCAGGACCTGTTTCAGCGTGGAAGTGAAG
CCCCAGACCATGAGCATGAGACAGGCCGAGAAGAAGCTGAGGGACGCCCTGGCTACAGGCAACGCC
GAATATCTTGATGGCTGGTGGTGGATGACGAGCTGGTGGTCGATACCAGCAAGATCGCCACCGAC
CAAGTGAAGGCTGTGGAAGCCGAAGTGGGAACCATCAGACGTTGGCGCGTGGACGGCTTTTTCAGC
CCCTCTAAGCTGAGACTGCGGCCCTGCAGATGAGCAAAGAGGGCATCAAGAAAGAGAGCGCCCT
GAGCTGTCCAAGATCATTGACAGACCTGGCTGGCTGCCCGCCGTGAACAAGCTGTTTTCTGACGGC
AACGTGACCGCTCGTGGGAGAGATTCTCTGGGCAGAGTGGCGCTGGAAAGCACAGCACATCTGCC
GTGACATGGAAGGTGCAGCCCAAGAAGAAGAGGAAGGTG

CdiCas9/NLS 단백질 서열 (서열 번호:30)

MKYHVGIDVGTFVSVGLAAIEVDDAGMPIKTLVSHIHDSGLDPDKIKSAVTRLASSGIARRTRRL
YRRKRRLQQLDKFIQRQGWPIELESDPLYPWKVRAELAASYIADEKERGEKLSVALRHIAH
RGWRNPYAKVSSLYLPDEPSDAFKAIREEIKRASGQVPETATVGMVTLCELGTLKLRGEGVLS
ARLQQSDHAREIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKASDAF
QRYRIAALIGNLRVRVDGEKRLSVEEKNLVFDHLVNLAPKKEPEWVTIAEILGIDRGQLIGTATM
TDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKALSNAEVDVDFDSPEGAKVQAF
FADLDDVHAKLDSLHLPLVGRAAYSEDTLVRLTRMLADGVLDLYTARLQEF'GIEPSWTPPAPRIGE
PVGNPVAVDRVLKTVSRWLESATKTWGAPERVIEHVREGFVTEKRAREMDGDMRRRAARNAKLFQE
MQEKLNVQGKPSRADLWRYQSVQRQNCQAYCGSPITFSNSEMDHIVPRAGQGSTNTRENLVAVCH
RCNQSKGNTFFAIWAKNTSIEGVSVKEAVERTRHWVTDGMRSTDFKKFTKAVVERFORATMDEEI
DARSMESVAVWMANELRSRVAQHFAHSGTTVRVYRGSALTAEARRASGISGKLEFLDGVGKSRLDRRH
HAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPOWREFTGKDAEHRAAWRVWCQKMEKLSALL
TEDLRDDRVRVMSNVRLRLGNGSAHEETIGKLSKVKLGSQLSVSDIDKASSEALWCALTREPDFDP
KDGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAEELGSSFFHHARVYKITSGKKPAFAM
LRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKLRDALATGNAEYLGWLVDDELVDVDTSKIATD
QVKAVEAEELGTIRRWVRVDGFFGDTRLRLRPLQMSKEGIKKESAPELSKIIDRPGWLPVANKLFSEG
NVTVVRDLSLGRVRLESTAHLPVTWKVQPKKKRKV

Bsm Cas9 sgRNA (서열 번호: 31)

NNNNNNNNNNNNNNNNNNNNNGUCAUAGUCCCCUAAGAUUUAUUGAAACAAUGAUCUUAGGGUUACU
AUGAUAGGGCUUUUCUACUUUAGGGGUAGAGAUGUCCCGCGCGUUGGGGAUCGCCUUAUUGCCCUU
AAAGGGCACUCCCCAUUUUUAUUUUUUU

Lrh Cas9 sgRNA (서열 번호:32)

CTGTGGCGGTATACACAGGATGCCCTACCGCCAAGAACGACGCCGAGAGAAAGTTTCGGCAAGATC
GCCAATCCTACCGTGACATCGGCCTGAATCAGCTGAGAAAGCTTGTCAATGCCCTGATCCAGAGA
TACGGCAAGCCCGCTCAGATCGTGGTGGAACTGGCCAGAAATCTGAAGGCTGGCCTGGAAGAGAAA
GAGCGGATCAAGAAACAGCAGACCGCCAACCTGGAACGGAACGAGAGAATCCGGCAGAAGCTGCAG
GACGCTGGCGTGCCCGACAACAGAGAAAACCGGCTGCGGATGCGGCTGTTTCGAGGAACCTCGGACAA
GGCAATGGACTGGGCACCCCTTGCTACTCCGGCAGACAGATCAGCCTGCAGAGACTGTTTCAGC
AACGACGTGCAGGTGCAGCACATCCTGCCTTTCAGCAAGACCCCTGGATGACAGCTTCGCCAACAAG
GTGCTCGCCAGCACGACGCCAACAGATACAAGGGCAACAGAGGCCCTTTCGAGGCCTTCGGAGCC
AACAGAGATGGCTACGCCTGGGACGACATTAGAGCCAGAGCAGCCGTGCTGCCCGGAACAAGAGA
AACAGATTTGCCGAGACAGCCATGCAGGACTGGCTGCACAACGAGACTGACTTTCTGGCTCGGCAG
CTGACCGATACCGCCTACCTTAGCAGAGTGGCCAGGCAGTACCTGACCGCCATCTGCAGCAAGGAC
GACGTGTACGTTAGCCCCGGCAGACTGACTGCCATGCTGAGAGCTAAGTGGGGCCTGAACAGAGTG
CTGGATGGCGTGATGGAAGAACAGGGCAGACCCGCCGTGAAGAACCAGGATGATCAGACACCAC
GCCATCGACGCCGTGGTTATTGGCGCCACAGATAGAGCCATGCTGCAACAGGTGGCCACACTGGCC
GCTAGAGCTAGAGAACAGGACGCCGAAAGGCTGATCGGCGACATGCCTACGCCTTGGCCTAATTTTC
CTTGAGGACGTGCGGGCTGCCGTGGCCAGATGTGTGGTTTCTCAAGCCCGACACGGACAGAA
GGCGGCTGCATAACGATACAGCCTACGGCATTGTGGCCGGACCATTCGAGGATGGCAGATACAGA
GTGCGGCACCGGGTGTCCTGTTTCGATCTGAAACCTGGCGACCTGAGCAACGTCCGCTGTGATGCT
CCTCTGCAAGCCGAGCTGGAACCCATCTTCGAGCAGGACGATGCCAGGGCCAGAGAAGTGGCTCTT
ACAGCCCTGGCTGAGCGGTACAGACAGCGGAAAGTGTGGCTGGAAGAACTGATGAGCGTGTGCT
ATCAGACCCAGAGCGAGGACGGAAGACCTGCCAGATAGCGCTCCTTACAAGGCCCTACAAGGGC
GACTCCCACTACTGCTATGAGCTGTTTCATCAATGAGCGCGGCAGATGGGATGGCGAGCTGATCTCT
ACCTTCCGGGCCAATCAGGCCGCTTACCGCGCGTTTCAGAAATGACCCAGCCAGGTTCAGAAGATAC
ACCGCTGGCGGTAGACCCCTGCTGATGAGACTGTGTATCAACGACTATATCGCCGTGGGCACAGCC
GCCGAGAGGACCATCTTTAGAGTGGTCAAGATGAGCGAGAACAAGATCACTCTGGCCGAGCACTTC
GAAGGCGGAACCTGAAACAGAGGGATGCCGACAAGGACGATCCCTTCAAGTATCTGACAAAGAGC
CCTGGCGCTCTGCGCGATCTGGGAGCTAGAAGAATCTTCGTGGACCTGATCGGCCGCGTGTGAGC
CCAGGCATTAAGGGCGATCCCAAGAAGAAGAGGAAGGTG

RsyCas9/NLS 단백질 서열 (서열 번호: 28)

MAEKQHRWGLDIGTNSIGWAVIALIEGRPAGLVATGSRIFSDGRNPKDGSSLAVERRGPRQMRRRR
DRYLRRRDRFMQALINVGLMPGDAAARKALVTENPYVLRQRGLDQALTLPEFGRALFHLNQRRGFQ
SNRKTDRTAKESGKVKNAIAAFRAGMGNARTVGEALARRLEDGRPVARMVGGKDEHYELYIAR
EWIAQEPDALWASQQRFAEVLADAARDRLRAILLFQRKLLPVPVGKCFLEPNQPRVAAALPSAQR
FRLMQELNHLRVMTLADKRERPLSFQERNDLAQLVARPKCGFDMRLKIVFGANKEAYRFTIESER
RKELKGCDTAAKLAKVNALGTRWQALSLEQDRLVCLLDGENDAVLADALREHYGLTDAQIDTLL
GLSFEDGHMRLGRSALLRVLDALESGRDEQGLPLSYDKAVVAAGYPAHTADLENGERDALFYYGEL
LWRYTQDAPTAKNDAERKFGKIANPTVHIGLNQLRKLVNALIQRYGKPAQIVVELARNLKAGLEEK
ERIKKQQTANLERNERIRQKLQDAGVPDNRENRLMRLEELGQGNLGTPCIYSGRQISLQRLFS
NDVQVDHILPFSKTLDDSFANKVLAQHDANRYKGNRGPFEEAFGANRDGYAWDDIRARA AVLPRNKR
NRFAETAMQDWLHNETDFLARQLTDAYLSRVARQYLTAICSKDDVYVSPGRLTAMLRKWLNRV
LDGMVEEQGRPAVKNRDDHRHHAIDAVVIGATDRAMLQQVATLAARAREQDAERLIGDMPTPWFNF
LEDVRAAVARCVVSHKPDHGPEGGLHNDTAYGIVAGPFEDGRYVRHRVSLFDLKPGLSNVRCD

[0147]

PLQAELEPIFEQDDARAREVALTALAERYRQRKVLWLEELMSVLPTRPRGEDGKTLFDSAPYKAYKG
DSNYCYELFINERGRWDGELISTFRANQAAYRRFRNDPARFRRYTAGGRPLLMRLCINDYIAVGTA
AERTIFRVVKMSSENKITLAEHFEGGTLKQRDADKDDPFKYLTKSPGALRDLGARRIFVDLIGRVLD
PGIKGDPKKRKV

CdiCas9/NLS DNA 서열 (서열 번호: 29)

ATGAAGTACCACGTGGGCATCGACGTGGGCACCTTTTCTGTTGGACTGGCCGCCATCGAAGTGGAC
GATGCCGGAATGCCTATCAAGACCCTGAGCCTGGTGTCCACATCCACGATTCTGGACTGGACCCC
GACGAGATCAAGAGCGCCGTTACAAGACTGGCCAGCAGCGGAATCGCCAGAAGAACCAGACGGCTG
TACCGGCGGAAGAGAAGAAGGCTGCAGCAGCTGGACAAGTTCATCCAGAGACAAGGCTGGCCCGTG
ATCGAGCTGGAAGATTACAGCGACCCTCTGTACCCCTGGAAAAGTGCGGGCTGAACTGGCTGCCAGC
TATATCGCCCGATGAGAAAAGAGCGGGGCGAGAAGCTGTCTGTGGCCCTGAGACACATTGCCAGACAC
AGAGGATGGCGGAACCCCTACGCCAAGGTGTCTCTCTGTATCTGCCTGACGGCCCTAGCGACGCC
TTCAAGGCCATCAGAGAGGAAATCAAGAGAGCCAGCGGCCAGCCTGTGCCTGAAACAGCTACAGTG
GGCCAGATGGTCACCCCTGTGTGAACTGGGCACCCCTGAAAGTTGAGAGGCGAAGGCGGAGTGCTGTCT
GCCAGACTCCAGCAGAGCGATTACGCCAGAGAGATCCAAGAGATTGCGCGGATGCAAGAGATCGGC
CAAGAGCTGTACAGAAAGATCATCGATGTGGTGTTCGCCGCCGAGTCTCCTAAGGGATCTGCCTCT
AGCAGAGTGGGCAAGACCCCTCTGCAGCCCGGCAAGAATAGAGCCCTGAAAGCCTCCGATGCCTTC
CAGAGATACCGGATCGCCGCTCTGATCGGCAACCTGAGAGTTAGAGTGGACGGCGAGAAGAGGATT
CTGAGCGTGGAAAGAGAAAAACCTGGTGTTCGACCACCTGGTCAATCTGACCCCTAAGAAAGAACCC
GAGTGGGTCAACAATCGCCGAGATCCTGGGAATCGACAGAGGCCAGCTGATCGGAACCGCCACCATG
ACAGATGATGGCGAAAGAGCCGGCGCTCGGCCTCCTACACATGACACCAATCGGAGCATCGTGAAC
AGCAGAATCGCCCCCTCTGGTGGACTGGTGGAAAACCGCCTCTGCTCTGGAACAGCACGCTATGGTC
AAGGCCCTGTCCAATGCCGAGGTGGACGACTTCGATTCTCCTGAGGGCGCCAAAGTGCAGGCCCTTC
TTTGGCGACCTGGACGACGATGTGCACGCCAAGCTGGATAGCCTGCATCTGCCTGTTGGCAGAGCC
GCCTACAGCGAGGATACACTTGTGCGGCTGACCAGACGGATGCTGAGTGATGGCGTGGACCTGTAT
ACCGCCAGACTGCAAGAGATTTGGCATCGAGCCTAGCTGGACCCCTCCAACACCTAGAATCGGAGAG
CCCGTGGGAAACCCCGCTGTGGACAGAGTGCTGAAAACCGTGTCCAGATGGCTGGAAAGCGCCACC
AAAACATGGGGCGCTCCCGAGAGAGTGATCATCGAACACGTGCGCGAGGGCTTCGTGACCGAGAAA
AGGGCCAGAGAAATGGATGGCGACATGCGGAGAAGGGCCGCCAGAAATGCCAAGCTGTTCCAAGAA
ATGCAAGAAAAGCTGAACGTGCAGGGCAAGCCCTCCAGAGCCGACCTTTGGAGATACCAGAGCGTG
CAGAGACAGAACTGCCAGTGCGCCTACTGTGGCAGCCCTATCACCTTCAGCAACAGCGAGATGGAC
CACATCGTGCTAGAGCCGGCCAGGGATCCACCAACACCAGAGAAAATCTGGTGGCCGTGTGCCAC
AGATGCAACCAGAGCAAGGGCAACACCCCATTCGCCATCTGGGCCAAGAACACCTCTATCGAGGGC
GTGTCCGTGAAAGAAGCCGTGGAAAGAACCAGGCACTGGGTCAACCGATACCGGCATGAGAAGCAC
GACTTCAAGAAATTACCAAGGCCGTGGTGGAAACGGTTCCAGAGGGCCACAATGGACGAGGAAATT
GACGCCCCGAGCATGGAAAGCGTGGCTGGATGGCCAATGAGCTGAGAAAGTAGAGTGGCCAGCAC
TTCGCCAGCCACGGCACAAACAGTCAGAGTGACAGAGGCAGCCTGACCGCCGAAGCTCGTAGAGCC
TCTGGAATCAGCGCAAGCTGAAGTTCTTTGACGGCGTGGGCAAGAGCAGACTGGACAGAAGGCAC
CACGCCATTGATGCCGCCGTGATCGCCTTACCAGCGACTATGTGGCCGAAACACTGGCCGTGCGG
AGCAACCTCAAACAGAGCCAGGCTCACAGACAAGAGGCTCCTCAGTGGCGCGAGTTCACAGGCAAA
GATGCCGAACACAGAGCCGCTTGGAGAGTGTGGTGCCAGAAGATGGAAAACTGAGCGCCCTGCTG
ACCGAGGACCTGAGAGATGATAGAGTGGTGGTCATGAGCAACGTGCGCCTGAGACTCGGAAATGGC

AGCGCCCAAAAGAGACAATCGGAAAGCTGAGCAAAGTGAAGCTGTCCAGCCAGCTGAGCGTGTCC
GACATCGATAAGGCCAGCTCTGAGGCCCTTTGGTGCGCCCTGACAAGAGAACCTGGCTTCGACCCC
AAAGAGGGACTGCCTGCCAATCCTGAGCGGCACATCAGAGTGAATGGCACCCATGTGTACGCCGGC
GACAAACATCGGCCTGTTTCCAGTGTCTGCCGGATCTATCGCTCTGAGAGGCGGATATGCCGAGCTG
GGCAGCTCTTTCCATCACGCCAGGGTGTACAAGATCACAAGCGGCAAGAAACCCGCCTTTGCCATG
CTGAGAGTGTATACCATCGACCTGCTGCCCTTACCGGAACCAGGACCTGTTCAGCGTGGAAGTGAAG
CCCCAGACCATGAGCATGAGACAGGCCGAGAAGAAGCTGAGGGACGCCCTGGCTACAGGCAACGCC
GAATATCTTGGATGGCTGGTGGTGGATGACGAGCTGGTGGTGGATACCAAGATCGCCACCGAC
CAAGTGAAGGCTGTGGAAGCCGAAGTGGGAACCATCAGACGTTGGCGCGTGGACGGCTTTTTCAGC
CCCTCTAAGCTGAGACTGCGGCCCTGCAGATGAGCAAAGAGGGCATCAAGAAAGAGAGCGCCCT
GAGCTGTCCAAGATCATTGACAGACCTGGCTGGCTGCCCGCCGTGAACAAGCTGTTTTCTGACGGC
AACGTGACCGCTCGTGGGAGAGATTCTCTGGGCAGAGTGGCCTGGAAAGCACAGCACATCTGCC
GTGACATGGAAGGTGCAGCCCAAGAAGAAGAGGAAGGTG

CdiCas9/NLS 단백질 서열 (서열 번호:30)

MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLISLVSHIHDSGLDPDKIKSAVTRLASSGIARRRRL
YRRKRRRLQQLDKFIQRQGWPIELEDYSDFLYPWKVRaelaASYIADEKERGEKLSVALRHIAH
RGWRNPYAKVSSLYLPDEPSDAFKAIREEIKRASGQVPETATVGQMVTLCELGLTKLRGEGGVL
ARLQQSDHAREIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKASDAF
QRYRIAALIGNLRVRVDGEKRLSVEEKNLVFDHLVNLAPKKEPEWVTIAEILGIDRGQLIGTATM
TDDGERAGARPPTHDTNRSIVNSRIAPLVDWWTASALEQHAMVKALSNAEVDDFDSPEGAKVQAF
FADLDDDVHAKLDSLHLPVGRAAYSEDTLVRLTRMLADGVLDLYTARLQEFIEPSWTPPAPRIGE
PVGNPVAVDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVTEKRAREMDGDMRRRAARNAKLFQE
MQEKLNVQGKPSRADLWRYQSVQRQNCQAYCGSPITFSNSEMDHIVPRAGQGSTNTRENLVAVCH
RCNQSKGNTFFAIWAKNTSIEGVSVKEAVERTRHWVTDGMRSTDFKKFTKAVVERFQRATMDEEI
DARSMESVANMANELRSRVAQHFAHSGTTVRVYRGSLTAEARRASGTSKLEFLDGVGKSRLDRRH
HAIDAAVIAFTSDYVAETLAVRSNLKQSQAHQEAPOWREFTGKDAEHRAAWRVWCOKMEKLSALL
TEDLRDDRVRVMSNVRLRLGNGSAHEETIGKLSKVKLGSQLSVSDIDKASSEALWCALTREPDFDP
KDGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAEFGSSFHARVYKITSGKKPAFAM
LRVYTIDLLPYRNQDLFSVELKPTMSMRQAEEKLRDALATGNAEYLGWLVDDELVDVDTSKIATD
QVKAVEAELGTIRRWVRVDGFFGDTRLRLRPLQMSKEGIKKEAPELSKIIDRPGWLPVANKLFSEG
NVTVVRDSLGRVRLESTAHLPVTWKVQPKKKRKV

Bsm Cas9 sgRNA (서열 번호: 31)

NNNNNNNNNNNNNNNNNNNNNGUCAUAGUCCCCUAAGAUUAUUGAAACAAUGAUCUUAGGGUUACU
AUGAUAAAGGCUUUCUACUUUAGGGGUAGAGAUGUCCCGCGCGUUGGGGAUCGCCUAAUUGCCCUU
AAAGGGCACUCCCCAUUUUAAUUUUUUU

Lrh Cas9 sgRNA (서열 번호:32)

NNNNNNNNNNNNNNNNNNNNNGUCUCAGGUAGAUGUCAGAUCAAUCAGAAAUGAUUGAUCUGACAUC
UACGAGUUGAGAUCAAACAAAGCUUCAGCUGAGUUUCAUUUCUGAGCCCAUGUUGGGCCAUACAU
AUGCCACCCGAGUGCAAUCGGGUGGCUUUUUUU

Pex Cas9 sgRNA (서열 번호:33)

NNNNNNNNNNNNNNNNNNNNNGUUUCAGUAGUUGUUAGAAGAAUGAAAAUUCUUUUAACAACGAAGU
CGCCUUCGGGCGAGCUGAAAUCAAUUUGAUUAAUUAUUAAGAUCCGGCUACUGAGGUCUUUGACCUU
AUCCGGAUUAAACGAAGAGCCUCCGAGGAGGCUUUUUUU

Mca Cas9 sgRNA (서열 번호:34)

NNNNNNNNNNNNNNNNNNNNNGUUUUAGUGUUGUACAAUUAUUUGGGUGAAAACCCAAUAUUGUACA
UCCUAAAUAAGGCGCUUAAUUGCUGCCGUAUUGCUGAAAGCGUAGCUUUCAGUUUUUUU

Mga Cas9 sgRNA (서열 번호:35)

NNNNNNNNNNNNNNNNNNNNNGUUUUAGCAGUACAAUACUUGUGUAAGCAAUAACGAAAAUUAUU
GCUUACACAAUUAUUGUCGUGCUAAAAUAAGGCGCUGUUAUUGCAGCUGCCGCAUCCGCCAGAGCA
UUUAUGCUCUGGCUUUUUUU

AgI Cas9 sgRNA (서열 번호:36)

NNNNNNNNNNNNNNNNNNNNNGUUUUGCCUUGAAUCCAAAGUAAGGCAUGGUAgaaaUAUUUAUCCU
GUGGAUUCAGACAAAAUUUGAAUGCAAACCGAUUCCCCGGCUGCAAGCCAGCCACACCGGUCUU
UCAAGCAUUUUUUU

Amu Cas9 sgRNA (서열 번호:37)

NNNNNNNNNNNNNNNNNNNNNGUUUUGCCUUGAAUCCAAAACGGAUUCAGACAAAAUUUGAAUGC
AAACCGAUUUUCCUGACUGCCAGCCAGUCACACCGGUAACAAAAGCAUUUUUUU

Oki Cas9 sgRNA (서열 번호:38)

NNNNNNNNNNNNNNNNNNNNNGCUUCAGAUUGUGUCAGAUCAAUGAgaaaUCAUUGAUCUGACACA
CAGCAUUGAAGUAAAGCAAGAUUAAUUUCAAGCUUAAUUUCUUCACAUUUUAUGUGCAGAAGGGC
UUAUGCCCACAAUACAUA AAAAGUCCGCAUUCACUUGCGGACUUUAUUUUUUU

Bbo Cas9 sgRNA (서열 번호: 39)

NNNNNNNNNNNNNNNNNNNNNGUUUCAAUUCAAUCUAAAAGCGAAAGCUAUACUUUUUUUUGAAUUU
GAAAUAGGCGUUCUUCGUUAGUUCAGUCGAUUGCUCUCCGUAUUGCUUAUGCAUGCCGGAG
UUUUUU

Ace Cas9 sgRNA (서열 번호: 40)

NNNNNNNNNNNNNNNNNNNNNGCUGGGGAGCCUGUCUGAAAAGACAGGCUACCUAGCAAGACCCCUU
CGUGGGGUGCGCAUUCUUCACCCCCUCGCAGCAGCGAGGGGGUUCGUUUUUUU

Ahe Cas9 sgRNA (서열 번호: 41)

NNNNNNNNNNNNNNNNNNNNNGUCAUAGUUCUCCUCACAAGCCUCCGAUGUGGAAACACAUCAAGGCUU
GCGAGGUUGCUAUGAUAAAGGCAACAGGCCGCAAAGCACUGACCCGCAUCCAAUGAAUGCGGGUCA
UCUACUUUUUUU

Wsu Cas9 sgRNA (서열 번호: 42)

NNNNNNNNNNNNNNNNNNNNNGUUUCACAGGCUAAGCGGAUUUGCGaaaGCAAUCCGUUCGAUGCC
UUGAAAUCAUCAAAGAUAAUAGACCCGCCACUGUAUUGUACAUGGCGGGACUUUUUUU

Nsa Cas9 sgRNA (서열 번호: 43)

NNNNNNNNNNNNNNNNNNNNNGUUUAAGACCCCUCAAAACCCGACCCUGUUACAAUGUUGUAACAG
GGUAGGGUUAUUUGAGGGGUCUUAUAUACAAGAACUGUUACAACAGUCCAUCUAGGGCCCAUCU
UCGGACGGGCCUCAGCCUUUUUUU

Rsy Cas9 sgRNA (서열 번호: 44)

NNNNNNNNNNNNNNNNNNNNNGUUGUAGCCAGAGCGCAAUCCCGAUCUGCUGAAAAGCAGAUCCGG
AAUUGCGCUUUGCUACUAACAAGCUGAAUCCGUUAGGAGUAAAUGCACCAAUGAGAGGGCCGGCU
UAUGCCGGCCCUUUGCUUUUUUU

Cdi Cas9 sgRNA (서열 번호: 45)

NNNNNNNNNNNNNNNNNNNNNACUGGGGUUCAGUUCUCAAACCCUGAUAGACUUCGAAAAGUCAC
UACUUAUUUUUAGAACUGAACUCAGUAAGCAUUGGCUCGUUCCAAUGUUGAUUGCUCGCC
GGUGCUCCUUAUUUAUAAAGGGCGCCGGCUUUCUUUUUUU

PexCas9-HN1HB1 융합 (서열 번호:117)

MPKRRVSSAEGAAKEEPPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVA
NQETKEDLPAENGETKTEESPASDEAGEKEAKSD TGSGMGKTHIIGVGLDLGGTYTGTFITSHPSD
 EAEHRDHSSAFTVVNSEKLSFSSKSRTAVRHRVRSYKGFDLRRRLLLLVAEYQLLQKKQTLAPEER
 ENLRIALSGYLKRRGYARTEAETDTSVLES LDPSVFSSAPSTNFFNDSEPLNIQWEAIANS PETT
 KALNKELSGQKEADFKKYIKTSFPEYSAKEILANYVEGRRAILDASKYIANLQSLGHKHKRSKYLS
 ILQDMKRDSRITRLSEAFGSTDNLWRIIGNISNLQERAVRWYFNDAKFEQGOEQQLDAVKLKNVLVR
 ALKYLRSDDKIEWSASQKQIIQSLEQSGDVLDVLAGLDPDRTIPPYEDQNNRRPPEDQTLTYLNP
 KALSS EYGEKWKSWANKFAGAYPLLTEDLTEILKNTDRKSRIKIRSDVLPDSDYRLAYILQRAFDRS
 IALDECSIRRTAEDFENG VVIKNEKLEDVLSGHQLEEFLEFANRYQETAKAKNGLWFPENALLERAD
 LHPPMKNKILNVIVGQALGVSPAEGTDFIEEIWN SKVKGRSTVRSICNAIENERKTYGPFYSEDYK
 FVK TALKEGKTEKELSKKFAAVIKVLKVMSEVVPFIGKELRLSDEAQSKFDNLYSLAQLYNLIETE
 RNF GSKVSLAAHLENAWRMTMTDGS AQCCRLPADCVRPFDGFIKKAIDRNSWEVAKRIAEEVKKSV
 DFTNGTVKIPVAIEANSFNFTASLTDLKYIQLKEQK LKKKLEDIQRNEENQEKRWLSKEERIRADS
 HGICAYTGRPLDDVGEIDHIIIPRSLTLKKSESIYNSEVNLIFVSAQGNQEKKNIIYLLSNLAKNYL
 AAVFGTSDLSQITNEIESTVLQKAAGRLGYFDLLSEKERACARHALFLNSDSEARRAVIDVLGSR
 RKASVNGTQAWFVRSIFSKVRQALAAWTQETGNELIFDAISVPAADSSEMRKRFAEYRPEFRKPKV
 QPVASHSIDAMCIYLAACSDPFKTKRMGSQLAIYEPINF DNLTGSCQVIQNTPRNFSDKTNIANS
 PIFKETIYAERFLDIIVSRGEIFIGYPSNM PFEEKPNRISIGGKDPFSILSVLGAYLDKAPSSEKE
 KLTIIYRVVKNKAFELFSKVAGSKFTAEDDKAAKILEALHFVTVKQDVAAATVSDLIKSKKELSKDSI
 ENLAKQKGLCKKVEYSSKEFKFKGSLIIPA AVEWGKVLWNVFKENTAEELKDENALRKALEAAWPS
 SFGTRNLHSAKRVFSLPVVATQSGAVRIRKTAFGDFVYQSQDTNNLYSSFPVKNGKLDWSSPII
 HPALQNRNLTAYGYRFVDHRSISMSEFREVYNKDDL MRIELAQGTSSRRYL RVEMPGEKFLAWFG
 ENSISLGSSPFKFSVSEVFDNKIYTENA EFTKFLPKPREDNKHNGTIFFELVGPRVIFNYIVGGAAS
 SLKEIFSEAGKERSPKKKRKVLEGGGGSGKGD PKKPRGKMSSYAFFVQTCREEHKKKHPDASVNF
SEFSSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGE

PexCas9-HN1H1G 융합 (서열 번호:118)

MPKRRVSSAEGAAKEEPPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVA
NQETKEDLPAENGETKTEESPASDEAGEKEAKSD TGSGMGKTHIIGVGLDLGGTYTGTFITSHPSD
 EAEHRDHSSAFTVVNSEKLSFSSKSRTAVRHRVRSYKGFDLRRRLLLLVAEYQLLQKKQTLAPEER
 ENLRIALSGYLKRRGYARTEAETDTSVLES LDPSVFSSAPSTNFFNDSEPLNIQWEAIANS PETT
 KALNKELSGQKEADFKKYIKTSFPEYSAKEILANYVEGRRAILDASKYIANLQSLGHKHKRSKYLS
 ILQDMKRDSRITRLSEAFGSTDNLWRIIGNISNLQERAVRWYFNDAKFEQGOEQQLDAVKLKNVLVR
 ALKYLRSDDKIEWSASQKQIIQSLEQSGDVLDVLAGLDPDRTIPPYEDQNNRRPPEDQTLTYLNP
 KALSS EYGEKWKSWANKFAGAYPLLTEDLTEILKNTDRKSRIKIRSDVLPDSDYRLAYILQRAFDRS
 IALDECSIRRTAEDFENG VVIKNEKLEDVLSGHQLEEFLEFANRYQETAKAKNGLWFPENALLERAD
 LHPPMKNKILNVIVGQALGVSPAEGTDFIEEIWN SKVKGRSTVRSICNAIENERKTYGPFYSEDYK
 FVK TALKEGKTEKELSKKFAAVIKVLKVMSEVVPFIGKELRLSDEAQSKFDNLYSLAQLYNLIETE
 RNF GSKVSLAAHLENAWRMTMTDGS AQCCRLPADCVRPFDGFIKKAIDRNSWEVAKRIAEEVKKSV
 DFTNGTVKIPVAIEANSFNFTASLTDLKYIQLKEQK LKKKLEDIQRNEENQEKRWLSKEERIRADS

[0148]

HGICAYTGRPLDDVGEIDHIIIPRSLTLKKSESIYNSEVNLI FVSAQGNQEKKNIIYLLSNLAKNYL
 AAVFGTSDLSQITNEIESTVLQKAAGRLGYFDLLSEKERACARHALFLNSDSEARRAVIDVLGSR
 RKASVNGTQAWFVRSIFSKVRQALAAWTQETGNELIFDAISVPAADSSEMRKRFAEYRPEFRKPKV
 QPVASHSIDAMCIYLAACSDPFKTKRMGSQLAIYEPINFNDLFTGSCQVIQNTPRNFSDKTNANS
 PIFKETIYAERFLDIIIVSRGEIFIGYPSNMPFEEKPNRISIGGKDPFSILSVLGAYLDKAPSSEKE
 KLTIIYRVVKNKAFELFSKVAGSKFTAEDDKAAKILEALHFVTVKQDVAAATVSDLIKSKKELSKDSI
 ENLAKQKGLKKVEYSSKEFKFKGSLIIPAAVEWGKVLWNVFKENTAEELKDENALRKALEAAWPS
 SFGTRNLHSAKRVFSLPVVATQSGAVRIRRKTAFGDFVYQSQDTNNLYSSFPVKNGKLDWSSPII
 HPALQNRNLTAIYGYRFVDHRSISMSEFREVYNKDDLMRIELAQGTSSRRYLKVEMPGKFLAWFG
 ENSISLGSSFKFSVSEVFDNKIYTENAEFTKFLPKPREDNKHNCTIFFELVGPVIFNYIVGGAAS
 SLKEIFSEAGKERSPKKKRKVLEGGGGSSTDHPKYSMDIVAAIQAEKNRAGSSRQSIQKYIKSHYK
 VGENADSQIKLSIKRLVTTGVVLKQTKGVGASGSFRLAKSDEP

BsmCas9-HN1HB1 융합 (서열 번호:119)

MPKRVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVA
 NQETKEDLPAENGETKTEESPASDEAGEKEAKSDTGSGMNYKMGLDIGIASVGWAVINLCLKRIED
 LGVRIFDKAEHPQNGESLALPRRIARSARRRLRRRKHRLERIRRLVSENVLTKEEMNLLFKQKKQ
 IDVWQLRVDALEKRLNDELARVLLHLAKRRGFKSNRKSEKSESEFLKNIEENQSILAQYRSV
 GEMIVKDSKFAYHKNKLDSSYNMIARDDLEREIKLIFEKQREFNPNVCTERLEEKYLNIIWSSQRP
 FASKEDIEKKVGFCTFEPKEKRAPKATYTFQSFIVWEHINKLRLVSPDETRALTEIERNLLYKQAF
 SKNKMTYYDIRKLLNLSDDIHFGLLYDPKSSLKQIENIRFLELDSYHKIRKCIENVYKGDGIRMF
 NETDIDTFGYALTIFKDDDEDIVAYLQNEYITKNGKRVSNLANKVYDKSLIDELNLSFSKFAHLSM
 KAIRNIPYMEQGEIYSKACELAGYNFTGPKKKEKALLLPVIPNIANPVVMRALTSRKVVNAIIK
 KYGSPVSIHIELARDLSHSFDERKKIQKDQ TENRKKNETAIKQLIEYELTKNPTGLDIVKFKLWSE
 QQGRCMYSLKPIELERLLEPGYVEVDHILPYSRSLDDSYANKVLVLTENREKGNHTPVEYLGLGS
 ERWKKFEKFLANKQFSSKKKKQNLRLRYEETEEKEFKERNLNDTRYISKFFANFIKEHLKFADGD
 GGQKVYTINGKITAHLRSRWDFNKNREESDLHHAVDIVACATQGMIKKITEFYKAREONKESAK
 KKEPIFPQWPWFADDELKARLSKFPQESIEAFALGNDRKKLESRLPVFVSRMPKRSVTGAHQET
 LRRCVGIDEQSGKIQTAVKTKLSDIKLDKDGHPMYQKESDPRTYEATIRQRLLEHNNDPKKAFQEP
 LYKPKKNGEFGPVIRTVKIIDTKNKVVHLDGSKTVAYNSNIVRTDVFEKDGKYYCVPVYTM DIMKG
 TLPNKAIEANKPYSEWKEMTEEYTFQFSLFPNDLVRIVLPREKTIKTSTNEEII IKDIFAYYKTID
 SATGGLELISHDRNFSLRGVGSKTLKRFKYQVDVLGNIHKVKGEKRVGLAAPTQKKGKTVDSLQ
 SVSDPKKKRKVLEGGGGS~~GKGD~~PKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKCSERW
 KTMSAKEKGFEDMAKADKARYEREMKTYIPPKGE

BsmCas9-HN1H1G 융합 (서열 번호:120)

MPKRVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVA
 NQETKEDLPAENGETKTEESPASDEAGEKEAKSDTGSGMNYKMGLDIGIASVGWAVINLCLKRIED
 LGVRIFDKAEHPQNGESLALPRRIARSARRRLRRRKHRLERIRRLVSENVLTKEEMNLLFKQKKQ
 IDVWQLRVDALEKRLNDELARVLLHLAKRRGFKSNRKSEKSESEFLKNIEENQSILAQYRSV
 GEMIVKDSKFAYHKNKLDSSYNMIARDDLEREIKLIFEKQREFNPNVCTERLEEKYLNIIWSSQRP
 FASKEDIEKKVGFCTFEPKEKRAPKATYTFQSFIVWEHINKLRLVSPDETRALTEIERNLLYKQAF
 SKNKMTYYDIRKLLNLSDDIHFGLLYDPKSSLKQIENIRFLELDSYHKIRKCIENVYKGDGIRMF

NETDIDTFGYALTIFKDDDEDIVAYLQNEYITKNGRVSNLANKVYDKSLIDELLNLSFSKFAHLSM
KAIRNIPYMEQGEIYSKACELAGYNFTGPKKKEKALLLPVIPNIANPVVMRALTSRQVNNAIK
KYGSPVSIHIELARDLSHSFDERKKIQKDQ TENRKKNETAIKQLIEYELTKNPTGLDIVKFKLWSE
QOGRCMYSLKPIELERLLEPGYVEVDHILPYSRSLDDSYANKVLVLTKENREKGNHTPVEYLGGLGS
ERWKKFEKFLVANKQFSKKKKQNLRLRYEETEEKEFKERNLNDTRYISKFFANFIKEHLKFADGD
GGQKVYTINGKITAHLRSRWDNFKNREESDLHHAVDVAVACATQGMIKKITEFYKAREONKESAK
KKEPIFPQWPHFADELKARLSKFPQESIEAFALGNDRKKLESRLPVFVSRMPKRSVTGAHQET
LRRCVGIDEQSGKIQTAVKTKLSDIKLDKDGHPMYQKESDPRTYBAIRQRLLEHNNDPKKAFQEP
LYKPKKNGEPPGVIRTVKIIDTKNKVVHLDGSKTVAYNSNIVRTDVFEDGKYCYCPVYTM DIMKG
TLPNKAIEANKPYSEWKEMTEEYTFQFSLFPNDLVRIVLPREKTIKTSTNEEIIKDI FAYYKTID
SATGGLELISHDRNFSLRGVGSKTLKRFKYQVDVLGNIHKVKGEKRVGLAAPTQKKKGKTVDSLQ
SVSDPKKKRKVLEGGGGSSTDHPKYSMDMIVAAIQAEKNRAGSSRQSIQKYIKSHYKVG ENADSQIK
LSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP

LrhCas9-HN1HB1 융합 (서열 번호:121)

MPKRRKVSSAEGAAKEEPPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVA
NQETKEDLPAENGETKTEESPASDEAGEKEAKSDTGSGMTKLNQPYGIGLDIGSNSIGFAVVDANS
HLLRLKGETAIGARLFREGQSAADRRGSRTRRLSRTRWRLSFLRDFFAPHITKIDPDDFFLRQKY
SEISPKDKDRFKYEKRLFNDRDTEAFYEDYPSMYHLRLHLMTHTHKADPREIFLAIHHILKSRGHF
LTPGAAKDFNTDKVDLEDIFPALTEAYAQVYPDELTDFDLAKADDFKAKLLDEQATPSDTQKALVN
LLSSDGEKEIVKRRKQVLTEFAKAITGLKTKFNALGTEVDEADASNWQFSMGQLDDKWSNIETS
MTDQGTETFEQIQELYRARLLNGIVPAGMSLSQAKVADYGGHKEDELEFKTYLKKLNDHELAKTIR
GLYDRYINGDDAKPFLREDFVKALTKEVTAHPNEVSEQLLNRMGQANFMLKQRTKANGAIP IQLQQ
RELDQIIANQSKYYDWLAAPNPVEAHRWKMPYQLDELLNFHIPYYVGPLITPKQQAESGENVFAMW
VRKDPGSGNITPYNFDEKVDREASANTFIQRMKTTDTYLIGEDVLPKQSLLYQKYEVNLNLRNVRIN
NECLGTDQKQRLIREVFERHSSVTIKQVADNLVAHGDFARRFEIRGLADEKRFSSLSSTYHQLKEI
LHEAIDDPKLLDIENIITWSTVFEDHTIFETKLAIEIWLDPKKINELSGIRYRGWGQFSRKLDDG
LKLGNHGTVIQELMLSNNHLMQILADETLKETMTLQDKLKTDDIEDVINDAYTSPSNKKALRQV
LRVVEDIKHAANGQDFSWLFIETADGTGTAGKRTQSRQKQIQTVYANAAQELIDSAVRGELEDKIA
DKASFTRDLVLYFMQGGRIYTGAPLNIDQLSHYDIDHILPQSLIKDDSLDNRVLVNATINREKNN
VFASTLFAGKMKATWRKWHEAGLISGRKLRNLMRLPDEIDKFAKGFVARQLVETRQIIKLTEQIAA
AQYPNTKIIAVKAGLSHQLREELDFPKNRDVNHYHHAFLAARIGTYLLKRYPKLAPFFTYGEF
AKVDVKKFRENFFIGALTHAKNNIIAKDTGEIVWDKERDIRELDRIYNFKRMLITHEVYFETADLF
KQTIYAAKDSKERGGSKQLIPKKQGYPTQVYGGYTQESGSYNALVRVAEADTTAYQVIKISAQNAS
KIASANLKSREKQKQLLNEIVVKQLAKRRKNWKPSANSFKIIVIPRFGMGTLFQNAKYGLFMVNSDT
YYRNYQELWLSRENQKLLKLLFSIKYEKTQMNHDALQVYKAIIDQVEKFPKLYDINOFRAKLSDAI
ERFEKLPINTDGNKIGKTETLRQILIGLQANGTRSNVKNLGIKTDLGLLQVGSIGKLDKDTQIVYQ
SPSGLPKRRRIPLADLPKKKRVLEGGGGS~~SGKGD~~PKKPRGKMSSYAFFVQTCREEHKKKH~~PDASVNF~~
SEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGE

LrhCas9-HN1H1G 융합 (서열 번호:122)

MPKRRKVSSAEGAAKEEPPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAKGKQAEVA
NQETKEDLPAENGETKTEESPASDEAGEKEAKSDTGSGMTKLNQPYGIGLDIGSNSIGFAVVDANS

HLLRLKGETAIGARLFREGQSAADRRGSRTTRRRLSRTRWRLSFLRDFPAPHITKIDPDDFFLRQKY
SEISPKDKDRFKYKRLFNDRDTAEFYEDYPSMYHLRLHLMTHTHKADPREIFLAIHHILKSRGHF
LTPGAAKDFNTDKVDLEDIFPALTEAYAQVYPDELETFDLAKADDFKAKLLDEQATPSDTPQKALVN
LLLSSDGEKEIVKRRKQVLTEFAKAITGLKTKFNLALGTEVDEADASNWQFSGMLDDKWSNIETS
MTDQGTETFEQIQELYRARLLNGIVFAGMSLSQAKVADYGQHKEDLELFKTYLKKLNDHELAKTIR
GLYDRYINGDDAKPFLREDFVKALTKEVTAHPNEVSEQLLNRMGQANFMLKQRTKANGAIPILQLOQ
RELDQIIANQSKYYDWLAAPNPVEAHRWKMPYQLDELLENFHIPYYVGPLITPKQQAESGENVFAWM
VRKDPQSGNITPYNFDEKVDREASANTFIQRMKTTDTYLIGEDVLPKQSLLYQKYEVNLNNVRIN
NECLGTDQKQRLIREVFERHSSVTIKQVADNLVAHGDFARRPEIRGLADEKRFLLSSSLSTYHQLKEI
LHEAIDDPKLLDIENIITWSTVFEDHTIFETKLAEIEWLDPKKINELSGIRYRGWGQFSRKLDDG
LKLNGHGTVIQELMLSNHNLMQILADETLKETMTELNQDKLKTDDIEDVINDAYTSPSNKKALRQV
LRVVEDIKHAANGQDPSWLFETADGTGTAGKRTQSRQKQIQTVYANAAQELIDSAVRGELEDKIA
DKASFTDRLVLYFMQGGGRDIYTGAPLNIDQLSHYDIDHILPQSLIKDDSLDNRVLVNATINREKNN
VFASTLPAGKMKATWRKWHEAGLISGRKLRNLMRPEIDKFAGGFVARQLVETRQIIKLTEQIAA
AQYPNTKIIAVKAGLSHQLREELDFPKNRDVENHYHHAFFDAFLAARIGTYLLKRYPKLAPFFTYGEF
AKVDVKKFREFNFIGALTHAKNNIIAKDTGEIVWDERDIRELBRIYNFKRMLITHEVYFETADLF
KOTIYAARDKSKERGGSKOLIPKKQGYPTQVYGGYTQESGSYNALVRVAEADTTAYQVIKISAQNAS
KIASANLSREKKGKQLLNEIVVKQLAKRRKNWKPANSFKIIVIPRFGMGTLFQNAKYGLFMVNSDT
YYRNYQELWSRENQKLLKLLFSIKYEKTQMNHDALQVYKAIIDQVEKFFKLYDINQFRAKLSDAI
ERFEKLPINTDGNKIGKTETLRQILIGLQANGTRSNVKNLGIKTDLGLLQVGSIGIKLDKDTQIVYQ
SPSGLFKRRIPLADLPKKKRKVLGGGGSSSTDPKYSMDMIVAAIQAENRAGSSRSQSIQKYIKSHY
KVGENADSQIKLSIKRLVTTGVLEKQTKGVGASGSFRLAKSDEP

McaCas9-HN1HB1 융합 (서열 번호:123)

MPKRRKVSSAEGAAKEEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAAGKQAEVA
NQETKEDLPAENGETKTEESPASDEAGEKEAKSDTGSGMEKKRKVTLGFDLGIASVGWAIVDSETN
QVYKLGSRFLDAPDTNLERRTORGTRRLLRRRKYRNQKFYNLVKRTVEVFLSSREAIENRFRELSI
KYPNIIELKTKALSQEVCPDEIAWILHDYLNKRGYFYDEKETKEDFDQQTVESMPQSYKLNEFYKKY
GYFKGALSQPTESSEMKDNKDLKEAFFFDENKEWLKEINYFFNVQKNILSETFIEEFKKIFSFTRD
ISKGPGSDNMPSPYGFIFGEFGDNGQGGRYEHIWDKNIGKCSIFTNEQRAPKYLPSALIFNFLNELA
NIRLYSTDKNKNIQPLWLKSSIDKLNILLNLFNLPISEKKKKLTSTNINDIVKKESIKSIMLSVEDI
DMIKDEWAGKEPNVYGVGLSGLNIEESAKENKFKFQDLKILNVLINLLDNVGIKFEFKDRSDIKN
LELLDNLYLFLIYQKESNNKDSSIDLFIANKNSLNIEENLKLKLEFLGAGNEFENHNSKTHSLSK
KAIDAILPKLLDNNEGWNLEAIKNYDEEIKSQIEDNSSLMKQDKKYLNDNFKDAILPPNVKVT
QQAILIFNKIIQKFSKDFEIDKVVIELAREMTQDQENDALKGIAKAQSKSKSLVEERLEANNIDKS
VFNDKYEKLIYKIFLWISQDFKDPYTGAKISANEIVDNKVEIDHIIIPYSLCFDDSSANKVLVHKQS
NQEKSNLPEYIYKQGHSGWNWDEFTKYVVRVFNVDLSLSKKERLKKSENLLTTSYDGYEKLGF
LARNLNDTRYATILFRDQLNNYAEEHLIDNKKMFKVIAMNGAVTSFIRKNMSYDNKLRKDRSDFS
HHAYDAAIALFSNKTLYNLIDPSLNGIISKRSYGVVIEDRYTGEIKELKKEDWTSIKNNVQA
RKIAKEIEEYLIDLDDEVFSSRKTRKTRNQLYNETIYGIAAKTDEDGITNYYKKEKFSILDDKDI
YLRLLREREREKPVINQSNPEVIDQIIIEIESYGKENNIPSRDEAINIKYTKNKINYNLYLKQYMRSL
TKSLDQFSEGFNQMIANKTFVLNPTKNTTRKIKFLRLVNDVKINDIRKNQVINKFNKGNNPKA
FYENINSLGAIVFKSSANNFKTSLINTQIAIFGDKNWDIEDFKTYNMEKIEKYKEIYIGIDKTYNFH
SFIFPGTILLDKQNKEFYIISIQTVNDQIELKFLNKIEFKNDNTSGANKPFRRLRFGIKSIMNN

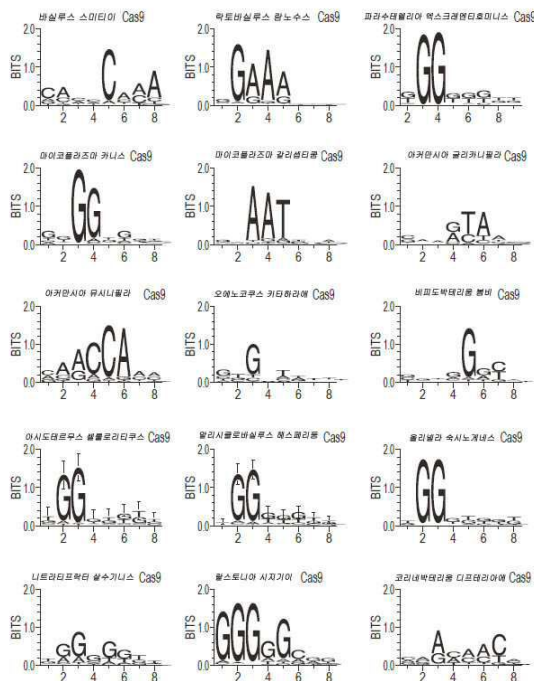
YEQVDISPFGINKKIFEPKKKRKVLGGGGSGKGDPPKPRGKMSSYAFFVQTCREEHKKKHPDASV
NFSEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGE

McaCas9-HN1H1G 융합 (서열 번호:124)

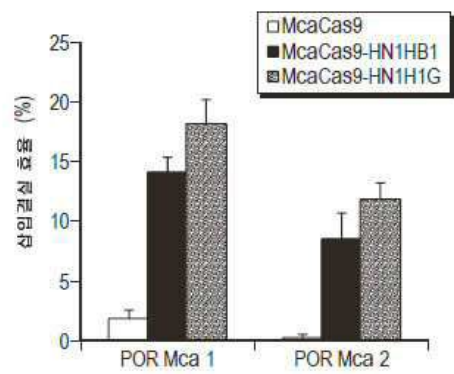
**MPKRKVSSAEGAAKEPKRRSARLSAKPPAKVEAKPKKAAAKDKSSDKKVQTKGKRGAAGKQAEVA
NQETKEDLPAENGETKTEESPASDEAGEKEAKSDTGSGMEKKRKVTLGFDLGIASVGWAIVDSETN
QVYKLGSRFLDAPDTNLERRTQGRTRLLRRRKYRNQKFYNLVKRTFVGLSSREAIENRFRELSI
KYPNIIELKTKALSQEVCPDEIAWILHDYLNKRGYFYDEKETKEDFDQQTVESMPYSYKLNFEYKYY
GYFKGALSQPTSEMKDNKDLKEAFFFDNKEWLKEINYFFNVQKNILSETFIEEFKKIFSFTRD
ISKGPGSDNMPSPYGI FGEFGDNGQGGRYEHIWDKNIGKCSIFTNEQRA PKYLPSALIFNFLNELA
NIRLYSTDKNKNIQPLWLKSSIDKLNILLNLFNLPISEKKKLTSTNINDIVKKEIKSIMLSVEDI
DMIKDEWAGKEPNVYGVGLSGLNIEESAKENKFKFQDLKILNVLINLLDNVGIKFEFKDRSDIKN
LELLDNLYLFLIYQKESNNKDSSIDLFIKKNKSLNIENLKLKLEFLLGAGNEFENHNSKTHSLSK
KAIDAILPKLLDNNEGWNLEAIKNYDEEIKSQIEDNSSLMKQDKKYLNDNFKDAILPPNVKVTF
QQAILIFNKIIQKFSKDFEIDKVVIELAREMTQDQENDALGIAKAQKSKKSLVEERLEANNIDKS
VFNDKYEKLIYKIFLWISQDFKDPYTGA KISANEIVDNKVEIDHII PYSLCFDDSSANKVLVHKQS
NQEKSNLSPY EYIKQGHSGWNWDEFTKYVKRVFVNNVDSILSKKERLKKSENLLTTSYDGYEKLGF
LARNLNDTRYATILFRDQLNNYAEHHLIDNKKMFKVIAMNGAVTSFIRKNMSYDNKRLKDRSDFS
HHAYDAAI IALFSNKT KTLYNLIDPSLNGI I SKRSEGYWVIEDRYTGEIKELKKEDWTSIKNNVQA
RKIAKEIEEYLIDLDDEVFFSRKTKRKTNRQLYNETIYGIAAKTDEDGITNYYKKEKFSILDDKDI
YLRLLREREK FVINQSNPEVIDQIIEIIESYK ENNIPSRDEAINIKYTKNKINYNL YLKQYMRSL
TKSLDQFSEGFINQMIANKTFVLYNP TKNTTRKIKFLRLVNDVKINDIRKNQVINKFNKGKNNPKA
FYENINSLGAIVFKSSANNFKTLSINTQIAIFGDKNWDIEDFKTYNMEKIEKYKEIYGIDKTYNFH
SFIFPGTILLDKQNK EYFISSIQTVNDQIELKFLNKIEFKNDNTSGANKPPRRLRFGIKSIMNN
YEQVDISPFGINKKIFEPKKKRKVLGGGGSSSTDHPKYS DMIVAAIQAEKNRAGSSRQSIQKYIKS
HYKVGENADSQIKLSIKRLVTTGV LKQTKGVGASGSFRLAKSDEP**

도면

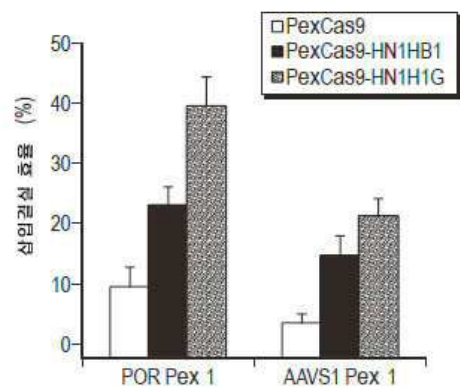
도면1



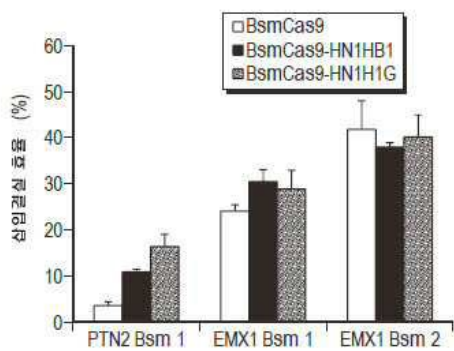
도면2a



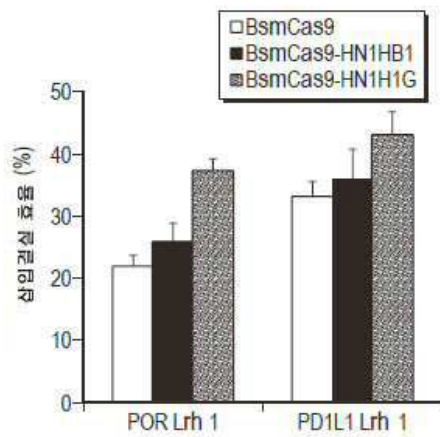
도면2b



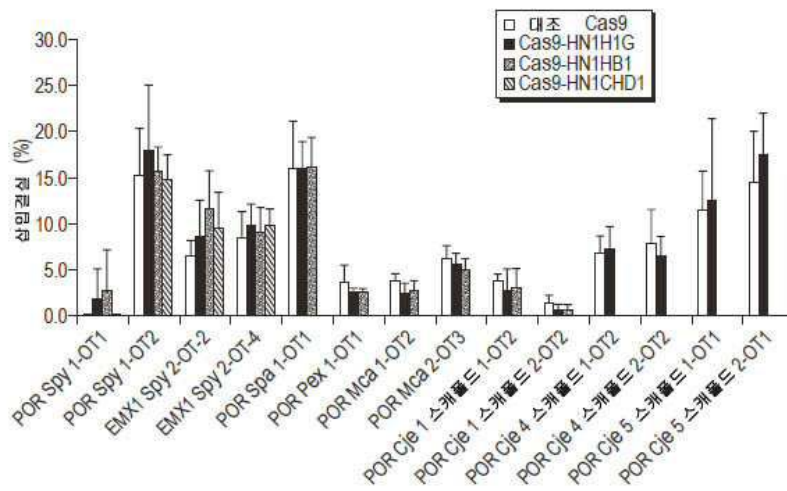
도면2c



도면2d



도면3



서열 목록

SEQUENCE LISTING

<110> SEEBECK, TIMOTHY

CHEN, FUQIANG

DAVIS, GREGORY

<120> ENGINEERED CAS9 SYSTEMS FOR EUKARYOTIC GENOME MODIFICATION

<130> P18/023PCT

<140><141><150> 62/720,525

<151> 2018-08-21

<150> 62/631,304

<151> 2018-02-15

<160> 134

<170> PatentIn version 3.5

<210> 1

<211> 3285

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 1

atgaactaca agatgggcct cgacatcgga atcgccctctg ttggatgggc cgtgatcaac	60
ctggacctga agagaatcga ggacctcggc gtgcggatct tcgacaaggc tgagcatcct	120
cagaacggcg agtctctggc cctgcctaga agaattgccga gaagcgccag acggcggctg	180
cggagaagaa agcacagact ggaacggatc agacggctgc tgggtgccga gaacgtgctg	240
accaaagaag agatgaacct gctgttcaag cagaaaaagc agatcgacgt gtggcagctg	300
agagtggacg ccctggaaag aaagctgaac aacgacgagc tggccagagt gctgctgcac	360
ctggccaaga gaagaggctt caagagcaac agaaagagcg agcggaacag caaagagagc	420
agcgagtcc tgaagaacat cgaagagaac cagagcattc tggcccagta cagatccgtg	480
ggcgagatga tcgtgaagga cagcaagttc gcctaccaca agcggaacaa gctggacagc	540
tacagcaaca tgatcgccag ggacgatctg gaaagagaga tcaagctgat cttcgagaag	600
cagcgcgagt tcaacaaccc cgtgtgcacc gagagactgg aagagaagta cctgaacatc	660
tggtcagcc agcggccttt cgcctccaaa gaggacatcg agaaaaaagt gggcttctgc	720
accttcgagc ccaaagagaa aagagccctt aaggccacct acaccttcca gagcttcac	780
gtgtgggagc acatcaacaa gctgcggctg gtgtctcccg acgagacaag agccctgacc	840
gagatcgagc ggaatctgct gtataagcag gccttcagca agaacaagat gacctactac	900
gacatccgga agctgctgaa cctgagcgac gacatccact tcaagggcct gctgtacgac	960
cccaagagca gcctgaagca gattgagaac atccggtttc tggaactgga ctcttaccac	1020
aagatccgga agtgcacga gaatgtgtac ggcaaggacg gcatccgcat gttcaacgag	1080
acagacatcg acaccttcgg ctacgccctg accatcttca aggacgacga ggatatcgtg	1140
gcctacctgc agaacgagta catcaccaag aacggcaagc ggggtgccaa tctggccaac	1200
aaggtgtacg acaagtcctt gatcgacgaa ctgctgaatc tgtccttctc caaatcggc	1260
cacctgagca tgaaggccat ccggaacatc ctgccttaca tggaacaggg cgaaatctac	1320

agcaaggcct gcgaactggc cggctacaac ttcacaggcc ccaagaagaa agagaaggcc	1380
ctgctgctgc ctgtgatccc caatatcgcc aatcctgtgg tcatgcgggc cctgacacag	1440
agcagaaagg tggtaacgc catcatcaag aaatacggat cccccgtgtc catccacatc	1500
gagctggcta gggatctgag ccacagcttc gacgagcggg agaagatcca gaaggaccag	1560
accgagaacc gcaagaagaa cgaaaccgcc atcaagcagc tgatcgagta cgagctgact	1620
aagaacccca cggcctgga catcgtgaag ttcaaacttt ggagcggagca gcaaggcaga	1680
tgcattgact ccctgaagcc tattgagctg gaaagactgc tggaaaccgg ctacgtggaa	1740
gtggaccaca ttctgccta cagcagaagc ctggacgaca gctacgcaa caaagtgtgt	1800
gtcctgacaa aagagaaccg cgaaaagggc aatcacaccc ctgtggaata tctcggcctg	1860
ggctctgagc ggtggaagaa attcgagaag ttctgtgtgg ctaacaagca gttctctaag	1920
aagaagaagc agaacctgtc cggctgaga tacgaggaaa ccgaggaaaa agagttcaaa	1980
gagcggaaac tgaacgacac ccggtacatc tccaagtctc tcgccaactt catcaaagag	2040
catctgaagt tgcgcgacgg cgacggcggc cagaaagtgt acacaatcaa cggcaagatc	2100
accgctcacc tgagaagcag atgggacttc aacaagaacc gggaagagag cgacctgcac	2160
cacgtgtgtg atgtgtgat tgtggcctgt gccacacagg gcatgatcaa gaagattacc	2220
gagttctaca aggcccgca gcagaacaaa ggtccgcca agaaaaaga acccatcttt	2280
ccccagctt ggctcactt cgccgatgag ctgaaggctc ggctgagcaa gttccctcaa	2340
gagtcacatc aggccttcgc tctgggcaac tacgacagaa agaagctgga atccctcggg	2400
cctgtgttcg tgtccagaat gccaagaga tccgtgacag gcgtgcca ccaagagaca	2460
ctgagaagat gcgtgggcat cgacgagcag tctggcaaga ttcagaccgc cgtgaaaaca	2520
aagctgagcg acatcaagct ggataaggac ggacacttcc ccatgtacca gaaagagtct	2580
gacccagaa cctacgagc catcagacag aggtgtctcg aacacaaca cgacctaa	2640
aaggcctttc aagagccact gtacaagccc aaaaagaatg gcgagcccg accagtgatc	2700
cggaccgtga agatcatga cacaagaac aaggtggtgc acctggacgg cagcaagaca	2760
gtggcctaca actccaacat cgtgcggacc gacgtgttcg agaaggatgg caagtactac	2820
tgcgtgcccg tgiacactat ggatatcatg aagggcaccc tgcctaaca ggccatcgaa	2880
gccaacaagc cctactccga gtggaaagag atgaccgaag agtacagtt ccagttcagt	2940
ctgttccca acgacctgt gcgcatctg ctgccaagag agaaaacat caagaccagc	3000
accaacgagg aaatcatcat taaggacatc ttgcctact acaagacat cgacagcgcc	3060
acaggcggcc tggaactgat ctcccacgat cggaacttca gcctgagagg cgtgggctct	3120
aagacactga agcgctttga gaagtatcag gtggacgtgc tgggcaacat ccacaaagt	3180

aagggcgaga agagagtcgg cctggccgct cctaccaacc agaaaaaggg aaagaccgtg 3240

gacagcctgc agagcgtgtc cgatcccaag aagaagagga aggtg 3285

<210> 2

<211> 1095

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

polypeptide"

<400> 2

Met Asn Tyr Lys Met Gly Leu Asp Ile Gly Ile Ala Ser Val Gly Trp

1 5 10 15

Ala Val Ile Asn Leu Asp Leu Lys Arg Ile Glu Asp Leu Gly Val Arg

20 25 30

Ile Phe Asp Lys Ala Glu His Pro Gln Asn Gly Glu Ser Leu Ala Leu

35 40 45

Pro Arg Arg Ile Ala Arg Ser Ala Arg Arg Arg Leu Arg Arg Arg Lys

50 55 60

His Arg Leu Glu Arg Ile Arg Arg Leu Leu Val Ser Glu Asn Val Leu

65 70 75 80

Thr Lys Glu Glu Met Asn Leu Leu Phe Lys Gln Lys Lys Gln Ile Asp

85 90 95

Val Trp Gln Leu Arg Val Asp Ala Leu Glu Arg Lys Leu Asn Asn Asp

100 105 110

Glu Leu Ala Arg Val Leu Leu His Leu Ala Lys Arg Arg Gly Phe Lys

115 120 125

Ser Asn Arg Lys Ser Glu Arg Asn Ser Lys Glu Ser Ser Glu Phe Leu

130 135 140

Lys Asn Ile Glu Glu Asn Gln Ser Ile Leu Ala Gln Tyr Arg Ser Val

145 150 155 160

Gly Glu Met Ile Val Lys Asp Ser Lys Phe Ala Tyr His Lys Arg Asn

165 170 175

Lys Leu Asp Ser Tyr Ser Asn Met Ile Ala Arg Asp Asp Leu Glu Arg
 180 185 190

 Glu Ile Lys Leu Ile Phe Glu Lys Gln Arg Glu Phe Asn Asn Pro Val
 195 200 205
 Cys Thr Glu Arg Leu Glu Glu Lys Tyr Leu Asn Ile Trp Ser Ser Gln
 210 215 220
 Arg Pro Phe Ala Ser Lys Glu Asp Ile Glu Lys Lys Val Gly Phe Cys
 225 230 235 240
 Thr Phe Glu Pro Lys Glu Lys Arg Ala Pro Lys Ala Thr Tyr Thr Phe
 245 250 255

 Gln Ser Phe Ile Val Trp Glu His Ile Asn Lys Leu Arg Leu Val Ser
 260 265 270
 Pro Asp Glu Thr Arg Ala Leu Thr Glu Ile Glu Arg Asn Leu Leu Tyr
 275 280 285
 Lys Gln Ala Phe Ser Lys Asn Lys Met Thr Tyr Tyr Asp Ile Arg Lys
 290 295 300
 Leu Leu Asn Leu Ser Asp Asp Ile His Phe Lys Gly Leu Leu Tyr Asp
 305 310 315 320

 Pro Lys Ser Ser Leu Lys Gln Ile Glu Asn Ile Arg Phe Leu Glu Leu
 325 330 335
 Asp Ser Tyr His Lys Ile Arg Lys Cys Ile Glu Asn Val Tyr Gly Lys
 340 345 350
 Asp Gly Ile Arg Met Phe Asn Glu Thr Asp Ile Asp Thr Phe Gly Tyr
 355 360 365
 Ala Leu Thr Ile Phe Lys Asp Asp Glu Asp Ile Val Ala Tyr Leu Gln
 370 375 380

 Asn Glu Tyr Ile Thr Lys Asn Gly Lys Arg Val Ser Asn Leu Ala Asn
 385 390 395 400
 Lys Val Tyr Asp Lys Ser Leu Ile Asp Glu Leu Leu Asn Leu Ser Phe
 405 410 415
 Ser Lys Phe Ala His Leu Ser Met Lys Ala Ile Arg Asn Ile Leu Pro

420 425 430
 Tyr Met Glu Gln Gly Glu Ile Tyr Ser Lys Ala Cys Glu Leu Ala Gly
 435 440 445

 Tyr Asn Phe Thr Gly Pro Lys Lys Lys Glu Lys Ala Leu Leu Leu Pro
 450 455 460
 Val Ile Pro Asn Ile Ala Asn Pro Val Val Met Arg Ala Leu Thr Gln
 465 470 475 480
 Ser Arg Lys Val Val Asn Ala Ile Ile Lys Lys Tyr Gly Ser Pro Val
 485 490 495
 Ser Ile His Ile Glu Leu Ala Arg Asp Leu Ser His Ser Phe Asp Glu
 500 505 510

 Arg Lys Lys Ile Gln Lys Asp Gln Thr Glu Asn Arg Lys Lys Asn Glu
 515 520 525
 Thr Ala Ile Lys Gln Leu Ile Glu Tyr Glu Leu Thr Lys Asn Pro Thr
 530 535 540
 Gly Leu Asp Ile Val Lys Phe Lys Leu Trp Ser Glu Gln Gln Gly Arg
 545 550 555 560
 Cys Met Tyr Ser Leu Lys Pro Ile Glu Leu Glu Arg Leu Leu Glu Pro
 565 570 575

 Gly Tyr Val Glu Val Asp His Ile Leu Pro Tyr Ser Arg Ser Leu Asp
 580 585 590
 Asp Ser Tyr Ala Asn Lys Val Leu Val Leu Thr Lys Glu Asn Arg Glu
 595 600 605
 Lys Gly Asn His Thr Pro Val Glu Tyr Leu Gly Leu Gly Ser Glu Arg
 610 615 620
 Trp Lys Lys Phe Glu Lys Phe Val Leu Ala Asn Lys Gln Phe Ser Lys
 625 630 635 640

 Lys Lys Lys Gln Asn Leu Leu Arg Leu Arg Tyr Glu Glu Thr Glu Glu
 645 650 655
 Lys Glu Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr Ile Ser Lys
 660 665 670

Phe Phe Ala Asn Phe Ile Lys Glu His Leu Lys Phe Ala Asp Gly Asp
 675 680 685
 Gly Gly Gln Lys Val Tyr Thr Ile Asn Gly Lys Ile Thr Ala His Leu
 690 695 700

 Arg Ser Arg Trp Asp Phe Asn Lys Asn Arg Glu Glu Ser Asp Leu His
 705 710 715 720
 His Ala Val Asp Ala Val Ile Val Ala Cys Ala Thr Gln Gly Met Ile
 725 730 735
 Lys Lys Ile Thr Glu Phe Tyr Lys Ala Arg Glu Gln Asn Lys Glu Ser
 740 745 750
 Ala Lys Lys Lys Glu Pro Ile Phe Pro Gln Pro Trp Pro His Phe Ala
 755 760 765

 Asp Glu Leu Lys Ala Arg Leu Ser Lys Phe Pro Gln Glu Ser Ile Glu
 770 775 780
 Ala Phe Ala Leu Gly Asn Tyr Asp Arg Lys Lys Leu Glu Ser Leu Arg
 785 790 795 800
 Pro Val Phe Val Ser Arg Met Pro Lys Arg Ser Val Thr Gly Ala Ala
 805 810 815
 His Gln Glu Thr Leu Arg Arg Cys Val Gly Ile Asp Glu Gln Ser Gly
 820 825 830

 Lys Ile Gln Thr Ala Val Lys Thr Lys Leu Ser Asp Ile Lys Leu Asp
 835 840 845
 Lys Asp Gly His Phe Pro Met Tyr Gln Lys Glu Ser Asp Pro Arg Thr
 850 855 860
 Tyr Glu Ala Ile Arg Gln Arg Leu Leu Glu His Asn Asn Asp Pro Lys
 865 870 875 880
 Lys Ala Phe Gln Glu Pro Leu Tyr Lys Pro Lys Lys Asn Gly Glu Pro
 885 890 895

 Gly Pro Val Ile Arg Thr Val Lys Ile Ile Asp Thr Lys Asn Lys Val
 900 905 910
 Val His Leu Asp Gly Ser Lys Thr Val Ala Tyr Asn Ser Asn Ile Val

915 920 925
 Arg Thr Asp Val Phe Glu Lys Asp Gly Lys Tyr Tyr Cys Val Pro Val
 930 935 940
 Tyr Thr Met Asp Ile Met Lys Gly Thr Leu Pro Asn Lys Ala Ile Glu
 945 950 955 960

Ala Asn Lys Pro Tyr Ser Glu Trp Lys Glu Met Thr Glu Glu Tyr Thr
 965 970 975
 Phe Gln Phe Ser Leu Phe Pro Asn Asp Leu Val Arg Ile Val Leu Pro
 980 985 990
 Arg Glu Lys Thr Ile Lys Thr Ser Thr Asn Glu Glu Ile Ile Ile Lys
 995 1000 1005
 Asp Ile Phe Ala Tyr Tyr Lys Thr Ile Asp Ser Ala Thr Gly Gly
 1010 1015 1020

Leu Glu Leu Ile Ser His Asp Arg Asn Phe Ser Leu Arg Gly Val
 1025 1030 1035
 Gly Ser Lys Thr Leu Lys Arg Phe Glu Lys Tyr Gln Val Asp Val
 1040 1045 1050
 Leu Gly Asn Ile His Lys Val Lys Gly Glu Lys Arg Val Gly Leu
 1055 1060 1065
 Ala Ala Pro Thr Asn Gln Lys Lys Gly Lys Thr Val Asp Ser Leu
 1070 1075 1080

Gln Ser Val Ser Asp Pro Lys Lys Lys Arg Lys Val
 1085 1090 1095

<210> 3

<211> 4110

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"

<400> 3

atgaccaagc tgaaccagcc ttacggcatc ggcttgaca tcggcagcaa tagcatcggc

60

ttgccgtgg tggacgcaa cagccatctg ctgagactga agggcgagac agccatcggc 120
 gccagactgt ttagagaggg acagagcgcc gctgacagac ggggaagcag aaccacaaga 180

 agggcgctgt ccagaaccag atggcggtg agcttctgc gggatttctt cggccctcac 240
 atcaccaaga tcgacccga cttctttctg cggcaaaaat actccgagat cagccccaag 300
 gacaaggaca ggtttaagta cgagaagcgg ctgttcaacg accggaccga cgccgagttc 360
 tacgaggact accccagcat gtaccacctg agactgcacc tgatgacca cacacacaag 420
 gccgatcctc gggaaatctt cctggccatc caccacatcc tgaagtccag agggcacttt 480
 ctgacacccg gcgtgcca ggacttcaac accgacaaag tggaccttga ggacatcttc 540
 cccgctctga cagaggctta cgcccagggtg taccctgatc tggaactgac cttcgatctg 600

 gccaaaggccg acgacttcaa ggccaagctg ctggacgaac aggccacacc tagcgacaca 660
 cagaaagccc tggtaacct gctgctgtct agcgacggcg agaaagaaat cgtgaagaag 720
 cggaagcagg tctgaccga gttcgccaag gccatcaccg gcctgaaaac aaagtccaat 780
 ctggccctgg gcaccgaggt ggacgaagct gatgcttcca actggcagtt cagcatgggc 840
 cagctggacg acaagtggtc caacatcgaa accagcatga ccgaccaggg caccgaaatc 900
 ttcgagcaga tccaagagct gtaccgggccc agactgctga acggaattgt gcctgccggc 960
 atgagcctgt ctcaggccaa agtggccgat tacggccagc acaaagagga cctggaactg 1020

 ttcaagacct acctgaagaa gctgaacgac cagagctgg ccaagacat caggggcctg 1080
 tacgatcggg acatcaacgg cgacgacgcc aagcctttcc tgcgcgagga ttttgtgaag 1140
 gccctgacca aagaagtgc agctcaccac aacgaggtgt ccgaacagct gctgaacagg 1200
 atgggccaag ccaacttcat gctgaagcag cggaccaagg ccaacggcgc cattcctatt 1260
 cagctgcagc agagagagct ggaccagatc attgccaacc agagcaagta ctacactgg 1320
 ctggcgcgtc ctaatcctgt ggaagccac agatggaaga tgcctacca gctggatgag 1380
 ctgctcaact ttacatccc ctactacgtg ggccctctga tcaccctaa acagcaggcc 1440

 gagagcggcg agaattgtgt cgcttggatg gtccgaaagg accccagcgg caacatcacc 1500
 ccttacaact tcgacgagaa ggtggacaga gaggccagcg ccaacacctt catccagaga 1560
 atgaagacca ccgacacata cctgatcggc gaggacgtgc tgcctaagca gacgtgctg 1620
 taccagaaat acgaggtgct gaacgagctg aacaacgtgc ggatcaaca cgagtgcctg 1680
 ggcacagacc agaagcagag actgatcaga gaggtgttcg agcggcacag cagcgtgacc 1740
 atcaaacagg tggccgacaa tctggtggcc cacggcgatt ttgccagacg gcctgagatt 1800

agaggactgg ccgatgagaa gcggttcctg agcagcctga gcacctacca ccagctgaaa 1860

gagatcctgc acgaggccat cgacgacccc accaaactgc tggatatcga gaacatcatc 1920

acctggtcca ccgtgttcga ggaccacacc atcttcgaga caaagctggc cgagatcgag 1980

tggctggacc ccaagaagat caacgagctg tctggcatca gatacagagg ctggggccag 2040

ttctcccga agctgctcga tggactgaag ctgggcaatg gccacaccgt gattcaagaa 2100

ctgatgctga gcaaccacaa cctgatgcag atcctggccg acgagacact gaaagaaacc 2160

atgacagagc tgaatcagga caagctgaaa accgacgaca tcgaggatgt gatcaacgac 2220

gcctacacaa gccccagcaa caaaaaggcc ctgagacagg tgctgagagt ggtcgaggat 2280

atcaagcacg ccgccaacgg acaggaccct agctggctgt ttatcgaaac cgccgatgga 2340

acaggcaccg ccggcaagag aacacagagc cggcagaaac agatccagac cgtgtacgcc 2400

aacgcgcctc aagagctgat cgattctgcc gtgcggggcg agctggaaga taagattgct 2460

gacaaggcca gcttcaccga ccggctgggtg ctgtacttta tgcaaggcgg cagagacatc 2520

tacacaggcg cccctctgaa catcgaccag ctgagccact acgatatcga ccacattctg 2580

ccccagagcc tgatcaagga cgacagcctg gacaaccggg tgctcgtgaa cgccaccatc 2640

aaccgcgaga agaacaatgt gtttgccagc aactgttcg ccggaaagat gaaggccacc 2700

tggcggaat ggcacgaagc cggactgatc tctggcagaa agctgcggaa tctgatgctg 2760

cggcccgacg agatcgacaa gtttgccaag ggcttcgtgg cccggcagct ggttgagaca 2820

agacagatca tcaagctgac agagcagatt gccgcgcctc agtaccceaa caccaagatt 2880

attgccgtga aggccggact gtcccatcag ctgagagagg aactggactt ccccaagaac 2940

cgggacgtga accactacca ccacgccttc gatgccttc tggccgctag aatcggcacc 3000

tacctgctga agagataccc caagctggcc ccattcttca cctacggcga gtttgctaag 3060

gtggacgtca agaagttccg cgagttcaac ttcacggag ccctgacaca cgccaagaag 3120

aacattatcg ccaaggacac cggcgagatc gtgtgggaca aagagcggga catcagagaa 3180

ctggaccgca tctacaactt caagcggatg ctgatcacac acgaggtgta cttcgagact 3240

gccgacctgt tcaagcagac catctacgcc gctaaggaca gcaaagagag aggcggcagc 3300

aagcagctga tcctaagaa gcagggtac cccactcagg tgtacggcgg ctacacacaa 3360

gagagcggct cttacaacgc cctcgtcaga gtggccgagg ccgatacaac agcctaccaa 3420

gtgatcaaga tcagcgccca gaacgccagc aagatgcct ccgccaacct gaaaagccgc 3480

gagaaaggca aacagctcct gaatgagatc gtcgtgaagc agctggctaa gcggcggaag 3540

aactggaagc ctacgcgcaa tagcttcaag atcgtgatcc ccagattcgg catgggcacc 3600
 ctgttccaga acgctaagta cggcctgttc atggtcaaca gcgacaccta ctaccggaac 3660
 taccaagaac tctggctgag ccgggaaaac cagaaactgc tgaaaaagct gttctccatc 3720
 aaatacgaga aaaccagat gaaccacgac gccctgcagg tctacaaggc cattatcgac 3780
 caggtggaaa agttcttcaa gctgtacgac atcaaccagt tccgcgcaa gctgagcgac 3840
 gccatcgaga gatttgagaa gctgcccac aataccgacg gcaacaagat cggcaagacc 3900
 gagactctga gacagatcct gatcggactg caggccaatg gcacccggtc caacgtgaag 3960

aacctgggca tcaagaccga tctgggcctg ctgcaagtcg gcagcggaat caagctggac 4020
 aaggatacc agatcgtgta tcagagcccc tccggcctgt ttaagcggag aatcccactg 4080
 gctgacctgc ccaagaagaa gaggaaggtg 4110

<210> 4

<211> 1370

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
 polypeptide"

<400> 4

Met Thr Lys Leu Asn Gln Pro Tyr Gly Ile Gly Leu Asp Ile Gly Ser
 1 5 10 15

Asn Ser Ile Gly Phe Ala Val Val Asp Ala Asn Ser His Leu Leu Arg
 20 25 30

Leu Lys Gly Glu Thr Ala Ile Gly Ala Arg Leu Phe Arg Glu Gly Gln
 35 40 45

Ser Ala Ala Asp Arg Arg Gly Ser Arg Thr Thr Arg Arg Arg Leu Ser
 50 55 60

Arg Thr Arg Trp Arg Leu Ser Phe Leu Arg Asp Phe Phe Ala Pro His
 65 70 75 80

Ile Thr Lys Ile Asp Pro Asp Phe Phe Leu Arg Gln Lys Tyr Ser Glu
 85 90 95

Ile Ser Pro Lys Asp Lys Asp Arg Phe Lys Tyr Glu Lys Arg Leu Phe
 100 105 110

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

 Glu Ile Phe Leu Ala Ile His His Ile Leu Lys Ser Arg Gly His Phe
 145 150 155 160
 Leu Thr Pro Gly Ala Ala Lys Asp Phe Asn Thr Asp Lys Val Asp Leu
 165 170 175
 Glu Asp Ile Phe Pro Ala Leu Thr Glu Ala Tyr Ala Gln Val Tyr Pro
 180 185 190
 Asp Leu Glu Leu Thr Phe Asp Leu Ala Lys Ala Asp Asp Phe Lys Ala
 195 200 205

 Lys Leu Leu Asp Glu Gln Ala Thr Pro Ser Asp Thr Gln Lys Ala Leu
 210 215 220
 Val Asn Leu Leu Leu Ser Ser Asp Gly Glu Lys Glu Ile Val Lys Lys
 225 230 235 240
 Arg Lys Gln Val Leu Thr Glu Phe Ala Lys Ala Ile Thr Gly Leu Lys
 245 250 255
 Thr Lys Phe Asn Leu Ala Leu Gly Thr Glu Val Asp Glu Ala Asp Ala
 260 265 270

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

 Asp Leu Glu Leu Phe Lys Thr Tyr Leu Lys Lys Leu Asn Asp His Glu
 340 345 350
 Leu Ala Lys Thr Ile Arg Gly Leu Tyr Asp Arg Tyr Ile Asn Gly Asp

355	360	365
Asp Ala Lys Pro Phe Leu Arg Glu Asp Phe Val Lys Ala Leu Thr Lys		
370	375	380
Glu Val Thr Ala His Pro Asn Glu Val Ser Glu Gln Leu Leu Asn Arg		
385	390	395
		400
Met Gly Gln Ala Asn Phe Met Leu Lys Gln Arg Thr Lys Ala Asn Gly		
405	410	415
Ala Ile Pro Ile Gln Leu Gln Gln Arg Glu Leu Asp Gln Ile Ile Ala		
420	425	430
Asn Gln Ser Lys Tyr Tyr Asp Trp Leu Ala Ala Pro Asn Pro Val Glu		
435	440	445
Ala His Arg Trp Lys Met Pro Tyr Gln Leu Asp Glu Leu Leu Asn Phe		
450	455	460
His Ile Pro Tyr Tyr Val Gly Pro Leu Ile Thr Pro Lys Gln Gln Ala		
465	470	475
		480
Glu Ser Gly Glu Asn Val Phe Ala Trp Met Val Arg Lys Asp Pro Ser		
485	490	495
Gly Asn Ile Thr Pro Tyr Asn Phe Asp Glu Lys Val Asp Arg Glu Ala		
500	505	510
Ser Ala Asn Thr Phe Ile Gln Arg Met Lys Thr Thr Asp Thr Tyr Leu		
515	520	525
Ile Gly Glu Asp Val Leu Pro Lys Gln Ser Leu Leu Tyr Gln Lys Tyr		
530	535	540
Glu Val Leu Asn Glu Leu Asn Asn Val Arg Ile Asn Asn Glu Cys Leu		
545	550	555
		560
Gly Thr Asp Gln Lys Gln Arg Leu Ile Arg Glu Val Phe Glu Arg His		
565	570	575
Ser Ser Val Thr Ile Lys Gln Val Ala Asp Asn Leu Val Ala His Gly		
580	585	590
Asp Phe Ala Arg Arg Pro Glu Ile Arg Gly Leu Ala Asp Glu Lys Arg		
595	600	605

Phe Leu Ser Ser Leu Ser Thr Tyr His Gln Leu Lys Glu Ile Leu His
 610 615 620
 Glu Ala Ile Asp Asp Pro Thr Lys Leu Leu Asp Ile Glu Asn Ile Ile
 625 630 635 640
 Thr Trp Ser Thr Val Phe Glu Asp His Thr Ile Phe Glu Thr Lys Leu
 645 650 655

 Ala Glu Ile Glu Trp Leu Asp Pro Lys Lys Ile Asn Glu Leu Ser Gly
 660 665 670
 Ile Arg Tyr Arg Gly Trp Gly Gln Phe Ser Arg Lys Leu Leu Asp Gly
 675 680 685
 Leu Lys Leu Gly Asn Gly His Thr Val Ile Gln Glu Leu Met Leu Ser
 690 695 700
 Asn His Asn Leu Met Gln Ile Leu Ala Asp Glu Thr Leu Lys Glu Thr
 705 710 715 720

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

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

 Asp Gln Leu Ser His Tyr Asp Ile Asp His Ile Leu Pro Gln Ser Leu

850	855	860	
Ile Lys Asp Asp Ser Leu Asp Asn Arg Val Leu Val Asn Ala Thr Ile			
865	870	875	880
Asn Arg Glu Lys Asn Asn Val Phe Ala Ser Thr Leu Phe Ala Gly Lys			
	885	890	895
Met Lys Ala Thr Trp Arg Lys Trp His Glu Ala Gly Leu Ile Ser Gly			
	900	905	910
Arg Lys Leu Arg Asn Leu Met Leu Arg Pro Asp Glu Ile Asp Lys Phe			
	915	920	925
Ala Lys Gly Phe Val Ala Arg Gln Leu Val Glu Thr Arg Gln Ile Ile			
	930	935	940
Lys Leu Thr Glu Gln Ile Ala Ala Ala Gln Tyr Pro Asn Thr Lys Ile			
945	950	955	960
Ile Ala Val Lys Ala Gly Leu Ser His Gln Leu Arg Glu Glu Leu Asp			
	965	970	975
Phe Pro Lys Asn Arg Asp Val Asn His Tyr His His Ala Phe Asp Ala			
	980	985	990
Phe Leu Ala Ala Arg Ile Gly Thr Tyr Leu Leu Lys Arg Tyr Pro Lys			
	995	1000	1005
Leu Ala Pro Phe Phe Thr Tyr Gly Glu Phe Ala Lys Val Asp Val			
	1010	1015	1020
Lys Lys Phe Arg Glu Phe Asn Phe Ile Gly Ala Leu Thr His Ala			
	1025	1030	1035
Lys Lys Asn Ile Ile Ala Lys Asp Thr Gly Glu Ile Val Trp Asp			
	1040	1045	1050
Lys Glu Arg Asp Ile Arg Glu Leu Asp Arg Ile Tyr Asn Phe Lys			
	1055	1060	1065
Arg Met Leu Ile Thr His Glu Val Tyr Phe Glu Thr Ala Asp Leu			
	1070	1075	1080
Phe Lys Gln Thr Ile Tyr Ala Ala Lys Asp Ser Lys Glu Arg Gly			
	1085	1090	1095

Gly Ser	Lys Gln Leu Ile Pro	Lys Lys Gln Gly Tyr	Pro Thr Gln
1100	1105	1110	
Val Tyr	Gly Gly Tyr Thr Gln	Glu Ser Gly Ser Tyr	Asn Ala Leu
1115	1120	1125	
Val Arg	Val Ala Glu Ala Asp	Thr Thr Ala Tyr Gln	Val Ile Lys
1130	1135	1140	
Ile Ser	Ala Gln Asn Ala Ser	Lys Ile Ala Ser Ala	Asn Leu Lys
1145	1150	1155	
Ser Arg	Glu Lys Gly Lys Gln	Leu Leu Asn Glu Ile	Val Val Lys
1160	1165	1170	
Gln Leu	Ala Lys Arg Arg Lys	Asn Trp Lys Pro Ser	Ala Asn Ser
1175	1180	1185	
Phe Lys	Ile Val Ile Pro Arg	Phe Gly Met Gly Thr	Leu Phe Gln
1190	1195	1200	
Asn Ala	Lys Tyr Gly Leu Phe	Met Val Asn Ser Asp	Thr Tyr Tyr
1205	1210	1215	
Arg Asn	Tyr Gln Glu Leu Trp	Leu Ser Arg Glu Asn	Gln Lys Leu
1220	1225	1230	
Leu Lys	Lys Leu Phe Ser Ile	Lys Tyr Glu Lys Thr	Gln Met Asn
1235	1240	1245	
His Asp	Ala Leu Gln Val Tyr	Lys Ala Ile Ile Asp	Gln Val Glu
1250	1255	1260	
Lys Phe	Phe Lys Leu Tyr Asp	Ile Asn Gln Phe Arg	Ala Lys Leu
1265	1270	1275	
Ser Asp	Ala Ile Glu Arg Phe	Glu Lys Leu Pro Ile	Asn Thr Asp
1280	1285	1290	
Gly Asn	Lys Ile Gly Lys Thr	Glu Thr Leu Arg Gln	Ile Leu Ile
1295	1300	1305	
Gly Leu	Gln Ala Asn Gly Thr	Arg Ser Asn Val Lys	Asn Leu Gly
1310	1315	1320	
Ile Lys	Thr Asp Leu Gly Leu	Leu Gln Val Gly Ser	Gly Ile Lys

1325 1330 1335

Leu Asp Lys Asp Thr Gln Ile Val Tyr Gln Ser Pro Ser Gly Leu

1340 1345 1350

Phe Lys Arg Arg Ile Pro Leu Ala Asp Leu Pro Lys Lys Lys Arg

1355 1360 1365

Lys Val

1370

<210> 5

<211> 4305

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 5

atgggcaaga cccacatcat cggcgttggc ctggatctcg gcggcacata cacaggcacc 60

ttcataacca gccatcctag cgacgaagcc gagcacagag atcacagcag cgccttcacc 120

gtggtcaaca gcgagaagct gagcttcagc agcaagagca gaacagccgt gcggcacaga 180

gtgcggagct acaagggtt cgacctgcgt agaaggctgc tgcttctggt ggccgagtat 240

cagctgctgc agaagaagca gacctggcc cctgaggaaa gagagaacct gagaatcgcc 300

ctgagcggct acctgaagag aagaggctac gccagaaccg aggccgagac agatacaagc 360

gtgctggaat ctctggacct cagcgtgttc agcagcgtc ccagcttcac caatttcttc 420

aacgacagcg agcccctgaa catccagtgg gaagccattg ccaactctcc cgagacaaca 480

aaggccctga acaaagagct gagcggccag aaagaggccg acttcaagaa gtacatcaag 540

accagcttcc ccgagtacag cgccaaagag attctggcca actacgtgga aggcagacgg 600

gccattctgg acgccagcaa gtatatcgcc aacctgcaga gcctgggcca caagcacaga 660

agcaagtacc tgagcgacat tctgcaggac atgaagcggg acagccggat caccagactg 720

agcgaagcct ttggcagcac cgacaacctg tggcggatca tcggcaacat cagcaatctg 780

caagaacggg ccgtgcggtg gtacttcaac gatgccaagt tcgagcaggg ccaagagcag 840

ctggatgccg tgaagctgaa gaatgtgtc gtgcgggccc tgaagtatct gcggagtgtac 900

gacaaagagt ggagcgcctc tcagaagcag atcatccagt ctctggaaca gagcggcgac 960

gtgctggatg tgctggctgg actcgacccc gacagaacaa tccctccata cgaggaccag 1020
aacaacagac ggctctcta ggatcagacc ctgtatctga accccaaggc tctgagcagc 1080
gagtacggcg agaagtggaa gtcttgggcc aacaagtgtg ccggcgctta ccctctgctg 1140
accgaggatc tgaccgagat cctgaagaac accgacagaa agtcccggat caagatcaga 1200
tccgatgtgc tgcccgacag cgactacaga ctggcctaca tctgcagag agccttcgat 1260
cggctctatcg ccctggacga gtgcagcatc agaagaaccg ccgaggactt cgagaacggc 1320

gtggtcatca agaacgagaa actggaagat gtgctgagcg gacaccagct ggaagagttt 1380
ctggaatttg ccaatcggtta ctaccaagag acagccaagg ccaagaacgg cctgtggttc 1440
ccagagaacg ccctgctgga aagagccgat ctgcaccctc ctatgaagaa caagattctg 1500
aacgtgatcg tcggacaggc cctgggagtg tctcctgctg agggcaccga tttcatcgag 1560
gaaatttggg acagcaaagt gaaaggccgg tccaccgtgc ggagcatctg taacgccatc 1620
gagaatgaga gaaagaccta cggaccctac ttcagcgagg actacaagtt cgtgaaaacg 1680
gccctgaaag agggcaaac cgagaaagag ctgtccaaga aattcgccgc cgtgatcaag 1740

gtgctgaaga tgggtgtctga ggtggtgccc ttatcggaa aagagctgcg gctgtctgac 1800
gaggcccaga gcaagttcga caatctgtac tctctggccc agctgtacaa cctgatcgag 1860
acagagcgga acggcttcag caaggtgtca ctggctgccc acctggaaaa tgcctggcgg 1920
atgaccatga cagatggatc cgcccagtgc tgtagactgc ctgccgattg tgtgcggccc 1980
ttcgacggct ttatccggaa ggccatcgac cggaactctt gggaagtcgc caagcgatt 2040
gccgaggaag tgaagaagtc cgtcgacttc accaacggca ccgtgaagat ccctgtggcc 2100
atcgaggcca acagcttcaa ctttaccgcc agcctgaccg acctgaagta cattcagctc 2160

aaagaacaga agctcaagaa gaagttggag gacatccagc ggaacgaaga gaatcaagag 2220
aagcgttgcc tgagcaaaga ggaacggatc agagccgaca gccacggcat ctgtgcctat 2280
actggcagac ccctggatga cgtgggcgag atcgatcaca tcatccccag aagcctgaca 2340
ctgaagaaaa gcgagagcat ctacaactcc gaagtgaacc tgatcttcgt gtctgccag 2400
ggcaatcaag aaaagaagaa caacatctac ctgctgagca acctcgcaa gaactacctg 2460
gccgccgtgt ttggcacaag cgacctgagc cagatcacca acgagatcga gagcacctg 2520
ctgcagctga aagctgctgg cagactgggc tacttcgac tgctgagcga aaaagagcgg 2580

gcctgcgcca gacatgcctt gtttctgaat agcgactccg aggccagacg cgccgtgatt 2640
gatgttcttg gctctcgag aaagccagc gtgaacggaa cccaggcttg gtttgtcgg 2700
tccatcttct ccaaagtgcg gcaggcactg gccgttggg cacaagaaac aggcaacgag 2760
ctgatctttg acgccatcag cgtgccagcc gccgatagct ctgagatgag gaagagattc 2820

gccgagtacc ggccctgagtt cagaaagccc aaagtgcagc ctgtggcctc tcacagcatc 2880

gacgccatgt gcatctatct ggccgcctgc agcgaccctt tcaagaccaa gagaatgggc 2940

tctcagctgg ccatctacga gcccataaac ttcgataacc tgttcaccgg cagctgtcaa 3000

gtgatccaga acaccctcgc gaacttctcc gacaagacca atatcgctaa cagcccatc 3060

ttcaaagaga caatctacgc cgagcggttc ctggacatca tcgtgtccag aggcgagatt 3120

ttcatcggtc accccagcaa catgcccttc gaggaagaag ccaaccggat cagcatcggc 3180

ggcaaggacc ctttcagcat cctgtctgtg ctgggcgcct acctggataa ggcccctagc 3240

agcgagaaag aaaagctcac catctaccgg gtcgtcaaga acaaagcctt cgagctgttc 3300

tccaaggtgg ccggcagcaa gtttaccgcc gaagaagata aggcgcgcaa gatcctggaa 3360

gccctgcact tcgtgaccgt gaaacaggat gtggccgcca ccgtgtccga tctgatcaag 3420

agcaagaaag aactgagcaa ggatagcatc gagaacctgg ccaagcagaa gggctgcctg 3480

aagaaggtgg aatctccag caaagagttc aagttcaagg gcagcctgat catccctgcc 3540

gccgtggaat ggggaaaagt gctgtggaac gtgttcaaag aaaacacggc cgaagaactg 3600

aaggacgaga acgtctgag gaagccctg gaagctgcct ggccctagtc ttccggcacc 3660

agaaacctgc actctaaggc caagcgggtg ttcagcctgc ctgtggtggc tacacaatct 3720

ggcgccgtgc ggatcagacg caagacagcc ttcggcgact tcgtgtacca gagccaggac 3780

acaaacaacc tgtacagcag ctccccctg aagaacggca agctggattg gagcagccct 3840

atcattcacc ccgtctgca gaaccggaac ctgaccgctt acggctacag attcgtggac 3900

cacgacagat ccatcagcat gagcgagttc agagaggtgt acaacaagga cgacctgatg 3960

cggatcgagc tggcccaggg aacaagcagc agacgctacc tgagagtgga aatgcccggc 4020

gagaaattcc tcgtttggtt tggcgagaac agcatcagcc tgggctccag cttcaagtcc 4080

tctgtgtccg aggtgttcga caacaaaatc tacaccgaga acgcccagtt taccaagtcc 4140

ctgcctaagc ctagagagga caacaagcac aacgggacca tctttttcga actcgtgggc 4200

cccagagtga tcttcaacta catcgttggc ggagccgcca gcagcctgaa agaaatcttt 4260

agcgaggccg gcaaagagcg gagccccaag aagaagagga aggtg 4305

<210> 6

<211> 1435

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

polypeptide"

<400> 6

Met Gly Lys Thr His Ile Ile Gly Val Gly Leu Asp Leu Gly Gly Thr

1 5 10 15

Tyr Thr Gly Thr Phe Ile Thr Ser His Pro Ser Asp Glu Ala Glu His

20 25 30

Arg Asp His Ser Ser Ala Phe Thr Val Val Asn Ser Glu Lys Leu Ser

35 40 45

Phe Ser Ser Lys Ser Arg Thr Ala Val Arg His Arg Val Arg Ser Tyr

50 55 60

Lys Gly Phe Asp Leu Arg Arg Arg Leu Leu Leu Leu Val Ala Glu Tyr

65 70 75 80

Gln Leu Leu Gln Lys Lys Gln Thr Leu Ala Pro Glu Glu Arg Glu Asn

85 90 95

Leu Arg Ile Ala Leu Ser Gly Tyr Leu Lys Arg Arg Gly Tyr Ala Arg

100 105 110

Thr Glu Ala Glu Thr Asp Thr Ser Val Leu Glu Ser Leu Asp Pro Ser

115 120 125

Val Phe Ser Ser Ala Pro Ser Phe Thr Asn Phe Phe Asn Asp Ser Glu

130 135 140

Pro Leu Asn Ile Gln Trp Glu Ala Ile Ala Asn Ser Pro Glu Thr Thr

145 150 155 160

Lys Ala Leu Asn Lys Glu Leu Ser Gly Gln Lys Glu Ala Asp Phe Lys

165 170 175

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

180 185 190

Ala Asn Tyr Val Glu Gly Arg Arg Ala Ile Leu Asp Ala Ser Lys Tyr

195 200 205

Ile Ala Asn Leu Gln Ser Leu Gly His Lys His Arg Ser Lys Tyr Leu

210 215 220

Ser Asp Ile Leu Gln Asp Met Lys Arg Asp Ser Arg Ile Thr Arg Leu

225 230 235 240
 Ser Glu Ala Phe Gly Ser Thr Asp Asn Leu Trp Arg Ile Ile Gly Asn
 245 250 255
 Ile Ser Asn Leu Gln Glu Arg Ala Val Arg Trp Tyr Phe Asn Asp Ala
 260 265 270
 Lys Phe Glu Gln Gly Gln Glu Gln Leu Asp Ala Val Lys Leu Lys Asn
 275 280 285

 Val Leu Val Arg Ala Leu Lys Tyr Leu Arg Ser Asp Asp Lys Glu Trp
 290 295 300
 Ser Ala Ser Gln Lys Gln Ile Ile Gln Ser Leu Glu Gln Ser Gly Asp
 305 310 315 320
 Val Leu Asp Val Leu Ala Gly Leu Asp Pro Asp Arg Thr Ile Pro Pro
 325 330 335
 Tyr Glu Asp Gln Asn Asn Arg Arg Pro Pro Glu Asp Gln Thr Leu Tyr
 340 345 350

 Leu Asn Pro Lys Ala Leu Ser Ser Glu Tyr Gly Glu Lys Trp Lys Ser
 355 360 365
 Trp Ala Asn Lys Phe Ala Gly Ala Tyr Pro Leu Leu Thr Glu Asp Leu
 370 375 380
 Thr Glu Ile Leu Lys Asn Thr Asp Arg Lys Ser Arg Ile Lys Ile Arg
 385 390 395 400
 Ser Asp Val Leu Pro Asp Ser Asp Tyr Arg Leu Ala Tyr Ile Leu Gln
 405 410 415

 Arg Ala Phe Asp Arg Ser Ile Ala Leu Asp Glu Cys Ser Ile Arg Arg
 420 425 430
 Thr Ala Glu Asp Phe Glu Asn Gly Val Val Ile Lys Asn Glu Lys Leu
 435 440 445
 Glu Asp Val Leu Ser Gly His Gln Leu Glu Glu Phe Leu Glu Phe Ala
 450 455 460
 Asn Arg Tyr Tyr Gln Glu Thr Ala Lys Ala Lys Asn Gly Leu Trp Phe
 465 470 475 480

Pro Glu Asn Ala Leu Leu Glu Arg Ala Asp Leu His Pro Pro Met Lys
485 490 495

Asn Lys Ile Leu Asn Val Ile Val Gly Gln Ala Leu Gly Val Ser Pro
500 505 510

Ala Glu Gly Thr Asp Phe Ile Glu Glu Ile Trp Asn Ser Lys Val Lys
515 520 525

Gly Arg Ser Thr Val Arg Ser Ile Cys Asn Ala Ile Glu Asn Glu Arg
530 535 540

Lys Thr Tyr Gly Pro Tyr Phe Ser Glu Asp Tyr Lys Phe Val Lys Thr
545 550 555 560

Ala Leu Lys Glu Gly Lys Thr Glu Lys Glu Leu Ser Lys Lys Phe Ala
565 570 575

Ala Val Ile Lys Val Leu Lys Met Val Ser Glu Val Val Pro Phe Ile
580 585 590

Gly Lys Glu Leu Arg Leu Ser Asp Glu Ala Gln Ser Lys Phe Asp Asn
595 600 605

Leu Tyr Ser Leu Ala Gln Leu Tyr Asn Leu Ile Glu Thr Glu Arg Asn
610 615 620

Gly Phe Ser Lys Val Ser Leu Ala Ala His Leu Glu Asn Ala Trp Arg
625 630 635 640

Met Thr Met Thr Asp Gly Ser Ala Gln Cys Cys Arg Leu Pro Ala Asp
645 650 655

Cys Val Arg Pro Phe Asp Gly Phe Ile Arg Lys Ala Ile Asp Arg Asn
660 665 670

Ser Trp Glu Val Ala Lys Arg Ile Ala Glu Glu Val Lys Lys Ser Val
675 680 685

Asp Phe Thr Asn Gly Thr Val Lys Ile Pro Val Ala Ile Glu Ala Asn
690 695 700

Ser Phe Asn Phe Thr Ala Ser Leu Thr Asp Leu Lys Tyr Ile Gln Leu
705 710 715 720

Lys Glu Gln Lys Leu Lys Lys Lys Leu Glu Asp Ile Gln Arg Asn Glu

725

730

735

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

740

745

750

Asp Ser His Gly Ile Cys Ala Tyr Thr Gly Arg Pro Leu Asp Asp Val

755

760

765

Gly Glu Ile Asp His Ile Ile Pro Arg Ser Leu Thr Leu Lys Lys Ser

770

775

780

Glu Ser Ile Tyr Asn Ser Glu Val Asn Leu Ile Phe Val Ser Ala Gln

785

790

795

800

Gly Asn Gln Glu Lys Lys Asn Asn Ile Tyr Leu Leu Ser Asn Leu Ala

805

810

815

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

820

825

830

Thr Asn Glu Ile Glu Ser Thr Val Leu Gln Leu Lys Ala Ala Gly Arg

835

840

845

Leu Gly Tyr Phe Asp Leu Leu Ser Glu Lys Glu Arg Ala Cys Ala Arg

850

855

860

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

865

870

875

880

Asp Val Leu Gly Ser Arg Arg Lys Ala Ser Val Asn Gly Thr Gln Ala

885

890

895

Trp Phe Val Arg Ser Ile Phe Ser Lys Val Arg Gln Ala Leu Ala Ala

900

905

910

Trp Thr Gln Glu Thr Gly Asn Glu Leu Ile Phe Asp Ala Ile Ser Val

915

920

925

Pro Ala Ala Asp Ser Ser Glu Met Arg Lys Arg Phe Ala Glu Tyr Arg

930

935

940

Pro Glu Phe Arg Lys Pro Lys Val Gln Pro Val Ala Ser His Ser Ile

945

950

955

960

Asp Ala Met Cys Ile Tyr Leu Ala Ala Cys Ser Asp Pro Phe Lys Thr

965

970

975

Lys Arg Met Gly Ser Gln Leu Ala Ile Tyr Glu Pro Ile Asn Phe Asp
 980 985 990

Asn Leu Phe Thr Gly Ser Cys Gln Val Ile Gln Asn Thr Pro Arg Asn
 995 1000 1005

Phe Ser Asp Lys Thr Asn Ile Ala Asn Ser Pro Ile Phe Lys Glu
 1010 1015 1020

Thr Ile Tyr Ala Glu Arg Phe Leu Asp Ile Ile Val Ser Arg Gly
 1025 1030 1035

Glu Ile Phe Ile Gly Tyr Pro Ser Asn Met Pro Phe Glu Glu Lys
 1040 1045 1050

Pro Asn Arg Ile Ser Ile Gly Gly Lys Asp Pro Phe Ser Ile Leu
 1055 1060 1065

Ser Val Leu Gly Ala Tyr Leu Asp Lys Ala Pro Ser Ser Glu Lys
 1070 1075 1080

Glu Lys Leu Thr Ile Tyr Arg Val Val Lys Asn Lys Ala Phe Glu
 1085 1090 1095

Leu Phe Ser Lys Val Ala Gly Ser Lys Phe Thr Ala Glu Glu Asp
 1100 1105 1110

Lys Ala Ala Lys Ile Leu Glu Ala Leu His Phe Val Thr Val Lys
 1115 1120 1125

Gln Asp Val Ala Ala Thr Val Ser Asp Leu Ile Lys Ser Lys Lys
 1130 1135 1140

Glu Leu Ser Lys Asp Ser Ile Glu Asn Leu Ala Lys Gln Lys Gly
 1145 1150 1155

Cys Leu Lys Lys Val Glu Tyr Ser Ser Lys Glu Phe Lys Phe Lys
 1160 1165 1170

Gly Ser Leu Ile Ile Pro Ala Ala Val Glu Trp Gly Lys Val Leu
 1175 1180 1185

Trp Asn Val Phe Lys Glu Asn Thr Ala Glu Glu Leu Lys Asp Glu
 1190 1195 1200

Asn Ala Leu Arg Lys Ala Leu Glu Ala Ala Trp Pro Ser Ser Phe

1205	1210	1215
Gly Thr Arg Asn Leu His Ser	Lys Ala Lys Arg Val	Phe Ser Leu
1220	1225	1230
Pro Val Val Ala Thr Gln Ser	Gly Ala Val Arg Ile	Arg Arg Lys
1235	1240	1245
Thr Ala Phe Gly Asp Phe Val	Tyr Gln Ser Gln Asp	Thr Asn Asn
1250	1255	1260
Leu Tyr Ser Ser Phe Pro Val	Lys Asn Gly Lys Leu	Asp Trp Ser
1265	1270	1275
Ser Pro Ile Ile His Pro Ala	Leu Gln Asn Arg Asn	Leu Thr Ala
1280	1285	1290
Tyr Gly Tyr Arg Phe Val Asp	His Asp Arg Ser Ile	Ser Met Ser
1295	1300	1305
Glu Phe Arg Glu Val Tyr Asn	Lys Asp Asp Leu Met	Arg Ile Glu
1310	1315	1320
Leu Ala Gln Gly Thr Ser Ser	Arg Arg Tyr Leu Arg	Val Glu Met
1325	1330	1335
Pro Gly Glu Lys Phe Leu Ala	Trp Phe Gly Glu Asn	Ser Ile Ser
1340	1345	1350
Leu Gly Ser Ser Phe Lys Phe	Ser Val Ser Glu Val	Phe Asp Asn
1355	1360	1365
Lys Ile Tyr Thr Glu Asn Ala	Glu Phe Thr Lys Phe	Leu Pro Lys
1370	1375	1380
Pro Arg Glu Asp Asn Lys His	Asn Gly Thr Ile Phe	Phe Glu Leu
1385	1390	1395
Val Gly Pro Arg Val Ile Phe	Asn Tyr Ile Val Gly	Gly Ala Ala
1400	1405	1410
Ser Ser Leu Lys Glu Ile Phe	Ser Glu Ala Gly Lys	Glu Arg Ser
1415	1420	1425
Pro Lys Lys Lys Arg Lys Val		
1430	1435	

<210> 7

<211> 3720

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 7

atggaaaaga agcggaaagt caccctgggc ttcgacctgg gaatcgctc tgttgatgg	60
gccatcgtgg acagcgagac aaaccaggtg tacaagctgg gcagcagact gttcgacgcc	120
cctgacacca acctggaaag aagaaccag cggggcacca gaaggctgct gcggagaaga	180
aagtaccgga accagaaatt ctacaacctg gtcaagcgga ccgaggtgtt cggcctgtct	240
agcagagagg ccatcgagaa cagattcaga gagctgagca tcaagtacc caacatcatc	300
gagctgaaaa caaaggccct gagccaagaa gtgtgccccg acgagattgc ctggattctg	360
cacgactacc tgaagaaccg gggctacttc tacgacgaga aagagacaaa agaggacttc	420
gaccagcaga ccgtggaatc catgcctagc tacaagctga acgagttcta caagaagtac	480
ggctacttca aaggcgccct gtctcagcct accgagagcg agatgaagga caacaaggac	540
ctgaaagagg cattcttctt cgacttctcc aacaaagagt ggctgaaaga gatcaactac	600
ttcttcaacg tgcagaagaa catcctgagc gagacattca tcgaagagtt caagaagatt	660
ttcagcttca cccgggacat cagcaaaggc ccaggcagcg acaatatgcc ctctccttac	720
ggcatcttcg gcgagttcgg cgacaatggc caaggcggca gatacgagca catctgggac	780
aagaacatcg gcaagtgcag catcttcacc aacgagcaga gagcccctaa gtacctgcct	840
agcgctctga tcttcaactt cctgaacgag ctggccaaca tcagactgta cagcaccgac	900
aagaagaata tccagcctct gtggaagctg agcagcatcg ataagctgaa tatcctgctg	960
aacctgttca acctgcctat cagcgagaag aagaaaaagc tgaccagcac caacatcaac	1020
gacatcgtga agaaagagtc catcaagagc atcatgctga gcgtcgagga catcgacatg	1080
atcaaggatg agtgggccgg caaagaaccc aacgtgtacg gcgttggaact gagcggcctg	1140
aacatcgagg aaagcgccaa agagaacaag ttcaagttcc aagacctgaa gatcctgaac	1200
gtcctgatca atctgctgga caacgtgggc atcaagttcg agttcaagga ccgcagcgac	1260
atcatcaaga acctggaact gctggataac ctgtacctgt tctgatcta ccagaaagag	1320
agcaacaaca aagacagctc catcgacctg ttatcgcca agaacaagtc cctgaatatc	1380

gagaacctga agctcaagct caaagagttc ctgctcggag ccggcaacga gttcgagaac 1440
 cacaacagca agaccacag cctgtccaag aaggccattg acgccatcct gcctaagctg 1500
 ctcgacaaca acgaaggctg gaatctggaa gccatcaaga attacgacga ggaaatcaag 1560
 agccagatcg aggacaactc cagcctgatg gccaaagcagg ataagaagta cctgaacgac 1620
 aacttcctca aggatgccat tctgccgcca aacgtgaaag tgaccttcca gcaggccatc 1680
 ctcatcttca acaagatcat ccagaagttc agcaaggatt tcgagatcga caaggtcgtg 1740
 atcgaactgg ccagagagat gaccaggac caagagaacg acgccctgaa gggaatcgct 1800

 aaggcccaga agtccaagaa aagcctggtg gaagagagac tcgaagccaa caacatcgac 1860
 aagagcgtgt tcaacgataa gtacgagaag cttatctaca agattttcct gtggatcagc 1920
 caggacttta aggacccta caccggcgcc aagatcagcg ccaatgagat cgtggataac 1980
 aaggtggaaa tcgaccacat catcccttac agcctgtgct tcgacgacag cagcgccaac 2040
 aaagtgctgg tgcacaagca gagcaatcaa gagaagtcta acagcctgcc gtacgagtac 2100
 atcaagcagg gccactccgg ctggaactgg gacgagtta ccaaatacgt gaagcgggtg 2160
 ttcgtgaaca acgtggactc tatctgagc aagaaagagc gcctgaagaa gtccgagaat 2220

 ctgctgacca ccagctacga cggctatgag aagctgggct tcctggccag aaacctgaat 2280
 gacaccagat acgccacat cctgttcgg gaccagctga acaattacgc cgagcaccac 2340
 ctgatcgata acaagaaaat gttcaaagt atcgccatga acggggccgt gaccagcttc 2400
 atccggaaga acatgagcta cgacaacaag ctgcggctga aggacagaag cgacttcagc 2460
 caccacgct acgacgccgc catcattgcc ctgttcagca acaagacaa gacgctgtac 2520
 aacctgattg accccagcct gaacggcatc atcagcaaga gaagcgaagg ctattgggtc 2580
 atcgaggatc ggtacacagg cgagatcaaa gagcttaaga aagaggattg gacctctatc 2640

 aagaacaatg tgcaggcccg gaagatcgcc aaagaaatcg aggaatatct gatcgacctg 2700
 gacgatgagg tgttcttcag ccggaaaact aagcgcaaga ccaaccggca gctgtacaat 2760
 gagacaatct acggaatcgc cgccaagacc gacgaggacg gcatcacaa ctactacaag 2820
 aaagaaaagt tctccatcct ggacgacaag gacatctacc tgcggctgct gagagaacgc 2880
 gagaagttcg tgatcaacca gagcaacccc gaagtgatcg accagattat cgagatcatc 2940
 gagagctacg ggaagaaaaa caacatcccc agccgcgacg aggccatcaa tatcaagtac 3000
 acgaagaaca agattaacta caacctctac ctcaagcagt acatgcggag cctgaccaag 3060

 agcctggacc agttcagcga gggcttcac aatcagatga tcgccaacaa gacgttcgtg 3120
 ctgtataacc ccaccaagaa cacaacgcgg aagatcaagt tctgcggct cgtgaacgat 3180
 gtgaagatca acgatatcgc caagaatcaa gtgatcaaca agtttaacgg gaagaacaac 3240

gagcccaagg ccttctacga gaatatcaac agcctgggcg ccatcgtgtt caagtcctcc 3300
gccaacaact tcaagaccct gtccatcaac acccagatcg ccatcttcgg agacaagaac 3360
tgggatatcg aggatttcaa gacctacaac atggaaaaaa tcgagaagta caaagagata 3420
tacggcatcg acaaaacctt caacttccac agctttatct tccccggcac aatcctgctc 3480

gataagcaga acaaagagtt ctactacatc agcagcatcc agaccgtgaa cgaccaaatt 3540
gagctgaagt ttctgaacaa gatcgagttt aagaacgacg acaacacctc cggggccaac 3600
aagcctctct ggagactgag attcggcatt aagtccatca tgaacaacta cgagcaggtc 3660
gacatcagcc ctttcggcat caacaagaag atattcgagc ccaagaagaa gaggaaggtg 3720

<210> 8

<211> 1247

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 8

Met Glu Lys Lys Arg Lys Val Thr Leu Gly Phe Asp Leu Gly Ile Ala

1 5 10 15
Ser Val Gly Trp Ala Ile Val Asp Ser Glu Thr Asn Gln Val Tyr Lys

20 25 30
Leu Gly Ser Arg Leu Phe Asp Ala Pro Asp Thr Asn Leu Glu Arg Arg

35 40 45
Thr Gln Arg Gly Thr Arg Arg Leu Leu Arg Arg Arg Lys Tyr Arg Asn

50 55 60
Gln Lys Phe Tyr Asn Leu Val Lys Arg Thr Glu Val Phe Gly Leu Ser

65 70 75 80
Ser Arg Glu Ala Ile Glu Asn Arg Phe Arg Glu Leu Ser Ile Lys Tyr

85 90 95
Pro Asn Ile Ile Glu Leu Lys Thr Lys Ala Leu Ser Gln Glu Val Cys

100 105 110
Pro Asp Glu Ile Ala Trp Ile Leu His Asp Tyr Leu Lys Asn Arg Gly

115 120 125

Tyr Phe Tyr Asp Glu Lys Glu Thr Lys Glu Asp Phe Asp Gln Gln Thr

130 135 140

Val Glu Ser Met Pro Ser Tyr Lys Leu Asn Glu Phe Tyr Lys Lys Tyr

145 150 155 160

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

165 170 175

Asp Asn Lys Asp Leu Lys Glu Ala Phe Phe Phe Asp Phe Ser Asn Lys

180 185 190

Glu Trp Leu Lys Glu Ile Asn Tyr Phe Phe Asn Val Gln Lys Asn Ile

195 200 205

Leu Ser Glu Thr Phe Ile Glu Glu Phe Lys Lys Ile Phe Ser Phe Thr

210 215 220

Arg Asp Ile Ser Lys Gly Pro Gly Ser Asp Asn Met Pro Ser Pro Tyr

225 230 235 240

Gly Ile Phe Gly Glu Phe Gly Asp Asn Gly Gln Gly Gly Arg Tyr Glu

245 250 255

His Ile Trp Asp Lys Asn Ile Gly Lys Cys Ser Ile Phe Thr Asn Glu

260 265 270

Gln Arg Ala Pro Lys Tyr Leu Pro Ser Ala Leu Ile Phe Asn Phe Leu

275 280 285

Asn Glu Leu Ala Asn Ile Arg Leu Tyr Ser Thr Asp Lys Lys Asn Ile

290 295 300

Gln Pro Leu Trp Lys Leu Ser Ser Ile Asp Lys Leu Asn Ile Leu Leu

305 310 315 320

Asn Leu Phe Asn Leu Pro Ile Ser Glu Lys Lys Lys Lys Leu Thr Ser

325 330 335

Thr Asn Ile Asn Asp Ile Val Lys Lys Glu Ser Ile Lys Ser Ile Met

340 345 350

Leu Ser Val Glu Asp Ile Asp Met Ile Lys Asp Glu Trp Ala Gly Lys

355 360 365

Glu Pro Asn Val Tyr Gly Val Gly Leu Ser Gly Leu Asn Ile Glu Glu

370 375 380
 Ser Ala Lys Glu Asn Lys Phe Lys Phe Gln Asp Leu Lys Ile Leu Asn

 385 390 395 400
 Val Leu Ile Asn Leu Leu Asp Asn Val Gly Ile Lys Phe Glu Phe Lys
 405 410 415
 Asp Arg Ser Asp Ile Ile Lys Asn Leu Glu Leu Leu Asp Asn Leu Tyr
 420 425 430
 Leu Phe Leu Ile Tyr Gln Lys Glu Ser Asn Asn Lys Asp Ser Ser Ile
 435 440 445
 Asp Leu Phe Ile Ala Lys Asn Lys Ser Leu Asn Ile Glu Asn Leu Lys

 450 455 460
 Leu Lys Leu Lys Glu Phe Leu Leu Gly Ala Gly Asn Glu Phe Glu Asn
 465 470 475 480
 His Asn Ser Lys Thr His Ser Leu Ser Lys Lys Ala Ile Asp Ala Ile
 485 490 495
 Leu Pro Lys Leu Leu Asp Asn Asn Glu Gly Trp Asn Leu Glu Ala Ile
 500 505 510
 Lys Asn Tyr Asp Glu Glu Ile Lys Ser Gln Ile Glu Asp Asn Ser Ser

 515 520 525
 Leu Met Ala Lys Gln Asp Lys Lys Tyr Leu Asn Asp Asn Phe Leu Lys
 530 535 540
 Asp Ala Ile Leu Pro Pro Asn Val Lys Val Thr Phe Gln Gln Ala Ile
 545 550 555 560
 Leu Ile Phe Asn Lys Ile Ile Gln Lys Phe Ser Lys Asp Phe Glu Ile
 565 570 575
 Asp Lys Val Val Ile Glu Leu Ala Arg Glu Met Thr Gln Asp Gln Glu

 580 585 590
 Asn Asp Ala Leu Lys Gly Ile Ala Lys Ala Gln Lys Ser Lys Lys Ser
 595 600 605
 Leu Val Glu Glu Arg Leu Glu Ala Asn Asn Ile Asp Lys Ser Val Phe
 610 615 620

Asn Asp Lys Tyr Glu Lys Leu Ile Tyr Lys Ile Phe Leu Trp Ile Ser
 625 630 635 640
 Gln Asp Phe Lys Asp Pro Tyr Thr Gly Ala Lys Ile Ser Ala Asn Glu

 645 650 655
 Ile Val Asp Asn Lys Val Glu Ile Asp His Ile Ile Pro Tyr Ser Leu
 660 665 670
 Cys Phe Asp Asp Ser Ser Ala Asn Lys Val Leu Val His Lys Gln Ser
 675 680 685
 Asn Gln Glu Lys Ser Asn Ser Leu Pro Tyr Glu Tyr Ile Lys Gln Gly
 690 695 700
 His Ser Gly Trp Asn Trp Asp Glu Phe Thr Lys Tyr Val Lys Arg Val

 705 710 715 720
 Phe Val Asn Asn Val Asp Ser Ile Leu Ser Lys Lys Glu Arg Leu Lys
 725 730 735
 Lys Ser Glu Asn Leu Leu Thr Thr Ser Tyr Asp Gly Tyr Glu Lys Leu
 740 745 750
 Gly Phe Leu Ala Arg Asn Leu Asn Asp Thr Arg Tyr Ala Thr Ile Leu
 755 760 765
 Phe Arg Asp Gln Leu Asn Asn Tyr Ala Glu His His Leu Ile Asp Asn

 770 775 780
 Lys Lys Met Phe Lys Val Ile Ala Met Asn Gly Ala Val Thr Ser Phe
 785 790 795 800
 Ile Arg Lys Asn Met Ser Tyr Asp Asn Lys Leu Arg Leu Lys Asp Arg
 805 810 815
 Ser Asp Phe Ser His His Ala Tyr Asp Ala Ala Ile Ile Ala Leu Phe
 820 825 830
 Ser Asn Lys Thr Lys Thr Leu Tyr Asn Leu Ile Asp Pro Ser Leu Asn

 835 840 845
 Gly Ile Ile Ser Lys Arg Ser Glu Gly Tyr Trp Val Ile Glu Asp Arg
 850 855 860
 Tyr Thr Gly Glu Ile Lys Glu Leu Lys Lys Glu Asp Trp Thr Ser Ile

865 870 875 880
 Lys Asn Asn Val Gln Ala Arg Lys Ile Ala Lys Glu Ile Glu Glu Tyr
 885 890 895
 Leu Ile Asp Leu Asp Asp Glu Val Phe Phe Ser Arg Lys Thr Lys Arg

 900 905 910
 Lys Thr Asn Arg Gln Leu Tyr Asn Glu Thr Ile Tyr Gly Ile Ala Ala
 915 920 925
 Lys Thr Asp Glu Asp Gly Ile Thr Asn Tyr Tyr Lys Lys Glu Lys Phe
 930 935 940
 Ser Ile Leu Asp Asp Lys Asp Ile Tyr Leu Arg Leu Leu Arg Glu Arg
 945 950 955 960
 Glu Lys Phe Val Ile Asn Gln Ser Asn Pro Glu Val Ile Asp Gln Ile

 965 970 975
 Ile Glu Ile Ile Glu Ser Tyr Gly Lys Glu Asn Asn Ile Pro Ser Arg
 980 985 990
 Asp Glu Ala Ile Asn Ile Lys Tyr Thr Lys Asn Lys Ile Asn Tyr Asn
 995 1000 1005
 Leu Tyr Leu Lys Gln Tyr Met Arg Ser Leu Thr Lys Ser Leu Asp
 1010 1015 1020
 Gln Phe Ser Glu Gly Phe Ile Asn Gln Met Ile Ala Asn Lys Thr

 1025 1030 1035
 Phe Val Leu Tyr Asn Pro Thr Lys Asn Thr Thr Arg Lys Ile Lys
 1040 1045 1050
 Phe Leu Arg Leu Val Asn Asp Val Lys Ile Asn Asp Ile Arg Lys
 1055 1060 1065
 Asn Gln Val Ile Asn Lys Phe Asn Gly Lys Asn Asn Glu Pro Lys
 1070 1075 1080
 Ala Phe Tyr Glu Asn Ile Asn Ser Leu Gly Ala Ile Val Phe Lys

 1085 1090 1095
 Ser Ser Ala Asn Asn Phe Lys Thr Leu Ser Ile Asn Thr Gln Ile
 1100 1105 1110

Ala Ile Phe Gly Asp Lys Asn Trp Asp Ile Glu Asp Phe Lys Thr
1115 1120 1125

Tyr Asn Met Glu Lys Ile Glu Lys Tyr Lys Glu Ile Tyr Gly Ile
1130 1135 1140

Asp Lys Thr Tyr Asn Phe His Ser Phe Ile Phe Pro Gly Thr Ile
1145 1150 1155

Leu Leu Asp Lys Gln Asn Lys Glu Phe Tyr Tyr Ile Ser Ser Ile
1160 1165 1170

Gln Thr Val Asn Asp Gln Ile Glu Leu Lys Phe Leu Asn Lys Ile
1175 1180 1185

Glu Phe Lys Asn Asp Asp Asn Thr Ser Gly Ala Asn Lys Pro Pro
1190 1195 1200

Arg Arg Leu Arg Phe Gly Ile Lys Ser Ile Met Asn Asn Tyr Glu

1205 1210 1215

Gln Val Asp Ile Ser Pro Phe Gly Ile Asn Lys Lys Ile Phe Glu
1220 1225 1230

Pro Lys Lys Lys Arg Lys Val Pro Lys Lys Lys Arg Lys Val
1235 1240 1245

<210> 9

<211> 3831

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 9

atgaacaaca gcatcaagag caagcccgaa gtgaccatcg gcctggatct cggcgttggc 60

tctgttgat gggccatcgt ggacaacgag acaaactca tccaccacct gggcagcaga 120

ctgttcagcc aggccaagac agctgaggac aggcggtctt tcagaggcgt gcggagactg 180

atccggcgga gaaagtacaa gctgaagaga ttcgtgaacc tgatctggaa gtacaacagc 240

tacttcggct tcaagaacaa agaggacatc ctgaacaact accaagagca gcagaaactg 300

cacaacaccg tgctgaacct gaagctcgaa gcctgaacg ccaagatcga cccaaggct 360

ctgagctgga ttctgcacga ctacctgaag aaccggggcc acttctacga ggacaaccgg 420
gacttcaacg tgtacccac agaggaactg gccactact tcgacgagtt cggctactac 480

aagggcacatca tcgacagcaa gaacgacgac gatgataagc tggaagaggg cctgaccaag 540
tacaagttca gcaaccagca ctggctggaa gaagtgaaga aggtcctgag caaccagacc 600
ggcctgcctg agaagttcaa agaggaatac gagagcctgt tcagctacgt gcggaactac 660
tctgaaggcc ctggcagcat caacagcgtg tccccatacg gcatctatca cctggacgag 720
aaagagggca aagtcgtcca gaagtataac aacatctggg acaagaccat cgggaagtgc 780
agcatcttcc ccgacgagta cagagccccct aagaacagcc ctatcgccat gatcttcaac 840
gagatcaacg agctgagcac catccggtcc tacagcatct acctgaccgg ctggttcac 900

aatcaagagt tcaagaaggc ctacctgaac aagctgctgg acctgctgat caagaccaac 960
agcgagaagc ccatcgacgc cgggcagttt aagaagctgc gggaagagac aatcgccgag 1020
agcatcggca aagaaacct gaaggacgtg gaaagcgagg aaaagctgga aaaggacgac 1080
cacaagtgga agctgaaggg cctgaagctg aacaccaacg gcaagatcca gtacaacgac 1140
ctgtctagcc tggccaagtt cgtgcacaaa ctgaagcagc acctgaaact ggactttctg 1200
ctggaagatc agtacacccc tctggacaag atcaacttcc tgcagagcct gtacgtgtac 1260
ctgggcaagc acctgagata cagcaacaga gtggacagcg ccaacctgaa agagttcagc 1320

gacagctccc ggctgttcga gagagtgtg caagagcaga aggacggcct gttcaagctg 1380
tttgagcaga ccgacaagga cgacgagaag atcctgacac agaccacag cctgtccacc 1440
aaggctatgc tctgtggcat caccagaatg accaacctgg acaatgacga ggataaccag 1500
aagaacaacg acaaaggctg gaacttcgag gccatcaaga acttcgacca gaagttcatc 1560
gacatcacca agacgaacaa caacctgagc ctgaagcagg acaagcgcta cctggatgac 1620
cagttcatca acgacgccat tctgagccct ggctgaaga gaatcctgag cgaggccacc 1680
aaggtgttca acgccatcct caagcagttc tccgaagagt acgacgtgac caaggtggtc 1740

atcgagctgg ccagagagct gagcgaagag aaagaactgg aaaacaccaa gaactacaag 1800
aagcttatca agaagaacgg cgataagatc agcgaggagc tgaaagccct ggggatcgcc 1860
gaggataaga tcgaagagat cctgaagtct cccaccaagt cctacaaagt gctgctgtgg 1920
ctgcagcagg accacatcga tcctacagc cagaaagaga tcgccttcga cgatatcctg 1980
acaaaaaccg aaaagaccga gatcgaccac atcattcctt actccatcag cttcgacgac 2040
agcagcagca acaaactgct ggtgctggcc gagtccaatc aggccaagtc caaccagaca 2100

ccctacgagt ttatcaactc cggcaaggcc gagatcacct gggaagtgtg cgaggcctac	2160
tgccacaagt tcaaaaacgg cgactccagc ctgctggaca gcaccagag aagcaagaaa	2220
ttcgccaaga tgatgaagac cgacaccagc tctaagtacg acatcggtt tctggcccgg	2280
aacctgaacg acaccagata cgccaccatc gtgttcggg acgctctgaa ggactacgcc	2340
aacaaccacc tgggtgaaga taagcccatg ttcaaggtcg tgtgcatcaa cggcggcgtg	2400
accagcttcc tgcggaagaa ctttgacccc aagtcttggc acgccaagaa ggacagagac	2460
aagaacattc accacgccgt ggacgccagc atcatctcca tcttcagcaa cgagactaag	2520
accctgttca accagctgac aaagtctgcc gactacaagc tgttcaagaa taccgacggc	2580
tcttgaaga agatcgatcc taagacaggc gtggtgtcag aagtaccga cgagaattgg	2640
aagcagatcc gcgtgcgcaa ccaggtgtcc gagatcgcca aagtatcga caagtacatc	2700
caggacagca acatcgagcg gaaggccaga tacagccgga agatcgagaa caagaccaat	2760
atcagcctgt ttaacgacac cgtgtactcc gccaagaaag tgggctacga ggatcagatc	2820
aagcgcaaga acctgaaaac cctggacatc cacgagagcg ccgaggaaaa caagaacagc	2880
aaagtgaaaa agcagttcgt gtaccggaag ctgctgaacg tgtccctgct gaacaatgac	2940
aagctggcgg acctgttcgc cgagaaagaa gatattctga tgtaccgggc caatccgtgg	3000
gtcatcaacc tggccgagca gattttcaac gactacaccg agaacaaaa gatcaagagc	3060
cagaactgtg tcgagaagta catgctggat ctgaccaaag agttccccga gaagtttagc	3120
gaggccttcg tgaagtccat gatcagaaac aagaccgcca tcatctaaa cgtcgagaag	3180
gatgtggtgc accggatcaa gcggctgaag attctgagca gcgagctgaa agaaaacaag	3240
tggccaacg tgatcatccg ctccaagaac gagagcggca ccaagctgag ctaccaggac	3300
accatcaact ctatgcctt gatgatcatg cggagcatcg acccaaccgc caaaaaacag	3360
tacatcaggg tgccctgaa caccctgaat ctgcacctgg gcgaccagga cttcgacctg	3420
cacaatatcg acgcctatct gaagaagcct aagtctgtca agtacctgaa ggccaatgag	3480
atcggcgacg agtataagcc ttggcgcgtg ctgacaagcg gcacactgct gatccacaag	3540
aaagacaaga aactcatgta catcagcagc ttccagaacc tcaacgacct catcgagatc	3600
aagaatctga tcgagacaga gtacaaagaa aacgtggact cagaccccaa gaagaagaaa	3660
aaggccagcc agatcctgag aagcctgagc gtgatcctga acgattacat cctgctggat	3720
gccaaagtata acttcgacat cctgggcctg tctaagaaca agattgacga gatcctcaac	3780
agcaagctgg acctcgacaa gattgccaag cccaagaaga agaggaaggt g	3831

<210> 10

<211> 1277

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 10

Met Asn Asn Ser Ile Lys Ser Lys Pro Glu Val Thr Ile Gly Leu Asp

1 5 10 15

Leu Gly Val Gly Ser Val Gly Trp Ala Ile Val Asp Asn Glu Thr Asn

20 25 30

Ile Ile His His Leu Gly Ser Arg Leu Phe Ser Gln Ala Lys Thr Ala

35 40 45

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

50 55 60

Lys Tyr Lys Leu Lys Arg Phe Val Asn Leu Ile Trp Lys Tyr Asn Ser

65 70 75 80

Tyr Phe Gly Phe Lys Asn Lys Glu Asp Ile Leu Asn Asn Tyr Gln Glu

85 90 95

Gln Gln Lys Leu His Asn Thr Val Leu Asn Leu Lys Leu Glu Ala Leu

100 105 110

Asn Ala Lys Ile Asp Pro Lys Ala Leu Ser Trp Ile Leu His Asp Tyr

115 120 125

Leu Lys Asn Arg Gly His Phe Tyr Glu Asp Asn Arg Asp Phe Asn Val

130 135 140

Tyr Pro Thr Glu Glu Leu Ala Asn Tyr Phe Asp Glu Phe Gly Tyr Tyr

145 150 155 160

Lys Gly Ile Ile Asp Ser Lys Asn Asp Asp Asp Asp Lys Leu Glu Glu

165 170 175

Gly Leu Thr Lys Tyr Lys Phe Ser Asn Gln His Trp Leu Glu Glu Val

180 185 190

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

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

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

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

Ser Ala Asn Leu Lys Glu Phe Ser Asp Ser Ser Arg Leu Phe Glu Arg

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

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

 Lys Val Phe Asn Ala Ile Leu Lys Gln Phe Ser Glu Glu Tyr Asp Val
 565 570 575
 Thr Lys Val Val Ile Glu Leu Ala Arg Glu Leu Ser Glu Glu Lys Glu
 580 585 590
 Leu Glu Asn Thr Lys Asn Tyr Lys Lys Leu Ile Lys Lys Asn Gly Asp
 595 600 605
 Lys Ile Ser Glu Gly Leu Lys Ala Leu Gly Ile Ala Glu Asp Lys Ile
 610 615 620

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

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

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

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

 Lys Gln Ile Arg Val Arg Asn Gln Val Ser Glu Ile Ala Lys Val Ile
 885 890 895
 Asp Lys Tyr Ile Gln Asp Ser Asn Ile Glu Arg Lys Ala Arg Tyr Ser
 900 905 910
 Arg Lys Ile Glu Asn Lys Thr Asn Ile Ser Leu Phe Asn Asp Thr Val
 915 920 925
 Tyr Ser Ala Lys Lys Val Gly Tyr Glu Asp Gln Ile Lys Arg Lys Asn

930	935	940	
Leu Lys Thr Leu Asp Ile His Glu Ser Ala Glu Glu Asn Lys Asn Ser			
945	950	955	960
Lys Val Lys Lys Gln Phe Val Tyr Arg Lys Leu Val Asn Val Ser Leu			
	965	970	975
Leu Asn Asn Asp Lys Leu Ala Asp Leu Phe Ala Glu Lys Glu Asp Ile			
	980	985	990
Leu Met Tyr Arg Ala Asn Pro Trp Val Ile Asn Leu Ala Glu Gln Ile			
	995	1000	1005
Phe Asn Glu Tyr Thr Glu Asn Lys Lys Ile Lys Ser Gln Asn Val			
1010	1015	1020	
Phe Glu Lys Tyr Met Leu Asp Leu Thr Lys Glu Phe Pro Glu Lys			
1025	1030	1035	
Phe Ser Glu Ala Phe Val Lys Ser Met Ile Arg Asn Lys Thr Ala			
1040	1045	1050	
Ile Ile Tyr Asn Val Glu Lys Asp Val Val His Arg Ile Lys Arg			
1055	1060	1065	
Leu Lys Ile Leu Ser Ser Glu Leu Lys Glu Asn Lys Trp Ser Asn			
1070	1075	1080	
Val Ile Ile Arg Ser Lys Asn Glu Ser Gly Thr Lys Leu Ser Tyr			
1085	1090	1095	
Gln Asp Thr Ile Asn Ser Ile Ala Leu Met Ile Met Arg Ser Ile			
1100	1105	1110	
Asp Pro Thr Ala Lys Lys Gln Tyr Ile Arg Val Pro Leu Asn Thr			
1115	1120	1125	
Leu Asn Leu His Leu Gly Asp Gln Asp Phe Asp Leu His Asn Ile			
1130	1135	1140	
Asp Ala Tyr Leu Lys Lys Pro Lys Phe Val Lys Tyr Leu Lys Ala			
1145	1150	1155	
Asn Glu Ile Gly Asp Glu Tyr Lys Pro Trp Arg Val Leu Thr Ser			
1160	1165	1170	

Gly Thr Leu Leu Ile His Lys Lys Asp Lys Lys Leu Met Tyr Ile
1175 1180 1185

Ser Ser Phe Gln Asn Leu Asn Asp Leu Ile Glu Ile Lys Asn Leu
1190 1195 1200

Ile Glu Thr Glu Tyr Lys Glu Asn Val Asp Ser Asp Pro Lys Lys
1205 1210 1215

Lys Lys Lys Ala Ser Gln Ile Leu Arg Ser Leu Ser Val Ile Leu
1220 1225 1230

Asn Asp Tyr Ile Leu Leu Asp Ala Lys Tyr Asn Phe Asp Ile Leu
1235 1240 1245

Gly Leu Ser Lys Asn Lys Ile Asp Glu Ile Leu Asn Ser Lys Leu
1250 1255 1260

Asp Leu Asp Lys Ile Ala Lys Pro Lys Lys Lys Arg Lys Val
1265 1270 1275

<210> 11

<211> 3378

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 11

atgcagaaca tcaccttcag cttcgacgtg ggctacgcct ctatcggatg ggctgttgtt 60

caggccccctg ctcagccaga gcaggaccct ggaatagtgg cctgtggcac cgtgctgttc 120

cctagcgcagc attgccaggc cttccagcgg agaaactacc ggcggtgcg gaggaacatc 180

cgggtccagaa gagtgcggat cgagcggatc ggaaagctgc tggttcaggc cggaatcctg 240

acactgagg aaaaggccac acctggacac cccgtccat tctttttggc agcccaggct 300

tggcagggca tcagacaact gtctccactg gaagtgtggc acatcctgcg ttggtacgcc 360

cacaacagag gctacgacaa caatgccgcc tgggccaccg tgtccaccaa agaggatacc 420

gagaaagtca acaacgcccc gcacctgatg cagaagtgtg gcgccgagac aatgtgcgcc 480

acactgtgcc atgctatgga actggacatg gacgtgcccc atgccgccat gacagtgtct 540

acaccagcct acagaaccct gaacagcgcc ttctctagag atgtggtgca gagagaggtg 600
ctggacatcc tgagacacag cgccagccac atcaaagagc tgaccctga gatcatccgg 660
ctgatcgccc agcaagagga tctgagcaca gagcagagaa gcgagctggc cgccaaggga 720
attagactgg ccagaagata ccggggcagc ctgctgtttg gacagctgct gccagattc 780
gacaaccgga tcatcagacg gtgccccatc atctgggccc acacatttga gcaggccaag 840
accagcggca tgagcgagaa agaagctcag gccctggctg acaagggtggc caaagtgcct 900
acagccgact gtcccagatt ctacgcctac agattcgccc gcatcctgaa caatctgaga 960

gccaacggac tgcccctgcc tgtggaagtt cgctgtgaac tgatgcaggc cgccagagcc 1020
gagggaaaac tgacagccgc caagatcaag aaagaaatca tgaggctgat gggcgacgtc 1080
gagagcaaca tcgaccacta cttccatctg caccgccaca gcgaggaagc cctgattctc 1140
gatcccgcta tggaatgcct gcaccggacc ggactgtacg atgccctcag ctctgtcgtg 1200
cgaagagtgg ccttgaccag actgcggaga ggcaaaatct gtaccctgc ctacctcgcg 1260
gacatgatgc tgagacacgg cgaggatacc caggctctgg atctggccat tgccaagcag 1320
caggggaagaa aggccccctg gcctagaaaag aacgacacag atgccagcgc cgacgccagc 1380

attgcatggc aagataagcc cctggctcct aagacagcct ctggcagagc cccttatgcc 1440
agaccagttc tgagacaggc cgtggacgag atcatgaatg gcgaggaccc taccaggcca 1500
gtcttggatg aacagcatcc cgacggcgag gacaagcctt ctacggctg tctgtatggc 1560
ctgctggacc ctgccagcaa agagaacgag tacctgaaca agcttcccct ggacgcctg 1620
acaaacaatc acctcgtgcg gcaccggatg ctgatcctgg atagactgac ccaggacctc 1680
gtcagagagt tcgtgacgg cgatcccagc agagtggaaac ggttctgtat cgaagtgggc 1740
agagagctga gcgccttctc tggcatgacc agcaagcaga tccagtccga gctgaacgag 1800

cggatgaagc acttcaagag cgccgtggcc tatctggcca aacacgcccc tgatatggcc 1860
acatctgccg gcctgatccg gaagtgcaga atcgctatgg acatgaactg gcagtgcctt 1920
ttcaccggcc agacctacat gccctacgac ctgcctaagc tggaacgcga gcacatcgtg 1980
ccctacgcca acagaaagac agatgccctg tctgccctgg tgctgacatg gctggccgtg 2040
aacaagatga agggcaagag aaccgcctac cagtttatca aagagtgcga gggccagagc 2100
gtgccggca gaaatcagaa tatcgtgtcc gtgaagcagt acgagacatt cgtggaaaag 2160
ctggacacca agggccacgc cgacgacgcc aagagaaaaa agacccggaa gaaactgatg 2220

atggtggaca gactgagcag ccagggaaca aacggcgagt ctgagctgga ttaccagag 2280
ggcatgatga cccagagcag ccacctgatg aagatcgccg ctgaggcgt gcggaagaac 2340
tttctcagc ccaccgtgga catgatccct ggcgctatta ctggcactgt gcgcaaggct 2400

tggaaggtgg caggatgcct ggccggcatt tgcctgaag ccgtcgatcc cgtgacacac 2460
agaatccagg acaaagagac actgcggcgg ctgacctatc tgcacatgc actggatgcc 2520
tgcgtgctgg gactgatccc tcacctgatt ccagagcaca gatccggcct gctgagaaaa 2580
gctctggccg ctagaaggct gcccgagaat gttcggcaag aggtggaaag cgccgtgtcc 2640

aagcgtact acaccatcac aaaagagagc aaactggaac tgcgggatct gcccaccaca 2700
ctgaagaact ctatgccgc caagctgagc gagggcagag tgggtcaaca catcctgcc 2760
gatatgagcg gagccaagct ggaagagaca atctggggaa ttgccctgg ccagcacatc 2820
gacgacaata gcgaggtggt catccggcag aagtcctga gcatcgcaa ggacggcaac 2880
agaatcagaa ccagaaagac cgacaagcag ggcaaccca tcaccagaa ggcctctaag 2940
ctcgtgggca tcaagcctac cggcaccagc aaactgcagc ccatcagagg cgtgatcatc 3000
atcaaggaca acttcgcat tgcctggac ccgtgcca ccatgattcc ccaccacaac 3060

gtgtacaagc ggctggaaga actgcggaag ctgaaccacg gtagacatgt gcggctgctg 3120
aaaaagggca tgctgatcag gctgagccac cagaagtccg gcgacaagaa cggcatgtgg 3180
aaagtgcgga gcatccagga ccagggtcc tctggcctga aagtgaatct gcagaggccc 3240
tactacgccg gcaagatcga ggacaccaga accgagaatt ggaagaacgt gtccatcaag 3300
gccctgctga gccaaggcat ggaaatcctg ccaaccacct actcggcac cacacctccc 3360
aagaagaaga ggaaggtg 3378

<210> 12

<211> 1126

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 12

Met Gln Asn Ile Thr Phe Ser Phe Asp Val Gly Tyr Ala Ser Ile Gly

1 5 10 15

Trp Ala Val Val Gln Ala Pro Ala Gln Pro Glu Gln Asp Pro Gly Ile

20 25 30

Val Ala Cys Gly Thr Val Leu Phe Pro Ser Asp Asp Cys Gln Ala Phe

35 40 45

Gln Arg Arg Asn Tyr Arg Arg Leu Arg Arg Asn Ile Arg Ser Arg Arg

50

55

60

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

65

70

75

80

Thr Pro Glu Glu Lys Ala Thr Pro Gly His Pro Ala Pro Phe Phe Leu

85

90

95

Ala Ala Gln Ala Trp Gln Gly Ile Arg Gln Leu Ser Pro Leu Glu Val

100

105

110

Trp His Ile Leu Arg Trp Tyr Ala His Asn Arg Gly Tyr Asp Asn Asn

115

120

125

Ala Ala Trp Ala Thr Val Ser Thr Lys Glu Asp Thr Glu Lys Val Asn

130

135

140

Asn Ala Arg His Leu Met Gln Lys Phe Gly Ala Glu Thr Met Cys Ala

145

150

155

160

Thr Leu Cys His Ala Met Glu Leu Asp Met Asp Val Pro Asp Ala Ala

165

170

175

Met Thr Val Ser Thr Pro Ala Tyr Arg Thr Leu Asn Ser Ala Phe Pro

180

185

190

Arg Asp Val Val Gln Arg Glu Val Leu Asp Ile Leu Arg His Ser Ala

195

200

205

Ser His Ile Lys Glu Leu Thr Pro Glu Ile Ile Arg Leu Ile Ala Gln

210

215

220

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

225

230

235

240

Ile Arg Leu Ala Arg Arg Tyr Arg Gly Ser Leu Leu Phe Gly Gln Leu

245

250

255

Leu Pro Arg Phe Asp Asn Arg Ile Ile Arg Arg Cys Pro Ile Ile Trp

260

265

270

Ala His Thr Phe Glu Gln Ala Lys Thr Ser Gly Met Ser Glu Lys Glu

275

280

285

Ala Gln Ala Leu Ala Asp Lys Val Ala Lys Val Pro Thr Ala Asp Cys

290 295 300
 Pro Glu Phe Tyr Ala Tyr Arg Phe Ala Arg Ile Leu Asn Asn Leu Arg

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

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

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

 500 505 510
 Pro Ser His Gly Cys Leu Tyr Gly Leu Leu Asp Pro Ala Ser Lys Glu
 515 520 525
 Asn Glu Tyr Leu Asn Lys Leu Pro Leu Asp Ala Leu Thr Asn Asn His
 530 535 540

Leu Val Arg His Arg Met Leu Ile Leu Asp Arg Leu Thr Gln Asp Leu
 545 550 555 560
 Val Arg Glu Phe Ala Asp Gly Asp Pro Ser Arg Val Glu Arg Phe Cys

 565 570 575
 Ile Glu Val Gly Arg Glu Leu Ser Ala Phe Ser Gly Met Thr Ser Lys
 580 585 590
 Gln Ile Gln Ser Glu Leu Asn Glu Arg Met Lys His Phe Lys Ser Ala
 595 600 605
 Val Ala Tyr Leu Ala Lys His Ala Pro Asp Met Ala Thr Ser Ala Gly
 610 615 620
 Leu Ile Arg Lys Cys Arg Ile Ala Met Asp Met Asn Trp Gln Cys Pro

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

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

 755 760 765
 Leu Met Lys Ile Ala Ala Arg Gly Val Arg Lys Asn Phe Pro His Ala
 770 775 780
 Thr Val Asp Met Ile Pro Gly Ala Ile Thr Gly Thr Val Arg Lys Ala

785 790 795 800
 Trp Lys Val Ala Gly Cys Leu Ala Gly Ile Cys Pro Glu Ala Val Asp
 805 810 815
 Pro Val Thr His Arg Ile Gln Asp Lys Glu Thr Leu Arg Arg Leu Thr

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

 885 890 895
 Leu Pro Thr Thr Leu Lys Asn Ser Ile Ala Ala Lys Leu Ser Glu Gly
 900 905 910
 Arg Val Val Gln His Ile Pro Ala Asp Met Ser Gly Ala Lys Leu Glu
 915 920 925
 Glu Thr Ile Trp Gly Ile Ala Pro Gly Gln His Ile Asp Asp Asn Ser
 930 935 940
 Glu Val Val Ile Arg Gln Lys Ser Leu Ser Ile Gly Lys Asp Gly Asn

 945 950 955 960
 Arg Ile Arg Thr Arg Lys Thr Asp Lys Gln Gly Asn Pro Ile Thr Glu
 965 970 975
 Lys Ala Ser Lys Leu Val Gly Ile Lys Pro Thr Gly Thr Ser Lys Leu
 980 985 990
 Gln Pro Ile Arg Gly Val Ile Ile Ile Lys Asp Asn Phe Ala Ile Ala
 995 1000 1005
 Leu Asp Pro Val Pro Thr Met Ile Pro His His Asn Val Tyr Lys

 1010 1015 1020
 Arg Leu Glu Glu Leu Arg Lys Leu Asn His Gly Arg His Val Arg
 1025 1030 1035

Leu Leu Lys Lys Gly Met Leu Ile Arg Leu Ser His Gln Lys Ser
1040 1045 1050

Gly Asp Lys Asn Gly Met Trp Lys Val Arg Ser Ile Gln Asp Gln
1055 1060 1065

Gly Ser Ser Gly Leu Lys Val Asn Leu Gln Arg Pro Tyr Tyr Ala
1070 1075 1080

Gly Lys Ile Glu Asp Thr Arg Thr Glu Asn Trp Lys Asn Val Ser
1085 1090 1095

Ile Lys Ala Leu Leu Ser Gln Gly Met Glu Ile Leu Pro Thr Thr
1100 1105 1110

Tyr Cys Gly Thr Thr Pro Pro Lys Lys Lys Arg Lys Val
1115 1120 1125

<210> 13

<211> 3324

<212> DNA

<213> Artificial Sequence

<220><

221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 13

atgagcagaa gcctgacctt cagcttcgac atcggctacg cctctatcgg ctgggccgtg 60
attgcctctg ccagccacga tgatgccgat cctagcgtgt gtggctgtgg caccgtgctg 120
ttccccaagg atgattgcca ggccttcaag cggagagagt accggcggct gcggagaaac 180
atccgggtcca gaagagtgcg gatcgagcgg attggttagac tgctggtgca ggcccagatc 240
atcacccctg agatgaagga aaccagcggga cccccgctc cattctacct ggcattctgag 300
gccctgaagg gccacagaac actggcccct attgaactgt ggcatgtgct gcgttggtac 360

gcccacaaca gaggtacga caacaacgcc agctggtcca acagcctgtc tgaggatggt 420
ggcaacggcg aggataccga gagagtgaac cagccccagg acctgatgga caagcacggc 480
acagctacaa tggccgagac aatctgcaga gagctgaagc tggaagaggg caaagccgac 540
gtcctatagg aagtgtctac cctgcctac aagaacctga acaccgctt tccacggctg 600
atcgtggaaa aagaagtgcg gagaatcctg gaactgagcg cccctctgat ccctggactg 660
acagccgaga tcatcgagct gatcgcccag catcacctc tgaccactga acagagaggc 720

gtgctgctcc agcacggcat taagctggcc agaagataca gaggcagcct gctgttcggc 780

cagctgatcc ctagattcga caacaggatc atcagcagat gccccgtgac atgggcccac 840

gtgtatgagg ccgagctgaa gaagggcaac agcgagcagt ctgccagaga gagagccgag 900

aagctgagca aggtgcccac cgccaattgt cccgagtctt acgagtaccg gatggccaga 960

atcctgtgca acatcagagc cgacggcgag cctctgagcg ccgagattag acgcgagctg 1020

atgaaccagg ccagacaaga gggaaagctg accaaggcca gcctggaaaa ggccatctct 1080

agccggctgg gcaaagaaac cgagacaaac gtgtccaact acttcacact gcaccccgac 1140

agcgaggaag cctgtatct gaatcctgcc gtggaagtgc tgcagagaag cggcatcggc 1200

cagattctga gccccagcgt gtacagaatc gccgccaaca gactgcggag aggcaagagc 1260

gtgacccta actacctgt gaatctgtg aagtcagag gcgagtctgg cgaggccctg 1320

gaaaaaaga tcgaaaaa gtccaagaag aaagaggccg actacgccga cacaccctg 1380

aagcctaagt acgccacagg cagagcccct tacgccaga ccgtgctgaa gaaagtgggtg 1440

gaagagatcc tggatggcga ggaccctacc agacctgcta gaggcgaagc tcaccctgac 1500

ggcgaactga aagcccacga tggctgcctg tactgcctgc tggataccga cagcagcgtg 1560

aaccagcacc agaaagagcg gagactggac accatgacca acaaccacct cgtgcggcac 1620

cggatgctga tcttgacag actcctgaag gatctgatcc aggacttcgc cgacggccag 1680

aaggacagaa tcagcagagt gtgctggaa gtgcgcaaag agctgaccac cttcagcgct 1740

atggacagca agaagatcca gcgggaactg accctgcggc agaagtctca taccgacgcc 1800

gtgaacagac tgaagagaaa gcttcaggc aaggccctga gcgccaacct gatcagaaaag 1860

tgcagaatcg caatggacat gaactggaca tgccccttca ccggcgccac atatggcgat 1920

cacgagctgg aaaaatctgga actggaacac atcgtgcccc acagcttcag acagagcaat 1980

gccctgtcta gcctggtgct gacatggcct ggcgtgaaca ggatgaaggg acagagaacc 2040

ggctacgact tcgtggaaca agagcaagag aaccccgctc ctgacaagcc caacctgcac 2100

atctgcagcc tgaacaacta tcgcgagctg gtggaaaagc tggacgacaa gaagggacac 2160

gaggacgaca gacggcgga gaagaaaaga aaggccctgc tgatgggtccg aggcctgtct 2220

cacaaacacc agagccagaa ccacgaggcc atgaaagaaa tcggcatgac cgagggcacg 2280

atgaccaga gcagccacct gatgaagctg gcctgcaaga gcatcaagac cagcctgcct 2340

gacgctcaca tcgacatgat tccaggcgcc gtgactgccg aagttcgcaa agcctgggat 2400

gtgttcggcg tgttcaaaga actgtgcccc gaagccgccg atcctgactc tggcaagatc 2460

ctgaaagaga acctgcggag cctgactcat ctgcatcacg ccctggatgc ctgtgtgctg 2520
ggactgatcc cctacatcat ccccgctcac cacaatggcc tgctgagaag agtcctggcc 2580
atgcgagaaa tccccgagaa actgatacct caagtgcggc ccgtggccaa ccagagacac 2640
tacgtgctga acgacgacgg ccggatgatg ctgagggatc tgagtgccag cctgaaagaa 2700
aacatccgcg agcagctgat ggaacagcga gtgatccagc acgtgcccgc tgatatgggc 2760
ggagcactgc tcaaagaaac aatgcagcgg gtgctgagcg tggacggctc tggcgaagat 2820
gctatggtgt ccctgtctaa gaagaaggac ggcaagaaag agaagaatca agtcaaggcc 2880

tccaagctcg tgggagtgtt tctgagggc cccagcaagc tgaaagctct gaaggccgcc 2940
atcgagatcg acggcaatta tggcgtggca ctggaccca agcctgtggt catcagacac 3000
atcaaggtgt tcaagaggat catggccctc aaagagcaga acggcggcaa gccagtgcgc 3060
atcctgaaaa agggcatgct gattcacctg accagcagca aggaccctaa gcacgtggc 3120
gtttggagaa tcgagagcat ccaggacagc aaaggcggcg tgaaactgga cctgcagagg 3180
gctcattgcg ccgtgcctaa gaacaagacc cagcagtga attggagaga ggtggacctg 3240
atctccctgc tgaaaaagta ccagatgaag cgctaccca ccagctacac cggcacacct 3300

agaccaaga agaagaggaa ggtg 3324

<210> 14

<211> 1108

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 14

Met Ser Arg Ser Leu Thr Phe Ser Phe Asp Ile Gly Tyr Ala Ser Ile

1 5 10 15

Gly Trp Ala Val Ile Ala Ser Ala Ser His Asp Asp Ala Asp Pro Ser

20 25 30

Val Cys Gly Cys Gly Thr Val Leu Phe Pro Lys Asp Asp Cys Gln Ala

35 40 45

Phe Lys Arg Arg Glu Tyr Arg Arg Leu Arg Arg Asn Ile Arg Ser Arg

50 55 60

Arg Val Arg Ile Glu Arg Ile Gly Arg Leu Leu Val Gln Ala Gln Ile

65 70 75 80
 Ile Thr Pro Glu Met Lys Glu Thr Ser Gly His Pro Ala Pro Phe Tyr
 85 90 95

 Leu Ala Ser Glu Ala Leu Lys Gly His Arg Thr Leu Ala Pro Ile Glu
 100 105 110
 Leu Trp His Val Leu Arg Trp Tyr Ala His Asn Arg Gly Tyr Asp Asn
 115 120 125
 Asn Ala Ser Trp Ser Asn Ser Leu Ser Glu Asp Gly Gly Asn Gly Glu
 130 135 140
 Asp Thr Glu Arg Val Lys His Ala Gln Asp Leu Met Asp Lys His Gly
 145 150 155 160

 Thr Ala Thr Met Ala Glu Thr Ile Cys Arg Glu Leu Lys Leu Glu Glu
 165 170 175
 Gly Lys Ala Asp Ala Pro Met Glu Val Ser Thr Pro Ala Tyr Lys Asn
 180 185 190
 Leu Asn Thr Ala Phe Pro Arg Leu Ile Val Glu Lys Glu Val Arg Arg
 195 200 205
 Ile Leu Glu Leu Ser Ala Pro Leu Ile Pro Gly Leu Thr Ala Glu Ile
 210 215 220

 Ile Glu Leu Ile Ala Gln His His Pro Leu Thr Thr Glu Gln Arg Gly
 225 230 235 240
 Val Leu Leu Gln His Gly Ile Lys Leu Ala Arg Arg Tyr Arg Gly Ser
 245 250 255
 Leu Leu Phe Gly Gln Leu Ile Pro Arg Phe Asp Asn Arg Ile Ile Ser
 260 265 270
 Arg Cys Pro Val Thr Trp Ala Gln Val Tyr Glu Ala Glu Leu Lys Lys
 275 280 285

 Gly Asn Ser Glu Gln Ser Ala Arg Glu Arg Ala Glu Lys Leu Ser Lys
 290 295 300
 Val Pro Thr Ala Asn Cys Pro Glu Phe Tyr Glu Tyr Arg Met Ala Arg
 305 310 315 320

Ile Leu Cys Asn Ile Arg Ala Asp Gly Glu Pro Leu Ser Ala Glu Ile
325 330 335

Arg Arg Glu Leu Met Asn Gln Ala Arg Gln Glu Gly Lys Leu Thr Lys
340 345 350

Ala Ser Leu Glu Lys Ala Ile Ser Ser Arg Leu Gly Lys Glu Thr Glu
355 360 365

Thr Asn Val Ser Asn Tyr Phe Thr Leu His Pro Asp Ser Glu Glu Ala
370 375 380

Leu Tyr Leu Asn Pro Ala Val Glu Val Leu Gln Arg Ser Gly Ile Gly
385 390 395 400

Gln Ile Leu Ser Pro Ser Val Tyr Arg Ile Ala Ala Asn Arg Leu Arg
405 410 415

Arg Gly Lys Ser Val Thr Pro Asn Tyr Leu Leu Asn Leu Leu Lys Ser
420 425 430

Arg Gly Glu Ser Gly Glu Ala Leu Glu Lys Lys Ile Glu Lys Glu Ser
435 440 445

Lys Lys Lys Glu Ala Asp Tyr Ala Asp Thr Pro Leu Lys Pro Lys Tyr
450 455 460

Ala Thr Gly Arg Ala Pro Tyr Ala Arg Thr Val Leu Lys Lys Val Val
465 470 475 480

Glu Glu Ile Leu Asp Gly Glu Asp Pro Thr Arg Pro Ala Arg Gly Glu
485 490 495

Ala His Pro Asp Gly Glu Leu Lys Ala His Asp Gly Cys Leu Tyr Cys
500 505 510

Leu Leu Asp Thr Asp Ser Ser Val Asn Gln His Gln Lys Glu Arg Arg
515 520 525

Leu Asp Thr Met Thr Asn Asn His Leu Val Arg His Arg Met Leu Ile
530 535 540

Leu Asp Arg Leu Leu Lys Asp Leu Ile Gln Asp Phe Ala Asp Gly Gln
545 550 555 560

Lys Asp Arg Ile Ser Arg Val Cys Val Glu Val Gly Lys Glu Leu Thr

565 570 575
 Thr Phe Ser Ala Met Asp Ser Lys Lys Ile Gln Arg Glu Leu Thr Leu
 580 585 590
 Arg Gln Lys Ser His Thr Asp Ala Val Asn Arg Leu Lys Arg Lys Leu
 595 600 605

 Pro Gly Lys Ala Leu Ser Ala Asn Leu Ile Arg Lys Cys Arg Ile Ala
 610 615 620
 Met Asp Met Asn Trp Thr Cys Pro Phe Thr Gly Ala Thr Tyr Gly Asp
 625 630 635 640
 His Glu Leu Glu Asn Leu Glu Leu Glu His Ile Val Pro His Ser Phe
 645 650 655
 Arg Gln Ser Asn Ala Leu Ser Ser Leu Val Leu Thr Trp Pro Gly Val
 660 665 670

 Asn Arg Met Lys Gly Gln Arg Thr Gly Tyr Asp Phe Val Glu Gln Glu
 675 680 685
 Gln Glu Asn Pro Val Pro Asp Lys Pro Asn Leu His Ile Cys Ser Leu
 690 695 700
 Asn Asn Tyr Arg Glu Leu Val Glu Lys Leu Asp Asp Lys Lys Gly His
 705 710 715 720
 Glu Asp Asp Arg Arg Arg Lys Lys Lys Arg Lys Ala Leu Leu Met Val
 725 730 735

 Arg Gly Leu Ser His Lys His Gln Ser Gln Asn His Glu Ala Met Lys
 740 745 750
 Glu Ile Gly Met Thr Glu Gly Met Met Thr Gln Ser Ser His Leu Met
 755 760 765
 Lys Leu Ala Cys Lys Ser Ile Lys Thr Ser Leu Pro Asp Ala His Ile
 770 775 780
 Asp Met Ile Pro Gly Ala Val Thr Ala Glu Val Arg Lys Ala Trp Asp
 785 790 795 800

 Val Phe Gly Val Phe Lys Glu Leu Cys Pro Glu Ala Ala Asp Pro Asp
 805 810 815

Ser Gly Lys Ile Leu Lys Glu Asn Leu Arg Ser Leu Thr His Leu His
820 825 830
His Ala Leu Asp Ala Cys Val Leu Gly Leu Ile Pro Tyr Ile Ile Pro
835 840 845
Ala His His Asn Gly Leu Leu Arg Arg Val Leu Ala Met Arg Arg Ile
850 855 860

Pro Glu Lys Leu Ile Pro Gln Val Arg Pro Val Ala Asn Gln Arg His
865 870 875 880
Tyr Val Leu Asn Asp Asp Gly Arg Met Met Leu Arg Asp Leu Ser Ala
885 890 895
Ser Leu Lys Glu Asn Ile Arg Glu Gln Leu Met Glu Gln Arg Val Ile
900 905 910
Gln His Val Pro Ala Asp Met Gly Gly Ala Leu Leu Lys Glu Thr Met
915 920 925

Gln Arg Val Leu Ser Val Asp Gly Ser Gly Glu Asp Ala Met Val Ser
930 935 940
Leu Ser Lys Lys Lys Asp Gly Lys Lys Glu Lys Asn Gln Val Lys Ala
945 950 955 960
Ser Lys Leu Val Gly Val Phe Pro Glu Gly Pro Ser Lys Leu Lys Ala
965 970 975
Leu Lys Ala Ala Ile Glu Ile Asp Gly Asn Tyr Gly Val Ala Leu Asp
980 985 990

Pro Lys Pro Val Val Ile Arg His Ile Lys Val Phe Lys Arg Ile Met
995 1000 1005
Ala Leu Lys Glu Gln Asn Gly Gly Lys Pro Val Arg Ile Leu Lys
1010 1015 1020
Lys Gly Met Leu Ile His Leu Thr Ser Ser Lys Asp Pro Lys His
1025 1030 1035
Ala Gly Val Trp Arg Ile Glu Ser Ile Gln Asp Ser Lys Gly Gly
1040 1045 1050

Val Lys Leu Asp Leu Gln Arg Ala His Cys Ala Val Pro Lys Asn

1055 1060 1065
Lys Thr His Glu Cys Asn Trp Arg Glu Val Asp Leu Ile Ser Leu
1070 1075 1080
Leu Lys Lys Tyr Gln Met Lys Arg Tyr Pro Thr Ser Tyr Thr Gly
1085 1090 1095
Thr Pro Arg Pro Lys Lys Lys Arg Lys Val
1100 1105
<210> 15
<211> 4188
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"
<400> 15
atggccagag attacagcgt cggcctggat atcggcacct cttctgttgg atgggccgcc 60
atcgacaaca agtaccacct gatcggggcc aagagcaaga acctgattgg cgtgcggctg 120
ttcgatagcg ccgtgaccgc cgagaagaga agaggctaca gaaccaccag acggcggctg 180
agcagacggc attggagact gagactgctg aacgacatct tcgccggacc tctgaccgat 240
ttcggcgacg agaatttctt ggccagactg aagtacagct gggttcaccc tcaagaccag 300
agcaatcagg ccacttttgc cgccggactg ctgttcgaca gcaaagagca ggacaaggac 360
ttctaccgga agtaccctac catctatcac ctgagactgg ccctgatgaa cgacgaccag 420
aagcacgacc tgagagaggt gtacctggcc atccaccacc tggtaagta cagaggccac 480
ttcctgatcg agggcgacgt gaaagccgac agcgcttttg atgtgcacac cttcgccgac 540
gccatccaga gatacgccga gagcaacaac tccgacgaga acctgctggg caagatcgac 600
gagaagaagc tgagcgctgc cctgaccgat aagcacggca gcaaaagcca gagagccgag 660
acagccgaaa ccgccttcga catcctggac ctgcagtcca agaagcagat ccaggccatc 720
ctgaagtccg tcgtgggcaa ccaggccaat ctgatggcca tttttggcct ggacagcagc 780
gccatcagca aggacgagca gaagaactac aagttcagct tcgacgacgc cgacatcgat 840
gagaagatcg ccgattctga ggccctgctg agcgataccg agttcgagtt cctgtgcgat 900
ctgaaggccg cctttgacgg cctgacactg aaaatgctgc tgggcgacga caagaccgtg 960
tccgctgcta tggttcgacg gttcaacgag caccagaagg actgggagta catcaagagc 1020

cacatccgga acgccaagaa cgccggcaat ggctgtacg agaagtctaa gaagttcgac	1080
ggcatcaacg cgcctatct ggctctgcag tccgacaacg aggacgacag aaagaaggcc	1140
aagaagattt tccaggacga gatcagctcc gccgacattc ccgatgatgt gaaggccgat	1200
ttcctgaaga agattgacga cgatcagttc ctgcctatcc agcggaccaa gaacaacggc	1260
acaatccctc accagctgca ccggaacgag ctggaacaga tcatcgagaa gcaggggatc	1320
tactacccat tectgaagga cacctaccaaa gagaacagcc acgagctgaa caaaatcaca	1380
gccctgatca acttcagggt gccctactac gtgggccctc tgggtgaaga ggaacagaaa	1440
atcgccgacg acggcaagaa catccccgat cctaccaacc actggatggg ccgaaagtc	1500
aacgacacca tcacacctg gaacctgagc cagggtggtc acctggataa gagcggcaga	1560
agattcatcg agcggctgac cggcaccgat acctatctga tcggagagcc cacactgccc	1620
aagaacagcc tgtgtacca gaaattcgac gtgctgcaag aactgaacaa catccgctg	1680
tccggcagac ggctggacat tagagccaag caggatgcct tcgagcacct gttcaagggtg	1740
cagaaaaccg tgtctgtac caatctgaag gacttcctgg tgcaagccgg ctacatcagc	1800
gaggacaccc agattgaagg actcgccgac gtgaacggaa agaacttcaa caacgccctg	1860
accacctaca actacctggt gtctgtgctg ggccgcgagt tcgtggaaaa cccagcaac	1920
gaggaactgc tggaagagat taccgagctg cagaccgtgt tcgaggacaa gaaggtgctg	1980
cggagacagc tggatcagct ggacggactg agcgaccaca acagagagaa gctttcccgg	2040
aagcactaca ccggctgggg cagaatcagc aagaagctgc tgaccaccaa gatcgtgcag	2100
aacgccgaca agatcgataa ccagaccttc gatgtgcccc ggatgaacca gagcatcatc	2160
gacacctgt acaacaccaa gatgaacctg atggaaatca tcaacaatgc cgaggatgac	2220
ttcggcgtca gagcctggat cgacaagcag aacaccaccg atggcgacga gcaggacgtg	2280
tacagcctga tcgatgaact ggctggcccc aaagagatca agcggggcat cgtgcagtec	2340
tttagaatcc tggacgacat caccaaggcc gtgggctacg ccctaaccg ggtgtacctc	2400
gaatttgcca gaaagaccca agagagccac ctgaccaaca gccggaagaa ccagctgagc	2460
acctgctga agaatgccgg cctgtctgag ctggtcacac aggtgtccca gtatgatgcc	2520
gccgctctgc agaacgaccg gctgtatctt tacttcctgc agcaaggcaa ggacatgtac	2580
tccggcgaga agctgaatct ggacaacctg agcaactacg acatcgacca catcatccct	2640
caggcttaca ccaaggacaa cagcctggac aacagagtgc tgggtgccaa tatcaccaac	2700
cggcggaagt ccgacagcag caactatctg cccgctctga tcgataagat gcggcccttt	2760

tggagcgtgc tgagcaagca ggggctgctg tctaagcaca agttcgccaa cctgaccaga 2820

accagagact tcgacgatat ggaaaaagag cggtttatcg cccgcagcct ggtggaacc 2880

cggcagatca ttaagaacgt ggccagcctg attgacagcc acttcggcgg agagacaaaa 2940

gccgtggcca ttagaagcag cctgacagcc gacatgcgga gatacgtgga catccccaag 3000

aaccgggaca tcaacgacta ccaccagcc ttcatgccc tgcgttttag cacagtgggc 3060

cagtacaccg agaacagcgg cctgatgaag aaggccagc tgtccgattc tgccgcaac 3120

cagtacaatc ggtacatcaa agagtggatt cacgccca ggctgaacgc acagtcccag 3180

agagtgaacc cttcggctt tgcgtgggc tccatgagaa atgctgccc tggcaagctg 3240

aaccccgaga caggggagat caccacagag gaaaacgccg actggtctat cgccgacctg 3300

gactacctgc acaaagtgat gaatttcgg aagatcaccg tgaccaggcg gctgaaggat 3360

cagaaaggac agctgtacga cgagagcaga taccctccg tgcgtcacga cgccaagtct 3420

aaggccagca tcaacttga caagcacaag cccgtggacc tgtacggcgg ctttagctct 3480

gccaaagcctg cctatgccgc actgatcaag ttcaagaaca agttccggct ggtcaacgtg 3540

ctgcggcagt ggacctacg cgacaagaac tccgaggact atatccttga gcagatcaga 3600

ggcaagtacc ctaaggccga gatggtgctg tctcacatcc cttacggcca gctggtcaag 3660

aaagatggcg ccttggtcac catctctagc gccacagagc tgcacaactt tgagcagctg 3720

tggctgcctc tggccgacta caagctgac aacacactgc ttaagaccaa agaggacaac 3780

ctcgtcgata tctgcacaa ccggctggat ctccccgaga tgacaatcga gagcgcttc 3840

tacaaagcct tcgactccat cctgagcttc gccttcaaca gatacgccct gcaccagaac 3900

gccctcgtga aactgcaggc ccacaggac gatttcaatg cctgaacta cgaggataag 3960

cagcagacc tggaaggat tctggacgct ctgcatgcct ctccagccag cagcgacctg 4020

aagaaaatca acctgtccag cggcttcggc cggtgtttt ccctagcca cttaccctg 4080

gccgacaccg acgagttcat cttccagagc gtgaccggcc tgttcagcac ccagaaaaca 4140

gtggctcagc tgtatcaaga gacaaagccc aagaagaaga ggaaggtg 4188

<210> 16

<211> 1396

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic polypeptide"

<400> 16

Met Ala Arg Asp Tyr Ser Val Gly Leu Asp Ile Gly Thr Ser Ser Val

1 5 10 15

Gly Trp Ala Ala Ile Asp Asn Lys Tyr His Leu Ile Arg Ala Lys Ser

20 25 30

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

35 40 45

Lys Arg Arg Gly Tyr Arg Thr Thr Arg Arg Arg Leu Ser Arg Arg His

50 55 60

Trp Arg Leu Arg Leu Leu Asn Asp Ile Phe Ala Gly Pro Leu Thr Asp

65 70 75 80

Phe Gly Asp Glu Asn Phe Leu Ala Arg Leu Lys Tyr Ser Trp Val His

85 90 95

Pro Gln Asp Gln Ser Asn Gln Ala His Phe Ala Ala Gly Leu Leu Phe

100 105 110

Asp Ser Lys Glu Gln Asp Lys Asp Phe Tyr Arg Lys Tyr Pro Thr Ile

115 120 125

Tyr His Leu Arg Leu Ala Leu Met Asn Asp Asp Gln Lys His Asp Leu

130 135 140

Arg Glu Val Tyr Leu Ala Ile His His Leu Val Lys Tyr Arg Gly His

145 150 155 160

Phe Leu Ile Glu Gly Asp Val Lys Ala Asp Ser Ala Phe Asp Val His

165 170 175

Thr Phe Ala Asp Ala Ile Gln Arg Tyr Ala Glu Ser Asn Asn Ser Asp

180 185 190

Glu Asn Leu Leu Gly Lys Ile Asp Glu Lys Lys Leu Ser Ala Ala Leu

195 200 205

Thr Asp Lys His Gly Ser Lys Ser Gln Arg Ala Glu Thr Ala Glu Thr

210 215 220

Ala Phe Asp Ile Leu Asp Leu Gln Ser Lys Lys Gln Ile Gln Ala Ile

225 230 235 240

Leu Lys Ser Val Val Gly Asn Gln Ala Asn Leu Met Ala Ile Phe Gly
 245 250 255
 Leu Asp Ser Ser Ala Ile Ser Lys Asp Glu Gln Lys Asn Tyr Lys Phe
 260 265 270
 Ser Phe Asp Asp Ala Asp Ile Asp Glu Lys Ile Ala Asp Ser Glu Ala
 275 280 285
 Leu Leu Ser Asp Thr Glu Phe Glu Phe Leu Cys Asp Leu Lys Ala Ala
 290 295 300
 Phe Asp Gly Leu Thr Leu Lys Met Leu Leu Gly Asp Asp Lys Thr Val
 305 310 315 320
 Ser Ala Ala Met Val Arg Arg Phe Asn Glu His Gln Lys Asp Trp Glu
 325 330 335
 Tyr Ile Lys Ser His Ile Arg Asn Ala Lys Asn Ala Gly Asn Gly Leu
 340 345 350
 Tyr Glu Lys Ser Lys Lys Phe Asp Gly Ile Asn Ala Ala Tyr Leu Ala
 355 360 365
 Leu Gln Ser Asp Asn Glu Asp Asp Arg Lys Lys Ala Lys Lys Ile Phe
 370 375 380
 Gln Asp Glu Ile Ser Ser Ala Asp Ile Pro Asp Asp Val Lys Ala Asp
 385 390 395 400
 Phe Leu Lys Lys Ile Asp Asp Asp Gln Phe Leu Pro Ile Gln Arg Thr
 405 410 415
 Lys Asn Asn Gly Thr Ile Pro His Gln Leu His Arg Asn Glu Leu Glu
 420 425 430
 Gln Ile Ile Glu Lys Gln Gly Ile Tyr Tyr Pro Phe Leu Lys Asp Thr
 435 440 445
 Tyr Gln Glu Asn Ser His Glu Leu Asn Lys Ile Thr Ala Leu Ile Asn
 450 455 460
 Phe Arg Val Pro Tyr Tyr Val Gly Pro Leu Val Glu Glu Glu Gln Lys
 465 470 475 480
 Ile Ala Asp Asp Gly Lys Asn Ile Pro Asp Pro Thr Asn His Trp Met

485 490 495
 Val Arg Lys Ser Asn Asp Thr Ile Thr Pro Trp Asn Leu Ser Gln Val
 500 505 510
 Val Asp Leu Asp Lys Ser Gly Arg Arg Phe Ile Glu Arg Leu Thr Gly
 515 520 525
 Thr Asp Thr Tyr Leu Ile Gly Glu Pro Thr Leu Pro Lys Asn Ser Leu

 530 535 540
 Leu Tyr Gln Lys Phe Asp Val Leu Gln Glu Leu Asn Asn Ile Arg Val
 545 550 555 560
 Ser Gly Arg Arg Leu Asp Ile Arg Ala Lys Gln Asp Ala Phe Glu His
 565 570 575
 Leu Phe Lys Val Gln Lys Thr Val Ser Ala Thr Asn Leu Lys Asp Phe
 580 585 590
 Leu Val Gln Ala Gly Tyr Ile Ser Glu Asp Thr Gln Ile Glu Gly Leu

 595 600 605
 Ala Asp Val Asn Gly Lys Asn Phe Asn Asn Ala Leu Thr Thr Tyr Asn
 610 615 620
 Tyr Leu Val Ser Val Leu Gly Arg Glu Phe Val Glu Asn Pro Ser Asn
 625 630 635 640
 Glu Glu Leu Leu Glu Glu Ile Thr Glu Leu Gln Thr Val Phe Glu Asp
 645 650 655
 Lys Lys Val Leu Arg Arg Gln Leu Asp Gln Leu Asp Gly Leu Ser Asp

 660 665 670
 His Asn Arg Glu Lys Leu Ser Arg Lys His Tyr Thr Gly Trp Gly Arg
 675 680 685
 Ile Ser Lys Lys Leu Leu Thr Thr Lys Ile Val Gln Asn Ala Asp Lys
 690 695 700
 Ile Asp Asn Gln Thr Phe Asp Val Pro Arg Met Asn Gln Ser Ile Ile
 705 710 715 720
 Asp Thr Leu Tyr Asn Thr Lys Met Asn Leu Met Glu Ile Ile Asn Asn

 725 730 735

Ala Glu Asp Asp Phe Gly Val Arg Ala Trp Ile Asp Lys Gln Asn Thr
740 745 750

Thr Asp Gly Asp Glu Gln Asp Val Tyr Ser Leu Ile Asp Glu Leu Ala
755 760 765

Gly Pro Lys Glu Ile Lys Arg Gly Ile Val Gln Ser Phe Arg Ile Leu
770 775 780

Asp Asp Ile Thr Lys Ala Val Gly Tyr Ala Pro Lys Arg Val Tyr Leu

785 790 795 800

Glu Phe Ala Arg Lys Thr Gln Glu Ser His Leu Thr Asn Ser Arg Lys
805 810 815

Asn Gln Leu Ser Thr Leu Leu Lys Asn Ala Gly Leu Ser Glu Leu Val
820 825 830

Thr Gln Val Ser Gln Tyr Asp Ala Ala Ala Leu Gln Asn Asp Arg Leu
835 840 845

Tyr Leu Tyr Phe Leu Gln Gln Gly Lys Asp Met Tyr Ser Gly Glu Lys

850 855 860

Leu Asn Leu Asp Asn Leu Ser Asn Tyr Asp Ile Asp His Ile Ile Pro
865 870 875 880

Gln Ala Tyr Thr Lys Asp Asn Ser Leu Asp Asn Arg Val Leu Val Ser
885 890 895

Asn Ile Thr Asn Arg Arg Lys Ser Asp Ser Ser Asn Tyr Leu Pro Ala
900 905 910

Leu Ile Asp Lys Met Arg Pro Phe Trp Ser Val Leu Ser Lys Gln Gly

915 920 925

Leu Leu Ser Lys His Lys Phe Ala Asn Leu Thr Arg Thr Arg Asp Phe
930 935 940

Asp Asp Met Glu Lys Glu Arg Phe Ile Ala Arg Ser Leu Val Glu Thr
945 950 955 960

Arg Gln Ile Ile Lys Asn Val Ala Ser Leu Ile Asp Ser His Phe Gly
965 970 975

Gly Glu Thr Lys Ala Val Ala Ile Arg Ser Ser Leu Thr Ala Asp Met

980	985	990
Arg Arg Tyr Val Asp Ile Pro Lys	Asn Arg Asp Ile Asn	Asp Tyr His
995	1000	1005
His Ala Phe Asp Ala Leu Leu	Phe Ser Thr Val Gly	Gln Tyr Thr
1010	1015	1020
Glu Asn Ser Gly Leu Met Lys	Lys Gly Gln Leu Ser	Asp Ser Ala
1025	1030	1035
Gly Asn Gln Tyr Asn Arg Tyr	Ile Lys Glu Trp Ile	His Ala Ala
1040	1045	1050
Arg Leu Asn Ala Gln Ser Gln	Arg Val Asn Pro Phe	Gly Phe Val
1055	1060	1065
Val Gly Ser Met Arg Asn Ala	Ala Pro Gly Lys Leu	Asn Pro Glu
1070	1075	1080
Thr Gly Glu Ile Thr Pro Glu	Glu Asn Ala Asp Trp	Ser Ile Ala
1085	1090	1095
Asp Leu Asp Tyr Leu His Lys	Val Met Asn Phe Arg	Lys Ile Thr
1100	1105	1110
Val Thr Arg Arg Leu Lys Asp	Gln Lys Gly Gln Leu	Tyr Asp Glu
1115	1120	1125
Ser Arg Tyr Pro Ser Val Leu	His Asp Ala Lys Ser	Lys Ala Ser
1130	1135	1140
Ile Asn Phe Asp Lys His Lys	Pro Val Asp Leu Tyr	Gly Gly Phe
1145	1150	1155
Ser Ser Ala Lys Pro Ala Tyr	Ala Ala Leu Ile Lys	Phe Lys Asn
1160	1165	1170
Lys Phe Arg Leu Val Asn Val	Leu Arg Gln Trp Thr	Tyr Ser Asp
1175	1180	1185
Lys Asn Ser Glu Asp Tyr Ile	Leu Glu Gln Ile Arg	Gly Lys Tyr
1190	1195	1200
Pro Lys Ala Glu Met Val Leu	Ser His Ile Pro Tyr	Gly Gln Leu
1205	1210	1215

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

1220 1225 1230

Leu His Asn Phe Glu Gln Leu Trp Leu Pro Leu Ala Asp Tyr Lys

1235 1240 1245

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

1250 1255 1260

Ile Leu His Asn Arg Leu Asp Leu Pro Glu Met Thr Ile Glu Ser

1265 1270 1275

Ala Phe Tyr Lys Ala Phe Asp Ser Ile Leu Ser Phe Ala Phe Asn

1280 1285 1290

Arg Tyr Ala Leu His Gln Asn Ala Leu Val Lys Leu Gln Ala His

1295 1300 1305

Arg Asp Asp Phe Asn Ala Leu Asn Tyr Glu Asp Lys Gln Gln Thr

1310 1315 1320

Leu Glu Arg Ile Leu Asp Ala Leu His Ala Ser Pro Ala Ser Ser

1325 1330 1335

Asp Leu Lys Lys Ile Asn Leu Ser Ser Gly Phe Gly Arg Leu Phe

1340 1345 1350

Ser Pro Ser His Phe Thr Leu Ala Asp Thr Asp Glu Phe Ile Phe

1355 1360 1365

Gln Ser Val Thr Gly Leu Phe Ser Thr Gln Lys Thr Val Ala Gln

1370 1375 1380

Leu Tyr Gln Glu Thr Lys Pro Lys Lys Lys Arg Lys Val

1385 1390 1395

<210> 17

<211> 3738

<212> DNA

<213> Artificial Sequence

<220><

221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 17

atgagccagc accggcggtg tagaatcggc atcgacgtgg gcctgaatag cgttgactg	60
gccgccgtgg aaatcgacgc caaccacgac aatcctctgg acgagatccc catcagcatc	120
ctgaatgccc agagcgtgat ccacgatggc ggagtggacc ctgatgaggc caagtctgct	180
acaagcagac gggcttctgc tggcgtggcc agaagaacaa gacggctgca caagagcaag	240
cggcagagac tggccaagct ggacgaggtg ctgaatgagc tgggtaccc cgtggaagat	300
gagagccagt ttccagccgg cagcaacccc tatatcgctt ggcaagtgcg ggccaaactg	360
gccgagacat tcatccccga cgtggaaacc cggaagcgga tgatctctat cgccatccgg	420
cacattgccc ggcatagagg atggcggaat cctactctt ctgtggccga cgccgagcgg	480
atgagccata caccttctcc attcatggtg gaatacgcca agaagctgga cttcgagatc	540
aacgacagac ggaccaacgg cttctatcac agcccttggc agagcgtgga cgaggaaggc	600
aagagactga gcaagagcga gctggaaaag cagcccaaga tcgaggactg gaacgacaac	660
cccatcaacg gcaagacaat cgcccagctg gtcgtgtcct ctctggaacc ccagaccaag	720
atcagacggg atctgacaca cggcctgcag accgagagca ccctgaatat ccagacagag	780
aagctgcacc agagcgacta catccacgaa ctggaaacca tcttcgagcg gcagcacgtg	840
gaccagacaa cccaagaaca gctgctggaa gccaccttcc acaccaagaa tcctaaggcc	900
gtgggagccg ccgctaagct cgttggaaaa gatgccctgg acagccggtg ctacagagcc	960
agcagagcca caccagcctt cgaagagtac agagtgatgg ccgccatcga caccctgcgg	1020
attagagagc acggcaccga gagacagctg accaccgacg agagaagaaa gctgttcgac	1080
ttcatcaagg ggctgcccag caaaaagacc aagaacgagc ccagcatcag ctccctgacc	1140
tggggagatg tggccgattt tctgggcatc cagcggatcg atctgagagg cctgggctct	1200
ctgaaagacg gcgaacctgt gtctgccaag cagcctcctg tgatcgagac aaacgacatc	1260
atgcagaagg cccctgatcc aatcgctgcc tgggtgtcac aggccaacac caaagaacgg	1320
gacagattcg tcgagttcat gagcaacgct ggcgccatca aggacacctc cgacgaagtg	1380
cggaacattg acgccgagat cagccagctg ctggaagaac tgaccggctc tgagctggaa	1440
tccttgata agatcacctt gaccttggc agagccgcct acagctctca gacctgaga	1500
aacatcacca actatatgta cgagacaggc tgcgacctga ccacagccag acaagagctg	1560
taccagtggt gcaagaattg ggccctcctt gctcctccta tctacgagca cacaggcaac	1620
cccagcgtgg acagaacctt cagcatcatc cacagatggc tgtgcaacat gcgggaccag	1680
tacggcgagc ccgagacagt gaatatcgag tacgtccgag acggttcag cagcacatct	1740
acacagctgg ccgagcagcg cgagcgggat agaagatagc ccgacaacct gaagatgctg	1800

agcaactacg agggcgccag cagcagatca gatgtgcgga gaatcaaggc cctgcagaga 1860
 cagaactgcc agtgcattcta ctgcggccgg accatcacct tcgagacatg ccagatggac 1920
 catgtgctgc cccggaaagg ccttgatcc gatagcaagt tcgagaacct ggtggccaca 1980
 tgcggcgagt gcaacaagtc caagagcgat accctgtaca tgaactgggc caagacatac 2040

 cccaatacca acctgcagga cgtgctgaga agaattcaag agtgggtcaa ggacggctgg 2100
 atgaccgaca aaagatggcg gcagtacaaa gaggccctga tctgagact ggaagctacc 2160
 gagaagcaag agccccctga caatcggagc atggaaagcg tgtcctacat ggccagagag 2220
 ctgcggaacc ggatctacgg cttttacggc tggcacgacc aggacgacgc cctgaaacaa 2280
 ggacagacaga ggggtgttct gtccagcggc agtatgacag ccgctgccag aaggaccctt 2340
 ttcgagtcct cactgattaa gggcgccgat gaggaaacct acgagagcag cctgccttgg 2400
 ctggatggca tgaagggcaa gaccagactg gatcggagac accatgccgt ggacgccagc 2460

 atcattgcca tgatgagggc ccagatcgtg aagatcctga cagaggccca agagatcaga 2520
 agcgagcagc acgacaagta ccggaagggc cagacacctg actacgtgtg caagcggcgg 2580
 gactactggc ggaattggag aggcacccct gacaccagag atgaggaagt gttcaactac 2640
 tgggctgggg agcagctgag aacctgacc gatctggtgt ccagaagat ggccgacgac 2700
 gaaatccccg tgatctaccc caccagactg agactcggca atggcagcgc ccacaaggat 2760
 accgtggtgt ccatgatgac ccggaaagtg ggcgacgagc tgagcatcac cgccatcaac 2820
 aaagccgaaa gcggagccct gtacacagcc ctgaccagag acagcgactt cgactggaaa 2880

 accggcctga gcgccaatcc taaccggcgg atcagagtgc acgataagtg gttcagggcc 2940
 gacgatacca tcaagtttct ggaacctgcc gtggaagtgg tgctgaagaa caacaccaga 3000
 gccagaatcg accccgagcg tctggataag gtgcacagca cactgtacgt gcccgtcaga 3060
 ggcggaatcg ccgaagccgg aaatagcatt caccacgtgc ggttttacia gatcccaag 3120
 ctgaacagca agggcaagca gaccggcagc atctacgcca tgctgagagt gctgaccatc 3180
 gacctggcca tgaaccagta cgacaaagag acaggcaaga agcaggacct gttcacctg 3240
 ccactgcctg aaagcagcct gagcagaaga ttcagcgagc ccaaactgcg gcaggtctg 3300

 atcgatggca cagccgaata tctcgatgg gccgtcgtgg acgatgagct tgagatcccc 3360
 gccttcgcca acgccgaat cacagaggaa caggccatta acggcagctt caccgacaga 3420
 ctgctgcaca gctttcccg caccacaaag ttcagattcg ccggtttctc ccggaacacc 3480
 gagatcgcca ttagacctgt gcagctggcc tctgagggcc tgatcgaaac cgatgagaac 3540
 cggaagagac agcagctcgc gctgaccag ctaacaccg agtacagca cagcatcaag 3600
 aacgtgctga agtccggcct gcacctgaaa gtgaacaccc tgttcagac aggcattctg 3660

gtcaccaggg ccaatagcca gggaaagcag agcatccggt tcagcacagt ggaagagccc 3720

aagaagaaga ggaaggtg 3738

<210> 18

<211> 1246

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 18

Met Ser Gln His Arg Arg Tyr Arg Ile Gly Ile Asp Val Gly Leu Asn

1 5 10 15

Ser Val Gly Leu Ala Ala Val Glu Ile Asp Ala Asn His Asp Asn Pro

20 25 30

Leu Asp Glu Ile Pro Ile Ser Ile Leu Asn Ala Gln Ser Val Ile His

35 40 45

Asp Gly Gly Val Asp Pro Asp Glu Ala Lys Ser Ala Thr Ser Arg Arg

50 55 60

Ala Ser Ala Gly Val Ala Arg Arg Thr Arg Arg Leu His Lys Ser Lys

65 70 75 80

Arg Gln Arg Leu Ala Lys Leu Asp Glu Val Leu Asn Glu Leu Gly Tyr

85 90 95

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

100 105 110

Ala Trp Gln Val Arg Ala Lys Leu Ala Glu Thr Phe Ile Pro Asp Val

115 120 125

Glu Thr Arg Lys Arg Met Ile Ser Ile Ala Ile Arg His Ile Ala Arg

130 135 140

His Arg Gly Trp Arg Asn Pro Tyr Ser Ser Val Ala Asp Ala Glu Arg

145 150 155 160

Met Ser His Thr Pro Ser Pro Phe Met Val Glu Tyr Ala Lys Lys Leu

165	170	175
Asp Phe Glu Ile Asn Asp Arg Arg Thr Asn Gly Phe Tyr His Ser Pro		
180	185	190
Trp Gln Ser Val Asp Glu Glu Gly Lys Arg Leu Ser Lys Ser Glu Leu		
195	200	205
Glu Lys Gln Pro Lys Ile Glu Asp Trp Asn Asp Asn Pro Ile Asn Gly		
210	215	220
Lys Thr Ile Ala Gln Leu Val Val Ser Ser Leu Glu Pro Gln Thr Lys		
225	230	235
Ile Arg Arg Asp Leu Thr His Gly Leu Gln Thr Glu Ser Thr Leu Asn		
245	250	255
Ile Gln Thr Glu Lys Leu His Gln Ser Asp Tyr Ile His Glu Leu Glu		
260	265	270
Thr Ile Phe Glu Arg Gln His Val Asp Gln Thr Thr Gln Glu Gln Leu		
275	280	285
Leu Glu Ala Thr Phe His Thr Lys Asn Pro Lys Ala Val Gly Ala Ala		
290	295	300
Ala Lys Leu Val Gly Lys Asp Ala Leu Asp Ser Arg Tyr Tyr Arg Ala		
305	310	315
Ser Arg Ala Thr Pro Ala Phe Glu Glu Tyr Arg Val Met Ala Ala Ile		
325	330	335
Asp Thr Leu Arg Ile Arg Glu His Gly Thr Glu Arg Gln Leu Thr Thr		
340	345	350
Asp Glu Arg Arg Lys Leu Phe Asp Phe Ile Lys Gly Leu Pro Ser Lys		
355	360	365
Lys Thr Lys Asn Glu Pro Ser Ile Ser Ser Leu Thr Trp Gly Asp Val		
370	375	380
Ala Asp Phe Leu Gly Ile Gln Arg Ile Asp Leu Arg Gly Leu Gly Ser		
385	390	395
Leu Lys Asp Gly Glu Pro Val Ser Ala Lys Gln Pro Pro Val Ile Glu		
405	410	415

Thr Asn Asp Ile Met Gln Lys Ala Pro Asp Pro Ile Ala Ala Trp Trp
 420 425 430
 Ser Gln Ala Asn Thr Lys Glu Arg Asp Arg Phe Val Glu Phe Met Ser
 435 440 445
 Asn Ala Gly Ala Ile Lys Asp Thr Ser Asp Glu Val Arg Asn Ile Asp
 450 455 460
 Ala Glu Ile Ser Gln Leu Leu Glu Glu Leu Thr Gly Ser Glu Leu Glu
 465 470 475 480

 Ser Leu Asp Lys Ile Thr Leu Thr Ser Gly Arg Ala Ala Tyr Ser Ser
 485 490 495
 Gln Thr Leu Arg Asn Ile Thr Asn Tyr Met Tyr Glu Thr Gly Cys Asp
 500 505 510
 Leu Thr Thr Ala Arg Gln Glu Leu Tyr His Val Gly Lys Asn Trp Ala
 515 520 525
 Pro Pro Ala Pro Pro Ile Tyr Glu His Thr Gly Asn Pro Ser Val Asp
 530 535 540

 Arg Thr Phe Ser Ile Ile His Arg Trp Leu Cys Asn Met Arg Asp Gln
 545 550 555 560
 Tyr Gly Glu Pro Glu Thr Val Asn Ile Glu Tyr Val Arg Asp Gly Phe
 565 570 575
 Ser Ser Thr Ser Thr Gln Leu Ala Glu Gln Arg Glu Arg Asp Arg Arg
 580 585 590
 Tyr Ala Asp Asn Leu Lys Met Leu Ser Asn Tyr Glu Gly Ala Ser Ser
 595 600 605

 Arg Ser Asp Val Arg Arg Ile Lys Ala Leu Gln Arg Gln Asn Cys Gln
 610 615 620
 Cys Ile Tyr Cys Gly Arg Thr Ile Thr Phe Glu Thr Cys Gln Met Asp
 625 630 635 640
 His Val Leu Pro Arg Lys Gly Pro Gly Ser Asp Ser Lys Phe Glu Asn
 645 650 655
 Leu Val Ala Thr Cys Gly Glu Cys Asn Lys Ser Lys Ser Asp Thr Leu

660	665	670	
Tyr Met Asn Trp Ala Lys Thr Tyr Pro Asn Thr Asn Leu Gln Asp Val			
675	680	685	
Leu Arg Arg Ile Gln Glu Trp Ser Lys Asp Gly Trp Met Thr Asp Lys			
690	695	700	
Arg Trp Arg Gln Tyr Lys Glu Ala Leu Ile Leu Arg Leu Glu Ala Thr			
705	710	715	720
Glu Lys Gln Glu Pro Leu Asp Asn Arg Ser Met Glu Ser Val Ser Tyr			
725	730	735	
Met Ala Arg Glu Leu Arg Asn Arg Ile Tyr Gly Phe Tyr Gly Trp His			
740	745	750	
Asp Gln Asp Asp Ala Leu Lys Gln Gly Arg Gln Arg Val Phe Val Ser			
755	760	765	
Ser Gly Ser Met Thr Ala Ala Ala Arg Arg Thr Pro Phe Glu Ser Pro			
770	775	780	
Leu Ile Lys Gly Ala Asp Glu Glu Thr Tyr Glu Ser Ser Leu Pro Trp			
785	790	795	800
Leu Asp Gly Met Lys Gly Lys Thr Arg Leu Asp Arg Arg His His Ala			
805	810	815	
Val Asp Ala Ser Ile Ile Ala Met Met Arg Pro Gln Ile Val Lys Ile			
820	825	830	
Leu Thr Glu Ala Gln Glu Ile Arg Ser Glu Gln His Asp Lys Tyr Arg			
835	840	845	
Lys Gly Gln Thr Pro Asp Tyr Val Cys Lys Arg Arg Asp Tyr Trp Arg			
850	855	860	
Asn Trp Arg Gly Thr Pro Asp Thr Arg Asp Glu Glu Val Phe Asn Tyr			
865	870	875	880
Trp Ala Gly Glu Gln Leu Arg Thr Leu Thr Asp Leu Val Ser Gln Lys			
885	890	895	
Met Ala Asp Asp Glu Ile Pro Val Ile Tyr Pro Thr Arg Leu Arg Leu			
900	905	910	

Gly Asn Gly Ser Ala His Lys Asp Thr Val Val Ser Met Met Thr Arg
915 920 925

Lys Val Gly Asp Glu Leu Ser Ile Thr Ala Ile Asn Lys Ala Glu Ser
930 935 940

Gly Ala Leu Tyr Thr Ala Leu Thr Arg Asp Ser Asp Phe Asp Trp Lys
945 950 955 960

Thr Gly Leu Ser Ala Asn Pro Asn Arg Arg Ile Arg Val His Asp Lys
965 970 975

Trp Phe Glu Ala Asp Asp Thr Ile Lys Phe Leu Glu Pro Ala Val Glu
980 985 990

Val Val Leu Lys Asn Asn Thr Arg Ala Arg Ile Asp Pro Glu Ala Leu
995 1000 1005

Asp Lys Val His Ser Thr Leu Tyr Val Pro Val Arg Gly Gly Ile
1010 1015 1020

Ala Glu Ala Gly Asn Ser Ile His His Val Arg Phe Tyr Lys Ile
1025 1030 1035

Pro Lys Leu Asn Ser Lys Gly Lys Gln Thr Gly Ser Ile Tyr Ala
1040 1045 1050

Met Leu Arg Val Leu Thr Ile Asp Leu Ala Met Asn Gln Tyr Asp
1055 1060 1065

Lys Glu Thr Gly Lys Lys Gln Asp Leu Phe Thr Leu Pro Leu Pro
1070 1075 1080

Glu Ser Ser Leu Ser Arg Arg Phe Ser Glu Pro Lys Leu Arg Gln
1085 1090 1095

Ala Leu Ile Asp Gly Thr Ala Glu Tyr Leu Gly Trp Ala Val Val
1100 1105 1110

Asp Asp Glu Leu Glu Ile Pro Ala Phe Ala Asn Ala Arg Ile Thr
1115 1120 1125

Glu Glu Gln Ala Ile Asn Gly Ser Phe Thr Asp Arg Leu Leu His
1130 1135 1140

Ser Phe Pro Gly Thr His Lys Phe Arg Phe Ala Gly Phe Ser Arg

1145 1150 1155
Asn Thr Glu Ile Ala Ile Arg Pro Val Gln Leu Ala Ser Glu Gly
1160 1165 1170

Leu Ile Glu Thr Asp Glu Asn Arg Lys Arg Gln Gln Leu Arg Leu
1175 1180 1185
Thr Gln Pro Asn Thr Glu Tyr Ser Asn Ser Ile Lys Asn Val Leu
1190 1195 1200
Lys Ser Gly Leu His Leu Lys Val Asn Thr Leu Phe Gln Thr Gly
1205 1210 1215
Ile Leu Val Thr Arg Ala Asn Ser Gln Gly Lys Gln Ser Ile Arg
1220 1225 1230

Phe Ser Thr Val Glu Glu Pro Lys Lys Lys Arg Lys Val
1235 1240 1245

<210> 19

<211> 3435

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 19

atgggcggat ctgaagtggg aaccgtgcct gtgacttggg gactgggagt cgatgtgggc	60
gagagatcca ttggactggc cgccgtgtcc tacgaagagg acaagcccaa agaaatcctg	120
gctgctgtgt cctggattca cgatggcgga gtgggcgacg aaagaagcgg agctagtaga	180
ctggccctga gaggcattggc cagaagggct agacggctgc ggagattccg tagagccaga	240
ctgcgcgacc tggacatgct gctgtctgaa ctcgatgga cccctctgcc tgacaagaac	300
gtgtcacctg tggatgcctg gctggccaga aagagactgg cagaggaata cgtggtggac	360
gagacagaga gaagaaggct gctgggctac gccgtgtctc acatggctag acatagaggc	420
tggcggaacc cctggaccac catcaaggac ctgaagaacc tgcctcagcc tagcgacagc	480
tgggagagaa ccagagaaag cctggaagcc cggtaactccg tgtctctgga acctggcaca	540
gttggacagt gggccggata cctgctgcag agagcccctg gcatcagact gaaccctaca	600

cagcagagcg ccggaagaag ggccgaactg tctaattgcca ccgccttcga gacaagactg 660
 cggcaagagg atgtgctgtg ggagctgaga tgtatcgccg acgttcaggg cctgcctgag 720
 gagtggtgt ccaatgtgat cgacgccgtg ttctgccaga aaagacctag cgtgcccgcc 780
 gagagaatcg gcagagatcc tctcgatccc agccagctga gagccagcag agcctgcctg 840
 gaatttcaag agtaccggat cgtggccgct gtggccaacc tgagaatcag agatggcagc 900
 ggacgagac cctgagctt ggaagaaaga aacgccgtga tcgaggccct gctggcccag 960
 acagaaagaa gcctcacttg gagcgacatt gccttggaag tccitgaagct gcccacagag 1020

 agcgacctga cctctgtgcc tgaagaggat ggcccaagca gcctggccta ctctcagttc 1080
 gcccctttcg atgagacaag cgcccgatc gccgagtta tcgccaagaa cagacggaag 1140
 atccccacat tcgcccagtg gtggcaagag caggatcgga ccagtagaag cgatctggtg 1200
 gctgcccctgg ccgacaattc tattgccggc gaggaagaac aagagctgct ggtgcatctg 1260
 cccgacgccg aacttgaagc tctggaagga ctggctctgc cctctggcag agtggcctat 1320
 agcagactga cactgagcgg cctgaccaga gtgatgagag atgatggcgt ggacgtgcac 1380
 aacgcccga agacatgctt cggagtggac gacaattggc ggctccact gcctgctctg 1440

 catgaagcta caggacaccc cgtggtggat agaaacctgg ctatcctgcg gaagtctctg 1500
 agcagcgcca ccatgagatg gggccctcca cagtctatcg tgggtggaact tgccagaggc 1560
 gccagcgaga gcagagaaag gcaggccgaa gaagaagccg ctcgagagc ccacagaaag 1620
 gccaacgaca gaattagagc cgaactcaga gcctccggcc tgagcgatcc ttctcctgcc 1680
 gatcttgta gagcccggt gctggaactg tacgactgcc actgtatgta ctgtggcgcc 1740
 cctatctcct gggagaacag cgagctggat cacatcgtgc ccagaacaga tggcgatcc 1800
 aacagacacg agaacctggc cattacatgc ggccgctgca acaagaaaa aggcagaagg 1860

 ccttcgcca gctgggcca gacaagcaat agagtgcagc tgcgggacgt gatcgaccgg 1920
 gtgcagaagc tgaagtacg cggcaacatg tactggacc gggacgagtt cagccggtac 1980
 aagaaaagcg tgggtggccg gctgaagcgg agaacctctg atcctgaagt gatccagagc 2040
 atcgagagca ccggctatgc tgccgtggct ctgagagata gactgctgag ctacggcgag 2100
 aagaatggcg tggcacaggt ggccgttttt agaggcgag tgacagcca gccagaaga 2160
 tggctggaca tctccatcga gcggctgttc agtagagtgg ccatcttcgc ccagagcacc 2220
 tccaccaaga ggctggatag aaggcaccac gccgtggatg ctgtggtgct gacaacactg 2280

 acaccggcg tggccaagac actggctgat gctagaagca gaagagtgtc cgccgagttc 2340
 tggcgagac caagcgacgt gaacagacac agcaccgagg aacctcagag ccccgctac 2400
 agacagtga aagagagctg ttctggcctg ggcgacctgc tgatttctac cgccgccaga 2460

gattctatcg ccgtggctgc tcctctgaga ctgaggccaa caggcgact gcacgaggaa 2520
accctgagag cctttagcga gcacacagtg ggagccgctt ggaagggcgc tgagctgaga 2580
agaatcgtgg aacccgaagt gtacgccgcc ttcttgccac ttacagatcc tggcggcaga 2640
ttctgaagg tgicccctag cgaagatgtg ctgcctgccg acgagaacag gcacattgtg 2700

ctgagcgaca gagtgtctggg cccagagac agagtgaac tgttccccga cgaccggggc 2760
agcatcagag tcagaggtag cgcagcctat atcgccagct ttcaccacgc cagagtgttc 2820
agatggggaa gcagccactc tcttagcttc gcctgtctga gagtctctct ggctgatctg 2880
gctgtggctg gcctgcttag agatggggtc gacgtgttca cagccgagct gccaccttgg 2940
actcccgctt ggagatatgc ctctatgcc ctggtcaagg ccgtggaaag cggcgacgct 3000
aagcaagttg gatggctggg gcctggcgac gaactggatt ttggacctga gggcgtgaca 3060
accgctgccg gcgatctgag catgttcttg aagtactttc ccgagcggca ctgggtcgtg 3120

accggcttcg aagatgacaa gaggatcaac ctgaagcctg ccttctctgc tgccgaacag 3180
gctgaggtgc tgaggactga gagaagcgac agaccgaca cactgacaga ggccggcgaa 3240
attctggccc agttcttccc tagatgttgg cgggccacag tggttaaggt gctgtgcat 3300
cctggcctga ccgtgatcag aagaacagcc ctgggacagc ctaggtggcg gagaggacat 3360
ctgccttatt catggcgccc ttggagcgcc gatccttggg gtggcggaac acctccaag 3420
aagaagagga aggtg 3435

<210> 20

<211> 1145

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 20

Met Gly Gly Ser Glu Val Gly Thr Val Pro Val Thr Trp Arg Leu Gly

1 5 10 15

Val Asp Val Gly Glu Arg Ser Ile Gly Leu Ala Ala Val Ser Tyr Glu

20 25 30

Glu Asp Lys Pro Lys Glu Ile Leu Ala Ala Val Ser Trp Ile His Asp

35 40 45

Gly Gly Val Gly Asp Glu Arg Ser Gly Ala Ser Arg Leu Ala Leu Arg

50

55

60

Gly Met Ala Arg Arg Ala Arg Arg Leu Arg Arg Phe Arg Arg Ala Arg

65

70

75

80

Leu Arg Asp Leu Asp Met Leu Leu Ser Glu Leu Gly Trp Thr Pro Leu

85

90

95

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

100

105

110

Leu Ala Glu Glu Tyr Val Val Asp Glu Thr Glu Arg Arg Arg Leu Leu

115

120

125

Gly Tyr Ala Val Ser His Met Ala Arg His Arg Gly Trp Arg Asn Pro

130

135

140

Trp Thr Thr Ile Lys Asp Leu Lys Asn Leu Pro Gln Pro Ser Asp Ser

145

150

155

160

Trp Glu Arg Thr Arg Glu Ser Leu Glu Ala Arg Tyr Ser Val Ser Leu

165

170

175

Glu Pro Gly Thr Val Gly Gln Trp Ala Gly Tyr Leu Leu Gln Arg Ala

180

185

190

Pro Gly Ile Arg Leu Asn Pro Thr Gln Gln Ser Ala Gly Arg Arg Ala

195

200

205

Glu Leu Ser Asn Ala Thr Ala Phe Glu Thr Arg Leu Arg Gln Glu Asp

210

215

220

Val Leu Trp Glu Leu Arg Cys Ile Ala Asp Val Gln Gly Leu Pro Glu

225

230

235

240

Asp Val Val Ser Asn Val Ile Asp Ala Val Phe Cys Gln Lys Arg Pro

245

250

255

Ser Val Pro Ala Glu Arg Ile Gly Arg Asp Pro Leu Asp Pro Ser Gln

260

265

270

Leu Arg Ala Ser Arg Ala Cys Leu Glu Phe Gln Glu Tyr Arg Ile Val

275

280

285

Ala Ala Val Ala Asn Leu Arg Ile Arg Asp Gly Ser Gly Ser Arg Pro

290 295 300
 Leu Ser Leu Glu Glu Arg Asn Ala Val Ile Glu Ala Leu Leu Ala Gln

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

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

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

 500 505 510
 Ile Val Val Glu Leu Ala Arg Gly Ala Ser Glu Ser Arg Glu Arg Gln
 515 520 525
 Ala Glu Glu Glu Ala Ala Arg Arg Ala His Arg Lys Ala Asn Asp Arg
 530 535 540

Ile Arg Ala Glu Leu Arg Ala Ser Gly Leu Ser Asp Pro Ser Pro Ala
545 550 555 560
Asp Leu Val Arg Ala Arg Leu Leu Glu Leu Tyr Asp Cys His Cys Met

 565 570 575
Tyr Cys Gly Ala Pro Ile Ser Trp Glu Asn Ser Glu Leu Asp His Ile
 580 585 590
Val Pro Arg Thr Asp Gly Gly Ser Asn Arg His Glu Asn Leu Ala Ile
 595 600 605
Thr Cys Gly Ala Cys Asn Lys Glu Lys Gly Arg Arg Pro Phe Ala Ser
 610 615 620
Trp Ala Glu Thr Ser Asn Arg Val Gln Leu Arg Asp Val Ile Asp Arg

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

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

 755 760 765
Ala Asp Ala Arg Ser Arg Arg Val Ser Ala Glu Phe Trp Arg Arg Pro
 770 775 780
Ser Asp Val Asn Arg His Ser Thr Glu Glu Pro Gln Ser Pro Ala Tyr

785 790 795 800
 Arg Gln Trp Lys Glu Ser Cys Ser Gly Leu Gly Asp Leu Leu Ile Ser
 805 810 815
 Thr Ala Ala Arg Asp Ser Ile Ala Val Ala Ala Pro Leu Arg Leu Arg

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

 885 890 895
 Arg His Ile Val Leu Ser Asp Arg Val Leu Gly Pro Arg Asp Arg Val
 900 905 910
 Lys Leu Phe Pro Asp Asp Arg Gly Ser Ile Arg Val Arg Gly Gly Ala
 915 920 925
 Ala Tyr Ile Ala Ser Phe His His Ala Arg Val Phe Arg Trp Gly Ser
 930 935 940
 Ser His Ser Pro Ser Phe Ala Leu Leu Arg Val Ser Leu Ala Asp Leu

 945 950 955 960
 Ala Val Ala Gly Leu Leu Arg Asp Gly Val Asp Val Phe Thr Ala Glu
 965 970 975
 Leu Pro Pro Trp Thr Pro Ala Trp Arg Tyr Ala Ser Ile Ala Leu Val
 980 985 990
 Lys Ala Val Glu Ser Gly Asp Ala Lys Gln Val Gly Trp Leu Val Pro
 995 1000 1005
 Gly Asp Glu Leu Asp Phe Gly Pro Glu Gly Val Thr Thr Ala Ala

 1010 1015 1020
 Gly Asp Leu Ser Met Phe Leu Lys Tyr Phe Pro Glu Arg His Trp
 1025 1030 1035

Val Val Thr Gly Phe Glu Asp Asp Lys Arg Ile Asn Leu Lys Pro
1040 1045 1050

Ala Phe Leu Ser Ala Glu Gln Ala Glu Val Leu Arg Thr Glu Arg
1055 1060 1065

Ser Asp Arg Pro Asp Thr Leu Thr Glu Ala Gly Glu Ile Leu Ala

1070 1075 1080

Gln Phe Phe Pro Arg Cys Trp Arg Ala Thr Val Ala Lys Val Leu
1085 1090 1095

Cys His Pro Gly Leu Thr Val Ile Arg Arg Thr Ala Leu Gly Gln
1100 1105 1110

Pro Arg Trp Arg Arg Gly His Leu Pro Tyr Ser Trp Arg Pro Trp
1115 1120 1125

Ser Ala Asp Pro Trp Ser Gly Gly Thr Pro Pro Lys Lys Lys Arg

1130 1135 1140

Lys Val

1145

<210> 21

<211> 3459

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 21

atggcctata gactgggcct cgacatcggc atcacatctg ttggatgggc cgtcgtggcc 60
ctggaaaagg atgagtctgg actgaagccc gtgcgcatcc aggatctggg cgtcagaatc 120
ttcgacaagg ccgaggatag caagaccggc gcttctctgg ctctgccag aagagaagcc 180
agaagcgcca gaagaagaac ccggcggaga aggcacagac tgtggcgct gaaaagactg 240

ctggaacagc acggcatcct gagcatggaa cagatcgagg ccctgtacgc ccagagaaca 300
agcagccctg atgtgtatgc cctgagagtg gccggcctgg acagatgtct gatcgccgaa 360
gagatcgccc ggggtctgat tcacattgcc cacagaagag gcttcagag caacagaaag 420
agcgagatca aggacagcga cgccggcaag ctgctgaagg ccgtgcaaga gaacgagaac 480

ctgatgcaga gcaagggcta cagaaccgtg gccgagatgc tgggtgtctga ggccacaaag 540
 acagacgccg agggaaagct ggtgcacggc aagaagcacg gctacgtcag caacgtgcgg 600
 aacaaggccg gcgagtacag acacacagtg tccagacagg ccatcgtgga cgaagtgcgg 660

 aagattttcg ccgtcagag agccctgggc aacgacgtga tgagcgagga actggaagat 720
 agctacctga agatcctgtg cagccagcgg aacttcgatg atggccctgg cggcgattct 780
 ccttatggac acggaagcgt tagcccgac ggcgctcagac agagcatcta cgagagaatg 840
 gtcggaagct gcaccttcga gacaggcgag aagagagccc ctagaagcag ctacagcttc 900
 gagcggtttc agctgctgac caaggtggtc aacctgcgga tctaccggca gcaagaggat 960
 ggccgagat acccttgtga actgaccag accgagcggg ccagagtgat cgattgtgcc 1020
 tacgagcaga ccaagatcac ctacggaaag ctgagaaagc tgctggacat gaaggacacc 1080

 gagagctttg ccggcctgac ctacggcctg aacagaagca gaaacaagac cgaggacacc 1140
 gtgttcgtgg aaatgaagtt ctaccacgaa gtccgcaagg ccctgcagag agccgggggtt 1200
 ttcatcagg acctgagcat cgagacactg gaccagatcg gctggattct gagcgtgtgg 1260
 aagtcgcagc acaaccggcg gaagaagctg tctacactgg gcctgagcga caacgtgatc 1320
 gaagaactgc tgccctgaa cggctccaag tttggccacc tgagcctgaa ggccatcaga 1380
 aagatcctgc ctttctgga agatgggtac agctacgacg tggcctgtga actggccggc 1440
 tatcagtttc agggcaagac agagtacgtg aagcagcggc tgctgcctcc acttgagaa 1500

 ggcgaaagtga caaacccgt tgtgcccaga gcactgagcc aggccatcaa ggttgtgaac 1560
 gccgtgatca gaaagcacgg cagcccagag agcatccaca tcgaactggc cagagagctg 1620
 agcaagaacc tggacgagcg gagaagatc gagaaggccc agaagaaaa tcagaagaac 1680
 aacgagcaga ttaaggacga gatccgcgag atcctgggat ccgccatgt gaccggaaga 1740
 gacatcgtga agtacaagct gttcaaacag caacaagagt tctgcatgta cagcggcgag 1800
 aagctggacg tgaccagact gtttgagcct ggctatgccg aggtggacca catcatcct 1860
 tacggcatca gtttcgacga ctctacgac aacaaggtgc tggtttaagac cgagcagaac 1920

 cggcagaagg gcaatagaac ccctctggaa tacctgcggg acaagcctga gcagaaggcc 1980
 aagtttatcg ccctggtgga atctatccct ctgagccaga aaaagaaaa ccacctcctg 2040
 atggacaagc gggccatcga cctggaacaa gagggcttca gagagcggaa cctgagcgat 2100
 acccggtaca tcacacgggc cctgatgaac cacatccagg cttggctgct gttcgacgag 2160
 acagccagca ccagatcaa gaggtcgtg tgtgtgaatg gcgccgtgac gcctacatg 2220
 agagctagat ggggcctgac aaaggataga gatgccggcg ataagcacca cgccgctgat 2280

gctgtggtgg tggcctgtat cggagacagc ctgatccaga gagtgaccaa atacgacaag 2340

ttcaagcggga acgccctggc cgaccggaac agatatgtgc agcaggtttc caagagcgag 2400

ggcatcaccc agtacgtgga caaagaaacc ggcgaggtgt tcacctggga gtccttcgat 2460

gagcgggaagt tcctgcctaa cgagcccttg gaaccttggc cattcttcag ggatgagctg 2520

ctggccagac tgagcgacga cccctccaag aacatcagag ccatcgccct gctgacctac 2580

agcgagactg agcagatcga tcccatcttc gtgtccagaa tgcccaccag aaaagtgacc 2640

ggcgcagccc acaaagagac aatcagatcc ccacggatcg tgaagggtgga cgataacaag 2700

ggcaccgaga tccaggtggt ggtgtctaag gtggccctga ccgagctgaa gctgacaaaa 2760

gacggcgaaa tcaaggatta cttcaggccc gaggacgacc ccagactgta caacaccctg 2820

agagaacggc tgggtgcagtt cggcggagat gccaaggccg ctttcaaaga acccgtgtac 2880

aagatcagca aggacggctc tgtgcggacc cctgtgcgga aagtgaagat tcaagagaag 2940

ctgacactgg gcgtgccagt gcatggcgga agaggaattg ccgagaatgg cggcatggtc 3000

cgaatcgacg tgttcgccaa aggcggcaag tactacttcg tgcccatcta cgtggccgac 3060

gtgtgaaga gagagctgcc caacagactg gccaccgctc acaagcctta cagcgaatgg 3120

cgcgtggtgg acgacagcta ccagttcaag ttctctctgt accccaacga tgccgtgatg 3180

atcaagccca gcagagaggt ggacatcacc tacaaggacc ggaaagagcc cgtcggctgc 3240

cggatcatgt actttgtgtc cgccaatatc gccagcgccct ccatcagcct gagaaccac 3300

gataactccg gcgagctgga aggactgggc atccaaggac tggaagtgtt cgagaaatac 3360

gtcgtgggcc ctctgggcga cacacacct gtgtacaaag aacggcggat gcccttcaga 3420

gtggaacgga agatgaaccc caagaagaag aggaaggtg 3459

<210> 22

<211> 1153

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

polypeptide"

<400> 22

Met Ala Tyr Arg Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly Trp

1 5 10 15

Ala Val Val Ala Leu Glu Lys Asp Glu Ser Gly Leu Lys Pro Val Arg

20	25	30
Ile Gln Asp Leu Gly Val Arg Ile Phe Asp Lys Ala Glu Asp Ser Lys		
35	40	45
Thr Gly Ala Ser Leu Ala Leu Pro Arg Arg Glu Ala Arg Ser Ala Arg		
50	55	60
Arg Arg Thr Arg Arg Arg Arg His Arg Leu Trp Arg Val Lys Arg Leu		
65	70	75
Leu Glu Gln His Gly Ile Leu Ser Met Glu Gln Ile Glu Ala Leu Tyr		
85	90	95
Ala Gln Arg Thr Ser Ser Pro Asp Val Tyr Ala Leu Arg Val Ala Gly		
100	105	110
Leu Asp Arg Cys Leu Ile Ala Glu Glu Ile Ala Arg Val Leu Ile His		
115	120	125
Ile Ala His Arg Arg Gly Phe Gln Ser Asn Arg Lys Ser Glu Ile Lys		
130	135	140
Asp Ser Asp Ala Gly Lys Leu Leu Lys Ala Val Gln Glu Asn Glu Asn		
145	150	155
Leu Met Gln Ser Lys Gly Tyr Arg Thr Val Ala Glu Met Leu Val Ser		
165	170	175
Glu Ala Thr Lys Thr Asp Ala Glu Gly Lys Leu Val His Gly Lys Lys		
180	185	190
His Gly Tyr Val Ser Asn Val Arg Asn Lys Ala Gly Glu Tyr Arg His		
195	200	205
Thr Val Ser Arg Gln Ala Ile Val Asp Glu Val Arg Lys Ile Phe Ala		
210	215	220
Ala Gln Arg Ala Leu Gly Asn Asp Val Met Ser Glu Glu Leu Glu Asp		
225	230	235
Ser Tyr Leu Lys Ile Leu Cys Ser Gln Arg Asn Phe Asp Asp Gly Pro		
245	250	255
Gly Gly Asp Ser Pro Tyr Gly His Gly Ser Val Ser Pro Asp Gly Val		
260	265	270

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

 Gly Gly Arg Tyr Pro Cys Glu Leu Thr Gln Thr Glu Arg Ala Arg Val
 325 330 335
 Ile Asp Cys Ala Tyr Glu Gln Thr Lys Ile Thr Tyr Gly Lys Leu Arg
 340 345 350
 Lys Leu Leu Asp Met Lys Asp Thr Glu Ser Phe Ala Gly Leu Thr Tyr
 355 360 365
 Gly Leu Asn Arg Ser Arg Asn Lys Thr Glu Asp Thr Val Phe Val Glu
 370 375 380

 Met Lys Phe Tyr His Glu Val Arg Lys Ala Leu Gln Arg Ala Gly Val
 385 390 395 400
 Phe Ile Gln Asp Leu Ser Ile Glu Thr Leu Asp Gln Ile Gly Trp Ile
 405 410 415
 Leu Ser Val Trp Lys Ser Asp Asp Asn Arg Arg Lys Lys Leu Ser Thr
 420 425 430
 Leu Gly Leu Ser Asp Asn Val Ile Glu Glu Leu Leu Pro Leu Asn Gly
 435 440 445

 Ser Lys Phe Gly His Leu Ser Leu Lys Ala Ile Arg Lys Ile Leu Pro
 450 455 460
 Phe Leu Glu Asp Gly Tyr Ser Tyr Asp Val Ala Cys Glu Leu Ala Gly
 465 470 475 480
 Tyr Gln Phe Gln Gly Lys Thr Glu Tyr Val Lys Gln Arg Leu Leu Pro
 485 490 495
 Pro Leu Gly Glu Gly Glu Val Thr Asn Pro Val Val Arg Arg Ala Leu
 500 505 510

 Ser Gln Ala Ile Lys Val Val Asn Ala Val Ile Arg Lys His Gly Ser

515	520	525	
Pro Glu Ser Ile His Ile Glu Leu Ala Arg Glu Leu Ser Lys Asn Leu			
530	535	540	
Asp Glu Arg Arg Lys Ile Glu Lys Ala Gln Lys Glu Asn Gln Lys Asn			
545	550	555	560
Asn Glu Gln Ile Lys Asp Glu Ile Arg Glu Ile Leu Gly Ser Ala His			
565	570	575	
Val Thr Gly Arg Asp Ile Val Lys Tyr Lys Leu Phe Lys Gln Gln Gln			
580	585	590	
Glu Phe Cys Met Tyr Ser Gly Glu Lys Leu Asp Val Thr Arg Leu Phe			
595	600	605	
Glu Pro Gly Tyr Ala Glu Val Asp His Ile Ile Pro Tyr Gly Ile Ser			
610	615	620	
Phe Asp Asp Ser Tyr Asp Asn Lys Val Leu Val Lys Thr Glu Gln Asn			
625	630	635	640
Arg Gln Lys Gly Asn Arg Thr Pro Leu Glu Tyr Leu Arg Asp Lys Pro			
645	650	655	
Glu Gln Lys Ala Lys Phe Ile Ala Leu Val Glu Ser Ile Pro Leu Ser			
660	665	670	
Gln Lys Lys Lys Asn His Leu Leu Met Asp Lys Arg Ala Ile Asp Leu			
675	680	685	
Glu Gln Glu Gly Phe Arg Glu Arg Asn Leu Ser Asp Thr Arg Tyr Ile			
690	695	700	
Thr Arg Ala Leu Met Asn His Ile Gln Ala Trp Leu Leu Phe Asp Glu			
705	710	715	720
Thr Ala Ser Thr Arg Ser Lys Arg Val Val Cys Val Asn Gly Ala Val			
725	730	735	
Thr Ala Tyr Met Arg Ala Arg Trp Gly Leu Thr Lys Asp Arg Asp Ala			
740	745	750	
Gly Asp Lys His His Ala Ala Asp Ala Val Val Val Ala Cys Ile Gly			
755	760	765	

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

770 775 780

Ala Leu Ala Asp Arg Asn Arg Tyr Val Gln Gln Val Ser Lys Ser Glu

785 790 795 800

Gly Ile Thr Gln Tyr Val Asp Lys Glu Thr Gly Glu Val Phe Thr Trp

805 810 815

Glu Ser Phe Asp Glu Arg Lys Phe Leu Pro Asn Glu Pro Leu Glu Pro

820 825 830

Trp Pro Phe Phe Arg Asp Glu Leu Leu Ala Arg Leu Ser Asp Asp Pro

835 840 845

Ser Lys Asn Ile Arg Ala Ile Gly Leu Leu Thr Tyr Ser Glu Thr Glu

850 855 860

Gln Ile Asp Pro Ile Phe Val Ser Arg Met Pro Thr Arg Lys Val Thr

865 870 875 880

Gly Ala Ala His Lys Glu Thr Ile Arg Ser Pro Arg Ile Val Lys Val

885 890 895

Asp Asp Asn Lys Gly Thr Glu Ile Gln Val Val Val Ser Lys Val Ala

900 905 910

Leu Thr Glu Leu Lys Leu Thr Lys Asp Gly Glu Ile Lys Asp Tyr Phe

915 920 925

Arg Pro Glu Asp Asp Pro Arg Leu Tyr Asn Thr Leu Arg Glu Arg Leu

930 935 940

Val Gln Phe Gly Gly Asp Ala Lys Ala Ala Phe Lys Glu Pro Val Tyr

945 950 955 960

Lys Ile Ser Lys Asp Gly Ser Val Arg Thr Pro Val Arg Lys Val Lys

965 970 975

Ile Gln Glu Lys Leu Thr Leu Gly Val Pro Val His Gly Gly Arg Gly

980 985 990

Ile Ala Glu Asn Gly Gly Met Val Arg Ile Asp Val Phe Ala Lys Gly

995 1000 1005

Gly Lys Tyr Tyr Phe Val Pro Ile Tyr Val Ala Asp Val Leu Lys

1010	1015	1020	
Arg Glu Leu Pro Asn Arg Leu	Ala Thr Ala His Lys	Pro Tyr Ser	
1025	1030	1035	
Glu Trp Arg Val Val Asp Asp	Ser Tyr Gln Phe Lys	Phe Ser Leu	
1040	1045	1050	
Tyr Pro Asn Asp Ala Val Met	Ile Lys Pro Ser Arg	Glu Val Asp	
1055	1060	1065	
Ile Thr Tyr Lys Asp Arg Lys	Glu Pro Val Gly Cys	Arg Ile Met	
1070	1075	1080	
Tyr Phe Val Ser Ala Asn Ile	Ala Ser Ala Ser Ile	Ser Leu Arg	
1085	1090	1095	
Thr His Asp Asn Ser Gly Glu	Leu Glu Gly Leu Gly	Ile Gln Gly	
1100	1105	1110	
Leu Glu Val Phe Glu Lys Tyr	Val Val Gly Pro Leu	Gly Asp Thr	
1115	1120	1125	
His Pro Val Tyr Lys Glu Arg	Arg Met Pro Phe Arg	Val Glu Arg	
1130	1135	1140	
Lys Met Asn Pro Lys Lys Lys	Arg Lys Val		
1145	1150		
<210> 23			
<211> 4248			
<212> DNA			
<213> Artificial Sequence			
<220><221> source			
<223> /note="Description of Artificial Sequence: Synthetic			
polynucleotide"			
<400> 23			
atgctggtgt ccctatctc tgtggatctc ggcggcaaga ataccggctt cttcagcttc	60		
accgacagcc tggacaatag ccagagcggc accgtgatct acgacgagag cttcgtgctg	120		
agccaagtgg gcagaagaag caagcggcac agcaagcgga acaacctgag aaacaagctg	180		
gtcaagcggc tgttcctgct gatcctgcaa gagcaccacg gcctgagcat cgacgttctg	240		

cccgatgaga tccggggcct gttcaacaag agaggctaca cctacgccgg cttcgagctg 300
 gacgagaaga agaaggacgc cctggaaagc gatacctga aagagttcct gagcgagaag 360
 ctgcagtcca tcgacagaga cagcgacgtg gaagatttcc tgaaccagat cgccagcaac 420
 gccgagagct ttaaggacta caagaaaggc ttcgaggccg tgttcgccag cgccacacac 480
 agccccaaca agaagctgga actgaaggac gagctgaagt ccgagtacgg cgagaacgcc 540
 aaagaactgc tggccggcct gagagtgacc aaagagatcc tggacgagtt cgacaagcaa 600
 gagaaccagg gcaacctgcc tcgggccaag tactttgagg aactgggcga gtatatcgcc 660

 accaacgaga aagtcaagag cttcttcgac agcaacagcc tgaagctgac cgacatgacc 720
 aagctgatcg gcaacatcag caactaccag ctgaaagagc tgcggcggta cttcaacgac 780
 aaagagatgg aaaaggcgca catctggatt cccaacaagc tgcacaagat caccgagaga 840
 tttgtcggga gctggcaccc caagaacgac gccgatagac agagaagggc cgagctgatg 900
 aaggacctga agtccaaaga aatcatggaa ctgctgacca ccaccgagcc tgtgatgaca 960
 atccctcctt acgacgacat gaacaacaga ggcgccgtga agtgtcagac cctgcggctg 1020
 aatgaggaat acctggacaa acatctgccc aactggcggg atatcgccaa gagactgaac 1080

 caccgcaagt tcaacgacga cctggccgac tctaccgtga agggctacag cgaggatagc 1140
 accctgctgc acagactgct ggacacctct aaagagatcg acatctacga gctgcggggc 1200
 aagaagccca acgagctgct ggttaagaca ctgggccaga gcgacgcaa cagactgtat 1260
 ggcttcgccc agaactacta tgagctgac cggcagaaag tgcgcgtgg catttgggtg 1320
 cccgtgaaga acaaggatga ctccctgaac ctggaagata actccaacat gctgaagcgg 1380
 tgcaaccaca atcctccaca caagaagaat cagatccaca acctggtggc cggcatcctg 1440
 ggagtgaac tggatgaggg caagttcgcc gagttcgaga aagagctttg gagcgccaaa 1500

 gtgggcaaca agaaactgag cgcctactgc aagaacatcg aggaactgag aaagaccac 1560
 ggcaacacct tcaagatcga tatagaggaa ctgcgcaaga aggacccgc cgagctgtcc 1620
 aaagaggaaa aggccaaagt gagactgacc gacgacgtga tcctgaatga gtggtcccag 1680
 aagatcgcca acttctttga catcgacgac aagcacgggc agcggttcaa caacctgttc 1740
 agcatggccc agctgcacac agtgatcgac acaccagaa gcggcttcag ctctacctgc 1800
 aaaagatgca ccgccgagaa caggttcaga agcgagacag ccttctacaa cgacgagaca 1860
 ggcgagttcc acaagaaggc cacagccacc tgcagagac tgcccgtga taccagagg 1920

 cctttcagcg gaaagatcga gcggtacatc gacaagctgg gatacgagct ggccaagatc 1980
 aaggctaaag aactggaagg catggaagct aaagaaatca aggtgcccac catcctggaa 2040
 cagaacgcct tcgagtacga ggaaagcctg cggaagtcta agaccgcatc caacgacaga 2100

gtgatcaact ccaagaaaga ccgcgacgga aagaaactgg ccaaggccaa agagaacgcc 2160
gaggacagcg tgaaggacaa ggacaagcgg atcaaggcct tcagcagcgg catctgccct 2220
tactgcggag ataccatcgg agatgacggc gagatcgacc acatcctgcc tagaagccac 2280
acactgaaaa tctacgggac cgtgttcaac cccgagggca atctgatcta cgtgcaccag 2340

aagtgaacc aggccaaagc cgacagcatc tacaagctga gcgatatcaa ggccggcgtg 2400
tcagcccagt ggattgaaga acaggtggcc aacattaagg ggtacaagac cttcagcgtg 2460
ctgtccgccg aacagcagaa ggccittaga tacgccctgt tcctccagaa cgacaacgag 2520
gcctacaaaa aggtgggtga ctggctgcgg accgaccagt ctgctagagt gaacggcaca 2580
cagaagtacc tggccaaaaa gatccaagag aagctacca agatgctgcc taacaagcac 2640
ctgagcttcg agttcatcct ggccgatgcc accgaggtgt cagagctgag aaggcagtac 2700
gccagacaga accctctgct ggctaaggcc gagaagcagg ccccttcttc tcacgccatt 2760

gatgccgtga tggccttcgt ggccagatac cagaaggtgt tcaaggacgg caccctcct 2820
aacgccgatg aggtggcaaa actggctatg ctggacagct ggaacccgc ctctaagag 2880
cctctgacaa agggcctgtc cacgaaccag aaaatcgaga agatgatcaa gagcgcgac 2940
tacggccaga aaaacatgag agaggtgttc ggcaagtcca tcttcggaga gaatgccatc 3000
ggcgagagat acaagcccat cgtggttcaa gaaggcggct actacatcgg ctacccgcc 3060
acagtgaaaa agggctacga actgaagaac tgcaaggtgg tcaccagcaa gaacgatatt 3120
gccaagctgg aaaagatcat caagaaccag gacctgatct ctctgaaaga gaatcagtac 3180

atcaaaatct tctccatcaa caagcagacc atcagcgagc tgagcaaccg ctacttcaac 3240
atgaattaca agaacctggt cgagcgggac aaagaaattg tgggactgct tgagtttacc 3300
gtcgagaact gccgttacta caccaagaaa gtggacgtga agttcgcccc taagtacatc 3360
cacgagacaa agtacccttt ctacgatgac tggcggagat tcgacgaggc ctggcggat 3420
ctgaagaaa accagaacaa gaccagctcc aaggaccgct tcgtgatcga taagagcagc 3480
ctgaacgagt actaccagcc agacaagaat gagtacaagc tggacgtgga caccagcct 3540
atctgggacg acttctgcgg gtggtacttc ctggacagat acaagaccgc caacgacaag 3600

aagtccatcc gcatcaaggc ccgcaagaca ttctccctgc tggctgagtc tggcgtgcag 3660
ggcaaagtgt tccgggccaa gagaaagatc cctaccggct acgctatca ggccctgcct 3720
atggacaaca acgtgatcgc tggcgattac gccaacattc tgctggaagc caacagcaag 3780
accctgagcc tggcgcctaa gagcggcatc agcattgaga agcagctgga caaaaagctc 3840
gacgtcatca aaaagaccga cgtgcgcggc ctggcaatcg acaacaactc cttcttcaac 3900
gccgacttcg acacacacgg catccggctg atcgtggaaa acaccagcgt gaaagtggga 3960

aacttcccca tcagcgccat cgataagtcc gccaaagcgga tgatcttcag agccctgttt 4020

gagaaagaga aggggaagcg caagaaaaag accaccatca gcttcaaaga aagcggcctt 4080

gtgcaggact acctcaaggt gttcctgaaa aagatcgtga agatccagct gagaaccgac 4140

ggctccatct ccaacatcgt cgtgcggaag aatgccgccg atttaccct gagctttaga 4200

agcgagcaca tccagaaact gctgaagccc aagaagaaga ggaagggtg 4248

<210> 24

<211> 1416

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

polypeptide"

<400> 24

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

1 5 10 15

Phe Phe Ser Phe Thr Asp Ser Leu Asp Asn Ser Gln Ser Gly Thr Val

20 25 30

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

35 40 45

Arg His Ser Lys Arg Asn Asn Leu Arg Asn Lys Leu Val Lys Arg Leu

50 55 60

Phe Leu Leu Ile Leu Gln Glu His His Gly Leu Ser Ile Asp Val Leu

65 70 75 80

Pro Asp Glu Ile Arg Gly Leu Phe Asn Lys Arg Gly Tyr Thr Tyr Ala

85 90 95

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

100 105 110

Leu Lys Glu Phe Leu Ser Glu Lys Leu Gln Ser Ile Asp Arg Asp Ser

115 120 125

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

130

135

140

Lys Asp Tyr Lys Lys Gly Phe Glu Ala Val Phe Ala Ser Ala Thr His
 145 150 155 160
 Ser Pro Asn Lys Lys Leu Glu Leu Lys Asp Glu Leu Lys Ser Glu Tyr
 165 170 175
 Gly Glu Asn Ala Lys Glu Leu Leu Ala Gly Leu Arg Val Thr Lys Glu
 180 185 190
 Ile Leu Asp Glu Phe Asp Lys Gln Glu Asn Gln Gly Asn Leu Pro Arg

 195 200 205
 Ala Lys Tyr Phe Glu Glu Leu Gly Glu Tyr Ile Ala Thr Asn Glu Lys
 210 215 220
 Val Lys Ser Phe Phe Asp Ser Asn Ser Leu Lys Leu Thr Asp Met Thr
 225 230 235 240
 Lys Leu Ile Gly Asn Ile Ser Asn Tyr Gln Leu Lys Glu Leu Arg Arg
 245 250 255
 Tyr Phe Asn Asp Lys Glu Met Glu Lys Gly Asp Ile Trp Ile Pro Asn

 260 265 270
 Lys Leu His Lys Ile Thr Glu Arg Phe Val Arg Ser Trp His Pro Lys
 275 280 285
 Asn Asp Ala Asp Arg Gln Arg Arg Ala Glu Leu Met Lys Asp Leu Lys
 290 295 300
 Ser Lys Glu Ile Met Glu Leu Leu Thr Thr Thr Glu Pro Val Met Thr
 305 310 315 320
 Ile Pro Pro Tyr Asp Asp Met Asn Asn Arg Gly Ala Val Lys Cys Gln

 325 330 335
 Thr Leu Arg Leu Asn Glu Glu Tyr Leu Asp Lys His Leu Pro Asn Trp
 340 345 350
 Arg Asp Ile Ala Lys Arg Leu Asn His Gly Lys Phe Asn Asp Asp Leu
 355 360 365
 Ala Asp Ser Thr Val Lys Gly Tyr Ser Glu Asp Ser Thr Leu Leu His
 370 375 380
 Arg Leu Leu Asp Thr Ser Lys Glu Ile Asp Ile Tyr Glu Leu Arg Gly

385 390 395 400
 Lys Lys Pro Asn Glu Leu Leu Val Lys Thr Leu Gly Gln Ser Asp Ala
 405 410 415
 Asn Arg Leu Tyr Gly Phe Ala Gln Asn Tyr Tyr Glu Leu Ile Arg Gln
 420 425 430
 Lys Val Arg Ala Gly Ile Trp Val Pro Val Lys Asn Lys Asp Asp Ser
 435 440 445
 Leu Asn Leu Glu Asp Asn Ser Asn Met Leu Lys Arg Cys Asn His Asn

 450 455 460
 Pro Pro His Lys Lys Asn Gln Ile His Asn Leu Val Ala Gly Ile Leu
 465 470 475 480
 Gly Val Lys Leu Asp Glu Ala Lys Phe Ala Glu Phe Glu Lys Glu Leu
 485 490 495
 Trp Ser Ala Lys Val Gly Asn Lys Lys Leu Ser Ala Tyr Cys Lys Asn
 500 505 510
 Ile Glu Glu Leu Arg Lys Thr His Gly Asn Thr Phe Lys Ile Asp Ile

 515 520 525
 Glu Glu Leu Arg Lys Lys Asp Pro Ala Glu Leu Ser Lys Glu Glu Lys
 530 535 540
 Ala Lys Leu Arg Leu Thr Asp Asp Val Ile Leu Asn Glu Trp Ser Gln
 545 550 555 560
 Lys Ile Ala Asn Phe Phe Asp Ile Asp Asp Lys His Arg Gln Arg Phe
 565 570 575
 Asn Asn Leu Phe Ser Met Ala Gln Leu His Thr Val Ile Asp Thr Pro

 580 585 590
 Arg Ser Gly Phe Ser Ser Thr Cys Lys Arg Cys Thr Ala Glu Asn Arg
 595 600 605
 Phe Arg Ser Glu Thr Ala Phe Tyr Asn Asp Glu Thr Gly Glu Phe His
 610 615 620
 Lys Lys Ala Thr Ala Thr Cys Gln Arg Leu Pro Ala Asp Thr Gln Arg
 625 630 635 640

Pro Phe Ser Gly Lys Ile Glu Arg Tyr Ile Asp Lys Leu Gly Tyr Glu

645 650 655

Leu Ala Lys Ile Lys Ala Lys Glu Leu Glu Gly Met Glu Ala Lys Glu

660 665 670

Ile Lys Val Pro Ile Ile Leu Glu Gln Asn Ala Phe Glu Tyr Glu Glu

675 680 685

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

690 695 700

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

705 710 715 720

Glu Asp Arg Leu Lys Asp Lys Asp Lys Arg Ile Lys Ala Phe Ser Ser

725 730 735

Gly Ile Cys Pro Tyr Cys Gly Asp Thr Ile Gly Asp Asp Gly Glu Ile

740 745 750

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

755 760 765

Phe Asn Pro Glu Gly Asn Leu Ile Tyr Val His Gln Lys Cys Asn Gln

770 775 780

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

785 790 795 800

Ser Ala Gln Trp Ile Glu Glu Gln Val Ala Asn Ile Lys Gly Tyr Lys

805 810 815

Thr Phe Ser Val Leu Ser Ala Glu Gln Gln Lys Ala Phe Arg Tyr Ala

820 825 830

Leu Phe Leu Gln Asn Asp Asn Glu Ala Tyr Lys Lys Val Val Asp Trp

835 840 845

Leu Arg Thr Asp Gln Ser Ala Arg Val Asn Gly Thr Gln Lys Tyr Leu

850 855 860

Ala Lys Lys Ile Gln Glu Lys Leu Thr Lys Met Leu Pro Asn Lys His

865 870 875 880

Leu Ser Phe Glu Phe Ile Leu Ala Asp Ala Thr Glu Val Ser Glu Leu

885 890 895
 Arg Arg Gln Tyr Ala Arg Gln Asn Pro Leu Leu Ala Lys Ala Glu Lys

 900 905 910
 Gln Ala Pro Ser Ser His Ala Ile Asp Ala Val Met Ala Phe Val Ala
 915 920 925
 Arg Tyr Gln Lys Val Phe Lys Asp Gly Thr Pro Pro Asn Ala Asp Glu
 930 935 940
 Val Ala Lys Leu Ala Met Leu Asp Ser Trp Asn Pro Ala Ser Asn Glu
 945 950 955 960
 Pro Leu Thr Lys Gly Leu Ser Thr Asn Gln Lys Ile Glu Lys Met Ile

 965 970 975
 Lys Ser Gly Asp Tyr Gly Gln Lys Asn Met Arg Glu Val Phe Gly Lys
 980 985 990
 Ser Ile Phe Gly Glu Asn Ala Ile Gly Glu Arg Tyr Lys Pro Ile Val
 995 1000 1005
 Val Gln Glu Gly Gly Tyr Tyr Ile Gly Tyr Pro Ala Thr Val Lys
 1010 1015 1020
 Lys Gly Tyr Glu Leu Lys Asn Cys Lys Val Val Thr Ser Lys Asn

 1025 1030 1035
 Asp Ile Ala Lys Leu Glu Lys Ile Ile Lys Asn Gln Asp Leu Ile
 1040 1045 1050
 Ser Leu Lys Glu Asn Gln Tyr Ile Lys Ile Phe Ser Ile Asn Lys
 1055 1060 1065
 Gln Thr Ile Ser Glu Leu Ser Asn Arg Tyr Phe Asn Met Asn Tyr
 1070 1075 1080
 Lys Asn Leu Val Glu Arg Asp Lys Glu Ile Val Gly Leu Leu Glu

 1085 1090 1095
 Phe Ile Val Glu Asn Cys Arg Tyr Tyr Thr Lys Lys Val Asp Val
 1100 1105 1110
 Lys Phe Ala Pro Lys Tyr Ile His Glu Thr Lys Tyr Pro Phe Tyr
 1115 1120 1125

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

1355 1360 1365
 Leu Lys Lys Ile Val Lys Ile Gln Leu Arg Thr Asp Gly Ser Ile
 1370 1375 1380
 Ser Asn Ile Val Val Arg Lys Asn Ala Ala Asp Phe Thr Leu Ser

1385 1390 1395
 Phe Arg Ser Glu His Ile Gln Lys Leu Leu Lys Pro Lys Lys Lys
 1400 1405 1410
 Arg Lys Val

1415

<210> 25

<211> 3417

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"

<400> 25

atgaagaaga tcttgggcgt cgacctgggc atcaccagct ttggatacgc catcctgcaa 60
 gagacaggca aggacctgta cagatgcctg gacaacagcg tggatcatgcg gaacaacccc 120

tacgacgaga agtctggcga gagcagccag agcatccgca gcacccagaa atccatgcgg 180
 cggctgatcg agaagcggaa gaaacggatc agatgcgtgg cccagacaat ggaacgctac 240
 ggcatcctgg actactccga gacaatgaag atcaacgacc ccaagaacaa cccgatcaag 300
 aacagatggc agctgagagc cgtggacgcc tggaaaagac ctctgagccc tcaagagctg 360
 ttcgccatct ttgcccatc ggccaagcac cggggctaca agtctatcgc caccgaggac 420
 ctgatctacg agctggaact ggaactcggc ctgaacgacc ctgagaaaaga gtccgagaag 480
 aaggccgacg agcggagaca ggtgtacaac gcctgagac acctggaaga actgcggaag 540

aagtacggcg gcgagacaat cgcccagacc atccacagag ctgtggaagc cggcgacctg 600
 cggagctaca gaaaccacga cgactacgag aagatgatcc gcagagagga catcgaggaa 660
 gagattgaga aggtcctgct gcggcaggct gaactgggag cacttggact gcctgaggaa 720
 caggtgtccg agctgatcga tgagctgaag gcctgcatca ccgaccaaga gatgcccacc 780
 atcgacgaga gcctgttcgg caagtgcacc ttctacaagg acgagctggc cgctcctgcc 840
 tacagctacc tgtacgacct gtaccggctg tacaagaagc tggccgacct gaacatcgac 900

ggctacgaag tgacccaaga ggaccgcgag aaagtgatcg agtgggtcga gaaaaagatc 960

 gcccgaggca agaacctgaa gaaaatcacc cacaaggacc tccggaagat cctcggactg 1020
 gcccttgagc agaagatttt cggcgtcgag gacgagagaa tcgtcaaggg aaagaaagaa 1080
 ccccggaact tcgtgccctt cttcttccig gccgatatcg ccaagttaa agaactgttt 1140
 gccagcatcc agaagcacc cgacgtctg cagattttca gagaactggc cgagatcctg 1200
 cagcggagca agacacctca agaggccctg gatagactga gagccctgat ggccggcaag 1260
 ggcatcgaca ccgatgacag agagctgctg gaactcttca agaacaagcg gagcggcaca 1320
 agagagctga gccaccgcta taccctggaa gccctgcctc tgttcttgga aggctatgac 1380

 gagaaagagg tgcagagaat cctgggcttt gacgaccgcg aggactacag cagatacccc 1440
 aagagcctgc ggcacttgca cctgagagag ggcaacctgt tcgagaaaga agagaatccc 1500
 atcaacaacc acgccgtgaa gtccctggct tcttgggccc tgggactgat cgctgacctg 1560
 tcttggagat acggccccctt cgatgagatc atcctggaaa ccaccaggga cggcctgcct 1620
 gagaagatcc ggaagaaat cgacaaggcc atgcgcgaga gagagaaagc cctggacaag 1680
 atcatcggca agtacaagaa agagttcccc agcatcgaca agcggctggc cagaaagatt 1740
 cagctgtggg agagacagaa aggcctcgat ctgtactccg gcaaagtgat caacctgagc 1800

 cagctgctcg atggatccgc cgacatcgag cacatcgtgc ctcagtctct cggcggcctg 1860
 agcaccgact acaataccat cgtgacctg aagtcctga acgccgcaa gggcaataga 1920
 ctgcttgagg atttggtggc cggaatccc gactacagag aacggatcgg catgctgtct 1980
 gagaagggcc tgatcgactg gaagaagagg aagaacctgc tggcccagag cctggacgaa 2040
 atctacaccg agaacaccca cagcaaaggc atccgggcca caagctacct ggaagctctg 2100
 gttgcccagg tgctgaagcg gtactacca ttctctgac ctgagctgcg caagaatggc 2160
 atcggcgtgc ggatgatccc cggaaaagt accagcaaga ccagaagcct gctgggaatc 2220

 aagagcaaga gccgcgagac aaacttcac cagccgagg atgcctgat tctgagcaca 2280
 ctgaccagag gctggcagaa ccggctgcac agaattgctga gagacaacta cggcaagagc 2340
 gaggccgagc tgaagaact ctggaaaaag tacatgcccc acatcgaggg cctgacactg 2400
 gccgactata tcgatgaggc cttccggcgg ttcatgagca agggcgaaga gtccctgttc 2460
 taccgggaca tgttcgacac catccggtcc atcagctact gggtcgacaa gaagcctctg 2520
 agcggcagca gccacaaaga aacctgttac agcagcagac acgagggtgc cacactgagg 2580
 aaaaacattc tggaagcctt cgacagcctg aacgtgatca aggaccggca caagctgacc 2640

accgaagagt tcatgaagcg ctacgacaaa gagatccggc agaagctgtg gctgcaccgc 2700
atcggcaaca ccaacgacga gtcttaccgc gccgtggaag agagagccac acagattgcc 2760
cagatcctga ccagatacca gctcatggac gcccagaatg acaaagaaat tgatgagaag 2820
tttcagcagg ccctgaaaga gctgatcaca agcccatcg aagtgactgg caagctgctg 2880
cggaaaatga gattcgtgta cgacaagctg aacgccatgc agatcgacag aggcctggtg 2940
gaaaccgaca agaacatgct gggcatccac atcagcaagg gccccaatga gaagctgac 3000
ttcagacgga tggacgtgaa caacgcccac gagctgcaaa aagaacgcag cggaatcctg 3060

tgctacctga acgagatgct gttcatcttc aacaagaagg ggctgattca ctacggctgc 3120
ctgcggtctt acctgaaaa aggccagggc agcaagtata tcgccctgtt caaccctcgg 3180
ttccccgcca atcctaaggc tcagcctagc aagttcacca gcgacagcaa gatcaagcaa 3240
gtcggcatcg gcagcgccac cggaatcatt aaggcccacc tggatctgga tggccacgtg 3300
cgctcttatg aggtgttcgg aacactgccc gagggcagca tcgagtgggtt caaagaggaa 3360
agcggctacg gcagagtgga agatgaccct caccaccca agaagaagag gaaggtg 3417

<210> 26

<211> 1139

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 26

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

1 5 10 15

Ala Ile Leu Gln Glu Thr Gly Lys Asp Leu Tyr Arg Cys Leu Asp Asn

20 25 30

Ser Val Val Met Arg Asn Asn Pro Tyr Asp Glu Lys Ser Gly Glu Ser

35 40 45

Ser Gln Ser Ile Arg Ser Thr Gln Lys Ser Met Arg Arg Leu Ile Glu

50 55 60

Lys Arg Lys Lys Arg Ile Arg Cys Val Ala Gln Thr Met Glu Arg Tyr

65 70 75 80

Gly Ile Leu Asp Tyr Ser Glu Thr Met Lys Ile Asn Asp Pro Lys Asn

85 90 95
 Asn Pro Ile Lys Asn Arg Trp Gln Leu Arg Ala Val Asp Ala Trp Lys
 100 105 110
 Arg Pro Leu Ser Pro Gln Glu Leu Phe Ala Ile Phe Ala His Met Ala

 115 120 125
 Lys His Arg Gly Tyr Lys Ser Ile Ala Thr Glu Asp Leu Ile Tyr Glu
 130 135 140
 Leu Glu Leu Glu Leu Gly Leu Asn Asp Pro Glu Lys Glu Ser Glu Lys
 145 150 155 160
 Lys Ala Asp Glu Arg Arg Gln Val Tyr Asn Ala Leu Arg His Leu Glu
 165 170 175
 Glu Leu Arg Lys Lys Tyr Gly Gly Glu Thr Ile Ala Gln Thr Ile His

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

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

 305 310 315 320
 Ala Gln Gly Lys Asn Leu Lys Lys Ile Thr His Lys Asp Leu Arg Lys
 325 330 335

Ile Leu Gly Leu Ala Pro Glu Gln Lys Ile Phe Gly Val Glu Asp Glu
340 345 350
Arg Ile Val Lys Gly Lys Lys Glu Pro Arg Thr Phe Val Pro Phe Phe
355 360 365
Phe Leu Ala Asp Ile Ala Lys Phe Lys Glu Leu Phe Ala Ser Ile Gln
370 375 380
Lys His Pro Asp Ala Leu Gln Ile Phe Arg Glu Leu Ala Glu Ile Leu
385 390 395 400
Gln Arg Ser Lys Thr Pro Gln Glu Ala Leu Asp Arg Leu Arg Ala Leu
405 410 415
Met Ala Gly Lys Gly Ile Asp Thr Asp Asp Arg Glu Leu Leu Glu Leu
420 425 430
Phe Lys Asn Lys Arg Ser Gly Thr Arg Glu Leu Ser His Arg Tyr Ile
435 440 445
Leu Glu Ala Leu Pro Leu Phe Leu Glu Gly Tyr Asp Glu Lys Glu Val
450 455 460
Gln Arg Ile Leu Gly Phe Asp Asp Arg Glu Asp Tyr Ser Arg Tyr Pro
465 470 475 480
Lys Ser Leu Arg His Leu His Leu Arg Glu Gly Asn Leu Phe Glu Lys
485 490 495
Glu Glu Asn Pro Ile Asn Asn His Ala Val Lys Ser Leu Ala Ser Trp
500 505 510
Ala Leu Gly Leu Ile Ala Asp Leu Ser Trp Arg Tyr Gly Pro Phe Asp
515 520 525
Glu Ile Ile Leu Glu Thr Thr Arg Asp Ala Leu Pro Glu Lys Ile Arg
530 535 540
Lys Glu Ile Asp Lys Ala Met Arg Glu Arg Glu Lys Ala Leu Asp Lys
545 550 555 560
Ile Ile Gly Lys Tyr Lys Lys Glu Phe Pro Ser Ile Asp Lys Arg Leu
565 570 575
Ala Arg Lys Ile Gln Leu Trp Glu Arg Gln Lys Gly Leu Asp Leu Tyr

580 585 590
 Ser Gly Lys Val Ile Asn Leu Ser Gln Leu Leu Asp Gly Ser Ala Asp
 595 600 605
 Ile Glu His Ile Val Pro Gln Ser Leu Gly Gly Leu Ser Thr Asp Tyr
 610 615 620
 Asn Thr Ile Val Thr Leu Lys Ser Val Asn Ala Ala Lys Gly Asn Arg

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

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

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

 820 825 830

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

1070 1075 1080
 Gly Ser Ala Thr Gly Ile Ile Lys Ala His Leu Asp Leu Asp Gly
 1085 1090 1095
 His Val Arg Ser Tyr Glu Val Phe Gly Thr Leu Pro Glu Gly Ser
 1100 1105 1110
 Ile Glu Trp Phe Lys Glu Glu Ser Gly Tyr Gly Arg Val Glu Asp
 1115 1120 1125
 Asp Pro His His Pro Lys Lys Lys Arg Lys Val
 1130 1135

<210> 27

<211> 3207

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"

<400> 27

atggccgaga agcagcacag atggggactc gacatcggca ccaattctat cggctgggcc	60
gtgatcgccc tgatcgaagg cagacctgct ggactggtgg ctaccggcag cagaatcttt	120
agcgacggca gaaaccccaa ggacggcagc tctctggccg tcgagagaag aggacctcgg	180
cagatcgggc ggagaagaga cagatatctc cggcggaggg acagattcat gcaggccctg	240
atcaacgtgg gcctgatgcc tggggatgcc gccgctagaa aagccctggt caccgagaat	300
ccctacgtgc tgagacagag aggccctggac caagctctga ccctgcctga atttggcaga	360
gccctgttcc acctgaacca gcggagagggc ttccagagca acagaaagac cgatcgggcc	420
accgccaag aaagcggcaa agtgaagaac gccattgccg ccttcagagc cggcatgggc	480
aatgccagaa cagtgggaga agccctggcc agacgactgg aagatggcag accagtgcgg	540
gccagaatgg tcggacaggg caaagatgag cactacgagc tgtatatcgc cagagagtgg	600
atcgcccaag agttcgatgc cctgtgggcc agccagcaga gatttcatgc tgaggtgctg	660
gccgacgccg ccagagatag actgagagcc atcctgctgt tccagcggaa gctgctgcct	720
gtgcctgtgg gcaagtgttt cctggaacct aaccagccta gattggccgc tgctctgcct	780
agcgctcaga gattcagact gatgcaagag ctgaaccacc tgagagtgat gaccctggcc	840
gacaagagag agaggccctt gagcttccaa gagagaaacg atctgctggc ccagctggtg	900

gccagaccta agtgcggctt cgacatgctg cggaagatcg tgttcggcgc caacaaagag 960
gcctacagat tcaccatcga gagcgagcgg cggaagaac tgaagggtg tgatacagcc 1020
gccaaagtgg ccaaagtga tgccttggga actagatggc aggcctctgtc cctggacgag 1080
caggatagac tcgtgtgcct gctgctggac ggcgagaatg atgctgtgct ggctgatgcc 1140

ctgcgggaac actatggact gacagacgcc cagatcgaca cactgctggg cctgtctttt 1200
gaggacggcc acatgagact ggggagaagc gctctgctga gagtccctgga tgccttgaa 1260
tccggaagag atgagcaggg actgccccctg tcctacgata aggcctgtgt ggctgccggc 1320
tatccagctc acacagccga tctgaaaaac ggcgagagag atgcactgcc ctactacggc 1380
gagctgctgt ggcggtatac acaggatgcc cctaccgcca agaacgacgc cgagagaaag 1440
ttcggcaaga tcgccaatcc taccgtgcac atcggcctga atcagctgag aaagcttgtc 1500
aatgccctga tccagagata cggcaagccc gctcagatcg tgggtggaact ggccagaaat 1560

ctgaaggctg gcctggaaga gaaagagcgg atcaagaaac agcagaccgc caacctggaa 1620
cggaacgaga gaatccggca gaagctgcag gacgctggcg tgcccagaca cagagaaaac 1680
cggctgcgga tgcggctgtt cgaggaactc ggacaaggca atggactggg cacccttgc 1740
atctactccg gcagacagat cagcctgcag agactgttca gcaacgacgt gcaggtcgac 1800
cacatcctgc ctttcagcaa gacctggat gacagcttcg ccaacaaggt gctcgcccag 1860
cacgacgcca acagatacaa gggcaacaga ggccctttcg aggccttcgg agccaacaga 1920
gatggctacg cctgggacga cattagagcc agagcagccg tgctgccccg gaacaagaga 1980

aacagatttg ccgagacagc catgcaggac tggctgcaca acgagactga ctttctggct 2040
cggcagctga ccgataccgc ctaccttagc agagtggcca ggcagtacct gaccgccatc 2100
tgcagcaagg acgacgtgta cgttagcccc ggcagactga ctgccatgct gagagctaag 2160
tggggcctga acagagtgtt ggatggcgtg atggaagaac agggcagacc cgccgtgaag 2220
aaccgggatg atcacagaca ccacgccatc gacgccgtgg ttattggcgc cacagataga 2280
gccatgctgc aacaggtggc cacttggtgc gctagagcta gagaacagga cgccgaaagg 2340
ctgatcggcg acatgcctac gccttggcct aatttccttg aggacgtgcg ggctgccgtg 2400

gccagatgtg tggtttttca caagcccac cacggaccag aaggcggcct gcataacgat 2460
acagcctacg gcattgtggc cggaccattc gaggatggca gatacagagt gcggcaccgg 2520
gtgtccctgt tcgatctgaa acctggcgac ctgagcaacg tccgtgtgta tgctcctctg 2580
caagccgagc tggaacccat cttcgagcag gacgatgcca gggccagaga agtggtcttt 2640
acagccctgg ctgagcggtg cagacagcgg aaagtgtggc tggaagaact gatgagcgtg 2700
ctgcctatca gaccagagg cgaggacgga aagaccctgc cagatagcgc tccttacaag 2760

gcctacaagg gcgactccaa ctactgctat gagctgttca tcaatgagcg cggcagatgg 2820

gatggcgagc tgatctctac cttccgggcc aatcaggccg cttaccggcg gttcagaaat 2880

gaccagccca ggttcagaag atacaccgct ggcggtagac ccctgctgat gagactgtgt 2940

atcaacgact atatcgccgt gggcacagcc gccgagagga ccatcttttag agtgggtcaag 3000

atgagcgaga acaagatcac tctggccgag cacttcgaag gcggaaccct gaaacagagg 3060

gatgccgaca aggacgatcc cttcaagtat ctgacaaaga gccttggcgc tctgcgcgat 3120

ctgggagcta gaagaatctt cgtggacctg atcgccgcg tgctggaccc aggcattaag 3180

ggcgatccca agaagaagag gaaggtg 3207

<210> 28

<211> 1069

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 28

Met Ala Glu Lys Gln His Arg Trp Gly Leu Asp Ile Gly Thr Asn Ser

1 5 10 15

Ile Gly Trp Ala Val Ile Ala Leu Ile Glu Gly Arg Pro Ala Gly Leu

20 25 30

Val Ala Thr Gly Ser Arg Ile Phe Ser Asp Gly Arg Asn Pro Lys Asp

35 40 45

Gly Ser Ser Leu Ala Val Glu Arg Arg Gly Pro Arg Gln Met Arg Arg

50 55 60

Arg Arg Asp Arg Tyr Leu Arg Arg Arg Asp Arg Phe Met Gln Ala Leu

65 70 75 80

Ile Asn Val Gly Leu Met Pro Gly Asp Ala Ala Ala Arg Lys Ala Leu

85 90 95

Val Thr Glu Asn Pro Tyr Val Leu Arg Gln Arg Gly Leu Asp Gln Ala

100 105 110

Leu Thr Leu Pro Glu Phe Gly Arg Ala Leu Phe His Leu Asn Gln Arg

115	120	125	
Arg Gly Phe Gln Ser Asn Arg Lys Thr Asp Arg Ala Thr Ala Lys Glu			
130	135	140	
Ser Gly Lys Val Lys Asn Ala Ile Ala Ala Phe Arg Ala Gly Met Gly			
145	150	155	160
Asn Ala Arg Thr Val Gly Glu Ala Leu Ala Arg Arg Leu Glu Asp Gly			
165	170	175	
Arg Pro Val Arg Ala Arg Met Val Gly Gln Gly Lys Asp Glu His Tyr			
180	185	190	
Glu Leu Tyr Ile Ala Arg Glu Trp Ile Ala Gln Glu Phe Asp Ala Leu			
195	200	205	
Trp Ala Ser Gln Gln Arg Phe His Ala Glu Val Leu Ala Asp Ala Ala			
210	215	220	
Arg Asp Arg Leu Arg Ala Ile Leu Leu Phe Gln Arg Lys Leu Leu Pro			
225	230	235	240
Val Pro Val Gly Lys Cys Phe Leu Glu Pro Asn Gln Pro Arg Val Ala			
245	250	255	
Ala Ala Leu Pro Ser Ala Gln Arg Phe Arg Leu Met Gln Glu Leu Asn			
260	265	270	
His Leu Arg Val Met Thr Leu Ala Asp Lys Arg Glu Arg Pro Leu Ser			
275	280	285	
Phe Gln Glu Arg Asn Asp Leu Leu Ala Gln Leu Val Ala Arg Pro Lys			
290	295	300	
Cys Gly Phe Asp Met Leu Arg Lys Ile Val Phe Gly Ala Asn Lys Glu			
305	310	315	320
Ala Tyr Arg Phe Thr Ile Glu Ser Glu Arg Arg Lys Glu Leu Lys Gly			
325	330	335	
Cys Asp Thr Ala Ala Lys Leu Ala Lys Val Asn Ala Leu Gly Thr Arg			
340	345	350	
Trp Gln Ala Leu Ser Leu Asp Glu Gln Asp Arg Leu Val Cys Leu Leu			
355	360	365	

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

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

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

 Arg Leu Arg Met Arg Leu Phe Glu Glu Leu Gly Gln Gly Asn Gly Leu
 565 570 575
 Gly Thr Pro Cys Ile Tyr Ser Gly Arg Gln Ile Ser Leu Gln Arg Leu
 580 585 590
 Phe Ser Asn Asp Val Gln Val Asp His Ile Leu Pro Phe Ser Lys Thr
 595 600 605
 Leu Asp Asp Ser Phe Ala Asn Lys Val Leu Ala Gln His Asp Ala Asn

610	615	620	
Arg Tyr Lys Gly Asn Arg Gly Pro Phe Glu Ala Phe Gly Ala Asn Arg			
625	630	635	640
Asp Gly Tyr Ala Trp Asp Asp Ile Arg Ala Arg Ala Ala Val Leu Pro			
	645	650	655
Arg Asn Lys Arg Asn Arg Phe Ala Glu Thr Ala Met Gln Asp Trp Leu			
	660	665	670
His Asn Glu Thr Asp Phe Leu Ala Arg Gln Leu Thr Asp Thr Ala Tyr			
	675	680	685
Leu Ser Arg Val Ala Arg Gln Tyr Leu Thr Ala Ile Cys Ser Lys Asp			
	690	695	700
Asp Val Tyr Val Ser Pro Gly Arg Leu Thr Ala Met Leu Arg Ala Lys			
705	710	715	720
Trp Gly Leu Asn Arg Val Leu Asp Gly Val Met Glu Glu Gln Gly Arg			
	725	730	735
Pro Ala Val Lys Asn Arg Asp Asp His Arg His His Ala Ile Asp Ala			
	740	745	750
Val Val Ile Gly Ala Thr Asp Arg Ala Met Leu Gln Gln Val Ala Thr			
	755	760	765
Leu Ala Ala Arg Ala Arg Glu Gln Asp Ala Glu Arg Leu Ile Gly Asp			
	770	775	780
Met Pro Thr Pro Trp Pro Asn Phe Leu Glu Asp Val Arg Ala Ala Val			
785	790	795	800
Ala Arg Cys Val Val Ser His Lys Pro Asp His Gly Pro Glu Gly Gly			
	805	810	815
Leu His Asn Asp Thr Ala Tyr Gly Ile Val Ala Gly Pro Phe Glu Asp			
	820	825	830
Gly Arg Tyr Arg Val Arg His Arg Val Ser Leu Phe Asp Leu Lys Pro			
	835	840	845
Gly Asp Leu Ser Asn Val Arg Cys Asp Ala Pro Leu Gln Ala Glu Leu			
850	855	860	

Glu Pro Ile Phe Glu Gln Asp Asp Ala Arg Ala Arg Glu Val Ala Leu
865 870 875 880

Thr Ala Leu Ala Glu Arg Tyr Arg Gln Arg Lys Val Trp Leu Glu Glu
885 890 895

Leu Met Ser Val Leu Pro Ile Arg Pro Arg Gly Glu Asp Gly Lys Thr
900 905 910

Leu Pro Asp Ser Ala Pro Tyr Lys Ala Tyr Lys Gly Asp Ser Asn Tyr
915 920 925

Cys Tyr Glu Leu Phe Ile Asn Glu Arg Gly Arg Trp Asp Gly Glu Leu
930 935 940

Ile Ser Thr Phe Arg Ala Asn Gln Ala Ala Tyr Arg Arg Phe Arg Asn
945 950 955 960

Asp Pro Ala Arg Phe Arg Arg Tyr Thr Ala Gly Gly Arg Pro Leu Leu
965 970 975

Met Arg Leu Cys Ile Asn Asp Tyr Ile Ala Val Gly Thr Ala Ala Glu
980 985 990

Arg Thr Ile Phe Arg Val Val Lys Met Ser Glu Asn Lys Ile Thr Leu
995 1000 1005

Ala Glu His Phe Glu Gly Gly Thr Leu Lys Gln Arg Asp Ala Asp
1010 1015 1020

Lys Asp Asp Pro Phe Lys Tyr Leu Thr Lys Ser Pro Gly Ala Leu
1025 1030 1035

Arg Asp Leu Gly Ala Arg Arg Ile Phe Val Asp Leu Ile Gly Arg
1040 1045 1050

Val Leu Asp Pro Gly Ile Lys Gly Asp Pro Lys Lys Lys Arg Lys
1055 1060 1065

Val

<210> 29

<211> 3273

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 29

atgaagtacc acgtgggcat cgacgtgggc accttttctg ttggactggc cgccatcgaa	60
gtggacgatg ccggaatgcc tatcaagacc ctgagcctgg tgtcccacat ccacgattct	120
ggactggacc ccgacgagat caagagcgcc gttacaagac tggccagcag cggaatcgcc	180
agaagaacca gacggctgta ccggcggaag agaagaaggc tgcagcagct ggacaagttc	240
atccagagac aaggctggcc cgtgatcgag ctggaagatt acagcgaccc tctgtacccc	300
tggaaagtgc gggctgaact ggctgccagc tatatcgccg atgagaaaaga gcggggcgag	360
aagctgtctg tggccctgag acacattgcc agacacagag gatggcggaa cccctacgcc	420
aaggtgtcct ctctgtatct gcctgacggc cctagcgacg ctttcaaggc catcagagag	480
gaaatcaaga gagccagcgg ccagcctgtg cctgaaacag ctacagtggg ccagatggtc	540
accctgtgtg aactgggcac cctgaagtig agaggcgaag gcggagtgtc gtctgccaga	600
ctccagcaga gcgattacgc cagagagatc caagagattt gccggatgca agagatcggc	660
caagagctgt acagaaagat catcgatgtg gtgttcgccg ccgagtctcc taagggatct	720
gcctctagca gagtgggcaa agaccctctg cagcccggca agaataagagc cctgaaagcc	780
tccgatgcct tccagagata ccggatcgcc gctctgatcg gcaacctgag agtttagagt	840
gacggcgaga agaggattct gagcgtggaa gagaaaaacc tgggtttcga ccacctggtc	900
aatctgacct ctaagaaaga acccgagtgg gtcacaatcg ccgagatcct gggaatcgac	960
agaggccagc tgatcggaac cgccaccatg acagatgatg gcgaaagagc cggcgctcgg	1020
cctcctacac atgacaccaa tcggagcatc gtgaacagca gaatcgcccc tctggtagac	1080
tgggtgaaaa ccgcctctgc tctggaacag cacgtatgg tcaaggccct gtccaatgcc	1140
gaggtggacg acttcgattc tcctgagggc gccaaagtgc aggccttctt tgccgacctg	1200
gacgacgatg tgcacgcaa gctggatagc ctgcatctgc ctgttggcag agccgcctac	1260
agcgaggata cacttgtcgc gctgaccaga cggatgtga gtgatggcgt ggacctgtac	1320
accgccagac tgcaagagtt tggcatcgag cctagctgga cccctccaac acctagaatc	1380
ggagagcccc tgggaaaccc cgctgtggac agagtgtga aaaccgtgtc cagatggctg	1440
gaaagcgcca ccaaacatg gggcgctccc gagagagtga tcatcgaaca cgtgcgcgag	1500
ggcttcgtga ccgagaaaag ggccagagaa atggatggcg acatcgagg aagggccgcc	1560

agaaatgcc a gctgttcca agaaatgcaa gaaaagctga acgtgcaggg caagccctcc 1620
 agagccgacc tttagagata ccagagcgtg cagagacaga actgccagtg cgcctactgt 1680
 ggagcccta tcaccttcag caacagcgag atggaccaca tcgtgcctag agccggccag 1740
 ggatccacca acaccagaga aaatctgggt gccgtgtgcc acagatgcaa ccagagcaag 1800
 ggcaacaccc cattcgccat ctgggccaag aacacctcta tcgagggcgt gtccgtgaaa 1860
 gaagccgtgg aaagaaccag gcaactgggt accgataccg gcatgagaag caccgacttc 1920
 aagaaattca ccaaggccgt ggtggaacgg ttccagaggg ccacaatgga cgaggaaatt 1980

 gacgcccga gcatggaaag cgtggcctgg atggccaatg agctgagaag tagagtggcc 2040
 cagcacttcg ccagccacgg cacaacagtc agagtgtaca gaggcagcct gaccgccgaa 2100
 gctcgtagag cctctggaat cagcggcaag ctgaagtctt ttgacggcgt gggcaagagc 2160
 agactggaca gaaggcacca cgccattgat gccgccgtga tcgccttcac cagcgactat 2220
 gtggccgaaa cactggccgt gcggagcaac ctcaaacaga gccaggtca cagacaagag 2280
 gctcctcagt ggcgcgagtt cacaggcaaa gatgccgaac acagagccgc ttggagagtg 2340
 tgggtgccaga agatggaaaa actgagcgcc ctgctgaccg aggacctgag agatgataga 2400

 gtggtggtca tgagcaacgt gcgcctgaga ctcggaatg gcagcgcca caaagagaca 2460
 atcgaaagc tgagcaaagt gaagctgtcc agccagctga gcgtgtccga catcgataag 2520
 gccagctctg aggccctttg gtgcgccctg acaagagaac ctggcttcga ccccaaagag 2580
 ggactgcctg ccaatcctga gcggcacatc agagtgaatg gcacccatgt gtacgccggc 2640
 gacaacatcg gctgtttcc agtgtctgcc ggatctatcg ctctgagagg cggatatgcc 2700
 gagctgggca gctctttcca tcacgccagg gtgtacaaga tcacaagcgg caagaaaccc 2760
 gcctttgcc tgcctgagagt gtataccatc gacctgtgc cttaccgaa ccaggacctg 2820

 ttcagcgtgg aactgaagcc ccagaccatg agcatgagac aggcggagaa gaagctgagg 2880
 gacgccctgg ctacaggcaa cgccgaatat ctggatggc tgggtgtgga tgacgagctg 2940
 gtggtcgata ccagcaaat cgccaccgac caagtgaagg ctgtggaagc cgaactggga 3000
 accatcagac gttggcgcgt ggacggcttt ttcagccct ctaagctgag actgcggccc 3060
 ctgcagatga gaaagaggg catcaagaaa gagagcgccc ctgagctgtc caagatcatt 3120
 gacagacctg gctggctgcc gcgcgtgaac aagctgttt ctgacggcaa cgtgaccgtc 3180
 gtgcggagag attctctggg cagagtgcgc ctggaaagca cagcacatct gcccgtgaca 3240

 tggaaaggtgc agcccaagaa gaagaggaag gtg 3273

<210> 30

<211> 1091

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic polypeptide"

<400> 30

Met Lys Tyr His Val Gly Ile Asp Val Gly Thr Phe Ser Val Gly Leu

1 5 10 15

Ala Ala Ile Glu Val Asp Asp Ala Gly Met Pro Ile Lys Thr Leu Ser

20 25 30

Leu Val Ser His Ile His Asp Ser Gly Leu Asp Pro Asp Lys Ile Lys

35 40 45

Ser Ala Val Thr Arg Leu Ala Ser Ser Gly Ile Ala Arg Arg Thr Arg

50 55 60

Arg Leu Tyr Arg Arg Lys Arg Arg Arg Leu Gln Gln Leu Asp Lys Phe

65 70 75 80

Ile Gln Arg Gln Gly Trp Pro Val Ile Glu Leu Glu Asp Tyr Ser Asp

85 90 95

Pro Leu Tyr Pro Trp Lys Val Arg Ala Glu Leu Ala Ala Ser Tyr Ile

100 105 110

Ala Asp Glu Lys Glu Arg Gly Glu Lys Leu Ser Val Ala Leu Arg His

115 120 125

Ile Ala Arg His Arg Gly Trp Arg Asn Pro Tyr Ala Lys Val Ser Ser

130 135 140

Leu Tyr Leu Pro Asp Glu Pro Ser Asp Ala Phe Lys Ala Ile Arg Glu

145 150 155 160

Glu Ile Lys Arg Ala Ser Gly Gln Pro Val Pro Glu Thr Ala Thr Val

165 170 175

Gly Gln Met Val Thr Leu Cys Glu Leu Gly Thr Leu Lys Leu Arg Gly

180 185 190

Glu Gly Gly Val Leu Ser Ala Arg Leu Gln Gln Ser Asp His Ala Arg

195 200 205

Glu Ile Gln Glu Ile Cys Arg Met Gln Glu Ile Gly Gln Glu Leu Tyr
 210 215 220

Arg Lys Ile Ile Asp Val Val Phe Ala Ala Glu Ser Pro Lys Gly Ser
 225 230 235 240

Ala Ser Ser Arg Val Gly Lys Asp Pro Leu Gln Pro Gly Lys Asn Arg
 245 250 255

Ala Leu Lys Ala Ser Asp Ala Phe Gln Arg Tyr Arg Ile Ala Ala Leu
 260 265 270

Ile Gly Asn Leu Arg Val Arg Val Asp Gly Glu Lys Arg Ile Leu Ser
 275 280 285

Val Glu Glu Lys Asn Leu Val Phe Asp His Leu Val Asn Leu Ala Pro
 290 295 300

Lys Lys Glu Pro Glu Trp Val Thr Ile Ala Glu Ile Leu Gly Ile Asp
 305 310 315 320

Arg Gly Gln Leu Ile Gly Thr Ala Thr Met Thr Asp Asp Gly Glu Arg
 325 330 335

Ala Gly Ala Arg Pro Pro Thr His Asp Thr Asn Arg Ser Ile Val Asn
 340 345 350

Ser Arg Ile Ala Pro Leu Val Asp Trp Trp Lys Thr Ala Ser Ala Leu
 355 360 365

Glu Gln His Ala Met Val Lys Ala Leu Ser Asn Ala Glu Val Asp Asp
 370 375 380

Phe Asp Ser Pro Glu Gly Ala Lys Val Gln Ala Phe Phe Ala Asp Leu
 385 390 395 400

Asp Asp Asp Val His Ala Lys Leu Asp Ser Leu His Leu Pro Val Gly
 405 410 415

Arg Ala Ala Tyr Ser Glu Asp Thr Leu Val Arg Leu Thr Arg Arg Met
 420 425 430

Leu Ala Asp Gly Val Asp Leu Tyr Thr Ala Arg Leu Gln Glu Phe Gly
 435 440 445

Ile Glu Pro Ser Trp Thr Pro Pro Ala Pro Arg Ile Gly Glu Pro Val

450 455 460
 Gly Asn Pro Ala Val Asp Arg Val Leu Lys Thr Val Ser Arg Trp Leu
 465 470 475 480

 Glu Ser Ala Thr Lys Thr Trp Gly Ala Pro Glu Arg Val Ile Ile Glu
 485 490 495
 His Val Arg Glu Gly Phe Val Thr Glu Lys Arg Ala Arg Glu Met Asp
 500 505 510
 Gly Asp Met Arg Arg Arg Ala Ala Arg Asn Ala Lys Leu Phe Gln Glu
 515 520 525
 Met Gln Glu Lys Leu Asn Val Gln Gly Lys Pro Ser Arg Ala Asp Leu
 530 535 540

 Trp Arg Tyr Gln Ser Val Gln Arg Gln Asn Cys Gln Cys Ala Tyr Cys
 545 550 555 560
 Gly Ser Pro Ile Thr Phe Ser Asn Ser Glu Met Asp His Ile Val Pro
 565 570 575
 Arg Ala Gly Gln Gly Ser Thr Asn Thr Arg Glu Asn Leu Val Ala Val
 580 585 590
 Cys His Arg Cys Asn Gln Ser Lys Gly Asn Thr Pro Phe Ala Ile Trp
 595 600 605

 Ala Lys Asn Thr Ser Ile Glu Gly Val Ser Val Lys Glu Ala Val Glu
 610 615 620
 Arg Thr Arg His Trp Val Thr Asp Thr Gly Met Arg Ser Thr Asp Phe
 625 630 635 640
 Lys Lys Phe Thr Lys Ala Val Val Glu Arg Phe Gln Arg Ala Thr Met
 645 650 655
 Asp Glu Glu Ile Asp Ala Arg Ser Met Glu Ser Val Ala Trp Met Ala
 660 665 670

 Asn Glu Leu Arg Ser Arg Val Ala Gln His Phe Ala Ser His Gly Thr
 675 680 685
 Thr Val Arg Val Tyr Arg Gly Ser Leu Thr Ala Glu Ala Arg Arg Ala
 690 695 700

Ser Gly Ile Ser Gly Lys Leu Glu Phe Leu Asp Gly Val Gly Lys Ser
705 710 715 720
Arg Leu Asp Arg Arg His His Ala Ile Asp Ala Ala Val Ile Ala Phe
725 730 735

Thr Ser Asp Tyr Val Ala Glu Thr Leu Ala Val Arg Ser Asn Leu Lys
740 745 750
Gln Ser Gln Ala His Arg Gln Glu Ala Pro Gln Trp Arg Glu Phe Thr
755 760 765
Gly Lys Asp Ala Glu His Arg Ala Ala Trp Arg Val Trp Cys Gln Lys
770 775 780
Met Glu Lys Leu Ser Ala Leu Leu Thr Glu Asp Leu Arg Asp Asp Arg
785 790 795 800

Val Val Val Met Ser Asn Val Arg Leu Arg Leu Gly Asn Gly Ser Ala
805 810 815
His Glu Glu Thr Ile Gly Lys Leu Ser Lys Val Lys Leu Gly Ser Gln
820 825 830
Leu Ser Val Ser Asp Ile Asp Lys Ala Ser Ser Glu Ala Leu Trp Cys
835 840 845
Ala Leu Thr Arg Glu Pro Asp Phe Asp Pro Lys Asp Gly Leu Pro Ala
850 855 860

Asn Pro Glu Arg His Ile Arg Val Asn Gly Thr His Val Tyr Ala Gly
865 870 875 880
Asp Asn Ile Gly Leu Phe Pro Val Ser Ala Gly Ser Ile Ala Leu Arg
885 890 895
Gly Gly Tyr Ala Glu Leu Gly Ser Ser Phe His His Ala Arg Val Tyr
900 905 910
Lys Ile Thr Ser Gly Lys Lys Pro Ala Phe Ala Met Leu Arg Val Tyr
915 920 925

Thr Ile Asp Leu Leu Pro Tyr Arg Asn Gln Asp Leu Phe Ser Val Glu
930 935 940
Leu Lys Pro Gln Thr Met Ser Met Arg Gln Ala Glu Lys Lys Leu Arg

945 950 955 960
 Asp Ala Leu Ala Thr Gly Asn Ala Glu Tyr Leu Gly Trp Leu Val Val
 965 970 975
 Asp Asp Glu Leu Val Val Asp Thr Ser Lys Ile Ala Thr Asp Gln Val
 980 985 990

Lys Ala Val Glu Ala Glu Leu Gly Thr Ile Arg Arg Trp Arg Val Asp
 995 1000 1005
 Gly Phe Phe Gly Asp Thr Arg Leu Arg Leu Arg Pro Leu Gln Met
 1010 1015 1020
 Ser Lys Glu Gly Ile Lys Lys Glu Ser Ala Pro Glu Leu Ser Lys
 1025 1030 1035
 Ile Ile Asp Arg Pro Gly Trp Leu Pro Ala Val Asn Lys Leu Phe
 1040 1045 1050

Ser Glu Gly Asn Val Thr Val Val Arg Arg Asp Ser Leu Gly Arg
 1055 1060 1065
 Val Arg Leu Glu Ser Thr Ala His Leu Pro Val Thr Trp Lys Val
 1070 1075 1080
 Gln Pro Lys Lys Lys Arg Lys Val
 1085 1090

<210> 31

<211> 160

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"

<220><221>

modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 31

nnnnnnnnnn nnnnnnnnnn gucauaguuc cccuaagauu auugaaacaa ugaucuuagg 60

guuacuauga uaagggcuuu cuacuuuagg gguagagaug ucccgcggcg ugggggaucg 120

ccuauugccc uaaaagggca cuccccauuu uaauuuuuuu 160

<210> 32

<211> 166

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222>

> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 32

nnnnnnnnnn nnnnnnnnnn gucucaggua gaugucagau caaucagaaa ugauugaucu 60

gacaucuacg aguugagauca aaacaaagcu ucagcugagu uucaauuuucu gagcccaugu 120

ugggcgauac auaugccacc cgagugcaaa ucggguggcgcu uuuuuu 166

<210> 33

<211> 170

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 33

nnnnnnnnnn nnnnnnnnnn guuucaguag uuguuagaag aaugaaaauu cuuuuaacaa 60

cgaagucgcc uucgggcgag cugaaaacaa uuugauuuaa uauuagaucg ggcucacugag 120

gucuugacc uuaucgggau uaacgaagag ccuccgagga ggcuuuuuuu 170

<210> 34

<211> 127

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 34

nnnnnnnnnn nnnnnnnnnn guuuuagugu uguacaauau uugggugaaa acccaauau 60

uguacauccu aaaucaaggc gcuuaauugc ugccguauuu gcugaaagcg uagcuuucag 120

uuuuuuuu 127

<210> 35

<211> 152

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 35

nnnnnnnnnn nnnnnnnnnn guuuuagcac uguacaauac uuguguaagc aaauacgaaa 60

auuauugcuu acacaauuau ugucgugcua aaauaaggcg cuguuaaagc agcugccgca 120

uccgccagag cauuuauugcu cuggcuuuuu uu 152

<210> 36

<211> 147

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other
 <400> 36
 nnnnnnnnnn nnnnnnnnnn guuuugccuu gaauccaaag uaaggcaugg uagaaauuu 60
 auuccugugg auucaagaca aaauuugaaa ugcaaaccga uuccccggcu gcaagccagc 120
 cacaccgguc uuucaaagca uuuuuuuu 147

<210> 37
 <211> 120
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <220><221> modified_base

<222> (1)..(20)
 <223> a, c, u, g, unknown or other
 <400> 37
 nnnnnnnnnn nnnnnnnnnn guuuugccuu gaauccaaaa cggaaucaag acaaaauuug 60
 aaaugcaaac cgauuuuccu gacugccagc cagucacacc gguaacaaaa gcauuuuuuu 120

<210> 38
 <211> 186
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"

<220><221> modified_base
 <222> (1)..(20)
 <223> a, c, u, g, unknown or other
 <400> 38
 nnnnnnnnnn nnnnnnnnnn gcuucagaug ugugucagau caaugagaaa ucauugaucu 60
 gacacacagc auugaagua agcaagauua auuucagcu uaaauuuccu cacauuuuuu 120
 gugcagaagg gcuuaugccc acaauacaua aaaaguccgc auucacuugc ggacuuuuuu 180
 uuuuuuuu 186

<210> 39

<211> 138

<212> RNA

<213> Artificial Sequence

<220><221> source

<223>

/note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 39

nnnnnnnnnn nnnnnnnnnn guuucaaaau caaucuaaag cgaaagcuau acuuauuuuu	60
gaauuugaaa uaaggcuguu ccuucguuag uucagucgau ugcuccuccg guauugcuua	120
ugcaugccgg aguuuuuu	138

<210> 40

<211> 118

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 40

nnnnnnnnnn nnnnnnnnnn gcuggggagc cugucugaaa agacaggcua ccuagcaaga	60
cccuucgug gggucgcauu cuucaccccc ugcagcagc gagggguuc guuuuuuu	118

<210> 41

<211> 144

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

```

        polynucleotide"
<220><221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other

<400> 41
nnnnnnnnnn nnnnnnnnnn gucauaguuc ccucacaagc cucgaugugg aaacacauca      60
aggcuugcga gguugcuaug auaaggcaac aggccgcaaa gcacugaccc gcauuccaau      120
gaaugcgggu caucuacuuu uuuu                                           144

<210> 42
<211> 129
<212> RNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
        polynucleotide"
<220><221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other
<400> 42
nnnnnnnnnn nnnnnnnnnn guuucacagg cuaagcggau uugcgaaagc aaauccguuc      60

gaugccuuga aaucacaaa aagauuaau agacccgccc acuguauugu acauggcggg      120
acuuuuuuu                                           129

<210> 43
<211> 156
<212> RNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
        polynucleotide"
<220><221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other
<400> 43

```

nnnnnnnnnn nnnnnnnnnn guuauaagac ccucaaaaac cccaccuugu uacaauguug 60

uaacagggua ggguuuuuug aggggucuaa uaaucaagaa cuguuacaac aguuccauuc 120

uagggcccau cuucggacgg gccucagccu uuuuuu 156

<210> 44

<211> 155

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 44

nnnnnnnnnn nnnnnnnnnn guuguagcca gagcgcaauu cccgaucugc ugaaaagcag 60

aucgggaauu gcgcuuugcu acuaacaagc ugaauccguu aggaguaaa ugcaccaaag 120

agagggccgg cuuaugccgg cccuuugcuu uuuuu 155

<210> 45

<211> 171

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<220><221> modified_base

<222> (1)..(20)

<223> a, c, u, g, unknown or other

<400> 45

nnnnnnnnnn nnnnnnnnnn acugggguuc aguucuaaaa aaccugaua gacuucgaaa 60

agucacuaac uuaauuaaa agaacugaac cucaguaagc auuggcucgu uuccaauguu 120

gauugcuccg ccggugcucc uuauuuuuu gggcgccggc uuuuuuuuu u 171

<210> 46

<211> 28

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<220><221> modified_base

<222> (21)..(28)

<223> a, c, t, g, unknown or other

<400> 46

gtacaaacgg cagaagctgg nnnnnnnn

28

<210> 47

<211> 27

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 47

ctgtttcctg ggtccaata acaagac

27

<210> 48

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 48

acaagggtc aagtgagtg

20

<210> 49

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 49

atgggagcag ctggtcagag 20

<210> 50

<211> 19

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 50

cagccattg cttgtcct 19

<210> 51

<211> 23

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 51

ctgccattc cagccactca aac 23

<210> 52

<211> 23

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 52

ggttaagtcg ggtttccttg cag 23

<210> 53

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 primer"
 <400> 53
 ttcgggtcac ctctcactcc 20
 <210> 54
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source

 <223> /note="Description of Artificial Sequence: Synthetic
 primer"
 <400> 54
 ggctccatcg taagcaaacc 20
 <210> 55
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 primer"
 <400> 55
 ctcccctgct tctgtcgta t 21
 <210> 56
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic

 primer"

<400> 56
acaggtcgtg gacactcaca 20
<210> 57
<211> 20
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400> 57
ggatcaagtc aaggcatgt 20
<210> 58
<211> 20
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"
<400>
> 58
atgtagctgg acaggcttgg 20
<210> 59
<211> 22
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 59
ctcatacatg gctataatag aa 22
<210> 60
<211> 22
<212> DNA
<213> Artificial Sequence
<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400>

60

gaaggtgtgg ttccagaacc gg 22

<210> 61

<211> 22

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 61

tggttccaga accggaggac aa 22

<210> 62

<211> 22

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 62

cccaggtgaa ggtgtggttc ca 22

<210> 63

<211> 22

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 63

agaaccggag gacaaagtac aa 22

<210> 64

<211> 20

```
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
        oligonucleotide"
<400> 64
cctctggcac atcctccaaa
```

```
<210> 65
<211> 20
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
        oligonucleotide"
<400> 65
ctagggacag gattggtgac
```

```
<210> 66
<211> 19
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
        oligonucleotide"
<400> 66
gctcgtactg gcgaatgct
```

```
<
210> 67
<211> 20
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
        oligonucleotide"
<400> 67
```

gctgaagagc tacgagaacc 20

<210> 68

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 68

catgggggag atgggccggc 20

<210>

> 69

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 69

agagactctc tagaaggac 20

<210> 70

<211> 21

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 70

gtgagagtct cctccccaat g 21

<210>

71

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 71

gggaggagac tctcacctga

20

<210> 72

<211> 100

<212> PRT

<213> Homo sapiens

<400> 72

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu

1 5 10 15

Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val

20 25 30

Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys

35 40 45

Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu

50 55 60

Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu

65 70 75 80

Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu

85 90 95

Ala Lys Ser Asp

100

<210> 73

<211> 84

<212> PRT

<213> Homo sapiens

<400> 73

Met Gly Lys Gly Asp Pro Lys Lys Pro Arg Gly Lys Met Ser Ser Tyr

1 5 10 15

Ala Phe Phe Val Gln Thr Cys Arg Glu Glu His Lys Lys Lys His Pro

20 25 30

Asp Ala Ser Val Asn Phe Ser Glu Phe Ser Lys Lys Cys Ser Glu Arg

35 40 45

Trp Lys Thr Met Ser Ala Lys Glu Lys Gly Lys Phe Glu Asp Met Ala

50 55 60

Lys Ala Asp Lys Ala Arg Tyr Glu Arg Glu Met Lys Thr Tyr Ile Pro

65 70 75 80

Pro Lys Gly Glu

<210> 74

<211> 80

<212> PRT

<213> Homo sapiens

<400> 74

Ser Thr Asp His Pro Lys Tyr Ser Asp Met Ile Val Ala Ala Ile Gln

1 5 10 15

Ala Glu Lys Asn Arg Ala Gly Ser Ser Arg Gln Ser Ile Gln Lys Tyr

20 25 30

Ile Lys Ser His Tyr Lys Val Gly Glu Asn Ala Asp Ser Gln Ile Lys

35 40 45

Leu Ser Ile Lys Arg Leu Val Thr Thr Gly Val Leu Lys Gln Thr Lys

50 55 60

Gly Val Gly Ala Ser Gly Ser Phe Arg Leu Ala Lys Ser Asp Glu Pro

65 70 75 80

<210> 75

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 75

tgtacatggg ggagatgggc

<210> 76

20

<211> 107

<212> RNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 76

guucuagugu uguacaauau uugggugaaa acccaauau uguacauccu agaucaaggc 60

gcuaauugc ugccguauu gcugaaagcg uagcuuucag uuuuuuu 107

<210> 77

<211> 22

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 77

atagatgcgg ccaaggtgta ca 22

<210> 78

<211> 7

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 78

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 79

<211> 7

<212>

PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 79

Pro Lys Lys Lys Arg Arg Val

1 5

<210> 80

<211> 16

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 80

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210>

> 81

<211> 11

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 81

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5 10

<210> 82

<211> 9

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 82

Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5

<210> 83

<211> 9

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 83

Pro Ala Ala Lys Arg Val Lys Leu Asp

1 5

<210> 84

<211> 11

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 84

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro

1 5 10

<210> 85

<211>

> 8

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 85

Val Ser Arg Lys Arg Pro Arg Pro

1 5

<210> 86

<211> 8

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 86

Pro Pro Lys Lys Ala Arg Glu Asp

1 5

<210> 87

<211> 8

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 87

Pro Gln Pro Lys Lys Lys Pro Leu

1 5

<210> 88

<211> 12

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 88

Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro

1 5 10

<210> 89

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 89

Pro Lys Gln Lys Lys Arg Lys

1 5

<210> 90

<211> 10

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 90

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu

1 5 10

<210> 91

<211> 10

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

peptide"

<400> 91

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg

1 5 10

<210> 92

<211> 20

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 92

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys

1 5 10 15

Lys Ser Lys Lys

20

<210> 93

<211> 17

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 93

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys

1 5 10 15

Lys

<210> 94

<211> 38

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 94

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly

1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro

20 25 30

Arg Asn Gln Gly Gly Tyr

35

<210> 95

<211> 42

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 95

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu

1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys

20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val

35 40

<210> 96

<211> 20

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 96

Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Pro Lys Lys

1 5 10 15

Lys Arg Lys Val

20

<210> 97

<211> 19

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 97

Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Pro Lys Lys Lys

1 5 10 15

Arg Lys Val

<210> 98

<211> 24

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 98

Gly Ala Leu Phe Leu Gly Trp Leu Gly Ala Ala Gly Ser Thr Met Gly

1 5 10 15

Ala Pro Lys Lys Lys Arg Lys Val

20

<210> 99

<211> 27

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 99

Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly

1 5 10 15

Ala Trp Ser Gln Pro Lys Lys Lys Arg Lys Val

20

25

<210> 100

<211> 21

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 100

Lys Glu Thr Trp Trp Glu Thr Trp Trp Thr Glu Trp Ser Gln Pro Lys

1 5 10 15

Lys Lys Arg Lys Val

20

<210> 101

<211> 11

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<

400> 101

Tyr Ala Arg Ala Ala Ala Arg Gln Ala Arg Ala

1 5 10

<210> 102

<211> 11

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 102

Thr His Arg Leu Pro Arg Arg Arg Arg Arg Arg

1 5 10

<210> 103

<211> 11

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

peptide"

<400> 103

Gly Gly Arg Arg Ala Arg Arg Arg Arg Arg

1 5 10

<210> 104

<211> 12

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 104

Arg Arg Gln Arg Arg Thr Ser Lys Leu Met Lys Arg

1 5 10

<210> 105

<211> 27

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 105

Gly Trp Thr Leu Asn Ser Ala Gly Tyr Leu Leu Gly Lys Ile Asn Leu

1 5 10 15

Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu

20 25

<210> 106

<211> 33

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic polypeptide"

<400> 106

Lys Ala Leu Ala Trp Glu Ala Lys Leu Ala Lys Ala Leu Ala Lys Ala

1 5 10 15

Leu Ala Lys His Leu Ala Lys Ala Leu Ala Lys Ala Leu Lys Cys Glu

20 25 30

Ala

<210> 107

<211> 16

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 107

Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys

1 5 10 15

<210> 108

<211> 6

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 108

Leu Glu Gly Gly Gly Ser

1 5

<210> 109

<211> 4

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 109

Thr Gly Ser Gly

1

<210> 110
 <211> 8
 <212> PRT
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 peptide"
 <400> 110
 Gly Gly Ser Gly Gly Gly Ser Gly
 1 5
 <210> 111
 <211> 20
 <212> PRT
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 peptide"
 <220><221> SITE
 <222> (1)..(20)
 <223> /note="This sequence may encompass 1-4 'Gly Gly Gly Gly Ser'
 repeating units"
 <400> 111
 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
 1 5 10 15

 Gly Gly Gly Ser
 20
 <210> 112
 <211> 8
 <212> PRT
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 peptide"

<220><221> SITE

<222> (1)..(8)

<223> /note="This sequence may encompass 6-8 residues"

<400> 112

Gly Gly Gly Gly Gly Gly Gly Gly

1 5

<210> 113

<211> 20

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

peptide"

<220><221> SITE

<222> (1)..(20)

<223> /note="This sequence may encompass 1-4 'Glu Ala Ala Ala Lys'

repeating units"

<400> 113

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu

1 5 10 15

Ala Ala Ala Lys

20

<210> 114

<211> 27

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

peptide"

<220><221> SITE

<222>

(2)..(26)

<223> /note="This region may encompass 2-5 'Glu Ala Ala Ala Lys'

repeating units"

<400> 114

Ala Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys

1 5 10 15

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Ala

20 25

<210> 115

<211> 5

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 115

Pro Ala Pro Ala Pro

1 5

<210> 116

<211> 16

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<220><221> SITE

<222> (1)..(16)

<223> /note="This sequence may encompass 6-8 'Ala Pro' repeating units"

<400> 116

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

1 5 10 15

<210> 117

<211> 1629

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic polypeptide"

<400> 117

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu

1 5 10 15

Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val

20 25 30

Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys

35 40 45

Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu

50 55 60

Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu

65 70 75 80

Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu

85 90 95

Ala Lys Ser Asp Thr Gly Ser Gly Met Gly Lys Thr His Ile Ile Gly

100 105 110

Val Gly Leu Asp Leu Gly Gly Thr Tyr Thr Gly Thr Phe Ile Thr Ser

115 120 125

His Pro Ser Asp Glu Ala Glu His Arg Asp His Ser Ser Ala Phe Thr

130 135 140

Val Val Asn Ser Glu Lys Leu Ser Phe Ser Ser Lys Ser Arg Thr Ala

145 150 155 160

Val Arg His Arg Val Arg Ser Tyr Lys Gly Phe Asp Leu Arg Arg Arg

165 170 175

Leu Leu Leu Leu Val Ala Glu Tyr Gln Leu Leu Gln Lys Lys Gln Thr

180 185 190

Leu Ala Pro Glu Glu Arg Glu Asn Leu Arg Ile Ala Leu Ser Gly Tyr

195 200 205

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

210 215 220

Val Leu Glu Ser Leu Asp Pro Ser Val Phe Ser Ser Ala Pro Ser Phe
225 230 235 240

Thr Asn Phe Phe Asn Asp Ser Glu Pro Leu Asn Ile Gln Trp Glu Ala
245 250 255

Ile Ala Asn Ser Pro Glu Thr Thr Lys Ala Leu Asn Lys Glu Leu Ser
260 265 270

Gly Gln Lys Glu Ala Asp Phe Lys Lys Tyr Ile Lys Thr Ser Phe Pro
275 280 285

Glu Tyr Ser Ala Lys Glu Ile Leu Ala Asn Tyr Val Glu Gly Arg Arg
290 295 300

Ala Ile Leu Asp Ala Ser Lys Tyr Ile Ala Asn Leu Gln Ser Leu Gly
305 310 315 320

His Lys His Arg Ser Lys Tyr Leu Ser Asp Ile Leu Gln Asp Met Lys
325 330 335

Arg Asp Ser Arg Ile Thr Arg Leu Ser Glu Ala Phe Gly Ser Thr Asp
340 345 350

Asn Leu Trp Arg Ile Ile Gly Asn Ile Ser Asn Leu Gln Glu Arg Ala
355 360 365

Val Arg Trp Tyr Phe Asn Asp Ala Lys Phe Glu Gln Gly Gln Glu Gln
370 375 380

Leu Asp Ala Val Lys Leu Lys Asn Val Leu Val Arg Ala Leu Lys Tyr
385 390 395 400

Leu Arg Ser Asp Asp Lys Glu Trp Ser Ala Ser Gln Lys Gln Ile Ile
405 410 415

Gln Ser Leu Glu Gln Ser Gly Asp Val Leu Asp Val Leu Ala Gly Leu
420 425 430

Asp Pro Asp Arg Thr Ile Pro Pro Tyr Glu Asp Gln Asn Asn Arg Arg
435 440 445

Pro Pro Glu Asp Gln Thr Leu Tyr Leu Asn Pro Lys Ala Leu Ser Ser
450 455 460

Glu Tyr Gly Glu Lys Trp Lys Ser Trp Ala Asn Lys Phe Ala Gly Ala

465 470 475 480
 Tyr Pro Leu Leu Thr Glu Asp Leu Thr Glu Ile Leu Lys Asn Thr Asp
 485 490 495

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

 Leu Glu Glu Phe Leu Glu Phe Ala Asn Arg Tyr Tyr Gln Glu Thr Ala
 565 570 575
 Lys Ala Lys Asn Gly Leu Trp Phe Pro Glu Asn Ala Leu Leu Glu Arg
 580 585 590
 Ala Asp Leu His Pro Pro Met Lys Asn Lys Ile Leu Asn Val Ile Val
 595 600 605
 Gly Gln Ala Leu Gly Val Ser Pro Ala Glu Gly Thr Asp Phe Ile Glu
 610 615 620

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

 Val Ser Glu Val Val Pro Phe Ile Gly Lys Glu Leu Arg Leu Ser Asp
 690 695 700
 Glu Ala Gln Ser Lys Phe Asp Asn Leu Tyr Ser Leu Ala Gln Leu Tyr
 705 710 715 720

Asn Leu Ile Glu Thr Glu Arg Asn Gly Phe Ser Lys Val Ser Leu Ala
 725 730 735
 Ala His Leu Glu Asn Ala Trp Arg Met Thr Met Thr Asp Gly Ser Ala
 740 745 750

 Gln Cys Cys Arg Leu Pro Ala Asp Cys Val Arg Pro Phe Asp Gly Phe
 755 760 765
 Ile Arg Lys Ala Ile Asp Arg Asn Ser Trp Glu Val Ala Lys Arg Ile
 770 775 780
 Ala Glu Glu Val Lys Lys Ser Val Asp Phe Thr Asn Gly Thr Val Lys
 785 790 795 800
 Ile Pro Val Ala Ile Glu Ala Asn Ser Phe Asn Phe Thr Ala Ser Leu
 805 810 815

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

 Arg Ser Leu Thr Leu Lys Lys Ser Glu Ser Ile Tyr Asn Ser Glu Val
 885 890 895
 Asn Leu Ile Phe Val Ser Ala Gln Gly Asn Gln Glu Lys Lys Asn Asn
 900 905 910
 Ile Tyr Leu Leu Ser Asn Leu Ala Lys Asn Tyr Leu Ala Ala Val Phe
 915 920 925
 Gly Thr Ser Asp Leu Ser Gln Ile Thr Asn Glu Ile Glu Ser Thr Val
 930 935 940

 Leu Gln Leu Lys Ala Ala Gly Arg Leu Gly Tyr Phe Asp Leu Leu Ser
 945 950 955 960
 Glu Lys Glu Arg Ala Cys Ala Arg His Ala Leu Phe Leu Asn Ser Asp

965	970	975
Ser Glu Ala Arg Arg Ala Val Ile Asp Val Leu Gly Ser Arg Arg Lys		
980	985	990
Ala Ser Val Asn Gly Thr Gln Ala Trp Phe Val Arg Ser Ile Phe Ser		
995	1000	1005
Lys Val Arg Gln Ala Leu Ala Ala Trp Thr Gln Glu Thr Gly Asn		
1010	1015	1020
Glu Leu Ile Phe Asp Ala Ile Ser Val Pro Ala Ala Asp Ser Ser		
1025	1030	1035
Glu Met Arg Lys Arg Phe Ala Glu Tyr Arg Pro Glu Phe Arg Lys		
1040	1045	1050
Pro Lys Val Gln Pro Val Ala Ser His Ser Ile Asp Ala Met Cys		
1055	1060	1065
Ile Tyr Leu Ala Ala Cys Ser Asp Pro Phe Lys Thr Lys Arg Met		
1070	1075	1080
Gly Ser Gln Leu Ala Ile Tyr Glu Pro Ile Asn Phe Asp Asn Leu		
1085	1090	1095
Phe Thr Gly Ser Cys Gln Val Ile Gln Asn Thr Pro Arg Asn Phe		
1100	1105	1110
Ser Asp Lys Thr Asn Ile Ala Asn Ser Pro Ile Phe Lys Glu Thr		
1115	1120	1125
Ile Tyr Ala Glu Arg Phe Leu Asp Ile Ile Val Ser Arg Gly Glu		
1130	1135	1140
Ile Phe Ile Gly Tyr Pro Ser Asn Met Pro Phe Glu Glu Lys Pro		
1145	1150	1155
Asn Arg Ile Ser Ile Gly Gly Lys Asp Pro Phe Ser Ile Leu Ser		
1160	1165	1170
Val Leu Gly Ala Tyr Leu Asp Lys Ala Pro Ser Ser Glu Lys Glu		
1175	1180	1185
Lys Leu Thr Ile Tyr Arg Val Val Lys Asn Lys Ala Phe Glu Leu		
1190	1195	1200

Phe Ser	Lys Val	Ala Gly	Ser	Lys Phe	Thr Ala	Glu	Glu Asp	Lys
1205			1210			1215		
Ala Ala	Lys Ile	Leu Glu	Ala	Leu His	Phe Val	Thr	Val Lys	Gln
1220			1225			1230		
Asp Val	Ala Ala	Thr Val	Ser	Asp Leu	Ile Lys	Ser	Lys Lys	Glu
1235			1240			1245		
Leu Ser	Lys Asp	Ser Ile	Glu	Asn Leu	Ala Lys	Gln	Lys Gly	Cys
1250			1255			1260		
Leu Lys	Lys Val	Glu Tyr	Ser	Ser Lys	Glu Phe	Lys	Phe Lys	Gly
1265			1270			1275		
Ser Leu	Ile Ile	Pro Ala	Ala	Val Glu	Trp Gly	Lys	Val Leu	Trp
1280			1285			1290		
Asn Val	Phe Lys	Glu Asn	Thr	Ala Glu	Glu Leu	Lys	Asp Glu	Asn
1295			1300			1305		
Ala Leu	Arg Lys	Ala Leu	Glu	Ala Ala	Trp Pro	Ser	Ser Phe	Gly
1310			1315			1320		
Thr Arg	Asn Leu	His Ser	Lys	Ala Lys	Arg Val	Phe	Ser Leu	Pro
1325			1330			1335		
Val Val	Ala Thr	Gln Ser	Gly	Ala Val	Arg Ile	Arg	Arg Lys	Thr
1340			1345			1350		
Ala Phe	Gly Asp	Phe Val	Tyr	Gln Ser	Gln Asp	Thr	Asn Asn	Leu
1355			1360			1365		
Tyr Ser	Ser Phe	Pro Val	Lys	Asn Gly	Lys Leu	Asp	Trp Ser	Ser
1370			1375			1380		
Pro Ile	Ile His	Pro Ala	Leu	Gln Asn	Arg Asn	Leu	Thr Ala	Tyr
1385			1390			1395		
Gly Tyr	Arg Phe	Val Asp	His	Asp Arg	Ser Ile	Ser	Met Ser	Glu
1400			1405			1410		
Phe Arg	Glu Val	Tyr Asn	Lys	Asp Asp	Leu Met	Arg	Ile Glu	Leu
1415			1420			1425		
Ala Gln	Gly Thr	Ser Ser	Arg	Arg Tyr	Leu Arg	Val	Glu Met	Pro

1430	1435	1440
Gly Glu Lys Phe Leu Ala Trp	Phe Gly Glu Asn Ser	Ile Ser Leu
1445	1450	1455
Gly Ser Ser Phe Lys Phe Ser	Val Ser Glu Val Phe	Asp Asn Lys
1460	1465	1470
Ile Tyr Thr Glu Asn Ala Glu	Phe Thr Lys Phe Leu	Pro Lys Pro
1475	1480	1485

Arg Glu Asp Asn Lys His Asn	Gly Thr Ile Phe Phe	Glu Leu Val
1490	1495	1500
Gly Pro Arg Val Ile Phe Asn	Tyr Ile Val Gly Gly	Ala Ala Ser
1505	1510	1515
Ser Leu Lys Glu Ile Phe Ser	Glu Ala Gly Lys Glu	Arg Ser Pro
1520	1525	1530
Lys Lys Lys Arg Lys Val Leu	Glu Gly Gly Gly Gly	Ser Gly Lys
1535	1540	1545

Gly Asp Pro Lys Lys Pro Arg	Gly Lys Met Ser Ser	Tyr Ala Phe
1550	1555	1560
Phe Val Gln Thr Cys Arg Glu	Glu His Lys Lys Lys	His Pro Asp
1565	1570	1575
Ala Ser Val Asn Phe Ser Glu	Phe Ser Lys Lys Cys	Ser Glu Arg
1580	1585	1590
Trp Lys Thr Met Ser Ala Lys	Glu Lys Gly Lys Phe	Glu Asp Met
1595	1600	1605

Ala Lys Ala Asp Lys Ala Arg	Tyr Glu Arg Glu Met	Lys Thr Tyr
1610	1615	1620
Ile Pro Pro Lys Gly Glu		
1625		

<210> 118

<211> 1626

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 118

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu

1 5 10 15

Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val

20 25 30

Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys

35 40 45

Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu

50 55 60

Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu

65 70 75 80

Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu

85 90 95

Ala Lys Ser Asp Thr Gly Ser Gly Met Gly Lys Thr His Ile Ile Gly

100 105 110

Val Gly Leu Asp Leu Gly Gly Thr Tyr Thr Gly Thr Phe Ile Thr Ser

115 120 125

His Pro Ser Asp Glu Ala Glu His Arg Asp His Ser Ser Ala Phe Thr

130 135 140

Val Val Asn Ser Glu Lys Leu Ser Phe Ser Ser Lys Ser Arg Thr Ala

145 150 155 160

Val Arg His Arg Val Arg Ser Tyr Lys Gly Phe Asp Leu Arg Arg Arg

165 170 175

Leu Leu Leu Leu Val Ala Glu Tyr Gln Leu Leu Gln Lys Lys Gln Thr

180 185 190

Leu Ala Pro Glu Glu Arg Glu Asn Leu Arg Ile Ala Leu Ser Gly Tyr

195 200 205

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

210 215 220

Val Leu Glu Ser Leu Asp Pro Ser Val Phe Ser Ser Ala Pro Ser Phe
225 230 235 240
Thr Asn Phe Phe Asn Asp Ser Glu Pro Leu Asn Ile Gln Trp Glu Ala
245 250 255
Ile Ala Asn Ser Pro Glu Thr Thr Lys Ala Leu Asn Lys Glu Leu Ser
260 265 270
Gly Gln Lys Glu Ala Asp Phe Lys Lys Tyr Ile Lys Thr Ser Phe Pro
275 280 285
Glu Tyr Ser Ala Lys Glu Ile Leu Ala Asn Tyr Val Glu Gly Arg Arg
290 295 300
Ala Ile Leu Asp Ala Ser Lys Tyr Ile Ala Asn Leu Gln Ser Leu Gly
305 310 315 320
His Lys His Arg Ser Lys Tyr Leu Ser Asp Ile Leu Gln Asp Met Lys
325 330 335
Arg Asp Ser Arg Ile Thr Arg Leu Ser Glu Ala Phe Gly Ser Thr Asp
340 345 350
Asn Leu Trp Arg Ile Ile Gly Asn Ile Ser Asn Leu Gln Glu Arg Ala
355 360 365
Val Arg Trp Tyr Phe Asn Asp Ala Lys Phe Glu Gln Gly Gln Glu Gln
370 375 380
Leu Asp Ala Val Lys Leu Lys Asn Val Leu Val Arg Ala Leu Lys Tyr
385 390 395 400
Leu Arg Ser Asp Asp Lys Glu Trp Ser Ala Ser Gln Lys Gln Ile Ile
405 410 415
Gln Ser Leu Glu Gln Ser Gly Asp Val Leu Asp Val Leu Ala Gly Leu
420 425 430
Asp Pro Asp Arg Thr Ile Pro Pro Tyr Glu Asp Gln Asn Asn Arg Arg
435 440 445
Pro Pro Glu Asp Gln Thr Leu Tyr Leu Asn Pro Lys Ala Leu Ser Ser
450 455 460
Glu Tyr Gly Glu Lys Trp Lys Ser Trp Ala Asn Lys Phe Ala Gly Ala

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

 530 535 540
 Val Val Ile Lys Asn Glu Lys Leu Glu Asp Val Leu Ser Gly His Gln
 545 550 555 560
 Leu Glu Glu Phe Leu Glu Phe Ala Asn Arg Tyr Tyr Gln Glu Thr Ala
 565 570 575
 Lys Ala Lys Asn Gly Leu Trp Phe Pro Glu Asn Ala Leu Leu Glu Arg
 580 585 590
 Ala Asp Leu His Pro Pro Met Lys Asn Lys Ile Leu Asn Val Ile Val

 595 600 605
 Gly Gln Ala Leu Gly Val Ser Pro Ala Glu Gly Thr Asp Phe Ile Glu
 610 615 620
 Glu Ile Trp Asn Ser Lys Val Lys Gly Arg Ser Thr Val Arg Ser Ile
 625 630 635 640
 Cys Asn Ala Ile Glu Asn Glu Arg Lys Thr Tyr Gly Pro Tyr Phe Ser
 645 650 655
 Glu Asp Tyr Lys Phe Val Lys Thr Ala Leu Lys Glu Gly Lys Thr Glu

 660 665 670
 Lys Glu Leu Ser Lys Lys Phe Ala Ala Val Ile Lys Val Leu Lys Met
 675 680 685
 Val Ser Glu Val Val Pro Phe Ile Gly Lys Glu Leu Arg Leu Ser Asp
 690 695 700
 Glu Ala Gln Ser Lys Phe Asp Asn Leu Tyr Ser Leu Ala Gln Leu Tyr
 705 710 715 720

Asn Leu Ile Glu Thr Glu Arg Asn Gly Phe Ser Lys Val Ser Leu Ala

725 730 735

Ala His Leu Glu Asn Ala Trp Arg Met Thr Met Thr Asp Gly Ser Ala

740 745 750

Gln Cys Cys Arg Leu Pro Ala Asp Cys Val Arg Pro Phe Asp Gly Phe

755 760 765

Ile Arg Lys Ala Ile Asp Arg Asn Ser Trp Glu Val Ala Lys Arg Ile

770 775 780

Ala Glu Glu Val Lys Lys Ser Val Asp Phe Thr Asn Gly Thr Val Lys

785 790 795 800

Ile Pro Val Ala Ile Glu Ala Asn Ser Phe Asn Phe Thr Ala Ser Leu

805 810 815

Thr Asp Leu Lys Tyr Ile Gln Leu Lys Glu Gln Lys Leu Lys Lys Lys

820 825 830

Leu Glu Asp Ile Gln Arg Asn Glu Glu Asn Gln Glu Lys Arg Trp Leu

835 840 845

Ser Lys Glu Glu Arg Ile Arg Ala Asp Ser His Gly Ile Cys Ala Tyr

850 855 860

Thr Gly Arg Pro Leu Asp Asp Val Gly Glu Ile Asp His Ile Ile Pro

865 870 875 880

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

885 890 895

Asn Leu Ile Phe Val Ser Ala Gln Gly Asn Gln Glu Lys Lys Asn Asn

900 905 910

Ile Tyr Leu Leu Ser Asn Leu Ala Lys Asn Tyr Leu Ala Ala Val Phe

915 920 925

Gly Thr Ser Asp Leu Ser Gln Ile Thr Asn Glu Ile Glu Ser Thr Val

930 935 940

Leu Gln Leu Lys Ala Ala Gly Arg Leu Gly Tyr Phe Asp Leu Leu Ser

945 950 955 960

Glu Lys Glu Arg Ala Cys Ala Arg His Ala Leu Phe Leu Asn Ser Asp

965	970	975
Ser Glu Ala Arg Arg Ala Val Ile Asp Val Leu Gly Ser Arg Arg Lys		
980	985	990
Ala Ser Val Asn Gly Thr Gln Ala Trp Phe Val Arg Ser Ile Phe Ser		
995	1000	1005
Lys Val Arg Gln Ala Leu Ala Ala Trp Thr Gln Glu Thr Gly Asn		
1010	1015	1020
Glu Leu Ile Phe Asp Ala Ile Ser Val Pro Ala Ala Asp Ser Ser		
1025	1030	1035
Glu Met Arg Lys Arg Phe Ala Glu Tyr Arg Pro Glu Phe Arg Lys		
1040	1045	1050
Pro Lys Val Gln Pro Val Ala Ser His Ser Ile Asp Ala Met Cys		
1055	1060	1065
Ile Tyr Leu Ala Ala Cys Ser Asp Pro Phe Lys Thr Lys Arg Met		
1070	1075	1080
Gly Ser Gln Leu Ala Ile Tyr Glu Pro Ile Asn Phe Asp Asn Leu		
1085	1090	1095
Phe Thr Gly Ser Cys Gln Val Ile Gln Asn Thr Pro Arg Asn Phe		
1100	1105	1110
Ser Asp Lys Thr Asn Ile Ala Asn Ser Pro Ile Phe Lys Glu Thr		
1115	1120	1125
Ile Tyr Ala Glu Arg Phe Leu Asp Ile Ile Val Ser Arg Gly Glu		
1130	1135	1140
Ile Phe Ile Gly Tyr Pro Ser Asn Met Pro Phe Glu Glu Lys Pro		
1145	1150	1155
Asn Arg Ile Ser Ile Gly Gly Lys Asp Pro Phe Ser Ile Leu Ser		
1160	1165	1170
Val Leu Gly Ala Tyr Leu Asp Lys Ala Pro Ser Ser Glu Lys Glu		
1175	1180	1185
Lys Leu Thr Ile Tyr Arg Val Val Lys Asn Lys Ala Phe Glu Leu		
1190	1195	1200

Phe Ser Lys Val Ala Gly Ser Lys Phe Thr Ala Glu Glu Asp Lys		
1205	1210	1215
Ala Ala Lys Ile Leu Glu Ala Leu His Phe Val Thr Val Lys Gln		
1220	1225	1230
Asp Val Ala Ala Thr Val Ser Asp Leu Ile Lys Ser Lys Lys Glu		
1235	1240	1245
Leu Ser Lys Asp Ser Ile Glu Asn Leu Ala Lys Gln Lys Gly Cys		
1250	1255	1260
Leu Lys Lys Val Glu Tyr Ser Ser Lys Glu Phe Lys Phe Lys Gly		
1265	1270	1275
Ser Leu Ile Ile Pro Ala Ala Val Glu Trp Gly Lys Val Leu Trp		
1280	1285	1290
Asn Val Phe Lys Glu Asn Thr Ala Glu Glu Leu Lys Asp Glu Asn		
1295	1300	1305
Ala Leu Arg Lys Ala Leu Glu Ala Ala Trp Pro Ser Ser Phe Gly		
1310	1315	1320
Thr Arg Asn Leu His Ser Lys Ala Lys Arg Val Phe Ser Leu Pro		
1325	1330	1335
Val Val Ala Thr Gln Ser Gly Ala Val Arg Ile Arg Arg Lys Thr		
1340	1345	1350
Ala Phe Gly Asp Phe Val Tyr Gln Ser Gln Asp Thr Asn Asn Leu		
1355	1360	1365
Tyr Ser Ser Phe Pro Val Lys Asn Gly Lys Leu Asp Trp Ser Ser		
1370	1375	1380
Pro Ile Ile His Pro Ala Leu Gln Asn Arg Asn Leu Thr Ala Tyr		
1385	1390	1395
Gly Tyr Arg Phe Val Asp His Asp Arg Ser Ile Ser Met Ser Glu		
1400	1405	1410
Phe Arg Glu Val Tyr Asn Lys Asp Asp Leu Met Arg Ile Glu Leu		
1415	1420	1425
Ala Gln Gly Thr Ser Ser Arg Arg Tyr Leu Arg Val Glu Met Pro		

1430 1435 1440
 Gly Glu Lys Phe Leu Ala Trp Phe Gly Glu Asn Ser Ile Ser Leu
 1445 1450 1455
 Gly Ser Ser Phe Lys Phe Ser Val Ser Glu Val Phe Asp Asn Lys

1460 1465 1470
 Ile Tyr Thr Glu Asn Ala Glu Phe Thr Lys Phe Leu Pro Lys Pro
 1475 1480 1485
 Arg Glu Asp Asn Lys His Asn Gly Thr Ile Phe Phe Glu Leu Val
 1490 1495 1500
 Gly Pro Arg Val Ile Phe Asn Tyr Ile Val Gly Gly Ala Ala Ser
 1505 1510 1515
 Ser Leu Lys Glu Ile Phe Ser Glu Ala Gly Lys Glu Arg Ser Pro

1520 1525 1530
 Lys Lys Lys Arg Lys Val Leu Glu Gly Gly Gly Gly Ser Ser Thr
 1535 1540 1545
 Asp His Pro Lys Tyr Ser Asp Met Ile Val Ala Ala Ile Gln Ala
 1550 1555 1560
 Glu Lys Asn Arg Ala Gly Ser Ser Arg Gln Ser Ile Gln Lys Tyr
 1565 1570 1575
 Ile Lys Ser His Tyr Lys Val Gly Glu Asn Ala Asp Ser Gln Ile

1580 1585 1590
 Lys Leu Ser Ile Lys Arg Leu Val Thr Thr Gly Val Leu Lys Gln
 1595 1600 1605
 Thr Lys Gly Val Gly Ala Ser Gly Ser Phe Arg Leu Ala Lys Ser
 1610 1615 1620
 Asp Glu Pro

1625

<210> 119

<211> 1289

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic polypeptide"

<400> 119

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu
 1 5 10 15
 Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val
 20 25 30
 Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys
 35 40 45
 Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu
 50 55 60

Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu
 65 70 75 80
 Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu
 85 90 95
 Ala Lys Ser Asp Thr Gly Ser Gly Met Asn Tyr Lys Met Gly Leu Asp
 100 105 110
 Ile Gly Ile Ala Ser Val Gly Trp Ala Val Ile Asn Leu Asp Leu Lys
 115 120 125

Arg Ile Glu Asp Leu Gly Val Arg Ile Phe Asp Lys Ala Glu His Pro
 130 135 140
 Gln Asn Gly Glu Ser Leu Ala Leu Pro Arg Arg Ile Ala Arg Ser Ala
 145 150 155 160
 Arg Arg Arg Leu Arg Arg Arg Lys His Arg Leu Glu Arg Ile Arg Arg
 165 170 175
 Leu Leu Val Ser Glu Asn Val Leu Thr Lys Glu Glu Met Asn Leu Leu
 180 185 190

Phe Lys Gln Lys Lys Gln Ile Asp Val Trp Gln Leu Arg Val Asp Ala
 195 200 205
 Leu Glu Arg Lys Leu Asn Asn Asp Glu Leu Ala Arg Val Leu Leu His
 210 215 220

Leu Ala Lys Arg Arg Gly Phe Lys Ser Asn Arg Lys Ser Glu Arg Asn
 225 230 235 240
 Ser Lys Glu Ser Ser Glu Phe Leu Lys Asn Ile Glu Glu Asn Gln Ser
 245 250 255

 Ile Leu Ala Gln Tyr Arg Ser Val Gly Glu Met Ile Val Lys Asp Ser
 260 265 270
 Lys Phe Ala Tyr His Lys Arg Asn Lys Leu Asp Ser Tyr Ser Asn Met
 275 280 285
 Ile Ala Arg Asp Asp Leu Glu Arg Glu Ile Lys Leu Ile Phe Glu Lys
 290 295 300
 Gln Arg Glu Phe Asn Asn Pro Val Cys Thr Glu Arg Leu Glu Glu Lys
 305 310 315 320

 Tyr Leu Asn Ile Trp Ser Ser Gln Arg Pro Phe Ala Ser Lys Glu Asp
 325 330 335
 Ile Glu Lys Lys Val Gly Phe Cys Thr Phe Glu Pro Lys Glu Lys Arg
 340 345 350
 Ala Pro Lys Ala Thr Tyr Thr Phe Gln Ser Phe Ile Val Trp Glu His
 355 360 365
 Ile Asn Lys Leu Arg Leu Val Ser Pro Asp Glu Thr Arg Ala Leu Thr
 370 375 380

 Glu Ile Glu Arg Asn Leu Leu Tyr Lys Gln Ala Phe Ser Lys Asn Lys
 385 390 395 400
 Met Thr Tyr Tyr Asp Ile Arg Lys Leu Leu Asn Leu Ser Asp Asp Ile
 405 410 415
 His Phe Lys Gly Leu Leu Tyr Asp Pro Lys Ser Ser Leu Lys Gln Ile
 420 425 430
 Glu Asn Ile Arg Phe Leu Glu Leu Asp Ser Tyr His Lys Ile Arg Lys
 435 440 445

 Cys Ile Glu Asn Val Tyr Gly Lys Asp Gly Ile Arg Met Phe Asn Glu
 450 455 460
 Thr Asp Ile Asp Thr Phe Gly Tyr Ala Leu Thr Ile Phe Lys Asp Asp

465 470 475 480
 Glu Asp Ile Val Ala Tyr Leu Gln Asn Glu Tyr Ile Thr Lys Asn Gly
 485 490 495
 Lys Arg Val Ser Asn Leu Ala Asn Lys Val Tyr Asp Lys Ser Leu Ile
 500 505 510

 Asp Glu Leu Leu Asn Leu Ser Phe Ser Lys Phe Ala His Leu Ser Met
 515 520 525
 Lys Ala Ile Arg Asn Ile Leu Pro Tyr Met Glu Gln Gly Glu Ile Tyr
 530 535 540
 Ser Lys Ala Cys Glu Leu Ala Gly Tyr Asn Phe Thr Gly Pro Lys Lys
 545 550 555 560
 Lys Glu Lys Ala Leu Leu Leu Pro Val Ile Pro Asn Ile Ala Asn Pro
 565 570 575

 Val Val Met Arg Ala Leu Thr Gln Ser Arg Lys Val Val Asn Ala Ile
 580 585 590
 Ile Lys Lys Tyr Gly Ser Pro Val Ser Ile His Ile Glu Leu Ala Arg
 595 600 605
 Asp Leu Ser His Ser Phe Asp Glu Arg Lys Lys Ile Gln Lys Asp Gln
 610 615 620
 Thr Glu Asn Arg Lys Lys Asn Glu Thr Ala Ile Lys Gln Leu Ile Glu
 625 630 635 640

 Tyr Glu Leu Thr Lys Asn Pro Thr Gly Leu Asp Ile Val Lys Phe Lys
 645 650 655
 Leu Trp Ser Glu Gln Gln Gly Arg Cys Met Tyr Ser Leu Lys Pro Ile
 660 665 670
 Glu Leu Glu Arg Leu Leu Glu Pro Gly Tyr Val Glu Val Asp His Ile
 675 680 685
 Leu Pro Tyr Ser Arg Ser Leu Asp Asp Ser Tyr Ala Asn Lys Val Leu
 690 695 700

 Val Leu Thr Lys Glu Asn Arg Glu Lys Gly Asn His Thr Pro Val Glu
 705 710 715 720

Tyr Leu Gly Leu Gly Ser Glu Arg Trp Lys Lys Phe Glu Lys Phe Val
 725 730 735
 Leu Ala Asn Lys Gln Phe Ser Lys Lys Lys Lys Gln Asn Leu Leu Arg
 740 745 750
 Leu Arg Tyr Glu Glu Thr Glu Glu Lys Glu Phe Lys Glu Arg Asn Leu
 755 760 765

 Asn Asp Thr Arg Tyr Ile Ser Lys Phe Phe Ala Asn Phe Ile Lys Glu
 770 775 780
 His Leu Lys Phe Ala Asp Gly Asp Gly Gly Gln Lys Val Tyr Thr Ile
 785 790 795 800
 Asn Gly Lys Ile Thr Ala His Leu Arg Ser Arg Trp Asp Phe Asn Lys
 805 810 815
 Asn Arg Glu Glu Ser Asp Leu His His Ala Val Asp Ala Val Ile Val
 820 825 830

 Ala Cys Ala Thr Gln Gly Met Ile Lys Lys Ile Thr Glu Phe Tyr Lys
 835 840 845
 Ala Arg Glu Gln Asn Lys Glu Ser Ala Lys Lys Lys Glu Pro Ile Phe
 850 855 860
 Pro Gln Pro Trp Pro His Phe Ala Asp Glu Leu Lys Ala Arg Leu Ser
 865 870 875 880
 Lys Phe Pro Gln Glu Ser Ile Glu Ala Phe Ala Leu Gly Asn Tyr Asp
 885 890 895

 Arg Lys Lys Leu Glu Ser Leu Arg Pro Val Phe Val Ser Arg Met Pro
 900 905 910
 Lys Arg Ser Val Thr Gly Ala Ala His Gln Glu Thr Leu Arg Arg Cys
 915 920 925
 Val Gly Ile Asp Glu Gln Ser Gly Lys Ile Gln Thr Ala Val Lys Thr
 930 935 940
 Lys Leu Ser Asp Ile Lys Leu Asp Lys Asp Gly His Phe Pro Met Tyr
 945 950 955 960

 Gln Lys Glu Ser Asp Pro Arg Thr Tyr Glu Ala Ile Arg Gln Arg Leu

965							970					975				
Leu	Glu	His	Asn	Asn	Asp	Pro	Lys	Lys	Ala	Phe	Gln	Glu	Pro	Leu	Tyr	
980							985					990				
Lys	Pro	Lys	Lys	Asn	Gly	Glu	Pro	Gly	Pro	Val	Ile	Arg	Thr	Val	Lys	
995							1000					1005				
Ile	Ile	Asp	Thr	Lys	Asn	Lys	Val	Val	His	Leu	Asp	Gly	Ser	Lys		
1010							1015					1020				
Thr	Val	Ala	Tyr	Asn	Ser	Asn	Ile	Val	Arg	Thr	Asp	Val	Phe	Glu		
1025							1030					1035				
Lys	Asp	Gly	Lys	Tyr	Tyr	Cys	Val	Pro	Val	Tyr	Thr	Met	Asp	Ile		
1040							1045					1050				
Met	Lys	Gly	Thr	Leu	Pro	Asn	Lys	Ala	Ile	Glu	Ala	Asn	Lys	Pro		
1055							1060					1065				
Tyr	Ser	Glu	Trp	Lys	Glu	Met	Thr	Glu	Glu	Tyr	Thr	Phe	Gln	Phe		
1070							1075					1080				
Ser	Leu	Phe	Pro	Asn	Asp	Leu	Val	Arg	Ile	Val	Leu	Pro	Arg	Glu		
1085							1090					1095				
Lys	Thr	Ile	Lys	Thr	Ser	Thr	Asn	Glu	Glu	Ile	Ile	Ile	Lys	Asp		
1100							1105					1110				
Ile	Phe	Ala	Tyr	Tyr	Lys	Thr	Ile	Asp	Ser	Ala	Thr	Gly	Gly	Leu		
1115							1120					1125				
Glu	Leu	Ile	Ser	His	Asp	Arg	Asn	Phe	Ser	Leu	Arg	Gly	Val	Gly		
1130							1135					1140				
Ser	Lys	Thr	Leu	Lys	Arg	Phe	Glu	Lys	Tyr	Gln	Val	Asp	Val	Leu		
1145							1150					1155				
Gly	Asn	Ile	His	Lys	Val	Lys	Gly	Glu	Lys	Arg	Val	Gly	Leu	Ala		
1160							1165					1170				
Ala	Pro	Thr	Asn	Gln	Lys	Lys	Gly	Lys	Thr	Val	Asp	Ser	Leu	Gln		
1175							1180					1185				
Ser	Val	Ser	Asp	Pro	Lys	Lys	Lys	Arg	Lys	Val	Leu	Glu	Gly	Gly		
1190							1195					1200				

Gly Gly Ser Gly Lys Gly Asp Pro Lys Lys Pro Arg Gly Lys Met
 1205 1210 1215
 Ser Ser Tyr Ala Phe Phe Val Gln Thr Cys Arg Glu Glu His Lys
 1220 1225 1230
 Lys Lys His Pro Asp Ala Ser Val Asn Phe Ser Glu Phe Ser Lys
 1235 1240 1245
 Lys Cys Ser Glu Arg Trp Lys Thr Met Ser Ala Lys Glu Lys Gly
 1250 1255 1260

Lys Phe Glu Asp Met Ala Lys Ala Asp Lys Ala Arg Tyr Glu Arg
 1265 1270 1275
 Glu Met Lys Thr Tyr Ile Pro Pro Lys Gly Glu
 1280 1285

<210> 120

<211> 1286

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
 polypeptide"

<400> 120

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu
 1 5 10 15

Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val
 20 25 30

Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys
 35 40 45

Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu
 50 55 60

Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu
 65 70 75 80

Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu
 85 90 95

Ala Lys Ser Asp Thr Gly Ser Gly Met Asn Tyr Lys Met Gly Leu Asp
100 105 110

Ile Gly Ile Ala Ser Val Gly Trp Ala Val Ile Asn Leu Asp Leu Lys
115 120 125

Arg Ile Glu Asp Leu Gly Val Arg Ile Phe Asp Lys Ala Glu His Pro
130 135 140

Gln Asn Gly Glu Ser Leu Ala Leu Pro Arg Arg Ile Ala Arg Ser Ala
145 150 155 160

Arg Arg Arg Leu Arg Arg Arg Lys His Arg Leu Glu Arg Ile Arg Arg
165 170 175

Leu Leu Val Ser Glu Asn Val Leu Thr Lys Glu Glu Met Asn Leu Leu
180 185 190

Phe Lys Gln Lys Lys Gln Ile Asp Val Trp Gln Leu Arg Val Asp Ala
195 200 205

Leu Glu Arg Lys Leu Asn Asn Asp Glu Leu Ala Arg Val Leu Leu His
210 215 220

Leu Ala Lys Arg Arg Gly Phe Lys Ser Asn Arg Lys Ser Glu Arg Asn
225 230 235 240

Ser Lys Glu Ser Ser Glu Phe Leu Lys Asn Ile Glu Glu Asn Gln Ser
245 250 255

Ile Leu Ala Gln Tyr Arg Ser Val Gly Glu Met Ile Val Lys Asp Ser
260 265 270

Lys Phe Ala Tyr His Lys Arg Asn Lys Leu Asp Ser Tyr Ser Asn Met
275 280 285

Ile Ala Arg Asp Asp Leu Glu Arg Glu Ile Lys Leu Ile Phe Glu Lys
290 295 300

Gln Arg Glu Phe Asn Asn Pro Val Cys Thr Glu Arg Leu Glu Glu Lys
305 310 315 320

Tyr Leu Asn Ile Trp Ser Ser Gln Arg Pro Phe Ala Ser Lys Glu Asp
325 330 335

Ile Glu Lys Lys Val Gly Phe Cys Thr Phe Glu Pro Lys Glu Lys Arg

340	345	350	
Ala Pro Lys Ala Thr Tyr Thr Phe Gln Ser Phe Ile Val Trp Glu His			
355	360	365	
Ile Asn Lys Leu Arg Leu Val Ser Pro Asp Glu Thr Arg Ala Leu Thr			
370	375	380	
Glu Ile Glu Arg Asn Leu Leu Tyr Lys Gln Ala Phe Ser Lys Asn Lys			
385	390	395	400
Met Thr Tyr Tyr Asp Ile Arg Lys Leu Leu Asn Leu Ser Asp Asp Ile			
405	410	415	
His Phe Lys Gly Leu Leu Tyr Asp Pro Lys Ser Ser Leu Lys Gln Ile			
420	425	430	
Glu Asn Ile Arg Phe Leu Glu Leu Asp Ser Tyr His Lys Ile Arg Lys			
435	440	445	
Cys Ile Glu Asn Val Tyr Gly Lys Asp Gly Ile Arg Met Phe Asn Glu			
450	455	460	
Thr Asp Ile Asp Thr Phe Gly Tyr Ala Leu Thr Ile Phe Lys Asp Asp			
465	470	475	480
Glu Asp Ile Val Ala Tyr Leu Gln Asn Glu Tyr Ile Thr Lys Asn Gly			
485	490	495	
Lys Arg Val Ser Asn Leu Ala Asn Lys Val Tyr Asp Lys Ser Leu Ile			
500	505	510	
Asp Glu Leu Leu Asn Leu Ser Phe Ser Lys Phe Ala His Leu Ser Met			
515	520	525	
Lys Ala Ile Arg Asn Ile Leu Pro Tyr Met Glu Gln Gly Glu Ile Tyr			
530	535	540	
Ser Lys Ala Cys Glu Leu Ala Gly Tyr Asn Phe Thr Gly Pro Lys Lys			
545	550	555	560
Lys Glu Lys Ala Leu Leu Leu Pro Val Ile Pro Asn Ile Ala Asn Pro			
565	570	575	
Val Val Met Arg Ala Leu Thr Gln Ser Arg Lys Val Val Asn Ala Ile			
580	585	590	

Ile Lys Lys Tyr Gly Ser Pro Val Ser Ile His Ile Glu Leu Ala Arg
595 600 605

Asp Leu Ser His Ser Phe Asp Glu Arg Lys Lys Ile Gln Lys Asp Gln
610 615 620

Thr Glu Asn Arg Lys Lys Asn Glu Thr Ala Ile Lys Gln Leu Ile Glu
625 630 635 640

Tyr Glu Leu Thr Lys Asn Pro Thr Gly Leu Asp Ile Val Lys Phe Lys
645 650 655

Leu Trp Ser Glu Gln Gln Gly Arg Cys Met Tyr Ser Leu Lys Pro Ile
660 665 670

Glu Leu Glu Arg Leu Leu Glu Pro Gly Tyr Val Glu Val Asp His Ile
675 680 685

Leu Pro Tyr Ser Arg Ser Leu Asp Asp Ser Tyr Ala Asn Lys Val Leu
690 695 700

Val Leu Thr Lys Glu Asn Arg Glu Lys Gly Asn His Thr Pro Val Glu
705 710 715 720

Tyr Leu Gly Leu Gly Ser Glu Arg Trp Lys Lys Phe Glu Lys Phe Val
725 730 735

Leu Ala Asn Lys Gln Phe Ser Lys Lys Lys Lys Gln Asn Leu Leu Arg
740 745 750

Leu Arg Tyr Glu Glu Thr Glu Glu Lys Glu Phe Lys Glu Arg Asn Leu
755 760 765

Asn Asp Thr Arg Tyr Ile Ser Lys Phe Phe Ala Asn Phe Ile Lys Glu
770 775 780

His Leu Lys Phe Ala Asp Gly Asp Gly Gly Gln Lys Val Tyr Thr Ile
785 790 795 800

Asn Gly Lys Ile Thr Ala His Leu Arg Ser Arg Trp Asp Phe Asn Lys
805 810 815

Asn Arg Glu Glu Ser Asp Leu His His Ala Val Asp Ala Val Ile Val
820 825 830

Ala Cys Ala Thr Gln Gly Met Ile Lys Lys Ile Thr Glu Phe Tyr Lys

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

Ser Leu Phe Pro Asn Asp Leu Val Arg Ile Val Leu Pro Arg Glu
1085 1090 1095

Lys Thr Ile Lys Thr Ser Thr Asn Glu Glu Ile Ile Ile Lys Asp
1100 1105 1110

Ile Phe Ala Tyr Tyr Lys Thr Ile Asp Ser Ala Thr Gly Gly Leu
1115 1120 1125

Glu Leu Ile Ser His Asp Arg Asn Phe Ser Leu Arg Gly Val Gly
1130 1135 1140

Ser Lys Thr Leu Lys Arg Phe Glu Lys Tyr Gln Val Asp Val Leu
1145 1150 1155

Gly Asn Ile His Lys Val Lys Gly Glu Lys Arg Val Gly Leu Ala
1160 1165 1170

Ala Pro Thr Asn Gln Lys Lys Gly Lys Thr Val Asp Ser Leu Gln
1175 1180 1185

Ser Val Ser Asp Pro Lys Lys Lys Arg Lys Val Leu Glu Gly Gly
1190 1195 1200

Gly Gly Ser Ser Thr Asp His Pro Lys Tyr Ser Asp Met Ile Val
1205 1210 1215

Ala Ala Ile Gln Ala Glu Lys Asn Arg Ala Gly Ser Ser Arg Gln
1220 1225 1230

Ser Ile Gln Lys Tyr Ile Lys Ser His Tyr Lys Val Gly Glu Asn
1235 1240 1245

Ala Asp Ser Gln Ile Lys Leu Ser Ile Lys Arg Leu Val Thr Thr
1250 1255 1260

Gly Val Leu Lys Gln Thr Lys Gly Val Gly Ala Ser Gly Ser Phe
1265 1270 1275

Arg Leu Ala Lys Ser Asp Glu Pro
1280 1285

<210> 121

<211> 1564

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic polypeptide"

<400> 121

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu

1 5 10 15

Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val

20 25 30

Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys

35 40 45

Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu

50 55 60

Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu

65 70 75 80

Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu

85 90 95

Ala Lys Ser Asp Thr Gly Ser Gly Met Thr Lys Leu Asn Gln Pro Tyr

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

Arg Thr Thr Arg Arg Arg Leu Ser Arg Thr Arg Trp Arg Leu Ser Phe

165 170 175

Leu Arg Asp Phe Phe Ala Pro His Ile Thr Lys Ile Asp Pro Asp Phe

180 185 190

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

195 200 205

Phe Lys Tyr Glu Lys Arg Leu Phe Asn Asp Arg Thr Asp Ala Glu Phe

210	215	220	
Tyr Glu Asp Tyr Pro Ser Met Tyr His Leu Arg Leu His Leu Met Thr			
225	230	235	240
His Thr His Lys Ala Asp Pro Arg Glu Ile Phe Leu Ala Ile His His			
	245	250	255
Ile Leu Lys Ser Arg Gly His Phe Leu Thr Pro Gly Ala Ala Lys Asp			
	260	265	270
Phe Asn Thr Asp Lys Val Asp Leu Glu Asp Ile Phe Pro Ala Leu Thr			
	275	280	285
Glu Ala Tyr Ala Gln Val Tyr Pro Asp Leu Glu Leu Thr Phe Asp Leu			
	290	295	300
Ala Lys Ala Asp Asp Phe Lys Ala Lys Leu Leu Asp Glu Gln Ala Thr			
305	310	315	320
Pro Ser Asp Thr Gln Lys Ala Leu Val Asn Leu Leu Leu Ser Ser Asp			
	325	330	335
Gly Glu Lys Glu Ile Val Lys Lys Arg Lys Gln Val Leu Thr Glu Phe			
	340	345	350
Ala Lys Ala Ile Thr Gly Leu Lys Thr Lys Phe Asn Leu Ala Leu Gly			
	355	360	365
Thr Glu Val Asp Glu Ala Asp Ala Ser Asn Trp Gln Phe Ser Met Gly			
	370	375	380
Gln Leu Asp Asp Lys Trp Ser Asn Ile Glu Thr Ser Met Thr Asp Gln			
385	390	395	400
Gly Thr Glu Ile Phe Glu Gln Ile Gln Glu Leu Tyr Arg Ala Arg Leu			
	405	410	415
Leu Asn Gly Ile Val Pro Ala Gly Met Ser Leu Ser Gln Ala Lys Val			
	420	425	430
Ala Asp Tyr Gly Gln His Lys Glu Asp Leu Glu Leu Phe Lys Thr Tyr			
	435	440	445
Leu Lys Lys Leu Asn Asp His Glu Leu Ala Lys Thr Ile Arg Gly Leu			
	450	455	460

Tyr Asp Arg Tyr Ile Asn Gly Asp Asp Ala Lys Pro Phe Leu Arg Glu
465 470 475 480

Asp Phe Val Lys Ala Leu Thr Lys Glu Val Thr Ala His Pro Asn Glu
485 490 495

Val Ser Glu Gln Leu Leu Asn Arg Met Gly Gln Ala Asn Phe Met Leu
500 505 510

Lys Gln Arg Thr Lys Ala Asn Gly Ala Ile Pro Ile Gln Leu Gln Gln
515 520 525

Arg Glu Leu Asp Gln Ile Ile Ala Asn Gln Ser Lys Tyr Tyr Asp Trp
530 535 540

Leu Ala Ala Pro Asn Pro Val Glu Ala His Arg Trp Lys Met Pro Tyr
545 550 555 560

Gln Leu Asp Glu Leu Leu Asn Phe His Ile Pro Tyr Tyr Val Gly Pro
565 570 575

Leu Ile Thr Pro Lys Gln Gln Ala Glu Ser Gly Glu Asn Val Phe Ala
580 585 590

Trp Met Val Arg Lys Asp Pro Ser Gly Asn Ile Thr Pro Tyr Asn Phe
595 600 605

Asp Glu Lys Val Asp Arg Glu Ala Ser Ala Asn Thr Phe Ile Gln Arg
610 615 620

Met Lys Thr Thr Asp Thr Tyr Leu Ile Gly Glu Asp Val Leu Pro Lys
625 630 635 640

Gln Ser Leu Leu Tyr Gln Lys Tyr Glu Val Leu Asn Glu Leu Asn Asn
645 650 655

Val Arg Ile Asn Asn Glu Cys Leu Gly Thr Asp Gln Lys Gln Arg Leu
660 665 670

Ile Arg Glu Val Phe Glu Arg His Ser Ser Val Thr Ile Lys Gln Val
675 680 685

Ala Asp Asn Leu Val Ala His Gly Asp Phe Ala Arg Arg Pro Glu Ile
690 695 700

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

705 710 715 720
 His Gln Leu Lys Glu Ile Leu His Glu Ala Ile Asp Asp Pro Thr Lys
 725 730 735

 Leu Leu Asp Ile Glu Asn Ile Ile Thr Trp Ser Thr Val Phe Glu Asp
 740 745 750
 His Thr Ile Phe Glu Thr Lys Leu Ala Glu Ile Glu Trp Leu Asp Pro
 755 760 765
 Lys Lys Ile Asn Glu Leu Ser Gly Ile Arg Tyr Arg Gly Trp Gly Gln
 770 775 780
 Phe Ser Arg Lys Leu Leu Asp Gly Leu Lys Leu Gly Asn Gly His Thr
 785 790 795 800

 Val Ile Gln Glu Leu Met Leu Ser Asn His Asn Leu Met Gln Ile Leu
 805 810 815
 Ala Asp Glu Thr Leu Lys Glu Thr Met Thr Glu Leu Asn Gln Asp Lys
 820 825 830
 Leu Lys Thr Asp Asp Ile Glu Asp Val Ile Asn Asp Ala Tyr Thr Ser
 835 840 845
 Pro Ser Asn Lys Lys Ala Leu Arg Gln Val Leu Arg Val Val Glu Asp
 850 855 860

 Ile Lys His Ala Ala Asn Gly Gln Asp Pro Ser Trp Leu Phe Ile Glu
 865 870 875 880
 Thr Ala Asp Gly Thr Gly Thr Ala Gly Lys Arg Thr Gln Ser Arg Gln
 885 890 895
 Lys Gln Ile Gln Thr Val Tyr Ala Asn Ala Ala Gln Glu Leu Ile Asp
 900 905 910
 Ser Ala Val Arg Gly Glu Leu Glu Asp Lys Ile Ala Asp Lys Ala Ser
 915 920 925

 Phe Thr Asp Arg Leu Val Leu Tyr Phe Met Gln Gly Gly Arg Asp Ile
 930 935 940
 Tyr Thr Gly Ala Pro Leu Asn Ile Asp Gln Leu Ser His Tyr Asp Ile
 945 950 955 960

Asp His Ile Leu Pro Gln Ser Leu Ile Lys Asp Asp Ser Leu Asp Asn			
	965	970	975
Arg Val Leu Val Asn Ala Thr Ile Asn Arg Glu Lys Asn Asn Val Phe			
	980	985	990
Ala Ser Thr Leu Phe Ala Gly Lys Met Lys Ala Thr Trp Arg Lys Trp			
	995	1000	1005
His Glu Ala Gly Leu Ile Ser Gly Arg Lys Leu Arg Asn Leu Met			
	1010	1015	1020
Leu Arg Pro Asp Glu Ile Asp Lys Phe Ala Lys Gly Phe Val Ala			
	1025	1030	1035
Arg Gln Leu Val Glu Thr Arg Gln Ile Ile Lys Leu Thr Glu Gln			
	1040	1045	1050
Ile Ala Ala Ala Gln Tyr Pro Asn Thr Lys Ile Ile Ala Val Lys			
	1055	1060	1065
Ala Gly Leu Ser His Gln Leu Arg Glu Glu Leu Asp Phe Pro Lys			
	1070	1075	1080
Asn Arg Asp Val Asn His Tyr His His Ala Phe Asp Ala Phe Leu			
	1085	1090	1095
Ala Ala Arg Ile Gly Thr Tyr Leu Leu Lys Arg Tyr Pro Lys Leu			
	1100	1105	1110
Ala Pro Phe Phe Thr Tyr Gly Glu Phe Ala Lys Val Asp Val Lys			
	1115	1120	1125
Lys Phe Arg Glu Phe Asn Phe Ile Gly Ala Leu Thr His Ala Lys			
	1130	1135	1140
Lys Asn Ile Ile Ala Lys Asp Thr Gly Glu Ile Val Trp Asp Lys			
	1145	1150	1155
Glu Arg Asp Ile Arg Glu Leu Asp Arg Ile Tyr Asn Phe Lys Arg			
	1160	1165	1170
Met Leu Ile Thr His Glu Val Tyr Phe Glu Thr Ala Asp Leu Phe			
	1175	1180	1185
Lys Gln Thr Ile Tyr Ala Ala Lys Asp Ser Lys Glu Arg Gly Gly			

1190	1195	1200
Ser Lys Gln Leu Ile Pro Lys	Lys Gln Gly Tyr Pro	Thr Gln Val
1205	1210	1215
Tyr Gly Gly Tyr Thr Gln Glu	Ser Gly Ser Tyr Asn	Ala Leu Val
1220	1225	1230
Arg Val Ala Glu Ala Asp Thr	Thr Ala Tyr Gln Val	Ile Lys Ile
1235	1240	1245
Ser Ala Gln Asn Ala Ser Lys	Ile Ala Ser Ala Asn	Leu Lys Ser
1250	1255	1260
Arg Glu Lys Gly Lys Gln Leu	Leu Asn Glu Ile Val	Val Lys Gln
1265	1270	1275
Leu Ala Lys Arg Arg Lys Asn	Trp Lys Pro Ser Ala	Asn Ser Phe
1280	1285	1290
Lys Ile Val Ile Pro Arg Phe	Gly Met Gly Thr Leu	Phe Gln Asn
1295	1300	1305
Ala Lys Tyr Gly Leu Phe Met	Val Asn Ser Asp Thr	Tyr Tyr Arg
1310	1315	1320
Asn Tyr Gln Glu Leu Trp Leu	Ser Arg Glu Asn Gln	Lys Leu Leu
1325	1330	1335
Lys Lys Leu Phe Ser Ile Lys	Tyr Glu Lys Thr Gln	Met Asn His
1340	1345	1350
Asp Ala Leu Gln Val Tyr Lys	Ala Ile Ile Asp Gln	Val Glu Lys
1355	1360	1365
Phe Phe Lys Leu Tyr Asp Ile	Asn Gln Phe Arg Ala	Lys Leu Ser
1370	1375	1380
Asp Ala Ile Glu Arg Phe Glu	Lys Leu Pro Ile Asn	Thr Asp Gly
1385	1390	1395
Asn Lys Ile Gly Lys Thr Glu	Thr Leu Arg Gln Ile	Leu Ile Gly
1400	1405	1410
Leu Gln Ala Asn Gly Thr Arg	Ser Asn Val Lys Asn	Leu Gly Ile
1415	1420	1425

Lys Thr Asp Leu Gly Leu Leu Gln Val Gly Ser Gly Ile Lys Leu
 1430 1435 1440
 Asp Lys Asp Thr Gln Ile Val Tyr Gln Ser Pro Ser Gly Leu Phe
 1445 1450 1455
 Lys Arg Arg Ile Pro Leu Ala Asp Leu Pro Lys Lys Lys Arg Lys
 1460 1465 1470

Val Leu Glu Gly Gly Gly Gly Ser Gly Lys Gly Asp Pro Lys Lys
 1475 1480 1485
 Pro Arg Gly Lys Met Ser Ser Tyr Ala Phe Phe Val Gln Thr Cys
 1490 1495 1500
 Arg Glu Glu His Lys Lys Lys His Pro Asp Ala Ser Val Asn Phe
 1505 1510 1515
 Ser Glu Phe Ser Lys Lys Cys Ser Glu Arg Trp Lys Thr Met Ser
 1520 1525 1530

Ala Lys Glu Lys Gly Lys Phe Glu Asp Met Ala Lys Ala Asp Lys
 1535 1540 1545
 Ala Arg Tyr Glu Arg Glu Met Lys Thr Tyr Ile Pro Pro Lys Gly
 1550 1555 1560

Glu

<210> 122

<211> 1561

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
 polypeptide"

<400> 122

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu

1 5 10 15
 Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val
 20 25 30

Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys
 35 40 45
 Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu
 50 55 60
 Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu

 65 70 75 80
 Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu
 85 90 95
 Ala Lys Ser Asp Thr Gly Ser Gly Met Thr Lys Leu Asn Gln Pro Tyr
 100 105 110
 Gly Ile Gly Leu Asp Ile Gly Ser Asn Ser Ile Gly Phe Ala Val Val
 115 120 125
 Asp Ala Asn Ser His Leu Leu Arg Leu Lys Gly Glu Thr Ala Ile Gly

 130 135 140
 Ala Arg Leu Phe Arg Glu Gly Gln Ser Ala Ala Asp Arg Arg Gly Ser
 145 150 155 160
 Arg Thr Thr Arg Arg Arg Leu Ser Arg Thr Arg Trp Arg Leu Ser Phe
 165 170 175
 Leu Arg Asp Phe Phe Ala Pro His Ile Thr Lys Ile Asp Pro Asp Phe
 180 185 190
 Phe Leu Arg Gln Lys Tyr Ser Glu Ile Ser Pro Lys Asp Lys Asp Arg

 195 200 205
 Phe Lys Tyr Glu Lys Arg Leu Phe Asn Asp Arg Thr Asp Ala Glu Phe
 210 215 220
 Tyr Glu Asp Tyr Pro Ser Met Tyr His Leu Arg Leu His Leu Met Thr
 225 230 235 240
 His Thr His Lys Ala Asp Pro Arg Glu Ile Phe Leu Ala Ile His His
 245 250 255
 Ile Leu Lys Ser Arg Gly His Phe Leu Thr Pro Gly Ala Ala Lys Asp

 260 265 270
 Phe Asn Thr Asp Lys Val Asp Leu Glu Asp Ile Phe Pro Ala Leu Thr

275 280 285
 Glu Ala Tyr Ala Gln Val Tyr Pro Asp Leu Glu Leu Thr Phe Asp Leu
 290 295 300
 Ala Lys Ala Asp Asp Phe Lys Ala Lys Leu Leu Asp Glu Gln Ala Thr
 305 310 315 320
 Pro Ser Asp Thr Gln Lys Ala Leu Val Asn Leu Leu Leu Ser Ser Asp

 325 330 335
 Gly Glu Lys Glu Ile Val Lys Lys Arg Lys Gln Val Leu Thr Glu Phe
 340 345 350
 Ala Lys Ala Ile Thr Gly Leu Lys Thr Lys Phe Asn Leu Ala Leu Gly
 355 360 365
 Thr Glu Val Asp Glu Ala Asp Ala Ser Asn Trp Gln Phe Ser Met Gly
 370 375 380
 Gln Leu Asp Asp Lys Trp Ser Asn Ile Glu Thr Ser Met Thr Asp Gln

 385 390 395 400
 Gly Thr Glu Ile Phe Glu Gln Ile Gln Glu Leu Tyr Arg Ala Arg Leu
 405 410 415
 Leu Asn Gly Ile Val Pro Ala Gly Met Ser Leu Ser Gln Ala Lys Val
 420 425 430
 Ala Asp Tyr Gly Gln His Lys Glu Asp Leu Glu Leu Phe Lys Thr Tyr
 435 440 445
 Leu Lys Lys Leu Asn Asp His Glu Leu Ala Lys Thr Ile Arg Gly Leu

 450 455 460
 Tyr Asp Arg Tyr Ile Asn Gly Asp Asp Ala Lys Pro Phe Leu Arg Glu
 465 470 475 480
 Asp Phe Val Lys Ala Leu Thr Lys Glu Val Thr Ala His Pro Asn Glu
 485 490 495
 Val Ser Glu Gln Leu Leu Asn Arg Met Gly Gln Ala Asn Phe Met Leu
 500 505 510
 Lys Gln Arg Thr Lys Ala Asn Gly Ala Ile Pro Ile Gln Leu Gln Gln

 515 520 525

Arg Glu Leu Asp Gln Ile Ile Ala Asn Gln Ser Lys Tyr Tyr Asp Trp
 530 535 540
 Leu Ala Ala Pro Asn Pro Val Glu Ala His Arg Trp Lys Met Pro Tyr
 545 550 555 560
 Gln Leu Asp Glu Leu Leu Asn Phe His Ile Pro Tyr Tyr Val Gly Pro
 565 570 575
 Leu Ile Thr Pro Lys Gln Gln Ala Glu Ser Gly Glu Asn Val Phe Ala
 580 585 590
 Trp Met Val Arg Lys Asp Pro Ser Gly Asn Ile Thr Pro Tyr Asn Phe
 595 600 605
 Asp Glu Lys Val Asp Arg Glu Ala Ser Ala Asn Thr Phe Ile Gln Arg
 610 615 620
 Met Lys Thr Thr Asp Thr Tyr Leu Ile Gly Glu Asp Val Leu Pro Lys
 625 630 635 640
 Gln Ser Leu Leu Tyr Gln Lys Tyr Glu Val Leu Asn Glu Leu Asn Asn
 645 650 655
 Val Arg Ile Asn Asn Glu Cys Leu Gly Thr Asp Gln Lys Gln Arg Leu
 660 665 670
 Ile Arg Glu Val Phe Glu Arg His Ser Ser Val Thr Ile Lys Gln Val
 675 680 685
 Ala Asp Asn Leu Val Ala His Gly Asp Phe Ala Arg Arg Pro Glu Ile
 690 695 700
 Arg Gly Leu Ala Asp Glu Lys Arg Phe Leu Ser Ser Leu Ser Thr Tyr
 705 710 715 720
 His Gln Leu Lys Glu Ile Leu His Glu Ala Ile Asp Asp Pro Thr Lys
 725 730 735
 Leu Leu Asp Ile Glu Asn Ile Ile Thr Trp Ser Thr Val Phe Glu Asp
 740 745 750
 His Thr Ile Phe Glu Thr Lys Leu Ala Glu Ile Glu Trp Leu Asp Pro
 755 760 765
 Lys Lys Ile Asn Glu Leu Ser Gly Ile Arg Tyr Arg Gly Trp Gly Gln

770	775	780	
Phe Ser Arg Lys Leu Leu Asp Gly Leu Lys Leu Gly Asn Gly His Thr			
785	790	795	800
Val Ile Gln Glu Leu Met Leu Ser Asn His Asn Leu Met Gln Ile Leu			
805	810	815	
Ala Asp Glu Thr Leu Lys Glu Thr Met Thr Glu Leu Asn Gln Asp Lys			
820	825	830	
Leu Lys Thr Asp Asp Ile Glu Asp Val Ile Asn Asp Ala Tyr Thr Ser			
835	840	845	
Pro Ser Asn Lys Lys Ala Leu Arg Gln Val Leu Arg Val Val Glu Asp			
850	855	860	
Ile Lys His Ala Ala Asn Gly Gln Asp Pro Ser Trp Leu Phe Ile Glu			
865	870	875	880
Thr Ala Asp Gly Thr Gly Thr Ala Gly Lys Arg Thr Gln Ser Arg Gln			
885	890	895	
Lys Gln Ile Gln Thr Val Tyr Ala Asn Ala Ala Gln Glu Leu Ile Asp			
900	905	910	
Ser Ala Val Arg Gly Glu Leu Glu Asp Lys Ile Ala Asp Lys Ala Ser			
915	920	925	
Phe Thr Asp Arg Leu Val Leu Tyr Phe Met Gln Gly Gly Arg Asp Ile			
930	935	940	
Tyr Thr Gly Ala Pro Leu Asn Ile Asp Gln Leu Ser His Tyr Asp Ile			
945	950	955	960
Asp His Ile Leu Pro Gln Ser Leu Ile Lys Asp Asp Ser Leu Asp Asn			
965	970	975	
Arg Val Leu Val Asn Ala Thr Ile Asn Arg Glu Lys Asn Asn Val Phe			
980	985	990	
Ala Ser Thr Leu Phe Ala Gly Lys Met Lys Ala Thr Trp Arg Lys Trp			
995	1000	1005	
His Glu Ala Gly Leu Ile Ser Gly Arg Lys Leu Arg Asn Leu Met			
1010	1015	1020	

Leu Arg Pro Asp Glu Ile Asp Lys Phe Ala Lys Gly Phe Val Ala

1025

1030

1035

Arg Gln Leu Val Glu Thr Arg Gln Ile Ile Lys Leu Thr Glu Gln

1040

1045

1050

Ile Ala Ala Ala Gln Tyr Pro Asn Thr Lys Ile Ile Ala Val Lys

1055

1060

1065

Ala Gly Leu Ser His Gln Leu Arg Glu Glu Leu Asp Phe Pro Lys

1070

1075

1080

Asn Arg Asp Val Asn His Tyr His His Ala Phe Asp Ala Phe Leu

1085

1090

1095

Ala Ala Arg Ile Gly Thr Tyr Leu Leu Lys Arg Tyr Pro Lys Leu

1100

1105

1110

Ala Pro Phe Phe Thr Tyr Gly Glu Phe Ala Lys Val Asp Val Lys

1115

1120

1125

Lys Phe Arg Glu Phe Asn Phe Ile Gly Ala Leu Thr His Ala Lys

1130

1135

1140

Lys Asn Ile Ile Ala Lys Asp Thr Gly Glu Ile Val Trp Asp Lys

1145

1150

1155

Glu Arg Asp Ile Arg Glu Leu Asp Arg Ile Tyr Asn Phe Lys Arg

1160

1165

1170

Met Leu Ile Thr His Glu Val Tyr Phe Glu Thr Ala Asp Leu Phe

1175

1180

1185

Lys Gln Thr Ile Tyr Ala Ala Lys Asp Ser Lys Glu Arg Gly Gly

1190

1195

1200

Ser Lys Gln Leu Ile Pro Lys Lys Gln Gly Tyr Pro Thr Gln Val

1205

1210

1215

Tyr Gly Gly Tyr Thr Gln Glu Ser Gly Ser Tyr Asn Ala Leu Val

1220

1225

1230

Arg Val Ala Glu Ala Asp Thr Thr Ala Tyr Gln Val Ile Lys Ile

1235

1240

1245

Ser Ala Gln Asn Ala Ser Lys Ile Ala Ser Ala Asn Leu Lys Ser

1250 1255 1260
Arg Glu Lys Gly Lys Gln Leu Leu Asn Glu Ile Val Val Lys Gln

1265 1270 1275
Leu Ala Lys Arg Arg Lys Asn Trp Lys Pro Ser Ala Asn Ser Phe

1280 1285 1290
Lys Ile Val Ile Pro Arg Phe Gly Met Gly Thr Leu Phe Gln Asn

1295 1300 1305
Ala Lys Tyr Gly Leu Phe Met Val Asn Ser Asp Thr Tyr Tyr Arg

1310 1315 1320
Asn Tyr Gln Glu Leu Trp Leu Ser Arg Glu Asn Gln Lys Leu Leu

1325 1330 1335
Lys Lys Leu Phe Ser Ile Lys Tyr Glu Lys Thr Gln Met Asn His

1340 1345 1350
Asp Ala Leu Gln Val Tyr Lys Ala Ile Ile Asp Gln Val Glu Lys

1355 1360 1365
Phe Phe Lys Leu Tyr Asp Ile Asn Gln Phe Arg Ala Lys Leu Ser

1370 1375 1380
Asp Ala Ile Glu Arg Phe Glu Lys Leu Pro Ile Asn Thr Asp Gly

1385 1390 1395
Asn Lys Ile Gly Lys Thr Glu Thr Leu Arg Gln Ile Leu Ile Gly

1400 1405 1410
Leu Gln Ala Asn Gly Thr Arg Ser Asn Val Lys Asn Leu Gly Ile

1415 1420 1425
Lys Thr Asp Leu Gly Leu Leu Gln Val Gly Ser Gly Ile Lys Leu

1430 1435 1440
Asp Lys Asp Thr Gln Ile Val Tyr Gln Ser Pro Ser Gly Leu Phe

1445 1450 1455
Lys Arg Arg Ile Pro Leu Ala Asp Leu Pro Lys Lys Lys Arg Lys

1460 1465 1470
Val Leu Glu Gly Gly Gly Gly Ser Ser Thr Asp His Pro Lys Tyr

1475 1480 1485

Ser Asp Met Ile Val Ala Ala Ile Gln Ala Glu Lys Asn Arg Ala
1490 1495 1500

Gly Ser Ser Arg Gln Ser Ile Gln Lys Tyr Ile Lys Ser His Tyr

1505 1510 1515

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

1520 1525 1530

Arg Leu Val Thr Thr Gly Val Leu Lys Gln Thr Lys Gly Val Gly

1535 1540 1545

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

1550 1555 1560

<210> 123

<211> 1434

<212> PRT

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 123

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu

1 5 10 15

Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val

20 25 30

Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys

35 40 45

Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu

50 55 60

Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu

65 70 75 80

Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu

85 90 95

Ala Lys Ser Asp Thr Gly Ser Gly Met Glu Lys Lys Arg Lys Val Thr

100 105 110

Leu Gly Phe Asp Leu Gly Ile Ala Ser Val Gly Trp Ala Ile Val Asp

115 120 125

Ser Glu Thr Asn Gln Val Tyr Lys Leu Gly Ser Arg Leu Phe Asp Ala

130 135 140

Pro Asp Thr Asn Leu Glu Arg Arg Thr Gln Arg Gly Thr Arg Arg Leu

145 150 155 160

Leu Arg Arg Arg Lys Tyr Arg Asn Gln Lys Phe Tyr Asn Leu Val Lys

165 170 175

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

180 185 190

Phe Arg Glu Leu Ser Ile Lys Tyr Pro Asn Ile Ile Glu Leu Lys Thr

195 200 205

Lys Ala Leu Ser Gln Glu Val Cys Pro Asp Glu Ile Ala Trp Ile Leu

210 215 220

His Asp Tyr Leu Lys Asn Arg Gly Tyr Phe Tyr Asp Glu Lys Glu Thr

225 230 235 240

Lys Glu Asp Phe Asp Gln Gln Thr Val Glu Ser Met Pro Ser Tyr Lys

245 250 255

Leu Asn Glu Phe Tyr Lys Lys Tyr Gly Tyr Phe Lys Gly Ala Leu Ser

260 265 270

Gln Pro Thr Glu Ser Glu Met Lys Asp Asn Lys Asp Leu Lys Glu Ala

275 280 285

Phe Phe Phe Asp Phe Ser Asn Lys Glu Trp Leu Lys Glu Ile Asn Tyr

290 295 300

Phe Phe Asn Val Gln Lys Asn Ile Leu Ser Glu Thr Phe Ile Glu Glu

305 310 315 320

Phe Lys Lys Ile Phe Ser Phe Thr Arg Asp Ile Ser Lys Gly Pro Gly

325 330 335

Ser Asp Asn Met Pro Ser Pro Tyr Gly Ile Phe Gly Glu Phe Gly Asp

340 345 350

Asn Gly Gln Gly Gly Arg Tyr Glu His Ile Trp Asp Lys Asn Ile Gly

355 360 365
 Lys Cys Ser Ile Phe Thr Asn Glu Gln Arg Ala Pro Lys Tyr Leu Pro

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

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

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

 565 570 575
 Gly Ala Gly Asn Glu Phe Glu Asn His Asn Ser Lys Thr His Ser Leu
 580 585 590
 Ser Lys Lys Ala Ile Asp Ala Ile Leu Pro Lys Leu Leu Asp Asn Asn
 595 600 605

Glu Gly Trp Asn Leu Glu Ala Ile Lys Asn Tyr Asp Glu Glu Ile Lys
 610 615 620
 Ser Gln Ile Glu Asp Asn Ser Ser Leu Met Ala Lys Gln Asp Lys Lys

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

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

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

 820 825 830
 Leu Ser Lys Lys Glu Arg Leu Lys Lys Ser Glu Asn Leu Leu Thr Thr
 835 840 845
 Ser Tyr Asp Gly Tyr Glu Lys Leu Gly Phe Leu Ala Arg Asn Leu Asn

850 855 860
 Asp Thr Arg Tyr Ala Thr Ile Leu Phe Arg Asp Gln Leu Asn Asn Tyr
 865 870 875 880
 Ala Glu His His Leu Ile Asp Asn Lys Lys Met Phe Lys Val Ile Ala

 885 890 895
 Met Asn Gly Ala Val Thr Ser Phe Ile Arg Lys Asn Met Ser Tyr Asp
 900 905 910
 Asn Lys Leu Arg Leu Lys Asp Arg Ser Asp Phe Ser His His Ala Tyr
 915 920 925
 Asp Ala Ala Ile Ile Ala Leu Phe Ser Asn Lys Thr Lys Thr Leu Tyr
 930 935 940
 Asn Leu Ile Asp Pro Ser Leu Asn Gly Ile Ile Ser Lys Arg Ser Glu

 945 950 955 960
 Gly Tyr Trp Val Ile Glu Asp Arg Tyr Thr Gly Glu Ile Lys Glu Leu
 965 970 975
 Lys Lys Glu Asp Trp Thr Ser Ile Lys Asn Asn Val Gln Ala Arg Lys
 980 985 990
 Ile Ala Lys Glu Ile Glu Glu Tyr Leu Ile Asp Leu Asp Asp Glu Val
 995 1000 1005
 Phe Phe Ser Arg Lys Thr Lys Arg Lys Thr Asn Arg Gln Leu Tyr

 1010 1015 1020
 Asn Glu Thr Ile Tyr Gly Ile Ala Ala Lys Thr Asp Glu Asp Gly
 1025 1030 1035
 Ile Thr Asn Tyr Tyr Lys Lys Glu Lys Phe Ser Ile Leu Asp Asp
 1040 1045 1050
 Lys Asp Ile Tyr Leu Arg Leu Leu Arg Glu Arg Glu Lys Phe Val
 1055 1060 1065
 Ile Asn Gln Ser Asn Pro Glu Val Ile Asp Gln Ile Ile Glu Ile

 1070 1075 1080
 Ile Glu Ser Tyr Gly Lys Glu Asn Asn Ile Pro Ser Arg Asp Glu
 1085 1090 1095

Ala Ile Asn Ile Lys Tyr Thr	Lys Asn Lys Ile Asn Tyr Asn Leu
1100	1105 1110
Tyr Leu Lys Gln Tyr Met Arg	Ser Leu Thr Lys Ser Leu Asp Gln
1115	1120 1125
Phe Ser Glu Gly Phe Ile Asn	Gln Met Ile Ala Asn Lys Thr Phe
1130	1135 1140
Val Leu Tyr Asn Pro Thr Lys	Asn Thr Thr Arg Lys Ile Lys Phe
1145	1150 1155
Leu Arg Leu Val Asn Asp Val	Lys Ile Asn Asp Ile Arg Lys Asn
1160	1165 1170
Gln Val Ile Asn Lys Phe Asn	Gly Lys Asn Asn Glu Pro Lys Ala
1175	1180 1185
Phe Tyr Glu Asn Ile Asn Ser	Leu Gly Ala Ile Val Phe Lys Ser
1190	1195 1200
Ser Ala Asn Asn Phe Lys Thr	Leu Ser Ile Asn Thr Gln Ile Ala
1205	1210 1215
Ile Phe Gly Asp Lys Asn Trp	Asp Ile Glu Asp Phe Lys Thr Tyr
1220	1225 1230
Asn Met Glu Lys Ile Glu Lys	Tyr Lys Glu Ile Tyr Gly Ile Asp
1235	1240 1245
Lys Thr Tyr Asn Phe His Ser	Phe Ile Phe Pro Gly Thr Ile Leu
1250	1255 1260
Leu Asp Lys Gln Asn Lys Glu	Phe Tyr Tyr Ile Ser Ser Ile Gln
1265	1270 1275
Thr Val Asn Asp Gln Ile Glu	Leu Lys Phe Leu Asn Lys Ile Glu
1280	1285 1290
Phe Lys Asn Asp Asp Asn Thr	Ser Gly Ala Asn Lys Pro Pro Arg
1295	1300 1305
Arg Leu Arg Phe Gly Ile Lys	Ser Ile Met Asn Asn Tyr Glu Gln
1310	1315 1320
Val Asp Ile Ser Pro Phe Gly	Ile Asn Lys Lys Ile Phe Glu Pro

1325 1330 1335
 Lys Lys Lys Arg Lys Val Leu Glu Gly Gly Gly Gly Ser Gly Lys
 1340 1345 1350
 Gly Asp Pro Lys Lys Pro Arg Gly Lys Met Ser Ser Tyr Ala Phe
 1355 1360 1365
 Phe Val Gln Thr Cys Arg Glu Glu His Lys Lys Lys His Pro Asp

1370 1375 1380
 Ala Ser Val Asn Phe Ser Glu Phe Ser Lys Lys Cys Ser Glu Arg
 1385 1390 1395
 Trp Lys Thr Met Ser Ala Lys Glu Lys Gly Lys Phe Glu Asp Met
 1400 1405 1410
 Ala Lys Ala Asp Lys Ala Arg Tyr Glu Arg Glu Met Lys Thr Tyr
 1415 1420 1425
 Ile Pro Pro Lys Gly Glu
 1430

<210> 124

<211> 1431

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
 polypeptide"

<400> 124

Met Pro Lys Arg Lys Val Ser Ser Ala Glu Gly Ala Ala Lys Glu Glu
 1 5 10 15
 Pro Lys Arg Arg Ser Ala Arg Leu Ser Ala Lys Pro Pro Ala Lys Val
 20 25 30
 Glu Ala Lys Pro Lys Lys Ala Ala Ala Lys Asp Lys Ser Ser Asp Lys
 35 40 45
 Lys Val Gln Thr Lys Gly Lys Arg Gly Ala Lys Gly Lys Gln Ala Glu
 50 55 60
 Val Ala Asn Gln Glu Thr Lys Glu Asp Leu Pro Ala Glu Asn Gly Glu

65	70	75	80
Thr Lys Thr Glu Glu Ser Pro Ala Ser Asp Glu Ala Gly Glu Lys Glu			
	85	90	95
Ala Lys Ser Asp Thr Gly Ser Gly Met Glu Lys Lys Arg Lys Val Thr			
	100	105	110
Leu Gly Phe Asp Leu Gly Ile Ala Ser Val Gly Trp Ala Ile Val Asp			
	115	120	125
Ser Glu Thr Asn Gln Val Tyr Lys Leu Gly Ser Arg Leu Phe Asp Ala			
	130	135	140
Pro Asp Thr Asn Leu Glu Arg Arg Thr Gln Arg Gly Thr Arg Arg Leu			
145	150	155	160
Leu Arg Arg Arg Lys Tyr Arg Asn Gln Lys Phe Tyr Asn Leu Val Lys			
	165	170	175
Arg Thr Glu Val Phe Gly Leu Ser Ser Arg Glu Ala Ile Glu Asn Arg			
	180	185	190
Phe Arg Glu Leu Ser Ile Lys Tyr Pro Asn Ile Ile Glu Leu Lys Thr			
	195	200	205
Lys Ala Leu Ser Gln Glu Val Cys Pro Asp Glu Ile Ala Trp Ile Leu			
	210	215	220
His Asp Tyr Leu Lys Asn Arg Gly Tyr Phe Tyr Asp Glu Lys Glu Thr			
225	230	235	240
Lys Glu Asp Phe Asp Gln Gln Thr Val Glu Ser Met Pro Ser Tyr Lys			
	245	250	255
Leu Asn Glu Phe Tyr Lys Lys Tyr Gly Tyr Phe Lys Gly Ala Leu Ser			
	260	265	270
Gln Pro Thr Glu Ser Glu Met Lys Asp Asn Lys Asp Leu Lys Glu Ala			
	275	280	285
Phe Phe Phe Asp Phe Ser Asn Lys Glu Trp Leu Lys Glu Ile Asn Tyr			
	290	295	300
Phe Phe Asn Val Gln Lys Asn Ile Leu Ser Glu Thr Phe Ile Glu Glu			
305	310	315	320

Phe Lys Lys Ile Phe Ser Phe Thr Arg Asp Ile Ser Lys Gly Pro Gly
 325 330 335
 Ser Asp Asn Met Pro Ser Pro Tyr Gly Ile Phe Gly Glu Phe Gly Asp
 340 345 350
 Asn Gly Gln Gly Gly Arg Tyr Glu His Ile Trp Asp Lys Asn Ile Gly
 355 360 365

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

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

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

 Ser Leu Asn Ile Glu Asn Leu Lys Leu Lys Leu Lys Glu Phe Leu Leu

565 570 575
 Gly Ala Gly Asn Glu Phe Glu Asn His Asn Ser Lys Thr His Ser Leu
 580 585 590
 Ser Lys Lys Ala Ile Asp Ala Ile Leu Pro Lys Leu Leu Asp Asn Asn
 595 600 605
 Glu Gly Trp Asn Leu Glu Ala Ile Lys Asn Tyr Asp Glu Glu Ile Lys
 610 615 620

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

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

 Gly Ala Lys Ile Ser Ala Asn Glu Ile Val Asp Asn Lys Val Glu Ile
 755 760 765
 Asp His Ile Ile Pro Tyr Ser Leu Cys Phe Asp Asp Ser Ser Ala Asn
 770 775 780
 Lys Val Leu Val His Lys Gln Ser Asn Gln Glu Lys Ser Asn Ser Leu
 785 790 795 800
 Pro Tyr Glu Tyr Ile Lys Gln Gly His Ser Gly Trp Asn Trp Asp Glu
 805 810 815

Phe Thr Lys Tyr Val Lys Arg Val Phe Val Asn Asn Val Asp Ser Ile
820 825 830

Leu Ser Lys Lys Glu Arg Leu Lys Lys Ser Glu Asn Leu Leu Thr Thr
835 840 845

Ser Tyr Asp Gly Tyr Glu Lys Leu Gly Phe Leu Ala Arg Asn Leu Asn
850 855 860

Asp Thr Arg Tyr Ala Thr Ile Leu Phe Arg Asp Gln Leu Asn Asn Tyr
865 870 875 880

Ala Glu His His Leu Ile Asp Asn Lys Lys Met Phe Lys Val Ile Ala
885 890 895

Met Asn Gly Ala Val Thr Ser Phe Ile Arg Lys Asn Met Ser Tyr Asp
900 905 910

Asn Lys Leu Arg Leu Lys Asp Arg Ser Asp Phe Ser His His Ala Tyr
915 920 925

Asp Ala Ala Ile Ile Ala Leu Phe Ser Asn Lys Thr Lys Thr Leu Tyr
930 935 940

Asn Leu Ile Asp Pro Ser Leu Asn Gly Ile Ile Ser Lys Arg Ser Glu
945 950 955 960

Gly Tyr Trp Val Ile Glu Asp Arg Tyr Thr Gly Glu Ile Lys Glu Leu
965 970 975

Lys Lys Glu Asp Trp Thr Ser Ile Lys Asn Asn Val Gln Ala Arg Lys
980 985 990

Ile Ala Lys Glu Ile Glu Glu Tyr Leu Ile Asp Leu Asp Asp Glu Val
995 1000 1005

Phe Phe Ser Arg Lys Thr Lys Arg Lys Thr Asn Arg Gln Leu Tyr
1010 1015 1020

Asn Glu Thr Ile Tyr Gly Ile Ala Ala Lys Thr Asp Glu Asp Gly
1025 1030 1035

Ile Thr Asn Tyr Tyr Lys Lys Glu Lys Phe Ser Ile Leu Asp Asp
1040 1045 1050

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

1055	1060	1065
Ile Asn Gln Ser Asn Pro Glu Val Ile Asp Gln Ile Ile Glu Ile		
1070	1075	1080
Ile Glu Ser Tyr Gly Lys Glu Asn Asn Ile Pro Ser Arg Asp Glu		
1085	1090	1095
Ala Ile Asn Ile Lys Tyr Thr Lys Asn Lys Ile Asn Tyr Asn Leu		
1100	1105	1110
Tyr Leu Lys Gln Tyr Met Arg Ser Leu Thr Lys Ser Leu Asp Gln		
1115	1120	1125
Phe Ser Glu Gly Phe Ile Asn Gln Met Ile Ala Asn Lys Thr Phe		
1130	1135	1140
Val Leu Tyr Asn Pro Thr Lys Asn Thr Thr Arg Lys Ile Lys Phe		
1145	1150	1155
Leu Arg Leu Val Asn Asp Val Lys Ile Asn Asp Ile Arg Lys Asn		
1160	1165	1170
Gln Val Ile Asn Lys Phe Asn Gly Lys Asn Asn Glu Pro Lys Ala		
1175	1180	1185
Phe Tyr Glu Asn Ile Asn Ser Leu Gly Ala Ile Val Phe Lys Ser		
1190	1195	1200
Ser Ala Asn Asn Phe Lys Thr Leu Ser Ile Asn Thr Gln Ile Ala		
1205	1210	1215
Ile Phe Gly Asp Lys Asn Trp Asp Ile Glu Asp Phe Lys Thr Tyr		
1220	1225	1230
Asn Met Glu Lys Ile Glu Lys Tyr Lys Glu Ile Tyr Gly Ile Asp		
1235	1240	1245
Lys Thr Tyr Asn Phe His Ser Phe Ile Phe Pro Gly Thr Ile Leu		
1250	1255	1260
Leu Asp Lys Gln Asn Lys Glu Phe Tyr Tyr Ile Ser Ser Ile Gln		
1265	1270	1275
Thr Val Asn Asp Gln Ile Glu Leu Lys Phe Leu Asn Lys Ile Glu		
1280	1285	1290

Phe Lys Asn Asp Asp Asn Thr Ser Gly Ala Asn Lys Pro Pro Arg
 1295 1300 1305

Arg Leu Arg Phe Gly Ile Lys Ser Ile Met Asn Asn Tyr Glu Gln
 1310 1315 1320

Val Asp Ile Ser Pro Phe Gly Ile Asn Lys Lys Ile Phe Glu Pro
 1325 1330 1335

Lys Lys Lys Arg Lys Val Leu Glu Gly Gly Gly Gly Ser Ser Thr
 1340 1345 1350

Asp His Pro Lys Tyr Ser Asp Met Ile Val Ala Ala Ile Gln Ala
 1355 1360 1365

Glu Lys Asn Arg Ala Gly Ser Ser Arg Gln Ser Ile Gln Lys Tyr
 1370 1375 1380

Ile Lys Ser His Tyr Lys Val Gly Glu Asn Ala Asp Ser Gln Ile
 1385 1390 1395

Lys Leu Ser Ile Lys Arg Leu Val Thr Thr Gly Val Leu Lys Gln
 1400 1405 1410

Thr Lys Gly Val Gly Ala Ser Gly Ser Phe Arg Leu Ala Lys Ser
 1415 1420 1425

Asp Glu Pro
 1430

<210> 125

<211> 26

<212> DNA

<213> Unknown

<220><221> source

<223> /note="Description of Unknown:

Cas9 target site"

<400> 125

atagatgcgg ccaaggtgta catggg

26

<210> 126

<211> 26

<212> DNA

<213> Unknown

<220><221> source

<223> /note="Description of Unknown:
 Cas9 target site"

<400> 126

ctacgagaac cagaagccgt gagtgg 26

<210> 127

<211> 30

<212> DNA

<213> Unknown

<220><221> source

<223> /note="Description of Unknown:
 Cas9 target site"

<400> 127

ctcatacatg gctataatag aaggagcaaa 30

<210> 128

<211> 30

<212> DNA

<213> Unknown

<220><221> source

<223> /note="Description of Unknown:
 Cas9 target site"

<400> 128

gaaggtgtgg ttccagaacc ggaggacaaa 30

<210> 129

<211> 30

<212> DNA

<213> Unknown

<220><221> source

<223> /note="Description of Unknown:
 Cas9 target site"

<400> 129

tggttccaga accggaggac aaagtacaaa 30

<210> 130
 <211> 23
 <212> DNA
 <213> Unknown
 <220><221> source
 <223> /note="Description of Unknown:
 Cas9 target site"
 <400> 130
 tgtacatggg ggagatgggc cgg 23
 <210> 131
 <211> 23
 <212> DNA
 <213> Unknown
 <220><221> source
 <223> /note="Description of Unknown:
 Cas9 target site"
 <400> 131
 ggggccacta gggacaggat tgg 23
 <210> 132
 <211> 25
 <212> DNA
 <213> Unknown
 <220><221> source
 <223> /note="Description of Unknown:
 Cas9 target site"
 <400> 132
 agctcgtact ggcgaatgct ggaaa 25
 <210> 133
 <211> 25
 <212> DNA
 <213> Unknown
 <220><221> source
 <223> /note="Description of Unknown:
 Cas9 target site"

<400> 133

cctctggcac atcctccaaa tgaaa

25

<210> 134

<211> 6

<212> PRT

<213> Artificial Sequence

<

220><221> source

<223> /note="Description of Artificial Sequence: Synthetic

6xHis tag"

<400> 134

His His His His His His

1

5