



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2025년02월18일

(11) 등록번호 10-2769515

(24) 등록일자 2025년02월13일

(51) 국제특허분류(Int. Cl.)
C12N 15/85 (2006.01) **C12N 15/90** (2006.01)
C12N 9/12 (2006.01) **C12N 9/22** (2006.01)
(52) CPC특허분류
C12N 15/85 (2013.01)
C12N 15/8509 (2013.01)
(21) 출원번호 10-2017-7031337
(22) 출원일자(국제) 2016년03월31일
심사청구일자 2021년02월26일
(85) 번역문제출일자 2017년10월30일
(65) 공개번호 10-2018-0029953
(43) 공개일자 2018년03월21일
(86) 국제출원번호 PCT/US2016/025426
(87) 국제공개번호 WO 2016/161207
국제공개일자 2016년10월06일
(30) 우선권주장
62/140,454 2015년03월31일 미국(US)
(뒷면에 계속)
(56) 선행기술조사문헌
미국 특허출원공개공보
US2015/071898(2015.3.12.)*
(뒷면에 계속)

(73) 특허권자
소흥, 인코포레이티드
미국, 캘리포니아 (우편번호 91709), 치노 힐스,
치노 힐스 파크웨이 4195, 스위트 675
(72) 발명자
세이크, 페러크
미국 캘리포니아 91362, 5505 레이크뷰 캐년 로드
웨스트레이크 빌리지
카와무라, 테츠야
미국 캘리포니아 92122, 8148 제네시 애비뉴
#137 샌 디에고
모, 글로리아
미국 캘리포니아 92122, 8148 제네시 애비뉴
#137 샌 디에고
(74) 대리인
김태홍, 김진희

전체 청구항 수 : 총 17 항

심사관 : 최성호

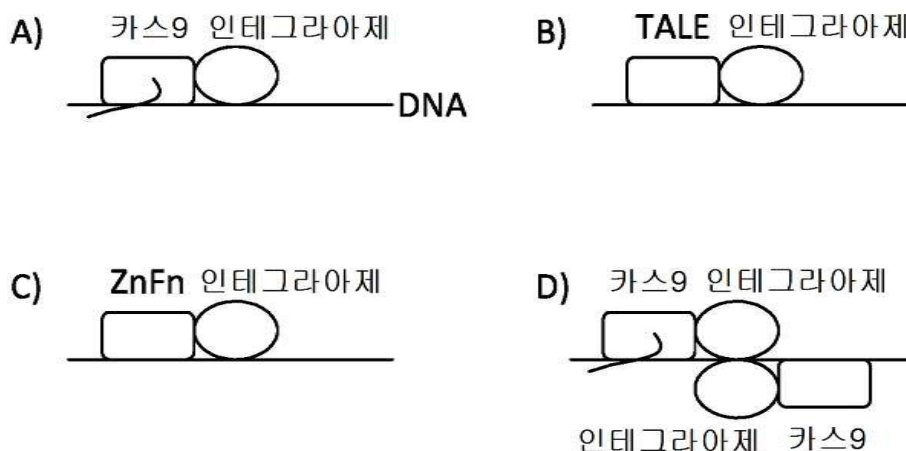
(54) 발명의 명칭 세포 또는 유기체의 게놈으로의 DNA 서열의 표적화 혼입을 위한 Cas 9 레트로바이러스 인테그라제 시스템 및 Cas 9 재조합효소 시스템

(57) 요약

본 개시 내용은, 관심 대상의 DNA 서열(또는 관심 대상의 유전자)을 세포 또는 유기체의 게놈 내 표적된 위치로 전달하기 위하여, 바이러스성 인테그라아제, 재조합효소, 또는 전이효소가 부착된 카스9, Cpf1, TALE 및 징크 핑거 단백질(zinc finger protein)과 같은 공학적으로 조작된 단백질의 이용에 관한 것이다. DNA 절단에 있어서 그

(뒷면에 계속)

대표도 - 도1



의 기능에 대해 비활성인 카스9의 이용은, 상동성 재조합을 위한 기타 다른 시스템에서 의도되는 바와 같은 DNA 파괴(break)를 유발하지 않으면서, RNA 가이드의 이용에 의해 DNA를 표적하는 카스9 단백질 능력의 이용을 가능하게 할 것이다. 바이러스성 인테그라아제 또는 재조합효소에 부착된 징크 핑거 단백질 또는 TALE(DNA의 특정 서열에 결합하는 공학적으로 조작된 단백질)의 이용이 또한 개시된다. 이러한 시스템은 실험실 및 치료 목적을 위해 사용될 수 있다. 예를 들어, 관심 대상의 유전자는, 세포 내 정상 유전자 산물을 회복하기 위하여 그의 유전자 산물을 생산하는 능력을 결여한 유전자를 갖는 세포 내로 포함될 수 있다(예를 들어, 유전자 산물은 단백질 또는 특수화된 RNA일 수 있다).

(52) CPC특허분류

C12N 15/907 (2013.01)

C12N 9/1241 (2013.01)

C12N 9/22 (2013.01)

C07K 2319/80 (2013.01)

C07K 2319/81 (2013.01)

(56) 선행기술조사문헌

W02014182700 A1

Methods., 47(4):269-276(2009.4.)*

US20060252140 A1

US20140193915 A1

Journal of Virology., 71(1):458-464(1997.)

*는 심사관에 의하여 인용된 문헌

(30) 우선권주장

62/210,451 2015년08월27일 미국(US)

62/240,359 2015년10월12일 미국(US)

명세서

청구범위

청구항 1

(a) (i) 가이드 RNA(gRNA)의 존재 하에 표적 DNA에 특이적으로 결합할 수 있고 뉴클레아제 활성이 결여된 카스9(Cas9) 단백질을 포함하는 제1 단백질 도메인,

(ii) 인테그라아제(integrase)를 포함하는 제2 단백질 도메인, 및

(iii) 제1 단백질 도메인을 제2 단백질 도메인에 연결하는 링커

를 포함하는 융합 단백질;

(b) 단지 하나의 gRNA; 및

(c) 도너 핵산, 제1 긴 말단 반복서열(LTR), 및 제2 LTR을 포함하는 폴리뉴클레오티드로서, 도너 핵산이 제1 LTR과 제2 LTR 사이에 위치하는 것인 폴리뉴클레오티드

를 포함하는, 게놈에 도너 핵산을 삽입하기 위한 시스템.

청구항 2

제1항에 있어서, 제1 단백질 도메인은 서열번호 58의 잔기 348 내지 1715에 대해 95% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하고, 서열번호 58의 잔기 348 내지 1715에 대해 95% 이상의 서열 동일성을 갖는 아미노산 서열 중 임의의 아미노산 치환은 보존적 아미노산 치환인 것인 시스템.

청구항 3

제1항에 있어서, 제1 단백질 도메인은 서열번호 58의 잔기 348 내지 1715의 아미노산 서열을 포함하는 것인 시스템.

청구항 4

제1항에 있어서, 제2 단백질 도메인은 서열번호 58의 잔기 44 내지 331에 대해 95% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하고, 서열번호 58의 잔기 44 내지 331에 대해 95% 이상의 서열 동일성을 갖는 아미노산 서열 중 임의의 아미노산 치환은 보존적 아미노산 치환인 것인 시스템.

청구항 5

제1항에 있어서, 제2 단백질 도메인은 서열번호 58의 잔기 44 내지 331의 아미노산 서열을 포함하는 것인 시스템.

청구항 6

제1항에 있어서, 인테그라아제는 HIV1 인테그라아제인 것인 시스템.

청구항 7

제1항에 있어서, 인테그라아제는 렌티바이러스성(lentiviral) 인테그라아제인 것인 시스템.

청구항 8

제1항에 있어서, 링커는 길이가 1개 내지 16개의 아미노산인 것인 시스템.

청구항 9

제1항에 있어서, 링커는 길이가 4개 내지 8개의 아미노산인 것인 시스템.

청구항 10

제1항에 있어서, 융합 단백질은 길이가 16개 내지 24개의 염기쌍인 표적 DNA 서열에 특이적으로 결합할 수 있는 것인 시스템.

청구항 11

제1항에 있어서, 가이드 RNA는 16개 내지 24개 염기의 길이를 갖는 것인 시스템.

청구항 12

제1항에 있어서, 제1 LTR 또는 제2 LTR은 서열번호 59 및 60으로부터 선택된 뉴클레오티드 서열을 포함하는 것인 시스템.

청구항 13

제1항에 있어서, 융합 단백질은 서열번호 58에 대해 95% 이상의 서열 동일성을 갖는 아미노산 서열을 포함하고, 서열번호 58에 대해 95% 이상의 서열 동일성을 갖는 아미노산 서열 중 임의의 아미노산 치환은 보존적 아미노산 치환인 것인 시스템.

청구항 14

제1항에 있어서, 융합 단백질은 서열번호 58의 아미노산 서열을 포함하고, 융합 단백질은 4량체인 것인 시스템.

청구항 15

DNA 서열을 세포의 표적 게놈 DNA로 삽입함으로써 세포를 유전적으로 변형시키는 방법으로서, 제1항 내지 제14항 중 어느 한 항의 시스템을 세포에 도입하는 단계를 포함하고, gRNA는 표적 게놈 DNA에 특이적으로 결합하는 것인 방법.

청구항 16

제15항에 있어서, 역전사효소(reverse transcriptase)를 암호화하는 핵산을 세포에 도입하는 단계를 더 포함하는 것인 방법.

청구항 17

세포에서 표적 유전자의 발현을 감소시키는 방법으로서, 제15항의 방법을 수행하는 단계를 포함하고, 표적 게놈 서열은 표적 유전자의 제1 코딩 엑손이고, 도너 핵산은 하나 이상의 연속적인 정지 코돈을 포함하는 것인 방법.

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

발명의 설명

기술 분야

[0001] 관련 출원에 대한 상호 참고문헌

[0002] 본 출원은, 2015년 3월 31일 출원된 미국 가출원 62,140,454, 2015년 8월 27일 출원된 미국 가출원 62,210,451, 및 2015년 10월 12일 출원된 미국 가출원 62,240,359의 이익을 주장하며, 이들 각각의 전체 내용은 모든 목적에 대해 참고문헌으로 포함된다.

[0003] 서론

[0004] 본 개시 내용은, 관심 대상의 DNA 서열(또는 관심 대상의 유전자)을 세포 또는 유기체의 게놈 내 표적된 위치로 전달하기 위하여, 링커에 의해 바이러스성 인테그라아제(integrase)(예를 들어, HIV 또는 MMTV 인테그라아제) 또는 재조합효소와 부착된, 카스(Cas)9(크리스퍼(CRISPR): 주기적으로 간격을 띠고 분포하는 짧은 회문구조의 반복서열(clustered regularly interspaced short palindromic repeats) 단백질), TALE 및 징크 핑거 단백질(Zinc finger protein)과 같이 게놈 특이성을 나타내는 DNA 결합 단백질을 갖는 공학적으로 조작된(engineered) 단백질의 이용에 관한 것이다. DNA를 절단하는데 있어서 그의 기능에 대해 비활성인 카스9의 이용은, 상동성 재조합을 위하여 기타 다른 시스템에서 의도되는 바와 같은 DNA 파괴를 유발하지 않으면서 RNA 가이드(gRNA)의 이용에 의해 DNA를 표적하는 카스9 단백질 능력의 이용을 가능하게 할 것이다. 바이러스성 인테그라아제 또는 재조합효소에 부착된 징크 핑거 단백질 또는 TALE(DNA의 특정 서열에 결합하는 공학적으로 조작된 단백질)의 이용이 또한 개시된다. 본 시스템은 실험실 및 치료 목적을 위해 사용될 수 있다. 예를 들어, 관심 대상의 유전자(들)를 함유하는 도너(donor) DNA는, 기존의 방법을 통해 비-표적(off target) 절단 가능성 없이 숙주 게놈 내로 쉽게 도입될 수 있다. 도너 DNA는 또한 "녹아웃(knock out)" 전략을 촉진하기 위해 공학적으로 조작될 수 있다. 카스9 표적화의 특이성을 개선하기 위한 새로운 전략이 또한 논의된다. 이 전략은, 어떤 가이드 RNA가 카스9의 특이적 표적화를 제공하는지 발견하기 위한 분석에서, 가이드 RNA와 게놈성 DNA와 함께 표면 결합된 d카스9(그의 DNA 절단능에 대해 비활성인 카스 9)를 이용한다. 이는 크리스퍼/카스9의 생체 내 적용에서 특히 중요할 것이며, 실리코(silico) 예측 모델의 현재 한계를 극복하지만, 이는 또한 실리코 예측 모델과 연결하여 어떤 gRNA가 분석에서 사용될 지를 지식적으로 결정하는 데에 사용될 수 있다.

배경 기술

[0005] 게놈 서열 분석(sequencing) 기법 및 분석 방법에서의 현재의 진보는, 다양한 범위의 생물학적 기능 및 질병과 연관된 유전적/게놈성 인자를 목록화하고 지도화하는 능력을 현저히 가속화시켜 왔다. 개별적인 유전적 요소의 선택적 교란(perturbation)을 가능하게 함으로써 원인(causal) 유전자 변이의 체계적인 역조작(reverse engineering)을 가능하게 하는데, 또한 합성 생물학, 생명공학, 및 의학 응용분야를 진보시키는 데, 정확한 게놈 표적화 기술이 요구된다. 설계자 징크 핑거, 전사 활성화자-유사 이펙터(TALE: transcription activator-like effector), 크리스퍼/카스9 또는 메가뉴클레아제(meganuclease)와 같은 게놈-편집(genome-editing) 기법이 표적된 게놈 교란을 생성하는데 이용가능하며, 소정의 게놈 내 특정 장소(location) 내로 DNA 서열(전체 유전자 서열 포함)의 혼입을 가능하게 할 신규 게놈 공학적 조작 기술에 대한 요구가 남아있다. 이는 공학적으로 조작된 유전자를 발현하는 세포주 또는 유전자이식(transgenic) 유기체의 생산, 또는 그를 필요로 하는 대상체에서 기능장애 유전자의 대체를 가능하게 할 것이다.

[0006] 인테그라아제는 숙주 게놈(포유동물, 인간, 마우스, 래트, 원숭이, 개구리, 물고기, 식물(작물 및 애기장대와 같은 실험용 식물), 실험실용 또는 생의학적 세포주 또는 일차 세포 배양물, C. 엘레간스(C. elegans), 파리(초파리), 등) 내로 바이러스성 핵산의 삽입을 가능하게 하는 바이러스성 단백질이다. 인테그라아제는, 바이러스성 핵산 서열을 숙주 게놈 내로 혼입시키기 위하여 인테그라아제를 숙주 게놈과 회합시키는데 숙주의 DNA 결합 단

백질을 이용한다. 인테그라아제는 HIV(인간 면역결핍 바이러스)와 같은 레트로바이러스에서 발견된다. 인테그라아제는 그들의 게놈이 숙주 DNA 내로 삽입되도록 하는 바이러스성 유전자에 대한 서열에 따라 달라진다. 문헌 [Leavitt et. al, Journal of Biological Chemistry, 1993, volume 268, pages 2113-2119]은 위치 지정 돌연변이 유발 및 시험관 내 연구를 이용함으로써 HIV1 인테그라아제의 기능을 시험하였다. Leavitt는 또한 바이러스성 인테그라아제에 의한 숙주 게놈 내로의 HIV1 DNA(역전사 후 생성됨)의 통합에 중요한 U5 및 U3 HIV1 att 위치의 서열을 표시한다.

[0007] 본 개시 내용은, 원하는 핵산(DNA) 서열을 게놈 내 특정 장소에서 그 게놈 내로 특이적으로 삽입하는 것을 가능하게 함으로써 현재의 게놈 편집 기술을 개선한다. DNA 결합능을 갖는 재조합 공학적으로 조작된 인테그라아제(또는 재조합효소)는 게놈 내 소정의 DNA 서열에 결합하고, 위치 특이적 방식으로 게놈 내로 소정의 핵산 서열을 삽입하기 위하여, 인테그라아제 인식 도메인(예컨대, HIV1(또는 기타 다른 레트로바이러스) att 위치)을 갖는 제공된 DNA 서열 및/또는 상동염기서열(homology arm)을 인식할 것이다. 본 개시 내용의 일 양태는 유전자의 전사 출발 위치 직후 정지 코돈(UAA, UAG 및/또는 UGA)의 DNA 서열을 삽입하는 것을 포함한다. 이는 세포 또는 유기체의 게놈 내 유전자 전사의 효과적인 저해를 가능하게 할 것이다.

발명의 내용

해결하려는 과제

[0008] 본 개시 내용은, 징크 핑거 단백질, 탈렌(TALEN) 및 크리스퍼/카스9, 또는 CpF1과 같은 기타 다른 크리스퍼 단백질 등을 포함하는 DNA 타겟팅 기술을, 레트로바이러스성 인테그라아제와 연결하여 DNA 타겟팅 인테그라아제를 형성한다. 그 다음, 관심 대상의 유전자(GOI)는 DNA 표적화 인테그라아제와 제공되어 이는 표적된 방식으로 게놈 내로 혼입될 수 있다. GOI는 상동염기서열을 이용하여 설계되어 그의 게놈 내 삽입에 대해 또 다른 수준의 특이성을 제공할 것이다.

[0009] 본 개시 내용은 특히 레트로바이러스성 인테그라아제와의 연결을 위해 DNA를 절단하는데 비활성인 변종 카스9의 이용에 관한 것이다.

[0010] 본 개시 내용은 다음을 포함하는 시스템을 포함한다: A) 예를 들어, DNA 절단능에 대해 비활성인 카스 단백질(예를 들어, 카스9)에 공유결합적으로 연결된 바이러스성 인테그라아제(또는 박테리아성 재조합효소)를 포함한다. 대안적으로, 바이러스성 인테그라아제(또는 재조합효소)는, TALE 단백질 또는 징크 핑거 단백질에 공유결합적으로 연결되며, 여기서 이들 단백질은 게놈에서 DNA의 특정 서열을 표적하도록 설계된다. 이는 발현 벡터 내에 또는 정제된 단백질로서 제공될 수 있다; B) 원하는 게놈 내에 혼입되는, 상동염기서열이 있거나 또는 없는 관심 대상의 유전자(또는 관심 대상의 DNA 서열). 관심 대상의 GOI 또는 DNA 서열은 필요에 따라 바이러스성 인테그라아제에 의해 인식되도록 변형될 수 있다. 기타 다른 시약이 폴리뉴클레오타이드 형질감염 및/또는 단백질의 세포 내로의 도입에 필요하였다. DNA 서열의 비표적 통합에 대한 분석. 일 양태에서, 삽입된 DNA 서열 내로 공학적으로 조작된 마커 서열을 이용한다.

[0011] 본 명세서는, 작동적으로 연결된, a) 카스9, 비활성 카스9, 또는 Cpf1, 또는 이의 일부를 암호화하는 제1 폴리뉴클레오타이드 서열; b) 인테그라아제, 재조합효소, 또는 전이효소(transposase), 또는 이의 일부를 암호화하는 제2 폴리뉴클레오타이드 서열; 및 c) 핵산 링커를 암호화하는 제3 폴리뉴클레오타이드를 포함하고; 이때 제1 폴리뉴클레오타이드 서열은 5' 및 3' 말단을 포함하고, 제2 폴리뉴클레오타이드 서열은 5' 및 3' 말단을 포함하고, 제1 폴리뉴클레오타이드의 3' 말단은 핵산 링커에 의해 제2 폴리뉴클레오타이드의 5' 말단에 연결되고, 제1 및 제2 폴리뉴클레오타이드는 세포 또는 유기체 내 융합 단백질로서 발현될 수 있는, 핵산 구조물이 제공된다. 일부 구현예에서, 제1 폴리뉴클레오타이드 서열은 서열번호 1, 3, 5, 7, 9, 11, 13, 27 내지 46, 49, 56, 또는 68, 또는 이에 대해 80% 이상, 85% 이상, 90% 이상, 95% 이상, 또는 99% 이상의 동일성을 갖는 서열 중 임의의 하나를 포함한다. 일부 구현예에서, 카스9, 비활성 카스9, 또는 Cpf1은 서열번호 2, 4, 6, 8, 10, 12, 14, 50, 52, 69, 72 내지 78, 또는 86 내지 92, 또는 이에 대해 80% 이상, 85% 이상, 90% 이상, 95% 이상, 또는 99% 이상의 동일성을 갖는 서열 중 임의의 하나를 포함한다. 일부 구현예에서, 제2 폴리뉴클레오타이드 서열은 서열번호 15, 17, 19, 21, 23, 47, 55, 62, 64, 66, 70, 또는 79, 또는 이에 대해 80% 이상, 85% 이상, 90% 이상, 95% 이상, 99% 이상의 동일성을 갖는 서열 중 임의의 하나를 포함한다. 일부 구현예에서, 인테그라아제, 재조합효소, 또는 전이효소는 서열번호 16, 18, 20, 22, 24, 25, 26, 48, 63, 65, 67, 71, 또는 80, 또는 이에 대해 80% 이상, 85% 이상, 90% 이상, 95% 이상 또는 99% 이상의 동일성을 갖는 서열 중 임의의 하나를 포함한다. 본 명세서는 또한 상기 핵산 구조물을 포함하는 유기체를 기재한다. 본 명세서는 또한 융합 단백질을 포함하는 유기체를 기

재하며 이때 유기체는 변형된 게놈을 갖는다.

[0012] 본 명세서에는, a) 카스9, 비활성 카스9, 또는 Cpf1, 또는 이의 일부를 암호화하는 제1 폴리뉴클레오타이드 서열; b) 인테그라아제, 재조합효소, 또는 전이효소, 또는 이의 일부를 암호화하는 제2 폴리뉴클레오타이드 서열; 및 c) 핵산 링커를 암호화하는 제3 폴리뉴클레오타이드 서열을 포함하는 유기체를 제공하고, 이때 제1 폴리뉴클레오타이드 서열은 5' 및 3' 말단을 포함하고, 제2 폴리뉴클레오타이드 서열은 5' 및 3' 말단을 포함하고, 제1 폴리뉴클레오타이드의 3' 말단은 핵산 링커에 의해 제2 폴리뉴클레오타이드의 5' 말단에 연결되고, 제1 및 제2 폴리뉴클레오타이드는 세포 또는 유기체 내 융합 단백질로서 발현될 수 있다.

[0013] 본 명세서에는 또한, a) 촉매적으로 비활성인 카스9, 카스9, TALE 단백질, 징크 핑거 단백질, 또는 Cpf1 단백질인, 표적 DNA 서열로 표적되는 제1 단백질; b) 인테그라아제, 재조합효소, 또는 전이효소인 제2 단백질; 및 c) 제2 단백질에 제1 단백질을 연결하는 링커를 포함하는 융합 단백질을 제공한다. 일부 구현예에서, 제2 단백질은 인테그라아제이거나; 인테그라아제는 HIV1 인테그라아제 또는 렌티바이러스성(lentiviral) 인테그라아제이거나; 링커 서열은 길이가 하나 이상인 아미노산이거나; 제1 단백질은 촉매적으로 비활성인 카스9이다. 일부 구현예에서, 링커 서열은 길이가 4 개 내지 8 개의 아미노산이거나; 제1 단백질은 TALE 단백질이거나; 제1 단백질은 징크 핑거 단백질이다. 융합 단백질이 TALE 또는 징크 핑거 단백질을 포함하는 일부 구현예에서, 표적 DNA 서열은 길이가 약 16 개 내지 약 24 개의 염기쌍이다. 일부 구현예에서, 제1 단백질은 카스9 또는 촉매적으로 비활성인 카스9이고, 이때 하나 이상의 가이드 RNA가 약 16 개 내지 약 24 개 염기쌍의 표적 DNA 서열의 표적화에 사용된다.

[0014] 본 명세서에는 또한, a) 게놈성 DNA 내 표적 서열을 확인하는 단계; b) 제1항에 따른 융합 단백질이 게놈성 DNA 내 표적 서열에 결합하도록 설계하는 단계; 3) 관심 대상의 DNA 서열이 게놈성 DNA 내로 혼입되도록 설계하는 단계; 및 d) 세포 또는 유기체 내로 융합 단백질 및 관심 대상의 DNA 서열의 도입을 가능하게 하는 기법에 의해, 세포 또는 유기체에 융합 단백질 및 관심 대상의 DNA 서열을 제공하는 단계를 포함하고, 이때 관심 대상의 DNA 서열은 게놈성 DNA의 표적 서열에 통합되는, 게놈 DNA 내로 DNA 서열을 삽입하는 방법을 제공한다.

[0015] 본 명세서에는 또한, a) 카스9, 촉매적으로 비활성인 카스9, TALE 단백질, 징크 핑거 단백질, 표적 DNA 서열에 결합하도록 공학적으로 조작된 Cpf1 단백질인 제1 단백질에 대한 제1 코딩 서열; b) 인테그라아제, 재조합효소, 또는 전이효소인 제2 단백질에 대한 제2 코딩 서열; c) 제1 및 제2 단백질 사이에 아미노산 링커를 형성하는 제1 및 제2 코딩 서열 사이의 DNA 서열; d) 선택적으로, 인테그라아제에 의해 인식된 att 위치에 의해 둘러싸인 관심 대상의 발현된 DNA 서열, 및 선택적으로 하나 이상의 가이드 RNA를 포함하고, 이때 제1 단백질은 결정된 DNA 서열에 표적되고, 제1 단백질은 아미노산 링커 서열에 의해 제2 단백질에 연결되는, 뉴클레오타이드 벡터를 제공한다.

[0016] 본 명세서에는, a) 유전자 내 ATG 시작 코돈을 확인하는 단계; b) 제1항에 따른 융합 단백질을 이용하여 유전자의 ATG 시작 코돈 직후의 표적 서열에 결합하도록 융합 단백질 시스템을 설계하는 단계; c) 하나 이상의 연속적인 정지 코돈인 관심 대상의 DNA 서열을 설계하는 단계; 및 d) 세포 또는 유기체 내로 융합 단백질 및 관심 대상의 DNA 서열의 도입을 가능하게 하는 기법에 의해, 세포 또는 유기체에 융합 단백질 및 관심 대상의 DNA 서열을 제공하는 단계를 포함하고, 이때 관심 대상의 DNA 서열은 게놈성 DNA의 표적 서열에 통합되고, 유전자의 전사는 저해되는, 세포 또는 유기체에서 유전자 전사를 저해하는 방법을 제공한다. 일부 구현예에서, 제2 단백질은 재조합효소이고; 재조합효소는 Cre 재조합효소 또는 이의 변형된 형태이며, 이때 변형된 Cre 재조합효소는 구성적(constitutive) 재조합효소 활성을 갖는다. 일부 구현예에서, 벡터는 세포 내에서 발현될 역전사효소 유전자를 추가로 포함한다.

[0017] 본 명세서에는 또한, DNA 결합 단백질/인테그라아제 융합 및 길이 약 15 개 내지 약 100 개의 염기쌍의 RNA의 정제된 단백질을 포함하고, 이때 DNA 결합 단백질은 게놈 내에서 표적된 DNA 서열에 대해 공학적으로 조작된 카스9, Cpf1, TALEN 및 징크 핑거 단백질로부터 선택되고, 이때 인테그라아제는 HIV 인테그라아제, 렌티바이러스성 인테그라아제, 아데노바이러스성 인테그라아제, 레트로바이러스성 인테그라아제, 또는 MMTV 인테그라아제인, 조성물을 제공한다.

과제의 해결 수단

[0018] 하기의 상세한 기재는 당업자가 본 개시 내용을 실시하는 것을 돕기 위하여 제공된다. 그렇다 하더라도, 당업자는 본 발견의 사상 또는 범주로부터 벗어나지 않으면서 본 명세서에서 논의된 구현예에서의 변형 및 변경을 만들어 낼 수 있으므로, 이러한 상세한 기재가 본 개시 내용을 과도하게 제한하는 것으로 해석되어서는 안 된다.

- [0019] 본 개시 내용 및 첨부된 청구범위에서 사용된 바와 같이, 단수 형태("a", "an" 및 "the")는 그 문맥이 명백하게 그렇지 않은 것을 지시하지 않는다면, 복수의 지칭을 포함한다. 본 개시 내용 및 첨부된 청구범위에서 사용된 바와 같이, 용어 "또는"은 단수 또는 포괄적일 수 있다. 예를 들어, A 또는 B는 A 및 B일 수 있다.
- [0020] 내인성
- [0021] 본 명세서에서 기재된 바와 같은 내인성 핵산, 뉴클레오타이드, 폴리펩티드, 또는 단백질은 숙주 유기체에 관련하여 정의된다. 내인성 핵산, 뉴클레오타이드, 폴리펩티드, 또는 단백질은 숙주 유기체에서 자연발생적인 것이다.
- [0022] 외인성
- [0023] 본 명세서에서 기재된 바와 같은 외인성 핵산, 뉴클레오타이드, 폴리펩티드, 또는 단백질은 숙주 유기체에 관련하여 정의된다. 외인성 핵산, 뉴클레오타이드, 폴리펩티드, 또는 단백질은 숙주 유기체에서 자연발생적이지 않은 것이거나, 숙주 유기체에서 다른 장소에 있는 것이다.
- [0024] 녹아웃(Knockout)
- [0025] 유전자는 외인성 핵산이 숙주 유기체 내로 형질전환되어(예를 들어, 랜덤 삽입 또는 상동성 재조합), (예를 들어, 결실, 삽입에 의한) 유전자의 붕괴를 초래하는 경우, 녹아웃된 것으로 여겨진다.
- [0026] 유전자의 녹아웃시, 상응하는 단백질의 활성은 감소될 수 있다. 예를 들어, 유전자가 녹아웃되지 않은 동일한 단백질의 활성에 비교시, 10% 이상, 20% 이상, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 90% 이상, 또는 100%이다.
- [0027] 유전자 외의 녹아웃시, 녹아웃되지 않았던 유전자에 비하여, 유전자의 전사는 20% 이상, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 90% 이상, 또는 100% 감소될 수 있다.
- [0028] 변형된
- [0029] 변형된 유기체는 변형되지 않은 유기체와 상이한 유기체이다. 예를 들어, 변형된 유기체는 표적된 유전자 서열의 녹아웃을 결과로서 초래하는 개시 내용의 융합 단백질을 포함할 수 있다. 변형된 유기체는 변형된 계통을 가질 수 있다.
- [0030] 변형된 핵산 서열 또는 아미노산 서열은 변형되지 않은 핵산 서열 또는 아미노산 서열과 상이하다. 예를 들어, 핵산 서열은 삽입된, 결실된 또는 부가된 하나 이상의 핵산을 가질 수 있다. 예를 들어, 아미노산 서열은 삽입된, 결실된, 또는 부가된 하나 이상의 아미노산을 가질 수 있다.
- [0031] 작동적으로 연결된
- [0032] 일부 구현예에서, 벡터는 하나 이상의 제어 요소에 작동적으로 연결된 폴리뉴클레오타이드, 예컨대 프로모터 및/또는 전사 종결자를 포함한다. 핵산 서열은 또 다른 핵산 서열과 기능적 관계에 배치되는 경우 작동적으로 연결된 것이다. 예를 들어, 전구서열 또는 분비성 리더(secretory leader)에 대한 DNA는 폴리펩티드의 분비에 참여하는 전구단백질로서 발현되는 경우 폴리펩티드에 대한 DNA에 작동적으로 연결되거나; 프로모터는 서열의 전사에 영향을 미치는 경우 코딩 서열에 작동적으로 연결되거나; 번역을 용이하게 하기 위하여 배치되는 경우 리보솜 결합 위치는 코딩 서열에 작동적으로 연결된다. 작동적으로 연결된 서열은 근접성일 수 있으며, 분비 리더의 경우 근접성 및 판독 상(reading phase)에 있다.
- [0033] 숙주 세포 또는 숙주 유기체
- [0034] 숙주 세포는 본 개시 내용의 폴리펩티드를 암호화하는 폴리뉴클레오타이드를 함유할 수 있다. 일부 구현예에서, 숙주 세포는 다세포성 유기체의 일부이다. 다른 구현예에서, 숙주 세포는 단세포성 유기체로서 배양된다.
- [0035] 숙주 유기체는 임의의 적합한 숙주, 예를 들어 미생물을 포함할 수 있다. 본 명세서에 기재된 방법에 유용한 미생물은, 예를 들어 박테리아(예를 들어, 대장균(*E. coli*)), 효모(예를 들어, 사카로마이세스 세레비시에(*Saccharomyces cerevisiae*)), 및 식물을 포함한다. 유기체는 원핵성 또는 진핵성일 수 있다. 유기체는 단세포성 또는 다세포성일 수 있다.
- [0036] 숙주 세포는 원핵성일 수 있다. 적합한 원핵성 세포로는, 이에 제한되지는 않지만, 대장균, 락토바실러스(*Lactobacillus*) 종, 살모넬라(*Salmonella*) 종, 및 쉬겔라(*Shigella*) 종의 임의의 다양한 실험실 균주를 포함한다(예를 들어, 문헌 [Carrier et al. (1992) J. Immunol. 148:1176-1181]; 미국 특허 6,447,784; 및 문헌

[Sizemore et al. (1995) Science 270:299-302]에 기재된 바와 같음). 본 개시 내용에 사용될 수 있는 살모넬라 균주의 예로는, 이에 제한되지는 않지만, 살모넬라 티피(*Salmonella typhi*) 및 S. 티피무리움(*S. typhimurium*)을 포함한다. 적합한 쉬겔라 균주는, 이에 제한되지는 않지만, 쉬겔라 플렉스네리(*Shigella flexneri*), 쉬겔라 소나이(*Shigella sonnei*), 및 쉬겔라 디센테리애(*Shigella dysenteriae*)를 포함한다. 통상적으로, 실험실 균주는 비병원성인 것이다. 기타 다른 적합한 박테리아의 비제한적인 예로는, 이에 제한되지는 않지만, 슈도모나스 푸디타(*Pseudomonas putida*), 슈도모나스 아에루기노사(*Pseudomonas aeruginosa*), 슈도모나스 메발로니(*Pseudomonas mevalonii*), 로도박터 스파에로이데스(*Rhodobacter sphaeroides*), 로도박터 캡슐라투스(*Rhodobacter capsulatus*), 로도스피릴룸 루브룸(*Rhodospirillum rubrum*), 및 로도코커스 종(*Rhodococcus* sp)을 포함한다.

[0037] 일부 구현예에서, 숙주 유기체는 진핵성이다. 적합한 진핵성 숙주 세포는, 이에 제한되지는 않지만, 효모 세포, 곤충 세포, 식물 세포, 진균류 세포, 및 조류 세포를 포함한다.

[0038] 폴리뉴클레오티드 및 폴리펩티드[핵산 및 단백질]

[0039] 본 개시 내용의 단백질은 당업계에 알려진 임의의 방법에 의해 제조될 수 있다. 단백질은 고체 상 펩티드 합성 또는 액체 상 펩티드 합성으로서도 알려져 있는 고전적인 방식인 펩티드 합성에 의한 것 중 어느 하나를 이용하여 합성될 수 있다. 출발 주형으로서 Val-Pro-Pro, 에날라프릴(Enalapril) 및 리시노프릴(Lisinopril)을 이용하여, X-Pro-Pro, X-Ala-Pro, 및 X-Lys-Pro와 같은 몇 가지 일련의 펩티드 유사체(이때, X는 임의의 아미노산 잔기를 나타냄)는 고체 상 또는 액체 상 펩티드 합성을 이용하여 합성될 수 있다. 가용성 올리고머성 지지체에 커플링된 펩티드 및 올리고뉴클레오티드의 라이브러리의 액체 상 합성을 실시하기 위한 방법이 또한 기재되어 왔다. 문헌 [Bayer, Ernst and Mutter, Manfred, Nature 237:512-513 (1972)]; [Bayer, Ernst, et al., J. Am. Chem. Soc. 96:7333-7336 (1974)]; [Bonora, Gian Maria, et al., Nucleic Acids Res. 18:3155-3159 (1990)]. 액체 상 합성 방법이 반응 물질을 고체 상에 부착시키는 데 적합한 제1 반응 물질 상에 존재하는 구조물을 필요로 하지 않는다는 점에서, 액체 상 합성 방법은 고체 상 합성 방법에 비해 장점을 갖는다. 또한, 액체 상 합성 방법은 고체 상과 제1 반응물질(또는 중간체 생성물) 간의 결합을 분할할 수 있는 화학 조건을 회피하는 것을 필요로 하지 않는다. 추가적으로, 균질 용액 내 반응은, 고체 상 합성에서 존재하는 것과 같은 불균질 고체 상/액체 상 시스템에서 수득되는 것보다 더 양호한 수율 및 더 완벽한 반응을 제공할 수 있다.

[0040] 올리고머-지지된 액체 상 합성에서, 증가하는 생성물은 큰 용해성 중합체성 기에 부착된다. 합성의 각 단계로부터의 생성물은, 비교적 큰 중합체-부착된 생성물과 미반응된 반응물질간에 크기에서의 큰 차이를 기준으로, 미반응된 반응물질로부터 이후 분리될 수 있다. 이는 반응이 균질 용액에서 일어나는 것을 가능하게 하며, 전통적인 액체 상 합성과 연관된 지루한 정제 단계를 제거한다. 올리고머-지지된 액체 상 합성은 펩티드의 자동적 액체 상 합성에 또한 맞추어진다. 문헌 [Bayer, Ernst, et al., Peptides: Chemistry, Structure, Biology, 426-432].

[0041] 고체 상 펩티드 합성의 경우, 절차는 적절한 아미노산의 바람직한 서열의 펩티드 내로의 순차적인 조립을 수반하는 한편, 성장하는 펩티드의 말단은 불용성 지지체로 연결된다. 일반적으로, 펩티드의 카르복실 말단은 중합체에 연결되며, 이는 분할 시약을 이용한 처리시 중합체로부터 자유롭게 될 수 있다. 일반적인 방법에서, 아미노산은 수지 입자에 결합되며, 펩티드는 보호된 아미노산의 연속적인 첨가에 의해 단계적인 방식으로 생성되어 아미노산의 사슬을 생산한다. Merrifield에 의해 기재된 기법의 변형이 일반적으로 사용된다. 예를 들어, 문헌 [Merrifield, J. Am. Chem. Soc. 96: 2989-93 (1964)]을 참조한다. 자동화된 고체 상 방법에서, 펩티드는 카르복시-말단 아미노산을 유기 링커(예를 들어, PAM, 4-옥시메틸페닐아세트아미도메틸)를 부하함으로써 합성되며, 이는 디비닐 벤젠으로 가교결합된 불용성 폴리스티렌 수지에 공유결합적으로 부착된다. 말단 아민은 t-부틸옥시 카르보닐을 차단함으로써 보호될 수 있다. 히드록실기 및 카르복실기는 일반적으로 O-벤질 기로 차단함으로써 보호된다. 합성은 Applied Biosystems사(미국 캘리포니아주 포포스터 시티 소재)로부터 입수가 가능한 것과 같은 자동화된 펩티드 합성기 내에서 달성된다. 합성 후, 생성물이 수지로부터 제거될 수 있다. 차단 기는, 수렴된 방법에 따라, 플루오르화수소산 또는 트리플루오로메틸 술폰산을 이용함으로써 제거된다. 관례적인 합성은 0.5 밀리몰의 펩티드 수지를 생산할 수 있다. 분할 및 정제 후, 대략 60% 내지 70%의 수율로 통상적으로 생산된다. 생성물 펩티드의 정제는, 예를 들어 메틸-부틸 에테르와 같은 유기 용매로부터 펩티드를 결정화하고, 이후 증류수 내에 용해하고, 투석(대상 펩티드의 분자량이 약 500 달톤보다 큰 경우)을 이용하거나 또는 펩티드의 분자량이 500 달톤 미만인 경우 역상 고압 액체 크로마토그래피(예를 들어, 0.1% 트리플루오로아세트산 및 아세트오닐릴을 용매로서 이용하여 C¹⁸ 컬럼을 사용)를 이용함으로써 달성된다. 정제된 펩티드는 동결건조될 수 있고, 사용

시까지 건조한 상태로 저장될 수 있다. 결과로서 생성된 펩티드의 분석은 분석적 고압 액체 크로마토그래피 (HPLC) 및 전자분무 질량 분석법(ES-MS)의 일반 방법을 이용하여 달성될 수 있다.

[0042] 기타 다른 경우에, 단백질, 예를 들어 단백질은 재조합 방법에 의해 생산된다. 본 명세서에서 기재된 임의의 단백질의 생산을 위하여, 그러한 단백질을 암호화하는 폴리뉴클레오티드를 함유하는 발현 벡터로 형질전환된 숙주 세포가 사용될 수 있다. 숙주 세포는 보다 고등의 진핵 세포, 예컨대 포유동물 세포, 또는 보다 하등의 진핵 세포, 예컨대 효모일 수 있거나, 또는 숙주는 박테리아 세포와 같은 원핵 세포일 수 있다. 발현 벡터의 숙주 세포 내로의 도입은 인산칼슘 형질감염, DEAE-텍스트란 매개된 형질감염, 폴리브렌, 원형질체 융합, 리포솜, 핵 내로의 직접 미세주입법(microinjection), 스크레이프 부하(scrape loading), 바이오리스팅 (biolistic) 형질전환 및 전기천공법(electroporation)을 포함하는 다양한 방법에 의해 달성될 수 있다. 재조합 유기체로부터의 단백질의 큰 규모의 생산은 상업적 규모로 실시된 잘 수립된 공정이고, 당업자의 능력 내에서 충분하다.

[0043] 코돈 최적화

[0044] 암호화하는 폴리뉴클레오티드의 하나 이상의 코돈은 "편향(biased)" 또는 "최적화"되어 숙주 유기체의 코돈 선호도(codon usage)를 반영할 수 있다. 예를 들어, 암호화 폴리뉴클레오티드의 하나 이상의 코돈은 "편향" 또는 "최적화"되어 염록체 코돈 선호도 또는 핵 코돈 선호도를 반영할 수 있다. 대부분의 아미노산은 둘 이상의 상이한 (퇴화) 코돈에 의해 암호화되며, 각종 유기체가 다른 것들에 선호하는 특정 코돈을 이용한다는 것은 잘 인식되어 있다. "편향된" 또는 "최적화된" 코돈은 명세서 전체에 걸쳐 상호변경가능하게 사용될 수 있다. 코돈 편향은, 예를 들어 조류를 포함한 상이한 식물에서 담배에 비해 다양하게 편향될 수 있다. 일반적으로, 선택된 코돈 편향은, 본 개시 내용의 핵산으로 형질전환되고 있는 식물(또는 그 안에 있는 세포소기관)의 코돈 선호도를 반영한다.

[0045] 특정 코돈 선호도에 대해 편향된 폴리뉴클레오티드는 새로이 합성될 수 있거나, 또는 예를 들어 위치 지정 돌연변이 유발 방법에 의해, 관련적인 재조합 DNA 기법을 이용하여 유전적으로 변형되어 그들이 염록체 코돈 선호도에 대해 편향되도록 하나 이상의 코돈을 변화시킬 수 있다.

[0046] 서열 동일성 백분율

[0047] 핵산 또는 폴리펩티드 서열들 간의 서열 동일성 백분율 또는 서열 유사성을 결정하기에 적합한 알고리즘의 일 예는 BLAST 알고리즘으로, 이는 예를 들어 문헌 [Altschul et al., *J. Mol. Biol.* 215:403-410 (1990)]에 기재되어 있다. BLAST 분석을 수행하기 위한 소프트웨어는 미국 국립 생명공학 정보센터(National Center for Biotechnology Information)를 통해 공개적으로 입수가능하다. BLAST 알고리즘 파라미터 W, T, 및 X는 정렬(alignment)의 민감도 및 속도를 결정한다. BLASTN 프로그램(뉴클레오티드 서열의 경우)은 11의 단어 길이(W), 10의 예측(E), 100의 컷오프(cutoff), M=5, N=4, 및 두 가닥 모두의 비교를 기본값(default)으로서 사용한다. 아미노산 서열의 경우, BLASTP 프로그램은 3의 단어 길이(W), 10의 예측(E), 및 BLOSUM62 스코어링 매트릭스(scoring matrix)를 기본값으로서 사용한다(예를 들어, 문헌 [Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA*, 89:10915]에 기재된 바와 같음). 서열 동일성 퍼센트의 계산에 추가하여, BLAST 알고리즘은 또한 두 개의 서열들 간의 유사성의 통계적 분석을 수행할 수 있다(예를 들어, 문헌 [Karlin & Altschul, *Proc. Natl. Acad. Sci. USA*, 90:5873-5877 (1993)]에 기재된 바와 같음). BLAST 알고리즘에 의해 제공된 유사성의 일 측정은 최소 합계 확률(P(N))로, 이는 두 개의 뉴클레오티드 또는 아미노산 서열들 간의 매치가 우연히 일어날 확률의 표시를 제공한다. 예를 들어, 시험 핵산 대 참조 핵산의 비교에서 최소 합계 확률이 약 0.1 미만, 약 0.01 미만, 또는 약 0.001 미만인 경우, 핵산은 참조 서열에 대해 유사한 것으로 여겨진다.

[0048] 본 개시 내용은 A) 예를 들어, DNA 절단능에 대해 비활성인, 카스 단백질(예를 들어, 카스9)에 공유결합적으로 연결된 바이러스성 인테그라아제(또는 재조합효소)를 포함하는 시스템을 포함한다. 대안적으로, 바이러스성 인테그라아제(또는 박테리아성 또는 파지 재조합효소)는 TALE 단백질 또는 징크 핑거 단백질에 공유결합적으로 연결되며, 여기서 이들 단백질은 게놈 내 DNA의 특정 서열을 표적화하도록 설계된다.

[0049] 이는 발현 벡터 내에 또는 정제된 단백질로서 제공될 수 있다. B) 원하는 게놈 내로 혼입되는 상동염기서열 없이 또는 그를 갖는 관심 대상의 유전자(또는 관심 대상의 DNA 서열). GOI 또는 관심 대상의 DNA 서열은 필요에 따라 바이러스성 인테그라아제에 의해 인식되도록 변형될 수 있다. 예를 들어, 바이러스성 att 위치는 DNA 서열의 말단에 첨가될 수 있다. C) 폴리뉴클레오티드 형질감염 및/또는 단백질의 세포로의 도입에 필요한 기타 다른 시약.

[0050] 핵산

- [0051] 용어 "폴리뉴클레오티드", "뉴클레오티드", "뉴클레오티드 서열", "핵산" 및 "올리고뉴클레오티드"는 본 개시 내용에서 상호교환가능하게 사용된다. 이들은 데옥시리보뉴클레오티드 또는 리보뉴클레오티드 또는 이의 유사체 중 어느 하나의 임의의 길이의 뉴클레오티드의 중합체 형태를 지칭한다. 폴리뉴클레오티드는 임의의 3 차원의 구조물을 가질 수 있으며, 알려진 또는 미지의, 임의의 기능을 수행할 수 있다. 하기는 폴리뉴클레오티드의 비 제한적인 예들이다: 유전자 또는 유전자 단편의 코딩 또는 비-코딩 영역, 연결 분석으로부터 정의된 유전자좌들 (유전자좌), 엑손, 인트론, 메신저(messenger) RNA(mRNA), 운반 RNA, 리보솜 RNA, 짧은 간섭(short interfering) RNA(siRNA), 짧은-헤어핀 RNA(shRNA), 마이크로-RNA(miRNA), 리보자임, cDNA, 재조합 폴리뉴클레오티드, 분지된 폴리뉴클레오티드, 플라스미드, 벡터, 임의의 서열의 분리된 DNA, 임의의 서열의 분리된 RNA, 핵산 프로브, 및 프라이머. 폴리뉴클레오티드는 하나 이상의 변형된 뉴클레오티드, 예컨대 메틸화 뉴클레오티드 및 뉴클레오티드 유사체를 포함할 수 있다. 존재하는 경우, 뉴클레오티드 구조물에 대한 변형은 중합체의 조립 전 또는 후에 부여될 수 있다. 뉴클레오티드의 서열은 비뉴클레오티드 성분에 의해 중단될 수 있다. 폴리뉴클레오티드는 중합 후에, 예컨대 표지화 성분과의 접합에 의해 추가로 변형될 수 있다.
- [0052] 가이드 RNA
- [0053] 본 개시 내용의 양태에서, 용어 "키메라성 RNA", "키메라성 가이드 RNA", "가이드 RNA", "단일 가이드 RNA" 및 "합성 가이드 RNA"는 상호교환가능하게 사용되고, 가이드 서열, tracr 서열 및 tracr 메이트(mate) 서열을 포함하는 폴리뉴클레오티드 서열을 지칭한다. 용어 "가이드 서열"은 표적 위치를 특정하는 가이드 RNA 내의 약 20 bp(12 bp 내지 30 bp)를 지칭하며, 이는 용어 "가이드" 또는 "스페이스 (spacer)"와 상호교환가능하게 사용될 수 있다. 용어 "tracr 메이트 서열"은 또한 "직접 반복 서열(들)"과 상호교환가능하게 사용될 수 있다.
- [0054] 야생형
- [0055] 본 명세서에서 사용된 바와 같은 용어 "야생형"은 숙련자가 이해하는 기술 용어이며, 돌연변이 또는 변종 형태와 구분되는 자연 중에 발생하는 바와 같은, 유기체, 균주(strain), 유전자 또는 특징의 통상적인 형태이다.
- [0056] 변종
- [0057] 본 명세서에서 사용된 바와 같은 용어 "변종" 또는 "돌연변이"는 자연 발생하는 것에서 벗어나는 패턴을 갖는 성질의 표출을 의미하는 것으로 받아들여져야 한다. 유전자와 관련하여, 이들 용어는 특히 단일 뉴클레오티드 다형성(SNP), 삽입, 결실, 유전자 이동(shift)을 포함하는 야생형 유전자와 상이하게 하는 유전자에서의 다수의 변화를 나타낸다.
- [0058] 공학적으로 조작된
- [0059] 용어 "자연발생적이지 않은" 또는 "공학적으로 조작된"은 상호교환가능하게 사용되며, 인공 기술의 수반을 나타낸다. 이 용어가 핵산 분자 또는 폴리펩티드를 지칭하는 경우, 그 핵산 분자 또는 폴리펩티드가, 자연에서 자연적으로 회합되고, 자연에서 발견되는 바와 같은 하나 이상의 기타 다른 성분이 적어도 실질적으로 없음을 의미한다.
- [0060] 상보적
- [0061] "상보성"은, 핵산이 전통적인 왓슨-크릭 또는 기타 다른 전통적이지 않은 유형 중 어느 하나에 의해 또 다른 핵산 서열과 수소 결합(들)을 형성하는 핵산의 능력을 지칭한다. 상보성 퍼센트는 제2 핵산 서열과 함께 수소 결합을 형성할 수 있는(예를 들어, 왓슨-크릭 염기쌍) 핵산 분자 내 잔기의 백분율을 표시한다(예를 들어, 10 개 중 5 개, 6 개, 7 개, 8 개, 9 개, 10 개인 50%, 60%, 70%, 80%, 90%, 및 100% 상보성). "완전히 상보적"은 핵산 서열의 모든 연속적인 잔기가 제2 핵산 서열에서 동일한 수의 연속적인 잔기와 수소 결합할 것임을 의미한다. 본 명세서에서 사용된 바와 같은 "실질적으로 상보적"은, 적어도 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 또는 100%인 상보성 정도, 또는 8 개, 9 개, 10 개, 11 개, 12 개, 13 개, 14 개, 15 개, 16 개, 17 개, 18 개, 19 개, 20 개, 21 개, 22 개, 23 개, 24 개, 25 개, 30 개, 35 개, 40 개, 45 개, 50 개 이상의 뉴클레오티드의 영역에 걸친 백분율을 지칭하거나, 또는 엄격한 조건 하에서 혼성화하는 두 개의 핵산을 지칭한다.
- [0062] 아미노산
- [0063] 전체 명, 3 문자 코드, 1 문자 코드
- [0064] 아스파르트산 Asp D

[0065]	글루탐산 Glu E
[0066]	라이신 Lys K
[0067]	아르기닌 Arg R
[0068]	히스티딘 His H
[0069]	티로신 Tyr Y
[0070]	시스테인 Cys C
[0071]	아스파라긴 Asn N
[0072]	글루타민 Gln Q
[0073]	세린 Ser S
[0074]	트레오닌 Thr T
[0075]	글리신 Gly G
[0076]	알라닌 Ala A
[0077]	발린 Val V
[0078]	류신 Leu L
[0079]	이소류신 Ile I
[0080]	메티오닌 Met M
[0081]	프롤린 Pro P
[0082]	페닐알라닌 Phe F
[0083]	트립토판 Trp W
[0084]	본 명세서에서 사용된 바와 같은 표현 "아미노산"은 천연 및 합성 아미노산을 모두 포함하는 것을 의미한다. "표준 아미노산"은 자연발생적인 단백질/펩티드에서 일반적으로 발견되는 20 개의 표준 L-아미노산 중 임의의 것을 의미한다. "비-표준 아미노산 잔기"는, 합성적으로 제조되거나 천연 공급원으로부터 유도되는지의 여부에 관계없이 표준 아미노산 외의 임의의 아미노산을 의미한다. 본 명세서에서 사용된 바와 같이, "합성 아미노산"은 화학적으로 변형된 아미노산을 포괄하며, 염, 아미노산 유도체(예컨대, 아미드), 및 치환기를 포함하지만 이에 제한되지는 않는다. 본 개시 내용의 펩티드 내에 함유된, 그리고 특히 카르복시-말단 또는 아미노-말단에서 아미노산은 메틸화, 아미드화, 아세틸화, 또는 그의 활성화에 부정적인 영향을 미치지 않으면서 펩티드의 순환 반감기를 변화시킬 수 있는 기타 다른 화학 기로 치환함으로써 변형될 수 있다. 추가적으로, 이황화물 연결은 펩티드 내에 존재 또는 부재할 수 있다.
[0085]	아미노산은 측쇄 R을 기준으로 7 개의 군으로 분류될 수 있다: (1) 지방족 측쇄; (2) 히드록실(OH) 기를 함유하는 측쇄; (3) 황 원자를 함유하는 측쇄; (4) 산성 기 또는 아미드 기를 함유하는 측쇄; (5) 염기성 기를 함유하는 측쇄; (6) 방향족 고리를 함유하는 측쇄; 및 (7) 프롤린, 즉 측쇄가 아미노기와 융합된 이미노산.
[0086]	본 명세서에서 사용된 바와 같은, 용어 "보존성 아미노산 치환"은 본 명세서에서 하기 5 개 군 중 하나 내에서 교환되는 바와 같이 정의된다:
[0087]	I. 작은 지방족, 비극성 또는 약간 극성인 잔기:
[0088]	Ala, Ser, Thr, Pro, Gly;
[0089]	II. 극성, 음으로 하전된 잔기 및 그들의 아미드;
[0090]	Asp, Asn, Glu, Gin;
[0091]	III. 극성, 양으로 하전된 잔기:
[0092]	His, Arg, Lys;

- [0093] IV. 큰, 지방족, 비극성 잔기;
- [0094] Met, Leu, He, Val, Cys(Ile; 자동 수정은 읽을 수 없음)
- [0095] V. 큰, 방향족 잔기:
- [0096] Phe, Tyr, Tip(Trp, 유사)
- [0097] 본 개시 내용은, 달리 제공되지 않는 경우, 당업계의 기술에 속하는, 면역학, 생화학, 화학, 분자 생물학, 미생물학, 세포 생물학, 게놈학 및 재조합 DNA의 통상의 기법을 이용한다. 문헌 [Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989)]; [CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987))]; [시리즈 METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M. J. MacPherson, B. D. Hames and G. R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R. I. Freshney, ed. (1987))]을 참조한다.
- [0098] 벡터
- [0099] 유전자 발현 벡터(DNA-계 또는 바이러스성)는 세포 또는 조직에서 융합 인테그라아제를 발현하고, 또한 관심 대상의 DNA 서열(또는 유전자)을 인테그라아제 또는 재조합효소가 숙주 종 또는 세포의 게놈 내로 그 DNA(또는 유전자)를 통합시키는 데 필요한 적절한 위치에 제공하는 데 사용될 것이다. 다수의 유전자 발현 벡터가 당업계에서 알려져 있다. 벡터는 관심 대상의 유전자(또는 관심 대상의 DNA 서열)를 위해 이용될 것이다. 벡터는 당업계에서 알려진 다수의 제한 효소를 이용하여 절단될 수 있다.
- [0100] 크리스퍼/카스9
- [0101] 크리스퍼/카스9는 미국 특허 8697359, 미국 특허 8889356 및 문헌 [Ran et al, Nature Protocols, 2013, volume 8, pages 2281-2308]에 기재되어 있다. 카스9 단백질은, 게놈 내 DNA의 특정 서열에 결합하기 위하여 RNA 가이드를 이용한다. RNA 가이드(가이드 RNA)는 길이가 10 개 내지 40 개, 12 개 내지 35 개, 15 개 내지 30 개, 또는 예를 들어, 18 개 내지 22 개, 또는 20 개인 뉴클레오티드로 설계될 수 있다. 스트렙토코커스 파이오제네스(*Streptococcus pyogenes*)로부터의 카스9를 이용하는, 문헌 [Hsu et al, Nature Biotechnology, September 2013, volume 31, pages 827-832]을 참조한다. 또 다른 주요 카스9는 스태필로코커스 아우레우스(*Staphylococcus Aureus*)로부터의 것이다(S. 파이오제네스의 것보다 더욱 작은 카스9). 카스9 단백질은 DNA 서열의 특정 영역에 결합하기 위하여 가이드 RNA를 이용한다.
- [0102] 촉매적으로 비활성 형태의 카스9가 문헌 [Guilinger et al, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification, Nature Biotechnology, April 25, 2014, volume 32, pages 577-582]에 기재되어 있다. Guilinger와 다수는 촉매적으로 비활성인 카스9를 FokI 효소에 부착시켜 게놈성 DNA에서 절단부를 형성하는 데 있어서 더 큰 특이성을 달성한다. 이러한 촉매적으로 비활성인 카스9는, DNA를 절단할 수는 없는 반면, 카스9가 게놈성 DNA의 결합에 대한 RNA 가이드를 이용하는 것을 가능하게 한다.
- [0103] 카스9는 또한 그의 자연의 야생형 형태, 및 또한 세포 내에서 카스9 구조물의 더 양호한 발현을 위해 인간 최적화된 코돈 형태로 이용가능하다(문헌 [Mali et al, Science, 2013, volume 339, pages 823-826] 참조). 카스9의 코돈 최적화는 그의 발현을 위한 중에 따라 수행될 수 있다. 인테그라아제/카스9 융합 단백질의 단백질 형태(애비1로서도 알려짐) 또는 뉴클레오티드 발현 벡터 형태가 생산되느냐에 따라, 최적화된 또는 비-최적화된(야생형) 형태가 사용될 수 있다.
- [0104] 특정 DNA 서열을 향한 RNA 가이드는 각종 컴퓨터-기반 도구에 의해 설계될 수 있다.
- [0105] 크리스퍼/CPF1
- [0106] Cpf1은, 게놈성 DNA에서 특정 서열을 결합하기 위하여 가이드 RNA를 이용하는 또 다른 단백질이다. Cpf1은 또한 DNA를 절단하여 엇갈림(staggered) 절단을 만든다. Cpf1은 절단능에 대해 촉매적으로 비활성으로 될 수 있다.
- [0107] 기타 다른 크리스퍼 단백질
- [0108] 이들은, DNA를 절단하는 능력을 갖는지의 여부에 관계없이, 특정 DNA 서열을 표적하기 위하여 가이드 RNA를 이용하는 단백질이다. 이들 단백질들 중 일부는 기타 다른 효소/촉매 기능을 자연적으로 가질 수 있다.
- [0109] 탈렌(TALEN)

- [0110] 전사 활성화자-유사 이펙터 뉴클레아제(탈렌)는 TAL 이펙터 DNA 결합 도메인을 DNA 분할 도메인에 융합시킴으로써 생성된 제한 효소를 갖는 융합 단백질이다. 이들 시약은 효율적이고, 프로그래밍가능하며, 특이적 DNA 분할을 가능하게 하고, 동일계내(in situ) 게놈 편집을 위한 강력한 도구를 나타낸다. 전사 활성화자-유사 이펙터(TALE)는 실질적으로 임의의 DNA 서열을 결합하기 위하여 신속하게 공학적으로 조작될 수 있다. 본 명세서에서 사용된 바와 같은 용어 탈렌은 광범위하고, 또 다른 탈렌으로부터의 보조 없이 이중 가닥 DNA를 분할할 수 있는 단량체성 탈렌을 포함한다. 용어 탈렌은 또한 동일한 위치에서 DNA를 분할하는데 함께 작용하도록 공학적으로 조작된 탈렌들의 쌍의 하나 또는 두 멤버(member) 모두를 지칭하는 데 사용된다. 함께 작용하는 탈렌은, DNA의 손잡이 방향(handedness)을 지칭하는, 왼쪽-탈렌 및 오른쪽-탈렌으로서 지칭될 수 있다. 미국 특허 8,440,432를 참조한다.
- [0111] TAL 이펙터는 크산토모나스(Xanthomonas) 박테리아에 의해 분비된 단백질이다. DNA 결합 도메인은 12 번째 및 13 번째 아미노산을 제외하고 고도로 보존된 33 개 내지 34 개 아미노산 서열을 함유한다. 이들 두 장소는 매우 가변적이고(반복 가변성 2잔기(RVD: Repeat Variable Diresidues), 특정 뉴클레오티드 인식과의 강한 상관관계를 나타낸다. 아미노산 서열과 DNA 인식 간의 이러한 단순한 관계는 적절한 RVD를 함유하는 반복 절편들(segment)의 조합을 선택함으로써 특정 DNA 결합 도메인의 공학적 조작을 가능하게 한다.
- [0112] 인테그라아제 또는 재조합효소는 효모 또는 세포 분석에서 활성인 혼성 인테그라아제 또는 재조합효소를 구축하는 데 사용될 수 있다. 이들 시약은 또한 식물 세포 및 동물 세포에서 활성이다. 탈렌 연구는 야생형 FokI 분할 도메인을 사용하였지만, 일부 이후의 탈렌 연구는 또한 분할 특이성 및 분할 활성을 개선시키기 위해서 설계된 돌연변이를 갖는 FokI 분할 도메인 변종을 사용하였다. 탈렌 DNA 결합 도메인과 인테그라아제 또는 재조합효소 도메인 간의 아미노산 잔기의 수와, 두 개의 개별적인 탈렌 결합 위치 간의 염기 수는 모두 고도의 활성을 달성하기 위한 파라미터이다. 탈렌 DNA 결합 도메인과 인테그라아제 또는 재조합효소 도메인간의 아미노산 잔기의 수는, 복수의 TAL 이펙터 반복 서열과 인테그라아제 또는 재조합효소 도메인 사이에 스페이서(스페이서 서열과 다름)의 도입에 의해 변형될 수 있다. 스페이서 서열은 6 개 내지 102 개 또는 9 개 내지 30 개의 뉴클레오티드 또는 15 개 내지 21 개의 뉴클레오티드일 수 있다. 이들 스페이서는 DNA 표적 단백질(카스9, TALE 또는 징크 핑거 단백질)과 인테그라아제 또는 재조합효소간에 연결을 제공하는 것 외에 기타 다른 활성을 혼성 단백질에 일반적으로 제공하지 않을 것이다. 스페이서에 대한 아미노산 및 본 개시 내용에서의 기타 다른 용도를 위한 아미노산이 존재한다.
- [0113] 아미노산 서열 및 탈렌 결합 도메인의 DNA 인식 간의 관계는 설계가능한 단백질을 가능하게 한다. 이 경우, TALE 결합 도메인에서 발견되는 반복적인 서열의 부적절한 어닐링으로 인해 인공 유전자 합성은 문제가 있다. 이에 대한 일 해결책은 올리고뉴클레오티드 조립에 뒤이어 전체 유전자 증폭이라는 2 단계 PCR에서의 조립에 적합한 올리고뉴클레오티드를 찾기 위해 공개적으로 입수가능한 소프트웨어 프로그램인 DNAWorks를 이용하는 것이다. 공학적으로 조작된 TALE 구조물을 생성하기 위한 다수의 모듈 조립 방법 또한 당업계에 보고되어 있다.
- [0114] 일단 탈렌 유전자가 함께 조립되면, 이들은 플라스미드 내로 삽입되고; 플라스미드는 이어서 표적 세포를 형질 감염시키는 데 사용되고, 여기서 유전자 산물이 발현되고 핵을 도입하여 게놈에 접근한다. 탈렌은, 세포들이 반응하는, 이중-가닥 파괴(DSB)를 유도함으로써 DNA 수복으로 게놈을 편집하는 데 사용되지만, 본 개시 내용은 바이러스성 인테그라아제 또는 박테리아성 또는 과거 재조합효소의 힘을 이용하여 관심 대상의 DNA 서열을 게놈 내 표적된 위치 내로 삽입하고자 한다. WO 2014134412 및 미국 특허 8748134의 개시 내용을 참조한다.
- [0115] 징크 핑거 단백질
- [0116] DNA 결합을 위한 징크 핑거 단백질 및 그의 설계는 미국 7928195, 미국 2009/0111188, 및 미국 7951925에 기재되어 있다. 징크 핑거 단백질은 다수의 연결된 징크 핑거 도메인을 명시된 순서로 DNA의 특정 서열에 결합한다. 징크 핑거 단백질 엔도뉴클레아제는 잘 확립되어 있다.
- [0117] 징크 핑거 단백질(ZFP)은 서열 특이적 방식으로 DNA에 결합할 수 있는 단백질이다. 징크 핑거들은 아프리카 발톱 두꺼비(제노푸스 래비스(Xenopus laevis))의 난자로부터의 전사 인자인 TFIIIA에서 최초로 확인되었다. 이러한 ZFP 부류의 단일 징크 핑거 도메인은 길이가 약 30 개 아미노산이고, 몇몇 구조 연구로부터 이는, 두 개의 시스테인과 두 개의 히스티딘에 의한 아연 원자의 배위를 통하여 특정 입체 배좌로 유지되는, 베타 회전(두 개의 보존된 시스테인 잔기를 함유) 및 알파 나선(두 개의 보존된 히스티딘 잔기를 함유)을 함유한다는 것이 증명되었다. ZFP의 이러한 부류는 또한 C2H2 ZFP로도 알려져 있다. ZFP의 추가의 부류가 또한 제안되어 왔다. 예를 들어, Cys-Cys-His-Cys (C3H) ZFP의 논의에 대한 문헌 [Jiang et al. (1996) J. Biol. Chem. 271:10723-10730]을 참조한다. 이제까지, 10,000 개가 넘는 징크 핑거 서열이 알려지거나 추정적인 수천 개의 전사 인자들

내에서 확인되어 왔다. 징크 핑거 도메인은 DNA 인식뿐만 아니라, RNA 결합 및 단백질-단백질 결합에도 연루된다. 이러한 부류의 분자는 모든 인간 유전자들 중 약 2%를 구성할 것으로 현재 추산되어 있다.

[0118] 많은 징크 핑거 단백질은, 각 핑거 도메인에서 단일 아연 원자를 정사면체-배위하는 시스테인 및 히스티딘 잔기를 보존하였다. 특히, 대부분의 ZFP는 일반 서열: -Cys-(X)₂₋₄-Cys-(X)₁₂-His-(X)₃₋₅-His-(서열번호 49, 여기서 X는 임의의 아미노산(C₂H₂ ZFPs)을 나타냄)의 핑거 성분에 의해 특징지어진다. 이렇게 가장 광범위하게 표시되는 부류의 아연-배위 서열은 특정 간격(spacing)을 갖는 두 개의 시스테인 및 두 개의 히스티딘을 함유한다. 각각의 핑거의 접힌 구조는 역평행 β-회전, 핑거 팁 영역 및 짧은 양친매성 α-나선을 함유한다. 금속 배위 리간드는 아연 이온에 결합하고, zif268-형 징크 핑거의 경우에서, 짧은 양친매성 α-나선은 DNA의 주요 그루브(groove)에 결합한다. 추가적으로, 징크 핑거의 구조는 특징의 보존된 소수성 아미노산 잔기에 의해(예를 들어, 제1 보존된 Cys에 바로 선행하는 잔기 및 핑거의 나선 절편의 +4 위치에서의 잔기) 그리고 보존된 시스테인 및 히스티딘 잔기를 통한 아연 배위에 의해 안정화된다.

[0119] 게놈성 DNA에서 특정 표적 서열에 결합할 수 있는 기타 다른 DNA 결합 단백질

[0120] 이런 단백질은, 각종 유기체의 게놈성 DNA의 특정 서열에 결합할 수 있는 징크 핑거 단백질, 탈렌 및 크리스퍼 단백질에 관련되지 않은 것들을 포함한다. 이들은 전사 인자, 전사 리프레서(repressor), 메가뉴클레아제, 엔도뉴클레아제 DNA 결합 도메인 및 기타를 포함할 수 있다.

[0121] 인테그라아제

[0122] 인테그라아제 및 이의 엔도뉴클레아제 융합 단백질은 미국 2009/0011509에 기재되어 있다. 도입된 인테그라아제는 렌티바이러스성 인테그라아제 및 HIV1(인간 면역결핍 바이러스 1) 인테그라아제이다. 본 개시 내용은 촉매적으로 비활성인(또는 활성인) 카스9, TALE 또는 징크 핑거 단백질을 인테그라아제에 융합시켜서, 그 인테그라아제를 사용자에게 의해 선택된 게놈에서 DNA의 특정 영역에 표적한다.

[0123] 기타 다른 레트로바이러스성 인테그라아제와 같이, HIV-1 인테그라아제는 긴 말단 반복서열(LTR)의 U3 및 U5 영역에 위치한 바이러스성 DNA의 말단에서 특별한 특징부를 인식할 수 있다((Brown, 1997). LTR 말단은 레트로바이러스의 통합 기계에 의한 인식을 위해 cis에서 요구되는 것으로 생각되는 유일한 바이러스 서열이다. 짧은 불완전한 전환된 반복서열은 쥐과 및 조류 레트로바이러스 모두에서의 LTR의 외측 가장자리에 존재한다(문헌 [Reicin et al., 1995]에 의해 검토됨). 레트로바이러스성 DNA 말단에서 최외측 위치 3 및 4에 위치된 부말단 CA(3' 말단 가공된 뉴클레오티드가 있는 1 및 2 위치)와 함께, 이들 서열은 시험관 내 및 생체 내 프로바이러스성(proviral) 통합을 결정하는 데 필요 충분하다. CA 디뉴클레오티드에 대한 내부 서열은 최적 인테그라아제 활성에 중요한 것으로 보인다(Brin & Leis, 2002a; Brin & Leis, 2002b; Brown, 1997). HIV-1 LTR의 말단 15 bp는 정확한 3' 말단 가공 및 시험관 내 가닥 전달 반응에 매우 중요한 것으로 나타났다(Reicin et al., 1995; Brown, 1997). HIV-1 IN에 의해, 더 긴 기질이 짧은 기질보다 더욱 효율적으로 사용되며, 이는 결합 상호작용이 바이러스성 DNA 말단으로부터 내측으로 적어도 14 bp 내지 21 bp를 연장함을 나타낸다. 문헌 [Brin and Leis (2002a)]은 HIV-1 LTR의 특이적 특징들을 분석하였으며, U5 LTR이 시험관 내 IN 가공에 더 효율적인 기질이라 하더라도(Bushman & Craigie, 1991; Sherman et al., 1992), U3 및 U5 LTR 인식 서열 둘 모두가 IN-촉진된 공동의 DNA 통합에 요구된다고 결론지었다. IN 인식 서열의 위치 17 내지 20은 공동의 DNA 통합 메커니즘에 필요하지만, HIV-1 IN은 불변성 부말단 CA 디뉴클레오티드로부터 연장하는 U3 및 U5 말단 모두에서 상당한 변화를 감내한다(Brin & Leis, 2002b). 본 개시 내용은, 게놈 내로 통합되는 관심 대상의 DNA 서열 또는 유전자를 수용하기 위한 장소인 5' 및 3' 말단에서 바이러스성(레트로바이러스성 또는 HIV) LTR 영역을 함유하는 DNA 벡터를 포함한다. LTR 영역은, 이들이 적절한 통합을 위해 인테그라아제와 상호작용하는 기능을 하는 한, 전장 LTR이어야 할 필요는 없다. LTR 영역은 검출가능성이거나(예를 들어, 형광), PCR 검출, 또는 선택성 마커(예를 들어, 항생제 내성)를 함유하도록 변형될 수 있다. 벡터는, LTR 영역이 (제한 엔도뉴클레아제에 대해 설계된 제한 위치를 통해) DNA 단편의 5' 및 3' 말단에 존재하도록, 절단되고 선형화되도록 설계된다.

[0124] 인테그라아제는 유연성 링커에 의해 연결된 3 개의 도메인으로 이루어진다. 이들 도메인은 N-말단 HHH-CC 아연-결합 도메인, 촉매성 코어 도메인 및 C-말단 DNA 결합 도메인이다(Lodi et al, Biochemistry, 1995, volume 34, pages 9826-9833). 본 개시 내용의 일부 양태에서, 카스9(또는 기타 다른 DNA 결합 분자)에 결합된 인테그라아제는 C-말단 결합 도메인을 갖지 않을 것이다. 본 개시 내용의 일 양태에서, 두 개의 상이한 융합 단백질이 생산될 것인데, 여기서 하나는 인테그라아제의 N-말단 아연 결합 도메인과 융합된 촉매적으로 비활성인 카스9(또는 TALE 또는 징크 핑거 단백질)이고, 다른 하나는 인테그라아제의 촉매적 코어 도메인과 융합된 촉매적으로 비활성인 카스9(또는 TALE 또는 징크 핑거 단백질)를 갖는다. 두 개의 상이한 융합 단백질은 TALE-FokI 또는 징

크 핑거-FokI 시스템으로 알려진 바와 같이 게놈성 DNA의 반대 가닥에 결합하도록 설계될 것이다. 이러한 방식으로, N-말단 도메인 및 촉매적 코어가 게놈성 DNA 상의 위치에서 접촉하는 경우, 이는 인테그라아제 활성을 나타낼 것이다. 인테그라아제의 완전한 활성은 또한 인테그라아제의 4량체를 연루하는 것으로 관찰되어 왔기 때문에, 융합 단백질은 길이가 1 개 내지 20 개의 아미노산 또는 길이가 4 개 내지 12 개의 아미노산일 수 있는 유연성 링커에 의해 연결된 1 개, 2 개, 3 개, 4 개 인테그라아제 단백질을 이용하여 설계될 수 있다.

[0125] 재조합효소

[0126] Cre, FLP, R, Dre, Kw, 및 Gin 재조합효소를 포함하는 재조합효소는 미국 8816153 및 미국 2004/0003420에 기재되어 있다. Cre 재조합효소와 같은 재조합효소는 게놈으로부터의 서열을 잘라내기 위하여 LoxP 위치를 사용한다. 재조합효소는 그의 재조합 활성에 대하여 구성적으로 활성화되고 덜 위치특이적이 되도록 변형될 수 있다. 이에 따라, 서열 특이성을 갖지 않는 그러한 구성적으로 활성화된 재조합효소 단백질을, 본 개시 내용의 융합 단백질 내로 포함시킴으로써 게놈 내 DNA의 특정 서열에 대해 표적하는 것이 가능하다. 이러한 방식으로, 크리스퍼/카스9, TALE 또는 징크 핑거 단백질 도메인은, 재조합효소가 그의 재조합 활성에 기여할 DNA 서열을 특정한다. 그러한 재조합효소 단백질은 야생형으로, 재조합효소 활성에 대하여 구성적으로 활성이거나 활성이 아닐 수 있다. 카스9-Gin 또는 카스9-Cre와 같은 카스9-재조합효소는 링커 서열의 이용 또는 직접 융합에 의해 생산될 수 있다.

[0127] 융합 단백질에 대한 핵 국소화 신호 서열(NLS: Nuclear Localization Signal Sequence)

[0128] 신호 펩티드 도메인(또한 "NLS"로도 지칭됨)은, 예를 들어 효모 GAL4, SKI3, L29 또는 히스톤 H2B 단백질, 폴리오마 바이러스 거대 T 단백질, VP1 또는 VP2 캡시드 단백질, SV40 VP1 또는 VP2 캡시드 단백질, 아데노바이러스 E1a 또는 DBP 단백질, 인플루엔자 바이러스 NS1 단백질, 간염 바이러스 코어 항원 또는 포유동물 라민(lamin), c-myc, max, c-myb, p53, c-erbA, jun, Tax, 스테로이드 수용체 또는 Mx 단백질(문헌 [Boulikas, Crit. Rev. Eucar. Gene Expression, 3, 193-227 (1993)] 참조), 유인원(simian) 바이러스 40("SV40") T-항원(Kalderon et. al, Cell, 39, 499-509 (1984)) 또는 알려진 핵 국소화를 갖는 기타 다른 단백질로부터 유래된다. NLS는, 예를 들어 SV40 T-항원으로부터 유래되며, 당업계에서 알려진 기타 다른 NLS 서열일 수 있다. 탠덤 NLS 서열이 사용될 수 있다.

[0129] 링커 영역

[0130] 합성되는 융합 단백질/펩티드 사이에 사용된 다양한 링커는 아미노산으로 구성될 것이다. DNA 수준에서, 이들은 유전자 코드에서 알려진 바와 같은 3 개의 염기쌍(bp) 코돈으로 나타내어진다. 링커는 길이가 1 개 내지 1000 개 및 그 사이의 임의의 정수 개의 아미노산일 수 있다. 예를 들어, 링커는 길이가 1 개 내지 200 개의 아미노산이거나, 링커는 길이가 1 개 내지 20 개의 아미노산이다.

[0131] 발현 벡터

[0132] 많은 핵산이 세포 내로 도입되어 유전자의 발현을 이끌 수 있다. 본 명세서에서 사용된 바와 같이, 용어 핵산은 DNA, RNA, 및 핵산 유사체를 포함하고, 핵산은 이중가닥 또는 단일가닥(즉, 센스 또는 안티센스 단일가닥)이다. 핵산 유사체는 염기 모이어티(moiety), 당 모이어티, 또는 포스페이트 골격에서 변형되어, 예를 들어 핵산의 안정성, 혼성화 또는 용해도를 개선할 수 있다. 염기 모이어티에서의 변형은 테옥시티미딘에 대해 테옥시우리딘, 및 테옥시시티딘에 대해 5-메틸-2'-테옥시시티딘 및 5-브로모-2'-독시시티딘을 포함한다. 당 모이어티의 변형은 2'-O-메틸 또는 2'-O-알릴 당을 형성하는 리보오스 당의 2'히드록실의 변형을 포함한다. 테옥시 리보오스 포스페이트 골격은 모르폴리노 핵산을 생산하도록 변형될 수 있으며, 여기서 각각의 염기 모이어티는 6-원 모르폴리노 고리, 또는 펩티드 핵산에 연결될 수 있는데, 여기서 테옥시포스페이트 골격은 슈도펩티드 골격으로 대체되고, 4 개의 염기는 보유된다. 문헌 [Summerton and Weller (1997) Antisense Nucleic Acid Drug Dev. 7(3): 187]; 및 [Hyrup et al. (1996) Bioorgan. Med. Chem. 4:5]을 참조한다. 추가적으로, 테옥시포스페이트 골격은, 예를 들어 포스포로티오에이트 또는 포스포로디티오에이트 골격, 포스포로아미다이트, 또는 알킬 포스포로트리에스테르 골격으로 대체될 수 있다. 핵산 서열은 프로모터와 같은 조절 영역에 작동적으로 연결될 수 있다. 조절 영역은 임의의 화학종으로부터의 것일 수 있다. 본 명세서에서 사용된 바와 같은, '작동적으로 연결된'은 표적 핵산의 전사를 허용 또는 촉진시키도록 하는 방식으로 핵산 서열에 대한 조절 영역의 배치(positioning)를 지칭한다. 임의의 형태의 프로모터가 핵산 서열에 대해 작동적으로 연결될 수 있다. 프로모터의 예로는, 이에 제한되지 않지만, 조직-특이적 프로모터, 구성 프로모터, 및 특정 자극에 대해 반응성 또는 비반응성인 프로모터(예를 들어, 유도성 프로모터)를 포함한다.

- [0133] 핵산 구조물에서 유용할 수 있는 추가의 영역으로는, 이에 제한되지 않지만, 폴리아데닐화 서열, 번역 제어 서열(예를 들어, 내부 리보솜 도입 절편, IRES), 인핸서(enhancer), 유도성 요소, 또는 인트론을 포함한다. 그러한 조절 영역은, 전사, mRNA의 안정성, 번역 효율 등에 영향을 미침으로써 발현을 증가시킬 수 있지만, 필수적이지는 않을 수 있다. 그러한 조절 영역은 세포(들) 내 핵산의 최적 발현을 수득하고자 함에 따라 핵산 구조물 내에 포함될 수 있다. 때때로 그러한 추가 요소 없이 충분한 발현이 수득될 수 있다.
- [0134] 핵산 구조물이 신호 펩티드 또는 선택성 마커를 암호화하는 데 사용될 수 있다. 신호화(마커) 펩티드는 암호화된 폴리펩티드가 특정 세포 장소(예를 들어, 세포 표면)를 지향하도록 이용될 수 있다. 그러한 선택성 마커의 비제한적인 예는 푸로마이신(puromycin), 간시클로버(ganciclovir), 아데노신 데아미나아제(ADA), 아미노글리코시드 인산기 전이효소(neo, G418, APH), 디하이드로폴레이트 환원효소(DHFR), 하이그로마이신-B-인산전달효소, 티미딘 키나아제(TK), 및 크산틴-구아닌 포스포리보실기 전이효소(XGPRT)를 포함한다. 이들 마커는 배양에서 안정적인 형질전환체를 선발하는 데 유용하다. 기타 다른 선택성 마커로는 형광 폴리펩티드, 예컨대 녹색 형광 단백질, 적색 형광 단백질, 또는 황색 형광 단백질을 포함한다.
- [0135] 핵산 구조물은, 당업계에서 알려진 각종 생물학적 기법을 이용하여 임의의 유형의 세포 내로 도입될 수 있다. 이들 기법의 비제한적인 예는, 트랜스포존(transposon) 시스템, 세포를 감염시킬 수 있는 재조합 바이러스, 또는 리포솜의 이용, 또는 핵산을 세포로 전달할 수 있는, 전기천공법, 미세주입법, 또는 인산칼슘 침전과 같은 기타 다른 비-바이러스 방법을 포함한다. 뉴클레오펙션(Nucleofection) TM으로 명명되는 시스템이 또한 사용될 수 있다.
- [0136] 핵산은 벡터 내로 혼입될 수 있다. 벡터는 운반체로부터 표적 DNA 내로 이동하도록 설계된 임의의 특정 DNA 절편을 포함하는 광범위한 용어이다. 벡터는 발현 벡터 또는 벡터 시스템으로 지칭될 수 있으며, 이는 계놈 또는 기타 다른 표적된 DNA 서열, 예컨대 에피솜, 플라스미드, 또는 심지어는 바이러스/파지 DNA 절편 내로 DNA 삽입을 일으키는 데 요구되는 일단의 성분들이다. 벡터는 거의 종종, 하나 이상의 발현 제어 서열을 포함하는 하나 이상의 발현 카세트를 함유하며, 이때 발현 제어 서열은 또 다른 DNA 서열 또는 mRNA 각각의 전사 및/또는 번역을 제어 및 조절하는 DNA 서열이다.
- [0137] 벡터의 많은 상이한 유형이 당업계에 알려져 있다. 예를 들어, 플라스미드 및 레트로바이러스 벡터를 비롯한 바이러스성 벡터가 알려져 있다. 포유동물 발현 플라스미드는 통상적으로 복제 기원, 적합한 프로모터 및 선택적인 인핸서를 갖고, 또한 임의의 필요한 리보솜 결합 위치, 폴리아데닐화 위치, 스플라이스 도너 및 억셉터 위치, 전사 종결 서열, 및 5' 플랭킹(flanking) 비-전사된 서열을 갖는다. 그러한 벡터로는 플라스미드(이는 또 다른 유형의 벡터의 운반체일 수 있음), 아데노바이러스, 아데노-연관 바이러스(AAV), 렌티바이러스(예를 들어, 변형된 HIV-1, SIV 또는 FIV), 레트로바이러스(예를 들어, ASV, ALV 또는 MoMLV), 및 트랜스포존(P-요소, Tol-2, 프로그 프린스(Frog Prince), 피기백(piggyBac) 또는 기타)이 포함된다.
- [0138] 본 개시 내용에서의 사용을 위한 박테리아 및 바이러스성 유전자 및 단백질은 아래 "본 개시 내용의 서열"이라는 제목의 섹션에서 열거된다. 기타 다른 바이러스성 인테그라아제, 예를 들어 마우스 유방 종양 바이러스(MMTV)로부터의 것들 및 아데노바이러스가 또한 본 명세서에서 개시된 방법 및 조성물 중에 사용될 수 있다.
- [0139] 편집된 세포들의 혼합된 개체군은 유전자 편집을 받은 세포들 및 받지 않은 세포들의 혼합물로 여겨진다.
- [0140] 예시적인 예비1 시험관 내 분석
- [0141] 1) 예비1 단백질을 가이드 RNA와 함께 인큐베이션시킨다;
- [0142] 2) 예비1/가이드 RNA를 부분적인 LTR을 갖는 도너 DNA와 함께 인큐베이션하여 전-개시(pre-initiation) 복합체를 형성한다;
- [0143] 3) 전-개시 복합체를 편집될 유전자(예를 들어, CXCR4)를 함유한 플라스미드와 인큐베이션시킨다; 및
- [0144] 4) 도너 DNA 통합을 위한 PCR 및 DNA 서열분석 확인.
- [0145] 카스9프로토콜은, 예를 들어 문헌 [Gagnon et al., 2014, http://labs.mcb.harvard.edu/schier/VertEmbryo/cas9_Protocols.pdf]에 기재되어 있다.
- [0146] 인테그라아제 활성화에 대한 분석은, 예를 들어 문헌 [Merkel et al., Methods, 2009, volume 47, pages 243-248]에 기재되어 있다.

도면의 간단한 설명

[0147]

본 개시 내용의 이들 및 기타 다른 특징, 양태, 및 장점은 하기 기재, 첨부된 청구범위 및 수반된 도면과 관련하여 더 잘 이해될 것이다.

도 1은 a) 예시적인 촉매적으로 비활성인 카스9/HIV1 인테그라아제 융합 단백질, b) 예시적인 TALE/HIV1 인테그라아제 융합 단백질, c) 예시적인 징크 핑거 단백질/HIV1 인테그라아제 융합 단백질, 및 d) 표적된 위치에서 DNA의 반대 면에 대해 설계된 예시적인 카스9/HIV1 인테그라아제 융합 단백질을 나타낸다. 융합 단백질 각각은 DNA의 특이적 표적 서열에 결합한다. "ZnFn"은 징크 핑거 단백질이다. "인테그라아제"는, 하나의 인테그라아제 단위 또는 예를 들어 짧은 아미노산 링커에 의해 연결된 두 개의 인테그라아제 단위를 나타낸다. 일부 구현예에서, 인테그라아제는 재조합효소에 의해 대체될 수 있다. 카스9는 촉매적으로 활성 또는 비활성일 수 있다.

도 2는, 촉매적으로 비활성인 카스9/인테그라아제 융합 단백질을 포함하는 벡터, 관심 대상의 DNA 서열을 포함하는 벡터, 및 역전사효소를 포함하는 벡터를 포함하는 DNA 플라스미드 시스템을 나타낸다. 가이드 RNA(gRNA) 또는 RNA는 별도로 제공될 수 있다. 또 다른 벡터는 gRNA를 발현하는 데 사용될 수 있다. "1 또는 2"는 하나의 인테그라아제 또는 예를 들어 아미노산 링커에 의해 연결된 두 개의 인테그라아제를 지칭한다.

도 3은 뉴클레오타이드 서열 촉매적으로 비활성인 카스9/인테그라아제 융합 단백질, 가이드 RNA, 관심 대상의 DNA(유전자) 서열, 및 역전사효소를 포함하는 예시적인 DNA 플라스미드를 나타낸다. 바이러스성 att 위치는 관심 대상의 DNA 서열에 제공되어, 인테그라아제의 세포의 게놈성 DNA 내로의 혼입을 가능하게 할 수 있다. 가이드 RNA(gRNA) 또는 RNA는 별도로 제공될 수 있다. 또 다른 벡터가 gRNA를 발현하는 데 사용될 수 있다. "1 또는 2"는 하나의 인테그라아제 또는 예를 들어, 아미노산 링커에 의해 연결된 두 개의 인테그라아제를 지칭한다.

도 4는 흐름도를 나타낸다. 도 2 및 도 3에 나타난 벡터를 이용하는 일 예시적인 방법을 도 4에 나타내며, 이는 다음과 같다: 1) 역전사효소는 벡터로부터 발현된 att 위치를 갖는 관심 대상의 DNA 서열을 전사한다(대안적으로 att 위치를 갖는 선형 DNA가 사용됨), 2) 융합 카스9/인테그라아제는 가이드 RNA를 기준으로 게놈성 DNA 상의 위치를 표적한다, 3) 인테그라아제는 관심 대상의 DNA 서열 상의 att(LTR) 위치를 인식하고, 그 DNA를 표적된 위치에서 게놈 내로 통합시키고, 4) 관심 대상의 DNA 서열의 적절한 삽입에 대해 확인하기 위하여, 분석(예를 들어, PCR(중합효소 연쇄 반응))을 수행한다. 분석은 비특이적 통합에 대해 확인하기 위하여 수행될 수 있다.

도 5는 가이드 NrF2-sgRNA2 및 sgRNA3을 이용한 NrF2의 애비1(Abbie1) 유전자 편집 표적화 엑손 2를 나타낸다.

도 6은 애비1 유전자 편집에 의해 생성된 이론적 데이터를 나타낸다.

도 7은 가이드 NrF2-sgRNA 3을 이용하여 NrF2의 애비1 유전자 편집 표적화 엑손 2를 나타낸다.

도 8은 혼합된(pooled) Hek293T 세포 내 NrF2의 애비1 녹아웃을 나타낸다.

도 9는 혼합된 Hek293T 세포 내 NrF2의 애비1 녹아웃을 나타낸다.

도 10은 애비1 유전자 편집 표적화 CXCR4 엑손 2를 나타낸다.

도 11은 대장균의 분리 및 정제 후 애비1 단백질의 검출을 나타낸다. 쿠마씨 (Coomassie) 염색된 겔.

발명을 실시하기 위한 구체적인 내용

[0148]

실시예

[0149]

하기 실시예는 본 개시 내용의 출원의 예시를 제공하고자 하는 것이다. 하기 실시예는 본 개시 내용의 범주를 완전히 정의하거나, 그렇지 않으면 그를 한정하고자 하는 것은 아니다. 당업자는 당업계에서 알려진 많은 기타 다른 방법들이 본 명세서에 구체적으로 기재되거나 참조된 예들을 대신하여 치환될 수 있음을 이해할 것이다.

[0150]

실시예 1: 카스9-인테그라아제 융합 단백질의 발현을 위한 DNA 벡터

[0151]

촉매적으로 비활성인 카스9의 DNA 서열은 12 bp, 15 bp, 18 bp, 21 bp, 24 bp, 27 bp 또는 30 bp 스페이서(카스9와 인테그라아제 간의 링커로서 4 개, 5 개, 6 개, 7 개, 8 개, 9 개 또는 10 개의 아미노산에 대한 코드) 및 HIV1 인테그라아제를 갖는 발현 벡터 내로 혼입된다. 기타 다른 실험에서, 인테그라아제보다는 박테리아 또는 파지 기원의 재조합효소가 사용된다. 이는, 이들을 임의의 기타 다른 위치의 DNA에서 재조합하는 것을 가능하게 하는 돌연변이를 갖거나 갖지 않는 Hin 재조합효소(서열번호 25) 및 Cre 재조합효소(서열번호 26)를 포함

한다. 융합 단백질을 분리하기 위하여 His 또는 cMyc 태그(또는 단백질 정제에 유용한 기타 다른 서열)가 포함될 수 있다. 발현 벡터는 벡터가 제공될 세포에서 활성화된 프로모터를 사용한다. CMV(거대세포바이러스 프로모터)는 포유동물 세포를 위한 발현 벡터에 흔히 사용된다. U6 프로모터 또한 흔히 사용된다. T7 프로모터는 특정 구현예에서 시험관 내 전사에 사용될 수 있다.

[0152] 실시예 2: 관심 대상의 DNA 서열(관심 대상의 유전자)의 발현을 위한 DNA 벡터

[0153] 관심 대상의 DNA 서열은 적절한 발현 벡터 내로 삽입될 것이고, 위치가 관심 대상의 DNA 서열에 적절히 부가되어서 HIV1 인테그라아제가 게놈 내로의 통합을 위한 서열을 인지할 것이다. 이러한 위치는 att 위치(U5 및 U3 att 위치)로 명명된다(문헌 [Masuda et al, Journal of Virology, 1998, volume 72, pages 8396-8402] 참조). 게놈 내 타겟 위치에 대한 상동염기서열은 관심 대상의 DNA(유전자) 서열의 5' 및 3' 말단에 축적하는 영역 내에 포함될 수 있다(문헌 [Ishii et al, PLOS ONE, September 24, 2014, DOI: 10.1371/journal.pone.0108236] 참조). 재조합효소를 사용하는 경우, 인테그라아제 인식 위치는 포함되지 않을 수 있다. 약물 내성 마커(예를 들어, 블라스티시딘 또는 푸로마이신)와 같은 마커가 관심 대상의 DNA 서열의 삽입에 대한 확인을 위해, 그리고 게놈 내 랜덤 삽입에 대한 분석을 돕기 위하여 포함될 것이다. 이러한 내성 마커는 표적된 게놈 랜딩 패드(landing pad)로부터 그를 제거하도록 공학적으로 조작될 수 있다. 예를 들어, LoxP 위치를 갖는 푸로마이신 내성 유전자를 축적하고 외생적으로 발현된 CRE를 도입하는 것은 내부 서열을 제거하여 LoxP 위치를 함유하는 상처를 남길 것이다.

[0154] 실시예 3: 역전사효소 발현을 위한 DNA 벡터

[0155] 벡터 내에 관심 대상의 설계된 DNA 서열(유전자)이 RNA로서 발현되고, 인테그라아제 효소에 의한 통합을 위해 DNA로 다시 전환되어야 할 것이므로 시스템에서 역전사효소가 또한 공동발현될 수 있다. 역전사 효소는 그 기원이 바이러스(예를 들어, HIV1와 같은 레트로바이러스)일 수 있다. 이는 관심 대상의 DNA 서열과 동일한 벡터 내에 통합될 수 있다.

[0156] 실시예 4: 관심 대상의 DNA 서열과 함께 DNA 표적-인테그라아제(또는 재조합효소)의 공동 발현

[0157] 세포를 게놈 내 표적 위치에 대해 요구되는 카스9 RNA 가이드와 함께, 상기 기재된 벡터에 대해 전기천공시켰다. 일부 실험에서, 모든 성분을 발현한 벡터가 만들어졌다(융합 카스9/인테그라아제(또는 재조합 효소), 카스9 RNA 가이드, 및 인테그라아제 인식 위치를 갖고, 상동염기서열은 갖거나 갖지 않은 관심 대상의 DNA 서열). 벡터 내에 관심 대상의 설계된 DNA 서열(Gene사)이 RNA로서 발현되고, 인테그라아제 효소에 의한 통합을 위해 DNA로 다시 전환되어야 할 것이므로 시스템에서 역전사효소가 또한 공동발현될 수 있다. 역전사 효소는 그 기원이 바이러스(예를 들어, HIV1와 같은 레트로바이러스)일 수 있다. 기타 다른 실험에서, 관심 대상의 DNA 서열은 세포 내로의 도입 전에 선형화된다. 카스9 RNA 가이드 서열 및 관심 대상의 DNA 서열은 표준 분자생물학 프로토콜에 의해 설계되고, 이용 전에 벡터 내로 삽입되어야 하였다.

[0158] 실시예 5: 비표적 삽입을 위한 시험 실험 및 분석

[0159] 특정 유전자의 발현이 빠진 세포, 예컨대 녹아웃 마우스 모델로부터의 마우스 배아 섬유아세포 또는 소정의 유전자에 대해 녹아웃이 되도록 유전적으로 유전공학적으로 조작된 세포를 관심 대상의 유전자가 포함된 상기 벡터로 형질감염 또는 전기천공시킨다. 삽입된 유전자를 커버하도록 설계된 키메라성 프라이머 세트 및 축적 게놈성 서열은 편집된 세포들의 초기 풀을 스크리닝하는 데 사용될 것이다. 제한 희석 클로닝(LDC) 및/또는 FACS 분석을 이후 수행하여 단일클론성을 보장한다. 차세대 서열분석(NGS) 또는 단일 뉴클레오타이드 다형성(SNP) 분석을, 분리된 클론이 설계된 편집에 대해 동질성인지 보장하기 위한 최종 품질 제어 단계로서 수행한다. 스크리닝을 위한 기타 다른 메커니즘은, 이에 제한되지는 않지만 qRT-PCR 및 적절한 항체를 이용하는 웨스턴 블로팅을 포함할 수 있다. 단백질이 세포의 특정 표현형과 연관되는 경우, 세포는 그 표현형의 구제에 대해 검사될 수 있다. 세포의 게놈은 DNA 삽입의 특이성에 대해, 그리고 존재하는 경우 비표적 삽입의 상대적인 수를 알아내기 위하여 분석된다.

[0160] 실시예 6: 카스9 연결된 인테그라아제 단백질 발현 및 분리

[0161] 대장균 또는 곤충 세포 내에서 유전자 발현을 위해 설계된 벡터는 대장균 또는 곤충 세포 내로 통합되고, 소정의 기간 동안 발현되도록 둘 것이다. 몇몇 설계는 카스9(또는 비활성 카스9) 연결된 인테그라아제 단백질을 생성하도록 이용될 것이다. 벡터는 또한 높은 순도 및 수율을 갖는 단백질의 궁극적인 분리를 위해 His 또는 cMyc 태그에 제한되지 않는 태그를 포함할 것이다. 키메라성 단백질의 제조는 이에 제한되지는 않지만 표준 크로마토그래피 기법을 포함할 것이다. 단백질은 또한 하나 이상의 NLS(핵 국소화 신호 서열) 및/또는 TAT 서열을 이용

하여 설계될 수 있다. 핵 국소화 신호는 단백질이 핵으로 들어가는 것을 가능하게 한다. TAT 서열은 단백질의 셀 내로의 더 용이한 도입을 가능하게 한다(이는 세포-침투 펩티드이다). 당업계에서의 기타 다른 세포 침투 펩티드가 고려될 수 있다. 발현을 위한 충분한 시간이 발생된 후, 단백질 용해물은 세포로부터 수집되어, 사용된 태그에 따라 달라지는 적절한 컬럼 내에서 정제될 것이다. 정제된 단백질은 이후 적절한 완충 용액 내에 배치되어 -20℃ 또는 -80℃ 중 어느 하나의 온도에서 저장될 것이다.

[0162] 실시예 7: 정지 코돈을 전사 시작 위치의 바로 업스트림에 혼입시키는 데 카스9-인테그라아제를 이용

[0163] 본 개시 내용은 녹아웃 세포주 또는 유기체를 만들기 위한 방법을 포함한다. 상기 시스템은 관심 대상의 DNA 서열과 함께 사용되어 타겟 유전자에 대한 ATG 시작 위치 직후 1 개, 3 개, 6 개, 10 개, 15 개 또는 20 개의 연속적인 정지 코돈이 위치된다. 이는 ATG 시작 위치 후에 즉각적인 정지 코돈에 도달하는 때에 전사/번역이 정지됨에 따라, 이는 효과적인 유전자 녹아웃을 만들 것이다. 추가의 정지 코돈은 전사효소의 가능한 번져나감을 방지하는 것을 도울 것이다(전사효소가 첫번째 정지 코돈을 지나치는 경우).

[0164] 실시예 8: 세포의 게놈을 편집하기 위하여 정제된 단백질로서 애비1(또는 기타 다른 특이적 DNA 결합 도메인을 갖는 기타 다른 변형)을 이용

[0165] 애비1 분리된 단백질(레트로바이러스성 인테그라아제에 연결된 기타 다른 특정 DNA 서열 결합 단백질)을 바이러스 LTR 영역을 갖는 삽입성/통합성 DNA와 함께 적합한 완충액 내에 인큐베이션한다. (경우에 따라 4량체 또는 기타 다른 다량체의 형성을 위해). 대안적으로, 가이드 RNA를 갖는 분리된 애비1 단백질의 예비제조된(premade) 조성물은 삽입성 DNA 서열과 조합될 수 있다. 가이드 RNA를 포함시키고, 인큐베이션하여 가이드 RNA를 혼입시킨다. 애비1/DNA 제조물을 세포 내로 형질감염 또는 전기천공시킨다(또는 단백질을 세포에 제공하는 기타 다른 기법). 게놈/DNA 편집이 일어나도록 시간을 허용한다. 설계된 삽입성 DNA 서열의 세포의 게놈성 DNA의 특정 위치 내로의 삽입에 대해 확인한다. PCR 및 DNA 서열분석에 의해 비특이적 삽입에 대해 확인한다.

[0166] 현재 계획된 바에 따르면, 박테리아 발현 벡터는 pMAL-c5e일 것이며, 이는 NEB로부터의 생산중단된 제품으로, Genscript사에 대한 사내 자체 클로닝 선택 중 하나이다. 코돈-최적화된 Spy 카스9는 his-태그 및 말토오스-결합 단백질(MBP) 태그를 갖는 프레임 내 TEV 프로테아제 분할 위치를 이용하여 클로닝된다. ORF는 유도성 Tac 프로모터 하에 존재하며, 벡터는 또한 더 강한 조절을 위해 lac 리프레서 (LacI)에 대해 코딩한다. 아밀로오스 수지는 상당히 비싸기 때문에, MBP는 단지 안정화 태그로서만 사용될 것이며, 정제 태그로는 사용되지 않을 것이다. 가용성의 발현된 재료는 Ni-친화성 크로마토그래피로 정제될 것이며, 이후 카스9는 TEV 프로테아제에 의해 MBP로부터 방출되고, 양이온 교환 크로마토그래피에 의해 정제되고, 겔 여과에 의해 정련된다(polished).

[0167] 실시예 9: 융합 단백질에 대한 구조물의 설계

[0168] 서열 특이적 징크 핑거 도메인, TALE, 또는 표적 DNA 서열을 향한 크리스퍼 기반 접근을 위한 가이드 RNA를 설계한다. 선택한 온라인 설계 소프트웨어를 이용한다.

[0169] 위치 특이적 융합 인테그라아제 단백질을 형성하기 위하여, 인테그라아제, 전이효소 또는 재조합효소에 대한 코딩 서열; 적합한 아미노산 링커; 적절한 징크 핑거, TALE 또는 크리스퍼 단백질(예를 들어, 카스9, Cpf1); 및 핵 국소화 신호(또는 미토콘드리아 국소화 신호)를 갖는 DNA 구조물을 생산한다. 이들은 다수의 정렬로 구체화된다. 원하는 경우 단백질 분리 및 정제(예를 들어, 말토오스 결합 단백질(MBP) 또는 His 태그)을 위한 적합한 태그가 포함될 수 있다.

[0170] DNA 구조물은 당업계에서 일반적인 포유동물 세포 프로모터 또는 박테리아 프로모터를 이용할 수 있다(예를 들어, CMV, T7 등).

[0171] 대장균을 공급원으로서 이용하여 재조합 융합 단백질을 생산할 수 있다. 단백질을 당업계의 표준 수단에 의해 분리한다(예를 들어, MBP 컬럼, 니켈-세파로오스 컬럼, 등).

[0172] 도너-RNP 복합체를 조립한다(RNA 올리고를 듀플렉스(duplex)하고(융합 단백질이 그의 DNA 결합능에 대해 엔도뉴클레아제 비활성인 크리스퍼 관련 단백질, 예를 들어, 애비1을 갖는 경우) 본 발명의 융합 단백질과 혼합) - RNP를 형성하는 이들 단계들은 징크 핑거 도메인 및 TALE에서는 필요하지 않다.

[0173] 1. 도너 DNA와 적절한 LRT 도메인 및 삽입성 서열, 및 융합 단백질을 혼합하고, 10 분 동안 인큐베이션한다(대안적으로, RNP 복합체 형성 후 도너 DNA를 첨가한다).

[0174] 2. 뉴클레아제가 없는 IDTE 완충액 내 각각의 RNA 올리고(crRNA 및 tracrRNA)를 재현탁시킨다. 예를 들어, 최종

농도 100 μM 를 이용한다.

- [0175] 3. 멸균 미세원심분리관 내 등몰량의 농도로 두 개의 RNA 올리고를 혼합한다. 예를 들어, 하기 표를 이용하여 3 μM 의 최종 듀플렉스 농도를 생성한다: 성분 양 100 μM crRNA 3 μl 100 μM tracrRNA 3 μl 뉴클레아제가 없는 듀플렉스 완충액 94 μl 최종 부피 100 μl .
- [0176] 4. 95°C에서 5 분 동안 가열한다.
- [0177] 5. 열을 제거하고, 벤치 탑에서 실온(15°C 내지 25°C)까지 냉각되도록 둔다.
- [0178] 6. 필요에 따라, 뉴클레아제가 없는 듀플렉스 완충액 내에서 RNA를 작업 농도로 희석 듀플렉스한다(예를 들어, 3 μM).
- [0179] 7. 융합 단백질을 작업 버퍼(20 mM HEPES, 150 mM KCl, 5% 글리세롤, 1 mM DTT, pH 7.5)내에서 작업 농도(예를 들어, 5 μM)로 희석한다.
- [0180] 8. 각각의 형질감염을 위해, 1.5 pmol의 듀플렉스된 RNA 올리고(단계 A5)를 1.5 pmol의 융합 단백질(단계 A6)과 Opti-MEM 매질 내에서 최종 부피 12.5 μl 로 조합한다.
- [0181] 9. 실온에서 5 분 동안 인큐베이션하여 RNP 복합체를 조립한다.
- [0182] 실시예 10: 96-웰 플레이트 내 역 형질감염 gRNA-융합 단백질
- [0183] 1. 하기를 실온에서 20 분 동안 인큐베이션하여 형질감염 복합체를 형성한다: 성분 양 RNP(단계 A8) 12.5 μl 리포펙타민(Lipofectamine)[®] RNAiMAX 형질감염 시약 1.2 μl Opti-MEM[®] 매질 11.3 μl 총 부피 25.0 μl
- [0184] 2. 인큐베이션 동안(단계 B1), 항생제 없이 완전 배지를 이용하여 배양 세포를 400,000 세포/mL로 희석한다.
- [0185] 3. 인큐베이션이 완료되면, (단계 B1으로부터) 25 μl 의 형질감염 복합체를 96-웰 조직 배양 플레이트에 첨가한다.
- [0186] 4. (단계 B2로부터) 125 μl 의 희석된 세포를 96-웰 조직 배양 플레이트(50,000 세포/웰; RNP의 최종 농도는 10 nM이 될 것이다)에 첨가한다.
- [0187] 5. 형질감염 복합체 및 세포를 담은 플레이트를 조직 배양 인큐베이터 내에서 48 시간 동안 인큐베이션한다(37°C, 5% CO₂). 표적상 돌연변이를 검출하기 위하여, 적절한 프라이머를 이용하여 PCR을 사용한다(도너 서열 내의 프라이머 및 표적 삽입 위치를 둘러싼 프라이머).
- [0188] 실시예 11: 크리스퍼/카스9의 특이성을 시험하기 위한 프로토콜
- [0189] 비오틴에 연결된 d카스9(DNA 절단 비활성 카스9)를 생산한다(d카스9-비오틴). 카스9(s 파이오제네스(s pyogenes), s 아우레우스(s aureus) 등). 비오틴화 방법은 하기 기재된다.
- [0190] 비오틴화 방법 #1: N- 또는 C-말단에서 avi-태그(약 15 잔기)를 공학적으로 조작하고, WT(태그되지 않음) 단백질로서 발현 및 정제한다. 대장균 비오틴 리가제(BirA) 및 비오틴을 이용하여 avi-태그된 카스9를 비오틴화한다. 본 발명자는 이 방법을 이용하여 케모카인(chemokine)을 비오틴화한다. avi-태그 기술에 대한 IP는 수년 전에 만료된 것으로 여긴다.
- [0191] 비오틴화 방법 #2.1: 석신이미드-에스테르로 작용화된 비오틴은 표면-노출된 리신 잔기에 혼입될 수 있다(효소 반응은 요구되지 않음). 카스9와 같이 큰 단백질의 경우, 이는 실현가능한 선택안일 수 있다.
- [0192] 비오틴화 방법 #2.2: 동일한 방식으로, 비오틴-말레이미드는 상업적으로 입수가능하고, 이들은 표면-노출된 시스테인(효소 없음)에서 접합될 수 있다.
- [0193] 시험은 DNA-결합 및 분할 면에서 비오틴화된 카스9를 특징화하기 위하여 달성될 것이다.
- [0194] 스트렙타비딘-코팅된 96-웰 플레이트는 상업적으로 입수가능하지만, 또한 사내에서 자체로 생산될 수도 있다.
- [0195] d카스9-비오틴을 플라스틱 플레이트(96-웰, 24-웰, 384-웰 등)에 결합시킨다.
- [0196] 설계된 가이드 RNA를 각각의 웰에 제공한다. 가이드 RNA가 카스9 단백질과 상호작용하는 시간을 허용한다.
- [0197] 게놈성 DNA를 각 웰 또는 표적된 서열을 갖는 DNA에 제공한다. 카스9가 DNA에 결합되도록 시간을 허용한다.

- [0198] 적절한 완충액을 이용하여 웰을 세척한다.
- [0199] 어댑터(adpater)(DNA 올리고머)를 제공한다. 결합되는 시간을 허용한다.
- [0200] 게놈성 DNA를 제한-소화하여 이를 더 다루기 쉽고, 어댑터에 결합시키기에 더 용이하게 만든다.
- [0201] 웰을 세척한다.
- [0202] 결합 위치를 확인하기 위하여 DNA 서열을 수행한다(표적 상 대 표적 외).
- [0203] 실시예 12: 예비 1을 통한 NRF2 편집
- [0204] **도 5**는 가이드 Nrf2-sgRNA2 및 sgRNA3을 이용하는 Nrf2의 예비1 유전자 편집 표적화 엑손 2를 나타낸다. 예비1 편집을 통한 녹아웃을 위해 Nrf2 유전자좌를 표적하는 엑손 2에 대해 PCR 스크리닝한다. 가이드 Nrf2-sgRNA 2 및 3을 사용하여 Nrf2의 예비1 형질감염 표적화 엑손 2는 표적된 영역에서 도너의 통합을 나타내었다. 독특한 밴드가 1 내지 8에서와 같이 확인된다.
- [0205] **도 6**은 예비1 유전자 편집에 의해 생성된 이론적 데이터를 나타낸다. sgRNA 1 내지 3을 이용하여 게놈 재료를 표적하기 위한 예비1 시스템을 통하여 삽입된 도너 DNA를 시각화하는 DNA 겔 전기영동의 표시. 검은 밴드는 PCR 방법론으로 인한 배경 생성물을 나타낸다. 적색 밴드는 삽입물 및 삽입물의 영역을 측정하는 유전 재료를 증폭 시킴으로써 생성된 독특한 생성물을 나타낸다. 다수의 밴드는 표적된 영역에서 가능한 다수의 삽입을 나타낸다.
- [0206] **도 7**은 가이드 Nrf2-sgRNA 3을 이용하는 Nrf2의 예비1 유전자 편집 표적화 엑손 2를 나타낸다. 예비1 편집을 통한 녹아웃을 위해 Nrf2 유전자좌를 표적하는 엑손 2에 대해 PCR 스크리닝한다. 가이드 Nrf2-sgRNA 3을 사용하여 Nrf2의 표적화 엑손 2는 도너 삽입을 암시하였으며, 이는 도너 서열 및 인접 위치가 예측된 삽입에 대해 설계된 PCR 프라이머에 의해 표시되는 바와 같다. 독특한 밴드가 1 내지 4에서와 같이 확인된다.
- [0207] **도 8**은 혼합된 Hek293T 세포에서 Nrf2의 예비1 녹아웃을 나타낸다. (A) 혼합된 HEK293T 개체군 내 Nrf2의 녹아웃을 나타내는 55 kD 동형단백질(isoform)(Santa Cruz Bio사)에 대한 다중클론성 항체를 이용한 웨스턴 블롯 분석. (B) GAPDH(Santa Cruz Bio사) 부하 제어군.
- [0208] **도 9**는 혼합된 Hek293T 세포에서 Nrf2의 예비1 녹아웃을 나타낸다. (A) HEK 293t 세포에서 Nrf2 혼합된 개체군의 녹아웃을 나타내는 Nrf2(Abcam)에 대한 단일클론성 항체를 이용하는 웨스턴 블롯 분석. (B) GAPDH 부하 제어군. (C) 대조군에 비교하여 발현 비율에서 감소를 나타내는 농도계측 분석의 평균.
- [0209] 예비1 처리된 세포는 도너 DNA의 통합을 나타내는 독특한 PCR 밴드를 생성한다. HEK293T 혼합된 세포주에서 녹아웃의 표현형 확인은, 독특하며 상이한 항체를 갖는 두 개의 동형단백질을 프로빙하는(probing) 웨스턴 블롯 분석을 통해 확인되었다. 통합에 의한 약 80% 녹아웃이 혼합된 개체군 내에서 2 주 내에 관찰되었다.
- [0210] 실시예 13: 예비1을 통한 CXCR4 편집
- [0211] **도 10**은 CXCR4 엑손 2를 표적하는 예비1 유전자 편집을 나타낸다. 예비1을 통해 편집된 CXCR4의 표적화 엑손 2를 PCR 스크리닝한다. 4 개 세트의 프라이머를 관심 대상의 영역에 대하여 설계하였다. 세트 번호 2 및 4는 관심 대상의 영역에서 도너 DNA의 통합을 암시하는 독특한 밴드를 생성하는 것으로 보인다.
- [0212] 실시예 14: 예비1을 이용하여 NRF2 유전자좌에서 녹-인(knock-in) 실험을 위한 형질감염
- [0213] 유의: 500 ng 단백질 및 120 ng sgRNA를 단일 반응을 위해 사용한다. DNA의 양은 도너 구조물의 크기에 따라 달라진다. 도너 DNA(LTR 서열을 갖는 DNA)는 세포에 대해 제공/형질감염/전기천공되기 전, 동안 또는 후에 예비1과 함께 인큐베이션될 수 있다. 모든 반응은 멸균 생물안전 캐비닛(cabinet) 내에서 제조된다.
- [0214] 제1 일: 인간 배아 신장(HEK 293T) 세포를 24-웰 배양 플레이트(Corning사) 내에 10% 소 태아 혈청(Omega Scientific사)으로 보충된 500 μ l DMEM(Gibco사) 중의 HEK293T 세포(ATCC)를 웰 당 200,000개로 접종시켰다. 24 시간 동안 세포가 회복되도록 하였다.
- [0215] 제2 일: 예비1 제조:
- [0216] 튜브 1:
- [0217] 실온에서 10분 동안, 감소된-혈청 형질감염 매질(OptiMEM, Life Technologies사) 내 1:1 물 비의 정제된 예비1 단백질(서열번호 58) 및 도너 DNA (서열번호 101). 1.3 배 물 과량(대략 120 ng)으로의 sgRNA를 단백질/DNA 복

합체에 첨가하고, 실온에서 추가 10 분 동안 인큐베이션을 계속하였다. 이 혼합물의 부피는 25 μ l이다.

- [0218] 튜브 2:
- [0219] 2 μ l의 형질감염 시약(RNAiMAX, Life Technologies사)을 23 μ l의 OptiMEM에 첨가하였다. 그리고, 실온에서 10 분 동안 인큐베이션하였다.
- [0220] 튜브 1 및 튜브 2(최종 부피 50 ml)를 합하고 실온에서 15 분 동안 인큐베이션하였다.
- [0221] 전체 50 μ l 형질감염 혼합물을 웰에 첨가하였다.
- [0222] 혼합된 개체군 내에서 게놈 DNA 편집의 확인을 위해 혼합 편집된 세포들의 절반을 형질감염 후 48 시간에 수확하였다. 편집된 게놈의 확인은 중합효소 연쇄 반응(PCR)에 의해 수행하였다. 본 발명자들은 하기 기재된 바와 같이(PCR 프로토콜 참조) 표적된 영역에 대하여 PCR을 수행하였으며, 나머지는 6 cm 배양접시(Corning사) 상에 접종하고 48 시간 동안 회복되도록 하였다.
- [0223] 제5 일: 웨스턴 블롯팅을 통한 표현형 변화의 스크리닝
- [0224] 표준 웨스턴 블롯 분석을, 55 kD 동형단백질(Santa Cruz Biotechnology사, sc-722) 및 98 kD 동형단백질(Abcam사, ab-62352)을 표적하는 일차 항체를 이용하여, Nrf2 동형단백질에 대하여 수행하였다. GAPDH(Santa Cruz Biotechnology사, sc-51907).
- [0225] 실시예 15: Nrf2 및 CXCR4 유전자좌에 대해 예비1을 이용한 유전자 편집의 검출을 위한 PCR 조건
- [0226] 인간 Nrf2에 대한 등록 번호(Accession number)
- [0227] 유니프로트(Uniprot): Q16236
- [0228] 엔SEMBL(Ensembl) 유전자 ID: ENSG00000116044
- [0229] Nrf2(엑손2)에 대한 편집 표적 서열 및 PAM: sgRNA 설계 1 내지 3에 대해 사용됨.
- GCGACGGAAAGAGTATGAGC TGG**
TATTTGACTTCAGTCAGCGA CGG
TGGAGGCAAGATATAGATCT TGG
- [0230]
- [0231] Nrf2 표적에서 통합의 검출을 위한 프라이머 키
- [0232] 프라이머 세트 1: 프라이머 1: 5'-GTGTTAATTTCAAACATCAGCAGC-3', 프라이머 2: 5'- GACAAGACATCCTTGATTG-3'
- [0233] 프라이머 세트 2: 프라이머 1: 5'-GAGGTTGACTGTGTAAATG-3', 프라이머 2: 5'- GATACCAGAGTCACACAACAG-3'
- [0234] 프라이머 세트 3: 프라이머 1: 5'-TCTACATTAATTCTCTTGTC-3', 프라이머 2: 5'- GATACCAGAGTCACACAACAG-3'
- [0235] 인간 CXCR4에 대한 등록 번호
- [0236] 유니프로트 P61073
- [0237] 엔SEMBL 유전자 ID: ENSG00000121966
- [0238] CXCR4(엑손 2)에 대한 편집 표적 서열 및 PAM: sgRNA 설계1에 대해 사용됨.
- [0239] **GGGCAATGGATTGGTCATCC TGG**
- [0240] CXCR4 타겟에서 통합의 검출을 위한 프라이머 키
- [0241] 프라이머 세트 1: 프라이머 1: 5'- TCTACATTAATTCTCTTGTC-3', 프라이머 2: 5'- GACAAGACATCCTTGATTG-3'
- [0242] 프라이머 세트 2: 프라이머 1: 5'- TCTACATTAATTCTCTTGTC-3', 프라이머 2: 5'- GATACCAGAGTCACACAACAG -3'
- [0243] 프라이머 세트 3: 프라이머 1: 5'- GAGGTTGACTGTGTAAATG -3', 프라이머 2: 5'- GACAAGACATCCTTGATTG-3'
- [0244] 프라이머 세트 4: 프라이머 1: 5'- GAGGTTGACTGTGTAAATG -3', 프라이머 2: 5'- GATACCAGAGTCACACAACAG -3'
- [0245] 통합된 도너 DNA의 검출을 위해 사용된 PCR 사이클링 조건

- [0246] * 어닐링(annealing) 온도는 프라이머 서열에 따라 달라질 것임에 유의한다.
- [0247] 1. 초기 변성: 4 분 94
- [0248] 2. 변성: 30 초 94
- [0249] 3. 어닐링: 30 초 55
- [0250] 4. 연장: 30 초 72
- [0251] 5. 단계 2로 감: 40 사이클
- [0252] 6. 최종 연장: 4 분 72
- [0253] 7. 최종 유지: ∞ 4
- [0254] 비오틴화에 대해 Avi-태그된 카스9
- [0255] 카스9 비오틴화에 사용된 avi-태그의 서열
- [0256] 아미노산 서열:
- [0257] G G D L E G S G L N D I F E A Q K I E W H E *
- [0258] 핵산 서열:
- [0259] GGCGGCGACCTCGAGGGTAGCGGTCTGAACGATATTTTGAAGCGCA
GAAAATTGAATGGCATGAATAA
- [0260] 첫 번째 밑줄 부분 = 카스9 C-말단
- [0261] 이탤릭체 부분 = 제한 위치/링커
- [0262] 두 번째 밑줄 부분 = avi-태그(비오틴화 위치 강조)
- [0263] 실시예 16: 애비1 융합 단백질에 대한 발현 프로토콜
- [0264] 전장 융합 단백질(서열번호 57)을 함유하는 발현 구조물의 형질전환.
- [0265] -80℃ 냉동고에서 컴피턴트(competent) 대장균 세포를 취한다.
- [0266] 수조를 42℃로 한다.
- [0267] 1.5 ml 튜브(에펜도르프 또는 유사물) 내에 컴피턴트 세포를 넣는다. DNA 구조물을 형질전환하기 위하여, 50 μ l의 컴피턴트 세포를 사용한다.
- [0268] 튜브를 얼음 중에 유지시킨다.
- [0269] 50 ng의 환형 DNA를 대장균 세포 내로 첨가한다. 얼음 상에서 10 분 동안 인큐베이션하여 컴피턴트 세포를 해동시킨다.
- [0270] DNA와 대장균이 있는 튜브(들)를 42℃의 수조 내에 45 초 동안 둔다. 튜브를 다시 얼음 상에 2 분 동안 두어 대장균 세포에 대한 손상을 감소시킨다.
- [0271] 1 ml의 LB(첨가된 항생제 없음)를 첨가한다. 튜브들을 1 시간 동안 37℃에서 인큐베이션시킨다(튜브를 30 분 동안 인큐베이션할 수 있음).
- [0272] 결과로 생성된 배양물 약 100 μ l를 LB 플레이트 상에 적절한 항생물질과 함께 도말한다.
- [0273] 약 12 시간 내지 16 시간 후 콜로니들(colonies)을 취한다.
- [0274] 접종 및 확장
- [0275] LB 및 항생물질을 함유하는 1 리터의 플라스크를 접종한다.
- [0276] 0.6 OD가 달성될 때까지 박테리아 배양이 성장하도록 두고, 이후 1 mM 최종 농도에서 이소프로필 β -D-1-티오갈락토피라노사이드(IPTG)를 이용하여 유도한다.

- [0277] 배양물이 6 시간 내지 8 시간 동안 확장되도록 하고, 현탁된 박테리아 배양물을 최소 2000 G 력으로 10 분 동안 원심분리한다.
- [0278] 이후에 추가로 가공하기 위하여 -80℃에서 펠렛을 동결시킨다.
- [0279] 단백질 제조 및 정제
- [0280] 모든 단계는 실온에서 수행된다.
- [0281] 20 mM 트리스 pH 8.0, 300 mM NaCl, 0.1 mg/ml 달걀 흰자 리소자임 내에서 동결-해동의 2 회 사이클에 의해 세포를 용해시킨다. 6,000 g에서 15 분 동안 원심분리하고 상청액을 유지한다.
- [0282] 20 mM Tris pH 8.0, 300 mM 염화나트륨 내에서 평형화된 Ni-IDA 아가로스 컬럼 상에 상청액을 부하한다. 이미다졸의 0 mM 내지 200 mM 구배를 이용하여 단백질을 용리한다. 7% SDS-PAGE에 의해 융합 단백질을 함유하는 분획을 확인한다.
- [0283] 분획들을 혼합하고 20 mM Tris pH 8.0로 희석하여 최종 NaCl 농도가 50 mM이 되도록 한다. Q-세파로스 컬럼 상에 부하하고, 염화나트륨의 0 mM 내지 500 mM 구배를 이용하여 용리한다. 7% SDS-PAGE에 의해 융합 단백질을 함유하는 분획을 확인한다.
- [0284] 분획들을 혼합하고 20 mM Tris pH8.0로 희석하여 최종 NaCl 농도가 100 mM이 되도록 한다. SP-세파로스 컬럼 상에 부하하고, 염화나트륨의 0 mM 내지 500 mM 구배를 이용하여 용리한다. 7% SDS-PAGE에 의해 융합 단백질을 함유하는 분획을 확인한다.
- [0285] 분획들을 혼합하고, UV 흡광도 280 nm에서 농도를 측정하고, 원심분리 필터에 의해 최종 농도 400 µg/ml로 농축한다. 최종 농도 50%로 글리세롤을 첨가한다. -20℃에서 저장한다.
- [0286] 특정 구현예들이 본 명세서에 나타내어지고 기재되었지만, 그러한 구현예들은 단지 예시를 위해 제공된다는 것은 당업자에게 명백할 것이다. 다양한 변형, 변화, 및 치환은 이제 본 개시 내용에서 벗어나지 않으면서 당업자에게 일어날 것이다. 본 명세서에 기재된 개시 내용의 구현예에 대한 각종 대안들이 본 개시 내용의 실시예에 사용될 수 있다는 것이 이해되어야 한다. 하기 청구범위는 본 개시 내용의 범주를 규정하고, 이들 청구범위 및 그의 균등물의 범위 내의 방법 및 구조는 그에 의해 커버되는 것으로 의도된다.
- [0287] **본 개시 내용의 서열들**
- [0288] 하기 제공된 각각의 서열에 대하여, 하기 정보가 제공된다: 서열 유형(핵산 또는 아미노산), 공급원(예를 들어, 대장균), 길이, 및 식별 번호(이용가능한 경우).
- [0289] 본 개시 내용의 제1 폴리뉴클레오티드는, 예를 들어 카스9, Cpf1, TALE, 또는 ZnFn 단백질을 암호화할 수 있다. 본 개시 내용의 제2 폴리뉴클레오티드는, 예를 들어 인테그라아제, 전이효소, 또는 재조합효소를 암호화할 수 있다. 하기에 예시적인 제1 및 제2 폴리뉴클레오티드 서열 및 단백질 서열이, 예시적인 링커 서열과 함께 열거되었으며, 이는 본 명세서에서 기재된 조성물(구조물, 융합 단백질) 및 방법에서 사용될 수 있다. 하기 표 1에서 열거되지 않았지만 본 명세서에서 기재된 조성물(구조물, 융합 단백질) 및 방법에서 사용될 수 있는 기타 다른 폴리뉴클레오티드 서열, 단백질 서열, 또는 링커 서열이 본 개시 내용에 제공될 수 있다. 예를 들어, 서열번호 49, 서열번호 57, 서열번호 58 및/또는 이의 일부.
- [0290] 링커는 임의의 길이일 수 있으며, 예를 들어 길이가 3 개 내지 300 개의 뉴클레오티드, 길이가 6 개 내지 60 개의 뉴클레오티드, 또는 제1 및 제2 폴리뉴클레오티드가 융합되는 것을 허용하는 임의의 길이이다. 폴리펩타이드는 유기체, 예를 들어 대장균에 의해 제조될 수 있거나, 합성 제조되거나, 이들 모두의 조합에 의해 제조될 수 있다.
- [0291] 예시적인 핵산 서열: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 27 내지 47, 49, 55, 56, 57, 62, 64, 66, 68, 70, 79, 82, 및 83.
- [0292] 예시적인 아미노산 서열: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 48, 50, 52, 58, 63, 65, 67, 69, 71, 72 내지 78, 및 80.

[0293]

제1 단백질, 제2 단백질, 또는 링커

제1 폴리뉴클레오타이드 또는 단백질 서열 서열번호	제2 폴리뉴클레오타이드 또는 단백질 서열 서열번호	링커 서열 서열번호 또는 서열
1 내지 14, 27 내지 46, 50, 52, 56, 57, 68, 69, 72 내지 78, 86 내지 92, 200 내지 253	15 내지 26, 47, 48, 55, 57, 62 내지 67, 70, 71, 79, 80	51, 54, 61, GGS

표 2

[0294]

서열의 일부 목록

유전자 (단백질)	박테리아/바이러스	DNA 서열	단백질 서열	서열번호 (DNA, 단백질)
카스9	S. 썬모필러스(S. thermophilus)	HQ712120.1	Q03JI6.1	서열번호 1, 2
	P. 멀토사이다(P. multocida)		Q9CLT2.1	서열번호 3, 4
	S. 뮈탄스(S. mutans)		Q8DTE3.1	서열번호 5, 6
	N. 메닝기타이데스(N. meningitides)		C9X1G5.1	서열번호 7, 8
	S. 미티스(S. mitis)		KJQ69483.1	서열번호 9, 10
	S. 마카카에(S. macacae)		EHJ52063.1	서열번호 11, 12
	스테필로코커스 아우레우스		KKJ92487.1	서열번호 49, 50
	S. 파이오제네스		AFV37892.1	서열번호 13, 14
인테그라아제	HIV1		ABR68182.1	서열번호 15, 16
	유인원 T-림프구 바이러스		AAA47841.1	서열번호 17, 18
	S. 뉴모니아(S. pneumonia)		CBW38769.1	서열번호 19, 20
	대장균		CAA41325.1	서열번호 21, 22
	렌티바이러스			서열번호 47, 48
재조합효소	썬모언에어로박테리움 파지(Thermoanaerobacterium phage)		YP_006546326.1	서열번호 23 24

[0295]

추가 서열들

[0296]

서열번호 1

[0297]

명칭: S. 썬모필러스 Csn1 cds HQ712120.1

[0298]

서열:

ATGACTAAGCCATACTCAATTGGACTTGATATTGGAACGAATAGTGTGGAT
 GGGCTGTAATAACTGATAATTACAAGGTTCCGTCTAAAAAATGAAAGTCTT
 AGGAAATACGAGTAAAAAGTATATCAAAAAGAACCTGTTAGGTGTATTACTC
 TTTGACTCTGGAATCACAGCAGAAGGAAGAAGATTGAAGCGTACTGCAAGAA
 GACGTTATACTAGACGCCGTAATCGTATCCTTTATTTGCAGGAAATTTTAGC
 ACGGAGATGGCTACATTAGATGATGCTTTCTTTCAAAGACTTGACGATTCGTT
 TTTAGTTCCTGATGATAAACGTGATAGTAAGTATCCGATATTTGGAACTTAG
 TAGAAGAAAAAGTCTATCATGATGAATTTCCAACCTATCTATCATTTAAGGAA
 ATATTTAGCAGATAGTACTAAAAAAGCAGATTTGCGTCTAGTTTATCTTGCAT
 TGGCTCATATGATTAAATATAGAGGTCACCTCTTAATTGAAGGAGAGTTTAAT
 TCAAAAAATAATGATATTCAGAAGAATTTTCAAGACTTTTGGACACTTATAA
 TGCTATTTTTGAATCGGATTTATCACTTGAGAATAGTAAACAACCTGAGGAAA
 TTGTTAAAGATAAGATTAGTAAATTAGAAAAGAAAGATCGTATTTTAAACT
 CTTCCCTGGGGAGAAGAATTCGGGGATTTTTTCAGAGTTTCTAAAGTTGATTG
 TAGGAAATCAAGCTGATTTTAGGAAATGTTTTAATTTAGACGAAAAAGCCTC
 CTTACATTTTTCCAAAGAAAGCTATGATGAAGATTTAGAGACTTTGTTAGGTT
 ATATTGGAGATGATTACAGTGATGTCTTTCTCAAAGCAAAGAACTTTATGAT
 GCTATTCTTTATCGGGTTTTCTGACTGTAAGTATAATGAGACAGAAGCACC
 TCTCTCTTCTGCTATGATAAAGCGATATAATGAACACAAAGAAGATTAGCGT
 TACTAAAGGAATATATAAGAAATATTTCACTAAAAACGTATAATGAAGTATT

[0299]

TAAAGATGACACCAAAAATGGTTATGCTGGTTATATTGATGGAAAAACAAAT
CAGGAAGATTTCTACGTATATCTAAAAACCTATTGGCTGAATTTGAAGGTG
CGGATTATTTTCTTGAAAAAATTGATCGAGAAGATTTTTTGAGAAAGCAACGT
ACATTTGACAATGGTTCGATACCATATCAGATTCATCTTCAAGAAATGAGAG
CAATTCCTTGATAAGCAAGCTAAATTTTATCCTTTCTTGGCTAAAAATAAAGAA
AGAATCGAGAAGATTTTAACTTCCGAATTCCTTATTATGTAGGTCCACTTGC
GAGAGGGAATAGTGATTTTGCCTGGTCAATAAGAAAACGAAATGAAAAAATT
ACACCTTGGAATTTTGAGGACGTTATTGACAAAGAATCTTCGGCAGAGGCTTT
CATTAAATCGAATGACTAGTTTTGATTTGTATTTGCCAGAAGAGAAGGTACTION
CAAAGCATAGTCTCTTATACGAACTTTTAATGTATATAATGAATTAACAAAA
GTTAGATTTATTGCCGAAAGTATGAGAGATTATCAATTTTTAGATAGTAAGCA
GAAGAAAGATATTGTTAGACTTTATTTTAAAGATAAAAGGAAAGTTACTGAT
AAGGATATTATTGAATATTTACATGCAATTTATGGGTATGATGGAATTGAATT
AAAAGGCATAGAGAAACAGTTTAATTCTAGTTTATCTACTTATCACGATCTTT
TAAATATTATTAATGATAAAGAGTTTTTGGATGATAGTTCAAATGAAGCGATT
ATCGAAGAAATTATCCATACTTTGACAATTTTTGAAGATAGAGAGATGATAA
AACAAACGTCTTTCAAAATTTGAGAATATATTCGATAAATCCGTTTTGAAAAAG
TTATCTCGTAGACATTACACTGGCTGGGGTAAGTTATCTGCTAAGCTTATTAA
TGGTATTCGAGATGAAAAATCTGGTAATACTATTCTTGATTACTTAATTGATG
ATGGTATTTCTAACCGTAATTTTCATGCAACTTATTCACGATGATGCTCTTTCTT
TAAAAAGAAGATACAGAAAGCACAAATTATTGGTGACGAAGATAAAGGTA
ATATTAAAGAGGTCGTTAAGTCTTTGCCAGGTAGTCCTGCGATTAAAAAAGG
TATTTTACAAAGCATAAAAAATTGTAGATGAATTGGTCAAAGTAATGGGAGGA
AGAAAACCCGAGTCAATTGTTGTTGAGATGGCTCGTGAAAATCAATATACCA
ATCAAGGTAAGTCTAATTCCCAACAACGCTTGAAACGTTTAGAAAAATCTCT
CAAAGAGTTAGGTAGTAAGATACTTAAGGAAAATATTCCTGCAAACTTTCT
AAAATAGACAATAACGCACTTCAAAATGATCGACTTTACTTATACTATCTTCA
AAATGGAAAAGATATGTATACCGGAGATGATTTAGATATTGATAGATTAAAGT
AATTATGATATTGATCATATTATTCCTCAAGCTTTTTTGAAAGATAATTCTATT
GACAATAAAGTACTTGTTCATCTGCTAGTAACCGTGGTAAATCAGATGATTT
TCCAAGTTTAGAGGTTGTCAAAAAAAGAAAGACATTTTGGTATCAATTATTG

[0300]

AAATCAAAATTAATTTCTCAACGAAAATTTGATAATCTGACAAAAGCTGAAC
GGGGAGGATTGTTACCTGAGGACAAAGCTGGTTTTATTCAACGCCAGTTGGT
TGAACACGTCAAATAACAAAACATGTAGCTCGTTTACTTGATGAGAAATTT
AATAATAAAAAAGATGAAAATAATAGAGCGGTACGAACAGTAAAAATTATT
ACCTTGAAATCTACCTTAGTTTCTCAATTTTCGTAAGGATTTTGAACTTTATAA
AGTTCGTGAAATCAATGATTTTCATCATGCTCATGATGCTTACTTGAATGCCG
TTATAGCAAGTGCTTTACTTAAGAAATACCCTAACTAGAGCCAGAATTTGTG
TACGGTGATTATCCAAAATACAATAGTTTGTAGAGAAAGAAAGTCCGCTACAG
AAAAGGTATATTTCTATTCAAATATCATGAATATCTTTAAAAAATCTATTTCT
TTAGCTGATGGTAGAGTTATTGAAAGACCACTTATTGAGGTAAATGAGGAGA
CCGGCGAATCCGTTTGAATAAAGAATCTGATTTAGCAACTGTAAGGAGAGT
ACTCTCTTATCCGCAAGTAAATGTTGTGAAAAAAGTTGAGGAACAGAATCAC
GGATTGGATAGAGGAAAACCAAAGGGATTGTTTAATGCAAATCTTTCCTCAA
AGCCAAAACCAAATAGTAATGAAAATTTAGTAGGTGCTAAAGAGTATCTTGA
CCCCAAAAGTATGGGGGGTATGCTGGAATTTCTAATTCTTTTGCTGTTCTTG
TTAAAGGGACAATTGAAAAAGGTGCTAAGAAAAAATAACAAATGTACTAG
AATTTCAAGGTATTTCTATTTTAGATAGGATTAATTATAGAAAAGATAAACTT
AATTTTTTACTTGAAAAAGGTTATAAAGATATTGAGTTAATTATTGAACTACC
TAAATATAGTTTATTTGAACTTTCAGATGGTTCACGTCGTATGTTGGCTAGTA
TTTTGTCAACGAATAATAAGAGGGGAGAGATTCACAAAGGAAATCAGATTTT
TCTTTCACAGAAGTTTGTGAAATTACTTTATCATGCTAAGAGAATAAGTAACA
CAATTAATGAGAATCATAGAAAATATGTTGAGAACCATAAAAAAGAGTTTGA
AGAATTATTTTACTACATTCTTGAGTTTAATGAGAATTATGTTGGAGCTAAAA
AGAATGGTAAACTTTTAACTCTGCCTTTCAATCTTGGCAAAATCATAGTATA
GATGAACTCTGTAGTAGTTTTATAGGACCTACCGGAAGTGAAAGAAAGGGGC
TATTTGAATTAACCTCTCGTGGAAGTGCTGCTGATTTTGAATTTTGTAGGTGTTA
AAATCCAAGGTATAGAGACTATACCCCATCATCCCTATTAAAAGATGCCAC
ACTTATTCATCAATCTGTTACAGGCCTCTATGAAACACGAATAGACCTTGCCA
AACTAGGAGAGGGTTAA

[0301]

[0302]

서열번호 2

[0303]

서열:

MTKPYSIGLDIGTNSVGWAVITDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFD
SGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFLVPDD
KRDSKYPIFGNLVEEKVYHDEFPTIYHLRKYLADESTKKADLRLVYLALAHMIKY
RGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEK
KDRILKLPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLET
LLGYIGDDYSDVFLKAKKLYDAILLSGFLTVDNETEAPLSSAMIKRYNEHKEDL
ALLKEYIRNISLKYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKNLLAEFEGA
DYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPLAKNKERIEK
ILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSF
DLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYF
KDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLSTYHDLLNIINDKEFLDDSS
NEAIIIEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLSRRHYTGWGKLSAKLIN
GIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEV
VKSLPGSPAIIKKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQ
QRLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTGD
DLIDRLSNYDIDHIIPQAFKDNSIDNKVLVSSASNRGKSDDFPSLEVVKRKRTF
WYQLLSKSLISQRKFDNLTKAERGGLLPEDKAGFIQRQLVETRQITKHVARLLDE
KFNNKKDENNRVVRTVKIITLTKSTLVSQFRKDFELYKVRINDFHHAHDAYLNA
VIASALLKKYPKLEPEFVYGDYPKYNSFRERKSATEKVYFYFYSNIMNIFKKSISLAD
GRVIERPLIEVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVEEQNHGLDRG
KPKGLFNANLSSKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFAVLVKGTIEK
GAKKKITNVLEFQGISILDRINYRKDKLNFLEKGYKDIELIIELPKYSLEFSDGSR
RMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKK
EFEELFYIILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSEKGL
FELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGE
G

[0304]

서열번호 3

[0305]

명칭: P. 멀토사이다 카스9

[0306]

서열:

ATGCAAACAACAAATTTAAGTTATATTTTAGGTTTAGATTGGGGATCGCTTC
TGTAGGTTGGGCTGTCGTTGAAATCAATGAAAATGAAGACCCTATCGGCTTG

[0308]

ATTGATGTAGGAGTAAGGATATTTGAGCGTGCTGAGGTACCCAAAACCTGGAG
AATCTTTAGCACTCTCTCGCCGTCTTGCAAGAAGTACTCGCCGTTTGATACGC
CGTCGTGCACACCGTTTACTCCTCGCAAAACGCTTCTTAAAACGTGAAGGTAT
ACTTTCCACAATCGACTTAGAAAAAGGATTACCCAACCAAGCTTGGGAATTA
CGTGTCGCCGGTCTTGAACGTCGGTTATCCGCCATAGAATGGGGTGCGGTTCT
GCTACATTTAATCAAGCATCGAGGTTATCTTTCTAAACGTAAAAATGAATCCC
AAACAAACAACAAAGAATTAGGAGCCTTACTCTCTGGAGTGGCACAAAACCA
TCAATTATTACAATCAGATGACTACCGAACACCAGCAGAGCTCGCACTGAAA
AAATTTGCTAAAGAAGAAGGGCATATCCGTAATCAACGAGGTGCCTATACAC
ATACATTTAATCGATTAGACTTATTAGCTGAACTTAACTTGCTTTTTGCTCAAC
AACATCAGTTTGGTAACCCTCACTGTAAAGAGCATATTCAACAATATATGAC
AGAATTGCTTATGTGGCAAAAGCCAGCCTTATCTGGTGAGGCAATTTTAAAA
ATGTTGGGTAAATGTACGCATGAAAAAATGAGTTTAAAGCAGCAAAACATA
CCTACAGTGCGGAGCGCTTTGTTTGGCTAACCAAAACCTCAATAACTTGCGCATT
TTAGAAGATGGGGCAGAACGAGCTCTTAATGAAGAAGAACGTCAACTATTGA
TAAATCATCCGTATGAGAAATCAAAATTAACCTATGCCCAAGTCAGAAAATT
GTTAGGGCTTTCCGAACAAGCGATTTTAAAGCATCTACGTTATAGTAAAGAA
AACGCAGAATCAGCTACTTTTATGGAGCTTAAAGCTTGGCATGCAATTCGTA
AAGCGTTAGAAAATCAAGGATTGAAGGATACTTGGCAAGATCTCGCTAAGAA
ACCTGACTTACTAGATGAAATTGGTACCGCATTTTCTCTTTATAAAACTGATG
AAGATATTAGCAATATTTGACAAATAAGGTACCGAACTCAGTCATCAATGC
ATTATTAGTTTCTCTGAATTTGATAAATTCATTGAGTTATCTTTGAAAAGTTT
ACGTAAAATCTTGCCCTAATGGAGCAAGGTAAGCGTTATGATCAAGCTTGT
CGTGAAATTTATGGGCATCATTATGGTGAGGCAAATCAAAAAACTTCTCAGC
TACTACCAGCTATTCCAGCCCAAGAAATTCGTAATCCTGTTGTTTTACGTACA
CTTTCACAAGCACGTAAAGTGATCAATGCCATTATTCGTCAATATGGTTCCCC
TGCTCGAGTCCATATTGAAACAGGAAGAGAACTTGGGAAATCTTTTAAAGAA
CGTCGTGAAATTCAAAAACAACAGGAAGATAATCGAACTAAGCGAGAAAGT
GCGGTACAAAAATTCAAAGAATTATTTTCTGACTTTTCAAGTGAACCCAAAA
GTAAAGATATTTTAAAATTCCGCTTATACGAACAACAGCATGGTAAATGCTT
ATACTCTGGAAAAGAGATCAATATTCATCGCTTAAATGAAAAGGGTTATGTG

[0309]

GAAATTGATCATGCTTTACCTTTCTCACGGACTTGGGATGATAGTTTTAATAA
TAAAGTATTAGTTCTTGCCAGCGAAAACCAAAACAAAGGGAATCAAACACCG
TATGAATGGCTACAAGGTAAAATAAATTCGGAACGTTGGAAAACTTTGTTG
CTTTAGTACTGGGTAGCCAGTGCAGTGCAGCCAAGAAACAACGATTACTCAC
TCAAGTTATTGATGATAATAAATTTATTGATAGAACTTAAATGATACTCGCT
ATATTGCCCCGATTCTATCCAACCTATATTCAAGAAAATTTGCTTTTGGTGGGT
AAAAATAAGAAAAATGTCTTTACACCAAACGGTCAAATTACTGCATTATTAA
GAAGTCGCTGGGGATTAATTAAGGCTCGTGAGAATAATAACCGTCATCATGC
TTTAGATGCGATAGTTGTGGCTTGTGCAACACCTTCTATGCAACAAAAAATTA
CCCGATTTATTGATTTAAAGAAGTGCATCCATACAAAATAGAAAATAGGTA
TGAAATGGTGGATCAAGAAAGCGGAGAAATTATTTACCTCATTTTCCTGAA
CCTTGGGCTTATTTTAGACAAGAGGTTAATATTCTGTGTTTTTGATAATCATCC
AGATACTGTCTTAAAAGAGATGCTACCTGATCGCCACAAGCAAATCACCAG
TTTGTACAGCCCCCTTTTGTCTCGTGCCCCAACTCGTAAAATGAGTGGTCA
AGGGCATATGGAAACAATTAAATCAGCTAAACGCTTAGCAGAAGGCATTAGC
GTTTTAAGAATTCCTCTCACGCAATTAACCTAATTTATTGGAAAATATGGT
GAATAAAGAACGTGAGCCAGCACTTTATGCAGGACTAAAAGCACGCTTGGCT
GAATTTAATCAAGATCCAGCAAAAGCGTTTGCTACGCCTTTTATAACAAG
GAGGGCAGCAGGTCAAAGCTATTCGTGTTGAACAGGTACAAAAATCAGGGGT
ATTAGTCAGAGAAAACAATGGGGTAGCAGATAATGCCTCTATCGTTCGAACA
GACGTATTTATCAAAAATAATAAATTTTTCCTTGTTCTATCTATACTTGGCA
AGTTGCGAAAGGCATCTTGCCAAATAAAGCTATTGTTGCTCATAAAAATGAA
GATGAATGGGAAGAAATGGATGAAGGTGCTAAGTTTAAATTCAGCCTTTTCC
CGAATGATCTTGTGAGCTAAAAACCAAAAAAGAATACTTTTTCGGCTATTA
CATCGGACTAGATCGTGCAACTGGAAACATTAGCCTAAAAGAACATGATGGT
GAGATATCAAAAGGTAAAGACGGTGTTTACCGTGTTGGTGTCAAGTTAGCTC
TTTCTTTTGAAAAATATCAAGTTGATGAGCTCGGTAAAAATAGACAAATTTGC
CGACCTCAGCAAAAGACAACCTGTGCGTTAA

[0310]

[0311]

서열번호 4

[0312] 서열:

MQTTNLSYILGLDLGIASVGWAVVEINENEDPIGLIDVGVRIFERAIEVPKTGESLA
LSRRLARSTRRLIRRRRAHRLLLAKRFLKREGILSTIDLEKGLPNQAWELRVAGLER
RLSAIEWGAVLLHLIKHRGYLSKRKNESQTNNKELGALLSGVAQNHQLLQSDDY
RTPAELALKKFAKEEGHIRNQRGAYTHTFNRLDLLAELNLLFAQQHQFGNPHCK
EHIQQYMTTELLMWQKPALSGEAILKMLGKCTHEKNEFKAAKHTYSAERFVWLT
KLNNLRILEDGAERALNEEERQLLINHPYEKSKLTYAQVRKLLGLSEQAIFKHLR
YSKENAESATFMELKAWHAIKALENQGLKDTWQDLAKKPDLLDEIGTAFSLY
KTDEDIQQYLTKVPNSVINALLVSLNFDKFIELSLKSLRKILPLMEQGKRYDQAC
REIYGHYGEANQKTSQLLPAIPAQEIRNPVVLRTLQARKVINAIIRQYGGPARV
HIETGRELKGSFKERREIQKQEDNRTKRESAVQKFELFSDFSSEPKSKDILKFR
LYEQHQHGKCLYSKGKINIHLNEKGYVEIDHALPFSRTWDDSFNNKVLVLASEN
QNKGNQTPYEWLQGKINSERWKNFVALVLGSQCSAAKKQRLLTQVIDDNKFID
RNLNDTRYIARFLSNYIQENLLL VGKNKKNVFTPNGQITALLRSRWGLIKARENN
NRHHALDAIVVACATPSMQQKITRIFRKEVHPYKIENRYEMVDQESGEIISPHFP
EPWAYFRQEVNIRVFDNHPDVLKEMLPDRPQANHGFVQPLFVSRAPTRKMSG
QGHMETIKSAKRLAEGISVLRIPLTQLKPNLLENMVNKEREPALYAGLKARLAEF
NQDPAKAFATPFYKQGGQVKAIRVEQVQKSGVLVRENNGVADNASIVRTDVF
KNNKFFLVPIYTWQVAKGILPNKAIVAHKNEDEWEEMDEGAKFKFSLFPNDLVE
LKTKKEYFFGYIIGLDRATGNISLKEHDGEISKGKDGVYRVGVKLALSFEKYQV
DELGKNRQICRPQQRQPV

[0313]

[0314] 서열번호 5

[0315] 명칭: S.뮤탄스 카스9

[0316] 서열:

ATGAAAAAACCTTACTCTATTGGACTTGATATTGGAACCAATTCTGTTGGTTG
GGCTGTTGTGACAGATGACTACAAAGTTCCTGCTAAGAAGATGAAGGTTCTG
GGAAATACAGATAAAAGTCATATCGAGAAAAATTTGCTTGGCGCTTTATTAT
TTGATAGCGGGAATACTGCAGAAGACAGACGGTTAAAGAGAACTGCTCGCCG
TCGTTACACACGTCGCAGAAATCGTATTTTATATTTGCAAGAGATTTTTTCAG
AAGAAATGGGCAAGGTAGATGATAGTTTCTTTCATCGTTTAGAGGATTCTTTT
CTTGTTACTGAGGATAAACGAGGAGAGCGCCATCCCATTTTTGGGAATCTTG
AAGAAGAAGTTAAGTATCATGAAAATTTTCCAACCATTTATCATTTGCGGCA

[0317]

ATATCTTGCGGATAATCCAGAAAAAGTTGATTTGCGTTTAGTTTATTTGGCTT
 TGGCACATATAATTAAGTTTAGAGGTCATTTTTTAATTGAAGGAAAGTTTGAT
 ACACGCAATAATGATGTACAAAGACTGTTTCAAGAATTTTAGCAGTCTATG
 ATAATACTTTTGAGAATAGTTCGCTTCAGGAGCAAAATGTTCAAGTTGAAGA
 AATTCTGACTGATAAAATCAGTAAATCTGCTAAGAAAGATAGAGTTTGAAA
 CTTTTTCCTAATGAAAAGTCTAATGGCCGCTTTGCAGAATTTCTAAAACTAAT
 TGTGTTGTAATCAAGCTGATTTTAAAAAGCATTTTGAATTAGAAGAGAAAGCA
 CCATTGCAATTTTCTAAAGATACTTATGAAGAAGAGTTAGAAGTACTATTAGC
 TCAAAATTGGAGATAATTACGCAGAGCTCTTTTTATCAGCAAAGAACTGTAT
 GATAGTATCCTTTTATCAGGGATTTTAACAGTTACTGATGTTGGTACCAAAGC
 GCCTTTATCTGCTTCGATGATTCAGCGATATAATGAACATCAGATGGATTTAG
 CTCAGCTTAAACAATTCATTCGTCAGAAATTATCAGATAAATATAACGAAGTT
 TTTTCTGATGTTTCAAAAGACGGCTATGCGGGTTATATTGATGGGAAAACAA
 ATCAAGAAGCTTTTTATAAATACCTTAAAGGTCTATTAAATAAGATTGAGGG
 AAGTGGCTATTTCTTGATAAAATTGAGCGTGAAGATTTTCTAAGAAAGCAA
 CGTACCTTTGACAATGGCTCTATTCCACATCAGATTCATCTTCAAGAAATGCG
 TGCTATCATTCGTAGACAGGCTGAATTTTATCCGTTTTTAGCAGACAATCAAG
 ATAGGATTGAGAAATTATTGACTTTCCGTATTCCTACTATGTTGGTCCATTA
 GCGCGCGGAAAAAGTGATTTTGCTTGGTTAAGTCGGAAATCGGCTGATAAAA
 TTACACCATGGAATTTTGATGAAATCGTTGATAAAGAATCCTCTGCAGAAGCT
 TTTATCAATCGTATGACAAATTATGATTTGTACTTGCCAAATCAAAAAGTTCT
 TCCTAAACATAGTTTATTATACGAAAAATTTACTGTTTACAATGAATTAACAA
 AGGTAAATATAAAACAGAGCAAGGAAAAACAGCATTTTTTGATGCCAATAT
 GAAGCAAGAAATCTTTGATGGCGTATTTAAGGTTTATCGAAAAGTAACTAAA
 GATAAATTAATGGATTTCTTGAAAAAGAATTTGATGAATTTCTGATTGTTGA
 TTTAACAGGTCTGGATAAAGAAAATAAAGTATTTAACGCTTCTTATGGAACCT
 ATCATGATTTGTGTAAAATTTTAGATAAAGATTTTCTCGATAATTCAAAGAAT
 GAAAAGATTTTAGAAGATATTGTGTTGACCTTAACGTTATTTGAAGATAGAG
 AAATGATTAGAAAACGTCTAGAAAATTACAGTGATTTATTGACCAAAGAACA
 AGTGAAAAAGCTGGAAAGACGTCATTATACTGGTTGGGGAAGATTATCAGCT
 GAGTTAATTCATGGTATTCGCAATAAAGAAAGCAGAAAAACAATTCTTGATT

[0318]

ATCTCATTGATGATGGCAATAGCAATCGGAACCTTTATGCAACTGATTAACGAT
 GATGCTCTTTCTTTCAAAGAAGAGATTGCTAAGGCACAAGTTATTGGAGAAA
 CAGACAATCTAAATCAAGTTGTTAGTGATATTGCTGGCAGCCCTGCTATTTAA
 AAAGGAATTTTACAAAGCTTGAAGATTGTTGATGAGCTTGTCAAAATTATGG
 GACATCAACCTGAAAATATCGTCGTGGAGATGGCGCGTGAAAACAGTTTAC
 CAATCAGGGACGACGAAATTCACAGCAACGTTTGAAAAGTTTGACAGATTCT
 ATTTAAAGAATTTGGAAGTCAAATTCCTTAAAGAACATCCGGTTGAGAATTCAC
 AGTTACAAAATGATAGATTGTTTCTATATTATTTACAAAACGGCAGAGATATG
 TATACTGGAGAAGAATTGGATATTGATTATCTAAGCCAGTATGATATAGACC
 ATATTATCCCGCAAGCTTTTATAAAGGATAATTCTATTGATAATAGAGTATTG
 ACTAGCTCAAAGGAAAATCGTGGAATAATCGGATGATGTACCAAGTAAAGAT
 GTTGTTCGTAAAATGAAATCCTATTGGAGTAAGCTACTTTTCGGCAAAGCTTAT
 TACACAACGTAAATTTGATAATTTGACAAAAGCTGAACGAGGTGGATTGACC
 GACGATGATAAAGCTGGATTTCATCAAGCGTCAATTAGTAGAAACACGACAAA
 TTACCAAACATGTAGCACGTATTCTGGACGAACGATTTAATACAGAAACAGA
 TGAACAACAAGAAAATTCGTCAAGTAAAAATTGTGACCTTGAAATCAAAT
 CTTGTTTCCAATTTCCGTAAAGAGTTTGAACCTCTACAAAGTGCGTGAAATTAA
 TGACTATCATCATGCACATGATGCCTATCTCAATGCTGTAATTGGAAAGGCTT
 TACTAGGTGTTTACCCACAATTGGAACCTGAATTTGTTTATGGTGATTATCCT
 CATTTTCATGGACATAAAGAAAAATAAAGCAACTGCTAAGAAATTTTCTATTC
 AAATATTATGAACCTCTTTAAAAAAGATGATGTCCGTACTGATAAAAAATGGT
 GAAATTATCTGGAAAAAAGATGAGCATATTTCTAATATTAAAAAAGTGCTTT
 CTTATCCACAAGTTAATATTGTAAAGAAAGTAGAGGAGCAAACGGGAGGATT
 TTCTAAAGAATCTATCTTGCCGAAAGGTAATTCTGACAAGCTTATTCCTCGAA
 AAACGAAGAAATTTTATTGGGATACCAAGAAATATGGAGGATTTGATAGCCC
 GATTGTTGCTTATTCTATTTTAGTTATTGCTGATATTGAAAAAGGTAAATCTA
 AAAAATTGAAAACAGTCAAAGCCTTAGTTGGTGTCACTATTATGGAAAAGAT
 GACTTTTGAAAGGGATCCAGTTGCTTTTCTTGAGCGAAAAGGCTATCGAAAT
 GTTCAAGAAGAAAATATTATAAAGTTACCAAAATATAGTTTATTTAACTAG
 AAAACGGACGAAAAAGGCTATTGGCAAGTGCTAGGGAACCTTCAAAGGGAA
 ATGAAATCGTTTTGCCAAATCATTTAGGAACCTTGCTTTATCACGCTAAAAAT

ATTCATAAAGTTGATGAACCAAGCATTGACTATGTTGATAAACATAAAG
 ATGAATTTAAGGAGTTGCTAGATGTTGTGTCAAACCTTTTCTAAAAAATATACT
 TTAGCAGAAGGAAATTTAGAAAAAATCAAAGAATTATATGCACAAAATAATG
 GTGAAGATCTTAAAGAATTAGCAAGTTCATTTATCAACTTATTAACATTTACT
 GCTATAGGAGCACCGGCTACTTTTAAATTCTTTGATAAAAAATATTGATCGAAA
 ACGATATACTTCAACTACTGAAATTCTCAACGCTACCCATCCACCAATCCA
 TCACCGGTCTTTATGAAACGCGGATTGATCTCAATAAGTTAGGAGGAGACTA

A

서열번호 6

[0322]

서열:

MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLF
 DSGNTAEDRRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFHRLSDSFLVT
 EDKRGERHPIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKF
 RGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKS
 AKKDRVLKLPNEKSNGRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEE
 LEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQ
 MDLAQLKQFIRQKLSDKYNEVFSVSKDGYAGYIDGKTNQEAFYKYLKGLLNKI
 EGSYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRQAEFYPFLADNQDR
 IEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITPWNFDEIVDKESSAEAFINRM
 TNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFD
 GVFKVYRKVTDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGTYHDLCKIL
 DKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLNYSDLLTKEQVKKLERRHYT
 GWGRLSAELIHGIRNKESRKTIIDYLDGNSNRNFMQLINDDALSFKEEIAKAQ
 VIGETDNLNQVSDIAGSPAIAKKGILQSLKIVDELVKIMGHQPENIVVEMARENQF
 TNQGRNSQQRLKGLTDSIKEFGSQILKEHPVENSQQLQNDRLFLYYLQNGRDMY
 TGEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKM
 KSYWSKLLSAKLITQRKFDNLTAKERGGLTDDDKAGFIKRQLVETRQITKHVARI
 LDERFNTETDENNKIRQVKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYL
 NAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDV
 RTDKNGEIIWKKDEHISNIKKVLSYPQVNIVKKVEEQTGGFSKESILPKGNSDKLIP
 RKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKM

[0323]

TFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLENGRKRLASARELQKGNEIVL
 PNHLGTLTYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSSKKTAEAGN
 LEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNIDRKRYTSTTEIL
 NATLIHQSIITGLYETRIDLNLGGD

[0324]

서열번호 7

[0325]

명칭: N. 메닌기타이데스 카스9

[0326]

[0327] 서열:

ATGGCTGCCTTCAAACCTAATTCAATCAACTACATCCTCGGCCTCGATATCGG
CATCGCATCCGTCGGCTGGGCGATGGTAGAAATTGACGAAGAAGAAAACCCC
ATCCGCTGATTGATTTGGGCGTGCGCGTATTTGAGCGTGCCGAAGTACCGA
AAACAGGCGACTCCCTTGCCATGGCAAGGCGTTTGGCGCGCAGTGTTGCGCG
CCTGACCCGCCGTGCGGCCCACCGCCTGCTTCGGACCCGCCGCCTATTGAAAC
GCGAAGGCGTATTACAAGCCGCCAATTTTGACGAAAACGGCTTGATTAAATC
CTTACCGAATACACCATGGCAACTTCGCGCAGCCGCATTAGACCGCAAACCTG
ACGCCTTTAGAGTGGTCGGCAGTCTTGTTGCATTTAATCAAACATCGCGGCTA
TTTATCGCAACGGAAAAACGAGGGCGAAACTGCCGATAAGGAGCTTGGCGCT
TTGCTTAAAGGCGTAGCCGGCAATGCCCATGCCTTACAGACAGGCGATTTCC
GCACACCGGCCGAATTGGCTTTAAATAAATTTGAGAAAGAAAGCGGCCATAT
CCGCAATCAGCGCAGCGATTATTCGCATACGTTTCAGCCGCAAAGATTACAG
GCGGAGCTGATTTTGCTGTTTGAAAAACAAAAAGAATTTGGCAATCCGCATG
TTTCAGGCGGCCCTTAAAGAAGGTATTGAAACCCTACTGATGACGCAACGCCC
TGCCCTGTCCGGCGATGCCGTTCAAAAAATGTTGGGGCATTGCACCTTCGAAC
CGGCAGAGCCGAAAGCCGCTAAAAACACCTACACAGCCGAACGTTTCATCTG
GCTGACCAAGCTGAACAACCTGCGTATTTTAGAGCAAGGCAGCGAGCGGCCA
TTGACCGATACCGAACGCGCCACGCTTATGGACGAGCCATACAGAAAATCCA
AACTGACTTACGCACAAGCCCGTAAGCTGCTGGGTTTAGAAGATACCGCCTT
TTTCAAAGGCTTGCGCTATGGTAAAGACAATGCCGAAGCCTCAACATTGATG
GAAATGAAGGCCTACCATGCCATCAGCCGTGCACTGGAAAAAGAAGGATTG
AAAGACAAAAAATCCCATTAACCTTTCTCCCGAATTACAAGACGAAATCG
GCACGGCATTCTCCCTGTTCAAAACCGATGAAGACATTACAGGCCGTCTGAA
AGACCGTATACAGCCCGAAATCTTAGAAGCGCTGTTGAAACACATCAGCTTC

[0328]

GATAAGTTCGTCCAAATTTCTTGAAGCATTGCGCCGAATTGTGCCTCTAAT
GGAACAAGGCAAACGTTACGATGAAGCCTGCGCCGAAATCTACGGAGACCA
TTACGGCAAGAAGAATACGGAAGAAAAGATTTATCTGCCGCCGATTCCCGCC
GACGAAATCCGCAACCCCGTCGTCTTGC GCGCCTTATCTCAAGCACGTAAGG
TCATTAACGGCGTGGTACGCCGTTACGGCTCCCCAGCTCGTATCCATATTGAA
ACTGCAAGGGAAGTAGGTAAATCGTTTAAAGACCGCAAAGAAATTGAGAAA
CGCCAAGAAGAAAACCGCAAAGACCGGGAAAAAGCCGCCGCAAATTCGGA
GAGTATTTCCCAATTTTGTCTGGAGAACCCAAATCCAAAGATATTCTGAAACT
GCGCCTGTACGAGCAACAACACGGCAAATGCCTGTATTGGGGCAAAGAAATC
AACTTAGGCCGTCTGAACGAAAAAGGCTATGTCGAAATCGACCATGCCCTGC
CGTTCTCGCGCACATGGGACGACAGTTTCAACAATAAAGTACTGGTATTGGG
CAGCGAAAACCAAAACAAAGGCAATCAAACCCCTTACGAATACTTCAACGG
CAAAGACAACAGCCGCGAATGGCAGGAATTTAAAGCGCGTGTCGAAACCAG
CCGTTTCCCGCGCAGTAAAAACAACGGATTCTGCTGCAAAAATTCGATGAA
GACGGCTTTAAAGAACGCAATCTGAACGACACGCGCTACGTCAACCGTTTCC
TGTGTCAATTTGTTGCCGACCGTATGCGGCTGACAGGTAAAGGCAAGAAACG
TGTCTTTGCATCCAACGGACAAATTACCAATCTGTTGCGCGGCTTTTGGGGAT
TGCGCAAAGTGCGTGCGGAAAACGACCGCCATCACGCCTTGACGCGCGTCGT
CGTTGCCCTGCTCGACCGTTGCCATGCAGCAGAAAATTACCCGTTTTGTACGCT
ATAAGAGATGAACGCGTTTGACGGTAAACCATAGACAAAGAAACAGGAG
AAGTGCTGCATCAAAAAACACACTTCCCACAACCTTGGAATTTTTCGCACA
AGAAGTCATGATTGCGCTTTCGGCAAACCGGACGGCAAACCCGAATTCGAA
GAAGCCGATACCCTAGAAAAACTGCGCACGTTGCTTGCCGAAAAATTATCAT
CTCGCCCCGAAGCCGTACACGAATACGTTACGCCACTGTTTGTTCACGCGCG
CCCAATCGGAAGATGAGCGGGCAAGGGCATATGGAGACCGTCAAATCCGCC
AAACGACTGGACGAAGGCGTCAGCGTGTTGCGCGTACCGCTGACACAGTTAA
AACTGAAAGACTTGGAaaaaatGGTCAATCGGGAGCGCGAACCTAAGCTAT
ACGAAGCACTGAAAGCACGGCTGGAAGCACATAAAGACGATCCTGCCAAAG
CCTTTGCCGAGCCGTTTTACAAATACGATAAAGCAGGCAACCGCACCCAACA
GGTAAAAGCCGTACGCGTAGAGCAAGTACAGAAAACCGGCGTATGGGTGCG
CAACCATAACGGTATTGCCGACAACGCAACCATGGTGCGCGTAGATGTGTTT

GAGAAAGGCGACAAGTATTATCTGGTACCGATTTACAGTTGGCAGGTAGCGA
AAGGGATTTTGCCGGATAGGGCTGTTGTACAAGGAAAAGATGAAGAAGATTG
GCAACTTATTGATGATAGTTTCAACTTTAAATTCTCATTACACCCTAATGATTT
AGTCGAGGTTATAACAAAAAAGCTAGAATGTTTGGTTACTTTGCCAGCTGC
CATCGAGGCACAGGTAATATCAATATACGCATTCATGATCTTGATCATAAAA
TTGGCAAAAATGGAATACTGGAAGGTATCGGCGTCAAAACCGCCCTTTCATT
CCAAAAATACCAAATTGACGAACTGGGCAAAGAAATCAGACCATGCCGTCTG
AAAAAACGCCCCGCTGTCCGTAA

서열번호 8

[0332] 서열:

MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGD
SLAMARRLARSVRRLTRRRARLLRTRRLKREGVLQAANFDENGLIKSLPNT
WQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVA
GNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYSHTFSRKDLQAEILLFEK
QKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAPKAAKNTY
TAERFIWLTCLNNLRILEQGSRPLTDERATLMDEPYRKSCLTYAQARKLLGLE
DTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDE
IGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKQVQISLKALRRIVPLMEQKG
RYDEACAEIYGDHYGKKNTTEKIYLPPIPADEIRNPVLRALSQARKVINGVVR
YGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGE
PKSKDILKLRLYEQQHGKCLYSGKEINLGRLEKGYVEIDHALPFSRTWDDSFNNK
VLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKF
DEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFW
GLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETG
EVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRP
EAVHEYVTPLFVSRAPNRKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLD
LEKMNVREREPKLYEALKARLEAHKDDPAKAFAPFYKYDKAGNRTQQVKAV
RVEQVQKTGVWVRNHNGIADNATMVRVDVFEGDKYYLVPIYSWQVAKGILP
DRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRTG
NINIRIHDLDHKIGKNGILEGIGVKTALSFQKYOIDELGKEIRPCRLKKRPPVR

[0333]

[0334] 서열번호 9

[0335] 서열:

ATGAACAATAACAATTACTCTATCGGACTCGATATCGGAACAAACAGCGTCG
GATGGGCCGTCATTACGGATGACTATAAGGTGCCATCGAAAAAGATGAAAGT
TCTAGGCAATACAGATAAACACTTTATCAAGAAAAATCTAATTGGAGCTTTA
TTATTTGATGAAGGAGCTACTGCTGAAGATAGACGTTTCAAACGAACAGCAC
GCCGTCGCTATACTCGTCGAAAAAATCGTCTTCGCTATCTTCAAGAAATCTTT
TCTGAGGAAATGAGCAAAGTGGATAGTAGTTTCTTTCATCGATTAGATGACTC
ATTCTTAGTTCCGTGAGGATAAAAGAGGAAGTAAATATCCTATTTTTGCTACCT
TGGCAGAAGAAAAAGAATATCACAAGAAATTTCCAACCTATCTATCATTTGAG
AAAACACCTTGCGGACTCAAAAGAAAAAACTGACTTGCGCTTGATCTATCTA
GCATTAGCGCATATGATTAAATACCGCGGACATTTTTTGTATGAAGAATCTTT
CGATATTA AAAACAATGATATCCAAAAAATCTTTAGCGAGTTTATAAGCATT
ACGACAACACCTTTGAAGGAAGTTCACCTTAGTGGACAAAATGCACAAGTAGA
AGCAATTTTTACTGATAAAATTAGTAAATCTGCTAAGAGAGAACGCATTCTA
AAACTCTTTGCTTATGAAAAATCCACTGATCTATTTTCAGAATTTCTCAAGCT
GATTGTAGGAAATCAAGCTGATTTTAAGAAACACTTTGACTTGGAAGAAAA
GCTCCACTACAATTCTCTAAAGATACCTATGATGAGGATTTGGAAAACTTACT
CGGACAAATTGGAGATGACTTTGCAGACCTTTTCCTAGTTGCTAAAAAACTCT
ATGATGCCATTCTTTTATCAGGAATCTTAACTGTTACAGATTCTTCAACTAAG
GCCCCACTATCAGCATCTATGATTGAGCGCTATGAAAACCACCAAAAAGACT
TAGCGGCTTTAAAAACAATTCATCCAAAACAATCTTCAAGAAAAATATGATGA
AGTTTTCTCTGACCAATCTAAAGATGGGTATGCTAGGTATATCAATGGCAAA
ACCACTCAAGAAGCATTTTACAAGTACATCAAAAATCTTCTCTCTAAATTCGA
AGGATCAGATTATTTCTTGATAAAATTGAACGTGAAGATTTCTTGAGAAAA
CAACGCACCTTTGATAATGGTTCTATCCCTCATCAAATTCATCTTCAAGAAAT
GAATGCCATTATCCGTCGGCAAGGAGAACATTATCCATTTCTGAAGGAATAT
AAAGAAAAGATAGAGACAATCTTGACTTTCCGTATTCCTTATTATGTTGGCCC
ATTGGCTCGTGGAATCGTAATTTTGCTTGCTTACTCGAAACTCTGACCAAG
CAATCCGACCTTGGAATTTGAAGAAATTGTTGATCAAGCAAGCTCTGCGGA
AGAATTCATCAATAAGATGACTAACTATGACTTGTATCTGCCAGAGGAAAAA
GTTTTGCCCAAGCATAGTCTCTTGTATGAAACATTTGCTGTCTACAATGAATT

[0336]

AACAAAAGTAAAATTTATTTTCAGAGGGATTGAGAGACTATCAATTCCTTGAT
AGTGGGCAAAAGAAGCAAATTGTCAATCAATTATTCAAAGAGAAAAGAAAA
GTAAGTAAAAAGACATCATTCAAGTATCTACACAATGTTGATGGCTACGATG
GAATCGAACTAAAAGGAATTGAAAAACAATTTAACGCTAGTCTTTCTACTTA
TCATGATTTACTCAAATAATCAAGGATAAAGAGTTTATGGATGATCCTAAA
AATGAAGAGATTCTTGAAAATATCGTCCACACACTAACTATCTTTGAAGATC
GTGAGATGATCAAGCAACGCCCTTGCTCAATATGCCTCTATCTTTGATAAAAAA
GTGATCAAGGCACTGACTCGTCGACATTATACTGGTTGGGGAAAACCTCTCTG
CTAAGCTAATCAACGGTATCTGTGATAAAAAAACTGGTAAAACAATTCTTGA
CTACTTGATTGATGACGGCTACAGCAATCGTAACTTTATGCAGTTAATCAATG
ATGACGGGCTTTCTTCAAAGATATTATTCAAAGCACAAGTGGTTGGTAA
GACAAACGATGTGAAGCAAGTTGTCCAAGAACTCCCAGGTAGTCCTGCTATT
AAAAAGGGAATTTTACAAAGTATCAAGCTTGTGCGATGAGCTTGTCAAAGTTA
TGGGCCATGCTCCCGAGTCCATTGTGATTGAAATTGCACGAGAAAATCAGAC
AACTGCCAGAGGGAAAAAGAATTCTCAACAAAGATATAAGCGCATTGAAGA
TGCACTAAAAAATTTAGCACCTGGGCTTGATTCAAATATATTAAGAAGACAT
CCAACAGATAATATTCAACTTCAAATGACCGTCTCTTCTTTACTATCTCCA
AAATGGGAAGGATATGTACACTGGAGAAGCTCTTGATATCAACCAACTGAGC
AGCTATGACATTGACCACATCGTCCCACAGGCCTTTATCAAGGATGATTCTCT
TGATAACCGTGTCTTGACTAGTTCAAAGGATAATCGTGGGAAATCCGATAAT
GTTCCAAGTTTAGAAGTCGTTCAAAAAAGAAAAGCTTTTTGGCAACAATTAC
TAGATTCCAAATTGATTTCAGAACATAAATTTAATAATTTAACCAAGGCTGAA
CGTGGTGGGCTAGATGAGCGAGATAAAGTTGGCTTTATCAGACGCCAACTAG
TTGAAACACGGCAAATCACAAAACATGTTGCTCAGATTTTGGATGCCCGTTTT
AATACAGAAGTGAATGAGAAAGATAAGAAGAACCGTACCGTCAAATTATC
ACTTTGAAATCCAATCTAGTTTCCAACCTCCGTAAAGAATTTAAGTTATATAA
GGTACGCGAAATCAATGACTACCACCATGCACATGATGCCTATTTAAATGCA
GTGGTGGCTAAGGCTATCCTTAAGAAATATCCTAACTAGAGCCTGAATTCG
TCTATGGTGACTATCAAAGTACGATATTAAGAGATATATTTCCAGATCCAA
AGATCCTAAAGAAGTTGAAAAAGCAACTGAAAAGTATTTCTTCTACTCAAAC
TTGTTGAACCTCTTTAAAGAAGAGGTGCATTACGCAGACGGAACCATCGTAA

[0337]

AACGAGAGAATATCGAATACTCTAAGGACACTGGAGAAATCGCTTGGAATAA
 AGAAAAAGATTTTCGCTACAATTAAGTTCTTTCACTTCCGCAGGTGAAT
 ATTGTGAAGAAAACAGAGATTCAAACACATGGTCTAGATAGAGGTAAACCTA
 GAGGATTGTTCAATTCCAATCCATCTCCTAAACCTTCAGAAGATCGTAAAGA
 AAACCTTGTTCCCAATTAAACAAGGGCTTGACCCACGAAAATACGGTGGTTAC
 GCTGGTATTTCTAACTCATAACGCGGTCTTAGTTAAAGCTATTATTGAAAAAGG
 AGCGAAAAACAACAAAAGACCGTTCTTGAATTTCAAGGTATCTCTATTTTA
 GATAAAATAAATTTTGAAAAGAACAAAGAAACTATCTTCTTGAAAAAGGAT
 ACATAAAATTTCTATCAACTATTACTTTACCTAAATATAGTTTGTGAGTTTC
 CTGATGGTACAAGAAGAAGACTAGCAAGTATTCTATCGACAAACAATAAACG
 AGGAGAAATTCATAAAGGTAATGAATTGGTCATCCCTGAAAAGTATACGACT
 CTTTTGTATCATGCTAAGAATATTAATAAAACACTTGAACCAGAACACTTAGA
 GTATGTTGAGAAACATCGAAATGATTTTGCTAAACTTTTAGAATATGTACTTA
 ACTTTAACGATAAGTATGTAGGCGCATTAAAAAATGGAGAAAGAATCAGACA
 AGCATTATTGATTGGGAAACAGTTGATATTGAAAAGTTATGTTTCAGTTTCA
 TTGGTCCAAGAAATAGTAAAAATGCTGGTTTATTCGAGTTAACTTCACAAGG
 AAGTGCTTCTGACTTCGAGTTCTTGGGAGTAAAAATCCACGATACAGAGAC
 TATACACCTTCGTCCTCAACGCCACCCTCATCCACCAATCCATCACTGG
 TCTTTACGAGACTCGGATTGACTTAAGCAAACCTGGGAGAAGACTGA

[0338]

[0339]

서열번호 10

[0340]

명칭: gi|777888062|gb|KJQ69483.1| 크리스퍼-회합된 엔도뉴클레아제 카스9 [스트렙토코커스 미티스]

[0341]

서열:

MNNNNYSIGLDIGTNSVGWAVITDDYKVPSKKMKVLGNTDKHFIKKNLIGALLF
 DEGATAEDRRFKRTARRRYTRRKNRLRYLQEIFSEEMSKVDSSFFHRLDDSFLVP
 EDKRGSKYPIFATLAEEKEYHKKFPTIYHLRKHLADSKEKTDLRLIYLALAHMIK
 YRGHFLYEESFDIKNNDIQKIFSEFISIYDNTFEGSSLGQNAQVEAIFDKISAK
 RERILKLFAYEKSTDLFSEFLKLIVGNQADFKKHFDLEEKAPLQFSKDTYDEDLEN
 LLGQIGDDFADFLVAKKLYDAILLSGILTVTDSSTKAPLSASMIERYENHQDLA
 ALKQFIQNNLQEKYDEVFSDQSKDGYARYINGKTTQEAFYKYIKNLLSKFEGSD
 YFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMNAIIRRQGEHYPFLKEYKEKIEITL

[0342]

TFRIPYYVGPLARGNRNFAWLTRNSDQAIRPWNFEEIVDQASSAEFINKMTNYD
 LYLPEEKVLPKHSLLYETFAVYNELTKVKFISEGLRDYQFLDSGQKKQIVNQLFK
 EKRKVTEKDIIQYLHNVDGYDGIELKGIEKQFNASLSTYHDLLKIIKDKEFMDDP
 KNEEILENIVHTLTIFEDREMIKQRLAQYASIFDKKVIKALTRRHYTGWGKLSAKL
 INGICDKKTGKTILDYLDGYSNRNFMQLINDDGLSFKDIIQKAQVVGKTNDVK
 QVVQELPGSPAIIKKGILQSIKLVDELVKVMGHAPESIVIEIARENQTTARGKKNSQ
 QRYKRIEDALKNLAPGLDSNILKEHPTDNIQLQNDRLFLYYLQNGKDMYTGEAL
 DINQLSSYDIDHIVPQAFIKDDSLDNRVLTSSKDNRGKSDNVPSLEVQKRKAFW
 QQLLDSKLISEHKFNNTKAERGGLDERDKVGFIRRLVETRQITKHVAQILDAR
 FNTEVNEKDKKNRVTVKIITLKSNLVSNFRKEFKLYKVREINDYHHAHDAYLNAV
 VAKAILKKYPKLEPEFVYGDYQKYDIKRYISRSKDPKEVEKATEKYFFYSNLLNF
 FKEEVHYADGTIVKRENIEYSKDTGEIAWNKEKDFATIKKVLSPQVNIVKKTEIQ
 THGLDRGKPRGLFNSNPSPKPSDRKENLVPIKQGLDPRKYGGYAGISNSYAVLV
 KAIIEKGAKKQQKTVLEFQGISILDKINFENKENYLLKGYIKILSTITLPKYSLFE
 FPDGTRRRRLASILSTNNKRGEIHKGNELVIPEKYTTLLYHAKNINKTLEPEHLEYV
 EKHRNDFAKLLEYVLNFNDKYVGALKNGERIRQAFIDWETVDIEKLCFSFIGPRN
 SKNAGLFELTSQGSASDFELGVKIPRYRDYTPSSLLNATLIHQSTGLYETRIDL
 S
 KLGED

[0343]

[0344]

서열번호 11

[0345]

서열:

ATGACAAAACCTTATTCTATTGGACTTGATATTGGGACTAACTCTGTTGGTTG
 GGCTGTTGTGACAGATGGCTACAAAGTTCCTGCTAAGAAGATGAAGGTTCTG
 GGAAATACAGATAAAAGCCATATCAAGAAAAATTTACTTGGAGCTTTATTGT
 TTGATAGCGGTAATACTGCAAAAGACAGACGTTTGAAGCGGACAGCTAGGCG
 TCGATATACACGTCGTAGAAACCGTATTTTATATTTGCAGGAAATTTTGCTG
 AAGAAATGGCTAAAGCAGACGAAAGTTTCTTCCAGCGCTTAAACGAATCGTT
 TTTAACAATGATGACAAAGAATTTGATTCTCATCCAATCTTTGGGAATAAAG
 CTGAAGAGGAGGCTCATCACCATAAATTTCCAACAATTTTTCATTTGCGAAAG
 CATTTAGCAGACTCAACCGAGAAATCTGATTTGCGCTTAATTTATCTAGCTTT
 AGCGCATATGATTAAATTCCGGGGACATTTCTTAATTGAAGGTCAGCTAAAA
 GCTGAAAATACAAATGTTCAAACATTATTTGACGATTTTGTAGAAGTATATGA

[0346]

TAAGACAGTTGAAGAAAGTCATTTATCAGAAATTAGTGTCTCCAGTATTCTGA
CAGAAAAAATTAGTAAATCGCGTCGCTTAGAAAAATCTTATAAAATACTATCC
CACTGAGAAGAAAAACACTCTCTTCGGAAATCTTATCGCCTTGTCTTTAGGAT
TACAGCCAAACTTTAAAACAAATTTTAAATTATCCGAAGATGCTAAACTACA
GTTTTCTAAGGATACTTATGAAGAAGATTTAGGAGAATTACTTGGA AAAATC
GGAGATAATTATGCAGATTTATTTATATCAGCTAAAAATCTTTATGATGCTAT
TTTGCTATCAGGAATTTTAAACAATAGATGACAACACGACAAAAGGCTCCGTTG
TCTGCTTCAATGATTAAACGTTATGAGGAACATCAGGAAGATTTAGCACAAC
TTAAGAAATTTATCCGTCAGAATTTACCAGATCAATATAGTGAGGTTTTTCT
GATAAAACAAAGGATGGCTATGCTGGTTATATTGATGGAAAAACGAATCAGG
AGGCCTTTTATAAATACATCAAAAATATGCTGTCAAAAACAGAAGGTGCAGA
TTATTTTCTTGACAAAATTGATCGTGAAGACTTTTGTAGAAAACAGAGAACGT
TTGATAATGGTTCCGTTCCGCATCAGATTCATCTGCAAGAGATGCATGCTATT
TTACGACGTCAGGGTGAATACTATCCATTCTTGAAAAGAAAATCAGGATAAAA
TTGAAAAAATCTTAACGTTTAGAATTCCTTACTACGTTGGTCCTTTGGCGCGA
AAAGGTAGCCGCTTTCCTGGGCAGAATACAAGGCGGATAAAAAAGTTACGC
CATGGAATTTTGATGATATTCTTGATAAAGAAAAATCAGCAGAAGAATTCAT
CACACGCATGACTTTAAATGATTTGTATTTACCTGAAGAAAAAGTCTTACCAA
AGCATAGTCTTGTTTATGAAACGTTTAATGTTTACAATGAGTTAACTAAAGTT
AAGTATGTCAATGAGCAAGGGAAAGCCATTTTCTTTGATGCCAATATGAAGC
AAGAGATTTTGTATCATGTTTTTAAAGAAAATCGGAAAGTTACTAAAGATAA
ACTTTTAAATTATTTGAATAAAGAGTTTGAAGAATTTAGAATTGTTAACTTAA
CTGGACTGGATAAGGAAAAATAAGCCTTTAATTCCAGTCTTGGAACCTATCA
TGATTTGCGTAAAATTTTAGATAAATCATTCTTAGATGATAAAGTAAATGAAA
AGATAATTGAGGATATCATTCAAACACTAACTCTGTTTGAAGACAGAGAAAT
GATTCGTCAGCGTCTTCAAAAGTATAGTGATATTTTACAACACAGCAATTGA
AAAAACTTGAACGCCGTCATTATACAGGTTGGGGAAGATTATCAGCGAAGTT
AATCAATGGTATTCGAGATAAACAGAGTAATAAGACTATCTGGGTTATTG
ATTGATGATGGTTATAGCAATCGTAACTTTATGCAGTTGATTAATGACGATTC
TCTTCCTTTTAAAGAAGAAATTGCTAGGGCACAAAGTCATTGGAGAAACAGAT
GACTTAAATCAACTTGTTAGTGATATTGCTGGCAGTCCTGCTATTAAAAAGGG

[0347]

AATTTTACAAAGTCTGAAAATTGTAGATGAGCTTGTTAAAGTCATGGGGCAT
AATCCTGCTAACATTGTTATCGAAATGGCGCGTGAAAATCAGACTACAGCCA
AAGGGCGTCGCAGTTCACAGCAACGTTATAAACGACTTGAGGAGGCAATAAA
AAATCTTGACCATGATTTAAATCATAAGATTTTAAAAGAACACCCAACAGAT
AATCAAGCTTTACAGAATGACCGTCTTTTCTTATATTATCTCCAAAATGGCCG
AGATATGTATACTGAAGATCCACTTGATATTAATCGTTTAAAGTGATTATGATA
TCGACCATATTATTCCACAATCTTTTATAAAAGATGACTCTATTGACAATAAG
GTTCTGGTTTCATCAGCTAAAAACCGTGGGAAATCGGATAATGTACCGAGTG
AAGATGTTGTCAATAGGATGAGACCGTTTTTGAATAAATTATTGAGCTGTGG
ATTGATTTCTCAACGGAAATACAGCAATCTAACCAAAAAAGAATTAAAACCA
GATGATAAGGCTGGTTTCATCAAACGTCAATTGGTTGAGACAAGACAAATTA
CAAAGCATGTTGCACAAATTTTAGACGCTCGTTTTAATACAAAACGTGATGA
AAATAAAAAAGTAATTCGTGATGTCAAATTATCACTTTAAAATCTAATTTAG
TTTCACAATTTTCGTAAAGACTTTAAATTTTACAAAGTACGTGAGATTAATGAT
TACCATCATGCGCATGACGCTTATCTTAATGCAGTTATAGGAAAAGCTTTATT
AGATGTTTATCCGCAGTTAGAGCCCGAATTTGTTTATGGTGAGTACCCTCATT
TTCATGGATATAAAGAAAATAAAGCAACTGCTAAGAAATTTTTCTATTCAA
TATTATGAATTTTTTTAAGAAAGATGATATCCGTACCGATGAAAATGGTGAG
ATTGTTTGGAAAAAAGATGAGCATATTTCTAATATTAAAAGGGTGCTTTCCTA
TCCCCAAGTTAATATTGTTAAGAAAGTAGAAATACAGACTGTTGGACAAAAT
GGGGGACTTTTTGACGATAATCCTAAATCACCATTAGAGGTTACACCTAGTA
AACTTGTTCCACTAAAAAAAGAATTAAACCCTAAAAAATATGGAGGATATCA
AAAACCGACGACAGCTTATCCTGTTTTACTGATAACAGATACTAAACAGCTA
ATTCCAATCTCAGTAATGAATAAGAAGCAATTTGAACAAAATCCGGTTAAAT
TTTTAAGAGATAGAGGCTATCAACAGGTAGGAAAGAATGACTTTATTAAATT
ACCCAAATATACCCTAGTTGATATCGGTGATGGGATTAAACGCCTATGGGCT
AGTTCGAAAGAAATACATAAAGGAAATCAATTAGTTGTATCTAAAAAATCTC
AAATTTTGCTTTATCATGCACATCACTTAGATAGTGATTTGAGTAATGATTAT
CTTCAAAATCATAATCAACAATTCGATGTTTTATTTAATGAAATTATTTCTTT
TCTAAAAAATGTAAATTGGGAAAAGAACATATTCAGAAAATTGAAAATGTTT
ACTCCAATAAGAAGAATAGTGCATCAATAGAAGAATTAGCAGAGAGTTTTAT

[0348]

TAAATTATTAGGATTTACACAATTAGGTGCAACTTCCCCATTAAATTTTTTAG
GGGTAAACTAAATCAAAAACAATATAAAGGTAAAAAAGATTATATTTTACC
GTGTACAGAGGGGACCCTTATCCGCCAATCTATCACTGGTCTTTACGAAACAC
GAGTTGATCTTAGTAAATAGGAGAAGACTAA

[0349]

[0350] 서열번호 12

[0350]

[0351] 명칭: gi|357584860|gb|EHJ52063.1| 크리스퍼-회합된 단백질 카스9/Csn1, 하위유형 II/NMEMI [스트렙토코커스 마카카에 NCTC 11558]

[0352] 서열:

MTKPYSIGLDIGTNSVGWAVVTDGYKVPAAKMKVLGNTDKSHIKKNLLGALLF
 DSGNTAKDRRLKRTARRRYTRRRNRILYLQEIFAEMAKADESFFQRLNESFLT
 DDKEFDSHPIFGNKAEEEEAHHKFPTIFHLRKHLADSTEKSDLRLIYLALAHMIK
 RGHFLIEGQLKAENTNVQTLFDDFVEVYDKTVEESHLSEISVSSILTEKISKSRRL
 NLIKYYPTKKNLTLFGNLIALSLGLQPNFKTNFKLSEDAKLQFSKDTYEEDLGELL
 GKIGDNYADLFISAKNLYDAILLSGILTIDDNTTKAPLSASMIKRYEEHQEDLAQL
 KKFIRQNLDPQYSEVFSKTKDGYAGYIDGKTNQEAFYKYIKNMLSKEGADYF
 LDKIDREDFLRKQRTFDNGSVPHQIHLQEMHAILRRQGEYYPFLKENQDKIEKILT
 FRIPYYVGPLARKGSRFAWAEYKADKKVTPWNFDDILDKEKSAAEFITRMTLND
 LYLPEEKVLPKHS�VYETFNVYNELTKVKYVNEQGKAIFFDANMKQEIFDHVFK
 ENRKVTKDKLLNYLNKEFEEFRIVNLTLGLDKENKAFNSSLGTYHDLRKILDKSFL
 DDKVNEKIIEDIIQTLTLFEDREMIRQRLQKYSDFTTQQLKKLERRHYTGWGRLS
 AKLINGIRDKQSNKTILGYLIDDGYSNRNFMQLINDDSLFPKKEIARAQVIGETDD
 LNQLVSDIAGSPAIAKKGILQSLKIVDELVKVMGHNPNANIVIAMARENQTTAKGRR
 SSQQRKYRLEEAIKNLDHDLNHKILKEHPTDNQALQNDRLFLYYLQNGRDMYTE
 DPLDINRLSDYDIDHIIPQSFIKDDSIDNKVLVSSAKNRGKSDNVPSEDVVNRMRP
 FWNKLLSCGLISQRKYSNLTKKELKPDDKAGFIKRQLVETRQITKHVAQILDARF
 NTKRDENKKVIRDVKIITLKSNLVSQFRKDFKFYKREINDYHHAHDAYLNAVIG
 KALLDVYPQLEPEFVYGEYPHFHGYKENKATAKKFFYSNIMNFFKKDDIRTDEN
 GEIVWKKDEHISNIKRVLSYPQVNIVKKVEIQTVGQNGGLFDDNPKSPLEVT
 PSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRD
 RGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSCKSQILLYHA
 HHLSDSLSDNYLQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASI
 EELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGL
 YETRVDLISKIGED

[0353]

[0354]

[0355] 서열번호 13

[0356] 서열:

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGAT
GGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCT
GGGAAATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGGCTCTTTTA
TTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTA
GAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCA
AATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTT
TTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATA
GTAGATGAAGTTGCTTATCATGAGAAATATCCAACCTATCTATCATCTGCGAAA
AAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCT
TAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAAT
CCTGATAATAGTGATGTGGACAACTATTTATCCAGTTGGTACAAACCTACA
ATCAATTATTTGAAGAAAACCTATTAACGCAAGTGGAGTAGATGCTAAAGC
GATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC
AGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCA
TTGGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAA
ATTACAGCTTTCAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGC
AAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGAT
GCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCC
CCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTC
TTTTAAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTT
TTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCC
AAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTAC
TGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGG
ACCTTTGACAACGGCTCTATTCCTCATCAAATTCACCTGGGTGAGCTGCATGC
TATTTTGAGAAGACAAGAAGACTTTTATCCATTTTAAAAGACAATCGTGAG
AAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGC
GCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATT

[0357]

ACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATT
 TATTGAACGCATGACAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTA
 CCAAAACATAGTTTGCTTTATGAGTATTTACGGTTTATAACGAATTGACAAA
 GGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAA
 CAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCG
 TTAAGCAATTAAGAAGATTTATTTCAAAAAAATAGAATGTTTTGATAGTGT
 TGAAATTTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATG
 ATTTGCTAAAAATTATTAAGATAAAGATTTTTTGGATAATGAAGAAAATGA
 AGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGA
 TGATTGAGGAAAAGACTTAAACATATGCTCACCTCTTTGATGATAAGGTGAT
 GAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAA
 TTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTT
 GAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATA
 GTTTGACATTTAAAGAAGACATTCAAAAAGCACAAAGTGTCTGGACAAGGCGA
 TAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTA AAAAAG
 GTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCG
 GCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACCT
 CAAAAGGGCCAGAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGT
 ATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTC
 AATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACAT
 GTATGTGGACCAAGAATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATC
 ACATTGTTCCACAAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTA
 ACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAG
 TAGTCAAAAAGATGAAAAACTATTGGAGACAACCTTCTAAACGCCAAGTTAAT
 CACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGT
 GAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAA
 TCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGA
 TGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAA
 TTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAA
 CAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAAGTGCTT
 TGATTAAGAAATATCCAAAACCTGAATCGGAGTTTGTCTATGGTGATTATAAA

[0358]

GTTTATGATGTTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAG
CAACCGCAAAATATTTCTTTTACTCTAATATCATGAACCTTCTTCAAAACAGAA
ATTACACTTGCAAATGGAGAGATTTCGCAAACGCCCTCTAATCGAAACTAATG
GGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCG
CAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAG
ACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGC
TTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAG
TCCAACGGTAGCTTATTCAGTCCTAGTGTTGCTAAGGTGGAAAAAGGGAAA
TCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAA
GAAGTTCCTTTGAAAAAATCCGATTGACTTTTTTAGAAGCTAAAGGATATAA
GGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGT
TAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAG
GAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTAGCTAGT
CATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGT
TTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGA
ATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTG
CATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTA
TTCATTTATTTACGTTGACGAATCTTGAGCTCCCGCTGCTTTTAAATATTTTG
ATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGC
CACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGA
GTCAGCTAGGAGGTGACTGA

[0359]

[0360]

서열번호 14

[0361]

명칭: gi|409693032|gb|AFV37892.1| 크리스퍼-회합된 단백질, Csn1 군 [스트렙토코커스 파이오제네스 A20]

[0362]

서열:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEED
KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
GHFLIEGDLNPDNSVDKLFQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDL

[0363]

TLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKIL
TFRIPYYVGPLARGNSRFawMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRL
SRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHHPENIVIAMARENQTTQK
GQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMK
NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYL
NAVVGTAIIKKYPKLESEFVYGDYKVDYVRKMIKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTE
VQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGK
SKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKDLIIKLPKYSLELENGR
KRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVEQHK
HYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGD

[0364]

[0365]

서열번호 15

[0366]

명칭: gi|150381361|gb|EF472760.1| HIV-1 클론 39B(미국 인테그라아제 (pol) 유전자 유래), 일부 cds

[0367]

서열:

TTTTTGGATGGAATAGATAGGGCCCAAGAAGAGCATGAGAAATATCACAATA
ATTGGAGAGCAATGGCTAGTGATTTTAACCTGCCACCTNTAGTAGCAAAGGA
GATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGA
CAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATNTAGAAG
GAAAAGTTATCCTGGTAGCAGTNCATGTAGCCAGTGGTTATATAGAAGCAGA
AGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATACTTCCTCTTAAAATTA
GCAGGAAGATGGCCAGTAAAAACAGTACATACAGACAATGGCAGCAACTTC
ACCAGTGCTGCGNTGAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAAT

[0368]

TTGGCATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAGTCTATGAATAA
TGAATTAAGAAAATTGTAGGACAAGTAAGAGATCAGGCTGAGCATCTCAAG
ACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGA
TTGGGGGGTACAGTGCAGGAGAAAGAATAGTAGACATAATAGCCACAGACA
TACAACTAAAGAACTACAAAAAATATTACAAAAATGCAAAATTTTCGGGT
CTATTTTCAGAGACAGCAGAGATCCACTTTGGAAAGGACCAGCAAAGCTTCTC
TGGAAAGGTGAAGGGGCAGTAGTAATACAAGATACCAATGACATAAARGTA
GTGCCARGAAGAAAAGCAAAGATCATAGAGATTATGGAACACAGATGGCA
GGTGATGATTGTGTGGCAAGTAGACAGGNTGAGGATTAG

[0369]

[0370]

서열번호 16

[0371]

명칭: gi|150381362|gb|ABR68182.1| 인테그라아제, 일부 [인간 면역결핍 바이러스 1]

- [0372] 서열:
FLDGIDRAQEEHEKYHNNWRAMASDFNLPPXVAKEIVASCDKCQLKGEAMHGQ
VDCSPGIWQLDCTXHEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGR
WPVKTVHTDNGSNFTSAAXKAACWWAGIKQEFGIPYNPQSQGVVESMNNELKK
IVGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQK
NITKMQNFRVYFRDSRDPLWKGPALKLLWKGEHAVVIQDTNDIKVVPXRKAKIIR
DYGKQMAGDDCVASRQXED
- [0373]
- [0374] 서열번호 17
- [0375] 명칭: gi|459980|gb|L20651.1|STLKIAPOL 유인원 T-세포 림프친화 바이러스 유형 I 인테그라아제 (pol) 유전자, 일부 cds
- [0376] 서열:
GACTTGTAGAACGCTCTAATGGCATTCTTAAAACCTATTATATAAGTACTTT
ACTGACAAACCCGACCT
ACCTATGGATAATGCTCTATCCATAGCCCTATGGACGATCAACCACCTGAATG
TGTTAACCCACTGCCAC
- [0377]
- [0378] 서열번호 18
- [0379] 명칭: gi|459981|gb|AAA47841.1| 인테그라아제, 일부 [유인원 T-림프친화 바이러스 1]
- [0380] 서열:
LVERSNGILKTLLYKYFTDKPDLPMDNALSIALWTINHLNVLTHCH
- [0381]
- [0382] 서열번호 19
- [0383] 명칭: gi|321156784:1-1509 스트렙토코커스 뉴모니에 통합성 및 접합성 요소 ICESpn11930, 균주 11930

[0384]

서열:

GAGTTTTTTTCCTTTCGTAGCAAGGGTTTAGAGCCCCTATTTTATTTTACTATT
GTCTAAACACCAAGCG
AACACCAAAACTACCATGCAATGGAAAAACCTCTGATTTGATTCTCACTTGAT
TTCACAATCTTTATATC
AAACTGTGGGTGGTATTTGACAATATCTTTTTTGATTTTAATAGTAAATTCTG
AAATAATATTTTATAGGT
GAGTAACGTGGACTAAGATGTAACAAGTCTTTGAACTCATCGACACTTAATT
CTACTTTATTGCTATTAT
CACTAGTTTCAATGAATTTTCAATTATTCTGGAATATTTACAGGTATAACTTT
TCAATTCTTCAAATG
GAAATTGTGATTTTCTACAAATTGATTTAAGGCTTTTACAGTATTTTCTTGTGA
ACGATTTATATTATGT
GTATAGCCCATTGTTGTCTCAAAGTTAGCGTGTCTACTCTAGTCATAATATC
TTTCACTGCTATGTGCA
TCTCATTACTTTGAAGGTAACATAATATGCATATGCCTAAACGAATGGGGAGT
AACATGTTTTACCCACTT
AAAACCATAGTCACTTAAACAATTTGTCAATAATTTTCTTCTATTTCGTTTCA
AAATTTGACGAAAAGTG
CTTGATGTTATTGGAGAGCCGTATTCTGTTCTAAATACACTTTCAGAATGTGT
AAAAGCAGGACAGGGAT
GTTTCTCCATATAAGCATCAAACCTCTTTATTTCTCTGTATTGTCCTTTTAATAG
CTTCGCTTGCAGCTTC
AGGCAAAGCTACTTCTCTAATTGAATTGAGTGTTTTAGTTGTATCAAAATGAA
ATTGTTTAACTTTTAAA
CAATGATATTGAAGTGCTTTATCAATATGCAAGATTCCTTTTTTCAAATCAAT
ATCTGATGGTAAAAATG

[0385]

CTGCTTCACTAATTCGAATACCTGTAAGCAACAATACTATAGCAAGATCATA
ATAGTTTGCATTTCTGCA
TTGGCGTAACACATCAAAAAATGCATGTAATTCATGGATTTCTAGAAATTTAG
AATCATGTCTTTCTTTT
GCTTTACGCCTTTTCTCTAGTGAAATATCTAGTTTACCGCAGTCATTGGAGA
AAACTTAATGACATTAT
ATAACACACCATGATTAATAAATCTTATTACAAGTACTTTTTATATGAGTCATT
GTTGAAGGCGATGCATC
ATACATTTCTAAATATTTATTGAGACTATTTTTCATCAGAAGTGGAGTAATCC
TGTCTAACAAAAATCA
TCTCCTATAATTTTCCCAAGACGCTTCATAACCAGTAGTTCTCTCTGAATTGTT
TGTGGTTTAAACAGAGA
CACACCAAGTCTGAAACCAATTTTCTTTTAACTCTCCAAATGTTGTAATCAGT
TCAGGACTATACTGACT
TTCAAATGAAGTAGTTAGTCTATCTATTTTATCAAGAACCTCTCTTTCAGCTTG
TTTCTCGCCCTACTA

[0386]

GTATTCTTAGTATAACTTACAGTTACTGATTTCCACTTT

- [0387] 서열번호 20
- [0388] 명칭: gi|321156785|emb|CBW38769.1| 인테그라아제 [스트렙토코커스 뉴모니에]
- [0389] 서열:
- MYVYVTKTNSKGQPLYQVVEKYKDPLTGKWKSVTVSYTKNTSRARKQAEREVL
DKIDRLTTSFESQYSPEL
ITTFGELKENWFQWCVSVKPQTIQRELLVMKRLGKIIGDDFLLDRITPLLMKNSL
NKYLEMYDASPSTM
THIKSTCNKIFNHGVLYNVIKFSPMTAVKLDISLEKRRKAKERHDSKFLEIHELHA
FFDVLQRNANYYY
DLAIVLLLTGIRISEAAFLPSDIDFEKGILHIDKALQYHCLKVKQFHFDTTKTLNSIR
EVALPEAASEAI
KRTIQRNKEFDAYMEKHPCPAFTHSESVFRTEYGGSPITSSTFRQILKRIEGKLLTNC
LSDYGFKWVKHVT
- [0390] PHSFRHMHISYLSNEMHIAVKDIMTRVGHANFETTMGYTHNINRSQENTVKAL
NQFVENHNFHFEELKS
YTCKYSRIIEKFIETSDNSNKVELSVDEFKDLLHLSPRYSPKNIISNLLKIKKDIVK
YHPQFDIKIVKS
- [0391] SENQIRGFSIAW
- [0392] 서열번호 21
- [0393] 명칭: gi|43090:1-436 대장균(Tn5086) 디하이드로폴레이트 환원효소 유형 VII에 대한 dhfrVII 유전자 및 sulI 유전자, 5' 말단(인테그라아제)
- [0394] 서열:
- GCATGCCCGTTCCATACAGAAGCTGGGCGAACAACGATGCTCGCCTTCCAG
AAAACCGAGGATGCGAAC
CACTTCATCCGGGGTCAGCACCAACCGCAAGCGCCGCGACGGCCGAGGTCTT
CCGATCTCCTGAAGCCAG
GGCAGATCCGTGCACAGCACCTTGCCGTAGAAGAAGCAAGGCCGCCAAT
GCCTGACGATGCGTGAGAGA
CCGAAACCTTGCGCTCGTTCGCCAGCCAGGACAGAAATGCCTCGACTTCGCT
GCTGCCCAAGGTTGCCGG
GTGACGCACACCGTGGAACGGATGAAGGCACGAACCCAGTGACATAAGC
CTGTTCGGTTCGTAAGCTG
TAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCTTGACCGAAC
GCAGCGGTGGTAACGGCG
CAGTGCGGTTTTTCAT
- [0395]
- [0396] 서열번호 22
- [0397] 명칭: gi|43091|emb|CAA41325.1| 인테그라아제, 일부(플라스미드) [대장균]

- [0398] 서열:
MKTATAPLPPLRSVKVLDQLRERIRYLHYSRLTEQAYVHWVRAFIRFHGVRHPA
TLGSSEVEAFLSWLAN
ERKVSVSTHRQALAALLFFYGKVLCTDLPWLQEIGRPRPSRRLPVVLTTPDEVVRI
LGFLEGEHRLFAQLL
- [0399] YGTGM
- [0400] 서열번호 23
- [0401] >gi|397912605:40372-41898 써모엔에어로박테리움 파지 THSA-485A, 완전한 게놈 -재조합효소
ATGAATCGTGTATGTATTTATCTTAGGAAGTCCCGAGCAGACGAAGAAATAG
AAAAAGAGCTTGGACAAG
GAGAAACACTCGCAAAACATCGTAAGGCCCTTCTTAAATTTGCAAAAGAGAA
AAATTTGAACATAGTAAA
AATCAGAGAGGAAATAGTATCAGGCGAAAGCCTTATCCATAGACCTGAAATG
TTGGAATTACTAAAAGAA
GTCGAACAAGGCATGTACGATGCTGTATTATGTATGGATCTACAGCGTTTAG
GGCGTGGCAACATGCAGG
AACAAAGGTCTCATTTTAGAAGCCTTTAAAAAGTCAAACACTAAAATTATAAC
GCTTCAAAAAACTTATGA
TTTGAACAATGATTTTGACGAAGAATATAGCGAATTGAAGCATTTATGAGC
CGAAAGGAACCTAAAATG
ATAAATAGAAGGCTACAAGGTGGCAGAGTACGCTCTATTCAGGAAGGTAATT
ATTTATCACCATTGCCAC
CTTATGGTTACTTAATACACGAAGAAAAATTTTCGCGCACTCTTGTGCCTAAT
CCTGAGCAAGCTGATGT
AGTTAAAATGATTTTTGATATGTATGTCAATAAACAGATGGGGTCTAGTGCTA
TAGCGAACGAACCTAAAC
AAAATGGGTATAAGACGTATACTGGCAGGAATTGGGCTTCAAGCTCTGTAA
TAAACATACTCAAGAATC
CAGTTTACATCGGTAAAATAACGTGGAAGAAGAAGGATATAAAGAAGTCTGC
TGACCCAAATAAAAAGCAA
AGATACACGTCAAAGACCACGCTCTGAATGGATTGTATCAGATGGCAAACAT
GAACCAATAGTGGGCAAA
GAGCTCTTTGCCAAGGCTCAAGAAATCATTAAAAACAAGTATCACATACCGT
ATCAGATCGTTAATGGTC
CACGTAACCCATTGGCAGGGCTTATTATATGCAAAATATGTGGCTCTAAAAT
GGTGTATAGACCCTACAA
- [0402]

AGATAAAGAAGCGCATATAATATGTCCAAACAAGTGCGGCAATAAAAGCAG
CAAATTTATCTATGTAGAA
AAAAGATTATTACAGGCTTTGGAGGAATGGATGCAAGGCTACGAGCTGGATC
TGCAAATAGAAGAAGATG
ACAGCTCTTTTGCAGAAGCACAAAGAGAAACAAAAAGAAGCTCTTGAAAGAG
AATTGCACGAGCTGCAAAA
GCAAAAGAACAATTTACACGATTTGCTCGAGCGTGGCATATACGATATAGAT
ACATTTGTGGAAAGATCT
ACAATTGTAGCACAGAGAATAGAAGAAACACAGAAAAGTATAGATGTGCTT
GTGCAAAAAATAGAAGAAG
AAAAGAATAAAAGAGACAAAGAAAAAATACTTCCGGAAATTCGGCATGTGT
TGGATCTATATTGGAAAAC
AGACGACATTGCACAAAAAATATGTTGTTAAAGAGCGTACTTGAAAAAGCA
GAATATCTAAAAGAAAAG
AAGCAGAGAGAAGACAACCTTCGAACCTTGGATTTATCCAAAGCTGCCTGAAA
AATAG

[0403]

[0404]

서열번호 24

[0405]

>gi|397912662|ref|YP_006546326.1| 제조합효소 [썬모언에어로박테리움 파지 THSA-485A]

MNRVCIYLKRSRADEEIEKELGQGETLAKHRKALLKFAKEKNLNIVKIREEIVSG
ESLIHRPEMLELLKE
VEQGMYPDVLCDLQRLGRGNMQEQGLILEAFKKSNTKIITLQKTYDLNNDFF
EEYSEFEAFMSRKELKM
INRRLQGGRVRSIQEGNYLSPLPPYGYLIHEEKFSRTLVPNPEQADVVKMIFDMY
VNKQMGSSAIANELN
KMGYKTYTGRNWASSVINILKNPVYIGKITWKKKDIKKSADPNKSKDTRQRPR
SEWIVSDGKHEPIVGK
ELFAKAQEIIKNKYHIPYQIVNGPRNPLAGLIICKICGSKMVYRYPYKDKEAHIICPN
KCGNKSSKFIYVE
KRLQALEEWMQGYELDLQIEEDDSSFAEAQEKQKEALERELHELQKQKNNLH
DLLERGIYDIDTFVERS

[0406]

TIVAQRIEETQKSIDVLVQKIEEEKNKRDKEKILPEIRHVLDLYWKTDDIAQKNML
LKSVLEKAEYLKEK
KQREDNFELWIYPKLPEK

[0407]

[0408]

서열번호 25

[0409]

Gin 제조합효소

[0410]

>gi|657193240|sp|Q38199.2|GIN_BPD10 RecName: 전체=세린 제조합효소 gin; AltName: 전체=G-절편 인버타제;
짧음=Gin

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQK
GDTLVVWKLDRGRSM
KHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFHVMGALAEMERELIERTMAG
LAAARNKGRIGGRP
PKLTKAEWEQAGRLLAQGIPRKQVALIYDVALSTLYKKHPAKRTHIENDDRINQI
DR

[0411]

서열번호 26

[0412]

Cre 재조합효소

[0413]

>gi|375331813|dbj|BAL61207.1| Cre 재조합효소 [Cre-발현 벡터 pHVX2-cre]

[0414]

MVQTSLLTVHQNLPALPVDATSDEVKRNLMDFRDRQAFSEHTWKMLLSVCRS
WAAWCKLNNRKWFPAEP
EDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRK
ENVDAGERAKQALAFE
RTDFDQVRSLMENS DRCQDIRNLAFLGIAYNTLLRIAEIARIRVKDISRTDGG RML
IHIGRTKTLVSTAG
VEKALSLGVTKLVERWISVSGVADDPNNYLF CRVRKNGVAAPSATSQLSTRALE
GIFEATHRLIYGAKDD
SGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSE
TGAMVRLLEDGD

[0415]

서열번호 27 내지 46

[0416]

이들은 본 발명에 기재된 바와 같은 인테그라아제 또는 재조합효소에의 연결에서의 사용을 위한 TALE 반복 모듈을 암호화하는 폴리뉴클레오티드의 예시적인 서열이다.

[0417]

서열번호 27

[0418]

명칭: NI

[0419]

서열:

[0420]

CTGACCCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG
GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTGTG
CCAAGCGCACGGA

[0421]

서열번호 28

[0422]

명칭: NG

[0423]

서열:

[0424]

CTGACCCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG
CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTGT
GCCAAGCGCACGGC

[0425]

서열번호 29

[0426]

명칭: HD

[0427]

- [0428] 서열:
TTGACCCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAGG
AAAGCAAGCCTTGGAACAGTACAGAGGCTGTTGCCTGTGCTGT
GCCAAGCGCACGGG
- [0429]
- [0430] 서열번호 30
- [0431] 명칭: NN
- [0432] 서열:
CTTACCCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG
AAAACAGGCTTTGGAACGGTGCAGAGGCTCCTTCCAGTGCTGT
GCCAAGCGCACGGG
- [0433]
- [0434] 서열번호 31
- [0435] 명칭: NI-NI
- [0436] 서열:
CTGACCCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG
GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCCTC
CAACATTGGCGGGAAACAGGCACTCGAGACTGTCCAGCGCCTGC
TTCCCGTGCTGTGCCAAGCGCACGGT
- [0437]
- [0438] 서열번호 32
- [0439] 명칭: NI-NG
- [0440] 서열:
CTGACCCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG
GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCCATTGCCTC
GAATGGAGGGGGCAAACAGGCGTTGGAAACCGTACAACGATTG
CTGCCGGTGCTGTGCCAAGCGCACGGT
- [0441]
- [0442] 서열번호 33
- [0443] 명칭: NI-HD
- [0444] 서열:
CTGACCCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG
GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCGATCGCAA
GCCACGACGGAGGAAAGCAAGCCTTGGAACAGTACAGAGGCT
GTTGCCTGTGCTGTGCCAAGCGCACGGT
- [0445]
- [0446] 서열번호 34
- [0447] 명칭: NI-NN

- [0448] 서열:
CTGACCCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG
GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCGA
GCAATAACGGCGGAAAACAGGCTTTGGAAACGGTGCAGAGGCT
CCTTCCAGTGCTGTGCCAAGCGCACGGT
- [0449]
- [0450] 서열번호 35
- [0451] 명칭: NG-NI
- [0452] 서열:
CTGACCCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG
CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCCTC
CAACATTGGCGGGAAAACAGGCACTCGAGACTGTCCAGCGCCTGC
TTCCCGTGCTGTGCCAAGCGCACGGT
- [0453]
- [0454] 서열번호 36
- [0455] 명칭: NG-NG
- [0456] 서열:
CTGACCCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG
CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCCATTGCCTC
GAATGGAGGGGGCAAACAGGCGTTGGAAACCGTACAACGATTG
CTGCCGGTGCTGTGCCAAGCGCACGGT
- [0457]
- [0458] 서열번호 37
- [0459] 명칭: NG-HD
- [0460] 서열:
CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCGATCGCAA
GCCACGACGGAGGAAAGCAAGCCTTGGAAACAGTACAGAGGCT
GTTGCCTGTGCTGTGCCAAGCGCACGGT
- [0461]
- [0462] 서열번호 38
- [0463] 명칭: NG-NN
- [0464] 서열:
CTGACCCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG
CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCGA
GCAATAACGGCGGAAAACAGGCTTTGGAAACGGTGCAGAGGCT
CCTTCCAGTGCTGTGCCAAGCGCACGGT
- [0465]
- [0466] 서열번호 39
- [0467] 명칭: HD-NI

- [0468] 서열:
CTGACCCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAG
GAAAGCAAGCCTTGGAACAGTACAGAGGCTGTTGCCTGTGCTTT
GTCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCCT
CCAACATTGGCGGGAAACAGGCACTCGAGACTGTCCAGCGCCTG
CTTCCCGTGCTGTGCCAAGCGCACGGT
- [0469]
- [0470] 서열번호 40
- [0471] 명칭: HD-NG
- [0472] 서열:
GAAAGCAAGCCTTGGAACAGTACAGAGGCTGTTGCCTGTGCTTT
GTCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCCATTGCCT
CGAATGGAGGGGGCAAACAGGCGTTGGAACCGTACAACGATT
GCTGCCGGTGCTGTGCCAAGCGCACGGT
- [0473]
- [0474] 서열번호 41
- [0475] 명칭: HD-HD
- [0476] 서열:
CTGACCCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAG
GAAAGCAAGCCTTGGAACAGTACAGAGGCTGTTGCCTGTGCTTT
GTCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCGATCGCA
AGCCACGACGGAGGAAAGCAAGCCTTGGAACAGTACAGAGGC
TGTTGCCTGTGCTGTGCCAAGCGCACGGT
- [0477]
- [0478] 서열번호 42
- [0479] 명칭: HD-NN
- [0480] 서열:
CTCACCCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAGG
AAAGCAAGCCTTGGAACAGTACAGAGGCTGTTGCCTGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCGA
GCAATAACGGCGGAAAACAGGCTTTGGAACGGTGCAGAGGCT
CCTTCCAGTGCTGTGCCAAGCGCACGGA
- [0481]
- [0482] 서열번호 43
- [0483] 명칭: NN-NI
- [0484] 서열:
CTGACCCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG
AAAACAGGCTTTGGAACGGTGCAGAGGCTCCTTCCAGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCCTC
CAACATTGGCGGGAAACAGGCACTCGAGACTGTCCAGCGCCTGC
TTCCCGTGCTGTGCCAAGCGCACGGT
- [0485]
- [0486] 서열번호 44
- [0487] 명칭: NN-NG

- [0488] 서열:
CTGACCCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG
AAAACAGGCTTTGGAAACGGTGCAGAGGCTCCTTCCAGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCCATTGCCTC
GAATGGAGGGGGCAAACAGGCGTTGGAAACCGTACAACGATTG
[0489] CTGCCGGTGCTGTGCCAAGCGCACGGT
- [0490] 서열번호 45
- [0491] 명칭: NN-HD
- [0492] 서열:
CTGACCCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG
AAAACAGGCTTTGGAAACGGTGCAGAGGCTCCTTCCAGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCGATCGCAA
GCCACGACGGAGGAAAGCAAGCCTTGGAAACAGTACAGAGGCT
[0493] GTTGCCTGTGCTGTGCCAAGCGCACGGT
- [0494] 서열번호 46
- [0495] 명칭: NN-NN
- [0496] 서열:
CTGACCCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG
AAAACAGGCTTTGGAAACGGTGCAGAGGCTCCTTCCAGTGCTTTG
TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCGA
[0497] GCAATAACGGCGGAAAACAGGCTTTGGAAACGGTGCAGAGGCT
- [0498] 서열번호 47
- [0499] 명칭: gi|71796612|gb|DQ084353.1| 양 렌티바이러스 분리물 0v10 인테그라아제 (pol) 유전자, 일부 cds
- [0500] 서열:
CATAGTAAATGGCATCAAGATGCTATGTCATTGCAGTTAGATTTTGGGATACC
GAAAGGTGCGGCAGAAG
ATATAGTACAACAATGTGAAGTATGTCAGGAAAATAAAATGCCTAGCACCAT
[0501] CAGAGGAAGTAACAAAAG
AGGGATAGATCATTGGCAGGTGGATTATACTCATTATAAAGACAAAATAATA
TTGGTATGGGTAGAAACA
[0502] AATTTCGGA
- [0503] 서열번호 48
- [0504] 명칭: gi|71796613|gb|AAZ41325.1| 인테그라아제, 일부 [양 렌티바이러스]
- [0505] 서열:
HSKWHQDAMSLQLDFGIPKGAAEDIVQQCEVCQENKMPSTIRGSNKRIGIDHWQ
[0506] VDYTHYKDKIILVWVET
- [0507] 서열번호 49
- [0508] >gb|AYLT01000127.1|:11804-12046 스타필로코커스 아우레우스 하위종 아우레우스 SK1585 contig000127, 전체 게놈 쏫건(shotgun) 서열

TTATAGATAGGTTAGTGACAAAATACATTTTTTCGTCTAGATTAACCGTGCCTC
TTAGATTATTAATATTT
TCGTTTAGATGTTTTTCAGAACTTTAGCAACTTCATAATCGTTCATGTAAAG
TGTTTGTTTTTTTATTG
TATAATTAAGTAATTCATAATCTTTGTATACTTCTTTTACTTTATCTATATCAA
CATTTTCAAGAACAAG

[0509] TTTTTTATGTTATTATAATTAAAGTTTTCCAT

[0510] 서열번호 50

[0511] >gi|669035130|gb|KFD30483.1| 가상 단백질 D484_02234 [스테필로코커스 아우레우스 하위종 아우레우스 SK1585] - s 아우레우스 카스9

MENFNYYNNIKKLVLNVNDIDKVKEVYKDYELLYTIKNQTLYMNDYEVAKVSE
KHLNENINNLRGTVNLD

[0512] EKCILSLTYL

[0513] 서열번호 51

[0514] 명칭: 링커2의 dna

[0515] 서열:

[0516] agcggcagcgaacccccgggcaccagcgaagcgcgacccccggaaagc

[0517] 서열번호 52

[0518] 명칭: d카스9 단백질

[0519] 서열:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDS
GETAEATR
LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIF
GNIVDEVA
YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQ
TYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSN
FDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEK
MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNR
EKIEKILTR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPK
HSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
KEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLKKIHKDKDFLDNEENEDILEDIVLTLTLFEDR
EMIEERLK
TYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHD
DSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDLVKVMGR
HKPENIVIE
MARENQTTQKGQKNSRERMKRIEELGSGILKEHPVENTQLQNEKLYLYYL
QNGRDMY
VDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN
TKYDENDK

[0520] LIREVKVITLTKSLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGITALIKKYPK
LESEFVYG
DYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG
ETGEIVWD
KGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSP
TVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKK
DLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK
QLFVEQHK
HYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRRDKPIREQAENIIHLFTLTNLGA
PAAFKYFD TTIDRKRYTSTKEVLDTLIHQSTGLYETRIDLSQLGGD

[0521]

- [0522] 서열번호 53
- [0523] 명칭: ATG를 갖는 NLS 뉴클레오티드
- [0524] 서열:
- ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACA
AGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGG
GTACCT
- [0525]
- [0526] 서열번호 54
- [0527] 명칭: GGS 링커 뉴클레오티드
- [0528] 서열:
- [0529] GGGGGAAGT
- [0530] 서열번호 55
- [0531] 명칭: 합성 인테그라아제
- [0532] 서열:
- ATGTTCTGGACGGTATCGACAAAGCTCAGGACGAGCACGAAAAGTACCATT
CTAACTGGCGCGCCATGG
CCTCTGACTTCAATCTCCCGCCGGTTGTTGCCAAGGAGATCGTGGCTTCTTGC
GACAAGTGCCAATTGAA
- [0533]
- GGGTGAGGCTATGCATGGTCAGGTCGATTGCTCTCCCGGTATCTGGCAGCTG
GACTGCACTCACCTCGAG
GGTAAGGTGATTCTCGTTGCTGTGCACGTGGCTTCCGGCTACATCGAGGCTGA
GGTCATCCCGGCTGAGA
CCGGTCAAGAGACTGCTTACTTCCTGCTCAAGCTGGCCGCGCGTTGGCCAGTT
AAGACTATTACACTGA
TAACGGTTCTAACTTTACTTCCGCAACTGTGAAAGCTGCATGCTGGTGGGCCG
GCATTAAACAAGAGTTC
GGAATCCGTATAACCCGCAGTCTCAGGGCGTTGTCGAGTCTATGAACAAGG
AGCTCAAAAAGATCATTG
GTCAAGTCCGTGACCAAGCTGAGCACCTTAAGACCGCTGTGCAGATGGCTGT
TTTTATTCACTAACTTCAA
GCGTAAGGGTGGTATCGGTGGTTATAGCGCTGGTGAGCGTATCGTAGACATC
ATCGCTACTGATATCCAG
ACAAAGGAGCTGCAGAAGCAGATCACTAAGATCCAGAACTTCCGTGTGTACT
ATCGGGACTCTAGGAACC
CGCTCTGGAAGGGTCTGCTAAACTGCTGTGGAAGGGAGAGGGTGTGTTGT
TATCCAGGACAACTCTGA
TATCAAGGTGGTTCCGCGTCGTAAGGCTAAAATTATCCGCGACTACGGCAAG
CAAATGGCTGGAGACGAC
- [0534] TCGGTTGCTAGCCGTCAAGACGAAGACTAA
- [0535] 서열번호 56

[0536] 명칭: ATG를 갖는 d카스9 뉴클레오티드

[0537] 서열:

ATGGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGGATG
GGCTGTCA
TAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACAC
AGACCGTC
ATTCGATTA AAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACG
GCAGAGG

[0538] CGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACC
GAATATGTT
ACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTT
CACCGTTT
GGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATC
TTTGGA
CATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTC
AGAAAAA
GCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTG
CCCATATG
ATAAAGTTCCGTGGGCACCTTCTCATTGAGGGTGATCTAAATCCGGACAACCTC
GGATGTC
GACAAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGA
ACCCTATA
AATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAAT
CCCGACGG
CTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCG
GTAACCTT
ATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGC
TGAAGAT
GCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTAC
TGGCACAA
ATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGC
AATCCTCC
TATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCT
TCAATGAT
CAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTC
CGTCAGCA
ACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTAC
GCAGGTTA
TATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATA
TTAGAGAA

[0539]

GATGGATGGGACGGAAGAGTTGCTTGTA AAACTCAATCGCGAAGATCTACTG
 CGAAAGC
 AGCGGACTTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATT
 GCATGCTA
 TACTTAGAAGGCAGGAGGATTTTTATCCGTTCTCTCAAAGACAATCGTGAAAA
 GATTGAGA
 AAATCCTAACCTTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAA
 CTCTCGGTT
 CGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAG
 GAAGTTGT
 CGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGAC
 AAGAATTT
 ACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTACA
 GTGTACAA
 TGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTT
 CTAAGCGG
 AGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGT
 GACAGTTA
 AGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGA
 GATCTCCG
 GGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAA
 GATAATTA
 AAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATAT
 AGTGTTGA
 CTCTTACCCCTTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATA
 CGCTCACC
 TGTTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTG
 GGGACGAT
 TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAACTAT
 TCTCGATT
 TTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGA

[0540]

TGACTCTTT
 AACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTC
 ATTGCACG
 AACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCA
 GACAGTCA
 AAGTAGTGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACAT
 TGTAATCG
 AGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCGAG
 AGCGGAT
 GAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGA
 GCATCCTG
 TGGAAAATACCCAATTGCAGAACGAGAACTTTACCTCTATTACCTACAAAA
 TGGAAAGGG
 ACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTC
 GATGCCAT
 TGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACAC
 GCTCGGAT
 AAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAA
 ATGAAGA
 ACTATTGGCGGCAGCTCCTAAATGCGAACTGATAACGCAAAGAAAGTTCGA
 TAACTTAA
 CTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAA
 ACGTCAGC
 TCGTGGAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCG
 AATGAATA
 CGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTT
 AAAGTCA
 AAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGA
 TAAATAAC
 TACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCAT
 TAAGAAA

[0541]

TACCCGAAGCTAGAAAAGTGAGTTTGTGTATGGTGATTACAAAAGTTTATGACG
TCCGTAAG
ATGATCGCGAAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTC
TTTTATTCT
AACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATAC
GCAAACGA
CCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCC
GGGACTTC
GCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAA
CTGAGGTG
CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATA
AGCTCATC
GCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTA
CAGTTGCC
TATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAAATCCAAGAACTGA
AGTCAGTC
AAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACC
CCATCGAC
TTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAAC
TACCAAAG
TATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCG
GAGAGCTT
CAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCTGTATT
TAGCGTCC
CATTACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTT
TTGTTGAG
CAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTTCGGAATTCAGTA
AGAGAGTC
ATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACA
GGGATAAA
[0542] CCCATACGTGAGCAGGCGGAAAAATATTATCCATTTGTTTACTCTTACCAACCT
CGGCGCTC
CAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTC
TACCAAGG
[0543] AGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAAC
TCGGATAG ATTTGTCACAGCTTGGGGGTGACTAA

[0544] 서열번호 57

[0545] 명칭: 애비1(NLS-링커1-인테그라아제-링커2-d카스9) -DNA 서열

[0546]

서열:

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACA
AGGATGAC
GATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGG
GTACCT
GGGGGAAGTATGTTCTGGACGGTATCGACAAAGCTCAGGACGAGCACGAA
AAGTACCATTCTAACTGGCGCGCCATGGCCTCTGACTTCAATCTCCCGCCGGT
TGTTGCCAAGGAGATCGTGGCTTCTTGCGACAAGTGCCAATTGAA
GGGTGAGGCTATGCATGGTCAGGTCGATTGCTCTCCCGGTATCTGGCAGCTG
GACTGCACTCACCTCGAG
GGTAAGGTGATTCTCGTTGCTGTGCACGTGGCTTCCGGCTACATCGAGGCTGA
GGTCATCCCGGCTGAGA
CCGGTCAAGAGACTGCTTACTTCTGCTCAAGCTGGCCGGCCGTTGGCCAGTT
AAGACTATTCACACTGA
TAACGGTTCTAACTTTACTTCCGCAACTGTGAAAGCTGCATGCTGGTGGGCCG
GCATTAAACAAGAGTTC
GGAATTCCGTATAACCCGCAGTCTCAGGGCGTTGTGCGAGTCTATGAACAAGG
AGCTCAAAAAGATCATTG
GTCAAGTCCGTGACCAAGCTGAGCACCTTAAGACCGCTGTGCAGATGGCTGT
TTTTATTCACTAACTTCAA
GCGTAAGGGTGGTATCGGTGGTTATAGCGCTGGTGAGCGTATCGTAGACATC
ATCGCTACTGATATCCAG
ACAAAGGAGCTGCAGAAGCAGATCACTAAGATCCAGAACTTCCGTGTGTACT
ATCGGGACTCTAGGAACC

[0547]

CGCTCTGGAAGGGTCCTGCTAAACTGCTGTGGAAGGGAGAGGGTGCTGTTGT
TATCCAGGACAACTCTGA
TATCAAGGTGGTTCCGCGTCGTAAGGCTAAAATTATCCGCGACTACGGCAAG
CAAATGGCTGGAGACGAC
TGCGTTGCTAGCCGTCAAGACGAAGACagcggcagcgaacccccgggcaccagcgaagcgca
ccccggaaagc
ATGGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGGATG
GGCTGTCA
TAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACAC
AGACCGTC
ATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACG
GCAGAGG
CGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACC
GAATATGTT
ACTTACAAGAAATTTTATAGCAATGAGATGGCCAAAGTTGACGATTCTTCTTT
CACCGTTT
GGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATC
TTTGAAAA
CATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTC
AGAAAAAA
GCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTG
CCCATATG
ATAAAGTTCGGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAATC
GGATGTC
GACAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGA
ACCTATA
AATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAAT
CCCGACGG
CTAGAAAACCTGATCGCACAAATTACCCGGAGAGAAGAAAAATGGGTTGTTTCG
GTAACCTT
ATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGC

[0548]

TGAAGAT
 GCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTAC
 TGGCACAA
 ATTGGAGATCAGTATGCGGACTTATTTTGGCTGCCAAAAACCTTAGCGATGC
 AATCCTCC
 TATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCT
 TCAATGAT
 CAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTC
 CGTCAGCA
 ACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTAC
 GCAGGTTA
 TATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATA
 TTAGAGAA
 GATGGATGGGACGGAAGAGTTGCTTGTA AAACTCAATCGCGAAGATCTACTG
 CGAAAGC
 AGCGGACTTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATT
 GCATGCTA
 TACTTAGAAGGCAGGAGGATTTTTATCCGTTCTCTCAAAGACAATCGTGAAAA
 GATTGAGA
 AAATCCTAACCTTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAA
 CTCTCGGTT
 CGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAG
 GAAGTTGT
 CGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGAC
 AAGAATTT
 ACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTACA
 GTGTACAA
 TGAATCACGAAAGTTAAGTATGTC ACTGAGGGCATGCGTAAACCCGCCTTT
 CTAAGCGG
 AGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGT
 GACAGTTA

[0549]

AGCAATTGAAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGA
 GATCTCCG
 GGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAA
 GATAATTA
 AAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATAT
 AGTGTTGA
 CTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATA
 CGCTCACC
 TGTTGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTG
 GGGACGAT
 TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAACTAT
 TCTCGATT
 TTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGA
 TGACTCTTT
 AACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTC
 ATTGCACG
 AACATATTGCGAATCTTGCTGGTTCCGCCAGCCATCAAAAAGGGCATACTCCA
 GACAGTCA
 AAGTAGTGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACAT
 TGTAATCG
 AGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAACAGTCGAG
 AGCGGAT
 GAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGA
 GCATCCTG
 TGGAAAATACCCAATTGCAGAACGAGAACTTTACCTCTATTACCTACAAAA
 TGGAAGGG
 ACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTC
 GATGCCAT
 TGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACAC
 GCTCGGAT
 AAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAA

[0550]

ATGAAGA
 ACTATTGGCGGCAGCTCCTAAATGCGAACTGATAACGCAAAGAAAGTTCGA
 TAACTTAA
 CTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAA
 ACGTCAGC
 TCGTGGAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCG
 AATGAATA
 CGAAATACGACGAGAACGATAAGCTGATTTCGGGAAGTCAAAGTAATCACTTT
 AAAGTCA
 AAATTGGTGTCTGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGA
 TAAATAAC
 TACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCAT
 TAAGAAA
 TACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACG
 TCCGTAAG
 ATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTC
 TTTTATTCT
 AACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATAC
 GCAAACGA
 CCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCC
 GGGACTTC
 GCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAA
 CTGAGGTG
 CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATA
 AGCTCATC
 GCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTA
 CAGTTGCC
 TATTCTGTCCTAGTAGTGGCAAAGTTGAGAAGGGAAAATCCAAGAAACTGA
 AGTCAGTC
 AAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACC
 CCATCGAC

[0551]

TTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAAC
TACCAAAG
TATAGTCTGTTTGTAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCG
GAGAGCTT
CAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTCCTGTATT
TAGCGTCC
CATTACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTT
TTGTTGAG
CAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTCGGAATTCAGTA
AGAGAGTC
ATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACA
GGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCT
CGGCGCTC
CAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTC
TACCAAGG
AGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAAC
TCGGATAG ATTTGTCACAGCTTGGGGGTGACTAA

[0552]

[0553]

서열번호 58

[0554]

명칭: 예비1(결합 기반 인테그라아제 편집기)의 번역

[0555]

서열:

MetDYKDHDGDYKDHDIDYKDDDDKMetAPKKKRKVGIHR
GVPGGSMetFLDGIDKAQDEHEKYHSNWRAMetASDFNLPP
VVAKEIVASCDKCQLKGEAMetHGQVDCSPGIWQLDCTHL
EGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLGRW
PVKTIHTDNGSNFTSATVKAACWWAGIKQEFGIPYNPQSQ
GVVESMetNKELEKKIIGQVRDQAEHLKTAVQMetAVFIHNF
KRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVY
YRDSRNPLWKGPALLWKGEAVVIQDNSDIKVVPKRKA
KIIRDYDGKQMetAGDDCVASRQDEDSGSETPGTSESATPES
MetDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNT

[0556]

DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNR
 ICYLQEIFSNE^{Met}AKVDDSFHRLEESFLVEEDKKHERHPI
 FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRRLIYLAL
 AH^{Met}IKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFE
 ENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGL
 FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLN
 LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
 AS^{Met}IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
 GYAGYIDGGASQEEFYKFIKPILEK^{Met}DGTEELLVKLNRE
 DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPPFLKDN
 REKIEKILTFRIPYYVGPLARGNSRFAW^{Met}TRKSEETITPW
 NFEEVVDKGASAQSFIER^{Met}TNFDKNLPNEKVLPKHSLLY
 EYFTVYNELTKVKYVTEG^{Met}RKPAFLSGEQKKAIVDLLF
 KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT
 YHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDRE^{Met}IE
 ERLKTYAHLFDDKVM^{Met}KQLKRRRYTGWGRLSRKLINGIR
 DKQSGKTILDFLKSDGFANRNF^{Met}QLIHDDSLTFKEDIQK
 AQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDDELVK
 V^{Met}GRHKPENIVIE^{Met}ARENQTTQKGQKNSRER^{Met}KRIEE
 GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM^{Met}YV
 DQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNR
 GKSDNVPSEEVVKK^{Met}KNYWRQLLNAKLITQRKFDNLTK
 AERGGLSELDKAGFIKQQLVETRQITKHVAQILDSR^{Met}NT
 KYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNY
 HHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVR
 K^{Met}IAKSEQEIGKATAKYFFYSNIM^{Met}NFFKTEITLANGEIR
 KRPLIETNGETGEIVWDKGRDFATVRKVLS^{Met}PQVNIVKK
 TEVQTGGFSKESILPKRNSDKLIARKKDWDPPKKYGGFDSP
 TVAYSVLVVAKVEKGKSKKLKSVKELLGITIM^{Met}ERSSEFEK
 NPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKR^{Met}LA
 SAGELQKGNELALPSKYVNFLYLASHYEKCLKGSPEDNEQ

[0557]

KQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY
 NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
 YTSTKEVLDA^TLIHQ^SITGLYETRIDLSQLGGD

[0558]

도너 DNA 경우(인테그라아제 인식을 위한 LTR 영역의 att 위치).

[0559]

서열번호 59

[0560]

명칭: U3att

[0561]

서열:

[0562]

ACTGGAAGGGCTAATTCAC^TCCCCAAAGAA

[0563]

서열번호 60

[0564]

- [0565] 명칭: U5att
- [0566] 서열:
- [0567] GACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGT
- [0568] NLS-링커1-인테그라아제-링커2-d카스9, 또는 인테그라아제-링커1-NLS-링커2-d카스9 또는 인테그라아제-링커2-d카스9-링커1-NLS 또는 인테그라아제-링커2-d카스9-NLS
- [0569] 링커 1 = GGS
- [0570] 서열번호 61
- [0571] 명칭: 링커 2
- [0572] 서열:
- [0573] SGSETPGTSESATPES
- [0574] 서열번호 62
- [0575] 명칭: MMTV 인테그라아제 cDNA, gb|AF071010.1|:16-1113 마우스 유방 종양 바이러스 추정 인테그라아제, env 다단백질, 및 초항원 mRNA, 완전한 cds
- [0576] 서열:
- ATGACAGGAAAGTGGCCTTGTATTTACTCCACTAACTGCAGAGATGTGTTGC
ATGGGACGGGGGGGCACTG
CACCAGCCCTCGTGCTGAATTCGGCACGAGGAAATGCCTATGCAGATTCTTTA
ACAAGAATTCTGACCGC
TTTAGAGTCAGCTCAAGAAAGCCACGCACTGCACCATCAAAATGCCGCGGCG
CTTAGGTTTCAGTTTCAC
- [0577]

ATCACTCGTGAACAAGCACGAGAAATAGTAAAATTATGTCCAAATTGCCCCG
 ACTGGGGACATGCACCAC
 AACTAGGAGTAAACCCTAGGGGCCTTAAGCCCGGGTTCTATGGCAAATGGA
 TGTTACTCATGTCTCAGA
 ATTTGGAAAATTAAAGTATGTACATGTGACAGTGGATACTTACTCTCATTTTA
 CTTTCGCTACCGCCCGG
 ACGGGCGAAGCAGCCAAAGATGTGTTACAACACTTGGCTCAAAGCTTGCAT
 ACATGGGCATTCTCTCAA
 AAATAAAAACAGATAATGCCCCTGCCTATGTGTCTCGTTCAATACAAGAATTT
 CTGGCCAGATGGAAAAT
 ATCTCACGTCACGGGGATCCCTTACAATCCCCAAGGACAGGCCATTGTTGAA
 CGAACGCACCAAAATATA
 AAGGCACAGATTAATAAACTTCAAAAGGCTGGAAAATACTATACACCCCACC
 ATCTATTGGCACATGCTC
 TTTTGTGCTGAATCATGTAAATATGGACAATCAAGGCCATACAGCGGCCGA
 AAGACATTGGGGTCCAAT
 CTCAGCCGATCCAAAACCTATGGTCATGTGGAAAGACCTTCTCACAGGGTCC
 TGGAAAGGACCCGATGTC
 CTAATAACAGCCGACGAGGCTATGCTTGTGTTTTTCCACAGGATGCCGAATC
 ACCAATCTGGGTCCCCG
 ACCGGTTCATCCGACCTTTTACTGAGCGGAAAGAAGCAACGCCACACCTGG
 CACTGCGGAGAAAACGCC
 GCCGCGAGATGAGAAAGATCAACAGGAAAGTCCGGAGGATGAATCTTGCCC
 CCATCAAAGAGAAGACGGC

[0578] TTGGCAACATCTGCAGGCGTTAATCTCCGAAGCGGAGGAGTTCTTAA

[0579] 서열번호 63

[0580] 명칭: gi|3273866|gb|AAC24859.1| 추정 인테그라아제 [마우스 유방 종양 바이러스]

[0581] 서열:

[0582] MTGKWPCISTNCRDVLHGTGGTAPALVLNSARGNAYADSLTRILTALESAQES
 HALHHQNAALRFQFH
 ITREQAREIVKLCPCPDWGHAPQLGVNPRGLKPGVLWQMDVTHVSEFGKLKY
 VHVTVDITYSHFTFATAR
 TGEAAKDVLQHLAQSFAYMGIPQKIKTDNAPAYVSRISQEFARWKISHVTGIPY
 NPQGQAIVERTHQNI
 KAQINKLQKAGKYYPHLLAHALFVLNVNMDNQGHATAERHWGPISADPKP
 MVMWKDLLTGSWKGPDV
 LITAGRGYACVFPQDAESPIWVPDRFIRPFTERKEATPTPGTAEKTPPRDEKDQQE
 SPEDESCPHQREDG
 [0583] LATSAGVNLRSGGGS

[0584] 서열번호 64

[0585] 명칭: gb|AXUN02000059.1|:5116-8850 용기박터 프래질리스(Youngiibacter fragilis) 232.1 contig_151, 전체 게놈 샷건 서열 - 재조합효소

[0586]

서열:

TTGAAAGATAACGATAAAAGGATGTGGGTTTCAGACTTTATGGAATCCCATCA
 ATGAAAGACATAAAAGTC
 CACTGGATAGCCCAGAACCAGGGATTAAAGTAGCGGCCTACTGCAGAGTAAG
 CATGAAAGAGGAGGAACA
 ACTCCGGTCATTGGAAAACCAGGTGCATCACTATACTCATTTTATCAAAAGTA
 AGCCGAATTGGAGATTT
 GTAGGGGTTTATTACGATGATGGCATAAGTGCAGCCATGGCAAGTGGGAGAA
 GAGGGTTCCAGCGGATTA
 TCCGTCATGCTGAAGAAGGTAAGGTTGATCTGATTCTAACAAGAATATTTC
 ACGGTTTTCCAGAAATTC
 CAAGGAGTTACTGGATATAATCAATCAACTGAAAGCTATCGGTGTGGGCATC
 TATTTTGAGAAAGAGAAT
 ATTGATACTTCAAGAGAGTACAATAAATTCCTCTTAAGCACTTATGCTGCGCT
 GGCACAGGAAGAGATAG
 AAAC TATTTCAA CTCTACGATGTGGGGTTATGAGAAAAGGTTTCTAAAGGG
 TATCCCAAAGTTCAACCG
 CTTATATGGATACAAAGTCATCCATGCAGGGGATGATTCCCAATTGATTGTTC
 TTGAAGATGAAGCAAAA

[0587]

ATCGTAAGAATGATGTATGAACAGTACCTTCAAGGGAAGACGTTCACTGATA
 TTGCAAGGGCGCTAACAG
 AAGCTGGAGTGAAAACAGCCAAAGGGAAGGATGTCTGGATAGGCGGCATGA
 TAAAGCATATTTTATCCAA
 CGTCACCTACACCGGTAACAAGCTTACACGAGAACTGAAAAGAGATTTATTT
 ACGAACAAAGTTAATAGC
 GGTGAACGGGATCAGGTTTTTATAGGAAACACTCACGAACCGATCATCAGCA
 ATGATATTTTCAATCTTG
 TTCAAAAGAAGCTTGAGGCCAATACGAAGGAAAGAAAGCCCAGTGAGAAGC
 GAGAGAAGAACCACATGTC
 TGGTCGGCTACTTTGCGGAAGATGTGGATACAGTTTTACCATAATTCACAATA
 GAGCTTCTCATCACTTT
 AAGTGTAGCCCTAAAATCATGGGGGTCTGTGATTCTGAACTTTATCGGGATGC
 GGATATTCGAGAAATGA
 TGATGAGGGCAATGTATATAAAATATGACTTCACCGATGAAGACATAGTACT
 AAAACTGCTGAAGGAACT
 CCAGGTCATCAATCAAAATGATCACTTTGAGTTTCATAGGCTAAAGTTTATCA
 CTGAAATTGAAATCGTA
 AAAAGGCAGCAGGCCATTTTACAGATAGATATTCAGCTATTAGCATAGAAAAAA
 TGGAAGAAGAATACCGCA
 CTTTTGAAAGCAAGATTGCGAAAATTGAGGATGACAGGTACATCAGAATCGA
 TGCAGTGGAGTGGTTAAA
 GAAAAACAAGACGCTGGATTCTTTTATCGCTCAGGTCACCACTAAAATATTG
 CGAGCTTGGGTTTCCGAG
 ATGACTGTTTATACACGAGATGACTTTTTAGTGCAGTGGATTGACGGAACTCA
 AACTGAGATAGGAAGCT
 GCGAGCATCATCTTGTGAAGGATAGAAATAGTAAGAGTTACGAGTCCGGTGA
 AGAAACGAGCAGGAGGGC
 CAAATTTGAAGTCAACCACATTAGTGAAACCACCGAAGGACAAGGAGAACTT
 GATCTCTTAAGCAAGAGT

[0588]

GCAAGTTCAAACAATGAAGATAGTAATCAACCAGAAAATAATTCTACGGGAA
 AGGAGGAGCTTGAATTGA
 ACTTAAACAGTAATGCAGAAATTATCAAAATTGAGCCCGGGCAAAGGGACTA
 TATTATGAAGAATTTGCA
 CAAGAGCCTGAGTGCAAATATGATGATGCAAAATGCTTCAGTACACACGGCA
 AGTATTAACAAACCTAGA
 CTTAAGACTGCTGCTTACTGCAGAATCTCAACAGATTCAGAAGAACAAAAGG
 TAAGCTTGAAAACCCAAG
 TAGCCTATTACACTTATCTGATTCTAAAGGATCCCCAATATGAATATGCAGGC
 ATCTATGCCGATGAAGG
 TATATCAGGGCGTTCTATGAAAAACCGTACAGAATTTCTCAAATACTCGAA
 GAATGTAAAGCCGGGAAT
 GTGGACTTGATTTTAAACCAAGTCAATCTCACGGTTTAGCAGAAACGCATTAG
 ATTGCTTGGAACAGATCA
 GGATGCTGAAGTCGCTGCCAAGTCCAGTTTATGTGTATTTTGAGAAAGAGAA
 TATTCATACAAAAGATGA
 GAAGAGTGAGCTGATGATTTCTATTTTGGAAAGTATCGCTCAGGAAGAGAGC
 GTAAACATGGGAGAAGCC
 ATGGCTTGGGGAAAACGGAGATATGCTGAGAGAGGGATAGTAAACCCAAGT
 GTTGACCTTATGGATATA
 GAACGGTCAGAAAAGGTGAATGGGAGGTGGTTGAAGAAGAAGCTACGATCA
 TTAGAAGAATTTATCGGAT
 GCTCCTAAGTGGAAGAGTATTCATGAAATCACAAAGGAGCTCTCCATGGAG
 AAGATAAAGGGTCCTGGC
 GGCAACGAGCAGTGGCATCTTCAAACCATTAGAAATATCTTGAGAAATGAAA
 TCTATAGGGGTAACCTACC
 TTTATCAAAAGGCTTATATCAAGGACACGATCGAGAAGAAGGTGGTAATGAA
 TCGAGGAGAACTGCCACA
 GTATCTCATAGAGAATCATCATAAAGCCATTGTTGACAATGAGACCTGGGAA
 AAGGTCCAGAAGGTACTA

[0589]

GAAGCCAGAAGGGAAAAATATGAGAATAAAAAGTCCATAACTTATCCTGAA
GACAAAATGAAAAACGCTT
CTCTTGAAGATATTTTTACCTGTGGAGAATGTGGAAGTAAAATAGGCCATAG
AAGGAGCATCCAGAGCTC
TAATGAGATTCAATTCCTGGATCTGCACAAAAGCCGCTAAGTCTTCTTGGTGG
ACTCGTGTAAGTCCACA
AGCGTATATCAGAAGCACCTGGAGCTGCATTTTATGAAGACTCTTCTCGATAT
TAAAAAGCATCGTTCTT
TCAAAGATGAGGTGCTCACCTATATTCGAACCCAAGAAGTAGATGAAAAGGA
AGAGTGGAGAATCAAAGT
CATAGAGAAACGAATCAAAGATCTTAACAGAGAGCTTTATAATGCGGTAGAC
CAGGAGCTCAATAAAAAA
GGTCAGGACTCCAGGAAAGTTGATGAGCTCACAGAGAAAATTGTGGATCTTC
AAGAGGAATTAAAGGTGT
TTAGGGACCGAAAGGCAAAGGTTGAGGATCTTAAAGCTGAGCTTGAATGGTT
CCTAAAGAAGCTGGAAAC
CATTGATGACGCTCGAGTAAAAAGAAATGAAGGAATAGGCCACGGTGAAGA
GATCTACTTCAGAGAAGAT
ATTTTTGAAAGAATAGTAAGGAGTGCACAGCTTTATAGCGATGGAAGGATCG
TCTACGAACTAAGCCTCG
GGATCCAGTGGTTCATTGACTTTAAATACAGCGCATTTTACAAGCTTCTTATA
AAGTGGAAGGATAAACA
AAGGGCAGAAGAAAAAGAGGCTTTTCTTGAGGGGCCGGAAGTTAAAGAGCT
GCTGGAATTTTGTAAGGAA
CCGAAGAGCTACTCTGATTTACATGCCTTCATGTGTGAGAGAAAAAGAGGTGT
CTTATAGCTATTTACAGGA
AATTGGTGATAAGACCTTTGATGAAGAAAGGAAAGCTGAAGTTCACCATACC
AGAAGATGTTATGAATAG
GCATCAGAGATACACATCAATCTAA

[0590]

[0591]

서열번호 65

[0592]

명칭: gi|564135645|gb|ETA81829.1| 재조합효소 [용기박터 프레질리스 232.1]

[0593]

서열:

MKDNDKRMWVQTLWNPINERHKSPLDSPEPGIKVAAAYCRVSMKEEEQLRSLEN
QVHHYTHFIKSKPNWRF
VGVEYDDGISAAMASGRRGFQRIIRHAEKGVDLILTKNISRFSRNSKELLDIINQ
LKAIGVGIYFEKEN
IDTSREYNKFLSTYAALAEETISNSTMWGYEKFLKGIPKFNRLYGYKVIHA
GDDSQLIVLEDEAK
IVRMMYEQYLQGKTFTDIARALTEAGVKTAKGKDVWIGGMIKHILSNVTYTGN
KLTRELKRDFTNKVNS
GERDQVFIGNTHEPIISNDIFNLVQKKLEANTKERKPSEKREKNHMSGRLLCGRG
GYSFTIIHNRASHHF
KCSPKIMGVCDSELRYDADIREMMMRAMYIKYDFTDEDIVLKLKELQVINQND
HFEFHRLKFITEIEIV
KRQQAISDRYSAISIEKMEEEEYRTFESKIAKIEDDRYIRIDAVEWLKKNKTLDSFIA
QVTTKILRAWVSE
MTVYTRDDFLVQWIDGTQTEIGSCEHHLVKDRNSKSYESGEETSRRAKFEVNHIS
ETTEGQGELDLLSKS
ASSNNEDSNQPENNSTGKEELELNLNSNAEIIKIEPGQRDYIMKNLHKSLSANMM
MQNASVHTASINKPR
LKTAAYCRISTDSEEQKVSLKTQVAYYTYLILKDPQYIYAGIYADEGISGRSMKN
RTEFLKLEECKAGN
VDLILTKSISRFSRNALDCLEQIRMLKSLPSPVYVYFEKENIHTKDEKSELMISIFGS
IAQEEVNMGEA
MAWGKRRYAERGIVNPSVAPYGYRTVRKGEWEVVEEEATIIRRIYRMLLSGKSI
HEITKELSMEEKIKGPG
GNEQWHLQTIRNILRNEIYRGNYLYQKAYIKDTIEKKVVMNRGELPQYLIENHH
KAIVDNETWEKVQKVL
EARREKYENKKSITYPEDKMKNASLEDIFTCGECGSKIGHRRSIQSSNEIHSWICT
KAAKSFLVDSCKST
SVYQKHLELHFMKTLLDIKKHRSFKDEVLTYIRTQEVDEKEEWRIKVIEKRIKDL
NRELYNAVDQELNKK

[0594]

GQDSRKVDELTEKIVDLQEELKVFRDRKAKVEDLKAELEWFLKKLETIDARVK
RNEGIGHGEEIYFRED
IFERIVRSAQLYSDGRIVYELSLGIQWFIDFKYSAFQKLLIKWKDKQRAEEKEAFL
EGPEVKELLEFCKE

[0595]

PKSYSDLHAFMCERKEVSYSYFRKLVRPLMKKGKLGKFTIPEDVMNRHQRYTSI

[0596]

서열번호 66

[0597]

명칭: gi|571264543:16423-16770 클로스트리듐 디피실 트랜스포존 Tn6218, 균주 0x42 전이효소

- [0598] 서열:
TTAGTCTTCAAAAGGTTTTGGACTAAATTTACTCTCGTAGTCAGGTCCAAGTG
TTTCTTCAGATTTTTTT
TTCAACCAATCCACCTGCATGGTGAGCTGGCCAACTTTTTTCGCATATTCAGC
TTTTTCCTTGCGTTCTA
AAGCGAGTTTTTCTTTCAGATTATCCTCTCGTGTGTCATTAAAAACCACGGAT
GCTTTATCGAGGAACTC
CTTCTTCCAGTTGCGGAGAAGATTTCGGCTGAATATTGTTTTCGGTTGCGATTG
TATTTAAGTCTTTTTCT
CCTTTGAGCAGTTCAATCACTAATTCTGATTGAATTTGGCAGAGAAATTTCT
TCTTGTTGAGACAT
- [0599]
- [0600] 서열번호 67
- [0601] 명칭: gi|571264559|emb|CDF47133.1| 전이효소 [헵토클로스트리움 디피실]
- [0602] 서열:
MSRTRRNFSAKFKSELVIELLKGEKDLNTIATENNIQPNLLRNWKKEFLDKASVV
FNDTREDNLKEKLAL
- [0603] ERKEKAHEYAKKVGOLTMOVDWLKKKSEETLGPDYESKFSPKPFED
- [0604] 서열번호 68
- [0605] 명칭: gb|CP009444.1|:1317724-1320543 프란시셀라 필로미라지아(*Francisella philomiragia*) 균주 GA01-2801, 완전한 게놈 Cpf1
- [0606] 서열:
ATGAATCTATATAGTAATCTAACAAATAAATATAGTTTAAGTAAAACTCTAA
GATTTGAGTTAATTCCAC
- [0607]

AGGGTGAAACACTTGAAAAATATAAAAGCAAGAGGTTTGATTTTAGATGATGA
 GAAAAGAGCTAAAGACTA
 TAAAAAAGCTAAACAAATCATTGATAAATATCATCAGTTTTTTATAGAGGAG
 ATATTAAGTTCGGTATGT
 ATTAGCGAAGATTTATTACAAAACCTATTCTGATGTTTATTTTAACTTAAAAA
 GAGTGATGATGATAATC
 TACAAAAAGATTTTAAAAGTGCAAAAGATACGATAAAGAAACACATATCTAG
 ATATATAAATGACTCGGA
 GAAATTTAAGAATTTGTTTAAATCAAAATCTTATAGATGCTAAAAAAGGGCAA
 GAGTCAGATTTAATTCTA
 TGGCTAAAGCAATCTAAGGATAATGGCATAGAACTATTTAAAGCTAACAGTG
 ATATCACAGACATAGATG
 AGGCGTTAGAAATAATCAAATCTTTTAAAGGTTGGACAACCTATTTTAAGGGT
 TTTTCATGAAAATAGAAA
 AAATGTCTATAGTAGTGATGATATCCCTACATCTATTATTTATAGAATAGTAG
 ATGATAATTTGCCTAAA
 TTTATAGAAAATAAAGCTAAGTATGAGAATTTAAAAGACAAAGCTCCAGAAG
 CTATAAACTATGAACAAA
 TTA AAAAAGATTTGGCAGAAGAGCTAACCTTTGATATTGACTACAAAACATC
 TGAAGTTAATCAAAGAGT
 TTTTTCACCTTGATGAAGTTTTTGAGATAGCAAACCTTTAATAATTATCTAAATC
 AAAGTGGTATTACTAAA
 TTTAATACTATTATTGGTGGTAAATTTGTTAATGGTGAAAATACAAAGAGAA
 AAGGTATAAATGAATATA
 TAAATCTATACTCACAGCAAATAAATGATAAAACACTTAAAAAATATAAAAT
 GAGTGTTTTATTTAAGCA
 AATTTTAAGTGATACAGAATCTAAATCTTTTGTAATTGATAAGTTAGAAGATG
 ATAGTGATGTAGTTACA
 ACGATGCAAAGTTTTTATGAGCAAATAGCAGCTTTTAAAACATTAGAAGAAA
 AGTCTATTAAGGAAACAT

[0608]

TATCTTTACTATTTGATGATTTAAAAGCTCAAAAACCTTGATTTGAGTAAAATT
 TATTTTAAAAATGATAA
 ATCTCTTACTGATCTATCACACAAGTTTTTGATGATTATAGTGTTATTGGTAC
 AGCGGTACTAGAATAT
 ATAACTCAACAAGTAGCACCTAAAAATCTTGATAACCCTAGTAAGAAAAGAGC
 AAGATTTAATAGCCAAAA
 AAAGTGAAGGCAAAATACCTTATCTCTAGAACTATAAAGCTTGCCTTAGA
 AGAATTTAATAAGTATAG
 AGATATAGATAAACAGTGAGGTTTGAAGAAATATTTGCAAGCTTTGCAGAT
 ATTCCGGTGCTATTTGAT
 GAAATAGCTCAAAAACAAAAACAATTTGGCACAGATATCTATCAAATATCAAA
 ATCAAGGTAAAAAAGACC
 TGCTTCAAAGTAGTGAGAGTAGATGTTAAAGCTATCAAGGATCTTTTGGAT
 CAACTAATAATCTCTT
 GCATAAACTAAAAATATTTTCATATTACGCAATCAGAAGATAAGGCAAATATT
 TTAGACAAGGATGAGCAT
 TTTTATTTAGTATTTGATGAGTGCTACTTTGAGCTAGCGAATATAGTGGCTCTT
 TATAACAAAATTAGAA
 ACTATATAACTCAAAAGCCATATAGTGATGAGAAATTTAAGCTCAATTTTGA
 GAACTCAACTTTAGCCAA
 TGGTTGGGATAAAAAATAAAGAGCCTGACAATACGGCAATTTTATTTATCAAA
 GATGATAAATATTATCTG
 GGTGTGATGAACAAGAAAAATAACAAAATATTTGATGATAAAGCTATCAAAG
 AAAATAAAGGTGAAGGAT
 ATAAGAAAGTTGTATATAAACTTTTACCCGGTGCAAATAAAATGTTACCTAA
 GGTTTTCTTTCTGCTAA
 ATCTATAAATTTTTATAATCCTAGTGAAGATATACTTAGAATAAGAAACCACT
 CAACACATACAAAAAAT
 GGTAGTCCTCAAAAAGGATATGAAAACTTGAGTTTAATATTGAAGATTGCC
 GAAAATTTATAGATTTT

[0609]

ATAAACATTCTATAAGTAGGCATCCAGAGTGGAAAGATTTTGGATTTAGATTT
TCTGATACTAAAAATA
CAACTCTATAGATGAATTTTATAGAGAAGTTGAAAATCAAGGCTACAACTA
ACTTTTGAAAATATATCA
GAAAGCTATATTGATAGTTTAGTCGATGAAGGCAAATTATACCTATTCCAAAT
CTATAATAAAGATTTCT
CAGTATATAGTAAGGGTAAACCAAATTTACATACGCTATATTGGAAGGCGTT
GTTTGATGAGAGAAATCT
CCAAGATGTAGTATATAAATTAATGGTGAAGCAGAACTCTTCTATCGTAAA
CAATCAATACCTAAGAAA
ATCACTCACCCAGCCAAAGAGGCAATAGCTAATAAAAAACAAAGATAATCCTA
AAAAAGAGAGTATTTTG
AATATGATTTAATCAAAGATAAACGCTTTACTGAAGATAAGTTTTTCTTTCAC
TGTCTTATTACAATCAA
TTTCAAATCTAGTGGAGCTAATAAGTTTAATGATGAAATCAATTTATTGCTAA
AAGAAAAAGCAAATGAT
GTTTATATCCTAAGTATAGATAGAGGAGAAAGACATTTAGCTTACTATACTTT
GGTAGATGGTAAAGGAA
ACATTATCTGTAAGAATTAA

[0610]

[0611]

서열번호 69

[0612]

명칭: gi|754264888|gb|AJ157252.1| 크리스퍼-회합된 단백질 Cpf1, 하위 유형 프레프란(PREFRAN) [프란시셀라 필로미라지아]

[0613]

서열:

MKTNYHDKLAAIEKDRESARKDWKKINNIKEMKEGYLSQVVHEIAKL VIGYNAI
VVFEDLNFGFKRGRFK
VEKQVYQKLEKMLIEKLNLYVFKDNEFDKAGGVLRAYQLTAPFETFKKMGKQT
GIYYVPADFTSKICPV
TGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKW
TIASFGSRLINFRNSD
KNHNWDTREVVYPTKELEKLLKDYSIEYGHGECIKAAIYAENDKKFFAKLTSILNS
ILQMRNSKTGTEDY

[0614]

LISPVADVNGNFFDSRHAPKNMPQDADANGAYHIGLKGLMLLYRIKNNQDGKK
LNLVIKNEEYFEFVQNR

[0615]

NKSSKI

[0616]

서열번호 70

[0617]

명칭: gi|438609|gb|L21188.1|HIV1NY5A 인간 면역결핍 바이러스 유형 1 인테그라아제 유전자, 3' 말단

[0618] 서열:

TTCCTGGACGGTATCGATAAAGCTCAGGAAGAACACGAAAAATACCACTCTA
 ACTGGCGCGCCATGGCTT
 CTGACTTCAACCTGCCGCCGTTGTTGCCAAGGAAATCGTGGCTTCTTGCGAC
 AAATGCCAATTGAAAGG
 TGAAGCTATGCATGGTCAGGTCGACTGCTCTCCAGGTATCTGGCAGCTGGACT
 GCACTCATCTCGAGGGT
 AAAGTTATCTGGTTGCTGTTACGTGGCTTCCGGATACATCGAAGCTGAAGT
 TATCCCGGCTGAAACCG
 GTCAGGAAACTGCTTACTTCTGCTTAAGCTGGCCGGCCGTTGGCCGGTTAAA
 ACTGTTCACTGACAA
 CGGTTCTAACTTCACTAGTACTACTGTTAAAGCTGCATGCTGGTGGGCCGGA
 TCAAACAGGAGTTCGGG
 ATCCCGTACAACCCGCAGTCTCAGGGCGTTATCGAATCTATGAACAAAGAGC
 TCAAAAAAATCATTGGCC
 AGGTACGTGATCAGGCTGAGCACCTGAAAACCGCGGTGCAGATGGCTGTTTT
 CATCCACAACCTTCAAACG
 TAAAGGTGGTATCGGTGGTTACAGCGCTGGTGAACGTATCGTTGACATCATC
 GCTACTGATATCCAGACT
 AAAGAACTGCAGAAACAGATCACTAAAATCCAGAACTTCCGTGTATACTACC
 GTGACTCTAGAGACCCGG
 TTTGGAAAGGTCCTGCTAAACTCCTGTGGAAGGGTGAAGGTGCTGTTGTTATC
 CAGGACAACCTCTGACAT
 CAAAGTGGTACCGCGTCGTAAAGCTAAAATCATTTCGCGACTACGGCAAACAG
 ATGGCTGGTGACGACTGC

[0619]

[0620] GTTGCTAGCCGTCAGGACGAAGACTAAAAGCTTCAGGC

[0621] 서열번호 71

[0622] 명칭: gi|438610|gb|AAC37875.1| 인테그라아제, 일부 [인간 면역결핍 바이러스 1]

[0623] 서열:

FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQ
 VDCSPGIWQLDCTHLEG
 KVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSNFTSTT
 VKAACWWAGIKQEFQ
 IPYNPQSQGVIESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYS
 AGERIVDIIATDIQT
 KELQKQITKIQNFRVYYRDSRDPVWKGPAKLLWKGEGAVVIQDNSDIKVPPRRK
 AKIIRDYGKQMAGDDC
 VASRQDED

[0624]

[0625] 서열번호 72

[0626] 명칭: gi|545612232|ref|WP_021736722.1| 유형 V 크리스퍼-회합된 단백질 Cpf1 [애시드아미노코커스 종, BV3L6]

[0627]

서열:

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRI
YKTYADQCLQLVQ
LDWENLSAAIDSYRKEKTEETRNLIEEQATYRNAIHDIYFIGRTDNLTDANKRH
AEIYKGLFKAELFNG
KVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFS AEDISTAIPHRIVQD
NFPKFKENCHIFTR
LITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISRE
AGTEKIKGLNEV
LNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTL
LRNENVLETAE
ALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKE
KVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAA LDQPLPTTLKKQEEKEILKSQDSSLGL
YHLLDWFAVDESN

[0628]

EVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKNFQMPTLASGW
DVNKEKNNGAILFVKN
GLYYLGIMPKQKGRYKALSFEPTTEKTSEGFDMYYDYFPDAAKMIPKCSTQLKA
VTAHFQTHHTPILLSN
NFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLS
KYTKTTSIDLSSLRP
SSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHH
GKPNLHTLYWTGLFS
PENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIDTLYQ
ELYDYVNHRLSHDLSD
EARALLPNVITKEVSHEIHKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYL
KEHPETPIIGIDRG
ERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAAARQAWSVVGTIK
DLKQGYLSQVIHEIV
DLMIHYQAVVVLENLNFQKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAE
KVGGVLNPYQLTDQFT
SFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLGFDLH
YDVKTGDFILHFKMN
RNL SFQRGLPGFMPAWDIVFEKNETQF DAKGTPFIAGKRIVPVIEHRFTGRYRD
LYPANELIALLEEK
IVFRDGSNILPKLLENDSDHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLN
GVCFDSRFQNPPEWPM
DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN

[0629]

서열번호 73

[0630]

명칭: gi|769142322|ref|WP_044919442.1| 유형 V 크리스퍼-회합된 단백질 Cpf1 [라크노스피라세 박테리움 MA2020]

[0631]

[0632] 서열:

MYYESLTKQYPVSKTIRNELPIGKTLDNIRQNNILESDVKKRQNYEHVKGILDEY
HKQLINEALDNCTL
PSLKIAAEIYLKNQKEVSDREDFNKTQDLLRKEVVEKLKAHENFTKIGKKDILDL
LEKLPSISEDYNAL

[0633]

ESFRNFYTYFTSYNKVRENLYSDKEKSSTVAYRLINENFPKFLDNVKSYPFVKTA
GILADGLGEEEQDSL
FIVETFNKTLTQDGIDTYSQVGKINSSINLYNQKNQKANGFRKIPKMMLYKQI
LSDREESFIDEFQSD
EVLIDNVESYGSVLIESLKSSKVSAFFDALRESKGKNVYVKNDLAKTAMSNIVFE
NWRTFDDLLNQEYDL
ANENKKKDDKYFEKRQKELKKNKSYSLEHLCNLSEDSCLNIENYIHQISDDIENIII
NNETFLRIVINEH
DRSRKLAKNRKAVKAIKDFLDSIKVLERELKLINSSGQELEKDLIVYSAHEELLVE
LKQVDSLNMTRNY
LTKKPFSTEKVKLNFNIRSTLLNGWDRNKETDNLGVLLKDGKYYLGIMNTSAN
KAFVNPPVAKTEKVFKK
VDYKLLPVPNQMLPKVFFAKSNIDFYNPSSEIYSNYKKGTHKKGNMFSLEDCHN
LIDFFKESISKHEDWS
KFGFKFSDTASYNDISEFYREVEKQGYKLTYTIDIDETYINDLIERNELYLFQIYNK
DFSMYSGKGLNLHT
LYFMMLFDQRNIDDVYKLNGEAEVFYRPASISEDELIHKAAGEEIKNKNPNRAR
TKETSTFSYDIVKDK
RYSKDKFTLHIPITMNFVDEVKRFNDVNSAIRIDENVNVIGIDRGERNLLYVV
VIDSKGNILEQISLN
SIINKEYDIETDYHALLDEREGGRDKARKDWNTVENIRDLKAGYLSQVVNVVAK
LVLKYNAIICLEDLNF
GFKRGRQKVEKQVYQKFEKMLIDKLNLYVIDKSREQTSPKELGGALNALQLTSK
FKSFKELGKQSGVIYY
VPAYLTSKIDPTTGANLFYMKCENVEKSKRFFDGFDFIRFNALENVFEFGFDYR
SFTQRACGINSKWTV
CTNGERIIKYRNPDKNNMFDEKVVVVTDENKLNFEQYKIPYEDGRNVKDMIISN
EEAEFYRRLYRLLQQT
LQMRNSTSDGTRDYIISPVKNKREAYFNSELSDGSVPKDADANGAYNIARKGLW
VLEQIRQKSEGEKINL
AMTNAEWLEYAQTHLL

[0634]

[0635] 서열번호 74

[0636] 명칭: gi|489130501|ref|WP_003040289.1| 유형 V 크리스퍼-회합된 단백질 Cpf1 [프란시셀라 툴라렌시스 (Francisella tularensis)]

[0637]

서열:

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQ
NLIDAKKGQESDLIL
WLKQSKDNGIELFKANSDITDIDEALEIISFKGWTTYFKGFHENRKNVYSSNDIP
TSIIYRIVDDNLPK
FLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEI
ANFNYYLNQSGITK
FNTIIGGKVFNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVT
TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDL SQQV
FDDYSVIGTAVLEY
ITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLAEFNKHRDIDKQCRFE
EILANFAAIPMIFD
EIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNLLHKLKIFHI
SQSEDKANILDKDEH
FYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYL
GVMNKKNNKIFDDKAIVENKGEYKKIVYKLLPGANKMLPKVFFSAKSIKFYNP
SEDILRIRNHSTHTKN
GSPQKGYEKFENIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYRE
VENQGYKLTFENIS
ESYIDSVVNQGKLYLFQIYNKDFSAYS KGRPNLHTLYWKALFDERNLQDVVYKL
NGEAELFYRKQSIPKK
ITHPAKEAIAKNKNDPNKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKF
NDEINLLLKEKAND

[0638]

VHILSIDRGERHLAYYTLVDGKGNIQKQDTFNIIGNDRMKTNYHDKLAAIEKDRD
SARKDWKKINNIKEM
KEGYLSQVVHEIAKL VIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKL
NYLVFKDNEFDKTGG
VLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKS
QEFFSKFDKICYNLD
KGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVPYPTKEL
EKLLKDYSIEYGHGEC
IKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTEDYLISPVADVNGNFFDSRQ
APKNMPQDADANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0639]

[0640]

서열번호 75

[0641]

명칭: gi|502240446|ref|WP_012739647.1| 유형 V 크리스퍼-회합된 단백질 Cpf1 [[유박테리움 엘리겐스 (Eubacterium) eligens]]

[0642]

서열:

MNGNRSIVYREFVGVIPVAKTLRNLRPVGHTQEHIIQNGLIQEDELREQEKSTELK
NIMDDYYREYIDKS
LSGVTDLDFTLFELMNLVQSSPSKDNKKALEKEQSKMREQICTHLQSDSNYKNI
FNAKLLKEILPDFIK
NYNQYDVKDKAGKLETALFNGFSTYFTDFFEKRKNVFTKEAVSTSIAYRIVHE
NSLIFLANMTSYKKIS
EKALDEIEVIEKNNQDKMGDWELNQIFNPDFYNMVLIQSGIDFYNEICGVVNAH
MNLVCQQTKNNYNLFK
MRKLHKQILAYTSTSFEVPMKFEDDMSVYNVNAFIDETEGNIIGKLKDIVNKY
DELDEKRIYISKDFY
ETLSCFMSGNWNLTGCVENFYDENIHAKGKSKEEKVKKAVKEDKYKSINDVND
LVEKYIDEKERNEFKN
SNAKQYIREISNIITDTETAHLEYDDHISLIESEEKADEMCKRLDMYMNMYHWA
KAFIVDEVLDREMFY
SDIDDIYNILENIVPLYNRVRYVTQKPYNKKIKLNFQSPTLANGWSQSKEFDN
NAIILIRDNKYYLAI

[0643]

FNAKNKPKKKIIQGNSDKKNDNDYKKMVYNLLPGANKMLPKVFLSKKGIETFK
PSDYIISGYNAHKHIKT
SENFDISFCRDLIDYFKNSIEKHAEWKRYEFKFSATDSYSDISEFYREVEMQGYRI
DWTYISEADINKLD
EEGKIYLFQIYNKDFEAENSTGKENLHTMYFKNIFSEENLKDIKLNQAEIFYRR
ASVKNPVKHKKDSV
LVNKTYKNQLDNGDVVRIPDDIYNEIYKMYNGYIKESDLSEAAKEYLDKVEV
RTAQKDIVKDYRYTVD
KYFIHTPITINYKVTARNNVNDMVVKYIAQNDDIHVIGIDRGERNLIYISVIDSHG
NIVKQKSYNLLNNY
DYKKKLVEKEKTREYARKNWKSGNIKELKEGYISGVVHEIAMLIVEYNAILAME
DLNYGFKRGRFKVER
QVYQKFESMLINKLNYFASKEKSVDEPGGLLKGYQLTYVPDNIKNLGKQCGVIF
YVPAAFTSKIDPSTGF
ISAFNFKSISTNASRKQFFMQFDEIRYCAEKDMFSFGFDYNNFDTYNITMGKTQW
TVYTNGERLQSEFNN
ARRTGKTKSINLTETIKLLLEDNEINYADGHDIRIDMEKMDDEDKKSEFFAQLLSLY
KLTVMQMRNSYTEAE
EQENGISYDKIISPVINDEGEFFDSNYKESDDKECKMPKADANGAYCIALKGL
YEVCLKIKSEWTEGDF
DRNCLKLPHAEWLDFIQNKRYE

[0644]

[0645]

서열번호 76

[0646]

명칭: gi|537834683|ref|WP_020988726.1| 유형 V 크리스퍼-회합된 단백질 Cpf1 [렙토스피라 이나다이 (Leptospira inada)]

[0647] 서열:
 MEDYSGFVNIYSIQKTLRFELKPVGKTLEHIEKKGFLKKDKIRAEDYKAVKKIIDK
 YHRAYIEEVFDSVL
 HQKKKKDKTRFSTQFIKEIKEFSELYYKTEKNIPDKERLEALSEKLRKMLVGAFK
 GEFSEEVAEKYKNLF
 SKELIRNEIEKFCETDEERKQVSNFKSFTTYFTGFHSNRQNIYSDEKKSTAIGYRII
 HQNLPKFLDNLKI

[0648]
 IESIQRRFKDFPWSDLKKNLKKIDKNIKLTEYFSIDGFVNVLNQKGIDAYNTILGG
 KSEESGEKIQGLNE
 YINLYRQKNNIDRKNLPNVKILFKQILGDRETKSFIPEAFPDDQSVLNSITEFAKYL
 KLDKKKKSIIAEL
 KKFLSSFNRYELDGIYLANDNSLASISTFLFDDWSFIKKSVSFKYDESVGDPKKKI
 KSPLKYEKEKEKWL
 KQKYYTISFLNDAIESYSKSSQDEKRVKIRLEAYFAEFKSKDDAKKQFDLLERIEEA
 YAIVEPLLGAEYPR
 DRNLKADKKEVGKIKDFLDSIKSLQFFLKPLL SAEIFDEKDLGFYNQLEGYYEEID
 SIGHLYNKVRNYLT
 GKIYSKEKFKLNFENSTLLKGWDENREVANLCVIFREDQKYVLGVMDKENNTIL
 SDIPKVKPNELFYEKM
 VYKLIPTPHMQLPRIIFSSDNL SIYNPSKSILKIREAKSFKEGKNFKLKDCHKFIDFY
 KESISKNE DWSR
 FDFKFSKTSSYENISEFYREVERQGYNLDFKKVSKFYIDSLVEDGKLYLFQIYNKD
 FSIFSKGKPNLHTI
 YFRSLFSKENLKDVCLKLNGEAEMFFRKKSYNYDEKKKREGHHPELFEKLKYPIL
 KDKRYSEDKFQFHLP
 ISLNFKSKERLNFNLKVNEFLKRNKDINIIGIDRGERNLLYLVMINQKGEILKQTLL
 DSMQSGKGRPEIN
 YKEKLQEKEIERDKARKSWGTVENIKELKEGYLSIVIHQISKLMVENNAIVVLED
 LNIGFKRGRQKVERQ
 VYQKFEKMLIDKLNFLVFKENKPTEPGGVLKAYQLTDEFQSFEKLSKQTGFLFY
 VPSWNTSKIDPRTGFI
 DFLHPAYENIEKAKQWINKFDSIRFNSKMDWFEFTADTRKFSENMLGKNRVW
 VICTTNVERYFTSKTAN
 SSIQYNSIQITEKLEKELFVDIPFSNGQDLKPEILRKNDVFFKSLLFYIKTTLSLRQN
 NGKKGEEEEKDFI
 LSPVVD SKGRFFNSLEASDDEPKDADANGAYHIALKGLMNLLVLNETKEENLSR
 PKWKIKNKDWLEFVWE

[0649] RNR

[0650] 서열번호 77

[0651] 명칭: gi|739008549|ref|WP_036890108.1| 유형 V 크리스퍼-회합된 단백질 Cpf1 [포르피로모나스 크레비오리카
 니스(Porphyromonas crevioricanis)]

[0652]

서열:

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRAESYRRVKKIIDT
YHKVFIDSSLENMA
KMGIEKAMLSQFCELYKKDHRTEGEDKALDKIRAVLRGLIVGAFTGVCGR
ENTVQNEKYESLFKEK
LIKEILPDFVLSTEAESLPFSVEEATRSKKEFDSFTSYFAGFYENRKNISTKQSTA
IAYRLIHENLPK
FIDNILVFQKIKEPIAKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEK
YNALIGKIVTEG
DGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQILSDREQLSYLPESFEKDEELL
RALKEFYDHIAEDIL
GRTQQLMTSISEYDLRSRIYVRNDSQLTDISKMLGDWNAIYMARERAYDHEQAP
KRITAKYERDRIKALK
GEESISLANLNSCIAFLDNVRDCRVDTYLSTLGQKEGPHGLSNLVENVFASYHEA
EQLLSPYPEENLI
QDKDNVVLIKNLLDNISDLQRFLKPLWGMGDEPKDERFYGEYNYIRGALDQVI
PLYNKVRNYLTRKPYS
TRKVKNLFGNSQLLSGWDNRNEKDNSCVILRKGQNFYLAIMNNRHKRSFENKM
LPEYKEGEPYFEKMDYK
FLPDPNKMLPKVFLSKKGIEIYKPSPKLLEQYGHGTHKKGDTFSMDDLHELIDFF
KHSIEAHEDWKQFGF
KFSDTATYENVSSFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFS
PCSKGTPNLHTLYWR
MLFDERNLADVYKLDGKAIEFFREKSLKNDHPTHPAGKPIKKKSRQKKGEESLF
EYDLVKDRRYTMDKF
QFHPITMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYICVIDSRGT
ILDQISLNTINDIDY

[0653]

HDLLESRDKDRQQEHRNWQTIEGIKELKQGYLSQAVHRIAELMVAYKAVVALE
DLNMGFKRGRQKVESSV
YQQFEKQLIDKLNLYVDKKKRPEDIGLLRAYQFTAPFKSFKEMGKQNGFLFYIP
AWNTSNIDPTTGfVN
LFHVQYENVDKAKSFFQKFDsisYNPKKDWFEFAFDYKNFTKKAEGSRSMWILC
THGSRIKNFRNSQKNG
QWDSEEFALTEAFKSLFVRYEIDYTADLKTAIVDEKQKDFVDLLKLFKLTVM
RNSWKEKDLDYLISPV
AGADGRFFDTREGNKS LPKDADANGAYNIALKGLWALRQIRQTSEGGKLLAIS
NKEWLQFVQERSYEKD

[0654]

[0655]

서열번호 78

[0656]

명칭: gi|517171043|ref|WP_018359861.1| 유형 V 크리스퍼-회합된 단백질 Cpf1 [포르피로모나스 마카카에]

[0657]

서열:

MKTQHFFEDFTSLYSLSKTIRFELKPIGKTLENIKKNGLIRRDEQRLDDYEKLKKV
 IDEYHEDFIANILS
 SFSFSEEILQSYIQNLSESEARAKIEKTMRDTLAKAFSEDERYKSIFKKELVKKDIP
 VWCPAYKSLCKKF
 DNFTTSLVPFHENRKNLYTSNEITASIPYRIVHVNLPKFIQNIEALCELQKKMGAD
 LYLEMMENLRNVWP
 SFVKTPDDLCLNKTYNHLMVQSSISEYNRFVGGYSTEDGTKHQGINEWINIYRQR
 NKEMRLPGLVFLHKQ
 ILAKVDSSSFISDTLENDQVFCVLRQFRKLFWNTVSSKEDDAASLKDLFCGLSG
 YDPEAIYVSDAHLAT
 ISKNIFDRWNYISDAIRRKTEVLMPRKKESVERYAEKISKQIKKRQSYSLAELDDL
 LAHYSEESLPAGFS
 LLSYFTSLGGQKYL VSDGEVILYEEGSNIWDEVLIAFRDLQVILDKDFTEKKLGK
 DEEAVSVIKKALDSA
 LRLRKFFDLLSGTGAEIRRDSSFYALYTD RMDKLKGLLKMYDKVRNYLTKKPYS
 IEKFKLHFDNPSLLSG

[0658]

WDKNKELNNLSVIFRQNGYYYLGIMTPKGKNLFKTL PKLGAEEMFYEKMEYKQ
 IAEPMLMLPKVFFPKKT
 KPAFAPDQSVVDIYNKKTFTKGQKGFNKKDLYRLIDFYKEALTVHEWKLFNFSF
 SPTEQYRNIGEFFDEV
 REQAYKVSMVNPASVYIDEAVENGKLYLFQIYNKDFSPYSGIPNLHTLYWKAL
 FSEQNQSRVYKLCGGG
 ELFYRKASLHMQD TTVHPKGISIHKKNLNKKGETSLFN YDLVKDKRFTEDKFFF
 HVPISINYKNKKITNV
 NQMVRDYIAQNDDLQIIGIDRGERNLLYISRIDTRGNLLEQFSLNVIESDKGDLRT
 DYQKILGDREQERL
 RRRQEWKSIESIKDLKDGYMSQVVHKICNMVVEHKAIVVLENLNL SFMKGRKK
 VEKSVYEKFERMLVDKL
 NYLVVDKKNLSNEPGGLY AAYQLTNPLFSFEELHRY PQSGILFFVDPWNTSLTDP
 STGFVNLLGRINYTN
 VGDARKFFDRFN AIRDGKGNILFDLDLSRFDVRVETQRKLWTLTTFGSRIAKSK
 KSGKWMVERIENLSL
 CFLELFEQFNIGYRVEKDLKKAILSQDRKEFYVRLIYLFNLMMQIRNSDGEEDYIL
 SPALNEKNLQFDSR
 LIEAKDLPVDADANGAYNVARKGLMVVQRIKRGDHESIHRIGRAQWLRYVQEGI
 VE

[0659]

서열번호 79

[0660]

명칭: 유니프로트 위치 상에서 발견되는 인테그라아제 단백질 서열. DNA 서열을 GenBank로부터 획득하였다.

[0661]

- [0662] 서열:
- TTTTTAGATGGAATAGATAAGGCCCAAGATGAACATGAGAAATATCACAGTA
ATTGGAGAGCAATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGA
AATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGA
CAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATTTAGAAG
GAAAAGTTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGA
AGTTATTCAGCAGAAACAGGGCAGGAAACAGCATATTTTCTTTTAAAATTA
GCAGGAAGATGGCCAGTAAAAACAATACATACTGACAATGGCAGCAATTTCA
- [0663]
- CCGGTGCTACGGTTAGGGCCGCTGTTGGTGGGCGGGAATCAAGCAGGAATT
TGGAATTCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAA
GAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAG
ACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGA
TTGGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACA
TACAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGT
TTATTACAGGGACAGCAGAAATCCACTTTGGAAAGGACCAGCAAAGCTCCTC
TGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA
GTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAACAGATGGCA
GGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAG
- [0664]
- [0665] 서열번호 80
- [0666] 명칭: sp|P04585|1148-1435
- [0667] 서열:
- FLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQ
VDCSPGIWQLDCTHLEGKVLVAVHVASGYIEAEVIPAETGQETAYFLLKLAGR
WPVKTIHTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKI
IGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQI
TKIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNSDIKVVP RRKAKIIRDY
GKQ MAGDDCVASRQDED
- [0668]
- [0669] 서열번호 81
- [0670] 징크 핑거 단백질을 특징화하는 단백질 도메인
- [0671] CX(2-4)CX(12)HX(3-5)H (예를 들어, X(2-4)는 XX 또는 XXX 또는 XXXX를 의미한다)
- [0672] 서열번호 82
- [0673] >gi|1616606|emb|X97044.1| 마우스 유방 종양 바이러스 5' LTR DNA
- ATGCCGCGCCTGCAGCAGAAATGGTTGAACTCCCGAGAGTGTCTTACACTTA
GGGGAGAAGCAGCCAAGG
GGTTGTTTCCACCCAGAACGACCCATCTGCGCACACACGGATGAGCCCGTC
AAACAAAGACATATTCAT
TCTCTGCTGCAAACCTTGGCATAGCTCTGCTTTGCCTGGGGCTATTGGGGGAAG
TTGCGGTTTCATGCTCGC
- [0674]

AGGGCTCTCACCCCTTGACTCTTTTAATAGCTCTTCTGTGCAAGATTACAATCT
 AAACAATTCGGAGAACT
 CGACCTTCCTCCTGAGGCAAGGACCACAGCCAACCTCCTCTTACAAGCCGCA
 TCGATTTAGTCCTTCAGA
 AATAGAAATAAGAATGCTTGCTAAAAATTATATTTTTACCAATGAGACCAAT
 CCAATAGGTCGATTATTA
 ATTACTATGTTAAGAAATGAATCATTATCTTTTAGTACTATTTTTACTCAAATT
 CAGAAGTTAGAAATGG
 GAATAGAAAAATAGAAAGAGACGCTCAGCCTCAGTTGAAGAACAGGTGCAAG
 GACTAAGGGCCTCAGGCCT
 AGAAGTAAAAAGGGGGAAGAGGAGTGCGCTTGTCAAAATAGGAGACAGGTG
 GTGGCAACCAAGGAACCTTAT
 AGGGGACCTTACATCTACAGACCAACAGACGCCCCCTTACCGTATACAGGAA
 GATATGACCTAAATTTTG
 ATAGGTGGGTCACAGTCAATGGCTATAAAGTGTTATACAGATCCCTCCCCTTT
 CGTGAAAGGCTCGCCAG
 AGCTAGACCTCCTTGGTGCGTGTTGTCTCAGGAAGAAAAAGACGACATGAAA
 CAACAGGTACATGATTAT
 ATTTATCTAGGAACAGGAATGAACTTTTGGAGATATTATACCAAGGAGGGGG
 CAGTGGCTAGACTATTAG
 AACACATTTCTGCAGATACTAATAGCATGAGTTATTATGATTAGCCTTTATTG
 GCCCAATCTTGTGGTTC
 CCAGGGTTCAAGTAGGTTTCATGGTCACAACTGTTCTTAAAAACAAGGATGT
 GAGACAAGTGGTTTCCTG
 GCTTGGTTTGGTATCAAATGTTTTGATCTGAGCTCTGAGTGTCTGTTTTCTTA
 TGTTCTTTTGGAATCT
 ATCCAAGTCTTATGTAAATGCTTATGTAAACCAAAGTATAAAAGAGTGCTGA
 TTTTTTGAGTAAACTTGC
 AACAGTCCTAACATTCACCTCTCGTGTGTTTGTGTCTGTTGCGCCATCCCGTCTC
 CGCTCGTCACTTATCC

[0675]

TTCACTTTCCAGAGGGTCCCCCGCAGACCCCGGTGACCCTCAGGTTGGCCG
 ACTGCGGCA

[0676]

[0677] 서열번호 83

[0678] >gi|1403387|emb|X98457.1| 마우스 유방 종양 바이러스 3' LTR

ATGCCGCGCCTGCAGCAGAAATGGTTGAACTCCCGAGAGTGCCTACACTTA
 GGAGAGAAGCAGCCAAGG
 GGTGTGTTCCACCAAGGACGACCCGTCTGCGTGCACGCGGATGAGCCCATC
 AGACAAAGACATACTCAT
 TCTCTGCTGCAAACCTTGGCATAGCTCTGCTTTGCCTGGGGCTATTGGGGGAAG
 TTGCGGTTTCGTGCTCGC
 AGGGCTCTCACCTTGATTCTTTTAATAACTCTTCTGTGCAAGATTACAATCT
 AAACGATTCGGAGAACT
 CGACCTTCCTCCTGGGGCAAGGACCACAGCCAACTTCCTCTTACAAGCCACA
 CCGACTTTGTCTTCAGA
 AATAGAAATAAGAATGCTTGCTAAAAATTATATTTTTACCAATGAGACCAAT
 CCAATAGGTCGATTATTA
 ATCATGATGTTTAGAAATGAATCTTTGTCTTTTAGCACTATATTTACTCAAATT
 CAAAGGTTAGAAATGG
 GAATAGAAAATAGAAAGAGACGCTCAACCTCAGTTGAAGAACAGGTGCAAG
 GACTAAGGGCCTCAGGCCT
 AGAAGTAAAAAGGGGAAAGAGGAGTGCGCTTGTCAAAATAGGAGACAGGTG
 GTGGCAACCAGGGACTTAT
 AGGGGACCTTACATCTACAGACCAACAGACGCCCCGCTACCATATACAGGAA
 GATACGATTTAAATTTTG
 ATAGGTGGGTCACAGTCAACGGCTATAAAGTGTTATACAGATCCCTCCCCCTT
 CGTGAAAGACTCGCCAG
 GGCTAGACCTCCTTGGTGTGTGTTAACTCAGGAAGAAAAAGACGACATGAAA
 CAACAGGTACATGATTAT
 ATTTATCTAGGAACAGGAATGAACTTCTGGGGAAAGATATTTGACTACACCG
 AAGAGGGAGCTATAGCAA

[0679]

AAATTATATATAATATGAAATATACTCATGGGGGTCGCATTGGCTTCGATCCC
 TTTTGAAACATTTATAA
 ATACAATTAGGTCTACCTTGCGGTTCCCAAGGTTTAAGTAAGTTCAGGGTCAC
 AAACCTGTTCTTAAAACA

[0680] AGGATGTGAGACAAGTGGTTTCCTGACTTGGT

[0681] 서열번호 84

[0682] >gi|119662099|emb|AM076881.1| 인간 면역결핍 바이러스 1 프로바이러스성(proviral) 5' LTR, TAR 요소 및 U3, U5 및 R 반복 영역, 클론 PG232.14

GGCAAGAAATCCTTGATTTGTGGGTCTACTACACACAAGGCTTCTTCCCTGAT
TGGCAAAACTACACACC
GGGACCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTG
CCAGTTGACCCAAAGGAA
GTAGAAGAGGCTAACCAAAGAGAAGACAACCTGTTTGCTACACCCTATGAGCC
TGCATGGAATAGAGGACG
AAGACAGAGAAGTATTAAGTGGCAGTTTGACAGCAGCCTAGCACGCAGAC
ACATGGCCCCGCGAGCTACA
TCCAGAGTATTACAAAGACTGCTGACACAGAAAAGACTTTCCGCTAGGACTT
TCCACTGAGGCGTTCCAG
GGGAGTGCTTAGGCAGGACTAGGAGTGGCCAACCCTCAGATGCTGCATAT
AAGCAGCTGCTTTTCGCC
TGTACTAGGTCTCTCTAGGTGGACCAGATCTGAGCCTAGGCGCTCTCTGGCTA
TCTAAGGAACCCACTGC
TTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCTAAGTAGTGTGTGCCCCGTCTG
TTGTGTGACTCTAGTAA
CTAGAGATCCCTCAGACCAACTTTAGTAGTGTAATAAATCTCTAGCAGTGGC
GCCCCAACAGGGACCCGA
AAGTGAAAGCAGGACCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGA
AAGTGCACTCGGCAAGAG
GCGAGAGCAGCGGCGACTGGTGAGTACGCCGAATTTTATTTTGACTAGCGGA
GGCTAGAAGGAGAGAGAT

A

서열번호 85

>gi|1072081|gb|U37267.1|HIV1U37267 인간 면역결핍 바이러스 유형 1 3' LTR 영역
ATGGGTGGCAAGTGGTCAGAAAGTAGTGTTAGAAGGCATGTACCTTAA
GACAAGGCAGCTATAGAT
CTTAGCCGCTTTTTAAAGAAAAGGGGGGACTGGAAGGGCTAATCACTCAC
AGAGAAGATCAGTTGAAC
CAGAAGAAGATAGAAGAGGCCATGAAGAAGAAAACAACAGATTGTTCCGTT
TGTTCCGTTGGGGACTTTC
CAGGAGACGTGGCCTGAGTGATAAGCCGCTGGGGACTTCCGAAGAGGCGTG
ACGGGACTTTCCAAGGCG
ACGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAA
GCAGCTGCTTTCTGCCTG
TACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAAC
TAGGGAACCCACTGCTT
AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCCGTCTGTT
GTGTGACTCTGGTATCT

AGA

서열번호 86 내지 99가 존재함

서열번호 100

[0689] neo의 세포 계놈 내로의 삽입을 위한 올리고(5' 및 3' HIV LTR의 완전한 서열 이용)
GACAAGACATCCTTGATTTGTGGGTCTATAACACACAAGGCTTCTTCCCTGAT
TGGCAAAACTACACACC
GGGACCAGGGACCAGATACCCACTGACCTTTGGATGGTGCTTCAAGCTAGTG
CCAGTTGACCCAAGGGAA
GTAGAAGAGGCCAATACAGGGGAAAACAACCTGTTTGCTCCACCCTATGAGCC
AGCATGGAATGGAAGATG
ACCATAGAGAAGTATTAAAGTGGAAGTTTGACAGTATGCTAGCACGCAGACA
CCTGGCCCCGCGAGCTACA

[0690]
TCCGGAGTACTACAAAACTGCTGACATGGAGGGACTTTCCGCTGGGACTTT
CCATTGGGGCGTTCCAGG
AGGTGTGGTCTGGGCGGGACAAGGGAGTGGTCAACCCTCAGATGCTGCATAT
AAGCAGCTGCTTTTCGCT
TGTACTGGGTCTCTTTAGGTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTA
CCTGAGGAACCCACTGC
TTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCTAAGTAGTGTGTGCCCCGTCTG
TTGTGTGACTCTGGTAA
CTAGAGATCCCTCAGACCCTTTTGGTAGTGTGGAAAATCTCTAGCAGATGATT
 GAACAAGATGGATTGCAC
 GCAGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCAC
 AACATGGGTGGCAAGTGGTCAG
 AAAGTAGTGTGGTTAGAAGGCATGTACCTTTAAGACAAGGCAGCTATAG
 ATCTTAGCCGCTTTTTTAAAAGAAAAG
 GGGGGACTGGAAGGGCTAATTCACTCACAGAGAAGATCAGTTGAACCAG
 AAGAAGATAGAAGAGGCCATGAAG
 AAGAAAACAACAGATTGTTCCGTTTGTTCGTTGGGGACTTTCCAGGAG
 ACGTGGCCTGAGTGATAAGCCGCTGGG
 GACTTTCCGAAGAGGCGTGACGGGACTTTCCAAGGCGACGTGGCCTGGG
 CGGGACTGGGGAGTGGCGAGCCCTC
 AGATGCTGCATATAAGCAGCTGCTTTCTGCCTGTACTGGGTCTCTCTGGT
 TAGACCAGATCTGAGCCTGGGAGCTCT
 CTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTG
 AGTGCTTCAAGTAGTGTGTGCCCCGTCTG
 TTGTGTGACTCTGGTATCTAGA

[0691]
 [0692] 첫번째 5'LTR을 밑줄표시하고, 일반 텍스트는 neo이며, 3'LTR은 굵은글씨로 표시된다(1179 bp)

[0693] 서열번호 101

[0694] 5'LTR 및 3'LTR의 약어 버전으로 그 안에 neo 서열을 갖는 것(224 bp)

[0695] 첫번째 5'LTR을 밑줄표시하고, 일반 텍스트는 neo이며, 3'LTR은 굵은글씨로 표시된다.

[0696] GACAAGACATCCTTGATTTGTGGGTCTATAACACACAAGGCTTCTTCCCTGAT
TGGCAAAACTACACACCATGATTGAACAAGATGGATTGCAC

GCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCAC
AACTTAAGCCTCAATAAAGCTTGCCCTTGAGTGCTTCAAGTAGTGTGTGCC
CGTCTG
TTGTGTGACTCTGGTATCTAGA

- [0697]
- [0698] **서열번호 72에 대하여**
- [0699] Genbank 단백질 ID: WP_021736722.1
- [0700] NR 데이터베이스로부터의 NCBI 단백질 GI 또는 국소 GI(WGS 데이터베이스로부터 유래된 단백질의 경우): 545612232
- [0701] WGS 데이터베이스에서 Contig ID: AWUR01000016.1
- [0702] Contig 설명: 애시드아미노코커스 종 BV3L6 contig00028, 전체 게놈 샷건 서열
- [0703] 단백질 완전성: 완성
- [0704] 실험적으로 분석된 단백질: 8
- [0705] 비-중복 세트: nr
- [0706] 유기체: 애시드아미노코커스_sp_BV3L6
- [0707] 분류학: 박테리아, 페르미쿠테스(Firmicutes), 네가티브쿠테스 (Negativicutes), 셀레노모나달레스 (Selenomonadales), 애시드아미노코카세아 (Acidaminococcaceae), 애시드아미노코커스, 애시드아미노코커스 종 BV3L6
- [0708] **서열번호 73에 대하여**
- [0709] Genbank 단백질 ID: WP_044919442.1
- [0710] NR 데이터베이스로부터의 NCBI 단백질 GI, 또는 국소 GI(WGS 데이터베이스로부터 유래된 단백질의 경우): 769142322
- [0711] WGS 데이터베이스에서 Contig ID: JQKK01000008.1
- [0712] Contig 설명: 라크노스피라세 박테리움(Lachnospiraceae bacterium) MA2020 T348DRAFT_스캐폴드00007.7_C, 전체 게놈 샷건서열
- [0713] 단백질 완전성: 완성
- [0714] 실험적으로 분석된 단백질: 9
- [0715] 비-중복 세트: nr
- [0716] 유기체: 라크노스피라세_박테리움_MA2020
- [0717] 분류학: 박테리아, 페르미쿠테스, 클로스트리디아, 라크노스피라세, 미분류된 라크노스피라세, 라크노스피라세 박테리움 MA2020
- [0718] **개시된 조성물 및 방법에서 사용될 수 있는 추가 핵산 서열 및 단백질 서열 - CPF 1 배열.** 서열번호 86-92; 차트의 위에서부터 바닥까지의 순서

```

g1:5456612282 ref:WP_021736722.1
g1:517371043 ref:WP_038350861.1
g1:502240445 ref:WP_012739647.1
g1:537034603 ref:WP_029006726.1
g1:760142322 ref:WP_044015442.1
g1:489130501 ref:WP_083046289.1
g1:739008540 ref:WP_036630108.1

```

PKQTLSQNTLAFPIEAFPKQDEYIDSPCKYKTLIRN-----ENVLTATGALPME-----LA
HKQTLLAKYDSTFISDTLENDGQVFCVLRQKFLPNTYSSK-EDDAASLKKLFCG-----LA
HKQTLLAVDSTFSEVPMKFIENYQVNAIPIDETEN-----KQIGLGLKIRN-----YQ
PKQTLSQDRETHSPFIDEAPDDQGLVLTGPAKYLKKKK-----KKGISLAEKAKFSL-----FM
HKQTLLADREAE-----FIDEFDSQVLTONTVSYSSVILIESL-----TSKVSAPFDALP
PKQTLSQDSTSPFIVMLDEDSQVLTITMQSFVEQIAPKTVKEESIKETLSLLPBDLKKAL
YKQTLSQDREGLSVLPSFPEKDEELLRALKFVYDHIAEGLIGETQRLM-----TSIG

g1|545612232|ref|WP_021736722.1|
g1|517171043|ref|WP_018359851.1|
g1|502240446|ref|WP_012739647.1|
g1|537834683|ref|WP_020988726.1|
g1|769142322|ref|WP_044919442.1|
g1|489130501|ref|WP_003040289.1|
g1|739088549|ref|WP_036890108.1|

g1|545612232|ref|WP_021736722.1|
g1|517171043|ref|WP_018359851.1|
g1|502240446|ref|WP_012739647.1|
g1|537834683|ref|WP_020988726.1|
g1|769142322|ref|WP_044919442.1|
g1|489130501|ref|WP_003040289.1|
g1|739088549|ref|WP_036890108.1|

g1|545612232|ref|WP_021736722.1|
g1|517171043|ref|WP_018359851.1|
g1|502240446|ref|WP_012739647.1|
g1|537834683|ref|WP_020988726.1|
g1|769142322|ref|WP_044919442.1|
g1|489130501|ref|WP_003040289.1|
g1|739088549|ref|WP_036890108.1|

g1|545612232|ref|WP_021736722.1|
g1|517171043|ref|WP_018359851.1|
g1|502240446|ref|WP_012739647.1|
g1|537834683|ref|WP_020988726.1|
g1|769142322|ref|WP_044919442.1|
g1|489130501|ref|WP_003040289.1|
g1|739088549|ref|WP_036890108.1|

g1|545612232|ref|WP_021736722.1|
g1|517171043|ref|WP_018359851.1|
g1|502240446|ref|WP_012739647.1|
g1|537834683|ref|WP_020988726.1|
g1|769142322|ref|WP_044919442.1|
g1|489130501|ref|WP_003040289.1|
g1|739088549|ref|WP_036890108.1|

g1|545612232|ref|WP_021736722.1|
g1|517171043|ref|WP_018359851.1|
g1|502240446|ref|WP_012739647.1|
g1|537834683|ref|WP_020988726.1|
g1|769142322|ref|WP_044919442.1|
g1|489130501|ref|WP_003040289.1|
g1|739088549|ref|WP_036890108.1|

g1|545612232|ref|WP_021736722.1|
g1|517171043|ref|WP_018359851.1|
g1|502240446|ref|WP_012739647.1|
g1|537834683|ref|WP_020988726.1|
g1|769142322|ref|WP_044919442.1|
g1|489130501|ref|WP_003040289.1|
g1|739088549|ref|WP_036890108.1|

GIDLTHIFISHK-KLETISGALCDHDTLALALYERRISELTGKIT-----KS
GYDPEATVYSDA-HLATISKMIPOBNYISDAIRAKTEVLMP---BKKEBVERYAEKISKG
ELDEKRIYISKDF-YETLSCPMGNNLITGCVENFYDENIHAKGSKS-----EEKVYKA
RYELDGIYLANDNGLASISTFLFDWSPFKKSVGFKYDESVDGPKKKIKGPKLYEKEKEK
EKGKNNVYKNDLAKTAMGNVFNHRTFHDLLNDEYDLANENKCKDDKYPKRDKELEKK
KLDLSKIYFKNDKSLTDLSDQVFDYSVISITAVLEYITDQIAPKNLWPSKKEGELIAKK
EYDLSRIYVNRDGLTDLISKNNLGDWNAIYMAHREYDHEDAPKAITAKYERDRIKALKG

AKKQVGRSLKHEDIN-----LQETISAAGKEL-----SE-----AFYQKTSF
IKKRSYSLAELDOLLANYSEGLPAGFS-----LLSYFTLGGGKYLVSQGEVILYEIGN
YKEDKYKSEINDVNDLVEKYIDEKERNEFYNSNAKQYI-----REISN
NLKQKYITISFLNDATESYKSKSQDEKRVKIR-LEAYFAEFKSKS-----DDAKKQPD
N---KQSYLHLCHLS-----EDSCHL-JENYI-----HQISB
TEKAKYLSLETIKLALEEPKNHRDIDKQCRF-----EETLANFAAI-----P---M
E---ESISLANLNSCI-----AFLENVDRDVR---DTYLSLGGK-----EGPHGLSN

ILSHAH-----AALDQPLP-----TTLKKGEKEILKSQDLSLLGLYHLLDWA---
JWDEVLIAFRDLQVILKDFY-----EKKLKGEKAVSVIKKALDGLRLKQFDLLG---
ITDTETA-----HLEYD-----DHISLISEEKADENKKBLOMYNNMYWAKAF---
LLERIEEAYAIPEPLGAEP-----RPNLKADEKKEVGKIDFLDSIKSLDFLKPIL---
DIENIINNE-----TFLRIVINE-HDRSRHLAKNRKAVKAIKDFLDSIKVLERELKIN---
IFDE-IAGMKDNLAQTGKYNQKQKDLQASAEQDVKAIDKLDQTNMLHKLKIFHTG
LVENVFASYHEAEQLSFPYP-----EENMLI-----QKDNVVLIKNLLONISDLQFLKPLW--

-----VQSENEVDPEFSAULTGKLEMEPSLSFYNNKARNYATKKPYSEKFLNFMPTLA
-----GTGAEIRRDSFYALYTDMDKIKGLKMYDKVRNYLTKKPYSTKFLHFDNPSLL
-----IVDEVLDRDEMFYGDIDDIYNIENIYPLYNVRNYYTQKPYNSKKIKLNFQEPFLA
-----SABIFDEKDLGPFYNGLEQYEEIDSIGHLNVRNYYLTKKPYSEKFLNFMPTLA
-----SSGDELEKDLIVYSAHEELLVELKRVDSLYNMTRNYLTKKPYSTKFLNFMPTLA
QSEUKANILDKDENFYLYVFEELCYFLANIVPLYNKIRNVTQKPYSEKFLNFMPTLA
-----QMSDEPKDERFYGEVNYIRGALDQVILPLYNKVRNYLTKKPYSTRKVLNFMPTLA

SQWQVNEKKNNGAILFVKNGLYYLGIMPKDKGRY-KALSFTEKTSSEGFQKMYDYFPD
SQWQVNEKKNNGAILFVKNGLYYLGIMPKDKGRY-KALSFTEKTSSEGFQKMYDYFPD
SQWQVNEKKNNGAILFVKNGLYYLGIMPKDKGRY-KALSFTEKTSSEGFQKMYDYFPD
SQWQVNEKKNNGAILFVKNGLYYLGIMPKDKGRY-KALSFTEKTSSEGFQKMYDYFPD
SQWQVNEKKNNGAILFVKNGLYYLGIMPKDKGRY-KALSFTEKTSSEGFQKMYDYFPD
SQWQVNEKKNNGAILFVKNGLYYLGIMPKDKGRY-KALSFTEKTSSEGFQKMYDYFPD
SQWQVNEKKNNGAILFVKNGLYYLGIMPKDKGRY-KALSFTEKTSSEGFQKMYDYFPD

AAKMIKPCSTGLKAVTAHFQHTHTPIILBNFIEPLEITKEIYDLNNPEKEPKKQTAYA
PMLMLPKVFPFKKTKPA-----FAP-----DQSVVDIYNKKT-----K
ANKMLPKVFLSKKQIEF-----FKP-----SDYIISGYNAHKH-----I
PMDLPRJIFGSDNLSI-----YNP-----GKSLIKIREAKSF-----K
PMDLPRJIFGSDNLSI-----YNP-----GKSLIKIREAKSF-----K
ANKMLPKVFPFAKSNIDF-----YNP-----SEIYENYKKGTH-----K
ANKMLPKVFPFAKSNIDF-----YNP-----SEIYENYKKGTH-----K
PMDLPRJIFGSDNLSI-----YNP-----GKSLIKIREAKSF-----K

KKTGQKQYR-----EALCKWIDFTBDFLSKYTKTTSIDLSSLAPSSQYKDLGEYYREL
TGG-----KGFNKKDLYNLTDFYKEALYH-EWKLPM-FSPSPTEQYRNLIGEPFDEY
KTS-----ENFDISFCRDLIDYFKNIGIEKHAERNYK-FKFSATGYSYSISEFYREV
EGK-----NPKLKDCKPFDYFKEISIKNEDWSPFD-FKFSATGYSYSISEFYREV
KGN-----NPKLKDCKPFDYFKEISIKNEDWSPFD-FKFSATGYSYSISEFYREV
HNGSPQKQYKFEFNIEDCPKFDYFKEISIKNEDWSPFD-FKFSATGYSYSISEFYREV
KGD-----TFSDDLHELIDFFKHSIEAHEWKKQFG-FKFSATGYSYSISEFYREV

[0720]

g1 545612232 ref WP_021736722.1	NPLLYHISFQRIAEKIMDAVETGKLYLFGIYNKDFPAKHGKPNLHTLYWTKLPSPENL
g1 517171043 ref WP_018359861.1	REDAAYKSMVWVPASYIDEAVENGKLYLFGIYNKDFSPYSGKIPNCHLYWTKLPSEOND
g1 502240446 ref WP_012739647.1	EMQGYRTDHTYISEADINKLDEEGKLYLFGIYNKDFPENGSTGKENLHTMYFKNIFSEENL
g1 537834683 ref WP_028988726.1	ERQGYNLQFKKVSFYIDSLVEDGKLYLFGIYNKDFGIFSKGKPNLHTIYFRELFSKENL
g1 769142322 ref WP_044919442.1	EKOGYKLYTDTIDETYNIDLERNELYLFGIYNKDFSMYSKGRNLHTLYWTKLPDGRNI
g1 489138581 ref WP_083048289.1	EADGYKLYTFENISEGYIDGVNKGKLYLFGIYNKDFSAYSKGRPNLHTLYWTKLPDERNL
g1 739808549 ref WP_036898108.1	EDDGYKLSFRKVSSEYVGLDGGKLYLFGIYNKDFSPGSGTPNCHLYWTKLPDERNL
g1 545612232 ref WP_021736722.1	AKTSIKLNGGAEFLYRPSKRMKR--MAHRLGKKHKKLX-----DQKTPIDPTLYQE
g1 517171043 ref WP_018359861.1	S-RVYKLCGGGELFYRKASLHMDDTTVHPKQISIKKK--
g1 502240446 ref WP_012739647.1	KDITIKLNGGAEFLYRNASVKNPVK--HKDSVLNKTYYKQLDNGGVYKIPIDDIYNE
g1 537834683 ref WP_028988726.1	KDVC LKINGAEENFFRKSSINYDEKK-----R-----
g1 769142322 ref WP_044919442.1	DDVYKLNAGEAEVFRPASISEDELIIHKAGEELKKNNP-----
g1 489138581 ref WP_083048289.1	DDVYKLNAGEAEFLYRKDSIPK-K-IHPAKEALANKM-----
g1 739808549 ref WP_036898108.1	ADVIVKLDGKACIFFRKSLKNDH-PHPAGKPIKKD-----
g1 545612232 ref WP_021736722.1	LYOYVNHRLS-HDLSDEARALLPNVITREVSHEIHKDRFTSOKFFHVPITLNYDAANS
g1 517171043 ref WP_018359861.1	-----LNKKGETSLFNVDVKDKRFTEDKFFHVPISINYKAK-K
g1 502240446 ref WP_012739647.1	IYKMYNGYIKESGLSAAKEYLDKVEVKTAKQDIYKDYRYTVQKYFHTPTITINYKYT-A
g1 537834683 ref WP_028988726.1	-----EGHMPFLFEKLYPIKDKRYSDEKFFHVPISLNFPSK-E
g1 769142322 ref WP_044919442.1	-----NRARTKETSIFSYDVKDKRYSDEKFFHVPITMNFQVD-E
g1 489138581 ref WP_083048289.1	-----KDNPKCESVFEYDLKDKRFTEDKFFHVPITINFKSS-G
g1 739808549 ref WP_036898108.1	-----RQKKGEGLFEYDLVKDARYTNOXFFHVPITMNFKCS-A
g1 545612232 ref WP_021736722.1	PSKFNQRYNAYLK-EHPETPIIGIDRGERNLIIYITVIDSTGKILEDRSLNTIG-----Q
g1 517171043 ref WP_018359861.1	ITNNGHVRDYIA-QNDOLQIIGIDRGERNLIIYISATOTRNLLEQFSLNVEISDCKDLR
g1 502240446 ref WP_012739647.1	ANVNVGMVYKYIA-QNDOTIYIGIDRGERNLIIYISVIOHGNIVKQKSYNILN-----M
g1 537834683 ref WP_028988726.1	RLANPLKVMNPLK-RKNDINIIGIDRGERNLIIYVINGKGETLRYLLDSMSGKQAPPE
g1 769142322 ref WP_044919442.1	VKRFNDAVNSAIR-TDMVNIYIGIDRGERNLIIYVVIDSKENILEQISLNSITMKYDIE
g1 489138581 ref WP_083048289.1	ANKYNDITNL LKKEKANDVHILSIGHGERMLAYTTLVDGKONI KDDTFNI IGNDR--MK
g1 739808549 ref WP_036898108.1	GSKVNDMMYNAHIR-EAKDMVYIGIDRGERNLIIYICVIDSGTILDQISLNTIN-----D
g1 545612232 ref WP_021736722.1	FDYQKKLNNRREKVAARQAWSSVGTIKDKQGYLGGVTHIEIVLMIHYQAVVLEALNF
g1 517171043 ref WP_018359861.1	TDYQKILGDEQERLNRQQWKSIESEKLDQGYNSGVVHKICNMVVENHATYVLEALNF
g1 502240446 ref WP_012739647.1	VDYKKLVEREKTRTYARNWKSIGIKELKEGYISGVVHIEAMLIVEYNAIIMEDLNY
g1 537834683 ref WP_028988726.1	INYKELQEREIERDKARKSWGTVENIKELKEGYLSVIMQISKLHVENNAIIVLEDLNI
g1 769142322 ref WP_044919442.1	TYHALLDEREGGRDKARKDNTVENIRDLKAGVLSQVNVAAKLVLKYNAIICLEDLNF
g1 489138581 ref WP_083048289.1	TMVHDLAAIEKDRQSAARKDWKIKINIKEMKEGYLGGVTHIEIAKLVIEYNAIIVFEDLNF
g1 739808549 ref WP_036898108.1	IDYHDLLESARKDRQGRHMQTIEGTKEKQGYLGGAVHIAELNVAYKAVVALEDLNM
g1 545612232 ref WP_021736722.1	GFPSKRTGTAEKAYQQFEKMLIDKLNCLVLKDY-----PAEKVGGVLNPFYQLTQGFSPA
g1 517171043 ref WP_018359861.1	SFMKGRKK--VEKSYVEKEFERMLVDKLNLYLVDKKK-----LSNEGGLVAAQGLTNPLFSPE
g1 502240446 ref WP_012739647.1	GFKRGRFK--VERQVYQKFEKMLIDKLNLYLVDKKK-----SVDEPGGLKQYGLTYYPONIK
g1 537834683 ref WP_028988726.1	GFKRGRQK--VERQVYQKFEKMLIDKLNLYLVDKKK-----KPTPEGGVLKAYQLTDEPQSP
g1 769142322 ref WP_044919442.1	GFKRGRQK--VEKQVYQKFEKMLIDKLNLYLVDKSRQTSFKELGGALNALQLTSKPSKPK
g1 489138581 ref WP_083048289.1	GFKRGRFK--VEKQVYQKLEKMLIDKLNLYLVDKKK-----EFOKTGGLVAAQGLTAPFETFK
g1 739808549 ref WP_036898108.1	GFKRGRQK--VESSYVQKFEKMLIDKLNLYLVDKKK-----RPEIGGLVAAQGLTAPFETFK
g1 545612232 ref WP_021736722.1	KNG---TQSGFLFYYPAPYTSKIDPLTGFPDFFVWKTIKNN--ESRKNFLGFPFLNYPVKT
g1 517171043 ref WP_018359861.1	ELMRYPQSGILFFYDPWNTSLTDPSTGFVNLGRINYNV--GDARKFFDRFNAIRYDGGK
g1 502240446 ref WP_012739647.1	NLG--KQGVIFVYPAATFTSKIDPSTGFISAFNPK--SISTNASKQFMQFDEIRYCAEK
g1 537834683 ref WP_028988726.1	KLS--KOTGFLFYYPWNTSKIDPSTGFISAFNPK--YENI-EKAKQINKFDSIRFNSKM
g1 769142322 ref WP_044919442.1	ELG--KQSGVIFVYYPAYLTSKIDPSTGFANLPHK--CENY-EKSKNFPDGFDFINFALE
g1 489138581 ref WP_083048289.1	KNG---KQGTITYYYPAGFTKICPYTGFPVNLQPK--YESV-SKSGFFPKFQKICYNLDR
g1 739808549 ref WP_036898108.1	ENG--KQGFLEFYPAWNTSNIDPTTGFPVNLPHVQ--YENV-DKAKSFFPKFQKISYNPKK
g1 545612232 ref WP_021736722.1	TGRYR-GLYPANELIALLEEKGIYFADGSNLPKLL-----ENDDSHAIDTMVALIRSVLQM
g1 517171043 ref WP_018359861.1	KMMVERIENLSLCLFELFEQFNIGYRVKDLKKAIL-----SGDRAKEFYVRLIYVNLMMOI
g1 502240446 ref WP_012739647.1	TGKTK-SINLTETIKILLFQNEINYADGHDIRIGMEKMDDEKKSEFFAQILSLYKLTVMQ
g1 537834683 ref WP_028988726.1	SDQYN-SIGITEKLKELFVD--IPFGSGGLKPEIL-----RKNDVAFPKSLFYKTTLSL
g1 769142322 ref WP_044919442.1	MFDEK-VVVVTDENKMLFEQYKIPYEDGSNNVDMII-----SNEEAEFYARLYRLDQTLQM
g1 489138581 ref WP_083048289.1	NNQTA-EVYPTKELEKLKQYSIEYGHSECTAAIC-----GESDKKFAKLTSLNTILQM
g1 739808549 ref WP_036898108.1	QNDSE-CPALTEAFKSLFYRYEIOYTA--DLKTAIV-----DEKQKDFFYDLLKLFKLTVMQ
g1 545612232 ref WP_021736722.1	RNSNAA-----TGEDYINSVPRDNGVCFDSRF-----QNPENPMADANGAYHIAL
g1 517171043 ref WP_018359861.1	RNS-----DGEEDYILSPALNEKMLQFOSRLI-----EAKDLFVDAANGAYNVAR
g1 502240446 ref WP_012739647.1	RNSYTAEEQENSISYDKIISPVINDEGEFFDSORNYEGSDKEENMPKADANGAYCIAI
g1 537834683 ref WP_028988726.1	RNSNNGKG-----EEKDFILSPYVDSKGRFNSLE-----ASUDEPKADANGAYHIAL
g1 769142322 ref WP_044919442.1	RNS-----TS-----DSTRDYIISPVVNRKRAYFNDEL-----SDGEVFKADANGAYHIAK
g1 489138581 ref WP_083048289.1	RNS-----RT-----GTEDLYLISPVADVNGVFFDSRD-----APKMPQADANGAYHIGL
g1 739808549 ref WP_036898108.1	RNS-----WK-----EKDLQYLSIPVAGADGRFFDTRE-----GKSLPFADANGAYHIAL
g1 545612232 ref WP_021736722.1	KGGLLNHLKESKD-----LKLNGTISGDMAYISIELNN-----
g1 517171043 ref WP_018359861.1	KGLMVVRKIKAGDH-----EGIHRTISAGMAYVRECEIVE--
g1 502240446 ref WP_012739647.1	KGLYEVKIKKSEHTEDGDFNKLKLPNAEWLDFIQMKRYE--
g1 537834683 ref WP_028988726.1	KGLMHLVLNET--KEENLSRQKIKKIDKDLFFVWERNR----
g1 769142322 ref WP_044919442.1	KGLWVLEQIRK--SEG--EKINLANTAEWLEYAQTML----
g1 489138581 ref WP_083048289.1	KGLMLGRIMN--SEG--KKLNLVINKEEYFEPVQNRNN----
g1 739808549 ref WP_036898108.1	KGLMALRQIRQT--SEG--GKLKLAISNKEMLQFVQERSYEKD

개시된 조성물 및 방법에서 사용될 수 있는 추가 핵산 서열 및 단백질 서열 - Cfp1 인간 분할 단백질 배열. 서열번호 86(제1열) 및 서열번호 90(제2열).

CLUSTAL O(1.2.1) 다수 서열 정렬

```

g1|545612232|ref|WP_021736722.1|      MTQFEGFTNLVQVSKTLRFELIPQCKTLKHIGEGGFIEEDKARNQHYKELKPTIDRIYKT
g1|769142322|ref|WP_044919442.1|      -MYVESLTKQYPPVSKTIKNELIPICKYLDNIIRNNILESDVKRKNQYEHVKGILDEYHKQ
      10.10  *  *****  *****  0.01  1.00  *  11.01.10  0.00  10
      |
g1|545612232|ref|WP_021736722.1|      YADDCLOLVOLDWENLSAAISYRKEKTEET-RNALIEEDATYRNAIHDFYIGRTDNLTD
g1|769142322|ref|WP_044919442.1|      LLINEALDNCYLPDLKI--AAE IYLNQKQVSDREDFNKTOBLRKEVVEKLGK--
      11.01  *  11  *  1  *  0.01  *  1  *  1  *  *  1  1  1
      |
g1|545612232|ref|WP_021736722.1|      AINERMAEIVKGLFKAEELPHQKVLKDLGT-VYITEHAKLIRSFQKFTTYFSQFYENRKN
g1|769142322|ref|WP_044919442.1|      ---AH-ENFTKIGK----KOILDLLEKLPISIEDDYNALESFNFYTYFTSYNKKVREN
      *  *  1  *  *  1.00  *  *  0  10.01  *  *  *  *  *  *  1  1  10
      |
g1|545612232|ref|WP_021736722.1|      VFSAEISATPHRIYGRNPCKFKBNCHIFTHLTAVPSLRNHFENKKAIGIFVSTSE
g1|769142322|ref|WP_044919442.1|      LYDQKCKSTVAYRLINENPKFLDNVKSYPKVTAGT--LADG-----GEEEGD
      12  *  11  *  *  *  10.01  *  *  1  1  00  *  1  1  1  1  1  1
      |
g1|545612232|ref|WP_021736722.1|      EVFSPFPFYNGILLTQTGIDLNGQLGGISKEAGTEKIKGLNEVNLAIQNQDETAMITASE
g1|769142322|ref|WP_044919442.1|      GLFIVETFNKTLTGGIDITNGQVQKINGGIN-----LYNQKQKAKN-GFKKI
      13  *  1  10  *  *  *  *  10  *  *  *  *  *  *  *  *  *  *  1  1
      |
g1|545612232|ref|WP_021736722.1|      PHAFIPILPKOILSDRNLYSPFLEEPKSDEEVISQSPCKYKTLNNENVLETAEALPNEIAD
g1|769142322|ref|WP_044919442.1|      P-KNNKMLYKQILSDREE--SFIDEFQSDGVLDNNEVSGVSLIEGLKSKKVSFAFFDALRE
      *  11  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      IDLTHIFISHKK-LEYISSALCDHWYLRNALYERHIGEL-TGKITSAKEKQVORSKLHE
g1|769142322|ref|WP_044919442.1|      SHQKNVYVNDLAKTAMSNIVFENMRTFDDLNQGYDLANENKQKDDKPYFERQKELKRN
      14  *  11  *  *  *  11  *  1  10  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      -DINLQHEII--SANGKELSEAFKPKYSGE---ILSHAAALDPL----PTTL-KKDEE
g1|769142322|ref|WP_044919442.1|      KSYLEHHCNLSEDSCKLIEYINQISHDENIIZNNKTFILRTVINENHSHKAKNRNA
      15  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      KEILKSDLDLSLILYHLLDWFADVEDGNEVD--PEFSARLTGKILENEPQLSFYNKARNYA
g1|769142322|ref|WP_044919442.1|      VNAIKKSLDSKVLREKLEIN-SGGRIKEDLIVYSANEEELVGLKQVDSLYNMTNYYL
      16  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      TKKPYGVVEKFLNFQMPILASGWDVNEKHNNGAILFYKNGLYYLGIHPKQMGVYKALSFE
g1|769142322|ref|WP_044919442.1|      YKKPPSYEKVKRLNPNRSTLLNGWQKKEIDNLGLVLLKDGKYYLGIPMTYSAN--KAPVNP
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      PTEKTEGEGFKQHYVDFPSAAKNIPKISTQLKANTAMFGYHTTPIILLGWPIEPIETKE
g1|769142322|ref|WP_044919442.1|      PVAKTEKVPKKVDYKLLPVPMNMLPKVPFAKSNID-----FYNP-----SSE
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      IYGLNYPKEKPKKFTQYAKKTGDRGYSALCNWIDFTDHFCSKYTKTISTGLSSCLPS
g1|769142322|ref|WP_044919442.1|      IYUNYKKG-----THKKGNMFS--LEDCHWIDFFRESISKHEDWSKFD--PKFSOT
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      SGVYKLDGEYKALNPLLKHSFGRIAEKIMDAVEIGKLYLQIYVNDPAAKHHSKPKMLH
g1|769142322|ref|WP_044919442.1|      ASYNDISSEFYNEVEKQGYHLYTYDIDETIINDLIEKNELYLQIYVNDPAAKHHSKPKMLH
      17  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      YLYWTGLPQPNLAKTSIKLNGGAEILPYKPKSRMKR--MAHRLGERMLKKKLDQKTPIP
g1|769142322|ref|WP_044919442.1|      TLYFMNLFDQNNIDGVVYKNGEAEVYFYPASISEDELITMKAGEEIKNNKPNR-----
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      DYLQDELVDYNNHKLSHDLSDEARALLPNVITYKEVSHIEIKDRRPTSDKPPFHPVITLAY
g1|769142322|ref|WP_044919442.1|      -----ARTKETSTFSDYIVDKKRYSKDKFTLHPIITMNF
      18  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      QANSPKPFNQRVNAYLKEHPETPIIGIIVGSEANLIYITVIDSTGKILEORDNITTOOF-
g1|769142322|ref|WP_044919442.1|      GVD-EVKRFNDVANSAILRIDNNVNVIGIDGSEANLIYVVIDSGKNILEDISLNSINKK
      19  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      -----DYORKLONREKERVAAAGANSVVGITKDKOGYLSQVIHEIVDLMINHYQAVVLE
g1|769142322|ref|WP_044919442.1|      YDIEYHALLDEREGQADKARDNITVENIRDLKAGYLSQVNVVAKLVLYNNAISCLE
      20  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

```

[0724]

```

g1|545612232|ref|WP_021736722.1|      MNAIFGFKSRTGIAEKAVYQGFENMLIDKLNELVKDY----PAEKVSGVLPVGLTDGF
g1|769142322|ref|WP_044919442.1|      DLNFGFKGRGQK-VEKQVYQKFERMLIDKLNELVICKSRGQTSPEKELGGALNALGLYSKF
      1  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      TSFAKNGTQGGFLFYVPAPYTSKIDPKTGFDVDFVWKTJKNHESAKHFLGFGFLHYDVK
g1|769142322|ref|WP_044919442.1|      KDFKELSNQSGVLYFPAYELTSKIDPITGFANLFYK-CENVEKSKRFPDGFIRFNL
      2  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      TGDFTLHPKNNBLSFGQGLPGFMPWDIVFEKNETDFAKSTPFIAKRIYVPVIE--H
g1|769142322|ref|WP_044919442.1|      ENVFEFGFBN--SFTDRACGINSKWTCTNS-----ENIIZKYNPDQKN
      3  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      HRFITGRYGLYPANELIALLEEKIVFRDGSNILPKLLENDOSHAIOTNVALISVLQNR
g1|769142322|ref|WP_044919442.1|      NMFGE--KVNVVTOENKMLFEQYKIPYEDGNVKNITISNEAFYKRLYRLIDQTLQNR
      4  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      NQMAATGEDYINSFYVDLSGVCFSRFRQPEWPMQDANGAYHIALKGLLNLHLES-H
g1|769142322|ref|WP_044919442.1|      NSTSDGTSDYIISPVNKKREAYFNSLEDGCVPKDADANGAYSTARKGLWLEGIQKKE
      5  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      |
g1|545612232|ref|WP_021736722.1|      DLKLNQISMDQWLAHYDELNN
g1|769142322|ref|WP_044919442.1|      GEKINSAMTNAEWLEYAQTHLL
      6  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

```

[0725]

[0726]

개시된 조성물 및 방법에서 사용될 수 있는 추가 핵산 서열 및 단백질 서열. 문헌 [Haft, D., et al. PLoS Computational Biology, November 2005, Vol. 1, Issue 6, pp. 474-483]으로부터 취한 표. 서열번호 200-253; 차트의 맨 위부터 바닥까지의 순서.

표 1. 새로운 Cas 단백질 패밀리들 기원으로 한, 상이한 카스 코어 유전자, 크리스피/카스 하위유형, 및 램프(RAMP) 모듈의 설명

카테고리	유전자	유전자좌 예	특이적 HMM	COG	추정 기능/패밀리	주
Core proteins	cas1	AF18378	TCAG62287	COG1538	Endonuclease domain ^a	---
	cas2	AF18376	TCAG62573	COG1542, CXX344	---	---
		CF18918	TCAG62873	COG1543	---	Core subtypes specific
	cas3	AF18374	TCAG62123	COG1523	Nucleic acid hydrolase	Core domain
		AF18375	TCAG62386	COG1574	Nucleic acid hydrolase	Not present
		VF17007	TCAG62562	COG1503	Nucleic acid hydrolase	Yeast subtype-specific
Core subtypes	cas4	AF18377	TCAG62372	COG1608	Endonuclease domain ^a	---
	cas5	AF18372	TCAG62593	---	---	Is terminal domain
	cas6	AF18398	TCAG62877	COG1581	Protease domain ^a	Not present, usually lost
	cas7	CF18972	TCAG62543	---	---	---
	cas8	CF18973	TCAG62548	---	---	---
	cas9	CF18974	TCAG62597	---	---	---
Yeast subtype	cas10	CF18975	TCAG62893	---	---	---
	cas11	CF18976	TCAG62898	---	Core N-terminal domain	---
	cas12	VF17008	TCAG62564	---	---	---
	cas13	VF17009	TCAG62565	---	---	---
	cas14	VF17010	TCAG62566	---	---	---
	cas15	VF17011	TCAG62567	---	---	---
Mycobacterium subtype	cas16	VF17012	TCAG62568	COG1571	HNH endonuclease ^a	---
	cas17	VF17013	TCAG62569	---	---	Not always present
Phage subtype	cas18	CF18977	TCAG62599	COG1582	---	---
	cas19	CF18978	TCAG62600	COG1583	---	---
	cas20	CF18979	TCAG62601	COG1584	---	---
	cas21	CF18980	TCAG62602	COG1585	---	---
	cas22	CF18981	TCAG62603	COG1586	---	---
	cas23	CF18982	TCAG62604	COG1587	---	---
Tnase subtype	cas24	CF18983	TCAG62605	COG1588	---	---
	cas25	CF18984	TCAG62606	COG1589	---	---
	cas26	CF18985	TCAG62607	COG1590	---	---
	cas27	CF18986	TCAG62608	COG1591	---	---
	cas28	CF18987	TCAG62609	COG1592	---	---
	cas29	CF18988	TCAG62610	COG1593	---	---
Apore subtype	cas30	AF18379	TCAG62124	COG1504	Core N-terminal domain	Usually present to repeat
	cas31	AF18380	TCAG62125	COG1505	Regulator (TCAG62125)	---
	cas32	AF18381	TCAG62126	COG1506	Regulator (TCAG62126)	Occasionally absent
	cas33	AF18382	TCAG62127	COG1507	Regulator (TCAG62127)	Occasionally absent
	cas34	AF18383	TCAG62128	COG1508	Regulator (TCAG62128)	Occasionally absent
	cas35	AF18384	TCAG62129	COG1509	Regulator (TCAG62129)	Occasionally absent
Mature subtype	cas36	AF18385	TCAG62130	COG1510	Core N-terminal domain	Occasionally absent
	cas37	AF18386	TCAG62131	COG1511	Core N-terminal domain	Occasionally absent
	cas38	AF18387	TCAG62132	COG1512	Core N-terminal domain	Occasionally absent
	cas39	AF18388	TCAG62133	COG1513	Core N-terminal domain	Occasionally absent
	cas40	AF18389	TCAG62134	COG1514	Core N-terminal domain	Occasionally absent
	cas41	AF18390	TCAG62135	COG1515	Core N-terminal domain	Occasionally absent
Ramp subtype	cas42	AF18391	TCAG62136	COG1516	Core N-terminal domain	Occasionally absent
	cas43	AF18392	TCAG62137	COG1517	Core N-terminal domain	Occasionally absent
	cas44	AF18393	TCAG62138	COG1518	Core N-terminal domain	Occasionally absent
	cas45	AF18394	TCAG62139	COG1519	Core N-terminal domain	Occasionally absent
	cas46	AF18395	TCAG62140	COG1520	Core N-terminal domain	Occasionally absent
	cas47	AF18396	TCAG62141	COG1521	Core N-terminal domain	Occasionally absent

표 2. 확인된 문맥상의 패턴이 없는 기타 다른 크리스피/카스 단백질 패밀리

유전자 기호	유전자좌 예	특이적 HMM	COG	추정 기능	발견되는 하위유형 여부				
					Apore	Tnase	Mtase	RAMP	기타
cas1	AF18378	TCAG62287	COG1538	Endonuclease domain ^a	+	+	+	+	+
cas2	AF18376	TCAG62573	COG1542	---	+	+	+	+	+
cas3	AF18374	TCAG62123	COG1523	---	+	+	+	+	+
cas4	AF18377	TCAG62372	COG1608	---	+	+	+	+	+
cas5	AF18372	TCAG62593	---	---	+	+	+	+	+
cas6	AF18398	TCAG62877	COG1581	Protease domain ^a	+	+	+	+	+
cas7	CF18972	TCAG62543	---	---	+	+	+	+	+

^aReference et al. 2006.

편집 표적 서열 및 Nr2(엑손 2)에 대한 PAM: sgRNA 설계 1-3을 위해 이용됨

서열번호 254

GCGACGGAAAGAGTATGAGC TGG

서열번호 255

TATTTGACTTCAGTCAGCGA CGG

서열번호 256

TGGAGGCAAGATATAGATCT TGG

Nr2 타겟에서 통합의 검출을 위한 프라이머 키

프라이머 세트1:

- [0738] **서열번호 257**
- [0739] 프라이머 1: 5'-GTGTTAATTTCAAACATCAGCAGC-3' ,
- [0740] **서열번호 258**
- [0741] 프라이머 2: 5'- GACAAGACATCCTTGATTTG-3'
- [0742] 프라이머 세트 2:
- [0743] **서열번호 259**
- [0744] 프라이머 1: 5'-GAGGTTGACTGTGTAAATG-3' ,
- [0745] **서열번호 260**
- [0746] **프라이머 2: 5'- GATACCAGAGTCACACAACAG-3'**
- [0747] 프라이머 세트 3:
- [0748] **서열번호 261**
- [0749] 프라이머 1: 5'-TCTACATTAATTCTCTTGTGC-3' ,
- [0750] **서열번호 262**
- [0751] 프라이머 2: 5'- GATACCAGAGTCACACAACAG-3'
- [0752] 인간 CXCR4에 대한 등록 번호
- [0753] 유니프로트 P61073
- [0754] 엔SEMBL 유전자 ID: ENSG00000121966
- [0755] CXCR4(엑손 2)에 대한 편집 표적 서열 및 PAM: sgRNA 설계1에 사용됨
- [0756] **서열번호 263**
- [0757] **GGGCAATGGATTGGTCATCC TGG**
- [0758] CXCR4 표적에서 통합의 검출을 위한 프라이머 키
- [0759] 프라이머 세트 1:
- [0760] **서열번호 264**
- [0761] 프라이머 1: 5'- TCTACATTAATTCTCTTGTGC-3' ,
- [0762] **서열번호 265**
- [0763] 프라이머 2: 5'- GACAAGACATCCTTGATTTG-3'
- [0764] 프라이머 세트 2:
- [0765] **서열번호 266**
- [0766] 프라이머 1: 5'- TCTACATTAATTCTCTTGTGC-3' ,
- [0767] **서열번호 267**
- [0768] 프라이머 2: 5'- GATACCAGAGTCACACAACAG -3'
- [0769] 프라이머 세트 3:
- [0770] **서열번호 268**

[0771] 프라이머 1: 5'- GAGGTTGACTGTGTAAATG -3' ,

[0772] 서열번호 269

[0773] 프라이머 2: 5'- GACAAGACATCCTTGATTG-3'

[0774] 프라이머 세트 4:

[0775] 서열번호 270

[0776] 프라이머 1: 5'- GAGGTTGACTGTGTAAATG -3' ,

[0777] 서열번호 271

[0778] 프라이머 2: 5'- GATACCAGAGTCACACAACAG -3'

[0779] 비오틴화를 위해 Avi-태그된 카스9

[0780] 카스9 비오틴화에 사용된 avi-태그의 서열

[0781] 아미노산 서열:

[0782] 서열번호 272

[0783] G G D L E G S G L N D I F E A O K I E W H E *

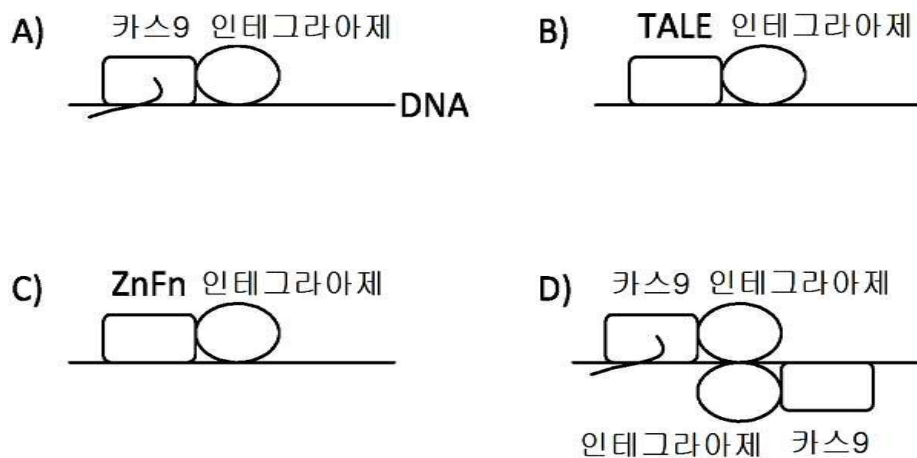
[0784] 핵산 서열:

[0785] 서열번호 273

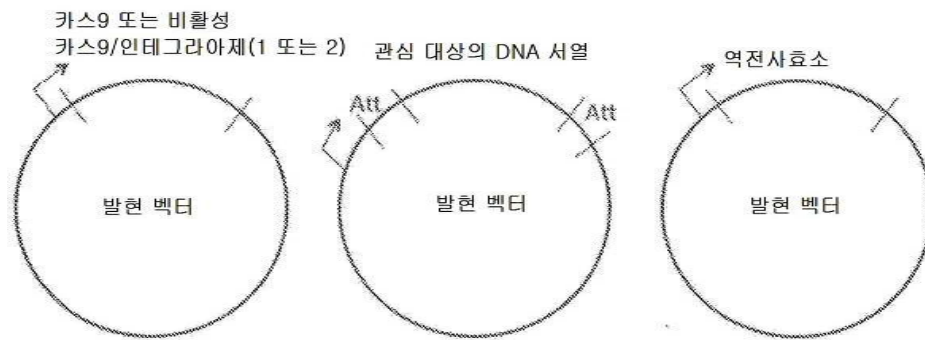
GGCGGCGACCTCGAGGGTAGCGGTCTGAACGATATTTTTGAAGCGCAGAAAA
TTGAATGGCATGAATAA

도면

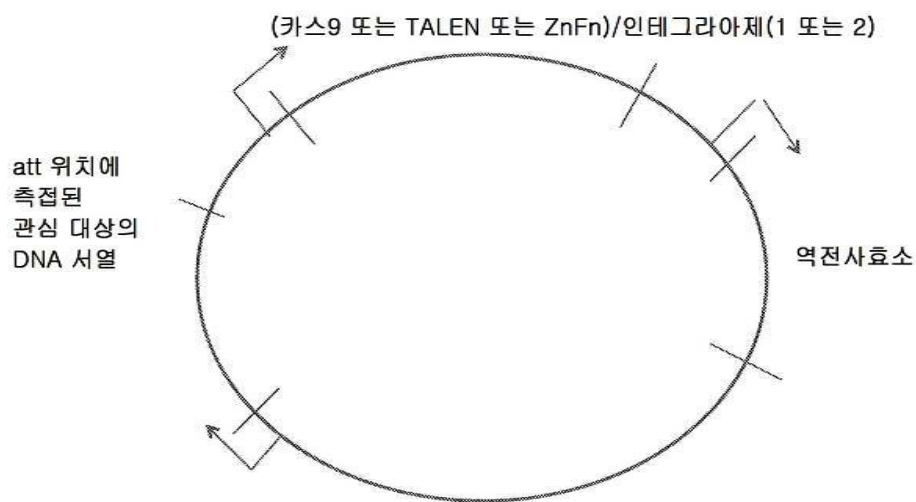
도면1



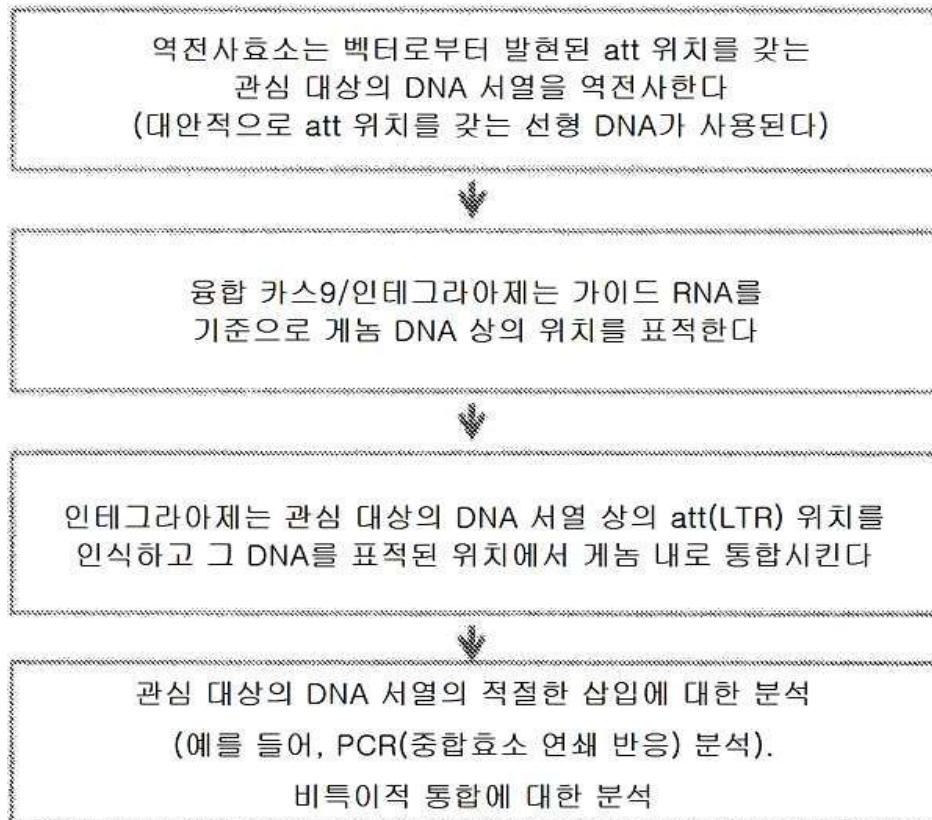
도면2



도면3

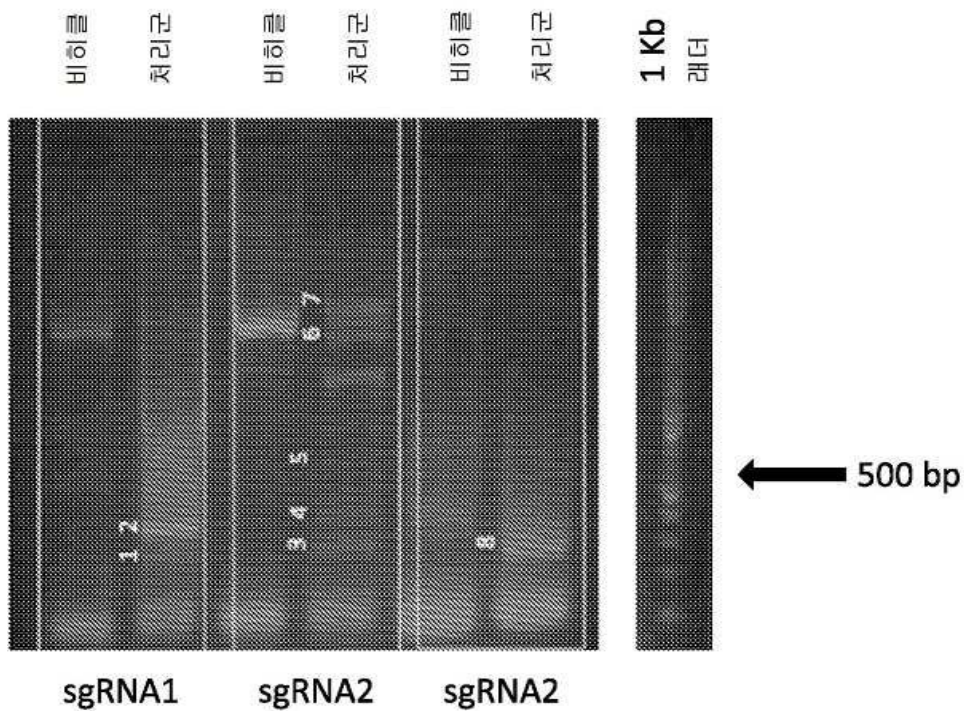


도면4

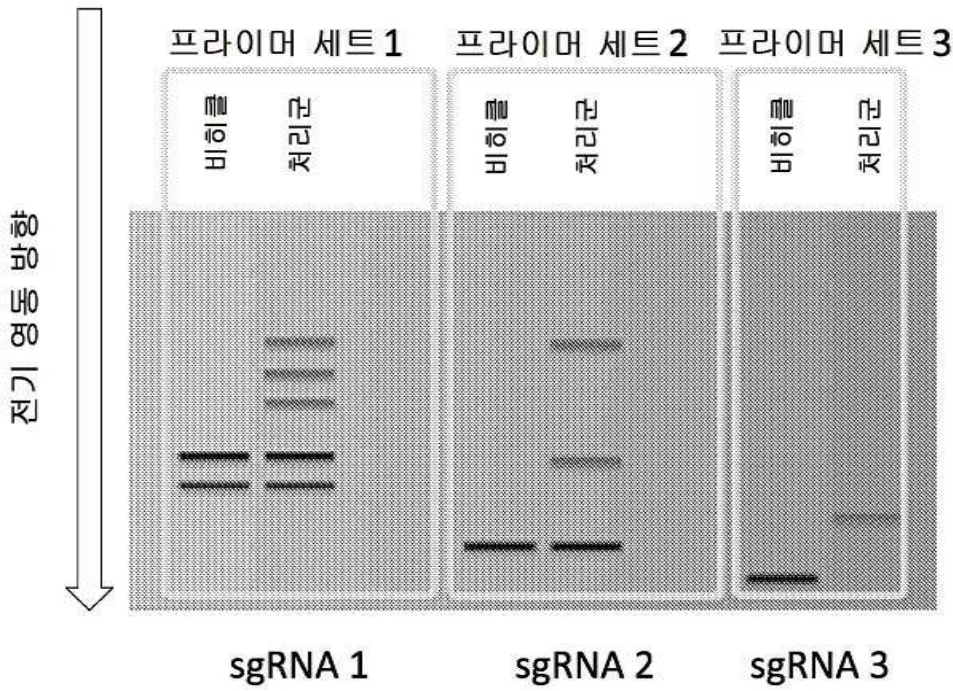


도면5

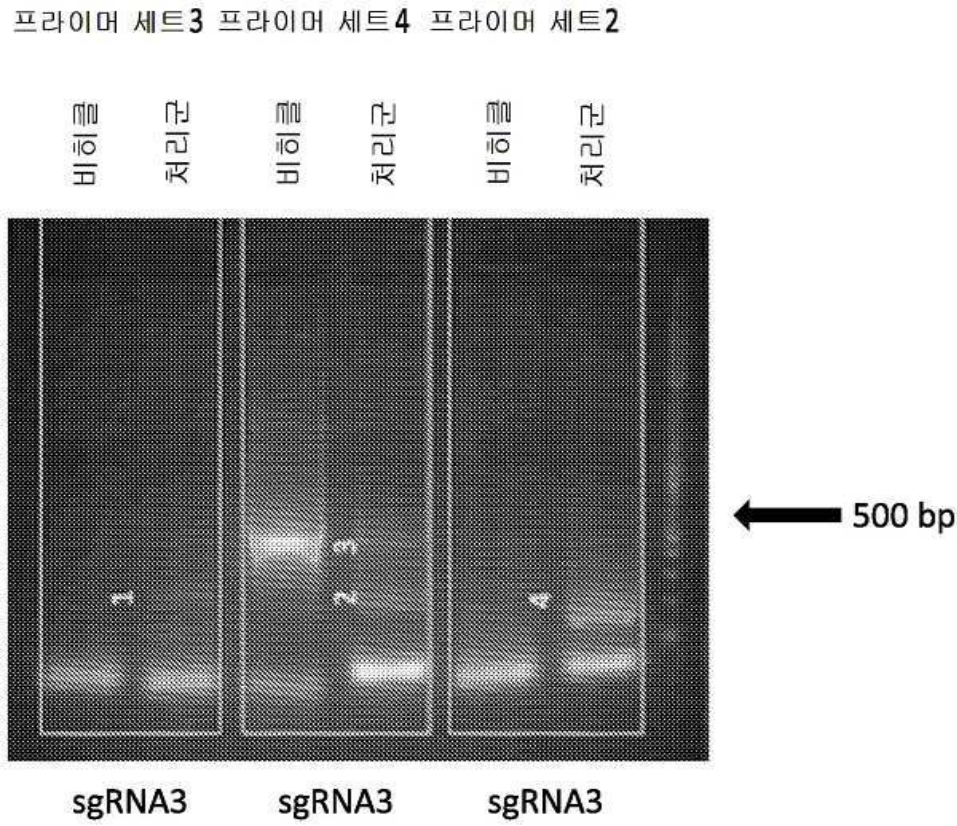
프라이머 세트1 프라이머 세트 1 프라이머 세트2



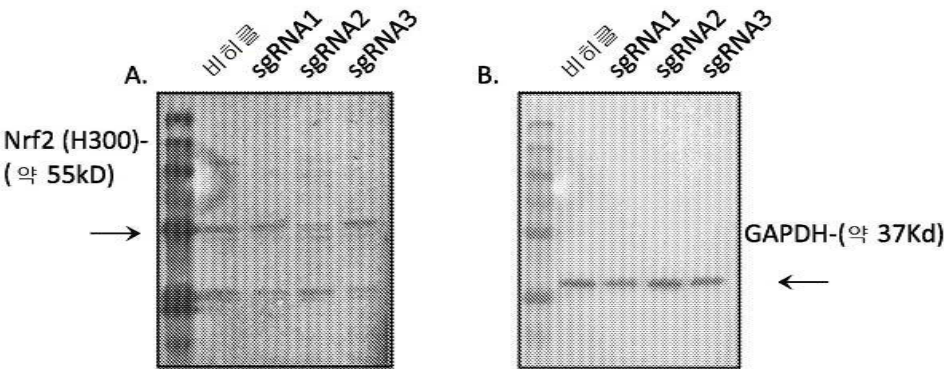
도면6



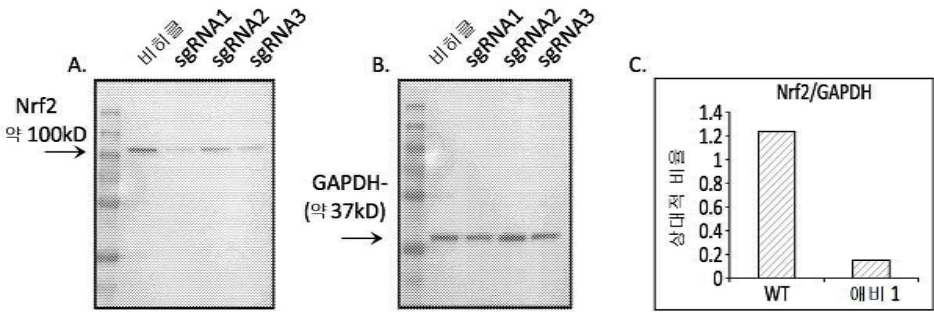
도면7



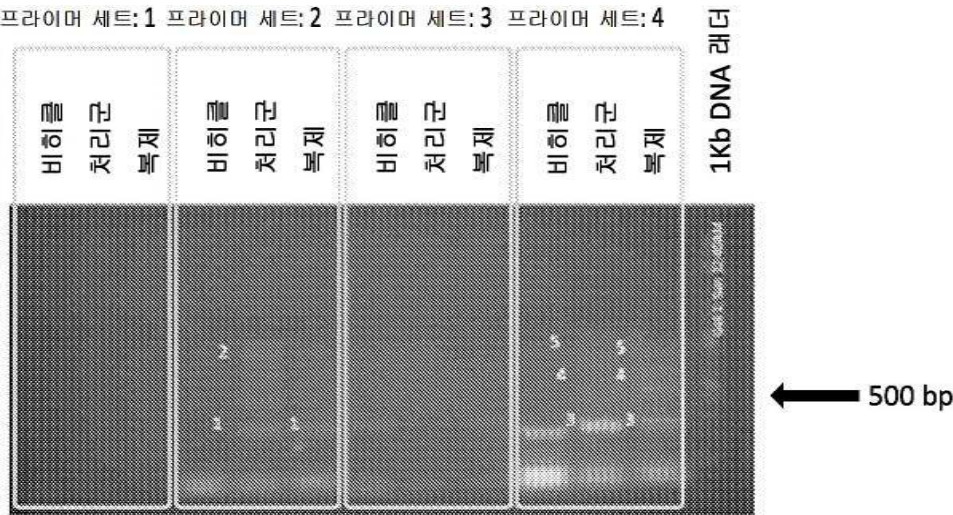
도면8



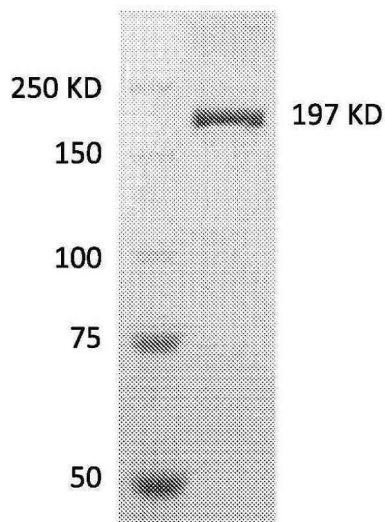
도면9



도면10



도면11



서열 목록

- <110> EXELIGEN SCIENTIFIC, INC.
SHEIKH, Ferrukh
KAWAMURA, Tetsuya
MO, Gloria
- <120> CAS 9 RETROVIRAL INTEGRASE AND CAS 9 RECOMBINASE SYSTEMS FOR
TARGETED INCORPORATION OF A DNA SEQUENCE INTO A GENOME OF A CELL
OR ORGANISM
- <130> 2000PCT
- <150> US 62/140,454
- <151> 2015-03-31
- <150> US 62/210,451
- <151> 2015-08-27
- <150> US 62/240,359
- <151> 2015-10-12
- <160> 274
- <170> PatentIn version 3.5
- <210> 1
- <211> 4167
- <212> DNA
- <213> S.thermophilus
- <400> 1

atgactaagc catactcaat tggacttgat attggaacga atagtgttgg atgggctgta 60

ataactgata attacaaggt tccgtctaaa aaaatgaaag tcttaggaaa tacgagtaaa	120
aagtatatca aaaagaacct gttaggtgta ttactctttg actctggaat cacagcagaa	180
ggaagaagat tgaagcgtac tgcaagaaga cgttatacta gacgccgtaa tcgtatcctt	240
tatttgcagg aaatttttag cacggagatg gctacattag atgatgcttt ctttcaaaga	300
cttgacgatt cgtttttagt tcctgatgat aaacgtgata gtaagtatcc gatatttga	360
aacttagtag aagaaaaagt ctatcatgat gaatttccaa ctatctatca tttaggaaa	420
tatttagcag atagtactaa aaaagcagat ttgcgtctag tttatcttgc attggctcat	480
atgattaaat atagaggta cttcttaatt gaaggagagt ttaattcaaa aaataatgat	540
attcagaaga attttcaaga ctttttgac acttataatg ctatttttga atcggtatta	600
tcacttgaga atagtaaca acttgaggaa attgttaaag ataagattag taaattagaa	660
aagaaagatc gtattttaaa actcttcctt ggggagaaga attcggggat ttttcagag	720
tttctaaagt tgattgtagg aaatcaagct gatttttaga aatgttttaa tttagacgaa	780
aaagcctcct tacatttttc caaagaaagc tatgatgaag atttagagac ttigttaggt	840
tatatggag atgattacag tgatgtcttt ctcaaagcaa agaaacttta tgatgtctatt	900
cttttatcgg gttttctgac tgtaactgat aatgagacag aagcacctct ctcttctgct	960
atgataaagc gatataatga acacaaagaa gatttagcgt tactaaagga atatataaga	1020
aatatttcac taaaacgta taatgaagta tttaaagatg acaccaaaaa tggttatgct	1080
ggttatattg atggaaaaac aaatcaggaa gatttctacg tataatctaa aaacctattg	1140
gctgaatttg aaggtgcgga ttattttctt gaaaaaattg atcgagaaga ttttttgaga	1200
aagcaacgta catttgacaa tggttcgata ccataatcaga ttcatcttca agaaatgaga	1260
gcaattcttg ataagcaagc taaattttat ctttcttgg ctaaaaataa agaaagaatc	1320
gagaagattt taaccttccg aattccttat tatgtaggtc cacttgcgag agggaatagt	1380
gattttgcct ggtcaataag aaaacgaaat gaaaaatta caccttgga ttttgaggac	1440
gttattgaca aagaatcttc ggcagaggct ttcattaatc gaatgactag ttttgatttg	1500
tatttgccag aagagaaggt acttccaaag catagtctct tatacgaaac ttttaatgta	1560
tataatgaat taacaaaagt tagatttatt gccgaaagta tgagagatta tcaattttta	1620
gatagtaagc agaagaaaga tattgttaga ctttatttta agataaaaag gaaagtact	1680
gataaggata ttattgaata ttacatgca atttatgggt atgatggaat tgaattaaaa	1740
ggcatagaga aacagtttaa ttctagtta tctacttate acgatctttt aaatattatt	1800
aatgataaag agtttttga tgatagtcca aatgaagcga ttatcgaaga aattatccat	1860
actttgacaa tttttgaaga tagagagatg ataaaacaac gtccttcaaa atttgagaat	1920

atattcgata aatccgtttt gaaaaagtta tctcgtagac attacactgg ctggggtaag	1980
ttatctgcta agcttattaa tggatttcga gatgaaaaat ctggtaatac tattcttgat	2040
tacttaattg atgatggat ttttaaccgt aatttcacgc aacttattca cgatgatgct	2100
ctttctttta aaaagaagat acagaaagca caaattattg gtgacgaaga taaaggtaat	2160
attaaagagg tcgttaagtc ttgtccaggt agtcctgcga ttaaaaaagg tattttacaa	2220
agcataaaaa ttgtagatga attggtcaaa gtaatgggag gaagaaaacc cgagtcaatt	2280
gttgttgaga tggctcgtga aaatcaatat accaatcaag gtaagtctaa ttccaacaa	2340
cgcttgaaac gtttagaaaa atctctcaaa gagttaggta gtaagatact taaggaaaat	2400
attcctgcaa aactttctaa aatagacaat aacgcacttc aaaatgatcg actttactta	2460
tactatcttc aaaatggaaa agatatgtat accggagatg atttagatat tgatagatta	2520
agtaattatg atattgatca tattattcct caagcttttt tgaaagataa ttctattgac	2580
aataaagtac ttgtttcatc tgctagtaac cgtggtaaat cagatgattt tccaagtta	2640
gaggttgtca aaaaaagaaa gacattttgg tatcaattat tgaaatcaaa attaatttct	2700
caacgaaaat ttgataatct gacaaaagct gaacggggag gattgttacc tgaggacaaa	2760
gctggtttta ttcaacgcca gttggttgaa acacgtcaaa taacaaaaca ttagctcgt	2820
ttacttgatg agaaatttaa taataaaaaa gatgaaaata atagagcggc acgaacagta	2880
aaaattatta ccttgaaatc taccttagtt tctcaatttc gtaaggattt tgaactttat	2940
aaagtctgtg aaatcaatga ttttcatcat gctcatgatg cttacttgaa tgccgttata	3000
gcaagtgcct tacttaagaa ataccctaaa ctagagccag aatttgtgta cggtgattat	3060
ccaaaataca atagttttag agaaagaaag tccgtacag aaaaggataa ttctattca	3120
aatatcatga atatctttta aaaatctatt tctttagctg atggtagagt tattgaaaga	3180
ccacttattg aggtaaatga ggagaccggc gaatccgttt ggaataaaga atctgattta	3240
gcaactgtaa ggagagtact ctcttatccg caagtaaatg ttgtgaaaaa agttgaggaa	3300
cagaatcacg gattggatag aggaaaacca aagggttgt ttaatgcaa tctttcctca	3360
aagccaaaac caaatagtaa tgaaaattta gtaggtgcta aagagtatct tgaccccaaa	3420
aagtatgggg ggtatgctgg aatttcta atcttttctg ttcttgtaa agggacaatt	3480
gaaaaagggt ctaagaaaaa aataacaaat gtactagaat ttcaaggat ttctatttta	3540
gataggatta attatagaaa agataaactt aattttttac ttgaaaaagg ttataaagat	3600
attgagttaa ttattgaact acctaaatat agtttatttg aactttcaga tggttcacgt	3660

cgtatgttgg ctagtatttt gtcaacgaat aataagaggg gagagattca caaaggaaat 3720

cagatttttc ttccacagaa gtttgtgaaa ttactttatc atgctaagag aataagtaac 3780

acaattaatg agaatcatag aaaatatgtt gagaaccata aaaaagagtt tgaagaatta 3840

ttttactaca ttcttgagtt taatgagaat tatgttggag ctaaaaagaa tggtaactt 3900

ttaaactctg cctttcaatc ttggcaaaat catagtatag atgaactctg tagtagtttt 3960

ataggaccta ccggaagtga aagaaagggg ctatttgaat taacctctcg tggaagtgct 4020

gctgattttg aatttttagg tggtaaaatt ccaaggtata gagactatac cccatcatcc 4080

ctattaaaag atgccacact tattcatcaa tctgttacag gcctctatga aacacgaata 4140

gaccttgcca aactaggaga gggtttaa 4167

<210> 2

<211> 1388

<212> PRT

<213> S. Thermophilus

<400> 2

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

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asn Tyr Lys Val Pro Ser Lys Lys Met

20 25 30

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

35 40 45

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

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala Thr Leu Asp Asp Ala

85 90 95

Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val Pro Asp Asp Lys Arg

100 105 110

Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val Glu Glu Lys Val Tyr

115 120 125

His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg Lys Tyr Leu Ala Asp

130	135	140	
Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His			
145	150	155	160
Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Glu Phe Asn Ser			
	165	170	175
Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp Phe Leu Asp Thr Tyr			
	180	185	190
Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu Asn Ser Lys Gln Leu			
	195	200	205
Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu Glu Lys Lys Asp Arg			
	210	215	220
Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser Gly Ile Phe Ser Glu			
225	230	235	240
Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Arg Lys Cys Phe			
	245	250	255
Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser Lys Glu Ser Tyr Asp			
	260	265	270
Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly Asp Asp Tyr Ser Asp			
	275	280	285
Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly			
	290	295	300
Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala Pro Leu Ser Ser Ala			
305	310	315	320
Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp Leu Ala Leu Leu Lys			
	325	330	335
Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr Asn Glu Val Phe Lys			
	340	345	350
Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn			
	355	360	365
Gln Glu Asp Phe Tyr Val Tyr Leu Lys Asn Leu Leu Ala Glu Phe Glu			
	370	375	380
Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg Glu Asp Phe Leu Arg			

385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro Tyr Gln Ile His Leu

 405 410 415
 Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala Lys Phe Tyr Pro Phe
 420 425 430
 Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Asp Phe Ala Trp
 450 455 460
 Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro Trp Asn Phe Glu Asp
 465 470 475 480

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

 545 550 555 560
 Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile Tyr Gly Tyr Asp Gly
 565 570 575
 Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ser Ser Leu Ser Thr
 580 585 590
 Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys Glu Phe Leu Asp Asp
 595 600 605
 Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile His Thr Leu Thr Ile
 610 615 620

 Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ser Lys Phe Glu Asn
 625 630 635 640

Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser Arg Arg His Tyr Thr
 645 650 655
 Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Arg Asp Glu
 660 665 670
 Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Ile Ser
 675 680 685
 Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ala Leu Ser Phe Lys
 690 695 700
 Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp Glu Asp Lys Gly Asn
 705 710 715 720
 Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser Pro Ala Ile Lys Lys
 725 730 735
 Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met
 740 745 750
 Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu Met Ala Arg Glu Asn
 755 760 765
 Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln Gln Arg Leu Lys Arg
 770 775 780
 Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys Ile Leu Lys Glu Asn
 785 790 795 800
 Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn Ala Leu Gln Asn Asp
 805 810 815
 Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly
 820 825 830
 Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr Asp Ile Asp His Ile
 835 840 845
 Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile Asp Asn Lys Val Leu
 850 855 860
 Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp Asp Phe Pro Ser Leu
 865 870 875 880
 Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr Gln Leu Leu Lys Ser
 885 890 895

Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
 900 905 910

 Gly Gly Leu Leu Pro Glu Asp Lys Ala Gly Phe Ile Gln Arg Gln Leu
 915 920 925
 Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Arg Leu Leu Asp Glu
 930 935 940
 Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg Ala Val Arg Thr Val
 945 950 955 960
 Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser Gln Phe Arg Lys Asp
 965 970 975
 Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Phe His His Ala His
 980 985 990
 Asp Ala Tyr Leu Asn Ala Val Ile Ala Ser Ala Leu Leu Lys Lys Tyr
 995 1000 1005
 Pro Lys Leu Glu Pro Glu Phe Val Tyr Gly Asp Tyr Pro Lys Tyr Asn
 1010 1015 1020
 Ser Phe Arg Glu Arg Lys Ser Ala Thr Glu Lys Val Tyr Phe Tyr Ser
 1025 1030 1035 1040
 Asn Ile Met Asn Ile Phe Lys Lys Ser Ile Ser Leu Ala Asp Gly Arg
 1045 1050 1055

 Val Ile Glu Arg Pro Leu Ile Glu Val Asn Glu Glu Thr Gly Glu Ser
 1060 1065 1070
 Val Trp Asn Lys Glu Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser
 1075 1080 1085
 Tyr Pro Gln Val Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly
 1090 1095 1100
 Leu Asp Arg Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser
 1105 1110 1115 1120
 Lys Pro Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr
 1125 1130 1135
 Leu Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe

1140 1145 1150
Ala Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys Ile
1155 1160 1165
Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg Ile Asn
1170 1175 1180
Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly Tyr Lys Asp
1185 1190 1195 1200

Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu Phe Glu Leu Ser
1205 1210 1215
Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu Ser Thr Asn Asn Lys
1220 1225 1230
Arg Gly Glu Ile His Lys Gly Asn Gln Ile Phe Leu Ser Gln Lys Phe
1235 1240 1245
Val Lys Leu Leu Tyr His Ala Lys Arg Ile Ser Asn Thr Ile Asn Glu
1250 1255 1260
Asn His Arg Lys Tyr Val Glu Asn His Lys Lys Glu Phe Glu Glu Leu

1265 1270 1275 1280
Phe Tyr Tyr Ile Leu Glu Phe Asn Glu Asn Tyr Val Gly Ala Lys Lys
1285 1290 1295
Asn Gly Lys Leu Leu Asn Ser Ala Phe Gln Ser Trp Gln Asn His Ser
1300 1305 1310
Ile Asp Glu Leu Cys Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg
1315 1320 1325
Lys Gly Leu Phe Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu
1330 1335 1340

Phe Leu Gly Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser
1345 1350 1355 1360
Leu Leu Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr
1365 1370 1375
Glu Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly
1380 1385

<210> 3

<211> 3171

<212> DNA

<213> P.multocida

<400> 3

atgcaaaca caaatttaag ttatatatta ggttttagatt tggggatcgc ttctgtaggt 60

tgggctgtcg ttgaaatcaa tgaaaatgaa gacctatcgc gcttgattga tgtaggagta 120

aggatatattg agcgtgtcga ggtacccaaa actggagaat ctttagcact ctctcgccgt 180

cttgcaagaa gtactcgccg ttgtatagcgc cgtcgtgcac accgtttact cctcgcaaaa 240

cgcttcttaa aacgtgaagg tatactttcc acaatcgact tagaaaaagg attacccaac 300

caagcttggg aattacgtgt cgccgggtcgt gaacgtcggg tatccgccat agaatggggg 360

gcggttctgc tacatttaac caagcatcga ggttatcttt ctaaacgtaa aatgaatcc 420

caaacaaca acaaagaatt aggagcctta ctctctggag tggcacaaaa ccatcaatta 480

ttacaatcag atgactaccg aacaccagca gagctcgcac tgaaaaaatt tgctaaagaa 540

gaagggcata tccgtaatca acgaggtgcc tatacacata catttaatcg attagactta 600

ttagctgaac ttaacttgct ttttctcaa caacatcagt ttggtaaccc tcaactgaaa 660

gagcatattc aacaatatat gacagaattg cttatgtggc aaaagccagc cttatctggg 720

gaggcaattt taaaatgtt gggtaaattg acgcatgaaa aaaatgagtt taaagcagca 780

aaacatacct acagtgcgga gcgctttgtt tggctaacca aactcaataa ctgctgcatt 840

ttagaagatg gggcagaacg agctcttaac gaagaagaac gtcaactatt gataaatcat 900

ccgtatgaga aatcaaaatt aacctatgcc caagtcagaa aattgttagg gctttccgaa 960

caagcgattt ttaagcatct acgttatagt aaagaaaacg cagaatcagc tacttttatg 1020

gagcttaaag cttggcatgc aattcgtaaa gcgttagaaa atcaaggatt gaaggatact 1080

tggcaagatc tcgctaagaa acctgactta ctagatgaaa ttggtaccgc attttctctt 1140

tataaaactg atgaagatat tcagcaatat ttgacaaata aggtaccgaa ctcaatcatt 1200

aatgcattat tagtttctct gaatttcgat aaattcattg agttatcttt gaaaagttaa 1260

cgtaaaatct tgcccctaac ggagcaaggc aagcgttatg atcaagcttg tcgtgaaatt 1320

tatgggcatac attatgggta ggcaaatcaa aaaacttctc agctactacc agctattcca 1380

gcccagaaa ttcgtaatcc tgttgtttta cgtacacttt cacaagcacg taaagtatc 1440

aatgccatta ttcgtcaata tggttccctt gctcgagtcc atattgaaac aggaagagaa 1500

cttgggaaat cttttaaaga acgtcgtgaa attcaaaaac aacaggaaga taatcgaact 1560

aagcgagaaa gtgcggtaca aaaattcaaa gaattatctt ctgacttttc aagtgaaccc 1620

aaaagtaaag atattttaaa attccgctta tacgaacaac agcatggtaa atgcttatac 1680
 tctggaaaag agatcaatat tcatcgctta aatgaaaagg gttatgtgga aattgatcat 1740

 gctttacctt tctcacggac ttgggatgat agttttaata ataaagtatt agttcttgcc 1800
 agcgaacc aaacaaagg gaatcaaaca ccgtatgaat ggctacaagg taaaataaat 1860
 tcggaacgtt ggaaaaactt tgttgcttta gtactgggta gccagtgcag tgcagccaag 1920
 aaacaacgat tactcactca agttattgat gataataaat ttattgatag aaacttaaat 1980
 gatactcgct atattgcccg attcctatcc aactatattc aagaaaattt gcttttggtg 2040
 ggtaaaaata agaaaaatgt ctttacacca aacggtcaaa ttactgcatt attaagaagt 2100
 cgctggggat taattaaggc tcgtgagaat aataaccgtc atcatgcttt agatgcgata 2160

 gttgtggctt gtgcaacacc ttctatgcaa caaaaaatta cccgatttat tcgatttaaa 2220
 gaagtgcac catacaaaat agaaaaagg tatgaaatgg tggatcaaga aagcggagaa 2280
 attatttcac ctcatcttc tgaaccttgg gcttatttta gacaagaggt taatattcgt 2340
 gtttttgata atcatccaga tactgtctta aaagagatgc tacctgatcg cccacaagca 2400
 aatcaccagt ttgtacagcc ctttttgggt tctcgtgccc caactcgtaa aatgagtgg 2460
 caagggcata tggaacaat taaatcagct aaacgcttag cagaaggcat tagcgtttta 2520
 agaattctc tcacgaatt aaaacctaatt ttattggaaa atatggtgaa taaagaacgt 2580

 gagccagcac ttatgcagg actaaaagca cgcttggtg aatttaatca agatccagca 2640
 aaagcgtttg ctacgccttt ttataaaca ggagggcagc aggtcaaagc tattcgtgtt 2700
 gaacaggtag aaaaatcagg ggtattagtc agagaaaaca atggggtagc agataatgcc 2760
 tctatcgttc gaacagacgt atttatcaaa aataataaat ttttccttgt tcctatctat 2820
 acttgccaag ttgcgaaagg catcttgcca aataaagcta ttgttgctca taaaaatgaa 2880
 gatgaatggg aagaatgga tgaagggtct aagtttaaat tcagcctttt cccgaatgat 2940
 cttgtcgagc taaaaccaa aaaagaatac ttttcggct attacatcgg actagatcgt 3000

 gcaactggaa acattagcct aaaagaacat gatggtgaga tatcaaaagg taaagacggt 3060
 gtttaccgtg ttggtgtcaa gttagctctt tcttttgaaa aatatcaagt tgatgagctc 3120
 ggtaaaaata gacaaatttg ccgacctcag caaagacaac ctgtgcgtta a 3171

 <210> 4
 <211> 1056
 <212> PRT
 <213> P.multocida
 <400> 4

Met Gln Thr Thr Asn Leu Ser Tyr Ile Leu Gly Leu Asp Leu Gly Ile
1 5 10 15
Ala Ser Val Gly Trp Ala Val Val Glu Ile Asn Glu Asn Glu Asp Pro
20 25 30
Ile Gly Leu Ile Asp Val Gly Val Arg Ile Phe Glu Arg Ala Glu Val
35 40 45
Pro Lys Thr Gly Glu Ser Leu Ala Leu Ser Arg Arg Leu Ala Arg Ser
50 55 60
Thr Arg Arg Leu Ile Arg Arg Arg Ala His Arg Leu Leu Leu Ala Lys
65 70 75 80
Arg Phe Leu Lys Arg Glu Gly Ile Leu Ser Thr Ile Asp Leu Glu Lys
85 90 95
Gly Leu Pro Asn Gln Ala Trp Glu Leu Arg Val Ala Gly Leu Glu Arg
100 105 110
Arg Leu Ser Ala Ile Glu Trp Gly Ala Val Leu Leu His Leu Ile Lys
115 120 125
His Arg Gly Tyr Leu Ser Lys Arg Lys Asn Glu Ser Gln Thr Asn Asn
130 135 140
Lys Glu Leu Gly Ala Leu Leu Ser Gly Val Ala Gln Asn His Gln Leu
145 150 155 160
Leu Gln Ser Asp Asp Tyr Arg Thr Pro Ala Glu Leu Ala Leu Lys Lys
165 170 175
Phe Ala Lys Glu Glu Gly His Ile Arg Asn Gln Arg Gly Ala Tyr Thr
180 185 190
His Thr Phe Asn Arg Leu Asp Leu Leu Ala Glu Leu Asn Leu Leu Phe
195 200 205
Ala Gln Gln His Gln Phe Gly Asn Pro His Cys Lys Glu His Ile Gln
210 215 220
Gln Tyr Met Thr Glu Leu Leu Met Trp Gln Lys Pro Ala Leu Ser Gly
225 230 235 240
Glu Ala Ile Leu Lys Met Leu Gly Lys Cys Thr His Glu Lys Asn Glu

245 250 255
 Phe Lys Ala Ala Lys His Thr Tyr Ser Ala Glu Arg Phe Val Trp Leu
 260 265 270
 Thr Lys Leu Asn Asn Leu Arg Ile Leu Glu Asp Gly Ala Glu Arg Ala
 275 280 285
 Leu Asn Glu Glu Glu Arg Gln Leu Leu Ile Asn His Pro Tyr Glu Lys
 290 295 300
 Ser Lys Leu Thr Tyr Ala Gln Val Arg Lys Leu Leu Gly Leu Ser Glu
 305 310 315 320

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

 385 390 395 400
 Asn Ala Leu Leu Val Ser Leu Asn Phe Asp Lys Phe Ile Glu Leu Ser
 405 410 415
 Leu Lys Ser Leu Arg Lys Ile Leu Pro Leu Met Glu Gln Gly Lys Arg
 420 425 430
 Tyr Asp Gln Ala Cys Arg Glu Ile Tyr Gly His His Tyr Gly Glu Ala
 435 440 445
 Asn Gln Lys Thr Ser Gln Leu Leu Pro Ala Ile Pro Ala Gln Glu Ile
 450 455 460

 Arg Asn Pro Val Val Leu Arg Thr Leu Ser Gln Ala Arg Lys Val Ile
 465 470 475 480
 Asn Ala Ile Ile Arg Gln Tyr Gly Ser Pro Ala Arg Val His Ile Glu
 485 490 495
 Thr Gly Arg Glu Leu Gly Lys Ser Phe Lys Glu Arg Arg Glu Ile Gln

500 505 510
 Lys Gln Gln Glu Asp Asn Arg Thr Lys Arg Glu Ser Ala Val Gln Lys
 515 520 525
 Phe Lys Glu Leu Phe Ser Asp Phe Ser Ser Glu Pro Lys Ser Lys Asp

 530 535 540
 Ile Leu Lys Phe Arg Leu Tyr Glu Gln Gln His Gly Lys Cys Leu Tyr
 545 550 555 560
 Ser Gly Lys Glu Ile Asn Ile His Arg Leu Asn Glu Lys Gly Tyr Val
 565 570 575
 Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser Phe
 580 585 590
 Asn Asn Lys Val Leu Val Leu Ala Ser Glu Asn Gln Asn Lys Gly Asn
 595 600 605

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

 675 680 685
 Thr Pro Asn Gly Gln Ile Thr Ala Leu Leu Arg Ser Arg Trp Gly Leu
 690 695 700
 Ile Lys Ala Arg Glu Asn Asn Asn Arg His His Ala Leu Asp Ala Ile
 705 710 715 720
 Val Val Ala Cys Ala Thr Pro Ser Met Gln Gln Lys Ile Thr Arg Phe
 725 730 735
 Ile Arg Phe Lys Glu Val His Pro Tyr Lys Ile Glu Asn Arg Tyr Glu
 740 745 750

Met Val Asp Gln Glu Ser Gly Glu Ile Ile Ser Pro His Phe Pro Glu
755 760 765

Pro Trp Ala Tyr Phe Arg Gln Glu Val Asn Ile Arg Val Phe Asp Asn
770 775 780

His Pro Asp Thr Val Leu Lys Glu Met Leu Pro Asp Arg Pro Gln Ala
785 790 795 800

Asn His Gln Phe Val Gln Pro Leu Phe Val Ser Arg Ala Pro Thr Arg
805 810 815

Lys Met Ser Gly Gln Gly His Met Glu Thr Ile Lys Ser Ala Lys Arg
820 825 830

Leu Ala Glu Gly Ile Ser Val Leu Arg Ile Pro Leu Thr Gln Leu Lys
835 840 845

Pro Asn Leu Leu Glu Asn Met Val Asn Lys Glu Arg Glu Pro Ala Leu
850 855 860

Tyr Ala Gly Leu Lys Ala Arg Leu Ala Glu Phe Asn Gln Asp Pro Ala
865 870 875 880

Lys Ala Phe Ala Thr Pro Phe Tyr Lys Gln Gly Gly Gln Gln Val Lys
885 890 895

Ala Ile Arg Val Glu Gln Val Gln Lys Ser Gly Val Leu Val Arg Glu
900 905 910

Asn Asn Gly Val Ala Asp Asn Ala Ser Ile Val Arg Thr Asp Val Phe
915 920 925

Ile Lys Asn Asn Lys Phe Phe Leu Val Pro Ile Tyr Thr Trp Gln Val
930 935 940

Ala Lys Gly Ile Leu Pro Asn Lys Ala Ile Val Ala His Lys Asn Glu
945 950 955 960

Asp Glu Trp Glu Glu Met Asp Glu Gly Ala Lys Phe Lys Phe Ser Leu
965 970 975

Phe Pro Asn Asp Leu Val Glu Leu Lys Thr Lys Lys Glu Tyr Phe Phe
980 985 990

Gly Tyr Tyr Ile Gly Leu Asp Arg Ala Thr Gly Asn Ile Ser Leu Lys
995 1000 1005

Glu His Asp Gly Glu Ile Ser Lys Gly Lys Asp Gly Val Tyr Arg Val

1010

1015

1020

Gly Val Lys Leu Ala Leu Ser Phe Glu Lys Tyr Gln Val Asp Glu Leu

1025

1030

1035

1040

Gly Lys Asn Arg Gln Ile Cys Arg Pro Gln Gln Arg Gln Pro Val Arg

1045

1050

1055

<210> 5

<211> 4038

<212> DNA

<213> S.mutans

<400> 5

atgaaaaaac cttactctat tggacttgat attggaacca attctgttgg ttgggctgtt 60

gtgacagatg actacaaagt tcctgctaag aagatgaagg ttctgggaaa tacagataaa 120

agtcatatcg agaaaaatit gcttggcgct ttattatttg atagcgggaa tactgcagaa 180

gacagacggt taaagagaac tgctcgccgt cgttacacac gtcgcagaaa tcgtatttta 240

tatttgcaag agattttttc agaagaaatg ggcaaggtag atgatatgtt ctttcatcgt 300

ttagaggatt cttttcttgt tactgaggat aaacgaggag agcgccatcc catttttggg 360

aatcttgaag aagaagttaa gtatcatgaa aattttccaa ccatttatca ttgcgggcaa 420

tatcttgcgg ataattcaga aaaagttgat ttgcgtttag tttatttggc ttggcacat 480

ataattaagt ttagaggica ttttttaatt gaaggaaagt ttgatacacg caataatgat 540

glacaaagac tgittcaaga attttttagca gtctatgata atacttttga gaatagttag 600

cttcaggagc aaaaatttca agttgaagaa attctgactg ataaaatcag taaatctgct 660

aagaaagata gagttttgaa actttttcct aatgaaaagt ctaatggccg ctttgcagaa 720

tttctaaaac taattgttgg taatcaagct gatttttaaaa agcattttga attagaagag 780

aaagcaccat tgcaattttc taaagatact tatgaagaag agttagaagt actattagct 840

caaattggag ataattacgc agagctcttt ttatcagcaa agaaactgta tgatagtatc 900

cttttatcag ggattttaac agttactgat gttggtacca aagcgctttt atctgcttcg 960

atgattcagc gatataatga acatcagatg gatttagctc agcttaaaca attcattcgt 1020

cagaaattat cagataaata taacgaagtt ttttctgatg tttcaaaaga cggctatgcg 1080

ggtttatattg atgggaaaac aaatcaagaa gctttttata aataccttaa aggtctatta 1140

aataagattg agggaagtgg ctatttcctt gataaaattg agcgtgaaga ttttctaaga 1200

aagcaacgta cctttgacaa tggctctatt ccacatcaga ttcaccttca agaaatgcgt	1260
gctatcattc gtagacagcg tgaattttat cggtttttag cagacaatca agataggatt	1320
gagaaattat tgactttccg tattccctac tatgttggtc cattagcgcg cggaaaaagt	1380
gattttgctt ggtaagtcg gaaatcggct gataaaatta caccatggaa ttttgatgaa	1440
atcgttgata aagaatcctc tgcagaagct tttatcaatc gtatgacaaa ttatgatttg	1500
tacttgccaa atcaaaaagt tcttcctaaa catagtttat tatacgaaaa atttactgtt	1560
tacaatgaat taacaaaggt taaatataaa acagagcaag gaaaaacagc attttttgat	1620
gccaatatga agcaagaaat ctttgatggc gtatttaagg tttatcgaaa agtaactaaa	1680
gataaattaa tggatttcct tgaaaaagaa tttgatgaat ttcgtattgt tgatttaaca	1740
ggtctggata aagaaaaata agtatattaac gcttcttatg gaacttatca tgatttgtgt	1800
aaaattttag ataaagattt tctcgataat tcaaagaatg aaaagatttt agaagatatt	1860
gtgttgacct taacgttatt tgaagataga gaaatgatta gaaaacgtct agaaaattac	1920
agtgatttat tgaccaaga acaagtgaag aagctggaaa gacgtcatta tactggttgg	1980
ggaagattat cagctgagtt aattcatggt attcgcaata aagaaagcag aaaaacaatt	2040
cttgattatc tcattgatga tggcaatagc aatcggaact ttatgcaact gattaacgat	2100
gatgctcttt ctttcaaaga agagattgct aaggcacaag ttattggaga aacagacaat	2160
ctaaatcaag ttgttagtga tattgctggc agccctgcta ttaaaaaagg aattttaca	2220
agcttgaaga ttgttgatga gcttgtcaaa attatgggac atcaacctga aaatatcgtc	2280
gtggagatgg cgcgtgaaaa ccagtttacc aatcagggac gacgaaattc acagcaacgt	2340
ttgaaaggtt tgacagattc tattaagaa tttggaagtc aaattcttaa agaacatccg	2400
gttgagaatt cacagttaca aaatgataga ttgtttctat attatttaca aaacggcaga	2460
gatatgtata ctggagaaga attggatatt gattatctaa gccagtatga tatagaccat	2520
attatcccg aagcttttat aaaggataat tctattgata atagagtatt gactagctca	2580
aaggaaaatc gtggaaaatc ggatgatgta ccaagtaaag atgttggtcg taaaatgaaa	2640
tcctattgga gtaagctact ttcggcaaag ctattacac aacgtaaatt tgataatttg	2700
acaaaagctg aacgaggtgg attgaccgac gatgataaag ctggattcat caagcgtcaa	2760
ttagtagaaa cagacaaat taccaacat gtagcacgta ttctggacga acgatttaat	2820
acagaaacag atgaaaacaa caagaaaatt cgtcaagtaa aaattgtgac cttgaaatca	2880
aatcttggtt ccaatttccg taaagagttt gaactctaca aagtgcgtga aattaatgac	2940
tatcatcatg cacatgatgc ctatctcaat gctgtaattg gaaaggcttt actaggtgtt	3000
taccacaat tggaacctga atttgtttat ggtgattatc ctcatcttca tggacataaa	3060

gaaaataaag caactgctaa gaaatttttc tattcaaata ttatgaactt ctttaaaaaa	3120
gatgatgtcc gtactgataa aaatggtgaa attatctgga aaaaagatga gcatatttct	3180
aatattaaaa aagtgttttc ttatccacaa gttaatatig ttaagaaagt agaggagcaa	3240
acgggaggat ttctaaaga atctatcttg ccgaaaggta attctgacaa gcttattcct	3300
cgaataacga agaaatttta ttgggatacc aagaaatag gaggatttga tagcccgatt	3360
gttgcttatt ctattttagt tattgctgat attgaaaag gtaaatctaa aaaattgaaa	3420
acagtcaaag ccttagttgg tgtcactatt atggaaaaga tgacttttga aagggatcca	3480
gttgcttttc ttgagcgaaa aggctatcga aatgttcaag aagaaatat tataaagtta	3540
ccaaaatata gtttatttaa actagaaaac ggacgaaaaa ggctattggc aagtgtagg	3600
gaacttcaaa agggaaatga aatcgttttg ccaaatcatt taggaacctt gctttatcac	3660
gctaaaaata ttcataaagt tgatgaacca aagcatttgg actatgttga taaacataaa	3720
gatgaattta aggagtgtct agatgttttg tcaaaccttt ctaaaaata tactttagca	3780
gaaggaaatt tagaaaaat caaagaatta tatgcacaaa ataatggtga agatcttaaa	3840
gaattagcaa gttcatttat caacttatta acatttactg ctataggagc accggtact	3900
tttaatttct ttgataaaaa tattgatcga aaacgatata cttcaactac tgaaattctc	3960
aacgtaccc tcatcacca atccatcacc ggtctttatg aaacgcggat tgatctcaat	4020
aagttaggag gagactaa	4038
<210> 6	
<211> 1345	
<212> PRT	
<213> S. mutans	
<400> 6	
Met Lys Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val	
1 5 10 15	
Gly Trp Ala Val Val Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys Met	
20 25 30	
Lys Val Leu Gly Asn Thr Asp Lys Ser His Ile Glu Lys Asn Leu Leu	
35 40 45	
Gly Ala Leu Leu Phe Asp Ser Gly Asn Thr Ala Glu Asp Arg Arg Leu	
50 55 60	
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu	

65						70						75						80	
Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Glu	Glu	Met	Gly	Lys	Val	Asp	Asp	Ser				
					85					90					95				
Phe	Phe	His	Arg	Leu	Glu	Asp	Ser	Phe	Leu	Val	Thr	Glu	Asp	Lys	Arg				
					100					105					110				
Gly	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Leu	Glu	Glu	Glu	Val	Lys	Tyr				
					115					120					125				
His	Glu	Asn	Phe	Pro	Thr	Ile	Tyr	His	Leu	Arg	Gln	Tyr	Leu	Ala	Asp				
					130					135					140				
Asn	Pro	Glu	Lys	Val	Asp	Leu	Arg	Leu	Val	Tyr	Leu	Ala	Leu	Ala	His				
145					150					155					160				
Ile	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Lys	Phe	Asp	Thr				
					165					170					175				
Arg	Asn	Asn	Asp	Val	Gln	Arg	Leu	Phe	Gln	Glu	Phe	Leu	Ala	Val	Tyr				
					180					185					190				
Asp	Asn	Thr	Phe	Glu	Asn	Ser	Ser	Leu	Gln	Glu	Gln	Asn	Val	Gln	Val				
					195					200					205				
Glu	Glu	Ile	Leu	Thr	Asp	Lys	Ile	Ser	Lys	Ser	Ala	Lys	Lys	Asp	Arg				
					210					215					220				
Val	Leu	Lys	Leu	Phe	Pro	Asn	Glu	Lys	Ser	Asn	Gly	Arg	Phe	Ala	Glu				
225					230					235					240				
Phe	Leu	Lys	Leu	Ile	Val	Gly	Asn	Gln	Ala	Asp	Phe	Lys	Lys	His	Phe				
					245					250					255				
Glu	Leu	Glu	Glu	Lys	Ala	Pro	Leu	Gln	Phe	Ser	Lys	Asp	Thr	Tyr	Glu				
					260					265					270				
Glu	Glu	Leu	Glu	Val	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Asn	Tyr	Ala	Glu				
					275					280					285				
Leu	Phe	Leu	Ser	Ala	Lys	Lys	Leu	Tyr	Asp	Ser	Ile	Leu	Leu	Ser	Gly				
					290					295					300				
Ile	Leu	Thr	Val	Thr	Asp	Val	Gly	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser				
305					310					315					320				
Met	Ile	Gln	Arg	Tyr	Asn	Glu	His	Gln	Met	Asp	Leu	Ala	Gln	Leu	Lys				

325	330	335	
Gln Phe Ile Arg Gln Lys Leu Ser Asp Lys Tyr Asn Glu Val Phe Ser			
340	345	350	
Asp Val Ser Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn			
355	360	365	
Gln Glu Ala Phe Tyr Lys Tyr Leu Lys Gly Leu Leu Asn Lys Ile Glu			
370	375	380	
Gly Ser Gly Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp Phe Leu Arg			
385	390	395	400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu			
405	410	415	
Gln Glu Met Arg Ala Ile Ile Arg Arg Gln Ala Glu Phe Tyr Pro Phe			
420	425	430	
Leu Ala Asp Asn Gln Asp Arg Ile Glu Lys Leu Leu Thr Phe Arg Ile			
435	440	445	
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Lys Ser Asp Phe Ala Trp			
450	455	460	
Leu Ser Arg Lys Ser Ala Asp Lys Ile Thr Pro Trp Asn Phe Asp Glu			
465	470	475	480
Ile Val Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr			
485	490	495	
Asn Tyr Asp Leu Tyr Leu Pro Asn Gln Lys Val Leu Pro Lys His Ser			
500	505	510	
Leu Leu Tyr Glu Lys Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys			
515	520	525	
Tyr Lys Thr Glu Gln Gly Lys Thr Ala Phe Phe Asp Ala Asn Met Lys			
530	535	540	
Gln Glu Ile Phe Asp Gly Val Phe Lys Val Tyr Arg Lys Val Thr Lys			
545	550	555	560
Asp Lys Leu Met Asp Phe Leu Glu Lys Glu Phe Asp Glu Phe Arg Ile			
565	570	575	

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

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

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

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

Ser Asp Leu Leu Thr Lys Glu Gln Val Lys Lys Leu Glu Arg Arg His
645 650 655

Tyr Thr Gly Trp Gly Arg Leu Ser Ala Glu Leu Ile His Gly Ile Arg
660 665 670

Asn Lys Glu Ser Arg Lys Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly
675 680 685

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

Phe Lys Glu Glu Ile Ala Lys Ala Gln Val Ile Gly Glu Thr Asp Asn
705 710 715 720

Leu Asn Gln Val Val Ser Asp Ile Ala Gly Ser Pro Ala Ile Lys Lys
725 730 735

Gly Ile Leu Gln Ser Leu Lys Ile Val Asp Glu Leu Val Lys Ile Met
740 745 750

Gly His Gln Pro Glu Asn Ile Val Val Glu Met Ala Arg Glu Asn Gln
755 760 765

Phe Thr Asn Gln Gly Arg Arg Asn Ser Gln Gln Arg Leu Lys Gly Leu
770 775 780

Thr Asp Ser Ile Lys Glu Phe Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Ser Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Thr Gly Glu Glu Leu Asp Ile Asp Tyr
820 825 830

Leu Ser Gln Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe Ile Lys

835 840 845

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

850 855 860

Gly Lys Ser Asp Asp Val Pro Ser Lys Asp Val Val Arg Lys Met Lys

865 870 875 880

Ser Tyr Trp Ser Lys Leu Leu Ser Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Val Tyr Gly Asp Tyr Pro His Phe His Gly His Lys Glu Asn Lys Ala

1010 1015 1020

Thr Ala Lys Lys Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Lys

1025 1030 1035 1040

Asp Asp Val Arg Thr Asp Lys Asn Gly Glu Ile Ile Trp Lys Lys Asp

1045 1050 1055

Glu His Ile Ser Asn Ile Lys Lys Val Leu Ser Tyr Pro Gln Val Asn

1060 1065 1070

Ile Val Lys Lys Val Glu Glu Gln Thr Gly Gly Phe Ser Lys Glu Ser

1075 1080 1085
 Ile Leu Pro Lys Gly Asn Ser Asp Lys Leu Ile Pro Arg Lys Thr Lys
 1090 1095 1100
 Lys Phe Tyr Trp Asp Thr Lys Lys Tyr Gly Gly Phe Asp Ser Pro Ile
 1105 1110 1115 1120
 Val Ala Tyr Ser Ile Leu Val Ile Ala Asp Ile Glu Lys Gly Lys Ser

 1125 1130 1135
 Lys Lys Leu Lys Thr Val Lys Ala Leu Val Gly Val Thr Ile Met Glu
 1140 1145 1150
 Lys Met Thr Phe Glu Arg Asp Pro Val Ala Phe Leu Glu Arg Lys Gly
 1155 1160 1165
 Tyr Arg Asn Val Gln Glu Glu Asn Ile Ile Lys Leu Pro Lys Tyr Ser
 1170 1175 1180
 Leu Phe Lys Leu Glu Asn Gly Arg Lys Arg Leu Leu Ala Ser Ala Arg
 1185 1190 1195 1200

 Glu Leu Gln Lys Gly Asn Glu Ile Val Leu Pro Asn His Leu Gly Thr
 1205 1210 1215
 Leu Leu Tyr His Ala Lys Asn Ile His Lys Val Asp Glu Pro Lys His
 1220 1225 1230
 Leu Asp Tyr Val Asp Lys His Lys Asp Glu Phe Lys Glu Leu Leu Asp
 1235 1240 1245
 Val Val Ser Asn Phe Ser Lys Lys Tyr Thr Leu Ala Glu Gly Asn Leu
 1250 1255 1260
 Glu Lys Ile Lys Glu Leu Tyr Ala Gln Asn Asn Gly Glu Asp Leu Lys

 1265 1270 1275 1280
 Glu Leu Ala Ser Ser Phe Ile Asn Leu Leu Thr Phe Thr Ala Ile Gly
 1285 1290 1295
 Ala Pro Ala Thr Phe Lys Phe Phe Asp Lys Asn Ile Asp Arg Lys Arg
 1300 1305 1310
 Tyr Thr Ser Thr Thr Glu Ile Leu Asn Ala Thr Leu Ile His Gln Ser
 1315 1320 1325
 Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Asn Lys Leu Gly Gly

1330	1335	1340	
Asp			
134			
<210>	7		
<211>	3249		
<212>	DNA		
<213>	N.meningitides		
<400>	7		
atggctgcct tcaaacctaa ttcaatcaac tacatcctcg gcctcgatat cggcatcgca		60	
tccgtcggct gggcgatggg agaaattgac gaagaagaaa accccatccg cctgattgat		120	
ttgggcgtgc gcgtatttga gcgtgccgaa gtaccgaaaa caggcgactc ccttgccatg		180	
gcaaggcgtt tggcgcgcag tggtcgccgc ctgaccgccg gtcgcgccca cgcctgctt		240	
cggaccgccg gcctattgaa acgcgaaggc gtattacaag ccgccaattt tgacgaaaac		300	
ggcttgatta aatccttacc gaatacacca tggcaacttc gcgcagccgc attagaccgc		360	
aaactgacgc ctttagagtg gtcggcagtc ttgttgcat taaatcaaca tcgcggctat		420	
ttatcgcaac ggaaaaacga gggcgaaact gccgataagg agcttggcgc ttgtcttaaa		480	
ggcgtagccg gcaatgccca tgccttacag acaggcgatt tccgcacacc ggccgaattg		540	
gctttaaata aatttgagaa agaaagcggc catatccgca atcagcgcag cgattattcg		600	
catacgttca gccgcaaaga ttacaggcgc gagctgattt tgctgtttga aaaacaaaaa		660	
gaatttggca atccgcatgt ttacggcggc cttaaagaag gtattgaaac cctactgatg		720	
acgcaacgcc ctgccctgtc cggcgatgcc gttcaaaaaa tgttggggca ttgcaccttc		780	
gaaccggcag agccgaaagc cgctaaaaac acctacacag ccgaacgttt catctggctg		840	
accaagctga acaacctgcg tattttagag caaggcagcg agcggccatt gaccgatacc		900	
gaacgcgcca cgcttatgga cgagccatac agaaaaatcca aactgactta cgcacaagcc		960	
cgtaagctgc tgggtttaga agataccgcc tttttcaaag gcttgcgcta tggtaaagac		1020	
aatgccgaag cctcaacatt gatggaaatg aaggcctacc atgccatcag ccgtgcactg		1080	
gaaaaagaag gattgaaaga caaaaaatcc ccattaaacc tttctcccga attacaagac		1140	
gaaatcggca cggcattctc cctgttcaaa accgatgaag acattacagg ccgtctgaaa		1200	
gaccgtatac agcccgaat cttagaagcg ctgttgaaac acatcagctt cgataagttc		1260	
gtccaaattt ccttgaaagc attgcgccga attgtgcctc taatggaaca aggcaaacgt		1320	
tacgatgaag cctgcgccga aatctacgga gaccattacg gcaagaagaa tacggaagaa		1380	

aagatttattc tgccgccgat tcccgccgac gaaatccgca accccgtcgt cttgcgcgcc	1440
ttatctcaag cacgtaaggt cattaacggc gtggtacgcc gttacggctc cccagctcgt	1500
atccatattg aaactgcaag ggaagtaggt aaatcgttta aagaccgcaa agaaattgag	1560
aaacgccaag aagaaaaccg caaagaccgg gaaaaagccg ccgccaaatt ccgagagtat	1620
ttccccaatt ttgtcggaga acccaaatcc aaagatattc tgaaactgcg cctgtacgag	1680
caacaacacg gcaaatgcct gtattcgggc aaagaaatca acttaggccg tctgaacgaa	1740
aaaggctatg tcgaaatcga ccatgccctg ccgttctcgc gcacatggga cgacagtttc	1800
aacaataaag tacttggtatt gggcagcgaa aaccaaaca aaggcaatca aacccttac	1860
gaatacttca acggcaaaga caacagccgc gaatggcagg aatttaaagc gcgtgtcgaa	1920
accagccgtt tcccgcgcag taaaaaaca cggattctgc tgcaaaaatt cgatgaagac	1980
ggctttaaag aacgcaatct gaacgacacg cgctacgtca accgtttcct gtgtcaattt	2040
gttgccgacc gtatgcggct gacaggtaaa ggcaagaaac gtgtctttgc atccaacgga	2100
caaattacca atctgttgcg cggcttttgg ggattgcgca aagtgcgtgc ggaaaacgac	2160
cgccatcacg ccttgagcg cgctcgtctt gcctgctcga ccgttgccat gcagcagaaa	2220
attaccgtt ttgtacgcta taaagagatg aacgcgtttg acgglaaaac catagacaaa	2280
gaaacaggag aagtgtgca tcaaaaaaca cacttccac aaccttgga atttttcgca	2340
caagaagtca tgattcgcgt cttcggcaaa ccggacggca aaccgaatt cgaagaagcc	2400
gataccctag aaaaactgcg cacgttgctt gccgaaaaat tatcatctcg cccgaagcc	2460
gtacacgaat acgttacgcc actgtttgtt tcacgcgcgc ccaatcgga gatgagcggg	2520
caagggcata tggagaccgt caaatccgcc aaacgactgg acgaaggcgt cagcgtgttg	2580
cgcgtaccgc tgacacagtt aaaactgaaa gacttgaaa aaatggtcaa tcgggagcgc	2640
gaacctaaagc tatacgaagc actgaaagca cggctggaag cacataaaga cgatcctgcc	2700
aaagcctttg ccgagccgtt ttacaaatac gataaagcag gcaaccgcac ccaacaggt	2760
aaagccgtac gcgtagagca agtacagaaa accggcgtat gggtgcgcaa ccataacggt	2820
attgccgaca acgcaacat ggtgcgcgta gatgtgtttg agaaaggcga caagtattat	2880
ctggtaccga tttacagttg gcaggtacg aaagggattt tgccggatag ggctgttgta	2940
caaggaaaag atgaagaaga ttggcaactt attgatgata gtttcaactt taaattctca	3000
ttacacccta atgatttagt cgaggttata acaaaaaaag ctagaatgtt tgggtacttt	3060
gccagctgcc atcgaggcac aggtaatatc aatatacgca ttcatgatct tgatcataaa	3120
attggcaaaa atggaatact ggaaggtatc ggcgtaaaaa ccgccctttc attccaaaaa	3180
taccaaatg acgaactggg caaagaaatc agaccatgcc gtctgaaaaa acgcccgcct	3240

gtccgttaa

3249

<210> 8

<211> 1082

<212> PRT

<213> N.meningitides

<400> 8

Met Ala Ala Phe Lys Pro Asn Ser Ile Asn Tyr Ile Leu Gly Leu Asp

1 5 10 15

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

20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg

35 40 45

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

50 55 60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu

65 70 75 80

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

85 90 95

Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln

100 105 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser

115 120 125

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

130 135 140

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

145 150 155 160

Gly Val Ala Gly Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr

165 170 175

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

180 185 190

Arg Asn Gln Arg Ser Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu

195	200	205
Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn		
210	215	220
Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met		
225	230	235
240		
Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly		
245	250	255
His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr		
260	265	270
Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile		
275	280	285
Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr		
290	295	300
Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala		
305	310	315
320		
Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg		
325	330	335
Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala		
340	345	350
Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys		
355	360	365
Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr		
370	375	380
Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys		
385	390	395
400		
Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser		
405	410	415
Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val		
420	425	430
Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile		
435	440	445
Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu		

450 455 460
 Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
 465 470 475 480
 Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly

 485 490 495
 Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
 500 505 510
 Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
 515 520 525
 Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
 530 535 540
 Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
 545 550 555 560

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

 625 630 635 640
 Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
 645 650 655
 Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
 660 665 670
 Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
 675 680 685
 Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
 690 695 700

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

 770 775 780
 Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
 785 790 795 800
 Asp Thr Leu Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
 805 810 815
 Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
 820 825 830
 Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
 835 840 845

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

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

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
965 970 975
Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp
980 985 990
Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu
995 1000 1005
Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys His
1010 1015 1020
Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp His Lys
1025 1030 1035 1040
Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys Thr Ala Leu
1045 1050 1055
Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys Glu Ile Arg Pro
1060 1065 1070
Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1075 1080
<210> 9
<211> 4179
<212> DNA
<213> Streptococcus mitis
<400> 9
atgaacaata acaattactc tatcggactc gatatcggaa caaacagcgt cggatgggcc 60
gtcattacgg atgactataa ggtgccatcg aaaaagatga aagttctagg caatacagat 120
aaacacttta tcaagaaaaa tctaattgga gctttattat ttgatgaagg agctactgct 180
gaagatagac gtttcaaacg aacagcacgc cgtcgtata ctcgtcgaaa aaatcgtctt 240
cgctatcttc aagaaatctt ttctgaggaa atgagcaaag tggatagtag tttctttcat 300
cgattagatg actcattctt agttcctgag gataaaagag gaagtaaata tcctatTTTT 360
gtacacttgg cagaagaaaa agaatacac aagaaatttc caactatcta tcatttgaga 420
aaacaccttg cggactcaaa agaaaaact gacttgcgct tgatctatct agcattagcg 480
catatgatta aataccgcgg acattttttg tatgaagaat ctttcgatat taaaaacaat 540
gatatccaaa aaatctttag cgagtttata agcatttacg acaacacctt tgaaggaagt 600

tcacttagtg gacaaaatgc acaagtagaa gcaattttta ctgataaaat tagtaaatct	660
gctaagagag aacgcattct aaaactcttt gcttatgaaa aatccactga tctattttca	720
gaattttctca agctgattgt aggaaatcaa gctgatttta agaaacactt tgacttgga	780
gaaaaagctc cactacaatt ctctaaagat acctatgatg aggatttgga aaacttactc	840
ggacaaattg gagatgactt tgcagacctt ttcctagtgt ctaaaaaact ctatgatgcc	900
attcttttat caggaatctt aactgttaca gattcttcaa ctaaggcccc actatcagca	960
tctatgattg agcgctatga aaaccaccaa aaagacttag cggcttttaa acaattcatc	1020
caaaacaatc ttcaagaaaa atatgatgaa gttttctctg accaatctaa agatgggtat	1080
gctaggtata tcaatggcaa aaccactcaa gaagcatttt acaagtacat caaaaatctt	1140
ctctctaaat tcgaaggatc agattatttc ctgataaaa ttgaacgtga agatttcttg	1200
agaaaacaac gcaccttga taatggttct atccctcatc aaattcatct tcaagaaatg	1260
aatgccatta tccgtcggca aggagaacat tatccatttc tgaaggaata taaagaaaag	1320
atagagacaa tcttgacttt ccgtattcct tattatgttg gccattggc tcgtggaaat	1380
cgtaattttg cttggttac tcgaaactct gaccaagcaa tccgacctg gaattttgaa	1440
gaaattgttg atcaagcaag ctctgcggaa gaattcatca ataagatgac taactatgac	1500
ttgtatctgc cagaggaaaa agttttgccc aagcatagtc tcttgtatga aacatttgct	1560
gtctacaatg aattaacaaa agtaaaattt atttcagagg gattgagaga ctatcaattc	1620
cttgatagtg ggcaaaagaa gcaaatgtc aatcaattat tcaaagagaa aagaaaagta	1680
actgaaaag acatcattca gtatctacac aatgttgatg gctacgatgg aatcgaacta	1740
aaaggaattg aaaaacaatt taacgctagt ctttctactt atcatgattt actcaaaata	1800
atcaaggata aagagtttat ggatgatcct aaaaatgaag agattcttga aaatatcgtc	1860
cacacactaa ctatctttga agatcgtgag atgatcaagc aacgccttgc tcaatatgcc	1920
tctatctttg ataaaaagt gatcaaggca ctgactcgtc gacattatc tggttgggga	1980
aaactctctg ctaagctaat caacggtatc tgtgataaaa aaactggtaa aacaattctt	2040
gactacttga ttgatgacgg ctacagcaat cgtaacttta tgcagttaat caatgatgac	2100
gggctttcct tcaaagatat tattcaaaaa gcacaagtgg ttggtgaagac aaacgatgtg	2160
aagcaagttg tccaagaact cccaggtagt cctgtatta aaaagggaat tttaaaagt	2220
atcaagcttg tcgatgagct tgtcaaagtt atgggccatg ctcccagatc cattgtgatt	2280
gaaattgcac gagaaaaatca gacaactgcc agagggaaaa agaattctca acaaagatat	2340

aagcgcatg aagatgcact aaaaaattta gcacctgggc ttgattcaaa tatattaaaa	2400
gaacatccaa cagataatat tcaacttcaa aatgaccgtc tcttccttta ctatctccaa	2460
aatgggaagg atatgtacac tggagaagct cttgatatca accaactgag cagctatgac	2520
attgaccaca tcgtccaca ggcctttatc aaggatgatt ctcttgataa cgtgtcttg	2580
actagttaa aggataatcg tgggaaatcc gataatgttc caagtttaga agtcgttcaa	2640
aaaagaaaag ctttttgca acaattacta gattccaaat tgatttcaga acataaattt	2700
aataatttaa ccaaggctga acgtggtggg ctgatgagc gagataaagt tggctttatc	2760
agacgccaac tagttgaaac acggcaaadc aaaaaacatg ttgctcagat ttgggatgcc	2820
cgttttaata cagaagtga tgagaaagat aagaagaacc gtaccgtcaa aattatcact	2880
ttgaaatcca atctagtctt caacttccgt aaagaattta agttatataa ggtacgcgaa	2940
atcaatgact accaccatgc acatgatgcc tatttaaatg cagtgggtggc taaggctatc	3000
cttaagaaat atcctaaact agagcctgaa ttcgtctatg gtgactatca aaagtacgat	3060
attaagagat atatttccag atccaaagat cctaaagaag ttgaaaaagc aactgaaaag	3120
tatttcttct actcaaactt gttgaacttc tttaaagaag aggtgcatta cgcagacgga	3180
accatcgtaa aacgagagaa tatcgaatac tctaaggaca ctggagaaat cgcttggaat	3240
aaagaaaaag atttcgtac aattaaaaaa gtctttcac ttcgcaggt gaatatgtg	3300
aagaaaacag agattcaaac acatgggtcta gatagaggta aacctagagg attgttcaat	3360
tccaatccat ctctaaacc ttcagaagat cgtaaagaaa acctgtgcc aattaacaa	3420
gggcttgacc cagaaaaata cgggtggttac gctggtatct ctaactcata cgcggtctta	3480
gttaaagcta ttattgaaaa aggagcgaaa aaacaacaaa agaccgttct tgaatttcaa	3540
ggtatctcta ttttagataa aataaatttt gaaaagaaca aagaaaacta tcttcttgaa	3600
aaaggataca taaaaattct atcaactatt actttaccta aatatagttt gtttgagttt	3660
cctgatggta caagaagaag actagcaagt attctatga caaacaataa acgaggagaa	3720
attcataaag gtaatgaatt ggtcatccct gaaaagtata cgactctttt gtatcatgct	3780
aagaatatta ataaaacact tgaaccagaa cacttagagt atgttgagaa acatcgaaat	3840
gattttgcta aacttttaga atatgtactt aactttaacg ataagtatgt aggcgcatta	3900
aaaaatggag aaagaatcag acaagcattt attgattggg aaacagttga tattgaaaag	3960
ttatgtttca gtttcattgg tccaagaaat agtaaaaatg ctggtttatt cgagttaact	4020
tcacaaggaa gtgcttctga cttcgagttc ttgggagtaa aaattccacg atacagagac	4080
tatacacctt cgtcactcct caacgccacc ctcatccacc aatccatcac tggcttttac	4140
gagactcgga ttgacttaag caaactggga gaagactga	4179

<210> 10
 <211> 1392
 <212> PRT
 <213> Streptococcus mitis
 <400> 10
 Met Asn Asn Asn Asn Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser
 1 5 10 15
 Val Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys
 20 25 30
 Met Lys Val Leu Gly Asn Thr Asp Lys His Phe Ile Lys Lys Asn Leu
 35 40 45
 Ile Gly Ala Leu Leu Phe Asp Glu Gly Ala Thr Ala Glu Asp Arg Arg
 50 55 60
 Phe Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Leu
 65 70 75 80
 Arg Tyr Leu Gln Glu Ile Phe Ser Glu Glu Met Ser Lys Val Asp Ser
 85 90 95
 Ser Phe Phe His Arg Leu Asp Asp Ser Phe Leu Val Pro Glu Asp Lys
 100 105 110
 Arg Gly Ser Lys Tyr Pro Ile Phe Ala Thr Leu Ala Glu Glu Lys Glu
 115 120 125
 Tyr His Lys Lys Phe Pro Thr Ile Tyr His Leu Arg Lys His Leu Ala
 130 135 140
 Asp Ser Lys Glu Lys Thr Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala
 145 150 155 160
 His Met Ile Lys Tyr Arg Gly His Phe Leu Tyr Glu Glu Ser Phe Asp
 165 170 175
 Ile Lys Asn Asn Asp Ile Gln Lys Ile Phe Ser Glu Phe Ile Ser Ile
 180 185 190
 Tyr Asp Asn Thr Phe Glu Gly Ser Ser Leu Ser Gly Gln Asn Ala Gln
 195 200 205
 Val Glu Ala Ile Phe Thr Asp Lys Ile Ser Lys Ser Ala Lys Arg Glu

210 215 220
 Arg Ile Leu Lys Leu Phe Ala Tyr Glu Lys Ser Thr Asp Leu Phe Ser
 225 230 235 240

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

 305 310 315 320
 Ser Met Ile Glu Arg Tyr Glu Asn His Gln Lys Asp Leu Ala Ala Leu
 325 330 335
 Lys Gln Phe Ile Gln Asn Asn Leu Gln Glu Lys Tyr Asp Glu Val Phe
 340 345 350
 Ser Asp Gln Ser Lys Asp Gly Tyr Ala Arg Tyr Ile Asn Gly Lys Thr
 355 360 365
 Thr Gln Glu Ala Phe Tyr Lys Tyr Ile Lys Asn Leu Leu Ser Lys Phe
 370 375 380

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

 450 455 460

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

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

 595 600 605
 Asp Pro Lys Asn Glu Glu Ile Leu Glu Asn Ile Val His Thr Leu Thr
 610 615 620
 Ile Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ala Gln Tyr Ala
 625 630 635 640
 Ser Ile Phe Asp Lys Lys Val Ile Lys Ala Leu Thr Arg Arg His Tyr
 645 650 655
 Thr Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Cys Asp
 660 665 670

 Lys Lys Thr Gly Lys Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Tyr
 675 680 685
 Ser Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Gly Leu Ser Phe
 690 695 700
 Lys Asp Ile Ile Gln Lys Ala Gln Val Val Gly Lys Thr Asn Asp Val
 705 710 715 720

Lys Gln Val Val Gln Glu Leu Pro Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735
 Ile Leu Gln Ser Ile Lys Leu Val Asp Glu Leu Val Lys Val Met Gly
 740 745 750
 His Ala Pro Glu Ser Ile Val Ile Glu Ile Ala Arg Glu Asn Gln Thr
 755 760 765
 Thr Ala Arg Gly Lys Lys Asn Ser Gln Gln Arg Tyr Lys Arg Ile Glu
 770 775 780
 Asp Ala Leu Lys Asn Leu Ala Pro Gly Leu Asp Ser Asn Ile Leu Lys
 785 790 795 800
 Glu His Pro Thr Asp Asn Ile Gln Leu Gln Asn Asp Arg Leu Phe Leu
 805 810 815
 Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly Glu Ala Leu Asp
 820 825 830
 Ile Asn Gln Leu Ser Ser Tyr Asp Ile Asp His Ile Val Pro Gln Ala
 835 840 845
 Phe Ile Lys Asp Asp Ser Leu Asp Asn Arg Val Leu Thr Ser Ser Lys
 850 855 860
 Asp Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Leu Glu Val Val Gln
 865 870 875 880
 Lys Arg Lys Ala Phe Trp Gln Gln Leu Leu Asp Ser Lys Leu Ile Ser
 885 890 895
 Glu His Lys Phe Asn Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Asp
 900 905 910
 Glu Arg Asp Lys Val Gly Phe Ile Arg Arg Gln Leu Val Glu Thr Arg
 915 920 925
 Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ala Arg Phe Asn Thr
 930 935 940
 Glu Val Asn Glu Lys Asp Lys Lys Asn Arg Thr Val Lys Ile Ile Thr
 945 950 955 960
 Leu Lys Ser Asn Leu Val Ser Asn Phe Arg Lys Glu Phe Lys Leu Tyr

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

 1025 1030 1035 1040
 Tyr Phe Phe Tyr Ser Asn Leu Leu Asn Phe Phe Lys Glu Glu Val His
 1045 1050 1055
 Tyr Ala Asp Gly Thr Ile Val Lys Arg Glu Asn Ile Glu Tyr Ser Lys
 1060 1065 1070
 Asp Thr Gly Glu Ile Ala Trp Asn Lys Glu Lys Asp Phe Ala Thr Ile
 1075 1080 1085
 Lys Lys Val Leu Ser Leu Pro Gln Val Asn Ile Val Lys Lys Thr Glu
 1090 1095 1100

 Ile Gln Thr His Gly Leu Asp Arg Gly Lys Pro Arg Gly Leu Phe Asn
 1105 1110 1115 1120
 Ser Asn Pro Ser Pro Lys Pro Ser Glu Asp Arg Lys Glu Asn Leu Val
 1125 1130 1135
 Pro Ile Lys Gln Gly Leu Asp Pro Arg Lys Tyr Gly Gly Tyr Ala Gly
 1140 1145 1150
 Ile Ser Asn Ser Tyr Ala Val Leu Val Lys Ala Ile Ile Glu Lys Gly
 1155 1160 1165
 Ala Lys Lys Gln Gln Lys Thr Val Leu Glu Phe Gln Gly Ile Ser Ile

 1170 1175 1180
 Leu Asp Lys Ile Asn Phe Glu Lys Asn Lys Glu Asn Tyr Leu Leu Glu
 1185 1190 1195 1200
 Lys Gly Tyr Ile Lys Ile Leu Ser Thr Ile Thr Leu Pro Lys Tyr Ser
 1205 1210 1215
 Leu Phe Glu Phe Pro Asp Gly Thr Arg Arg Arg Leu Ala Ser Ile Leu

1220 1225 1230
 Ser Thr Asn Asn Lys Arg Gly Glu Ile His Lys Gly Asn Glu Leu Val
 1235 1240 1245
 Ile Pro Glu Lys Tyr Thr Thr Leu Leu Tyr His Ala Lys Asn Ile Asn
 1250 1255 1260
 Lys Thr Leu Glu Pro Glu His Leu Glu Tyr Val Glu Lys His Arg Asn
 1265 1270 1275 1280
 Asp Phe Ala Lys Leu Leu Glu Tyr Val Leu Asn Phe Asn Asp Lys Tyr
 1285 1290 1295
 Val Gly Ala Leu Lys Asn Gly Glu Arg Ile Arg Gln Ala Phe Ile Asp
 1300 1305 1310
 Trp Glu Thr Val Asp Ile Glu Lys Leu Cys Phe Ser Phe Ile Gly Pro

1315 1320 1325
 Arg Asn Ser Lys Asn Ala Gly Leu Phe Glu Leu Thr Ser Gln Gly Ser
 1330 1335 1340
 Ala Ser Asp Phe Glu Phe Leu Gly Val Lys Ile Pro Arg Tyr Arg Asp
 1345 1350 1355 1360
 Tyr Thr Pro Ser Ser Leu Leu Asn Ala Thr Leu Ile His Gln Ser Ile
 1365 1370 1375
 Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Lys Leu Gly Glu Asp
 1380 1385 1390

<210> 11

<211> 4017

<212> DNA

<213> Streptococcus macacae

<400> 11

atgacaaaac cttattctat tggacttgat attgggacta actctgttgg ttgggctgtt 60
 gtgacagatg gctacaaagt tcctgctaag aagatgaagg ttctgggaaa tacagataaa 120
 agccatatca agaaaaattt acttggagct ttattgtttg atagcggtaa tactgcaaaa 180
 gacagacgtt tgaagcggac agctaggcgt cgatatacac gtcgtagaaa ccgtatttta 240
 tatttgcagg aaatttttgc tgaagaaatg gctaaagcag acgaaagttt cttccagcgc 300

ttaaacgaat cgttttaac aaatgatgac aaagaatttg attctcatcc aatctttggg	360
aataaagctg aagaggaggc tcatcacat aaatttccaa caatttttca ttgcgaaag	420
catttagcag actcaaccga gaaatctgat ttgcgcttaa tttatctagc tttagcgcat	480
atgattaaat tccggggaca tttcttaatt gaaggtcagc taaaagctga aaatacaaat	540
gttcaaacat tatttgacga tttttagaa gtatatgata agacagttga agaaagtcac	600
ttatcagaaa ttagtgtctc cagtattctg acagaaaaa ttagtaaatc gcgtcgctta	660
gaaaatctta taaaatacta tccactgag aagaaaaaca ctctcttcgg aatcttacc	720
gccttgctct taggattaca gccaaacttt aaaacaaatt ttaaattatc cgaagatgct	780
aaactacagt ttictaagga tacttatgaa gaagatttag gagaattact tggaaaaatc	840
ggagataatt atgcagattt atttatatca gctaaaaatc tttatgatgc tattttgcta	900
tcaggaaatt taacaataga tgacaacacg acaaggctc cgttgtctgc ttcaatgatt	960
aaacgttatg aggaacatca ggaagattta gcacaactta agaaatttat ccgtcagaat	1020
ttaccagatc aatatagtga ggttttttct gataaaacaa aggatggcta tgctggttat	1080
attgatgaa aaacgaatca ggaggccttt tataaatata tcaaaaatat gctgtcaaaa	1140
acagaagggtg cagattattt tcttgacaaa attgatcgtg aagacttttt gagaaaacag	1200
agaacgtttg ataatggttc cgttccgat cagattcatc tgcaagagat gcatgctatt	1260
ttacgacgtc aggttgaata ctatccattc ttgaaagaaa atcaggataa aattgaaaaa	1320
atcttaacgt ttagaattcc ttactagtt ggtccttttg cgcgaaaagg tagccgcttt	1380
gcctgggcag aatacaaggc ggataaaaaa gttacgcat ggaattttga tgatattctt	1440
gataaagaaa aatcagcaga agaattcatc acacgcatga ctttaaatga ttgtattta	1500
cctgaagaaa aagtcctacc aaagcatagt ctgtttatg aaacgtttta tgtttacaat	1560
gagttaacta aagttaagta tgtcaatgag caagggaag ccattttctt tgatgccaat	1620
atgaagcaag agatttttga tcatgttttt aaagaaaatc ggaaagttac taaagataaa	1680
cttttaaat atttgaataa agagtttgaa gaatttagaa ttgttaactt aactggactg	1740
gataaggaaa ataaagcctt taattccagt ctggaacct atcatgattt gcgtaaaatt	1800
ttagataaat cattcttaga tgataaagta aatgaaaaga taattgagga tatcattcaa	1860
acactaactc tgtttgaaga cagagaaatg attcgtcagc gtcttcaaaa gtatagtgat	1920
atttttacaa cacagcaatt gaaaaaactt gaacgccgtc attatacagg ttggggaaga	1980
ttatcagcga agttaatcaa tggatttcga gataaacaga gtaataagac tattctgggt	2040

tatttgattg atgatggta tagcaatcgt aactttatgc agttgattaa tgacgattct	2100
cttcctttta aagaagaaat tgctagggca caagtcattg gagaacaga tgacttaaat	2160
caacttgta gtgatattgc tggcagtcct gctattaaaa aggggaatttt acaaagtctg	2220
aaaattgtag atgagcttgt taaagtcag gggcataatc ctgctaaca tgttatcgaa	2280
atggcgctg aaaatcagac tacagccaaa gggcgctgca gttcacagca acgttataaa	2340
cgacttgagg aggcaataaa aaatcttgac catgatttaa atcataagat tttaaaagaa	2400
caccaacag ataataagc tttacagaat gaccgtcttt tcttatatta tctccaaaat	2460
ggccgagata tgtatactga agatccactt gatattaatc gtttaagtga ttatgatatc	2520
gaccatatta ttccacaatc ttttataaaa gatgactcta ttgacaataa ggttctggtt	2580
tcatcagcta aaaaccgtgg gaaatcggat aatgtaccga gtgaagatgt tgtcaatagg	2640
atgagaccgt tttggaataa attattgagc tgtggattga tttctcaacg gaaatacagc	2700
aatctaacca aaaaagaatt aaaaccagat gataaggctg gtttcatcaa acgtcaattg	2760
gttgagacaa gacaaattac aaagcatgtt gcacaaattt tagacgctcg ttttaataca	2820
aaacgtgatg aaaataaaaa agtaattcgt gatgtcaaaa ttatcacttt aaaatctaata	2880
ttagtttcac aatttcgtaa agactttaaa ttttacaag tacgtgagat taatgattac	2940
catcatgcgc atgacgctta tcttaatgca gttataggaa aagctttatt agatgtttat	3000
ccgcagttag agcccgaaat tgtttatggt gactaccctc attttcatgg atataaagaa	3060
aataaagcaa ctgctaagaa atttttctat tcaaatatta tgaatTTTTT taagaaagat	3120
gatatccgta ccgatgaaaa tggtagatgt gtttgaaaa aagatgagca tatttctaata	3180
attaaaaggg tgctttccta tccccaagtt aatattgtta agaaagtaga aatacagact	3240
gttgacaaa atgggggact ttttgacgat aatcctaata caccattaga ggttacacct	3300
agtaaacctg ttccactaaa aaaagaatta aaccctaaaa aatatggagg atatcaaaaa	3360
ccgacgacag cttatcctgt tttactgata acagatacta aacagctaata tccaatctca	3420
gtaatgaata agaagcaatt tgaacaaaat ccggttaaat ttttaagaga tagaggctat	3480
caacaggtag gaaagaatga ctttattaaa ttacccaaat ataccctagt tgatatcgggt	3540
gatgggatta aacgcctatg ggctagtctg aaagaaatac ataaaggaaa tcaattagtt	3600
gtatctaaaa aatctcaaat tttgctttat catgcacatc acttagatag tgatttgagt	3660
aatgattatc ttcaaaatca taatcaacaa ttcgatgttt tatttaatga aattatttct	3720
ttttctaaaa aatgtaaat gggaaaagaa catattcaga aaattgaaaa tgtttactcc	3780
aataagaaga atagtgcac aatagaagaa ttagcagaga gttttattaa attattagga	3840
tttacacaat taggtgcaac ttccccattt aatttttttag gggtaaaact aatcaaaaa	3900

caatataaag gtaaaaaaga ttatatTTTA ccgtgtacag aggggaccct tatccgcaa 3960

tctatcactg gtctttacga aacacgagtt gatcttagta aaataggaga agactaa 4017

<210> 12

<211> 1338

<212> PRT

<213> Streptococcus macacae NCTC 11558

<400> 12

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

1 5 10 15
Gly Trp Ala Val Val Thr Asp Gly Tyr Lys Val Pro Ala Lys Lys Met

20 25 30
Lys Val Leu Gly Asn Thr Asp Lys Ser His Ile Lys Lys Asn Leu Leu

35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Asn Thr Ala Lys Asp Arg Arg Leu

50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ala Glu Glu Met Ala Lys Ala Asp Glu Ser
85 90 95

Phe Phe Gln Arg Leu Asn Glu Ser Phe Leu Thr Asn Asp Asp Lys Glu
100 105 110

Phe Asp Ser His Pro Ile Phe Gly Asn Lys Ala Glu Glu Glu Ala His
115 120 125

His His Lys Phe Pro Thr Ile Phe His Leu Arg Lys His Leu Ala Asp
130 135 140

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

145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Gln Leu Lys Ala

165 170 175
Glu Asn Thr Asn Val Gln Thr Leu Phe Asp Asp Phe Val Glu Val Tyr

180 185 190
Asp Lys Thr Val Glu Glu Ser His Leu Ser Glu Ile Ser Val Ser Ser

195	200	205	
Ile Leu Thr Glu Lys Ile Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile			
210	215	220	
Lys Tyr Tyr Pro Thr Glu Lys Lys Asn Thr Leu Phe Gly Asn Leu Ile			
225	230	235	240
Ala Leu Ser Leu Gly Leu Gln Pro Asn Phe Lys Thr Asn Phe Lys Leu			
	245	250	255
Ser Glu Asp Ala Lys Leu Gln Phe Ser Lys Asp Thr Tyr Glu Glu Asp			
	260	265	270
Leu Gly Glu Leu Leu Gly Lys Ile Gly Asp Asn Tyr Ala Asp Leu Phe			
	275	280	285
Ile Ser Ala Lys Asn Leu Tyr Asp Ala Ile Leu Leu Ser Gly Ile Leu			
290	295	300	
Thr Ile Asp Asp Asn Thr Thr Lys Ala Pro Leu Ser Ala Ser Met Ile			
305	310	315	320
Lys Arg Tyr Glu Glu His Gln Glu Asp Leu Ala Gln Leu Lys Lys Phe			
	325	330	335
Ile Arg Gln Asn Leu Pro Asp Gln Tyr Ser Glu Val Phe Ser Asp Lys			
	340	345	350
Thr Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn Gln Glu			
	355	360	365
Ala Phe Tyr Lys Tyr Ile Lys Asn Met Leu Ser Lys Thr Glu Gly Ala			
	370	375	380
Asp Tyr Phe Leu Asp Lys Ile Asp Arg Glu Asp Phe Leu Arg Lys Gln			
385	390	395	400
Arg Thr Phe Asp Asn Gly Ser Val Pro His Gln Ile His Leu Gln Glu			
	405	410	415
Met His Ala Ile Leu Arg Arg Gln Gly Glu Tyr Tyr Pro Phe Leu Lys			
	420	425	430
Glu Asn Gln Asp Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr			
435	440	445	

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

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

 580 585 590
 Thr Tyr His Asp Leu Arg Lys Ile Leu Asp Lys Ser Phe Leu Asp Asp
 595 600 605
 Lys Val Asn Glu Lys Ile Ile Glu Asp Ile Ile Gln Thr Leu Thr Leu
 610 615 620
 Phe Glu Asp Arg Glu Met Ile Arg Gln Arg Leu Gln Lys Tyr Ser Asp
 625 630 635 640
 Ile Phe Thr Thr Gln Gln Leu Lys Lys Leu Glu Arg Arg His Tyr Thr
 645 650 655

 Gly Trp Gly Arg Leu Ser Ala Lys Leu Ile Asn Gly Ile Arg Asp Lys
 660 665 670
 Gln Ser Asn Lys Thr Ile Leu Gly Tyr Leu Ile Asp Asp Gly Tyr Ser
 675 680 685
 Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Ser Leu Pro Phe Lys
 690 695 700

Glu Glu Ile Ala Arg Ala Gln Val Ile Gly Glu Thr Asp Asp Leu Asn
 705 710 715 720
 Gln Leu Val Ser Asp Ile Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

 725 730 735
 Leu Gln Ser Leu Lys Ile Val Asp Glu Leu Val Lys Val Met Gly His
 740 745 750
 Asn Pro Ala Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr
 755 760 765
 Ala Lys Gly Arg Arg Ser Ser Gln Gln Arg Tyr Lys Arg Leu Glu Glu
 770 775 780
 Ala Ile Lys Asn Leu Asp His Asp Leu Asn His Lys Ile Leu Lys Glu
 785 790 795 800

 His Pro Thr Asp Asn Gln Ala Leu Gln Asn Asp Arg Leu Phe Leu Tyr
 805 810 815
 Tyr Leu Gln Asn Gly Arg Asp Met Tyr Thr Glu Asp Pro Leu Asp Ile
 820 825 830
 Asn Arg Leu Ser Asp Tyr Asp Ile Asp His Ile Ile Pro Gln Ser Phe
 835 840 845
 Ile Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Val Ser Ser Ala Lys
 850 855 860
 Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Asp Val Val Asn Arg

 865 870 875 880
 Met Arg Pro Phe Trp Asn Lys Leu Leu Ser Cys Gly Leu Ile Ser Gln
 885 890 895
 Arg Lys Tyr Ser Asn Leu Thr Lys Lys Glu Leu Lys Pro Asp Asp Lys
 900 905 910
 Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
 915 920 925
 His Val Ala Gln Ile Leu Asp Ala Arg Phe Asn Thr Lys Arg Asp Glu
 930 935 940

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

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

 1010 1015 1020
 Ala Lys Lys Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Lys Asp
 1025 1030 1035 1040
 Asp Ile Arg Thr Asp Glu Asn Gly Glu Ile Val Trp Lys Lys Asp Glu
 1045 1050 1055
 His Ile Ser Asn Ile Lys Arg Val Leu Ser Tyr Pro Gln Val Asn Ile
 1060 1065 1070
 Val Lys Lys Val Glu Ile Gln Thr Val Gly Gln Asn Gly Gly Leu Phe
 1075 1080 1085

 Asp Asp Asn Pro Lys Ser Pro Leu Glu Val Thr Pro Ser Lys Leu Val
 1090 1095 1100
 Pro Leu Lys Lys Glu Leu Asn Pro Lys Lys Tyr Gly Gly Tyr Gln Lys
 1105 1110 1115 1120
 Pro Thr Thr Ala Tyr Pro Val Leu Leu Ile Thr Asp Thr Lys Gln Leu
 1125 1130 1135
 Ile Pro Ile Ser Val Met Asn Lys Lys Gln Phe Glu Gln Asn Pro Val
 1140 1145 1150
 Lys Phe Leu Arg Asp Arg Gly Tyr Gln Gln Val Gly Lys Asn Asp Phe

 1155 1160 1165
 Ile Lys Leu Pro Lys Tyr Thr Leu Val Asp Ile Gly Asp Gly Ile Lys
 1170 1175 1180
 Arg Leu Trp Ala Ser Ser Lys Glu Ile His Lys Gly Asn Gln Leu Val
 1185 1190 1195 1200
 Val Ser Lys Lys Ser Gln Ile Leu Leu Tyr His Ala His His Leu Asp

1205				1210				1215							
Ser	Asp	Leu	Ser	Asn	Asp	Tyr	Leu	Gln	Asn	His	Asn	Gln	Gln	Phe	Asp
1220				1225				1230							
Val	Leu	Phe	Asn	Glu	Ile	Ile	Ser	Phe	Ser	Lys	Lys	Cys	Lys	Leu	Gly
1235				1240				1245							
Lys	Glu	His	Ile	Gln	Lys	Ile	Glu	Asn	Val	Tyr	Ser	Asn	Lys	Lys	Asn
1250				1255				1260							
Ser	Ala	Ser	Ile	Glu	Glu	Leu	Ala	Glu	Ser	Phe	Ile	Lys	Leu	Leu	Gly
1265				1270				1275				1280			
Phe	Thr	Gln	Leu	Gly	Ala	Thr	Ser	Pro	Phe	Asn	Phe	Leu	Gly	Val	Lys
1285				1290				1295							
Leu	Asn	Gln	Lys	Gln	Tyr	Lys	Gly	Lys	Lys	Asp	Tyr	Ile	Leu	Pro	Cys
1300				1305				1310							
Thr	Glu	Gly	Thr	Leu	Ile	Arg	Gln	Ser	Ile	Thr	Gly	Leu	Tyr	Glu	Thr
1315				1320				1325							
Arg	Val	Asp	Leu	Ser	Lys	Ile	Gly	Glu	Asp						
1330				1335											
<210>	13														
<211>	4107														
<212>	DNA														
<213>	Streptococcus pyogenes														
<400>	13														
atggataaga aataactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60															
atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120															
cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180															
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240															
tatctacagg agattttttc aaatgagatg gcgaaagtag atgatatgtt ctttcatcga 300															
cttgaagagt ctttttttgt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360															
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420															
aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcgt 480															
atgattaagt ttcgtgggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540															

gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct	600
attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga	660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttggaat	720
ctcattgctt tgcattggg tttgaccct aattttaaat caaatTTTga tttggcagaa	780
gatgctaaat tacagcttcc aaaagatact tacgatgatg atttagataa tttattggcg	840
caaatggag atcaatatgc tgatttgtt tttggcagcta agaatttatc agatgctatt	900
ttactttcag atatcctaag agtaaatact gaaataacta aggctccct atcagcttca	960
atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaagc tttagtctga	1020
caacaacttc cagaaaagta taaagaaatc tttttgatc aatcaaaaaa cggatatgca	1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta	1140
gaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgtcgcgc	1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat	1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt	1320
gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt	1380
cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa	1440
gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa	1500
aatcttccaa atgaaaaagt actacaaaaa catagtttgc tttatgagta tttacgggt	1560
tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt	1620
tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc	1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt	1740
tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt	1800
attaaagata aagatTTTTT ggataatgaa gaaaatgaag atatcttaga ggatattgtt	1860
ttaacattga cctattttga agataggag atgattgagg aaagacttaa aacatatgct	1920
cacctcttg atgataaggt gatgaacag cttaaactgc gccgttatc tggttgggga	1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta	2040
gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat	2100
agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagtta	2160
catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggat tttacagact	2220
gtaaaagtgt ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt	2280
attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt	2340
atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct	2400

gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460
gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac 2520
attgttccac aaagtttctt taaagacgat tcaatagaca ataaggtctt aacgcgttct 2580
gataaaaatc gtggtaaadc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa 2640

aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700
acgaaagctg aacgtggagg tttagagtga cttgataaag ctggttttat caaacgcca 2760
ttggttgaaa ctgcctaaat cactaagcat gtggcacaaa ttttgatag tcgcatgaat 2820
actaaatagc atgaaaatga taaacttatt cgagaggtta aagtattac cttaaaatct 2880
aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940
taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000
tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga gtctcgtaaa 3060

atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120
aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180
cctctaatacg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatatt 3240
gccacagtgc gcaaagtatt gtccatgcc caagtcaata ttgtcaagaa aacagaagta 3300
cagacaggcg gattctccaa ggagtcaatt ttacaaaaa gaaattcgga caagcttatt 3360
gtctgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420
tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt 3480

aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac 3540
tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600
tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660
caaaaaggaa atgagctggc tctgccaagc aaatatgtga attttttata tttagctagt 3720
cattatgaaa agttgaagg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780
cagcataagc attattttaga tgagattatt gagcaaatca gtgaatttct taagcgtgtt 3840
attttagcag atgccaattt agataaagtt cttagtgcac ataacaaaca tagagacaaa 3900

ccaatcgtg aacaagcaga aaatattatt cattatttta cgttgacgaa tcttgagct 3960
cccgtcgtt ttaaataatt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080
gatttgagtc agctaggagg tgactga 4107

<210> 14

<211> 1368

<212> PRT

<213> Streptococcus pyogenes A20]

<400> 14

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

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

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

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

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

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

 515 520 525
 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575
 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590

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

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

	725	730	735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly			
	740	745	750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln			
	755	760	765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile			
	770	775	780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro			
785	790	795	800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu			
	805	810	815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg			
	820	825	830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys			
	835	840	845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg			
	850	855	860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys			
865	870	875	880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys			
	885	890	895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp			
	900	905	910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr			
	915	920	925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp			
	930	935	940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser			
945	950	955	960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg			
	965	970	975

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

 Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
 1025 1030 1035 1040
 Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
 1045 1050 1055
 Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile
 1060 1065 1070
 Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
 1075 1080 1085
 Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly

 1090 1095 1100
 Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
 1105 1110 1115 1120
 Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
 1125 1130 1135
 Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
 1140 1145 1150
 Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
 1155 1160 1165

 Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
 1170 1175 1180
 Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
 1185 1190 1195 1200
 Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser
 1205 1210 1215
 Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
 1220 1225 1230

Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

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

1265 1270 1275 1280
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys

1285 1290 1295
His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

1300 1305 1310
Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp

1315 1320 1325
Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp

1330 1335 1340
Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1345 1350 1355 1360
Asp Leu Ser Gln Leu Gly Gly Asp

1365

<210> 15
<211> 867
<212> DNA
<213> Human immunodeficiency virus 1

<220><221> misc_feature
<222> (91)
<223> n is a, c, g, or t
<220><221> misc_feature
<222> (202)
<223> n is a, c, g, or t
<220><221> misc_feature
<222> (231)
<223> n is a, c, g, or t
<220><221> misc_feature

<222> (376)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (857)

<223> n is a, c, g, or t

<400> 15

tttttggatg gaatagatag ggccaagaa gagcatgaga aatatcacia taattggaga 60

gcaatggcta gtgattttaa cctgccacct ntagtagcaa aggagatagt agccagctgt 120

gataaatgtc agctaaaagg agaagccatg catggacaag tagactgtag tccaggaata 180

tggcaactag attgtacaca tntagaagga aaagtatatcc tggtagcagt ncatgtagcc 240

agtgggtata tagaagcaga agttattcca gcagagacag ggagagaaac agcatacttc 300

ctcttaaaat tagcaggaag atggccagta aaaacagtac atacagacia tggcagcaac 360

ttcaccagtg ctgcgntgaa ggccgcctgt tggtaggcag ggatcaagca ggaatttggc 420

attcctaca atcccaaag tcaaggagta gtagagtcta tgaataatga attaaagaaa 480

attgtaggac aagtaagaga tcaggctgag catctcaaga cagcagtaca aatggcagta 540

ttcatccaca attttaaaag aaaagggggg attggggggt acagtgcagg agaaagaata 600

gtagacataa tagccacaga catacaaaact aaagaactac aaaaaaatat tacaaaaatg 660

caaaattttc gggctctattt cagagacagc agagatccac tttggaaagg accagcaaag 720

cttctctgga aaggtgaagg ggcagtagta atacaagata ccaatgacat aaargtagtg 780

ccargaagaa aagcaaagat cattagagat tatggaaaac agatggcagg tgatgattgt 840

gtggcaagta gacaggnitga ggattag 867

<210> 16

<211> 288

<212> PRT

<213> Human immunodeficiency virus 1

<220><221> MISC_FEATURE

<222> (31)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (68)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (126)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (262)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (286)

<223> Xaa can be any naturally occurring amino acid

<400> 16

Phe Leu Asp Gly Ile Asp Arg Ala Gln Glu Glu His Glu Lys Tyr His

1 5 10 15

Asn Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Xaa Val

20 25 30

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

35 40 45

Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp

50 55 60

Cys Thr His Xaa Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala

65 70 75 80

Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu

85 90 95

Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr

100 105 110

Val His Thr Asp Asn Gly Ser Asn Phe Thr Ser Ala Ala Xaa Lys Ala

115 120 125

Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn

130 135 140

Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Asn Glu Leu Lys Lys

145 150 155 160

Ile Val Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val

165 170 175

Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly

180 185 190
 Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile
 195 200 205
 Gln Thr Lys Glu Leu Gln Lys Asn Ile Thr Lys Met Gln Asn Phe Arg

210 215 220
 Val Tyr Phe Arg Asp Ser Arg Asp Pro Leu Trp Lys Gly Pro Ala Lys
 225 230 235 240
 Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Thr Asn Asp
 245 250 255
 Ile Lys Val Val Pro Xaa Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly
 260 265 270
 Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Xaa Glu Asp
 275 280 285

<210> 17
 <211> 140
 <212> DNA
 <213> Simian T-lymphotropic virus 1
 <400> 17
 gacttgtaga acgctctaataa ggcattctta aaacctatt atataagtac ttactgaca 60
 aaccgcacct acctatggat aatgctctat ccatagccct atggacgac aaccacctga 120
 atgtgttaac ccactgccac 140

<210> 18
 <211> 46
 <212> PRT
 <213> Simian T-lymphotropic virus 1
 <400> 18
 Leu Val Glu Arg Ser Asn Gly Ile Leu Lys Thr Leu Leu Tyr Lys Tyr
 1 5 10 15

Phe Thr Asp Lys Pro Asp Leu Pro Met Asp Asn Ala Leu Ser Ile Ala
 20 25 30
 Leu Trp Thr Ile Asn His Leu Asn Val Leu Thr His Cys His
 35 40 45

<210> 19
 <211> 1509
 <212> DNA
 <213> Streptococcus pneumoniae
 <400> 19

gagttttttt ctttcgtag caagggttta gagccctat tttattttac tattgtctaa 60
 acaccaagcg aacacaaaa ctacatgca atggaaaaac ctctgatttg attctcactt 120
 gatttcacaa tctttatatac aaactgtggg tggattttga caatatcttt ttgattttt 180

aatagtaaat tcgaaataat atttttaggt gagtaacgtg gactaagatg taacaagtct 240
 ttgaactcat cgacacttaa ttctacttta ttgctattat cactagtctt aatgaatctt 300
 tcaattattc tggaatatct acaggtataa cttttcaatt cttcaaaatg gaaattgtga 360
 ttttctacaa attgatttaa ggctttttaca gtattttctt gtgaacgatt tatattatgt 420
 gatatgccca ttgttgtctc aaagtttagcg tgcctactc tagtcataat atctttcact 480
 gctatgtgca tctcattact ttgaaggtaa ctaatatgca tatgcctaaa cgaatgggga 540
 glaacatgtt ttaccactt aaaaccatag tcacttaaac aatttgtcaa taattttcct 600

tctattcggt tcaaaatttg acgaaaagtg ctgtatgtta ttggagagcc gtattctgtt 660
 ctaaatacac tticagaatg tgtaaaagca ggacagggat gtttctccat ataagcatca 720
 aactctttat ttctctgtat tgccttttta atagcttcgc ttgcagcttc aggcaaagct 780
 acttctctaa ttgaattgag tgttttagtt gtatcaaaat gaaattgttt aactttttaa 840
 caatgatatt gaagtgtctt atcaatatgc aagattcctt tttcaaaatc aatatctgat 900
 ggtaaaaatg ctgcttact aattcgaata cctgtaagca acaatactat agcaagatca 960
 taatagtttg catttctgca ttggcgtaac acatcaaaaa atgcatgtaa ttcattggatt 1020

tctagaaatt tagaatcatg tctttctttt gctttacgcc ttttctctag tgaaatatct 1080
 agttttaccg cagtcattgg agaaaactta atgacattat ataacacacc atgattaaaa 1140
 atcttattac aagtactttt tatatgagtc attgttgaag gcgatgcac atacatttct 1200
 aaatatctat tgagactatt ttcatcaga agtggagtaa tcctgtctaa caaaaaatca 1260
 tctctataa ttttccaaag acgttccata accagtagtt ctctctgaat tgtttgtggt 1320
 ttaacagaga cacaccaagt ctgaaaccaa ttttctttta actctccaaa tgttgtaatc 1380
 agttcaggac tatactgact ttcaaatgaa gtagttagtc tatctatctt atcaagaacc 1440

tctctttcag cttgtttcct cgccctacta gtattcttag tataacttac agttactgat 1500
 ttccacttt 1509

<210> 20
 <211> 502
 <212> PRT
 <213> Streptococcus pneumoniae
 <400> 20
 Met Tyr Tyr Val Thr Lys Thr Asn Ser Lys Gly Gln Pro Leu Tyr Gln
 1 5 10 15
 Val Val Glu Lys Tyr Lys Asp Pro Leu Thr Gly Lys Trp Lys Ser Val
 20 25 30
 Thr Val Ser Tyr Thr Lys Asn Thr Ser Arg Ala Arg Lys Gln Ala Glu
 35 40 45
 Arg Glu Val Leu Asp Lys Ile Asp Arg Leu Thr Thr Ser Phe Glu Ser
 50 55 60
 Gln Tyr Ser Pro Glu Leu Ile Thr Thr Phe Gly Glu Leu Lys Glu Asn
 65 70 75 80
 Trp Phe Gln Thr Trp Cys Val Ser Val Lys Pro Gln Thr Ile Gln Arg
 85 90 95
 Glu Leu Leu Val Met Lys Arg Leu Gly Lys Ile Ile Gly Asp Asp Phe
 100 105 110
 Leu Leu Asp Arg Ile Thr Pro Leu Leu Met Lys Asn Ser Leu Asn Lys
 115 120 125
 Tyr Leu Glu Met Tyr Asp Ala Ser Pro Ser Thr Met Thr His Ile Lys
 130 135 140
 Ser Thr Cys Asn Lys Ile Phe Asn His Gly Val Leu Tyr Asn Val Ile
 145 150 155 160
 Lys Phe Ser Pro Met Thr Ala Val Lys Leu Asp Ile Ser Leu Glu Lys
 165 170 175
 Arg Arg Lys Ala Lys Glu Arg His Asp Ser Lys Phe Leu Glu Ile His
 180 185 190
 Glu Leu His Ala Phe Phe Asp Val Leu Arg Gln Cys Arg Asn Ala Asn
 195 200 205
 Tyr Tyr Asp Leu Ala Ile Val Leu Leu Leu Thr Gly Ile Arg Ile Ser

210	215	220	
Glu Ala Ala Phe Leu Pro Ser Asp Ile Asp Phe Glu Lys Gly Ile Leu			
225	230	235	240
His Ile Asp Lys Ala Leu Gln Tyr His Cys Leu Lys Val Lys Gln Phe			
245	250	255	
His Phe Asp Thr Thr Lys Thr Leu Asn Ser Ile Arg Glu Val Ala Leu			
260	265	270	
Pro Glu Ala Ala Ser Glu Ala Ile Lys Arg Thr Ile Gln Arg Asn Lys			
275	280	285	
Glu Phe Asp Ala Tyr Met Glu Lys His Pro Cys Pro Ala Phe Thr His			
290	295	300	
Ser Glu Ser Val Phe Arg Thr Glu Tyr Gly Ser Pro Ile Thr Ser Ser			
305	310	315	320
Thr Phe Arg Gln Ile Leu Lys Arg Ile Glu Gly Lys Leu Leu Thr Asn			
325	330	335	
Cys Leu Ser Asp Tyr Gly Phe Lys Trp Val Lys His Val Thr Pro His			
340	345	350	
Ser Phe Arg His Met His Ile Ser Tyr Leu Gln Ser Asn Glu Met His			
355	360	365	
Ile Ala Val Lys Asp Ile Met Thr Arg Val Gly His Ala Asn Phe Glu			
370	375	380	
Thr Thr Met Gly Tyr Thr His Asn Ile Asn Arg Ser Gln Glu Asn Thr			
385	390	395	400
Val Lys Ala Leu Asn Gln Phe Val Glu Asn His Asn Phe His Phe Glu			
405	410	415	
Glu Leu Lys Ser Tyr Thr Cys Lys Tyr Ser Arg Ile Ile Glu Lys Phe			
420	425	430	
Ile Glu Thr Ser Asp Asn Ser Asn Lys Val Glu Leu Ser Val Asp Glu			
435	440	445	
Phe Lys Asp Leu Leu His Leu Ser Pro Arg Tyr Ser Pro Lys Asn Ile			
450	455	460	

Ile Ser Asn Leu Leu Leu Lys Ile Lys Lys Asp Ile Val Lys Tyr His

465 470 475 480

Pro Gln Phe Asp Ile Lys Ile Val Lys Ser Ser Glu Asn Gln Ile Arg

485 490 495

Gly Phe Ser Ile Ala Trp

500

<210> 21

<211> 436

<212> DNA

<213> Escherichia coli

<400> 21

gcatgcccgt tccatacaga agctgggcga acaaacgatg ctgccttcc agaaaaccga 60

ggatgcgaac cacttcatcc ggggtcagca ccaccggcaa gcgccgcgac ggccgaggtc 120

ttccgatctc ctgaagccag ggcagatccg tgcacagcac cttgccgtag aagaacagca 180

aggccgccaa tgctgacga tgcgtggaga ccgaaacctt gcgctcgttc gccagccagg 240

acagaaatgc ctgcacttcg ctgctgcca aggttgccgg gtgacgcaca ccgtggaaac 300

ggatgaaggc acgaaccagc tggacataag cctgttcggt tcgtaagctg taatgcaagt 360

agcgtatgcg ctacgcgaac tgggtccagaa ccttgaccga acgcagcggt ggtaacggcg 420

cagtggcgggt tttcat 436

<210> 22

<211> 145

<212> PRT

<213> Escherichia coli

<400> 22

Met Lys Thr Ala Thr Ala Pro Leu Pro Pro Leu Arg Ser Val Lys Val

1 5 10 15

Leu Asp Gln Leu Arg Glu Arg Ile Arg Tyr Leu His Tyr Ser Leu Arg

20 25 30

Thr Glu Gln Ala Tyr Val His Trp Val Arg Ala Phe Ile Arg Phe His

35 40 45

Gly Val Arg His Pro Ala Thr Leu Gly Ser Ser Glu Val Glu Ala Phe

50 55 60

Leu Ser Trp Leu Ala Asn Glu Arg Lys Val Ser Val Ser Thr His Arg
65 70 75 80

Gln Ala Leu Ala Ala Leu Leu Phe Phe Tyr Gly Lys Val Leu Cys Thr
85 90 95

Asp Leu Pro Trp Leu Gln Glu Ile Gly Arg Pro Arg Pro Ser Arg Arg
100 105 110

Leu Pro Val Val Leu Thr Pro Asp Glu Val Val Arg Ile Leu Gly Phe
115 120 125

Leu Glu Gly Glu His Arg Leu Phe Ala Gln Leu Leu Tyr Gly Thr Gly
130 135 140

Met

145

<210> 23

<211> 1527

<212> DNA

<213> Thermoanaerobacterium phage THSA-485A

<400> 23

atgaatcgtg tatgtattta tcttaggaag tcccgagcag acgaagaaat agaaaaagag 60
cttggacaag gagaaacact cgcaaaacat cgtaaggccc ttcttaaatt tgcaaaagag 120
aaaaatttga acatagtaaa aatcagagag gaaatagtat caggcgaaag ccttatccat 180
agacctgaaa tgttggatt actaaaagaa gtgcaacaag gcatgtacga tgctgtatta 240
tgtatggatc tacagcgttt agggcgtggc aacatgcagg aacaaggctt cattttagaa 300
gcctttaaaa agtcaaacac taaaattata acgcttcaaa aaacttatga ttggaacaat 360

gattttgacg aagaatatag cgaatttgaa gcatttatga gccgaaagga acttaaaatg 420
ataaatagaa ggctacaagg tggcagagta cgtctattc aggaaggtaa ttatttatca 480
ccattgccac cttatggtta cttaatacac gaagaaaaat ttctgcgcac tcttgtcct 540
aatctgagc aagctgatgt agttaaatg atttttgata tgtatgtcaa taaacagatg 600
gggtctagt ctatagcgaa cgaactaaac aaaatgggtt ataagacgta tactggcagg 660
aattgggctt caagctctgt aataaacata ctcaagaatc cagtttacat cggtaaaata 720
acgtggaaga agaaggatat aaagaagtct gctgacccaa ataaaagcaa agatacacgt 780

caaagaccac gctctgaatg gattgtatca gatggcaaac atgaaccaat agtgggcaaa 840

gagctctttg ccaaggctca agaaatcatt aaaaacaagt atcacatacc gtatcagatc 900
 gttaatggtc cacgtaaccc attggcaggg cttattatat gcaaaatatg tggctctaaa 960
 atggtgtata gaccctacaa agataaagaa gcgcatataa tatgtccaaa caagtgcggc 1020
 aataaaagca gcaaatttat ctatgtagaa aaaagattat tacaggcttt ggaggaatgg 1080
 atgcaaggct acgagctgga tctgcaaata gaagaagatg acagctcttt tgcagaagca 1140
 caagagaaac aaaaagaagc tcttgaaaga gaattgcacg agctgcaaaa gcaaaagaac 1200

aatttacacg atttgctcga gcgtggcata tacgatatag atacatttgt ggaaagatct 1260
 acaattgtag cacagagaat agaagaaaca cagaaaagta tagatgtgct tgtgcaaaaa 1320
 atagaagaag aaaagaataa aagagacaaa gaaaaaatac ttccggaaat tcggcatgtg 1380
 ttggatctat attggaaaac agacgacatt gcacaaaaaa atatgttgtt aaagagcgtg 1440
 cttgaaaaag cagaatatct aaaagaaaag aagcagagag aagacaactt cgaactttgg 1500
 atttatccaa agctgcctga aaaatag 1527

<210> 24
 <211> 508
 <212> PRT

<213> Thermoanaerobacterium phage THSA-485A

<400> 24

Met Asn Arg Val Cys Ile Tyr Leu Arg Lys Ser Arg Ala Asp Glu Glu

1 5 10 15

Ile Glu Lys Glu Leu Gly Gln Gly Glu Thr Leu Ala Lys His Arg Lys

20 25 30

Ala Leu Leu Lys Phe Ala Lys Glu Lys Asn Leu Asn Ile Val Lys Ile

35 40 45

Arg Glu Glu Ile Val Ser Gly Glu Ser Leu Ile His Arg Pro Glu Met

50 55 60

Leu Glu Leu Leu Lys Glu Val Glu Gln Gly Met Tyr Asp Ala Val Leu

65 70 75 80

Cys Met Asp Leu Gln Arg Leu Gly Arg Gly Asn Met Gln Glu Gln Gly

85 90 95

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

100 105 110

Gln Lys Thr Tyr Asp Leu Asn Asn Asp Phe Asp Glu Glu Tyr Ser Glu

115 120 125
Phe Glu Ala Phe Met Ser Arg Lys Glu Leu Lys Met Ile Asn Arg Arg

130 135 140
Leu Gln Gly Gly Arg Val Arg Ser Ile Gln Glu Gly Asn Tyr Leu Ser
145 150 155 160

Pro Leu Pro Pro Tyr Gly Tyr Leu Ile His Glu Glu Lys Phe Ser Arg
165 170 175

Thr Leu Val Pro Asn Pro Glu Gln Ala Asp Val Val Lys Met Ile Phe
180 185 190

Asp Met Tyr Val Asn Lys Gln Met Gly Ser Ser Ala Ile Ala Asn Glu
195 200 205

Leu Asn Lys Met Gly Tyr Lys Thr Tyr Thr Gly Arg Asn Trp Ala Ser
210 215 220

Ser Ser Val Ile Asn Ile Leu Lys Asn Pro Val Tyr Ile Gly Lys Ile
225 230 235 240

Thr Trp Lys Lys Lys Asp Ile Lys Lys Ser Ala Asp Pro Asn Lys Ser
245 250 255

Lys Asp Thr Arg Gln Arg Pro Arg Ser Glu Trp Ile Val Ser Asp Gly
260 265 270

Lys His Glu Pro Ile Val Gly Lys Glu Leu Phe Ala Lys Ala Gln Glu

275 280 285
Ile Ile Lys Asn Lys Tyr His Ile Pro Tyr Gln Ile Val Asn Gly Pro

290 295 300
Arg Asn Pro Leu Ala Gly Leu Ile Ile Cys Lys Ile Cys Gly Ser Lys

305 310 315 320
Met Val Tyr Arg Pro Tyr Lys Asp Lys Glu Ala His Ile Ile Cys Pro

325 330 335
Asn Lys Cys Gly Asn Lys Ser Ser Lys Phe Ile Tyr Val Glu Lys Arg

340 345 350
Leu Leu Gln Ala Leu Glu Glu Trp Met Gln Gly Tyr Glu Leu Asp Leu

355 360 365

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

Lys Glu Ala Leu Glu Arg Glu Leu His Glu Leu Gln Lys Gln Lys Asn
385 390 395 400

Asn Leu His Asp Leu Leu Glu Arg Gly Ile Tyr Asp Ile Asp Thr Phe
405 410 415

Val Glu Arg Ser Thr Ile Val Ala Gln Arg Ile Glu Glu Thr Gln Lys
420 425 430

Ser Ile Asp Val Leu Val Gln Lys Ile Glu Glu Glu Lys Asn Lys Arg
435 440 445

Asp Lys Glu Lys Ile Leu Pro Glu Ile Arg His Val Leu Asp Leu Tyr
450 455 460

Trp Lys Thr Asp Asp Ile Ala Gln Lys Asn Met Leu Leu Lys Ser Val
465 470 475 480

Leu Glu Lys Ala Glu Tyr Leu Lys Glu Lys Lys Gln Arg Glu Asp Asn
485 490 495

Phe Glu Leu Trp Ile Tyr Pro Lys Leu Pro Glu Lys
500 505

<210> 25
<211> 197
<212> PRT
<213> Escherichia phage D108
<400> 25

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu

65 70 75 80
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95
Ser Ser Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala
100 105 110
Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala

115					120					125						
Ala	Ala	Arg	Asn	Lys	Gly	Arg	Ile	Gly	Gly	Arg	Pro	Pro	Lys	Leu	Thr	
130					135					140						
Lys	Ala	Glu	Trp	Glu	Gln	Ala	Gly	Arg	Leu	Leu	Ala	Gln	Gly	Ile	Pro	
145					150					155					160	
Arg	Lys	Gln	Val	Ala	Leu	Ile	Tyr	Asp	Val	Ala	Leu	Ser	Thr	Leu	Tyr	
165					170					175						
Lys	Lys	His	Pro	Ala	Lys	Arg	Thr	His	Ile	Glu	Asn	Asp	Asp	Arg	Ile	
180					185					190						

Asn Gln Ile Asp Arg

195

<210> 26

<211> 345

<212> PRT

<213>	Unknown
-------	---------

<220><223> P1 bacteriophage

<400> 26

Met Val Gln Thr Ser Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu

1 5 10 15

Pro Val Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met

20 25 30

Phe Arg Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu

35 40 45

Ser Val Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys

50 55 60

Trp Phe Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu

65					70				75				80			
Gln	Ala	Arg	Gly	Leu	Ala	Val	Lys	Thr	Ile	Gln	Gln	His	Leu	Gly	Gln	
				85				90				95				
Leu	Asn	Met	Leu	His	Arg	Arg	Ser	Gly	Leu	Pro	Arg	Pro	Ser	Asp	Ser	
				100				105				110				
Asn	Ala	Val	Ser	Leu	Val	Met	Arg	Arg	Ile	Arg	Lys	Glu	Asn	Val	Asp	
				115				120				125				
Ala	Gly	Glu	Arg	Ala	Lys	Gln	Ala	Leu	Ala	Phe	Glu	Arg	Thr	Asp	Phe	
				130				135				140				
Asp	Gln	Val	Arg	Ser	Leu	Met	Glu	Asn	Ser	Asp	Arg	Cys	Gln	Asp	Ile	
145				150				155				160				
Arg	Asn	Leu	Ala	Phe	Leu	Gly	Ile	Ala	Tyr	Asn	Thr	Leu	Leu	Arg	Ile	
				165				170				175				
Ala	Glu	Ile	Ala	Arg	Ile	Arg	Val	Lys	Asp	Ile	Ser	Arg	Thr	Asp	Gly	
				180				185				190				
Gly	Arg	Met	Leu	Ile	His	Ile	Gly	Arg	Thr	Lys	Thr	Leu	Val	Ser	Thr	
				195				200				205				
Ala	Gly	Val	Glu	Lys	Ala	Leu	Ser	Leu	Gly	Val	Thr	Lys	Leu	Val	Glu	
				210				215				220				
Arg	Trp	Ile	Ser	Val	Ser	Gly	Val	Ala	Asp	Asp	Pro	Asn	Asn	Tyr	Leu	
225				230				235				240				
Phe	Cys	Arg	Val	Arg	Lys	Asn	Gly	Val	Ala	Ala	Pro	Ser	Ala	Thr	Ser	
				245				250				255				
Gln	Leu	Ser	Thr	Arg	Ala	Leu	Glu	Gly	Ile	Phe	Glu	Ala	Thr	His	Arg	
				260				265				270				
Leu	Ile	Tyr	Gly	Ala	Lys	Asp	Asp	Ser	Gly	Gln	Arg	Tyr	Leu	Ala	Trp	
				275				280				285				
Ser	Gly	His	Ser	Ala	Arg	Val	Gly	Ala	Ala	Arg	Asp	Met	Ala	Arg	Ala	
				290				295				300				
Gly	Val	Ser	Ile	Pro	Glu	Ile	Met	Gln	Ala	Gly	Gly	Trp	Thr	Asn	Val	
305				310				315				320				
Asn	Ile	Val	Met	Asn	Tyr	Ile	Arg	Asn	Leu	Asp	Ser	Glu	Thr	Gly	Ala	

325

330

335

Met Val Arg Leu Leu Glu Asp Gly Asp

340

345

<210> 27

<211> 102

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 27

ctgacccag agcaggtcgt ggcaatgcc tccaacattg gcgggaaaca ggcactcgag 60

actgtccagc gcctgcttcc cgtgctgtgc caagcgacg ga 102

<210> 28

<211> 102

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 28

ctgacccag agcaggtcgt ggccattgcc tcgaatggag ggggcaaaca ggcgttggaa 60

accgtacaac gattgctgcc ggtgctgtgc caagcgacg gc 102

<210> 29

<211> 102

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 29

ttgacccag agcaggtcgt ggcgatcgca agccacgacg gaggaagca agccttggaa 60

acagtacaga ggctgttgcc tgtgctgtgc caagcgacg gg 102

<210> 30

<211> 102

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 30

cttaccacag agcaggtcgt ggcaatcgcg agcaataacg gcggaaaaca ggctttggaa	60
acggtgcaga ggctccttcc agtgctgtgc caagcgacg gg	102
<210> 31	
<211> 204	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 31	
ctgacccag agcaggtcgt ggcaatgcc tccaacattg gcgggaaaca ggcactcgag	60
actgtccagc gcctgcttcc cgtgctttgt caggcacacg gcctcactcc ggaacaagtg	120
gtcgcaatcg cctccaacat tggcgggaaa caggcactcg agactgtcca gcgcctgctt	180
cccgtgctgt gccaagcgca cggt	204
<210>	
> 32	
<211> 204	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 32	
ctgacccag agcaggtcgt ggcaatgcc tccaacattg gcgggaaaca ggcactcgag	60
actgtccagc gcctgcttcc cgtgctttgt caggcacacg gcctcactcc ggaacaagtg	120
gtcgccattg cctcgaatgg agggggcaaa caggcgttgg aaaccgtaca acgattgctg	180
ccggtgctgt gccaagcgca cggt	204
<210> 33	
<211> 204	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 33	
ctgacccag agcaggtcgt ggcaatgcc tccaacattg gcgggaaaca ggcactcgag	60
actgtccagc gcctgcttcc cgtgctttgt caggcacacg gcctcactcc ggaacaagtg	120
gtcgcgatcg caagccacga cggaggaaaag caagccttgg aaacagtaca gaggtgttg	180
cctgtgctgt gccaagcgca cggt	204

<210> 34
 <211> 204
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 34
 ctgacccag agcaggtcgt ggcaatgcc tccaacattg gcgggaaaca ggcactcgag 60

 actgtccagc gcctgcttcc cgtgctttgt caggcacacg gcctcactcc ggaacaagtg 120
 gtcgcaatcg cgagcaataa cggcggaaaa caggctttgg aaacgggtgca gaggtcctt 180
 ccagtgtgtg gccaagcgca cggt 204

 <210> 35
 <211> 204
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 35
 ctgacccag agcaggtcgt ggccattgcc tcgaatggag ggggcaaaca ggcgttggaa 60
 accgtacaac gattgctgcc ggtgctttgt caggcacacg gcctcactcc ggaacaagtg 120
 gtcgcaatcg cctccaacat tggcgggaaa caggcactcg agactgtcca gcgcctgctt 180

 cccgtgctgt gccaagcgca cggt 204

 <210> 36
 <211> 204
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 36
 ctgacccag agcaggtcgt ggccattgcc tcgaatggag ggggcaaaca ggcgttggaa 60
 accgtacaac gattgctgcc ggtgctttgt caggcacacg gcctcactcc ggaacaagtg 120
 gtcgccattg cctcgaatgg agggggcaaa caggcgttgg aaacgtaca acgattgtg 180
 ccggtgctgt gccaagcgca cggt 204

 <210>
 > 37
 <211> 160

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 37
 caaacaggcg ttggaaccg tacaacgatt gctgccggtg cttgtcagg cacacgcct 60
 cactccggaa caagtggtcg cgatcgcaag ccacgacgga ggaaagcaag ccttggaac 120
 agtacagagg ctgttgctg tgctgtgcca agcgacggt 160
 <210> 38
 <211> 204
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 38
 ctgacccag agcaggtcgt ggccattgcc tcgaatggag ggggcaaaca ggcgttgaa 60
 accgtacaac gattgctgcc ggtgctttgt caggcacacg gcctcactcc ggaacaagtg 120
 gtcgcaatcg cgagcaataa cggcggaaaa caggctttgg aaacggtgca gaggtcctt 180
 ccagtgtgt gccaaagcga cggt 204
 <210> 39
 <211> 204
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 39
 ctgacccag agcaggtcgt ggcgatcgca agccacgacg gaggaagca agccttgaa 60
 acagtacaga ggctgttgcc tgtgctttgt caggcacacg gcctcactcc ggaacaagtg 120
 gtcgcaatcg cctccaacat tggcgggaaa caggcactcg agactgtcca ggcctgctt 180
 cccgtgctgt gccaaagcga cggt 204
 <210> 40
 <211> 161
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 40

gaaagcaagc cttggaaaca gtacagaggc tgttgctgt gctttgtcag gcacacggcc 60
tcactccgga acaagtggc gccattgcct cgaatggagg gggcaaacag gcgttgga 120
ccgtacaacg attgtgtccg gtgctgtgcc aagcgacgg t 161

<210> 41

<211> 204

<212> DNA

<213> Artificial Sequence

<220><223>

> Synthetic

<400> 41

ctgacccag agcaggtcgt ggcgatcgca agccacgacg gaggaagca agccttgga 60
acagtacaga ggctgttgcc tgtgctttgt caggcacacg gcctcactcc ggaacaagt 120
gtcgcatcg caagccacga cggaggaaag caagccttg aaacagtaca gaggtgttg 180
cctgtgctgt gccaaagcgca cggt 204

<210> 42

<211> 204

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 42

ctcaccag agcaggtcgt ggcgatcgca agccacgacg gaggaagca agccttgga 60

acagtacaga ggctgttgcc tgtgctttgt caggcacacg gcctcactcc ggaacaagt 120
gtcgcaatcg cgagcaataa cggcggaaaa caggctttg aaacggtgca gaggtcctt 180
ccagtgtgt gccaaagcgca cgga 204

<210> 43

<211> 204

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 43

ctgacccag agcaggtcgt ggcaatcgcg agcaataacg gcgaaaaca ggctttgga 60
acggtgcaga ggctccttcc agtgctttgt caggcacacg gcctcactcc ggaacaagt 120
gtcgcaatcg cctccaacat tggcgggaaa caggcactcg agactgtcca gcgcctgctt 180

cccgtgctgt gccaaagcgca cgggt 204

<210> 44

<211> 204

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 44

ctgaccccag agcaggctcgt ggcaatcgcg agcaataacg gcggaaaaca ggctttggaa 60

acggtgcaga ggctccttcc agtgctttgt caggcacacg gcctcactcc ggaacaagtg 120

gtcgccattg cctcgaatgg agggggcaaa caggcgttgg aaacagtaca acgattgctg 180

ccggtgctgt gccaaagcgca cgggt 204

<210

> 45

<211> 204

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 45

ctgaccccag agcaggctcgt ggcaatcgcg agcaataacg gcggaaaaca ggctttggaa 60

acggtgcaga ggctccttcc agtgctttgt caggcacacg gcctcactcc ggaacaagtg 120

gtcgcatcgc caagccacga cggaggaaaag caagccttgg aaacagtaca gaggctgttg 180

cctgtgctgt gccaaagcgca cgggt 204

<210> 46

<211> 176

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 46

ctgaccccag agcaggctcgt ggcaatcgcg agcaataacg gcggaaaaca ggctttggaa 60

acggtgcaga ggctccttcc agtgctttgt caggcacacg gcctcactcc ggaacaagtg 120

gtcgcaatcg cgagcaataa cggcggaaaa caggctttgg aaacggtgca gaggct 176

<210> 47

<211> 219

<212> DNA

<213> Ovine lentivirus

<400> 47

catagtaaat ggcataaga tgctatgtca ttgcagttag attttgggat accgaaaggt 60

gcggcagaag ataatgtaca acaatgtgaa gtatgtcagg aaaataaaat gcctagcacc 120

atcagaggaa gtaacaaaag agggatagat cattggcagg tggattatac tcattataaa 180

gacaaaataa tattggtatg ggtagaaaca aattcggga 219

<210> 48

<211> 73

<212> PRT

<213> Ovine lentivirus

<400> 48

His Ser Lys Trp His Gln Asp Ala Met Ser Leu Gln Leu Asp Phe Gly

1 5 10 15

Ile Pro Lys Gly Ala Ala Glu Asp Ile Val Gln Gln Cys Glu Val Cys

20 25 30

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

35 40 45

Ile Asp His Trp Gln Val Asp Tyr Thr His Tyr Lys Asp Lys Ile Ile

50 55 60

Leu Val Trp Val Glu Thr Asn Ser Gly

65 70

<210> 49

<211> 243

<212> DNA

<213> Staphylococcus aureus subsp. aureus SK1585

<400> 49

ttatagatag gttagtga aaatacattt ttcgtctaga ttaaccgtgc ctcttagatt 60

attaatattt tcgttttagat gtttttcaga aacttttagca acttcataat cgttcattga 120

aagtgtttgg ttttttattg tataattaag taattcataa tctttgtata cttcttttac 180

tttatctata tcaacatttt caagaacaag tttttttatg ttattataat taaagttttc 240

cat 243

<210> 50

<211> 80
 <212> PRT
 <213> Staphylococcus aureus subsp. aureus SK1585
 <400> 50
 Met Glu Asn Phe Asn Tyr Asn Asn Ile Lys Lys Leu Val Leu Glu Asn
 1 5 10 15
 Val Asp Ile Asp Lys Val Lys Glu Val Tyr Lys Asp Tyr Glu Leu Leu
 20 25 30

Asn Tyr Thr Ile Lys Asn Gln Thr Leu Tyr Met Asn Asp Tyr Glu Val
 35 40 45
 Ala Lys Val Ser Glu Lys His Leu Asn Glu Asn Ile Asn Asn Leu Arg
 50 55 60
 Gly Thr Val Asn Leu Asp Glu Lys Cys Ile Leu Ser Leu Thr Tyr Leu
 65 70 75 80

<210> 51
 <211> 48
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 51

agcggcagcg aaaccccggg caccagcgaa agcgcgaccc cggaaagc 48

<210> 52
 <211> 1368
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 52

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

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

50

55

60

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

65

70

75

80

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

85

90

95

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

100

105

110

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

115

120

125

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

130

135

140

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

145

150

155

160

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

165

170

175

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

180

185

190

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

195

200

205

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

210

215

220

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

225

230

235

240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245

250

255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260

265

270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275

280

285

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

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

 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380
 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415

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

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

545 550 555 560
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575
 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590
 Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605
 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

 625 630 635 640
 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 645 650 655
 Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670
 Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 675 680 685
 Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 690 695 700

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

 770 775 780
 Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

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

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
1025 1030 1035 1040

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
1045 1050 1055

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

1060 1065 1070
Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser

1075 1080 1085
Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly

1090 1095 1100
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile

1105 1110 1115 1120
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser

1125 1130 1135

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

1140 1145 1150

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

1155 1160 1165

Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1170 1175 1180

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys

1185 1190 1195 1200

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

1205 1210 1215

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

1220 1225 1230

Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

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

1265 1270 1275 1280

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

1285 1290 1295

His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

1300				1305				1310							
Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro	Ala	Ala	Phe	Lys	Tyr	Phe	Asp
1315				1320				1325							
Thr	Thr	Ile	Asp	Arg	Lys	Arg	Tyr	Thr	Ser	Thr	Lys	Glu	Val	Leu	Asp
1330				1335				1340							
Ala	Thr	Leu	Ile	His	Gln	Ser	Ile	Thr	Gly	Leu	Tyr	Glu	Thr	Arg	Ile
1345				1350				1355				1360			
Asp	Leu	Ser	Gln	Leu	Gly	Gly	Asp								
1365															
<210>	53														
<211>	117														
<212>	DNA														
<213>	Artificial Sequence														
<220><223>	Synthetic														
<400>	53														
atggactaca aagaccatga cggtgattat aaagatcatg acatcgatta caaggatgac 60															
gatgacaaga tggcccccaa gaagaagagg aaggtgggca ttcaccgcgg ggtacct 117															
<210>	54														
<211>	9														
<212>	DNA														
<213>	Artificial Sequence														
<220><223>	Synthetic														
<400>	54														
gggggaagt 9															
<210>	55														
<211>	870														
<212>	DNA														
<213>	Artificial Sequence														
<220><223>	Synthetic														
<400>	55														
atgttcctgg acggtatcga caaagctcag gacgagcacg aaaagtacca ttctaactgg 60															
cgcgccatgg cctctgactt caatctcccg ccggttggtg ccaaggagat cgtggcttct 120															
tgcgacaagt gccaatgaa gggtgaggct atgcatggtc aggtcgattg ctctcccggt 180															

atctggcagc tggactgcac tcacctcgag ggtaaggatga ttctcgttgc tgtgcacgtg	240
gcttccggct acatcgagcg tgaggtcatc ccggctgaga ccggtcaaga gactgcttac	300
ttctctgctca agctggccgg ccgttggcca gtaagacta ttcacactga taacggttct	360
aactttactt ccgcaactgt gaaagctgca tgctgggtggg ccggcattaa acaagagttc	420
ggaattccgt ataaccgca gtctcagggc gttgtcgagt ctatgaacaa ggagctcaaa	480
aagatcattg gtcaagtccg tgaccaagct gagcacctta agaccgctgt gcagatggct	540
gtttttattc ataacttcaa gcgtaagggt ggtatcgggtg gttatagcgc tggtagcgt	600
atcgtagaca tcctcgctac tgatatccag acaaaggagc tgcagaagca gatcactaag	660
atccagaact tccgtgtgta ctatcgggac tctaggaacc cgctctggaa gggtcctgct	720
aaactgctgt ggaagggaga ggggtgctgtt gttatccagg acaactctga tatcaagggtg	780
gttccgctgc gtaaggctaa aattatccgc gactacggca agcaaatggc tggagacgac	840
tgcgttgcta gccgtcaaga cgaagactaa	870
<210> 56	
<211> 4107	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 56	
atggataaaa agtattctat tggtttagct atcggcacta attccgttgg atgggctgtc	60
ataaccgatg aatacaaatg accctcaaaag aaatttaagg tgttggggaa cacagaccgt	120
cattcgatta aaaagaatct tatcggtgcc ctctatttcg atagtggcga aacggcagag	180
gcgactcgcc tgaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt	240
tacttacaag aaatttttag caatgagatg gccaaagtig acgattcttt ctttcaccgt	300
ttggaagagt ccttccttgt cgaagaggac aagaaacatg aacggcacc cacttttga	360
aacatagtag atgaggtggc atatcatgaa aagtaccaa cgatttatca cctcagaaaa	420
aagctagtgt actcaactga taaagcggac ctgagggtta tctacttggc tcttgccat	480
atgataaagt tccgtgggca ctttctcatt gaggtgatc taaatccga caactcggat	540
gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaacct	600
ataaatgcaa gtggcgtgga tgcgaaggct attcttagcg cccgcctctc taaatcccga	660
cggctagaaa acctgatcg acaattaccc ggagagaaga aaaatgggtt gttcggtaac	720
cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa	780

gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca	840
caaattggag atcagtatgc ggacttattt ttggctgccca aaaaccttag cgatgcaatc	900
ctcctatctg acatactgag agttaatact gagattacca aggcgccgtt atccgcttca	960
atgatcaaaa ggiacgatga acatcaccaa gacttgacac ttctcaaggc cctagtcctg	1020
cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca	1080
ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta	1140
gagaagatgg atgggacgga agagttgctt gtaaaactca atcgcaaga tctactgcga	1200
aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat	1260
gctatactta gaaggcagga ggatttttat cgttctca aagacaatcg tgaagatt	1320
gagaaatcc taaccttctg cataccttac tatgtgggac ccctggcccg agggactct	1380
cggttcgcat ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa	1440
gttgtcgata aagggtcgctc agctcaatcg ttcacgaga ggatgacaa ctttgacaag	1500
aatttaccga acgaaaaagt attgcctaag cacagtttac ttacgagta ttccacagt	1560
tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta	1620
agcggagAAC agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca	1680
gttaagcaat tgaagagga ctactttaag aaaattgaat gcttcgattc tgcgagatc	1740
tccgggtag aagatcgatt taatgcgtca ctgggtacgt atcatgacct cctaaagata	1800
attaaagata aggacttctt ggataacgaa gagaatgaag atatcttaga agatatagt	1860
ttgactctta cctcttttga agatcgggaa atgattgagg aaagactaaa aacatacgt	1920
cacctgttcg acgataaggt tatgaacag ttaaagaggc gtcgctatc gggctgggga	1980
cgattgtcgc gaaacttat caacgggata agagacaagc aaagtggtaa aactattctc	2040
gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac	2100
tctttaacct tcaaagagga tatacaaaag gcacaggttt cggacaagg ggactcattg	2160
cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaaggcat actccagaca	2220
gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaacgga aacattgta	2280
atcgagatgg cagcgaaaa tcaaacgact cagaaggggc aaaaaacag tcgagagcgg	2340
atgaagagaa tagaagaggg tattaagaa ctgggcagcc agatcttaa ggagatcct	2400
gtggaaaata ccaattgca gaacgagaaa cttacctct attacctaca aaatggaagg	2460
gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatgcc	2520
attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgt tacacgtctg	2580
gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag	2640

aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagt t cgataactta 2700

actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccgatttat taaacgtcag 2760

ctcgtggaaa cccgccaaat cacaaagcat gttgcacaga tactagattc ccgaatgaat 2820

acgaaatagc acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca 2880

aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940

taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000

taccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060

atgatcgca aaagcgaaca ggagataggc aaggctacag ccaaatactt cttttattct 3120

aacattatga atttctttaa gacggaaatc actctggcaa acggagagat acgcaaacga 3180

cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc 3240

gcgacggtga gaaaagtttt gtccatgcc caagtcaaca tagtaagaa aactgaggtg 3300

cagaccggag ggttttcaaa ggaatcgatt ctccaaaaa ggaatagtga taagctcatc 3360

gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc 3420

tattctgtcc tagtagtggc aaaagttgag aagggaatat ccaagaaact gaagtcagtc 3480

aaagaattat tggggataac gattatggag cgctcgtctt ttgaaaagaa ccccatcgac 3540

ttccttgagg cgaaggtta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600

tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660

caaaagggga acgaactgc actaccgtct aaatacgtga atttctgta tttagcgtcc 3720

cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact tttgttgag 3780

cagcacaac attatctga cgaaatcata gagcaaat t cggaattcag taagagagtc 3840

atcctagctg atgccaatct ggacaaagta ttaagcgcac acaacaagca cagggataaa 3900

cccatagctg agcaggcgga aaatattatc catttgttta ctcttaccac cctcggcgt 3960

ccagccgcat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020

gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080

gatttgcac agcttggggg tgactaa 4107

<210> 57

<211> 5148

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 57

atggactaca aagaccatga cggtgattat aaagatcatg acatcgatta caaggatgac	60
gatgacaaga tggccccaa gaagaagagg aaggtgggca ttcaccgcgg ggtacctggg	120
ggaagtatgt tcctggacgg tatcgacaaa gctcaggacg agcacgaaaa gtaccattct	180
aactggcgcg ccatggcctc tgacttcaat ctcccgcgg ttgttgcaa ggagatcgtg	240
gcttcttgcg acaagtgcc attgaagggt gaggtatgc atggtcaggt cgattgctct	300
cccggatatc ggagctgga ctgcactcac ctcgagggtta aggtgattct cgttgctgtg	360
cacgtggctt ccggctacat cgaggctgag gtcacccgg ctgagaccgg tcaagagact	420
gcttacttcc tgctcaagct ggccggccgt tggccagtta agactattca cactgataac	480
ggttctaact ttacttccgc aactgtgaaa gctgcatgct ggtgggcccgg cattaacaa	540
gagttcggaa ttccgtataa cccgcagtct cagggcgttg tcgagtctat gaacaaggag	600
ctcaaaaaga tcattggta agtcctgac caagctgagc accttaagac cgctgtgcag	660
atggctgttt ttattcataa cttcaagcgt aagggtggta tcggtggta tagcgctggt	720
gagcgtatcg tagacatcat cgctactgat atccagacaa aggagctgca gaagcagatc	780
actaagatcc agaacttccg tgtgtactat cgggactcta ggaaccgcct ctggaagggt	840
cctgctaacc tgctgtgaa gggagagggt gctgttgta tccaggacaa ctctgatatc	900
aaggtggttc cgcgtcgtaa ggctaaaatt atccgcgact acggcaagca aatggctgga	960
gacgactcgc ttgctagccg tcaagacgaa gacagcggca gcgaaacccc gggcaccagc	1020
gaaagcgca ccccgaaag catggataaa aagtattcta ttggtttagc tatcggcact	1080
aattccgttg gatgggtgt cataaccgat gaatacaaag taccttcaa gaaatttaag	1140
gtgttggga acacagaccg tcattcgatt aaaaagaatc ttatcggtgc cctcctatc	1200
gatagtggcg aaacggcaga ggcgactcgc ctgaaacgaa ccgctcggag aaggtatata	1260
cgctcgcaaga accgaatatg ttacttacia gaaattttta gcaatgagat ggccaaagtt	1320
gacgattctt tctttaccg tttggaagag tccttccttg tcgaagagga caagaacat	1380
gaacggcacc ccatctttgg aaacatagta gatgaggtgg catatcatga aaagtacca	1440
acgatttata acctcagaaa aaagctagt gactcaactg ataaagcgga cctgaggtta	1500
atctacttgg ctcttgccca tatgataaag ttccgtgggc actttctcat tgagggtgat	1560
ctaaatccgg acaactcgga tgctgacaaa ctgttcatcc agttagtaca aacctataat	1620
cagttgtttg aagagaaccc tataaatgca agtggcgtgg atgcgaaggc tattcttagc	1680
gccccctct ctaaatcccg acggctagaa aacctgatcg cacaattacc cggagagaag	1740
aaaaatgggt tgttcggtaa ccttatagcg ctctcactag gcctgacacc aaattttaag	1800

tcgaacttcg acttagctga agatgccaaa ttgcagctta gtaaggacac gtacgatgac	1860
gatctcgaca atctactggc acaaattgga gatcagtatg cggacttatt tttggctgcc	1920
aaaaacctta gcgatgcaat cctcctatct gacatactga gagttaatac tgagattacc	1980
aaggcgccgt tatccgcttc aatgatcaaa aggtacgatg aacatcacca agacttgaca	2040
cttctcaagg ccctagtccg tcagcaactg cctgagaaat ataaggaaat attctttgat	2100
cagtcgaaaa acgggtacgc aggttatatt gacggcggag cgagtcaaga ggaattctac	2160
aagtttatca aacctatatt agagaagatg gatgggacgg aagagttgct tgtaaaactc	2220
aatcgcgaaag atctactgcg aaagcagcgg actttcgaca acggtagcat tccacatcaa	2280
atccacttag gcgaattgca tgctatactt agaaggcagg aggattttta tccgttcctc	2340
aaagacaatc gtgaaaagat tgagaaaatc ctaacctttc gcataacctta ctatgtggga	2400
cccctggccc gagggaaactc tcggttcgca tggatgacaa gaaagtccga agaaacgatt	2460
actccatgga attttgagga agttgtcgat aaagggtcgt cagctcaatc gttcatcgag	2520
aggatgacca actttgacaa gaatttaccg aacgaaaaag tattgcctaa gcacagtta	2580
ctttacgagt atttcacagt gtacaatgaa ctcacgaaag ttaagtatgt cactgagggc	2640
atgcgtaaac ccgcctttct aagcggagaa cagaagaaag caatagtaga tctgtttatc	2700
aagaccaacc gcaaagtgac agttaagcaa ttgaaagagg actactttaa gaaaattgaa	2760
tgcttcgatt ctgtcgagat ctccggggta gaagatcgat ttaatgcgtc acttggtacg	2820
tatcatgacc tcctaaagat aattaaagat aaggacttcc tggataacga agagaatgaa	2880
gatatcttag aagatatagt gttgactctt accctctttg aagatcgga aatgattgag	2940
gaaagactaa aaacatacgc tcacctgttc gacgataagg ttatgaaaca gttaaagagg	3000
cgtcgtata cgggctgggg acgattgtcg cggaaactta tcaacgggat aagagacaag	3060
caaagtggta aaactattct cgattttcta aagagcgacg gcttcgcaa taggaacttt	3120
atgcagctga tccatgatga ctctttaacc ttcaaagagg atatacaaaa ggcacaggtt	3180
tccggacaag gggactcatt gcacgaacat attgcgaatc ttgctggttc gccagccatc	3240
aaaaagggca tactccagac agtcaaagta gtggatgagc tagttaaggt catgggacgt	3300
cacaaaccgg aaaacattgt aatcgagatg gcacgcgaaa atcaaacgac tcagaagggg	3360
caaaaaaaca gtcgagagcg gatgaagaga atagaagagg gtattaaaga actgggcagc	3420
cagatcttaa aggagcatcc tgtggaaaat acccaattgc agaacgagaa actttacctc	3480
tattacctac aaaaatggaag ggacatgtat gttgatcagg aactggacat aaaccgttta	3540

tctgattacg acgtcgatgc cattgtaccc caatcctttt tgaaggacga ttcaatcgac 3600
aataaagtgc ttacacgctc ggataagaac cgagggaaaa gtgacaatgt tccaagcgag 3660
gaagtcgtaa agaaaatgaa gaactattgg cggcagctcc taaatgcgaa actgataacg 3720
caaagaaagt tcgataactt aactaaagct gagaggggtg gcttgtctga acttgacaag 3780
gccggattta ttaaagctca gctcgtggaa acccgccaaa tcacaaagca tgttgcacag 3840
atactagatt cccgaatgaa tacgaatac gacgagaacg ataagctgat tcgggaagtc 3900
aaagtaatca ctttaaagtc aaaattggig tcggacttca gaaaggattt tcaattctat 3960

aaagttaggg agataaataa ctaccacat gcgcacgacg cttatcttaa tgccgtcgta 4020
gggaccgcac tcattaagaa ataccgaag ctagaaagtg agtttgtgta tgggtattac 4080
aaagtttatg acgtccgtaa gatgatcgcg aaaagcgaac aggagatagg caaggctaca 4140
gccaaatact tcttttattc taacattatg aatttcttta agacggaaat cactctggca 4200
aacggagaga tacgcaaagc acctttaatt gaaaccaatg gggagacagg tgaaatcgta 4260
tgggataagg gccgggactt cgcgacggtg agaaaagttt tgtccatgcc ccaagtcaac 4320
atagtaaaga aaactgaggt gcagaccgga gggttttcaa aggaatcgat tcttccaaaa 4380

aggaatagtg ataagctcat cgctcgtaaa aaggactggg acccgaaaaa gtacggtggc 4440
ttcgatagcc ctacagttgc ctattctgtc ctagtagtgg caaaagtga gaaggaaaaa 4500
tccaagaaac tgaagtcagt caaagaatta ttggggataa cgattatgga gcgctcgtct 4560
tttgaagaaga accccatcga cttccttgag gcgaaagggtt acaaggaagt aaaaaaggat 4620
ctcataatta aactacaaa gtatagtctg tttagattag aaaatggccg aaacggatg 4680
ttggctagcg ccggagagct tcaaaagggg aacgaactcg cactaccgtc taaatacgtg 4740
aatttcctgt atttagcgtc ccattacgag aagttgaaag gttcacctga agataacgaa 4800

cagaagcaac tttttgttga gcagcacaaa cattatctcg acgaaatcat agagcaaatt 4860
tcggaattca gtaagagagt catcctagct gatgccaatc tggacaaagt attaagcgca 4920
tacaacaagc acagggataa acccatacgt gagcagcgcg aaaatattat ccatttgttt 4980
actcttacca acctcggcgc tccagccgca ttcaagtatt ttgacacaac gatagatcgc 5040
aaacgataca ctctaccaa ggaggtgcta gacgcgacac tgattcacca atccatcacg 5100
ggatttatg aaactcgat agatttgtca cagcttgggg gtgactaa 5148

<210> 58

<211> 1715

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 58

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp

1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val

20 25 30

Gly Ile His Arg Gly Val Pro Gly Gly Ser Met Phe Leu Asp Gly Ile

35 40 45

Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala

50 55 60

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

65 70 75 80

Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln

85 90 95

Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu

100 105 110

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

115 120 125

Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu

130 135 140

Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn

145 150 155 160

Gly Ser Asn Phe Thr Ser Ala Thr Val Lys Ala Ala Cys Trp Trp Ala

165 170 175

Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly

180 185 190

Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val

195 200 205

Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe

210 215 220

Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly

225 230 235 240

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

485 490 495
 Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg
 500 505 510
 Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val
 515 520 525
 Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu
 530 535 540
 Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser
 545 550 555 560
 Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu

 565 570 575
 Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser
 580 585 590
 Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp
 595 600 605
 Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn
 610 615 620
 Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala
 625 630 635 640

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

 705 710 715 720
 Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu
 725 730 735

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

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

 850 855 860
 Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly
 865 870 875 880
 Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val
 885 890 895
 Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys
 900 905 910
 Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser
 915 920 925

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

Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035 1040

Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln

1045 1050 1055

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

1060 1065 1070

Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val

1075 1080 1085

Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu

1090 1095 1100

Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly

1105 1110 1115 1120

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

1125 1130 1135

Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln

1140 1145 1150

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

1155 1160 1165

Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp

1170 1175 1180

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

1185 1190 1195 1200

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

1205 1210 1215

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln

1220 1225 1230

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

1235 1240 1245
 Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile
 1250 1255 1260
 Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln
 1265 1270 1275 1280
 Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu

 1285 1290 1295
 Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp
 1300 1305 1310
 Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr
 1315 1320 1325
 His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu
 1330 1335 1340
 Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr
 1345 1350 1355 1360

 Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile
 1365 1370 1375
 Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe
 1380 1385 1390
 Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro
 1395 1400 1405
 Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly
 1410 1415 1420
 Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn

 1425 1430 1435 1440
 Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
 1445 1450 1455
 Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
 1460 1465 1470
 Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr
 1475 1480 1485
 Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu

1490 1495 1500

 Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
 1505 1510 1515 1520
 Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
 1525 1530 1535
 Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu
 1540 1545 1550
 Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln
 1555 1560 1565
 Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr

 1570 1575 1580
 Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu
 1585 1590 1595 1600
 Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile
 1605 1610 1615
 Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala
 1620 1625 1630
 Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro
 1635 1640 1645

 Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
 1650 1655 1660
 Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg
 1665 1670 1675 1680
 Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
 1685 1690 1695
 Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
 1700 1705 1710
 Gly Gly Asp

1715
 <210> 59
 <211> 29

<212>	DNA	
<213>	Human immunodeficiency virus 1	
<400>	59	
	actggaaggg ctaattcact cccaaagaa	29
<210>	60	
<211>	35	
<212>	DNA	
<213>	Human immunodeficiency virus 1	
<400>	60	
	gaccctttta gtcagtgtgg aaaatctcta gcagt	35
<210>	61	
<211>	16	
<212>	PRT	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	61	
	Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser	
	1 5 10 15	
<210>	62	
<211>	1098	
<212>	DNA	
<213>	Mouse mammary tumor virus	
<400>	62	
	atgacaggaa agtggccttg tatttactcc actaactgca gagatgtgtt gcatgggacg	60
	gggggcactg caccagccct cgtgctgaat tcggcacgag gaaatgccta tgcagattct	120
	ttaacaagaa ttctgaccgc tttagagtca gctcaagaaa gccacgcact gcaccatcaa	180
	aatgccgcgg cgcttaggtt tcagtttcac atcactcgtg aacaagcacg agaaatagta	240
	aaattatgtc caaattgccc cgactgggga catgcaccac aactaggagt aaaccctagg	300
	ggccttaagc ccggggttct atggcaaatg gatgttactc atgtctcaga atttggaana	360
	ttaaagtatg tacatgtgac agtggatact tactctcatt ttactttcgc taccgcccgg	420
	acgggcgaag cagccaaaga tgtgttacaa cacttggctc aaagctttgc atacatgggc	480
	attcctcaaa aaataaaaac agataatgcc cctgcctatg tgtctcgttc aatacaagaa	540
	tttctggcca gatggaaaat atctcagtc acggggatcc cttacaatcc ccaaggacag	600

gccattgttg aacgaacgca ccaaaatata aaggcacaga ttaataaact tcaaaaggct 660
 ggaaaatact atacaccca ccattctattg gcacatgctc tttttgtgct gaatcatgta 720
 aatatggaca atcaaggcca tacagcggcc gaaagacatt ggggtccaat ctcagccgat 780
 ccaaaaccta tggatcatgtg gaaagacctt ctcacagggt cctggaaagg acccgatgtc 840
 ctaataacag ccggacgagg ctatgcttgt gtttttcac aggatgccga atcaccaatc 900
 tgggtcccg accggttcat ccgacctttt actgagcgga aagaagcaac gcccacacct 960
 ggcaactgcgg agaaaacgcc gccgcgagat gagaaagatc aacaggaaag tccggaggat 1020
 gaatcttgcc cccatcaaag agaagacggc ttggcaacat ctgcaggcgt taatctccga 1080
 agcggaggag gttcttaa 1098

<210> 63
 <211> 365
 <212> PRT
 <213> Mouse mammary tumor virus
 <400> 63

Met Thr Gly Lys Trp Pro Cys Ile Tyr Ser Thr Asn Cys Arg Asp Val
 1 5 10 15
 Leu His Gly Thr Gly Gly Thr Ala Pro Ala Leu Val Leu Asn Ser Ala
 20 25 30
 Arg Gly Asn Ala Tyr Ala Asp Ser Leu Thr Arg Ile Leu Thr Ala Leu
 35 40 45
 Glu Ser Ala Gln Glu Ser His Ala Leu His His Gln Asn Ala Ala Ala
 50 55 60
 Leu Arg Phe Gln Phe His Ile Thr Arg Glu Gln Ala Arg Glu Ile Val
 65 70 75 80
 Lys Leu Cys Pro Asn Cys Pro Asp Trp Gly His Ala Pro Gln Leu Gly
 85 90 95
 Val Asn Pro Arg Gly Leu Lys Pro Gly Val Leu Trp Gln Met Asp Val
 100 105 110
 Thr His Val Ser Glu Phe Gly Lys Leu Lys Tyr Val His Val Thr Val
 115 120 125
 Asp Thr Tyr Ser His Phe Thr Phe Ala Thr Ala Arg Thr Gly Glu Ala

130 135 140
Ala Lys Asp Val Leu Gln His Leu Ala Gln Ser Phe Ala Tyr Met Gly
145 150 155 160
Ile Pro Gln Lys Ile Lys Thr Asp Asn Ala Pro Ala Tyr Val Ser Arg
165 170 175
Ser Ile Gln Glu Phe Leu Ala Arg Trp Lys Ile Ser His Val Thr Gly
180 185 190
Ile Pro Tyr Asn Pro Gln Gly Gln Ala Ile Val Glu Arg Thr His Gln

195 200 205
Asn Ile Lys Ala Gln Ile Asn Lys Leu Gln Lys Ala Gly Lys Tyr Tyr
210 215 220
Thr Pro His His Leu Leu Ala His Ala Leu Phe Val Leu Asn His Val
225 230 235 240
Asn Met Asp Asn Gln Gly His Thr Ala Ala Glu Arg His Trp Gly Pro
245 250 255
Ile Ser Ala Asp Pro Lys Pro Met Val Met Trp Lys Asp Leu Leu Thr
260 265 270

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

340 345 350
Thr Ser Ala Gly Val Asn Leu Arg Ser Gly Gly Gly Ser
355 360 365

<210> 64
<211> 3735
<212> DNA

<213> Youngiibacter fragilis 232.1

<400> 64

ttgaaagata acgataaaag gatgtgggtt cagactttat ggaatcccat caatgaaaga	60
cataaaagt cactggatag cccagaacca gggattaaag tagcggccta ctgcagagta	120
agcatgaaag aggaggaaca actccggtca ttggaaaacc aggtgcatca ctatactcat	180
tttatcaaaa gtaagccgaa ttggagattt gtaggggttt attacgatga tggcataagt	240
gcagccatgg caagtgggag aagagggttc cagcggatta tccgtcatgc tgaagaaggt	300
aaggttgatc tgattctaac aaagaatatt tcacggtttt ccagaaattc caaggagtta	360
ctggatataa tcaatcaact gaaagctatc ggtgtgggca tctattttga gaaagagaat	420
attgatactt caagagagta caataaattc ctcttaagca cttatgctgc gctggcacag	480
gaagagatag aaactatttc aaactctacg atgtgggtt atgagaaaag gtttctaag	540
ggtatcccaa agttcaaccg cttatatgga tacaagtc tccatgcagg ggatgattcc	600
caattgattg ttcttgaaga tgaagcaaaa atcgtaagaa tgatgtatga acagtacctt	660
caagggaaga cgttactga tattgcaagg gcgctaacag aagctggagt gaaaacagcc	720
aaagggaagg atgtctggat aggcggcatg ataaagcata ttttatccaa cgtcacctac	780
accggttaaca agcttacacg agaactgaaa agagatttat ttacgaaca agttaatagc	840
ggtgaacggg atcaggtttt tataggaaac actcacgaac cgatcatcag caatgatatt	900
ttcaatcttg ttcaaaagaa gcttgaggcc aatacgaagg aaagaaagcc cagtgagaag	960
cgagagaaga accacatgtc tggtcggcta ctttgcggaa gatgtggata cagttttacc	1020
ataattcaca atagagcttc tcatcacttt aagtgtagcc ctaaaatcat gggggtctgt	1080
gattctgaac tttatcggga tgcggatatt cgagaaatga tgatgagggc aatgtatata	1140
aaatatgact tcaccgatga agacatagta ctaaaactgc tgaaggaaact ccaggtcac	1200
aatcaaaatg atcacttga gtttcatagg ctaaagtta tcaactgaaat tgaaatcgta	1260
aaaaggcagc aggccatttc agatagatat tcagctatta gcatagaaaa aatggaagaa	1320
gaataccgca cttttgaaag caagattgcg aaaattgagg atgacaggta catcagaatc	1380
gatgcagtgg agtgggttaa gaaaaacaag acgtggatt cttttatcgc tcaggtcacc	1440
actaaaatat tgcgagcttg ggtttccgag atgactgttt atacacgaga tgacttttta	1500
gtgcagtgga ttgacggaac tcaactgag ataggaagct gcgagcatca tcttgtgaag	1560
gatagaaata gtaagagtta cgagtccggt gaagaaacga gcaggagggc caaatttgaa	1620
gtcaaccaca ttagtgaac caccgaagga caaggagaac ttgatctctt aagcaagagt	1680
gcaagttcaa acaatgaaga tagtaatcaa ccagaaaata attctacggg aaaggaggag	1740

cttgaattga acttaaacag taatgcagaa attatcaaaa ttgagcccgg gcaaaggac	1800
tatattatga agaatttga caagagcctg agtgcaaata tgatgatgca aaatgcttca	1860
gtacacacgg caagtattaa caaacctaga cttagactg ctgcttactg cagaatctca	1920
acagattcag aagaacaaaa ggtaagcttg aaaaccaag tagcctatta cacttatctg	1980
attctaaagg atccccaata tgaatatgca ggcattctatg ccgatgaagg tatatcaggg	2040
cgttctatga aaaacgctac agaatttctc aaactactcg aagaatgtaa agccgggaat	2100
gtggacttga ttttaacaa gtcaatctca cggtttagca gaaacgcatt agattgcttg	2160
gaacagatca ggatgctgaa gtcgctgcc agtccagttt atgtgtatgt tgagaaagag	2220
aatattcata caaaagatga gaagagtga ctgatgattt ctatttttgg aagtatcgct	2280
caggaagaga gcgtaaacat gggagaagcc atggcttggg gaaaacggag atatgctgag	2340
agagggatag taaaccaag tgttcacct tatggatata gaacggtcag aaaagtgaa	2400
tgggaggtgg ttgaagaaga agctacgatc attagaaga tttatcggat gctcctaagt	2460
ggaaagagta ttcatgaaat cacaaggag ctctccatgg agaagataaa gggctctggc	2520
ggcaacgagc agtggcatct tcaaacctt agaatatct tgagaaatga aatctatagg	2580
ggtaactacc tttatcaaaa ggcttatatc aaggacacga tcgagaaga ggtggtaatg	2640
aatcgaggag aactgccaca gtatctcata gagaatcatc ataaagccat tgttgacaat	2700
gagacctggg aaaaggtcca gaaggtacta gaagccagaa gggaaaaata tgagaataaa	2760
aagtccataa cttatctga agacaaaatg aaaaacgctt ctcttgaaga tattttacc	2820
tgtggagaat gtggaagtaa aataggccat agaaggagca tccagagctc taatgagatt	2880
cattctctga tcigcacaaa agccgctaag tctttcttgg tggactcgtg taagtccaca	2940
agcgtatatc agaagcacct ggagctgcat tttatgaaga ctcttctcga tattaanaag	3000
catcgttctt tcaaagatga ggtgctcacc tatattcgaa ccaagaagt agatgaaaag	3060
gaagagtga gaatcaaagt catagagaaa cgaatcaaag atcttaacag agagctttat	3120
aatgcggtag accaggagct caataaaaaa ggtcaggact ccaggaaagt tgatgagctc	3180
acagagaaaa ttgtggatct tcaagaggaa ttaaaggtgt ttagggaccg aaaggcaaag	3240
gttgaggatc ttaaagctga gcttgaatgg ttcttaaga agctggaaac cattgatgac	3300
gctcgagtaa aaagaaatga aggaataggc cacggtgaag agatctactt cagagaagat	3360
atttttgaaa gaatagtaag gagtgcacag ctttatagcg atggaaggat cgtctacgaa	3420
ctaagcctcg ggatccagtg gttcattgac tttaaataga gcgcatttca gaagcttctt	3480
ataaagtga aggataaaca aagggcagaa gaaaaagagg cttttcttga ggggccggaa	3540

gttaaagagc tgctggaatt ttgtaaggaa ccgaagagct actctgattt acatgccttc 3600

atgtgtgaga gaaaagaggt gtcttatagc tatttcagga aattggtgat aagacctttg 3660

atgaagaaag gaaagctgaa gttcaccata ccagaagatg ttatgaatag gcatcagaga 3720

tacacatcaa tctaa 3735

<210> 65

<211> 1244

<212> PRT

<213> Youngiibacter fragilis 232.1

<400> 65

Met Lys Asp Asn Asp Lys Arg Met Trp Val Gln Thr Leu Trp Asn Pro

1 5 10 15

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

20 25 30

Lys Val Ala Ala Tyr Cys Arg Val Ser Met Lys Glu Glu Glu Gln Leu

35 40 45

Arg Ser Leu Glu Asn Gln Val His His Tyr Thr His Phe Ile Lys Ser

50 55 60

Lys Pro Asn Trp Arg Phe Val Gly Val Tyr Tyr Asp Asp Gly Ile Ser

65 70 75 80

Ala Ala Met Ala Ser Gly Arg Arg Gly Phe Gln Arg Ile Ile Arg His

85 90 95

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

100 105 110

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

115 120 125

Ala Ile Gly Val Gly Ile Tyr Phe Glu Lys Glu Asn Ile Asp Thr Ser

130 135 140

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

145 150 155 160

Glu Glu Ile Glu Thr Ile Ser Asn Ser Thr Met Trp Gly Tyr Glu Lys

165 170 175

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

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

 305 310 315 320
 Arg Glu Lys Asn His Met Ser Gly Arg Leu Leu Cys Gly Arg Cys Gly
 325 330 335
 Tyr Ser Phe Thr Ile Ile His Asn Arg Ala Ser His His Phe Lys Cys
 340 345 350
 Ser Pro Lys Ile Met Gly Val Cys Asp Ser Glu Leu Tyr Arg Asp Ala
 355 360 365
 Asp Ile Arg Glu Met Met Met Arg Ala Met Tyr Ile Lys Tyr Asp Phe
 370 375 380

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

Ile Ser Ile Glu Lys Met Glu Glu Glu Tyr Arg Thr Phe Glu Ser Lys
435 440 445
Ile Ala Lys Ile Glu Asp Asp Arg Tyr Ile Arg Ile Asp Ala Val Glu
450 455 460
Trp Leu Lys Lys Asn Lys Thr Leu Asp Ser Phe Ile Ala Gln Val Thr
465 470 475 480
Thr Lys Ile Leu Arg Ala Trp Val Ser Glu Met Thr Val Tyr Thr Arg
485 490 495
Asp Asp Phe Leu Val Gln Trp Ile Asp Gly Thr Gln Thr Glu Ile Gly
500 505 510
Ser Cys Glu His His Leu Val Lys Asp Arg Asn Ser Lys Ser Tyr Glu
515 520 525
Ser Gly Glu Glu Thr Ser Arg Arg Ala Lys Phe Glu Val Asn His Ile
530 535 540
Ser Glu Thr Thr Glu Gly Gln Gly Glu Leu Asp Leu Leu Ser Lys Ser
545 550 555 560
Ala Ser Ser Asn Asn Glu Asp Ser Asn Gln Pro Glu Asn Asn Ser Thr
565 570 575
Gly Lys Glu Glu Leu Glu Leu Asn Leu Asn Ser Asn Ala Glu Ile Ile
580 585 590
Lys Ile Glu Pro Gly Gln Arg Asp Tyr Ile Met Lys Asn Leu His Lys
595 600 605
Ser Leu Ser Ala Asn Met Met Met Gln Asn Ala Ser Val His Thr Ala
610 615 620
Ser Ile Asn Lys Pro Arg Leu Lys Thr Ala Ala Tyr Cys Arg Ile Ser
625 630 635 640
Thr Asp Ser Glu Glu Gln Lys Val Ser Leu Lys Thr Gln Val Ala Tyr
645 650 655
Tyr Thr Tyr Leu Ile Leu Lys Asp Pro Gln Tyr Glu Tyr Ala Gly Ile
660 665 670
Tyr Ala Asp Glu Gly Ile Ser Gly Arg Ser Met Lys Asn Arg Thr Glu

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

 740 745 750
 Ile Ser Ile Phe Gly Ser Ile Ala Gln Glu Glu Ser Val Asn Met Gly
 755 760 765
 Glu Ala Met Ala Trp Gly Lys Arg Arg Tyr Ala Glu Arg Gly Ile Val
 770 775 780
 Asn Pro Ser Val Ala Pro Tyr Gly Tyr Arg Thr Val Arg Lys Gly Glu
 785 790 795 800
 Trp Glu Val Val Glu Glu Glu Ala Thr Ile Ile Arg Arg Ile Tyr Arg
 805 810 815

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

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

930 935 940
 Gly Ser Lys Ile Gly His Arg Arg Ser Ile Gln Ser Ser Asn Glu Ile
 945 950 955 960

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

 1025 1030 1035 1040
 Asn Ala Val Asp Gln Glu Leu Asn Lys Lys Gly Gln Asp Ser Arg Lys
 1045 1050 1055
 Val Asp Glu Leu Thr Glu Lys Ile Val Asp Leu Gln Glu Glu Leu Lys
 1060 1065 1070
 Val Phe Arg Asp Arg Lys Ala Lys Val Glu Asp Leu Lys Ala Glu Leu
 1075 1080 1085
 Glu Trp Phe Leu Lys Lys Leu Glu Thr Ile Asp Asp Ala Arg Val Lys
 1090 1095 1100

 Arg Asn Glu Gly Ile Gly His Gly Glu Glu Ile Tyr Phe Arg Glu Asp
 1105 1110 1115 1120
 Ile Phe Glu Arg Ile Val Arg Ser Ala Gln Leu Tyr Ser Asp Gly Arg
 1125 1130 1135
 Ile Val Tyr Glu Leu Ser Leu Gly Ile Gln Trp Phe Ile Asp Phe Lys
 1140 1145 1150
 Tyr Ser Ala Phe Gln Lys Leu Leu Ile Lys Trp Lys Asp Lys Gln Arg
 1155 1160 1165
 Ala Glu Glu Lys Glu Ala Phe Leu Glu Gly Pro Glu Val Lys Glu Leu

 1170 1175 1180

Leu Glu Phe Cys Lys Glu Pro Lys Ser Tyr Ser Asp Leu His Ala Phe
 1185 1190 1195 1200
 Met Cys Glu Arg Lys Glu Val Ser Tyr Ser Tyr Phe Arg Lys Leu Val
 1205 1210 1215
 Ile Arg Pro Leu Met Lys Lys Gly Lys Leu Lys Phe Thr Ile Pro Glu
 1220 1225 1230
 Asp Val Met Asn Arg His Gln Arg Tyr Thr Ser Ile
 1235 1240

<210> 66
 <211> 348
 <212> DNA
 <213> Clostridium difficile
 <400> 66
 ttagtcttca aaaggttttg gactaaattt actctcgtag tcaggccaa gtgtttcttc 60
 agattttttt ttcaaccaat ccacctgcat ggtgagctgg ccaacttttt tcgcatattc 120
 agctttttcc ttgcgttcta aagcgagttt ttctttcaga ttatcctctc gtgtgtcatt 180
 aaaaaccacg gatgctttat cgaggaaactc cttcttcag ttgcggagaa gattcggctg 240
 aatattgttt tcggttgcca ttgtatttaa gtctttttct cctttgagca gttcaatcac 300
 taattctgat ttgaatttgg cagagaaatt tcttcttggt cgagacat 348

<210> 67
 <211> 115
 <212> PRT
 <213> Peptoclostridium difficile
 <400> 67
 Met Ser Arg Thr Arg Arg Asn Phe Ser Ala Lys Phe Lys Ser Glu Leu
 1 5 10 15
 Val Ile Glu Leu Leu Lys Gly Glu Lys Asp Leu Asn Thr Ile Ala Thr
 20 25 30
 Glu Asn Asn Ile Gln Pro Asn Leu Leu Arg Asn Trp Lys Lys Glu Phe
 35 40 45
 Leu Asp Lys Ala Ser Val Val Phe Asn Asp Thr Arg Glu Asp Asn Leu
 50 55 60

Lys Glu Lys Leu Ala Leu Glu Arg Lys Glu Lys Ala Glu Tyr Ala Lys
65 70 75 80
Lys Val Gly Gln Leu Thr Met Gln Val Asp Trp Leu Lys Lys Lys Ser
85 90 95
Glu Glu Thr Leu Gly Pro Asp Tyr Glu Ser Lys Phe Ser Pro Lys Pro
100 105 110
Phe Glu Asp

115

<210> 68

<211> 2820

<212> DNA

<213> Francisella philomiragia

<400> 68

atgaatctat atagtaatct aacaaataaa tatagtttaa gtaaaactct aagatttgag 60

ttaattccac aggggtgaaac acttgaaaat ataaaagcaa gaggtttgat ttagatgat 120

gagaaaagag ctaaagacta taaaaaagct aaacaaatca ttgataaata tcatcagttt 180

tttatagagg agatattaag ttcggtatgt attagcgaag atttattaca aaactattct 240

gatgtttatt ttaaacttaa aaagagtgat gatgataatc tacaaaaaga ttttaaaagt 300

gcaaaagata cgataaagaa acacatatct agatatataa atgactcgga gaaatttaag 360

aatttgttta atcaaaatct tatagatgct aaaaaagggc aagagtcaga ttttaattcta 420

tggctaagc aatctaagga taatggcata gaactattta aagctaacag tgatcacaca 480

gacatagatg aggcggttaga aataatcaaa tcttttaaag gttggacaac ttattttaag 540

ggttttcatg aaaatagaaa aaatgtctat agtagtgatg ataccctac atctattatt 600

tatagaatag tagatgataa tttgcctaaa tttatagaaa ataaagctaa gtatgagaat 660

ttaaaagaca aagctccaga agctataaac tatgaacaaa ttaaaaaaga tttggcagaa 720

gagctaacct ttgatattga ctacaaaaca tctgaagtta atcaaagagt tttttcactt 780

gatgaagttt ttgagatagc aaactttaat aattatctaa atcaaagtgg tattactaaa 840

tttaatacta ttattgggtgg taaatttggt aatggtgaaa atacaaagag aaaaggtata 900

aatgaatata taaatctata ctacacgcaa ataaatgata aaacacttaa aaaatataaa 960

atgagtgttt tatttaagca aattttaagt gatacagaat ctaaatcttt tgtaattgat 1020

aagttagaag atgatatgta tgtagttaca acgatgcaaa gtttttatga gcaaatagca 1080

gcttttataaa cattagaaga aaagtctatt aaggaaacat tatctttact atttgatgat 1140

ttaaaagctc aaaaacttga tttagtaaa atttatitta aaaatgataa atctcttact	1200
gatctatcac aacaagtttt tgatgattat agtgttattg gtacagcggg actagaatat	1260
ataactcaac aagtagcacc taaaaatctt gataacccta gtaagaaaga gcaagattta	1320
atagccaaaa aaactgaaaa agcaaaatac ttatctctag aaactataaa gcttgcctta	1380
gaagaattta ataagtatag agatatagat aaacagtga ggtttgaaga aatatttgca	1440
agctttgcag atattccggg gctatttgat gaaatagctc aaaacaaaaa caatttgca	1500
cagatatcta tcaaatatca aaatcaaggg aaaaaagacc tgcttcaaac tagtgcagaa	1560
gtagatgtta aagctatcaa ggatcttttg gatcaaaacta ataatctctt gcataaacta	1620
aaaatatttc atattacgca atcagaagat aaggcaaata ttttagacaa ggatgagcat	1680
ttttatttag tatttgatga gtgctacttt gagctagcga atatagtggc tctttataac	1740
aaaattagaa actatataac tcaaaagcca tatagtgatg agaaatttaa gctcaatttt	1800
gagaactcaa ctttagccaa tggttgggat aaaaataaag agcctgacaa tacggcaatt	1860
ttatttatca aagatgataa atattatctg ggtgtgatga acaagaaaaa taacaaaata	1920
tttgatgata aagctatcaa agaaaaataa ggtgaaggat ataagaaagt tgtatataaa	1980
cttttaccgg gtgcaataa aatgttacct aaggttttct tttctgctaa atctataaat	2040
ttttataatc ctagtgaaga tatacttaga ataagaaacc actcaacaca tacaaaaaat	2100
ggtagtcctc aaaaaggata tgaaaaactt gagtttaata ttgaagattg ccgaaaattt	2160
atagattttt ataacattc tataagtagg catccagagt ggaaagattt tggatttaga	2220
ttttctgata ctaaaaaata caactctata gatgaatttt atagagaagt tgaaaatcaa	2280
ggctacaaac taacttttga aaatatatca gaaagctata ttgatagttt agtcgatgaa	2340
ggcaaattat acctattcca aatctataat aaagatttct cagtatatag taagggtaaa	2400
ccaaatttac atacgtata ttggaaggcg ttgtttgatg agagaaatct ccaagatgta	2460
glatataaat taaatgggta agcagaactc ttctatcgta aacaatcaat acctaagaaa	2520
atcactcacc cagccaaaga ggcaatagct aataaaaaca aagataatcc taaaaaagag	2580
agtatttttg aatatgattt aatcaaagat aaacgcttta ctgaagataa gtttttcttt	2640
cactgtccta ttacaatcaa tttcaaatct agtggagcta ataagtttaa tgatgaaatc	2700
aattttattgc taaaagaaaa agcaaatgat gttcatatcc taagtataga tagaggagaa	2760
agacatttag ctactatac tttagtagat ggtaaaggaa acattatctg taagaattaa	2820
	2820
<210> 69	
<211> 356	

<212> PRT

<213> Francisella philomiragia

<400> 69

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

1 5 10 15

Glu Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn Ile Lys Glu Met

20 25 30

Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile Ala Lys Leu Val

35 40 45

Ile Gly Tyr Asn Ala Ile Val Val Phe Glu Asp Leu Asn Phe Gly Phe

50 55 60

Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val Tyr Gln Lys Leu Glu

65 70 75 80

Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu Val Phe Lys Asp Asn Glu

85 90 95

Phe Asp Lys Ala Gly Gly Val Leu Arg Ala Tyr Gln Leu Thr Ala Pro

100 105 110

Phe Glu Thr Phe Lys Lys Met Gly Lys Gln Thr Gly Ile Ile Tyr Tyr

115 120 125

Val Pro Ala Asp Phe Thr Ser Lys Ile Cys Pro Val Thr Gly Phe Val

130 135 140

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

145 150 155 160

Phe Ser Lys Phe Asp Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe

165 170 175

Glu Phe Ser Phe Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly

180 185 190

Lys Trp Thr Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn

195 200 205

Ser Asp Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys

210 215 220

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

 275 280 285
 Asn Gly Asn Phe Phe Asp Ser Arg His Ala Pro Lys Asn Met Pro Gln
 290 295 300
 Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly Leu Lys Gly Leu Met
 305 310 315 320
 Leu Leu Tyr Arg Ile Lys Asn Asn Gln Asp Gly Lys Lys Leu Asn Leu
 325 330 335
 Val Ile Lys Asn Glu Glu Tyr Phe Glu Phe Val Gln Asn Arg Asn Lys
 340 345 350

Ser Ser Lys Ile

355

<210> 70

<211> 878

<212> DNA

<213> Human immunodeficiency virus 1

<400> 70

ttcctggacg gtatcgataa agctcaggaa gaacacgaaa aataccactc taactggcgc	60
gccatggctt ctgacttcaa cctgccgccg gttgttgcca aggaaatcgt ggcttcttgc	120
gacaaatgcc aattgaaagg tgaagctatg catggtcagg tcgactgctc tccaggtatc	180
tggcagctgg actgcactca tctcgagggt aaagttatcc tggttgctgt tcacgtggct	240
tccggataca tcgaagetga agttatcccg gctgaaaccg gtcaggaaac tgcttacttc	300
ctgcttaagc tggccggccg ttggccggtt aaaactgttc aactgacaa cggttctaac	360
ttcactagta ctactgttaa agctgcatgc tgggtgggccg gcatcaaaca ggagttcggg	420
atcccgta acccgcagtc tcaggcggtt atcgaatcta tgaacaaaga gtcacaaaaa	480
atcattggcc aggtacgtga tcaggctgag cacctgaaaa ccgcggtgca gatggctgtt	540

ttcatccaca acttcaaacg taaagggtgt atcggtggtt acagcgctgg tgaacgtatc 600
gttgacatca tcgtactga tatccagact aaagaactgc agaaacagat cactaaaatc 660
cagaacttcc gtgtatacta ccgtgactct agagaccggg tttggaaagg tcctgctaaa 720

ctcctgtgga aggggtgaagg tgctgttgtt atccaggaca actctgacat caaagtggta 780
ccgcgtcgta aagctaaaat cattcgcgac tacggcaaac agatggctgg tgacgactgc 840
gttgctagcc gtcaggacga agactaaaag cttcaggc 878

<210> 71
<211> 288
<212> PRT
<213> Human immunodeficiency virus 1
<400> 71

Phe Leu Asp Gly Ile Asp Lys Ala Gln Glu Glu His Glu Lys Tyr His
1 5 10 15
Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val

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

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

Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val

165 170 175

Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly

180 185 190

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

195 200 205

Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg

210 215 220

Val Tyr Tyr Arg Asp Ser Arg Asp Pro Val Trp Lys Gly Pro Ala Lys

225 230 235 240

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

245 250 255

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

260 265 270

Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp

275 280 285

<210> 72

<211> 1307

<212> PRT

<213> Acidaminococcus sp. BV3L6

<400> 72

Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr

1 5 10 15

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

20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys

35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln

50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile

65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
 85 90 95
 Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
 100 105 110
 Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
 115 120 125
 Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
 130 135 140
 Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
 145 150 155 160
 Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
 165 170 175
 Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
 180 185 190
 Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
 195 200 205
 Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
 210 215 220
 Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
 225 230 235 240
 Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
 245 250 255
 Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
 260 265 270
 Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
 275 280 285
 Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
 290 295 300
 Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
 305 310 315 320
 Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
 325 330 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
 340 345 350
 Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
 355 360 365

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

 435 440 445
 Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
 450 455 460
 Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
 465 470 475 480
 Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
 485 490 495
 Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
 500 505 510

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

580	585	590
Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys		
595	600	605
Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr		
610	615	620
Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys		
625	630	635
Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln		
645	650	655
Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala		
660	665	670
Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr		
675	680	685
Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr		
690	695	700
Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His		
705	710	715
Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu		
725	730	735
Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys		
740	745	750
Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu		
755	760	765
Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln		
770	775	780
Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His		
785	790	795
Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr		
805	810	815
Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His		
820	825	830
Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn		

835	840	845	
Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe			
850	855	860	
Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln			
865	870	875	880
Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu			
	885	890	895
Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg			
	900	905	910
Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu			
	915	920	925
Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu			
	930	935	940
Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val			
945	950	955	960
Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile			
	965	970	975
His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu			
	980	985	990
Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu			
	995	1000	1005
Lys Ala Val Tyr Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu Asn			
1010	1015	1020	
Cys Leu Val Leu Lys Asp Tyr Pro Ala Glu Lys Val Gly Gly Val Leu			
1025	1030	1035	1040
Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala Lys Met Gly			
	1045	1050	1055
Thr Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro Tyr Thr Ser Lys			
	1060	1065	1070
Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe Val Trp Lys Thr Ile			
	1075	1080	1085

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

<210> 73

<211> 1206

<212> PRT

<213> Lachnospiraceae bacterium MA2020

<400> 73

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

1 5 10 15

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

20 25 30

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

35 40 45

Val Lys Gly Ile Leu Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala

50 55 60

Leu Asp Asn Cys Thr Leu Pro Ser Leu Lys Ile Ala Ala Glu Ile Tyr

65 70 75 80

Leu Lys Asn Gln Lys Glu Val Ser Asp Arg Glu Asp Phe Asn Lys Thr

85 90 95

Gln Asp Leu Leu Arg Lys Glu Val Val Glu Lys Leu Lys Ala His Glu

100 105 110

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

115 120 125

Leu Pro Ser Ile Ser Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Arg

130 135 140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu

145 150 155 160

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

165 170 175

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

180 185 190

Lys Thr Ala Gly Ile Leu Ala Asp Gly Leu Gly Glu Glu Glu Gln Asp

195 200 205

Ser Leu Phe Ile Val Glu Thr Phe Asn Lys Thr Leu Thr Gln Asp Gly

210 215 220

Ile Asp Thr Tyr Asn Ser Gln Val Gly Lys Ile Asn Ser Ser Ile Asn

225 230 235 240

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

245 250 255
Lys Met Lys Met Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser

260 265 270
Phe Ile Asp Glu Phe Gln Ser Asp Glu Val Leu Ile Asp Asn Val Glu

275 280 285
Ser Tyr Gly Ser Val Leu Ile Glu Ser Leu Lys Ser Ser Lys Val Ser

290 295 300
Ala Phe Phe Asp Ala Leu Arg Glu Ser Lys Gly Lys Asn Val Tyr Val

305 310 315 320
Lys Asn Asp Leu Ala Lys Thr Ala Met Ser Asn Ile Val Phe Glu Asn

325 330 335

Trp Arg Thr Phe Asp Asp Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
340 345 350

Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
355 360 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Glu His Leu Cys Asn Leu Ser
370 375 380

Glu Asp Ser Cys Asn Leu Ile Glu Asn Tyr Ile His Gln Ile Ser Asp
385 390 395 400

Asp Ile Glu Asn Ile Ile Ile Asn Asn Glu Thr Phe Leu Arg Ile Val

405 410 415
Ile Asn Glu His Asp Arg Ser Arg Lys Leu Ala Lys Asn Arg Lys Ala

420 425 430
Val Lys Ala Ile Lys Asp Phe Leu Asp Ser Ile Lys Val Leu Glu Arg

435 440 445
Glu Leu Lys Leu Ile Asn Ser Ser Gly Gln Glu Leu Glu Lys Asp Leu

450 455 460
Ile Val Tyr Ser Ala His Glu Glu Leu Leu Val Glu Leu Lys Gln Val

465 470 475 480

Asp Ser Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
485 490 495

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

 545 550 555 560
 Val Asp Tyr Lys Leu Leu Pro Val Pro Asn Gln Met Leu Pro Lys Val
 565 570 575
 Phe Phe Ala Lys Ser Asn Ile Asp Phe Tyr Asn Pro Ser Ser Glu Ile
 580 585 590
 Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Asn Met Phe Ser
 595 600 605
 Leu Glu Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Ser
 610 615 620

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

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

Arg Ala Arg Thr Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
755 760 765

Asp Lys Arg Tyr Ser Lys Asp Lys Phe Thr Leu His Ile Pro Ile Thr
770 775 780

Met Asn Phe Gly Val Asp Glu Val Lys Arg Phe Asn Asp Ala Val Asn
785 790 795 800

Ser Ala Ile Arg Ile Asp Glu Asn Val Asn Val Ile Gly Ile Asp Arg
805 810 815

Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Ser Lys Gly Asn
820 825 830

Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp

835 840 845

Ile Glu Thr Asp Tyr His Ala Leu Leu Asp Glu Arg Glu Gly Gly Arg
850 855 860

Asp Lys Ala Arg Lys Asp Trp Asn Thr Val Glu Asn Ile Arg Asp Leu
865 870 875 880

Lys Ala Gly Tyr Leu Ser Gln Val Val Asn Val Val Ala Lys Leu Val
885 890 895

Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
900 905 910

Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
915 920 925

Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
930 935 940

Glu Gln Thr Ser Pro Lys Glu Leu Gly Gly Ala Leu Asn Ala Leu Gln
945 950 955 960

Leu Thr Ser Lys Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Ser Gly
965 970 975

Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr

980 985 990

Thr Gly Phe Ala Asn Leu Phe Tyr Met Lys Cys Glu Asn Val Glu Lys

995 1000 1005
 Ser Lys Arg Phe Phe Asp Gly Phe Asp Phe Ile Arg Phe Asn Ala Leu
 1010 1015 1020
 Glu Asn Val Phe Glu Phe Gly Phe Asp Tyr Arg Ser Phe Thr Gln Arg
 1025 1030 1035 1040
 Ala Cys Gly Ile Asn Ser Lys Trp Thr Val Cys Thr Asn Gly Glu Arg
 1045 1050 1055

 Ile Ile Lys Tyr Arg Asn Pro Asp Lys Asn Asn Met Phe Asp Glu Lys
 1060 1065 1070
 Val Val Val Val Thr Asp Glu Met Lys Asn Leu Phe Glu Gln Tyr Lys
 1075 1080 1085
 Ile Pro Tyr Glu Asp Gly Arg Asn Val Lys Asp Met Ile Ile Ser Asn
 1090 1095 1100
 Glu Glu Ala Glu Phe Tyr Arg Arg Leu Tyr Arg Leu Leu Gln Gln Thr
 1105 1110 1115 1120
 Leu Gln Met Arg Asn Ser Thr Ser Asp Gly Thr Arg Asp Tyr Ile Ile

 1125 1130 1135
 Ser Pro Val Lys Asn Lys Arg Glu Ala Tyr Phe Asn Ser Glu Leu Ser
 1140 1145 1150
 Asp Gly Ser Val Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile
 1155 1160 1165
 Ala Arg Lys Gly Leu Trp Val Leu Glu Gln Ile Arg Gln Lys Ser Glu
 1170 1175 1180
 Gly Glu Lys Ile Asn Leu Ala Met Thr Asn Ala Glu Trp Leu Glu Tyr
 1185 1190 1195 1200

 Ala Gln Thr His Leu Leu
 1205

 <210> 74
 <211> 1300
 <212> PRT
 <213> Francisella tularensis
 <400> 74

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15
Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30
Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45
Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80
Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95
Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

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

260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys

275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile

290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys

305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser

325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

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

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400

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

405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn

420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala

435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn

450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495

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

500	505	510	
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp			
515	520	525	
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His			
530	535	540	
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His			
545	550	555	560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val			
565	570	575	
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser			
580	585	590	
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly			
595	600	605	
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys			
610	615	620	
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile			
625	630	635	640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys			
645	650	655	
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val			
660	665	670	
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile			
675	680	685	
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln			
690	695	700	
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe			
705	710	715	720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp			
725	730	735	
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu			
740	745	750	
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn			

755 760 765
 Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
 770 775 780
 Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
 785 790 795 800
 Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
 805 810 815
 Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
 820 825 830
 Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
 835 840 845
 Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
 850 855 860
 Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
 865 870 875 880
 His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
 885 890 895
 Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
 900 905 910
 Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
 915 920 925
 Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
 930 935 940
 Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
 945 950 955 960
 Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
 965 970 975
 Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
 980 985 990
 Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
 995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val Tyr
 1010 1015 1020
 Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu Val Phe
 1025 1030 1035 1040
 Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg Ala Tyr Gln
 1045 1050 1055

 Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly Lys Gln Thr Gly
 1060 1065 1070
 Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser Lys Ile Cys Pro Val
 1075 1080 1085
 Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys Tyr Glu Ser Val Ser Lys
 1090 1095 1100
 Ser Gln Glu Phe Phe Ser Lys Phe Asp Lys Ile Cys Tyr Asn Leu Asp
 1105 1110 1115 1120
 Lys Gly Tyr Phe Glu Phe Ser Phe Asp Tyr Lys Asn Phe Gly Asp Lys

 1125 1130 1135
 Ala Ala Lys Gly Lys Trp Thr Ile Ala Ser Phe Gly Ser Arg Leu Ile
 1140 1145 1150
 Asn Phe Arg Asn Ser Asp Lys Asn His Asn Trp Asp Thr Arg Glu Val
 1155 1160 1165
 Tyr Pro Thr Lys Glu Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu
 1170 1175 1180
 Tyr Gly His Gly Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp
 1185 1190 1195 1200

 Lys Lys Phe Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln
 1205 1210 1215
 Met Arg Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro
 1220 1225 1230
 Val Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
 1235 1240 1245
 Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly Leu
 1250 1255 1260

Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu Gly Lys

1265 1270 1275 1280

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

1285 1290 1295

Asn Arg Asn Asn

1300

<210> 75

<211> 1282

<212> PRT

<213> Eubacterium eligens

<400> 75

Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Ile

1 5 10 15

Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr

20 25 30

Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg

35 40 45

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

50 55 60

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

65 70 75 80

Leu Leu Phe Glu Leu Met Asn Leu Val Gln Ser Ser Pro Ser Lys Asp

85 90 95

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

100 105 110

Cys Thr His Leu Gln Ser Asp Ser Asn Tyr Lys Asn Ile Phe Asn Ala

115 120 125

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

130 135 140

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

145 150 155 160

Asn Gly Phe Ser Thr Tyr Phe Thr Asp Phe Phe Glu Lys Arg Lys Asn

165 170 175

Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val

180 185 190

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

195 200 205

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

210 215 220

Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp

225 230 235 240

Phe Tyr Asn Met Val Leu Ile Gln Ser Gly Ile Asp Phe Tyr Asn Glu

245 250 255

Ile Cys Gly Val Val Asn Ala His Met Asn Leu Tyr Cys Gln Gln Thr

260 265 270

Lys Asn Asn Tyr Asn Leu Phe Lys Met Arg Lys Leu His Lys Gln Ile

275 280 285

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

290 295 300

Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu

305 310 315 320

Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp

325 330 335

Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr

340 345 350

Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val

355 360 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu

370 375 380

Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn

385 390 395 400

Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn

405 410 415
 Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
 420 425 430
 Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Asp His Ile
 435 440 445
 Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Met Lys Lys Arg Leu

 450 455 460
 Asp Met Tyr Met Asn Met Tyr His Trp Ala Lys Ala Phe Ile Val Asp
 465 470 475 480
 Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
 485 490 495
 Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
 500 505 510
 Tyr Val Thr Gln Lys Pro Tyr Asn Ser Lys Lys Ile Lys Leu Asn Phe
 515 520 525

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

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

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

Lys Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys
 915 920 925
 Asn Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
 930 935 940
 Ser Gly Val Val His Glu Ile Ala Met Leu Ile Val Glu Tyr Asn Ala
 945 950 955 960

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

 1025 1030 1035 1040
 Thr Ser Lys Ile Asp Pro Ser Thr Gly Phe Ile Ser Ala Phe Asn Phe
 1045 1050 1055
 Lys Ser Ile Ser Thr Asn Ala Ser Arg Lys Gln Phe Phe Met Gln Phe
 1060 1065 1070
 Asp Glu Ile Arg Tyr Cys Ala Glu Lys Asp Met Phe Ser Phe Gly Phe
 1075 1080 1085
 Asp Tyr Asn Asn Phe Asp Thr Tyr Asn Ile Thr Met Gly Lys Thr Gln
 1090 1095 1100

 Trp Thr Val Tyr Thr Asn Gly Glu Arg Leu Gln Ser Glu Phe Asn Asn
 1105 1110 1115 1120
 Ala Arg Arg Thr Gly Lys Thr Lys Ser Ile Asn Leu Thr Glu Thr Ile
 1125 1130 1135
 Lys Leu Leu Leu Glu Asp Asn Glu Ile Asn Tyr Ala Asp Gly His Asp
 1140 1145 1150
 Ile Arg Ile Asp Met Glu Lys Met Asp Glu Asp Lys Lys Ser Glu Phe
 1155 1160 1165

Phe Ala Gln Leu Leu Ser Leu Tyr Lys Leu Thr Val Gln Met Arg Asn

1170 1175 1180
 Ser Tyr Thr Glu Ala Glu Glu Gln Glu Asn Gly Ile Ser Tyr Asp Lys
 1185 1190 1195 1200
 Ile Ile Ser Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp
 1205 1210 1215
 Asn Tyr Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala
 1220 1225 1230
 Asp Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val
 1235 1240 1245

Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn Cys
 1250 1255 1260
 Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn Lys Arg
 1265 1270 1275 1280
 Tyr Glu

<210> 76
 <211> 1263
 <212> PRT
 <213> Leptospira inadai
 <400> 76

Met Glu Asp Tyr Ser Gly Phe Val Asn Ile Tyr Ser Ile Gln Lys Thr
 1 5 10 15
 Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu His Ile Glu

20 25 30
 Lys Lys Gly Phe Leu Lys Lys Asp Lys Ile Arg Ala Glu Asp Tyr Lys
 35 40 45
 Ala Val Lys Lys Ile Ile Asp Lys Tyr His Arg Ala Tyr Ile Glu Glu
 50 55 60
 Val Phe Asp Ser Val Leu His Gln Lys Lys Lys Lys Asp Lys Thr Arg
 65 70 75 80
 Phe Ser Thr Gln Phe Ile Lys Glu Ile Lys Glu Phe Ser Glu Leu Tyr

85

90

95

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

100

105

110

Ser Glu Lys Leu Arg Lys Met Leu Val Gly Ala Phe Lys Gly Glu Phe

115

120

125

Ser Glu Glu Val Ala Glu Lys Tyr Lys Asn Leu Phe Ser Lys Glu Leu

130

135

140

Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg Lys

145

150

155

160

Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His

165

170

175

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

180

185

190

Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp Asn Leu

195

200

205

Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro Trp Ser

210

215

220

Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys Leu Thr

225

230

235

240

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

245

250

255

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

260

265

270

Glu Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Arg Gln Lys

275

280

285

Asn Asn Ile Asp Arg Lys Asn Leu Pro Asn Val Lys Ile Leu Phe Lys

290

295

300

Gln Ile Leu Gly Asp Arg Glu Thr Lys Ser Phe Ile Pro Glu Ala Phe

305

310

315

320

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

325

330

335

Leu Lys Leu Asp Lys Lys Lys Lys Ser Ile Ile Ala Glu Leu Lys Lys
 340 345 350
 Phe Leu Ser Ser Phe Asn Arg Tyr Glu Leu Asp Gly Ile Tyr Leu Ala
 355 360 365
 Asn Asp Asn Ser Leu Ala Ser Ile Ser Thr Phe Leu Phe Asp Asp Trp
 370 375 380

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

 450 455 460
 Lys Gln Phe Asp Leu Leu Glu Arg Ile Glu Glu Ala Tyr Ala Ile Val
 465 470 475 480
 Glu Pro Leu Leu Gly Ala Glu Tyr Pro Arg Asp Arg Asn Leu Lys Ala
 485 490 495
 Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser Ile Lys
 500 505 510
 Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile Phe Asp
 515 520 525

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

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

595 600 605

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

610 615 620

Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro His Met

625 630 635 640

Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile Tyr Asn

645 650 655

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

660 665 670

Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp Phe Tyr

675 680 685

Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp Phe Lys

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn Leu His

755 760 765

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

770 775 780

Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys Lys Ser

785 790 795 800

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

805 810 815

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

820 825 830

Asp Lys Phe Gln Phe His Leu Pro Ile Ser Leu Asn Phe Lys Ser Lys

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

 885 890 895
 Leu Asp Ser Met Gln Ser Gly Lys Gly Arg Pro Glu Ile Asn Tyr Lys
 900 905 910
 Glu Lys Leu Gln Glu Lys Glu Ile Glu Arg Asp Lys Ala Arg Lys Ser
 915 920 925
 Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser
 930 935 940
 Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn Ala Ile
 945 950 955 960

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

 1025 1030 1035 1040
 Ser Lys Ile Asp Pro Arg Thr Gly Phe Ile Asp Phe Leu His Pro Ala
 1045 1050 1055
 Tyr Glu Asn Ile Glu Lys Ala Lys Gln Trp Ile Asn Lys Phe Asp Ser
 1060 1065 1070
 Ile Arg Phe Asn Ser Lys Met Asp Trp Phe Glu Phe Thr Ala Asp Thr
 1075 1080 1085
 Arg Lys Phe Ser Glu Asn Leu Met Leu Gly Lys Asn Arg Val Trp Val

1090	1095	1100	
Ile Cys Thr Thr Asn Val Glu Arg Tyr Phe Thr Ser Lys Thr Ala Asn			
1105	1110	1115	1120
Ser Ser Ile Gln Tyr Asn Ser Ile Gln Ile Thr Glu Lys Leu Lys Glu			
	1125	1130	1135
Leu Phe Val Asp Ile Pro Phe Ser Asn Gly Gln Asp Leu Lys Pro Glu			
	1140	1145	1150
Ile Leu Arg Lys Asn Asp Ala Val Phe Phe Lys Ser Leu Leu Phe Tyr			
	1155	1160	1165
Ile Lys Thr Thr Leu Ser Leu Arg Gln Asn Asn Gly Lys Lys Gly Glu			
1170	1175	1180	
Glu Glu Lys Asp Phe Ile Leu Ser Pro Val Val Asp Ser Lys Gly Arg			
1185	1190	1195	1200
Phe Phe Asn Ser Leu Glu Ala Ser Asp Asp Glu Pro Lys Asp Ala Asp			
	1205	1210	1215
Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Met Asn Leu Leu			
	1220	1225	1230
Val Leu Asn Glu Thr Lys Glu Glu Asn Leu Ser Arg Pro Lys Trp Lys			
	1235	1240	1245
Ile Lys Asn Lys Asp Trp Leu Glu Phe Val Trp Glu Arg Asn Arg			
1250	1255	1260	
<210> 77			
<211> 1260			
<212> PRT			
<213> Porphyromonas crevioricanis			
<400> 77			
Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr			
1 5 10 15			
Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu			
20 25 30			
Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg			

35	40	45	
Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser			
50	55	60	
Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala			
65	70	75	80
Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu			
85	90	95	
Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu			
100	105	110	
Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val			
115	120	125	
Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu			
130	135	140	
Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe			
145	150	155	160
Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr			
165	170	175	
Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr			
180	185	190	
Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu			
195	200	205	
Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro			
210	215	220	
Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly			
225	230	235	240
Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr			
245	250	255	
Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu			
260	265	270	
Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn			
275	280	285	
Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu			

290 295 300
 Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
 305 310 315 320
 Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg

 325 330 335
 Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
 340 345 350
 Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
 355 360 365
 Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
 370 375 380
 Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
 385 390 395 400

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

 465 470 475 480
 Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
 485 490 495
 Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
 500 505 510
 Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
 515 520 525
 Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
 530 535 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545 550 555 560
Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
565 570 575
Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
580 585 590
Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
595 600 605
Phe Glu Asn Lys Met Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
610 615 620
Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625 630 635 640
Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
645 650 655
Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
660 665 670
Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
675 680 685
Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
690 695 700
Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705 710 715 720
Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
725 730 735
Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
740 745 750
Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
755 760 765
Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
770 775 780
Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785 790 795 800

Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
 805 810 815
 Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
 820 825 830

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

 900 905 910
 Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu His Arg Asn Trp
 915 920 925
 Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
 930 935 940
 Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
 945 950 955 960
 Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
 965 970 975

 Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
 980 985 990
 Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
 995 1000 1005
 Leu Arg Ala Tyr Gln Phe Thr Ala Pro Phe Lys Ser Phe Lys Glu Met
 1010 1015 1020
 Gly Lys Gln Asn Gly Phe Leu Phe Tyr Ile Pro Ala Trp Asn Thr Ser
 1025 1030 1035 1040
 Asn Ile Asp Pro Thr Thr Gly Phe Val Asn Leu Phe His Val Gln Tyr

1045 1050 1055
 Glu Asn Val Asp Lys Ala Lys Ser Phe Phe Gln Lys Phe Asp Ser Ile
 1060 1065 1070
 Ser Tyr Asn Pro Lys Lys Asp Trp Phe Glu Phe Ala Phe Asp Tyr Lys
 1075 1080 1085
 Asn Phe Thr Lys Lys Ala Glu Gly Ser Arg Ser Met Trp Ile Leu Cys
 1090 1095 1100
 Thr His Gly Ser Arg Ile Lys Asn Phe Arg Asn Ser Gln Lys Asn Gly
 1105 1110 1115 1120

 Gln Trp Asp Ser Glu Glu Phe Ala Leu Thr Glu Ala Phe Lys Ser Leu
 1125 1130 1135
 Phe Val Arg Tyr Glu Ile Asp Tyr Thr Ala Asp Leu Lys Thr Ala Ile
 1140 1145 1150
 Val Asp Glu Lys Gln Lys Asp Phe Phe Val Asp Leu Leu Lys Leu Phe
 1155 1160 1165
 Lys Leu Thr Val Gln Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp
 1170 1175 1180
 Tyr Leu Ile Ser Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr

 1185 1190 1195 1200
 Arg Glu Gly Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala
 1205 1210 1215
 Tyr Asn Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg Gln
 1220 1225 1230
 Thr Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys Glu Trp
 1235 1240 1245
 Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp
 1250 1255 1260

<210> 78
 <211> 1246
 <212> PRT
 <213> Porphyromonas macacae
 <400> 78

Met Lys Thr Gln His Phe Phe Glu Asp Phe Thr Ser Leu Tyr Ser Leu
1 5 10 15
Ser Lys Thr Ile Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu
20 25 30
Asn Ile Lys Lys Asn Gly Leu Ile Arg Arg Asp Glu Gln Arg Leu Asp
35 40 45
Asp Tyr Glu Lys Leu Lys Lys Val Ile Asp Glu Tyr His Glu Asp Phe
50 55 60
Ile Ala Asn Ile Leu Ser Ser Phe Ser Phe Ser Glu Glu Ile Leu Gln
65 70 75 80
Ser Tyr Ile Gln Asn Leu Ser Glu Ser Glu Ala Arg Ala Lys Ile Glu
85 90 95
Lys Thr Met Arg Asp Thr Leu Ala Lys Ala Phe Ser Glu Asp Glu Arg
100 105 110
Tyr Lys Ser Ile Phe Lys Lys Glu Leu Val Lys Lys Asp Ile Pro Val
115 120 125
Trp Cys Pro Ala Tyr Lys Ser Leu Cys Lys Lys Phe Asp Asn Phe Thr
130 135 140
Thr Ser Leu Val Pro Phe His Glu Asn Arg Lys Asn Leu Tyr Thr Ser
145 150 155 160
Asn Glu Ile Thr Ala Ser Ile Pro Tyr Arg Ile Val His Val Asn Leu
165 170 175
Pro Lys Phe Ile Gln Asn Ile Glu Ala Leu Cys Glu Leu Gln Lys Lys
180 185 190
Met Gly Ala Asp Leu Tyr Leu Glu Met Met Glu Asn Leu Arg Asn Val
195 200 205
Trp Pro Ser Phe Val Lys Thr Pro Asp Asp Leu Cys Asn Leu Lys Thr
210 215 220
Tyr Asn His Leu Met Val Gln Ser Ser Ile Ser Glu Tyr Asn Arg Phe
225 230 235 240
Val Gly Gly Tyr Ser Thr Glu Asp Gly Thr Lys His Gln Gly Ile Asn
245 250 255

Glu Trp Ile Asn Ile Tyr Arg Gln Arg Asn Lys Glu Met Arg Leu Pro
260 265 270

Gly Leu Val Phe Leu His Lys Gln Ile Leu Ala Lys Val Asp Ser Ser
275 280 285

Ser Phe Ile Ser Asp Thr Leu Glu Asn Asp Asp Gln Val Phe Cys Val
290 295 300

Leu Arg Gln Phe Arg Lys Leu Phe Trp Asn Thr Val Ser Ser Lys Glu
305 310 315 320

Asp Asp Ala Ala Ser Leu Lys Asp Leu Phe Cys Gly Leu Ser Gly Tyr
325 330 335

Asp Pro Glu Ala Ile Tyr Val Ser Asp Ala His Leu Ala Thr Ile Ser
340 345 350

Lys Asn Ile Phe Asp Arg Trp Asn Tyr Ile Ser Asp Ala Ile Arg Arg
355 360 365

Lys Thr Glu Val Leu Met Pro Arg Lys Lys Glu Ser Val Glu Arg Tyr
370 375 380

Ala Glu Lys Ile Ser Lys Gln Ile Lys Lys Arg Gln Ser Tyr Ser Leu
385 390 395 400

Ala Glu Leu Asp Asp Leu Leu Ala His Tyr Ser Glu Glu Ser Leu Pro
405 410 415

Ala Gly Phe Ser Leu Leu Ser Tyr Phe Thr Ser Leu Gly Gly Gln Lys
420 425 430

Tyr Leu Val Ser Asp Gly Glu Val Ile Leu Tyr Glu Glu Gly Ser Asn
435 440 445

Ile Trp Asp Glu Val Leu Ile Ala Phe Arg Asp Leu Gln Val Ile Leu
450 455 460

Asp Lys Asp Phe Thr Glu Lys Lys Leu Gly Lys Asp Glu Glu Ala Val
465 470 475 480

Ser Val Ile Lys Lys Ala Leu Asp Ser Ala Leu Arg Leu Arg Lys Phe
485 490 495

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

500	505	510	
Phe Tyr Ala Leu Tyr Thr Asp Arg Met Asp Lys Leu Lys Gly Leu Leu			
515	520	525	
Lys Met Tyr Asp Lys Val Arg Asn Tyr Leu Thr Lys Lys Pro Tyr Ser			
530	535	540	
Ile Glu Lys Phe Lys Leu His Phe Asp Asn Pro Ser Leu Leu Ser Gly			
545	550	555	560
Trp Asp Lys Asn Lys Glu Leu Asn Asn Leu Ser Val Ile Phe Arg Gln			
565	570	575	
Asn Gly Tyr Tyr Tyr Leu Gly Ile Met Thr Pro Lys Gly Lys Asn Leu			
580	585	590	
Phe Lys Thr Leu Pro Lys Leu Gly Ala Glu Glu Met Phe Tyr Glu Lys			
595	600	605	
Met Glu Tyr Lys Gln Ile Ala Glu Pro Met Leu Met Leu Pro Lys Val			
610	615	620	
Phe Phe Pro Lys Lys Thr Lys Pro Ala Phe Ala Pro Asp Gln Ser Val			
625	630	635	640
Val Asp Ile Tyr Asn Lys Lys Thr Phe Lys Thr Gly Gln Lys Gly Phe			
645	650	655	
Asn Lys Lys Asp Leu Tyr Arg Leu Ile Asp Phe Tyr Lys Glu Ala Leu			
660	665	670	
Thr Val His Glu Trp Lys Leu Phe Asn Phe Ser Phe Ser Pro Thr Glu			
675	680	685	
Gln Tyr Arg Asn Ile Gly Glu Phe Phe Asp Glu Val Arg Glu Gln Ala			
690	695	700	
Tyr Lys Val Ser Met Val Asn Val Pro Ala Ser Tyr Ile Asp Glu Ala			
705	710	715	720
Val Glu Asn Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe			
725	730	735	
Ser Pro Tyr Ser Lys Gly Ile Pro Asn Leu His Thr Leu Tyr Trp Lys			
740	745	750	
Ala Leu Phe Ser Glu Gln Asn Gln Ser Arg Val Tyr Lys Leu Cys Gly			

755 760 765
Gly Gly Glu Leu Phe Tyr Arg Lys Ala Ser Leu His Met Gln Asp Thr

770 775 780
Thr Val His Pro Lys Gly Ile Ser Ile His Lys Lys Asn Leu Asn Lys
785 790 795 800

Lys Gly Glu Thr Ser Leu Phe Asn Tyr Asp Leu Val Lys Asp Lys Arg
805 810 815

Phe Thr Glu Asp Lys Phe Phe Phe His Val Pro Ile Ser Ile Asn Tyr
820 825 830

Lys Asn Lys Lys Ile Thr Asn Val Asn Gln Met Val Arg Asp Tyr Ile
835 840 845

Ala Gln Asn Asp Asp Leu Gln Ile Ile Gly Ile Asp Arg Gly Glu Arg
850 855 860

Asn Leu Leu Tyr Ile Ser Arg Ile Asp Thr Arg Gly Asn Leu Leu Glu
865 870 875 880

Gln Phe Ser Leu Asn Val Ile Glu Ser Asp Lys Gly Asp Leu Arg Thr
885 890 895

Asp Tyr Gln Lys Ile Leu Gly Asp Arg Glu Gln Glu Arg Leu Arg Arg
900 905 910

Arg Gln Glu Trp Lys Ser Ile Glu Ser Ile Lys Asp Leu Lys Asp Gly

915 920 925
Tyr Met Ser Gln Val Val His Lys Ile Cys Asn Met Val Val Glu His

930 935 940
Lys Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly

945 950 955 960
Arg Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu

965 970 975
Val Asp Lys Leu Asn Tyr Leu Val Val Asp Lys Lys Asn Leu Ser Asn

980 985 990

Glu Pro Gly Gly Leu Tyr Ala Ala Tyr Gln Leu Thr Asn Pro Leu Phe
995 1000 1005

Ser Phe Glu Glu Leu His Arg Tyr Pro Gln Ser Gly Ile Leu Phe Phe
1010 1015 1020
Val Asp Pro Trp Asn Thr Ser Leu Thr Asp Pro Ser Thr Gly Phe Val
1025 1030 1035 1040
Asn Leu Leu Gly Arg Ile Asn Tyr Thr Asn Val Gly Asp Ala Arg Lys
1045 1050 1055
Phe Phe Asp Arg Phe Asn Ala Ile Arg Tyr Asp Gly Lys Gly Asn Ile
1060 1065 1070
Leu Phe Asp Leu Asp Leu Ser Arg Phe Asp Val Arg Val Glu Thr Gln
1075 1080 1085
Arg Lys Leu Trp Thr Leu Thr Thr Phe Gly Ser Arg Ile Ala Lys Ser
1090 1095 1100
Lys Lys Ser Gly Lys Trp Met Val Glu Arg Ile Glu Asn Leu Ser Leu
1105 1110 1115 1120
Cys Phe Leu Glu Leu Phe Glu Gln Phe Asn Ile Gly Tyr Arg Val Glu
1125 1130 1135
Lys Asp Leu Lys Lys Ala Ile Leu Ser Gln Asp Arg Lys Glu Phe Tyr
1140 1145 1150
Val Arg Leu Ile Tyr Leu Phe Asn Leu Met Met Gln Ile Arg Asn Ser
1155 1160 1165
Asp Gly Glu Glu Asp Tyr Ile Leu Ser Pro Ala Leu Asn Glu Lys Asn
1170 1175 1180
Leu Gln Phe Asp Ser Arg Leu Ile Glu Ala Lys Asp Leu Pro Val Asp
1185 1190 1195 1200
Ala Asp Ala Asn Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val
1205 1210 1215
Val Gln Arg Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly
1220 1225 1230
Arg Ala Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu
1235 1240 1245
<210> 79
<211> 867

<212> DNA

<213> Human immunodeficiency virus 1

<400> 79

tttttagatg gaatagataa ggccaagat gaacatgaga aatatcacag taattggaga 60

gcaatggcta gtgattttaa cctgccacct gtagtagcaa aagaatagt agccagctgt 120

gataaatgtc agctaaaagg agaagccatg catggacaag tagactgtag tccaggaata 180

tggcaactag attgtacaca tttagaagga aaagtatatcc tggtagcagt tcatgtagcc 240

agtggatata tagaagcaga agttattcca gcagaaacag ggcaggaaac agcatatattt 300

cttttaaaat tagcaggaag atggccagta aaaacaatac atactgacaa tggcagcaat 360

ttcaccggtg ctacggttag ggccgcctgt tggtagggcg gaatcaagca ggaatttgga 420

attcctaca atcccaaag tcaaggagta gtagaatcta tgaataaaga attaaagaaa 480

attataggac aggtaagaga tcaggctgaa catcttaaga cagcagtaca aatggcagta 540

ttcattccaca attttaaaag aaaagggggg attggggggt acagtgcagg ggaaagaata 600

gtagacataa tagcaacaga catacaaaact aaagaattac aaaaacaaat tacaaaaatt 660

caaaattttc gggtttatta caggacagc agaaatccac tttggaaagg accagcaaag 720

ctcctctgga aaggtgaagg ggcagtagta atacaagata atagtacat aaaagtagtg 780

ccaagaagaa aagcaaagat cattagggat tatggaaaac agatggcagg tgatgattgt 840

gtggcaagta gacaggatga ggattag 867

<210> 80

<211> 288

<212> PRT

<213> Human immunodeficiency virus 1

<400> 80

Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His

1 5 10 15

Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val

20 25 30

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

35 40 45

Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp

50 55 60

Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala
 65 70 75 80
 Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu
 85 90 95
 Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr
 100 105 110
 Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala
 115 120 125
 Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn

 130 135 140
 Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys
 145 150 155 160
 Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val
 165 170 175
 Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly
 180 185 190
 Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile
 195 200 205

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

 275 280 285
 <210> 81
 <211> 25
 <212> PRT
 <213> Artificial Sequence

<220><223> Synthetic

<220><221> MISC_FEATURE

<222> (2)..(5)

<223> At least two Xaa are present; if present, can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (7)..(18)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (20)..(24)

<223> At least three Xaa are present; if present, can be any naturally

occurring amino acid

<400> 81

Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

1 5 10 15

Xaa Xaa His Xaa Xaa Xaa Xaa Xaa His

20 25

<210> 82

<211> 1321

<212> DNA

<213> Mouse mammary tumor virus

<400> 82

atgccgcgcc tgcagcagaa atggttgaac tcccagagagt gtcttacact taggggagaa 60

gcagccaagg ggttgtttcc caccagaaac gacccatctg cgcacacacg gatgagcccc 120

tcaaacaaag acatattcat tctctgctgc aaacttggca tagctctgct ttgcctgggg 180

ctattggggg aagttgcggt tcatgctcgc agggctctca cccttgactc ttttaatagc 240

tcttctgtgc aagattacaa tctaacaat tcggagaact cgaccttcct cctgaggcaa 300

ggaccacagc caacttcctc ttacaagccg catcgattta gtcttcaga aatagaaata 360

agaatgcttg ctaaaaatta tatttttacc aatgagacca atccaatagg tcgattatta 420

attactatgt taagaaatga atcattatct tttagtacta tttttactca aattcagaag 480

ttagaaatgg gaatagaaaa tagaaagaga cgtcagcct cagttgaaga acaggtgcaa 540

ggactaaggg cctcaggcct agaagtaaaa agggggaaga ggagtgcgct tgtcaaaata 600

ggagacaggt ggtggcaacc aggaacttat aggggacctt acatctacag accaacagac 660
gcccccttac cgtatacagg aagatatgac cttaaattttg ataggtgggt cacagtcaat 720
ggctataaag tgttatacag atccctcccc ttctgtgaaa ggctcgccag agctagacct 780
ccttgggtgcg tgttgtctca ggaagaaaaa gacgacatga aacaacaggt acatgattat 840
atztatctag gaacaggaat gaacttttgg agatattata ccaaggaggg ggcagtggct 900
agactattag aacacatttc tgcagatact aatagcatga gttattatga ttagccttta 960
ttggcccaat ctgtgggttc ccagggttca agtaggttca tggtcacaaa ctgtttctaa 1020

aaacaaggat gtgagacaag tggtttctctg gcttggtttg gtatcaaag ttttgatctg 1080
agctctgagt gtctgtttt cctatgttct tttggaatct atccaagtct tatgtaaag 1140
cttatgtaaa ccaaagtata aaagagtgc tttttttga gtaaacttgc aacagtccta 1200
acattcacct ctctgtgttt tgtgtctgtt cgccatcccc tctccgctcg tcaacttacc 1260
ttcactttcc agaggggtccc cccgcagacc ccggtgaccc tcaggttggc cgactgcggc 1320
a 1321

<210> 83
<211> 1082
<212> DNA

<213> Mouse mammary tumor virus
<400> 83

atgccgcgcc tgcagcagaa atggttgaac tcccagagat gtcctacact taggagagaa 60
gcagccaagg ggttgtttcc caccaaggac gacccgtctg cgtgcacgcg gatgagccca 120
tcagacaaag acatactcat tctctgctgc aaacttggca tagctctgct ttgcctgggg 180
ctattggggg aagttgcggt tctgtctgc agggctctca ccttgattc ttttaataac 240
tcttctgtgc aagattacaa tctaaacgat tcggagaact cgaccttct cctggggcaa 300
ggaccacagc caacttctc ttacaagcca caccgacttt gtccttcaga aatagaaata 360

agaatgcttg ctaaaaatta tatttttacc aatgagacca atccaatagg tcgattatta 420
atcatgatgt ttagaaatga atctttgtct ttagcacta tatttactca aattcaaagg 480
ttagaaatgg gaatagaaaa tagaaagaga cgtcaacct cagtgaaga acaggtgcaa 540
ggactaaggg cctcaggcct agaagtaaaa aggggaaaga ggagtgcgct tgtcaaaata 600
ggagacaggt ggtggcaacc agggacttat aggggacctt acatctacag accaacagac 660
gccccgtac catatacagg aagatacga ttaaatattg ataggtgggt cacagtcaac 720
ggctataaag tgttatacag atccctcccc ctctgtgaaa gactcgccag ggctagacct 780

ccttggtgtg tgtaactca ggaagaaaaa gacgacatga aacaacaggt acatgattat 840
 atttatctag gaacaggaat gaacttctgg ggaagatat ttgactacac cgaagaggga 900
 gctatagcaa aaattatata taatatgaaa tatactcatg ggggtcgcac ttgcttcgat 960
 cccttttgaa acatttataa atacaattag gtctaccttg cggttcccaa ggtttaagta 1020
 agttcagggt cacaaactgt tcttaaaaca aggatgtgag acaagtgggt tctgacttg 1080
 gt 1082

<210> 84

<211> 771

<212> DNA

<213> Human immunodeficiency virus 1

<400> 84

ggcaagaaat ccttgatttg tgggtctact acacacaagg cttcttcct gattggcaaa 60
 actacacacc gggaccaggg gtcagatc cactgacctt tggatggtgc tacaagctag 120
 tgccagtga cccaagga gtagaagg ctaaccaaag agaagacaac tgtttgctac 180
 accctatgag cctgcatgga atagaggacg aagacagaga agtattaaag tggcagtttg 240
 acagcagcct agcacgcaga cacatggccc gcgagctaca tccagagtat tacaagact 300
 gctgacacag aaaagacttt ccgctaggac ttccactga ggcgttcag ggggagtgg 360

ctaggcagga ctaggagtg ccaaccctca gatgctgcat ataagcagct gcttttcgcc 420
 tgtactaggt ctctctaggt ggaccagatc tgagcctagg cgctctctgg ctatctaagg 480
 aaccactgc ttaagcctca ataaagcttg ccttgagtgc tctaagtagt gtgtgcccg 540
 ctgttgtgtg actctagtaa ctagagatcc ctacagacaa ctttagtagt gtaaaaaatc 600
 tctagcagtg gcgcccgaac agggaccga aagtgaagc aggaccagag gagatctctc 660
 gacgcaggac tcggcttgct gaaagtgcac tcggcaagag gcgagagcag cggcgactgg 720
 tgagtacgcc gaattttatt ttgactagcg gaggctagaa ggagagagat a 771

<210> 85

<211> 493

<212> DNA

<213> Human immunodeficiency virus 1

<400> 85

atgggtggca agtggtcaga aagtagtgtg gtagaaggc atgtacctt aagacaaggc 60
 agctatagat cttagccgct ttttaaaga aaaggggga ctggaaggc taattcactc 120
 acagagaaga tcagttgaac cagaagaaga tagaaggc catgaagaag aaaacaacag 180

attgttccgt ttgttccgtt ggggactttc caggagacgt ggcctgagtg ataagccgct 240
 ggggactttc cgaagaggcg tgacgggact ttccaaggcg acgtggcctg ggcgggactg 300
 gggagtggcg agccctcaga tgctgcatat aagcagctgc tttctgcctg tactgggtct 360

ctctggtttag accagatctg agcctgggag ctctctggct aactagggaa cccactgctt 420
 aagcctcaat aaagcttgcc ttgagtgtt caagtagtgt gtgccgtct gttgtgtgac 480
 tctggatatct aga 493

<210> 86

<211> 1307

<212> PRT

<213> Acidaminococcus sp. BV3L6

<400> 86

Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr

1 5 10 15

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

20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys

35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln

50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile

65 70 75 80

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

85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly

100 105 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile

115 120 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys

130 135 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg

145 150 155 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg

165 170 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg

180 185 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe

195 200 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn

210 215 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val

225 230 235 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

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

305 310 315 320

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

325 330 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu

340 345 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His

355 360 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr

370 375 380

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

385 390 395 400

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

405 410 415
 Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
 420 425 430
 Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
 435 440 445
 Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys

 450 455 460
 Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
 465 470 475 480
 Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
 485 490 495
 Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
 500 505 510
 Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
 515 520 525

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

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

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

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

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

 1025 1030 1035 1040
 Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala Lys Met Gly
 1045 1050 1055
 Thr Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro Tyr Thr Ser Lys
 1060 1065 1070
 Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe Val Trp Lys Thr Ile
 1075 1080 1085
 Lys Asn His Glu Ser Arg Lys His Phe Leu Glu Gly Phe Asp Phe Leu
 1090 1095 1100

 His Tyr Asp Val Lys Thr Gly Asp Phe Ile Leu His Phe Lys Met Asn
 1105 1110 1115 1120
 Arg Asn Leu Ser Phe Gln Arg Gly Leu Pro Gly Phe Met Pro Ala Trp
 1125 1130 1135
 Asp Ile Val Phe Glu Lys Asn Glu Thr Gln Phe Asp Ala Lys Gly Thr
 1140 1145 1150
 Pro Phe Ile Ala Gly Lys Arg Ile Val Pro Val Ile Glu Asn His Arg
 1155 1160 1165

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

1170 1175 1180
 Leu Leu Glu Glu Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu
 1185 1190 1195 1200
 Pro Lys Leu Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met Val
 1205 1210 1215
 Ala Leu Ile Arg Ser Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr
 1220 1225 1230
 Gly Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys
 1235 1240 1245

Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp Ala
 1250 1255 1260
 Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu Asn His
 1265 1270 1275 1280
 Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile Ser Asn Gln
 1285 1290 1295
 Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
 1300 1305

<210> 87
 <211> 1246
 <212> PRT

<213> Porphyromonas macacae
 <400> 87

Met Lys Thr Gln His Phe Phe Glu Asp Phe Thr Ser Leu Tyr Ser Leu
 1 5 10 15
 Ser Lys Thr Ile Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu
 20 25 30
 Asn Ile Lys Lys Asn Gly Leu Ile Arg Arg Asp Glu Gln Arg Leu Asp
 35 40 45
 Asp Tyr Glu Lys Leu Lys Lys Val Ile Asp Glu Tyr His Glu Asp Phe
 50 55 60

Ile	Ala	Asn	Ile	Leu	Ser	Ser	Phe	Ser	Phe	Ser	Glu	Glu	Ile	Leu	Gln
65				70				75				80			
Ser	Tyr	Ile	Gln	Asn	Leu	Ser	Glu	Ser	Glu	Ala	Arg	Ala	Lys	Ile	Glu
85				90				95							
Lys	Thr	Met	Arg	Asp	Thr	Leu	Ala	Lys	Ala	Phe	Ser	Glu	Asp	Glu	Arg
100				105				110							
Tyr	Lys	Ser	Ile	Phe	Lys	Lys	Glu	Leu	Val	Lys	Lys	Asp	Ile	Pro	Val
115				120				125							
Trp	Cys	Pro	Ala	Tyr	Lys	Ser	Leu	Cys	Lys	Lys	Phe	Asp	Asn	Phe	Thr
130				135				140							
Thr	Ser	Leu	Val	Pro	Phe	His	Glu	Asn	Arg	Lys	Asn	Leu	Tyr	Thr	Ser
145				150				155				160			
Asn	Glu	Ile	Thr	Ala	Ser	Ile	Pro	Tyr	Arg	Ile	Val	His	Val	Asn	Leu
165				170				175							
Pro	Lys	Phe	Ile	Gln	Asn	Ile	Glu	Ala	Leu	Cys	Glu	Leu	Gln	Lys	Lys
180				185				190							
Met	Gly	Ala	Asp	Leu	Tyr	Leu	Glu	Met	Met	Glu	Asn	Leu	Arg	Asn	Val
195				200				205							
Trp	Pro	Ser	Phe	Val	Lys	Thr	Pro	Asp	Asp	Leu	Cys	Asn	Leu	Lys	Thr
210				215				220							
Tyr	Asn	His	Leu	Met	Val	Gln	Ser	Ser	Ile	Ser	Glu	Tyr	Asn	Arg	Phe
225				230				235				240			
Val	Gly	Gly	Tyr	Ser	Thr	Glu	Asp	Gly	Thr	Lys	His	Gln	Gly	Ile	Asn
245				250				255							
Glu	Trp	Ile	Asn	Ile	Tyr	Arg	Gln	Arg	Asn	Lys	Glu	Met	Arg	Leu	Pro
260				265				270							
Gly	Leu	Val	Phe	Leu	His	Lys	Gln	Ile	Leu	Ala	Lys	Val	Asp	Ser	Ser
275				280				285							
Ser	Phe	Ile	Ser	Asp	Thr	Leu	Glu	Asn	Asp	Asp	Gln	Val	Phe	Cys	Val
290				295				300							
Leu	Arg	Gln	Phe	Arg	Lys	Leu	Phe	Trp	Asn	Thr	Val	Ser	Ser	Lys	Glu
305				310				315				320			

Asp Asp Ala Ala Ser Leu Lys Asp Leu Phe Cys Gly Leu Ser Gly Tyr
 325 330 335
 Asp Pro Glu Ala Ile Tyr Val Ser Asp Ala His Leu Ala Thr Ile Ser
 340 345 350

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

 420 425 430
 Tyr Leu Val Ser Asp Gly Glu Val Ile Leu Tyr Glu Glu Gly Ser Asn
 435 440 445
 Ile Trp Asp Glu Val Leu Ile Ala Phe Arg Asp Leu Gln Val Ile Leu
 450 455 460
 Asp Lys Asp Phe Thr Glu Lys Lys Leu Gly Lys Asp Glu Glu Ala Val
 465 470 475 480
 Ser Val Ile Lys Lys Ala Leu Asp Ser Ala Leu Arg Leu Arg Lys Phe
 485 490 495

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

565	570	575
Asn Gly Tyr Tyr Tyr Leu Gly Ile Met Thr Pro Lys Gly Lys Asn Leu		
580	585	590
Phe Lys Thr Leu Pro Lys Leu Gly Ala Glu Glu Met Phe Tyr Glu Lys		
595	600	605
Met Glu Tyr Lys Gln Ile Ala Glu Pro Met Leu Met Leu Pro Lys Val		
610	615	620
Phe Phe Pro Lys Lys Thr Lys Pro Ala Phe Ala Pro Asp Gln Ser Val		
625	630	635
		640
Val Asp Ile Tyr Asn Lys Lys Thr Phe Lys Thr Gly Gln Lys Gly Phe		
645	650	655
Asn Lys Lys Asp Leu Tyr Arg Leu Ile Asp Phe Tyr Lys Glu Ala Leu		
660	665	670
Thr Val His Glu Trp Lys Leu Phe Asn Phe Ser Phe Ser Pro Thr Glu		
675	680	685
Gln Tyr Arg Asn Ile Gly Glu Phe Phe Asp Glu Val Arg Glu Gln Ala		
690	695	700
Tyr Lys Val Ser Met Val Asn Val Pro Ala Ser Tyr Ile Asp Glu Ala		
705	710	715
		720
Val Glu Asn Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe		
725	730	735
Ser Pro Tyr Ser Lys Gly Ile Pro Asn Leu His Thr Leu Tyr Trp Lys		
740	745	750
Ala Leu Phe Ser Glu Gln Asn Gln Ser Arg Val Tyr Lys Leu Cys Gly		
755	760	765
Gly Gly Glu Leu Phe Tyr Arg Lys Ala Ser Leu His Met Gln Asp Thr		
770	775	780
Thr Val His Pro Lys Gly Ile Ser Ile His Lys Lys Asn Leu Asn Lys		
785	790	795
		800
Lys Gly Glu Thr Ser Leu Phe Asn Tyr Asp Leu Val Lys Asp Lys Arg		
805	810	815
Phe Thr Glu Asp Lys Phe Phe Phe His Val Pro Ile Ser Ile Asn Tyr		

820	825	830	
Lys Asn Lys Lys Ile Thr Asn Val Asn Gln Met Val Arg Asp Tyr Ile			
835	840	845	
Ala Gln Asn Asp Asp Leu Gln Ile Ile Gly Ile Asp Arg Gly Glu Arg			
850	855	860	
Asn Leu Leu Tyr Ile Ser Arg Ile Asp Thr Arg Gly Asn Leu Leu Glu			
865	870	875	880
Gln Phe Ser Leu Asn Val Ile Glu Ser Asp Lys Gly Asp Leu Arg Thr			
885	890	895	
Asp Tyr Gln Lys Ile Leu Gly Asp Arg Glu Gln Glu Arg Leu Arg Arg			
900	905	910	
Arg Gln Glu Trp Lys Ser Ile Glu Ser Ile Lys Asp Leu Lys Asp Gly			
915	920	925	
Tyr Met Ser Gln Val Val His Lys Ile Cys Asn Met Val Val Glu His			
930	935	940	
Lys Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly			
945	950	955	960
Arg Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu			
965	970	975	
Val Asp Lys Leu Asn Tyr Leu Val Val Asp Lys Lys Asn Leu Ser Asn			
980	985	990	
Glu Pro Gly Gly Leu Tyr Ala Ala Tyr Gln Leu Thr Asn Pro Leu Phe			
995	1000	1005	
Ser Phe Glu Glu Leu His Arg Tyr Pro Gln Ser Gly Ile Leu Phe Phe			
1010	1015	1020	
Val Asp Pro Trp Asn Thr Ser Leu Thr Asp Pro Ser Thr Gly Phe Val			
1025	1030	1035	1040
Asn Leu Leu Gly Arg Ile Asn Tyr Thr Asn Val Gly Asp Ala Arg Lys			
1045	1050	1055	
Phe Phe Asp Arg Phe Asn Ala Ile Arg Tyr Asp Gly Lys Gly Asn Ile			
1060	1065	1070	

Leu Phe Asp Leu Asp Leu Ser Arg Phe Asp Val Arg Val Glu Thr Gln
 1075 1080 1085
 Arg Lys Leu Trp Thr Leu Thr Thr Phe Gly Ser Arg Ile Ala Lys Ser
 1090 1095 1100
 Lys Lys Ser Gly Lys Trp Met Val Glu Arg Ile Glu Asn Leu Ser Leu
 1105 1110 1115 1120
 Cys Phe Leu Glu Leu Phe Glu Gln Phe Asn Ile Gly Tyr Arg Val Glu
 1125 1130 1135
 Lys Asp Leu Lys Lys Ala Ile Leu Ser Gln Asp Arg Lys Glu Phe Tyr
 1140 1145 1150
 Val Arg Leu Ile Tyr Leu Phe Asn Leu Met Met Gln Ile Arg Asn Ser
 1155 1160 1165
 Asp Gly Glu Glu Asp Tyr Ile Leu Ser Pro Ala Leu Asn Glu Lys Asn
 1170 1175 1180
 Leu Gln Phe Asp Ser Arg Leu Ile Glu Ala Lys Asp Leu Pro Val Asp
 1185 1190 1195 1200
 Ala Asp Ala Asn Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val
 1205 1210 1215
 Val Gln Arg Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly
 1220 1225 1230
 Arg Ala Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu
 1235 1240 1245
 <210> 88
 <211> 1282
 <212> PRT
 <213> Eubacterium eligens
 <400> 88
 Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Ile
 1 5 10 15
 Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr
 20 25 30
 Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg

35 40 45
 Gln Glu Lys Ser Thr Glu Leu Lys Asn Ile Met Asp Asp Tyr Tyr Arg
 50 55 60
 Glu Tyr Ile Asp Lys Ser Leu Ser Gly Val Thr Asp Leu Asp Phe Thr
 65 70 75 80
 Leu Leu Phe Glu Leu Met Asn Leu Val Gln Ser Ser Pro Ser Lys Asp
 85 90 95

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

 165 170 175
 Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val
 180 185 190
 His Glu Asn Ser Leu Ile Phe Leu Ala Asn Met Thr Ser Tyr Lys Lys
 195 200 205
 Ile Ser Glu Lys Ala Leu Asp Glu Ile Glu Val Ile Glu Lys Asn Asn
 210 215 220
 Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp
 225 230 235 240

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

290 295 300
 Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu

 305 310 315 320
 Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp

 325 330 335
 Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr

 340 345 350
 Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val

 355 360 365
 Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu

 370 375 380

 Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
 385 390 395 400
 Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn

 405 410 415
 Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn

 420 425 430
 Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Asp His Ile

 435 440 445
 Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Met Lys Lys Arg Leu

 450 455 460
 Asp Met Tyr Met Asn Met Tyr His Trp Ala Lys Ala Phe Ile Val Asp
 465 470 475 480
 Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile

 485 490 495
 Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn

 500 505 510
 Tyr Val Thr Gln Lys Pro Tyr Asn Ser Lys Lys Ile Lys Leu Asn Phe

 515 520 525

 Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp
 530 535 540

Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
 545 550 555 560
 Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
 565 570 575
 Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
 580 585 590
 Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly

 595 600 605
 Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala
 610 615 620
 His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
 625 630 635 640
 Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
 645 650 655
 Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Ser Asp Ile
 660 665 670

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

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

Met Tyr Asn Gly Tyr Ile Lys Glu Ser Asp Leu Ser Glu Ala Ala Lys
805 810 815

Glu Tyr Leu Asp Lys Val Glu Val Arg Thr Ala Gln Lys Asp Ile Val
820 825 830

Lys Asp Tyr Arg Tyr Thr Val Asp Lys Tyr Phe Ile His Thr Pro Ile
835 840 845

Thr Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Val
850 855 860

Val Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp
865 870 875 880

Arg Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly
885 890 895

Asn Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr
900 905 910

Lys Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys
915 920 925

Asn Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile
930 935 940

Ser Gly Val Val His Glu Ile Ala Met Leu Ile Val Glu Tyr Asn Ala
945 950 955 960

Ile Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe
965 970 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn
980 985 990

Lys Leu Asn Tyr Phe Ala Ser Lys Glu Lys Ser Val Asp Glu Pro Gly
995 1000 1005

Gly Leu Leu Lys Gly Tyr Gln Leu Thr Tyr Val Pro Asp Asn Ile Lys
1010 1015 1020

Asn Leu Gly Lys Gln Cys Gly Val Ile Phe Tyr Val Pro Ala Ala Phe
1025 1030 1035 1040

Thr Ser Lys Ile Asp Pro Ser Thr Gly Phe Ile Ser Ala Phe Asn Phe

1045 1050 1055
 Lys Ser Ile Ser Thr Asn Ala Ser Arg Lys Gln Phe Phe Met Gln Phe
 1060 1065 1070
 Asp Glu Ile Arg Tyr Cys Ala Glu Lys Asp Met Phe Ser Phe Gly Phe
 1075 1080 1085
 Asp Tyr Asn Asn Phe Asp Thr Tyr Asn Ile Thr Met Gly Lys Thr Gln
 1090 1095 1100

 Trp Thr Val Tyr Thr Asn Gly Glu Arg Leu Gln Ser Glu Phe Asn Asn
 1105 1110 1115 1120
 Ala Arg Arg Thr Gly Lys Thr Lys Ser Ile Asn Leu Thr Glu Thr Ile
 1125 1130 1135
 Lys Leu Leu Leu Glu Asp Asn Glu Ile Asn Tyr Ala Asp Gly His Asp
 1140 1145 1150
 Ile Arg Ile Asp Met Glu Lys Met Asp Glu Asp Lys Lys Ser Glu Phe
 1155 1160 1165
 Phe Ala Gln Leu Leu Ser Leu Tyr Lys Leu Thr Val Gln Met Arg Asn

 1170 1175 1180
 Ser Tyr Thr Glu Ala Glu Glu Gln Glu Asn Gly Ile Ser Tyr Asp Lys
 1185 1190 1195 1200
 Ile Ile Ser Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp
 1205 1210 1215
 Asn Tyr Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala
 1220 1225 1230
 Asp Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val
 1235 1240 1245

 Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn Cys
 1250 1255 1260
 Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn Lys Arg
 1265 1270 1275 1280
 Tyr Glu

<210> 89

<211> 1263

<212> PRT

<213> *Leptospira inadai*

<400> 89

Met Glu Asp Tyr Ser Gly Phe Val Asn Ile Tyr Ser Ile Gln Lys Thr

1 5 10 15

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

20 25 30

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

35 40 45

Ala Val Lys Lys Ile Ile Asp Lys Tyr His Arg Ala Tyr Ile Glu Glu

50 55 60

Val Phe Asp Ser Val Leu His Gln Lys Lys Lys Lys Asp Lys Thr Arg

65 70 75 80

Phe Ser Thr Gln Phe Ile Lys Glu Ile Lys Glu Phe Ser Glu Leu Tyr

85 90 95

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

100 105 110

Ser Glu Lys Leu Arg Lys Met Leu Val Gly Ala Phe Lys Gly Glu Phe

115 120 125

Ser Glu Glu Val Ala Glu Lys Tyr Lys Asn Leu Phe Ser Lys Glu Leu

130 135 140

Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg Lys

145 150 155 160

Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His

165 170 175

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

180 185 190

Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp Asn Leu

195 200 205

Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro Trp Ser

210 215 220

Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys Leu Thr
225 230 235 240

Glu Tyr Phe Ser Ile Asp Gly Phe Val Asn Val Leu Asn Gln Lys Gly
245 250 255

Ile Asp Ala Tyr Asn Thr Ile Leu Gly Gly Lys Ser Glu Glu Ser Gly
260 265 270

Glu Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Arg Gln Lys
275 280 285

Asn Asn Ile Asp Arg Lys Asn Leu Pro Asn Val Lys Ile Leu Phe Lys
290 295 300

Gln Ile Leu Gly Asp Arg Glu Thr Lys Ser Phe Ile Pro Glu Ala Phe

305 310 315 320

Pro Asp Asp Gln Ser Val Leu Asn Ser Ile Thr Glu Phe Ala Lys Tyr
325 330 335

Leu Lys Leu Asp Lys Lys Lys Lys Ser Ile Ile Ala Glu Leu Lys Lys
340 345 350

Phe Leu Ser Ser Phe Asn Arg Tyr Glu Leu Asp Gly Ile Tyr Leu Ala
355 360 365

Asn Asp Asn Ser Leu Ala Ser Ile Ser Thr Phe Leu Phe Asp Asp Trp
370 375 380

Ser Phe Ile Lys Lys Ser Val Ser Phe Lys Tyr Asp Glu Ser Val Gly
385 390 395 400

Asp Pro Lys Lys Lys Ile Lys Ser Pro Leu Lys Tyr Glu Lys Glu Lys
405 410 415

Glu Lys Trp Leu Lys Gln Lys Tyr Tyr Thr Ile Ser Phe Leu Asn Asp
420 425 430

Ala Ile Glu Ser Tyr Ser Lys Ser Gln Asp Glu Lys Arg Val Lys Ile
435 440 445

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

450 455 460

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

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

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

 595 600 605
 Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro Asn Glu
 610 615 620
 Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro His Met
 625 630 635 640
 Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile Tyr Asn
 645 650 655
 Pro Ser Lys Ser Ile Leu Lys Ile Arg Glu Ala Lys Ser Phe Lys Glu
 660 665 670

 Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp Phe Tyr
 675 680 685
 Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp Phe Lys
 690 695 700
 Phe Ser Lys Thr Ser Ser Tyr Glu Asn Ile Ser Glu Phe Tyr Arg Glu
 705 710 715 720
 Val Glu Arg Gln Gly Tyr Asn Leu Asp Phe Lys Lys Val Ser Lys Phe

725 730 735
Tyr Ile Asp Ser Leu Val Glu Asp Gly Lys Leu Tyr Leu Phe Gln Ile

740 745 750
Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn Leu His

755 760 765
Thr Ile Tyr Phe Arg Ser Leu Phe Ser Lys Glu Asn Leu Lys Asp Val

770 775 780
Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys Lys Ser

785 790 795 800
Ile Asn Tyr Asp Glu Lys Lys Lys Arg Glu Gly His His Pro Glu Leu

805 810 815

Phe Glu Lys Leu Lys Tyr Pro Ile Leu Lys Asp Lys Arg Tyr Ser Glu
820 825 830

Asp Lys Phe Gln Phe His Leu Pro Ile Ser Leu Asn Phe Lys Ser Lys
835 840 845

Glu Arg Leu Asn Phe Asn Leu Lys Val Asn Glu Phe Leu Lys Arg Asn
850 855 860

Lys Asp Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu
865 870 875 880

Tyr Leu Val Met Ile Asn Gln Lys Gly Glu Ile Leu Lys Gln Thr Leu

885 890 895
Leu Asp Ser Met Gln Ser Gly Lys Gly Arg Pro Glu Ile Asn Tyr Lys

900 905 910
Glu Lys Leu Gln Glu Lys Glu Ile Glu Arg Asp Lys Ala Arg Lys Ser

915 920 925
Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser

930 935 940
Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn Ala Ile

945 950 955 960

Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Arg Gly Arg Gln Lys
965 970 975

Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys
980 985 990

Leu Asn Phe Leu Val Phe Lys Glu Asn Lys Pro Thr Glu Pro Gly Gly
995 1000 1005

Val Leu Lys Ala Tyr Gln Leu Thr Asp Glu Phe Gln Ser Phe Glu Lys
1010 1015 1020

Leu Ser Lys Gln Thr Gly Phe Leu Phe Tyr Val Pro Ser Trp Asn Thr

1025 1030 1035 1040

Ser Lys Ile Asp Pro Arg Thr Gly Phe Ile Asp Phe Leu His Pro Ala
1045 1050 1055

Tyr Glu Asn Ile Glu Lys Ala Lys Gln Trp Ile Asn Lys Phe Asp Ser
1060 1065 1070

Ile Arg Phe Asn Ser Lys Met Asp Trp Phe Glu Phe Thr Ala Asp Thr
1075 1080 1085

Arg Lys Phe Ser Glu Asn Leu Met Leu Gly Lys Asn Arg Val Trp Val
1090 1095 1100

Ile Cys Thr Thr Asn Val Glu Arg Tyr Phe Thr Ser Lys Thr Ala Asn
1105 1110 1115 1120

Ser Ser Ile Gln Tyr Asn Ser Ile Gln Ile Thr Glu Lys Leu Lys Glu
1125 1130 1135

Leu Phe Val Asp Ile Pro Phe Ser Asn Gly Gln Asp Leu Lys Pro Glu
1140 1145 1150

Ile Leu Arg Lys Asn Asp Ala Val Phe Phe Lys Ser Leu Leu Phe Tyr
1155 1160 1165

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

1170 1175 1180

Glu Glu Lys Asp Phe Ile Leu Ser Pro Val Val Asp Ser Lys Gly Arg
1185 1190 1195 1200

Phe Phe Asn Ser Leu Glu Ala Ser Asp Asp Glu Pro Lys Asp Ala Asp
1205 1210 1215

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Met Asn Leu Leu
1220 1225 1230

Val Leu Asn Glu Thr Lys Glu Glu Asn Leu Ser Arg Pro Lys Trp Lys

1235

1240

1245

Ile Lys Asn Lys Asp Trp Leu Glu Phe Val Trp Glu Arg Asn Arg

1250

1255

1260

<210> 90

<211> 1206

<212> PRT

<213> Lachnospiraceae bacterium MA2020

<400> 90

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

1

5

10

15

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

20

25

30

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

35

40

45

Val Lys Gly Ile Leu Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala

50

55

60

Leu Asp Asn Cys Thr Leu Pro Ser Leu Lys Ile Ala Ala Glu Ile Tyr

65

70

75

80

Leu Lys Asn Gln Lys Glu Val Ser Asp Arg Glu Asp Phe Asn Lys Thr

85

90

95

Gln Asp Leu Leu Arg Lys Glu Val Val Glu Lys Leu Lys Ala His Glu

100

105

110

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

115

120

125

Leu Pro Ser Ile Ser Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Arg

130

135

140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu

145

150

155

160

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

165

170

175

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

180 185 190
Lys Thr Ala Gly Ile Leu Ala Asp Gly Leu Gly Glu Glu Glu Gln Asp
195 200 205
Ser Leu Phe Ile Val Glu Thr Phe Asn Lys Thr Leu Thr Gln Asp Gly
210 215 220
Ile Asp Thr Tyr Asn Ser Gln Val Gly Lys Ile Asn Ser Ser Ile Asn
225 230 235 240
Leu Tyr Asn Gln Lys Asn Gln Lys Ala Asn Gly Phe Arg Lys Ile Pro
245 250 255

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

325 330 335
Trp Arg Thr Phe Asp Asp Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
340 345 350
Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
355 360 365
Leu Lys Lys Asn Lys Ser Tyr Ser Leu Glu His Leu Cys Asn Leu Ser
370 375 380
Glu Asp Ser Cys Asn Leu Ile Glu Asn Tyr Ile His Gln Ile Ser Asp
385 390 395 400

Asp Ile Glu Asn Ile Ile Ile Asn Asn Glu Thr Phe Leu Arg Ile Val
405 410 415
Ile Asn Glu His Asp Arg Ser Arg Lys Leu Ala Lys Asn Arg Lys Ala

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

 465 470 475 480
 Asp Ser Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
 485 490 495
 Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Arg Ser Thr Leu Leu Asn
 500 505 510
 Gly Trp Asp Arg Asn Lys Glu Thr Asp Asn Leu Gly Val Leu Leu Leu
 515 520 525
 Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ala Asn Lys
 530 535 540

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

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

675 680 685
 Ser Met Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
 690 695 700
 Met Leu Phe Asp Gln Arg Asn Ile Asp Asp Val Val Tyr Lys Leu Asn
 705 710 715 720
 Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ser Glu Asp Glu
 725 730 735
 Leu Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Pro Asn
 740 745 750
 Arg Ala Arg Thr Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
 755 760 765
 Asp Lys Arg Tyr Ser Lys Asp Lys Phe Thr Leu His Ile Pro Ile Thr
 770 775 780
 Met Asn Phe Gly Val Asp Glu Val Lys Arg Phe Asn Asp Ala Val Asn
 785 790 795 800
 Ser Ala Ile Arg Ile Asp Glu Asn Val Asn Val Ile Gly Ile Asp Arg
 805 810 815
 Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Ser Lys Gly Asn
 820 825 830
 835 840 845
 Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp
 850 855 860
 Ile Glu Thr Asp Tyr His Ala Leu Leu Asp Glu Arg Glu Gly Gly Arg
 865 870 875 880
 Asp Lys Ala Arg Lys Asp Trp Asn Thr Val Glu Asn Ile Arg Asp Leu
 Lys Ala Gly Tyr Leu Ser Gln Val Val Asn Val Val Ala Lys Leu Val
 885 890 895
 Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
 900 905 910
 Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
 915 920 925

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

 Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
 980 985 990
 Thr Gly Phe Ala Asn Leu Phe Tyr Met Lys Cys Glu Asn Val Glu Lys
 995 1000 1005
 Ser Lys Arg Phe Phe Asp Gly Phe Asp Phe Ile Arg Phe Asn Ala Leu
 1010 1015 1020
 Glu Asn Val Phe Glu Phe Gly Phe Asp Tyr Arg Ser Phe Thr Gln Arg
 1025 1030 1035 1040
 Ala Cys Gly Ile Asn Ser Lys Trp Thr Val Cys Thr Asn Gly Glu Arg

 1045 1050 1055
 Ile Ile Lys Tyr Arg Asn Pro Asp Lys Asn Asn Met Phe Asp Glu Lys
 1060 1065 1070
 Val Val Val Val Thr Asp Glu Met Lys Asn Leu Phe Glu Gln Tyr Lys
 1075 1080 1085
 Ile Pro Tyr Glu Asp Gly Arg Asn Val Lys Asp Met Ile Ile Ser Asn
 1090 1095 1100
 Glu Glu Ala Glu Phe Tyr Arg Arg Leu Tyr Arg Leu Leu Gln Gln Thr
 1105 1110 1115 1120

 Leu Gln Met Arg Asn Ser Thr Ser Asp Gly Thr Arg Asp Tyr Ile Ile
 1125 1130 1135
 Ser Pro Val Lys Asn Lys Arg Glu Ala Tyr Phe Asn Ser Glu Leu Ser
 1140 1145 1150
 Asp Gly Ser Val Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile
 1155 1160 1165
 Ala Arg Lys Gly Leu Trp Val Leu Glu Gln Ile Arg Gln Lys Ser Glu
 1170 1175 1180

Gly Glu Lys Ile Asn Leu Ala Met Thr Asn Ala Glu Trp Leu Glu Tyr

1185 1190 1195 1200

Ala Gln Thr His Leu Leu

1205

<210> 91

<211> 1300

<212> PRT

<213> Francisella tularensis

<400> 91

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr

1 5 10 15

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

20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys

35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu

50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser

180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu

195 200 205

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

210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu

225 230 235 240

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

245 250 255

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

260 265 270

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

275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile

290 295 300

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

305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser

325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met

340 345 350

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

355 360 365

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

370 375 380

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

385 390 395 400

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

405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn

420 425 430
 Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
 435 440 445
 Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
 450 455 460
 Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

 465 470 475 480
 Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
 485 490 495
 Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
 500 505 510
 Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
 515 520 525
 Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
 530 535 540

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

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

675 680 685
 Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
 690 695 700
 Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
 705 710 715 720
 Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
 725 730 735
 Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
 740 745 750
 Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

 755 760 765
 Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
 770 775 780
 Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
 785 790 795 800
 Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
 805 810 815
 Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
 820 825 830

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

 900 905 910
 Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
 915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val Tyr
1010 1015 1020

Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu Val Phe
1025 1030 1035 1040

Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg Ala Tyr Gln
1045 1050 1055

Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly Lys Gln Thr Gly
1060 1065 1070

Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser Lys Ile Cys Pro Val
1075 1080 1085

Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys Tyr Glu Ser Val Ser Lys
1090 1095 1100

Ser Gln Glu Phe Phe Ser Lys Phe Asp Lys Ile Cys Tyr Asn Leu Asp
1105 1110 1115 1120

Lys Gly Tyr Phe Glu Phe Ser Phe Asp Tyr Lys Asn Phe Gly Asp Lys
1125 1130 1135

Ala Ala Lys Gly Lys Trp Thr Ile Ala Ser Phe Gly Ser Arg Leu Ile
1140 1145 1150

Asn Phe Arg Asn Ser Asp Lys Asn His Asn Trp Asp Thr Arg Glu Val
1155 1160 1165

Tyr Pro Thr Lys Glu Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu
1170 1175 1180

Tyr Gly His Gly Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp

1185 1190 1195 1200

Lys Lys Phe Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln

1205 1210 1215

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

1220 1225 1230

Val Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly Leu

1250 1255 1260

Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu Gly Lys

1265 1270 1275 1280

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

1285 1290 1295

Asn Arg Asn Asn

1300

<210> 92

<211> 1260

<212> PRT

<213> Porphyromonas crevioricanis

<400> 92

Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr

1 5 10 15

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

20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg

35 40 45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser

50 55 60

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

65 70 75 80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu

85 90 95

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

100 105 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro

210 215 220

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

225 230 235 240

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

245 250 255

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

260 265 270

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

275 280 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu

290 295 300

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

305 310 315 320

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

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

 370 375 380
 Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
 385 390 395 400
 Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
 405 410 415
 Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
 420 425 430
 Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
 435 440 445

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

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

580	585	590
Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser		
595	600	605
Phe Glu Asn Lys Met Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe		
610	615	620
Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro		
625	630	635
Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro		
645	650	655
Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr		
660	665	670
Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser		
675	680	685
Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp		
690	695	700
Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp		
705	710	715
Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr		
725	730	735
Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys		
740	745	750
Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr		
755	760	765
Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys		
770	775	780
Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn		
785	790	795
Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg		
805	810	815
Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp		
820	825	830

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

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

 945 950 955 960
 Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
 965 970 975
 Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
 980 985 990
 Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu
 995 1000 1005
 Leu Arg Ala Tyr Gln Phe Thr Ala Pro Phe Lys Ser Phe Lys Glu Met
 1010 1015 1020

 Gly Lys Gln Asn Gly Phe Leu Phe Tyr Ile Pro Ala Trp Asn Thr Ser
 1025 1030 1035 1040
 Asn Ile Asp Pro Thr Thr Gly Phe Val Asn Leu Phe His Val Gln Tyr
 1045 1050 1055
 Glu Asn Val Asp Lys Ala Lys Ser Phe Phe Gln Lys Phe Asp Ser Ile
 1060 1065 1070
 Ser Tyr Asn Pro Lys Lys Asp Trp Phe Glu Phe Ala Phe Asp Tyr Lys
 1075 1080 1085

Asn Phe Thr Lys Lys Ala Glu Gly Ser Arg Ser Met Trp Ile Leu Cys

1090	1095	1100	
Thr His Gly Ser Arg Ile Lys Asn Phe Arg Asn Ser Gln Lys Asn Gly			
1105	1110	1115	1120
Gln Trp Asp Ser Glu Glu Phe Ala Leu Thr Glu Ala Phe Lys Ser Leu			
	1125	1130	1135
Phe Val Arg Tyr Glu Ile Asp Tyr Thr Ala Asp Leu Lys Thr Ala Ile			
	1140	1145	1150
Val Asp Glu Lys Gln Lys Asp Phe Phe Val Asp Leu Leu Lys Leu Phe			
1155	1160	1165	

Lys Leu Thr Val Gln Met Arg Asn Ser Trp Lys Glu Lys Asp Leu Asp			
1170	1175	1180	
Tyr Leu Ile Ser Pro Val Ala Gly Ala Asp Gly Arg Phe Phe Asp Thr			
1185	1190	1195	1200
Arg Glu Gly Asn Lys Ser Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala			
	1205	1210	1215
Tyr Asn Ile Ala Leu Lys Gly Leu Trp Ala Leu Arg Gln Ile Arg Gln			
	1220	1225	1230
Thr Ser Glu Gly Gly Lys Leu Lys Leu Ala Ile Ser Asn Lys Glu Trp			

1235	1240	1245	
Leu Gln Phe Val Gln Glu Arg Ser Tyr Glu Lys Asp			

1250	1255	1260	
------	------	------	--

<210> 93

<400> 93

000

<210> 94

<400> 94

000

<210> 95

<400> 95

000

<210> 96

<400> 96
 000
 <210> 97
 <400> 97
 000
 <210> 98
 <400> 98
 000
 <210> 99
 <400> 99
 000
 <210> 100
 <211> 1179
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 100

gacaagacat ccttgatttg tgggtctata acacacaagg cttcttcctt gattggcaaa 60

actacacacc gggaccaggg accagatacc cactgacctt tggatggtgc ttcaagctag 120
 tgccagtga cccaaggga gtagaagagg ccaatacagg ggaaaacaac tgtttgctcc 180
 accctatgag ccagcatgga atggaagatg accatagaga agtattaaag tggaaagttg 240
 acagtatgct agcacgcaga cacctggccc gcgagctaca tccggagtac tacaaaaact 300
 gctgacatgg agggactttc cgctgggact ttccattggg gcgttccagg aggtgtggtc 360
 tgggcgggac aaggagatgg tcaaccctca gatgctgcat ataagcagct gcttttcgct 420
 tgtactgggt ctcttttagt agaccagatc tgaacctggg agctctctgg ctacctgagg 480

aaccactgc ttaagcctca ataaagcttg ccttgagtgc tctaagtagt gtgtgcccgt 540
 ctgttgtgtg actctggtta ctagagatcc ctacagacct tttgtagtg tggaaaatct 600
 ctagcagatg attgaacaag atggattgca cgcaggttct ccggccgctt gggatggagag 660
 gctattcggc tatgactggg cacaacatgg gtggcaagtg gtcagaaaagt agtgtggtta 720
 gaaggcatgt acctttaaga caaggcagct atagatctta gccgcttttt aaaagaaaag 780
 gggggactgg aagggtta tcactcacag agaagatcag ttgaaccaga agaagataga 840
 agaggccatg aagaagaaaa caacagattg ttccgtttgt tccgttgggg actttccagg 900

agacgtggcc tgagtgataa gccgctgggg actttccgaa gaggcgtgac gggactttcc 960
aaggcgacgt ggcttgggcg ggactgggga gtggcgagcc ctcagatgct gcatataagc 1020
agctgttttc tgcctgtact gggctctctt ggtagacca gatctgagcc tgggagctct 1080
ctggctaact aggaaccca ctgcttaagc ctcaataaag cttgccttga gtgcttcaag 1140
tagtgtgtgc ccgtctgttg tgtgactctg gtatctaga 1179

<210> 101

<211> 224

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 101

gacaagacat cttgatttg tgggtctata acacacaagg cttcttcct gattggcaaa 60

actacacacc atgattgaac aagatggatt gcacgcaggt tctccggccg cttgggtgga 120

gaggctattc ggctatgact gggcacaact taagcctcaa taaagcttgc cttgagtgtc 180

tcaagtagtg tgtgcccgtc tgttgtgtga cttggtatc taga 224

<210> 102

<400> 102

000

<210> 103

<400> 103

000

<210> 104

<400> 104

000

<210> 105

<400> 105

000

<210> 106

<400> 106

000

<210> 107

<400> 107

000

<210> 108

<400>	108
000	
<210>	109
<400>	109
000	
<210>	110
<400>	110
000	
<210>	111
<400>	111
000	
<210>	112
<400>	112
000	
<210>	113
<400>	113
000	
<210>	114
<400>	114
000	
<210>	115
<400>	115
000	
<210>	116
<400>	116
000	
<210>	117
<400>	117
000	
<210>	118
<400>	118
000	
<210>	119
<400>	119

000
 <210> 120
 <400> 120
 000
 <210> 121
 <400> 121
 000
 <210> 122
 <400> 122
 000
 <210> 123
 <400> 123
 000
 <210
 > 124
 <400> 124
 000
 <210> 125
 <400> 125
 000
 <210> 126
 <400> 126
 000
 <210> 127
 <400> 127
 000
 <210> 128
 <400> 128
 000
 <210> 129
 <400> 129
 000
 <210> 130
 <400> 130
 000

<210>	131
<400>	131
000	
<210>	132
<400>	132
000	
<210>	133
<400>	133
000	
<210>	134
<400>	134
000	
<210>	135
<400>	135
000	
<210>	136
<400>	136
000	
<210>	137
<400>	137
000	
<210>	138
<400>	138
000	
<210>	139
<400>	139
000	
<210>	140
<400>	140
000	
<210>	141
<400>	141
000	
<210>	142

<400>	142
000	
<210>	143
<400>	143
000	
<210>	144
<400>	144
000	
<210>	145
<400>	145
000	
<210>	146
<400>	146
000	
<210>	147
<400>	147
000	
<210>	148
<400>	148
000	
<210>	149
<400>	149
000	
<210>	150
<400>	150
000	
<210>	151
<400>	151
000	
<210>	152
<400>	152
000	
<210>	153
<400>	153

000	
<210>	154
<400>	154
000	
<210>	155
<400>	155
000	
<210>	156
<400>	156
000	
<210>	157
<400>	157
000	
<210>	158
<400>	158
000	
<210>	159
<400>	159
000	
<210>	160
<400>	160
000	
<210>	161
<400>	161
000	
<210>	162
<400>	162
000	
<210>	163
<400>	163
000	
<210>	164
<400>	164
000	
<210>	165
<400>	165

000
 <210> 166
 <400> 166
 000
 <210
 > 167
 <400> 167
 000
 <210> 168
 <400> 168
 000
 <210> 169
 <400> 169
 000
 <210> 170
 <400> 170
 000
 <210> 171
 <400> 171
 000
 <210> 172
 <400> 172
 000
 <210> 173
 <400> 173
 000
 <210> 174
 <400> 174
 000
 <210> 175
 <400> 175
 000
 <210> 176
 <400> 176
 000

<210>	177
<400>	177
000	
<210>	178
<400>	178
000	
<210>	179
<400>	179
000	
<210>	180
<400>	180
000	
<210>	181
<400>	181
000	
<210>	182
<400>	182
000	
<210>	183
<400>	183
000	
<210>	184
<400>	184
000	
<210>	185
<400>	185
000	
<210>	186
<400>	186
000	
<210>	187
<400>	187
000	
<210>	188

<400>	188
000	
<210>	189
<400>	189
000	
<210>	190
<400>	190
000	
<210>	191
<400>	191
000	
<210>	192
<400>	192
000	
<210>	193
<400>	193
000	
<210>	194
<400>	194
000	
<210>	195
<400>	195
000	
<210>	196
<400>	196
000	
<210>	197
<400>	197
000	
<210>	198
<400>	198
000	
<210>	199
<400>	199

000
 <210> 200
 <400> 200
 000
 <210> 201
 <400> 201
 000
 <210> 202
 <400> 202
 000
 <210> 203
 <400> 203
 000
 <210> 204
 <400> 204
 000
 <210> 205
 <400> 205
 000
 <210> 206
 <400> 206
 000
 <210> 207
 <400> 207
 000
 <210> 208
 <400> 208
 000
 <210> 209
 <400> 209
 000
 <210
 > 210
 <400> 210
 000

<210>	211
<400>	211
000	
<210>	212
<400>	212
000	
<210>	213
<400>	213
000	
<210>	214
<400>	214
000	
<210>	215
<400>	215
000	
<210>	216
<400>	216
000	
<210>	217
<400>	217
000	
<210>	218
<400>	218
000	
<210>	219
<400>	219
000	
<210>	220
<400>	220
000	
<210>	221
<400>	221
000	
<210>	222
<400>	222
000	

<210>	223
<400>	223
000	
<210>	224
<400>	224
000	
<210>	225
<400>	225
000	
<210>	226
<400>	226
000	
<210>	227
<400>	227
000	
<210>	228
<400>	228
000	
<210>	229
<400>	229
000	
<210>	230
<400>	230
000	
<210>	231
<400>	231
000	
<210>	232
<400>	232
000	
<210>	233
<400>	233
000	
<210>	234

<400> 234
 000
 <210> 235
 <400> 235
 000
 <210> 236
 <400> 236
 000
 <210> 237
 <400> 237
 000
 <210> 238
 <400> 238

 000
 <210> 239
 <400> 239
 000
 <210> 240
 <400> 240
 000
 <210> 241
 <400> 241
 000
 <210> 242
 <400> 242
 000
 <210> 243
 <400> 243
 000
 <210> 244
 <400> 244
 000
 <210> 245
 <400> 245

000

<210> 246

<400> 246

000

<210> 247

<400> 247

000

<210> 248

<400> 248

000

<210> 249

<400> 249

000

<210> 250

<400> 250

000

<210> 251

<400> 251

000

<210> 252

<400> 252

000

<210

> 253

<400> 253

000

<210> 254

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 254

gcgacggaaa gagtatgagc tgg

<210> 255

<211> 23

<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	255	
tatttgactt cagtcagcga cgg		23
<210>	256	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	256	
tggaggcaag atatagatct tgg		23
<210>	257	
<211>	24	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	257	
gigttaattt caaacatcag cagc		24
<210>	258	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	258	
gacaagacat ccttgatttg		20
<210>	259	
<211>	19	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	259	
gaggttgact gtgtaaatg		19

<210>	260	
<211>	21	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	260	
	gataccagag tcacacaaca g	21
<210>	261	
<211>	21	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	261	
	tctacattaa ttctcttg c	21
<210>	262	
<211>	21	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	262	
	gataccagag tcacacaaca g	21
<210>	263	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	263	
	gggcaatgga ttggtcatcc tgg	23
<210>	264	
<211>	21	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Synthetic	
<400>	264	

tctacattaa ttctcttg c	21
<210> 265	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 265	
gacaagacat ccttgattg	20
<210> 266	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 266	
tctacattaa ttctcttg c	21
<210> 267	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 267	
gataccagag tcacacaaca g	21
<210> 268	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 268	
gaggttgact gtgtaaatg	19
<210> 269	
<211> 20	
<212> DNA	
<213> Artificial Sequence	

<220><223> Synthetic

<400> 269

gacaagacat ccttgatttg 20

<210> 270

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 270

gaggttgact gtgtaaag 19

<210> 271

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 271

gataccagag tcacacaaca g 21

<210> 272

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 272

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

1 5 10 15

Lys Ile Glu Trp His Glu

20

<210> 273

<211> 69

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 273

ggcggcgacc tcgagggtag cggctctgaac gatatttttg aagcgcagaa aattgaatgg 60

catgaataa 69

<210> 274

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 274

Cys Cys His Cys

1