(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년12월03일
(11) 등록번호 10-2185464
(24) 등록일자 2020년11월26일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) *C12N 15/10* (2017.01)
C12N 15/113 (2010.01)
(52) CPC특허분류
C12N 9/22 (2013.01)
C12N 15/102 (2013.01)
(21) 출원번호 10-2019-7030158
(22) 출원일자(국제) 2018년03월15일
심사청구일자 2019년11월28일
(85) 번역문제출일자 2019년10월14일
(65) 공개번호 10-2019-0129088
(43) 공개일자 2019년11월19일
(86) 국제출원번호 PCT/US2018/022751
(87) 국제공개번호 WO 2018/170333
국제공개일자 2018년09월20일
(30) 우선권주장
62/471,710 2017년03월15일 미국(US)
62/566,829 2017년10월02일 미국(US)
(56) 선행기술조사문헌
Molecular Cell, Vol.65, pp.618-630 (2017.02.16) 1부.*
WO2017219027 A1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
매사추세츠 인스티튜트 오브 테크놀로지
미국 02139 매사추세츠주 캠브리지 매사추세츠 애비뉴 77
더 브로드 인스티튜트, 인코퍼레이티드
미국 02142 매사 추세츠주 캠브리지 메인 스트리트 415
(72) 발명자
콕스 데이비드 벤자민 털리츠
미국 02139 매사추세츠주 캠브리지 매사추세츠 애비뉴 950 아파트먼트 217
피조차 니나
미국 03070 뉴 햄프셔주 뉴 보스턴 2번 엔에이치 턴파이크 1937
장 펑
미국 02139 매사추세츠주 캠브리지 퍼시픽 스트리트 100 아파트먼트 11
(74) 대리인
특허법인코리아나

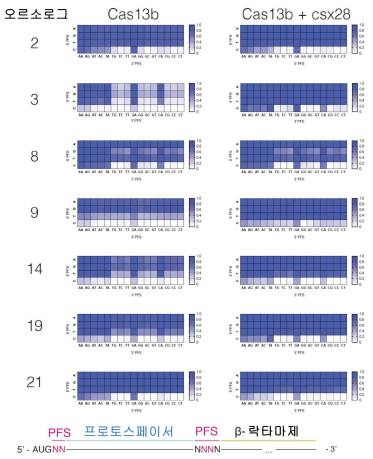
전체 청구항 수 : 총 30 항 심사관 : 한지혜

(54) 발명의 명칭 **신규 CAS13B 오르소로그 CRISPR 효소 및 시스템**

(57) 요 약

본 발명은 핵산을 표적화하기 위한 시스템, 방법 및 조성물을 제공한다. 특히, 본 발명은 신규 RNA-표적화 Cas13b 이펙터 단백질 및 가이드 RNA 또는 crRNA와 같은 적어도 하나의 표적화 핵산 성분을 포함하는 비-천연 발생 또는 조작된 RNA-표적화 시스템을 제공한다.

대 표 도 - 도8



(52) CPC특허분류
C12N 15/113 (2013.01)
C12N 2310/20 (2017.05)

명 세 서

청구범위

청구항 1

하기 i) 및 ii)를 포함하는 비-천연 발생 또는 조작된 조성물:

i) SEQ No. 31 내지 38 로 이루어진 군에서 선택되는 아미노산 서열로 이루어진 Cas13b 이펙터 단백질, 및

ii) crRNA,

crRNA는 a) 표적 RNA 서열에 혼성화할 수 있는 가이드 서열, 및 b) 직접 반복부 서열을 포함하고,

그리하여 표적 RNA 서열과 혼성화하는 가이드 서열과 복합체를 형성하는 Cas13b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성됨.

청구항 2

제 1 항에 있어서, Cas13b 이펙터 단백질 활성을 향상시키는 부속 단백질인 csx28 단백질을 포함하는 비-천연 발생 또는 조작된 조성물.

청구항 3

삭제

청구항 4

제 1 항에 있어서, Cas13b 이펙터 단백질 활성을 억제하는 부속 단백질인 csx27 단백질을 포함하는 비-천연 발생 또는 조작된 조성물.

청구항 5

삭제

청구항 6

제 1 항에 있어서, 2 개 이상의 crRNA를 포함하는 비-천연 발생 또는 조작된 조성물.

청구항 7

제 1 항에 있어서, 가이드 서열이 원핵생물 세포 내 표적 RNA 서열과 혼성화하는 비-천연 발생 또는 조작된 조성물.

청구항 8

제 1 항에 있어서, 가이드 서열이 진핵생물 세포 내 표적 RNA 서열과 혼성화하는 비-천연 발생 또는 조작된 조성물.

청구항 9

제 1 항에 있어서, Cas13b 이펙터 단백질이 하나 이상의 핵 국재화 신호 (NLS)를 포함하는 비-천연 발생 또는 조작된 조성물.

청구항 10

제 1 항에 있어서, Cas13b 이펙터 단백질이 하나 이상의 기능성 도메인과 회합(associated)되는 비-천연 발생 또는 조작된 조성물로서, 상기 기능성 도메인이 표적 RNA 서열을 절단하거나, 표적 RNA 서열의 전사 또는 번역을 변형시키는 비-천연 발생 또는 조작된 조성물.

청구항 11

삭제

청구항 12

삭제

청구항 13

제 1 항에 있어서, Cas13b 이펙터 단백질이 하나 이상의 기능성 도메인과 회합(associated)되고; 이펙터 단백질은 HEPN 도메인 내에 하나 이상의 하기 돌연변이를 함유하며:

R116A, H121A, R1177A, H1182A(여기서 아미노산 위치는 베르게옐라 주헬쿰(*Bergeyella zoohelcum*) ATCC 43767에서 유래한 그룹 29 단백질의 아미노산 위치에 해당),

그리하여 복합체는 후생유전적 변형인자 또는 전사 또는 번역 활성화 또는 억제 신호를 전달할 수 있는 비-천연 발생 또는 조작된 조성물로서, 상기 기능성 도메인이 표적 RNA 서열을 절단하거나, 표적 RNA 서열의 전사 또는 번역을 변형시키는 비-천연 발생 또는 조작된 조성물.

청구항 14

제 2 항에 있어서, Cas13b 이펙터 단백질 및 부속 단백질이 동일한 유기체로부터 유래하는 비-천연 발생 또는 조작된 조성물.

청구항 15

제 2 항에 있어서, Cas13b 이펙터 단백질 및 부속 단백질이 상이한 유기체로부터 유래하는 비-천연 발생 또는 조작된 조성물.

청구항 16

하기를 포함하는 하나 이상의 벡터를 포함하는, 제 1 항의 조성물을 제공하기 위한 Cas13b 벡터 시스템:

SEQ No. 31 내지 38 로 이루어진 군에서 선택되는 아미노산 서열로 이루어진 Cas13b 이펙터 단백질을 인코딩하는 뉴클레오티드 서열에 작동가능하게 연결된 제 1 조절 요소, 및

crRNA 를 인코딩하는 뉴클레오티드 서열에 작동가능하게 연결된 제 2 조절 요소.

청구항 17

제 16 항에 있어서, 부속 단백질인 csx28 단백질 또는 csx27 단백질의 뉴클레오티드 서열에 작동가능하게 연결된 조절 요소를 더 포함하는 벡터 시스템.

청구항 18

제 16 항에 있어서, Cas13b 이펙터 단백질을 인코딩하는 뉴클레오티드 서열이 진핵생물 세포에서 발현을 위해 코돈 최적화되는 벡터 시스템.

청구항 19

제 17 항에 있어서, Cas13b 이펙터 단백질; 및 부속 단백질인 csx28 단백질 또는 csx27 단백질;을 코딩하는 뉴클레오티드 서열이 진핵생물 세포에서 발현을 위해 코돈 최적화되는 벡터 시스템.

청구항 20

제 16 항에 있어서, 단일 벡터에 포함되는 벡터 시스템.

청구항 21

제 16 항에 있어서, 하나 이상의 벡터가 바이러스 벡터를 포함하는 벡터 시스템.

청구항 22

제 16 항에 있어서, 하나 이상의 벡터가 하나 이상의 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연합 또는 헤르페스 심플렉스 바이러스 벡터를 포함하는 벡터 시스템.

청구항 23

하기를 포함하는 비-천연 발생 또는 조작된 조성물의 Cas13b 이펙터 단백질 및 하나 이상의 핵산 성분을 전달하도록 구성된 전달 시스템:

i) SEQ No. 31 내지 38 로 이루어진 군에서 선택되는 아미노산 서열로 이루어진 Cas13b 이펙터 단백질, 및

ii) crRNA,

crRNA는 a) 세포 내 표적 RNA 서열과 혼성화하는 가이드 서열, 및 b) 직접 반복부 서열을 포함하고,

Cas13b 이펙터 단백질은 crRNA와 복합체를 형성하고,

가이드 서열은 표적 RNA 서열과 서열-특이적 결합을 유도하고,

그리하여 표적 RNA 서열과 혼성화하는 가이드 서열과 복합체를 형성하는 Cas13b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성됨.

청구항 24

제 23 항에 있어서, 하나 이상의 벡터 또는 하나 이상의 폴리뉴클레오티드 분자를 포함하고, 하나 이상의 벡터 또는 폴리뉴클레오티드 분자는 비-천연 발생 또는 조작된 조성물의 하나 이상의 핵산 성분 및 Cas13b 이펙터 단백질을 인코딩하는 하나 이상의 폴리뉴클레오티드 분자를 포함하는 전달 시스템.

청구항 25

제 23 항에 있어서, 리포좀(들), 입자(들), 엑소좀(들), 미소소낭(들), 유전자-총 또는 하나 이상의 바이러스 벡터(들)를 포함하는 전달 비히클을 포함하는 전달 시스템.

청구항 26

삭제

청구항 27

관심 표적 유전자의 발현을 변형시키는 방법으로서, 표적 유전자는 원핵생물 세포 또는 인간을 제외한 진핵생물 세포에 존재하고, 표적 RNA를 하기를 포함하는 하나 이상의 비-천연 발생 또는 조작된 조성물과 접촉시키는 것을 포함하는 방법:

i) SEQ No. 31 내지 38 로 이루어진 군에서 선택되는 아미노산 서열로 이루어진 Cas13b 이펙터 단백질, 및

ii) crRNA,

crRNA는 a) 세포 내 표적 RNA 서열과 혼성화하는 가이드 서열, 및 b) 직접 반복부 서열을 포함하고,

Cas13b 이펙터 단백질은 crRNA와 복합체를 형성하고,

가이드 서열은 세포 내 표적 RNA 서열과 서열-특이적 결합을 유도하고,

그리하여 표적 RNA 서열과 혼성화하는 가이드 서열과 복합체를 형성하는 Cas13b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성됨,

그리하여 관심 표적 유전자좌의 발현이 변형됨.

청구항 28

제 27 항에 있어서, 표적 RNA를 Cas13b 이펙터 단백질 활성을 향상시키는 부속 단백질인 csx28 단백질과 접촉시키는 단계를 더 포함하는 방법.

청구항 29

삭제

청구항 30

제 27 항에 있어서, 표적 RNA를 Cas13b 이펙터 단백질 활성을 억제하는 부속 단백질인 csx27 단백질과 접촉시키는 단계를 더 포함하는 방법.

청구항 31

삭제

청구항 32

제 27 항에 있어서, 표적 유전자의 발현의 변형은 표적 RNA의 절단을 포함하는 방법.

청구항 33

제 27 항에 있어서, 표적 유전자의 발현의 변형은 표적 RNA의 발현의 증가 또는 감소를 포함하는 방법.

청구항 34

삭제

청구항 35

삭제

청구항 36

제 16 항 내지 제 22 항 중 어느 한 항에 있어서, 치료의 치료적 방법에서 사용하기 위한 벡터 시스템.

청구항 37

제 23 항 내지 제 25 항 중 어느 한 항에 있어서, 치료의 치료적 방법에서 사용하기 위한 전달 시스템.

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

SEQ No. 31 내지 38 로 이루어진 군에서 선택되는 아미노산 서열로 이루어진 단리된 Cas13b 이펙터 단백질을 인코딩하는 단리된 핵산, 또는 DNA이고 crRNA를 인코딩하는 서열을 더 포함하는 SEQ No. 31 내지 38 로 이루어진 군에서 선택되는 아미노산 서열로 이루어진 단리된 Cas13b 이펙터 단백질을 인코딩하는 단리된 핵산, 또는 SEQ No. 31 내지 38 로 이루어진 군에서 선택되는 아미노산 서열로 이루어진 단리된 Cas13b 이펙터 단백질의 Cas13b 를 포함하는 단리된 진핵생물 세포.

청구항 50

하기 i) 및 ii)를 포함하는 비-천연 발생 또는 조작된 조성물:

i) SEQ No. 31 내지 38 로 이루어진 군에서 선택되는 아미노산 서열로 이루어진 Cas13b 이펙터 단백질을 인코딩하는 mRNA, 및

ii) crRNA, crRNA는 a) 표적 RNA 서열에 혼성화할 수 있는 가이드 서열, 및 b) 직접 반복부 서열을 포함함.

발명의 설명

기 술 분 야

[0001] 관련 출원 및 참조에 의한 통합

[0002] 우선권은 2017 년 3 월 15 일에 출원된 미국 가출원 62/471,710 및 2017 년 10 월 2 일에 출원된 미국 가출원 62/566,829 에 청구된다.

[0003] 지정국으로 미국을 포함하여 PCT 출원, *그 중에서도* 2016 년 10 월 21 일 출원된 출원 번호 PCT/US2016/058302 를 참조한다. 2015 년 10 월 22 일에 출원된 미국 가특허 출원 62/245,270, 2016 년 2 월 17 일에 출원된 미국 가특허 출원 62/296,548 및 2016 년 8 월 17 일에 출원된 미국 가특허 출원 62/376,367 및 62/376,382 를 참고한다. 우선권은 2017 년 3 월 15 일에 출원된 미국 가출원 62/471,792 및 2017 년 4 월 12 일에 출원된 미국 가출원 62/484,786 을 참고한다. 하기가 언급된다: Smargon et al. (2017), "Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28," Molecular Cell 65, 618-630 (Feb. 16, 2017) doi: 10.1016/j.molcel.2016.12.023. Epub Jan 5, 2017 및 Smargon et al. (2017), "Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28," bioRxiv 092577; doi: https://doi.org/10.1101/092577. Posted December 9, 2017. 상기 출원 및 문헌 인용 각각은 본원에 참고로 포함된다.

[0004] 게다가, 본 명세서에서 참조로 편입한 임의 문서 또는 본 명세서에 언급된 임의 제품에 관한 임의의 제조사 지시서, 서술서, 제품 사양서, 및 제품 설명서와 함께, 본 명세서에서 인용하는 문서에서 참조하거나 또는 인용하는 모든 문서는 참조로 본 명세서에 편입되고, 본 발명의 실시에 적용될 수 있다. 더 구체적으로, 모든 참조된 문헌은, 마치 각각의 개별 문헌을 참고로 포함하는 것으로 특정적으로 그리고 개별적으로 나타내는 것과 동일한 정도로 참고로 포함된다.

[0005] 주정부 후원 연구에 관한 진술

[0006] 본 발명은 미국 국립 보건원에서 수여하는 보조금 번호 MH100706 및 MH110049 하의 정부 지원으로 수행되었다. 정부는 본 발명에 대해 특정 권리를 갖는다.

[0007] 발명의 분야

[0008] 본 발명은 일반적으로 클러스터링되고 규칙적으로 산재된 짧은 팔린드로믹 반복부(Clustered Regularly Interspaced Short Palindromic Repeats; CRISPR)와 관련된 벡터 시스템 및 이의 성분을 사용할 수 있는, 서열 표적화를 수반하는 유전자 발현의 제어, 예컨대 유전자 전사물의 변동(perturbation) 또는 핵산 편집에 사용되는 시스템, 방법 및 조성물에 관한 것이다.

배 경 기 술

[0009] 박테리아 및 고세균 적응 면역의 CRISPR-CRISPR 연합 (Cas) 시스템은 단백질 조성 및 게놈 유전자좌 아키텍처의 극도의 다양성을 보여주는 일부 그러한 시스템이다. CRISPR-Cas 시스템 유전자좌는 50 개 초과의 유전자 패밀리를 가지며, 유전자좌 구조의 빠른 진화 및 극단적인 다양성을 나타내는 엄격한 보편적 유전자는 존재하지 않는다. 지금까지는, 다면적인 접근법을 채택하여, 93 개의 Cas 단백질에 대해 약 395 개의 프로파일의 포괄적인 *cas* 유전자 확인이 있다. 분류는 특징적인 유전자 프로파일 및 유전자좌 구성의 특징을 포함한다.
CRISPR-Cas 시스템의 새로운 분류가 제안되어 있으며, 여기서 이들 시스템은 광범위하게 2 개의 부류, 즉 다수 서브유닛 이펙터 복합체를 가지는 클래스 1 및 Cas9 단백질로 예시되는 단일-서브유닛 이펙터 모듈을 가지는 클래스 2로 나뉜다. 클래스 2 CRISPR-Cas 시스템과 연관된 신규 이펙터 단백질은 강력한 게놈 조작 도구로서 개발될 수 있으며, 추정상의 신규 이펙터 단백질 및 이의 조작 및 최적화의 예측이 중요하다. 신규 Cas13b 오르소로그 및 이의 용도가 바람직하다.

[0010] 본 출원에서 임의의 문헌의 인용 또는 확인은 상기 문헌이 본 출원에 대한 선행 기술로서 이용가능하다는 인정이 아니다.

발명의 내용

해결하려는 과제

[0011] 발명의 요약

[0012] 이펙터 단백질은 2개 서브그룹, 타입 VI-B1 및 타입 VI-B2를 포함하고, RNA-프로그램가능한 뉴클레아제, RNA-간섭의 구성원을 포함하고 RNA 파지에 대항하는 박테리아 양자 면역에 관여될 수 있다. Cas13b 시스템은 거대한 단일 이펙터 (대략 1100개 아미노산 길이) 및 CRISPR 어레이 근처 2개의 소형 추정 부속 단백질 (대략 200개 아미노산 길이) 중 하나를 포함하거나 또는 그 중 하나도 포함하지 않는다. 근처의 작은 단백질에 기초하여, 시스템은 2 개의 유전자좌 (Loci) A 및 B로 분기된다. 어레이로부터 상류 또는 하류에 25 킬로베이스 쌍까지의 추가 단백질은 각각의 유전자좌를 갖는 종에 걸쳐 보존되지 않는다. 소수의 예외로서, CRISPR 어레이는 36 뉴클레오티드 길이의 직접 반복부 서열 및 30 뉴클레오티드 길이의 스페이서 서열을 포함한다. 직접 반복부는 일반적으로, 3' 말단의 CAAC에 역 상보적인 5' 말단의 GTTG/GUUG를 가져서, 특히 말단에서 잘 보존된다. 이러한 보존성은 유전자좌 내 단백질(들)과 잠재적으로 상호작용하는 RNA 루프 구조의 강력한 염기 쌍형성을 시사한다. 직접 반복부에 상보적인 모티프 검색은 어레이 근처에 어떠한 후보 tracrRNA도 밝히지 못하여서, 아마도 Cpf1 유전자좌에 존재하는 것과 유사한 단일 crRNA를 의미하는 듯하다.

[0013] 본 발명의 실시형태에서, 타입 VI-B 시스템은 신규 Cas13b 이펙터 단백질 및 임의로 Cas13b 이펙터 단백질의 상류 또는 하류에서 인코딩된 소형 부속 단백질을 포함한다. 일정 실시형태에서, 소형 부속 단백질은 RNA를 표적으로 하는 Cas13b 이펙터의 능력을 향상시킨다.

과제의 해결 수단

[0014] 본 발명은 하기를 포함하는 비-천연 발생 또는 조작된 조성물을 제공한다:

[0015] i) 특정의 신규 Cas13b 이펙터 단백질, 및

[0016] ii) crRNA,

[0017] crRNA는 a) 표적 RNA 서열에 혼성화할 수 있는 가이드 서열, 및 b) 직접 반복부 서열을 포함하고,

[0018] 그리하여 표적 RNA 서열과 혼성화하는 가이드 서열과 복합체를 형성하는 Cas13b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성됨. 복합체는 시험관 내 또는 생체 외에서 형성될 수 있고, 세포 내로 도입되거나 RNA와 접촉될 수 있으며; 또는 생체 내에서 형성될 수 있다.

[0019] 일부 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 조성물은 타입 VI-B CRISPR-Cas 이펙터 단백질 활성을 향상시키는 부속 단백질을 포함할 수 있다.

[0020] 이러한 일정 실시형태에서, Cas13b 이펙터 단백질 활성을 향상시키는 부속 단백질은 csx28 단백질이다. 이러한 실시형태에서, 타입 VI-B CRISPR-Cas 이펙터 단백질 및 타입 VI-B CRISPR-Cas 부속 단백질은 동일한 공급원으로부터 유래되거나 또는 상이한 공급원으로부터 유래될 수 있다.

[0021] 일부 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 조성물은 Cas13b 이펙터 단백질 활성을 억제하는 부속 단백질을 포함할 수 있다.

[0022] 이러한 일정 실시형태에서, Cas13b 이펙터 단백질 활성을 억제하는 부속 단백질은 csx27 단백질이다. 이러한 실시형태에서, 타입 VI-B CRISPR-Cas 이펙터 단백질 및 타입 VI-B CRISPR-Cas 부속 단백질은 동일한 공급원으로부터 유래되거나 또는 상이한 공급원으로부터 유래될 수 있다. 본 발명의 일정 실시형태에서, 타입 VI-B CRISPR-Cas 이펙터 단백질은 표 1에서 확인한다. 일정 실시형태에서, 타입 VI-B CRISPR-Cas 부속 단백질은 표 1에서 확인한다.

[0023] 일부 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 조성물은 2 개 이상의 crRNA를 포함한다.

[0024] 일부 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 조성물은 원핵생물 세포에서 표적 RNA 서열과 혼성화하는 가이드 서열을 포함한다.

[0025] 일부 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 조성물은 진핵생물 세포에서 표적 RNA 서열과 혼성화하는 가이드 서열을 포함한다.

[0026] 일부 실시형태에서, Cas13b 이펙터 단백질은 하나 이상의 핵 국재화 신호(NLS)를 포함한다.

[0027] 본 발명의 Cas13b 이펙터 단백질은 표 1에 제시된 바와 같은 또는 이와 같은 단백질이거나, 이를 포함하거나, 또는 그로 본질적으로 이루어지거나, 또는 이들로 이루어지거나 또는 이를 수반하거나 또는 이와 관련되거나 또는 이에 언급된다. 본 발명은 본원에 제시된 바와 같은 돌연변이 또는 변형을 포함하여, 표 1에 제시된 단백질을 제공하거나, 이에 관한 것이거나, 이를 수반하거나 포함하거나, 이로 본질적으로 이루어지거나, 이루어진다. 표 1 Cas13b 이펙터 단백질은 표 1과 관련하여 본원에서 보다 상세하게 논의된다.

[0028] 본 발명의 비-천연 발생 또는 조작된 조성물의 일부 실시형태에서, Cas13b 이펙터 단백질은 하나 이상의 기능성 도메인과 연합된다. 연합은 crRNA와의 연합에 의하거나, 또는 기능성 도메인에 대한 이펙터 단백질의 직접 연결에 의할 수 있다. 비-제한적인 예에서, crRNA는 예를 들어 핵산 결합 어댑터 단백질에 결합하는 뉴클레오티드 또는 앱타머를 포함하는, 관심 기능성 도메인과 연합될 수 있는 첨가 또는 삽입 서열을 포함한다.

[0029] 일정 비제한적 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 조성물은 표적 RNA 서열을 절단하는 기능성 도메인을 포함한다.

[0030] 일정 비제한적 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 조성물은 표적 RNA 서열의 전사 또는 번역을 변형시키는 기능성 도메인을 포함한다.

[0031] 본 발명의 조성물의 일부 실시형태에서, Cas13b 이펙터 단백질은 하나 이상의 기능성 도메인과 연합되고, 이펙터 단백질은 HEPN 도메인 내에 하나 이상의 돌연변이를 함유하며, 그리하여 복합체가 후생유전적 변형인자 또는 전사 또는 번역 활성화 또는 억제 신호를 전달할 수 있다. 복합체는 시험관 내 또는 생체 외에서 형성될 수 있고, 세포 내로 도입되거나 RNA와 접촉될 수 있으며; 또는 생체 내에서 형성될 수 있다.

[0032] 본 발명의 비-천연 발생 또는 조작된 조성물의 일부 실시형태에서, Cas13b 이펙터 단백질 및 부속 단백질은 동일한 유기체로부터 유래한다.

[0033] 본 발명의 비-천연 발생 또는 조작된 조성물의 일부 실시형태에서, Cas13b 이펙터 단백질 및 부속 단백질은 상이한 유기체로부터 유래한다.

[0034] 본 발명은 또한 타입 VI-B CRISPR-Cas 벡터 시스템를 제공하고, 이것은 하기를 포함하는 하나 이상의 벡터를 포함한다:

[0035] Cas13b 이펙터 단백질을 인코딩하는 뉴클레오티드 서열에 작동가능하게 연결된 제 1 조절 요소, 및

[0036] crRNA 를 인코딩하는 뉴클레오티드 서열에 작동가능하게 연결된 제 2 조절 요소.

[0037] 일정 실시형태에서, 본 발명의 벡터 시스템은 타입 VI-B CRISPR-Cas 부속 단백질의 뉴클레오티드 서열에 작동가능하게 연결된 조절 요소를 더 포함한다.

[0038] 적절한 경우, 타입 VI-B CRISPR-Cas 이펙터 단백질을 인코딩하는 뉴클레오티드 서열 및/또는 타입 VI-B CRISPR-Cas 부속 단백질을 인코딩하는 뉴클레오티드 서열은 진핵생물 세포에서의 발현을 위해 코돈 최적화된다.

[0039] 본 발명의 벡터 시스템의 일부 실시형태에서, Cas13b 이펙터 단백질 및 보조 단백질을 인코딩하는 뉴클레오티드 서열은 진핵생물 세포에서의 발현을 위해 코돈 최적화된다.

[0040] 일부 실시형태에서, 본 발명의 벡터 시스템은 단일 벡터를 포함한다.

[0041] 본 발명의 벡터 시스템의 일부 실시형태에서, 하나 이상의 벡터는 바이러스 벡터를 포함한다.

[0042] 본 발명의 벡터 시스템의 일부 실시형태에서, 하나 이상의 벡터는 하나 이상의 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연합 또는 헤르페스 심플렉스 바이러스 벡터를 포함한다.

[0043] 본 발명은 하기를 포함하는 비-천연 발생 또는 조작된 조성물의 Cas13b 이펙터 단백질 및 하나 이상의 핵산 성분을 전달하도록 구성된 전달 시스템을 제공한다:

[0044] i) Cas13b 이펙터 단백질, 및

[0045] ii) crRNA,

[0046] crRNA는 a) 세포 내 표적 RNA 서열과 혼성화하는 가이드 서열, 및 b) 직접 반복부 서열을 포함하고,

[0047] Cas13b 이펙터 단백질은 crRNA와 복합체를 형성하고,

[0048] 가이드 서열은 표적 RNA 서열과 서열-특이적 결합을 유도하고,

[0049] 그리하여 표적 RNA 서열과 혼성화하는 가이드 서열과 복합체를 형성하는 Cas13b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성됨. 복합체는 시험관 내 또는 생체 외에서 형성될 수 있고, 세포 내로 도입되거나 RNA와 접촉될 수 있으며; 또는 생체 내에서 형성될 수 있다.

[0050] 본 발명의 벡터 시스템의 일부 실시형태에서, 시스템은 하나 이상의 벡터 또는 하나 이상의 폴리뉴클레오티드 분자를 포함하고, 하나 이상의 벡터 또는 폴리뉴클레오티드 분자는 비-천연 발생 또는 조작된 조성물의 하나 이상의 핵산 성분 및 Cas13b 이펙터 단백질을 인코딩하는 하나 이상의 폴리뉴클레오티드 분자를 포함한다.

[0051] 일부 실시형태에서, 본 발명의 전달 시스템은 리포솜(들), 입자(들), 엑소솜(들), 미소소낭(들), 유전자-총 또는 하나 이상의 바이러스 벡터(들)를 포함하는 전달 비히클을 포함한다.

[0052] 일부 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 조성물은 치료의 치료적 방법에서 또는 조사 프로그램에서 사용을 위한 것이다.

[0053] 일부 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 벡터 시스템은 치료의 치료적 방법에서 또는 조사 프로그램에서 사용을 위한 것이다.

[0054] 일부 실시형태에서, 본 발명의 비-천연 발생 또는 조작된 전달 시스템은 처리의 치료적 방법에서 또는 조사 프로그램에서 사용을 위한 것이다.

[0055] 본 발명은 관심 표적 유전자의 발현을 변형시키는 방법을 제공하며, 표적 RNA를 하기를 포함하는 하나 이상의 비-천연 발생 또는 조작된 조성물과 접촉시키는 것을 포함하는 방법:

[0056] i) Cas13b 이펙터 단백질, 및

[0057] ii) crRNA,

[0058] crRNA는 a) 세포 내 표적 RNA 서열과 혼성화하는 가이드 서열, 및 b) 직접 반복부 서열을 포함하고,

[0059] Cas13b 이펙터 단백질은 crRNA와 복합체를 형성하고,

[0060] 가이드 서열은 세포 내 표적 RNA 서열과 서열-특이적 결합을 유도하고,

[0061] 그리하여 표적 RNA 서열과 혼성화하는 가이드 서열과 복합체를 형성하는 Cas13b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성되고,

[0062] 그리하여 관심 표적 유전자좌의 발현이 변형됨. 복합체는 시험관 내 또는 생체 외에서 형성될 수 있고, 세포 내로 도입되거나 RNA와 접촉될 수 있으며; 또는 생체 내에서 형성될 수 있다.

[0063] 일부 실시형태에서, 관심 표적 유전자의 발현을 변형시키는 방법은 표적 RNA를 Cas13b 이펙터 단백질 활성을 향상시키는 보조 단백질과 접촉시키는 단계를 추가로 포함한다.

[0064] 관심 표적 유전자의 발현을 변형시키는 방법의 일부 실시형태에서, Cas13b 이펙터 단백질 활성을 향상시키는 보조 단백질은 csx28 단백질이다.

[0065] 일부 실시형태에서, 관심 표적 유전자의 발현을 변형시키는 방법은 표적 RNA를 Cas13b 이펙터 단백질 활성을 억제하는 보조 단백질과 접촉시키는 단계를 추가로 포함한다.

[0066] 관심 표적 유전자의 발현을 변형시키는 방법의 일부 실시형태에서, Cas13b 이펙터 단백질 활성을 억제하는 보조 단백질은 csx27 단백질이다.

[0067] 일부 실시형태에서, 관심 표적 유전자의 발현을 변형시키는 방법은 표적 RNA의 절단을 포함한다.

[0068] 일부 실시형태에서, 관심 표적 유전자의 발현을 변형시키는 방법은 표적 RNA의 발현의 증가 또는 감소를 포함한다.

[0069] 관심 표적 유전자의 발현을 변형시키는 방법의 일부 실시형태에서, 표적 유전자는 원핵생물 세포이다.

[0070] 관심 표적 유전자의 발현을 변형시키는 방법의 일부 실시형태에서, 표적 유전자는 진핵생물 세포이다.

[0071] 본 발명은 변형된 관심 표적을 포함하는 세포를 제공하고, 관심 표적은 본 명세서에 개시된 임의 방법에 따라서 변형된다.

[0072] 본 발명의 일부 실시형태에서, 세포는 원핵생물 세포이다.

[0073] 본 발명의 일부 실시형태에서, 세포는 진핵생물 세포이다.

[0074] 일부 실시형태에서, 세포에서 관심 표적의 변형은 다음을 초래한다:

[0075] 적어도 하나의 유전자 산물의 변경된 발현을 포함하는 세포;

[0076] 적어도 하나의 유전자 산물의 변경된 발현을 포함하는 세포로서, 적어도 하나의 유전자 산물의 발현이 증가되는 세포; 또는

[0077] 적어도 하나의 유전자 산물의 변경된 발현을 포함하는 세포로서, 적어도 하나의 유전자 산물의 발현이 감소되는 세포.

[0078] 일부 실시형태에서, 세포는 포유동물 세포 또는 인간 세포이다.

[0079] 본 발명은 또한 본 명세서에 개시된 세포의 또는 이를 포함하는 세포주 또는 본 명세서에 개시된 임의의 방법에 의해 변형된 세포, 또는 이의 자손을 제공한다.

[0080] 본 발명은 본 명세서에 개시된 하나 이상의 세포를 포함하는 복수세포 유기체, 또는 본 명세서에 개시된 임의 방법에 따라 변형된 하나 이상의 세포를 제공한다.

[0081] 본 발명은 본 명세서에 개시된 하나 이상의 세포 또는 본 명세서에 개시된 임의 방법에 따라 변형된 하나 이상의 세포를 포함하는 식물 또는 동물 모델을 제공한다.

[0082] 본 발명은 본 명세서에 개시된 세포 또는 세포주 또는 유기체 또는 식물 또는 동물 모델로부터 유래된 유전자 산물을 제공한다.

[0083] 일부 실시형태에서, 발현되는 유전자 산물의 양은 발현이 변경되지 않은 세포로부터 유래되는 유전자 산물의 양을 초과하거나 또는 그 미만이다.

[0084] 본 발명은 표 1에 제시되거나 이를 포함하거나 이로 본질적으로 이루어지거나, 이것으로 이루어진 단리된 Cas13b 이펙터 단백질을 제공한다. 표 1 Cas13b 이펙터 단백질은 표 1과 관련하여 본원에서 보다 상세하게 논의된다. 본 발명은 Cas13b 이펙터 단백질을 인코딩하는 단리된 핵산을 제공한다. 본 발명의 일부 실시형태에서, 단리된 핵산은 DNA 서열을 포함하고 crRNA를 인코딩하는 서열을 추가로 포함한다. 본 발명은 Cas13b 이펙터 단백질을 인코딩하는 핵산을 포함하는 단리된 진핵생물 세포를 제공한다. 따라서, 본원에서 "Cas13b 이펙터 단백질" 또는 "이펙터 단백질" 또는 "Cas"또는 "Cas 단백질" 또는 "RNA 표적화 이펙터 단백질" 또는 "RNA 표적화 단백질" 또는 유사한 표현은 표 1을 참조하여 이해될 수 있고 표 1 Cas13b 이펙터 단백질로서 판독될 수 있고; "CRISPR 시스템을 표적으로 하는 RNA" 와 같은 표현은 표 1을 참조하여 이해될 수 있고 표 1 Cas13b 이펙터 단백질 CRISPR 시스템으로서 판독될 수 있고; 가이드 RNA 또는 sgRNA에 대한 참고문헌은 Cas13b 시스템 crRNA의 본원의 논의와 함께 판독되어야 하며, 예를 들어, 다른 시스템에서 sgRNA인 것은 본 발명에서 crRNA인 것으로 또는 이와 유사한 것으로 고려될 수 있다.

[0085] 본 발명은 본 발명의 Cas13b 이펙터 단백질 (예를 들어, 표 1)을 위한 적합한 가이드 서열의 요건을 동정하는 방법을 제공하고, 상기 방법은 하기 단계를 포함한다:

[0086] (a) 유기체 내에서 필수 유전자의 세트를 선택하는 단계,

[0087] (b) 이들 유전자의 코딩 영역을 비롯하여 이들 유전자의 5' 및 3' UTR의 영역에 혼성화할 수 있는 표적화 가이드 서열의 라이브러리를 디자인하는 단계,

[0088] (c) 대조군 가이드로서 상기 유기체의 게놈 내 임의 영역에 혼성화하지 않는 무작위 가이드 서열을 생성시키는 단계,

[0089] (d) RNA-표적화 단백질 및 제1 내성 유전자를 포함하는 플라스미드 및 상기 표적화 가이드의 라이브러리 및 상기 대조군 가이드 및 제2 내성 유전자를 포함하는 가이드 플라스미드 라이브러리를 제조하는 단계,

[0090] (e) 상기 플라스미드를 숙주 세포에 공동도입시키는 단계,

[0091] (f) 상기 제1 및 제2 내성 유전자에 대한 선택 배지 상에 상기 숙주 세포를 도입시키는 단계,

[0092] (g) 성장하는 숙주 세포의 필수 유전자를 서열분석하는 단계,

[0093] (h) 대조군 가이드에 의한 세포의 감손을 비교하여 표적화 가이드로 형질전환된 세포의 감손의 유의성을 결정하는 단계; 및

[0094] (i) 감손된 가이드 서열을 기반으로 적합한 가이드 서열의 요건을 결정하는 단계.

[0095] 이러한 방법의 일 측면에서, RNA-표적화 단백질의 적합한 가이드 서열을 위한 PFS 서열의 결정은 감손된 세포에서 가이드에 의해 표적화된 서열의 비교에 의한다. 이러한 방법의 일 측면에서, 방법은 상이한 반복 실험으로 상이한 조건에 대해 가이드 존재비를 비교하는 단계를 더 포함한다. 이러한 방법의 일 측면에서, 대조군 가이드는 그들이 반복 실험으로 가이드 감손에서 한계 편차를 보이는 것을 결정하도록 선택된다. 이러한 방법의 일 측면에서, 감손의 중요성은 (a) 가장 감손된 대조 가이드보다 더 큰 감손; 또는 (b) 평균 감손보다 대조 가이드에 대한 표준 편차의 두 배를 더한 감손으로서 결정된다. 이러한 방법의 일 측면에서, 숙주 세포는 박테리아 숙주 세포이다. 이러한 방법의 일 측면에서, 플라스미드를 공동도입시키는 단계는 전기천공에 의하며 숙주 세포는 전기천공 적격한 숙주 세포이다.

[0096] Cas13b

[0097] 본 발명은 관심 표적 유전자좌의 서열 또는 그와 연합된 서열을 변형시키는 방법을 제공하고, 방법은 상기 유전자좌에 Cas13b 이펙터 단백질 및 하나 이상의 핵산 성분을 포함하는 비-천연 발생 또는 조작된 조성물을 전달하는 단계를 포함하고, 이펙터 단백질은 하나 이상의 핵산 성분과 복합체를 형성하고 관심 유전자좌에 상기 복합체의 결합시 이펙터 단백질은 관심 표적 유전자좌의 서열 또는 그와 연합된 서열의 변형을 유도한다. 바람직한 실시형태에서, 변형은 가닥 파손의 도입이다. 바람직한 실시형태에서, 관심 표적 유전자좌의 서열 또는 그와 연합된 서열은 RNA를 포함하거나 또는 RNA로 이루어진다.

[0098] 본 발명은 관심 표적 유전자좌의 서열 또는 그와 연합된 서열을 변형시키는 방법을 제공하고, 방법은 상기 유전자좌에 Cas13b 이펙터 단백질, 임의로 소형 부속 단백질, 및 하나 이상의 핵산 성분을 포함하는 비-천연 발생 또는 조작된 조성물을 전달하는 단계를 포함하고, 이펙터 단백질은 하나 이상의 핵산 성분과 복합체를 형성하고 관심 유전자좌에 상기 복합체의 결합시 이펙터 단백질은 관심 표적 유전자좌의 서열 또는 그와 연합된 서열의

변형을 유도한다. 바람직한 실시형태에서, 변형은 가닥 파손의 도입이다. 바람직한 실시형태에서, 관심 표적 유전자좌의 서열 또는 그와 연합된 서열은 RNA를 포함하거나 또는 RNA로 이루어진다.

[0099] 본 발명은 관심 표적 유전자좌의 서열 또는 그와 연합된 서열을 변형시키는 방법을 제공하고, 방법은 유전자좌 또는 그와 연합된 상기 서열에 Cas13b 유전자좌 이펙터 단백질 및 하나 이상의 핵산 성분을 포함하는 비-천연 발생 또는 조작된 조성물을 전달하는 단계를 포함하고, Cas13b 이펙터 단백질은 하나 이상의 핵산 성분과 복합체를 형성하고 상기 복합체가 관심 유전자에 결합시 이펙터 단백질은 관심 표적 유전자좌의 서열 또는 그와 연합된 서열의 변형을 유도한다. 바람직한 실시형태에서, 변형은 가닥 파손의 도입이다. 바람직한 실시형태에서 Cas13b 이펙터 단백질은 하나의 핵산 성분; 유리하게는 조작되거나 또는 비-천연 발생인 핵산 성분과 복합체를 형성한다. 관심 표적 유전자좌와 연관되거나 상기 표적 유전자좌에서의 서열의 변형의 유도는 Cas13b 이펙터 단백질-핵산으로 가이드될 수 있다. 바람직한 구현예에서, 하나의 핵산 성분은 CRISPR RNA(crRNA)이다. 바람직한 실시형태에서 하나의 핵산 성분은 성숙한 crRNA 또는 가이드 RNA이고, 성숙한 crRNA 또는 가이드 RNA는 스페이서 서열 (또는 가이드 서열) 및 직접 반복부 (DR) 서열 또는 이의 유도체를 포함한다. 바람직한 구현예에서, 스페이서 서열 또는 이의 유도체는 씨드(seed) 서열을 포함하며, 상기 시드 서열은 표적 유전자좌에서 서열에 대한 인지 및/또는 혼성화에 중요하다. 본 발명의 바람직한 실시형태에서, crRNA는 단형 DR 서열과 연합될 수 있는 짧은 crRNA이다. 본 발명의 다른 실시형태에서, crRNA는 장형 DR 서열 (또는 이중 DR)과 연합될 수 있는 긴 crRNA이다. 본 발명의 측면은 하나 이상의 비-천연 발생 또는 조작 또는 변형 또는 최적화된 핵산 성분을 갖는 Cas13b 이펙터 단백질 복합체에 관한 것이다. 바람직한 실시형태에서 핵산 성분은 RNA를 포함한다. 바람직한 구현예에서, 상기 복합체의 핵산 성분은 직접 반복부 서열에 연결된 가이드 서열을 포함할 수 있으며, 상기 직접 반복부 서열은 하나 이상의 줄기 루프(stem loop) 또는 최적화된 2차 구조를 포함한다. 본 발명의 바람직한 실시형태에서, 직접 반복부는 단형 DR 또는 장형 DR (이중 DR)일 수 있다. 바람직한 구현예에서, 직접 반복부는 변형되어 하나 이상의 단백질-결합 RNA 앱타머를 포함할 수 있다. 바람직한 구현예에서, 최적화된 2차 구조의 일부와 같은 하나 이상의 앱타머가 포함될 수 있다. 이러한 앱타머는 박테리오파지 코트 단백질에 결합할 수 있다. 박테리오파지 코트 단백질은 Qβ, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ΦCb5, ΦCb8r, ΦCb12r, ΦCb23r, 7s 및 PRR1을 포함하는 군으로부터 선택될 수 있다. 바람직한 구현예에서, 박테리오파지 코트 단백질은 MS2이다. 본 발명은 또한 길이가 30 개 이상, 40 개 이상 또는 50 개 이상의 뉴클레오티드인 복합체의 핵산 성분을 제공한다.

[0100] 본 발명은 관심 표적 유전자좌와 연관된 또는 상기 표적 유전자좌에서의 서열을 변형시키거나 게놈 편집하는 방법을 제공하며, 상기 방법은 임의의 원하는 세포 유형, 원핵생물 또는 진핵생물 세포내로 Cas13b 복합체를 도입하여 Cas13b 이펙터 단백질 복합체가 효과적으로 진핵생물 또는 원핵생물 세포 내로 RNA을 간섭하는 작용을 하는 단계를 포함한다. 바람직한 실시형태에서, 세포는 진핵생물 세포이고, RNA는 포유동물 게놈으로부터 전사되거나 또는 포유동물 세포에 존재한다. 인간 세포에서 RNA 편집 또는 게놈 편집의 바람직한 방법에서, Cas13b 이펙터 단백질은 제한없이, 본 명세서에 개시된 Cas13b 이펙터 단백질의 특정 종을 포함할 수 있다.

[0101] 본 발명은 또한 관심 표적 유전자좌를 변형시키는 방법을 제공하고, 방법은 상기 유전자좌에 Cas13b 이펙터 단백질 및 하나 이상의 핵산 성분을 포함하는 비-천연 발생 또는 조작된 조성물을 전달하는 단계를 포함하고, Cas13b 이펙터 단백질은 하나 이상의 핵산 성분과 복합체를 형성하고 상기 복합체가 관심 유전자좌에 결합시 이펙터 단백질은 관심 표적 유전자좌의 변형을 유도한다. 바람직한 실시형태에서, 변형은 가닥 파손의 도입이다.

[0102] 이러한 방법에서, 관심 표적 유전자좌는 RNA 분자 내에 포함될 수 있다. 이러한 방법에서, 관심 표적 유전자좌는 시험관 내에서 RNA 분자에 포함될 수 있다.

[0103] 이러한 방법에서, 관심 표적 유전자좌는 세포 내 RNA 분자에 포함될 수 있다. 세포는 원핵생물 세포 또는 진핵생물 세포일 수 있다. 세포는 포유동물 세포일 수 있다. 본 발명에 의해 세포에 도입된 변형은, 세포 및 세포의 자손이 항체, 전분, 알코올 또는 기타 바람직한 세포 산출물과 같은 생물학적 생성물의 개선된 생산에 대해 변경되도록 할 수 있다. 본 발명에 의해 세포에 도입된 변형은 세포 및 세포의 자손이 생성된 생물학적 생성물을 변화시키는 변경을 포함하도록 하는 것일 수 있다.

[0104] 포유동물 세포는 비인간 포유동물, 예를 들어 영장류, 소, 양, 돼지, 개, 설치류, 토끼과, 예컨대 원숭이, 소, 양, 돼지, 개, 토끼, 래트 또는 마우스 세포일수 있다. 세포는 비-인간 진핵생물 세포, 예컨대 가금류 조류(예를 들어, 닭), 척추동물 어류(예를 들어, 연어) 또는 패류(예를 들어, 굴, 조개, 바닷가재, 새우) 세포일

수 있다. 세포는 식물일 수도 있다. 식물 세포는 단자엽 또는 쌍자엽의 세포, 또는 작물 또는 곡물, 예컨대 카사바, 옥수수, 수수, 대두, 밀, 귀리 또는 쌀의 세포일 수 있다. 식물 세포는 또한 조류, 나무 또는 생산 식물, 과일 또는 야채 (예를 들어, 감귤 나무와 같은 나무, 예를 들어 오렌지, 자몽 또는 레몬 나무; 복숭아 또는 넥타린 나무; 사과 또는 배 나무; 너트 나무, 예컨대 아몬드 또는 호두 또는 피스타치오 나무; 가지속 나무; *브라시카* 속의 식물; *락투카* 속의 식물; *스피나시아* 속의 식물; *캅시쿰* 속의 식물; 면화, 담배, 아스파라거스, 당근, 양배추, 브로콜리, 콜리플라워, 토마토, 가지, 후추, 상추, 시금치, 딸기, 블루베리, 라즈베리, 블랙베리, 포도, 커피, 코코아 등) 일 수 있다.

[0105] 본 발명은 관심 표적 유전자좌를 변형시키는 방법을 제공하고, 방법은 상기 유전자좌에 Cas13b 이펙터 단백질 및 하나 이상의 핵산 성분을 포함하는 비-천연 발생 또는 조작된 조성물을 전달하는 단계를 포함하고, 이펙터 단백질은 하나 이상의 핵산 성분과 복합체를 형성하고 상기 복합체가 관심 유전자좌에 결합시 이펙터 단백질은 관심 표적 유전자좌의 변형을 유도한다. 바람직한 실시형태에서, 변형은 가닥 파손의 도입이다.

[0106] 이러한 방법에서, 관심 표적 유전자좌는 RNA 분자 내에 포함될 수 있다. 바람직한 실시형태에서, 관심 표적 유전자좌는 RNA를 포함하거나 또는 그로 이루어진다.

[0107] 본 발명은 또한 관심 표적 유전자좌를 변형시키는 방법을 제공하고, 방법은 상기 유전자좌에 Cas13b 이펙터 단백질 및 하나 이상의 핵산 성분을 포함하는 비-천연 발생 또는 조작된 조성물을 전달하는 단계를 포함하고, Cas13b 이펙터 단백질은 하나 이상의 핵산 성분과 복합체를 형성하고 상기 복합체가 관심 유전자좌에 결합시 이펙터 단백질은 관심 표적 유전자좌의 변형을 유도한다. 바람직한 실시형태에서, 변형은 가닥 파손의 도입이다.

[0108] 바람직하게, 이러한 방법에서 관심 표적 유전자좌는 시험관 내에서 RNA 분자에 포함될 수 있다. 또한 바람직하게, 이러한 방법에서 관심 표적 유전자좌는 세포 내 RNA 분자에 포함될 수 있다. 세포는 원핵생물 세포 또는 진핵생물 세포일 수 있다. 세포는 포유동물 세포일 수 있다. 세포는 설치류 세포일 수 있다. 세포는 마우스 세포일 수 있다.

[0109] 기재된 방법 중 임의의 것에서, 관심 표적 유전자좌는 관심 게놈 또는 후생유전체 유전자좌일 수 있다. 기재된 방법 중 임의의 것에서, 복합체는 다중화된 사용을 위하여 다수의 가이드와 함께 전달될 수 있다. 기재된 방법 중 임의의 것에서 하나 초과의 단백질(들)이 사용될 수 있다.

[0110] 본 발명의 추가 측면에서, 핵산 성분은 CRISPR RNA (crRNA) 서열을 포함할 수 있다. 이펙터 단백질이 Cas13b 이펙터 단백질이므로, 핵산 성분은 CRISPR RNA (crRNA) 서열을 포함할 수 있고 임의의 트랜스-활성화 crRNA (tracr RNA) 서열을 포함하지 않을 수 있다.

[0111] 임의의 기술된 방법에서, 이펙터 단백질 및 핵산 성분은 단백질 및/또는 핵산 성분(들)을 인코딩하는 하나 이상의 폴리뉴클레오티드 분자를 통해 제공될 수 있고, 여기서 하나 이상의 폴리뉴클레오티드 분자는 단백질 및/또는 핵산 성분(들)을 발현하도록 작동가능하게 구성된다. 하나 이상의 폴리뉴클레오티드 분자는 단백질 및/또는 핵산 성분(들)을 발현하도록 작동가능하게 구성된 하나 이상의 조절 요소를 포함할 수 있다. 하나 이상의 폴리뉴클레오티드 분자는 하나 이상의 벡터 내에 포함될 수 있다. 임의의 기술된 방법에서 관심 표적 유전자좌는 관심 게놈, 후생유전, 또는 전사체 유전자좌일 수 있다. 기재된 방법 중 임의의 것에서, 복합체는 다중화된 사용을 위하여 다수의 가이드와 함께 전달될 수 있다. 기재된 방법 중 임의의 것에서 하나 초과의 단백질(들)이 사용될 수 있다.

[0112] 기재된 방법 중 임의의 것에서, 가닥 파손은 단일 가닥 파손 또는 이중 가닥 파손일 수 있다. 바람직한 실시형태에서, 이중 가닥 파손은 RNA의 2개 부문, 예컨대 단일 가닥 RNA 분자가 그 자체로 폴딩되거나 또는 자기 상보성 서열을 함유하여 RNA의 일부가 폴딩되어 그 자체로 쌍을 형성하게 되는 RNA 분자와 형성된 추정 이중 나선으로 폴딩된 경우 형성된 RNA의 2개 부문의 파손을 의미할 수 있다.

[0113] 조절 요소는 유도 프로모터를 포함할 수 있다. 폴리뉴클레오티드 및/또는 벡터 시스템은 유도 시스템을 포함할 수 있다.

[0114] 기재된 방법 중 임의의 것에서, 하나 이상의 폴리뉴클레오티드 분자가 전달 시스템에 포함될 수 있거나, 하나 이상의 벡터가 전달 시스템에 포함될 수 있다.

[0115] 임의의 기술된 방법에서 비-천연 발생 또는 조작된 조성물은 리포솜, 나노입자를 포함하는 입자, 엑소솜, 미소소낭, 유전자-총 또는 하나 이상의 바이러스 벡터를 통해 전달될 수 있다.

[0116] 본 발명은 또한 본원에서 논의되거나 본원에 기재된 방법 중 임의의 것에서 정의된 바와 같은 특징을 가지는 조성물인 비-천연 발생 또는 조작된 조성물을 제공한다.

[0117] 일정 실시형태에서, 본 발명은 따라서 비-천연 발생 또는 조작된 조성물, 예컨대 특히 관심 표적 유전자좌를 변형시킬 수 있거나 또는 변형시키도록 구성된 조성물을 제공하고, 상기 조성물은 Cas13b 이펙터 단백질 및 하나 이상의 핵산 성분을 포함하며, 여기서 이펙터 단백질은 하나 이상의 핵산 성분과 복합체를 형성하고, 관심 유전자좌에 상기 복합체의 결합시 이펙터 단백질은 관심 표적 유전자좌의 변형을 유도한다. 일정 실시형태에서, 이펙터 단백질은 Cas13b 이펙터 단백질일 수 있다.

[0118] 본 발명은 또한 추가의 측면에서 비-천연 발생 또는 조작된 조성물, 예를 들어 관심 표적 유전자좌를 변형시킬 수 있거나 변형하도록 구성되는 조성물을 제공하며, 상기 조성물은 하기를 포함한다: (a) 가이드 RNA 분자 (또는 가이드 RNA 분자의 조합, 예를 들어, 제 1 가이드 RNA 분자 및 제 2 가이드 RNA 분자) 또는 가이드 RNA 분자를 인코딩하는 핵산 (또는 가이드 RNA 분자의 조합을 인코딩하는 하나 이상의 핵산); (b) Cas13b 이펙터 단백질. 일정 실시형태에서, 이펙터 단백질은 Cas13b 이펙터 단백질일 수 있다.

[0119] 본 발명은 또한 하기를 포함하는 비-천연 발생 또는 조작된 조성물을 제공한다: (I.) (a) 폴리뉴클레오티드 유전자좌에서 표적 서열과 혼성화할 수 있는 가이드 서열, (b) tracr 메이트 서열, 및 (c) tracrRNA 서열을 포함하는, 하나 이상의 CRISPR-Cas 시스템 폴리뉴클레오티드 서열, 및 (II.) Cas13b 이펙터 단백질을 인코딩하는 제2 폴리뉴클레오티드 서열, 여기서 전사될 때, tracr 메이트 서열은 tracrRNA 서열과 혼성화하고, 가이드 서열은 CRISPR 복합체의 서열-특이적 결합을 표적 서열로 유도하며, CRISPR 복합체는 (1) 표적 서열과 혼성화하는 가이드 서열, 및 (2) tracrRNA 서열과 혼성화하는 tracr 메이트 서열과 복합체 형성하는 Cas13b 이펙터 단백질을 포함한다. 일정 실시형태에서, 이펙터 단백질은 Cas13b 이펙터 단백질일 수 있다.

[0120] 일정 실시형태에서, tracrRNA가 요구되지 않을 수 있다. 그러므로, 본 발명은 또한 일정 실시형태에서 하기를 포함하는 비-천연 발생 또는 조작된 조성물을 제공한다: (I.) (a) 폴리뉴클레오티드 유전자좌에서 표적 서열과 혼성화할 수 있는 가이드 서열, 및 (b) 직접 반복부 서열을 포함하는, 하나 이상의 CRISPR-Cas 시스템 폴리뉴클레오티드 서열, 및 (II.) Cas13b 이펙터 단백질을 인코딩하는 제2 폴리뉴클레오티드 서열, 여기서 전사될 때, 가이드 서열은 CRISPR 복합체의 서열-특이적 결합을 표적 서열로 유도하며, CRISPR 복합체는 (1) 표적 서열과 혼성화하는 가이드 서열, 및 (2) 직접 반복부 서열과 복합체 형성하는 Cas13b 이펙터 단백질을 포함한다. 바람직하게, 이펙터 단백질은 Cas13b 이펙터 단백질일 수 있다. 제한없이, 출원인은 이러한 예에서, 직접 반복부 서열이 이펙터 단백질 상에 적재되는 crRNA에 충분한 이차 구조를 포함할 수 있다고 가정한다. 예로서 제한없이, 이러한 이차 구조는 직접 반복부 내에 스템 루프 (예컨대 하나 이상의 스템 루프)를 포함할 수 있거나, 그로 본질적으로 이루어질 수 있거나 또는 그로 이루어질 수 있다.

[0121] 본 발명은 또한 하나 이상의 벡터를 포함하는 벡터 시스템을 제공하고, 하나 이상의 벡터는 임의의 본 명세서에 기술된 방법에 정의된 바와 같은 특징을 갖는 조성물인 비-천연 발생 또는 조작된 조성물의 성분을 인코딩하는 하나 이상의 폴리뉴클레오티드 분자를 포함한다.

[0122] 본 발명은 또한 하나 이상의 벡터 또는 하나 이상의 폴리뉴클레오티드 분자를 포함하는 전달 시스템을 제공하고, 하나 이상의 벡터 또는 폴리뉴클레오티드 분자는 본 명세서에 기술되거나 또는 임의의 본원에 기술된 방법에서 정의된 바와 같은 특징을 갖는 조성물인 비-천연 발생 또는 조작된 조성물의 성분을 인코딩하는 하나 이상의 폴리뉴클레오티드 분자를 포함한다.

[0123] 본 발명은 또한 치료의 치료적 방법에서 사용을 위한 비-천연 발생 또는 조작된 조성물, 또는 상기 조성물의 성분을 인코딩하는 하나 이상의 폴리뉴클레오티드, 또는 상기 조성물의 성분을 인코딩하는 하나 이상의 폴리뉴클레오티드를 포함하는 벡터 또는 전달 시스템을 제공한다. 치료적 처리 방법은 유전자 또는 게놈 편집, 또는 유전자 치료법을 포함할 수 있다.

[0124] 본 발명은 또한 방법 및 조성물을 제공하고, 여기서 이펙터 단백질의 하나 이상의 아미노산 잔기는 변형, 예를 들어 조작되거나 또는 비-천연 발생된 Cas13b 이펙터 단백질일 수 있거나 표 1 단백질을 포함하거나 이것으로 이루어지거나 또는 이것으로 본질적으로 이루어진다. 일 실시형태에서, 변형은 이펙터 단백질의 하나 이상의 아미노산 잔기의 돌연변이를 포함할 수 있다. 하나 이상의 돌연변이는 이펙터 단백질의 하나 이상의 촉매적으로 활성인 도메인에 존재할 수 있다. 이펙터 단백질은 상기 하나 이상의 돌연변이가 없는 이펙터 단백질과 비교하여 감소되거나 제거된 뉴클레아제 활성을 가질 수 있다. 이펙터 단백질은 관심 표적 유전자좌에서 하나의 RNA 가닥의 절단을 유도하지 않을 수 있다. 바람직한 실시형태에서, 하나 이상의 돌연변이는 2

개의 돌연변이를 포함할 수 있다. 바람직한 실시형태에서, 하나 이상의 아미노산 잔기는 Cas13b 이펙터 단백질, 예를 들어 조작되거나 비-천연적으로 발생하는 Cas13b 이펙터 단백질에서 변형된다. 본 발명의 일정 실시형태에서 이펙터 단백질은 하나 이상의 HEPN 도메인을 포함한다. 바람직한 실시형태에서, 이펙터 단백질은 2개의 HEPN 도메인을 포함한다. 다른 바람직한 실시형태에서, 이펙터 단백질은 C-말단에 하나의 HEPN 도메인 및 단백질의 N-말단에 다른 HEPN 도메인을 포함하다. 일정 실시형태에서, 하나 이상의 돌연변이 또는 둘 이상의 돌연변이는 HEPN 도메인을 포함하는 이펙터 단백질의 촉매적 활성 도메인, 또는 HEPN 도메인에 상동성인 촉매적 활성 도메인에 존재할 수 있다. 일정 실시형태에서, 이펙터 단백질은 하나 이상의 하기 돌연변이를 포함한다: R116A, H121A, R1177A, H1182A (여기서 아미노산 위치는 Bergeyella zoohelcum ATCC 43767에서 유래한 그룹 29 단백질의 아미노산 위치에 해당). 당업자는 상이한 Cas13b 단백질 내 상응하는 아미노산 위치가 동일한 효과로 돌연변이될 수 있다는 것을 이해할 것이다. 일정 실시형태예에서, 하나 이상의 돌연변이는 단백질의 촉매 활성을 완전히 또는 부분적으로 폐지한다 (예를 들어, 변형된 절단 속도, 변형된 특이성 등). 일정 실시형태에서, 본원에 기술된 이펙터 단백질은 "사멸" 이펙터 단백질, 예컨대 사멸 Cas13b 이펙터 단백질 (즉, dCas13b) 이다. 일정 실시형태에서, 이펙터 단백질은 HEPN 도메인 1에 하나 이상의 돌연변이를 갖는다. 일정 실시형태에서, 이펙터 단백질은 HEPN 도메인 2에 하나 이상의 돌연변이를 갖는다.

일정 실시형태에서, 이펙터 단백질은 HEPN 도메인 1 및 HEPN 도메인 2에 하나 이상의 돌연변이를 갖는다.이펙터 단백질은 하나 이상의 이종 기능성 도메인을 포함할 수 있다. 하나 이상의 이종 기능성 도메인은 하나 이상의 핵 국재화 신호 (nuclear localization signal; NLS) 도메인을 포함할 수 있다. 하나 이상의 이종성 기능성 도메인은 적어도 둘 이상의 NLS 도메인을 포함할 수 있다. 하나 이상의 NLS 도메인(들)은 이펙터 단백질 (예를 들어, Cas13b 이펙터 단백질)의 말단에 또는 상기 말단 근처에 또는 상기 말단에 근접하여 위치할 수 있으며, 둘 이상의 NLS가 있는 경우, 2 개의 NLS 각각은 이펙터 단백질 (예를 들어, Cas13b 이펙터 단백질)의 말단에 또는 상기 말단 근처에 또는 상기 말단에 근접하여 위치할 수 있다. 하나 이상의 이종성 기능성 도메인은 하나 이상의 전사 활성화 도메인을 포함할 수 있다. 바람직한 구현에에서, 전사 활성 도메인은 VP64를 포함할 수 있다. 하나 이상의 이종성 기능성 도메인은 하나 이상의 전사 억제 도메인을 포함할 수 있다. 바람직한 구현예에서, 전사 억제 도메인은 KRAB 도메인 또는 SID 도메인(예를 들어, SID4X)을 포함한다. 하나 이상의 이종성 기능성 도메인은 하나 이상의 뉴클레아제 도메인을 포함할 수 있다. 바람직한 구현예에서, 뉴클레아제 도메인은 Fok1을 포함한다.

[0125] 본 발명은 또한 다음의 활성 중 하나 이상을 가지는 하나 이상의 이종성 기능성 도메인을 제공한다:메틸라제 활성, 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, 뉴클레아제 활성, 단일-가닥 RNA 절단 활성, 이중-가닥 RNA 절단 활성, 단일-가닥 DNA 절단 활성, 이중-가닥 DNA 활성 및 핵산 결합 활성. 적어도 하나 이상의 이종성 기능성 도메인은 이펙터 단백질의 아미노-말단에 또는 상기 말단 근처에 존재할 수 있고, 있거나 여기서 적어도 하나 이상의 이종성 기능성 도메인은 이펙터 단백질의 카르복시-말단에 또는 상기 말단 근처에 존재한다. 하나 이상의 이종성 기능성 도메인은 이펙터 단백질과 융합될 수 있다. 하나 이상의 이종성 기능성 도메인은 이펙터 단백질에 묶일 수 있다. 하나 이상의 이종성 기능성 도메인은 링커 모이어티에 의해 이펙터 단백질에 연결될 수 있다.

[0126] 일정 실시형태에서, 본 명세서에서 의도하는 바와 같은 Cas13b 이펙터 단백질은 30 내지 40 bp 길이, 보다 전형적으로 34 내지 38 bp 길이, 보다 더 전형적으로 36 내지 37 bp 길이, 예를 들어, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 또는 40 bp 길이의 단형 CRISPR을 포함하는 유전자좌와 연합될 수 있다. 일정 실시형태에서 CRISPR 반복부는 80 내지 350 bp 길이 예컨대 80 내지 200 bp 길이, 보다 더 전형적으로 86 내지 88 bp 길이, 예, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 또는 90 bp 길이의 장형 또는 이중 반복부이다.

[0127] 일정 실시형태에서, 프로토스페이서 인접 모티프(protospacer adjacent motif) (PAM) 또는 PAM-유사 모티프는 관심 표적 유전자좌로 본 명세서에 개시된 바와 같은 이펙터 단백질(예를 들어, Cas13b 이펙터 단백질) 복합체의 결합을 유도한다. 일부 실시형태에서, PAM는 5' PAM (즉, 프로토스페이서의 5' 말단의 상류에 위치)일 수 있다. 다른 실시형태에서, PAM는 3' PAM (즉, 프로토스페이서의 5' 말단의 하류에 위치)일 수 있다. 다른 실시형태에서, 5' PAM 및 3' PAM 둘 모두가 요구된다. 본 발명의 일정 실시형태에서, PAM 또는 PAM-유사 모티프는 이펙터 단백질(예, Cas13b 이펙터 단백질)의 결합을 유도하는데 요구되지 않을 수 있다. 일정 실시형태에서, 5' PAM는 D (즉, A, G, 또는 U)이다. 일정 실시형태에서, 5' PAM는 Cas13b 이펙터에 대해 D이다. 본 발명의 일정 실시형태에서, 반복부 서열에서 절단은 5' 말단에 단형 뉴클레오티드 (예를 들어, 이중 반복부이면 5, 6, 7, 8, 9, 또는 10 nt 또는 그 이상의 길이) 반복부 서열(이는 crRNA "태그"로서 언급될 수 있음) 및 3' 말단에 반복부의 나머지가 측접된 전체 스페이서 서열을 함유하는 crRNA (예를 들어, 단형 또는 장형 crRNA)를 생성시킬 수 있다. 일정 실시형태에서, 본 명세서에 기술된 이펙터 단백질에 의한

표적화는 crRNA 태그 및 표적 5' 측접 서열 간 상동성의 결여를 요구할 수 있다. 이러한 요건은 Samai 등의 문헌 ["Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity" Cell 161, 1164-1174, May 21, 2015]에 더욱 기술된 것과 유사할 수 있는데, 여기서 요건은 CRISPR 어레이 자체로부터 침입 핵산 상의 진짜 표적을 구별하는 것으로 여겨지며, 반복부 서열의 존재는 crRNA 태그와 완전한 상동성을 야기하여 자가면역을 방지하게 될 것이다.

[0128] 일정 실시형태에서, Cas13b 이펙터 단백질은 조작되고 뉴클레아제 활성을 감소시키거나 또는 제거시키는 하나 이상의 돌연변이를 포함할 수 있어서, RNA 간섭 활성을 감소시키거나 또는 제거한다. 돌연변이는 또한 인접한 잔기에서, 예를 들어 뉴클레아제 활성에 참여하는 것들 가까이의 아미노산에서 만들어질 수 있다. 일부 실시형태에서, 하나 이상의 추정 촉매적 뉴클레아제 도메인은 불활성화되고 이펙터 단백질 복합체는 절단 활성이 결여되고, RNA 결합 복합체로서 기능한다. 바람직한 실시형태에서, 최종 RNA 결합 복합체는 본 명세서에 기술된 바와 같은 하나 이상의 기능성 도메인과 연결될 수 있다.

[0129] 일정 실시형태에서, 하나 이상의 기능성 도메인는 제어가능하며, 즉 유도가능하다.

[0130] 본 발명의 특정 구현예에서, 가이드 RNA 또는 성숙 crRNA는 직접 반복부 서열 및 가이드 서열 또는 스페이서 서열을 포함하거나, 본질적으로 이루어지거나, 이루어진다. 특정 구현예에서, 가이드 RNA 또는 성숙 crRNA는 가이드 서열 또는 스페이서 서열에 연결된 직접 반복부 서열을 포함하거나, 본질적으로 이루어지거나, 이루어진다. 본 발명의 바람직한 구현예에서, 성숙 crRNA는 스템 루프 또는 최적화된 스템 루프 구조 또는 최적화된 2차 구조를 포함한다. 바람직한 구현예에서, 성숙 crRNA는 직접 반복부 서열에서 스템 루프 또는 최적화된 스템 루프 구조를 포함하며, 여기서 스템 루프 또는 최적화된 스템 루프 구조는 절단 활성에 중요하다. 특정 구현예에서, 성숙 crRNA는 바람직하게 단일 스템 루프를 포함한다. 특정 구현예에서, 직접 반복부 서열은 바람직하게 단일 스템 루프를 포함한다. 특정 구현예에서, 이펙터 단백질 복합체의 절단 활성은 스템 루프 RNA 이중나선 구조에 영향을 미치는 돌연변이를 도입함으로써 변형된다. 바람직한 구현예에서, 스템 루프의 RNA 이중나선을 유지하는 돌연변이가 도입될 수 있으며, 이에 의해 이펙터 단백질 복합체의 절단 활성이 유지된다. 다른 바람직한 구현예에서, 스템 루프의 RNA 이중나선 구조를 파괴하는 돌연변이가 도입될 수 있으며, 이에 의해 이펙터 단백질 복합체의 절단 활성이 완전히 없어진다.

[0131] 본원에 제공된 CRISPR 시스템은 crRNA 또는 가이드 서열을 포함하는 유사한 폴리뉴클레오티드를 사용할 수 있으며, 여기서 폴리뉴클레오티드는 RNA, DNA 또는 RNA와 DNA의 혼합물이고/거나, 폴리뉴클레오티드는 하나 이상의 뉴클레오티드 유사체를 포함한다. 서열은 벌지 (bulge), 헤어핀 또는 스템 루프 구조와 같은 천연 crRNA의 구조를 포함하지만 이에 제한되지 않는 임의의 구조를 포함할 수 있다. 일정 실시형태에서, 가이드 서열을 포함하는 폴리뉴클레오티드는 RNA 또는 DNA 서열일 수 있는 제 2 폴리 뉴클레오티드 서열과의 이중체를 형성한다.

[0132] 일정 실시형태에서, 방법은 화학적으로 변형된 가이드 RNA를 사용한다. 가이드 RNA 화학 변형의 예로는, 이로 제한되지는 않지만, 하나 이상의 말단 뉴클레오티드에서 2'-O-메틸(M), 2'-O-메틸 3'포스포로티오에이트(MS), 또는 2'-O-메틸 3'티오PACE(MSP)의 혼입을 포함한다. 이러한 화학적으로 변형된 가이드 RNA는, 온-표적 (on-target; 표적상, 표적-적중) 특이성 대 오프-표적 (off-target; 표적외, 표적-외) 특이성은 예측가능하지 않지만, 비변형된 가이드 RNA에 비하여 증가된 안정성 및 증가된 활성을 포함할 수 있다. (참고, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015). 화학적으로 변형된 가이드 RNA는 포스포로티오에이트 연결 및 리보스 고리의 2 '및 4' 탄소 사이의 메틸렌 브릿지를 포함하는 잠금 핵산 (LNA) 뉴클레오티드를 갖는 RNA를 제한없이 포함한다.

[0133] 본 발명은 또한 본 명세서에 기재된 임의의 방법 또는 조성물에서 진핵생물 또는 진핵생물 세포에서의 발현을 위해 코돈 최적화된 이펙터 단백질을 인코딩하는 뉴클레오티드 서열을 제공한다. 본 발명의 실시형태에서, 코돈 최적화된 이펙터 단백질은 본 명세서에 기술된 임의의 Cas13b 이펙터 단백질이고 진핵생물 세포 또는 유기체, 예를 들어 본 명세서의 다른 곳에서 언급된 이러한 세포 또는 유기체, 예를 들어, 제한없이, 마우스 세포, 래트 세포, 및 인간 세포 또는 인간이외 진핵생물 유기체, 예를 들어 식물을 포함하는, 효모 세포, 또는 포유동물 세포 또는 유기체에서 작동능을 위해 코돈 최적화된다.

[0134] 본 발명의 특정 구현예에서, 적어도 하나의 핵 국재화 신호(nuclear localization signal; NLS)가 Cas13b 이펙터 단백질을 인코딩하는 핵산 서열에 부착된다. 바람직한 구현예에서, 적어도 하나 이상의 C-말단 또는 N-말단 NLS가 부착된다(그리고 이에 따라 Cas13b 이펙터 단백질을 인코딩하는 핵산 분자(들)는, 발현된 생성물이 부착되거나 연결된 NLS(들)를 갖도록 NLS(들)에 대한 코딩을 포함할 수 있다). 바람직한 구현예에서, C-말

단 NLS는 진핵생물 세포, 바람직하게는 인간 세포에서 최적 발현 및 핵 표적화를 위해 부착된다. 본 발명은 또한 다수의 핵산 성분을 전달하기 위한 방법을 포함하며, 여기서 각각의 핵산 성분은 상이한 관심 표적 유전자좌에 특이적이고, 이에 의해 다수의 관심 표적 유전자좌를 변형시킨다. 복합체의 핵산 성분은 하나 이상의 단백질-결합 RNA 앱타머를 포함할 수 있다. 하나 이상의 앱타머는 박테리오파지 코트 단백질을 결합할 수 있다. 박테리오파지 코트 단백질은 Qβ, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ΦCb5, ΦCb8r, ΦCb12r, ΦCb23r, 7s 및 PRR1을 포함하는 군으로부터 선택될 수 있다. 바람직한 구현예에서, 박테리오파지 코트 단백질은 MS2이다. 본 발명은 또한 길이가 30 개 이상, 40 개 이상 또는 50 개 이상의 뉴클레오티드인 복합체의 핵산 성분을 제공한다.

[0135] 추가 측면에서, 본 발명은 변형된 관심 표적 유전자좌를 포함하는 진핵생물 세포를 제공하고, 여기서 관심 표적 유전자좌는 임의의 본 명세서에 기술된 방법에 따라 변형된다. 추가 측면은 상기 세포의 세포주를 제공한다. 다른 측면은 하나 이상의 상기 세포를 포함하는 다수세포 유기체를 제공한다.

[0136] 알정 실시형태에서, 관심 표적 유전좌위의 변형은 하기를 산출할 수 있다: 적어도 하나의 유전자 산물의 변경된 발현을 포함하는 진핵생물 세포; 적어도 하나의 유전자 산물의 변형된 발현을 포함하는 진핵생물 세포, 적어도 하나의 유전자 산물의 발현은 증가됨; 적어도 하나의 유전자 산물의 변형된 발현을 포함하는 진핵생물 세포, 적어도 하나의 유전자 산물의 발현은 감소됨; 또는 편집된 게놈을 포함하는 진핵생물 세포.

[0137] 일정 실시형태에서, 진핵생물 세포는 포유동물 세포 또는 인간 세포일 수 있다.

[0138] 추가 실시형태에서, 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템은, 부위-특이적 유전자 녹아웃; 부위-특이적 게놈 편집; RNA 서열-특이적 간섭; 또는 멀티플렉스 게놈 조작에 사용될 수 있다.

[0139] 또한 본 명세서에 기술된 바와 같은 세포, 세포주 또는 유기체 유래의 유전자 산물을 제공한다. 일정 실시형태에서, 발현된 유전자 산물의 양은 변경된 발현 또는 편집된 게놈을 갖지 않는 세포 유래의 유전자 산물의 양을 초과하거나 또는 그 미만일 수 있다. 일정 실시형태에서, 유전자 산물은 변경된 발현 또는 편집된 게놈을 갖지 않는 세포 유래의 유전자 산물과 비교하여 변경될 수 있다.

[0140] 또 다른 측면에서, 본 발명은 하기 단계를 포함하는, 신규 핵산 변형 이펙터를 식별하는 방법을 제공한다: 정의된 크기 한계 초과의 적어도 하나의 단백질, 또는 둘 다를 포함하는, 유전자좌의 보존된 게놈 엘리먼트로부터 정의된 거리 내에 있는 추정 핵산 변형 효소 유전자좌를 인코딩하는 핵산 서열 세트로부터 추정 핵산 변형 유전자좌를 확인하는 단계; 동종 단백질을 포함하는 서브세트로 확인된 추정 핵산 변형 유전자좌를 그룹화하는 단계; 다음 중 하나 이상에 기초하여 하나 이상의 서브세트로부터 핵산 변형 유전자좌를 선택함으로써 후보 핵산 변형 유전자좌의 최종 세트를 확인하는 단계; 다른 서브세트 내의 유전자좌에 비해 공지된 단백질 도메인과 낮은 도메인 상동성 매치를 갖는 추정 이펙터 단백질을 갖는 유전자좌를 포함하는 서브 세트, 다른 서브세트 내의 유전자좌에 비해 보존된 게놈 엘리먼트와 최소 거리를 갖는 추정 단백질을 포함하는 서브세트, 다른 서브세트 내의 큰 이펙터 단백질에 비해 추정 인접 보조 단백질과 동일한 방향을 갖는 큰 이팩터 단백질을 포함하는 유전자좌를 갖는 서브세트, 다른 유전자좌에 비해 기존의 핵산 변형 분류가 낮은 추정 이펙터 단백질을 포함하는 서브세트, 다른 서브세트에 비해 공지된 핵산 변형 유전자좌에 대해 낮은 근접성을 갖는 유전자좌를 포함하는 서브세트, 및 각 서브세트 내의 후보 유전자좌의 총 수.

[0141] 일 실시형태에서, 핵산 서열의 세트는 게놈 또는 메타게놈 데이타베이스, 예컨대 원핵생물 게놈 또는 메타게놈 서열을 포함하는 게놈 또는 메타게놈 데이타베이스로부터 수득된다.

[0142] 일 실시형태에서, 보존된 게놈 엘리먼트로부터 정해진 거리는 1 kb 내지 25 kb이다.

[0143] 일 실시형태에서, 보존된 게놈 엘리먼트는 반복적 엘리먼트, 예컨대 CRISPR 어레이를 포함한다. 특정 실시형태에서, 보존된 게놈 엘리먼트로부터 정해진 거리는 CRISPR 어레이의 10 kb 이내이다.

[0144] 일 실시형태에서, 추정 핵산 변형 (이팩터) 유전자좌 내에 포함된 단백질의 정해진 크기 한계치는 200개 아미노산을 초과하거나, 또는 그 이상이고, 특히 정해진 크기 한계치는 700개 아미노산을 초과한다. 일 실시형태에서, 추정 핵산 변형 유전자좌는 900 내지 1800 개의 아미노산이다.

[0145] 일 실시형태에서, 보존된 게놈 엘리먼트는 핵산 세트의 반복 또는 패턴 발견 분석, 예컨대 PILER-CR을 사용하여 동정된다.

[0146] 일 실시형태에서, 본 명세서에 기술된 방법의 그룹분류 단계는 적어도 부분적으로, 도메인 상동성 검색 또는

HHpred 단백질 도메인 상동성 검색의 결과를 기반으로 한다.

[0147] 일 실시형태에서, 정의된 한계치는 0 내지 1e-7의 BLAST 최근접 이웃 컷오프값이다.

[0148] 일 실시형태에서, 본 명세서에 기술된 방법은 900 내지 1800개 아미노산의 추정 단백질을 갖는 유전자좌만을 오직 포함하는 필터링 단계를 더 포함한다.

[0149] 일 실시형태에서, 본 명세서에 기술된 방법은 핵산 변형 이펙터를 인코딩하는 핵산 구축물의 세트를 생성시키는 단계, 및 예컨대 박테리아 콜로니에서 PAM 검증, 시험관내 절단 어세이, 서베이어 방법, 포유동물 세포에서의 실험, PFS 검증, 또는 이의 조합의 사용을 통해서, 하나 이상의 생화학적 검증 어세이를 수행하는 단계를 포함하는, 후보 핵산 변형 이펙터의 핵산 변형 기능의 실험 검증을 더 포함한다.

[0150] 일 실시형태에서, 본 명세서에 기술된 방법은 동정된 핵산 변형 유전자좌로부터 하나 이상의 단백질을 포함하는 비-천연 발생 또는 조작된 조성물을 제조하는 단계를 더 포함한다.

[0151] 일 실시형태에서, 동정된 유전자좌는 클래스 2 CRISPR 이펙터를 포함하거나, 또는 동정된 유전자좌는 Cas1 또는 Cas2가 결여되거나, 또는 동정된 유전자좌는 단일 이펙터를 포함한다.

[0152] 일 실시형태에서, 단일 거대 이펙터 단백질은 900개 아미노산 길이를 초과하거나, 또는 1100개 아미노산 길이를 초과하거나, 또는 적어도 하나의 HEPN 도메인을 포함한다.

[0153] 일 실시형태에서, 적어도 하나의 HEPN 도메인은 이펙터 단백질의 N-말단 또는 C-말단에 가까이 있거나, 또는 이펙터 단백질의 내부 위치에 위치된다.

[0154] 일 실시형태에서, 단일 거대 이펙터 단백질은 N-말단 및 C-말단에 HEPN 도메인을 포함하고, 2개 HEPN 도메인은 단백질의 내부에 있다.

[0155] 일 실시형태에서, 동정된 유전자좌는 CRISPR 어레이의 2 kb 내지 10 kb 이내에 하나 이상의 소형 추정 부속 단백질을 더 포함한다.

[0156] 일 실시형태에서, 소형 부속 단백질은 700 개 미만의 아미노산이다. 일 실시형태에서, 소형 부속 단백질은 50 내지 300 개 길이의 아미노산이다.

[0157] 일 실시형태에서, 소형 부속 단백질은 다수의 예측된 경막 도메인을 포함하거나, 또는 4개의 예측된 경막 도메인을 포함하거나, 또는 적어도 하나의 HEPN 도메인을 포함한다.

[0158] 일 실시형태에서, 소형 부속 단백질은 적어도 하나의 HEPN 도메인 및 적어도 하나의 경막 도메인을 포함한다.

[0159] 일 실시형태에서, 유전자좌는 CRISPR 어레이로부터 25 kb 밖으로 추가 단백질을 포함하지 않는다.

[0160] 일 실시형태에서, CRISPR 어레이는 약 36개 뉴클레오티드 길이를 포함하는 직접 반복부 서열을 포함한다. 특정 실시형태에서, 직접 반복부는 3' 말단의 CAAC에 역 상보적인 5' 말단의 GTTG/GUUG를 포함한다.

[0161] 일 실시형태에서, CRISPR 어레이는 약 30개 뉴클레오티드 길이를 포함하는 스페이서 서열을 포함한다.

[0162] 일 실시형태에서, 동정된 유전자좌는 소형 부속 단백질이 결여된다.

[0163] 본 발명은 다음을 포함하는 신규 CRISPR 이펙터를 식별하는 방법을 제공한다: a) CRISPR 어레이를 인코딩하는 게놈 또는 메타게놈 데이터베이스에서 서열을 식별하는 단계; b) CRISPR 어레이의 10kb 내의 상기 선택된 서열에서 하나 이상의 오픈 리딩 프레임 (ORF)을 식별하는 단계; c) 크기가 900-1800 개 아미노산인 추정 CRISPR 이펙터 단백질의 존재에 기초하여 유전자좌를 선택하는 단계, d) 50-300 개 아미노산의 추정 보조 단백질을 인코딩하는 유전자좌를 선택하는 단계; 및 e) 추정 CRISPR 이펙터 및 CRISPR 보조 단백질을 인코딩하는 유전자좌를 확인하고 선택적으로 구조 분석에 기초하여 이들을 분류하는 단계.

[0164] 일 실시형태에서, CRISPR 이펙터는 유형 VI CRISPR 이펙터이다. 일 실시형태에서, 단계 (a)는 i) 게놈 및/또는 메타게놈 데이터베이스의 서열을 CRISPR 어레이를 인코딩하는 적어도 하나의 사전-식별된 시드 서열과 비교하고, 상기 시드 서열을 포함하는 서열을 선택하는 단계; 또는 ii) CRISPR 알고리즘에 기초하여 CRISPR 어레이를 식별하는 단계.

[0165] 일 실시형태에서, 단계 (d)는 뉴클레아제 도메인을 식별하는 것을 포함한다. 일 실시형태에서, 단계 (d)는 RuvC, HPN, 및/또는 HEPN 도메인을 식별하는 것을 포함한다.

[0166] 일 실시형태에서, Cas1 또는 Cas2를 인코딩하는 ORF는 CRISPR 어레이의 10kb 내에 존재하지 않는다.

[0167] 일 실시형태에서, 단계 (b)에서의 ORF는 50-300 아미노산의 추정 보조 단백질을 인코딩한다.

[0168] 일 실시형태에서, 단계 (d)에서 수득된 추정 신규 CRISPR 이펙터는 게놈 및/또는 메타게놈 서열을 추가로 비교하고 제 1 항의 단계 a) 내지 d)에 기재된 관심 유전자좌를 후속 선택하기 위한 시드 서열로서 사용된다. 일 실시형태에서, 사전-식별된 시드 서열은 다음을 포함하는 방법에 의해 수득된다: (a) 게놈 또는 메타게놈 데이터베이스에서 CRISPR 모티프를 식별하는 단계, (b) 상기 식별된 CRISPR 모티프에서 여러 기능을 추출하는 단계, (c) 비지도 학습을 사용하여 CRISPR 유전자좌를 분류하는 단계, (d) 상기 분류를 기반으로 보존된 유전자좌 요소를 식별하는 단계, 및 e) 시드 서열로서 적합한 추정 CRISPR 이펙터를 선택하는 단계.

[0169] 일 실시형태에서, 특징은 단백질 엘리먼트, 반복 구조, 반복 서열, 스페이서 서열 및 스페이서 맵핑을 포함한다. 일 실시형태에서, 게놈 및 메타게놈 데이터베이스는 박테리아 및/또는 고세균 게놈이다. 일 실시형태에서, 게놈 및 메타게놈 서열은 Ensembl 및/또는 NCBI 게놈 데이터베이스로부터 수득된다. 일 실시형태에서, 단계 (d)에서의 구조 분석은 2 차 구조 예측 및/또는 서열 정렬에 기초한다. 일 실시형태에서, 단계 (d)는 이들이 인코딩하는 단백질 및 수득된 클러스터의 수동 큐레이션에 기초하여 나머지 유전자좌를 클러스터링함으로써 달성된다.

[0170] 따라서, 본 발명의 목적은 출원인이 권리를 보유하도록 임의의 이전에 공지된 생성물, 그 생성물의 제조 공정, 또는 그 생성물의 사용 방법을 본 발명 내에 포함시키지 않으며, 이에 의하여 임의의 이전에 공지된 생성물, 공정, 또는 방법의 권리 포기를 개시한다. 출원인(들)이 권리를 보유하고 이에 의해 이미 기재된 임의의 생성물, 생성물의 제조 방법 또는 생성물의 사용 방법의 포기를 개시하는, USPTO (35 U.S.C. 112, 제 1 단락) 또는 EPO (EPC 제 83 조) 의 명시적 기재 및 실시 가능 요건을 충족시키는, 이미 개시된 임의의 생성물, 생성물 제조 방법 또는 생성물 사용 방법을 본 발명의 범위 내에 포함시키는 의도가 아님이 추가로 주목된다. 조항 53(c) EPC 및 규칙 28(b) 및 (c) EPC에 따르는 것이 본 발명의 실시에 유리할 수 있다. 본 명세서에서 어떤 것도 약속으로 해석되어서는 안 된다.

[0171] 이러한 개시 및 특히 청구항 및/또는 단락에서, 용어 예컨대 "포함하다", "포함된", "포함하는" 등은 미국 특허법에서 그에 부여하는 의미를 가질 수 있다는 것을 주의하며, 예를 들어 그들은 "포괄하다", "포괄되는", "포괄하는" 등을 의미할 수 있으며, 용어 예컨대 "본질적으로 이루어지는" 및 "본질적으로 이루어지다"는 미국 특허법에서 그들에 부여하는 의미를 가지며, 예를 들어 그들은 성분들을 명백하게 인용하지 않는 것을 허용하지만, 종래에 존재하였거나 또는 본 발명의 기본 또는 신규 특징에 영향을 미치는 성분들은 배제한다.

[0172] 이들 및 다른 구현예는 하기 상세한 설명에 개시되어 있거나 이로부터 명백 해지고 이에 의해 포함된다.

도면의 간단한 설명

[0173] 본 발명의 신규 특징은 특히 첨부된 청구범위에 제시되어 있다. 본 발명의 원리가 이용된 예시적인 구현예를 제시하는 하기의 상세한 설명을 참조함으로써 본 발명의 특징 및 이점을 더욱 잘 이해할 것이며, 첨부된 도면은 다음과 같다:

도 1A-1B는 Cas13b 오르소로그의 트리 정렬을 보여준다.

도 2A-2C는 C2c2 및 Cas13b 오르소로그의 트리 정렬을 보여준다.

도 3은 이. 콜라이에서의 활성에 대한 Cas13b 오르소로그의 시험의 예시적인 결과를 보여주며, 이에 의해 B. 주헬컴으로부터의 Cas13b의 도입은 빈 벡터의 도입과 비교한다.

도 4는 표 1의 상이한 오르소로그로 수득된 특정 RNA 절단 활성의 일반적인 비교를 보여준다.

도 5는 도 1 및 2에 제공된 바와 같은 상이한 Cas13b 오르소로그의 정렬을 보여준다.

도 6A-6B (A)는 RNA 간섭을 시험하고 PFS를 확인하기 위한 MS2 파지 드롭 플라크 어세이를 위한 프로토스페이서 디자인을 보여준다. (B)는 RNA 간섭 어세이 개략도를 보여준다. 표적 서열은 암피실린 내성을 부여하는 전사된 bla 유전자의 출발부에서 인프레임으로 위치되거나 또는 동일한 표적 플라스미드의 반대 가닥 상의 비전사된 영역에 위치된다. 표적 플라스미드는 bzcas13b 플라스미드 또는 클로람페니콜 내성을 부여하는 빈 벡터와 공동형질전환되었고 이중 선택 항생제 플레이트 상에 플레이팅되었다. 감손된 콜로니가 확인되었고 상응하는 표적이 PFS 식별을 위해 서열분석되었다.

도 7은 csx27 보조 단백질의 존재 및 부재 하에서 오르소로그 1, 13 및 16에 대해 안전하게 감손된 스페이서로부터 정규화된 PFS 점수의 히트맵을 보여준다.

도 8은 csx28 보조 단백질의 존재 및 부재 하에서 오르소로그 2, 3, 8, 9, 14, 19 및 21에 대해 안전하게 감손된 스페이서로부터 정규화된 PFS 점수의 히트맵을 보여준다.

도 9는 오르소로그 5, 6, 7, 10, 12 및 15에 대해 안전하게 감손된 스페이서로부터 정규화된 PFS 점수의 히트맵을 보여준다.

도 10A-10BB는 상이한 오르소로그 및 유도된 PFS에 대해 안전하게 감손된 스페이서로부터 정규화된 PFS 점수의 히트맵을 보여준다.

도 11은 시험된 가이드와 함께 포유동물 세포에서 덜 활성인 Cas13b 오르소로그에 대한 루시퍼라제 간섭 데이터의 개요를 제공한다.

도 12는 시험된 가이드와 함께 포유동물 세포에서 낮은 내지 중간 활성을 나타낸 Cas13b 오르소로그에 대한 루시퍼라제 간섭 데이터의 개요를 제공한다.

도 13은 시험된 가이드와 함께 포유동물 세포에서 상당한 활성을 나타낸 Cas13b 오르소로그 일부에 대한 루시퍼라제 간섭 데이터의 개요를 제공한다.

도 14A-14B는 시험된 가이드와 함께 포유동물 세포에서 상당한 활성을 나타낸 Cas13b 오르소로그 일부에 대한 루시퍼라제 간섭 데이터의 개요를 제공한다.

도 15는 시험된 가이드와 함께 포유동물 세포에서 상당한 활성을 나타낸 Cas13b 오르소로그 일부에 대한 루시퍼라제 간섭 데이터의 개요 및 C2c2 활성과의 비교를 제공한다.

도 16은 동일한 가이드 서열과 함께 진핵생물 세포에서 상당한 활성을 갖는 오르소로그로부터의 복합 데이터를 보여준다.

도 17A-17G는 2 개의 상이한 리포터 유전자, G-루시퍼라제 및 C-루시퍼라제를 사용하는 포유동물 세포에서 Cas13b 오르소로그의 부수 효과를 보여준다.

도 18은 RNA 녹다운에 대한 고 활성 Cas13b 오르소로그의 특징분석을 보여준다. A) 정형화된 Cas13 유전자좌 및 상응하는 crRNA 구조의 개략도. B) 2 개의 상이한 가이드를 사용하는 루시페라제 녹다운에 대한 19 Cas13a, 15 Cas13b, 및 7 Cas13c 오르소로그의 평가. 두 가이드를 모두 사용하는 효율적인 녹다운이 있는 오르소로그에 숙주 유기체 이름으로 표지한다. C) PspCas13b 및 LwaCas13a 녹다운 활성은 Gluc에 대해 가이드를 타일링하고 루시퍼라제 발현을 측정함으로써 비교된다. D) PspCas13b 및 LwaCas13a 녹다운 활성은 Cluc에 대해 가이드를 타일링하고 루시퍼라제 발현을 측정함으로써 비교된다. E) LwaCas13a (적색) 및 shRNA (검정색)에 대한 Gluc-표적화 조건 (y-축)과 비교하여 비-표적화 대조 (x-축)의 RNA-seq 라이브러리에서 검출된 모든 유전자의 log2 (백만 당 전사물 (TPM)) 값의 발현 수준. 세 개의 생물학적 복제물의 평균이 제시된다. Gluc 전사물 데이터 포인트가 표지된다. F) PspCas13b (청색) 및 shRNA (검정색)에 대한 Gluc-표적화 조건 (y-축)과 비교하여 비-표적화 대조 (x-축)의 RNA-seq 라이브러리에서 검출된 모든 유전자의 log2 (백만 당 전사물 (TPM)) 값의 발현 수준. 세 개의 생물학적 복제물의 평균이 제시된다. Gluc 전사물 데이터 포인트가 표지된다. G) E 및 F의 전사체 전반 분석에서 LwaCas13a, PspCas13b 및 shRNA에 대한 Gluc 녹다운에서 유의한 오프-표적의 수.

도 19는 RNA 편집을 위한 dCas13b-ADAR 융합의 조작을 보여준다. A) dCas13b-ADAR 융합 단백질에 의한 RNA 편집의 개략도. B) *사이프리디나* 루시퍼라제 W85X 표적 및 표적화 가이드 디자인의 개략도. C) 길이 30, 50, 70, 또는 84 nt의 타일링 가이드와 함께 Cas13b-dADAR1 (왼쪽) 및 Cas13b-ADAR2-cd (오른쪽)에 대한 루시퍼라제 활성 복원의 정량. D) *사이프리디니아* 루시퍼라제 W85X 표적화를 위한 표적 부위의 개략도. E) *사이프리디니아* 루시퍼라제 W85X 표적화를 위한 50 nt 가이드의 A->I 편집의 서열분석 정량.

도 20은 REPAIRv1에 의한 RNA 편집을 위한 서열 유연성 측정을 보여준다. A) REPAIRv1에 의한 RNA 편집의 PSP (Protospacer Flanking Site: 프로토스페이서 측면 부위) 선호도를 결정하기 위한 스크린의 개략도. B) 2 개의 상이한 편집 부위에서 모든 4-N PFS 조합에 대한 RNA 편집 효율의 분포 C) 모든 가능한 3 개 염기 모티프에 걸친 Cluc W85에서 REPAIRv1의 편집 백분율의 정량. D) 모든 가능한 3 개 염기 모티프에 대한 Cluc W85에서 RNA 편집의 5 '및 3' 염기 선호도의 히트맵

도 21은 REPAIRv1에 의한 질환-관련 돌연변이의 교정을 보여준다. A) *AVPR2* 878G>A 표적화를 위한 표적 및 가이드 디자인의 개략도. B) *AVPR2* 중의 878G>A 돌연변이는 3 개의 상이한 가이드 디자인과 함께 REPAIRv1을 사용하여 백분율을 변화시키도록 교정되었다. C) *FANCC* 1517G>A 표적화를 위한 표적 및 가이드 디자인의 개략도. D) *FANCC* 중의 1517G>A 돌연변이는 3 개의 상이한 가이드 디자인과 함께 REPAIRv1을 사용하여 백분율을 변화시키도록 교정되었다. E) REPAIRv1을 사용하여 34 개의 상이한 질병-관련 G>A 돌연변이의 백분율 편집의 정량. F) ClinVar 데이터베이스에 의해 주석처리된 바와 같이 교정될 수 있는 모든 가능한 G>A 돌연변이의 분석. G) ClinVar에서 모든 G>A 돌연변이에 대한 편집 모티프의 분포는 Gluc 전사물에서 정량화된 모티브 당 REPAIRv1에 의한 편집 효율에 비해 제시된다.

도 22는 REPAIRv1의 특이성의 특징분석을 보여준다. A) *KRAS* 표적 부위 및 가이드 디자인의 개략도. B) 타일화 *KRAS*-표적화 가이드에 대한 퍼센트 편집의 정량. 편집 백분율은 표적-적중 및 인접 아데노신 부위에 제시된다. 각 가이드에 대해, 이중 RNA의 영역은 빨간색 사각형으로 표시된다. C) Cluc 표적화 가이드와 함께 REPAIRv1에 의한 상당한 RNA 편집의 전사체-전반 부위. 표적-적중 부위 Cluc 부위 (254 A>G)가 주황색으로 강조 표시된다. D) 비-표적화 가이드와 함께 REPAIRv1에 의한 상당한 RNA 편집의 전사체-전반 부위.

도 23은 REPAIRv1의 특이성을 개선하기 위한 ADAR2의 유리한 돌연변이유발을 보여준다. A) 다양한 dCas13-ADAR2 돌연변이체에 의한 루시퍼라제 신호 복원의 정량 및 그의 특이성 점수는 핵심 ADAR2 데아미나제 잔기와 dsRNA 표적 사이의 접촉에 대한 개략도를 따라 플롯팅되었다. 특이성 점수는 표적화 가이드와 비-표적화 가이드 조건 사이의 루시퍼라제 신호의 비율로 정의된다. B) 다양한 dCas13-ADAR2 돌연변이체에 의한 루시퍼라제 신호 복원의 정량 대 이들의 특이성 점수. C) mRNA의 전사체 전반 서열분석에 의해 각 dCas13-ADAR2 돌연변이체에 대한 표적-적중 편집 분율 및 유의한 표적-외 수의 측정. D) Cluc 내 사전종결 부위를 표적으로 하는 가이드와 함께 REPAIRv1 및 REPAIRv2에 의한 상당한 RNA 편집의 전사체-전반 부위. 표적-적중 Cluc 부위 (254 A>G)가 주황색으로 강조 표시된다. E) REPAIRv1 및 REPAIRv2 사이의 표적-외 편집에서의 차이점을 강조하는 표적-적중 Cluc 편집 부위 (254 A>G)를 둘러싼 RNA 서열분석 판독. 모든 A>G 편집은 빨간색으로 강조 표시되는 반면, 서열분석 오류는 파란색으로 강조 표시된다. F) 내인성 *KRAS* 및 *PPIB* 전사물에서 프레임-외 UAG 부위를 표적으로 하는 가이드와 함께 REPAIRv1 및 REPAIRv2에 의한 RNA 편집. 표적-적중 편집 비율은 각 조건 행의 오른쪽에 측면 막대 차트로 표시된다. 가이드 RNA에 의해 형성된 이중체 영역은 적색 외곽선 박스로 표시된다.

도 24는 *생체 내* 효율성 및 PFS 결정에 대한 Cas13b 오르소로그의 박테리아 스크리닝을 보여준다. A) Cas13b 오르소로그의 PFS를 결정하기 위한 박테리아 분석의 개략도. 베타-락타마제 표적화 스페이서와 함께 Cas13b 오르소로그는 베타-락타마제 발현 플라스미드와 공동-형질전환되고 이중 선별에 적용된다. B) 콜로니 형성 단위 (cfu)에 의해 측정된 베타-락타마제를 표적으로 하는 Cas13b 오르소로그의 간섭 활성의 정량. C) 박테리아 어세이로부터 감손된 서열에 의해 결정된 Cas13b 오르소로그에 대한 PFS 로고.

도 25는 Cas13b 녹다운의 최적화 및 미스매치 특이성의 추가 특징분석을 나타낸다. A) 2 개의 상이한 가이드와 함께 Gluc 녹다운은 다양한 핵 국재화 및 핵 유출 태그에 융합된 상단 2 Cas13a 및 상단 4 Cas13b 오르소로그를 사용하여 측정된다. B) *KRAS* 의 녹다운은 4 개의 상이한 가이드와 함께 LwaCas13a, RanCas13b, PguCas13b, 및 PspCas13b에 대해 측정되고 4 개의 위치-매칭된 shRNA 대조군과 비교하였다. C) LwaCas13a 및 PspCas13b 녹다운의 특이성을 평가하기 위해 사용된 단일 및 이중 미스매치 플라스미드 라이브러리의 개략도. 모든 가능한 단일 및 이중 미스매치는 표적 서열뿐만 아니라 표적 부위의 5 '및 3'말단에 직접 인접한 3 개의 위치에 존재한다. D) 표시된 단일 미스매치가 있는 전사체의 감손 수준은 LwaCas13a 및 PspCas13b 조건 모두에 대한 히트맵으로 플롯팅된다. E) 표시된 이중 미스매치가 있는 전사체의 감손 수준은 LwaCas13a 및 PspCas13b 조건 모두에 대한 히트맵으로 플롯팅된다.

도 26은 dCas13-ADAR2 RNA 편집을 위한 디자인 파라미터의 특징분석을 보여준다. A) 야생형 Cas13b 및 촉매적으로 불활성인 H133A/H1058A Cas13b (dCas13b)에 대한 Gluc 표적화의 녹다운 효율. B) 타일링 Cluc 표적화 가이드로 시험한, 야생형 ADAR2 촉매 도메인 또는 과활성 E488Q 돌연변이체 ADAR2 촉매 촉매 도메인에 융합된 dCas13b 에 의한 루시퍼라제 활성 복원의 정량. C) *사이프리디니아* 루시퍼라제 W85X 표적화 30 nt 가이드를 위한 A->I 편집의 가이드 디자인 및 서열분석 정량. D) *PPIB* 표적화 50 nt 가이드를 위한 A->I 편집의 가이드 디자인 및 서열분석 정량. E) REPAIRv1에 의한 루시퍼라제 활성 복원에 대한 링커 선택의 영향. F) REPAIRv1에 의한 루시퍼라제 활성 복원에 대한 표적화된 아데노신에 반대되는 염기 일치성의 영향.

도 27은 G>A 돌연변이에 대한 clinVar 모티프 분포를 보여준다. 모든 G>A 돌연변이에 대해 ClinVar 데이터베이스에서 관찰된 각각의 가능한 삼중항 모티프의 수.

도 28A-28B는 dCas13b의 절단에 의한 RNA 결합을 보여준다. dCas13b의 다양한 N-말단 및 C-말단 절단이 도시되어 있다. RNA 결합은 표적화 및 비-표적화 가이드를 사용하여 활성을 비교하는, 루시퍼라제 신호의 복원에 의해 측정된 바와 같이 ADAR-의존적 RNA 편집이 있을 때 개시된다. 아미노산 위치는 Prevotella sp. P5-125 Cas13b 단백질의 아미노산 위치에 상응한다.

도 29는 다른 프로그램 가능한 ADAR 시스템과 dCas13-ADAR2 편집기의 비교를 보여준다. A) 두 가지 프로그램 가능한 ADAR 체계의 개략도: BoxB-기반 표적화 및 전체 길이 ADAR2 표적화. BoxB 체계 (상단)에서, ADAR2 데아미나제 도메인 (ADAR2$_{DD}$(E488Q))는 람다 N (N)으로 불리는 소형 박테리아 바이러스 단백질에 융합되는데, 이는 BoxB-로 불리는 소형 RNA 서열에 특이적으로 결합한다. 2 개의 BoxB- 헤어핀을 포함하는 가이드 RNA는 부위 특이적 편집을 위해 ADAR2 $_{DD}$(E488Q), -N를 가이드할 수 있다. 전체 길이 ADAR2 개략도 (하단)에서, ADAR2의 dsRNA 결합 도메인은 가이드 RNA 내 헤어핀에 결합하여, 프로그램 가능한 ADAR2 편집을 가능하게 한다. B) Cluc 를 표적하는 가이드 및 비-표적화 가이드와 함께, BoxB-ADAR2 $_{DD}$(E488Q)에 의한 상당한 RNA 편집의 전사체-전반 부위. 표적-적중 Cluc 부위 (254 A>G)가 주황색으로 강조 표시된다. C) Cluc 를 표적하는 가이드 및 비-표적화 가이드와 함께, ADAR2에 의한 상당한 RNA 편집의 전사체-전반 부위. 표적-적중 Cluc 부위 (254 A>G)가 주황색으로 강조 표시된다. D) Cluc 를 표적하는 가이드 및 비-표적화 가이드와 함께, REPAIRv1에 의한 상당한 RNA 편집의 전사체-전반 부위. 표적-적중 Cluc 부위 (254 A>G)가 주황색으로 강조 표시된다. E) Cluc에 대항하는 가이드 표적화를 위한 BoxB-ADAR2 $_{DD}$(E488Q), ADAR2, 및 REPAIRv1에 대한 표적-적중 편집률 백분율의 정량. F) 프로그램 가능한 ADAR 시스템을 위한 상이한 표적화와 비-표적화 조건 사이의 표적-외 부위의 중복.

도 30은 dCas13b-ADAR2 돌연변이체의 효율성 및 특이성을 보여준다. A) Cluc-표적화 및 비-표적화 가이드를 위한 dCas13b-ADAR2 $_{DD}$(E488Q) 돌연변이체 의한 루시퍼라제 활성 복원의 정량. B) 전사체-전반 서열분석에 의해 정량화된 표적화 및 비-표적화 가이드의 비율과 RNA-편집 표적-외의 수의 관련성. C) dCas13b-ADAR2 $_{DD}$(E488Q) 돌연변이체에 대한 전사체-전반 표적-외 RNA 편집 부위의 수 대 표적-적중 Cluc 편집 효율의 정량.

도 31은 dCas13b-ADAR2 $_{DD}$(E488Q) 돌연변이체에 의한 RNA 편집의 전사체-전반 특이성을 보여준다. A) Cluc를 표적하는 가이드와 함께, dCas13b-ADAR2 $_{DD}$(E488Q) 돌연변이체에 의한 상당한 RNA 편집의 전사체-전반 부위. 표적-적중 Cluc 부위 (254 A>G)가 주황색으로 강조 표시된다. B) 비-표적화 가이드와 함께, dCas13b-ADAR2 $_{DD}$(E488Q) 돌연변이체에 의한 상당한 RNA 편집의 전사체-전반 부위.

도 32는 dCas13b-ADAR2 $_{DD}$(E488Q) 편집의 표적-외에서 모티프 바이어스의 특징분석을 보여준다. A) 각 dCas13b-ADAR2 $_{DD}$(E488Q) 돌연변이에 대해, 전사체에서 모든 A>G 표적-외 편집물에 존재하는 모티프가 제시되어 있다. B) 모티프 일치성 당 표적-외 A>G 편집의 분포는 표적화 및 비-표적화 가이드와 함께 REPAIRv1에 대해 제시된다. C) 모티프 일치성 당 표적-외 A>G 편집의 분포는 표적화 및 비-표적화 가이드와 함께 REPAIRv2에 대해 제시된다.

도 33은 REPAIRv1 및 REPAIRv2 표적-외의 추가 특징분석을 보여준다. A) REPAIRv1에 대한 전사물 당 표적-외의 수의 히스토그램. B) REPAIRv2에 대한 전사물 당 표적-외의 수의 히스토그램. C) REPAIRv1 표적-외의 변형 효과 예측. D) REPAIRv1 표적-외의 잠재적 발암 효과의 분포. E) REPAIRv2 표적-외의 변형 효과 예측. F) REPAIRv2 표적-외의 잠재적 발암 효과의 분포.

도 34는 REPAIRv1 및 REPAIRv2의 RNA 편집 효율성 및 특이성을 보여준다. A) REPAIRv1 및 REPAIRv2에 대한 표적화된 아데노신 및 이웃 부위에 *KRAS*-표적화 가이드 1과 함께 *KRAS*의 백분율 편집의 정량. B) REPAIRv1 및 REPAIRv2에 대한 표적화된 아데노신 및 이웃 부위에 *KRAS*-표적화 가이드 3과 함께 *KRAS*의 백분율 편집의 정량. C) REPAIRv1 및 REPAIRv2에 대한 표적화된 아데노신 및 이웃 부위에 *PPIB*-표적화 가이드 2와 함께 *PPIB*의 백분율 편집의 정량.

도 35는 A>G RNA 편집기를 사용한 모든 잠재적 코돈 변화의 입증을 보여준다. A) A>I 편집에 의해 가능한 모든 잠재적 코돈 전환 표. B) A>I 편집에 의해 가능한 모든 잠재적 코돈 전환을 나타내는 코돈 표. C)

도 36A-36B는 RNA 간섭에 대한 Csx 단백질의 효과를 보여준다. A) Cas13b 및 Cas13b + Csx27 간섭의 비교; B) Cas13b 및 Cas13b + Csx28 간섭의 비교;

도 37A-37F는 루시퍼라제 어세이를 통한 Cas13b 및 Cas13b + Csx28에 의한 전사물 녹다운 비교를 나타낸다. A) Pin Cas13b (WP_036860899); B) Pbu Cas13b (WP_004343973); C) Rin Cas13b (WP_004919755); D) Pau Cas13b (WP025000926); E) Pgu Cas13b (WP_039434803); F) Pig Cas13b (WP_053444417)

본원에서 도면은 오직 예시의 목적을 위한 것이며, 반드시 척도에 따라 도시된 것은 아니다.

발명을 실시하기 위한 구체적인 내용

[0174] 일반적으로, CRISPR-Cas 또는 CRISPR 시스템은 Cas 유전자를 인코딩하는 서열, tracr (트랜스-활성화 CRISPR) 서열 (예를 들어, tracrRNA 또는 활성 부분 tracrRNA), tracr-메이트 서열 (내인성 CRISPR 시스템의 경우 "직접 반복부" 및 tracrRNA-처리된 부분 직접 반복부를 포함), 가이드 서열 (또한 내인성 CRISPR 시스템의 경우 "스페이서"라고도 함), 또는 그 용어가 본 명세서에서 사용시 "RNA(들)" (예, 가이드 Cas, 예컨대 Cas9, 예를 들어 CRISPR RNA 및 트랜스활성화 (tracr) RNA 또는 단일 가이드 RNA (sgRNA) (키메라 RNA)에 대해 RNA(들)) 또는 다른 서열, 및 CRISPR 유전자좌 유래의 전사물을 포함하여, CRISPR-연합 ("Cas") 유전자의 활성의 유도 또는 발현에 관여하는 전사물 및 다른 엘리먼트를 종합적으로 의미한다. 일반적으로, CRISPR 시스템은 표적 서열의 부위에서 CRISPR 복합체의 형성을 촉진하는 엘리먼트(내생의 CRISPR 시스템의 맥락에서 프로토스페이서(protospacer)로서도 지칭됨)에 의해 특징지어진다. CRISPR 단백질이 클래스 2 타입 VI-B 이펙터인 경우, tracrRNA가 요구되지 않는다. 본 발명의 조작된 시스템에서, 직접 반복부는 천연 발생 서열 또는 비-천연 발생 서열을 포함할 수 있다. 본 발명의 직접 반복부는 제한없이 천연 발생 길이 및 서열을 포함한다. 직접 반복부는 36 nt 길이일 수 있지만, 더 길거나 또는 더 짧게 직접 반복부는 다양할 수 있다. 예를 들어, 직접 반복부는 30 nt 이상, 예컨대 30 내지 100 nt 이상일 수 있다. 예를 들어, 직접 반복부는 길이가 30 nt, 40nt, 50nt, 60nt, 70nt, 70nt, 80nt, 90nt, 100nt, 또는 그 이상일 수 있다. 일부 실시형태에서, 본 발명의 직접 반복부는 천연 발생 직접 반복부의 5' 및 3' 말단 사이에 삽입된 합성 뉴클레오티드 서열을 포함할 수 있다. 일정 실시형태에서, 삽입된 서열은 자기 상보성일 수 있고, 예를 들어 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 또는 100% 자기 상보성일 수 있다. 더 나아가서, 본 발명의 직접 반복부는 뉴클레오티드 예컨대 어댑터 단백질에 결합하는 서열 또는 앱타머의 삽입을 포함할 수 있다 (기능성 도메인과의 연합을 위함). 일정 실시형태에서, 이러한 삽입부를 함유하는 직접 반복부의 한쪽 말단은 대체로 단형 DR의 처음 절반이고 말단은 대체로 단형 CR의 두번째 절반이다.

[0175] CRISPR 복합체의 형성의 맥락에서, "표적 서열"은 가이드 서열이 상보성을 갖도록 디자인된 서열을 지칭하며, 여기서, 표적 서열과 가이드 서열 간의 혼성화는 CRISPR 복합체의 형성을 증진시킨다. 표적 서열은 임의의 폴리뉴클레오티드, 예를 들어, DNA 또는 RNA 폴리뉴클레오티드를 포함할 수 있다. 일부 구현예에서, 표적 서열은 세포의 핵 또는 세포질 내에 위치된다. 일부 실시형태에서, 직접 반복은 다음 기준 중 어느 하나 또는 모두를 충족시키는 반복적인 모티프를 검색함으로써 인 실리코 (in silico) 에서 식별될 수 있다: 1. 타입 II CRISPR 유전자좌의 측면에 있는 게놈 서열의 2Kb 윈도우에서 발견됨; 2. 20 내지 50 bp 의 간격; 및 3. 20 내지 50 bp 간격을 둠. 일부 실시형태에서, 이들 기준 중 두 가지가 사용될 수 있으며, 예를 들어 1과 2, 2와 3, 또는 1과 3이 사용될 수 있다. 일부 실시형태에서, 모든 3개 기준이 사용될 수 있다.

[0176] 본 발명의 실시형태에서 용어 가이드 서열 및 가이드 RNA, 즉, Cas13b 이펙터 단백질을 표적 유전자좌로 가이드할 수 있는 RNA가 앞서 인용된 문헌 예컨대 WO 2014/093622 (PCT/US2013/074667)에서 처럼 상호교환적으로 사용된다. 일반적으로, 가이드 서열 (또는 스페이서 서열)은 표적 서열과 혼성화하여 CRSPR 복합체의 서열-특이적 결합을 표적 서열로 유도하도록 표적 폴리뉴클레오티드 서열과 충분한 상보성을 갖는 임의의 폴리뉴클레오티드 서열이다. 일부 구현예에서, 가이드 서열과 그의 상응하는 표적 서열 간의 상보성의 정도는 적절한 정렬 알고리즘을 사용하여 최적으로 정렬되는 경우, 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% 이상이다. 최적의 정렬은 서열을 정렬하기에 적절한 임의의 알고리즘의 사용으로 결정될 수 있으며, 그의 비-제한적인 예는 스미스-워터만(Smith-Waterman) 알고리즘, 니들만-분쉬(Needleman-Wunsch) 알고리즘, 버로우즈-휠러 트랜스폼(Burrows-Wheeler Transform)에 기초한 알고리즘(예를 들어, 버로우즈 휠러 얼라이너(Burrows Wheeler Aligner)), ClustalW, Clustal X, BLAT, 노보얼라인(Novoalign)(노보크라프트 테크놀로지즈(Novocraft Technologies)(www.novocraft.com에서 이용가능함), ELAND(일루미나(Illumina), 미국 캘리포니아주 샌디에고), SOAP(soap.genomics.org.cn에서 이용가능) 및 Maq(maq.sourceforge.net에서 이용가능)를 포함한다. 일부 실시형태에서, 가이드 서열 (또는 스페이서 서열)은 약 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,

22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, 또는 그 이상의 뉴클레오티드 길이이거나, 또는 이를 초과한다. 일부 구현예에서, 가이드 서열은 길이가 약 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 개 이하의 뉴클레오티드이다. 바람직하게 가이드 서열은 10-40 뉴클레오티드 길이, 예컨대 20-30 또는 20-40 뉴클레오티드 길이 또는 그 이상, 예컨대 30 뉴클레오티드 길이 또는 약 30 뉴클레오티드 길이이다. 일정 실시형태에서, 가이드 서열은 Cas13b 이펙터에 대해 10-30 뉴클레오티드 길이, 예컨대 20-30 또는 20-40 뉴클레오티드 길이 또는 그 이상, 예컨대 30 뉴클레오티드 길이 또는 약 30 뉴클레오티드 길이이다. 일정 실시형태에서, 가이드 서열은 10-30 뉴클레오티드 길이, 예컨대 또는 20-30 뉴클레오티드 길이, 예컨대 30 뉴클레오티드 길이이다. 표적 서열로의 CRISPR 복합체의 서열-특이적 결합을 유도하는 가이드 서열의 능력은 임의의 적절한 어세이에 의해 평가될 수 있다. 예를 들어, 시험되는 가이드 서열을 포함하는 CRISPR 복합체를 형성하기에 충분한 CRISPR 시스템의 성분은 예를 들어, CRISPR 서열의 성분을 인코딩하는 벡터로의 트랜스펙션 후에, 예를 들어, 본원에 기술된 바와 같은 서베이어 어세이에 의한 표적 서열 내의 우선적인 절단의 평가에 의해서, 상응하는 표적 서열을 갖는 숙주 세포로 제공될 수 있다. 유사하게, 표적 폴리뉴클레오티드 서열의 절단은 표적 서열, 시험되는 가이드 서열 및 시험 가이드 서열과 상이한 대조군 가이드 서열을 포함하는 CRISPR 복합체의 성분을 제공하고, 표적 서열에서 시험 및 대조군 가이드 서열 반응 간의 결합 또는 절단 비율을 비교함으로써 시험관에서 평가될 수 있다. 다른 어세이가 가능하며, 해당 분야의 숙련자에게 일어날 것이다.

[0177] 본 발명은 특정 Cas13b 이펙터, 핵산, 시스템, 벡터, 및 사용 방법을 제공한다. 모든 VI-B 는 구조에 의해, 그리고 또한 HEPN 도메인(들)의 위치에 의해 VI-A (Cas13b)와 구별가능하다. VI-B2 유전자좌가 작은 보조 단백질이 훨씬 더 보존된 것으로 보이는 것을 제외하고, 작은 보조 단백질이 결여된 Cas13b를 그것을 가지고 있는 Cas13b로부터 분리할 수 있는 것은 거의 없다.

[0178] 본 명세서에서 사용시, 용어 Cas13b-s1 부속 단백질, Cas13b-s1 단백질, Cas13b-s1, Csx27, 및 Csx27 단백질은 상호교환적으로 사용되며, 용어 Cas13b-s2 부속 단백질, Cas13b-s2 단백질, Cas13b-S2, Csx28, 및 Csx28 단백질은 상호교환적으로 사용된다.

[0179] 고전적인 CRISPR-Cas 시스템에서, 가이드 서열과 이에 상응하는 표적 서열 사이의 상보성 정도는 약 또는 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, 또는 100% 이상일 수 있고; 가이드 또는 RNA 또는 crRNA는 약 또는 약 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 또는 이상의 뉴클레오티드 길이일 수 있고; 또는 가이드 또는 RNA 또는 crRNA는 약 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 개 이하의 뉴클레오티드 길이 미만일 수 있고; 유리하게는 tracr RNA는 30 또는 50 개의 뉴클레오티드 길이이다. 그러나, 본 발명의 측면은 표적외 상호작용을 감소시키는 것이며, 예를 들어 낮은 상보성을 갖는 표적 서열과 가이드 상호작용을 감소시키는 것이다. 실제로, 예에서, 본 발명은 80% 초과 내지 약 95% 상보성, 예를 들어 83% 내지 84% 또는 88 내지 89% 또는 94 내지 95% 상보성을 갖는 표적외 서열과 표적 서열을 구별 (예를 들어, 18개 뉴클레오티드를 갖는 표적을 1, 2 또는 3개 미스매치를 갖는 18개 뉴클레오티드의 표적외와 구별)할 수 있는 CRISPR-Cas 시스템을 생성시키는 돌연변이를 포함하는 것으로 확인되었다. 따라서, 본 발명의 상황에서, 가이드 서열과 이의 상응하는 표적 서열 사이의 상보성 정도는 94.5% 또는 95% 또는 95.5% 또는 96% 또는 96.5% 또는 97% 또는 97.5% 또는 98% 또는 98.5% 또는 99% 또는 99.5% 또는 99.9%, 또는 100% 초과이다. 표적외는 서열과 가이드 사이의 상보성이 100% 또는 99.9% 또는 99.5% 또는 99% 또는 99% 또는 98.5% 또는 98% 또는 97.5% 또는 97% 또는 96.5% 또는 96% 또는 95.5% 또는 95% 또는 94.5% 또는 94% 또는 93% 또는 92% 또는 91% 또는 90% 또는 89% 또는 88% 또는 87% 또는 86% 또는 85% 또는 84% 또는 83% 또는 82% 또는 81% 또는 80% 미만이고, 서열과 가이드 사이의 상보성이 표적외는 100% 또는 99.9% 또는 99.5% 또는 99% 또는 99% 또는 98.5% 또는 98% 또는 97.5% 또는 97% 또는 96.5% 또는 96% 또는 95.5% 또는 95% 또는 94.5%인 것이 유리하다.

[0180] 본 발명에 따른 특히 바람직한 실시형태에서, (Cas를 표적 유전자좌로 가이드할 수 있는) 가이드 RNA는 (1) 진핵생물 세포에서 표적 유전자좌 (폴리뉴클레오티드 표적 유전자좌, 예컨대 RNA 표적 유전자좌)와 혼성화할 수 있는 가이드 서열; (2) 단일 RNA, 즉, sgRNA (5’ 내지 3’ 배향으로 정렬됨) 또는 crRNA 중에 잔류하는 직접 반복부 (DR) 서열) 을 포함할 수 있다.

[0181] 특정 실시형태에서, 야생형 Cas13b 이펙터 단백질은 RNA 결합 및 절단 기능을 갖는다.

[0182] 특정 실시형태에서, Cas13b 이펙터 단백질은 DNA 절단 기능을 가질 수 있다. 이들 실시형태에서, 방법은 본 명세서에서 기술하는 바와 같은 벡터를 세포에 도입시키는 것을 포함하는 본 명세서에 기술된 바와 같은 진핵생물 세포 (시험관내, 즉 단리된 진핵생물 세포)에 하나 이상의 돌연변이를 유도시키는 것으로 이해되는 본 명세

서에 제공된 이펙터 단백질을 기반으로 제공될 수 있다. 돌연변이(들)는 가이드(들) RNA(들) 또는 sgRNA(들) 또는 crRNA(들)을 통해 세포(들)의 각 표적 서열에서 하나 이상의 뉴클레오티드의 도입, 결실, 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들) 또는 crRNA(들)를 통해 상기 세포(들)의 각 표적 서열에 1 내지 75개 뉴클레오티드의 도입, 결실, 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들) 또는 crRNA(들)를 통해 상기 세포(들)의 각 표적 서열에 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 또는 75개 뉴클레오티드의 도입, 결실, 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들) 또는 crRNA(들)를 통해 상기 세포(들)의 각 표적 서열에 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 또는 75개 뉴클레오티드의 도입, 결실, 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들) 또는 crRNA(들)를 통해 상기 세포(들)의 각 표적 서열에 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 또는 75개 뉴클레오티드의 도입, 결실, 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들) 또는 crRNA(들)를 통해 상기 세포(들)의 각 표적 서열에 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 또는 75개 뉴클레오티드의 도입, 결실, 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들) 또는 crRNA(들)를 통해서 상기 세포(들)의 각 표적 서열에서 40, 45, 50, 75, 100, 200, 300, 400 또는 500개 뉴클레오티드의 도입, 결실, 또는 치환을 포함할 수 있다.

[0183] 독성 및 표적외 효과를 최소화하기 위해서, 전달된 Cas13b mRNA 및 가이드 RNA의 농도를 제어하는 것이 중요하다. Cas13b mRNA 및 가이드 RNA의 최적 농도는 세포 또는 인간이외의 진핵생물 동물 모델에서 상이한 농도를 시험하고 잠재적인 표적외 게놈 유전자좌에서 변형 정도를 분석하기 위해 심층 서열분석을 사용하여 결정할 수 있다. 대안적으로, 독성 수준 및 표적을 벗어난 효과를 최소화하기 위해, Cas13b 닉카제 (nickase) mRNA(예를 들어 D10A 돌연변이를 지니는 스트렙토코커스 피오게네스 Cas9)는 관심 대상의 부위를 표적화하는 가이드 RNA의 쌍에 의해 전달될 수 있다. 독성 및 표적외 효과를 최소화하기 위한 가이드 서열 및 전략 표적은 WO 2014/093622 (PCT/US2013/074667)에 기술된 바와 같이, 또는 본 명세서에서와 같은 돌연변이를 통해서일 수 있다.

[0184] 전형적으로, 내인성 CRISPR 시스템의 상황에서, CRISPR 복합체 (표적 서열과 혼성화하고 하나 이상의 Cas 단백질과 복합체 형성되는 가이드 서열을 포함)의 형성은 표적 서열에서 또는 그 가까이 (예를 들어, 표적 서열로부터 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50개, 또는 그 이상의 염기쌍이내)에서 한쪽 또는 양쪽 가닥(적용가능한 경우)의 절단을 야기시킨다.

[0185] Cas13b를 인코딩하는 핵산 분자는 유리하게 코돈 최적화된다. 코돈 최적화된 서열의 예는 이러한 예에서, 진핵생물, 예, 인간 (즉, 인간에서 발현을 위해 최적화됨), 또는 다른 진핵생물의 경우, 본 명세서에서 기술된 바와 같이, 동물 또는 포유동물에서 발현을 위해 최적화된 서열이고, 예를 들어, WO 2014/093622 (PCT/US2013/074667)의 SaCas9 인간 코돈 최적화된 서열을 참조한다. 이것이 바람직하지만, 기타 다른 예들도 가능하며, 인간 이외의 숙주 종에 대한 코돈 최적화, 또는 특정 기관에 대한 코돈 최적화가 공지되어 있다는 것이 이해될 것이다. 일부 실시형태에서, Cas를 인코딩하는 효소 코딩 서열은 특정 세포, 예컨대 진핵생물 세포에서 발현을 위해 코돈 최적화된다. 진핵생물 세포는 제한없이 인간, 또는 인간이외의 진핵생물 또는 본 명세서에 기술된 바와 같은 동물 또는 포유동물, 예를 들어 마우스, 래트, 토끼, 개, 가축 또는 인간이외의 포유동물 또는 영장류를 포함하는 특정 유기체의 것일 수 있거나 또는 그로부터 유래될 수 있다. 일부 구현예에서, 인간이나 동물에 임의의 실질적인 의학적 이익 없이 그들을 고통스럽게 할 수 있는 인간의 생식계열 유전자 아이덴티티를 변형하기 위한 과정 및/또는 동물의 유전자 아이덴티티를 변형하기 위한 과정과, 또한 이러한 과정으로부터 생긴 동물이 배제될 수 있다. 일반적으로, 코돈 최적화는 고유 서열의 적어도 하나의 코돈(예를 들어, 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 개 이상의 코돈)을 숙주 세포의 유전자에 더욱 빈번하게 또는 가장 빈번하게 사용되는 코돈으로 대체하면서, 고유 아미노산 서열을 유지함으로써 대상 숙주 세포에서의 발현의 증진을 위해 핵산 서열을 변형시키는 과정을 지칭한다. 다양한 종은 특정 아미노산의 특정 코돈에 대하여 특정 편향을 나타낸다. 코돈 편향(유기체 간의 코돈 사용의 차이)은 종종 메신저 RNA(mRNA)의 번역의 효율과 상관관계가 있으며, 이는 결국, 특히, 번역되는 코돈의 특성 및 특정 운반 RNA(tRNA) 분자의 이용 가능성에 좌우되는 것으로 여겨진다. 세포에서의 선택된 tRNA의 우세는 일반적으로 펩티드 합성에 가장 빈번하게 사용되는 코돈을 반영하는 것이다. 따라서, 유전자는 코돈 최적화를 기반으로 하여 주어진 유기체에서의 최적의 유전자 발현을 위해 맞춤화될 수 있다. 코돈 사용 표는, 예를 들어 www.kazusa.orjp/codon/에서 이용 가능한 "코돈 사용 데이터베이스"에서 용이하게 이용 가능하며, 이 표는 다수의 방식으로 적합하게 될 수 있다. 문헌[Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence

databases:status for the year 2000" Nucl. Acids Res. 28:292 (2000)]을 참조한다. 특정 숙주 세포에서의 발현을 위해 특정 서열을 코돈 최적화시키는 컴퓨터 알고리즘이 또한 이용 가능하며, 예를 들어 Gene Forge (Aptagen; Jacobus, PA) 이 또한 이용가능하다. 일부 실시형태에서, Cas를 인코딩하는 서열에서 하나 이상의 코돈 (예를 들어, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, 또는 그 이상, 또는 모든 코돈)은 특정 아미노산에 대해 가장 빈번하게 사용되는 코돈에 해당된다.

[0186] 일정 실시형태에서, 본 명세서에 기술된 바와 같은 방법은 하나 이상의 가이드 RNA를 인코딩하는 하나 이상의 핵산이 하나 이상의 관심 유전자의 프로모터를 포함하는 조절 요소와 세포에서 작동가능하게 연결되게 제공되거나 또는 도입된 Cas13b 유전자이식 세포의 제공을 포함할 수 있다. 본 명세서에서 사용시, 용어 "Cas13b 유전자이식 세포"는 Cas13b 유전자가 게놈에 통합된 세포, 예컨대 진핵생물 세포를 의미한다. 세포의 성질, 유형 또는 기원은 본 발명에 따라 특별히 제한되지 않는다. 또한 Cas13b 이식유전자를 세포에 도입시키는 방법은 다양할 수 있고 당분야에 공지된 바와 같은 임의 방법일 수 있다. 일정 실시형태에서, Cas13b 유전자이식 세포는 단리된 세포에 Cas13b 이식유전자를 도입하여 수득된다. 일정 다른 실시형태에서, Cas13b 유전자이식 세포는 Cas13b 유전자이식 유기체로부터 세포를 단리하여 수득된다. 예로서, 제한없이, 본 명세서에서 언급시 Cas13b 유전자이식 세포는 Cas13b 유전자이식 진핵생물, 예컨대 Cas13b 녹-인(knock-in) 진핵생물로부터 유래될 수 있다. 참조로 본 명세서에 편입된, WO 2014/093622 (PCT/US13/74667)를 참조한다. Rosa 유전자좌를 표적화하도록 지시된 Sangamo BioSciences, Inc.가 출원한 미국 특허 공개 제20120017290호 및 20110265198호의 방법은 본 발명의 CRISPR Cas 시스템을 이용하도록 변형될 수 있다. Rosa 유전자좌를 표적화하는 것에 관한 셀렉티스(Cellectis)에 양도된 미국 공개 특허 출원 제20130236946의 방법은 또한 본 발명의 CRISPR Cas 시스템을 이용하여 변형될 수 있다. Cas9 녹-인 마우스를 설명하는 Platt et. al. (Cell; 159(2):440-455 (2014)) 이 추가의 예로 참조되며, 이는 본원에 참고로 포함된다. Cas13b 이식유전자는 Lox-Stop-polyA-Lox(LSL) 카세트를 더 포함할 수 있어서 Cre 리콤비나제에 의해 Cas13b 발현을 유도시킬 수 있다. 대안적으로, Cas13b 유전자이식 세포는 단리된 세포에서 Cas13b 이식유전자를 도입하여 수득될 수 있다. 이식유전자를 위한 전달 시스템은 당분야에 잘 공지되어 있다. 예로서, Cas13b 이식유전자는 역시 본 명세서의 다른 곳에 기술된 바와 같이, 벡터 (예, AAV, 아데노바이러스, 렌티바이러스) 및/또는 입자 및/또는 입자 전달을 통해 예를 들어 진핵생물 세포로 전달될 수 있다.

[0187] 세포, 예컨대 본 명세서에서 언급시 Cas13b 유전자이식 세포는 예를 들어 제한없이 Platt 등 (2014), Chen 등 (2014) 또는 Kumar 등 (2009)이 기술한 바와 같이, 표적 유전자좌에 Cas13b를 가이드할 수 있는 RNA와 복합체형성시 Cas13b의 서열 특이적 작용으로 발생되는 돌연변이 또는 통합된 Cas13b 유전자를 갖는 것 이외의 게놈 변경, 예컨대 예를 들어 하나 이상의 종양발생성 돌연변이를 더 포함할 수 있다는 것을 당업자는 이해할 것이다.

[0188] 일부 실시형태에서, Cas13b 서열은 하나 이상의 핵 국재화 서열 (NLS), 예컨대 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 또는 그 이상의 NLS 또는 이를 초과하는 NLS에 융합된다. 일부 실시형태에서, Cas13b는 아미노 말단에 또는 그 가까이에 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 또는 그 이상의 NLS 또는 이를 초과하는 NLS, 또는 카복시 말단에 또는 그 가까이에 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 또는 그 이상의 NLS 또는 이를 초과하는 NLS, 또는 이들의 조합 (예를 들어, 아미노 말단에 0 또는 적어도 하나 이상의 NLS 및 카복시 말단에 0 또는 하나 이상의 NLS)을 포함한다. 1 개 초과의 NLS가 존재하는 경우, 각각은 단일의 NLS가 1 개 초과의 복제물로 존재하고/존재하거나 1 개 이상의 복제물로 존재하는 1 개 이상의 기타 다른 NLS와 조합될 수 있도록, 다른 것들로부터 독립적으로 선택될 수 있다. 본 발명의 바람직한 실시형태에서, Cas13b는 6개 이하의 NLS를 포함한다.

일부 구현예에서, NLS는 NLS의 가장 가까운 아미노산이 N- 또는 C-말단으로부터 폴리펩티드 사슬을 따라 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50 개 이상의 아미노산 내에 존재하는 경우, N- 또는 C-말단 근처에 있는 것으로 간주된다. NLS의 비-제한적 예는 하기로부터 유래된 NLS 서열을 포함한다: 아미노산 서열 PKKKRKV(SEQ ID NO: X)를 갖는, SV40 바이러스 대형 T-항원의 NLS; 뉴클레오플라스민으로부터의 NLS (예를 들어, 서열 KRPAATKKAGQAKKKK (SEQ ID NO: X)를 갖는 뉴클레오플라스민 양분(bipartite) NLS); 아미노산 서열 PAAKRVKLD (SEQ ID NO: X) 또는 RQRRNELKRSP(SEQ ID NO: X)를 갖는 c-myc NLS; 서열 NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(SEQ ID NO: X)를 갖는 hRNPA1 M9 NLS; 임포틴-알파로부터의 IBB 도메인의 서열 RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: X); 근종 T 단백질의 서열 VSRKRPRP (SEQ ID NO: X) 및 PPKKARED (SEQ ID NO: X); 인간 p53의 서열 POPKKKPL (SEQ ID NO: X); 마우스 c-abl IV의 서열 SALIKKKKKMAP (SEQ ID NO: X); 인플루엔자 바이러스 NS1의 서열 DRLRR (SEQ ID NO: X) 및 PKQKKRK (SEQ ID NO: X); 간염 바이러스 델타 항원의 서열 RKLKKKIKKL (SEQ ID NO: X); 마우스 Mx1 단백질의 서열 REKKKFLKRR (SEQ ID NO: X); 인간 폴리(ADP-리보스) 폴리머라제의 서열 KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: X); 및 스테로이드

호르몬 수용체 (인간) 글루코코르티코이드의 서열 RKCLQAGMNLEARKTKK (SEQ ID NO: X). 일반적으로, 하나 이상의 NLS는 진핵생물 세포의 핵에서 검출가능한 양으로 Cas의 축적을 구동시키는데 충분한 강도이다. 일반적으로, 핵 국재화 활성의 강도는 Cas 내 NLS의 수, 사용된 특정 NLS(들), 또는 이들 인자의 조합으로부터 유래될 수 있다. 핵에서의 축적의 검출은 임의의 적합한 기법에 의해 실행될 수 있다. 예를 들어, 검출가능한 마커가 Cas에 융합될 수 있어서, 세포 내 위치를 예컨대 핵의 위치를 검출하기 위한 수단 (예를 들어, 핵에 특이적인 착색제 예컨대 DAPI)과 조합하여 가시화시킬 수 있다. 세포 핵을 또한 세포로부터 분리할 수 있으며, 그 다음 이의 내용물을 단백질을 검출하기 위한 임의의 적합한 과정, 예컨대 면역조직화학, 웨스턴 블롯, 또는 효소 활성 분석에 의해 분석할 수 있다. 핵 내에 축적은 또한 예컨대 Cas 또는 복합체에 노출되지 않았거나, 또는 하나 이상의 NLS가 결여된 Cas에 노출된 대조군과 비교하여, CRISPR 복합체 형성의 효과에 대한 어세이 (예를 들어, 표적 서열에서 DNA 절단 또는 돌연변이에 대한 어세이, 또는 CRISPR 복합체 형성 및/또는 Cas 효소 활성에 의해 영향받는 변경된 유전자 발현 활성에 대한 어세이)를 통해 간접적으로 결정할 수 있다.

[0189] 일정 측면에서, 본 발명은 예를 들어, Cas13b 및/또는 Cas13b를 표적 유전자좌로 가이드할 수 있는 RNA (즉, 가이드 RNA)를 세포에 전달하거나 또는 도입시킬뿐만 아니라, 이들 성분을 전파 (예를 들어, 원핵생물 세포에서) 시킬 수 있는 벡터를 포함한다. 본원에 사용되는 바와 같이, "벡터"는 엔티티를 하나의 환경에서 또 다른 환경으로 전달하는 것을 허용 또는 촉진하는 도구이다. 이는 레플리콘, 예컨대 플라스미드, 파지, 또는 코스미드로, 또 다른 DNA 절편이 그 안에 삽입될 수 있어서, 삽입된 절편의 복제를 유발한다. 일반적으로, 벡터는 적절한 제어 엘리먼트와 관련되었을 때 복제가 가능하다. 일반적으로, 용어 "벡터"는 그것이 연결되어 있는 다른 핵산을 전달할 수 있는 핵산 분자를 지칭한다. 벡터는 제한 없이, 단일 가닥, 이중 가닥 또는 부분 이중 가닥인 핵산 분자; 하나 이상의 자유 말단을 포함하는, 자유 말단을 포함하지 않는 (예 : 환형) 핵산 분자; DNA, RNA 또는 둘 다를 포함하는 핵산 분자; 및 당업계에 공지된 다른 종류의 폴리뉴클레오티드를 포함한다. 하나의 유형의 벡터는 "플라스미드"이며, 이는 추가의 DNA 분절이 예를 들어, 표준 분자 클로닝 기술에 의해 삽입될 수 있는 환형 이중 가닥 DNA 루프를 지칭한다. 벡터의 다른 종류는 바이러스 벡터로서, 바이러스-유래된 DNA 또는 RNA 서열이 바이러스에 봉입되는 벡터에 존재한다(예를 들어, 레트로바이러스, 복제 결함 레트로바이러스, 아데노바이러스, 복제 결함 아데노바이러스, 및 아데노-연관 바이러스(AAV)). 바이러스 벡터는 또한 숙주 세포 내로의 트랜스펙션을 위한, 바이러스가 보유하는 폴리뉴클레오티드를 포함한다. 특정 벡터는 그것이 도입된 숙주 세포에서 자율적 복제가 가능하다(예를 들어, 박테리아 복제 기원을 가진 박테리아 벡터 및 에피솜 포유동물 벡터). 다른 벡터(예를 들어, 비-에피솜 포유동물 벡터)는 숙주 세포 내로 도입시 숙주 세포의 게놈에 통합되며, 이에 의해 숙주 게놈과 함께 복제된다. 더욱이, 특정 벡터는 그것이 작동가능하게 연결된 유전자의 발현을 지시할 수 있다. 이러한 벡터는 본원에서 "발현 벡터"로 지칭된다. 재조합 DNA 기술에 유용한 통상적인 발현 벡터는 종종 플라스미드의 형태로 존재한다.

[0190] 재조합 발현 벡터는 숙주 세포에서 핵산의 발현에 적합한 형태로 본 발명의 핵산을 포함할 수 있는데, 이는 재조합 발현 벡터가 하나 이상의 조절 요소를 포함하는 것을 의미하며, 하나 이상의 조절 요소는 발현에 사용될 숙주 세포에 기반하여 선택될 수 있고, 발현될 핵산 서열에 작동가능하게 연결된다. 재조합 발현 벡터 내에서, "작동가능하게 연결된"은 대상 뉴클레오티드 서열이 (예를 들어, 시험관내 전사/번역 시스템 내에서, 또는 벡터가 숙주 세포 내로 도입되는 경우 숙주 세포 내에서) 뉴클레오티드 서열의 발현을 가능하게 하는 방식으로 조절 요소(들)에 연결된 것을 의미하는 의도이다. 재조합 및 클로닝 방법에 관련하여, US 2004-0171156 A1 호로서 2004년 9월 2일 간행된 미국 특허 출원 10/815,730호가 언급될 수 있으며, 이의 모든 내용은 그 전체로서 본원에 참조로 포함된다.

[0191] 벡터(들)는 조절 요소(들), 예, 프로모터(들)를 포함할 수 있다. 벡터(들)는 Cas13b 엔코딩 서열, 및/또는 단독을 포함할 수 있지만, 또한 가능하게 적어도 2, 3 또는 8 또는 16 또는 32 또는 48 또는 50 가이드 RNA(들) (예, crRNAs) 엔코딩 서열, 예컨대 1-2, 1-3, 1-4 1-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-8, 3-16, 3-30, 3-32, 3-48, 3-50 RNA(들) (예, crRNAs)를 포함할 수 있다. 각 RNA에 대한 프로모터가 있는 (예, crRNA(들)) 단일 벡터에서, 유리하게는 약 16 초과의 RNA(들)이 있는 경우(예, crRNA(들)), 단일 벡터가 16 초과의 RNA(들) (예, crRNA(들))을 제공하는 경우, 하나 이상의 프로모터(들)는 하나를 초과하는 RNA(들) (예, crRNA(들))의 발현을 구동시킬 수 있으며, 예를 들어 32 RNA(들) (예, sgRNA 또는 crRNA(들))가 존재할 때, 각각의 프로모터는 2 RNA(들) (예, sgRNA 또는 crRNA(들))의 발현을 구동시킬 수 있고, 48 RNA(들) (예, sgRNA 또는 crRNA(들))가 존재할 때, 각 프로모터는 3 RNA(들) (예, sgRNA 또는 crRNA(들))의 발현을 구동시킬 수 있다. 단순한 산술 및 잘 확립된 클로닝 프로토콜 및 본 개시내용의 교시를 통해서 당업자는 적합한 예시적 벡터 예컨대 AAV를 위한 RNA(들), 예를 들어, sgRNA(들) 또는 crRNA(들), 및 적합한 프로모터 예컨대 U6 프로모터, 예를 들어, U6-sgRNA 또는 -crRNA(들)에 관해 본 발명을 쉽게 실시할 수 있다. 예를 들어, AAV의 포장 한계는 ~4.7 kb이

다. 당업자는 단일 벡터에 쉽게 약 12-16, 예를 들어, 13 U6-sgRNA 또는 crRNA(들) 카세트를 핏팅시킬 수 있다. 이는 임의의 적합한 수단, 예컨대 TALE 조립에 사용되는 골든 게이트 전략을 통해 조립될 수 있다 (http://www.genome-engineering.org/taleffectors/). 당업자는 또한 탠덤 가이드 전략을 사용하여 U6-sgRNA 또는 -crRNA(들)의 개수를 대략 1.5배 증가시킬 수 있고, 예를 들어 12 내지 16, 예를 들어, 13 내지 대략 18 내지 24, 예를 들어, 약 19개 U6-sgRNA 또는 -crRNA(들)까지 증가시킬 수 있다. 따라서, 당업자는 단일 벡터, 예를 들어 AAV 벡터에 대략 18 내지 24, 예를 들어, 약 19개 프로모터-RNA, 예를 들어, U6-sgRNA 또는 -crRNA(들)에 쉽게 도달할 수 있다. 벡터 내에 프로모터 및 RNA, 예를 들어 sgRNA(들) 또는 crRNA(들)의 수를 증가시키기 위한 추가 수단은 절단성 서열에 의해 분리된 RNA, 예를 들어 sgRNA 또는 crRNA(들)의 어레이를 발현시키기 위해 단일 프로모터 (예, U6)를 사용하는 것이다. 그리고 벡터에 프로모터-RNA, 예, sgRNA 또는 crRNA(들)의 수를 증가시키기 위한 추가의 수단은 코딩 서열 또는 유전자의 인트론에 절단성 서열에 의해 분리된 프로모터-RNAs, 예, sgRNA 또는 crRNA(들)의 어레이를 발현시키는 것이고, 이러한 예에서, 조직 특이적 방식으로 긴 RNA의 전사를 할 수 있고 발현을 증가시킬 수 있는, 중합효소 II 프로모터를 사용하는 것이 유리하다. (참조, 예를 들어, nar.oxfordjournals.org/content/34/7/e53.short, www.nature.com/mt/journal/v16/n9/abs/mt2008144a.html). 유리한 실시형태에서, AAV는 최대 약 50개 유전자를 표적화하는 U6 탠덤 sgRNA를 포장할 수 있다. 따라서, 당분야의 지식 및 본 개시의 교시로부터 당업자는 임의의 과도한 실험없이, 본 명세서에 기술된 RNA 또는 가이드 또는 sgRNA 또는 crRNA(들)의 수와 특히 관련하여, 하나 이상의 프로모터에 작동적으로 또는 기능적으로 연결되거나 또는 그 제어 하에서 다수의 RNA 또는 가이드 또는 sgRNA 또는 crRNA(들)를 발현하는, 벡터(들), 예를 들어 단일 벡터를 쉽게 만들고 이용할 수 있다.

[0192] 가이드 RNA(들), 예, sgRNA(들) 또는 crRNA(들) 엔코딩 서열 및/또는 Cas13b 엔코딩 서열은 조절 요소(들)에 기능적으로 또는 작동적으로 연결될 수 있어서 조절 요소(들)가 발현을 구동한다. 프로모터(들)는 항상성 프로모터(들) 및/또는 조건적 프로모터(들) 및/또는 유도성 프로모터(들) 및/또는 조직 특이적 프로모터(들)일 수 있다. 프로모터는 RNA 중합효소, pol I, pol II, pol III, T7, U6, H1, 레트로바이러스 루이스 육종 바이러스 (RSV) LTR 프로모터, 사이토메갈로바이러스 (CMV) 프로모터, SV40 프로모터, 디히드로폴레이트 리덕타제 프로모터, β-액틴 프로모터, 포스포글리세롤 키나제 (PGK) 프로모터, 및 EF1α 프로모터로 이루어진 군으로부터 선택될 수 있다. 유리한 프로모터는 U6인 프로모터이다.

[0193] 본 발명의 일 측면에서, 본 출원의 RNA- 또는 RNA-표적화 CRISPR 시스템이라고도 하는 신규 RNA 표적화 시스템은 표적 특이적 RNA 서열에 대한 맞춤형 단백질의 생성을 요구하지 않고 대신 단일 효소가 특이적 RNA 표적을 인식하기 위해 RNA 분자에 의해 프로그램될 수 있는, 달리 말해서 효소가 상기 RNA 분자를 사용해 특이적 RNA 표적으로 동원될 수 있는 본원에서 확인된 Cas13b 단백질을 기반으로 한다.

[0194] 일부 구현예에서, 핵산-표적화 시스템의 하나 이상의 엘리먼트는 내인성 CRISPR RNA-표적화 시스템을 포함하는 특정 유기체로부터 유래한다. 특정 실시형태에서, CRISPR RNA-표적화 시스템은 *유박테리움* 및 *루미노코커스*에서 발견된다. 특정 실시형태에서, 이펙터 단백질은 표적화 및 부수적 ssRNA 절단 활성을 포함한다. 특정 실시형태에서, 이펙터 단백질은 이중 HEPN 도메인을 포함한다. 특정 실시형태에서, 이펙터 단백질은 Cas13a의 Helical-1 도메인에 대한 대응부가 결여되어 있다. 특정 실시형태에서, 이펙터 단백질은 928 aa의 중간 크기를 갖는 이전에 특징화된 클래스 2 CRISPR 이펙터보다 작다. 이 중간 크기는 Cas13c보다 190aa (17%) 미만이고, Cas13b보다 200aa (18%) 미만이고, Cas13a보다 300aa (26%) 미만이고. 특정 실시형태에서, 이펙터 단백질은 측면 서열 (예를 들어, PFS, PAM)을 필요로 하지 않는다.

[0195] 특정 실시형태에서, 이펙터 단백질 유전자좌 구조는 WYL 도메인 함유 보조 단백질을 포함한다 (이러한 도메인의 원래 확인된 그룹에 보존된 3 개의 아미노산 이후에 표시됨; 예를 들어, WYL 도메인 IPR026881 참조). 특정 실시형태에서, WYL 도메인 보조 단백질은 하나 이상의 헬릭스-턴-헬릭스 (HTH) 또는 리본-헬릭스-헬릭스 (RHH) DNA-결합 도메인을 포함한다. 특정 실시형태에서, WYL 도메인 함유 보조 단백질은 RNA-표적화 이펙터 단백질의 표적화 및 부수적 ssRNA 절단 활성 둘 모두를 증가시킨다. 특정 실시형태에서, WYL 도메인 함유 보조 단백질은 N-말단 RHH 도메인 뿐만 아니라, 원래의 WYL 모티프에 상응하는 불변 티로신-류신 이중선을 포함하여, 주로 소수성 보존된 잔기의 패턴을 포함한다. 특정 실시형태에서, WYL 도메인 함유 보조 단백질은 WYL1이다. WYL1은 주로 *루미노코커스*와 관련된 단일 WYL-도메인 단백질이다.

[0196] 다른 예시적인 실시형태에서, 타입 VI RNA-표적화 Cas 효소는 Cas 13d이다. 특정 실시형태에서, Cas13d는 *유박테리움 시라에움*(*Eubacterium siraeum*) DSM 15702 (EsCas13d) 또는 *루미노코커스* (*Ruminococcus*) sp. N15.MGS-57 (RspCas13d) 이다 (참조, 예, Yan et al., Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein, Molecular Cell (2018),

doi.org/10.1016/j.molcel.2018.02.028). RspCas13d 및 EsCas13d에는 측면 서열 요구사항이 없다 (예, PFS, PAM).

[0197] 본 명세서에 기재된 핵산-표적화 시스템, 벡터 시스템, 벡터 및 조성물은 다양한 핵산-표적화 응용, 유전자 산물, 예컨대 단백질의 합성 변경 또는 변형, 핵산 절단, 핵산 편집, 핵산 스플라이싱; 표적 핵산의 트래피킹, 표적 핵산의 추적, 표적 핵산의 단리, 표적 핵산의 시각화 등에 사용될 수 있다.

[0198] 유리한 실시형태에서, 본 발명은 표 1을 참조하여 Cas13b 이펙터 단백질을 포함한다. 표 1 Cas13b 이펙터 단백질은 표 1과 관련하여 본원에서 보다 상세하게 논의된다.

[0199] **Cas13b 뉴클레아제**

[0200] 본 발명의 Cas13b 이펙터 단백질은 표 1에 제시된 바와 같은 또는 이와 같은 단백질이거나, 이를 포함하거나, 또는 그로 본질적으로 이루어지거나, 또는 이들로 이루어지거나 또는 이를 수반하거나 또는 이와 관련되거나 또는 이에 언급된다. 본 발명은 본원에 제시된 바와 같은 돌연변이 또는 변형을 포함하여, 표 1에 제시된 단백질을 제공하거나, 이에 관한 것이거나, 이를 수반하거나 포함하거나, 이로 본질적으로 이루어지거나, 이루어진다. 표 1 Cas13b 이펙터 단백질은 표 1과 관련하여 본원에서 보다 상세하게 논의된다.

[0201] 따라서, 일부 실시형태에서, 이펙터 단백질은 예를 들어 임의로 본원에 기술된 바와 같이 전사 활성인자 또는 억제인자 도메인, NLS 또는 다른 기능성 도메인으로 기능할 수 있는, RNA-결합 단백질, 예컨대 데드-Cas 유형 이펙터 단백질일 수 있다. 일부 실시형태에서, 이펙터 단백질은 RNA의 단일 가닥을 절단하는 RNA-결합 단백질일 수 있다. 결합된 RNA가 ssRNA이라면, ssRNA은 완전하게 절단된다. 일부 실시형태에서, 이펙터 단백질은 예를 들어, 2개의 RNase 도메인을 포함한다면, RNA의 이중 가닥을 절단하는 RNA-결합 단백질일 수 있다. 결합된 RNA가 dsRNA이라면, dsRNA는 완전하게 절단된다. 일부 실시형태에서, 이펙터 단백질은 닉카제 활성을 갖는 RNA-결합 단백질일 수 있는데, 즉 dsRNA에 결합하지만, RNA 가닥의 하나만을 절단한다.

[0202] CRISPR 시스템에서의 RNase 기능은 공지되어 있으며, 예를 들어 특정 타입 III CRISPR-Cas 시스템에 대한 mRNA 표적화가 보고되어 있고 (Hale et al., 2014, Genes Dev, vol. 28, 2432-2443; Hale et al., 2009, Cell, vol. 139, 945-956; Peng et al., 2015, Nucleic acids research, vol. 43, 406-417) 및 상당한 이점을 제공한다. 본 발명의 이펙터 단백질을 통한 CRISPR-Cas 시스템, 조성물 또는 RNA의 표적화 방법이 따라서 제공된다.

[0203] 표적 RNA, 즉 관심 RNA는 표적 RNA 상의 관심 표적 부위로 동원되어 이펙터 단백질의 결합을 야기시키는 본 발명에 의해 표적화되는 RNA이다. 표적 RNA는 RNA의 임의의 적합한 형태일 수 있다. 이는 일부 실시형태에서, mRNA를 포함한다. 다른 실시형태에서, 표적 RNA는 tRNA 또는 rRNA를 포함할 수 있다.

[0204] **간섭 RNA (RNAi) 및 마이크로RNA (miRNA)**

[0205] 다른 실시형태에서, 표적 RNA는 간섭 RNA, 즉 RNA 간섭 경로에 관여하는 RNA, 예컨대 shRNA, siRNA 등을 포함할 수 있다. 다른 실시형태에서, 표적 RNA는 마이크로RNA (miRNA)를 포함할 수 있다. 간섭 RNA 또는 miRNA에 대한 제어는 생체내 또는 시험관내 간섭 RNA 또는 miRNA의 수명을 감소시켜 그들 접근법에서 확인되는 표적외 효과 (OTE)를 줄이는데 도움을 줄 수 있다.

[0206] 이펙터 단백질 및 적합한 가이드가 (예를 들어 적합한 프로모터, 예를 들어 조직 특이적 또는 세포 주기 특이적 프로모터 및/또는 인핸서의 제어 하에서 공간적으로 또는 일시적으로) 발현되면, 이를 사용하여 세포 또는 그들 세포 내 시스템을 (생체내 또는 시험관내에서) RNAi로부터 '보호'할 수 있다. 이것은 RNAi가 이펙터 단백질 및 적합한 가이드가 발현되지 않는 세포 또는 조직의 비교를 위해서 또는 RNAi가 요구되지 않는 이웃한 조직 또는 세포에서 유용할 수 있다 (즉, 개별적으로 RNAi가 제어되지 않는 곳 및 제어되는 곳). 이펙터 단백질은 RNA, 예컨대 리보자임, 리보솜, 또는 리보스위치를 포함하거나 또는 그로 이루어진 분자를 제어하거나 또는 그에 결합시키는데 사용될 수 있다. 본 발명의 실시형태에서, RNA 가이드가 이들 분자로 이펙터 단백질을 동원시켜서 이펙터 단백질이 그들에 결합할 수 있게 한다.

[0207] **리보솜 RNA (rRNA)**

[0208] 예를 들어, 아잘리드 항생제 예컨대 아지트로마이신이 잘 알려져 있다. 그들은 50S 리보솜 서브유닛을 표적으로 하여 파괴한다. 본 발명의 이펙터 단백질은 50S 리보솜 서브유닛을 표적으로 하는 적합한 가이드 RNA와 함께, 일부 실시형태에서, 50S 리보솜 서브유닛에 동원되어 결합될 수 있다. 따라서, 본 발명의 이

펙터 단백질은 리보솜 (특히 50s 리보솜 서브유닛) 표적에 유도되는 적합한 가이드와 함께 제공된다. 리보솜 (특히 50s 리보솜 서브유닛) 표적에 유도되는 적합한 가이드와 함께 이펙터 단백질의 사용은 항생제 사용을 포함할 수 있다. 특히, 항생제 사용은 아잘리드 항생제, 예컨대 아지트로마이신의 작용과 유사하다. 일부 실시형태에서, 원핵생물 리보솜 서브유닛, 예컨대 원핵생물 내 70S 서브유닛, 상기 언급된 50S 서브유닛, 30S 서브유닛을 비롯하여 16S 및 5S 서브유닛이 표적화될 수 있다. 다른 실시형태에서, 진핵생물 리보솜 서브유닛, 예컨대 진핵생물에서 80S 서브유닛, 60S 서브유닛, 40S 서브유닛을 비롯하여, 28S, 18S. 5.8S 및 5S 서브유닛이 표적이 될 수 있다.

[0209] 일부 실시형태에서, 이펙터 단백질은 본 명세서에 기술된 바와 같이, 임의로 작용화된, RNA-결합 단백질일 수 있다. 일부 실시형태에서, 이펙터 단백질은 RNA의 단일 가닥을 절단하는 RNA-결합 단백질일 수 있다. 양쪽 경우에서, 특히 RNA-결합 단백질이 RNA의 단일 가닥을 절단하는 경우에, 리보솜 기능은 조정될 수 있고 특히 감소되거나 또는 파괴될 수 있다. 이는 리보솜 RNA 및 임의의 리보솜 서브유닛에 적용될 수 있고, rRNA의 서열은 충분히 공지되어 있다.

[0210] 리보솜 활성의 대조군은 따라서 리보솜 표적에 대한 적합한 가이드와 함께 본 발명의 이펙터 단백질의 사용을 통해 예상한다. 이는 리보솜의 절단, 또는 리보솜과의 결합을 통해서일 수 있다. 특히, 리보솜 활성의 감소가 예상된다. 이는 생체내 또는 시험관내에서 리보솜 기능을 어세이하는데 유용할 뿐만 아니라 생체내 또는 시험관내에서 리보솜 활성을 기반으로 요법을 제어하는 수단으로서 유용할 수 있다. 더 나아가서, 생체내 또는 시험관내 시스템에서 단백질 합성의 제어 (즉, 감소)가 예상되며, 이러한 제어는 항생제 및 연구 및 진단 용도를 포함한다.

[0211] **리보스위치**

[0212] 리보스위치 (압토자임이라고도 알려짐)는 소형 분자에 결합하는 메신저 RNA 분자의 조절 절편이다. 이는 전형적으로 mRNA에 의해 인코딩되는 단백질 생산의 변화를 일으킨다. 따라서, 리보스위치 활성의 제어는 리보스위치 표적에 적합한 가이드와 함께 본 발명의 이펙터 단백질의 사용을 통해 예상된다. 이는 리보스위치의 절단, 또는 리보스위치와의 결합을 통해서일 수 있다. 특히, 리보스위치 활성의 감소가 예상된다. 생체내 또는 시험관내 리보스위치 기능을 어세이하는데 유용할뿐만 아니라, 생체내 또는 시험관내에서 리보스위치 활성을 기반으로 하는 요법을 제어하는 수단으로서 유용할 수 있다. 더 나아가서, 생체내 또는 시험관내 시스템에서 단백질 합성의 제어 (즉, 감소)가 예상된다. rRNA의 경우처럼, 이러한 제어는 항생제 및 연구 및 진단 용도를 포함할 수 있다.

[0213] **리보자임**

[0214] 리보자임은 효소 (물론 단백질임)와 유사한, 촉매 특성을 갖는 RNA 분자이다. 천연 발생 및 조작된 리보자임은 RNA를 포함하거나 또는 그로 이루어지므로, 그들은 또한 본 발명의 RNA-결합 이펙터 단백질에 의해 표적화될 수 있다. 일부 실시형태에서, 이펙터 단백질은 리보자임을 절단하여 이를 무능하게 만드는 RNA-결합 단백질일 수 있다. 따라서 리보자임 활성의 제어는 리보자임 표적에 적합한 가이드와 본 발명의 이펙터 단백질의 사용을 통해 예상된다. 이는 리보자임의 절단 또는 리보자이모가의 결합을 통해서일 수 있다. 특히, 리보자임 활성의 감소가 예상된다. 이는 생체내 또는 시험관내 리보자임 기능을 어세이하는데 유용할 뿐만 아니라, 생체내 또는 시험관내에서 리보자임 활성을 기반으로 요법을 제어하는 수단으로서 유용할 수 있다.

[0215] **RNA 프로세싱을 포함하는, 유전자 발현**

[0216] 이펙터 단백질이 또한, RNA 프로세싱의 제어를 통한것을 포함하여, 표적 유전자 발현에 적합한 가이드와 함께 사용될 수 있다. RNA 프로세싱의 제어는 RNApol의 표적화를 통한 대안적 스플라이싱을 포함하는 RNA 스플라이싱; 식물에서의 비리오이드를 포함하는 바이러스 복제 (특히 위성 바이러스, 박테리오파지 및 레트로바이러스, 예를 들어 HBV, HBC 및 HIV 및 본원에 열거된 다른 것); 및 tRNA 생합성과 같은 RNA 프로세싱 반응을 포함할 수 있다. 이펙터 단백질 및 적합한 가이드는 또한 RNA활성화 (RNAa)를 제어하는데 사용될 수 있다. RNAa는 유전자 발현의 촉진을 야기시키므로, 유전자 발현의 제어는 RNAa의 파괴 또는 감소를 통해서 그에 따라 유전자 발현의 촉진을 덜하는 방식으로 달성될 수 있다.

[0217] **RNAi 스크린**

[0218] 그의 녹다운이 표현형 변화와 연관된 유전자 산물의 동정은, RNAi 스크린을 통해서 생물학적 경로에 대한 정보를 얻고 성분 부분을 동정한다. 제어는 또한 스크린에서 RNAi의 활성을 제거시키거나 또는 감소시켜서 (간

섭/억제를 제거 또는 감소시켜) (이전에 간섭된) 유전자 산물의 활성을 복귀시키도록 이펙터 단백질 및 적합한 가이드를 사용하여 이들 스크린 시에 또는 그 동안 수행될 수 있다.

[0219] 위성 RNA (satRNAs) 및 위성 바이러스가 또한 처리될 수 있다.

[0220] RNase 활성화 관련된 본 명세서에서의 제어는 일반적으로 감소, 음성적 파괴 또는 녹다운 또는 녹아웃을 의미한다.

[0221] ***생체 내 RNA 적용***

[0222] **유전자 발현의 억제**

[0223] 본 명세서에 제공되는 표적-특이적 RNAse는 표적 RNA의 매우 특이적 절단을 가능하게 한다. RNA 수준에서 간섭은 게놈이 변형되지 않으므로, 비침윤적 방식으로 공간적이고 일시적으로 조정을 가능하게 한다.

[0224] 다수의 질환이 mRNA 표적화를 통해 처리가능한 것으로 입증되었다. 이들 연구의 대부분이 siRNA의 투여와 관련되지만, 본 명세서에 제공된 RNA 표적화 이펙터 단백질이 유사한 방식으로 적용될 수 있다는 것이 분명하다.

[0225] mRNA 표적의 예 (및 상응하는 질환 치료)는 VEGF, VEGF-R1 및 RTP801 (AMD 및/또는 DME의 치료 시), 캐스파제 2 (비동맥 전방국소빈혈성 시신경장애(Naion)의 치료 시), ADRB2 (안내압의 치료 시), TRPVI (안구 건조 증후군의 치료시), Syk 키나제 (천식의 치료 시), Apo B (과콜레스테롤혈증의 치료 시), PLK1, KSP 및 VEGF (고형 종양의 치료 시), Ber-Abl (CML의 치료 시)(Burnett and Rossi Chem Biol. 2012, 19(1): 60-71)) 이다. 유사하게, RNA 표적화는 RNA-바이러스 매개 질환 예컨대 HIV (HIV Tet 및 Rev의 표적화), RSV (RSV 뉴클레오캡시드의 표적화) 및 HCV (miR-122의 표적화) (Burnett and Rossi Chem Biol. 2012, 19(1):60-71)의 치료에서 효과적인 것으로 입증되었다.

[0226] 본 발명의 RNA 표적화 이펙터 단백질가 돌연변이 특이적 또는 대립유전자 특이적 녹다운에 사용될 수 있다는 것을 더욱 예상한다. 가이드 RNA는 돌연변이 또는 대립유전자 특이적 서열을 포함하는 전사된 mRNA 내 서열을 특이적으로 표적화하도록 디자인될 수 있다. 이러한 특이적 녹다운은 특히 돌연변이되거나 또는 대립유전자 특이적 유전자 산물과 연관된 질환과 관련된 치료적 용도에 적합하다. 예를 들어, 가족성 저베타지단백질혈증 (FHBL)의 대부분의 사례는 ApoB 유전자의 돌연변이로 인해 초래된다. 이 유전자는 아포지단백질 B 단백질의 2가지 버전; 단형 (ApoB-48) 및 장형 (ApoB-100)을 인코딩한다. FHBL을 초래하는 몇몇 ApoB 유전자 돌연변이는 ApoB의 두가지 버전을 비정상적으로 짧게 만들어버린다. 본 발명의 RNA 표적화 이펙터 단백질에 의한 돌연변이된 ApoB mRNA 전 사물의 특이적 표적화 및 녹다운은 FHBL의 치료에서 유리할 것이다. 다른 예로서, 헌팅톤 질환 (HD)은 비정상적인 단백질을 야기시키는 헌팅틴 단백질을 인코딩하는 유전자 내 CAG 트리플렛 반복부의 확장에 의해 초래된다. 본 발명의 RNA 표적화 이펙터 단백질을 사용하여 헌팅틴 단백질을 인코딩하는 돌연변이되거나 또는 대립유전자-특이적 mRNA 전사물의 특이적 표적화 및 녹다운은 HD의 치료에서 유리할 것이다.

[0227] 이러한 내용에서, 본 명세서에 기술된 바와 같은 다양한 용도에 대해 보다 일반적으로, RNA 표적화 이펙터 단백질의 분할 형태의 사용을 예상할 수 있다는 것을 주목한다. 실제로, 이는 특이성을 증가시킬 수 있을 뿐만 아니라 전달에 유익할 수 있다. Cas13b는 Cas13b 효소의 2부분이 기능성 Cas13b를 실질적으로 포함한다는 의미에서 분할된다. 이상적으로 분할은 항상 촉매 도메인(들)이 영향받지 않도록 해야만 한다. 그 Cas13b는 뉴클레아제로서 기능할 수 있거나 또는 이의 촉매 도메인에 전형적인 돌연변이(들)로 인해, 촉매 활성이 매우 적거나 또는 전혀 없는 본질적으로 RNa-결합 단백질인 데드-Cas13b일 수 있다.

[0228] 분할 Cas13b의 각각의 절반은 이량체와 파트너와 융합될 수 있다. 예로서, 제한없이, 라파마이신 감수성 이량체화 도메인의 적용은 Cas13b 활성의 일시적 제어를 위해 화학적으로 유도가능한 분할 Cas13b를 생성시킬 수 있게 한다. 따라서, Cas13b는 2개 단편으로 분할되어 화학적으로 유도가능하게 될 수 있고 라파마이신-감수성 이량체화 도메인은 Cas13b의 제어된 재조립을 위해 사용될 수 있다. 분할 Cas13b의 2개 부분은 분할 Cas13b의 N 말단 부분 및 C 말단 부분으로서 여겨질 수 있다. 융합은 통상적으로 Cas13b 스플릿 지점에서이다. 달리 말해서, 분할 Cas13b의 N 말단 부분의 C 말단이 이량체 절반부의 하나에 융합되는 한편, C 말단 부분의 N 말단이 다른 이량체 절반에 융합된다.

[0229] Cas13b는 파괴를 새롭게 생성시키지 않는다는 의미에서 분할될 필요는 없다. 스플릿 지점은 통상적으로 가상 환경에서 디자인되고, 구축물 내로 클로닝된다. 함께, 분할 Cas13b의 두 부분, N 말단 부분 및 C 말단

부분은 바람직하게 야생형 아미노산 (또는 이를 인코딩하는 뉴클레오티드)의 적어도 70% 이상, 바람직하게 야생형 아미노산 (또는 이를 인코딩하는 뉴클레오티드)의 적어도 80% 이상, 바람직하게 적어도 90% 이상, 바람직하게 적어도 95% 이상, 가장 바람직하게 적어도 99% 이상을 포함하는 전체 Cas13b를 형성한다. 일부 트리밍(trimming)이 가능할 수 있으며, 돌연변이체가 고려된다. 비-기능성 도메인은 완전히 제거될 수 있다. 중요한 것은 2개 부분이 함께할 수 있고 바람직한 Cas13b 기능이 복구되거나 또는 재구성된다는 것이다. 이량체는 동형이량체 또는 이형이량체일 수 있다.

[0230] 일정 실시형태에서, 본 명세서에서 기술된 바와 같은 Cas13b 이펙터는 돌연변이-특이적, 또는 대립유전자-특이적 표적화, 예컨대 돌연변이-특이적, 또는 대립유전자-특이적 녹다운에 사용될 수 있다.

[0231] 게다가, RNA 표적화 이펙터 단백질은 RNAse 활성을 증가시키거나 또는 메시지의 추가적인 분해를 보장하도록 상승적으로 작용하는, 다른 기능성 RNAse 도메인, 예컨대 비특이적 RNase 또는 아르고노트(Argonaute) 2에 융합될 수 있다.

[0232] **RNA 기능의 조정을 통한 유전자 발현의 조정**

[0233] mRNA의 절단을 통해 유전자 발현에 대한 직접 효과 이외에도, RNA 표적화는 또한 유전자 발현의 보다 미세한 조정을 가능하게 할 수 있는, 세포 내 RNA 프로세싱의 특정 측면에 영향을 미치는데 사용될 수 있다. 일반적으로, 조정은 예를 들어, RNA에 대한 단백질 결합을 방해, 예컨대, 예를 들어 단백질의 결합 차단, 또는 RNA 결합 단백질의 동원에 의해 매개될 수 있다. 실제로, 조정은 상이한 수준 예컨대 mRNA의 스플라이싱, 수송, 국재화, 번역 및 턴오버 시에 보장될 수 있다. 요법과 관련하여 유사하게, RNA-특이적 표적화 분자를 사용하여 이들 각각의 수준에서 (병원성) 이상기능을 해결하는 것을 예상할 수 있다. 이들 실시형태에서 RNA 표적화 단백질이 RNA 표적을 절단하는 능력은 상실하였지만, 예컨대 본 명세서에 기술된 Cas13b의 돌연변이된 형태에 결합하는 그 능력은 유지하는 "데드" Cas13b인 것이 많은 경우에서 바람직하다.

[0234] ***a) 선택적 스플라이싱***

[0235] 많은 인간 유전자는 선택적 스플라이싱의 결과로서 다수의 mRNA를 발현한다. 상이한 질환들이 발현된 유전자의 기능 상실 또는 기능 획득을 초래하는 이상 스플라이싱과 연관된 것으로 확인되었다. 이들 질환의 일부가 스플라이싱 결합을 야기하는 돌연변이에 의해 야기되지만, 이들 중 많은 것들은 그렇지 않다. 하나의 치료적 옵션은 스플라이싱 기전을 직접적으로 표적화하는 것이다. 본 명세서에 기술된 RNA 표적화 이펙터 단백질은 예를 들어 스플라이싱을 차단 또는 촉진, 엑손을 포함 또는 배제, 특이적 이소폼의 발현에 영향 및/또는 대체 단백질 생성물의 발현의 자극을 위해 사용될 수 있다. 이러한 적용분야는 이하에서 보다 상세하게 기술한다.

[0236] 표적 RNA에 RNA 표적화 이펙터 단백질의 결합은 RNA 서열에 대한 스플라이싱 인자의 접근은 입체적으로 차단할 수 있다. 스플라이싱 부위를 표적으로 하는 RNA 표적화 이펙터 단백질은 그 부위에서 스플라이싱을 차단할 수 있고, 임의로 인접한 부위로 스플라이싱을 재유도할 수 있다. 예를 들어, 5' 스플라이싱 부위 결합에 RNA 표적화 이펙터 단백질 결합은 스플라이시오솜의 U1 성분의 동원을 차단할 수 있고, 그 엑손을 건너뛰는것이 선호된다. 대안적으로, 스플라이싱 인핸서 또는 사일렌서를 표적으로 하는 RNA 표적화 이펙터 단백질은 표적 부위에서 트랜스작용 조절성 스플라이싱 인자의 결합을 방지하여서 효과적으로 스플라이싱을 차단하거나 또는 촉진할 수 있다. 엑손 배제는 본 명세서에 기술된 바와 같은 RNA 표적화 이펙터 단백질에 의해 엑손 근처의 전구체 mRNA로 ILF2/3의 동원에 의해 달성될 수 있다. 또 다른 예로서, 글리신 풍부 도메인이 hnRNP A1의 동원 및 엑손 배제를 위해 부착될 수 있다 (Del Gatto-Konczak et al. Mol Cell Biol. 1999 Jan;19(1):251-60).

[0237] 일정 실시형태에서, gRNA의 적절한 선택을 통해서, 특이적 스플라이싱 변이체가 표적화될 수 있는 한편, 다른 스플라이싱 변이체는 표적화되지 않을 것이다.

[0238] 일부 경우에서, RNA 표적화 이펙터 단백질을 사용하여 스플라이싱을 촉징할 수 있다 (예를 들어, 스플라이싱이 결함이 있는 경우). 예를 들어, RNA 표적화 이펙터 단백질은 추가의 스플라이싱을 위해서 스플라이싱 조절 스템-루프를 안정화시킬 수 있는 이펙터와 연합될 수 있다. RNA 표적화 이펙터 단백질은 표적 DNA로 단백질을 동원하기 위해서 특이적 스플라이싱 인자를 위한 공통 결합 부위 서열에 연결될 수 있다.

[0239] 이상 스플라이싱과 연관된 질환의 예는 제한없이, 시냅스에서 기능하는 단백질의 스플라이싱을 조절하는 Nova 단백질의 상실로 인한, 안구간대경련-근간대경련 증후군 (Paraneoplastic Opsoclonus Myoclonus Ataxia)(POMA), 및 낭포성 섬유증 경막 전도도 조절인자의 결함적 스플라이싱으로 인해, 비기능성 클로라이드

채널이 생성되어 야기되는 낭포성 섬유증을 포함한다. 다른 질환에서, 이상 RNA 스플라이싱은 기능 획득을 일으킨다. 이는 예를 들어 mRNa의 3' UTR에서 CUG 트리플렛-반복부 확대 (50 내지 > 1500 반복부)로 인해 스플라이싱 결함이 야기되어 초래된 근긴장성 이영양증이 그 경우이다.

[0240] RNA 표적화 이펙터 단백질은 바람직한 엑손 주변의 인트론의 절제를 촉진하기 위해 5' 스플라이싱 부위로 스플라이싱 인자 (예컨대 U1)를 동원하여 엑손을 포함시키는데 사용될 수 있다. 이러한 동원은 스플라이싱 활성 인자로서 기능하는, 아르기닌/세린 풍부 도메인과 융합을 통해 매개될 수 있다 (Gravely BR and Maniatis T, Mol Cell. 1998 (5): 765-71).

[0241] RNA 표적화 이펙터 단백질은 바람직한 유전자좌에서 스플라이싱 기구를 차단하여, 상이한 단백질 생성물의 엑손 인식 및 발현을 방지하도록 사용될 수 있다는 것을 예상한다. 치료할 수 있는 질환의 예는 디스트로핀 단백질을 인코딩하는 유전자의 돌연변이에 의해 초래된 뒤센 근이영양증(Duchenne muscular dystrophy)(DMD)이다. 거의 모든 DMD 돌연변이가 프레임시프트를 초래하여, 그 결과 손상된 디스트로핀 번역을 야기한다. RNA 표적화 이펙터 단백질은 스플라이싱 접합부 또는 엑손 스플라이싱 인핸서 (ESE)와 쌍형성할 수 있어서, 엑손 인식을 방지하여, 그 결과 부분적으로 기능성인 단백질의 번역이 일어난다. 이는 치명적인 뒤센 표현형을 덜 심각한 벡커 표현형으로 전환시킨다.

[0242] ***b) RNA 변형***

[0243] RNA 편집은 소정 서열의 유전자 산물의 다양성이 RNA 내 소수의 변형으로 인해 증가되는 자연적 과정이다. 전형적으로, 변형은 게놈에 의해 인코딩되는 것과 상이한 RNA 서열을 생성시키는, 아데노신 (A)의 이노신 (I)으로의 전환을 포함한다. RNA 변형은 인반적으로, ADAR 효소에 의해 보장되며, 그리하여 프리-RNA 표적은 인트론 논-코딩 엘리먼트 및 편집된 아데노신을 함유하는 엑손 사이의 염기쌍형성을 통해 불완전한 듀플렉스 RNA를 형성한다. A-I 편집의 고전적인 예는 글루타메이트 수용체GluR-B mRNA인데, 그로 인한 변화는 채널의 변형된 전도도 특성을 야기시킨다 (Higuchi M, et al. Cell. 1993;75:1361-70] 참조).

[0244] 인간에서, *ADAR1* 유전자의 이형접합 기능성-무효 돌연변이는 피부 질환, 인간 색소성 유전피부병을 초래한다 (Miyamura Y, et al. Am J Hum Genet. 2003;73:693-9] 참조). 본 발명의 RNA 표적화 이펙터 단백질은 이상 기능성 RNA 변형을 보정하는데 사용될 수 있다는 것을 예상한다.

[0245] RNA 아데노신 메틸라제 (N(6)-메틸아데노신)은 본 발명의 RNA 표적화 이펙터 단백질에 융합되어 관심 전사물을 표적으로 할 수 있다는 것을 더욱 예상한다. 이러한 메틸라제는 가역적 메틸화를 야기하며, 조절성 역할을 갖고, 다수의 RNA-관련 세포 경로를 조정하여 유전자 발현 및 세포 운명 결정에 영향을 줄 수 있다 (Fu et al Nat Rev Genet. 2014;15(5):293-306).

[0246] ***c) 폴리아데닐화***

[0247] mRNA의 폴리아데닐화는 mRNA의 핵 수송, 번역 효율 및 안정성에 중요하며, 이들 모두뿐만 아니라 폴리아데닐화 과정은 특이적 RBP에 의존된다. 대부분의 진핵생물 mRNA는 전사 후에 약 200개 뉴클레오티드의 3' 폴리(A) 테일을 받는다. 폴리아데닐화는 폴리(A) 중합효소의 활성을 자극하는 상이한 RNA-결합 단백질 복합체가 관여된다 (Minvielle-Sebastia L et al. Curr Opin Cell Biol. 1999;11:352-7] 참조). 본 명세서에서 제공되는 RNA-표적화 이펙터 단백질은 RNA-결합 단백질과 RNA 사이의 상호작용을 방해하거나 또는 촉진시키는데 사용될 수 있다는 것을 예상한다.

[0248] 폴리아데닐화에 관여하는 결함성 단백질과 연관된 질환의 예는 안구인두근 이영양증 (OPMD)이다 (Brais B, et al. Nat Genet. 1998;18:164-7] 참조).

[0249] ***d) RNA 이출***

[0250] 프리-mRNA 프로세싱 후, mRNA는 핵에서 세포질로 이출된다. 이는 캐리어 복합체의 생성을 포함하는 세포 기전에 의해 보장되며, 이후에 핵 공극을 통해 전위되고 세포질로 mRNA를 방출하고, 캐리어는 이후에 재생된다.

[0251] RNA의 이출에서 역할을 하는 단백질 (예컨대 TAP)의 과발현은 그렇지 않으면 개구리 (Xenopus)에서 비효율적으로 이출되는 전사물의 이출을 증가시키는 것으로 확인되었다 (Katahira J, et al. EMBO J. 1999;18:2593-609).

[0252] ***e) mRNA 국재화***

[0253] mRNA 국재화는 공간적으로 조절되는 단백질 생산을 보장한다. 세포의 특정 영역으로 전사물의 국재화는 국

재화 엘리먼트에 의해 보장될 수 있다. 특정 실시형태에서, 본 명세서에 기술된 이펙터 단백질은 관심 RNA로 국재화 엘리먼트를 표적화시키는데 사용될 수 있다는 것을 예상한다. 이펙터 단백질은 표적 전사물에 결합하여 그들을 이의 펩티드 신호전달 태그에 의해 결정된 세포 내 위치로 왕복이동시키도록 디자인될 수 있다.

보다 특히, 예를 들면, 핵 국재화 신호 (NLS)에 융합된 RNA 표적화 이펙터 단백질을 사용하여 RNA 국재화를 변경시킬 수 있다.

[0254] 국재화 신호의 추가적인 예는 몇몇 비대칭 세포 유형에서 세포질에 β-액틴의 국재화를 보장하는 집코드 결합 단백질 (ZBP1), KDEL 보유 서열 (세포질 망상 구조에 국재화), 핵 이출 신호 (세포질에 국재화) 미토콘드리아 표적화 신호 (미토콘드리아에 국재화), 퍼옥시솜 표적화 신호 (퍼옥시솜에 국재화) 및 m6A 마킹/YTHDF2 (p-바디에 국재화)를 포함한다. 예상되는 다른 접근법은 RNA 표적화 이펙터 단백질의 공지의 국재화 단백질 (예를 들어, 막, 시냅스)과의 융합이다.

[0255] 대안적으로, 본 발명에 따른 이펙터 단백질은 예를 들어 국재화-의존적 녹아웃에서 사용될 수 있다. 적절한 국재화 신호와 이펙터 단백질을 융합시켜서, 이펙터 세포가 특정 세포 구획에 표적화된다. 이 구획에 존재하는 표적 RNA만이 효과적으로 표적화되는데 반해, 그렇지 않은 동일한 표적이지만 다른 세포 구획에 존재하는 것은 표적화되지 않아서, 국재화 의존적 녹아웃이 확정될 수 있다.

[0256] ***f) 번역***

[0257] 본 명세서에 기술된 RNA 표적화 이펙터 단백질은 번역을 향상시키거나 또는 억제하는데 사용될 수 있다. 상향조절 번역은 세포 회로를 제어하는 매우 강력한 방식으로 예상된다. 게다가, 기능성 연구의 경우, 단백질 번역 스크린은 전사물의 상향조절이 증가된 단백질 생산으로 번역되지 않는다는 단점을 갖는 전사 상향조절 스크린에 비해 선호될 수 있다.

[0258] 본 명세서에 기술된 RNA 표적화 이펙터 단백질은 번역 개시 인자, 예컨대 EIF4G 를 번역을 구동시키기 위해 관심 메신저 RNA의 5' 미번역 반복부(5'UTR)에 인접하도록 하는데 사용될 수 있다는 것을 예상한다(비-재프로그램가능 RNA 결합 단백질에 대한 문헌 [De Gregorio et al. EMBO J. 1999;18(17):4865-74 에 기재된 바와 같음).

다른 예시적인 GLD2로서, 세포질 폴리(A) 중합효소는 RNA 표적화 이펙터 단백질에 의해 표적 mRNA로 동원될 수 있다. 이는 표적 mRNA의 지정된 폴리아데닐화를 가능하게 하여서 번역을 자극할 수 있다.

[0259] 유사하게, 본 명세서에서 예상되는 RNA 표적화 이펙터 단백질은 mRNA의 번역 억제인자, 예컨대 ZBP1을 차단하는데 사용될 수 있다 (Huttelmaier S, et al. Nature. 2005;438:512-5] 참조). 표적 RNA의 번역 개시 부위에 결합하여, 번역이 직접적으로 영향받을 수 있다.

[0260] 또한, 예를 들어 이의 분해를 억제하여, mRNA를 안정화시키는 단백질, 예컨대 RNase 억제제와 RNA 표적화 이펙터 단백질의 융합은, 관심 전사물로부터 단백질 생산을 증가시키는 것이 가능하다.

[0261] 본 명세서에 기술된 RNA 표적화 이펙터 단백질은 RNA 전사물의 5' UTR 영역에서 결합하여 리보솜이 형성되어 번역을 시작하는 것을 방지하여 번역을 억제하는데 사용될 수 있다는 것을 예상한다.

[0262] 게다가, RNA 표적화 이펙터 단백질은 CCR4-NOT 디아데닐라제 복합체의 성분인 Caf1을 표적 mRNA로 동원시켜서, 탈아데닐화 또는 표적 전사물 및 단백질 번역의 억제를 일으킬 수 있다.

[0263] 예를 들어, 본 발명의 RNA 표적화 이펙터 단백질은 치료적으로 관련된 단백질의 번역을 증가시키거나 또는 감소시키는데 사용될 수 있다. RNA 표적화 이펙터 단백질을 사용하여 번역을 하향조절 또는 상향조절시킬 수 있는 치료적 적용분야의 예는 근위축성 측상 경화증 (ALS) 및 심혈관 질환이다. 신경교 글루타메이트 수송체 EAAT2의 감소된 수준이 ALS 운동 피질 및 척수를 비롯하여, ALS 뇌 조직에서 다수의 비정상적 EAAT2 mRNA 전사물에서 보고되었다. EAAT2 단백질 및 기능의 상실은 ALS에서 흥분독성의 주요 원인이라고 여겨진다. EAAT2 단백질 수준 및 기능의 복원은 치료적 이득을 제공할 수 있다. 그런 이유로, RNA 표적화 이펙터 단백질은 예를 들어, 상기 기술된 바와 같이 번역 억제인자를 차단하거나 또는 mRNA를 안정화시켜서, EAAT2 단백질의 발현을 상향조절시키는데 유리하게 사용될 수 있다. 아포지단백질 A1은 고밀도 지단백질 (HDL)의 주요 단백질 성분이고 ApoA1 및 HDL은 일반적으로 동맥경화 예방적이라고 간주된다. RNA 표적화 이펙터 단백질은 상기 기술된 바와 같이 예를 들어, 번역 억제인자를 차단하거나 또는 mRNA를 안정화시켜, ApoA1의 발현을 상향조절시키는데 유리하게 사용될 수 있다는 것을 예상한다.

[0264] ***g) mRNA 턴오버***

[0265] 번역은 mRNA 턴오버 및 조절된 mRNA 안정성과 단단하게 결합되어 있다. 특이적 단백질이 전사물의 안정화

(예컨대, 뉴런의 ELAV/Hu 단백질, Keene JD, 1999, Proc Natl Acad Sci U S A. 96:5-7) 및 트리스테트라프롤린(TTP)에 관여하는 것으로 기재되어 있다. 이들 단백질은 세포질에서 분해로부터 메신저를 보호하여 표적 mRNA를 안정화시킨다 (Peng SS et al., 1988, EMBO J. 17:3461-70).

[0266] 본 발명의 RNA-표적화 이펙터 단백질을 사용하여 mRNA 전사물을 안정화시키도록 작용하는 단백질의 활성을 방해하거나 또는 촉진시킬 수 있어서, mRNA 턴오버가 영향받는다. 예를 들어, RNA 표적화 이펙터 단백질을 사용하여 표적 RNA로 인간 TTP의 동원은 아데닐레이트-우리딜레이트-풍부 엘리먼트 (AU-풍부 엘리먼트) 매개 번역 억제 및 표적 분해를 가능하게 한다. AU-풍부 엘리먼트는 원발암유전자, 핵 전사 인자, 및 사이토카인을 코딩하는 많은 mRNA의 3' UTR에서 확인되며, RNA 안정성을 촉진한다. 다른 예로서, RNA 표적화 이펙터 단백질은 다른 mRNA 안정화 단백질인 HuR에 융합될 수 있고(Hinman MN and Lou H, Cell Mol Life Sci 2008;65:3168-81), 이것을 표적 전사물에 동원시켜 이의 수명을 연장시키거나 또는 단수명 mRNA를 안정화시킨다.

[0267] 본 명세서에 기술된 RNA-표적화 이펙터 단백질은 표적 전사물의 분해를 촉진하는데 사용될 수 있다고 더욱 예상된다. 예를 들어, m6A 메틸트랜스퍼라제는 표적 전사물에 동원되어, P-바디로 전사물을 국재화시켜서 표적의 분해를 초래할 수 있다.

[0268] 또 다른 예로서, 본 명세서에 기술된 바와 같은 RNA 표적화 이펙터 단백질은 표적 전사물로 동원시켜서, 이의 분해를 가능하게 하도록, 비특이적 엔도뉴클레아제 도메인 PilT N-말단 (PIN)에 융합시킬 수 있다.

[0269] 부신생물성 신경계 장애 (PND)-연관 뇌척수염 및 신경병증을 갖는 환자는 이후에 혈관-뇌 장벽을 가로지르는 중추 신경계 밖 종양 내 Hu-단백질에 대한 자가 항체를 발생시키는 환자이다 (Szabo A et al. 1991, Cell.;67:325-33). 본 발명의 RNA-표적화 이펙터 단백질은 mRNA 전사물에 자가 항체의 결합을 방해하는데 사용될 수 있다는 것을 예상한다.

[0270] 근긴장성 이영양증-단백질 키나제 (DMPK) 유전자의 3' UTR에서 (CUG)n의 확장에 의해 야기된, 이영양증 타입 1(dystrophy type 1) (DM1)을 갖는 환자는 핵 내 이러한 전사물의 축적으로 특징된다. (CUG)n 반복부를 표적으로 하는 엔도뉴클레아제와 융합된 본 발명의 RNA 표적화 이펙터 단백질은 이상 전사물의 이러한 축적을 억제할 수 있다고 예상된다.

[0271] ***h) 다중-기능성 단백질과 상호작용***

[0272] 일부 RNA-결합 단백질은 다양한 과정으로 기능하도록 수많은 RNA 상의 다수 부위에 결합된다. 예를 들어, hnRNP A1 단백질은 스플라이싱 인자를 길항하는, 엑손 스플라이싱 사일렌서 서열에 결합하고, 텔로미어 말단과 연합되며 (그리하여 텔로미어 활성을 자극), mRNA 결합하여 드로샤 (Drosha)-매개 프로세싱을 촉진해 성숙화에 영향을 미치는 것으로 확인되었다. 본 발명의 RNA-결합 이펙터 단백질은 하나 이상의 위치에서 RNA-결합 단백질의 결합을 방해할 수 있다고 예상된다.

[0273] ***i) RNA 폴딩***

[0274] RNA는 이의 생물학적 활성을 수행하기 위해 한정된 구조를 채택한다. 대안적인 3차 구조 중에서 입체형태의 전이는 대부분의 RNA-매개 과정에 핵심적이다. 그러나, RNA 폴딩은 몇몇 문제와 연관된다. 예를 들어, RNA는 부적절한 대체적인 입체형태로 폴딩되어 유지되는 경향을 가질 수 있고/있거나 올바른 3차 구조는 대체적인 구조에 대해 충분하게 열역학적으로 호의적이지 않을 수 있다. 본 발명의 RNA 표적화 이펙터 단백질, 특히 절단-결핍 또는 데드 RNA 표적화 단백질은 (m)RNA의 폴딩을 지정하고/하거나 이의 올바른 3차 구조를 포착하는데 사용될 수 있다.

[0275] **세포 상태의 조정에서 RNA-표적화 이펙터 단백질의 용도**

[0276] 일정 실시형태에서, cRNA와 복합체 내 Cas13b는 표적 RNA와 결합시 활성화되고 이후에 임의의 가까이의 ssRNA 표적을 절단한다 (즉, "부수적" 또는 "방관자" 효과). 동족 표적에 의해 프라이밍되면, Cas13b는 다른 (비상보성) RNA 분자를 절단할 수 있다. 이러한 불규칙한 RNA 절단은 잠재적으로 세포 독성을 야기할 수 있거나, 또는 아니면 세포 생리학 또는 세포 상태에 영향을 미칠 수 있다.

[0277] 따라서, 일정 실시형태에서, 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템은 세포 휴면에서 사용을 위한 것이거나 또는 세포 휴면의 유도를 위해 사용된다. 일정 실시형태에서, 본 명세서에서 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템은 세포 주기 정지의 유도에서 사용을 위한 것이거나 또는 세포 주기 정지의 유도를 위해 사용된다. 일정 실시형태에서, 본 명세서에서 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시

스템은 세포 성장 및/또는 세포 증식의 감소에서 사용을 위한 것이거나 또는 그를 위해 사용되고, 일정 실시형태에서, 본 명세서에서 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템은 세포 아네르기의 유도에서 사용을 위하거나 또는 그를 위해 사용된다. 일정 실시형태에서, 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템은 세포 아폽토시스에서 사용을 위한 것이거나 또는 그를 위해 사용된다. 일정 실시형태에서, 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템은 세포 괴사의 유도에서 사용을 위한 것이거나 또는 그를 위해 사용된다. 일정 실시형태에서, 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템은 세포 사멸의 유도에서 사용을 위한 것이거나 또는 그를 위해 사용된다. 일정 실시형태에서, 본 명세서에서 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템은 프로그램된 세포 사멸의 유도에서 사용을 위한 것이거나 또는 그를 위해 사용된다.

[0278] 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 세포 휴면의 유도를 위한 방법에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 세포 주기 정지의 유도를 위한 방법에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 세포 성장 및/또는 세포 증식의 감소를 위한 방법에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 세포 아네르기의 유도를 위한 방법에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 세포 아폽토시스의 유도를 위한 방법에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 세포 괴사의 유도를 위한 방법에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 세포 사멸의 유도를 위한 방법에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 프로그램된 세포 사멸의 유도를 위한 방법에 관한 것이다.

[0279] 본 명세서에 기술된 바와 같은 방법 및 용도는 치료적 또는 예방적일 수 있고 특정 세포, 세포 (하위)개체군, 또는 세포/조직 유형을 표적으로 할 수 있다. 특히, 본 명세서에 기술된 바와 같은 방법 및 용도는 치료적 또는 예방적일 수 있고 하나 이상의 표적 서열, 예컨대 하나 이상의 특정 표적 RNA (예를 들어, ssRNA)를 발현하는 특정 세포, 세포 (하위)개체군, 또는 세포/조직 유형을 표적으로 할 수 있다. 제한없이, 표적 세포는 예를 들어, 특정 전사물을 발현하는 암 세포, 예를 들어 소정 클래스의 뉴런, 예를 들어 자가면역을 야기하는 (면역) 세포, 또는 특이적 (예를 들어, 바이러스) 병원체에 의해 감염된 세포 등일 수 있다.

[0280] 따라서, 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는, 바람직하지 않은 세포 (숙주 세포)의 존재를 특징으로 하는 병리학적 병태를 치료하기 위한 방법에 관한 것이다. 일정 실시형태에서, 본 발명은 바람직하지 않은 세포 (숙주 세포)의 존재를 특징으로 하는 병리학적 병태를 치료하기 위한 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템의 용도에 관한 것이다. 일정 실시형태에서, 본 발명은 바람직하지 않은 세포 (숙주 세포)의 존재를 특징으로 하는 병리학적 병태를 치료하는데 사용을 위한 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템에 관한 것이다. 바람직하게 CRISPR-Cas 시스템은 바람직하지 않은 세포에 특이적인 표적을 표적으로 한다는 것을 이해한다. 일정 실시형태에서, 본 발명은 암을 치료하거나, 예방하거나, 또는 완화시키기 위한 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템의 용도에 관한 것이다. 일정 실시형태에서, 본 발명은 암을 치료하거나, 예방하거나, 또는 완화시키는데 사용을 위한 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는, 암을 치료하거나, 예방하거나, 또는 완화시키기 위한 방법에 관한 것이다. 바람직하게 CRISPR-Cas 시스템은 암 세포에 특이적인 표적을 표적으로 한다는 것을 이해해야 한다. 일정 실시형태에서, 본 발명은 병원체에 의한 세포의 감염을 치료

하거나, 예방하거나, 또는 완화시키기 위한 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템의 용도에 관한 것이다. 일정 실시형태에서, 본 발명은 병원체에 의한 세포의 감염을 치료하거나, 예방하거나, 또는 완화시키는데 사용을 위한 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는, 병원체에 의한 세포의 감염을 치료하거나, 예방하거나 또는 완화시키기 위한 방법에 관한 것이다. 바람직하게 CRISPR-Cas 시스템은 병원체에 의해 감염된 세포에 특이적인 표적 (예를 들어, 병원체 유래된 표적)을 표적으로 한다는 것을 이해해야 한다. 일정 실시형태에서, 본 발명은 자가면역 질환을 치료하거나, 예방하거나, 또는 완화시키기 위한 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템의 용도에 관한 것이다. 일정 실시형태에서, 본 발명은 자가면역 질환을 치료하거나, 예방하거나, 또는 완화시키는데 사용을 위한 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템에 관한 것이다. 일정 실시형태에서, 본 발명은 본 명세서에 기술된 바와 같은 비-천연 발생 또는 조작된 조성물, 벡터 시스템, 또는 전달 시스템을 도입시키거나 또는 유도시키는 단계를 포함하는 자가면역 질환을 치료하거나, 예방하거나, 또는 완화시키기 위한 방법에 관한 것이다. 바람직하게 CRISPR-Cas 시스템은 자가면역 질환의 원인이 되는 세포 (예를 들어, 특이적 면역세포)에 특이적인 표적을 표적으로 한다.

[0281] **RNA 검출에서 RNA-표적화 이펙터 단백질의 용도**

[0282] RNA 표적화 이펙터 단백질은 노던 블롯 어세이에서 사용될 수 있다는 것을 더욱 예상한다. 노던 블롯팅은 크기에 의해 RNA 샘플을 분리시키기 위한 전기영동의 사용을 포함한다. RNA 표적화 이펙터 단백질은 표적 RNA 서열에 특이적으로 결합하여 검출하는데 사용될 수 있다.

[0283] RNA 표적화 이펙터 단백질은 형광발광성 단백질 (예컨대, GFP)에 융합되어 살아있는 세포에서 RNA 국재화를 추적하는데 사용될 수 있다. 보다 특히, RNA 표적화 이펙터 단백질은 RNA를 더 이상 절단하지 않는다는 점에서 불활성화될 수 있다. 특정 실시형태에서, 분할 RNA 표적화 이펙터 단백질을 사용할 수 있어서, 보다 정밀한 가시화를 보장하기 위해서, 신호는 양쪽 하위단백질의 결합에 의존적이라는 것을 고려한다. 대안적으로, 분할 형광발광성 단백질은 다수의 RNA 표적화 이펙터 단백질 복합체가 표적 전사물에 결합될 때 재구성되는 것을 사용할 수 있다. 전사물은 mRNA를 따라서 다수의 결합 부위에서 표적화되어, 형광발광 신호가 진짜 신호를 증폭시킬 수 있고 병소 동정을 가능케한다는 것을 더욱 고려한다. 또 다른 대안으로서, 형광발광성 단백질은 분할 인테인으로부터 재구성될 수 있다.

[0284] RNA 표적화 이펙터 단백질은 예를 들어 RNA 또는 특이적 스플라이싱 변이체의 국재화, mRNA 전사물의 수준, 전사물의 상향 또는 하향 조절 및 질환 특이적 진단을 결정하는데 적합하게 사용된다. RNA 표적화 이펙터 단백질은 세포의 고수율 스크리닝 및 세포 분류 수 생존 세포의 회수를 가능하게 하는 형광발광-활성화 세포 분류법 (FACS)과 같은, 예를 들어 유세포측정법 또는 형광 현비경을 사용하여 (생존) 세포에서 RNA를 가시화하는데 사용될 수 있다. 더욱이, 상이한 전사물의 발현도는 스트레스, 예를 들어 세포에 대해 저산소 상태 또는 분자 억제제를 사용하는 암 성장의 억제 하에서 동시에 평가될 수 있다. 다른 적용분야는 이광자 현미경을 사용하는 신경 자극 동안 시냅스 연결부로 전사물의 국재화를 추적하는 것일 수 있다.

[0285] 일정 실시형태에서, 본 명세서에 기술된 바와 같은 본 발명에 따른 성분 또는 복합체는 예컨대 예를 들어 (형광발광으로) 표지된 Cas13b 이펙터를 사용한, 복합 오차-강건한 형광발광 인 시츄 혼성화에서 사용될 수 있다 (MERFISH; Chen et al. Science; 2015; 348(6233)).

[0286] **시험관 내 향점(apex) 표지화**

[0287] 세포 프로세스는 단백질, RNA 및 DNA 간의 분자 상호작용의 네트워크에 의존적이다. 단백질-DNA 및 단백질-RNA 상호작용의 정확한 검출은 이러한 과정을 이해하기 위한 열쇠이다. 시험관내 근접 표지화 기술은 예를 들어 시험관 내에서 관심 단백질 또는 RNA에 근접하여 RNA 및 폴리펩티드를 표지하기 위한 광활성화 프로브와 조합된 친화성 태그를 적용한다. UV 조사 후, 광활성화가능한 기는 태그화된 분자에 매우 근접해 있는 단백질 및 다른 분자와 상호작용하여, 그들을 표지화한다. 표지된 상호작용 분자를 이후에 회수하고 동정할 수 있다. 본 발명의 RNA 표적화 이펙터 단백질은 예를 들어 RNA 서열을 선택하기 위한 프로브를 표적화하는데 사용될 수 있다.

[0288] 이들 적용분야는 또한 배양이 어려운 세포 유형 또는 질환 관련 적용 분야의 생체내 이미지화를 위한 동물 모델

에 적용될 수 있다.

[0289] **RNA 오리가미 (origami)/시험관 내 조립 라인에서 RNA-표적화 이펙터 단백질의 사용 - 조합론**

[0290] RNA 오리가미는 통합된 주형으로서 RNA를 사용하여 2 차원 또는 3 차원 구조를 생성하기 위한 나노스케일 폴딩된 구조를 지칭한다. 폴딩된 구조는 RNA에서 인코딩되고 따라서 최종 RNA의 형상은 합성된 RNA 서열에 의해 결정된다 (Geary, et al. 2014. Science, 345 (6198). pp. 799-804). RNA 오리가미는 다른 성분들, 예컨대 단백질을 복합체로 배열시키기 위한 스캐폴드로서 작용할 수 있다. 본 발명의 RNA 표적화 이펙터 단백질은 예를 들어 적합한 가이드 RNA를 사용하는 RNA 오리가미에 대해 관심 단백질을 표적화하는데 사용될 수 있다.

[0291] 이들 적용분야는 또한 배양이 어려운 세포 유형 또는 질환 관련 적용 분야의 생체내 이미지화를 위한 동물 모델에 적용될 수 있다.

[0292] **RNA 단리 또는 정제, 농축 또는 감손에서 RNA-표적화 이펙터 단백질의 용도**

[0293] RNA와 복합체형성시 RNA 표적화 이펙터 단백질은 RNA를 단리하고/하거나 정제하는데 사용될 수 있다는 것을 더욱 예상한다. RNA 표적화 이펙터 단백질은 예를 들어 RNA-RNA 표적화 이펙터 단백질 복합체를 단리 및/또는 정제하는데 사용할 수 있는 친화성 태그에 융합될 수 있다. 이러한 적용분야는 예를 들어 세포에서 유전자 발현 프로파일의 분석에 유용하다.

[0294] 특정 실시형태에서, RNA 표적화 이펙터 단백질은 특이적 논코딩 RNA (ncRNA)를 표적으로 하여 이의 활성을 차단하는데 사용할 수 있어서, 유용한 기능적 프로브를 제공할 수 있다는 것을 예상할 수 있다. 일정 실시형태에서, 본 명세서에 기술된 이펙터 단백질은 특정 RNA에 대해 특별하게 농축하거나 (제한없이, 안정성 증가 등을 포함), 대안적으로 특정 RNA (예컨대, 제한없이 예를 들어 특정 스플라이싱 변이체, 이소폼 등)를 특이적으로 감손시키는데 사용될 수 있다.

[0295] **lincRNA 기능 및 다른 핵 RNA의 문의**

[0296] 현행 RNA 녹다운 전략, 예컨대 siRNA는 단백질 기구가 세포액에 있으므로 그들이 대체로 세포액 전사물을 표적화하는데 국한된다는 단점을 갖는다. 세포 기능에 필수적이지 않은 외인성 시스템인, 본 발명의 RNA 표적화 이펙터 단백질의 장점은 세포 내 임의 구획에서 사용할 수 있다는 것이다. NLS 신호를 RNA 표적화 이펙터 단백질에 융합시켜서, 핵으로 가이드할 수 있어, 핵 RNA를 표적화하는 것이 가능하다. 예를 들어, lincRNA의 기능을 프로브하는 것을 예상한다. 장형 유전자간 논-코딩 RNA (Long intergenic non-coding RNA) (lincRNA)는 많이 조사되지 않은 연구 영역이다. 대부분의 lincRNA는 본 발명의 RNA 표적화 이펙터 단백질을 사용하여 연구할 수 있는 아직 알려지지 않은 기능을 갖는다.

[0297] **RNA 결합 단백질의 동정**

[0298] 특이적 RNA에 결합하는 단백질의 동정은 많은 RNA의 규칙을 이해하기 위해 유용할 수 있다. 예를 들어, 많은 lincRNA는 전사를 제어하기 위한 전사 및 후생유전자 조절인자와 연관되어 있다. 소정 lincRNA에 어떠한 단백질이 결합하는 가에 대한 이해는 소정 조절 경로의 성분을 밝히는데 도움을 줄 수 있다. 본 발명의 RNA 표적화 이펙터 단백질은 바이오틴으로 국소 결합된 단백질을 표지화시키기 위해 특이적 전사물로 바이오틴 리가제를 동원하도록 디자인될 수 있다. 다음으로, 단백질을 풀다운시킬 수 있고 질량 분광법으로 분석하여 그들을 동정할 수 있다.

[0299] **RNA 상에서 복합체의 조립 및 기질 셔틀링**

[0300] 본 발명의 RNA 표적화 이펙터 단백질은 RNA 상에서 복합체를 조립하는데 더욱 사용될 수 있다. 이는 다수의 관련 단백질 (예를 들어, 특정 합성 경로의 성분)으로 RNA 표적화 이펙터 단백질을 작용화시켜 달성될 수 있다. 대안적으로, 다수의 RNA 표적화 이펙터 단백질은 이러한 상이한 관련 단백질로 작용화되어 동일하거나 또는 인접한 표적 RNA를 표적화할 수 있다. RNA 상에서 복합체를 조립하는 유용한 적용분야는 예를 들어 단백질 사이에서 기질 셔틀링을 촉진하는 것이다.

[0301] **합성 생물학**

[0302] 생물학적 시스템의 개발은 임상적 적용분야를 포함하여, 폭넓은 유용성을 갖는다. 본 발명의 프로그램가능한 RNA 표적화 이펙터 단백질은 예를 들어 표적 전사물로서 암-연결된 RNA를 사용하여, 표적화된 세포 사멸을 위해 독성 도메인의 분할 단백질에 융합되는데 사용될 수 있다는 것을 고려한다. 더욱이, 단백질-단백질 상

호작용을 포함하는 경로는 적절한 이펙터 예컨대 키나제 또는 다른 효소와 예를 들어 융합 복합체에 의해 합성 생물학적 시스템에서 영향받을 수 있다.

[0303] **단백질 스플라이싱:인테인**

[0304] 단백질 스플라이싱은 인테인이라고 하는, 개재 폴리펩티드가 엑스테인이라고 하는, 이에 측접된 폴리펩티드로부터 그 자신의 절제를 비롯하여, 엑스테인의 후속 결찰을 촉매하는 번역 후 과정이다. 표적 전사물 상에서 본 명세서에 기술된 바와 같은 둘 이상의 RNA 표적화 이펙터 단백질의 조립은 분할 인테인의 방출을 유도하는데 사용될 수 있어서 (Topilina and Mills Mob DNA. 2014 Feb 4;5(1):5), mRNA 전사물의 존재의 직접 컴퓨터화 및 단백질 생성물, 예컨대 대사 효소 또는 전사 인자(전사 경로의 하류 발동을 위함)의 후속 방출을 가능하게 한다. 이러한 적용분야는 합성 생물학 (상기 참조) 또는 대규모 생물생산 (일정 조건 하에서만 오직 생성물을 생산)에서 상당한 관련성을 가질 수 있다.

[0305] **유도성, 용량화, 및 자가-불활성화 시스템**

[0306] 일 실시형태에서, 본 발명의 RNA 표적화 이펙터 단백질 및 이펙터 성분을 포함하는 융합 복합체가 유도성, 예를 들어 광유도성 또는 화학적 유도성으로 디자인된다. 이러한 유도성은 바람직한 시기에 이펙터 성분의 활성화를 가능하게 한다.

[0307] 광유도성은 예를 들어 CRY2PHR/CIBN 쌍형성을 융합에 사용하는 융합 복합체를 디자인하여 달성된다. 이러한 시스템은 살아있는 세포 내에서 단백질 상호작용의 광유도에 특히 유용하다 (Konermann S, et al. Nature. 2013;500:472-476] 참조).

[0308] 화학 유도성은 예를 들어 FKBP/FRB (FK506 결합 단백질 / FKBP 라파마이신 결합) 쌍형성이 융합에 사용되는 융합 복합체를 디자인하여 제공된다. 이러한 시스템을 사용시 라파마이신이 단백질의 결합에 필요하다 (Zetsche et al. Nat Biotechnol. 2015;33(2):139-42는 Cas9 에 대한 상기 시스템의 용도를 기술한다).

[0309] 더욱이, DNA로서 세포에 도입시, 본 발명의 RNA 표적화 이펙터 단백질은 유도성 프로모터, 예컨대 테트라사이클린 또는 독시사이클린 제어된 전사 활성화 (Tet-온 및 Tet-오프 발현 시스템), 호르몬 유도성 유전자 발현 시스템 예컨대 예를 들어 엑디손 유도성 유전자 발현 시스템 및 아라비노스-유도성 유전자 발현 시스템에 의해 조정될 수 있다. RNA로서 전달 시, RNA 표적화 이펙터 단백질의 발현은 테트라사이클린과 같은 소형 분자를 감지할 수 있는 리보스위치를 통해 조정될 수 있다 (문헌 [Goldfless et al. Nucleic Acids Res. 2012;40(9):e64] 참조).

[0310] 일 실시형태에서, 본 발명의 RNA 표적화 이펙터 단백질의 전달은 세포 내에서 단백질 또는 crRNA의 양을 변화시키도록 조정되어서, 바람직한 효과 또는 임의의 바람직하지 않은 표적외 효과의 규모를 변화시킬 수 있다.

[0311] 일 실시형태에서, 본 명세서에 기술된 RNA 표적화 이펙터 단백질은 자가 불활성화되도록 디자인될 수 있다. 복제 RNA 치료제로서 또는 mRNA의 RNA로서 세포로 전달 시 (Wrobleska et al Nat Biotechnol. 2015 Aug; 33(8): 839-841), 그들은 그 자신의 RNA를 파괴하여 발현 및 이후의 효과를 자가 불활성화시켜서, 거주성 및 잠재적인 바람직하지 않은 효과를 감소시킬 수 있다.

[0312] 본 명세서에 기술된 바와 같은 RNA 표적화 이펙터 단백질의 추가적인 생체 내 적용분야를 위해서, 참조로 본 명세서에 편입되는, 다음의 문헌들 [Mackay JP et al (Nat Struct Mol Biol. 2011 Mar;18(3):256-61)], [Nelles et al (Bioessays. 2015 Jul;37(7):732-9)] 및 [Abil Z and Zhao H (Mol Biosyst. 2015 Oct;11(10):2658-65)]을 참조한다. 특히, 다음의 적용은 본 발명의 특정 실시형태, 바람직하게는 촉매적으로 불활성인 Cas13b를 사용하고: 번역을 향상시키고 (예, Cas13b - 번역 촉진 인자 융합 (예, eIF4 융합)); 번역을 억제하고 (예, 리보솜 결합 부위를 표적으로하는 gRNA); 엑손을 건너뛰고 (예, 스플라이스 공여자 및/또는 수용체 부위를 표적으로하는 gRNA); 엑손을 포함하고 (예, 특정 엑손 스플라이스 공여자 및/또는 수용체 부위를 포함시키거나 Cas13b 융합시키거나 스플라이스좀 성분을 동원하는 것을 표적으로 하는 gRNA (예, U1 snRNA)에); RNA 국재화에 접근하고 (예, Cas13b - 마커 융합 (예, EGFP 융합)); RNA 국재화를 변경하고 (예, Cas13b - 국재화 신호 융합 (예, NLS 또는 NES 융합)); RNA 분해 (이 경우, 촉매적으로 불활성인 Cas13b는 Cas13b의 활성에 의존하는 경우 사용되지 않아야 하고, 대안 적으로 그리고 증가된 특이성을 위해 스플릿 Cas13b가 사용될 수 있음); 기능성 부위에 대한 gRNA의 분해 또는 결합에 의한 것과 같은 비-코딩 RNA 기능의 억제 (예, miRNA)의 억제 (Cas13b-신호 서열 융합에 의한 재국재화에 의해 특정 부위에서 적정될 수 있음) 에 의해 특정 실시형태에서 예상된다.

[0313] 본 명세서에서 앞서 기술하고 실시예에서 입증하는 바와 같이, Cas13b 기능은 crRNA의 5' 또는 3' 연장 및

crRNA 루프의 연장에 대해 강력하다. 따라서 MS2 루프 및 다른 동원 도메인이 복합체 형성 및 표적 전사물과의 결합에 영향없이 crRNA에 첨가될 수 있다는 것을 예상한다. 다양한 이펙터 도메인의 동원을 위한 crRNA에 대한 이러한 변형은 상기 기술된 RNA 표적화된 이펙터 단백질의 사용에서 적용가능하다.

[0314] Cas13b는 RNA 파지에 대한 내성을 매개할 수 있다. 따라서, Cas13b는 제한없이 에볼라 바이러스 및 지카 바이러스를 포함한, RNA-단독 병원체에 대항하여, 예를 들어 동물, 인간 및 식물을 면역시키는데 사용될 수 있다는 것을 예상한다.

[0315] 일정 실시형태에서, Cas13b는 그 자신의 어레이를 프로세싱 (절단)할 수 있다. 이는 본원에 논의된 바와 같은 하나 이상의 돌연변이된 아미노산 잔기를 함유하는 야생형 Cas13b 단백질 및 돌연변이된 Cas13b 단백질 모두에 적용된다. 따라서, 상이한 표적 전사물 및/또는 적용분야에 대해 디자인된 다수의 crRNA는 하나의 프로모터에 의해 구동된 단일 전사물로서 또는 단일 프리-crRNA로서 전달될 수 있다는 것을 예상한다. 이러한 전달 방법은 실질적으로 보다 촘촘하고, 더 쉽게 합성되고 더 쉽게 바이러스 시스템으로 전달된다는 장점을 갖는다. 정확한 아미노산 위치는 본원의 오르소로그에 따라 달라질 수 있음을 이해해야 한다. Cas13b는 당업계에 공지되어 있고 본원의 다른 곳에 기재된 바와 같이 단백질 정렬에 의해 적절히 결정될 수 있다.본 발명의 측면은 또한, 예를 들어 시험관 내, *생체 내* 또는 *생체 외에서* 원핵생물 또는 진핵생물 세포에서 하나 이상의 유전자 또는 하나 이상의 유전자 산물의 발현을 변경 또는 조작하기 위한 게놈 공학에서의 본원에 기재된 조성물 및 시스템의 용도 및 방법을 포함한다.

[0316] 일 측면에서, 본 발명은 세포에서 표적 RNA의 발현을 조정, 예를 들어 감소시키기 위한 방법 및 조성물을 제공한다. 주제 방법에서, 본 발명의 Cas13b 시스템은 RNA의 전사, 안정성, 및/또는 번역을 간섭하는 것이 제공된다.

[0317] 일정 실시형태에서, Cas13b 시스템의 유효량은 RNA를 절단하거나 또는 아니면 RNA 발현을 억제하는데 사용된다. 이러한 점에서, 시스템은 siRNA 및 shRNA와 유사한 용도를 가지며, 따라서 또한 이러한 방법을 대체할 수 있다. 방법은 제한없이, 리보핵산 (예컨대, siRNA 또는 shRNA) 또는 이의 전사 주형, 예를 들어, shRNA를 인코딩하는 DNA를 간섭하기 위한 대체물로서 Cas13b 시스템의 용도를 포함한다. Cas13b 시스템은 예를 들어 표적 세포를 포함하는 포유동물에 투여하여, 표적 세포에 도입된다.

[0318] 유리하게, 본 발명의 Cas13b 시스템은 특이적이다. 예를 들어, 간섭 리보핵산 (예컨대 siRNA 또는 shRNA) 폴리뉴클레오티드 시스템은 디자인 및 안정성 문제 및 표적외 결합으로 인해 번거롭지만, 본 발명의 Cas13b 시스템은 높은 특이성을 갖게 디자인될 수 있다.

[0319] **탈안정화된 Cas13b**

[0320] 일 실시형태에서, 본원에 기재된 바와 같은 본 발명에 따른 이펙터 단백질은 탈안정화 도메인 (destabilization domain) (DD)과 연합되거나 또는 융합된다. 일부 실시형태에서, DD은 ER50이다. 이 DD에 대한 상응하는 안정화 리간드는, 일부 실시형태에서, 4HT이다. 이와 같이, 일부 실시형태에서, 적어도 하나의 DD 중 하나는 ER50이고 이의 안정화 리간드는 4HT 또는 CMP8이다. 일부 실시형태에서, DD는 DHFR50이다. 이 DD에 대한 상응하는 안정화 리간드는, 일부 실시형태에서, TMP이다. 이와 같이, 일부 실시형태에서, 적어도 하나의 DD 중 하나는 DHFR50이고, 이의 안정화 리간드는 TMP이다. 일부 실시형태에서, DD은 ER50이다. 이 DD에 대한 상응하는 안정화 리간드는, 일부 실시형태에서, CMP8이다. 그러므로, CMP8은 ER50 시스템에서 4HT에 대한 대안적인 안정화 리간드일 수 있다. CMP8 및 4HT가 경쟁적 방식으로 사용될 수 있고/사용되어야 하는 것이 가능할 수 있지만, 일부 세포 유형은 이들 2개 리간드 중 하나 또는 다른 하나에 더 감수성일 수 있으며, 본 개시 및 당분야의 지식으로부터 당업자는 CMP8 및/또는 4HT를 사용할 수 있다.

[0321] 일부 실시형태에서, 하나 또는 두개의 DD는 Cas13b의 C 말단에 융합된 하나 또는 두개의 DD와 Cas13b의 N 말단에 융합될 수 있다. 일부 실시형태에서, 적어도 2개의 DD는 Cas13b와 연합되고 DD는 동일한 DD이고, 즉 DD는 상동성이다. 따라서, DD 둘 모두 (또는 둘 이상)은 ER50 DD일 수 있다. 이는 일부 실시형태에서 바람직하다. 대안적으로, DD 둘 모두 (또는 둘 이상)은 DHFR50 DD일 수 있다. 이는 또한 일부 실시형태에서 바람직하다. 일부 실시형태에서, 적어도 2개의 DD는 CRISPR 효소와 연합되고 DD는 상이한 DD이며, 즉 DD는 이종성이다. 따라서, DD 중 하나는 ER50일 수 있는 한편 DD 중 하나 이상 또는 임의의 다른 DD는 DHFR50일 수 있다. 더 높은 수준의 분해 제어도를 제공하므로 이종성인 2 이상의 DD를 갖는 것이 유리할 수 있다. N 또는 C-말단에서 하나 초과의 DD의 탠덤 융합은 분해를 향상시킬 수 있으며; 이러한 탠덤 융합은 예를 들어 ER50-ER50-Cas13b 또는 DHFR-DHFR-Cas13b 일 수 있다. 이는, 어느 하나의 안정화 리간드 부

재시 높은 수준의 분해가 발생할 것이며, 하나의 안정화 리간드 부재 및 기타 다른(또는 또 다른) 안정화 리간드의 존재시 중간 수준의 분해가 발생할 것인 한편, 2 개(또는 둘 이상)의 모든 안정화 리간드의 존재 하에서 낮은 수준의 분해가 발생할 것임이 생각된다. 또한 N-말단 ER50 DD 및 C-말단 DHFR50 DD의 소유에 의해 제어가 부여될 수 있다.

[0322] 일부 구현예에서, DD와 Cas13b의 융합은 DD와 Cas13b 사이의 링커를 포함한다. 일부 구현예에서, 링커는 GlySer 링커이다. 일부 구현예에서, DD-Cas13b는 적어도 하나의 핵외 수송 신호(NES)를 추가로 포함한다. 일부 구현예에서, DD-Cas13b는 둘 이상의 NES를 포함한다. 일부 구현예에서, DD-Cas13b는 적어도 하나의 핵 국재화 신호(NLS)를 포함한다. 이는 NES에 추가적일 수 있다. 일부 구현예에서, Cas13b는, Cas13b와 DD 사이의 링커로서 또는 그 일부로서, 국소화(핵 이동 또는 핵외 수송) 신호를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다. HA 또는 플랙 태그(Flag tag) 또한 링커로서 본 발명의 영역 내에 있다. 출원인은 링커로서 NLS 및/또는 NES를 사용하고, 또한 GS 내지 (GGGGS)3 정도로 짧은 글리신 세린 링커를 사용한다.

[0323] 탈안정화 도메인을 광범위한 단백질에 대해 불안정성을 부여하는데 일반적인 유용성을 가지며, 예를 들어 참조로 본 명세서에 편입되는 문헌 [Miyazaki, J Am Chem Soc. Mar 7, 2012; 134(9): 3942-3945]을 참조한다. CMP8 또는 4-하이드록시타목시펜은 탈안정화 도메인일 수 있다. 보다 일반적으로, N-말단 규칙에 의해 불안정화 잔기인 포유동물 DHFR (DHFRts)의 온도-민감성 돌연변이체는 허용 온도에서는 안정하지만 37 ℃에서는 불안정한 것으로 밝혀졌다. 포유동물 DHFR에 대한 고-친화도 리간드인 메토트렉세이트의 DHFRts를 발현하는 세포에 대한 첨가는 단백질의 분해를 부분적으로 억제하였다. 이는 작은 분자 리간드가, 그렇지 않으면 세포 내에서 분해를 위해 표적화될 단백질을 안정화할 수 있다는 중요한 설명이었다. 라파마이신 유도체를 사용하여 mTOR의 FRB 도메인의 불안정한 돌연변이체 (FRB*)를 안정화시키고 융합된 키나아제, GSK-3β.6,7의 기능을 회복시켰다. 이 시스템은 리간드-의존적 안정성이 복잡한 생물학적 환경에서 특정 단백질의 기능을 조절하는 매력적인 전략을 나타냈다는 것을 입증했다. 단백질 활성을 제어하기 위한 시스템은 유비퀴틴 상보성이, FK506-결합 단백질 및 FKBP12의 라파마이신 유도된 이량체에 의해 발생하는 경우, 작용성이 되는 DD를 수반할 수 있다. 인간 FKBP12 또는 ecDHFR 단백질의 돌연변이체는 이의 고친화성 리간드, 실드(Shield)-1 또는 트리메토프림(TMP) 각각의 부재 하에서 대사적으로 불안정하도록 조작될 수 있다. 이들 돌연변이체는 본 발명의 실시에 유용한 가능한 탈안정화 도메인(DD)의 일부이며, Cas13b와의 융합물로서 DD의 불안정성이 프로테아솜에 의한 전체 융합 단백질의 Cas13b 분해에 부여된다. 실드-1 및 TMP는 용량-의존적 방식으로 DD에 결합하고 이를 안정화시킨다. 에스트로겐 수용체 리간드 결합 도메인 (ERLBD, ERS1의 잔기 305-549)는 또한 탈안정화 도메인으로서 조작될 수 있다. 에스트로겐 수용체 신호화 경로는 다양한 질병, 예컨대 유방암에 수반되기 때문에, 이 경로는 널리 연구되었고, 에스트로겐 수용체의 다양한 작용제(agonist) 및 길항제가 개발되어 왔다. 따라서, ERLBD와 약물의 상용성 쌍이 알려져 있다. 돌연변이체에 결합하지만 ERLBD의 야생 형태에는 결합하지 않는 리간드가 존재한다. 3 개의 돌연변이를 인코딩하는 이들 돌연변이체 도메인들 (L384M, M421G, G521R)12 중 하나를 사용함으로써, 내인성 에스트로겐-민감성 네트워크를 동요시키지 않는 리간드를 사용하여 ERLBD-유도된 DD의 안정성을 조절하는 것이 가능하다. 추가적인 돌연변이(Y537S)가 도입되어 ERLBD를 추가로 불안정화시키고, 잠재적 DD 후보물로서 이를 구성할 수 있다. 이러한 테트라-돌연변이체는 유리한 DD 개발이다. 돌연변이체 ERLBD는 Cas13b에 융합될 수 있고, 이의 안정성은 리간드를 사용하여 조절 또는 동요될 수 있으며, 이로 인해 Cas13b는 DD를 갖는다. 다른 DD는 쉴드-1 리간드에 의해 안정화된 돌연변이된 FKBP 단백질을 기반으로 12-kDa (107-아미노산) 태그일 수 있으며, 예를 들어, 문헌 [Nature Methods 5, (2008)]을 참조한다. 예를 들어, DD는 합성된, 생물학적으로 불활성인 소형 분자, 쉴드-1에 결합하여 그에 의해 가역적으로 안정화되는 변형된 FK506 결합 단백질 12 (FKBP12)일 수 있다; 예를 들어, 참고: Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell. 2006;126:995-1004; Banaszynski LA, Sellmyer MA, Contag CH, Wandless TJ, Thorne SH. Chemical control of protein stability and function in living mice. Nat Med. 2008;14:1123-1127; Maynard-Smith LA, Chen LC, Banaszynski LA, Ooi AG, Wandless TJ. A directed approach for engineering conditional protein stability using biologically silent small molecules. The Journal of biological chemistry. 2007;282:24866-24872; 및 Rodriguez, Chem Biol. Mar 23, 2012; 19(3): 391-398 - 이들 모두는 본 명세서에 참고로서 포함되어 있으며, 본 발명의 실시에서 Cas13b와 연합되는 DD를 선택하는 것에 있어서 본 발명의 실시에 이용될 수 있다. 확인할 수 있는 바와 같이, 당분야의 지식은 수많은 DD를 포함하고, DD는 유리하게 링커와 함께, Cas13b와 연합, 예를 들어 융합되고, 그리하여 DD는 리간드의 존재 하에서 안정화될 수 있고 이의 부재시에 DD는 탈안정

화될 수 있고, 그리하여 Cas13b가 전체적으로 탈안정화되거나, 또는 DD가 리간드의 부재 하에서 안정화될 수 있고 리간드가 존재하는 경우 DD는 탈안정화될 수 있으며, DD는 Cas13b 및 그에 따라 CRISPR-Cas13b 복합체 또는 시스템이 조절되거나 또는 제어될 수 있어서, 말하자면 켜지거나 또는 꺼질 수 있어서, 그렇게 함으로써 예를 들어 생체내 또는 시험관 환경에서 시스템의 조절 또는 제어를 위한 수단을 제공할 수 있다. 예를 들어, 관심 단백질이 DD 태그를 이용하여 융합물로서 발현된 경우, 이는 세포 내에서, 예를 들어 프로테아솜에 의해 탈안정화되고 신속히 분해된다. 따라서, 안정화 리간드의 부재는 분해되는 D 연합된 Cas를 초래한다. 새로운 DD가 관심 단백질에 융합된 경우, 이의 불안정성이 관심 단백질에 부여되어, 전체 융합 단백질의 신속한 분해를 초래한다. Cas에 대한 피크 활성도는 때때로 표적외 효과를 감소시키는 데 유익하다. 따라서, 높은 활성의 짧은 파열이 바람직하다. 본 발명은 이러한 피크를 제공할 수 있다. 일부 의미에서, 시스템은 유도성이다. 일부 기타 다른 의미에서, 안정화 리간드의 부재 하에서 시스템은 억제되고, 안정화 리간드의 존재 하에서 탈-억제된다.

[0324] **Cas13 돌연변이**

[0325] 특정 실시형태에서, 본원에 기재된 본 발명에 따른 이펙터 단백질 (CRISPR 효소; Cas13; 이펙터 단백질)은 촉매적으로 불활성이거나 또는 죽은 Cas13 이펙터 단백질 (dCas13)이다. 일부 실시형태에서, dCas13 이펙터는 뉴클레아제 도메인에 돌연변이를 포함한다. 일부 실시형태에서, dCas13 이펙터 단백질은 절단되었다. Cas13 이펙터 및 하나 이상의 기능성 도메인의 융합 단백질의 크기를 감소시키기 위해, Cas13 이펙터의 C-말단은 RNA 결합 기능을 여전히 유지하면서 절단될 수 있다. 예를 들어, 20 개 이상의 아미노산, 50 개 이상의 아미노산, 80 개 이상의 아미노산, 또는 100 개 이상의 아미노산, 또는 150 개 이상의 아미노산, 또는 200 개 이상의 아미노산, 또는 250 개 이상의 아미노산, 또는 300 개 이상의 아미노산, 또는 350 개 이상의 아미노산, 또는 최대 120 개의 아미노산, 또는 최대 140 개의 아미노산, 또는 최대 160 개의 아미노산, 또는 최대 180 개의 아미노산, 또는 최대 200 개의 아미노산, 또는 최대 250 개 이하의 아미노산, 또는 300 개 이하의 아미노산, 또는 350 개 이하의 아미노산, 또는 400 개 이하의 아미노산이, Cas13b 이펙터의 C-말단에서 절단될 수 있다.

Cas13 절단의 구체적인 예는 C-말단 Δ984-1090, C-말단 Δ1026-1090 및 C-말단 Δ1053-1090, C-말단 Δ934-1090, C-말단 Δ884-1090, C-말단 Δ834-1090, C-말단 Δ784-1090 및 C-말단 Δ734-1090을 포함하고, 여기서 아미노산 위치는 Prevotella sp. P5-125 Cas13b 단백질의 아미노산 위치에 상응한다. 도 28 을 참고한다.

[0326] **Cas13 이펙터 단백질의 조절**

[0327] 본 발명은 CRISPR 단백질 기능을 조절하는 보조 단백질을 제공한다. 특정 실시형태에서, 보조 단백질은 CRISPR 단백질의 촉매 활성을 조절한다. 본 발명의 일 실시형태에서, 보조 단백질은 표적화된, 또는 서열 특이적인, 뉴클레아제 활성을 조절한다. 본 발명의 일 실시형태에서, 보조 단백질은 부수적인 뉴클레아제 활성을 조절한다. 본 발명의 일 실시형태에서, 보조 단백질은 표적 핵산에 대한 결합을 조절한다.

[0328] 본 발명에 따르면, 조절될 뉴클레아제 활성은 mRNA, miRNA, siRNA 및 절단가능한 RNA 연결을 포함하는 핵산을 뉴클레오티드 유사체와 함께 제한 없이 포함하는 RNA를 포함하거나 이로 구성된 핵산에 대해 지시될 수 있다.

본 발명에 따르면, 조절될 뉴클레아제 활성은 절단가능한 DNA 연결을 포함하는 핵산 및 핵산 유사체를 제한 없이 포함하는 DNA를 포함하거나 이로 구성된 핵산에 대해 지시될 수 있다.

[0329] 본 발명의 일 실시형태에서, 보조 단백질은 CRISPR 단백질의 활성을 향상시킨다. 이러한 특정 실시형태에서, 보조 단백질은 HEPN 도메인을 포함하고 RNA 절단을 향상시킨다. 특정 실시형태에서, 보조 단백질은 CRISPR 단백질의 활성을 억제한다. 이러한 특정 실시형태에서, 보조 단백질은 비활성화된 HEPN 도메인을 포함하고 HEPN 도메인이 함께 결핍된다.

[0330] 본 발명에 따르면, 타입 VI CRISPR 시스템의 자연 발생 보조 단백질은 CRISPR 단백질의 활성을 변형시키는 기능을 하는 CRISPR 유전자좌에서 또는 그 근처에 인코딩된 소형 단백질을 포함한다. 일반적으로 CRISPR 유전자좌는 추정 CRISPR 어레이를 포함하고/하거나 추정 CRISPR 이펙터 단백질을 인코딩하는 것으로 식별될 수 있다. 한 실시형태에서, 이펙터 단백질은 800 내지 2000 개 아미노산, 또는 900 내지 1800 개 아미노산, 또는 950 내지 1300 개 아미노산일 수 있다. 한 실시형태에서, 보조 단백질은 추정 CRISPR 이펙터 단백질 또는 어레이의 25kb 내, 또는 20kb 내 또는 15kb 내, 또는 10kb 내에, 또는 추정 CRISPR 이펙터 단백질 또는 어레이로부터 2kb 내지 10kb 내에 인코딩될 수 있다.

[0331] 한 실시형태에서, 보조 단백질은 50 내지 300 개 아미노산, 또는 100 내지 300 개 아미노산, 또는 150 내지 250

개 아미노산 또는 약 200 개 아미노산이다. 보조 단백질의 비-제한적인 예는 본원에서 확인된 csx27 및 csx28 단백질을 포함한다.

[0332] 본 발명의 CRISPR 보조 단백질의 확인 및 사용은 CRISPR 이펙터 단백질 분류와 무관하다. 본 발명의 보조 단백질은 다양한 CRISPR 이펙터 단백질과 관련되는 것으로 발견되거나 기능하도록 조작될 수 있다. 본원에서 확인되고 사용되는 보조 단백질의 예는 일반적으로 CRISPR 이펙터 단백질을 대표한다. CRISPR 이펙터 단백질 분류는 상동성, 특징 위치 (예를 들어, REC 도메인, NUC 도메인, HEPN 서열의 위치), 핵산 표적 (예를 들어, DNA 또는 RNA), 추적자 RNA의 부재 또는 존재, 직접 반복부의 가이드/스페이서 서열 5 '또는 3'의 위치, 또는 다른 기준을 포함할 수 있는 것으로 이해된다. 본 발명의 실시형태에서, 보조 단백질 식별 및 사용은 이러한 분류를 초월한다.

[0333] RNA를 표적화하는 타입 VI CRISPR-Cas 시스템에서, Cas 단백질은 일반적으로 RNA 절단에 관여하는 2 개의 보존된 HEPN 도메인을 포함한다. 특정 실시형태에서, Cas 단백질은 crRNA를 프로세싱하여 성숙한 crRNA를 생성한다. crRNA의 가이드 서열은 상보적 서열로 표적 RNA를 인지하고 Cas 단백질은 표적 가닥을 분해한다. 보다 구체적으로, 특정 실시형태에서, 표적 결합시, Cas 단백질은 2 개의 HEPN 도메인을 함께 가져 와서 활성 HEPN 촉매 부위를 형성하고 이어서 표적 RNA가 절단되는 구조적 재배열을 겪는다. Cas 단백질의 표면 근처의 촉매 부위의 위치는 비-특이적 부수적 ssRNA 절단을 허용한다.

[0334] 특정 실시형태에서, 보조 단백질은 표적 및/또는 부수적 RNA 절단을 증가시키거나 감소시키는데 도움이 된다. 이론에 구속되지 않고, CRISPR 활성을 활성화시키는 보조 단백질 (예를 들어, csx28 단백질 또는 오르소로그 또는 HEPN 도메인을 포함하는 변이체)은 Cas 단백질과 상호작용할 수 있고 그의 HEPN 도메인을 Cas 단백질의 HEPN 도메인과 조합시켜 활성 HEPN 촉매 부위를 형성하는 반면, 억제 보조 단백질 (예를 들어 HEPN 도메인이 결여된 csx27)은 Cas 단백질과 상호작용할 수 있고 2 개의 HEPN 도메인과 함께 Cas 단백질의 형태를 감소시키거나 차단할 수 있는 것으로 고려된다.

[0335] 본 발명에 따르면, 특정 실시형태에서, 타입 VI Cas 단백질 또는 이의 복합체의 활성을 향상시키는 것은 타입 VI Cas 단백질 또는 이의 복합체를 Cas 단백질을 활성화시키는 동일한 유기체로부터의 보조 단백질과 접촉시키는 것을 포함한다. 다른 실시형태에서, 타입 VI Cas 단백질 또는 이의 복합체의 활성을 향상시키는 것은 타입 VI Cas 단백질 또는 이의 복합체를 동일한 서브클래스 내의 (예, 타입 VI-b) 상이한 유기체로부터의 활성화제 보조 단백질과 접촉시키는 것을 포함한다. 다른 실시형태에서, 타입 VI Cas 단백질 또는 이의 복합체의 활성을 향상시키는 것은 타입 VI Cas 단백질 또는 이의 복합체를 서브클래스 내에 있지 않은 보조 단백질 (예, 타입 VI-b 이외의 타입 VI Cas 단백질과 타입 VI-b 보조 단백질 또는 그 반대로)과 접촉시키는 것을 포함한다.

[0336] 본 발명에 따르면, 특정 실시형태에서, 타입 VI Cas 단백질 또는 이의 복합체의 활성을 억제하는 것은 타입 VI Cas 단백질 또는 이의 복합체를 Cas 단백질을 억제하는 동일한 유기체로부터의 보조 단백질과 접촉시키는 것을 포함한다. 다른 실시형태에서, 타입 VI Cas 단백질 또는 이의 복합체의 활성을 억제하는 것은 타입 VI Cas 단백질 또는 이의 복합체를 동일한 서브클래스 내의 (예, 타입 VI-b) 상이한 유기체로부터의 억제제 보조 단백질과 접촉시키는 것을 포함한다. 다른 실시형태에서, 타입 VI Cas 단백질 또는 이의 복합체의 활성을 억제하는 것은 타입 VI Cas 단백질 또는 이의 복합체를 서브클래스 내에 있지 않은 억제제 보조 단백질 (예, 타입 VI-b 이외의 타입 VI Cas 단백질과 타입 VI-b 억제제 보조 단백질 또는 그 반대로)과 접촉시키는 것을 포함한다.

[0337] 타입 VI Cas 단백질 및 타입 VI 보조 단백질이 동일한 유기체로부터 유래된 특정 실시형태에서, 두 단백질은 조작된 CRISPR 시스템에서 함께 기능할 것이다. 특정 실시형태에서, 조작된 CRISPR 시스템의 기능, 예를 들어 단백질 중 하나 또는 둘 다 또는 그의 발현을 변형시킴으로써 기능을 변경하는 것이 바람직할 것이다. 타입 VI Cas 단백질 및 타입 VI 보조 단백질이 동일한 계열 또는 상이한 계열 내에 있을 수 있는 상이한 유기체로부터 유래된 실시형태에서, 단백질은 조작된 CRISPR 시스템에서 함께 기능할 수 있지만 종종 함께 작용하는 단백질 중 하나 또는 둘 다를 변형시키는 것이 바람직하거나 필요할 것이다.

[0338] 따라서, 본 발명의 특정 실시형태에서 Cas 단백질과 보조 단백질 중 하나 또는 둘 모두는 Cas 단백질과 보조 단백질 사이의 단백질-단백질 상호작용의 측면을 조정하도록 변형될 수 있다. 특정 실시형태에서, Cas 단백질 및 보조 단백질 중 하나 또는 둘 다는 단백질-핵산 상호작용의 측면을 조정하도록 변형될 수 있다. 단백질-단백질 상호작용 및 단백질-핵산 상호작용을 조정하는 방식은 제한 없이, 분자 표면 적합, 극성 상호작용, 수소 결합 및 반 데르 발스 상호작용 조절을 포함한다. 특정 실시형태에서, 단백질-단백질 상호작용 또는

단백질-핵산 결합을 조정하는 것은 결합 상호작용을 증가 또는 감소시키는 것을 포함한다. 특정 실시형태에서, 단백질-단백질 상호작용 또는 단백질-핵산 결합을 조정하는 것은 단백질 또는 핵산의 형태를 선호하거나 싫어하는 변형을 포함한다.

[0339] "피팅 (fitting)"은 자동 또는 반자동 수단에 의해, Cas13 단백질의 하나 이상의 원자와 Cas13 보조 단백질의 하나 이상의 원자 사이, 또는 Cas13 단백질의 하나 이상의 원자와 핵산의 하나 이상의 원자 사이, 또는 Cas13 보조 단백질의 하나 이상의 원자와 핵산 사이의 상호작용을 포함함으로써 결정하고, 이러한 상호작용이 안정적인 정도를 계산하는 것을 의미한다. 상호작용은 하전, 입체배치 고려 등에 의해 유발되는, 인력과 척력을 포함한다.

[0340] 타입 VI CRISPR 단백질 또는 이의 복합체 또는 타입 VI CRISPR 보조 단백질 또는 그의 복합체의 3 차원 구조는 본 발명의 맥락에서 Cas13의 오르소로그에서 추가 돌연변이를 확인하기 위한 추가 도구를 제공한다. 결정 구조는 또한 신규하고 특이적인 Cas13 및 Cas13 보조 단백질의 디자인을 위한 기초가 될 수 있다. 피팅을 위한 각종 컴퓨터-기반 방법이 추가로 설명된다. Cas13, 보조 단백질 및 핵산의 결합 상호작용은 도킹 프로그램을 사용하는 컴퓨터 모델링을 사용하여 검사할 수 있다. 도킹 프로그램은 공지되어 있다; 예를 들어 GRAM, DOCK 또는 AUTODOCK (참조, Walters et al. Drug Discovery Today, vol. 3, no. 4 (1998), 160-178, 및 Dunbrack et al. Folding and Design 2 (1997), 27-42). 이 절차는 결합 파트너의 형상 및 화학적 구조가 얼마나 잘되는지 확인하기 위한 컴퓨터 피팅을 포함할 수 있다. 타입 VI 시스템의 활성 부위 또는 결합 부위에 대한 컴퓨터 지원, 수동 검사가 수행될 수 있다. GRID (P. Goodford, J. Med. Chem, 1985, 28, 849-57)와 같은 프로그램 (다양한 작용기를 가진 분자 간의 가능한 상호작용 부위를 결정하는 프로그램)을 사용하여, 결합 화합물의 부분 구조를 예측하기 위해 활성 부위 또는 결합 부위를 분석할 수 있다. 컴퓨터 프로그램을 사용하여 2 개의 결합 파트너, 예를 들어 타입 VI CRISPR 시스템의 성분, 또는 핵산 분자 및 타입 VI CRISPR 시스템의 성분의 인력, 반발 또는 입체 장애를 추정할 수 있다.

[0341] 아미노산 특성(예컨대 잔기의 극성, 전하, 용해성, 소수성, 친수성 및/또는 양친매성)의 상이성 또는 유사성을 기초로 아미노산 치환이 만들어질 수 있으며, 따라서 이것은 아미노산을 기능적 그룹으로 함께 그룹화하는데 유용하다. 아미노산은 그것의 측쇄만의 특성에 기초하여 그룹화될 수 있다. 오르소로그를 비교할 때, 구조적 또는 촉매적 이유로 보존된 잔기가 있을 수 있다. 이들 세트는 벤 다이어그램의 형태로 기술될 수 있다 (Livingstone C.D. and Barton G.J. (1993) “Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation” *Comput. Appl. Biosci.* 9: 745-756) (Taylor W.R. (1986) “The classification of amino acid conservation” *J. Theor. Biol.* 119; 205-218). 보존적 치환은, 예를 들어 일반적으로 용인된 벤다이어그램 아미노산 그룹화를 설명한 아래 표에 따라서 만들어질 수 있다.

[0342]

세트		서브-세트	
소수성	FWYHKMILVAGC	방향족	FWYH
		지방족	ILV
극성	WYHKREDCSTNQ	하전됨	HKRED
		양전하	HKR
		음전하	ED
소형	VCAGSPTND	아주 작음	AGS

[0343] 조작된 Cas13 시스템에서, 변형은 Cas13 단백질의 하나 이상의 아미노산 잔기의 변형을 포함할 수 있고/거나 Cas13 보조 단백질의 하나 이상의 아미노산 잔기의 변형을 포함할 수 있다.

[0344] 조작된 Cas13 시스템에서, 변형은 비변형된 Cas13 단백질 및/또는 Cas13 보조 단백질에서 양으로 하전된 잔기를 포함하는 영역에 위치된 하나 이상의 아미노산 잔기의 변형을 포함할 수 있다.

[0345] 조작된 Cas13 시스템에서, 변형은 비변형된 Cas13 단백질 및/또는 Cas13 보조 단백질에서 양으로 하전된 하나 이상의 아미노산 잔기의 변형을 포함할 수 있다.

[0346] 조작된 Cas13 시스템에서, 변형은 비변형된 Cas13 단백질 및/또는 Cas13 보조 단백질에서 양으로 하전되지 않은

하나 이상의 아미노산 잔기의 변형을 포함할 수 있다.

[0347] 변형은 비변형된 Cas13 단백질 및/또는 Cas13 보조 단백질에서 미하전된 하나 이상의 아미노산 잔기의 변형을 포함할 수 있다.

[0348] 변형은 비변형된 Cas13 단백질 및/또는 Cas13 보조 단백질에서 음으로 하전된 하나 이상의 아미노산 잔기의 변형을 포함할 수 있다.

[0349] 변형은 비변형된 Cas13 단백질 및/또는 Cas13 보조 단백질에서 소수성인 하나 이상의 아미노산 잔기의 변형을 포함할 수 있다.

[0350] 변형은 비변형된 Cas13 단백질 및/또는 Cas13 보조 단백질에서 극성인 하나 이상의 아미노산 잔기의 변형을 포함할 수 있다.

[0351] 변형은 소수성 아미노산 또는 극성 아미노산의 하전된 아미노산으로의 치환을 포함할 수 있으며, 이는 음으로 하전된 또는 양으로 하전된 아미노산일 수 있다. 변형은 음으로 하전된 아미노산을 양으로 하전되거나 극성 또는 소수성 아미노산으로 치환하는 것을 포함할 수 있다. 변형은 양으로 하전된 아미노산을 음으로 하전되거나 극성 또는 소수성 아미노산으로 치환하는 것을 포함할 수 있다.

[0352] 본 발명의 실시형태는 일어날 수 있는 상동 치환(치환 및 대체는 본 명세서에서 기존 아미노산 잔기 또는 뉴클레오티드의, 대안적인 잔기 또는 뉴클레오티드와의 교환을 의미하기 위해 본원에서 사용됨), 즉 염기성에 염기성, 산성에 산성, 극성에 극성과 같은 아미노산의 경우 유사-대-유사 치환을 포함할 수 있는 서열 (폴리뉴클레오티드 또는 폴리펩티드 둘 다)을 포함한다. 비-상동 치환은 또한 한 계열의 잔기에서 다른 계열로 또는 대안적으로 오르니틴 (이하 Z로 지칭됨), 디아미노부티르산 오르니틴 (이하 B로 지칭 됨), 노르류신 오르니틴 (이하 O로 지칭됨), 피리릴알라닌, 티에닐알라닌, 나프틸알라닌 및 페닐글리신과 같은 비-천연 아미노산의 혼입을 수반하여 일어날 수 있다. 변이체 아미노산 서열은 글리신 또는 β-알라닌 잔기와 같은 아미노산 스페이서에 더하여 메틸, 에틸 또는 프로필 기와 같은 알킬 기를 포함하는 서열의 임의의 두 아미노산 잔기 사이에 삽입될 수 있는 적합한 스페이서 기를 포함할 수 있다. 펩토이드 형태로 하나 이상의 아미노산 잔기의 존재를 포함하는 추가의 형태의 변형도 당업자에게 잘 이해될 수 있다. 의심을 피하기 위해서, "펩토이드 형태"는 α-탄소 치환체 기가 α-탄소가 아니라 잔기의 질소 원자에 존재하는 변이체 아미노산 잔기를 말하기 위해서 사용된다. 펩토이드 형태로 펩티드를 제조하는 과정은 해당 분야에 공지되어 있다, 예를 들어: Simon RJ et al., *PNAS* (1992) 89(20), 9367-9371 and Horwell DC, *Trends Biotechnol*. (1995) 13(4), 132-134.

[0353] 상동성 모델링: 다른 Cas13 오르소로그에서 상응하는 잔기는 Zhang et al., 2012 (Nature; 490(7421): 556-60) 및 Chen et al., 2015 (PLoS Comput Biol; 11(5): e1004248)의 방법 -도메인-모티브 계면에 의해 매개되는 상호작용을 예측하기 위한 전산 단백질-단백질 상호작용 (PPI) 방법- 으로 확인할 수 있다. PrePPI(예측 PPI)는 구조 기준 PPI 예측 방법으로, 베이지안(Bayesian) 통계 프레임워크를 이용하여 비-구조적 증거와 구조적 증거를 조합한다. 이 방법은 문의되는 단백질의 쌍을 취하고, 구조적 정렬을 이용하여 그의 실험적으로 결정된 구조 또는 상동성 모델 중 어느 하나에 상응하는 구조적 대표물을 확인하는 것을 포함한다. 구조적 정렬은 전반적인 및 국소의 기하 관계를 고려함으로써 가까운 구조적 이웃 및 먼 구조적 이웃 모두를 확인하는 데 추가로 사용된다. 구조적 대표물의 두 이웃이 단백질 데이터 뱅크에 보고된 복합체를 형성할 때마다, 이는 두 개의 문의되는 단백질 사이의 상호작용을 모델링하기 위한 주형을 정의한다. 복합체의 모델은 대표 구조물을 주형 내에서 그의 상응하는 구조적 이웃에 겹쳐놓음으로써 창출된다. 이러한 접근법은 Dey et al., 2013 (Prot Sci; 22: 359-66)에 있다.

[0354] **식물 및 효모에 RNA 표적화-CRISPR 시스템의 적용**

[0355] 정의들:

[0356] 일반적으로, "식물"이라는 용어는 특징적으로 세포 분열에 의해 성장하고, 엽록소를 함유하고, 셀룰로오스로 구성된 세포벽을 갖는, 식물계의 임의의 각종 광합성, 진핵성, 단세포 또는 다세포 생물에 관한 것이다. 식물이라는 용어는 단자엽 식물 및 쌍자엽 식물을 포함한다. 구체적으로, 식물은 제한 없이 다음을 포함하고자 한다:속씨 식물 및 겉씨 식물, 예컨대 아카시아, 알팔파, 아마란스, 사과, 살구, 아티초크, 물푸레나무, 아스파라거스, 아보카도, 바나나, 보리, 콩, 비트, 자작나무, 너도밤나무, 블랙베리, 블루베리, 브로콜리, 방울 양배추, 양배추, 카놀라, 칸탈루프, 당근, 카사바, 콜리플라워, 삼나무, 곡물, 셀러리, 밤, 체리, 중국 배추, 감귤류, 클레멘타인, 클로버, 커피, 옥수수, 목화, 동부, 오이, 사이프러스, 가지, 느릅나무, 꽃상추, 유칼립투스, 회향, 무화과, 전나무, 제라늄, 포도, 자몽, 땅콩, 꽈리, 검 헴록(gum hemlock), 히코리, 케일, 키위, 콜라비,

낙엽송, 상추, 대파, 레몬, 라임, 로커스트(locust), 소나무, 공작고사리, 옥수수, 망고, 단풍나무, 멜론, 기장, 버섯, 겨자, 견과류, 참나무, 귀리, 기름야자, 오크라, 양파, 오렌지, 장식용 식물 또는 꽃 또는 나무, 파파야, 야자, 파슬리, 파스닙, 콩, 복숭아, 땅콩, 배, 피트(peat), 후추, 감, 나무콩, 소나무, 파인애플, 플랜테인, 자두, 석류, 감자, 호박, 라디키오, 무, 유채, 라스베리, 쌀, 호밀, 수수, 홍화, 갯버들, 대두, 시금치, 가문비 나무, 호박, 딸기, 사탕무, 사탕수수, 해바라기, 고구마, 스위트콘, 귤, 녹차, 담배, 토마토, 나무, 라이밀, 잔디풀, 순무, 포도나무, 호두, 물냉이, 수박, 밀, 참마, 주목나무, 주키니 호박. 식물이라는 용어는, 뿌리, 잎 및 기타 고등 식물을 특징짓는 기타 기관이 적어서 주로 일차적으로 연합된 광독립영양생물인, 조류를 또한 포함한다.

[0357] 본 명세서에 기술된 바와 같은 RNA 표적화 시스템을 사용하는 유전자 발현을 조정하기 위한 방법은 바람직한 특성을 본질적으로 임의의 식물에 부여하는데 사용될 수 있다. 광범위한 종류의 식물 및 식물 세포 시스템은, 본 개시 내용의 핵산 구조물과 상기 언급된 각종 형질전환 방법을 이용하여 본 명세서에 설명된 바람직한 생리학적 및 작물학적 특징들에 대해, 조작될 수 있다. 바람직한 실시형태에서, 조작을 위한 표적 식물 및 식물 세포는 제한 없이, 이들 외떡잎식물 및 쌍떡잎식물, 예컨대 곡물 (예를 들어, 밀, 옥수수, 쌀, 기장, 보리)을 포함하는 작물, 과일 작물 (예를 들어, 토마토, 사과, 배, 딸기, 오렌지), 사료 작물 (예를 들어, 알팔파), 뿌리 채소 작물 (예를 들어, 당근, 감자, 사탕무, 참마), 잎이 많은 채소 작물 (예를 들어, 상추, 시금치); 꽃 피는 식물 (예를 들어, 피튜니아, 장미, 국화), 침엽수 및 소나무 (예를 들어, 소나무 전나무, 가문비 나무); 식물치료에 사용되는 식물 (예를 들어, 중금속 축적 식물); 오일 작물 (예를 들어, 해바라기, 유채 씨) 및 실험 목적으로 사용되는 식물 (예를 들어, *애기장대*) 를 포함한다. 따라서, 방법 및 CRISPR-Cas 시스템은 광범위한 식물, 예컨대 예를 들어, Magniolales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papeverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, Santales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales, 및 Asterales 목에 속하는 쌍떡잎식물에 걸쳐 사용될 수 있고; 방법 및 CRISPR-Cas 시스템은 Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales, 및 Orchid ales 목에 속하는 외떡잎식물, 또는 Gymnospermae 에 속하는 식물, 예를 들어 Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales 및 Gnetales 목에 속하는 식물에 걸쳐 사용될 수 있다.

[0358] 본원에 기재된 RNA 표적화 CRISPR 시스템 및 사용 방법은 하기 쌍떡잎, 외떡잎 또는 겉씨식물 속의 비-제한적인 목록에 포함되는, 광범위한 식물 종에 걸쳐 사용될 수 있다: *Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glycine, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, Litsea, Lycopersicon, Lupinus, Manihot, Majorana, Malus, Medicago, Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Senecio, Sinomenium, Stephania, Sinapis, Solanum, Theobroma, Trifolium, Trigonella, Vicia, Vinca, Vilis,* 및 *Vigna; 및 속 Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, Zea, Abies, Cunninghamia, Ephedra, Picea, Pinus,* 및 *Pseudotsuga.*

[0359] RNA 표적화 CRISPR 시스템 및 사용 방법은 또한 광범위한 "조류" 또는 "조류 세포" 에 걸쳐 사용될 수 있으며; 예를 들어 Rhodophyta (적색 조류), Chlorophyta (녹색 조류), Phaeophyta (갈색 조류), Bacillariophyta (규조류), Eustigmatophyta 및 Dinoflagellates를 포함하여 여러 진핵생물 문(phyla) 뿐 아니라 원핵생물 문 Cyanobacteria (청색-녹색 조류)에서 선택된 조류를 포함한다. 용어 "조류"는 예를 들어 하기로부터 선택된 조류를 포함한다: Amphora, Anabaena, Anikstrodesmis, Botryococcus, Chaetoceros, Chlamydomonas, Chlorella, Chlorococcum, Cyclotella, Cylindrotheca, Dunaliella, Emiliana, Euglena, Hematococcus, Isochrysis, Monochrysis, Monoraphidium, Nannochloris, Nannnochloropsis, Navicula, Nephrochloris, Nephroselmis, Nitzschia, Nodularia, Nostoc, Oochromonas, Oocystis, Oscillartoria, Pavlova,

Phaeodactylum, Playtmonas, Pleurochrysis, Porhyra, Pseudoanabaena, Pyramimonas, Stichococcus, Synechococcus, Synechocystis, Tetraselmis, Thalassiosira, 및 Trichodesmium.

[0360] 식물의 부분, 즉 "식물 조직"은 본 발명의 방법에 따라 처리되어 개선된 식물을 생산할 수 있다. 식물 조직은 또한 식물 세포를 포함한다.본원에 사용된 용어 "식물 세포"는 온전한 전체 식물 또는 배지 또는 아가 상에, 성장 배지 또는 완충액 중의 현탁액에서, 고차 조직화된 단위의 일부, 예를 들어, 식물 조직, 식물 기관 또는 전체 식물로서 시험관 내 조직 배양에서 성장한 단리된 형태로, 살아있는 식물의 개별 단위를 지칭한다.

[0361] "원형질"이라는 용어는, 예를 들어 기계적 또는 효소적 수단을 이용하여 그의 보호성 세포 벽이 완전히 또는 부분적으로 제거된 식물 세포를 지칭하며, 그 결과 살아있는 식물의 온전한 생화학 컴피턴트 단위가 결과로서 초래되며, 이는 적절한 성장 조건 하에서 그의 세포 벽을 재형성하고, 번식 및 재생하여 전체 식물로 성장할 수 있다.

[0362] 용어 "형질전환"은 광범위하게 아그로박테리아 또는 다양한 화학적 또는 물리적 방법 중 하나에 의해 DNA를 도입함으로써 식물 숙주가 유전자 변형되는 과정을 의미한다. 본 명세서에 사용된 바와 같은 용어 "식물 숙주"는, 식물의 임의의 세포, 조직, 기관 또는 자손을 포함하는 식물을 지칭한다. 많은 적합한 식물 조직 또는 식물 세포가 형질전환될 수 있으며, 이로 제한되지는 않지만, 원형질, 체세포배(somatic embryo), 화분, 잎, 묘목, 줄기, 가피(calli), 기는 줄기(stolon), 미소괴경, 및 싹을 포함할 수 있다. 식물 조직은 또한, 유성 또는 무성으로 생성된 여부를 떠나, 그러한 식물, 종자, 자손, 번식체의 임의의 클론, 및 절단물 또는 종자와 같은 이들의 임의의 후손을 지칭한다.

[0363] 본 명세서에 사용된 바와 같은 용어 "형질전환된"은, 외래 DNA 분자, 예컨대 구축물이 그 안으로 도입된 세포, 조직, 기관 또는 생물을 지칭한다. 도입된 DNA 분자는 수령 세포, 조직, 기관 또는 생물의 게놈 DNA 내로 통합될 수 있어서, 도입된 DNA 분자가 이후의 자손에게 전달되도록 한다. 이들 구현예에서, "형질전환된" 또는 "유전자이식" 세포 또는 식물은, 교배에서 부모로서 그러한 형질전환된 식물을 이용하고, 도입된 DNA분자의 존재로부터 초래되는 변경된 표현형을 나타내는, 그 세포 또는 식물의 자손 및 육종 계획으로부터 생산된 자손을 또한 포함할 수 있다. 바람직하게는, 유전자이식 식물은 생식력이 있으며, 유성 생식을 통하여 도입된 DNA를 자손에게 전달할 수 있다.

[0364] 용어 "자손", 예컨대 유전자이식 식물의 자손은, 식물 또는 유전자이식 식 식물로부터 태어난, 얻은, 또는 유래된 것이다. 도입된 DNA 분자는 수령 세포 내로 일시적으로 도입될 수도 있어서, 도입된 DNA 분자가 이후의 자손에 의해 유전되지 않아 "유전자이식"이라고 여겨지지는 않는다. 이에 따라, 본 명세서에서 사용된 바와 같은, "비-유전자이식" 식물 또는 식물 세포는 그의 게놈 내로 안정하게 통합된 외래 DNA를 함유하지 않는 것이다.

[0365] 본 명세서에 사용된 바와 같은 용어 "식물 프로모터"는, 그의 기원이 식물 세포인지 아닌지의 여부를 떠나, 식물 세포에서 전사를 개시할 수 있는 프로모터이다. 예시적인 적합한 식물 프로모터는, 이로 제한되지는 않지만, 식물, 식물 바이러스 및 박테리아, 예컨대 식물 세포에서 발현되는 유전자를 포함하는 아그로박테리움 또는 리조비움으로부터 수득되는 것들이다.

[0366] 본 명세서에 사용된 바와 같은 용어 "균류(fungal) 세포"는 균류 계 내에서 임의의 유형의 진핵 세포를 지칭한다. 균류 계 내의 문(phyla)은 자낭균문, 담자균문, 블라스토클라디아균문(Blastocladiomycota), 호상균문, 글로메로균문(Glomeromycota), 미포자충문 및 네오칼리마스트릭스균문(Neocallimastigomycota)를 포함한다. 균류 세포는 효모, 곰팡이, 섬유성 균류이다. 일부 구현예에서, 균류 세포는 효모 세포이다.

[0367] 본 명세서에 사용된 바와 같은 용어 "효모 세포"는 자낭균류 및 담자균류 문 내의 임의의 균류 세포를 지칭한다. 효모 세포는 출아 효모 세포, 분열 효모 세포, 및 곰팡이 세포를 포함할 수 있다. 이들 생물에 제한되지 않으면서, 실험실 및 산업적 환경에서 사용되는 많은 유형의 효모는 자낭균문의 일부이다. 일부 구현예에서, 효모 세포는 사카로마이세스 세레비시에, 클루이베로마이세스 마르시아누스(Kluyveromyces marxianus), 또는 이삿첸키아 오리엔탈리스(Issatchenkia orientalis) 세포이다. 다른 효모 세포는 제한없이, 칸디다 종(예를 들어, 칸디다 알비칸스(Candida albicans)), 야로이아 종(예를 들어, 야로이아 리폴리티카(Yarrowia lipolytica)), 피키아 종(예를 들어, 피키아 파스토리스(Pichia pastoris)), 클루이베로마이세스 종(예를 들어, 클루이베로마이세스 락티스(Kluyveromyces lactis) 및 클루이베로마이세스 마르시아누스), 뉴로스포라 종 (예를 들어, 뉴로스포라 크라사(Neurospora crassa)), 푸사리움 종 (예를 들어, 푸사리움 옥시스포룸(Fusarium oxysporum)), 및 이삿첸키아 종 (예를 들어, 이삿첸키아 오리엔탈리스, a.k.a. 피키아 쿠드리아바제

빌(Pichia kudriavzevii) 및 칸디다 아시도써모필룸(Candida acidothermophilum))을 포함할 수 있다. 일부 구현예에서, 균류 세포는 섬유성 균류 세포이다. 본 명세서에 사용된 바와 같은 용어 "섬유성 균류 세포"는 섬유 상으로 자라는 임의의 유형의 균류 세포, 균사 또는 균사체를 지칭한다. 섬유성 균류 세포의 예로는 이로 제한되지 않지만, 누룩곰팡이 종 (Aspergillus spp.)(예를 들어, 아스페르길루스 나이거(Aspergillus niger)), 트리코더마 종(Trichoderma spp.)(예를 들어, 트리코더마 리세이(Trichoderma reesei)), 리조푸스 종(Rhizopus spp.)(예를 들어, 리조푸스 오리재(Rhizopus oryzae)), 및 모르티에렐라 종(Mortierella spp.)(예를 들어, 모르티에렐라 이사벨리나)가 포함될 수 있다.

[0368] 일부 구현예에서, 균류 세포는 산업용 균주이다. 본 명세서에서 사용된 바와 같은 "산업용 균주"는 산업적 공정, 예를 들어 상업적 또는 산업적 규모에서의 생성물의 생산으로부터 사용 또는 분리된 균주 세포의 임의의 균주를 지칭한다. 산업용 균주는 산업적 공정에서 통상적으로 사용되는 균류 종을 지칭할 수 있거나, 비-산업적 목적에도 사용될 수 있는(예를 들어, 실험실 연구) 균류 종의 분리물을 지칭할 수 있다. 산업적 공정의 예는 (예를 들어, 식품 또는 음료 제품의 생산에서) 발효, 증류, 생물연료 생산, 화합물의 생산 및 폴리펩티드의 생산을 포함할 수 있다. 산업용 균주의 예는 제한 없이, JAY270 및 ATCC4124를 포함할 수 있다.

[0369] 일부 구현예에서, 균류 세포는 배수체 세포이다. 본 명세서에서 사용된 바와 같은 "배수체" 세포는 그의 게놈이 하나 초과의 복사물로 존재하는 임의의 세포를 지칭할 수 있다. 배수체 세포는 자연에서 배수체 상태로 발견되는 세포의 유형을 지칭할 수 있거나, 배수체 상태로 존재하도록 유도된 세포를 지칭할 수 있다(예를 들어, 특이적 조절, 변경, 비활성화, 활성화, 또는 감수분열의 변형, 세포질 분열, 또는 DNA 복제를 통해). 배수체 세포는 그의 전체 게놈이 배수체인 세포를 지칭할 수 있거나, 특정한 관심 게놈 유전자좌 내 배수체인 세포를 지칭할 수 있다. 이론에 한정하고자 하지는 않지만, 가이드RNA의 존재비는 반수체 세포보다 배수체 세포의 게놈 조작에서 종종 더욱 속도-제한적인 성분일 수 있고, 따라서 본 명세서에 기술된 Cas13b CRISPR 시스템을 사용하는 방법은 일정 진균류 세포 유형을 사용하는 것을 활용할 수 있다.

[0370] 일부 구현예에서, 균류 세포는 이배체 세포이다. 본 명세서에서 사용된 바와 같은 "이배체" 세포는 그의 게놈이 두 개의 복사물로 존재하는 임의의 세포를 지칭할 수 있다. 이배체 세포는 자연에서 이배체 상태로 발견되는 세포의 유형을 지칭할 수 있거나, 이배체 상태로 존재하도록 유도된 세포를 지칭할 수 있다 (예를 들어, 특이적 조절, 변경, 비활성화, 활성화, 또는 감수분열의 변형, 세포질 분열, 또는 DNA 복제를 통해). 예를 들어, 사카로마이세스 세레비시에 균주 S228C는 반수체 또는 이배체 상태로 유지될 수 있다. 이배체 세포는 그의 전체 게놈이 이배체인 세포를 지칭할 수 있거나, 특정한 관심 게놈 유전자좌 내 이배체인 세포를 지칭할 수 있다. 일부 구현예에서, 균류 세포는 반수체 세포이다. 본 명세서에서 사용된 바와 같은 "반수체" 세포는 그의 게놈이 하나의 복사물로 존재하는 임의의 세포를 지칭할 수 있다. 반수체 세포는 자연에서 반수체 상태로 발견되는 세포의 유형을 지칭할 수 있거나, 반수체 상태로 존재하도록 유도된 세포를 지칭할 수 있다 (예를 들어, 특이적 조절, 변경, 비활성화, 활성화, 또는 감수분열의 변형, 세포질 분열, 또는 DNA 복제를 통해). 예를 들어, 사카로마이세스 세레비시에 균주 S228C는 반수체 또는 이배체 상태로 유지될 수 있다.

반수체 세포는 그의 전체 게놈이 반수체인 세포를 지칭할 수 있거나, 특정한 관심 게놈 유전자좌 내 반수체인 세포를 지칭할 수 있다.

[0371] 본 명세서에서 사용된 바와 같은 "효모 발현 벡터"는 RNA 및/또는 폴리펩티드를 인코딩하는 하나 이상의 서열을 함유하는 핵산을 지칭하며, 핵산(들)의 발현을 제어하는 임의의 바람직한 요소뿐만 아니라, 효모 세포 내부의 발현 벡터의 복제 및 유지를 가능하게 하는 임의의 요소를 추가로 함유할 수 있다. 많은 적합한 효모 발현 벡터 및 이의 특징은 당업계에 공지되어 있으며; 예를 들어, 다양한 벡터 및 기술은 하기 문헌에 설명된다: Yeast Protocols, 2nd edition, Xiao, W., ed. (Humana Press, New York, 2007) 및 Buckholz, R.G. and Gleeson, M.A. (1991) Biotechnology (NY) 9(11): 1067-72. 효모 벡터는 제한 없이, 동원체 (CEN) 서열, 자율 증식 서열 (ARS), 프로모터, 예컨대, 관심 서열 또는 유전자에 작동 가능하게 연결된, RNA 폴리머라제 III 프로모터, RNA 폴라머라제 III 종결자와 같은 종결자, 복제 기원, 및 마커 유전자 (예를 들어, 영양요구성, 항생성, 또는 다른 선택가능한 마커)를 함유할 수 있다. 효모에서의 이용을 위한 발현 벡터의 예로는, 플라스미드, 효모 인공 염색체, 2μ 플라스미드, 효모 편입형 플라스미드, 효모 자가증식형 플라스미드, 셔틀 벡터, 및 에피솜 유사 플라스미드를 포함할 수 있다.

[0372] **식물 및 식물 세포의 게놈으로 RNA 표적화 CRISP 시스템 성분의 안정한 통합**

[0373] 특정 실시형태에서, RNA 표적화 CRISPR 시스템의 성분을 인코딩하는 폴리뉴클레오티드가 식물 세포의 게놈 내로 안정한 통합을 위해 유도되는 것이 고려된다. 이들 실시형태에서, 형질전환 벡터 또는 발현 시스템의 디자

인은, 가이드 RNA 및/또는 RNA 표적화 유전자(들)가 언제, 어디에서 어떤 조건 하에서 발현되는지에 따라 조절될 수 있다.

[0374] 특정 실시형태에서, RNA 표적화 CRISPR 시스템의 성분을 식물 세포의 게놈 DNA 내로 도입하는 것이 고려된다. 추가적으로 또는 대안적으로, 안정한 통합을 위한 RNA 표적화 CRISPR 시스템의 성분을 식물 세포소기관, 예컨대 이로 제한되지는 않지만 색소체, 미토콘드리아 또는 엽록체의 DNA 내로 도입하는 것이 고려된다.

[0375] 식물 세포의 게놈으로의 안정적인 통합을 위한 발현 시스템은 하기 엘리먼트 중 하나 이상을 함유할 수 있다: 식물 세포에서 가이드 RNA 및/또는 RNA 표적화 효소를 발현시키기 위해 사용될 수 있는 프로모터 요소; 발현을 향상시키기 위한 5' 비번역 영역; 모노코트 세포와 같은 특정 세포에서의 발현을 추가로 향상시키는 인트론 엘리먼트; 하나 이상의 가이드 RNA 및/또는 RNA 표적화 유전자 서열 및 다른 원하는 엘리먼트를 삽입하기 위한 편리한 제한 부위를 제공하는 다중-클로닝 부위; 및 발현된 전사체의 효율적인 종결을 제공하기 위한 3'비번역 영역.

[0376] 발현 시스템의 요소는 플라스미드 또는 형질전환 벡터와 같은 환형, 또는 선형 이중 가닥 DNA와 같은 비-환형 중 어느 하나인 하나 이상의 발현 구축물 상에 있을 수 있다.

[0377] 특정 실시형태에서, RNA 표적화 CRISPR 발현 시스템은 적어도 하기를 포함한다:

[0378] 식물에서 표적 서열과 혼성화하는 가이드 RNA(gRNA)를 인코딩하는 뉴클레오티드 서열로, 여기서 가이드 RNA는 가이드 서열 및 직접 반복부 서열을 포함하는 서열, 및

[0379] RNA 표적화 단백질을 인코딩하는 뉴클레오티드 서열,

[0380] 여기서 성분 (a) 또는 성분 (b)는 동일하거나 상이한 구축물 상에 위치되며, 이로 인해 상이한 뉴클레오티드 서열은 식물 세포 내에서 동일하거나 상이한 조절 요소의 제어 하에 있을 수 있다.

[0381] RNA 표적화 CRISPR 시스템의 성분을 함유하는 DNA 구축물(들)는 다양한 통상의 기술에 의해 식물의 게놈, 식물 일부, 또는 식물 세포 내로 도입될 수 있다. 프로세스는 일반적으로 적합한 숙주 세포 또는 숙주 조직을 선택하는 단계, 구축물(들)을 숙주 세포 또는 숙주 조직 내로 도입하는 단계, 식물 세포 또는 그로부터의 식물을 재생하는 단계를 포함한다.

[0382] 특정 구현예에서, DNA 구축물은, 이로 제한되지는 않지만 예컨대 식물 세포 원형질체의 전기천공, 미세주입, 에어로졸 빔 주사 기술을 이용하여 식물 세포 내로 도입될 수 있거나, DNA 구축물은 DNA 유전자 총(particle bombardment)과 같은 바이오리스틱 방법을 이용하여 식물 조직에 직접 도입될 수 있다(문헌 [Fu et al., Transgenic Res. 2000 Feb;9(1):11-9] 참조). 유전자 총의 기초는 관심 유전자/들로 코팅된 입자를 세포를 향해 가속화하는 것으로, 이는 그 입자에 의한 원형질의 침투 및 통상적으로 게놈 내로의 안정한 통합을 결과로서 초래한다. (예를 들어, 문헌 [Klein et al, Nature (1987), Klein et ah, Bio/Technology (1992), Casas et ah, Proc. Natl. Acad. Sci. USA (1993).] 참조).

[0383] 특정 구현예에서, RNA 표적화 CRISPR 시스템의 성분을 함유하는 DNA 구축물은 아그로박테리움-매개된 형질전환에 의해 식물 내로 도입될 수 있다. DNA 구축물은 적합한 T-DNA 측접 영역과 함께 조합될 수 있으며, 통상의 아그로박테리움 투메파시엔스 (Agrobacterium tumefaciens) 숙주 벡터 내로 도입될 수 있다. 외래 DNA는 식물을 감염시킴으로써 또는 하나 이상의 Ti (종양-유도) 플라스미드를 함유하는, 아그로박테리움 박테리아와 함께 식물 원형질체를 항온반응시킴으로써 식물의 게놈 내로 통합될 수 있다 (예를 들어, 문헌 [Fraley et al., (1985), Rogers et al., (1987)] 및 미국 특허 제5,563,055호 참조).

[0384] **식물 프로모터**

[0385] 식물 세포 내 적절한 발현을 보장하기 위하여, 본 명세서에서 설명된 Cas13b CRISPR 시스템의 성분은 통상적으로 식물 프로모터, 즉 식물 세포 내 작동가능한 프로모터의 제어 하에 위치된다. 상이한 유형의 프로모터들의 이용이 고려된다.

[0386] 구조적 식물 프로모터는, 오픈 리딩 프레임(open reading frame)(ORF)를 발현할 수 있는 프로모터로, 이는 식물의 모든 또는 거의 모든 발생 단계 동안, 모든 또는 거의 모든 식물 조직을 제어한다 ("구조적 발현"으로도 지칭됨). 구조적 프로모터의 비-제한적인 일 예는 콜리플라워 모자이크 바이러스 35S 프로모터이다. 본 발명은 RNA 서열을 변형시키는 방법을 고려하며 그와 같이 또한 식물 생물분자의 발현을 조절하는 것을 고려한다. 따라서, 본 발명의 특정 실시형태는 조절될 수 있는 프로모터의 제어 하에 RNA 표적화 CRISPR 시스템의

하나 이상의 엘리먼트를 배치하는 것이 유리하다."조절된 프로모터"는 유전자 발현을 구성적으로가 아니라 일시적으로 및/또는 공간적으로-조절되는 방식으로 유도하고 조직-특이적, 조직-선호적 및 유도성 프로모터를 포함하는 프로모터를 지칭한다. 상이한 프로모터는 상이한 조직 또는 세포 유형에서, 또는 상이한 발생 단계에서, 또는 상이한 환경 조건에 반응하여 유전자의 발현을 유도할 수 있다. 특정 구현예에서, 하나 이상의 RNA 표적화 CRISPR 성분은 구조적 프로모터, 예컨대 콜리플라워 모자이크 바이러스 35S 프로모터의 제어 하에 발현된다. 조직-선호된 프로모터는 특정 식물 조직 내에서 소정의 세포 유형 내, 예를 들어, 잎 또는 뿌리 내의 맥관 세포 또는 종자의 특정 세포 내 증진된 발현을 표적화하는데 사용될 수 있다. RNA 표적화 CRISPR 시스템에 사용하기 위한 특정 프로모터의 예는 하기 문헌에서 발견된다: Kawamata et al., (1997) Plant Cell Physiol 38:792-803; Yamamoto et al., (1997) Plant J 12:255-65; Hire et al, (1992) Plant Mol Biol 20:207-18,Kuster et al, (1995) Plant Mol Biol 29:759-72, 및 Capana et al., (1994) Plant Mol Biol 25:681 -91.

[0387] 유도성이며 유전자 편집 또는 유전자 발현의 시공간적 제어를 가능하게 하는 프로모터의 예는 에너지의 형태를 이용할 수 있다. 에너지의 형태는 이로 제한되지는 않지만, 소리 에너지, 전자기 방사, 화학 에너지 및/또는 열 에너지를 포함할 수 있다. 유도성 시스템의 예로는 테트라사이클린 유도성 프로모터(Tet-온 또는 Tet-오프), 소분자 2-혼성체 전사 활성화 시스템(FKBP, ABA, 등), 또는 광 유도성 시스템(피토크롬, LOV 도메인, 또는 크립토크롬), 예컨대 서열-특이적 방식으로 전사 활성화에서의 변화를 유도하는 광 유도성 전사 이펙터(LITE)가 포함된다. 광 유도성 시스템의 성분은 RNA 표적화 Cas13b, 광-반응성 시토크롬 이종이량체(예를 들어, 아라비돕시스 탈리아나 (Arabidopsis thaliana) 유래), 및 전사 활성화/억제 도메인을 포함할 수 있다. 유도성 DNA 결합 단백질 및 그의 사용 방법의 추가적인 예는 US 61/736465 및 US 61/721,283에 제공되며, 이들은 그 전체로 본 명세서에 참고문헌으로 통합된다.

[0388] 특정 구현예에서, 일시적 또는 유도성 발현은 예를 들어, 화학-조절된 프로모터를 이용함으로써 달성될 수 있으며, 즉 이에 인해 외인성 화학물질의 적용이 유전자 발현을 유도한다. 유전자 발현의 조정은 화학물질-억제성 프로모터에 의해 수득될 수도 있으며, 여기서 화학물질의 적용은 유전자 발현을 억제한다. 화학-유도성 프로모터는, 이로 제한되지는 않지만, 벤젠 술폰아미드 제초제 약해경감제에 의해 활성화되는, 메이즈 In2-2 프로모터 [문헌 (De Veylder et al., (1997) Plant Cell Physiol 38:568-77)], 발아전 처리(pre-emergent) 제초제로서 사용된 소수성 친전자성 화합물에 의해 활성화된, 메이즈 GST 프로모터 (GST-11-27, WO93/01294), 및 살리실산에 의해 활성화된 담배 PR-1 프로모터 [문헌 (Ono et al., (2004) Biosci Biotechnol Biochem 68:803-7)]를 포함한다. 항생체, 예컨대 테트라사이클린-유도성 및 테트라사이클린-억제성 프로모터에 의해 조절되는 프로모터(Gatz et al., (1991) Mol Gen Genet 227:229-37; 미국 특허 5,814,618 및 5,789,156)가 또한 본원에 사용될 수 있다.

[0389] **<u>특정 식물 세포소기관으로의 전위 및/또는 거기에서의 발현</u>**

[0390] 발현 시스템은 특정 식물 세포소기관으로의 전위를 위한 및/또는 식물 세포소기관에서의 발현을 위한 요소를 포함할 수 있다.

[0391] *엽록체 표적화*

[0392] 특정 실시형태에서, RNA 표적화 CRISPR 시스템은 엽록체 유전의 발현 및/또는 번역을 특이적으로 변형시키거나, 또는 엽록체에서의 발현을 보장하는데 사용되는 것을 고려한다. 이러한 목적을 위해, 엽록체 형질전환 방법 또는 RNA 표적화 CRISPR 시스템 성분의 엽록체로의 구획화를 이용한다. 예를 들어, 색소체 게놈 내 유전적 변형의 도입은 화분을 통한 유전자 흐름과 같은 생물안전성 문제를 감소시킬 수 있다.

[0393] 엽록체 형질전환 방법은 해당 기술 분야에 알려져 있으며, 유전자 총법, PEG 처리, 및 미세주입을 포함한다. 추가적으로, 형질전환 카세트를 핵 게놈으로부터 색소체로의 전위를 수반하는 방법이 WO2010061186에 설명된 바와 같이 사용될 수 있다.

[0394] 대안적으로, 하나 이상의 RNA 표적화 CRISPR 시스템 성분을 식물 엽록체로 표적화하는 것이 고려된다. 이는 발현 구축물 내로, RNA 표적화 단백질을 인코딩하는 서열의 5' 영역에 작동 가능하게 연결된, 엽록체 이동 펩티드(chloroplast transit peptide) (CTP) 또는 색소체 이동 펩티드를 인코딩하는 서열을 통합시킴으로써 달성된다. CTP는 엽록체 내로의 전위 동안 가공 단계에서 제거된다. 발현된 단백질의 엽록체 표적화는 숙련자에게 공지이다(예를 들어, 문헌 [Protein Transport into Chaloroplasts, 2010, Annual Review of Plant Biology,Vol. 61: 157-180). 그러한 실시형태에서, 하나 이상의 가이드 RNA를 식물 엽록체로 표적화하는 것

이 또한 바람직하다. 엽록체 국재화 서열을 이용하여 가이드 RNA를 엽록체 내로 전위하는데 사용될 수 있는 방법 및 구축물이 예를 들어 20040142476에 설명되어 있으며, 이는 본 명세서에 참고문헌으로 통합된다. 구축물의 그러한 변이는 본 발명의 발현 시스템 내로 통합되어 RNA 표적화-가이드 RNA(들)을 효율적으로 전위시킬 수 있다.

[0395] **조류 세포에 CRISPR-RNA 표적화 시스템을 인코딩하는 폴리뉴클레오티드의 도입**

[0396] 유전자이식 조류(또는 유채와 같은 다른 식물)는 채소유 또는 생물연료 예컨대 알코올(특히 메탄올 및 에탄올) 또는 다른 생성물의 생산에서 특히 유용할 수 있다. 이들은 오일 또는 생체연료 산업에서 사용을 위한 고수준의 오일 또는 알코올을 발현 또는 과발현시키도록 유전자조작될 수 있다.

[0397] US 8945839호는 Cas9를 이용하여 미세조류(클라미도모나스 레인하티(Chlamydomonas reinhardtii) 세포) 종을 조작하는 방법을 설명한다. 유사한 도구를 사용하여, 본 명세서에서 기재된 RNA 표적화 CRISPR 시스템의 방법은 클라미도모나스 종 및 기타 다른 조류 상에 적용될 수 있다. 특정 실시형태에서, RNA 표적화 및 가이드 RNA(들)는, Hsp70A-Rbc S2 또는 Beta2 -튜불린과 같은 항상성 프로모터의 제어 하에서 RNA 표적화 단백질을 발현하는 벡터를 이용하여 발현된 조류 내에 도입된다. 가이드 RNA는 T7 프로모터를 함유하는 벡터를 사용하여 선택적으로 전달된다. 대안적으로, RNA 표적화 mRNA 및 시험관 내 전사된 가이드 RNA는 조류 세포로 전달될 수 있다. 전기천공 프로토콜은 진아트 클라미도모나스 조작 키트(GeneArt Chlamydomonas Engineering kit)로부터 권장된 표준 프로토콜과 같이, 숙련자에게 이용가능하다.

[0398] **효모 세포로 RNA 표적화 성분을 인코딩하는 폴리뉴클레오티드의 도입**

[0399] 특정 실시형태에서, 본 발명은 효모 세포에서 RNA 편집을 위한 RNA 표적화 CRISPR 시스템의 용도에 관한 것이다. RNA 표적화 CRISPR 시스템 성분을 인코딩하는 폴리뉴클레오티드를 도입하는데 사용될 수 있는 효모 세포를 형질전환하는 방법은 당업자에게 공지이며, 하기 문헌에서 검토된다 (Kawai et al., 2010, Bioeng Bugs. 2010 Nov-Dec; 1(6): 395-403). 비-제한적인 예로는 리튬 아세테이트 처리에 의해 (이는 캐리어(carrier) DNA 및 PEG 처리를 추가로 포함할 수 있음), 총 또는 전기천공에 의한 효모 세포의 형질전환을 포함한다.

[0400] **식물 및 식물 세포에서 RNA 표적화 CRISP 시스템 성분의 일시적 발현**

[0401] 특정 실시형태에서, 가이드 RNA 및/또는 RNA 표적화 유전자는 식물 세포 내에서 일시적으로 발현된다고 생각된다. 이들 실시형태에서, RNA 표적화 CRISPR 시스템은, 가이드 RNA 및 RNA 표적화 단백질 모두가 세포 내에 존재하는 때에만 RNA 표적 분자의 변형을 보장할 수 있어서, 게놈 변형은 추가로 제어될 수 있다. RNA 표적화 효소의 발현이 일시적임에 따라, 그러한 식물 세포로부터 생성된 식물은 통상적으로 외래 DNA를 함유하지 않는다. 특정 실시형태에서 RNA 표적화 효소는 식물 세포에 의해 안정하게 발현되고, 가이드 서열이 일시적으로 발현된다.

[0402] 특정 실시형태에서, RNA 표적화 CRISPR 시스템 성분은 식물 바이러스 벡터를 이용하여 식물 세포 내로 도입될 수 있다(문헌 [Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299-323] 참조). 추가의 특정 구현예에서, 상기 바이러스 벡터는 DNA 바이러스로부터의 벡터이다. 예를 들어, 제미니바이러스(geminivirus)(예를 들어, 양배추 잎말림 바이러스, 콩 황색 왜소 바이러스, 밀 왜소 바이러스, 토마토 잎말림 바이러스, 메이즈 줄무늬병 바이러스, 담배 잎말림 바이러스, 또는 토마토 골든 모자이크 바이러스) 또는 나노바이러스(예를 들어, 누에콩 괴사성 황색 바이러스). 다른 특정 구현예에서, 상기 바이러스 벡터는 RNA 바이러스로부터의 벡터이다. 예를 들어, 토브라바이러스(tobravirus)(예를 들어, 담배 얼룩 바이러스, 담배 모자이크 바이러스), 포텍스바이러스(potexvirus)(예를 들어, 감자 바이러스 X), 또는 호르데이(hordeivirus) (예를 들어, 보리 반엽 모자이크 바이러스). 식물 바이러스의 복제 게놈은 비-통합 벡터가 GMO 식물의 생산을 피하려는 경우에 관심이 있는 것이다.

[0403] 특정 실시형태에서, RNA 표적화 CRISPR 구축물의 일시 발현에 사용된 벡터는 예를 들어, pEAQ 벡터이고, 이는 원형질체 내에서 아그로박테리움-매개 일시 발현에 대해 조절된다(Sainsbury F. et al., Plant Biotechnol J. 2009 Sep;7(7):682-93). 게놈 위치의 정확한 표적화는, 변형된 양배추 잎말림 바이러스(CaLCuV) 벡터를 이용하여, Cas13b를 발현하는 안정한 유전자이식 식물에서 gRNA를 발현하는 것을 증명하였다(문헌[Scientific Reports 5, Article number:14926 (2015), doi:10.1038/srep14926)].

[0404] 특정 실시형태에서, 가이드 RNA 또는 crRNA 및/또는 RNA 표적화 유전자를 인코딩하는 이중-가닥 DNA 단편은 식물 세포 내로 일시적으로 도입될 수 있다. 그러한 실시형태에서, 도입된 이중-가닥 DNA 단편은 세포를 변형하기에는 충분하지만, 고려된 기간이 지나거나 한번 이상의 세포 분열 후에는 지속되지 않는 양으로 제공된다.

식물에서 직접 DNA 전달 방법은 당업자에게 알려져 있다(예를 들어, 문헌[Davey et al. Plant Mol Biol. 1989 Sep;13(3):273-85.] 참조).

[0405] 다른 실시형태에서, RNA 표적화 단백질을 인코딩하는 RNA 폴리뉴클레오티드가 식물 세포 내로 도입되고, 이는 이후 번역되고 숙주 세포에 의해 가공되어(하나 이상의 가이드 RNA 존재 하에서) 세포를 변형시키기에는 충분한 양이지만 고려된 기간이 지나거나 한번 이상의 세포 분열 후에는 지속되지 않는 양의 단백질을 생성한다. 일시 발현을 위하여 mRNA를 식물 원형질체로 도입하는 방법은 숙련자에게 알려져 있다(예를 들어, 문헌 [Gallie, Plant Cell Reports (1993), 13;119-122] 참조). 상기 설명된 상이한 방법들의 조합이 또한 고려된다.

[0406] **식물 세포로 RNA 표적화 CRISPR 성분의 전달**

[0407] 특정 실시형태에서, RNA 표적화 CRISPR 시스템의 하나 이상의 성분을 식물 세포로 직접 전달하는 것에 관심을 갖는다. 이는 그중에서도 비-유전자이식 식물의 생성에 유익하다. 특정 실시형태에서, 하나 이상의 RNA 표적화 성분이 식물 또는 식물 세포 외부에서 제조되고 세포로 전달된다. 예를 들어 특정 실시형태에서, RNA 표적화 단백질은 식물 세포로의 도입 전에 시험관 내에서 제조된다. RNA 표적화 단백질은 당업자에게 알려진 각종 방법에 의해 제조될 수 있으며, 재조합 생산을 포함한다. 발현 후, RNA 표적화 단백질이 단리되고, 필요에 따라 리폴딩되고, 정제 및 임의로 처리되어 임의의 정제 태그, 예컨대 His-태그를 제거할수 있다.

조생의, 부분적으로 정제된, 또는 더욱 완전하게 정제된 RNA 표적화 단백질이 일단 수득되면, 단백질은 식물 세포로 도입될 수 있다.

[0408] 특정 실시형태에서, RNA 표적화 단백질은 관심 RNA를 표적화하는 가이드 RNA와 혼합되어 사전-조립된 리보뉴클레오단백질을 형성한다.

[0409] 개별적인 성분 또는 사전-조립된 리보뉴클레오단백질은 전기 천공을 통해, RNA 표적화-연관된 유전자 산물 코팅된 입자를 이용한 충격에 의해, 화학적 트랜스펙션에 의해, 또는 세포막을 지나 운송하기 위한 일부 다른 수단에 의해 식물 세포 내로 도입될 수 있다. 예를 들어, 사전-조립된 CRISPR 리보뉴클레오단백질을 이용한 식물 원형질체의 트랜스펙션은, 식물 게놈이 표적된 변형을 보장하는 것으로 증명되었다(이는 하기 문헌에 의해 기술됨 [Woo et al. *Nature Biotechnology*, 2015; DOI: 10.1038/nbt.3389). 이들 방법은 식물에서 RNA 분자의 표적화된 변형을 획득하도록 변형될 수 있다.

[0410] 특정 실시형태에서, RNA 표적화 CRISPR 시스템 성분은 나노입자를 이용하여 식물 세포 내로 도입된다. 단백질 또는 핵산으로서, 또는 이의 조합 중 어느 하나인 성분은 나노입자 위에 업로드(upload)되거나, 나노입자 내에 패키지되고, 식물에 적용될 수 있다(예를 들어, WO 2008042156 및 US 20130185823에 설명된 것과 같음).

특히, 본 발명의 실시형태는, WO2015089419에 설명된 바와 같이, RNA 표적화 단백질을 인코딩하는 DNA 분자(들), 가이드 RNA 및/또는 단리된 가이드 RNA를 인코딩하는 DNA 분자가 업로드되거나 또는 그로 패킹된 나노입자를 포함한다.

[0411] RNA 표적화 CRISPR 시스템의 하나 이상의 성분을 식물 세포 내로 도입하는 추가의 수단은 세포 침투 펩티드(CPP)의 이용에 의한 것이다. 따라서, 특히 본 발명에서 실시형태는 RNA 표적화 단백질에 연결된 세포 침투 펩티드를 포함하는 조성물을 포함한다. 본 발명의 특정 실시형태에서, RNA 표적화 단백질 및/또는 가이드 RNA는 하나 이상의 CPP에 커플링되어 이들을 식물 원형질체 내부로 효과적으로 운송한다; 또한 문헌 [Ramakrishna (2014, Genome Res. 2014 Jun;24(6):1020-7 for Cas9 in human cells)]을 참조한다. 다른 실시형태에서, RNA 표적화 유전자 및/또는 가이드 RNA는 식물 원형질체 전달을 위한 하나 이상의 CPP에 커플링되는 하나 이상의 환형 또는 비-환형 DNA 분자(들)에 의해 인코딩된다. 식물 원형질체는 이후 식물 세포로, 그리고 추가로 식물에 재생된다. CPP는, 단백질로부터 유도된 또는 수용체 독립적인 방식으로 세포막을 지나 생물분자를 운송할 수 있는 키메라 서열로부터 유도된 어느 하나인, 일반적으로 35 개 미만의 아미노산의 짧은 펩티드로서 설명된다. CPP는 양이온 펩티드, 소수성 서열을 갖는 펩티드, 양친매성 펩티드, 프롤린-풍부 및 항균성 서열을 갖는 펩티드, 및 키메라 또는 이분(bipartite) 펩티드(문헌 [Pooga and Langel 2005])일 수 있다. CPP는 생물학적 막을 침투할 수 있고, 그와 같이 각종 생물 분자의 세포막을 거친 세포질 내로의 이동을 촉발하고, 그의 세포내 여정을 개선시키고, 이에 따라 생물 분자의 표적과의 상호작용을 촉진한다. CPP의 예는 그중에서도 다음을 포함한다: HIV 타입 1 에 의한 바이러스 복제에 필요한 핵 전사 활성화제 단백질인 Tat, 페네트라틴, 카포시 섬유아세포 성장 인자 (FGF) 신호 펩티드 서열, 인테그린 β3 신호 펩티드 서열; 폴리아르기닌 펩티드 Args 서열, 구아닌 풍부-분자 수송체, 스위트 화살표 펩티드 등.

[0412] **식물, 조류 또는 진균 응용에 고려되는 표적 RNA**

[0413] 표적 RNA, 즉 관심 RNA는 본 발명에 의해 표적화되어, 표적 RNA 상의 관심 표적 부위에서, 동원 및 RNA 표적화 단백질의 결합을 일으키는 RNA이다. 표적 RNA는 RNA의 임의의 적합한 형태일 수 있다. 이는 일부 실시형태에서, mRNA를 포함한다. 다른 실시형태에서, 표적 RNA는 전달 RNA (tRNA) 또는 리보좀 RNA (rRNA)를 포함할 수 있다. 다른 실시형태에서 표적 RNA는 간섭 RNA (RNAi), 마이크로RNA (miRNA), 마이크로스위치, 마이크로자임, 위성 RNA 및 RNA 바이러스를 포함할 수 있다. 표적 RNA는 식물 세포의 세포질에, 세포 핵에 또는 식물 세포 세포소기관, 예컨대 미토콘드리아, 엽록체 또는 색소체에 위치할 수 있다.

[0414] 특정 실시형태에서, RNA 표적화 CRISPR 시스템은 RNA를 절단하거나 또는 아니면 RNA 발현을 억제하는데 사용된다.

[0415] **RNA 조정을 통한 식물 유전자 발현을 위한 RNA 표적화 CRISPR 시스템의 용도**

[0416] RNA 표적화 단백질은 또한 RNA 프로세싱의 제어를 통해 유전자 발현을 표적화기 위해, 적합한 가이드 RNA와 함께 사용될 수 있다. RNA 프로세싱의 제어는 대안적인 스플라이싱을 포함하는 RNA 스플라이싱과 같은 RNA 프로세싱 반응; 바이러스 복제 (특히 식물의 비리오이드 및 tRNA 생합성을 포함한 식물 바이러스의 경우)를 포함할 수 있다. 적합한 가이드 RNA와 조합하여 RNA 표적화 단백질은 또한 RNA 활성화 (RNAa)를 제어하는 데 사용될 수 있다. RNAa는 유전자 발현의 촉진을 야기시키므로, 유전자 발현의 제어는 RNAa의 파괴 또는 감소를 통해서 그에 따라 유전자 발현의 촉진을 덜하는 방식으로 달성될 수 있다.

[0417] 본 발명의 RNA 표적화 이펙터 단백질은 특히 RNA 바이러스에 대항하여, 식물에서 항바이러스 활성을 위해 더욱 사용될 수 있다. 이펙터 단백질은 선택된 바이러스 RNA 서열에 대해 선택적인 적합한 가이드 RNA를 사용하여 바이러스 RNA에 대해 표적화할 수 있다. 특히, 이펙터 단백질은 RNA, 예컨대 단일 가닥 RNA를 절단하는 활성 뉴클레아제일 수 있다. 따라서, 항바이러스제로서 본 발명의 RNA 표적화 이펙터 단백질의 용도가 제공된다. 이러한 방식으로 대응할 수 있는 바이러스의 예는 제한없이, 담배 모자이크 바이러스(Tobacco mosaic virus) (TMV), 토마토 반점 위조 바이러스(Tomato spotted wilt virus) (TSWV), 오이 모자이크 바이러스(Cucumber mosaic virus) (CMV), 감자 바이러스(Potato virus Y) (PVY), 콜리플라워 모자이크 바이러스(Cauliflower mosaic virus) (CaMV) (RT 바이러스), 자두 곰보 바이러스(Plum pox virus) (PPV), 브롬 모자이크 바이러스(Brome mosaic virus) (BMV) 및 감자 바이러스(Potato virus X) (PVX)를 포함한다.

[0418] 표적화 유전자 변형의 대안으로서, 식물, 조류 또는 진균에서 RNA 발현을 조정하는 예는 본명세서에서 더욱 기술한다.

[0419] 특히 관심은 mRNA의 조절된 절단을 통한 유전자 발현의 조절된 제어이다. 이는 본 명세서에 기술된 바와 같은 조절된 프로모터의 제어 하에 RNA 표적화의 엘리먼트를 위치시켜 달성할 수 있다.

[0420] **tRNA 분자의 기능성을 복원하기 위한 RNA 표적화 CRISPR 시스템의 용도**

[0421] Pring 등은 DNA와 상이한 단백질을 인코딩하도록 mRNA 서열을 변경시키는 식물 미토콘드리아 및 엽록체에서 RNA 편집을 기술한다. (Plant Mol. Biol. (1993) 21 (6): 1163-1170. doi:10.1007/BF00023611). 본 발명의 특정 실시형태에서, 미토콘드리아 및 엽록체 mRNA를 특이적으로 표적화하는 RNA 표적화 CRISPR 시스템의 엘리먼트는 식물 또는 식물 세포에 도입되어 생체내에서 발생되는 과정을 모방하는 이러한 식물 세포 소기관에서 상이한 단백질을 발현시킬 수 있다.

[0422] **RNA 발현을 억제하기 위한 RNA 간섭의 대안으로서 RNA 표적화 CRISPR 시스템의 용도**

[0423] RNA 표적화 CRISPR 시스템은 RNA 억제 또는 RNA 간섭과 유사한 용도를 갖고, 따라서 또한 그러한 방법을 대체할 수 있다. 특정 실시형태에서, 본 발명의 방법은 예를 들어 간섭 리보핵산 (예컨대, siRNA 또는 shRNA 또는 dsRNA)의 대체물로서 RNA 표적화 CRISPR의 용도를 포함한다. 표적화 유전자 변형의 대안으로서 식물, 조류 또는 진균에서 RNA 발현의 억제의 예는 본 명세서에서 더욱 설명한다.

[0424] **RNA 간섭을 제어하기 위한 RNA 표적화 CRISPR 시스템의 용도**

[0425] 간섭 RNA 또는 miRNA에 대한 제어는 생체내 또는 시험관내 간섭 RNA 또는 miRNA의 수명을 감소시켜 그들 접근법에서 확인되는 표적외 효과 (OTE)를 줄이는데 도움을 줄 수 있다. 특정 실시형태에서, 표적 RNA는 간섭 RNA, 즉 간섭 경로에 관여하는 RNA, 예컨대 shRNA, siRNA 등을 포함할 수 있다. 다른 실시형태에서, 표적

RNA는 마이크로RNA (miRNA) 또는 이중 가닥 RNA (dsRNA)를 포함할 수 있다.

[0426] 다른 특정 실시형태에서, RNA 표적화 단백질 및 적합한 가이드 RNA(들)가 선택적으로 발현되면 (예를 들어, 조절된 프로모터, 예를 들어 조직 특이적 또는 세포 주기 특이적 프로모터 및/또는 인핸서의 제어 하에서 공간적으로 또는 시간적으로), 이것은 그들 세포에서 RNAi로부터 세포 또는 시스템을 (생체내 또는 시험관내에서) '보호'하는데 사용될 수 있다. 이거은 RNAi가 이펙터 단백질 및 적합한 가이드가 발현되지 않는 세포 또는 조직의 비교를 위해서 또는 RNAi가 요구되지 않는 이웃한 조직 또는 세포에서 유용할 수 있다 (즉, 개별적으로 RNAi가 제어되지 않는 곳 및 제어되는 곳). RNA 표적화 단백질은 RNA, 예컨대 리보자임, 리보좀 또는 리보스위치를 포함하거나 또는 그로 이루어지는 분자를 제어하거나 또는 그에 결합하기 위해 사용될 수 있다. 본 발명의 실시형태에서, 가이드 RNA는 RNA 표적화 단백질이 그들에 결합할 수 있도록 이들 분자로 RNA 표적화 단백질을 동원할 수 있다.

[0427] 본 출원은 시스템의 알려진 정보에 대한 기초를 제공하기 때문에, 본 발명의 RNA 표적화 CRISPR 시스템은 해충 관리, 식물 질환 관리 및 제초제 내성의 관리를 비롯한 식물 어세이 및 다른 응용분야를 포함하여, 본 개시로부터, 과도한 실험없이, 식물체내 RNAi 영역에서 적용될 수 있다 (예를 들어, Kim et al., in Pesticide Biochemistry and Physiology (Impact Factor: 2.01). 01/2015; 120. DOI: 10.1016/j.pestbp.2015.01.002; Sharma et al. in Academic Journals (2015), Vol.12(18) pp2303-2312); Green J.M, inPest Management Science, Vol 70(9), pp 1351-1357) 참고).

[0428] **식물, 조류 및 진균에서 리보스위치를 변형시키고 대사 조절을 제어하기 위한 RNA 표적화 CRISPR 시스템의 용도**

[0429] 리보스위치 (압토자임이라고도 알려짐)은 소형 분자에 결합하고 그 다음으로 유전자 발현을 조절하는 메신저 RNA의 조절 절편이다. 이러한 기전은 이들 소형 분자의 세포내 농도를 세포가 감지할 수 있게 한다. 특정 리보시위치는 전형적으로 이러한 유전자의 전사, 번역 또는 스플라이싱을 변경시켜 이의 인접한 유전자를 조절한다. 따라서, 본 발명의 특정 실시형태에서, 리보스위치 활성의 제어는 리보스위치를 표적으로 하는 적합한 가이드 RNA와 조합하여 RNA 표적화 단백질의 사용을 통해 고려된다. 이는 리보스위치의 절단, 또는 리보스위치와의 결합을 통해서일 수 있다. 특정 실시형태에서, 리보스위치 활성의 감소가 고려된다. 최근에, 티아민 파이로포스페이트 (TPP)에 결합하는 리보스위치가 특징규명되었고 식물 및 조류에서 티아민 생합성을 조절하는 것으로 확인되었다. 더 나아가서 이 엘리먼트는 식물에서 주요 물질대사의 필수 조절인자인 것으로 나타났다 (Bocobza and Aharoni, Plant J. 2014 Aug; 79(4):693-703. doi: 10.1111/tpj.12540. Epub 2014 Jun 17). TPP 리보스위치는 또 한 일정 진균, 예컨대 뉴로스포라 크라사 (Neurospora crassa)에서 발견되었는데, 여기서 상류 오픈 리딩 프레임 (uORF)를 조건적으로 생성시키도록 선택적 스플라이싱을 제어하여서, 하류 유전자의 발현에 영향을 미친다 (Cheah MT et al., (2007)Nature 447 (7143):497-500. doi:10.1038/nature05769) 본 명세서에 기술된 RNA 표적화 CRISPR 시스템은 식물, 조류 또는 진균에서 내인성 리보스위치 활성을 조종하는데 사용될 수 있고 그리하여 이를 통해 제어되는 하류 유전자의 발현을 변경시킬 수 있다. 특정 실시형태에서, RNA 표적화 CRISP 시스템은 물질대사 네트워크와 이의 관련성을 연구하고 생체내 또는 시험관내 리보스위치 기능을 어세이하는데 사용될 수 있다. 특정 실시형태에서 RNA 표적화 CRISPR 시스템은 유전자 제어를 위한 플랫폼 및 식물에서 대사산물 센서로서 리보스위치의 조작에 잠재적으로 사용될 수 있다.

[0430] **식물, 조류 또는 진균에 대한 RNAi 스크린에서 RNA 표적화 CRISPR 시스템의 용도**

[0431] 그의 녹다운이 표현형 변화와 연관된 유전자 산물의 동정은, RNAi 스크린을 통해서 생물학적 경로에 대한 정보를 얻고 성분 부분을 동정한다. 본 발명의 특정 실시형태에서, 제어는 또한 이들 스크린 상에서 또는 이들 스크린 동안 본 명세서에 기술된 가이드 29 또는 가이드 30 단백질 및 가이드 RNA의 사용을 통해 스크린에서 RNAi의 활성을 제거하거나 또는 감소시키고 그에 따라 (이전에 간섭된) 유전자 산물의 활성을 (간섭/억제를 제거 또는 감소시켜) 회복시키도록 발휘될 수도 있다.

[0432] **생체내 및 시험관내 RNA 분자의 가시화를 위한 RNA 표적화 단백질의 용도**

[0433] 특정 실시형태에서, 본 발명은 핵산 결합 시스템을 제공한다. 상보성 프로브와 RNA의 인시츄 혼성화는 강력한 기술이다. 전형적으로 형광발광성 DNA 올리고뉴클레오티드를 사용하여 혼성화에 의한 핵산을 검출한다. 증가된 효율성은 일정 변형, 예컨대 잠김 핵산 (locked nucleic acid) (LNA)을 통해 얻어졌지만, 효율적이고 다재다능한 대안에 대한 요구가 여전히 존재한다. 이와 같이, RNA 표적화 시스템의 표지된 엘리먼트가 인시츄 혼성화를 위한 효율적이고 적합화가능한 시스템에 대한 대안으로서 사용될 수 있다.

[0434] **식물 및 효모에서 RNA 표적화 CRISPR 시스템의 추가적 응용**

[0435] *생물연료 생산에서 RNA 표적화 CRISPR 시스템의 이용*

[0436] 본 명세서에서 설명된 것과 같은 용어 "생물연료(biofuel)"는 식물 및 식물-유도된 자원으로부터 제조된 대안적인 연료이다. 재생성 생물연료는, 그 에너지가 탄소 고정 과정을 통해 수득되거나 바이오매스의 이용 또는 전환을 통하여 만들어진 유기 물질로부터 추출될 수 있다. 이러한 바이오매스는 생물연료에 대해 직접 사용될 수 있거나, 열 전환, 화학 전환 및 생화학 전환에 의한 성분을 함유하는 편리한 에너지로 전환될 수 있다. 이러한 바이오매스는 고체, 액체, 또는 가스 형태의 연료를 결과로서 초래할 수 있다. 생물연료에는 두 가지 유형:바이오에탄올 및 바이오디젤이 존자한다. 바이오에탄올은 셀룰로오스의 당 발효 공정에 의해 주로 생산되며 (전분), 이는 메이즈 및 사탕수수로부터 대부분 유도된다. 한편, 바이오디젤은 유채씨, 야자 및 대두와 같은 유료 작물로부터 주로 생산된다. 생물연료는 주로 수송에 사용된다.

[0437] *생물연료 생산을 위한 식물 특성의 증진*

[0438] 특정 실시형태에서, 본 명세서에서 설명된 바와 같은 RNA 표적화 CRISPR 시스템을 이용하는 본 방법은, 발효를 위한 당의 더욱 효율적인 방출을 위해 주요 가수분해제에 의한 접근을 촉진하기 위해, 세포벽의 특성을 변경하는데 사용된다. 특정 구현예에서, 셀룰로오스 및/또는 리그닌의 생합성이 변형된다. 셀룰로오스는 세포벽의 주 성분이다. 셀룰로오스 및 리그닌의 생합성은 공동조절된다. 식물에서 리그닌의 비를 감소시킴으로써 셀룰로오스의 비가 증가될 수 있다. 특정 구현예에서, 본 명세서에서 설명된 방법은, 발효가능한 탄수화물을 증가시키기 위하여 식물에서 리그닌 생합성을 하향조절하는데 사용된다. 더욱 특히, 본 명세서에서 설명된 방법은, WO 2008064289 A2에 개시된 바와 같은, 4-쿠마레이트 3-수산화효소(C3H), 페닐알라닌 암모니아-리아제(PAL), 신나메이트 4-수산화효소(C4H), 하이드록시신나모일 트랜스퍼라제(HCT), 카페인산 O-메틸트랜스퍼라제(COMT), 카페오일 CoA 3-O-메틸트랜스퍼라제(CCoAOMT), 페룰레이트 5-수산화효소(F5H), 신나밀 알코올 탈수소효소(CAD), 신나모일 CoA-환원효소(CCR), 4-쿠마레이트-CoA 리가제(4CL), 모노리그놀-리그닌-특이적 글리코실트랜스퍼라제, 및 알데히드 탈수소효소(ALDH)로 이루어진 군으로부터 선택된 적어도 제1 리그닌 생합성 유전자를 하향조절하는데 사용된다.

[0439] 특정 구현예에서, 본 명세서에서 설명된 방법은 발효 동안 더욱 낮은 수준의 아세트산을 생산하는 식물 무리(mass)를 생산하는데 사용된다 (WO 제200096488호 또한 참조).

[0440] *생물연료 생산을 위한 효모의 변형*

[0441] 특정 실시형태에서, 본 명세서에 제공되는 RNA 표적화 효소는 재조합 미생물에 의한 생물에탄올 생산에 사용된다. 예를 들어, RNA 표적화 효소는 발효가능한 당류로부터 생물연료 또는 생물중합체를 발생시키도록, 그리고 임의로 발효가능한 당류의 공급원으로서 농업 폐기물에서 유래되는 식물-유래 리그노셀룰로스를 분해할 수 있도록 미생물, 예컨대 효모를 조작하는데 사용될 수 있다. 보다 특히, 본 발명은 RNA 표적화 CRISPR 복합체가 생물연료 생산에 요구되는 내인성 유전자의 발현을 변형시키고/시키거나 생물연료 합성을 간섭할 수 있는 내인성 유전자를 변형시키는데 사용되는 방법을 제공한다. 보다 특히, 방법은 피루베이트를 에탄올 또는 다른 관심 생성물로 전환시키는데 관여하는 효소를 인코딩하는 하나 이상의 뉴클레오티드 서열을 미생물, 예컨대 효모에서 발현을 자극시키는 단계를 포함한다. 특정 실시형태에서, 방법은 미생물이 셀룰로스, 예컨대 셀룰라제를 분해할 수 있게 하는 하나 이상의 효소, 예컨대 셀룰라제의 자극을 보장한다. 더 추가의 실시형태에서, RNA 표적화 CRISPR 복합체는 생물연료 생산 경로와 경쟁하는 내인성 물질대사 경로를 억제하는데 사용된다.

[0442] *채소유 또는 생물연료의 생산을 위한 조류 및 식물의 변형*

[0443] 유전자이식 조류 또는 다른 식물, 예컨대 평지는, 예를 들어 알코올(특히 메탄올 및 에탄올)과 같은 식물성 오일 또는 생체연료의 생성에서 특히 유용할 수 있다. 이들은 오일 또는 생체연료 산업에서 사용을 위한 고수준의 오일 또는 알코올을 발현 또는 과발현시키도록 유전자조작될 수 있다.

[0444] US 8945839호는 Cas9를 이용하여 미세조류(클라미도모나스 레인하티(Chlamydomonas reinhardtii) 세포) 종을 조작하는 방법을 설명한다. 유사한 도구를 사용하여, 본 명세서에서 기재된 RNA 표적화 CRISPR 시스템의 방법은 클라미도모나스 종 및 기타 다른 조류 상에 적용될 수 있다. 특정 실시형태에서, RNA 표적화 이펙터 단백질 및 가이드 RNA는, Hsp70A-Rbc S2 또는 Beta2 -튜불린과 같은 항상성 프로모터의 제어 하에서 RNA 표적화 이펙터 단백질을 발현하는 벡터를 사용하여 발현된 조류 내에 도입된다. 가이드 RNA는 T7 프로모터를 함유

하는 벡터를 사용하여 전달될 것이다. 대안적으로, 시험관 내 전사된 가이드 RNA는 조류 세포로 전달될 수 있다. 전기천공법 프로토콜은 진아트(GeneArt) 클라미도모나스 조작 키트로부터의 표준 권고 프로토콜을 따른다.

[0445] *식물에서 RNA 표적화 효소의 특정 응용*

[0446] 특정 실시형태에서, 본 발명은 바이러스 RNA를 절단할 수 있는 것처럼 식물 시스템에서 바이러스 제거를 위한 요법으로서 사용될 수 있다. 인간 시스템에서 이전의 연구들은 단일 가닥 RNA 바이러스, C형 간염 바이러스를 표적화하는데서 CRISPR을 사용한 성공을 입증하였다 (A. Price, et al., Proc. Natl. Acad. Sci, 2015). 이들 방법은 또한 식물에서 RNA 표적화 CRISPR 시스템을 사용하기 위해 적합화될 수 있다.

[0447] **개선된 식물**

[0448] 본 발명은 또한 본 명세서에서 제공된 방법에 의해 수득가능하거나 수득된 식물 및 효모 세포를 제공한다. 본 명세서에서 설명된 방법에 의해 수득된 개선된 식물은, 예를 들어 식물 해충, 제초제, 가뭄, 저온 또는 고온, 과다한 물, 등에 대한 내성을 보장하는 유전자의 변형된 발현을 통해 식물 또는 사료 생산에 유용할 수 있다.

[0449] 본 명세서에서 설명된 방법에 의해 수득된 개선된 식물은 특히 작물이며, 야생형에서 정상적으로 보일 수 있는 것보다 더 많은 예를 들어 단백질, 탄수화물, 영양소 또는 비타민 수준의 발현을 통해 식물 또는 사료 생산에 유용할 수 있다. 이와 관련하여, 개선된 식물, 특히 콩 종류 및 괴경이 바람직하다.

[0450] 유전자이식 조류 또는 다른 식물, 예컨대 평지는, 예를 들어 알코올(특히 메탄올 및 에탄올)과 같은 식물성 오일 또는 생체연료의 생성에서 특히 유용할 수 있다. 이들은 오일 또는 생체연료 산업에서 사용을 위한 고수준의 오일 또는 알코올을 발현 또는 과발현시키도록 유전자조작될 수 있다.

[0451] 본 발명은 또한 식물의 개선된 부분을 위해 제공된다. 식물 부분은 이로 제한되지는 않지만, 잎, 줄기, 뿌리, 괴경, 종자, 배젖, 배주, 및 화분을 포함한다. 여기서 고려되는 식물 부분은 자생성, 비자생성, 재생성, 및/또는 비재생성일 수 있다.

[0452] 본 발명의 방법에 따라 생성된 식물 세포 및 식물을 제공하는 것 또한 본 발명은 고려한다. 전통적인 육종 방법에 의해 생산된, 배우체, 종자, 접합성 또는 체세포성 중 하나의, 배아, 유전적 변형을 포함하는 식물의 자손 또는 잡종 또한 본 발명의 범주 내에 포함된다. 그러한 식물은 표적 서열 대신 또는 표적 서열에 삽입된 이종성 또는 외래의 DNA 서열을 함유할 수 있다. 대안적으로, 그러한 식물은 하나 이상의 뉴클레오티드에서 단지 하나의 변경(돌연변이, 결실, 삽입, 치환)을 함유할 수 있다. 그와 같이, 그러한 식물은 그들의 조상 식물과 특정 변형의 존재에 의해서만 상이할 것이다.

[0453] 본 발명의 실시형태에서, Cas13b 시스템은 예를 들어 박테리아, 진균 또는 바이러스에 의해 야기되는 질환에 대항하여 내성을 생성시켜서, 병원체 내성 식물을 조작하는데 사용된다. 일정 실시형태에서, 병원체 내성은 해충이 섭취하여, 죽음을 초래하게 되는 Cas13b 시스템을 생산하도록 작물을 조작하여 수행될 수 있다. 본 발명의 실시형태에서, Cas13b 시스템은 무생물적 스트레스 내성을 조작하는데 사용된다. 다른 실시형태에서, Cas13b 시스템은 가뭄 스트레스 내성 또는 염 스트레스 내성, 또는 저온 또는 고온 스트레스 내성을 조작하는데 사용된다. Younis et al. 2014, Int. J. Biol. Sci. 10;1150은 식물 육종 방법의 잠재적 표적을 고찰하고 있는데, 이들 모두는 본 명세서에 기술된 Cas13b 시스템의 사용을 통해 보정 또는 개선될 수 있다. 일부 비-제한적인 표적 작물은 아라비돕스(Arabidops) 옥수수(Zea mays) 탈리아나(thaliana), 오리자 사티바 엘(Oryza sativa L), 프루너스 도메스티카 엘.(Prunus domestica L.), 고시피움 히르수툼(Gossypium hirsutum), 니코티아나 루스티카(Nicotiana rustica), 옥수수(Zea mays), 메디카고 사티바(Medicago sativa), 니코티아나 벤타미아나(Nicotiana benthamiana) 및 아라비돕시스 탈리아나(Arabidopsis thaliana)를 포함한다.

[0454] 본 발명의 실시형태에서, Cas13b 시스템은 작물 해충의 관리에 사용된다. 예를 들어, 작물 해충에서 작동 가능한 Cas13b 시스템은 예를 들어 바이러스 벡터를 사용해, 표적에 직접적으로 형질전달되거나 또는 식물 숙주로부터 발현될 수 있다.

[0455] 일 실시형태에서, 본 발명은 이형접합 인간외 출발 유기체로부터 동형접합 유기체를 효율적으로 생산하는 방법을 제공한다. 일 실시형태에서, 본 발명은 식물 육종에서 사용된다. 다른 실시형태에서, 본 발명은 동물 육종에서 사용된다. 이러한 실시형태에서, 동형접합 유기체 예컨대 식물 또는 동물은 이중 가닥 파손, 염색체 쌍형성 및/또는 가닥 교환에 관여하는 적어도 하나의 표적 유전자에 의한 간섭을 통해 재조합을

방지하거나 또는 억제하여 만들어진다.

[0456] *최적화된 기능성 RNA 표적화 시스템에서 CAS13B 단백질의 응용*

[0457] 일 측면에서, 본 발명은 RNA 환경으로 기능성 성분의 특이적 전달을 위한 시스템을 제공한다. 이는 RNA로 상이한 성분의 특이적 표적화를 가능하게 하는 본 발명의 RNA 표적화 이펙터 단백질을 포함하는 CRISPR 시스템을 사용하여 보장된다. 보다 특히 이러한 성분은 활성인자 또는 억제인자, 예컨대 RNA 번역, 분해 등의 활성인자 또는 억제인자를 포함한다. 이러한 시스템의 적용은 본 명세서의 다른 곳에 기재되어 있다.

[0458] 일 측면에 따라서, 본 발명은 세포 내에서 관심 게놈 유전자좌 내 표적 서열과 혼성화할 수 있는 가이드 서열을 포함하는 가이드 RNA를 포함하는 비-천연 발생 또는 조작된 조성물을 제공하고, 가이드 RNA는 어댑터 단백질에 결합하는 하나 이상의 개별 RNA 서열 (들)의 삽입에 의해 변형된다. 특정 실시형태에서, RNA 서열은 둘 이상의 어댑터 단백질 (예를 들어, 앱타머)에 결합할 수 있고, 각 어댑터 단백질은 하나 이상의 기능성 도메인과 연관된다. 본 명세서에서 기술된 바와 같은 Cas13b 효소의 가이드 RNA는 가이드 서열의 변형으로 처리할 수 있는 것으로 확인되었다. 특정 구현예에서, 가이드 RNA는 직접 반복부의 5', 직접 반복부 내측, 또는 가이드 서열의 3'의 별개의 RNA 서열(들)의 삽입에 의해 변형된다. 1개 초과의 기능성 도메인이 존재하는 경우, 기능성 도메인은 동일하거나 상이할 수 있으며, 예를 들어, 2개가 동일하거나, 2개가 상이한 활성인자 또는 억제인자이다. 한 측면에서, 본 발명은 본원에 논의된 조성물을 제공하며, 여기서 하나 이상의 기능성 도메인은 RNA 표적화 효소에 부착되어, 표적 RNA에 결합시 기능성 도메인은 기능성 도메인이 그의 해당 기능에서 기능을 하도록 하는 공간 배열에 있다; 한 측면에서, 본 발명은 본원에 논의된 조성물을 제공하며, 여기서 조성물은 3 개 이상의 기능성 도메인을 갖는 CRISPR-Cas 복합체를 포함하며, 이들 중 1 개 이상은 RNA 표적화 효소와 관련되고 2 개 이상은 gRNA와 관련된다.

[0459] 따라서, 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 가이드 RNA 및 RNA 표적화 효소인 Cas13b를 포함하는 비-천연 발생 또는 조작된 CRISPR-Cas13b 복합체 조성물을 포함하고, 여기서 임의로 RNA 표적화 효소는 적어도 하나의 돌연변이를 포함하고, 그에 따라 RNA 표적화 효소는 적어도 하나의 돌연변이를 갖지 않는 효소의 뉴클레아제 활성의 5% 이하를 가지며, 임의로 적어도 하나 이상은적어도 하나 이상의 핵 국재화 서열을 포함한다. 특정 실시형태에서, 가이드 RNA는 여전히 RNA 표적화 효소의 결합을 보장하지만 RNA 표적화 효소에 의한 절단은 방지하도록 추가적으로 또는 대안적으로 변경된다 (본 명세서의 다른 곳에 상세하게 설명된 바와 같음).

[0460] 특정 실시형태에서, RNA 표적화 효소는 적어도 하나의 돌연변이를 갖지 않는 Cas13b 효소와 비교하여 적어도 97% 또는 100%의 감소된 뉴클레아제 활성을 갖는 Cas13b 효소이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 조성물을 제공하고, 여기서 Cas13b 효소는 다르게 본원에서 논의되지 않는다면 둘 이상의 돌연변이를 포함한다.

[0461] 특정 실시형태에서, RNA 표적화 시스템은 둘 이상의 기능성 도메인을 포함하는 본 명세서에 기술된 바와 같이 제공된다. 특정 실시형태에서, 둘 이상의 기능성 도메인은 이종성 기능성 도메인이다. 특정 실시형태에서, 시스템은 기능성 도메인을 포함하는 융합 단백질인 어댑터 단백질을 포함하고, 융합 단백질은 임의로 어댑터 단백질과 기능성 도메인 사이에 링커를 포함한다. 특정 실시형태에서, 링커는 GlySer 링커를 포함한다. 추가적으로 또는 부가적으로 하나 이상의 기능성 도메인은 링커, 임의로 GlySer 링커에 의해 RNA 이펙터 단백질에 부착된다. 특정 실시형태에서, 하나 이상의 기능성 도메인은 HEPN 도메인 중 하나 또는 둘 모두를 통해 RNA 표적화 효소에 부착된다.

[0462] 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 조성물을 제공하고, 어댑터 단백질 또는 RNA 표적화 효소와 연관된 하나 이상의 기능성 도메인은 RNA 번역을 활성화시키거나 또는 억제시키는 도메인이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 조성물을 제공하고, 어댑터 단백질과 연관된 하나 이상의 기능성 도메인의 적어도 하나는 메틸라제 활성, 탈메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, DNA 통합 활성, RNA 절단 활성, DNA 절단 활성 또는 핵산 결합 활성, 또는 분자 스위치 활성 또는 화학적 유도성 또는 광 유도성을 포함하는 하나 이상의 활성을 갖는다.

[0463] 일 측면에서, 본 발명은 앱타머 서열을 포함하는 본 명세서에 기술된 조성물을 제공한다. 특정 실시형태에서, 앱타머 서열은 동일한 어댑터 단백질에 특이적인 둘 이상의 앱타머 서열이다. 일 측면에서, 본 발명은 본 명세서에 기술된 조성물을 제공하고, 여기서 앱타머 서열은 상이한 어댑터 단백질에 특이적인 둘 이상의 앱타머 서열이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 조성물을 제공하고, 여기서 어댑터

단백질은 MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ΦCb5, ΦCb8r, ΦCb12r, ΦCb23r, 7s, PRR1을 포함한다.따라서, 특정 실시형태에서, 앱타머는 상기 열거된 어댑터 단백질 중 어느 하나에 특이적으로 결합하는 결합 단백질로부터 선택된다.

일 측면에서, 본 발명은 본 명 세서에 기술된 조성물을 제공하고, 세포는 진핵생물 세포이다. 일 측면에서, 본발명은 본 발명에 기술된 조성물을 제공하고, 여기서 진핵생물 세포는 포유동물 세포, 식물 세포 또는 효모 세포이고, 그리하여 포유동물 세포는 임의로 마우스 세포이다. 일 측면에서, 본 발명은 본 발명에 기술된 조성물을 제공하고, 여기서 포유동물 세포는 인간 세포이다.

[0464] 일 측면에서, 본 발명은 본 명세서에 기술된 조성물을 제공하고, 여기서 하나를 초과하는 가이드 RNA 또는 gRNA 또는 crRNA가 존재하고, 이들은 상이한 서열을 표적으로 하고 그리하여 조성물이 적용될 때, 다중화가 존재한다. 일 측면에서, 본 발명은 하나 이상의 어댑터 단백질에 결합하는 별개의 RNA 서열(들)의 삽입에 의해 변형되는 1가지 초과의 가이드 RNA 또는 gRNA 또는 crRNA가 존재하는 조성물을 제공한다.

[0465] 일 측면에서, 본 발명은 본 발명에 기술된 조성물을 제공하고, 여기서 하나 이상의 기능성 도메인과 연관된 하나 이상의 어댑터 단백질이 존재하고 가이드 RNA(들)에 삽입된 별개의 RNA 서열(들)에 결합된다.

[0466] 일 측면에서, 본 발명은 본 발명에 기술된 조성물을 제공하고, 여기서 가이드 RNA는 적어도 하나의 논-코딩 기능성 루프를 갖도록 변형되고, 예를 들어 적어도 하나의 논-코딩 기능성 루프는 억제성이고, 예를 들어 적어도 하나의 논-코딩 기능성 루프는 Alu를 포함한다.

[0467] 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 하나 이상의 조성물의 생체내 숙주 세포에서 발현 또는 숙주에 투여를 포함하는 유전자 발현을 변형하기 위한 방법을 제공한다.

[0468] 일 측면에서, 본 발명은 조성물 또는 이를 코딩하는 핵산 분자(들)의 전달을 포함하는 본 명세서에 기술된 방법을 제공하고, 여기서 상기 핵산 분자(들)는 조절 서열(들)에 작동적으로 연결되고 생체내에서 발현된다. 일 측면에서, 본 발명은 본 명세서에 기술된 방법을 제공하고, 여기서 생체내 발현은 렌티바이러스, 아데노바이러스, 또는 AAV를 통한다.

[0469] 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 세포의 포유동물 세포주를 제공하고, 여기서 세포주는 임의로 인간 세포주 또는 마우스 세포주이다. 일 측면에서, 본 발명은 유전자이식 포유동물 모델, 임의로 마우스를 제공하고, 여기서 모델은 본 발명에 기술된 조성물로 형질전환되거나 또는 상기 형질전환체의 자손이다.

[0470] 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같이 가이드 RNA 또는 RNA 표적화 CRISPR-Cas13b 복합체 또는 조성물을 인코딩하는 핵산 분자(들)를 제공한다. 일 측면에서, 본 발명은 세포에서 RNA 표적 서열과 혼성화할 수 있는 가이드 서열을 포함하는 가이드 RNA (gRNA) 또는 crRNA를 인코딩하는 핵산 분자를 포함하는 벡터를 제공하고, 여기서 gRNA 또는 crRNA의 직접 반복부는 둘 이상의 어댑터 단백질에 결합하는 별개 RNA 서열(들)의 삽입에 의해 변형되고, 각각의 어댑터 단백질은 하나 이상의 기능성 도메인과 연관되거나, 또는 gRNA는 적어도 하나의 논-코딩 기능성 루프를 갖도록 변형된다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 gRNA 또는 crRNA, 및 RNA 표적화 효소를 포함하는 비-천연 발생 또는 조작된 CRISPR-Cas13b 복합체 조성물을 엔코딩하는 핵산 분자(들)를 포함하는 벡터(들)를 제공하고, 임의로 RNA 표적화 효소는 적어도 하나의 돌연변이를 포함하여서, RNA 표적화 효소는 적어도 하나의 돌연변이를 갖지 않는 RNA 표적화 효소의 뉴클레아제 활성의 5% 이하를 갖고, 임의로 하나 이상은 적어도 하나 이상의 핵 국재화 서열을 포함한다. 일 측면에서, 벡터는 가이드 RNA (gRNA) 또는 crRNA를 인코딩하는 핵산 분자 및/또는 RNA 표적화 효소를 엔코딩하는 핵산 분자 및/또는 선택적 핵 국재화 서열(들)에 작동가능하게 연결된 진핵생물 세포에서 작동가능한 조절 요소(들)를 더 포함할 수 있다.

[0471] 일 양태에서, 본 발명은 본 명세서에서 기재된 하나 이상의 성분을 포함하는 키트를 제공한다. 일부 구현예에서, 키트는 본 명세서에 기재된 바와 같은 벡터 시스템 및 키트를 사용하기 위한 안내서를 포함한다.

[0472] 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 세포주 또는 RNA 표적화 효소를 함유하거나 또는 발현하는 본 명세서에 기술된 모델의 세포를 포함하고, 세포주 또는 모델의 세포본 명세서에 기술된 바와 같은 조성물을 도입하여, gRNA 또는 crRNA가 활성인자 또는 억제인자를 포함하게되고, 도입된 gRNA 또는 crRNA가 활성인자를 포함하는가에 대해 세포에 대해 또는 도입된 gRNA 또는 crRNA가 억제인자를 포함하는가에 대해 세포에 대해 개별적으로 GOF 또는 LOF에 대해 모니터링하는, 기능 획득 (GOF) 또는 기능 상실 (LOF)을 스크리닝하거나 또는

논-코딩 RNA 또는 잠재적 조절 영역 (예를 들어, 인핸서, 억제인자)를 스크리닝하는 방법을 제공한다.

[0473] 일 측면에서, 본 발명은 비-천연 발생 또는 조작된 조성물의 라이브러리를 제공하고, 각각은 세포에서 관심 표적 RNA 서열과 혼성화할 수 있는 가이드 서열, RNA 표적화 효소를 포함하는 RNA 표적화 CRISPR 가이드 RNA (gRNA) 또는 crRNA를 포함하고, 여기서 RNA 표적화 효소는 적어도 하나의 돌연변이를 포함하여, RNA 표적화 효소가 적어도 하나의 돌연변이를 갖지 않는 RNA 표적화 효소의 뉴클레아제 활성의 5% 이하를 갖고, gRNA 또는 crRNA는 하나 이상의 어댑터 단백질에 결합하는 별개의 RNA 서열(들)의 삽입에 의해 변형되고, 어댑터 단백질은 하나 이상의 기능성 도메인과 연관되며, 조성물은 하나 이상 또는 둘 이상의 어댑터 단백질을 포함하며, 여기서 각각의 단백질은 하나 이상의 기능성 도메인과 연관되고, gRNA 또는 crRNA는 다수의 RNA 표적화 가이드 RNA (gRNA) 또는 crRNA를 포함하는 게놈 전반 라이브러리를 포함한다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 RNA 표적화 RNA 표적화 효소는 적어도 하나의 돌연변이를 갖지 않는 RNA 표적화 효소와 비교하여 적어도 97%, 또는 100%의 감소된 뉴클레아제 활성을 갖는다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 어댑터 단백질은 기능성 도메인을 포함하는 융합 단백질이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 gRNA 또는 crRNA는 하나 또는 둘 또는 그 이상의 어댑터 단백질에 결합하는 별개의 RNA 서열(들)의 삽입에 의해 변형되지 않는다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 하나 또는 둘 또는 그 이상의 기능성 도메인은 RNA 표적화 효소와 연관된다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 세포의 세포 개체군은 진핵생물 세포의 개체군이다. 일 측면에서, 본 발명은 본 명세서에서 제공되는 바와 같은 라이브러리를 제공하고, 여기서 진핵생물 세포는 포유동물 세포, 식물 세포 또는 효모 세포이다. 일 측면에서, 본 발명은 본 발명에 기술된 바와 같은 라이브러리를 제공하고, 여기서 포유동물 세포는 인간 세포이다. 일 측면에서, 본 발명은 본 발명에 기술된 바와 같은 라이브러리를 제공하고, 여기서 세포의 개체군은 배아 줄기 (ES) 세포의 개체군이다.

[0474] 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 표적화는 약 100 이상의 RNA 서열에 대한 것이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 표적화는 약 1000 이상의 RNA 서열에 대한 것이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 표적화는 약 20,000 또는 그 이상의 서열에 대한 것이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 표적화는 전체 전사체에 대한 것이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 표적화는 관련 또는 바람직한 경로에 집중하는 표적 서열의 패널에 대한 것이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 경로는 면역 경로이다. 일 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 라이브러리를 제공하고, 여기서 경로는 세포 분열 경로이다.

[0475] 일정 측면에서, 본 발명은 발현이 변형된 유전자를 포함하는 모델 진핵생물 세포를 생성시키는 방법을 제공한다. 일부 구현예에서, 질병 유전자는 질병을 갖거나 발생시키는 위험에서의 증가와 연관된 임의의 유전자이다. 일부 실시형태에서, 방법은 (a) 진핵생물 세포로 상기 본 명세서에 기술된 시스템의 성분을 인코딩하는 하나 이상의 벡터를 도입시키는 단계, 및 (b) CRISPR 복합체가 표적 폴리뉴클레오티드에 결합할 수 있게 하여, 유전자의 발현을 변형시켜, 변형된 유전자 발현을 포함하는 모델 진핵생물 세포를 생성시키는 단계를 포함한다.

[0476] 본 명세서에 제공되는 구조적 정보는 전체 RNA 표적화 CRISPR-Cas13b 시스템의 작용성을 최적화하기 위해 가이드 RNA 구조의 조작 또는 변경을 가능하게 하는 표적 RNA 및 RNA 표적화 효소와 가이드 RNA 또는 crRNA 상호작용에 관한 질문을 가능하게 한다. 예를 들어, 가이드 RNA 또는 crRNA는 RNA에 결합할 수 있는 어댑터 단백질의 삽입에 의해 RNA 표적화 단백질과 충돌없이 확장될 수 있다. 이들 어댑터 단백질은 하나 이상의 기능성 도메인을 포함하는 이펙터 단백질 또는 융합체를 추가로 동원할 수 있다.

[0477] 본 발명의 일 양태는 상기 요소가 단일의 조성물에 포함되거나, 개별 조성물에 포함되는 것이다. 이들 조성물을 유리하게는 숙주에 적용하여, 게놈 수준에서 기능적 효과를 유도할 수 있다.

[0478] 숙련자는 어댑터 + 기능성 도메인의 결합을 허용하지만, (예를 들어, CRISPR Cas13b 복합체의 3차원 구조 내의 입체 장해로 인하여) 어댑터 + 기능성 도메인의 적절한 배치를 허용하지 않는 가이드 RNA 또는 crRNA에 대한 변형은 의도되지 않는 변형임을 이해할 것이다. 하나 이상의 가이드 RNA 또는 crRNA는 직접 반복부의 5', 직접 반복부 내, 또는 가이드 서열의 3'에서의 별개의 RNA 서열(들)의 도입에 의해 변형될 수 있다.

[0479] 변형된 가이드 RNA 또는 crRNA, 불활성화된 RNA 표적화 효소 (기능성 도메인이 있거나 또는 없음), 및 하나 이상의 기능성 도메인을 갖는 결합 단백질은 각각 개별적으로 조성물에 포함될 수 있고 숙주에 개별적으로 또는 종합적으로 투여될 수 있다. 대안적으로, 이들 성분은 숙주에의 투여를 위해 단일 조성물로 제공될 수 있다. 숙주에 투여는 숙주에게 전달을 위해 본 명세서에 기술되거나 또는 당업자에게 공지된 바이러스 벡터 (예를 들어, 렌티바이러스 벡터, 아데노바이러스 벡터, AAV 벡터)를 통해 수행될 수 있다. 본 명세서에 설명하는 바와 같이, 상이한 선택 마커 (예를 들어, 렌티바이러스 gRNA 또는 crRNA 선택용) 및 gRNA 또는 crRNA의 농도 (예를 들어, 다수 gRNA 또는 crRNA가 사용되는지 여부에 의존적)의 사용은 개선된 효과를 유발시키기 위해 유리할 수 있다.

[0480] 제공된 조성물을 사용하여, 당업자는 하나 이상의 게놈 사건을 유발시키기 위해 동일하거나 또는 상이한 기능성 도메인을 갖는 단일 또는 다수 유전자좌를 유리하게 그리고 특이적으로 표적화할 수 있다. 조성물은 세포 내 라이브러리의 스크리닝 및 *생체 내* 기능적 모델링을 위한 다양한 방법으로 적용될 수 있다 (예를 들어, lincRNA의 유전자 활성화 및 기능의 식별; 기능의 획득 모델링; 기능의 손실 모델링; 최적화 및 스크리닝 목적을 위해 세포주 및 형질전환 동물을 확립하기 위해 본 발명의 조성물의 사용).

[0481] 본 발명은 조건적 또는 유도성 CRISPR Cas13b RNA 표적화 사건을 확립하고 활용하기 위한 본 발명의 조성물의 용도를 이해한다. (예를 들어, 문헌 [Platt et al., Cell (2014), http://dx.doi.org/10.1016/j.cell.2014.09.014], 또는 본 명세서에서 인용하는 PCT 특허 공개, 예컨대 WO 2014/093622 (PCT/US2013/074667)를 참조하지만, 본 발명 또는 출원의 선행이라고 여기지는 않음). .

[0482] **데드 가이드 서열을 포함하는 본 발명에 따른 가이드 RNA**

[0483] 일 양태에서, 본 발명은 CRISPR Cas13b 복합체의 형성 및 표적으로의 성공적인 결합을 허용함과 동시에, 성공적인 뉴클레아제 활성을 허용하지 않는(즉, 뉴클레아제 활성 부재/삽입결실 활성 부재)방식으로 변형되는 가이드 서열을 제공한다. 설명을 위하여, 이러한 변형된 가이드 서열은 "데드 가이드" 또는 "데드 가이드 서열"로 지칭된다. 이들 데드 가이드 또는 데드 가이드 서열은 뉴클레아제 활성에 관하여 촉매적으로 비활성이거나 입체형태적으로 비활성인 것으로 생각될 수 있다. 실제로, 데드 가이드 서열은 표적-적중 및 표적-이외 결합 활성을 구별하거나 또는 촉매 활성을 촉진하는 능력에 대해 생산적인 염기 쌍형성에 충분하게 관여하지 않을 수 있다. 간략하게, 어세이는 CRISPR 표적 RNA 및 표적 RNA와 불일치를 포함하는 가이드 RNA를 합성하는 단계, 이들을 RNA 표적화 효소와 조합하는 단계 및 절단 생성물에 의해 생성된 밴드의 존재를 기반으로 겔을 기반하여 절단을 분석하는 단계 및 상대적 밴드 강도를 기반으로 절단을 정량하는 단계를 포함한다.

[0484] 그런 이유로, 관련 측면에서, 본 발명은 본 명세서에 기술된 바와 같은 기능성 RNA 표적화 효소, 및 가이드 RNA (gRNA) 또는 crRNA를 포함하는 비-천연 발생 또는 조작된 조성물 RNA 표적화 CRISPR-Cas 시스템을 제공하고, 여기서 gRNA 또는 crRNA는 데드 가이드 서열을 포함하고, 그리하여 gRNA는 표적 서열과 혼성화할 수 있어서 RNA 표적화 CRISPR-Cas 시스템이 시스템의 비돌연변이체 RNA 표적화 효소의 검출가능한 RNA 절단 활성없이 세포 내 관심 게놈 유전자로 유도된다. 본원의 다른 곳에 기재된 바와 같은 본 발명에 따른 gRNA 또는 crRNA 중 임의의 것이 데드 gRNA/데드 가이드 서열을 포함하는 crRNA로서 사용될 수 있는 것이 이해될 것이다.

[0485] RNA 표적 서열로 CRISPR 복합체의 서열-특이적 결합을 유도하는 데드 가이드 서열의 능력은 임의 적합한 어세이로 평가될 수 있다. 예를 들어, 시험하려는 데드 가이드 서열을 포함하는, CRISPR Cas13b 복합체를 형성하는데 충분한 CRISPR Cas13b 시스템의 성분이 예컨대 시스템의 성분을 인코딩하는 벡터로 트랜스펙션 후, 표적 서열 내 선호되는 절단의 평가를 통해서 상응하는 표적 서열을 갖는 숙주 세포에 제공될 수 있다.

[0486] 본원에 추가로 설명된 바와 같이, 몇몇의 구조적 파라미터는 적절한 프레임워크가 이러한 데드 가이드에 도달하게 한다. 데드 가이드 서열은 활성 RNA 절단을 일으키는 개별 가이드 서열보다 전형적으로 더 짧을 수 있다. 특정 실시형태에서, 데드 가이드는 동일하게 지정되는 개별 가이드보다 5%, 10%, 20%, 30%, 40%, 50% 더 짧다.

[0487] 하기에 설명되고 당분야에 공지된 바와 같이, gRNA 또는 crRNA - RNA 표적화 특이성의 일 측면은 직접 반복부 서열로서, 이러한 가이드에 적절하게 연결된다. 특히 이는 직접 반복부 서열이 RNA 표적화 효소의 기원에 의존적으로 디자인된다는 것을 추론한다. 검증된 데드 가이드 서열에 대해 입수할 수 있는 구조적 데이타는 Cas13b 특이적 균등물을 디자인하는데 사용될 수 있다. 예를 들어 둘 이상의 Cas13b 이펙터 단백질의 동족체 뉴클레아제 도메인 HEPN 사이의 구조적 유사성은 균등한 데드 가이드를 디자인을 전달하는데 사용될 수 있다. 따라서, 본 명세서의 데드 가이드는 CRISPR Cas13b 복합체의 형성 및 표적 RNA와 성공적인 결합을 가

능하게 하는 한편, 동시에 성공적인 뉴클레아제 활성은 허용하지 않는, 이러한 Cas13b 특이적 균등물을 반영하도록 길이 및 서열을 적절하게 변형시킬 수 있다.

[0488] 데드 가이드는 뉴클레아제 활성의 결과없이, 유전자 표적화를 수단으로서 gRNA 또는 crRNA를 사용할 수 있게 하는 동시에 활성화 또는 억제를 위한 지정된 수단을 제공한다. 데드 가이드를 포함하는 가이드 RNA 또는 crRNA는 유전자 활성의 활성화 또는 억제를 허용하는 방식으로 요소, 특히, 본원의 다른 곳에 기재된 바와 같은 단백질 어댑터(예를 들어, 앱타머)를 추가로 포함하여, 유전자 이펙터(예를 들어, 유전자 활성의 활성인자 또는 억제인자)의 기능적 배치를 허용하도록 변형될 수 있다. 일 예는 본원에 설명된, 해당 분야의 상태의 앱타머의 혼입이다. 단백질-상호작용 앱타머를 도입하기 위해 데드 가이드를 포함하는 gRNA 또는 crRNA를 조작하기 위해서(참조로 본 명세서에 편입된, 문헌 [Konermann et al., "Genome-scale transcription activation by an engineered CRISPR-Cas9 complex", doi:10.1038/nature14136] 참조), 다수의 별개 이펙터 도메인을 조립할 수 있다. 이와 같이 천연 과정 후에 모델링할 수 있다.

[0489] **일반 정보**

[0490] 본 발명의 실시형태에서, 용어 성숙 가이드 서열 및 가이드 RNA는 상기 언급된 문헌, 예컨대 WO 2014/093622(PCT/US2013/074667)에서 상호교환 가능하게 사용된다. 일반적으로, 가이드 서열은 표적 서열과 혼성화하고, 표적 서열로의 CRISPR 복합체의 서열-특이적 결합을 유도하기에 충분한, 표적 폴리뉴클레오티드 서열과의 상보성을 갖는 임의의 폴리뉴클레오티드 서열이다. 일부 구현예에서, 가이드 서열과 그의 상응하는 표적 서열 간의 상보성의 정도는 적절한 정렬 알고리즘을 사용하여 최적으로 정렬되는 경우, 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% 이상이다. 최적의 정렬은 서열을 정렬하기에 적절한 임의의 알고리즘의 사용으로 결정될 수 있으며, 그의 비-제한적인 예는 스미스-워터만(Smith-Waterman) 알고리즘, 니들만-분쉬(Needleman-Wunsch) 알고리즘, 버로우즈-휠러 트랜스폼(Burrows-Wheeler Transform)에 기초한 알고리즘(예를 들어, 버로우즈 휠러 얼라이너(Burrows Wheeler Aligner)), ClustalW, Clustal X, BLAT, 노보얼라인(Novoalign)(노보크라프트 테크놀로지즈(Novocraft Technologies)(www.novocraft.com에서 이용가능함), ELAND(일루미나(Illumina), 미국 캘리포니아주 샌디에고), SOAP(soap.genomics.org.cn에서 이용가능) 및 Maq(maq.sourceforge.net에서 이용가능)를 포함한다. 일부 구현예에서, 가이드 서열은 길이가 약 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 개 이상이거나 이를 초과한다. 일부 구현예에서, 가이드 서열은 길이가 약 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 개 이하의 뉴클레오티드이다. 바람직하게 가이드 서열은 길이가 10 개 내지 30 개 뉴클레오티드, 예컨대 30 개 뉴클레오티드 길이이다. 표적 서열로의 CRISPR 복합체의 서열-특이적 결합을 유도하는 가이드 서열의 능력은 임의의 적절한 어세이에 의해 평가될 수 있다. 예를 들어, 시험되는 가이드 서열을 포함하는 CRISPR 복합체를 형성하기에 충분한 CRISPR 시스템의 성분은 예를 들어, CRISPR 서열의 성분을 인코딩하는 벡터로의 트랜스펙션 후에, 예를 들어, 본원에 기술된 바와 같은 서베이어 어세이에 의한 표적 서열 내의 우선적인 절단의 평가에 의해서, 상응하는 표적 서열을 갖는 숙주 세포로 제공될 수 있다. 유사하게, 표적 폴리뉴클레오티드 서열의 절단은 표적 서열, 시험되는 가이드 서열 및 시험 가이드 서열과 상이한 대조군 가이드 서열을 포함하는 CRISPR 복합체의 성분을 제공하고, 표적 서열에서 시험 및 대조군 가이드 서열 반응 간의 결합 또는 절단 비율을 비교함으로써 시험관에서 평가될 수 있다. 다른 어세이가 가능하며, 해당 분야의 숙련자에게 일어날 것이다.임의의 표적 서열을 표적으로 하기 위해 가이드 서열이 선택될 수 있다. 일부 구현예에서, 표적 서열은 세포의 게놈 내 서열이다. 예시적인 표적 서열은 표적 게놈 내의 독특한 것들을 포함한다.

[0491] 일반적으로, 그리고 본원에서, 용어 "벡터"는 그것이 연결된 다른 핵산을 수송할 수 있는 핵산 분자를 지칭한다. 벡터는 제한 없이, 단일 가닥, 이중 가닥 또는 부분 이중 가닥인 핵산 분자; 하나 이상의 자유 말단을 포함하는, 자유 말단을 포함하지 않는 (예 : 환형) 핵산 분자; DNA, RNA 또는 둘 다를 포함하는 핵산 분자; 및 당업계에 공지된 다른 종류의 폴리뉴클레오티드를 포함한다. 하나의 유형의 벡터는 "플라스미드"이며, 이는 추가의 DNA 절편이 예를 들어, 표준 분자 클로닝 기술에 의해 삽입될 수 있는 환형 이중 가닥 DNA 루프를 지칭한다. 벡터의 다른 종류는 바이러스 벡터로서, 바이러스-유래된 DNA 또는 RNA 서열이 바이러스에 봉입되는 벡터에 존재한다(예를 들어, 레트로바이러스, 복제 결함 레트로바이러스, 아데노바이러스, 복제 결함 아데노바이러스, 및 아데노-관련 바이러스(AAV)). 바이러스 벡터는 또한 숙주 세포 내로의 트랜스펙션을 위한, 바이러스가 보유하는 폴리뉴클레오티드를 포함한다. 특정 벡터는 그것이 도입된 숙주 세포에서 자율적 복제가 가능하다(예를 들어, 박테리아 복제 기원을 가진 박테리아 벡터 및 에피솜 포유류 벡터). 다른 벡터(예를 들어, 비-에피솜 포유동물 벡터)는 숙주 세포 내로 도입시 숙주 세포의 게놈에 통합되며, 이에 의해

숙주 게놈과 함께 복제된다. 더욱이, 특정 벡터는 그것이 작동가능하게 연결된 유전자의 발현을 지시할 수 있다. 이러한 벡터는 본원에서 "발현 벡터"로 지칭된다. 진핵 세포에서의 발현을 위한 벡터 및 이를 초래하는 벡터는 본 명세서에서 "진핵세포 발현 벡터"로 지칭될 수 있다. 재조합 DNA 기술에 유용한 통상적인 발현 벡터는 종종 플라스미드의 형태로 존재한다.

[0492] 재조합 발현 벡터는 숙주 세포에서 핵산의 발현에 적합한 형태로 본 발명의 핵산을 포함할 수 있는데, 이는 재조합 발현 벡터가 하나 이상의 조절 요소를 포함하는 것을 의미하며, 하나 이상의 조절 요소는 발현에 사용될 숙주 세포에 기반하여 선택될 수 있고, 발현될 핵산 서열에 작동가능하게 연결된다. 재조합 발현 벡터 내에서, "작동가능하게 연결된"은 대상 뉴클레오티드 서열이 (예를 들어, 시험관내 전사/번역 시스템 내에서, 또는 벡터가 숙주 세포 내로 도입되는 경우 숙주 세포 내에서) 뉴클레오티드 서열의 발현을 가능하게 하는 방식으로 조절 요소(들)에 연결된 것을 의미하는 의도이다.

[0493] 용어 "조절 요소"는 프로모터, 인핸서, 내부 리보솜 엔트리 부위(internal ribosomal entry site; IRES), 및 기타 다른 발현 제어 요소(예를 들어, 전사 종결 신호, 예컨대 폴리아데닐화 신호 및 폴리-U 서열)를 포함하는 것으로 의도된다. 이와 같은 조절 요소는, 예를 들어 문헌 [Goeddel, GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)]에 기재되어 있다. 조절 요소는 많은 유형의 숙주 세포에서 뉴클레오티드 서열의 구성적 발현을 지시하는 것들과 특정 숙주 세포에서만 뉴클레오티드 서열의 발현을 지시하는 것들(예를 들어, 조직-특이적 조절 서열)을 포함한다. 조직-특이적 프로모터는 근육, 뉴런, 뼈, 피부, 혈액, 특정 장기(예를 들어, 간, 췌장)와 같은 요망되는 관심 조직, 또는 특정 세포 유형(예를 들어, 림프구)에서 주로 발현을 유도할 수 있다. 조절 요소는 시간-의존적 방식으로, 예컨대 세포-주기 의존적 또는 발생 단계 의존적 방식으로 발현을 지시할 수 있으며, 이것은 조직 특이적이거나 세포-유형 특이적일 수 있거나, 또한 그렇지 않을 수 있다. 일부 구현예에서, 벡터는 하나 이상의 pol III 프로모터 (예를 들어, 1, 2, 3, 4, 5 개 또는 그 이상의 pol III 프로모터), 하나 이상의 pol II 프로모터(예를 들어, 1, 2, 3, 4, 5 개 또는 그 이상의 pol II 프로모터), 하나 이상의 pol I 프로모터(예를 들어, 1, 2, 3, 4, 5 개 또는 그 이상의 pol I 프로모터), 또는 이들의 조합을 포함한다. pol III 프로모터의 예들은, 제한은 아니지만, U6 및 H1 프로모터를 포함한다. pol II 프로모터의 예들은, 이로 제한되지는 않지만, 레트로바이러스 라우스 육종 바이러스(Rous sarcoma virus; RSV) LTR 프로모터(선택적으로 RSV 인핸서와 함께), 사이토메갈로바이러스(CMV) 프로모터(선택적으로 CMV 인핸서와 함께)[예를 들어, 문헌[Boshart et al, Cell, 41:521-530 (1985)] 참조], SV40 프로모터, 디하이드로폴레이트 환원효소 프로모터, β-액틴 프로모터, 포스포글리세롤 키나아제(phosphoglycerol kinase; PGK) 프로모터, 및 EF1α 프로모터를 포함한다. 용어 "조절 요소"는 인핸서 엘리먼트, 예컨대 WPRE; CMV 인핸서; HTLV-1의 LTR 내의 R-U5 '분절 (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 인핸서; 및 토끼 β-글로빈의 엑손 2와 3 사이의 인트론 서열 (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981)이다. 발현 벡터의 디자인은 형질전환될 숙주 세포의 선택, 원하는 발현 수준 등과 같은 인자에 의존할 수 있다는 것을 당업자는 이해할 것이다. 벡터는 숙주 세포 내로 도입되어 본원에 기재된 바와 같이 핵산에 의해 인코딩된 융합 단백질 또는 펩티드를 포함하는 전사체, 단백질 또는 펩티드를 생산할 수 있다 (예를 들어, 클러스터링되고 규칙적으로 산재된 짧은 팔린드로믹 반복부(Clustered Regularly Interspaced Short Palindromic Repeats; CRISPR) 전사체, 단백질, 효소, 이의 돌연변이체 형태, 이의 융합 단백질 등).

[0494] 유리한 벡터는 렌티바이러스 및 아데노-연관 바이러스를 포함하며, 이러한 벡터의 유형은 또한 특정 유형의 세포를 표적화하기 위해 선택될 수 있다.

[0495] 본 명세서에서 사용된 바와 같이, 유형 VI CRISPR-Cas 유전자좌 이펙터 단백질의 용어 "crRNA" 또는 "가이드 RNA" 또는 "단일 가이드 RNA" 또는 "sgRNA" 또는 "하나 이상의 핵산 성분"은 표적 핵산 서열과 혼성화하고, 표적 RNA 서열로의 RNA-표적화 복합체의 서열-특이적 결합을 지시하기에, 표적 핵산 서열과의 충분한 상보성을 갖는 임의의 폴리뉴클레오티드 서열을 포함한다.

[0496] 일정 실시형태에서, 본원에 제공된 CRISPR 시스템은 crRNA 또는 가이드 서열을 포함하는 유사한 폴리뉴클레오티드를 사용할 수 있으며, 여기서 폴리뉴클레오티드는 RNA, DNA 또는 RNA와 DNA의 혼합물이고/거나, 폴리뉴클레오티드는 하나 이상의 뉴클레오티드 유사체를 포함한다. 서열은 벌지 (bulge), 헤어핀 또는 스템 루프 구조와 같은 천연 crRNA의 구조를 포함하지만 이에 제한되지 않는 임의의 구조를 포함할 수 있다. 일정 실시형태에서, 가이드 서열을 포함하는 폴리뉴클레오티드는 RNA 또는 DNA 서열일 수 있는 제 2 폴리 뉴클레오티드 서열과의 이중체를 형성한다.

[0497] 일정 실시형태에서, 본 발명의 가이드는 비-천연 발생 핵산 및/또는 비-천연 발생 뉴클레오티드 및/또는 뉴클레오티드 유사체, 및/또는 화학적 변형을 포함한다. 비-천연 발생 핵산은 예를 들어 천연 및 비-천연 뉴클레오티드의 혼합물을 포함할 수 있다. 비-천연 발생 뉴클레오티드 및/또는 뉴클레오티드 유사체는 리보스, 포스페이트 및/또는 염기 모이어티에서 변형될 수 있다. 본 발명의 일 구현예에서, 가이드 핵산은 리보뉴클레오티드 및 비-리보뉴클레오티드를 포함한다. 이러한 일 구현예에서, 가이드는 하나 이상의 리보뉴클레오티드 및 하나 이상의 데옥시리보뉴클레오티드를 포함한다. 본 발명의 일 구현예에서, 가이드는 하나 이상의 비-천연 발생 뉴클레오티드 또는 뉴클레오티드 유사체, 예컨대 포스포로티오에이트 결합, 보라노포스페이트 결합이 있는 뉴클레오티드, 리보스 고리의 2' 및 4' 탄소 사이에 메틸렌 브릿지를 포함하는 잠금 핵산(LNA) 뉴클레오티드 또는 브릿징된 핵산(BNA)을 포함한다. 변형된 뉴클레오티드의 다른 예는 2'-O-메틸 유사체, 2'-데옥시 유사체, 2-티오우리딘 유사체, N6-메틸아디노신 유사체 또는 2'-플루오로 유사체를 포함한다. 변형된 염기의 추가의 예에는 2-아미노퓨린, 5-브로모-우리딘, 슈도우리딘 (Ψ), N1-메틸슈도우리딘 (me1Ψ), 5-메톡시우리딘(5moU), 이노신, 7-메틸구아노신이 포함되나 이들에 한정되지 않는다. 가이드 RNA 화학 변형의 예로는, 이로 제한되지는 않지만, 하나 이상의 말단 뉴클레오티드에서 2'-O-메틸(M), 2'-O-메틸 3'포스포로티오에이트(MS), S-속박된 에틸 (cEt), 또는 2'-O-메틸 3'티오PACE(MSP)의 혼입을 포함한다. 이러한 화학적으로 변형된 가이드 RNA는, 표적상 특이성 대 표적외 특이성은 예측가능하지 않지만, 비변형된 가이드 RNA에 비하여 증가된 안정성 및 증가된 활성을 포함할 수 있다. (참조, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015; Allerson et al., J. Med. Chem. 2005, 48:901-904; Bramsen et al., Front. Genet., 2012, 3:154; Deng et al., PNAS, 2015, 112:11870-11875; Sharma et al., MedChemComm., 2014, 5:1454-1471; Li et al., Nature Biomedical Engineering, 2017, 1, 0066 DOI:10.1038/s41551-017-0066).

[0498] 일부 실시형태에서, 가이드 RNA의 5 '및/또는 3' 말단은 형광 염료, 폴리에틸렌 글리콜, 콜레스테롤, 단백질 또는 검출 태그를 포함하는 다양한 기능성 부분에 의해 변형된다. (참조, Kelly et al., 2016, J. Biotech. 233:74-83). 특정 구현예에서, 가이드는 표적 DNA에 결합하는 영역 내에 리보뉴클레오티드 및 Cas9, Cpf1, 또는 C2c1에 결합하는 영역 내에 하나 이상의 데옥시리보뉴클레오티드 및/또는 뉴클레오티드 유사체를 포함한다. 본 발명의 일 실시형태에서, 데옥시리보뉴클레오티드 및/또는 뉴클레오티드 유사체는 조작된 가이드 구조, 예컨대, 제한 없이 5' 및/또는 3' 말단, 스템-루프 영역, 및 시드 영역에 혼입된다. 특정 실시형태에서, 변형은 스템-루프 영역의 5'-핸들에 있지 않다. 가이드의 스템-루프 영역의 5'-핸들에서의 화학적 변형은 그 기능을 소멸시킬 수 있다 (참조, Li, et al., Nature Biomedical Engineering, 2017, 1:0066).

특정 실시형태에서, 가이드의 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 또는 75개의 뉴클레오티드가 화학적으로 변형된다. 일부 실시형태에서, 가이드의 3 '또는 5'말단에서 3-5 개의 뉴클레오티드가 화학적으로 변형된다. 일부 실시형태에서, 2'-F 변형과 같은 작은 변형 만이 시드 영역에 도입된다. 일부 실시형태에서, 2'-F 변형은 가이드의 3 '말단에 도입된다. 특정 실시형태에서, 가이드의 5' 및/또는 3' 말단에 3 내지 5 개의 뉴클레오티드가 2'-O-메틸 (M), 2'-O-메틸-3'-포스포로티오에이트 (MS), S-속박된 에틸 (cEt), 또는 2'-O-메틸-3'-티오PACE (MSP)로 화학적으로 변형된다. 이러한 변형은 게놈 편집 효율을 향상시킬 수 있다 (참조, Hendel et al., Nat. Biotechnol. (2015) 33 (9) : 985-989). 특정 실시형태에서, 가이드의 모든 포스포디에스테르 결합은 유전자 파괴 수준을 향상시키기 위해 포스포로티오에이트 (PS)로 치환된다. 특정 실시형태에서, 가이드의 5' 및/또는 3' 말단에 5 개 초과의 뉴클레오티드가 2'-O-Me, 2'-F 또는 S-속박된 에틸(cEt)로 화학적으로 변형된다. 이러한 화학적으로 변형된 가이드는 강화된 수준의 유전자 파괴를 매개할 수 있다 (참조, Ragdarm et al., 0215, PNAS, E7110-E7111). 본 발명의 한 실시형태에서, 가이드는 3 '및/또는 5'말단에 화학적 모이어티를 포함하도록 변형된다. 이러한 모이어티에는 아민, 아지드, 알킨, 티오, 디벤조시클로옥틴 (DBCO) 또는 로다민이 포함되나 이에 제한되지는 않는다. 특정 실시형태에서, 화학적 모이어티는 알킬 쇄와 같은 링커에 의해 가이드에 콘쥬게이션된다. 특정 실시형태에서, 변형된 가이드의 화학적 모이어티는 가이드를 DNA, RNA, 단백질 또는 나노입자와 같은 또다른 분자에 부착하는데 사용될 수 있다.

이러한 화학적으로 변형된 가이드는 CRISPR 시스템에 의해 일반적으로 편집된 세포를 식별하거나 풍부하게 하기 위해 사용될 수 있다 (참조, Lee et al., eLife, 2017, 6:e25312, DOI:10.7554)

[0499] 본 발명의 한 측면에서, 가이드는 5'-핸들 및 시드 영역 및 3'-말단을 추가로 포함하는 가이드 분절을 갖는, Cpf1에 대한 변형된 crRNA를 포함한다. 일부 실시형태에서, 변형된 가이드는 Acidaminococcus sp. BV3L6 Cpf1 (AsCpf1); Francisella tularensis subsp. Novicida U112 Cpf1 (FnCpf1); L. 박테리움 MC2017 Cpf1 (Lb3Cpf1); Butyrivibrio proteoclasticus Cpf1 (BpCpf1); Parcubacteria bacterium GWC2011_GWC2_44_17 Cpf1

(PbCpf1); Peregrinibacteria bacterium GW2011_GWA_33_10 Cpf1 (PeCpf1); Leptospira inadai Cpf1 (LiCpf1); Smithella sp. SC_K08D17 Cpf1 (SsCpf1); L. bacterium MA2020 Cpf1 (Lb2Cpf1); Porphyromonas crevioricanis Cpf1 (PcCpf1); Porphyromonas macacae Cpf1 (PmCpf1); Candidatus Methanoplasma termitum Cpf1 (CMtCpf1); Eubacterium eligens Cpf1 (EeCpf1); Moraxella bovoculi 237 Cpf1 (MbCpf1); Prevotella disiens Cpf1 (PdCpf1); 또는 L. bacterium ND2006 Cpf1 (LbCpf1) 중 어느 하나의 Cpf1과 함께 사용될 수 있다.

[0500] 일부 실시형태에서, 가이드에 대한 변형은 화학적 변형, 삽입, 결실 또는 스플릿이다. 일부 실시형태에서, 화학적 변형은 2'-O-메틸 (M) 유사체, 2'-데옥시 유사체, 2- 티오우리딘 유사체, N6-메틸아데노신 유사체, 2'-플루오로 유사체, 2-아미노푸린, 5-브로모-우리딘, 슈도우리딘 (Ψ), N1-메틸슈도우리딘 (me1Ψ), 5-메톡시우리딘 (5moU), 이노신, 7- 메틸구아노신, 2'-O-메틸-3'-포스포로티오에이트 (MS), S-속박된 에틸 (cEt), 포스포로티오에이트 (PS) 또는 2'-O-메틸-3'-티오PACE (MSP)의 혼입을 포함하나, 이에 제한되지는 않는다. 일부 구현예에서, 가이드는 하나 이상의 포스포로티오에이트 개질을 포함한다. 특정 실시형태에서, 가이드의 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 25개의 뉴클레오티드가 화학적으로 변형된다. 특정 실시형태에서, 시드 영역 내의 하나 이상의 뉴클레오티드가 화학적으로 변형된다. 특정 실시형태에서, 3' -말단 내의 하나 이상의 뉴클레오티드가 화학적으로 변형된다. 특정 실시형태에서, 5' -핸들 내의 뉴클레오티드 중 어느 것도 화학적으로 변형되지 않는다. 일부 실시형태에서, 시드 영역 내의 화학적 변형은 2'-플루오로 유사체의 혼입과 같은 작은 변형이다. 특정 실시형태에서, 시드 영역의 하나의 뉴클레오티드는 2'-플루오로 유사체로 대체된다. 일부 실시형태에서, 3' -말단 내의 5 또는 10 개의 뉴클레오티드가 화학적으로 변형된다. Cpf1 CrRNA의 3'-말단에서의 이러한 화학적 변형은 유전자 절단 효율을 향상시킨다 (참조, Li, et al., Nature Biomedical Engineering, 2017, 1:0066). 특정 실시형태에서, 3' -말단 내의 5개의 뉴클레오티드는 2'-플루오로 유사체로 대체된다. 특정 실시형태에서, 3' -말단 내의 10개의 뉴클레오티드는 2'-플루오로 유사체로 대체된다. 특정 실시형태에서, 3' -말단 내의 5개의 뉴클레오티드는 2'-O-메틸 (M) 유사체로 대체된다.

[0501] 일부 실시형태에서, 가이드의 5'-핸들의 루프가 변형된다. 일부 실시형태에서, 가이드의 5'-핸들의 루프는 결실, 삽입, 분할 또는 화학적 변형을 갖도록 변형된다. 특정 실시형태에서, 루프는 3, 4 또는 5 개의 뉴클레오티드를 포함한다. 특정 실시형태에서, 루프는 UCUU, UUUU, UAUU, 또는 UGUU의 서열을 포함한다.

[0502] 일 측면에서, 가이드는 비-포스포디에스테르 결합을 통해 화학적으로 연결되거나 콘쥬게이션된 부분을 포함한다. 일 측면에서, 가이드는 비 제한적인 예에서, 비-뉴클레오티드 루프를 통해 화학적으로 연결되거나 콘쥬게이션된 추적자 서열 및 추적자 메이트 서열 부분 또는 직접 반복부 및 표적화 서열 부분을 포함한다.

일부 실시형태에서, 부분은 비-포스포디에스테르 공유 링커를 통해 연결된다. 화학적 링커의 예는 제한 없이, 카르바메이트, 에테르, 에스테르, 아미드, 이민, 아미드, 아미노트리진, 히드로존, 디설피드, 티오에테르, 티오에스테르, 포스포로티오에이트, 포스포로디티오에이트, 설폰아미드, 설포네이트, 풀폰, 설폭시드, 우레아, 티오우레아, 히드라지드, 옥심, 트리아졸, 광불안정성 결합, 딜스-알더 시클로-부가 쌍 또는 고리-폐쇄 복분해 쌍, 및 마이클 반응 쌍으로 이루어지는 군으로부터 선택되는 화학적 모이어티를 포함한다.

[0503] 일부 실시형태에서, 가이드의 일부는 먼저 표준 포스포르아미다이트 합성 프로토콜을 사용하여 합성된다. (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). 일부 실시형태에서, 비-표적화 가이드 부분은 당업계에 공지된 표준 프로토콜 (Hermanson, GT, Bioconjugate Techniques, Academic Press (2013))을 사용하여 라이게이션을 위한 적절한 작용기를 함유하도록 작용기화될 수 있다. 작용기의 예는 제한 없이, 히드록실, 아민, 카르복실산, 카르복실산 할라이드, 카르복실산 활성 에스테르, 알데히드, 카르보닐, 클로로카르보닐, 이미다졸릴카르보닐, 히드로지드, 세미카르바지드, 티오 세미카르바지드, 티올, 말레이미드, 할로알킬, 수포닐, 알릴, 프로파르길, 디엔, 알킨 및 아지드를 포함한다. 일단 가이드의 비-표적화 부분이 작용기화되면, 2 개의 올리고뉴클레오티드 사이에 공유 화학 결합 또는 연결이 형성될 수 있다. 화학적 결합의 예는 제한 없이, 카르바메이트, 에테르, 에스테르, 아미드, 이민, 아미드, 아미노트리진, 히드로존, 디설피드, 티오에테르, 티오에스테르, 포스포로티오에이트, 포스포로디티오에이트, 설폰아미드, 설포네이트, 풀폰, 설폭시드, 우레아, 티오우레아, 히드라지드, 옥심, 트리아졸, 광불안정성 결합, 딜스-알더 시클로-부가 쌍 또는 고리-폐쇄 복분해 쌍, 및 마이클 반응 쌍과 같은 C-C 결합 형성기를 기본으로 하는 것을 포함한다.

[0504] 일부 실시형태에서, 가이드의 하나 이상의 부분이 화학적으로 합성될 수 있다. 일부 실시형태에서, 화학 합성은 2'-아세톡시에틸 오르토에스테르 (2'-ACE)를 갖는 자동화된, 고상 올리고뉴클레오티드 합성 기계 (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-

18) 또는 2'-티오노카바메이트 (2'-TC) 화학 (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989)를 사용한다.

[0505] 일부 실시형태에서, 가이드 부분은 당, 뉴클레오티드 포스포디에스테르 결합, 퓨린 및 피리미딘 잔기의 변형을 통해 다양한 바이오콘쥬게이션 반응, 루프, 브릿지 및 비-뉴클레오티드 연결을 사용하여 공유적으로 연결될 수 있다. Sletten et al., Angew. Chem. Int. Ed. (2009) 48:6974-6998; Manoharan, M. Curr. Opin. Chem. Biol. (2004) 8: 570-9; Behlke et al., Oligonucleotides (2008) 18: 305-19; Watts, et al., Drug. Discov. Today (2008) 13: 842-55; Shukla, et al., ChemMedChem (2010) 5: 328-49.

[0506] 일부 실시형태에서, 가이드 부분은 클릭 화학을 사용하여 공유적으로 연결될 수 있다. 일부 실시형태에서, 가이드 부분은 트리아졸 링커를 사용하여 공유적으로 연결될 수 있다. 일부 실시형태에서, 가이드 부분은 알킨 및 아지드를 포함하는 Huisgen 1,3-이극성 시클로 부가 반응을 사용하여 공유적으로 연결되어 매우 안정한 트리아졸 링커를 생성할 수 있다 (He et al., ChemBioChem (2015) 17: 1809-1812; WO 2016/186745). 일부 실시형태에서, 가이드 부분은 5'-헥신 부분 및 3'-아지드 부분을 라이게이션시킴으로써 공유 연결된다. 일부 실시형태에서, 5'-헥신 가이드 부분 및 3'-아지드 가이드 부분 중 하나 또는 둘 모두는 2'-아세톡시에틸 오르토에스테르 (2'-ACE) 기로 보호될 수 있고, 이는 이후 Dharmacon 프로토콜을 사용하여 제거될 수 있다 (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18).

[0507] 일부 실시형태에서, 가이드 부분은 스페이서, 부착물, 바이오콘쥬게이트, 발색단, 리포터 그룹, 염료 표지된 RNA 및 비-천연 발생 뉴클레오티드 유사체와 같은 모이어티를 포함하는 링커 (예를 들어, 비-뉴클레오티드 루프)를 통해 공유 연결될 수 있다. 보다 구체적으로, 본 발명의 목적에 적합한 스페이서는 제한 없이, 폴리에테르 (예를 들어, 폴리에틸렌 글리콜, 폴리알코올, 폴리프로필렌 글리콜 또는 에틸렌 및 프로필렌 글리콜의 혼합물), 폴리아민기 (예를 들어, 스퀜닌, 스페르미딘 및 이의 중합체 유도체), 폴리에스테르 (예를 들어, 폴리(에틸 아크릴레이트)), 폴리포스포디에스테르, 알킬렌 및 이들의 조합을 포함한다. 적합한 부착물은, 예컨대 그러나 제한 없이 형광 표지와 같은 링커에 추가적인 특성을 추가하기 위해 링커에 첨가될 수 있는 임의의 모이어티를 포함할 수 있다. 적합한 바이오콘쥬게이트는 펩티드, 글리코시드, 지질, 콜레스테롤, 인지질, 디아실 글리세롤 및 디알킬 글리세롤, 지방산, 탄화수소, 효소 기질, 스테로이드, 비오틴, 디곡시제닌, 탄수화물, 다당류를 포함하지만 이에 제한되지는 않는다. 적합한 발색단, 리포터 그룹 및 염료-표지된 RNA는 형광 염료, 예컨대 플루오레세인 및 로다민, 화학발광, 전기화학발광 및 생물발광 마커 화합물을 포함하지만 이에 제한되지는 않는다. 2 개의 RNA 성분을 콘쥬게이션시키는 예시적인 링커의 디자인은 또한 WO 2004/015075에 기재되어 있다.

[0508] 링커 (예를 들어, 비-뉴클레오티드 루프)는 임의의 길이일 수 있다. 일부 실시형태에서, 링커는 약 0-16 뉴클레오티드와 동등한 길이를 갖는다. 일부 실시형태에서, 링커는 약 0-8 뉴클레오티드와 동등한 길이를 갖는다. 일부 실시형태에서, 링커는 약 0-4 뉴클레오티드와 동등한 길이를 갖는다. 일부 실시형태에서, 링커는 약 2 뉴클레오티드와 동등한 길이를 갖는다. 링커 디자인의 예는 또한 WO2011/008730에 기재되어 있다.

[0509] 일부 구현예에서, 상보성의 정도는 적합한 정렬 알고리즘을 사용하여 최적으로 정렬되는 경우, 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% 이상 또는 이를 초과한다. 최적의 정렬은 서열을 정렬하기에 적절한 임의의 알고리즘의 사용으로 결정될 수 있으며, 그의 비-제한적인 예는 스미스-워터만(Smith-Waterman) 알고리즘, 니들만-분쉬(Needleman-Wunsch) 알고리즘, 버로우즈-휠러 트랜스폼(Burrows-Wheeler Transform)에 기초한 알고리즘(예를 들어, 버로우즈 휠러 얼라이너(Burrows Wheeler Aligner)), ClustalW, Clustal X, BLAT, 노보얼라인(Novoalign)(노보크라프트 테크놀로지즈(Novocraft Technologies)(www.novocraft.com에서 이용가능함), ELAND(일루미나(Illumina), 미국 캘리포니아주 샌디에고), SOAP(soap.genomics.org.cn에서 이용가능) 및 Maq(maq.sourceforge.net에서 이용가능)를 포함한다. 표적 핵산 서열로의 핵산-표적화 복합체의 서열-특이적 결합을 지시하는 (RNA-표적화 가이드 RNA 또는 crRNA 내) 가이드 서열의 능력은 임의의 적합한 분석에 의해 평가될 수 있다. 예를 들어, 시험되는 가이드 서열을 포함하는 핵산-표적화 복합체를 형성하기에 충분한 RNA-표적화 CRISPR Cas13b 시스템의 성분은, 핵산-표적화 복합체의 성분을 인코딩하는 벡터를 이용한 트랜스펙션에 의해서와 같이 해당하는 표적 핵산 서열을 갖는 숙주 세포로 제공될 수 있으며, 본 명세서에 기재된 바와 같은 Surveyor 분석에 의해서와 같이, 표적 핵산 서열 내의 우선적인 표적화(예를 들어, 절단)의 평가로 이어질 수 있다. 유사하게, 표적 핵산 서열의 절단은 표적 핵산 서열, 즉 시험 가이드 서열 및 시험 가이드 서열과 상이한 대조군 가이드 서열을 포함하는 핵산-표적화 복합체의 성분을 제공하고, 표적 서열에서 시험 및 대조

군 가이드 서열 반응 간의 결합 또는 절단 속도를 비교함으로써 시험관에서 평가될 수 있다. 다른 어세이가 가능하며, 해당 분야의 숙련자에게 일어날 것이다. 가이드 서열, 따라서 RNA-표적화 가이드 RNA 또는 crRNA는 임의의 표적 핵산 서열을 표적화하도록 선택될 수 있다. 표적 서열은 DNA일 수 있다. 표적 서열은 임의의 RNA 서열일 수 있다. 일부 실시형태에서, 표적 서열은 메신저 RNA(mRNA), 프리-mRNA, 리보솜 RNA(rRNA), 운반 RNA(tRNA), 마이크로-RNA(miRNA), 작은 간섭 RNA(siRNA), 소형 핵 RNA(snRNA), 소형 핵소체 RNA(snoRNA), 이중 가닥화 RNA(dsRNA), 비-코딩 RNA(ncRNA), 긴 비-코딩 RNA(lncRNA), 및 소형 세포질 RNA(scRNA)로 이루어진 군으로부터 선택되는 RNA 분자 내의 서열일 수 있다. 일부 바람직한 실시형태에서, 표적 서열은 mRNA, 프리-mRNA, 및 rRNA로 이루어진 군으로부터 선택되는 RNA 분자 내의 서열일 수 있다. 일부 바람직한 실시형태에서, 표적 서열은 ncRNA, 및 lncRNA로 이루어진 군으로부터 선택되는 RNA 분자 내의 서열일 수 있다. 일부 더 바람직한 실시형태에서, 표적 서열은 mRNA 분자 또는 프리-mRNA 분자 내의 서열일 수 있다.

[0510] 일부 구현예에서, RNA-표적화 가이드 RNA 또는 crRNA는 RNA-표적화 가이드 RNA 또는 crRNA 내의 2차 구조 정도를 감소시키도록 선택된다. 일부 구현예에서, RNA-표적화 가이드 RNA의 뉴클레오티드의 약 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1% 이하 또는 이의 미만이 최적으로 폴딩된 경우 자기-상보성 염기 쌍형성에 참여한다. 최적 폴딩은 어떤 적합한 폴리뉴클레오티드 폴딩 알고리즘에 의해서 결정될 수 있다. 일부 프로그램은 최소 깁스 자유 에너지를 계산하는 것에 기초한다. 이러한 한 알고리즘의 예는 mFold이며, 이는 Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148) 에 기재된 바와 같다. 또 다른 폴딩 알고리즘의 예는 중심 구조 예측 알고리즘을 사용하는, 비엔나 대학의 이론 화학 연구소에서 개발한 온라인 웹서버 RNAfold이다. (참조, 예를 들어, A.R. Gruber et al., 2008, Cell 106(1): 23-24; 및 PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0511] 일정 실시형태에서, 가이드 RNA 또는 crRNA는 직접 반복부 (DR) 서열 및 가이드 서열 또는 스페이서 서열을 포함할 수 있거나, 그로 본질적으로 이루어질 수 있거나, 또는 그로 이루어질 수 있다. 일정 실시형태에서, 가이드 RNA 또는 crRNA는 가이드 서열 또는 스페이서 서열에 융합되거나 또는 연결된 직접 반복부 서열을 포함할 수 있거나, 그로 본질적으로 이루어질 수 있거나, 또는 그로 이루어질 수 있다. 일정 실시형태에서, 직접 반복부 서열은 가이드 서열 또는 스페이서 서열로부터 상류 (즉, 5')에 위치될 수 있다. 다른 실시형태에서, 직접 반복부 서열은 가이드 서열 또는 스페이서 서열로부터 하류 (즉, 3')에 위치될 수 있다. 다른 실시형태에서, 다수의 DR (예컨대, 이중 DR)이 존재할 수 있다.

[0512] 일정 실시형태에서, crRNA는 스템 루프, 바람직하게 단일 스템 루프를 포함할 수 있다. 일정 실시형태에서, 직접 반복부 서열은 스템 루프, 바람직하게 단일 스템 루프를 형성한다.

[0513] 일정 실시형태에서, 가이드 RNA의 스페이서 길이는 15 내지 35 nt이다. 일정 실시형태에서, 가이드 RNA의 스페이서 길이는 적어도 15 뉴클레오티드이다. 일정 실시형태에서, 스페이서 길이는 15 내지 17 nt, 예, 15, 16, 또는 17 nt,17 내지 20 nt, 예, 17, 18, 19, 또는 20 nt, 20 내지 24 nt, 예, 20, 21, 22, 23, 또는 24 nt, 23 내지 25 nt, 예를 들어, 23, 24, 또는 25 nt, 24 내지 27 nt, 예를 들어, 24, 25, 26, 또는 27 nt, 27-30 nt, 예를 들어, 27, 28, 29, 또는 30 nt, 30 내지 35 nt, 예를 들어, 30, 31, 32, 33, 34, 또는 35 nt, 또는 35 nt 또는 그 이상이다.

[0514] "tracrRNA" 서열 또는 유사체라는 용어는 혼성화를 위해 crRNA 서열과 충분한 상보성을 갖는 임의의 폴리뉴클레오티드 서열을 포함한다. 일반적으로, 상보성 정도는 2종의 서열의 보다 짧은 길이를 따라서, sca 서열 및 tracr 서열의 최적 정렬을 참조한다. 최적 정렬은 임의의 적합한 정렬 알고리즘에 의해 결정될 수 있고, 이차 구조, 예컨대 sca 서열 또는 tracr 서열 내 자가-상보성에 대해 더욱 설명될 수 있다. 일부 실시형태에서, 최적으로 정렬시 2종의 보다 짧은 것의 길이를 따라서 tracr 서열 및 sca 서열 간 상보성 정도는 약 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, 또는 그 이상이거나 또는 이를 초과한다. 일정 실시형태에서, tracrRNA는 요구되지 않을 수 있다. 실제로, 베르게이엘라 주헬쿰 및 이의 동족체 유래의 Cas13b 이펙터 단백질은 RNA 표적의 절단을 보장하기 위해 tracrRNA를 요구하지 않는다.

[0515] 보다 상세하게, PAM 서열이 인식을 유도하는데 필요하다면, RNA 표적에 대한 어세이는 다음과 같다. 2 가지이. 콜라이 균주가 이 분석에서 사용된다. 하나는 박테리아 균주 유래의 내인성 이펙터 단백질 유전자좌를 인코딩하는 플라스미드를 보유한다. 나머지 균주는 빈 플라스미드(예를 들어, pACYC184, 대조 균주)를 보유한다. 가능한 모든 7 bp 또는 8 bp PAM 서열을 항생제 저항성 플라스미드(앰피실린 저항성 유전자를 갖는 pUC19) 상에 제시한다. PAM은 프로토스페이서 1의 서열(내인성 이펙터 단백질 유전자좌에서 제1 스페이서에

대한 RNA 표적) 옆에 위치한다. 2 개의 PAM 라이브러리를 클로닝하였다. 하나는 프로토-스페이서의 8 랜덤 bp 5'를 갖는다(예를 들어, 총 65536 개의 상이한 PAM 서열 = 복잡성). 나머지 라이브러리는 프로토-스페이서의 7 랜덤 bp 3'를 갖는다(예를 들어, 총 복잡성은 16384 개의 상이한 PAM임). 2 가지 라이브러리 모두 가능한 PAM 당 평균 500 개의 플라스미드를 갖도록 클로닝되었다. 시험 균주 및 대조 균주를 별도의 형질전환에서 5'PAM 및 3'PAM 라이브러리로 형질전환하고, 형질전환된 세포를 앰피실린 플레이트 상에 별도로 도말하였다. 플라스미드를 이용한 인식 및 후속 절단/간섭은 세포를 앰피실린에 취약하게 만들고 성장을 방지한다. 대략 형질전환 12 시간 후, 시험 및 대조 균주에 의해 형성된 모든 콜로니를 수확하고 플라스미드 RNA를 단리하였다. 플라스미드 RNA를 PCR 증폭 및 후속 심층 서열분석을 위한 주형으로서 사용하였다. 비형질전환 라이브러리에서 모든 PAM의 표현은 형질전환된 세포에서 PAM의 예상된 표현을 나타내었다. 대조 균주에서 발견된 모든 PAM의 표현은 실제 표현을 나타내었다. 시험 균주에서 모든 PAM의 표현은 어떤 PAM이 효소에 의해 인식되지 않는지를 나타내었고 대조 균주와의 비교는 감손된 PAM의 서열을 추출할 수 있게 한다.

특정 실시형태에서, 절단, 예컨대 RNA 절단은 PAM 의존적이지 않다. 실제로, *Bergeyella Zoohelcum* 이펙터 단백질 및 그의 오르소로그에서, RNA 표적 절단은 PAM 독립적인 것으로 보이므로, 본 발명의 표 1 Cas13b는 PAM 독립적인 방식으로 작용할 수 있다.

[0516] 독성 및 표적외 효과의 최소화를 위해, 전달되는 RNA-표적화 가이드 RNA의 농도를 제어하는 것이 중요할 것이다. 핵산-표적화 가이드 RNA의 최적 농도는 세포 또는 비-인간 진핵 생물 동물 모델에서 상이한 농도를 시험함으로써, 그리고 잠재적인 표적외 게놈 유전자좌에서의 변형 정도를 분석하는 심층 서열분석을 사용함으로써 결정될 수 있다. 표적외 변형 수준을 최소화하면서 표적상 변형의 가장 높은 수준을 제공하는 농도는 생체 내 전달을 위해 선택되어야 한다. RNA-표적화 시스템은 유리하게 CRISPR-Cas13b 시스템으로부터 유래된다. 일부 구현예에서, RNA-표적화 시스템의 하나 이상의 엘리먼트는 본원에 기재된 표 1 Cas13b 이펙터 단백질 시스템의 내인성 RNA-표적화 시스템을 포함하는 특정 유기체로부터 유래한다.

[0517] 용어 "오르소로그(orthologue)"(또한 본 명세서에서 "오르소로그(ortholog)"로 지칭됨) 및 "동족체(homologue)"(또한 본 명세서에서 "동족체(homolog)"로 지칭됨)는 당업계에 잘 공지되어 있다. 추가 지침에 의해, 본 명세서에서 사용된 바와 같은 단백질의 "동족체"는 이의 동족체인 단백질과 동일 또는 유사한 기능을 실행하는 동일 종의 단백질이다. 상동성 단백질은 구조적으로 관련될 수 있지만 관련될 필요가 없거나, 부분적으로만 구조적으로 관련된다. 본 명세서에서 사용된 바와 같은 단백질의 "오르소로그"는 이의 오르소로그인 단백질과 동일 또는 유사한 기능을 실행하는 상이한 종의 단백질이다. 오르소로그 단백질은 구조적으로 관련될 수 있지만 관련될 필요가 없거나, 부분적으로만 구조적으로 관련된다. 특정 구현예에서, 본 명세서에 언급된 바와 같은 Cas13b 단백질의 동족체 또는 오르소로그는 표 1에 언급된 Cas13b 이펙터 단백질과 적어도 80%, 더 바람직하게는 적어도 85%, 훨씬 더 바람직하게는 적어도 90%, 예컨대 적어도 95%의 서열 상동성 또는 동일성을 갖는다. 바람직한 실시형태에서, Cas13b 이펙터 단백질은 표 1에서 확인된 유기체 또는 유기체가 속하는 속일 수 있다.

[0518] 다수의 Cas13b 오르소로그가 공통 모티프를 특징으로 하는 것으로 밝혀졌다. 따라서, 특정 실시형태에서, Cas13b 이펙터 단백질은 DKHXFGAFLNLARHN (SEQ ID NO:XX), GLLFFVSLFLDK (SEQ ID NO:XX), SKIXGFK (SEQ ID NO:XX), DMLNELXRCP (SEQ ID NO:XX), RXZDRFPYFALRYXD (SEQ ID NO: XX) 및 LRFQVBLGXY (SEQ ID NO:XX)로 이루어진 서열 하나 이상과 적어도 7% 서열 동일성을 갖는 서열을 포함하는 단백질이다. 추가의 특정 실시형태에서, Cas13b 이펙터 단백질은 이들 서열 중 적어도 2, 3, 4, 5 또는 모든 6 개에서 적어도 70% 서열 동일성을 갖는 서열을 포함한다. 추가의 특정 실시형태에서, 이들 서열과 서열 동일성은 적어도 75%, 80%, 85%, 90%, 95% 또는 100%이다. 추가의 특정 실시형태에서, Cas13b 이펙터 단백질은 GLLFFVSLFL (SEQ ID NO:XX) 및 RHQXRFPYF (SEQ ID NO:XX)와 100% 서열 동일성을 갖는 서열을 포함하는 단백질이다. 추가의 특정 실시형태에서, Cas13b 이펙터는 RHQDRFPY (SEQ ID NO:XX)와 100% 서열 동일성을 갖는 서열을 포함하는 Cas13b 이펙터 단백질이다.

[0519] 특정 실시형태에서, Cas13b 이펙터 단백질은 *Prevotella buccae, Porphyromonas gingivales, Prevotella saccharolytica, Riemerella antipestifer*로부터의 Cas13b 단백질과 적어도 65%, 바람직하게는 적어도 70%, 75%, 80%, 85%, 90%, 95% 이상의 서열 동일성을 갖는 Cas13b 이펙터 단백질이다. 추가의 특정 실시형태에서, Cas13b 이펙터는 *Bacteroides pyogenes*, *Prevotella sp. MA2016*, *Riemerella anatipestifer, Porphyromonas gulae*, *Porphyromonas gingivalis*, 및 *Porphyromonas sp.COT-052OH4946*로부터의 Cas13b 단백질로부터 선택된다.

[0520] 본 발명 내에 존재할 수 있는 표 1 Cas13b 효소의 오르소로그는 표 1 Cas13b 효소의 다중 오르소로그의 단편을

포함하는 키메라 효소를 포함할 수 있음을 이해할 것이다. 이러한 오르소로그의 예는 본 명세서의 다른 곳에 기재되어 있다. 키메라 효소는 표 1 Cas13b 효소의 단편 및 다른 CRISPR 효소의 단편, 예를 들어 하기를 포함하나 이에 제한되지 않는 유기체의 표 1 Cas13b 효소의 오르소로그를 포함할 수 있다: *Bergeyella, Prevotella, Porphyromonas, Bacteroides, Alistipes, Riemerella, Myroides, Flavobacterium, Capnocytophaga, Chryseobacterium, Phaeodactylibacter, Paludibacter* 또는 *Psychroflexus*. 키메라 효소는 제 1 단편 및 제 2 단편을 포함할 수 있으며, 상기 단편은 제 1 및 제 2 단편 중 하나가 표 1 Cas13b 효소이거나 또는 그 유래의 것이고 다른 단편은 상이한 종의 CRISPR 효소 오르소로그이거나 또는 그 유래의 것이다.

[0521] 실시형태에서, 본 명세서에서 언급하는 Cas13b RNA-표적화 Cas13b 이펙터 단백질은 또한 이펙터 단백질의 기능성 변이체 또는 이의 상동체 또는 오르소로그를 포함한다. 본 명세서에서 사용된 바와 같은 단백질의 "기능적 변이체"는 그 단백질의 활성을 적어도 부분적으로 보유하는 단백질의 변이체를 지칭한다. 기능적 변이체는 돌연변이체(이는 삽입, 결실, 또는 치환 돌연변이체일 수 있음)를 포함할 수 있으며, 이는 예를 들어 표 1과 조합하여 본원에 논의된 것을 포함하는 다형체 등을 포함한다. 또한, 기능적 변이체 내에는 또 다른, 보통 관련되지 않은, 핵산, 단백질, 폴리펩티드 또는 펩티드와 이러한 단백질의 융합 생성물이 포함된다. 기능적 변이체는 천연적으로 발생할 수 있거나 인공적일 수 있다. 일 실시형태에서, Cas13b RNA-표적화 이펙터 단백질, 또는 이의 오르소로그 또는 상동체를 인코딩하는 핵산 분자(들)는 진핵생물 세포에서 발현을 위해 코돈 최적화될 수 있다. 진핵생물은 본 명세서에 논의된 바와 같을 수 있다. 핵산 분자(들)는 조작되거나 비-천연적으로 발생할 수 있다.

[0522] 일 실시형태에서, Cas13b RNA-표적화 이펙터 단백질 또는 이의 오르소로그 또는 상동체는 하나 이상의 돌연변이를 포함할 수 있다. 돌연변이는 인위적으로 도입된 돌연변이일 수 있으며, 예를 들어, 촉매 도메인에 하나 이상의 돌연변이를 포함할 수 있으나 이에 제한되지 않고, 예를 들어, 하나 이상의 돌연변이가 하나 이상의 HEPN 도메인 내에 도입된다,

[0523] 일 실시형태에서, Cas13b 단백질 또는 이의 오르소로그 또는 상동체는 기능성 도메인과 융합되거나 또는 작동적으로 연결되어 제네릭 핵산 결합 단백질로서 사용될 수 있다. 예시적인 작용 도메인은 이로 제한되지는 않지만, 번역 개시제, 번역 활성화제, 번역 억제제, 뉴클레아제, 특히 리보뉴클레아제, 스플라이세오좀, 비즈, 광유도성/제어성 도메인 또는 화학적으로 유도성/제어성 도메인을 포함할 수 있다.

[0524] 일부 실시형태에서, 비-변형된 RNA-표적화 이펙터 단백질(Cas13b)은 절단 활성을 가질 수 있다. 일부 실시형태에서, Cas13b는 표적 서열 내 및/또는 표적 서열의 상보체 내 또는 표적 서열과 연관된 서열에서, 예를 들어, 표적 서열의 처음 또는 마지막 뉴클레오티드로부터 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, 또는 그 이상의 염기쌍 이내에서 표적 서열의 위치 또는 근처에서 1 또는 2 개의 핵산 가닥의 직접 절단을 유도할 수 있다. 일부 실시형태에서, Cas13b 단백질은 표적 서열 내 및/또는 표적 서열의 상보체 내 또는 표적 서열과 연관된 서열에서 및/또는 표적 서열의 처음 또는 마지막 뉴클레오티드로부터 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, 또는 그 이상의 염기쌍 이내에서 가닥 중 1 또는 2 개 가닥의 하나 이상의 절단 (예컨대, 1, 2, 3, 4, 5, 또는 그 이상의 절단)을 유도할 수 있다. 일부 실시형태에서, 절단은 평활할 수 있고, 즉 평활 말단을 생성시킬 수 있다. 일부 실시형태에서, 절단은 엇갈릴 수 있고, 즉 점착 말단을 생성시킬 수 있다. 일부 실시형태에서, 벡터는 상응하는 야생형 효소에 대해, 돌연변이된 핵산-표적화 Cas13b 단백질이 표적 서열을 함유하는 표적 폴리뉴클레오티드의 하나 또는 두 가닥을 절단하는 능력이 결여되도록 돌연변이될 수 있는 (예를 들어, 모든 RNA 절단 활성이 실질적으로 결여된 돌연변이된 Cas13b를 생성하기 위한 HEPN 도메인 내의 변형 또는 돌연변이) 핵산-표적화 Cas13b 단백질을 인코딩하여, 예를 들어, 돌연변이된 효소의 RNA 절단 활성은 약 25%, 10%, 5%, 1%, 0.1%, 0.01% 이하의, 효소의 비-돌연변이된 형태의 핵산 절단 활성이고; 그 예는돌연변이된 형태의 핵산 절단 활성이 비-돌연변이된 형태와 비교하여 0이거나 무시할 수 있는 경우일 수 있다. "유래"란, 출원인은 유래된 효소가 야생형 효소와 높은 정도의 서열 상동성을 갖는 의미에서 크게 기반으로 하지만, 당업계에 공지되거나 본 명세서에 기재된 바와 같은 일부 방법으로 돌연변이된(변형된) 것을 의미한다.

[0525] 통상적으로, 내인성 RNA-표적화 시스템의 맥락에서, RNA-표적화 복합체(표적 서열에 혼성화되고, 하나 이상의 RNA-표적화 이펙터 단백질과 복합체화된 가이드 RNA 또는 crRNA를 포함함)의 형성은 표적 서열 내 또는 가까이에서(예를 들어, 이로부터 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 개 이상의 염기쌍) RNA 가닥(들)의 절단을 초래한다. 본 명세서에서 사용된 바와 같이, 용어 "관심 표적 유전자와 연관된 서열(들)"은 표적 서열의 부근 가까이의(예를 들어, 표적 서열로부터 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 개 이상의 염기쌍 내, 여

기서 표적 서열은 관심 표적 유전자좌 내에 포함됨) 서열을 지칭한다.

[0526] 코돈 표적화된 서열의 예는 이러한 예에서, 진핵생물, 예를 들어 인간 (즉, 인간에서 발현을 위해 최적화됨)에서, 또는 다른 진핵생물의 경우, 본 명세서에 기술된 바와 같은 동물 또는 포유동물에서 발현을 위해 최적화된 서열이고, 예를 들어, 코돈 최적화된 서열의 예로서, WO 2014/093622 (PCT/US2013/074667)의 SaCas9 인간 코돈 최적화된 서열 (당분야 및 본 개시의 지식으로부터, 특히 이펙터 단백질 (예를 들어, Cas13b)에 대한, 코돈 최적화되는 코딩 핵산 분자(들)는 당업자의 영역에 속함)을 참조한다. 이것이 바람직하지만, 기타 다른 예들도 가능하며, 인간 이외의 숙주 종에 대한 코돈 최적화, 또는 특정 기관에 대한 코돈 최적화가 공지되어 있다는 것이 이해될 것이다. 일부 구현예에서, RNA-표적화 Cas13b 단백질을 인코딩하는 효소 코딩 서열은 진핵 세포와 같은 특정 세포에서의 발현에 대해 코돈 최적화된다. 진핵생물 세포는 제한없이 인간, 또는 인간이외의 진핵생물 또는 본 명세서에 기술된 바와 같은 동물 또는 포유동물, 예를 들어 마우스, 래트, 토끼, 개, 가축 또는 인간이외의 포유동물 또는 영장류를 포함하는 특정 유기체의 것일 수 있거나 또는 그로부터 유래될 수 있다. 일부 구현예에서, 인간이나 동물에 임의의 실질적인 의학적 이익 없이 그들을 고통스럽게 할 수 있는 인간의 생식계열 유전자 아이덴티티를 변형하기 위한 과정 및/또는 동물의 유전자 아이덴티티를 변형하기 위한 과정과, 또한 이러한 과정으로부터 생긴 동물이 배제될 수 있다. 일반적으로, 코돈 최적화는 고유 서열의 적어도 하나의 코돈(예를 들어, 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 개 이상의 코돈)을 숙주 세포의 유전자에 더욱 빈번하게 또는 가장 빈번하게 사용되는 코돈으로 대체하면서, 고유 아미노산 서열을 유지함으로써 대상 숙주 세포에서의 발현의 증진을 위해 핵산 서열을 변형시키는 과정을 지칭한다. 다양한 종은 특정 아미노산의 특정 코돈에 대하여 특정 편향을 나타낸다. 코돈 편향(유기체 간의 코돈 사용의 차이)은 종종 메신저 RNA(mRNA)의 번역의 효율과 상관관계가 있으며, 이는 결국, 특히, 번역되는 코돈의 특성 및 특정 운반 RNA(tRNA) 분자의 이용 가능성에 좌우되는 것으로 여겨진다. 세포에서의 선택된 tRNA의 우세는 일반적으로 펩티드 합성에 가장 빈번하게 사용되는 코돈을 반영하는 것이다. 따라서, 유전자는 코돈 최적화를 기반으로 하여 주어진 유기체에서의 최적의 유전자 발현을 위해 맞춤화될 수 있다. 코돈 사용 표는, 예를 들어 www.kazusa.orjp/codon/에서 이용 가능한 "코돈 사용 데이터베이스"에서 용이하게 이용 가능하며, 이 표는 다수의 방식으로 적합하게 될 수 있다. 문헌[Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence databases:status for the year 2000” Nucl. Acids Res. 28:292 (2000)]을 참조한다. 특정 숙주 세포에서의 발현을 위해 특정 서열을 코돈 최적화시키는 컴퓨터 알고리즘이 또한 이용 가능하며, 예를 들어 Gene Forge (Aptagen; Jacobus, PA) 이 또한 이용가능하다. 일부 구현예에서, DNA/RNA-표적화 Cas 단백질을 인코딩하는 서열 내의 하나 이상의 코돈(예를 들어, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 개 이상의, 또는 모든 코돈)은 특정 아미노산에 대하여 가장 빈번하게 사용되는 코돈에 해당한다.

[0527] (i) Cas13b 또는 이를 인코딩하는 핵산 분자(들) 또는 (ii) crRNA는 별도로 전달될 수 있으며; 유리하게는 (i) 및 (ii) 중 하나 이상 또는 모두, 예를 들어 조립된 복합체는 입자 또는 나노입자 복합체를 통해 전달된다. RNA-표적화 이펙터 단백질 mRNA는 핵산-표적화 이펙터 단백질이 발현될 시간을 제공하기 위하여 RNA-표적화 가이드 RNA 또는 crRNA 이전에 전달될 수 있다. RNA-표적화 이펙터 단백질 (Cas13b) mRNA는 RNA-표적화 가이드 RNA 또는 crRNA 투여 1 시간 내지 12 시간(바람직하게는, 약 2 시간 내지 6 시간) 전에 투여될 수 있다. 대안적으로, RNA-표적화 이펙터 단백질 mRNA 및 RNA-표적화 가이드 RNA 또는 crRNA는 함께 투여될 수 있다. 유리하게, 가이드 RNA 또는 crRNA의 제2 부스터 용량은 RNA-표적화 이펙터 (Cas13b) 단백질 mRNA + 가이드 RNA의 초기 투여 1 시간 내지 12 시간(바람직하게는, 약 2 시간 내지 6 시간) 이후에 투여될 수 있다. RNA-표적화 이펙터 단백질 mRNA 및/또는 가이드 RNA 또는 crRNA의 추가적인 투여는 게놈 변형의 가장 효율적인 수준을 달성하는 데 유용할 수 있다.

[0528] 일 양태에서, 본 발명은 RNA-표적화 시스템의 하나 이상의 엘리먼트를 사용하는 방법을 제공한다. 본 발명의 RNA-표적화 복합체는 표적 RNA 단일 또는 이중 가닥, 선형 또는 수퍼코일을 변형시키기 위한 효율적인 수단을 제공한다. 본 발명의 RNA-표적화 복합체는 다수의 세포 유형에서 표적 RNA를 변형시키는(예를 들어, 결실시키는, 삽입하는, 전위시키는, 불활성화시키는, 활성화시키는) 것을 포함하는 광범위한 다양한 유용성을 갖는다. 이와 같이, 본 발명의 RNA-표적화 복합체는, 예를 들어 유전자 치료법, 약물 스크리닝, 질병 진단, 및 예후에서 광범위한 적용을 갖는다. 예시적인 RNA-표적화 복합체는 관심 표적 유전자좌 내의 표적 서열에 혼성화되는 가이드 RNA 또는 crRNA와 복합체화되는 RNA-표적화 이펙터 단백질을 포함한다.

[0529] 일 구현예에서, 본 발명은 표적 RNA를 절단하는 방법을 제공한다. 방법은 표적 RNA와 결합하여 상기 표적 RNA의 절단을 행하는 RNA-표적화 복합체를 사용하여 표적 RNA를 변형하는 단계를 포함한다. 일 실시형태에서, 본 발명의 RNA-표적화 복합체는 세포 내로 도입된 경우, RNA 서열에서 파손(예를 들어, 단일 또는 이중 가

닥 파손)을 생성할 수 있다. 예를 들어, 본 방법은 세포 내에서 질병 RNA를 절단하는 데 사용될 수 있다.

예를 들어, 상류 서열 및 하류 서열에 의해 측접된 통합될 서열을 포함하는 외인성 RNA 주형은 세포 내로 도입될 수 있다. 상류 및 하류 서열은 RNA 내 통합 부위의 어느 한 측면과의 서열 유사성을 공유한다. 원하는 경우, 공여자 RNA는 mRNA일 수 있다. 외생의 RNA 주형은 통합될 서열(예를 들어, 돌연변이된 RNA)을 포함한다. 통합을 위한 서열은 세포에 대해 내생 또는 외인성 서열일 수 있다. 통합될 서열의 예로는 단백질 또는 비-코딩 RNA(예를 들어, 마이크로RNA)를 인코딩하는 RNA를 포함한다. 따라서, 통합을 위한 서열은 적절한 제어 서열 또는 서열들에 작동 가능하게 연결될 수 있다. 대안적으로, 통합될 서열은 조절 작용을 제공할 수 있다. 외생의 RNA 주형 내 상류 및 하류 서열이 선택되어 관심 RNA 서열과 공여자 RNA 사이의 재조합을 촉진한다. 상류 서열은 통합을 위해 표적된 부위의 RNA 서열 상류와의 서열 유사성을 공유하는 RNA 서열이다. 유사하게, 하류 서열은 통합을 위한 표적된 부위의 하류의 RNA 서열과 서열 유사성을 공유하는 RNA 서열이다. 외인성 RNA 주형의 상류 및 하류 서열은 표적화된 RNA 서열과 75%, 80%, 85%, 90%, 95%, 또는 100% 서열 동일성을 가질 수 있다. 바람직하게, 외인성 RNA 주형의 상류 및 하류 서열은 표적화된 RNA 서열과 약 95%, 96%, 97%, 98%, 99%, 또는 100% 서열 동일성을 가진다. 일부 방법에서, 외인성 RNA 주형의 상류 및 하류 서열은 표적화된 RNA 서열과 약 99% 또는 100% 서열 동일성을 가진다. 상류 또는 하류 서열은 약 20 bp 내지 약 2500 bp, 예를 들어, 약 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 또는 2500 bp를 포함할 수 있다. 일부 방법에서, 예시적인 상류 또는 하류 서열은 약 200 bp 내지 약 2000 bp, 약 600 bp 내지 약 1000 bp, 또는 더욱 구체적으로는 약 700 bp 내지 약 1000 bp를 갖는다. 일부 방법에서, 외인성 RNA 주형은 마커를 추가로 포함할 수 있다. 그러한 마커는 표적된 통합을 스크리닝하기 쉽게 만들 수 있다.

적합한 마커의 예로는 제한 부위, 형광 단백질 또는 선택성 마커를 포함한다. 본 발명의 외인성 RNA 주형은 재조합 기술을 이용하여 구축될 수 있다(예를 들어, 문헌 [Sambrook et al., 2001 및 Ausubel et al., 1996] 참조) 외생의 RNA 주형을 통합함으로써 표적 RNA를 변형하기 위한 예시적인 방법에서, 파손 (예를 들어, 이중 또는 단일 가닥 RNA 내의 이중 또는 단일 가닥 파손)은 핵산-표적화 복합체에 의해 RNA 서열 내로 도입되고, 그 파손은 그 주형이 RNA 표적 내로 통합되도록 하는 외인성 RNA 주형을 이용한 상동성 재조합을 통해 수복된다. 이중가닥 브레이크의 존재는 주형의 통합을 촉진한다. 다른 실시형태에서, 본 발명은 진핵생물 세포에서 RNA의 발현을 변형시키는 방법을 제공한다. 방법은 DNA 또는 RNA (예를 들어, mRNA 또는 프리-mRNA)에 결합하는 핵산-표적화 복합체를 사용하여 표적 폴리뉴클레오티드의 발현을 감소시키거나 또는 증가시키는 것을 포함한다. 일부 방법에서, 표적 RNA는 세포에서 발현의 변형을 실시하기 위해 불활성화될 수 있다.

예를 들어, 세포 내 표적 서열에 RNA-표적화 복합체의 결합시, 표적 RNA는 불활성화되어서 서열이 번역되지 않고, 코딩된 단백질은 생산되지 않거나, 또는 서열은 야생형 서열이 하는 것처럼 기능하지 않는다. 예를 들어, 단백질 또는 마이크로RNA 코딩 서열은 불활성화될 수 있어서 단백질 또는 마이크로RNA 또는 프리-마이크로RNA 전사물이 생성되지 않는다. RNA-표적화 복합체의 표적 RNA는 진핵생물 세포에 내인성 또는 외인성인 임의의 RNA일 수 있다. 예를 들어, 표적 RNA는 진핵생물 세포의 핵에 존재하는 RNA일 수 있다. 표적 RNA는 유전자 산물 (예를 들어, 단백질)을 코딩하는 서열 (예를 들어, mRNA 또는 프리-mRNA)이거나 또는 논-코딩 서열 (예를 들어, ncRNA, lncRNA, tRNA, 또는 rRNA)일 수 있다. 표적 RNA의 예는 신호전달 생화학적 경로와 연관된 서열, 예를 들어 신호전달 생화학적 경로-연관 RNA를 포함한다. 표적 RNA의 예는 질환 연관 RNA를 포함한다. "질환-연관" RNA는 비질환 대조군의 조직 또는 세포와 비교하여 질환 영향받은 조직에서 유래하는 세포에서 비정상적인 형태 또는 비정상적인 수준으로 변역 생성물을 산출하는 임의의 RNA를 의미한다.

이는 비정상적으로 고 수준으로 발현되는 유전자로부터 전사되는 RNA일 수 있고, 비정상적으로 낮은 수준으로 발현되는 유전자로부터 전사되는 RNA일 수 있는데, 여기서 변경된 발현은 질환의 발생 및/또는 진행과 관련된다. 질환-연관된 RNA는 또한 질환의 병인론의 원인인 유전자(들)과 연관 불균형이거나 또는 직접적으로 원인인 유전자 프로세싱 돌연변이(들) 또는 유전자 변이로부터 전사되는 RNA를 의미한다. 번역된 생성물은 기지이거나 또는 미지일 수 있고, 정상 또는 비정상 수준일 수 있다. RNA-표적화 복합체의 표적 RNA는 진핵생물 세포에 내인성 또는 외인성인 임의의 RNA일 수 있다. 예를 들어, 표적 RNA는 진핵생물 세포의 핵에 존재하는 RNA일 수 있다. 표적 RNA는 융전자 생성물 (예를 들어, 단백질)을 코딩하는 서열 (예를 들어, mRNA 또는 프리-mRNA)이거나 또는 논-코딩 서열 (예를 들어, ncRNA, lncRNA, tRNA, 또는 rRNA)일 수 있다.

[0530] 일부 실시형태에서, 방법은 RNA-표적화 복합체가 표적 RNA에 결합하여 상기 표적 RNA의 절단을 실시하도록 하여 표적 RNA를 변형시키는 단계를 포함하고, 여기서 RNA-표적화 복합체는 상기 표적 RNA 내 표적 서열과 혼성화된 가이드 RNA 또는 crRNA와 복합체 형성하는 핵산-표적화 이펙터 (Cas13b) 단백질을 포함한다. 일 양태에서, 본 발명은 진핵 세포에서 RNA의 발현을 변형하는 방법을 제공한다. 일부 실시형태에서, 방법은 RNA-표적화

복합체가 RNA에 결합하도록 하여 상기 결합 결과로 상기 RNA의 증가되거나 또는 감소된 발현이 일어나는 단계를 포함하고; 여기서 RNA-표적화 복합체는 가이드 RNA 복합체형성된 핵산-표적화 이펙터 (Cas13b) 단백질을 포함한다. 표적 RNA를 변형시키는 방법은 생체내, 생체외 또는 시험관내일 수 있는, 진핵생물 세포에서 행해질 수 있다. 일부 구현예에서, 본 방법은 인간 또는 비-인간 동물로부터 세포 또는 세포 집단을 샘플링하는 단계 및 세포 또는 세포들을 변형하는 단계를 포함한다. 배양은 생체 외 임의의 단계에서 발생할 수 있다. 세포 또는 세포들은 비-인간 동물 또는 식물 내로 재도입될 수도 있다. 재도입된 세포의 경우, 세포는 줄기 세포인 것이 바람직할 수 있다.

[0531] 2 개의 상이한 앱타머(각각 별개의 RNA-표적화 가이드 RNA)의 사용은 상이한 RNA-표적화 가이드 RNA 또는 crRNA를 이용하여, 사용될 활성화제-어댑터 단백질 융합 및 억제제-어댑터 단백질 융합이, 다른 하나는 억제하는 한편 RNA의 발현은 활성화시키는 것을 가능하게 한다. 이들은 그들의 상이한 가이드 RNA 또는 crRNA와 함께, 다중화된 접근법에서 함께, 또는 실질적으로 함께 투여될 수 있다. 이러한 변형된 RNA-표적화 가이드 RNA 또는 crRNA의 다수, 예를 들어 10 개 또는 20 개 또는 30 개 등이 모두 동시에 사용될 수 있고, 한편 비교적 적은 수의 이펙터 단백질 (Cas13b) 분자로서, 전달되는 이펙터 단백질 분자 중 단지 하나(또는 적어도 최소 수)가 다수의 변형된 가이드와 함께 사용될 수 있다. 어댑터 단백질은 하나 이상의 활성화제 또는 하나 이상의 억제제에 연관(바람직하게는 연결 또는 융합)될 수 있다. 예를 들어, 어댑터 단백질은 제1 활성화제 및 제2 활성화제와 연관될 수 있다. 제1 및 제2 활성화제는 동일할 수 있지만, 이들은 바람직하게 상이한 활성화제이다. 3 개 이상 또는 심지어 4 개 이상의 활성화제(또는 억제제)가 사용될 수 있지만, 패키지 크기는 5 개를 초과하는 개수의 상이한 기능성 도메인은 제한할 수 있다. 어댑터 단백질에 대한 직접 융합에 대하여, 링커가 바람직하게 사용되며, 여기서 2 개 이상의 기능성 도메인은 어댑터 단백질에 연결된다. 적합한 링커는 GlySer 링커를 포함할 수 있다.

[0532] 또한 RNA-표적화 이펙터 단백질-가이드 RNA 복합체는 전체로서 둘 이상의 기능성 도메인과 연합될 수 있음이 고려된다. 예를 들어, RNA-표적화 이펙터 단백질과 연합된 둘 이상의 기능성 도메인이 존재할 수 있거나, (하나 이상의 어댑터 단백질을 통해) 가이드 RNA 또는 crRNA와 연합된 둘 이상의 기능성 도메인이 존재할 수 있거나, (하나 이상의 어댑터 단백질을 통해) RNA-표적화 이펙터 단백질과 연합된 하나 이상의 기능성 도메인 및 가이드 RNA 또는 crRNA와 연합된 하나 이상의 기능성 도메인이 존재할 수 있다.

[0533] 어댑터 단백질과 활성화제 또는 억제제 사이의 융합은 링커를 포함할 수 있다. 예를 들어, GlySer 링커 GGGS가 사용될 수 있다. 이들은 3($(GGGGS)_3$) 또는 6, 9 또는 심지어는 12 이상의 반복부로 사용되어, 필요에 따라 적합한 길이를 제공할 수 있다. 링커는 가이드 RNA와 기능성 도메인 (활성인자 또는 억제인자) 사이에 또는 핵산-표적화 이펙터 단백질 및 기능성 도메인 (활성인자 또는 억제인자) 사이에 사용될 수 있다. 사용자는 링커를 "기계적 유연성"의 적절한 양으로 조작한다.

[0534] **<u>Cas13b 이펙터 단백질 복합체는 작용 이펙터를 전달할 수 있다</u>**

[0535] RNA 수준에서 돌연변이에 의해 발현을 제거하는 CRISPR-Cas13b-매개 녹아웃과 달리, CRISPR-Cas13b 녹다운은 인공 전사 인자의 사용을 통한, 예를 들어, 촉매적으로 불활성인 Cas13b 단백질을 생성하는 Cas13b 단백질의 절단 도메인(들) 내의 잔기 돌연변이를 통한 유전자 발현의 일시적 감소를 허용한다. 촉매적으로 비활성인 Cas13b은 가이드 RNA 또는 crRNA와 복합체화하고, 가이드 RNA 또는 crRNA의 표적 도메인에 의해 특정된 RNA 서열에 국재화되지만, 이는 표적을 절단하지 않는다. 불활성 Cas13b 단백질의 이펙터 도메인, 예를 들어 전사 억제 도메인에 대한 융합은 가이드 RNA에 의해 특정되는 임의의 부위로 이펙터의 동원을 가능하게 한다.

[0536] ***최적화된 기능성 RNA 표적화 시스템***

[0537] 일 측면에서 따라서 본 발명은 RNA 환경으로 기능성 성분의 특이적 전달을 위한 시스템을 제공하다. 이는 RNA로 상이한 성분의 특이적 표적화를 가능하게 하는 본 발명의 RNA 표적화 이펙터 단백질 (표 1 Cas13b) 을 포함하는 CRISPR 시스템을 사용하여 보장된다. 보다 특히 이러한 성분은 활성인자 또는 억제인자, 예컨대 RNA 번역, 분해 등의 활성인자 또는 억제인자를 포함한다.

[0538] 일 측면에 따라서, 본 발명은 세포 내에서 관심 게놈 유전자좌 내 표적 서열과 혼성화할 수 있는 가이드 서열을 포함하는 가이드 RNA 또는 crRNA를 포함하는 비-천연 발생 또는 조작된 조성물을 제공하고, 가이드 RNA 또는 crRNA는 어댑터 단백질에 결합하는 하나 이상의 개별 RNA 서열 (들)의 삽입에 의해 변형된다. 특정 실시형태에서, RNA 서열은 둘 이상의 어댑터 단백질 (예를 들어, 앱타머)에 결합할 수 있고, 각 어댑터 단백질은 하나 이상의 기능성 도메인과 연관된다. 1개 초과의 기능성 도메인이 존재하는 경우, 기능성 도메인은 동일하거

나 상이할 수 있으며, 예를 들어, 2개가 동일하거나, 2개가 상이한 활성인자 또는 억제인자이다. 한 측면에서, 본 발명은 본원에 논의된 조성물을 제공하며, 여기서 하나 이상의 기능성 도메인은 RNA 표적화 효소에 부착되어, 표적 RNA에 결합시 기능성 도메인은 기능성 도메인이 그의 해당 기능에서 기능을 하도록 하는 공간 배열에 있다; 한 측면에서, 본 발명은 본원에 논의된 조성물을 제공하며, 여기서 조성물은 3 개 이상의 기능성 도메인을 갖는 CRISPR-Cas13b 복합체를 포함하며, 이들 중 1 개 이상은 RNA 표적화 효소와 관련되고 2 개 이상은 gRNA와 관련된다.

[0539] **Cas13b 이펙터 단백질 복합체 또는 이의 성분의 전달**

[0540] 본 개시 내용 및 당업계의 지식을 통하여, TALE, CRISPR-Cas 시스템, 또는 이의 성분 또는 이의 핵산 분자(예를 들어, HDR 주형을 포함함) 또는 이의 성분을 인코딩하거나 제공하는 핵산 분자는 본 명세서에서 일반적으로 그리고 상세하게 기재된 전달 시스템에 의해 전달될 수 있다.

[0541] 벡터 전달, 예를 들어 플라스미드, 바이러스 전달: CRISPR 효소, 및/또는 본 발명의 임의의 RNA, 예를 들어 가이드 RNA는 임의의 적합한 벡터, 예를 들어 플라스미드 또는 바이러스 벡터, 예컨대 아데노 연합 바이러스(AAV), 렌티바이러스, 아데노바이러스 또는 기타 다른 바이러스 벡터 유형, 또는 이의 조합을 사용하여 전달될 수 있다. 이펙터 단백질 및 하나 이상의 가이드 RNA는 하나 이상의 벡터, 예를 들어 플라스미드 또는 바이러스 벡터 내로 패키지될 수 있다. 일부 구현예에서, 벡터, 예를 들어 플라스미드 또는 바이러스 벡터는, 예를 들어 근육내 주사에 의해 관심 조직으로 전달되는 한편, 기타 다른 경우에 전달은 정맥내, 경피, 비강내, 구강, 점막 또는 기타 다른 전달 방법을 통하여 이루어진다. 이러한 전달은 단일 용량 또는 다중 용량 중 어느 하나를 통하여 이루어질 수 있다. 당업자는, 본 명세서에서 전달되는 실제 투여량은 다양한 인자, 예컨대 벡터 선택, 표적 세포, 유기체, 또는 조직, 치료될 대상체의 일반 상태, 구하는 형질전환/변형의 정도, 투여 경로, 투여 방법, 구하는 형질전환/변형의 형태 등에 따라 크게 달라질 수 있다.

[0542] 이러한 투여량은, 예를 들어 담체(물, 식염수, 에탄올, 글리세롤, 락토오스, 수크로오스, 인산칼슘, 젤라틴, 덱스트란, 한천, 펙틴, 땅콩유, 참기름 등), 희석제, 약학적으로 허용가능한 담체(예를 들어, 인산염 완충 식염수), 약학적으로 허용가능한 부형제, 및/또는 당업계에 알려진 기타 다른 화합물을 추가로 함유할 수 있다.

투여량은 하나 이상의 약학적으로 허용가능한 염, 예컨대 예를 들어, 히드로클로라이드, 히드로브로마이드, 포스페이트, 술 페이트 등과 같은 무기산 염; 및 아세테이트, 프로피오네이트, 말로네이트, 벤조에이트 등과 같은 유기산의 염을 추가로 함유할 수 있다.부가적으로, 습윤제 또는 유화제, pH 완충 물질, 겔 또는 겔화 물질, 향미제, 착색제, 미소구체, 중합체, 현탁화제 등과 같은 보조 물질이 또한 본원에 존재할 수 있다. 추가적으로, 하나 이상의 기타 다른 통상의 약학 성분들, 예컨대 보존제, 보수제, 현탁화제, 계면활성제, 산화방지제, 케이킹방지제(anticaking agent), 충전제, 킬레이트화제, 코팅제, 화학 안정화제 등이 또한 존재할 수 있으며, 특히 투약 형태는 재구성가능한 형태이다. 적합한 예시적 성분은 미정질 셀룰로스, 카복시메틸셀룰로스 나트륨, 폴리솔베이트 80, 페닐에틸 알코올, 클로로부탄올, 솔브산칼륨, 솔브산, 이산화황, 갈산프로필, 파라벤, 에틸 바닐린, 글리세린, 페놀, 파라클로로페놀, 젤라틴, 알부민 및 이들의 조합을 포함한다. 약제학적으로 허용가능한 부형제의 철저한 논의는 본 명세서에 참고로 포함되는 문헌 REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991 에서 이용가능하다.

[0543] 본 명세서의 일 구현예에서, 전달은 아데노바이러스를 통한 것으로, 이는 적어도 1×10^{5} 개의 입자(입자 단위, pu로서도 지칭됨)의 아데노바이러스 벡터를 함유하는 단일 부스터(booster) 용량일 수 있다. 본 명세서의 일 구현예에서, 용량은 바람직하게 적어도 약 1×10^{6} 개의 입자(예를 들어, 약 1×10^{6} 개 내지 1×10^{12} 개 입자), 더욱 바람직하게는 적어도 약 1×10^{7} 개의 입자, 더욱 바람직하게는 적어도 약 1×10^{8} 개 이상의 입자(예를 들어, 약, 1×10^{8} 개 내지 1×10^{11} 개 입자 또는 약 1×10^{8} 개 내지 1×10^{12} 개 입자), 가장 바람직하게는 적어도 약 1×10^{0} 개의 입자(예를 들어, 약 1×10^{9} 개 내지 1×10^{10} 개 입자 또는 약 1×10^{9} 개 내지 1×10^{12} 입자), 또는 심지어 적어도 약 1×10^{10} 개의 입자(예를 들어, 약 1×10^{10} 개 내지 1×10^{12} 개 입자)의 아데노바이러스 벡터이다. 대안적으로, 용량은 1×10^{14} 개 이하의 입자, 바람직하게는 약 1×10^{13} 개 이하의 입자, 훨씬 더 바람직하게는 약 1×10^{12} 개 이하의 입자, 훨씬 더 바람직하게는 약 1×10^{11} 개 이하의 입자, 및 가장 바람직하게는 약 1×10^{10} 개 이하의 입자(예를 들어, 약 1×10^{9} 개 이하의 입자)를 포함한다. 따라서, 용량은, 예를 들어 약 1×10^{6} 개 입자 단위(pu), 약 2×10^{6} pu, 약 4×10^{6} pu, 약 1×10^{7} pu, 약 2×10^{7} pu, 약 4 x

10^{7} pu, 약 1 x 10^{8} pu, 약 2 x 10^{8} pu, 약 4 x 10^{8} pu, 약 1 x 10^{9} pu, 약 2 x 10^{9} pu, 약 4 x 10^{9} pu, 약 1 x 10^{10} pu, 약 2 x 10^{10} pu, 약 4 x 10^{10} pu, 약 1 x 10^{11} pu, 약 2 x 10^{11} pu, 약 4 x 10^{11} pu, 약 1 x 10^{12} pu, 약 2 x 10^{12} pu, 또는 약 4 x 10^{12} pu의 아데노바이러스 벡터에 의한 단일 용량의 아데노바이러스 벡터를 함유할 수 있다. 예를 들어, 아데노바이러스 벡터에 대해, Nabel외 다수에게, 2013년 6월 4일 허여된, 미국 특허 8,454,972 B2를 참조하며, 이는 본 명세서에 참고로 포함되고, 이의 컬럼 29, 36 내지 58 줄의 투여량을 참조한다. 본 명세서의 일 구현예에서, 아데노바이러스는 다중 용량을 통해 전달된다.

[0544] 본 명세서의 일 구현예에서, 전달은 AAV를 통해 이루어진다. 인간에 대한 AAV의 생체내 전달을 위한 치료적 유효량은 약 1 x 10^{10} 내지 약 1 x 10^{10} 기능성 AAV/ml 용액을 함유하는 식염수 용액의 약 20 내지 약 50 ml의 범위에 있는 것으로 믿어진다. 투약량은 임의의 부작용에 대해 치료적 이점의 균형을 맞추도록 조절될 수 있다. 본 명세서의 구현예에서, AAV 용량은 일반적으로 약 1 x 10^{5} 내지 1 x 10^{50} 개 게놈 AAV, 약 1 x 10^{8} 개 내지 1 x 10^{20} 개 게놈 AAV, 약 1 x 10^{10} 내지 약 1 x 10^{16} 개 게놈, 또는 약 1 x 10^{11} 내지 약 1 x 10^{16} 개 게놈 AAV의 농도 범위에 있다. 인간 투약량은 약 1 x 10^{13} 개 게놈 AAV일 수 있다. 이러한 농도는 약 0.001 ml 내지 약 100 ml, 약 0.05 내지 약 50 ml, 또는 약 10 내지 약 25 ml의 담체 용액으로 전달될 수 있다. 다른 유효한 투약량은 용량 반응 곡선을 확립하는 일상적 시도를 통해 당업자에 의해 용이하게 확립될 수 있다. 예를 들어, 2013년 3월 26일 등록된 Hajjar외 다수의 미국 특허 8,404,658 B2, 27 칼럼 45 내지 60 줄을 참조한다.

[0545] 본원의 일 구현예에서, 전달은 플라스미드를 통해 이루어진다. 이러한 플라스미드 조성물에서, 투여량은 반응을 유발하는 데 충분한 양의 플라스미드이어gi한다. 예를 들어, 플라스미드 조성물 중의 플라스미드 DNA의 적합한 양은 70 kg의 개체 당 약 0.1 mg 내지 약 2 mg, 또는 약 1 ㎍ 내지 약 10 ㎍일 수 있다. 본 발명의 플라스미드는 일반적으로 (i) 프로모터; (ii) 상기 프로모터에 작동 가능하게 연결된 핵산-표적화 CRISPR 효소를 인코딩하는 서열; (iii) 선별가능한 마커; (iv) 복제 기원; 및 (v) (ii)의 하류에 작동가능하게 연결된 전사 터미네이터를 포함할 것이다. 플라스미드는 또한 CRISPR 복합체의 RNA 성분을 인코딩할 수 있지만, 대신에 이들 중 하나 이상은 상이한 벡터 상에서 인코딩될 수 있다.

[0546] 본 명세서의 용량은 평균 70 kg 개체를 기준으로 한다. 투여 빈도는 의학적 또는 수의학적 실행자(예를 들어, 의사, 수의사)또는 당업자의 영역 내이다. 또한, 실험에 사용된 마우스는 통상적으로 약 20 g이며, 마우스 실험으로부터 70 kg 개체까지 증대될 수 있음이 주목된다.

[0547] 몇몇 구현예에서 본 발명의 RNA 분자는 리포좀 또는 리포펙틴 제형 등으로 전달되며 당 분야의 숙련자에게 널리 공지된 방법에 의해 제조될 수 있다. 이러한 방법은 예를 들어 미국 특허 5,593,972, 5,589,466 및 5,580,859에 기재되며, 이들은 본원에 참조로 포함된다. 특별히 포유동물 세포 내로 siRNA의 향상되고 개선된 전달을 목표로 하는 전달 시스템이 개발되었으며 (참조, 예를 들어, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-1010; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen. 2002, 32: 107-108 및 Simeoni et al., NAR 2003, 31, 11: 2717-2724), 본 발명에 적용될 수있다. siRNA는 최근 영장류에서 유전자 발현의 억제에 성공적으로 사용되었고 (참조, 예를 들어, Tolentino et al., Retina 24(4):660), 이것은 또한 본 발명에 적용될 수 있다.

[0548] 실제로, RNA 전달은 생체내 전달의 유용한 방법이다. 리포좀 또는 입자를 사용하여 핵산-표적화 Cas 단백질 및 가이드 RNA (예를 들어, HR 복구 주형)를 세포 내로 전달할 수 있다. 따라서, 핵산-표적화 Cas 단백질/CRISPR 효소, 예컨대 CasCas9의 전달 및/또는 본 발명의 가이드 RNA의 전달은 RNA 형태로 일 수 있고, 미소소낭, 리포좀 또는 입자를 통해서 일 수 있다. 예를 들어, Cas13b mRNA 및 가이드 RNA 또는 crRNA는 생체내 전달을 위해 리포좀 입자 내로 패키징될 수 있다. Life Technologies로부터의 리포펙타민(Lipofectamine)과 같은 리포좀 트랜스펙션 시약 및 시판 중인 기타 다른 시약은 RNA 분자를 간에 효과적으로 전달할 수 있다.

[0549] RNA의 전달 수단은 또한 바람직하게는 나노 입자 (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, Advanced Functional Materials, 19: 3112-3118, 2010) 또는 엑소좀 (Schroeder, A., Levins, C., Cortez, C., Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641)을 통한 RNA의 전달을 포

함한다. 실제로, 엑소좀은 RNA-표적화 시스템과 일부 유사한 시스템인 전달 siRNA에서 특히 유용한 것으로 나타났다. 예를 들어, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA in vitro and in vivo." Nat Protoc. 2012 Dec;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov 15.)은 엑소좀이 상이한 생물학적 장벽을 통과하는 약물 전달을 위해 유망한 도구이며, 시험관 내 및 생체 내에서 siRNA의 전달을 위해 이용될 수 있다고 설명한다. 이들의 접근법은 펩티드 리간드와 융합된 엑소좀 단백질을 포함하여, 발현 벡터의 트랜스펙션을 통해 표적화된 엑소좀을 생성하는 것이다. 이어서, 엑소좀은 트랜스펙션된 세포 상청액으로부터 정제 및 특성화된 다음, RNA는 엑소좀에 로딩된다. 본 발명에 따른 전달 또는 투여는 엑소좀에 의해 수행될 수 있으며, 특히 뇌로 제한되지 않는다. 비타민 E(α-토코페롤)는 핵산-표적화 Cas 단백질과 접합될 수 있고, 예를 들어 뇌에 짧은 간섭 RNA(siRNA)를 전달하기 위해 Uno 등(문헌 [HUMAN GENE THERAPY 22:711-719 (June 2011)])에 의해 행해지는 것과 유사한 방식으로 고밀도 지단백질(HDL)과 함께 뇌에 전달될 수 있다. 마우스에 인산염 완충 식염수(PBS) 또는 유리 TocsiBACE 또는 Toc-siBACE/HDL이 충전되고, 뇌 주입 키트 3(Brain Infusion Kit 3)(Alzet)과 연결된 삼투 미니펌프(Osmotic minipump)(모델 1007D; Alzet, 미국 캘리포니아주 쿠퍼티노 소재)를 통해 주입하였다. 뇌 주입 캐뉼라를 등의 제3뇌실 내로 주입을 위해 중앙에서 브레그마(bregma)에 대해 약 0.5 mm 뒤에 위치시켰다. Uno외 다수는 HDL과 함께 3 nmol만큼 적은 Toc-siRNA가 동일한 ICV 주입 방법에 의해 유사한 정도로 표적 감소를 유도할 수 있었다는 것을 발견하였다. α-토코페롤에 접합되고 뇌에 표적화된 HDL과 공동-투여되는 유사한 투여량의 핵산-표적화 이펙터 단백질이 본 발명에서 인간에 대해 고려될 수 있고, 예를 들어 약 3 nmol 내지 약 3 μmol의 뇌로 표적화된 핵산-표적화 이펙터 단백질이 고려될 수 있다. Zou외 다수(문헌[HUMAN GENE THERAPY 22:465-475 (April 2011)])는 래트의 척수에서 생체내 유전자 침묵화를 위한 PKCγ를 표적화하는 짧은 헤어핀 RNA의 렌티바이러스-매개의 전달 방법을 기재한다. Zou외 다수는 척추강내 카테터에 의해 1×10^9 형질도입 단위(TU)/㎖의 역가를 갖는 재조합 렌티바이러스 약 10 ㎕를 투여하였다. 뇌에 표적화된 렌티바이러스 벡터에서 발현되는 핵산-표적화 이펙터 단백질의 유사한 투여량이 본 발명에서 인간에 대해 고려될 수 있는데, 예를 들어 1×10^9 형질도입 단위(TU)/㎖의 역가를 갖는 렌티바이러스에서 뇌로 표적화된 약 10 ㎖ 내지 50 ㎖의 핵산-표적화 이펙터 단백질이 고려될 수 있다.

[0550] 뇌로의 국소 전달에 관하여, 이는 다양한 방법으로 달성될 수 있다. 예를 들어, 물질은 예를 들어, 주사에 의해 선상체내로 전달될 수 있다. 주사는 개두술을 통해 정위적으로 실행될 수 있다.

[0551] **일반적으로 패키징 및 프로모터**

[0552] 생체내에서 게놈 변형을 매개하기 위해 RNA-표적화 이펙터 단백질 (Cas13b 단백질) 코딩 핵산 분자, 예를 들어 DNA를 벡터, 예를 들어 바이러스 벡터 내로 패키징하는 방법은 다음을 포함한다:

[0553] 단일 바이러스 벡터:

[0554] 2 개 이상의 발현 카세트를 함유하는 벡터:

[0555] 프로모터- 핵산-표적화 이펙터 단백질 코딩 핵산 분자- 터미네이터

[0556] 프로모터- 가이드 RNA1- 터미네이터

[0557] 프로모터- 가이드 RNA (N)-터미네이터 (벡터의 크기 제한까지)

[0558] 이중 바이러스 벡터:

[0559] RNA-표적화 이펙터 단백질 (Cas13b)의 발현을 유도하기 위한 하나의 발현 카세트를 함유하는 벡터 1

[0560] 프로모터- RNA-표적화 이펙터 (Cas13b) 단백질 코딩 핵산 분자- 터미네이터

[0561] 하나 이상의 가이드RNA 또는 crRNA의 발현을 유도하기 위한 하나의 발현 카세트를 함유하는 벡터 2

[0562] 프로모터- 가이드 RNA1 또는 crRNA1- 터미네이터

[0563] 프로모터- 가이드 RNA1 (N) 또는 crRNA1 (N)- 터미네이터 (벡터의 크기 제한까지)

[0564] RNA-표적화 이펙터 단백질 코딩 핵산 분자 발현을 구동하기 위해 사용되는 프로모터는 AAV ITR이 프로모터로서 기능할 수 있음을 포함할 수 있고: 이는 추가적인 프로모터 엘리먼트(벡터에서 공간을 차지할 수 있음)의 필요성을 제거하는데 유리하다. 추가적인 비어 있는 공간은 추가적인 요소(gRNA 등)의 발현을 유도하는 데 사용될 수 있다. 또한, ITR 활성은 비교적 더 약하므로, 핵산-표적화 이펙터 단백질의 과발현으로 인한 잠재적

독성을 감소시키는데 사용될 수 있다. 유비퀴터스 발현을 위해 CMV, CAG, CBh, PGK, SV40, 페리틴 중쇄 또는 경쇄 등의 프로모터를 사용할 수 있다. 뇌 또는 다른 CNS 발현을 위해서는: 모든 뉴런의 경우 SynapsinI, 흥분성 뉴런의 경우 CaMKIIalpha; GABAergic 뉴런의 경우 GAD67 또는 GAD65 또는 VGAT 등의 프로모터를 사용할 수 있다. 간 발현을 위해, 알부민 프로모터를 사용할 수 있다. 폐 발현을 위해, SP-B를 사용할 수 있다. 내피 세포를 위해, ICAM을 사용할 수 있다. 조혈 세포에서 IFN베타 또는 CD45를 사용할 수 있다. 조골세포에 대해 OG-2를 사용할 수 있다. 가이드 RNA를 구동하기 위해 사용되는 프로모터는, U6 또는 H1과 같은 Pol III 프로모터; 가이드 RNA 또는 crRNA를 발현하기 위한 Pol II 프로모터 및 인트론 카세트를 포함할 수 있다.

[0565] **아데노 연관 바이러스(AAV)**

[0566] Cas13b 및 하나 이상의 가이드 RNA 또는 crRNA는 아데노 연관 바이러스(AAV), 렌티바이러스, 아데노바이러스 또는 기타 다른 플라스미드 또는 바이러스 벡터 유형을 사용하여, 특히, 예를 들어, 미국 특허 8,454,972(아데노바이러스에 대한 제형, 용량), 8,404,658(AAV에 대한 제형, 용량) 및 5,846,946(DNA 플라스미드에 대한 제형, 용량)으로부터의 제형 및 용량, 및 렌티바이러스, AAV 및 아데노바이러스를 수반하는 임상 시험 및 임상 시험에 관한 간행물로부터의 제형 및 용량을 사용하여 전달될 수 있다. 예를 들어, AAV에 대해, 투여 경로, 제형 및 용량은 미국 특허 제8,454,972호에서와 같고, AAV를 수반하는 임상시험에서와 같을 수 있다. 아데노바이러스에 대해, 투여 경로, 제형 및 용량은 미국 특허 제8,404,658호에서와 같을 수 있고 아데노바이러스를 수반하는 임상시험에서와 같을 수 있다. 플라스미드 전달에 대해, 투여 경로, 제형 및 용량은 미국 특허 제5,846,946호에서와 같고, 플라스미드를 수반하는 임상 연구에서와 같을 수 있다. 용량은 평균 70 kg 개체(예를 들어, 성인 남성)를 기준으로 하거나 또는 이로 추정될 수 있고, 상이한 체중 및 종의 환자, 대상체, 포유동물에 대해 조정될 수 있다. 투여 빈도는 환자 또는 대상체의 연령, 성별, 일반적 건강상태, 다른 병태 및 처리될 특정 병태 또는 증상을 포함하는 보통의 인자에 따라서 의학적 또는 수의학적 실행자(예를 들어, 의사, 수의사)의 영역 내에 있다. 바이러스 벡터는 관심 대상의 조직 내에 주사될 수 있다. 세포 유형 특이적 게놈 변형을 위해, RNA-표적화 이펙터 단백질 (Cas13b 이펙터 단백질)의 발현은 세포 유형 특이적 프로모터에 의해 유도될 수 있다. 예를 들어, 간 특이적 발현은 알부민 프로모터를 사용할 수 있고, 뉴런 특이적 발현(예를 들어, CNS 장애를 표적화하기 위함)은 시냅신 I 프로모터를 사용할 수 있다. *생체 내* 전달의 관점에서, AAV는 몇 가지 이유로 다른 바이러스 벡터보다 유리하다: 낮은 독성(이것은 면역 반응을 활성화시킬 수 있는 세포 입자의 초원심분리를 요구하지 않는 정제 방법에 기인한 것일 수 있음), 및 . 숙주 게놈으로 통합되지 않기 때문에 삽입 돌연변이유발을 야기할 낮은 확률.

[0567] AAV는 4.5 또는 4.75 Kb의 패키징 제한을 가진다. 이는 RNA-표적화 이펙터 단백질 (Cas13b 이펙터 단백질) 코딩 서열을 비롯하여 프로모터 및 전사 종결자가 모두 동일한 바이러스 벡터에 맞아들어가야 하는 것을 의미한다. AAV에 관해서, AAV는 AAV1, AAV2, AAV5 또는 이들의 임의의 조합일 수 있다. 표적화될 세포와 관련하여 AAV들 중 AAV를 선택할 수 있는데, 예를 들어 뇌 또는 뉴런 세포를 표적화하기 위해 AAV 혈청형 1, 2, 5 또는 하이브리드 또는 캡시드 AAV1, AAV2, AAV5 또는 이들의 임의의 조합을 선택할 수 있고; 심장 조직을 표적화하기 위해 AAV4를 선택할 수 있다. AAV8은 간에 대한 전달에 유용하다. 본 명세서의 프로모터 및 벡터는 개별적으로 바람직하다. 이들 세포에 대한 특정 AAV 혈청형의 표는 다음과 같다 (참조 Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)):

세포주	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0
911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND
A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
단핵구	1111	100	ND	ND	125	1429	ND	ND
미성숙 DC	2500	100	ND	ND	222	2857	ND	ND
성숙 DC	2222	100	ND	ND	333	3333	ND	ND

[0568]

[0569] **렌티바이러스**

[0570] 렌티바이러스는 유사분열과 유사분열 후 세포 둘 다에서 그들의 유전자를 감염 및 발현시키는 능력을 갖는 복합 레트로바이러스이다. 가장 흔히 알려진 렌티바이러스는 광범위한 세포 유형을 표적화하기 위해 다른 바이러스의 외피 당단백질을 사용하는 인간 면역결핍바이러스(HIV)이다. 렌티바이러스 다음과 같이 제조될 수 있다. (렌티 바이러스 전달 플라스미드 백본을 함유하는) pCasES10의 클로닝 후에, 낮은 계대(p=5)에서 HEK293FT를 10% 소태아 혈청과 함께 그리고 항생제 없이 DMEM 중의 트렌스펙션 전날에 50% 컨플류언스로 T-75 플라스크에 파종하였다. 20시간 후에, 배지를 OptiMEM(무혈청) 배지로 바꾸고, 4 시간 후에 트랜스펙션을 행하였다. 세포를 10 ㎍의 렌티바이러스 전달 플라스미드(pCasES10), 및 5 ㎍의 pMD2.G(VSV-g 위형), 및 7.5 ㎍의 psPAX2(gag/pol/rev/tat)의 패키징 플라스미드를 이용하여 트랜스펙션시켰다. 양이온성 지질 전달제(50uL 리포펙타민 2000 및 100ul 플러스 시약)을 이용하여 4 mL OptiMEM 중에서 트렌스펙션을 행하였다. 6 시간 후, 배지를 10% 우태아혈청이 있는 무항생제 DMEM으로 바꾸었다. 이들 방법은 세포 배양 동안 혈청을 사용하지만, 무혈청 방법이 바람직하다.

[0571] 렌티바이러스는 다음과 같이 정제될 수 있다. 바이러스 상청액을 48 시간 후에 채취하였다. 상청액을 먼저 데브리스(debris)를 없애고, 0.45 ㎛ 저 단백질 결합(PVDF) 필터를 통해 여과시켰다. 이어서, 그들을 24,000 rpm에서 2 시간 동안 원심분리로 교반시켰다. 바이러스 펠렛을 4C에서 밤새 50 ul의 DMEM 중에서 재현탁시켰다. 이어서 이들을 분주하고 즉시 -80℃에서 즉시 동결시켰다.

[0572] 다른 구현예에서, 말 전염성 빈혈증 바이러스(EIAV)에 근거한 최소한의 비-영장류 렌티바이러스 벡터가 또한 특히 안구 유전자 치료에서 고려되었다(예를 들어, Balagaan, J Gene Med 2006; 8: 275 - 285 참조). 또 다른 실시형태에서, RetinoStat®, 망상 형태의 노인성 황반변성의 치료를 위한 망막하 주사를 통해 전달되는 혈관신생억제 단백질 엔도스타틴 및 앤지오스타틴을 발현시키는 말 감염성 빈혈 바이러스-기반의 렌티바이러스 유전자 치료법 벡터가 또한 고려되며(예를 들어, 문헌 [Binley et al., HUMAN GENE THERAPY 23:980-991 (September 2012)]), 이러한 벡터는 본 발명의 핵산-표적화 시스템을 위해 변형될 수 있다.

[0573] 또다른 실시형태에서, HIV tat/rev에 의해 공유되는 공통 엑손, 핵-국재화 TAR 디코이 및 항-CCR5-특이적 해머헤드 리보자임을 표적하는 siRNA를 갖는 자가-불활성화 렌티바이러스 벡터 (참조, 예를 들어, DiGiusto et al. (2010) Sci Transl Med 2:36ra43)는 본 발명의 핵산-표적화 시스템에 사용 및/또는 적응될 수 있다. 환자 중량 kg 당 최소 2.5 × 10^6 CD34+ 세포가 수집될 수 있고, 이들은 2 μmol/L-글루타민, 줄기 세포 인자 (100 ng/ml), Flt-3 리간드 (Flt-3L) (100 ng/ml), 및 트롬보포이에틴 (10 ng/ml) (CellGenix)을 함유하는 X-VIVO 15 매질 (Lonza) 내에서 2 × 10^6 세포/ml의 밀도로 16 시간 내지 20 시간 동안 예비자극되었다. 예비자극된 세포는 피브로넥틴으로 코팅된 75-cm^2 조직 배양 플라스크 내에서 16 시간 내지 24 시간 동안, 감염 다중도

5로 렌티바이러스를 이용하여 형질도입될 수 있다 (25 mg/cm^2) (RetroNectin,Takara Bio Inc.).

[0574] 렌티바이러스 벡터는 파킨슨병의 치료를 위한 치료의 경우와 같이 개시되었다, 예를 들어, 미국 특허 공개 제 20120295960호 및 미국 특허 제7303910호 및 제 7351585호 참조. 렌티바이러스 벡터는 또한 안질환의 치료용으로 개시되었으며, 예를 들어 미국 특허 공개 20060281180, 20090007284, US20110117189; US20090017543; US20070054961, US20100317109 참고. 렌티바이러스 벡터는 또한 뇌로의 전달을 위해 개시되어 있다, 참조, 예를 들어, 미국 특허 공개번호 US20110293571; US20110293571, US20040013648, US20070025970, US20090111106 및 미국 특허 US7259015.

[0575] **RNA 전달**

[0576] RNA 전달: 핵산-표적화 Cas13b 단백질 및/또는 가이드 RNA는 또한 RNA 형태로 전달될 수 있다. mRNA는 하기 엘리먼트를 포함하는 PCR 카세트를 사용하여 합성할 수 있다: 베타 글로빈-polyA 테일 (120 개 이상의 아데닌의 스트링)로부터의 T7_프로모터-코작 서열 (GCCACC)-이펙터 프로트레인-3 'UTR. 카세트는 T7 중합효소의 전사를 위해 사용될 수 있다. 가이드 RNA 또는 crRNA는 또한 T7_프로모터-GG-가이드 RNA 또는 crRNA 서열을 함유하는 카세트로부터의 시험관 내 전사를 사용하여 전사될 수 있다.

[0577] **입자 전달 시스템 및/또는 제형:**

[0578] 여러 유형의 입자 전달 시스템 및/또는 제형은 다양한 범위의 생체의학 적용에서 유용하다고 알려져 있다. 일반적으로, 입자는 이의 운반 및 특성에 대해서 전체 유닛으로서 거동하는 작은 물체로서 규정된다. 입자는 직경에 따라서 추가로 분류된다. 조립 입자는 2,500 나노미터 내지 10,000 나노미터 사이의 범위를 망라한다. 미세 입자는 100 나노미터 내지 2,500 나노미터 사이의 크기이다. 초미세 입자, 또는 나노입자는 일반적으로 1 나노미터 내지 100 나노미터 크기 사이이다. 100-nm 한계의 근거는 벌크 물질로부터 입자를 구별하는 신규 특성이 통상적으로 100 nm 미만의 역치 길이 규모에서 발생한다는 사실이다.

[0579] 본원에서 사용된 바와 같이, 입자 전달 시스템/제형은 본 발명에 따른 입자를 포함하는 임의의 생물학적 전달 시스템/제형으로서 규정된다. 본 발명에 따른 입자는 100 미크론(㎛) 미만의 가장 큰 크기(예를 들어, 직경)를 갖는 임의의 개체(entity)이다. 일부 실시형태에서, 본 발명의 입자는 10 ㎛ 미만의 가장 큰 치수를 갖는다. 일부 구현예에서, 본 발명의 입자는 2000 나노미터(nm) 미만의 가장 큰 치수를 갖는다. 일부 구현예에서, 본 발명의 입자는 1000 나노미터(nm) 미만의 가장 큰 치수를 갖는다. 일부 구현예에서, 본 발명의 입자는 900 nm, 800 nm, 700 nm, 600 nm, 500 nm, 400 nm, 300 nm, 200 nm, 또는 100 nm 미만의 가장 큰 치수를 갖는다. 통상적으로, 본 발명의 입자는 500 nm 이하의 가장 큰 치수(예를 들어, 직경)를 갖는다. 일부 구현예에서, 본 발명의 입자는 250 nm 이하의 가장 큰 치수(예를 들어, 직경)를 갖는다. 일부 구현예에서, 본 발명의 입자는 200 nm 이하의 가장 큰 치수(예를 들어, 직경)를 갖는다. 일부 구현예에서, 본 발명의 입자는 150 nm 이하의 가장 큰 치수(예를 들어, 직경)를 갖는다. 일부 구현예에서, 본 발명의 입자는 100 nm 이하의 가장 큰 치수(예를 들어, 직경)를 갖는다. 더 작은 입자, 예를 들어 50 nm 이하의 가장 큰 치수를 갖는 더 작은 입자가 본 발명의 일부 구현예에서 사용된다. 일부 구현예에서, 본 발명의 입자는 25 nm 및 200 nm 사이 범위의 가장 큰 치수를 갖는다.

[0580] 입자 특성화(예를 들어, 형태, 치수 등의 특성화를 포함함)는 상이한 기법을 사용하여 수행된다. 일반적인 기법은 전자 현미경검사법(TEM, SEM), 원자력 현미경검사법(AFM), 동적 광 산란(DLS), X-선 광전자 분광법(XPS), 분말 X-선 회절(XRD), 푸리에 형질전환 적외선 분광광도법(FTIR), 매트릭스-보조 레이저 탈착/이온화 비행시간형 질량 분석기(MALDI-TOF), 자외선-가시 분광법, 이중 분극 간섭계 및 핵 자기 공명(NMR)이다. 특성화(치수 측정)는 본 발명의 임의의 시험관 내, 생체 외 및/또는 생체내 적용을 위한 전달에 있어서 최적인 크기의 입자를 제공하기 위해 고유 입자(즉, 로딩 이전)에 대해 또는 카고의 로딩 후(본원에서 카고는, 예를 들어, 하나 이상의 성분의 CRISPR-Cas13b 시스템, 예를 들어, Cas13b 효소 또는 mRNA 또는 가이드 RNA, 또는 이들의 임의의 조합을 지칭하며, 추가적인 담체 및/또는 부형제를 포함할 수 있음)에 이루어질 수 있다. 바람직한 특정 구현예에서, 입자 치수(예를 들어, 직경) 특성화는 동적 레이저 산란(DLS)을 사용한 측정을 기반으로 한다. 미국 특허 제 8,709,843 호; 미국 특허 번호 6,007,845; 미국 특허 번호 5,855,913; 미국 특허 번호 5,985,309; 미국 특허 번호 5,543,158; 및 입자, 입자의 제조 및 사용 방법 및 측정 방법에 관한 공개문헌 James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84 를 참조한다. 또한 Dahlman et al. “Orthogonal gene control with a catalytically active Cas9 nuclease,” Nature Biotechnology 33, 1159-1161 (November, 2015) 를 참조한다.

[0581] 본 발명의 범주 내의 입자 전달 시스템은 고체, 반고체, 에멀젼 또는 콜로이드 입자를 포함하지만 이에 제한되지 않는 임의의 형태로 제공될 수 있다. 예를 들어, 지질-기반 시스템, 리포좀, 미셀, 미소소낭, 엑소좀, 또는 유전자-총을 포함하지만 이에 제한되지는 않는 본 명세서에 기재된 전달 시스템 중 임의의 전달 시스템이 본 발명의 범주 내의 입자 전달 시스템으로서 제공될 수 있다.

[0582] **입자**

[0583] Cas13b mRNA 및 가이드 RNA 또는 crRNA는 입자 또는 지질 외피를 사용하여 동시에 전달될 수 있으며; 예를 들어, 복합체로서 예를 들어, CRISPR 효소 및 본 발명의 RNA는 Dahlman et al., WO2015089419 A2 및 그곳에 인용된 문헌에서와 같이 입자, 예컨대 7C1(참조, 예를 들어, James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84), 예를 들어 지질 또는 리피도이드 및 친수성 중합체, 예를 들어 양이온성 지질 및 친수성 중합체를 포함하는 전달 입자를 통해 전달될 수 있고, 예를 들어 양이온성 지질은 1,2-디올레오일-3-트리메틸암모늄-프로판 (DOTAP) 또는 1,2-디테트라데카노일-*sn*-글리세로-3-포스포콜린 (DMPC)을 포함하고 및/또는 친수성 중합체는 에틸렌 글리콜 또는 폴리에틸렌 글리콜 (PEG)을 포함하고; 및/또는 입자는 콜레스테롤을 추가로 포함하고 (예를 들어, 제제 1 = DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0; 제제 번호 2 = DOTAP 90, DMPC 0, PEG 10, 콜레스테롤 0; 제제 번호 3 = DOTAP) 90, DMPC 0, PEG 5, 콜레스테롤 5 로부터의 입자), 입자는 먼저, 이펙터 단백질 및 RNA가 예를 들어 1 : 1 몰비로, 예를 들어 실온에서, 예를 들어 30 분 동안 예를 들어, 멸균된, 뉴클레아제가 없는 1X PBS에서 함께 혼합되고; 별도로, 제형에 적용가능한 DOTAP, DMPC, PEG 및 콜레스테롤을 알코올, 예를 들어 100% 에탄올에 용해시키고; 두 용액을 함께 혼합하여 복합체를 함유하는 입자를 형성하는 효율적인, 다단계 공정을 사용하여 형성된다. Cas13b 이펙터 단백질 mRNA 및 가이드 RNA는 입자 또는 지질 외피를 사용하여 동시에 전달될 수 있다. 상기 Dahlman et al 의 기술은 본 발명에 적용될 수 있다. 폐, 심혈관 또는 신장 세포로 본 발명의 핵산-표적화 시스템을 전달하기 위해 에폭시드-변형된 지질-중합체가 이용될 수 있으나, 당업자는 기타 다른 표적 기관으로 전달하기 위해 시스템을 적합하게 할 수 있다. 약 0.05 mg/kg 내지 약 0.6 mg/kg 범위의 투여량이 계획된다. 약 2 mg/kg의 전체 투여량의 수 일 또는 수 주에 걸친 투여량이 또한 계획된다. 예를 들어, Su X, Fricke J, Kavanagh DG, Irvine DJ ("In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. 2011 Jun 6;8(3):774-87. doi: 10.1021/mp100390w. Epub 2011 Apr 1) 은 인지질 이중층 쉘로 둘러싸인 폴리(β-아미노 에스테르) (PBAE) 코어를 갖는 생분해성 코어-쉘 구조 입자를 기술한다. 이들은 생체내 mRNA 전달을 위해 개발되었다. pH-반응성 PBAE 구성성분은 엔도좀 파괴를 촉진하도록 선택되는 한편, 지질표면층은 폴리양이온 코어의 독성을 최소화하도록 선택되었다. 따라서 이러한 것은 본 발명의 RNA를 전달하는 데 바람직하다.

[0584] 일 실시형태에서, 모두 뇌로의 펩티드의 구강 전달, 펩티드의 정맥내 전달 및 펩티드의 비강 전달에 적용될 수 있는 생접착성 중합체의 자가 조립에 기반한 입자가 고려된다. 다른 구현예, 예컨대 소수성 약물의 구강 흡수 및 눈 전달이 또한 고려된다. 분자 외피 기술은 보호되고 질병의 부위에 전달되는 조작된 중합체 외피를 수반한다 (참조, 예를 들어, Mazza, M. et al. ACSNano, 2013. 7(2): 1016-1026; Siew, A., et al. Mol Pharm, 2012. 9(1):14-28; Lalatsa, A., et al. J Contr Rel, 2012. 161(2):523-36; Lalatsa, A., et al., Mol Pharm, 2012. 9(6):1665-80; Lalatsa, A., et al. Mol Pharm, 2012. 9(6):1764-74; Garrett, N.L., et al. J Biophotonics, 2012. 5(5-6):458-68; Garrett, N.L., et al. J Raman Spect, 2012. 43(5):681-688; Ahmad, S., et al. J Royal Soc Interface 2010. 7:S423-33; Uchegbu, I.F. Expert Opin Drug Deliv, 2006. 3(5):629-40; Qu, X.,et al. Biomacromolecules, 2006. 7(12):3452-9 및 Uchegbu, I.F., et al. Int J Pharm, 2001. 224:185-199). 표적 조직에 따라서 약 5 mg/kg의 단일 또는 다중 용량이 고려된다.

[0585] 입자에 대해서는 또한 하기 문헌을 참조한다: Alabi et al., Proc Natl Acad Sci U S A. 2013 Aug 6;110(32):12881-6; Zhang et al., Adv Mater. 2013 Sep 6;25(33):4641-5; Jiang et al., Nano Lett. 2013 Mar 13;13(3):1059-64; Karagiannis et al., ACS Nano. 2012 Oct 23;6(10):8484-7; Whitehead et al., ACS Nano. 2012 Aug 28;6(8):6922-9 및 Lee et al., Nat Nanotechnol. 2012 Jun 3;7(6):389-93.

[0586] 미국 특허 출원 20110293703은 또한 본 발명의 핵산-표적화 시스템을 전달하기 위해 적용될 수 있는 폴리뉴클레오티드의 투여에서 특히 유용한 리피도이드 화합물에 관한 것이다. 일 양태에서, 아미노알코올 리피도이드 화합물은 작용제와 병용되어 세포 또는 대상체에게 전달되어 마이크로입자, 나노입자, 리포좀, 또는 미셀을 형성한다. 입자, 리포좀 또는 미셀에 의해 전달될 작용제는 기체, 액체 또는 고체의 형태일 수 있고, 작용제는 폴리뉴클레오티드, 단백질, 펩티드 또는 소형 분자일 수 있다. 미노알코올 리피도이드 화합물은 다른 아미노알코올 리피도이드 화합물, 중합체(합성 또는 천연), 계면활성제, 콜레스테롤, 탄수화물, 단백질, 지질 등

과 조합되어 입자를 형성할 수 있다. 이어서 이들 입자는 선택적으로 약학적 부형제와 조합되어 약학적 조성물을 형성할 수 있다. 미국 특허 공개 제20110293703호는 또한 아미노알코올 리피도이드 화합물을 준비하는 방법을 제공한다. 1 이상의 당량의 아민은 본 발명의 아미노알코올 리피도이드 화합물을 형성하기에 적합한 조건 하에서 에폭시드-종결 화합물의 1 당량 이상과 반응된다. 특정 구현예에서, 아민의 모든 아미노기는 에폭시드-말단 화합물과 완전히 반응되어 3차 아민을 형성한다. 다른 구현예에서, 아민의 모든 아미노기는 3차 아민을 형성하기 위하여 에폭시드-말단 화합물과 완전히 반응되지 않으며, 이에 의해 아미노알코올 리피도이드 화합물에서 1차 또는 2차 아민이 초래된다. 이들 1차 또는 2차 아민은 상이한 에폭시드-종결 화합물과 같은 다른 친핵체로서 남아있거나 또는 이런 다른 친핵체와 반응될 수 있다. 당업자에 의해 인식될 바와 같이, 아민을 과량 미만의 에폭시드-종결 화합물과 반응시키는 것은 다수의 꼬리를 지니는 복수의 상이한 아미노알코올 리피도이드 화합물을 초래할 것이다. 특정 아민은 2개의 에폭시드 유래 화합물 꼬리에 의해 완전히 기능화될 수 있는 반면, 다른 분자는 에폭시드 유래 화합물 꼬리에 의해 완전히 기능화되지 않을 것이다. 예를 들어, 디아민 또는 폴리아민은 1차, 2차 및 3차 아민을 초래하는 분자의 다양한 아미노산 모이어티의를 감소시킨 1, 2, 3 또는 4개의 에폭시드 유래 화합물을 포함할 수 있다. 특정 구현예에서, 모든 아미노기는 완전히 작용화되지 않는다. 특정 구현예에서, 2 가지 동일 유형의 에폭시드 말단 화합물이 사용된다. 다른 구현예에서, 2 개 이상의 상이한 에폭시드 말단 화합물이 사용된다. 아미노알코올 리피도이드 화합물의 합성은 용매와 함께 또는 용매 없이 수행되고, 합성은 30℃ 내지 100℃, 바람직하게는 대략 50℃ 내지 90℃ 범위의 고온에서 실행될 수 있다. 제조된 아미노알코올 리피도이드 화합물은 임의로 정제될 수 있다. 예를 들어, 아미노알코올 리피도이드 화합물의 혼합물은 정제되어 특정 수의 에폭시드 유래 화합물 꼬리를 지니는 아미노알코올 리피도이드 화합물을 수득할 수 있다. 또는 혼합물을 정제하여 특정 입체 이성질체 또는 구조 이성질체를 수득할 수 있다. 아미노알코올 리피도이드 화합물은 또한 알킬 할로겐화물(예를 들어, 요오드화메틸) 또는 다른 알킬화제를 사용하여 알킬화될 수 있고/있거나 그들은 아실화될 수 있다.

[0587] 미국 특허 공개 20110293703은 또한 본 발명의 방법에 의해 제조된 아미노알코올 리피도이드 화합물의 라이브러리를 제공한다. 이들 아미노알코올 리피도이드 화합물은 액체 처리기, 로봇, 마이크로타이터 플레이트, 컴퓨터 등을 포함하는 고-처리량 기술을 사용하여 제조 및/또는 스크리닝될 수 있다. 특정 실시형태에서, 아미노알코올 리피도이드 화합물은 폴리뉴클레오티드 또는 다른 작용제 (예를 들어, 단백질, 펩티드, 소분자)를 세포 내로 트랜스펙션시키는 능력에 대해 스크리닝된다. 미국 특허 공개 제20130302401호는 조합적 중합화를 사용하여 제조된 폴리(베타-아미노 알코올)(PBAA)의 부류에 관한 것이다. 본 발명의 PBAA는 코팅(예컨대, 의료장치 또는 이식물에 대한 필름 또는 다중층 필름의 코팅), 첨가제, 재료, 부형제, 비생물오손제(non-biofouling agent), 미세패터닝제 및 세포 캡슐화제로서 생명공학 및 생체의학 적용분야에서 사용될 수 있다. 표면 코팅으로서 사용될 때, 이들 PBAA는 그들의 화학적 구조에 따라서 시험관내와 생체내 둘 다의 상이한 염증 수준을 유발하였다. 이 재료 부류의 큰 화학적 다양성은 본 발명자들이 시험관내 대식세포 활성화를 저해하는 중합체 코팅을 동정하게 하였다. 더 나아가, 이들 코팅은 카복실화된 폴리스티렌의 피하 이식 후에 염증 세포의 동원을 감소시키고, 섬유증을 감소시킨다. 이들 중합체를 사용하여 세포 캡슐화를 위한 고분자전해질 복합체 캡슐을 형성한다. 본 발명은 또한 미생물방지 코팅, DNA 또는 siRNA 전달, 및 줄기 세포 조직 유전자조작과 같은 다수의 다른 생물학적 적용분야를 가질 수 있다. 미국 특허 공개 20130302401의 교시는 본 발명의 핵산-표적화 시스템에 적용될 수 있다.

[0588] 또 다른 구현예에서, 지질 나노입자(LNP)가 고려된다. 항트렌스티레틴 소형 간섭 RNA는 지질 나노입자에 캡슐화되고, 인간에게 전달되며(예를 들어, 문헌 [Coelho et al., N Engl J Med 2013;369:819-29] 참조), 이러한 시스템은 본 발명의 핵산-표적화 시스템에 적합화되어 적용될 수 있다. 정맥내로 투여된 체중의 kg 당 약 0.01 내지 약 1 mg의 용량이 고려된다. 주입 관련 반응의 위험을 감소시키기 위한 의약이 고려되며, 예컨대 덱사메타손, 아세탐피노펜, 디펜하이드라민 또는 세티리진 및 라니티딘이 고려된다. 5 회 용량에 있어서 4 주마다 킬로그램 당 약 0.3 mg의 다회 용량이 또한 고려된다. LNP는 siRNA를 간으로 전달하는데 매우 효과적인 것으로 나타 났으며 (참조, 예를 들어, Tabernero et al., Cancer Discovery, April 2013, Vol. 3, No. 4, pages 363-470), 핵산-표적화 이펙터 단백질을 인코딩하는 RNA를 간으로 전달하는 것으로 고려된다. 2 주마다 6 mg/kg의 LNP의 약 4 회 용량의 투약량이 고려될 수 있다. Tabernero외 다수는 0.7 mg/kg으로 투약한 LNP의 처음 2 사이클 후에 종양 퇴행을 관찰하였고, 6 사이클의 종료까지 환자는 림프절 전이의 완전한 퇴행 및 간 종양의 실질적 수축으로 부분적 반응을 달성하였음을 입증하였다. 26 개월에 걸쳐 용량을 받은 후에 관해가 남아있고 치료를 완료한 이 환자에게 40 용량 후에 완전한 반응을 얻었다. VEGF 경로 억제제를 이용한 사전 치료법 후에 진행한 신장, 폐 및 림프절을 포함하는 RCC 및 간외영역 질병을 지니는 2 명의 환자는 대략 8 개월 내지 12 개월 동안 모든 부위에서 안정적인 질병을 가졌고, PNET 및 간 전이를 지니는 환자는 안정

적인 질병과 함께 18 개월(36 회 용량) 동안 연장 연구를 계속하였다. 그러나, LNP의 변화를 고려하여야 한다. 양이온성 지질은 세포내 전달을 용이하게 하는 비이중층 구조를 유발하기 위해 음으로 하전된 지질과 조합되기 때문이다. 하전된 LNP는 정맥내 주사 후 순환으로부터 빠르게 제거되기 때문에, 7 미만의 pKa 값을 지니는 이온화 가능한 양이온성 지질을 개발하였다(참조, 예를 들어, Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). 음으로 하전된 중합체, 예컨대 RNA는 이온화 가능한 지질이 양전하를 나타내는 경우 낮은 pH 값(예를 들어, pH 4)에서 LNP에 로딩될 수 있다. 그러나, 생리적 pH 값에서, LNP는 더 긴 순환시간과 양립가능한 낮은 표면 전하를 나타낸다. 이온화 가능한 양이온성 지질의 4 개 종, 즉, 1,2-디리네오일-3-디메틸암모늄-프로판(DLinDAP), 1,2-디리놀레일옥시-3-N,N-디메틸아미노프로판(DLinDMA), 1,2-디리놀레일옥시-케토-N,N-디메틸-3-아미노프로판(DLinKDMA), 및 1,2-디리놀레일-4-(2-디메틸아미노에틸)-[1,3]-디옥솔란(DLinKC2-DMA)에 중점을 두었다. 이들 지질을 함유하는 LNP siRNA 시스템은 생체내 간세포에서 현저하게 상이한 유전자 침묵화 특성을 나타내는 것으로 나타났고, 효능은 인자 VII 유전자 침묵화 모델을 이용하는 시리즈 DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP에 따라 다르다(참조, 예를 들어, Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). 특히 DLinKC2-DMA를 함유하는 제형에 대하여, 1 ㎍/㎖의 투여량의 LNP 또는 LNP 중의 또는 이와 연합된 CRISPR Cas RNA가 고려될 수 있다.

[0589] LNP의 제조 및 CRISPR-Cas13b 캡슐화는 Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011로부터 사용 및/또는 적응될 수 있다. 양이온성 지질 1,2-디리네오일-3-디메틸암모늄-프로판(DLinDAP), 1,2-디리놀레일옥시-3-N,N-디메틸아미노프로판(DLinDMA), 1,2-디리놀레일옥시케토-N,N-디메틸-3-아미노프로판(DLinK-DMA), 1,2-디리놀레일-4-(2-디메틸아미노에틸)-[1,3]-디옥솔란(DLinKC2-DMA), (3-o-[2"-(메톡시폴리에틸렌글리콜 2000) 숙시노일]-1,2-디미리스토일-sn-글리콜(PEG-S-DMG), 및 R-3-[(ω-메톡시-폴리(에틸렌 글리콜)2000) 카르바모일]-1,2-디미리스틸옥시프로필-3-아민(PEG-C-DOMG)은 Tekmira Pharmaceuticals(캐나다 밴쿠버에 소재)에 의해 제공될 수 있거나 합성될 수 있다. 콜레스테롤은 Sigma사(미조리주 세인트루이스에 소재)로부터 구입할 수 있다. 특정 핵산-표적화 복합체 (CRISPR Cas) RNA는 DLinDAP, DLinDMA, DLinK-DMA 및 DLinKC2-DMA (40:10:40:10 몰비의 양이온성 지질:DSPC:CHOL:PEGS-DMG 또는 PEG-C-DOMG)를 함유하는 LNP 내에 캡슐화될 수 있다. 필요하다면, 0.2% SP-DiOC18(Invitrogen, 캐나다 벌링턴 소재)을 혼입시켜 세포 흡수, 세포내 전달, 및 생체분포를 평가할 수 있다. 캡슐화는 10 mmol/l의 최종 지질 농도로 에탄올 중에서 양이온성 지질:DSPC:콜레스테롤:PEG-c-DOMG (40:10:40:10 몰비)을 포함한 지질 혼합물을 용해함으로써 수행될 수 있다. 지질의 이 에탄올 용액을 50 mmol/l 시트르산염, pH 4.0에 적가하여 다중 라멜라 소수포를 형성하여 30% 에탄올 vol/vol의 최종 농도를 생성할 수 있다. 압출기(Northern Lipids, 캐나다 벤쿠버 소재)를 사용하여 2 개의 적층 80 nm Nuclepore 폴리카르보네이트 필터를 통해 다중 라멜라 소낭을 압출한 후 거대 단일 라멜라 소낭이 형성될 수 있다. 30% 에탄올 vol/vol을 함유하는 50 mmol/l 시트르산염, pH 4.0 중에 2 mg/㎖로 용해시킨 RNA를 압출된 사전형성된 거대 단일 라멜라 소낭에 적가하고, 0.06/1 wt/wt의 최종 RNA/지질 중량비로 일정하게 혼합하면서 30 분 동안 31℃에서 인큐베이션함으로써 캡슐화를 달성할 수 있다. Spectra/Por 2 재생 셀룰로스 투석막을 사용하여 16 시간 동안 인산염 완충 식염수(PBS), pH 7.4에 대해 투석에 의해 에탄올의 제거 및 제형 완충제의 중화를 수행하여다. NICOMP 370 입도 분석기, 소낭/강도 모드, 및 가우시안(Gaussian) 핏팅(Nicomp Particle Sizing, 미국 캘리포니아주 산타 바바라 소재)을 사용하여 동적 광 산란에 의해 입자 크기 분포를 결정할 수 있다. 3 개 모두의 LNP 시스템에 대한 입자 크기는 직경이 약 70 nm일 수 있다. VivaPureD MiniH 컬럼(Sartorius Stedim Biotech)을 사용하여 투석 전 및 후에 수집한 샘플로부터 유리 RNA를 제거함으로써 RNA 캡슐화 효율을 결정할 수 있다. 캡슐화된 RNA는 용리된 입자로부터 추출되고, 260 nm에서 정량화될 수 있다. RNA 대 지질 비는 Wako Chemicals USA(미국 버지니아주 리치몬드 소재)로부터의 콜레스테롤 E 효소 어세이를 사용하여 소낭 내 콜레스테롤 함량의 측정에 의해 결정되었다. LNP 및 PEG 지질의 본 명세서에서의 논의와 함께, PEG화 리포좀 또는 LNP는 마찬가지로 핵산-표적화 시스템 또는 이의 성분의 전달에 적합하다. 대형 LNP의 제조는 Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011로부터 사용 및/또는 적응될 수 있다. 지질 프레믹스 용액(20.4 mg/ml 총 지질 농도)는 50:10:38.5 몰비에서 DLinKC2-DMA, DSPC 및 콜레스테롤을 함유하는 에탄올 중에서 제조될 수 있다. 아세트산나트륨은 0.75:1의 몰비(아세트산나트륨:DLinKC2-DMA)로 지질 예비혼합물에 첨가될 수 있다. 지질은 격렬한 교반과 함께 1.85 부피의 시트르산염 완충제(10 mmol/l, pH 3.0)와 혼합물을 합함으로써 이후에 수화되어, 35% 에탄올을 함유하는 수성 완충제 중의 자발적 리포좀 형성을 초래할 수 있다. 리포좀 용액을 37℃에서 인큐베이션시켜 시간-의존적 입자 크기의 증가를 가능하게 할 수 있다. 분취물을 인큐베이션 동안 다양한 시간에 제거하여, 동적 광 산란에 의해 리포좀 크기의 변화를 조사할 수 있다(Zetasizer Nano ZS, Malvern Instruments, 영국 우스터셔 소재). 일단 목적으로 하는 입자 크기가 달성되면, 수성 PEG 지질 용액(저장액

= 35%(vol/vol) 에탄올 중의 10 mg/ml PEG-DMG)은 리포좀 혼합물에 첨가되어 3.5%의 총 지질의 최종 PEG 몰 농도를 수득할 수 있다. PEG-지질의 첨가 시, 리포좀을 추가 성장을 효과적으로 퀀칭시켜 그들의 크기를 해야 한다. 이어서, RNA를 대략 1:10 (wt:wt)의 RNA 대 총 지질 비율로 빈 리포좀에 첨가한 후에 37℃에서 30 분 동안 항온반응시켜 로딩된 LNP를 형성할 수 있다. 혼합물을 후속적으로 PBS 중에 밤새 투석시키고, 0.45-㎛ 주사기 필터로 여과할 수 있다.

[0590] 구체 핵산(Spherical Nucleic Acid; SNA™) 구축물 및 다른 입자(특히, 금 입자)는 또한 의도된 표적에 핵산-표적화 시스템을 전달하기 위한 수단으로서 고려된다. 유의한 데이터는 핵산 작용화된 금 입자에 기반한 AuraSense Therapeutics의 구체 핵산(SNA™) 구축물이 유용한 것을 나타낸다.

[0591] 본 명세서의 교시와 함께 이용될 수 있는 문헌은 하기를 포함한다: Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) 및 Mirkin, et al., Small, 10:186-192.

[0592] RNA를 갖는 자가 조립 입자는, 폴리에틸렌 글리콜(PEG)의 원위 말단에 부착된 Arg-Gly-Asp(RGD) 펩티드 리간드로 PEG화된 폴리에틸렌이민(PEI)을 이용하여 작제될 수 있다. 이러한 시스템은, 예를 들어 인테그린을 발현하는 종양 신생혈관을 표적화하고, 혈관 내피 성장 인자 수용체-2(VEGF R2) 발현을 억제하는 siRNA를 전달하여, 이에 의해, 종양 혈관형성을 달성하기 위한 수단으로 사용된다(참조, 예를 들어, Schiffelers et al., Nucleic Acids Research, 2004, Vol. 32, No. 19). 나노복합체는 동일한 용적의 양이온성 중합체 및 핵산의 수용액을 혼합함으로써 제조되어 2 내지 6의 범위에 걸쳐 인산염(핵산)에 대한 순 몰과량의 이온화가능한 질소(중합체)를 제공할 수 있다. 양이온성 중합체와 핵산 간의 정전기적 상호작용은 약 100 nm의 평균 입자 크기를 지니는 다중복합체 형성을 야기하며, 따라서 본 명세서에서 나노복합체로서 지칭된다. Schiffelers 등의 자가 조립 입자로 전달을 위해 약 100 mg 내지 200 mg의 핵산-표적화 복합체 RNA의 투여량이 예상된다.

[0593] Bartlett 등의 나노플렉스 (PNAS, September 25, 2007,vol. 104, no. 39)도 본 발명에 적용될 수 있다. Bartlett외 다수의 나노플렉스는 동일 부피의 양이온성 중합체 및 핵산의 수용액을 혼합함으로써 제조되어 2 내지 6의 범위에 걸쳐 인산염(핵산)에 대한 순 몰 과량의 이온화 가능한 질소(중합체)를 제공한다. 양이온성 중합체와 핵산 간의 정전기적 상호작용은 약 100 nm의 평균 입자 크기를 지니는 다중복합체 형성을 야기하며, 따라서 본 명세서에서 나노복합체로서 지칭된다. Bartlett 등의 DOTA-siRNA를 다음과 같이 합성하였다: 1,4,7,10-테트라아자사이클로도데칸-1,4,7,10-테트라아세트산 모노(N-하이드록시숙신이미드 에스테르)(DOTA-NHS 에스테르)를 Macrocyclics(미국 텍사스주 달라스 소재)에서 주문하였다. 탄산염 완충제 중의 100 배 몰 과량의 DOTA-NHS-에스터에 의한 아민 변형 RNA 센스 가닥(pH 9)을 마이크로원심분리관에 첨가하였다. 내용물을 실온에서 4 시간 동안 교반시킴으로써 반응시켰다. DOTA-RNA센스 콘쥬게이트를 에탄올 침전시키고, 수중에 재현탁화한 다음, 미변형 안티센스 가닥에 어닐링시켜 DOTA-siRNA를 수득하였다. 모든 액체를 Chelex-100(캘리포니아주 허큘러스에 소재한 Bio-Rad)으로 전처리하여 미량의 금속 오염물질을 제거하였다. Tf-표적화 및 비표적화 siRNA 입자를 사이클로덱스트린-함유 다가양이온을 사용함으로써 형성할 수 있다. 통상적으로, 입자를 전하비 3(+/-) 및 siRNA 농도 0.5 g/리터로 수 중에서 형성하였다. 표적화된 입자의 표면 상의 아다만탄-PEG 분자의 1%를 Tf로 변형시켰다(아다만탄-PEG-Tf). 입자를 주사용 5%(wt/vol) 글루코스 담체 용액 중에서 현탁화시켰다.

[0594] Davis 등(문헌 [Nature, Vol 464, 15 April 2010])은 표적화된 입자-전달 시스템(임상 시험 등록 번호 NCT00689065)을 사용하는 RNA 임상 시험을 수행한다. 표준 치료 치료법에 난치성인 고형암을 지니는 환자에게 30 분 정맥내 주입에 의해 21 일 사이클 중 제1 일, 제3 일, 제8 일 및 제10 일에 표적화된 입자의 용량을 투여한다. 입자는 하기를 함유하는 합성 전달 시스템을 포함하거나, 그로 본질적으로 이루어지거나, 또는 그로로 이루어진다:(1) 선형, 사이클로덱스트린-기반 중합체(CDP), (2) 암 세포의 표면 상의 TF 수용체(TFR)에 맞물리는 나노입자 외부에 나타나는 인간 트랜스페린 단백질(TF) 표적화 리간드, (3) 친수성 중합체(생물학적 유체 중에서 나노입자 안정성을 촉진하기 위해 사용되는 폴리에틸렌 글리콜(PEG)), 및 (4) RRM2의 발현을 감소시키도록 디자인된 siRNA(임상에서 사용한 서열은 앞서 정의한 siR2B+5였음). TFR는 악성 세포에서 상향조절되는 것으로 알려져 있고, RRM2는 확립된 항암 표적이다. 이들 나노입자(CALAA-01로서 정의된 임상 형태)는 비-인간 영장류에서의 다회 투약 연구에서 잘 용인되는 것으로 나타났다. 만성 골수성 백혈병을 지니

는 단일 환자에게 리포좀 전달에 의해 siRNA가 투여되었지만, Davis외 다수의 임상 시험은 표적화된 전달 시스템을 이용하여 siRNA를 전신으로 전달하기 위한 그리고 고형암을 지니는 환자를 치료하기 위한 초기 인간 시험이다. 표적화된 전달 시스템이 인간 종양으로의 작용성 siRNA의 효과적인 전달을 제공할 수 있는지 여부를 확인하기 위해, Davis외 다수는 3 개의 상이한 투약 코호트로부터의 3 명 환자로부터의 생검을 조사하였으며; 환자 A, B 및 C는 모두 전이성 흑색종을 가졌고, 각각 18, 24 및 30 mg m^{-2} siRNA의 CALAA-01 용량을 받았다.

유사한 용량이 또한 본 발명의 핵산-표적화 시스템을 위해 고려될 수 있다. 본 발명의 전달은 선형, 사이클로덱스트린-기반 중합체(CDP), 암세포의 표면 상의 TF 수용체(TFR)에 맞물리는 입자 외부에 나타나는 인간 트랜스페린 단백질(TF) 표적화 리간드 및/또는 친수성 중합체(예를 들어, 생물학적 유체 중에서 입자 안정성을 촉진하기 위해 사용되는 폴리에틸렌 글리콜(PEG))를 함유하는 입자에 의해 달성될 수 있다.

[0595] 본 발명과 관련하여, 입자 또는 지질 엔벨로프를 사용하여 RNA-표적화 복합체의 하나 이상의 성분, 예를 들어 핵산-표적화 이펙터 (Cas13b) 단백질 또는 이에 대한 mRNA 또는 가이드 RNA 또는 crRNA를 전달하는 것이 바람직하다. 다른 전달 시스템 또는 벡터가 본 발명의 입자 측면과 함께 사용될 수 있다. 본 발명에 포함된 입자는 상이한 형태, 예를 들어 고체 입자(예를 들어, 금속, 예컨대 은, 금, 철, 티타늄), 비-금속, 지질-기반 고체, 중합체), 입자의 현탁액, 또는 이들의 조합물로 제공될 수 있다. 금속, 유전체, 및 반도체 입자뿐만 아니라 하이브리드 구조(예를 들어, 코어-쉘 입자)가 제조될 수 있다. 반도체 물질로 제조된 입자는 또한 이들이 전자 에너지 수준의 양자화가 발생하기에 충분히 작은(통상적으로, 10 nm 이하) 경우 양자점으로 표지화될 수 있다. 이러한 나노규모 입자는 약물 담체 또는 영상화 작용제로서 생물의학 응용에서 사용되며, 본 발명에서 유사한 목적을 위해 적합하게 될 수 있다.

[0596] 반고체 및 연성 입자가 제조되었고, 이들은 본 발명의 범주 내에 속한다. 반고체 성질의 프로토타입 입자는 리포좀이다. 다양한 유형의 리포좀 입자가 항암 약물 및 백신을 위한 전달 시스템으로서 현재 임상적으로 사용된다. 절반이 친수성이고, 나머지 절반이 소수성인 입자는 야누스 입자로 명명되며, 에멀젼을 안정화시키는데 특히 효과적이다. 이들은 물/오일 계면에서 자가-조립되며, 고체 계면활성제로 작용할 수 있다.

[0597] 참고로 본 명세서에 포함되는 미국 특허 8,709,843은 조직, 세포, 및 세포내 구획으로의 치료제-함유 입자의 표적화된 전달을 위한 약물 전달 시스템을 제공한다. 상기 발명은 계면활성제, 친수성 중합체 또는 지질에 접합된 중합체를 포함하는 포함하는 표적화된 입자를 제공한다. 참고로 본 명세서에 포함되는 미국 특허 6,007,845는 다작용성 화합물과 하나 이상의 소수성 중합체 및 하나 이상의 친수성 중합체를 공유적으로 연결시킴으로써 형성된 멀티블록 공중합체의 코어를 갖고, 생물학적으로 활성인 물질을 함유하는 입자를 제공한다.

참고로 본 명세서에 포함되는 미국 특허 5,855,913은 폐기관계로의 약물 전달을 위해 표면 상에 계면활성제가 혼입된, 0.4 g/㎤ 미만의 탭 밀도(tap density)와 함께 5 ㎛ 내지 30 ㎛의 평균 직경을 갖는 공기역학적으로 가벼운 입자를 갖는 미립자 조성물을 제공한다. 참고로 본 명세서에 포함되는 미국 특허 5,985,309는 폐기관계로의 전달을 위한 계면활성제 및/또는 양으로 또는 음으로 하전된 치료제 또는 진단제 및 반대 전하로 하전된 분자의 친수성 또는 소수성 복합체가 혼입된 입자를 제공한다. 참고로 본 명세서에 포함되는 미국 특허 5,543,158은 표면 상에 폴리(알킬렌 글리콜) 모이어티 및 생물학적 활성 물질을 함유하는 생분해성 고체 코어를 갖는 생분해성 주사가능한 입자를 제공한다. 참고로 본 명세서에 포함되는 WO2012135025(또한 US20120251560으로 공개됨)는 콘쥬게이트된 폴리에틸렌이민(PEI) 중합체 및 콘쥬게이트된 아자-마크로사이클(집합적으로, "콘쥬게이트된 리포머" 또는 "리포머"로 지칭됨)이 기재한다. 특정 실시형태에서, 본 명세서에서 인용된 문헌의 이러한 방법 및 재료, 예를 들어 접합된 리포머는 단백질 발현의 조정을 포함하여, 유전자 발현을 변형시키기 위한 시험관 내, 생체 외 및 생체내 게놈 교란을 달성하기 위해 핵산-표적화 시스템의 상황에서 사용될 수 있음이 계획될 수 있다.

[0598] **엑소좀**

[0599] 엑소좀은 뇌 및 기타 다른 표적 기관에 RNA를 전달할 수 있는, 단백질 및 RNA를 전달하는 내인성 나노-소낭이다. 면역원성을 감소시키기 위해, Alvarez-Erviti외 다수(문헌[2011, Nat Biotechnol 29:341])는 엑소좀 생성을 위해 자가-유래 수상돌기 세포를 사용하였다. 뉴런-특이적 RVG 펩티드에 융합된 엑소좀 막 단백질인 Lamp2b를 발현하도록 수상돌기 세포를 조작함으로써 뇌로의 표적화를 달성하였다. 정제된 엑소좀에 전기천공법에 의해 외인성 RNA를 로딩하였다. 정맥내로 주사한 RVG-표적화 엑소좀은 뇌 내의 뉴런, 미세아교세포, 희소돌기신경교에 특이적으로 GAPDH siRNA를 전달하여, 특이적 유전자 녹다운을 초래하였다. RVG 엑소좀에 대한 사전 노출은 녹다운을 약화시키지 않았고, 기타 다른 조직에서의 비-특이적 흡수는 관찰되지 않았다. 엑소좀-매개 siRNA 전달의 치료 잠재력은 알츠하이머병에서의 치료 표적인 BACE1의 강력

한 mRNA(60%) 및 단백질(62%) 녹다운에 의해 입증되었다.

[0600] 면역학적으로 비활성인 엑소좀의 풀을 얻기 위해, Alvarez-Erviti외 다수는 상동성 주조직적합성 복합체(MHC) 단상형을 지니는 근친교배 C57BL/6 마우스로부터 골수를 수집하였다. 미성숙 수상돌기 세포가 MHC-II 및 CD86과 같은 T-세포 활성화제가 없는 다량의 엑소좀을 생성하기 때문에, Alvarez-Erviti외 다수는 7 일 동안 과립구/대식세포-콜로니 자극 인자(GM-CSF)를 지니는 수상돌기 세포에 대해 선택하였다. 잘-확립된 초원심분리 프로토콜을 사용하여 다음날 배양 상청액으로부터 엑소좀을 정제하였다. 생성된 엑소좀은 물리적으로 균질하였고, 크기 분포는 입자 추적 분석(NTA) 및 전자 현미경에 의해 결정하였을 때, 직경 80 nm에서 최고였다.

Alvarez-Erviti외 다수는 10^6 개 세포당 6 ㎍ 내지 12 ㎍의 엑소좀(단백질 농도를 기준으로 측정)을 얻었다.

다음에, Alvarez-Erviti외 다수는 나노규모 적용에 적합한 전기천공법 프로토콜을 사용하여 외인성 카고로의 변형된 엑소좀의 로딩 가능성을 조사하였다. 나노미터 규모에서 막 입자에 대한 전기천공법은 제대로 특성화되지 않기 때문에, 비특이적 Cy5-표지화 RNA를 전기천공법 프로토콜의 경험적 최적화를 위해 사용하였다.

캡슐화된 RNA의 양을 엑소좀의 초원심분리 및 용해 후에 분석하였다. 400 V 및 125 μF에서의 전기천공법은 RNA의 가장 큰 보유를 초래하였고, 모든 후속 실험에 대해 사용하였다. Alvarez-Erviti외 다수는 150 ㎍의 RVG 엑소좀 중에 캡슐화된 150 ㎍의 각각의 BACE1 siRNA를 정상 C57BL/6 마우스에게 투여하고, 녹다운 효율을 4 개의 대조군, 즉 미처리 마우스, RVG 엑소좀만을 주사한 마우스, 생체내 양이온성 리포좀 시약에 복합체화된 BACE1 siRNA를 주사한 마우스, 및 siRNA에 정전기적으로 결합된 9 개의 D-아르기닌에 콘쥬게이트된 RVG 펩티드인 RVG-9R에 복합체화된 BACE1 siRNA를 주사한 마우스에 대하여 비교하였다. 투여 3 일 후 피질 조직 샘플을 분석하였고, siRNA-RVG-9R-처리 마우스와 siRNARVG 엑소좀-처리 마우스 둘 다에서 BACE1 mRNA 수준의 상당한 감소(각각 66% [+ 또는 -] 15%, $P < 0.001$ 및 61% [+ 또는 -] 13%, $P < 0.01$)로부터 초래되는 상당한 단백질 녹다운(45%, $P < 0.05$, 대 62%, $P < 0.01$)을 관찰하였다. 게다가, 본 발명자들은 RVG-엑소좀 처리 동물에서 알츠하이머 병리에서의 아밀로이드 플라크의 주요 구성성분인 총 [베타]-아밀로이드 1 내지 42 수준에서 상당한 감소(55%, $P < 0.05$)를 입증하였다. 관찰한 감소는 BACE1 저해제의 정맥내 주사 후 정상 마우스에서 입증된 β-아밀로이드 1 내지 40 감소보다 더 컸다. Alvarez-Erviti외 다수는 BACE1 절단 생성물에 대해 cDNA 말단의 5'-신속한 증폭(RACE)을 수행하였는데, 이는 siRNA에 의한 RNAi-매개 녹다운의 증거를 제공하였다.

최종적으로, Alvarez-Erviti 등은 IL-6, IP-10, TNFα 및 IFN-α 혈청 농도를 평가함으로써 RNA-RVG 엑소좀이 생체내 면역 반응을 유도하였는지 여부를 조사하였다. 엑소좀 처리 후에, 모든 사이토카인에서의 중요하지 않은 변화가 IL-6 분비를 강력하게 자극한 siRNA-RVG-9R과 대조적으로 siRNA-트랜스펙션 시약 처리와 유사하게 등록되었는데, 이는 엑소좀 처리의 면역학적 비활성 프로파일을 확인시켜준다. 엑소좀이 단지 20%의 siRNA를 캡슐화하는 것을 고려해 볼 때, RVG-엑소좀에 의한 전달은 비슷한 mRNA 녹다운으로서 RVG-9R 전달보다 더 효율적인 것으로 나타나며, 대응하는 면역 자극 수준 없이, 더 큰 단백질 녹다운이 5 배 더 적은 siRNA에 의해 달성되었다. 이 실험은 신경변성 질병과 관련된 유전자의 장기간 침묵화에 잠재적으로 적합한 RVG-엑소좀 기술의 치료 잠재력을 입증하였다. Alvarez-Erviti 등의 엑소좀 전달 시스템은 치료 표적, 특히 신경변성 질환에 본 발명의 핵산-표적화 시스템을 전달하기 위해 적용될 수 있다. 약 100 mg 내지 1000 mg의 RVG 엑소좀에서 캡슐화된 약 100 mg 내지 1000 mg의 핵산-표적화 시스템의 투여량은 본 발명을 위해 고려될 수 있다.

[0601] El-Andaloussi외 다수(Nature Protocols 7,2112-2126(2012))는 시험관 내 및 생체내 RNA의 전달을 위해 이용된 배양 세포로부터 유래된 엑소좀을 제공한다. 이 프로토콜은 펩티드 리간드와 융합된 엑소좀 단백질을 포함하는 발현 벡터의 트랜스펙션을 통한 표적화된 엑소좀의 생성을 처음으로 기재한다. 다음에, El-Andaloussi외 다수는 트랜스펙션된 세포 상청액으로부터 엑소좀을 정제하고 특성화하는 방법을 설명한다. 다음에, El-Andaloussi외 다수는 RNA를 엑소좀에 로딩하기 위한 결정적인 단계를 상술한다. 최종적으로, El-Andaloussi외 다수는 마우스 뇌에서 시험관 내 및 생체내 RNA를 효율적으로 전달하기 위해 엑소좀을 사용하는 방법을 약술한다. 엑소좀-매개 RNA 전달이 작용성 분석 및 영상에 의해 평가되는 예상된 결과의 예가 또한 제공된다.

전체 프로토콜은 약 3 주가 걸린다. 본 발명에 따른 전달 또는 투여는 자기 유래 수지상 세포로부터 생성된 엑소좀을 사용하여 수행될 수 있다. 본 명세서의 교시로부터, 이는 본 발명의 실시에 이용될 수 있다.

[0602] 다른 실시형태에서, Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130)의 혈장 엑소좀이 고려된다. 엑소좀은 수지상 세포(DC), B 세포, T 세포, 비만 세포, 상피 세포 및 종양 세포를 포함하는 다수의 세포 유형에 의해 생성된 나노 크기 소수포(30 내지 90 nm 크기)이다. 이들 소수포는 후기 엔도솜의 안쪽으로 향하는 출아에 의해 형성되며, 이어서 혈장막과 융합시 세포밖 환경으로 방출된다. 엑소좀은 천연적으로 세포 간에 RNA를 운반하기 때문에, 이 특성은 유전자 치료법에 유용할 수 있으며, 이러한 개시 내용으로

부터, 본 발명의 실시에 이용될 수 있다. 혈장으로부터의 엑소좀은 20 분 동안 900 g에서 버피 코트의 원심분리로 혈장을 분리시킨 다음에 세포 상청액을 수집하고 10 분 동안 300 g에서 그리고 30 분 동안 16 500 g에서 원심분리하여 세포를 제거한 후 0.22 mm 필터를 통해 여과시킴으로써 제조될 수 있다. 엑소좀은 70 분 동안 120 000g에서 초원심분리에 의해 펠렛화된다. siRNA의 엑소좀으로의 화학적 트랜스펙션은 RNAi Human/Mouse Starter Kit (Quiagen, Hilden, Germany)의 제조사 지침에 따라 수행된다. siRNA를 2 mmol/ml의 최종 농도로 100 ml PBS에 첨가하였다. HiPerFect 트랜스펙션 시약을 첨가한 후에, 혼합물을 실온에서 10 분 동안 인큐베이션시킨다. 과량의 미셀을 제거하기 위해, 알데하이드/설페이트 라텍스 비드를 사용하여 엑소좀을 재단리시킨다. 엑소좀 내로의 핵산-표적화 시스템의 화학적 트랜스펙션은 siRNA와 유사하게 수행될 수 있다. 엑소좀은 단핵구와 공동배양될 수 있고, 림프구는 건강한 공여체의 말초혈액으로부터 단리되었다. 따라서, 핵산-표적화 시스템을 함유하는 엑소좀은 인간의 단핵구 및 림프구에 도입될 수 있고 인간에게 자가 조직으로 재도입될 수도 있는 것이 고려될 수 있다. 따라서, 본 발명에 따른 전달 또는 투여는 혈장 엑소좀을 사용하여 실행될 수 있다.

[0603] **리포좀**

[0604] 본 발명에 따른 전달 또는 투여는 리포좀에 의해 수행될 수 있다. 리포좀은 내부 수성 격막을 둘러싸는 단일 또는 다중 라멜라 지질 이중층 및 상대적으로 불침투성인 외부 친유성 인지질 이중층으로 구성된 구체의 소수포 구조이다. 리포좀은 이들이 생체적합성, 비독성이기 때문에 약물 전달 담체로서 상당한 관심을 얻었고, 친수성과 친지성 약물 분자를 둘 다 전달할 수 있으며, 혈장 효소에 의한 분해로부터 그들의 카고를 보호하고, 생물학적 막 및 혈액 뇌 장벽(BBB)을 가로질러 그들의 로드를 전달할 수 있다(예를 들어, 검토를 위해 Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 참조). 리포좀은 여러 가지 다른 유형의 지질로 만들 수 있다; 그러나 인지질이 약물 담체로서 리포좀을 생성하는데 가장 일반적으로 사용된다. 지질 필름이 수용액과 혼합될 때 리포좀 형성은 자발적이지만, 또한 균질화기, 초음파처리기, 또는 압출 장치를 사용하는 것에 의한 진탕 형태로 힘을 적용함으로써 더 신속히 처리될 수 있다 (검토를 위해 참조, 예를 들어, Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679).

[0605] 그들의 구조 및 특성을 변형시키기 위해 몇몇 다른 첨가제가 리포좀에 첨가될 수 있다. 예를 들어, 콜레스테롤 또는 스핑고미엘린이 리포좀 혼합물에 첨가되어 리포좀 구조를 안정화시키고 리포좀 내부 화물의 누출을 방지한다. 추가로, 리포좀은 수소화된 난황 포스파티딜콜린 또는 난황 포스파티딜콜린, 콜레스테롤 및 디세틸 포스페이트로부터 제조되고, 그들의 평균 소수포 크기는 약 50 및 100 nm로 조절되었다. (검토를 위해 참조, 예를 들어, Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679). 리포좀 제형은 주로 천연 인지질 및 지질, 예컨대 1,2-디스테아로릴-sn-글리세로-3-포스파티딜 콜린(DSPC), 스핑고미엘린, 난 포스파티딜콜린 및 모노시알로강글리오사이드로 구성될 수 있다. 이 제형은 인지질 만으로 구성되기 때문에, 리포좀 제형은 다수의 도전, 그 중에서도 혈장 내 불안정성과 접하게 되었다. 특이적으로 지질막의조작에서 이들 도전을 극복하기 위한 몇몇 시도가 있었다.
이들 시도 중 하나는 콜레스테롤의 조작에 집중되어 있다. 통상의 제형으로의 콜레스테롤의 첨가는 캡슐화된 생활성 화합물의 혈장 내로의 신속한 방출을 감소시키거나 또는 1,2-디올레오일-sn-글리세로-3-포스포에탄올아민(DOPE)은 안정성을 증가시킨다(예를 들어, 검토를 위해 Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 참조). 특히 유리한 구현예에서, 트로이 목마 리포좀(또한 분자 트로이 목마로 알려져 있음)이 바람직하며, 프로토콜은 http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long에서 찾을 수 있다. 이들 입자는 혈관내 주사 후 전체 뇌에 이식유전자를 전달하게 한다. 제한에 의해 구속되지 않고, 표면에 콘쥬게이트된 특이적 항체를 지니는 중성 지질 입자는 엔도시토시스를 통해 혈액 뇌 장벽을 가로지르는 것으로 여겨진다. 본 발명자들은 혈관내 주사를 통해 뇌에 CRISPR-Cas13b 복합체를 전달하기 위해 트로이목마 리포좀을 이용하는 것을 상정하는데, 이는 배아 조작에 대한 필요 없이 전체 뇌 유전자이식 동물을 허용한다. 약 1 g 내지 5 g의 DNA 또는 RNA가 리포좀 내 생체내 투여를 위해 고려될 수 있다.

[0606] 또 다른 실시형태에서, 핵산-표적화 시스템 또는 이의 성분은 안정적인 핵산-지질 입자(SNALP)와 같은 리포좀으로 투여될 수 있다(예를 들어, Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005 참조) SNALP에서 표적화된 특이적 핵산-표적화 시스템의 약 1 mg/kg/일, 3 mg/kg/일 또는 5 mg/kg/일의 매일의 정맥내 주사가 고려된다. 일일 처리는 약 3일 넘게, 그 후 약 5주 동안 주 1회일 수 있다. 또 다른 실시형태에서, 약 1 mg/kg 또는 2.5 mg/kg의 용량에서 정맥내 주사에 의해 투여되는 특정 핵산-표적화 시스템 캡슐화된

SNALP가 또한 고려된다(참조, 예를 들어, Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006). SNALP 제형은 지질 3-N-[(w메톡시폴리(에틸렌 글리콜) 2000) 카바모일]-1,2-디미리스틸옥시-프로필아민(PEG-C-DMA), 1,2-디리놀레일옥시-N,N-디메틸-3-아미노프로판(DLinDMA), 1,2-디스테아로일-sn-글리세로-3-포스포콜린(DSPC) 및 콜레스테롤을 2:40:10:48 몰 백분율 비로 함유할 수 있다(참조, 예를 들어, Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006). 또 다른 구현예에서, 안정적인 핵산-지질 입자(SNALP)는 고도로 혈관화된 HepG2-유래 간 종양에 대해 효과적인 전달 분자인 것으로 입증되었지만, 불량하게 혈관화된 HCT-116 유래 간 종양에 대해서는 그렇지 않다(예를 들어, 문헌[Li, Gene Therapy (2012) 19, 775-780] 참조). SNALP 리포좀은 콜레스테롤/D-Lin-DMA/DSPC/PEG-C-DMA의 25:1 지질/siRNA 비 및 48/40/10/2 몰비를 사용하여 디스테아로일포스파티딜콜린(DSPC), 콜레스테롤 및 siRNA로 D-Lin-DMA 및 PEG-C-DMA를 제형화함으로써 제조될 수 있다. 얻어진 SNALP 리포좀은 크기가 약 80 내지 100 nm이다. 또 다른 구현예에서, SNALP는 합성 콜레스테롤(Sigma-Aldrich, 미국 미주리주 세인트루이스 소재), 디팔미토일포스파티딜콜린(Avanti Polar Lipids, 미국 앨라배마주 앨러배스터 소재), 3-N-[(w-메톡시 폴리(에틸렌 글리콜)2000)카바모일]-1,2-디미레스틸옥시프로필아민, 및 양이온성 1,2-디리놀레일옥시-3-N,N디메틸아미노프로판을 포함할 수 있다(참조, 예를 들어, Geisbert et al., Lancet 2010; 375: 1896-905). 예를 들어, 볼루스 정맥내 주입으로서 투여되는 용량당 약 2 mg/kg의 총 핵산-표적화 시스템의 투여량이 고려될 수 있다. 또 다른 실시형태에서, SNALP는 합성 콜레스테롤 (Sigma-Aldrich), 1,2-디스테아로일-sn-글리세로-3-포스포콜린 (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA 및 1,2-디리놀레일옥시-3-(N; N-디메틸)아미노프로판 (DLinDMA)을 포함할 수 있다(참조, 예를 들어, Judge, J. Clin. Invest. 119:661-673 (2009)). 생체내 연구를 위해 사용되는 제형은 약 9:1의 최종 지질/RNA 질량비를 포함할 수 있다.

[0607] RNAi 나노의학의 안전성 프로파일은 Alnylam Pharmaceuticals의 Barros 및 Gollob에 의해 검토되었다(예를 들어, 문헌[Advanced Drug Delivery Reviews 64 (2012) 1730-1737] 참조). 안정항 핵산 지질 입자(SNALP)는 4 가지 상이한 지질 - 낮은 pH에서 양이온성인 이온화 가능한 지질(DLinDMA), 중성 헬퍼 지질, 콜레스테롤 및 확산성 폴리 에틸렌 글리콜(PEG)-지질을 포함한다. 입자는 직경이 대략 80 nm이고, 생리학적 pH에서 하전 중성이다. 제형화 동안, 이온화 가능한 지질은 입자 형성 동안 음이온성 RNA와 지질을 축합하는 작용을 한다. 점점 더 산성인 엔도좀 조건 하에 양으로 하전될 때, 이온화 가능한 지질은 또한 RNA의 세포질 내로의 방출을 가능하게 하는 엔도좀 막과 SNALP의 융합을 매개한다. PEG-지질은 입자를 안정화하고, 제형 동안 응집을 감소시키며, 후속적으로 약동학적 특성을 개선시키는 외부의 중성 친수성을 제공한다. 지금까지, RNA가 있는 SNALP 제형을 사용하여 2 가지 임상 프로그램을 개시하였다. Tekmira Pharmaceuticals는 최근에 상승된 LDL 콜레스테롤을 지니는 성인 지원자에서 SNALP-ApoB의 I상 단일 용량 연구를 완료하였다. ApoB는 간 및 공장에서 대부분 발현되며, VLDL 및 LDL의 어셈블리 및 분비에 필수적이다. 17 명의 대상체에 단일 용량의 SNALP-ApoB를 투여하였다(7 용량 수준에 걸친 용량 상승). 간 독성의 증거는 없었다(전임상 연구에 기반한 잠재적 용량 제한 독성으로서 예상). 가장 높은 용량에서 (둘 중) 한 명의 대상체는 면역계 자극과 일치되는 감기-유사 증상을 경험하였고, 시험을 결론 짓는 결정을 하였다. Alnylam Pharmaceuticals는 상기 기재한 SNALP 기법을 사용하는 유사하게 진보된 ALN-TTR01을 가지며, TTR 아밀로이드증(ATTR)을 치료하기 위한 돌연변이체와 야생형 TTR 둘 다의 간세포 생성을 표적화한다. 3 가지 ATTR 증후군이 기재되었다:가족성 아밀로이드성 다발신경병증(FAP) 및 가족성 아밀로이드성 심근증(FAC) - 둘 다 TTR에서의 상염색체 우성 돌연변이에 의해 야기됨; 및 야생형 TTR에 의해 야기되는 노인 전신성 아밀로이드증(SSA). ALN-TTR01의 위약 제어, 단일 용량 상승 I상 시험을 ATTR을 지니는 환자에서 최근에 완료하였다. ALN-TTR01을 0.01 내지 1.0 mg/kg(siRNA 기준)의 용량 범위 내에서 31 명의 환자(연구 약물에 의한 23 명 및 위약에 의한 8 명)에 대한 15 분 IV 주입으로서 투여하였다. 간 기능 시험에서 유의한 증가 없이 치료는 잘 용인되었다. 주입-관련 반응은 0.4 mg/kg 이상에서 23 명의 환자 중 3 명에서 관찰되었으며; 모두 주입 속도가 느려지는 것에 반응했으며 모두 연구를 계속했다. 혈청 사이토카인 IL-6, IP-10 및 IL-1ra의 최소 및 일시적 상승은 1 mg/kg의 가장 높은 용량에서 2 명의 환자에서 주목되었다(전임상 및 NHP 연구로부터 참여). 혈청 TTR의 저하, ALN-TTR01의 예상된 약동학적 효과를 1 mg/kg에 관찰하였다.

[0608] 또 다른 구현예에서, SNALP는 양이온성 지질, DSPC, 콜레스테롤 및 PEG-지질을, 예를 들어 각각 40:10:40:10의 몰비로, 예를 들어 에탄올 중에 용해시킴으로써 제조될 수 있다(참조, Semple et al., Nature Niotechnology, Volume 28 Number 2 February 2010, pp. 172-177). 지질 혼합물을 30% (vol/vol) 및 6.1 mg/ml의 최종 에탄올 및 지질 농도로 각각 혼합하면서 수성 완충제(50 mM 시트르산염, pH 4)에 첨가하였고, 2 분 동안 22℃에서 평형상태로 만든 후 압출시켰다. 수화된 지질을 역동적 광 산란에 의해 결정된 바와 같은 소수포 직경 70 내지 90 nm까지 Lipex Extruder(Northern Lipids)를 사용하여 22℃에서 2층 80 nm 기공 크기 필터(Nuclepore)

를 통해 압출하였고, 얻었다. 이는 일반적으로 1 내지 3 회 통과를 필요로 하였다. siRNA(50 mM 시트르산염 중에서 가용화됨, 30% 에탄올을 함유하는 pH 4 수용액)를 혼합하면서 약 5 ml/분의 속도로 사전 평형상태로 만든(35℃) 소수포에 첨가하였다. 0.06(wt/wt)의 최종 표적 siRNA/지질 비에 도달한 후, 혼합물을 35℃에서 추가 30 분 동안 인큐베이션시켜 소낭 재조직화 및 siRNA의 캡슐화를 가능하게 하였다. 이어서, 에탄올을 제거하고, 투석 또는 접선 유동 정용여과에 의해 외부 완충제를 PBS로 대체하였다(155 mM NaCl, 3 mM Na_2HPO_4, 1 mM KH_2PO_4, pH 7.5). 통제된 단계적 희석법 공정을 사용하여 siRNA를 SNALP에 캡슐화하였다.

KC2-SNALP의 지질 성분은 57.1:7.1:34.3:1.4의 몰비로 사용된 DLin-KC2-DMA (양이온 지질), 디팔미토일포스파티딜콜린 (DPPC; Avanti Polar Lipids), 합성 콜레스테롤 (Sigma) 및 PEG-C-DMA였다. 로딩된 입자의 형성 시, SNALP를 PBS에 대해 투석하였고, 사용 전 0.2 μm 필터를 통해 멸균여과시켰다. 평균 입자 크기는 75 내지 85 nm였고, siRNA의 90 내지 95%는 지질 입자 내에서 캡슐화되었다. 생체내 시험을 위해 사용한 제형 내 최종 siRNA/지질비는 약 0.15(wt/wt)였다. 인자 VII siRNA를 함유하는 LNP-siRNA 시스템을 사용 직전에 멸균 PBS 중에서 적절한 농도로 희석시켰고, 제형을 총 10 ml/kg 용적으로 옆 꼬리 정맥을 통해 정맥내로 투여하였다. 이러한 방법 및 이들 전달 시스템은 본 발명의 핵산-표적화 시스템에 대해 추론될 수 있다.

[0609] **다른 지질**

[0610] 다른 양이온성 지질, 예컨대 아미노 지질 2,2-디리놀레일-4-디메틸아미노에틸-[1,3]-디옥솔란(DLin-KC2-DMA)은 예를 들어, siRNA와 유사한 핵산-표적화 시스템 또는 이의 성분 또는 이를 코딩하는 핵산 분자(들)를 캡슐화하기 위해 이용될 수 있으며(예를 들어, 문헌 [Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 - 8533] 참조), 이런 이유로, 본 발명의 실시에 이용될 수 있다. 다음의 지질 조성물을 이용하여 사전 형성한 소수포가 고려될 수 있다:40/10/40/10의 몰비 및 대략0.05(w/w) FVII siRNA/총 지질비로 각각 아미노지질, 디스테아로일포스파티딜콜린(DSPC), 콜레스테롤 및 (R)-2,3-비스(옥타데실옥시) 프로필-1-(메톡시 폴리(에틸렌 글리콜)2000)프로필카바메이트(PEG-지질). 70 nm 내지 90 nm 범위의 좁은 입자 크기 분포 및 0.11±0.04(n=56)의 낮은 다분산지수를 보장하기 위해, 입자를 가이드 RNA를 첨가하기 전에 80 nm 막을 통해 최대 3 회까지 압출시킬 수 있다. 고도로 강한 아미노 지질 16(생체내 활성을 활성화하도록 추가로 최적화될 수 있는 4 가지 지질 구성성분 16, DSPC, 콜레스테롤 및 PEG-지질의 몰비(50/10/38.5/1.5))을 함유하는 입자가 사용될 수 있다.

[0611] Michael S D Kormann외 다수(문헌["Expression of therapeutic proteins after delivery of chemically modified mRNA in mice:Nature Biotechnology, Volume:29, Pages:154-157 (2011)])는 RNA를 전달하기 위한 지질 외피의 사용을 기재한다. 지질 외피의 사용은 또한 본 발명에서 바람직하다.

[0612] 또 다른 구현예에서, 지질은 본 발명의 RNA-표적화 시스템 (CRISPR-Cas13b 복합체,즉, crRNA와 복합체를 형성한 Cas13b) 또는 이의 성분(들) 또는 이를 코딩하는 핵산 분자(들)와 함께 제형화되어 지질 나노입자(LNP)를 형성할 수 있다. 지질은 DLin-KC2-DMA4, C12-200 및 공지질 디스테로일포스파티딜 콜린을 포함하지만 이들로 제한되지 않으며, 콜레스테롤, 및 PEG-DMG는 자발적 소낭 형성 절차를 사용하여 siRNA 대신 핵산-표적화 시스템을 이용하여 제형화될 수 있다(참조, 예를 들어, e.g., Novobrantseva, Molecular Therapy-Nucleic Acids (2012) 1, e4; doi:10.1038/mtna.2011.3). 구성성분 몰비는 약 50/10/38.5/1.5(DLin-KC2-DMA 또는 C12-200/디스테로일포스파티딜 콜린/콜레스테롤/PEG-DMG)일 수 있다. 최종 지질:siRNA 중량비는 DLin-KC2-DMA 및 C12-200 지질 입자(LNP)의 경우에 각각 약 12:1 및 9:1일 수 있다. 제형은 90% 초과의 포착 효율을 지니는 약 80 nm의 평균 입자 직경을 가질 수 있다. 3 mg/kg 용량이 고려될 수 있다. Tekmira는 LNP 및 LNP 제제의 다양한 측면에 관한 미국 및 해외에 약 95 개의 특허 패밀리 포트폴리오를 보유하고 있고 (참조, 예를 들어, 미국 특허번호 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 및 7,838,658 및 유럽 특허 번호 1766035; 1519714; 1781593 및 1664316), 이들 모두는 본 발명에 사용 및/또는 적응될 수 있다.

[0613] RNA-표적화 시스템 또는 이의 성분 또는 이를 코딩하는 핵산 분자(들)는 단백질, 단백질 전구체, 또는 단백질 또는 단백질 전구체의 부분적으로 또는 완전히 가공된 형태를 인코딩할 수 있는 변형된 핵산 분자를 포함하는 조성물의 제형의 양태에 관한 미국 출원 공개 20130252281 및 20130245107 및 20130244279(Moderna Therapeutics 출원)에서 추가로 기재된 것과 같은 PLGA 미소구체에서 캡슐화되어 전달될 수 있다. 제형은 몰비 50:10:38.5:1.5-3.0(양이온성 지질:융합생성 지질:콜레스테롤:PEG 지질)을 가질 수 있다. PEG 지질은 PEG-c-DOMG, PEG-DMG로부터 선택될 수 있지만, 이들로 제한되지 않는다. 융합생성 지질은 DSPC일 수 있다.

또한 Schrum 등의 미국 출원 공개 제20120251618호(Delivery and Formulation of Engineered Nucleic Acids)를 참조한다.

[0614] 나노머 기술은 저분자량 소수성 약물, 펩티드 및 핵산 기반 치료제(플라스미드, siRNA, miRNA)를 포함하는 광범위 치료제에 대한 생체이용가능성 도전을 처리한다. 기술이 명확한 이점을 입증한 구체적 투여 경로는 경구 경로, 혈액뇌장벽에 걸친 수송, 고형 종양에 대한 전달뿐만 아니라 눈에 대한 전달을 포함한다. 참조, 예를 들어, Mazza et al., 2013, ACS Nano. 2013 Feb 26;7(2):1016-26; Uchegbu and Siew, 2013, J Pharm Sci. 102(2):305-10 및 Lalatsa et al., 2012, J Control Release. 2012 Jul 20; 161(2):523-36.

[0615] 미국 특허 공개 제20050019923호는 폴리뉴클레오티드 분자, 펩티 및 폴리펩티드 및/또는 약제학적 작용제(agent)와 같은 생활성 분자를 포유류 신체에 전달하기 위한 양이온성 덴드리머를 기재한다. 덴드리머는, 예를 들어 간, 비장, 폐, 신장 또는 심장(또는 심지어 뇌)으로의 생활성 분자의 전달을 표적화하는 데 적합하다. 덴드리머는 단순한 분지 단량체 단위로부터 단계적 방식으로 제조된 합성 3-차원 거대분자이며, 이의 특성 및 기능성은 용이하게 제어되고 변할 수 있다. 덴드리머는 다작용성 코어에 대한(합성에 대한 확산적 접근), 또는 다작용성 코어를 향한(합성에 대한 수렴적 접근) 빌딩 블록의 반복된 첨가로부터 합성되고, 빌딩 블록의 3차원 쉘의 각각의 첨가는 덴드리머의 더 고차 세대의 형성을 야기한다. 폴리프로필렌이민 덴드리머는 디아미노부탄 코어로부터 출발하며, 이것에 일차 아민으로의 아크릴로니트릴의 이중 마이클 첨가(Michael addition)에 의해 아미노기 수의 2 배가 첨가된 후, 니트릴이 수소화된다. 이는 아미노기의 배증을 초래한다. 폴리프로필렌이민 덴드리머는 100% 양성자화 가능한 질소 및 64 개까지의 말단 아미노기를 함유한다(5세대, DAB 64). 양성자화 가능한기는 보통 중성 pH에서 양성자를 수용할 수 있는 아민기이다. 유전자 전달제로서 덴드리머의 용도는 콘쥬게이팅 단위로서 각각 아민/아미드 또는 N--P(O_2)S의 혼합물과 함께 폴리아미도아민 및 인 함유 화합물의 용도에 크게 집중되어 있었는데 유전자 전달을 위한 더 낮은 세대 폴리프로필렌이민 덴드리머의 용도에 대해 보고한 연구는 없었다. 폴리프로필렌이민 덴드리머는 또한 주위 아미노기에 의해 화학적으로 변형될 때 약물 전달을 위해 그리고 게스트 분자의 캡슐화를 위해 pH 민감 제어 방출 시스템으로서 연구되었다. 세포독성 및 DNA와 폴리프로필렌이민 덴드리머의 상호작용뿐만 아니라 DAB 64의 트랜스펙션 효능이 또한 연구되었다. 미국 특허 공개 제20050019923호는 더 조기의 보고와 상반되게 양이온성 덴드리머, 예컨대 폴리프로필렌이민 덴드리머가 유전물질과 같은 생활성 분자의 표적화된 전달에서 사용을 위한 적합한 특성, 예컨대 특이적 표적화 및 낮은 독성을 나타낸다는 관찰에 기반한다. 추가로, 양이온성 덴드리머의 유도체는 또한 생활성 분자의 표적화된 전달을 위한 적합한 특성을 나타낸다. 또한, 양이온성 폴리아민 중합체 및 덴드리머 중합체를 포함하는 "다양한 중합체"를 개시하는 미국 특허 공개 제20080267903호의 생활성 중합체는 항증식 활성을 갖는 것으로 나타나며, 따라서 신생물 및 종양과 같은 바람직하지 않은 세포 증식, 염증 장애(자가면역 장애를 포함), 건선 및 아테롬성 동맥 경화증을 특징으로 하는 장애의 치료에 유용할 수 있다. 중합체는 단독으로 활성제로서 또는 다른 치료제용 전달 비히클, 예컨대 유전자 요법을 위한 약물 분자 또는 핵산으로서 사용될 수 있다. 이러한 경우에, 중합체 자체의 고유 항종양 활성은 전달될 작용제의 활성을 보완할 수 있다"고 개시한다. 이들 특허 간행물의 개시 내용은 핵산-표적화 시스템(들) 또는 이의 성분(들) 또는 이를 코딩하는 핵산 분자(들)의 전달을 위한 본 발명의 교시와 함께 이용될 수 있다.

[0616] **초하전(supercharged) 단백질**

[0617] 초하전 단백질은 대단히 높은 양 또는 음의 이론적 순전하를 지니는 조작된 또는 천연적으로 발생하는 단백질의 부류이며, 핵산-표적화 시스템(들) 또는 이의 성분(들) 또는 이를 코딩하는 핵산 분자(들)의 전달에 이용될 수 있다. 초 음으로 하전된 단백질과 초 양으로 하전된 단백질은 둘 다 열적으로 또는 화학적으로 유도된 응집을 견뎌내는 현저한 능력을 나타낸다. 초 양으로 하전된 단백질은 또한 포유류 세포를 침투할 수 있다. 카고와 이들 단백질, 예컨대 플라스미드 DNA, RNA 또는 기타 다른 단백질의 연합은 시험관 내와 생체내 둘 다에서 포유동물 세포 내로 이들 거대분자의 작용성 전달을 가능하게 할 수 있다. David Liu 연구소는 2007년에 초 하전 단백질의 생성 및 특성규명을 보고하였다(Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110-10112).

[0618] 포유동물 세포 내로의 RNA 및 플라스미드 DNA의 비바이러스 전달은 연구와 치료 응용 둘 다에 대해 가치가 있다(Akinc et al., 2010, Nat. Biotech. 26, 561-569). 정제된 +36 GFP 단백질(또는 기타 다른 초양성으로 하전된 단백질)은 적절한 무혈청 배지에서 RNA와 혼합되고, 세포에 첨가 전 복합체화되게 한다. 이 단계에서 혈청의 포함은 초하전 단백질-RNA 복합체의 형성을 억제하고, 치료 효능을 감소시킨다. 다음의 프로토콜은 다양한 세포주에 대해 효과적이 되는 것으로 관찰되었다(McNaughton et al., 2009, Proc. Natl. Acad. Sci. USA 106, 6111-6116). 그러나, 단백질 및 RNA의 용량을 변화시키는 파일럿 실험은 특정 세포주에 대하여 절차를 최적화하도록 수행되어야 한다: (1) 치료 하루 전, 48-웰 플레이트에 웰당 1 × 10^5 개 세포를 도말함.

(2) 치료일에, 최종 농도 200 nM로 무혈청 배지에서 정제된 +36 GFP 단백질을 희석시킴. RNA를 50 nM의 최종 농도로 첨가함. 와류시켜 혼합하고 실온에서 10 분 동안 인큐베이션시킴. (3) 인큐베이션 동안, 세포로부터 배지를 흡입하고, PBS로 1 회 세척함. (4) +36 GFP 및 RNA의 인큐베이션 후에, 단백질-RNA 복합체를 세포에 첨가함. (5) 37℃에서 4 시간 동안 복합체와 세포를 인큐베이션시킴. (6) 인큐베이션 후에, 배지를 흡입하고, 20 U/㎖ 헤파린 PBS로 3 회 세척함. 활성에 대한 어세이에 따라 추가 48 시간 또는 그 이상 동안 혈청-함유 배지와 세포를 인큐베이션시킴. (7) 세포를 면역블롯, qPCR, 표현형 분석 또는 기타 다른 적절한 방법에 의해 분석함.

[0619] 다양한 세포에서 +36 GFP가 효과적인 플라스미드 전달 시약인 것으로 추가로 발견하였다. 또한 참조, 예를 들어, McNaughton et al., Proc. Natl. Acad. Sci. USA 106, 6111-6116 (2009); Cronican et al., ACS Chemical Biology 5, 747-752 (2010); Cronican et al., Chemistry & Biology 18, 833-838 (2011); Thompson et al., Methods in Enzymology 503, 293-319 (2012); Thompson, D.B., et al., Chemistry & Biology 19 (7), 831-843 (2012). 초-하전 단백질의 방법은 본 발명의 RNA-표적화 시스템(들) 또는 이의 성분(들) 또는 이를 코딩하는 핵산 분자(들)의 전달에 사용 및/또는 적응될 수 있다.

[0620] **세포 투과성 펩티드(CPP)**

[0621] 또 다른 구현예에서, 세포 투과성 펩티드(CPP)가 CRISPR Cas 시스템의 전달을 위해 고려된다. CPP는 다양한 분자 카고(나노크기 입자로부터 작은 화학 분자 및 DNA의 큰 단편까지)의 세포 흡수를 용이하게 하는 짧은 펩티드이다. 본 명세서에서 사용된 바와 같은 용어 "카고"는 치료제, 진단 프로브, 펩티드, 핵산, 안티센스 올리고뉴클레오티드, 플라스미드, 단백질, 입자(나노입자를 포함함), 리포좀, 발색단, 소분자 및 방사성 물질로 이루어진 군을 포함하지만 이에 제한되지 않는다. 본 발명의 양태에서, 카고는 또한 CRISPR Cas 시스템의 임의의 성분 또는 전체 작용성 CRISPR Cas 시스템을 포함할 수 있다. 본 발명의 양태는 하기를 포함하는, 원하는 카고를 대상체로 전달하는 방법을 제공한다:(a) 본 발명의 세포 투과성 펩티드 및 원하는 카고를 포함하는 복합체를 제조하는 단계, 및 (b) 대상체에게 복합체를 경구, 관절내, 복강내, 척추강내, 동맥내, 비강내, 뇌실내, 피하내, 근육내, 정맥내, 피부, 직장내, 또는 국소투여하는 단계. 카고는 공유 결합을 통한 화학적 결합 또는 비-공유 상호작용을 통해 펩티드와 연합된다. CPP의 기능은 카고를 세포로 전달하는 것이며, 살아있는 포유동물 세포의 엔도솜으로 전달된 카고의 엔도시토시스를 통해 일반적으로 발생하는 공정이다. 세포-침투성 펩티드는 상이한 크기, 아미노산 서열, 및 전하를 가지지만, 모든 CPP가, 혈장 막을 전위시키고 세포질 또는 세포소기관으로 다양한 분자 카고의 전달을 용이하게 하는 능력인 하나의 별개 특징을 갖는다. CPP 전위는 3 개의 주요 엔트리 메커니즘으로 분류될 수 있다:막에서의 직접 투과, 엔도시토시스-매개 엔트리, 및 일시적 구조의 형성을 통한 전위. CPP는 암을 포함하는 상이한 질병의 치료에서 약물 전달제, 바이러스 억제제뿐만 아니라, 세포 표지화를 위한 조영제로서 의약에서의 많은 적용이 발견되었다. 후자의 예는 GFP에 대한 담체, MRI 조영제, 또는 양자점으로서 작용하는 것을 포함한다. CPP는 연구 및 의약에서의 용도에서 시험관내 및 생체내 전달 벡터로서 훌륭한 잠재성을 보유한다. CPP는 통상적으로 리신 또는 아르기닌과 같은 상대적으로 풍부한 양전하 아미노산을 함유하거나 극성/전하 아미노산 및 비-극성 소수성 아미노산의 교대 패턴을 함유하는 서열을 갖는 아미노산 조성물을 갖는다. 이들 2 가지 유형의 구조는 각각 다가이온성 또는 양친매성으로서 지칭된다. CPP의 세번째 부류는 낮은 순 전하를 갖는 극성 잔기만을 함유하는 소수성 펩티드이거나, 세포 흡수에 결정적인 소수성 아미노산 그룹을 갖는다. 발견된 초기 CPP 중 하나는 배양에서 많은 세포 유형에 의해 배지 주위로부터 효율적으로 취득되어 발견되는 인간 면역결핍 바이러스 1(HIV-1)로부터의 트랜스-활성화 전사 활성화제(Tat)였다. 그 이후부터, 많은 공지된 CPP는 상당히 확장되었으며 더 효과적인 단백질 형질도입 특성을 갖는 소분자 합성 유사체가 생성되었다. CPP는 페너트라틴, Tat(48-60), 트란스포르탄, 및 (R-AhX-R4)(Ahx=아미노헥사노일)을 포함하지만 이에 제한되지 않는다.

[0622] 미국 특허 8,372,951은 높은 세포-투과 효율 및 낮은 독성을 나타내는 호산구 양이온성 단백질(ECP)으로부터 유래된 CPP를 제공한다. 척추동물 대상체로 CPP를 이의 카고와 함께 전달하는 양태가 또한 제공된다. CPP 및 이의 전달의 추가 양태는 미국 특허 8,575,305; 8;614,194 및 8,044,019 에 기재된다. CPP는 CRISPR-Cas 시스템 또는 이의 성분을 전달하기 위해 사용될 수 있다. CPP가 CRISPR-Cas 시스템 또는 이의 성분을 전달하기 위해 이용될 수 있는 것은 또한 본원에 전체가 참고로 포함된, Suresh Ramakrishna, Abu-Bonsrah Kwaku Dad, Jagadish Beloor외 다수의 문헌["Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA", Gees.nome R 2014 Apr 2. [인쇄 전 전자논문상태]]에서 제공되며, 여기에서 CPP-콘쥬게이트된 재조합 Cas9 단백질 및 CPP-복합체화 가이드 RNA를 이용한 치료가 인간 세포주에서의 내인성 유전자 파괴를 이끌어냄이 입증되었다. 이 논문에서 Cas9 단백질은 티오에스테르 결합을 통해 CPP에 콘쥬게이트되는 반면, 가이드 RNA는 CPP와 복합체화되어, 축합된 양으로 하전된 입자를 형성한다. 변

형된 Cas9 및 가이드 RNA와 함께, 배아 줄기 세포, 피부 섬유아세포, HEK293T 세포, HeLa 세포, 및 배아 암종 세포를 포함하는 인간 세포의 동시 및 순차적 치료가 플라스미드 트랜스펙션에 비하여 감소된 표적외 돌연변이로 효율적인 유전자 파괴를 이끌어냄을 나타내었다. CPP 전달이 본 발명의 실시에서 사용될 수 있다.

[0623] **이식가능한 장치**

[0624] 또 다른 구현예에서, RNA-표적화 시스템 또는 이의 성분(들) 또는 이를 코딩하는 핵산 분자(들)의 전달을 위해 이식 가능한 장치가 또한 고려된다. 예를 들어, 미국 특허 공개 20110195123은 국소적으로 그리고 연장된 기간 동안 약물을 용리하는 이식 가능한 의료 장치(이러한 장치의 몇 가지 유형, 실행의 처리 방식 및 이식 방법을 포함함)가 제공됨을 개시한다. 장치는 중합체성 기질, 예를 들어 장치 몸체로서 사용되는 매트릭스, 및 약물, 및 일부 경우에 추가적인 스캐폴딩 재료, 예컨대 금속 또는 추가적인 중합체, 및 가시성 및 영상화를 향상시키는 재료를 포함한다. 이식 가능한 전달 장치는 국소적인, 그리고 연장된 기간에 걸친 방출을 제공하는 데 유리할 수 있으며, 여기서 약물은, 예컨대 종양, 염증, 변성과 같은 병에 걸린 영역의 세포외 매트릭스(ECM)로 직접 방출되거나, 증상 목표를 위해, 또는 손상된 평활근 세포에 또는 예방을 위해 방출된다. 한 종류의 약물은 상기 개시된 바와 같은 RNA이며, 이러한 시스템은 본 발명의 핵산-표적화 시스템에 사용될 수 있고/있거나 적합하게 될 수 있다. 일부 구현예에서 이식 방식은 근접 치료법 및 바늘 생검을 포함하는 기타 다른 치료를 위해 오늘날 개발되고 사용되는 기존의 이식 절차이다. 이러한 경우에, 본 발명에 기재된 새로운 이식물의 치수는 본래의 이식물과 유사하다. 전형적으로 소수의 장치는 동일한 치료 절차 동안 이식된다. 미국 특허 공개 20110195123은, 복강과 같은 강에 적용 가능한 시스템을 포함하는 약물 전달 이식 가능한 또는 삽입 가능한 시스템 및/또는 임의의 기타 다른 유형의 투여를 제공하며, 여기서, 예를 들어 선택적으로 매트릭스일 수 있는 생체안정성 및/또는 분해가능 및/또는 생체흡수성 중합체성 기질을 포함하는 약물 전달 시스템은 고정 또는 부착되지 않는다. 용어 "삽입"은 또한 이식을 포함하는 것으로 언급되어야 한다.

약물 전달 시스템은 바람직하게는 미국 특허 공개 제20110195123호에 기재되어 있는 바와 같은 "로더(Loder)"로서 실행된다. 중합체 또는 복수의 중합체는 생체적합성이며, 제어된 속도로 작용제를 방출시킬 수 있는 작용제 및/또는 복수의 작용제를 혼입하며, 중합체성 기질, 예컨대 일부 구현예에서 매트릭스의 총 부피는, 선택적으로 그리고 바람직하게는 작용제의 치료 수준이 도달되게 하는 최대 부피 이하이다. 비제한적 예로서, 이러한 용적은 작용제 장입을 위한 용적으로 필요한 바와 같은 바람직하게는 0.1 m^3 내지 1000 mm^3 범위 내에 있다. 로더는, 예를 들어 크기가 작용성에 의해 결정되는 장치, 예를 들어 제한 없이, 무릎 관절, 자궁내 또는 자궁경부 링 등과 함께 혼입될 때, 선택적으로 더 클 수 있다. 약물 전달 시스템 (조성물 전달을 위한)은 일부 실시형태에서 바람직하게는 분해성 중합체를 사용하도록 디자인되며, 여기서 주요 방출 메커니즘은 벌크 침식이고; 또는 일부 구현예에서, 비분해성 또는 서서히 분해되는 중합체가 사용되며(주요 방출 메카니즘은 벌크 침식보다는 확산임), 따라서 외부는 막으로서 기능하고, 내부는 약물 저장소로서 기능하는데, 이는 연장된 기간(예를 들어, 약 1 주 내지 약 수 개월) 동안 주변에 의해 실제로 영향받지 않는다. 상이한 방출 메카니즘을 지니는 상이한 중합체의 조합이 또한 선택적으로 사용될 수 있다. 표면에서 농도 기울기는 바람직하게 총 약물 방출 기간의 상당한 기간 동안 효과적으로 일정하게 유지되고, 따라서 확산 속도는 효과적으로 일정하다("제로 방식" 확신으로 지칭됨). 용어 "일정한"은 더 낮은 치료적 유효성의 역치 초과로 바람직하게 유지되지만, 여전히 선택적으로 초기 버스트를 특징으로 하고/하거나 변동할 수 있는, 예를 들어 특정 정도로 증가 및 감소할 수 있는 확산 속도를 의미한다. 확산 속도는 바람직하게는 연장된 기간 동안 이렇게 유지되며, 치료적으로 효과적인 기간, 예를 들어 효과적인 침묵 기간을 최적화하는 특정 수준으로 일정하게 고려될 수 있다. 약물 전달 시스템은 선택적으로 그리고 바람직하게는 성질이 화학물질이든 대상체의 체내의 효소 및 기타 다른 인자로부터의 공격에 기인하든, 분해로부터 뉴클레오티드 기반 치료제를 보호하도록 디자인된다. 미국 특허 공개 20110195123의 약물 전달 시스템은 장치의 이식 시 및/또는 이식 후에, 활성화 및/또는 가속화/감속화의 비침습적 및/또는 최소 침습적 방법에 의해 작동되는 감지 및/또는 활성화 기기와 선택적으로 연동되며, 예를 들어 선택적으로 제한되는 것은 아니지만, 열적인 가열 및 냉각, 레이저 빔 및 집속 초음파 및/또는 RF(무선주파수)를 포함하는 초음파 방법 또는 장치를 포함한다. 미국 특허 공개 20110195123의 일부 구현예에 따르면, 국소 전달을 위한 부위는 선택적으로 종양을 포함하는 세포의 고도 비정상 증식, 및 억제된 아폽토시스, 자가면역질병 상태를 포함하는 활성 및 또는 만성 염증 및 감염, 근육 및 신경 조직을 포함하는 변성 조직, 만성 통증, 변성 부위, 및 뼈 골절 위치 및 조직 재생의 향상을 위한 기타 다른 상처 위치, 및 손상된 심근, 평활근 및 횡문근을 특징으로 하는 표적 부위를 포함할 수 있다. 조성물의 이식을 위한 부위, 또는 표적 부위는 바람직하게는 표적화된 국소 전달에 충분히 작은 반경, 면적 및/또는 용적을 특징으로 한다. 예를 들어, 표적 부위는 임의로 약 0.1 mm 내지 약 5 cm 범위의 직경을 갖는다. 표적 부위의 위치는 바람직하게는 최대 치료 효능을 위해 선택된다. 예를 들어, 약물 전달 시스템의 조성물(선택적으로 상

기 기재한 바와 같은 이식을 위한 장치와 함께)은 선택적으로 그리고 바람직하게는 종양 환경, 또는 이들의 연관된 혈액 공급 내에서 또는 이에 근접하여 이식된다. 예를 들어 조성물은 (선택적으로 장치와 함께) 선택적으로 췌장, 전립선, 유방, 간 내에 또는 그에 근접하여, 니플을 통해, 혈관계 내 등에 이식된다. 표적 위치는 (비-제한적인 예로서, 신체 내의 임의의 부위가 Loder를 이식하기에 적합할 수 있으므로) 다음을 포함하거나, 이로 이루어지거나 또는 이것으로 본질적으로 이루어진 군으로부터 선택적으로 선택된다: 1. 기저핵, 백질 및 회백질의 파킨슨 병 또는 알츠하이머 병과 같은 퇴행성 부위의 뇌; 2. 근위축성 측삭 경화증 (ALS)의 경우와 같은 척추; 3. HPV 감염을 예방하는 자궁 경부; 4. 활성 및 만성 염증성 관절; 5. 건선의 경우와 같은 피부; 6. 진통 효과에 대한 교감 및 감각 신경 부위; 7. 골내 이식; 8. 급성 및 만성 감염 부위; 9. 질내; 10. 내이-청각 시스템, 내이의 미로, 전정 시스템; 11. 기관 내; 12. 심장 내; 관상동맥, 심장외막; 13. 방광; 14. 담도 시스템; 15. 신장, 간, 비장을 포함하나 이에 제한되지 않는 실질 조직; 16. 림프절; 17. 침샘; 18. 잇몸; 19. 관절내 (관절 내로); 20. 안구내; 21. 뇌조직; 22. 뇌실; 23. 복강을 포함하는 강(예를 들어, 제한되는 것은 아니지만, 난소암); 24. 내부식도 및 25. 직장내.

[0625] 선택적으로 시스템의 삽입(예를 들어, 조성물을 함유하는 장치)은 표적 부위의 및 이러한 부위 부근의, 국소 pH 및/또는 온도 및/또는 ECM 내 약물의 확산 및/또는 약물 역학에 영향을 미치는 기타 다른 생물학적 인자에 영향을 미치도록 하는 표적 부위 및 해당 부위 부근에서 ECM에 대한 물질의 주사와 연관된다. 선택적으로, 일부 구현예에 따르면, 상기 작용제의 방출은 삽입 전 및/또는 삽입 시 및/또는 삽입 후, 레이저빔, 조사, 열적 가열 및 냉각, 및 집속 초음파 및/또는 RF(무선주파수)를 포함하는 초음파 방법 또는 장치 및 화학적 활성화제를 포함하는 활성화 및/또는 가속화/감속화의 비침습적 및/또는 최소 침습적 및/또는 그 밖의 방법에 의해 작동되는 감지 및/또는 활성화 기기와 연관될 수 있었다.

[0626] 본 발명의 실시에 사용될 수 있는 미국 특허 공개 20110195123의 구현예에 따르면, 약물은 바람직하게, 예를 들어 이하에 기재되는 바와 같이 유방, 췌장, 뇌, 신장, 방광, 폐 및 전립선에서 국재화된 암 경우를 위한 RNA를 포함한다. RNAi로 예시되지만, 약물이 로더 기질, 예컨대 매트릭스와 함께 캡슐화될 수 있는 한, 다수 약물이 로더에서 캡슐화되어 적용 가능하고, 본 발명과 함께 사용될 수 있으며, 이러한 시스템이 본 발명의 핵산-표적화 시스템을 전달하는 데 사용되고/사용되거나 그에 적합하게 될 수 있다. 구체적 적용의 또 다른 예로서, 신경 및 근육 변성 질병은 비정상 유전자 발현에 기인하여 발생한다. RNA의 국소 전달은 이러한 비정상 유전자 발현을 방해하는 치료적 특성을 가질 수 있다. 작은 약물 및 거대분자를 포함하는 아폽토시스방지, 염증방지 및 변성방지 약물의 국소 전달은 또한 선택적으로 치료적일 수 있다. 이러한 경우에, 로더는 일정한 속도로 및/또는 별도로 이식되는 전용 장치를 통해 연장된 방출을 위해 적용된다.

[0627] 이것 모두는 본 발명의 RNA-표적화 시스템에 사용될 수 있고/있거나 적합하게 될 수 있다. 본 명세서에서 논의된 이식가능한 장치 기술이 본 명세서의 교시와 함께 이용될 수 있으며, 이에 따라 본 개시 내용 및 당업계의 지식에 의해, CRISPR-Cas13b 시스템 또는 복합체 또는 이의 성분 또는 이의 핵산 분자 또는 성분의 인코딩 또는 제공이 이식가능한 장치를 통해 전달될 수 있다.

[0628] **환자-특이적 스크리닝 방법**

[0629] RNA를 표적화하는 핵산-표적화 시스템을 사용하여 특정 RNA의 존재에 대하여 환자 또는 환자 샘플을 스크리닝할 수 있다.

[0630] **CRISPR 이펙터 단백질 mRNA 및 가이드 RNA**

[0631] CRISPR 이펙터 (Cas13b) 단백질 또는 그에 대한 mRNA (또는 더 일반적으로 이를 위한 핵산 분자) 및 가이드 RNA 또는 crRNA는 또한 예를 들어, 가이드 RNA 또는 crRNA, 또는 함께 투여하기 전, 이전 1 내지 12 시간 (바람직하게는 대략 2 내지 6 시간) 개별적으로 전달될 수 있다. 가이드 RNA 또는 crRNA의 제2 부스터 용량은 초기 투여 1 시간 내지 12 시간(바람직하게는, 약 2 시간 내지 6 시간) 이후에 투여될 수 있다.

[0632] Cas13b 이펙터 단백질은 때때로 본 명세서에서 CRISPR 효소로 지칭된다. 이펙터 단백질은 효소를 기반으로 하거나 효소로부터 유래되어, 따라서 용어 '이펙터 단백질'은 일부 구현예에서 확실히 '효소'를 포함하는 것으로 이해될 것이다. 그러나, 또한 이펙터 단백질은, 일부 실시형태에서 요구되는 바와 같이, DNA 또는 RNA 결합을 가질 수 있지만, 반드시 데드-Cas 이펙터 단백질 기능을 포함하여 절단 또는 닉킹, 활성을 필요로 하는 것은 아님을 이해할 것이다.

[0633] 세포 표적은 조혈 줄기/전구 세포 (CD34 +); 인간 T 세포; 및 눈 (망막 세포) - 예를 들어 광수용기 전구체 세포를 포함한다.

[0634] 본 발명의 방법은 주형의 전달을 추가로 포함할 수 있다. 주형의 전달은 임의의 또는 모든 CRISPR 이펙터 단백질 (Cas13b) 또는 가이드 또는 crRNA와 함께 또는 그와 분리되어, 그리고 동일하거나 상이한 전달 메커니즘을 통해 전달될 수 있다.

[0635] **유도성 시스템**

[0636] 일부 실시형태에서, CRISPR 이펙터 (Cas 13n) 단백질은 유도성 시스템의 성분을 형성할 수 있다. 시스템의 유도성 성질은 에너지 형태를 사용하여 유전자 편집 또는 유전자 발현의 시공간적 제어를 가능하게 할 것이다. 에너지 형태는, 제한은 아니지만, 전자기선, 소리 에너지, 화학 에너지 및 열 에너지를 포함할 수 있다. 유도성 시스템의 예들은 테트라사이클린 유도성 프로모터(Tet-On 또는 Tet-Off), 소 분자 2-하이브리드 전사 활성화 시스템(FKBP, ABA 등) 또는 광 유도성 시스템(피토크롬, LOV 도메인 또는 크립토크롬)을 포함한다. 일 실시형태에서, CRISPR 이펙터 단백질은 서열-특이적 방식으로 전사 활성에서의 변화를 유도하기 위한 광 유도성 전사 이펙터(LITE)의 일부일 수 있다. 광의 성분은 CRISPR 이펙터 단백질, 광-반응성 시토크롬 이형이량체(예를 들어, 아라비돕시스 탈리아나(Arabidopsis thaliana) 유래), 및 전사 활성화/억제 도메인을 포함할 수 있다. 유도성 DNA 결합 단백질 및 이의 사용을 위한 방법의 추가의 예는 미국 61/736,465, 미국 61/721,283 및 WO 2014/018423 A2에서 제공되며, 이는 전체가 참고로 본 명세서에 포함된다.

[0637] **자가-비활성화 시스템**

[0638] 세포 내 RNA의 모든 복제물이 편집되면, 그 세포에서 계속적인 Cas13b 이펙터 단백질 발현 또는 활성은 더 이상 필요하지 않다. 가이드 표적 서열로서 Cas13b 또는 crRNA에 대한 RNA의 사용에 의존하는 자가-비활성화 시스템은 Cas13b의 발현 또는 복합체 형성을 방지함으로써 시스템을 정지시킬 수 있다.

[0639] **키트**

[0640] 일 양태에서, 본 발명은 상기 방법 및 조성물에서 개시된 요소들 중 어느 하나 이상을 함유하는 키트를 제공한다. 일부 실시형태에서, 키트는 본 명세서에 교시된 바와 같은 벡터 시스템 또는 본 명세서에 교시된 바와 같은 CRISPR/Cas13b 시스템 또는 복합체의 성분 중 하나 이상, 예컨대 crRNA 및/또는 Cas13b 이펙터 단백질 또는 mRNA를 인코딩하는 Cas13b 이펙터 단백질, 및 키트를 사용하기 위한 안내서를 포함한다. 요소는 개별적으로 또는 조합하여 제공될 수 있으며, 임의의 적절한 용기, 예를 들어, 바이얼, 병 또는 튜브에 제공될 수 있다. 일부 구현예에서, 키트는 하나 이상의 언어, 예를 들어 하나 초과의 언어로 적힌 안내서를 포함한다. 안내서는 본 명세서에 기재된 적용 및 방법에 특이적일 수 있다. 일부 구현예에서, 본 명세서에서 설명된 하나 이상의 요소를 이용하는 공정에서의 사용을 위한 하나 이상의 시약을 포함한다. 시약은 임의의 적절한 용기에 제공될 수 있다. 예를 들어, 키트는 하나 이상의 반응 또는 저장 완충제를 제공할 수 있다. 시약은 특정 분석에서 사용가능한 형태로, 또는 사용 전에 하나 이상의 기타 성분의 첨가를 필요로 하는 형태로 제공될 수 있다 (예를 들어, 농축 또는 동결건조된 형태로). 완충액은 임의의 완충액일 수 있으며, 이로 제한되지는 않지만, 탄산나트륨 완충액, 중탄산나트륨 완충액, 보레이트 완충액, 트리스(Tris) 완충액, MOPS 완충액, HEPES 완충액, 및 이들의 조합일 수 있다. 일부 구현예에서, 완충액은 알칼리성이다. 일부 구현예에서, 완충액은 약 7 내지 약 10의 pH를 갖는다. 일부 구현예에서, 키트는, 가이드 서열 및 조절 요소에 작동 가능하게 연결되도록, 하나 이상의 벡터 내로의 삽입을 위한 가이드 또는 crRNA 서열에 상응하는 올리고뉴클레오티드를 포함한다. 일부 구현예에서, 키트는 상동성 재조합 주형 폴리뉴클레오티드를 포함한다. 일부 구현예에서, 키트는, 본 명세서에서 설명된, 하나 이상의 벡터 및/또는 하나 이상의 폴리뉴클레오티드를 포함한다. 키트는 유리하게는 본 발명의 시스템의 모든 요소를 제공하는 것을 가능하게 할 수 있다.

[0641] 본 발명은, 예를 들어 유전자 치료법, 약물 스크리닝, 질병 진단, 및 예후에서 광범위한 적용을 갖는다.

[0642] 용어 "폴리뉴클레오티드", "뉴클레오티드", "뉴클레오티드 서열", "핵산" 및 "올리고뉴클레오티드"는 상호교환 가능하게 사용된다. 그것들은 임의의 길이의 뉴클레오티드, 데옥시리보뉴클레오티드 또는 리보뉴클레오티드 중 어느 하나, 또는 그의 유사체의 중합체 형태를 말한다. 폴리뉴클레오티드는 임의의 3차원 구조를 가질 수 있으며, 기지의 또는 미지의 임의의 기능을 수행할 수 있다. 다음은 폴리뉴클레오티드의 비-제한적인 예이다:유전자 또는 유전자 단편의 코딩 또는 비-코딩 영역, 연관 분석으로부터 정의된 유전자좌들(유전자좌), 엑손, 인트론, 전령 RNA(mRNA), 운반 RNA, 리보솜 RNA, 짧은 간섭 RNA(siRNA), 짧은 헤어핀 RNA(shRNA), 마이크로-RNA(miRNA), 리보자임, cDNA, 재조합 폴리뉴클레오티드, 분지형 폴리뉴클레오티드, 플라스미드, 벡터, 임의의 서열의 분리된 DNA, 임의의 서열의 분리된 RNA, 핵산 프로브 및 프라이머. 이 용어는

또한 합성 백본을 갖는 핵산-유사 구조를 포함한다, 참조, 예를 들어, Eckstein, 1991; Baserga et al., 1992; Milligan, 1993; WO 97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; 및 Samstag, 1996. 폴리뉴클레오티드는 하나 이상의 변형된 뉴클레오티드, 예를 들어, 메틸화 뉴클레오티드 및 뉴클레오티드 유사체를 포함할 수 있다. 뉴클레오티드 구조에 대한 변형이 존재한다면, 중합체의 조립 전에 또는 후에 부여될 수 있다. 뉴클레오티드의 서열은 비-뉴클레오티드 성분에 의해 단속될 수 있다. 폴리뉴클레오티드는 중합 후에, 예를 들어, 표지화 성분과의 콘쥬게이션에 의해 추가로 변형될 수 있다. 본원에 사용되는 바와 같이, 용어 "야생형"은 당업자에 의해 이해되는 해당 분야의 용어이며, 그것이 돌연변이체 또는 변이체 형태로부터 구별되는 정도로 천연에서 발생하는 것과 같은 전형적인 형태의 유기체, 균주, 유전자 또는 특징을 의미한다. "야생형"은 기준선일 수 있다. 본원에 사용되는 바와 같이, 용어 "변이체"는 천연에서 발생하는 것에서 벗어난 패턴을 갖는 특성의 표현을 의미하는 것으로 이해해야 한다. 용어 "비-천연 발생" 또는 "조작된"은 상호교환 가능하게 사용되며, 인간의 손의 개입을 나타낸다. 상기 용어는 핵산 분자 또는 폴리펩티드에 대하여 언급되는 경우, 핵산 분자 또는 폴리펩티드에 천연에서 천연적으로 관련되어 있고, 천연에서 관찰되는 적어도 하나의 다른 성분이 적어도 실질적으로 없음을 의미한다. "상보성"은 통상의 왓슨-크릭(Watson-Crick) 염기 쌍형성 또는 기타 비-통상적 유형에 의해 다른 핵산 서열과 수소 결합(들)을 형성하는 핵산의 능력을 지칭한다. 상보성 백분율은 제2 핵산 서열과 수소 결합(예를 들어, 왓슨-크릭 염기 쌍형성)을 형성할 수 있는 핵산 분자 내의 잔기의 백분율을 나타낸다(예를 들어, 10 개 중 5, 6, 7, 8, 9, 10 개는 50%, 60%, 70%, 80%, 90% 및 100% 상보성임). "완전한 상보성"은 핵산 서열의 모든 연속 잔기가 동일한 수의 제2 핵산 서열 내의 연속 잔기와 수소 결합할 것임을 의미한다. 본원에 사용되는 바와 같이, "실질적인 상보성"은 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 개 이상의 뉴클레오티드의 영역에 걸쳐 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% 또는 100%인 상보성 정도를 지칭하거나, 엄격한 조건 하에서 혼성화하는 2개의 핵산을 지칭한다. 본원에 사용되는 바와 같이, 혼성화를 위한 "엄격한 조건"은 표적 서열에 대하여 상보성을 갖는 핵산 서열이 대개 표적 서열과 혼성화하며, 비-표적 서열에는 실질적으로 혼성화하지 않는 조건을 지칭한다. 엄격한 조건은 일반적으로 서열-의존적이며, 다수의 요인에 따라 달라진다. 일반적으로, 서열이 길수록, 서열이 그의 표적 서열에 특이적으로 혼성화하는 온도가 더 높아진다. 엄격한 조건의 비-제한적인 예는 문헌[Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N. Y]에 상세히 기재되어 있다. 폴리뉴클레오티드 서열이 언급될 때, 상보성 또는 부분 상보성 서열도 생각된다. 이들은 바람직하게 높은 엄격성 조건하에 기준 서열에 혼성화할 수 있다. 일반적으로, 혼성화 비율을 최대화하기 위해서, 상대적으로 저-긴축 혼성화 조건이 선택된다:열 용융점 (T_m)보다 약 20 내지 25℃ 낮은 온도. T_m은 특정 표적 서열의 50%가 규정된 이온 강도와 pH에서 용액 중에서 완전히 상보성인 프로브와 혼성화하는 온도이다. 일반적으로, 혼성화된 서열의 적어도 약 85% 뉴클레오티드 상보성을 요구하기 위해서, 고 긴축 세척 조건은 T_m보다 약 5 내지 15℃ 낮은 온도가 되도록 선택된다. 혼성화된 서열의 적어도 약 70% 뉴클레오티드 상보성을 요구하기 위해서, 중도-긴축 세척 조건은 T_m보다 약 15 내지 30℃ 낮은 온도가 되도록 선택된다. 아주 관대한(매우 낮은 엄격성) 세척 조건은 T_m보다 50℃ 정도 낮을 수 있으며, 이것은 혼성화된 서열들 간에 높은 수준의 미스매칭을 허용한다. 당업자는 표적 서열과 프로브 서열 간 상동성의 특정 수준으로부터 검출가능한 혼성화 신호의 결과에 영향을 미치기 위해서 혼성화 및 세척 단계에서 다른 물리화학적 변수들도 변경될 수 있다는 것을 인정할 것이다. 바람직한 매우 엄격한 조건은 42 ℃에서 50% 포름아미드, 5xSSC 및 1% SDS에서 인큐베이션, 또는 65℃에서 5xSSC 및 1% SDS에서 인큐베이션, 65 ℃에서 0.2xSSC 및 0.1% SDS에서 세척을 포함한다. "혼성화"는 하나 이상의 폴리 뉴클레오티드가 반응하여 뉴클레오티드 잔기의 염기 사이의 수소 결합을 통해 안정화된 복합체를 형성하는 반응을 지칭한다. 수소 결합은 왓슨 크릭 염기 쌍형성, 후그스타인(Hoogstein) 결합 또는 임의의 다른 서열 특이적 방식으로 발생할 수 있다.

복합체는 듀플렉스 구조를 형성하는 2 개의 가닥, 다중 가닥 복합체를 형성하는 3 개 이상의 가닥, 단일의 자가 혼성화 가닥 또는 이들의 임의의 조합을 포함할 수 있다. 혼성화 반응은 PCR의 개시 또는 효소에 의한 폴리뉴클레오티드의 절단과 같은 보다 광범위한 과정에서 하나의 단계를 이룰 수 있다. 주어진 서열과 혼성화할 수 있는 서열은 주어진 서열의 "상보물"로 지칭된다. 본 명세서에서 사용된 용어 "게놈 유전자좌" 또는 "유전자좌"(복수의 유전자좌)는 염색체에서 유전자 또는 DNA 서열의 특정 위치이다. "유전자"는 유기체에서 활동하는 기능적 역할을 가진 RNA 사슬 또는 폴리펩티드를 인코딩하는 DNA 또는 RNA의 스트레치로서, 이에 따라, 살아있는 유기체에서 유전의 분자 단위를 말한다. 본 발명의 목적을 위해서, 유전자는 이러한 조

절 서열이 코딩 서열 및/또는 전사된 서열에 인접하든 아니든 유전자 산물의 생성을 조절하는 영역을 포함하는 것으로 간주될 수 있다. 따라서, 반드시 제한되는 것은 아니지만, 유전자는 프로모터 서열, 종결자, 리보솜 결합 부위 및 내부 리보솜 진입 부위와 같은 번역 조절 서열, 인핸서, 사일런서, 인슐레이터, 경계 요소, 복제 원점, 매트릭스 부착 부위 및 유전자좌 조절 영역을 포함한다. 본 명세서에서 사용된 "게놈 유전자좌의 발현" 또는 "유전자 발현"은 유전자로부터의 정보가 기능적 유전자 산물의 합성에서 사용되는 과정이다. 유전자 발현의 산물은 주로 단백질이지만, rRNA 유전자 또는 tRNA 유전자와 같은 비-단백질 코딩 유전자에서는 이 산물은 기능적 RNA이다. 유전자 발현 과정은 생존을 위한 기능적 산물들을 생성하기 위해서 모든 알려진 생명체, 즉 진핵생물 (다세포 유기체 포함), 원핵생물 (박테리아 및 고세균) 및 바이러스에 의해서 사용된다. 본 명세서에서 사용된 유전자 또는 핵산의 "발현"은 세포 유전자 발현뿐만 아니라 클로닝 시스템 및 임의의 다른 맥락에서 핵산(들)의 전사 및 번역을 포괄한다. 본 명세서에서 사용된 "발현"은 또한 폴리뉴클레오티드가 DNA 주형으로부터(예컨대, mRNA 또는 다른 RNA 전사물로) 전사되는 과정 및/또는 전사된 mRNA가 이어서 펩티드, 폴리펩티드, 또는 단백질로 번역되는 과정을 말한다. 전사물 및 인코딩된 폴리펩티드는 집합적으로 "유전자 산물"이라고 언급될 수 있다. 폴리뉴클레오티드가 게놈 DNA로부터 유래되면, 발현은 진핵생물 세포에서 mRNA의 스플라이싱을 포함할 수 있다. 용어 "폴리펩티드", "펩티드" 및 "단백질"은 본원에서 임의의 길이의 아미노산의 중합체를 지칭하기 위해 상호교환가능하게 사용된다. 중합체는 선형 또는 분지형일 수 있으며, 그것은 변형된 아미노산을 포함할 수 있고, 그것은 비-아미노산에 의해 단속될 수 있다. 또한, 상기 용어는 변형된 아미노산 중합체, 예를 들어, 이황화 결합 형성, 글리코실화, 지질화(lipidation), 아세틸화, 인산화 또는 임의의 기타 조작, 예를 들어, 표지화 성분과의 콘쥬게이션을 포함한다. 본원에 사용되는 바와 같이, 용어 "아미노산"은 글리신 및 D 또는 L 광학 이성질체 및 아미노산 유사체 및 펩티도미메틱을 포함하는 천연 및/또는 비-천연 또는 합성 아미노산을 포함한다. 본 명세서에서 사용된 용어 "도메인" 또는 "단백질 도메인"은 단백질 사슬의 나머지 부분과 독립적으로 존재하며 기능할 수 있는 단백질 서열의 일부를 말한다.

본 발명의 양태들에서 설명된 대로, 서열 동일성은 서열 상동성과 연관된다. 상동성 비교는 눈으로, 또는 더 일반적으로는 쉽게 이용가능한 서열 비교 프로그램의 도움하에 수행될 수 있다. 이들 상업적으로 이용가능한 컴퓨터 프로그램은 둘 이상의 서열 사이의 상동성 퍼센트(%)를 계산할 수 있고, 또한 둘 이상의 아미노산 또는 핵산 서열에 의해서 공유된 서열 동일성을 계산할 수 있다.

[0643] 본원에 사용되는 바와 같이, 용어 "야생형"은 당업자에 의해 이해되는 해당 분야의 용어이며, 그것이 돌연변이체 또는 변이체 형태로부터 구별되는 정도로 천연에서 발생하는 것과 같은 전형적인 형태의 유기체, 균주, 유전자 또는 특징을 의미한다. "야생형"은 기준선일 수 있다.

[0644] 본원에 사용되는 바와 같이, 용어 "변이체"는 천연에서 발생하는 것에서 벗어난 패턴을 갖는 특성의 표현을 의미하는 것으로 이해해야 한다. 용어 "비-천연 발생" 또는 "조작된"은 상호교환 가능하게 사용되며, 인간의 손의 개입을 나타낸다. 상기 용어는 핵산 분자 또는 폴리펩티드에 대하여 언급되는 경우, 핵산 분자 또는 폴리펩티드에 천연에서 천연적으로 관련되어 있고, 천연에서 관찰되는 적어도 하나의 다른 성분이 적어도 실질적으로 없음을 의미한다. 모든 양태 및 구현예에서, 이들 용어가 포함되는지 아닌지의 여부에 관계없이, 바람직하게는 이는 선택적일 수 있으며, 이에 따라 바람직하게는 포함되거나 또는 바람직하게는 포함되지 않을 수 있는 것으로 이해될 것이다. 추가로, 용어 "비-천연 발생" 및 "조작된"은 상호교환 가능하게 사용될 수 있으며, 따라서 단독으로 또는 조합되어 사용될 수 있고 하나 또는 다른 것들이 함께 언급된 것들을 대체할 수 있다. 특히, "비-천연 발생" 또는 "비-천연 발생 및/또는 조작된" 대신에 "조작된"이 바람직하다.

[0645] 서열 상동성은 당업계에 공지된 다수의 컴퓨터 프로그램, 예를 들어 BLAST 또는 FASTA 등에 의해 생성될 수 있다. 이러한 정렬을 수행하기 위한 적합한 컴퓨터 프로그램은 GCG 위스콘신 베스트핏 패키지이다(University of Wisconsin, U.S.A; Devereux et al., 1984, Nucleic Acids Research 12:387). 서열 비교를 수행할 수 있는 다른 소프트웨어의 예들은, 이로 제한되지는 않지만 BLAST 패키지 (참조, Ausubel et al., 1999 ibid - Chapter 18), FASTA (Atschul et al., 1990, J. Mol. Biol., 403-410) 및 비교 툴의 GENEWORKS 슈트이다. BLAST와 FASTA는 모두 오프라인과 온라인 검색을 위해 이용할 수 있다(상기 문헌[Ausubel et al., 1999], 페이지 7-58 내지 7-60 참조). 그러나, GCG Bestfit 프로그램을 사용하는 것이 바람직하다. 서열 상동성 퍼센트(%)는 연속 서열에 걸쳐서 계산될 수 있는데, 즉 하나의 서열이 다른 서열과 정렬되고, 하나의 서열에서 각 아미노산 또는 뉴클레오티드가 다른 서열에서 상응하는 아미노산 또는 뉴클레오티드와 한번에 하나의 잔기씩 직접 비교된다. 이것은 "비갭"(ungapped) 정렬이라고 불린다. 통상적으로, 이러한 비갭 정렬은 상대적으로 짧은 수의 잔기에 걸쳐서만 수행된다. 이것은 매우 간단하고 일관된 방법이지만, 예를 들어 다른 동일한 쌍의 서열에서, 하나의 삽입 또는 결실이 이후의 아미노산 잔기를 정렬되지 않도록 할 수 있다는 것을 고려하지 않으므로, 전반적 정렬이 수행될 때는 상동성 %에 잠재적으로 상당한 감소가 있게 된다. 결론적으로, 대부

분의 서열 비교 방법은 전체 상동성 또는 동일성 점수에 지나친 벌점부과 없이 가능한 삽입과 결실을 고려하여 최적 정렬을 생성하도록 디자인된다. 이것은 국소 상동성 또는 동일성을 최대화하도록 시도하도록 서열 정렬에 "갭"을 삽입함으로써 달성된다. 그러나, 이런 더 복잡한 방법은 정렬에서 발생한 각 갭에 "갭 패널티"를 배정하며, 이로써 동일한 수의 동일한 아미노산에 대해, 가능한 적은 갭을 가진 서열 정렬 - 비교된 두 서열 간의 더 높은 연관성 반영 - 이 많은 갭을 가진 것보다 더 높은 점수를 달성할 수 있다. "친화성 갭 점수"가 통상적으로 사용되는데, 이것은 갭의 존재에는 상대적으로 높은 점수를 매기고, 갭의 각 연속 잔기에는 더 적은 패널티를 매긴다. 이것은 가장 흔히 사용되는 갭 점수배정 시스템이다. 높은 갭 패널티는 당연히 갭이 적을수록 최적화된 정렬을 생성할 수 있다. 대부분의 정렬 프로그램은 갭 패널티의 변형을 허용한다. 그러나, 서열 비교에 이러한 소프트웨어를 사용할 때는 디폴트 값을 사용하는 것이 바람직하다. 예를 들어, GCG Wisconsin Bestfit 패키지를 사용할 때, 아미노산 서열에 대한 디폴트 갭 패널티는 갭 당 -12, 각 연장 당 -4이다. 따라서, 최대 상동성 %의 계산은 갭 패널티를 고려한 최적 정렬의 생성을 요구한다. 이러한 정렬을 수행하기 위한 적합한 컴퓨터 프로그램은 GCG 위스콘신 베스트핏 패키지이다(Devereux et al., 1984 *Nuc. Acids Research* 12 p387). 서열 비교를 수행할 수 있는 다른 소프트웨어의 예들은, 이로 제한되지는 않지만 BLAST 패키지 (참조, Ausubel et al., 1999 *Short Protocols in Molecular Biology*, 4th Ed. - Chapter 18), FASTA (Atschul et al., 1990, J. Mol. Biol., 403-410) 및 비교 툴의 GENEWORKS 슈트이다. BLAST와 FASTA는 모두 오프라인과 온라인 검색으로 이용할 수 있다(문헌[Ausubel et al., 1999, Short Protocols in Molecular Biology, pages 7-58 to 7-60] 참조). 그러나, 일부 용도에서, GCG Bestfit 프로그램을 사용하는 것이 바람직하다. BLAST 2 Sequences라고 불리는 새로운 툴을 사용하여 단백질 및 뉴클레오티드 서열을 비교하는 것이 또한 가능하다 (참조, *FEMS Microbiol Lett.* 1999 174(2): 247-50; *FEMS Microbiol Lett.* 1999 177(1): 187-8 및 the National Center for Biotechnology information at the website of the National Institutes for Health 의 웹사이트). 최종 상동성%는 동일성 항목으로 측정될 수 있지만, 전형적으로 정렬 과정 자체는 모두-아니면-전부 쌍 비교에 기초하지 않는다. 대신, 축척된 유사성 점수 행렬이 일반적으로 사용되는데, 이것은 화학적 유사성 또는 진화적 거리에 기초하여 각 쌍-방식 비교에 점수를 배정한다. 흔히 사용되는 이러한 행렬의 예는 BLOSUM62 매트릭스로서, 이것은 BLAST 프로그램 일습의 디폴트 행렬이다. GCG Wisconsin 프로그램은 일반적으로 공개된 디폴트 값이나 제공되는 경우에는 사용자 지정 기호비교표 (추가의 상세한 설명을 위한 사용자 매뉴얼 참조)를 사용한다. 일부 용도에서, GCG 패키지에 대한 공개된 디폴트 값을 사용하거나, 또는 다른 소프트웨어의 경우에는 BLOSUM62와 같은 디폴트 행렬을 사용하는 것이 바람직하다. 또는 달리, 상동성 퍼센트는 CLUSTAL와 유사한 알고리즘에 기초한, DNASIS™(Hitachi Software)의 다중 정렬 특징을 사용하여 계산될 수 있다(Higgins DG & Sharp PM (1988), *Gene* 73(1), 237-244). 일단 소프트웨어가 최적 정렬을 생성하면, 상동성 %, 바람직하게는 서열 동일성 %를 계산하는 것이 가능하다. 이 소프트웨어는 통상적으로 서열 비교의 일부로서 이것을 행? 10하고, 수치 결과를 생성한다. 서열은 또한 침묵성 변화를 야기하여 기능적으로 동등한 물질을 가져오는 아미노산 잔기의 결실, 삽입 또는 치환을 가질 수 있다. 아미노산 특성(예컨대 잔기의 극성, 전하, 용해성, 소수성, 친수성 및/또는 양친매성)의 유사성을 기초로 의도적인 아미노산 치환이 만들어질 수 있으며, 따라서 이것은 아미노산을 기능적 그룹으로 함께 그룹화하는데 유용하다. 아미노산은 그것의 측쇄만의 특성에 기초하여 그룹화될 수 있다. 그러나, 돌연변이 데이터를 포함하는 것도 역시 더 유용하다. 이렇게 유도된 아미노산 세트는 구조적 이유로 보존될 수 있다. 이들 세트는 벤 다이어그램의 형태로 기술될 수 있다 (Livingstone C.D. and Barton G.J. (1993) “Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation” *Comput. Appl. Biosci.* 9: 745-756) (Taylor W.R. (1986) “The classification of amino acid conservation” *J. Theor. Biol.* 119; 205-218). 보존적 치환 (또한 표 1과 함께 본원에서 동일한 논의를 참조함) 은, 예를 들어 일반적으로 용인된 벤다이어그램 아미노산 그룹화를 설명한 아래 표에 따라서 만들어질 수 있다.

세트		서브-세트	
소수성	FWYHKMILVAGC	방향족	FWYH
		지방족	ILV
극성	WYHKREDCSTNQ	하전됨	HKRED
		양전하	HKR
		음전하	ED
소형	VCAGSPTND	아주 작음	AGS

[0646]

[0647] 또한, 표 1과 관련하여 보존적 치환에 관한 인용은 보존적 치환과 동일한 것으로 언급된다.

[0648] 용어 "대상체", "개체" 및 "환자"는 척추동물, 바람직하게는 포유동물, 더욱 바람직하게는 인간을 지칭하기 위해 본원에서 상호교환가능하게 사용된다. 포유동물은 이로 제한되지는 않지만 뮤린, 유인원, 인간, 농장 동물, 경기용 동물, 및 반려동물을 포함한다. 생체내에서 수득되거나 시험관내에서 배양된 생물학적 엔티티(entity)의 조직, 세포 및 그들의 자손도 또한 포함된다.

[0649] 용어 "치료제", "치료가능한 작용제" 또는 "치료 작용제"는 상호교환가능하게 사용되며, 대상체로의 투여 시에 몇몇 유리한 효과를 부여하는 분자 또는 화합물을 지칭한다. 유익한 효과는 진단 결정의 가능; 질병, 증상, 장애 또는 병리학적 상태의 개선; 질병, 증상, 장애 또는 상태의 발병의 감소 또는 예방; 및 일반적으로 질병, 증상, 장애 또는 병리학적 상태에 대한 대항을 포함한다. 본원에 사용되는 바와 같이, "치료" 또는 "치료하는" 또는 "완화하는" 또는 "개선하는"은 상호교환가능하게 사용된다. 이들 용어는 치료 이익 및/또는 예방 이익을 포함하나 이들에 한정되지 않는 유리한 또는 요망되는 결과를 수득하는 방법을 지칭한다. 치료적 이익이라 함은, 치료 하에 있는 하나 이상의 질병, 질환, 또는 증상에서의 임의의 치료 관련 개선 또는 이에 대한 효과를 의미한다. 예방적 이익을 위해, 조성물은 특정 질병, 질환, 또는 증상 발전의 위험에 있는 대상체에, 또는 질병, 질환, 또는 증상이 아직 표명되지 않을 수 있다 하더라도 질병의 하나 이상의 생리학적 증상이 보고되는 대상체에게 투여될 수 있다. 용어 "유효량" 또는 "치료적 유효량"은 유리한 또는 요망되는 결과를 야기하기에 충분한 작용제의 양을 지칭한다. 치료적 유효량은, 치료될 대상체 및 질병 상태, 대상체의 체중 및 연령, 질병 상태의 심각도, 투여 방법 등 중 하나 이상에 따라 달라질 수 있으며, 이는 당업자에 의해 쉽게 결정될 수 있다. 또한, 상기 용어는 본원에 기술된 영상화 방법 중 임의의 것에 의한 검출을 위한 이미지를 제공할 용량에 적용된다. 특정 용량은 선택된 특정 작용제, 뒤따르는 투여 요법, 그것이 다른 화합물과 병용하여 투여되는지 여부, 투여 시기, 영상화되는 조직 및 그것을 운반하는 신체 전달 시스템 중 하나 이상에 따라 달라질 수 있다.

[0650] 본 발명의 실시는 달리 나타내지 않는 한, 당업계의 기술 내에 있는 면역학, 생화학, 화학, 분자 생물학, 미생물학, 세포 생물학, 유전체학 및 재조합 DNA의 통상의 기술을 사용한다. 참조: Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series 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)). 본 발명의 일부 양태는 하나 이상의 벡터를 포함하는 벡터 시스템, 또는 이러한 벡터에 관한 것이다. 벡터는 원핵 또는 진핵 세포에서 CRISPR 전사물(예를 들어, 핵산 전사물, 단백질 또는 효소)의 발현을 위해 디자인될 수 있다. 예를 들어, CRISPR 전사물은 박테리아 세포, 예를 들어, 에스케리키아 콜라이, 곤충 세포(배큘로바이러스 발현 벡터 사용), 효모 세포 또는 포유동물 세포에서 발현될 수 있다. 적합한 숙주 세포는 Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 에 추가로 논의된다. 대안적으로, 재조합 발현 벡터는 예를 들어, T7 프로모터 조절 서열 및 T7 폴리머라제를 사용하여 시험관내에서 전사되고 번역될 수 있다. 본 발명의 실시형태는 일어날 수 있는 상동 치환(치환 및 대체는 본 명세서에서 기존 아미노산 잔기 또는 뉴클레오티드의, 대안적인 잔기 또는 뉴클레오티드와의 교환을 의미하기 위해 본원에서 사용됨), 즉 염기성에 염기성, 산성에 산성, 극성에 극성과 같은 아미노산의 경우 유사-대-유사 치환을 포함할 수 있는 서열 (폴리뉴클레오티드 또는 폴리펩티드 둘 다)을 포

함한다. 비-상동 치환은 또한 한 계열의 잔기에서 다른 계열로 또는 대안적으로 오르니틴 (이하 Z로 지칭됨), 디아미노부티르산 오르니틴 (이하 B로 지칭 됨), 노르류신 오르니틴 (이하 O로 지칭됨), 피리릴알라닌, 티에닐알라닌, 나프틸알라닌 및 페닐글리신과 같은 비-천연 아미노산의 혼입을 수반하여 일어날 수 있다. 변이체 아미노산 서열은 글리신 또는 β-알라닌 잔기와 같은 아미노산 스페이서에 더하여 메틸, 에틸 또는 프로필 기와 같은 알킬 기를 포함하는 서열의 임의의 두 아미노산 잔기 사이에 삽입될 수 있는 적합한 스페이서 기를 포함할 수 있다. 펩토이드 형태로 하나 이상의 아미노산 잔기의 존재를 포함하는 추가의 형태의 변형도 당업자에게 잘 이해될 수 있다. 의심을 피하기 위해서, "펩토이드 형태"는 α-탄소 치환체 기가 α-탄소가 아니라 잔기의 질소 원자에 존재하는 변이체 아미노산 잔기를 말하기 위해서 사용된다. 펩토이드 형태로 펩티드를 제조하는 과정은 해당 분야에 공지되어 있다, 예를 들어: Simon RJ et al., *PNAS* (1992) 89(20), 9367-9371 and Horwell DC, *Trends Biotechnol*. (1995) 13(4), 132-134. 상동성 모델링: 다른 Cas13b 오르소로그에서 상응하는 잔기는 Zhang et al., 2012 (Nature; 490(7421): 556-60) 및 Chen et al., 2015 (PLoS Comput Biol; 11(5): e1004248)의 방법 -도메인-모티브 계면에 의해 매개되는 상호작용을 예측하기 위한 전산 단백질-단백질 상호작용 (PPI) 방법- 으로 확인할 수 있다. PrePPI(예측 PPI)는 구조 기준 PPI 예측 방법으로, 베이지안(Bayesian) 통계 프레임워크를 이용하여 비-구조적 증거와 구조적 증거를 조합한다.

이 방법은 문의되는 단백질의 쌍을 취하고, 구조적 정렬을 이용하여 그의 실험적으로 결정된 구조 또는 상동성 모델 중 어느 하나에 상응하는 구조적 대표물을 확인하는 것을 포함한다. 구조적 정렬은 전반적인 및 국소의 기하 관계를 고려함으로써 가까운 구조적 이웃 및 먼 구조적 이웃 모두를 확인하는데 추가로 사용된다.

구조적 대표물의 두 이웃이 단백질 데이터 뱅크에 보고된 복합체를 형성할 때마다, 이는 두 개의 문의되는 단백질 사이의 상호작용을 모델링하기 위한 주형을 정의한다. 복합체의 모델은 대표 구조물을 주형 내에서 그의 상응하는 구조적 이웃에 겹쳐놓음으로써 창출된다. 이러한 접근법은 Dey et al., 2013 (Prot Sci; 22: 359-66)에 추가로 기술된다.

[0651] 본 발명의 목적을 위하여, 증폭은 타당한 충실도로 표적 서열을 복제할 수 있는 폴리머라제 및 프라이머를 사용하는 임의의 방법을 의미한다. 증폭은 천연 또는 재조합 DNA 폴리머라제, 예를 들어, TaqGold™, T7 DNA 폴리머라제, 대장균 DNA 폴리머라제의 클레나우 단편 및 역전사효소에 의해 수행될 수 있다. 바람직한 증폭 방법은 PCR이다. 특정 양태에서, 본 발명은 벡터를 포함한다. 본원에 사용되는 바와 같이, "벡터"는 엔티티를 하나의 환경에서 또 다른 환경으로 전달하는 것을 허용 또는 촉진하는 도구이다. 이는 레플리콘, 예컨대 플라스미드, 파지, 또는 코스미드로, 또 다른 DNA 분절이 그 안에 삽입될 수 있어서, 삽입된 분절의 복제를 유발한다. 일반적으로, 벡터는 적절한 제어 요소와 관련되었을 때 복제가 가능하다. 일반적으로, 용어 "벡터"는 그것이 연결되어 있는 다른 핵산을 전달할 수 있는 핵산 분자를 지칭한다. 벡터는 제한 없이, 단일 가닥, 이중 가닥 또는 부분 이중 가닥인 핵산 분자; 하나 이상의 자유 말단을 포함하는, 자유 말단을 포함하지 않는 (예 : 환형) 핵산 분자; DNA, RNA 또는 둘 다를 포함하는 핵산 분자; 및 당업계에 공지된 다른 종류의 폴리뉴클레오티드를 포함한다. 하나의 유형의 벡터는 "플라스미드"이며, 이는 추가의 DNA 절편이 예를 들어, 표준 분자 클로닝 기술에 의해 삽입될 수 있는 환형 이중 가닥 DNA 루프를 지칭한다.

벡터의 다른 종류는 바이러스 벡터로서, 바이러스-유래된 DNA 또는 RNA 서열이 바이러스에 봉입되는 벡터에 존재한다(예를 들어, 레트로바이러스, 복제 결함 레트로바이러스, 아데노바이러스, 복제 결함 아데노바이러스, 및 아데노-관련 바이러스(AAV)). 바이러스 벡터는 또한 숙주 세포 내로의 트랜스펙션을 위한, 바이러스가 보유하는 폴리뉴클레오티드를 포함한다. 특정 벡터는 그것이 도입된 숙주 세포에서 자율적 복제가 가능하다(예를 들어, 박테리아 복제 기원을 가진 박테리아 벡터 및 에피솜 포유류 벡터). 다른 벡터(예를 들어, 비-에피솜 포유동물 벡터)는 숙주 세포 내로 도입시 숙주 세포의 게놈에 통합되며, 이에 의해 숙주 게놈과 함께 복제된다. 더욱이, 특정 벡터는 그것이 작동가능하게 연결된 유전자의 발현을 지시할 수 있다. 이러한 벡터는 본원에서 "발현 벡터"로 지칭된다. 재조합 DNA 기술에 유용한 통상적인 발현 벡터는 종종 플라스미드의 형태로 존재한다. 재조합 발현 벡터는 숙주 세포에서 핵산의 발현에 적합한 형태로 본 발명의 핵산을 포함할 수 있는데, 이는 재조합 발현 벡터가 하나 이상의 조절 요소를 포함하는 것을 의미하며, 하나 이상의 조절 요소는 발현에 사용될 숙주 세포에 기반하여 선택될 수 있고, 발현될 핵산 서열에 작동 가능하게 연결된다. 재조합 발현 벡터 내에서, "작동가능하게 연결된"은 대상 뉴클레오티드 서열이 (예를 들어, 시험관내 전사/번역 시스템 내에서, 또는 벡터가 숙주 세포 내로 도입되는 경우 숙주 세포 내에서) 뉴클레오티드 서열의 발현을 가능하게 하는 방식으로 조절 요소(들)에 연결된 것을 의미하는 의도이다. 재조합 및 클로닝 방법에 관련하여, US 2004-0171156 A1호로서 2004년 9월 2일 간행된 미국 특허 출원 10/815,730호가 언급될 수 있으며, 이의 모든 내용은 그 전체로서 본원에 참조로 포함된다. 본 발명의 측면은 가이드 RNA 및 야생형, 변형 또는 돌연변이된 CRISPR 이펙터 단백질/효소 (예를 들어, Cas13b 이펙터 단백질)를 위한 이중시스트론 벡

터에 관한 것이다. 이중시스트론 발현 벡터 가이드 RNA 및 야생형, 변형 또는 돌연변이된 CRISPR 이펙터 단백질/효소 (예를 들어, Cas13b 이펙터 단백질)가 바람직하다. 일반적으로 특히 이 실시형태 및 야생형에서, 변형 또는 돌연변이된 CRISPR 이펙터 단백질/효소 (예를 들어, Cas13b 이펙터 단백질)이 바람직하게 CBh 프로모터에 의해 구동된다. RNA는 바람직하게 Pol III 프로모터, 예를 들어, U6 프로모터에 의해 유도된다. 이상적으로 2 개가 조합된다.

[0652] 일부 구현예에서, 가이드 RNA 또는 crRNA에서 루프가 제공된다. 이는 스템 루프 도는 테트라 루프일 수 있다. 이 루프는 바람직하게 GAAA이지만, 이 서열에만 제한되는 것은 아니며, 실제로 단지 4 bp 길이에만 제한되는 것도 아니다. 실제로, 헤어핀 구조에서 사용되는 바람직한 루프 형성 서열은 4 뉴클레오티드 길이이고, 가장 바람직하게 서열 GAAA를 가진다. 그러나, 더 길거나 짧은 루프 서열도 사용될 수 있으며, 대용 서열들도 사용될 수 있다. 서열은 바람직하게 뉴클레오티드 트리플(예를 들어, AAA), 및 추가의 뉴클레오티드(예를 들어, C 또는 G)를 포함한다. 루프 형성 서열의 예는 CAAA 및 AAAG를 포함한다.

[0653] 본 명세서에 개시된 임의의 방법의 실시에 있어서, 적합한 벡터가, 이로 제한되지는 않지만, 미세주입, 전기천공, 소노포레이션(sonoporation), 바이오리스틱스, 칼슘 포스페이트-매개 트랜스펙션, 양이온 트랜스펙션, 리포좀 트랜스펙션, 덴드리머 트랜스펙션, 열 충격 트랜스펙션, 뉴클레오펙션 트랜스펙션, 마그네토펙션, 리포펙션, 임페일펙션, 광학 트랜스펙션, 핵산의 전용 제제-증진 흡수, 및 리포좀, 면역리포좀, 비로좀 또는 인공 비리온을 통한 전달을 포함하는 해당 분야에 알려진 하나 이상의 방법을 통해서 세포 또는 배아에 도입될 수 있다.
일부 방법에서, 벡터는 마이크로인젝션에 의해서 배아에 도입된다. 벡터 또는 벡터들은 배아의 핵 또는 세포질에 마이크로인젝션될 수 있다. 일부 방법에서, 벡터 또는 벡터들은 뉴클레오펙션에 의해서 세포에 도입될 수 있다.

[0654] 벡터는 원핵 또는 진핵 세포에서 CRISPR 전사물(예를 들어, 핵산 전사물, 단백질 또는 효소)의 발현을 위해 디자인될 수 있다. 예를 들어, CRISPR 전사물은 박테리아 세포, 예를 들어, 에스케리키아 콜라이, 곤충 세포(배큘로바이러스 발현 벡터 사용), 효모 세포 또는 포유동물 세포에서 발현될 수 있다. 적합한 숙주 세포는 Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 에 추가로 논의된다. 대안적으로, 재조합 발현 벡터는 예를 들어, T7 프로모터 조절 서열 및 T7 폴리머라제를 사용하여 시험관내에서 전사되고 번역될 수 있다.

[0655] 벡터는 원핵생물, 또는 원핵생물 세포에 도입되고, 그에서 증식될 수 있다. 일부 실시형태에서, 원핵생물은 진핵 세포로 도입될 벡터 또는 진핵 세포로 도입되는 벡터의 생성에서 중간체 벡터(예를 들어, 바이러스 벡터 패키징 시스템의 일부로서 플라스미드 증폭)로서 벡터의 복사물을 증폭시키기 위해서 사용된다. 일부 구현예에서, 원핵생물은 벡터의 복사를 증폭시키고, 하나 이상의 핵산을 발현하기 위해, 예를 들어, 숙주 세포 또는 숙주 생물로의 전달을 위한 하나 이상의 단백질의 공급원을 제공하기 위해 사용된다. 원핵생물에서의 단백질의 발현은 자주 융합 또는 비-융합 단백질 중 어느 하나의 발현을 유도하는 구성성 또는 유도성 프로모터를 함유하는 벡터를 사용하여 에스케리키아 콜라이에서 수행된다. 융합 벡터는 거기에서 인코딩된 단백질에, 예컨대 재조합 단백질의 아미노 말단에 수많은 아미노산을 부가한다. 이러한 융합 벡터는 다음과 같은, 하나 이상의 목적을 제공할 수 있다: (i) 재조합 단백질의 발현을 증가시키기 위해; (ii) 재조합 단백질의 용해도를 증가시키기 위해; 및 (iii) 친화성 정제에서 리간드로서 작용함으로써 재조합 단백질의 정제를 돕기 위해. 종종, 융합 발현 벡터에서, 단백질분해 절단 부위는 융합 모이어티와 재조합 단백질의 연접부에 도입되어, 융합 단백질의 정제 이후에 융합 모이어티로부터 재조합 단백질의 분리를 가능하게 한다. 이러한 효소 및 그들의 동족 인식 서열은 인자 Xa, 트롬빈 및 엔테로키나아제를 포함한다. 융합 발현 벡터의 예는 pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) 및 pRIT5 (Pharmacia, Piscataway, N.J.) 를 포함하며, 이는 각각 글루타티온 S-트랜스퍼라제(GST), 말토스 E 결합 단백질 또는 단백질 A를 표적 재조합 단백질에 융합시킨다. 적절한 유도성 비-융합 대장균 발현 벡터의 예는 pTrc(Amrann et al., (1988) *Gene* 69:301-315) 및 pET 11d(Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89) 를 포함한다. 일부 구현예에서, 벡터는 효모 발현 벡터이다. 효모 *사카로마이세스 세리비시아에*에서의 발현을 위한 벡터의 예는 pYepSec1 (Baldari, et al., 1987. *EMBO J.* 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz et al., 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), 및 picZ (InVitrogen Corp, San Diego, Calif.) 를 포함한다.
일부 구현예에서, 벡터는 배큘로바이러스 발현 벡터를 사용하여 곤충 세포에서 단백질 발현을 유도한다. 배양 곤충 세포 (예를 들어, SF9 세포)에서 단백질의 발현에 이용가능한 배큘로바이러스 벡터는 pAc 시리즈

(Smith, et al., 1983. *Mol. Cell. Biol.* 3: 2156-2165) 및 pVL 시리즈 (Lucklow and Summers, 1989. *Virology* 170: 31-39)를 포함한다. 일부 구현예에서, 벡터는 포유 동물 발현 벡터를 사용하여 포유 동물 세포에서 하나 이상의 서열의 발현을 유도할 수 있다. 포유동물 발현 벡터의 예는 pCDM8 (Seed, 1987. *Nature* 329: 840) 및 pMT2PC (Kaufman, et al., 1987. *EMBO J.* 6: 187-195)를 포함한다. 포유 동물 세포에서 사용되는 경우, 발현 벡터의 제어 작용은 통상적으로 하나 이상의 조절 요소에 의해 제공된다. 예를 들어, 통상적으로 사용되는 프로모터는 폴리오마, 아데노바이러스 2, 사이토메갈로바이러스, 유인원 바이러스 40 및 본원에 개시되고 당업계에 공지되어 있는 기타의 것으로부터 유래된다. 원핵생물 및 진핵생물 세포 둘 다에 대한 다른 적합한 발현 시스템에 대해서는, 예를 들어, Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989를 참조한다. 일부 구현예에서, 재조합 포유류 발현 벡터는 특정 세포 유형에서 우선적으로 핵산의 발현을 유도할 수 있다(예를 들어, 핵산을 발현하기 위하여 조직-특이적 조절 요소가 사용됨). 조직-특이적 조절 요소는 해당 분야에 공지되어 있다. 적합한 조직-특이적 프로모터의 비-제한적 예는 알부민 프로모터 (간-특이적; Pinkert, et al., 1987. Genes Dev. 1: 268-277), 림프-특이적 프로모터 (Calame and Eaton, 1988. Adv. Immunol. 43: 235-275), 특히 T 세포 수용체의 프로모터 (Winoto and Baltimore, 1989. EMBO J. 8: 729-733) 및 면역글로불린 (Baneiji, et al., 1983. Cell 33: 729-740; Queen and Baltimore, 1983. Cell 33: 741-748), 뉴런-특이적 프로모터 (예를 들어, 뉴로필라멘트 프로모터; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), 췌장-특이적 프로모터 (Edlund, et al., 1985. *Science* 230: 912-916), 및 유선-특이적 프로모터 (예를 들어, 우유 유청 프로모터; 미국 특허 번호 4,873,316 및 유럽 출원 공개 번호 264,166)를 포함한다. 발달-조절 프로모터, 예를 들어 뮤린 혹스 프로모터 (Kessel and Gruss, 1990. *Science* 249: 374-379) 및 α-태아단백질 프로모터 (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546)도 포함된다. 원핵세포 및 진핵세포 벡터와 관련하여, 미국특허 6,750,059를 참조하며, 이것의 내용은 그 전체가 여기 참고로 포함된다. 본 발명의 다른 구현예들은 바이러스 벡터의 사용에 관한 것일 수 있으며, 이에 관해서는 미국 특허 출원 13/092,085가 언급되며, 그의 내용은 본 명세서에 그 전체로서 참고문헌으로 통합된다. 조직-특이적 조절 요소는 해당 분야에 공지이며, 이에 관해서는 미국 특허 7,776,321이 언급되며, 그의 내용은 본 명세서에 그 전체로서 참고문헌으로 통합된다.

[0656] 일부 구현예에서, 조절 요소는 CRISPR Cas13b 시스템 또는 복합체의 하나 이상의 엘리먼트에 작동가능하게 연결되어 또는 이를 인코딩하여 CRISPR 시스템의 하나 이상의 엘리먼트의 발현을 유도한다. 일반적으로, SPIDR(스페이서 산재된 직접 반복부)로도 알려진 CRISPR(클러스터링되고 규칙적으로 산재된 짧은 팔린드로믹 반복부)은 통상 특정 박테리아 종에 특이적인 DNA 유전자좌의 패밀리를 구성한다. CRISPR 유전자좌는 대장균에서 인지되는 구별되는 계열의 산재된 짧은 서열 반복부(SSR)(Ishino et al., J. Bacteriol., 169:5429-5433 [1987]; 및 Nakata et al., J. Bacteriol., 171:3553-3556 [1989]) 및 관련 유전자를 포함한다. 유사한 산재된 SSR이 할로페락스 메디테라네이(Haloferax mediterranei), 스트렙토코커스 피오게네스, 아나바에나(Anabaena) 및 마이코박테리움 튜베르큘로시스(Mycobacterium tuberculosis)에서 확인되었다(참조, Groenen et al., Mol. Microbiol., 10:1057-1065 [1993]; Hoe et al., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl et al., Biochim. Biophys. Acta 1307:26-30 [1996]; 및 Mojica et al., Mol. Microbiol., 17:85-93 [1995]).

CRISPR 유전자좌는 통상적으로 SRSR(규칙적으로 산재된 짧은 반복부(short regularly spaced repeat))로 명명된 반복부의 구조가 다른 SSR과 상이하다(Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; 및 Mojica et al., Mol. Microbiol., 36:244-246 [2000]). 일반적으로, 반복부는 실질적으로 고정된 길이를 갖는 독특한 개재 서열에 의해 규칙적으로 산재된 클러스터에 존재하는 짧은 요소이다(상기 문헌[Mojica et al., [2000]]). 반복 서열이 균주들 간에 고도로 보존되어 있지만, 산재된 반복부의 수와 스페이서 영역의 서열은 통상적으로 균주마다 상이하다 (van Embden et al., J. Bacteriol., 182:2393-2401 [2000]). CRISPR 유전자좌는 아에로피룸(*Aeropyrum*), 피로바쿨룸(*Pyrobaculum*), 술폴로부스(Sulfolobus), 아캐오글로부스(Archaeoglobus), 할로카르쿨라(Halocarcula), 메타노박테리움(Methanobacterium), 메타노코커스(Methanococcus), 메타노사르시나(Methanosarcina), 메타노피러스(Methanopyrus), 피로코커스(Pyrococcus), 피크로필러스(Picrophilus), 써모플라스마(Thermoplasma), 코리네박테리움(Corynebacterium), 마이코박테리움(Mycobacterium), 스트렙토마이세스(Streptomyces), 아퀴펙스(Aquifex), 포르피로모나스(Porphyromonas), 클로로비움(Chlorobium), 써머스(Thermus), 바실러스(Bacillus), 리스테리아(Listeria), 스타필로코커스(Staphylococcus), 클로스트리디움(Clostridium), 써모아나에로박터(Thermoanaerobacter), 마이코플라스마(Mycoplasma), 푸소박테리움(Fusobacterium), 아자쿠스(Azarcus), 크로모박테리움(Chromobacterium), 네이세리아(Neisseria), 니트로소모나스(Nitrosomonas), 데설포비브리오(Desulfovibrio), 게오박터(Geobacter), 믹소코

커스(Myxococcus), 캄필로박터, 볼리넬라(Wolinella), 아시네토박터(Acinetobacter), 에르위니아(Erwinia), 에스케리키아, 레지오넬라(Legionella), 메틸로코커스(Methylococcus), 파스퇴렐라(Pasteurella), 포토박테리움(Photobacterium), 살모넬라(Salmonella), 잔토모나스(Xanthomonas), 예르시니아(Yersinia), 트레포네마(Treponema) 및 써모토가(Thermotoga)를 포함하나 이들에 한정되지 않는 40개 초과의 원핵생물에서 확인되었다(참조, 예를 들어, Jansen et al., Mol. Microbiol., 43:1565-1575 [2002]; 및 Mojica et al., [2005]).

[0657] 일반적으로, 본 출원에서 사용되는 바와 같은 "RNA-표적화 시스템"은 RNA-표적화 CRISPR-연합된 13b("Cas13b") 유전자(또한, 본 명세서에서 이펙터 단백질로 지칭됨)의 발현 또는 활성을 유도하는 데 수반된 기타 다른 엘리먼트 및 전사물을 종합적으로 지칭하며, 이는 본원에 논의된 표 1을 참조하여 RNA-표적화 Cas (이펙터) 단백질 및 가이드 RNA (또는 crRNA 서열)를 인코딩하는 서열을 포함한다. 일반적으로, RNA-표적화 시스템은 표적 서열의 부위에서 RNA-표적화 복합체의 형성을 촉진하는 엘리먼트에 의해 특징지어진다. RNA-표적화 복합체의 형성의 맥락에서, "표적 서열"은 가이드 서열이 이에 대해 상보성을 갖도록 디자인되는 RNA 서열을 지칭하며, 여기서 표적 서열과 가이드 RNA 사이의 혼성화는 RNA-표적화 복합체의 형성을 촉진한다. 본질적으로 완전한 상보성이 필요하지 않지만, 혼성화를 야기하고, RNA-표적화 복합체의 형성을 증진시키는 충분한 상보성이 존재한다. 일부 구현예에서, 표적 서열은 세포의 핵 또는 세포질 내에 위치된다. 일부 구현예에서, 표적 서열은 진핵 세포의 세포소기관 내에 있을 수 있다. 표적 서열을 포함하는 표적화된 유전자좌 내로의 재조합을 위해 사용될 수 있는 서열 또는 주형은 "편집 주형" 또는 "편집 RNA" 또는 "편집 서열"로 지칭된다. 본 발명의 양태에서, 외인성 주형 RNA는 편집 주형으로서 지칭될 수 있다. 본 발명의 일 양태에서 재조합은 상동성 재조합이다. 일반적으로, 가이드 서열은 표적 서열과 혼성화하고, 핵산-표적화 복합체의 표적 서열로의 서열-특이적 결합을 유도하기에 충분한 표적 폴리뉴클레오티드와의 상보성을 갖는 임의의 폴리뉴클레오티드 서열이다. 일부 구현예에서, 가이드 서열과 그의 상응하는 표적 서열 간의 상보성의 정도는 적절한 정렬 알고리즘을 사용하여 최적으로 정렬되는 경우, 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% 이상이다. 최적의 정렬은 서열을 정렬하기에 적절한 임의의 알고리즘의 사용으로 결정될 수 있으며, 그의 비-제한적인 예는 스미스-워터만(Smith-Waterman) 알고리즘, 니들만-분쉬(Needleman-Wunsch) 알고리즘, 버로우즈-휠러 트랜스폼(Burrows-Wheeler Transform)에 기초한 알고리즘(예를 들어, 버로우즈 휠러 얼라이너(Burrows Wheeler Aligner)), ClustalW, Clustal X, BLAT, 노보얼라인(Novoalign)(노보크라프트 테크놀로지즈(Novocraft Technologies), ELAND(일루미나(Illumina), 미국 캘리포니아주 샌디에고), SOAP(soap.genomics.org.cn에서 이용가능) 및 Maq(maq.sourceforge.net에서 이용가능)를 포함한다. 일부 구현예에서, 가이드 서열은 길이가 약 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 개 이상이거나 이를 초과한다. 일부 구현예에서, 가이드 서열은 길이가 약 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 개 이하의 뉴클레오티드이다. 표적 서열로의 RNA-표적화 복합체의 서열-특이적 결합을 유도하는 가이드 서열의 능력은 임의의 적절한 어세이에 의해 평가될 수 있다. 주형 폴리뉴클레오티드는 임의의 적합한 길이를 가질 수 있으며, 예컨대 길이가 약 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000 개 이상의 뉴클레오티드, 또는 이를 초과할 수 있다. 일부 구현예에서, 주형 폴리뉴클레오티드는 표적 서열을 포함하는 폴리뉴클레오티드의 일부에 상보적이다. 최적으로 정렬된 경우, 주형 폴리뉴클레오티드는 표적 서열의 하나 이상의 뉴클레오티드(예를 들어, 약 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 개 이상 또는 이를 초과하는 뉴클레오티드)와 중첩될 수 있다. 일부 구현예에서, 표적 서열 및 표적 서열을 포함하는 폴리뉴클레오티드의 경우, 표적 서열로부터 주형 폴리뉴클레오티드의 최근접 뉴클레오티드는 약 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000 개 이상의 뉴클레오티드 내에 속한다. 일부 구현예에서, RNA-표적화 이펙터 단백질은 하나 이상의 이종성 단백질 도메인(예를 들어, 핵산-표적화 이펙터 단백질에 추가적으로, 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 개 이상 또는 이를 초과하는 도메인)을 포함하는 융합 단백질의 일부이다. 일부 구현예에서, CRISPR Cas13b 이펙터 단백질/효소는 하나 이상의 이종성 단백질 도메인(예를 들어, CRISPR Cas13b 효소에 추가적으로, 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 개 이상 또는 이를 초과하는 도메인)을 포함하는 융합 단백질의 일부이다. CRISPR 이펙터 단백질에 융합될 수 있는 단백질 도메인의 예는 제한 없이, 에피토프 태그, 리포터 유전자 서열 및 하기의 활성 중 하나 이상을 갖는 단백질 도메인을 포함한다:메틸라제 활성, 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, RNA 절단 활성, 및 핵산 결합 활성. 에피토프 태그의 비-제한적인 예는 히스티딘(His) 태그, V5 태그, FLAG 태그, 인플루엔자 헤마글루티닌(HA) 태그, Myc 태그, VSV-G 태그 및 티오레독신(Trx) 태그를 포함한다. 리포터 유전자의 예는, 이로 한정되지 않지만, 글루타티온-S-트랜스퍼라제(GST), 호스라디시(horseradish) 과산화효소(HRP), 클로람페니콜 아세틸트랜스퍼라제(CAT) 베타-갈락토시다제, 베타-글루쿠로니다제, 루시퍼라제, 녹색 형광 단백질(GFP), HcRed,

DsRed, 청록색 형광 단백질(CFP), 황색 형광 단백질(YFP) 및 청색 형광 단백질(BFP)을 포함하는 자가형광 단백질을 포함한다. 핵산 표적화 이펙터 단백질은 DNA 분자에 결합하거나, 이로 제한되지는 않지만, 말토스 결합 단백질(MBP), S-태그, Lex A DNA 결합 도메인(DBD) 융합물, GAL4 DNA 결합 도메인 융합물 및 단순 포진 바이러스(HSV) BP16 단백질 융합체물을 포함하는 다른 세포 분자에 결합하는 단백질 또는 단백질의 단편을 인코딩하는 유전자 서열에 융합될 수 있다. 핵산-표적화 이펙터 단백질은 포함하는 융합 단백질의 부분을 형성할 수 있는 추가의 도메인은 본 명세서에 참고 문헌으로 포함되는 US20110059502에 기재되어 있다. 일부 실시형태에서, 태그화된 핵산-표적화 이펙터 단백질을 사용하여 표적 서열의 위치를 동정한다. 일부 구현예에서, CRISPR Cas13b 효소는 유도성 시스템의 성분을 형성할 수 있다. 시스템의 유도성 성질은 에너지 형태를 사용하여 유전자 편집 또는 유전자 발현의 시공간적 제어를 가능하게 할 것이다. 에너지 형태는, 제한은 아니지만, 전자기선, 소리 에너지, 화학 에너지 및 열 에너지를 포함할 수 있다. 유도성 시스템의 예들은 테트라사이클린 유도성 프로모터(Tet-On 또는 Tet-Off), 소 분자 2-하이브리드 전사 활성화 시스템(FKBP, ABA 등) 또는 광 유도성 시스템(피토크롬, LOV 도메인 또는 크립토크롬)을 포함한다. 일 구현예에서, CRISPR Cas13b 효소는 서열-특이적 방식으로 전사 활성에서의 변화를 유도하기 위한 광 유도성 전사 이펙터(LITE)의 일부일 수 있다. 광의 성분은 CRISPR 효소, 광-반응성 시토크롬 이형이량체(예를 들어, 아라비돕시스 탈리아나(Arabidopsis thaliana) 유래), 및 전사 활성화/억제 도메인을 포함할 수 있다. 유도성 DNA 결합 단백질 및 이의 사용을 위한 방법의 추가의 예는 US 61/736465 및 US 61/721,283 및 WO 2014/018423 및 US8889418, US8895308, US20140186919, US20140242700, US20140273234, US20140335620, WO2014093635에서 제공되며, 이는 전체가 참고로 본 명세서에 포함된다. 일부 양태에서, 본 발명은 본 명세서에 기재된 바와 같은 하나 이상의 폴리뉴클레오티드, 하나 이상의 벡터, 하나 이상의 이의 전사물, 및/또는 이로부터의 하나 이상의 단백질을 숙주 세포에 전달하는 것을 포함하는 방법을 제공한다. 일부 양태에서, 본 발명은 이러한 방법에 의해 생산된 이러한 세포, 및 상기 세포를 포함하거나 상기 세포로부터 생산된 유기체(예컨대, 동물, 식물, 또는 균류)를 제공한다. 일부 구현예에서, 가이드 RNA 또는 crRNA와 조합된(및 선택적으로 이와 복합체화된) RNA-표적화 이펙터 단백질은 세포로 전달된다. 통상의 바이러스 및 비-바이러스 기반의 유전자 운반 방법을 사용하여 핵산을 포유동물 세포 또는 표적 조직에 도입할 수 있다. 이러한 방법을 사용하여 RNA-표적화 시스템의 성분을 인코딩하는 핵산을 배양물 중의 또는 숙주 유기체 내의 세포로 투여할 수 있다. 비-바이러스 벡터 전달 시스템은 DNA 플라스미드, RNA(예를 들어, 본 명세서에 기재된 벡터의 전사물), 네이키드 핵산, 및 리포좀과 같은 전달 비히클과 복합체화된 핵산을 포함한다. 바이러스 벡터 전달 시스템은 DNA 및 RNA 바이러스를 포함하며, 이는 세포로의 전달 후에 에피솜 또는 통합된 게놈을 갖는다.

유전자 치료 절차에 대한 검토는 하기를 참조한다: Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology, Doerfler and BΦhm (eds) (1995); 및 Yu et al., Gene Therapy 1:13-26 (1994). 핵산의 비-바이러스 전달 방법은 리포펙션, 뉴클레오펙션, 미세주입, 바이오리스틱스, 비로좀, 리포좀, 면역리포좀, 다가양이온 또는 지질:핵산 콘쥬게이트, 네이키드 DNA, 인공 비리온, 및 DNA의 제제-향상 흡수를 포함한다. 리포펙션은 예를 들어 미국 특허 번호 5,049,386, 4,946,787; 및 4,897,355에 기재되고, 리포펙션 시약은 상업적으로 판매된다 (예를 들어, Transfectam™ 및 Lipofectin™). 폴리뉴클레오티드의 효율적인 수용체-인식 리포펙션에 적절한 양이온 및 중성 지질은 Felgner, WO 91/17424; WO 91/16024의 것을 포함한다. 전달은 세포로(예를 들어, 시험관내 또는 생체외 투여) 또는 표적 조직으로(예를 들어, 생체내 투여) 이루어질 수 있다.

[0658] **조건 모델**

[0659] 본 발명의 방법은, 예컨대 관심 돌연변이의 모델 또는 질환 모델을 통하여, 질환 모델로서 사용될 수 있는 식물, 동물 또는 세포를 형성하고/형성하거나 관심 유전적 또는 후생유전적 병태를 연구하는 데 사용될 수 있다. 본 명세서에서 사용된 바와 같이 "질병"은 대상체에서 질병, 장애 또는 징후를 지칭한다. 예를 들어, 본 발명의 방법은 질병과 연관된 하나 이상의 핵산 서열에서의 변형을 포함하는 동물 또는 세포, 또는 질병과 연관된 하나 이상의 핵산 서열의 발현이 변경된 식물, 동물 또는 세포를 만드는데 사용될 수 있다. 그러한 핵산 서열은 질병 연관 단백질 서열을 인코딩할 수 있거나, 질병 연관 제어 서열일 수 있다. 이에 따라, 본 발명의 구현예에서, 식물, 대상체, 환자, 생물 또는 세포는 비-인간 대상, 환자, 생물 또는 세포일 수 있는 것으로 이해된다. 따라서, 본 발명은 본 방법에 의해서 생성된 식물, 동물 또는 세포, 또는 그것의 자손을 제공한다. 그 자손은 생산된 식물 또는 동물의 클론일 수 있거나, 추가의 바람직한 특성을 그의 자손

에게 유전자 이입(introgress)하기 위해 동일 종의 다른 개체와의 교잡에 의한 유성 생식으로부터의 것일 수 있다. 세포는 다세포 생물, 특히 동물 또는 식물의 경우 생체 내 또는 생체 외일 수 있다. 세포가 배양중인 경우, 적절한 배양 조건이 충족된다면, 바람직하게 세포가 이 목적을 위해 적절히 개조된다면 세포주가 확립될 수 있다(예를 들어, 줄기세포). 본 발명에 의해서 생성된 박테리아 세포주가 또한 고려된다. 따라서, 세포주가 또한 고려된다. 일부 방법에서, 질병 모델은 동물 또는 세포에서 돌연변이의 효과 및 질병 연구에서 흔히 사용되는 기준을 이용하여 질병의 발생 및/또는 진전을 연구하는데 사용될 수 있다. 대안적으로, 그러한 질병 모델은 약학 활성 화합물의 질병에 대한 효과를 연구하는데 유용하다. 일부 방법에서, 질병 모델은 잠재적인 유전자 치료 전략의 효능을 평가하는데 사용될 수 있다. 즉, 질환-관련 RNA는 질환 발달 및/또는 진행이 표시되거나 억제되거나 감소되도록 변형될 수 있고, 이어서 진행 또는 억제 또는 감소에 대한 화합물의 효과가 시험된다.

[0660] 표 1 Cas13b 이펙터 단백질 및 이의 복합체 및 이를 인코딩하는 핵산 분자 및 이를 사용하는 방법을 사용하는 본 발명의 실시에 유용한 것은, 다음을 참조한다: Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013). [인쇄 전 전자 출판]; 최종 편집 된 형태로 다음과 같이 공개됨: Science. 2014 Jan 3; 343(6166): 84-87. Shalem 등은 게놈-전반 규모에서 유전자 작용을 질의하기 위한 새로운 방식을 이용한다. 그들의 연구는 64,751 개의 독특한 가이드 서열을 이용하여 18,080 개 유전자를 표적화시킨 게놈-규모의 CRISPR-Cas9 낙아웃(GeCKO) 라이브러리의 전달이 인간 세포에서 음성 및 양성 선택 스크리닝 둘 모두를 가능하게 하는 것을 보여주었다. 먼저, 상기 저자는 암 및 만능성 줄기 세포에서세포 생존력에 필수적인 유전자를 확인하기 위한 GeCKO 라이브러리의 사용을 보여주었다. 다음으로, 흑색종 모델에서, 상기 저자는 유전자의 소실이 돌연변이 단백질 키나제 BRAF를 억제하는 치료제인 베무라페닙(vemurafenib)에 대한 내성과 관련된 유전자를 스크리닝하였다. 그들의 연구는 가장 높은 순위의 후보물질이 이전에 확인된 유전자 NF1 및 MED12뿐만 아니라 새로운 히트(hit) NF2, CUL3, TADA2B, 및 TADA1을 포함한 것을 나타내었다. 저자들은 동일한 유전자를 표적화하는 독립적인 가이드 RNA 및 높은 비율의 히트 확증 간 높은 수준의 일관성을 관찰하였고, 따라서, Cas9에 의한 게늄 규모 스크리닝의 유방성을 입증하였다. 미국 특허 공개 번호 US20140357530; 및 PCT 특허 공개 WO2014093701 (본 명세서에 참조로 포함됨)가 또한 참조된다.

[0661] 용어 "연관된"이 작용 도메인의 Cas13b 이펙터 단백질 또는 어댑터 단백질로의 연결에 관하여 본 명세서에서 사용된다. 이는 하나의 분자가 다른 분자에 대해, 예를 들어 어댑터 단백질과 기능성 도메인 사이, 또는 Cas13b 이펙터 단백질과 기능성 도메인 사이에 어떻게 "연관"되는지와 관련하여 사용된다. 이러한 단백질-단백질 상호작용의 경우, 이라힌 연합은 항체가 에피토프를 인식하는 방식에 있어서 인식에 관하여 생각될 수 있다. 대안적으로, 하나의 단백질은 2 개의 융합, 예를 들어 또 다른 서브유닛에 융합된 하나의 서브유닛을 통해, 또 다른 단백질과 연합될 수 있다. 융합은 예를 들어 각각의 단백질 또는 서브유닛을 인코딩하는 뉴클레오티드 서열을 함께 스플라이싱을 통해, 통상적으로 하나의 아미노산 서열의 다른 아미노산 서열로의 부가에 의해 일어난다. 대안적으로, 이는 본질적으로 2 개의 분자 간의 결합 또는 직접 연결, 예컨대 융합 단백질로 볼 수 있다. 어쨌든, 융합 단백질은 관심의 두 서브유닛 사이(즉, 효소와 작용 도메인 또는 어댑터 단백질와 작용 도메인 사이)의 링커를 포함할 수 있다. 따라서, 일부 구현예에서, Cas13b 이펙터 단백질 또는 어댑터 단백질은 거기 결합함으로써 작용 도메인과 연관된다. 다른 구현예에서, Cas13b 이펙터 단백질 또는 어댑터 단백질은 작용 도메인과 함께 연관되는데, 이는 이들 둘이 함께, 선택적으로 중간체 링커를 통해 융합되기 때문이다.

[0662] **<u>Cas13b 이펙터 단백질 복합체는 식물에서 사용될 수 있다</u>**

[0663] 일부 구현예에서 본 발명은 세포 또는 생물을 변형시키는 방법을 포함한다. 세포는 원핵생물 세포 또는 진핵생물 세포일 수 있다. 세포는 포유동물 세포일 수 있다. 포유 동물 세포는 많은 경우 비-인간 영장류, 소, 돼지, 설치류 또는 마우스 세포일 수 있다. 세포는 가금류, 어류 또는 새우와 같은 비-포유 동물 진핵 세포일 수 있다. 세포는 식물일 수도 있다. 식물 세포는 카사바, 옥수수, 수수, 밀 또는 쌀과 같은 작물 식물의 것일 수 있다. 식물 세포는 또한 조류, 수목 또는 채소의 것일 수 있다. 본 발명에 의해 세포에 도입된 변형은, 세포 및 세포의 자손이 항체, 전분, 알코올 또는 기타 바람직한 세포 산출물과 같은 생물학적 생성물의 개선된 생산에 대해 변경되도록 할 수 있다. 본 발명에 의해 세포에 도입된 변형은 세포 및 세포의 자손이 생성된 생물학적 생성물을 변화시키는 변경을 포함하도록 하는 것일 수 있다. 시스템은 하나 이상의 상이한 벡터를 포함할 수 있다. 본 발명의 측면에서, 이펙터 단백질은 바람직한 세포 유

형, 우선적으로 진핵생물 세포, 바람직하게 포유동물 세포 또는 인간 세포에서 발현을 위해 코돈 최적화된다.

Cas13b 시스템(들)(예를 들어, 단일 또는 다중화)은 작물 유전체학에서의 최근의 진전과 함께 사용될 수 있다. 이러한 CRISPR 시스템(들)은 예를 들어, 식물 세놈을 형질전환하거나, 또는 식물(들)을 생성시키거나, 동정하거나, 개발하거나, 최적화하거나, 또는 그에 특질(들) 또는 특징(들)을 부여하기 위한, 식물 유전자 또는 게놈의 신속한 조사 및/또는 선택 및/또는 문의 및/또는 비교 및/또는 조작 및/또는 형질전환을 위해, 효율적이고 비용 효과적으로 식물 유전자 또는 게놈 또는 전사체 문의 또는 편집 또는 조작을 수행하는데 사용할 수 있다. 따라서 식물, 특성 또는 특징의 새로운 조합을 갖는 새로운 식물 또는 향상된 특성을 갖는 새로운 식물의 개선된 생산이 있을 수 있다. 그러한 CRISPR 이펙터 단백질 시스템(들)은 식물에 대하여 부위-지정된 통합(SDI) 또는 유전자 편집(GE) 또는 임의의 근친 역육종(Near Reverse Breeding)(NRB) 또는 역육종(Reverse Breeding)(RB) 기법에서 사용될 수 있다. 따라서, 동물 세포에 대한 본 명세서의 참고문헌 또한 달리 명백하지 않은 경우, 식물 세포에 또한 적절한 변형을 가하여 적용될 수 있고, 본 명세서에서 감소된 표적외 효과를 갖는 효소 및 그러한 효소를 사용하는 시스템은 본 명세서에서 언급된 것들을 포함하여, 식물 응용분야에서 사용될 수 있다. 제공된 이펙터 단백질 (표 1 Cas13b) 및 적절한 가이드 (crRNA) 및 그의 자손에 의해 변형된 조작된 식물이 제공된다. 여기에는 이펙터 단백질 및 적합한 가이드를 통한 표적화에 의해, 밀, 보리, 쌀, 대두 또는 옥수수와 같은 질병 또는 가뭄에 강한 작물; 자가 수분 능력을 제거하거나 감소시키도록 변형된 식물(그러나 대신에 선택적으로 잡종이 될 수 있음); 및 면역원성 단백질이 비활성화, 파괴 또는 붕괴된 땅콩 및 견과류와 같은 알러지성 식품이 포함된다. 고전적인 CRISPR-Cas 시스템을 사용하는 임의 측면은 Cas 단백질 애그노스틱, 예를 들어 Cas13b 이펙터 단백질 시스템인 CRISPR 시스템을 사용하는데 적합화될 수 있다.

[0664] **치료적 처치**

[0665] 본 발명의 시스템은 본 출원이 시스템의 알려진 조작에 대한 기초를 제공하기 때문에, 치료법, 어세이 및 다른 응용법을 포함하여, 본 개시로부터, 과도한 실험없이, 이전의 RNA 절단 기술 분야에 적용될 수 있다. 본 발명은 RNA의 과발현, 독성 RNA 및/또는 돌연변이된 RNA (예컨대, 예를 들어, 스플라이싱 결함 또는 절두)로 초래된 질환의 치료적 처치를 제공한다. 독성 RNA의 발현은 뇌, 심장 또는 골격근에서 핵 봉입체의 형성 및 후기-발병 퇴행성 변화와 연관될 수 있다. 최고로 연구된 예인, 근긴장성 이영양증에서, 독성 RNA의 주요 병원성 효과는 결합 단백질을 격리시켜서 선택적 스플라이싱의 조절을 손상시키는 것으로 보인다 (Hum. Mol. Genet. (2006) 15 (suppl 2): R162-R169). 근긴장성 이영양증 [dystrophia myotonica (DM)]은 이것이 극도로 광범위한 임상적 특성을 일으키기 때문에 유전학자들에게 특히 흥미롭다. 부분적인 목록에는 근육 소모, 백내장, 인슐린 저항성, 정소 위축, 심장 정도의 지연, 피하 종양 및 인지에 대한 효과가 포함된다. 이제 DM 타입 1 (Dm1)이라고 하는 DM의 고전적인 형태는 세포질 단백질 키나제를 인코딩하는 유전자인 DMPK의 3-미번역 영역 (UTR) 에서 CTG 반복부의 확장에 의해 초래된다.

[0666] 선천성 면역 시스템은 DNA 또는 RNA 센싱이라고 하는, 감염된 세포 내부의 바이러스 핵산을 주로 인식하여 바이러스 감염을 검출한다. 시험관내 RNA 센싱 어세이는 특이적 RNA 기질을 검출하기 위해 사용될 수 있다. RNA 표적화 이펙터 단백질은 예를 들어 살아있는 세포에서 RNA-기반 센싱에 사용될 수 있다. 응용의 예는 예를 들어, 질환-특이적 RNA의 센싱에 의한 진단이다. 본 발명의 RNA 표적화 이펙터 단백질 (표 1 Cas13b)은 특히 RNA 바이러스에 대항한, 항바이러스 활성에 대해 더욱 사용될 수 있다. 이펙터 단백질 (표 1 Cas13b)은 선택된 바이러스 RNA 서열에 대해 선택적인 적합한 가이드 RNA를 사용하여 바이러스 RNA에 대해 표적화할 수 있다. 특히, 이펙터 단백질은 RNA, 예컨대 단일 가닥 RNA를 절단하는 활성 뉴클레아제일 수 있다.

상기 기준 RNA, 표적 RNA에 대한 본 발명의 효소 시스템의 치료적 용량은 약 0.1 내지 약 2 mg/kg으로 고려되고 용량은 반응을 모니터링하면서 순차적으로 투여될 수 있고 필요하다면, 환자 당 최대 약 7 내지 10 용량으로 반복된 용량일 수 있다. 유리하게, 샘플은 치료의 효율성을 알아내기 위한 치료 섭생 동안 각 환자로부터 수집된다. 예를 들어, RNA 샘플은 단리되어 발현이 감소 또는 완화되는지 여부를 결정하기 위해 정량된다. 이러한 진단은 당업자의 이해 범위 내이다.

[0667] **전사물 검출 방법**

[0668] 본 발명의 이펙터 단백질 (표 1 Cas13b) 및 시스템은 세포 또는 다른 샘플 중 RNA의 특이적 검출에 유용하다.

관심 RNA 표적의 존재 하에서, 가이드-의존적 Cas13b 뉴클레아제 활성은 이차적인 표적에 대한 비특이적 RNAse 활성에 의해 수행될 수 있다. RNAse 활성을 이용하기 위해, 필요한 모든 것은 검출가능하게 절단할 수 있는 리포터 기질이다. 예를 들어, 리포터 분자는 한쪽 말단은 형광발광 리포터 분자(플루오르) 및 다른 쪽에는 소광제가 태그화된, RNA를 포함할 수 있다. Cas13b RNase 활성의 부재하에서, 소광제의 물리적 근접성은 플루오르 유래의 형광발광을 낮은 수준으로 약화시킨다. Cas13b 표적 특이적 절단이 관심 RNA 표적

및 적합한 가이드 RNA의 존재에 의해 활성화될 때, RNA-함유 리포터 분자는 비특이적으로 절단되고 플루오르 및 소광제가 공간적으로 분리된다. 이는 적절한 파장의 광에 의해 여기될 때 플루오르가 검출가능한 신호를 발광하도록 한다. 하나의 예시적인 어세이 방법에서, Cas13b 이펙터, 관심 표적-특이적 가이드 RNA, 및 리포터 분자는 세포 샘플에 첨가된다. 형광발광성의 증가는 관심 RNA 표적의 존재를 의미한다. 다른 예시적인 방법에서, 검출 어레이가 제공된다. 어레이의 각 위치는 Cas13b 이펙터, 리포터 분자, 및 관심 표적-특이적 가이드 RNA가 제공된다. 수행하려는 어레이에 의존적으로, 어레이의 각 위치에서 관심 표적-특이적 가이드 RNA는 동일할 수 있거나, 상이할 수 있거나, 또는 이의 조합일 수 있다. 상이한 관심 표적-특이적 가이드 RNA는 예를 들어 단일 공급원 샘플 중에서 하나 이상의 표적에 대해 시험되는 것이 바람직할 때 제공될 수 있다. 동일한 관심 표적-특이적 가이드 RNA는 예를 들어, 동일한 표적에 대해 다수 샘플을 시험하는 것이 바람직할 때, 각 위치에서 제공될 수 있다.

[0669] 특정 실시형태에서, Cas13b는 시험관 내 시스템 또는 세포에서 일시적으로 또는 안정적으로 제공되거나 발현되고, 비-특이적으로 절단성 세포 핵산을 표적으로 하거나 촉발된다. 한 실시형태에서, Cas13b는 ssDNA, 예를 들어 바이러스성 ssDNA를 녹다운시키도록 조작된다. 다른 실시형태에서, Cas13b는 RNA를 녹다운시키도록 조작된다. 녹다운이 세포 또는 시험관 내 시스템에 존재하는 표적 DNA에 의존하거나, 시스템 또는 세포에 표적 핵산의 첨가에 의해 촉발되도록 시스템을 고안할 수 있다.

[0670] 한 실시형태에서, Cas13b 시스템은 비정상 DNA 서열의 존재에 의해 구별될 수 있는 세포의 서브세트에서 RNA를 비-특이적으로 절단하도록, 예를 들어 비정상 DNA의 절단이 불완전하거나 비효과적일 수 있도록 조작된다. 하나의 비-제한적인 예에서, 암 세포에 존재하고 세포 형질전환을 유도하는 DNA 전좌가 표적화된다. 염색체 DNA를 겪고 복구되는 세포의 소집단이 생존할 수 있는 반면, 비-특이적 부수적인 리보뉴클레아제 활성은 유리하게는 잠재적 생존자의 세포 사멸을 초래한다.

[0671] 부수적 활동은 최근 많은 임상 진단에 유용한 SHERLOCK이라고하는 매우 민감하고 특이적인 핵산 검출 플랫폼에 활용되었다. (Gootenberg, J. S. et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356, 438-442 (2017)).

[0672] 본 발명에 따르면, 조작된 Cas13b 시스템은 RNA 엔도뉴클레아제 활성에 최적화되어 있으며 포유동물 세포에서 발현될 수 있고 세포에서 리포터 분자 또는 전사체를 효과적으로 녹다운시키는 것을 목표로 한다.

[0673] 등온 증폭을 갖는 조작된 Cas13b의 부수 효과는 높은 민감성 및 단일-염기 미스매치 특이성을 갖는 신속한 DNA 또는 RNA 검출을 제공하는 CRISPR-기반 진단을 제공한다. Cas13b-기반 분자 검출 플랫폼은 바이러스의 특정 균주를 검출하고, 병원성 박테리아, 유전자형 인간 DNA를 구별하고, 무-세포 종양 DNA 돌연변이를 식별하는데 사용된다. 또한, 반응 시약은 저온-사슬 독립 및 장기 저장을 위해 동결건조될 수 있고, 현장 적용을 위해 서류상 쉽게 재구성될 수 있다.

[0674] 휴대용 플랫폼에서 높은 민감성 및 단일-염기 미스매치 특이성을 갖는 핵산을 신속하게 검출하는 능력은 질병 진단 및 모니터링, 역학 및 일반적인 실험실 작업에 도움이 될 수 있다. 핵산을 검출하기 위한 방법이 존재하지만, 민감성, 특이성, 단순성, 비용 및 속도간에 절충점이 있다.

[0675] Microbial Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) 및 CRISPR-연관 (CRISPR-Cas) 적응 면역 시스템은 CRISPR-기반 진단 (CRISPR-Dx)에 활용할 수 있는 프로그램가능한 엔도뉴클레아제를 포함한다. Cas13b는 특정 DNA 감지를 위한 플랫폼을 제공하기 위해 CRISPR RNAs (crRNAs)로 재프로그램될 수 있다. DNA 표적의 인식에 따라, 활성화된 Cas13b는 근처의 비-표적 핵산 (즉, RNA 및/또는 ssDNA)의 "부수적" 절단에 관여한다. 상기 crRNA-프로그램된 부수적 절단 활성은 Cas13b가 프로그램된 세포 사멸을 유발하거나 표지된 RNA 또는 ssDNA의 비특이적 분해에 의해 생체 내에서 특정 DNA의 존재를 검출할 수 있게 한다. 여기에 표적의 실시간 검출을 가능하게 하는, 상업적인 리포터 RNA의 핵산 증폭 및 Cas13b-매개된 부수적 절단에 기초하여 높은 민감성을 갖는 시험관 내 핵산 검출 플랫폼이 기재되어 있다.

[0676] 특정 예시적 실시형태에서, Cas13b 이펙터 단백질은 표 1에서 확인된 유기체로부터 유래된다. 특정 예시적 실시형태에서, Cas13b 이펙터 단백질은 Bergeyella zoohelcum, Prevotella intermedia, Prevotella buccae, Porphyromonas gingivalis, Bacteroides pyogenes, Alistipes sp. ZOR0009, Prevotella sp. MA2016, Riemerella anatipestifer, Prevotella aurantiaca, Prevotella saccharolytica, Myroides odoratimimus CCUG 10230, Capnocytophaga canimorsus, Porphyromonas gulae, Prevotella sp. P5-125, Flavobacterium branchiophilum, Myroides odoratimimus, Flavobacterium columnare, 또는 Porphyromonas sp. COT-052 OH4946

로부터 선택되는 유기체로부터 유래된다. 다른 실시형태에서, 하나 이상의 가이드 RNA는 질병 상태를 진단하는 하나 이상의 표적 RNA 서열에 결합하도록 디자인된다.

[0677] 특정 예시적 실시형태에서, RNA-기반 마스킹 구축물은 검출가능한 양성 신호의 생성을 억제하거나, 또는 RNA-기반 마스킹 구축물은 검출가능한 양성 신호를 마스킹하거나, 대신 검출가능한 음성 신호를 생성함으로써 검출가능한 양성 신호의 생성을 억제하거나, 또는 RNA-기반 마스킹 구축물은 리포트 구축물에 의해 인코딩된 유전자 산물의 생성을 억제하는 침묵 RNA를 포함하고, 여기서 유전자 산물은 발현될 때 검출가능한 양성 신호를 생성한다.

[0678] 또다른 예시적 실시형태에서, RNA-기반 마스킹 구축물은 음성 검출가능한 신호를 생성하는 리보자임이고, 양성 검출가능한 신호는 리보자임이 비활성화될 때 생성된다. 하나의 예시적 실시형태에서, 리보자임은 기질을 제 1 색상으로 변환하고, 리보자임이 비활성화될 때 기질은 제 2 색상으로 변환된다. 다른 예시적 실시형태에서, RNA-기반 마스킹제는 효소를 격리시키는 앱타머이고, 여기서 효소는 기질에 작용함으로써 앱타머로부터 방출시 검출가능한 신호를 생성하거나, 앱타머는 앱타머로부터 방출시 결합하여 검출가능한 신호를 생성하는 작용제의 쌍을 격리한다.

[0679] 다른 예시적 실시형태에서, RNA-기반 마스킹 구축물은 검출가능한 리간드 올리고뉴클레오티드 및 마스킹 성분이 부착된 RNA 올리고뉴클레오티드를 포함한다. 특정 예시적 실시형태에서, 검출가능한 리간드는 형광단이고 마스킹 성분은 소광제 분자이다.

[0680] 또 다른 측면에서, 본 발명은 하기 단계를 포함하는, 샘플에서 표적 RNA를 검출하는 방법을 제공한다: 샘플 또는 샘플 세트를 하나 이상의 개별 이산 부피로 분배하는 단계로서, 개별 이산 부피는 이펙터 단백질, 하나 이상의 가이드 RNA, 및 RNA-기반 마스킹 구축물을 포함하는 단계; 하나 이상의 표적 분자에 하나 이상의 가이드 RNA의 결합을 허용하기에 충분한 조건 하에서 샘플 또는 샘플 세트를 배양하는 단계; 하나 이상의 표적 분자에 대한 하나 이상의 가이드 RNA의 결합을 통해 CRISPR 이펙터 단백질을 활성화시키는 단계로서, CRISPR 이펙터 단백질의 활성화가 검출 가능한 양성 신호가 생성되도록 RNA-기반 마스킹 구축물의 변형을 도출하는 단계; 및 검출가능한 양성 신호를 검출하는 단계로서, 검출가능한 양성 신호의 검출은 샘플에서 하나 이상의 표적 분자의 존재를 나타내는 단계.

[0681] 또 다른 측면에서, 본 발명은 하기 단계를 포함하는, 샘플에서 펩티드를 검출하는 방법을 제공한다: 샘플 또는 샘플 세트를 하나 이상의 개별 이산 부피로 분배하는 단계로서, 개별 이산 부피는 펩티드 검출 앱타머를 포함하고, CRISPR 시스템은 이펙터 단백질, 하나 이상의 가이드 RNA, 및 RNA-기반 마스킹 구축물을 포함하고, 펩티드 검출 앱타머는 마킹된 RNA 폴리머라제 부위를 포함하고 하나 이상의 표적 분자에 결합하도록 배열되는 단계; 하나 이상의 표적 분자에 펩티드 검출 앱타머의 결합을 허용하기에 충분한 조건 하에서 샘플 또는 샘플 세트를 배양하는 단계로서, 상응하는 표적 분자에 대한 앱타머의 결합이 RNA 폴리머라제 부위를 노출시켜 촉발자 RNA의 RNA 합성을 도출하는 단계; 촉발자 RNA에 대한 하나 이상의 가이드 RNA의 결합을 통해 CRISPR 이펙터 단백질을 활성화시키는 단계로서, CRISPR 이펙터 단백질의 활성화가 검출 가능한 양성 신호가 생성되도록 RNA-기반 마스킹 구축물의 변형을 도출하는 단계; 및 검출가능한 양성 신호를 검출하는 단계로서, 검출가능한 양성 신호의 검출은 샘플에서 하나 이상의 표적 분자의 존재를 나타내는 단계.

[0682] 특정 예시적 실시형태에서, 하나 이상의 가이드 RNA는 질병 상태를 진단하는 하나 이상의 표적 분자에 결합하도록 디자인된다. 특정 다른 예시적 실시형태에서, 질환 상태는 감염, 장기 질환, 혈액 질환, 면역계 질환, 암, 뇌 및 신경계 질환, 내분비 질환, 임신 또는 출산-관련 질환, 유전 질환, 또는 환경적으로 획득된 질환, 암 또는 진균 감염, 박테리아 감염, 기생충 감염 또는 바이러스 감염이다.

[0683] 특정 예시적 실시형태에서, RNA-기반 마스킹 구축물은 검출가능한 양성 신호의 생성을 억제하거나, 또는 RNA-기반 마스킹 구축물은 검출가능한 양성 신호를 마스킹하거나, 대신 검출가능한 음성 신호를 생성함으로써 검출가능한 양성 신호의 생성을 억제하거나, 또는 RNA-기반 마스킹 구축물은 리포트 구축물에 의해 인코딩된 유전자 산물의 생성을 억제하는 침묵 RNA를 포함하고, 여기서 유전자 산물은 발현될 때 검출가능한 양성 신호를 생성하거나, 또는 RNA-기반 마스킹 구축물은 음성 검출가능한 신호를 생성하는 리보자임이고, 양성 검출가능한 신호는 리보자임이 비활성화될 때 생성된다. 다른 예시적 실시형태에서, 리보자임은 기질을 제 1 색상으로 변환하고, 리보자임이 비활성화될 때 기질은 제 2 색상으로 변환되고, RNA-기반 마스킹제는 앱타머이거나, 앱타머는 효소를 격리시키고, 여기서 효소는 기질에 작용함으로써 앱타머로부터 방출시 검출가능한 신호를 생성하거나, 앱타머는 앱타머로부터 방출시 결합하여 검출가능한 신호를 생성하는 작용제의 쌍을 격리한다. 또 다른 실시형태에서, RNA-기반 마스킹 구축물은 RNA 올리고뉴클레오티드의 제 1 말단에 검출가능한 리간드 및 RNA 올

리고뉴클레오티드의 제 2 말단에 마스킹 성분을 갖는 RNA 올리고뉴클레오티드를 포함하거나, 검출가능한 리간드는 형광단이고 마스킹 성분은 소광제 분자이다.

[0684] 양 및 제제에 대한 것을 포함하여, 방법, 재료, 전달 비히클, 벡터, 입자, AAB, 및 이의 제조 및 사용을 포함한, CRISPR-Cas 시스템, 이의 성분, 및 이러한 성분의 전달에 대한 일반 정보와 관련하여, 본 발명의 실시에 유용한 모든 것은 다음을 참조한다: 미국 특허 번호 8,999,641, 8,993,233, 8,945,839, 8,932,814, 8,906,616, 8,895,308, 8,889,418, 8,889,356, 8,871,445, 8,865,406, 8,795,965, 8,771,945 및 8,697,359; 미국 특허 공개 US 2014-0310830 (US APP. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) 및 US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); 유럽 특허 EP 2 784 162 B1 및 EP 2 771 468 B1; 유럽 특허 출원 EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), 및 EP 2 784 162 (EP14170383.5); 및 PCT 특허 공개 PCT 특허 공개 WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO 2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO 2014/093701 (PCT/US2013/074800), WO 2014/018423 (PCT/US2013/051418), WO 2014/204723 (PCT/US2014/041790), WO 2014/204724 (PCT/US2014/041800), WO 2014/204725 (PCT/US2014/041803), WO 2014/204726 (PCT/US2014/041804), WO 2014/204727 (PCT/US2014/041806), WO 2014/204728 (PCT/US2014/041808), WO 2014/204729 (PCT/US2014/041809). 또한 미국 가특허 출원 61/758,468; 61/802,174; 61/806,375; 61/814,263; 61/819,803 및 61/828,130 (각각 January 30, 2013; March 15, 2013; March 28, 2013; April 20, 2013; May 6, 2013 및 May 28, 2013 에 출원) 을 참조한다. 또한 미국 가특허 출원 61/836,123 (June 17, 2013 에 출원) 을 참조한다. 부가적으로 미국 가특허 출원 61/835,931, 61/835,936, 61/836,127, 61/836, 101, 61/836,080 및 61/835,973 (각각 June 17, 2013 에 출원) 을 참조한다. 추가로 미국 가특허 출원 61/862,468 및 61/862,355 (August 5, 2013 에 출원); 61/871,301 (August 28, 2013 에 출원); 61/960,777 (September 25, 2013 에 출원) 및 61/961,980 (October 28, 2013 에 출원) 을 참조한다. 추가로 다음을 참조한다: PCT 특허 출원 번호: PCT/US2014/041803, PCT/US2014/041800, PCT/US2014/041809, PCT/US2014/041804 및 PCT/US2014/041806 (각각 June 10, 2014 에 출원됨- 6/10/14); PCT/US2014/041808 (June 11, 2014 에 출원); 및 PCT/US2014/62558 (October 28, 2014 에 출원), 및 미국 가특허 출원 일련 번호: 61/915,150, 61/915,301, 61/915,267 및 61/915,260 (각각, December 12, 2013 에 출원); 61/757,972 및 61/768,959 (January 29, 2013 및 February 25, 2013 에 출원); 61/835,936, 61/836,127, 61/836,101, 61/836,080, 61/835,973, 및 61/835,931 (June 17, 2013 에 출원); 62/010,888 및 62/010,879 (둘다 June 11, 2014 에 출원); 62/010,329 및 62/010,441 (각각 June 10, 2014 에 출원); 61/939,228 및 61/939,242 (각각 February 12, 2014 에 출원); 61/980,012 (April 15,2014 에 출원); 62/038,358 (August 17, 2014 에 출원); 62/054,490, 62/055,484, 62/055,460 및 62/055,487 (각각 September 25, 2014 에 출원); 및 62/069,243 (October 27, 2014 에 출원). 또한 미국 가특허 출원 번호 62/055,484, 62/055,460, 및 62/055,487 (September 25, 2014 에 출원); 미국 가특허 출원 61/980,012 (April 15, 2014 에 출원); 및 미국 가특허 출원 61/939,242 (February 12, 2014 에 출원) 을 참조한다. PCT 출원, 특히 미국을 지정한, 출원 번호 PCT/US14/41806 (June 10, 2014 에 출원) 를 참조한다. 미국 가특허 출원 61/930,214 (January 22, 2014 에 출원) 을 참조한다. 미국 가특허 출원 61/915,251; 61/915,260 및 61/915,267 (각각 December 12, 2013 에 출원) 을 참조한다. 미국 가특허 출원 USSN 61/980,012 (April 15, 2014 에 출원) 을 참조한다. PCT 출원, 특히 미국을 지정한, 출원 번호 PCT/US14/41806 (June 10, 2014 에 출원) 를 참조한다. 미국 가특허 출원 61/930,214 (January 22, 2014 에 출원) 을 참조한다. 미국 가특허 출원 61/915,251; 61/915,260 및 61/915,267 (각각 December 12, 2013 에 출원) 을 참조한다.

[0685] 또한 하기를 언급한다: US 출원 62/091,455, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US 출원 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US 출원 62/091,462, 12-Dec-14, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US 출원 62/096,324, 23-Dec-14, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US 출원 62/091,456, 12-Dec-14, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; US 출원 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US 출원 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; US 출원 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US 출원 62/098,059, 30-Dec-14, RNA-TARGETING SYSTEM; US 출원 62/096,656, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US 출원 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US 출원 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US 출원 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US 출원 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US 출원 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US 출원 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US 출원 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US 출원 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US 출원 62/054,675, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US 출원 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US 출원 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US 출원 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US 출원 62/087,475, 4-Dec-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US 출원 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US 출원 62/087,546, 4-Dec-14, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; 및 US 출원 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[0686] 이들 특허, 특허 공보 및 출원 각각, 및 거기에 또는 그들 절차 중에 인용된 모든 문헌("출원 인용 문헌") 및 상기 출원 인용 문헌에 인용되거나 참조된 모든 문헌은 상기 출원 인용 문헌에 또는 상기 출원 인용 문헌의 임의의 문헌 및 본원에 참조로서 포함된 임의의 문헌에서 언급된 임의의 제품에 대한 임의의 지침서, 설명서, 제품 명세서, 및 제품 시이트와 함께, 참조로서 본원에 포함되며, 본 발명의 실시에 이용될 수 있다. 모든 문헌(예를 들어, 이들 특허, 특허 공보 및 출원 및 출원 인용 문헌)은 각각의 개별적 문헌이 참조로서 포함하는 것이 구체적으로 그리고 개별적으로 표시된 것과 동일한 정도로 참조로서 본원에 포함된다.

[0687] CRISPR-Cas 시스템에 대한 일반적인 정보는 다음의 문헌들을 언급한다 (참조로 본 명세서에 편입됨):

[0688] ▷ Multiplex genome engineering using CRISPR/Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. Science Feb 15;339(6121):819-23 (2013);

[0689] ▷ RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013);

[0690] ▷ One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013);

[0691] ▷ Optical control of mammalian endogenous transcription and epigenetic states. Konermann S,

Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Nature. Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23 (2013);

[0692] ▷ Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. Cell Aug 28. pii: S0092-8674(13)01015-5 (2013-A);

[0693] ▷ DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013);

[0694] ▷ Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. Nature Protocols Nov;8(11):2281-308 (2013-B);

[0695] ▷ Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013). [Epub ahead of print];

[0696] ▷ Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Cell Feb 27, 156(5):935-49 (2014);

[0697] ▷ Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. Nat Biotechnol. Apr 20. doi: 10.1038/nbt.2889 (2014);

[0698] ▷ CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F. Cell 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014);

[0699] ▷ Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu PD, Lander ES, Zhang F., Cell. Jun 5;157(6):1262-78 (2014).

[0700] ▷ Genetic screens in human cells using the CRISPR/Cas9 system, Wang T, Wei JJ, Sabatini DM, Lander ES., Science. January 3; 343(6166): 80-84. doi:10.1126/science.1246981 (2014);

[0701] ▷ Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation, Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE., (published online 3 September 2014) Nat Biotechnol. Dec;32(12):1262-7 (2014);

[0702] ▷ *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F., (published online 19 October 2014) Nat Biotechnol. Jan;33(1):102-6 (2015);

[0703] ▷ Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015).

[0704] ▷ A split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (published online 02 February 2015) Nat Biotechnol. Feb;33(2):139-42 (2015);

[0705] ▷ Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246-1260, March 12, 2015 (multiplex screen in mouse), and

[0706] ▷ *In vivo* genome editing using *Staphylococcus aureus* Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (published online 01 April 2015), Nature. Apr 9;520(7546):186-91 (2015).

[0707] ▷ Shalem et al., "High-throughput functional genomics using CRISPR-Cas9," Nature Reviews Genetics 16, 299-311 (May 2015).

[0708] ▷ Xu et al., "Sequence determinants of improved CRISPR sgRNA design," Genome Research 25, 1147-1157 (August 2015).

[0709] ▷ Parnas et al., "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks," Cell 162, 675-686 (July 30, 2015).

[0710] ▷ Ramanan et al., CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus," Scientific Reports 5:10833. doi: 10.1038/srep10833 (June 2, 2015)

[0711] ▷ Nishimasu et al., Crystal Structure of Staphylococcus aureus Cas9," Cell 162, 1113-1126 (Aug. 27, 2015)

[0712] ▷ Zetsche *et al*. (2015), "Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system," Cell 163, 759-771 (Oct. 22, 2015) doi: 10.1016/j.cell.2015.09.038. Epub Sep. 25, 2015

[0713] ▷ Shmakov *et al*. (2015), "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems," Molecular Cell 60, 385-397 (Nov. 5, 2015) doi: 10.1016/j.molcel.2015.10.008. Epub Oct 22, 2015

[0714] ▷ Dahlman et al., "Orthogonal gene control with a catalytically active Cas9 nuclease," Nature Biotechnology 33, 1159-1161 (November, 2015)

[0715] ▷ Gao *et al*, "Engineered Cpf1 Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: http://dx.doi.org/10.1101/091611 Epub Dec. 4, 2016

[0716] ▷ Smargon et al. (2017), "Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28," Molecular Cell 65, 618-630 (Feb. 16, 2017) doi: 10.1016/j.molcel.2016.12.023. Epub Jan 5, 2017

[0717] 이들 각각은 본 명세서에 참고문헌으로 통합되고, 본 발명의 실시에서 고려될 수 있으며, 아래 간략히 설명된다:

[0718] ▷ Cong 등은 스트렙토코커스 써모필러스 Cas9 및 또한 스트렙토코커스 피오게네스 Cas9 둘 모두를 기초로 하여 진핵 세포에서 사용하기 위해 II형 CRISPR/Cas 시스템을 조작하였고, Cas9 뉴클레아제가 짧은 RNA에 의해 유도되어 인간 및 마우스 세포에서 DNA의 정밀한 절단을 유도할 수 있는 것을 입증하였다. 이들의 연구는 닉킹 효소로 전환됨에 따라 Cas9가 최소의 돌연변이유발 활성을 갖는 진핵 세포에서 상동성-유도된 수복을 촉진하는데 이용될 수 있음을 추가로 보여준다. 또한, 그들의 연구는 다수의 가이드 서열이 단일의 CRISPR 어레이로 인코딩되어 포유류 게놈 내의 내인성 게놈 유전자좌 위치에서 몇몇의 동시 편집을 가능하게 할 수 있음을 입증하였으며, 이는 RNA-안내된 뉴클레아제 기술의 용이한 프로그램화가능성 및 폭넓은 응용 가능성을 입증한다. 세포에서 서열 특이적 DNA 절단을 프로그램화하기 위해 RNA를 이용하는 이러한 능력은 게놈 조작 도구의 새로운 부류를 규정하였다. 이들 연구는 다른 CRISPR 유전자좌가 포유류 세포로 이식될 수 있고, 또한 포유류 게놈 절단을 매개할 수 있음을 추가로 보여준다. 중요하게는, CRISPR-Cas 시스템의 여러 양태가 그의 효능 및 다능성을 증가시키기 위해 추가로 개선될 수 있음이 예견될 수 있다.

[0719] ▷ Jiang 등은 스트렙토코커스 뉴모니애 및 에스케리키아 콜라이의 게놈에서 정밀한 돌연변이를 도입시키기 위해 이중-RNA와 복합체화된 클러스터링되고 규칙적으로 산재된 짧은 팔린드로믹 반복부(CRISPR)-연합된 Cas9 엔도뉴클레아제를 이용하였다. 상기 접근법은 돌연변이되지 않은 세포를 사멸시키기 위해 표적화된 게놈 부위에서의 이중-RNA:Cas9-유도된 절단에 의존하며, 선별 마커 또는 반대-선택 시스템의 필요를 회피한다.

상기 연구는 편집 주형에서 수행되는 단일- 및 다중 뉴클레오티드 변화를 만들기 위해 짧은 CRISPR RNA(crRNA)의 서열을 변화시킴에 의한 이중-RNA:Cas9 특이성의 재프로그램화를 보고하였다. 상기 연구는 2개의 crRNA의 동시 사용이 멀티플렉스 돌연변이유발을 가능하게 하는 것을 보여준다. 또한, 상기 접근법이 리컴비니어링(recombineering)과 병용하여 이용된 경우, 스트렙토코커스 뉴모니애에서 기재된 접근법을 이용하여 회수된 세포의 거의 100%가 요망되는 돌연변이를 함유하였고, 에스케리키아 콜라이에서, 회수된 65%가 돌연변이를 함유하였다.

[0720] ▷ Wang 등(2013)은 배아 줄기 세포에서 순차적 재조합 및/또는 단일 돌연변이를 갖는 마우스의 시간 소모적 이종교배에 의한 복수 단계로 전통적으로 생성된 다수 유전자에 돌연변이를 보유하는 마우스의 한 단계 생성을 위해 CRISPR/Cas 시스템을 사용하였다. CRISPR/Cas 시스템은 기능적으로 중복인 유전자 및 상위 유전자 상호작용의 생체내 연구를 상당히 가속시킬 것이다.

[0721] ▷ Konermann 등 (2013) 은 DNA-결합 도메인 기반 CRISPR Cas9 효소 및 또한 전사 활성화제 유사 이펙터의 광학적 및 화학적 조절을 가능하게 하는 다능성이고 강력한 기술에 대한 해당 분야의 요구를 다루었다.

[0722] ▷ Ran 등 (2013-A)은 표적화된 이중-가닥 파손을 도입시키기 위해 쌍을 형성하는 가이드 RNA와 Cas9 닉카제 돌연변이체를 병용한 접근법을 기재하였다. 이는 가이드 서열에 의해 특정 게놈 유전자좌로 표적되는 미생물 CRISPR-Cas 시스템으로부터의 Cas9 뉴클레아제의 문제를 다루며, 이는 DNA 표적으로의 특정 미스매치를 용인할 수 있고, 이에 의해 바람직하지 않은 표적외 돌연변이유발을 촉진할 수 있다. 게놈 내의 개별의 닉은 높은 정확도로 수복되므로, 적절한 오프셋 가이드 RNA를 통한 동시 닉킹이 이중-가닥 파손에 필요하고, 표적 절단을 위한 특이적으로 인식된 염기의 수를 연장시킨다. 상기 저자는 쌍을 형성한 닉킹을 이용하는 것이 세포주에서 50 내지 1,500배까지 표적외 활성을 감소시킬 수 있고, 표적에 대한 절단 효율을 희생시키지 않고 마우스 접합체에서 유전자 낙아웃을 촉진하는 것을 입증하였다. 이러한 다능성 전략은 높은 특이성을 필요로 하는 매우 다양한 게놈 편집 응용을 가능하게 한다.

[0723] ▷ Hsu 등 (2013)은 표적 부위의 선택을 알아내고, 표적외 효과를 피하기 위해 인간 세포에서 SpCas9 표적화 특이성을 특성화하였다. 상기 연구는 293T 및 293FT 세포에서 100 개 초과의 예측 게놈 표적외 유전자좌에서 SpCas9-유도된 삽입-결실 돌연변이 수준 및 700 개 초과의 가이드 RNA 변이체를 평가하였다. 상기 저자는 SpCas9가 미스매치의 수, 위치 및 분포에 민감한 서열-의존성 방식으로 상이한 위치에서 가이드 RNA와 표적 DNA 사이의 미스매치를 용인하는 것을 보고하였다. 저자들은 SpCas9-매개 절단이 DNA 메틸화에 의해 영향받지 않고 SpCas9 및 sgRNA의 용량이 표적외 변형을 최소화하도록 적정될 수 있다는 것을 더욱 보여주었다.
또한, 포유류 게놈 조작 응용을 용이하게 하기 위해, 상기 저자는 표적 서열의 선택 및 확인뿐만 아니라 표적외 분석을 안내하기 위해 웹-기반 소프트웨어 도구를 제공한 것을 보고하였다.

[0724] ▷ Ran 등(2013-B)은 포유류 세포에서 비-상동성 말단 연결(NHEJ) 또는 상동성-유도 수복(HDR)뿐만 아니라 하류 작용성 연구를 위한 변형된 세포주의 생성을 통한 Cas9-매개 게놈 편집을 위한 일련의 도구를 기재하였다.
표적외 절단을 최소화시키기 위해, 상기 저자는 쌍을 형성한 가이드 RNA와 함께 Cas9 닉카제 돌연변이체를 이용한 이중-닉킹 전략을 추가로 기재하였다. 상기 저자에 의해 제공된 프로토콜은 표적 부위의 선택, 절단 효율의 평가 및 표적외 활성의 분석을 위한 가이드를 실험적으로 유래하였다. 상기 연구는 표적 디자인로 시작하여, 유전자 변형은 1 내지 2주만큼 적은 기간 내에 달성될 수 있고, 변형된 클론성 세포주가 2 내지 3주 내에 유래될 수 있음을 나타내었다.

[0725] ▷ Shalem 등은 게놈-전반 규모에서 유전자 작용을 질의하기 위한 새로운 방식을 기재한다. 그들의 연구는 64,751 개의 독특한 가이드 서열을 이용하여 18,080 개 유전자를 표적화시킨 게놈-규모의 CRISPR-Cas9 낙아웃(GeCKO) 라이브러리의 전달이 인간 세포에서 음성 및 양성 선택 스크리닝 둘 모두를 가능하게 하는 것을 보여주었다. 먼저, 상기 저자는 암 및 만능성 줄기 세포에서세포 생존력에 필수적인 유전자를 확인하기 위한 GeCKO 라이브러리의 사용을 보여주었다. 다음으로, 흑색종 모델에서, 상기 저자는 유전자의 소실이 돌연변이 단백질 키나제 BRAF를 억제하는 치료제인 베무라페닙(vemurafenib)에 대한 내성과 관련된 유전자를 스크리닝하였다. 그들의 연구는 가장 높은 순위의 후보물질이 이전에 확인된 유전자 NF1 및 MED12뿐만 아니라 새로운 히트(hit) NF2, CUL3, TADA2B, 및 TADA1을 포함한 것을 나타내었다. 저자들은 동일한 유전자를 표적화하는 독립적인 가이드 RNA 및 높은 비율의 히트 확증 간 높은 수준의 일관성을 관찰하였고, 따라서, Cas9에 의한 게놈 규모 스크리닝의 유망성을 입증하였다.

[0726] ▷ Nishimasu 등은 2.5 A° 해상도로 sgRNA 및 그의 표적 DNA와의 복합체에서 스트렙토코커스 피오게네스 Cas9의 결정 구조를 보고하였다. 상기 구조는 그들의 계면에서 양으로 하전된 홈에서 sgRNA:DNA 이형이중가닥을 수용하는 표적 인식 및 뉴클레아제 엽으로 구성된 2엽 구조를 나타내었다. 인식 엽이 sgRNA 및 DNA 결합에 필수적인 반면, 뉴클레아제 엽은 각각 표적 DNA의 상보적 및 비-상보적 가닥의 절단을 위해 적절하게 위치된 HNH 및 RuvC 뉴클레아제 도메인을 함유한다. 뉴클레아제 엽은 또한 프로토스페이서 인접 모티프(PAM)와의 상호작용을 담당하는 카르복실-말단 도메인을 함유한다. 이러한 고-해상도 구조 및 수반하는 기능 분석은 Cas9에 의한 RNA-안내된 DNA 표적화의 분자 메카니즘을 나타내었고, 이에 따라 새로운 다능성 게놈-편집 기

술의 합리적 디자인을 용이하게 한다.

[0727] ▷ Wu 등은 마우스 배아 줄기 세포(mESC)에서 단일 가이드 RNA(sgRNA)가 로딩된 스트렙토코커스 피오게네스 유래의 촉매적 비활성 Cas9(dCas9)의 게놈-전반 결합 부위를 맵핑하였다. 저자는 4 개의 sgRNA 각각이 NGG 프로토스페이서 인접 모티프(PAM) 및 sgRNA 내의 5-뉴클레오티드 시드 영역을 종종 특징으로 하는 수십 내지 수천 개의 게놈 부위에 대하여 표적 dCas9를 시험한 것을 보여주었다. 염색질 비접근성은 일치하는 시드 서열과 다른 부위에 대한 dCas9 결합을 감소시키고, 따라서 70%의 표적외 부위가 유전자와 연관되었다. 상기 저자는 촉매적 활성 Cas9로 트랜스펙션된 mESC 내의 295 개의 dCas9 결합 부위의 표적화된 서열분석에 의해, 백그라운드 수준 초과로 돌연변이된 단지 하나의 부위가 확인된 것을 나타내었다. 상기 저자는 씨드 매치가 결합을 촉발하나, 표적 DNA와의 광범위한 쌍형성이 절단에 필요한 Cas9 결합 및 절단을 위한 2-상태 모델을 제안하였다.

[0728] ▷ Platt 등은 Cre-의존성 Cas9 낙인 마우스를 수립하였다. 저자는 뉴런, 면역 세포, 및 상피 세포에서 가이드 RNA의 아데노-연관 바이러스 (AAV)-, 렌티바이러스-, 또는 입자-매개된 전달을 이용하여 생체 내 및 생체 외 게놈 편집을 증명하였다.

[0729] ▷ Hsu 등 (2014)은 요거트부터 세포의 유전적 스크리닝을 포함하는 게놈 편집까지 CRISPR-Cas9 전력를 일반적으로 논의하는 리뷰 기사이다.

[0730] ▷ Wang 등 (2014)은 게놈-규모 렌티바이러스 단일 가이드 RNA (sgRNA) 라이브러리를 이용하는 양성 및 음성 선택 모두에 적합한 풀화된(pooled) 기능 손실 유전학적 스크리닝 시도에 관한 것이다.

[0731] ▷ Doench 등은 6 개의 내인성 마우스 및 3 개의 내인성 인간 유전자의 패널의 모든 가능한 표적 부위에 걸쳐 타일링하는(tiling) sgRNA의 풀을 만들고, 항체 염색 및 유세포분석기에 의해 표적 유전자의 널 대립형질을 생산하는 그의 능력을 정량적으로 평가하였다. 저자들은 PAM 개선된 활성의 최적화를 보였으며, sgRNA의 디자인을 위한 온라인 도구를 또한 제공하였다.

[0732] ▷ Swiech 등은 AAV-매개된 SpCas9 게놈 편집이 뇌 내 유전자 작용의 역전 유전적 구조를 가능하게 할 수 있음을 증명하였다.

[0733] ▷ Konermann 등 (2015)은 다수의 이펙터 도메인, 예를 들어 전사 활성화제, 예를 들어 작용적 및 후생유전자적 조절제를 링커 없이 또는 링커와 함께 줄기 또는 테트라루프와 같은 가이드 상에서 적절한 위치에 부착시키는 능력을 논의한다.

[0734] ▷ Zetsche 등은 Cas9 효소가 둘로 분열될 수 있고, 이에 따라 활성화를 위한 Cas9 조립이 제어될 수 있음을 증명한다.

[0735] ▷ Chen 등은 마우스들에서 게놈-전반 생체 내 CRISPR-Cas9 스크린은 폐 전이를 조절하는 유전자를 드러냄을 증명함에 의한 멀티플렉스 스크리닝에 관한 것이다.

[0736] ▷ Ran 등 (2015)은 SaCas9 및 게놈을 편집하는 그의 능력에 관한 것으로, 생화학 분석으로부터 추론할 수 없음을 증명한다. Shalem 등 (2015)은, 촉매적으로 비활성인 Cas9 (dCas9) 융합물이 발현을 합성적으로 저해 (CRISPRi) 또는 활성 (CRISPRa)하는데 사용되는 방법을 설명하며, 이는 배열되고 풀화된 스크린을 포함하는, 게놈-규모 스크린을 위해 Cas9를 이용하는 진전, 비활성 게놈 유전자좌를 비활성화하는 낙아웃 시도 및 전사 활성을 조절하는 전략을 보여준다.

[0737] 최종 편집

[0738] ▷ Shalem 등 (2015)은, 촉매적으로 비활성인 Cas9 (dCas9) 융합물이 발현을 합성적으로 저해 (CRISPRi) 또는 활성 (CRISPRa)하는데 사용되는 방법을 설명하며, 이는 배열되고 풀화된 스크린을 포함하는, 게놈-규모 스크린을 위해 Cas9를 이용하는 진전, 비활성 게놈 유전자좌를 비활성화하는 낙아웃 시도 및 전사 활성을 조절하는 전략을 보여준다.

[0739] ▷ Xu 등 (2015)은 CRISPR-기반 스크린에서 단일 가이드 RNA (sgRNA) 효율에 기여하는 DNA 서열 특징을 평가하였다. 저자들은 CRISPR/Cas9 낙아웃 및 절단 부위에서 뉴클레오티드 선호성을 탐구하였다. 저자들은 또한 CRISPRi/a에 대한 서열 선호성이 CRISPR/Cas9 낙아웃에 대한 선호와 실질적으로 상이함을 발견하였다.

[0740] ▷ Parnas 등 (2015)은 박테리아성 지질다당류 (LPS)에 의해 종양 괴사 인자 (Tnf)의 유도를 제어하는 유

전자를 확인하기 위하여 게놈-전반 풀화된 CRISPR-Cas9 라이브러리를 수지상 세포 (DC) 내로 도입하였다. Tlr4 신호화의 알려진 조절자 및 이전에 알려지지 않은 후보물질들이 확인되었으며, LPS에 대한 정규 반응에 대한 구별된 효과를 갖는 세 개의 작용 모듈로 확인되었다.

[0741] ▷ Ramanan 등 (2015)은 감염된 세포에서 바이러스 에피좀성 DNA (cccDNA)의 절단을 증명하였다. HBV 게놈은 공유결합 폐환형 DNA (cccDNA)로도 명명되는 3.2kb의 이중-가닥 에피좀 DNA 종으로서 감염된 간세포의 핵에 존재하고, 그의 복제가 현재 치료법에 의해 억제되지 않는 HBV 생애 주기에서 주요 성분이다. HBV의 고도로 보존된 영역을 특이적으로 표적화하는 sgRNA가 바이러스 복제 및 감손된 cccDNA를 저해함을 보여주었다.

[0742] ▷ Nishimasu 등 (2015)은 5'-TTGAAT-3' PAM 및 5'-TTGGGT-3' PAM을 함유하는 단일 가이드 RNA (sgRNA) 및 그의 이중-가닥 DNA 표적과 함께 복합체화되는 SaCas9의 결정 구조를 보고하였다. SpCas9 와 SaCas9의 구조 비교는 구조적 보존 및 다양성 모두에서 강조표시하였으며, 이는 그들의 구별되는 PAM 특이성 및 오솔로그성 sgRNA 인식을 설명한다.

[0743] 또한, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014)은 연장된 서열을 인식하고 인간 세포에서 높은 효율로 내인성 유전자를 편집할 수 있는 이량체 RNA-가이드 FokI 뉴클레아제에 관한 것이다. 또한, 본원에 참조로서 인용된 PCT 출원 PCT/US14/70057, Attorney Reference 47627.99.2060 및 BI-2013/107, 발명의 제목 "DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS (US 가특허 출원: 62/054,490 (September 24, 2014 에 출원); 62/010,441 (June 10, 2014 에 출원); 및 61/915,118, 61/915,215 및 61/915,148 (각각 December 12, 2013 에 출원) 중 하나 이상 또는 모두의 우선권 주장) ("the Particle Delivery PCT") 로서, sgRNA 및 Cas9 단백질 (및 임의로 HDR 주형)을 포함하는 혼합물을 계면 활성제, 인지질, 생분해성 중합체, 지단백질 및 알코올을 포함하거나 이것으로 본질적으로 구성되는 혼합물과 혼련하는 것을 포함하는 sgRNA-및-Cas9 단백질 함유 입자의 제조 방법; 및 그러한 방법으로부터의 입자에 관한 것을 언급한다.

예를 들어, Cas9 단백질 및 sgRNA는 적합하게, 예를 들어 3:1 내지 1:3 또는 2:1 내지 1:2 또는 1:1의 몰비율로, 적합한 온도, 예를 들어 15-30℃, 예를 들어 20-25℃, 예를 들어 실온에서, 적합한 시간, 예를 들어 15 내지 45분, 예컨대 39분간, 유리하게 멸균된, 뉴클레아제 무함유 완충액, 예를 들어 1×PBS에서 함께 혼합되었다. 별도로, 하기와 같은 또는 하기를 포함하는 입자 성분: 계면활성제, 예를 들어, 양이온성 지질, 예를 들어, 1,2-디올레오일-3-트리메틸암모늄-프로판 (DOTAP); 인지질, 예를 들어, 디미리스토일포스파티딜콜린 (DMPC); 생분해성 중합체, 예컨대 에틸렌-글리콜 중합체 또는 PEG, 및 지단백질, 예컨대 저밀도 지단백질, 예를 들어 콜레스테롤과 같은 또는 이를 포함하는 입자 성분을 알콜, 유리하게 C_{1-6} 알킬 알콜, 예컨대 메탄올, 에탄올, 이소프로판올, 예를 들어 100% 에탄올에 용해시켰다. 2가지 용액을 함께 혼합하여 Cas9-sgRNA 복합체를 함유하는 입자를 형성시켰다. 따라서, sgRNA는 입자에 전체 복합체를 제제화시키기 전에 Cas9 단백질과 예비 복합체화될 수 있다. 제제는 세포 내로 핵산의 전달을 촉진하는 것으로 알려진 상이한 성분의 상이한 몰비로 제조될 수 있다 (예를 들어, 1,2-디올레오일-3-트리메틸암모늄프로판 (DOTAP), 1,2-디테트라데카노일-*sn*-글리세로-3-포스포콜린 (DMPC), 폴리에틸렌 글리콜 (PEG), 및 콜레스테롤). 에를 들어 DOTAP : DMPC : PEG : 콜레스테롤 몰비는 하기와 같을 것이다: DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0; 또는 DOTAP 90, DMPC 0, PEG 10, 콜레스테롤 0; 또는 DOTAP 90, DMPC 0, PEG 5, 콜레스테롤 5. DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0. 따라서 이 출원은 sgRNA, Cas9 단백질 및 입자를 형성하는 성분을 혼합하는 것 뿐 아니라; 이러한 혼합으로부터의 입자로 이해된다. 본 발명의 측면은 예를 들어, 예로서 본 발명에서와 같은 crRNA 및/또는 Cas13b를 포함하는 혼합물 및 입자를 형성하는 성분을 혼합하여, 예를 들어 입자 전달 PCT에서 처럼, 입자를 형성하기 위한, 입자 전달 PCT의 것과 유사한 방법을 사용한 입자 및 그러한 혼합 유래의 입자 (또는 물론, 본 발명에서와 같은 crRNA 및/또는 Cas13b를 포함하는 다른 입자)를 포함할 수 있다.

[0744] 본 발명은 하기 실시예에서 추가로 예시될 것이며, 실시예는 단지 예시의 목적으로 제공된 것이고 어떠한 방식으로 본 발명을 제한하는 것으로 의도되지 않는다.

[0745] **실시예**

[0746] *실시예 1: Cas13b 오르소로그의 동정*

[0747] 하기 표 1에 표시된 Cas13b 단백질은 유리하게 포유동물 세포에서의 발현을 위한 코돈 최적화로부터 유래된다. 하기 표 1에 표시된 단백질은 유리하게 포유동물 세포에서의 발현을 위해 코돈 최적화된 것이다.

mCherry 및 임의로 NLS 또는 NES와 각각의 Cas13b 오르소로그의 융합 구축물을 만들어서 포유동물 발현 벡터에 클로닝한다. 다양한 Cas13b 오르소로그를 HEK293T 세포에 트랜스펙션시키고 세포 국재화는 mCherry 발현을 기반으로 평가된다. NES 또는 NLS 융합없거나, 또는 C-말단 및 N-말단 NLS에 융합되고, C-말단 및 N-말단 NES에 융합된 상이한 Cas13b 오르소로그의 국재화가 결정된다. NES 융합은 그 결과로서 효율적으로 Cas13b 단백질의 세포질 국재화를 일으킨다. NLS 융합은 효율적으로 Cas13b 단백질의 핵 국재화를 일으킨다. 가변적으로, 또한 핵소체 국재화가 NLS 융합에서 관찰될 수 있다.

표 1A

Sinomicrobium oceani	WP_072319476.1	MESTTTLGLHLKYQHDLFEDKHYFGGGVNLAVQNIESIFQAFA ERYGIQNPLRKNGVPAINNIFHDNISISNYKEYLKFLKQYLPVVG FLEKSNEINIFEFREDFEILINAIYKLRHFYTHYYHSPIKLEDRFYT CLNELFVAVAIQVKKHKMKSDKTRQLLNKNLHQLLQQLIEQKR EKLKDKKAEGEKVSLDTKSIENAVLNDAFVHLLDKDENIRLNY SSRLSEDIITKNGITLSISGLLFLLSLFLQRKEAEDLRSRIEGFKGK GNELRFMATHWVFSYLNVKRIKHRLNTDFQKETLLIQIADELSK VPDEVYKTLDHENRSKFLEDINEYIREGNEDASLNESTVVHGVI RKRYENKFHYLVLRYLDEFVDFPSLRFQVHLGNYIHDRRDKVI DGTNFITNRVIKEPIKVFGKLSHVSKLKSDYMESLSREHKNGWD VFPNPSYNFVGHNIPIFINLRSASSKGKELYRDLMKIKSEKKKKS REEGIPMERRDGKPTKIEISNQIDRNIKDNNFKDIYPGEPLAMLS LNELPALLFELLRRPSITPQDIEDRMVEKLYERFQIIRDYKPGDG LSTSKISKKLRKADNSTRLDGKKLLRAIQTETRNAREKLHTLEE NKALQKNRKRRTVYTTREQGREASWLAQDLKRFMPIASRKEW RGYHHSQLQQILAFYDQNPKQPLELLEQFWDLKEDTYVWNSWI HKSLSQHNGFVPMYEGYLKGRLGYYKKLESDIIGFLEEHKVLK RYYTQQHLNVIFRERLYFIKTETKQKLELLARPLVFPRGIFDDKP TFVQDKKVVDHPELFADWYVYSYKDDHSFQEFYHYKRDYNEI FETELSWDIDFKDNKRQLNPSEQMDLFRMKWDLKIKKIKIQDIF LKIVAEDIYLKIFGHKIPLSLSDFYISRQERLTLDEQAVAQSMRLP GDTSENQIKESNLWQTTVPYEKEQIREPKIKLKDIGKFKYFLQQ QKVLNLLKYDPQHVWTKAELEEELYIGKHSYEVVRREMLLQK CHQLEKHILEQFRFDGSNHPRELEQGNHPNFKMYIVNGILTKRG ELEIEAENWWLELGNSKNSLDKVEVELLTMKTIPEQKAFLLILIR NKFAHNQLPADNYFHYASNLMNLKKSDTYSLFWFTVADTIVQ EFMSL

[0748]

Prevotella intermedia	12	MEDDKKTTDSIRYELKDKHFWAAFLNLARHNVYITVNHINKIL EEDEINRDGYENTLENSWNEIKDINKKDRLSKLIIKHFPFLEATT YRQNPTDTTKQKEEKQAEAQSLESLKKSFFVFIYKLRDLRNHYS HYKHSKSLERPKFEEDLQNKMYNIFDVSIQFVKEDYKHNTDINP KKDFKHLDRKRKGKFHYSFADNEGNITESGLLFFVSLFLEKKDA IWVQKKLEGFKCSNKSYQKMTNEVFCRSRMLLPKLRLESTQTQ DWILLDMLNELIRCPKSLYERLQGVNRKKFYVSFDPADEDYDA EQEPFKNTLVRHQDRFPYFALRYFDYNEVFANLRFQIDLGTYHF SIYKKLIGGQKEDRHLTHKLYGFERIQEFDKQNRPDEWKAIVKD SDTFKKKEEKEEEKPYISETTPHYHLENKKIGIAFKNHNIWPSTQ TELTNNKRKKYNLGTSIKAEAFLSVHELLPMMFYYLLLKTENT KNDNKVGGKKETKKQGKHKIEAIIESKIKDIYALYDAFANGEIN SEDELKEYLKGKDIKIVHLPKQMIAILKNEHKDMAEKAEAKQE KMKLATENRLKTLDKQLKGKIQNGKRYNSAPKSGEIASWLVN DMMRFQPVQKDENGESLNNSKANSTEYQLLQRTLAFFGSEHER LAPYFKQTKLIESSNPHPFLNDTEWEKCSNILSFYRSYLKARKNF LESLKPEDWEKNQYFLMLKEPKTNRETLVQGWKNGFNLPRGFF TEPIRKWFMEHWKSIKVDDLKRVGLVAKVTPLFFSEKYKDSVQ PFYNYPFNVGDVNKPKEEDFLHREERIELWDKKKDKFKGYKA KKKFKEMTDKEKEEHRSYLEFQSWNKFERELRLVRNQDIVTWL LCTELIDKLKIDELNIKELKKLRLKDINTDTAKKEKNNILNRVMP MELPVTVYKVNKGGYIIKNKPLHTIYIKEAETKLLKQGNFKALV KDRRLNGLFSFVKTPSEAESESNPISKLRVEYELGKYQNARLDII EDMLALEKKLIDKYNSLDTDNFHNMLTGWLELKGEAKKARFQ NDVKLLTAVRNAFSHNQYPMYDENLFGNIERFSLSSSNIIESKGL DIAAKLKEEVSKAAKKIQNEEDNKKEKET
Porphyromonas gingivalis	19	MTEQNEKPYNGTYYTLEDKHFWAAFLNLARHNAYITLAHIDR QLAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFS FLEGAAYGKKLFESQSSGNKSSKKKELSKKEKEELQANALSLD NLKSILFDFLQKLKDFRNYYSHYRHPESSELPLFDGNMLQRLYN VFDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDKYGNND NPFFKHHFVDREGTVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTEAYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTL VRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGE QPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGD KPYITQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGATRTGRSK YAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKVQ GRIKRVIEDVYAVYDAFARDEINTRDELDACLADKGIRRGHLPR QMIAILSQEHKDMEEKVRKKLQEMIADTDHRLDMLDRQTDRKI RIGRKNAGLPKSGVVADWLVRDMMRFQPVAKDTSGKPLNNSK ANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLH ETRWESHTNILSFYRSYLEARKAFLQSIGRSDRVENHRFLLLKEP KTDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEMGYDEVGSYKE VGFMAKAVPLYFERASKDRVQPFYDYPFNVGNSLKPKKGRFLS KEKRAEEWESGKERFRLAKLKKEILEAKEHPYHDFKSWQKFER

[0749]

		ELRLVKNQDIITWMMCRDLMEENKVEGLDTGTLYLKDIRTDV QEQGSLNVLNRVKPMRLPVVVYRADSRGHVHKEQAPLATVYI EERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAMEQYPISK LRVEYELAKYQTARVCAFEQTLELEESLLTRYPHLPDKNFRKM LESWSDPLLDKWPDLHGNVRLLIAVRNAFSHNQYPMYDETLFS SIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAKEMVERIIQA
A2033_10205 [Bacteroidetes bacterium GWA2_31_9]	OFX18020.1	MENQTQKGKGIYYYYTKNEDKHYFGSFLNLANNNIEQIIEEFRI RLSLKDEKNIKEIINNYFTDKKSYTDWERGINILKEYLPVIDYLD LAITDKEFEKIDLKQKETAKRKYFRTNFSLLIDTIIDLRNFYTHYF HKPISINPDVAKFLDKNLLNVCLDIKKQKMKTDKTKQALKDGL DKELKKLIELKKAELKEKKIKTWNITENVEGAVYNDAFNHMVY KNNAGVTILKDYHKSILPDDKIDSELKLNFSISGLVFLLSMFLSK KEIEQFKSNLEGFKGKVIGENGEYEISKFNNSLKYMATHWIFSY LTFKGLKQRVKNTFDKETLLMQMIDELNKVPHEVYQTLSKEQQ NEFLEDINEYVQDNEENKKSMENSIVVHPVIRKRYDDKFNYFAI RFLDEFANFPTLKFFVTAGNFVHDKREKQIQGSMLTSDRMIKEK INVFGKLTEIAKYKSDYFSNENTLETSEWELFPNPSYLLIQNNIPV HIDLIHNTEEAKQCQIAIDRIKCTTNPAKKRNTRKSKEEIIKIIYQ KNKNIKYGDPTALLSSNELPALIYELLVNKKSGKELENIIVEKIV NQYKTIAGFEKGQNLSNSLITKKLKKSEPNEDKINAEKIILAINRE LEITENKLNIIKNNRAEFRTGAKRKHIFYSKELGQEATWIAYDLK RFMPEASRKEWKGFHHSELQKFLAFYDRNKNDAKALLNMFW NFDNDQLIGNDLNSAFREFHFDKFYEKYLIKRDEILEGFKSFISN FKDEPKLLKKGIKDIYRVFDKRYYIIKSTNAQKEQLLSKPICLPR GIFDNKPTYIEGVKVESNSALFADWYQYTYSDKHEFQSFYDMP RDYKEQFEKFELNNIKSIQNKKNLNKSDKFIYFRYKQDLKIKQIK SQDLFIKLMVDELFNVVFKNNIELNLKKLYQTSDERFKNQLIAD VQKNREKGDTSDNKMNENFIWNMTIPLSLCNGQIEEPKVKLKD IGKFRKLETDDKVIQLLEYDKSKVWKKLEIEDELENMPNSYERI RREKLLKGIQEFEHFLLEKEKFDGINHPKHFEQDLNPNFKTYVIN GVLRKNSKLNYTEIDKLLDLEHISIKDIETSAKEIHLAYFLIHVRN KFGHNQLPKLEAFELMKKYYKKNNEETYAEYFHKVSSQIVNEF KNSLEKHS
SAMN05421542_0666 [Chryseobacterium jejuense]	SDI27289.1	MEKTQTGLGIYYDHTKLQDKYFFGGFFNLAQNNIDNVIKAFIIK FFPERKDKDINIAQFLDICFKDNDADSDFQKKNKFLRIHFPVIGF LTSDNDKAGFKKKFALLLKTISELRNFYTHYYHKSIEFPSELFEL LDDIFVKTTSEIKKLKKKDDKTQQLLNKNLSEEYDIRYQQQIER LKELKAQGKRVSLTDETAIRNGVFNAAFNHLIYRDGENVKPSR LYQSSYSEPDPAENGISLSQNSILFLLSMFLERKETEDLKSRVKG FKAKIIKQGEEQISGLKFMATHWVFSYLCFKGIKQKLSTEFHEET LLIQIIDELSKVPDEVYSAFDSKTKEKFLEDINEYMKEGNADLSL EDSKVIHPVIRKRYENKFNYFAIRFLDEYLSSTSLKFQVHVGNY VHDRRVKHINGTGFQTERIVKDRIKVFGRLSNISNLKADYIKEQ LELPNDSNGWEIFPNPSYIFIDNNVPIHVLADEATKKGIELFKDK RRKEQPEELQKRKGKISKYNIVSMIYKEAKGKDKLRIDEPLALL

[0750]

		SLNEIPALLYQILEKGATPKDIELIIKNKLTERFEKIKNYDPETPAP ASQISKRLRNNTTAKGQEALNAEKLSLLIEREIENTETKLSSIEEK RLKAKKEQRRNTPQRSIFSNSDLGRIAAWLADDIKRFMPAEQRK NWKGYQHSQLQQSLAYFEKRPQEAFLLLKEGWDTSDGSSYWN NWVMNSFLENNHFEKFYKNYLMKRVKYFSELAGNIKQHTHNT KFLRKFIKQQMPADLFPKRHYILKDLETEKNKVLSKPLVFSRGL FDNNPTFIKGVKVTENPELFAEWYSYGYKTEHVFQHFYGWERD YNELLDSELQKGNSFAKNSIYYNRESQLDLIKLKQDLKIKKIKIQ DLFLKRIAEKLFENVFNYPTTLSLDEFYLTQEERAEKERIALAQS LREEGDNSPNIIKDDFIWSKTIAFRSKQIYEPAIKLKDIGKFNRFV LDDEESKASKLLSYDKNKIWNKEQLERELSIGENSYEVIRREKL FKEIQNLELQILSNWSWDGINHPREFEMEDQKNTRHPNFKMYL VNGILRKNINLYKEDEDFWLESLKENDFKTLPSEVLETKSEMVQ LLFLVILIRNQFAHNQLPEIQFYNFIRKNYPEIQNNTVAELYLNLI KLAVQKLKDNS
SAMN054 44360_113 66 [Chryseoba cterium carnipullor um]	SHM52 812.1	MNTRVTGMGVSYDHTKKEDKHFFGGFLNLAQDNITAVIKAFCI KFDKNPMSSVQFAESCFTDKDSDTDFQNKVRYVRTHLPVIGYL NYGGDRNTFRQKLSTLLKAVDSLRNFYTHYYHSPLALSTELFEL LDTVFASVAVEVKQHKMKDDKTRQLLSKSLAEELDIRYKQQLE RLKELKEQGKNIDLRDEAGIRNGVLNAAFNHLIYKEGEIAKPTL SYSSFYYGADSAENGITISQSGLLFLLSMFLGKKEIEDLKSRIRGF KAKIVRDGEENISGLKFMATHWIFSYLSFKGMKQRLSTDFHEET LLIQIIDELSKVPDEVYHDFDTATREKFVEDINEYIREGNEDFSLG DSTIIHPVIRKRYENKFNYFAVRFLDEFIKFPSLRFQVHLGNFVH DRRIKDIHGTGFQTERVVKDRIKVFGKLSEISSLKTEYIEKELDL DSDTGWEIFPNPSYVFIDNNIPIYISTNKTFKNGSSEFIKLRRKEKP EEMKMRGEDKKEKRDIASMIGNAGSLNSKTPLAMLSLNEMPAL LYEILVKKTTPEEIELIIKEKLDSHFENIKNYDPEKPLPASQISKRL RNNTTDKGKKVINPEKLIHLINKEIDATEAKFALLAKNRKELKE KFRGKPLRQTIFSNMELGREATWLADDIKRFMPDILRKNWKGY QHNQLQQSLAFFNSRPKEAFTILQDGWDFADGSSFWNGWIINSF VKNRSFEYFYEAYFEGRKEYFSSLAENIKQHTSNHRNLRRFIDQ QMPKGLFENRHYLLENLETEKNKILSKPLVFPRGLFDTKPTFIKG IKVDEQPELFAEWYQYGYSTEHVFQNFYGWERDYNDLLESELE KDNDFSKNSIHYSRTSQLELIKLKQDLKIKKIKIQDLFLKLIAGHI FENIFKYPASFSLDELYLTQEERLNKEQEALIQSQRKEGDHSDNII KDNFIGSKTVTYESKQISEPNVKLKDIGKFNRFLLDDKVKTLLS YNEDKVWNKNDLDLELSIGENSYEVIRREKLFKKIQNFELQTLT DWPWNGTDHPEEFGTTDNKGVNHPNFKMYVVNGILRKHTDW FKEGEDNWLENLNETHFKNLSFQELETKSKSIQTAFLIIMIRNQF AHNQLPAVQFFEFIQKKYPEIQGSTTSELYLNFINLAVVELLELL EK
SAMN054 21786_101 1119	SIS704 81.1	METQILGNGISYDHTKTEDKHFFGGFLNTAQNNIDLLIKAYISKF ESSPRKLNSVQFPDVCFKKNDSDADFQHKLQFIRKHLPVIQYLK YGGNREVLKEKFRLLLQAVDSLRNFYTHFYHKPIQLPNELLTLL

[0751]

[Chryseobacterium ureilyticum]		DTIFGEIGNEVRQNKMKDDKTRHLLKKNLSEELDFRYQEQLER LRKLKSEGKKVDLRDTEAIRNGVLNAAFNHLIFKDAEDFKPTVS YSSYYYDSDTAENGISISQSGLLFLLSMFLGRREMEDLKSRVRG FKARIIKHEEQHVSGLKFMATHWVFSEFCFKGIKTRLNADYHEE TLLIQLIDELSKVPDELYRSFDVATRERFIEDINEYIRDGKEDKSL IESKIVHPVIRKRYESKFNYFAIRFLDEFVNFPTLRFQVHAGNYV HDRRIKSIEGTGFKTERLVKDRIKVFGKLSTISSLKAEYLAKAVN ITDDTGWELLPHPSYVFIDNNIPIHLTVDPSFKNGVKEYQEKRKL QKPEEMKNRQGGDKMHKPAISSKIGKSKDINPESPVALLSMNEI PALLYEILVKKASPEEVEAKIRQKLTAVFERIRDYDPKVPLPASQ VSKRLRNNTDTLSYNKEKLVELANKEVEQTERKLALITKNRRE CREKVKGKFKRQKVFKNAELGTEATWLANDIKRFMPEEQKKN WKGYQHSQLQQSLAFFESRPGEARSLLQAGWDFSDGSSFWNG WVMNSFARDNTFDGFYESYLNGRMKYFLRLADNIAQQSSTNK LISNFIKQQMPKGLFDRRLYMLEDLATEKNKILSKPLIFPRGIFD DKPTFKKGVQVSEEPEAFADWYSYGYDVKHKFQEFYAWDRD YEELLREELEKDTAFTKNSIHYSRESQIELLAKKQDLKVKKVRI QDLYLKLMAEFLFENVFGHELALPLDQFYLTQEERLKQEQEAIV QSQRPKGDDSPNIVKENFIWSKTIPFKSGRVFEPNVKLKDIGKFR NLLTDEKVDILLSYNNTEIGKQVIENELIIGAGSYEFIRREQLFKEI QQMKRLSLRSVRGMGVPIRLNLK
Reichenbachiella agariperforans	WP_07312444 1.1	MKTNPLIASSGEKPNYKKFNTESDKSFKKIFQNKGSIAPIAEKAC KNFEIKSKSPVNRDGRLHYFSVGHAFKNIDSKNVFRYELDESQM DMKPTQFLALQKEFFDFQGALNGLLKHIRNVNSHYVHTFEKLEI QSINQKLITFLIEAFELAVIHSYLNEEELSYEAYKDDPQSGQKLV QFLCDKFYPNKEHEVEERKTILAKNKRQALEHLLFIEVTSDIDW KLFEKHKVFTISNGKYLSFHACLFLLSLFLYKSEANQLISKIKGF KRNDDNQYRSKRQIFTFFSKKFTSQDVNSEEQHLVKFRDVIQYL NHYPSAWNKHLELKSGYPQMTDKLMRYIVEAEIYRSFPDQTDN HRFLLFAIREFFGQSCLDTWTGNTPINFSNQEQKGFSYEINTSAEI KDIETKLKALVLKGPLNFKEKKEQNRLEKDLRREKKEQPTNRV KEKLLTRIQHNMLYVSYGRNQDRFMDFAARFLAETDYFGKDA KFKMYQFYTSDEQRDHLKEQKKELPKKEFEKLKYHQSKLVDY FTYAEQQARYPDWDTPFVVENNAIQIKVTLFNGAKKIVSVQRN LMLYLLEDALYSEKRENAGKGLISGYFVHHQKELKDQLDILEK ETEISREQKREFKKLLPKRLLHRYSPAQINDTTEWNPMEVILEEA KAQEQRYQLLLEKAILHQTEEDFLKRNKGKQFKLRFVRKAWH LMYLKELYMNKVAEHGHHKSFHITKEEFNDFCRWMFAFDEVP KYKEYLCDYFSQKGFFNNAEFKDLIESSTSLNDLYEKTKQRFEG WSKDLTKQSDENKYLLANYESMLKDDMLYVNISHFISYLESKG KINRNAHGHIAYKALNNVPHLIEEYYYKDRLAPEEYKSHGKLY NKLKTVKLEDALLYEMAMHYLSLEPALVPKVKTKVKDILSSNI AFDIKDAAGHHLYHLLIPFHKIDSFVALINHQSQQEKDPDKTSFL AKIQPYLEKVKNSKDLKAVYHYYKDTPHTLRYEDLNMIHSHIV SQSVQFTKVALKLEEYFIAKKSITLQIARQISYSEIADLSNYFTDE VRNTAFHFDVPETAYSMILQGIESEFLDREIKPQKPKSLSELSTQ

[0752]

		QVSVCTAFLETLHNNLFDRKDDKKERLSKARERYFEQIN

[0753]

[0754]

표 1B		
Bergeyella zoohelcum	1	MENKTSLGNNIYYNPFKPQDKSYFAGYFNAAMENTDSVFRELG KRLKGKEYTSENFFDAIFKENISLVEYERYVKLLSDYFPMARLL DKKEVPIKERKENFKKNFKGIIKAVRDLRNFYTHKEHGEVEITD EIFGVLDEMLKSTVLTVKKKKVKTDKTKEILKKSIEKQLDILCQ KKLEYLRDTARKIEEKRRNQRERGEKELVAPFKYSDKRDDLIA AIYNDAFDVYIDKKKDSLKESSKAKYNTKSDPQQEEGDLKIPIS KNGVVFLLSLFLTKQEIHAFKSKIAGFKATVIDEATVSEATVSHG KNSICFMATHEIFSHLAYKKLKRKVRTAEINYGEAENAEQLSVY AKETLMMQMLDELSKVPDVVYQNLSEDVQKTFIEDWNEYLKE NNGDVGTMEEEQVIHPVIRKRYEDKFNYFAIRFLDEFAQFPTLR FQVHLGNYLHDSRPKENLISDRRIKEKITVFGRLSELEHKKALFI KNTETNEDREHYWEIFPNPNYDFPKENISVNDKDFPIAGSILDRE KQPVAGKIGIKVKLLNQQYVSEVDKAVKAHQLKQRKASKPSIQ NIIEEIVPINESNPKEAIVFGGQPTAYLSMNDIHSILYEFFDKWEK KKEKLEKKGEKELRKEIGKELEKKIVGKIQAQIQQIIDKDTNAKI LKPYQDGNSTAIDKEKLIKDLKQEQNILQKLKDEQTVREKEYN DFIAYQDKNREINKVRDRNHKQYLKDNLKRKYPEAPARKEVL YYREKGKVAVWLANDIKRFMPTDFKNEWKGEQHSLLQKSLAY YEQCKEELKNLLPEKVFQHLPFKLGGYFQQKYLYQFYTCYLDK RLEYISGLVQQAENFKSENKVFKKVENECFKFLKKQNYTHKEL DARVQSILGYPIFLERGFMDEKPTIIKGKTFKGNEALFADWFRY YKEYQNFQTFYDTENYPLVELEKKQADRKRKTKIYQQKKNDV FTLLMAKHIFKSVFKQDSIDQFSLEDLYQSREERLGNQERARQT GERNTNYIWNKTVDLKLCDGKITVENVKLKNVGDFIKYEYDQR VQAFLKYEENIEWQAFLIKESKEEENYPYVVEREIEQYEKVRRE ELLKEVHLIEEYILEKVKDKEILKKGDNQNFKYYILNGLLKQLK NEDVESYKVFNLNTEPEDVNINQLKQEATDLEQKAFVLTYIRN KFAHNQLPKKEFWDYCQEKYGKIEKEKTYAEYFAEVFKKEKE ALIK
Prevotella intermedia	2	MEDDKKTTDSIRYELKDKHFWAAFLNLARHNVYITVNHINKIL EEGEINRDGYETTLKNTWNEIKDINKKDRLSKLIIKHFPFLEAAT YRLNPTDTTKQKEEKQAEAQSLESLRKSFFVFIYKLRDLRNHYS HYKHSKSLERPKFEEGLLEKMYNIFNASIRLVKEDYQYNKDINP DEDFKHLDRTEEEFNYYFTKDNEGNITESGLLFFVSLFLEKKDAI WMQQKLRGFKDNRENKKKMTNEVFCRSRMLLPKLRLQSTQTQ DWILLDMLNELIRCPKSLYERLREEDREKFRVPIEIADEDYDAEQ EPFKNTLVRHQDRFPYFALRYFDYNEIFTNLRFQIDLGTYHFSIY KKQIGDYKESHHLTHKLYGFERIQEFTKQNRPDEWRKFVKTFN SFETSKEPYIPETTPHYHLENQKIGIRFRNDNDKIWPSLKTNSEK NEKSKYKLDKSFQAEAFLSVHELLPMMFYYLLLKTENTDNDNE IETKKKENKNDKQEKHKIEEIIENKITEIYALYDTFANGEIKSIDE

		LEEYCKGKDIEIGHLPKQMIAILKDEHKVMATEAERKQEEMLV DVQKSLESLDNQINEEIENVERKNSSLKSGKIASWLVNDMMRF QPVQKDNEGKPLNNSKANSTEYQLLQRTLAFFGSEHERLAPYF KQTKLIESSNPHPFLKDTEWEKCNNILSFYRSYLEAKKNFLESLK PEDWEKNQYFLKLKEPKTKPKTLVQGWKNGFNLPRGIFTEPIRK WFMKHRENITVAELKRVGLVAKVIPLFFSEEYKDSVQPFYNYH FNVGNINKPDEKNFLNCEERRELLRKKKDEFKKMTDKEKEENP SYLEFKSWNKFERELRLVRNQDIVTWLLCMELFNKKKIKELNV EKIYLKNINTNTTKKEKNTEEKNGEEKNIKEKNNILNRIMPMRL PIKVYGRENFSKNKKKKIRRNTFFTVYIEEKGTKLLKQGNFKAL ERDRRLGGLFSFVKTPSKAESKSNTISKLRVEYELGEYQKARIEII KDMLALEKTLIDKYNSLDTDNFNKMLTDWLELKGEPDKASFQ NDVDLLIAVRNAFSHNQYPMRNRIAFANINPFSLSSANTSEEKG LGIANQLKDKTHKTIEKIIEIEKPIETKE
Prevotella buccae	3	MQKQDKLFVDRKKNAIFAFPKYITIMENKEKPEPIYYELTDKHF WAAFLNLARHNVYTTINHINRRLEIAELKDDGYMMGIKGSWNE QAKKLDKKVRLRDLIMKHFPFLEAAAYEMTNSKSPNNKEQRE KEQSEALSLNNLKNVLFIFLEKLQVLRNYYSHYKYSEESPKPIFE TSLLKNMYKVFDANVRLVKRDYMHHENIDMQRDFTHLNRKK QVGRTKNIIDSPNFHYHFADKEGNMTIAGLLFFVSLFLDKKDAI WMQKKLKGFKDGRNLREQMTNEVFCRSRISLPKLKLENVQTK DWMQLDMLNELVRCPKSLYERLREKDRESFKVPFDIFSDDYNA EEEPFKNTLVRHQDRFPYFVLRYFDLNEIFEQLRFQIDLGTYHFS IYNKRIGDEDEVRHLTHHLYGFARIQDFAPQNQPEEWRKLVKD LDHFETSQEPYISKTAPHYHLENEKIGIKFCSAHNNLFPSLQTDK TCNGRSKFNLGTQFTAEAFLSVHELLPMMFYYLLLTKDYSRKE SADKVEGIIRKEISNIYAIYDAFANNEINSIADLTRRLQNTNILQG HLPKQMISILKGRQKDMGKEAERKIGEMIDDTQRRLDLLCKQT NQKIRIGKRNAGLLKSGKIADWLVNDMMRFQPVQKDQNNIPIN NSKANSTEYRMLQRALALFGSENFRLKAYFNQMNLVGNDNPH PFLAETQWEHQTNILSFYRNYLEARKKYLKGLKPQNWKQYQH FLILKVQKTNRNTLVTGWKNSFNLPRGIFTQPIREWFEKHNNSK RIYDQILSFDRVGFVAKAIPLYFAEEYKDNVQPFYDYPFNIGNRL KPKKRQFLDKKERVELWQKNKELFKNYPSEKKKTDLAYLDFLS WKKFERELRLIKNQDIVTWLMFKELFNMATVEGLKIGEIHLRDI DTNTANEESNNILNRIMPMKLPVKTYETDNKGNILKERPLATFY IEETETKVLKQGNFKALVKDRRLNGLFSFAETTDLNLEEHPISKL SVDLELIKYQTTRISIFEMTLGLEKKLIDKYSTLPTDSFRNMLER WLQCKANRPELKNYVNSLIAVRNAFSHNQYPMYDATLFAEVK KFTLFPSVDTKKIELNIAPQLLEIVGKAIKEIEKSENKN
Porphyromonas gingivalis	4	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFAVFFKPDDFVLA KNRKEQLISVADGKECLTVSGFAFFICLFLDREQASGMLSRIRGF

[0755]

		KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLDEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPQSMGFISVHDLRKLLLMELLCEGSFSRMQSDFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMDQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLRKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRRQFRAIV AELRLLDPSSGHPFLSATMETAHRYTEGFYKCYLEKKREWLAK IFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKVMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVRDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDEENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL
Bacteroides pyogenes	5	MESIKNSQKSTGKTLQKDPPYFGLYLNMALLNVRKVENHIRKW LGDVALLPEKSGFHSLLTTDNLSSAKWTRFYYKSRKFLPFLEMF DSDKKSYENRRETAECLDTIDRQKISSLLKEVYGKLQDIRNAFS HYHIDDQSVKHTALIISSEMHRFIENAYSFALQKTRARFTGVFVE TDFLQAEEKGDNKKFFAIGGNEGIKLKDNALIFLICLFLDREEAF KFLSRATGFKSTKEKGFLAVRETFCALCCRQPHERLLSVNPREA LLMDMLNELNRCPDILFEMLDEKDQKSFLPLLGEEEQAHILENS LNDELCEAIDDPFEMIASLSKRVRYKNRFPYLMLRYIEEKNLLPF IRFRIDLGCLELASYPKKMGEENNYERSVTDHAMAFGRLTDFH NEDAVLQQITKGITDEVRFSLYAPRYAIYNNKIGFVRTSGSDKIS FPTLKKKGGEGHCVAYTLQNTKSFGFISIYDLRKILLLSFLDKDK AKNIVSGLLEQCEKHWKDLSENLFDAIRTELQKEFPVPLIRYTLP RSKGGKLVSSKLADKQEKYESEFERRKEKLTEILSEKDFDLSQIP RRMIDEWLNVLPTSREKKLKGYVETLKLDCRERLRVFEKREKG EHPLPPRIGEMATDLAKDIIRMVIDQGVKQRITSAYYSEIQRCLA QYAGDDNRRHLDSIIRELRLKDTKNGHPFLGKVLRPGLGHTEK LYQRYFEEKKEWLEATFYPAASPKRVPRFVNPPTGKQKELPLII RNLMKERPEWRDWKQRKNSHPIDLPSQLFENEICRLLKDKIGKE PSGKLKWNEMFKLYWDKEFPNGMQRFYRCKRRVEVFDKVVE YEYSEEGGNYKKYYEALIDEVVRQKISSSKEKSKLQVEDLTLSV RRVFKRAINEKEYQLRLLCEDDRLLFMAVRDLYDWKEAQLDL DKIDNMLGEPVSVSQVIQLEGGQPDAVIKAECKLKDVSKLMRY CYDGRVKGLMPYFANHEATQEQVEMELRHYEDHRRRVFNWV FALEKSVLKNEKLRRFYEESQGGCEHRRCIDALRKASLVSEEEY EFLVHIRNKSAHNQFPDLEIGKLPPNVTSGFCECIWSKYKAIICRI

[0756]

		IPFIDPERRFFGKLLEQK
Alistipes sp. ZOR0009	6	MSNEIGAFREHQFAYAPGNEKQEEATFATYFNLALSNVEGMMF GEVESNPDKIEKSLDTLPPAILRQIASFIWLSKEDHPDKAYSTEE VKVIVTDLVRRLCFYRNYFSHCFYLDTQYFYSDELVDTTAIGEK LPYNFHHFITNRLFRYSLPEITLFRWNEGERKYEILRDGLIFFCCL FLKRGQAERFLNELRFFKRTDEEGRIKRTIFTKYCTRESHKHIGIE EQDFLIFQDIIGDLNRVPKVCDGVVDLSKENERYIKNRETSNESD ENKARYRLLIREKDKFPYYLMRYIVDFGVLPCITFKQNDYSTKE GRGQFHYQDAAVAQEERCYNFVVRNGNVYYSYMPQAQNVVR ISELQGTISVEELRNMVYASINGKDVNKSVEQYLYHLHLLYEKI LTISGQTIKEGRVDVEDYRPLLDKLLLRPASNGEELRRELRKLLP KRVCDLLSNRFDCSEGVSAVEKRLKAILLRHEQLLLSQNPALHI DKIKSVIDYLYLFFSDDEKFRQQPTEKAHRGLKDEEFQMYHYL VGDYDSHPLALWKELEASGRLKPEMRKLTSATSLHGLYMLCL KGTVEWCRKQLMSIGKGTAKVEAIADRVGLKLYDKLKEYTPE QLEREVKLVVMHGYAAAATPKPKAQAAIPSKLTELRFYSFLGK REMSFAAFIRQDKKAQKLWLRNFYTVENIKTLQKRQAAADAA CKKLYNLVGEVERVHTNDKVLVLVAQRYRERLLNVGSKCAVT LDNPERQQKLADVYEVQNAWLSIRFDDLDFTLTHVNLSNLRKA YNLIPRKHILAFKEYLDNRVKQKLCEECRNVRRKEDLCTCCSPR YSNLTSWLKENHSESSIEREAATMMLLDVERKLLSFLLDERRKA IIEYGKFIPFSALVKECRLADAGLCGIRNDVLHDNVISYADAIGK LSAYFPKEASEAVEYIRRTKEVREQRREELMANSSQ
Prevotella sp. MA2016	7a	MSKECKKQRQEKKRRLQKANFSISLTGKHVFGAYFNMARTNF VKTINYILPIAGVRGNYSENQINKMLHALFLIQAGRNEELTTEQK QWEKKLRLNPEQQTKFQKLLFKHFPVLGPMMADVADHKAYL NKKKSTVQTEDETFAMLKGVSLADCLDIICLMADTLTECRNFY THKDPYNKPSQLADQYLHQEMIAKKLDKVVVASRRILKDREGL SVNEVEFLTGIDHLHQEVLKDEFGNAKVKDGKVMKTFVEYDD FYFKISGKRLVNGYTVTTKDDKPVNVNTMLPALSDFGLLYFCV LFLSKPYAKLFIDEVRLFEYSPFDDKENMIMSEMLSIYRIRTPRL HKIDSHDSKATLAMDIFGELRRCPMELYNLLDKNAGQPFFHDE VKHPNSHTPDVSKRLRYDDRFPTLALRYIDETELFKRIRFQLQL GSFRYKFYDKENCIDGRVRVRRIQKEINGYGRMQEVADKRMD KWGDLIQKREERSVKLEHEELYINLDQFLEDTADSTPYVTDRRP AYNIHANRIGLYWEDSQNPKQYKVFDENGMYIPELVVTEDKKA PIKMPAPRCALSVYDLPAMLFYEYLREQQDNEFPSAEQVIIEYE DDYRKFFKAVAEGKLKPFKRPKEFRDFLKKEYPKLRMADIPKK LQLFLCSHGLCYNNKPETVYERLDRLTLQHLEERELHIQNRLEH YQKDRDMIGNKDNQYGKKSFSDVRHGALARYLAQSMMEWQP TKLKDKEKGHDKLTGLNYNVLTAYLATYGHPQVPEEGFTPRTL EQVLINAHLIGGSNPHPFINKVLALGNRNIEELYLHYLEEELKHI RSRIQSLSSNPSDKALSALPFIHHDRMRYHERTSEEMMALAARY TTIQLPDGLFTPYILEILQKHYTENSDLQNALSQDVPVKLNPTCN AAYLITLFYQTVLKDNAQPFYLSDKTYTRNKDGEKAESFSFKR

[0757]

		AYELFSVLNNNKKDTFPFEMIPLFLTSDEIQERLSAKLLDGDGNP VPEVGEKGKPATDSQGNTIWKRRIYSEVDDYAEKLTDRDMKIS FKGEWEKLPRWKQDKIIKRRDETRRQMRDELLQRMPRYIRDIK DNERTLRRYKTQDMVLFLLAEKMFTNIISEQSSEFNWKQMRLS KVCNEAFLRQTLTFRVPVTVGETTIYVEQENMSLKNYGEFYRFL TDDRLMSLLNNIVETLKPNENGDLVIRHTDLMSELAAYDQYRS TIFMLIQSIENLIITNNAVLDDPDADGFWVREDLPKRNNFASLLE LINQLNNVELTDDERKLLVAIRNAFSHNSYNIDFSLIKDVKHLPE VAKGILQHLQSMLGVEITK
Prevotella sp. MA2016	7b	MSKECKKQRQEKKRRLQKANFSISLTGKHVFGAYFNMARTNF VKTINYILPIAGVRGNYSENQINKMLHALFLIQAGRNEELTTEQK QWEKKLRLNPEQQTKFQKLLFKHFPVLGPMMADVADHKAYL NKKKSTVQTEDETFAMLKGVSLADCLDIICLMADTLTECRNFY THKDPYNKPSQLADQYLHQEMIAKKLDKVVVASRRILKDREGL SVNEVEFLTGIDHLHQEVLKDEFGNAKVKDGKVMKTFVEYDD FYFKISGKRLVNGYTVTTKDDKPVNVNTMLPALSDFGLLYFCV LFLSKPYAKLFIDEVRLFEYSPFDDKENMIMSEMLSIYRIRTPRL HKIDSHDSKATLAMDIFGELRRCPMELYNLLDKNAGQPFFHDE VKHPNSHTPDVSKRLRYDDRFPTLALRYIDETELFKRIRFQLQL GSFRYKFYDKENCIDGRVRVRRIQKEINGYGRMQEVADKRMD KWGDLIQKREERSVKLEHEELYINLDQFLEDTADSTPYVTDRRP AYNIHANRIGLYWEDSQNPKQYKVFDENGMYIPELVVTEDKKA PIKMPAPRCALSVYDLPAMLFYEYLREQQDNEFPSAEQVIIEYE DDYRKFFKAVAEGKLKPFKRPKEFRDFLKKEYPKLRMADIPKK LQLFLCSHGLCYNNKPETVYERLDRLTLQHLEERELHIQNRLEH YQKDRDMIGNKDNQYGKKSFSDVRHGALARYLAQSMMEWQP TKLKDKEKGHDKLTGLNYNVLTAYLATYGHPQVPEEGFTPRTL EQVLINAHLIGGSNPHPFINKVLALGNRNIEELYLHYLEEELKHI RSRIQSLSSNPSDKALSALPFIHHDRMRYHERTSEEMMALAARY TTIQLPDGLFTPYILEILQKHYTENSDLQNALSQDVPVKLNPTCN AAYLITLFYQTVLKDNAQPFYLSDKTYTRNKDGEKAESFSFKR AYELFSVLNNNKKDTFPFEMIPLFLTSDEIQERLSAKLLDGDGNP VPEVGEKGKPATDSQGNTIWKRRIYSEVDDYAEKLTDRDMKIS FKGEWEKLPRWKQDKIIKRRDETRRQMRDELLQRMPRYIRDIK DNERTLRRYKTQDMVLFLLAEKMFTNIISEQSSEFNWKQMRLS KVCNEAFLRQTLTFRVPVTVGETTIYVEQENMSLKNYGEFYRFL TDDRLMSLLNNIVETLKPNENGDLVIRHTDLMSELAAYDQYRS TIFMLIQSIENLIITNNAVLDDPDADGFWVREDLPKRNNFASLLE LINQLNNVELTDDERKLLVAIRNAFSHNSYNIDFSLIKDVKHLPE VAKGILQHLQSMLGVEITK
Riemerella anatipestifer	8	MEKPLLPNVYTLKHKFFWGAFLNIARHNAFITICHINEQLGLKT PSNDDKIVDVVCETWNNILNNDHDLLKKSQLTELILKHFPFLTA MCYHPPKKEGKKKGHQKEQQKEKESEAQSQAEALNPSKLIEAL EILVNQLHSLRNYYSHYKHKKPDAEKDIFKHLYKAFDASLRMV KEDYKAHFTVNLTRDFAHLNRKGKNKQDNPDFNRYRFEKDGF

[0758]

		FTESGLLFFTNLFLDKRDAYWMLKKVSGFKASHKQREKMTTE VFCRSRILLPKLRLESRYDHNQMLLDMLSELSRCPKLLYEKLSE ENKKHFQVEADGFLDEIEEEQNPFKDTLIRHQDRFPYFALRYLD LNESFKSIRFQVDLGTYHYCIYDKKIGDEQEKRHLTRTLLSFGRL QDFTEINRPQEWKALTKDLDYKETSNQPFISKTTPHYHITDNKIG FRLGTSKELYPSLEIKDGANRIAKYPYNSGFVAHAFISVHELLPL MFYQHLTGKSEDLLKETVRHIQRIYKDFEEERINTIEDLEKANQ GRLPLGAFPKQMLGLLQNKQPDLSEKAKIKIEKLIAETKLLSHR LNTKLKSSPKLGKRREKLIKTGVLADWLVKDFMRFQPVAYDA QNQPIKSSKANSTEFWFIRRALALYGGEKNRLEGYFKQTNLIGN TNPHPFLNKFNWKACRNLVDFYQQYLEQREKFLEAIKNQPWEP YQYCLLLKIPKENRKNLVKGWEQGGISLPRGLFTEAIRETLSED LMLSKPIRKEIKKHGRVGFISRAITLYFKEKYQDKHQSFYNLSY KLEAKAPLLKREEHYEYWQQNKPQSPTESQRLELHTSDRWKD YLLYKRWQHLEKKLRLYRNQDVMLWLMTLELTKNHFKELNL NYHQLKLENLAVNVQEADAKLNPLNQTLPMVLPVKVYPATAF GEVQYHKTPIRTVYIREEHTKALKMGNFKALVKDRRLNGLFSFI KEENDTQKHPISQLRLRRELEIYQSLRVDAFKETLSLEEKLLNKH TSLSSLENEFRALLEEWKKEYAASSMVTDEHIAFIASVRNAFCH NQYPFYKEALHAPIPLFTVAQPTTEEKDGLGIAEALLKVLREYC EIVKSQI
Prevotella aurantiaca	9	MEDDKKTTGSISYELKDKHFWAAFLNLARHNVYITINHINKLLE IREIDNDEKVLDIKTLWQKGNKDLNQKARLRELMTKHFPFLET AIYTKNKEDKKEVKQEKQAEAQSLESLKDCLFLFLDKLQEARN YYSHYKYSEFSKEPEFEEGLLEKMYNIFGNNIQLVINDYQHNKD INPDEDFKHLDRKGQFKYSFADNEGNITESGLLFFVSLFLEKKD AIWMQQKLNGFKDNLENKKKMTHEVFCRSRILMPKLRLESTQT QDWILLDMLNELIRCPKSLYERLQGDDREKFKVPFDPADEDYN AEQEPFKNTLIRHQDRFPYFVLRYFDYNEIFKNLRFQIDLGTYHF SIYKKLIGGQKEDRHLTHKLYGFERIQEFAKQNRPDEWKAIVKD LDTYETSNKRYISETTPHYHLENQKIGIRFRNGNKEIWPSLKTND ENNEKSKYKLDKQYQAEAFLSVHELLPMMFYYLLLKKEKPNN DEINASIVEGFIKREIRNIFKLYDAFANGEINNIDDLEKYCADKGI PKRHLPKQMVAILYDEHKDMVKEAKRKQKEMVKDTKKLLAT LEKQTQKEKEDDGRNVKLLKSGEIARWLVNDMMRFQPVQKD NEGKPLNNSKANSTEYQMLQRSLALYNNEEKPTRYFRQVNLIE SNNPHPFLKWTKWEECNNILTFYYSYLTKKIEFLNKLKPEDWK KNQYFLKLKEPKTNRETLVQGWKNGFNLPRGIFTEPIREWFKR HQNNSKEYEKVEALDRVGLVTKVIPLFFKEEYFKDKEENFKED TQKEINDCVQPFYNFPYNVGNIHKPKEKDFLHREERIELWDKKK DKFKGYKEKIKSKKLTEKDKEEFRSYLEFQSWNKFERELRLVR NQDIVTWLLCKELIDKLKIDELNIEELKKLRLNNIDTDTAKKEK NNILNRVMPMELPVTVYEIDDSHKIVKDKPLHTIYIKEAETKLL KQGNFKALVKDRRLNGLFSFVKTNSEAESKRNPISKLRVEYELG EYQEARIEIIQDMLALEEKLINKYKDLPTNKFSEMLNSWLEGKD EADKARFQNDVDFLIAVRNAFSHNQYPMHNKIEFANIKPFSLYT

[0759]

		ANNSEEKGLGIANQLKDKTKETTDKIKKIEKPIETKE
Prevotella saccharolytica	10	MEDKPFWAAFFNLARHNVYLTVNHINKLLDLEKLYDEGKHKEI FEREDIFNISDDVMNDANSNGKKRKLDIKKIWDDLDTDLTRKY QLRELILKHFPFIQPAIIGAQTKERTTIDKDKRSTSTSNDSLKQTG EGDINDLLSLSNVKSMFFRLLQILEQLRNYYSHVKHSKSATMPN FDEDLLNWMRYIFIDSVNKVKEDYSSNSVIDPNTSFSHLIYKDE QGKIKPCRYPFTSKDGSINAFGLLFFVSLFLEKQDSIWMQKKIPG FKKASENYMKMTNEVFCRNHILLPKIRLETVYDKDWMLLDML NEVVRCPLSLYKRLTPAAQNKFKVPEKSSDNANRQEDDNPFSRI LVRHQNRFPYFVLRFFDLNEVFTTLRFQINLGCYHFAICKKQIGD KKEVHHLIRTLYGFSRLQNFTQNTRPEEWNTLVKTTEPSSGNDG KTVQGVPLPYISYTIPHYQIENEKIGIKIFDGDTAVDTDIWPSVST EKQLNKPDKYTLTPGFKADVFLSVHELLPMMFYYQLLLCEGML KTDAGNAVEKVLIDTRNAIFNLYDAFVQEKINTITDLENYLQDK PILIGHLPKQMIDLLKGHQRDMLKAVEQKKAMLIKDTERRLKL LDKQLKQETDVAAKNTGTLLKNGQIADWLVNDMMRFQPVKR DKEGNPINCSKANSTEYQMLQRAFAFYATDSCRLSRYFTQLHLI HSDNSHLFLSRFEYDKQPNLIAFYAAYLKAKLEFLNELQPQNW ASDNYFLLLRAPKNDRQKLAEGWKNGFNLPRGLFTEKIKTWFN EHKTIVDISDCDIFKNRVGQVARLIPVFFDKKFKDHSQPFYRYDF NVGNVSKPTEANYLSKGKREELFKSYQNKFKNNIPAEKTKEYR EYKNFSLWKKFERELRLIKNQDILIWLMCKNLFDEKIKPKKDIL EPRIAVSYIKLDSLQTNTSTAGSLNALAKVVPMTLAIHIDSPKPK GKAGNNEKENKEFTVYIKEEGTKLLKWGNFKTLLADRRIKGLF SYIEHDDIDLKQHPLTKRRVDLELDLYQTCRIDIFQQTLGLEAQL LDKYSDLNTDNFYQMLIGWRKKEGIPRNIKEDTDFLKDVRNAF SHNQYPDSKKIAFRRIRKFNPKELILEEEEGLGIATQMYKEVEKV VNRIKRIELFD
HMPREF9712_03108 [Myroides odoratimimus CCUG 10230]	11	MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEE VNKRNTFGKLAKRDNGNLKNYIIHVFKDELSISDFEKRVAIFAS YFPILETVDKKSIKERNRTIDLTLSQRIRQFREMLISLVTAVDQLR NFYTHYHHSDIVIENKVLDFLNSSFVSTALHVKDKYLKTDKTKE FLKETIAAELDILIEAYKKKQIEKKNTRFKANKREDILNAIYNEA FWSFINDKDKDKDKETVVAKGADAYFEKNHHKSNDPDFALNIS EKGIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIKYMA TQRIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVPNVVY QHLSTTQQNSFIEDWNEYYKDYEDDVETDDLSRVIHPVIRKRYE DRFNYFAIRFLDEFFDFPTLRFQVHLGDYVHDRRTKQLGKVESD RIIKEKVTVFARLKDINSAKASYFHSLEEQDKEELDNKWTLFPN PSYDFPKEHTLQHQGEQKNAGKIGIYVKLRDTQYKEKAALEEA RKSLNPKERSATKASKYDIITQIIEANDNVKSEKPLVFTGQPIAY LSMNDIHSMLFSLLTDNAELKKTPEEVEAKLIDQIGKQINEILSK DTDTKILKKYKDNDLKETDTDKITRDLARDKEEIEKLILEQKQR ADDYNYTSSTKFNIDKSRKRKHLLFNAEKGKIGVWLANDIKRF MFKESKSKWKGYQHTELQKLFAYFDTSKSDLELILSNMVMVK

[0760]

		DYPIELIDLVKKSRTLVDFLNKYLEARLEYIENVITRVKNSIGTP QFKTVRKECFTFLKKSNYTVVSLDKQVERILSMPLFIERGFMDD KPTMLEGKSYKQHKEKFADWFVHYKENSNYQNFYDTEVYEIT TEDKREKAKVTKKIKQQQKNDVFTLMMVNYMLEEVLKLSSND RLSLNELYQTKEERIVNKQVAKDTQERNKNYIWNKVVDLQLC DGLVHIDNVKLKDIGNFRKYENDSRVKEFLTYQSDIVWSAYLS NEVDSNKLYVIERQLDNYESIRSKELLKEVQEIECSVYNQVANK ESLKQSGNENFKQYVLQGLLPIGMDVREMLILSTDVKFKKEEII QLGQAGEVEQDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISD NEYYAEYYMEIFRSIKEKYAN
Capnocytophaga canimorsus	13	MKNIQRLGKGNEFSPFKKEDKFYFGGFLNLANNNIEDFFKEIITR FGIVITDENKKPKETFGEKILNEIFKKDISIVDYEKWVNIFADYFP FTKYLSLYLEEMQFKNRVICFRDVMKELLKTVEALRNFYTHYD HEPIKIEDRVFYFLDKVLLDVSLTVKNKYLKTDKTKEFLNQHIG EELKELCKQRKDYLVGKGKRIDKESEIINGIYNNAFKDFICKREK QDDKENHNSVEKILCNKEPQNKKQKSSATVWELCSKSSSKYTE KSFPNRENDKHCLEVPISQKGIVFLLSFFLNKGEIYALTSNIKGFK AKITKEEPVTYDKNSIRYMATHRMFSFLAYKGLKRKIRTSEINY NEDGQASSTYEKETLMLQMLDELNKVPDVVYQNLSEDVQKTFI EDWNEYLKENNGDVGTMEEEQVIHPVIRKRYEDKFNYFAIRFL DEFAQFPTLRFQVHLGNYLCDKRTKQICDTTTEREVKKKITVFG RLSELENKKAIFLNEREEIKGWEVFPNPSYDFPKENISVNYKDFP IVGSILDREKQPVSNKIGIRVKIADELQREIDKAIKEKKLRNPKNR KANQDEKQKERLVNEIVSTNSNEQGEPVVFIGQPTAYLSMNDIH SVLYEFLINKISGEALETKIVEKIETQIKQIIGKDATTKILKPYTNA NSNSINREKLLRDLEQEQQILKTLLEEQQQREKDKKDKKSKRK HELYPSEKGKVAVWLANDIKRFMPKAFKEQWRGYHHSLLQKY LAYYEQSKEELKNLLPKEVFKHFPFKLKGYFQQQYLNQFYTDY LKRRLSYVNELLLNIQNFKNDKDALKATEKECFKFFRKQNYIIN PINIQIQSILVYPIFLKRGFLDEKPTMIDREKFKENKDTELADWF MHYKNYKEDNYQKFYAYPLEKVEEKEKFKRNKQINKQKKND VYTLMMVEYIIQKIFGDKFVEENPLVLKGIFQSKAERQQNNTHA ATTQERNLNGILNQPKDIKIQGKITVKGVKLKDIGNFRKYEIDQR VNTFLDYEPRKEWMAYLPNDWKEKEKQGQLPPNNVIDRQISK YETVRSKILLKDVQELEKIISDEIKEEHRHDLKQGKYYNFKYYIL NGLLRQLKNENVENYKVFKLNTNPEKVNITQLKQEATDLEQKA FVLTYIRNKFAHNQLPKKEFWDYCQEKYGKIEKEKTYAEYFAE VFKREKEALIK
Porphyromonas gulae	14	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDQDVLSFKALWKNFDNDLERKSRLRSLILKHFSF LEGAAYGKKLFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSGSSELPLFDGNMLQRLYNV FDVSVQRVKIDHEHNDEVDPHYHFNHLVRKGKKDRYGHNDNP SFKHHFVDGEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKG GTETYQQMTNEVFCRSRISLPKLKLESLRMDDWMLLDMLNEL

[0761]

		VRCPKPLYDRLREDDRACFRVPVDILPDEDDTDGGGEDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETG DKPYISQTSPHYHIEKGKIGLRFMPEGQHLWPSPEVGTTRTGRS KYAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAERV QGRIKRVIEDVYAVYDAFARDEINTRDELDACLADKGIRRGHLP RQMIAILSQEHKDMEEKIRKKLQEMMADTDHRLDMLDRQTDR KIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDASGKPLNNS KANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFL HETRWESHTNILSFYRSYLRARKAFLERIGRSDRVENRPFLLLKE PKTDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEMGHDEVASYK EVGFMAKAVPLYFERACEDRVQPFYDSPFNVGNSLKPKKGRFL SKEERAEEWERGKERFRDLEAWSYSAARRIEDAFAGIEYASPG NKKKIEQLLRDLSLWEAFESKLKVRADRINLAKLKKEILEAQEH PYHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLD TGTLYLKDIRPNVQEQGSLNVLNRVKPMRLPVVVYRADSRGH VHKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDT GGLAMEQYPISKLRVEYELAKYQTARVCVFELTLRLEESLLTRY PHLPDESFREMLESWSDPLLAKWPELHGKVRLLIAVRNAFSHN QYPMYDEAVFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAK ETVERIIQA
Prevotella sp. P5-125	15	MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQ NENNENLWFHPVMSHLYNAKNGYDKQPEKTMFIIERLQSYFPF LKIMAENQREYSNGKYKQNRVEVNSNDIFEVLKRAFGVLKMY RDLTNHYKTYEEKLNDGCEFLTSTEQPLSGMINNYYTVALRNM NERYGYKTEDLAFIQDKRFKFVKDAYGKKKSQVNTGFFLSLQD YNGDTQKKLHLSGVGIALLICLFLDKQYINIFLSRLPIFSSYNAQS EERRIIIRSFGINSIKLPKDRIHSEKSNKSVAMDMLNEVKRCPDEL FTTLSAEKQSRFRIISDDHNEVLMKRSSDRFVPLLLQYIDYGKLF DHIRFHVNMGKLRYLLKADKTCIDGQTRVRVIEQPLNGFGRLE EAETMRKQENGTFGNSGIRIRDFENMKRDDANPANYPYIVDTY THYILENNKVEMFINDKEDSAPLLPVIEDDRYVVKTIPSCRMSTL EIPAMAFHMFLFGSKKTEKLIVDVHNRYKRLFQAMQKEEVTAE NIASFGIAESDLPQKILDLISGNAHGKDVDAFIRLTVDDMLTDTE RRIKRFKDDRKSIRSADNKMGKRGFKQISTGKLADFLAKDIVLF QPSVNDGENKITGLNYRIMQSAIAVYDSGDDYEAKQQFKLMFE KARLIGKGTTEPHPFLYKVFARSIPANAVEFYERYLIERKFYLTG LSNEIKKGNRVDVPFIRRDQNKWKTPAMKTLGRIYSEDLPVELP RQMFDNEIKSHLKSLPQMEGIDFNNANVTYLIAEYMKRVLDDD FQTFYQWNRNYRYMDMLKGEYDRKGSLQHCFTSVEEREGLW KERASRTERYRKQASNKIRSNRQMRNASSEEIETILDKRLSNSR NEYQKSEKVIRRYRVQDALLFLLAKKTLTELADFDGERFKLKEI MPDAEKGILSEIMPMSFTFEKGGKKYTITSEGMKLKNYGDFFVL ASDKRIGNLLELVGSDIVSKEDIMEEFNKYDQCRPEISSIVFNLE KWAFDTYPELSARVDREEKVDFKSILKILLNNKNINKEQSDILR KIRNAFDHNNYPDKGVVEIKALPEIAMSIKKAFGEYAIMK

[0762]

Flavobacterium branchiophilum	16	MENLNKILDKENEICISKIFNTKGIAAPITEKALDNIKSKQKNDL NKEARLHYFSIGHSFKQIDTKKVFDYVLIEELKDEKPLKFITLQK DFFTKEFSIKLQKLINSIRNINNHYVHNFNDINLNKIDSNVFHFLK ESFELAIIEKYYKVNKKYPLDNEIVLFLKELFIKDENTALLNYFT NLSKDEAIEYILTFTITENKIWNINNEHNILNIEKGKYLTFEAMLF LITIFLYKNEANHLLPKLYDFKNNKSKQELFTFFSKKFTSQDIDA EEGHLIKFRDMIQYLNHYPTAWNNDLKLESENKNKIMTTKLIDS IIEFELNSNYPSFATDIQFKKEAKAFLFASNKKRNQTSFSNKSYN EEIRHNPHIKQYRDEIASALTPISFNVKEDKFKIFVKKHVLEEYFP NSIGYEKFLEYNDFTEKEKEDFGLKLYSNPKTNKLIERIDNHKL VKSHGRNQDRFMDFSMRFLAENNYFGKDAFFKCYKFYDTQEQ DEFLQSNENNDDVKFHKGKVTTYIKYEEHLKNYSYWDCPFVEE NNSMSVKISIGSEEKILKIQRNLMIYFLENALYNENVENQGYKL VNNYYRELKKDVEESIASLDLIKSNPDFKSKYKKILPKRLLHNY APAKQDKAPENAFETLLKKADFREEQYKKLLKKAEHEKNKED FVKRNKGKQFKLHFIRKACQMMYFKEKYNTLKEGNAAFEKKD PVIEKRKNKEHEFGHHKNLNITREEFNDYCKWMFAFNGNDSYK KYLRDLFSEKHFFDNQEYKNLFESSVNLEAFYAKTKELFKKWIE TNKPTNNENRYTLENYKNLILQKQVFINVYHFSKYLIDKNLLNS ENNVIQYKSLENVEYLISDFYFQSKLSIDQYKTCGKLFNKLKSN KLEDCLLYEIAYNYIDKKNVHKIDIQKILTSKIILTINDANTPYKIS VPFNKLERYTEMIAIKNQNNLKARFLIDLPLYLSKNKIKKGKDS AGYEIIIKNDLEIEDINTINNKIINDSVKFTEVLMELEKYFILKDKC ILSKNYIDNSEIPSLKQFSKVWIKENENEIINYRNIACHFHLPLLET FDNLLLNVEQKFIKEELQNVSTINDLSKPQEYLILLFIKFKHNNF YLNLFNKNESKTIKNDKEVKKNRVLQKFINQVILKKK
Myroides odoratimimus	17	MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEE VNKRNTFGKLAKRDNGNLKNYIIHVFKDELSISDFEKRVAIFAS YFPILETVDKKSIKERNRTIDLTLSQRIRQFREMLISLVTAVDQLR NFYTHYHHSDIVIENKVLDFLNSSFVSTALHVKDKYLKTDKTKE FLKETIAAELDILIEAYKKKQIEKKNTRFKANKREDILNAIYNEA FWSFINDKDKDKDKETVVAKGADAYFEKNHHKSNDPDFALNIS EKGIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIKYMA TQRIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVPNVVY QHLSTTQQNSFIEDWNEYYKDYEDDVETDDLSRVTHPVIRKRY EDRFNYFAIRFLDEFFDFPTLRFQVHLGDYVHDRRTKQLGKVES DRIIKEKVTVFARLKDINSAKASYFHSLEEQDKEELDNKWTLFP NPSYDFPKEHTLQHQGEQKNAGKIGIYVKLRDTQYKEKAALEE ARKSLNPKERSATKASKYDIITQIIEANDNVKSEKPLVFTGQPIA YLSMNDIHSMLFSLLTDNAELKKTPEEVEAKLIDQIGKQINEILS KDTDTKILKKYKDNDLKETDTDKITRDLARDKEEIEKLILEQKQ RADDYNYTSSTKFNIDKSRKRKHLLFNAEKGKIGVWLANDIKR FMFKESKSKWKGYQHIELQKLFAYFDTSKSDLELILSNMVMVK DYPIELIDLVKKSRTLVDFLNKYLEARLEYIENVITRVKNSIGTP QFKTVRKECFTFLKKSNYTVVSLDKQVERILSMPLFIERGFMDD KPTMLEGKSYKQHKEKFADWFVHYKENSNYQNFYDTEVYEIT

[0763]

		TEDKREKAKVTKKIKQQQKNDVFTLMMVNYMLEEVLKLSSND RLSLNELYQTKEERIVNKQVAKDTQERNKNYIWNKVVDLQLC DGLVHIDNVKLKDIGNFRKYENDSRVKEFLTYQSDIVWSAYLS NEVDSNKLYVIERQLDNYESIRSKELLKEVQEIECSVYNQVANK ESLKQSGNENFKQYVLQGLLPIGMDVREMLILSTDVKFKKEEII QLGQAGEVEQDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISD NEYYAEYYMEIFRSIKEKYAN
Flavobacterium columnare	18	MSSKNESYNKQKTFNHYKQEDKYFFGGFLNNADDNLRQVGKE FKTRINFNHNNNELASVFKDYFNKEKSVAKREHALNLLSNYFP VLERIQKHTNHNFEQTREIFELLLDTIKKLRDYYTHHYHKPITIN PKIYDFLDDTLLDVLITIKKKKVKNDTSRELLKEKLRPELTQLKN QKREELIKKGKKLLEENLENAVFNHCLIPFLEENKTDDKQNKTV SLRKYRKSKPNEETSITLTQSGLVFLMSFFLHRKEFQVFTSGLER FKAKVNTIKEEEISLNKNNIVYMITHWSYSYYNFKGLKHRIKTD QGVSTLEQNNTTHSLTNTNTKEALLTQIVDYLSKVPNEIYETLSE KQQKEFEEDINEYMRENPENEDSTFSSIVSHKVIRKRYENKFNY FAMRFLDEYAELPTLRFMVNFGDYIKDRQKKILESIQFDSERIIK KEIHLFEKLSLVTEYKKNVYLKETSNIDLSRFPLFPNPSYVMAN NNIPFYIDSRSNNLDEYLNQKKKAQSQNKKRNLTFEKYNKEQS KDAIIAMLQKEIGVKDLQQRSTIGLLSCNELPSMLYEVIVKDIKG AELENKIAQKIREQYQSIRDFTLDSPQKDNIPTTLIKTINTDSSVT FENQPIDIPRLKNALQKELTLTQEKLLNVKEHEIEVDNYNRNKN TYKFKNQPKNKVDDKKLQRKYVFYRNEIRQEANWLASDLIHF MKNKSLWKGYMHNELQSFLAFFEDKKNDCIALLETVFNLKED CILTKGLKNLFLKHGNFIDFYKEYLKLKEDFLSTESTFLENGFIG LPPKILKKELSKRLKYIFIVFQKRQFIIKELEEKKNNLYADAINLS RGIFDEKPTMIPFKKPNPDEFASWFVASYQYNNYQSFYELTPDI VERDKKKKYKNLRAINKVKIQDYYLKLMVDTLYQDLFNQPLD KSLSDFYVSKAEREKIKADAKAYQKLNDSSLWNKVIHLSLQNN RITANPKLKDIGKYKRALQDEKIATLLTYDARTWTYALQKPEK ENENDYKELHYTALNMELQEYEKVRSKELLKQVQELEKKILDK FYDFSNNASHPEDLEIEDKKGKRHPNFKLYITKALLKNESEIINL ENIDIEILLKYYDYNTEELKEKIKNMDEDEKAKIINTKENYNKIT NVLIKKALVLIIIRNKMAHNQYPPKFIYDLANRFVPKKEEEYFAT YFNRVFETITKELWENKEKKDKTQV
Porphyromonas sp. COT-052 OH4946	20	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDQDVLSFKALWKNFDNDLERKSRLRSLILKHFSF LEGAAYGKKLFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRYGHNDN PSFKHHFVDSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTETYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKPLYDRLREDDRACFRVPVDILPDEDDTDGGGEDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETG

[0764]

		DKPYISQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGTTRTGRSK YAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKVQ GRIKRVIEDVYAIYDAFARDEINTLKELDACLADKGIRRGHLPK QMIGILSQERKDMEEKVRKKLQEMIADTDHRLDMLDRQTDRKI RIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSK ANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLH ETRWESHTNILSFYRSYLRARKAFLERIGRSDRVENCPFLLLKEP KTDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEMGYDEVGSYRE VGFMAKAVPLYFERACEDRVQPFYDSPFNVGNSLKPKKGRFLS KEDRAEEWERGKERFRDLEAWSHSAARRIKDAFAGIEYASPGN KKKIEQLLRDLSLWEAFESKLKVRADKINLAKLKKEILEAQEHP YHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLDT GTLYLKDIRPNVQEQGSLNVLNRVKPMRLPVVVYRADSRGHV HKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTG GLAMEQYPISKLRVEYELAKYQTARVCVFELTLRLEESLLSRYP HLPDESFREMLESWSDPLLAKWPELHGKVRLLIAVRNAFSHNQ YPMYDEAVFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAKE TVERIIQA
Prevotella intermedia	21	MEDDKKTKESTNMLDNKHFWAAFLNLARHNVYITVNHINKVL ELKNKKDQDIIIDNDQDILAIKTHWEKVNGDLNKTERLRELMTK HFPFLETAIYTKNKEDKEEVKQEKQAKAQSFDSLKHCLFLFLEK LQEARNYYSHYKYSESTKEPMLEKELLKKMYNIFDDNIQLVIK DYQHNKDINPDEDFKHLDRTEEEFNYYFTTNKKGNITASGLLFF VSLFLEKKDAIWMQQKLRGFKDNRESKKKMTHEVFCRSRMLL PKLRLESTQTQDWILLDMLNELIRCPKSLYERLQGEYRKKFNVP FDSADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNLR FQIDLGTYHFSIYKKLIGGQKEDRHLTHKLYGFERIQEFAKQNR TDEWKAIVKDFDTYETSEEPYISETAPHYHLENQKIGIRFRNDN DEIWPSLKTNGENNEKRKYKLDKQYQAEAFLSVHELLPMMFY YLLLKKEEPNNDKKNASIVEGFIKREIRDIYKLYDAFANGEINNI DDLEKYCEDKGIPKRHLPKQMVAILYDEHKDMAEEAKRKQKE MVKDTKKLLATLEKQTQGEIEDGGRNIRLLKSGEIARWLVNDM MRFQPVQKDNEGNPLNNSKANSTEYQMLQRSLALYNKEEKPT RYFRQVNLINSSNPHPFLKWTKWEECNNILSFYRSYLTKKIEFLN KLKPEDWEKNQYFLKLKEPKTNRETLVQGWKNGFNLPRGIFTE PIREWFKRHQNDSEEYEKVETLDRVGLVTKVIPLFFKKEDSKDK EEYLKKDAQKEINNCVQPFYGFPYNVGNIHKPDEKDFLPSEERK KLWGDKKYKFKGYKAKVKSKKLTDKEKEEYRSYLEFQSWNK FERELRLVRNQDIVTWLLCTELIDKLKVEGLNVEELKKLRLKDI DTDTAKQEKNNILNRVMPMQLPVTVYEIDDSHNIVKDRPLHTV YIEETKTKLLKQGNFKALVKDRRLNGLFSFVDTSSETELKSNPIS KSLVEYELGEYQNARIETIKDMLLLEETLIEKYKTLPTDNFSDM LNGWLEGKDEADKARFQNDVKLLVAVRNAFSHNQYPMRNRIA FANINPFSLSSADTSEEKKLDIANQLKDKTHKIIKRIIEIEKPIETK E

[0765]

PIN17_0200 [Prevotella intermedia 17]	AFJ07523	MKMEDDKKTKESTNMLDNKHFWAAFLNLARHNVYITVNHIN KVLELKNKKDQDIIIDNDQDILAIKTHWEKVNGDLNKTERLREL MTKHFPFLETAIYTKNKEDKEEVKQEKQAKAQSFDSLKHCLFL FLEKLQEARNYYSHYKYSESTKEPMLEKELLKKMYNIFDDNIQ LVIKDYQHNKDINPDEDFKHLDRTEEEFNYYFTTNKKGNITASG LLFFVSLFLEKKDAIWMQQKLRGFKDNRESKKKMTHEVFCRSR MLLPKLRLESTQTQDWILLDMLNELIRCPKSLYERLQGEYRKKF NVPFDSADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFT NLRFQIDLGTYHFSIYKKLIGGQKEDRHLTHKLYGFERIQEFAK QNRTDEWKAIVKDFDTYETSEEPYISETAPHYHLENQKIGIRFRN DNDEIWPSLKTNGENNEKRKYKLDKQYQAEAFLSVHELLPMM FYYLLLKKEEPNNDKKNASIVEGFIKREIRDIYKLYDAFANGEIN NIDDLEKYCEDKGIPKRHLPKQMVAILYDEHKDMAEEAKRKQ KEMVKDTKKLLATLEKQTQGEIEDGGRNIRLLKSGEIARWLVN DMMRFQPVQKDNEGNPLNNSKANSTEYQMLQRSLALYNKEEK PTRYFRQVNLINSSNPHPFLKWTKWEECNNILSFYRSYLTKKIEF LNKLKPEDWEKNQYFLKLKEPKTNRETLVQGWKNGFNLPRGIF TEPIREWFKRHQNDSEEYEKVETLDRVGLVTKVIPLFFKKEDSK DKEEYLKKDAQKEINNCVQPFYGFPYNVGNIHKPDEKDFLPSEE RKKLWGDKKYKFKGYKAKVKSKKLTDKEKEEYRSYLEFQSW NKFERELRLVRNQDIVTWLLCTELIDKLKVEGLNVEELKKLRLK DIDTDTAKQEKNNILNRVMPMQLPVTVYEIDDSHNIVKDRPLHT VYIEETKTKLLKQGNFKALVKDRRLNGLFSFVDTSSETELKSNPI SKSLVEYELGEYQNARIETIKDMLLLEETLIEKYKTLPTDNFSDM LNGWLEGKDEADKARFQNDVKLLVAVRNAFSHNQYPMRNRIA FANINPFSLSSADTSEEKKLDIANQLKDKTHKIIKRIIEIEKPIETK E
Prevotella intermedia	BAU18623	MEDDKKTTDSISYELKDKHFWAAFLNLARHNVYITVNHINKVL ELKNKKDQDIIIDNDQDILAIKTHWEKVNGDLNKTERLRELMTK HFPFLETAIYSKNKEDKEEVKQEKQAKAQSFDSLKHCLFLFLEK LQETRNYYSHYKYSESTKEPMLEKELLKKMYNIFDDNIQLVIKD YQHNKDINPDEDFKHLDRTEEDFNYYFTRNKKGNITESGLLFFV SLFLEKKDAIWMQQKLRGFKDNRESKKKMTHEVFCRSRMLLP KLRLESTQTQDWILLDMLNELIRCPKSLYERLQGEDREKFKVPF DPADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNLRF QIDLGTFHFSIYKKLIGGQKEDRHLTHKLYGFERIQEFAKQNRPD EWKAIVKDLDTYETSNERYISETTPHYHLENQKIGIRFRNDNDEI WPSLKTNGENNEKSKYKLDKQYQAEAFLSVHELLPMMFYYLL LKKEEPNNDKKNASIVEGFIKREIRDMYKLYDAFANGEINNIDD LEKYCEDKGIPKRHLPKQMVAILYDEHKDMVKEAKRKQRKMV KDTEKLLAALEKQTQEKTEDGGRNIRLLKSGEIARWLVNDMM RFQPVQKDNEGNPLNNSKANSTEYQMLQRSLALYNKEEKPTRY FRQVNLINSSNPHPFLKWTKWEECNNILSFYRSYLTKKIEFLNKL KPEDWEKNQYFLKLKEPKTNRETLVQGWKNGFNLPRGIFTEPIR EWFKRHQNDSKEYEKVEALDRVGLVTKVIPLFFKKEDSKDKEE DLKKDAQKEINNCVQPFYSFPYNVGNIHKPDEKDFLHREERIEL

[0766]

		WDKKKDKFKGYKAKVKSKKLTDKEKEEYRSYLEFQSWNKFER ELRLVRNQDIVTWLLCTELIDKLKVEGLNVEELKKLRLKDIDTD TAKQEKNNILNRVMPMQLPVTVYEIDDSHNIVKDRPLHTVYIEE TKTKLLKQGNFKALVKDRRLNGLFSFVDTSSEAELKSNPISKSL VEYELGEYQNARIETIKDMLLLEETLIEKYKNLPTDNFSDMLNG WLEGKDEADKARFQNDVKLLVAVRNAFSHNQYPMRNRIAFAN INPFSLSSADTSEEKKLDIANQLKDKTHKIIKRIIEIEKPIETKE
HMPREF6485_0083 [Prevotella buccae ATCC 33574]	EFU31981	MQKQDKLFVDRKKNAIFAFPKYITIMENKEKPEPIYYELTDKHF WAAFLNLARHNVYTTINHINRRLEIAELKDDGYMMGIKGSWNE QAKKLDKKVRLRDLIMKHFPFLEAAAYEMTNSKSPNNKEQRE KEQSEALSLNNLKNVLFIFLEKLQVLRNYYSHYKYSEESPKPIFE TSLLKNMYKVFDANVRLVKRDYMHHENIDMQRDFTHLNRKK QVGRTKNIIDSPNFHYHFADKEGNMTIAGLLFFVSLFLDKKDAI WMQKKLKGFKDGRNLREQMTNEVFCRSRISLPKLKLENVQTK DWMQLDMLNELVRCPKSLYERLREKDRESFKVPFDIFSDDYNA EEEPFKNTLVRHQDRFPYFVLRYFDLNEIFEQLRFQIDLGTYHFS IYNKRIGDEDEVRHLTHHLYGFARIQDFAPQNQPEEWRKLVKD LDHFETSQEPYISKTAPHYHLENEKIGIKFCSAHNNLFPSLQTDK TCNGRSKFNLGTQFTAEAFLSVHELLPMMFYYLLLTKDYSRKE SADKVEGIIRKEISNIYAIYDAFANNEINSIADLTRRLQNTNILQG HLPKQMISILKGRQKDMGKEAERKIGEMIDDTQRRLDLLCKQT NQKIRIGKRNAGLLKSGKIADWLVNDMMRFQPVQKDQNNIPIN NSKANSTEYRMLQRALALFGSENFRLKAYFNQMNLVGNDNPH PFLAETQWEHQTNILSFYRNYLEARKKYLKGLKPQNWKQYQH FLILKVQKTNRNTLVTGWKNSFNLPRGIFTQPIREWFEKHNNSK RIYDQILSFDRVGFVAKAIPLYFAEEYKDNVQPFYDYPFNIGNRL KPKKRQFLDKKERVELWQKNKELFKNYPSEKKKTDLAYLDFLS WKKFERELRLIKNQDIVTWLMFKELFNMATVEGLKIGEIHLRDI DTNTANEESNNILNRIMPMKLPVKTYETDNKGNILKERPLATFY IEETETKVLKQGNFKALVKDRRLNGLFSFAETTDLNLEEHPISKL SVDLELIKYQTTRISIFEMTLGLEKKLIDKYSTLPTDSFRNMLER WLQCKANRPELKNYVNSLIAVRNAFSHNQYPMYDATLFAEVK KFTLFPSVDTKKIELNIAPQLLEIVGKAIKEIEKSENKN
HMPREF9144_1146 [Prevotella pallens ATCC 700821]	EGQ18444	MKEEEKGKTPVVSTYNKDDKHFWAAFLNLARHNVYITVNHIN KILGEGEINRDGYENTLEKSWNEIKDINKKDRLSKLIIKHFPFLE VTTYQRNSADTTKQKEEKQAEAQSLESLKKSFFVFIYKLRDLRN HYSHYKHSKSLERPKFEEDLQEKMYNIFDASIQLVKEDYKHNT DIKTEEDFKHLDRKGQFKYSFADNEGNITESGLLFFVSLFLEKK DAIWVQKKLEGFKCSNESYQKMTNEVFCRSRMLLPKLRLQSTQ TQDWILLDMLNELIRCPKSLYERLREEDRKKFRVPIEIADEDYD AEQEPFKNALVRHQDRFPYFALRYFDYNEIFTNLRFQIDLGTYH FSIYKKQIGDYKESHHLTHKLYGFERIQEFTKQNRPDEWRKFVK TFNSFETSKEPYIPETTPHYHLENQKIGIRFRNDNDKIWPSLKTNS EKNEKSKYKLDKSFQAEAFLSVHELLPMMFYYLLLKTENTDND NEIETKKKENKNDKQEKHKIEEIIENKITEIYALYDAFANGKINSI

[0767]

		DKLEEYCKGKDIEIGHLPKQMIAILKSEHKDMATEAKRKQEEM LADVQKSLESLDNQINEEIENVERKNSSLKSGEIASWLVNDMM RFQPVQKDNEGNPLNNSKANSTEYQMLQRSLALYNKEEKPTRY FRQVNLIESSNPHPFLNNTEWEKCNNILSFYRSYLEAKKNFLESL KPEDWEKNQYFLMLKEPKTNCETLVQGWKNGFNLPRGIFTEPI RKWFMEHRKNITVAELKRVGLVAKVIPLFFSEEYKDSVQPFYN YLFNVGNINKPDEKNFLNCEERRELLRKKKDEFKKMTDKEKEE NPSYLEFQSWNKFERELRLVRNQDIVTWLLCMELFNKKKIKEL NVEKIYLKNINTNTTKKEKNTEEKNGEEKIIKEKNNILNRIMPMR LPIKVYGRENFSKNKKKKIRRNTFFTVYIEEKGTKLLKQGNFKA LERDRRLGGLFSFVKTHSKAESKSNTISKSRVEYELGEYQKARIE IIKDMLALEETLIDKYNSLDTDNFHNMLTGWLKLKDEPDKASF QNDVDLLIAVRNAFSHNQYPMRNRIAFANINPFSLSSANTSEEK GLGIANQLKDKTHKTIEKIIEIEKPIETKE
HMPREF9714_02132 [Myroides odoratimimus CCUG 12901]	EHO08761	MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEE VNKRNTFGKLAKRDNGNLKNYIIHVFKDELSISDFEKRVAIFAS YFPILETVDKKSIKERNRTIDLTLSQRIRQFREMLISLVTAVDQLR NFYTHYHHSEIVIENKVLDFLNSSLVSTALHVKDKYLKTDKTKE FLKETIAAELDILIEAYKKKQIEKKNTRFKANKREDILNAIYNEA FWSFINDKDKDKETVVAKGADAYFEKNHHKSNDPDFALNISEK GIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIKYMATQ RIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVPNVVYQH LSTTQQNSFIEDWNEYYKDYEDDVETDDLSRVIHPVIRKRYEDR FNYFAIRFLDEFFDFPTLRFQVHLGDYVHDRRTKQLGKVESDRII KEKVTVFARLKDINSAKANYFHSLEEQDKEELDNKWTLFPNPS YDFPKEHTLQHQGEQKNAGKIGIYVKLRDTQYKEKAALEEARK SLNPKERSATKASKYDIITQIIEANDNVKSEKPLVFTGQPIAYLS MNDIHSMLFSLLTDNAELKKTPEEVEAKLIDQIGKQINEILSKDT DTKILKKYKDNDLKETDTDKITRDLARDKEEIEKLILEQKQRAD DYNYTSSTKFNIDKSRKRKHLLFNAEKGKIGVWLANDIKRFMT EEFKSKWKGYQHTELQKLFAYYDTSKSDLDLILSDMVMVKDY PIELIALVKKSRTLVDFLNKYLEARLGYMENVITRVKNSIGTPQF KTVRKECFTFLKKSNYTVVSLDKQVERILSMPLFIERGFMDDKP TMLEGKSYQQHKEKFADWFVHYKENSNYQNFYDTEVYEITTE DKREKAKVTKKIKQQQKNDVFTLMMVNYMLEEVLKLSSNDRL SLNELYQTKEERIVNKQVAKDTQERNKNYIWNKVVDLQLCEG LVRIDKVKLKDIGNFRKYENDSRVKEFLTYQSDIVWSAYLSNEV DSNKLYVIERQLDNYESIRSKELLKEVQEIECSVYNQVANKESL KQSGNENFKQYVLQGLVPIGMDVREMLILSTDVKFIKEEIIQLG QAGEVEQDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISDNEY YAEYYMEIFRSIKEKYTS
HMPREF9711_00870 [Myroides odoratimim	EKB06014	MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEE VNKRNTFGKLAKRDNGNLKNYIIHVFKDELSISDFEKRVAIFAS YFPILETVDKKSIKERNRTIDLTLSQRIRQFREMLISLVTAVDQLR NFYTHYHHSEIVIENKVLDFLNSSLVSTALHVKDKYLKTDKTKE

[0768]

us CCUG 3837]		FLKETIAAELDILIEAYKKKQIEKKNTRFKANKREDILNAIYNEA FWSFINDKDKDKETVVAKGADAYFEKNHHKSNDPDFALNISEK GIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIKYMATQ RIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVPNVVYQH LSTTQQNSFIEDWNEYYKDYEDDVETDDLSRVIHPVIRKRYEDR FNYFAIRFLDEFFDFPTLRFQVHLGDYVHDRRTKQLGKVESDRII KEKVTVFARLKDINSAKASYFHSLEEQDKEELDNKWTLFPNPS YDFPKEHTLQHQGEQKNAGKIGIYVKLRDTQYKEKAALEEARK SLNPKERSATKASKYDIITQIIEANDNVKSEKPLVFTGQPIAYLS MNDIHSMLFSLLTDNAELKKTPEEVEAKLIDQIGKQINEILSKDT DTKILKKYKDNDLKETDTDKITRDLARDKEEIEKLILEQKQRAD DYNYTSSTKFNIDKSRKRKHLLFNAEKGKIGVWLANDIKRFMF KESKSKWKGYQHTELQKLFAYFDTSKSDLELILSDMVMVKDYP IELIDLVRKSRTLVDFLNKYLEARLGYIENVITRVKNSIGTPQFKT VRKECFAFLKESNYTVASLDKQIERILSMPLFIERGFMDSKPTML EGKSYQQHKEDFADWFVHYKENSNYQNFYDTEVYEIITEDKRE QAKVTKKIKQQQKNDVFTLMMVNYMLEEVLKLPSNDRLSLNE LYQTKEERIVNKQVAKDTQERNKNYIWNKVVDLQLCEGLVRID KVKLKDIGNFRKYENDSRVKEFLTYQSDIVWSGYLSNEVDSNK LYVIERQLDNYESIRSKELLKEVQEIECIVYNQVANKESLKQSGN ENFKQYVLQGLLPRGTDVREMLILSTDVKFKKEEIMQLGQVRE VEQDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRPISDNEYYAEY YMEIFRSIKEKYAS
HMPREF9 699_02005 [Bergeyella zoohelcum ATCC 43767]	EKB54 193	MENKTSLGNNIYYNPFKPQDKSYFAGYFNAAMENTDSVFRELG KRLKGKEYTSENFFDAIFKENISLVEYERYVKLLSDYFPMARLL DKKEVPIKERKENFKKNFKGIIKAVRDLRNFYTHKEHGEVEITD EIFGVLDEMLKSTVLTVKKKKVKTDKTKEILKKSIEKQLDILCQ KKLEYLRDTARKIEEKRRNQRERGEKELVAPFKYSDKRDDLIA AIYNDAFDVYIDKKKDSLKESSKAKYNTKSDPQQEEGDLKIPIS KNGVVFLLSLFLTKQEIHAFKSKIAGFKATVIDEATVSEATVSHG KNSICFMATHEIFSHLAYKKLKRKVRTAEINYGEAENAEQLSVY AKETLMMQMLDELSKVPDVVYQNLSEDVQKTFIEDWNEYLKE NNGDVGTMEEEQVIHPVIRKRYEDKFNYFAIRFLDEFAQFPTLR FQVHLGNYLHDSRPKENLISDRRIKEKITVFGRLSELEHKKALFI KNTETNEDREHYWEIFPNPNYDFPKENISVNDKDFPIAGSILDRE KQPVAGKIGIKVKLLNQQYVSEVDKAVKAHQLKQRKASKPSIQ NIIEEIVPINESNPKEAIVFGGQPTAYLSMNDIHSILYEFFDKWEK KKEKLEKKGEKELRKEIGKELEKKIVGKIQAQIQQIIDKDTNAKI LKPYQDGNSTAIDKEKLIKDLKQEQNILQKLKDEQTVREKEYN DFIAYQDKNREINKVRDRNHKQYLKDNLKRKYPEAPARKEVL YYREKGKVAVWLANDIKRFMPTDFKNEWKGEQHSLLQKSLAY YEQCKEELKNLLPEKVFQHLPFKLGGYFQQKYLYQFYTCYLDK RLEYISGLVQQAENFKSENKVFKKVENECFKFLKKQNYTHKEL DARVQSILGYPIFLERGFMDEKPTIIKGKTFKGNEALFADWFRY YKEYQNFQTFYDTENYPLVELEKKQADRKRKTKIYQQKKNDV FTLLMAKHIFKSVFKQDSIDQFSLEDLYQSREERLGNQERARQT

[0769]

		GERNTNYIWNKTVDLKLCDGKITVENVKLKNVGDFIKYEYDQR VQAFLKYEENIEWQAFLIKESKEEENYPYVVEREIEQYEKVRRE ELLKEVHLIEEYILEKVKDKEILKKGDNQNFKYYILNGLLKQLK NEDVESYKVFNLNTEPEDVNINQLKQEATDLEQKAFVLTYIRN KFAHNQLPKKEFWDYCQEKYGKIEKEKTYAEYFAEVFKKEKE ALIK
HMPREF9 151_01387 [Prevotella saccharolyt ica F0055]	EKY00 089	MMEKENVQGSHIYYEPTDKCFWAAFYNLARHNAYLTIAHINSF VNSKKGINNDDKVLDIIDDWSKFDNDLLMGARLNKLILKHFPFL KAPLYQLAKRKTRKQQGKEQQDYEKKGDEDPEVIQEAIANAFK MANVRKTLHAFLKQLEDLRNHFSHYNYNSPAKKMEVKFDDGF CNKLYYVFDAALQMVKDDNRMNPEINMQTDFEHLVRLGRNR KIPNTFKYNFTNSDGTINNNGLLFFVSLFLEKRDAIWMQKKIKG FKGGTENYMRMTNEVFCRNRMVIPKLRLETDYDNHQLMFDML NELVRCPLSLYKRLKQEDQDKFRVPIEFLDEDNEADNPYQENA NSDENPTEETDPLKNTLVRHQHRFPYFVLRYFDLNEVFKQLRFQ INLGCYHFSIYDKTIGERTEKRHLTRTLFGFDRLQNFSVKLQPEH WKNMVKHLDTEESSDKPYLSDAMPHYQIENEKIGIHFLKTDTE KKETVWPSLEVEEVSSNRNKYKSEKNLTADAFLSTHELLPMMF YYQLLSSEEKTRAAAGDKVQGVLQSYRKKIFDIYDDFANGTINS MQKLDERLAKDNLLRGNMPQQMLAILEHQEPDMEQKAKEKL DRLITETKKRIGKLEDQFKQKVRIGKRRADLPKVGSIADWLVND MMRFQPAKRNADNTGVPDSKANSTEYRLLQEALAFYSAYKDR LEPYFRQVNLIGGTNPHPFLHRVDWKKCNHLLSFYHDYLEAKE QYLSHLSPADWQKHQHFLLLKVRKDIQNEKKDWKKSLVAGW KNGFNLPRGLFTESIKTWFSTDADKVQITDTKLFENRVGLIAKLI PLYYDKVYNDKPQPFYQYPFNINDRYKPEDTRKRFTAASSKLW NEKKMLYKNAQPDSSDKIEYPQYLDFLSWKKLERELRMLRNQ DMMVWLMCKDLFAQCTVEGVEFADLKLSQLEVDVNVQDNLN VLNNVSSMILPLSVYPSDAQGNVLRNSKPLHTVYVQENNTKLL KQGNFKSLLKDRRLNGLFSFIAAEGEDLQQHPLTKNRLEYELSI YQTMRISVFEQTLQLEKAILTRNKTLCGNNFNNLLNSWSEHRTD KKTLQPDIDFLIAVRNAFSHNQYPMSTNTVMQGIEKFNIQTPKL EEKDGLGIASQLAKKTKDAASRLQNIINGGTN
A343_1752 [Porphyro monas gingivalis JCVI SC001]	EOA10 535	MTEQNEKPYNGTYYTLEDKHFWAAFFNLARHNAYITLTHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSF LEGAAYGKKLFESQSSGNKSSKKKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHPESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRCGNNDN PFFKHHFVDREEKVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKG GTETYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNELV RCPKSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTLV RHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQ PEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDK PYITQTTPHYHIEKGKIGLRFVPEGQLLWPSPEVGATRTGRSKY AQDKRFTAEAFLSVHELMPMMFYYFLLREKYSEEASAERVQGR

[0770]

		IKRVIEDVYAVYDAFARGEIDTLDRLDACLADKGIRRGHLPRQ MIAILSQEHKDMEEKVRKKLQEMIADTDHRLDMLDRQTDRKIR IGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSKA NSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLHE TRWESHTNILSFYRSYLKARKAFLQSIGRSDRVENHRFLLLKEP KTDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEMGLDEVGSYKE VGFMAKAVPLYFERACKDRVQPFYDYPFNVGNSLKPKKGRFLS KEKRAEEWESGKERFRDLEAWSHSAARRIEDAFAGIENASREN KKKIEQLLQDLSLWETFESKLKVKADKINIAKLKKEILEAKEHP YLDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLDT GTLYLKDIRTDVHEQGSLNVLNRVKPMRLPVVVYRADSRGHV HKEQAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDT GALAMEQYPISKLRVEYELAKYQTARVCAFEQTLELEESLLTRY PHLPDKNFRKMLESWSDPLLDKWPDLHGNVRLLIAVRNAFSHN QYPMYDETLFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAK EMVERIIQA
HMPREF1 981_03090 [Bacteroide s pyogenes F0041]	ERI817 00	MESIKNSQKSTGKTLQKDPPYFGLYLNMALLNVRKVENHIRKW LGDVALLPEKSGFHSLLTTDNLSSAKWTRFYYKSRKFLPFLEMF DSDKKSYENRRETTECLDTIDRQKISSLLKEVYGKLQDIRNAFS HYHIDDQSVKHTALIISSEMHRFIENAYSFALQKTRARFTGVFVE TDFLQAEEKGDNKKFFAIGGNEGIKLKDNALIFLICLFLDREEAF KFLSRATGFKSTKEKGFLAVRETFCALCCRQPHERLLSVNPREA LLMDMLNELNRCPDILFEMLDEKDQKSFLPLLGEEEQAHILENS LNDELCEAIDDPFEMIASLSKRVRYKNRFPYLMLRYIEEKNLLPF IRFRIDLGCLELASYPKKMGEENNYERSVTDHAMAFGRLTDFH NEDAVLQQITKGITDEVRFSLYAPRYAIYNNKIGFVRTGGSDKIS FPTLKKKGGEGHCVAYTLQNTKSFGFISIYDLRKILLLSFLDKDK AKNIVSGLLEQCEKHWKDLSENLFDAIRTELQKEFPVPLIRYTLP RSKGGKLVSSKLADKQEKYESEFERRKEKLTEILSEKDFDLSQIP RRMIDEWLNVLPTSREKKLKGYVETLKLDCRERLRVFEKREKG EHPVPPRIGEMATDLAKDIIRMVIDQGVKQRITSAYYSEIQRCLA QYAGDDNRRHLDSIIRELRLKDTKNGHPFLGKVLRPGLGHTEK LYQRYFEEKKEWLEATFYPAASPKRVPRFVNPPTGKQKELPLII RNLMKERPEWRDWKQRKNSHPIDLPSQLFENEICRLLKDKIGKE PSGKLKWNEMFKLYWDKEFPNGMQRFYRCKRRVEVFDKVVE YEYSEEGGNYKKYYEALIDEVVRQKISSSKEKSKLQVEDLTLSV RRVFKRAINEKEYQLRLLCEDDRLLFMAVRDLYDWKEAQLDL DKIDNMLGEPVSVSQVIQLEGGQPDAVIKAECKLKDVSKLMRY CYDGRVKGLMPYFANHEATQEQVEMELRHYEDHRRRVFNWV FALEKSVLKNEKLRRFYEESQGGCEHRRCIDALRKASLVSEEEY EFLVHIRNKSAHNQFPDLEIGKLPPNVTSGFCECIWSKYKAIICRI IPFIDPERRFFGKLLEQK
HMPREF1 553_02065 [Porphyro	ERJ656 37	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL

[0771]

monas gingivalis F0568]		DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPRSMGFISVHDLRKLLLMELLCEGSFSRMQSDFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMDQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLQKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRHQFRAIV AELRLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDEENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL
HMPREF1988_01768 [Porphyromonas gingivalis F0185]	ERJ81987	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPQSMGFISVHDLRKLLLMELLCEGSFSRMQSGFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMNQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLRKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRRQFRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDEENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI

[0772]

		MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDHENRFFGKLLNNMSQPINDL
HMPREF1990_01800 [Porphyromonas gingivalis W4087]	ERJ87335	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPRSMGFISVHDLRKLLLMELLCEGSFSRMQSDFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMDQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLQKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRHQFRAIV AELRLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKVMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVRDKKRELRTAGKPVPPDLAAYIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKIMTDREEDILPGLKNIDSILDKENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEIPLIYRDVSAKVGSIEGSSAKDLPEG SSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL
M573_117042 [Prevotella intermedia ZT]	KJJ86756	MKMEDDKKTTESTNMLDNKHFWAAFLNLARHNVYITVNHINK VLELKNKKDQDIIIDNDQDILAIKTHWEKVNGDLNKTERLRELM TKHFPFLETAIYTKNKEDKEEVKQEKQAEAQSLESLKDCLFLFL EKLQEARNYYSHYKYSESTKEPMLEEGLLEKMYNIFDDNIQLVI KDYQHNKDINPDEDFKHLDRKGQFKYSFADNEGNITESGLLFF VSLFLEKKDAIWMQQKLTGFKDNRESKKKMTHEVFCRRRMLL PKLRLESTQTQDWILLDMLNELIRCPKSLYERLQGEYRKKFNVP FDSADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNLR FQIDLGTYHFSIYKKLIGGQKEDRHLTHKLYGFERIQEFAKQNRP DEWKALVKDLDTYETSNERYISETTPHYHLENQKIGIRFRNGNK EIWPSLKTNGENNEKSKYKLDKPYQAEAFLSVHELLPMMFYYL LLKKEEPNNDKKNASIVEGFIKREIRDMYKLYDAFANGEINNIG DLEKYCEDKGIPKRHLPKQMVAILYDEPKDMVKEAKRKQKEM VKDTKKLLATLEKQTQEEIEDGGRNIRLLKSGEIARWLVNDMM RFQPVQKDNEGNPLNNSKANSTEYQMLQRSLALYNKEEKPTRY FRQVNLINSSNPHPFLKWTKWEECNNILSFYRNYLTKKIEFLNK

[0773]

		LKPEDWEKNQYFLKLKEPKTNRETLVQGWKNGFNLPRGIFTEPI REWFKRHQNDSKEYEKVEALKRVGLVTKVIPLFFKEEYFKEDA QKEINNCVQPFYSFPYNVGNIHKPDEKDFLPSEERKKLWGDKK DKFKGYKAKVKSKKLTDKEKEEYRSYLEFQSWNKFERELRLV RNQDIVTWLLCTELIDKMKVEGLNVEELQKLRLKDIDTDTAKQ EKNNILNRIMPMQLPVTVYEIDDSHNIVKDRPLHTVYIEETKTKL LKQGNFKALVKDRRLNGLFSFVDTSSKAELKDKPISKSVVEYEL GEYQNARIETIKDMLLLEKTLIKKYEKLPTDNFSDMLNGWLEG KDESDKARFQNDVKLLVAVRNAFSHNQYPMRNRIAFANINPFS LSSADISEEKKLDIANQLKDKTHKIIKKIIEIEKPIETKE
Prevotella buccae	WP_00434358 1	MQKQDKLFVDRKKNAIFAFPKYITIMENQEKPEPIYYELTDKHF WAAFLNLARHNVYTTINHINRRLEIAELKDDGYMMDIKGSWNE QAKKLDKKVRLRDLIMKHFPFLEAAAYEITNSKSPNNKEQREK EQSEALSLNNLKNVLFIFLEKLQVLRNYYSHYKYSEESPKPIFET SLLKNMYKVFDANVRLVKRDYMHHENIDMQRDFTHLNRKKQ VGRTKNIIDSPNFHYHFADKEGNMTIAGLLFFVSLFLDKKDAIW MQKKLKGFKDGRNLREQMTNEVFCRSRISLPKLKLENVQTKD WMQLDMLNELVRCPKSLYERLREKDRESFKVPFDIFSDDYDAE EEPFKNTLVRHQDRFPYFVLRYFDLNEIFEQLRFQIDLGTYHFSI YNKRIGDEDEVRHLTHHLYGFARIQDFAQQNQPEVWRKLVKD LDYFEASQEPYIPKTAPHYHLENEKIGIKFCSTHNNLFPSLKTEK TCNGRSKFNLGTQFTAEAFLSVHELLPMMFYYLLLTKDYSRKE SADKVEGIIRKEISNIYAIYDAFANGEINSIADLTCRLQKTNILQG HLPKQMISILEGRQKDMEKEAERKIGEMIDDTQRRLDLLCKQTN QKIRIGKRNAGLLKSGKIADWLVNDMMRFQPVQKDQNNIPINN SKANSTEYRMLQRALALFGSENFRLKAYFNQMNLVGNDNPHP FLAETQWEHQTNILSFYRNYLEARKKYLKGLKPQNWKQYQHF LILKVQKTNRNTLVTGWKNSFNLPRGIFTQPIREWFEKHNNSKR IYDQILSFDRVGFVAKAIPLYFAEEYKDNVQPFYDYPFNIGNKL KPQKGQFLDKKERVELWQKNKELFKNYPSEKKKTDLAYLDFL SWKKFERELRLIKNQDIVTWLMFKELFNMATVEGLKIGEIHLRD IDTNTANEESNNILNRIMPMKLPVKTYETDNKGNILKERPLATF YIEETETKVLKQGNFKVLAKDRRLNGLLSFAETTDIDLEKNPITK LSVDHELIKYQTTRISIFEMTLGLEKKLINKYPTLPTDSFRNMLE RWLQCKANRPELKNYVNSLIAVRNAFSHNQYPMYDATLFAEV KKFTLFPSVDTKKIELNIAPQLLEIVGKAIKEIEKSENKN
Porphyromonas gingivalis	WP_00587351 1	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR

[0774]

		MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPQSMGFISVHNLRKLLLMELLCEGSFSRMQSDFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMNQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLRKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRRQFRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDEENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL
Porphyromonas gingivalis	WP_005874195	MTEQNEKPYNGTYYTLEDKHFWAAFFNLARHNAYITLAHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSF LEGAAYGKKLFESQSSGNKSSKKKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHPESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDKYGNNDN PFFKHHFVDREEKVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKG GTEAYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNELV RCPKSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTLV RHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQ PEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDK PYITQTTPHYHIEKGKIGLRFVPEGQLLWPSPEVGATRTGRSKY AQDKRFTAEAFLSVHELMPMMFYYFLLREKYSEEASAEKVQG RIKRVIEDVYAVYDAFARDEINTRDELDACLADKGIRRGHLPRQ MIAILSQEHKDMEEKVRKKLQEMIADTDHRLDMLDRQTDRKIR IGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSKA NSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLHE TRWESHTNILSFYRSYLKARKAFLQSIGRSDREENHRFLLLKEPK TDRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEV GFMAKAVPLYFERACKDRVQPFYDYPFNVGNSLKPKKGRFLSK EKRAEEWESGKERFRDLEAWSHSAARRIEDAFVGIEYASWENK KKIEQLLQDLSLWETFESKLKVKADKINIAKLKKEILEAKEHPY HDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLDTG TLYLKDIRTDVQEQGSLNVLNHVKPMRLPVVVYRADSRGHVH KEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGA LAMEQYPISKLRVEYELAKYQTARVCAFEQTLELEESLLTRYPH LPDESFREMLESWSDPLLDKWPDLQREVRLLIAVRNAFSHNQY PMYDETIFSSIRKYDPSSLDAIEERMGLNIAHRLSEEVKLAKEMV ERIIQA

[0775]

Prevotella pallens	WP_006044833	MKEEEKGKTPVVSTYNKDDKHFWAAFLNLARHNVYITVNHIN KILGEGEINRDGYENTLEKSWNEIKDINKKDRLSKLIIKHFPFLE VTTYQRNSADTTKQKEEKQAEAQSLESLKKSFFVFIYKLRDLRN HYSHYKHSKSLERPKFEEDLQEKMYNIFDASIQLVKEDYKHNT DIKTEEDFKHLDRKGQFKYSFADNEGNITESGLLFFVSLFLEKK DAIWVQKKLEGFKCSNESYQKMTNEVFCRSRMLLPKLRLQSTQ TQDWILLDMLNELIRCPKSLYERLREEDRKKFRVPIEIADEDYD AEQEPFKNALVRHQDRFPYFALRYFDYNEIFTNLRFQIDLGTYH FSIYKKQIGDYKESHHLTHKLYGFERIQEFTKQNRPDEWRKFVK TFNSFETSKEPYIPETTPHYHLENQKIGIRFRNDNDKIWPSLKTNS EKNEKSKYKLDKSFQAEAFLSVHELLPMMFYYLLLKTENTDND NEIETKKKENKNDKQEKHKIEEIIENKITEIYALYDAFANGKINSI DKLEEYCKGKDIEIGHLPKQMIAILKSEHKDMATEAKRKQEEM LADVQKSLESLDNQINEEIENVERKNSSLKSGEIASWLVNDMM RFQPVQKDNEGNPLNNSKANSTEYQMLQRSLALYNKEEKPTRY FRQVNLIESSNPHPFLNNTEWEKCNNILSFYRSYLEAKKNFLESL KPEDWEKNQYFLMLKEPKTNCETLVQGWKNGFNLPRGIFTEPI RKWFMEHRKNITVAELKRVGLVAKVIPLFFSEEYKDSVQPFYN YLFNVGNINKPDEKNFLNCEERRELLRKKKDEFKKMTDKEKEE NPSYLEFQSWNKFERELRLVRNQDIVTWLLCMELFNKKKIKEL NVEKIYLKNINTNTTKKEKNTEEKNGEEKIIKEKNNILNRIMPMR LPIKVYGRENFSKNKKKKIRRNTFFTVYIEEKGTKLLKQGNFKA LERDRRLGGLFSFVKTHSKAESKSNTISKSRVEYELGEYQKARIE IIKDMLALEETLIDKYNSLDTDNFHNMLTGWLKLKDEPDKASF QNDVDLLIAVRNAFSHNQYPMRNRIAFANINPFSLSSANTSEEK GLGIANQLKDKTHKTIEKIIEIEKPIETKE
Myroides odoratimimus	WP_006261414	MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEE VNKRNTFGKLAKRDNGNLKNYIIHVFKDELSISDFEKRVAIFAS YFPILETVDKKSIKERNRTIDLTLSQRIRQFREMLISLVTAVDQLR NFYTHYHHSEIVIENKVLDFLNSSLVSTALHVKDKYLKTDKTKE FLKETIAAELDILIEAYKKKQIEKKNTRFKANKREDILNAIYNEA FWSFINDKDKDKETVVAKGADAYFEKNHHKSNDPDFALNISEK GIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIKYMATQ RIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVPNVVYQH LSTTQQNSFIEDWNEYYKDYEDDVETDDLSRVIHPVIRKRYEDR FNYFAIRFLDEFFDFPTLRFQVHLGDYVHDRRTKQLGKVESDRII KEKVTVFARLKDINSAKANYFHSLEEQDKEELDNKWTLFPNPS YDFPKEHTLQHQGEQKNAGKIGIYVKLRDTQYKEKAALEEARK SLNPKERSATKASKYDIITQIIEANDNVKSEKPLVFTGQPIAYLS MNDIHSMLFSLLTDNAELKKTPEEVEAKLIDQIGKQINEILSKDT DTKILKKYKDNDLKETDTDKITRDLARDKEEIEKLILEQKQRAD DYNYTSSTKFNIDKSRKRKHLLFNAEKGKIGVWLANDIKRFMT EEFKSKWKGYQHTELQKLFAYYDTSKSDLDLILSDMVMVKDY PIELIALVKKSRTLVDFLNKYLEARLGYMENVITRVKNSIGTPQF KTVRKECFTFLKKSNYTVVSLDKQVERILSMPLFIERGFMDDKP TMLEGKSYQQHKEKFADWFVHYKENSNYQNFYDTEVYEITTE

[0776]

		DKREKAKVTKKIKQQQKNDVFTLMMVNYMLEEVLKLSSNDRL SLNELYQTKEERIVNKQVAKDTQERNKNYIWNKVVDLQLCEG LVRIDKVKLKDIGNFRKYENDSRVKEFLTYQSDIVWSAYLSNEV DSNKLYVIERQLDNYESIRSKELLKEVQEIECSVYNQVANKESL KQSGNENFKQYVLQGLVPIGMDVREMLILSTDVKFIKEEIIQLG QAGEVEQDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISDNEY YAEYYMEIFRSIKEKYTS
Myroides odoratimimus	WP_006265509	MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEE VNKRNTFGKLAKRDNGNLKNYIIHVFKDELSISDFEKRVAIFAS YFPILETVDKKSIKERNRTIDLTLSQRIRQFREMLISLVTAVDQLR NFYTHYHHSEIVIENKVLDFLNSSLVSTALHVKDKYLKTDKTKE FLKETIAAELDILIEAYKKKQIEKKNTRFKANKREDILNAIYNEA FWSFINDKDKDKETVVAKGADAYFEKNHHKSNDPDFALNISEK GIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIKYMATQ RIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVPNVVYQH LSTTQQNSFIEDWNEYYKDYEDDVETDDLSRVIHPVIRKRYEDR FNYFAIRFLDEFFDFPTLRFQVHLGDYVHDRRTKQLGKVESDRII KEKVTVFARLKDINSAKASYFHSLEEQDKEELDNKWTLFPNPS YDFPKEHTLQHQGEQKNAGKIGIYVKLRDTQYKEKAALEEARK SLNPKERSATKASKYDIITQIIEANDNVKSEKPLVFTGQPIAYLS MNDIHSMLFSLLTDNAELKKTPEEVEAKLIDQIGKQINEILSKDT DTKILKKYKDNDLKETDTDKITRDLARDKEEIEKLILEQKQRAD DYNYTSSTKFNIDKSRKRKHLLFNAEKGKIGVWLANDIKRFMF KESKSKWKGYQHTELQKLFAYFDTSKSDLELILSDMVMVKDYP IELIDLVRKSRTLVDFLNKYLEARLGYIENVITRVKNSIGTPQFKT VRKECFAFLKESNYTVASLDKQIERILSMPLFIERGFMDSKPTML EGKSYQQHKEDFADWFVHYKENSNYQNFYDTEVYEIITEDKRE QAKVTKKIKQQQKNDVFTLMMVNYMLEEVLKLPSNDRLSLNE LYQTKEERIVNKQVAKDTQERNKNYIWNKVVDLQLCEGLVRID KVKLKDIGNFRKYENDSRVKEFLTYQSDIVWSGYLSNEVDSNK LYVIERQLDNYESIRSKELLKEVQEIECIVYNQVANKESLKQSGN ENFKQYVLQGLLPRGTDVREMLILSTDVKFKKEEIMQLGQVRE VEQDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRPISDNEYYAEY YMEIFRSIKEKYAS
Prevotella sp. MSX73	WP_007412163	MQKQDKLFVDRKKNAIFAFPKYITIMENQEKPEPIYYELTDKHF WAAFLNLARHNVYTTINHINRRLEIAELKDDGYMMGIKGSWNE QAKKLDKKVRLRDLIMKHFPFLEAAAYEITNSKSPNNKEQREK EQSEALSLNNLKNVLFIFLEKLQVLRNYYSHYKYSEESPKPIFET SLLKNMYKVFDANVRLVKRDYMHHENIDMQRDFTHLNRKKQ VGRTKNIIDSPNFHYHFADKEGNMTIAGLLFFVSLFLDKKDAIW MQKKLKGFKDGRNLREQMTNEVFCRSRISLPKLKLENVQTKD WMQLDMLNELVRCPKSLYERLREKDRESFKVPFDIFSDDYDAE EEPFKNTLVRHQDRFPYFVLRYFDLNEIFEQLRFQIDLGTYHFSI YNKRIGDEDEVRHLTHHLYGFARIQDFAPQNQPEEWRKLVKDL DHFETSQEPYISKTAPHYHLENEKIGIKFCSTHNNLFPSLKREKT

[0777]

		CNGRSKFNLGTQFTAEAFLSVHELLPMMFYYLLLTKDYSRKES ADKVEGIIRKEISNIYAIYDAFANNEINSIADLTCRLQKTNILQGH LPKQMISILEGRQKDMEKEAERKIGEMIDDTQRRLDLLCKQTNQ KIRIGKRNAGLLKSGKIADWLVSDMMRFQPVQKDTNNAPINNS KANSTEYRMLQHALALFGSESSRLKAYFRQMNLVGNANPHPFL AETQWEHQTNILSFYRNYLEARKKYLKGLKPQNWKQYQHFLIL KVQKTNRNTLVTGWKNSFNLPRGIFTQPIREWFEKHNNSKRIYD QILSFDRVGFVAKAIPLYFAEEYKDNVQPFYDYPFNIGNKLKPQ KGQFLDKKERVELWQKNKELFKNYPSEKNKTDLAYLDFLSWK KFERELRLIKNQDIVTWLMFKELFKTTTVEGLKIGEIHLRDIDTN TANEESNNILNRIMPMKLPVKTYETDNKGNILKERPLATFYIEET ETKVLKQGNFKVLAKDRRLNGLLSFAETTDIDLEKNPITKLSVD YELIKYQTTRISIFEMTLGLEKKLIDKYSTLPTDSFRNMLERWLQ CKANRPELKNYVNSLIAVRNAFSHNQYPMYDATLFAEVKKFTL FPSVDTKKIELNIAPQLLEIVGKAIKEIEKSENKN
Porphyromonas gingivalis	WP_012458414	MTEQNERPYNGTYYTLEDKHFWAAFFNLARHNAYITLAHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSF LEGAAYGKKLFESQSSGNKSSKKKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHPESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRYGNNDN PFFKHHFVDREEKVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKG GTETYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNELV RCPKSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTLV RHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQ PEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDK PYITQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGATRTGRSKY AQDKRLTAEAFLSVHELMPMMFYYFLLREKYSDEASAERVQG RIKRVIEDVYAVYDAFARGEINTRDELDACLADKGIRRGHLPRQ MIGILSQEHKDMEEKVRKKLQEMIVDTDHRLDMLDRQTDRKIR IGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSKA NSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLHE TRWESHTNILSFYRSYLKARKAFLQSIGRSDRVENHRFLLLKEP KTDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEMGLDEVGSYKE VGFMAKAVPLYFERACKDRVQPFYDYPFNVGNSLKPKKGRFLS KEKRAEEWESGKERFRLAKLKKEILEAKEHPYLDFKSWQKFER ELRLVKNQDIITWMICRDLMEENKVEGLDTGTLYLKDIRTDVQ EQGNLNVLNRVKPMRLPVVVYRADSRGHVHKEQAPLATVYIE ERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAMEQYPISKL RVEYELAKYQTARVCAFEQTLELEESLLTRYPHLPDKNFRKML ESWSDPLLDKWPDLHGNVRLLIAVRNAFSHNQYPMYDEAVFSS IRKYDPSSPDAIEERMGLNIAHRLSEEVKQAKEMAERIIQA
Paludibacter propionicigenes	WP_013446107	MKTSANNIYFNGINSFKKIFDSKGAIAPIAEKSCRNFDIKAQNDV NKEQRIHYFAVGHTFKQLDTENLFEYVLDENLRAKRPTRFISLQ QFDKEFIENIKRLISDIRNINSHYIHRFDPLKIDAVPTNIIDFLKESF ELAVIQIYLKEKGINYLQFSENPHADQKLVAFLHDKFLPLDEKK

[0778]

		TSMLQNETPQLKEYKEYRKYFKTLSKQAAIDQLLFAEKETDYI WNLFDSHPVLTISAGKYLSFYSCLFLLSMFLYKSEANQLISKIKG FKKNTTEEEKSKREIFTFFSKRFNSMDIDSEENQLVKFRDLILYL NHYPVAWNKDLELDSSNPAMTDKLKSKIIELEINRSFPLYEGNE RFATFAKYQIWGKKHLGKSIEKEYINASFTDEEITAYTYETDTCP ELKDAHKKLADLKAAKGLFGKRKEKNESDIKKTETSIRELQHEP NPIKDKLIQRIEKNLLTVSYGRNQDRFMDFSARFLAEINYFGQD ASFKMYHFYATDEQNSELEKYELPKDKKKYDSLKFHQGKLVH FISYKEHLKRYESWDDAFVIENNAIQLKLSFDGVENTVTIQRAL LIYLLEDALRNIQNNTAENAGKQLLQEYYSHNKADLSAFKQILT QQDSIEPQQKTEFKKLLPRRLLNNYSPAINHLQTPHSSLPLILEK ALLAEKRYCSLVVKAKAEGNYDDFIKRNKGKQFKLQFIRKAW NLMYFRNSYLQNVQAAGHHKSFHIERDEFNDFSRYMFAFEELS QYKYYLNEMFEKKGFFENNEFKILFQSGTSLENLYEKTKQKFEI WLASNTAKTNKPDNYHLNNYEQQFSNQLFFINLSHFINYLKSTG KLQTDANGQIIYEALNNVQYLIPEYYYTDKPERSESKSGNKLYN KLKATKLEDALLYEMAMCYLKADKQIADKAKHPITKLLTSDVE FNITNKEGIQLYHLLVPFKKIDAFIGLKMHKEQQDKKHPTSFLA NIVNYLELVKNDKDIRKTYEAFSTNPVKRTLTYDDLAKIDGHLI SKSIKFTNVTLELERYFIFKESLIVKKGNNIDFKYIKGLRNYYNN EKKKNEGIRNKAFHFGIPDSKSYDQLIRDAEVMFIANEVKPTHA TKYTDLNKQLHTVCDKLMETVHNDYFSKEGDGKKKREAAGQ KYFENIISAK
Porphyromonas gingivalis	WP_013816155	MTEQNEKPYNGTYYTLEDKHFWAAFFNLARHNAYITLAHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSF LEGAAYGKKLFESQSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHPESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRYGNNDN PFFKHHFVDREGTVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKG GTETYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNELV RCPKSLYDRLREEDRARFRVPVDILSDEEDTDGAEEDPFKNTLV RHQDRFPYFALRYFDLKKVFTSLRFQIDLGTYHFAIYKKNIGEQ PEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDK PYITQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGATRTGRSKY AQDKRFTAEAFLSAHELMPMMFYYFLLREKYSEEASAERVQGR IKRVIEDVYAVYDAFARDEINTRDELDACLADKGIRRGHLPRQ MIGILSQEHKDMEEKIRKKLQEMMADTDHRLDMLDRQTDRKIR IGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSKA NSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLHE TRWESHTNILSFYRSYLKARKAFLQSIGRSDRVENHRFLLLKEP KTDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEMGLDEVGSYKE VGFMAKAVPLYFERACKDWVQPFYNYPFNVGNSLKPKKGRFL SKEKRAEEWESGKERFRLAKLKKEILEAKEHPYLDFKSWQKFE RELRLVKNQDIITWMICGDLMEENKVEGLDTGTLYLKDIRTDV QEQGSLNVLNRVKPMRLPVVVYRADSRGHVHKEQAPLATVYI EERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAMEQYPISK

[0779]

		LRVEYELAKYQTARVCAFEQTLELEESLLTRCPHLPDKNFRKM LESWSDPLLDKWPDLHRKVRLLIAVRNAFSHNQYPMYDEAVFS SIRKYDPSFPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA
Flavobacterium columnare	WP_014165541	MSSKNESYNKQKTFNHYKQEDKYFFGGFLNNADDNLRQVGKE FKTRINFNHNNNELASVFKDYFNKEKSVAKREHALNLLSNYFP VLERIQKHTNHNFEQTREIFELLLDTIKKLRDYYTHHYHKPITIN PKIYDFLDDTLLDVLITIKKKKVKNDTSRELLKEKLRPELTQLKN QKREELIKKGKKLLEENLENAVFNHCLRPFLEENKTDDKQNKT VSLRKYRKSKPNEETSITLTQSGLVFLMSFFLHRKEFQVFTSGLE GFKAKVNTIKEEEISLNKNNIVYMITHWSYSYYNFKGLKHRIKT DQGVSTLEQNNTTHSLTNTNTKEALLTQIVDYLSKVPNEIYETL SEKQQKEFEEDINEYMRENPENEDSTFSSIVSHKVIRKRYENKFN YFAMRFLDEYAELPTLRFMVNFGDYIKDRQKKILESIQFDSERII KKEIHLFEKLSLVTEYKKNVYLKETSNIDLSRFPLFPNPSYVMA NNNIPFYIDSRSNNLDEYLNQKKKAQSQNKKRNLTFEKYNKEQ SKDAIIAMLQKEIGVKDLQQRSTIGLLSCNELPSMLYEVIVKDIK GAELENKIAQKIREQYQSIRDFTLDSPQKDNIPTTLIKTINTDSSV TFENQPIDIPRLKNAIQKELTLTQEKLLNVKEHEIEVDNYNRNKN TYKFKNQPKNKVDDKKLQRKYVFYRNEIRQEANWLASDLIHF MKNKSLWKGYMHNELQSFLAFFEDKKNDCIALLETVFNLKED CILTKGLKNLFLKHGNFIDFYKEYLKLKEDFLNTESTFLENGLIG LPPKILKKELSKRFKYIFIVFQKRQFIIKELEEKKNNLYADAINLS RGIFDEKPTMIPFKKPNPDEFASWFVASYQYNNYQSFYELTPDI VERDKKKKYKNLRAINKVKIQDYYLKLMVDTLYQDLFNQPLD KSLSDFYVSKAEREKIKADAKAYQKRNDSSLWNKVIHLSLQNN RITANPKLKDIGKYKRALQDEKIATLLTYDDRTWTYALQKPEK ENENDYKELHYTALNMELQEYEKVRSKELLKQVQELEKQILEE YTDFLSTQIHPADFEREGNPNFKKYLAHSILENEDDLDKLPEKV EAMRELDETITNPIIKKAIVLIIIRNKMAHNQYPPKFIYDLANRFV PKKEEEYFATYFNRVFETITKELWENKEKKDKTQV
Psychroflexus torquis	WP_015024765	MESIIGLGLSFNPYKTADKHYFGSFLNLVENNLNAVFAEFKERIS YKAKDENISSLIEKHFIDNMSIVDYEKKISILNGYLPIIDFLDDELE NNLNTRVKNFKKNFIILAEAIEKLRDYYTHFYHDPITFEDNKEPL LELLDEVLLKTILDVKKKYLKTDKTKEILKDSLREEMDLLVIRK TDELREKKKTNPKIQHTDSSQIKNSIFNDAFQGLLYEDKGNNKK TQVSHRAKTRLNPKDIHKQEERDFEIPLSTSGLVFLMSLFLSKKE IEDFKSNIKGFKGKVVKDENHNSLKYMATHRVYSILAFKGLKY RIKTDTFSKETLMMQMIDELSKVPDCVYQNLSETKQKDFIEDW NEYFKDNEENTENLENSRVVHPVIRKRYEDKFNYFAIRFLDEFA NFKTLKFQVFMGYYIHDQRTKTIGTTNITTERTVKEKINVFGKL SKMDNLKKHFFSQLSDDENTDWEFFPNPSYNFLTQADNSPANN IPIYLELKNQQIIKEKDAIKAEVNQTQNRNPNKPSKRDLLNKILK TYEDFHQGDPTAILSLNEIPALLHLFLVKPNNKTGQQIENIIRIKIE KQFKAINHPSKNNKGIPKSLFADTNVRVNAIKLKKDLEAELDM LNKKHIAFKENQKASSNYDKLLKEHQFTPKNKRPELRKYVFYK

[0780]

		SEKGEEATWLANDIKRFMPKDFKTKWKGCQHSELQRKLAFYD RHTKQDIKELLSGCEFDHSLLDINAYFQKDNFEDFFSKYLENRIE TLEGVLKKLHDFKNEPTPLKGVFKNCFKFLKRQNYVTESPEIIK KRILAKPTFLPRGVFDERPTMKKGKNPLKDKNEFAEWFVEYLE NKDYQKFYNAEEYRMRDADFKKNAVIKKQKLKDFYTLQMVN YLLKEVFGKDEMNLQLSELFQTRQERLKLQGIAKKQMNKETG DSSENTRNQTYIWNKDVPVSFFNGKVTIDKVKLKNIGKYKRYE RDERVKTFIGYEVDEKWMMYLPHNWKDRYSVKPINVIDLQIQE YEEIRSHELLKEIQNLEQYIYDHTTDKNILLQDGNPNFKMYVLN GLLIGIKQVNIPDFIVLKQNTNFDKIDFTGIASCSELEKKTIILIAIR NKFAHNQLPNKMIYDLANEFLKIEKNETYANYYLKVLKKMISD LA
Riemerella anatipestifer	WP_015345620	MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYKAHFTVNLTRDF AHLNRKGKNKQDNPDFNRYRFEKDGFFTESGLLFFTNLFLDKR DAYWMLKKVSGFKASHKQREKMTTEVFCRSRILLPKLRLESRY DHNQMLLDMLSELSRCPKLLYEKLSEENKKHFQVEADGFLDEI EEEQNPFKDTLIRHQDRFPYFALRYLDLNESFKSIRFQVDLGTYH YCIYDKKIGDEQEKRHLTRTLLSFGRLQDFTEINRPQEWKALTK DLDYKETSNQPFISKTTPHYHITDNKIGFRLGTSKELYPSLEIKDG ANRIAKYPYNSGFVAHAFISVHELLPLMFYQHLTGKSEDLLKET VRHIQRIYKDFEEERINTIEDLEKANQGRLPLGAFPKQMLGLLQ NKQPDLSEKAKIKIEKLIAETKLLSHRLNTKLKSSPKLGKRREKL IKTGVLADWLVKDFMRFQPVAYDAQNQPIKSSKANSTEFWFIR RALALYGGEKNRLEGYFKQTNLIGNTNPHPFLNKFNWKACRNL VDFYQQYLEQREKFLEAIKHQPWEPYQYCLLLKVPKENRKNLV KGWEQGGISLPRGLFTEAIRETLSKDLTLSKPIRKEIKKHGRVGFI SRAITLYFKEKYQDKHQSFYNLSYKLEAKAPLLKKEEHYEYWQ QNKPQSPTESQRLELHTSDRWKDYLLYKRWQHLEKKLRLYRN QDIMLWLMTLELTKNHFKELNLNYHQLKLENLAVNVQEADAK LNPLNQTLPMVLPVKVYPTTAFGEVQYHETPIRTVYIREEQTKA LKMGNFKALVKDRRLNGLFSFIKEENDTQKHPISQLRLRRELEI YQSLRVDAFKETLSLEEKLLNKHASLSSLENEFRTLLEEWKKKY AASSMVTDKHIAFIASVRNAFCHNQYPFYKETLHAPILLFTVAQ PTTEEKDGLGIAEALLKVLREYCEIVKSQI
Prevotella pleuritidis	WP_021584635	MENDKRLEESACYTLNDKHFWAAFLNLARHNVYITVNHINKTL ELKNKKNQEIIIDNDQDILAIKTHWAKVNGDLNKTDRLRELMIK HFPFLEAAIYSNNKEDKEEVKEEKQAKAQSFKSLKDCLFLFLEK LQEARNYYSHYKYSESSKEPEFEEGLLEKMYNTFDASIRLVKED YQYNKDIDPEKDFKHLERKEDFNYLFTDKDNKGKITKNGLLFF VSLFLEKKDAIWMQQKFRGFKDNRGNKEKMTHEVFCRSRMLL PKIRLESTQTQDWILLDMLNELIRCPKSLYERLQGAYREKFKVP FDSIDEDYDAEQEPFRNTLVRHQDRFPYFALRYFDYNEIFKNLR FQIDLGTYHFSIYKKLIGGKKEDRHLTHKLYGFERIQEFTKQNRP DKWQAIIKDLDTYETSNERYISETTPHYHLENQKIGIRFRNDNN DIWPSLKTNGEKNEKSKYNLDKPYQAEAFLSVHELLPMMFYYL

[0781]

		LLKMENTDNDKEDNEVGTKKKGNKNNKQEKHKIEEIIENKIKDI YALYDAFTNGEINSIDELAEQREGKDIEIGHLPKQLIVILKNKSK DMAEKANRKQKEMIKDTKKRLATLDKQVKGEIEDGGRNIRLL KSGEIARWLVNDMMRFQPVQKDNEGKPLNNSKANSTEYQMLQ RSLALYNKEEKPTRYFRQVNLIKSSNPHPFLEDTKWEECYNILSF YRNYLKAKIKFLNKLKPEDWKKNQYFLMLKEPKTNRKTLVQG WKNGFNLPRGIFTEPIKEWFKRHQNDSEEYKKVEALDRVGLVA KVIPLFFKEEYFKEDAQKEINNCVQPFYSFPYNVGNIHKPEEKNF LHCEERRKLWDKKKDKFKGYKAKEKSKKMTDKEKEEHRSYLE FQSWNKFERELRLVRNQDILTWLLCTKLIDKLKIDELNIEELQKL RLKDIDTDTAKKEKNNILNRVMPMRLPVTVYEIDKSFNIVKDKP LHTVYIEETGTKLLKQGNFKALVKDRRLNGLFSFVKTSSEAESK SKPISKLRVEYELGAYQKARIDIIKDMLALEKTLIDNDENLPTNK FSDMLKSWLKGKGEANKARLQNDVGLLVAVRNAFSHNQYPM YNSEVFKGMKLLSLSSDIPEKEGLGIAKQLKDKIKETIERIIEIEKE IRN
Porphyromonas gingivalis	WP_02166319 7	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPRSMGFISVHDLRKLLLMELLCEGSFSRMQSDFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMDQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLQKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRHQFRAIV AELRLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDEENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL
Porphyromonas gingivalis	WP_02166547 5	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL

[0782]

		DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTNENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPQSMGFISVHDLRKLLLMELLCEGSFSRMQSGFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMNQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLRKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRRQFRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDKENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDHENRFFGKLLNNMSQPINDL
Porphyromonas gingivalis	WP_021677657	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPQSMGFISVHDLRKLLLMELLCEGSFSRMQSGFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMNQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLRKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRRQFRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDEENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI

[0783]

		MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDHENRFFGKLLNNMSQPINDL
Porphyromonas gingivalis	WP_021680012	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPRSMGFISVHDLRKLLLMELLCEGSFSRMQSDFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMDQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLQKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRHQFRAIV AELRLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKVMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVRDKKRELRTAGKPVPPDLAAYIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKIMTDREEDILPGLKNIDSILDKENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEIPLIYRDVSAKVGSIEGSSAKDLPEG SSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL
Porphyromonas gingivalis	WP_021384676 7	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPRSMGFISVHDLRKLLLMELLCEGSFSRMQSDFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMNQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLRKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRRQFRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK

[0784]

		TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDEENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL
Prevotella falsenii	WP_03688492 9	MKNDNNSTKSTDYTLGDKHFWAAFLNLARHNVYITVNHINKV LELKNKKDQEIIIDNDQDILAIKTLWGKVDTDINKKDRLRELIM KHFPFLEAATYQQSSTNNTKQKEEEQAKAQSFESLKDCLFLFLE KLREARNYYSHYKHSKSLEEPKLEEKLLENMYNIFDTNVQLVIK DYEHNKDINPEEDFKHLGRAEGEFNYYFTRNKKGNITESGLLFF VSLFLEKKDAIWAQTKIKGFKDNRENKQKMTHEVFCRSRMLLP KLRLESTQTQDWILLDMLNELIRCPKSLYKRLQGEKREKFRVPF DPADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNLRF QIDLGTYHFSIYKKQIGDKKEDRHLTHKLYGFERIQEFAKENRP DEWKALVKDLDTFEESNEPYISETTPHYHLENQKIGIRNKNKKK KKTIWPSLETKTTVNERSKYNLGKSFKAEAFLSVHELLPMMFY YLLLNKEEPNNGKINASKVEGIIEKKIRDIYKLYGAFANEEINNE EELKEYCEGKDIAIRHLPKQMIAILKNEYKDMAKKAEDKQKKM IKDTKKRLAALDKQVKGEVEDGGRNIKPLKSGRIASWLVNDM MRFQPVQRDRDGYPLNNSKANSTEYQLLQRTLALFGSERERLA PYFRQMNLIGKDNPHPFLKDTKWKEHNNILSFYRSYLEAKKNF LGSLKPEDWKKNQYFLKLKEPKTNRETLVQGWKNGFNLPRGIF TEPIREWFIRHQNESEEYKKVKDFDRIGLVAKVIPLFFKEDYQKE IEDYVQPFYGYPFNVGNIHNSQEGTFLNKKEREELWKGNKTKF KDYKTKEKNKEKTNKDKFKKKTDEEKEEFRSYLDFQSWKKFE RELRLVRNQDIVTWLLCMELIDKLKIDELNIEELQKLRLKDIDTD TAKKEKNNILNRIMPMELPVTVYETDDSNNIIKDKPLHTIYIKEA ETKLLKQGNFKALVKDRRLNGLFSFVETSSEAELKSKPISKSLVE YELGEYQRARVEIIKDMLRLEETLIGNDEKLPTNKFRQMLDKW LEHKKETDDTDLKNDVKLLTEVRNAFSHNQYPMRDRIAFANIK PFSLSSANTSNEEGLGIAKKLKDKTKETIDRIIEIEEQTATKR
Prevotella pleuritidis	WP_03693148 5	MENDKRLEESTCYTLNDKHFWAAFLNLARHNVYITINHINKLL EIRQIDNDEKVLDIKALWQKVDKDINQKARLRELMIKHFPFLEA AIYSNNKEDKEEVKEEKQAKAQSFKSLKDCLFLFLEKLQEARN YYSHYKSSESSKEPEFEEGLLEKMYNTFGVSIRLVKEDYQYNKD IDPEKDFKHLERKEDFNYLFTDKDNKGKITKNGLLFFVSLFLEK KDAIWMQQKLRGFKDNRGNKEKMTHEVFCRSRMLLPKIRLES TQTQDWILLDMLNELIRCPKSLYERLQGAYREKFKVPFDSIDED YDAEQEPFRNTLVRHQDRFPYFALRYFDYNEIFKNLRFQIDLGT YHFSIYKKLIGDNKEDRHLTHKLYGFERIQEFAKQKRPNEWQA

[0785]

		LVKDLDIYETSNEQYISETTPHYHLENQKIGIRFKNKKDKIWPSL ETNGKENEKSKYNLDKSFQAEAFLSIHELLPMMFYDLLLKKEEP NNDEKNASIVEGFIKKEIKRMYAIYDAFANEEINSKEGLEEYCK NKGFQERHLPKQMIAILTNKSKNMAEKAKRKQKEMIKDTKKR LATLDKQVKGEIEDGGRNIRLLKSGEIARWLVNDMMRFQSVQK DKEGKPLNNSKANSTEYQMLQRSLALYNKEQKPTPYFIQVNLI KSSNPHPFLEETKWEECNNILSFYRSYLEAKKNFLESLKPEDWK KNQYFLMLKEPKTNRKTLVQGWKNGFNLPRGIFTEPIKEWFKR HQNDSEEYKKVEALDRVGLVAKVIPLFFKEEYFKEDAQKEINN CVQPFYSFPYNVGNIHKPEEKNFLHCEERRKLWDKKKDKFKGY KAKEKSKKMTDKEKEEHRSYLEFQSWNKFERELRLVRNQDIVT WLLCTELIDKLKIDELNIEELQKLRLKDIDTDTAKKEKNNILNRI MPMQLPVTVYEIDKSFNIVKDKPLHTIYIEETGTKLLKQGNFKA LVKDRRLNGLFSFVKTSSEAESKSKPISKLRVEYELGAYQKARI DIIKDMLALEKTLIDNDENLPTNKFSDMLKSWLKGKGEANKAR LQNDVDLLVAIRNAFSHNQYPMYNSEVFKGMKLLSLSSDIPEKE GLGIAKQLKDKIKETIERIIEIEKEIRN
[Porphyromonas gingivalis	WP_039417390	MTEQNERPYNGTYYTLEDKHFWAAFFNLARHNAYITLAHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSF LEGAAYGKKLFESQSSGNKSSKKKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHPESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRYGNNDN PFFKHHFVDREGTVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKG GTEAYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNELV RCPKSLYDRLREEDRARFRVPIDILSDEDDTDGTEEDPFKNTLVR HQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQPE DRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDKPY ITQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGATRTGRSKYAQ DKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKVQGRIK RVIEDVYAVYDAFARGEIDTLDRLDACLADKGIRRGHLPRQMI AILSQEHKDMEEKVRKKLQEMIADTDHRLDMLDRQTDRKIRIG RKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSKANS TEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLHETR WESHTNILSFYRSYLKARKAFLQSIGRSDREENHRFLLLKEPKT DRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEVG FMAKAVPLYFERACKDRVQPFYDYPFNVGNSLKPKKGRFLSKE KRAEEWESGKERFRLAKLKKEILEAKEHPYLDFKSWQKFEREL RLVKNQDIITWMMCRDLMEENKVEGLDTGTLYLKDIRTDVHE QGSLNVLNRVKPMRLPVVVYRADSRGHVHKEQAPLATVYIEE RDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAMEQYPISKLR VEYELAKYQTARVCAFEQTLELEESLLTRYPHLPDKNFRKMLE SWSDPLLDKWPDLHRKVRLLIAVRNAFSHNQYPMYDEAVFSSI RKYDPSSPDAIEERMGLNIAHRLSEEVKQAKEMAERIIQV
Porphyromonas gulae	WP_039418910	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDQDVLSFKALWKNLDNDLERKSRLRSLILKHFSF

[0786]

	2	LEGAAYGKKLFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSGSSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRYGHNDN PSFKHHFVDSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTETYQQMTNEVFCRSRISLPKLKLESLRMDDWMLLDMLNE LVRCPKPLYDRLREDDRACFRVPVDILPDEDDTDGGGEDPFKN TLVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMI GEQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFET GDKPYISQTSPHYHIEKGKIGLRFMPEGQHLWPSPEVGTTRTGR SKYAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKV QGRIKRVIEDVYAIYDAFARDEINTLKELDACLADKGIRRGHLP KQMIAILSQEHKNMEEKVRKKLQEMIADTDHRLDMLDRQTDR KIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDASGKPLNNS KANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFL HDTRWESHTNILSFYRSYLRARKAFLERIGRSDRMENRPFLLLK EPKTDRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSY REVGFMAKAVPLYFERACEDRVQPFYDSPFNVGNSLKPKKGRF LSKEERAEEWERGKERFRDLEAWSHSAARRIEDAFAGIEYASPG NKKKIEQLLRDLSLWEAFESKLKVRADKINLAKLKKEILEAQEH PYHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLD TGTLYLKDIRTNVQEQGSLNVLNHVKPMRLPVVVYRADSRGH VHKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDT GGLAMEQYPISKLRVEYELAKYQTARVCAFEQTLELEESLLTRY PHLPDKNFRKMLESWSDPLLAKWPELHGKVRLLIAVRNAFSHN QYPMYDEAVFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAK ETVERIIQA
Porphyromonas gulae	WP_039419792	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDQDVLSFKALWKNLDNDLERKSRLRSLILKHFSF LEGAAYGKKLFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSGSSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRYGHNDN PSFKHHFVDGEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTETYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKPLYDRLREKDRARFRVPVDILPDEDDTDGGGEDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKVIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETG DKPYISQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGTTRTGRSK YAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKVQ GRIKRVIEDVYAIYDAFARDEINTRDELDACLADKGIRRGHLPK QMIGILSQEHKNMEEKVRKKLQEMIADTDHRLDMLDRQTDRKI RIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSK ANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLD ETRWESHTNILSFYRSYLRARKAFLERIGRSDRVENRPFLLLKEP KTDRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKE VGFMAKAVPLYFERACKDRVQPFYDSPFNVGNSLKPKKGRFLS KEKRAEEWESGKERFRLAKLKKEILEAQEHPYHDFKSWQKFER

[0787]

		ELRLVKNQDIITWMMCRDLMEENKVEGLDTGTLYLKDIRPNVQ EQGSLNVLNRVKPMRLPVVVYRADSRGHVHKEEAPLATVYIEE RDTKLLKQGNFKSFVKDRRLNGLFSFVDTGGLAMEQYPISKLR VEYELAKYQTARVCVFELTLRLEESLLSRYPHLPDESFREMLES WSDPLLAKWPELHGKVRLLIAVRNAFSHNQYPMYDEAVFSSIR KYDPSSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA
Porphyromonas gulae	WP_03942617 6	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDQDVLSFKALWKNFDNDLERKSRLRSLILKHFSF LEGAAYGKKLFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSGSSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHYHFNHLVRKGKKDRYGHNDN PSFKHHFVDSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTGPYEQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKPLYDRLREKDRACFRVPVDILPDEDDTDGGGEDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETG DKPYISQTTPHYHIEKGKIGLRFMPEGQHLWPSPEVGTTRTGRS KYAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKV QGRIKRVIKDVYAIYDAFARDEINTLKELDACSADKGIRRGHLP KQMIGILSQEHKNMEEKVRKKLQEMIADTDHRLDMLDRQTDR KIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNS KANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFL DETRWESHTNILSFYRSYLRARKAFLERIGRSDRVENRPFLLLKE PKNDRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYK EVGFMAKAVPLYFERACKDRVQPFYDSPFNVGNSLKPKKGRFL SKEKRAEEWESGKERFRLAKLKKEILEAKEHPYHDFKSWQKFE RELRLVKNQDIITWMMCRDLMEENKVEGLDTGTLYLKDIRTDV HEQGSLNVLNRVKPMRLPVVVYRADSRGHVHKEQAPLATVYI EERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGGLAMEQYPISK LRVEYELAKYQTARVCAFEQTLELEESLLTRYPHLPDENFREML ESWSDPLLGKWPDLHGKVRLLIAVRNAFSHNQYPMYDEAVFSS IRKYDPSSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA
Porphyromonas gulae	WP_03943177 8	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDQDVLSFKALWKNFDNDLERKSRLRSLILKHFSF LEGAAYGKKLFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRYGHNDN PSFKHHFVDGEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTETYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKPLYDRLREDDRACFRVPVDILPDEDDTDGGGEDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETG DKPYISQTSPHYHIEKGKIGLRFMPEGQHLWPSPEVGTTRTGRS KYAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKV QGRIKRVIEDVYAIYDAFARDEINTLKELDACLADKGIRRGHLP

[0788]

		KQMIAILSQEHKDMEEKIRKKLQEMIADTDHRLDMLDRQTDRK IRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSK ANSTEYRMLQRALALFGGEKKRLTPYFRQMNLTGGNNPHPFLH ETRWESHTNILSFYRSYLRARKAFLERIGRSDRMENRPFLLLKEP KTDRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYRE VGFMAKAVPLYFERACEDRVQPFYDSPFNVGNSLKPKKGRFLS KEERAEEWERGKERFRDLEAWSHSAARRIEDAFAGIEYASPGN KKKIEQLLRDLSLWEAFESKLKVRADKINLAKLKKEILEAQEHP YHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLDT GTLYLKDIRPNVQEQGSLNVLNRVKPMRLPVVVYRADSRGHV HKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTG GLAMEQYPISKLRVEYELAKYQTARVCVFELTLRLEESLLTRYP HLPDESFRKMLESWSDPLLAKWPELHGKVRLLIAVRNAFSHNQ YPMYDEAVFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAKE TVERIIQV
Porphyromonas gulae	WP_03943719	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKSRLRSLILKHFSF LEGAAYGKKFFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSGSSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDEVDPHYHFNHLVRKGKKDRYGHNDN PSFKHHFVDGEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTEPYEQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKPLYDRLREKDRACFRVPVDILPDEDDTDGGGEDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETG DKPYISQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGTTRTGRSK YAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKVQ GRIKRVIEDVYAIYDAFARDEINTLKELDACLADKGIRRGHLPK QMIGILSQERKDMEEKVRKKLQEMIADTDHRLDMLDRQTDRKI RIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSK ANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLH ETRWESHTNILSFYRSYLRARKAFLERIGRSDRVENCPFLLLKEP KTDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEMGYDEVGSYRE VGFMAKAVPLYFERACEDRVQPFYDSPFNVGNSLKPKKGRFLS KEKRAEEWESGKERFRLAKLKKEILEAQEHPYHDFKSWQKFER ELRLVKNQDIITWMMCRDLMEENKVEGLDTGTLYLKDIRPNVQ EQGSLNVLNRVKPMRLPVVVYRADSRGHVHKEEAPLATVYIEE RDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAMEQYPISKLR VEYELAKYQTARVCAFEQTLELEESLLTRYPHLPDESFREMLES WSDPLLTKWPELHGKVRLLIAVRNAFSHNQYPMYDEAVFSSIW KYDPSSPDAIEERMGLNIAHRLSEEVKQAKETIERIIQA
Porphyromonas gulae	WP_03944217	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDQDVLSFKALWKNLDNDLERKSRLRSLILKHFSF LEGAAYGKKLFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSGSSELPLFDGNMLQRLYNV

[0789]

		FDVSVQRVKRDHEHNDKVDPHYHFNHLVRKGKKDRYGHNDN PSFKHHFVDSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTGPYEQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKPLYDRLREKDRACFRVPVDILPDEDDTDGGGEDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYLETG DKPYISQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGTTRTGRSK CAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKVQ GRIKRVIEDVYAIYDAFARDEINTLKELDTCLADKGIRRGHLPK QMITILSQERKDMKEKIRKKLQEMIADTDHRLDMLDRQTDRKI RIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDASGKPLNNSK ANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLH ETRWESHTNILSFYRSYLRARKAFLERIGRSDRVENCPFLLLKEP KTDRQTLVAGWKDEFHLPRGIFTEAVRDCLIEMGYDEVGSYRE VGFMAKAVPLYFERACEDRVQPFYDSPFNVGNSLKPKKGRFLS KEDRAEEWERGMERFRDLEAWSHSAARRIKDAFAGIEYASPGN KKKIEQLLRDLSLWEAFESKLKVRADKINLAKLKKEILEAQEHP YHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLDT GTLYLKDIRPNVQEQGSLNVLNRVKPMRLPVVVYRADSRGHV HKEAPLATVYIEERNTKLLKQGNFKSFVKDRRLNGLFSFVDTG GLAMEQYPISKLRVEYELAKYQTARVCVFELTLRLEESLLSRYP HLPDESFREMLESWSDPLLAKWPELHGKVRLLIAVRNAFSHNQ YPMYDEAVFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAKE TVERIIQA
Porphyromonas gulae	WP_039445055	MNTVPATENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRI KFGKKKLNEESLKQSLLCDHLLSIDRWTKVYGHSRRYLPFLHCF DPDSGIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLKVSPDISSFITGAYTFACERAQSRFADFFKPDDFLLA KNRKEQLISVADGKECLTVSGFAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPQSMGFISVHDLRKLLLMELLCEGSFSRMQSDFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMNQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLRKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRRQFRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVRDKKRELRTAGKPVPPDLAAYIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDEENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR

[0790]

		RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDHENRFFGKLLNNMSQPINDL
Capnocyto phaga cynodegmi	WP_04 198958 1	MENKTSLGNNIYYNPFKPQDKSYFAGYLNAAMENIDSVFRELG KRLKGKEYTSENFFDAIFKENISLVEYERYVKLLSDYFPMARLL DKKEVPIKERKENFKKNFRGIIKAVRDLRNFYTHKEHGEVEITD EIFGVLDEMLKSTVLTVKKKKIKTDKTKEILKKSIEKQLDILCQK KLEYLKDTARKIEEKRRNQRERGEKKLVPRFEYSDRRDDLIAAI YNDAFDVYIDKKKDSLKESSKTKYNTESYPQQEEGDLKIPISKN GVVFLLSLFLSKQEVHAFKSKIAGFKATVIDEATVSHRKNSICF MATHEIFSHLAYKKLKRKVRTAEINYSEAENAEQLSIYAKETLM MQMLDELSKVPDVVYQNLSEDVQKTFIEDWNEYLKENNGDVG TMEEEQVIHPVIRKRYEDKFNYFAIRFLDEFAQFPTLRFQVHLG NYLHDSRPKEHLISDRRIKEKITVFGRLSELEHKKALFIKNTETN EDRKHYWEVFPNPNYDFPKENISVNDKDFPIAGSILDREKQPTA GKIGIKVNLLNQKYISEVDKAVKAHQLKQRNNKPSIQNIIEEIVPI NGSNPKEIIVFGGQPTAYLSMNDIHSILYEFFDKWEKKKEKLEK KGEKELRKEIGKELEEKIVGKIQTQIQQIIDKDINAKILKPYQDDD STAIDKEKLIKDLKQEQKILQKLKNEQTAREKEYQECIAYQEES RKIKRSDKSRQKYLRNQLKRKYPEVPTRKEILYYQEKGKVAVW LANDIKRFMPTDFKNEWKGEQHSLLQKSLAYYEQCKEELKNLL PQQKVFKHLPFELGGHFQQKYLYQFYTRYLDKRLEHISGLVQQ AENFKNENKVFKKVENECFKFLKKQNYTHKGLDAQAQSVLGY PIFLERGFMDEKPTIIKGKTFKGNESLFTDWFRYYKEYQNFQTF YDTENYPLVELEKKQADRKRETKIYQQKKNDVFTLLMAKHIFK SVFKQDSIDRFSLEDLYQSREERLENQEKAKQTGERNTNYIWNK TVDLNLCDGKVTVENVKLKNVGNFIKYEYDQRVQTFLKYEENI KWQAFLIKESKEEENYPYIVEREIEQYEKVRREELLKEVHLIEEY ILEKVKDKEILKKGDNQNFKYYILNGLLKQLKNEDVESYKVFN LNTKPEDVNINQLKQEATDLEQKAFVLTYIRNKFAHNQLPKKEF WDYCQEKYGKIEKEKTYAEYFAEVFKREKEALMK
Prevotella sp. P5-119	WP_04 251816 9	MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQ NENNENLWFHPVMSHLYNAKNGYDKQPEKTMFIIERLQSYFPF LKIMAENQREYSNGKYKQNRVEVNSNDIFEVLKRAFGVLKMY RDLTNHYKTYEEKLIDGCEFLTSTEQPLSGMISKYYTVALRNTK ERYGYKTEDLAFIQDNIKKITKDAYGKRKSQVNTGFFLSLQDYN GDTQKKLHLSGVGIALLICLFLDKQYINIFLSRLPIFSSYNAQSEE RRIIIRSFGINSIKLPKDRIHSEKSNKSVAMDMLNEVKRCPDELFT TLSAEKQSRFRIISDDHNEVLMKRSTDRFVPLLLQYIDYGKLFD HIRFHVNMGKLRYLLKADKTCIDGQTRVRVIEQPLNGFGRLEE AETMRKQENGTFGNSGIRIRDFENVKRDDANPANYPYIVDTYT HYILENNKVEMFISDKGSSAPLLPLIEDDRYVVKTIPSCRMSTLEI PAMAFHMFLFGSKKTEKLIVDVHNRYKRLFQAMQKEEVTAENI ASFGIAESDLPQKILDLISGNAHGKDVDAFIRLTVDDMLTDTER

[0791]

		RIKRFKDDRKSIRSADNKMGKRGFKQISTGKLADFLAKDIVLFQ PSVNDGENKITGLNYRIMQSAIAVYDSGDDYEAKQQFKLMFEK ARLIGKGTTEPHPFLYKVFARSIPANAVDFYERYLIERKFYLTGL CNEIKRGNRVDVPFIRRDQNKWKTPAMKTLGRIYSEDLPVELPR QMFDNEIKSHLKSLPQMEGIDFNNANVTYLIAEYMKRVLNDDF QTFYQWKRNYHYMDMLKGEYDRKGSLQHCFTSVEEREGLWK ERASRTERYRKLASNKIRSNRQMRNASSEEIETILDKRLSNCRNE YQKSEKVIRRYRVQDALLFLLAKKTLTELADFDGERFKLKEIMP DAEKGILSEIMPMSFTFEKGGKKYTITSEGMKLKNYGDFFVLAS DKRIGNLLELVGSDIVSKEDIMEEFNKYDQCRPEISSIVFNLEKW AFDTYPELSARVDREEKVDFKSILKILLNNKNINKEQSDILRKIR NAFDHNNYPDKGIVEIKALPEIAMSIKKAFGEYAIMK
Prevotella sp. P4-76	WP_04407214 7	MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQ NENNENLWFHPVMSHLYNAKNGYDKQPEKTMFIIERLQSYFPF LKIMAENQREYSNGKYKQNRVEVNSNDIFEVLKRAFGVLKMY RDQASHYKTYDEKLIDGCEFLTSTEQPLSGMINNYYTVALRNM NERYGYKTEDLAFIQDKRFKFVKDAYGKKKSQVNTGFFLSLQD YNGDTQKKLHLSGVGIALLICLFLDKQYINIFLSRLPIFSSYNAQS EERRIIIRSFGINSIKQPKDRIHSEKSNKSVAMDMLNEIKRCPNEL FETLSAEKQSRFRIISNDHNEVLMKRSSDRFVPLLLQYIDYGKLF DHIRFHVNMGKLRYLLKADKTCIDGQTRVRVIEQPLNGFGRLE EVETMRKQENGTFGNSGIRIRDFENMKRDDANPANYPYIVDTY THYILENNKVEMFISDEETPAPLLPVIEDDRYVVKTIPSCRMSTL EIPAMAFHMFLFGSKKTEKLIVDVHNRYKRLFKAMQKEEVTAE NIASFGIAESDLPQKIIDLISGNAHGKDVDAFIRLTVDDMLADTE RRIKRFKDDRKSIRSADNKMGKRGFKQISTGKLADFLAKDIVLF QPSVNDGENKITGLNYRIMQSAIAVYNSGDDYEAKQQFKLMFE KARLIGKGTTEPHPFLYKVFVRSIPANAVDFYERYLIERKFYLIG LSNEIKKGNRVDVPFIRRDQNKWKTPAMKTLGRIYDEDLPVELP RQMFDNEIKSHLKSLPQMEGIDFNNANVTYLIAEYMKRVLNDD FQTFYQWKRNYRYMDMLRGEYDRKGSLQSCFTSVEEREGLWK ERASRTERYRKLASNKIRSNRQMRNASSEEIETILDKRLSNSRNE YQKSEKVIRRYRVQDALLFLLAKKTLTELADFDGERFKLKEIMP DAEKGILSEIMPMSFTFEKGGKKYTITSEGMKLKNYGDFFVLAS DKRIGNLLELVGSDTVSKEDIMEEFKKYDQCRPEISSIVFNLEKW AFDTYPELSARVDREEKVDFKSILKILLNNKNINKEQSDILRKIR NAFDHNNYPDKGVVEIRALPEIAMSIKKAFGEYAIMK
Prevotella sp. P5-60	WP_04407478 0	MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQ NENNENLWFHPVMSHLYNAKNGYDKQPEKTMFIIERLQSYFPF LKIMAENQREYSNGKYKQNRVEVNSNDIFEVLKRAFGVLKMY RDLTNHYKTYEEKLIDGCEFLTSTEQPFSGMISKYYTVALRNTK ERYGYKAEDLAFIQDNRYKFTKDAYGKRKSQVNTGSFLSLQDY NGDTTKKLHLSGVGIALLICLFLDKQYINLFLSRLPIFSSYNAQSE ERRIIIRSFGINSIKQPKDRIHSEKSNKSVAMDMLNEVKRCPDELF TTLSAEKQSRFRIISDDHNEVLMKRSSDRFVPLLLQYIDYGKLFD

[0792]

		HIRFHVNMGKLRYLLKADKTCIDGQTRVRVIEQPLNGFGRLEE VETMRKQENGTFGNSGIRIRDFENMKRDDANPANYPYIVETYT HYILENNKVEMFISDEENPTPLLPVIEDDRYVVKTIPSCRMSTLEI PAMAFHMFLFGSEKTEKLIIDVHDRYKRLFQAMQKEEVTAENI ASFGIAESDLPQKIMDLISGNAHGKDVDAFIRLTVDDMLTDTER RIKRFKDDRKSIRSADNKMGKRGFKQISTGKLADFLAKDIVLFQ PSVNDGENKITGLNYRIMQSAIAVYDSGDDYEAKQQFKLMFEK ARLIGKGTTEPHPFLYKVFVRSIPANAVDFYERYLIERKFYLIGL SNEIKKGNRVDVPFIRRDQNKWKTPAMKTLGRIYSEDLPVELPR QMFDNEIKSHLKSLPQMEGIDFNNANVTYLIAEYMKRVLNDDF QTFYQWKRNYRYMDMLRGEYDRKGSLQHCFTSIEEREGLWKE RASRTERYRKLASNKIRSNRQMRNASSEEIETILDKRLSNCRNE YQKSEKIIRRYRVQDALLFLLAKKTLTELADFDGERFKLKEIMP DAEKGILSEIMPMSFTFEKGGKIYTITSGGMKLKNYGDFFVLAS DKRIGNLLELVGSNTVSKEDIMEEFKKYDQCRPEISSIVFNLEKW AFDTYPELPARVDRKEKVDFWSILDVLSNNKDINNEQSYILRKI RNAFDHNNYPDKGIVEIKALPEIAMSIKKAFGEYAIMK
Phaeodactylibacter xiamenensis	WP_044218239	MTNTPKRRTLHRHPSYFGAFLNIARHNAFMIMEHLSTKYDMED KNTLDEAQLPNAKLFGCLKKRYGKPDVTEGVSRDLRRYFPFLN YPLFLHLEKQQNAEQAATYDINPEDIEFTLKGFFRLLNQMRNNY SHYISNTDYGKFDKLPVQDIYEAAIFRLLDRGKHTKRFDVFESK HTRHLESNNSEYRPRSLANSPDHENTVAFVTCLFLERKYAFPFL SRLDCFRSTNDAAEGDPLIRKASHECYTMFCCRLPQPKLESSDIL LDMVNELGRCPSALYNLLSEEDQARFHIKREEITGFEEDPDEELE QEIVLKRHSDRFPYFALRYFDDTEAFQTLRFDVYLGRWRTKPV YKKRIYGQERDRVLTQSIRTFTRLSRLLPIYENVKHDAVRQNEE DGKLVNPDVTSQFHKSWIQIESDDRAFLSDRIEHFSPHYNFGDQ VIGLKFINPDRYAAIQNVFPKLPGEEKKDKDAKLVNETADAIIST HEIRSLFLYHYLSKKPISAGDERRFIQVDTETFIKQYIDTIKLFFED IKSGELQPIADPPNYQKNEPLPYVRGDKEKTQEERAQYRERQKE IKERRKELNTLLQNRYGLSIQYIPSRLREYLLGYKKVPYEKLAL QKLRAQRKEVKKRIKDIEKMRTPRVGEQATWLAEDIVFLTPPK MHTPERKTTKHPQKLNNDQFRIMQSSLAYFSVNKKAIKKFFQK ETGIGLSNRETSHPFLYRIDVGRCRGILDFYTGYLKYKMDWLDD AIKKVDNRKHGKKEAKKYEKYLPSSIQHKTPLELDYTRLPVYLP RGLFKKAIVKALAAHADFQVEPEEDNVIFCLDQLLDGDTQDFY NWQRYYRSALTEKETDNQLVLAHPYAEQILGTIKTLEGKQKNN KLGNKAKQKIKDELIDLKRAKRRLLDREQYLRAVQAEDRALW LMIQERQKQKAEHEEIAFDQLDLKNITKILTESIDARLRIPDTKV DITDKLPLRRYGDLRRVAKDRRLVNLASYYHVAGLSEIPYDLV KKELEEYDRRRVAFFEHVYQFEKEVYDRYAAELRNENPKGEST YFSHWEYVAVAVKHSADTHFNELFKEKVMQLRNKFHHNEFPY FDWLLPEVEKASAALYADRVFDVAEGYYQKMRKLMRQ
Flavobacterium sp. 316	WP_044596837	MDNNITVEKTELGLGITYNHDKVEDKHYFGGFFNLAQNNIDLV AQEFKKRLLIQGKDSINIFANYFSDQCSITNLERGIKILAEYFPVV

[0793]

	7	SYIDLDEKNKSKSIREHLILLLETINNLRNYYTHYYHKKIIIDGSL FPLLDTILLKVVLEIKKKKLKEDKTKQLLKKGLEKEMTILFNLM KAEQKEKKIKGWNIDENIKGAVLNRAFSHLLYNDELSDYRKSK YNTEDETLKDTLTESGILFLLSFFLNKKEQEQLKANIKGYKGKIA SIPDEEITLKNNSLRNMATHWTYSHLTYKGLKHRIKTDHEKETL LVNMVDYLSKVPHEIYQNLSEQNKSLFLEDINEYMRDNEENHD SSEASRVIHPVIRKRYENKFAYFAIRFLDEFAEFPTLRFMVNVGN YIHDNRKKDIGGTSLITNRTIKQQINVFGNLTEIHKKKNDYFEKE ENKEKTLEWELFPNPSYHFQKENIPIFIDLEKSKETNDLAKEYAK EKKKIFGSSRKKQQNTAKKNRETIINLVFDKYKTSDRKTVTFEQ PTALLSFNELNSFLYAFLVENKTGKELEKIIIEKIANQYQILKNCS STVDKTNDNIPKSIKKIVNTTTDSFYFEGKKIDIEKLEKDITIEIEK TNEKLETIKENEESAQNYKRNERNTQKRKLYRKYVFFTNEIGIE ATWITNDILRFLDNKENWKGYQHSELQKFISQYDNYKKEALGL LESEWNLESDAFFGQNLKRMFQSNSTFETFYKKYLDNRKNTLE TYLSAIENLKTMTDVRPKVLKKKWTELFRFFDKKIYLLSTIETKI NELITKPINLSRGIFEEKPTFINGKNPNKENNQHLFANWFIYAKK QTILQDFYNLPLEQPKAITNLKKHKYKLERSINNLKIEDIYIKQM VDFLYQKLFEQSFIGSLQDLYTSKEKREIEKGKAKNEQTPDESFI WKKQVEINTHNGRIIAKTKIKDIGKFKNLLTDNKIAHLISYDDRI WDFSLNNDGDITKKLYSINTELESYETIRREKLLKQIQQFEQFLL EQETEYSAERKHPEKFEKDCNPNFKKYIIEGVLNKIIPNHEIEEIEI LKSKEDVFKINFSDILILNNDNIKKGYLLIMIRNKFAHNQLIDKN LFNFSLQLYSKNENENFSEYLNKVCQNIIQEFKEKLK
Porphyromonas gulae	WP_046201018	MTEQSERPYNGTYYTLEDKHFWAAFLNLARHNAYITLTHIDRQ LAYSKADITNDQDVLSFKALWKNFDNDLERKSRLRSLILKHFSF LEGAAYGKKLFESKSSGNKSSKNKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHSESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRYGHNDN PSFKHHFVDSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTETYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKPLYDRLREKDRARFRVPVDILPDEDDTDGGGEDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETG DKPYISQTTPHYHIEKGKIGLRFMPEGQHLWPSPEVGTTRTGRS KYAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKV QGRIKRVIEDVYAIYDAFARDEINTLKELDACLADKGIRRGHLP KQMIAILSQEHKDMEEKIRKKLQEMIADTDHRLDMLDRQTDRK IRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSK ANSTEYRMLQRALALFGGEKKRLTPYFRQMNLTGGNNPHPFLH ETRWESHTNILSFYRSYLRARKAFLERIGRSDRMENRPFLLLKEP KTDRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYRE VGFMAKAVPLYFERACEDRVQPFYDSPFNVGNSLKPKKGRFLS KEERAEEWERGKERFRDLEAWSHSAARRIEDAFAGIEYASPGN KKKIEQLLRDLSLWEAFESKLKVRADKINLAKLKKEILEAQEHP YHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLDT

[0794]

		GTLYLKDIRPNVQEQGSLNVLNRVKPMRLPVVVYRADSRGHV HKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTG GLAMEQYPISKLRVEYELAKYQTARVCVFELTLRLEESLLTRYP HLPDESFRKMLESWSDPLLAKWPELHGKVRLLIAVRNAFSHNQ YPMYDEAVFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAKE TVERIIQV
WP_04743 1796	Chryse obacteri um sp. YR477	METQTIGHGIAYDHSKIQDKHFFGGFLNLAENNIKAVLKAFSEK FNVGNVDVKQFADVSLKDNLPDNDFQKRVSFLKMYFPVVDFIN IPNNRAKFRSDLTTLFKSVDQLRNFYTHYYHKPLDFDASLFILLD DIFARTAKEVRDQKMKDDKTRQLLSKSLSEELQKGYELQLERL KELNRLGKKVNIHDQLGIKNGVLNNAFNHLIYKDGESFKTKLT YSSALTSFESAENGIEISQSGLLFLLSMFLKRKEIEDLKNRNKGF KAKVVIDEDGKVNGLKFMATHWVFSYLCFKGLKSKLSTEFHEE TLLIQIIDELSKVPDELYCAFDKETRDKFIEDINEYVKEGHQDFSL EDAKVIHPVIRKRYENKFNYFAIRFLDEFVKFPSLRFQVHVGNY VHDRRIKNIDGTTFETERVVKDRIKVFGRLSEISSYKAQYLSSVS DKHDETGWEIFPNPSYVFINNNIPIHISVDTSFKKEIADFKKLRRA QVPDELKIRGAEKKRKFEITQMIGSKSVLNQEEPIALLSLNEIPAL LYEILINGKEPAEIERIIKDKLNERQDVIKNYNPENWLPASQISRR LRSNKGERIINTDKLLQLVTKELLVTEQKLKIISDNREALKQKKE GKYIRKFIFTNSELGREAIWLADDIKRFMPADVRKEWKGYQHS QLQQSLAFYNSRPKEALAILESSWNLKDEKIIWNEWILKSFTQN KFFDAFYNEYLKGRKKYFAFLSEHIVQYTSNAKNLQKFIKQQM PKDLFEKRHYIIEDLQTEKNKILSKPFIFPRGIFDKKPTFIKGVKV EDSPESFANWYQYGYQKDHQFQKFYDWKRDYSDVFLEHLGKP FINNGDRRTLGMEELKERIIIKQDLKIKKIKIQDLFLRLIAENLFQ KVFKYSAKLPLSDFYLTQEERMEKENMAALQNVREEGDKSPNI IKDNFIWSKMIPYKKGQIIENAVKLKDIGKLNVLSLDDKVQTLL SYDDAKPWSKIALENEFSIGENSYEVIRREKLFKEIQQFESEILFR SGWDGINHPAQLEDNRNPKFKMYIVNGILRKSAGLYSQGEDIW FEYNADFNNLDADVLETKSELVQLAFLVTAIRNKFAHNQLPAK EFYFYIRAKYGFADEPSVALVYLNFTKYAINEFKKVMI
Riemerella anatipestife r	WP_04 935426 3	MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYKAHFTVNLTRDF AHLNRKGKNKQDNPDFNRYRFEKDGFFTESGLLFFTNLFLDKR DAYWMLKKVSGFKASHKQREKMTTEVFCRSRILLPKLRLESRY DHNQMLLDMLSELSRCPKLLYEKLSEENKKHFQVEADGFLDEI EEEQNPFKDTLIRHQDRFPYFALRYLDLNESFKSIRFQVDLGTYH YCIYDKKIGDEQEKRHLTRTLLSFGRLQDFTEINRPQEWKALTK DLDYKETSNQPFISKTTPHYHITDNKIGFRLGTSKELYPSLEIKDG ANRIAKYPYNSGFVAHAFISVHELLPLMFYQHLTGKSEDLLKET VRHIQRIYKDFEEERINTIEDLEKANQGRLPLGAFPKQMLGLLQ NKQPDLSEKAKIKIEKLIAETKLLSHRLNTKLKSSPKLGKRREKL IKTGVLADWLVKDFMRFQPVAYDAQNQPIKSSKANSTEFWFIR RALALYGGEKNRLEGYFKQTNLIGNTNPHPFLNKFNWKACRNL VDFYQQYLEQREKFLEAIKNQPWEPYQYCLLLKIPKENRKNLV

[0795]

		KGWEQGGISLPRGLFTEAIRETLSEDLMLSKPIRKEIKKHGRVGF ISRAITLYFKEKYQDKHQSFYNLSYKLEAKAPLLKREEHYEYW QQNKPQSPTESQRLELHTSDRWKDYLLYKRWQHLEKKLRLYR NQDVMLWLMTLELTKNHFKELNLNYHQLKLENLAVNVQEAD AKLNPLNQTLPMVLPVKVYPATAFGEVQYHKTPIRTVYIREEHT KALKMGNFKALVKDRRLNGLFSFIKEENDTQKHPISQLRLRREL EIYQSLRVDAFKETLSLEEKLLNKHTSLSSLENEFRALLEEWKK EYAASSMVTDEHIAFIASVRNAFCHNQYPFYKEALHAPIPLFTV AQPTTEEKDGLGIAEALLKVLREYCEIVKSQI
Porphyromonas gingivalis	WP_052912312	MTEQNEKPYNGTYYTLEDKHFWAAFFNLARHNAYITLAHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSF LEGAAYGKKLFESQSSGNKSSKKKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHPESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDKYGNNDN PFFKHHFVDREEKVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKG GTEAYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNELV RCPKLLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTLV RHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQ PEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDK PYITQTTPHYHIEKGKIGLRFVPEGQLLWPSPEVGATRTGRSKY AQDKRFTAEAFLSVHELMPMMFYYFLLREKYSEEASAEKVQG RIKRVIEDVYAVYDAFARDEINTRDELDACLADKGIRRGHLPRQ MIAILSQEHKDMEEKVRKKLQEMIADTDHRLDMLDRQTDRKIR IGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSKA NSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLHE TRWESHTNILSFYRSYLKARKAFLQSIGRSDREENHRFLLLKEPK TDRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEV GFMAKAVPLYFERACKDRVQPFYDYPFNVGNSLKPKKGRFLSK EKRAEEWESGKERFRDLEAWSHSAARRIEDAFVGIEYASWENK KKIEQLLQDLSLWETFESKLKVKADKINIAKLKKEILEAKEHPY HDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGLDTG TLYLKDIRTDVQEQGSLNVLNHVKPMRLPVVVYRADSRGHVH KEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGA LAMEQYPISKLRVEYELAKYQTARVCAFEQTLELEESLLTRYPH LPDESFREMLESWSDPLLDKWPDLQREVRLLIAVRNAFSHNQY PMYDETIFSSIRKYDPSSLDAIEERMGLNIAHRLSEEVKLAKEMV ERIIQA
Porphyromonas gingivalis	WP_058019250	MTEQNEKPYNGTYYTLKDKHFWAAFFNLARHNAYITLTHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSF LEGAAYGKKLFESQSSGNKSSKKKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHPESSELPMFDGNMLQRLYN VFDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRCGNND NPFFKHHFVDREGKVTEAGLLFFVSLFLEKRDAIWMQKKIRGF KGGTETYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNE LVRCPKSLYDRLREEDRACFRVPVDILSDEDDTDGAEEDPFKNT

[0796]

		LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIG EQPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDCFETG DKPYITQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGATRTGRS KYAQDKRFTAEAFLSVHELMPMMFYYFLLREKYSEEVSAERV QGRIKRVIEDVYAVYDAFARDEINTRDELDACLADKGIRRGHLP RQMIAILSQKHKDMEEKVRKKLQEMIADTDHRLDMLDRQTDR KIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNS KANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFL HETRWESHTNILSFYRSYLKARKAFLQSIGRSDRVENHRFLLLK EPKTDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEMGLDEVGSY KEVGFMAKAVPLYFERACKDRVQPFYDYPFNVGNSLKPKKGR FLSKEKRAEEWESGKERFRDLEAWSHSAARRIEDAFAGIENASR ENKKKIEQLLQDLSLWETFESKLKVKADKINIAKLKKEILEAKE HPYLDFKSWQKFERELRLVKNQDIITWMMCRDLMEENKVEGL DTGTLYLKDIRTDVQEQGSLNVLNHVKPMRLPVVVYRADSRG HVHKEQAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFV DTGALAMEQYPISKLRVEYELAKYQTARVCAFEQTLELEESLLT RYPHLPDENFRKMLESWSDPLLDKWPDLHRKVRLLIAVRNAFS HNQYPMYDEAVFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQ AKEMAERIIQA
Flavobacterium columnare	WP_060381855	MSSKNESYNKQKTFNHYKQEDKYFFGGFLNNADDNLRQVGKE FKTRINFNHNNNELASVFKDYFNKEKSVAKREHALNLLSNYFP VLERIQKHTNHNFEQTREIFELLLDTIKKLRDYYTHHYHKPITIN PKVYDFLDDTLLDVLITIKKKKVKNDTSRELLKEKFRPELTQLK NQKREELIKKGKKLLEENLENAVFNHCLRPFLEENKTDDKQNK TVSLRKYRKSKPNEETSITLTQSGLVFLISFFLHRKEFQVFTSGLE GFKAKVNTIKEEEISLNKNNIVYMITHWSYSYYNFKGLKHRIKT DQGVSTLEQNNTTHSLTNTNTKEALLTQIVDYLSKVPNEIYETL SEKQQKEFEEDINEYMRENPENEDSTFSSIVSHKVIRKRYENKFN YFAMRFLDEYAELPTLRFMVNFGDYIKDRQKKILESIQFDSERII KKEIHLFEKLGLVTEYKKNVYLKETSNIDLSRFPLFPSPSYVMA NNNIPFYIDSRSNNLDEYLNQKKKAQSQNRKRNLTFEKYNKEQ SKDAIIAMLQKEIGVKDLQQRSTIGLLSCNELPSMLYEVIVKDIK GAELENKIAQKIREQYQSIRDFTLDSPQKDNIPTTLTKTISTDTSV TFENQPIDIPRLKNALQKELTLTQEKLLNVKQHEIEVDNYNRNK NTYKFKNQPKDKVDDNKLQRKYVFYRNEIGQEANWLASDLIH FMKNKSLWKGYMHNELQSFLAFFEDKKNDCIALLETVFNLKED CILTKDLKNLFLKHGNFIDFYKEYLKLKEDFLNTESTFLENGFIG LPPKILKKELSKRLNYIFIVFQKRQFIIKELEEKKNNLYADAINLS RGIFDEKPTMIPFKKPNPDEFASWFVASYQYNNYQSFYELTPDK IENDKKKKYKNLRAINKVKIQDYYLKLMVDTLYQDLFNQPLDK SLSDFYVSKTDREKIKADAKAYQKRNDSFLWNKVIHLSLQNNR ITANPKLKDIGKYKRALQDEKIATLLTYDDRTWTYALQKPEKE NENDYKELHYTALNMELQEYEKVRSKKLLKQVQELEKQILDKF YDFSNNATHPEDLEIEDKKGKRHPNFKLYITKALLKNESEIINLE NIDIEILIKYYDYNTEKLKEKIKNMDEDEKAKIVNTKENYNKITN

[0797]

		VLIKKALVLIIIRNKMAHNQYPPKFIYDLATRFVPKKEEEYFACY FNRVFETITTELWENKKKAKEIV
Porphyromonas gingivalis	WP_061156470	MTEQNERPYNGTYYTLEDKHFWAAFFNLARHNAYITLTHIDRQ LAYSKADITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSF LEGAAYGKKLFENKSSGNKSSKKKELTKKEKEELQANALSLDN LKSILFDFLQKLKDFRNYYSHYRHPESSELPLFDGNMLQRLYNV FDVSVQRVKRDHEHNDKVDPHRHFNHLVRKGKKDRCGNNDN PFFKHHFVDREGKVTEAGLLFFVSLFLEKRDAIWMQKKIRGFK GGTEAYQQMTNEVFCRSRISLPKLKLESLRTDDWMLLDMLNEL VRCPKSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTL VRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGE QPEDRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGD KPYITQTTPHYHIEKGKIGLRFVPEGQHLWPSPEVGATRTGRSK YAQDKRLTAEAFLSVHELMPMMFYYFLLREKYSEEVSAEKVQ GRIKRVIEDVYAVYDAFARGEIDTLDRLDACLADKGIRRGHLPR QMIAILSQEHKDMEEKVRKKLQEMIADTDHRLDMLDRQTDRKI RIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGKPLNNSK ANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLH ETRWESHTNILSFYRSYLKARKAFLQSIGRSDREENHRFLLLKEP KTDRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKE VGFMAKAVPLYFERACKDRVQPFYDYPFNVGNSLKPKKGRFLS KEKRAEEWESGKERFRLAKLKKEILEAKEHPYLDFKSWQKFER ELRLVKNQDIITWMMCRDLMEENKVEGLDTGTLYLKDIRTEVQ EQGSLNVLNRVKPMRLPVVVYRADSRGHVHKEQAPLATVYIEE RDTKLLKQGNFKSFVKDRRLNGLFSFVDTGGLAMEQYPISKLR VEYELAKYQTARVCAFEQTLELEESLLTRCPHLPDKNFRKMLES WSDPLLDKWPDLQREVWLLIAVRNAFSHNQYPMYDEAVFSSIR KYDPSSPDAIEERMGLNIAHRLSEEVKQAKEMAERIIQA
Porphyromonas gingivalis	WP_061156637	MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIK FGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSRRYLPFLHYF DPDSQIEKDHDSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNRL DGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLA KNRKEQLISVADGKECLTVSGLAFFICLFLDREQASGMLSRIRGF KRTDENWARAVHETFCDLCIRHPHDRLESSNTKEALLLDMLNE LNRCPRILYDMLPEEERAQFLPALDENSMNNLSENSLNEESRLL WDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKGIRFRVD LGEIELDSYSKKVGRNGEYDRTITDHALAFGKLSDFQNEEEVSR MISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEK RALSNPQSMGFISVHDLRKLLLMELLCEGSFSRMQSGFLRKANR ILDETAEGKLQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKY KQEIKGRKDKLNSQLLSAFDMNQRQLPSRLLDEWMNIRPASHS VKLRTYVKQLNEDCRLRLRKFRKDGDGKARAIPLVGEMATFLS QDIVRMIISEETKKLITSAYYNEMQRSLAQYAGEENRRQFRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAK TFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQ

[0798]

		DWIRNKQAHPIDLPSHLFDSKIMELLKVKDGKKKWNEAFKDW WSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHL MEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLR LVQEDDRLMLMAINKMMTDREEDILPGLKNIDSILDKENQFSLA VHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDR RVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAI MSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQ YLILIRNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPE GSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL
Riemerella anatipestifer	WP_061710138	MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYKAHFTVNLTRDF AHLNRKGKNKQDNPDFNRYRFEKDGFFTESGLLFFTNLFLDKR DAYWMLKKVSGFKASHKQSEKMTTEVFCRSRILLPKLRLESRY DHNQMLLDMLSELSRCPKLLYEKLSEKDKKCFQVEADGFLDEI EEEQNPFKDTLIRHQDRFPYFALRYLDLNESFKSIRFQVDLGTYH YCIYDKKIGYEQEKRHLTRTLLNFGRLQDFTEINRPQEWKALTK DLDYNETSNQPFISKTTPHYHITDNKIGFRLRTSKELYPSLEVKD GANRIAKYPYNSDFVAHAFISISVHELLPLMFYQHLTGKSEDLL KETVRHIQRIYKDFEEERINTIEDLEKANQGRLPLGAFPKQMLGL LQNKQPDLSEKAKIKIEKLIAETKLLSHRLNTKLKSSPKLGKRRE KLIKTGVLADWLVKDFMRFQPVVYDAQNQPIKSSKANSTESRLI RRALALYGGEKNRLEGYFKQTNLIGNTNPHPFLNKFNWKACRN LVDFYQQYLEQREKFLEAIKHQPWEPYQYCLLLKVPKENRKNL VKGWEQGGISLPRGLFTEAIRETLSKDLTLSKPIRKEIKKHGRVG FISRAITLYFKEKYQDKHQSFYNLSYKLEAKAPLLKKEEHYEYW QQNKPQSPTESQRLELHTSDRWKDYLLYKRWQHLEKKLRLYR NQDIMLWLMTLELTKNHFKELNLNYHQLKLENLAVNVQEADA KLNPLNQTLPMVLPVKVYPTTAFGEVQYHETPIRTVYIREEQTK ALKMGNFKALVKDRHLNGLFSFIKEENDTQKHPISQLRLRRELE IYQSLRVDAFKETLSLEEKLLNKHASLSSLENEFRTLLEEWKKK YAASSMVTDKHIAFIASVRNAFCHNQYPFYKETLHAPILLFTVA QPTTEEKDGLGIAEALLRVLREYCEIVKSQI
Flavobacterium columnare	WP_063744070	MSSKNESYNKQKTFNHYKQEDKYFFGGFLNNADDNLRQVGKE FKTRINFNHNNNELASVFKDYFNKEKSVAKREHALNLLSNYFP VLERIQKHTNHNFEQTREIFELLLDTIKKLRDYYTHHYHKPITIN PKIYDFLDDTLLDVLITIKKKKVKNDTSRELLKEKLRPELTQLKN QKREELIKKGKKLLEENLENAVFNHCLRPFLEENKTDDKQNKT VSLRKYRKSKPNEETSITLTQSGLVFLMSFFLHRKEFQVFTSGLE GFKAKVNTIKEEKISLNKNNIVYMITHWSYSYYNFKGLKHRIKT DQGVSTLEQNNTTHSLTNTNTKEALLTQIVDYLSKVPNEIYETL SEKQQKEFEEDINEYMRENPENEDSTFSSIVSHKVIRKRYENKFN YFAMRFLDEYAELPTLRFMVNFGDYIKDRQKKILESIQFDSERII KKEIHLFEKLGLVTEYKKNVYLKETSNIDLSRFPLFPSPSYVMA NNNIPFYIDSRSNNLDEYLNQKKKAQSQNRKRNLTFEKYNKEQ SKDAIIAMLQKEIGVKDLQQRSTIGLLSCNELPSMLYEVIVKDIK GAELENKIAQKIREQYQSIRDFTLNSPQKDNIPTTLIKTISTDTSV

[0799]

		TFENQPIDIPRLKNAIQKELALTQEKLLNVKQHEIEVNNYNRNK NTYKFKNQPKDKVDDNKLQRKYVFYRNEIGQEANWLASDLIH FMKNKSLWKGYMHNELQSFLAFFEDKKNDCIALLETVFNLKED CILTKDLKNLFLKHGNFIDFYKEYLKLKEDFLNTESTFLENGFIG LPPKILKKELSKRLNYIFIVFQKRQFIIKELEEKKNNLYADAINLS RGIFDEKPTMIPFKKPNPDEFASWFVASYQYNNYQSFYELTPDK IENDKKKKYKNLRAINKVKIQDYYLKLMVDTLYQDLFNQPLDK SLSDFYVSKTDREKIKADAKAYQKRNDSFLWNKVIHLSLQNNR ITANPKLKDIGKYKRALQDEKIATLLTYDDRTWTYALQKPEKE NENDYKELHYTALNMELQEYEKVRSKKLLKQVQELEKQILDKF YDFSNNATHPEDLEIEDKKGKRHPNFKLYITKALLKNESEIINLE NIDIEILIKYYDYNTEKLKEKIKNMDEDEKAKIVNTKENYNKITN VLIKKALVLIIIRNKMAHNQYPPKFIYDLATRFVPKKEEEYFACY FNRVFETITTELWENKKKAKEIV
Riemerella anatipestifer	WP_064970887	MEKPLPPNVYTLKHKFFWGAFLNIARHNAFITICHINEQLGLTTP PNDDKIADVVCGTWNNILNNDHDLLKKSQLTELILKHFPFLAA MCYHPPKKEGKKKGSQKEQQKEKENEAQSQAEALNPSELIKVL KTLVKQLRTLRNYYSHHSHKKPDAEKDIFKHLYKAFDASLRMV KEDYKAHFTVNLTQDFAHLNRKGKNKQDNPDFDRYRFEKDGF FTESGLLFFTNLFLDKRDAYWMLKKVSGFKASHKQSEKMTTEV FCRSRILLPKLRLESRYDHNQMLLDMLSELSRYPKLLYEKLSEE DKKRFQVEADGFLDEIEEEQNPFKDTLIRHQDRFPYFALRYLDL NESFKSIRFQVDLGTYHYCIYDKKIGDEQEKRHLTRTLLSFGRL QDFTEINRPQEWKALTKDLDYKETSKQPFISKTTPHYHITDNKIG FRLGTSKELYPSLEVKDGANRIAQYPYNSDFVAHAFISVHELLP LMFYQHLTGKSEDLLKETVRHIQRIYKDFEEERINTIEDLEKANQ GRLPLGAFPKQMLGLLQNKQPDLSEKAKIKIEKLIAETKLLSHR LNTKLKSSPKLGKRREKLIKTGVLADWLVKDFMRFQPVAYDA QNQPIESSKANSTEFQLIQRALALYGGEKNRLEGYFKQTNLIGN TNPHPFLNKFNWKACRNLVDFYQQYLEQREKFLEAIKNQPWEP YQYCLLLKIPKENRKNLVKGWEQGGISLPRGLFTEAIRETLSKD LTLSKPIRKEIKKHGRVGFISRAITLYFREKYQDDHQSFYDLPYK LEAKASPLPKKEHYEYWQQNKPQSPTELQRLELHTSDRWKDYL LYKRWQHLEKKLRLYRNQDVMLWLMTLELTKNHFKELNLNY HQLKLENLAVNVQEADAKLNPLNQTLPMVLPVKVYPATAFGE VQYQETPIRTVYIREEQTKALKMGNFKALVKDRRLNGLFSFIKE ENDTQKHPISQLRLRRELEIYQSLRVDAFKETLNLEEKLLKKHTS LSSVENKFRILLEEWKKEYAASSMVTDEHIAFIASVRNAFCHNQ YPFYEEALHAPIPLFTVAQQTTEEKDGLGIAEALLRVLREYCEIV KSQI

[0800]

[0801] 표 1 본 발명의 Cas13b 이펙터 단백질은 상기 표 1로부터의 또는 그에 기재된 바와 같은 단백질 또는 표 1에 기재된 유기체의 Cas13b 이펙터 단백질을 비롯하여, 그와 90, 91, 92, 93, 94, 95 96, 97, 98, 99 또는 100%의 동일성을 갖는 단백질, 예를 들어 전술한 백분율 내에서 하나 이상의 변화를 갖는 단백질이거나, 또는 그를 포함하거나, 또는 그로 본질적으로 이루어지거나, 또는 그로 이루어지거나, 또는 그를 수반하거나, 또는 그와 관련되고, 여기서 변화는 보존성 치환 그룹이거나, 그를 포함하거나, 그로 본질적으로 이루어지거나, 또는 그로 이루어지고, 보존성 치환은 하기 표를 참조할 수 있다:

표 1

[0802]

부류	**아미노산 명칭**
지방족	글리신, 알라닌, 발린, 류신, 이소류신
히드록실 또는 황/셀레늄-함유	세린, 시스테인, 셀레노시스테인, 트레오닌, 메티오닌
환형	프롤린
방향족	페닐알라닌, 티로신, 트립토판
염기성	히스티딘, 리신, 아르기닌
산성 및 그들의 아미드	아스파테이트, 글루타메이트, 아스파라긴, 글루타민

[0803] 또는, 보존성 치환에 관한 당분야의 지식, 예를 들어, [French et al. "What is a Conservative Substitution?", J. Molecular Evolution, 19(2):171-175 (March 1983)]; [Yampolsky et al., "The

Exchangeability of Amino Acids in Proteins," Genetics 170(4): 1459-1472 (Aug 2005), doi: 10.1534/genetics.104.039107]; [Gemovic, "Feature-Based Classification of Amino Acid Substitutions outside Conserved Functional Protein Domains," The Scientific World Journal, Volume 2013 (2013), Article ID 948617, 10 pages dx.doi.org/10.1155/2013/948617]를 참조한다. 또한, 표 1 본 발명의 Cas13b 이펙터 단백질은 본 명세서에 기술된 바와 같은 돌연변이, 변경 또는 융합을 포함하는, 상기 표 1로부터의 또는 그에 기재된 바와 같은 단백질 또는 표 1에 기재된 바와 같은 유기체의 Cas13b 이펙터 단백질을 비롯하여, 본 명세서에 기술된 바와 같은 돌연변이, 변경 또는 융합을 포함하고, 이러한 돌연변이, 변경 또는 융합을 고려하거나 또는 고려하지 않고, 변경 또는 융합이 표 1에 기재된 것 또는 표 1에 기재된 유기체의 Cas13b 이펙터 단백질과 90, 91, 92, 93, 94, 95 96, 97, 98, 99 또는 100%의 동일성을 갖는 단백질이거나, 또는 그를 포함하거나, 또는 그로 본질적으로 이루어지거나, 또는 그로 이루어지거나, 또는 그를 수반하거나, 또는 그와 관련된다. 또한, (본 명세서에 기술된 돌연변이, 변경 또는 융합 및/또는 표 1에 기재된 것 또는 표 1에 기재된 유기체의 Cas13b 이펙터 단백질과 90, 91, 92, 93, 94, 95 96, 97, 98, 99 또는 100% 동일성을 포함할 수 있는) 표 1 Cas13b 이펙터 단백질은 표 1에 기재된 것 또는 표 1에 기재된 유기체의 Cas13b 이펙터 단백질과 적어도 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95 96, 97, 98, 99 또는 100% 동일성 및 표 1의 단백질(들) 및 표 3의 단백질(들) 사이에 표시된 바와 같은 표 3의 C2c2 단백질과의 유사성, 상동성 또는 동일성을 갖는 Cas13b 이펙터 단백질일 수 있다. 도 1, 도 2를 또한 참조한다.

[0804] *실시예 2: 진핵생물 세포에서 Cas13b의 활성*

[0805] 루시퍼라제 표적화 어세이가 Gluc에 대한 상이한 gRNA를 사용해 수행되었다. Cas13b 오르소로그는 NLS 또는 NES에 융합되거나 또는 대안적으로 국재화 신호에 융합되지 않았다. 루시퍼라제의 정규화된 단백질 발현은 비표적화 (NT) gRNA와 비교하였다. 상이한 Cas13b 오르소로그에 대한 결과는 도 11-15에 제공된다. 도 16에서, 동일한 가이드를 갖는 가장 활성있는 오르소로그로부터의 복합 데이타가 편집되어 있다. 박테로이데스 피오제네스 (*Bacteroides pyogenes*), 프레보텔라 (*Prevotella*) *sp.* MA2016, 리에메렐라 아나티페스티퍼 (*Riemerella anatipestifer*), 폴피로모나스 굴라에 (*Porphyromonas gulae*), 폴피로모나스 진지발리스 (*Porphyromonas gingivalis*), 및 폴피로모나스 (*Porphyromonas*) *sp.* COT-052OH4946 유래의 Cas13b 오르소로그는 특히 진핵생물 세포에서 활성이 있는 것으로 확인되었다.

[0806] *실시예 3: C2c2 오르소로그*

[0807] 일정한 Cas13b 오르소로그는 놀랍게도 C2c2와 유사하다. 도 2는 C2c2 및 Cas13b 단백질의 계통수 정렬을 제공한다.

[0808] 표 2: Cas13b 단백질과 유사한 C2c2 오르소로그

[0809]

표 2: Cas13b 단백질과 유사한 C2c2 오르소로그		
C2c2 오르소로그	코드	멀티 글자
Leptotrichia shahii	C2-2	Lsh
L wadei F0279 (Lw2)	C2-3	Lw2
Listeria seeligeri	C2-4	Lse

Lachnospiraceae bacterium MA2020	C2-5	LbM
Lachnospiraceae bacterium NK4A179	C2-6	LbNK179
[Clostridium] aminophilum DSM 10710	C2-7	Ca
Carnobacterium gallinarum DSM 4847	C2-8	Cg
Carnobacterium gallinarum DSM 4847	C2-9	Cg2
Paludibacter propionicigenes WB4	C2-10	Pp
Listeria weihenstephanensis FSL R9-0317	C2-11	Lwei
Listeriaceae bacterium FSL M6-0635	C2-12	LbFSL
Leptotrichia wadei F0279	C2-13	Lw
Rhodobacter capsulatus SB 1003	C2-14	Rc
Rhodobacter capsulatus R121	C2-15	Rc
Rhodobacter capsulatus DE442	C2-16	Rc
Leptotrichia buccalis C-1013-b	C2-17	Lbu
Herbinix hemicellulosilytica	C2-18	Hhe
[Eubacterium] rectale	C2-19	Ere
Eubacteriaceae bacterium CHKCI004	C2-20	Eba
Blautia sp. Marseille-P2398	C2-21	BSm
Leptotrichia sp. oral taxon 879 str. F0557	C2-22	Lsp
Lachnospiraceae bacterium NK4A144	C2-23	NK4A144
RNA-binding protein S1 Chloroflexus aggregans	C2-24	
Demequina aurantiaca	C2-25	
Thalassospira sp. TSL5-1	C2-26	
SAMN04487830_13920 Pseudobutyrivibrio sp. OR37	C2-27	
SAMN02910398_00008 Butyrivibrio sp. YAB3001	C2-28	
Blautia sp. Marseille-P2398	C2-29	
Leptotrichia sp. Marseille-P3007	C2-30	
Bacteroides ihuae	C2-31	
SAMN05216357_1045 Porphyromonadaceae bacterium KH3CP3RA	C2-32	
Listeria riparia	C2-33	
Insolitispirillum peregrinum	C2-34	

[0810]

[0811] 상기 C2c2 종의 단백질 서열은 하기 표 3에 열거되어 있다.

표 3: 본원에서 확인된 C2c2 종의 서열 (상기, 표 2 참조)

c2c2-5	1	Lachnospiraceae bacterium MA2020	MQISKVNHKHVAVGQKDRERITGFIYNDPVGDEKSLEDVVA KRANDTKVLFNVFNTKDLYDSQESDKSEKDKEIISKGAKFV AKSFNSAITILKKQNKIYSTLTSQQVIKELKDKFGGARIYDDD IEEALTETLKKSFRKENVRNSIKVLIENAAGIRSSLSKDEEELI QEYFVKQLVEEYTKTKLQKNVVKSIKNQNMVIQPDSDSQVL SLSESRREKQSSAVSSDTLVNCKEKDVLKAFLTDYAVLDEDE RNSLLWKLRNLVNLYFYGSESIRDYSYTKEKSVWKEHDEQK ANKTLFIDEICHITKIGKNGKEQKVLDYEENRSRCRKQNINY YRSALNYAKNNTSGIFENEDSNHFWIHLIENEVERLYNGIEN GEEFKFETGYISEKVWKAVINHLSIKYIALGKAVYNYAMKEL SSPGDIEPGKIDDSYINGITSFDYEIIKAEESLQRDISMNVVFAT NYLACATVDTDKDFLLFSKEDIRSCTKKDGNLCKNIMQFWG GYSTWKNFCEEYLKDDKDALELLYSLKSMLYSMRNSSFHFS

[0812]

			TENVDNGSWDTELIGKLFEEDCNRAARIEKEKFYNNNLHMF YSSSLLEKVLERLYSSHHERASQVPSFNRVFVRKNFPSSLSEQ RITPKFTDSKDEQIWQSAVYYLCKEIYYNDFLQSKEAYKLFR EGVKNLDKNDINNQKAADSFKQAVVYYGKAIGNATLSQVC QAIMTEYNRQNNDGLKKKSAYAEKQNSNKYKHYPLFLKQV LQSAFWEYLDENKEIYGFISAQIHKSNVEIKAEDFIANYSSQQ YKKLVDKVKKTPELQKWYTLGRLINPRQANQFLGSIRNYVQ FVKDIQRRAKENGNPIRNYYEVLESDSIIKILEMCTKLNGTTS NDIHDYFRDEDEYAEYISQFVNFGDVHSGAALNAFCNSESEG KKNGIYYDGINPIVNRNWVLCKLYGSPDLISKIISRVNENMIH DFHKQEDLIREYQIKGICSNKKEQQDLRTFQVLKNRVELRDI VEYSEIINELYGQLIKWCYLRERDLMYFQLGFHYLCLNNASS KEADYIKINVDDRNISGAILYQIAAMYINGLPVYYKKDDMY VALKSGKKASDELNSNEQTSKKINYFLKYGNNILGDKKDQL YLAGLELFENVAEHENIIIFRNEIDHFHYFYDRDRSMLDLYSE VFDRFFTYDMKLRKNVVNMLYNILLDHNIVSSFVFETGEKK VGRGDSEVIKPSAKIRLRANNGVSSDVFTYKVGSKDELKIAT LPAKNEEFLLNVARLIYYPDMEAVSENMVREGVVKVEKSND KKGKISRGSNTRSSNQSKYNNKSKNRMNYSMGSIFEKMDLK FD
c2c2-6	2	Lachnospiraceae bacterium NK4A179	MKISKVREENRGAKLTVNAKTAVVSENRSQEGILYNDPSRY GKSRKNDEDRDRYIESRLKSSGKLYRIFNEDKNKRETDELQ WFLSEIVKKINRRNGLVLSDMLSVDDRAFEKAFEKYAELSYT NRRNKVSGSPAFETCGVDAATAERLKGIISETNFINRIKNNID NKVSEDIIDRIIAKYLKKSLCRERVKRGLKKLLMNAFDLPYS DPDIDVQRDFIDYVLEDFYHVRAKSQVSRSIKNMNMPVQPE GDGKFAITVSKGGTESGNKRSAEKEAFKKFLSDYASLDERV RDDMLRRMRRLVVLYFYGSDDSKLSDVNEKFDVWEDHAA RRVDNREFIKLPLENKLANGKTDKDAERIRKNTVKELYRNQ NIGCYRQAVKAVEEDNNGRYFDDKMLNMFFIHRIEYGVEKI YANLKQVTEFKARTGYLSEKIWKDLINYISIKYIAMGKAVYN YAMDELNASDKKEIELGKISEEYLSGISSFDYELIKAEEMLQR ETAVYVAFAARHLSSQTVELDSENSDFLLLKPKGTMDKNDK NKLASNNILNFLKDKETLRDTILQYFGGHSLWTDFPFDKYLA GGKDDVDFLTDLKDVIYSMRNDSFHYATENHNNGKWNKEL ISAMFEHETERMTVVMKDKFYSNNLPMFYKNDDLKKLLIDL YKDNVERASQVPSFNKVFVRKNFPALVRDKDNLGIELDLKA DADKGENELKFYNALYYMFKEIYYNAFLNDKNVRERFITKA TKVADNYDRNKERNLKDRIKSAGSDEKKKLREQLQNYIAEN DFGQRIKNIVQVNPDYTLAQICQLIMTEYNQQNNGCMQKKS AARKDINKDSYQHYKMLLLVNLRKAFLEFIKENYAFVLKPY KHDLCDKADFVPDFAKYVKPYAGLISRVAGSSELQKWYIVS RFLSPAQANHMLGFLHSYKQYVWDIYRRASETGTEINHSIAE DKIAGVDITDVDAVIDLSVKLCGTISSEISDYFKDDEVYAEYI SSYLDFEYDGGNYKDSLNRFCNSDAVNDQKVALYYDGEHP

[0813]

			KLNRNIILSKLYGERRFLEKITDRVSRSDIVEYYKLKKETSQY QTKGIFDSEDEQKNIKKFQEMKNIVEFRDLMDYSEIADELQG QLINWIYLRERDLMNFQLGYHYACLNNDSNKQATYVTLDY QGKKNRKINGAILYQICAMYINGLPLYYVDKDSSEWTVSDG KESTGAKIGEFYRYAKSFENTSDCYASGLEIFENISEHDNITEL RNYIEHFRYYSSFDRSFLGIYSEVFDRFFTYDLKYRKNVPTIL YNILLQHFVNVRFEFVSGKKMIGIDKKDRKIAKEKECARITIR EKNGVYSEQFTYKLKNGTVYVDARDKRYLQSIIRLLFYPEK VNMDEMIEVKEKKKPSDNNTGKGYSKRDRQQDRKEYDKY KEKKKKEGNFLSGMGGNINWDEINAQLKN
c2c2-7	3	[Clostridium] aminophilum DSM 10710	MKFSKVDHTRSAVGIQKATDSVHGMLYTDPKKQEVNDLDK RFDQLNVKAKRLYNVFNQSKAEEDDDEKRFGKVVKKLNRE LKDLLFHREVSRYNSIGNAKYNYYGIKSNPEEIVSNLGMVES LKGERDPQKVISKLLLYYLRKGLKPGTDGLRMILEASCGLRK LSGDEKELKVFLQTLDEDFEKKTFKKNLIRSIENQNMAVQPS NEGDPIIGITQGRFNSQKNEEKSAIERMMSMYADLNEDHRED VLRKLRRLNVLYFNVDTEKTEEPTLPGEVDTNPVFEVWHDH EKGKENDRQFATFAKILTEDRETRKKEKLAVKEALNDLKSAI RDHNIMAYRCSIKVTEQDKDGLFFEDQRINRFWIHHIESAVE RILASINPEKLYKLRIGYLGEKVWKDLLNYLSIKYIAVGKAV FHFAMEDLGKTGQDIELGKLSNSVSGGLTSFDYEQIRADETL QRQLSVEVAFAANNLFRAVVGQTGKKIEQSKSEENEEDFLL WKAEKIAESIKKEGEGNTLKSILQFFGGASSWDLNHFCAAYG NESSALGYETKFADDLRKAIYSLRNETFHFTTLNKGSFDWNA KLIGDMFSHEAATGIAVERTRFYSNNLPMFYRESDLKRIMDH LYNTYHPRASQVPSFNSVFVRKNFRLFLSNTLNTNTSFDTEV YQKWESGVYYLFKEIYYNSFLPSGDAHHLFFEGLRRIRKEAD NLPIVGKEAKKRNAVQDFGRRCDELKNLSLSAICQMIMTEY NEQNNGNRKVKSTREDKRKPDIFQHYKMLLLRTLQEAFAIYI RREEFKFIFDLPKTLYVMKPVEEFLPNWKSGMFDSLVERVK QSPDLQRWYVLCKFLNGRLLNQLSGVIRSYIQFAGDIQRRAK ANHNRLYMDNTQRVEYYSNVLEVVDFCIKGTSRFSNVFSDY FRDEDAYADYLDNYLQFKDEKIAEVSSFAALKTFCNEEEVK AGIYMDGENPVMQRNIVMAKLFGPDEVLKNVVPKVTREEIE EYYQLEKQIAPYRQNGYCKSEEDQKKLLRFQRIKNRVEFQTI TEFSEIINELLGQLISWSFLRERDLLYFQLGFHYLCLHNDTEK PAEYKEISREDGTVIRNAILHQVAAMYVGGLPVYTLADKKL AAFEKGEADCKLSISKDTAGAGKKIKDFFRYSKYVLIKDRML TDQNQKYTIYLAGLELFENTDEHDNITDVRKYVDHFKYYAT SDENAMSILDLYSEIHDRFFTYDMKYQKNVANMLENILLRH FVLIRPEFFTGSKKVGEGKKITCKARAQIEIAENGMRSEDFTY KLSDGKKNISTCMIAARDQKYLNTVARLLYYPHEAKKSIVD TREKKNNKKTNRGDGTFNKQKGTARKEKDNGPREFNDTGF SNTPFAGFDPFRNS

[0814]

[0815]

c2c2-8	5	Carnobacterium gallinarum DSM 4847	MRITKVKIKLDNKLYQVTMQKEEKYGTLKLNEESRKSTAEIL RLKKASFNKSFHSKTINSQKENKNATIKKNGDYISQIFEKLVG VDTNKNIRKPKMSLTDLKDLPKKDLALFIKRKFKNDDIVEIK NLDLISLFYNALQKVPGEHFTDESWADFCQEMMPYREYKNK FIERKIILLANSIEQNKGFSINPETFSKRKRVLHQWAIEVQERG DFSILDEKLSKLAEIYNFKKMCKRVQDELNDLEKSMKKGKN PEKEKEAYKKQKNFKIKTIWKDYPYKTHIGLIEKIKENEELN QFNIEIGKYFEHYFPIKKERCTEDEPYYLNSETIATTVNYQLK NALISYLMQIGKYKQFGLENQVLDSKKLQEIGIYEGFQTKFM DACVFATSSLKNIIEPMRSGDILGKREFKEAIATSSFVNYHHF FPYFPFELKGMKDRESELIPFGEQTEAKQMQNIWALRGSVQQ IRNEIFHSFDKNQKFNLPQLDKSNFEFDASENSTGKSQSYIET DYKFLFEAEKNQLEQFFIERIKSSGALEYYPLKSLEKLFAKKE MKFSLGSQVVAFAPSYKKLVKKGHSYQTATEGTANYLGLS YYNRYELKEESFQAQYYLLKLIYQYVFLPNFSQGNSPAFRET VKAILRINKDEARKKMKKNKKFLRKYAFEQVREMEFKETPD QYMSYLQSEMREEKVRKAEKNDKGFEKNITMNFEKLLMQIF VKGFDVFLTTFAGKELLLSSEEKVIKETEISLSKKINEREKTLK ASIQVEHQLVATNSAISYWLFCKLLDSRHLNELRNEMIKFKQ SRIKFNHTQHAELIQNLLPIVELTILSNDYDEKNDSQNVDVSA YFEDKSLYETAPYVQTDDRTRVSFRPILKLEKYHTKSLIEALL KDNPQFRVAATDIQEWMHKREEIGELVEKRKNLHTEWAEG QQTLGAEKREEYRDYCKKIDRFNWKANKVTLTYLSQLHYLI TDLLGRMVGFSALFERDLVYFSRSFSELGGETYHISDYKNLS GVLRLNAEVKPIKIKNIKVIDNEENPYKGNEPEVKPFLDRLH AYLENVIGIKAVHGKIRNQTAHLSVLQLELSMIESMNNLRDL MAYDRKLKNAVTKSMIKILDKHGMILKLKIDENHKNFEIESL IPKEIIHLKDKAIKTNQVSEEYCQLVLALLTTNPGNQLN
c2c2-9	6	Carnobacterium gallinarum DSM 4847	MRMTKVKINGSPVSMNRSKLNGHLVWNGTTNTVNILTKKE QSFAASFLNKTLVKADQVKGYKVLAENIFIIFEQLEKSNSEKP SVYLNNIRRLKEAGLKRFFKSKYHEEIKYTSEKNQSVPTKLN LIPLFFNAVDRIQEDKFDEKNWSYFCKEMSPYLDYKKSYLNR KKEILANSIQQNRGFSMPTAEEPNLLSKRKQLFQQWAMKFQ ESPLIQQNNFAVEQFNKEFANKINELAAVYNVDELCTAITEK LMNFDKDKSNKTRNFEIKKLWKQHPHNKDKALIKLFNQEG NEALNQFNIELGKYFEHYFPKTGKKESAESYYLNPQTIIKTVG YQLRNAFVQYLLQVGKLHQYNKGVLDSQTLQEIGMYEGFQ TKFMDACVFASSSLRNIIQATTNEDILTREKFKKELEKNVELK HDLFFKTEIVEERDENPAKKIAMTPNELDLWAIRGAVQRVR NQIFHQQINKRHEPNQLKVGSFENGDLGNVSYQKTIYQKLFD AEIKDIEIYFAEKIKSSGALEQYSMKDLEKLFSNKELTLSLGG QVVAFAPSYKKLYKQGYFYQNEKTIELEQFTDYDFSNDVFK ANYYLIKLIYHYVFLPQFSQANNKLFKDTVHYVIQQNKELNT TEKDKKNNKKIRKYAFEQVKLMKNESPEKYMQYLQREMQE ERTIKEAKKTNEEKPNYNFEKLLIQIFIKGFDTFLRNFDLNLNP

			AEELVGTVKEKAEGLRKRKERIAKILNVDEQIKTGDEEIAFW IFAKLLDARHLSELRNEMIKFKQSSVKKGLIKNGDLIEQMQPI LELCILSNDSESMEKESFDKIEVFLEKVELAKNEPYMQEDKL TPVKFRFMKQLEKYQTRNFIENLVIENPEFKVSEKIVLNWHE EKEKIADLVDKRTKLHEEWASKAREIEEYNEKIKKNKSKKL DKPAEFAKFAEYKIICEAIENFNRLDHKVRLTYLKNLHYLMI DLMGRMVGFSVLFERDFVYMGRSYSALKKQSIYLNDYDTF ANIRDWEVNENKHLFGTSSSDLTFQETAEFKNLKKPMENQL KALLGVTNHSFEIRNNIAHLHVLRNDGKGEGVSLLSCMNDL RKLMSYDRKLKNAVTKAIIKILDKHGMILKLTNNDHTKPFEI ESLKPKKIIHLEKSNHSFPMDQVSQEYCDLVKKMLVFTN
c2c2-10	7	Paludibacter propionicigenes WB4	MRVSKVKVKDGGKDKMVLVHRKTTGAQLVYSGQPVSNET SNILPEKKRQSFDLSTLNKTIIKFDTAKKQKLNVDQYKIVEKI FKYPKQELPKQIKAEEILPFLNHKFQEPVKYWKNGKEESFNL TLLIVEAVQAQDKRKLQPYYDWKTWYIQTKSDLLKKSIENN RIDLTENLSKRKKALLAWETEFTASGSIDLTHYHKVYMTDV LCKMLQDVKPLTDDKGKINTNAYHRGLKKALQNHQPAIFGT REVPNEANRADNQLSIYHLEVVKYLEHYFPIKTSKRRNTADD IAHYLKAQTLKTTIEKQLVNAIRANIIQQGKTNHHELKADTT SNDLIRIKTNEAFVLNLTGTCAFAANNIRNMVDNEQTNDILG KGDFIKSLLKDNTNSQLYSFFFGEGLSTNKAEKETQLWGIRG AVQQIRNNVNHYKKDALKTVFNISNFENPTITDPKQQTNYA DTIYKARFINELEKIPEAFAQQLKTGGAVSYYTIENLKSLLTT FQFSLCRSTIPFAPGFKKVFNGGINYQNAKQDESFYELMLEQ YLRKENFAEESYNARYFMLKLIYNNLFLPGFTTDRKAFADSV GFVQMQNKKQAEKVNPRKKEAYAFEAVRPMTAADSIADY MAYVQSELMQEQNKKEEKVAEETRINFEKFVLQVFIKGFDS FLRAKEFDFVQMPQPQLTATASNQQKADKLNQLEASITADC KLTPQYAKADDATHIAFYVFCKLLDAAHLSNLRNELIKFRES VNEFKFHHLLEIIEICLLSADVVPTDYRDLYSSEADCLARLRP FIEQGADITNWSDLFVQSDKHSPVIHANIELSVKYGTTKLLEQ IINKDTQFKTTEANFTAWNTAQKSIEQLIKQREDHHEQWVK AKNADDKEKQERKREKSNFAQKFIEKHGDDYLDICDYINTY NWLDNKMHFVHLNRLHGLTIELLGRMAGFVALFDRDFQFF DEQQIADEFKLHGFVNLHSIDKKLNEVPTKKIKEIYDIRNKIIQ INGNKINESVRANLIQFISSKRNYYNNAFLHVSNDEIKEKQM YDIRNHIAHFNYLTKDAADFSLIDLINELRELLHYDRKLKNA VSKAFIDLFDKHGMILKLKLNADHKLKVESLEPKKIYHLGSS AKDKPEYQYCTNQVMMAYCNMCRSLLEMKK
c2c2-11	9	Listeria weihenstephanensis FSL R9-0317	MLALLHQEVPSQKLHNLKSLNTESLTKLFKPKFQNMISYPPS KGAEHVQFCLTDIAVPAIRDLDEIKPDWGIFFEKLKPYTDWA ESYIHYKQTTIQKSIEQNKIQSPDSPRKLVLQKYVTAFLNGEP LGLDLVAKKYKLADLAESFKVVDLNEDKSANYKIKACLQQ HQRNILDELKEDPELNQYGIEVKKYIQRYFPIKRAPNRSKHA

[0816]

			RADFLKKELIESTVEQQFKNAVYHYVLEQGKMEAYELTDPK TKDLQDIRSGEAFSFKFINACAFASNNLKMILNPECEKDILGK GDFKKNLPNSTTQSDVVKKMIPFFSDEIQNVNFDEAIWAIRG SIQQIRNEVYHCKKHSWKSILKIKGFEFEPNNMKYTDSDMQK LMDKDIAKIPDFIEEKLKSSGIIRFYSHDKLQSIWEMKQGFSL LTTNAPFVPSFKRVYAKGHDYQTSKNRYYDLGLTTFDILEY GEEDFRARYFLTKLVYYQQFMPWFTADNNAFRDAANFVLR LNKNRQQDAKAFINIREVEEGEMPRDYMGYVQGQIAIHEDS TEDTPNHFEKFISQVFIKGFDSHMRSADLKFIKNPRNQGLEQS EIEEMSFDIKVEPSFLKNKDDYIAFWTFCKMLDARHLSELRN EMIKYDGHLTGEQEIIGLALLGVDSRENDWKQFFSSEREYEK IMKGYVGEELYQREPYRQSDGKTPILFRGVEQARKYGTETVI QRLFDASPEFKVSKCNITEWERQKETIEETIERRKELHNEWE KNPKKPQNNAFFKEYKECCDAIDAYNWHKNKTTLVYVNEL HHLLIEILGRYVGYVAIADRDFQCMANQYFKHSGITERVEY WGDNRLKSIKKLDTFLKKEGLFVSEKNARNHIAHLNYLSLK SECTLLYLSERLREIFKYDRKLKNAVSKSLIDILDRHGMSVVF ANLKENKHRLVIKSLEPKKLRHLGEKKIDNGYIETNQVSEEY CGIVKRLLEI
c2c2-12	10	Listeriaceae bacterium FSL M6-0635 = Listeria newyorkensis FSL M6-0635	MKITKMRVDGRTIVMERTSKEGQLGYEGIDGNKTTEIIFDKK KESFYKSILNKTVRKPDEKEKNRRKQAINKAINKEITELMLA VLHQEVPSQKLHNLKSLNTESLTKLFKPKFQNMISYPPSKGA EHVQFCLTDIAVPAIRDLDEIKPDWGIFFEKLKPYTDWAESYI HYKQTTIQKSIEQNKIQSPDSPRKLVLQKYVTAFLNGEPLGL DLVAKKYKLADLAESFKLVDLNEDKSANYKIKACLQQHQR NILDELKEDPELNQYGIEVKKYIQRYFPIKRAPNRSKHARADF LKKELIESTVEQQFKNAVYHYVLEQGKMEAYELTDPKTKDL QDIRSGEAFSFKFINACAFASNNLKMILNPECEKDILGKGNFK KNLPNSTTRSDVVKKMIPFFSDELQNVNFDEAIWAIRGSIQQI RNEVYHCKKHSWKSILKIKGFEFEPNNMKYADSDMQKLMD KDIAKIPEFIEEKLKSSGVVRFYRHDELQSIWEMKQGFSLLTT NAPFVPSFKRVYAKGHDYQTSKNRYYNLDLTTFDILEYGEE DFRARYFLTKLVYYQQFMPWFTADNNAFRDAANFVLRLNK NRQQDAKAFINIREVEEGEMPRDYMGYVQGQIAIHEDSIEDT PNHFEKFISQVFIKGFDRHMRSANLKFIKNPRNQGLEQSEIEE MSFDIKVEPSFLKNKDDYIAFWIFCKMLDARHLSELRNEMIK YDGHLTGEQEIIGLALLGVDSRENDWKQFFSSEREYEKIMKG YVVEELYQREPYRQSDGKTPILFRGVEQARKYGTETVIQRLF DANPEFKVSKCNLAEWERQKETIEETIKRRKELHNEWAKNP KKPQNNAFFKEYKECCDAIDAYNWHKNKTTLAYVNELHHL LIEILGRYVGYVAIADRDFQCMANQYFKHSGITERVEYWGD NRLKSIKKLDTFLKKEGLFVSEKNARNHIAHLNYLSLKSECT LLYLSERLREIFKYDRKLKNAVSKSLIDILDRHGMSVVFANL KENKHRLVIKSLEPKKLRHLGGKKIDGGYIETNQVSEEYCGI VKRLLEM

[0817]

c2c2-13	12	Leptotrichia wadei F0279	MKVTKVDGISHKKYIEEGKLVKSTSEENRTSERLSELLSIRLDIYIKNPDNASEEENRIRRENLKKFFSNKVLHLKDSVLYLKNRKEKNAVQDKNYSEEDISEYDLKNKNSFSVLKKILLNEDVNSEELEIFRKDVEAKLNKINSLKYSFEENKANYQKINENNVEKVGGKSKRNIIYDYYRESAKRNDYINNVQEAFDKLYKKEDIEKLFFLIENSKKHEKYKIREYYHKIIGRKNDKENFAKIIYEEIQNVNNIKELIEKIPDMSELKKSQVFYKYYLDKEELNDKNIKYAFCHFVEIEMSQLLKNYVYKRLSNISNDKIKRIFEYQNLKKLIENKLLNKLDTYVRNCGKYNYYLQVGEIATSDFIARNRQNEAFLRNIIGVSSVAYFSLRNILETENENDITGRMRGKTVKNNKGEEKYVSGEVDKIYNENKQNEVKENLKMFYSYDFNMDNKNEIEDFFANIDEAISSIRHGIVHFNLELEGKDIFAFKNIAPSEISKKMFQNEINEKKLKLKIFKQLNSANVFNYYEKDVIIKYLKNTKFNFVNKNIPFVPSFTKLYNKIEDLRNTLKFFWSVPKDKEEKDAQIYLLKNIYYGEFLNKFVKNSKVFFKITNEVIKINKQRNQKTGHYKYQKFENIEKTVPVEYLAIIQSREMINNQDKEEKNTYIDFIQQIFLKGFIDYLNKNNLKYIESNNNNDNNDIFSKIKIKKDNKEKYDKILKNYEKHNRNKEIPHEINEFVREIKLGKILKYTENLNMFYLILKLLNHKELTNLKGSLEKYQSANKEETFSDELELINLLNLDNNRVTEDFELEANEIGKFLDFNENKIKDRKELKKFDTNKIYFDGENIIKHRAFYNIKKYGMLNLLEKIADKAKYKISLKELKEYSNKKNEIEKNYTMQQNLHRKYARPKKDEKFNDEDYKEYEKAIGNIQKYTHLKNKVEFNELNLLQGLLLKILHRLVGYTSIWERDLRFRLKGEFPENHYIEEIFNFDNSKNVKYKSGQIVEKYINFYKELYKDNVEKRSIYSDKKVKKLKQEKKDLYIRNYIAHFNYIPHAEISLLEVLENLRKLLSYDRKLKNAIMKSIVDILKEYGFVATFKIGADKKIEIQTLESEKIVHLKNLKKKKLMTDRNSEELCELVKVMFEYKALE
c2c2-14	15	Rhodobacter capsulatus SB 1003	MQIGKVQGRTISEFGDPAGGLKRKISTDGKNRKELPAHLSSDPKALIGQWISGIDKIYRKPDSRKSDGKAIHSPTPSKMQFDARDDLGEAFWKLVSEAGLAQDSDYDQFKRRLHPYGDKFQPADSGAKLKFEADPPEPQAFHGRWYGAMSKRGNDAKELAAALYEHLHVDEKRIDGQPKRNPKTDKFAPGLVVARALGIESSVLPRGMARLARNWGEEEIQTYFVVDVAASVKEVAKAAVSAAQAFDPPRQVSGRSLSPKVGFALAEHLERVTGSKRCSFDPAAGPSVLALHDEVKKTYKRLCARGKNAARAFPADKTELLALMRHTHENRVRNQMVRMGRVSEYRGQQAGDLAQSHYWTSAGQTEIKESEIFVRLWVGAFALAGRSMKAWIDPMGKIVNTEKNDRDLTAAVNIRQVISNKEMVAEAMARRGIYFGETPELDRLGAEGNEGFVFALLRYLRGCRNQTFHLGARAGFLKEIRKELEKTRWGKAKEAEHVVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFVAHYASKEHFSTLYSEIVKAVKDAPEVSSGLPRLKLLLKRADGVRGYVHGLRDTRKHAFATKLPPPPAPRELDDPATKARYIALLRLYDGPFRAYASGITGTALAGPAARAKEAATALAQSVNVTKAYSDVMEGRTSRLRPPNDGETLREYLSALTGETATEFRVQIG

[0818]

			YESDSENARKQAEFIENYRRDMLAFMFEDYIRAKGFDWILKI EPGATAMTRAPVLPEPIDTRGQYEHWQAALYLVMHFVPASD VSNLLHQLRKWEALQGKYELVQDGDATDQADARREALDL VKRFRDVLVLFLKTGEARFEGRAAPFDLKPFRALFANPATFD RLFMATPTTARPAEDDPEGDGASEPELRVARTLRGLRQIARY NHMAVLSDLFAKHKVRDEEVARLAEIEDETQEKSQIVAAQE LRTDLHDKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGRV YLGDHLRLHRLMMDVIGRLIDYAGAYERDTGTFLINASKQL GAGADWAVTIAGAANTDARTQTRKDLAHFNVLDRADGTPD LTALVNRAREMMAYDRKRKNAVPRSILDMLARLGLTLKWQ MKDHLLQDATITQAAIKHLDKVRLTVGGPAAVTEARFSQDY LQMVAAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQSQPD QKPPNKAPSAGSRLPPPQVGEVYEGVVVKVIDTGSLGFLAVE GVAGNIGLHISRLRRIREDAIIVGRRYRFRVEIYVPPKSNTSKL NAADLVRID
c2c2-15	16	Rhodobacter capsulatus R121	MQIGKVQGRTISEFGDPAGGLKRKISTDGKNRKELPAHLSSD PKALIGQWISGIDKIYRKPDSRKSDGKAIHSPTPSKMQFDARD DLGEAFWKLVSEAGLAQDSDYDQFKRRLHPYGDKFQPADS GAKLKFEADPPEPQAFHGRWYGAMSKRGNDAKELAAALYE HLHVDEKRIDGQPKRNPKTDKFAPGLVVARALGIESSVLPRG MARLARNWGEEEIQTYFVVDVAASVKEVAKAAVSAAQAFD PPRQVSGRSLSPKVGFALAEHLERVTGSKRCSFDPAAGPSVL ALHDEVKKTYKRLCARGKNAARAFPADKTELLALMRHTHE NRVRNQMVRMGRVSEYRGQQAGDLAQSHYWTSAGQTEIK ESEIFVRLWVGAFALAGRSMKAWIDPMGKIVNTEKNDRDLT AAVNIRQVISNKEMVAEAMARRGIYFGETPELDRLGAEGNE GFVFALLRYLRGCRNQTFHLGARAGFLKEIRKELEKTRWGK AKEAEHVVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFV AHYASKEHFSTLYSEIVKAVKDAPEVSSGLPRLKLLLKRADG VRGYVHGLRDTRKHAFATKLPPPPAPRELDDPATKARYIALL RLYDGPFRAYASGITGTALAGPAARAKEAATALAQSVNVTK AYSDVMEGRSSRLRPPNDGETLREYLSALTGETATEFRVQIG YESDSENARKQAEFIENYRRDMLAFMFEDYIRAKGFDWILKI EPGATAMTRAPVLPEPIDTRGQYEHWQAALYLVMHFVPASD VSNLLHQLRKWEALQGKYELVQDGDATDQADARREALDL VKRFRDVLVLFLKTGEARFEGRAAPFDLKPFRALFANPATFD RLFMATPTTARPAEDDPEGDGASEPELRVARTLRGLRQIARY NHMAVLSDLFAKHKVRDEEVARLAEIEDETQEKSQIVAAQE LRTDLHDKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGRV YLGDHLRLHRLMMDVIGRLIDYAGAYERDTGTFLINASKQL GAGADWAVTIAGAANTDARTQTRKDLAHFNVLDRADGTPD LTALVNRAREMMAYDRKRKNAVPRSILDMLARLGLTLKWQ MKDHLLQDATITQAAIKHLDKVRLTVGGPAAVTEARFSQDY LQMVAAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQSQPD QKPPNKAPSAGSRLPPPQVGEVYEGVVVKVIDTGSLGFLAVE

[0819]

			GVAGNIGLHISRLRRIREDAIIVGRRYRFRVEIYVPPKSNTSKL NAADLVRID
c2c2-16	17	Rhodobacter capsulatus DE442	MQIGKVQGRTISEFGDPAGGLKRKISTDGKNRKELPAHLSSD PKALIGQWISGIDKIYRKPDSRKSDGKAIHSPTPSKMQFDARD DLGEAFWKLVSEAGLAQDSDYDQFKRRLHPYGDKFQPADS GAKLKFEADPPEPQAFHGRWYGAMSKRGNDAKELAAALYE HLHVDEKRIDGQPKRNPKTDKFAPGLVVARALGIESSVLPRG MARLARNWGEEEIQTYFVVDVAASVKEVAKAAVSAAQAFD PPRQVSGRSLSPKVGFALAEHLERVTGSKRCSFDPAAGPSVL ALHDEVKKTYKRLCARGKNAARAFPADKTELLALMRHTHE NRVRNQMVRMGRVSEYRGQQAGDLAQSHYWTSAGQTEIK ESEIFVRLWVGAFALAGRSMKAWIDPMGKIVNTEKNDRDLT AAVNIRQVISNKEMVAEAMARRGIYFGETPELDRLGAEGNE GFVFALLRYLRGCRNQTFHLGARAGFLKEIRKELEKTRWGK AKEAEHVVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFV AHYASKEHFSTLYSEIVKAVKDAPEVSSGLPRLKLLLKRADG VRGYVHGLRDTRKHAFATKLPPPPAPRELDDPATKARYIALL RLYDGPFRAYASGITGTALAGPAARAKEAATALAQSVNVTK AYSDVMEGRSSRLRPPNDGETLREYLSALTGETATEFRVQIG YESDSENARKQAEFIENYRRDMLAFMFEDYIRAKGFDWILKI EPGATAMTRAPVLPEPIDTRGQYEHWQAALYLVMHFVPASD VSNLLHQLRKWEALQGKYELVQDGDATDQADARREALDL VKRFRDVLVLFLKTGEARFEGRAAPFDLKPFRALFANPATFD RLFMATPTTARPAEDDPEGDGASEPELRVARTLRGLRQIARY NHMAVLSDLFAKHKVRDEEVARLAEIEDETQEKSQIVAAQE LRTDLHDKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGRV YLGDHLRLHRLMMDVIGRLIDYAGAYERDTGTFLINASKQL GAGADWAVTIAGAANTDARTQTRKDLAHFNVLDRADGTPD LTALVNRAREMMAYDRKRKNAVPRSILDMLARLGLTLKWQ MKDHLLQDATITQAAIKHLDKVRLTVGGPAAVTEARFSQDY LQMVAAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQSQPD QKPPNKAPSAGSRLPPPQVGEVYEGVVVKVIDTGSLGFLAVE GVAGNIGLHISRLRRIREDAIIVGRRYRFRVEIYVPPKSNTSKL NAADLVRID
c2-2			MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNI NENNNKEKIDNNKFIRKYINYKKNDNILKEFTRKFHAGNILF KLKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKLKALGITK KKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLND CSIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFE NRYYEEHLREKLLKDDKIDVILTNFMEIREKIKSNLEILGFVK FYLNVGGDKKKSKNKKMLVEKILNINVDLTVEDIADFVIKEL EFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFK IERENKKDKIVKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEK ELKKGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKELYKIIY

[0820]

			RYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKILKRV KQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLE LITFFASTNMELNKIFSRENINNDENIDFFGGDREKNYVLDKK ILNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRILHAISK ERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKK NIITKINDIKISEENNNDIKYLPSFSKVLPEILNLYRNNPKNEPF DTIETEKIVLNALIYVNKELYKKLILEDDLEENESKNIFLQELK KTLGNIDEIDENIIENYYKNAQISASKGNNKAIKKYQKKVIEC YIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITV KTSDKTIVINDDFEYIISIFALLNSNAVINKIRNRFFATSVWLN TSEYQNIIDILDEIMQLNTLRNECITENWNLNLEEFIQKMKEIE KDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLEKK LEKIVIFDDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQ YIKDKDQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENEN KFQEIYYPKERKNELYIYKKNLFLNIGNPNFDKIYGLISNDIK MADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYK EKYIKKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRD LVEFNYLNKIESYLIDINWKLAIQMARFERDMHYIVNGLREL GIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKKF EKICYGFGIDLSENSEINKPENESIRNYISHFYIVRNPFADYSIA EQIDRVSNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKK KFKLIGNNDILERLMKPKKVSVLELESYNSDYIKNLIIELLTKI ENTNDTL
c2-3		L wadei (Lw2)	MKVTKVDGISHKKYIEEGKLVKSTSEENRTSERLSELLSIRLD IYIKNPDNASEEENRIRRENLKKFFSNKVLHLKDSVLYLKNR KEKNAVQDKNYSEEDISEYDLKNKNSFSVLKKILLNEDVNSE ELEIFRKDVEAKLNKINSLKYSFEENKANYQKINENNVEKVG GKSKRNIIYDYYRESAKRNDYINNVQEAFDKLYKKEDIEKLF FLIENSKKHEKYKIREYYHKIIGRKNDKENFAKIIYEEIQNVN NIKELIEKIPDMSELKKSQVFYKYYLDKEELNDKNIKYAFCH FVEIEMSQLLKNYVYKRLSNISNDKIKRIFEYQNLKKLIENKL LNKLDTYVRNCGKYNYYLQVGEIATSDFIARNRQNEAFLRNI IGVSSVAYFSLRNILETENENDITGRMRGKTVKNNKGEEKYV SGEVDKIYNENKQNEVKENLKMFYSYDFNMDNKNEIEDFFA NIDEAISSIRHGIVHFNLELEGKDIFAFKNIAPSEISKKMFQNEI NEKKLKLKIFKQLNSANVFNYYEKDVIIKYLKNTKFNFVNK NIPFVPSFTKLYNKIEDLRNTLKFFWSVPKDKEEKDAQIYLLK NIYYGEFLNKFVKNSKVFFKITNEVIKINKQRNQKTGHYKYQ KFENIEKTVPVEYLAIIQSREMINNQDKEEKNTYIDFIQQIFLK GFIDYLNKNNLKYIESNNNNDNNDIFSKIKIKKDNKEKYDKIL KNYEKHNRNKEIPHEINEFVREIKLGKILKYTENLNMFYLILK LLNHKELTNLKGSLEKYQSANKEETFSDELELINLLNLDNNR VTEDFELEANEIGKFLDFNENKIKDRKELKKFDTNKIYFDGE NIIKHRAFYNIKKYGMLNLLEKIADKAKYKISLKELKEYSNK KNEIEKNYTMQQNLHRKYARPKKDEKFNDEDYKEYEKAIG

[0821]

			NIQKYTHLKNKVEFNELNLLQGLLLKILHRLVGYTSIWERDL RFRLKGEFPENHYIEEIFNFDNSKNVKYKSGQIVEKYINFYKE LYKDNVEKRSIYSDKKVKKLKQEKKDLYIRNYIAHFNYIPHA EISLLEVLENLRKLLSYDRKLKNAIMKSIVDILKEYGFVATFK IGADKKIEIQTLESEKIVHLKNLKKKKLMTDRNSEELCELVK VMFEYKALEKRPAATKKAGQAKKKKGSYPYDVPDYAYPY DVPDYAYPYDVPDYA*
c2-4		Listeria seeligeri	MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTMRITKVEV DRKKVLISRDKNGGKLVYENEMQDNTEQIMHHKKSSFYKS VVNKTICRPEQKQMKKLVHGLLQENSQEKIKVSDVTKLNIS NFLNHRFKKSLYYFPENSPDKSEEYRIEINLSQLLEDSLKKQQ GTFICWESFSKDMELYINWAENYISSKTKLIKKSIRNNRIQST ESRSGQLMDRYMKDILNKNKPFDIQSVSEKYQLEKLTSALK ATFKEAKKNDKEINYKLKSTLQNHERQIIEELKENSELNQFNI EIRKHLETYFPIKKTNRKVGDIRNLEIGEIQKIVNHRLKNKIVQ RILQEGKLASYEIESTVNSNSLQKIKIEEAFALKFINACLFASN NLRNMVYPVCKKDILMIGEFKNSFKEIKHKKFIRQWSQFFSQ EITVDDIELASWGLRGAIAPIRNEIIHLKKHSWKKFFNNPTFK VKKSKIINGKTKDVTSEFLYKETLFKDYFYSELDSVPELIINK MESSKILDYYSSDQLNQVFTIPNFELSLLTSAVPFAPSFKRVY LKGFDYQNQDEAQPDYNLKLNIYNEKAFNSEAFQAQYSLFK MVYYQVFLPQFTTNNDLFKSSVDFILTLNKERKGYAKAFQDI RKMNKDEKPSEYMSYIQSQLMLYQKKQEEKEKINHFEKFIN QVFIKGFNSFIEKNRLTYICHPTKNTVPENDNIEIPFHTDMDDS NIAFWLMCKLLDAKQLSELRNEMIKFSCSLQSTEEISTFTKAR EVIGLALLNGEKGCNDWKELFDDKEAWKKNMSLYVSEELL QSLPYTQEDGQTPVINRSIDLVKKYGTETILEKLFSSSDDYKV SAKDIAKLHEYDVTEKIAQQESLHKQWIEKPGLARDSAWTK KYQNVINDISNYQWAKTKVELTQVRHLHQLTIDLLSRLAGY MSIADRDFQFSSNYILERENSEYRVTSWILLSENKNKNKYND YELYNLKNASIKVSSKNDPQLKVDLKQLRLTLEYLELFDNRL KEKRNNISHFNYLNGQLGNSILELFDDARDVLSYDRKLKNA VSKSLKEILSSHGMEVTFKPLYQTNHHLKIDKLQPKKIHHLG EKSTVSSNQVSNEYCQLVRTLLTMK
C2-17		Leptotrichia buccalis C-1013-b	MKVTKVGGISHKKYTSEGRLVKSESEENRTDERLSALLNMR LDMYIKNPSSTETKENQKRIGKLKKFFSNKMVYLKDNTLSL KNGKKENIDREYSETDILESDVRDKKNFAVLKKIYLNENVNS EELEVFRNDIKKKLNKINSLKYSFEKNKANYQKINENNIEKV EGKSKRNIIYDYYRESAKRDAYVSNVKEAFDKLYKEEDIAK LVLEIENLTKLEKYKIREFYHEIIGRKNDKENFAKIIYEEIQNV NNMKELIEKVPDMSELKKSQVFYKYYLDKEELNDKNIKYAF CHFVEIEMSQLLKNYVYKRLSNISNDKIKRIFEYQNLKKLIEN KLLNKLDTYVRNCGKYNYYLQDGEIATSDFIARNRQNEAFL RNIIGVSSVAYFSLRNILETENENDITGRMRGKTVKNNKGEE

[0822]

			KYVSGEVDKIYNENKKNEVKENLKMFYSYDFNMDNKNEIE DFFANIDEAISSIRHGIVHFNLELEGKDIFAFKNIAPSEISKKMF QNEINEKKLKLKIFRQLNSANVFRYLEKYKILNYLKRTRFEF VNKNIPFVPSFTKLYSRIDDLKNSLGIYWKTPKTNDDNKTKEI IDAQIYLLKNIYYGEFLNYFMSNNGNFFEISKEIIELNKNDKR NLKTGFYKLQKFEDIQEKIPKEYLANIQSLYMINAGNQDEEE KDTYIDFIQKIFLKGFMTYLANNGRLSLIYIGSDEETNTSLAE KKQEFDKFLKKYEQNNNIKIPYEINEFLREIKLGNILKYTERL NMFYLILKLLNHKELTNLKGSLEKYQSANKEEAFSDQLELIN LLNLDNNRVTEDFELEADEIGKFLDFNGNKVKDNKELKKFD TNKIYFDGENIIKHRAFYNIKKYGMLNLLEKIADKAGYKISIE ELKKYSNKKNEIEKNHKMQENLHRKYARPRKDEKFTDEDY ESYKQAIENIEEYTHLKNKVEFNELNLLQGLLLRILHRLVGY TSIWERDLRFRLKGEFPENQYIEEIFNFENKKNVKYKGGQIVE KYIKFYKELHQNDEVKINKYSSANIKVLKQEKKDLYIRNYIA HFNYIPHAEISLLEVLENLRKLLSYDRKLKNAVMKSVVDILK EYGFVATFKIGADKKIGIQTLESEKIVHLKNLKKKKLMTDRN SEELCKLVKIMFEYKMEEKKSEN
C2-18		Herbinix hemicellulosilytica	MKLTRRRISGNSVDQKITAAFYRDMSQGLLYYDSEDNDCTD KVIESMDFERSWRGRILKNGEDDKNPFYMFVKGLVGSNDKI VCEPIDVDSDPDNLDILINKNLTGFGRNLKAPDSNDTLENLIR KIQAGIPEEEVLPELKKIKEMIQKDIVNRKEQLLKSIKNNRIPF SLEGSKLVPSTKKMKWLFKLIDVPNKTFNEKMLEKYWEIYD YDKLKANITNRLDKTDKKARSISRAVSEELREYHKNLRTNY NRFVSGDRPAAGLDNGGSAKYNPDKEEFLLFLKEVEQYFKK YFPVKSKHSNKSKDKSLVDKYKNYCSYKVVKKEVNRSIINQ LVAGLIQQGKLLYYFYYNDTWQEDFLNSYGLSYIQVEEAFK KSVMTSLSWGINRLTSFFIDDSNTVKFDDITTKKAKEAIESNY FNKLRTCSRMQDHFKEKLAFFYPVYVKDKKDRPDDDIENLI VLVKNAIESVSYLRNRTFHFKESSLLELLKELDDKNSGQNKI DYSVAAEFIKRDIENLYDVFREQIRSLGIAEYYKADMISDCFK TCGLEFALYSPKNSLMPAFKNVYKRGANLNKAYIRDKGPKE TGDQGQNSYKALEEYRELTWYIEVKNNDQSYNAYKNLLQLI YYHAFLPEVRENEALITDFINRTKEWNRKETEERLNTKNNKK HKNFDENDDITVNTYRYESIPDYQGESLDDYLKVLQRKQMA RAKEVNEKEEGNNNYIQFIRDVVVWAFGAYLENKLKNYKN ELQPPLSKENIGLNDTLKELFPEEKVKSPFNIKCRFSISTFIDNK GKSTDNTSAEAVKTDGKEDEKDKKNIKRKDLLCFYLFLRLL DENEICKLQHQFIKYRCSLKERRFPGNRTKLEKETELLAELEE LMELVRFTMPSIPEISAKAESGYDTMIKKYFKDFIEKKVFKNP KTSNLYYHSDSKTPVTRKYMALLMRSAPLHLYKDIFKGYYL ITKKECLEYIKLSNIIKDYQNSLNELHEQLERIKLKSEKQNGK DSLYLDKKDFYKVKEYVENLEQVARYKHLQHKINFESLYRI FRIHVDIAARMVGYTQDWERDMHFLFKALVYNGVLEERRF EAIFNNNDDNNDGRIVKKIQNNLNNKNRELVSMLCWNKKL

[0823]

			NKNEFGAIIWKRNPIAHLNHFTQTEQNSKSSLESLINSLRILLA YDRKRQNAVTKTINDLLLNDYHIRIKWEGRVDEGQIYFNIKE KEDIENEPIIHLKHLHKKDCYIYKNSYMFDKQKEWICNGIKE EVYDKSILKCIGNLFKFDYEDKNKSSANPKHT
C2-19		[Eubacterium] rectale	MLRRDKEVKKLYNVFNQIQVGTKPKKWNNDEKLSPEENER RAQQKNIKMKNYKWREACSKYVESSQRIINDVIFYSYRKAK NKLRYMRKNEDILKKMQEAEKLSKFSGGKLEDFVAYTLRKS LVVSKYDTQEFDSLAAMVVFLECIGKNNISDHEREIVCKLLE LIRKDFSKLDPNVKGSQGANIVRSVRNQNMIVQPQGDRFLFP QVYAKENETVTNKNVEKEGLNEFLLNYANLDDEKRAESLR KLRRILDVYFSAPNHYEKDMDITLSDNIEKEKFNVWEKHEC GKKETGLFVDIPDVLMEAEAENIKLDAVVEKRERKVLNDRV RKQNIICYRYTRAVVEKYNSNEPLFFENNAINQYWIHHIENA VERILKNCKAGKLFKLRKGYLAEKVWKDAINLISIKYIALGK AVYNFALDDIWKDKKNKELGIVDERIRNGITSFDYEMIKAHE NLQRELAVDIAFSVNNLARAVCDMSNLGNKESDFLLWKRN DIADKLKNKDDMASVSAVLQFFGGKSSWDINIFKDAYKGKK KYNYEVRFIDDLRKAIYCARNENFHFKTALVNDEKWNTELF GKIFERETEFCLNVEKDRFYSNNLYMFYQVSELRNMLDHLY SRSVSRAAQVPSYNSVIVRTAFPEYITNVLGYQKPSYDADTL GKWYSACYYLLKEIYYNSFLQSDRALQLFEKSVKTLSWDDK KQQRAVDNFKDHFSDIKSACTSLAQVCQIYMTEYNQQNNQI KKVRSSNDSIFDQPVYQHYKVLLKKAIANAFADYLKNNKDL FGFIGKPFKANEIREIDKEQFLPDWTSRKYEALCIEVSGSQEL QKWYIVGKFLNARSLNLMVGSMRSYIQYVTDIKRRAASIGN ELHVSVHDVEKVEKWVQVIEVCSLLASRTSNQFEDYFNDKD DYARYLKSYVDFSNVDMPSEYSALVDFSNEEQSDLYVDPKN PKVNRNIVHSKLFAADHILRDIVEPVSKDNIEEFYSQKAEIAY CKIKGKEITAEEQKAVLKYQKLKNRVELRDIVEYGEIINELLG QLINWSFMRERDLLYFQLGFHYDCLRNDSKKPEGYKNIKVD ENSIKDAILYQIIGMYVNGVTVYAPEKDGDKLKEQCVKGGV GVKVSAFHRYSKYLGLNEKTLYNAGLEIFEVVAEHEDIINLR NGIDHFKYYLGDYRSMLSIYSEVFDRFFTYDIKYQKNVLNLL QNILLRHNVIVEPILESGFKTIGEQTKPGAKLSIRSIKSDTFQY KVKGGTLITDAKDERYLETIRKILYYAENEEDNLKKSVVVTN ADKYEKNKESDDQNKQKEKKNKDNKGKKNEETKSDAEKN NNERLSYNPFANLNFKLSN
C2-20		Eubacteriaceae bacterium CHKCI004	MKISKESHKRTAVAVMEDRVGGVVYVPGGSGIDLSNNLKK RSMDTKSLYNVFNQIQAGTAPSEYEWKDYLSEAENKKREAQ KMIQKANYELRRECEDYAKKANLAVSRIIFSKKPKKIFSDDDI ISHMKKQRLSKFKGRMEDFVLIALRKSLVVSTYNQEVFDSR KAATVFLKNIGKKNISADDERQIKQLMALIREDYDKWNPDK DSSDKKESSGTKVIRSIEHQNMVIQPEKNKLSLSKISNVGKKT KTKQKEKAGLDAFLKEYAQIDENSRMEYLKKLRRLLDTYFA

[0824]

			APSSYIKGAAVSLPENINFSSELNVWERHEAAKKVNINFVEIP ESLLNAEQNNNKINKVEQEHSLEQLRTDIRRRNITCYHFANA LAADERYHTLFFENMAMNQFWIHHMENAVERILKKCNVGT LFKLRIGYLSEKVWKDMLNLLSIKYIALGKAVYHFALDDIW KADIWKDASDKNSGKINDLTLKGISSFDYEMVKAQEDLQRE MAVGVAFSTNNLARVTCKMDDLSDAESDFLLWNKEAIRRH VKYTEKGEILSAILQFFGGRSLWDESLFEKAYSDSNYELKFL DDLKRAIYAARNETFHFKTAAIDGGSWNTRLFGSLFEKEAGL CLNVEKNKFYSNNLVLFYKQEDLRVFLDKLYGKECSRAAQI PSYNTILPRKSFSDFMKQLLGLKEPVYGSAILDQWYSACYYL FKEVYYNLFLQDSSAKALFEKAVKALKGADKKQEKAVESFR KRYWEISKNASLAEICQSYITEYNQQNNKERKVRSANDGMF NEPIYQHYKMLLKEALKMAFASYIKNDKELKFVYKPTEKLF EVSQDNFLPNWNSEKYNTLISEVKNSPDLQKWYIVGKFMNA RMLNLLLGSMRSYLQYVSDIQKRAAGLGENQLHLSAENVG QVKKWIQVLEVCLLLSVRISDKFTDYFKDEEEYASYLKEYV DFEDSAMPSDYSALLAFSNEGKIDLYVDASNPKVNRNIIQAK LYAPDMVLKKVVKKISQDECKEFNEKKEQIMQFKNKGDEVS WEEQQKILEYQKLKNRVELRDLSEYGELINELLGQLINWSYL RERDLLYFQLGFHYSCLMNESKKPDAYKTIRRGTVSIENAVL YQIIAMYINGFPVYAPEKGELKPQCKTGSAGQKIRAFCQWAS MVEKKKYELYNAGLELFEVVKEHDNIIDLRNKIDHFKYYQG NDSILALYGEIFDRFFTYDMKYRNNVLNHLQNILLRHNVIIKP IISKDKKEVGRGKMKDRAAFLLEEVSSDRFTYKVKEGERKID AKNRLYLETVRDILYFPNRAVNDKGEDVIICSKKAQDLNEK KADRDKNHDKSKDTNQKKEGKNQEEKSENKEPYSDRMTW KPFAGIKLE
C2-21		Blautia sp. Marseille-P2398	MKISKVDHVKSGIDQKLSSQRGMLYKQPQKKYEGKQLEEH VRNLSRKAKALYQVFPVSGNSKMEKELQIINSFIKNILLRLDS GKTSEEIVGYINTYSVASQISGDHIQELVDQHLKESLRKYTCV GDKRIYVPDIIVALLKSKFNSETLQYDNSELKILIDFIREDYLK EKQIKQIVHSIENNSTPLRIAEINGQKRLIPANVDNPKKSYIFE FLKEYAQSDPKGQESLLQHMRYLILLYLYGPDKITDDYCEEI EAWNFGSIVMDNEQLFSEEASMLIQDRIYVNQQIEEGRQSKD TAKVKKNKSKYRMLGDKIEHSINESVVKHYQEACKAVEEK DIPWIKYISDHVMSVYSSKNRVDLDKLSLPYLAKNTWNTWI SFIAMKYVDMGKGVYHFAMSDVDKVGKQDNLIIGQIDPKFS DGISSFDYERIKAEDDLHRSMSGYIAFAVNNFARAICSDEFRK KNRKEDVLTVGLDEIPLYDNVKRKLLQYFGGASNWDDSIIDI IDDKDLVACIKENLYVARNVNFHFAGSEKVQKKQDDILEEIV RKETRDIGKHYRKVFYSNNVAVFYCDEDIIKLMNHLYQREK PYQAQIPSYNKVISKTYLPDLIFMLLKGKNRTKISDPSIMNMF RGTFYFLLKEIYYNDFLQASNLKEMFCEGLKNNVKNKKSEK PYQNFMRRFEELENMGMDFGEICQQIMTDYEQQNKQKKKT ATAVMSEKDKKIRTLDNDTQKYKHFRTLLYIGLREAFIIYLK

[0825]

[0826]

			DEKNKEWYEFLREPVKREQPEEKEFVNKWKLNQYSDCSELI LKDSLAAAWYVVAHFINQAQLNHLIGDIKNYIQFISDIDRRA KSTGNPVSESTEIQIERYRKILRVLEFAKFFCGQITNVLTDYY QDENDFSTHVGHYVKFEKKNMEPAHALQAFSNSLYACGKE KKKAGFYYDGMNPIVNRNITLASMYGNKKLLENAMNPVTE QDIRKYYSLMAELDSVLKNGAVCKSEDEQKNLRHFQNLKN RIELVDVLTLSELVNDLVAQLIGWVYIRERDMMYLQLGLHY IKLYFTDSVAEDSYLRTLDLEEGSIADGAVLYQIASLYSFNLP MYVKPNKSSVYCKKHVNSVATKFDIFEKEYCNGDETVIENG LRLFENINLHKDMVKFRDYLAHFKYFAKLDESILELYSKAYD FFFSYNIKLKKSVSYVLTNVLLSYFINAKLSFSTYKSSGNKTV QHRTTKISVVAQTDYFTYKLRSIVKNKNGVESIENDDRRCEV VNIAARDKEFVDEVCNVINYNSDK
C2-22		Leptotrichia sp. oral taxon 879 str. F0557	MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNI NENNNKEKIDNNKFIGEFVNYKKNNNVLKEFKRKFHAGNIL FKLKGKEEIIRIENNDDFLETEEVVLYIEVYGKSEKLKALEITK KKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLND CSIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFE NRYYEEHLREKLLKDNKIDVILTNFMEIREKIKSNLEIMGFVK FYLNVSGDKKKSENKKMFVEKILNTNVDLTVEDIVDFIVKEL KFWNITKRIEKVKKFNNEFLENRRNRTYIKSYVLLDKHEKFK IERENKKDKIVKFFVENIKNNSIKEKIEKILAEFKINELIKKLEK ELKKGNCDTEIFGIFKKHYKVNFDSKKFSNKSDEEKELYKIIY RYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKILKRV KQYTLEHIMYLGKLRHNDIVKMTVNTDDFSRLHAKEELDLE LITFFASTNMELNKIFNGKEKVTDFFGFNLNGQKITLKEKVPS FKLNILKKLNFINNENNIDEKLSHFYSFQKEGYLLRNKILHNS YGNIQETKNLKGEYENVEKLIKELKVSDEEISKSLSLDVIFEG KVDIINKINSLKIGEYKDKKYLPSFSKIVLEITRKFREINKDKL FDIESEKIILNAVKYVNKILYEKITSNEENEFLKTLPDKLVKKS NNKKENKNLLSIEEYYKNAQVSSSKGDKKAIKKYQNKVTNA YLEYLENTFTEIIDFSKFNLNYDEIKTKIEERKDNKSKIIIDSIST NINITNDIEYIISIFALLNSNTYINKIRNRFFATSVWLEKQNGTK EYDYENIISILDEVLLINLLRENNITDILDLKNAIIDAKIVENDE TYIKNYIFESNEEKLKKRLFCEELVDKEDIRKIFEDENFKFKSF IKKNEIGNFKINFGILSNLECNSEVEAKKIIGKNSKKLESFIQNI IDEYKSNIRTLFSSEFLEKYKEEIDNLVEDTESENKNKFEKIYY PKEHKNELYIYKKNLFLNIGNPNFDKIYGLISKDIKNVDTKIL FDDDIKKNKISEIDAILKNLNDKLNGYSNDYKAKYVNKLKE NDDFFAKNIQNENYSSFGEFEKDYNKVSEYKKIRDLVEFNYL NKIESYLIDINWKLAIQMARFERDMHYIVNGLRELGIIKLSGY NTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKKFEKICYGF GIDLSENSEINKPENESIRNYISHFYIVRNPFADYSIAEQIDRVS NLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFRLIGN NDILERLMKPKKVSVLELESYNSDYIKNLIIELLTKIENTNDTL

C2-23		Lachnospiraceae bacterium NK4A144	MKISKVDHTRMAVAKGNQHRRDEISGILYKDPTKTGSIDFDE RFKKLNCSAKILYHVFNGIAEGSNKYKNIVDKVNNNLDRVL FTGKSYDRKSIIDIDTVLRNVEKINAFDRISTEEREQIIDDLLEI QLRKGLRKGKAGLREVLLIGAGVIVRTDKKQEIADFLEILDE DFNKTNQAKNIKLSIENQGLVVSPVSRGEERIFDVSGAQKGK SSKKAQEKEALSAFLLDYADLDKNVRFEYLRKIRRLINLYFY VKNDDVMSLTEIPAEVNLEKDFDIWRDHEQRKEENGDFVGC PDILLADRDVKKSNSKQVKIAERQLRESIREKNIKRYRFSIKTI EKDDGTYFFANKQISVFWIHRIENAVERILGSINDKKLYRLRL GYLGEKVWKDILNFLSIKYIAVGKAVFNFAMDDLQEKDRDI EPGKISENAVNGLTSFDYEQIKADEMLQREVAVNVAFAANN LARVTVDIPQNGEKEDILLWNKSDIKKYKKNSKKGILKSILQ FFGGASTWNMKMFEIAYHDQPGDYEENYLYDIIQIIYSLRNK SFHFKTYDHGDKNWNRELIGKMIEHDAERVISVEREKFHSN NLPMFYKDADLKKILDLLYSDYAGRASQVPAFNTVLVRKNF PEFLRKDMGYKVHFNNPEVENQWHSAVYYLYKEIYYNLFL RDKEVKNLFYTSLKNIRSEVSDKKQKLASDDFASRCEEIEDR SLPEICQIIMTEYNAQNFGNRKVKSQRVIEKNKDIFRHYKML LIKTLAGAFSLYLKQERFAFIGKATPIPYETTDVKNFLPEWKS GMYASFVEEIKNNLDLQEWYIVGRFLNGRMLNQLAGSLRSY IQYAEDIERRAAENRNKLFSKPDEKIEACKKAVRVLDLCIKIS TRISAEFTDYFDSEDDYADYLEKYLKYQDDAIKELSGSSYAA LDHFCNKDDLKFDIYVNAGQKPILQRNIVMAKLFGPDNILSE VMEKVTESAIREYYDYLKKVSGYRVRGKCSTEKEQEDLLKF QRLKNAVEFRDVTEYAEVINELLGQLISWSYLRERDLLYFQL GFHYMCLKNKSFKPAEYVDIRRNNGTIIHNAILYQIVSMYIN GLDFYSCDKEGKTLKPIETGKGVGSKIGQFIKYSQYLYNDPS YKLEIYNAGLEVFENIDEHDNITDLRKYVDHFKYYAYGNKM SLLDLYSEFFDRFFTYDMKYQKNVVNVLENILLRHFVIFYPK FGSGKKDVGIRDCKKERAQIEISEQSLTSEDFMFKLDDKAGE EAKKFPARDERYLQTIAKLLYYPNEIEDMNRFMKKGETINK KVQFNRKKKITRKQKNNSSNEVLSSTMGYLFKNIKL
C2-24		Chloroflexus aggregans	MTDQVRREEVAAGELADTPLAAAQTPAADAAVAATPAPAE AVAPTPEQAVDQPATTGESEAPVTTAQAAAHEAEPAEATGA SFTPVSEQQPQKPRRLKDLQPGMELEGKVTSIALYGIFVDVG VGRDGLVHISEMSDRRIDTPSELVQIGDTVKVWVKSVDLDA RRISLTMLNPSRGEKPRRSRQSQPAQPQPRRQEVDREKLASL KVGEIVEGVITGFAPFGAFADIGVGKDGLIHISELSEGRVEKP EDAVKVGERYQFKVLEIDGEGTRISLSLRRAQRTQRMQQLEP GQIIEGTVSGIATFGAFVDIGVGRDGLVHISALAPHRVAKVED VVKVGDKVKVKVLGVDPQSKRISLTMRLEEEQPATTAGDEA AEPAEEVTPTRRGNLERFAAAAQTARERSERGERSERGERRE RRERRPAQSSPDTYIVGEDDDESFEGNATIEDLLTKFGGSSSR RDRDRRRRHEDDDDEEMERPSNRRQREAIRRTLQQIGYDE

[0827]

C2-25		Demequina aurantiaca	MDLTWHALLILFIVALLAGFLDTLAGGGGLLTVPALLLTGIP PLQALGTNKLQSSFGTGMATYQVIRKKRVHWRDVRWPMV WAFLGSAAGAVAVQFIDTDALLIIIPVVLALVAAYFLFVPKS HLPPPEPRMSDPAYEATLVPIIGAYDGAFGPGTGSLYALSGV ALRAKTLVQSTAIAKTLNFATNFAALLVFAFAGHMLWTVGA VMIAGQLIGAYAGSHMLFRVNPLVLRVLIVVMSLGMLIRVL LD
C2-26		Thalassospira sp. TSL5-1	MRIIKPYGRSHVEGVATQEPRRKLRLNSSPDISRDIPGFAQSH DALIIAQWISAIDKIATKPKPDKKPTQAQINLRTTLGDAAWQ HVMAENLLPAATDPAIREKLHLIWQSKIAPWGTARPQAEKD GKPTPKGGWYERFCGVLSPEAITQNVARQIAKDIYDHLHVA AKRKGREPAKQGESSNKPGKFKPDRKRGLIEERAESIAKNAL RPGSHAPCPWGPDDQATYEQAGDVAGQIYAAARDCLEEKK RRSGNRNTSSVQYLPRDLAAKILYAQYGRVFGPDTTIKAALD EQPSLFALHKAIKDCYHRLINDARKRDILRILPRNMAALFRL VRAQYDNRDINALIRLGKVIHYHASEQGKSEHHGIRDYWPS QQDIQNSRFWGSDGQADIKRHEAFSRIWRHIIALASRTLHDW ADPHSQKFSGENDDILLLAKDAIEDDVFKAGHYERKCDVLF GAQASLFCGAEDFEKAILKQAITGTGNLRNATFHFKGKVRFE KELQELTKDVPVEVQSAIAALWQKDAEGRTRQIAETLQAVL AGHFLTEEQNRHIFAALTAAMAQPGDVPLPRLRRVLARHDSI CQRGRILPLSPCPDRAKLEESPALTCQYTVLKMLYDGPFRAW LAQQNSTILNHYIDSTIARTDKAARDMNGRKLAQAEKDLITS RAADLPRLSVDEKMGDFLARLTAATATEMRVQRGYQSDGE NAQKQAAFIGQFECDVIGRAFADFLNQSGFDFVLKLKADTP QPDAAQCDVTALIAPDDISVSPPQAWQQVLYFILHLVPVDDA SHLLHQIRKWQVLEGKEKPAQIAHDVQSVLMLYLDMHDAK FTGGAALHGIEKFAEFFAHAADFRAVFPPQSLQDQDRSIPRR GLREIVRFGHLPLLQHMSGTVQITHDNVVAWQAARTAGAT GMSPIARRQKQREELHALAVERTARFRNADLQNYMHALVD VIKHRQLSAQVTLSDQVRLHRLMMGVLGRLVDYAGLWERD LYFVVLALLYHHGATPDDVFKGQGKKNLADGQVVAALKPK NRKAAAPVGVFDDLDHYGIYQDDRQSIRNGLSHFNMLRGG KAPDLSHWVNQTRSLVAHDRKLKNAVAKSVIEMLAREGFD LDWGIQTDRGQHILSHGKIRTRQAQHFQKSRLHIVKKSAKPD KNDTVKIRENLHGDAMVERVVQLFAAQVQKRYDITVEKRL DHLFLKPQDQKGKNGIHTHNGWSKTEKKRRPSRENRKGNH EN
C2-27		SAMN04487830_13920 [Pseudobutyrivibrio sp. OR37]	MKFSKESHRKTAVGVTESNGIIGLLYKDPLNEKEKIEDVVNQ RANSTKRLFNLFGTEATSKDISRASKDLAKVVNKAIGNLKGN KKFNKKEQITKGLNTKIIVEELKNVLKDEKKLIVNKDIIDEAC SRLLKTSFRTAKTKQAVKMILTAVLIENTNLSKEDEAFVHEY FVKKLVNEYNKTSVKKQIPVALSNQNMVIQPNSVNGTLEISE TKKSKETKTTEKDAFRAFLRDYATLDENRRHKMRLCLRNLV

[0828]

[0829]

			NLYFYGETSVSKDDFDEWRDHEDKKQNDELFVKKIVSIKTD RKGNVKEVLDVDATIDAIRTNNIACYRRALAYANENPDVFF SDTMLNKFWIHHVENEVERIYGHINNNTGDYKYQLGYLSEK VWKGIINYLSIKYIAEGKAVYNYAMNALAKDNNSNAFGKLD EKFVNGITSFEYERIKAEETLQRECAVNIAFAANHLANATVD LNEKDSDFLLLKHEDNKDTLGAVARPNILRNILQFFGGKSRW NDFDFSGIDEIQLLDDLRKMIYSLRNSSFHFKTENIDNDSWNT KLIGDMFAYDFNMAGNVQKDKMYSNNVPMFYSTSDIEKML DRLYAEVHERASQVPSFNSVFVRKNFPDYLKNDLKITSAFGV DDALKWQSAVYYVCKEIYYNDFLQNPETFTMLKDYVQCLPI DIDKSMDQKLKSERNAHKNFKEAFATYCKECDSLSAICQMI MTEYNNQNKGNRKVISARTKDGDKLIYKHYKMILFEALKN VFTIYLEKNINTYGFLKKPKLINNVPAIEEFLPNYNGRQYETL VNRITEETELQKWYIVGRLLNPKQVNQLIGNFRSYVQYVND VARRAKQTGNNLSNDNIAWDVKNIIQIFDVCTKLNGVTSNIL EDYFDDGDDYARYLKNFVDYTNKNNDHSATLLGDFCAKEI DGIKIGIYHDGTNPIVNRNIIQCKLYGATGIISDLTKDGSILSV DYEIIKKYMQMQKEIKVYQQKGICKTKEEQQNLKKYQELKN IVELRNIIDYSEILDELQGQLINWGYLRERDLMYFQLGFHYLC LHNESKKPVGYNNAGDISGAVLYQIVAMYTNGLSLIDANGK SKKNAKASAGAKVGSFCSYSKEIRGVDKDTKEDDDPIYLAG VELFENINEHQQCINLRNYIEHFHYYAKHDRSMLDLYSEVFD RFFTYDMKYTKNVPNMMYNILLQHLVVPAFEFGSSEKRLDD NDEQTKPRAMFTLREKNGLSSEQFTYRLGDGNSTVKLSARG DDYLRAVASLLYYPDRAPEGLIRDAEAEDKFAKINHSNPKSD NRNNRGNFKNPKVQWYNNKTKRK
C2-28		SAMN029 10398_000 08 [Butyrivibri o sp. YAB3001]	MKISKVDHRKTAVKITDNKGAEGFIYQDPTRDSSTMEQIISN RARSSKVLFNIFGDTKKSKDLNKYTESLIIYVNKAIKSLKGDK RNNKYEEITESLKTERVLNALIQAGNEFTCSENNIEDALNKY LKKSFRVGNTKSALKKLLMAAYCGYKLSIEEKEEIQNYFVD KLVKEYNKDTVLKYTAKSLKHQNMVVQPDTDNHVFLPSRI AGATQNKMSEKEALTEFLKAYAVLDEEKRHNLRIILRKLVN LYFYESPDFIYPENNEWKEHDDRKNKTETFVSPVKVNEEKN GKTFVKIDVPATKDLIRLKNIECYRRSVAETAGNPITYFTDHN ISKFWIHHIENEVEKIFALLKSNWKDYQFSVGYISEKVWKEII NYLSIKYIAIGKAVYNYALEDIKKNDGTLNFGVIDPSFYDGIN SFEYEKIKAEETFQREVAVYVSFAVNHLSSATVKLSEAQSDM LVLNKNDIEKIAYGNTKRNILQFFGGQSKWKEFDFDRYINPV NYTDIDFLFDIKKMVYSLRNESFHFTTTDTESDWNKNLISAM FEYECRRISTVQKNKFFSNNLPLFYGENSLERVLHKLYDDYV DRMSQVPSFGNVFVRKKFPDYMKEIGIKHNLSSEDNLKLQG ALYFLYKEIYYNAFISSEKAMKIFVDLVNKLDTNARDDKGRI THEAMAHKNFKDAISHYMTHDCSLADICQKIMTEYNQQNT GHRKKQTTYSSEKNPEIFRHYKMILFMLLQKAMTEYISSEEIF DFIMKPNSPKTDIKEEEFLPQYKSCAYDNLIKLIADNVELQK

			WYITARLLSPREVNQLIGSFRSYKQFVSDIERRAKETNNSLSK SGMTVDVENITKVLDLCTKLNGRFSNELTDYFDSKDDYAVY VSKFLDFGFKIDEKFPAALLGEFCNKEENGKKIGIYHNGTEPI LNSNIIKSKLYGITDVVSRAVKPVSEKLIREYLQQEVKIKPYL ENGVCKNKEEQAALRKYQELKNRIEFRDIVEYSEIINELMGQ LINFSYLRERDLMYFQLGFHYLCLNNYGAKPEGYYSIVNDK RTIKGAILYQIVAMYTYGLPIYHYVDGTISDRRKNKKTVLDT LNSSETVGAKIKYFIYYSDELFNDSLILYNAGLELFENINEHE NIVNLRKYIDHFKYYVSQDRSLLDIYSEVFDRYFTYDRKYKK NVMNLFSNIMLKHFIITDFEFSTGEKTIGEKNTAKKECAKVRI KRGGLSSDKFTYKFKDAKPIELSAKNTEFLDGVARILYYPEN VVLTDLVRNSEVEDEKRIEKYDRNHNSSPTRKDKTYKQDVK KNYNKKTSKAFDSSKLDTKSVGNNLSDNPVLKQFLSESKKK R
C2-29		Blautia sp. Marseille-P2398	MKISKVDHVKSGIDQKLSSQRGMLYKQPQKKYEGKQLEEH VRNLSRKAKALYQVFPVSGNSKMEKELQIINSFIKNILLRLDS GKTSEEIVGYINTYSVASQISGDHIQELVDQHLKESLRKYTCV GDKRIYVPDIIVALLKSKFNSETLQYDNSELKILIDFIREDYLK EKQIKQIVHSIENNSTPLRIAEINGQKRLIPANVDNPKKSYIFE FLKEYAQSDPKGQESLLQHMRYLILLYLYGPDKITDDYCEEI EAWNFGSIVMDNEQLFSEEASMLIQDRIYVNQQIEEGRQSKD TAKVKKNKSKYRMLGDKIEHSINESVVKHYQEACKAVEEK DIPWIKYISDHVMSVYSSKNRVDLDKLSLPYLAKNTWNTWI SFIAMKYVDMGKGVYHFAMSDVDKVGKQDNLIIGQIDPKFS DGISSFDYERIKAEDDLHRSMSGYIAFAVNNFARAICSDEFRK KNRKEDVLTVGLDEIPLYDNVKRKLLQYFGGASNWDDSIIDI IDDKDLVACIKENLYVARNVNFHFAGSEKVQKKQDDILEEIV RKETRDIGKHYRKVFYSNNVAVFYCDEDIIKLMNHLYQREK PYQAQIPSYNKVISKTYLPDLIFMLLKGKNRTKISDPSIMNMF RGTFYFLLKEIYYNDFLQASNLKEMFCEGLKNNVKNKKSEK PYQNFMRRFEELENMGMDFGEICQQIMTDYEQQNKQKKKT ATAVMSEKDKKIRTLDNDTQKYKHFRTLLYIGLREAFIIYLK DEKNKEWYEFLREPVKREQPEEKEFVNKWKLNQYSDCSELI LKDSLAAAWYVVAHFINQAQLNHLIGDIKNYIQFISDIDRRA KSTGNPVSESTEIQIERYRKILRVLEFAKFFCGQITNVLTDYY QDENDFSTHVGHYVKFEKKNMEPAHALQAFSNSLYACGKE KKKAGFYYDGMNPIVNRNITLASMYGNKKLLENAMNPVTE QDIRKYYSLMAELDSVLKNGAVCKSEDEQKNLRHFQNLKN RIELVDVLTLSELVNDLVAQLIGWVYIRERDMMYLQLGLHY IKLYFTDSVAEDSYLRTLDLEEGSIADGAVLYQIASLYSFNLP MYVKPNKSSVYCKKHVNSVATKFDIFEKEYCNGDETVIENG LRLFENINLHKDMVKFRDYLAHFKYFAKLDESILELYSKAYD FFFSYNIKLKKSVSYVLTNVLLSYFINAKLSFSTYKSSGNKTV QHRTTKISVVAQTDYFTYKLRSIVKNKNGVESIENDDRRCEV VNIAARDKEFVDEVCNVINYNSDK

[0830]

C2-30		Leptotrichia sp. Marseille-P3007	MKITKIDGISHKKYIKEGKLVKSTSEENKTDERLSELLTIRLD TYIKNPDNASEEENRIRRENLKEFFSNKVLYLKDGILYLKDR REKNQLQNKNYSEEDISEYDLKNKNNFLVLKKILLNEDINSE ELEIFRNDFEKKLDKINSLKYSLEENKANYQKINENNIKKVE GKSKRNIFYNYYKDSAKRNDYINNIQEAFDKLYKKEDIENLF FLIENSKKHEKYKIRECYHKIIGRKNDKENFATIIYEEIQNVNN MKELIEKVPNVSELKKSQVFYKYYLNKEKLNDENIKYVFCH FVEIEMSKLLKNYVYKKPSNISNDKVKRIFEYQSLKKLIENKL LNKLDTYVRNCGKYSFYLQDGEIATSDFIVGNRQNEAFLRNI IGVSSTAYFSLRNILETENENDITGRMRGKTVKNNKGEEKYIS GEIDKLYDNNKQNEVKKNLKMFYSYDFNMNSKKEIEDFFSN IDEAISSIRHGIVHFNLELEGKDIFTFKNIVPSQISKKMFHDEIN EKKLKLKIFKQLNSANVFRYLEKYKILNYLNRTRFEFVNKNI PFVPSFTKLYSRIDDLKNSLGIYWKTPKTNDDNKTKEITDAQI YLLKNIYYGEFLNYFMSNNGNFFEITKEIIELNKNDKRNLKT GFYKLQKFENLQEKTPKEYLANIQSLYMINAGNQDEEEKDT YIDFIQKIFLKGFMTYLANNGRLSLIYIGSDEETNTSLAEKKQ EFDKFLKKYEQNNNIEIPYEINEFVREIKLGKILKYTERLNMF YLILKLLNHKELTNLKGSLEKYQSANKEEAFSDQLELINLLN LDNNRVTEDFELEADEIGKFLDFNGNKVKDNKELKKFDTNK IYFDGENIIKHRAFYNIKKYGMLNLLEKISDEAKYKISIEELKN YSKKKNEIEENHTTQENLHRKYARPRKDEKFTDEDYKKYEK AIRNIQQYTHLKNKVEFNELNLLQSLLLRILHRLVGYTSIWER DLRFRLKGEFPENQYIEEIFNFDNSKNVKYKNGQIVEKYINFY KELYKDDTEKISIYSDKKVKELKKEKKDLYIRNYIAHFNYIPN AEISLLEMLENLRKLLSYDRKLKNAIMKSIVDILKEYGFVVTF KIEKDKKIRIESLKSEEVVHLKKLKLKDNDKKKEPIKTYRNS KELCKLVKVMFEYKMKEKKSEN
C2-31		Bacteroides ihuae	MRITKVKVKESSDQKDKMVLIHRKVGEGTLVLDENLADLTA PIIDKYKDKSFELSLLKQTLVSEKEMNIPKCDKCTAKERCLSC KQREKRLKEVRGAIEKTIGAVIAGRDIIPRLNIFNEDEICWLIK PKLRNEFTFKDVNKQVVKLNLPKVLVEYSKKNDPTLFLAYQ QWIAAYLKNKKGHIKKSILNNRVVIDYSDESKLSKRKQALEL WGEEYETNQRIALESYHTSYNIGELVTLLPNPEEYVSDKGEIR PAFHYKLKNVLQMHQSTVFGTNEILCINPIFNENRANIQLSAY NLEVVKYFEHYFPIKKKKKNLSLNQAIYYLKVETLKERLSLQ LENALRMNLLQKGKIKKHEFDKNTCSNTLSQIKRDEFFVLNL VEMCAFAANNIRNIVDKEQVNEILSKKDLCNSLSKNTIDKEL CTKFYGADFSQIPVAIWAMRGSVQQIRNEIVHYKAEAIDKIF ALKTFEYDDMEKDYSDTPFKQYLELSIEKIDSFFIEQLSSNDV LNYYCTEDVNKLLNKCKLSLRRTSIPFAPGFKTIYELGCHLQ DSSNTYRIGHYLMLIGGRVANSTVTKASKAYPAYRFMLKLI YNHLFLNKFLDNHNKRFFMKAVAFVLKDNRENARNKFQYA FKEIRMMNNDESIASYMSYIHSLSVQEQEKKGDKNDKVRYN TEKFIEKVFVKGFDDFLSWLGVEFILSPNQEERDKTVTREEYE

[0831]

			NLMIKDRVEHSINSNQESHIAFFTFCKLLDANHLSDLRNEWI KFRSSGDKEGFSYNFAIDIIELCLLTVDRVEQRRDGYKEQTEL KEYLSFFIKGNESENTVWKGFYFQQDNYTPVLYSPIELIRKY GTLELLKLIIVDEDKITQGEFEEWQTLKKVVEDKVTRRNELH QEWEDMKNKSSFSQEKCSIYQKLCRDIDRYNWLDNKLHLV HLRKLHNLVIQILSRMARFIALWDRDFVLLDASRANDDYKL LSFFNFRDFINAKKTKTDDELLAEFGSKIEKKNAPFIKAEDVP LMVECIEAKRSFYQKVFFRNNLQVLADRNFIAHYNYISKTAK CSLFEMIIKLRTLMYYDRKLRNAVVKSIANVFDQNGMVLQL SLDDSHELKVDKVISKRIVHLKNNNIMTDQVPEEYYKICRRL LEMKK
C2-32		SAMN052 16357_104 5 [Porphyro monadacea e bacterium KH3CP3R A]	MEFRDSIFKSLLQKEIEKAPLCFAEKLISGGVFSYYPSERLKEF VGNHPFSLFRKTMPFSPGFKRVMKSGGNYQNANRDGRFYD LDIGVYLPKDGFGDEEWNARYFLMKLIYNQLFLPYFADAEN HLFRECVDFVKRVNRDYNCKNNNSEEQAFIDIRSMREDESIA DYLAFIQSNIIIEENKKKETNKEGQINFNKFLLQVFVKGFDSFL KDRTELNFLQLPELQGDGTRGDDLESLDKLGAVVAVDLKLD ATGIDADLNENISFYTFCKLLDSNHLSRLRNEIIKYQSANSDF SHNEDFDYDRIISIIELCMLSADHVSTNDNESIFPNNDKDFSGI RPYLSTDAKVETFEDLYVHSDAKTPITNATMVLNWKYGTDK LFERLMISDQDFLVTEKDYFVWKELKKDIEEKIKLREELHSL WVNTPKGKKGAKKKNGRETTGEFSEENKKEYLEVCREIDRY VNLDNKLHFVHLKRMHSLLIELLGRFVGFTYLFERDYQYYH LEIRSRRNKDAGVVDKLEYNKIKDQNKYDKDDFFACTFLYE KANKVRNFIAHFNYLTMWNSPQEEEHNSNLSGAKNSSGRQN LKCSLTELINELREVMSYDRKLKNAVTKAVIDLFDKHGMVI KFRIVNNNNNDNKNKHHLELDDIVPKKIMHLRGIKLKRQDG KPIPIQTDSVDPLYCRMWKKLLDLKPTPF
C2-33		Listeria riparia	MHDAWAENPKKPQSDAFLKEYKACCEAIDTYNWHKNKAT LVYVNELHHLLIDILGRLVGYVAIADRDFQCMANQYLKSSG HTERVDSWINTIRKNRPDYIEKLDIFMNKAGLFVSEKNGRNY IAHLNYLSPKHKYSLLYLFEKLREMLKYDRKLKNAVTKSLID LLDKHGMCVVFANLKNNKHRLVIASLKPKKIETFKWKKIK
C2-34		Insolitispiri llum peregrinum	MRIIRPYGSSTVASPSPQDAQPLRSLQRQNGTFDVAEFSRRHP ELVLAQWVAMLDKIIRKPAPGKNSTALPRPTAEQRRLRQQV GAALWAEMQRHTPVPPELKAVWDSKVHPYSKDNAPATAKT PSHRGRWYDRFGDPETSAATVAEGVRRHLLDSAQPFRANGG QPKGKGVIEHRALTIQNGTLLHHHQSEKAGPLPEDWSTYRA DELVSTIGKDARWIKVAASLYQHYGRIFGPTTPISEAQTRPEF VLHTAVKAYYRRLFKERKLPAERLERLLPRTGEALRHAVTV QHGNRSLADAVRIGKILHYGWLQNGEPDPWPDDAALYSSR YWGSDGQTDIKHSEAVSRVWRRALTAAQRTLTSWLYPAGT DAGDILLIGQKPDSIDRNRLPLLYGDSTRHWTRSPGDVWLFL

[0832]

			KQTLENLRNSSFHFKTLSAFTSHLDGTCESEPAEQQAAQALW QDDRQQDHQQVFLSLRALDATTYLPTGPLHRIVNAVQSTDA TLPLPRFRRVVTRAANTRLKGFPVEPVNRRTMEDDPLLRCR YGVLKLLYERGFRAWLETRPSIASCLDQSLKRSTKAAQTING KNSPQGVEILSRATKLLQAEGGGGHGIHDLFDRLYAATARE MRVQVGYHHDAEAARQQAEFIEDLKCEVVARAFCAYLKTL GIQGDTFRRQPEPLPTWPDLPDLPSSTIGTAQAALYSVLHLMP VEDVGSLLHQLRRWLVALQARGGEDGTAITATIPLLELYLN RHDAKFSGGGAGTGLRWDDWQVFFDCQATFDRVFPPGPAL DSHRLPLRGLREVLRFGRVNDLAALIGQDKITAAEVDRWHT AEQTIAAQQQRREALHEQLSRKKGTDAEVDEYRALVTAIAD HRHLTAHVTLSNVVRLHRLMTTVLGRLVDYGGLWERDLTF VTLYEAHRLGGLRNLLSESRVNKFLDGQTPAALSKKNNAEE NGMISKVLGDKARRQIRNDFAHFNMLQQGKKTINLTDEINN ARKLMAHDRKLKNAITRSVTTLLQQDGLDIVWTMDASHRL TDAKIDSRNAIHLHKTHNRANIREPLHGKSYCRWVAALFGA TSTPSATKKSDKIR

[0833]

[0834] *실시예 3: Cas13b 오르소로그의 특징규명*

[0835] PFS 모티프는 [Smargon et al. 2016]에 기술되고 도 6에 에시된 바와 같은 상이한 Cas13b 오르소로그에 대해 결정되었다. 스크린에서 모든 스페이서에 대해 안전하게 고갈된 (> 5σ는 비표적화 스페이서의 평균 고갈 이상) 스페이서의 비율의 히트맵을 생성시켰다. 일부 오르소로그의 경우에, 실험은 보조 단백질 csx27

또는 csx28의 존재 하에서 반복되었다.

[0836] 결과는 도 7-10에 제공된다. 하기 PFS는 상이한 오르소로그에 대해 유래되었다:

표 4

오르소로그	번호	보조 단백질	5'PFS	3'PFS
Bergeyella zoohelcum	1		A	NGA
Bergeyella zoohelcum	1	csx27	A	NGA
Prevotella intermedia	2		A	NGA
Prevotella intermedia	2	csx28	A	NGA
Prevotella buccae	3		A	NGA
Prevotella buccae	3	csx28	A	NGA
Bacteroides pyogenes	5		A	NGA
Alistipes sp. ZOR0009	6		TG	NA(G)
Prevotella sp. MA2016	7		AT	
Riemerella anatipestifer	8		A	NGA
Riemerella anatipestifer	8	csx28	A	NGA
Prevotella aurantiaca	9		G	NAA
Prevotella aurantiaca	9	csx28		
Prevotella saccharolytica	10		AG	
Prevotella intermedia	12		AG	
Capnocytophaga canimorsus	13		A	NAA
Capnocytophaga canimorsus	13	csx27	A	NAA
Porphyromonas gulae	14		A	NAA
Porphyromonas gulae	14	csx28	A	NGA
Prevotella sp. P5-125	15		AT	
Flavobacterium branchiophilum	16		TG	
Flavobacterium branchiophilum	16	csx27	TA	
Myroides odoratimimus	17		T	NAA
Porphyromonas gingivalis	19		A	NAA
Porphyromonas gingivalis	19	csx28	A	
Prevotella intermedia	21		A	NGA
Prevotella intermedia	21	csx28	A	NGA

[0837]

[0838] *실시예 4: Cas13b 오르소로그의 부차적 활성*

[0839] 진핵생물 세포에서 Cas13b 오르소로그의 부차적 활성을 결정하기 위해서, 진핵생물 세포에서 표적 특이적 가이드를 사용한 실험을 G-루시퍼라제 r C-루시퍼라제 리포터 구축물의 존재 하에서 수행되었다. 결과는 도 17A-G에 제공된다. 생체 내 부차적 활성이 상이한 Cas13b 오르소로그에 대해 효과적으로 관찰되었다.

[0840] *실시예 5: 추가 Cas13b 오르소로그의 동정*

[0841] 추가 Cas13b 오르소로그는 CRISPR 연관된 단백질과의 근접성 또는 그 자체로 CRISPR 어레이에 의존하는 생물계산 파이프라인을 사용하여 동정할 수 있다. 예시적인 방법은 (NCBI 데이타베이스로부터) 모든 원핵생물 게놈을 다운로딩시키는 단계, CRISPR 어레이를 발견하기 위해서 이들 게놈 상에서 알고리즘을 실행시키는 단계, 추정 CRISPR 시스템으로서 발견된 어레이의 10 kb 이내에 클러스터링된 단백질을 저장하는 단계 및 크기가 700개 아미노산을 초과하는 단백질의 부재를 기반으로 후보 시스템을 선택하는 단계를 포함할 수 있다. 추가의 선택 기준은 크기가 900개 내지 1800개 아미노산인 유일한 하나의 이펙터 단백질의 존재 및 기능성 및 상용성 고려 사항을 기반으로 할 수 있다.

[0842] *실시예 6:*

[0843] 효율적이고 상세한 핵산 편집은 특히 RNA의 수준에서, 유전자 질환의 표적화에 대해 상당한 유망성을 유지하고, 여기서 질환-관련 전사물은 기능성 단백질 생성물을 산출하도록 구제될 수 있다. VI형 CRISPR-Cas 시스템은 프로그램가능한 단일-이펙터 RNA-가이딩된 RNases Cas13을 함유한다. 여기서, 우리는 포유동물 세포에서 전사물에 대한 아데노신 디아미나제 활성을 유도시키는 촉매적으로 불활성화된 Cas18 (dCas13)을 사용하여 RNA 편집을 입증하고 강건한 녹다운을 할 수 있는 Cas13 오르소로그를 조작하기 위해 VI 시스템의 다양성을 프로파일링한다. ADAR2 디아미나제 도메인을 dCas13에 융합시키고 가이드 RNA를 조작하여 최적의 RNA 듀플렉스 기질을 생성시켜서, 우리는 대체로 20% 내지 40% 범위 및 최대 89%까지의 효율로 이노신 (번역 동안 구아노신으로서 판독)에 대한 특이적 단일 아데노신의 표적화 편집을 획득한다. REPAIR (RNA Editing for Programmable A to I Replacement)라고 하는 이러한 시스템은 높은 특이성을 획득하도록 더욱 조작될 수 있다. 조작된 변이체, REPAIRv2는 강건한 표적 적중 A 내지 I 편집을 유지시키면서 170배 초과로 증가된 특이성을 나타낸다. 우리는 기지의 병원성 돌연변이를 함유하는 전체 길이 전사물을 편집하고 아데노-연합 바이러스 (AAV) 벡터로의 패키징에 적합한 기능성 절두 형태를 생성시키기 위해 REPAIRv2를 사용한다. REPAIR는 연구, 치료제 및 생명공학에 대한 광범위한 적용가능성으로 유망한 RNA 편집 플랫폼을 제시한다. 정밀한 핵산 편집 기술은 신규한 치료제로서 그리고 세포 기능을 연구하기 위해 가치있다. 현행 편집 도구, 예컨대 Cas9 뉴클레아제가 게놈 유전자좌의 프로그램가능한 변형을 획득할 수 있지만, 종종 편집물은 삽입 또는 결실로 인해 이종성이거나 또는 정확한 편집을 위해 도너 주형을 요구한다. 염기 편집자, 예컨대 dCas9-APOBEC 융합체는 이중 가닥 파쇄를 발생시키지 않으면서 편집을 가능하게 하지만, 표적창 내 임의 시티딘을 편집하는, 시티딘 디아미나제 활성의 성질로 인해 정확성이 결여될 수 있다. 뿐만 아니라, 프로토스페이서 인접 모티프 (PAM)에 대한 요건은 가능한 편집 부위의 수를 제한한다. 여기서, 우리는 VI형 원핵생물 CRISPR (clustered regularly interspaced short palindromic repeats) 획득성 면역계로부터의 RNA-가이딩된 RNA 표적화 Cas13 효소를 사용한 정확하고 탄력적인 RNA 염기 편집 도구의 개발을 기술한다.

[0844] 정확한 핵산 편집 기술은 신규한 치료제로서 그리고 세포 기능을 연구하기 위해 가치있다. 프로그램가능한 뉴클레아제 예컨대 원핵생물 CRISPR-연관 뉴클레아제 Cas9 (1-4) 또는 Cpf1(5)을 기반으로 하는 현행 편집 도구는, 이후에 비상동성 말단 연결 (non-homologous end joining) (NHEJ)을 통한 표적화 유전자 파괴 또는 주형-의존적 상동성-지정 복구 (homology-directed repair) (HDR)를 통한 정확한 유전자 편집을 구동시키는, 표적화 DNA 절단을 매개하기 위해 광범위하게 채택되었다 (6). NHEJ는 분열 및 유사분열 후 세포 둘 모두에서 활성있는 숙주 기구를 이용하고 단백질 코딩 유전자에서 프레임 시프트를 야기시킬 수 있는 삽입 또는 결실 (indel) 돌연변이의 혼합을 발생시켜 효율적인 유전자 파괴를 제공한다. 대조적으로, HDR은 대체로 그 발현이 복제 세포에 제한적인 숙주 기구에 의해 매개된다. 이와 같이, 유사분열후 세포에서 유전자-편집 능력의 발생은 주요한 도전으로 남아있다. 최근에, DNA 염기 편집자, 예컨대 표적창 내에서 시토신의 티민으로의 전환을 실시하기 위해 특이적 게놈 표적에 대해 시티딘 디아미나제 활성을 표적화하기 위한 촉매적으로 불활성인 Cas9 (dCas9)의 사용은 DNA 이중 가닥 파괴를 발생시키지 않으면서 편집을 가능하게 하고 유의하게 indel의 형성을 감소시킨다 (7, 8). 그러나, DNA 염기 편집자의 표적화 범위는 편집 부위에서 프로토스페이서 인접 모티프 (PAM)에 대한 Cas9의 요구로 인해 제한적이다 (9). 여기서, 우리는 VI형 CRISPR-연관 RNA-가이딩된 RNase Cas13를 사용한 정확하고 탄력적인 RNA 염기 편집 기술의 개발을 기술한다 (10-13).

[0845] Cas13 효소는 정확한 RNA 절단을 매개하는 2개의 HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) 엔도RNase 도메인을 갖는다 (10, 11). 다음의 3개의 Cas13 단백질 패밀리가 지금까지 동정되었다: Cas13a (이전에 C2c2로 공지), Cas13b, 및 Cas13c(12, 13). 우리는 이전에 핵산 검출 (14)을 비롯하여 포유동물 식물 세포 RNA 녹다운 및 전사물 추적 (15)을 위한 도구로서 적합화될 수 있다고 보고하였다. Cas13 효소의 RNA-가이딩 성질은 그들을 RNA 결합 및 섭동 적용을 위한 매력적인 도구로 만든다.

[0846] 효소의 ADAR (adenosine deaminase acting on RNA) 패밀리는 번역 및 스플라이싱에서 구아노신과 기능적으로 동등한 핵염기인 이노신으로 아데노신의 가수분해적 탈아민화를 통해 전사물의 내생성 편집을 매개한다 (16). N-말단 이중 가닥 RNA-결합 도메인 및 C-말단 촉매적 탈아민화 도메인으로 이루어진 2개의 기능적 인간 ADAR 오르소로그, ADAR1 및 ADAR2가 존재한다. ADAR1 및 ADAR2의 내생성 표적 부위는 실질적인 이중 가닥 동일성을 함유하고, 촉매적 도메인은 생체 내 및 시험관내에서 효율적인 편집을 위해 듀플렉스 영역을 요구한다 (17, 18). ADAR 단백질이 표적화의 잠재적인 탄력성을 제한할 수 있는 편집을 위한 바람직한 모티프를 갖지만, 과활성 돌연변이체, 예컨대 ADAR(E488Q)(19)은 서열 제한성을 완화시키고, 아데노신의 이노신으

로의 편집 속도를 개선시킨다. ADAR은 RNA 듀플렉스에서 시티딘 염기 반대의 아데노신을 우선적으로 탈아민화 (20)시켜서, 정확한 염기 편집을 위한 유망한 기회를 제공한다. 이전의 접근법이 RNA 가이드를 통해 표적화된 ADAR 융합체를 조작하였지만 (21-24), 이들 접근법의 특이성은 보고되지 않았고 그들의 개별 표적화 기전은 표적 인식 및 엄격도를 증강시킬 수 있는 단백질 파트너의 보조없이 RNA-RNA 하이브리드화에 의존한다.

[0847] 여기서 우리는 포유동물 세포에서 RNA 녹다운 활성에 대해 Cas13 표소의 전체 패밀리를 어세이하고 포유동물 세포 적용을 위해 가장 효율적이고 특이적인 것으로서 프레보텔라 sp. P5-125 (PspCas13b) 유래의 Cas13b 오르소로그를 동정한다. 그 다음으로 우리는 ADAR2 디아미나제 도메인 (ADARDD)을 촉매적으로 불활성인 PspCas13b와 융합시키고 질환-관련 돌연변이를 비롯하여 내생성 전사물 및 리포터의 REPAIR (RNA editing for programmable A to I (G) replacement)를 입증한다. 마지막으로, 우리는 특이성이 170배 초과로 증가된 REPAIRv2를 생성시키기 위해 dCas13b-ADAR2DD 융합체의 특이성을 개선시키도록 합리적인 돌연변이유발 계획을 적용한다.

[0848] 방법

[0849] 박테리아 구축물의 디자인 및 클로닝

[0850] 포유동물 코돈 최적화된 Cas13b 구축물을 Lac 프로모터의 제어 하의 클로람페니콜 내성 pACYC184 벡터에 클로닝시켰다. BsaI 제한효소 부위에 의해 분리된 2개의 상응하는 직접-반복부 (DR) 서열은 pJ23119 프로모터의 제어 하의 Cas13b의 하류에 삽입되었다. 마지막으로, 표적화 스페이서를 위한 올리고는 T4 PNK (New England Biolabs)를 사용해 인산화시켰고, 어닐링시키고, T7 리가제 (Enzymatics)를 사용해 BsaI 분해된 벡터에 결찰시켜 표적화 Cas13b 벡터를 생성시켰다.

[0851] 박테리아 PFS 스크린

[0852] PFS 스크린을 위한 앰피실린 내성 플라스미드는 NEB Gibson Assembly (New England Biolabs)를 사용하여 앰피실린 내성 유전자 bla의 출발 코돈의 바로 하류에 표적 부위에 의해 분리시킨 2개 5' 무작위 뉴클레오티드 및 4개 3' 무작위 뉴클레오티드를 갖는 Cas13b 표적 함유 PCR 생성물을 삽입시켜 클로닝되었다. 100 ng의 앰피실린-내성 표적 플라스미드를 Endura 전기적수용능력 세포로 65-100 ng 클로람페니콜-내성 Cas13b 박테리아 표적화 플라스미드와 함께 전기영동시켰다. 플라스미드를 세포에 첨가하였고, 얼음 상에서 15분간 인큐베이션시켰고, 제조사의 프로토콜을 사용해 전기천공하고 나서, 950 ㎕의 회수 배지를 세포에 첨가하여 1시간 동안 37℃에서 증식시켰다. 증식물을 클로람페니콜 및 앰피실린 이중 선별 플레이트 상에 플레이팅하였다. 증식물의 연속 희석물을 사용하여 cfu/ng DNA를 평가하였다. 플레이팅 후 16시간에, 세포를 플레이트로부터 긁어 모으고 생존 플라스미드 DNA를 Qiagen 플라스미드 플러스 맥시 키트 (Qiagen)를 사용해 회수하였다. 생존 Cas13b 표적 서열 및 그들의 측접 영역을 PCR을 통해 증폭시켰고 Illumina NextSeq를 사용해 서열분석하였다. PFS 선호도를 평가하기 위해서, 본래 라이브러리 중 무작위 뉴클레오티드를 함유하는 위치를 추출하였고, 양쪽 생물복제물이 존제하는 복제 단독 조건에 비해 고갈된 서열은 고객 파이썬 스크립트를 사용하여 추출하였다. 벡터 단독 조건과 비교하여 Cas13b 조건에서 PFS 존재비의 비율의 -log2을 사용하여 바람직한 모티프를 계산하였다. 특히, 특별한 한계치 이상의-log2(샘플/벡터) 고갈 비율을 갖는 모든 서열을 사용하여 서열 모티프의 웹로고를 생성시켰다 (weblogo.berkeley.edu). 각각의 Cas13b 오르소로그에 대한 웹로고를 생성시키는데 사용된 특별한 고갈 비율 값은 표 9에 열거되어 있다.

[0853] RNA 간섭을 위한 포유동물 구축물의 디자인 및 클로닝

[0854] 포유동물 세포에서 Cas13 오르소로그를 시험하기 위한 벡터를 생성시키기 위해서, 포유동물 코돈 최적화된 Cas13a, Cas13b, 및 Cas13c 유전자를 PCR 증폭시켰고 EF1알파 프로모터의 제어 하에 이중 NLS 서열 및 C-말단 msfGFP를 함유하는 포유동물 발현 벡터에 골든-게이트 클로닝되었다. 추가 최적화를 위해서 Cas13 오르소로그는 EF1알파 프로모터의 제어 하에 상이한 C-말단 국재화 태그를 함유하는 목적 벡터에 골든 게이트 클로닝시켰다.

[0855] 이중 루시퍼라제 리포터는 가우시아 (Gaussia) 및 시프리디니아 (Cypridinia) 루시퍼라제 코딩 DNA, EF1알파 및 CMV 프로모터를 PCR 증폭시키고 NEB Gibson Assembly (New England Biolabs)를 사용한 조립에 의해 클로닝되었다.

[0856] Cas13a, Cas13b, 또는 Cas13c 오르소로그에 대한 포유동물 가이드 RNA의 발현을 위해서, 상응하는 직접 반복부 서열은 골든-게이트 억셉터 부위를 갖게 합성되었고 제한효소 분해 클로닝을 통해 U6 발현 하에 클로닝되었다.

이후에 개별 가이드는 골든 게이트 클로닝을 통해 각각의 오르소로그에 대한 상응하는 발현 골격에 클로닝되었다.

[0857] Cas13 간섭 특이성에 대해 풀링된 미스매치 라이브러리의 클로닝

[0858] 풀링된 미스매치 라이브러리 표적 부위는 PCR을 통해 생성시켰다. 표적 내 각 위치에서 94%의 올바른 염기 및 2%의 각각의 올바르지 않은 염기의 혼합을 함유하는 G-루시퍼라제 내 반-축퇴성 표적 서열을 함유하는 올리고가 제1 프라이머로서 사용되었고, G-루시퍼라제의 비표적화 영역에 상응하는 올리고가 PCR 반응의 제2 프라이머로서 사용되었다. 그 다음으로 미스매치 라이브러리 표적을 NEB Gibson assembly (New England Biolabs)를 사용하여 야생형 G-루시퍼라제 대신에 이중 루시퍼라제 리포터에 클로닝시켰다.

[0859] RNA 편집을 위한 포유동물 구축물의 디자인 및 클로닝

[0860] PspCas13b는 HEPN 도메인의 촉매 부위에서 2개 히스티딘의 알라닌으로의 돌연변이 (H133A/H1058A)를 통해서 촉매적으로 불활성 (dPspCas13b)으로 만들었다. 인간 ADAR1 및 ADAR2의 디아미나제 도메인은 pcDNA-CMV 벡터 골격으로 gibson 클로닝을 위해 합성하고 PCR 증폭시켰으며 dPspCas13b와 C-말단에서 GS 또는 GSGGGGS 링커를 통해 융합시켰다. 우리는 상이한 링커를 시험하는 실험을 위해서, dPspCas13b 및 ADAR2dd 사이에 하기의 추가적인 링커를 클로닝시켰다: GGGGSGGGGSGGGGS, EAAAK, GGSGGSGGSGGSGGSGGS, 및 SGSETPGTSESATPES (XTEN). 특이성 돌연변이체는 적절한 돌연변이체를 dPspCas13b-GSGGGGS 골격에 gibson 클로닝을 통해 생성시켰다.

[0861] RNA 편집 활성을 측정하기 위한 루시퍼라제 리포터 벡터는 녹다운 실험에 사용된 루시퍼라제 리포터 벡터에 W85X 돌연변이 (TGG>TAG)를 생성시켜서 만들었다. 이 리포터 벡터는 정규화 대조군으로서 기능적 Gluc를 발현하지만, 사전종결 부위의 첨가로 인해 결함성 Cluc를 발현한다. ADAR 편집 모티프 선호도를 시험하기 위해서, 우리는 Cluc의 코돈 85 (XAX)의 아데노신 주변의 모든 가능한 모티프를 클로닝하였다.

[0862] REPAIR의 PFS 선호도를 시험하기 위해서, 우리는 아데노신 편집 부위 및 표적 영역의 상류에 6개 염기쌍 축퇴성 PFS 서열을 함유하는 풀링된 플라스미드 라이브러리를 클로닝하였다. 라이브러리는 Integrated DNA Technologies (IDT)의 울트라머로서 합성되었고 프라이머를 어닐링시키고 DNA 중합효소 I의 클레노우 단편 (New England Biolabs)으로 서열을 채워서 이중 가닥으로 만들었다. 이러한 축퇴성 서열을 함유하는 dsDNA 단편을 분해된 리포터 벡터로 gibson 클로닝시켰고 그 다음으로 이소프로판올 침전시키고 정제하였다. 클로닝된 라이브러리로 Endura 수용성 이. 콜라이 세포 (Lucigen)를 전기천공하였고 245 mm x 245 mm 사각형 바이오어세이 플레이트 (Nunc)에 플레이팅하였다. 16시간 이후에, 콜로니를 회수하고 내독소-무함유 MACHEREY-NAGEL 미디프렙 키트를 사용해 미디프렙하였다. 클로닝된 라이브러리는 차세대 서열분석을 통해 검증하였다.

[0863] REPAIR 활성을 시험하기 위해 질환-관련 돌연변이를 클로닝하기 위해서, ClinVar에서 정의된 바와 같이 질환 발병기전과 관련된 34개의 G>A 돌연변이를 선별하였고 이들 돌연변이 주변의 200 bp 영역은 CMV 프로모터 하에서 mScarlett 및 EGFP 사이에 golden gate 클로닝시켰다. AVPR2 및 FANCC에 2개의 추가적인 G>A 돌연변이를 EF1알파의 발현 하에서 전체 유전자 서열의 Gibson 클로닝을 위해 선별하였다.

[0864] REPAIR를 위한 포유동물 가이드 RNA의 발현을 위해서, PspCas13b 직접 반복부 서열은 golden-gate 억셉터 부위를 갖게 합성하여 제한효소 분해 클로닝을 통해 U6 발현 하에 클로닝시켰다. 그 다음으로 개별 가이드를 golden gate 클로닝을 통해 이 발현 골격에 클로닝시켰다.

[0865] 포유동물 세포 배양

[0866] 포유동물 세포 배양 실험은 고농도 포도당, 소듐 피루베이트, 및 GlutaMAX (Thermo Fisher Scientific)와 추가적으로 1× 페니실린-스트렙토마이신 (Thermo Fisher Scientific) 및 10% 태아 소 혈청 (VWR Seradigm)이 보충된 둘베코의 변형 이글 배지 (Dulbecco's Modified Eagle Medium)에서 성장시킨 HEK293FT 세포주 (American Type Culture Collection (ATCC))에서 수행하였다. 세포는 80% 이하의 합류성으로 유지시켰다.

[0867] 달리 표시하지 않으면, 모든 트랜스펙션은 폴리-D-리신 (BD Biocoat)으로 코팅된 96웰 플레이트 중에서 Lipofectamine 2000 (Thermo Fisher Scientific)을 사용해 수행되었다. 세포는 트랜스펙션의 시점에 90%의 함류성을 보장하기 위해 트랜스펙션 전 16시간에 대략 20,000 세포/웰로 플레이팅되었다. 플레이트의 각 웰에 대해, 트랜스펙션 플라스미드는 총 25 ㎕까지 Opti-MEM I 감량 혈청 배지 (Thermo Fisher)와 배합되었다. 개별적으로, 24.5 ㎕의 Opti-MEM이 0.5 ㎕의 Lipofectamine 2000과 배합되었다. 플라

스미드 및 Lipofectamine 용액을 배합하고 5분 동안 인큐베이션시킨 후에, 그들을 세포 상에 파이펫팅해 넣었다.

[0868] RNA 녹다운 포유동물 세포 어세이

[0869] 리포터 구축물을 갖는 포유동물 세포에서 RNA 표적화를 평가하기 위해서, 150 ng의 Cas13 구축물을 300 ng의 가이드 발현 플라스미드 및 12.5 ng의 녹다운 리포터 구축물과 공동-트랜스펙션시켰다. 트랜스펙션 후 48시간에, 분비된 루시퍼라제를 함유하는 배지를 세포로부터 제거하였고, PBS에 1:5로 희석하여 주입 프로토콜에 따라 플레이트 판독기 (Biotek Synergy Neo2) 상에서 BioLux Cypridinia 및 Biolux Gaussia 루시퍼라제 어세이 키트 (New England Biolabs)를 사용해 활성에 대해 측정하였다. 수행된 모든 복제물은 생물학적 복제물이다.

[0870] 내생성 유전자의 표적화를 위해서, 150 ng의 Cas13 구축물을 300 ng의 가이드 발현 플라스미드와 공동-트랜스펙션시켰다. 트랜스펙션 후 48시간에, 세포를 용해시켰고 RNA를 회수하여 Cells-to-Ct 키트 (Thermo Fisher Scientific)의 이전에 기술된 [CITE PROTOCOLS] 변형법을 사용해 역전사시켰다. cDNA 발현은 KRAS 전사물에 대한 TaqMan qPCR 프로브 (Thermo Fisher Scientific), GAPDH 대조군 프로브 (Thermo Fisher Scientific), 및 Fast Advanced Master Mix (Thermo Fisher Scientific)를 사용한 qPCR을 통해 측정하였다. qPCR 반응은 384웰 구성으로 4개의 5 ㎕ 기술적 복제물로, LightCycler 480 Instrument II (Roche)에서 판독하였다.

[0871] 미스매치된 표적의 풀링된 라이브러리를 사용한 RNA 특이성의 평가

[0872] 미스매치된 표적 라이브러리를 간섭하는 Cas13의 능력은 6웰 플레이트에 파종된 HEK293FT 세포를 사용해 시험되었다. ~70% 합류성 세포를 2400 ng의 Cas13 벡터, 4800 ng의 가이드 및 240 ng의 미스매치된 표적 라이브러리를 사용해 트랜스펙션시켰다. 트랜스펙션 후 48시간에, 세포를 회수하였고 RNA는 QIAshredder (Qiagen) 및 Qiagen RNeasy 미니 키트를 사용해 추출하였다. 1 ㎍의 추출된 RNA를 제조사의 유전자-특이적 프라이밍 프로토콜에 따라 qScript Flex cDNA 합성 키트 (Quantabio) 및 Gluc 특이적 RT 프라이머를 사용해 역전사시켰다. 다음으로 cDNA를 증폭시키고 Illumina NextSeq에서 서열분석하였다.

[0873] 서열분석은 서열 당 판독치를 계측하여 분석되었고 고갈 점수는 log2(-판독 계측 비율) 값을 결정하여 계산하였으며, 여기서 판독 계측 비율은 비표적화 가이드 조건에 대한 표적화 가이드 조건에서 판독 계측치의 비율이다. 이러한 점수 값은 서열에 대한 Cas13 활성의 수준을 의미하며, 값이 높을수록 더 강력한 고갈 및 그에 따라 더 높은 Cas13 절단 활성을 의미한다. 단일 미스매치 및 이중 미스매치 서열에 대한 개별 분포를 결정하였고 각각의 미스매치 동일성에 대한 고갈 점수를 사용해 히트맵으로서 도식화하였다. 이중 미스매치 서열의 경우 소정 위치에서 모든 가능한 이중 미스매치의 평균을 도식화하였다.

[0874] RNA 서열분석에 의한 포유동물에서 Cas13의 전사체-전반 프로파일링

[0875] 전사체-전반 특이성의 측정을 위해, 150 ng의 Cas13 구축물, 300 ng의 가이드 발현 플라스미드 및 15 ng의 녹다운 리포터 구축물을 공동-트랜스펙션시켰고; shRNA 조건의 경우, 300 ng의 shRNA 표적화 플라스미드, 15 ng의 녹다운 리포터 구축물, 및 150 ng의 EF1-알파 구동 mCherry (리포터 로딩량 맞춤용)를 공동-트랜스펙션시켰다. 트랜스펙션 후 48시간에, RNA는 RNeasy 플러스 미니 키트 (Qiagen)를 사용해 정제하였고, mRNA는 NEBNext Poly(A) mRNA 자성 단리 모듈 (New England Biolabs)를 사용해 선택하였고 Illumina 용 NEBNext Ultra RNA 라이브러리 프렙 키트 (New England Biolabs)를 사용해 서열분석을 위해 준비하였다. RNA 서열분석 라이브러리는 NextSeq (Illumina)에서 서열분석하였다.

[0876] 전사체-전반 서열분석 데이타를 분석하기 위해서, 디폴트 매개변수에 의해서 Bowtie 및 RSEM 버전 1.2.31를 사용하는 RefSeq GRCh38로 판독치를 정렬시켰다 [CITE RSEM: 기준 게놈과 함께 또는 없이 RNA-Seq 데이타로부터 정확한 전사물 정량]. 전사물 발현은 log2(TPM + 1)로서 정량하였고, 유전자는 log2(TPM + 1) >2.5에 대해 필터링하였다. 차등적으로 발현된 유전자의 선별을 위해서, >2 또는 <.75의 차등적 변화가 있는 유전자만을 고려하였다. 차등적 발현의 통계적 유의성은 비표적화 복제물에 대한 3개의 표적화 복제물에 대한 스튜던트 t-검정법으로 평가하였고, 벤자미니-호치버그 (Benjamini-Hochberg) 보정법을 통해서 <0.01%의 거짓 발견율을 필터링하였다.

[0877] 포유동물 세포 트랜스펙션에서 ADAR RNA 편집

[0878] 포유동물 세포에서 REPAIR 활성을 평가하기 위해서, 우리는 150 ng의 REPAIR 벡터, 300 ng의 가이드 발현 플라

스미드, 및 40 ng의 RNA 편집 리포터를 트랜스펙션시켰다. 48시간 후에, 세포로부터 RNA를 회수하고 유전자 특이적 역전사 프라이머를 사용하여 이전에 기술된 방법을 사용해 역전사시켰다 [cite JJ]. 추출된 cDNA는 Illumina 어댑터 및 샘플 바코드를 첨가하기 위해 NEBNext High-Fidelity 2X PCR 마스터 믹스를 사용해 2회의 PCR을 수행하였다. 그 다음으로 라이브러리는 Illumina NextSeq 또는 MiSeq에서 차세대 서열분석을 수행하였다. RNA 편집율은 서열분석창 내의 모든 아데노신에서 평가되었다.

[0879] 루시퍼라제 리포터가 RNA 편집에 의해 표적화되는 실험에서, 또한 RNA 회수 전에 루시퍼라제가 분비된 배지를 회수하였다. 이 경우에, 올바른 Cluc는 낮은 수준일 것이기 때문에, 배지를 희석하지 않았다. 주입 프로토몰을 사용해 플레이트 판독기 (Biotek Synergy Neo2)에서 BioLux Cypridinia 및 Biolux Gaussia 루시퍼라제 어세이 키트 (New England Biolabs)로 루시퍼라제 활성을 측정하였다. 수행된 모든 복제는 생물학적 복제물이다.

[0880] PFS 결합 포유동물 스크린

[0881] 편집 효율에 대한 PFS의 기여도를 결정하기 위해서, 625 ng의 PFS 표적 라이브러리, 4.7 ㎍의 가이드, 및 2.35 ㎍의 REPAIR로 225 ㎠ 플레이트에 플레이팅된 HEK293FT 세포를 공동-트랜스펙션시켰다. 플라스미드를 33 ㎕의 PLUS 시약 (Thermo Fisher Scientific)과 혼합하였고, Opti-MEM와 533 ㎕가 되게하고 5분간 인큐베이션시켰으며, 30 ㎕의 Lipofectamine 2000 및 500 ㎕의 Opti-MEM과 배합하고, 추가 5분 동안 인큐베이션시킨 후에, 세포 상에 파이펫팅해 넣었다. 트랜스펙션 후 48시간에, RNA는 RNeasy 플러스 미니 키트 (Qiagen)를 사용해 회수하고, 유전자 특이적 프라이머를 사용해 qScript Flex (Quantabio)로 역전사시키고, Illumina 어댑터 및 샘플 바코드를 첨가하기 위해 NEBNext High-Fidelity 2X PCR 마스터 믹스 (New England Biolabs)를 사용해 2회의 PCR을 수행하였다. 라이브러리는 Illumina NextSeq에서 서열분석하였고, 표적 아데노신에서 RNA 편집 속도는 PFS 동일성에 대해 맵핑되었다. 적용범위를 증가시키기 위해서, PFS는 4개 뉴클레오티드까지 계산적으로 붕괴시켰다. REPAIR 편집율을 각 PFS에 대해 계산하였고, 감산된 상응하는 PFS에 대한 비표적화율로 생물학적 복제물에 대해 평균내었다.

[0882] ADAR 편집 특이성을 평가하기 위한 전체-전사체 서열분석

[0883] 전사체 전반에서 표적-외 RNA 편집 부위를 분석하기 위해서, RNeasy 플러스 미니프렙 키트 (Qiagen)를 사용해 트랜스펙션 후 48시간에 세포로부터 전체 RNA를 회수하였다. mRNA 분획을 NEBNext Poly(A) mRNA 자성 단리 모듈 (NEB)을 사용해 농축시키고 이 RNA는 Illumina용 NEBNext Ultra RNA 라이브러리 프렙 키트 (NEB)를 사용해 서열분석을 위해 준비하였다. 다음으로 라이브러리는 Illumina NextSeq에서 서열분석하였고 로딩하여 샘플 당 적어도 5백만 판독치가 존재하였다.

[0884] 표적화 및 전사체 전반 실험을 위한 RNA 편집 분석

[0885] 전사체-전반 RNA 편집 RNA 서열분석 데이타를 분석하기 위해서, 서열 파일은 무작위로 5백만 판독치까지 다운샘플링하였다. 첨가된 Gluc 및 Cluc 서열과 함께 RefSeq GRCh38 어셈블리를 사용해 인덱스를 생성시켰고 판독치는 Bowtie/RSEM 버전 1.3.0을 사용해 정렬시키고 정량하였다. 다음으로 정렬 BAM은 다음의 매개변수와 REDitools [cite]를 사용해 RNA 편집 부위에 대해 분류하고 분석하였다: -t 8 -e -d -l -U [AG 또는 TC] -p -u -m20 -T6-0 -W -v 1 -n 0.0. 비트랜스펙션 또는 EGFP-트랜스펙션 상태에서 발견된 임의의 유의한 편집은 SNP 또는 트랜스펙션의 인공물로 간주되었고 표적-외 분석으로부터 필터링해 버렸다. 표적-외는 피셔 정확 검정이 0.5 미만의 p-값을 산출하고 3개 생물학적 복제물 중 적어도 2개가 편집 부위를 확인하면 유의한 것으로 간주하였다.

[0886] 각각의 표적-외의 예측된 변이 효과를 분석하기 위해서, 표적-외 편집 부위의 목록은 SIFT 및 PolyPhen-2 주석을 사용하는 도구의 UCSC 게놈 브라우저 스위트의 일부로서 변이체 주석 완성자 (https://genome.ucsc.edu/cgi-bin/hgVai)를 사용해 분석하였다. 표적-외 유전자가 종양발생성인지 여부를 명확히 하기 위해서, 암에서 체세포 돌연변이의 COSMIC 카탈로그로부터의 종양발생 주석의 데이타베이스 (cancer.sanger.ac.uk)를 사용하였다.

[0887] REPAIR 구축물이 RNA 수준을 교란시키는지 여부를 분석하기 위해서, SEM 분석에 의한 TPM (transcript per million) 값 출력치를 발현 계측에 사용하였고 log2(TPM+1)를 취하여 로그-공간으로 전환시켰다. 차등적으로 조절된 유전자를 찾기 위해서, 스튜던트 t-검정을 3개의 비표적화 가이드 복제물 대비 3개의 표적화 가이드 복제물에 대해 수행하였다. 통계적 분석은 오직 2.5 초과의 log2(TPM+1) 값을 갖는 유전자에 대해서만 수행하였고 유전자는 그들이 2 초과 또는 0.8 미만의 배수 변화를 가지면 단지 차등적으로 조절되는 것으

로 간주되었다. 유전자는 그들이 0.01 미만의 거짓 발견율을 가지면 기록하였다.

[0888] 결과

[0889] 포유동물 세포에서 Cas13 패밀리 구성원의 종합적 특징규명

[0890] 우리는 이전에 포유동물 녹다운 적용을 위해 LwaCas13a를 개발하였지만, 효율적인 녹다운을 위해 msfGFP 안정화 도메인이 요구되고, 특이성은 높았지만, 녹다운 효율은 일관적으로 50% 이하가 아니었다 (15). 우리는 포유동물 세포에서 그들의 RNA 녹다운 활성을 평가하기 위해서 Cas13 패밀리 구성원의 유전적으로 다양한 세트를 특징으로 하는 보다 강건한 RNA-표적화 CRISPR 시스템을 동정하고자 하였다 (도 31A). 우리는 21 Cas13a, 15 Cas13b, 및 7 Cas13c 포유동물 코돈-최적화된 오르소로그 (표 6)를 단백질 안정성을 증강시키기 위해서 N-말단 및 C-말단 핵 이출 신호 (NES) 서열 및 C-말단 msfGFP를 갖는 발현 벡터에 클로닝시켰다. 포유동물 세포에서 간섭을 어세이하기 위해서, 우리는 하나의 루시퍼라제가 Cas13 간섭 활성의 척도로서 기능할 수 있게 하고 나머지는 내부 대조군으로서 제공되는, 개별 프로모터 하의 오르소고날 가우시아 (*Gaussia*) (Gluc) 및 시프리디니아 (*Cypridinia*) (Cluc) 루시퍼라제를 발현하는 이중 리포터 구축물을 디자인하였다.
각각의 오르소로그에 대해서 우리는 앰피실린 간섭 어세이로부터의 Cas13b PFS 모티프 (도 55; 표 7; 부기 1) 및 Cas13a 활성의 이전 보고로부터의 3' H PFS를 사용하는, PFS-상용성 가이드 RNA를 디자인하였다 (10).

[0891] 우리는 HEK293FT 세포를 Cas13 발현, 가이드 RNA 및 리포터 플라스미드로 트랜스펙션시켰고 48시간 이후에 표적화된 Gluc 48의 수준을 정량하였다. 각각의 Cas13 오르소로그에 대한 2개 가이드 RNA의 시험은 최근에 특징규명된 LwaCas13a에 비해 양쪽 가이드 RNA에 걸쳐 유사하거나 또는 증가된 간섭을 갖는 5종의 Cas13b 오르소로그를 포함한, 활성 수준의 범위를 밝혀주었다 (도 49B). 우리는 이들 5종의 Cas13b 오르소로그를 비롯하여, 상위 2종 Cas13a 오르소로그를 추가 조작을 위해 선택하였다.

[0892] 우리는 다음으로 강건한 활성을 위해 안정화 도메인을 요구하지 않는 오르소로그를 선택하기 위해서, msfGFP 없이 Gluc의 Cas13-매개 녹다운에 대해 시험하였다. 우리는 msfGFP 이외에도, Cas13 활성이 LwaCas13a의 최적화에 대해 이전에 보고된 바와 같이, 세포내 국재화에 의해 영향받을 수 있다고 가정하였다 (15). 그러므로, 우리는 msfGFP 없이 6종의 상이한 국재화 태그 중 하나와 C-말단에서 융합된 7종의 선택된 Cas13 오르소로그의 간섭 활성을 시험하였다. 루시퍼라제 리포터 어세이를 사용하여, 우리는 HIV Rev 유전자 NES와 C-말단에서 융합된 PspCas13b 및 PguCas13b 및 MAPK NES와 C-말단에서 융합된 RanCas13b가 최고 수준의 간섭 활성을 갖는다는 것을 확인하였다 (도 56A). 상위 오르소로그의 활성 수준을 더욱 구별하기 위해서, 우리는 위치-매칭된 가이드를 사용하여 *KRAS* 전사물을 녹다운시키는 그들 능력에 대해서 3종의 최적화된 Cas13b 구축물을 최적 LwaCas13a-msfGFP 융합 및 shRNA와 비교하였다 (도 56B). 우리는 PspCas13b에 대한 최고 수준 간섭을 관찰하였고 (평균 녹다운 62.9%), 따라서 LwaCas13a와의 추가 비교를 위해 이것을 선택하였다.

[0893] PspCas13b 및 LwaCas13a의 활성 수준을 보다 엄격하게 정의하기 위해서 우리는 Gluc 및 Cluc 둘 모두를 따라서 타일링된 위치 매칭된 가이드를 디자인하였고 우리의 루시퍼라제 리포터 어세이를 사용해 그들 활성을 어세이하였다. 우리는 각각 Gluc 및 Cluc를 표적화하는 93 및 20개의 위치 매칭된 가이드를 시험하였고, LwaCas13a에 비해서 일관적으로 증가된 녹다운 수준을 갖는다는 것을 확인하였다 (PspCas13b의 경우 평균 92.3% 대 LwaCas13a의 경우 40.1% 녹다운) (도 49C,D).

[0894] Cas13 포유동물 간섭 활성의 특이성

[0895] PspCas13b 및 LwaCas13a의 간섭 특이성을 특징규명하기 위해서 우리는 표적 서열 전반에서 단일 미스매치 및 이중 미스매치 및 3개의 측접된 5' 및 3' 염기쌍을 함유하는 루시퍼라제 표적의 플라스미드 라이브러리를 디자인하였다 (도 56C). 우리는 HEK293FT 세포를 LwaCas13a 또는 PspCas13b, 미변형된 표적 서열을 표적화하는 고정된 가이드 RNA, 및 적절한 시스템에 상응하는 미스매치된 표적 라이브러리로 트랜스펙션시켰다.
다음으로 우리는 미스매치된 표적 서열의 고갈을 정량하기 위해 미절단된 전사물의 표적화된 RNA 서열분석을 수행하였다. 우리는 LwaCas13a 및 PspCas13b가 PspCas13b 표적의 경우 염기쌍 12-26 및 LwaCas13a 표적의 경우 염기쌍 13-24로부터 연장된, 단일 미스매치에 비교적 비내성인 중심 영역을 갖는다는 것을 발견하였다 (도 56D). 이중 미스매치는 단일 돌연변이보다도 덜 내성이었고, 그들 개별 표적에서 PspCas13b의 경우 12-29 염기쌍 및 LwaCas13a에 대해 8-27 염기쌍으로부터 연장된 더 큰 창에서 녹다운 활성이 거의 관찰되지 않았다 (도 56E). 추가로, 표적 서열의 5' 및 3' 말단에 측접하는 3개 뉴클레오티드에 포함된 미스매치가 존재하기 때문에, 우리는 Cas13 녹다운 활성에 대한 PFS 제한성을 평가할 수 있었다. 서열분석

은 거의 모든 PFS 조합이 강건한 녹다운을 허용하였다는 것을 보여주어서, 포유동물 세포에서 간섭에 대한 PFS 제한성이 아마도 시험된 효소에 대해 존재하지 않는다는 것을 시사한다. 이들 결과는 Cas13a 및 Cas13b가 미스매치에 대한 유사한 서열 제한성 및 감도를 나타낸다는 것을 의미한다.

[0896] 다음으로 우리는 전사체의 mRNA 분획에 걸쳐 PspCas13b 및 LwaCas13a의 간섭 특이성을 특징규명하였다. 우리는 유의한 차등적으로 발현된 유전자를 검출하기 위해 전사체-전반 mRNA 서열분석을 수행하였다. LwaCas13a 및 PspCas13b는 Gluc의 강건한 녹다운이 입증되었고 (도 49E,F) 수백개의 표적-외 효과를 보인 위치-매칭된 shRNA와 비교하여 고도로 특이적이었다 (도 49G).

[0897] Cas13-ADAR 융합체는 표적화된 RNA 편집을 가능하게 한다

[0898] PspCas13b가 포유동물에서 mRNA의 일관적이고, 강건하며 특이적인 녹다운을 획득했다는 것을 고려하여, 우리는 프로그램가능한 RNA 편집을 위해 ADAR의 디아미나제 도메인 ($ADAR_{DD}$)을 동원하도록 RNA 결합 플랫폼으로서 적합화시킬 수 있다는 것을 구상하였다. 뉴클레아제 활성이 결여된 PspCas13b (dPspCas13b, 이하 dCas13b라고 함)를 조작하기 위해서, 우리는 HEPN 도메인 내 보존된 촉매성 잔기를 돌연변이시켰고 루시퍼라제 RNA 녹다운 활성의 상실을 관찰하였다 (도 57A). 우리는 $dCas13b\text{-}ADAR_{DD}$ 융합체가 아데노신을 표적화하기 위해 가이드 RNA에 의해 동원될 수 있는데, 하이브리드화된 RNA가 ADAR 활성을 위해 필요한 듀플렉스 기질을 생성시킬 수 있다고 가정하였다 (도 50A). 표적 아데노신 탈아민화를 증강시키기 위해서, 우리는 우리의 초기 RNA 편집 디자인에 2개의 추가 변형을 도입시켰고, 우리는 탈아민화 빈도를 증가시킨다고 이전에 보고된, 표적 아데노신 반대에 미스매치된 시티딘을 도입시켰고, 촉매 활성을 증강시키기 위해서 초활성화 돌연변이를 함유하는 인간 ADAR1 또는 ADAR2의 디아미나제 도메인과 dCas13b를 융합시켰다 ($ADAR1_{DD}$(E1008Q)(*25*) 또는 $ADAR2_{DD}$(E488Q)(*19*)).

[0899] $dCas13b\text{-}ADAR_{DD}$ 의 활성을 시험하기 위해서, 우리는 A->I 편집을 통해서 기능적으로 야생형 코돈으로 복구될 수 있고 그 다음으로 Cluc 발광성의 복원으로서 검출될 수 있는, 넌센스 돌연변이 (W85X (UGG->UAG))를 도입시켜 Cluc에 대한 RNA-편집 리포터를 생성시켰다 (도 50B). 우리는 최적 가이드 위치 및 디자인을 결정하기 위해서 표적 아데노신에 걸쳐 30, 50, 70 또는 84개 뉴클레오티드 길이의 스페이서를 사용해 가이드를 고르게 타일링시켰다 (도 50C). 우리는 $dCas13b\text{-}ADAR1_{DD}$가 Cluc 리포터를 복구하기 위해 더 긴 가이드를 필요로 하는 반면, $dCas13b\text{-}ADAR2_{DD}$는 시험된 모든 가이드 길이에서 기능적이었다는 것을 확인하였다 (도 50C).

우리는 또한 야생형 $ADAR2_{DD}$ 에 의한 루시퍼라제 복원이 감소되었기 때문에, 초활성 E488Q 돌연변이가 편집 효율을 개선시켰다는 것을 확인하였다 (도 57B). 이러한 활성의 입증으로부터, 우리는 추가 특징규명을 위해 $dCas13b\text{-}ADAR2_{DD}$(E488Q)를 선택하였고 이 접근법을 REPAIRv1 (RNA Editing for Programmable A to I Replacement version 1)로 명명하였다.

[0900] 루시퍼라제 활성의 복원이 진짜 편집 사건에 기인한 것인지를 검증하기 위해서, 우리는 역전사 및 표적화 차세대 서열분석을 통해 적접적으로 REPAIRv1가 수행된 Cluc 전사물의 편집을 측정하였다. 우리는 표적 부위 주변의 30-nt 및 50-nt 스페이서를 시험하였고 양쪽 가이드 길이가 예상되는 A에서 I로의 편집을 일으키고, 50-nt 스페이서가 더 높은 편집율을 달성했다는 것을 확인하였다 (도 50D,E, 도 57C). 우리는 또한 50-nt 스페이서가 아마도 증가된 듀플렉스 RNA 영역에 기인하여, 비표적화된 아데노신에서 편집을 위한 증가된 경향을 갖는다는 것을 관찰하였다 (도 50E, 도 57C).

[0901] 우리는 다음으로 내생성 유전자, *PPIB* 를 표적화하였다. 우리는 *PPIB* 를 타일링하는 50-nt 스페이서를 디자인하였고 *PPIB* 전사물을 최대 28% 편집 효율로 편집시킬 수 있다는 것을 확인하였다 (도 57D). REPAIR가 더욱 최적화될 수 있는가를 시험하기 위해서, 우리는 dCas13b 및 $ADAR2_{DD}$(E488Q) 사이의 링커를 변형시켰고 (도 57E, 표 8), 링커 선택이 적당하게 루시퍼라제 활성 복원에 영향을 미친다는 것을 확인하였다.

[0902] RNA 편집을 위한 서열 매개변수의 규정

[0903] 우리가 시험 부위에서 정확한 RNA 편집을 달성할 수 있었던 것을 고려하여, 우리는 전사체의 임의의 RNA 표적에 대해 시스템을 프로그래밍하기 위한 서열 제한성을 특징규명하고 싶었다. 서열 제한성은 dCas13b 표적화 제한, 예컨대 PFS, 또는 ADAR 서열 선호성으로부터 발생될 수 있었다 (26). REPAIRv1에 대한 PFS 제한성을 조사하기 위해서, 우리는 Cluc 전사물 상의 표적 부위의 5' 말단에 일련의 4개의 무작위 뉴클레오티드를 보유하는 플라스미드 라이브러리를 디자인하였다 (도 51A). 우리는 UAG 또는 AAC 모티프 내에 중심

아데노신을 표적화하였고, 양쪽 모티프의 경우에, 모든 PFS가 검출가능한 수준의 RNA 편집을 입증하였고, 대부분의 PFS는 표적 부위에서 편집이 50%를 초과한다는 것을 확인하였다 (도 51B). 다음으로, 우리는 $ADAR2_{DD}$ 에 대해 이전에 보고된 바와 같이, 표적화된 염기에 바로 측접된 임의 서열 제한성을 갖는지 여부를 결정하고자 하였다 (26). 우리는 표적 아데노신을 직접적으로 둘러싼 5' 및 3' 측접 뉴클레오티드의 모든 가능한 조합을 시험하였고 (도 51C), REPAIRv1이 모든 모티프를 편집할 수 있다는 것을 확인하였다 (도 51D). 마지막으로, 우리는 스페이서 서열의 표적 A 반대의 염기의 동일성이 편집 효율에 영향을 미치는지 여부를 분석하였고 A-C 미스매치가 최고 루시퍼라제 복원을 가졌고 A-G, A-U, 및 A-A는 REPAIRv1 활성을 극적으로 감소시킨다는 것을 확인하였다 (도 57F).

[0904] REPAIRv1을 사용한 질환-관련 인간 돌연변이의 교정

[0905] 포유동물에서 RAN 편집을 위한 REPAIRv1 시스템의 광범위한 적용가능성을 입증하기 위해서, 우리는 2종의 질환 관련 돌연변이: X-연관 신장성 요붕증에서 878G>A (*AVPR2* W293X) 및 판코니 빈혈의 1517G>A (*FANCC* W506X)에 대한 REPAIRv1 가이드를 디자인하였다. 우리는 HEK293FT 세포를 이들 돌연변이를 보유하는 유전자의 cDNA를 위한 발현 구축물로 트랜스펙션시켰고, REPAIRv1이 돌연변이를 교정할 수 있는지 여부를 시험하였다. 50-nt 스페이서를 함유하는 가이드 RNA를 사용하여, 우리는 *AVPR2* 의 35% 교정율 및 *FANCC* 의 23% 교정률을 획득할 수 있었다 (도 52A-D). 다음으로 우리는 34종의 상이한 질환-관련 G>A 돌연변이를 교정하는 REPAIRv1의 능력을 시험하였고 (표 9) 최대 28%의 편집 효율로 33개 부위에서 유의한 편집을 달성할 수 있다는 것을 확인하였다 (도 52E). 우리가 선택한 돌연변이는 또한 추가의 11,943종의 G에서 A로의 변이체를 포함하는, ClinVar 데이타베이스에서 병원성 G에서 A로의 돌연변이 (5,739)의 오직 일부분이다 (도 52F 및 도 58). 서열 제한이 없기 때문에, REPAIRv1은 특히 우리가 표적 모티프와 무관하게 유의한 편집을 관찰했다는 것을 고려하면, 모든 이들 질환 관련 돌연변이를 잠재적으로 편집할 수 있다 (도 51C 및 도 52G).

[0906] 질환 세포로 REPAIRv1 시스템의 전달은 치료 용도를 위한 전제조건이므로, 우리는 치료적으로 관련된 바이러스 벡터, 예컨대 아데노-연합 바이러스 (AAV) 벡터로 패키징될 수 있는 REPAIRv1 구축물을 디자인하고자 하였다. AAV 벡터는 4.7kb의 패키징 한계를 가져서, 프로모터 및 발현 조절 엘리먼트와 함께 $dCas13b\text{-}ADAR_{DD}$ (4473 bp)의 거대한 크기를 수용할 수 없다. 크기를 축소시키기 위해서, 우리는 RNA 편집 활성에 대해 $ADAR2_{DD}$(E488Q)에 융합된 dCas13의 다양한 N-말단 및 C-말단 절두를 시험하였다. 우리는 시험된 모든 C-말단 절두가 여전히 기능성이었고 루시퍼라제 신호를 복원할 수 있고 (도 59), 최대 절두, C-말단 Δ84-1090 (융합 단백질의 총 크기 4,152bp)이 AAV 벡터의 패키징 한계 내에서 맞추기에 충분히 작다는 것을 확인하였다.

[0907] REPAIRv1의 전사체-전반 특이성

[0908] PspCas13b으로 RNA 녹다운은 고도로 특이적이지만, 우리의 루시퍼라제 타일링 실험에서, 우리는 가이드:표적 듀플렉스 내에서 표적외 아데노신 편집을 관찰하였다 (도 50E). 이것이 광범위한 현상인지를 확인하기 위해서, 우리는 내생성 전사물, *KRAS* 을 타일링시켰고, 표적 아데노신 근처에서 표적외 편집 정도를 측정하였다 (도 53A). 우리는 *KRAS* 의 경우에, 표적-적중 편집율이 23%였지만, 역시 검출가능한 A에서 G로의 편집을 갖는 표적 부위 주변의 많은 부위가 존재하였다 (도 53B).

[0909] 가이드:표적 듀플렉스 내에서 관찰된 표적외 편집때문에, 우리는 모든 mRNA 상에서 RNA 서열분석을 수행하여 모든 가능한 전사체 표적-외를 평가하였다. RNA 서열분석은 많은 수의 A에서 G로의 표적-외 사건이 존재하였고, 표적화 조건에서 1,732 표적외 및 비표적화 조건에서 925 표적외가 존재하고, 828 표적-외 중복이 있다는 것을 밝혔다 (도 53C,D). 전사체에 걸쳐 모든 편집 부위 중에서, 표적-적중 편집 부위는 최고 편집율을 가졌는데, 89%로 A가 G로 전환되었다.

[0910] Cas13 표적화의 높은 특이성을 고려하여, 우리는 표적-외가 ADAR로부터 일어날 수 있다고 추리하였다. 2개 RNA-가이딩된 ADAR 시스템은 이전에 기술되었다 (도 60A). 첫번째는 BoxB-λ RNA 헤어핀에 결합하는, 소형 바이러스 단백질 람다 (λ)와 $ADAR2_{DD}$ 의 융합체를 이용한다 (22). 이중 BoxB-λ 헤어핀을 갖는 가이드 RNA는 가이드 RNA에서 코딩되는 편집 부위로 $ADAR2_{DD}$를 가이딩한다 (23). 두번째는 전체 길이 ADAR2 (ADAR2) 및 ADAR2의 이중 가닥 RNA 결합 도메인 (dsRBD)을 인식하는 헤어핀을 갖는 가이드 RNA를 이용한다 (21, 24). 우리는 REPAIRv1와 비교하여 이들 2개 시스템의 편집 효율을 분석하였고 BoxB-ADAR2 및 ADAR2 시스템이 REPAIRv1에 의해 획득된 89% 편집율과 비교하여 각각 63% 및 36% 편집율을 입증하였다는 것을 확인하였다 (도 60B-E). 추가로, BoxB 및 ADAR2 시스템은 REPAIRv1 표적화 가이드 조건에서 1,229

표적외와 비교하여, 표적화 가이드 조건에서, 개별적으로 2018 및 174 관찰 표적외를 생성하였다. 특히, 2개 $ADAR2_{DD}$-기반 시스템 (REPAIRv1 및 BoxB)에 의한 모든 조건은 그들의 표적-외에서 높은 중복 백분율을 나타내지만 ADAR2 시스템은 대체로 별개의 표적-외 세트를 가졌다 (도 60F). 표적화 및 비표적화 조건 간 및 REPAIRv1 및 BoxB 조건 간 표적-외의 중복은 $ADAR2_{DD}$는 dCas13 표적화와 독립적으로 표적-외를 구동했다는 것을 시사한다 (도 60F).

[0911] 이성적 단백질 조작을 통한 REPAIRv1의 특이성의 개선

[0912] REPAIR의 특이성을 개선시키기 위해서, 우리는 $ADAR2_{DD}$ (E488Q)의 구조-가이딩된 단백질 조작을 적용하였다. 표적-외의 가이드-독립적 성질때문에, 우리는 탈안정화 $ADAR2_{DD}$(E488Q)-RNA 결합은 선택적으로 표적-외 편집을 감소시키지만, $ADAR2_{DD}$(E488Q)의 dCas13b 속박부터 표적 부위까지 증가되는 국소 농도에 기인하여 표적-적중 편집을 유지한다고 가정하였다. 우리는 $ADAR2_{DD}$(E488Q) 배경 상에서 표적 RNA (도 54A) (18)의 듀플렉스 영역에 접촉하는 것으로 이전에 결정된 $ADAR2_{DD}$(E488Q) 잔기를 돌연변이유발시켰다. 효율 및 특이성을 평가하기 위해서, 우리는 검출된 비표적화 조건에서 배경 루시퍼라제 복원이 더 광범위한 표적-외 활성을 의미한다는 조건 하에서, 표적화 및 비표적화 가이드 둘 모두로 17 단일 돌연변이체를 시험하였다. 우리는 선택된 잔기에서 돌연변이가 표적화 및 비표적화 가이드에 대한 루시퍼라제 활성에 대한 유의한 효과를 갖는다는 것을 확인하였다 (도 54A,B, 도 61A). 대부분의 돌연변이체는 표적화 가이드에 대한 루시퍼라제 활성을 유의하게 개선시켰거나 또는 비표적화 가이드 활성에 대한 표적화의 비율을 증가시켰고, 이것을 우리는 특이적 점수라고 하였다 (도 54A,B). 우리는 차세대 서열분석에 의해서 전사체-전반 특이성 프로파일링에 대해 이들 돌연변이체의 서브셋를 선택하였다 (도 54B). 예상한 바와 같이, 표적-외는 변이체에 대해 우리의 특이적 점수와 상관있는 전사체-전반 서열분석으로부터 측정되었다 (도 61B). 우리는 $ADAR2_{DD}$(E488Q/R455E)를 제외하고, 모든 서열분석된 REPAIRv1 돌연변이체는 리포터 전사물을 효과적으로 편집할 수 있고 (도 54C), 많은 돌연변이체는 표적-외의 수의 감소를 보였다는 것을 확인하였다 (도 61C, 62). 우리는 특이성 돌연변이체에 대한 표적외의 주변 모티프를 조사하였고, REPAIRv1 및 대부분의 조작된 돌연변이체가 특징적인 ADAR2 모티프와 일관되게, 그들 편집에 대해 강력한 3' G 선호도를 나타냈다는 것을 확인하였다 (도 63A)(26). 우리는 향후 실험을 위해서 돌연변이체 $ADAR2_{DD}$(E488Q/T375G)를 선택하였고, 전사체-전반 표적외의 최저 개수를 갖는 4개 변이체의 최고 편집 백분율을 갖기 때문이고 REPAIRv2라고 하였다. REPAIRv1과 비교하여, REPAIRv2는 증가된 특이성을 나타냈고, 1732에서 10으로의 전사체 표적외 감소가 있었다 (도 54D). Cluc에서 표적화 아데노신 주변의 영역에서, REPAIRv2는 서열분석 추적으로 가시화된, 감소된 표적외 편집을 가졌다 (도 54E). 차세대 서열분석으로부터 유래된 모티프에서, REPAIRv1은 3' G에 대해 강력한 선호도를 보이지만, 모든 모티프에 대해 표적-외 편집을 보였으며 (도 63B); 대조적으로, REPAIRv2는 오직 가장 강력한 표적외-모티프를 편집하였다 (도 63C). 전사물 상의 편집 분포는 심하게 왜곡되었는데, 고도로-편집된 유전자는 60 편집 이상을 갖지만 (도 64A,B), REPAIRv2는 오직 하나의 전사물 (*EEF1A1*)을 다수회 편집하였다 (도 64D-F). REPAIRv1 표적-외 편집은 93 종양발생성 사건 (도 64D)과, 1000 미스센스 돌연변이 (도 64C)를 포함하여, 수많은 변이체를 야기시키는 것으로 예상되었다. 대조적으로, REPAIRv2는 오직 6 미스센스 돌연변이를 가졌고 (도 64E), 그 어떠한 것도 종양발생 결과를 갖지않았다 (도 64F). 예상되는 표적-외 효과의 이러한 감소는 REPAIRv2를 다른 RNA 편집 접근법과 구별되게 한다.

[0913] 우리는 높은-효율의 표적-적중 편집을 유지하면서 표적 전사물의 근처 편집부를 특이성-증강 돌연변이가 감소시키는지 여부를 시험하기 위해서 내생성 유전자에 대해 REPAIRv2를 표적화하였다. *KRAS* 또는 *PPIB* 를 표적화하는 가이드의 경우에, 우리는 REPAIRv1과 달리, 검출가능한 표적외 편집을 갖지 않았고, 각각 27.1% 및 13%로 표적-적중 아데노신을 효과적으로 편집할 수 있다는 것을 확인하였다 (도 54F,G). 이러한 특이성은 REPAIRv1, 예컨대 다른 *KRAS* 또는 *PPIB* 표적 부위에 대한 비표적화 조건에서 높은 수준의 배경을 입증하는 영역을 포함하여, 추가의 표적 부위로 확장되었다 (도 65). 전체적으로, REPAIRv2는 편집된 아데노신 주변의 듀플렉스된 영역에서 표적-외를 제거하였고 극적으로 증강된 전사체-전반 특이성을 보였다.

[0914] 결론

[0915] 우리는 여기서 RNA-가이딩된 RNA-표적화 VI-B형 이펙터 Cas13b가 고도로 효율적이고 특이적으로 RNA 녹다운을 할 수 있다는 것을 확인하여, 전사체 수준에서 세포 프로세스를 제어할 뿐만 아니라 필수 유전자 및 넌코딩 RNA

를 심문하는 개선된 도구의 기초를 제공한다. 촉매적으로 불활성인 Cas13b (dCas13b)는 프로그램가능한 RNA 결합 능력을 보유하고, 이것은 정확한 A에서 I 편집을 달성하기 위해 우리는 dCas13b를 아데노신 디아미나제 ADAR2에 융합시켜 영향을 미쳤고, 이 시스템을 REPAIRv1 (RNA Editing for Programmable A to I Replacement version 1)라고 하였다. 시스템의 추가 조작은 극적으로 개선된 특이성을 갖는 현행 편집 플랫폼에 비해서 비슷하거나 또는 증가된 활성을 갖는 방법인 REPAIRv2를 생성시켰다.

[0916] Cas13b가 높은 신뢰성을 나타내지만, dCas13b-ADAR2$_{DD}$ 융합체에 의한 우리의 초기 결과는 수천개의 표적-외를 밝혀주었다. 이것을 해결하기 위해서, 우리는 RNA 듀플렉스와 접촉하는 ADAR2$_{DD}$ 잔기를 변이시키고, dCas13b와 융합시 정확하고, 효율적이며 고도로 특이적인 편집을 할 수 있는, 변이체 ADAR2$_{DD}$(E488Q/T375G)를 동정하기 위한 합리적인 돌연변이유발 전략을 적용하였다. 이러한 변이체에 의한 편집 효율은 2개의 현재 이용가능한 시스템, BoxB-ADAR$_{DD}$ 또는 ADAR2 편집으로 획득되는 것과 비슷하거나 또는 더 양호하였다.

게다가, REPAIRv2 시스템은 양쪽 대안적인 편집 기술보다 적어도 한자릿수가 더 양호하게, 전체 전사체에서 오직 10의 관찰가능한 표적-외를 생성시켰다.

[0917] REPAIR 시스템은 다른 핵산 편집 도구와 비교하여 많은 장점을 제공한다. 첫째로, 정확한 표적 부위는 ADAR 편집 활성에 이상적인 바람직한 A-C 미스매치를 생성시키도록 바람직한 아데노신으로부터 가이드 연장 내에 시티딘을 위치시킴으로써 가이드에서 코딩될 수 있다. 두번째로, Cas13은 표적화 서열 제한, 예컨대 PFS 또는 PAM이 없고, 표적 아데노신 주변에 모티프 선호도가 없어서, 전사체 내 임의 아데노신을 REPAIR 시스템에 의해 잠재적으로 표적화할 수 있게 한다. 그러나, DNA 염기 편집자는 센스 또는 안티센스 가닥을 표적화할 수 있는 한편, REPAIR 시스템은 전사된 서열에 제한적이어서, 우리가 표적으로 할 수 있는 가능한 편집 부위의 총 개수를 제한한다는 것을 주목한다. 그러나, REPAIR에 의한 표적화의 보다 탄력적인 성질에 기인하여, 이 시스템은 Cas9-DNA 염기 편집자보다 ClinVar 내에서 더 많은 편집을 실시할 수 있다 (도 52C).

세째로, REPAIR 시스템은 표적 아데노신을 이노신으로 직접적으로 탈아민화시키고 바람직한 편집 결과를 생성시키기 위해서, 염기-절제 또는 미스매치 복구와 같은 내생성 복구 경로에 의존하지 않는다. 따라서, REPAIR는 다른 편집 형태를 지원할 수 없는 비분열 세포에서 가능할 수 있다. 네째로, RNA 편집은 일시적일 수 있어서, 편집 결과에 대한 일시적 제어에 대한 잠재성을 가능하게 한다. 이러한 특성은 아마도 국소 염증과 같은, 세포 상태의 일시적 변화에 의해 야기된 질환을 치료하는데 유용할 것이다.

[0918] REPAIR 시스템은 추가의 조작을 위한 다수의 기회를 제공한다. Cas13b는 프리-crRNA 프로세싱 활성을 보유하여서(13), 단독으로 질환 위험성을 변경시키지 않지만, 함께 가산적 효과 및 질환-변형 잠재성을 가질 수 있는, 복수 변이체의 다중 편집을 가능하게 한다. 유망한 돌연변이를 조합하는 것과 같이, 우리의 합리적인 디자인 접근법의 연장은 시스템의 특이성 및 효율성을 더 증가시킬 수 있는 한편, 비편향된 스크리닝 접근법은 REPAIR 활성 및 특이성을 개선시키기 위한 추가의 잔기를 동정할 수 있다.

[0919] 현재, REPAIR에 의해 획득가능한 염기 전환은 아데노신으로부터 이노신을 생성시키는데 제한되고, 다른 촉매적 RNA 편집 도메인, 예컨대 APOBEC과 dCas13의 추가 융합은 시티딘을 우리딘으로 편집할 수 있게 한다. 추가적으로, ADAR의 돌연변이유발은 표적 시티딘에 대한 기질 선호도를 완화시킬 수 있어서, 듀플렉스 RNA 기질 요건에 의해 부여되는 증강된 특이성으로 C->U 편집자에 의해 탐색될 수 있게 한다. DNA 기질 상에서 아데노신에서 이노신으로의 편집은 또한 DNA-RNA 헤테로듀플렉스 표적의 형성 (27) 또는 ADAR 도메인의 돌연변이유발을 통해서, 촉매적으로 불활성인 DNA-표적화 CRISPR 이펙터, 예컨대 dCas9 또는 dCpf1를 사용해 가능할 수 있다.

[0920] REPAIR은 A에서 I (A에서 G) 편집이 질환 진행을 반전시킬 수 있거나 또는 둔화시킬 수 있는 광범위한 치료적 징후에 대해 적용될 수 있다 (도 66). 첫째로, 질환-관련 조직에서 일상적인 멘델의 G에서 A로의 돌연변이를 표적화하기 위한 REPAIR의 발현은 유해한 돌연변이를 반전시키고 질환을 치료하는데 사용될 수 있다.

예를 들어, 뇌 조직에서 AAV를 통한 안정한 REPAIR 발현은 초점성 간질을 일으키는 *GRIN2A* 미스센스 돌연변이 c.2191G>A (Asp731Asn) (28) 또는 알츠하이머 질환의 초기 발생을 야기시키는 *APP* 미스센스 돌연변이 c.2149G>A (Val717Ile) (29)를 교정하는데 사용될 수 있다. 두번째로, REPAIR는 질환-관련 신호전달에 관여되는 단백질의 기능을 변형시켜서 질환을 치료하는데 사용될 수 있다. 예를 들어, REPAIR 편집은 키나제의 표적인 일부 세린, 트레오닌 및 티로신 잔기의 재-코딩을 가능하게 한다 (도 66). 질환-관련 단백질에서 이들 잔기의 인산화는 알츠하이머 질환 및 다발성 신경 퇴행성 병태를 포함한 많은 질병에 대한 질환 진행에 영향을 미친다 (30). 세째로, REPAIR는 환자에 대해 질환 상태로 진입하는 기회를 미리 감소시키기 위해 발현된, 위험-수정 G에서 A로의 변이체의 서열을 변화시키는데 사용될 수 있다. 가장 흥미

로운 경우는 질환 상태로 진입할 기회를 극적으로 감소시키고, 일부 경우에, 추가의 건강 이득을 부여하는, '보호적' 위험-변형 대립유전자이다. 예를 들어, REPAIR는 각각 심혈관 질환 및 건선성 관절염에 대해 보호하는 *PCSK9* 및 *IFIH1* 의 A에서 G 대립유전자를 기능적으로 모방하는데 사용될 수 있다 (31). 마지막으로, REPAIR는 엑손 조절 요법을 위해 슬라이스 억셉터 및 도너 부위를 치료적으로 변형시키는데 사용될 수 있다. REPAIR는 AU를 IU로 또는 AA를 AI로 변화시킬 수 있는데, 각각 공통 5' 스플라이스 도너 또는 3' 스플라이스 억셉터 부위와 기능적으로 등가이고, 새로운 슬라이스 접합부를 생성시킨다. 추가적으로, REPAIR 편집은 공통 3' 스플라이스 억셉터 부위를 AG->IG로 돌연변이시켜서 인접한 하류 엑손의 스킵핑을 촉진할 수 있고, DMD의 치료를 위한 상당힌 관심을 받는 치료적 전략일 수 있다. 슬라이스 부위의 조절은 예컨대 척수성 근위축증의 치료를 위한 SMN2 스플라이싱의 조절을 위해, 안티센스 올리고가 일부 성공을 거둔 질환에서 광범위 적용성을 가질 수 있다 (32).

[0921] 우리는 RNA 녹다운 및 RNA 편집 도구로서 PspCas13b 효소의 용도를 입증하였다. 프로그램가능한 RNA 결합을 위한 dCas13b 플랫폼은 간 전사물 이미지화, 스플라이싱 변형, 전사물의 표적화 국재화, RNa-결합 단백질의 풀다운, 및 후성전사체 변형을 포함한, 많은 적용성을 갖는다. 여기서 우리는 dCas13을 사용하여 핵산 편집 기술의 현행 제품에 첨가되는 REPAIR를 생성시켰다. REPAIR는 유전자 질환을 치료하고 보호적 대립유전자를 모장하기 위한 신규한 접근법을 제공하고, 유전자 기능을 변형시키기 위한 유용한 도구로서 RNA 편집을 확립시킨다.

표 6 – 본 연구에서 사용된 **Cas13** 오르소로그

Cas13 ID	Cas13 약어	숙주 유기체	단백질 접근번호
Cas13a1	LshCas13a	Leptotrichia shahii	WP_018451595.1
Cas13a2	LwaCas13a	Leptotrichia wadei (Lw2)	WP_021746774.1
Cas13a3	LseCas13a	Listeria seeligeri	WP_012985477.1
Cas13a4	LbmCas13a	Lachnospiraceae bacterium MA2020	WP_044921188.1
Cas13a5	LbnCas13a	Lachnospiraceae bacterium NK4A179	WP_022785443.1
Cas13a6	CamCas13a	[Clostridium] aminophilum DSM 10710	WP_031473346.1
Cas13a7	CgaCas13a	Carnobacterium gallinarum DSM 4847	WP_034560163.1
Cas13a8	Cga2Cas13a	Carnobacterium gallinarum DSM 4847	WP_034563842.1
Cas13a9	Pprcas13a	Paludibacter propionicigenes WB4	WP_013443710.1
Cas13a10	LweCas13a	Listeria weihenstephanensis FSL R9-0317	WP_036059185.1
Cas13a11	LbfCas13a	Listeriaceae bacterium FSL M6-0635	WP_036091002.1
Cas13a12	Lwa2cas13a	Leptotrichia wadei F0279	WP_021746774.1
Cas13a13	RcsCas13a	Rhodobacter capsulatus SB 1003	WP_013067728.1
Cas13a14	RcrCas13a	Rhodobacter capsulatus R121	WP_023911507.1
Cas13a15	RcdCas13a	Rhodobacter capsulatus DE442	WP_023911507.1
Cas13a16	LbuCas13a	Leptotrichia buccalis C-1013-b	WP_015770004.1
Cas13a17	HheCas13a	Herbinix hemicellulosilytica	CRZ35554.1
Cas13a18	EreCas13a	[Eubacterium] rectale	WP_055061018.1
Cas13a19	EbaCas13a	Eubacteriaceae bacterium CHKCI004	WP_090127496.1
Cas13a20	BmaCas13a	Blautia sp. Marseille-P2398	WP_062808098.1
Cas13a21	LspCas13a	Leptotrichia sp. oral taxon 879 str. F0557	WP_021744063.1
Cas13b1	BzoCas13b	Bergeyella zoohelcum	WP_002664492
Cas13b2	PinCas13b	Prevotella intermedia	WP_036860899
Cas13b3	PbuCas13b	Prevotella buccae	WP_004343973
Cas13b4	AspCas13b	Alistipes sp. ZOR0009	WP_047447901
Cas13b5	PsmCas13b	Prevotella sp. MA2016	WP_036929175

[0922]

Cas13b6	RanCas13b	Riemerella anatipestifer	WP_004919755
Cas13b7	PauCas13b	Prevotella aurantiaca	WP_025000926
Cas13b8	PsaCas13b	Prevotella saccharolytica	WP_051522484
Cas13b9	Pin2Cas13b	Prevotella intermedia	WP_061868553
Cas13b10	CcaCas13b	Capnocytophaga canimorsus	WP_013997271
Cas13b11	PguCas13b	Porphyromonas gulae	WP_039434803
Cas13b12	PspCas13b	Prevotella sp. P5-125	WP_044065294
Cas13b13	FbrCas13b	Flavobacterium branchiophilum	WP_014084666
Cas13b14	PgiCas13b	Porphyromonas gingivalis	WP_053444417
Cas13b15	Pin3Cas13b	Prevotella intermedia	WP_050955369
Cas13c1	FnsCas13c	Fusobacterium necrophorum subsp. funduliforme ATCC 51357 contig00003	WP_005959231.1
Cas13c2	FndCas13c	Fusobacterium necrophorum DJ-2 contig0065, whole genome shotgun sequence	WP_035906563.1
Cas13c3	FnbCas13c	Fusobacterium necrophorum BFTR-1 contig0068	WP_035935671.1
Cas13c4	FnfCas13c	Fusobacterium necrophorum subsp. funduliforme 1_1_36S cont1.14	EHO19081.1
Cas13c5	FpeCas13c	Fusobacterium perfoetens ATCC 29250 T364DRAFT_scaffold00009.9_C	WP_027128616.1
Cas13c6	FulCas13c	Fusobacterium ulcerans ATCC 49185 cont2.38	WP_040490876.1
Cas13c7	AspCas13c	Anaerosalibacter sp. ND1 genome assembly Anaerosalibacter massiliensis ND1	WP_042678931.1

[0923]

표 7 박테리아 스크린에서의 **PFS** 컷오프

Cas13b 오르소로그	핵심	**PFS** 모티프를 생성하는데 사용된 - **Log_2** 감손 점수
Bergeyella zoohelcum	1	2
Prevotella intermedia locus 1	2	1
Prevotella buccae	3	3
Alistipes sp. ZOR0009	4	1
Prevotella sp. MA2016	5	2
Riemerella anatipestifer	6	4
Prevotella aurantiaca	7	1
Prevotella saccharolytica	8	0
Prevotella intermedia locus 2	9	0
Capnocytophaga canimorsus	10	3
Porphyromonas gulae	11	4
Prevotella sp. P5-125	12	2.1
Flavobacterium branchiophilum	13	1
Porphyromonas gingivalis	14	3
Prevotella intermedia locus 2	15	4

[0924]

[0925]

표 8 - 포유동물 세포에서 RNA 편집을 위해 본 연구에서 사용된 dCas13b-ADAR 링커 서열	
도	링커
50C	GSGGGGS
50E	GS
57B	GSGGGGS
57C	GS
57D	GS
57E: GS	GS
57E: GSGGGGS	GSGGGGS
57E: (GGGS)3	GGGGSGGGGSGGGGS
57E: Rigid	EAAAK
57E: (GGS)6	GGSGGSGGSGGSGGSGGS
57E: XTEN	SGSETPGTSESATPES
51B	GS
57F	GS

[0926]

51C	GS
52B	GS
52D	GS
52E	GS
51A: Δ984-1090, Δ1026-1090, Δ1053-1090	GS
51A: Δ1-125, Δ1-88, Δ1-72	GSGGGGS
53B	GS
53C	GS
53D	GS
60A	GS
60C	GS
60D	GS
61D	GS
54A	GS
62A	GS
54B	GS
62B	GS
62C	GS
63A	GS
63B	GS
54C	GS
54D	GS
54E	GS
54F	GS
66A	GS
66A	GS

[0927]

표 9 - 질환-관련 돌연변이에 대한 질환 정보		
전체 길이 후보자	유전자	질환
NM_000054.4(AVPR2):c.878G>A (p.Trp293Ter)	AVPR2	신성요붕증, X-연관됨
NM_000136.2(FANCC):c.1517G>A (p.Trp506Ter)	FANCC	판코니 빈혈, 보완군 C
부가적인 시뮬레이션된 후보자		
후보자	유전자	질환
NM_000206.2(IL2RG):c.710G>A (p.Trp237Ter)	IL2RG	X-연관 심각한 복합 면역결핍
NM_000132.3(F8):c.3144G>A (p.Trp1048Ter)	F8	유전 인자 VIII 결핍성 질환
NM_000527.4(LDLR):c.1449G>A (p.Trp483Ter)	LDLR	가족성 고콜레스테롤혈증
NM_000071.2(CBS):c.162G>A (p.Trp54Ter)	CBS	CBS 결핍증으로 인한 호모시스틴뇨증
NM_000518.4(HBB):c.114G>A (p.Trp38Ter)	HBB	베타^0^ 지중해빈혈|베타 지중해빈혈
NM_000035.3(ALDOB):c.888G>A (p.Trp296Ter)	ALDOB	유전성 과당뇨증
NM_004006.2(DMD):c.3747G>A (p.Trp1249Ter)	DMD	듀켄씨근이영양증
NM_005359.5(SMAD4):c.906G>A (p.Trp302Ter)	SMAD4	청소년 폴립증 증후군
NM_000059.3(BRCA2):c.582G>A (p.Trp194Ter)	BRCA2	가족력 유방암|유방-난소암, 가족력 2
NM_000833.4(GRIN2A):c.3813G>A (p.Trp1271Ter)	GRIN2A	간질, 초점성, 언어 장애가 있고 정신 지체가 있거나 없음
NM_002977.3(SCN9A):c.2691G>A (p.Trp897Ter)	SCN9A	통증에 대한 무감각, 선천성, 상염색체 열성
NM_007375.3(TARDBP):c.943G>A (p.Ala315Thr)	TARDBP	근위축성 측색 경화증 타입 10
NM_000492.3(CFTR):c.3846G>A (p.Trp1282Ter)	CFTR	낭포성 섬유증|유전성 췌장염|제공되지 않음|아탈루렌(ataluren) 반응 - 효능
NM_130838.1(UBE3A):c.2304G>A (p.Trp768Ter)	UBE3A	안젤만 증후군
NM_000543.4(SMPD1):c.168G>A (p.Trp56Ter)	SMPD1	니만-피크병, 타입 A
NM_206933.2(USH2A):c.9390G>A	USH2A	어셔 증후군, 타입 2A

(p.Trp3130Ter)		
NM_130799.2(MEN1):c.1269G>A (p.Trp423Ter)	MEN1	유전성 암-소인적 증후군
NM_177965.3(C8orf37):c.555G>A (p.Trp185Ter)	C8orf37	색소성 망막염 64
NM_000249.3(MLH1):c.1998G>A (p.Trp666Ter)	MLH1	린치 증후군
NM_000548.4(TSC2):c.2108G>A (p.Trp703Ter)	TSC2	결절성 경화증 2|결절성 경화증 증후군
NM_000267.3(NF1):c.7044G>A (p.Trp2348Ter)	NF1	신경섬유종증, 타입 1
NM_000179.2(MSH6):c.3020G>A (p.Trp1007Ter)	MSH6	린치 증후군
NM_000344.3(SMN1):c.305G>A (p.Trp102Ter)	SMN1	척수성 근위축, 타입 II| 쿠겔베르그-웰란더병
NM_024577.3(SH3TC2):c.920G>A (p.Trp307Ter)	SH3TC2	샤르코-마리-치아 질환, 타입 4C
NM_001369.2(DNAH5):c.8465G>A (p.Trp2822Ter)	DNAH5	원발성 섬모운동 이상증
NM_004992.3(MECP2):c.311G>A (p.Trp104Ter)	MECP2	레트 증후군
NM_032119.3(ADGRV1):c.7406G>A (p.Trp2469Ter)	ADGRV1	어셔 증후군, 타입 2C
NM_017651.4(AHI1):c.2174G>A (p.Trp725Ter)	AHI1	주버트 증후군 3
NM_004562.2(PRKN):c.1358G>A (p.Trp453Ter)	PRKN	파킨슨병 2
NM_000090.3(COL3A1):c.3833G>A (p.Trp1278Ter)	COL3A1	엘러스-단로스 증후군, 타입 4
NM_007294.3(BRCA1):c.5511G>A (p.Trp1837Ter)	BRCA1	가족력 유방암|유방-난소암, 가족력 1
NM_000256.3(MYBPC3):c.3293G>A (p.Trp1098Ter)	MYBPC3	원발성 가족력 비대 심근증
NM_000038.5(APC):c.1262G>A (p.Trp421Ter)	APC	가족력 대장 폴립증 1
NM_001204.6(BMPR2):c.893G>A (p.W298*)	BMPR2	원발성 폐 고혈압

[0928]

표 10 – 본 연구에서 사용된 핵심 플라스미드	
플라스미드 명칭	설명
pAB0006	CMV-C 루시페라제-polyA EF1a-G-루시페라제-polyA
pAB0040	CMV-C 루시페라제(STOP85)-polyA EF1a-G-루시페라제-polyA
pAB0048	pCDNA-ADAR2-N-말단 B12-HIV NES
pAB0050	pAB0001-A02 (crRNA 백본)
pAB0051	pAB0001-B06 (crRNA 백본)
pAB0052	pAB0001-B11 (crRNA 백본)
pAB0053	pAB0001-B12 (crRNA 백본)
pAB0014.B6	EF1a-BsiWI-Cas13b6-NES-mapk
pAB0013.B11	EF1a-BsiWI-Cas13b11-NES-HIV
pAB0013.B12	EF1a-BsiWI-Cas13b12-NES-HIV
pAB0051	pAB0001-B06 (crRNA 백본)
pAB0052	pAB0001-B11 (crRNA 백본)
pAB0053	pAB0001-B12 (crRNA 백본)
pAB0079	pCDNA-ADAR1hu-EQ 돌연변이체-N-말단 도착 벡터
pAB0085	pCDNA-ADAR2 (E488Q)hu-EQ 돌연변이체-N-말단 도착 벡터
pAB0095	EF1a-BsiWI-Cas13-B12-NES-HIV, 이중 H HEPN 돌연변이체를 가짐
pAB0114	pCDNA-wtADAR2hu-EQ 돌연변이체-N-말단 도착 벡터
pAB0120	루시페라제 ADAR 가이드 최적 (wC0054 유래 가이드 24)
pAB0122	ADAR 실험 을 위한 pAB0001-B12 NT 가이드
pAB0151	pCDNA-ADAR2hu-EQ 돌연변이체-N-말단 도착 벡터 C-말단 델타 984-1090
pAB0180	T375G 특이성 돌연변이체
pAB0181	T375G Cas13b C-말단 델타 984-1090

[0929]

표 11 - 포유동물 세포에서 녹다운을 위해 본 연구에서 사용된 가이드/shRNA 서열				
명칭	스페이서 서열	간섭 기작	주석	첫번째 도면
박테리아 PFS 가이드	GCCAGCUUUCCGGGCAUUGGCUUCCAUC	Cas13b	모든 오르소로그에 대해 사용됨	
Cas13a-Gluc 가이드 1	GCCAGCTTTCCGGGCATTGGCTTCCATC	Cas13a	모든 Cas13a 오르소로그에 대해 사용됨	도 49B
Cas13a-	ACCCAGGAATCTCAGG	Cas13a	모든 Cas13a	도 49B

[0930]

Gluc 가이드 2	AATGTCGACGAT		오르소로그에 대해 사용됨	
Cas13a-비-표적화 가이드 (LacZ)	AGGGTTTTCCCAGTCACGACGTTGTAAA	Cas13a	모든 Cas13a 오르소로그에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 1.1	GGGCATTGGCTTCCATCTCTTTGAGCACCT	Cas13b	오르소로그 1-3, 6, 7, 10, 11, 12, 14, 15 에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 1.2	GUGCAGCCAGCUUUCCGGGCAUUGGCUUCC	Cas13b	오르소로그 4 에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 1.3	GCAGCCAGCUUUCCGGGCAUUGGCUUCCAU	Cas13b	오르소로그 5 에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 1.4	GGCUUCCAUCUCUUUGAGCACCUCCAGCGG	Cas13b	오르소로그 8, 9 에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 1.5	GGAAUGUCGACGAUCGCCUCGCCUAUGCCG	Cas13b	오르소로그 13 에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 2.1	GAAUGUCGACGAUCGCCUCGCCUAUGCCGC	Cas13b	오르소로그 1-3, 6, 7, 10, 11, 14, 15 에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 2.2	GACCUGUGCGAUGAACUGCUCCAUGGGCUC	Cas13b	오르소로그 12 에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 2.2	GUGUGGCAGCGUCCUGGGAUGAACUUCUUC	Cas13b	오르소로그 4 에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 2.3	GUGGCAGCGUCCUGGGAUGAACUUCUUCAU	Cas13b	오르소로그 5 에 대해 사용됨	도 49B

[0931]

Cas13b-Gluc 가이드 2.4	GCUUCUUGCCGGGCAACUUCCCGCGGUCAG	Cas13b	오르소로그 8, 9에 대해 사용됨	도 49B
Cas13b-Gluc 가이드 2.6	GCAGGGUUUUCCCAGUCACGACGUUGUAAAA	Cas13b	오르소로그 13에 대해 사용됨	도 49B
Cas13b-비 표적화 가이드	GCAGGGUUUUCCCAGUCACGACGUUGUAAAA	Cas13b	모든 오르소로그에 대해 사용됨	도 49B
Cas13a-Gluc 가이드-RNASeq	ACCCAGGAAUCUCAGGAAUGUCGACGAU	Cas13a		도 49E
shRNA-Gluc 가이드	CAGCTTTCCGGGCATTGGCTT	shRNA		도 49F
Cas13b-Gluc 가이드-RNASeq	CCGCUGGAGGUGCUCAAAGAGAUGGAAGCC	Cas13b		도 49F
Cas13a-Gluc-가이드-1	GCCAGCTTTCCGGGCATTGGCTTCCATC	Cas13a		도 56A
Cas13a-Gluc-가이드-2	ACCCAGGAATCTCAGGAATGTCGACGAT	Cas13a		도 56A
Cas13b-Gluc-opt-가이드-1	GGGCATTGGCTTCCATCTCTTTGAGCACCT	Cas13b		도 56A
Cas13b-Gluc-opt-가이드-2	GAAUGUCGACGAUCGCCUCGCCUAUGCCGC	Cas13b		도 56A
Cas13a KRAS 가이드 1	CAAGGCACTCTTGCCTACGCCACCAGCT	Cas13a		도 56B

[0932]

Cas13a KRAS 가이드 2	TCATATTCGTCCACAAA ATGATTCTGAA	Cas13a		도 56B
Cas13a KRAS 가이드 3	ATTATTTATGGCAAATA CACAAAGAAAG	Cas13a		도 56B
Cas13a KRAS 가이드 4	GAATATCTTCAAATGAT TTAGTATTATT	Cas13a		도 56B
Cas13a KRAS 가이드 5	ACCATAGGTACATCTTC AGAGTCCTTAA	Cas13a		도 56B
Cas13b KRAS 가이드 1	GTCAAGGCACTCTTGCC TACGCCACCAGCT	Cas13b		도 56B
Cas13b KRAS 가이드 2	GATCATATTCGTCCACA AAATGATTCTGAA	Cas13b		도 56B
Cas13b KRAS 가이드 3	GTATTATTTATGGCAAA TACACAAAGAAAG	Cas13b		도 56B
Cas13b KRAS 가이드 4	GTGAATATCTTCAAATG ATTTAGTATTATT	Cas13b		도 56B
Cas13b KRAS 가이드 5	GGACCATAGGTACATCT TCAGAGTCCTTAA	Cas13b		도 56B
shRNA KRAS 가이드 1	aagagtgccttgacgatacagcCTC GAGgctgtatcgtcaaggcactctt	shRNA		도 56B
shRNA KRAS 가이드 2	aatcattttgtggacgaatatCTCGA Gatattcgtccacaaaatgatt	shRNA		도 56B
shRNA KRAS 가이드 3	aaataatactaaatcatttgaCTCGA Gtcaaatgatttagtattattt	shRNA		도 56B
shRNA KRAS 가이드 4	aataatactaaatcatttgaaCTCGA Gttcaaatgatttagtattatt	shRNA		도 56B
shRNA KRAS 가이드 5	aaggactctgaagatgtacctCTCG AGaggtacatcttcagagtcctt	shRNA		도 56B

[0933]

[0934]

표 12 - Gluc 녹다운을 위해 사용된 가이드 서열				
명칭	스페이서 서열	위치	주석	첫번째 도면
Gluc 타일링 가이드 1	GAGATCAGGGCAAAC AGAACTTTGACTCCC	2	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 2	GGATGCAGATCAGGGC AAACAGAACTTTGA	7	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 3	GCACAGCGATGCAGAT CAGGGCAAACAGAA	13	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 4	GCTCGGCCACAGCGAT GCAGATCAGGGCAA	19	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 5	GGGGCTTGGCCTCGGC CACAGCGATGCAGA	28	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 6	GTGGGCTTGGCCTCGG CCACAGCGATGCAG	29	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 7	GTCTCGGTGGGCTTGG CCTCGGCCACAGCG	35	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 8	GTTCGTTGTTCTCGGTG GGCTTGGCCTCGG	43	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 9	GGAAGTCTTCGTTGTT CTCGGTGGGCTTGG	49	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해	49C

			절단됨에 주의	
Gluc 타일링 가이드 10	GATGTTGAAGTCTTCG TTGTTCTCGGTGGG	54	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 11	GCGGCCACGATGTTGA AGTCTTCGTTGTTC	62	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 12	GTGGCCACGGCCACGA TGTTGAAGTCTTCG	68	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 13	GGTTGCTGGCCACGGC CACGATGTTGAAGT	73	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 14	GTCGCGAAGTTGCTGG CCACGGCCACGATG	80	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 15	GCCGTGGTCGCGAAGT TGCTGGCCACGGCC	86	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 16	GCGAGATCCGTGGTCG CGAAGTTGCTGGCC	92	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 17	GCAGCATCGAGATCCG TGGTCGCGAAGTTG	98	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 18	GGGTCAGCATCGAGAT CCGTGGTCGCGAAG	101	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 19	GCTTCCCGCGGTCAGC ATCGAGATCCGTGG	109	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C

[0935]

Gluc 타일링 가이드 20	GGGGCAACTTCCCGCG GTCAGCATCGAGAT	115	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 21	GTCTTGCCGGGCAACT TCCCGCGGTCAGCA	122	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 22	GGCAGCTTCTTGCCGG GCAACTTCCCGCGG	128	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 23	GCCAGCGGCAGCTTCT TGCCGGGCAACTTC	134	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 24	GCACCTCCAGCGGCAG CTTCTTGCCGGGCA	139	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 25	GCTTTGAGCACCTCCA GCGGCAGCTTCTTG	146	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 26	GCATCTCTTTGAGCAC CTCCAGCGGCAGCT	151	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 27	GTCCATCTCTTTGAGC ACCTCCAGCGGCAG	153	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 28	GGGCATTGGCTTCCAT CTCTTTGAGCACCT	163	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 29	GTCCGGGCATTGGCTT CCATCTCTTTGAGC	167	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc	GGCCAGCTTTCCGGGC	175	Cas13a 스페이서가 5'	49C

[0936]

타일링 가이드 30	ATTGGCTTCCATCT		말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	
Gluc 타일링 가이드 31	GGGTGCAGCCAGCTTTCCGGGCATTGGCTT	181	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 32	GAGCCCCTGGTGCAGCCAGCTTTCCGGGCA	188	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 33	GATCAGACAGCCCCTGGTGCAGCCAGCTTT	195	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 34	GGCAGATCAGACAGCCCCTGGTGCAGCCAG	199	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 35	GACAGGCAGATCAGACAGCCCCTGGTGCAG	203	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 36	GTGATGTGGGACAGGCAGATCAGACAGCCC	212	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 37	GACTTGATGTGGGACAGGCAGATCAGACAG	215	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 38	GGGGCGTGCACTTGATGTGGGACAGGCAGA	223	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 39	GCTTCATCTTGGGCGTGCACTTGATGTGGG	232	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링	GTGAACTTCTTCATCTTGGGCGTGCACTTG	239	Cas13a 스페이서가 5' 말단에서 2 개의	49C

[0937]

가이드 40			뉴클레오티드에 의해 절단됨에 주의	
Gluc 타일링 가이드 41	GGGATGAACTTCTTCA TCTTGGGCGTGCAC	242	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 42	GTGGGATGAACTTCTT CATCTTGGGCGTGC	244	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 43	GGGCAGCGTCCTGGGA TGAACTTCTTCATC	254	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 44	GGGTGTGGCAGCGTCC TGGGATGAACTTCT	259	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 45	GTTCGTAGGTGTGGCA GCGTCCTGGGATGA	265	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 46	GCGCCTTCGTAGGTGT GGCAGCGTCCTGGG	269	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 47	GTCTTTGTCGCCTTCGT AGGTGTGGCAGCG	276	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 48	GCTTTGTCGCCTTCGTA GGTGTGGCAGCGT	275	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 49	GTGCCGCCCTGTGCGG ACTCTTTGTCGCCT	293	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드	GTATGCCGCCCTGTGC GGACTCTTTGTCGC	295	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해	49C

[0938]

50			절단됨에 주의	
Gluc 타일링 가이드 51	GCCTCGCCTATGCCGC CCTGTGCGGACTCT	302	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 52	GGATCGCCTCGCCTAT GCCGCCCTGTGCGG	307	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 53	GATGTCGACGATCGCC TCGCCTATGCCGCC	315	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 54	GCAGGAATGTCGACGA TCGCCTCGCCTATG	320	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 55	GAATCTCAGGAATGTC GACGATCGCCTCGC	325	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 56	GCCCAGGAATCTCAGG AATGTCGACGATCG	331	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 57	GCCTTGAACCCAGGAA TCTCAGGAATGTCG	338	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 58	GCCAAGTCCTTGAACC CAGGAATCTCAGGA	344	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 59	GTGGGCTCCAAGTCCT TGAACCCAGGAATC	350	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 60	GCCATGGGCTCCAAGT CCTTGAACCCAGGA	353	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C

[0939]

Gluc 타일링 가이드 61	GGAACTGCTCCATGGG CTCCAAGTCCTTGA	361	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 62	GTGCGATGAACTGCTC CATGGGCTCCAAGT	367	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 63	GGACCTGTGCGATGAA CTGCTCCATGGGCT	373	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 64	GACAGATCGACCTGTG CGATGAACTGCTCC	380	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 65	GACACACAGATCGACC TGTGCGATGAACTG	384	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 66	GTGCAGTCCACACACA GATCGACCTGTGCG	392	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 67	GCCAGTTGTGCAGTCC ACACACAGATCGAC	399	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 68	GGGCAGCCAGTTGTGC AGTCCACACACAGA	404	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 69	GTTTGAGGCAGCCAGT TGTGCAGTCCACAC	409	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 70	GAAGCCCTTTGAGGCA GCCAGTTGTGCAGT	415	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc	GCACGTTGGCAAGCCC	424	Cas13a 스페이서가 5'	49C

[0940]

타일링 가이드 71	TTTGAGGCAGCCAG		말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	
Gluc 타일링 가이드 72	GACTGCACGTTGGCAA GCCCTTTGAGGCAG	428	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 73	GGGTCAGAACACTGCA CGTTGGCAAGCCCT	437	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 74	GCAGGTCAGAACACTG CACGTTGGCAAGCC	439	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 75	GAGCAGGTCAGAACAC TGCACGTTGGCAAG	441	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 76	GGCCACTTCTTGAGCA GGTCAGAACACTGC	452	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 77	GCGGCAGCCACTTCTT GAGCAGGTCAGAAC	457	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 78	GTGCGGCAGCCACTTC TTGAGCAGGTCAGA	459	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 79	GAGCGTTGCGGCAGCC ACTTCTTGAGCAGG	464	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 80	GAAAGGTCGCACAGCG TTGCGGCAGCCACT	475	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링	GCTGGCAAAGGTCGCA CAGCGTTGCGGCAG	480	Cas13a 스페이서가 5' 말단에서 2 개의	49C

[0941]

가이드 81			뉴클레오티드에 의해 절단됨에 주의	
Gluc 타일링 가이드 82	GGGCAAAGGTCGCACA GCGTTGCGGCAGCC	478	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 83	GTGGATCTTGCTGGCA AAGGTCGCACAGCG	489	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 84	GCACCTGGCCCTGGAT CTTGCTGGCAAAGG	499	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 85	GTGGCCCTGGATCTTG CTGGCAAAGGTCGC	495	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 86	GTGATCTTGTCCACCT GGCCCTGGATCTTG	509	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 87	GCCCCTTGATCTTGTCC ACCTGGCCCTGGA	514	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 88	GCCCTTGATCTTGTCC ACCTGGCCCTGGAT	513	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 89	GCCTTGATCTTGTCCA CCTGGCCCTGGATC	512	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 90	GGCAAAGGTCGCACAG CGTTGCGGCAGCCA	477	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드	GCAAAGGTCGCACAGC GTTGCGGCAGCCAC	476	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해	49C

[0942]

91			절단됨에 주의	
Gluc 타일링 가이드 92	GAAGGTCGCACAGCGT TGCGGCAGCCACTT	474	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
Gluc 타일링 가이드 93	GAGGTCGCACAGCGTT GCGGCAGCCACTTC	473	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
비-표적화 가이드 1	GGTAATGCCTGGCTTG TCGACGCATAGTCTG	N/A	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
비-표적화 가이드 2	GGGAACCTTGGCCGTT ATAAAGTCTGACCAG	N/A	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C
비-표적화 가이드 3	GGAGGGTGAGAATTTA GAACCAAGATTGTTG	N/A	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49C

[0943]

표 13 - Cluc 녹다운을 위해 사용된 가이드 서열

명칭	스페이서 서열	위치	주석	첫번째 도면
Cluc 타일링 가이드 1	GAGTCCTGGCAATGA ACAGTGGCGCAGTAG	32	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 2	GGGTGCCACAGCTGC TATCAATACATTCTC	118	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 3	GTTACATACTGACAC ATTCGGCAACATGTT	197	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc	GTATGTACCAGGTTCC	276	Cas13a 스페이서가 5'	49D

[0944]

타일링 가이드 4	TGGAACTGGAATCT		말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	
Cluc 타일링 가이드 5	GCCTTGGTTCCATCCA GGTTCTCCAGGGTG	350	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 6	GCAGTGATGGGATTC TCAGTAGCTTGAGCG	431	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 7	GAGCCTGGCATCTCA ACAACAGCGATGGTG	512	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 8	GTGTCTGGGGCGATTC TTACAGATCTTCCT	593	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 9	GCTGGATCTGAAGTG AAGTCTGTATCTTCC	671	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 10	GGCAACGTCATCAGG ATTTCCATAGAGTGG	747	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 11	GAGGCGCAGGAGATG GTGTAGTAGTAGAAG	830	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 13	GAGGGACCCTGGAAT TGGTATCTTGCTTTG	986	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 14	GGTAAGAGTCAACAT TCCTGTGTGAAACCT	1066	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링	GACCAGAATCTGTTTT CCATCAACAATGAG	1143	Cas13a 스페이서가 5' 말단에서 2 개의	49D

[0945]

가이드 15			뉴클레오티드에 의해 절단됨에 주의	
Cluc 타일링 가이드 16	GATGGCTGTAGTCAG TATGTCACCATCTTG	1227	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 17	GTACCATCGAATGGA TCTCTAATATGTACG	1304	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 18	GAGATCACAGGCTCC TTCAGCATCAAAAGA	1380	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 19	GCTTTGACCGGCGAA GAGACTATTGCAGAG	1461	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 20	GCCCCTCAGGCAATA CTCGTACATGCATCG	1539	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
Cluc 타일링 가이드 21	GCTGGTACTTCTAGGG TGTCTCCATGCTTT	1619	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
비-표적화 가이드 1	GGTAATGCCTGGCTTG TCGACGCATAGTCTG	N/A	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
비-표적화 가이드 2	GGGAACCTTGGCCGT TATAAAGTCTGACCA G	N/A	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D
비-표적화 가이드 3	GGAGGGTGAGAATTT AGAACCAAGATTGTT G	N/A	Cas13a 스페이서가 5' 말단에서 2 개의 뉴클레오티드에 의해 절단됨에 주의	49D

[0946]

표 14 - 포유동물 세포에서 RNA 편집을 위해 본 연구에서 사용되는 가이드 서열. 미스매치된 염기 플립은 대문자로 표시됨

명칭	스페이서 서열	주석	첫번째 도면
타일링 30 nt 30 미스매치 거리	gCatcctgcggcctctactctgcattcaatt	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 28 미스매치 거리	gacCatcctgcggcctctactctgcattcaa	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 26 미스매치 거리	gaaacCatcctgcggcctctactctgcattc	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 24 미스매치 거리	gctaaacCatcctgcggcctctactctgcat	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 22 미스매치 거리	gttctaaacCatcctgcggcctctactctgc	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 20 미스매치 거리	gtgttctaaacCatcctgcggcctctactct	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 18 미스매치 거리	gaatgttctaaacCatcctgcggcctctact	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 16 미스매치 거리	gagaatgttctaaacCatcctgcggcctcta	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 14 미스매치 거리	gatagaatgttctaaacCatcctgcggcctc	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 12 미스매치 거리	gccatagaatgttctaaacCatcctgcggcc	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 10 미스매치 거리	gttccatagaatgttctaaacCatcctgcgg	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 8 미스매치 거리	gctttccatagaatgttctaaacCatcctgc	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 6 미스매치 거리	gctctttccatagaatgttctaaacCatcct	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 4 미스매치 거리	gatctctttccatagaatgttctaaacCatc	U6 발현을 위해 5' G 를 가짐	50C
타일링 30 nt 2 미스매치 거리	ggaatctctttccatagaatgttctaaacCa	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 50 미스매치 거리	gCatcctgcggcctctactctgcattcaatt acatactgacacattcggca	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 48 미스매치 거리	gacCatcctgcggcctctactctgcattca attacatactgacacattcgg	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 46	gaaacCatcctgcggcctctactctgcatt caattacatactgacacattc	U6 발현을 위해 5' G 를	50C

[0947]

미스매치 거리		가짐	
타일링 50 nt 44 미스매치 거리	gctaaacCatcctgcggcctctactctgcattcaattacatactgacacat	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 42 미스매치 거리	gttctaaacCatcctgcggcctctactctgcattcaattacatactgacac	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 40 미스매치 거리	gtgttctaaacCatcctgcggcctctactctgcattcaattacatactgac	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 38 미스매치 거리	gaatgttctaaacCatcctgcggcctctactctgcattcaattacatactg	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 36 미스매치 거리	gagaatgttctaaacCatcctgcggcctctactctgcattcaattacatac	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 34 미스매치 거리	gatagaatgttctaaacCatcctgcggcctctactctgcattcaattacat	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 32 미스매치 거리	gccatagaatgttctaaacCatcctgcggcctctactctgcattcaattac	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 30 미스매치 거리	gttccatagaatgttctaaacCatcctgcggcctctactctgcattcaatt	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 28 미스매치 거리	gctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaa	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 26 미스매치 거리	gctctttccatagaatgttctaaacCatcctgcggcctctactctgcattc	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 24 미스매치 거리	gatctctttccatagaatgttctaaacCatcctgcggcctctactctgcat	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 22 미스매치 거리	ggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgc	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 20 미스매치 거리	gtggaatctctttccatagaatgttctaaacCatcctgcggcctctactct	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 18 미스매치 거리	gactggaatctctttccatagaatgttctaaacCatcctgcggcctctact	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 16 미스매치 거리	ggaactggaatctctttccatagaatgttctaaacCatcctgcggcctcta	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 14 미스매치 거리	gtggaactggaatctctttccatagaatgttctaaacCatcctgcggcctc	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 12 미스매치 거리	gcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcc	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 10 미스매치 거리	gttcctggaactggaatctctttccatagaatgttctaaacCatcctgcgg	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 8 미스매치 거리	gggttcctggaactggaatctctttccatagaatgttctaaacCatcctgc	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 6	gcaggttcctggaactggaatctctttccatagaatgttctaaacCatcct	U6 발현을 위해 5' G 를	50C

[0948]

미스매치 거리		가짐	
타일링 50 nt 4 미스매치 거리	gaccaggttcctggaactggaatctctttcc atagaatgttctaaacCatc	U6 발현을 위해 5' G 를 가짐	50C
타일링 50 nt 2 미스매치 거리	ggtaccaggttcctggaactggaatctcttt ccatagaatgttctaaacCa	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 70 미스매치 거리	gCatcctgcggcctctactctgcattcaatt acatactgacacattcggcaacatgtttttc ctggtttat	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 68 미스매치 거리	gacCatcctgcggcctctactctgcattca attacatactgacacattcggcaacatgttttt tcctggttt	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 66 미스매치 거리	gaaacCatcctgcggcctctactctgcatt caattacatactgacacattcggcaacatgt ttttcctggt	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 64 미스매치 거리	gctaaacCatcctgcggcctctactctgca ttcaattacatactgacacattcggcaacat gtttttcctg	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 62 미스매치 거리	gttctaaacCatcctgcggcctctactctg cattcaattacatactgacacattcggcaac atgttttcc	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 60 미스매치 거리	gtgttctaaacCatcctgcggcctctactct gcattcaattacatactgacacattcggcaa catgttttt	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 58 미스매치 거리	gaatgttctaaacCatcctgcggcctctac tctgcattcaattacatactgacacattcgg caacatgttt	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 56 미스매치 거리	gagaatgttctaaacCatcctgcggcctct actctgcattcaattacatactgacacattc ggcaacatgt	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 54 미스매치 거리	gatagaatgttctaaacCatcctgcggcct ctactctgcattcaattacatactgacacatt cggcaacat	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 52 미스매치 거리	gccatagaatgttctaaacCatcctgcgg cctctactctgcattcaattacatactgacac attcggcaac	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 50 미스매치 거리	gttccatagaatgttctaaacCatcctgcg gcctctactctgcattcaattacatactgac acattcggca	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 48 미스매치 거리	gctttccatagaatgttctaaacCatcctgc ggcctctactctgcattcaattacatactga cacattcgg	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 46 미스매치 거리	gctctttccatagaatgttctaaacCatcct gcggcctctactctgcattcaattacatact gacacattc	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 44	gatctctttccatagaatgttctaaacCatc	U6 발현을 위해 5' G 를	50C

[0949]

미스매치 거리	ctgcggcctctactctgcattcaattacata ctgacacat	가짐	
타일링 70 nt 42 미스매치 거리	ggaatctctttccatagaatgttctaaacCa tcctgcggcctctactctgcattcaattacat actgacac	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 40 미스매치 거리	gtggaatctctttccatagaatgttctaaac Catcctgcggcctctactctgcattcaatta catactgac	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 38 미스매치 거리	gactggaatctctttccatagaatgttctaaa cCatcctgcggcctctactctgcattcaatt acatactg	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 36 미스매치 거리	ggaactggaatctctttccatagaatgttct aaacCatcctgcggcctctactctgcattc aattacatac	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 34 미스매치 거리	gtggaactggaatctctttccatagaatgtt ctaaacCatcctgcggcctctactctgcat tcaattacat	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 32 미스매치 거리	gcctggaactggaatctctttccatagaatg ttctaaacCatcctgcggcctctactctgc attcaattac	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 30 미스매치 거리	gttcctggaactggaatctctttccatagaa tgttctaaacCatcctgcggcctctactctg cattcaatt	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 28 미스매치 거리	gggttcctggaactggaatctctttccatag aatgttctaaacCatcctgcggcctctact ctgcattcaa	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 26 미스매치 거리	gcaggttcctggaactggaatctctttccat agaatgttctaaacCatcctgcggcctcta ctctgcattc	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 24 미스매치 거리	gaccaggttcctggaactggaatctctttcc atagaatgttctaaacCatcctgcggcctc tactctgcat	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 22 미스매치 거리	ggtaccaggttcctggaactggaatctcttt ccatagaatgttctaaacCatcctgcggcc tctactctgc	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 20 미스매치 거리	gatgtaccaggttcctggaactggaatctc tttccatagaatgttctaaacCatcctgcgg cctctactct	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 18 미스매치 거리	ggtatgtaccaggttcctggaactggaatc tctttccatagaatgttctaaacCatcctgc ggcctctact	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 16 미스매치 거리	gacgtatgtaccaggttcctggaactggaa tctctttccatagaatgttctaaacCatcctg cggcctcta	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 14 미스매치 거리	gacacgtatgtaccaggttcctggaactgg aatctctttccatagaatgttctaaacCatc	U6 발현을 위해 5' G 를 가짐	50C

[0950]

	ctgcggcctc		
타일링 70 nt 12 미스매치 거리	gcaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcc	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 10 미스매치 거리	gcccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcgg	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 8 미스매치 거리	ggacccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgc	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 6 미스매치 거리	gttgacccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcct	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 4 미스매치 거리	gccttgacccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatc	U6 발현을 위해 5' G 를 가짐	50C
타일링 70 nt 2 미스매치 거리	gttccttgacccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCa	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 84 미스매치 거리	gCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctggtttatttttcacacagtcca	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 82 미스매치 거리	gacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctggtttatttttcacacagtc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 80 미스매치 거리	gaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcctggtttatttttcacacag	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 78 미스매치 거리	gctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcctggtttatttttcacac	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 76 미스매치 거리	gttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcctggtttatttttcac	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 74 미스매치 거리	gtgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcctggtttatttttc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 72 미스매치 거리	gaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcctggtttattt	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 70 미스매치 거리	gagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcctggtttat	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 68 미스매치 거리	gatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcctggttt	U6 발현을 위해 5' G 를 가짐	50C

[0951]

타일링 84 nt 66 미스매치 거리	gccatagaatgttctaaacCatcctgcgg cctctactctgcattcaattacatactgacac attcggcaacatgtttttcctggt	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 64 미스매치 거리	gttccatagaatgttctaaacCatcctgcg gcctctactctgcattcaattacatactgac acattcggcaacatgtttttcctg	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 62 미스매치 거리	gctttccatagaatgttctaaacCatcctgc ggcctctactctgcattcaattacatactga cacattcggcaacatgtttttcc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 60 미스매치 거리	gctctttccatagaatgttctaaacCatcct gcggcctctactctgcattcaattacatact gacacattcggcaacatgttttt	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 58 미스매치 거리	gatctctttccatagaatgttctaaacCatc ctgcggcctctactctgcattcaattacata ctgacacattcggcaacatgttt	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 56 미스매치 거리	ggaatctctttccatagaatgttctaaacCa tcctgcggcctctactctgcattcaattacat actgacacattcggcaacatgt	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 54 미스매치 거리	gtggaatctctttccatagaatgttctaaac Catcctgcggcctctactctgcattcaatta catactgacacattcggcaacat	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 52 미스매치 거리	gactggaatctctttccatagaatgttctaaa cCatcctgcggcctctactctgcattcaatt acatactgacacattcggcaac	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 50 미스매치 거리	ggaactggaatctctttccatagaatgttct aaacCatcctgcggcctctactctgcattc aattacatactgacacattcggca	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 48 미스매치 거리	gtggaactggaatctctttccatagaatgtt ctaaacCatcctgcggcctctactctgcat tcaattacatactgacacattcgg	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 46 미스매치 거리	gcctggaactggaatctctttccatagaatg ttctaaacCatcctgcggcctctactctgc attcaattacatactgacacattc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 44 미스매치 거리	gttcctggaactggaatctctttccatagaa tgttctaaacCatcctgcggcctctactctg cattcaattacatactgacacat	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 42 미스매치 거리	gggttcctggaactggaatctctttccatag aatgttctaaacCatcctgcggcctctact ctgcattcaattacatactgacac	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 40 미스매치 거리	gcaggttcctggaactggaatctctttccat agaatgttctaaacCatcctgcggcctcta ctctgcattcaattacatactgac	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 38 미스매치 거리	gaccaggttcctggaactggaatctctttcc atagaatgttctaaacCatcctgcggcctc tactctgcattcaattacatactg	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 36	ggtaccaggttcctggaactggaatctcttt	U6 발현을 위해 5' G 를	50C

[0952]

미스매치 거리	ccatagaatgttctaaacCatcctgcggcc tctactctgcattcaattacatac	가짐	
타일링 84 nt 34 미스매치 거리	gatgtaccaggttcctggaactggaatctc tttccatagaatgttctaaacCatcctgcgg cctctactctgcattcaattacat	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 32 미스매치 거리	ggtatgtaccaggttcctggaactggaatc tctttccatagaatgttctaaacCatcctgc ggcctctactctgcattcaattac	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 30 미스매치 거리	gacgtatgtaccaggttcctggaactggaa tctctttccatagaatgttctaaacCatcctg cggcctctactctgcattcaatt	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 28 미스매치 거리	gacacgtatgtaccaggttcctggaactgg aatctctttccatagaatgttctaaacCatc ctgcggcctctactctgcattcaa	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 26 미스매치 거리	gcaacacgtatgtaccaggttcctggaact ggaatctctttccatagaatgttctaaacCa tcctgcggcctctactctgcattc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 24 미스매치 거리	gcccaacacgtatgtaccaggttcctgga actggaatctctttccatagaatgttctaaac Catcctgcggcctctactctgcat	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 22 미스매치 거리	ggacccaacacgtatgtaccaggttcctg gaactggaatctctttccatagaatgttctaa acCatcctgcggcctctactctgc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 20 미스매치 거리	gttgacccaacacgtatgtaccaggttcct ggaactggaatctctttccatagaatgttct aaacCatcctgcggcctctactct	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 18 미스매치 거리	gccttgacccaacacgtatgtaccaggttc ctggaactggaatctctttccatagaatgtt ctaaacCatcctgcggcctctact	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 16 미스매치 거리	gttccttgacccaacacgtatgtaccaggtt cctggaactggaatctctttccatagaatgt tctaaacCatcctgcggcctcta	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 14 미스매치 거리	gggttccttgacccaacacgtatgtaccag gttcctggaactggaatctctttccatagaa tgttctaaacCatcctgcggcctc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 12 미스매치 거리	gttggttccttgacccaacacgtatgtacca ggttcctggaactggaatctctttccataga atgttctaaacCatcctgcggcc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 10 미스매치 거리	gccttggttccttgacccaacacgtatgtac caggttcctggaactggaatctctttccata gaatgttctaaacCatcctgcgg	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 8 미스매치 거리	ggcccttggttccttgacccaacacgtatgt accaggttcctggaactggaatctctttcca tagaatgttctaaacCatcctgc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 6 미스매치 거리	gccgcccttggttccttgacccaacacgta tgtaccaggttcctggaactggaatctctttt	U6 발현을 위해 5' G 를 가짐	50C

[0953]

	ccatagaatgttctaaacCatcct		
타일링 84 nt 4 미스매치 거리	gcgccgccttggttccttgacccaacac gtatgtaccaggttcctggaactggaatct ctttccatagaatgttctaaacCatc	U6 발현을 위해 5' G 를 가짐	50C
타일링 84 nt 2 미스매치 거리	ggtcgccgccttggttccttgacccaaca cgtatgtaccaggttcctggaactggaatc tctttccatagaatgttctaaacCa	U6 발현을 위해 5' G 를 가짐	50C
ADAR 비-표적화 가이드	GTAATGCCTGGCTTGTCG ACGCATAGTCTG	U6 발현을 위해 5' G 를 가짐	50C
TAG 모티프에 대한 PFS 결합 스크린 가이드	gaaaacgcaggttcctcCagtttcgggag cagcgcacgtctccctgtagtc	U6 발현을 위해 5' G 를 가짐	51B
AAC 모티프에 대한 PFS 결합 스크린 가이드	gacgcaggttcctctagCttcgggagcag cgcacgtctccctgtagtcaag	U6 발현을 위해 5' G 를 가짐	51B
PFS 결합 스크린 비-표적화	GTAATGCCTGGCTTGTCG ACGCATAGTCTG	U6 발현을 위해 5' G 를 가짐	51B
모티프 선호 표적화 가이드	gatagaatgttctaaacCatcctgcggcct ctactctgcattcaattacat	U6 발현을 위해 5' G 를 가짐	51C
모티프 선호 비-표적화 가이드	GTAATGCCTGGCTTGTCG ACGCATAGTCTG	U6 발현을 위해 5' G 를 가짐	51C
PPIB 타일링 가이드 50 미스매치 거리	gCaaggccacaaaattatccactgttttg gaacagtctttccgaagagac	U6 발현을 위해 5' G 를 가짐	57D
PPIB 타일링 가이드 42 미스매치 거리	gcctgtagcCaaggccacaaaattatcca ctgtttttggaacagtctttcc	U6 발현을 위해 5' G 를 가짐	57D
PPIB 타일링 가이드 34 미스매치 거리	gctttctctcctgtagcCaaggccacaaaa ttatccactgtttttggaaca	U6 발현을 위해 5' G 를 가짐	57D
PPIB 타일링 가이드 26 미스매치 거리	ggccaaatcctttctctcctgtagcCaagg ccacaaaattatccactgttt	U6 발현을 위해 5' G 를 가짐	57D
PPIB 타일링 가이드 18 미스매치 거리	gtttttgtagccaaatcctttctctcctgtagc Caaggccacaaaattatc	U6 발현을 위해 5' G 를 가짐	57D
PPIB 타일링 가이드 10 미스매치 거리	gatttgctgtttttgtagccaaatcctttctct cctgtagcCaaggccaca	U6 발현을 위해 5' G 를 가짐	57D
PPIB 타일링 가이드 2	gacgatggaatttgctgtttttgtagccaaat cctttctctcctgtagcCa	U6 발현을 위해 5' G 를 가짐	57D

[0954]

미스매치 거리			
표적화 가이드, 반대 염기 G	gatagaatgttctaaacGatcctgcggcct ctactctgcattcaattacat	U6 발현을 위해 5' G 를 가짐	57D
표적화 가이드, 반대 염기 A	gatagaatgttctaaacAatcctgcggcct ctactctgcattcaattacat	U6 발현을 위해 5' G 를 가짐	57D
표적화 가이드, 반대 염기 C	gatagaatgttctaaacTatcctgcggcct ctactctgcattcaattacat	U6 발현을 위해 5' G 를 가짐	57D
AVPR2 가이드 37 미스매치 거리	ggtcccacgcggccCacagctgcacca ggaagaagggtgcccagcacagca	U6 발현을 위해 5' G 를 가짐	52A
AVPR2 가이드 35 미스매치 거리	ggggtcccacgcggccCacagctgcac caggaagaagggtgcccagcacag	U6 발현을 위해 5' G 를 가짐	52A
AVPR2 가이드 33 미스매치 거리	gccgggtcccacgcggccCacagctgc accaggaagaagggtgcccagcac	U6 발현을 위해 5' G 를 가짐	52A
FANCC 가이드 37 미스매치 거리	gggtgatgacatccCaggcgatcgtgtg gcctccaggagcccagagcagga	U6 발현을 위해 5' G 를 가짐	52B
FANCC 가이드 35 미스매치 거리	gagggtgatgacatccCaggcgatcgtg tggcctccaggagcccagagcag	U6 발현을 위해 5' G 를 가짐	52B
FANCC 가이드 32 미스매치 거리	gatcagggtgatgacatccCaggcgatc gtgtggcctccaggagcccagag	U6 발현을 위해 5' G 를 가짐	52B
합성 질환 유전자 표적 IL2RG	ggtggctccattcactcCaatgctgagca cttccacagagtgggttaaagc	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 F8	gtttctaatatattttgCcagactgatggact attctcaattaataatgat	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 LDLR	gagatgttgctgtggatCcagtccacagc cagcccgtcgggggcctggatg	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 CBS	gcaggccggcccagctgCcaggtgcac ctgctcggagcatcgggccggatc	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 HBB	gcaaagaacctctgggtCcaagggtaga ccaccagcagcctgcccagggcc	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적	gaagagaaacttagtttCcagggctttggt agagggcaaaggttgatagca	U6 발현을 위해 5' G 를 가짐	52E

[0955]

ALDOB			
합성 질환 유전자 표적 DMD	gtcagcctagtgcagagCcactggtagtt ggtggttagagtttcaagttcc	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 SMAD4	ggctcattgtgaacaggCcagtaatgtcc gggatggggcggcataggcggg	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 BRCA2	gtagctaaagaacttgaCcaagacatatc aggatccacctcagctcctaga	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 GRIN2A	ggggcattgttctgtgcCcagtcctgctgg tagacctgctccccggtggct	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 SCN9A	gagaagtcgttcatgtgCcaccgtggga gcgtacagtcatcattgatcttg	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 TARDBP	gggattaatgctgaacgCaccaaagttca tcccaccacccatattactacc	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 CFTR	gctccaaaggctttcctCcactgttgcaaa gttattgaatcccaagacaca	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 UBE3A	gatgaatgaacgatttcCcagaactcccta atcagaacagagtccctggta	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 SMPD1	ggagcctctgccggagcCcagagaacc cgagagtcagacagagccagcgcc	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 USH2A	ggcttccgtggagacacCcaatcaatttg aagagatcttgaagtgatgcca	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 MEN1	gtgggactgccctcctcCcatttgcagatg ccgtcgtagaatcgcagcagg	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 C8orf37	gcttcttcaatagttctCcagctacactggc aggcatatgcccgtgttcct	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 MLH1	gattccttttcttcgtcCcaattcacctcagt ggctagtcgaagaatgaag	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적	gcagcttcagcaccttcCagtcagactcct gcttcaagcactgcagcagga	U6 발현을 위해 5' G 를 가짐	52E

[0956]

TSC2			
합성 질환 유전자 표적 NF1	gccatttgcttgcagtgCcactccagaggattccggattgccataaatact	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 MSH6	gttcaatagttttggtcCagtatcgtttacagcccttcttggtagatttca	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 SMN1	ggcaaccgtcttctgacCaaatggcagaacatttgtccccaactttccact	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 SH3TC2	gcgactttccaatgaacCactgaagcccaggtatgacaaagccgatgatct	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 DNAH5	gtttacactcatgcttcCacagctttaacagatcatttggttccttgatga	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 MECP2	gcttaagcttccgtgtcCagccttcaggcagggtggggtcatcatacatgg	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 ADGRV1	ggacagctgggctgatcCatgatgtcatccagaaacactggggaccctcag	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 AHI1	gtctcatctcaactttcCatatccgtatcatggaatcatagcatcctgtaa	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 PRKN	gcatgcagacgcggttcCactcgcagccacagttccagcaccactcgagcc	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 COL3A1	gttggttagggtcaaccCagtattctccactcttgagttcaggatggcaga	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 BRCA1	gctacactgtccaacacCcactctcgggtcaccacaggtgcctcacacatc	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 MYBPC3	gctgcactgtgtaccccCagagctccgtgttgccgacatcctggggtggct	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 APC	gagcttcctgccactccCaacaggtttcacagtaagcgcgtatctgttcca	U6 발현을 위해 5' G 를 가짐	52E
합성 질환 유전자 표적 BMPR2	gacggcaagagcttaccCagtcacttgtgtggagacttaaatacttgcata	U6 발현을 위해 5' G 를 가짐	52E

[0957]

KRAS 타일링 가이드 50 미스매치 거리	gCaaggccacaaaattatccactgtttttg gaacagtctttccgaagagac	U6 발현을 위해 5' G 를 가짐	53A
KRAS 타일링 가이드 42 미스매치 거리	gcctgtagcCaaggccacaaaattatcca ctgtttttggaacagtctttcc	U6 발현을 위해 5' G 를 가짐	53A
KRAS 타일링 가이드 34 미스매치 거리	gctttctctcctgtagcCaaggccacaaaa ttatccactgtttttggaaca	U6 발현을 위해 5' G 를 가짐	53A
KRAS 타일링 가이드 26 미스매치 거리	ggccaaatcctttctctcctgtagcCaagg ccacaaaattatccactgttt	U6 발현을 위해 5' G 를 가짐	53A
KRAS 타일링 가이드 18 미스매치 거리	gtttttgtagccaaatcctttctctcctgtagc Caaggccacaaaattatc	U6 발현을 위해 5' G 를 가짐	53A
KRAS 타일링 가이드 10 미스매치 거리	gatttgctgtttttgtagccaaatcctttctct cctgtagcCaaggccaca	U6 발현을 위해 5' G 를 가짐	53A
KRAS 타일링 가이드 2 미스매치 거리	gacgatggaatttgctgtttttgtagccaaat cctttctctcctgtagcCa	U6 발현을 위해 5' G 를 가짐	53A
KRAS 타일링 비-표적화 가이드	GTAATGCCTGGCTTGTCG ACGCATAGTCTG	U6 발현을 위해 5' G 를 가짐	53A
전사체 특이성에 대한 루시페라제 W85X 표적화 가이드	gatagaatgttctaaacCatcctgcggcct ctactctgcattcaattacat	U6 발현을 위해 5' G 를 가짐	53B
전사체 특이성에 대한 비-표적화 가이드	GCAGGGTTTTCCCAGTCA CGACGTTGTAAAGTTG	U6 발현을 위해 5' G 를 가짐	53C
내인성 KRAS 가이드 2	gtcaaggcactcttgccCacgccaccag ctccaactaccacaagtttatat	U6 발현을 위해 5' G 를 가짐	54F
내인성 PPIB 가이드 1	gcaaagatcacccggccCacatcttcatc tccaattcgtaggtcaaaatac	U6 발현을 위해 5' G 를 가짐	54G
내인성 KRAS 가이드 1	GcgccaccagctccaacCaccacaagtt tatattcagtcattttcagcagg	U6 발현을 위해 5' G 를 가짐	54F
내인성 KRAS 가이드 3	GtttctccatcaattacCacttgcttcctgta ggaatcctctattGTtgga	U6 발현을 위해 5' G 를 가짐	54F
내인성 PPIB 가이드 2	GctttctctcctgtagcCaaggccacaaa attatccactgtttttggaaca	U6 발현을 위해 5' G 를 가짐	54G

[0958]

내인성 비-표적화 가이드	GTAATGCCTGGCTTGTCG ACGCATAGTCTG	U6 발현을 위해 5' G 를 가짐	54F
BoxB Cluc 가이드	tctttccataGGCCCTGAAAAA GGGCCtgttctaaacCatcctgcggc ctctactcGGCCCTGAAAAAG GGCCattcaattac	U6 발현을 위해 5' G 를 가짐	62B
BoxB 비-표적화 가이드	cagctggcgaGGCCCTGAAAA AGGGCCggggatgtgcCgcaaggc gattaagttggGGCCCTGAAAA AGGGCCacgccagggt	U6 발현을 위해 5' G 를 가짐	62B
Stafforst 전체 길이 ADAR2 가이드 1	GTGGAATAGTATAACAAT ATGCTAAATGTTGTTATA GTATCCCACtctaaaCCAtcctg cgGGGCCCTCTTCAGGGCC C	U6 발현을 위해 5' G 를 가짐	62C
Stafforst 전체 길이 ADAR2 비-표적화 가이드	GTGGAATAGTATAACAAT ATGCTAAATGTTGTTATA GTATCCCACaccctggcgttaccc aGGGCCCTCTTCAGGGCCC	U6 발현을 위해 5' G 를 가짐	62C

[0959]

[0960] **참조문헌**

[0961] 1. P. D. Hsu, E. S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262-1278 (2014).

[0962] 2. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell 168, 20-36 (2017).

[0963] 3. L. Cong et al., Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013).

[0964] 4. P. Mali et al., RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013).

[0965] 5. B. Zetsche et al., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759-771 (2015).

[0966] 6. H. Kim, J. S. Kim, A guide to genome engineering with programmable nucleases. Nat Rev Genet 15, 321-334 (2014).

[0967] 7. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-424 (2016).

[0968] 8. K. Nishida et al., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353, (2016).

[0969] 9. Y. B. Kim et al., Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol 35, 371-376 (2017).

[0970] 10. O. O. Abudayyeh et al., C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353, aaf5573 (2016).

[0971] 11. S. Shmakov et al., Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell 60, 385-397 (2015).

[0972] 12. S. Shmakov et al., Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol 15, 169-182 (2017).

[0973] 13. A. A. Smargon et al., Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Mol Cell 65, 618-630 e617 (2017).

[0974] 14. J. S. Gootenberg et al., Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356, 438-442 (2017).

[0975] 15. O. O. Abudayyeh et al., RNA targeting with CRISPR-Cas13a. Nature in press, (2017).

[0976] 16. K. Nishikura, Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 79, 321-349 (2010).

[0977] 17. B. L. Bass, H. Weintraub, An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell 55, 1089-1098 (1988).

[0978] 18. M. M. Matthews et al., Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. Nat Struct Mol Biol 23, 426-433 (2016).

[0979] 19. A. Kuttan, B. L. Bass, Mechanistic insights into editing-site specificity of ADARs. Proc Natl Acad Sci U S A 109, E3295-3304 (2012).

[0980] 20. S. K. Wong, S. Sato, D. W. Lazinski, Substrate recognition by ADAR1 and ADAR2. RNA 7, 846-858 (2001).

[0981] 21. M. Fukuda et al., Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. Sci Rep 7, 41478 (2017).

[0982] 22. M. F. Montiel-Gonzalez, I. Vallecillo-Viejo, G. A. Yudowski, J. J. Rosenthal, Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. Proc Natl Acad Sci U S A 110, 18285-18290 (2013).

[0983] 23. M. F. Montiel-Gonzalez, I. C. Vallecillo-Viejo, J. J. Rosenthal, An efficient system for selectively altering genetic information within mRNAs. Nucleic Acids Res 44, e157 (2016).

[0984] 24. J. Wettengel, P. Reautschnig, S. Geisler, P. J. Kahle, T. Stafforst, Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. Nucleic Acids Res 45, 2797-2808 (2017).

[0985] 25. Y. Wang, J. Havel, P. A. Beal, A Phenotypic Screen for Functional Mutants of Human Adenosine Deaminase Acting on RNA 1. ACS Chem Biol 10, 2512-2519 (2015).

[0986] 26. K. A. Lehmann, B. L. Bass, Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry 39, 12875-12884 (2000).

[0987] 27. Y. Zheng, C. Lorenzo, P. A. Beal, DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. Nucleic Acids Res 45, 3369-3377 (2017).

[0988] 28. K. Gao et al., A de novo loss-of-function GRIN2A mutation associated with childhood focal epilepsy and acquired epileptic aphasia. PLoS One 12, e0170818 (2017).

[0989] 29. H. M. Lanoiselee et al., APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. PLoS Med 14, e1002270 (2017).

[0990] 30. C. Ballatore, V. M. Lee, J. Q. Trojanowski, Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. Nat Rev Neurosci 8, 663-672 (2007).

[0991] 31. Y. Li et al., Carriers of rare missense variants in IFIH1 are protected from psoriasis. J Invest Dermatol 130, 2768-2772 (2010).

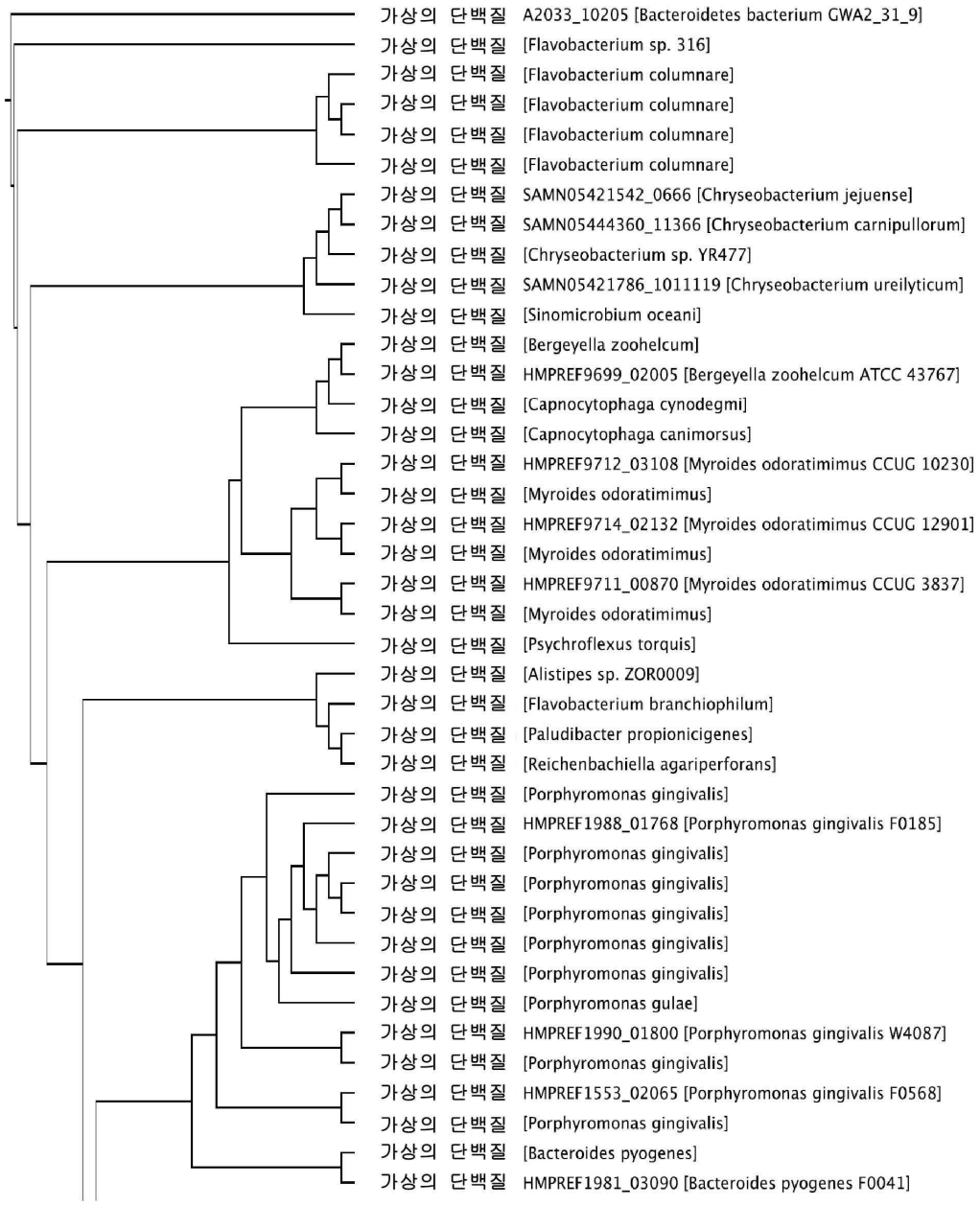
[0992] 32. R. S. Finkel et al., Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. Lancet 388, 3017-3026 (2016).

[0993] * * *

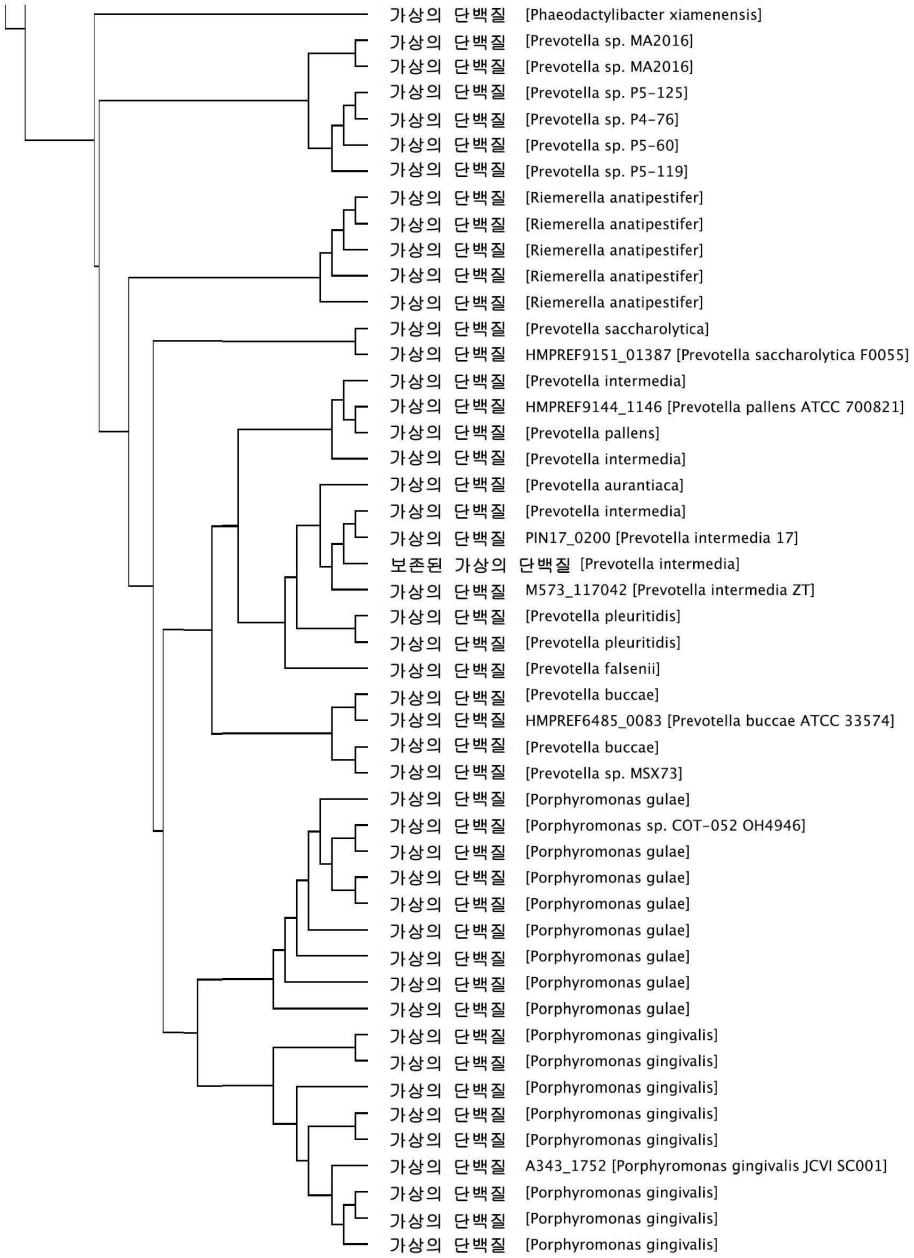
[0994] 본 발명의 바람직한 구현예가 본 명세서에 나타내어져 있고 기재되었지만, 이러한 구현예가 단지 예로서 제공된 것임이 당업자에게 자명할 것이다. 이제 본 발명을 벗어나지 않으면서 수많은 변형, 변경, 및 치환이 당업자에게 일어날 것이다. 본 명세서에 기재된 본 발명의 구현예에 대한 다양한 대안이 본 발명을 실시함에 있어서 이용될 수 있음이 이해되어야 한다. 하기 청구항이 본 발명의 범주를 규정하고, 이들 청구항의 범주 내의 방법 및 구조 및 이들의 균등물은 이에 의해 포괄되로 의도된다.

도면

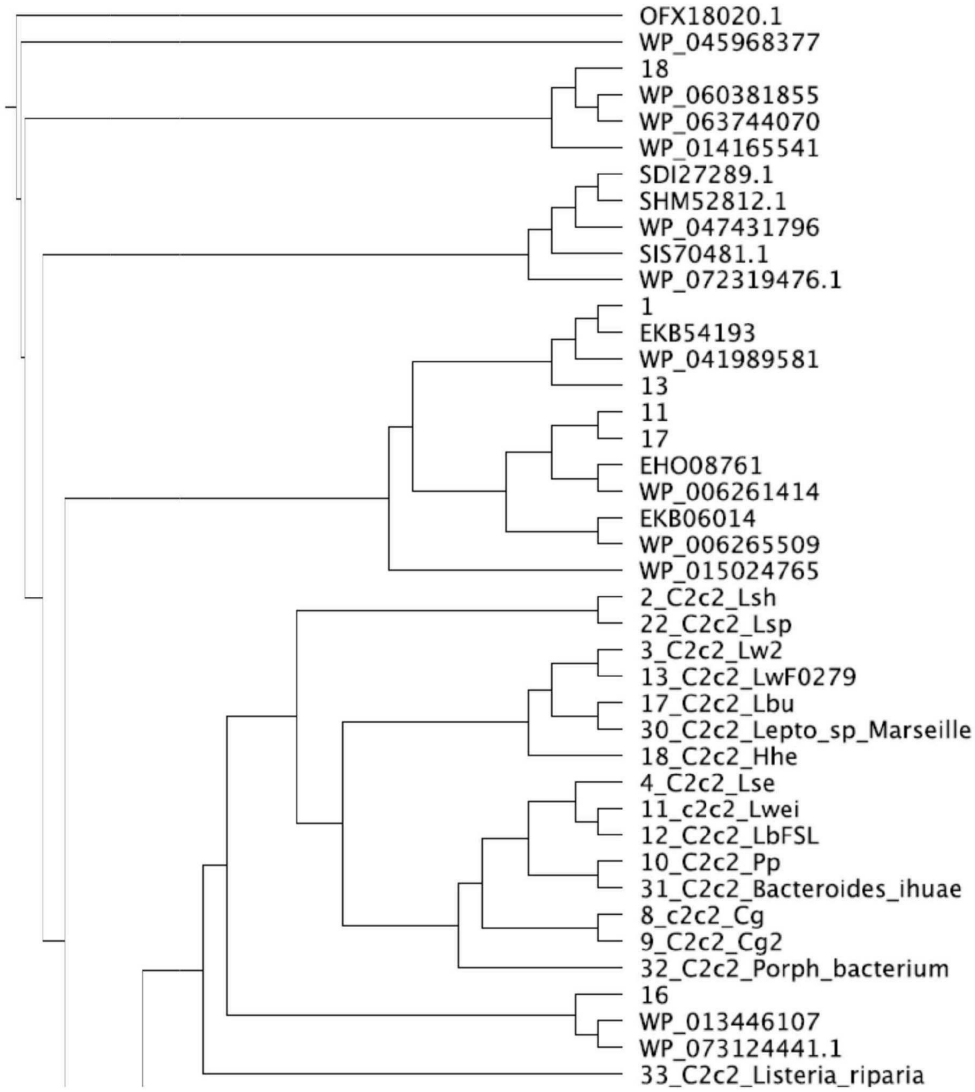
도면1a


가상의 단백질 A2033_10205 [Bacteroidetes bacterium GWA2_31_9]
가상의 단백질 [Flavobacterium sp. 316]
가상의 단백질 [Flavobacterium columnare]
가상의 단백질 [Flavobacterium columnare]
가상의 단백질 [Flavobacterium columnare]
가상의 단백질 [Flavobacterium columnare]
가상의 단백질 SAMN05421542_0666 [Chryseobacterium jejuense]
가상의 단백질 SAMN05444360_11366 [Chryseobacterium carnipullorum]
가상의 단백질 [Chryseobacterium sp. YR477]
가상의 단백질 SAMN05421786_1011119 [Chryseobacterium ureilyticum]
가상의 단백질 [Sinomicrobium oceani]
가상의 단백질 [Bergeyella zoohelcum]
가상의 단백질 HMPREF9699_02005 [Bergeyella zoohelcum ATCC 43767]
가상의 단백질 [Capnocytophaga cynodegmi]
가상의 단백질 [Capnocytophaga canimorsus]
가상의 단백질 HMPREF9712_03108 [Myroides odoratimimus CCUG 10230]
가상의 단백질 [Myroides odoratimimus]
가상의 단백질 HMPREF9714_02132 [Myroides odoratimimus CCUG 12901]
가상의 단백질 [Myroides odoratimimus]
가상의 단백질 HMPREF9711_00870 [Myroides odoratimimus CCUG 3837]
가상의 단백질 [Myroides odoratimimus]
가상의 단백질 [Psychroflexus torquis]
가상의 단백질 [Alistipes sp. ZOR0009]
가상의 단백질 [Flavobacterium branchiophilum]
가상의 단백질 [Paludibacter propionicigenes]
가상의 단백질 [Reichenbachiella agariperforans]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 HMPREF1988_01768 [Porphyromonas gingivalis F0185]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 HMPREF1990_01800 [Porphyromonas gingivalis W4087]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 HMPREF1553_02065 [Porphyromonas gingivalis F0568]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Bacteroides pyogenes]
가상의 단백질 HMPREF1981_03090 [Bacteroides pyogenes F0041]

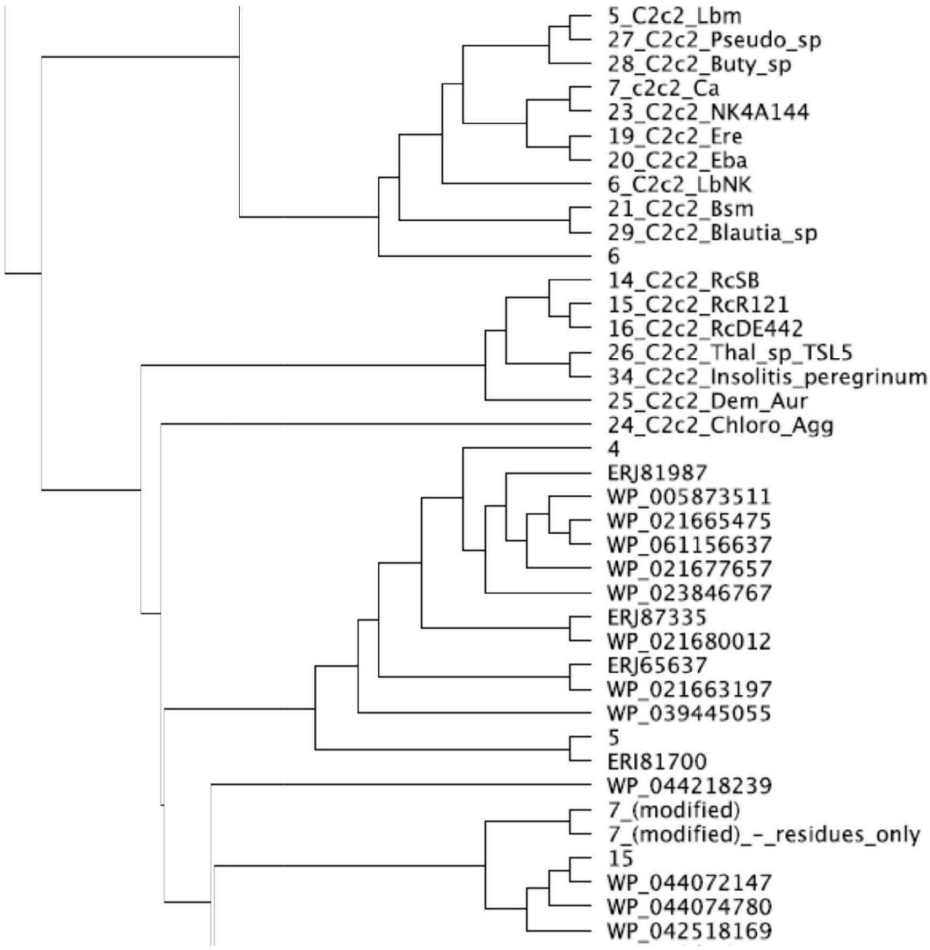
도면1b


가상의 단백질 [Phaeodactylibacter xiamenensis]
가상의 단백질 [Prevotella sp. MA2016]
가상의 단백질 [Prevotella sp. MA2016]
가상의 단백질 [Prevotella sp. P5-125]
가상의 단백질 [Prevotella sp. P4-76]
가상의 단백질 [Prevotella sp. P5-60]
가상의 단백질 [Prevotella sp. P5-119]
가상의 단백질 [Riemerella anatipestifer]
가상의 단백질 [Riemerella anatipestifer]
가상의 단백질 [Riemerella anatipestifer]
가상의 단백질 [Riemerella anatipestifer]
가상의 단백질 [Riemerella anatipestifer]
가상의 단백질 [Prevotella saccharolytica]
가상의 단백질 HMPREF9151_01387 [Prevotella saccharolytica F0055]
가상의 단백질 [Prevotella intermedia]
가상의 단백질 HMPREF9144_1146 [Prevotella pallens ATCC 700821]
가상의 단백질 [Prevotella pallens]
가상의 단백질 [Prevotella intermedia]
가상의 단백질 [Prevotella aurantiaca]
가상의 단백질 [Prevotella intermedia]
가상의 단백질 PIN17_0200 [Prevotella intermedia 17]
보존된 가상의 단백질 [Prevotella intermedia]
가상의 단백질 M573_117042 [Prevotella intermedia ZT]
가상의 단백질 [Prevotella pleuritidis]
가상의 단백질 [Prevotella pleuritidis]
가상의 단백질 [Prevotella falsenii]
가상의 단백질 [Prevotella buccae]
가상의 단백질 HMPREF6485_0083 [Prevotella buccae ATCC 33574]
가상의 단백질 [Prevotella buccae]
가상의 단백질 [Prevotella sp. MSX73]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 [Porphyromonas sp. COT-052 OH4946]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 [Porphyromonas gulae]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 A343_1752 [Porphyromonas gingivalis JCVI SC001]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]
가상의 단백질 [Porphyromonas gingivalis]

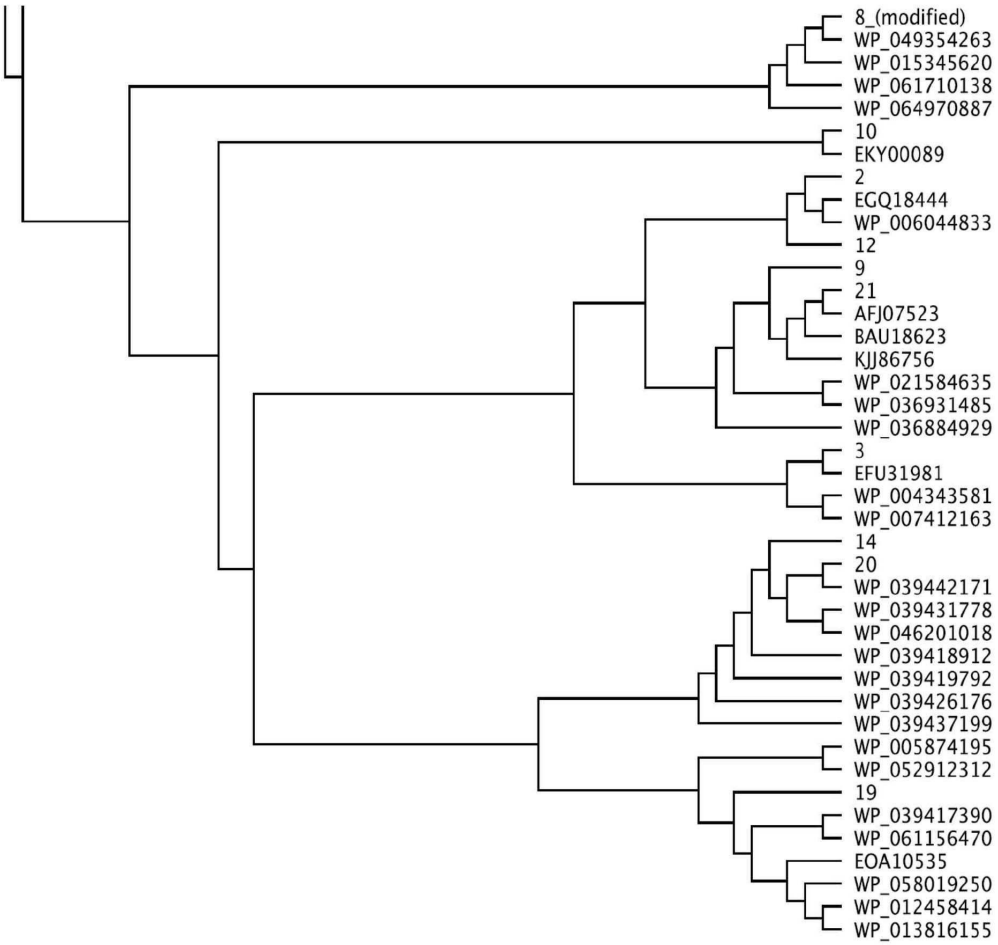
도면2a


OFX18020.1
WP_045968377
18
WP_060381855
WP_063744070
WP_014165541
SDI27289.1
SHM52812.1
WP_047431796
SIS70481.1
WP_072319476.1
1
EKB54193
WP_041989581
13
11
17
EHO08761
WP_006261414
EKB06014
WP_006265509
WP_015024765
2_C2c2_Lsh
22_C2c2_Lsp
3_C2c2_Lw2
13_C2c2_LwF0279
17_C2c2_Lbu
30_C2c2_Lepto_sp_Marseille
18_C2c2_Hhe
4_C2c2_Lse
11_c2c2_Lwei
12_C2c2_LbFSL
10_C2c2_Pp
31_C2c2_Bacteroides_ihuae
8_c2c2_Cg
9_C2c2_Cg2
32_C2c2_Porph_bacterium
16
WP_013446107
WP_073124441.1
33_C2c2_Listeria_riparia

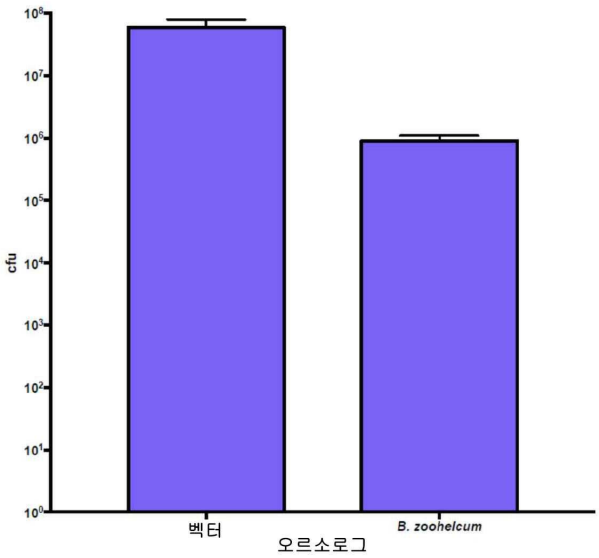
도면2b


5_C2c2_Lbm
27_C2c2_Pseudo_sp
28_C2c2_Buty_sp
7_c2c2_Ca
23_C2c2_NK4A144
19_C2c2_Ere
20_C2c2_Eba
6_C2c2_LbNK
21_C2c2_Bsm
29_C2c2_Blautia_sp
6
14_C2c2_RcSB
15_C2c2_RcR121
16_C2c2_RcDE442
26_C2c2_Thal_sp_TSL5
34_C2c2_Insolitis_peregrinum
25_C2c2_Dem_Aur
24_C2c2_Chloro_Agg
4
ERJ81987
WP_005873511
WP_021665475
WP_061156637
WP_021677657
WP_023846767
ERJ87335
WP_021680012
ERJ65637
WP_021663197
WP_039445055
5
ERI81700
WP_044218239
7_(modified)
7_(modified)_-_residues_only
15
WP_044072147
WP_044074780
WP_042518169

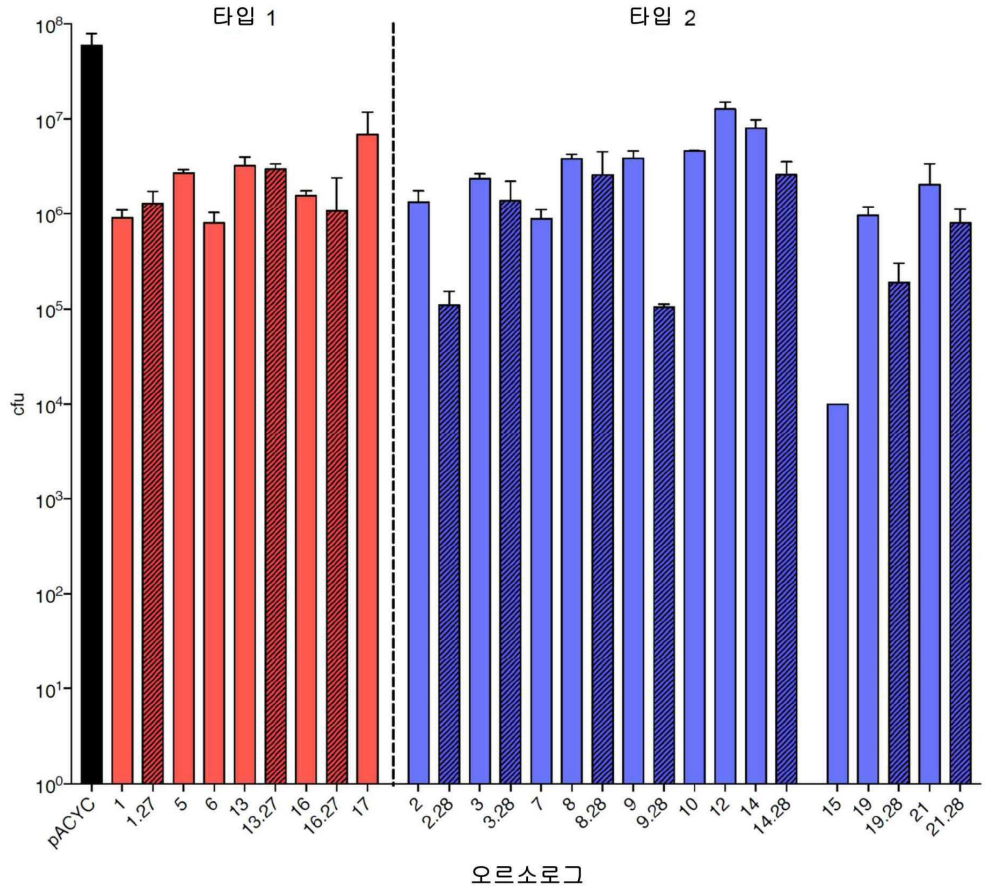
도면2c


8_(modified)
WP_049354263
WP_015345620
WP_061710138
WP_064970887
10
EKY00089
2
EGQ18444
WP_006044833
12
9
21
AFJ07523
BAU18623
KJJ86756
WP_021584635
WP_036931485
WP_036884929
3
EFU31981
WP_004343581
WP_007412163
14
20
WP_039442171
WP_039431778
WP_046201018
WP_039418912
WP_039419792
WP_039426176
WP_039437199
WP_005874195
WP_052912312
19
WP_039417390
WP_061156470
EOA10535
WP_058019250
WP_012458414
WP_013816155

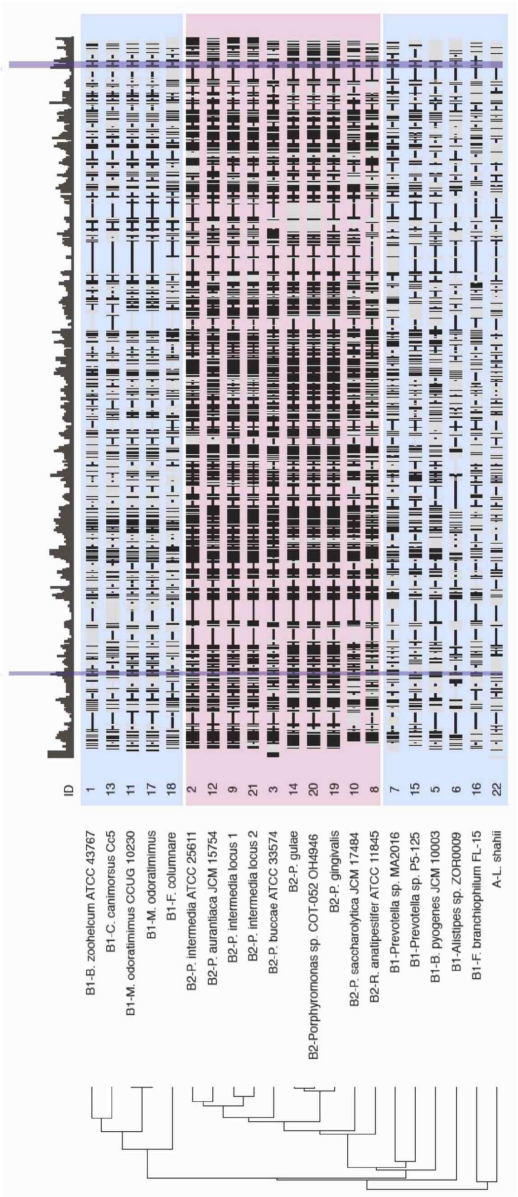
도면3


cfu
벡터
B. zoohelcum
오르소로그

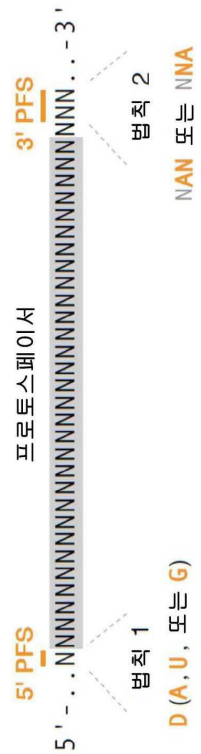
도면4


타입 1
타입 2
cfu
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
pACYC
1
1.27
5
6
13
13.27
16
16.27
17
2
2.28
3
3.28
7
8
8.28
9
9.28
10
12
14
14.28
15
19
19.28
21
21.28
오르소로그

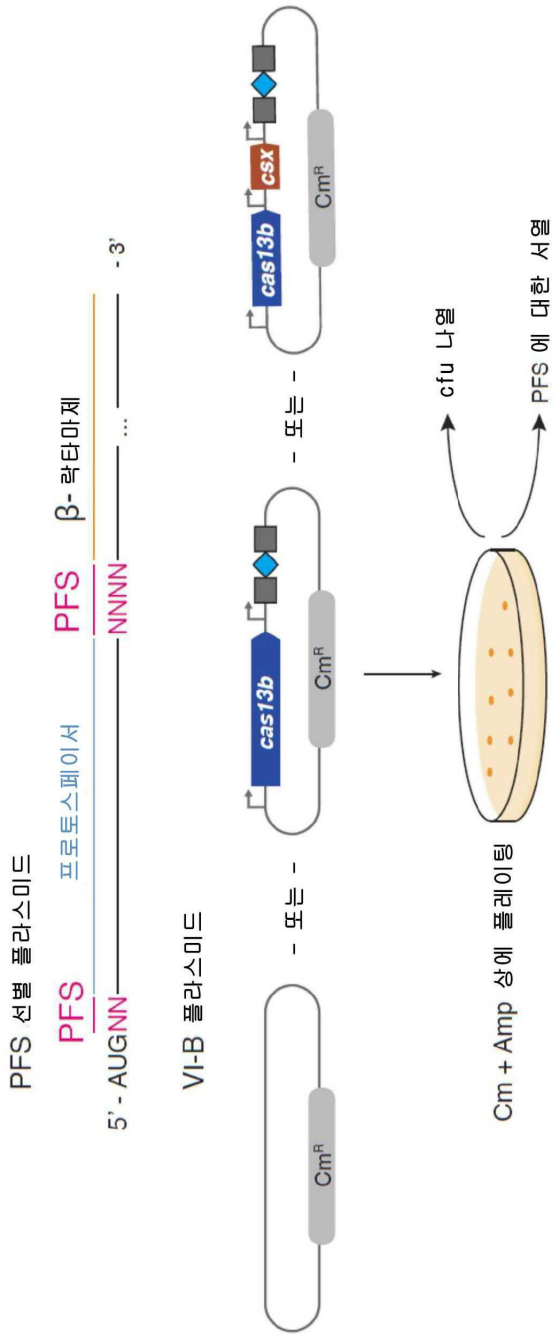
도면5


ID
1
13
11
17
18
2
12
9
21
3
14
20
19
10
8
7
15
5
6
16
22
B1-B. zoohelcum ATCC 43767
B1-C. canimorsus Cc5
B1-M. odoratimimus CCUG 10230
B1-M. odoratimimus
B1-F. columnare
B2-P. intermedia ATCC 25611
B2-P. aurantiaca JCM 15754
B2-P. intermedia locus 1
B2-P. intermedia locus 2
B2-P. buccae ATCC 33574
B2-P. gulae
B2-Porphyromonas sp. COT-052 OH4946
B2-P. gingivalis
B2-P. saccharolytica JCM 17484
B2-R. analipesifer ATCC 11845
B1-Prevotella sp. MA2016
B1-Prevotella sp. P5-125
B1-B. pyogenes JCM 10003
B1-Alistipes sp. ZOR0009
B1-F. branchiophilum FL-15
A-L. shahii

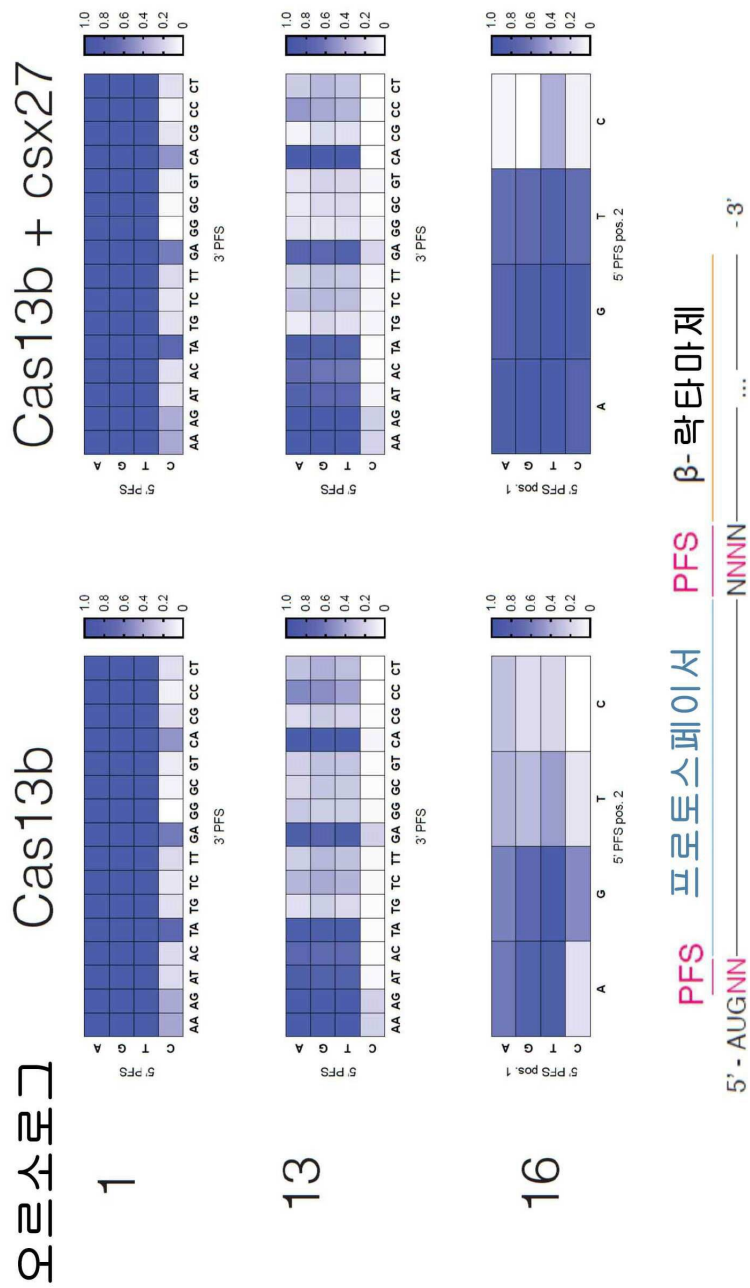
도면6a


5' PFS
프로토스페이서
3' PFS
5' - . . NNNN

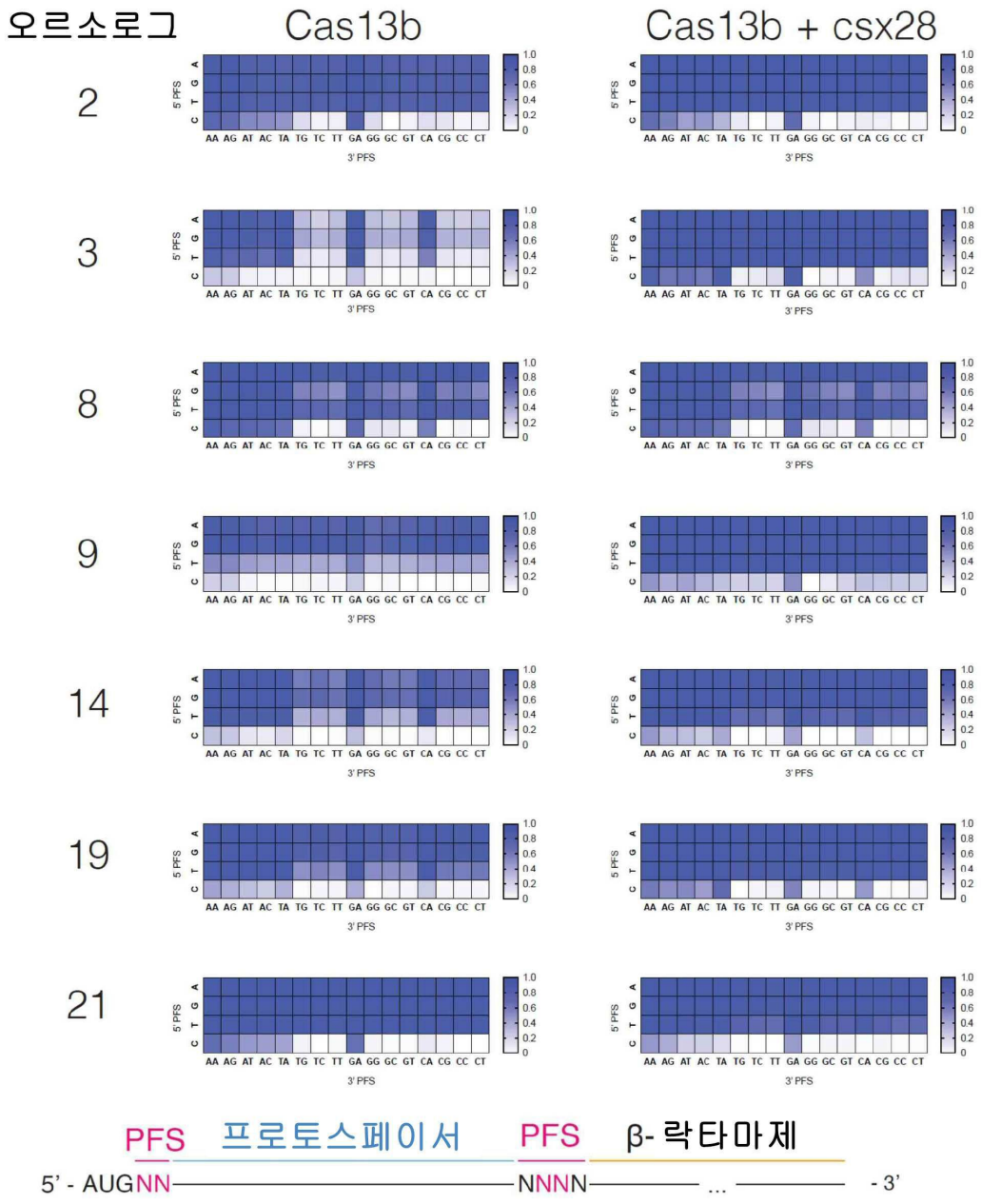
도면6b


PFS 선별 플라스미드
PFS
프로토스페이서
PFS
β-락타마제
5'- AUGNN
NNNN
...
-3'
VI-B 플라스미드
- 또는 -
cas13b
Cm^R
csx
- 또는 -
Cm + Amp 상에 플레이팅
cfu 나열
PFS 에 대한 서열

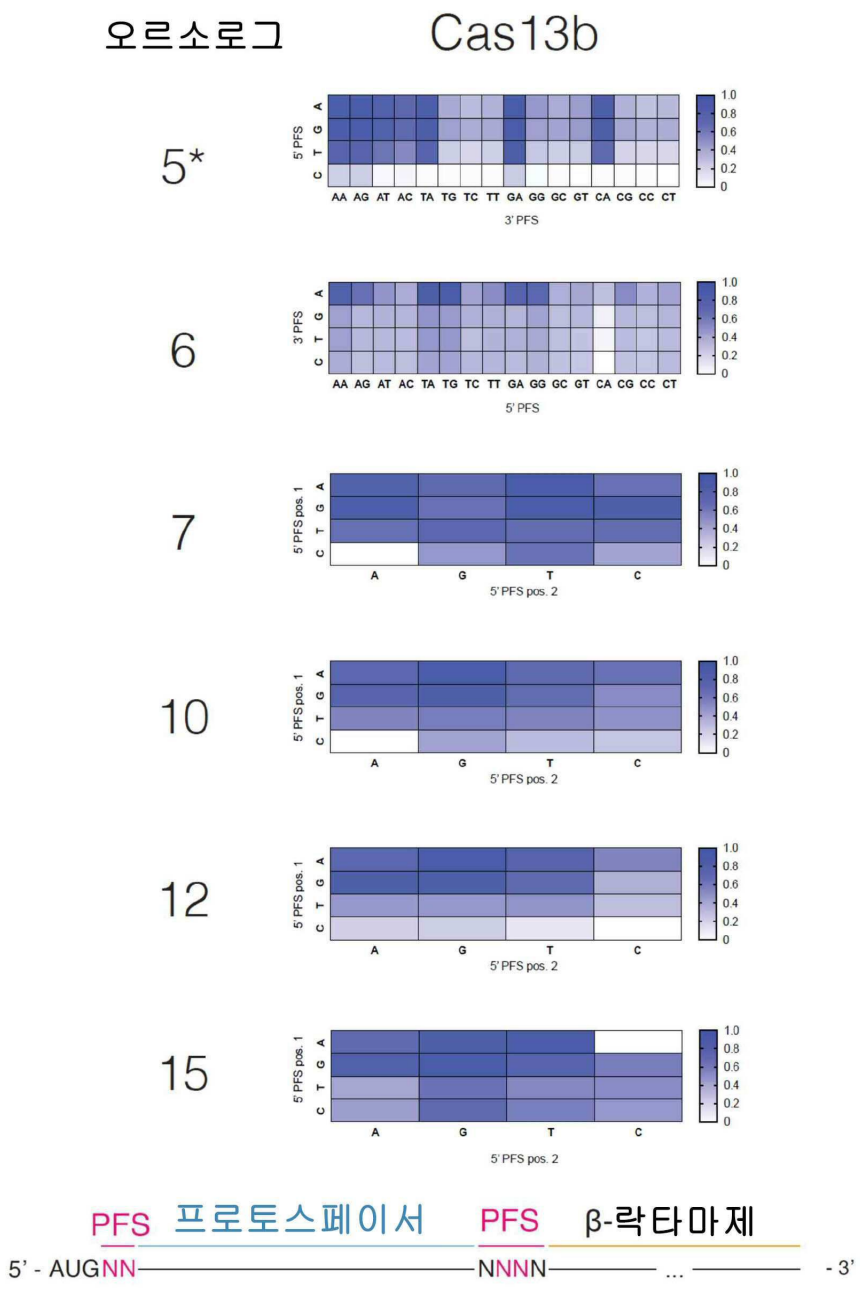
도면7


오르소로그
Cas13b
Cas13b + csx27
1
13
16
5' PFS
3' PFS
5' PFS pos. 1
5' PFS pos. 2
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
A G T C
1.0 0.8 0.6 0.4 0.2 0
5' - AUGNN
PFS
프로토스페이서
PFS
NNNN
β-락타마제
...
- 3'

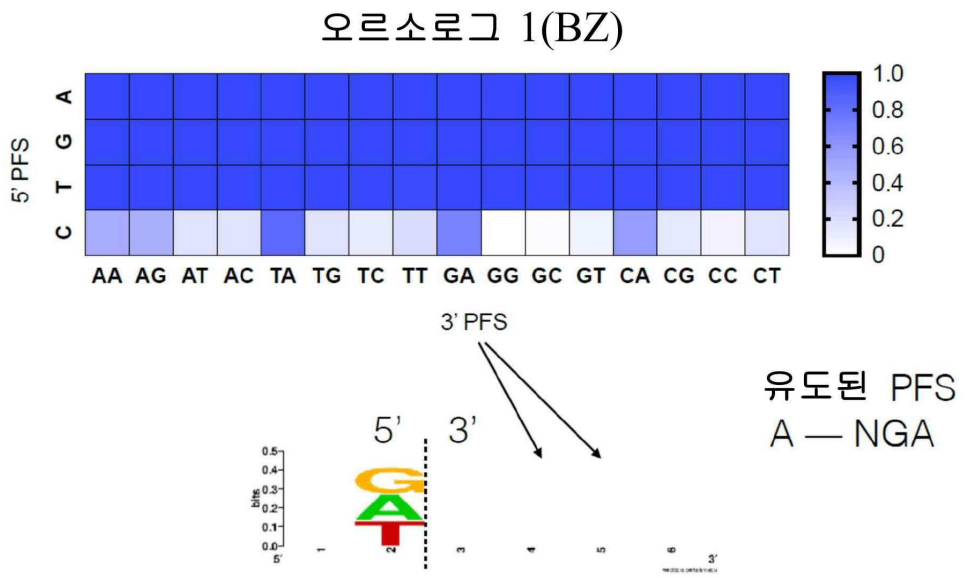
도면8


오르소로그
Cas13b
Cas13b + csx28
2
3
8
9
14
19
21
5' PFS
3' PFS
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
PFS
프로토스페이서
PFS
β-락타마제
5' - AUGNN
NNNN
...
- 3'

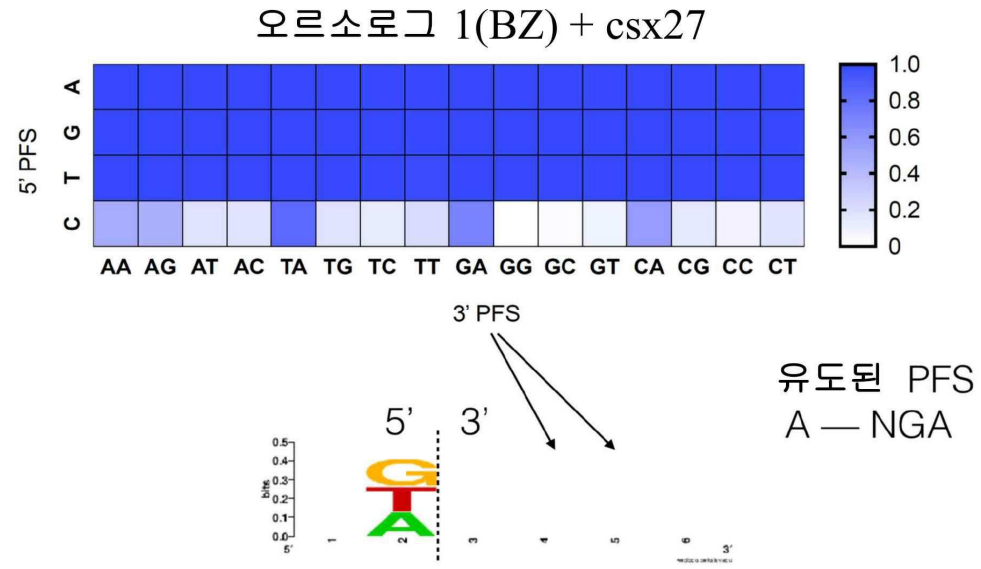
도면9


오르소로그
Cas13b
5*
6
7
10
12
15
5' PFS
3' PFS
5' PFS pos. 1
5' PFS pos. 2
A G T C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
1.0 0.8 0.6 0.4 0.2 0
PFS
프로토스페이서
PFS
β-락타마제
5' - AUGNN
NNNN
...
- 3'

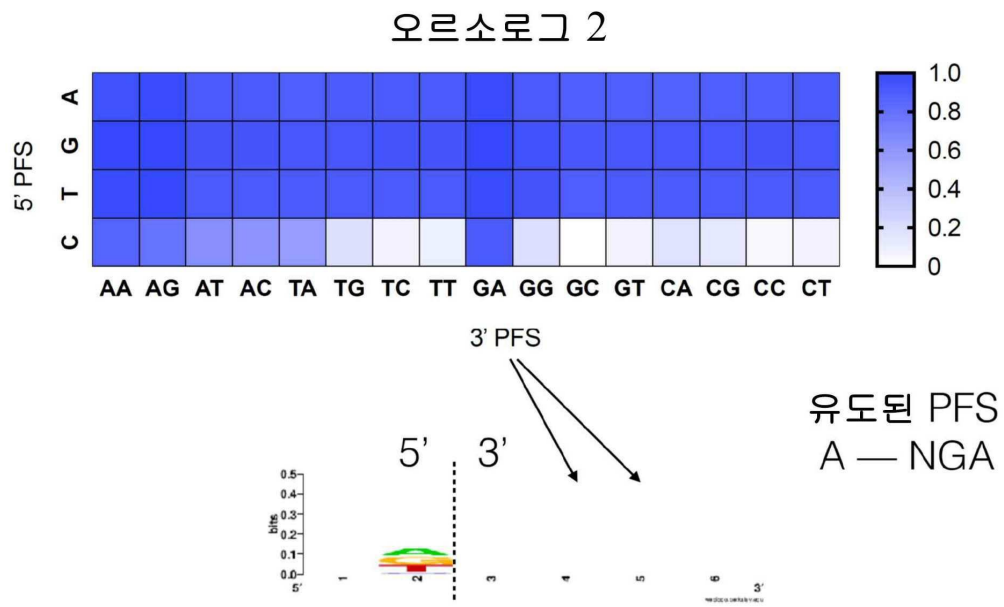
도면10a


오르소로그 1(BZ)
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
1.0
0.8
0.6
0.4
0.2
0
3' PFS
5'
3'
유도된 PFS
A — NGA

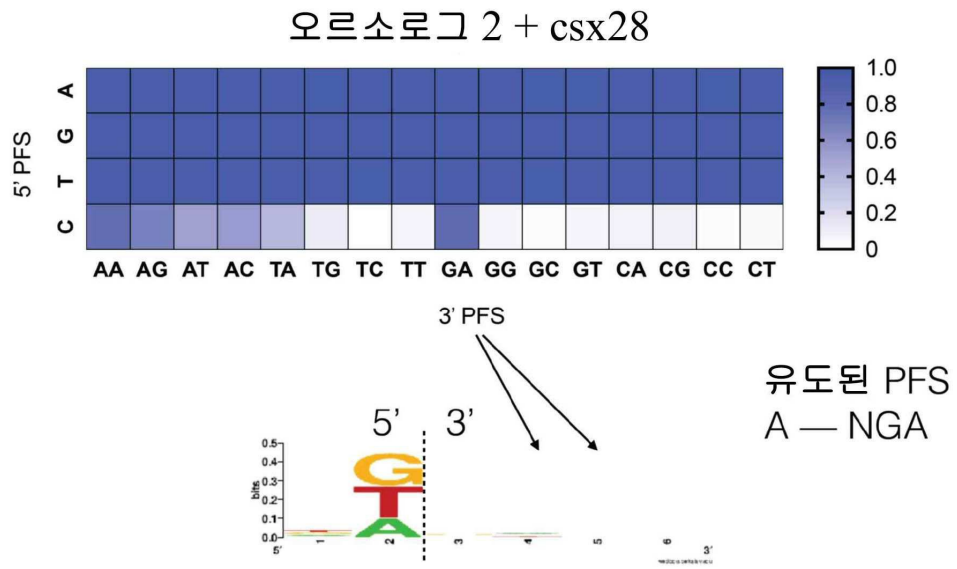
도면10b


오르소로그 1(BZ) + csx27
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
1.0
0.8
0.6
0.4
0.2
0
3' PFS
5'
3'
유도된 PFS
A — NGA

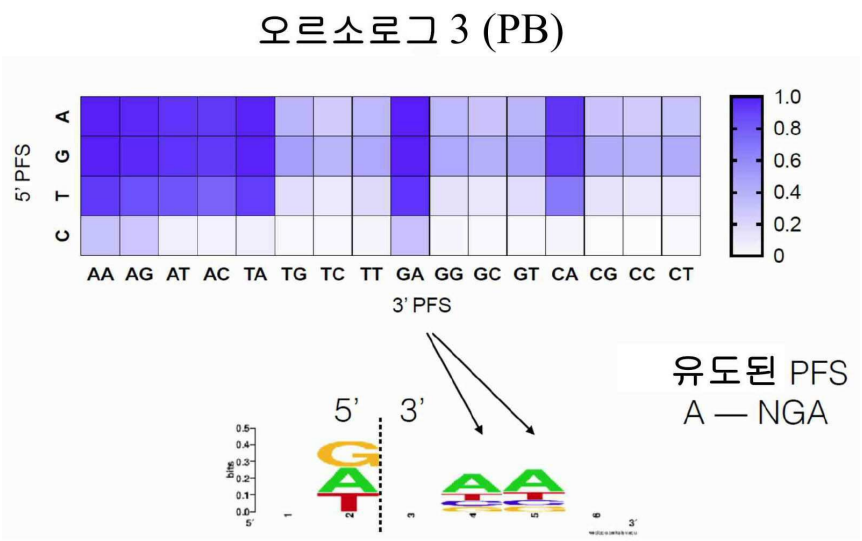
도면10c


오르소로그 2
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
1.0
0.8
0.6
0.4
0.2
0
3' PFS
5'
3'
유도된 PFS
A — NGA

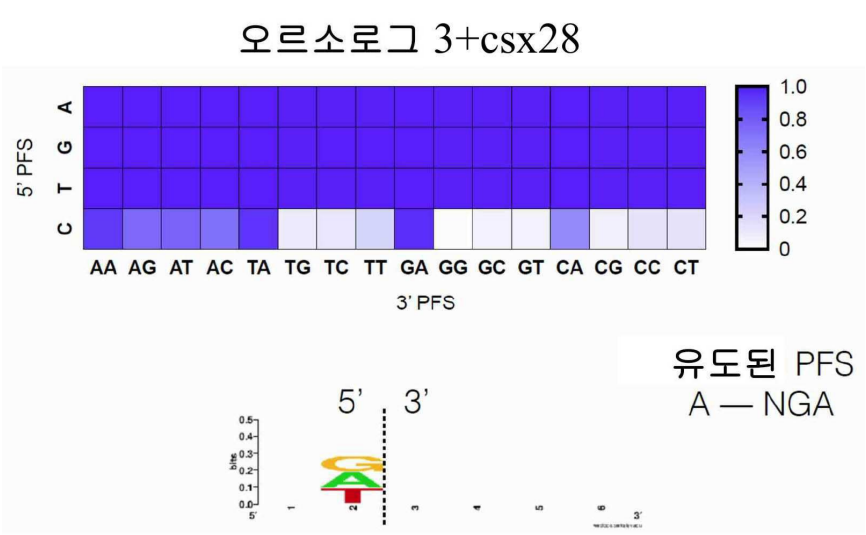
도면10d


오르소로그 2 + csx28
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
1.0
0.8
0.6
0.4
0.2
0
3' PFS
5'
3'
유도된 PFS
A — NGA

도면10e

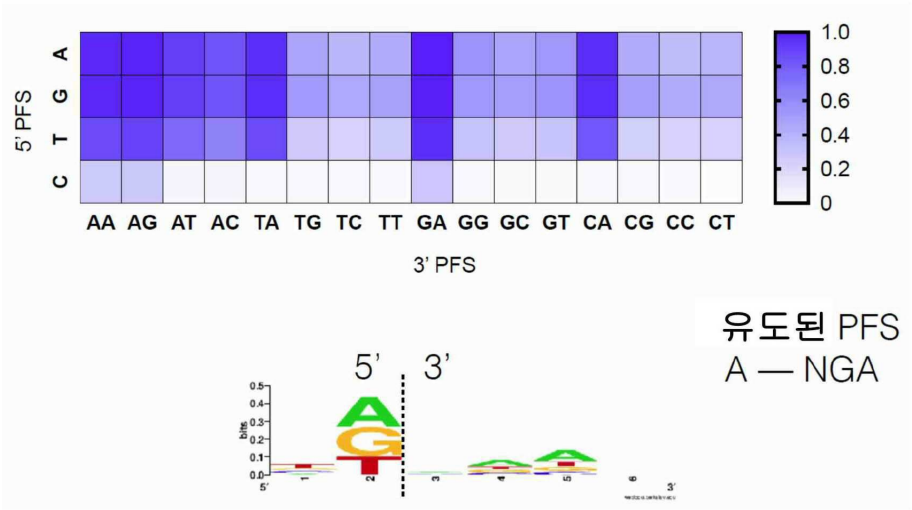

오르소로그 3 (PB)
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
3' PFS
1.0
0.8
0.6
0.4
0.2
0
5'
3'
유도된 PFS
A — NGA

도면10f


오르소로그 3+csx28
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
3' PFS
1.0
0.8
0.6
0.4
0.2
0
5'
3'
유도된 PFS
A — NGA

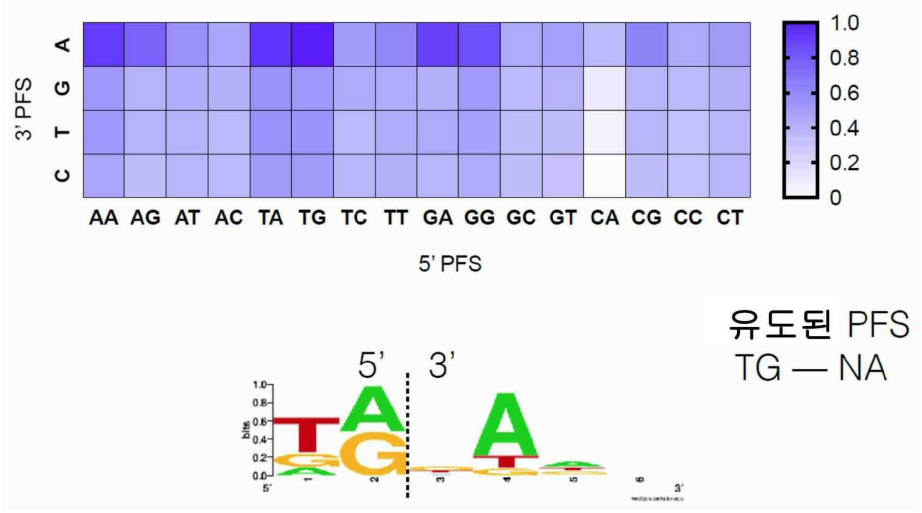
도면10g

오르소로그 5

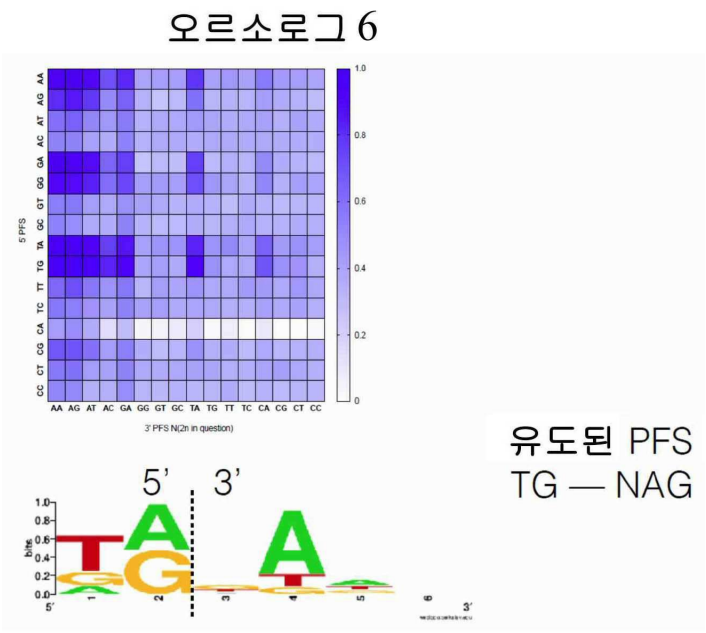


도면10h

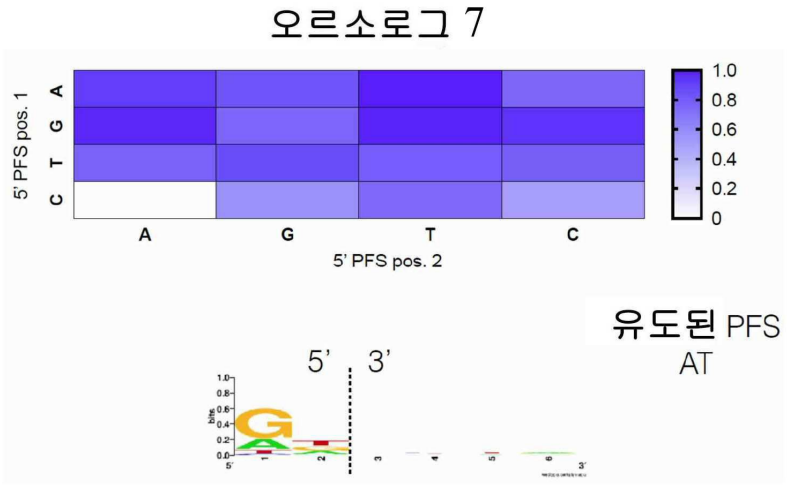
오르소로그6



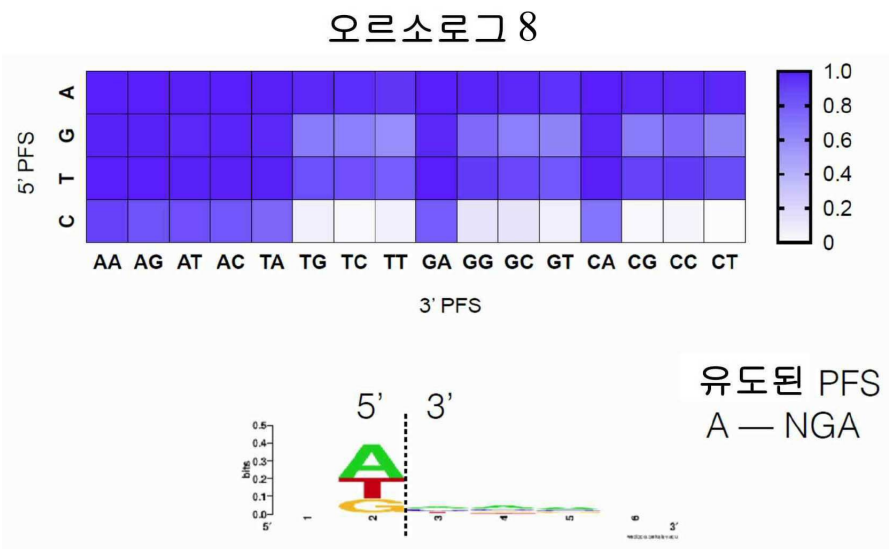
도면10i


오르소로그 6
5' PFS
3' PFS N(2n in question)
AA AG AT AC GA GG GT GC TA TG TT TC CA CG CT CC
1.0
0.8
0.6
0.4
0.2
0
유도된 PFS
TG — NAG
5'
3'
bits

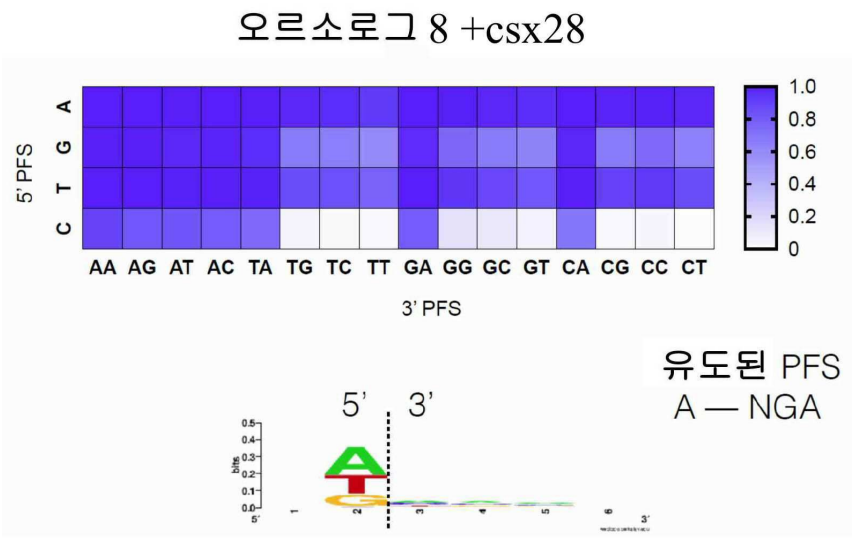
도면10j


오르소로그 7
5' PFS pos. 1
A
G
T
C
5' PFS pos. 2
1.0
0.8
0.6
0.4
0.2
0
유도된 PFS
AT
5'
3'
bits

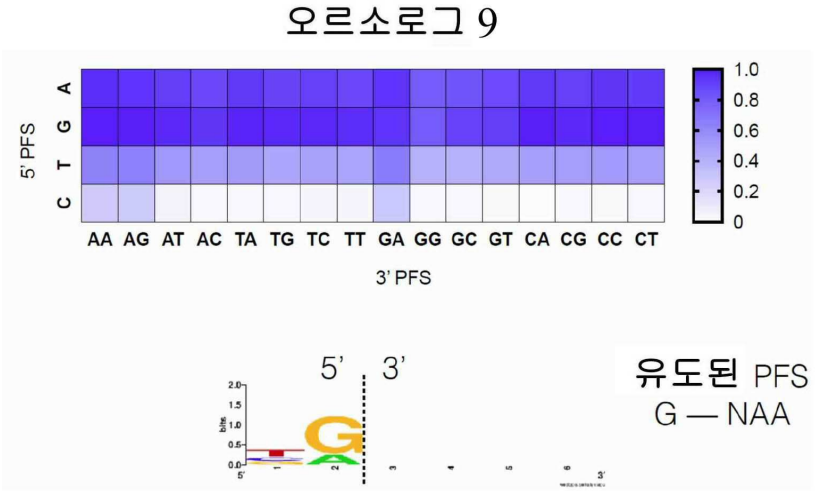
도면10k



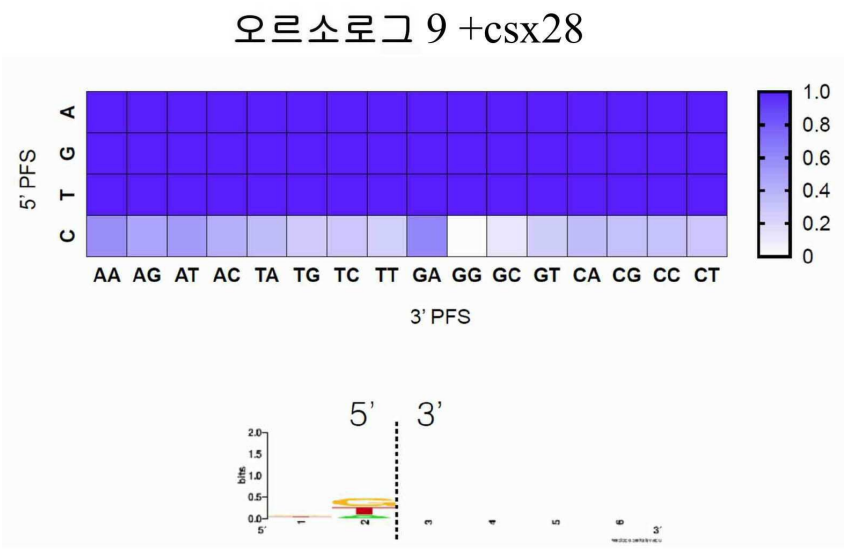
도면101



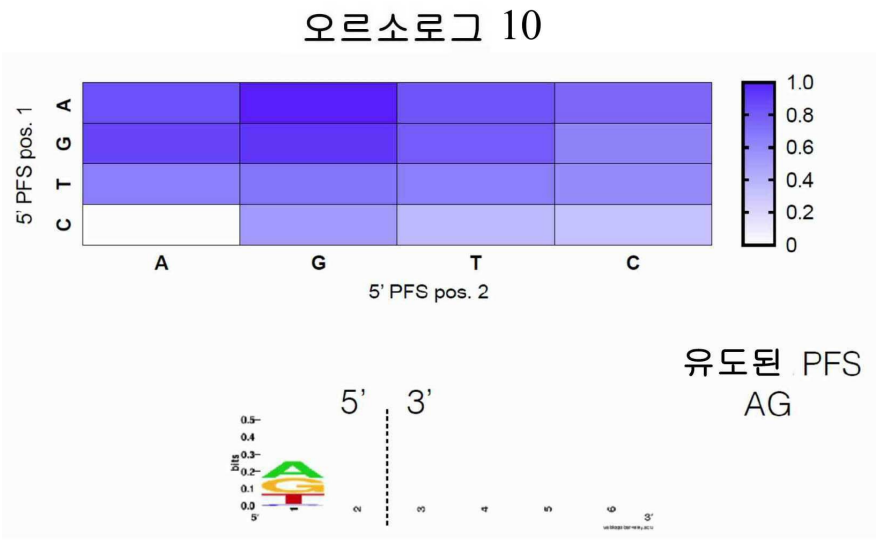
도면10m



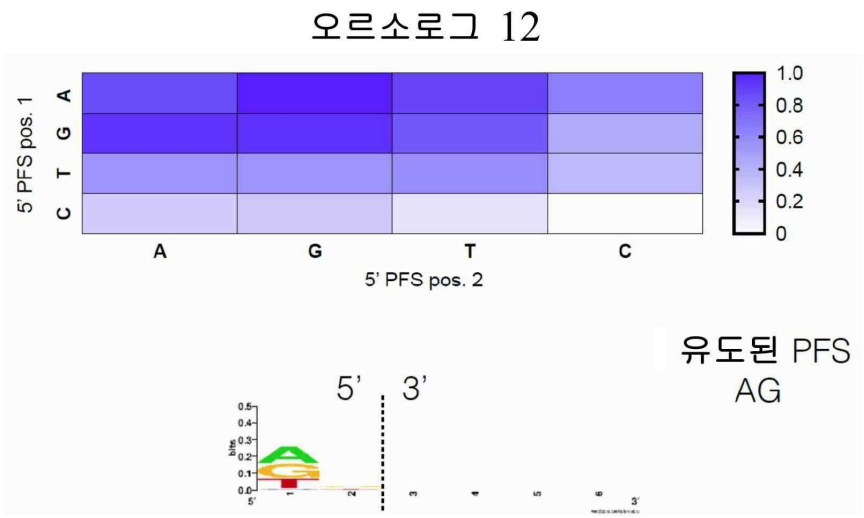
도면10n


오르소로그 9 +csx28
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
3' PFS
1.0
0.8
0.6
0.4
0.2
0
5'
3'
bits

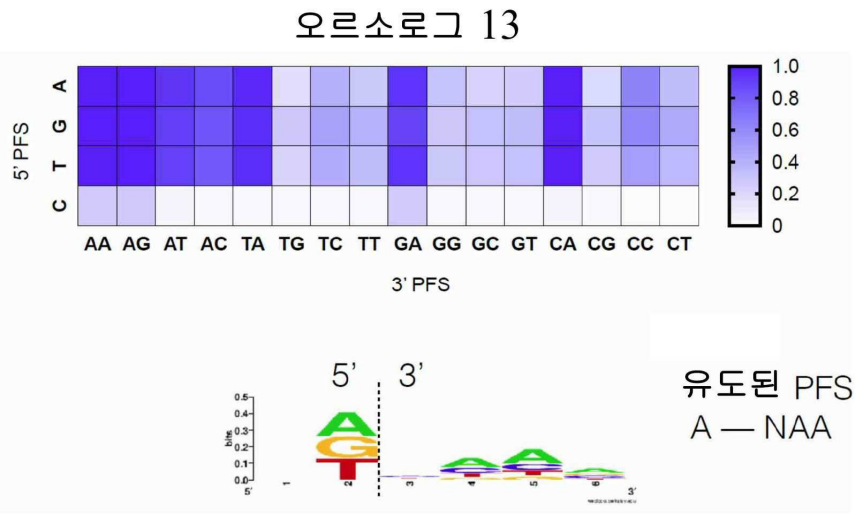
도면10o


오르소로그 10
5' PFS pos. 1
A
G
T
C
A
G
T
C
5' PFS pos. 2
1.0
0.8
0.6
0.4
0.2
0
유도된 PFS
AG
5'
3'
bits

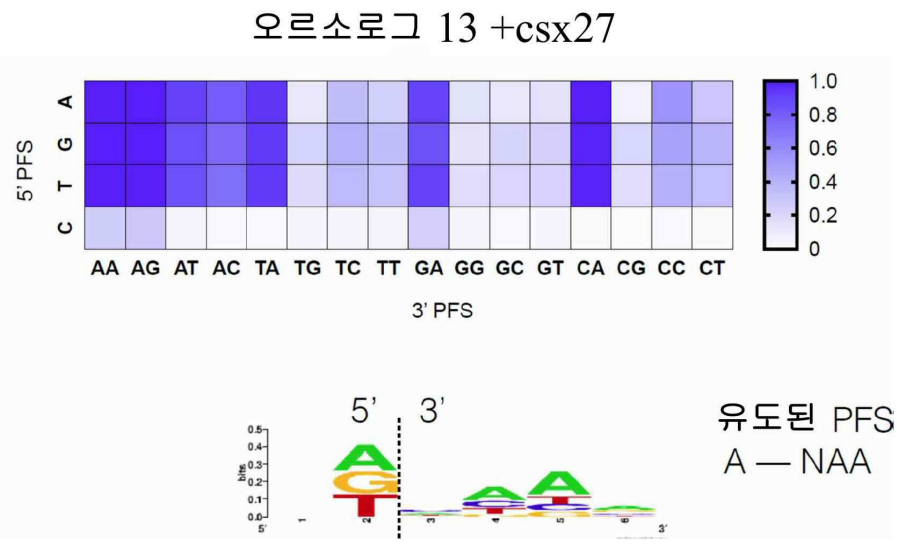
도면10p



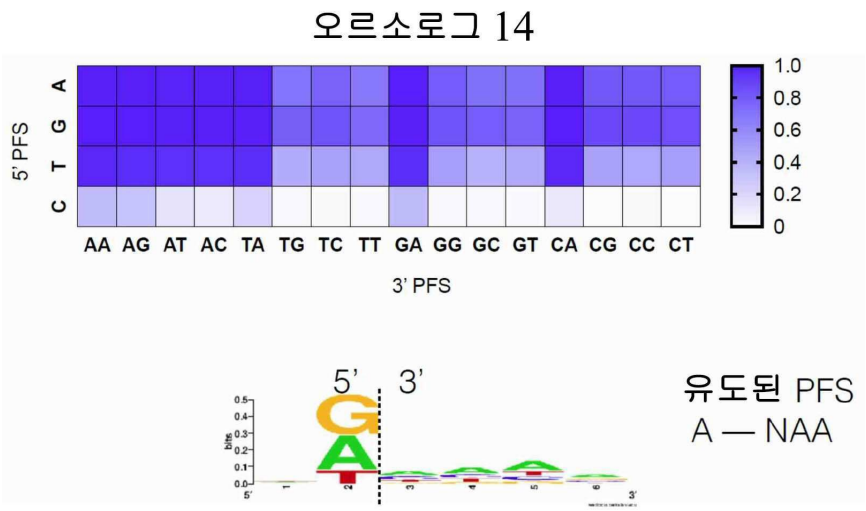
도면10q



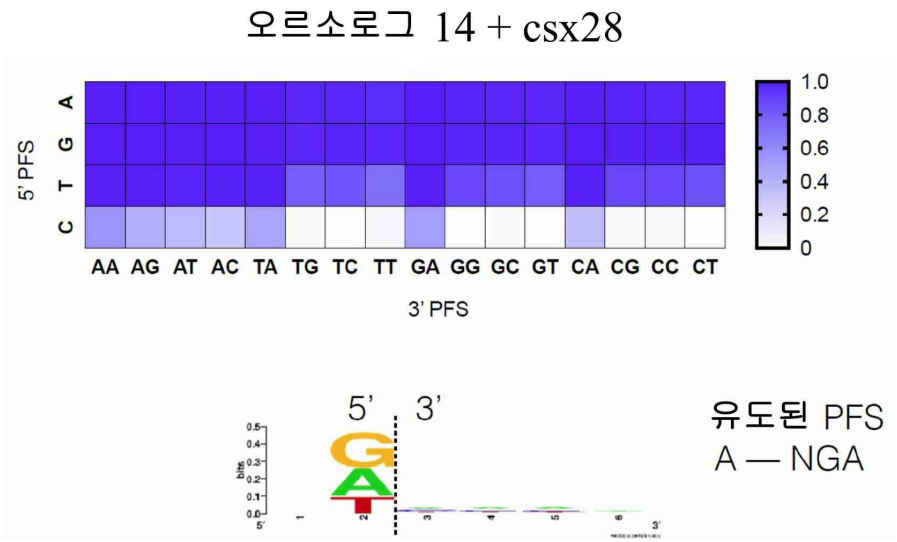
도면10r



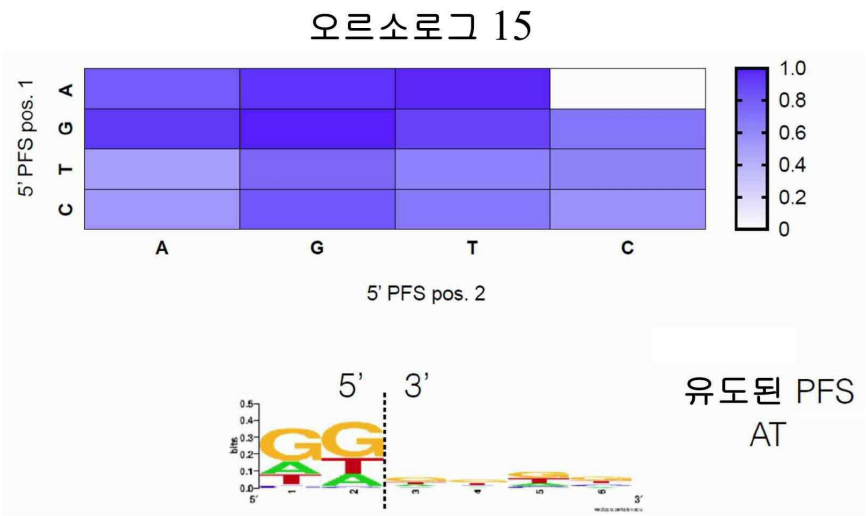
도면10s



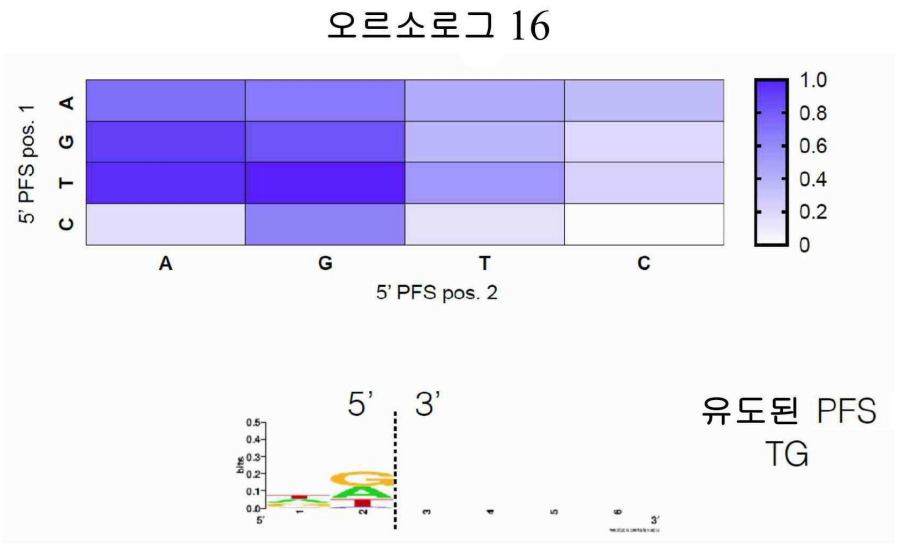
도면10t



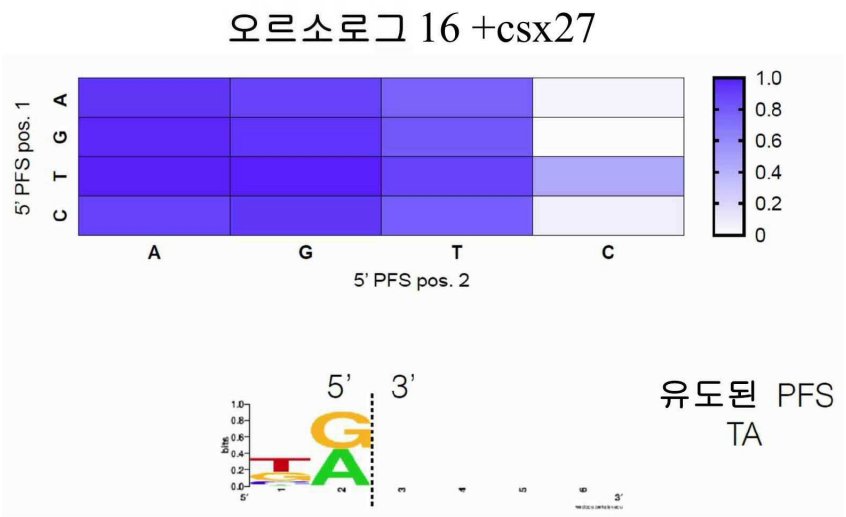
도면10u



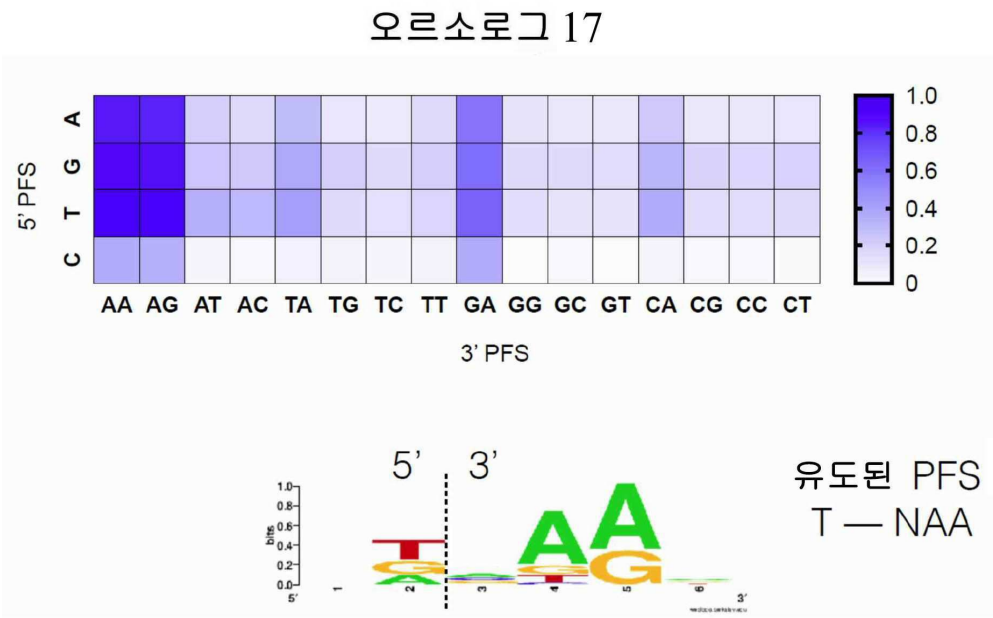
도면10v


오르소로그 16
5' PFS pos. 1
A
G
T
C
1.0
0.8
0.6
0.4
0.2
0
A
G
T
C
5' PFS pos. 2
5'
3'
유도된 PFS
TG

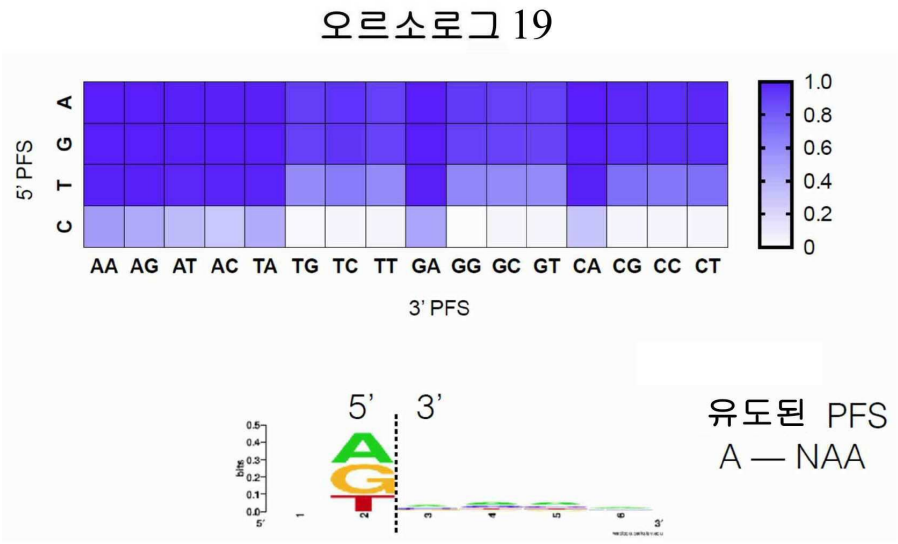
도면10w


오르소로그 16 +csx27
5' PFS pos. 1
A
G
T
C
1.0
0.8
0.6
0.4
0.2
0
A
G
T
C
5' PFS pos. 2
5'
3'
유도된 PFS
TA

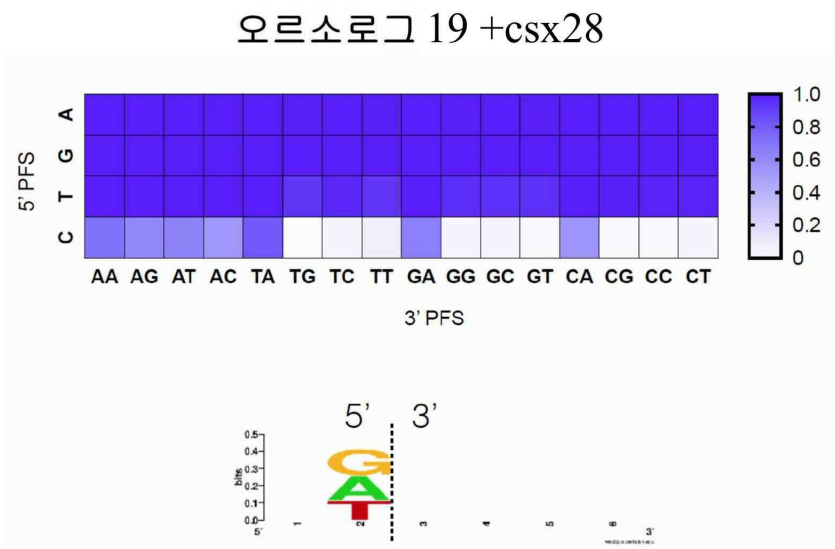
도면10x


오르소로그 17
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
3' PFS
1.0
0.8
0.6
0.4
0.2
0
5'
3'
유도된 PFS
T — NAA

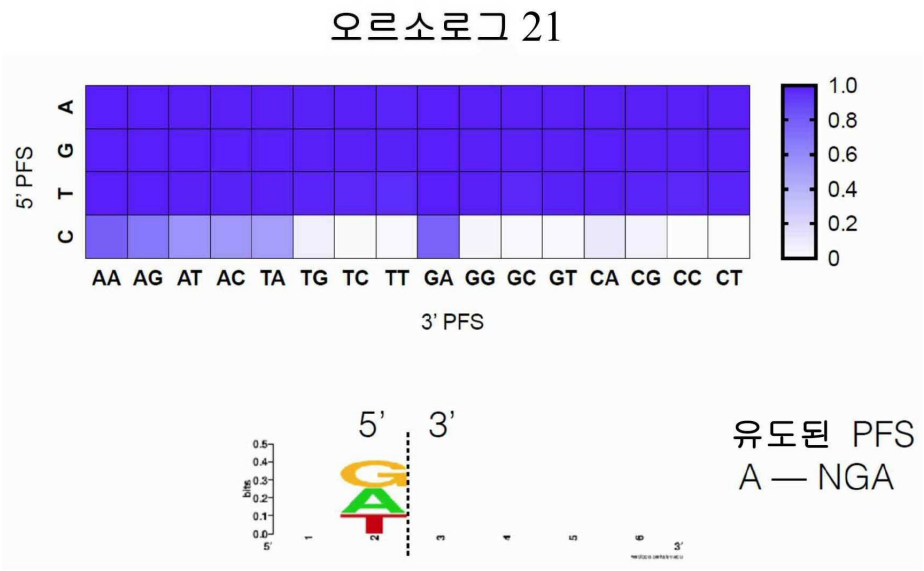
도면10y


오르소로그 19
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
3' PFS
1.0
0.8
0.6
0.4
0.2
0
5'
3'
유도된 PFS
A — NAA

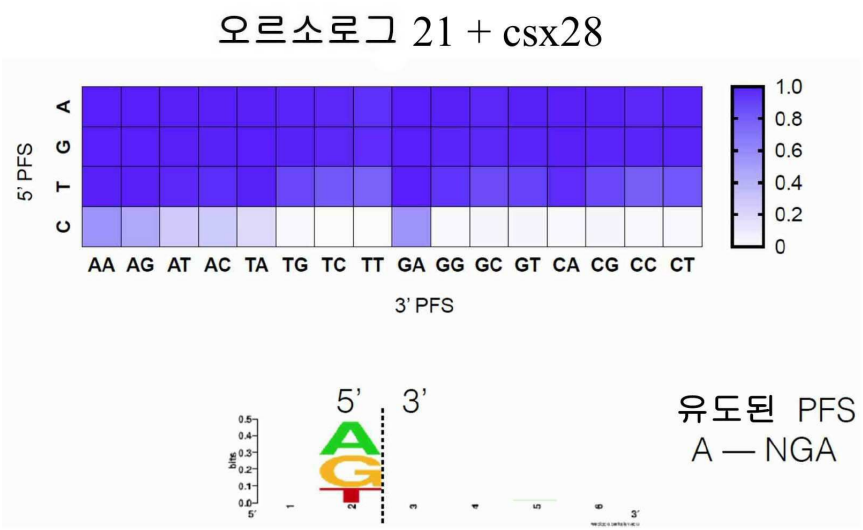
도면10z



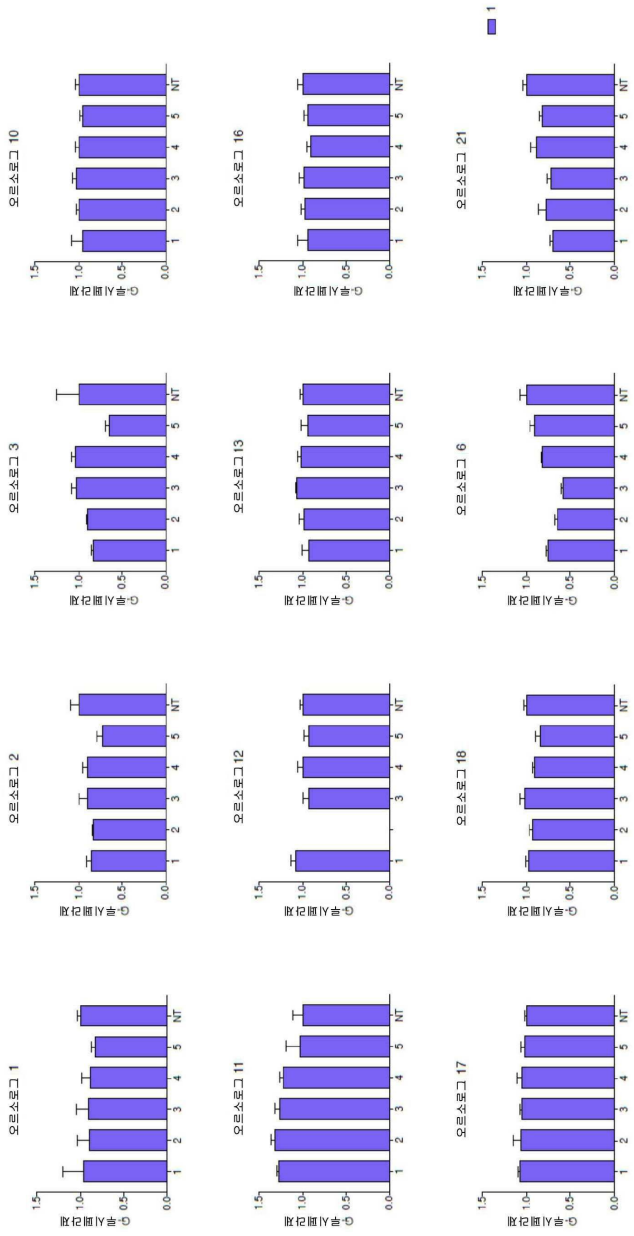
도면10aa



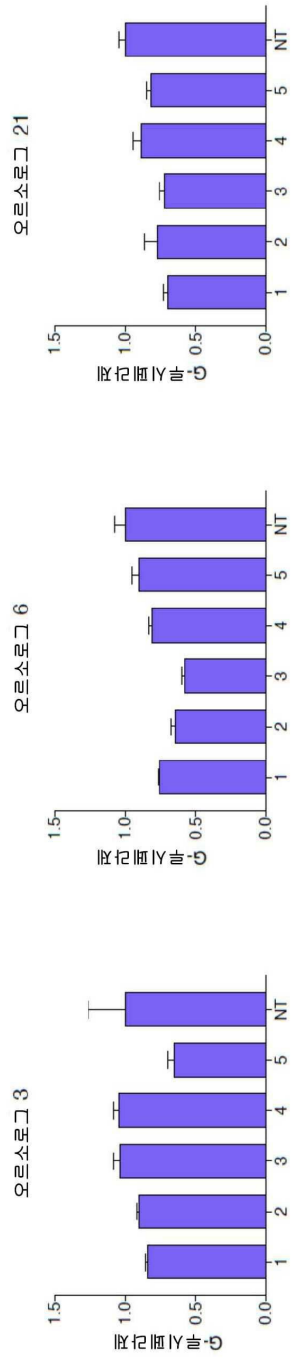
도면10bb


오르소로그 21 + csx28
5' PFS
A
G
T
C
AA AG AT AC TA TG TC TT GA GG GC GT CA CG CC CT
3' PFS
1.0
0.8
0.6
0.4
0.2
0
5'
3'
유도된 PFS
A — NGA

도면11

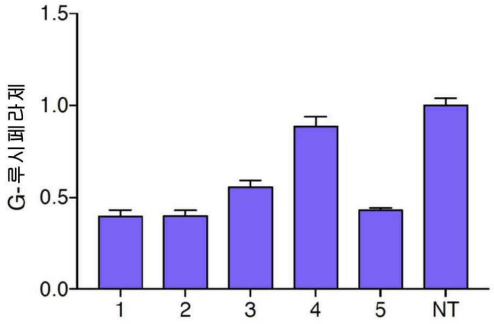

1
오르소로그 10
오르소로그 16
오르소로그 21
오르소로그 3
오르소로그 13
오르소로그 6
오르소로그 2
오르소로그 12
오르소로그 18
오르소로그 1
오르소로그 11
오르소로그 17
G. 루시퍼라제
NT

도면12

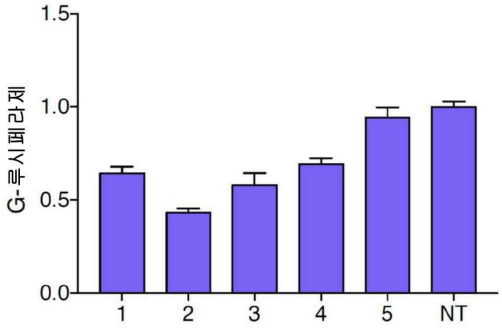


도면13

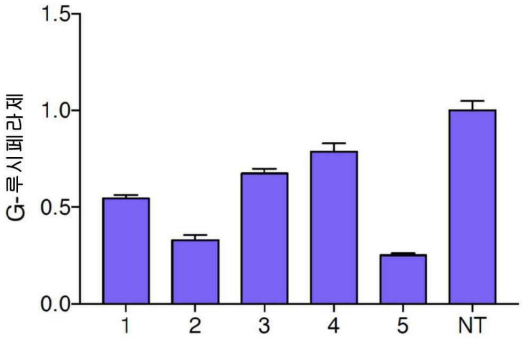
오르소로그 5



오르소로그 7

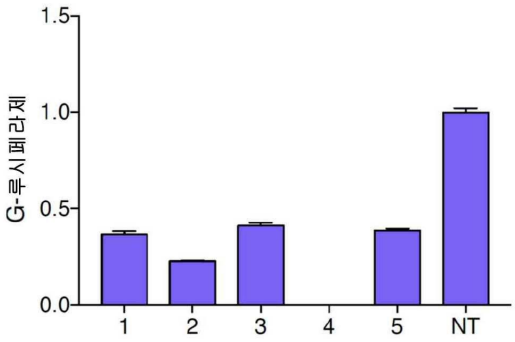


오르소로그 8

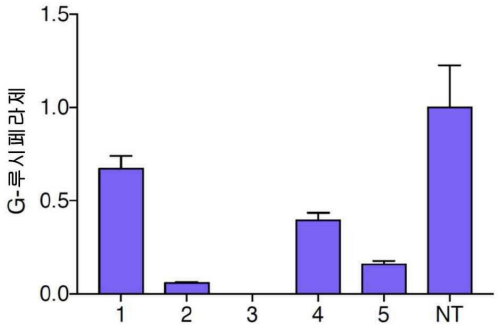


도면14a

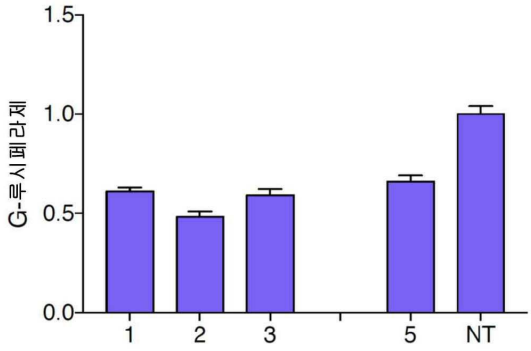
오르소로그 14



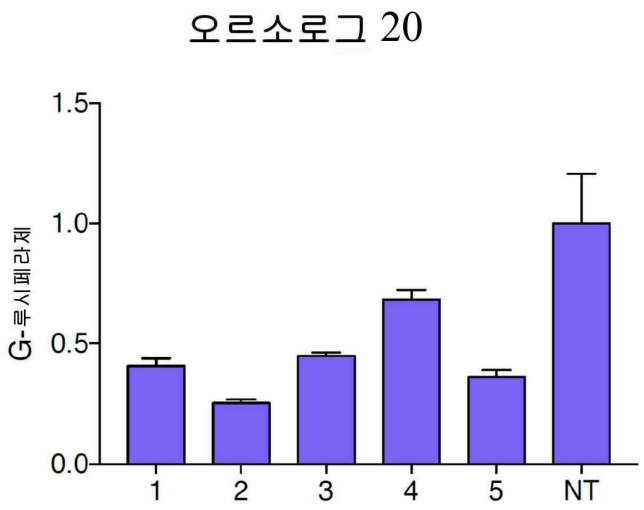
오르소로그 15



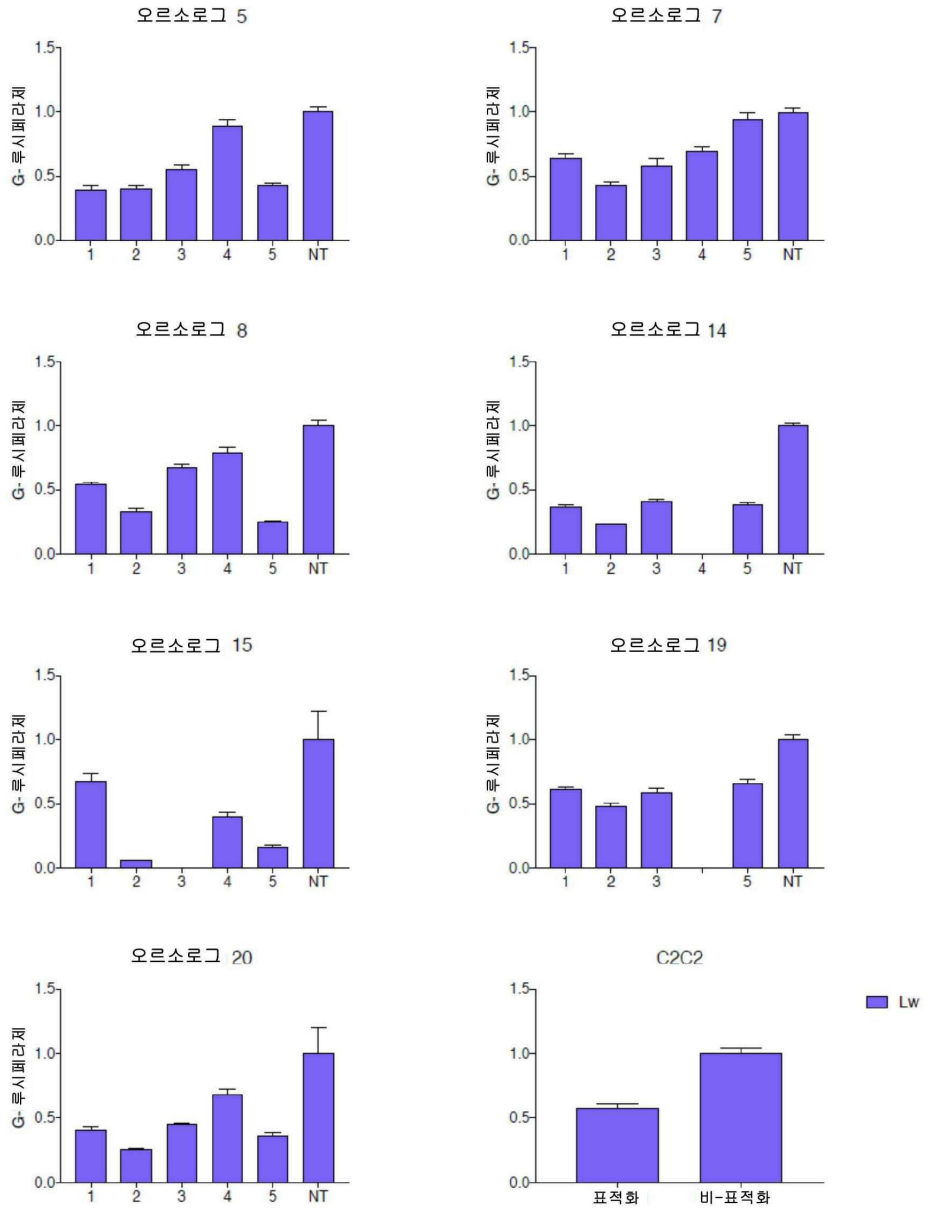
오르소로그 19



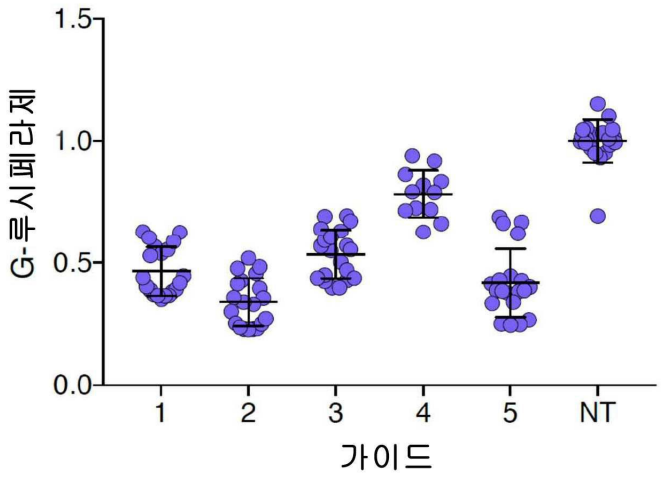
도면14b



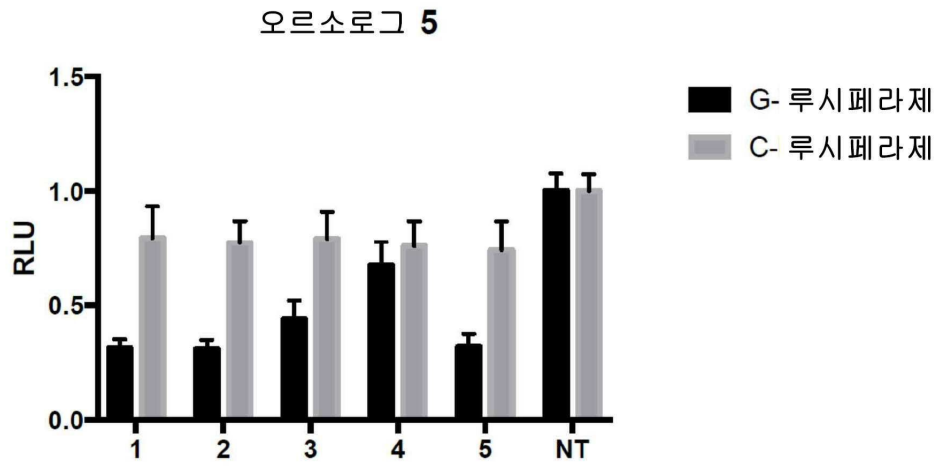
도면15


오르소로그 5
오르소로그 7
오르소로그 8
오르소로그 14
오르소로그 15
오르소로그 19
오르소로그 20
C2C2
G-루시페라제
1.5
1.0
0.5
0.0
1
2
3
4
5
NT
Lw
표적화
비-표적화

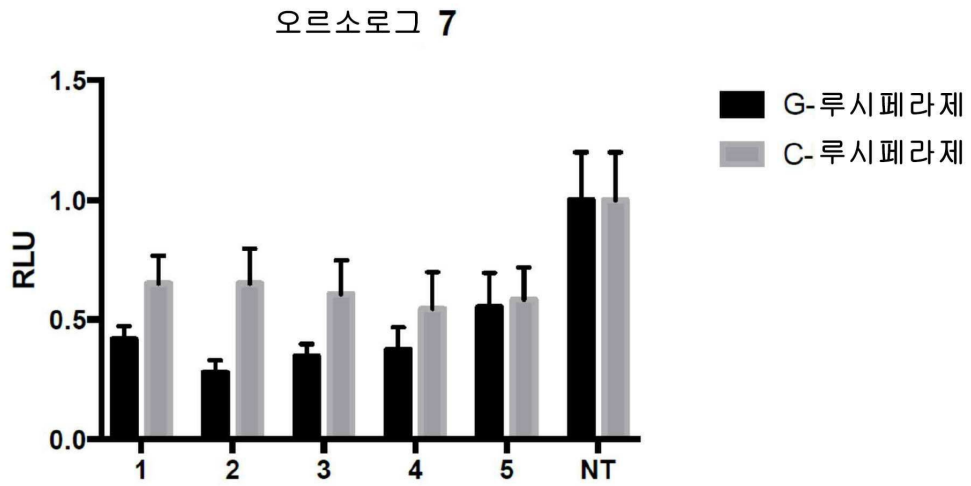
도면16


G-루시페라제
1.5
1.0
0.5
0.0
1
2
3
4
5
NT
가이드

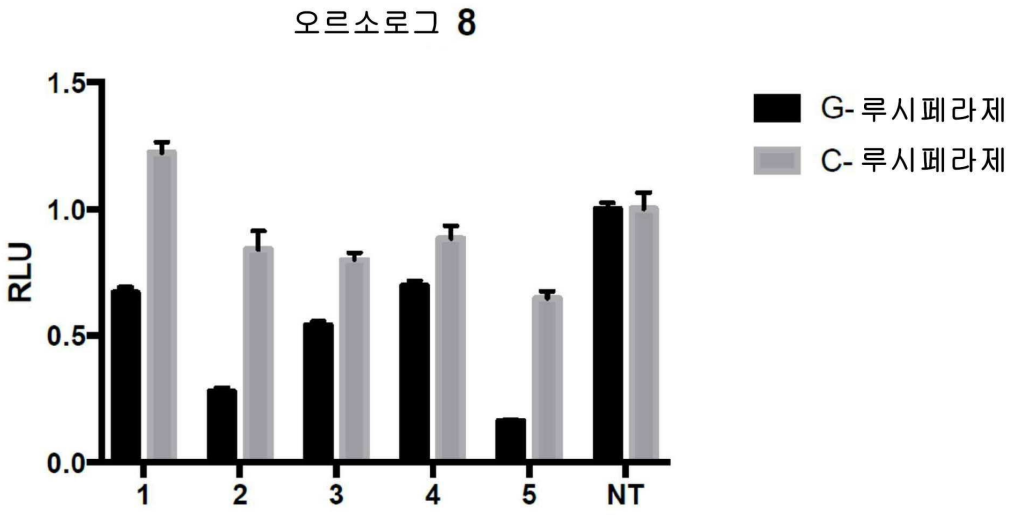
도면17a



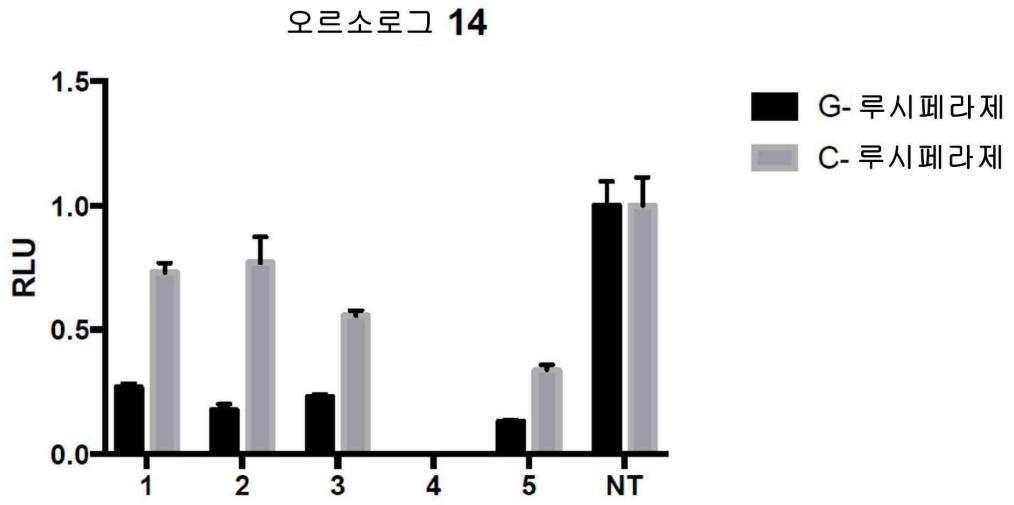
도면17b



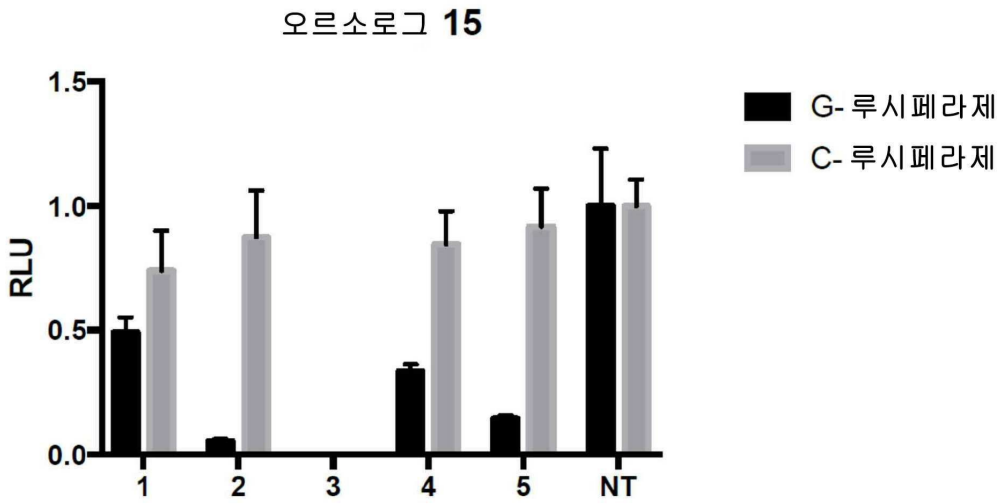
도면17c



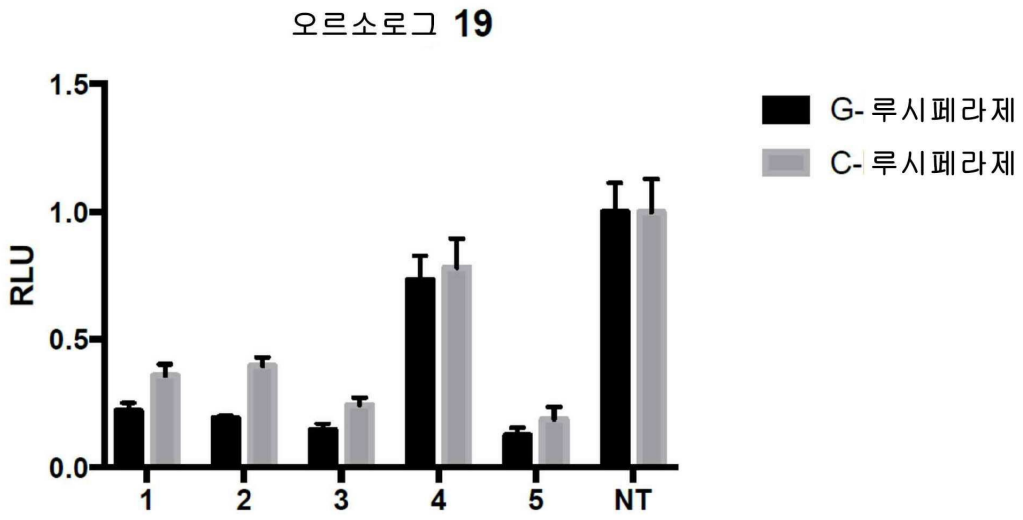
도면17d



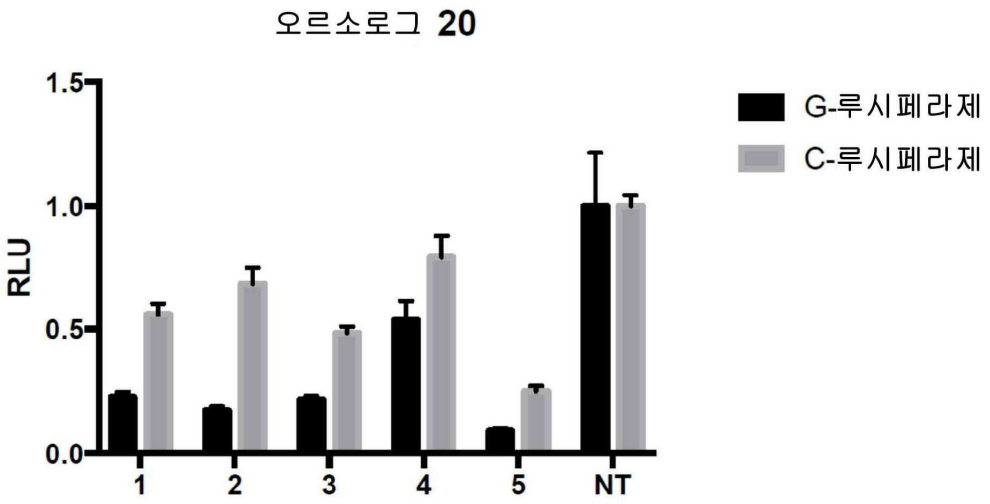
도면17e



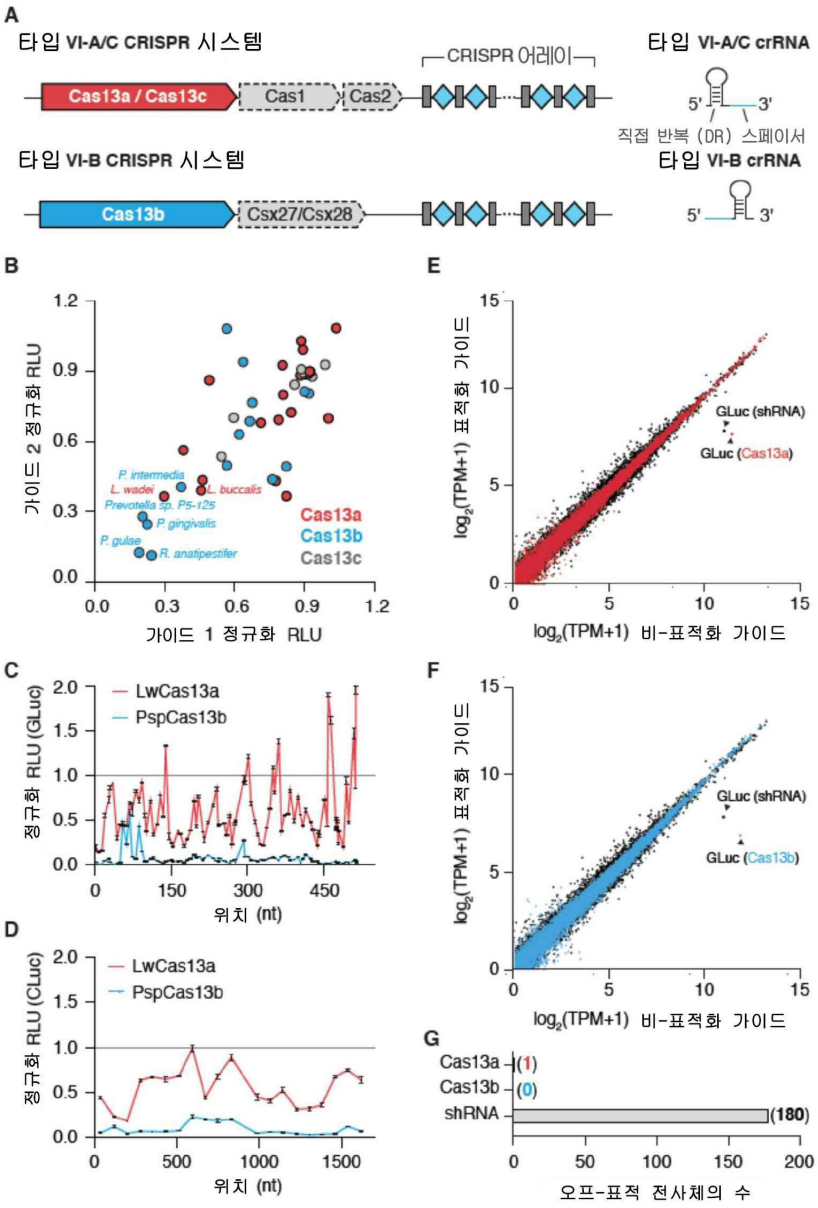
도면17f



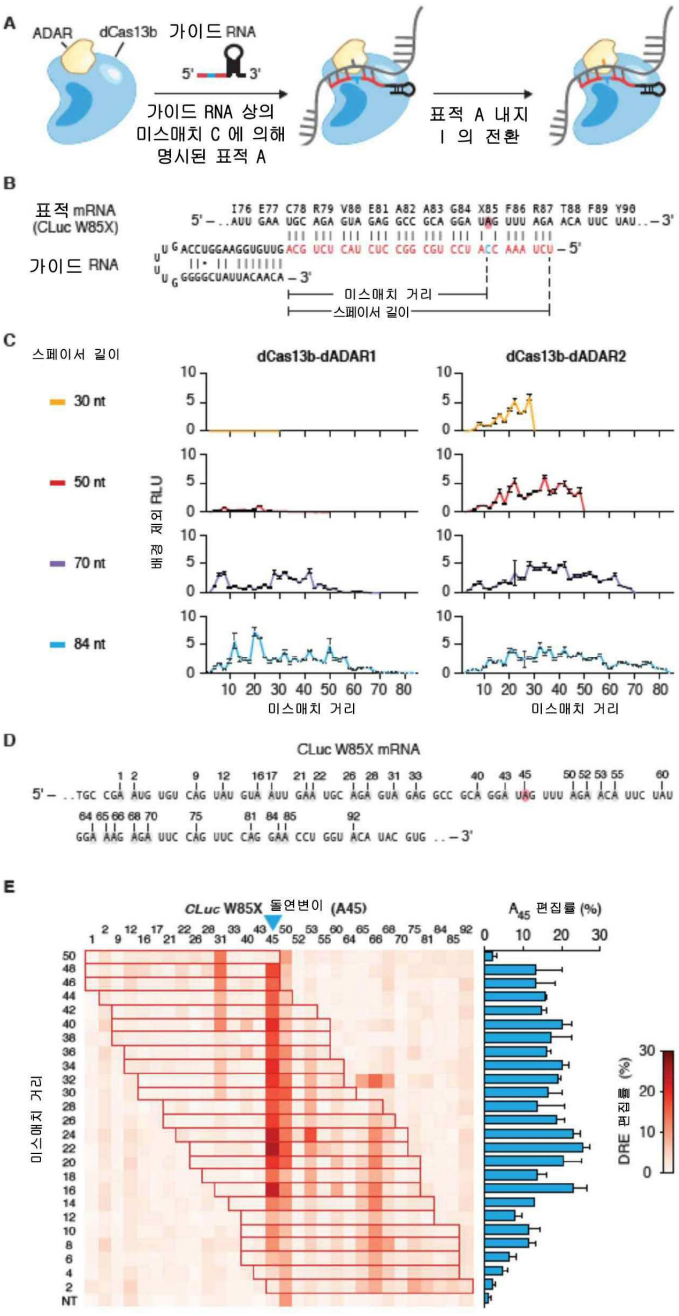
도면17g



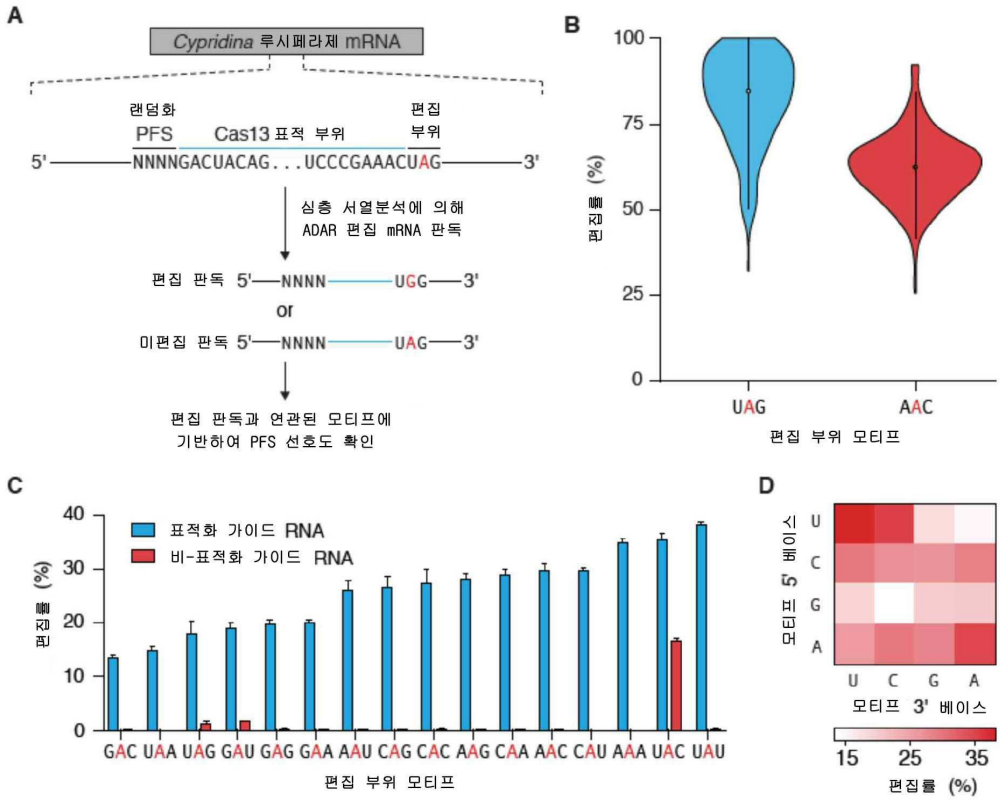
도면18


A
타입 VI-A/C CRISPR 시스템
CRISPR 어레이
Cas13a / Cas13c
Cas1
Cas2
타입 VI-A/C crRNA
5'
3'
직접 반복 (DR) 스페이서
타입 VI-B CRISPR 시스템
Cas13b
Csx27/Csx28
타입 VI-B crRNA
B
가이드 2 정규화 RLU
가이드 1 정규화 RLU
P. intermedia
L. wadei
L. buccalis
Prevotella sp. P5-125
P. gingivalis
P. gulae
R. anatipestifer
Cas13a
Cas13b
Cas13c
C
정규화 RLU (GLuc)
LwCas13a
PspCas13b
위치 (nt)
D
정규화 RLU (CLuc)
LwCas13a
PspCas13b
위치 (nt)
E
log2(TPM+1) 표적화 가이드
log2(TPM+1) 비-표적화 가이드
GLuc (shRNA)
GLuc (Cas13a)
F
log2(TPM+1) 표적화 가이드
log2(TPM+1) 비-표적화 가이드
GLuc (shRNA)
GLuc (Cas13b)
G
Cas13a
Cas13b
shRNA
(1)
(0)
(180)
오프-표적 전사체의 수

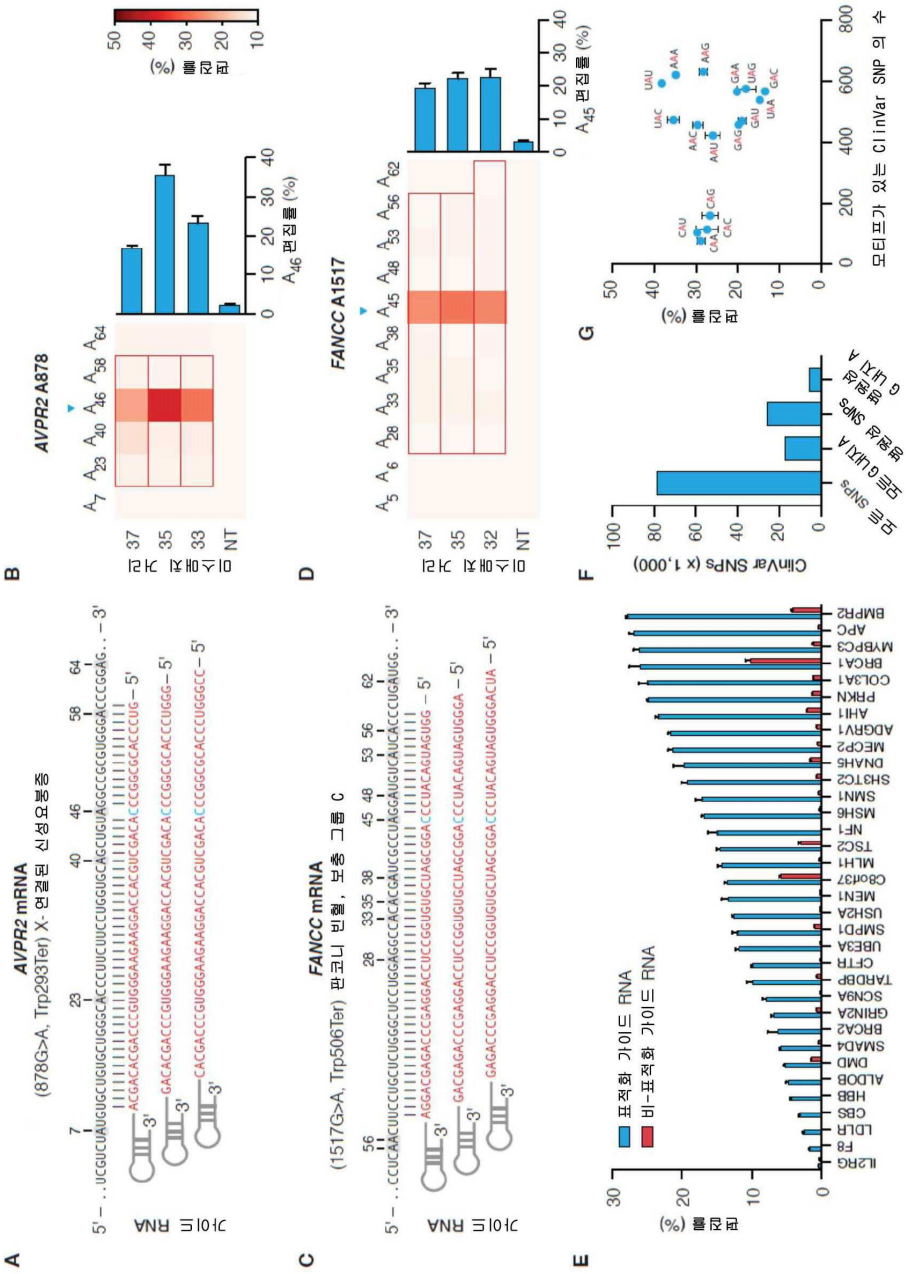
도면19


A
ADAR
dCas13b
가이드 RNA
5'
3'
가이드 RNA 상의
미스매치 C 에 의해
명시된 표적 A
표적 A 내지
I 의 전환
B
표적 mRNA
(CLuc W85X)
가이드 RNA
미스매치 거리
스페이서 길이
C
스페이서 길이
30 nt
50 nt
70 nt
84 nt
dCas13b-dADAR1
dCas13b-dADAR2
배경 제외 RLU
미스매치 거리
D
CLuc W85X mRNA
E
CLuc W85X 돌연변이 (A45)
A45 편집률 (%)
미스매치 거리
DRE 편집률 (%)
NT

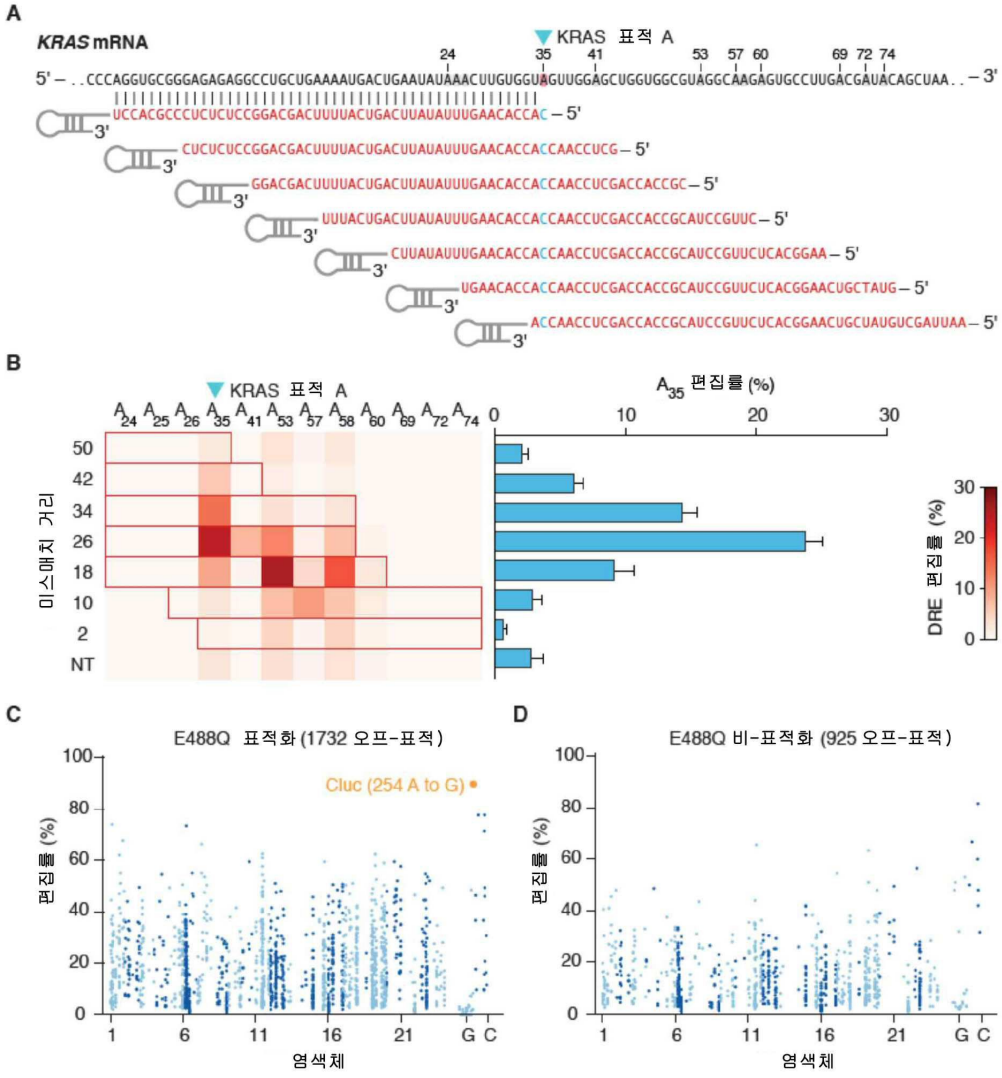
도면20


A
Cypridina 루시페라제 mRNA
랜덤화
PFS
Cas13 표적 부위
편집 부위
5'—NNNNGACUACAG...UCCCGAAACUAG—3'
심층 서열분석에 의해 ADAR 편집 mRNA 판독
편집 판독 5'—NNNN—UGG—3'
or
미편집 판독 5'—NNNN—UAG—3'
편집 판독과 연관된 모티프에 기반하여 PFS 선호도 확인
B
편집률 (%)
100
75
50
25
0
UAG
AAC
편집 부위 모티프
C
편집률 (%)
40
30
20
10
0
표적화 가이드 RNA
비-표적화 가이드 RNA
GAC UAA UAG GAU GAG GAA AAU CAG CAC AAG CAA AAC CAU AAA UAC UAU
편집 부위 모티프
D
모티프 5' 베이스
U
C
G
A
U C G A
모티프 3' 베이스
15 25 35
편집률 (%)

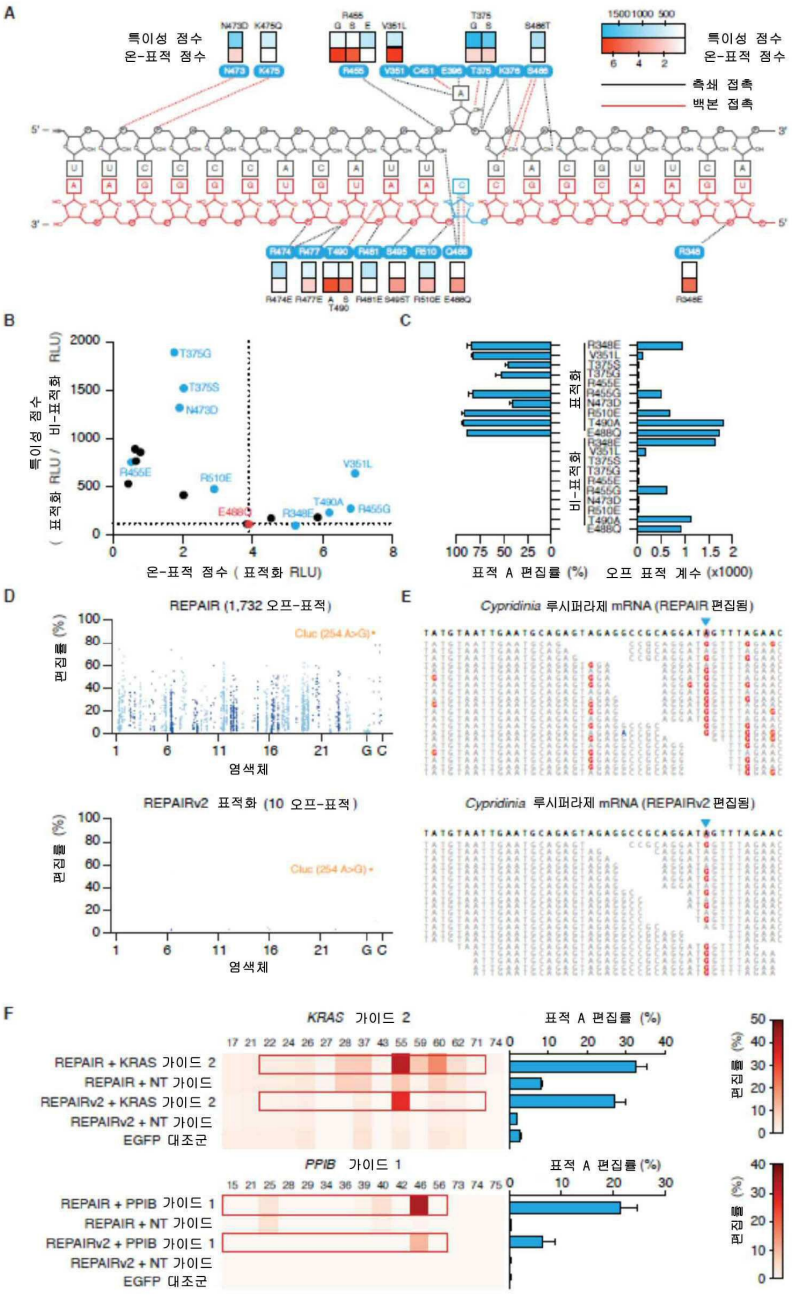
도면21


A
AVPR2 mRNA
(878G>A, Trp293Ter) X-연결된 신성요붕증
가이드 RNA
B
AVPR2 A878
편집 효율 (%)
A46 편집률 (%)
간 폐 비장 신장 NT
C
FANCC mRNA
(1517G>A, Trp506Ter) 판코니 빈혈, 보충 그룹 C
가이드 RNA
D
FANCC A1517
A45 편집률 (%)
간 폐 비장 신장 NT
E
표적화 가이드 RNA
비-표적화 가이드 RNA
편집률 (%)
IL2RG
F8
LDLR
CBS
HBB
ALDOB
DMD
SMAD4
BRCA2
GRIN2A
SCN9A
TARDBP
CFTR
UBE3A
SMPD1
USH2A
MEN1
C9orf37
MLH1
TSC2
NF1
MSH6
SMN1
SH3TC2
DNAH5
MECP2
ADGRV1
AHI1
PRKN
COL3A1
BRCA1
MYBPC3
APC
BMPR2
F
ClinVar SNPs (x 1,000)
모든 SNPs
병원성 SNPs
G 넉지 A
병원성 SNPs
G 넉지 A
G
편집률 (%)
모티프가 있는 ClinVar SNP 의 수

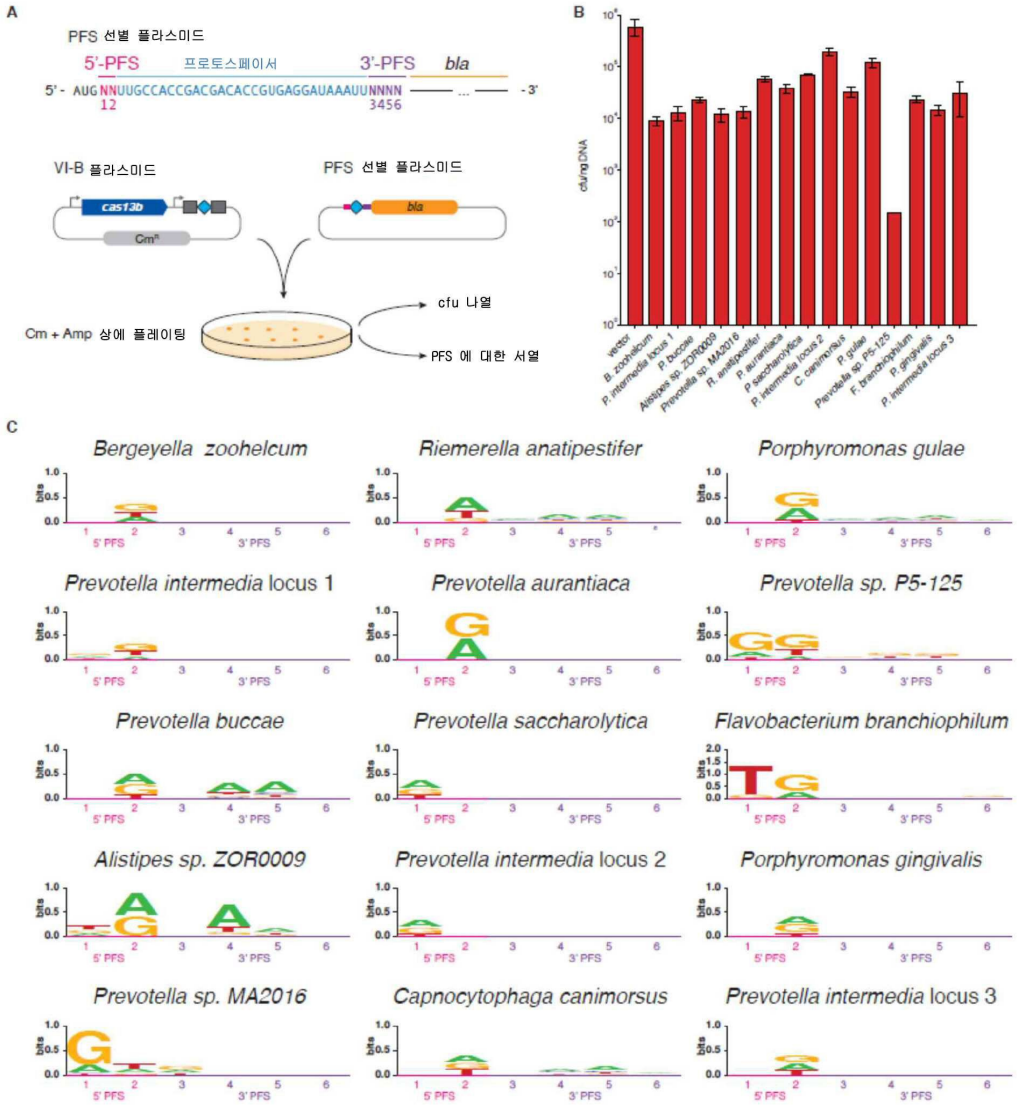
도면22


A
KRAS mRNA
KRAS 표적 A
B
KRAS 표적 A
A35 편집률 (%)
미스매치 거리
DRE 편집률 (%)
C
E488Q 표적화 (1732 오프-표적)
Cluc (254 A to G)
편집률 (%)
염색체
D
E488Q 비-표적화 (925 오프-표적)
편집률 (%)
염색체

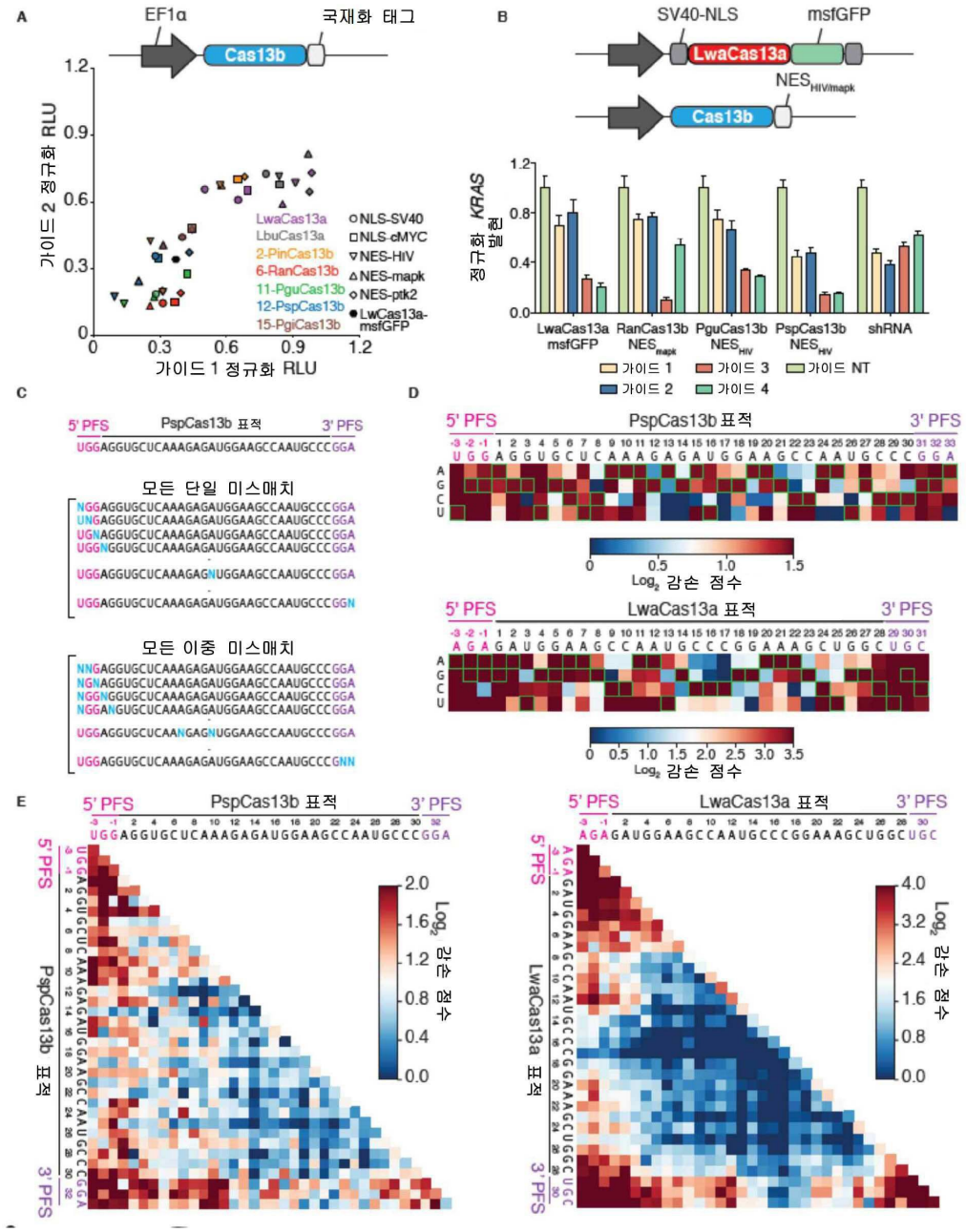
도면23


A
특이성 점수
온-표적 점수
측쇄 접촉
백본 접촉
B
특이성 점수 (표적화 RLU / 비-표적화 RLU)
온-표적 점수 (표적화 RLU)
C
표적화
비-표적화
표적 A 편집률 (%)
오프 표적 계수 (x1000)
D
REPAIR (1,732 오프-표적)
REPAIRv2 표적화 (10 오프-표적)
편집률 (%)
염색체
E
Cypridinia 루시퍼라제 mRNA (REPAIR 편집됨)
Cypridinia 루시퍼라제 mRNA (REPAIRv2 편집됨)
F
KRAS 가이드 2
PPIB 가이드 1
표적 A 편집률 (%)
REPAIR + KRAS 가이드 2
REPAIR + NT 가이드
REPAIRv2 + KRAS 가이드 2
REPAIRv2 + NT 가이드
EGFP 대조군
REPAIR + PPIB 가이드 1
REPAIRv2 + PPIB 가이드 1

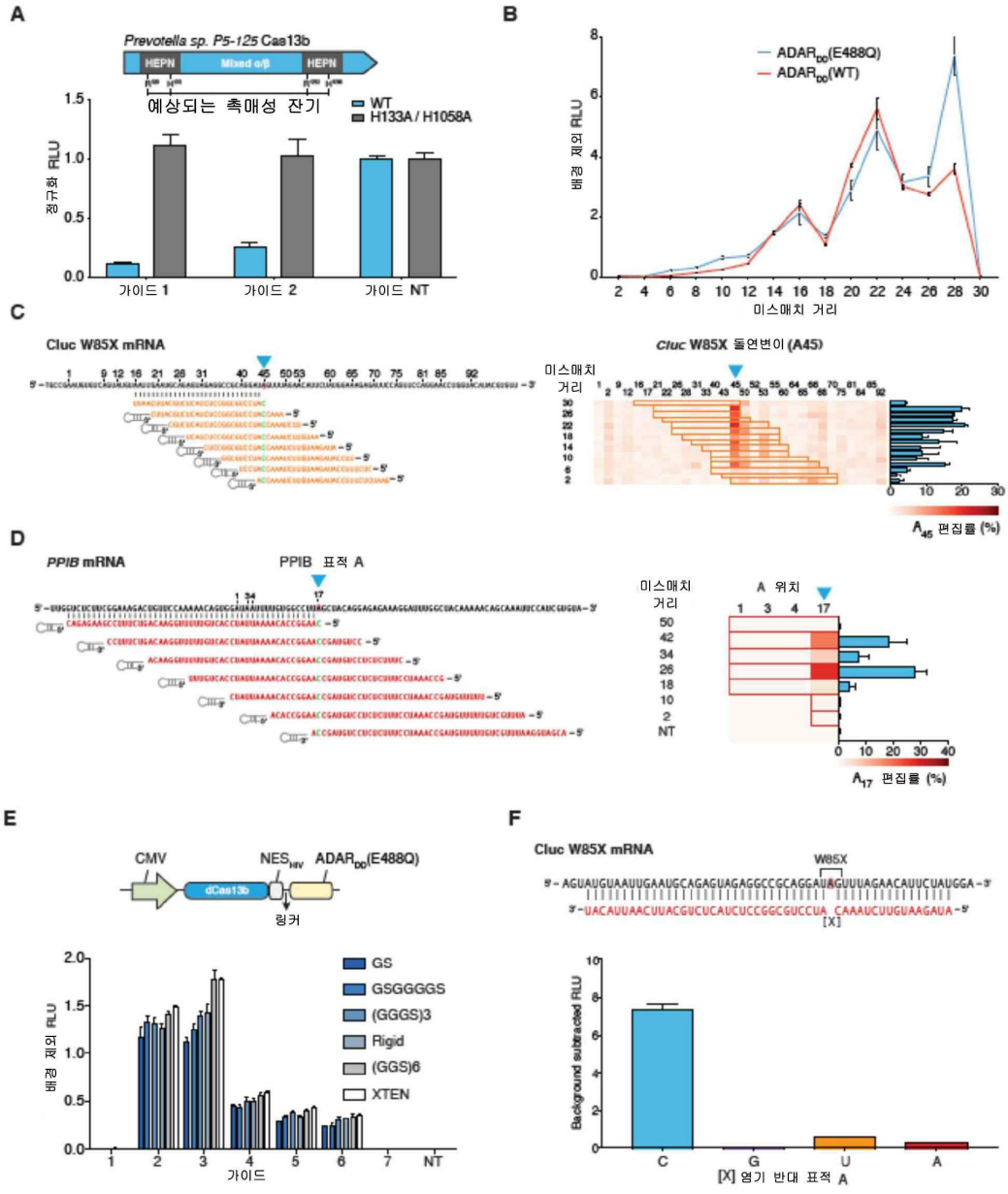
도면24


A
PFS 선별 플라스미드
5'-PFS
프로토스페이서
3'-PFS
bla
5' - AUG NNUUGCCACCGACGACACCGUGAGGAUAAAUUNNNN - 3'
12
3456
VI-B 플라스미드
PFS 선별 플라스미드
cas13b
Cm^R
bla
Cm + Amp 상에 플레이팅
cfu 나열
PFS 에 대한 서열
B
cfu/ng DNA
vector
B. zoohelcum
P. intermedia locus 1
P. buccae
Alistipes sp. ZOR0009
Prevotella sp. MA2016
R. anatipestifer
P. aurantiaca
P. saccharolytica
P. intermedia locus 2
C. canimorsus
P. gulae
Prevotella sp. P5-125
F. branchiophilum
P. gingivalis
P. intermedia locus 3
C
Bergeyella zoohelcum
Riemerella anatipestifer
Porphyromonas gulae
Prevotella intermedia locus 1
Prevotella aurantiaca
Prevotella sp. P5-125
Prevotella buccae
Prevotella saccharolytica
Flavobacterium branchiophilum
Alistipes sp. ZOR0009
Prevotella intermedia locus 2
Porphyromonas gingivalis
Prevotella sp. MA2016
Capnocytophaga canimorsus
Prevotella intermedia locus 3
bits
5' PFS
3' PFS

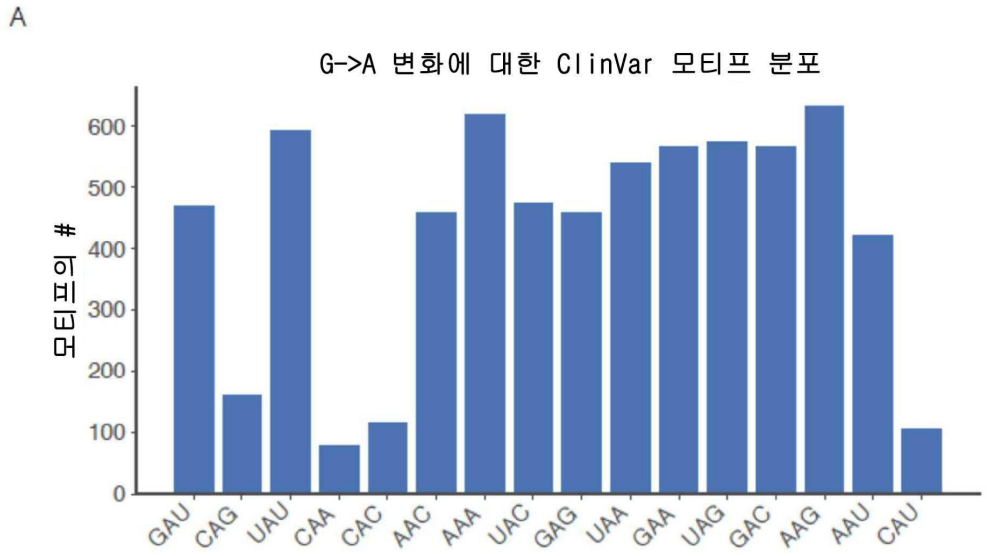
도면25


A
EF1α
국재화 태그
Cas13b
가이드 2 정규화 RLU
가이드 1 정규화 RLU
LwaCas13a
LbuCas13a
2-PinCas13b
6-RanCas13b
11-PguCas13b
12-PspCas13b
15-PgiCas13b
NLS-SV40
NLS-cMYC
NES-HIV
NES-mapk
NES-ptk2
LwaCas13a-msfGFP
B
SV40-NLS
msfGFP
LwaCas13a
NES HIV/mapk
Cas13b
정규화된 KRAS
LwaCas13a msfGFP
RanCas13b NES mapk
PguCas13b NES HIV
PspCas13b NES HIV
shRNA
가이드 1
가이드 2
가이드 3
가이드 4
가이드 NT
C
5' PFS
PspCas13b 표적
3' PFS
모든 단일 미스매치
모든 이중 미스매치
D
PspCas13b 표적
LwaCas13a 표적
Log2 감손 점수
E
PspCas13b 표적
LwaCas13a 표적

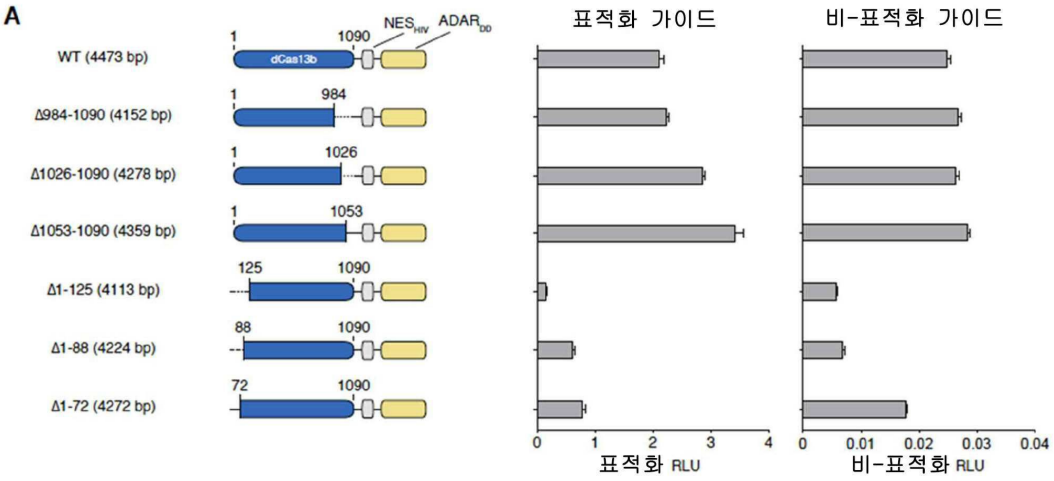
도면26

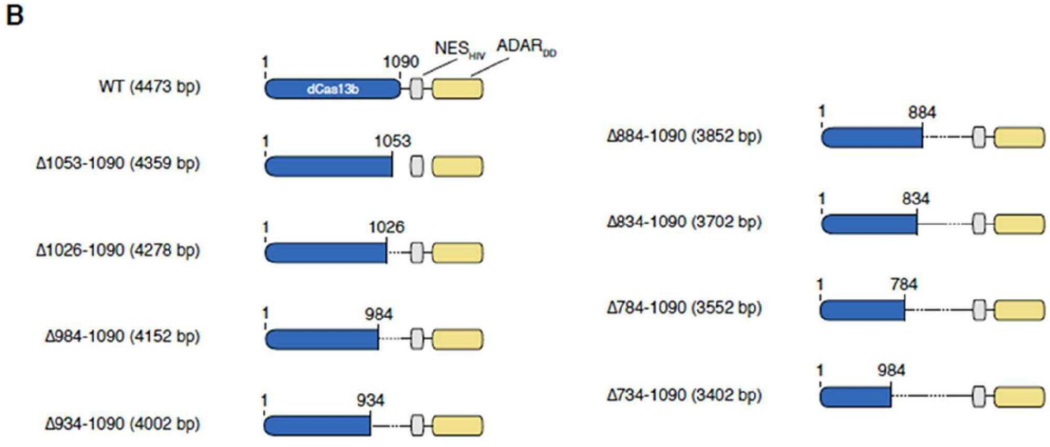

A
Prevotella sp. P5-125 Cas13b
HEPN
Mixed α/β
HEPN
예상되는 촉매성 잔기
WT
H133A / H1058A
정규화 RLU
가이드 1
가이드 2
가이드 NT
B
ADARDD(E488Q)
ADARDD(WT)
배경 제외 RLU
미스매치 거리
C
Cluc W85X mRNA
Cluc W85X 돌연변이 (A45)
미스매치 거리
A45 편집률 (%)
D
PPIB mRNA
PPIB 표적 A
미스매치 거리
A 위치
NT
A17 편집률 (%)
E
CMV
NESHIV
ADARDD(E488Q)
dCas13b
링커
GS
GSGGGGS
(GGGS)3
Rigid
(GGS)6
XTEN
배경 제외 RLU
가이드
F
Cluc W85X mRNA
W85X
[X]
Background subtracted RLU
C
G
U
A
[X] 염기 반대 표적 A

도면27

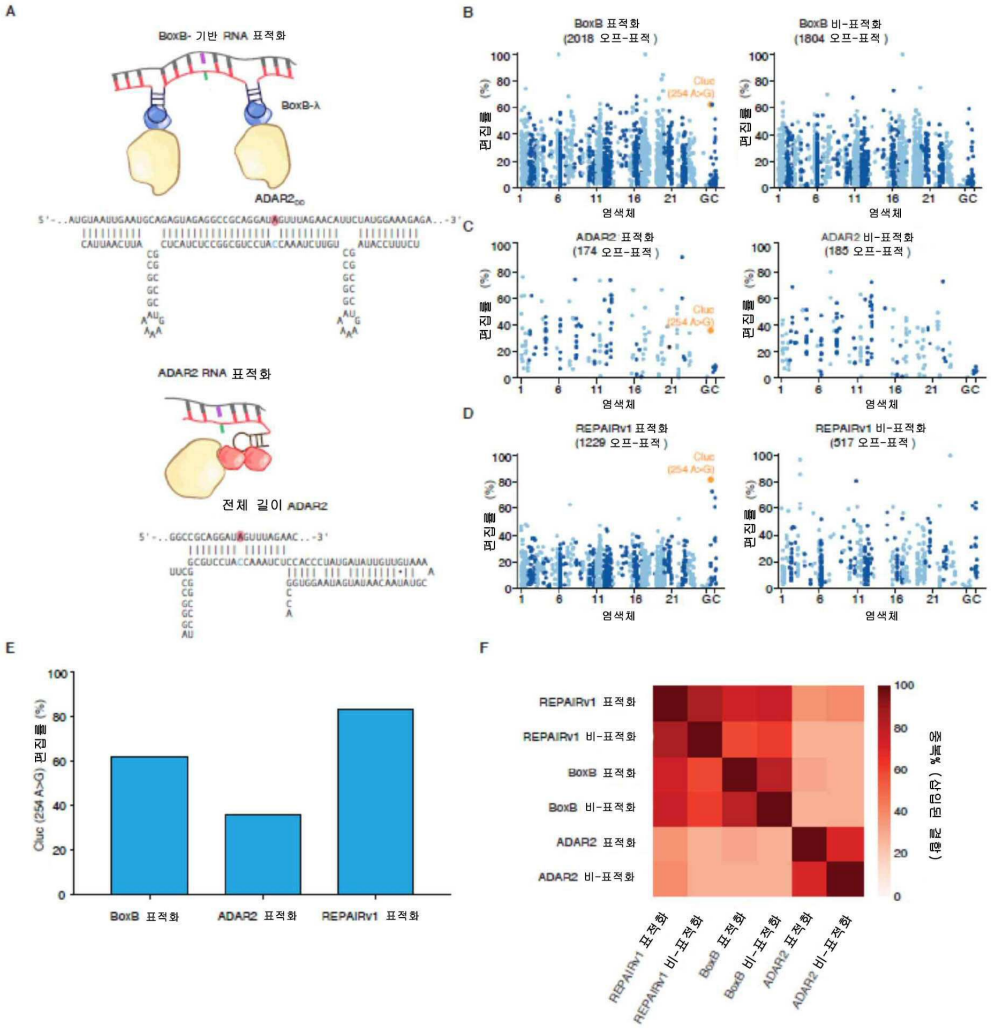

A
G->A 변화에 대한 ClinVar 모티프 분포
모티프의 #
0
100
200
300
400
500
600
GAU
CAG
UAU
CAA
CAC
AAC
AAA
UAC
GAG
UAA
GAA
UAG
GAC
AAG
AAU
CAU

도면28

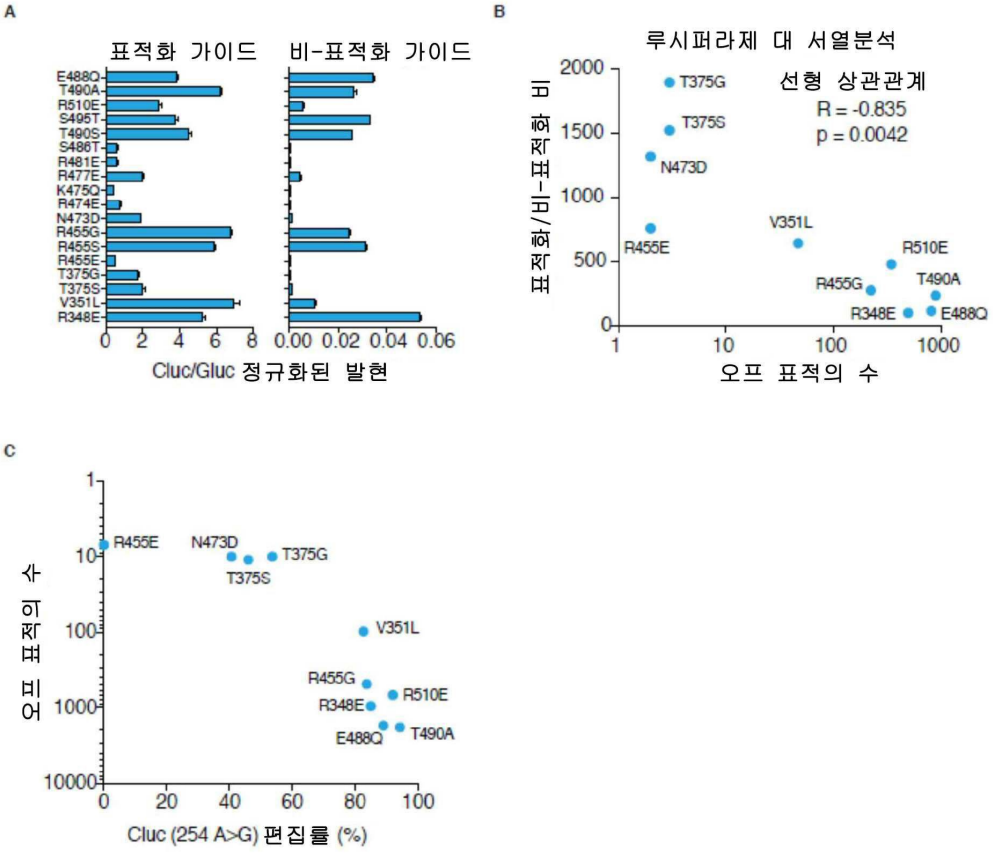

A
WT (4473 bp)
Δ984-1090 (4152 bp)
Δ1026-1090 (4278 bp)
Δ1053-1090 (4359 bp)
Δ1-125 (4113 bp)
Δ1-88 (4224 bp)
Δ1-72 (4272 bp)
dCas13b
NES HIV
ADAR DD
표적화 가이드
비-표적화 가이드
표적화 RLU
비-표적화 RLU


B
WT (4473 bp)
Δ1053-1090 (4359 bp)
Δ1026-1090 (4278 bp)
Δ984-1090 (4152 bp)
Δ934-1090 (4002 bp)
Δ884-1090 (3852 bp)
Δ834-1090 (3702 bp)
Δ784-1090 (3552 bp)
Δ734-1090 (3402 bp)
dCas13b
NES HIV
ADAR DD

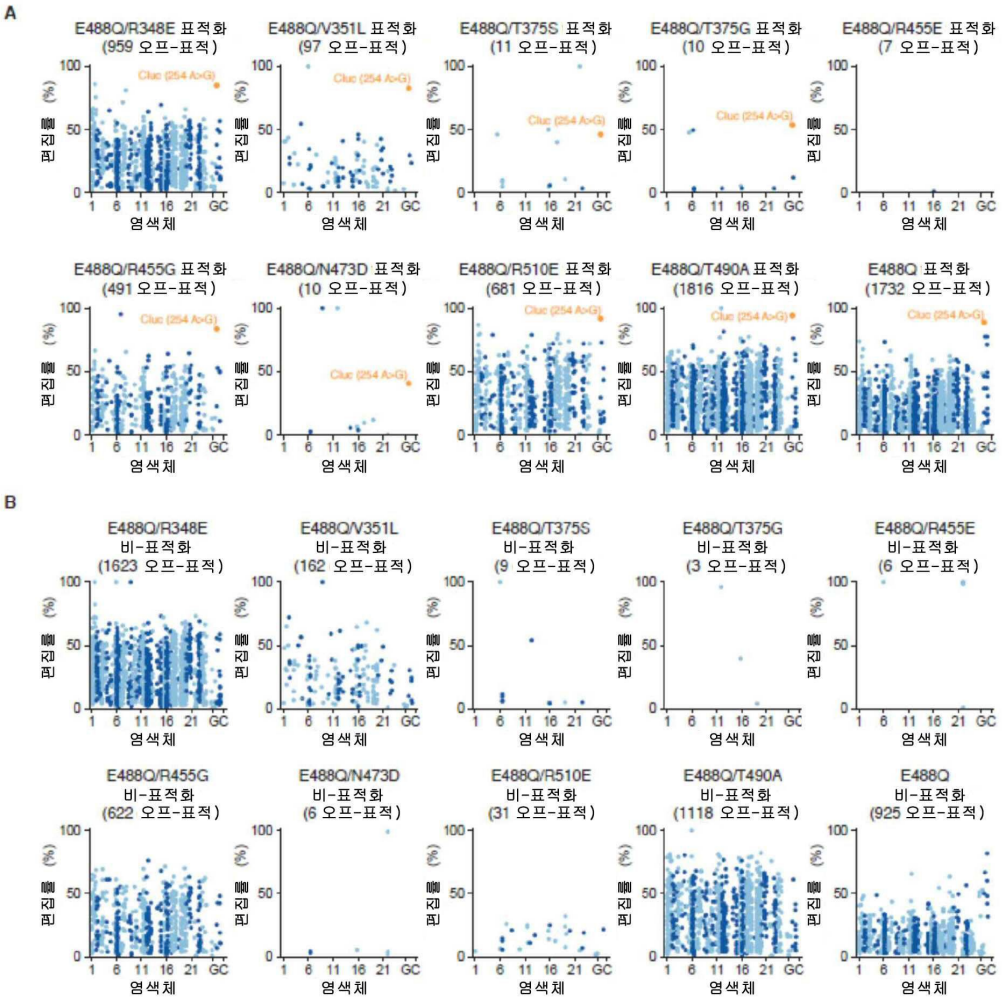
도면29


A
BoxB- 기반 RNA 표적화
BoxB-λ
ADAR2$_{DD}$
ADAR2 RNA 표적화
전체 길이 ADAR2
B
BoxB 표적화
(2018 오프-표적)
BoxB 비-표적화
(1804 오프-표적)
C
ADAR2 표적화
(174 오프-표적)
ADAR2 비-표적화
(185 오프-표적)
D
REPAIRv1 표적화
(1229 오프-표적)
REPAIRv1 비-표적화
(517 오프-표적)
편집률 (%)
염색체
Cluc (254 A>G)
E
Cluc (254 A>G) 편집률 (%)
BoxB 표적화
ADAR2 표적화
REPAIRv1 표적화
F
REPAIRv1 표적화
REPAIRv1 비-표적화
BoxB 표적화
BoxB 비-표적화
ADAR2 표적화
ADAR2 비-표적화
상관계수 (Pearson correlation)

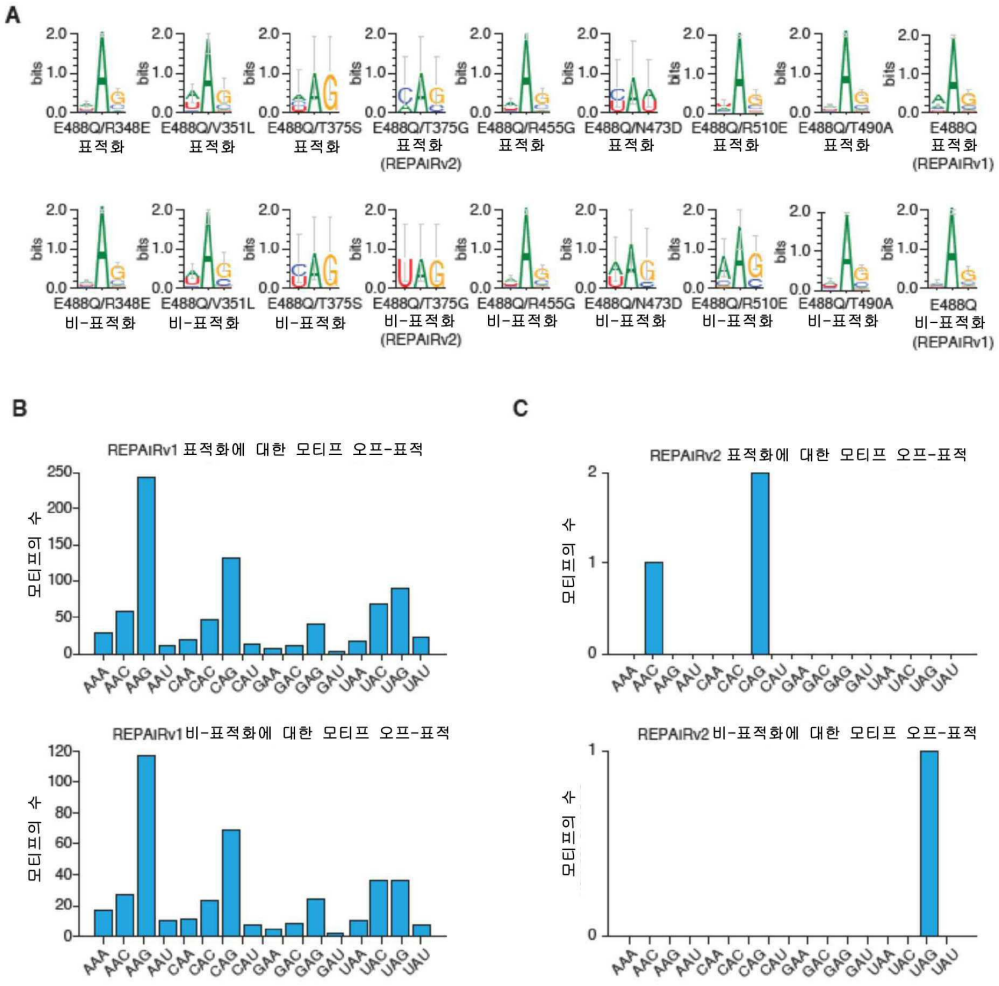
도면30


A
표적화 가이드
비-표적화 가이드
E488Q
T490A
R510E
S495T
T490S
S486T
R481E
R477E
K475Q
R474E
N473D
R455G
R455S
R455E
T375G
T375S
V351L
R348E
0 2 4 6 8
0.00 0.02 0.04 0.06
Cluc/Gluc 정규화된 발현
B
루시퍼라제 대 서열분석
선형 상관관계
R = -0.835
p = 0.0042
표적화/비-표적화 비
2000
1500
1000
500
0
T375G
T375S
N473D
R455E
V351L
R510E
R455G
T490A
R348E
E488Q
1
10
100
1000
오프 표적의 수
C
오프 표적의 수
1
10
100
1000
10000
R455E
N473D
T375G
T375S
V351L
R455G
R348E
R510E
E488Q
T490A
0 20 40 60 80 100
Cluc (254 A>G) 편집률 (%)

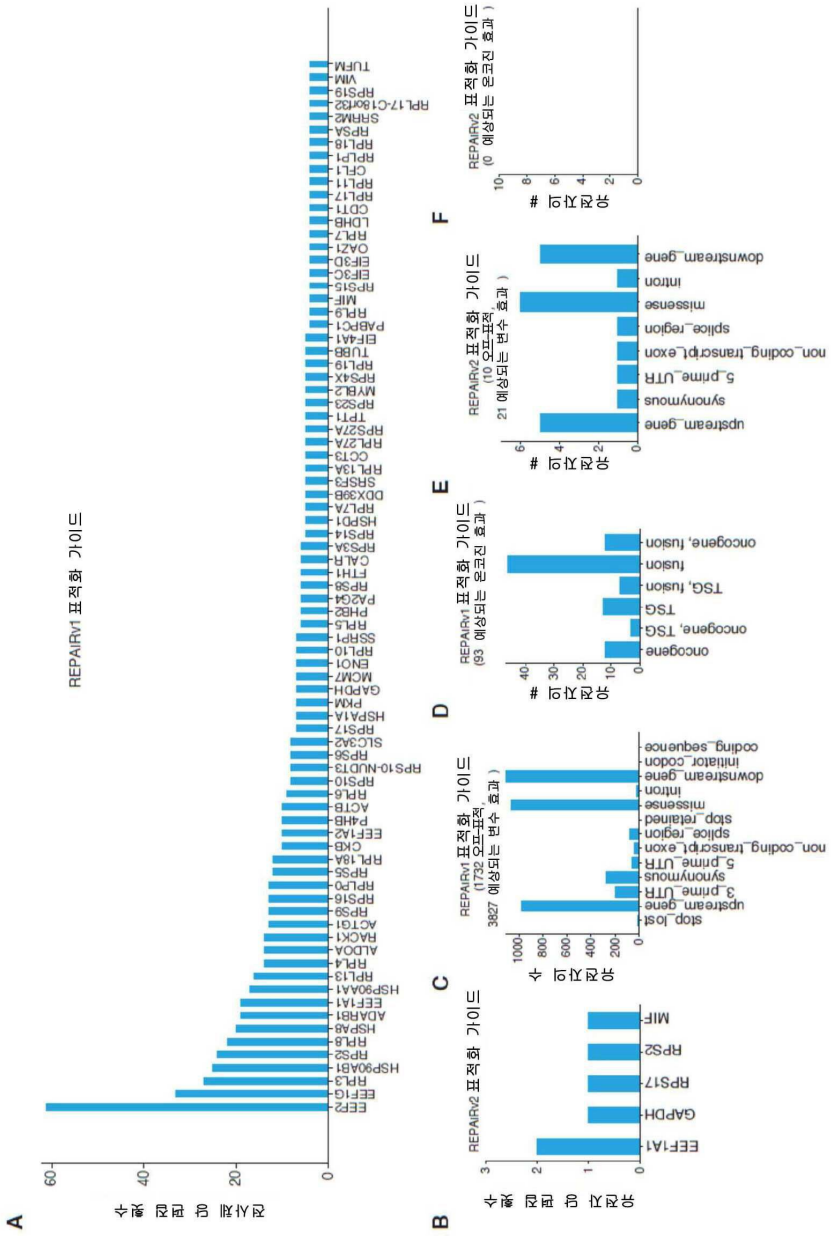
도면31


A
E488Q/R348E 표적화
(959 오프-표적)
E488Q/V351L 표적화
(97 오프-표적)
E488Q/T375S 표적화
(11 오프-표적)
E488Q/T375G 표적화
(10 오프-표적)
E488Q/R455E 표적화
(7 오프-표적)
E488Q/R455G 표적화
(491 오프-표적)
E488Q/N473D 표적화
(10 오프-표적)
E488Q/R510E 표적화
(681 오프-표적)
E488Q/T490A 표적화
(1816 오프-표적)
E488Q 표적화
(1732 오프-표적)
B
E488Q/R348E
비-표적화
(1623 오프-표적)
E488Q/V351L
비-표적화
(162 오프-표적)
E488Q/T375S
비-표적화
(9 오프-표적)
E488Q/T375G
비-표적화
(3 오프-표적)
E488Q/R455E
비-표적화
(6 오프-표적)
E488Q/R455G
비-표적화
(622 오프-표적)
E488Q/N473D
비-표적화
(6 오프-표적)
E488Q/R510E
비-표적화
(31 오프-표적)
E488Q/T490A
비-표적화
(1118 오프-표적)
E488Q
비-표적화
(925 오프-표적)
Cluc (254 A>G)
편집률 (%)
염색체
1 6 11 16 21 GC

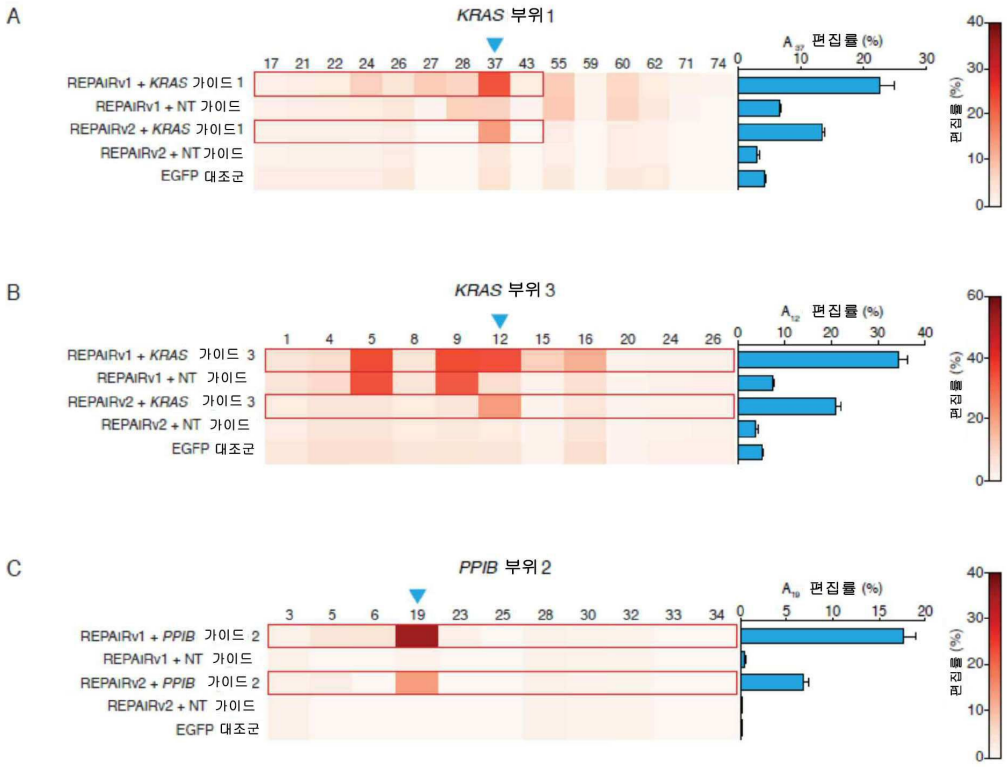
도면32


A
E488Q/R348E 표적화
E488Q/V351L 표적화
E488Q/T375S 표적화
E488Q/T375G 표적화 (REPAIRv2)
E488Q/R455G 표적화
E488Q/N473D 표적화
E488Q/R510E 표적화
E488Q/T490A 표적화
E488Q 표적화 (REPAIRv1)
E488Q/R348E 비-표적화
E488Q/V351L 비-표적화
E488Q/T375S 비-표적화
E488Q/T375G 비-표적화 (REPAIRv2)
E488Q/R455G 비-표적화
E488Q/N473D 비-표적화
E488Q/R510E 비-표적화
E488Q/T490A 비-표적화
E488Q 비-표적화 (REPAIRv1)
bits
B
REPAIRv1 표적화에 대한 모티프 오프-표적
REPAIRv1 비-표적화에 대한 모티프 오프-표적
C
REPAIRv2 표적화에 대한 모티프 오프-표적
REPAIRv2 비-표적화에 대한 모티프 오프-표적
모티프의 수
AAA AAC AAG AAU CAA CAC CAG CAU GAA GAC GAG GAU UAA UAC UAG UAU

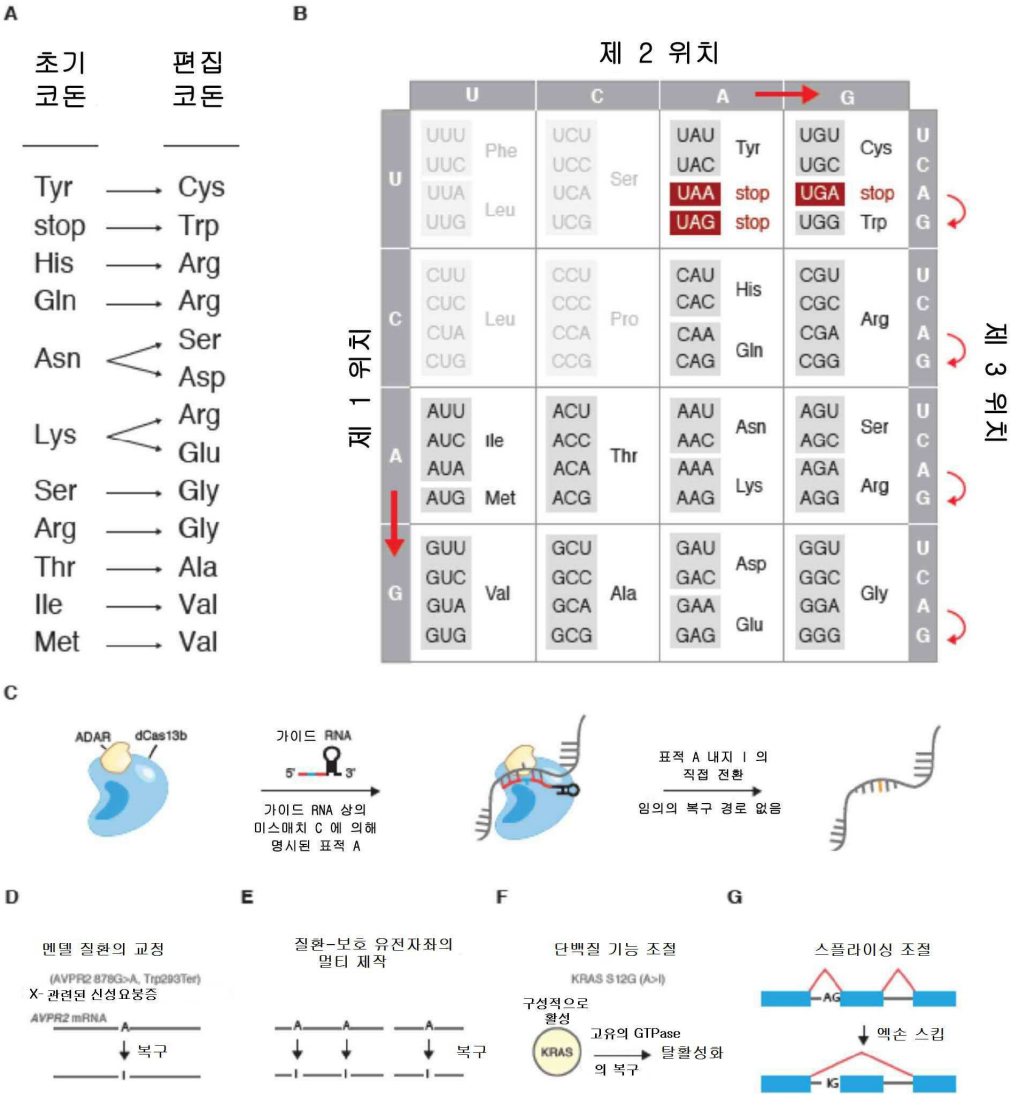
도면33


A
REPAIRv1 표적화 가이드
B
REPAIRv2 표적화 가이드
C
REPAIRv1 표적화 가이드
(1732 오프표적 예상되는 변수 효과)
D
REPAIRv1 표적화 가이드
(93 예상되는 온코진 효과)
E
REPAIRv2 표적화 가이드
(10 오프표적 예상되는 변수 효과)
21
F
REPAIRv2 표적화 가이드
(0 예상되는 온코진 효과)

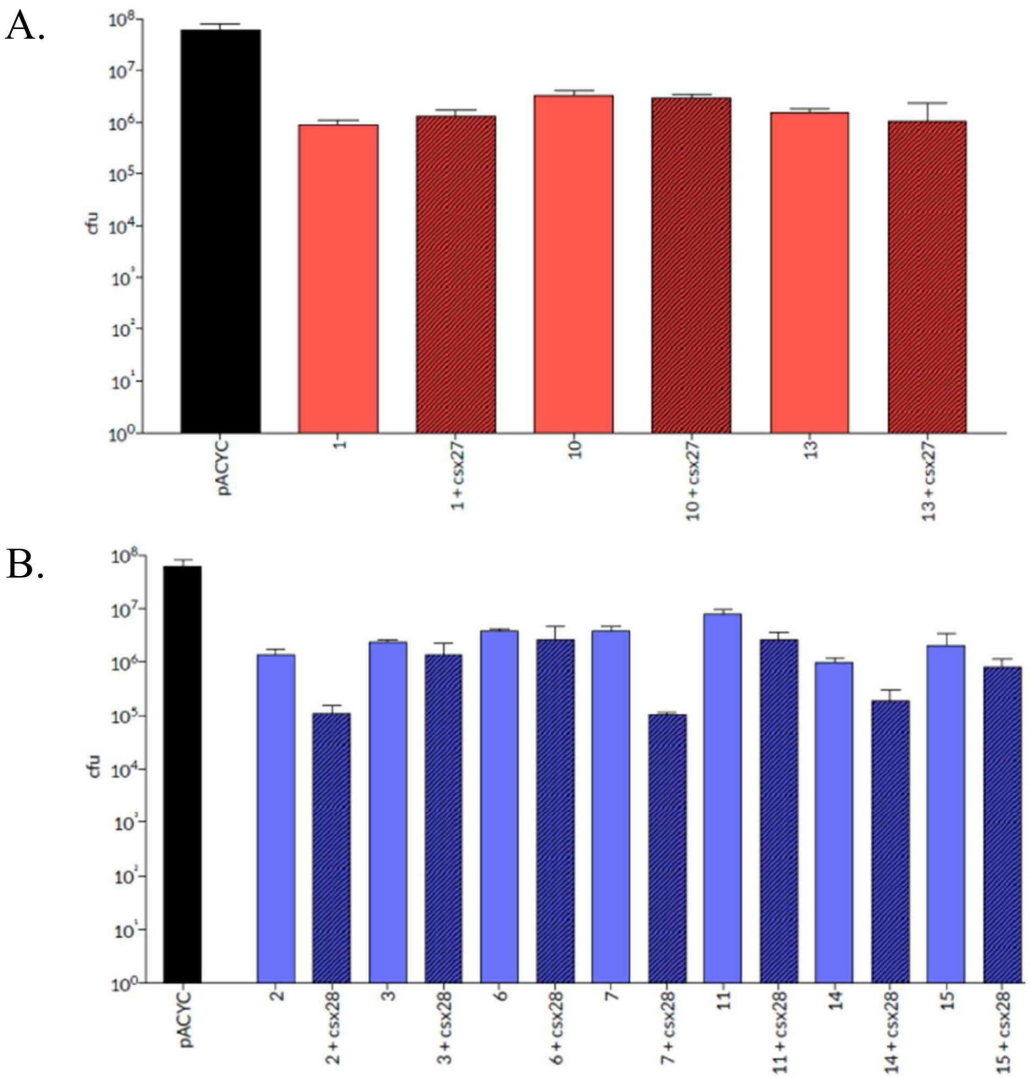
도면34


A
KRAS 부위 1
17 21 22 24 26 27 28 37 43 55 59 60 62 71 74
REPAIRv1 + KRAS 가이드 1
REPAIRv1 + NT 가이드
REPAIRv2 + KRAS 가이드 1
REPAIRv2 + NT 가이드
EGFP 대조군
편집률 (%)
0 10 20 30
편집률 (%)
0 10 20 30 40
B
KRAS 부위 3
1 4 5 8 9 12 15 16 20 24 26
REPAIRv1 + KRAS 가이드 3
REPAIRv1 + NT 가이드
REPAIRv2 + KRAS 가이드 3
REPAIRv2 + NT 가이드
EGFP 대조군
편집률 (%)
0 10 20 30 40
편집률 (%)
0 20 40 60
C
PPIB 부위 2
3 5 6 19 23 25 28 30 32 33 34
REPAIRv1 + PPIB 가이드 2
REPAIRv1 + NT 가이드
REPAIRv2 + PPIB 가이드 2
REPAIRv2 + NT 가이드
EGFP 대조군
편집률 (%)
0 5 10 15 20
편집률 (%)
0 10 20 30 40

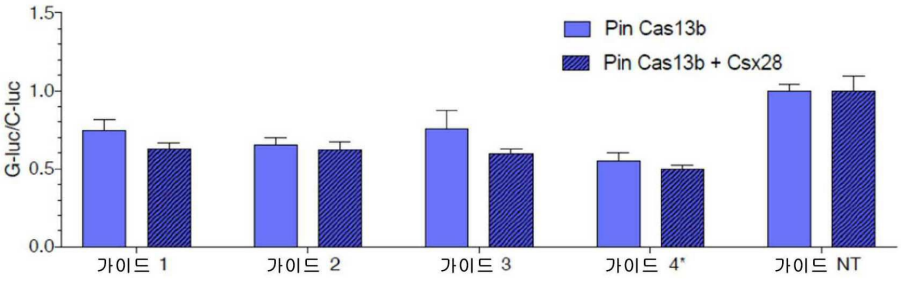
도면35


A
초기 코돈
편집 코돈
Tyr → Cys
stop → Trp
His → Arg
Gln → Arg
Asn → Ser / Asp
Lys → Arg / Glu
Ser → Gly
Arg → Gly
Thr → Ala
Ile → Val
Met → Val
B
제 2 위치
제 1 위치
제 3 위치
U C A G
UUU Phe
UUC Phe
UUA Leu
UUG Leu
UCU Ser
UCC Ser
UCA Ser
UCG Ser
UAU Tyr
UAC Tyr
UAA stop
UAG stop
UGU Cys
UGC Cys
UGA stop
UGG Trp
CUU Leu
CUC Leu
CUA Leu
CUG Leu
CCU Pro
CCC Pro
CCA Pro
CCG Pro
CAU His
CAC His
CAA Gln
CAG Gln
CGU Arg
CGC Arg
CGA Arg
CGG Arg
AUU Ile
AUC Ile
AUA Ile
AUG Met
ACU Thr
ACC Thr
ACA Thr
ACG Thr
AAU Asn
AAC Asn
AAA Lys
AAG Lys
AGU Ser
AGC Ser
AGA Arg
AGG Arg
GUU Val
GUC Val
GUA Val
GUG Val
GCU Ala
GCC Ala
GCA Ala
GCG Ala
GAU Asp
GAC Asp
GAA Glu
GAG Glu
GGU Gly
GGC Gly
GGA Gly
GGG Gly
C
ADAR
dCas13b
가이드 RNA
5'
3'
가이드 RNA 상의 미스매치 C 에 의해 명시된 표적 A
표적 A 내지 I 의 직접 전환
임의의 복구 경로 없음
D
멘델 질환의 교정
(AVPR2 878G>A, Trp293Ter)
X-관련된 신성요붕증
AVPR2 mRNA
복구
E
질환-보호 유전자좌의 멀티 제작
복구
F
단백질 기능 조절
KRAS S12G (A>I)
구성적으로 활성
KRAS
고유의 GTPase 의 복구
탈활성화
G
스플라이싱 조절
AG
엑손 스킵
IG

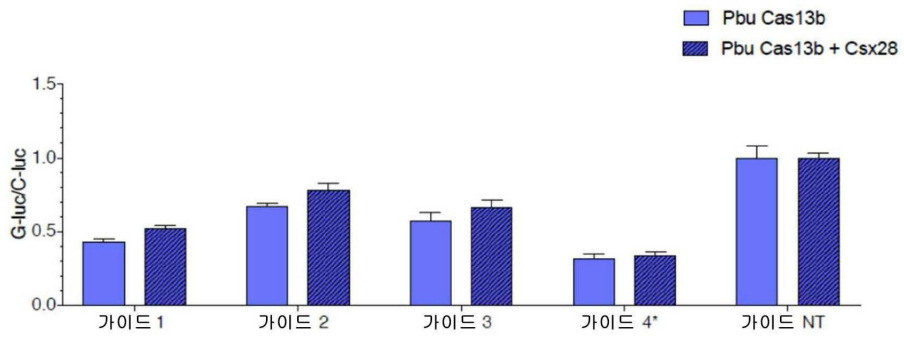
도면36


A.
cfu
pACYC
1
1 + csx27
10
10 + csx27
13
13 + csx27
B.
cfu
pACYC
2
2 + csx28
3
3 + csx28
6
6 + csx28
7
7 + csx28
11
11 + csx28
14
14 + csx28
15
15 + csx28

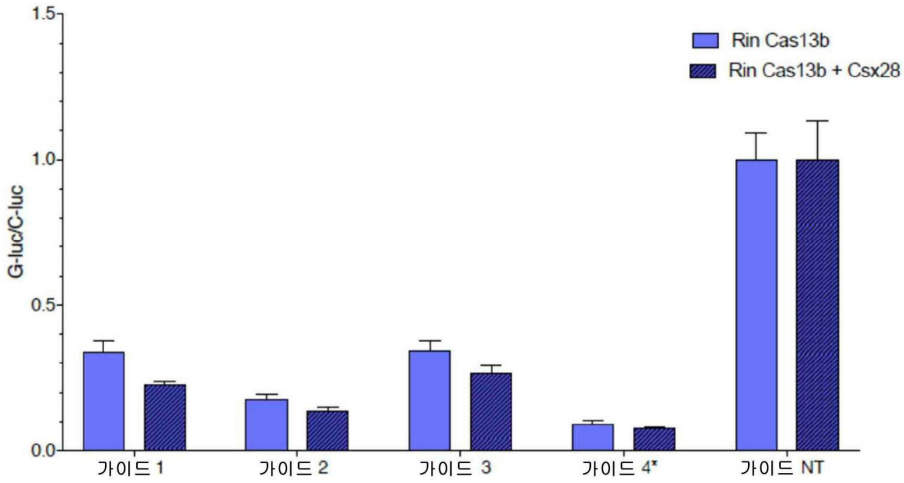
도면37a


Pin Cas13b
Pin Cas13b + Csx28
G-luc/C-luc
가이드 1
가이드 2
가이드 3
가이드 4*
가이드 NT

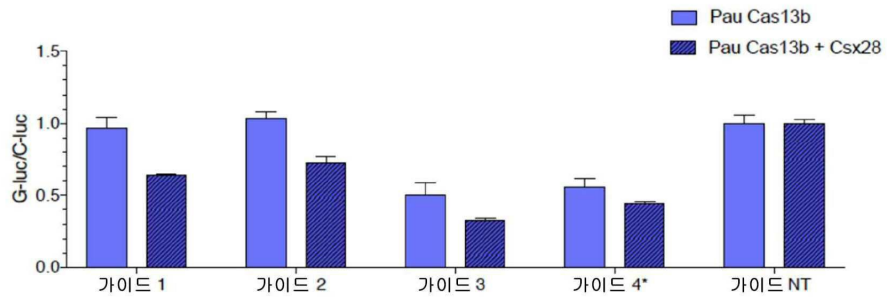
도면37b



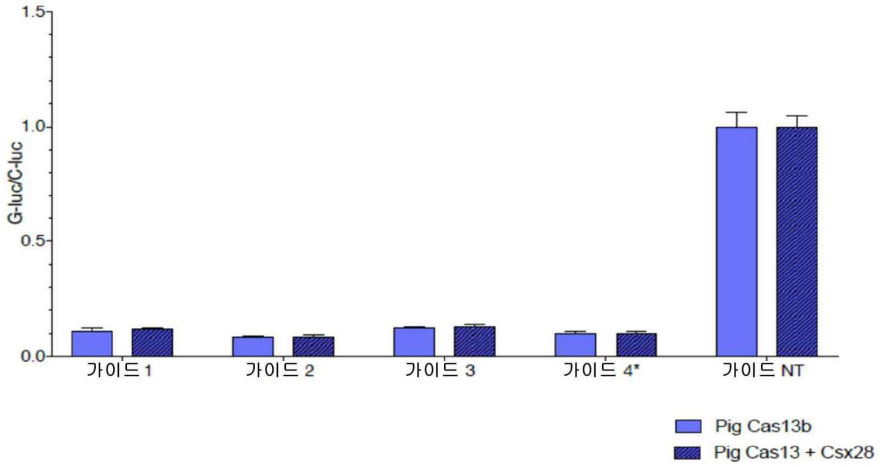
도면37c



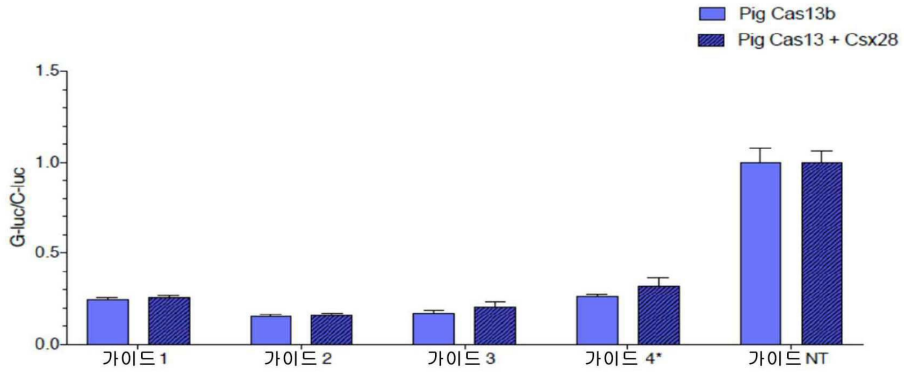
도면37d



도면37e



도면37f



서 열 목 록

<110> The BROAD Institute, Inc.

Massachusetts Institute of Technology

Cox, David Benjamin Turitz

Pyzocha, Neena

Zhang, Feng

<120> Novel CAS13B Orthologues CRISPR Enzymes and Systems

<130> BROD-2780WP (BI-10157) (10157-47626-99)

<140> PCT/US2018/022751

<141> 2018-03-15

<150> US 62/471,710

<151> 2017-03-15

<150> US 62/566829

<151> 2017-10-02

<160> 546

<170> PatentIn version 3.5

<210> 1

<211> 7

<212> PRT

<213> Simian virus 40

<400> 1

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 2

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 2

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 3

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 3

Pro Ala Ala Lys Arg Val Lys Leu Asp

1 5

<210> 4

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 4

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro

1 5 10

<210> 5

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 5

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
20 25 30

Arg Asn Gln Gly Gly Tyr
35

<210> 6

<211> 42

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 6

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
35 40

<210> 7

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 7

Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> 8

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 8

Pro Pro Lys Lys Ala Arg Glu Asp

1 5

<210> 9

<211> 8

<212> PRT

<213> Homo sapiens

<220><221> MISC_FEATURE

<222> (2)

<223> Xaa can be any naturally occurring amino acid

<400

> 9

Pro Xaa Pro Lys Lys Lys Pro Leu

1 5

<210> 10

<211> 12

<212> PRT

<213> Mus

<400> 10

Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro

1 5 10

<210> 11

<211> 5

<212> PRT

<213> Influenza NS1

<400> 11

Asp Arg Leu Arg Arg

1 5

<210> 12

<211> 7

<212> PRT

<213> Influenza NS1

<400> 12

Pro Lys Gln Lys Lys Arg Lys

1 5

<210> 13

<211> 10

<212> PRT

<213> Hepatitis virus

<400> 13

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu

1 5 10

<210> 14

<211> 10

<212> PRT

<213> Mus

<400> 14

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg

1 5 10

<210> 15

<211> 20

<212> PRT

<213> Homo sapiens

<400> 15

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys

1 5 10 15

Lys Ser Lys Lys

20

<210> 16

<211> 17

<212> PRT

<213> Homo sapiens

<400> 16

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

1 5 10 15

Lys

<210> 17

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 17

Phe Trp Tyr His Lys Met Ile Leu Val Ala Gly Cys

1 5 10

<210> 18

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 18

Trp Tyr His Lys Arg Glu Asp Cys Ser Thr Asn Gln

1 5 10

<210> 19

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 19

Val Cys Ala Gly Ser Pro Thr Asn Asp

1 5

<210> 20

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 20

Phe Trp Tyr His

1

<210
> 21
<211> 5
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 21
His Lys Arg Glu Asp
1 5
<210> 22
<211> 15
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<220><221> MISC_FEATURE
<222> (4)
<223> Xaa can be any naturally occurring amino acid
<400> 22
Asp Lys His Xaa Phe Gly Ala Phe Leu Asn Leu Ala Arg His Asn
1 5 10 15
<210> 23
<211> 12

<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<400> 23
Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Asp Lys
1 5 10
<210> 24
<211> 7
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic
<220><221> MISC_FEATURE

<222> (4)

<223> Xaa can be any naturally occurring amino acid

<400> 24

Ser Lys Ile Xaa Gly Phe Lys

1 5

<210> 25

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> MISC_FEATURE

<222> (7)

<223> Xaa can be any naturally occurring amino acid

<400> 25

Asp Met Leu Asn Glu Leu Xaa Arg Cys Pro

1 5 10

<210> 26

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> MISC_FEATURE

<222> (2)

<223> Xaa can be any naturally occurring amino acid

<220><221> MISC_FEATURE

<222> (14)

<223> Xaa can be any naturally occurring amino acid

<400> 26

Arg Xaa Glx Asp Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Xaa Asp

1 5 10 15

<210> 27

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> MISC_FEATURE

<222> (9)

<223> Xaa can be any naturally occurring amino acid

<400> 27

Leu Arg Phe Gln Val Asx Leu Gly Xaa Tyr
1 5 10

<210> 28

<211> 10

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 28

Gly Leu Leu Phe Phe Val Ser Leu Phe Leu
1 5 10

<210> 29

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<220><221> MISC_FEATURE

<222> (4)

<223> Xaa can be any naturally occurring amino acid

<400> 29

Arg His Gln Xaa Arg Phe Pro Tyr Phe
1 5

<210> 30

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400

> 30

Arg His Gln Asp Arg Phe Pro Tyr

1 5

<210> 31

<211> 1149

<212> PRT

<213> Sinomicrobial oceani

<400> 31

Met Glu Ser Thr Thr Thr Leu Gly Leu His Leu Lys Tyr Gln His Asp

1 5 10 15

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

20 25 30

Gln Asn Ile Glu Ser Ile Phe Gln Ala Phe Ala Glu Arg Tyr Gly Ile

35 40 45

Gln Asn Pro Leu Arg Lys Asn Gly Val Pro Ala Ile Asn Asn Ile Phe

50 55 60

His Asp Asn Ile Ser Ile Ser Asn Tyr Lys Glu Tyr Leu Lys Phe Leu

65 70 75 80

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

85 90 95

Asn Ile Phe Glu Phe Arg Glu Asp Phe Glu Ile Leu Ile Asn Ala Ile

100 105 110

Tyr Lys Leu Arg His Phe Tyr Thr His Tyr Tyr His Ser Pro Ile Lys

115 120 125

Leu Glu Asp Arg Phe Tyr Thr Cys Leu Asn Glu Leu Phe Val Ala Val

130 135 140

Ala Ile Gln Val Lys Lys His Lys Met Lys Ser Asp Lys Thr Arg Gln

145 150 155 160

Leu Leu Asn Lys Asn Leu His Gln Leu Leu Gln Gln Leu Ile Glu Gln

165 170 175

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

180 185 190

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

195 200 205

His Leu Leu Asp Lys Asp Glu Asn Ile Arg Leu Asn Tyr Ser Ser Arg

210 215 220

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

225 230 235 240

Gly Leu Leu Phe Leu Leu Ser Leu Phe Leu Gln Arg Lys Glu Ala Glu

245 250 255

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

260 265 270

Arg Phe Met Ala Thr His Trp Val Phe Ser Tyr Leu Asn Val Lys Arg

275 280 285

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

290 295 300

Gln Ile Ala Asp Glu Leu Ser Lys Val Pro Asp Glu Val Tyr Lys Thr

305 310 315 320

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

325 330 335

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

340 345 350

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

355 360 365

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

370 375 380

Val His Leu Gly Asn Tyr Ile His Asp Arg Arg Asp Lys Val Ile Asp

385 390 395 400

Gly Thr Asn Phe Ile Thr Asn Arg Val Ile Lys Glu Pro Ile Lys Val

405 410 415

Phe Gly Lys Leu Ser His Val Ser Lys Leu Lys Ser Asp Tyr Met Glu

420 425 430

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

435 440 445

Ser Tyr Asn Phe Val Gly His Asn Ile Pro Ile Phe Ile Asn Leu Arg

450 455 460

Ser Ala Ser Ser Lys Gly Lys Glu Leu Tyr Arg Asp Leu Met Lys Ile

465 470 475 480

Lys Ser Glu Lys Lys Lys Lys Ser Arg Glu Glu Gly Ile Pro Met Glu

485 490 495

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

500 505 510

Arg Asn Ile Lys Asp Asn Asn Phe Lys Asp Ile Tyr Pro Gly Glu Pro

515 520 525

Leu Ala Met Leu Ser Leu Asn Glu Leu Pro Ala Leu Leu Phe Glu Leu

530 535 540

Leu Arg Arg Pro Ser Ile Thr Pro Gln Asp Ile Glu Asp Arg Met Val

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

Glu Thr Arg Asn Ala Arg Glu Lys Leu His Thr Leu Glu Glu Asn Lys

610 615 620

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

625 630 635 640

Gln Gly Arg Glu Ala Ser Trp Leu Ala Gln Asp Leu Lys Arg Phe Met

645 650 655

Pro Ile Ala Ser Arg Lys Glu Trp Arg Gly Tyr His His Ser Gln Leu

660 665 670

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

675 680 685

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

690 695 700

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

Ile Lys Thr Glu Thr Lys Gln Lys Leu Glu Leu Leu Ala Arg Pro Leu
770 775 780
Val Phe Pro Arg Gly Ile Phe Asp Asp Lys Pro Thr Phe Val Gln Asp
785 790 795 800
Lys Lys Val Val Asp His Pro Glu Leu Phe Ala Asp Trp Tyr Val Tyr
805 810 815
Ser Tyr Lys Asp Asp His Ser Phe Gln Glu Phe Tyr His Tyr Lys Arg
820 825 830
Asp Tyr Asn Glu Ile Phe Glu Thr Glu Leu Ser Trp Asp Ile Asp Phe
835 840 845
Lys Asp Asn Lys Arg Gln Leu Asn Pro Ser Glu Gln Met Asp Leu Phe
850 855 860
Arg Met Lys Trp Asp Leu Lys Ile Lys Lys Ile Lys Ile Gln Asp Ile
865 870 875 880
Phe Leu Lys Ile Val Ala Glu Asp Ile Tyr Leu Lys Ile Phe Gly His
885 890 895
Lys Ile Pro Leu Ser Leu Ser Asp Phe Tyr Ile Ser Arg Gln Glu Arg
900 905 910

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

Asp Ile Gly Lys Phe Lys Tyr Phe Leu Gln Gln Gln Lys Val Leu Asn
965 970 975
Leu Leu Lys Tyr Asp Pro Gln His Val Trp Thr Lys Ala Glu Leu Glu

980 985 990
Glu Glu Leu Tyr Ile Gly Lys His Ser Tyr Glu Val Val Arg Arg Glu
995 1000 1005
Met Leu Leu Gln Lys Cys His Gln Leu Glu Lys His Ile Leu Glu Gln
1010 1015 1020
Phe Arg Phe Asp Gly Ser Asn His Pro Arg Glu Leu Glu Gln Gly Asn
1025 1030 1035 1040
His Pro Asn Phe Lys Met Tyr Ile Val Asn Gly Ile Leu Thr Lys Arg
1045 1050 1055

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

1125 1130 1135
Thr Val Ala Asp Thr Ile Val Gln Glu Phe Met Ser Leu
1140 1145

<210> 32
<211> 1120
<212> PRT
<213> Prevotella intermedia
<400> 32
Met Glu Asp Asp Lys Lys Thr Thr Asp Ser Ile Arg Tyr Glu Leu Lys
1 5 10 15
Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val

20 25 30

Tyr Ile Thr Val Asn His Ile Asn Lys Ile Leu Glu Glu Asp Glu Ile
35 40 45
Asn Arg Asp Gly Tyr Glu Asn Thr Leu Glu Asn Ser Trp Asn Glu Ile
50 55 60
Lys Asp Ile Asn Lys Lys Asp Arg Leu Ser Lys Leu Ile Ile Lys His
65 70 75 80
Phe Pro Phe Leu Glu Ala Thr Thr Tyr Arg Gln Asn Pro Thr Asp Thr
85 90 95
Thr Lys Gln Lys Glu Glu Lys Gln Ala Glu Ala Gln Ser Leu Glu Ser

100 105 110
Leu Lys Lys Ser Phe Phe Val Phe Ile Tyr Lys Leu Arg Asp Leu Arg
115 120 125
Asn His Tyr Ser His Tyr Lys His Ser Lys Ser Leu Glu Arg Pro Lys
130 135 140
Phe Glu Glu Asp Leu Gln Asn Lys Met Tyr Asn Ile Phe Asp Val Ser
145 150 155 160
Ile Gln Phe Val Lys Glu Asp Tyr Lys His Asn Thr Asp Ile Asn Pro
165 170 175

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

245 250 255
Leu Glu Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn
260 265 270

Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Gln Gly Val
275 280 285

Asn Arg Lys Lys Phe Tyr Val Ser Phe Asp Pro Ala Asp Glu Asp Tyr
290 295 300

Asp Ala Glu Gln Glu Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp
305 310 315 320

Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Val Phe
325 330 335

Ala Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile
340 345 350

Tyr Lys Lys Leu Ile Gly Gly Gln Lys Glu Asp Arg His Leu Thr His
355 360 365

Lys Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Asp Lys Gln Asn Arg
370 375 380

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

385 390 395 400

Lys Glu Glu Lys Glu Glu Glu Lys Pro Tyr Ile Ser Glu Thr Thr Pro
405 410 415

His Tyr His Leu Glu Asn Lys Lys Ile Gly Ile Ala Phe Lys Asn His
420 425 430

Asn Ile Trp Pro Ser Thr Gln Thr Glu Leu Thr Asn Asn Lys Arg Lys
435 440 445

Lys Tyr Asn Leu Gly Thr Ser Ile Lys Ala Glu Ala Phe Leu Ser Val
450 455 460

His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu Lys Thr Glu
465 470 475 480

Asn Thr Lys Asn Asp Asn Lys Val Gly Gly Lys Lys Glu Thr Lys Lys
485 490 495

Gln Gly Lys His Lys Ile Glu Ala Ile Ile Glu Ser Lys Ile Lys Asp
500 505 510

Ile Tyr Ala Leu Tyr Asp Ala Phe Ala Asn Gly Glu Ile Asn Ser Glu
515 520 525

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

530 535 540

Leu Pro Lys Gln Met Ile Ala Ile Leu Lys Asn Glu His Lys Asp Met
545 550 555 560

Ala Glu Lys Ala Glu Ala Lys Gln Glu Lys Met Lys Leu Ala Thr Glu
565 570 575

Asn Arg Leu Lys Thr Leu Asp Lys Gln Leu Lys Gly Lys Ile Gln Asn
580 585 590

Gly Lys Arg Tyr Asn Ser Ala Pro Lys Ser Gly Glu Ile Ala Ser Trp
595 600 605

Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Gln Lys Asp Glu Asn
610 615 620

Gly Glu Ser Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Leu
625 630 635 640

Leu Gln Arg Thr Leu Ala Phe Phe Gly Ser Glu His Glu Arg Leu Ala
645 650 655

Pro Tyr Phe Lys Gln Thr Lys Leu Ile Glu Ser Ser Asn Pro His Pro
660 665 670

Phe Leu Asn Asp Thr Glu Trp Glu Lys Cys Ser Asn Ile Leu Ser Phe

675 680 685

Tyr Arg Ser Tyr Leu Lys Ala Arg Lys Asn Phe Leu Glu Ser Leu Lys
690 695 700

Pro Glu Asp Trp Glu Lys Asn Gln Tyr Phe Leu Met Leu Lys Glu Pro
705 710 715 720

Lys Thr Asn Arg Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn
725 730 735

Leu Pro Arg Gly Phe Phe Thr Glu Pro Ile Arg Lys Trp Phe Met Glu
740 745 750

His Trp Lys Ser Ile Lys Val Asp Asp Leu Lys Arg Val Gly Leu Val

755 760 765

Ala Lys Val Thr Pro Leu Phe Phe Ser Glu Lys Tyr Lys Asp Ser Val

770 775 780

Gln Pro Phe Tyr Asn Tyr Pro Phe Asn Val Gly Asp Val Asn Lys Pro

785 790 795 800

Lys Glu Glu Asp Phe Leu His Arg Glu Glu Arg Ile Glu Leu Trp Asp

805 810 815

Lys Lys Lys Asp Lys Phe Lys Gly Tyr Lys Ala Lys Lys Lys Phe Lys

820 825 830

Glu Met Thr Asp Lys Glu Lys Glu Glu His Arg Ser Tyr Leu Glu Phe

835 840 845

Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln

850 855 860

Asp Ile Val Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys

865 870 875 880

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

885 890 895

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

900 905 910

Val Met Pro Met Glu Leu Pro Val Thr Val Tyr Lys Val Asn Lys Gly

915 920 925

Gly Tyr Ile Ile Lys Asn Lys Pro Leu His Thr Ile Tyr Ile Lys Glu

930 935 940

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

945 950 955 960

Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Val Lys Thr Pro Ser Glu

965 970 975

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

980 985 990

Leu Gly Lys Tyr Gln Asn Ala Arg Leu Asp Ile Ile Glu Asp Met Leu

995 1000 1005

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

1010 1015 1020

Asn Phe His Asn Met Leu Thr Gly Trp Leu Glu Leu Lys Gly Glu Ala

1025 1030 1035 1040

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

1105 1110 1115 1120

<210> 33
<211> 1119
<212> PRT
<213> Porphyromonas gingivalis
<400> 33

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

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

115 120 125

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

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

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

210 215 220

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

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

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

385 390 395 400

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

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460

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

465 470 475 480

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

485 490 495

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

500 505 510

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

515 520 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

690 695 700
Arg Ser Tyr Leu Glu Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720
Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
740 745 750
Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765

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

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

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
885 890 895
Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val Gln Glu Gln Gly Ser Leu
900 905 910

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

980 985 990
Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala
995 1000 1005
Phe Glu Gln Thr Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Tyr Pro
1010 1015 1020
His Leu Pro Asp Lys Asn Phe Arg Lys Met Leu Glu Ser Trp Ser Asp
1025 1030 1035 1040
Pro Leu Leu Asp Lys Trp Pro Asp Leu His Gly Asn Val Arg Leu Leu
1045 1050 1055

Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp
1060 1065 1070
Glu Thr Leu Phe Ser Ser Ile Arg Lys Tyr Asp Pro Ser Ser Pro Asp
1075 1080 1085
Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu Ser Glu
1090 1095 1100
Glu Val Lys Gln Ala Lys Glu Met Val Glu Arg Ile Ile Gln Ala
1105 1110 1115
<210> 34

<211> 1161

<212> PRT

<213> Bacteroidetes bacterium

<400> 34

Met Glu Asn Gln Thr Gln Lys Gly Lys Gly Ile Tyr Tyr Tyr Tyr Thr
1 5 10 15

Lys Asn Glu Asp Lys His Tyr Phe Gly Ser Phe Leu Asn Leu Ala Asn
20 25 30

Asn Asn Ile Glu Gln Ile Ile Glu Glu Phe Arg Ile Arg Leu Ser Leu
35 40 45

Lys Asp Glu Lys Asn Ile Lys Glu Ile Ile Asn Asn Tyr Phe Thr Asp
50 55 60

Lys Lys Ser Tyr Thr Asp Trp Glu Arg Gly Ile Asn Ile Leu Lys Glu
65 70 75 80

Tyr Leu Pro Val Ile Asp Tyr Leu Asp Leu Ala Ile Thr Asp Lys Glu
85 90 95

Phe Glu Lys Ile Asp Leu Lys Gln Lys Glu Thr Ala Lys Arg Lys Tyr
100 105 110

Phe Arg Thr Asn Phe Ser Leu Leu Ile Asp Thr Ile Ile Asp Leu Arg
115 120 125

Asn Phe Tyr Thr His Tyr Phe His Lys Pro Ile Ser Ile Asn Pro Asp

130 135 140

Val Ala Lys Phe Leu Asp Lys Asn Leu Leu Asn Val Cys Leu Asp Ile
145 150 155 160

Lys Lys Gln Lys Met Lys Thr Asp Lys Thr Lys Gln Ala Leu Lys Asp
165 170 175

Gly Leu Asp Lys Glu Leu Lys Lys Leu Ile Glu Leu Lys Lys Ala Glu
180 185 190

Leu Lys Glu Lys Lys Ile Lys Thr Trp Asn Ile Thr Glu Asn Val Glu
195 200 205

Gly Ala Val Tyr Asn Asp Ala Phe Asn His Met Val Tyr Lys Asn Asn

210 215 220

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

225 230 235 240

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

245 250 255

Val Phe Leu Leu Ser Met Phe Leu Ser Lys Lys Glu Ile Glu Gln Phe

260 265 270

Lys Ser Asn Leu Glu Gly Phe Lys Gly Lys Val Ile Gly Glu Asn Gly

275 280 285

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

290 295 300

His Trp Ile Phe Ser Tyr Leu Thr Phe Lys Gly Leu Lys Gln Arg Val

305 310 315 320

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

325 330 335

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

340 345 350

Gln Asn Glu Phe Leu Glu Asp Ile Asn Glu Tyr Val Gln Asp Asn Glu

355 360 365

Glu Asn Lys Lys Ser Met Glu Asn Ser Ile Val Val His Pro Val Ile

370 375 380

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

385 390 395 400

Asp Glu Phe Ala Asn Phe Pro Thr Leu Lys Phe Phe Val Thr Ala Gly

405 410 415

Asn Phe Val His Asp Lys Arg Glu Lys Gln Ile Gln Gly Ser Met Leu

420 425 430

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

435 440 445

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

450 455 460

Leu Glu Thr Ser Glu Trp Glu Leu Phe Pro Asn Pro Ser Tyr Leu Leu

465 470 475 480

Ile Gln Asn Asn Ile Pro Val His Ile Asp Leu Ile His Asn Thr Glu

485 490 495

Glu Ala Lys Gln Cys Gln Ile Ala Ile Asp Arg Ile Lys Cys Thr Thr

500 505 510

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

515 520 525

Lys Ile Ile Tyr Gln Lys Asn Lys Asn Ile Lys Tyr Gly Asp Pro Thr

530 535 540

Ala Leu Leu Ser Ser Asn Glu Leu Pro Ala Leu Ile Tyr Glu Leu Leu

545 550 555 560

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

565 570 575

Ile Val Asn Gln Tyr Lys Thr Ile Ala Gly Phe Glu Lys Gly Gln Asn

580 585 590

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

595 600 605

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

610 615 620

Leu Glu Ile Thr Glu Asn Lys Leu Asn Ile Ile Lys Asn Asn Arg Ala

625 630 635 640

Glu Phe Arg Thr Gly Ala Lys Arg Lys His Ile Phe Tyr Ser Lys Glu

645 650 655

Leu Gly Gln Glu Ala Thr Trp Ile Ala Tyr Asp Leu Lys Arg Phe Met

660 665 670

Pro Glu Ala Ser Arg Lys Glu Trp Lys Gly Phe His His Ser Glu Leu

675 680 685

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

690 695 700

Leu Leu Asn Met Phe Trp Asn Phe Asp Asn Asp Gln Leu Ile Gly Asn

705 710 715 720

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

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

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

Leu Ile Ala Asp Val Gln Lys Asn Arg Glu Lys Gly Asp Thr Ser Asp
930 935 940
Asn Lys Met Asn Glu Asn Phe Ile Trp Asn Met Thr Ile Pro Leu Ser
945 950 955 960
Leu Cys Asn Gly Gln Ile Glu Glu Pro Lys Val Lys Leu Lys Asp Ile
965 970 975

Gly Lys Phe Arg Lys Leu Glu Thr Asp Asp Lys Val Ile Gln Leu Leu
980 985 990
Glu Tyr Asp Lys Ser Lys Val Trp Lys Lys Leu Glu Ile Glu Asp Glu

995 1000 1005
Leu Glu Asn Met Pro Asn Ser Tyr Glu Arg Ile Arg Arg Glu Lys Leu
1010 1015 1020
Leu Lys Gly Ile Gln Glu Phe Glu His Phe Leu Leu Glu Lys Glu Lys
1025 1030 1035 1040
Phe Asp Gly Ile Asn His Pro Lys His Phe Glu Gln Asp Leu Asn Pro
1045 1050 1055
Asn Phe Lys Thr Tyr Val Ile Asn Gly Val Leu Arg Lys Asn Ser Lys
1060 1065 1070

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

1140 1145 1150
Phe Lys Asn Ser Leu Glu Lys His Ser
1155 1160

<210> 35
<211> 1165
<212> PRT
<213> Chryseobacterium jejuense
<400> 35
Met Glu Lys Thr Gln Thr Gly Leu Gly Ile Tyr Tyr Asp His Thr Lys
1 5 10 15
Leu Gln Asp Lys Tyr Phe Phe Gly Gly Phe Phe Asn Leu Ala Gln Asn

20 25 30

Asn Ile Asp Asn Val Ile Lys Ala Phe Ile Ile Lys Phe Phe Pro Glu

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

Leu Arg Asn Phe Tyr Thr His Tyr Tyr His Lys Ser Ile Glu Phe Pro

115 120 125

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

130 135 140

Glu Ile Lys Lys Leu Lys Lys Lys Asp Asp Lys Thr Gln Gln Leu Leu

145 150 155 160

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

165 170 175

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

180 185 190

Asp Glu Thr Ala Ile Arg Asn Gly Val Phe Asn Ala Ala Phe Asn His

195 200 205

Leu Ile Tyr Arg Asp Gly Glu Asn Val Lys Pro Ser Arg Leu Tyr Gln

210 215 220

Ser Ser Tyr Ser Glu Pro Asp Pro Ala Glu Asn Gly Ile Ser Leu Ser

225 230 235 240

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

245 250 255

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

260 265 270

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

325 330 335
Glu Lys Phe Leu Glu Asp Ile Asn Glu Tyr Met Lys Glu Gly Asn Ala
340 345 350
Asp Leu Ser Leu Glu Asp Ser Lys Val Ile His Pro Val Ile Arg Lys
355 360 365
Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu
370 375 380
Tyr Leu Ser Ser Thr Ser Leu Lys Phe Gln Val His Val Gly Asn Tyr
385 390 395 400

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

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

Pro Leu Ala Leu Leu Ser Leu Asn Glu Ile Pro Ala Leu Leu Tyr Gln
530 535 540

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

610 615 620
Glu Lys Arg Leu Lys Ala Lys Lys Glu Gln Arg Arg Asn Thr Pro Gln
625 630 635 640
Arg Ser Ile Phe Ser Asn Ser Asp Leu Gly Arg Ile Ala Ala Trp Leu
645 650 655
Ala Asp Asp Ile Lys Arg Phe Met Pro Ala Glu Gln Arg Lys Asn Trp
660 665 670
Lys Gly Tyr Gln His Ser Gln Leu Gln Gln Ser Leu Ala Tyr Phe Glu
675 680 685

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

755 760 765
Leu Phe Pro Lys Arg His Tyr Ile Leu Lys Asp Leu Glu Thr Glu Lys

770 775 780

Asn Lys Val Leu Ser Lys Pro Leu Val Phe Ser Arg Gly Leu Phe Asp

785 790 795 800

Asn Asn Pro Thr Phe Ile Lys Gly Val Lys Val Thr Glu Asn Pro Glu

805 810 815

Leu Phe Ala Glu Trp Tyr Ser Tyr Gly Tyr Lys Thr Glu His Val Phe

820 825 830

Gln His Phe Tyr Gly Trp Glu Arg Asp Tyr Asn Glu Leu Leu Asp Ser

835 840 845

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

850 855 860

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

865 870 875 880

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

885 890 895

Leu Phe Glu Asn Val Phe Asn Tyr Pro Thr Thr Leu Ser Leu Asp Glu

900 905 910

Phe Tyr Leu Thr Gln Glu Glu Arg Ala Glu Lys Glu Arg Ile Ala Leu

915 920 925

Ala Gln Ser Leu Arg Glu Glu Gly Asp Asn Ser Pro Asn Ile Ile Lys

930 935 940

Asp Asp Phe Ile Trp Ser Lys Thr Ile Ala Phe Arg Ser Lys Gln Ile

945 950 955 960

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

965 970 975

Val Leu Asp Asp Glu Glu Ser Lys Ala Ser Lys Leu Leu Ser Tyr Asp

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ile Gln Asn Leu Glu Leu Gln Ile Leu Ser Asn Trp Ser Trp Asp Gly

1025 1030 1035 1040

Ile Asn His Pro Arg Glu Phe Glu Met Glu Asp Gln Lys Asn Thr Arg

1045 1050 1055

His Pro Asn Phe Lys Met Tyr Leu Val Asn Gly Ile Leu Arg Lys Asn

1060 1065 1070

Ile Asn Leu Tyr Lys Glu Asp Glu Asp Phe Trp Leu Glu Ser Leu Lys

1075 1080 1085

Glu Asn Asp Phe Lys Thr Leu Pro Ser Glu Val Leu Glu Thr Lys Ser

1090 1095 1100

Glu Met Val Gln Leu Leu Phe Leu Val Ile Leu Ile Arg Asn Gln Phe

1105 1110 1115 1120

Ala His Asn Gln Leu Pro Glu Ile Gln Phe Tyr Asn Phe Ile Arg Lys

1125 1130 1135

Asn Tyr Pro Glu Ile Gln Asn Asn Thr Val Ala Glu Leu Tyr Leu Asn

1140 1145 1150

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

1155 1160 1165

<210> 36

<211> 1156

<212> PRT

<213> Chryseobacterium carnipullorum

<400> 36

Met Asn Thr Arg Val Thr Gly Met Gly Val Ser Tyr Asp His Thr Lys

1 5 10 15

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

20 25 30

Asn Ile Thr Ala Val Ile Lys Ala Phe Cys Ile Lys Phe Asp Lys Asn

35 40 45

Pro Met Ser Ser Val Gln Phe Ala Glu Ser Cys Phe Thr Asp Lys Asp

50 55 60

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

65 70 75 80

Pro Val Ile Gly Tyr Leu Asn Tyr Gly Gly Asp Arg Asn Thr Phe Arg
85 90 95
Gln Lys Leu Ser Thr Leu Leu Lys Ala Val Asp Ser Leu Arg Asn Phe
100 105 110
Tyr Thr His Tyr Tyr His Ser Pro Leu Ala Leu Ser Thr Glu Leu Phe
115 120 125
Glu Leu Leu Asp Thr Val Phe Ala Ser Val Ala Val Glu Val Lys Gln
130 135 140
His Lys Met Lys Asp Asp Lys Thr Arg Gln Leu Leu Ser Lys Ser Leu

145 150 155 160
Ala Glu Glu Leu Asp Ile Arg Tyr Lys Gln Gln Leu Glu Arg Leu Lys
165 170 175
Glu Leu Lys Glu Gln Gly Lys Asn Ile Asp Leu Arg Asp Glu Ala Gly
180 185 190
Ile Arg Asn Gly Val Leu Asn Ala Ala Phe Asn His Leu Ile Tyr Lys
195 200 205
Glu Gly Glu Ile Ala Lys Pro Thr Leu Ser Tyr Ser Ser Phe Tyr Tyr
210 215 220

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

290 295 300
Glu Glu Thr Leu Leu Ile Gln Ile Ile Asp Glu Leu Ser Lys Val Pro
305 310 315 320

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

Lys Phe Asn Tyr Phe Ala Val Arg Phe Leu Asp Glu Phe Ile Lys Phe
370 375 380
Pro Ser Leu Arg Phe Gln Val His Leu Gly Asn Phe Val His Asp Arg
385 390 395 400
Arg Ile Lys Asp Ile His Gly Thr Gly Phe Gln Thr Glu Arg Val Val
405 410 415
Lys Asp Arg Ile Lys Val Phe Gly Lys Leu Ser Glu Ile Ser Ser Leu
420 425 430
Lys Thr Glu Tyr Ile Glu Lys Glu Leu Asp Leu Asp Ser Asp Thr Gly
435 440 445
Trp Glu Ile Phe Pro Asn Pro Ser Tyr Val Phe Ile Asp Asn Asn Ile
450 455 460
Pro Ile Tyr Ile Ser Thr Asn Lys Thr Phe Lys Asn Gly Ser Ser Glu
465 470 475 480
Phe Ile Lys Leu Arg Arg Lys Glu Lys Pro Glu Glu Met Lys Met Arg
485 490 495
Gly Glu Asp Lys Lys Glu Lys Arg Asp Ile Ala Ser Met Ile Gly Asn
500 505 510

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

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

580 585 590

Asn Pro Glu Lys Leu Ile His Leu Ile Asn Lys Glu Ile Asp Ala Thr

595 600 605

Glu Ala Lys Phe Ala Leu Leu Ala Lys Asn Arg Lys Glu Leu Lys Glu

610 615 620

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

625 630 635 640

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

645 650 655

Pro Asp Ile Leu Arg Lys Asn Trp Lys Gly Tyr Gln His Asn Gln Leu

660 665 670

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

675 680 685

Ile Leu Gln Asp Gly Trp Asp Phe Ala Asp Gly Ser Ser Phe Trp Asn

690 695 700

Gly Trp Ile Ile Asn Ser Phe Val Lys Asn Arg Ser Phe Glu Tyr Phe

705 710 715 720

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

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

Val Phe Pro Arg Gly Leu Phe Asp Thr Lys Pro Thr Phe Ile Lys Gly

785 790 795 800

Ile Lys Val Asp Glu Gln Pro Glu Leu Phe Ala Glu Trp Tyr Gln Tyr

805 810 815

Gly Tyr Ser Thr Glu His Val Phe Gln Asn Phe Tyr Gly Trp Glu Arg

820 825 830

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

835 840 845

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

850 855 860

Lys Leu Lys Gln Asp Leu Lys Ile Lys Lys Ile Lys Ile Gln Asp Leu

865 870 875 880

Phe Leu Lys Leu Ile Ala Gly His Ile Phe Glu Asn Ile Phe Lys Tyr

885 890 895

Pro Ala Ser Phe Ser Leu Asp Glu Leu Tyr Leu Thr Gln Glu Glu Arg

900 905 910

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

915 920 925

Asp His Ser Asp Asn Ile Ile Lys Asp Asn Phe Ile Gly Ser Lys Thr

930 935 940

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

945 950 955 960

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

965 970 975

Leu Leu Ser Tyr Asn Glu Asp Lys Val Trp Asn Lys Asn Asp Leu Asp

980 985 990

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

995 1000 1005

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

1010 1015 1020

Trp Pro Trp Asn Gly Thr Asp His Pro Glu Glu Phe Gly Thr Thr Asp

1025 1030 1035 1040

Asn Lys Gly Val Asn His Pro Asn Phe Lys Met Tyr Val Val Asn Gly

1045 1050 1055

Ile Leu Arg Lys His Thr Asp Trp Phe Lys Glu Gly Glu Asp Asn Trp

1060 1065 1070

Leu Glu Asn Leu Asn Glu Thr His Phe Lys Asn Leu Ser Phe Gln Glu

1075 1080 1085

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

1155

<210> 37
<211> 1036
<212> PRT
<213> Chryseobacterium ureilyticum
<400> 37
Met Glu Thr Gln Ile Leu Gly Asn Gly Ile Ser Tyr Asp His Thr Lys
1 5 10 15
Thr Glu Asp Lys His Phe Phe Gly Gly Phe Leu Asn Thr Ala Gln Asn
20 25 30
Asn Ile Asp Leu Leu Ile Lys Ala Tyr Ile Ser Lys Phe Glu Ser Ser
35 40 45
Pro Arg Lys Leu Asn Ser Val Gln Phe Pro Asp Val Cys Phe Lys Lys

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

115 120 125

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

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

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

340 345 350
Ser Leu Ile Glu Ser Lys Ile Val His Pro Val Ile Arg Lys Arg Tyr
355 360 365

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

Leu Val Lys Asp Arg Ile Lys Val Phe Gly Lys Leu Ser Thr Ile Ser
420 425 430
Ser Leu Lys Ala Glu Tyr Leu Ala Lys Ala Val Asn Ile Thr Asp Asp
435 440 445
Thr Gly Trp Glu Leu Leu Pro His Pro Ser Tyr Val Phe Ile Asp Asn
450 455 460
Asn Ile Pro Ile His Leu Thr Val Asp Pro Ser Phe Lys Asn Gly Val
465 470 475 480
Lys Glu Tyr Gln Glu Lys Arg Lys Leu Gln Lys Pro Glu Glu Met Lys
485 490 495
Asn Arg Gln Gly Gly Asp Lys Met His Lys Pro Ala Ile Ser Ser Lys
500 505 510
Ile Gly Lys Ser Lys Asp Ile Asn Pro Glu Ser Pro Val Ala Leu Leu
515 520 525
Ser Met Asn Glu Ile Pro Ala Leu Leu Tyr Glu Ile Leu Val Lys Lys
530 535 540
Ala Ser Pro Glu Glu Val Glu Ala Lys Ile Arg Gln Lys Leu Thr Ala
545 550 555 560

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

Lys Val Lys Gly Lys Phe Lys Arg Gln Lys Val Phe Lys Asn Ala Glu

625 630 635 640

Leu Gly Thr Glu Ala Thr Trp Leu Ala Asn Asp Ile Lys Arg Phe Met

645 650 655

Pro Glu Glu Gln Lys Lys Asn Trp Lys Gly Tyr Gln His Ser Gln Leu

660 665 670

Gln Gln Ser Leu Ala Phe Phe Glu Ser Arg Pro Gly Glu Ala Arg Ser

675 680 685

Leu Leu Gln Ala Gly Trp Asp Phe Ser Asp Gly Ser Ser Phe Trp Asn

690 695 700

Gly Trp Val Met Asn Ser Phe Ala Arg Asp Asn Thr Phe Asp Gly Phe

705 710 715 720

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

725 730 735

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

740 745 750

Ile Lys Gln Gln Met Pro Lys Gly Leu Phe Asp Arg Arg Leu Tyr Met

755 760 765

Leu Glu Asp Leu Ala Thr Glu Lys Asn Lys Ile Leu Ser Lys Pro Leu

770 775 780

Ile Phe Pro Arg Gly Ile Phe Asp Asp Lys Pro Thr Phe Lys Lys Gly

785 790 795 800

Val Gln Val Ser Glu Glu Pro Glu Ala Phe Ala Asp Trp Tyr Ser Tyr

805 810 815

Gly Tyr Asp Val Lys His Lys Phe Gln Glu Phe Tyr Ala Trp Asp Arg

820 825 830

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

835 840 845

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

850 855 860

Ala Lys Lys Gln Asp Leu Lys Val Lys Lys Val Arg Ile Gln Asp Leu

865 870 875 880

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

885 890 895

Glu Leu Ala Leu Pro Leu Asp Gln Phe Tyr Leu Thr Gln Glu Glu Arg

900 905 910

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

915 920 925

Asp Asp Ser Pro Asn Ile Val Lys Glu Asn Phe Ile Trp Ser Lys Thr

930 935 940

Ile Pro Phe Lys Ser Gly Arg Val Phe Glu Pro Asn Val Lys Leu Lys

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Leu Phe Lys Glu Ile Gln Gln Met Lys Arg Leu Ser Leu Arg Ser Val

1010 1015 1020

Arg Gly Met Gly Val Pro Ile Arg Leu Asn Leu Lys

1025 1030 1035

<210> 38

<211> 1133

<212> PRT

<213> Reichenbachiella agariperforans

<400> 38

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

1 5 10 15

Lys Lys Phe Asn Thr Glu Ser Asp Lys Ser Phe Lys Lys Ile Phe Gln

20 25 30

Asn Lys Gly Ser Ile Ala Pro Ile Ala Glu Lys Ala Cys Lys Asn Phe

35 40 45

Glu Ile Lys Ser Lys Ser Pro Val Asn Arg Asp Gly Arg Leu His Tyr
50 55 60
Phe Ser Val Gly His Ala Phe Lys Asn Ile Asp Ser Lys Asn Val Phe
65 70 75 80
Arg Tyr Glu Leu Asp Glu Ser Gln Met Asp Met Lys Pro Thr Gln Phe

85 90 95
Leu Ala Leu Gln Lys Glu Phe Phe Asp Phe Gln Gly Ala Leu Asn Gly
100 105 110
Leu Leu Lys His Ile Arg Asn Val Asn Ser His Tyr Val His Thr Phe
115 120 125
Glu Lys Leu Glu Ile Gln Ser Ile Asn Gln Lys Leu Ile Thr Phe Leu
130 135 140
Ile Glu Ala Phe Glu Leu Ala Val Ile His Ser Tyr Leu Asn Glu Glu
145 150 155 160

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

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

290 295 300

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

370 375 380
Gln Glu Gln Lys Gly Phe Ser Tyr Glu Ile Asn Thr Ser Ala Glu Ile
385 390 395 400
Lys Asp Ile Glu Thr Lys Leu Lys Ala Leu Val Leu Lys Gly Pro Leu
405 410 415
Asn Phe Lys Glu Lys Lys Glu Gln Asn Arg Leu Glu Lys Asp Leu Arg
420 425 430
Arg Glu Lys Lys Glu Gln Pro Thr Asn Arg Val Lys Glu Lys Leu Leu
435 440 445

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

515 520 525
Thr Tyr Ala Glu Gln Gln Ala Arg Tyr Pro Asp Trp Asp Thr Pro Phe
530 535 540

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

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

660 665 670
Ile Leu His Gln Thr Glu Glu Asp Phe Leu Lys Arg Asn Lys Gly Lys
675 680 685
Gln Phe Lys Leu Arg Phe Val Arg Lys Ala Trp His Leu Met Tyr Leu
690 695 700
Lys Glu Leu Tyr Met Asn Lys Val Ala Glu His Gly His His Lys Ser
705 710 715 720
Phe His Ile Thr Lys Glu Glu Phe Asn Asp Phe Cys Arg Trp Met Phe
725 730 735

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

Leu Leu Ala Asn Tyr Glu Ser Met Leu Lys Asp Asp Met Leu Tyr Val

805 810 815

Asn Ile Ser His Phe Ile Ser Tyr Leu Glu Ser Lys Gly Lys Ile Asn

820 825 830

Arg Asn Ala His Gly His Ile Ala Tyr Lys Ala Leu Asn Asn Val Pro

835 840 845

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

850 855 860

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

865 870 875 880

Glu Asp Ala Leu Leu Tyr Glu Met Ala Met His Tyr Leu Ser Leu Glu

885 890 895

Pro Ala Leu Val Pro Lys Val Lys Thr Lys Val Lys Asp Ile Leu Ser

900 905 910

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

915 920 925

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

930 935 940

Asn His Gln Ser Gln Gln Glu Lys Asp Pro Asp Lys Thr Ser Phe Leu

945 950 955 960

Ala Lys Ile Gln Pro Tyr Leu Glu Lys Val Lys Asn Ser Lys Asp Leu

965 970 975

Lys Ala Val Tyr His Tyr Tyr Lys Asp Thr Pro His Thr Leu Arg Tyr

980 985 990

Glu Asp Leu Asn Met Ile His Ser His Ile Val Ser Gln Ser Val Gln

995 1000 1005

Phe Thr Lys Val Ala Leu Lys Leu Glu Glu Tyr Phe Ile Ala Lys Lys

1010 1015 1020

Ser Ile Thr Leu Gln Ile Ala Arg Gln Ile Ser Tyr Ser Glu Ile Ala

1025 1030 1035 1040

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

1045 1050 1055

Phe Asp Val Pro Glu Thr Ala Tyr Ser Met Ile Leu Gln Gly Ile Glu

1060 1065 1070

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

1075 1080 1085

Ser Glu Leu Ser Thr Gln Gln Val Ser Val Cys Thr Ala Phe Leu Glu

1090 1095 1100

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

1105 1110 1115 1120

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

1125 1130

<210> 39

<211> 1224

<212> PRT

<213> Bergeyella zoohelcum

<400> 39

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

1 5 10 15

Lys Pro Gln Asp Lys Ser Tyr Phe Ala Gly Tyr Phe Asn Ala Ala Met

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

Phe Pro Met Ala Arg Leu Leu Asp Lys Lys Glu Val Pro Ile Lys Glu

85 90 95

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

100 105 110

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

115 120 125

Ile Thr Asp Glu Ile Phe Gly Val Leu Asp Glu Met Leu Lys Ser Thr
130 135 140
Val Leu Thr Val Lys Lys Lys Lys Val Lys Thr Asp Lys Thr Lys Glu
145 150 155 160
Ile Leu Lys Lys Ser Ile Glu Lys Gln Leu Asp Ile Leu Cys Gln Lys

165 170 175
Lys Leu Glu Tyr Leu Arg Asp Thr Ala Arg Lys Ile Glu Glu Lys Arg
180 185 190
Arg Asn Gln Arg Glu Arg Gly Glu Lys Glu Leu Val Ala Pro Phe Lys
195 200 205
Tyr Ser Asp Lys Arg Asp Asp Leu Ile Ala Ala Ile Tyr Asn Asp Ala
210 215 220
Phe Asp Val Tyr Ile Asp Lys Lys Lys Asp Ser Leu Lys Glu Ser Ser
225 230 235 240

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

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

370 375 380

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

450 455 460
Ile Thr Val Phe Gly Arg Leu Ser Glu Leu Glu His Lys Lys Ala Leu
465 470 475 480
Phe Ile Lys Asn Thr Glu Thr Asn Glu Asp Arg Glu His Tyr Trp Glu
485 490 495
Ile Phe Pro Asn Pro Asn Tyr Asp Phe Pro Lys Glu Asn Ile Ser Val
500 505 510
Asn Asp Lys Asp Phe Pro Ile Ala Gly Ser Ile Leu Asp Arg Glu Lys
515 520 525

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

595 600 605
Glu Phe Phe Asp Lys Trp Glu Lys Lys Lys Glu Lys Leu Glu Lys Lys
610 615 620

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

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

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

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

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

885 890 895

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

900 905 910

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

915 920 925

Tyr Asp Thr Glu Asn Tyr Pro Leu Val Glu Leu Glu Lys Lys Gln Ala

930 935 940

Asp Arg Lys Arg Lys Thr Lys Ile Tyr Gln Gln Lys Lys Asn Asp Val

945 950 955 960

Phe Thr Leu Leu Met Ala Lys His Ile Phe Lys Ser Val Phe Lys Gln

965 970 975

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

980 985 990

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

995 1000 1005

Thr Asn Tyr Ile Trp Asn Lys Thr Val Asp Leu Lys Leu Cys Asp Gly

1010 1015 1020

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

1025 1030 1035 1040

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

1045 1050 1055

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

1060 1065 1070

Asn Tyr Pro Tyr Val Val Glu Arg Glu Ile Glu Gln Tyr Glu Lys Val

1075 1080 1085

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

1090 1095 1100

Leu Glu Lys Val Lys Asp Lys Glu Ile Leu Lys Lys Gly Asp Asn Gln

1105 1110 1115 1120

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

1125 1130 1135

Glu Asp Val Glu Ser Tyr Lys Val Phe Asn Leu Asn Thr Glu Pro Glu

1140 1145 1150

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

1155 1160 1165

Lys Ala Phe Val Leu Thr Tyr Ile Arg Asn Lys Phe Ala His Asn Gln

1170 1175 1180

Leu Pro Lys Lys Glu Phe Trp Asp Tyr Cys Gln Glu Lys Tyr Gly Lys

1185 1190 1195 1200

Ile Glu Lys Glu Lys Thr Tyr Ala Glu Tyr Phe Ala Glu Val Phe Lys

1205 1210 1215

Lys Glu Lys Glu Ala Leu Ile Lys

1220

<210> 40

<211> 1126

<212> PRT

<213> Prevotella intermedia

<400> 40

Met Glu Asp Asp Lys Lys Thr Thr Asp Ser Ile Arg Tyr Glu Leu Lys

1 5 10 15

Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val

20 25 30

Tyr Ile Thr Val Asn His Ile Asn Lys Ile Leu Glu Glu Gly Glu Ile

35 40 45

Asn Arg Asp Gly Tyr Glu Thr Thr Leu Lys Asn Thr Trp Asn Glu Ile

50 55 60

Lys Asp Ile Asn Lys Lys Asp Arg Leu Ser Lys Leu Ile Ile Lys His

65 70 75 80

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

85 90 95

Thr Lys Gln Lys Glu Glu Lys Gln Ala Glu Ala Gln Ser Leu Glu Ser

100 105 110

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

145 150 155 160
Ile Arg Leu Val Lys Glu Asp Tyr Gln Tyr Asn Lys Asp Ile Asn Pro
165 170 175
Asp Glu Asp Phe Lys His Leu Asp Arg Thr Glu Glu Glu Phe Asn Tyr
180 185 190
Tyr Phe Thr Lys Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu
195 200 205
Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Met Gln
210 215 220

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

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

355 360 365

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

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

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

580 585 590
Val Glu Arg Lys Asn Ser Ser Leu Lys Ser Gly Lys Ile Ala Ser Trp
595 600 605

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

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

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

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

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

865 870 875 880

Tyr Leu Lys Asn Ile Asn Thr Asn Thr Thr Lys Lys Glu Lys Asn Thr

885 890 895

Glu Glu Lys Asn Gly Glu Glu Lys Asn Ile Lys Glu Lys Asn Asn Ile

900 905 910

Leu Asn Arg Ile Met Pro Met Arg Leu Pro Ile Lys Val Tyr Gly Arg

915 920 925

Glu Asn Phe Ser Lys Asn Lys Lys Lys Lys Ile Arg Arg Asn Thr Phe

930 935 940

Phe Thr Val Tyr Ile Glu Glu Lys Gly Thr Lys Leu Leu Lys Gln Gly

945 950 955 960

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

965 970 975

Phe Val Lys Thr Pro Ser Lys Ala Glu Ser Lys Ser Asn Thr Ile Ser

980 985 990

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

995 1000 1005

Glu Ile Ile Lys Asp Met Leu Ala Leu Glu Lys Thr Leu Ile Asp Lys

1010 1015 1020

Tyr Asn Ser Leu Asp Thr Asp Asn Phe Asn Lys Met Leu Thr Asp Trp

1025 1030 1035 1040

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

1045 1050 1055

Asp Leu Leu Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro

1060 1065 1070

Met Arg Asn Arg Ile Ala Phe Ala Asn Ile Asn Pro Phe Ser Leu Ser

1075 1080 1085

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

1090 1095 1100

Lys Asp Lys Thr His Lys Thr Ile Glu Lys Ile Ile Glu Ile Glu Lys

1105 1110 1115 1120

Pro Ile Glu Thr Lys Glu

1125

<210> 41

<211> 1127

<212> PRT

<213> Prevotella buccae

<400> 41

Met Gln Lys Gln Asp Lys Leu Phe Val Asp Arg Lys Lys Asn Ala Ile

1 5 10 15

Phe Ala Phe Pro Lys Tyr Ile Thr Ile Met Glu Asn Lys Glu Lys Pro

20 25 30

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

35 40 45

Leu Asn Leu Ala Arg His Asn Val Tyr Thr Thr Ile Asn His Ile Asn

50 55 60

Arg Arg Leu Glu Ile Ala Glu Leu Lys Asp Asp Gly Tyr Met Met Gly

65 70 75 80

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

85 90 95

Arg Leu Arg Asp Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala

100 105 110

Ala Tyr Glu Met Thr Asn Ser Lys Ser Pro Asn Asn Lys Glu Gln Arg

115 120 125

Glu Lys Glu Gln Ser Glu Ala Leu Ser Leu Asn Asn Leu Lys Asn Val

130 135 140

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

145 150 155 160

His Tyr Lys Tyr Ser Glu Glu Ser Pro Lys Pro Ile Phe Glu Thr Ser

165 170 175

Leu Leu Lys Asn Met Tyr Lys Val Phe Asp Ala Asn Val Arg Leu Val

180 185 190

Lys Arg Asp Tyr Met His His Glu Asn Ile Asp Met Gln Arg Asp Phe
195 200 205
Thr His Leu Asn Arg Lys Lys Gln Val Gly Arg Thr Lys Asn Ile Ile
210 215 220
Asp Ser Pro Asn Phe His Tyr His Phe Ala Asp Lys Glu Gly Asn Met

225 230 235 240
Thr Ile Ala Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Asp Lys Lys
245 250 255
Asp Ala Ile Trp Met Gln Lys Lys Leu Lys Gly Phe Lys Asp Gly Arg
260 265 270
Asn Leu Arg Glu Gln Met Thr Asn Glu Val Phe Cys Arg Ser Arg Ile
275 280 285
Ser Leu Pro Lys Leu Lys Leu Glu Asn Val Gln Thr Lys Asp Trp Met
290 295 300

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

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

435 440 445

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

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

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

660 665 670
Ser Glu Asn Phe Arg Leu Lys Ala Tyr Phe Asn Gln Met Asn Leu Val
675 680 685

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

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

805 810 815
Tyr Pro Phe Asn Ile Gly Asn Arg Leu Lys Pro Lys Lys Arg Gln Phe
820 825 830
Leu Asp Lys Lys Glu Arg Val Glu Leu Trp Gln Lys Asn Lys Glu Leu
835 840 845
Phe Lys Asn Tyr Pro Ser Glu Lys Lys Lys Thr Asp Leu Ala Tyr Leu
850 855 860
Asp Phe Leu Ser Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Ile Lys
865 870 875 880

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

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

945 950 955 960

Glu Thr Lys Val Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp

965 970 975

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

980 985 990

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

995 1000 1005

Lys Tyr Gln Thr Thr Arg Ile Ser Ile Phe Glu Met Thr Leu Gly Leu

1010 1015 1020

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

1025 1030 1035 1040

Arg Asn Met Leu Glu Arg Trp Leu Gln Cys Lys Ala Asn Arg Pro Glu

1045 1050 1055

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

1060 1065 1070

His Asn Gln Tyr Pro Met Tyr Asp Ala Thr Leu Phe Ala Glu Val Lys

1075 1080 1085

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

1090 1095 1100

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

1105 1110 1115 1120

Glu Lys Ser Glu Asn Lys Asn

1125

<210> 42

<211> 1135

<212> PRT

<213> Porphyromonas gingivalis

<400> 42

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val

1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu

20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys

35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu

50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr

65 70 75 80

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

85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile

100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His

115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp

130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr

210 215 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys

225 230 235 240

Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala

245 250 255

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu

260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu

275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asp Glu Glu

290 295 300

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

305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe

325 330 335

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

340 345 350

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

355 360 365

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

370 375 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu

385 390 395 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile

405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro

420 425 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met

435 440 445

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

450 455 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys

465 470 475 480

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

485 490 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro

500 505 510

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

515 520 525

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

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

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

740 745 750
Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Val Met
755 760 765

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

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
820 825 830
Lys Thr Val Arg Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
835 840 845
Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
850 855 860
Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865 870 875 880
Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
885 890 895
Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
900 905 910
Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
915 920 925
Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
930 935 940
Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945 950 955 960

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

Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys

1025 1030 1035 1040

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

1045 1050 1055

Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg

1060 1065 1070

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

1075 1080 1085

Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu

1090 1095 1100

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

1105 1110 1115 1120

Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu

1125 1130 1135

<210> 43

<211> 1115

<212> PRT

<213> Bacteroides pyogenes

<400> 43

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

1 5 10 15

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

20 25 30

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

35 40 45

Leu Leu Pro Glu Lys Ser Gly Phe His Ser Leu Leu Thr Thr Asp Asn

50 55 60

Leu Ser Ser Ala Lys Trp Thr Arg Phe Tyr Tyr Lys Ser Arg Lys Phe

65 70 75 80

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

85 90 95

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

165 170 175
Thr Asp Phe Leu Gln Ala Glu Glu Lys Gly Asp Asn Lys Lys Phe Phe
180 185 190
Ala Ile Gly Gly Asn Glu Gly Ile Lys Leu Lys Asp Asn Ala Leu Ile
195 200 205
Phe Leu Ile Cys Leu Phe Leu Asp Arg Glu Glu Ala Phe Lys Phe Leu
210 215 220
Ser Arg Ala Thr Gly Phe Lys Ser Thr Lys Glu Lys Gly Phe Leu Ala
225 230 235 240

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

305 310 315 320
Pro Phe Glu Met Ile Ala Ser Leu Ser Lys Arg Val Arg Tyr Lys Asn
325 330 335
Arg Phe Pro Tyr Leu Met Leu Arg Tyr Ile Glu Glu Lys Asn Leu Leu
340 345 350

Pro Phe Ile Arg Phe Arg Ile Asp Leu Gly Cys Leu Glu Leu Ala Ser
355 360 365
Tyr Pro Lys Lys Met Gly Glu Glu Asn Asn Tyr Glu Arg Ser Val Thr
370 375 380

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

450 455 460
Ser Phe Gly Phe Ile Ser Ile Tyr Asp Leu Arg Lys Ile Leu Leu Leu
465 470 475 480
Ser Phe Leu Asp Lys Asp Lys Ala Lys Asn Ile Val Ser Gly Leu Leu
485 490 495
Glu Gln Cys Glu Lys His Trp Lys Asp Leu Ser Glu Asn Leu Phe Asp
500 505 510
Ala Ile Arg Thr Glu Leu Gln Lys Glu Phe Pro Val Pro Leu Ile Arg
515 520 525

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

595 600 605

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

610 615 620

Pro Pro Arg Ile Gly Glu Met Ala Thr Asp Leu Ala Lys Asp Ile Ile

625 630 635 640

Arg Met Val Ile Asp Gln Gly Val Lys Gln Arg Ile Thr Ser Ala Tyr

645 650 655

Tyr Ser Glu Ile Gln Arg Cys Leu Ala Gln Tyr Ala Gly Asp Asp Asn

660 665 670

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

675 680 685

Lys Asn Gly His Pro Phe Leu Gly Lys Val Leu Arg Pro Gly Leu Gly

690 695 700

His Thr Glu Lys Leu Tyr Gln Arg Tyr Phe Glu Glu Lys Lys Glu Trp

705 710 715 720

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

725 730 735

Phe Val Asn Pro Pro Thr Gly Lys Gln Lys Glu Leu Pro Leu Ile Ile

740 745 750

Arg Asn Leu Met Lys Glu Arg Pro Glu Trp Arg Asp Trp Lys Gln Arg

755 760 765

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

770 775 780

Ile Cys Arg Leu Leu Lys Asp Lys Ile Gly Lys Glu Pro Ser Gly Lys

785 790 795 800

Leu Lys Trp Asn Glu Met Phe Lys Leu Tyr Trp Asp Lys Glu Phe Pro

805 810 815

Asn Gly Met Gln Arg Phe Tyr Arg Cys Lys Arg Arg Val Glu Val Phe

820 825 830

Asp Lys Val Val Glu Tyr Glu Tyr Ser Glu Glu Gly Gly Asn Tyr Lys

835 840 845

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

850 855 860

Ser Ser Lys Glu Lys Ser Lys Leu Gln Val Glu Asp Leu Thr Leu Ser

865 870 875 880

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

885 890 895

Arg Leu Leu Cys Glu Asp Asp Arg Leu Leu Phe Met Ala Val Arg Asp

900 905 910

Leu Tyr Asp Trp Lys Glu Ala Gln Leu Asp Leu Asp Lys Ile Asp Asn

915 920 925

Met Leu Gly Glu Pro Val Ser Val Ser Gln Val Ile Gln Leu Glu Gly

930 935 940

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

945 950 955 960

Ser Lys Leu Met Arg Tyr Cys Tyr Asp Gly Arg Val Lys Gly Leu Met

965 970 975

Pro Tyr Phe Ala Asn His Glu Ala Thr Gln Glu Gln Val Glu Met Glu

980 985 990

Leu Arg His Tyr Glu Asp His Arg Arg Arg Val Phe Asn Trp Val Phe

995 1000 1005

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

1010 1015 1020

Glu Glu Ser Gln Gly Gly Cys Glu His Arg Arg Cys Ile Asp Ala Leu

1025 1030 1035 1040

Arg Lys Ala Ser Leu Val Ser Glu Glu Glu Tyr Glu Phe Leu Val His

1045 1050 1055

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

1060 1065 1070

Lys Leu Pro Pro Asn Val Thr Ser Gly Phe Cys Glu Cys Ile Trp Ser

1075 1080 1085

Lys Tyr Lys Ala Ile Ile Cys Arg Ile Ile Pro Phe Ile Asp Pro Glu

1090 1095 1100

Arg Arg Phe Phe Gly Lys Leu Leu Glu Gln Lys
1105 1110 1115

<210> 44

<211> 953

<212> PRT

<213> Alistipes sp. ZOR0009

<400> 44

Met Ser Asn Glu Ile Gly Ala Phe Arg Glu His Gln Phe Ala Tyr Ala
1 5 10 15

Pro Gly Asn Glu Lys Gln Glu Glu Ala Thr Phe Ala Thr Tyr Phe Asn
20 25 30

Leu Ala Leu Ser Asn Val Glu Gly Met Met Phe Gly Glu Val Glu Ser
35 40 45

Asn Pro Asp Lys Ile Glu Lys Ser Leu Asp Thr Leu Pro Pro Ala Ile
50 55 60

Leu Arg Gln Ile Ala Ser Phe Ile Trp Leu Ser Lys Glu Asp His Pro
65 70 75 80

Asp Lys Ala Tyr Ser Thr Glu Glu Val Lys Val Ile Val Thr Asp Leu
85 90 95

Val Arg Arg Leu Cys Phe Tyr Arg Asn Tyr Phe Ser His Cys Phe Tyr
100 105 110

Leu Asp Thr Gln Tyr Phe Tyr Ser Asp Glu Leu Val Asp Thr Thr Ala

115 120 125

Ile Gly Glu Lys Leu Pro Tyr Asn Phe His His Phe Ile Thr Asn Arg
130 135 140

Leu Phe Arg Tyr Ser Leu Pro Glu Ile Thr Leu Phe Arg Trp Asn Glu
145 150 155 160

Gly Glu Arg Lys Tyr Glu Ile Leu Arg Asp Gly Leu Ile Phe Phe Cys
165 170 175

Cys Leu Phe Leu Lys Arg Gly Gln Ala Glu Arg Phe Leu Asn Glu Leu
180 185 190

Arg Phe Phe Lys Arg Thr Asp Glu Glu Gly Arg Ile Lys Arg Thr Ile

195 200 205

Phe Thr Lys Tyr Cys Thr Arg Glu Ser His Lys His Ile Gly Ile Glu

210 215 220

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

225 230 235 240

Val Pro Lys Val Cys Asp Gly Val Val Asp Leu Ser Lys Glu Asn Glu

245 250 255

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

260 265 270

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

275 280 285

Leu Met Arg Tyr Ile Val Asp Phe Gly Val Leu Pro Cys Ile Thr Phe

290 295 300

Lys Gln Asn Asp Tyr Ser Thr Lys Glu Gly Arg Gly Gln Phe His Tyr

305 310 315 320

Gln Asp Ala Ala Val Ala Gln Glu Glu Arg Cys Tyr Asn Phe Val Val

325 330 335

Arg Asn Gly Asn Val Tyr Tyr Ser Tyr Met Pro Gln Ala Gln Asn Val

340 345 350

Val Arg Ile Ser Glu Leu Gln Gly Thr Ile Ser Val Glu Glu Leu Arg

355 360 365

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

370 375 380

Glu Gln Tyr Leu Tyr His Leu His Leu Leu Tyr Glu Lys Ile Leu Thr

385 390 395 400

Ile Ser Gly Gln Thr Ile Lys Glu Gly Arg Val Asp Val Glu Asp Tyr

405 410 415

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

420 425 430

Glu Leu Arg Arg Glu Leu Arg Lys Leu Leu Pro Lys Arg Val Cys Asp

435 440 445

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

450 455 460

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

465 470 475 480

Gln Asn Pro Ala Leu His Ile Asp Lys Ile Lys Ser Val Ile Asp Tyr

485 490 495

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

500 505 510

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

515 520 525

Tyr Leu Val Gly Asp Tyr Asp Ser His Pro Leu Ala Leu Trp Lys Glu

530 535 540

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

545 550 555 560

Ala Thr Ser Leu His Gly Leu Tyr Met Leu Cys Leu Lys Gly Thr Val

565 570 575

Glu Trp Cys Arg Lys Gln Leu Met Ser Ile Gly Lys Gly Thr Ala Lys

580 585 590

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

595 600 605

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

610 615 620

Met His Gly Tyr Ala Ala Ala Ala Thr Pro Lys Pro Lys Ala Gln Ala

625 630 635 640

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

645 650 655

Lys Arg Glu Met Ser Phe Ala Ala Phe Ile Arg Gln Asp Lys Lys Ala

660 665 670

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

675 680 685

Leu Gln Lys Arg Gln Ala Ala Ala Asp Ala Ala Cys Lys Lys Leu Tyr

690 695 700

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

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

835 840 845
Ala Ala Thr Met Met Leu Leu Asp Val Glu Arg Lys Leu Leu Ser Phe
850 855 860
Leu Leu Asp Glu Arg Arg Lys Ala Ile Ile Glu Tyr Gly Lys Phe Ile
865 870 875 880
Pro Phe Ser Ala Leu Val Lys Glu Cys Arg Leu Ala Asp Ala Gly Leu
885 890 895
Cys Gly Ile Arg Asn Asp Val Leu His Asp Asn Val Ile Ser Tyr Ala
900 905 910

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

<210> 45

<211> 1322

<212> PRT

<213> Prevotella sp. MA2016

<400> 45

Met Ser Lys Glu Cys Lys Lys Gln Arg Gln Glu Lys Lys Arg Arg Leu
1 5 10 15

Gln Lys Ala Asn Phe Ser Ile Ser Leu Thr Gly Lys His Val Phe Gly
20 25 30

Ala Tyr Phe Asn Met Ala Arg Thr Asn Phe Val Lys Thr Ile Asn Tyr
35 40 45

Ile Leu Pro Ile Ala Gly Val Arg Gly Asn Tyr Ser Glu Asn Gln Ile
50 55 60

Asn Lys Met Leu His Ala Leu Phe Leu Ile Gln Ala Gly Arg Asn Glu
65 70 75 80

Glu Leu Thr Thr Glu Gln Lys Gln Trp Glu Lys Lys Leu Arg Leu Asn
85 90 95

Pro Glu Gln Gln Thr Lys Phe Gln Lys Leu Leu Phe Lys His Phe Pro
100 105 110

Val Leu Gly Pro Met Met Ala Asp Val Ala Asp His Lys Ala Tyr Leu
115 120 125

Asn Lys Lys Lys Ser Thr Val Gln Thr Glu Asp Glu Thr Phe Ala Met
130 135 140

Leu Lys Gly Val Ser Leu Ala Asp Cys Leu Asp Ile Ile Cys Leu Met
145 150 155 160

Ala Asp Thr Leu Thr Glu Cys Arg Asn Phe Tyr Thr His Lys Asp Pro
165 170 175

Tyr Asn Lys Pro Ser Gln Leu Ala Asp Gln Tyr Leu His Gln Glu Met
180 185 190

Ile Ala Lys Lys Leu Asp Lys Val Val Val Ala Ser Arg Arg Ile Leu
195 200 205

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

210 215 220

Ile Asp His Leu His Gln Glu Val Leu Lys Asp Glu Phe Gly Asn Ala

225 230 235 240

Lys Val Lys Asp Gly Lys Val Met Lys Thr Phe Val Glu Tyr Asp Asp

245 250 255

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

260 265 270

Thr Thr Lys Asp Asp Lys Pro Val Asn Val Asn Thr Met Leu Pro Ala

275 280 285

Leu Ser Asp Phe Gly Leu Leu Tyr Phe Cys Val Leu Phe Leu Ser Lys

290 295 300

Pro Tyr Ala Lys Leu Phe Ile Asp Glu Val Arg Leu Phe Glu Tyr Ser

305 310 315 320

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

325 330 335

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

340 345 350

Lys Ala Thr Leu Ala Met Asp Ile Phe Gly Glu Leu Arg Arg Cys Pro

355 360 365

Met Glu Leu Tyr Asn Leu Leu Asp Lys Asn Ala Gly Gln Pro Phe Phe

370 375 380

His Asp Glu Val Lys His Pro Asn Ser His Thr Pro Asp Val Ser Lys

385 390 395 400

Arg Leu Arg Tyr Asp Asp Arg Phe Pro Thr Leu Ala Leu Arg Tyr Ile

405 410 415

Asp Glu Thr Glu Leu Phe Lys Arg Ile Arg Phe Gln Leu Gln Leu Gly

420 425 430

Ser Phe Arg Tyr Lys Phe Tyr Asp Lys Glu Asn Cys Ile Asp Gly Arg

435 440 445

Val Arg Val Arg Arg Ile Gln Lys Glu Ile Asn Gly Tyr Gly Arg Met

450 455 460

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

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

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

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

Arg Tyr Leu Ala Gln Ser Met Met Glu Trp Gln Pro Thr Lys Leu Lys
725 730 735

Asp Lys Glu Lys Gly His Asp Lys Leu Thr Gly Leu Asn Tyr Asn Val
740 745 750

Leu Thr Ala Tyr Leu Ala Thr Tyr Gly His Pro Gln Val Pro Glu Glu
755 760 765

Gly Phe Thr Pro Arg Thr Leu Glu Gln Val Leu Ile Asn Ala His Leu
770 775 780

Ile Gly Gly Ser Asn Pro His Pro Phe Ile Asn Lys Val Leu Ala Leu
785 790 795 800

Gly Asn Arg Asn Ile Glu Glu Leu Tyr Leu His Tyr Leu Glu Glu Glu

805 810 815

Leu Lys His Ile Arg Ser Arg Ile Gln Ser Leu Ser Ser Asn Pro Ser
820 825 830

Asp Lys Ala Leu Ser Ala Leu Pro Phe Ile His His Asp Arg Met Arg
835 840 845

Tyr His Glu Arg Thr Ser Glu Glu Met Met Ala Leu Ala Ala Arg Tyr
850 855 860

Thr Thr Ile Gln Leu Pro Asp Gly Leu Phe Thr Pro Tyr Ile Leu Glu
865 870 875 880

Ile Leu Gln Lys His Tyr Thr Glu Asn Ser Asp Leu Gln Asn Ala Leu
885 890 895

Ser Gln Asp Val Pro Val Lys Leu Asn Pro Thr Cys Asn Ala Ala Tyr
900 905 910

Leu Ile Thr Leu Phe Tyr Gln Thr Val Leu Lys Asp Asn Ala Gln Pro
915 920 925

Phe Tyr Leu Ser Asp Lys Thr Tyr Thr Arg Asn Lys Asp Gly Glu Lys
930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

Gly Asp Gly Asn Pro Val Pro Glu Val Gly Glu Lys Gly Lys Pro Ala

995 1000 1005

Thr Asp Ser Gln Gly Asn Thr Ile Trp Lys Arg Arg Ile Tyr Ser Glu

1010 1015 1020

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

1025 1030 1035 1040

Phe Lys Gly Glu Trp Glu Lys Leu Pro Arg Trp Lys Gln Asp Lys Ile

1045 1050 1055

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

1060 1065 1070

Gln Arg Met Pro Arg Tyr Ile Arg Asp Ile Lys Asp Asn Glu Arg Thr

1075 1080 1085

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

1090 1095 1100

Lys Met Phe Thr Asn Ile Ile Ser Glu Gln Ser Ser Glu Phe Asn Trp

1105 1110 1115 1120

Lys Gln Met Arg Leu Ser Lys Val Cys Asn Glu Ala Phe Leu Arg Gln

1125 1130 1135

Thr Leu Thr Phe Arg Val Pro Val Thr Val Gly Glu Thr Thr Ile Tyr

1140 1145 1150

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

1155 1160 1165

Phe Leu Thr Asp Asp Arg Leu Met Ser Leu Leu Asn Asn Ile Val Glu

1170 1175 1180

Thr Leu Lys Pro Asn Glu Asn Gly Asp Leu Val Ile Arg His Thr Asp

1185 1190 1195 1200

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

1205 1210 1215

Met Leu Ile Gln Ser Ile Glu Asn Leu Ile Ile Thr Asn Asn Ala Val

1220 1225 1230

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

1235 1240 1245

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

1250 1255 1260

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

1265 1270 1275 1280

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

1285 1290 1295

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

1300 1305 1310

Gln Ser Met Leu Gly Val Glu Ile Thr Lys

1315 1320

<210> 46

<211> 1322

<212> PRT

<213> Prevotella sp. MA2016

<400> 46

Met Ser Lys Glu Cys Lys Lys Gln Arg Gln Glu Lys Lys Arg Arg Leu

1 5 10 15

Gln Lys Ala Asn Phe Ser Ile Ser Leu Thr Gly Lys His Val Phe Gly

20 25 30

Ala Tyr Phe Asn Met Ala Arg Thr Asn Phe Val Lys Thr Ile Asn Tyr

35 40 45

Ile Leu Pro Ile Ala Gly Val Arg Gly Asn Tyr Ser Glu Asn Gln Ile

50 55 60

Asn Lys Met Leu His Ala Leu Phe Leu Ile Gln Ala Gly Arg Asn Glu

65 70 75 80

Glu Leu Thr Thr Glu Gln Lys Gln Trp Glu Lys Lys Leu Arg Leu Asn

85 90 95

Pro Glu Gln Gln Thr Lys Phe Gln Lys Leu Leu Phe Lys His Phe Pro

100 105 110

Val Leu Gly Pro Met Met Ala Asp Val Ala Asp His Lys Ala Tyr Leu

115 120 125

Asn Lys Lys Lys Ser Thr Val Gln Thr Glu Asp Glu Thr Phe Ala Met

130 135 140

Leu Lys Gly Val Ser Leu Ala Asp Cys Leu Asp Ile Ile Cys Leu Met

145 150 155 160

Ala Asp Thr Leu Thr Glu Cys Arg Asn Phe Tyr Thr His Lys Asp Pro

165 170 175

Tyr Asn Lys Pro Ser Gln Leu Ala Asp Gln Tyr Leu His Gln Glu Met

180 185 190

Ile Ala Lys Lys Leu Asp Lys Val Val Val Ala Ser Arg Arg Ile Leu

195 200 205

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

210 215 220

Ile Asp His Leu His Gln Glu Val Leu Lys Asp Glu Phe Gly Asn Ala

225 230 235 240

Lys Val Lys Asp Gly Lys Val Met Lys Thr Phe Val Glu Tyr Asp Asp

245 250 255

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

260 265 270

Thr Thr Lys Asp Asp Lys Pro Val Asn Val Asn Thr Met Leu Pro Ala

275 280 285

Leu Ser Asp Phe Gly Leu Leu Tyr Phe Cys Val Leu Phe Leu Ser Lys

290 295 300

Pro Tyr Ala Lys Leu Phe Ile Asp Glu Val Arg Leu Phe Glu Tyr Ser

305 310 315 320

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

325 330 335

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

340 345 350

Lys Ala Thr Leu Ala Met Asp Ile Phe Gly Glu Leu Arg Arg Cys Pro

355 360 365

Met Glu Leu Tyr Asn Leu Leu Asp Lys Asn Ala Gly Gln Pro Phe Phe

370 375 380

His Asp Glu Val Lys His Pro Asn Ser His Thr Pro Asp Val Ser Lys

385 390 395 400

Arg Leu Arg Tyr Asp Asp Arg Phe Pro Thr Leu Ala Leu Arg Tyr Ile

405 410 415

Asp Glu Thr Glu Leu Phe Lys Arg Ile Arg Phe Gln Leu Gln Leu Gly

420 425 430

Ser Phe Arg Tyr Lys Phe Tyr Asp Lys Glu Asn Cys Ile Asp Gly Arg

435 440 445

Val Arg Val Arg Arg Ile Gln Lys Glu Ile Asn Gly Tyr Gly Arg Met

450 455 460

Gln Glu Val Ala Asp Lys Arg Met Asp Lys Trp Gly Asp Leu Ile Gln

465 470 475 480

Lys Arg Glu Glu Arg Ser Val Lys Leu Glu His Glu Glu Leu Tyr Ile

485 490 495

Asn Leu Asp Gln Phe Leu Glu Asp Thr Ala Asp Ser Thr Pro Tyr Val

500 505 510

Thr Asp Arg Arg Pro Ala Tyr Asn Ile His Ala Asn Arg Ile Gly Leu

515 520 525

Tyr Trp Glu Asp Ser Gln Asn Pro Lys Gln Tyr Lys Val Phe Asp Glu

530 535 540

Asn Gly Met Tyr Ile Pro Glu Leu Val Val Thr Glu Asp Lys Lys Ala

545 550 555 560

Pro Ile Lys Met Pro Ala Pro Arg Cys Ala Leu Ser Val Tyr Asp Leu

565 570 575

Pro Ala Met Leu Phe Tyr Glu Tyr Leu Arg Glu Gln Gln Asp Asn Glu

580 585 590

Phe Pro Ser Ala Glu Gln Val Ile Ile Glu Tyr Glu Asp Asp Tyr Arg

595 600 605

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

610 615 620

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

690 695 700
Gln Tyr Gly Lys Lys Ser Phe Ser Asp Val Arg His Gly Ala Leu Ala
705 710 715 720
Arg Tyr Leu Ala Gln Ser Met Met Glu Trp Gln Pro Thr Lys Leu Lys
725 730 735
Asp Lys Glu Lys Gly His Asp Lys Leu Thr Gly Leu Asn Tyr Asn Val
740 745 750
Leu Thr Ala Tyr Leu Ala Thr Tyr Gly His Pro Gln Val Pro Glu Glu
755 760 765

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

835 840 845
Tyr His Glu Arg Thr Ser Glu Glu Met Met Ala Leu Ala Ala Arg Tyr
850 855 860

Thr Thr Ile Gln Leu Pro Asp Gly Leu Phe Thr Pro Tyr Ile Leu Glu
865 870 875 880
Ile Leu Gln Lys His Tyr Thr Glu Asn Ser Asp Leu Gln Asn Ala Leu
885 890 895
Ser Gln Asp Val Pro Val Lys Leu Asn Pro Thr Cys Asn Ala Ala Tyr
900 905 910

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

980 985 990
Gly Asp Gly Asn Pro Val Pro Glu Val Gly Glu Lys Gly Lys Pro Ala
995 1000 1005
Thr Asp Ser Gln Gly Asn Thr Ile Trp Lys Arg Arg Ile Tyr Ser Glu
1010 1015 1020
Val Asp Asp Tyr Ala Glu Lys Leu Thr Asp Arg Asp Met Lys Ile Ser
1025 1030 1035 1040
Phe Lys Gly Glu Trp Glu Lys Leu Pro Arg Trp Lys Gln Asp Lys Ile
1045 1050 1055

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

Lys Gln Met Arg Leu Ser Lys Val Cys Asn Glu Ala Phe Leu Arg Gln

1125 1130 1135

Thr Leu Thr Phe Arg Val Pro Val Thr Val Gly Glu Thr Thr Ile Tyr

1140 1145 1150

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

1155 1160 1165

Phe Leu Thr Asp Asp Arg Leu Met Ser Leu Leu Asn Asn Ile Val Glu

1170 1175 1180

Thr Leu Lys Pro Asn Glu Asn Gly Asp Leu Val Ile Arg His Thr Asp

1185 1190 1195 1200

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

1205 1210 1215

Met Leu Ile Gln Ser Ile Glu Asn Leu Ile Ile Thr Asn Asn Ala Val

1220 1225 1230

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

1235 1240 1245

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

1250 1255 1260

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

1265 1270 1275 1280

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

1285 1290 1295

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

1300 1305 1310

Gln Ser Met Leu Gly Val Glu Ile Thr Lys

1315 1320

<210> 47

<211> 1095

<212> PRT

<213> Riemerella anatipestifer

<400> 47

Met Glu Lys Pro Leu Leu Pro Asn Val Tyr Thr Leu Lys His Lys Phe

1 5 10 15

Phe Trp Gly Ala Phe Leu Asn Ile Ala Arg His Asn Ala Phe Ile Thr

20 25 30

Ile Cys His Ile Asn Glu Gln Leu Gly Leu Lys Thr Pro Ser Asn Asp

35 40 45

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

50 55 60

Asn Asp His Asp Leu Leu Lys Lys Ser Gln Leu Thr Glu Leu Ile Leu

65 70 75 80

Lys His Phe Pro Phe Leu Thr Ala Met Cys Tyr His Pro Pro Lys Lys

85 90 95

Glu Gly Lys Lys Lys Gly His Gln Lys Glu Gln Gln Lys Glu Lys Glu

100 105 110

Ser Glu Ala Gln Ser Gln Ala Glu Ala Leu Asn Pro Ser Lys Leu Ile

115 120 125

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

130 135 140

Tyr Ser His Tyr Lys His Lys Lys Pro Asp Ala Glu Lys Asp Ile Phe

145 150 155 160

Lys His Leu Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu

165 170 175

Asp Tyr Lys Ala His Phe Thr Val Asn Leu Thr Arg Asp Phe Ala His

180 185 190

Leu Asn Arg Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asn Arg

195 200 205

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

210 215 220

Phe Thr Asn Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys

225 230 235 240

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

245 250 255

Thr Glu Val Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu

260 265 270

Glu Ser Arg Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu

275 280 285

Leu Ser Arg Cys Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Glu Asn

290 295 300

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

305 310 315 320

Glu Glu Gln Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg

325 330 335

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

340 345 350

Ser Ile Arg Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr

355 360 365

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

370 375 380

Leu Leu Ser Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro

385 390 395 400

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

405 410 415

Asn Gln Pro Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp

420 425 430

Asn Lys Ile Gly Phe Arg Leu Gly Thr Ser Lys Glu Leu Tyr Pro Ser

435 440 445

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

450 455 460

Ser Gly Phe Val Ala His Ala Phe Ile Ser Val His Glu Leu Leu Pro

465 470 475 480

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

485 490 495

Glu Thr Val Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu

500 505 510

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

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

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

725 730 735
Glu Thr Leu Ser Glu Asp Leu Met Leu Ser Lys Pro Ile Arg Lys Glu
740 745 750

Ile Lys Lys His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu
755 760 765

Tyr Phe Lys Glu Lys Tyr Gln Asp Lys His Gln Ser Phe Tyr Asn Leu
770 775 780

Ser Tyr Lys Leu Glu Ala Lys Ala Pro Leu Leu Lys Arg Glu Glu His
785 790 795 800

Tyr Glu Tyr Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Ser Gln
805 810 815

Arg Leu Glu Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr
820 825 830

Lys Arg Trp Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln
835 840 845

Asp Val Met Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe
850 855 860

Lys Glu Leu Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala

865 870 875 880

Val Asn Val Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr
885 890 895

Leu Pro Met Val Leu Pro Val Lys Val Tyr Pro Ala Thr Ala Phe Gly
900 905 910

Glu Val Gln Tyr His Lys Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu
915 920 925

Glu His Thr Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys
930 935 940

Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp
945 950 955 960

Thr Gln Lys His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu
965 970 975

Ile Tyr Gln Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Ser Leu
980 985 990

Glu Glu Lys Leu Leu Asn Lys His Thr Ser Leu Ser Ser Leu Glu Asn
995 1000 1005

Glu Phe Arg Ala Leu Leu Glu Glu Trp Lys Lys Glu Tyr Ala Ala Ser

1010 1015 1020

Ser Met Val Thr Asp Glu His Ile Ala Phe Ile Ala Ser Val Arg Asn

1025 1030 1035 1040

Ala Phe Cys His Asn Gln Tyr Pro Phe Tyr Lys Glu Ala Leu His Ala

1045 1050 1055

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

1060 1065 1070

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

1075 1080 1085

Glu Ile Val Lys Ser Gln Ile

1090 1095

<210> 48

<211> 1124

<212> PRT

<213> Prevotella aurantiaca

<400> 48

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

1 5 10 15

Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val

20 25 30

Tyr Ile Thr Ile Asn His Ile Asn Lys Leu Leu Glu Ile Arg Glu Ile

35 40 45

Asp Asn Asp Glu Lys Val Leu Asp Ile Lys Thr Leu Trp Gln Lys Gly

50 55 60

Asn Lys Asp Leu Asn Gln Lys Ala Arg Leu Arg Glu Leu Met Thr Lys

65 70 75 80

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

85 90 95

Lys Lys Glu Val Lys Gln Glu Lys Gln Ala Glu Ala Gln Ser Leu Glu

100 105 110

Ser Leu Lys Asp Cys Leu Phe Leu Phe Leu Asp Lys Leu Gln Glu Ala

115 120 125

Arg Asn Tyr Tyr Ser His Tyr Lys Tyr Ser Glu Phe Ser Lys Glu Pro

130 135 140

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

145 150 155 160

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

165 170 175

Pro Asp Glu Asp Phe Lys His Leu Asp Arg Lys Gly Gln Phe Lys Tyr

180 185 190

Ser Phe Ala Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu Phe

195 200 205

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

210 215 220

Lys Leu Asn Gly Phe Lys Asp Asn Leu Glu Asn Lys Lys Lys Met Thr

225 230 235 240

His Glu Val Phe Cys Arg Ser Arg Ile Leu Met Pro Lys Leu Arg Leu

245 250 255

Glu Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn Glu

260 265 270

Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Gln Gly Asp Asp

275 280 285

Arg Glu Lys Phe Lys Val Pro Phe Asp Pro Ala Asp Glu Asp Tyr Asn

290 295 300

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

305 310 315 320

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

325 330 335

Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile Tyr

340 345 350

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

355 360 365

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

370 375 380

Asp Glu Trp Lys Ala Ile Val Lys Asp Leu Asp Thr Tyr Glu Thr Ser

385 390 395 400

Asn Lys Arg Tyr Ile Ser Glu Thr Thr Pro His Tyr His Leu Glu Asn

405 410 415

Gln Lys Ile Gly Ile Arg Phe Arg Asn Gly Asn Lys Glu Ile Trp Pro

420 425 430

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

435 440 445

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

450 455 460

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

465 470 475 480

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

485 490 495

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

500 505 510

Asn Ile Asp Asp Leu Glu Lys Tyr Cys Ala Asp Lys Gly Ile Pro Lys

515 520 525

Arg His Leu Pro Lys Gln Met Val Ala Ile Leu Tyr Asp Glu His Lys

530 535 540

Asp Met Val Lys Glu Ala Lys Arg Lys Gln Lys Glu Met Val Lys Asp

545 550 555 560

Thr Lys Lys Leu Leu Ala Thr Leu Glu Lys Gln Thr Gln Lys Glu Lys

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

690 695 700
Pro Lys Thr Asn Arg Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe
705 710 715 720
Asn Leu Pro Arg Gly Ile Phe Thr Glu Pro Ile Arg Glu Trp Phe Lys
725 730 735
Arg His Gln Asn Asn Ser Lys Glu Tyr Glu Lys Val Glu Ala Leu Asp
740 745 750
Arg Val Gly Leu Val Thr Lys Val Ile Pro Leu Phe Phe Lys Glu Glu
755 760 765

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

835 840 845
Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu Arg
850 855 860

Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Lys Glu Leu
865 870 875 880
Ile Asp Lys Leu Lys Ile Asp Glu Leu Asn Ile Glu Glu Leu Lys Lys
885 890 895
Leu Arg Leu Asn Asn Ile Asp Thr Asp Thr Ala Lys Lys Glu Lys Asn
900 905 910

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

980 985 990
Arg Val Glu Tyr Glu Leu Gly Glu Tyr Gln Glu Ala Arg Ile Glu Ile
995 1000 1005
Ile Gln Asp Met Leu Ala Leu Glu Glu Lys Leu Ile Asn Lys Tyr Lys
1010 1015 1020
Asp Leu Pro Thr Asn Lys Phe Ser Glu Met Leu Asn Ser Trp Leu Glu
1025 1030 1035 1040
Gly Lys Asp Glu Ala Asp Lys Ala Arg Phe Gln Asn Asp Val Asp Phe
1045 1050 1055

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

Glu Thr Lys Glu

<210> 49

<211> 1151

<212> PRT

<213> Prevotella saccharolytica

<400> 49

Met Glu Asp Lys Pro Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His
1 5 10 15

Asn Val Tyr Leu Thr Val Asn His Ile Asn Lys Leu Leu Asp Leu Glu
20 25 30

Lys Leu Tyr Asp Glu Gly Lys His Lys Glu Ile Phe Glu Arg Glu Asp
35 40 45

Ile Phe Asn Ile Ser Asp Asp Val Met Asn Asp Ala Asn Ser Asn Gly
50 55 60

Lys Lys Arg Lys Leu Asp Ile Lys Lys Ile Trp Asp Asp Leu Asp Thr
65 70 75 80

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

Pro Phe Ile Gln Pro Ala Ile Ile Gly Ala Gln Thr Lys Glu Arg Thr
100 105 110

Thr Ile Asp Lys Asp Lys Arg Ser Thr Ser Thr Ser Asn Asp Ser Leu
115 120 125

Lys Gln Thr Gly Glu Gly Asp Ile Asn Asp Leu Leu Ser Leu Ser Asn
130 135 140

Val Lys Ser Met Phe Phe Arg Leu Leu Gln Ile Leu Glu Gln Leu Arg
145 150 155 160

Asn Tyr Tyr Ser His Val Lys His Ser Lys Ser Ala Thr Met Pro Asn
165 170 175

Phe Asp Glu Asp Leu Leu Asn Trp Met Arg Tyr Ile Phe Ile Asp Ser
180 185 190

Val Asn Lys Val Lys Glu Asp Tyr Ser Ser Asn Ser Val Ile Asp Pro

195 200 205

Asn Thr Ser Phe Ser His Leu Ile Tyr Lys Asp Glu Gln Gly Lys Ile

210 215 220

Lys Pro Cys Arg Tyr Pro Phe Thr Ser Lys Asp Gly Ser Ile Asn Ala

225 230 235 240

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

245 250 255

Ile Trp Met Gln Lys Lys Ile Pro Gly Phe Lys Lys Ala Ser Glu Asn

260 265 270

Tyr Met Lys Met Thr Asn Glu Val Phe Cys Arg Asn His Ile Leu Leu

275 280 285

Pro Lys Ile Arg Leu Glu Thr Val Tyr Asp Lys Asp Trp Met Leu Leu

290 295 300

Asp Met Leu Asn Glu Val Val Arg Cys Pro Leu Ser Leu Tyr Lys Arg

305 310 315 320

Leu Thr Pro Ala Ala Gln Asn Lys Phe Lys Val Pro Glu Lys Ser Ser

325 330 335

Asp Asn Ala Asn Arg Gln Glu Asp Asp Asn Pro Phe Ser Arg Ile Leu

340 345 350

Val Arg His Gln Asn Arg Phe Pro Tyr Phe Val Leu Arg Phe Phe Asp

355 360 365

Leu Asn Glu Val Phe Thr Thr Leu Arg Phe Gln Ile Asn Leu Gly Cys

370 375 380

Tyr His Phe Ala Ile Cys Lys Lys Gln Ile Gly Asp Lys Lys Glu Val

385 390 395 400

His His Leu Ile Arg Thr Leu Tyr Gly Phe Ser Arg Leu Gln Asn Phe

405 410 415

Thr Gln Asn Thr Arg Pro Glu Glu Trp Asn Thr Leu Val Lys Thr Thr

420 425 430

Glu Pro Ser Ser Gly Asn Asp Gly Lys Thr Val Gln Gly Val Pro Leu

435 440 445

Pro Tyr Ile Ser Tyr Thr Ile Pro His Tyr Gln Ile Glu Asn Glu Lys

450 455 460

Ile Gly Ile Lys Ile Phe Asp Gly Asp Thr Ala Val Asp Thr Asp Ile

465 470 475 480

Trp Pro Ser Val Ser Thr Glu Lys Gln Leu Asn Lys Pro Asp Lys Tyr

485 490 495

Thr Leu Thr Pro Gly Phe Lys Ala Asp Val Phe Leu Ser Val His Glu

500 505 510

Leu Leu Pro Met Met Phe Tyr Tyr Gln Leu Leu Leu Cys Glu Gly Met

515 520 525

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

530 535 540

Arg Asn Ala Ile Phe Asn Leu Tyr Asp Ala Phe Val Gln Glu Lys Ile

545 550 555 560

Asn Thr Ile Thr Asp Leu Glu Asn Tyr Leu Gln Asp Lys Pro Ile Leu

565 570 575

Ile Gly His Leu Pro Lys Gln Met Ile Asp Leu Leu Lys Gly His Gln

580 585 590

Arg Asp Met Leu Lys Ala Val Glu Gln Lys Lys Ala Met Leu Ile Lys

595 600 605

Asp Thr Glu Arg Arg Leu Lys Leu Leu Asp Lys Gln Leu Lys Gln Glu

610 615 620

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

625 630 635 640

Ile Ala Asp Trp Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Lys

645 650 655

Arg Asp Lys Glu Gly Asn Pro Ile Asn Cys Ser Lys Ala Asn Ser Thr

660 665 670

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

675 680 685

Cys Arg Leu Ser Arg Tyr Phe Thr Gln Leu His Leu Ile His Ser Asp

690 695 700

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

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

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

915 920 925
Lys Leu Asp Ser Leu Gln Thr Asn Thr Ser Thr Ala Gly Ser Leu Asn
930 935 940

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

Asn Phe Lys Thr Leu Leu Ala Asp Arg Arg Ile Lys Gly Leu Phe Ser
995 1000 1005
Tyr Ile Glu His Asp Asp Ile Asp Leu Lys Gln His Pro Leu Thr Lys
1010 1015 1020
Arg Arg Val Asp Leu Glu Leu Asp Leu Tyr Gln Thr Cys Arg Ile Asp
1025 1030 1035 1040
Ile Phe Gln Gln Thr Leu Gly Leu Glu Ala Gln Leu Leu Asp Lys Tyr
1045 1050 1055
Ser Asp Leu Asn Thr Asp Asn Phe Tyr Gln Met Leu Ile Gly Trp Arg

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

Val Glu Lys Val Val Asn Arg Ile Lys Arg Ile Glu Leu Phe Asp
1140 1145 1150

<210> 50
<211> 1159
<212> PRT
<213> Myroides odoratimimus
<400> 50
Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu

65 70 75 80

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

85 90 95

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

100 105 110

Ser Gln Arg Ile Arg Gln Phe Arg Glu Met Leu Ile Ser Leu Val Thr

115 120 125

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Asp

130 135 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Phe Val

145 150 155 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr

165 170 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile

180 185 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys

195 200 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe

210 215 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Asp Lys Glu Thr Val

225 230 235 240

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

245 250 255

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

325 330 335
Leu Leu Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val
340 345 350
Tyr Gln His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp
355 360 365
Asn Glu Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu
370 375 380
Ser Arg Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe
385 390 395 400

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

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

Lys Leu Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala
515 520 525
Arg Lys Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys
530 535 540

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

610 615 620
Leu Lys Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys
625 630 635 640
Ile Thr Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile
645 650 655
Leu Glu Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr
660 665 670
Lys Phe Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn
675 680 685

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

```
        755                 760                 765
Lys Tyr Leu Glu Ala Arg Leu Glu Tyr Ile Glu Asn Val Ile Thr Arg
    770                 775                 780
Val Lys Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu
785                 790                 795                 800
Cys Phe Thr Phe Leu Lys Lys Ser Asn Tyr Thr Val Val Ser Leu Asp
                805                 810                 815
Lys Gln Val Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly
            820                 825                 830

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

            900                 905                 910
Val Leu Lys Leu Ser Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr
        915                 920                 925
Gln Thr Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr
    930                 935                 940
Gln Glu Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln
945                 950                 955                 960
Leu Cys Asp Gly Leu Val His Ile Asp Asn Val Lys Leu Lys Asp Ile
                965                 970                 975

Gly Asn Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu
            980                 985                 990
Thr Tyr Gln Ser Asp Ile Val Trp Ser Ala Tyr Leu Ser Asn Glu Val
        995                1000                1005
Asp Ser Asn Lys Leu Tyr Val Ile Glu Arg Gln Leu Asp Asn Tyr Glu
```

1010 1015 1020

Ser Ile Arg Ser Lys Glu Leu Leu Lys Glu Val Gln Glu Ile Glu Cys

1025 1030 1035 1040

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

1045 1050 1055

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

1060 1065 1070

Met Asp Val Arg Glu Met Leu Ile Leu Ser Thr Asp Val Lys Phe Lys

1075 1080 1085

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

1090 1095 1100

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

1105 1110 1115 1120

Pro Ile Lys Glu Phe Phe Asp Phe Cys Glu Asn Asn Tyr Arg Ser Ile

1125 1130 1135

Ser Asp Asn Glu Tyr Tyr Ala Glu Tyr Tyr Met Glu Ile Phe Arg Ser

1140 1145 1150

Ile Lys Glu Lys Tyr Ala Asn

1155

<210> 51

<211> 1199

<212> PRT

<213> Capnocytophaga canimorsus

<400> 51

Met Lys Asn Ile Gln Arg Leu Gly Lys Gly Asn Glu Phe Ser Pro Phe

1 5 10 15

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

20 25 30

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

35 40 45

Val Ile Thr Asp Glu Asn Lys Lys Pro Lys Glu Thr Phe Gly Glu Lys

50 55 60

Ile Leu Asn Glu Ile Phe Lys Lys Asp Ile Ser Ile Val Asp Tyr Glu
65 70 75 80
Lys Trp Val Asn Ile Phe Ala Asp Tyr Phe Pro Phe Thr Lys Tyr Leu

85 90 95
Ser Leu Tyr Leu Glu Glu Met Gln Phe Lys Asn Arg Val Ile Cys Phe
100 105 110
Arg Asp Val Met Lys Glu Leu Leu Lys Thr Val Glu Ala Leu Arg Asn
115 120 125
Phe Tyr Thr His Tyr Asp His Glu Pro Ile Lys Ile Glu Asp Arg Val
130 135 140
Phe Tyr Phe Leu Asp Lys Val Leu Leu Asp Val Ser Leu Thr Val Lys
145 150 155 160

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

225 230 235 240
Glu Pro Gln Asn Lys Lys Gln Lys Ser Ser Ala Thr Val Trp Glu Leu
245 250 255
Cys Ser Lys Ser Ser Ser Lys Tyr Thr Glu Lys Ser Phe Pro Asn Arg
260 265 270
Glu Asn Asp Lys His Cys Leu Glu Val Pro Ile Ser Gln Lys Gly Ile
275 280 285
Val Phe Leu Leu Ser Phe Phe Leu Asn Lys Gly Glu Ile Tyr Ala Leu
290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

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

355 360 365

Glu Thr Leu Met Leu Gln Met Leu Asp Glu Leu Asn Lys Val Pro Asp

370 375 380

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

385 390 395 400

Asp Trp Asn Glu Tyr Leu Lys Glu Asn Asn Gly Asp Val Gly Thr Met

405 410 415

Glu Glu Glu Gln Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp

420 425 430

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

435 440 445

Pro Thr Leu Arg Phe Gln Val His Leu Gly Asn Tyr Leu Cys Asp Lys

450 455 460

Arg Thr Lys Gln Ile Cys Asp Thr Thr Thr Glu Arg Glu Val Lys Lys

465 470 475 480

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

485 490 495

Ile Phe Leu Asn Glu Arg Glu Glu Ile Lys Gly Trp Glu Val Phe Pro

500 505 510

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

515 520 525

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

530 535 540

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

545 550 555 560

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

565 570 575

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

580 585 590

Ile Val Ser Thr Asn Ser Asn Glu Gln Gly Glu Pro Val Val Phe Ile

595 600 605

Gly Gln Pro Thr Ala Tyr Leu Ser Met Asn Asp Ile His Ser Val Leu

610 615 620

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

625 630 635 640

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

645 650 655

Ala Thr Thr Lys Ile Leu Lys Pro Tyr Thr Asn Ala Asn Ser Asn Ser

660 665 670

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

675 680 685

Leu Lys Thr Leu Leu Glu Glu Gln Gln Gln Arg Glu Lys Asp Lys Lys

690 695 700

Asp Lys Lys Ser Lys Arg Lys His Glu Leu Tyr Pro Ser Glu Lys Gly

705 710 715 720

Lys Val Ala Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met Pro Lys

725 730 735

Ala Phe Lys Glu Gln Trp Arg Gly Tyr His His Ser Leu Leu Gln Lys

740 745 750

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

755 760 765

Pro Lys Glu Val Phe Lys His Phe Pro Phe Lys Leu Lys Gly Tyr Phe

770 775 780

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

785 790 795 800

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

805 810 815

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

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

945 950 955 960
Ala Glu Arg Gln Gln Asn Asn Thr His Ala Ala Thr Thr Gln Glu Arg
965 970 975
Asn Leu Asn Gly Ile Leu Asn Gln Pro Lys Asp Ile Lys Ile Gln Gly
980 985 990
Lys Ile Thr Val Lys Gly Val Lys Leu Lys Asp Ile Gly Asn Phe Arg
995 1000 1005
Lys Tyr Glu Ile Asp Gln Arg Val Asn Thr Phe Leu Asp Tyr Glu Pro
1010 1015 1020

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

Leu Glu Lys Ile Ile Ser Asp Glu Ile Lys Glu Glu His Arg His Asp

1075 1080 1085

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

1090 1095 1100

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

1105 1110 1115 1120

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

1125 1130 1135

Glu Ala Thr Asp Leu Glu Gln Lys Ala Phe Val Leu Thr Tyr Ile Arg

1140 1145 1150

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

1155 1160 1165

Cys Gln Glu Lys Tyr Gly Lys Ile Glu Lys Glu Lys Thr Tyr Ala Glu

1170 1175 1180

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

1185 1190 1195

<210> 52

<211> 1175

<212> PRT

<213> Porphyromonas gulae

<400> 52

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu

1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn

20 25 30

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

35 40 45

Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp

50 55 60

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

65 70 75 80

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

85 90 95

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

165 170 175
Val Ser Val Gln Arg Val Lys Ile Asp His Glu His Asn Asp Glu Val
180 185 190
Asp Pro His Tyr His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205
Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
210 215 220
Gly Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240

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

305 310 315 320
Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
325 330 335

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

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

450 455 460
Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465 470 475 480
Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
485 490 495
Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
500 505 510
Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
515 520 525

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

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

595 600 605

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

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Ala Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

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

705 710 715 720

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

725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly His Asp Glu Val Ala Ser Tyr Lys Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro

785 790 795 800

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

805 810 815

Gly Arg Phe Leu Ser Lys Glu Glu Arg Ala Glu Glu Trp Glu Arg Gly

820 825 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser Tyr Ser Ala Ala Arg

835 840 845

Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn

850 855 860

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

865 870 875 880

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

885 890 895

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

900 905 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln

915 920 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys

930 935 940

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

945 950 955 960

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

965 970 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val

980 985 990

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

995 1000 1005

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

1010 1015 1020

Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Gly Leu Ala Met

1025 1030 1035 1040

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

1045 1050 1055

Tyr Gln Thr Ala Arg Val Cys Val Phe Glu Leu Thr Leu Arg Leu Glu

1060 1065 1070

Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro Asp Glu Ser Phe Arg

1075 1080 1085

Glu Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Ala Lys Trp Pro Glu

1090 1095 1100

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

1170 1175

<210> 53
<211> 1090
<212> PRT
<213> Prevotella sp. P5-125
<400> 53
Met Asn Ile Pro Ala Leu Val Glu Asn Gln Lys Lys Tyr Phe Gly Thr
1 5 10 15
Tyr Ser Val Met Ala Met Leu Asn Ala Gln Thr Val Leu Asp His Ile
20 25 30
Gln Lys Val Ala Asp Ile Glu Gly Glu Gln Asn Glu Asn Asn Glu Asn
35 40 45
Leu Trp Phe His Pro Val Met Ser His Leu Tyr Asn Ala Lys Asn Gly

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

115 120 125

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

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

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

340 345 350
Ile Arg Phe His Val Asn Met Gly Lys Leu Arg Tyr Leu Leu Lys Ala
355 360 365

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

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

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

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

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

625 630 635 640

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

645 650 655

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

660 665 670

Tyr Lys Val Phe Ala Arg Ser Ile Pro Ala Asn Ala Val Glu Phe Tyr

675 680 685

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

690 695 700

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

705 710 715 720

Gln Asn Lys Trp Lys Thr Pro Ala Met Lys Thr Leu Gly Arg Ile Tyr

725 730 735

Ser Glu Asp Leu Pro Val Glu Leu Pro Arg Gln Met Phe Asp Asn Glu

740 745 750

Ile Lys Ser His Leu Lys Ser Leu Pro Gln Met Glu Gly Ile Asp Phe

755 760 765

Asn Asn Ala Asn Val Thr Tyr Leu Ile Ala Glu Tyr Met Lys Arg Val

770 775 780

Leu Asp Asp Asp Phe Gln Thr Phe Tyr Gln Trp Asn Arg Asn Tyr Arg

785 790 795 800

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

805 810 815

His Cys Phe Thr Ser Val Glu Glu Arg Glu Gly Leu Trp Lys Glu Arg

820 825 830

Ala Ser Arg Thr Glu Arg Tyr Arg Lys Gln Ala Ser Asn Lys Ile Arg

835 840 845

Ser Asn Arg Gln Met Arg Asn Ala Ser Ser Glu Glu Ile Glu Thr Ile

850 855 860

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

865 870 875 880

Lys Val Ile Arg Arg Tyr Arg Val Gln Asp Ala Leu Leu Phe Leu Leu

885 890 895

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

900 905 910

Lys Leu Lys Glu Ile Met Pro Asp Ala Glu Lys Gly Ile Leu Ser Glu

915 920 925

Ile Met Pro Met Ser Phe Thr Phe Glu Lys Gly Gly Lys Lys Tyr Thr

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

Gln Cys Arg Pro Glu Ile Ser Ser Ile Val Phe Asn Leu Glu Lys Trp

995 1000 1005

Ala Phe Asp Thr Tyr Pro Glu Leu Ser Ala Arg Val Asp Arg Glu Glu

1010 1015 1020

Lys Val Asp Phe Lys Ser Ile Leu Lys Ile Leu Leu Asn Asn Lys Asn

1025 1030 1035 1040

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

1045 1050 1055

Asp His Asn Asn Tyr Pro Asp Lys Gly Val Val Glu Ile Lys Ala Leu

1060 1065 1070

Pro Glu Ile Ala Met Ser Ile Lys Lys Ala Phe Gly Glu Tyr Ala Ile

1075 1080 1085

Met Lys

1090

<210> 54

<211> 1150

<212> PRT

<213> Flavobacterium branchiophilum

<400> 54

Met Glu Asn Leu Asn Lys Ile Leu Asp Lys Glu Asn Glu Ile Cys Ile
1 5 10 15

Ser Lys Ile Phe Asn Thr Lys Gly Ile Ala Ala Pro Ile Thr Glu Lys
20 25 30

Ala Leu Asp Asn Ile Lys Ser Lys Gln Lys Asn Asp Leu Asn Lys Glu
35 40 45

Ala Arg Leu His Tyr Phe Ser Ile Gly His Ser Phe Lys Gln Ile Asp
50 55 60

Thr Lys Lys Val Phe Asp Tyr Val Leu Ile Glu Glu Leu Lys Asp Glu
65 70 75 80

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

Phe Ser Ile Lys Leu Gln Lys Leu Ile Asn Ser Ile Arg Asn Ile Asn
100 105 110

Asn His Tyr Val His Asn Phe Asn Asp Ile Asn Leu Asn Lys Ile Asp
115 120 125

Ser Asn Val Phe His Phe Leu Lys Glu Ser Phe Glu Leu Ala Ile Ile
130 135 140

Glu Lys Tyr Tyr Lys Val Asn Lys Lys Tyr Pro Leu Asp Asn Glu Ile
145 150 155 160

Val Leu Phe Leu Lys Glu Leu Phe Ile Lys Asp Glu Asn Thr Ala Leu
165 170 175

Leu Asn Tyr Phe Thr Asn Leu Ser Lys Asp Glu Ala Ile Glu Tyr Ile
180 185 190

Leu Thr Phe Thr Ile Thr Glu Asn Lys Ile Trp Asn Ile Asn Asn Glu
195 200 205

His Asn Ile Leu Asn Ile Glu Lys Gly Lys Tyr Leu Thr Phe Glu Ala
210 215 220

Met Leu Phe Leu Ile Thr Ile Phe Leu Tyr Lys Asn Glu Ala Asn His
225 230 235 240

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

245 250 255

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

260 265 270

Glu Glu Gly His Leu Ile Lys Phe Arg Asp Met Ile Gln Tyr Leu Asn

275 280 285

His Tyr Pro Thr Ala Trp Asn Asn Asp Leu Lys Leu Glu Ser Glu Asn

290 295 300

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

305 310 315 320

Glu Leu Asn Ser Asn Tyr Pro Ser Phe Ala Thr Asp Ile Gln Phe Lys

325 330 335

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

340 345 350

Thr Ser Phe Ser Asn Lys Ser Tyr Asn Glu Glu Ile Arg His Asn Pro

355 360 365

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

370 375 380

Ser Phe Asn Val Lys Glu Asp Lys Phe Lys Ile Phe Val Lys Lys His

385 390 395 400

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

405 410 415

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

420 425 430

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

435 440 445

His Lys Leu Val Lys Ser His Gly Arg Asn Gln Asp Arg Phe Met Asp

450 455 460

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

465 470 475 480

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

485 490 495

Leu Gln Ser Asn Glu Asn Asn Asp Asp Val Lys Phe His Lys Gly Lys

500 505 510

Val Thr Thr Tyr Ile Lys Tyr Glu Glu His Leu Lys Asn Tyr Ser Tyr

515 520 525

Trp Asp Cys Pro Phe Val Glu Glu Asn Asn Ser Met Ser Val Lys Ile

530 535 540

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

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

Lys Ser Lys Tyr Lys Lys Ile Leu Pro Lys Arg Leu Leu His Asn Tyr

610 615 620

Ala Pro Ala Lys Gln Asp Lys Ala Pro Glu Asn Ala Phe Glu Thr Leu

625 630 635 640

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

645 650 655

Lys Ala Glu His Glu Lys Asn Lys Glu Asp Phe Val Lys Arg Asn Lys

660 665 670

Gly Lys Gln Phe Lys Leu His Phe Ile Arg Lys Ala Cys Gln Met Met

675 680 685

Tyr Phe Lys Glu Lys Tyr Asn Thr Leu Lys Glu Gly Asn Ala Ala Phe

690 695 700

Glu Lys Lys Asp Pro Val Ile Glu Lys Arg Lys Asn Lys Glu His Glu

705 710 715 720

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

725 730 735

Tyr Cys Lys Trp Met Phe Ala Phe Asn Gly Asn Asp Ser Tyr Lys Lys

740 745 750

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

820 825 830
Asp Lys Asn Leu Leu Asn Ser Glu Asn Asn Val Ile Gln Tyr Lys Ser
835 840 845
Leu Glu Asn Val Glu Tyr Leu Ile Ser Asp Phe Tyr Phe Gln Ser Lys
850 855 860
Leu Ser Ile Asp Gln Tyr Lys Thr Cys Gly Lys Leu Phe Asn Lys Leu
865 870 875 880
Lys Ser Asn Lys Leu Glu Asp Cys Leu Leu Tyr Glu Ile Ala Tyr Asn
885 890 895

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

965 970 975
Tyr Glu Ile Ile Ile Lys Asn Asp Leu Glu Ile Glu Asp Ile Asn Thr
980 985 990

Ile Asn Asn Lys Ile Ile Asn Asp Ser Val Lys Phe Thr Glu Val Leu
995 1000 1005

Met Glu Leu Glu Lys Tyr Phe Ile Leu Lys Asp Lys Cys Ile Leu Ser
1010 1015 1020

Lys Asn Tyr Ile Asp Asn Ser Glu Ile Pro Ser Leu Lys Gln Phe Ser
1025 1030 1035 1040

Lys Val Trp Ile Lys Glu Asn Glu Asn Glu Ile Ile Asn Tyr Arg Asn
1045 1050 1055

Ile Ala Cys His Phe His Leu Pro Leu Leu Glu Thr Phe Asp Asn Leu
1060 1065 1070

Leu Leu Asn Val Glu Gln Lys Phe Ile Lys Glu Glu Leu Gln Asn Val
1075 1080 1085

Ser Thr Ile Asn Asp Leu Ser Lys Pro Gln Glu Tyr Leu Ile Leu Leu
1090 1095 1100

Phe Ile Lys Phe Lys His Asn Asn Phe Tyr Leu Asn Leu Phe Asn Lys

1105 1110 1115 1120

Asn Glu Ser Lys Thr Ile Lys Asn Asp Lys Glu Val Lys Lys Asn Arg
1125 1130 1135

Val Leu Gln Lys Phe Ile Asn Gln Val Ile Leu Lys Lys Lys
1140 1145 1150

<210> 55

<211> 1159

<212> PRT

<213> Myroides odoratimimus

<400> 55

Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1 5 10 15

Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
20 25 30

Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg
35 40 45

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

50 55 60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu

65 70 75 80

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

85 90 95

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

100 105 110

Ser Gln Arg Ile Arg Gln Phe Arg Glu Met Leu Ile Ser Leu Val Thr

115 120 125

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Asp

130 135 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Phe Val

145 150 155 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr

165 170 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile

180 185 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys

195 200 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe

210 215 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Asp Lys Glu Thr Val

225 230 235 240

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

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

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

370 375 380
Ser Arg Val Thr His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe
385 390 395 400
Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr
405 410 415
Leu Arg Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr
420 425 430
Lys Gln Leu Gly Lys Val Glu Ser Asp Arg Ile Ile Lys Glu Lys Val
435 440 445

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

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

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

Leu Lys Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile
595 600 605
Gly Lys Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile
610 615 620
Leu Lys Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys
625 630 635 640
Ile Thr Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile
645 650 655
Leu Glu Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr
660 665 670
Lys Phe Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn
675 680 685
Ala Glu Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg
690 695 700
Phe Met Phe Lys Glu Ser Lys Ser Lys Trp Lys Gly Tyr Gln His Ile
705 710 715 720
Glu Leu Gln Lys Leu Phe Ala Tyr Phe Asp Thr Ser Lys Ser Asp Leu
725 730 735

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

805 810 815

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

820 825 830

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

835 840 845

His Lys Glu Lys Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser

850 855 860

Asn Tyr Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Thr Thr Glu

865 870 875 880

Asp Lys Arg Glu Lys Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln

885 890 895

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

900 905 910

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

915 920 925

Gln Thr Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr

930 935 940

Gln Glu Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln

945 950 955 960

Leu Cys Asp Gly Leu Val His Ile Asp Asn Val Lys Leu Lys Asp Ile

965 970 975

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

980 985 990

Thr Tyr Gln Ser Asp Ile Val Trp Ser Ala Tyr Leu Ser Asn Glu Val

995 1000 1005

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

1010 1015 1020

Ser Ile Arg Ser Lys Glu Leu Leu Lys Glu Val Gln Glu Ile Glu Cys

1025 1030 1035 1040

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

1045 1050 1055

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

1060 1065 1070

Met Asp Val Arg Glu Met Leu Ile Leu Ser Thr Asp Val Lys Phe Lys

1075 1080 1085

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

1090 1095 1100

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

1105 1110 1115 1120

Pro Ile Lys Glu Phe Phe Asp Phe Cys Glu Asn Asn Tyr Arg Ser Ile

1125 1130 1135

Ser Asp Asn Glu Tyr Tyr Ala Glu Tyr Tyr Met Glu Ile Phe Arg Ser

1140 1145 1150

Ile Lys Glu Lys Tyr Ala Asn

1155

<210> 56

<211>

1214

<212> PRT

<213> Flavobacterium columnare

<400> 56

Met Ser Ser Lys Asn Glu Ser Tyr Asn Lys Gln Lys Thr Phe Asn His

1 5 10 15

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

20 25 30

Asp Asp Asn Leu Arg Gln Val Gly Lys Glu Phe Lys Thr Arg Ile Asn

35 40 45

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

50 55 60

Asn Lys Glu Lys Ser Val Ala Lys Arg Glu His Ala Leu Asn Leu Leu

65 70 75 80

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

85 90 95

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

100 105 110

Lys Lys Leu Arg Asp Tyr Tyr Thr His His Tyr His Lys Pro Ile Thr
115 120 125

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

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

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

340 345 350
Glu Lys Gln Gln Lys Glu Phe Glu Glu Asp Ile Asn Glu Tyr Met Arg

355 360 365

Glu Asn Pro Glu Asn Glu Asp Ser Thr Phe Ser Ser Ile Val Ser His

370 375 380

Lys Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Met

385 390 395 400

Arg Phe Leu Asp Glu Tyr Ala Glu Leu Pro Thr Leu Arg Phe Met Val

405 410 415

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

420 425 430

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

435 440 445

Glu Lys Leu Ser Leu Val Thr Glu Tyr Lys Lys Asn Val Tyr Leu Lys

450 455 460

Glu Thr Ser Asn Ile Asp Leu Ser Arg Phe Pro Leu Phe Pro Asn Pro

465 470 475 480

Ser Tyr Val Met Ala Asn Asn Asn Ile Pro Phe Tyr Ile Asp Ser Arg

485 490 495

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

500 505 510

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

515 520 525

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

530 535 540

Asp Leu Gln Gln Arg Ser Thr Ile Gly Leu Leu Ser Cys Asn Glu Leu

545 550 555 560

Pro Ser Met Leu Tyr Glu Val Ile Val Lys Asp Ile Lys Gly Ala Glu

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

625 630 635 640

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

645 650 655

Asp Asn Tyr Asn Arg Asn Lys Asn Thr Tyr Lys Phe Lys Asn Gln Pro

660 665 670

Lys Asn Lys Val Asp Asp Lys Lys Leu Gln Arg Lys Tyr Val Phe Tyr

675 680 685

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

690 695 700

His Phe Met Lys Asn Lys Ser Leu Trp Lys Gly Tyr Met His Asn Glu

705 710 715 720

Leu Gln Ser Phe Leu Ala Phe Phe Glu Asp Lys Lys Asn Asp Cys Ile

725 730 735

Ala Leu Leu Glu Thr Val Phe Asn Leu Lys Glu Asp Cys Ile Leu Thr

740 745 750

Lys Gly Leu Lys Asn Leu Phe Leu Lys His Gly Asn Phe Ile Asp Phe

755 760 765

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

770 775 780

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

785 790 795 800

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

805 810 815

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

820 825 830

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

835 840 845

Met Ile Pro Phe Lys Lys Pro Asn Pro Asp Glu Phe Ala Ser Trp Phe

850 855 860

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

915 920 925
Ser Asp Phe Tyr Val Ser Lys Ala Glu Arg Glu Lys Ile Lys Ala Asp
930 935 940
Ala Lys Ala Tyr Gln Lys Leu Asn Asp Ser Ser Leu Trp Asn Lys Val
945 950 955 960
Ile His Leu Ser Leu Gln Asn Asn Arg Ile Thr Ala Asn Pro Lys Leu
965 970 975
Lys Asp Ile Gly Lys Tyr Lys Arg Ala Leu Gln Asp Glu Lys Ile Ala
980 985 990

Thr Leu Leu Thr Tyr Asp Ala Arg Thr Trp Thr Tyr Ala Leu Gln Lys
995 1000 1005
Pro Glu Lys Glu Asn Glu Asn Asp Tyr Lys Glu Leu His Tyr Thr Ala
1010 1015 1020
Leu Asn Met Glu Leu Gln Glu Tyr Glu Lys Val Arg Ser Lys Glu Leu
1025 1030 1035 1040
Leu Lys Gln Val Gln Glu Leu Glu Lys Lys Ile Leu Asp Lys Phe Tyr
1045 1050 1055
Asp Phe Ser Asn Asn Ala Ser His Pro Glu Asp Leu Glu Ile Glu Asp

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

Lys Ile Lys Asn Met Asp Glu Asp Glu Lys Ala Lys Ile Ile Asn Thr
1125 1130 1135

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

1205 1210

<210> 57
<211> 1175
<212> PRT
<213> Porphyromonas sp.
<400> 57
Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15
Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
20 25 30
Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45

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

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

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

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Glu Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp

210 215 220

Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

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

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Asp Asp Arg Ala Cys Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly

325 330 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

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

385 390 395 400

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

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460

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

465 470 475 480

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

485 490 495

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

500 505 510

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

515 520 525

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

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Lys Gln Met Ile Gly Ile Leu Ser Gln Glu Arg Lys Asp Met Glu

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

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

690 695 700
Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705 710 715 720
Ser Asp Arg Val Glu Asn Cys Pro Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
740 745 750
Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765

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

835 840 845
Arg Ile Lys Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
850 855 860

Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865 870 875 880
Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys
885 890 895
Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
900 905 910

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

980 985 990
His Lys Glu Glu Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asp
995 1000 1005
Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg
1010 1015 1020
Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Gly Leu Ala Met
1025 1030 1035 1040
Glu Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu Leu Ala Lys
1045 1050 1055

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

His Asn Gln Tyr Pro Met Tyr Asp Glu Ala Val Phe Ser Ser Ile Arg

1125 1130 1135

Lys Tyr Asp Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu

1140 1145 1150

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

1155 1160 1165

Val Glu Arg Ile Ile Gln Ala

1170 1175

<210> 58

<211> 1132

<212> PRT

<213> Prevotella intermedia

<400> 58

Met Glu Asp Asp Lys Lys Thr Lys Glu Ser Thr Asn Met Leu Asp Asn

1 5 10 15

Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val Tyr

20 25 30

Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn Lys Lys

35 40 45

Asp Gln Asp Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala Ile Lys

50 55 60

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

65 70 75 80

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

85 90 95

Thr Lys Asn Lys Glu Asp Lys Glu Glu Val Lys Gln Glu Lys Gln Ala

100 105 110

Lys Ala Gln Ser Phe Asp Ser Leu Lys His Cys Leu Phe Leu Phe Leu

115 120 125

Glu Lys Leu Gln Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys Tyr Ser

130 135 140

Glu Ser Thr Lys Glu Pro Met Leu Glu Lys Glu Leu Leu Lys Lys Met

145 150 155 160

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

165 170 175

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

180 185 190

Thr Glu Glu Glu Phe Asn Tyr Tyr Phe Thr Thr Asn Lys Lys Gly Asn

195 200 205

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

210 215 220

Lys Asp Ala Ile Trp Met Gln Gln Lys Leu Arg Gly Phe Lys Asp Asn

225 230 235 240

Arg Glu Ser Lys Lys Lys Met Thr His Glu Val Phe Cys Arg Ser Arg

245 250 255

Met Leu Leu Pro Lys Leu Arg Leu Glu Ser Thr Gln Thr Gln Asp Trp

260 265 270

Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser Leu

275 280 285

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

290 295 300

Asp Ser Ala Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe Lys Asn

305 310 315 320

Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu Arg Tyr

325 330 335

Phe Asp Tyr Asn Glu Ile Phe Thr Asn Leu Arg Phe Gln Ile Asp Leu

340 345 350

Gly Thr Tyr His Phe Ser Ile Tyr Lys Lys Leu Ile Gly Gly Gln Lys

355 360 365

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

370 375 380

Glu Phe Ala Lys Gln Asn Arg Thr Asp Glu Trp Lys Ala Ile Val Lys

385 390 395 400

Asp Phe Asp Thr Tyr Glu Thr Ser Glu Glu Pro Tyr Ile Ser Glu Thr

405 410 415

Ala Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg Phe Arg

420 425 430

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

435 440 445

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

450 455 460

Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu

465 470 475 480

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

485 490 495

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

500 505 510

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

515 520 525

Cys Glu Asp Lys Gly Ile Pro Lys Arg His Leu Pro Lys Gln Met Val

530 535 540

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

545 550 555 560

Lys Gln Lys Glu Met Val Lys Asp Thr Lys Lys Leu Leu Ala Thr Leu

565 570 575

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

580 585 590

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

595 600 605

Arg Phe Gln Pro Val Gln Lys Asp Asn Glu Gly Asn Pro Leu Asn Asn

610 615 620

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

625 630 635 640

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

645 650 655

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

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

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

865 870 875 880
Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys Val Glu Gly
885 890 895

Leu Asn Val Glu Glu Leu Lys Lys Leu Arg Leu Lys Asp Ile Asp Thr
900 905 910

Asp Thr Ala Lys Gln Glu Lys Asn Asn Ile Leu Asn Arg Val Met Pro
915 920 925

Met Gln Leu Pro Val Thr Val Tyr Glu Ile Asp Asp Ser His Asn Ile
930 935 940

Val Lys Asp Arg Pro Leu His Thr Val Tyr Ile Glu Glu Thr Lys Thr
945 950 955 960

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

Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Ser Ser Glu Thr Glu Leu
980 985 990

Lys Ser Asn Pro Ile Ser Lys Ser Leu Val Glu Tyr Glu Leu Gly Glu
995 1000 1005

Tyr Gln Asn Ala Arg Ile Glu Thr Ile Lys Asp Met Leu Leu Leu Glu
1010 1015 1020

Glu Thr Leu Ile Glu Lys Tyr Lys Thr Leu Pro Thr Asp Asn Phe Ser
1025 1030 1035 1040

Asp Met Leu Asn Gly Trp Leu Glu Gly Lys Asp Glu Ala Asp Lys Ala
1045 1050 1055

Arg Phe Gln Asn Asp Val Lys Leu Leu Val Ala Val Arg Asn Ala Phe
1060 1065 1070

Ser His Asn Gln Tyr Pro Met Arg Asn Arg Ile Ala Phe Ala Asn Ile
1075 1080 1085

Asn Pro Phe Ser Leu Ser Ser Ala Asp Thr Ser Glu Glu Lys Lys Leu
1090 1095 1100

Asp Ile Ala Asn Gln Leu Lys Asp Lys Thr His Lys Ile Ile Lys Arg
1105 1110 1115 1120

Ile Ile Glu Ile Glu Lys Pro Ile Glu Thr Lys Glu
1125 1130

<210> 59

<211> 1134

<212> PRT

<213> Prevotella intermedia

<400> 59

Met Lys Met Glu Asp Asp Lys Lys Thr Lys Glu Ser Thr Asn Met Leu
1 5 10 15

Asp Asn Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
20 25 30

Val Tyr Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn
35 40 45

Lys Lys Asp Gln Asp Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala
50 55 60

Ile Lys Thr His Trp Glu Lys Val Asn Gly Asp Leu Asn Lys Thr Glu
65 70 75 80

Arg Leu Arg Glu Leu Met Thr Lys His Phe Pro Phe Leu Glu Thr Ala

85 90 95

Ile Tyr Thr Lys Asn Lys Glu Asp Lys Glu Glu Val Lys Gln Glu Lys
100 105 110

Gln Ala Lys Ala Gln Ser Phe Asp Ser Leu Lys His Cys Leu Phe Leu
115 120 125

Phe Leu Glu Lys Leu Gln Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys
130 135 140

Tyr Ser Glu Ser Thr Lys Glu Pro Met Leu Glu Lys Glu Leu Leu Lys
145 150 155 160

Lys Met Tyr Asn Ile Phe Asp Asp Asn Ile Gln Leu Val Ile Lys Asp
165 170 175

Tyr Gln His Asn Lys Asp Ile Asn Pro Asp Glu Asp Phe Lys His Leu
180 185 190

Asp Arg Thr Glu Glu Glu Phe Asn Tyr Tyr Phe Thr Thr Asn Lys Lys
195 200 205

Gly Asn Ile Thr Ala Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu
210 215 220

Glu Lys Lys Asp Ala Ile Trp Met Gln Gln Lys Leu Arg Gly Phe Lys

225 230 235 240

Asp Asn Arg Glu Ser Lys Lys Lys Met Thr His Glu Val Phe Cys Arg

245 250 255

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

260 265 270

Asp Trp Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys

275 280 285

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

290 295 300

Pro Phe Asp Ser Ala Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe

305 310 315 320

Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu

325 330 335

Arg Tyr Phe Asp Tyr Asn Glu Ile Phe Thr Asn Leu Arg Phe Gln Ile

340 345 350

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

355 360 365

Gln Lys Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg

370 375 380

Ile Gln Glu Phe Ala Lys Gln Asn Arg Thr Asp Glu Trp Lys Ala Ile

385 390 395 400

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

405 410 415

Glu Thr Ala Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg

420 425 430

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

435 440 445

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

450 455 460

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

465 470 475 480

Leu Leu Leu Lys Lys Glu Glu Pro Asn Asn Asp Lys Lys Asn Ala Ser

485 490 495

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

500 505 510

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

515 520 525

Lys Tyr Cys Glu Asp Lys Gly Ile Pro Lys Arg His Leu Pro Lys Gln

530 535 540

Met Val Ala Ile Leu Tyr Asp Glu His Lys Asp Met Ala Glu Glu Ala

545 550 555 560

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

565 570 575

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

580 585 590

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

595 600 605

Met Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu Gly Asn Pro Leu

610 615 620

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

625 630 635 640

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

645 650 655

Val Asn Leu Ile Asn Ser Ser Asn Pro His Pro Phe Leu Lys Trp Thr

660 665 670

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

675 680 685

Thr Lys Lys Ile Glu Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Glu

690 695 700

Lys Asn Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn Arg Glu

705 710 715 720

Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile

725 730 735

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

805 810 815
Glu Lys Asp Phe Leu Pro Ser Glu Glu Arg Lys Lys Leu Trp Gly Asp
820 825 830
Lys Lys Tyr Lys Phe Lys Gly Tyr Lys Ala Lys Val Lys Ser Lys Lys
835 840 845
Leu Thr Asp Lys Glu Lys Glu Glu Tyr Arg Ser Tyr Leu Glu Phe Gln
850 855 860
Ser Trp Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln Asp
865 870 875 880

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

945 950 955 960
Lys Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp
965 970 975

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

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

1090 1095 1100
Lys Leu Asp Ile Ala Asn Gln Leu Lys Asp Lys Thr His Lys Ile Ile
1105 1110 1115 1120
Lys Arg Ile Ile Glu Ile Glu Lys Pro Ile Glu Thr Lys Glu
1125 1130

<210> 60
<211> 1133
<212> PRT
<213> Prevotella intermedia
<400> 60
Met Glu Asp Asp Lys Lys Thr Thr Asp Ser Ile Ser Tyr Glu Leu Lys
1 5 10 15

Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
20 25 30
Tyr Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn Lys
35 40 45
Lys Asp Gln Asp Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala Ile

50 55 60

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

65 70 75 80

Leu Arg Glu Leu Met Thr Lys His Phe Pro Phe Leu Glu Thr Ala Ile

85 90 95

Tyr Ser Lys Asn Lys Glu Asp Lys Glu Glu Val Lys Gln Glu Lys Gln

100 105 110

Ala Lys Ala Gln Ser Phe Asp Ser Leu Lys His Cys Leu Phe Leu Phe

115 120 125

Leu Glu Lys Leu Gln Glu Thr Arg Asn Tyr Tyr Ser His Tyr Lys Tyr

130 135 140

Ser Glu Ser Thr Lys Glu Pro Met Leu Glu Lys Glu Leu Leu Lys Lys

145 150 155 160

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

165 170 175

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

180 185 190

Arg Thr Glu Glu Asp Phe Asn Tyr Tyr Phe Thr Arg Asn Lys Lys Gly

195 200 205

Asn Ile Thr Glu Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu

210 215 220

Lys Lys Asp Ala Ile Trp Met Gln Gln Lys Leu Arg Gly Phe Lys Asp

225 230 235 240

Asn Arg Glu Ser Lys Lys Lys Met Thr His Glu Val Phe Cys Arg Ser

245 250 255

Arg Met Leu Leu Pro Lys Leu Arg Leu Glu Ser Thr Gln Thr Gln Asp

260 265 270

Trp Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser

275 280 285

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

290 295 300

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

370 375 380
Gln Glu Phe Ala Lys Gln Asn Arg Pro Asp Glu Trp Lys Ala Ile Val
385 390 395 400
Lys Asp Leu Asp Thr Tyr Glu Thr Ser Asn Glu Arg Tyr Ile Ser Glu
405 410 415
Thr Thr Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg Phe
420 425 430
Arg Asn Asp Asn Asp Glu Ile Trp Pro Ser Leu Lys Thr Asn Gly Glu
435 440 445

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

515 520 525
Tyr Cys Glu Asp Lys Gly Ile Pro Lys Arg His Leu Pro Lys Gln Met
530 535 540
Val Ala Ile Leu Tyr Asp Glu His Lys Asp Met Val Lys Glu Ala Lys
545 550 555 560

Arg Lys Gln Arg Lys Met Val Lys Asp Thr Glu Lys Leu Leu Ala Ala
565 570 575
Leu Glu Lys Gln Thr Gln Glu Lys Thr Glu Asp Gly Gly Arg Asn Ile
580 585 590

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

660 665 670
Trp Glu Glu Cys Asn Asn Ile Leu Ser Phe Tyr Arg Ser Tyr Leu Thr
675 680 685
Lys Lys Ile Glu Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Glu Lys
690 695 700
Asn Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn Arg Glu Thr
705 710 715 720
Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile Phe
725 730 735

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

805 810 815

Lys Asp Phe Leu His Arg Glu Glu Arg Ile Glu Leu Trp Asp Lys Lys

820 825 830

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

835 840 845

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

850 855 860

Trp Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln Asp Ile

865 870 875 880

Val Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys Val Glu

885 890 895

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

900 905 910

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

915 920 925

Pro Met Gln Leu Pro Val Thr Val Tyr Glu Ile Asp Asp Ser His Asn

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Glu Glu Thr Leu Ile Glu Lys Tyr Lys Asn Leu Pro Thr Asp Asn Phe

1025 1030 1035 1040

Ser Asp Met Leu Asn Gly Trp Leu Glu Gly Lys Asp Glu Ala Asp Lys

1045 1050 1055

Ala Arg Phe Gln Asn Asp Val Lys Leu Leu Val Ala Val Arg Asn Ala

1060 1065 1070

Phe Ser His Asn Gln Tyr Pro Met Arg Asn Arg Ile Ala Phe Ala Asn

1075 1080 1085

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

1090 1095 1100

Leu Asp Ile Ala Asn Gln Leu Lys Asp Lys Thr His Lys Ile Ile Lys

1105 1110 1115 1120

Arg Ile Ile Glu Ile Glu Lys Pro Ile Glu Thr Lys Glu

1125 1130

<210> 61

<211> 1127

<212> PRT

<213> Prevotella buccae

<400> 61

Met Gln Lys Gln Asp Lys Leu Phe Val Asp Arg Lys Lys Asn Ala Ile

1 5 10 15

Phe Ala Phe Pro Lys Tyr Ile Thr Ile Met Glu Asn Lys Glu Lys Pro

20 25 30

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

35 40 45

Leu Asn Leu Ala Arg His Asn Val Tyr Thr Thr Ile Asn His Ile Asn

50 55 60

Arg Arg Leu Glu Ile Ala Glu Leu Lys Asp Asp Gly Tyr Met Met Gly

65 70 75 80

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

85 90 95

Arg Leu Arg Asp Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala

100 105 110

Ala Tyr Glu Met Thr Asn Ser Lys Ser Pro Asn Asn Lys Glu Gln Arg

115 120 125

Glu Lys Glu Gln Ser Glu Ala Leu Ser Leu Asn Asn Leu Lys Asn Val

130 135 140

Leu Phe Ile Phe Leu Glu Lys Leu Gln Val Leu Arg Asn Tyr Tyr Ser
145 150 155 160
His Tyr Lys Tyr Ser Glu Glu Ser Pro Lys Pro Ile Phe Glu Thr Ser

165 170 175
Leu Leu Lys Asn Met Tyr Lys Val Phe Asp Ala Asn Val Arg Leu Val
180 185 190
Lys Arg Asp Tyr Met His His Glu Asn Ile Asp Met Gln Arg Asp Phe
195 200 205
Thr His Leu Asn Arg Lys Lys Gln Val Gly Arg Thr Lys Asn Ile Ile
210 215 220
Asp Ser Pro Asn Phe His Tyr His Phe Ala Asp Lys Glu Gly Asn Met
225 230 235 240

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

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

Thr Tyr His Phe Ser Ile Tyr Asn Lys Arg Ile Gly Asp Glu Asp Glu

385 390 395 400

Val Arg His Leu Thr His His Leu Tyr Gly Phe Ala Arg Ile Gln Asp

405 410 415

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

420 425 430

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

435 440 445

Pro His Tyr His Leu Glu Asn Glu Lys Ile Gly Ile Lys Phe Cys Ser

450 455 460

Ala His Asn Asn Leu Phe Pro Ser Leu Gln Thr Asp Lys Thr Cys Asn

465 470 475 480

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

485 490 495

Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu

500 505 510

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

515 520 525

Ile Arg Lys Glu Ile Ser Asn Ile Tyr Ala Ile Tyr Asp Ala Phe Ala

530 535 540

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

545 550 555 560

Thr Asn Ile Leu Gln Gly His Leu Pro Lys Gln Met Ile Ser Ile Leu

565 570 575

Lys Gly Arg Gln Lys Asp Met Gly Lys Glu Ala Glu Arg Lys Ile Gly

580 585 590

Glu Met Ile Asp Asp Thr Gln Arg Arg Leu Asp Leu Leu Cys Lys Gln

595 600 605

Thr Asn Gln Lys Ile Arg Ile Gly Lys Arg Asn Ala Gly Leu Leu Lys

610 615 620

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

625 630 635 640

Pro Val Gln Lys Asp Gln Asn Asn Ile Pro Ile Asn Asn Ser Lys Ala

645 650 655

Asn Ser Thr Glu Tyr Arg Met Leu Gln Arg Ala Leu Ala Leu Phe Gly

660 665 670

Ser Glu Asn Phe Arg Leu Lys Ala Tyr Phe Asn Gln Met Asn Leu Val

675 680 685

Gly Asn Asp Asn Pro His Pro Phe Leu Ala Glu Thr Gln Trp Glu His

690 695 700

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

705 710 715 720

Lys Tyr Leu Lys Gly Leu Lys Pro Gln Asn Trp Lys Gln Tyr Gln His

725 730 735

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

740 745 750

Gly Trp Lys Asn Ser Phe Asn Leu Pro Arg Gly Ile Phe Thr Gln Pro

755 760 765

Ile Arg Glu Trp Phe Glu Lys His Asn Asn Ser Lys Arg Ile Tyr Asp

770 775 780

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

785 790 795 800

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

805 810 815

Tyr Pro Phe Asn Ile Gly Asn Arg Leu Lys Pro Lys Lys Arg Gln Phe

820 825 830

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

835 840 845

Phe Lys Asn Tyr Pro Ser Glu Lys Lys Lys Thr Asp Leu Ala Tyr Leu

850 855 860

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

865 870 875 880

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

885 890 895

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

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

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

Ile Ala Pro Gln Leu Leu Glu Ile Val Gly Lys Ala Ile Lys Glu Ile
1105 1110 1115 1120
Glu Lys Ser Glu Asn Lys Asn
1125

<210> 62
<211> 1125

<212> PRT

<213> Prevotella pallens

<400> 62

Met Lys Glu Glu Glu Lys Gly Lys Thr Pro Val Val Ser Thr Tyr Asn
1 5 10 15

Lys Asp Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His
20 25 30

Asn Val Tyr Ile Thr Val Asn His Ile Asn Lys Ile Leu Gly Glu Gly
35 40 45

Glu Ile Asn Arg Asp Gly Tyr Glu Asn Thr Leu Glu Lys Ser Trp Asn
50 55 60

Glu Ile Lys Asp Ile Asn Lys Lys Asp Arg Leu Ser Lys Leu Ile Ile
65 70 75 80

Lys His Phe Pro Phe Leu Glu Val Thr Thr Tyr Gln Arg Asn Ser Ala
85 90 95

Asp Thr Thr Lys Gln Lys Glu Glu Lys Gln Ala Glu Ala Gln Ser Leu
100 105 110

Glu Ser Leu Lys Lys Ser Phe Phe Val Phe Ile Tyr Lys Leu Arg Asp
115 120 125

Leu Arg Asn His Tyr Ser His Tyr Lys His Ser Lys Ser Leu Glu Arg
130 135 140

Pro Lys Phe Glu Glu Asp Leu Gln Glu Lys Met Tyr Asn Ile Phe Asp
145 150 155 160

Ala Ser Ile Gln Leu Val Lys Glu Asp Tyr Lys His Asn Thr Asp Ile
165 170 175

Lys Thr Glu Glu Asp Phe Lys His Leu Asp Arg Lys Gly Gln Phe Lys
180 185 190

Tyr Ser Phe Ala Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu
195 200 205

Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Val Gln
210 215 220

Lys Lys Leu Glu Gly Phe Lys Cys Ser Asn Glu Ser Tyr Gln Lys Met

225 230 235 240

Thr Asn Glu Val Phe Cys Arg Ser Arg Met Leu Leu Pro Lys Leu Arg

245 250 255

Leu Gln Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn

260 265 270

Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Arg Glu Glu

275 280 285

Asp Arg Lys Lys Phe Arg Val Pro Ile Glu Ile Ala Asp Glu Asp Tyr

290 295 300

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

305 310 315 320

Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Ile Phe

325 330 335

Thr Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile

340 345 350

Tyr Lys Lys Gln Ile Gly Asp Tyr Lys Glu Ser His His Leu Thr His

355 360 365

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

370 375 380

Pro Asp Glu Trp Arg Lys Phe Val Lys Thr Phe Asn Ser Phe Glu Thr

385 390 395 400

Ser Lys Glu Pro Tyr Ile Pro Glu Thr Thr Pro His Tyr His Leu Glu

405 410 415

Asn Gln Lys Ile Gly Ile Arg Phe Arg Asn Asp Asn Asp Lys Ile Trp

420 425 430

Pro Ser Leu Lys Thr Asn Ser Glu Lys Asn Glu Lys Ser Lys Tyr Lys

435 440 445

Leu Asp Lys Ser Phe Gln Ala Glu Ala Phe Leu Ser Val His Glu Leu

450 455 460

Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu Lys Thr Glu Asn Thr Asp

465 470 475 480

Asn Asp Asn Glu Ile Glu Thr Lys Lys Lys Glu Asn Lys Asn Asp Lys
485 490 495
Gln Glu Lys His Lys Ile Glu Glu Ile Ile Glu Asn Lys Ile Thr Glu
500 505 510
Ile Tyr Ala Leu Tyr Asp Ala Phe Ala Asn Gly Lys Ile Asn Ser Ile
515 520 525
Asp Lys Leu Glu Glu Tyr Cys Lys Gly Lys Asp Ile Glu Ile Gly His

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

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

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

```
Pro Arg Gly Ile Phe Thr Glu Pro Ile Arg Lys Trp Phe Met Glu His
            740                 745                 750

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

            820                 825                 830
Pro Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu
        835                 840                 845
Arg Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Met Glu
    850                 855                 860
Leu Phe Asn Lys Lys Lys Ile Lys Glu Leu Asn Val Glu Lys Ile Tyr
865                 870                 875                 880
Leu Lys Asn Ile Asn Thr Asn Thr Thr Lys Lys Glu Lys Asn Thr Glu
                885                 890                 895

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

                965                 970                 975
Val Lys Thr His Ser Lys Ala Glu Ser Lys Ser Asn Thr Ile Ser Lys
```

980 985 990

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

995 1000 1005

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

1010 1015 1020

Asn Ser Leu Asp Thr Asp Asn Phe His Asn Met Leu Thr Gly Trp Leu

1025 1030 1035 1040

Lys Leu Lys Asp Glu Pro Asp Lys Ala Ser Phe Gln Asn Asp Val Asp

1045 1050 1055

Leu Leu Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met

1060 1065 1070

Arg Asn Arg Ile Ala Phe Ala Asn Ile Asn Pro Phe Ser Leu Ser Ser

1075 1080 1085

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

1090 1095 1100

Asp Lys Thr His Lys Thr Ile Glu Lys Ile Ile Glu Ile Glu Lys Pro

1105 1110 1115 1120

Ile Glu Thr Lys Glu

1125

<210> 63

<211> 1157

<212> PRT

<213> Myroides odoratimimus

<400> 63

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

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

180 185 190
Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys
195 200 205
Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe
210 215 220
Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Glu Thr Val Val Ala
225 230 235 240
Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser Asn Asp
245 250 255

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

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

325 330 335

Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val Tyr Gln

340 345 350

His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp Asn Glu

355 360 365

Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu Ser Arg

370 375 380

Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe Asn Tyr

385 390 395 400

Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr Leu Arg

405 410 415

Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr Lys Gln

420 425 430

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

435 440 445

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

450 455 460

Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp Thr Leu

465 470 475 480

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

485 490 495

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

500 505 510

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

515 520 525

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

530 535 540

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

545 550 555 560

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

565 570 575

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

580 585 590

Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile Gly Lys

595 600 605

Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile Leu Lys

610 615 620

Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys Ile Thr

625 630 635 640

Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile Leu Glu

645 650 655

Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr Lys Phe

660 665 670

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

675 680 685

Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met

690 695 700

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

705 710 715 720

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

725 730 735

Ile Leu Ser Asp Met Val Met Val Lys Asp Tyr Pro Ile Glu Leu Ile

740 745 750

Ala Leu Val Lys Lys Ser Arg Thr Leu Val Asp Phe Leu Asn Lys Tyr

755 760 765

Leu Glu Ala Arg Leu Gly Tyr Met Glu Asn Val Ile Thr Arg Val Lys

770 775 780

Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu Cys Phe

785 790 795 800

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

805 810 815

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

820 825 830

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

900 905 910
Lys Leu Ser Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr Gln Thr
915 920 925
Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr Gln Glu
930 935 940
Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln Leu Cys
945 950 955 960
Glu Gly Leu Val Arg Ile Asp Lys Val Lys Leu Lys Asp Ile Gly Asn
965 970 975

Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu Thr Tyr
980 985 990
Gln Ser Asp Ile Val Trp Ser Ala Tyr Leu Ser Asn Glu Val Asp Ser
995 1000 1005
Asn Lys Leu Tyr Val Ile Glu Arg Gln Leu Asp Asn Tyr Glu Ser Ile
1010 1015 1020
Arg Ser Lys Glu Leu Leu Lys Glu Val Gln Glu Ile Glu Cys Ser Val
1025 1030 1035 1040
Tyr Asn Gln Val Ala Asn Lys Glu Ser Leu Lys Gln Ser Gly Asn Glu

1045 1050 1055
Asn Phe Lys Gln Tyr Val Leu Gln Gly Leu Val Pro Ile Gly Met Asp
1060 1065 1070

Val Arg Glu Met Leu Ile Leu Ser Thr Asp Val Lys Phe Ile Lys Glu
1075 1080 1085

Glu Ile Ile Gln Leu Gly Gln Ala Gly Glu Val Glu Gln Asp Leu Tyr
1090 1095 1100

Ser Leu Ile Tyr Ile Arg Asn Lys Phe Ala His Asn Gln Leu Pro Ile
1105 1110 1115 1120

Lys Glu Phe Phe Asp Phe Cys Glu Asn Asn Tyr Arg Ser Ile Ser Asp
1125 1130 1135

Asn Glu Tyr Tyr Ala Glu Tyr Tyr Met Glu Ile Phe Arg Ser Ile Lys
1140 1145 1150

Glu Lys Tyr Thr Ser
1155

<210> 64

<211> 1157

<212> PRT

<213> Myroides odoratimimus

<400> 64

Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1 5 10 15

Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
20 25 30

Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg
35 40 45

Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn
50 55 60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu
65 70 75 80

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

85 90 95

Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu
100 105 110

Ser Gln Arg Ile Arg Gln Phe Arg Glu Met Leu Ile Ser Leu Val Thr

115 120 125

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Glu

130 135 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Leu Val

145 150 155 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr

165 170 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile

180 185 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys

195 200 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe

210 215 220

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

225 230 235 240

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

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg Gly Leu

305 310 315 320

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

325 330 335

Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val Tyr Gln

340 345 350

His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp Asn Glu

355 360 365

Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu Ser Arg

370 375 380

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

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

515 520 525

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

Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile Gly Lys
595 600 605
Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile Leu Lys

610 615 620

Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys Ile Thr

625 630 635 640

Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile Leu Glu

645 650 655

Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr Lys Phe

660 665 670

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

675 680 685

Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met

690 695 700

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

705 710 715 720

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

725 730 735

Ile Leu Ser Asp Met Val Met Val Lys Asp Tyr Pro Ile Glu Leu Ile

740 745 750

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

755 760 765

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

770 775 780

Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu Cys Phe

785 790 795 800

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

805 810 815

Ile Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly Phe Met

820 825 830

Asp Ser Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Gln Gln His Lys

835 840 845

Glu Asp Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser Asn Tyr

850 855 860

Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Ile Thr Glu Asp Lys

865 870 875 880

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

945 950 955 960
Glu Gly Leu Val Arg Ile Asp Lys Val Lys Leu Lys Asp Ile Gly Asn
965 970 975
Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu Thr Tyr
980 985 990
Gln Ser Asp Ile Val Trp Ser Gly Tyr Leu Ser Asn Glu Val Asp Ser
995 1000 1005
Asn Lys Leu Tyr Val Ile Glu Arg Gln Leu Asp Asn Tyr Glu Ser Ile
1010 1015 1020

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

1090 1095 1100
Ser Leu Ile Tyr Ile Arg Asn Lys Phe Ala His Asn Gln Leu Pro Ile
1105 1110 1115 1120

Lys Glu Phe Phe Asp Phe Cys Glu Asn Asn Tyr Arg Pro Ile Ser Asp
1125 1130 1135

Asn Glu Tyr Tyr Ala Glu Tyr Tyr Met Glu Ile Phe Arg Ser Ile Lys
1140 1145 1150

Glu Lys Tyr Ala Ser
1155

<210> 65

<211> 1224

<212>
PRT

<213> Bergeyella zoohelcum

<400> 65

Met Glu Asn Lys Thr Ser Leu Gly Asn Asn Ile Tyr Tyr Asn Pro Phe
1 5 10 15

Lys Pro Gln Asp Lys Ser Tyr Phe Ala Gly Tyr Phe Asn Ala Ala Met
20 25 30

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

Lys Glu Tyr Thr Ser Glu Asn Phe Phe Asp Ala Ile Phe Lys Glu Asn
50 55 60

Ile Ser Leu Val Glu Tyr Glu Arg Tyr Val Lys Leu Leu Ser Asp Tyr
65 70 75 80

Phe Pro Met Ala Arg Leu Leu Asp Lys Lys Glu Val Pro Ile Lys Glu
85 90 95

Arg Lys Glu Asn Phe Lys Lys Asn Phe Lys Gly Ile Ile Lys Ala Val
100 105 110

Arg Asp Leu Arg Asn Phe Tyr Thr His Lys Glu His Gly Glu Val Glu
115 120 125

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

130 135 140

Val Leu Thr Val Lys Lys Lys Lys Val Lys Thr Asp Lys Thr Lys Glu
145 150 155 160

Ile Leu Lys Lys Ser Ile Glu Lys Gln Leu Asp Ile Leu Cys Gln Lys

165 170 175

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

180 185 190

Arg Asn Gln Arg Glu Arg Gly Glu Lys Glu Leu Val Ala Pro Phe Lys

195 200 205

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

210 215 220

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

225 230 235 240

Lys Ala Lys Tyr Asn Thr Lys Ser Asp Pro Gln Gln Glu Glu Gly Asp

245 250 255

Leu Lys Ile Pro Ile Ser Lys Asn Gly Val Val Phe Leu Leu Ser Leu

260 265 270

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

275 280 285

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

290 295 300

Ser His Gly Lys Asn Ser Ile Cys Phe Met Ala Thr His Glu Ile Phe

305 310 315 320

Ser His Leu Ala Tyr Lys Lys Leu Lys Arg Lys Val Arg Thr Ala Glu

325 330 335

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

340 345 350

Lys Glu Thr Leu Met Met Gln Met Leu Asp Glu Leu Ser Lys Val Pro

355 360 365

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

370 375 380

Glu Asp Trp Asn Glu Tyr Leu Lys Glu Asn Asn Gly Asp Val Gly Thr

385 390 395 400

Met Glu Glu Glu Gln Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu

405 410 415

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

420 425 430

Phe Pro Thr Leu Arg Phe Gln Val His Leu Gly Asn Tyr Leu His Asp

435 440 445

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

450 455 460

Ile Thr Val Phe Gly Arg Leu Ser Glu Leu Glu His Lys Lys Ala Leu

465 470 475 480

Phe Ile Lys Asn Thr Glu Thr Asn Glu Asp Arg Glu His Tyr Trp Glu

485 490 495

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

500 505 510

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

515 520 525

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

530 535 540

Gln Tyr Val Ser Glu Val Asp Lys Ala Val Lys Ala His Gln Leu Lys

545 550 555 560

Gln Arg Lys Ala Ser Lys Pro Ser Ile Gln Asn Ile Ile Glu Glu Ile

565 570 575

Val Pro Ile Asn Glu Ser Asn Pro Lys Glu Ala Ile Val Phe Gly Gly

580 585 590

Gln Pro Thr Ala Tyr Leu Ser Met Asn Asp Ile His Ser Ile Leu Tyr

595 600 605

Glu Phe Phe Asp Lys Trp Glu Lys Lys Lys Glu Lys Leu Glu Lys Lys

610 615 620

Gly Glu Lys Glu Leu Arg Lys Glu Ile Gly Lys Glu Leu Glu Lys Lys

625 630 635 640

Ile Val Gly Lys Ile Gln Ala Gln Ile Gln Gln Ile Ile Asp Lys Asp

645 650 655

Thr Asn Ala Lys Ile Leu Lys Pro Tyr Gln Asp Gly Asn Ser Thr Ala

660 665 670

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

675 680 685

Leu Gln Lys Leu Lys Asp Glu Gln Thr Val Arg Glu Lys Glu Tyr Asn

690 695 700

Asp Phe Ile Ala Tyr Gln Asp Lys Asn Arg Glu Ile Asn Lys Val Arg

705 710 715 720

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

725 730 735

Pro Glu Ala Pro Ala Arg Lys Glu Val Leu Tyr Tyr Arg Glu Lys Gly

740 745 750

Lys Val Ala Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met Pro Thr

755 760 765

Asp Phe Lys Asn Glu Trp Lys Gly Glu Gln His Ser Leu Leu Gln Lys

770 775 780

Ser Leu Ala Tyr Tyr Glu Gln Cys Lys Glu Glu Leu Lys Asn Leu Leu

785 790 795 800

Pro Glu Lys Val Phe Gln His Leu Pro Phe Lys Leu Gly Gly Tyr Phe

805 810 815

Gln Gln Lys Tyr Leu Tyr Gln Phe Tyr Thr Cys Tyr Leu Asp Lys Arg

820 825 830

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

835 840 845

Glu Asn Lys Val Phe Lys Lys Val Glu Asn Glu Cys Phe Lys Phe Leu

850 855 860

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

865 870 875 880

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

885 890 895

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

900 905 910

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

915 920 925

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

995 1000 1005
Thr Asn Tyr Ile Trp Asn Lys Thr Val Asp Leu Lys Leu Cys Asp Gly
1010 1015 1020
Lys Ile Thr Val Glu Asn Val Lys Leu Lys Asn Val Gly Asp Phe Ile
1025 1030 1035 1040
Lys Tyr Glu Tyr Asp Gln Arg Val Gln Ala Phe Leu Lys Tyr Glu Glu
1045 1050 1055
Asn Ile Glu Trp Gln Ala Phe Leu Ile Lys Glu Ser Lys Glu Glu Glu
1060 1065 1070

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

1140 1145 1150
Asp Val Asn Ile Asn Gln Leu Lys Gln Glu Ala Thr Asp Leu Glu Gln
1155 1160 1165

Lys Ala Phe Val Leu Thr Tyr Ile Arg Asn Lys Phe Ala His Asn Gln
1170 1175 1180
Leu Pro Lys Lys Glu Phe Trp Asp Tyr Cys Gln Glu Lys Tyr Gly Lys
1185 1190 1195 1200
Ile Glu Lys Glu Lys Thr Tyr Ala Glu Tyr Phe Ala Glu Val Phe Lys
1205 1210 1215

Lys Glu Lys Glu Ala Leu Ile Lys
1220

<210> 66
<211> 1150
<212> PRT
<213> Prevotella saccharolytica
<400> 66

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

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

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

145 150 155 160

Ala Lys Lys Met Glu Val Lys Phe Asp Asp Gly Phe Cys Asn Lys Leu

165 170 175

Tyr Tyr Val Phe Asp Ala Ala Leu Gln Met Val Lys Asp Asp Asn Arg

180 185 190

Met Asn Pro Glu Ile Asn Met Gln Thr Asp Phe Glu His Leu Val Arg

195 200 205

Leu Gly Arg Asn Arg Lys Ile Pro Asn Thr Phe Lys Tyr Asn Phe Thr

210 215 220

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

225 230 235 240

Leu Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Lys

245 250 255

Gly Phe Lys Gly Gly Thr Glu Asn Tyr Met Arg Met Thr Asn Glu Val

260 265 270

Phe Cys Arg Asn Arg Met Val Ile Pro Lys Leu Arg Leu Glu Thr Asp

275 280 285

Tyr Asp Asn His Gln Leu Met Phe Asp Met Leu Asn Glu Leu Val Arg

290 295 300

Cys Pro Leu Ser Leu Tyr Lys Arg Leu Lys Gln Glu Asp Gln Asp Lys

305 310 315 320

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

325 330 335

Pro Tyr Gln Glu Asn Ala Asn Ser Asp Glu Asn Pro Thr Glu Glu Thr

340 345 350

Asp Pro Leu Lys Asn Thr Leu Val Arg His Gln His Arg Phe Pro Tyr

355 360 365

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

370 375 380

Phe Gln Ile Asn Leu Gly Cys Tyr His Phe Ser Ile Tyr Asp Lys Thr

385 390 395 400

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

405 410 415

Phe Asp Arg Leu Gln Asn Phe Ser Val Lys Leu Gln Pro Glu His Trp

420 425 430

Lys Asn Met Val Lys His Leu Asp Thr Glu Glu Ser Ser Asp Lys Pro

435 440 445

Tyr Leu Ser Asp Ala Met Pro His Tyr Gln Ile Glu Asn Glu Lys Ile

450 455 460

Gly Ile His Phe Leu Lys Thr Asp Thr Glu Lys Lys Glu Thr Val Trp

465 470 475 480

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

485 490 495

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

500 505 510

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

515 520 525

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

530 535 540

Lys Lys Ile Phe Asp Ile Tyr Asp Asp Phe Ala Asn Gly Thr Ile Asn

545 550 555 560

Ser Met Gln Lys Leu Asp Glu Arg Leu Ala Lys Asp Asn Leu Leu Arg

565 570 575

Gly Asn Met Pro Gln Gln Met Leu Ala Ile Leu Glu His Gln Glu Pro

580 585 590

Asp Met Glu Gln Lys Ala Lys Glu Lys Leu Asp Arg Leu Ile Thr Glu

595 600 605

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

610 615 620

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

625 630 635 640

Asp Trp Leu Val Asn Asp Met Met Arg Phe Gln Pro Ala Lys Arg Asn

645 650 655

Ala Asp Asn Thr Gly Val Pro Asp Ser Lys Ala Asn Ser Thr Glu Tyr

660 665 670

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

675 680 685

Leu Glu Pro Tyr Phe Arg Gln Val Asn Leu Ile Gly Gly Thr Asn Pro

690 695 700

His Pro Phe Leu His Arg Val Asp Trp Lys Lys Cys Asn His Leu Leu

705 710 715 720

Ser Phe Tyr His Asp Tyr Leu Glu Ala Lys Glu Gln Tyr Leu Ser His

725 730 735

Leu Ser Pro Ala Asp Trp Gln Lys His Gln His Phe Leu Leu Leu Lys

740 745 750

Val Arg Lys Asp Ile Gln Asn Glu Lys Lys Asp Trp Lys Lys Ser Leu

755 760 765

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

770 775 780

Glu Ser Ile Lys Thr Trp Phe Ser Thr Asp Ala Asp Lys Val Gln Ile

785 790 795 800

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

805 810 815

Ile Pro Leu Tyr Tyr Asp Lys Val Tyr Asn Asp Lys Pro Gln Pro Phe

820 825 830

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

835 840 845

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

850 855 860

Met Leu Tyr Lys Asn Ala Gln Pro Asp Ser Ser Asp Lys Ile Glu Tyr

865 870 875 880

Pro Gln Tyr Leu Asp Phe Leu Ser Trp Lys Lys Leu Glu Arg Glu Leu

885 890 895

Arg Met Leu Arg Asn Gln Asp Met Met Val Trp Leu Met Cys Lys Asp

900 905 910

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

980 985 990
Ser Leu Leu Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Ile Ala
995 1000 1005
Ala Glu Gly Glu Asp Leu Gln Gln His Pro Leu Thr Lys Asn Arg Leu
1010 1015 1020
Glu Tyr Glu Leu Ser Ile Tyr Gln Thr Met Arg Ile Ser Val Phe Glu
1025 1030 1035 1040
Gln Thr Leu Gln Leu Glu Lys Ala Ile Leu Thr Arg Asn Lys Thr Leu
1045 1050 1055

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

1125 1130 1135
Ala Ala Ser Arg Leu Gln Asn Ile Ile Asn Gly Gly Thr Asn
1140 1145 1150

<210> 67

<211> 1175

<212> PRT

<213> Porphyromonas gingivalis

<400> 67

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
50 55 60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65 70 75 80

Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95

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

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205

Arg Cys Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp

210 215 220

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

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

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

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

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

385 390 395 400

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

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445

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

450 455 460

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

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

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

675 680 685
Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
690 695 700
Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720

Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
740 745 750

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

820 825 830
Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
835 840 845
Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Asn Ala Ser Arg Glu Asn
850 855 860
Lys Lys Lys Ile Glu Gln Leu Leu Gln Asp Leu Ser Leu Trp Glu Thr
865 870 875 880
Phe Glu Ser Lys Leu Lys Val Lys Ala Asp Lys Ile Asn Ile Ala Lys
885 890 895

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

965 970 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val

980 985 990

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

995 1000 1005

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

1010 1015 1020

Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Ala Leu Ala Met

1025 1030 1035 1040

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

1045 1050 1055

Tyr Gln Thr Ala Arg Val Cys Ala Phe Glu Gln Thr Leu Glu Leu Glu

1060 1065 1070

Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro Asp Lys Asn Phe Arg

1075 1080 1085

Lys Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Asp Lys Trp Pro Asp

1090 1095 1100

Leu His Gly Asn Val Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser

1105 1110 1115 1120

His Asn Gln Tyr Pro Met Tyr Asp Glu Thr Leu Phe Ser Ser Ile Arg

1125 1130 1135

Lys Tyr Asp Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu

1140 1145 1150

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

1155 1160 1165

Val Glu Arg Ile Ile Gln Ala

1170 1175

<210> 68

<

211> 1115

<212> PRT

<213> Bacteroides pyogenes

<400> 68

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

50 55 60
Leu Ser Ser Ala Lys Trp Thr Arg Phe Tyr Tyr Lys Ser Arg Lys Phe
65 70 75 80
Leu Pro Phe Leu Glu Met Phe Asp Ser Asp Lys Lys Ser Tyr Glu Asn
85 90 95
Arg Arg Glu Thr Thr Glu Cys Leu Asp Thr Ile Asp Arg Gln Lys Ile
100 105 110
Ser Ser Leu Leu Lys Glu Val Tyr Gly Lys Leu Gln Asp Ile Arg Asn
115 120 125

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

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

Leu Leu Ser Val Asn Pro Arg Glu Ala Leu Leu Met Asp Met Leu Asn
260 265 270

Glu Leu Asn Arg Cys Pro Asp Ile Leu Phe Glu Met Leu Asp Glu Lys
275 280 285

Asp Gln Lys Ser Phe Leu Pro Leu Leu Gly Glu Glu Glu Gln Ala His
290 295 300

Ile Leu Glu Asn Ser Leu Asn Asp Glu Leu Cys Glu Ala Ile Asp Asp
305 310 315 320

Pro Phe Glu Met Ile Ala Ser Leu Ser Lys Arg Val Arg Tyr Lys Asn
325 330 335

Arg Phe Pro Tyr Leu Met Leu Arg Tyr Ile Glu Glu Lys Asn Leu Leu
340 345 350

Pro Phe Ile Arg Phe Arg Ile Asp Leu Gly Cys Leu Glu Leu Ala Ser
355 360 365

Tyr Pro Lys Lys Met Gly Glu Glu Asn Asn Tyr Glu Arg Ser Val Thr
370 375 380

Asp His Ala Met Ala Phe Gly Arg Leu Thr Asp Phe His Asn Glu Asp
385 390 395 400

Ala Val Leu Gln Gln Ile Thr Lys Gly Ile Thr Asp Glu Val Arg Phe
405 410 415

Ser Leu Tyr Ala Pro Arg Tyr Ala Ile Tyr Asn Asn Lys Ile Gly Phe
420 425 430

Val Arg Thr Gly Gly Ser Asp Lys Ile Ser Phe Pro Thr Leu Lys Lys
435 440 445

Lys Gly Gly Glu Gly His Cys Val Ala Tyr Thr Leu Gln Asn Thr Lys
450 455 460

Ser Phe Gly Phe Ile Ser Ile Tyr Asp Leu Arg Lys Ile Leu Leu Leu
465 470 475 480

Ser Phe Leu Asp Lys Asp Lys Ala Lys Asn Ile Val Ser Gly Leu Leu
485 490 495

Glu Gln Cys Glu Lys His Trp Lys Asp Leu Ser Glu Asn Leu Phe Asp

500 505 510

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

515 520 525

Tyr Thr Leu Pro Arg Ser Lys Gly Gly Lys Leu Val Ser Ser Lys Leu

530 535 540

Ala Asp Lys Gln Glu Lys Tyr Glu Ser Glu Phe Glu Arg Arg Lys Glu

545 550 555 560

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

565 570 575

Pro Arg Arg Met Ile Asp Glu Trp Leu Asn Val Leu Pro Thr Ser Arg

580 585 590

Glu Lys Lys Leu Lys Gly Tyr Val Glu Thr Leu Lys Leu Asp Cys Arg

595 600 605

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

610 615 620

Pro Pro Arg Ile Gly Glu Met Ala Thr Asp Leu Ala Lys Asp Ile Ile

625 630 635 640

Arg Met Val Ile Asp Gln Gly Val Lys Gln Arg Ile Thr Ser Ala Tyr

645 650 655

Tyr Ser Glu Ile Gln Arg Cys Leu Ala Gln Tyr Ala Gly Asp Asp Asn

660 665 670

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

675 680 685

Lys Asn Gly His Pro Phe Leu Gly Lys Val Leu Arg Pro Gly Leu Gly

690 695 700

His Thr Glu Lys Leu Tyr Gln Arg Tyr Phe Glu Glu Lys Lys Glu Trp

705 710 715 720

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

725 730 735

Phe Val Asn Pro Pro Thr Gly Lys Gln Lys Glu Leu Pro Leu Ile Ile

740 745 750

Arg Asn Leu Met Lys Glu Arg Pro Glu Trp Arg Asp Trp Lys Gln Arg

```
        755                 760                 765
Lys Asn Ser His Pro Ile Asp Leu Pro Ser Gln Leu Phe Glu Asn Glu

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

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

        915                 920                 925
Met Leu Gly Glu Pro Val Ser Val Ser Gln Val Ile Gln Leu Glu Gly
    930                 935                 940
Gly Gln Pro Asp Ala Val Ile Lys Ala Glu Cys Lys Leu Lys Asp Val
945                 950                 955                 960
Ser Lys Leu Met Arg Tyr Cys Tyr Asp Gly Arg Val Lys Gly Leu Met
                965                 970                 975
Pro Tyr Phe Ala Asn His Glu Ala Thr Gln Glu Gln Val Glu Met Glu
            980                 985                 990

Leu Arg His Tyr Glu Asp His Arg Arg Arg Val Phe Asn Trp Val Phe
        995                1000                1005
```

Ala Leu Glu Lys Ser Val Leu Lys Asn Glu Lys Leu Arg Arg Phe Tyr
1010 1015 1020
Glu Glu Ser Gln Gly Gly Cys Glu His Arg Arg Cys Ile Asp Ala Leu
1025 1030 1035 1040
Arg Lys Ala Ser Leu Val Ser Glu Glu Glu Tyr Glu Phe Leu Val His
1045 1050 1055
Ile Arg Asn Lys Ser Ala His Asn Gln Phe Pro Asp Leu Glu Ile Gly
1060 1065 1070
Lys Leu Pro Pro Asn Val Thr Ser Gly Phe Cys Glu Cys Ile Trp Ser
1075 1080 1085
Lys Tyr Lys Ala Ile Ile Cys Arg Ile Ile Pro Phe Ile Asp Pro Glu
1090 1095 1100
Arg Arg Phe Phe Gly Lys Leu Leu Glu Gln Lys
1105 1110 1115

<210> 69
<211> 1135
<212> PRT
<213> Porphyromonas gingivalis
<400> 69

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

100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His

115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp

130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr

210 215 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys

225 230 235 240

Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala

245 250 255

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu

260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu

275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu

290 295 300

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

305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe

325 330 335

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

340 345 350

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

355 360 365

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

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

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

580 585 590
Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
595 600 605

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

Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe
660 665 670
Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
675 680 685
Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
690 695 700
Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705 710 715 720
Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
725 730 735
Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln
740 745 750
Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
755 760 765
Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
770 775 780
Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785 790 795 800

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

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg

865 870 875 880

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

885 890 895

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

900 905 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925

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

930 935 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg

945 950 955 960

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

965 970 975

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

980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg

995 1000 1005

Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser

1010 1015 1020

Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys

1025 1030 1035 1040

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

1045 1050 1055

Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg

1060 1065 1070

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

1075 1080 1085

Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu

1090 1095 1100

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

1105 1110 1115 1120

Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu

1125 1130 1135

<210> 70

<211> 1135

<212> PRT

<213> Porphyromonas gingivalis

<400> 70

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val

1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu

20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys

35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu

50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr

65 70 75 80

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

85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile

100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His

115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp

130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala

145 150 155 160

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

165 170 175

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

180 185 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
195 200 205
Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
210 215 220

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

290 295 300
Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305 310 315 320
Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
325 330 335
Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
340 345 350
Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
355 360 365

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

435 440 445

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

450 455 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Gly Phe Leu Arg Lys

465 470 475 480

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

485 490 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro

500 505 510

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

515 520 525

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

530 535 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp

545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr

565 570 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe

580 585 590

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

595 600 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu

610 615 620

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

625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala

645 650 655

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

660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr

675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr

690 695 700

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

705 710 715 720

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

725 730 735

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

740 745 750

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met

755 760 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr

785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile

805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu

820 825 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro

835 840 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala

850 855 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg

865 870 875 880

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

885 890 895

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

900 905 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925

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

930 935 940

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

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

Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu
1090 1095 1100
Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp His Glu Asn Arg Phe
1105 1110 1115 1120
Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu
1125 1130 1135

<210> 71
<211> 1135
<212> PRT
<213> Porphyromonas gingivalis
<400> 71
Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val

1 5 10 15
Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu

20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys

35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu

50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr

65 70 75 80

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

85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile

100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His

115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp

130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr

210 215 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys

225 230 235 240

Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala

245 250 255

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu

260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu

275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu

290 295 300

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

305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe

325 330 335

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

340 345 350

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

355 360 365

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

370 375 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu

385 390 395 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile

405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro

420 425 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met

435 440 445

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

450 455 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys

465 470 475 480

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

485 490 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro

500 505 510

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

515 520 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
530 535 540

Ser Ala Phe Asp Met Asp Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
565 570 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Gln Lys Phe

580 585 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
595 600 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
610 615 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg His Gln Phe Arg Ala
645 650 655

Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe
660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705 710 715 720

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

725 730 735

Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln
740 745 750

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Val Met
755 760 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
770 775 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
820 825 830

Lys Thr Val Arg Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
835 840 845

Val Pro Pro Asp Leu Ala Ala Tyr Ile Lys Arg Ser Phe His Arg Ala
850 855 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg

865 870 875 880
Leu Met Leu Met Ala Ile Asn Lys Ile Met Thr Asp Arg Glu Glu Asp
885 890 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Lys Glu Asn
900 905 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
915 920 925

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

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
965 970 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg
995 1000 1005

Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser

1010 1015 1020
Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys

1025 1030 1035 1040

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

1045 1050 1055

Ala His Asn Gln Phe Pro Cys Ala Ala Glu Ile Pro Leu Ile Tyr Arg

1060 1065 1070

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

1075 1080 1085

Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu

1090 1095 1100

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

1105 1110 1115 1120

Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu

1125 1130 1135

<210> 72

<211> 1125

<212> PRT

<213> Prevotella intermedia

<400> 72

Met Lys Met Glu Asp Asp Lys Lys Thr Thr Glu Ser Thr Asn Met Leu

1 5 10 15

Asp Asn Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn

20 25 30

Val Tyr Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn

35 40 45

Lys Lys Asp Gln Asp Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala

50 55 60

Ile Lys Thr His Trp Glu Lys Val Asn Gly Asp Leu Asn Lys Thr Glu

65 70 75 80

Arg Leu Arg Glu Leu Met Thr Lys His Phe Pro Phe Leu Glu Thr Ala

85 90 95

Ile Tyr Thr Lys Asn Lys Glu Asp Lys Glu Glu Val Lys Gln Glu Lys

100 105 110

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

145 150 155 160
Lys Met Tyr Asn Ile Phe Asp Asp Asn Ile Gln Leu Val Ile Lys Asp
165 170 175
Tyr Gln His Asn Lys Asp Ile Asn Pro Asp Glu Asp Phe Lys His Leu
180 185 190
Asp Arg Lys Gly Gln Phe Lys Tyr Ser Phe Ala Asp Asn Glu Gly Asn
195 200 205
Ile Thr Glu Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu Lys
210 215 220

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

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

355 360 365

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

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

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

580 585 590
Leu Leu Lys Ser Gly Glu Ile Ala Arg Trp Leu Val Asn Asp Met Met
595 600 605

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

Leu Ile Asn Ser Ser Asn Pro His Pro Phe Leu Lys Trp Thr Lys Trp
660 665 670
Glu Glu Cys Asn Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Thr Lys
675 680 685
Lys Ile Glu Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Glu Lys Asn
690 695 700
Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn Arg Glu Thr Leu
705 710 715 720
Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile Phe Thr
725 730 735
Glu Pro Ile Arg Glu Trp Phe Lys Arg His Gln Asn Asp Ser Lys Glu
740 745 750
Tyr Glu Lys Val Glu Ala Leu Lys Arg Val Gly Leu Val Thr Lys Val
755 760 765
Ile Pro Leu Phe Phe Lys Glu Glu Tyr Phe Lys Glu Asp Ala Gln Lys
770 775 780
Glu Ile Asn Asn Cys Val Gln Pro Phe Tyr Ser Phe Pro Tyr Asn Val
785 790 795 800

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

Arg Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Thr Glu

865 870 875 880

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

885 890 895

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

900 905 910

Asn Asn Ile Leu Asn Arg Ile Met Pro Met Gln Leu Pro Val Thr Val

915 920 925

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

930 935 940

Thr Val Tyr Ile Glu Glu Thr Lys Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ala Leu Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe

965 970 975

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

980 985 990

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

995 1000 1005

Thr Ile Lys Asp Met Leu Leu Leu Glu Lys Thr Leu Ile Lys Lys Tyr

1010 1015 1020

Glu Lys Leu Pro Thr Asp Asn Phe Ser Asp Met Leu Asn Gly Trp Leu

1025 1030 1035 1040

Glu Gly Lys Asp Glu Ser Asp Lys Ala Arg Phe Gln Asn Asp Val Lys

1045 1050 1055

Leu Leu Val Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met

1060 1065 1070

Arg Asn Arg Ile Ala Phe Ala Asn Ile Asn Pro Phe Ser Leu Ser Ser

1075 1080 1085

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

1090 1095 1100

Asp Lys Thr His Lys Ile Ile Lys Lys Ile Ile Glu Ile Glu Lys Pro

1105 1110 1115 1120

Ile Glu Thr Lys Glu

1125

<210> 73

<211> 1127

<212> PRT

<213> Prevotella buccae

<400> 73

Met Gln Lys Gln Asp Lys Leu Phe Val Asp Arg Lys Lys Asn Ala Ile

1 5 10 15

Phe Ala Phe Pro Lys Tyr Ile Thr Ile Met Glu Asn Gln Glu Lys Pro

20 25 30

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

35 40 45

Leu Asn Leu Ala Arg His Asn Val Tyr Thr Thr Ile Asn His Ile Asn

50 55 60

Arg Arg Leu Glu Ile Ala Glu Leu Lys Asp Asp Gly Tyr Met Met Asp

65 70 75 80

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

85 90 95

Arg Leu Arg Asp Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala

100 105 110

Ala Tyr Glu Ile Thr Asn Ser Lys Ser Pro Asn Asn Lys Glu Gln Arg

115 120 125

Glu Lys Glu Gln Ser Glu Ala Leu Ser Leu Asn Asn Leu Lys Asn Val

130 135 140

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

145 150 155 160

His Tyr Lys Tyr Ser Glu Glu Ser Pro Lys Pro Ile Phe Glu Thr Ser

165 170 175

Leu Leu Lys Asn Met Tyr Lys Val Phe Asp Ala Asn Val Arg Leu Val

180 185 190

Lys Arg Asp Tyr Met His His Glu Asn Ile Asp Met Gln Arg Asp Phe
195 200 205

Thr His Leu Asn Arg Lys Lys Gln Val Gly Arg Thr Lys Asn Ile Ile
210 215 220

Asp Ser Pro Asn Phe His Tyr His Phe Ala Asp Lys Glu Gly Asn Met
225 230 235 240

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

Asp Ala Ile Trp Met Gln Lys Lys Leu Lys Gly Phe Lys Asp Gly Arg
260 265 270

Asn Leu Arg Glu Gln Met Thr Asn Glu Val Phe Cys Arg Ser Arg Ile
275 280 285

Ser Leu Pro Lys Leu Lys Leu Glu Asn Val Gln Thr Lys Asp Trp Met
290 295 300

Gln Leu Asp Met Leu Asn Glu Leu Val Arg Cys Pro Lys Ser Leu Tyr

305 310 315 320

Glu Arg Leu Arg Glu Lys Asp Arg Glu Ser Phe Lys Val Pro Phe Asp
325 330 335

Ile Phe Ser Asp Asp Tyr Asp Ala Glu Glu Glu Pro Phe Lys Asn Thr
340 345 350

Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Val Leu Arg Tyr Phe
355 360 365

Asp Leu Asn Glu Ile Phe Glu Gln Leu Arg Phe Gln Ile Asp Leu Gly
370 375 380

Thr Tyr His Phe Ser Ile Tyr Asn Lys Arg Ile Gly Asp Glu Asp Glu
385 390 395 400

Val Arg His Leu Thr His His Leu Tyr Gly Phe Ala Arg Ile Gln Asp
405 410 415

Phe Ala Gln Gln Asn Gln Pro Glu Val Trp Arg Lys Leu Val Lys Asp
420 425 430

Leu Asp Tyr Phe Glu Ala Ser Gln Glu Pro Tyr Ile Pro Lys Thr Ala
435 440 445

Pro His Tyr His Leu Glu Asn Glu Lys Ile Gly Ile Lys Phe Cys Ser

450 455 460

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

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

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

Ser Glu Asn Phe Arg Leu Lys Ala Tyr Phe Asn Gln Met Asn Leu Val
675 680 685
Gly Asn Asp Asn Pro His Pro Phe Leu Ala Glu Thr Gln Trp Glu His

690 695 700

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

705 710 715 720

Lys Tyr Leu Lys Gly Leu Lys Pro Gln Asn Trp Lys Gln Tyr Gln His

725 730 735

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

740 745 750

Gly Trp Lys Asn Ser Phe Asn Leu Pro Arg Gly Ile Phe Thr Gln Pro

755 760 765

Ile Arg Glu Trp Phe Glu Lys His Asn Asn Ser Lys Arg Ile Tyr Asp

770 775 780

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

785 790 795 800

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

805 810 815

Tyr Pro Phe Asn Ile Gly Asn Lys Leu Lys Pro Gln Lys Gly Gln Phe

820 825 830

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

835 840 845

Phe Lys Asn Tyr Pro Ser Glu Lys Lys Lys Thr Asp Leu Ala Tyr Leu

850 855 860

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

865 870 875 880

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

885 890 895

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

900 905 910

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

915 920 925

Met Pro Met Lys Leu Pro Val Lys Thr Tyr Glu Thr Asp Asn Lys Gly

930 935 940

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

945 950 955 960

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

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

Ile Ala Pro Gln Leu Leu Glu Ile Val Gly Lys Ala Ile Lys Glu Ile
1105 1110 1115 1120
Glu Lys Ser Glu Asn Lys Asn
1125

<210> 74
<211> 1135
<212> PRT
<213> Porphyromonas gingivalis
<400> 74
Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1 5 10 15
Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu

20 25 30

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

100 105 110
Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
115 120 125
Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
130 135 140
Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145 150 155 160
Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
165 170 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
180 185 190
Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
195 200 205
Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
210 215 220
Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225 230 235 240
Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala

245 250 255
Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
290 295 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
325 330 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
340 345 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
355 360 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
370 375 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu

385 390 395 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
420 425 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
435 440 445

Gly Phe Ile Ser Val His Asn Leu Arg Lys Leu Leu Leu Met Glu Leu
450 455 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465 470 475 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
485 490 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
500 505 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
515 520 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu

530 535 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
565 570 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
580 585 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
595 600 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
610 615 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
645 650 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr

675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705 710 715 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
725 730 735

Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln
740 745 750

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met

755 760 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr

785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile

805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu

820 825 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro

835 840 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala

850 855 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg

865 870 875 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp

885 890 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn

900 905 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys

930 935 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg

945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu

965 970 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys

980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg

995 1000 1005

Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser

1010 1015 1020

Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys

1025 1030 1035 1040

Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg Asn Lys Ala
1045 1050 1055
Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg
1060 1065 1070
Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly Ser Ser Ala Lys Asp
1075 1080 1085
Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu
1090 1095 1100
Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp Pro Glu Asn Arg Phe

1105 1110 1115 1120
Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu
1125 1130 1135

<210> 75
<211> 1175
<212> PRT
<213> Porphyromonas gingivalis
<400> 75

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15
Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
20 25 30

Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45
Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
50 55 60
Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95
Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Lys Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp

210 215 220

Arg Glu Glu Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr

325 330 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu

385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln Leu Leu Trp Pro Ser

450 455 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp

465 470 475 480

Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro

485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Ala

500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val

515 520 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu

565 570 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
610 615 620
Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625 630 635 640
Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
645 650 655
Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
660 665 670
Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685
Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
690 695 700
Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720
Ser Asp Arg Glu Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765
Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
770 775 780
Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785 790 795 800
Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
805 810 815
Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly

820 825 830
Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
835 840 845
Arg Ile Glu Asp Ala Phe Val Gly Ile Glu Tyr Ala Ser Trp Glu Asn
850 855 860

Lys Lys Lys Ile Glu Gln Leu Leu Gln Asp Leu Ser Leu Trp Glu Thr
865 870 875 880
Phe Glu Ser Lys Leu Lys Val Lys Ala Asp Lys Ile Asn Ile Ala Lys
885 890 895

Leu Lys Lys Glu Ile Leu Glu Ala Lys Glu His Pro Tyr His Asp Phe
900 905 910
Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
915 920 925
Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
930 935 940
Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945 950 955 960
Asp Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn His Val Lys Pro
965 970 975
Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
980 985 990
His Lys Glu Glu Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asp
995 1000 1005
Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg
1010 1015 1020
Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Ala Leu Ala Met
1025 1030 1035 1040

Glu Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu Leu Ala Lys
1045 1050 1055
Tyr Gln Thr Ala Arg Val Cys Ala Phe Glu Gln Thr Leu Glu Leu Glu
1060 1065 1070
Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro Asp Glu Ser Phe Arg
1075 1080 1085
Glu Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Asp Lys Trp Pro Asp
1090 1095 1100
Leu Gln Arg Glu Val Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser

1105 1110 1115 1120

His Asn Gln Tyr Pro Met Tyr Asp Glu Thr Ile Phe Ser Ser Ile Arg

1125 1130 1135

Lys Tyr Asp Pro Ser Ser Leu Asp Ala Ile Glu Glu Arg Met Gly Leu

1140 1145 1150

Asn Ile Ala His Arg Leu Ser Glu Glu Val Lys Leu Ala Lys Glu Met

1155 1160 1165

Val Glu Arg Ile Ile Gln Ala

1170 1175

<210> 76

<

211> 1125

<212> PRT

<213> Prevotella pallens

<400> 76

Met Lys Glu Glu Glu Lys Gly Lys Thr Pro Val Val Ser Thr Tyr Asn

1 5 10 15

Lys Asp Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His

20 25 30

Asn Val Tyr Ile Thr Val Asn His Ile Asn Lys Ile Leu Gly Glu Gly

35 40 45

Glu Ile Asn Arg Asp Gly Tyr Glu Asn Thr Leu Glu Lys Ser Trp Asn

50 55 60

Glu Ile Lys Asp Ile Asn Lys Lys Asp Arg Leu Ser Lys Leu Ile Ile

65 70 75 80

Lys His Phe Pro Phe Leu Glu Val Thr Thr Tyr Gln Arg Asn Ser Ala

85 90 95

Asp Thr Thr Lys Gln Lys Glu Glu Lys Gln Ala Glu Ala Gln Ser Leu

100 105 110

Glu Ser Leu Lys Lys Ser Phe Phe Val Phe Ile Tyr Lys Leu Arg Asp

115 120 125

Leu Arg Asn His Tyr Ser His Tyr Lys His Ser Lys Ser Leu Glu Arg

130 135 140

Pro Lys Phe Glu Glu Asp Leu Gln Glu Lys Met Tyr Asn Ile Phe Asp
145 150 155 160
Ala Ser Ile Gln Leu Val Lys Glu Asp Tyr Lys His Asn Thr Asp Ile
165 170 175
Lys Thr Glu Glu Asp Phe Lys His Leu Asp Arg Lys Gly Gln Phe Lys
180 185 190
Tyr Ser Phe Ala Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu
195 200 205

Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Val Gln
210 215 220
Lys Lys Leu Glu Gly Phe Lys Cys Ser Asn Glu Ser Tyr Gln Lys Met
225 230 235 240
Thr Asn Glu Val Phe Cys Arg Ser Arg Met Leu Leu Pro Lys Leu Arg
245 250 255
Leu Gln Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn
260 265 270
Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Arg Glu Glu
275 280 285

Asp Arg Lys Lys Phe Arg Val Pro Ile Glu Ile Ala Asp Glu Asp Tyr
290 295 300
Asp Ala Glu Gln Glu Pro Phe Lys Asn Ala Leu Val Arg His Gln Asp
305 310 315 320
Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Ile Phe
325 330 335
Thr Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile
340 345 350

Tyr Lys Lys Gln Ile Gly Asp Tyr Lys Glu Ser His His Leu Thr His
355 360 365
Lys Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Thr Lys Gln Asn Arg
370 375 380
Pro Asp Glu Trp Arg Lys Phe Val Lys Thr Phe Asn Ser Phe Glu Thr
385 390 395 400

Ser Lys Glu Pro Tyr Ile Pro Glu Thr Thr Pro His Tyr His Leu Glu
405 410 415
Asn Gln Lys Ile Gly Ile Arg Phe Arg Asn Asp Asn Asp Lys Ile Trp

420 425 430
Pro Ser Leu Lys Thr Asn Ser Glu Lys Asn Glu Lys Ser Lys Tyr Lys
435 440 445
Leu Asp Lys Ser Phe Gln Ala Glu Ala Phe Leu Ser Val His Glu Leu
450 455 460
Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu Lys Thr Glu Asn Thr Asp
465 470 475 480
Asn Asp Asn Glu Ile Glu Thr Lys Lys Lys Glu Asn Lys Asn Asp Lys
485 490 495

Gln Glu Lys His Lys Ile Glu Glu Ile Ile Glu Asn Lys Ile Thr Glu
500 505 510
Ile Tyr Ala Leu Tyr Asp Ala Phe Ala Asn Gly Lys Ile Asn Ser Ile
515 520 525
Asp Lys Leu Glu Glu Tyr Cys Lys Gly Lys Asp Ile Glu Ile Gly His
530 535 540
Leu Pro Lys Gln Met Ile Ala Ile Leu Lys Ser Glu His Lys Asp Met
545 550 555 560
Ala Thr Glu Ala Lys Arg Lys Gln Glu Glu Met Leu Ala Asp Val Gln

565 570 575
Lys Ser Leu Glu Ser Leu Asp Asn Gln Ile Asn Glu Glu Ile Glu Asn
580 585 590
Val Glu Arg Lys Asn Ser Ser Leu Lys Ser Gly Glu Ile Ala Ser Trp
595 600 605
Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu
610 615 620
Gly Asn Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met
625 630 635 640

Leu Gln Arg Ser Leu Ala Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg

645 650 655

Tyr Phe Arg Gln Val Asn Leu Ile Glu Ser Ser Asn Pro His Pro Phe

660 665 670

Leu Asn Asn Thr Glu Trp Glu Lys Cys Asn Asn Ile Leu Ser Phe Tyr

675 680 685

Arg Ser Tyr Leu Glu Ala Lys Lys Asn Phe Leu Glu Ser Leu Lys Pro

690 695 700

Glu Asp Trp Glu Lys Asn Gln Tyr Phe Leu Met Leu Lys Glu Pro Lys

705 710 715 720

Thr Asn Cys Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu

725 730 735

Pro Arg Gly Ile Phe Thr Glu Pro Ile Arg Lys Trp Phe Met Glu His

740 745 750

Arg Lys Asn Ile Thr Val Ala Glu Leu Lys Arg Val Gly Leu Val Ala

755 760 765

Lys Val Ile Pro Leu Phe Phe Ser Glu Glu Tyr Lys Asp Ser Val Gln

770 775 780

Pro Phe Tyr Asn Tyr Leu Phe Asn Val Gly Asn Ile Asn Lys Pro Asp

785 790 795 800

Glu Lys Asn Phe Leu Asn Cys Glu Glu Arg Arg Glu Leu Leu Arg Lys

805 810 815

Lys Lys Asp Glu Phe Lys Lys Met Thr Asp Lys Glu Lys Glu Glu Asn

820 825 830

Pro Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu

835 840 845

Arg Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Met Glu

850 855 860

Leu Phe Asn Lys Lys Lys Ile Lys Glu Leu Asn Val Glu Lys Ile Tyr

865 870 875 880

Leu Lys Asn Ile Asn Thr Asn Thr Thr Lys Lys Glu Lys Asn Thr Glu

885 890 895

Glu Lys Asn Gly Glu Glu Lys Ile Ile Lys Glu Lys Asn Asn Ile Leu

900 905 910

Asn Arg Ile Met Pro Met Arg Leu Pro Ile Lys Val Tyr Gly Arg Glu

915 920 925

Asn Phe Ser Lys Asn Lys Lys Lys Lys Ile Arg Arg Asn Thr Phe Phe

930 935 940

Thr Val Tyr Ile Glu Glu Lys Gly Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ala Leu Glu Arg Asp Arg Arg Leu Gly Gly Leu Phe Ser Phe

965 970 975

Val Lys Thr His Ser Lys Ala Glu Ser Lys Ser Asn Thr Ile Ser Lys

980 985 990

Ser Arg Val Glu Tyr Glu Leu Gly Glu Tyr Gln Lys Ala Arg Ile Glu

995 1000 1005

Ile Ile Lys Asp Met Leu Ala Leu Glu Glu Thr Leu Ile Asp Lys Tyr

1010 1015 1020

Asn Ser Leu Asp Thr Asp Asn Phe His Asn Met Leu Thr Gly Trp Leu

1025 1030 1035 1040

Lys Leu Lys Asp Glu Pro Asp Lys Ala Ser Phe Gln Asn Asp Val Asp

1045 1050 1055

Leu Leu Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met

1060 1065 1070

Arg Asn Arg Ile Ala Phe Ala Asn Ile Asn Pro Phe Ser Leu Ser Ser

1075 1080 1085

Ala Asn Thr Ser Glu Glu Lys Gly Leu Gly Ile Ala Asn Gln Leu Lys

1090 1095 1100

Asp Lys Thr His Lys Thr Ile Glu Lys Ile Ile Glu Ile Glu Lys Pro

1105 1110 1115 1120

Ile Glu Thr Lys Glu

1125

<210> 77

<211> 1157

<212> PRT

<213> Myroides odoratimimus

<400> 77

Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe

1 5 10 15

Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn

20 25 30

Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg

35 40 45

Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn

50 55 60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu

65 70 75 80

Lys Arg Val Ala Ile Phe Ala Ser Tyr Phe Pro Ile Leu Glu Thr Val

85 90 95

Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu

100 105 110

Ser Gln Arg Ile Arg Gln Phe Arg Glu Met Leu Ile Ser Leu Val Thr

115 120 125

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Glu

130 135 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Leu Val

145 150 155 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr

165 170 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile

180 185 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys

195 200 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe

210 215 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Glu Thr Val Val Ala

225 230 235 240

Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser Asn Asp

245 250 255

Pro Asp Phe Ala Leu Asn Ile Ser Glu Lys Gly Ile Val Tyr Leu Leu

260 265 270

Ser Phe Phe Leu Thr Asn Lys Glu Met Asp Ser Leu Lys Ala Asn Leu

275 280 285

Thr Gly Phe Lys Gly Lys Val Asp Arg Glu Ser Gly Asn Ser Ile Lys

290 295 300

Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg Gly Leu

305 310 315 320

Lys Gln Lys Ile Arg Thr Ser Glu Glu Gly Val Lys Glu Thr Leu Leu

325 330 335

Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val Tyr Gln

340 345 350

His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp Asn Glu

355 360 365

Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu Ser Arg

370 375 380

Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe Asn Tyr

385 390 395 400

Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr Leu Arg

405 410 415

Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr Lys Gln

420 425 430

Leu Gly Lys Val Glu Ser Asp Arg Ile Ile Lys Glu Lys Val Thr Val

435 440 445

Phe Ala Arg Leu Lys Asp Ile Asn Ser Ala Lys Ala Asn Tyr Phe His

450 455 460

Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp Thr Leu

465 470 475 480

Phe Pro Asn Pro Ser Tyr Asp Phe Pro Lys Glu His Thr Leu Gln His

485 490 495

Gln Gly Glu Gln Lys Asn Ala Gly Lys Ile Gly Ile Tyr Val Lys Leu

500 505 510

Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala Arg Lys

515 520 525

Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys Tyr Asp

530 535 540

Ile Ile Thr Gln Ile Ile Glu Ala Asn Asp Asn Val Lys Ser Glu Lys

545 550 555 560

Pro Leu Val Phe Thr Gly Gln Pro Ile Ala Tyr Leu Ser Met Asn Asp

565 570 575

Ile His Ser Met Leu Phe Ser Leu Leu Thr Asp Asn Ala Glu Leu Lys

580 585 590

Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile Gly Lys

595 600 605

Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile Leu Lys

610 615 620

Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys Ile Thr

625 630 635 640

Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile Leu Glu

645 650 655

Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr Lys Phe

660 665 670

Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn Ala Glu

675 680 685

Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met

690 695 700

Thr Glu Glu Phe Lys Ser Lys Trp Lys Gly Tyr Gln His Thr Glu Leu

705 710 715 720

Gln Lys Leu Phe Ala Tyr Tyr Asp Thr Ser Lys Ser Asp Leu Asp Leu

725 730 735

Ile Leu Ser Asp Met Val Met Val Lys Asp Tyr Pro Ile Glu Leu Ile
740 745 750

Ala Leu Val Lys Lys Ser Arg Thr Leu Val Asp Phe Leu Asn Lys Tyr
755 760 765

Leu Glu Ala Arg Leu Gly Tyr Met Glu Asn Val Ile Thr Arg Val Lys
770 775 780

Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu Cys Phe
785 790 795 800

Thr Phe Leu Lys Lys Ser Asn Tyr Thr Val Val Ser Leu Asp Lys Gln
805 810 815

Val Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly Phe Met
820 825 830

Asp Asp Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Gln Gln His Lys
835 840 845

Glu Lys Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser Asn Tyr
850 855 860

Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Thr Thr Glu Asp Lys

865 870 875 880

Arg Glu Lys Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln Lys Asn
885 890 895

Asp Val Phe Thr Leu Met Met Val Asn Tyr Met Leu Glu Glu Val Leu
900 905 910

Lys Leu Ser Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr Gln Thr
915 920 925

Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr Gln Glu
930 935 940

Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln Leu Cys
945 950 955 960

Glu Gly Leu Val Arg Ile Asp Lys Val Lys Leu Lys Asp Ile Gly Asn
965 970 975

Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu Thr Tyr
980 985 990

Gln Ser Asp Ile Val Trp Ser Ala Tyr Leu Ser Asn Glu Val Asp Ser
995 1000 1005
Asn Lys Leu Tyr Val Ile Glu Arg Gln Leu Asp Asn Tyr Glu Ser Ile

1010 1015 1020
Arg Ser Lys Glu Leu Leu Lys Glu Val Gln Glu Ile Glu Cys Ser Val
1025 1030 1035 1040
Tyr Asn Gln Val Ala Asn Lys Glu Ser Leu Lys Gln Ser Gly Asn Glu
1045 1050 1055
Asn Phe Lys Gln Tyr Val Leu Gln Gly Leu Val Pro Ile Gly Met Asp
1060 1065 1070
Val Arg Glu Met Leu Ile Leu Ser Thr Asp Val Lys Phe Ile Lys Glu
1075 1080 1085

Glu Ile Ile Gln Leu Gly Gln Ala Gly Glu Val Glu Gln Asp Leu Tyr
1090 1095 1100
Ser Leu Ile Tyr Ile Arg Asn Lys Phe Ala His Asn Gln Leu Pro Ile
1105 1110 1115 1120
Lys Glu Phe Phe Asp Phe Cys Glu Asn Asn Tyr Arg Ser Ile Ser Asp
1125 1130 1135
Asn Glu Tyr Tyr Ala Glu Tyr Tyr Met Glu Ile Phe Arg Ser Ile Lys
1140 1145 1150
Glu Lys Tyr Thr Ser

1155

<210> 78
<211> 1157
<212> PRT
<213> Myroides odoratimimus
<400> 78
Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1 5 10 15
Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
20 25 30
Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg

35 40 45

Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn

50 55 60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu

65 70 75 80

Lys Arg Val Ala Ile Phe Ala Ser Tyr Phe Pro Ile Leu Glu Thr Val

85 90 95

Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu

100 105 110

Ser Gln Arg Ile Arg Gln Phe Arg Glu Met Leu Ile Ser Leu Val Thr

115 120 125

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Glu

130 135 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Leu Val

145 150 155 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr

165 170 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile

180 185 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys

195 200 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe

210 215 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Glu Thr Val Val Ala

225 230 235 240

Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser Asn Asp

245 250 255

Pro Asp Phe Ala Leu Asn Ile Ser Glu Lys Gly Ile Val Tyr Leu Leu

260 265 270

Ser Phe Phe Leu Thr Asn Lys Glu Met Asp Ser Leu Lys Ala Asn Leu

275 280 285

Thr Gly Phe Lys Gly Lys Val Asp Arg Glu Ser Gly Asn Ser Ile Lys
290 295 300
Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg Gly Leu
305 310 315 320
Lys Gln Lys Ile Arg Thr Ser Glu Glu Gly Val Lys Glu Thr Leu Leu
325 330 335
Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val Tyr Gln

340 345 350
His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp Asn Glu
355 360 365
Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu Ser Arg
370 375 380
Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe Asn Tyr
385 390 395 400
Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr Leu Arg
405 410 415

Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr Lys Gln
420 425 430
Leu Gly Lys Val Glu Ser Asp Arg Ile Ile Lys Glu Lys Val Thr Val
435 440 445
Phe Ala Arg Leu Lys Asp Ile Asn Ser Ala Lys Ala Ser Tyr Phe His
450 455 460
Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp Thr Leu
465 470 475 480
Phe Pro Asn Pro Ser Tyr Asp Phe Pro Lys Glu His Thr Leu Gln His

485 490 495
Gln Gly Glu Gln Lys Asn Ala Gly Lys Ile Gly Ile Tyr Val Lys Leu
500 505 510
Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala Arg Lys
515 520 525
Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys Tyr Asp
530 535 540

Ile Ile Thr Gln Ile Ile Glu Ala Asn Asp Asn Val Lys Ser Glu Lys
545 550 555 560

Pro Leu Val Phe Thr Gly Gln Pro Ile Ala Tyr Leu Ser Met Asn Asp
565 570 575
Ile His Ser Met Leu Phe Ser Leu Leu Thr Asp Asn Ala Glu Leu Lys
580 585 590
Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile Gly Lys
595 600 605
Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile Leu Lys
610 615 620
Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys Ile Thr

625 630 635 640
Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile Leu Glu
645 650 655
Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr Lys Phe
660 665 670
Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn Ala Glu
675 680 685
Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met
690 695 700

Phe Lys Glu Ser Lys Ser Lys Trp Lys Gly Tyr Gln His Thr Glu Leu
705 710 715 720
Gln Lys Leu Phe Ala Tyr Phe Asp Thr Ser Lys Ser Asp Leu Glu Leu
725 730 735
Ile Leu Ser Asp Met Val Met Val Lys Asp Tyr Pro Ile Glu Leu Ile
740 745 750
Asp Leu Val Arg Lys Ser Arg Thr Leu Val Asp Phe Leu Asn Lys Tyr
755 760 765
Leu Glu Ala Arg Leu Gly Tyr Ile Glu Asn Val Ile Thr Arg Val Lys

770 775 780
Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu Cys Phe

785 790 795 800

Ala Phe Leu Lys Glu Ser Asn Tyr Thr Val Ala Ser Leu Asp Lys Gln

805 810 815

Ile Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly Phe Met

820 825 830

Asp Ser Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Gln Gln His Lys

835 840 845

Glu Asp Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser Asn Tyr

850 855 860

Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Ile Thr Glu Asp Lys

865 870 875 880

Arg Glu Gln Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln Lys Asn

885 890 895

Asp Val Phe Thr Leu Met Met Val Asn Tyr Met Leu Glu Glu Val Leu

900 905 910

Lys Leu Pro Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr Gln Thr

915 920 925

Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr Gln Glu

930 935 940

Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln Leu Cys

945 950 955 960

Glu Gly Leu Val Arg Ile Asp Lys Val Lys Leu Lys Asp Ile Gly Asn

965 970 975

Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu Thr Tyr

980 985 990

Gln Ser Asp Ile Val Trp Ser Gly Tyr Leu Ser Asn Glu Val Asp Ser

995 1000 1005

Asn Lys Leu Tyr Val Ile Glu Arg Gln Leu Asp Asn Tyr Glu Ser Ile

1010 1015 1020

Arg Ser Lys Glu Leu Leu Lys Glu Val Gln Glu Ile Glu Cys Ile Val

1025 1030 1035 1040

Tyr Asn Gln Val Ala Asn Lys Glu Ser Leu Lys Gln Ser Gly Asn Glu

1045 1050 1055

Asn Phe Lys Gln Tyr Val Leu Gln Gly Leu Leu Pro Arg Gly Thr Asp

1060 1065 1070

Val Arg Glu Met Leu Ile Leu Ser Thr Asp Val Lys Phe Lys Lys Glu

1075 1080 1085

Glu Ile Met Gln Leu Gly Gln Val Arg Glu Val Glu Gln Asp Leu Tyr

1090 1095 1100

Ser Leu Ile Tyr Ile Arg Asn Lys Phe Ala His Asn Gln Leu Pro Ile

1105 1110 1115 1120

Lys Glu Phe Phe Asp Phe Cys Glu Asn Asn Tyr Arg Pro Ile Ser Asp

1125 1130 1135

Asn Glu Tyr Tyr Ala Glu Tyr Tyr Met Glu Ile Phe Arg Ser Ile Lys

1140 1145 1150

Glu Lys Tyr Ala Ser

1155

<210> 79

<211> 1127

<212> PRT

<213> Prevotella sp.

<400> 79

Met Gln Lys Gln Asp Lys Leu Phe Val Asp Arg Lys Lys Asn Ala Ile

1 5 10 15

Phe Ala Phe Pro Lys Tyr Ile Thr Ile Met Glu Asn Gln Glu Lys Pro

20 25 30

Glu Pro Ile Tyr Tyr Glu Leu Thr Asp Lys His Phe Trp Ala Ala Phe

35 40 45

Leu Asn Leu Ala Arg His Asn Val Tyr Thr Thr Ile Asn His Ile Asn

50 55 60

Arg Arg Leu Glu Ile Ala Glu Leu Lys Asp Asp Gly Tyr Met Met Gly

65 70 75 80

Ile Lys Gly Ser Trp Asn Glu Gln Ala Lys Lys Leu Asp Lys Lys Val

85 90 95

```
Arg Leu Arg Asp Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala
            100                 105                 110

Ala Tyr Glu Ile Thr Asn Ser Lys Ser Pro Asn Asn Lys Glu Gln Arg
        115                 120                 125
Glu Lys Glu Gln Ser Glu Ala Leu Ser Leu Asn Asn Leu Lys Asn Val
    130                 135                 140
Leu Phe Ile Phe Leu Glu Lys Leu Gln Val Leu Arg Asn Tyr Tyr Ser
145                 150                 155                 160
His Tyr Lys Tyr Ser Glu Glu Ser Pro Lys Pro Ile Phe Glu Thr Ser
                165                 170                 175
Leu Leu Lys Asn Met Tyr Lys Val Phe Asp Ala Asn Val Arg Leu Val

            180                 185                 190
Lys Arg Asp Tyr Met His His Glu Asn Ile Asp Met Gln Arg Asp Phe
        195                 200                 205
Thr His Leu Asn Arg Lys Lys Gln Val Gly Arg Thr Lys Asn Ile Ile
    210                 215                 220
Asp Ser Pro Asn Phe His Tyr His Phe Ala Asp Lys Glu Gly Asn Met
225                 230                 235                 240
Thr Ile Ala Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Asp Lys Lys
                245                 250                 255

Asp Ala Ile Trp Met Gln Lys Lys Leu Lys Gly Phe Lys Asp Gly Arg
            260                 265                 270
Asn Leu Arg Glu Gln Met Thr Asn Glu Val Phe Cys Arg Ser Arg Ile
        275                 280                 285
Ser Leu Pro Lys Leu Lys Leu Glu Asn Val Gln Thr Lys Asp Trp Met
    290                 295                 300
Gln Leu Asp Met Leu Asn Glu Leu Val Arg Cys Pro Lys Ser Leu Tyr
305                 310                 315                 320
Glu Arg Leu Arg Glu Lys Asp Arg Glu Ser Phe Lys Val Pro Phe Asp

                325                 330                 335
Ile Phe Ser Asp Asp Tyr Asp Ala Glu Glu Glu Pro Phe Lys Asn Thr
```

340 345 350

Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Val Leu Arg Tyr Phe

355 360 365

Asp Leu Asn Glu Ile Phe Glu Gln Leu Arg Phe Gln Ile Asp Leu Gly

370 375 380

Thr Tyr His Phe Ser Ile Tyr Asn Lys Arg Ile Gly Asp Glu Asp Glu

385 390 395 400

Val Arg His Leu Thr His His Leu Tyr Gly Phe Ala Arg Ile Gln Asp

405 410 415

Phe Ala Pro Gln Asn Gln Pro Glu Glu Trp Arg Lys Leu Val Lys Asp

420 425 430

Leu Asp His Phe Glu Thr Ser Gln Glu Pro Tyr Ile Ser Lys Thr Ala

435 440 445

Pro His Tyr His Leu Glu Asn Glu Lys Ile Gly Ile Lys Phe Cys Ser

450 455 460

Thr His Asn Asn Leu Phe Pro Ser Leu Lys Arg Glu Lys Thr Cys Asn

465 470 475 480

Gly Arg Ser Lys Phe Asn Leu Gly Thr Gln Phe Thr Ala Glu Ala Phe

485 490 495

Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu

500 505 510

Thr Lys Asp Tyr Ser Arg Lys Glu Ser Ala Asp Lys Val Glu Gly Ile

515 520 525

Ile Arg Lys Glu Ile Ser Asn Ile Tyr Ala Ile Tyr Asp Ala Phe Ala

530 535 540

Asn Asn Glu Ile Asn Ser Ile Ala Asp Leu Thr Cys Arg Leu Gln Lys

545 550 555 560

Thr Asn Ile Leu Gln Gly His Leu Pro Lys Gln Met Ile Ser Ile Leu

565 570 575

Glu Gly Arg Gln Lys Asp Met Glu Lys Glu Ala Glu Arg Lys Ile Gly

580 585 590

Glu Met Ile Asp Asp Thr Gln Arg Arg Leu Asp Leu Leu Cys Lys Gln

595 600 605

Thr Asn Gln Lys Ile Arg Ile Gly Lys Arg Asn Ala Gly Leu Leu Lys

610 615 620

Ser Gly Lys Ile Ala Asp Trp Leu Val Ser Asp Met Met Arg Phe Gln

625 630 635 640

Pro Val Gln Lys Asp Thr Asn Asn Ala Pro Ile Asn Asn Ser Lys Ala

645 650 655

Asn Ser Thr Glu Tyr Arg Met Leu Gln His Ala Leu Ala Leu Phe Gly

660 665 670

Ser Glu Ser Ser Arg Leu Lys Ala Tyr Phe Arg Gln Met Asn Leu Val

675 680 685

Gly Asn Ala Asn Pro His Pro Phe Leu Ala Glu Thr Gln Trp Glu His

690 695 700

Gln Thr Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Glu Ala Arg Lys

705 710 715 720

Lys Tyr Leu Lys Gly Leu Lys Pro Gln Asn Trp Lys Gln Tyr Gln His

725 730 735

Phe Leu Ile Leu Lys Val Gln Lys Thr Asn Arg Asn Thr Leu Val Thr

740 745 750

Gly Trp Lys Asn Ser Phe Asn Leu Pro Arg Gly Ile Phe Thr Gln Pro

755 760 765

Ile Arg Glu Trp Phe Glu Lys His Asn Asn Ser Lys Arg Ile Tyr Asp

770 775 780

Gln Ile Leu Ser Phe Asp Arg Val Gly Phe Val Ala Lys Ala Ile Pro

785 790 795 800

Leu Tyr Phe Ala Glu Glu Tyr Lys Asp Asn Val Gln Pro Phe Tyr Asp

805 810 815

Tyr Pro Phe Asn Ile Gly Asn Lys Leu Lys Pro Gln Lys Gly Gln Phe

820 825 830

Leu Asp Lys Lys Glu Arg Val Glu Leu Trp Gln Lys Asn Lys Glu Leu

835 840 845

Phe Lys Asn Tyr Pro Ser Glu Lys Asn Lys Thr Asp Leu Ala Tyr Leu
850 855 860
Asp Phe Leu Ser Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Ile Lys
865 870 875 880
Asn Gln Asp Ile Val Thr Trp Leu Met Phe Lys Glu Leu Phe Lys Thr
885 890 895
Thr Thr Val Glu Gly Leu Lys Ile Gly Glu Ile His Leu Arg Asp Ile

900 905 910
Asp Thr Asn Thr Ala Asn Glu Glu Ser Asn Asn Ile Leu Asn Arg Ile
915 920 925
Met Pro Met Lys Leu Pro Val Lys Thr Tyr Glu Thr Asp Asn Lys Gly
930 935 940
Asn Ile Leu Lys Glu Arg Pro Leu Ala Thr Phe Tyr Ile Glu Glu Thr
945 950 955 960
Glu Thr Lys Val Leu Lys Gln Gly Asn Phe Lys Val Leu Ala Lys Asp
965 970 975

Arg Arg Leu Asn Gly Leu Leu Ser Phe Ala Glu Thr Thr Asp Ile Asp
980 985 990
Leu Glu Lys Asn Pro Ile Thr Lys Leu Ser Val Asp Tyr Glu Leu Ile
995 1000 1005
Lys Tyr Gln Thr Thr Arg Ile Ser Ile Phe Glu Met Thr Leu Gly Leu
1010 1015 1020
Glu Lys Lys Leu Ile Asp Lys Tyr Ser Thr Leu Pro Thr Asp Ser Phe
1025 1030 1035 1040
Arg Asn Met Leu Glu Arg Trp Leu Gln Cys Lys Ala Asn Arg Pro Glu

1045 1050 1055
Leu Lys Asn Tyr Val Asn Ser Leu Ile Ala Val Arg Asn Ala Phe Ser
1060 1065 1070
His Asn Gln Tyr Pro Met Tyr Asp Ala Thr Leu Phe Ala Glu Val Lys
1075 1080 1085
Lys Phe Thr Leu Phe Pro Ser Val Asp Thr Lys Lys Ile Glu Leu Asn
1090 1095 1100

Ile Ala Pro Gln Leu Leu Glu Ile Val Gly Lys Ala Ile Lys Glu Ile
1105 1110 1115 1120

Glu Lys Ser Glu Asn Lys Asn
1125

<210> 80
<211> 1119
<212> PRT
<213> Porphyromonas gingivalis
<400> 80

Met Thr Glu Gln Asn Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15
Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
20 25 30
Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
50 55 60
Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95
Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
100 105 110
Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125
Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140
Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145 150 155 160
Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175
Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205
Arg Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
210 215 220
Arg Glu Glu Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240
Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255
Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe

260 265 270
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285
Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
290 295 300
Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305 310 315 320
Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
325 330 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
340 345 350
Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365
Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380
Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385 390 395 400
Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415
Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
420 425 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
435 440 445
Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
450 455 460
Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
485 490 495
Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Asp Glu Ala
500 505 510
Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
515 520 525
Tyr Ala Val Tyr Asp Ala Phe Ala Arg Gly Glu Ile Asn Thr Arg Asp
530 535 540
Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560
Pro Arg Gln Met Ile Gly Ile Leu Ser Gln Glu His Lys Asp Met Glu
565 570 575
Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Val Asp Thr Asp His
580 585 590
Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
595 600 605
Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625 630 635 640
Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
645 650 655
Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
660 665 670
Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720
Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
740 745 750
Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765

Gly Leu Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
770 775 780
Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785 790 795 800
Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
805 810 815
Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
820 825 830
Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala

835 840 845

Lys Glu His Pro Tyr Leu Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
850 855 860
Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Ile Cys
865 870 875 880
Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
885 890 895
Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val Gln Glu Gln Gly Asn Leu
900 905 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr

915 920 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala

930 935 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe

965 970 975

Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu

980 985 990

Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala

995 1000 1005

Phe Glu Gln Thr Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Tyr Pro

1010 1015 1020

His Leu Pro Asp Lys Asn Phe Arg Lys Met Leu Glu Ser Trp Ser Asp

1025 1030 1035 1040

Pro Leu Leu Asp Lys Trp Pro Asp Leu His Gly Asn Val Arg Leu Leu

1045 1050 1055

Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp

1060 1065 1070

Glu Ala Val Phe Ser Ser Ile Arg Lys Tyr Asp Pro Ser Ser Pro Asp

1075 1080 1085

Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu Ser Glu

1090 1095 1100

Glu Val Lys Gln Ala Lys Glu Met Ala Glu Arg Ile Ile Gln Ala

1105 1110 1115

<210> 81

<211> 1154

<212> PRT

<213> Paludibacter propionicigenes

<400> 81

Met Lys Thr Ser Ala Asn Asn Ile Tyr Phe Asn Gly Ile Asn Ser Phe

1 5 10 15

Lys Lys Ile Phe Asp Ser Lys Gly Ala Ile Ala Pro Ile Ala Glu Lys

20 25 30

Ser Cys Arg Asn Phe Asp Ile Lys Ala Gln Asn Asp Val Asn Lys Glu
35 40 45
Gln Arg Ile His Tyr Phe Ala Val Gly His Thr Phe Lys Gln Leu Asp
50 55 60

Thr Glu Asn Leu Phe Glu Tyr Val Leu Asp Glu Asn Leu Arg Ala Lys
65 70 75 80
Arg Pro Thr Arg Phe Ile Ser Leu Gln Gln Phe Asp Lys Glu Phe Ile
85 90 95
Glu Asn Ile Lys Arg Leu Ile Ser Asp Ile Arg Asn Ile Asn Ser His
100 105 110
Tyr Ile His Arg Phe Asp Pro Leu Lys Ile Asp Ala Val Pro Thr Asn
115 120 125
Ile Ile Asp Phe Leu Lys Glu Ser Phe Glu Leu Ala Val Ile Gln Ile

130 135 140
Tyr Leu Lys Glu Lys Gly Ile Asn Tyr Leu Gln Phe Ser Glu Asn Pro
145 150 155 160
His Ala Asp Gln Lys Leu Val Ala Phe Leu His Asp Lys Phe Leu Pro
165 170 175
Leu Asp Glu Lys Lys Thr Ser Met Leu Gln Asn Glu Thr Pro Gln Leu
180 185 190
Lys Glu Tyr Lys Glu Tyr Arg Lys Tyr Phe Lys Thr Leu Ser Lys Gln
195 200 205

Ala Ala Ile Asp Gln Leu Leu Phe Ala Glu Lys Glu Thr Asp Tyr Ile
210 215 220
Trp Asn Leu Phe Asp Ser His Pro Val Leu Thr Ile Ser Ala Gly Lys
225 230 235 240
Tyr Leu Ser Phe Tyr Ser Cys Leu Phe Leu Leu Ser Met Phe Leu Tyr
245 250 255
Lys Ser Glu Ala Asn Gln Leu Ile Ser Lys Ile Lys Gly Phe Lys Lys
260 265 270
Asn Thr Thr Glu Glu Glu Lys Ser Lys Arg Glu Ile Phe Thr Phe Phe

275 280 285

Ser Lys Arg Phe Asn Ser Met Asp Ile Asp Ser Glu Glu Asn Gln Leu

290 295 300

Val Lys Phe Arg Asp Leu Ile Leu Tyr Leu Asn His Tyr Pro Val Ala

305 310 315 320

Trp Asn Lys Asp Leu Glu Leu Asp Ser Ser Asn Pro Ala Met Thr Asp

325 330 335

Lys Leu Lys Ser Lys Ile Ile Glu Leu Glu Ile Asn Arg Ser Phe Pro

340 345 350

Leu Tyr Glu Gly Asn Glu Arg Phe Ala Thr Phe Ala Lys Tyr Gln Ile

355 360 365

Trp Gly Lys Lys His Leu Gly Lys Ser Ile Glu Lys Glu Tyr Ile Asn

370 375 380

Ala Ser Phe Thr Asp Glu Glu Ile Thr Ala Tyr Thr Tyr Glu Thr Asp

385 390 395 400

Thr Cys Pro Glu Leu Lys Asp Ala His Lys Lys Leu Ala Asp Leu Lys

405 410 415

Ala Ala Lys Gly Leu Phe Gly Lys Arg Lys Glu Lys Asn Glu Ser Asp

420 425 430

Ile Lys Lys Thr Glu Thr Ser Ile Arg Glu Leu Gln His Glu Pro Asn

435 440 445

Pro Ile Lys Asp Lys Leu Ile Gln Arg Ile Glu Lys Asn Leu Leu Thr

450 455 460

Val Ser Tyr Gly Arg Asn Gln Asp Arg Phe Met Asp Phe Ser Ala Arg

465 470 475 480

Phe Leu Ala Glu Ile Asn Tyr Phe Gly Gln Asp Ala Ser Phe Lys Met

485 490 495

Tyr His Phe Tyr Ala Thr Asp Glu Gln Asn Ser Glu Leu Glu Lys Tyr

500 505 510

Glu Leu Pro Lys Asp Lys Lys Lys Tyr Asp Ser Leu Lys Phe His Gln

515 520 525

Gly Lys Leu Val His Phe Ile Ser Tyr Lys Glu His Leu Lys Arg Tyr

530 535 540

Glu Ser Trp Asp Asp Ala Phe Val Ile Glu Asn Asn Ala Ile Gln Leu

545 550 555 560

Lys Leu Ser Phe Asp Gly Val Glu Asn Thr Val Thr Ile Gln Arg Ala

565 570 575

Leu Leu Ile Tyr Leu Leu Glu Asp Ala Leu Arg Asn Ile Gln Asn Asn

580 585 590

Thr Ala Glu Asn Ala Gly Lys Gln Leu Leu Gln Glu Tyr Tyr Ser His

595 600 605

Asn Lys Ala Asp Leu Ser Ala Phe Lys Gln Ile Leu Thr Gln Gln Asp

610 615 620

Ser Ile Glu Pro Gln Gln Lys Thr Glu Phe Lys Lys Leu Leu Pro Arg

625 630 635 640

Arg Leu Leu Asn Asn Tyr Ser Pro Ala Ile Asn His Leu Gln Thr Pro

645 650 655

His Ser Ser Leu Pro Leu Ile Leu Glu Lys Ala Leu Leu Ala Glu Lys

660 665 670

Arg Tyr Cys Ser Leu Val Val Lys Ala Lys Ala Glu Gly Asn Tyr Asp

675 680 685

Asp Phe Ile Lys Arg Asn Lys Gly Lys Gln Phe Lys Leu Gln Phe Ile

690 695 700

Arg Lys Ala Trp Asn Leu Met Tyr Phe Arg Asn Ser Tyr Leu Gln Asn

705 710 715 720

Val Gln Ala Ala Gly His His Lys Ser Phe His Ile Glu Arg Asp Glu

725 730 735

Phe Asn Asp Phe Ser Arg Tyr Met Phe Ala Phe Glu Glu Leu Ser Gln

740 745 750

Tyr Lys Tyr Tyr Leu Asn Glu Met Phe Glu Lys Lys Gly Phe Phe Glu

755 760 765

Asn Asn Glu Phe Lys Ile Leu Phe Gln Ser Gly Thr Ser Leu Glu Asn

770 775 780

Leu Tyr Glu Lys Thr Lys Gln Lys Phe Glu Ile Trp Leu Ala Ser Asn
785 790 795 800
Thr Ala Lys Thr Asn Lys Pro Asp Asn Tyr His Leu Asn Asn Tyr Glu
805 810 815
Gln Gln Phe Ser Asn Gln Leu Phe Phe Ile Asn Leu Ser His Phe Ile
820 825 830
Asn Tyr Leu Lys Ser Thr Gly Lys Leu Gln Thr Asp Ala Asn Gly Gln
835 840 845
Ile Ile Tyr Glu Ala Leu Asn Asn Val Gln Tyr Leu Ile Pro Glu Tyr

850 855 860
Tyr Tyr Thr Asp Lys Pro Glu Arg Ser Glu Ser Lys Ser Gly Asn Lys
865 870 875 880
Leu Tyr Asn Lys Leu Lys Ala Thr Lys Leu Glu Asp Ala Leu Leu Tyr
885 890 895
Glu Met Ala Met Cys Tyr Leu Lys Ala Asp Lys Gln Ile Ala Asp Lys
900 905 910
Ala Lys His Pro Ile Thr Lys Leu Leu Thr Ser Asp Val Glu Phe Asn
915 920 925

Ile Thr Asn Lys Glu Gly Ile Gln Leu Tyr His Leu Leu Val Pro Phe
930 935 940
Lys Lys Ile Asp Ala Phe Ile Gly Leu Lys Met His Lys Glu Gln Gln
945 950 955 960
Asp Lys Lys His Pro Thr Ser Phe Leu Ala Asn Ile Val Asn Tyr Leu
965 970 975
Glu Leu Val Lys Asn Asp Lys Asp Ile Arg Lys Thr Tyr Glu Ala Phe
980 985 990
Ser Thr Asn Pro Val Lys Arg Thr Leu Thr Tyr Asp Asp Leu Ala Lys

995 1000 1005
Ile Asp Gly His Leu Ile Ser Lys Ser Ile Lys Phe Thr Asn Val Thr
1010 1015 1020
Leu Glu Leu Glu Arg Tyr Phe Ile Phe Lys Glu Ser Leu Ile Val Lys
1025 1030 1035 1040

Lys Gly Asn Asn Ile Asp Phe Lys Tyr Ile Lys Gly Leu Arg Asn Tyr
1045 1050 1055
Tyr Asn Asn Glu Lys Lys Lys Asn Glu Gly Ile Arg Asn Lys Ala Phe
1060 1065 1070

His Phe Gly Ile Pro Asp Ser Lys Ser Tyr Asp Gln Leu Ile Arg Asp
1075 1080 1085
Ala Glu Val Met Phe Ile Ala Asn Glu Val Lys Pro Thr His Ala Thr
1090 1095 1100
Lys Tyr Thr Asp Leu Asn Lys Gln Leu His Thr Val Cys Asp Lys Leu
1105 1110 1115 1120
Met Glu Thr Val His Asn Asp Tyr Phe Ser Lys Glu Gly Asp Gly Lys
1125 1130 1135
Lys Lys Arg Glu Ala Ala Gly Gln Lys Tyr Phe Glu Asn Ile Ile Ser

1140 1145 1150
Ala Lys

<210> 82
<211> 1119
<212> PRT
<213> Porphyromonas gingivalis
<400> 82
Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15
Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
20 25 30
Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
50 55 60
Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys

85 90 95

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp

210 215 220

Arg Glu Gly Thr Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Glu Asp Thr Asp Gly Ala

325 330 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
340 345 350
Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365
Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380
Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385 390 395 400
Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415
Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
420 425 430
Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
435 440 445
Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
450 455 460
Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465 470 475 480

Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Ala His Glu Leu Met Pro
485 490 495
Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Ala
500 505 510
Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
515 520 525
Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
530 535 540
Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560
Pro Arg Gln Met Ile Gly Ile Leu Ser Gln Glu His Lys Asp Met Glu
565 570 575
Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Met Ala Asp Thr Asp His
580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
595 600 605
Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625 630 635 640
Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
645 650 655
Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
660 665 670
Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
675 680 685
Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700
Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720
Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
740 745 750
Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765

Gly Leu Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
770 775 780
Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Trp Val Gln Pro
785 790 795 800
Phe Tyr Asn Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
805 810 815
Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
820 825 830
Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala

835 840 845

Lys Glu His Pro Tyr Leu Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg

850 855 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Ile Cys

865 870 875 880

Gly Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr

885 890 895

Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val Gln Glu Gln Gly Ser Leu

900 905 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr

915 920 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala

930 935 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe

965 970 975

Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu

980 985 990

Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala

995 1000 1005

Phe Glu Gln Thr Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Cys Pro

1010 1015 1020

His Leu Pro Asp Lys Asn Phe Arg Lys Met Leu Glu Ser Trp Ser Asp

1025 1030 1035 1040

Pro Leu Leu Asp Lys Trp Pro Asp Leu His Arg Lys Val Arg Leu Leu

1045 1050 1055

Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp

1060 1065 1070

Glu Ala Val Phe Ser Ser Ile Arg Lys Tyr Asp Pro Ser Phe Pro Asp

1075 1080 1085

Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu Ser Glu

1090 1095 1100

Glu Val Lys Gln Ala Lys Glu Thr Val Glu Arg Ile Ile Gln Ala

1105 1110 1115

<210> 83

<211> 1179

<212> PRT

<213> Flavobacterium columnare

<400> 83

Met Ser Ser Lys Asn Glu Ser Tyr Asn Lys Gln Lys Thr Phe Asn His

1 5 10 15

Tyr Lys Gln Glu Asp Lys Tyr Phe Phe Gly Gly Phe Leu Asn Asn Ala

20 25 30

Asp Asp Asn Leu Arg Gln Val Gly Lys Glu Phe Lys Thr Arg Ile Asn

35 40 45

Phe Asn His Asn Asn Asn Glu Leu Ala Ser Val Phe Lys Asp Tyr Phe

50 55 60

Asn Lys Glu Lys Ser Val Ala Lys Arg Glu His Ala Leu Asn Leu Leu

65 70 75 80

Ser Asn Tyr Phe Pro Val Leu Glu Arg Ile Gln Lys His Thr Asn His

85 90 95

Asn Phe Glu Gln Thr Arg Glu Ile Phe Glu Leu Leu Leu Asp Thr Ile

100 105 110

Lys Lys Leu Arg Asp Tyr Tyr Thr His His Tyr His Lys Pro Ile Thr

115 120 125

Ile Asn Pro Lys Ile Tyr Asp Phe Leu Asp Asp Thr Leu Leu Asp Val

130 135 140

Leu Ile Thr Ile Lys Lys Lys Lys Val Lys Asn Asp Thr Ser Arg Glu

145 150 155 160

Leu Leu Lys Glu Lys Leu Arg Pro Glu Leu Thr Gln Leu Lys Asn Gln

165 170 175

Lys Arg Glu Glu Leu Ile Lys Lys Gly Lys Lys Leu Leu Glu Glu Asn

180 185 190

Leu Glu Asn Ala Val Phe Asn His Cys Leu Arg Pro Phe Leu Glu Glu
195 200 205

Asn Lys Thr Asp Asp Lys Gln Asn Lys Thr Val Ser Leu Arg Lys Tyr
210 215 220
Arg Lys Ser Lys Pro Asn Glu Glu Thr Ser Ile Thr Leu Thr Gln Ser
225 230 235 240
Gly Leu Val Phe Leu Met Ser Phe Phe Leu His Arg Lys Glu Phe Gln
245 250 255
Val Phe Thr Ser Gly Leu Glu Gly Phe Lys Ala Lys Val Asn Thr Ile
260 265 270
Lys Glu Glu Glu Ile Ser Leu Asn Lys Asn Asn Ile Val Tyr Met Ile

275 280 285
Thr His Trp Ser Tyr Ser Tyr Tyr Asn Phe Lys Gly Leu Lys His Arg
290 295 300
Ile Lys Thr Asp Gln Gly Val Ser Thr Leu Glu Gln Asn Asn Thr Thr
305 310 315 320
His Ser Leu Thr Asn Thr Asn Thr Lys Glu Ala Leu Leu Thr Gln Ile
325 330 335
Val Asp Tyr Leu Ser Lys Val Pro Asn Glu Ile Tyr Glu Thr Leu Ser
340 345 350

Glu Lys Gln Gln Lys Glu Phe Glu Glu Asp Ile Asn Glu Tyr Met Arg
355 360 365
Glu Asn Pro Glu Asn Glu Asp Ser Thr Phe Ser Ser Ile Val Ser His
370 375 380
Lys Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Met
385 390 395 400
Arg Phe Leu Asp Glu Tyr Ala Glu Leu Pro Thr Leu Arg Phe Met Val
405 410 415
Asn Phe Gly Asp Tyr Ile Lys Asp Arg Gln Lys Lys Ile Leu Glu Ser

420 425 430
Ile Gln Phe Asp Ser Glu Arg Ile Ile Lys Lys Glu Ile His Leu Phe

435 440 445

Glu Lys Leu Ser Leu Val Thr Glu Tyr Lys Lys Asn Val Tyr Leu Lys

450 455 460

Glu Thr Ser Asn Ile Asp Leu Ser Arg Phe Pro Leu Phe Pro Asn Pro

465 470 475 480

Ser Tyr Val Met Ala Asn Asn Asn Ile Pro Phe Tyr Ile Asp Ser Arg

485 490 495

Ser Asn Asn Leu Asp Glu Tyr Leu Asn Gln Lys Lys Lys Ala Gln Ser

500 505 510

Gln Asn Lys Lys Arg Asn Leu Thr Phe Glu Lys Tyr Asn Lys Glu Gln

515 520 525

Ser Lys Asp Ala Ile Ile Ala Met Leu Gln Lys Glu Ile Gly Val Lys

530 535 540

Asp Leu Gln Gln Arg Ser Thr Ile Gly Leu Leu Ser Cys Asn Glu Leu

545 550 555 560

Pro Ser Met Leu Tyr Glu Val Ile Val Lys Asp Ile Lys Gly Ala Glu

565 570 575

Leu Glu Asn Lys Ile Ala Gln Lys Ile Arg Glu Gln Tyr Gln Ser Ile

580 585 590

Arg Asp Phe Thr Leu Asp Ser Pro Gln Lys Asp Asn Ile Pro Thr Thr

595 600 605

Leu Ile Lys Thr Ile Asn Thr Asp Ser Ser Val Thr Phe Glu Asn Gln

610 615 620

Pro Ile Asp Ile Pro Arg Leu Lys Asn Ala Ile Gln Lys Glu Leu Thr

625 630 635 640

Leu Thr Gln Glu Lys Leu Leu Asn Val Lys Glu His Glu Ile Glu Val

645 650 655

Asp Asn Tyr Asn Arg Asn Lys Asn Thr Tyr Lys Phe Lys Asn Gln Pro

660 665 670

Lys Asn Lys Val Asp Asp Lys Lys Leu Gln Arg Lys Tyr Val Phe Tyr

675 680 685

Arg Asn Glu Ile Arg Gln Glu Ala Asn Trp Leu Ala Ser Asp Leu Ile

690 695 700

His Phe Met Lys Asn Lys Ser Leu Trp Lys Gly Tyr Met His Asn Glu

705 710 715 720

Leu Gln Ser Phe Leu Ala Phe Phe Glu Asp Lys Lys Asn Asp Cys Ile

725 730 735

Ala Leu Leu Glu Thr Val Phe Asn Leu Lys Glu Asp Cys Ile Leu Thr

740 745 750

Lys Gly Leu Lys Asn Leu Phe Leu Lys His Gly Asn Phe Ile Asp Phe

755 760 765

Tyr Lys Glu Tyr Leu Lys Leu Lys Glu Asp Phe Leu Asn Thr Glu Ser

770 775 780

Thr Phe Leu Glu Asn Gly Leu Ile Gly Leu Pro Pro Lys Ile Leu Lys

785 790 795 800

Lys Glu Leu Ser Lys Arg Phe Lys Tyr Ile Phe Ile Val Phe Gln Lys

805 810 815

Arg Gln Phe Ile Ile Lys Glu Leu Glu Glu Lys Lys Asn Asn Leu Tyr

820 825 830

Ala Asp Ala Ile Asn Leu Ser Arg Gly Ile Phe Asp Glu Lys Pro Thr

835 840 845

Met Ile Pro Phe Lys Lys Pro Asn Pro Asp Glu Phe Ala Ser Trp Phe

850 855 860

Val Ala Ser Tyr Gln Tyr Asn Asn Tyr Gln Ser Phe Tyr Glu Leu Thr

865 870 875 880

Pro Asp Ile Val Glu Arg Asp Lys Lys Lys Lys Tyr Lys Asn Leu Arg

885 890 895

Ala Ile Asn Lys Val Lys Ile Gln Asp Tyr Tyr Leu Lys Leu Met Val

900 905 910

Asp Thr Leu Tyr Gln Asp Leu Phe Asn Gln Pro Leu Asp Lys Ser Leu

915 920 925

Ser Asp Phe Tyr Val Ser Lys Ala Glu Arg Glu Lys Ile Lys Ala Asp

930 935 940

Ala Lys Ala Tyr Gln Lys Arg Asn Asp Ser Ser Leu Trp Asn Lys Val
945 950 955 960
Ile His Leu Ser Leu Gln Asn Asn Arg Ile Thr Ala Asn Pro Lys Leu
965 970 975
Lys Asp Ile Gly Lys Tyr Lys Arg Ala Leu Gln Asp Glu Lys Ile Ala
980 985 990
Thr Leu Leu Thr Tyr Asp Asp Arg Thr Trp Thr Tyr Ala Leu Gln Lys

995 1000 1005
Pro Glu Lys Glu Asn Glu Asn Asp Tyr Lys Glu Leu His Tyr Thr Ala
1010 1015 1020
Leu Asn Met Glu Leu Gln Glu Tyr Glu Lys Val Arg Ser Lys Glu Leu
1025 1030 1035 1040
Leu Lys Gln Val Gln Glu Leu Glu Lys Gln Ile Leu Glu Glu Tyr Thr
1045 1050 1055
Asp Phe Leu Ser Thr Gln Ile His Pro Ala Asp Phe Glu Arg Glu Gly
1060 1065 1070

Asn Pro Asn Phe Lys Lys Tyr Leu Ala His Ser Ile Leu Glu Asn Glu
1075 1080 1085
Asp Asp Leu Asp Lys Leu Pro Glu Lys Val Glu Ala Met Arg Glu Leu
1090 1095 1100
Asp Glu Thr Ile Thr Asn Pro Ile Ile Lys Lys Ala Ile Val Leu Ile
1105 1110 1115 1120
Ile Ile Arg Asn Lys Met Ala His Asn Gln Tyr Pro Pro Lys Phe Ile
1125 1130 1135
Tyr Asp Leu Ala Asn Arg Phe Val Pro Lys Lys Glu Glu Glu Tyr Phe

1140 1145 1150
Ala Thr Tyr Phe Asn Arg Val Phe Glu Thr Ile Thr Lys Glu Leu Trp
1155 1160 1165
Glu Asn Lys Glu Lys Lys Asp Lys Thr Gln Val
1170 1175
<210> 84
<211> 1145

<212> PRT

<213> Psychroflexus torquis

<400> 84

Met Glu Ser Ile Ile Gly Leu Gly Leu Ser Phe Asn Pro Tyr Lys Thr
1 5 10 15

Ala Asp Lys His Tyr Phe Gly Ser Phe Leu Asn Leu Val Glu Asn Asn
20 25 30

Leu Asn Ala Val Phe Ala Glu Phe Lys Glu Arg Ile Ser Tyr Lys Ala
35 40 45

Lys Asp Glu Asn Ile Ser Ser Leu Ile Glu Lys His Phe Ile Asp Asn
50 55 60

Met Ser Ile Val Asp Tyr Glu Lys Lys Ile Ser Ile Leu Asn Gly Tyr
65 70 75 80

Leu Pro Ile Ile Asp Phe Leu Asp Asp Glu Leu Glu Asn Asn Leu Asn
85 90 95

Thr Arg Val Lys Asn Phe Lys Lys Asn Phe Ile Ile Leu Ala Glu Ala
100 105 110

Ile Glu Lys Leu Arg Asp Tyr Tyr Thr His Phe Tyr His Asp Pro Ile
115 120 125

Thr Phe Glu Asp Asn Lys Glu Pro Leu Leu Glu Leu Leu Asp Glu Val
130 135 140

Leu Leu Lys Thr Ile Leu Asp Val Lys Lys Lys Tyr Leu Lys Thr Asp
145 150 155 160

Lys Thr Lys Glu Ile Leu Lys Asp Ser Leu Arg Glu Glu Met Asp Leu
165 170 175

Leu Val Ile Arg Lys Thr Asp Glu Leu Arg Glu Lys Lys Lys Thr Asn
180 185 190

Pro Lys Ile Gln His Thr Asp Ser Ser Gln Ile Lys Asn Ser Ile Phe
195 200 205

Asn Asp Ala Phe Gln Gly Leu Leu Tyr Glu Asp Lys Gly Asn Asn Lys
210 215 220

Lys Thr Gln Val Ser His Arg Ala Lys Thr Arg Leu Asn Pro Lys Asp

225 230 235 240

Ile His Lys Gln Glu Glu Arg Asp Phe Glu Ile Pro Leu Ser Thr Ser
245 250 255
Gly Leu Val Phe Leu Met Ser Leu Phe Leu Ser Lys Lys Glu Ile Glu
260 265 270
Asp Phe Lys Ser Asn Ile Lys Gly Phe Lys Gly Lys Val Val Lys Asp
275 280 285
Glu Asn His Asn Ser Leu Lys Tyr Met Ala Thr His Arg Val Tyr Ser
290 295 300
Ile Leu Ala Phe Lys Gly Leu Lys Tyr Arg Ile Lys Thr Asp Thr Phe

305 310 315 320
Ser Lys Glu Thr Leu Met Met Gln Met Ile Asp Glu Leu Ser Lys Val
325 330 335
Pro Asp Cys Val Tyr Gln Asn Leu Ser Glu Thr Lys Gln Lys Asp Phe
340 345 350
Ile Glu Asp Trp Asn Glu Tyr Phe Lys Asp Asn Glu Glu Asn Thr Glu
355 360 365
Asn Leu Glu Asn Ser Arg Val Val His Pro Val Ile Arg Lys Arg Tyr
370 375 380

Glu Asp Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Ala
385 390 395 400
Asn Phe Lys Thr Leu Lys Phe Gln Val Phe Met Gly Tyr Tyr Ile His
405 410 415
Asp Gln Arg Thr Lys Thr Ile Gly Thr Thr Asn Ile Thr Thr Glu Arg
420 425 430
Thr Val Lys Glu Lys Ile Asn Val Phe Gly Lys Leu Ser Lys Met Asp
435 440 445
Asn Leu Lys Lys His Phe Phe Ser Gln Leu Ser Asp Asp Glu Asn Thr

450 455 460
Asp Trp Glu Phe Phe Pro Asn Pro Ser Tyr Asn Phe Leu Thr Gln Ala
465 470 475 480

Asp Asn Ser Pro Ala Asn Asn Ile Pro Ile Tyr Leu Glu Leu Lys Asn
485 490 495
Gln Gln Ile Ile Lys Glu Lys Asp Ala Ile Lys Ala Glu Val Asn Gln
500 505 510
Thr Gln Asn Arg Asn Pro Asn Lys Pro Ser Lys Arg Asp Leu Leu Asn
515 520 525

Lys Ile Leu Lys Thr Tyr Glu Asp Phe His Gln Gly Asp Pro Thr Ala
530 535 540
Ile Leu Ser Leu Asn Glu Ile Pro Ala Leu Leu His Leu Phe Leu Val
545 550 555 560
Lys Pro Asn Asn Lys Thr Gly Gln Gln Ile Glu Asn Ile Ile Arg Ile
565 570 575
Lys Ile Glu Lys Gln Phe Lys Ala Ile Asn His Pro Ser Lys Asn Asn
580 585 590
Lys Gly Ile Pro Lys Ser Leu Phe Ala Asp Thr Asn Val Arg Val Asn

595 600 605
Ala Ile Lys Leu Lys Lys Asp Leu Glu Ala Glu Leu Asp Met Leu Asn
610 615 620
Lys Lys His Ile Ala Phe Lys Glu Asn Gln Lys Ala Ser Ser Asn Tyr
625 630 635 640
Asp Lys Leu Leu Lys Glu His Gln Phe Thr Pro Lys Asn Lys Arg Pro
645 650 655
Glu Leu Arg Lys Tyr Val Phe Tyr Lys Ser Glu Lys Gly Glu Glu Ala
660 665 670

Thr Trp Leu Ala Asn Asp Ile Lys Arg Phe Met Pro Lys Asp Phe Lys
675 680 685
Thr Lys Trp Lys Gly Cys Gln His Ser Glu Leu Gln Arg Lys Leu Ala
690 695 700
Phe Tyr Asp Arg His Thr Lys Gln Asp Ile Lys Glu Leu Leu Ser Gly
705 710 715 720
Cys Glu Phe Asp His Ser Leu Leu Asp Ile Asn Ala Tyr Phe Gln Lys
725 730 735

Asp Asn Phe Glu Asp Phe Phe Ser Lys Tyr Leu Glu Asn Arg Ile Glu

740 745 750

Thr Leu Glu Gly Val Leu Lys Lys Leu His Asp Phe Lys Asn Glu Pro

755 760 765

Thr Pro Leu Lys Gly Val Phe Lys Asn Cys Phe Lys Phe Leu Lys Arg

770 775 780

Gln Asn Tyr Val Thr Glu Ser Pro Glu Ile Ile Lys Lys Arg Ile Leu

785 790 795 800

Ala Lys Pro Thr Phe Leu Pro Arg Gly Val Phe Asp Glu Arg Pro Thr

805 810 815

Met Lys Lys Gly Lys Asn Pro Leu Lys Asp Lys Asn Glu Phe Ala Glu

820 825 830

Trp Phe Val Glu Tyr Leu Glu Asn Lys Asp Tyr Gln Lys Phe Tyr Asn

835 840 845

Ala Glu Glu Tyr Arg Met Arg Asp Ala Asp Phe Lys Lys Asn Ala Val

850 855 860

Ile Lys Lys Gln Lys Leu Lys Asp Phe Tyr Thr Leu Gln Met Val Asn

865 870 875 880

Tyr Leu Leu Lys Glu Val Phe Gly Lys Asp Glu Met Asn Leu Gln Leu

885 890 895

Ser Glu Leu Phe Gln Thr Arg Gln Glu Arg Leu Lys Leu Gln Gly Ile

900 905 910

Ala Lys Lys Gln Met Asn Lys Glu Thr Gly Asp Ser Ser Glu Asn Thr

915 920 925

Arg Asn Gln Thr Tyr Ile Trp Asn Lys Asp Val Pro Val Ser Phe Phe

930 935 940

Asn Gly Lys Val Thr Ile Asp Lys Val Lys Leu Lys Asn Ile Gly Lys

945 950 955 960

Tyr Lys Arg Tyr Glu Arg Asp Glu Arg Val Lys Thr Phe Ile Gly Tyr

965 970 975

Glu Val Asp Glu Lys Trp Met Met Tyr Leu Pro His Asn Trp Lys Asp

980 985 990

Arg Tyr Ser Val Lys Pro Ile Asn Val Ile Asp Leu Gln Ile Gln Glu

995 1000 1005

Tyr Glu Glu Ile Arg Ser His Glu Leu Leu Lys Glu Ile Gln Asn Leu

1010 1015 1020

Glu Gln Tyr Ile Tyr Asp His Thr Thr Asp Lys Asn Ile Leu Leu Gln

1025 1030 1035 1040

Asp Gly Asn Pro Asn Phe Lys Met Tyr Val Leu Asn Gly Leu Leu Ile

1045 1050 1055

Gly Ile Lys Gln Val Asn Ile Pro Asp Phe Ile Val Leu Lys Gln Asn

1060 1065 1070

Thr Asn Phe Asp Lys Ile Asp Phe Thr Gly Ile Ala Ser Cys Ser Glu

1075 1080 1085

Leu Glu Lys Lys Thr Ile Ile Leu Ile Ala Ile Arg Asn Lys Phe Ala

1090 1095 1100

His Asn Gln Leu Pro Asn Lys Met Ile Tyr Asp Leu Ala Asn Glu Phe

1105 1110 1115 1120

Leu Lys Ile Glu Lys Asn Glu Thr Tyr Ala Asn Tyr Tyr Leu Lys Val

1125 1130 1135

Leu Lys Lys Met Ile Ser Asp Leu Ala

1140 1145

<210> 85

<211> 948

<212> PRT

<213> Riemerella anatipestifer

<400> 85

Met Phe Phe Ser Phe His Asn Ala Gln Arg Val Ile Phe Lys His Leu

1 5 10 15

Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu Asp Tyr Lys

20 25 30

Ala His Phe Thr Val Asn Leu Thr Arg Asp Phe Ala His Leu Asn Arg

35 40 45

Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asn Arg Tyr Arg Phe
50 55 60
Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe Phe Thr Asn
65 70 75 80
Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys Lys Val Ser

85 90 95
Gly Phe Lys Ala Ser His Lys Gln Arg Glu Lys Met Thr Thr Glu Val
100 105 110
Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu Glu Ser Arg
115 120 125
Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu Leu Ser Arg
130 135 140
Cys Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Glu Asn Lys Lys His
145 150 155 160

Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu Glu Glu Gln
165 170 175
Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg Phe Pro Tyr
180 185 190
Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys Ser Ile Arg
195 200 205
Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr Asp Lys Lys
210 215 220
Ile Gly Asp Glu Gln Glu Lys Arg His Leu Thr Arg Thr Leu Leu Ser

225 230 235 240
Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro Gln Glu Trp
245 250 255
Lys Ala Leu Thr Lys Asp Leu Asp Tyr Lys Glu Thr Ser Asn Gln Pro
260 265 270
Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp Asn Lys Ile
275 280 285
Gly Phe Arg Leu Gly Thr Ser Lys Glu Leu Tyr Pro Ser Leu Glu Ile

290 295 300

Lys Asp Gly Ala Asn Arg Ile Ala Lys Tyr Pro Tyr Asn Ser Gly Phe
305 310 315 320
Val Ala His Ala Phe Ile Ser Val His Glu Leu Leu Pro Leu Met Phe
325 330 335
Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys Glu Thr Val
340 345 350
Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu Arg Ile Asn
355 360 365
Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu Pro Leu Gly

370 375 380
Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys Gln Pro Asp
385 390 395 400
Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile Ala Glu Thr
405 410 415
Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser Ser Pro Lys
420 425 430
Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val Leu Ala Asp
435 440 445

Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Ala Tyr Asp Ala
450 455 460
Gln Asn Gln Pro Ile Lys Ser Ser Lys Ala Asn Ser Thr Glu Phe Trp
465 470 475 480
Phe Ile Arg Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys Asn Arg Leu
485 490 495
Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr Asn Pro His
500 505 510
Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn Leu Val Asp

515 520 525
Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu Glu Ala Ile
530 535 540

Lys His Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu Leu Lys Val
545 550 555 560
Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu Gln Gly Gly
565 570 575
Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg Glu Thr Leu
580 585 590

Ser Lys Asp Leu Thr Leu Ser Lys Pro Ile Arg Lys Glu Ile Lys Lys
595 600 605
His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu Tyr Phe Lys
610 615 620
Glu Lys Tyr Gln Asp Lys His Gln Ser Phe Tyr Asn Leu Ser Tyr Lys
625 630 635 640
Leu Glu Ala Lys Ala Pro Leu Leu Lys Lys Glu Glu His Tyr Glu Tyr
645 650 655
Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Ser Gln Arg Leu Glu

660 665 670
Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr Lys Arg Trp
675 680 685
Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln Asp Ile Met
690 695 700
Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe Lys Glu Leu
705 710 715 720
Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala Val Asn Val
725 730 735

Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr Leu Pro Met
740 745 750
Val Leu Pro Val Lys Val Tyr Pro Thr Thr Ala Phe Gly Glu Val Gln
755 760 765
Tyr His Glu Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu Glu Gln Thr
770 775 780
Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys Asp Arg Arg
785 790 795 800

Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp Thr Gln Lys

805 810 815

His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu Ile Tyr Gln

820 825 830

Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Ser Leu Glu Glu Lys

835 840 845

Leu Leu Asn Lys His Ala Ser Leu Ser Ser Leu Glu Asn Glu Phe Arg

850 855 860

Thr Leu Leu Glu Glu Trp Lys Lys Lys Tyr Ala Ala Ser Ser Met Val

865 870 875 880

Thr Asp Lys His Ile Ala Phe Ile Ala Ser Val Arg Asn Ala Phe Cys

885 890 895

His Asn Gln Tyr Pro Phe Tyr Lys Glu Thr Leu His Ala Pro Ile Leu

900 905 910

Leu Phe Thr Val Ala Gln Pro Thr Thr Glu Glu Lys Asp Gly Leu Gly

915 920 925

Ile Ala Glu Ala Leu Leu Lys Val Leu Arg Glu Tyr Cys Glu Ile Val

930 935 940

Lys Ser Gln Ile

945

<210> 86

<211> 1139

<212> PRT

<213> Prevotella pleuritidis

<400> 86

Met Glu Asn Asp Lys Arg Leu Glu Glu Ser Ala Cys Tyr Thr Leu Asn

1 5 10 15

Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val

20 25 30

Tyr Ile Thr Val Asn His Ile Asn Lys Thr Leu Glu Leu Lys Asn Lys

35 40 45

Lys Asn Gln Glu Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala Ile

50 55 60

Lys Thr His Trp Ala Lys Val Asn Gly Asp Leu Asn Lys Thr Asp Arg
65 70 75 80
Leu Arg Glu Leu Met Ile Lys His Phe Pro Phe Leu Glu Ala Ala Ile
85 90 95
Tyr Ser Asn Asn Lys Glu Asp Lys Glu Glu Val Lys Glu Glu Lys Gln
100 105 110
Ala Lys Ala Gln Ser Phe Lys Ser Leu Lys Asp Cys Leu Phe Leu Phe
115 120 125

Leu Glu Lys Leu Gln Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys Tyr
130 135 140
Ser Glu Ser Ser Lys Glu Pro Glu Phe Glu Glu Gly Leu Leu Glu Lys
145 150 155 160
Met Tyr Asn Thr Phe Asp Ala Ser Ile Arg Leu Val Lys Glu Asp Tyr
165 170 175
Gln Tyr Asn Lys Asp Ile Asp Pro Glu Lys Asp Phe Lys His Leu Glu
180 185 190
Arg Lys Glu Asp Phe Asn Tyr Leu Phe Thr Asp Lys Asp Asn Lys Gly

195 200 205
Lys Ile Thr Lys Asn Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu
210 215 220
Lys Lys Asp Ala Ile Trp Met Gln Gln Lys Phe Arg Gly Phe Lys Asp
225 230 235 240
Asn Arg Gly Asn Lys Glu Lys Met Thr His Glu Val Phe Cys Arg Ser
245 250 255
Arg Met Leu Leu Pro Lys Ile Arg Leu Glu Ser Thr Gln Thr Gln Asp
260 265 270

Trp Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser
275 280 285
Leu Tyr Glu Arg Leu Gln Gly Ala Tyr Arg Glu Lys Phe Lys Val Pro

290 295 300

Phe Asp Ser Ile Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe Arg

305 310 315 320

Asn Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu Arg

325 330 335

Tyr Phe Asp Tyr Asn Glu Ile Phe Lys Asn Leu Arg Phe Gln Ile Asp

340 345 350

Leu Gly Thr Tyr His Phe Ser Ile Tyr Lys Lys Leu Ile Gly Gly Lys

355 360 365

Lys Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg Ile

370 375 380

Gln Glu Phe Thr Lys Gln Asn Arg Pro Asp Lys Trp Gln Ala Ile Ile

385 390 395 400

Lys Asp Leu Asp Thr Tyr Glu Thr Ser Asn Glu Arg Tyr Ile Ser Glu

405 410 415

Thr Thr Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg Phe

420 425 430

Arg Asn Asp Asn Asn Asp Ile Trp Pro Ser Leu Lys Thr Asn Gly Glu

435 440 445

Lys Asn Glu Lys Ser Lys Tyr Asn Leu Asp Lys Pro Tyr Gln Ala Glu

450 455 460

Ala Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu

465 470 475 480

Leu Leu Lys Met Glu Asn Thr Asp Asn Asp Lys Glu Asp Asn Glu Val

485 490 495

Gly Thr Lys Lys Lys Gly Asn Lys Asn Asn Lys Gln Glu Lys His Lys

500 505 510

Ile Glu Glu Ile Ile Glu Asn Lys Ile Lys Asp Ile Tyr Ala Leu Tyr

515 520 525

Asp Ala Phe Thr Asn Gly Glu Ile Asn Ser Ile Asp Glu Leu Ala Glu

530 535 540

Gln Arg Glu Gly Lys Asp Ile Glu Ile Gly His Leu Pro Lys Gln Leu

545 550 555 560

Ile Val Ile Leu Lys Asn Lys Ser Lys Asp Met Ala Glu Lys Ala Asn
565 570 575
Arg Lys Gln Lys Glu Met Ile Lys Asp Thr Lys Lys Arg Leu Ala Thr
580 585 590
Leu Asp Lys Gln Val Lys Gly Glu Ile Glu Asp Gly Gly Arg Asn Ile
595 600 605
Arg Leu Leu Lys Ser Gly Glu Ile Ala Arg Trp Leu Val Asn Asp Met
610 615 620
Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu Gly Lys Pro Leu Asn

625 630 635 640
Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met Leu Gln Arg Ser Leu
645 650 655
Ala Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg Tyr Phe Arg Gln Val
660 665 670
Asn Leu Ile Lys Ser Ser Asn Pro His Pro Phe Leu Glu Asp Thr Lys
675 680 685
Trp Glu Glu Cys Tyr Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Lys
690 695 700

Ala Lys Ile Lys Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Lys Lys
705 710 715 720
Asn Gln Tyr Phe Leu Met Leu Lys Glu Pro Lys Thr Asn Arg Lys Thr
725 730 735
Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile Phe
740 745 750
Thr Glu Pro Ile Lys Glu Trp Phe Lys Arg His Gln Asn Asp Ser Glu
755 760 765
Glu Tyr Lys Lys Val Glu Ala Leu Asp Arg Val Gly Leu Val Ala Lys

770 775 780
Val Ile Pro Leu Phe Phe Lys Glu Glu Tyr Phe Lys Glu Asp Ala Gln
785 790 795 800

Lys Glu Ile Asn Asn Cys Val Gln Pro Phe Tyr Ser Phe Pro Tyr Asn
805 810 815
Val Gly Asn Ile His Lys Pro Glu Glu Lys Asn Phe Leu His Cys Glu
820 825 830
Glu Arg Arg Lys Leu Trp Asp Lys Lys Lys Asp Lys Phe Lys Gly Tyr
835 840 845

Lys Ala Lys Glu Lys Ser Lys Lys Met Thr Asp Lys Glu Lys Glu Glu
850 855 860
His Arg Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu
865 870 875 880
Leu Arg Leu Val Arg Asn Gln Asp Ile Leu Thr Trp Leu Leu Cys Thr
885 890 895
Lys Leu Ile Asp Lys Leu Lys Ile Asp Glu Leu Asn Ile Glu Glu Leu
900 905 910
Gln Lys Leu Arg Leu Lys Asp Ile Asp Thr Asp Thr Ala Lys Lys Glu

915 920 925
Lys Asn Asn Ile Leu Asn Arg Val Met Pro Met Arg Leu Pro Val Thr
930 935 940
Val Tyr Glu Ile Asp Lys Ser Phe Asn Ile Val Lys Asp Lys Pro Leu
945 950 955 960
His Thr Val Tyr Ile Glu Glu Thr Gly Thr Lys Leu Leu Lys Gln Gly
965 970 975
Asn Phe Lys Ala Leu Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser
980 985 990

Phe Val Lys Thr Ser Ser Glu Ala Glu Ser Lys Ser Lys Pro Ile Ser
995 1000 1005
Lys Leu Arg Val Glu Tyr Glu Leu Gly Ala Tyr Gln Lys Ala Arg Ile
1010 1015 1020
Asp Ile Ile Lys Asp Met Leu Ala Leu Glu Lys Thr Leu Ile Asp Asn
1025 1030 1035 1040
Asp Glu Asn Leu Pro Thr Asn Lys Phe Ser Asp Met Leu Lys Ser Trp
1045 1050 1055

Leu Lys Gly Lys Gly Glu Ala Asn Lys Ala Arg Leu Gln Asn Asp Val

1060 1065 1070

Gly Leu Leu Val Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro

1075 1080 1085

Met Tyr Asn Ser Glu Val Phe Lys Gly Met Lys Leu Leu Ser Leu Ser

1090 1095 1100

Ser Asp Ile Pro Glu Lys Glu Gly Leu Gly Ile Ala Lys Gln Leu Lys

1105 1110 1115 1120

Asp Lys Ile Lys Glu Thr Ile Glu Arg Ile Ile Glu Ile Glu Lys Glu

1125 1130 1135

Ile Arg Asn

<210> 87

<211> 1135

<212> PRT

<213> Porphyromonas gingivalis

<400> 87

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val

1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu

20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys

35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu

50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr

65 70 75 80

Leu Pro Phe Leu His Tyr Phe Asp Pro Asp Ser Gln Ile Glu Lys Asp

85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile

100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His

115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
130 135 140
Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145 150 155 160
Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
165 170 175
Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
180 185 190
Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg

195 200 205
Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
210 215 220
Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225 230 235 240
Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala
245 250 255
Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
275 280 285
Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
290 295 300
Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305 310 315 320
Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
325 330 335
Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu

340 345 350
Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
355 360 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
370 375 380
Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385 390 395 400
Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
420 425 430
Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met
435 440 445
Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
450 455 460
Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465 470 475 480
Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
485 490 495
Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
500 505 510
Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
515 520 525
Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
530 535 540
Ser Ala Phe Asp Met Asp Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
565 570 575
Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Gln Lys Phe
580 585 590
Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
595 600 605
Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
610 615 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg

625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg His Gln Phe Arg Ala

645 650 655

Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe

660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr

675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr

690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe

705 710 715 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg

725 730 735

Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln

740 745 750

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met

755 760 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr

785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile

805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu

820 825 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro

835 840 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala

850 855 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg

865 870 875 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp

885 890 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn

900 905 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys

930 935 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg

945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu

965 970 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys

980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg

995 1000 1005

Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser

1010 1015 1020

Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys

1025 1030 1035 1040

Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg Asn Lys Ala

1045 1050 1055

Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg

1060 1065 1070

Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly Ser Ser Ala Lys Asp

1075 1080 1085

Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu

1090 1095 1100

Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp Pro Glu Asn Arg Phe

1105 1110 1115 1120

Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu

1125 1130 1135

<210> 88

<211> 1135

<212> PRT

<213> Porphyromonas gingivalis

<400> 88

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65 70 75 80

Leu Pro Phe Leu His Tyr Phe Asp Pro Asp Ser Gln Ile Glu Lys Asp
85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145 150 155 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
165 170 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
180 185 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg

195 200 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr

210 215 220

Asn Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys

225 230 235 240

Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala

245 250 255

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu

260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu

275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu

290 295 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr

305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe

325 330 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu

340 345 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly

355 360 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu

370 375 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu

385 390 395 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile

405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro

420 425 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met

435 440 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu

450 455 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Gly Phe Leu Arg Lys

465 470 475 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser

485 490 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro

500 505 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr

515 520 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu

530 535 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp

545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr

565 570 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe

580 585 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu

595 600 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu

610 615 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg

625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala

645 650 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe

660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr

675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr

690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705 710 715 720
Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
725 730 735
Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln
740 745 750
Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
755 760 765
Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780
Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785 790 795 800
Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
805 810 815
Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
820 825 830
Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
835 840 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
850 855 860
Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865 870 875 880
Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
885 890 895
Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Lys Glu Asn
900 905 910
Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925
Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
930 935 940
Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
965 970 975
Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg
995 1000 1005
Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser
1010 1015 1020
Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys
1025 1030 1035 1040
Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg Asn Lys Ala
1045 1050 1055
Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg

1060 1065 1070
Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly Ser Ser Ala Lys Asp
1075 1080 1085
Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu
1090 1095 1100
Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp His Glu Asn Arg Phe
1105 1110 1115 1120
Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu
1125 1130 1135

<210> 89
<211> 1135
<212> PRT
<213> Porphyromonas gingivalis
<400> 89
Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1 5 10 15
Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
20 25 30
Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys

35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu

50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr

65 70 75 80

Leu Pro Phe Leu His Tyr Phe Asp Pro Asp Ser Gln Ile Glu Lys Asp

85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile

100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His

115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp

130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala

145 150 155 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala

165 170 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys

180 185 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg

195 200 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr

210 215 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys

225 230 235 240

Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala

245 250 255

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu

260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu

275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
290 295 300
Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305 310 315 320
Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
325 330 335
Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu

340 345 350
Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
355 360 365
Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
370 375 380
Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385 390 395 400
Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
420 425 430
Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
435 440 445
Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
450 455 460
Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Gly Phe Leu Arg Lys
465 470 475 480
Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser

485 490 495
Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
500 505 510
Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
515 520 525
Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
530 535 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
565 570 575
Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
580 585 590
Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
595 600 605
Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
610 615 620
Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg

625 630 635 640
Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
645 650 655
Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
660 665 670
Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
675 680 685
Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705 710 715 720
Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
725 730 735
Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln
740 745 750
Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
755 760 765
Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780
Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr

785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile

805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu

820 825 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro

835 840 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala

850 855 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg

865 870 875 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp

885 890 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn

900 905 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys

930 935 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg

945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu

965 970 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys

980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg

995 1000 1005

Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser

1010 1015 1020

Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys

1025 1030 1035 1040

Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg Asn Lys Ala

1045 1050 1055

Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg

1060 1065 1070

Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly Ser Ser Ala Lys Asp

1075 1080 1085

Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu

1090 1095 1100

Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp His Glu Asn Arg Phe

1105 1110 1115 1120

Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu

1125 1130 1135

<210> 90

<211> 1135

<212> PRT

<213> Porphyromonas gingivalis

<400> 90

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val

1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu

20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys

35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu

50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr

65 70 75 80

Leu Pro Phe Leu His Tyr Phe Asp Pro Asp Ser Gln Ile Glu Lys Asp

85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile

100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His

115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
130 135 140
Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145 150 155 160
Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
165 170 175
Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
180 185 190
Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg

195 200 205
Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
210 215 220
Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225 230 235 240
Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala
245 250 255
Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
275 280 285
Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
290 295 300
Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305 310 315 320
Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
325 330 335
Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu

340 345 350
Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
355 360 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
370 375 380
Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385 390 395 400
Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
420 425 430
Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met
435 440 445
Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
450 455 460
Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465 470 475 480
Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
485 490 495
Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
500 505 510
Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
515 520 525
Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
530 535 540
Ser Ala Phe Asp Met Asp Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
565 570 575
Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Gln Lys Phe
580 585 590
Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
595 600 605
Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
610 615 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg

625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg His Gln Phe Arg Ala

645 650 655

Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe

660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr

675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr

690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe

705 710 715 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg

725 730 735

Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln

740 745 750

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Val Met

755 760 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr

785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile

805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu

820 825 830

Lys Thr Val Arg Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro

835 840 845

Val Pro Pro Asp Leu Ala Ala Tyr Ile Lys Arg Ser Phe His Arg Ala

850 855 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg

865 870 875 880

Leu Met Leu Met Ala Ile Asn Lys Ile Met Thr Asp Arg Glu Glu Asp

885 890 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Lys Glu Asn

900 905 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys

930 935 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg

945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu

965 970 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys

980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg

995 1000 1005

Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser

1010 1015 1020

Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys

1025 1030 1035 1040

Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg Asn Lys Ala

1045 1050 1055

Ala His Asn Gln Phe Pro Cys Ala Ala Glu Ile Pro Leu Ile Tyr Arg

1060 1065 1070

Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly Ser Ser Ala Lys Asp

1075 1080 1085

Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu

1090 1095 1100

Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp Pro Glu Asn Arg Phe

1105 1110 1115 1120

Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu

1125 1130 1135

<210> 91

<211> 1135

<212> PRT

<213> Porphyromonas gingivalis

<400> 91

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65 70 75 80

Leu Pro Phe Leu His Tyr Phe Asp Pro Asp Ser Gln Ile Glu Lys Asp
85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145 150 155 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
165 170 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
180 185 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg

195 200 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr

210 215 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys

225 230 235 240

Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala

245 250 255

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu

260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu

275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu

290 295 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr

305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe

325 330 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu

340 345 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly

355 360 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu

370 375 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu

385 390 395 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile

405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro

420 425 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met

435 440 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu

450 455 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys

465 470 475 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser

485 490 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro

500 505 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr

515 520 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu

530 535 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp

545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr

565 570 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe

580 585 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu

595 600 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu

610 615 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg

625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala

645 650 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe

660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr

675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr

690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705 710 715 720
Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
725 730 735
Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln
740 745 750
Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
755 760 765
Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780
Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785 790 795 800
Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
805 810 815
Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
820 825 830
Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
835 840 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
850 855 860
Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865 870 875 880
Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
885 890 895
Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
900 905 910
Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925
Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
930 935 940
Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
965 970 975
Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg
995 1000 1005
Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser
1010 1015 1020
Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys
1025 1030 1035 1040
Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg Asn Lys Ala
1045 1050 1055
Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg

1060 1065 1070
Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly Ser Ser Ala Lys Asp
1075 1080 1085
Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu
1090 1095 1100
Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp Pro Glu Asn Arg Phe
1105 1110 1115 1120
Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu
1125 1130 1135

<210> 92
<211> 1133
<212> PRT
<213> Prevotella falsenii
<400> 92
Met Lys Asn Asp Asn Asn Ser Thr Lys Ser Thr Asp Tyr Thr Leu Gly
1 5 10 15
Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
20 25 30
Tyr Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn Lys

35 40 45

Lys Asp Gln Glu Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala Ile

50 55 60

Lys Thr Leu Trp Gly Lys Val Asp Thr Asp Ile Asn Lys Lys Asp Arg
65 70 75 80
Leu Arg Glu Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala Thr
85 90 95
Tyr Gln Gln Ser Ser Thr Asn Asn Thr Lys Gln Lys Glu Glu Glu Gln
100 105 110
Ala Lys Ala Gln Ser Phe Glu Ser Leu Lys Asp Cys Leu Phe Leu Phe
115 120 125

Leu Glu Lys Leu Arg Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys His
130 135 140
Ser Lys Ser Leu Glu Glu Pro Lys Leu Glu Glu Lys Leu Leu Glu Asn
145 150 155 160
Met Tyr Asn Ile Phe Asp Thr Asn Val Gln Leu Val Ile Lys Asp Tyr
165 170 175
Glu His Asn Lys Asp Ile Asn Pro Glu Glu Asp Phe Lys His Leu Gly
180 185 190
Arg Ala Glu Gly Glu Phe Asn Tyr Tyr Phe Thr Arg Asn Lys Lys Gly

195 200 205

Asn Ile Thr Glu Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu
210 215 220
Lys Lys Asp Ala Ile Trp Ala Gln Thr Lys Ile Lys Gly Phe Lys Asp
225 230 235 240
Asn Arg Glu Asn Lys Gln Lys Met Thr His Glu Val Phe Cys Arg Ser
245 250 255
Arg Met Leu Leu Pro Lys Leu Arg Leu Glu Ser Thr Gln Thr Gln Asp
260 265 270

Trp Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser
275 280 285

Leu Tyr Lys Arg Leu Gln Gly Glu Lys Arg Glu Lys Phe Arg Val Pro
290 295 300
Phe Asp Pro Ala Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe Lys
305 310 315 320
Asn Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu Arg
325 330 335
Tyr Phe Asp Tyr Asn Glu Ile Phe Thr Asn Leu Arg Phe Gln Ile Asp

340 345 350
Leu Gly Thr Tyr His Phe Ser Ile Tyr Lys Lys Gln Ile Gly Asp Lys
355 360 365
Lys Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg Ile
370 375 380
Gln Glu Phe Ala Lys Glu Asn Arg Pro Asp Glu Trp Lys Ala Leu Val
385 390 395 400
Lys Asp Leu Asp Thr Phe Glu Glu Ser Asn Glu Pro Tyr Ile Ser Glu
405 410 415

Thr Thr Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg Asn
420 425 430
Lys Asn Lys Lys Lys Lys Lys Thr Ile Trp Pro Ser Leu Glu Thr Lys
435 440 445
Thr Thr Val Asn Glu Arg Ser Lys Tyr Asn Leu Gly Lys Ser Phe Lys
450 455 460
Ala Glu Ala Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr
465 470 475 480
Tyr Leu Leu Leu Asn Lys Glu Glu Pro Asn Asn Gly Lys Ile Asn Ala

485 490 495
Ser Lys Val Glu Gly Ile Ile Glu Lys Lys Ile Arg Asp Ile Tyr Lys
500 505 510
Leu Tyr Gly Ala Phe Ala Asn Glu Glu Ile Asn Asn Glu Glu Glu Leu
515 520 525
Lys Glu Tyr Cys Glu Gly Lys Asp Ile Ala Ile Arg His Leu Pro Lys
530 535 540

Gln Met Ile Ala Ile Leu Lys Asn Glu Tyr Lys Asp Met Ala Lys Lys
545 550 555 560

Ala Glu Asp Lys Gln Lys Lys Met Ile Lys Asp Thr Lys Lys Arg Leu
565 570 575
Ala Ala Leu Asp Lys Gln Val Lys Gly Glu Val Glu Asp Gly Gly Arg
580 585 590
Asn Ile Lys Pro Leu Lys Ser Gly Arg Ile Ala Ser Trp Leu Val Asn
595 600 605
Asp Met Met Arg Phe Gln Pro Val Gln Arg Asp Arg Asp Gly Tyr Pro
610 615 620
Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Leu Leu Gln Arg

625 630 635 640
Thr Leu Ala Leu Phe Gly Ser Glu Arg Glu Arg Leu Ala Pro Tyr Phe
645 650 655
Arg Gln Met Asn Leu Ile Gly Lys Asp Asn Pro His Pro Phe Leu Lys
660 665 670
Asp Thr Lys Trp Lys Glu His Asn Asn Ile Leu Ser Phe Tyr Arg Ser
675 680 685
Tyr Leu Glu Ala Lys Lys Asn Phe Leu Gly Ser Leu Lys Pro Glu Asp
690 695 700

Trp Lys Lys Asn Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn
705 710 715 720
Arg Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg
725 730 735
Gly Ile Phe Thr Glu Pro Ile Arg Glu Trp Phe Ile Arg His Gln Asn
740 745 750
Glu Ser Glu Glu Tyr Lys Lys Val Lys Asp Phe Asp Arg Ile Gly Leu
755 760 765
Val Ala Lys Val Ile Pro Leu Phe Phe Lys Glu Asp Tyr Gln Lys Glu

770 775 780
Ile Glu Asp Tyr Val Gln Pro Phe Tyr Gly Tyr Pro Phe Asn Val Gly

785 790 795 800

Asn Ile His Asn Ser Gln Glu Gly Thr Phe Leu Asn Lys Lys Glu Arg

805 810 815

Glu Glu Leu Trp Lys Gly Asn Lys Thr Lys Phe Lys Asp Tyr Lys Thr

820 825 830

Lys Glu Lys Asn Lys Glu Lys Thr Asn Lys Asp Lys Phe Lys Lys Lys

835 840 845

Thr Asp Glu Glu Lys Glu Glu Phe Arg Ser Tyr Leu Asp Phe Gln Ser

850 855 860

Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln Asp Ile

865 870 875 880

Val Thr Trp Leu Leu Cys Met Glu Leu Ile Asp Lys Leu Lys Ile Asp

885 890 895

Glu Leu Asn Ile Glu Glu Leu Gln Lys Leu Arg Leu Lys Asp Ile Asp

900 905 910

Thr Asp Thr Ala Lys Lys Glu Lys Asn Asn Ile Leu Asn Arg Ile Met

915 920 925

Pro Met Glu Leu Pro Val Thr Val Tyr Glu Thr Asp Asp Ser Asn Asn

930 935 940

Ile Ile Lys Asp Lys Pro Leu His Thr Ile Tyr Ile Lys Glu Ala Glu

945 950 955 960

Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp Arg

965 970 975

Arg Leu Asn Gly Leu Phe Ser Phe Val Glu Thr Ser Ser Glu Ala Glu

980 985 990

Leu Lys Ser Lys Pro Ile Ser Lys Ser Leu Val Glu Tyr Glu Leu Gly

995 1000 1005

Glu Tyr Gln Arg Ala Arg Val Glu Ile Ile Lys Asp Met Leu Arg Leu

1010 1015 1020

Glu Glu Thr Leu Ile Gly Asn Asp Glu Lys Leu Pro Thr Asn Lys Phe

1025 1030 1035 1040

Arg Gln Met Leu Asp Lys Trp Leu Glu His Lys Lys Glu Thr Asp Asp

1045 1050 1055

Thr Asp Leu Lys Asn Asp Val Lys Leu Leu Thr Glu Val Arg Asn Ala

1060 1065 1070

Phe Ser His Asn Gln Tyr Pro Met Arg Asp Arg Ile Ala Phe Ala Asn

1075 1080 1085

Ile Lys Pro Phe Ser Leu Ser Ser Ala Asn Thr Ser Asn Glu Glu Gly

1090 1095 1100

Leu Gly Ile Ala Lys Lys Leu Lys Asp Lys Thr Lys Glu Thr Ile Asp

1105 1110 1115 1120

Arg Ile Ile Glu Ile Glu Glu Gln Thr Ala Thr Lys Arg

1125 1130

<210> 93

<211> 1116

<212> PRT

<213> Prevotella pleuritidis

<400> 93

Met Glu Asn Asp Lys Arg Leu Glu Glu Ser Thr Cys Tyr Thr Leu Asn

1 5 10 15

Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val

20 25 30

Tyr Ile Thr Ile Asn His Ile Asn Lys Leu Leu Glu Ile Arg Gln Ile

35 40 45

Asp Asn Asp Glu Lys Val Leu Asp Ile Lys Ala Leu Trp Gln Lys Val

50 55 60

Asp Lys Asp Ile Asn Gln Lys Ala Arg Leu Arg Glu Leu Met Ile Lys

65 70 75 80

His Phe Pro Phe Leu Glu Ala Ala Ile Tyr Ser Asn Asn Lys Glu Asp

85 90 95

Lys Glu Glu Val Lys Glu Glu Lys Gln Ala Lys Ala Gln Ser Phe Lys

100 105 110

Ser Leu Lys Asp Cys Leu Phe Leu Phe Leu Glu Lys Leu Gln Glu Ala

115 120 125

Arg Asn Tyr Tyr Ser His Tyr Lys Ser Ser Glu Ser Ser Lys Glu Pro
130 135 140
Glu Phe Glu Glu Gly Leu Leu Glu Lys Met Tyr Asn Thr Phe Gly Val
145 150 155 160
Ser Ile Arg Leu Val Lys Glu Asp Tyr Gln Tyr Asn Lys Asp Ile Asp
165 170 175
Pro Glu Lys Asp Phe Lys His Leu Glu Arg Lys Glu Asp Phe Asn Tyr
180 185 190
Leu Phe Thr Asp Lys Asp Asn Lys Gly Lys Ile Thr Lys Asn Gly Leu

195 200 205
Leu Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Met
210 215 220
Gln Gln Lys Leu Arg Gly Phe Lys Asp Asn Arg Gly Asn Lys Glu Lys
225 230 235 240
Met Thr His Glu Val Phe Cys Arg Ser Arg Met Leu Leu Pro Lys Ile
245 250 255
Arg Leu Glu Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu
260 265 270

Asn Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Gln Gly
275 280 285
Ala Tyr Arg Glu Lys Phe Lys Val Pro Phe Asp Ser Ile Asp Glu Asp
290 295 300
Tyr Asp Ala Glu Gln Glu Pro Phe Arg Asn Thr Leu Val Arg His Gln
305 310 315 320
Asp Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Ile
325 330 335
Phe Lys Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser

340 345 350
Ile Tyr Lys Lys Leu Ile Gly Asp Asn Lys Glu Asp Arg His Leu Thr
355 360 365

His Lys Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Ala Lys Gln Lys
370 375 380
Arg Pro Asn Glu Trp Gln Ala Leu Val Lys Asp Leu Asp Ile Tyr Glu
385 390 395 400
Thr Ser Asn Glu Gln Tyr Ile Ser Glu Thr Thr Pro His Tyr His Leu
405 410 415

Glu Asn Gln Lys Ile Gly Ile Arg Phe Lys Asn Lys Lys Asp Lys Ile
420 425 430
Trp Pro Ser Leu Glu Thr Asn Gly Lys Glu Asn Glu Lys Ser Lys Tyr
435 440 445
Asn Leu Asp Lys Ser Phe Gln Ala Glu Ala Phe Leu Ser Ile His Glu
450 455 460
Leu Leu Pro Met Met Phe Tyr Asp Leu Leu Leu Lys Lys Glu Glu Pro
465 470 475 480
Asn Asn Asp Glu Lys Asn Ala Ser Ile Val Glu Gly Phe Ile Lys Lys
485 490 495
Glu Ile Lys Arg Met Tyr Ala Ile Tyr Asp Ala Phe Ala Asn Glu Glu
500 505 510
Ile Asn Ser Lys Glu Gly Leu Glu Glu Tyr Cys Lys Asn Lys Gly Phe
515 520 525
Gln Glu Arg His Leu Pro Lys Gln Met Ile Ala Ile Leu Thr Asn Lys
530 535 540
Ser Lys Asn Met Ala Glu Lys Ala Lys Arg Lys Gln Lys Glu Met Ile
545 550 555 560

Lys Asp Thr Lys Lys Arg Leu Ala Thr Leu Asp Lys Gln Val Lys Gly
565 570 575
Glu Ile Glu Asp Gly Gly Arg Asn Ile Arg Leu Leu Lys Ser Gly Glu
580 585 590
Ile Ala Arg Trp Leu Val Asn Asp Met Met Arg Phe Gln Ser Val Gln
595 600 605
Lys Asp Lys Glu Gly Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr
610 615 620

Glu Tyr Gln Met Leu Gln Arg Ser Leu Ala Leu Tyr Asn Lys Glu Gln

625 630 635 640

Lys Pro Thr Pro Tyr Phe Ile Gln Val Asn Leu Ile Lys Ser Ser Asn

645 650 655

Pro His Pro Phe Leu Glu Glu Thr Lys Trp Glu Glu Cys Asn Asn Ile

660 665 670

Leu Ser Phe Tyr Arg Ser Tyr Leu Glu Ala Lys Lys Asn Phe Leu Glu

675 680 685

Ser Leu Lys Pro Glu Asp Trp Lys Lys Asn Gln Tyr Phe Leu Met Leu

690 695 700

Lys Glu Pro Lys Thr Asn Arg Lys Thr Leu Val Gln Gly Trp Lys Asn

705 710 715 720

Gly Phe Asn Leu Pro Arg Gly Ile Phe Thr Glu Pro Ile Lys Glu Trp

725 730 735

Phe Lys Arg His Gln Asn Asp Ser Glu Glu Tyr Lys Lys Val Glu Ala

740 745 750

Leu Asp Arg Val Gly Leu Val Ala Lys Val Ile Pro Leu Phe Phe Lys

755 760 765

Glu Glu Tyr Phe Lys Glu Asp Ala Gln Lys Glu Ile Asn Asn Cys Val

770 775 780

Gln Pro Phe Tyr Ser Phe Pro Tyr Asn Val Gly Asn Ile His Lys Pro

785 790 795 800

Glu Glu Lys Asn Phe Leu His Cys Glu Glu Arg Arg Lys Leu Trp Asp

805 810 815

Lys Lys Lys Asp Lys Phe Lys Gly Tyr Lys Ala Lys Glu Lys Ser Lys

820 825 830

Lys Met Thr Asp Lys Glu Lys Glu Glu His Arg Ser Tyr Leu Glu Phe

835 840 845

Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln

850 855 860

Asp Ile Val Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys

865 870 875 880

Ile Asp Glu Leu Asn Ile Glu Glu Leu Gln Lys Leu Arg Leu Lys Asp

885 890 895

Ile Asp Thr Asp Thr Ala Lys Lys Glu Lys Asn Asn Ile Leu Asn Arg

900 905 910

Ile Met Pro Met Gln Leu Pro Val Thr Val Tyr Glu Ile Asp Lys Ser

915 920 925

Phe Asn Ile Val Lys Asp Lys Pro Leu His Thr Ile Tyr Ile Glu Glu

930 935 940

Thr Gly Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys

945 950 955 960

Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Val Lys Thr Ser Ser Glu

965 970 975

Ala Glu Ser Lys Ser Lys Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu

980 985 990

Leu Gly Ala Tyr Gln Lys Ala Arg Ile Asp Ile Ile Lys Asp Met Leu

995 1000 1005

Ala Leu Glu Lys Thr Leu Ile Asp Asn Asp Glu Asn Leu Pro Thr Asn

1010 1015 1020

Lys Phe Ser Asp Met Leu Lys Ser Trp Leu Lys Gly Lys Gly Glu Ala

1025 1030 1035 1040

Asn Lys Ala Arg Leu Gln Asn Asp Val Asp Leu Leu Val Ala Ile Arg

1045 1050 1055

Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asn Ser Glu Val Phe

1060 1065 1070

Lys Gly Met Lys Leu Leu Ser Leu Ser Ser Asp Ile Pro Glu Lys Glu

1075 1080 1085

Gly Leu Gly Ile Ala Lys Gln Leu Lys Asp Lys Ile Lys Glu Thr Ile

1090 1095 1100

Glu Arg Ile Ile Glu Ile Glu Lys Glu Ile Arg Asn

1105 1110 1115

<210> 94

<211> 1119

<212> PRT

<213> Porphyromonas gingivalis

<400> 94

Met Thr Glu Gln Asn Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu

1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn

20 25 30

Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys

35 40 45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp

50 55 60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu

65 70 75 80

Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys

85 90 95

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp

210 215 220

Arg Glu Gly Thr Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240
Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255
Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe
260 265 270
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285
Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300
Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305 310 315 320
Arg Val Pro Ile Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
325 330 335
Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
340 345 350
Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380
Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385 390 395 400
Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
405 410 415
Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
420 425 430
Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445
Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
450 455 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465 470 475 480
Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
485 490 495
Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
515 520 525
Tyr Ala Val Tyr Asp Ala Phe Ala Arg Gly Glu Ile Asp Thr Leu Asp
530 535 540
Arg Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545 550 555 560
Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
565 570 575
Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590
Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
595 600 605
Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
610 615 620
Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625 630 635 640
Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
660 665 670
Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
675 680 685
Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
690 695 700
Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720

Ser Asp Arg Glu Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys

725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro

785 790 795 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly

820 825 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala

835 840 845

Lys Glu His Pro Tyr Leu Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg

850 855 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys

865 870 875 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr

885 890 895

Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val His Glu Gln Gly Ser Leu

900 905 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr

915 920 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala

930 935 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe

965 970 975

Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu

980 985 990

Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala

995 1000 1005

Phe Glu Gln Thr Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Tyr Pro

1010 1015 1020

His Leu Pro Asp Lys Asn Phe Arg Lys Met Leu Glu Ser Trp Ser Asp

1025 1030 1035 1040

Pro Leu Leu Asp Lys Trp Pro Asp Leu His Arg Lys Val Arg Leu Leu

1045 1050 1055

Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp

1060 1065 1070

Glu Ala Val Phe Ser Ser Ile Arg Lys Tyr Asp Pro Ser Ser Pro Asp

1075 1080 1085

Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu Ser Glu

1090 1095 1100

Glu Val Lys Gln Ala Lys Glu Met Ala Glu Arg Ile Ile Gln Val

1105 1110 1115

<210> 95

<211> 1175

<212> PRT

<213> Porphyromonas gulae

<400> 95

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu

1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn

20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys

35 40 45

Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp

50 55 60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
100 105 110
Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125
Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140
Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu
145 150 155 160
Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175
Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
180 185 190
Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205
Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
210 215 220
Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240
Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255
Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
260 265 270
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285
Met Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
290 295 300
Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Asp Asp Arg Ala Cys Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly

325 330 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu

385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Ser Gln Thr Ser Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp

465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro

485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val

500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val

515 520 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Lys Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asn Met Glu

565 570 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Ala Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685

Leu His Asp Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg

705 710 715 720

Ser Asp Arg Met Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys

725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro

785 790 795 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Glu Arg Ala Glu Glu Trp Glu Arg Gly
820 825 830
Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
835 840 845
Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
850 855 860
Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865 870 875 880
Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys

885 890 895
Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
900 905 910
Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
915 920 925
Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
930 935 940
Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945 950 955 960

Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn His Val Lys Pro
965 970 975
Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
980 985 990
His Lys Glu Glu Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asp
995 1000 1005
Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg
1010 1015 1020
Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Gly Leu Ala Met

1025 1030 1035 1040
Glu Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu Leu Ala Lys
1045 1050 1055
Tyr Gln Thr Ala Arg Val Cys Ala Phe Glu Gln Thr Leu Glu Leu Glu
1060 1065 1070

Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro Asp Lys Asn Phe Arg
1075 1080 1085
Lys Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Ala Lys Trp Pro Glu
1090 1095 1100

Leu His Gly Lys Val Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser
1105 1110 1115 1120
His Asn Gln Tyr Pro Met Tyr Asp Glu Ala Val Phe Ser Ser Ile Arg
1125 1130 1135
Lys Tyr Asp Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu
1140 1145 1150
Asn Ile Ala His Arg Leu Ser Glu Glu Val Lys Gln Ala Lys Glu Thr
1155 1160 1165
Val Glu Arg Ile Ile Gln Ala

1170 1175

<210> 96
<211> 1119
<212> PRT
<213> Porphyromonas gulae
<400> 96
Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15
Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
20 25 30
Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45
Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp

50 55 60
Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95
Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp

210 215 220

Gly Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Arg Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly

325 330 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380

Lys Val Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
420 425 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
450 455 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
515 520 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545 550 555 560

Pro Lys Gln Met Ile Gly Ile Leu Ser Gln Glu His Lys Asn Met Glu
565 570 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
610 615 620
Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly

625 630 635 640
Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
645 650 655
Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
660 665 670
Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
675 680 685
Leu Asp Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
690 695 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705 710 715 720
Ser Asp Arg Val Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
740 745 750
Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765
Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys

770 775 780
Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785 790 795 800
Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
805 810 815
Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
820 825 830
Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
835 840 845

Gln Glu His Pro Tyr His Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg

850 855 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys

865 870 875 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr

885 890 895

Leu Tyr Leu Lys Asp Ile Arg Pro Asn Val Gln Glu Gln Gly Ser Leu

900 905 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr

915 920 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Glu Ala Pro Leu Ala

930 935 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe

965 970 975

Val Asp Thr Gly Gly Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu

980 985 990

Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Val

995 1000 1005

Phe Glu Leu Thr Leu Arg Leu Glu Glu Ser Leu Leu Ser Arg Tyr Pro

1010 1015 1020

His Leu Pro Asp Glu Ser Phe Arg Glu Met Leu Glu Ser Trp Ser Asp

1025 1030 1035 1040

Pro Leu Leu Ala Lys Trp Pro Glu Leu His Gly Lys Val Arg Leu Leu

1045 1050 1055

Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp

1060 1065 1070

Glu Ala Val Phe Ser Ser Ile Arg Lys Tyr Asp Pro Ser Ser Pro Asp

1075 1080 1085

Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu Ser Glu

1090 1095 1100

Glu Val Lys Gln Ala Lys Glu Thr Val Glu Arg Ile Ile Gln Ala

1105 1110 1115

<210> 97

<211> 1119

<212> PRT

<213> Porphyromonas gulae

<400> 97

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu

1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn

20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys

35 40 45

Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp

50 55 60

Lys Asn Phe Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu

65 70 75 80

Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys

85 90 95

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Tyr His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
210 215 220

Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240
Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255
Phe Lys Gly Gly Thr Gly Pro Tyr Glu Gln Met Thr Asn Glu Val Phe
260 265 270
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285
Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300
Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Cys Phe
305 310 315 320
Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
325 330 335
Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
340 345 350
Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380
Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385 390 395 400
Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
405 410 415
Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
420 425 430
Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445
Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp

465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro

485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val

500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Lys Asp Val

515 520 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys

530 535 540

Glu Leu Asp Ala Cys Ser Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Lys Gln Met Ile Gly Ile Leu Ser Gln Glu His Lys Asn Met Glu

565 570 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685

Leu Asp Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg

705 710 715 720

Ser Asp Arg Val Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys

725 730 735

Asn Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro

785 790 795 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly

820 825 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala

835 840 845

Lys Glu His Pro Tyr His Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg

850 855 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys

865 870 875 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr

885 890 895

Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val His Glu Gln Gly Ser Leu

900 905 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr

915 920 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala

930 935 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
965 970 975
Val Asp Thr Gly Gly Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
980 985 990
Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala
995 1000 1005
Phe Glu Gln Thr Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Tyr Pro

1010 1015 1020
His Leu Pro Asp Glu Asn Phe Arg Glu Met Leu Glu Ser Trp Ser Asp
1025 1030 1035 1040
Pro Leu Leu Gly Lys Trp Pro Asp Leu His Gly Lys Val Arg Leu Leu
1045 1050 1055
Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp
1060 1065 1070
Glu Ala Val Phe Ser Ser Ile Arg Lys Tyr Asp Pro Ser Ser Pro Asp
1075 1080 1085

Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu Ser Glu
1090 1095 1100
Glu Val Lys Gln Ala Lys Glu Thr Val Glu Arg Ile Ile Gln Ala
1105 1110 1115

<210> 98
<211> 1175
<212> PRT
<213> Porphyromonas gulae

<400> 98
Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15
Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn

20 25 30
Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45
Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp

50 55 60

Lys Asn Phe Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
100 105 110
Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125
Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140
Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Glu Ser Ser Glu Leu
145 150 155 160
Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175
Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
180 185 190
Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205
Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
210 215 220
Gly Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255
Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
260 265 270
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285
Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
290 295 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Asp Asp Arg Ala Cys Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly

325 330 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu

385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Ser Gln Thr Ser Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp

465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro

485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val

500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val

515 520 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Lys Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu

565 570 575

Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Lys Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg

705 710 715 720

Ser Asp Arg Met Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys

725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro

785 790 795 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Glu Arg Ala Glu Glu Trp Glu Arg Gly
820 825 830
Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
835 840 845
Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
850 855 860
Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865 870 875 880
Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys

885 890 895
Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
900 905 910
Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
915 920 925
Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
930 935 940
Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Pro
945 950 955 960

Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro
965 970 975
Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
980 985 990
His Lys Glu Glu Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asp
995 1000 1005
Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg
1010 1015 1020
Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Gly Leu Ala Met

1025 1030 1035 1040
Glu Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu Leu Ala Lys
1045 1050 1055

Tyr Gln Thr Ala Arg Val Cys Val Phe Glu Leu Thr Leu Arg Leu Glu
1060 1065 1070
Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro Asp Glu Ser Phe Arg
1075 1080 1085
Lys Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Ala Lys Trp Pro Glu
1090 1095 1100

Leu His Gly Lys Val Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser
1105 1110 1115 1120
His Asn Gln Tyr Pro Met Tyr Asp Glu Ala Val Phe Ser Ser Ile Arg
1125 1130 1135
Lys Tyr Asp Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu
1140 1145 1150
Asn Ile Ala His Arg Leu Ser Glu Glu Val Lys Gln Ala Lys Glu Thr
1155 1160 1165
Val Glu Arg Ile Ile Gln Val

1170 1175

<210> 99
<211> 1119
<212> PRT
<213> Porphyromonas gulae
<400> 99
Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15
Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
20 25 30
Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45
Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp

50 55 60
Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys

85 90 95

Phe Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Glu Val

180 185 190

Asp Pro His Tyr His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp

210 215 220

Gly Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

Phe Lys Gly Gly Thr Glu Pro Tyr Glu Gln Met Thr Asn Glu Val Phe

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Cys Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly

325 330 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu

385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp

465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro

485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val

500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val

515 520 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Lys Gln Met Ile Gly Ile Leu Ser Gln Glu Arg Lys Asp Met Glu

565 570 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg

705 710 715 720

Ser Asp Arg Val Glu Asn Cys Pro Phe Leu Leu Leu Lys Glu Pro Lys

725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro

785 790 795 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly

820 825 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala

835 840 845

Gln Glu His Pro Tyr His Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
850 855 860
Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys
865 870 875 880
Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
885 890 895
Leu Tyr Leu Lys Asp Ile Arg Pro Asn Val Gln Glu Gln Gly Ser Leu
900 905 910
Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr

915 920 925
Arg Ala Asp Ser Arg Gly His Val His Lys Glu Glu Ala Pro Leu Ala
930 935 940
Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn
945 950 955 960
Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
965 970 975
Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
980 985 990

Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala
995 1000 1005
Phe Glu Gln Thr Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Tyr Pro
1010 1015 1020
His Leu Pro Asp Glu Ser Phe Arg Glu Met Leu Glu Ser Trp Ser Asp
1025 1030 1035 1040
Pro Leu Leu Thr Lys Trp Pro Glu Leu His Gly Lys Val Arg Leu Leu
1045 1050 1055
Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp

1060 1065 1070
Glu Ala Val Phe Ser Ser Ile Trp Lys Tyr Asp Pro Ser Ser Pro Asp
1075 1080 1085

Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu Ser Glu

1090 1095 1100

Glu Val Lys Gln Ala Lys Glu Thr Ile Glu Arg Ile Ile Gln Ala

1105 1110 1115

<210> 100

<211> 1174

<212> PRT

<213> Porphyromonas gulae

<400> 100

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu

1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn

20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys

35 40 45

Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp

50 55 60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu

65 70 75 80

Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys

85 90 95

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Tyr His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp

210 215 220

Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

Phe Lys Gly Gly Thr Gly Pro Tyr Glu Gln Met Thr Asn Glu Val Phe

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Cys Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly

325 330 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu

385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Leu Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Cys Ala Gln Asp

465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro

485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val

500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val

515 520 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys

530 535 540

Glu Leu Asp Thr Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Lys Gln Met Ile Thr Ile Leu Ser Gln Glu Arg Lys Asp Met Lys

565 570 575

Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Ala Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg

705 710 715 720

Ser Asp Arg Val Glu Asn Cys Pro Phe Leu Leu Leu Lys Glu Pro Lys

725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Asp Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro

785 790 795 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Asp Arg Ala Glu Glu Trp Glu Arg Gly

820 825 830

Met Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg

835 840 845

Arg Ile Lys Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn

850 855 860

Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala

865 870 875 880

Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys

885 890 895

Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe

900 905 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln

915 920 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys

930 935 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Pro
945 950 955 960
Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro
965 970 975
Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
980 985 990
His Lys Glu Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asn Thr
995 1000 1005
Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg Arg

1010 1015 1020
Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Gly Leu Ala Met Glu
1025 1030 1035 1040
Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu Leu Ala Lys Tyr
1045 1050 1055
Gln Thr Ala Arg Val Cys Val Phe Glu Leu Thr Leu Arg Leu Glu Glu
1060 1065 1070
Ser Leu Leu Ser Arg Tyr Pro His Leu Pro Asp Glu Ser Phe Arg Glu
1075 1080 1085

Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Ala Lys Trp Pro Glu Leu
1090 1095 1100
His Gly Lys Val Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser His
1105 1110 1115 1120
Asn Gln Tyr Pro Met Tyr Asp Glu Ala Val Phe Ser Ser Ile Arg Lys
1125 1130 1135
Tyr Asp Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu Asn
1140 1145 1150
Ile Ala His Arg Leu Ser Glu Glu Val Lys Gln Ala Lys Glu Thr Val

1155 1160 1165
Glu Arg Ile Ile Gln Ala
1170

<210> 101

<211> 1135

<212> PRT

<213> Porphyromonas gulae

<400> 101

Met Asn Thr Val Pro Ala Thr Glu Asn Lys Gly Gln Ser Arg Thr Val
1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
50 55 60

Leu Ser Ile Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65 70 75 80

Leu Pro Phe Leu His Cys Phe Asp Pro Asp Ser Gly Ile Glu Lys Asp
85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Lys Val Ser Pro Asp
130 135 140

Ile Ser Ser Phe Ile Thr Gly Ala Tyr Thr Phe Ala Cys Glu Arg Ala
145 150 155 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Leu Leu Ala
165 170 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
180 185 190

Leu Thr Val Ser Gly Phe Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
195 200 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr

210 215 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys

225 230 235 240

Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala

245 250 255

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu

260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu

275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu

290 295 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr

305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe

325 330 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu

340 345 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly

355 360 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu

370 375 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu

385 390 395 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile

405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro

420 425 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met

435 440 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu

450 455 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys

465 470 475 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser

485 490 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro

500 505 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr

515 520 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu

530 535 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp

545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr

565 570 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe

580 585 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu

595 600 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu

610 615 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg

625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala

645 650 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe

660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr

675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr

690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe

705 710 715 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg

725 730 735

Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln

740 745 750

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met

755 760 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr

785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile

805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu

820 825 830

Lys Thr Val Arg Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro

835 840 845

Val Pro Pro Asp Leu Ala Ala Tyr Ile Lys Arg Ser Phe His Arg Ala

850 855 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg

865 870 875 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp

885 890 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn

900 905 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu

915 920 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys

930 935 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg

945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu

965 970 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
980 985 990
Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg
995 1000 1005
Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser
1010 1015 1020
Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys
1025 1030 1035 1040
Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg Asn Lys Ala

1045 1050 1055
Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg
1060 1065 1070
Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly Ser Ser Ala Lys Asp
1075 1080 1085
Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu
1090 1095 1100
Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp His Glu Asn Arg Phe
1105 1110 1115 1120

Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu
1125 1130 1135

<210> 102
<211> 1218
<212> PRT
<213> Capnocytophaga cynodegmi
<400> 102
Met Glu Asn Lys Thr Ser Leu Gly Asn Asn Ile Tyr Tyr Asn Pro Phe
1 5 10 15
Lys Pro Gln Asp Lys Ser Tyr Phe Ala Gly Tyr Leu Asn Ala Ala Met
20 25 30
Glu Asn Ile Asp Ser Val Phe Arg Glu Leu Gly Lys Arg Leu Lys Gly

35 40 45

Lys Glu Tyr Thr Ser Glu Asn Phe Phe Asp Ala Ile Phe Lys Glu Asn

50 55 60

Ile Ser Leu Val Glu Tyr Glu Arg Tyr Val Lys Leu Leu Ser Asp Tyr

65 70 75 80

Phe Pro Met Ala Arg Leu Leu Asp Lys Lys Glu Val Pro Ile Lys Glu

85 90 95

Arg Lys Glu Asn Phe Lys Lys Asn Phe Arg Gly Ile Ile Lys Ala Val

100 105 110

Arg Asp Leu Arg Asn Phe Tyr Thr His Lys Glu His Gly Glu Val Glu

115 120 125

Ile Thr Asp Glu Ile Phe Gly Val Leu Asp Glu Met Leu Lys Ser Thr

130 135 140

Val Leu Thr Val Lys Lys Lys Lys Ile Lys Thr Asp Lys Thr Lys Glu

145 150 155 160

Ile Leu Lys Lys Ser Ile Glu Lys Gln Leu Asp Ile Leu Cys Gln Lys

165 170 175

Lys Leu Glu Tyr Leu Lys Asp Thr Ala Arg Lys Ile Glu Glu Lys Arg

180 185 190

Arg Asn Gln Arg Glu Arg Gly Glu Lys Lys Leu Val Pro Arg Phe Glu

195 200 205

Tyr Ser Asp Arg Arg Asp Asp Leu Ile Ala Ala Ile Tyr Asn Asp Ala

210 215 220

Phe Asp Val Tyr Ile Asp Lys Lys Lys Asp Ser Leu Lys Glu Ser Ser

225 230 235 240

Lys Thr Lys Tyr Asn Thr Glu Ser Tyr Pro Gln Gln Glu Glu Gly Asp

245 250 255

Leu Lys Ile Pro Ile Ser Lys Asn Gly Val Val Phe Leu Leu Ser Leu

260 265 270

Phe Leu Ser Lys Gln Glu Val His Ala Phe Lys Ser Lys Ile Ala Gly

275 280 285

Phe Lys Ala Thr Val Ile Asp Glu Ala Thr Val Ser His Arg Lys Asn

290 295 300

Ser Ile Cys Phe Met Ala Thr His Glu Ile Phe Ser His Leu Ala Tyr

305 310 315 320

Lys Lys Leu Lys Arg Lys Val Arg Thr Ala Glu Ile Asn Tyr Ser Glu

325 330 335

Ala Glu Asn Ala Glu Gln Leu Ser Ile Tyr Ala Lys Glu Thr Leu Met

340 345 350

Met Gln Met Leu Asp Glu Leu Ser Lys Val Pro Asp Val Val Tyr Gln

355 360 365

Asn Leu Ser Glu Asp Val Gln Lys Thr Phe Ile Glu Asp Trp Asn Glu

370 375 380

Tyr Leu Lys Glu Asn Asn Gly Asp Val Gly Thr Met Glu Glu Glu Gln

385 390 395 400

Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Lys Phe Asn Tyr

405 410 415

Phe Ala Ile Arg Phe Leu Asp Glu Phe Ala Gln Phe Pro Thr Leu Arg

420 425 430

Phe Gln Val His Leu Gly Asn Tyr Leu His Asp Ser Arg Pro Lys Glu

435 440 445

His Leu Ile Ser Asp Arg Arg Ile Lys Glu Lys Ile Thr Val Phe Gly

450 455 460

Arg Leu Ser Glu Leu Glu His Lys Lys Ala Leu Phe Ile Lys Asn Thr

465 470 475 480

Glu Thr Asn Glu Asp Arg Lys His Tyr Trp Glu Val Phe Pro Asn Pro

485 490 495

Asn Tyr Asp Phe Pro Lys Glu Asn Ile Ser Val Asn Asp Lys Asp Phe

500 505 510

Pro Ile Ala Gly Ser Ile Leu Asp Arg Glu Lys Gln Pro Thr Ala Gly

515 520 525

Lys Ile Gly Ile Lys Val Asn Leu Leu Asn Gln Lys Tyr Ile Ser Glu

530 535 540

Val Asp Lys Ala Val Lys Ala His Gln Leu Lys Gln Arg Asn Asn Lys
545 550 555 560
Pro Ser Ile Gln Asn Ile Ile Glu Glu Ile Val Pro Ile Asn Gly Ser
565 570 575
Asn Pro Lys Glu Ile Ile Val Phe Gly Gly Gln Pro Thr Ala Tyr Leu
580 585 590
Ser Met Asn Asp Ile His Ser Ile Leu Tyr Glu Phe Phe Asp Lys Trp
595 600 605
Glu Lys Lys Lys Glu Lys Leu Glu Lys Lys Gly Glu Lys Glu Leu Arg

610 615 620
Lys Glu Ile Gly Lys Glu Leu Glu Glu Lys Ile Val Gly Lys Ile Gln
625 630 635 640
Thr Gln Ile Gln Gln Ile Ile Asp Lys Asp Ile Asn Ala Lys Ile Leu
645 650 655
Lys Pro Tyr Gln Asp Asp Asp Ser Thr Ala Ile Asp Lys Glu Lys Leu
660 665 670
Ile Lys Asp Leu Lys Gln Glu Gln Lys Ile Leu Gln Lys Leu Lys Asn
675 680 685

Glu Gln Thr Ala Arg Glu Lys Glu Tyr Gln Glu Cys Ile Ala Tyr Gln
690 695 700
Glu Glu Ser Arg Lys Ile Lys Arg Ser Asp Lys Ser Arg Gln Lys Tyr
705 710 715 720
Leu Arg Asn Gln Leu Lys Arg Lys Tyr Pro Glu Val Pro Thr Arg Lys
725 730 735
Glu Ile Leu Tyr Tyr Gln Glu Lys Gly Lys Val Ala Val Trp Leu Ala
740 745 750
Asn Asp Ile Lys Arg Phe Met Pro Thr Asp Phe Lys Asn Glu Trp Lys

755 760 765
Gly Glu Gln His Ser Leu Leu Gln Lys Ser Leu Ala Tyr Tyr Glu Gln
770 775 780
Cys Lys Glu Glu Leu Lys Asn Leu Leu Pro Gln Gln Lys Val Phe Lys
785 790 795 800

His Leu Pro Phe Glu Leu Gly Gly His Phe Gln Gln Lys Tyr Leu Tyr
805 810 815
Gln Phe Tyr Thr Arg Tyr Leu Asp Lys Arg Leu Glu His Ile Ser Gly
820 825 830

Leu Val Gln Gln Ala Glu Asn Phe Lys Asn Glu Asn Lys Val Phe Lys
835 840 845
Lys Val Glu Asn Glu Cys Phe Lys Phe Leu Lys Lys Gln Asn Tyr Thr
850 855 860
His Lys Gly Leu Asp Ala Gln Ala Gln Ser Val Leu Gly Tyr Pro Ile
865 870 875 880
Phe Leu Glu Arg Gly Phe Met Asp Glu Lys Pro Thr Ile Ile Lys Gly
885 890 895
Lys Thr Phe Lys Gly Asn Glu Ser Leu Phe Thr Asp Trp Phe Arg Tyr

900 905 910
Tyr Lys Glu Tyr Gln Asn Phe Gln Thr Phe Tyr Asp Thr Glu Asn Tyr
915 920 925
Pro Leu Val Glu Leu Glu Lys Lys Gln Ala Asp Arg Lys Arg Glu Thr
930 935 940
Lys Ile Tyr Gln Gln Lys Lys Asn Asp Val Phe Thr Leu Leu Met Ala
945 950 955 960
Lys His Ile Phe Lys Ser Val Phe Lys Gln Asp Ser Ile Asp Arg Phe
965 970 975

Ser Leu Glu Asp Leu Tyr Gln Ser Arg Glu Glu Arg Leu Glu Asn Gln
980 985 990
Glu Lys Ala Lys Gln Thr Gly Glu Arg Asn Thr Asn Tyr Ile Trp Asn
995 1000 1005
Lys Thr Val Asp Leu Asn Leu Cys Asp Gly Lys Val Thr Val Glu Asn
1010 1015 1020
Val Lys Leu Lys Asn Val Gly Asn Phe Ile Lys Tyr Glu Tyr Asp Gln
1025 1030 1035 1040
Arg Val Gln Thr Phe Leu Lys Tyr Glu Glu Asn Ile Lys Trp Gln Ala

1045 1050 1055

Phe Leu Ile Lys Glu Ser Lys Glu Glu Glu Asn Tyr Pro Tyr Ile Val

1060 1065 1070

Glu Arg Glu Ile Glu Gln Tyr Glu Lys Val Arg Arg Glu Glu Leu Leu

1075 1080 1085

Lys Glu Val His Leu Ile Glu Glu Tyr Ile Leu Glu Lys Val Lys Asp

1090 1095 1100

Lys Glu Ile Leu Lys Lys Gly Asp Asn Gln Asn Phe Lys Tyr Tyr Ile

1105 1110 1115 1120

Leu Asn Gly Leu Leu Lys Gln Leu Lys Asn Glu Asp Val Glu Ser Tyr

1125 1130 1135

Lys Val Phe Asn Leu Asn Thr Lys Pro Glu Asp Val Asn Ile Asn Gln

1140 1145 1150

Leu Lys Gln Glu Ala Thr Asp Leu Glu Gln Lys Ala Phe Val Leu Thr

1155 1160 1165

Tyr Ile Arg Asn Lys Phe Ala His Asn Gln Leu Pro Lys Lys Glu Phe

1170 1175 1180

Trp Asp Tyr Cys Gln Glu Lys Tyr Gly Lys Ile Glu Lys Glu Lys Thr

1185 1190 1195 1200

Tyr Ala Glu Tyr Phe Ala Glu Val Phe Lys Arg Glu Lys Glu Ala Leu

1205 1210 1215

Met Lys

<210> 103

<211> 1090

<212> PRT

<213> Prevotella sp.

<400> 103

Met Asn Ile Pro Ala Leu Val Glu Asn Gln Lys Lys Tyr Phe Gly Thr

1 5 10 15

Tyr Ser Val Met Ala Met Leu Asn Ala Gln Thr Val Leu Asp His Ile

20 25 30

Gln Lys Val Ala Asp Ile Glu Gly Glu Gln Asn Glu Asn Asn Glu Asn
35 40 45
Leu Trp Phe His Pro Val Met Ser His Leu Tyr Asn Ala Lys Asn Gly
50 55 60
Tyr Asp Lys Gln Pro Glu Lys Thr Met Phe Ile Ile Glu Arg Leu Gln
65 70 75 80
Ser Tyr Phe Pro Phe Leu Lys Ile Met Ala Glu Asn Gln Arg Glu Tyr
85 90 95
Ser Asn Gly Lys Tyr Lys Gln Asn Arg Val Glu Val Asn Ser Asn Asp

100 105 110
Ile Phe Glu Val Leu Lys Arg Ala Phe Gly Val Leu Lys Met Tyr Arg
115 120 125
Asp Leu Thr Asn His Tyr Lys Thr Tyr Glu Glu Lys Leu Ile Asp Gly
130 135 140
Cys Glu Phe Leu Thr Ser Thr Glu Gln Pro Leu Ser Gly Met Ile Ser
145 150 155 160
Lys Tyr Tyr Thr Val Ala Leu Arg Asn Thr Lys Glu Arg Tyr Gly Tyr
165 170 175

Lys Thr Glu Asp Leu Ala Phe Ile Gln Asp Asn Ile Lys Lys Ile Thr
180 185 190
Lys Asp Ala Tyr Gly Lys Arg Lys Ser Gln Val Asn Thr Gly Phe Phe
195 200 205
Leu Ser Leu Gln Asp Tyr Asn Gly Asp Thr Gln Lys Lys Leu His Leu
210 215 220
Ser Gly Val Gly Ile Ala Leu Leu Ile Cys Leu Phe Leu Asp Lys Gln
225 230 235 240
Tyr Ile Asn Ile Phe Leu Ser Arg Leu Pro Ile Phe Ser Ser Tyr Asn

245 250 255
Ala Gln Ser Glu Glu Arg Arg Ile Ile Ile Arg Ser Phe Gly Ile Asn
260 265 270
Ser Ile Lys Leu Pro Lys Asp Arg Ile His Ser Glu Lys Ser Asn Lys
275 280 285

Ser Val Ala Met Asp Met Leu Asn Glu Val Lys Arg Cys Pro Asp Glu
290 295 300
Leu Phe Thr Thr Leu Ser Ala Glu Lys Gln Ser Arg Phe Arg Ile Ile
305 310 315 320

Ser Asp Asp His Asn Glu Val Leu Met Lys Arg Ser Thr Asp Arg Phe
325 330 335
Val Pro Leu Leu Leu Gln Tyr Ile Asp Tyr Gly Lys Leu Phe Asp His
340 345 350
Ile Arg Phe His Val Asn Met Gly Lys Leu Arg Tyr Leu Leu Lys Ala
355 360 365
Asp Lys Thr Cys Ile Asp Gly Gln Thr Arg Val Arg Val Ile Glu Gln
370 375 380
Pro Leu Asn Gly Phe Gly Arg Leu Glu Glu Ala Glu Thr Met Arg Lys

385 390 395 400
Gln Glu Asn Gly Thr Phe Gly Asn Ser Gly Ile Arg Ile Arg Asp Phe
405 410 415
Glu Asn Val Lys Arg Asp Asp Ala Asn Pro Ala Asn Tyr Pro Tyr Ile
420 425 430
Val Asp Thr Tyr Thr His Tyr Ile Leu Glu Asn Asn Lys Val Glu Met
435 440 445
Phe Ile Ser Asp Lys Gly Ser Ser Ala Pro Leu Leu Pro Leu Ile Glu
450 455 460

Asp Asp Arg Tyr Val Val Lys Thr Ile Pro Ser Cys Arg Met Ser Thr
465 470 475 480
Leu Glu Ile Pro Ala Met Ala Phe His Met Phe Leu Phe Gly Ser Lys
485 490 495
Lys Thr Glu Lys Leu Ile Val Asp Val His Asn Arg Tyr Lys Arg Leu
500 505 510
Phe Gln Ala Met Gln Lys Glu Glu Val Thr Ala Glu Asn Ile Ala Ser
515 520 525
Phe Gly Ile Ala Glu Ser Asp Leu Pro Gln Lys Ile Leu Asp Leu Ile

530 535 540

Ser Gly Asn Ala His Gly Lys Asp Val Asp Ala Phe Ile Arg Leu Thr

545 550 555 560

Val Asp Asp Met Leu Thr Asp Thr Glu Arg Arg Ile Lys Arg Phe Lys

565 570 575

Asp Asp Arg Lys Ser Ile Arg Ser Ala Asp Asn Lys Met Gly Lys Arg

580 585 590

Gly Phe Lys Gln Ile Ser Thr Gly Lys Leu Ala Asp Phe Leu Ala Lys

595 600 605

Asp Ile Val Leu Phe Gln Pro Ser Val Asn Asp Gly Glu Asn Lys Ile

610 615 620

Thr Gly Leu Asn Tyr Arg Ile Met Gln Ser Ala Ile Ala Val Tyr Asp

625 630 635 640

Ser Gly Asp Asp Tyr Glu Ala Lys Gln Gln Phe Lys Leu Met Phe Glu

645 650 655

Lys Ala Arg Leu Ile Gly Lys Gly Thr Thr Glu Pro His Pro Phe Leu

660 665 670

Tyr Lys Val Phe Ala Arg Ser Ile Pro Ala Asn Ala Val Asp Phe Tyr

675 680 685

Glu Arg Tyr Leu Ile Glu Arg Lys Phe Tyr Leu Thr Gly Leu Cys Asn

690 695 700

Glu Ile Lys Arg Gly Asn Arg Val Asp Val Pro Phe Ile Arg Arg Asp

705 710 715 720

Gln Asn Lys Trp Lys Thr Pro Ala Met Lys Thr Leu Gly Arg Ile Tyr

725 730 735

Ser Glu Asp Leu Pro Val Glu Leu Pro Arg Gln Met Phe Asp Asn Glu

740 745 750

Ile Lys Ser His Leu Lys Ser Leu Pro Gln Met Glu Gly Ile Asp Phe

755 760 765

Asn Asn Ala Asn Val Thr Tyr Leu Ile Ala Glu Tyr Met Lys Arg Val

770 775 780

Leu Asn Asp Asp Phe Gln Thr Phe Tyr Gln Trp Lys Arg Asn Tyr His

785 790 795 800

Tyr Met Asp Met Leu Lys Gly Glu Tyr Asp Arg Lys Gly Ser Leu Gln

805 810 815

His Cys Phe Thr Ser Val Glu Glu Arg Glu Gly Leu Trp Lys Glu Arg

820 825 830

Ala Ser Arg Thr Glu Arg Tyr Arg Lys Leu Ala Ser Asn Lys Ile Arg

835 840 845

Ser Asn Arg Gln Met Arg Asn Ala Ser Ser Glu Glu Ile Glu Thr Ile

850 855 860

Leu Asp Lys Arg Leu Ser Asn Cys Arg Asn Glu Tyr Gln Lys Ser Glu

865 870 875 880

Lys Val Ile Arg Arg Tyr Arg Val Gln Asp Ala Leu Leu Phe Leu Leu

885 890 895

Ala Lys Lys Thr Leu Thr Glu Leu Ala Asp Phe Asp Gly Glu Arg Phe

900 905 910

Lys Leu Lys Glu Ile Met Pro Asp Ala Glu Lys Gly Ile Leu Ser Glu

915 920 925

Ile Met Pro Met Ser Phe Thr Phe Glu Lys Gly Gly Lys Lys Tyr Thr

930 935 940

Ile Thr Ser Glu Gly Met Lys Leu Lys Asn Tyr Gly Asp Phe Phe Val

945 950 955 960

Leu Ala Ser Asp Lys Arg Ile Gly Asn Leu Leu Glu Leu Val Gly Ser

965 970 975

Asp Ile Val Ser Lys Glu Asp Ile Met Glu Glu Phe Asn Lys Tyr Asp

980 985 990

Gln Cys Arg Pro Glu Ile Ser Ser Ile Val Phe Asn Leu Glu Lys Trp

995 1000 1005

Ala Phe Asp Thr Tyr Pro Glu Leu Ser Ala Arg Val Asp Arg Glu Glu

1010 1015 1020

Lys Val Asp Phe Lys Ser Ile Leu Lys Ile Leu Leu Asn Asn Lys Asn

1025 1030 1035 1040

Ile Asn Lys Glu Gln Ser Asp Ile Leu Arg Lys Ile Arg Asn Ala Phe
1045 1050 1055

Asp His Asn Asn Tyr Pro Asp Lys Gly Ile Val Glu Ile Lys Ala Leu
1060 1065 1070

Pro Glu Ile Ala Met Ser Ile Lys Lys Ala Phe Gly Glu Tyr Ala Ile
1075 1080 1085

Met Lys
1090

<210> 104

<211> 1090

<212> PRT

<213> Porphyromonas gulae

<400> 104

Met Asn Ile Pro Ala Leu Val Glu Asn Gln Lys Lys Tyr Phe Gly Thr
1 5 10 15

Tyr Ser Val Met Ala Met Leu Asn Ala Gln Thr Val Leu Asp His Ile
20 25 30

Gln Lys Val Ala Asp Ile Glu Gly Glu Gln Asn Glu Asn Asn Glu Asn
35 40 45

Leu Trp Phe His Pro Val Met Ser His Leu Tyr Asn Ala Lys Asn Gly
50 55 60

Tyr Asp Lys Gln Pro Glu Lys Thr Met Phe Ile Ile Glu Arg Leu Gln
65 70 75 80

Ser Tyr Phe Pro Phe Leu Lys Ile Met Ala Glu Asn Gln Arg Glu Tyr
85 90 95

Ser Asn Gly Lys Tyr Lys Gln Asn Arg Val Glu Val Asn Ser Asn Asp
100 105 110

Ile Phe Glu Val Leu Lys Arg Ala Phe Gly Val Leu Lys Met Tyr Arg
115 120 125

Asp Gln Ala Ser His Tyr Lys Thr Tyr Asp Glu Lys Leu Ile Asp Gly
130 135 140

Cys Glu Phe Leu Thr Ser Thr Glu Gln Pro Leu Ser Gly Met Ile Asn

145 150 155 160

Asn Tyr Tyr Thr Val Ala Leu Arg Asn Met Asn Glu Arg Tyr Gly Tyr

165 170 175

Lys Thr Glu Asp Leu Ala Phe Ile Gln Asp Lys Arg Phe Lys Phe Val

180 185 190

Lys Asp Ala Tyr Gly Lys Lys Lys Ser Gln Val Asn Thr Gly Phe Phe

195 200 205

Leu Ser Leu Gln Asp Tyr Asn Gly Asp Thr Gln Lys Lys Leu His Leu

210 215 220

Ser Gly Val Gly Ile Ala Leu Leu Ile Cys Leu Phe Leu Asp Lys Gln

225 230 235 240

Tyr Ile Asn Ile Phe Leu Ser Arg Leu Pro Ile Phe Ser Ser Tyr Asn

245 250 255

Ala Gln Ser Glu Glu Arg Arg Ile Ile Ile Arg Ser Phe Gly Ile Asn

260 265 270

Ser Ile Lys Gln Pro Lys Asp Arg Ile His Ser Glu Lys Ser Asn Lys

275 280 285

Ser Val Ala Met Asp Met Leu Asn Glu Ile Lys Arg Cys Pro Asn Glu

290 295 300

Leu Phe Glu Thr Leu Ser Ala Glu Lys Gln Ser Arg Phe Arg Ile Ile

305 310 315 320

Ser Asn Asp His Asn Glu Val Leu Met Lys Arg Ser Ser Asp Arg Phe

325 330 335

Val Pro Leu Leu Leu Gln Tyr Ile Asp Tyr Gly Lys Leu Phe Asp His

340 345 350

Ile Arg Phe His Val Asn Met Gly Lys Leu Arg Tyr Leu Leu Lys Ala

355 360 365

Asp Lys Thr Cys Ile Asp Gly Gln Thr Arg Val Arg Val Ile Glu Gln

370 375 380

Pro Leu Asn Gly Phe Gly Arg Leu Glu Glu Val Glu Thr Met Arg Lys

385 390 395 400

Gln Glu Asn Gly Thr Phe Gly Asn Ser Gly Ile Arg Ile Arg Asp Phe

405 410 415

Glu Asn Met Lys Arg Asp Asp Ala Asn Pro Ala Asn Tyr Pro Tyr Ile

420 425 430

Val Asp Thr Tyr Thr His Tyr Ile Leu Glu Asn Asn Lys Val Glu Met

435 440 445

Phe Ile Ser Asp Glu Glu Thr Pro Ala Pro Leu Leu Pro Val Ile Glu

450 455 460

Asp Asp Arg Tyr Val Val Lys Thr Ile Pro Ser Cys Arg Met Ser Thr

465 470 475 480

Leu Glu Ile Pro Ala Met Ala Phe His Met Phe Leu Phe Gly Ser Lys

485 490 495

Lys Thr Glu Lys Leu Ile Val Asp Val His Asn Arg Tyr Lys Arg Leu

500 505 510

Phe Lys Ala Met Gln Lys Glu Glu Val Thr Ala Glu Asn Ile Ala Ser

515 520 525

Phe Gly Ile Ala Glu Ser Asp Leu Pro Gln Lys Ile Ile Asp Leu Ile

530 535 540

Ser Gly Asn Ala His Gly Lys Asp Val Asp Ala Phe Ile Arg Leu Thr

545 550 555 560

Val Asp Asp Met Leu Ala Asp Thr Glu Arg Arg Ile Lys Arg Phe Lys

565 570 575

Asp Asp Arg Lys Ser Ile Arg Ser Ala Asp Asn Lys Met Gly Lys Arg

580 585 590

Gly Phe Lys Gln Ile Ser Thr Gly Lys Leu Ala Asp Phe Leu Ala Lys

595 600 605

Asp Ile Val Leu Phe Gln Pro Ser Val Asn Asp Gly Glu Asn Lys Ile

610 615 620

Thr Gly Leu Asn Tyr Arg Ile Met Gln Ser Ala Ile Ala Val Tyr Asn

625 630 635 640

Ser Gly Asp Asp Tyr Glu Ala Lys Gln Gln Phe Lys Leu Met Phe Glu

645 650 655

Lys Ala Arg Leu Ile Gly Lys Gly Thr Thr Glu Pro His Pro Phe Leu
660 665 670

Tyr Lys Val Phe Val Arg Ser Ile Pro Ala Asn Ala Val Asp Phe Tyr
675 680 685

Glu Arg Tyr Leu Ile Glu Arg Lys Phe Tyr Leu Ile Gly Leu Ser Asn
690 695 700

Glu Ile Lys Lys Gly Asn Arg Val Asp Val Pro Phe Ile Arg Arg Asp
705 710 715 720

Gln Asn Lys Trp Lys Thr Pro Ala Met Lys Thr Leu Gly Arg Ile Tyr
725 730 735

Asp Glu Asp Leu Pro Val Glu Leu Pro Arg Gln Met Phe Asp Asn Glu
740 745 750

Ile Lys Ser His Leu Lys Ser Leu Pro Gln Met Glu Gly Ile Asp Phe
755 760 765

Asn Asn Ala Asn Val Thr Tyr Leu Ile Ala Glu Tyr Met Lys Arg Val
770 775 780

Leu Asn Asp Asp Phe Gln Thr Phe Tyr Gln Trp Lys Arg Asn Tyr Arg
785 790 795 800

Tyr Met Asp Met Leu Arg Gly Glu Tyr Asp Arg Lys Gly Ser Leu Gln
805 810 815

Ser Cys Phe Thr Ser Val Glu Glu Arg Glu Gly Leu Trp Lys Glu Arg
820 825 830

Ala Ser Arg Thr Glu Arg Tyr Arg Lys Leu Ala Ser Asn Lys Ile Arg
835 840 845

Ser Asn Arg Gln Met Arg Asn Ala Ser Ser Glu Glu Ile Glu Thr Ile
850 855 860

Leu Asp Lys Arg Leu Ser Asn Ser Arg Asn Glu Tyr Gln Lys Ser Glu
865 870 875 880

Lys Val Ile Arg Arg Tyr Arg Val Gln Asp Ala Leu Leu Phe Leu Leu
885 890 895

Ala Lys Lys Thr Leu Thr Glu Leu Ala Asp Phe Asp Gly Glu Arg Phe
900 905 910

Lys Leu Lys Glu Ile Met Pro Asp Ala Glu Lys Gly Ile Leu Ser Glu
915 920 925
Ile Met Pro Met Ser Phe Thr Phe Glu Lys Gly Gly Lys Lys Tyr Thr
930 935 940

Ile Thr Ser Glu Gly Met Lys Leu Lys Asn Tyr Gly Asp Phe Phe Val
945 950 955 960
Leu Ala Ser Asp Lys Arg Ile Gly Asn Leu Leu Glu Leu Val Gly Ser
965 970 975
Asp Thr Val Ser Lys Glu Asp Ile Met Glu Glu Phe Lys Lys Tyr Asp
980 985 990
Gln Cys Arg Pro Glu Ile Ser Ser Ile Val Phe Asn Leu Glu Lys Trp
995 1000 1005
Ala Phe Asp Thr Tyr Pro Glu Leu Ser Ala Arg Val Asp Arg Glu Glu

1010 1015 1020
Lys Val Asp Phe Lys Ser Ile Leu Lys Ile Leu Leu Asn Asn Lys Asn
1025 1030 1035 1040
Ile Asn Lys Glu Gln Ser Asp Ile Leu Arg Lys Ile Arg Asn Ala Phe
1045 1050 1055
Asp His Asn Asn Tyr Pro Asp Lys Gly Val Val Glu Ile Arg Ala Leu
1060 1065 1070
Pro Glu Ile Ala Met Ser Ile Lys Lys Ala Phe Gly Glu Tyr Ala Ile
1075 1080 1085

Met Lys
1090

<210> 105
<211> 1090
<212> PRT
<213> Porphyromonas gulae
<400> 105
Met Asn Ile Pro Ala Leu Val Glu Asn Gln Lys Lys Tyr Phe Gly Thr
1 5 10 15
Tyr Ser Val Met Ala Met Leu Asn Ala Gln Thr Val Leu Asp His Ile

20 25 30

Gln Lys Val Ala Asp Ile Glu Gly Glu Gln Asn Glu Asn Asn Glu Asn

35 40 45

Leu Trp Phe His Pro Val Met Ser His Leu Tyr Asn Ala Lys Asn Gly

50 55 60

Tyr Asp Lys Gln Pro Glu Lys Thr Met Phe Ile Ile Glu Arg Leu Gln

65 70 75 80

Ser Tyr Phe Pro Phe Leu Lys Ile Met Ala Glu Asn Gln Arg Glu Tyr

85 90 95

Ser Asn Gly Lys Tyr Lys Gln Asn Arg Val Glu Val Asn Ser Asn Asp

100 105 110

Ile Phe Glu Val Leu Lys Arg Ala Phe Gly Val Leu Lys Met Tyr Arg

115 120 125

Asp Leu Thr Asn His Tyr Lys Thr Tyr Glu Glu Lys Leu Ile Asp Gly

130 135 140

Cys Glu Phe Leu Thr Ser Thr Glu Gln Pro Phe Ser Gly Met Ile Ser

145 150 155 160

Lys Tyr Tyr Thr Val Ala Leu Arg Asn Thr Lys Glu Arg Tyr Gly Tyr

165 170 175

Lys Ala Glu Asp Leu Ala Phe Ile Gln Asp Asn Arg Tyr Lys Phe Thr

180 185 190

Lys Asp Ala Tyr Gly Lys Arg Lys Ser Gln Val Asn Thr Gly Ser Phe

195 200 205

Leu Ser Leu Gln Asp Tyr Asn Gly Asp Thr Thr Lys Lys Leu His Leu

210 215 220

Ser Gly Val Gly Ile Ala Leu Leu Ile Cys Leu Phe Leu Asp Lys Gln

225 230 235 240

Tyr Ile Asn Leu Phe Leu Ser Arg Leu Pro Ile Phe Ser Ser Tyr Asn

245 250 255

Ala Gln Ser Glu Glu Arg Arg Ile Ile Ile Arg Ser Phe Gly Ile Asn

260 265 270

Ser Ile Lys Gln Pro Lys Asp Arg Ile His Ser Glu Lys Ser Asn Lys
275 280 285
Ser Val Ala Met Asp Met Leu Asn Glu Val Lys Arg Cys Pro Asp Glu
290 295 300
Leu Phe Thr Thr Leu Ser Ala Glu Lys Gln Ser Arg Phe Arg Ile Ile
305 310 315 320
Ser Asp Asp His Asn Glu Val Leu Met Lys Arg Ser Ser Asp Arg Phe
325 330 335
Val Pro Leu Leu Leu Gln Tyr Ile Asp Tyr Gly Lys Leu Phe Asp His

340 345 350
Ile Arg Phe His Val Asn Met Gly Lys Leu Arg Tyr Leu Leu Lys Ala
355 360 365
Asp Lys Thr Cys Ile Asp Gly Gln Thr Arg Val Arg Val Ile Glu Gln
370 375 380
Pro Leu Asn Gly Phe Gly Arg Leu Glu Glu Val Glu Thr Met Arg Lys
385 390 395 400
Gln Glu Asn Gly Thr Phe Gly Asn Ser Gly Ile Arg Ile Arg Asp Phe
405 410 415

Glu Asn Met Lys Arg Asp Asp Ala Asn Pro Ala Asn Tyr Pro Tyr Ile
420 425 430
Val Glu Thr Tyr Thr His Tyr Ile Leu Glu Asn Asn Lys Val Glu Met
435 440 445
Phe Ile Ser Asp Glu Glu Asn Pro Thr Pro Leu Leu Pro Val Ile Glu
450 455 460
Asp Asp Arg Tyr Val Val Lys Thr Ile Pro Ser Cys Arg Met Ser Thr
465 470 475 480
Leu Glu Ile Pro Ala Met Ala Phe His Met Phe Leu Phe Gly Ser Glu

485 490 495
Lys Thr Glu Lys Leu Ile Ile Asp Val His Asp Arg Tyr Lys Arg Leu
500 505 510
Phe Gln Ala Met Gln Lys Glu Glu Val Thr Ala Glu Asn Ile Ala Ser
515 520 525

Phe Gly Ile Ala Glu Ser Asp Leu Pro Gln Lys Ile Met Asp Leu Ile
530 535 540
Ser Gly Asn Ala His Gly Lys Asp Val Asp Ala Phe Ile Arg Leu Thr
545 550 555 560

Val Asp Asp Met Leu Thr Asp Thr Glu Arg Arg Ile Lys Arg Phe Lys
565 570 575
Asp Asp Arg Lys Ser Ile Arg Ser Ala Asp Asn Lys Met Gly Lys Arg
580 585 590
Gly Phe Lys Gln Ile Ser Thr Gly Lys Leu Ala Asp Phe Leu Ala Lys
595 600 605
Asp Ile Val Leu Phe Gln Pro Ser Val Asn Asp Gly Glu Asn Lys Ile
610 615 620
Thr Gly Leu Asn Tyr Arg Ile Met Gln Ser Ala Ile Ala Val Tyr Asp

625 630 635 640
Ser Gly Asp Asp Tyr Glu Ala Lys Gln Gln Phe Lys Leu Met Phe Glu
645 650 655
Lys Ala Arg Leu Ile Gly Lys Gly Thr Thr Glu Pro His Pro Phe Leu
660 665 670
Tyr Lys Val Phe Val Arg Ser Ile Pro Ala Asn Ala Val Asp Phe Tyr
675 680 685
Glu Arg Tyr Leu Ile Glu Arg Lys Phe Tyr Leu Ile Gly Leu Ser Asn
690 695 700

Glu Ile Lys Lys Gly Asn Arg Val Asp Val Pro Phe Ile Arg Arg Asp
705 710 715 720
Gln Asn Lys Trp Lys Thr Pro Ala Met Lys Thr Leu Gly Arg Ile Tyr
725 730 735
Ser Glu Asp Leu Pro Val Glu Leu Pro Arg Gln Met Phe Asp Asn Glu
740 745 750
Ile Lys Ser His Leu Lys Ser Leu Pro Gln Met Glu Gly Ile Asp Phe
755 760 765
Asn Asn Ala Asn Val Thr Tyr Leu Ile Ala Glu Tyr Met Lys Arg Val

770 775 780

Leu Asn Asp Asp Phe Gln Thr Phe Tyr Gln Trp Lys Arg Asn Tyr Arg

785 790 795 800

Tyr Met Asp Met Leu Arg Gly Glu Tyr Asp Arg Lys Gly Ser Leu Gln

805 810 815

His Cys Phe Thr Ser Ile Glu Glu Arg Glu Gly Leu Trp Lys Glu Arg

820 825 830

Ala Ser Arg Thr Glu Arg Tyr Arg Lys Leu Ala Ser Asn Lys Ile Arg

835 840 845

Ser Asn Arg Gln Met Arg Asn Ala Ser Ser Glu Glu Ile Glu Thr Ile

850 855 860

Leu Asp Lys Arg Leu Ser Asn Cys Arg Asn Glu Tyr Gln Lys Ser Glu

865 870 875 880

Lys Ile Ile Arg Arg Tyr Arg Val Gln Asp Ala Leu Leu Phe Leu Leu

885 890 895

Ala Lys Lys Thr Leu Thr Glu Leu Ala Asp Phe Asp Gly Glu Arg Phe

900 905 910

Lys Leu Lys Glu Ile Met Pro Asp Ala Glu Lys Gly Ile Leu Ser Glu

915 920 925

Ile Met Pro Met Ser Phe Thr Phe Glu Lys Gly Gly Lys Ile Tyr Thr

930 935 940

Ile Thr Ser Gly Gly Met Lys Leu Lys Asn Tyr Gly Asp Phe Phe Val

945 950 955 960

Leu Ala Ser Asp Lys Arg Ile Gly Asn Leu Leu Glu Leu Val Gly Ser

965 970 975

Asn Thr Val Ser Lys Glu Asp Ile Met Glu Glu Phe Lys Lys Tyr Asp

980 985 990

Gln Cys Arg Pro Glu Ile Ser Ser Ile Val Phe Asn Leu Glu Lys Trp

995 1000 1005

Ala Phe Asp Thr Tyr Pro Glu Leu Pro Ala Arg Val Asp Arg Lys Glu

1010 1015 1020

Lys Val Asp Phe Trp Ser Ile Leu Asp Val Leu Ser Asn Asn Lys Asp

1025 1030 1035 1040

Ile Asn Asn Glu Gln Ser Tyr Ile Leu Arg Lys Ile Arg Asn Ala Phe

1045 1050 1055

Asp His Asn Asn Tyr Pro Asp Lys Gly Ile Val Glu Ile Lys Ala Leu

1060 1065 1070

Pro Glu Ile Ala Met Ser Ile Lys Lys Ala Phe Gly Glu Tyr Ala Ile

1075 1080 1085

Met Lys

1090

<210> 106

<211> 1131

<212> PRT

<213> Porphyromonas gulae

<400> 106

Met Thr Asn Thr Pro Lys Arg Arg Thr Leu His Arg His Pro Ser Tyr

1 5 10 15

Phe Gly Ala Phe Leu Asn Ile Ala Arg His Asn Ala Phe Met Ile Met

20 25 30

Glu His Leu Ser Thr Lys Tyr Asp Met Glu Asp Lys Asn Thr Leu Asp

35 40 45

Glu Ala Gln Leu Pro Asn Ala Lys Leu Phe Gly Cys Leu Lys Lys Arg

50 55 60

Tyr Gly Lys Pro Asp Val Thr Glu Gly Val Ser Arg Asp Leu Arg Arg

65 70 75 80

Tyr Phe Pro Phe Leu Asn Tyr Pro Leu Phe Leu His Leu Glu Lys Gln

85 90 95

Gln Asn Ala Glu Gln Ala Ala Thr Tyr Asp Ile Asn Pro Glu Asp Ile

100 105 110

Glu Phe Thr Leu Lys Gly Phe Phe Arg Leu Leu Asn Gln Met Arg Asn

115 120 125

Asn Tyr Ser His Tyr Ile Ser Asn Thr Asp Tyr Gly Lys Phe Asp Lys

130 135 140

Leu Pro Val Gln Asp Ile Tyr Glu Ala Ala Ile Phe Arg Leu Leu Asp
145 150 155 160
Arg Gly Lys His Thr Lys Arg Phe Asp Val Phe Glu Ser Lys His Thr
165 170 175

Arg His Leu Glu Ser Asn Asn Ser Glu Tyr Arg Pro Arg Ser Leu Ala
180 185 190
Asn Ser Pro Asp His Glu Asn Thr Val Ala Phe Val Thr Cys Leu Phe
195 200 205
Leu Glu Arg Lys Tyr Ala Phe Pro Phe Leu Ser Arg Leu Asp Cys Phe
210 215 220
Arg Ser Thr Asn Asp Ala Ala Glu Gly Asp Pro Leu Ile Arg Lys Ala
225 230 235 240
Ser His Glu Cys Tyr Thr Met Phe Cys Cys Arg Leu Pro Gln Pro Lys

245 250 255
Leu Glu Ser Ser Asp Ile Leu Leu Asp Met Val Asn Glu Leu Gly Arg
260 265 270
Cys Pro Ser Ala Leu Tyr Asn Leu Leu Ser Glu Glu Asp Gln Ala Arg
275 280 285
Phe His Ile Lys Arg Glu Glu Ile Thr Gly Phe Glu Glu Asp Pro Asp
290 295 300
Glu Glu Leu Glu Gln Glu Ile Val Leu Lys Arg His Ser Asp Arg Phe
305 310 315 320

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Asp Thr Glu Ala Phe Gln Thr
325 330 335
Leu Arg Phe Asp Val Tyr Leu Gly Arg Trp Arg Thr Lys Pro Val Tyr
340 345 350
Lys Lys Arg Ile Tyr Gly Gln Glu Arg Asp Arg Val Leu Thr Gln Ser
355 360 365
Ile Arg Thr Phe Thr Arg Leu Ser Arg Leu Leu Pro Ile Tyr Glu Asn
370 375 380
Val Lys His Asp Ala Val Arg Gln Asn Glu Glu Asp Gly Lys Leu Val

385 390 395 400

Asn Pro Asp Val Thr Ser Gln Phe His Lys Ser Trp Ile Gln Ile Glu

405 410 415

Ser Asp Asp Arg Ala Phe Leu Ser Asp Arg Ile Glu His Phe Ser Pro

420 425 430

His Tyr Asn Phe Gly Asp Gln Val Ile Gly Leu Lys Phe Ile Asn Pro

435 440 445

Asp Arg Tyr Ala Ala Ile Gln Asn Val Phe Pro Lys Leu Pro Gly Glu

450 455 460

Glu Lys Lys Asp Lys Asp Ala Lys Leu Val Asn Glu Thr Ala Asp Ala

465 470 475 480

Ile Ile Ser Thr His Glu Ile Arg Ser Leu Phe Leu Tyr His Tyr Leu

485 490 495

Ser Lys Lys Pro Ile Ser Ala Gly Asp Glu Arg Arg Phe Ile Gln Val

500 505 510

Asp Thr Glu Thr Phe Ile Lys Gln Tyr Ile Asp Thr Ile Lys Leu Phe

515 520 525

Phe Glu Asp Ile Lys Ser Gly Glu Leu Gln Pro Ile Ala Asp Pro Pro

530 535 540

Asn Tyr Gln Lys Asn Glu Pro Leu Pro Tyr Val Arg Gly Asp Lys Glu

545 550 555 560

Lys Thr Gln Glu Glu Arg Ala Gln Tyr Arg Glu Arg Gln Lys Glu Ile

565 570 575

Lys Glu Arg Arg Lys Glu Leu Asn Thr Leu Leu Gln Asn Arg Tyr Gly

580 585 590

Leu Ser Ile Gln Tyr Ile Pro Ser Arg Leu Arg Glu Tyr Leu Leu Gly

595 600 605

Tyr Lys Lys Val Pro Tyr Glu Lys Leu Ala Leu Gln Lys Leu Arg Ala

610 615 620

Gln Arg Lys Glu Val Lys Lys Arg Ile Lys Asp Ile Glu Lys Met Arg

625 630 635 640

Thr Pro Arg Val Gly Glu Gln Ala Thr Trp Leu Ala Glu Asp Ile Val

645 650 655

Phe Leu Thr Pro Pro Lys Met His Thr Pro Glu Arg Lys Thr Thr Lys

660 665 670

His Pro Gln Lys Leu Asn Asn Asp Gln Phe Arg Ile Met Gln Ser Ser

675 680 685

Leu Ala Tyr Phe Ser Val Asn Lys Lys Ala Ile Lys Lys Phe Phe Gln

690 695 700

Lys Glu Thr Gly Ile Gly Leu Ser Asn Arg Glu Thr Ser His Pro Phe

705 710 715 720

Leu Tyr Arg Ile Asp Val Gly Arg Cys Arg Gly Ile Leu Asp Phe Tyr

725 730 735

Thr Gly Tyr Leu Lys Tyr Lys Met Asp Trp Leu Asp Asp Ala Ile Lys

740 745 750

Lys Val Asp Asn Arg Lys His Gly Lys Lys Glu Ala Lys Lys Tyr Glu

755 760 765

Lys Tyr Leu Pro Ser Ser Ile Gln His Lys Thr Pro Leu Glu Leu Asp

770 775 780

Tyr Thr Arg Leu Pro Val Tyr Leu Pro Arg Gly Leu Phe Lys Lys Ala

785 790 795 800

Ile Val Lys Ala Leu Ala Ala His Ala Asp Phe Gln Val Glu Pro Glu

805 810 815

Glu Asp Asn Val Ile Phe Cys Leu Asp Gln Leu Leu Asp Gly Asp Thr

820 825 830

Gln Asp Phe Tyr Asn Trp Gln Arg Tyr Tyr Arg Ser Ala Leu Thr Glu

835 840 845

Lys Glu Thr Asp Asn Gln Leu Val Leu Ala His Pro Tyr Ala Glu Gln

850 855 860

Ile Leu Gly Thr Ile Lys Thr Leu Glu Gly Lys Gln Lys Asn Asn Lys

865 870 875 880

Leu Gly Asn Lys Ala Lys Gln Lys Ile Lys Asp Glu Leu Ile Asp Leu

885 890 895

Lys Arg Ala Lys Arg Arg Leu Leu Asp Arg Glu Gln Tyr Leu Arg Ala
900 905 910
Val Gln Ala Glu Asp Arg Ala Leu Trp Leu Met Ile Gln Glu Arg Gln
915 920 925
Lys Gln Lys Ala Glu His Glu Glu Ile Ala Phe Asp Gln Leu Asp Leu
930 935 940
Lys Asn Ile Thr Lys Ile Leu Thr Glu Ser Ile Asp Ala Arg Leu Arg
945 950 955 960
Ile Pro Asp Thr Lys Val Asp Ile Thr Asp Lys Leu Pro Leu Arg Arg

965 970 975
Tyr Gly Asp Leu Arg Arg Val Ala Lys Asp Arg Arg Leu Val Asn Leu
980 985 990
Ala Ser Tyr Tyr His Val Ala Gly Leu Ser Glu Ile Pro Tyr Asp Leu
995 1000 1005
Val Lys Lys Glu Leu Glu Glu Tyr Asp Arg Arg Arg Val Ala Phe Phe
1010 1015 1020
Glu His Val Tyr Gln Phe Glu Lys Glu Val Tyr Asp Arg Tyr Ala Ala
1025 1030 1035 1040

Glu Leu Arg Asn Glu Asn Pro Lys Gly Glu Ser Thr Tyr Phe Ser His
1045 1050 1055
Trp Glu Tyr Val Ala Val Ala Val Lys His Ser Ala Asp Thr His Phe
1060 1065 1070
Asn Glu Leu Phe Lys Glu Lys Val Met Gln Leu Arg Asn Lys Phe His
1075 1080 1085
His Asn Glu Phe Pro Tyr Phe Asp Trp Leu Leu Pro Glu Val Glu Lys
1090 1095 1100
Ala Ser Ala Ala Leu Tyr Ala Asp Arg Val Phe Asp Val Ala Glu Gly

1105 1110 1115 1120
Tyr Tyr Gln Lys Met Arg Lys Leu Met Arg Gln
1125 1130
<210> 107
<211> 1155

<212> PRT

<213> Porphyromonas gulae

<400> 107

Met Asp Asn Asn Ile Thr Val Glu Lys Thr Glu Leu Gly Leu Gly Ile
1 5 10 15

Thr Tyr Asn His Asp Lys Val Glu Asp Lys His Tyr Phe Gly Gly Phe
20 25 30

Phe Asn Leu Ala Gln Asn Asn Ile Asp Leu Val Ala Gln Glu Phe Lys
35 40 45

Lys Arg Leu Leu Ile Gln Gly Lys Asp Ser Ile Asn Ile Phe Ala Asn
50 55 60

Tyr Phe Ser Asp Gln Cys Ser Ile Thr Asn Leu Glu Arg Gly Ile Lys
65 70 75 80

Ile Leu Ala Glu Tyr Phe Pro Val Val Ser Tyr Ile Asp Leu Asp Glu
85 90 95

Lys Asn Lys Ser Lys Ser Ile Arg Glu His Leu Ile Leu Leu Leu Glu
100 105 110

Thr Ile Asn Asn Leu Arg Asn Tyr Tyr Thr His Tyr Tyr His Lys Lys
115 120 125

Ile Ile Ile Asp Gly Ser Leu Phe Pro Leu Leu Asp Thr Ile Leu Leu
130 135 140

Lys Val Val Leu Glu Ile Lys Lys Lys Lys Leu Lys Glu Asp Lys Thr
145 150 155 160

Lys Gln Leu Leu Lys Lys Gly Leu Glu Lys Glu Met Thr Ile Leu Phe
165 170 175

Asn Leu Met Lys Ala Glu Gln Lys Glu Lys Lys Ile Lys Gly Trp Asn
180 185 190

Ile Asp Glu Asn Ile Lys Gly Ala Val Leu Asn Arg Ala Phe Ser His
195 200 205

Leu Leu Tyr Asn Asp Glu Leu Ser Asp Tyr Arg Lys Ser Lys Tyr Asn
210 215 220

Thr Glu Asp Glu Thr Leu Lys Asp Thr Leu Thr Glu Ser Gly Ile Leu

225 230 235 240

Phe Leu Leu Ser Phe Phe Leu Asn Lys Lys Glu Gln Glu Gln Leu Lys

245 250 255

Ala Asn Ile Lys Gly Tyr Lys Gly Lys Ile Ala Ser Ile Pro Asp Glu

260 265 270

Glu Ile Thr Leu Lys Asn Asn Ser Leu Arg Asn Met Ala Thr His Trp

275 280 285

Thr Tyr Ser His Leu Thr Tyr Lys Gly Leu Lys His Arg Ile Lys Thr

290 295 300

Asp His Glu Lys Glu Thr Leu Leu Val Asn Met Val Asp Tyr Leu Ser

305 310 315 320

Lys Val Pro His Glu Ile Tyr Gln Asn Leu Ser Glu Gln Asn Lys Ser

325 330 335

Leu Phe Leu Glu Asp Ile Asn Glu Tyr Met Arg Asp Asn Glu Glu Asn

340 345 350

His Asp Ser Ser Glu Ala Ser Arg Val Ile His Pro Val Ile Arg Lys

355 360 365

Arg Tyr Glu Asn Lys Phe Ala Tyr Phe Ala Ile Arg Phe Leu Asp Glu

370 375 380

Phe Ala Glu Phe Pro Thr Leu Arg Phe Met Val Asn Val Gly Asn Tyr

385 390 395 400

Ile His Asp Asn Arg Lys Lys Asp Ile Gly Gly Thr Ser Leu Ile Thr

405 410 415

Asn Arg Thr Ile Lys Gln Gln Ile Asn Val Phe Gly Asn Leu Thr Glu

420 425 430

Ile His Lys Lys Lys Asn Asp Tyr Phe Glu Lys Glu Glu Asn Lys Glu

435 440 445

Lys Thr Leu Glu Trp Glu Leu Phe Pro Asn Pro Ser Tyr His Phe Gln

450 455 460

Lys Glu Asn Ile Pro Ile Phe Ile Asp Leu Glu Lys Ser Lys Glu Thr

465 470 475 480

Asn Asp Leu Ala Lys Glu Tyr Ala Lys Glu Lys Lys Lys Ile Phe Gly
485 490 495
Ser Ser Arg Lys Lys Gln Gln Asn Thr Ala Lys Lys Asn Arg Glu Thr
500 505 510
Ile Ile Asn Leu Val Phe Asp Lys Tyr Lys Thr Ser Asp Arg Lys Thr
515 520 525
Val Thr Phe Glu Gln Pro Thr Ala Leu Leu Ser Phe Asn Glu Leu Asn
530 535 540

Ser Phe Leu Tyr Ala Phe Leu Val Glu Asn Lys Thr Gly Lys Glu Leu
545 550 555 560
Glu Lys Ile Ile Ile Glu Lys Ile Ala Asn Gln Tyr Gln Ile Leu Lys
565 570 575
Asn Cys Ser Ser Thr Val Asp Lys Thr Asn Asp Asn Ile Pro Lys Ser
580 585 590
Ile Lys Lys Ile Val Asn Thr Thr Thr Asp Ser Phe Tyr Phe Glu Gly
595 600 605
Lys Lys Ile Asp Ile Glu Lys Leu Glu Lys Asp Ile Thr Ile Glu Ile

610 615 620
Glu Lys Thr Asn Glu Lys Leu Glu Thr Ile Lys Glu Asn Glu Glu Ser
625 630 635 640
Ala Gln Asn Tyr Lys Arg Asn Glu Arg Asn Thr Gln Lys Arg Lys Leu
645 650 655
Tyr Arg Lys Tyr Val Phe Phe Thr Asn Glu Ile Gly Ile Glu Ala Thr
660 665 670
Trp Ile Thr Asn Asp Ile Leu Arg Phe Leu Asp Asn Lys Glu Asn Trp
675 680 685

Lys Gly Tyr Gln His Ser Glu Leu Gln Lys Phe Ile Ser Gln Tyr Asp
690 695 700
Asn Tyr Lys Lys Glu Ala Leu Gly Leu Leu Glu Ser Glu Trp Asn Leu
705 710 715 720
Glu Ser Asp Ala Phe Phe Gly Gln Asn Leu Lys Arg Met Phe Gln Ser
725 730 735

```
Asn Ser Thr Phe Glu Thr Phe Tyr Lys Lys Tyr Leu Asp Asn Arg Lys
            740                 745                 750
Asn Thr Leu Glu Thr Tyr Leu Ser Ala Ile Glu Asn Leu Lys Thr Met

        755                 760                 765
Thr Asp Val Arg Pro Lys Val Leu Lys Lys Lys Trp Thr Glu Leu Phe
    770                 775                 780
Arg Phe Phe Asp Lys Lys Ile Tyr Leu Leu Ser Thr Ile Glu Thr Lys
785                 790                 795                 800
Ile Asn Glu Leu Ile Thr Lys Pro Ile Asn Leu Ser Arg Gly Ile Phe
                805                 810                 815
Glu Glu Lys Pro Thr Phe Ile Asn Gly Lys Asn Pro Asn Lys Glu Asn
            820                 825                 830

Asn Gln His Leu Phe Ala Asn Trp Phe Ile Tyr Ala Lys Lys Gln Thr
        835                 840                 845
Ile Leu Gln Asp Phe Tyr Asn Leu Pro Leu Glu Gln Pro Lys Ala Ile
    850                 855                 860
Thr Asn Leu Lys Lys His Lys Tyr Lys Leu Glu Arg Ser Ile Asn Asn
865                 870                 875                 880
Leu Lys Ile Glu Asp Ile Tyr Ile Lys Gln Met Val Asp Phe Leu Tyr
                885                 890                 895
Gln Lys Leu Phe Glu Gln Ser Phe Ile Gly Ser Leu Gln Asp Leu Tyr

            900                 905                 910
Thr Ser Lys Glu Lys Arg Glu Ile Glu Lys Gly Lys Ala Lys Asn Glu
        915                 920                 925
Gln Thr Pro Asp Glu Ser Phe Ile Trp Lys Lys Gln Val Glu Ile Asn
    930                 935                 940
Thr His Asn Gly Arg Ile Ile Ala Lys Thr Lys Ile Lys Asp Ile Gly
945                 950                 955                 960
Lys Phe Lys Asn Leu Leu Thr Asp Asn Lys Ile Ala His Leu Ile Ser
                965                 970                 975

Tyr Asp Asp Arg Ile Trp Asp Phe Ser Leu Asn Asn Asp Gly Asp Ile
```

980 985 990

Thr Lys Lys Leu Tyr Ser Ile Asn Thr Glu Leu Glu Ser Tyr Glu Thr

995 1000 1005

Ile Arg Arg Glu Lys Leu Leu Lys Gln Ile Gln Gln Phe Glu Gln Phe

1010 1015 1020

Leu Leu Glu Gln Glu Thr Glu Tyr Ser Ala Glu Arg Lys His Pro Glu

1025 1030 1035 1040

Lys Phe Glu Lys Asp Cys Asn Pro Asn Phe Lys Lys Tyr Ile Ile Glu

1045 1050 1055

Gly Val Leu Asn Lys Ile Ile Pro Asn His Glu Ile Glu Glu Ile Glu

1060 1065 1070

Ile Leu Lys Ser Lys Glu Asp Val Phe Lys Ile Asn Phe Ser Asp Ile

1075 1080 1085

Leu Ile Leu Asn Asn Asp Asn Ile Lys Lys Gly Tyr Leu Leu Ile Met

1090 1095 1100

Ile Arg Asn Lys Phe Ala His Asn Gln Leu Ile Asp Lys Asn Leu Phe

1105 1110 1115 1120

Asn Phe Ser Leu Gln Leu Tyr Ser Lys Asn Glu Asn Glu Asn Phe Ser

1125 1130 1135

Glu Tyr Leu Asn Lys Val Cys Gln Asn Ile Ile Gln Glu Phe Lys Glu

1140 1145 1150

Lys Leu Lys

1155

<210> 108

<211> 1175

<212> PRT

<213> Porphyromonas gulae

<400> 108

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu

1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn

20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45
Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
50 55 60
Lys Asn Phe Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
100 105 110
Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125
Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140
Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Glu Ser Ser Glu Leu
145 150 155 160
Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175
Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
180 185 190
Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205
Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
210 215 220
Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255
Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
260 265 270
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
290 295 300
Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Arg Phe

305 310 315 320
Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
325 330 335
Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
340 345 350
Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365
Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385 390 395 400
Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
405 410 415
Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
420 425 430
Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
435 440 445
Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460
Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465 470 475 480
Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
485 490 495
Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
500 505 510
Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
515 520 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Lys Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu

565 570 575

Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Lys Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg

705 710 715 720

Ser Asp Arg Met Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys

725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro

785 790 795 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Glu Arg Ala Glu Glu Trp Glu Arg Gly

820 825 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg

835 840 845

Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn

850 855 860

Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala

865 870 875 880

Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys

885 890 895

Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe

900 905 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln

915 920 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys

930 935 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Pro

945 950 955 960

Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro

965 970 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val

980 985 990

His Lys Glu Glu Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asp

995 1000 1005

Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg

1010 1015 1020

Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Gly Leu Ala Met

1025 1030 1035 1040

Glu Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu Leu Ala Lys
1045 1050 1055
Tyr Gln Thr Ala Arg Val Cys Val Phe Glu Leu Thr Leu Arg Leu Glu
1060 1065 1070
Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro Asp Glu Ser Phe Arg
1075 1080 1085
Lys Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Ala Lys Trp Pro Glu
1090 1095 1100

Leu His Gly Lys Val Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser
1105 1110 1115 1120
His Asn Gln Tyr Pro Met Tyr Asp Glu Ala Val Phe Ser Ser Ile Arg
1125 1130 1135
Lys Tyr Asp Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu
1140 1145 1150
Asn Ile Ala His Arg Leu Ser Glu Glu Val Lys Gln Ala Lys Glu Thr
1155 1160 1165
Val Glu Arg Ile Ile Gln Val

1170 1175

<210> 109
<211> 1145
<212> PRT
<213> Porphyromonas gulae
<400> 109
Met Glu Thr Gln Thr Ile Gly His Gly Ile Ala Tyr Asp His Ser Lys
1 5 10 15
Ile Gln Asp Lys His Phe Phe Gly Gly Phe Leu Asn Leu Ala Glu Asn
20 25 30
Asn Ile Lys Ala Val Leu Lys Ala Phe Ser Glu Lys Phe Asn Val Gly
35 40 45
Asn Val Asp Val Lys Gln Phe Ala Asp Val Ser Leu Lys Asp Asn Leu

50 55 60
Pro Asp Asn Asp Phe Gln Lys Arg Val Ser Phe Leu Lys Met Tyr Phe

65 70 75 80

Pro Val Val Asp Phe Ile Asn Ile Pro Asn Asn Arg Ala Lys Phe Arg

85 90 95

Ser Asp Leu Thr Thr Leu Phe Lys Ser Val Asp Gln Leu Arg Asn Phe

100 105 110

Tyr Thr His Tyr Tyr His Lys Pro Leu Asp Phe Asp Ala Ser Leu Phe

115 120 125

Ile Leu Leu Asp Asp Ile Phe Ala Arg Thr Ala Lys Glu Val Arg Asp

130 135 140

Gln Lys Met Lys Asp Asp Lys Thr Arg Gln Leu Leu Ser Lys Ser Leu

145 150 155 160

Ser Glu Glu Leu Gln Lys Gly Tyr Glu Leu Gln Leu Glu Arg Leu Lys

165 170 175

Glu Leu Asn Arg Leu Gly Lys Lys Val Asn Ile His Asp Gln Leu Gly

180 185 190

Ile Lys Asn Gly Val Leu Asn Asn Ala Phe Asn His Leu Ile Tyr Lys

195 200 205

Asp Gly Glu Ser Phe Lys Thr Lys Leu Thr Tyr Ser Ser Ala Leu Thr

210 215 220

Ser Phe Glu Ser Ala Glu Asn Gly Ile Glu Ile Ser Gln Ser Gly Leu

225 230 235 240

Leu Phe Leu Leu Ser Met Phe Leu Lys Arg Lys Glu Ile Glu Asp Leu

245 250 255

Lys Asn Arg Asn Lys Gly Phe Lys Ala Lys Val Val Ile Asp Glu Asp

260 265 270

Gly Lys Val Asn Gly Leu Lys Phe Met Ala Thr His Trp Val Phe Ser

275 280 285

Tyr Leu Cys Phe Lys Gly Leu Lys Ser Lys Leu Ser Thr Glu Phe His

290 295 300

Glu Glu Thr Leu Leu Ile Gln Ile Ile Asp Glu Leu Ser Lys Val Pro

305 310 315 320

Asp Glu Leu Tyr Cys Ala Phe Asp Lys Glu Thr Arg Asp Lys Phe Ile

325 330 335

Glu Asp Ile Asn Glu Tyr Val Lys Glu Gly His Gln Asp Phe Ser Leu

340 345 350

Glu Asp Ala Lys Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asn

355 360 365

Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Val Lys Phe

370 375 380

Pro Ser Leu Arg Phe Gln Val His Val Gly Asn Tyr Val His Asp Arg

385 390 395 400

Arg Ile Lys Asn Ile Asp Gly Thr Thr Phe Glu Thr Glu Arg Val Val

405 410 415

Lys Asp Arg Ile Lys Val Phe Gly Arg Leu Ser Glu Ile Ser Ser Tyr

420 425 430

Lys Ala Gln Tyr Leu Ser Ser Val Ser Asp Lys His Asp Glu Thr Gly

435 440 445

Trp Glu Ile Phe Pro Asn Pro Ser Tyr Val Phe Ile Asn Asn Asn Ile

450 455 460

Pro Ile His Ile Ser Val Asp Thr Ser Phe Lys Lys Glu Ile Ala Asp

465 470 475 480

Phe Lys Lys Leu Arg Arg Ala Gln Val Pro Asp Glu Leu Lys Ile Arg

485 490 495

Gly Ala Glu Lys Lys Arg Lys Phe Glu Ile Thr Gln Met Ile Gly Ser

500 505 510

Lys Ser Val Leu Asn Gln Glu Glu Pro Ile Ala Leu Leu Ser Leu Asn

515 520 525

Glu Ile Pro Ala Leu Leu Tyr Glu Ile Leu Ile Asn Gly Lys Glu Pro

530 535 540

Ala Glu Ile Glu Arg Ile Ile Lys Asp Lys Leu Asn Glu Arg Gln Asp

545 550 555 560

Val Ile Lys Asn Tyr Asn Pro Glu Asn Trp Leu Pro Ala Ser Gln Ile

565 570 575

Ser Arg Arg Leu Arg Ser Asn Lys Gly Glu Arg Ile Ile Asn Thr Asp
580 585 590
Lys Leu Leu Gln Leu Val Thr Lys Glu Leu Leu Val Thr Glu Gln Lys
595 600 605
Leu Lys Ile Ile Ser Asp Asn Arg Glu Ala Leu Lys Gln Lys Lys Glu
610 615 620
Gly Lys Tyr Ile Arg Lys Phe Ile Phe Thr Asn Ser Glu Leu Gly Arg

625 630 635 640
Glu Ala Ile Trp Leu Ala Asp Asp Ile Lys Arg Phe Met Pro Ala Asp
645 650 655
Val Arg Lys Glu Trp Lys Gly Tyr Gln His Ser Gln Leu Gln Gln Ser
660 665 670
Leu Ala Phe Tyr Asn Ser Arg Pro Lys Glu Ala Leu Ala Ile Leu Glu
675 680 685
Ser Ser Trp Asn Leu Lys Asp Glu Lys Ile Ile Trp Asn Glu Trp Ile
690 695 700

Leu Lys Ser Phe Thr Gln Asn Lys Phe Phe Asp Ala Phe Tyr Asn Glu
705 710 715 720
Tyr Leu Lys Gly Arg Lys Lys Tyr Phe Ala Phe Leu Ser Glu His Ile
725 730 735
Val Gln Tyr Thr Ser Asn Ala Lys Asn Leu Gln Lys Phe Ile Lys Gln
740 745 750
Gln Met Pro Lys Asp Leu Phe Glu Lys Arg His Tyr Ile Ile Glu Asp
755 760 765
Leu Gln Thr Glu Lys Asn Lys Ile Leu Ser Lys Pro Phe Ile Phe Pro

770 775 780
Arg Gly Ile Phe Asp Lys Lys Pro Thr Phe Ile Lys Gly Val Lys Val
785 790 795 800
Glu Asp Ser Pro Glu Ser Phe Ala Asn Trp Tyr Gln Tyr Gly Tyr Gln
805 810 815
Lys Asp His Gln Phe Gln Lys Phe Tyr Asp Trp Lys Arg Asp Tyr Ser
820 825 830

Asp Val Phe Leu Glu His Leu Gly Lys Pro Phe Ile Asn Asn Gly Asp
835 840 845

Arg Arg Thr Leu Gly Met Glu Glu Leu Lys Glu Arg Ile Ile Ile Lys
850 855 860
Gln Asp Leu Lys Ile Lys Lys Ile Lys Ile Gln Asp Leu Phe Leu Arg
865 870 875 880
Leu Ile Ala Glu Asn Leu Phe Gln Lys Val Phe Lys Tyr Ser Ala Lys
885 890 895
Leu Pro Leu Ser Asp Phe Tyr Leu Thr Gln Glu Glu Arg Met Glu Lys
900 905 910
Glu Asn Met Ala Ala Leu Gln Asn Val Arg Glu Glu Gly Asp Lys Ser

915 920 925
Pro Asn Ile Ile Lys Asp Asn Phe Ile Trp Ser Lys Met Ile Pro Tyr
930 935 940
Lys Lys Gly Gln Ile Ile Glu Asn Ala Val Lys Leu Lys Asp Ile Gly
945 950 955 960
Lys Leu Asn Val Leu Ser Leu Asp Asp Lys Val Gln Thr Leu Leu Ser
965 970 975
Tyr Asp Asp Ala Lys Pro Trp Ser Lys Ile Ala Leu Glu Asn Glu Phe
980 985 990

Ser Ile Gly Glu Asn Ser Tyr Glu Val Ile Arg Arg Glu Lys Leu Phe
995 1000 1005
Lys Glu Ile Gln Gln Phe Glu Ser Glu Ile Leu Phe Arg Ser Gly Trp
1010 1015 1020
Asp Gly Ile Asn His Pro Ala Gln Leu Glu Asp Asn Arg Asn Pro Lys
1025 1030 1035 1040
Phe Lys Met Tyr Ile Val Asn Gly Ile Leu Arg Lys Ser Ala Gly Leu
1045 1050 1055
Tyr Ser Gln Gly Glu Asp Ile Trp Phe Glu Tyr Asn Ala Asp Phe Asn

1060 1065 1070
Asn Leu Asp Ala Asp Val Leu Glu Thr Lys Ser Glu Leu Val Gln Leu

1075 1080 1085

Ala Phe Leu Val Thr Ala Ile Arg Asn Lys Phe Ala His Asn Gln Leu

1090 1095 1100

Pro Ala Lys Glu Phe Tyr Phe Tyr Ile Arg Ala Lys Tyr Gly Phe Ala

1105 1110 1115 1120

Asp Glu Pro Ser Val Ala Leu Val Tyr Leu Asn Phe Thr Lys Tyr Ala

1125 1130 1135

Ile Asn Glu Phe Lys Lys Val Met Ile

1140 1145

<210> 110

<211> 948

<212> PRT

<213> Porphyromonas gulae

<400> 110

Met Phe Phe Ser Phe His Asn Ala Gln Arg Val Ile Phe Lys His Leu

1 5 10 15

Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu Asp Tyr Lys

20 25 30

Ala His Phe Thr Val Asn Leu Thr Arg Asp Phe Ala His Leu Asn Arg

35 40 45

Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asn Arg Tyr Arg Phe

50 55 60

Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe Phe Thr Asn

65 70 75 80

Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys Lys Val Ser

85 90 95

Gly Phe Lys Ala Ser His Lys Gln Arg Glu Lys Met Thr Thr Glu Val

100 105 110

Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu Glu Ser Arg

115 120 125

Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu Leu Ser Arg

130 135 140

Cys Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Glu Asn Lys Lys His
145 150 155 160
Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu Glu Glu Gln
165 170 175
Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg Phe Pro Tyr
180 185 190

Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys Ser Ile Arg
195 200 205
Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr Asp Lys Lys
210 215 220
Ile Gly Asp Glu Gln Glu Lys Arg His Leu Thr Arg Thr Leu Leu Ser
225 230 235 240
Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro Gln Glu Trp
245 250 255
Lys Ala Leu Thr Lys Asp Leu Asp Tyr Lys Glu Thr Ser Asn Gln Pro

260 265 270
Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp Asn Lys Ile
275 280 285
Gly Phe Arg Leu Gly Thr Ser Lys Glu Leu Tyr Pro Ser Leu Glu Ile
290 295 300
Lys Asp Gly Ala Asn Arg Ile Ala Lys Tyr Pro Tyr Asn Ser Gly Phe
305 310 315 320
Val Ala His Ala Phe Ile Ser Val His Glu Leu Leu Pro Leu Met Phe
325 330 335

Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys Glu Thr Val
340 345 350
Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu Arg Ile Asn
355 360 365
Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu Pro Leu Gly
370 375 380
Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys Gln Pro Asp
385 390 395 400

Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile Ala Glu Thr

405 410 415

Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser Ser Pro Lys

420 425 430

Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val Leu Ala Asp

435 440 445

Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Ala Tyr Asp Ala

450 455 460

Gln Asn Gln Pro Ile Lys Ser Ser Lys Ala Asn Ser Thr Glu Phe Trp

465 470 475 480

Phe Ile Arg Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys Asn Arg Leu

485 490 495

Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr Asn Pro His

500 505 510

Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn Leu Val Asp

515 520 525

Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu Glu Ala Ile

530 535 540

Lys Asn Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu Leu Lys Ile

545 550 555 560

Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu Gln Gly Gly

565 570 575

Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg Glu Thr Leu

580 585 590

Ser Glu Asp Leu Met Leu Ser Lys Pro Ile Arg Lys Glu Ile Lys Lys

595 600 605

His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu Tyr Phe Lys

610 615 620

Glu Lys Tyr Gln Asp Lys His Gln Ser Phe Tyr Asn Leu Ser Tyr Lys

625 630 635 640

Leu Glu Ala Lys Ala Pro Leu Leu Lys Arg Glu Glu His Tyr Glu Tyr

645 650 655

Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Ser Gln Arg Leu Glu

660 665 670

Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr Lys Arg Trp

675 680 685

Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln Asp Val Met

690 695 700

Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe Lys Glu Leu

705 710 715 720

Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala Val Asn Val

725 730 735

Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr Leu Pro Met

740 745 750

Val Leu Pro Val Lys Val Tyr Pro Ala Thr Ala Phe Gly Glu Val Gln

755 760 765

Tyr His Lys Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu Glu His Thr

770 775 780

Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys Asp Arg Arg

785 790 795 800

Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp Thr Gln Lys

805 810 815

His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu Ile Tyr Gln

820 825 830

Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Ser Leu Glu Glu Lys

835 840 845

Leu Leu Asn Lys His Thr Ser Leu Ser Ser Leu Glu Asn Glu Phe Arg

850 855 860

Ala Leu Leu Glu Glu Trp Lys Lys Glu Tyr Ala Ala Ser Ser Met Val

865 870 875 880

Thr Asp Glu His Ile Ala Phe Ile Ala Ser Val Arg Asn Ala Phe Cys

885 890 895

His Asn Gln Tyr Pro Phe Tyr Lys Glu Ala Leu His Ala Pro Ile Pro

900 905 910

Leu Phe Thr Val Ala Gln Pro Thr Thr Glu Glu Lys Asp Gly Leu Gly
915 920 925
Ile Ala Glu Ala Leu Leu Lys Val Leu Arg Glu Tyr Cys Glu Ile Val
930 935 940
Lys Ser Gln Ile
945

<210> 111
<211> 1175
<212> PRT
<213> Porphyromonas gulae
<400> 111

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15
Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn

20 25 30
Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45
Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
50 55 60
Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65 70 75 80
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
100 105 110
Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125
Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140
Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Lys Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp

210 215 220

Arg Glu Glu Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

Pro Lys Leu Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr

325 330 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu

385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln Leu Leu Trp Pro Ser

450 455 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp

465 470 475 480

Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro

485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Ala

500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val

515 520 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu

565 570 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
675 680 685
Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
690 695 700
Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720
Ser Asp Arg Glu Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu

740 745 750
Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765
Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
770 775 780
Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785 790 795 800
Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
805 810 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
820 825 830
Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
835 840 845
Arg Ile Glu Asp Ala Phe Val Gly Ile Glu Tyr Ala Ser Trp Glu Asn
850 855 860
Lys Lys Lys Ile Glu Gln Leu Leu Gln Asp Leu Ser Leu Trp Glu Thr
865 870 875 880
Phe Glu Ser Lys Leu Lys Val Lys Ala Asp Lys Ile Asn Ile Ala Lys

885 890 895
Leu Lys Lys Glu Ile Leu Glu Ala Lys Glu His Pro Tyr His Asp Phe
900 905 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
915 920 925
Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
930 935 940
Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945 950 955 960

Asp Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn His Val Lys Pro
965 970 975
Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
980 985 990
His Lys Glu Glu Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asp
995 1000 1005
Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg
1010 1015 1020
Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Ala Leu Ala Met

1025 1030 1035 1040
Glu Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu Leu Ala Lys
1045 1050 1055
Tyr Gln Thr Ala Arg Val Cys Ala Phe Glu Gln Thr Leu Glu Leu Glu
1060 1065 1070
Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro Asp Glu Ser Phe Arg
1075 1080 1085
Glu Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Asp Lys Trp Pro Asp
1090 1095 1100

Leu Gln Arg Glu Val Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser
1105 1110 1115 1120
His Asn Gln Tyr Pro Met Tyr Asp Glu Thr Ile Phe Ser Ser Ile Arg
1125 1130 1135
Lys Tyr Asp Pro Ser Ser Leu Asp Ala Ile Glu Glu Arg Met Gly Leu
1140 1145 1150
Asn Ile Ala His Arg Leu Ser Glu Glu Val Lys Leu Ala Lys Glu Met
1155 1160 1165

Val Glu Arg Ile Ile Gln Ala

1170 1175

<210> 112

<211> 1175

<212> PRT

<213> Porphyromonas gulae

<400> 112

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15

Lys Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
50 55 60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65 70 75 80

Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
85 90 95

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145 150 155 160

Pro Met Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Cys Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp

210 215 220

Arg Glu Gly Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu

225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly

245 250 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe

260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Cys Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Ala

325 330 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe

340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser

355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys

370 375 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu

385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu

405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Cys Phe Glu Thr Gly Asp

420 425 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser

450 455 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp

465 470 475 480

Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro

485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val

500 505 510

Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val

515 520 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp

530 535 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu

545 550 555 560

Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Lys His Lys Asp Met Glu

565 570 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu

610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly

625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu

645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe

675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr

690 695 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720
Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735
Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
740 745 750
Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765
Gly Leu Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys

770 775 780
Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785 790 795 800
Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
805 810 815
Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
820 825 830
Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
835 840 845

Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Asn Ala Ser Arg Glu Asn
850 855 860
Lys Lys Lys Ile Glu Gln Leu Leu Gln Asp Leu Ser Leu Trp Glu Thr
865 870 875 880
Phe Glu Ser Lys Leu Lys Val Lys Ala Asp Lys Ile Asn Ile Ala Lys
885 890 895
Leu Lys Lys Glu Ile Leu Glu Ala Lys Glu His Pro Tyr Leu Asp Phe
900 905 910
Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln

915 920 925
Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
930 935 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945 950 955 960
Asp Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn His Val Lys Pro
965 970 975
Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
980 985 990

His Lys Glu Gln Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asp
995 1000 1005
Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg
1010 1015 1020
Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Ala Leu Ala Met
1025 1030 1035 1040
Glu Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu Leu Ala Lys
1045 1050 1055
Tyr Gln Thr Ala Arg Val Cys Ala Phe Glu Gln Thr Leu Glu Leu Glu

1060 1065 1070
Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro Asp Glu Asn Phe Arg
1075 1080 1085
Lys Met Leu Glu Ser Trp Ser Asp Pro Leu Leu Asp Lys Trp Pro Asp
1090 1095 1100
Leu His Arg Lys Val Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser
1105 1110 1115 1120
His Asn Gln Tyr Pro Met Tyr Asp Glu Ala Val Phe Ser Ser Ile Arg
1125 1130 1135

Lys Tyr Asp Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu
1140 1145 1150
Asn Ile Ala His Arg Leu Ser Glu Glu Val Lys Gln Ala Lys Glu Met
1155 1160 1165
Ala Glu Arg Ile Ile Gln Ala
1170 1175
<210> 113
<211> 1213

<212> PRT

<213> Porphyromonas gulae

<400> 113

Met Ser Ser Lys Asn Glu Ser Tyr Asn Lys Gln Lys Thr Phe Asn His
1 5 10 15

Tyr Lys Gln Glu Asp Lys Tyr Phe Phe Gly Gly Phe Leu Asn Asn Ala
20 25 30

Asp Asp Asn Leu Arg Gln Val Gly Lys Glu Phe Lys Thr Arg Ile Asn
35 40 45

Phe Asn His Asn Asn Asn Glu Leu Ala Ser Val Phe Lys Asp Tyr Phe
50 55 60

Asn Lys Glu Lys Ser Val Ala Lys Arg Glu His Ala Leu Asn Leu Leu
65 70 75 80

Ser Asn Tyr Phe Pro Val Leu Glu Arg Ile Gln Lys His Thr Asn His

85 90 95

Asn Phe Glu Gln Thr Arg Glu Ile Phe Glu Leu Leu Leu Asp Thr Ile
100 105 110

Lys Lys Leu Arg Asp Tyr Tyr Thr His His Tyr His Lys Pro Ile Thr
115 120 125

Ile Asn Pro Lys Val Tyr Asp Phe Leu Asp Asp Thr Leu Leu Asp Val
130 135 140

Leu Ile Thr Ile Lys Lys Lys Lys Val Lys Asn Asp Thr Ser Arg Glu
145 150 155 160

Leu Leu Lys Glu Lys Phe Arg Pro Glu Leu Thr Gln Leu Lys Asn Gln

165 170 175

Lys Arg Glu Glu Leu Ile Lys Lys Gly Lys Lys Leu Leu Glu Glu Asn
180 185 190

Leu Glu Asn Ala Val Phe Asn His Cys Leu Arg Pro Phe Leu Glu Glu
195 200 205

Asn Lys Thr Asp Asp Lys Gln Asn Lys Thr Val Ser Leu Arg Lys Tyr
210 215 220

Arg Lys Ser Lys Pro Asn Glu Glu Thr Ser Ile Thr Leu Thr Gln Ser

225 230 235 240

Gly Leu Val Phe Leu Ile Ser Phe Phe Leu His Arg Lys Glu Phe Gln

245 250 255

Val Phe Thr Ser Gly Leu Glu Gly Phe Lys Ala Lys Val Asn Thr Ile

260 265 270

Lys Glu Glu Glu Ile Ser Leu Asn Lys Asn Asn Ile Val Tyr Met Ile

275 280 285

Thr His Trp Ser Tyr Ser Tyr Tyr Asn Phe Lys Gly Leu Lys His Arg

290 295 300

Ile Lys Thr Asp Gln Gly Val Ser Thr Leu Glu Gln Asn Asn Thr Thr

305 310 315 320

His Ser Leu Thr Asn Thr Asn Thr Lys Glu Ala Leu Leu Thr Gln Ile

325 330 335

Val Asp Tyr Leu Ser Lys Val Pro Asn Glu Ile Tyr Glu Thr Leu Ser

340 345 350

Glu Lys Gln Gln Lys Glu Phe Glu Glu Asp Ile Asn Glu Tyr Met Arg

355 360 365

Glu Asn Pro Glu Asn Glu Asp Ser Thr Phe Ser Ser Ile Val Ser His

370 375 380

Lys Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Met

385 390 395 400

Arg Phe Leu Asp Glu Tyr Ala Glu Leu Pro Thr Leu Arg Phe Met Val

405 410 415

Asn Phe Gly Asp Tyr Ile Lys Asp Arg Gln Lys Lys Ile Leu Glu Ser

420 425 430

Ile Gln Phe Asp Ser Glu Arg Ile Ile Lys Lys Glu Ile His Leu Phe

435 440 445

Glu Lys Leu Gly Leu Val Thr Glu Tyr Lys Lys Asn Val Tyr Leu Lys

450 455 460

Glu Thr Ser Asn Ile Asp Leu Ser Arg Phe Pro Leu Phe Pro Ser Pro

465 470 475 480

Ser Tyr Val Met Ala Asn Asn Asn Ile Pro Phe Tyr Ile Asp Ser Arg

485 490 495

Ser Asn Asn Leu Asp Glu Tyr Leu Asn Gln Lys Lys Lys Ala Gln Ser

500 505 510

Gln Asn Arg Lys Arg Asn Leu Thr Phe Glu Lys Tyr Asn Lys Glu Gln

515 520 525

Ser Lys Asp Ala Ile Ile Ala Met Leu Gln Lys Glu Ile Gly Val Lys

530 535 540

Asp Leu Gln Gln Arg Ser Thr Ile Gly Leu Leu Ser Cys Asn Glu Leu

545 550 555 560

Pro Ser Met Leu Tyr Glu Val Ile Val Lys Asp Ile Lys Gly Ala Glu

565 570 575

Leu Glu Asn Lys Ile Ala Gln Lys Ile Arg Glu Gln Tyr Gln Ser Ile

580 585 590

Arg Asp Phe Thr Leu Asp Ser Pro Gln Lys Asp Asn Ile Pro Thr Thr

595 600 605

Leu Thr Lys Thr Ile Ser Thr Asp Thr Ser Val Thr Phe Glu Asn Gln

610 615 620

Pro Ile Asp Ile Pro Arg Leu Lys Asn Ala Leu Gln Lys Glu Leu Thr

625 630 635 640

Leu Thr Gln Glu Lys Leu Leu Asn Val Lys Gln His Glu Ile Glu Val

645 650 655

Asp Asn Tyr Asn Arg Asn Lys Asn Thr Tyr Lys Phe Lys Asn Gln Pro

660 665 670

Lys Asp Lys Val Asp Asp Asn Lys Leu Gln Arg Lys Tyr Val Phe Tyr

675 680 685

Arg Asn Glu Ile Gly Gln Glu Ala Asn Trp Leu Ala Ser Asp Leu Ile

690 695 700

His Phe Met Lys Asn Lys Ser Leu Trp Lys Gly Tyr Met His Asn Glu

705 710 715 720

Leu Gln Ser Phe Leu Ala Phe Phe Glu Asp Lys Lys Asn Asp Cys Ile

725 730 735

Ala Leu Leu Glu Thr Val Phe Asn Leu Lys Glu Asp Cys Ile Leu Thr
740 745 750

Lys Asp Leu Lys Asn Leu Phe Leu Lys His Gly Asn Phe Ile Asp Phe
755 760 765

Tyr Lys Glu Tyr Leu Lys Leu Lys Glu Asp Phe Leu Asn Thr Glu Ser
770 775 780

Thr Phe Leu Glu Asn Gly Phe Ile Gly Leu Pro Pro Lys Ile Leu Lys
785 790 795 800

Lys Glu Leu Ser Lys Arg Leu Asn Tyr Ile Phe Ile Val Phe Gln Lys

805 810 815

Arg Gln Phe Ile Ile Lys Glu Leu Glu Glu Lys Lys Asn Asn Leu Tyr
820 825 830

Ala Asp Ala Ile Asn Leu Ser Arg Gly Ile Phe Asp Glu Lys Pro Thr
835 840 845

Met Ile Pro Phe Lys Lys Pro Asn Pro Asp Glu Phe Ala Ser Trp Phe
850 855 860

Val Ala Ser Tyr Gln Tyr Asn Asn Tyr Gln Ser Phe Tyr Glu Leu Thr
865 870 875 880

Pro Asp Lys Ile Glu Asn Asp Lys Lys Lys Lys Tyr Lys Asn Leu Arg

885 890 895

Ala Ile Asn Lys Val Lys Ile Gln Asp Tyr Tyr Leu Lys Leu Met Val
900 905 910

Asp Thr Leu Tyr Gln Asp Leu Phe Asn Gln Pro Leu Asp Lys Ser Leu
915 920 925

Ser Asp Phe Tyr Val Ser Lys Thr Asp Arg Glu Lys Ile Lys Ala Asp
930 935 940

Ala Lys Ala Tyr Gln Lys Arg Asn Asp Ser Phe Leu Trp Asn Lys Val

945 950 955 960

Ile His Leu Ser Leu Gln Asn Asn Arg Ile Thr Ala Asn Pro Lys Leu

965 970 975

Lys Asp Ile Gly Lys Tyr Lys Arg Ala Leu Gln Asp Glu Lys Ile Ala
980 985 990
Thr Leu Leu Thr Tyr Asp Asp Arg Thr Trp Thr Tyr Ala Leu Gln Lys
995 1000 1005
Pro Glu Lys Glu Asn Glu Asn Asp Tyr Lys Glu Leu His Tyr Thr Ala
1010 1015 1020

Leu Asn Met Glu Leu Gln Glu Tyr Glu Lys Val Arg Ser Lys Lys Leu
1025 1030 1035 1040
Leu Lys Gln Val Gln Glu Leu Glu Lys Gln Ile Leu Asp Lys Phe Tyr
1045 1050 1055
Asp Phe Ser Asn Asn Ala Thr His Pro Glu Asp Leu Glu Ile Glu Asp
1060 1065 1070
Lys Lys Gly Lys Arg His Pro Asn Phe Lys Leu Tyr Ile Thr Lys Ala
1075 1080 1085
Leu Leu Lys Asn Glu Ser Glu Ile Ile Asn Leu Glu Asn Ile Asp Ile

1090 1095 1100
Glu Ile Leu Ile Lys Tyr Tyr Asp Tyr Asn Thr Glu Lys Leu Lys Glu
1105 1110 1115 1120
Lys Ile Lys Asn Met Asp Glu Asp Glu Lys Ala Lys Ile Val Asn Thr
1125 1130 1135
Lys Glu Asn Tyr Asn Lys Ile Thr Asn Val Leu Ile Lys Lys Ala Leu
1140 1145 1150
Val Leu Ile Ile Ile Arg Asn Lys Met Ala His Asn Gln Tyr Pro Pro
1155 1160 1165

Lys Phe Ile Tyr Asp Leu Ala Thr Arg Phe Val Pro Lys Lys Glu Glu
1170 1175 1180
Glu Tyr Phe Ala Cys Tyr Phe Asn Arg Val Phe Glu Thr Ile Thr Thr
1185 1190 1195 1200
Glu Leu Trp Glu Asn Lys Lys Lys Ala Lys Glu Ile Val
1205 1210
<210> 114
<211> 1119

<212> PRT

<213> Porphyromonas gulae

<400> 114

Met Thr Glu Gln Asn Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu

1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn

20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys

35 40 45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp

50 55 60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu

65 70 75 80

Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys

85 90 95

Leu Phe Glu Asn Lys Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu

115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp

130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu

145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp

165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp

195 200 205

Arg Cys Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp

210 215 220

Arg Glu Gly Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240
Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255
Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe
260 265 270
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285
Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys

290 295 300
Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305 310 315 320
Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
325 330 335
Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
340 345 350
Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380
Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385 390 395 400
Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
405 410 415
Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
420 425 430
Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly

435 440 445
Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
450 455 460
Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
485 490 495
Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
515 520 525
Tyr Ala Val Tyr Asp Ala Phe Ala Arg Gly Glu Ile Asp Thr Leu Asp
530 535 540
Arg Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545 550 555 560
Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
565 570 575
Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His

580 585 590
Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
595 600 605
Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
610 615 620
Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625 630 635 640
Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
660 665 670
Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
675 680 685
Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
690 695 700
Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720
Ser Asp Arg Glu Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys

725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met

755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys

770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro

785 790 795 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly

820 825 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala

835 840 845

Lys Glu His Pro Tyr Leu Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg

850 855 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys

865 870 875 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr

885 890 895

Leu Tyr Leu Lys Asp Ile Arg Thr Glu Val Gln Glu Gln Gly Ser Leu

900 905 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr

915 920 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala

930 935 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe

965 970 975

Val Asp Thr Gly Gly Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu

980 985 990

Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala

995 1000 1005

Phe Glu Gln Thr Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Cys Pro

1010 1015 1020

His Leu Pro Asp Lys Asn Phe Arg Lys Met Leu Glu Ser Trp Ser Asp

1025 1030 1035 1040

Pro Leu Leu Asp Lys Trp Pro Asp Leu Gln Arg Glu Val Trp Leu Leu

1045 1050 1055

Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp

1060 1065 1070

Glu Ala Val Phe Ser Ser Ile Arg Lys Tyr Asp Pro Ser Ser Pro Asp

1075 1080 1085

Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu Ser Glu

1090 1095 1100

Glu Val Lys Gln Ala Lys Glu Met Ala Glu Arg Ile Ile Gln Ala

1105 1110 1115

<210> 115

<211> 1135

<212> PRT

<213> Porphyromonas gulae

<400> 115

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val

1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu

20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys

35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu

50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr

65 70 75 80

Leu Pro Phe Leu His Tyr Phe Asp Pro Asp Ser Gln Ile Glu Lys Asp
85 90 95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145 150 155 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
165 170 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
180 185 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
195 200 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
210 215 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225 230 235 240

Ile Arg His Pro His Asp Arg Leu Glu Ser Ser Asn Thr Lys Glu Ala
245 250 255

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
290 295 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe

325 330 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu

340 345 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly

355 360 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu

370 375 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu

385 390 395 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile

405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro

420 425 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met

435 440 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu

450 455 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Gly Phe Leu Arg Lys

465 470 475 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser

485 490 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro

500 505 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr

515 520 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu

530 535 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp

545 550 555 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr

565 570 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe

580 585 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu

595 600 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu

610 615 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg

625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala

645 650 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe

660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr

675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr

690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe

705 710 715 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg

725 730 735

Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln

740 745 750

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met

755 760 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe

770 775 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr

785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile

805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu

820 825 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
835 840 845
Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
850 855 860
Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865 870 875 880
Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp

885 890 895
Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Lys Glu Asn
900 905 910
Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
915 920 925
Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
930 935 940
Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945 950 955 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
965 970 975
Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
980 985 990
Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg
995 1000 1005
Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys Ser
1010 1015 1020
Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys Gly Cys

1025 1030 1035 1040
Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg Asn Lys Ala
1045 1050 1055
Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro Leu Ile Tyr Arg
1060 1065 1070
Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly Ser Ser Ala Lys Asp
1075 1080 1085

Leu Pro Glu Gly Ser Ser Leu Val Asp Ser Leu Trp Lys Lys Tyr Glu
1090 1095 1100

Met Ile Ile Arg Lys Ile Leu Pro Ile Leu Asp Pro Glu Asn Arg Phe
1105 1110 1115 1120
Phe Gly Lys Leu Leu Asn Asn Met Ser Gln Pro Ile Asn Asp Leu
1125 1130 1135

<210> 116
<211> 950
<212> PRT
<213> Porphyromonas gulae
<400> 116
Met Phe Phe Ser Phe His Asn Ala Gln Arg Val Ile Phe Lys His Leu
1 5 10 15
Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu Asp Tyr Lys

20 25 30
Ala His Phe Thr Val Asn Leu Thr Arg Asp Phe Ala His Leu Asn Arg
35 40 45
Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asn Arg Tyr Arg Phe
50 55 60
Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe Phe Thr Asn
65 70 75 80
Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys Lys Val Ser
85 90 95

Gly Phe Lys Ala Ser His Lys Gln Ser Glu Lys Met Thr Thr Glu Val
100 105 110
Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu Glu Ser Arg
115 120 125
Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu Leu Ser Arg
130 135 140
Cys Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Lys Asp Lys Lys Cys
145 150 155 160

Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu Glu Glu Gln

165 170 175

Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg Phe Pro Tyr

180 185 190

Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys Ser Ile Arg

195 200 205

Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr Asp Lys Lys

210 215 220

Ile Gly Tyr Glu Gln Glu Lys Arg His Leu Thr Arg Thr Leu Leu Asn

225 230 235 240

Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro Gln Glu Trp

245 250 255

Lys Ala Leu Thr Lys Asp Leu Asp Tyr Asn Glu Thr Ser Asn Gln Pro

260 265 270

Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp Asn Lys Ile

275 280 285

Gly Phe Arg Leu Arg Thr Ser Lys Glu Leu Tyr Pro Ser Leu Glu Val

290 295 300

Lys Asp Gly Ala Asn Arg Ile Ala Lys Tyr Pro Tyr Asn Ser Asp Phe

305 310 315 320

Val Ala His Ala Phe Ile Ser Ile Ser Val His Glu Leu Leu Pro Leu

325 330 335

Met Phe Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys Glu

340 345 350

Thr Val Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu Arg

355 360 365

Ile Asn Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu Pro

370 375 380

Leu Gly Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys Gln

385 390 395 400

Pro Asp Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile Ala

405 410 415

Glu Thr Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser Ser

420 425 430

Pro Lys Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val Leu

435 440 445

Ala Asp Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Val Tyr

450 455 460

Asp Ala Gln Asn Gln Pro Ile Lys Ser Ser Lys Ala Asn Ser Thr Glu

465 470 475 480

Ser Arg Leu Ile Arg Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys Asn

485 490 495

Arg Leu Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr Asn

500 505 510

Pro His Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn Leu

515 520 525

Val Asp Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu Glu

530 535 540

Ala Ile Lys His Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu Leu

545 550 555 560

Lys Val Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu Gln

565 570 575

Gly Gly Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg Glu

580 585 590

Thr Leu Ser Lys Asp Leu Thr Leu Ser Lys Pro Ile Arg Lys Glu Ile

595 600 605

Lys Lys His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu Tyr

610 615 620

Phe Lys Glu Lys Tyr Gln Asp Lys His Gln Ser Phe Tyr Asn Leu Ser

625 630 635 640

Tyr Lys Leu Glu Ala Lys Ala Pro Leu Leu Lys Lys Glu Glu His Tyr

645 650 655

Glu Tyr Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Ser Gln Arg

660 665 670

Leu Glu Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr Lys
675 680 685
Arg Trp Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln Asp
690 695 700
Ile Met Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe Lys
705 710 715 720
Glu Leu Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala Val
725 730 735
Asn Val Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr Leu

740 745 750
Pro Met Val Leu Pro Val Lys Val Tyr Pro Thr Thr Ala Phe Gly Glu
755 760 765
Val Gln Tyr His Glu Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu Glu
770 775 780
Gln Thr Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys Asp
785 790 795 800
Arg His Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp Thr
805 810 815

Gln Lys His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu Ile
820 825 830
Tyr Gln Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Ser Leu Glu
835 840 845
Glu Lys Leu Leu Asn Lys His Ala Ser Leu Ser Ser Leu Glu Asn Glu
850 855 860
Phe Arg Thr Leu Leu Glu Glu Trp Lys Lys Lys Tyr Ala Ala Ser Ser
865 870 875 880
Met Val Thr Asp Lys His Ile Ala Phe Ile Ala Ser Val Arg Asn Ala

885 890 895
Phe Cys His Asn Gln Tyr Pro Phe Tyr Lys Glu Thr Leu His Ala Pro
900 905 910

Ile Leu Leu Phe Thr Val Ala Gln Pro Thr Thr Glu Glu Lys Asp Gly
915 920 925

Leu Gly Ile Ala Glu Ala Leu Leu Arg Val Leu Arg Glu Tyr Cys Glu
930 935 940

Ile Val Lys Ser Gln Ile
945 950

<210> 117

<211> 1213

<212> PRT

<213> Porphyromonas gulae

<400> 117

Met Ser Ser Lys Asn Glu Ser Tyr Asn Lys Gln Lys Thr Phe Asn His
1 5 10 15

Tyr Lys Gln Glu Asp Lys Tyr Phe Phe Gly Gly Phe Leu Asn Asn Ala
20 25 30

Asp Asp Asn Leu Arg Gln Val Gly Lys Glu Phe Lys Thr Arg Ile Asn
35 40 45

Phe Asn His Asn Asn Asn Glu Leu Ala Ser Val Phe Lys Asp Tyr Phe
50 55 60

Asn Lys Glu Lys Ser Val Ala Lys Arg Glu His Ala Leu Asn Leu Leu
65 70 75 80

Ser Asn Tyr Phe Pro Val Leu Glu Arg Ile Gln Lys His Thr Asn His
85 90 95

Asn Phe Glu Gln Thr Arg Glu Ile Phe Glu Leu Leu Leu Asp Thr Ile
100 105 110

Lys Lys Leu Arg Asp Tyr Tyr Thr His His Tyr His Lys Pro Ile Thr
115 120 125

Ile Asn Pro Lys Ile Tyr Asp Phe Leu Asp Asp Thr Leu Leu Asp Val

130 135 140

Leu Ile Thr Ile Lys Lys Lys Lys Val Lys Asn Asp Thr Ser Arg Glu
145 150 155 160

Leu Leu Lys Glu Lys Leu Arg Pro Glu Leu Thr Gln Leu Lys Asn Gln

165 170 175

Lys Arg Glu Glu Leu Ile Lys Lys Gly Lys Lys Leu Leu Glu Glu Asn

180 185 190

Leu Glu Asn Ala Val Phe Asn His Cys Leu Arg Pro Phe Leu Glu Glu

195 200 205

Asn Lys Thr Asp Asp Lys Gln Asn Lys Thr Val Ser Leu Arg Lys Tyr

210 215 220

Arg Lys Ser Lys Pro Asn Glu Glu Thr Ser Ile Thr Leu Thr Gln Ser

225 230 235 240

Gly Leu Val Phe Leu Met Ser Phe Phe Leu His Arg Lys Glu Phe Gln

245 250 255

Val Phe Thr Ser Gly Leu Glu Gly Phe Lys Ala Lys Val Asn Thr Ile

260 265 270

Lys Glu Glu Lys Ile Ser Leu Asn Lys Asn Asn Ile Val Tyr Met Ile

275 280 285

Thr His Trp Ser Tyr Ser Tyr Tyr Asn Phe Lys Gly Leu Lys His Arg

290 295 300

Ile Lys Thr Asp Gln Gly Val Ser Thr Leu Glu Gln Asn Asn Thr Thr

305 310 315 320

His Ser Leu Thr Asn Thr Asn Thr Lys Glu Ala Leu Leu Thr Gln Ile

325 330 335

Val Asp Tyr Leu Ser Lys Val Pro Asn Glu Ile Tyr Glu Thr Leu Ser

340 345 350

Glu Lys Gln Gln Lys Glu Phe Glu Glu Asp Ile Asn Glu Tyr Met Arg

355 360 365

Glu Asn Pro Glu Asn Glu Asp Ser Thr Phe Ser Ser Ile Val Ser His

370 375 380

Lys Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Met

385 390 395 400

Arg Phe Leu Asp Glu Tyr Ala Glu Leu Pro Thr Leu Arg Phe Met Val

405 410 415

Asn Phe Gly Asp Tyr Ile Lys Asp Arg Gln Lys Lys Ile Leu Glu Ser

420 425 430

Ile Gln Phe Asp Ser Glu Arg Ile Ile Lys Lys Glu Ile His Leu Phe

435 440 445

Glu Lys Leu Gly Leu Val Thr Glu Tyr Lys Lys Asn Val Tyr Leu Lys

450 455 460

Glu Thr Ser Asn Ile Asp Leu Ser Arg Phe Pro Leu Phe Pro Ser Pro

465 470 475 480

Ser Tyr Val Met Ala Asn Asn Asn Ile Pro Phe Tyr Ile Asp Ser Arg

485 490 495

Ser Asn Asn Leu Asp Glu Tyr Leu Asn Gln Lys Lys Lys Ala Gln Ser

500 505 510

Gln Asn Arg Lys Arg Asn Leu Thr Phe Glu Lys Tyr Asn Lys Glu Gln

515 520 525

Ser Lys Asp Ala Ile Ile Ala Met Leu Gln Lys Glu Ile Gly Val Lys

530 535 540

Asp Leu Gln Gln Arg Ser Thr Ile Gly Leu Leu Ser Cys Asn Glu Leu

545 550 555 560

Pro Ser Met Leu Tyr Glu Val Ile Val Lys Asp Ile Lys Gly Ala Glu

565 570 575

Leu Glu Asn Lys Ile Ala Gln Lys Ile Arg Glu Gln Tyr Gln Ser Ile

580 585 590

Arg Asp Phe Thr Leu Asn Ser Pro Gln Lys Asp Asn Ile Pro Thr Thr

595 600 605

Leu Ile Lys Thr Ile Ser Thr Asp Thr Ser Val Thr Phe Glu Asn Gln

610 615 620

Pro Ile Asp Ile Pro Arg Leu Lys Asn Ala Ile Gln Lys Glu Leu Ala

625 630 635 640

Leu Thr Gln Glu Lys Leu Leu Asn Val Lys Gln His Glu Ile Glu Val

645 650 655

Asn Asn Tyr Asn Arg Asn Lys Asn Thr Tyr Lys Phe Lys Asn Gln Pro

660 665 670

Lys Asp Lys Val Asp Asp Asn Lys Leu Gln Arg Lys Tyr Val Phe Tyr

675 680 685

Arg Asn Glu Ile Gly Gln Glu Ala Asn Trp Leu Ala Ser Asp Leu Ile

690 695 700

His Phe Met Lys Asn Lys Ser Leu Trp Lys Gly Tyr Met His Asn Glu

705 710 715 720

Leu Gln Ser Phe Leu Ala Phe Phe Glu Asp Lys Lys Asn Asp Cys Ile

725 730 735

Ala Leu Leu Glu Thr Val Phe Asn Leu Lys Glu Asp Cys Ile Leu Thr

740 745 750

Lys Asp Leu Lys Asn Leu Phe Leu Lys His Gly Asn Phe Ile Asp Phe

755 760 765

Tyr Lys Glu Tyr Leu Lys Leu Lys Glu Asp Phe Leu Asn Thr Glu Ser

770 775 780

Thr Phe Leu Glu Asn Gly Phe Ile Gly Leu Pro Pro Lys Ile Leu Lys

785 790 795 800

Lys Glu Leu Ser Lys Arg Leu Asn Tyr Ile Phe Ile Val Phe Gln Lys

805 810 815

Arg Gln Phe Ile Ile Lys Glu Leu Glu Glu Lys Lys Asn Asn Leu Tyr

820 825 830

Ala Asp Ala Ile Asn Leu Ser Arg Gly Ile Phe Asp Glu Lys Pro Thr

835 840 845

Met Ile Pro Phe Lys Lys Pro Asn Pro Asp Glu Phe Ala Ser Trp Phe

850 855 860

Val Ala Ser Tyr Gln Tyr Asn Asn Tyr Gln Ser Phe Tyr Glu Leu Thr

865 870 875 880

Pro Asp Lys Ile Glu Asn Asp Lys Lys Lys Lys Tyr Lys Asn Leu Arg

885 890 895

Ala Ile Asn Lys Val Lys Ile Gln Asp Tyr Tyr Leu Lys Leu Met Val

900 905 910

Asp Thr Leu Tyr Gln Asp Leu Phe Asn Gln Pro Leu Asp Lys Ser Leu

915 920 925

Ser Asp Phe Tyr Val Ser Lys Thr Asp Arg Glu Lys Ile Lys Ala Asp
930 935 940
Ala Lys Ala Tyr Gln Lys Arg Asn Asp Ser Phe Leu Trp Asn Lys Val
945 950 955 960
Ile His Leu Ser Leu Gln Asn Asn Arg Ile Thr Ala Asn Pro Lys Leu
965 970 975
Lys Asp Ile Gly Lys Tyr Lys Arg Ala Leu Gln Asp Glu Lys Ile Ala
980 985 990
Thr Leu Leu Thr Tyr Asp Asp Arg Thr Trp Thr Tyr Ala Leu Gln Lys

995 1000 1005
Pro Glu Lys Glu Asn Glu Asn Asp Tyr Lys Glu Leu His Tyr Thr Ala
1010 1015 1020
Leu Asn Met Glu Leu Gln Glu Tyr Glu Lys Val Arg Ser Lys Lys Leu
1025 1030 1035 1040
Leu Lys Gln Val Gln Glu Leu Glu Lys Gln Ile Leu Asp Lys Phe Tyr
1045 1050 1055
Asp Phe Ser Asn Asn Ala Thr His Pro Glu Asp Leu Glu Ile Glu Asp
1060 1065 1070

Lys Lys Gly Lys Arg His Pro Asn Phe Lys Leu Tyr Ile Thr Lys Ala
1075 1080 1085
Leu Leu Lys Asn Glu Ser Glu Ile Ile Asn Leu Glu Asn Ile Asp Ile
1090 1095 1100
Glu Ile Leu Ile Lys Tyr Tyr Asp Tyr Asn Thr Glu Lys Leu Lys Glu
1105 1110 1115 1120
Lys Ile Lys Asn Met Asp Glu Asp Glu Lys Ala Lys Ile Val Asn Thr
1125 1130 1135
Lys Glu Asn Tyr Asn Lys Ile Thr Asn Val Leu Ile Lys Lys Ala Leu

1140 1145 1150
Val Leu Ile Ile Ile Arg Asn Lys Met Ala His Asn Gln Tyr Pro Pro
1155 1160 1165

Lys Phe Ile Tyr Asp Leu Ala Thr Arg Phe Val Pro Lys Lys Glu Glu
1170 1175 1180
Glu Tyr Phe Ala Cys Tyr Phe Asn Arg Val Phe Glu Thr Ile Thr Thr
1185 1190 1195 1200
Glu Leu Trp Glu Asn Lys Lys Lys Ala Lys Glu Ile Val
1205 1210

<210> 118
<211> 1095
<212> PRT
<213> Porphyromonas gulae
<400> 118
Met Glu Lys Pro Leu Pro Pro Asn Val Tyr Thr Leu Lys His Lys Phe
1 5 10 15
Phe Trp Gly Ala Phe Leu Asn Ile Ala Arg His Asn Ala Phe Ile Thr
20 25 30
Ile Cys His Ile Asn Glu Gln Leu Gly Leu Thr Thr Pro Pro Asn Asp
35 40 45
Asp Lys Ile Ala Asp Val Val Cys Gly Thr Trp Asn Asn Ile Leu Asn

50 55 60
Asn Asp His Asp Leu Leu Lys Lys Ser Gln Leu Thr Glu Leu Ile Leu
65 70 75 80
Lys His Phe Pro Phe Leu Ala Ala Met Cys Tyr His Pro Pro Lys Lys
85 90 95
Glu Gly Lys Lys Lys Gly Ser Gln Lys Glu Gln Gln Lys Glu Lys Glu
100 105 110
Asn Glu Ala Gln Ser Gln Ala Glu Ala Leu Asn Pro Ser Glu Leu Ile
115 120 125

Lys Val Leu Lys Thr Leu Val Lys Gln Leu Arg Thr Leu Arg Asn Tyr
130 135 140
Tyr Ser His His Ser His Lys Lys Pro Asp Ala Glu Lys Asp Ile Phe
145 150 155 160
Lys His Leu Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu

165 170 175

Asp Tyr Lys Ala His Phe Thr Val Asn Leu Thr Gln Asp Phe Ala His

180 185 190

Leu Asn Arg Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asp Arg

195 200 205

Tyr Arg Phe Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe

210 215 220

Phe Thr Asn Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys

225 230 235 240

Lys Val Ser Gly Phe Lys Ala Ser His Lys Gln Ser Glu Lys Met Thr

245 250 255

Thr Glu Val Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu

260 265 270

Glu Ser Arg Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu

275 280 285

Leu Ser Arg Tyr Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Glu Asp

290 295 300

Lys Lys Arg Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu

305 310 315 320

Glu Glu Gln Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg

325 330 335

Phe Pro Tyr Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys

340 345 350

Ser Ile Arg Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr

355 360 365

Asp Lys Lys Ile Gly Asp Glu Gln Glu Lys Arg His Leu Thr Arg Thr

370 375 380

Leu Leu Ser Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro

385 390 395 400

Gln Glu Trp Lys Ala Leu Thr Lys Asp Leu Asp Tyr Lys Glu Thr Ser

405 410 415

Lys Gln Pro Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp
420 425 430
Asn Lys Ile Gly Phe Arg Leu Gly Thr Ser Lys Glu Leu Tyr Pro Ser
435 440 445
Leu Glu Val Lys Asp Gly Ala Asn Arg Ile Ala Gln Tyr Pro Tyr Asn
450 455 460
Ser Asp Phe Val Ala His Ala Phe Ile Ser Val His Glu Leu Leu Pro
465 470 475 480
Leu Met Phe Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys

485 490 495
Glu Thr Val Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu
500 505 510
Arg Ile Asn Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu
515 520 525
Pro Leu Gly Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys
530 535 540
Gln Pro Asp Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile
545 550 555 560

Ala Glu Thr Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser
565 570 575
Ser Pro Lys Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val
580 585 590
Leu Ala Asp Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Ala
595 600 605
Tyr Asp Ala Gln Asn Gln Pro Ile Glu Ser Ser Lys Ala Asn Ser Thr
610 615 620
Glu Phe Gln Leu Ile Gln Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys

625 630 635 640
Asn Arg Leu Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr
645 650 655
Asn Pro His Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn
660 665 670

Leu Val Asp Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu
675 680 685
Glu Ala Ile Lys Asn Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu
690 695 700

Leu Lys Ile Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu
705 710 715 720
Gln Gly Gly Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg
725 730 735
Glu Thr Leu Ser Lys Asp Leu Thr Leu Ser Lys Pro Ile Arg Lys Glu
740 745 750
Ile Lys Lys His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu
755 760 765
Tyr Phe Arg Glu Lys Tyr Gln Asp Asp His Gln Ser Phe Tyr Asp Leu

770 775 780
Pro Tyr Lys Leu Glu Ala Lys Ala Ser Pro Leu Pro Lys Lys Glu His
785 790 795 800
Tyr Glu Tyr Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Leu Gln
805 810 815
Arg Leu Glu Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr
820 825 830
Lys Arg Trp Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln
835 840 845

Asp Val Met Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe
850 855 860
Lys Glu Leu Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala
865 870 875 880
Val Asn Val Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr
885 890 895
Leu Pro Met Val Leu Pro Val Lys Val Tyr Pro Ala Thr Ala Phe Gly
900 905 910
Glu Val Gln Tyr Gln Glu Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu

915 920 925

Glu Gln Thr Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys

930 935 940

Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp

945 950 955 960

Thr Gln Lys His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu

965 970 975

Ile Tyr Gln Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Asn Leu

980 985 990

Glu Glu Lys Leu Leu Lys Lys His Thr Ser Leu Ser Ser Val Glu Asn

995 1000 1005

Lys Phe Arg Ile Leu Leu Glu Glu Trp Lys Lys Glu Tyr Ala Ala Ser

1010 1015 1020

Ser Met Val Thr Asp Glu His Ile Ala Phe Ile Ala Ser Val Arg Asn

1025 1030 1035 1040

Ala Phe Cys His Asn Gln Tyr Pro Phe Tyr Glu Glu Ala Leu His Ala

1045 1050 1055

Pro Ile Pro Leu Phe Thr Val Ala Gln Gln Thr Thr Glu Glu Lys Asp

1060 1065 1070

Gly Leu Gly Ile Ala Glu Ala Leu Leu Arg Val Leu Arg Glu Tyr Cys

1075 1080 1085

Glu Ile Val Lys Ser Gln Ile

1090 1095

<210> 119

<211> 1340

<212> PRT

<213> Lachnospiraceae bacterium

<400> 119

Met Gln Ile Ser Lys Val Asn His Lys His Val Ala Val Gly Gln Lys

1 5 10 15

Asp Arg Glu Arg Ile Thr Gly Phe Ile Tyr Asn Asp Pro Val Gly Asp

20 25 30

Glu Lys Ser Leu Glu Asp Val Val Ala Lys Arg Ala Asn Asp Thr Lys
35 40 45
Val Leu Phe Asn Val Phe Asn Thr Lys Asp Leu Tyr Asp Ser Gln Glu
50 55 60
Ser Asp Lys Ser Glu Lys Asp Lys Glu Ile Ile Ser Lys Gly Ala Lys
65 70 75 80
Phe Val Ala Lys Ser Phe Asn Ser Ala Ile Thr Ile Leu Lys Lys Gln
85 90 95

Asn Lys Ile Tyr Ser Thr Leu Thr Ser Gln Gln Val Ile Lys Glu Leu
100 105 110
Lys Asp Lys Phe Gly Gly Ala Arg Ile Tyr Asp Asp Asp Ile Glu Glu
115 120 125
Ala Leu Thr Glu Thr Leu Lys Lys Ser Phe Arg Lys Glu Asn Val Arg
130 135 140
Asn Ser Ile Lys Val Leu Ile Glu Asn Ala Ala Gly Ile Arg Ser Ser
145 150 155 160
Leu Ser Lys Asp Glu Glu Glu Leu Ile Gln Glu Tyr Phe Val Lys Gln
165 170 175
Leu Val Glu Glu Tyr Thr Lys Thr Lys Leu Gln Lys Asn Val Val Lys
180 185 190
Ser Ile Lys Asn Gln Asn Met Val Ile Gln Pro Asp Ser Asp Ser Gln
195 200 205
Val Leu Ser Leu Ser Glu Ser Arg Arg Glu Lys Gln Ser Ser Ala Val
210 215 220
Ser Ser Asp Thr Leu Val Asn Cys Lys Glu Lys Asp Val Leu Lys Ala
225 230 235 240
Phe Leu Thr Asp Tyr Ala Val Leu Asp Glu Asp Glu Arg Asn Ser Leu
245 250 255
Leu Trp Lys Leu Arg Asn Leu Val Asn Leu Tyr Phe Tyr Gly Ser Glu
260 265 270
Ser Ile Arg Asp Tyr Ser Tyr Thr Lys Glu Lys Ser Val Trp Lys Glu
275 280 285

His Asp Glu Gln Lys Ala Asn Lys Thr Leu Phe Ile Asp Glu Ile Cys
290 295 300
His Ile Thr Lys Ile Gly Lys Asn Gly Lys Glu Gln Lys Val Leu Asp

305 310 315 320
Tyr Glu Glu Asn Arg Ser Arg Cys Arg Lys Gln Asn Ile Asn Tyr Tyr
325 330 335
Arg Ser Ala Leu Asn Tyr Ala Lys Asn Asn Thr Ser Gly Ile Phe Glu
340 345 350
Asn Glu Asp Ser Asn His Phe Trp Ile His Leu Ile Glu Asn Glu Val
355 360 365
Glu Arg Leu Tyr Asn Gly Ile Glu Asn Gly Glu Glu Phe Lys Phe Glu
370 375 380

Thr Gly Tyr Ile Ser Glu Lys Val Trp Lys Ala Val Ile Asn His Leu
385 390 395 400
Ser Ile Lys Tyr Ile Ala Leu Gly Lys Ala Val Tyr Asn Tyr Ala Met
405 410 415
Lys Glu Leu Ser Ser Pro Gly Asp Ile Glu Pro Gly Lys Ile Asp Asp
420 425 430
Ser Tyr Ile Asn Gly Ile Thr Ser Phe Asp Tyr Glu Ile Ile Lys Ala
435 440 445
Glu Glu Ser Leu Gln Arg Asp Ile Ser Met Asn Val Val Phe Ala Thr

450 455 460
Asn Tyr Leu Ala Cys Ala Thr Val Asp Thr Asp Lys Asp Phe Leu Leu
465 470 475 480
Phe Ser Lys Glu Asp Ile Arg Ser Cys Thr Lys Lys Asp Gly Asn Leu
485 490 495
Cys Lys Asn Ile Met Gln Phe Trp Gly Gly Tyr Ser Thr Trp Lys Asn
500 505 510
Phe Cys Glu Glu Tyr Leu Lys Asp Asp Lys Asp Ala Leu Glu Leu Leu
515 520 525

Tyr Ser Leu Lys Ser Met Leu Tyr Ser Met Arg Asn Ser Ser Phe His

530 535 540

Phe Ser Thr Glu Asn Val Asp Asn Gly Ser Trp Asp Thr Glu Leu Ile

545 550 555 560

Gly Lys Leu Phe Glu Glu Asp Cys Asn Arg Ala Ala Arg Ile Glu Lys

565 570 575

Glu Lys Phe Tyr Asn Asn Asn Leu His Met Phe Tyr Ser Ser Ser Leu

580 585 590

Leu Glu Lys Val Leu Glu Arg Leu Tyr Ser Ser His His Glu Arg Ala

595 600 605

Ser Gln Val Pro Ser Phe Asn Arg Val Phe Val Arg Lys Asn Phe Pro

610 615 620

Ser Ser Leu Ser Glu Gln Arg Ile Thr Pro Lys Phe Thr Asp Ser Lys

625 630 635 640

Asp Glu Gln Ile Trp Gln Ser Ala Val Tyr Tyr Leu Cys Lys Glu Ile

645 650 655

Tyr Tyr Asn Asp Phe Leu Gln Ser Lys Glu Ala Tyr Lys Leu Phe Arg

660 665 670

Glu Gly Val Lys Asn Leu Asp Lys Asn Asp Ile Asn Asn Gln Lys Ala

675 680 685

Ala Asp Ser Phe Lys Gln Ala Val Val Tyr Tyr Gly Lys Ala Ile Gly

690 695 700

Asn Ala Thr Leu Ser Gln Val Cys Gln Ala Ile Met Thr Glu Tyr Asn

705 710 715 720

Arg Gln Asn Asn Asp Gly Leu Lys Lys Lys Ser Ala Tyr Ala Glu Lys

725 730 735

Gln Asn Ser Asn Lys Tyr Lys His Tyr Pro Leu Phe Leu Lys Gln Val

740 745 750

Leu Gln Ser Ala Phe Trp Glu Tyr Leu Asp Glu Asn Lys Glu Ile Tyr

755 760 765

Gly Phe Ile Ser Ala Gln Ile His Lys Ser Asn Val Glu Ile Lys Ala

770 775 780

Glu Asp Phe Ile Ala Asn Tyr Ser Ser Gln Gln Tyr Lys Lys Leu Val

785 790 795 800

Asp Lys Val Lys Lys Thr Pro Glu Leu Gln Lys Trp Tyr Thr Leu Gly

805 810 815

Arg Leu Ile Asn Pro Arg Gln Ala Asn Gln Phe Leu Gly Ser Ile Arg

820 825 830

Asn Tyr Val Gln Phe Val Lys Asp Ile Gln Arg Arg Ala Lys Glu Asn

835 840 845

Gly Asn Pro Ile Arg Asn Tyr Tyr Glu Val Leu Glu Ser Asp Ser Ile

850 855 860

Ile Lys Ile Leu Glu Met Cys Thr Lys Leu Asn Gly Thr Thr Ser Asn

865 870 875 880

Asp Ile His Asp Tyr Phe Arg Asp Glu Asp Glu Tyr Ala Glu Tyr Ile

885 890 895

Ser Gln Phe Val Asn Phe Gly Asp Val His Ser Gly Ala Ala Leu Asn

900 905 910

Ala Phe Cys Asn Ser Glu Ser Glu Gly Lys Lys Asn Gly Ile Tyr Tyr

915 920 925

Asp Gly Ile Asn Pro Ile Val Asn Arg Asn Trp Val Leu Cys Lys Leu

930 935 940

Tyr Gly Ser Pro Asp Leu Ile Ser Lys Ile Ile Ser Arg Val Asn Glu

945 950 955 960

Asn Met Ile His Asp Phe His Lys Gln Glu Asp Leu Ile Arg Glu Tyr

965 970 975

Gln Ile Lys Gly Ile Cys Ser Asn Lys Lys Glu Gln Gln Asp Leu Arg

980 985 990

Thr Phe Gln Val Leu Lys Asn Arg Val Glu Leu Arg Asp Ile Val Glu

995 1000 1005

Tyr Ser Glu Ile Ile Asn Glu Leu Tyr Gly Gln Leu Ile Lys Trp Cys

1010 1015 1020

Tyr Leu Arg Glu Arg Asp Leu Met Tyr Phe Gln Leu Gly Phe His Tyr

1025 1030 1035 1040

Leu Cys Leu Asn Asn Ala Ser Ser Lys Glu Ala Asp Tyr Ile Lys Ile
1045 1050 1055
Asn Val Asp Asp Arg Asn Ile Ser Gly Ala Ile Leu Tyr Gln Ile Ala
1060 1065 1070
Ala Met Tyr Ile Asn Gly Leu Pro Val Tyr Tyr Lys Lys Asp Asp Met
1075 1080 1085
Tyr Val Ala Leu Lys Ser Gly Lys Lys Ala Ser Asp Glu Leu Asn Ser
1090 1095 1100

Asn Glu Gln Thr Ser Lys Lys Ile Asn Tyr Phe Leu Lys Tyr Gly Asn
1105 1110 1115 1120
Asn Ile Leu Gly Asp Lys Lys Asp Gln Leu Tyr Leu Ala Gly Leu Glu
1125 1130 1135
Leu Phe Glu Asn Val Ala Glu His Glu Asn Ile Ile Ile Phe Arg Asn
1140 1145 1150
Glu Ile Asp His Phe His Tyr Phe Tyr Asp Arg Asp Arg Ser Met Leu
1155 1160 1165
Asp Leu Tyr Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Met Lys

1170 1175 1180
Leu Arg Lys Asn Val Val Asn Met Leu Tyr Asn Ile Leu Leu Asp His
1185 1190 1195 1200
Asn Ile Val Ser Ser Phe Val Phe Glu Thr Gly Glu Lys Lys Val Gly
1205 1210 1215
Arg Gly Asp Ser Glu Val Ile Lys Pro Ser Ala Lys Ile Arg Leu Arg
1220 1225 1230
Ala Asn Asn Gly Val Ser Ser Asp Val Phe Thr Tyr Lys Val Gly Ser
1235 1240 1245

Lys Asp Glu Leu Lys Ile Ala Thr Leu Pro Ala Lys Asn Glu Glu Phe
1250 1255 1260
Leu Leu Asn Val Ala Arg Leu Ile Tyr Tyr Pro Asp Met Glu Ala Val
1265 1270 1275 1280
Ser Glu Asn Met Val Arg Glu Gly Val Val Lys Val Glu Lys Ser Asn
1285 1290 1295

Asp Lys Lys Gly Lys Ile Ser Arg Gly Ser Asn Thr Arg Ser Ser Asn

1300 1305 1310

Gln Ser Lys Tyr Asn Asn Lys Ser Lys Asn Arg Met Asn Tyr Ser Met

1315 1320 1325

Gly Ser Ile Phe Glu Lys Met Asp Leu Lys Phe Asp

1330 1335 1340

<210> 120

<211> 1437

<212> PRT

<213> Lachnospiraceae bacterium

<400> 120

Met Lys Ile Ser Lys Val Arg Glu Glu Asn Arg Gly Ala Lys Leu Thr

1 5 10 15

Val Asn Ala Lys Thr Ala Val Val Ser Glu Asn Arg Ser Gln Glu Gly

20 25 30

Ile Leu Tyr Asn Asp Pro Ser Arg Tyr Gly Lys Ser Arg Lys Asn Asp

35 40 45

Glu Asp Arg Asp Arg Tyr Ile Glu Ser Arg Leu Lys Ser Ser Gly Lys

50 55 60

Leu Tyr Arg Ile Phe Asn Glu Asp Lys Asn Lys Arg Glu Thr Asp Glu

65 70 75 80

Leu Gln Trp Phe Leu Ser Glu Ile Val Lys Lys Ile Asn Arg Arg Asn

85 90 95

Gly Leu Val Leu Ser Asp Met Leu Ser Val Asp Asp Arg Ala Phe Glu

100 105 110

Lys Ala Phe Glu Lys Tyr Ala Glu Leu Ser Tyr Thr Asn Arg Arg Asn

115 120 125

Lys Val Ser Gly Ser Pro Ala Phe Glu Thr Cys Gly Val Asp Ala Ala

130 135 140

Thr Ala Glu Arg Leu Lys Gly Ile Ile Ser Glu Thr Asn Phe Ile Asn

145 150 155 160

Arg Ile Lys Asn Asn Ile Asp Asn Lys Val Ser Glu Asp Ile Ile Asp

165 170 175

Arg Ile Ile Ala Lys Tyr Leu Lys Lys Ser Leu Cys Arg Glu Arg Val

180 185 190

Lys Arg Gly Leu Lys Lys Leu Leu Met Asn Ala Phe Asp Leu Pro Tyr

195 200 205

Ser Asp Pro Asp Ile Asp Val Gln Arg Asp Phe Ile Asp Tyr Val Leu

210 215 220

Glu Asp Phe Tyr His Val Arg Ala Lys Ser Gln Val Ser Arg Ser Ile

225 230 235 240

Lys Asn Met Asn Met Pro Val Gln Pro Glu Gly Asp Gly Lys Phe Ala

245 250 255

Ile Thr Val Ser Lys Gly Gly Thr Glu Ser Gly Asn Lys Arg Ser Ala

260 265 270

Glu Lys Glu Ala Phe Lys Lys Phe Leu Ser Asp Tyr Ala Ser Leu Asp

275 280 285

Glu Arg Val Arg Asp Asp Met Leu Arg Arg Met Arg Arg Leu Val Val

290 295 300

Leu Tyr Phe Tyr Gly Ser Asp Asp Ser Lys Leu Ser Asp Val Asn Glu

305 310 315 320

Lys Phe Asp Val Trp Glu Asp His Ala Ala Arg Arg Val Asp Asn Arg

325 330 335

Glu Phe Ile Lys Leu Pro Leu Glu Asn Lys Leu Ala Asn Gly Lys Thr

340 345 350

Asp Lys Asp Ala Glu Arg Ile Arg Lys Asn Thr Val Lys Glu Leu Tyr

355 360 365

Arg Asn Gln Asn Ile Gly Cys Tyr Arg Gln Ala Val Lys Ala Val Glu

370 375 380

Glu Asp Asn Asn Gly Arg Tyr Phe Asp Asp Lys Met Leu Asn Met Phe

385 390 395 400

Phe Ile His Arg Ile Glu Tyr Gly Val Glu Lys Ile Tyr Ala Asn Leu

405 410 415

Lys Gln Val Thr Glu Phe Lys Ala Arg Thr Gly Tyr Leu Ser Glu Lys
420 425 430
Ile Trp Lys Asp Leu Ile Asn Tyr Ile Ser Ile Lys Tyr Ile Ala Met
435 440 445
Gly Lys Ala Val Tyr Asn Tyr Ala Met Asp Glu Leu Asn Ala Ser Asp
450 455 460
Lys Lys Glu Ile Glu Leu Gly Lys Ile Ser Glu Glu Tyr Leu Ser Gly

465 470 475 480
Ile Ser Ser Phe Asp Tyr Glu Leu Ile Lys Ala Glu Glu Met Leu Gln
485 490 495
Arg Glu Thr Ala Val Tyr Val Ala Phe Ala Ala Arg His Leu Ser Ser
500 505 510
Gln Thr Val Glu Leu Asp Ser Glu Asn Ser Asp Phe Leu Leu Leu Lys
515 520 525
Pro Lys Gly Thr Met Asp Lys Asn Asp Lys Asn Lys Leu Ala Ser Asn
530 535 540

Asn Ile Leu Asn Phe Leu Lys Asp Lys Glu Thr Leu Arg Asp Thr Ile
545 550 555 560
Leu Gln Tyr Phe Gly Gly His Ser Leu Trp Thr Asp Phe Pro Phe Asp
565 570 575
Lys Tyr Leu Ala Gly Gly Lys Asp Asp Val Asp Phe Leu Thr Asp Leu
580 585 590
Lys Asp Val Ile Tyr Ser Met Arg Asn Asp Ser Phe His Tyr Ala Thr
595 600 605
Glu Asn His Asn Asn Gly Lys Trp Asn Lys Glu Leu Ile Ser Ala Met

610 615 620
Phe Glu His Glu Thr Glu Arg Met Thr Val Val Met Lys Asp Lys Phe
625 630 635 640
Tyr Ser Asn Asn Leu Pro Met Phe Tyr Lys Asn Asp Asp Leu Lys Lys
645 650 655
Leu Leu Ile Asp Leu Tyr Lys Asp Asn Val Glu Arg Ala Ser Gln Val
660 665 670

Pro Ser Phe Asn Lys Val Phe Val Arg Lys Asn Phe Pro Ala Leu Val
675 680 685

Arg Asp Lys Asp Asn Leu Gly Ile Glu Leu Asp Leu Lys Ala Asp Ala
690 695 700

Asp Lys Gly Glu Asn Glu Leu Lys Phe Tyr Asn Ala Leu Tyr Tyr Met
705 710 715 720

Phe Lys Glu Ile Tyr Tyr Asn Ala Phe Leu Asn Asp Lys Asn Val Arg
725 730 735

Glu Arg Phe Ile Thr Lys Ala Thr Lys Val Ala Asp Asn Tyr Asp Arg
740 745 750

Asn Lys Glu Arg Asn Leu Lys Asp Arg Ile Lys Ser Ala Gly Ser Asp
755 760 765

Glu Lys Lys Lys Leu Arg Glu Gln Leu Gln Asn Tyr Ile Ala Glu Asn
770 775 780

Asp Phe Gly Gln Arg Ile Lys Asn Ile Val Gln Val Asn Pro Asp Tyr
785 790 795 800

Thr Leu Ala Gln Ile Cys Gln Leu Ile Met Thr Glu Tyr Asn Gln Gln
805 810 815

Asn Asn Gly Cys Met Gln Lys Lys Ser Ala Ala Arg Lys Asp Ile Asn
820 825 830

Lys Asp Ser Tyr Gln His Tyr Lys Met Leu Leu Leu Val Asn Leu Arg
835 840 845

Lys Ala Phe Leu Glu Phe Ile Lys Glu Asn Tyr Ala Phe Val Leu Lys
850 855 860

Pro Tyr Lys His Asp Leu Cys Asp Lys Ala Asp Phe Val Pro Asp Phe
865 870 875 880

Ala Lys Tyr Val Lys Pro Tyr Ala Gly Leu Ile Ser Arg Val Ala Gly
885 890 895

Ser Ser Glu Leu Gln Lys Trp Tyr Ile Val Ser Arg Phe Leu Ser Pro
900 905 910

Ala Gln Ala Asn His Met Leu Gly Phe Leu His Ser Tyr Lys Gln Tyr

915 920 925

Val Trp Asp Ile Tyr Arg Arg Ala Ser Glu Thr Gly Thr Glu Ile Asn

930 935 940

His Ser Ile Ala Glu Asp Lys Ile Ala Gly Val Asp Ile Thr Asp Val

945 950 955 960

Asp Ala Val Ile Asp Leu Ser Val Lys Leu Cys Gly Thr Ile Ser Ser

965 970 975

Glu Ile Ser Asp Tyr Phe Lys Asp Asp Glu Val Tyr Ala Glu Tyr Ile

980 985 990

Ser Ser Tyr Leu Asp Phe Glu Tyr Asp Gly Gly Asn Tyr Lys Asp Ser

995 1000 1005

Leu Asn Arg Phe Cys Asn Ser Asp Ala Val Asn Asp Gln Lys Val Ala

1010 1015 1020

Leu Tyr Tyr Asp Gly Glu His Pro Lys Leu Asn Arg Asn Ile Ile Leu

1025 1030 1035 1040

Ser Lys Leu Tyr Gly Glu Arg Arg Phe Leu Glu Lys Ile Thr Asp Arg

1045 1050 1055

Val Ser Arg Ser Asp Ile Val Glu Tyr Tyr Lys Leu Lys Lys Glu Thr

1060 1065 1070

Ser Gln Tyr Gln Thr Lys Gly Ile Phe Asp Ser Glu Asp Glu Gln Lys

1075 1080 1085

Asn Ile Lys Lys Phe Gln Glu Met Lys Asn Ile Val Glu Phe Arg Asp

1090 1095 1100

Leu Met Asp Tyr Ser Glu Ile Ala Asp Glu Leu Gln Gly Gln Leu Ile

1105 1110 1115 1120

Asn Trp Ile Tyr Leu Arg Glu Arg Asp Leu Met Asn Phe Gln Leu Gly

1125 1130 1135

Tyr His Tyr Ala Cys Leu Asn Asn Asp Ser Asn Lys Gln Ala Thr Tyr

1140 1145 1150

Val Thr Leu Asp Tyr Gln Gly Lys Lys Asn Arg Lys Ile Asn Gly Ala

1155 1160 1165

Ile Leu Tyr Gln Ile Cys Ala Met Tyr Ile Asn Gly Leu Pro Leu Tyr

1170 1175 1180

Tyr Val Asp Lys Asp Ser Ser Glu Trp Thr Val Ser Asp Gly Lys Glu

1185 1190 1195 1200

Ser Thr Gly Ala Lys Ile Gly Glu Phe Tyr Arg Tyr Ala Lys Ser Phe

1205 1210 1215

Glu Asn Thr Ser Asp Cys Tyr Ala Ser Gly Leu Glu Ile Phe Glu Asn

1220 1225 1230

Ile Ser Glu His Asp Asn Ile Thr Glu Leu Arg Asn Tyr Ile Glu His

1235 1240 1245

Phe Arg Tyr Tyr Ser Ser Phe Asp Arg Ser Phe Leu Gly Ile Tyr Ser

1250 1255 1260

Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Leu Lys Tyr Arg Lys Asn

1265 1270 1275 1280

Val Pro Thr Ile Leu Tyr Asn Ile Leu Leu Gln His Phe Val Asn Val

1285 1290 1295

Arg Phe Glu Phe Val Ser Gly Lys Lys Met Ile Gly Ile Asp Lys Lys

1300 1305 1310

Asp Arg Lys Ile Ala Lys Glu Lys Glu Cys Ala Arg Ile Thr Ile Arg

1315 1320 1325

Glu Lys Asn Gly Val Tyr Ser Glu Gln Phe Thr Tyr Lys Leu Lys Asn

1330 1335 1340

Gly Thr Val Tyr Val Asp Ala Arg Asp Lys Arg Tyr Leu Gln Ser Ile

1345 1350 1355 1360

Ile Arg Leu Leu Phe Tyr Pro Glu Lys Val Asn Met Asp Glu Met Ile

1365 1370 1375

Glu Val Lys Glu Lys Lys Lys Pro Ser Asp Asn Asn Thr Gly Lys Gly

1380 1385 1390

Tyr Ser Lys Arg Asp Arg Gln Gln Asp Arg Lys Glu Tyr Asp Lys Tyr

1395 1400 1405

Lys Glu Lys Lys Lys Lys Glu Gly Asn Phe Leu Ser Gly Met Gly Gly

1410 1415 1420

Asn Ile Asn Trp Asp Glu Ile Asn Ala Gln Leu Lys Asn

1425 1430 1435

<210> 121

<211> 1385

<212> PRT

<213> Clostridium] aminophilum

<400> 121

Met Lys Phe Ser Lys Val Asp His Thr Arg Ser Ala Val Gly Ile Gln

1 5 10 15

Lys Ala Thr Asp Ser Val His Gly Met Leu Tyr Thr Asp Pro Lys Lys

20 25 30

Gln Glu Val Asn Asp Leu Asp Lys Arg Phe Asp Gln Leu Asn Val Lys

35 40 45

Ala Lys Arg Leu Tyr Asn Val Phe Asn Gln Ser Lys Ala Glu Glu Asp

50 55 60

Asp Asp Glu Lys Arg Phe Gly Lys Val Val Lys Lys Leu Asn Arg Glu

65 70 75 80

Leu Lys Asp Leu Leu Phe His Arg Glu Val Ser Arg Tyr Asn Ser Ile

85 90 95

Gly Asn Ala Lys Tyr Asn Tyr Tyr Gly Ile Lys Ser Asn Pro Glu Glu

100 105 110

Ile Val Ser Asn Leu Gly Met Val Glu Ser Leu Lys Gly Glu Arg Asp

115 120 125

Pro Gln Lys Val Ile Ser Lys Leu Leu Leu Tyr Tyr Leu Arg Lys Gly

130 135 140

Leu Lys Pro Gly Thr Asp Gly Leu Arg Met Ile Leu Glu Ala Ser Cys

145 150 155 160

Gly Leu Arg Lys Leu Ser Gly Asp Glu Lys Glu Leu Lys Val Phe Leu

165 170 175

Gln Thr Leu Asp Glu Asp Phe Glu Lys Lys Thr Phe Lys Lys Asn Leu

180 185 190

Ile Arg Ser Ile Glu Asn Gln Asn Met Ala Val Gln Pro Ser Asn Glu

195 200 205

Gly Asp Pro Ile Ile Gly Ile Thr Gln Gly Arg Phe Asn Ser Gln Lys

210 215 220

Asn Glu Glu Lys Ser Ala Ile Glu Arg Met Met Ser Met Tyr Ala Asp

225 230 235 240

Leu Asn Glu Asp His Arg Glu Asp Val Leu Arg Lys Leu Arg Arg Leu

245 250 255

Asn Val Leu Tyr Phe Asn Val Asp Thr Glu Lys Thr Glu Glu Pro Thr

260 265 270

Leu Pro Gly Glu Val Asp Thr Asn Pro Val Phe Glu Val Trp His Asp

275 280 285

His Glu Lys Gly Lys Glu Asn Asp Arg Gln Phe Ala Thr Phe Ala Lys

290 295 300

Ile Leu Thr Glu Asp Arg Glu Thr Arg Lys Lys Glu Lys Leu Ala Val

305 310 315 320

Lys Glu Ala Leu Asn Asp Leu Lys Ser Ala Ile Arg Asp His Asn Ile

325 330 335

Met Ala Tyr Arg Cys Ser Ile Lys Val Thr Glu Gln Asp Lys Asp Gly

340 345 350

Leu Phe Phe Glu Asp Gln Arg Ile Asn Arg Phe Trp Ile His His Ile

355 360 365

Glu Ser Ala Val Glu Arg Ile Leu Ala Ser Ile Asn Pro Glu Lys Leu

370 375 380

Tyr Lys Leu Arg Ile Gly Tyr Leu Gly Glu Lys Val Trp Lys Asp Leu

385 390 395 400

Leu Asn Tyr Leu Ser Ile Lys Tyr Ile Ala Val Gly Lys Ala Val Phe

405 410 415

His Phe Ala Met Glu Asp Leu Gly Lys Thr Gly Gln Asp Ile Glu Leu

420 425 430

Gly Lys Leu Ser Asn Ser Val Ser Gly Gly Leu Thr Ser Phe Asp Tyr

435 440 445

Glu Gln Ile Arg Ala Asp Glu Thr Leu Gln Arg Gln Leu Ser Val Glu

450 455 460

Val Ala Phe Ala Ala Asn Asn Leu Phe Arg Ala Val Val Gly Gln Thr
465 470 475 480
Gly Lys Lys Ile Glu Gln Ser Lys Ser Glu Glu Asn Glu Glu Asp Phe
485 490 495
Leu Leu Trp Lys Ala Glu Lys Ile Ala Glu Ser Ile Lys Lys Glu Gly
500 505 510
Glu Gly Asn Thr Leu Lys Ser Ile Leu Gln Phe Phe Gly Gly Ala Ser
515 520 525

Ser Trp Asp Leu Asn His Phe Cys Ala Ala Tyr Gly Asn Glu Ser Ser
530 535 540
Ala Leu Gly Tyr Glu Thr Lys Phe Ala Asp Asp Leu Arg Lys Ala Ile
545 550 555 560
Tyr Ser Leu Arg Asn Glu Thr Phe His Phe Thr Thr Leu Asn Lys Gly
565 570 575
Ser Phe Asp Trp Asn Ala Lys Leu Ile Gly Asp Met Phe Ser His Glu
580 585 590
Ala Ala Thr Gly Ile Ala Val Glu Arg Thr Arg Phe Tyr Ser Asn Asn

595 600 605

Leu Pro Met Phe Tyr Arg Glu Ser Asp Leu Lys Arg Ile Met Asp His
610 615 620
Leu Tyr Asn Thr Tyr His Pro Arg Ala Ser Gln Val Pro Ser Phe Asn
625 630 635 640
Ser Val Phe Val Arg Lys Asn Phe Arg Leu Phe Leu Ser Asn Thr Leu
645 650 655
Asn Thr Asn Thr Ser Phe Asp Thr Glu Val Tyr Gln Lys Trp Glu Ser
660 665 670

Gly Val Tyr Tyr Leu Phe Lys Glu Ile Tyr Tyr Asn Ser Phe Leu Pro

675 680 685

Ser Gly Asp Ala His His Leu Phe Phe Glu Gly Leu Arg Arg Ile Arg

690 695 700

Lys Glu Ala Asp Asn Leu Pro Ile Val Gly Lys Glu Ala Lys Lys Arg

705 710 715 720

Asn Ala Val Gln Asp Phe Gly Arg Arg Cys Asp Glu Leu Lys Asn Leu

725 730 735

Ser Leu Ser Ala Ile Cys Gln Met Ile Met Thr Glu Tyr Asn Glu Gln

740 745 750

Asn Asn Gly Asn Arg Lys Val Lys Ser Thr Arg Glu Asp Lys Arg Lys

755 760 765

Pro Asp Ile Phe Gln His Tyr Lys Met Leu Leu Leu Arg Thr Leu Gln

770 775 780

Glu Ala Phe Ala Ile Tyr Ile Arg Arg Glu Glu Phe Lys Phe Ile Phe

785 790 795 800

Asp Leu Pro Lys Thr Leu Tyr Val Met Lys Pro Val Glu Glu Phe Leu

805 810 815

Pro Asn Trp Lys Ser Gly Met Phe Asp Ser Leu Val Glu Arg Val Lys

820 825 830

Gln Ser Pro Asp Leu Gln Arg Trp Tyr Val Leu Cys Lys Phe Leu Asn

835 840 845

Gly Arg Leu Leu Asn Gln Leu Ser Gly Val Ile Arg Ser Tyr Ile Gln

850 855 860

Phe Ala Gly Asp Ile Gln Arg Arg Ala Lys Ala Asn His Asn Arg Leu

865 870 875 880

Tyr Met Asp Asn Thr Gln Arg Val Glu Tyr Tyr Ser Asn Val Leu Glu

885 890 895

Val Val Asp Phe Cys Ile Lys Gly Thr Ser Arg Phe Ser Asn Val Phe

900 905 910

Ser Asp Tyr Phe Arg Asp Glu Asp Ala Tyr Ala Asp Tyr Leu Asp Asn

915 920 925

Tyr Leu Gln Phe Lys Asp Glu Lys Ile Ala Glu Val Ser Ser Phe Ala

930 935 940

Ala Leu Lys Thr Phe Cys Asn Glu Glu Glu Val Lys Ala Gly Ile Tyr

945 950 955 960

Met Asp Gly Glu Asn Pro Val Met Gln Arg Asn Ile Val Met Ala Lys
965 970 975
Leu Phe Gly Pro Asp Glu Val Leu Lys Asn Val Val Pro Lys Val Thr
980 985 990
Arg Glu Glu Ile Glu Glu Tyr Tyr Gln Leu Glu Lys Gln Ile Ala Pro
995 1000 1005
Tyr Arg Gln Asn Gly Tyr Cys Lys Ser Glu Glu Asp Gln Lys Lys Leu
1010 1015 1020
Leu Arg Phe Gln Arg Ile Lys Asn Arg Val Glu Phe Gln Thr Ile Thr

1025 1030 1035 1040
Glu Phe Ser Glu Ile Ile Asn Glu Leu Leu Gly Gln Leu Ile Ser Trp
1045 1050 1055
Ser Phe Leu Arg Glu Arg Asp Leu Leu Tyr Phe Gln Leu Gly Phe His
1060 1065 1070
Tyr Leu Cys Leu His Asn Asp Thr Glu Lys Pro Ala Glu Tyr Lys Glu
1075 1080 1085
Ile Ser Arg Glu Asp Gly Thr Val Ile Arg Asn Ala Ile Leu His Gln
1090 1095 1100

Val Ala Ala Met Tyr Val Gly Gly Leu Pro Val Tyr Thr Leu Ala Asp
1105 1110 1115 1120
Lys Lys Leu Ala Ala Phe Glu Lys Gly Glu Ala Asp Cys Lys Leu Ser
1125 1130 1135
Ile Ser Lys Asp Thr Ala Gly Ala Gly Lys Lys Ile Lys Asp Phe Phe
1140 1145 1150
Arg Tyr Ser Lys Tyr Val Leu Ile Lys Asp Arg Met Leu Thr Asp Gln
1155 1160 1165
Asn Gln Lys Tyr Thr Ile Tyr Leu Ala Gly Leu Glu Leu Phe Glu Asn

1170 1175 1180
Thr Asp Glu His Asp Asn Ile Thr Asp Val Arg Lys Tyr Val Asp His
1185 1190 1195 1200

Phe Lys Tyr Tyr Ala Thr Ser Asp Glu Asn Ala Met Ser Ile Leu Asp
1205 1210 1215
Leu Tyr Ser Glu Ile His Asp Arg Phe Phe Thr Tyr Asp Met Lys Tyr
1220 1225 1230
Gln Lys Asn Val Ala Asn Met Leu Glu Asn Ile Leu Leu Arg His Phe
1235 1240 1245

Val Leu Ile Arg Pro Glu Phe Phe Thr Gly Ser Lys Lys Val Gly Glu
1250 1255 1260
Gly Lys Lys Ile Thr Cys Lys Ala Arg Ala Gln Ile Glu Ile Ala Glu
1265 1270 1275 1280
Asn Gly Met Arg Ser Glu Asp Phe Thr Tyr Lys Leu Ser Asp Gly Lys
1285 1290 1295
Lys Asn Ile Ser Thr Cys Met Ile Ala Ala Arg Asp Gln Lys Tyr Leu
1300 1305 1310
Asn Thr Val Ala Arg Leu Leu Tyr Tyr Pro His Glu Ala Lys Lys Ser

1315 1320 1325
Ile Val Asp Thr Arg Glu Lys Lys Asn Asn Lys Lys Thr Asn Arg Gly
1330 1335 1340
Asp Gly Thr Phe Asn Lys Gln Lys Gly Thr Ala Arg Lys Glu Lys Asp
1345 1350 1355 1360
Asn Gly Pro Arg Glu Phe Asn Asp Thr Gly Phe Ser Asn Thr Pro Phe
1365 1370 1375
Ala Gly Phe Asp Pro Phe Arg Asn Ser
1380 1385

<210> 122

<211> 1175
<212> PRT
<213> Carnobacterium gallinarum
<400> 122

Met Arg Ile Thr Lys Val Lys Ile Lys Leu Asp Asn Lys Leu Tyr Gln
1 5 10 15
Val Thr Met Gln Lys Glu Glu Lys Tyr Gly Thr Leu Lys Leu Asn Glu

20 25 30

Glu Ser Arg Lys Ser Thr Ala Glu Ile Leu Arg Leu Lys Lys Ala Ser

35 40 45

Phe Asn Lys Ser Phe His Ser Lys Thr Ile Asn Ser Gln Lys Glu Asn

50 55 60

Lys Asn Ala Thr Ile Lys Lys Asn Gly Asp Tyr Ile Ser Gln Ile Phe

65 70 75 80

Glu Lys Leu Val Gly Val Asp Thr Asn Lys Asn Ile Arg Lys Pro Lys

85 90 95

Met Ser Leu Thr Asp Leu Lys Asp Leu Pro Lys Lys Asp Leu Ala Leu

100 105 110

Phe Ile Lys Arg Lys Phe Lys Asn Asp Asp Ile Val Glu Ile Lys Asn

115 120 125

Leu Asp Leu Ile Ser Leu Phe Tyr Asn Ala Leu Gln Lys Val Pro Gly

130 135 140

Glu His Phe Thr Asp Glu Ser Trp Ala Asp Phe Cys Gln Glu Met Met

145 150 155 160

Pro Tyr Arg Glu Tyr Lys Asn Lys Phe Ile Glu Arg Lys Ile Ile Leu

165 170 175

Leu Ala Asn Ser Ile Glu Gln Asn Lys Gly Phe Ser Ile Asn Pro Glu

180 185 190

Thr Phe Ser Lys Arg Lys Arg Val Leu His Gln Trp Ala Ile Glu Val

195 200 205

Gln Glu Arg Gly Asp Phe Ser Ile Leu Asp Glu Lys Leu Ser Lys Leu

210 215 220

Ala Glu Ile Tyr Asn Phe Lys Lys Met Cys Lys Arg Val Gln Asp Glu

225 230 235 240

Leu Asn Asp Leu Glu Lys Ser Met Lys Lys Gly Lys Asn Pro Glu Lys

245 250 255

Glu Lys Glu Ala Tyr Lys Lys Gln Lys Asn Phe Lys Ile Lys Thr Ile

260 265 270

Trp Lys Asp Tyr Pro Tyr Lys Thr His Ile Gly Leu Ile Glu Lys Ile
275 280 285
Lys Glu Asn Glu Glu Leu Asn Gln Phe Asn Ile Glu Ile Gly Lys Tyr
290 295 300
Phe Glu His Tyr Phe Pro Ile Lys Lys Glu Arg Cys Thr Glu Asp Glu
305 310 315 320
Pro Tyr Tyr Leu Asn Ser Glu Thr Ile Ala Thr Thr Val Asn Tyr Gln
325 330 335
Leu Lys Asn Ala Leu Ile Ser Tyr Leu Met Gln Ile Gly Lys Tyr Lys

340 345 350
Gln Phe Gly Leu Glu Asn Gln Val Leu Asp Ser Lys Lys Leu Gln Glu
355 360 365
Ile Gly Ile Tyr Glu Gly Phe Gln Thr Lys Phe Met Asp Ala Cys Val
370 375 380
Phe Ala Thr Ser Ser Leu Lys Asn Ile Ile Glu Pro Met Arg Ser Gly
385 390 395 400
Asp Ile Leu Gly Lys Arg Glu Phe Lys Glu Ala Ile Ala Thr Ser Ser
405 410 415

Phe Val Asn Tyr His His Phe Phe Pro Tyr Phe Pro Phe Glu Leu Lys
420 425 430
Gly Met Lys Asp Arg Glu Ser Glu Leu Ile Pro Phe Gly Glu Gln Thr
435 440 445
Glu Ala Lys Gln Met Gln Asn Ile Trp Ala Leu Arg Gly Ser Val Gln
450 455 460
Gln Ile Arg Asn Glu Ile Phe His Ser Phe Asp Lys Asn Gln Lys Phe
465 470 475 480
Asn Leu Pro Gln Leu Asp Lys Ser Asn Phe Glu Phe Asp Ala Ser Glu

485 490 495
Asn Ser Thr Gly Lys Ser Gln Ser Tyr Ile Glu Thr Asp Tyr Lys Phe
500 505 510
Leu Phe Glu Ala Glu Lys Asn Gln Leu Glu Gln Phe Phe Ile Glu Arg
515 520 525

Ile Lys Ser Ser Gly Ala Leu Glu Tyr Tyr Pro Leu Lys Ser Leu Glu
530 535 540
Lys Leu Phe Ala Lys Lys Glu Met Lys Phe Ser Leu Gly Ser Gln Val
545 550 555 560

Val Ala Phe Ala Pro Ser Tyr Lys Lys Leu Val Lys Lys Gly His Ser
565 570 575
Tyr Gln Thr Ala Thr Glu Gly Thr Ala Asn Tyr Leu Gly Leu Ser Tyr
580 585 590
Tyr Asn Arg Tyr Glu Leu Lys Glu Glu Ser Phe Gln Ala Gln Tyr Tyr
595 600 605
Leu Leu Lys Leu Ile Tyr Gln Tyr Val Phe Leu Pro Asn Phe Ser Gln
610 615 620
Gly Asn Ser Pro Ala Phe Arg Glu Thr Val Lys Ala Ile Leu Arg Ile

625 630 635 640
Asn Lys Asp Glu Ala Arg Lys Lys Met Lys Lys Asn Lys Lys Phe Leu
645 650 655
Arg Lys Tyr Ala Phe Glu Gln Val Arg Glu Met Glu Phe Lys Glu Thr
660 665 670
Pro Asp Gln Tyr Met Ser Tyr Leu Gln Ser Glu Met Arg Glu Glu Lys
675 680 685
Val Arg Lys Ala Glu Lys Asn Asp Lys Gly Phe Glu Lys Asn Ile Thr
690 695 700

Met Asn Phe Glu Lys Leu Leu Met Gln Ile Phe Val Lys Gly Phe Asp
705 710 715 720
Val Phe Leu Thr Thr Phe Ala Gly Lys Glu Leu Leu Leu Ser Ser Glu
725 730 735
Glu Lys Val Ile Lys Glu Thr Glu Ile Ser Leu Ser Lys Lys Ile Asn
740 745 750
Glu Arg Glu Lys Thr Leu Lys Ala Ser Ile Gln Val Glu His Gln Leu
755 760 765
Val Ala Thr Asn Ser Ala Ile Ser Tyr Trp Leu Phe Cys Lys Leu Leu

770 775 780

Asp Ser Arg His Leu Asn Glu Leu Arg Asn Glu Met Ile Lys Phe Lys

785 790 795 800

Gln Ser Arg Ile Lys Phe Asn His Thr Gln His Ala Glu Leu Ile Gln

805 810 815

Asn Leu Leu Pro Ile Val Glu Leu Thr Ile Leu Ser Asn Asp Tyr Asp

820 825 830

Glu Lys Asn Asp Ser Gln Asn Val Asp Val Ser Ala Tyr Phe Glu Asp

835 840 845

Lys Ser Leu Tyr Glu Thr Ala Pro Tyr Val Gln Thr Asp Asp Arg Thr

850 855 860

Arg Val Ser Phe Arg Pro Ile Leu Lys Leu Glu Lys Tyr His Thr Lys

865 870 875 880

Ser Leu Ile Glu Ala Leu Leu Lys Asp Asn Pro Gln Phe Arg Val Ala

885 890 895

Ala Thr Asp Ile Gln Glu Trp Met His Lys Arg Glu Glu Ile Gly Glu

900 905 910

Leu Val Glu Lys Arg Lys Asn Leu His Thr Glu Trp Ala Glu Gly Gln

915 920 925

Gln Thr Leu Gly Ala Glu Lys Arg Glu Glu Tyr Arg Asp Tyr Cys Lys

930 935 940

Lys Ile Asp Arg Phe Asn Trp Lys Ala Asn Lys Val Thr Leu Thr Tyr

945 950 955 960

Leu Ser Gln Leu His Tyr Leu Ile Thr Asp Leu Leu Gly Arg Met Val

965 970 975

Gly Phe Ser Ala Leu Phe Glu Arg Asp Leu Val Tyr Phe Ser Arg Ser

980 985 990

Phe Ser Glu Leu Gly Gly Glu Thr Tyr His Ile Ser Asp Tyr Lys Asn

995 1000 1005

Leu Ser Gly Val Leu Arg Leu Asn Ala Glu Val Lys Pro Ile Lys Ile

1010 1015 1020

Lys Asn Ile Lys Val Ile Asp Asn Glu Glu Asn Pro Tyr Lys Gly Asn

1025 1030 1035 1040

Glu Pro Glu Val Lys Pro Phe Leu Asp Arg Leu His Ala Tyr Leu Glu

1045 1050 1055

Asn Val Ile Gly Ile Lys Ala Val His Gly Lys Ile Arg Asn Gln Thr

1060 1065 1070

Ala His Leu Ser Val Leu Gln Leu Glu Leu Ser Met Ile Glu Ser Met

1075 1080 1085

Asn Asn Leu Arg Asp Leu Met Ala Tyr Asp Arg Lys Leu Lys Asn Ala

1090 1095 1100

Val Thr Lys Ser Met Ile Lys Ile Leu Asp Lys His Gly Met Ile Leu

1105 1110 1115 1120

Lys Leu Lys Ile Asp Glu Asn His Lys Asn Phe Glu Ile Glu Ser Leu

1125 1130 1135

Ile Pro Lys Glu Ile Ile His Leu Lys Asp Lys Ala Ile Lys Thr Asn

1140 1145 1150

Gln Val Ser Glu Glu Tyr Cys Gln Leu Val Leu Ala Leu Leu Thr Thr

1155 1160 1165

Asn Pro Gly Asn Gln Leu Asn

1170 1175

<210> 123

<211> 1164

<212> PRT

<213> Carnobacterium gallinarum

<400> 123

Met Arg Met Thr Lys Val Lys Ile Asn Gly Ser Pro Val Ser Met Asn

1 5 10 15

Arg Ser Lys Leu Asn Gly His Leu Val Trp Asn Gly Thr Thr Asn Thr

20 25 30

Val Asn Ile Leu Thr Lys Lys Glu Gln Ser Phe Ala Ala Ser Phe Leu

35 40 45

Asn Lys Thr Leu Val Lys Ala Asp Gln Val Lys Gly Tyr Lys Val Leu

50 55 60

Ala Glu Asn Ile Phe Ile Ile Phe Glu Gln Leu Glu Lys Ser Asn Ser
65 70 75 80
Glu Lys Pro Ser Val Tyr Leu Asn Asn Ile Arg Arg Leu Lys Glu Ala

85 90 95
Gly Leu Lys Arg Phe Phe Lys Ser Lys Tyr His Glu Glu Ile Lys Tyr
100 105 110
Thr Ser Glu Lys Asn Gln Ser Val Pro Thr Lys Leu Asn Leu Ile Pro
115 120 125
Leu Phe Phe Asn Ala Val Asp Arg Ile Gln Glu Asp Lys Phe Asp Glu
130 135 140
Lys Asn Trp Ser Tyr Phe Cys Lys Glu Met Ser Pro Tyr Leu Asp Tyr
145 150 155 160

Lys Lys Ser Tyr Leu Asn Arg Lys Lys Glu Ile Leu Ala Asn Ser Ile
165 170 175
Gln Gln Asn Arg Gly Phe Ser Met Pro Thr Ala Glu Glu Pro Asn Leu
180 185 190
Leu Ser Lys Arg Lys Gln Leu Phe Gln Gln Trp Ala Met Lys Phe Gln
195 200 205
Glu Ser Pro Leu Ile Gln Gln Asn Asn Phe Ala Val Glu Gln Phe Asn
210 215 220
Lys Glu Phe Ala Asn Lys Ile Asn Glu Leu Ala Ala Val Tyr Asn Val

225 230 235 240
Asp Glu Leu Cys Thr Ala Ile Thr Glu Lys Leu Met Asn Phe Asp Lys
245 250 255
Asp Lys Ser Asn Lys Thr Arg Asn Phe Glu Ile Lys Lys Leu Trp Lys
260 265 270
Gln His Pro His Asn Lys Asp Lys Ala Leu Ile Lys Leu Phe Asn Gln
275 280 285
Glu Gly Asn Glu Ala Leu Asn Gln Phe Asn Ile Glu Leu Gly Lys Tyr
290 295 300

Phe Glu His Tyr Phe Pro Lys Thr Gly Lys Lys Glu Ser Ala Glu Ser

305 310 315 320

Tyr Tyr Leu Asn Pro Gln Thr Ile Ile Lys Thr Val Gly Tyr Gln Leu

325 330 335

Arg Asn Ala Phe Val Gln Tyr Leu Leu Gln Val Gly Lys Leu His Gln

340 345 350

Tyr Asn Lys Gly Val Leu Asp Ser Gln Thr Leu Gln Glu Ile Gly Met

355 360 365

Tyr Glu Gly Phe Gln Thr Lys Phe Met Asp Ala Cys Val Phe Ala Ser

370 375 380

Ser Ser Leu Arg Asn Ile Ile Gln Ala Thr Thr Asn Glu Asp Ile Leu

385 390 395 400

Thr Arg Glu Lys Phe Lys Lys Glu Leu Glu Lys Asn Val Glu Leu Lys

405 410 415

His Asp Leu Phe Phe Lys Thr Glu Ile Val Glu Glu Arg Asp Glu Asn

420 425 430

Pro Ala Lys Lys Ile Ala Met Thr Pro Asn Glu Leu Asp Leu Trp Ala

435 440 445

Ile Arg Gly Ala Val Gln Arg Val Arg Asn Gln Ile Phe His Gln Gln

450 455 460

Ile Asn Lys Arg His Glu Pro Asn Gln Leu Lys Val Gly Ser Phe Glu

465 470 475 480

Asn Gly Asp Leu Gly Asn Val Ser Tyr Gln Lys Thr Ile Tyr Gln Lys

485 490 495

Leu Phe Asp Ala Glu Ile Lys Asp Ile Glu Ile Tyr Phe Ala Glu Lys

500 505 510

Ile Lys Ser Ser Gly Ala Leu Glu Gln Tyr Ser Met Lys Asp Leu Glu

515 520 525

Lys Leu Phe Ser Asn Lys Glu Leu Thr Leu Ser Leu Gly Gly Gln Val

530 535 540

Val Ala Phe Ala Pro Ser Tyr Lys Lys Leu Tyr Lys Gln Gly Tyr Phe

545 550 555 560

Tyr Gln Asn Glu Lys Thr Ile Glu Leu Glu Gln Phe Thr Asp Tyr Asp

565 570 575

Phe Ser Asn Asp Val Phe Lys Ala Asn Tyr Tyr Leu Ile Lys Leu Ile

580 585 590

Tyr His Tyr Val Phe Leu Pro Gln Phe Ser Gln Ala Asn Asn Lys Leu

595 600 605

Phe Lys Asp Thr Val His Tyr Val Ile Gln Gln Asn Lys Glu Leu Asn

610 615 620

Thr Thr Glu Lys Asp Lys Lys Asn Asn Lys Lys Ile Arg Lys Tyr Ala

625 630 635 640

Phe Glu Gln Val Lys Leu Met Lys Asn Glu Ser Pro Glu Lys Tyr Met

645 650 655

Gln Tyr Leu Gln Arg Glu Met Gln Glu Glu Arg Thr Ile Lys Glu Ala

660 665 670

Lys Lys Thr Asn Glu Glu Lys Pro Asn Tyr Asn Phe Glu Lys Leu Leu

675 680 685

Ile Gln Ile Phe Ile Lys Gly Phe Asp Thr Phe Leu Arg Asn Phe Asp

690 695 700

Leu Asn Leu Asn Pro Ala Glu Glu Leu Val Gly Thr Val Lys Glu Lys

705 710 715 720

Ala Glu Gly Leu Arg Lys Arg Lys Glu Arg Ile Ala Lys Ile Leu Asn

725 730 735

Val Asp Glu Gln Ile Lys Thr Gly Asp Glu Glu Ile Ala Phe Trp Ile

740 745 750

Phe Ala Lys Leu Leu Asp Ala Arg His Leu Ser Glu Leu Arg Asn Glu

755 760 765

Met Ile Lys Phe Lys Gln Ser Ser Val Lys Lys Gly Leu Ile Lys Asn

770 775 780

Gly Asp Leu Ile Glu Gln Met Gln Pro Ile Leu Glu Leu Cys Ile Leu

785 790 795 800

Ser Asn Asp Ser Glu Ser Met Glu Lys Glu Ser Phe Asp Lys Ile Glu

805 810 815

Val Phe Leu Glu Lys Val Glu Leu Ala Lys Asn Glu Pro Tyr Met Gln
820 825 830
Glu Asp Lys Leu Thr Pro Val Lys Phe Arg Phe Met Lys Gln Leu Glu
835 840 845
Lys Tyr Gln Thr Arg Asn Phe Ile Glu Asn Leu Val Ile Glu Asn Pro
850 855 860
Glu Phe Lys Val Ser Glu Lys Ile Val Leu Asn Trp His Glu Glu Lys
865 870 875 880

Glu Lys Ile Ala Asp Leu Val Asp Lys Arg Thr Lys Leu His Glu Glu
885 890 895
Trp Ala Ser Lys Ala Arg Glu Ile Glu Glu Tyr Asn Glu Lys Ile Lys
900 905 910
Lys Asn Lys Ser Lys Lys Leu Asp Lys Pro Ala Glu Phe Ala Lys Phe
915 920 925
Ala Glu Tyr Lys Ile Ile Cys Glu Ala Ile Glu Asn Phe Asn Arg Leu
930 935 940
Asp His Lys Val Arg Leu Thr Tyr Leu Lys Asn Leu His Tyr Leu Met

945 950 955 960
Ile Asp Leu Met Gly Arg Met Val Gly Phe Ser Val Leu Phe Glu Arg
965 970 975
Asp Phe Val Tyr Met Gly Arg Ser Tyr Ser Ala Leu Lys Lys Gln Ser
980 985 990
Ile Tyr Leu Asn Asp Tyr Asp Thr Phe Ala Asn Ile Arg Asp Trp Glu
995 1000 1005
Val Asn Glu Asn Lys His Leu Phe Gly Thr Ser Ser Ser Asp Leu Thr
1010 1015 1020

Phe Gln Glu Thr Ala Glu Phe Lys Asn Leu Lys Lys Pro Met Glu Asn
1025 1030 1035 1040
Gln Leu Lys Ala Leu Leu Gly Val Thr Asn His Ser Phe Glu Ile Arg
1045 1050 1055
Asn Asn Ile Ala His Leu His Val Leu Arg Asn Asp Gly Lys Gly Glu
1060 1065 1070

Gly Val Ser Leu Leu Ser Cys Met Asn Asp Leu Arg Lys Leu Met Ser

1075 1080 1085

Tyr Asp Arg Lys Leu Lys Asn Ala Val Thr Lys Ala Ile Ile Lys Ile

1090 1095 1100

Leu Asp Lys His Gly Met Ile Leu Lys Leu Thr Asn Asn Asp His Thr

1105 1110 1115 1120

Lys Pro Phe Glu Ile Glu Ser Leu Lys Pro Lys Lys Ile Ile His Leu

1125 1130 1135

Glu Lys Ser Asn His Ser Phe Pro Met Asp Gln Val Ser Gln Glu Tyr

1140 1145 1150

Cys Asp Leu Val Lys Lys Met Leu Val Phe Thr Asn

1155 1160

<210> 124

<211> 1154

<212> PRT

<213> Paludibacter propionicigenes

<400> 124

Met Arg Val Ser Lys Val Lys Val Lys Asp Gly Gly Lys Asp Lys Met

1 5 10 15

Val Leu Val His Arg Lys Thr Thr Gly Ala Gln Leu Val Tyr Ser Gly

20 25 30

Gln Pro Val Ser Asn Glu Thr Ser Asn Ile Leu Pro Glu Lys Lys Arg

35 40 45

Gln Ser Phe Asp Leu Ser Thr Leu Asn Lys Thr Ile Ile Lys Phe Asp

50 55 60

Thr Ala Lys Lys Gln Lys Leu Asn Val Asp Gln Tyr Lys Ile Val Glu

65 70 75 80

Lys Ile Phe Lys Tyr Pro Lys Gln Glu Leu Pro Lys Gln Ile Lys Ala

85 90 95

Glu Glu Ile Leu Pro Phe Leu Asn His Lys Phe Gln Glu Pro Val Lys

100 105 110

Tyr Trp Lys Asn Gly Lys Glu Glu Ser Phe Asn Leu Thr Leu Leu Ile

115 120 125

Val Glu Ala Val Gln Ala Gln Asp Lys Arg Lys Leu Gln Pro Tyr Tyr
130 135 140
Asp Trp Lys Thr Trp Tyr Ile Gln Thr Lys Ser Asp Leu Leu Lys Lys
145 150 155 160
Ser Ile Glu Asn Asn Arg Ile Asp Leu Thr Glu Asn Leu Ser Lys Arg
165 170 175
Lys Lys Ala Leu Leu Ala Trp Glu Thr Glu Phe Thr Ala Ser Gly Ser
180 185 190
Ile Asp Leu Thr His Tyr His Lys Val Tyr Met Thr Asp Val Leu Cys

195 200 205
Lys Met Leu Gln Asp Val Lys Pro Leu Thr Asp Asp Lys Gly Lys Ile
210 215 220
Asn Thr Asn Ala Tyr His Arg Gly Leu Lys Lys Ala Leu Gln Asn His
225 230 235 240
Gln Pro Ala Ile Phe Gly Thr Arg Glu Val Pro Asn Glu Ala Asn Arg
245 250 255
Ala Asp Asn Gln Leu Ser Ile Tyr His Leu Glu Val Val Lys Tyr Leu
260 265 270

Glu His Tyr Phe Pro Ile Lys Thr Ser Lys Arg Arg Asn Thr Ala Asp
275 280 285
Asp Ile Ala His Tyr Leu Lys Ala Gln Thr Leu Lys Thr Thr Ile Glu
290 295 300
Lys Gln Leu Val Asn Ala Ile Arg Ala Asn Ile Ile Gln Gln Gly Lys
305 310 315 320
Thr Asn His His Glu Leu Lys Ala Asp Thr Thr Ser Asn Asp Leu Ile
325 330 335
Arg Ile Lys Thr Asn Glu Ala Phe Val Leu Asn Leu Thr Gly Thr Cys

340 345 350
Ala Phe Ala Ala Asn Asn Ile Arg Asn Met Val Asp Asn Glu Gln Thr
355 360 365

Asn Asp Ile Leu Gly Lys Gly Asp Phe Ile Lys Ser Leu Leu Lys Asp
370 375 380
Asn Thr Asn Ser Gln Leu Tyr Ser Phe Phe Phe Gly Glu Gly Leu Ser
385 390 395 400
Thr Asn Lys Ala Glu Lys Glu Thr Gln Leu Trp Gly Ile Arg Gly Ala
405 410 415

Val Gln Gln Ile Arg Asn Asn Val Asn His Tyr Lys Lys Asp Ala Leu
420 425 430
Lys Thr Val Phe Asn Ile Ser Asn Phe Glu Asn Pro Thr Ile Thr Asp
435 440 445
Pro Lys Gln Gln Thr Asn Tyr Ala Asp Thr Ile Tyr Lys Ala Arg Phe
450 455 460
Ile Asn Glu Leu Glu Lys Ile Pro Glu Ala Phe Ala Gln Gln Leu Lys
465 470 475 480
Thr Gly Gly Ala Val Ser Tyr Tyr Thr Ile Glu Asn Leu Lys Ser Leu
485 490 495
Leu Thr Thr Phe Gln Phe Ser Leu Cys Arg Ser Thr Ile Pro Phe Ala
500 505 510
Pro Gly Phe Lys Lys Val Phe Asn Gly Gly Ile Asn Tyr Gln Asn Ala
515 520 525
Lys Gln Asp Glu Ser Phe Tyr Glu Leu Met Leu Glu Gln Tyr Leu Arg
530 535 540
Lys Glu Asn Phe Ala Glu Glu Ser Tyr Asn Ala Arg Tyr Phe Met Leu
545 550 555 560

Lys Leu Ile Tyr Asn Asn Leu Phe Leu Pro Gly Phe Thr Thr Asp Arg
565 570 575
Lys Ala Phe Ala Asp Ser Val Gly Phe Val Gln Met Gln Asn Lys Lys
580 585 590
Gln Ala Glu Lys Val Asn Pro Arg Lys Lys Glu Ala Tyr Ala Phe Glu
595 600 605
Ala Val Arg Pro Met Thr Ala Ala Asp Ser Ile Ala Asp Tyr Met Ala
610 615 620

Tyr Val Gln Ser Glu Leu Met Gln Glu Gln Asn Lys Lys Glu Glu Lys

625 630 635 640

Val Ala Glu Glu Thr Arg Ile Asn Phe Glu Lys Phe Val Leu Gln Val

645 650 655

Phe Ile Lys Gly Phe Asp Ser Phe Leu Arg Ala Lys Glu Phe Asp Phe

660 665 670

Val Gln Met Pro Gln Pro Gln Leu Thr Ala Thr Ala Ser Asn Gln Gln

675 680 685

Lys Ala Asp Lys Leu Asn Gln Leu Glu Ala Ser Ile Thr Ala Asp Cys

690 695 700

Lys Leu Thr Pro Gln Tyr Ala Lys Ala Asp Asp Ala Thr His Ile Ala

705 710 715 720

Phe Tyr Val Phe Cys Lys Leu Leu Asp Ala Ala His Leu Ser Asn Leu

725 730 735

Arg Asn Glu Leu Ile Lys Phe Arg Glu Ser Val Asn Glu Phe Lys Phe

740 745 750

His His Leu Leu Glu Ile Ile Glu Ile Cys Leu Leu Ser Ala Asp Val

755 760 765

Val Pro Thr Asp Tyr Arg Asp Leu Tyr Ser Ser Glu Ala Asp Cys Leu

770 775 780

Ala Arg Leu Arg Pro Phe Ile Glu Gln Gly Ala Asp Ile Thr Asn Trp

785 790 795 800

Ser Asp Leu Phe Val Gln Ser Asp Lys His Ser Pro Val Ile His Ala

805 810 815

Asn Ile Glu Leu Ser Val Lys Tyr Gly Thr Thr Lys Leu Leu Glu Gln

820 825 830

Ile Ile Asn Lys Asp Thr Gln Phe Lys Thr Thr Glu Ala Asn Phe Thr

835 840 845

Ala Trp Asn Thr Ala Gln Lys Ser Ile Glu Gln Leu Ile Lys Gln Arg

850 855 860

Glu Asp His His Glu Gln Trp Val Lys Ala Lys Asn Ala Asp Asp Lys

865 870 875 880

Glu Lys Gln Glu Arg Lys Arg Glu Lys Ser Asn Phe Ala Gln Lys Phe

885 890 895

Ile Glu Lys His Gly Asp Asp Tyr Leu Asp Ile Cys Asp Tyr Ile Asn

900 905 910

Thr Tyr Asn Trp Leu Asp Asn Lys Met His Phe Val His Leu Asn Arg

915 920 925

Leu His Gly Leu Thr Ile Glu Leu Leu Gly Arg Met Ala Gly Phe Val

930 935 940

Ala Leu Phe Asp Arg Asp Phe Gln Phe Phe Asp Glu Gln Gln Ile Ala

945 950 955 960

Asp Glu Phe Lys Leu His Gly Phe Val Asn Leu His Ser Ile Asp Lys

965 970 975

Lys Leu Asn Glu Val Pro Thr Lys Lys Ile Lys Glu Ile Tyr Asp Ile

980 985 990

Arg Asn Lys Ile Ile Gln Ile Asn Gly Asn Lys Ile Asn Glu Ser Val

995 1000 1005

Arg Ala Asn Leu Ile Gln Phe Ile Ser Ser Lys Arg Asn Tyr Tyr Asn

1010 1015 1020

Asn Ala Phe Leu His Val Ser Asn Asp Glu Ile Lys Glu Lys Gln Met

1025 1030 1035 1040

Tyr Asp Ile Arg Asn His Ile Ala His Phe Asn Tyr Leu Thr Lys Asp

1045 1050 1055

Ala Ala Asp Phe Ser Leu Ile Asp Leu Ile Asn Glu Leu Arg Glu Leu

1060 1065 1070

Leu His Tyr Asp Arg Lys Leu Lys Asn Ala Val Ser Lys Ala Phe Ile

1075 1080 1085

Asp Leu Phe Asp Lys His Gly Met Ile Leu Lys Leu Lys Leu Asn Ala

1090 1095 1100

Asp His Lys Leu Lys Val Glu Ser Leu Glu Pro Lys Lys Ile Tyr His

1105 1110 1115 1120

Leu Gly Ser Ser Ala Lys Asp Lys Pro Glu Tyr Gln Tyr Cys Thr Asn

1125 1130 1135

Gln Val Met Met Ala Tyr Cys Asn Met Cys Arg Ser Leu Leu Glu Met
1140 1145 1150
Lys Lys

<210> 125
<211> 970
<212> PRT
<213> Listeria weihenstephanensis
<400> 125
Met Leu Ala Leu Leu His Gln Glu Val Pro Ser Gln Lys Leu His Asn
1 5 10 15
Leu Lys Ser Leu Asn Thr Glu Ser Leu Thr Lys Leu Phe Lys Pro Lys
20 25 30
Phe Gln Asn Met Ile Ser Tyr Pro Pro Ser Lys Gly Ala Glu His Val

35 40 45
Gln Phe Cys Leu Thr Asp Ile Ala Val Pro Ala Ile Arg Asp Leu Asp
50 55 60
Glu Ile Lys Pro Asp Trp Gly Ile Phe Phe Glu Lys Leu Lys Pro Tyr
65 70 75 80
Thr Asp Trp Ala Glu Ser Tyr Ile His Tyr Lys Gln Thr Thr Ile Gln
85 90 95
Lys Ser Ile Glu Gln Asn Lys Ile Gln Ser Pro Asp Ser Pro Arg Lys
100 105 110

Leu Val Leu Gln Lys Tyr Val Thr Ala Phe Leu Asn Gly Glu Pro Leu
115 120 125
Gly Leu Asp Leu Val Ala Lys Lys Tyr Lys Leu Ala Asp Leu Ala Glu
130 135 140
Ser Phe Lys Val Val Asp Leu Asn Glu Asp Lys Ser Ala Asn Tyr Lys
145 150 155 160
Ile Lys Ala Cys Leu Gln Gln His Gln Arg Asn Ile Leu Asp Glu Leu
165 170 175

Lys Glu Asp Pro Glu Leu Asn Gln Tyr Gly Ile Glu Val Lys Lys Tyr

180 185 190

Ile Gln Arg Tyr Phe Pro Ile Lys Arg Ala Pro Asn Arg Ser Lys His

195 200 205

Ala Arg Ala Asp Phe Leu Lys Lys Glu Leu Ile Glu Ser Thr Val Glu

210 215 220

Gln Gln Phe Lys Asn Ala Val Tyr His Tyr Val Leu Glu Gln Gly Lys

225 230 235 240

Met Glu Ala Tyr Glu Leu Thr Asp Pro Lys Thr Lys Asp Leu Gln Asp

245 250 255

Ile Arg Ser Gly Glu Ala Phe Ser Phe Lys Phe Ile Asn Ala Cys Ala

260 265 270

Phe Ala Ser Asn Asn Leu Lys Met Ile Leu Asn Pro Glu Cys Glu Lys

275 280 285

Asp Ile Leu Gly Lys Gly Asp Phe Lys Lys Asn Leu Pro Asn Ser Thr

290 295 300

Thr Gln Ser Asp Val Val Lys Lys Met Ile Pro Phe Phe Ser Asp Glu

305 310 315 320

Ile Gln Asn Val Asn Phe Asp Glu Ala Ile Trp Ala Ile Arg Gly Ser

325 330 335

Ile Gln Gln Ile Arg Asn Glu Val Tyr His Cys Lys Lys His Ser Trp

340 345 350

Lys Ser Ile Leu Lys Ile Lys Gly Phe Glu Phe Glu Pro Asn Asn Met

355 360 365

Lys Tyr Thr Asp Ser Asp Met Gln Lys Leu Met Asp Lys Asp Ile Ala

370 375 380

Lys Ile Pro Asp Phe Ile Glu Glu Lys Leu Lys Ser Ser Gly Ile Ile

385 390 395 400

Arg Phe Tyr Ser His Asp Lys Leu Gln Ser Ile Trp Glu Met Lys Gln

405 410 415

Gly Phe Ser Leu Leu Thr Thr Asn Ala Pro Phe Val Pro Ser Phe Lys

420 425 430

Arg Val Tyr Ala Lys Gly His Asp Tyr Gln Thr Ser Lys Asn Arg Tyr

435 440 445

Tyr Asp Leu Gly Leu Thr Thr Phe Asp Ile Leu Glu Tyr Gly Glu Glu

450 455 460

Asp Phe Arg Ala Arg Tyr Phe Leu Thr Lys Leu Val Tyr Tyr Gln Gln

465 470 475 480

Phe Met Pro Trp Phe Thr Ala Asp Asn Asn Ala Phe Arg Asp Ala Ala

485 490 495

Asn Phe Val Leu Arg Leu Asn Lys Asn Arg Gln Gln Asp Ala Lys Ala

500 505 510

Phe Ile Asn Ile Arg Glu Val Glu Glu Gly Glu Met Pro Arg Asp Tyr

515 520 525

Met Gly Tyr Val Gln Gly Gln Ile Ala Ile His Glu Asp Ser Thr Glu

530 535 540

Asp Thr Pro Asn His Phe Glu Lys Phe Ile Ser Gln Val Phe Ile Lys

545 550 555 560

Gly Phe Asp Ser His Met Arg Ser Ala Asp Leu Lys Phe Ile Lys Asn

565 570 575

Pro Arg Asn Gln Gly Leu Glu Gln Ser Glu Ile Glu Glu Met Ser Phe

580 585 590

Asp Ile Lys Val Glu Pro Ser Phe Leu Lys Asn Lys Asp Asp Tyr Ile

595 600 605

Ala Phe Trp Thr Phe Cys Lys Met Leu Asp Ala Arg His Leu Ser Glu

610 615 620

Leu Arg Asn Glu Met Ile Lys Tyr Asp Gly His Leu Thr Gly Glu Gln

625 630 635 640

Glu Ile Ile Gly Leu Ala Leu Leu Gly Val Asp Ser Arg Glu Asn Asp

645 650 655

Trp Lys Gln Phe Phe Ser Ser Glu Arg Glu Tyr Glu Lys Ile Met Lys

660 665 670

Gly Tyr Val Gly Glu Glu Leu Tyr Gln Arg Glu Pro Tyr Arg Gln Ser

675 680 685

Asp Gly Lys Thr Pro Ile Leu Phe Arg Gly Val Glu Gln Ala Arg Lys
690 695 700
Tyr Gly Thr Glu Thr Val Ile Gln Arg Leu Phe Asp Ala Ser Pro Glu
705 710 715 720
Phe Lys Val Ser Lys Cys Asn Ile Thr Glu Trp Glu Arg Gln Lys Glu
725 730 735
Thr Ile Glu Glu Thr Ile Glu Arg Arg Lys Glu Leu His Asn Glu Trp
740 745 750
Glu Lys Asn Pro Lys Lys Pro Gln Asn Asn Ala Phe Phe Lys Glu Tyr

755 760 765
Lys Glu Cys Cys Asp Ala Ile Asp Ala Tyr Asn Trp His Lys Asn Lys
770 775 780
Thr Thr Leu Val Tyr Val Asn Glu Leu His His Leu Leu Ile Glu Ile
785 790 795 800
Leu Gly Arg Tyr Val Gly Tyr Val Ala Ile Ala Asp Arg Asp Phe Gln
805 810 815
Cys Met Ala Asn Gln Tyr Phe Lys His Ser Gly Ile Thr Glu Arg Val
820 825 830

Glu Tyr Trp Gly Asp Asn Arg Leu Lys Ser Ile Lys Lys Leu Asp Thr
835 840 845
Phe Leu Lys Lys Glu Gly Leu Phe Val Ser Glu Lys Asn Ala Arg Asn
850 855 860
His Ile Ala His Leu Asn Tyr Leu Ser Leu Lys Ser Glu Cys Thr Leu
865 870 875 880
Leu Tyr Leu Ser Glu Arg Leu Arg Glu Ile Phe Lys Tyr Asp Arg Lys
885 890 895
Leu Lys Asn Ala Val Ser Lys Ser Leu Ile Asp Ile Leu Asp Arg His

900 905 910
Gly Met Ser Val Val Phe Ala Asn Leu Lys Glu Asn Lys His Arg Leu
915 920 925

Val Ile Lys Ser Leu Glu Pro Lys Lys Leu Arg His Leu Gly Glu Lys
930 935 940
Lys Ile Asp Asn Gly Tyr Ile Glu Thr Asn Gln Val Ser Glu Glu Tyr
945 950 955 960
Cys Gly Ile Val Lys Arg Leu Leu Glu Ile
965 970

<210
> 126

<211> 1051

<212> PRT

<213> Listeria newyorkensis

<400> 126

Met Lys Ile Thr Lys Met Arg Val Asp Gly Arg Thr Ile Val Met Glu
1 5 10 15
Arg Thr Ser Lys Glu Gly Gln Leu Gly Tyr Glu Gly Ile Asp Gly Asn
20 25 30
Lys Thr Thr Glu Ile Ile Phe Asp Lys Lys Lys Glu Ser Phe Tyr Lys
35 40 45
Ser Ile Leu Asn Lys Thr Val Arg Lys Pro Asp Glu Lys Glu Lys Asn

50 55 60
Arg Arg Lys Gln Ala Ile Asn Lys Ala Ile Asn Lys Glu Ile Thr Glu
65 70 75 80
Leu Met Leu Ala Val Leu His Gln Glu Val Pro Ser Gln Lys Leu His
85 90 95
Asn Leu Lys Ser Leu Asn Thr Glu Ser Leu Thr Lys Leu Phe Lys Pro
100 105 110
Lys Phe Gln Asn Met Ile Ser Tyr Pro Pro Ser Lys Gly Ala Glu His
115 120 125

Val Gln Phe Cys Leu Thr Asp Ile Ala Val Pro Ala Ile Arg Asp Leu
130 135 140
Asp Glu Ile Lys Pro Asp Trp Gly Ile Phe Phe Glu Lys Leu Lys Pro
145 150 155 160
Tyr Thr Asp Trp Ala Glu Ser Tyr Ile His Tyr Lys Gln Thr Thr Ile

165 170 175

Gln Lys Ser Ile Glu Gln Asn Lys Ile Gln Ser Pro Asp Ser Pro Arg

180 185 190

Lys Leu Val Leu Gln Lys Tyr Val Thr Ala Phe Leu Asn Gly Glu Pro

195 200 205

Leu Gly Leu Asp Leu Val Ala Lys Lys Tyr Lys Leu Ala Asp Leu Ala

210 215 220

Glu Ser Phe Lys Leu Val Asp Leu Asn Glu Asp Lys Ser Ala Asn Tyr

225 230 235 240

Lys Ile Lys Ala Cys Leu Gln Gln His Gln Arg Asn Ile Leu Asp Glu

245 250 255

Leu Lys Glu Asp Pro Glu Leu Asn Gln Tyr Gly Ile Glu Val Lys Lys

260 265 270

Tyr Ile Gln Arg Tyr Phe Pro Ile Lys Arg Ala Pro Asn Arg Ser Lys

275 280 285

His Ala Arg Ala Asp Phe Leu Lys Lys Glu Leu Ile Glu Ser Thr Val

290 295 300

Glu Gln Gln Phe Lys Asn Ala Val Tyr His Tyr Val Leu Glu Gln Gly

305 310 315 320

Lys Met Glu Ala Tyr Glu Leu Thr Asp Pro Lys Thr Lys Asp Leu Gln

325 330 335

Asp Ile Arg Ser Gly Glu Ala Phe Ser Phe Lys Phe Ile Asn Ala Cys

340 345 350

Ala Phe Ala Ser Asn Asn Leu Lys Met Ile Leu Asn Pro Glu Cys Glu

355 360 365

Lys Asp Ile Leu Gly Lys Gly Asn Phe Lys Lys Asn Leu Pro Asn Ser

370 375 380

Thr Thr Arg Ser Asp Val Val Lys Lys Met Ile Pro Phe Phe Ser Asp

385 390 395 400

Glu Leu Gln Asn Val Asn Phe Asp Glu Ala Ile Trp Ala Ile Arg Gly

405 410 415

Ser Ile Gln Gln Ile Arg Asn Glu Val Tyr His Cys Lys Lys His Ser
420 425 430
Trp Lys Ser Ile Leu Lys Ile Lys Gly Phe Glu Phe Glu Pro Asn Asn
435 440 445
Met Lys Tyr Ala Asp Ser Asp Met Gln Lys Leu Met Asp Lys Asp Ile
450 455 460
Ala Lys Ile Pro Glu Phe Ile Glu Glu Lys Leu Lys Ser Ser Gly Val
465 470 475 480
Val Arg Phe Tyr Arg His Asp Glu Leu Gln Ser Ile Trp Glu Met Lys

485 490 495
Gln Gly Phe Ser Leu Leu Thr Thr Asn Ala Pro Phe Val Pro Ser Phe
500 505 510
Lys Arg Val Tyr Ala Lys Gly His Asp Tyr Gln Thr Ser Lys Asn Arg
515 520 525
Tyr Tyr Asn Leu Asp Leu Thr Thr Phe Asp Ile Leu Glu Tyr Gly Glu
530 535 540
Glu Asp Phe Arg Ala Arg Tyr Phe Leu Thr Lys Leu Val Tyr Tyr Gln
545 550 555 560

Gln Phe Met Pro Trp Phe Thr Ala Asp Asn Asn Ala Phe Arg Asp Ala
565 570 575
Ala Asn Phe Val Leu Arg Leu Asn Lys Asn Arg Gln Gln Asp Ala Lys
580 585 590
Ala Phe Ile Asn Ile Arg Glu Val Glu Glu Gly Glu Met Pro Arg Asp
595 600 605
Tyr Met Gly Tyr Val Gln Gly Gln Ile Ala Ile His Glu Asp Ser Ile
610 615 620
Glu Asp Thr Pro Asn His Phe Glu Lys Phe Ile Ser Gln Val Phe Ile

625 630 635 640
Lys Gly Phe Asp Arg His Met Arg Ser Ala Asn Leu Lys Phe Ile Lys
645 650 655
Asn Pro Arg Asn Gln Gly Leu Glu Gln Ser Glu Ile Glu Glu Met Ser
660 665 670

Phe Asp Ile Lys Val Glu Pro Ser Phe Leu Lys Asn Lys Asp Asp Tyr
675 680 685
Ile Ala Phe Trp Ile Phe Cys Lys Met Leu Asp Ala Arg His Leu Ser
690 695 700

Glu Leu Arg Asn Glu Met Ile Lys Tyr Asp Gly His Leu Thr Gly Glu
705 710 715 720
Gln Glu Ile Ile Gly Leu Ala Leu Leu Gly Val Asp Ser Arg Glu Asn
725 730 735
Asp Trp Lys Gln Phe Phe Ser Ser Glu Arg Glu Tyr Glu Lys Ile Met
740 745 750
Lys Gly Tyr Val Val Glu Glu Leu Tyr Gln Arg Glu Pro Tyr Arg Gln
755 760 765
Ser Asp Gly Lys Thr Pro Ile Leu Phe Arg Gly Val Glu Gln Ala Arg

770 775 780
Lys Tyr Gly Thr Glu Thr Val Ile Gln Arg Leu Phe Asp Ala Asn Pro
785 790 795 800
Glu Phe Lys Val Ser Lys Cys Asn Leu Ala Glu Trp Glu Arg Gln Lys
805 810 815
Glu Thr Ile Glu Glu Thr Ile Lys Arg Arg Lys Glu Leu His Asn Glu
820 825 830
Trp Ala Lys Asn Pro Lys Lys Pro Gln Asn Asn Ala Phe Phe Lys Glu
835 840 845

Tyr Lys Glu Cys Cys Asp Ala Ile Asp Ala Tyr Asn Trp His Lys Asn
850 855 860
Lys Thr Thr Leu Ala Tyr Val Asn Glu Leu His His Leu Leu Ile Glu
865 870 875 880
Ile Leu Gly Arg Tyr Val Gly Tyr Val Ala Ile Ala Asp Arg Asp Phe
885 890 895
Gln Cys Met Ala Asn Gln Tyr Phe Lys His Ser Gly Ile Thr Glu Arg
900 905 910
Val Glu Tyr Trp Gly Asp Asn Arg Leu Lys Ser Ile Lys Lys Leu Asp

915 920 925

Thr Phe Leu Lys Lys Glu Gly Leu Phe Val Ser Glu Lys Asn Ala Arg

930 935 940

Asn His Ile Ala His Leu Asn Tyr Leu Ser Leu Lys Ser Glu Cys Thr

945 950 955 960

Leu Leu Tyr Leu Ser Glu Arg Leu Arg Glu Ile Phe Lys Tyr Asp Arg

965 970 975

Lys Leu Lys Asn Ala Val Ser Lys Ser Leu Ile Asp Ile Leu Asp Arg

980 985 990

His Gly Met Ser Val Val Phe Ala Asn Leu Lys Glu Asn Lys His Arg

995 1000 1005

Leu Val Ile Lys Ser Leu Glu Pro Lys Lys Leu Arg His Leu Gly Gly

1010 1015 1020

Lys Lys Ile Asp Gly Gly Tyr Ile Glu Thr Asn Gln Val Ser Glu Glu

1025 1030 1035 1040

Tyr Cys Gly Ile Val Lys Arg Leu Leu Glu Met

1045 1050

<210> 127

<211> 1152

<212> PRT

<213> Leptotrichia wadei

<400> 127

Met Lys Val Thr Lys Val Asp Gly Ile Ser His Lys Lys Tyr Ile Glu

1 5 10 15

Glu Gly Lys Leu Val Lys Ser Thr Ser Glu Glu Asn Arg Thr Ser Glu

20 25 30

Arg Leu Ser Glu Leu Leu Ser Ile Arg Leu Asp Ile Tyr Ile Lys Asn

35 40 45

Pro Asp Asn Ala Ser Glu Glu Glu Asn Arg Ile Arg Arg Glu Asn Leu

50 55 60

Lys Lys Phe Phe Ser Asn Lys Val Leu His Leu Lys Asp Ser Val Leu

65 70 75 80

Tyr Leu Lys Asn Arg Lys Glu Lys Asn Ala Val Gln Asp Lys Asn Tyr
85 90 95
Ser Glu Glu Asp Ile Ser Glu Tyr Asp Leu Lys Asn Lys Asn Ser Phe
100 105 110
Ser Val Leu Lys Lys Ile Leu Leu Asn Glu Asp Val Asn Ser Glu Glu
115 120 125
Leu Glu Ile Phe Arg Lys Asp Val Glu Ala Lys Leu Asn Lys Ile Asn
130 135 140

Ser Leu Lys Tyr Ser Phe Glu Glu Asn Lys Ala Asn Tyr Gln Lys Ile
145 150 155 160
Asn Glu Asn Asn Val Glu Lys Val Gly Gly Lys Ser Lys Arg Asn Ile
165 170 175
Ile Tyr Asp Tyr Tyr Arg Glu Ser Ala Lys Arg Asn Asp Tyr Ile Asn
180 185 190
Asn Val Gln Glu Ala Phe Asp Lys Leu Tyr Lys Lys Glu Asp Ile Glu
195 200 205
Lys Leu Phe Phe Leu Ile Glu Asn Ser Lys Lys His Glu Lys Tyr Lys

210 215 220
Ile Arg Glu Tyr Tyr His Lys Ile Ile Gly Arg Lys Asn Asp Lys Glu
225 230 235 240
Asn Phe Ala Lys Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Ile
245 250 255
Lys Glu Leu Ile Glu Lys Ile Pro Asp Met Ser Glu Leu Lys Lys Ser
260 265 270
Gln Val Phe Tyr Lys Tyr Tyr Leu Asp Lys Glu Glu Leu Asn Asp Lys
275 280 285

Asn Ile Lys Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln
290 295 300
Leu Leu Lys Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp
305 310 315 320
Lys Ile Lys Arg Ile Phe Glu Tyr Gln Asn Leu Lys Lys Leu Ile Glu
325 330 335

Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys
340 345 350
Tyr Asn Tyr Tyr Leu Gln Val Gly Glu Ile Ala Thr Ser Asp Phe Ile

355 360 365
Ala Arg Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val
370 375 380
Ser Ser Val Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn
385 390 395 400
Glu Asn Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn
405 410 415
Lys Gly Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn
420 425 430

Glu Asn Lys Gln Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser
435 440 445
Tyr Asp Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala
450 455 460
Asn Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe
465 470 475 480
Asn Leu Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala
485 490 495
Pro Ser Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys

500 505 510
Lys Leu Lys Leu Lys Ile Phe Lys Gln Leu Asn Ser Ala Asn Val Phe
515 520 525
Asn Tyr Tyr Glu Lys Asp Val Ile Ile Lys Tyr Leu Lys Asn Thr Lys
530 535 540
Phe Asn Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys
545 550 555 560
Leu Tyr Asn Lys Ile Glu Asp Leu Arg Asn Thr Leu Lys Phe Phe Trp
565 570 575

Ser Val Pro Lys Asp Lys Glu Glu Lys Asp Ala Gln Ile Tyr Leu Leu

580 585 590

Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Lys Phe Val Lys Asn Ser

595 600 605

Lys Val Phe Phe Lys Ile Thr Asn Glu Val Ile Lys Ile Asn Lys Gln

610 615 620

Arg Asn Gln Lys Thr Gly His Tyr Lys Tyr Gln Lys Phe Glu Asn Ile

625 630 635 640

Glu Lys Thr Val Pro Val Glu Tyr Leu Ala Ile Ile Gln Ser Arg Glu

645 650 655

Met Ile Asn Asn Gln Asp Lys Glu Glu Lys Asn Thr Tyr Ile Asp Phe

660 665 670

Ile Gln Gln Ile Phe Leu Lys Gly Phe Ile Asp Tyr Leu Asn Lys Asn

675 680 685

Asn Leu Lys Tyr Ile Glu Ser Asn Asn Asn Asn Asp Asn Asn Asp Ile

690 695 700

Phe Ser Lys Ile Lys Ile Lys Lys Asp Asn Lys Glu Lys Tyr Asp Lys

705 710 715 720

Ile Leu Lys Asn Tyr Glu Lys His Asn Arg Asn Lys Glu Ile Pro His

725 730 735

Glu Ile Asn Glu Phe Val Arg Glu Ile Lys Leu Gly Lys Ile Leu Lys

740 745 750

Tyr Thr Glu Asn Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu Leu Asn

755 760 765

His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr Gln Ser

770 775 780

Ala Asn Lys Glu Glu Thr Phe Ser Asp Glu Leu Glu Leu Ile Asn Leu

785 790 795 800

Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu Glu Ala

805 810 815

Asn Glu Ile Gly Lys Phe Leu Asp Phe Asn Glu Asn Lys Ile Lys Asp

820 825 830

Arg Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe Asp Gly

835 840 845

Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys Tyr Gly

850 855 860

Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Lys Tyr Lys Ile

865 870 875 880

Ser Leu Lys Glu Leu Lys Glu Tyr Ser Asn Lys Lys Asn Glu Ile Glu

885 890 895

Lys Asn Tyr Thr Met Gln Gln Asn Leu His Arg Lys Tyr Ala Arg Pro

900 905 910

Lys Lys Asp Glu Lys Phe Asn Asp Glu Asp Tyr Lys Glu Tyr Glu Lys

915 920 925

Ala Ile Gly Asn Ile Gln Lys Tyr Thr His Leu Lys Asn Lys Val Glu

930 935 940

Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Lys Ile Leu His

945 950 955 960

Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg Phe Arg

965 970 975

Leu Lys Gly Glu Phe Pro Glu Asn His Tyr Ile Glu Glu Ile Phe Asn

980 985 990

Phe Asp Asn Ser Lys Asn Val Lys Tyr Lys Ser Gly Gln Ile Val Glu

995 1000 1005

Lys Tyr Ile Asn Phe Tyr Lys Glu Leu Tyr Lys Asp Asn Val Glu Lys

1010 1015 1020

Arg Ser Ile Tyr Ser Asp Lys Lys Val Lys Lys Leu Lys Gln Glu Lys

1025 1030 1035 1040

Lys Asp Leu Tyr Ile Arg Asn Tyr Ile Ala His Phe Asn Tyr Ile Pro

1045 1050 1055

His Ala Glu Ile Ser Leu Leu Glu Val Leu Glu Asn Leu Arg Lys Leu

1060 1065 1070

Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala Ile Met Lys Ser Ile Val

1075 1080 1085

Asp Ile Leu Lys Glu Tyr Gly Phe Val Ala Thr Phe Lys Ile Gly Ala
1090 1095 1100
Asp Lys Lys Ile Glu Ile Gln Thr Leu Glu Ser Glu Lys Ile Val His
1105 1110 1115 1120
Leu Lys Asn Leu Lys Lys Lys Lys Leu Met Thr Asp Arg Asn Ser Glu
1125 1130 1135
Glu Leu Cys Glu Leu Val Lys Val Met Phe Glu Tyr Lys Ala Leu Glu
1140 1145 1150

<210> 128
<211> 1285
<212> PRT
<213> Rhodobacter capsulatus
<400> 128
Met Gln Ile Gly Lys Val Gln Gly Arg Thr Ile Ser Glu Phe Gly Asp
1 5 10 15
Pro Ala Gly Gly Leu Lys Arg Lys Ile Ser Thr Asp Gly Lys Asn Arg
20 25 30
Lys Glu Leu Pro Ala His Leu Ser Ser Asp Pro Lys Ala Leu Ile Gly
35 40 45
Gln Trp Ile Ser Gly Ile Asp Lys Ile Tyr Arg Lys Pro Asp Ser Arg

50 55 60
Lys Ser Asp Gly Lys Ala Ile His Ser Pro Thr Pro Ser Lys Met Gln
65 70 75 80
Phe Asp Ala Arg Asp Asp Leu Gly Glu Ala Phe Trp Lys Leu Val Ser
85 90 95
Glu Ala Gly Leu Ala Gln Asp Ser Asp Tyr Asp Gln Phe Lys Arg Arg
100 105 110
Leu His Pro Tyr Gly Asp Lys Phe Gln Pro Ala Asp Ser Gly Ala Lys
115 120 125

Leu Lys Phe Glu Ala Asp Pro Pro Glu Pro Gln Ala Phe His Gly Arg
130 135 140
Trp Tyr Gly Ala Met Ser Lys Arg Gly Asn Asp Ala Lys Glu Leu Ala

145 150 155 160

Ala Ala Leu Tyr Glu His Leu His Val Asp Glu Lys Arg Ile Asp Gly

165 170 175

Gln Pro Lys Arg Asn Pro Lys Thr Asp Lys Phe Ala Pro Gly Leu Val

180 185 190

Val Ala Arg Ala Leu Gly Ile Glu Ser Ser Val Leu Pro Arg Gly Met

195 200 205

Ala Arg Leu Ala Arg Asn Trp Gly Glu Glu Glu Ile Gln Thr Tyr Phe

210 215 220

Val Val Asp Val Ala Ala Ser Val Lys Glu Val Ala Lys Ala Ala Val

225 230 235 240

Ser Ala Ala Gln Ala Phe Asp Pro Pro Arg Gln Val Ser Gly Arg Ser

245 250 255

Leu Ser Pro Lys Val Gly Phe Ala Leu Ala Glu His Leu Glu Arg Val

260 265 270

Thr Gly Ser Lys Arg Cys Ser Phe Asp Pro Ala Ala Gly Pro Ser Val

275 280 285

Leu Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala

290 295 300

Arg Gly Lys Asn Ala Ala Arg Ala Phe Pro Ala Asp Lys Thr Glu Leu

305 310 315 320

Leu Ala Leu Met Arg His Thr His Glu Asn Arg Val Arg Asn Gln Met

325 330 335

Val Arg Met Gly Arg Val Ser Glu Tyr Arg Gly Gln Gln Ala Gly Asp

340 345 350

Leu Ala Gln Ser His Tyr Trp Thr Ser Ala Gly Gln Thr Glu Ile Lys

355 360 365

Glu Ser Glu Ile Phe Val Arg Leu Trp Val Gly Ala Phe Ala Leu Ala

370 375 380

Gly Arg Ser Met Lys Ala Trp Ile Asp Pro Met Gly Lys Ile Val Asn

385 390 395 400

Thr Glu Lys Asn Asp Arg Asp Leu Thr Ala Ala Val Asn Ile Arg Gln

405 410 415

Val Ile Ser Asn Lys Glu Met Val Ala Glu Ala Met Ala Arg Arg Gly
420 425 430
Ile Tyr Phe Gly Glu Thr Pro Glu Leu Asp Arg Leu Gly Ala Glu Gly
435 440 445
Asn Glu Gly Phe Val Phe Ala Leu Leu Arg Tyr Leu Arg Gly Cys Arg
450 455 460
Asn Gln Thr Phe His Leu Gly Ala Arg Ala Gly Phe Leu Lys Glu Ile
465 470 475 480
Arg Lys Glu Leu Glu Lys Thr Arg Trp Gly Lys Ala Lys Glu Ala Glu

485 490 495
His Val Val Leu Thr Asp Lys Thr Val Ala Ala Ile Arg Ala Ile Ile
500 505 510
Asp Asn Asp Ala Lys Ala Leu Gly Ala Arg Leu Leu Ala Asp Leu Ser
515 520 525
Gly Ala Phe Val Ala His Tyr Ala Ser Lys Glu His Phe Ser Thr Leu
530 535 540
Tyr Ser Glu Ile Val Lys Ala Val Lys Asp Ala Pro Glu Val Ser Ser
545 550 555 560

Gly Leu Pro Arg Leu Lys Leu Leu Leu Lys Arg Ala Asp Gly Val Arg
565 570 575
Gly Tyr Val His Gly Leu Arg Asp Thr Arg Lys His Ala Phe Ala Thr
580 585 590
Lys Leu Pro Pro Pro Pro Ala Pro Arg Glu Leu Asp Asp Pro Ala Thr
595 600 605
Lys Ala Arg Tyr Ile Ala Leu Leu Arg Leu Tyr Asp Gly Pro Phe Arg
610 615 620
Ala Tyr Ala Ser Gly Ile Thr Gly Thr Ala Leu Ala Gly Pro Ala Ala

625 630 635 640
Arg Ala Lys Glu Ala Ala Thr Ala Leu Ala Gln Ser Val Asn Val Thr
645 650 655

Lys Ala Tyr Ser Asp Val Met Glu Gly Arg Thr Ser Arg Leu Arg Pro
660 665 670

Pro Asn Asp Gly Glu Thr Leu Arg Glu Tyr Leu Ser Ala Leu Thr Gly
675 680 685

Glu Thr Ala Thr Glu Phe Arg Val Gln Ile Gly Tyr Glu Ser Asp Ser
690 695 700

Glu Asn Ala Arg Lys Gln Ala Glu Phe Ile Glu Asn Tyr Arg Arg Asp
705 710 715 720

Met Leu Ala Phe Met Phe Glu Asp Tyr Ile Arg Ala Lys Gly Phe Asp
725 730 735

Trp Ile Leu Lys Ile Glu Pro Gly Ala Thr Ala Met Thr Arg Ala Pro
740 745 750

Val Leu Pro Glu Pro Ile Asp Thr Arg Gly Gln Tyr Glu His Trp Gln
755 760 765

Ala Ala Leu Tyr Leu Val Met His Phe Val Pro Ala Ser Asp Val Ser
770 775 780

Asn Leu Leu His Gln Leu Arg Lys Trp Glu Ala Leu Gln Gly Lys Tyr
785 790 795 800

Glu Leu Val Gln Asp Gly Asp Ala Thr Asp Gln Ala Asp Ala Arg Arg
805 810 815

Glu Ala Leu Asp Leu Val Lys Arg Phe Arg Asp Val Leu Val Leu Phe
820 825 830

Leu Lys Thr Gly Glu Ala Arg Phe Glu Gly Arg Ala Ala Pro Phe Asp
835 840 845

Leu Lys Pro Phe Arg Ala Leu Phe Ala Asn Pro Ala Thr Phe Asp Arg
850 855 860

Leu Phe Met Ala Thr Pro Thr Thr Ala Arg Pro Ala Glu Asp Asp Pro
865 870 875 880

Glu Gly Asp Gly Ala Ser Glu Pro Glu Leu Arg Val Ala Arg Thr Leu
885 890 895

Arg Gly Leu Arg Gln Ile Ala Arg Tyr Asn His Met Ala Val Leu Ser
900 905 910

Asp Leu Phe Ala Lys His Lys Val Arg Asp Glu Glu Val Ala Arg Leu

915 920 925

Ala Glu Ile Glu Asp Glu Thr Gln Glu Lys Ser Gln Ile Val Ala Ala

930 935 940

Gln Glu Leu Arg Thr Asp Leu His Asp Lys Val Met Lys Cys His Pro

945 950 955 960

Lys Thr Ile Ser Pro Glu Glu Arg Gln Ser Tyr Ala Ala Ala Ile Lys

965 970 975

Thr Ile Glu Glu His Arg Phe Leu Val Gly Arg Val Tyr Leu Gly Asp

980 985 990

His Leu Arg Leu His Arg Leu Met Met Asp Val Ile Gly Arg Leu Ile

995 1000 1005

Asp Tyr Ala Gly Ala Tyr Glu Arg Asp Thr Gly Thr Phe Leu Ile Asn

1010 1015 1020

Ala Ser Lys Gln Leu Gly Ala Gly Ala Asp Trp Ala Val Thr Ile Ala

1025 1030 1035 1040

Gly Ala Ala Asn Thr Asp Ala Arg Thr Gln Thr Arg Lys Asp Leu Ala

1045 1050 1055

His Phe Asn Val Leu Asp Arg Ala Asp Gly Thr Pro Asp Leu Thr Ala

1060 1065 1070

Leu Val Asn Arg Ala Arg Glu Met Met Ala Tyr Asp Arg Lys Arg Lys

1075 1080 1085

Asn Ala Val Pro Arg Ser Ile Leu Asp Met Leu Ala Arg Leu Gly Leu

1090 1095 1100

Thr Leu Lys Trp Gln Met Lys Asp His Leu Leu Gln Asp Ala Thr Ile

1105 1110 1115 1120

Thr Gln Ala Ala Ile Lys His Leu Asp Lys Val Arg Leu Thr Val Gly

1125 1130 1135

Gly Pro Ala Ala Val Thr Glu Ala Arg Phe Ser Gln Asp Tyr Leu Gln

1140 1145 1150

Met Val Ala Ala Val Phe Asn Gly Ser Val Gln Asn Pro Lys Pro Arg

1155 1160 1165

Arg Arg Asp Asp Gly Asp Ala Trp His Lys Pro Pro Lys Pro Ala Thr

1170 1175 1180

Ala Gln Ser Gln Pro Asp Gln Lys Pro Pro Asn Lys Ala Pro Ser Ala

1185 1190 1195 1200

Gly Ser Arg Leu Pro Pro Pro Gln Val Gly Glu Val Tyr Glu Gly Val

1205 1210 1215

Val Val Lys Val Ile Asp Thr Gly Ser Leu Gly Phe Leu Ala Val Glu

1220 1225 1230

Gly Val Ala Gly Asn Ile Gly Leu His Ile Ser Arg Leu Arg Arg Ile

1235 1240 1245

Arg Glu Asp Ala Ile Ile Val Gly Arg Arg Tyr Arg Phe Arg Val Glu

1250 1255 1260

Ile Tyr Val Pro Pro Lys Ser Asn Thr Ser Lys Leu Asn Ala Ala Asp

1265 1270 1275 1280

Leu Val Arg Ile Asp

1285

<210> 129

<211> 1285

<212> PRT

<213> Rhodobacter capsulatus

<400> 129

Met Gln Ile Gly Lys Val Gln Gly Arg Thr Ile Ser Glu Phe Gly Asp

1 5 10 15

Pro Ala Gly Gly Leu Lys Arg Lys Ile Ser Thr Asp Gly Lys Asn Arg

20 25 30

Lys Glu Leu Pro Ala His Leu Ser Ser Asp Pro Lys Ala Leu Ile Gly

35 40 45

Gln Trp Ile Ser Gly Ile Asp Lys Ile Tyr Arg Lys Pro Asp Ser Arg

50 55 60

Lys Ser Asp Gly Lys Ala Ile His Ser Pro Thr Pro Ser Lys Met Gln

65 70 75 80

Phe Asp Ala Arg Asp Asp Leu Gly Glu Ala Phe Trp Lys Leu Val Ser
85 90 95
Glu Ala Gly Leu Ala Gln Asp Ser Asp Tyr Asp Gln Phe Lys Arg Arg
100 105 110
Leu His Pro Tyr Gly Asp Lys Phe Gln Pro Ala Asp Ser Gly Ala Lys
115 120 125

Leu Lys Phe Glu Ala Asp Pro Pro Glu Pro Gln Ala Phe His Gly Arg
130 135 140
Trp Tyr Gly Ala Met Ser Lys Arg Gly Asn Asp Ala Lys Glu Leu Ala
145 150 155 160
Ala Ala Leu Tyr Glu His Leu His Val Asp Glu Lys Arg Ile Asp Gly
165 170 175
Gln Pro Lys Arg Asn Pro Lys Thr Asp Lys Phe Ala Pro Gly Leu Val
180 185 190
Val Ala Arg Ala Leu Gly Ile Glu Ser Ser Val Leu Pro Arg Gly Met

195 200 205
Ala Arg Leu Ala Arg Asn Trp Gly Glu Glu Glu Ile Gln Thr Tyr Phe
210 215 220
Val Val Asp Val Ala Ala Ser Val Lys Glu Val Ala Lys Ala Ala Val
225 230 235 240
Ser Ala Ala Gln Ala Phe Asp Pro Pro Arg Gln Val Ser Gly Arg Ser
245 250 255
Leu Ser Pro Lys Val Gly Phe Ala Leu Ala Glu His Leu Glu Arg Val
260 265 270

Thr Gly Ser Lys Arg Cys Ser Phe Asp Pro Ala Ala Gly Pro Ser Val
275 280 285
Leu Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala
290 295 300
Arg Gly Lys Asn Ala Ala Arg Ala Phe Pro Ala Asp Lys Thr Glu Leu
305 310 315 320
Leu Ala Leu Met Arg His Thr His Glu Asn Arg Val Arg Asn Gln Met
325 330 335

Val Arg Met Gly Arg Val Ser Glu Tyr Arg Gly Gln Gln Ala Gly Asp

340 345 350

Leu Ala Gln Ser His Tyr Trp Thr Ser Ala Gly Gln Thr Glu Ile Lys

355 360 365

Glu Ser Glu Ile Phe Val Arg Leu Trp Val Gly Ala Phe Ala Leu Ala

370 375 380

Gly Arg Ser Met Lys Ala Trp Ile Asp Pro Met Gly Lys Ile Val Asn

385 390 395 400

Thr Glu Lys Asn Asp Arg Asp Leu Thr Ala Ala Val Asn Ile Arg Gln

405 410 415

Val Ile Ser Asn Lys Glu Met Val Ala Glu Ala Met Ala Arg Arg Gly

420 425 430

Ile Tyr Phe Gly Glu Thr Pro Glu Leu Asp Arg Leu Gly Ala Glu Gly

435 440 445

Asn Glu Gly Phe Val Phe Ala Leu Leu Arg Tyr Leu Arg Gly Cys Arg

450 455 460

Asn Gln Thr Phe His Leu Gly Ala Arg Ala Gly Phe Leu Lys Glu Ile

465 470 475 480

Arg Lys Glu Leu Glu Lys Thr Arg Trp Gly Lys Ala Lys Glu Ala Glu

485 490 495

His Val Val Leu Thr Asp Lys Thr Val Ala Ala Ile Arg Ala Ile Ile

500 505 510

Asp Asn Asp Ala Lys Ala Leu Gly Ala Arg Leu Leu Ala Asp Leu Ser

515 520 525

Gly Ala Phe Val Ala His Tyr Ala Ser Lys Glu His Phe Ser Thr Leu

530 535 540

Tyr Ser Glu Ile Val Lys Ala Val Lys Asp Ala Pro Glu Val Ser Ser

545 550 555 560

Gly Leu Pro Arg Leu Lys Leu Leu Leu Lys Arg Ala Asp Gly Val Arg

565 570 575

Gly Tyr Val His Gly Leu Arg Asp Thr Arg Lys His Ala Phe Ala Thr

580 585 590

Lys Leu Pro Pro Pro Pro Ala Pro Arg Glu Leu Asp Asp Pro Ala Thr

595 600 605

Lys Ala Arg Tyr Ile Ala Leu Leu Arg Leu Tyr Asp Gly Pro Phe Arg

610 615 620

Ala Tyr Ala Ser Gly Ile Thr Gly Thr Ala Leu Ala Gly Pro Ala Ala

625 630 635 640

Arg Ala Lys Glu Ala Ala Thr Ala Leu Ala Gln Ser Val Asn Val Thr

645 650 655

Lys Ala Tyr Ser Asp Val Met Glu Gly Arg Ser Ser Arg Leu Arg Pro

660 665 670

Pro Asn Asp Gly Glu Thr Leu Arg Glu Tyr Leu Ser Ala Leu Thr Gly

675 680 685

Glu Thr Ala Thr Glu Phe Arg Val Gln Ile Gly Tyr Glu Ser Asp Ser

690 695 700

Glu Asn Ala Arg Lys Gln Ala Glu Phe Ile Glu Asn Tyr Arg Arg Asp

705 710 715 720

Met Leu Ala Phe Met Phe Glu Asp Tyr Ile Arg Ala Lys Gly Phe Asp

725 730 735

Trp Ile Leu Lys Ile Glu Pro Gly Ala Thr Ala Met Thr Arg Ala Pro

740 745 750

Val Leu Pro Glu Pro Ile Asp Thr Arg Gly Gln Tyr Glu His Trp Gln

755 760 765

Ala Ala Leu Tyr Leu Val Met His Phe Val Pro Ala Ser Asp Val Ser

770 775 780

Asn Leu Leu His Gln Leu Arg Lys Trp Glu Ala Leu Gln Gly Lys Tyr

785 790 795 800

Glu Leu Val Gln Asp Gly Asp Ala Thr Asp Gln Ala Asp Ala Arg Arg

805 810 815

Glu Ala Leu Asp Leu Val Lys Arg Phe Arg Asp Val Leu Val Leu Phe

820 825 830

Leu Lys Thr Gly Glu Ala Arg Phe Glu Gly Arg Ala Ala Pro Phe Asp

835 840 845

Leu Lys Pro Phe Arg Ala Leu Phe Ala Asn Pro Ala Thr Phe Asp Arg
850 855 860
Leu Phe Met Ala Thr Pro Thr Thr Ala Arg Pro Ala Glu Asp Asp Pro
865 870 875 880
Glu Gly Asp Gly Ala Ser Glu Pro Glu Leu Arg Val Ala Arg Thr Leu
885 890 895
Arg Gly Leu Arg Gln Ile Ala Arg Tyr Asn His Met Ala Val Leu Ser
900 905 910
Asp Leu Phe Ala Lys His Lys Val Arg Asp Glu Glu Val Ala Arg Leu

915 920 925
Ala Glu Ile Glu Asp Glu Thr Gln Glu Lys Ser Gln Ile Val Ala Ala
930 935 940
Gln Glu Leu Arg Thr Asp Leu His Asp Lys Val Met Lys Cys His Pro
945 950 955 960
Lys Thr Ile Ser Pro Glu Glu Arg Gln Ser Tyr Ala Ala Ala Ile Lys
965 970 975
Thr Ile Glu Glu His Arg Phe Leu Val Gly Arg Val Tyr Leu Gly Asp
980 985 990

His Leu Arg Leu His Arg Leu Met Met Asp Val Ile Gly Arg Leu Ile
995 1000 1005
Asp Tyr Ala Gly Ala Tyr Glu Arg Asp Thr Gly Thr Phe Leu Ile Asn
1010 1015 1020
Ala Ser Lys Gln Leu Gly Ala Gly Ala Asp Trp Ala Val Thr Ile Ala
1025 1030 1035 1040
Gly Ala Ala Asn Thr Asp Ala Arg Thr Gln Thr Arg Lys Asp Leu Ala
1045 1050 1055
His Phe Asn Val Leu Asp Arg Ala Asp Gly Thr Pro Asp Leu Thr Ala

1060 1065 1070
Leu Val Asn Arg Ala Arg Glu Met Met Ala Tyr Asp Arg Lys Arg Lys
1075 1080 1085

Asn Ala Val Pro Arg Ser Ile Leu Asp Met Leu Ala Arg Leu Gly Leu
1090 1095 1100

Thr Leu Lys Trp Gln Met Lys Asp His Leu Leu Gln Asp Ala Thr Ile
1105 1110 1115 1120

Thr Gln Ala Ala Ile Lys His Leu Asp Lys Val Arg Leu Thr Val Gly
1125 1130 1135

Gly Pro Ala Ala Val Thr Glu Ala Arg Phe Ser Gln Asp Tyr Leu Gln
1140 1145 1150

Met Val Ala Ala Val Phe Asn Gly Ser Val Gln Asn Pro Lys Pro Arg
1155 1160 1165

Arg Arg Asp Asp Gly Asp Ala Trp His Lys Pro Pro Lys Pro Ala Thr
1170 1175 1180

Ala Gln Ser Gln Pro Asp Gln Lys Pro Pro Asn Lys Ala Pro Ser Ala
1185 1190 1195 1200

Gly Ser Arg Leu Pro Pro Pro Gln Val Gly Glu Val Tyr Glu Gly Val

1205 1210 1215

Val Val Lys Val Ile Asp Thr Gly Ser Leu Gly Phe Leu Ala Val Glu
1220 1225 1230

Gly Val Ala Gly Asn Ile Gly Leu His Ile Ser Arg Leu Arg Arg Ile
1235 1240 1245

Arg Glu Asp Ala Ile Ile Val Gly Arg Arg Tyr Arg Phe Arg Val Glu
1250 1255 1260

Ile Tyr Val Pro Pro Lys Ser Asn Thr Ser Lys Leu Asn Ala Ala Asp
1265 1270 1275 1280

Leu Val Arg Ile Asp
1285

<210> 130
<211> 1285
<212> PRT
<213> Rhodobacter capsulatus
<400> 130
Met Gln Ile Gly Lys Val Gln Gly Arg Thr Ile Ser Glu Phe Gly Asp

1 5 10 15

Pro Ala Gly Gly Leu Lys Arg Lys Ile Ser Thr Asp Gly Lys Asn Arg

20 25 30

Lys Glu Leu Pro Ala His Leu Ser Ser Asp Pro Lys Ala Leu Ile Gly

35 40 45

Gln Trp Ile Ser Gly Ile Asp Lys Ile Tyr Arg Lys Pro Asp Ser Arg

50 55 60

Lys Ser Asp Gly Lys Ala Ile His Ser Pro Thr Pro Ser Lys Met Gln

65 70 75 80

Phe Asp Ala Arg Asp Asp Leu Gly Glu Ala Phe Trp Lys Leu Val Ser

85 90 95

Glu Ala Gly Leu Ala Gln Asp Ser Asp Tyr Asp Gln Phe Lys Arg Arg

100 105 110

Leu His Pro Tyr Gly Asp Lys Phe Gln Pro Ala Asp Ser Gly Ala Lys

115 120 125

Leu Lys Phe Glu Ala Asp Pro Pro Glu Pro Gln Ala Phe His Gly Arg

130 135 140

Trp Tyr Gly Ala Met Ser Lys Arg Gly Asn Asp Ala Lys Glu Leu Ala

145 150 155 160

Ala Ala Leu Tyr Glu His Leu His Val Asp Glu Lys Arg Ile Asp Gly

165 170 175

Gln Pro Lys Arg Asn Pro Lys Thr Asp Lys Phe Ala Pro Gly Leu Val

180 185 190

Val Ala Arg Ala Leu Gly Ile Glu Ser Ser Val Leu Pro Arg Gly Met

195 200 205

Ala Arg Leu Ala Arg Asn Trp Gly Glu Glu Glu Ile Gln Thr Tyr Phe

210 215 220

Val Val Asp Val Ala Ala Ser Val Lys Glu Val Ala Lys Ala Ala Val

225 230 235 240

Ser Ala Ala Gln Ala Phe Asp Pro Pro Arg Gln Val Ser Gly Arg Ser

245 250 255

Leu Ser Pro Lys Val Gly Phe Ala Leu Ala Glu His Leu Glu Arg Val

260 265 270

Thr Gly Ser Lys Arg Cys Ser Phe Asp Pro Ala Ala Gly Pro Ser Val
275 280 285
Leu Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala
290 295 300
Arg Gly Lys Asn Ala Ala Arg Ala Phe Pro Ala Asp Lys Thr Glu Leu
305 310 315 320
Leu Ala Leu Met Arg His Thr His Glu Asn Arg Val Arg Asn Gln Met
325 330 335
Val Arg Met Gly Arg Val Ser Glu Tyr Arg Gly Gln Gln Ala Gly Asp

340 345 350
Leu Ala Gln Ser His Tyr Trp Thr Ser Ala Gly Gln Thr Glu Ile Lys
355 360 365
Glu Ser Glu Ile Phe Val Arg Leu Trp Val Gly Ala Phe Ala Leu Ala
370 375 380
Gly Arg Ser Met Lys Ala Trp Ile Asp Pro Met Gly Lys Ile Val Asn
385 390 395 400
Thr Glu Lys Asn Asp Arg Asp Leu Thr Ala Ala Val Asn Ile Arg Gln
405 410 415

Val Ile Ser Asn Lys Glu Met Val Ala Glu Ala Met Ala Arg Arg Gly
420 425 430
Ile Tyr Phe Gly Glu Thr Pro Glu Leu Asp Arg Leu Gly Ala Glu Gly
435 440 445
Asn Glu Gly Phe Val Phe Ala Leu Leu Arg Tyr Leu Arg Gly Cys Arg
450 455 460
Asn Gln Thr Phe His Leu Gly Ala Arg Ala Gly Phe Leu Lys Glu Ile
465 470 475 480
Arg Lys Glu Leu Glu Lys Thr Arg Trp Gly Lys Ala Lys Glu Ala Glu

485 490 495
His Val Val Leu Thr Asp Lys Thr Val Ala Ala Ile Arg Ala Ile Ile
500 505 510

Asp Asn Asp Ala Lys Ala Leu Gly Ala Arg Leu Leu Ala Asp Leu Ser
515 520 525
Gly Ala Phe Val Ala His Tyr Ala Ser Lys Glu His Phe Ser Thr Leu
530 535 540
Tyr Ser Glu Ile Val Lys Ala Val Lys Asp Ala Pro Glu Val Ser Ser
545 550 555 560

Gly Leu Pro Arg Leu Lys Leu Leu Leu Lys Arg Ala Asp Gly Val Arg
565 570 575
Gly Tyr Val His Gly Leu Arg Asp Thr Arg Lys His Ala Phe Ala Thr
580 585 590
Lys Leu Pro Pro Pro Pro Ala Pro Arg Glu Leu Asp Asp Pro Ala Thr
595 600 605
Lys Ala Arg Tyr Ile Ala Leu Leu Arg Leu Tyr Asp Gly Pro Phe Arg
610 615 620
Ala Tyr Ala Ser Gly Ile Thr Gly Thr Ala Leu Ala Gly Pro Ala Ala

625 630 635 640
Arg Ala Lys Glu Ala Ala Thr Ala Leu Ala Gln Ser Val Asn Val Thr
645 650 655
Lys Ala Tyr Ser Asp Val Met Glu Gly Arg Ser Ser Arg Leu Arg Pro
660 665 670
Pro Asn Asp Gly Glu Thr Leu Arg Glu Tyr Leu Ser Ala Leu Thr Gly
675 680 685
Glu Thr Ala Thr Glu Phe Arg Val Gln Ile Gly Tyr Glu Ser Asp Ser
690 695 700

Glu Asn Ala Arg Lys Gln Ala Glu Phe Ile Glu Asn Tyr Arg Arg Asp
705 710 715 720
Met Leu Ala Phe Met Phe Glu Asp Tyr Ile Arg Ala Lys Gly Phe Asp
725 730 735
Trp Ile Leu Lys Ile Glu Pro Gly Ala Thr Ala Met Thr Arg Ala Pro
740 745 750
Val Leu Pro Glu Pro Ile Asp Thr Arg Gly Gln Tyr Glu His Trp Gln
755 760 765

Ala Ala Leu Tyr Leu Val Met His Phe Val Pro Ala Ser Asp Val Ser

770 775 780

Asn Leu Leu His Gln Leu Arg Lys Trp Glu Ala Leu Gln Gly Lys Tyr
785 790 795 800
Glu Leu Val Gln Asp Gly Asp Ala Thr Asp Gln Ala Asp Ala Arg Arg
805 810 815
Glu Ala Leu Asp Leu Val Lys Arg Phe Arg Asp Val Leu Val Leu Phe
820 825 830
Leu Lys Thr Gly Glu Ala Arg Phe Glu Gly Arg Ala Ala Pro Phe Asp
835 840 845

Leu Lys Pro Phe Arg Ala Leu Phe Ala Asn Pro Ala Thr Phe Asp Arg
850 855 860
Leu Phe Met Ala Thr Pro Thr Thr Ala Arg Pro Ala Glu Asp Asp Pro
865 870 875 880
Glu Gly Asp Gly Ala Ser Glu Pro Glu Leu Arg Val Ala Arg Thr Leu
885 890 895
Arg Gly Leu Arg Gln Ile Ala Arg Tyr Asn His Met Ala Val Leu Ser
900 905 910
Asp Leu Phe Ala Lys His Lys Val Arg Asp Glu Glu Val Ala Arg Leu

915 920 925

Ala Glu Ile Glu Asp Glu Thr Gln Glu Lys Ser Gln Ile Val Ala Ala
930 935 940
Gln Glu Leu Arg Thr Asp Leu His Asp Lys Val Met Lys Cys His Pro
945 950 955 960
Lys Thr Ile Ser Pro Glu Glu Arg Gln Ser Tyr Ala Ala Ala Ile Lys
965 970 975
Thr Ile Glu Glu His Arg Phe Leu Val Gly Arg Val Tyr Leu Gly Asp
980 985 990

His Leu Arg Leu His Arg Leu Met Met Asp Val Ile Gly Arg Leu Ile

995 1000 1005

Asp Tyr Ala Gly Ala Tyr Glu Arg Asp Thr Gly Thr Phe Leu Ile Asn

1010 1015 1020

Ala Ser Lys Gln Leu Gly Ala Gly Ala Asp Trp Ala Val Thr Ile Ala

1025 1030 1035 1040

Gly Ala Ala Asn Thr Asp Ala Arg Thr Gln Thr Arg Lys Asp Leu Ala

1045 1050 1055

His Phe Asn Val Leu Asp Arg Ala Asp Gly Thr Pro Asp Leu Thr Ala

1060 1065 1070

Leu Val Asn Arg Ala Arg Glu Met Met Ala Tyr Asp Arg Lys Arg Lys

1075 1080 1085

Asn Ala Val Pro Arg Ser Ile Leu Asp Met Leu Ala Arg Leu Gly Leu

1090 1095 1100

Thr Leu Lys Trp Gln Met Lys Asp His Leu Leu Gln Asp Ala Thr Ile

1105 1110 1115 1120

Thr Gln Ala Ala Ile Lys His Leu Asp Lys Val Arg Leu Thr Val Gly

1125 1130 1135

Gly Pro Ala Ala Val Thr Glu Ala Arg Phe Ser Gln Asp Tyr Leu Gln

1140 1145 1150

Met Val Ala Ala Val Phe Asn Gly Ser Val Gln Asn Pro Lys Pro Arg

1155 1160 1165

Arg Arg Asp Asp Gly Asp Ala Trp His Lys Pro Pro Lys Pro Ala Thr

1170 1175 1180

Ala Gln Ser Gln Pro Asp Gln Lys Pro Pro Asn Lys Ala Pro Ser Ala

1185 1190 1195 1200

Gly Ser Arg Leu Pro Pro Pro Gln Val Gly Glu Val Tyr Glu Gly Val

1205 1210 1215

Val Val Lys Val Ile Asp Thr Gly Ser Leu Gly Phe Leu Ala Val Glu

1220 1225 1230

Gly Val Ala Gly Asn Ile Gly Leu His Ile Ser Arg Leu Arg Arg Ile

1235 1240 1245

Arg Glu Asp Ala Ile Ile Val Gly Arg Arg Tyr Arg Phe Arg Val Glu

1250 1255 1260

Ile Tyr Val Pro Pro Lys Ser Asn Thr Ser Lys Leu Asn Ala Ala Asp

1265 1270 1275 1280

Leu Val Arg Ile Asp
1285

<210> 131

<211> 1389

<212> PRT

<213> Unknown

<220><223> Synthetic

<400> 131

Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys
1 5 10 15

Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp
20 25 30

Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys
35 40 45

Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn
50 55 60

Asp Asn Ile Leu Lys Glu Phe Thr Arg Lys Phe His Ala Gly Asn Ile
65 70 75 80

Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn
85 90 95

Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr
100 105 110

Gly Lys Ser Glu Lys Leu Lys Ala Leu Gly Ile Thr Lys Lys Lys Ile

115 120 125

Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile
130 135 140

Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg
145 150 155 160

Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg
165 170 175

Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile

180 185 190

Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile
195 200 205
Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His
210 215 220
Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr
225 230 235 240
Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu
245 250 255
Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser

260 265 270
Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp
275 280 285
Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe
290 295 300
Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu
305 310 315 320
Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
325 330 335

Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
340 345 350
Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
355 360 365
Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile
370 375 380
Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
385 390 395 400
Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys

405 410 415
Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
420 425 430

Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys
435 440 445

Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
450 455 460

Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
465 470 475 480

Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp
485 490 495

Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
500 505 510

Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu
515 520 525

Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile
530 535 540

Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys

545 550 555 560

Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp
565 570 575

Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile
580 585 590

Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg
595 600 605

Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile
610 615 620

Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu
625 630 635 640

Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp
645 650 655

Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser
660 665 670

Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro
675 680 685

Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn

690 695 700

Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu
705 710 715 720
Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu
725 730 735
Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
740 745 750
Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys
755 760 765

Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
770 775 780
Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
785 790 795 800
Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
805 810 815
Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile
820 825 830
Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser

835 840 845

Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val
850 855 860
Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu
865 870 875 880
Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp
885 890 895
Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys
900 905 910

Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn
915 920 925
Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp

930 935 940

Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile

945 950 955 960

Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu

965 970 975

Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys

980 985 990

Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser

995 1000 1005

Lys Ile Leu Cys Arg Ile Ile Phe Asn Ser Asp Phe Leu Lys Lys Tyr

1010 1015 1020

Lys Lys Glu Ile Asp Asn Leu Ile Glu Asp Met Glu Ser Glu Asn Glu

1025 1030 1035 1040

Asn Lys Phe Gln Glu Ile Tyr Tyr Pro Lys Glu Arg Lys Asn Glu Leu

1045 1050 1055

Tyr Ile Tyr Lys Lys Asn Leu Phe Leu Asn Ile Gly Asn Pro Asn Phe

1060 1065 1070

Asp Lys Ile Tyr Gly Leu Ile Ser Asn Asp Ile Lys Met Ala Asp Ala

1075 1080 1085

Lys Phe Leu Phe Asn Ile Asp Gly Lys Asn Ile Arg Lys Asn Lys Ile

1090 1095 1100

Ser Glu Ile Asp Ala Ile Leu Lys Asn Leu Asn Asp Lys Leu Asn Gly

1105 1110 1115 1120

Tyr Ser Lys Glu Tyr Lys Glu Lys Tyr Ile Lys Lys Leu Lys Glu Asn

1125 1130 1135

Asp Asp Phe Phe Ala Lys Asn Ile Gln Asn Lys Asn Tyr Lys Ser Phe

1140 1145 1150

Glu Lys Asp Tyr Asn Arg Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu

1155 1160 1165

Val Glu Phe Asn Tyr Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile

1170 1175 1180

Asn Trp Lys Leu Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His

1185 1190 1195 1200

Tyr Ile Val Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly
1205 1210 1215
Tyr Asn Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp
1220 1225 1230
Gly Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser
1235 1240 1245
Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu Ser
1250 1255 1260
Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg Asn Tyr

1265 1270 1275 1280
Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp Tyr Ser Ile
1285 1290 1295
Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser Tyr Ser Thr Arg
1300 1305 1310
Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu Val Phe Lys Lys Asp
1315 1320 1325
Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys Lys Phe Lys Leu Ile Gly
1330 1335 1340

Asn Asn Asp Ile Leu Glu Arg Leu Met Lys Pro Lys Lys Val Ser Val
1345 1350 1355 1360
Leu Glu Leu Glu Ser Tyr Asn Ser Asp Tyr Ile Lys Asn Leu Ile Ile
1365 1370 1375
Glu Leu Leu Thr Lys Ile Glu Asn Thr Asn Asp Thr Leu
1380 1385

<210> 132
<211> 1197
<212> PRT
<213> L wadei
<400> 132
Met Lys Val Thr Lys Val Asp Gly Ile Ser His Lys Lys Tyr Ile Glu

1 5 10 15

Glu Gly Lys Leu Val Lys Ser Thr Ser Glu Glu Asn Arg Thr Ser Glu

20 25 30

Arg Leu Ser Glu Leu Leu Ser Ile Arg Leu Asp Ile Tyr Ile Lys Asn

35 40 45

Pro Asp Asn Ala Ser Glu Glu Glu Asn Arg Ile Arg Arg Glu Asn Leu

50 55 60

Lys Lys Phe Phe Ser Asn Lys Val Leu His Leu Lys Asp Ser Val Leu

65 70 75 80

Tyr Leu Lys Asn Arg Lys Glu Lys Asn Ala Val Gln Asp Lys Asn Tyr

85 90 95

Ser Glu Glu Asp Ile Ser Glu Tyr Asp Leu Lys Asn Lys Asn Ser Phe

100 105 110

Ser Val Leu Lys Lys Ile Leu Leu Asn Glu Asp Val Asn Ser Glu Glu

115 120 125

Leu Glu Ile Phe Arg Lys Asp Val Glu Ala Lys Leu Asn Lys Ile Asn

130 135 140

Ser Leu Lys Tyr Ser Phe Glu Glu Asn Lys Ala Asn Tyr Gln Lys Ile

145 150 155 160

Asn Glu Asn Asn Val Glu Lys Val Gly Gly Lys Ser Lys Arg Asn Ile

165 170 175

Ile Tyr Asp Tyr Tyr Arg Glu Ser Ala Lys Arg Asn Asp Tyr Ile Asn

180 185 190

Asn Val Gln Glu Ala Phe Asp Lys Leu Tyr Lys Lys Glu Asp Ile Glu

195 200 205

Lys Leu Phe Phe Leu Ile Glu Asn Ser Lys Lys His Glu Lys Tyr Lys

210 215 220

Ile Arg Glu Tyr Tyr His Lys Ile Ile Gly Arg Lys Asn Asp Lys Glu

225 230 235 240

Asn Phe Ala Lys Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Ile

245 250 255

Lys Glu Leu Ile Glu Lys Ile Pro Asp Met Ser Glu Leu Lys Lys Ser

260 265 270

Gln Val Phe Tyr Lys Tyr Tyr Leu Asp Lys Glu Glu Leu Asn Asp Lys

275 280 285

Asn Ile Lys Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln

290 295 300

Leu Leu Lys Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp

305 310 315 320

Lys Ile Lys Arg Ile Phe Glu Tyr Gln Asn Leu Lys Lys Leu Ile Glu

325 330 335

Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys

340 345 350

Tyr Asn Tyr Tyr Leu Gln Val Gly Glu Ile Ala Thr Ser Asp Phe Ile

355 360 365

Ala Arg Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val

370 375 380

Ser Ser Val Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn

385 390 395 400

Glu Asn Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn

405 410 415

Lys Gly Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn

420 425 430

Glu Asn Lys Gln Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser

435 440 445

Tyr Asp Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala

450 455 460

Asn Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe

465 470 475 480

Asn Leu Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala

485 490 495

Pro Ser Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys

500 505 510

Lys Leu Lys Leu Lys Ile Phe Lys Gln Leu Asn Ser Ala Asn Val Phe
515 520 525
Asn Tyr Tyr Glu Lys Asp Val Ile Ile Lys Tyr Leu Lys Asn Thr Lys
530 535 540
Phe Asn Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys
545 550 555 560
Leu Tyr Asn Lys Ile Glu Asp Leu Arg Asn Thr Leu Lys Phe Phe Trp
565 570 575
Ser Val Pro Lys Asp Lys Glu Glu Lys Asp Ala Gln Ile Tyr Leu Leu

580 585 590
Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Lys Phe Val Lys Asn Ser
595 600 605
Lys Val Phe Phe Lys Ile Thr Asn Glu Val Ile Lys Ile Asn Lys Gln
610 615 620
Arg Asn Gln Lys Thr Gly His Tyr Lys Tyr Gln Lys Phe Glu Asn Ile
625 630 635 640
Glu Lys Thr Val Pro Val Glu Tyr Leu Ala Ile Ile Gln Ser Arg Glu
645 650 655

Met Ile Asn Asn Gln Asp Lys Glu Glu Lys Asn Thr Tyr Ile Asp Phe
660 665 670
Ile Gln Gln Ile Phe Leu Lys Gly Phe Ile Asp Tyr Leu Asn Lys Asn
675 680 685
Asn Leu Lys Tyr Ile Glu Ser Asn Asn Asn Asn Asp Asn Asn Asp Ile
690 695 700
Phe Ser Lys Ile Lys Ile Lys Lys Asp Asn Lys Glu Lys Tyr Asp Lys
705 710 715 720
Ile Leu Lys Asn Tyr Glu Lys His Asn Arg Asn Lys Glu Ile Pro His

725 730 735
Glu Ile Asn Glu Phe Val Arg Glu Ile Lys Leu Gly Lys Ile Leu Lys
740 745 750
Tyr Thr Glu Asn Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu Leu Asn
755 760 765

His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr Gln Ser
770 775 780
Ala Asn Lys Glu Glu Thr Phe Ser Asp Glu Leu Glu Leu Ile Asn Leu
785 790 795 800

Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu Glu Ala
805 810 815
Asn Glu Ile Gly Lys Phe Leu Asp Phe Asn Glu Asn Lys Ile Lys Asp
820 825 830
Arg Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe Asp Gly
835 840 845
Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys Tyr Gly
850 855 860
Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Lys Tyr Lys Ile

865 870 875 880
Ser Leu Lys Glu Leu Lys Glu Tyr Ser Asn Lys Lys Asn Glu Ile Glu
885 890 895
Lys Asn Tyr Thr Met Gln Gln Asn Leu His Arg Lys Tyr Ala Arg Pro
900 905 910
Lys Lys Asp Glu Lys Phe Asn Asp Glu Asp Tyr Lys Glu Tyr Glu Lys
915 920 925
Ala Ile Gly Asn Ile Gln Lys Tyr Thr His Leu Lys Asn Lys Val Glu
930 935 940

Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Lys Ile Leu His
945 950 955 960
Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg Phe Arg
965 970 975
Leu Lys Gly Glu Phe Pro Glu Asn His Tyr Ile Glu Glu Ile Phe Asn
980 985 990
Phe Asp Asn Ser Lys Asn Val Lys Tyr Lys Ser Gly Gln Ile Val Glu
995 1000 1005
Lys Tyr Ile Asn Phe Tyr Lys Glu Leu Tyr Lys Asp Asn Val Glu Lys

1010 1015 1020

Arg Ser Ile Tyr Ser Asp Lys Lys Val Lys Lys Leu Lys Gln Glu Lys

1025 1030 1035 1040

Lys Asp Leu Tyr Ile Arg Asn Tyr Ile Ala His Phe Asn Tyr Ile Pro

1045 1050 1055

His Ala Glu Ile Ser Leu Leu Glu Val Leu Glu Asn Leu Arg Lys Leu

1060 1065 1070

Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala Ile Met Lys Ser Ile Val

1075 1080 1085

Asp Ile Leu Lys Glu Tyr Gly Phe Val Ala Thr Phe Lys Ile Gly Ala

1090 1095 1100

Asp Lys Lys Ile Glu Ile Gln Thr Leu Glu Ser Glu Lys Ile Val His

1105 1110 1115 1120

Leu Lys Asn Leu Lys Lys Lys Lys Leu Met Thr Asp Arg Asn Ser Glu

1125 1130 1135

Glu Leu Cys Glu Leu Val Lys Val Met Phe Glu Tyr Lys Ala Leu Glu

1140 1145 1150

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1155 1160 1165

Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val

1170 1175 1180

Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1185 1190 1195

<210> 133

<211> 1120

<212> PRT

<213> Listeria seeligeri

<400> 133

Met Trp Ile Ser Ile Lys Thr Leu Ile His His Leu Gly Val Leu Phe

1 5 10 15

Phe Cys Asp Tyr Met Tyr Asn Arg Arg Glu Lys Lys Ile Ile Glu Val

20 25 30

Lys Thr Met Arg Ile Thr Lys Val Glu Val Asp Arg Lys Lys Val Leu
35 40 45
Ile Ser Arg Asp Lys Asn Gly Gly Lys Leu Val Tyr Glu Asn Glu Met
50 55 60
Gln Asp Asn Thr Glu Gln Ile Met His His Lys Lys Ser Ser Phe Tyr
65 70 75 80
Lys Ser Val Val Asn Lys Thr Ile Cys Arg Pro Glu Gln Lys Gln Met
85 90 95

Lys Lys Leu Val His Gly Leu Leu Gln Glu Asn Ser Gln Glu Lys Ile
100 105 110
Lys Val Ser Asp Val Thr Lys Leu Asn Ile Ser Asn Phe Leu Asn His
115 120 125
Arg Phe Lys Lys Ser Leu Tyr Tyr Phe Pro Glu Asn Ser Pro Asp Lys
130 135 140
Ser Glu Glu Tyr Arg Ile Glu Ile Asn Leu Ser Gln Leu Leu Glu Asp
145 150 155 160
Ser Leu Lys Lys Gln Gln Gly Thr Phe Ile Cys Trp Glu Ser Phe Ser
165 170 175
Lys Asp Met Glu Leu Tyr Ile Asn Trp Ala Glu Asn Tyr Ile Ser Ser
180 185 190
Lys Thr Lys Leu Ile Lys Lys Ser Ile Arg Asn Asn Arg Ile Gln Ser
195 200 205
Thr Glu Ser Arg Ser Gly Gln Leu Met Asp Arg Tyr Met Lys Asp Ile
210 215 220
Leu Asn Lys Asn Lys Pro Phe Asp Ile Gln Ser Val Ser Glu Lys Tyr
225 230 235 240

Gln Leu Glu Lys Leu Thr Ser Ala Leu Lys Ala Thr Phe Lys Glu Ala
245 250 255
Lys Lys Asn Asp Lys Glu Ile Asn Tyr Lys Leu Lys Ser Thr Leu Gln
260 265 270
Asn His Glu Arg Gln Ile Ile Glu Glu Leu Lys Glu Asn Ser Glu Leu
275 280 285

Asn Gln Phe Asn Ile Glu Ile Arg Lys His Leu Glu Thr Tyr Phe Pro
290 295 300
Ile Lys Lys Thr Asn Arg Lys Val Gly Asp Ile Arg Asn Leu Glu Ile

305 310 315 320
Gly Glu Ile Gln Lys Ile Val Asn His Arg Leu Lys Asn Lys Ile Val
325 330 335
Gln Arg Ile Leu Gln Glu Gly Lys Leu Ala Ser Tyr Glu Ile Glu Ser
340 345 350
Thr Val Asn Ser Asn Ser Leu Gln Lys Ile Lys Ile Glu Glu Ala Phe
355 360 365
Ala Leu Lys Phe Ile Asn Ala Cys Leu Phe Ala Ser Asn Asn Leu Arg
370 375 380

Asn Met Val Tyr Pro Val Cys Lys Lys Asp Ile Leu Met Ile Gly Glu
385 390 395 400
Phe Lys Asn Ser Phe Lys Glu Ile Lys His Lys Lys Phe Ile Arg Gln
405 410 415
Trp Ser Gln Phe Phe Ser Gln Glu Ile Thr Val Asp Asp Ile Glu Leu
420 425 430
Ala Ser Trp Gly Leu Arg Gly Ala Ile Ala Pro Ile Arg Asn Glu Ile
435 440 445
Ile His Leu Lys Lys His Ser Trp Lys Lys Phe Phe Asn Asn Pro Thr

450 455 460
Phe Lys Val Lys Lys Ser Lys Ile Ile Asn Gly Lys Thr Lys Asp Val
465 470 475 480
Thr Ser Glu Phe Leu Tyr Lys Glu Thr Leu Phe Lys Asp Tyr Phe Tyr
485 490 495
Ser Glu Leu Asp Ser Val Pro Glu Leu Ile Ile Asn Lys Met Glu Ser
500 505 510
Ser Lys Ile Leu Asp Tyr Tyr Ser Ser Asp Gln Leu Asn Gln Val Phe
515 520 525

Thr Ile Pro Asn Phe Glu Leu Ser Leu Leu Thr Ser Ala Val Pro Phe

530 535 540

Ala Pro Ser Phe Lys Arg Val Tyr Leu Lys Gly Phe Asp Tyr Gln Asn

545 550 555 560

Gln Asp Glu Ala Gln Pro Asp Tyr Asn Leu Lys Leu Asn Ile Tyr Asn

565 570 575

Glu Lys Ala Phe Asn Ser Glu Ala Phe Gln Ala Gln Tyr Ser Leu Phe

580 585 590

Lys Met Val Tyr Tyr Gln Val Phe Leu Pro Gln Phe Thr Thr Asn Asn

595 600 605

Asp Leu Phe Lys Ser Ser Val Asp Phe Ile Leu Thr Leu Asn Lys Glu

610 615 620

Arg Lys Gly Tyr Ala Lys Ala Phe Gln Asp Ile Arg Lys Met Asn Lys

625 630 635 640

Asp Glu Lys Pro Ser Glu Tyr Met Ser Tyr Ile Gln Ser Gln Leu Met

645 650 655

Leu Tyr Gln Lys Lys Gln Glu Glu Lys Glu Lys Ile Asn His Phe Glu

660 665 670

Lys Phe Ile Asn Gln Val Phe Ile Lys Gly Phe Asn Ser Phe Ile Glu

675 680 685

Lys Asn Arg Leu Thr Tyr Ile Cys His Pro Thr Lys Asn Thr Val Pro

690 695 700

Glu Asn Asp Asn Ile Glu Ile Pro Phe His Thr Asp Met Asp Asp Ser

705 710 715 720

Asn Ile Ala Phe Trp Leu Met Cys Lys Leu Leu Asp Ala Lys Gln Leu

725 730 735

Ser Glu Leu Arg Asn Glu Met Ile Lys Phe Ser Cys Ser Leu Gln Ser

740 745 750

Thr Glu Glu Ile Ser Thr Phe Thr Lys Ala Arg Glu Val Ile Gly Leu

755 760 765

Ala Leu Leu Asn Gly Glu Lys Gly Cys Asn Asp Trp Lys Glu Leu Phe

770 775 780

Asp Asp Lys Glu Ala Trp Lys Lys Asn Met Ser Leu Tyr Val Ser Glu

785 790 795 800

Glu Leu Leu Gln Ser Leu Pro Tyr Thr Gln Glu Asp Gly Gln Thr Pro

805 810 815

Val Ile Asn Arg Ser Ile Asp Leu Val Lys Lys Tyr Gly Thr Glu Thr

820 825 830

Ile Leu Glu Lys Leu Phe Ser Ser Ser Asp Asp Tyr Lys Val Ser Ala

835 840 845

Lys Asp Ile Ala Lys Leu His Glu Tyr Asp Val Thr Glu Lys Ile Ala

850 855 860

Gln Gln Glu Ser Leu His Lys Gln Trp Ile Glu Lys Pro Gly Leu Ala

865 870 875 880

Arg Asp Ser Ala Trp Thr Lys Lys Tyr Gln Asn Val Ile Asn Asp Ile

885 890 895

Ser Asn Tyr Gln Trp Ala Lys Thr Lys Val Glu Leu Thr Gln Val Arg

900 905 910

His Leu His Gln Leu Thr Ile Asp Leu Leu Ser Arg Leu Ala Gly Tyr

915 920 925

Met Ser Ile Ala Asp Arg Asp Phe Gln Phe Ser Ser Asn Tyr Ile Leu

930 935 940

Glu Arg Glu Asn Ser Glu Tyr Arg Val Thr Ser Trp Ile Leu Leu Ser

945 950 955 960

Glu Asn Lys Asn Lys Asn Lys Tyr Asn Asp Tyr Glu Leu Tyr Asn Leu

965 970 975

Lys Asn Ala Ser Ile Lys Val Ser Ser Lys Asn Asp Pro Gln Leu Lys

980 985 990

Val Asp Leu Lys Gln Leu Arg Leu Thr Leu Glu Tyr Leu Glu Leu Phe

995 1000 1005

Asp Asn Arg Leu Lys Glu Lys Arg Asn Asn Ile Ser His Phe Asn Tyr

1010 1015 1020

Leu Asn Gly Gln Leu Gly Asn Ser Ile Leu Glu Leu Phe Asp Asp Ala

1025 1030 1035 1040

Arg Asp Val Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala Val Ser Lys
1045 1050 1055

Ser Leu Lys Glu Ile Leu Ser Ser His Gly Met Glu Val Thr Phe Lys
1060 1065 1070

Pro Leu Tyr Gln Thr Asn His His Leu Lys Ile Asp Lys Leu Gln Pro
1075 1080 1085

Lys Lys Ile His His Leu Gly Glu Lys Ser Thr Val Ser Ser Asn Gln
1090 1095 1100

Val Ser Asn Glu Tyr Cys Gln Leu Val Arg Thr Leu Leu Thr Met Lys
1105 1110 1115 1120

<210> 134

<211> 1159

<212> PRT

<213> Leptotrichia buccalis

<400> 134

Met Lys Val Thr Lys Val Gly Gly Ile Ser His Lys Lys Tyr Thr Ser
1 5 10 15

Glu Gly Arg Leu Val Lys Ser Glu Ser Glu Glu Asn Arg Thr Asp Glu
20 25 30

Arg Leu Ser Ala Leu Leu Asn Met Arg Leu Asp Met Tyr Ile Lys Asn
35 40 45

Pro Ser Ser Thr Glu Thr Lys Glu Asn Gln Lys Arg Ile Gly Lys Leu
50 55 60

Lys Lys Phe Phe Ser Asn Lys Met Val Tyr Leu Lys Asp Asn Thr Leu
65 70 75 80

Ser Leu Lys Asn Gly Lys Lys Glu Asn Ile Asp Arg Glu Tyr Ser Glu
85 90 95

Thr Asp Ile Leu Glu Ser Asp Val Arg Asp Lys Lys Asn Phe Ala Val
100 105 110

Leu Lys Lys Ile Tyr Leu Asn Glu Asn Val Asn Ser Glu Glu Leu Glu
115 120 125

Val Phe Arg Asn Asp Ile Lys Lys Lys Leu Asn Lys Ile Asn Ser Leu

130 135 140

Lys Tyr Ser Phe Glu Lys Asn Lys Ala Asn Tyr Gln Lys Ile Asn Glu

145 150 155 160

Asn Asn Ile Glu Lys Val Glu Gly Lys Ser Lys Arg Asn Ile Ile Tyr

165 170 175

Asp Tyr Tyr Arg Glu Ser Ala Lys Arg Asp Ala Tyr Val Ser Asn Val

180 185 190

Lys Glu Ala Phe Asp Lys Leu Tyr Lys Glu Glu Asp Ile Ala Lys Leu

195 200 205

Val Leu Glu Ile Glu Asn Leu Thr Lys Leu Glu Lys Tyr Lys Ile Arg

210 215 220

Glu Phe Tyr His Glu Ile Ile Gly Arg Lys Asn Asp Lys Glu Asn Phe

225 230 235 240

Ala Lys Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Met Lys Glu

245 250 255

Leu Ile Glu Lys Val Pro Asp Met Ser Glu Leu Lys Lys Ser Gln Val

260 265 270

Phe Tyr Lys Tyr Tyr Leu Asp Lys Glu Glu Leu Asn Asp Lys Asn Ile

275 280 285

Lys Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln Leu Leu

290 295 300

Lys Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp Lys Ile

305 310 315 320

Lys Arg Ile Phe Glu Tyr Gln Asn Leu Lys Lys Leu Ile Glu Asn Lys

325 330 335

Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys Tyr Asn

340 345 350

Tyr Tyr Leu Gln Asp Gly Glu Ile Ala Thr Ser Asp Phe Ile Ala Arg

355 360 365

Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val Ser Ser

370 375 380

Val Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn Glu Asn

385 390 395 400

Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn Lys Gly
405 410 415
Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn Glu Asn
420 425 430
Lys Lys Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser Tyr Asp
435 440 445
Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala Asn Ile
450 455 460
Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe Asn Leu

465 470 475 480
Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala Pro Ser
485 490 495
Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys Lys Leu
500 505 510
Lys Leu Lys Ile Phe Arg Gln Leu Asn Ser Ala Asn Val Phe Arg Tyr
515 520 525
Leu Glu Lys Tyr Lys Ile Leu Asn Tyr Leu Lys Arg Thr Arg Phe Glu
530 535 540

Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys Leu Tyr
545 550 555 560
Ser Arg Ile Asp Asp Leu Lys Asn Ser Leu Gly Ile Tyr Trp Lys Thr
565 570 575
Pro Lys Thr Asn Asp Asp Asn Lys Thr Lys Glu Ile Ile Asp Ala Gln
580 585 590
Ile Tyr Leu Leu Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Tyr Phe
595 600 605
Met Ser Asn Asn Gly Asn Phe Phe Glu Ile Ser Lys Glu Ile Ile Glu

610 615 620
Leu Asn Lys Asn Asp Lys Arg Asn Leu Lys Thr Gly Phe Tyr Lys Leu
625 630 635 640

Gln Lys Phe Glu Asp Ile Gln Glu Lys Ile Pro Lys Glu Tyr Leu Ala
645 650 655
Asn Ile Gln Ser Leu Tyr Met Ile Asn Ala Gly Asn Gln Asp Glu Glu
660 665 670
Glu Lys Asp Thr Tyr Ile Asp Phe Ile Gln Lys Ile Phe Leu Lys Gly
675 680 685

Phe Met Thr Tyr Leu Ala Asn Asn Gly Arg Leu Ser Leu Ile Tyr Ile
690 695 700
Gly Ser Asp Glu Glu Thr Asn Thr Ser Leu Ala Glu Lys Lys Gln Glu
705 710 715 720
Phe Asp Lys Phe Leu Lys Lys Tyr Glu Gln Asn Asn Asn Ile Lys Ile
725 730 735
Pro Tyr Glu Ile Asn Glu Phe Leu Arg Glu Ile Lys Leu Gly Asn Ile
740 745 750
Leu Lys Tyr Thr Glu Arg Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu

755 760 765
Leu Asn His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr
770 775 780
Gln Ser Ala Asn Lys Glu Glu Ala Phe Ser Asp Gln Leu Glu Leu Ile
785 790 795 800
Asn Leu Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu
805 810 815
Glu Ala Asp Glu Ile Gly Lys Phe Leu Asp Phe Asn Gly Asn Lys Val
820 825 830

Lys Asp Asn Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe
835 840 845
Asp Gly Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys
850 855 860
Tyr Gly Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Gly Tyr
865 870 875 880
Lys Ile Ser Ile Glu Glu Leu Lys Lys Tyr Ser Asn Lys Lys Asn Glu
885 890 895

Ile Glu Lys Asn His Lys Met Gln Glu Asn Leu His Arg Lys Tyr Ala

900 905 910

Arg Pro Arg Lys Asp Glu Lys Phe Thr Asp Glu Asp Tyr Glu Ser Tyr

915 920 925

Lys Gln Ala Ile Glu Asn Ile Glu Glu Tyr Thr His Leu Lys Asn Lys

930 935 940

Val Glu Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Arg Ile

945 950 955 960

Leu His Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg

965 970 975

Phe Arg Leu Lys Gly Glu Phe Pro Glu Asn Gln Tyr Ile Glu Glu Ile

980 985 990

Phe Asn Phe Glu Asn Lys Lys Asn Val Lys Tyr Lys Gly Gly Gln Ile

995 1000 1005

Val Glu Lys Tyr Ile Lys Phe Tyr Lys Glu Leu His Gln Asn Asp Glu

1010 1015 1020

Val Lys Ile Asn Lys Tyr Ser Ser Ala Asn Ile Lys Val Leu Lys Gln

1025 1030 1035 1040

Glu Lys Lys Asp Leu Tyr Ile Arg Asn Tyr Ile Ala His Phe Asn Tyr

1045 1050 1055

Ile Pro His Ala Glu Ile Ser Leu Leu Glu Val Leu Glu Asn Leu Arg

1060 1065 1070

Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala Val Met Lys Ser

1075 1080 1085

Val Val Asp Ile Leu Lys Glu Tyr Gly Phe Val Ala Thr Phe Lys Ile

1090 1095 1100

Gly Ala Asp Lys Lys Ile Gly Ile Gln Thr Leu Glu Ser Glu Lys Ile

1105 1110 1115 1120

Val His Leu Lys Asn Leu Lys Lys Lys Lys Leu Met Thr Asp Arg Asn

1125 1130 1135

Ser Glu Glu Leu Cys Lys Leu Val Lys Ile Met Phe Glu Tyr Lys Met

1140 1145 1150

Glu Glu Lys Lys Ser Glu Asn

1155

<210> 135

<211> 1285

<212> PRT

<213> Herbinix hemicellulosilytica

<400> 135

Met Lys Leu Thr Arg Arg Arg Ile Ser Gly Asn Ser Val Asp Gln Lys
1 5 10 15

Ile Thr Ala Ala Phe Tyr Arg Asp Met Ser Gln Gly Leu Leu Tyr Tyr
20 25 30

Asp Ser Glu Asp Asn Asp Cys Thr Asp Lys Val Ile Glu Ser Met Asp
35 40 45

Phe Glu Arg Ser Trp Arg Gly Arg Ile Leu Lys Asn Gly Glu Asp Asp
50 55 60

Lys Asn Pro Phe Tyr Met Phe Val Lys Gly Leu Val Gly Ser Asn Asp
65 70 75 80

Lys Ile Val Cys Glu Pro Ile Asp Val Asp Ser Asp Pro Asp Asn Leu

85 90 95

Asp Ile Leu Ile Asn Lys Asn Leu Thr Gly Phe Gly Arg Asn Leu Lys
100 105 110

Ala Pro Asp Ser Asn Asp Thr Leu Glu Asn Leu Ile Arg Lys Ile Gln
115 120 125

Ala Gly Ile Pro Glu Glu Glu Val Leu Pro Glu Leu Lys Lys Ile Lys
130 135 140

Glu Met Ile Gln Lys Asp Ile Val Asn Arg Lys Glu Gln Leu Leu Lys
145 150 155 160

Ser Ile Lys Asn Asn Arg Ile Pro Phe Ser Leu Glu Gly Ser Lys Leu

165 170 175

Val Pro Ser Thr Lys Lys Met Lys Trp Leu Phe Lys Leu Ile Asp Val
180 185 190

Pro Asn Lys Thr Phe Asn Glu Lys Met Leu Glu Lys Tyr Trp Glu Ile
195 200 205
Tyr Asp Tyr Asp Lys Leu Lys Ala Asn Ile Thr Asn Arg Leu Asp Lys
210 215 220
Thr Asp Lys Lys Ala Arg Ser Ile Ser Arg Ala Val Ser Glu Glu Leu

225 230 235 240
Arg Glu Tyr His Lys Asn Leu Arg Thr Asn Tyr Asn Arg Phe Val Ser
245 250 255
Gly Asp Arg Pro Ala Ala Gly Leu Asp Asn Gly Gly Ser Ala Lys Tyr
260 265 270
Asn Pro Asp Lys Glu Glu Phe Leu Leu Phe Leu Lys Glu Val Glu Gln
275 280 285
Tyr Phe Lys Lys Tyr Phe Pro Val Lys Ser Lys His Ser Asn Lys Ser
290 295 300

Lys Asp Lys Ser Leu Val Asp Lys Tyr Lys Asn Tyr Cys Ser Tyr Lys
305 310 315 320
Val Val Lys Lys Glu Val Asn Arg Ser Ile Ile Asn Gln Leu Val Ala
325 330 335
Gly Leu Ile Gln Gln Gly Lys Leu Leu Tyr Tyr Phe Tyr Tyr Asn Asp
340 345 350
Thr Trp Gln Glu Asp Phe Leu Asn Ser Tyr Gly Leu Ser Tyr Ile Gln
355 360 365
Val Glu Glu Ala Phe Lys Lys Ser Val Met Thr Ser Leu Ser Trp Gly

370 375 380
Ile Asn Arg Leu Thr Ser Phe Phe Ile Asp Asp Ser Asn Thr Val Lys
385 390 395 400
Phe Asp Asp Ile Thr Thr Lys Lys Ala Lys Glu Ala Ile Glu Ser Asn
405 410 415
Tyr Phe Asn Lys Leu Arg Thr Cys Ser Arg Met Gln Asp His Phe Lys
420 425 430
Glu Lys Leu Ala Phe Phe Tyr Pro Val Tyr Val Lys Asp Lys Lys Asp

435 440 445

Arg Pro Asp Asp Asp Ile Glu Asn Leu Ile Val Leu Val Lys Asn Ala
450 455 460
Ile Glu Ser Val Ser Tyr Leu Arg Asn Arg Thr Phe His Phe Lys Glu
465 470 475 480
Ser Ser Leu Leu Glu Leu Leu Lys Glu Leu Asp Asp Lys Asn Ser Gly
485 490 495
Gln Asn Lys Ile Asp Tyr Ser Val Ala Ala Glu Phe Ile Lys Arg Asp
500 505 510
Ile Glu Asn Leu Tyr Asp Val Phe Arg Glu Gln Ile Arg Ser Leu Gly

515 520 525
Ile Ala Glu Tyr Tyr Lys Ala Asp Met Ile Ser Asp Cys Phe Lys Thr
530 535 540
Cys Gly Leu Glu Phe Ala Leu Tyr Ser Pro Lys Asn Ser Leu Met Pro
545 550 555 560
Ala Phe Lys Asn Val Tyr Lys Arg Gly Ala Asn Leu Asn Lys Ala Tyr
565 570 575
Ile Arg Asp Lys Gly Pro Lys Glu Thr Gly Asp Gln Gly Gln Asn Ser
580 585 590

Tyr Lys Ala Leu Glu Glu Tyr Arg Glu Leu Thr Trp Tyr Ile Glu Val
595 600 605
Lys Asn Asn Asp Gln Ser Tyr Asn Ala Tyr Lys Asn Leu Leu Gln Leu
610 615 620
Ile Tyr Tyr His Ala Phe Leu Pro Glu Val Arg Glu Asn Glu Ala Leu
625 630 635 640
Ile Thr Asp Phe Ile Asn Arg Thr Lys Glu Trp Asn Arg Lys Glu Thr
645 650 655
Glu Glu Arg Leu Asn Thr Lys Asn Asn Lys Lys His Lys Asn Phe Asp

660 665 670
Glu Asn Asp Asp Ile Thr Val Asn Thr Tyr Arg Tyr Glu Ser Ile Pro
675 680 685

Asp Tyr Gln Gly Glu Ser Leu Asp Asp Tyr Leu Lys Val Leu Gln Arg
690 695 700
Lys Gln Met Ala Arg Ala Lys Glu Val Asn Glu Lys Glu Glu Gly Asn
705 710 715 720
Asn Asn Tyr Ile Gln Phe Ile Arg Asp Val Val Val Trp Ala Phe Gly
725 730 735

Ala Tyr Leu Glu Asn Lys Leu Lys Asn Tyr Lys Asn Glu Leu Gln Pro
740 745 750
Pro Leu Ser Lys Glu Asn Ile Gly Leu Asn Asp Thr Leu Lys Glu Leu
755 760 765
Phe Pro Glu Glu Lys Val Lys Ser Pro Phe Asn Ile Lys Cys Arg Phe
770 775 780
Ser Ile Ser Thr Phe Ile Asp Asn Lys Gly Lys Ser Thr Asp Asn Thr
785 790 795 800
Ser Ala Glu Ala Val Lys Thr Asp Gly Lys Glu Asp Glu Lys Asp Lys

805 810 815
Lys Asn Ile Lys Arg Lys Asp Leu Leu Cys Phe Tyr Leu Phe Leu Arg
820 825 830
Leu Leu Asp Glu Asn Glu Ile Cys Lys Leu Gln His Gln Phe Ile Lys
835 840 845
Tyr Arg Cys Ser Leu Lys Glu Arg Arg Phe Pro Gly Asn Arg Thr Lys
850 855 860
Leu Glu Lys Glu Thr Glu Leu Leu Ala Glu Leu Glu Glu Leu Met Glu
865 870 875 880

Leu Val Arg Phe Thr Met Pro Ser Ile Pro Glu Ile Ser Ala Lys Ala
885 890 895
Glu Ser Gly Tyr Asp Thr Met Ile Lys Lys Tyr Phe Lys Asp Phe Ile
900 905 910
Glu Lys Lys Val Phe Lys Asn Pro Lys Thr Ser Asn Leu Tyr Tyr His
915 920 925
Ser Asp Ser Lys Thr Pro Val Thr Arg Lys Tyr Met Ala Leu Leu Met
930 935 940

Arg Ser Ala Pro Leu His Leu Tyr Lys Asp Ile Phe Lys Gly Tyr Tyr

945 950 955 960

Leu Ile Thr Lys Lys Glu Cys Leu Glu Tyr Ile Lys Leu Ser Asn Ile

965 970 975

Ile Lys Asp Tyr Gln Asn Ser Leu Asn Glu Leu His Glu Gln Leu Glu

980 985 990

Arg Ile Lys Leu Lys Ser Glu Lys Gln Asn Gly Lys Asp Ser Leu Tyr

995 1000 1005

Leu Asp Lys Lys Asp Phe Tyr Lys Val Lys Glu Tyr Val Glu Asn Leu

1010 1015 1020

Glu Gln Val Ala Arg Tyr Lys His Leu Gln His Lys Ile Asn Phe Glu

1025 1030 1035 1040

Ser Leu Tyr Arg Ile Phe Arg Ile His Val Asp Ile Ala Ala Arg Met

1045 1050 1055

Val Gly Tyr Thr Gln Asp Trp Glu Arg Asp Met His Phe Leu Phe Lys

1060 1065 1070

Ala Leu Val Tyr Asn Gly Val Leu Glu Glu Arg Arg Phe Glu Ala Ile

1075 1080 1085

Phe Asn Asn Asn Asp Asp Asn Asn Asp Gly Arg Ile Val Lys Lys Ile

1090 1095 1100

Gln Asn Asn Leu Asn Asn Lys Asn Arg Glu Leu Val Ser Met Leu Cys

1105 1110 1115 1120

Trp Asn Lys Lys Leu Asn Lys Asn Glu Phe Gly Ala Ile Ile Trp Lys

1125 1130 1135

Arg Asn Pro Ile Ala His Leu Asn His Phe Thr Gln Thr Glu Gln Asn

1140 1145 1150

Ser Lys Ser Ser Leu Glu Ser Leu Ile Asn Ser Leu Arg Ile Leu Leu

1155 1160 1165

Ala Tyr Asp Arg Lys Arg Gln Asn Ala Val Thr Lys Thr Ile Asn Asp

1170 1175 1180

Leu Leu Leu Asn Asp Tyr His Ile Arg Ile Lys Trp Glu Gly Arg Val

1185 1190 1195 1200

Asp Glu Gly Gln Ile Tyr Phe Asn Ile Lys Glu Lys Glu Asp Ile Glu

1205 1210 1215

Asn Glu Pro Ile Ile His Leu Lys His Leu His Lys Lys Asp Cys Tyr

1220 1225 1230

Ile Tyr Lys Asn Ser Tyr Met Phe Asp Lys Gln Lys Glu Trp Ile Cys

1235 1240 1245

Asn Gly Ile Lys Glu Glu Val Tyr Asp Lys Ser Ile Leu Lys Cys Ile

1250 1255 1260

Gly Asn Leu Phe Lys Phe Asp Tyr Glu Asp Lys Asn Lys Ser Ser Ala

1265 1270 1275 1280

Asn Pro Lys His Thr

1285

<210> 136

<211> 1344

<212> PRT

<213> Eubacterium] rectale

<400> 136

Met Leu Arg Arg Asp Lys Glu Val Lys Lys Leu Tyr Asn Val Phe Asn

1 5 10 15

Gln Ile Gln Val Gly Thr Lys Pro Lys Lys Trp Asn Asn Asp Glu Lys

20 25 30

Leu Ser Pro Glu Glu Asn Glu Arg Arg Ala Gln Gln Lys Asn Ile Lys

35 40 45

Met Lys Asn Tyr Lys Trp Arg Glu Ala Cys Ser Lys Tyr Val Glu Ser

50 55 60

Ser Gln Arg Ile Ile Asn Asp Val Ile Phe Tyr Ser Tyr Arg Lys Ala

65 70 75 80

Lys Asn Lys Leu Arg Tyr Met Arg Lys Asn Glu Asp Ile Leu Lys Lys

85 90 95

Met Gln Glu Ala Glu Lys Leu Ser Lys Phe Ser Gly Gly Lys Leu Glu

100 105 110

Asp Phe Val Ala Tyr Thr Leu Arg Lys Ser Leu Val Val Ser Lys Tyr
115 120 125
Asp Thr Gln Glu Phe Asp Ser Leu Ala Ala Met Val Val Phe Leu Glu
130 135 140
Cys Ile Gly Lys Asn Asn Ile Ser Asp His Glu Arg Glu Ile Val Cys
145 150 155 160

Lys Leu Leu Glu Leu Ile Arg Lys Asp Phe Ser Lys Leu Asp Pro Asn
165 170 175
Val Lys Gly Ser Gln Gly Ala Asn Ile Val Arg Ser Val Arg Asn Gln
180 185 190
Asn Met Ile Val Gln Pro Gln Gly Asp Arg Phe Leu Phe Pro Gln Val
195 200 205
Tyr Ala Lys Glu Asn Glu Thr Val Thr Asn Lys Asn Val Glu Lys Glu
210 215 220
Gly Leu Asn Glu Phe Leu Leu Asn Tyr Ala Asn Leu Asp Asp Glu Lys

225 230 235 240
Arg Ala Glu Ser Leu Arg Lys Leu Arg Arg Ile Leu Asp Val Tyr Phe
245 250 255
Ser Ala Pro Asn His Tyr Glu Lys Asp Met Asp Ile Thr Leu Ser Asp
260 265 270
Asn Ile Glu Lys Glu Lys Phe Asn Val Trp Glu Lys His Glu Cys Gly
275 280 285
Lys Lys Glu Thr Gly Leu Phe Val Asp Ile Pro Asp Val Leu Met Glu
290 295 300

Ala Glu Ala Glu Asn Ile Lys Leu Asp Ala Val Val Glu Lys Arg Glu
305 310 315 320
Arg Lys Val Leu Asn Asp Arg Val Arg Lys Gln Asn Ile Ile Cys Tyr
325 330 335
Arg Tyr Thr Arg Ala Val Val Glu Lys Tyr Asn Ser Asn Glu Pro Leu
340 345 350
Phe Phe Glu Asn Asn Ala Ile Asn Gln Tyr Trp Ile His His Ile Glu
355 360 365

Asn Ala Val Glu Arg Ile Leu Lys Asn Cys Lys Ala Gly Lys Leu Phe

370 375 380

Lys Leu Arg Lys Gly Tyr Leu Ala Glu Lys Val Trp Lys Asp Ala Ile
385 390 395 400
Asn Leu Ile Ser Ile Lys Tyr Ile Ala Leu Gly Lys Ala Val Tyr Asn
405 410 415
Phe Ala Leu Asp Asp Ile Trp Lys Asp Lys Lys Asn Lys Glu Leu Gly
420 425 430
Ile Val Asp Glu Arg Ile Arg Asn Gly Ile Thr Ser Phe Asp Tyr Glu
435 440 445

Met Ile Lys Ala His Glu Asn Leu Gln Arg Glu Leu Ala Val Asp Ile
450 455 460
Ala Phe Ser Val Asn Asn Leu Ala Arg Ala Val Cys Asp Met Ser Asn
465 470 475 480
Leu Gly Asn Lys Glu Ser Asp Phe Leu Leu Trp Lys Arg Asn Asp Ile
485 490 495
Ala Asp Lys Leu Lys Asn Lys Asp Asp Met Ala Ser Val Ser Ala Val
500 505 510
Leu Gln Phe Phe Gly Gly Lys Ser Ser Trp Asp Ile Asn Ile Phe Lys

515 520 525

Asp Ala Tyr Lys Gly Lys Lys Lys Tyr Asn Tyr Glu Val Arg Phe Ile
530 535 540
Asp Asp Leu Arg Lys Ala Ile Tyr Cys Ala Arg Asn Glu Asn Phe His
545 550 555 560
Phe Lys Thr Ala Leu Val Asn Asp Glu Lys Trp Asn Thr Glu Leu Phe
565 570 575
Gly Lys Ile Phe Glu Arg Glu Thr Glu Phe Cys Leu Asn Val Glu Lys
580 585 590

Asp Arg Phe Tyr Ser Asn Asn Leu Tyr Met Phe Tyr Gln Val Ser Glu

595 600 605

Leu Arg Asn Met Leu Asp His Leu Tyr Ser Arg Ser Val Ser Arg Ala

610 615 620

Ala Gln Val Pro Ser Tyr Asn Ser Val Ile Val Arg Thr Ala Phe Pro

625 630 635 640

Glu Tyr Ile Thr Asn Val Leu Gly Tyr Gln Lys Pro Ser Tyr Asp Ala

645 650 655

Asp Thr Leu Gly Lys Trp Tyr Ser Ala Cys Tyr Tyr Leu Leu Lys Glu

660 665 670

Ile Tyr Tyr Asn Ser Phe Leu Gln Ser Asp Arg Ala Leu Gln Leu Phe

675 680 685

Glu Lys Ser Val Lys Thr Leu Ser Trp Asp Asp Lys Lys Gln Gln Arg

690 695 700

Ala Val Asp Asn Phe Lys Asp His Phe Ser Asp Ile Lys Ser Ala Cys

705 710 715 720

Thr Ser Leu Ala Gln Val Cys Gln Ile Tyr Met Thr Glu Tyr Asn Gln

725 730 735

Gln Asn Asn Gln Ile Lys Lys Val Arg Ser Ser Asn Asp Ser Ile Phe

740 745 750

Asp Gln Pro Val Tyr Gln His Tyr Lys Val Leu Leu Lys Lys Ala Ile

755 760 765

Ala Asn Ala Phe Ala Asp Tyr Leu Lys Asn Asn Lys Asp Leu Phe Gly

770 775 780

Phe Ile Gly Lys Pro Phe Lys Ala Asn Glu Ile Arg Glu Ile Asp Lys

785 790 795 800

Glu Gln Phe Leu Pro Asp Trp Thr Ser Arg Lys Tyr Glu Ala Leu Cys

805 810 815

Ile Glu Val Ser Gly Ser Gln Glu Leu Gln Lys Trp Tyr Ile Val Gly

820 825 830

Lys Phe Leu Asn Ala Arg Ser Leu Asn Leu Met Val Gly Ser Met Arg

835 840 845

Ser Tyr Ile Gln Tyr Val Thr Asp Ile Lys Arg Arg Ala Ala Ser Ile

850 855 860

Gly Asn Glu Leu His Val Ser Val His Asp Val Glu Lys Val Glu Lys

865 870 875 880

Trp Val Gln Val Ile Glu Val Cys Ser Leu Leu Ala Ser Arg Thr Ser
885 890 895
Asn Gln Phe Glu Asp Tyr Phe Asn Asp Lys Asp Asp Tyr Ala Arg Tyr
900 905 910
Leu Lys Ser Tyr Val Asp Phe Ser Asn Val Asp Met Pro Ser Glu Tyr
915 920 925
Ser Ala Leu Val Asp Phe Ser Asn Glu Glu Gln Ser Asp Leu Tyr Val
930 935 940
Asp Pro Lys Asn Pro Lys Val Asn Arg Asn Ile Val His Ser Lys Leu

945 950 955 960
Phe Ala Ala Asp His Ile Leu Arg Asp Ile Val Glu Pro Val Ser Lys
965 970 975
Asp Asn Ile Glu Glu Phe Tyr Ser Gln Lys Ala Glu Ile Ala Tyr Cys
980 985 990
Lys Ile Lys Gly Lys Glu Ile Thr Ala Glu Glu Gln Lys Ala Val Leu
995 1000 1005
Lys Tyr Gln Lys Leu Lys Asn Arg Val Glu Leu Arg Asp Ile Val Glu
1010 1015 1020

Tyr Gly Glu Ile Ile Asn Glu Leu Leu Gly Gln Leu Ile Asn Trp Ser
1025 1030 1035 1040
Phe Met Arg Glu Arg Asp Leu Leu Tyr Phe Gln Leu Gly Phe His Tyr
1045 1050 1055
Asp Cys Leu Arg Asn Asp Ser Lys Lys Pro Glu Gly Tyr Lys Asn Ile
1060 1065 1070
Lys Val Asp Glu Asn Ser Ile Lys Asp Ala Ile Leu Tyr Gln Ile Ile
1075 1080 1085
Gly Met Tyr Val Asn Gly Val Thr Val Tyr Ala Pro Glu Lys Asp Gly

1090 1095 1100
Asp Lys Leu Lys Glu Gln Cys Val Lys Gly Gly Val Gly Val Lys Val
1105 1110 1115 1120

Ser Ala Phe His Arg Tyr Ser Lys Tyr Leu Gly Leu Asn Glu Lys Thr
1125 1130 1135
Leu Tyr Asn Ala Gly Leu Glu Ile Phe Glu Val Val Ala Glu His Glu
1140 1145 1150
Asp Ile Ile Asn Leu Arg Asn Gly Ile Asp His Phe Lys Tyr Tyr Leu
1155 1160 1165

Gly Asp Tyr Arg Ser Met Leu Ser Ile Tyr Ser Glu Val Phe Asp Arg
1170 1175 1180
Phe Phe Thr Tyr Asp Ile Lys Tyr Gln Lys Asn Val Leu Asn Leu Leu
1185 1190 1195 1200
Gln Asn Ile Leu Leu Arg His Asn Val Ile Val Glu Pro Ile Leu Glu
1205 1210 1215
Ser Gly Phe Lys Thr Ile Gly Glu Gln Thr Lys Pro Gly Ala Lys Leu
1220 1225 1230
Ser Ile Arg Ser Ile Lys Ser Asp Thr Phe Gln Tyr Lys Val Lys Gly

1235 1240 1245
Gly Thr Leu Ile Thr Asp Ala Lys Asp Glu Arg Tyr Leu Glu Thr Ile
1250 1255 1260
Arg Lys Ile Leu Tyr Tyr Ala Glu Asn Glu Glu Asp Asn Leu Lys Lys
1265 1270 1275 1280
Ser Val Val Val Thr Asn Ala Asp Lys Tyr Glu Lys Asn Lys Glu Ser
1285 1290 1295
Asp Asp Gln Asn Lys Gln Lys Glu Lys Lys Asn Lys Asp Asn Lys Gly
1300 1305 1310

Lys Lys Asn Glu Glu Thr Lys Ser Asp Ala Glu Lys Asn Asn Asn Glu
1315 1320 1325
Arg Leu Ser Tyr Asn Pro Phe Ala Asn Leu Asn Phe Lys Leu Ser Asn
1330 1335 1340
<210> 137
<211> 1373
<212> PRT
<213> Eubacteriaceae bacterium

<400> 137

Met Lys Ile Ser Lys Glu Ser His Lys Arg Thr Ala Val Ala Val Met
1 5 10 15

Glu Asp Arg Val Gly Gly Val Val Tyr Val Pro Gly Gly Ser Gly Ile
20 25 30

Asp Leu Ser Asn Asn Leu Lys Lys Arg Ser Met Asp Thr Lys Ser Leu
35 40 45

Tyr Asn Val Phe Asn Gln Ile Gln Ala Gly Thr Ala Pro Ser Glu Tyr
50 55 60

Glu Trp Lys Asp Tyr Leu Ser Glu Ala Glu Asn Lys Lys Arg Glu Ala
65 70 75 80

Gln Lys Met Ile Gln Lys Ala Asn Tyr Glu Leu Arg Arg Glu Cys Glu
85 90 95

Asp Tyr Ala Lys Lys Ala Asn Leu Ala Val Ser Arg Ile Ile Phe Ser
100 105 110

Lys Lys Pro Lys Lys Ile Phe Ser Asp Asp Asp Ile Ile Ser His Met
115 120 125

Lys Lys Gln Arg Leu Ser Lys Phe Lys Gly Arg Met Glu Asp Phe Val
130 135 140

Leu Ile Ala Leu Arg Lys Ser Leu Val Val Ser Thr Tyr Asn Gln Glu
145 150 155 160

Val Phe Asp Ser Arg Lys Ala Ala Thr Val Phe Leu Lys Asn Ile Gly
165 170 175

Lys Lys Asn Ile Ser Ala Asp Asp Glu Arg Gln Ile Lys Gln Leu Met
180 185 190

Ala Leu Ile Arg Glu Asp Tyr Asp Lys Trp Asn Pro Asp Lys Asp Ser
195 200 205

Ser Asp Lys Lys Glu Ser Ser Gly Thr Lys Val Ile Arg Ser Ile Glu
210 215 220

His Gln Asn Met Val Ile Gln Pro Glu Lys Asn Lys Leu Ser Leu Ser
225 230 235 240

Lys Ile Ser Asn Val Gly Lys Lys Thr Lys Thr Lys Gln Lys Glu Lys
245 250 255
Ala Gly Leu Asp Ala Phe Leu Lys Glu Tyr Ala Gln Ile Asp Glu Asn
260 265 270
Ser Arg Met Glu Tyr Leu Lys Lys Leu Arg Arg Leu Leu Asp Thr Tyr
275 280 285
Phe Ala Ala Pro Ser Ser Tyr Ile Lys Gly Ala Ala Val Ser Leu Pro
290 295 300
Glu Asn Ile Asn Phe Ser Ser Glu Leu Asn Val Trp Glu Arg His Glu

305 310 315 320
Ala Ala Lys Lys Val Asn Ile Asn Phe Val Glu Ile Pro Glu Ser Leu
325 330 335
Leu Asn Ala Glu Gln Asn Asn Asn Lys Ile Asn Lys Val Glu Gln Glu
340 345 350
His Ser Leu Glu Gln Leu Arg Thr Asp Ile Arg Arg Arg Asn Ile Thr
355 360 365
Cys Tyr His Phe Ala Asn Ala Leu Ala Ala Asp Glu Arg Tyr His Thr
370 375 380

Leu Phe Phe Glu Asn Met Ala Met Asn Gln Phe Trp Ile His His Met
385 390 395 400
Glu Asn Ala Val Glu Arg Ile Leu Lys Lys Cys Asn Val Gly Thr Leu
405 410 415
Phe Lys Leu Arg Ile Gly Tyr Leu Ser Glu Lys Val Trp Lys Asp Met
420 425 430
Leu Asn Leu Leu Ser Ile Lys Tyr Ile Ala Leu Gly Lys Ala Val Tyr
435 440 445
His Phe Ala Leu Asp Asp Ile Trp Lys Ala Asp Ile Trp Lys Asp Ala

450 455 460
Ser Asp Lys Asn Ser Gly Lys Ile Asn Asp Leu Thr Leu Lys Gly Ile
465 470 475 480
Ser Ser Phe Asp Tyr Glu Met Val Lys Ala Gln Glu Asp Leu Gln Arg
485 490 495

Glu Met Ala Val Gly Val Ala Phe Ser Thr Asn Asn Leu Ala Arg Val
500 505 510
Thr Cys Lys Met Asp Asp Leu Ser Asp Ala Glu Ser Asp Phe Leu Leu
515 520 525

Trp Asn Lys Glu Ala Ile Arg Arg His Val Lys Tyr Thr Glu Lys Gly
530 535 540
Glu Ile Leu Ser Ala Ile Leu Gln Phe Phe Gly Gly Arg Ser Leu Trp
545 550 555 560
Asp Glu Ser Leu Phe Glu Lys Ala Tyr Ser Asp Ser Asn Tyr Glu Leu
565 570 575
Lys Phe Leu Asp Asp Leu Lys Arg Ala Ile Tyr Ala Ala Arg Asn Glu
580 585 590
Thr Phe His Phe Lys Thr Ala Ala Ile Asp Gly Gly Ser Trp Asn Thr
595 600 605
Arg Leu Phe Gly Ser Leu Phe Glu Lys Glu Ala Gly Leu Cys Leu Asn
610 615 620
Val Glu Lys Asn Lys Phe Tyr Ser Asn Asn Leu Val Leu Phe Tyr Lys
625 630 635 640
Gln Glu Asp Leu Arg Val Phe Leu Asp Lys Leu Tyr Gly Lys Glu Cys
645 650 655
Ser Arg Ala Ala Gln Ile Pro Ser Tyr Asn Thr Ile Leu Pro Arg Lys
660 665 670

Ser Phe Ser Asp Phe Met Lys Gln Leu Leu Gly Leu Lys Glu Pro Val
675 680 685
Tyr Gly Ser Ala Ile Leu Asp Gln Trp Tyr Ser Ala Cys Tyr Tyr Leu
690 695 700
Phe Lys Glu Val Tyr Tyr Asn Leu Phe Leu Gln Asp Ser Ser Ala Lys
705 710 715 720
Ala Leu Phe Glu Lys Ala Val Lys Ala Leu Lys Gly Ala Asp Lys Lys
725 730 735
Gln Glu Lys Ala Val Glu Ser Phe Arg Lys Arg Tyr Trp Glu Ile Ser

740 745 750

Lys Asn Ala Ser Leu Ala Glu Ile Cys Gln Ser Tyr Ile Thr Glu Tyr

755 760 765

Asn Gln Gln Asn Asn Lys Glu Arg Lys Val Arg Ser Ala Asn Asp Gly

770 775 780

Met Phe Asn Glu Pro Ile Tyr Gln His Tyr Lys Met Leu Leu Lys Glu

785 790 795 800

Ala Leu Lys Met Ala Phe Ala Ser Tyr Ile Lys Asn Asp Lys Glu Leu

805 810 815

Lys Phe Val Tyr Lys Pro Thr Glu Lys Leu Phe Glu Val Ser Gln Asp

820 825 830

Asn Phe Leu Pro Asn Trp Asn Ser Glu Lys Tyr Asn Thr Leu Ile Ser

835 840 845

Glu Val Lys Asn Ser Pro Asp Leu Gln Lys Trp Tyr Ile Val Gly Lys

850 855 860

Phe Met Asn Ala Arg Met Leu Asn Leu Leu Leu Gly Ser Met Arg Ser

865 870 875 880

Tyr Leu Gln Tyr Val Ser Asp Ile Gln Lys Arg Ala Ala Gly Leu Gly

885 890 895

Glu Asn Gln Leu His Leu Ser Ala Glu Asn Val Gly Gln Val Lys Lys

900 905 910

Trp Ile Gln Val Leu Glu Val Cys Leu Leu Leu Ser Val Arg Ile Ser

915 920 925

Asp Lys Phe Thr Asp Tyr Phe Lys Asp Glu Glu Glu Tyr Ala Ser Tyr

930 935 940

Leu Lys Glu Tyr Val Asp Phe Glu Asp Ser Ala Met Pro Ser Asp Tyr

945 950 955 960

Ser Ala Leu Leu Ala Phe Ser Asn Glu Gly Lys Ile Asp Leu Tyr Val

965 970 975

Asp Ala Ser Asn Pro Lys Val Asn Arg Asn Ile Ile Gln Ala Lys Leu

980 985 990

Tyr Ala Pro Asp Met Val Leu Lys Lys Val Val Lys Lys Ile Ser Gln

995 1000 1005

Asp Glu Cys Lys Glu Phe Asn Glu Lys Lys Glu Gln Ile Met Gln Phe

1010 1015 1020

Lys Asn Lys Gly Asp Glu Val Ser Trp Glu Glu Gln Gln Lys Ile Leu

1025 1030 1035 1040

Glu Tyr Gln Lys Leu Lys Asn Arg Val Glu Leu Arg Asp Leu Ser Glu

1045 1050 1055

Tyr Gly Glu Leu Ile Asn Glu Leu Leu Gly Gln Leu Ile Asn Trp Ser

1060 1065 1070

Tyr Leu Arg Glu Arg Asp Leu Leu Tyr Phe Gln Leu Gly Phe His Tyr

1075 1080 1085

Ser Cys Leu Met Asn Glu Ser Lys Lys Pro Asp Ala Tyr Lys Thr Ile

1090 1095 1100

Arg Arg Gly Thr Val Ser Ile Glu Asn Ala Val Leu Tyr Gln Ile Ile

1105 1110 1115 1120

Ala Met Tyr Ile Asn Gly Phe Pro Val Tyr Ala Pro Glu Lys Gly Glu

1125 1130 1135

Leu Lys Pro Gln Cys Lys Thr Gly Ser Ala Gly Gln Lys Ile Arg Ala

1140 1145 1150

Phe Cys Gln Trp Ala Ser Met Val Glu Lys Lys Lys Tyr Glu Leu Tyr

1155 1160 1165

Asn Ala Gly Leu Glu Leu Phe Glu Val Val Lys Glu His Asp Asn Ile

1170 1175 1180

Ile Asp Leu Arg Asn Lys Ile Asp His Phe Lys Tyr Tyr Gln Gly Asn

1185 1190 1195 1200

Asp Ser Ile Leu Ala Leu Tyr Gly Glu Ile Phe Asp Arg Phe Phe Thr

1205 1210 1215

Tyr Asp Met Lys Tyr Arg Asn Asn Val Leu Asn His Leu Gln Asn Ile

1220 1225 1230

Leu Leu Arg His Asn Val Ile Ile Lys Pro Ile Ile Ser Lys Asp Lys

1235 1240 1245

Lys Glu Val Gly Arg Gly Lys Met Lys Asp Arg Ala Ala Phe Leu Leu

1250 1255 1260

Glu Glu Val Ser Ser Asp Arg Phe Thr Tyr Lys Val Lys Glu Gly Glu

1265 1270 1275 1280

Arg Lys Ile Asp Ala Lys Asn Arg Leu Tyr Leu Glu Thr Val Arg Asp

1285 1290 1295

Ile Leu Tyr Phe Pro Asn Arg Ala Val Asn Asp Lys Gly Glu Asp Val

1300 1305 1310

Ile Ile Cys Ser Lys Lys Ala Gln Asp Leu Asn Glu Lys Lys Ala Asp

1315 1320 1325

Arg Asp Lys Asn His Asp Lys Ser Lys Asp Thr Asn Gln Lys Lys Glu

1330 1335 1340

Gly Lys Asn Gln Glu Glu Lys Ser Glu Asn Lys Glu Pro Tyr Ser Asp

1345 1350 1355 1360

Arg Met Thr Trp Lys Pro Phe Ala Gly Ile Lys Leu Glu

1365 1370

<210> 138

<211> 1276

<212> PRT

<213> Blautia sp.

<400> 138

Met Lys Ile Ser Lys Val Asp His Val Lys Ser Gly Ile Asp Gln Lys

1 5 10 15

Leu Ser Ser Gln Arg Gly Met Leu Tyr Lys Gln Pro Gln Lys Lys Tyr

20 25 30

Glu Gly Lys Gln Leu Glu Glu His Val Arg Asn Leu Ser Arg Lys Ala

35 40 45

Lys Ala Leu Tyr Gln Val Phe Pro Val Ser Gly Asn Ser Lys Met Glu

50 55 60

Lys Glu Leu Gln Ile Ile Asn Ser Phe Ile Lys Asn Ile Leu Leu Arg

65 70 75 80

Leu Asp Ser Gly Lys Thr Ser Glu Glu Ile Val Gly Tyr Ile Asn Thr

85 90 95

Tyr Ser Val Ala Ser Gln Ile Ser Gly Asp His Ile Gln Glu Leu Val

100 105 110

Asp Gln His Leu Lys Glu Ser Leu Arg Lys Tyr Thr Cys Val Gly Asp

115 120 125

Lys Arg Ile Tyr Val Pro Asp Ile Ile Val Ala Leu Leu Lys Ser Lys

130 135 140

Phe Asn Ser Glu Thr Leu Gln Tyr Asp Asn Ser Glu Leu Lys Ile Leu

145 150 155 160

Ile Asp Phe Ile Arg Glu Asp Tyr Leu Lys Glu Lys Gln Ile Lys Gln

165 170 175

Ile Val His Ser Ile Glu Asn Asn Ser Thr Pro Leu Arg Ile Ala Glu

180 185 190

Ile Asn Gly Gln Lys Arg Leu Ile Pro Ala Asn Val Asp Asn Pro Lys

195 200 205

Lys Ser Tyr Ile Phe Glu Phe Leu Lys Glu Tyr Ala Gln Ser Asp Pro

210 215 220

Lys Gly Gln Glu Ser Leu Leu Gln His Met Arg Tyr Leu Ile Leu Leu

225 230 235 240

Tyr Leu Tyr Gly Pro Asp Lys Ile Thr Asp Asp Tyr Cys Glu Glu Ile

245 250 255

Glu Ala Trp Asn Phe Gly Ser Ile Val Met Asp Asn Glu Gln Leu Phe

260 265 270

Ser Glu Glu Ala Ser Met Leu Ile Gln Asp Arg Ile Tyr Val Asn Gln

275 280 285

Gln Ile Glu Glu Gly Arg Gln Ser Lys Asp Thr Ala Lys Val Lys Lys

290 295 300

Asn Lys Ser Lys Tyr Arg Met Leu Gly Asp Lys Ile Glu His Ser Ile

305 310 315 320

Asn Glu Ser Val Val Lys His Tyr Gln Glu Ala Cys Lys Ala Val Glu

325 330 335

Glu Lys Asp Ile Pro Trp Ile Lys Tyr Ile Ser Asp His Val Met Ser

340 345 350

Val Tyr Ser Ser Lys Asn Arg Val Asp Leu Asp Lys Leu Ser Leu Pro

355 360 365

Tyr Leu Ala Lys Asn Thr Trp Asn Thr Trp Ile Ser Phe Ile Ala Met

370 375 380

Lys Tyr Val Asp Met Gly Lys Gly Val Tyr His Phe Ala Met Ser Asp

385 390 395 400

Val Asp Lys Val Gly Lys Gln Asp Asn Leu Ile Ile Gly Gln Ile Asp

405 410 415

Pro Lys Phe Ser Asp Gly Ile Ser Ser Phe Asp Tyr Glu Arg Ile Lys

420 425 430

Ala Glu Asp Asp Leu His Arg Ser Met Ser Gly Tyr Ile Ala Phe Ala

435 440 445

Val Asn Asn Phe Ala Arg Ala Ile Cys Ser Asp Glu Phe Arg Lys Lys

450 455 460

Asn Arg Lys Glu Asp Val Leu Thr Val Gly Leu Asp Glu Ile Pro Leu

465 470 475 480

Tyr Asp Asn Val Lys Arg Lys Leu Leu Gln Tyr Phe Gly Gly Ala Ser

485 490 495

Asn Trp Asp Asp Ser Ile Ile Asp Ile Ile Asp Asp Lys Asp Leu Val

500 505 510

Ala Cys Ile Lys Glu Asn Leu Tyr Val Ala Arg Asn Val Asn Phe His

515 520 525

Phe Ala Gly Ser Glu Lys Val Gln Lys Lys Gln Asp Asp Ile Leu Glu

530 535 540

Glu Ile Val Arg Lys Glu Thr Arg Asp Ile Gly Lys His Tyr Arg Lys

545 550 555 560

Val Phe Tyr Ser Asn Asn Val Ala Val Phe Tyr Cys Asp Glu Asp Ile

565 570 575

Ile Lys Leu Met Asn His Leu Tyr Gln Arg Glu Lys Pro Tyr Gln Ala

580 585 590

Gln Ile Pro Ser Tyr Asn Lys Val Ile Ser Lys Thr Tyr Leu Pro Asp
595 600 605

Leu Ile Phe Met Leu Leu Lys Gly Lys Asn Arg Thr Lys Ile Ser Asp
610 615 620

Pro Ser Ile Met Asn Met Phe Arg Gly Thr Phe Tyr Phe Leu Leu Lys
625 630 635 640

Glu Ile Tyr Tyr Asn Asp Phe Leu Gln Ala Ser Asn Leu Lys Glu Met
645 650 655

Phe Cys Glu Gly Leu Lys Asn Asn Val Lys Asn Lys Lys Ser Glu Lys
660 665 670

Pro Tyr Gln Asn Phe Met Arg Arg Phe Glu Glu Leu Glu Asn Met Gly
675 680 685

Met Asp Phe Gly Glu Ile Cys Gln Gln Ile Met Thr Asp Tyr Glu Gln
690 695 700

Gln Asn Lys Gln Lys Lys Lys Thr Ala Thr Ala Val Met Ser Glu Lys
705 710 715 720

Asp Lys Lys Ile Arg Thr Leu Asp Asn Asp Thr Gln Lys Tyr Lys His
725 730 735

Phe Arg Thr Leu Leu Tyr Ile Gly Leu Arg Glu Ala Phe Ile Ile Tyr
740 745 750

Leu Lys Asp Glu Lys Asn Lys Glu Trp Tyr Glu Phe Leu Arg Glu Pro
755 760 765

Val Lys Arg Glu Gln Pro Glu Glu Lys Glu Phe Val Asn Lys Trp Lys
770 775 780

Leu Asn Gln Tyr Ser Asp Cys Ser Glu Leu Ile Leu Lys Asp Ser Leu
785 790 795 800

Ala Ala Ala Trp Tyr Val Val Ala His Phe Ile Asn Gln Ala Gln Leu
805 810 815

Asn His Leu Ile Gly Asp Ile Lys Asn Tyr Ile Gln Phe Ile Ser Asp
820 825 830

Ile Asp Arg Arg Ala Lys Ser Thr Gly Asn Pro Val Ser Glu Ser Thr
835 840 845

Glu Ile Gln Ile Glu Arg Tyr Arg Lys Ile Leu Arg Val Leu Glu Phe
850 855 860
Ala Lys Phe Phe Cys Gly Gln Ile Thr Asn Val Leu Thr Asp Tyr Tyr

865 870 875 880
Gln Asp Glu Asn Asp Phe Ser Thr His Val Gly His Tyr Val Lys Phe
885 890 895
Glu Lys Lys Asn Met Glu Pro Ala His Ala Leu Gln Ala Phe Ser Asn
900 905 910
Ser Leu Tyr Ala Cys Gly Lys Glu Lys Lys Lys Ala Gly Phe Tyr Tyr
915 920 925
Asp Gly Met Asn Pro Ile Val Asn Arg Asn Ile Thr Leu Ala Ser Met
930 935 940

Tyr Gly Asn Lys Lys Leu Leu Glu Asn Ala Met Asn Pro Val Thr Glu
945 950 955 960
Gln Asp Ile Arg Lys Tyr Tyr Ser Leu Met Ala Glu Leu Asp Ser Val
965 970 975
Leu Lys Asn Gly Ala Val Cys Lys Ser Glu Asp Glu Gln Lys Asn Leu
980 985 990
Arg His Phe Gln Asn Leu Lys Asn Arg Ile Glu Leu Val Asp Val Leu
995 1000 1005
Thr Leu Ser Glu Leu Val Asn Asp Leu Val Ala Gln Leu Ile Gly Trp

1010 1015 1020
Val Tyr Ile Arg Glu Arg Asp Met Met Tyr Leu Gln Leu Gly Leu His
1025 1030 1035 1040
Tyr Ile Lys Leu Tyr Phe Thr Asp Ser Val Ala Glu Asp Ser Tyr Leu
1045 1050 1055
Arg Thr Leu Asp Leu Glu Glu Gly Ser Ile Ala Asp Gly Ala Val Leu
1060 1065 1070
Tyr Gln Ile Ala Ser Leu Tyr Ser Phe Asn Leu Pro Met Tyr Val Lys
1075 1080 1085

Pro Asn Lys Ser Ser Val Tyr Cys Lys Lys His Val Asn Ser Val Ala

1090 1095 1100

Thr Lys Phe Asp Ile Phe Glu Lys Glu Tyr Cys Asn Gly Asp Glu Thr

1105 1110 1115 1120

Val Ile Glu Asn Gly Leu Arg Leu Phe Glu Asn Ile Asn Leu His Lys

1125 1130 1135

Asp Met Val Lys Phe Arg Asp Tyr Leu Ala His Phe Lys Tyr Phe Ala

1140 1145 1150

Lys Leu Asp Glu Ser Ile Leu Glu Leu Tyr Ser Lys Ala Tyr Asp Phe

1155 1160 1165

Phe Phe Ser Tyr Asn Ile Lys Leu Lys Lys Ser Val Ser Tyr Val Leu

1170 1175 1180

Thr Asn Val Leu Leu Ser Tyr Phe Ile Asn Ala Lys Leu Ser Phe Ser

1185 1190 1195 1200

Thr Tyr Lys Ser Ser Gly Asn Lys Thr Val Gln His Arg Thr Thr Lys

1205 1210 1215

Ile Ser Val Val Ala Gln Thr Asp Tyr Phe Thr Tyr Lys Leu Arg Ser

1220 1225 1230

Ile Val Lys Asn Lys Asn Gly Val Glu Ser Ile Glu Asn Asp Asp Arg

1235 1240 1245

Arg Cys Glu Val Val Asn Ile Ala Ala Arg Asp Lys Glu Phe Val Asp

1250 1255 1260

Glu Val Cys Asn Val Ile Asn Tyr Asn Ser Asp Lys

1265 1270 1275

<210> 139

<211> 1385

<212> PRT

<213> Leptotrichia sp.

<400> 139

Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys

1 5 10 15

Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp

20 25 30

Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys
35 40 45
Ile Asp Asn Asn Lys Phe Ile Gly Glu Phe Val Asn Tyr Lys Lys Asn
50 55 60
Asn Asn Val Leu Lys Glu Phe Lys Arg Lys Phe His Ala Gly Asn Ile
65 70 75 80
Leu Phe Lys Leu Lys Gly Lys Glu Glu Ile Ile Arg Ile Glu Asn Asn

85 90 95
Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Val Tyr
100 105 110
Gly Lys Ser Glu Lys Leu Lys Ala Leu Glu Ile Thr Lys Lys Lys Ile
115 120 125
Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile
130 135 140
Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg
145 150 155 160

Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg
165 170 175
Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile
180 185 190
Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile
195 200 205
Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His
210 215 220
Leu Arg Glu Lys Leu Leu Lys Asp Asn Lys Ile Asp Val Ile Leu Thr

225 230 235 240
Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Met
245 250 255
Gly Phe Val Lys Phe Tyr Leu Asn Val Ser Gly Asp Lys Lys Lys Ser
260 265 270
Glu Asn Lys Lys Met Phe Val Glu Lys Ile Leu Asn Thr Asn Val Asp
275 280 285

Leu Thr Val Glu Asp Ile Val Asp Phe Ile Val Lys Glu Leu Lys Phe
290 295 300

Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Phe Asn Asn Glu
305 310 315 320
Phe Leu Glu Asn Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
325 330 335
Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
340 345 350
Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
355 360 365
Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asn Glu Leu Ile

370 375 380
Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
385 390 395 400
Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys
405 410 415
Phe Ser Asn Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
420 425 430
Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys
435 440 445

Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
450 455 460
Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
465 470 475 480
Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Val
485 490 495
Lys Met Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
500 505 510
Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu

515 520 525
Leu Asn Lys Ile Phe Asn Gly Lys Glu Lys Val Thr Asp Phe Phe Gly

530 535 540

Phe Asn Leu Asn Gly Gln Lys Ile Thr Leu Lys Glu Lys Val Pro Ser

545 550 555 560

Phe Lys Leu Asn Ile Leu Lys Lys Leu Asn Phe Ile Asn Asn Glu Asn

565 570 575

Asn Ile Asp Glu Lys Leu Ser His Phe Tyr Ser Phe Gln Lys Glu Gly

580 585 590

Tyr Leu Leu Arg Asn Lys Ile Leu His Asn Ser Tyr Gly Asn Ile Gln

595 600 605

Glu Thr Lys Asn Leu Lys Gly Glu Tyr Glu Asn Val Glu Lys Leu Ile

610 615 620

Lys Glu Leu Lys Val Ser Asp Glu Glu Ile Ser Lys Ser Leu Ser Leu

625 630 635 640

Asp Val Ile Phe Glu Gly Lys Val Asp Ile Ile Asn Lys Ile Asn Ser

645 650 655

Leu Lys Ile Gly Glu Tyr Lys Asp Lys Lys Tyr Leu Pro Ser Phe Ser

660 665 670

Lys Ile Val Leu Glu Ile Thr Arg Lys Phe Arg Glu Ile Asn Lys Asp

675 680 685

Lys Leu Phe Asp Ile Glu Ser Glu Lys Ile Ile Leu Asn Ala Val Lys

690 695 700

Tyr Val Asn Lys Ile Leu Tyr Glu Lys Ile Thr Ser Asn Glu Glu Asn

705 710 715 720

Glu Phe Leu Lys Thr Leu Pro Asp Lys Leu Val Lys Lys Ser Asn Asn

725 730 735

Lys Lys Glu Asn Lys Asn Leu Leu Ser Ile Glu Glu Tyr Tyr Lys Asn

740 745 750

Ala Gln Val Ser Ser Ser Lys Gly Asp Lys Lys Ala Ile Lys Lys Tyr

755 760 765

Gln Asn Lys Val Thr Asn Ala Tyr Leu Glu Tyr Leu Glu Asn Thr Phe

770 775 780

Thr Glu Ile Ile Asp Phe Ser Lys Phe Asn Leu Asn Tyr Asp Glu Ile

785 790 795 800

Lys Thr Lys Ile Glu Glu Arg Lys Asp Asn Lys Ser Lys Ile Ile Ile

805 810 815

Asp Ser Ile Ser Thr Asn Ile Asn Ile Thr Asn Asp Ile Glu Tyr Ile

820 825 830

Ile Ser Ile Phe Ala Leu Leu Asn Ser Asn Thr Tyr Ile Asn Lys Ile

835 840 845

Arg Asn Arg Phe Phe Ala Thr Ser Val Trp Leu Glu Lys Gln Asn Gly

850 855 860

Thr Lys Glu Tyr Asp Tyr Glu Asn Ile Ile Ser Ile Leu Asp Glu Val

865 870 875 880

Leu Leu Ile Asn Leu Leu Arg Glu Asn Asn Ile Thr Asp Ile Leu Asp

885 890 895

Leu Lys Asn Ala Ile Ile Asp Ala Lys Ile Val Glu Asn Asp Glu Thr

900 905 910

Tyr Ile Lys Asn Tyr Ile Phe Glu Ser Asn Glu Glu Lys Leu Lys Lys

915 920 925

Arg Leu Phe Cys Glu Glu Leu Val Asp Lys Glu Asp Ile Arg Lys Ile

930 935 940

Phe Glu Asp Glu Asn Phe Lys Phe Lys Ser Phe Ile Lys Lys Asn Glu

945 950 955 960

Ile Gly Asn Phe Lys Ile Asn Phe Gly Ile Leu Ser Asn Leu Glu Cys

965 970 975

Asn Ser Glu Val Glu Ala Lys Lys Ile Ile Gly Lys Asn Ser Lys Lys

980 985 990

Leu Glu Ser Phe Ile Gln Asn Ile Ile Asp Glu Tyr Lys Ser Asn Ile

995 1000 1005

Arg Thr Leu Phe Ser Ser Glu Phe Leu Glu Lys Tyr Lys Glu Glu Ile

1010 1015 1020

Asp Asn Leu Val Glu Asp Thr Glu Ser Glu Asn Lys Asn Lys Phe Glu

1025 1030 1035 1040

Lys Ile Tyr Tyr Pro Lys Glu His Lys Asn Glu Leu Tyr Ile Tyr Lys
1045 1050 1055
Lys Asn Leu Phe Leu Asn Ile Gly Asn Pro Asn Phe Asp Lys Ile Tyr
1060 1065 1070
Gly Leu Ile Ser Lys Asp Ile Lys Asn Val Asp Thr Lys Ile Leu Phe
1075 1080 1085
Asp Asp Asp Ile Lys Lys Asn Lys Ile Ser Glu Ile Asp Ala Ile Leu

1090 1095 1100
Lys Asn Leu Asn Asp Lys Leu Asn Gly Tyr Ser Asn Asp Tyr Lys Ala
1105 1110 1115 1120
Lys Tyr Val Asn Lys Leu Lys Glu Asn Asp Asp Phe Phe Ala Lys Asn
1125 1130 1135
Ile Gln Asn Glu Asn Tyr Ser Ser Phe Gly Glu Phe Glu Lys Asp Tyr
1140 1145 1150
Asn Lys Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu Val Glu Phe Asn
1155 1160 1165

Tyr Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu
1170 1175 1180
Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val Asn
1185 1190 1195 1200
Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn Thr Gly
1205 1210 1215
Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly Phe Tyr Thr
1220 1225 1230
Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser Tyr Lys Lys Phe

1235 1240 1245
Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu Ser Glu Asn Ser Glu
1250 1255 1260
Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg Asn Tyr Ile Ser His Phe
1265 1270 1275 1280
Tyr Ile Val Arg Asn Pro Phe Ala Asp Tyr Ser Ile Ala Glu Gln Ile
1285 1290 1295

Asp Arg Val Ser Asn Leu Leu Ser Tyr Ser Thr Arg Tyr Asn Asn Ser
1300 1305 1310

Thr Tyr Ala Ser Val Phe Glu Val Phe Lys Lys Asp Val Asn Leu Asp
1315 1320 1325

Tyr Asp Glu Leu Lys Lys Lys Phe Arg Leu Ile Gly Asn Asn Asp Ile
1330 1335 1340

Leu Glu Arg Leu Met Lys Pro Lys Lys Val Ser Val Leu Glu Leu Glu
1345 1350 1355 1360

Ser Tyr Asn Ser Asp Tyr Ile Lys Asn Leu Ile Ile Glu Leu Leu Thr
1365 1370 1375

Lys Ile Glu Asn Thr Asn Asp Thr Leu

1380 1385

<210> 140

<211> 1334

<212> PRT

<213> Lachnospiraceae bacterium

<400> 140

Met Lys Ile Ser Lys Val Asp His Thr Arg Met Ala Val Ala Lys Gly
1 5 10 15

Asn Gln His Arg Arg Asp Glu Ile Ser Gly Ile Leu Tyr Lys Asp Pro
20 25 30

Thr Lys Thr Gly Ser Ile Asp Phe Asp Glu Arg Phe Lys Lys Leu Asn
35 40 45

Cys Ser Ala Lys Ile Leu Tyr His Val Phe Asn Gly Ile Ala Glu Gly

50 55 60

Ser Asn Lys Tyr Lys Asn Ile Val Asp Lys Val Asn Asn Asn Leu Asp
65 70 75 80

Arg Val Leu Phe Thr Gly Lys Ser Tyr Asp Arg Lys Ser Ile Ile Asp
85 90 95

Ile Asp Thr Val Leu Arg Asn Val Glu Lys Ile Asn Ala Phe Asp Arg
100 105 110

Ile Ser Thr Glu Glu Arg Glu Gln Ile Ile Asp Asp Leu Leu Glu Ile

115 120 125

Gln Leu Arg Lys Gly Leu Arg Lys Gly Lys Ala Gly Leu Arg Glu Val
130 135 140
Leu Leu Ile Gly Ala Gly Val Ile Val Arg Thr Asp Lys Lys Gln Glu
145 150 155 160
Ile Ala Asp Phe Leu Glu Ile Leu Asp Glu Asp Phe Asn Lys Thr Asn
165 170 175
Gln Ala Lys Asn Ile Lys Leu Ser Ile Glu Asn Gln Gly Leu Val Val
180 185 190
Ser Pro Val Ser Arg Gly Glu Glu Arg Ile Phe Asp Val Ser Gly Ala

195 200 205
Gln Lys Gly Lys Ser Ser Lys Lys Ala Gln Glu Lys Glu Ala Leu Ser
210 215 220
Ala Phe Leu Leu Asp Tyr Ala Asp Leu Asp Lys Asn Val Arg Phe Glu
225 230 235 240
Tyr Leu Arg Lys Ile Arg Arg Leu Ile Asn Leu Tyr Phe Tyr Val Lys
245 250 255
Asn Asp Asp Val Met Ser Leu Thr Glu Ile Pro Ala Glu Val Asn Leu
260 265 270

Glu Lys Asp Phe Asp Ile Trp Arg Asp His Glu Gln Arg Lys Glu Glu
275 280 285
Asn Gly Asp Phe Val Gly Cys Pro Asp Ile Leu Leu Ala Asp Arg Asp
290 295 300
Val Lys Lys Ser Asn Ser Lys Gln Val Lys Ile Ala Glu Arg Gln Leu
305 310 315 320
Arg Glu Ser Ile Arg Glu Lys Asn Ile Lys Arg Tyr Arg Phe Ser Ile
325 330 335
Lys Thr Ile Glu Lys Asp Asp Gly Thr Tyr Phe Phe Ala Asn Lys Gln

340 345 350
Ile Ser Val Phe Trp Ile His Arg Ile Glu Asn Ala Val Glu Arg Ile
355 360 365

Leu Gly Ser Ile Asn Asp Lys Lys Leu Tyr Arg Leu Arg Leu Gly Tyr
370 375 380
Leu Gly Glu Lys Val Trp Lys Asp Ile Leu Asn Phe Leu Ser Ile Lys
385 390 395 400
Tyr Ile Ala Val Gly Lys Ala Val Phe Asn Phe Ala Met Asp Asp Leu
405 410 415

Gln Glu Lys Asp Arg Asp Ile Glu Pro Gly Lys Ile Ser Glu Asn Ala
420 425 430
Val Asn Gly Leu Thr Ser Phe Asp Tyr Glu Gln Ile Lys Ala Asp Glu
435 440 445
Met Leu Gln Arg Glu Val Ala Val Asn Val Ala Phe Ala Ala Asn Asn
450 455 460
Leu Ala Arg Val Thr Val Asp Ile Pro Gln Asn Gly Glu Lys Glu Asp
465 470 475 480
Ile Leu Leu Trp Asn Lys Ser Asp Ile Lys Lys Tyr Lys Lys Asn Ser
485 490 495
Lys Lys Gly Ile Leu Lys Ser Ile Leu Gln Phe Phe Gly Gly Ala Ser
500 505 510
Thr Trp Asn Met Lys Met Phe Glu Ile Ala Tyr His Asp Gln Pro Gly
515 520 525
Asp Tyr Glu Glu Asn Tyr Leu Tyr Asp Ile Ile Gln Ile Ile Tyr Ser
530 535 540
Leu Arg Asn Lys Ser Phe His Phe Lys Thr Tyr Asp His Gly Asp Lys
545 550 555 560

Asn Trp Asn Arg Glu Leu Ile Gly Lys Met Ile Glu His Asp Ala Glu
565 570 575
Arg Val Ile Ser Val Glu Arg Glu Lys Phe His Ser Asn Asn Leu Pro
580 585 590
Met Phe Tyr Lys Asp Ala Asp Leu Lys Lys Ile Leu Asp Leu Leu Tyr
595 600 605
Ser Asp Tyr Ala Gly Arg Ala Ser Gln Val Pro Ala Phe Asn Thr Val
610 615 620

Leu Val Arg Lys Asn Phe Pro Glu Phe Leu Arg Lys Asp Met Gly Tyr

625 630 635 640

Lys Val His Phe Asn Asn Pro Glu Val Glu Asn Gln Trp His Ser Ala

645 650 655

Val Tyr Tyr Leu Tyr Lys Glu Ile Tyr Tyr Asn Leu Phe Leu Arg Asp

660 665 670

Lys Glu Val Lys Asn Leu Phe Tyr Thr Ser Leu Lys Asn Ile Arg Ser

675 680 685

Glu Val Ser Asp Lys Lys Gln Lys Leu Ala Ser Asp Asp Phe Ala Ser

690 695 700

Arg Cys Glu Glu Ile Glu Asp Arg Ser Leu Pro Glu Ile Cys Gln Ile

705 710 715 720

Ile Met Thr Glu Tyr Asn Ala Gln Asn Phe Gly Asn Arg Lys Val Lys

725 730 735

Ser Gln Arg Val Ile Glu Lys Asn Lys Asp Ile Phe Arg His Tyr Lys

740 745 750

Met Leu Leu Ile Lys Thr Leu Ala Gly Ala Phe Ser Leu Tyr Leu Lys

755 760 765

Gln Glu Arg Phe Ala Phe Ile Gly Lys Ala Thr Pro Ile Pro Tyr Glu

770 775 780

Thr Thr Asp Val Lys Asn Phe Leu Pro Glu Trp Lys Ser Gly Met Tyr

785 790 795 800

Ala Ser Phe Val Glu Glu Ile Lys Asn Asn Leu Asp Leu Gln Glu Trp

805 810 815

Tyr Ile Val Gly Arg Phe Leu Asn Gly Arg Met Leu Asn Gln Leu Ala

820 825 830

Gly Ser Leu Arg Ser Tyr Ile Gln Tyr Ala Glu Asp Ile Glu Arg Arg

835 840 845

Ala Ala Glu Asn Arg Asn Lys Leu Phe Ser Lys Pro Asp Glu Lys Ile

850 855 860

Glu Ala Cys Lys Lys Ala Val Arg Val Leu Asp Leu Cys Ile Lys Ile

865 870 875 880

Ser Thr Arg Ile Ser Ala Glu Phe Thr Asp Tyr Phe Asp Ser Glu Asp

885 890 895

Asp Tyr Ala Asp Tyr Leu Glu Lys Tyr Leu Lys Tyr Gln Asp Asp Ala

900 905 910

Ile Lys Glu Leu Ser Gly Ser Ser Tyr Ala Ala Leu Asp His Phe Cys

915 920 925

Asn Lys Asp Asp Leu Lys Phe Asp Ile Tyr Val Asn Ala Gly Gln Lys

930 935 940

Pro Ile Leu Gln Arg Asn Ile Val Met Ala Lys Leu Phe Gly Pro Asp

945 950 955 960

Asn Ile Leu Ser Glu Val Met Glu Lys Val Thr Glu Ser Ala Ile Arg

965 970 975

Glu Tyr Tyr Asp Tyr Leu Lys Lys Val Ser Gly Tyr Arg Val Arg Gly

980 985 990

Lys Cys Ser Thr Glu Lys Glu Gln Glu Asp Leu Leu Lys Phe Gln Arg

995 1000 1005

Leu Lys Asn Ala Val Glu Phe Arg Asp Val Thr Glu Tyr Ala Glu Val

1010 1015 1020

Ile Asn Glu Leu Leu Gly Gln Leu Ile Ser Trp Ser Tyr Leu Arg Glu

1025 1030 1035 1040

Arg Asp Leu Leu Tyr Phe Gln Leu Gly Phe His Tyr Met Cys Leu Lys

1045 1050 1055

Asn Lys Ser Phe Lys Pro Ala Glu Tyr Val Asp Ile Arg Arg Asn Asn

1060 1065 1070

Gly Thr Ile Ile His Asn Ala Ile Leu Tyr Gln Ile Val Ser Met Tyr

1075 1080 1085

Ile Asn Gly Leu Asp Phe Tyr Ser Cys Asp Lys Glu Gly Lys Thr Leu

1090 1095 1100

Lys Pro Ile Glu Thr Gly Lys Gly Val Gly Ser Lys Ile Gly Gln Phe

1105 1110 1115 1120

Ile Lys Tyr Ser Gln Tyr Leu Tyr Asn Asp Pro Ser Tyr Lys Leu Glu

1125 1130 1135

Ile Tyr Asn Ala Gly Leu Glu Val Phe Glu Asn Ile Asp Glu His Asp
1140 1145 1150
Asn Ile Thr Asp Leu Arg Lys Tyr Val Asp His Phe Lys Tyr Tyr Ala
1155 1160 1165
Tyr Gly Asn Lys Met Ser Leu Leu Asp Leu Tyr Ser Glu Phe Phe Asp
1170 1175 1180
Arg Phe Phe Thr Tyr Asp Met Lys Tyr Gln Lys Asn Val Val Asn Val
1185 1190 1195 1200
Leu Glu Asn Ile Leu Leu Arg His Phe Val Ile Phe Tyr Pro Lys Phe

1205 1210 1215
Gly Ser Gly Lys Lys Asp Val Gly Ile Arg Asp Cys Lys Lys Glu Arg
1220 1225 1230
Ala Gln Ile Glu Ile Ser Glu Gln Ser Leu Thr Ser Glu Asp Phe Met
1235 1240 1245
Phe Lys Leu Asp Asp Lys Ala Gly Glu Glu Ala Lys Lys Phe Pro Ala
1250 1255 1260
Arg Asp Glu Arg Tyr Leu Gln Thr Ile Ala Lys Leu Leu Tyr Tyr Pro
1265 1270 1275 1280

Asn Glu Ile Glu Asp Met Asn Arg Phe Met Lys Lys Gly Glu Thr Ile
1285 1290 1295
Asn Lys Lys Val Gln Phe Asn Arg Lys Lys Lys Ile Thr Arg Lys Gln
1300 1305 1310
Lys Asn Asn Ser Ser Asn Glu Val Leu Ser Ser Thr Met Gly Tyr Leu
1315 1320 1325
Phe Lys Asn Ile Lys Leu
1330

<210> 141
<211> 500
<212> PRT
<213> Chloroflexus aggregans

<400> 141

Met Thr Asp Gln Val Arg Arg Glu Glu Val Ala Ala Gly Glu Leu Ala
1 5 10 15
Asp Thr Pro Leu Ala Ala Ala Gln Thr Pro Ala Ala Asp Ala Ala Val
20 25 30
Ala Ala Thr Pro Ala Pro Ala Glu Ala Val Ala Pro Thr Pro Glu Gln
35 40 45
Ala Val Asp Gln Pro Ala Thr Thr Gly Glu Ser Glu Ala Pro Val Thr
50 55 60
Thr Ala Gln Ala Ala Ala His Glu Ala Glu Pro Ala Glu Ala Thr Gly

65 70 75 80
Ala Ser Phe Thr Pro Val Ser Glu Gln Gln Pro Gln Lys Pro Arg Arg
85 90 95
Leu Lys Asp Leu Gln Pro Gly Met Glu Leu Glu Gly Lys Val Thr Ser
100 105 110
Ile Ala Leu Tyr Gly Ile Phe Val Asp Val Gly Val Gly Arg Asp Gly
115 120 125
Leu Val His Ile Ser Glu Met Ser Asp Arg Arg Ile Asp Thr Pro Ser
130 135 140

Glu Leu Val Gln Ile Gly Asp Thr Val Lys Val Trp Val Lys Ser Val
145 150 155 160
Asp Leu Asp Ala Arg Arg Ile Ser Leu Thr Met Leu Asn Pro Ser Arg
165 170 175
Gly Glu Lys Pro Arg Arg Ser Arg Gln Ser Gln Pro Ala Gln Pro Gln
180 185 190
Pro Arg Arg Gln Glu Val Asp Arg Glu Lys Leu Ala Ser Leu Lys Val
195 200 205
Gly Glu Ile Val Glu Gly Val Ile Thr Gly Phe Ala Pro Phe Gly Ala

210 215 220
Phe Ala Asp Ile Gly Val Gly Lys Asp Gly Leu Ile His Ile Ser Glu
225 230 235 240

Leu Ser Glu Gly Arg Val Glu Lys Pro Glu Asp Ala Val Lys Val Gly
245 250 255
Glu Arg Tyr Gln Phe Lys Val Leu Glu Ile Asp Gly Glu Gly Thr Arg
260 265 270
Ile Ser Leu Ser Leu Arg Arg Ala Gln Arg Thr Gln Arg Met Gln Gln
275 280 285

Leu Glu Pro Gly Gln Ile Ile Glu Gly Thr Val Ser Gly Ile Ala Thr
290 295 300
Phe Gly Ala Phe Val Asp Ile Gly Val Gly Arg Asp Gly Leu Val His
305 310 315 320
Ile Ser Ala Leu Ala Pro His Arg Val Ala Lys Val Glu Asp Val Val
325 330 335
Lys Val Gly Asp Lys Val Lys Val Lys Val Leu Gly Val Asp Pro Gln
340 345 350
Ser Lys Arg Ile Ser Leu Thr Met Arg Leu Glu Glu Glu Gln Pro Ala

355 360 365
Thr Thr Ala Gly Asp Glu Ala Ala Glu Pro Ala Glu Glu Val Thr Pro
370 375 380
Thr Arg Arg Gly Asn Leu Glu Arg Phe Ala Ala Ala Ala Gln Thr Ala
385 390 395 400
Arg Glu Arg Ser Glu Arg Gly Glu Arg Ser Glu Arg Gly Glu Arg Arg
405 410 415
Glu Arg Arg Glu Arg Arg Pro Ala Gln Ser Ser Pro Asp Thr Tyr Ile
420 425 430

Val Gly Glu Asp Asp Asp Glu Ser Phe Glu Gly Asn Ala Thr Ile Glu
435 440 445
Asp Leu Leu Thr Lys Phe Gly Gly Ser Ser Ser Arg Arg Asp Arg Asp
450 455 460
Arg Arg Arg Arg His Glu Asp Asp Asp Asp Glu Glu Met Glu Arg Pro
465 470 475 480
Ser Asn Arg Arg Gln Arg Glu Ala Ile Arg Arg Thr Leu Gln Gln Ile
485 490 495

Gly Tyr Asp Glu

500

<210> 142

<211> 249

<212> PRT

<213> Demequina aurantiaca

<400> 142

Met Asp Leu Thr Trp His Ala Leu Leu Ile Leu Phe Ile Val Ala Leu
1 5 10 15

Leu Ala Gly Phe Leu Asp Thr Leu Ala Gly Gly Gly Gly Leu Leu Thr
20 25 30

Val Pro Ala Leu Leu Leu Thr Gly Ile Pro Pro Leu Gln Ala Leu Gly
35 40 45

Thr Asn Lys Leu Gln Ser Ser Phe Gly Thr Gly Met Ala Thr Tyr Gln

50 55 60

Val Ile Arg Lys Lys Arg Val His Trp Arg Asp Val Arg Trp Pro Met
65 70 75 80

Val Trp Ala Phe Leu Gly Ser Ala Ala Gly Ala Val Ala Val Gln Phe
85 90 95

Ile Asp Thr Asp Ala Leu Leu Ile Ile Ile Pro Val Val Leu Ala Leu
100 105 110

Val Ala Ala Tyr Phe Leu Phe Val Pro Lys Ser His Leu Pro Pro Pro
115 120 125

Glu Pro Arg Met Ser Asp Pro Ala Tyr Glu Ala Thr Leu Val Pro Ile

130 135 140

Ile Gly Ala Tyr Asp Gly Ala Phe Gly Pro Gly Thr Gly Ser Leu Tyr
145 150 155 160

Ala Leu Ser Gly Val Ala Leu Arg Ala Lys Thr Leu Val Gln Ser Thr
165 170 175

Ala Ile Ala Lys Thr Leu Asn Phe Ala Thr Asn Phe Ala Ala Leu Leu
180 185 190

Val Phe Ala Phe Ala Gly His Met Leu Trp Thr Val Gly Ala Val Met

195 200 205

Ile Ala Gly Gln Leu Ile Gly Ala Tyr Ala Gly Ser His Met Leu Phe

210 215 220

Arg Val Asn Pro Leu Val Leu Arg Val Leu Ile Val Val Met Ser Leu

225 230 235 240

Gly Met Leu Ile Arg Val Leu Leu Asp

245

<210> 143

<211> 1235

<212> PRT

<213> Thalassospira sp.

<400> 143

Met Arg Ile Ile Lys Pro Tyr Gly Arg Ser His Val Glu Gly Val Ala

1 5 10 15

Thr Gln Glu Pro Arg Arg Lys Leu Arg Leu Asn Ser Ser Pro Asp Ile

20 25 30

Ser Arg Asp Ile Pro Gly Phe Ala Gln Ser His Asp Ala Leu Ile Ile

35 40 45

Ala Gln Trp Ile Ser Ala Ile Asp Lys Ile Ala Thr Lys Pro Lys Pro

50 55 60

Asp Lys Lys Pro Thr Gln Ala Gln Ile Asn Leu Arg Thr Thr Leu Gly

65 70 75 80

Asp Ala Ala Trp Gln His Val Met Ala Glu Asn Leu Leu Pro Ala Ala

85 90 95

Thr Asp Pro Ala Ile Arg Glu Lys Leu His Leu Ile Trp Gln Ser Lys

100 105 110

Ile Ala Pro Trp Gly Thr Ala Arg Pro Gln Ala Glu Lys Asp Gly Lys

115 120 125

Pro Thr Pro Lys Gly Gly Trp Tyr Glu Arg Phe Cys Gly Val Leu Ser

130 135 140

Pro Glu Ala Ile Thr Gln Asn Val Ala Arg Gln Ile Ala Lys Asp Ile

145 150 155 160

Tyr Asp His Leu His Val Ala Ala Lys Arg Lys Gly Arg Glu Pro Ala

165 170 175

Lys Gln Gly Glu Ser Ser Asn Lys Pro Gly Lys Phe Lys Pro Asp Arg

180 185 190

Lys Arg Gly Leu Ile Glu Glu Arg Ala Glu Ser Ile Ala Lys Asn Ala

195 200 205

Leu Arg Pro Gly Ser His Ala Pro Cys Pro Trp Gly Pro Asp Asp Gln

210 215 220

Ala Thr Tyr Glu Gln Ala Gly Asp Val Ala Gly Gln Ile Tyr Ala Ala

225 230 235 240

Ala Arg Asp Cys Leu Glu Glu Lys Lys Arg Arg Ser Gly Asn Arg Asn

245 250 255

Thr Ser Ser Val Gln Tyr Leu Pro Arg Asp Leu Ala Ala Lys Ile Leu

260 265 270

Tyr Ala Gln Tyr Gly Arg Val Phe Gly Pro Asp Thr Thr Ile Lys Ala

275 280 285

Ala Leu Asp Glu Gln Pro Ser Leu Phe Ala Leu His Lys Ala Ile Lys

290 295 300

Asp Cys Tyr His Arg Leu Ile Asn Asp Ala Arg Lys Arg Asp Ile Leu

305 310 315 320

Arg Ile Leu Pro Arg Asn Met Ala Ala Leu Phe Arg Leu Val Arg Ala

325 330 335

Gln Tyr Asp Asn Arg Asp Ile Asn Ala Leu Ile Arg Leu Gly Lys Val

340 345 350

Ile His Tyr His Ala Ser Glu Gln Gly Lys Ser Glu His His Gly Ile

355 360 365

Arg Asp Tyr Trp Pro Ser Gln Gln Asp Ile Gln Asn Ser Arg Phe Trp

370 375 380

Gly Ser Asp Gly Gln Ala Asp Ile Lys Arg His Glu Ala Phe Ser Arg

385 390 395 400

Ile Trp Arg His Ile Ile Ala Leu Ala Ser Arg Thr Leu His Asp Trp

405 410 415

Ala Asp Pro His Ser Gln Lys Phe Ser Gly Glu Asn Asp Asp Ile Leu

420 425 430

Leu Leu Ala Lys Asp Ala Ile Glu Asp Asp Val Phe Lys Ala Gly His

435 440 445

Tyr Glu Arg Lys Cys Asp Val Leu Phe Gly Ala Gln Ala Ser Leu Phe

450 455 460

Cys Gly Ala Glu Asp Phe Glu Lys Ala Ile Leu Lys Gln Ala Ile Thr

465 470 475 480

Gly Thr Gly Asn Leu Arg Asn Ala Thr Phe His Phe Lys Gly Lys Val

485 490 495

Arg Phe Glu Lys Glu Leu Gln Glu Leu Thr Lys Asp Val Pro Val Glu

500 505 510

Val Gln Ser Ala Ile Ala Ala Leu Trp Gln Lys Asp Ala Glu Gly Arg

515 520 525

Thr Arg Gln Ile Ala Glu Thr Leu Gln Ala Val Leu Ala Gly His Phe

530 535 540

Leu Thr Glu Glu Gln Asn Arg His Ile Phe Ala Ala Leu Thr Ala Ala

545 550 555 560

Met Ala Gln Pro Gly Asp Val Pro Leu Pro Arg Leu Arg Arg Val Leu

565 570 575

Ala Arg His Asp Ser Ile Cys Gln Arg Gly Arg Ile Leu Pro Leu Ser

580 585 590

Pro Cys Pro Asp Arg Ala Lys Leu Glu Glu Ser Pro Ala Leu Thr Cys

595 600 605

Gln Tyr Thr Val Leu Lys Met Leu Tyr Asp Gly Pro Phe Arg Ala Trp

610 615 620

Leu Ala Gln Gln Asn Ser Thr Ile Leu Asn His Tyr Ile Asp Ser Thr

625 630 635 640

Ile Ala Arg Thr Asp Lys Ala Ala Arg Asp Met Asn Gly Arg Lys Leu

645 650 655

Ala Gln Ala Glu Lys Asp Leu Ile Thr Ser Arg Ala Ala Asp Leu Pro
660 665 670
Arg Leu Ser Val Asp Glu Lys Met Gly Asp Phe Leu Ala Arg Leu Thr
675 680 685
Ala Ala Thr Ala Thr Glu Met Arg Val Gln Arg Gly Tyr Gln Ser Asp
690 695 700
Gly Glu Asn Ala Gln Lys Gln Ala Ala Phe Ile Gly Gln Phe Glu Cys
705 710 715 720
Asp Val Ile Gly Arg Ala Phe Ala Asp Phe Leu Asn Gln Ser Gly Phe

725 730 735
Asp Phe Val Leu Lys Leu Lys Ala Asp Thr Pro Gln Pro Asp Ala Ala
740 745 750
Gln Cys Asp Val Thr Ala Leu Ile Ala Pro Asp Asp Ile Ser Val Ser
755 760 765
Pro Pro Gln Ala Trp Gln Gln Val Leu Tyr Phe Ile Leu His Leu Val
770 775 780
Pro Val Asp Asp Ala Ser His Leu Leu His Gln Ile Arg Lys Trp Gln
785 790 795 800

Val Leu Glu Gly Lys Glu Lys Pro Ala Gln Ile Ala His Asp Val Gln
805 810 815
Ser Val Leu Met Leu Tyr Leu Asp Met His Asp Ala Lys Phe Thr Gly
820 825 830
Gly Ala Ala Leu His Gly Ile Glu Lys Phe Ala Glu Phe Phe Ala His
835 840 845
Ala Ala Asp Phe Arg Ala Val Phe Pro Pro Gln Ser Leu Gln Asp Gln
850 855 860
Asp Arg Ser Ile Pro Arg Arg Gly Leu Arg Glu Ile Val Arg Phe Gly

865 870 875 880
His Leu Pro Leu Leu Gln His Met Ser Gly Thr Val Gln Ile Thr His
885 890 895

Asp Asn Val Val Ala Trp Gln Ala Ala Arg Thr Ala Gly Ala Thr Gly
900 905 910
Met Ser Pro Ile Ala Arg Arg Gln Lys Gln Arg Glu Glu Leu His Ala
915 920 925
Leu Ala Val Glu Arg Thr Ala Arg Phe Arg Asn Ala Asp Leu Gln Asn
930 935 940

Tyr Met His Ala Leu Val Asp Val Ile Lys His Arg Gln Leu Ser Ala
945 950 955 960
Gln Val Thr Leu Ser Asp Gln Val Arg Leu His Arg Leu Met Met Gly
965 970 975
Val Leu Gly Arg Leu Val Asp Tyr Ala Gly Leu Trp Glu Arg Asp Leu
980 985 990
Tyr Phe Val Val Leu Ala Leu Leu Tyr His His Gly Ala Thr Pro Asp
995 1000 1005
Asp Val Phe Lys Gly Gln Gly Lys Lys Asn Leu Ala Asp Gly Gln Val

1010 1015 1020
Val Ala Ala Leu Lys Pro Lys Asn Arg Lys Ala Ala Ala Pro Val Gly
1025 1030 1035 1040
Val Phe Asp Asp Leu Asp His Tyr Gly Ile Tyr Gln Asp Asp Arg Gln
1045 1050 1055
Ser Ile Arg Asn Gly Leu Ser His Phe Asn Met Leu Arg Gly Gly Lys
1060 1065 1070
Ala Pro Asp Leu Ser His Trp Val Asn Gln Thr Arg Ser Leu Val Ala
1075 1080 1085

His Asp Arg Lys Leu Lys Asn Ala Val Ala Lys Ser Val Ile Glu Met
1090 1095 1100
Leu Ala Arg Glu Gly Phe Asp Leu Asp Trp Gly Ile Gln Thr Asp Arg
1105 1110 1115 1120
Gly Gln His Ile Leu Ser His Gly Lys Ile Arg Thr Arg Gln Ala Gln
1125 1130 1135
His Phe Gln Lys Ser Arg Leu His Ile Val Lys Lys Ser Ala Lys Pro
1140 1145 1150

Asp Lys Asn Asp Thr Val Lys Ile Arg Glu Asn Leu His Gly Asp Ala

1155 1160 1165

Met Val Glu Arg Val Val Gln Leu Phe Ala Ala Gln Val Gln Lys Arg

1170 1175 1180

Tyr Asp Ile Thr Val Glu Lys Arg Leu Asp His Leu Phe Leu Lys Pro

1185 1190 1195 1200

Gln Asp Gln Lys Gly Lys Asn Gly Ile His Thr His Asn Gly Trp Ser

1205 1210 1215

Lys Thr Glu Lys Lys Arg Arg Pro Ser Arg Glu Asn Arg Lys Gly Asn

1220 1225 1230

His Glu Asn

1235

<210> 144

<211> 1312

<212> PRT

<213> Pseudobutyrivibrio sp.

<400> 144

Met Lys Phe Ser Lys Glu Ser His Arg Lys Thr Ala Val Gly Val Thr

1 5 10 15

Glu Ser Asn Gly Ile Ile Gly Leu Leu Tyr Lys Asp Pro Leu Asn Glu

20 25 30

Lys Glu Lys Ile Glu Asp Val Val Asn Gln Arg Ala Asn Ser Thr Lys

35 40 45

Arg Leu Phe Asn Leu Phe Gly Thr Glu Ala Thr Ser Lys Asp Ile Ser

50 55 60

Arg Ala Ser Lys Asp Leu Ala Lys Val Val Asn Lys Ala Ile Gly Asn

65 70 75 80

Leu Lys Gly Asn Lys Lys Phe Asn Lys Lys Glu Gln Ile Thr Lys Gly

85 90 95

Leu Asn Thr Lys Ile Ile Val Glu Glu Leu Lys Asn Val Leu Lys Asp

100 105 110

Glu Lys Lys Leu Ile Val Asn Lys Asp Ile Ile Asp Glu Ala Cys Ser

115 120 125

Arg Leu Leu Lys Thr Ser Phe Arg Thr Ala Lys Thr Lys Gln Ala Val
130 135 140
Lys Met Ile Leu Thr Ala Val Leu Ile Glu Asn Thr Asn Leu Ser Lys
145 150 155 160
Glu Asp Glu Ala Phe Val His Glu Tyr Phe Val Lys Lys Leu Val Asn
165 170 175
Glu Tyr Asn Lys Thr Ser Val Lys Lys Gln Ile Pro Val Ala Leu Ser
180 185 190
Asn Gln Asn Met Val Ile Gln Pro Asn Ser Val Asn Gly Thr Leu Glu

195 200 205
Ile Ser Glu Thr Lys Lys Ser Lys Glu Thr Lys Thr Thr Glu Lys Asp
210 215 220
Ala Phe Arg Ala Phe Leu Arg Asp Tyr Ala Thr Leu Asp Glu Asn Arg
225 230 235 240
Arg His Lys Met Arg Leu Cys Leu Arg Asn Leu Val Asn Leu Tyr Phe
245 250 255
Tyr Gly Glu Thr Ser Val Ser Lys Asp Asp Phe Asp Glu Trp Arg Asp
260 265 270

His Glu Asp Lys Lys Gln Asn Asp Glu Leu Phe Val Lys Lys Ile Val
275 280 285
Ser Ile Lys Thr Asp Arg Lys Gly Asn Val Lys Glu Val Leu Asp Val
290 295 300
Asp Ala Thr Ile Asp Ala Ile Arg Thr Asn Asn Ile Ala Cys Tyr Arg
305 310 315 320
Arg Ala Leu Ala Tyr Ala Asn Glu Asn Pro Asp Val Phe Phe Ser Asp
325 330 335
Thr Met Leu Asn Lys Phe Trp Ile His His Val Glu Asn Glu Val Glu

340 345 350
Arg Ile Tyr Gly His Ile Asn Asn Asn Thr Gly Asp Tyr Lys Tyr Gln
355 360 365

Leu Gly Tyr Leu Ser Glu Lys Val Trp Lys Gly Ile Ile Asn Tyr Leu
370 375 380
Ser Ile Lys Tyr Ile Ala Glu Gly Lys Ala Val Tyr Asn Tyr Ala Met
385 390 395 400
Asn Ala Leu Ala Lys Asp Asn Asn Ser Asn Ala Phe Gly Lys Leu Asp
405 410 415

Glu Lys Phe Val Asn Gly Ile Thr Ser Phe Glu Tyr Glu Arg Ile Lys
420 425 430
Ala Glu Glu Thr Leu Gln Arg Glu Cys Ala Val Asn Ile Ala Phe Ala
435 440 445
Ala Asn His Leu Ala Asn Ala Thr Val Asp Leu Asn Glu Lys Asp Ser
450 455 460
Asp Phe Leu Leu Leu Lys His Glu Asp Asn Lys Asp Thr Leu Gly Ala
465 470 475 480
Val Ala Arg Pro Asn Ile Leu Arg Asn Ile Leu Gln Phe Phe Gly Gly
485 490 495
Lys Ser Arg Trp Asn Asp Phe Asp Phe Ser Gly Ile Asp Glu Ile Gln
500 505 510
Leu Leu Asp Asp Leu Arg Lys Met Ile Tyr Ser Leu Arg Asn Ser Ser
515 520 525
Phe His Phe Lys Thr Glu Asn Ile Asp Asn Asp Ser Trp Asn Thr Lys
530 535 540
Leu Ile Gly Asp Met Phe Ala Tyr Asp Phe Asn Met Ala Gly Asn Val
545 550 555 560

Gln Lys Asp Lys Met Tyr Ser Asn Asn Val Pro Met Phe Tyr Ser Thr
565 570 575
Ser Asp Ile Glu Lys Met Leu Asp Arg Leu Tyr Ala Glu Val His Glu
580 585 590
Arg Ala Ser Gln Val Pro Ser Phe Asn Ser Val Phe Val Arg Lys Asn
595 600 605
Phe Pro Asp Tyr Leu Lys Asn Asp Leu Lys Ile Thr Ser Ala Phe Gly
610 615 620

Val Asp Asp Ala Leu Lys Trp Gln Ser Ala Val Tyr Tyr Val Cys Lys

625 630 635 640

Glu Ile Tyr Tyr Asn Asp Phe Leu Gln Asn Pro Glu Thr Phe Thr Met

645 650 655

Leu Lys Asp Tyr Val Gln Cys Leu Pro Ile Asp Ile Asp Lys Ser Met

660 665 670

Asp Gln Lys Leu Lys Ser Glu Arg Asn Ala His Lys Asn Phe Lys Glu

675 680 685

Ala Phe Ala Thr Tyr Cys Lys Glu Cys Asp Ser Leu Ser Ala Ile Cys

690 695 700

Gln Met Ile Met Thr Glu Tyr Asn Asn Gln Asn Lys Gly Asn Arg Lys

705 710 715 720

Val Ile Ser Ala Arg Thr Lys Asp Gly Asp Lys Leu Ile Tyr Lys His

725 730 735

Tyr Lys Met Ile Leu Phe Glu Ala Leu Lys Asn Val Phe Thr Ile Tyr

740 745 750

Leu Glu Lys Asn Ile Asn Thr Tyr Gly Phe Leu Lys Lys Pro Lys Leu

755 760 765

Ile Asn Asn Val Pro Ala Ile Glu Glu Phe Leu Pro Asn Tyr Asn Gly

770 775 780

Arg Gln Tyr Glu Thr Leu Val Asn Arg Ile Thr Glu Glu Thr Glu Leu

785 790 795 800

Gln Lys Trp Tyr Ile Val Gly Arg Leu Leu Asn Pro Lys Gln Val Asn

805 810 815

Gln Leu Ile Gly Asn Phe Arg Ser Tyr Val Gln Tyr Val Asn Asp Val

820 825 830

Ala Arg Arg Ala Lys Gln Thr Gly Asn Asn Leu Ser Asn Asp Asn Ile

835 840 845

Ala Trp Asp Val Lys Asn Ile Ile Gln Ile Phe Asp Val Cys Thr Lys

850 855 860

Leu Asn Gly Val Thr Ser Asn Ile Leu Glu Asp Tyr Phe Asp Asp Gly

865 870 875 880

Asp Asp Tyr Ala Arg Tyr Leu Lys Asn Phe Val Asp Tyr Thr Asn Lys

885 890 895

Asn Asn Asp His Ser Ala Thr Leu Leu Gly Asp Phe Cys Ala Lys Glu

900 905 910

Ile Asp Gly Ile Lys Ile Gly Ile Tyr His Asp Gly Thr Asn Pro Ile

915 920 925

Val Asn Arg Asn Ile Ile Gln Cys Lys Leu Tyr Gly Ala Thr Gly Ile

930 935 940

Ile Ser Asp Leu Thr Lys Asp Gly Ser Ile Leu Ser Val Asp Tyr Glu

945 950 955 960

Ile Ile Lys Lys Tyr Met Gln Met Gln Lys Glu Ile Lys Val Tyr Gln

965 970 975

Gln Lys Gly Ile Cys Lys Thr Lys Glu Glu Gln Gln Asn Leu Lys Lys

980 985 990

Tyr Gln Glu Leu Lys Asn Ile Val Glu Leu Arg Asn Ile Ile Asp Tyr

995 1000 1005

Ser Glu Ile Leu Asp Glu Leu Gln Gly Gln Leu Ile Asn Trp Gly Tyr

1010 1015 1020

Leu Arg Glu Arg Asp Leu Met Tyr Phe Gln Leu Gly Phe His Tyr Leu

1025 1030 1035 1040

Cys Leu His Asn Glu Ser Lys Lys Pro Val Gly Tyr Asn Asn Ala Gly

1045 1050 1055

Asp Ile Ser Gly Ala Val Leu Tyr Gln Ile Val Ala Met Tyr Thr Asn

1060 1065 1070

Gly Leu Ser Leu Ile Asp Ala Asn Gly Lys Ser Lys Lys Asn Ala Lys

1075 1080 1085

Ala Ser Ala Gly Ala Lys Val Gly Ser Phe Cys Ser Tyr Ser Lys Glu

1090 1095 1100

Ile Arg Gly Val Asp Lys Asp Thr Lys Glu Asp Asp Asp Pro Ile Tyr

1105 1110 1115 1120

Leu Ala Gly Val Glu Leu Phe Glu Asn Ile Asn Glu His Gln Gln Cys

1125 1130 1135

Ile Asn Leu Arg Asn Tyr Ile Glu His Phe His Tyr Tyr Ala Lys His
1140 1145 1150
Asp Arg Ser Met Leu Asp Leu Tyr Ser Glu Val Phe Asp Arg Phe Phe
1155 1160 1165
Thr Tyr Asp Met Lys Tyr Thr Lys Asn Val Pro Asn Met Met Tyr Asn
1170 1175 1180
Ile Leu Leu Gln His Leu Val Val Pro Ala Phe Glu Phe Gly Ser Ser
1185 1190 1195 1200
Glu Lys Arg Leu Asp Asp Asn Asp Glu Gln Thr Lys Pro Arg Ala Met

1205 1210 1215
Phe Thr Leu Arg Glu Lys Asn Gly Leu Ser Ser Glu Gln Phe Thr Tyr
1220 1225 1230
Arg Leu Gly Asp Gly Asn Ser Thr Val Lys Leu Ser Ala Arg Gly Asp
1235 1240 1245
Asp Tyr Leu Arg Ala Val Ala Ser Leu Leu Tyr Tyr Pro Asp Arg Ala
1250 1255 1260
Pro Glu Gly Leu Ile Arg Asp Ala Glu Ala Glu Asp Lys Phe Ala Lys
1265 1270 1275 1280

Ile Asn His Ser Asn Pro Lys Ser Asp Asn Arg Asn Asn Arg Gly Asn
1285 1290 1295
Phe Lys Asn Pro Lys Val Gln Trp Tyr Asn Asn Lys Thr Lys Arg Lys
1300 1305 1310

<210> 145
<211> 1343
<212> PRT
<213> Butyrivibrio sp.
<400> 145

Met Lys Ile Ser Lys Val Asp His Arg Lys Thr Ala Val Lys Ile Thr
1 5 10 15
Asp Asn Lys Gly Ala Glu Gly Phe Ile Tyr Gln Asp Pro Thr Arg Asp

20 25 30

Ser Ser Thr Met Glu Gln Ile Ile Ser Asn Arg Ala Arg Ser Ser Lys

35 40 45

Val Leu Phe Asn Ile Phe Gly Asp Thr Lys Lys Ser Lys Asp Leu Asn

50 55 60

Lys Tyr Thr Glu Ser Leu Ile Ile Tyr Val Asn Lys Ala Ile Lys Ser

65 70 75 80

Leu Lys Gly Asp Lys Arg Asn Asn Lys Tyr Glu Glu Ile Thr Glu Ser

85 90 95

Leu Lys Thr Glu Arg Val Leu Asn Ala Leu Ile Gln Ala Gly Asn Glu

100 105 110

Phe Thr Cys Ser Glu Asn Asn Ile Glu Asp Ala Leu Asn Lys Tyr Leu

115 120 125

Lys Lys Ser Phe Arg Val Gly Asn Thr Lys Ser Ala Leu Lys Lys Leu

130 135 140

Leu Met Ala Ala Tyr Cys Gly Tyr Lys Leu Ser Ile Glu Glu Lys Glu

145 150 155 160

Glu Ile Gln Asn Tyr Phe Val Asp Lys Leu Val Lys Glu Tyr Asn Lys

165 170 175

Asp Thr Val Leu Lys Tyr Thr Ala Lys Ser Leu Lys His Gln Asn Met

180 185 190

Val Val Gln Pro Asp Thr Asp Asn His Val Phe Leu Pro Ser Arg Ile

195 200 205

Ala Gly Ala Thr Gln Asn Lys Met Ser Glu Lys Glu Ala Leu Thr Glu

210 215 220

Phe Leu Lys Ala Tyr Ala Val Leu Asp Glu Glu Lys Arg His Asn Leu

225 230 235 240

Arg Ile Ile Leu Arg Lys Leu Val Asn Leu Tyr Phe Tyr Glu Ser Pro

245 250 255

Asp Phe Ile Tyr Pro Glu Asn Asn Glu Trp Lys Glu His Asp Asp Arg

260 265 270

Lys Asn Lys Thr Glu Thr Phe Val Ser Pro Val Lys Val Asn Glu Glu

275 280 285

Lys Asn Gly Lys Thr Phe Val Lys Ile Asp Val Pro Ala Thr Lys Asp

290 295 300

Leu Ile Arg Leu Lys Asn Ile Glu Cys Tyr Arg Arg Ser Val Ala Glu

305 310 315 320

Thr Ala Gly Asn Pro Ile Thr Tyr Phe Thr Asp His Asn Ile Ser Lys

325 330 335

Phe Trp Ile His His Ile Glu Asn Glu Val Glu Lys Ile Phe Ala Leu

340 345 350

Leu Lys Ser Asn Trp Lys Asp Tyr Gln Phe Ser Val Gly Tyr Ile Ser

355 360 365

Glu Lys Val Trp Lys Glu Ile Ile Asn Tyr Leu Ser Ile Lys Tyr Ile

370 375 380

Ala Ile Gly Lys Ala Val Tyr Asn Tyr Ala Leu Glu Asp Ile Lys Lys

385 390 395 400

Asn Asp Gly Thr Leu Asn Phe Gly Val Ile Asp Pro Ser Phe Tyr Asp

405 410 415

Gly Ile Asn Ser Phe Glu Tyr Glu Lys Ile Lys Ala Glu Glu Thr Phe

420 425 430

Gln Arg Glu Val Ala Val Tyr Val Ser Phe Ala Val Asn His Leu Ser

435 440 445

Ser Ala Thr Val Lys Leu Ser Glu Ala Gln Ser Asp Met Leu Val Leu

450 455 460

Asn Lys Asn Asp Ile Glu Lys Ile Ala Tyr Gly Asn Thr Lys Arg Asn

465 470 475 480

Ile Leu Gln Phe Phe Gly Gly Gln Ser Lys Trp Lys Glu Phe Asp Phe

485 490 495

Asp Arg Tyr Ile Asn Pro Val Asn Tyr Thr Asp Ile Asp Phe Leu Phe

500 505 510

Asp Ile Lys Lys Met Val Tyr Ser Leu Arg Asn Glu Ser Phe His Phe

515 520 525

Thr Thr Thr Asp Thr Glu Ser Asp Trp Asn Lys Asn Leu Ile Ser Ala
530 535 540
Met Phe Glu Tyr Glu Cys Arg Arg Ile Ser Thr Val Gln Lys Asn Lys
545 550 555 560
Phe Phe Ser Asn Asn Leu Pro Leu Phe Tyr Gly Glu Asn Ser Leu Glu
565 570 575
Arg Val Leu His Lys Leu Tyr Asp Asp Tyr Val Asp Arg Met Ser Gln
580 585 590
Val Pro Ser Phe Gly Asn Val Phe Val Arg Lys Lys Phe Pro Asp Tyr
595 600 605
Met Lys Glu Ile Gly Ile Lys His Asn Leu Ser Ser Glu Asp Asn Leu
610 615 620
Lys Leu Gln Gly Ala Leu Tyr Phe Leu Tyr Lys Glu Ile Tyr Tyr Asn
625 630 635 640
Ala Phe Ile Ser Ser Glu Lys Ala Met Lys Ile Phe Val Asp Leu Val
645 650 655
Asn Lys Leu Asp Thr Asn Ala Arg Asp Asp Lys Gly Arg Ile Thr His
660 665 670
Glu Ala Met Ala His Lys Asn Phe Lys Asp Ala Ile Ser His Tyr Met
675 680 685
Thr His Asp Cys Ser Leu Ala Asp Ile Cys Gln Lys Ile Met Thr Glu
690 695 700
Tyr Asn Gln Gln Asn Thr Gly His Arg Lys Lys Gln Thr Thr Tyr Ser
705 710 715 720
Ser Glu Lys Asn Pro Glu Ile Phe Arg His Tyr Lys Met Ile Leu Phe
725 730 735
Met Leu Leu Gln Lys Ala Met Thr Glu Tyr Ile Ser Ser Glu Glu Ile
740 745 750
Phe Asp Phe Ile Met Lys Pro Asn Ser Pro Lys Thr Asp Ile Lys Glu
755 760 765
Glu Glu Phe Leu Pro Gln Tyr Lys Ser Cys Ala Tyr Asp Asn Leu Ile
770 775 780

Lys Leu Ile Ala Asp Asn Val Glu Leu Gln Lys Trp Tyr Ile Thr Ala
785 790 795 800
Arg Leu Leu Ser Pro Arg Glu Val Asn Gln Leu Ile Gly Ser Phe Arg
805 810 815

Ser Tyr Lys Gln Phe Val Ser Asp Ile Glu Arg Arg Ala Lys Glu Thr
820 825 830
Asn Asn Ser Leu Ser Lys Ser Gly Met Thr Val Asp Val Glu Asn Ile
835 840 845
Thr Lys Val Leu Asp Leu Cys Thr Lys Leu Asn Gly Arg Phe Ser Asn
850 855 860
Glu Leu Thr Asp Tyr Phe Asp Ser Lys Asp Asp Tyr Ala Val Tyr Val
865 870 875 880
Ser Lys Phe Leu Asp Phe Gly Phe Lys Ile Asp Glu Lys Phe Pro Ala
885 890 895
Ala Leu Leu Gly Glu Phe Cys Asn Lys Glu Glu Asn Gly Lys Lys Ile
900 905 910
Gly Ile Tyr His Asn Gly Thr Glu Pro Ile Leu Asn Ser Asn Ile Ile
915 920 925
Lys Ser Lys Leu Tyr Gly Ile Thr Asp Val Val Ser Arg Ala Val Lys
930 935 940
Pro Val Ser Glu Lys Leu Ile Arg Glu Tyr Leu Gln Gln Glu Val Lys
945 950 955 960

Ile Lys Pro Tyr Leu Glu Asn Gly Val Cys Lys Asn Lys Glu Glu Gln
965 970 975
Ala Ala Leu Arg Lys Tyr Gln Glu Leu Lys Asn Arg Ile Glu Phe Arg
980 985 990
Asp Ile Val Glu Tyr Ser Glu Ile Ile Asn Glu Leu Met Gly Gln Leu
995 1000 1005
Ile Asn Phe Ser Tyr Leu Arg Glu Arg Asp Leu Met Tyr Phe Gln Leu
1010 1015 1020
Gly Phe His Tyr Leu Cys Leu Asn Asn Tyr Gly Ala Lys Pro Glu Gly

1025 1030 1035 1040

Tyr Tyr Ser Ile Val Asn Asp Lys Arg Thr Ile Lys Gly Ala Ile Leu

1045 1050 1055

Tyr Gln Ile Val Ala Met Tyr Thr Tyr Gly Leu Pro Ile Tyr His Tyr

1060 1065 1070

Val Asp Gly Thr Ile Ser Asp Arg Arg Lys Asn Lys Lys Thr Val Leu

1075 1080 1085

Asp Thr Leu Asn Ser Ser Glu Thr Val Gly Ala Lys Ile Lys Tyr Phe

1090 1095 1100

Ile Tyr Tyr Ser Asp Glu Leu Phe Asn Asp Ser Leu Ile Leu Tyr Asn

1105 1110 1115 1120

Ala Gly Leu Glu Leu Phe Glu Asn Ile Asn Glu His Glu Asn Ile Val

1125 1130 1135

Asn Leu Arg Lys Tyr Ile Asp His Phe Lys Tyr Tyr Val Ser Gln Asp

1140 1145 1150

Arg Ser Leu Leu Asp Ile Tyr Ser Glu Val Phe Asp Arg Tyr Phe Thr

1155 1160 1165

Tyr Asp Arg Lys Tyr Lys Lys Asn Val Met Asn Leu Phe Ser Asn Ile

1170 1175 1180

Met Leu Lys His Phe Ile Ile Thr Asp Phe Glu Phe Ser Thr Gly Glu

1185 1190 1195 1200

Lys Thr Ile Gly Glu Lys Asn Thr Ala Lys Lys Glu Cys Ala Lys Val

1205 1210 1215

Arg Ile Lys Arg Gly Gly Leu Ser Ser Asp Lys Phe Thr Tyr Lys Phe

1220 1225 1230

Lys Asp Ala Lys Pro Ile Glu Leu Ser Ala Lys Asn Thr Glu Phe Leu

1235 1240 1245

Asp Gly Val Ala Arg Ile Leu Tyr Tyr Pro Glu Asn Val Val Leu Thr

1250 1255 1260

Asp Leu Val Arg Asn Ser Glu Val Glu Asp Glu Lys Arg Ile Glu Lys

1265 1270 1275 1280

Tyr Asp Arg Asn His Asn Ser Ser Pro Thr Arg Lys Asp Lys Thr Tyr

1285 1290 1295

Lys Gln Asp Val Lys Lys Asn Tyr Asn Lys Lys Thr Ser Lys Ala Phe

1300 1305 1310

Asp Ser Ser Lys Leu Asp Thr Lys Ser Val Gly Asn Asn Leu Ser Asp

1315 1320 1325

Asn Pro Val Leu Lys Gln Phe Leu Ser Glu Ser Lys Lys Lys Arg

1330 1335 1340

<210> 146

<211> 1276

<212> PRT

<213> Blautia sp.

<400> 146

Met Lys Ile Ser Lys Val Asp His Val Lys Ser Gly Ile Asp Gln Lys

1 5 10 15

Leu Ser Ser Gln Arg Gly Met Leu Tyr Lys Gln Pro Gln Lys Lys Tyr

20 25 30

Glu Gly Lys Gln Leu Glu Glu His Val Arg Asn Leu Ser Arg Lys Ala

35 40 45

Lys Ala Leu Tyr Gln Val Phe Pro Val Ser Gly Asn Ser Lys Met Glu

50 55 60

Lys Glu Leu Gln Ile Ile Asn Ser Phe Ile Lys Asn Ile Leu Leu Arg

65 70 75 80

Leu Asp Ser Gly Lys Thr Ser Glu Glu Ile Val Gly Tyr Ile Asn Thr

85 90 95

Tyr Ser Val Ala Ser Gln Ile Ser Gly Asp His Ile Gln Glu Leu Val

100 105 110

Asp Gln His Leu Lys Glu Ser Leu Arg Lys Tyr Thr Cys Val Gly Asp

115 120 125

Lys Arg Ile Tyr Val Pro Asp Ile Ile Val Ala Leu Leu Lys Ser Lys

130 135 140

Phe Asn Ser Glu Thr Leu Gln Tyr Asp Asn Ser Glu Leu Lys Ile Leu

145 150 155 160

Ile Asp Phe Ile Arg Glu Asp Tyr Leu Lys Glu Lys Gln Ile Lys Gln
165 170 175
Ile Val His Ser Ile Glu Asn Asn Ser Thr Pro Leu Arg Ile Ala Glu

180 185 190
Ile Asn Gly Gln Lys Arg Leu Ile Pro Ala Asn Val Asp Asn Pro Lys
195 200 205
Lys Ser Tyr Ile Phe Glu Phe Leu Lys Glu Tyr Ala Gln Ser Asp Pro
210 215 220
Lys Gly Gln Glu Ser Leu Leu Gln His Met Arg Tyr Leu Ile Leu Leu
225 230 235 240
Tyr Leu Tyr Gly Pro Asp Lys Ile Thr Asp Asp Tyr Cys Glu Glu Ile
245 250 255

Glu Ala Trp Asn Phe Gly Ser Ile Val Met Asp Asn Glu Gln Leu Phe
260 265 270
Ser Glu Glu Ala Ser Met Leu Ile Gln Asp Arg Ile Tyr Val Asn Gln
275 280 285
Gln Ile Glu Glu Gly Arg Gln Ser Lys Asp Thr Ala Lys Val Lys Lys
290 295 300
Asn Lys Ser Lys Tyr Arg Met Leu Gly Asp Lys Ile Glu His Ser Ile
305 310 315 320
Asn Glu Ser Val Val Lys His Tyr Gln Glu Ala Cys Lys Ala Val Glu

325 330 335
Glu Lys Asp Ile Pro Trp Ile Lys Tyr Ile Ser Asp His Val Met Ser
340 345 350
Val Tyr Ser Ser Lys Asn Arg Val Asp Leu Asp Lys Leu Ser Leu Pro
355 360 365
Tyr Leu Ala Lys Asn Thr Trp Asn Thr Trp Ile Ser Phe Ile Ala Met
370 375 380
Lys Tyr Val Asp Met Gly Lys Gly Val Tyr His Phe Ala Met Ser Asp
385 390 395 400

Val Asp Lys Val Gly Lys Gln Asp Asn Leu Ile Ile Gly Gln Ile Asp

405 410 415

Pro Lys Phe Ser Asp Gly Ile Ser Ser Phe Asp Tyr Glu Arg Ile Lys

420 425 430

Ala Glu Asp Asp Leu His Arg Ser Met Ser Gly Tyr Ile Ala Phe Ala

435 440 445

Val Asn Asn Phe Ala Arg Ala Ile Cys Ser Asp Glu Phe Arg Lys Lys

450 455 460

Asn Arg Lys Glu Asp Val Leu Thr Val Gly Leu Asp Glu Ile Pro Leu

465 470 475 480

Tyr Asp Asn Val Lys Arg Lys Leu Leu Gln Tyr Phe Gly Gly Ala Ser

485 490 495

Asn Trp Asp Asp Ser Ile Ile Asp Ile Ile Asp Asp Lys Asp Leu Val

500 505 510

Ala Cys Ile Lys Glu Asn Leu Tyr Val Ala Arg Asn Val Asn Phe His

515 520 525

Phe Ala Gly Ser Glu Lys Val Gln Lys Lys Gln Asp Asp Ile Leu Glu

530 535 540

Glu Ile Val Arg Lys Glu Thr Arg Asp Ile Gly Lys His Tyr Arg Lys

545 550 555 560

Val Phe Tyr Ser Asn Asn Val Ala Val Phe Tyr Cys Asp Glu Asp Ile

565 570 575

Ile Lys Leu Met Asn His Leu Tyr Gln Arg Glu Lys Pro Tyr Gln Ala

580 585 590

Gln Ile Pro Ser Tyr Asn Lys Val Ile Ser Lys Thr Tyr Leu Pro Asp

595 600 605

Leu Ile Phe Met Leu Leu Lys Gly Lys Asn Arg Thr Lys Ile Ser Asp

610 615 620

Pro Ser Ile Met Asn Met Phe Arg Gly Thr Phe Tyr Phe Leu Leu Lys

625 630 635 640

Glu Ile Tyr Tyr Asn Asp Phe Leu Gln Ala Ser Asn Leu Lys Glu Met

645 650 655

Phe Cys Glu Gly Leu Lys Asn Asn Val Lys Asn Lys Lys Ser Glu Lys

660 665 670

Pro Tyr Gln Asn Phe Met Arg Arg Phe Glu Glu Leu Glu Asn Met Gly

675 680 685

Met Asp Phe Gly Glu Ile Cys Gln Gln Ile Met Thr Asp Tyr Glu Gln

690 695 700

Gln Asn Lys Gln Lys Lys Lys Thr Ala Thr Ala Val Met Ser Glu Lys

705 710 715 720

Asp Lys Lys Ile Arg Thr Leu Asp Asn Asp Thr Gln Lys Tyr Lys His

725 730 735

Phe Arg Thr Leu Leu Tyr Ile Gly Leu Arg Glu Ala Phe Ile Ile Tyr

740 745 750

Leu Lys Asp Glu Lys Asn Lys Glu Trp Tyr Glu Phe Leu Arg Glu Pro

755 760 765

Val Lys Arg Glu Gln Pro Glu Glu Lys Glu Phe Val Asn Lys Trp Lys

770 775 780

Leu Asn Gln Tyr Ser Asp Cys Ser Glu Leu Ile Leu Lys Asp Ser Leu

785 790 795 800

Ala Ala Ala Trp Tyr Val Val Ala His Phe Ile Asn Gln Ala Gln Leu

805 810 815

Asn His Leu Ile Gly Asp Ile Lys Asn Tyr Ile Gln Phe Ile Ser Asp

820 825 830

Ile Asp Arg Arg Ala Lys Ser Thr Gly Asn Pro Val Ser Glu Ser Thr

835 840 845

Glu Ile Gln Ile Glu Arg Tyr Arg Lys Ile Leu Arg Val Leu Glu Phe

850 855 860

Ala Lys Phe Phe Cys Gly Gln Ile Thr Asn Val Leu Thr Asp Tyr Tyr

865 870 875 880

Gln Asp Glu Asn Asp Phe Ser Thr His Val Gly His Tyr Val Lys Phe

885 890 895

Glu Lys Lys Asn Met Glu Pro Ala His Ala Leu Gln Ala Phe Ser Asn

900 905 910

Ser Leu Tyr Ala Cys Gly Lys Glu Lys Lys Lys Ala Gly Phe Tyr Tyr
915 920 925

Asp Gly Met Asn Pro Ile Val Asn Arg Asn Ile Thr Leu Ala Ser Met
930 935 940

Tyr Gly Asn Lys Lys Leu Leu Glu Asn Ala Met Asn Pro Val Thr Glu
945 950 955 960

Gln Asp Ile Arg Lys Tyr Tyr Ser Leu Met Ala Glu Leu Asp Ser Val
965 970 975

Leu Lys Asn Gly Ala Val Cys Lys Ser Glu Asp Glu Gln Lys Asn Leu
980 985 990

Arg His Phe Gln Asn Leu Lys Asn Arg Ile Glu Leu Val Asp Val Leu
995 1000 1005

Thr Leu Ser Glu Leu Val Asn Asp Leu Val Ala Gln Leu Ile Gly Trp
1010 1015 1020

Val Tyr Ile Arg Glu Arg Asp Met Met Tyr Leu Gln Leu Gly Leu His
1025 1030 1035 1040

Tyr Ile Lys Leu Tyr Phe Thr Asp Ser Val Ala Glu Asp Ser Tyr Leu

1045 1050 1055

Arg Thr Leu Asp Leu Glu Glu Gly Ser Ile Ala Asp Gly Ala Val Leu
1060 1065 1070

Tyr Gln Ile Ala Ser Leu Tyr Ser Phe Asn Leu Pro Met Tyr Val Lys
1075 1080 1085

Pro Asn Lys Ser Ser Val Tyr Cys Lys Lys His Val Asn Ser Val Ala
1090 1095 1100

Thr Lys Phe Asp Ile Phe Glu Lys Glu Tyr Cys Asn Gly Asp Glu Thr
1105 1110 1115 1120

Val Ile Glu Asn Gly Leu Arg Leu Phe Glu Asn Ile Asn Leu His Lys
1125 1130 1135

Asp Met Val Lys Phe Arg Asp Tyr Leu Ala His Phe Lys Tyr Phe Ala
1140 1145 1150

Lys Leu Asp Glu Ser Ile Leu Glu Leu Tyr Ser Lys Ala Tyr Asp Phe
1155 1160 1165

Phe Phe Ser Tyr Asn Ile Lys Leu Lys Lys Ser Val Ser Tyr Val Leu
1170 1175 1180
Thr Asn Val Leu Leu Ser Tyr Phe Ile Asn Ala Lys Leu Ser Phe Ser

1185 1190 1195 1200
Thr Tyr Lys Ser Ser Gly Asn Lys Thr Val Gln His Arg Thr Thr Lys
1205 1210 1215
Ile Ser Val Val Ala Gln Thr Asp Tyr Phe Thr Tyr Lys Leu Arg Ser
1220 1225 1230
Ile Val Lys Asn Lys Asn Gly Val Glu Ser Ile Glu Asn Asp Asp Arg
1235 1240 1245
Arg Cys Glu Val Val Asn Ile Ala Ala Arg Asp Lys Glu Phe Val Asp
1250 1255 1260

Glu Val Cys Asn Val Ile Asn Tyr Asn Ser Asp Lys
1265 1270 1275
<210> 147
<211> 1168
<212> PRT
<213> Leptotrichia sp.
<400> 147
Met Lys Ile Thr Lys Ile Asp Gly Ile Ser His Lys Lys Tyr Ile Lys
1 5 10 15
Glu Gly Lys Leu Val Lys Ser Thr Ser Glu Glu Asn Lys Thr Asp Glu
20 25 30
Arg Leu Ser Glu Leu Leu Thr Ile Arg Leu Asp Thr Tyr Ile Lys Asn
35 40 45

Pro Asp Asn Ala Ser Glu Glu Glu Asn Arg Ile Arg Arg Glu Asn Leu
50 55 60
Lys Glu Phe Phe Ser Asn Lys Val Leu Tyr Leu Lys Asp Gly Ile Leu
65 70 75 80
Tyr Leu Lys Asp Arg Arg Glu Lys Asn Gln Leu Gln Asn Lys Asn Tyr
85 90 95
Ser Glu Glu Asp Ile Ser Glu Tyr Asp Leu Lys Asn Lys Asn Asn Phe

100 105 110

Leu Val Leu Lys Lys Ile Leu Leu Asn Glu Asp Ile Asn Ser Glu Glu

115 120 125

Leu Glu Ile Phe Arg Asn Asp Phe Glu Lys Lys Leu Asp Lys Ile Asn

130 135 140

Ser Leu Lys Tyr Ser Leu Glu Glu Asn Lys Ala Asn Tyr Gln Lys Ile

145 150 155 160

Asn Glu Asn Asn Ile Lys Lys Val Glu Gly Lys Ser Lys Arg Asn Ile

165 170 175

Phe Tyr Asn Tyr Tyr Lys Asp Ser Ala Lys Arg Asn Asp Tyr Ile Asn

180 185 190

Asn Ile Gln Glu Ala Phe Asp Lys Leu Tyr Lys Lys Glu Asp Ile Glu

195 200 205

Asn Leu Phe Phe Leu Ile Glu Asn Ser Lys Lys His Glu Lys Tyr Lys

210 215 220

Ile Arg Glu Cys Tyr His Lys Ile Ile Gly Arg Lys Asn Asp Lys Glu

225 230 235 240

Asn Phe Ala Thr Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Met

245 250 255

Lys Glu Leu Ile Glu Lys Val Pro Asn Val Ser Glu Leu Lys Lys Ser

260 265 270

Gln Val Phe Tyr Lys Tyr Tyr Leu Asn Lys Glu Lys Leu Asn Asp Glu

275 280 285

Asn Ile Lys Tyr Val Phe Cys His Phe Val Glu Ile Glu Met Ser Lys

290 295 300

Leu Leu Lys Asn Tyr Val Tyr Lys Lys Pro Ser Asn Ile Ser Asn Asp

305 310 315 320

Lys Val Lys Arg Ile Phe Glu Tyr Gln Ser Leu Lys Lys Leu Ile Glu

325 330 335

Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys

340 345 350

Tyr Ser Phe Tyr Leu Gln Asp Gly Glu Ile Ala Thr Ser Asp Phe Ile
355 360 365
Val Gly Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val
370 375 380
Ser Ser Thr Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn
385 390 395 400
Glu Asn Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn

405 410 415
Lys Gly Glu Glu Lys Tyr Ile Ser Gly Glu Ile Asp Lys Leu Tyr Asp
420 425 430
Asn Asn Lys Gln Asn Glu Val Lys Lys Asn Leu Lys Met Phe Tyr Ser
435 440 445
Tyr Asp Phe Asn Met Asn Ser Lys Lys Glu Ile Glu Asp Phe Phe Ser
450 455 460
Asn Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe
465 470 475 480

Asn Leu Glu Leu Glu Gly Lys Asp Ile Phe Thr Phe Lys Asn Ile Val
485 490 495
Pro Ser Gln Ile Ser Lys Lys Met Phe His Asp Glu Ile Asn Glu Lys
500 505 510
Lys Leu Lys Leu Lys Ile Phe Lys Gln Leu Asn Ser Ala Asn Val Phe
515 520 525
Arg Tyr Leu Glu Lys Tyr Lys Ile Leu Asn Tyr Leu Asn Arg Thr Arg
530 535 540
Phe Glu Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys

545 550 555 560
Leu Tyr Ser Arg Ile Asp Asp Leu Lys Asn Ser Leu Gly Ile Tyr Trp
565 570 575
Lys Thr Pro Lys Thr Asn Asp Asp Asn Lys Thr Lys Glu Ile Thr Asp
580 585 590
Ala Gln Ile Tyr Leu Leu Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn
595 600 605

Tyr Phe Met Ser Asn Asn Gly Asn Phe Phe Glu Ile Thr Lys Glu Ile
610 615 620

Ile Glu Leu Asn Lys Asn Asp Lys Arg Asn Leu Lys Thr Gly Phe Tyr
625 630 635 640
Lys Leu Gln Lys Phe Glu Asn Leu Gln Glu Lys Thr Pro Lys Glu Tyr
645 650 655
Leu Ala Asn Ile Gln Ser Leu Tyr Met Ile Asn Ala Gly Asn Gln Asp
660 665 670
Glu Glu Glu Lys Asp Thr Tyr Ile Asp Phe Ile Gln Lys Ile Phe Leu
675 680 685
Lys Gly Phe Met Thr Tyr Leu Ala Asn Asn Gly Arg Leu Ser Leu Ile

690 695 700
Tyr Ile Gly Ser Asp Glu Glu Thr Asn Thr Ser Leu Ala Glu Lys Lys
705 710 715 720
Gln Glu Phe Asp Lys Phe Leu Lys Lys Tyr Glu Gln Asn Asn Asn Ile
725 730 735
Glu Ile Pro Tyr Glu Ile Asn Glu Phe Val Arg Glu Ile Lys Leu Gly
740 745 750
Lys Ile Leu Lys Tyr Thr Glu Arg Leu Asn Met Phe Tyr Leu Ile Leu
755 760 765

Lys Leu Leu Asn His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu
770 775 780
Lys Tyr Gln Ser Ala Asn Lys Glu Glu Ala Phe Ser Asp Gln Leu Glu
785 790 795 800
Leu Ile Asn Leu Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe
805 810 815
Glu Leu Glu Ala Asp Glu Ile Gly Lys Phe Leu Asp Phe Asn Gly Asn
820 825 830
Lys Val Lys Asp Asn Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile

835 840 845
Tyr Phe Asp Gly Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile

850 855 860

Lys Lys Tyr Gly Met Leu Asn Leu Leu Glu Lys Ile Ser Asp Glu Ala

865 870 875 880

Lys Tyr Lys Ile Ser Ile Glu Glu Leu Lys Asn Tyr Ser Lys Lys Lys

885 890 895

Asn Glu Ile Glu Glu Asn His Thr Thr Gln Glu Asn Leu His Arg Lys

900 905 910

Tyr Ala Arg Pro Arg Lys Asp Glu Lys Phe Thr Asp Glu Asp Tyr Lys

915 920 925

Lys Tyr Glu Lys Ala Ile Arg Asn Ile Gln Gln Tyr Thr His Leu Lys

930 935 940

Asn Lys Val Glu Phe Asn Glu Leu Asn Leu Leu Gln Ser Leu Leu Leu

945 950 955 960

Arg Ile Leu His Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp

965 970 975

Leu Arg Phe Arg Leu Lys Gly Glu Phe Pro Glu Asn Gln Tyr Ile Glu

980 985 990

Glu Ile Phe Asn Phe Asp Asn Ser Lys Asn Val Lys Tyr Lys Asn Gly

995 1000 1005

Gln Ile Val Glu Lys Tyr Ile Asn Phe Tyr Lys Glu Leu Tyr Lys Asp

1010 1015 1020

Asp Thr Glu Lys Ile Ser Ile Tyr Ser Asp Lys Lys Val Lys Glu Leu

1025 1030 1035 1040

Lys Lys Glu Lys Lys Asp Leu Tyr Ile Arg Asn Tyr Ile Ala His Phe

1045 1050 1055

Asn Tyr Ile Pro Asn Ala Glu Ile Ser Leu Leu Glu Met Leu Glu Asn

1060 1065 1070

Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala Ile Met

1075 1080 1085

Lys Ser Ile Val Asp Ile Leu Lys Glu Tyr Gly Phe Val Val Thr Phe

1090 1095 1100

Lys Ile Glu Lys Asp Lys Lys Ile Arg Ile Glu Ser Leu Lys Ser Glu

1105 1110 1115 1120

Glu Val Val His Leu Lys Lys Leu Lys Leu Lys Asp Asn Asp Lys Lys

1125 1130 1135

Lys Glu Pro Ile Lys Thr Tyr Arg Asn Ser Lys Glu Leu Cys Lys Leu

1140 1145 1150

Val Lys Val Met Phe Glu Tyr Lys Met Lys Glu Lys Lys Ser Glu Asn

1155 1160 1165

<210> 148

<211> 1138

<212> PRT

<213> Bacteroides ihuae

<400> 148

Met Arg Ile Thr Lys Val Lys Val Lys Glu Ser Ser Asp Gln Lys Asp

1 5 10 15

Lys Met Val Leu Ile His Arg Lys Val Gly Glu Gly Thr Leu Val Leu

20 25 30

Asp Glu Asn Leu Ala Asp Leu Thr Ala Pro Ile Ile Asp Lys Tyr Lys

35 40 45

Asp Lys Ser Phe Glu Leu Ser Leu Leu Lys Gln Thr Leu Val Ser Glu

50 55 60

Lys Glu Met Asn Ile Pro Lys Cys Asp Lys Cys Thr Ala Lys Glu Arg

65 70 75 80

Cys Leu Ser Cys Lys Gln Arg Glu Lys Arg Leu Lys Glu Val Arg Gly

85 90 95

Ala Ile Glu Lys Thr Ile Gly Ala Val Ile Ala Gly Arg Asp Ile Ile

100 105 110

Pro Arg Leu Asn Ile Phe Asn Glu Asp Glu Ile Cys Trp Leu Ile Lys

115 120 125

Pro Lys Leu Arg Asn Glu Phe Thr Phe Lys Asp Val Asn Lys Gln Val

130 135 140

Val Lys Leu Asn Leu Pro Lys Val Leu Val Glu Tyr Ser Lys Lys Asn

145 150 155 160

Asp Pro Thr Leu Phe Leu Ala Tyr Gln Gln Trp Ile Ala Ala Tyr Leu
165 170 175
Lys Asn Lys Lys Gly His Ile Lys Lys Ser Ile Leu Asn Asn Arg Val
180 185 190
Val Ile Asp Tyr Ser Asp Glu Ser Lys Leu Ser Lys Arg Lys Gln Ala
195 200 205
Leu Glu Leu Trp Gly Glu Glu Tyr Glu Thr Asn Gln Arg Ile Ala Leu
210 215 220
Glu Ser Tyr His Thr Ser Tyr Asn Ile Gly Glu Leu Val Thr Leu Leu

225 230 235 240
Pro Asn Pro Glu Glu Tyr Val Ser Asp Lys Gly Glu Ile Arg Pro Ala
245 250 255
Phe His Tyr Lys Leu Lys Asn Val Leu Gln Met His Gln Ser Thr Val
260 265 270
Phe Gly Thr Asn Glu Ile Leu Cys Ile Asn Pro Ile Phe Asn Glu Asn
275 280 285
Arg Ala Asn Ile Gln Leu Ser Ala Tyr Asn Leu Glu Val Val Lys Tyr
290 295 300

Phe Glu His Tyr Phe Pro Ile Lys Lys Lys Lys Lys Asn Leu Ser Leu
305 310 315 320
Asn Gln Ala Ile Tyr Tyr Leu Lys Val Glu Thr Leu Lys Glu Arg Leu
325 330 335
Ser Leu Gln Leu Glu Asn Ala Leu Arg Met Asn Leu Leu Gln Lys Gly
340 345 350
Lys Ile Lys Lys His Glu Phe Asp Lys Asn Thr Cys Ser Asn Thr Leu
355 360 365
Ser Gln Ile Lys Arg Asp Glu Phe Phe Val Leu Asn Leu Val Glu Met

370 375 380
Cys Ala Phe Ala Ala Asn Asn Ile Arg Asn Ile Val Asp Lys Glu Gln
385 390 395 400

Val Asn Glu Ile Leu Ser Lys Lys Asp Leu Cys Asn Ser Leu Ser Lys
405 410 415
Asn Thr Ile Asp Lys Glu Leu Cys Thr Lys Phe Tyr Gly Ala Asp Phe
420 425 430
Ser Gln Ile Pro Val Ala Ile Trp Ala Met Arg Gly Ser Val Gln Gln
435 440 445

Ile Arg Asn Glu Ile Val His Tyr Lys Ala Glu Ala Ile Asp Lys Ile
450 455 460
Phe Ala Leu Lys Thr Phe Glu Tyr Asp Asp Met Glu Lys Asp Tyr Ser
465 470 475 480
Asp Thr Pro Phe Lys Gln Tyr Leu Glu Leu Ser Ile Glu Lys Ile Asp
485 490 495
Ser Phe Phe Ile Glu Gln Leu Ser Ser Asn Asp Val Leu Asn Tyr Tyr
500 505 510
Cys Thr Glu Asp Val Asn Lys Leu Leu Asn Lys Cys Lys Leu Ser Leu

515 520 525
Arg Arg Thr Ser Ile Pro Phe Ala Pro Gly Phe Lys Thr Ile Tyr Glu
530 535 540
Leu Gly Cys His Leu Gln Asp Ser Ser Asn Thr Tyr Arg Ile Gly His
545 550 555 560
Tyr Leu Met Leu Ile Gly Gly Arg Val Ala Asn Ser Thr Val Thr Lys
565 570 575
Ala Ser Lys Ala Tyr Pro Ala Tyr Arg Phe Met Leu Lys Leu Ile Tyr
580 585 590

Asn His Leu Phe Leu Asn Lys Phe Leu Asp Asn His Asn Lys Arg Phe
595 600 605
Phe Met Lys Ala Val Ala Phe Val Leu Lys Asp Asn Arg Glu Asn Ala
610 615 620
Arg Asn Lys Phe Gln Tyr Ala Phe Lys Glu Ile Arg Met Met Asn Asn
625 630 635 640
Asp Glu Ser Ile Ala Ser Tyr Met Ser Tyr Ile His Ser Leu Ser Val
645 650 655

Gln Glu Gln Glu Lys Lys Gly Asp Lys Asn Asp Lys Val Arg Tyr Asn

660 665 670

Thr Glu Lys Phe Ile Glu Lys Val Phe Val Lys Gly Phe Asp Asp Phe

675 680 685

Leu Ser Trp Leu Gly Val Glu Phe Ile Leu Ser Pro Asn Gln Glu Glu

690 695 700

Arg Asp Lys Thr Val Thr Arg Glu Glu Tyr Glu Asn Leu Met Ile Lys

705 710 715 720

Asp Arg Val Glu His Ser Ile Asn Ser Asn Gln Glu Ser His Ile Ala

725 730 735

Phe Phe Thr Phe Cys Lys Leu Leu Asp Ala Asn His Leu Ser Asp Leu

740 745 750

Arg Asn Glu Trp Ile Lys Phe Arg Ser Ser Gly Asp Lys Glu Gly Phe

755 760 765

Ser Tyr Asn Phe Ala Ile Asp Ile Ile Glu Leu Cys Leu Leu Thr Val

770 775 780

Asp Arg Val Glu Gln Arg Arg Asp Gly Tyr Lys Glu Gln Thr Glu Leu

785 790 795 800

Lys Glu Tyr Leu Ser Phe Phe Ile Lys Gly Asn Glu Ser Glu Asn Thr

805 810 815

Val Trp Lys Gly Phe Tyr Phe Gln Gln Asp Asn Tyr Thr Pro Val Leu

820 825 830

Tyr Ser Pro Ile Glu Leu Ile Arg Lys Tyr Gly Thr Leu Glu Leu Leu

835 840 845

Lys Leu Ile Ile Val Asp Glu Asp Lys Ile Thr Gln Gly Glu Phe Glu

850 855 860

Glu Trp Gln Thr Leu Lys Lys Val Val Glu Asp Lys Val Thr Arg Arg

865 870 875 880

Asn Glu Leu His Gln Glu Trp Glu Asp Met Lys Asn Lys Ser Ser Phe

885 890 895

Ser Gln Glu Lys Cys Ser Ile Tyr Gln Lys Leu Cys Arg Asp Ile Asp

900 905 910

Arg Tyr Asn Trp Leu Asp Asn Lys Leu His Leu Val His Leu Arg Lys

915 920 925

Leu His Asn Leu Val Ile Gln Ile Leu Ser Arg Met Ala Arg Phe Ile

930 935 940

Ala Leu Trp Asp Arg Asp Phe Val Leu Leu Asp Ala Ser Arg Ala Asn

945 950 955 960

Asp Asp Tyr Lys Leu Leu Ser Phe Phe Asn Phe Arg Asp Phe Ile Asn

965 970 975

Ala Lys Lys Thr Lys Thr Asp Asp Glu Leu Leu Ala Glu Phe Gly Ser

980 985 990

Lys Ile Glu Lys Lys Asn Ala Pro Phe Ile Lys Ala Glu Asp Val Pro

995 1000 1005

Leu Met Val Glu Cys Ile Glu Ala Lys Arg Ser Phe Tyr Gln Lys Val

1010 1015 1020

Phe Phe Arg Asn Asn Leu Gln Val Leu Ala Asp Arg Asn Phe Ile Ala

1025 1030 1035 1040

His Tyr Asn Tyr Ile Ser Lys Thr Ala Lys Cys Ser Leu Phe Glu Met

1045 1050 1055

Ile Ile Lys Leu Arg Thr Leu Met Tyr Tyr Asp Arg Lys Leu Arg Asn

1060 1065 1070

Ala Val Val Lys Ser Ile Ala Asn Val Phe Asp Gln Asn Gly Met Val

1075 1080 1085

Leu Gln Leu Ser Leu Asp Asp Ser His Glu Leu Lys Val Asp Lys Val

1090 1095 1100

Ile Ser Lys Arg Ile Val His Leu Lys Asn Asn Asn Ile Met Thr Asp

1105 1110 1115 1120

Gln Val Pro Glu Glu Tyr Tyr Lys Ile Cys Arg Arg Leu Leu Glu Met

1125 1130 1135

Lys Lys

<210> 149

<211> 697

<212> PRT

<213> Porphyromonadaceae bacterium

<400> 149

Met Glu Phe Arg Asp Ser Ile Phe Lys Ser Leu Leu Gln Lys Glu Ile
1 5 10 15

Glu Lys Ala Pro Leu Cys Phe Ala Glu Lys Leu Ile Ser Gly Gly Val
20 25 30

Phe Ser Tyr Tyr Pro Ser Glu Arg Leu Lys Glu Phe Val Gly Asn His
35 40 45

Pro Phe Ser Leu Phe Arg Lys Thr Met Pro Phe Ser Pro Gly Phe Lys
50 55 60

Arg Val Met Lys Ser Gly Gly Asn Tyr Gln Asn Ala Asn Arg Asp Gly
65 70 75 80

Arg Phe Tyr Asp Leu Asp Ile Gly Val Tyr Leu Pro Lys Asp Gly Phe

85 90 95

Gly Asp Glu Glu Trp Asn Ala Arg Tyr Phe Leu Met Lys Leu Ile Tyr
100 105 110

Asn Gln Leu Phe Leu Pro Tyr Phe Ala Asp Ala Glu Asn His Leu Phe
115 120 125

Arg Glu Cys Val Asp Phe Val Lys Arg Val Asn Arg Asp Tyr Asn Cys
130 135 140

Lys Asn Asn Asn Ser Glu Glu Gln Ala Phe Ile Asp Ile Arg Ser Met
145 150 155 160

Arg Glu Asp Glu Ser Ile Ala Asp Tyr Leu Ala Phe Ile Gln Ser Asn
165 170 175

Ile Ile Ile Glu Glu Asn Lys Lys Lys Glu Thr Asn Lys Glu Gly Gln
180 185 190

Ile Asn Phe Asn Lys Phe Leu Leu Gln Val Phe Val Lys Gly Phe Asp
195 200 205

Ser Phe Leu Lys Asp Arg Thr Glu Leu Asn Phe Leu Gln Leu Pro Glu
210 215 220

Leu Gln Gly Asp Gly Thr Arg Gly Asp Asp Leu Glu Ser Leu Asp Lys

225 230 235 240

Leu Gly Ala Val Val Ala Val Asp Leu Lys Leu Asp Ala Thr Gly Ile

245 250 255

Asp Ala Asp Leu Asn Glu Asn Ile Ser Phe Tyr Thr Phe Cys Lys Leu

260 265 270

Leu Asp Ser Asn His Leu Ser Arg Leu Arg Asn Glu Ile Ile Lys Tyr

275 280 285

Gln Ser Ala Asn Ser Asp Phe Ser His Asn Glu Asp Phe Asp Tyr Asp

290 295 300

Arg Ile Ile Ser Ile Ile Glu Leu Cys Met Leu Ser Ala Asp His Val

305 310 315 320

Ser Thr Asn Asp Asn Glu Ser Ile Phe Pro Asn Asn Asp Lys Asp Phe

325 330 335

Ser Gly Ile Arg Pro Tyr Leu Ser Thr Asp Ala Lys Val Glu Thr Phe

340 345 350

Glu Asp Leu Tyr Val His Ser Asp Ala Lys Thr Pro Ile Thr Asn Ala

355 360 365

Thr Met Val Leu Asn Trp Lys Tyr Gly Thr Asp Lys Leu Phe Glu Arg

370 375 380

Leu Met Ile Ser Asp Gln Asp Phe Leu Val Thr Glu Lys Asp Tyr Phe

385 390 395 400

Val Trp Lys Glu Leu Lys Lys Asp Ile Glu Glu Lys Ile Lys Leu Arg

405 410 415

Glu Glu Leu His Ser Leu Trp Val Asn Thr Pro Lys Gly Lys Lys Gly

420 425 430

Ala Lys Lys Lys Asn Gly Arg Glu Thr Thr Gly Glu Phe Ser Glu Glu

435 440 445

Asn Lys Lys Glu Tyr Leu Glu Val Cys Arg Glu Ile Asp Arg Tyr Val

450 455 460

Asn Leu Asp Asn Lys Leu His Phe Val His Leu Lys Arg Met His Ser

465 470 475 480

Leu Leu Ile Glu Leu Leu Gly Arg Phe Val Gly Phe Thr Tyr Leu Phe

485 490 495

Glu Arg Asp Tyr Gln Tyr Tyr His Leu Glu Ile Arg Ser Arg Arg Asn

500 505 510

Lys Asp Ala Gly Val Val Asp Lys Leu Glu Tyr Asn Lys Ile Lys Asp

515 520 525

Gln Asn Lys Tyr Asp Lys Asp Asp Phe Phe Ala Cys Thr Phe Leu Tyr

530 535 540

Glu Lys Ala Asn Lys Val Arg Asn Phe Ile Ala His Phe Asn Tyr Leu

545 550 555 560

Thr Met Trp Asn Ser Pro Gln Glu Glu Glu His Asn Ser Asn Leu Ser

565 570 575

Gly Ala Lys Asn Ser Ser Gly Arg Gln Asn Leu Lys Cys Ser Leu Thr

580 585 590

Glu Leu Ile Asn Glu Leu Arg Glu Val Met Ser Tyr Asp Arg Lys Leu

595 600 605

Lys Asn Ala Val Thr Lys Ala Val Ile Asp Leu Phe Asp Lys His Gly

610 615 620

Met Val Ile Lys Phe Arg Ile Val Asn Asn Asn Asn Asn Asp Asn Lys

625 630 635 640

Asn Lys His His Leu Glu Leu Asp Asp Ile Val Pro Lys Lys Ile Met

645 650 655

His Leu Arg Gly Ile Lys Leu Lys Arg Gln Asp Gly Lys Pro Ile Pro

660 665 670

Ile Gln Thr Asp Ser Val Asp Pro Leu Tyr Cys Arg Met Trp Lys Lys

675 680 685

Leu Leu Asp Leu Lys Pro Thr Pro Phe

690 695

<210> 150

<211> 204

<212> PRT

<213> Listeria riparia

<400> 150

Met His Asp Ala Trp Ala Glu Asn Pro Lys Lys Pro Gln Ser Asp Ala
1 5 10 15

Phe Leu Lys Glu Tyr Lys Ala Cys Cys Glu Ala Ile Asp Thr Tyr Asn
20 25 30

Trp His Lys Asn Lys Ala Thr Leu Val Tyr Val Asn Glu Leu His His
35 40 45

Leu Leu Ile Asp Ile Leu Gly Arg Leu Val Gly Tyr Val Ala Ile Ala
50 55 60

Asp Arg Asp Phe Gln Cys Met Ala Asn Gln Tyr Leu Lys Ser Ser Gly
65 70 75 80

His Thr Glu Arg Val Asp Ser Trp Ile Asn Thr Ile Arg Lys Asn Arg
85 90 95

Pro Asp Tyr Ile Glu Lys Leu Asp Ile Phe Met Asn Lys Ala Gly Leu
100 105 110

Phe Val Ser Glu Lys Asn Gly Arg Asn Tyr Ile Ala His Leu Asn Tyr
115 120 125

Leu Ser Pro Lys His Lys Tyr Ser Leu Leu Tyr Leu Phe Glu Lys Leu
130 135 140

Arg Glu Met Leu Lys Tyr Asp Arg Lys Leu Lys Asn Ala Val Thr Lys
145 150 155 160

Ser Leu Ile Asp Leu Leu Asp Lys His Gly Met Cys Val Val Phe Ala
165 170 175

Asn Leu Lys Asn Asn Lys His Arg Leu Val Ile Ala Ser Leu Lys Pro
180 185 190

Lys Lys Ile Glu Thr Phe Lys Trp Lys Lys Ile Lys
195 200

<210> 151

<211> 1082

<212> PRT

<213> Insolitispirillum peregrinum

<400> 151

Met Arg Ile Ile Arg Pro Tyr Gly Ser Ser Thr Val Ala Ser Pro Ser
1 5 10 15

Pro Gln Asp Ala Gln Pro Leu Arg Ser Leu Gln Arg Gln Asn Gly Thr
20 25 30

Phe Asp Val Ala Glu Phe Ser Arg Arg His Pro Glu Leu Val Leu Ala
35 40 45

Gln Trp Val Ala Met Leu Asp Lys Ile Ile Arg Lys Pro Ala Pro Gly
50 55 60

Lys Asn Ser Thr Ala Leu Pro Arg Pro Thr Ala Glu Gln Arg Arg Leu
65 70 75 80

Arg Gln Gln Val Gly Ala Ala Leu Trp Ala Glu Met Gln Arg His Thr

85 90 95

Pro Val Pro Pro Glu Leu Lys Ala Val Trp Asp Ser Lys Val His Pro
100 105 110

Tyr Ser Lys Asp Asn Ala Pro Ala Thr Ala Lys Thr Pro Ser His Arg
115 120 125

Gly Arg Trp Tyr Asp Arg Phe Gly Asp Pro Glu Thr Ser Ala Ala Thr
130 135 140

Val Ala Glu Gly Val Arg Arg His Leu Leu Asp Ser Ala Gln Pro Phe
145 150 155 160

Arg Ala Asn Gly Gly Gln Pro Lys Gly Lys Gly Val Ile Glu His Arg
165 170 175

Ala Leu Thr Ile Gln Asn Gly Thr Leu Leu His His His Gln Ser Glu
180 185 190

Lys Ala Gly Pro Leu Pro Glu Asp Trp Ser Thr Tyr Arg Ala Asp Glu
195 200 205

Leu Val Ser Thr Ile Gly Lys Asp Ala Arg Trp Ile Lys Val Ala Ala
210 215 220

Ser Leu Tyr Gln His Tyr Gly Arg Ile Phe Gly Pro Thr Thr Pro Ile

225 230 235 240

Ser Glu Ala Gln Thr Arg Pro Glu Phe Val Leu His Thr Ala Val Lys
245 250 255
Ala Tyr Tyr Arg Arg Leu Phe Lys Glu Arg Lys Leu Pro Ala Glu Arg
260 265 270
Leu Glu Arg Leu Leu Pro Arg Thr Gly Glu Ala Leu Arg His Ala Val
275 280 285
Thr Val Gln His Gly Asn Arg Ser Leu Ala Asp Ala Val Arg Ile Gly
290 295 300

Lys Ile Leu His Tyr Gly Trp Leu Gln Asn Gly Glu Pro Asp Pro Trp
305 310 315 320
Pro Asp Asp Ala Ala Leu Tyr Ser Ser Arg Tyr Trp Gly Ser Asp Gly
325 330 335
Gln Thr Asp Ile Lys His Ser Glu Ala Val Ser Arg Val Trp Arg Arg
340 345 350
Ala Leu Thr Ala Ala Gln Arg Thr Leu Thr Ser Trp Leu Tyr Pro Ala
355 360 365
Gly Thr Asp Ala Gly Asp Ile Leu Leu Ile Gly Gln Lys Pro Asp Ser

370 375 380
Ile Asp Arg Asn Arg Leu Pro Leu Leu Tyr Gly Asp Ser Thr Arg His
385 390 395 400
Trp Thr Arg Ser Pro Gly Asp Val Trp Leu Phe Leu Lys Gln Thr Leu
405 410 415
Glu Asn Leu Arg Asn Ser Ser Phe His Phe Lys Thr Leu Ser Ala Phe
420 425 430
Thr Ser His Leu Asp Gly Thr Cys Glu Ser Glu Pro Ala Glu Gln Gln
435 440 445

Ala Ala Gln Ala Leu Trp Gln Asp Asp Arg Gln Gln Asp His Gln Gln
450 455 460
Val Phe Leu Ser Leu Arg Ala Leu Asp Ala Thr Thr Tyr Leu Pro Thr
465 470 475 480
Gly Pro Leu His Arg Ile Val Asn Ala Val Gln Ser Thr Asp Ala Thr
485 490 495

Leu Pro Leu Pro Arg Phe Arg Arg Val Val Thr Arg Ala Ala Asn Thr
500 505 510
Arg Leu Lys Gly Phe Pro Val Glu Pro Val Asn Arg Arg Thr Met Glu

515 520 525
Asp Asp Pro Leu Leu Arg Cys Arg Tyr Gly Val Leu Lys Leu Leu Tyr
530 535 540
Glu Arg Gly Phe Arg Ala Trp Leu Glu Thr Arg Pro Ser Ile Ala Ser
545 550 555 560
Cys Leu Asp Gln Ser Leu Lys Arg Ser Thr Lys Ala Ala Gln Thr Ile
565 570 575
Asn Gly Lys Asn Ser Pro Gln Gly Val Glu Ile Leu Ser Arg Ala Thr
580 585 590

Lys Leu Leu Gln Ala Glu Gly Gly Gly Gly His Gly Ile His Asp Leu
595 600 605
Phe Asp Arg Leu Tyr Ala Ala Thr Ala Arg Glu Met Arg Val Gln Val
610 615 620
Gly Tyr His His Asp Ala Glu Ala Ala Arg Gln Gln Ala Glu Phe Ile
625 630 635 640
Glu Asp Leu Lys Cys Glu Val Val Ala Arg Ala Phe Cys Ala Tyr Leu
645 650 655
Lys Thr Leu Gly Ile Gln Gly Asp Thr Phe Arg Arg Gln Pro Glu Pro

660 665 670
Leu Pro Thr Trp Pro Asp Leu Pro Asp Leu Pro Ser Ser Thr Ile Gly
675 680 685
Thr Ala Gln Ala Ala Leu Tyr Ser Val Leu His Leu Met Pro Val Glu
690 695 700
Asp Val Gly Ser Leu Leu His Gln Leu Arg Arg Trp Leu Val Ala Leu
705 710 715 720
Gln Ala Arg Gly Gly Glu Asp Gly Thr Ala Ile Thr Ala Thr Ile Pro
725 730 735

Leu Leu Glu Leu Tyr Leu Asn Arg His Asp Ala Lys Phe Ser Gly Gly

740 745 750

Gly Ala Gly Thr Gly Leu Arg Trp Asp Asp Trp Gln Val Phe Phe Asp

755 760 765

Cys Gln Ala Thr Phe Asp Arg Val Phe Pro Pro Gly Pro Ala Leu Asp

770 775 780

Ser His Arg Leu Pro Leu Arg Gly Leu Arg Glu Val Leu Arg Phe Gly

785 790 795 800

Arg Val Asn Asp Leu Ala Ala Leu Ile Gly Gln Asp Lys Ile Thr Ala

805 810 815

Ala Glu Val Asp Arg Trp His Thr Ala Glu Gln Thr Ile Ala Ala Gln

820 825 830

Gln Gln Arg Arg Glu Ala Leu His Glu Gln Leu Ser Arg Lys Lys Gly

835 840 845

Thr Asp Ala Glu Val Asp Glu Tyr Arg Ala Leu Val Thr Ala Ile Ala

850 855 860

Asp His Arg His Leu Thr Ala His Val Thr Leu Ser Asn Val Val Arg

865 870 875 880

Leu His Arg Leu Met Thr Thr Val Leu Gly Arg Leu Val Asp Tyr Gly

885 890 895

Gly Leu Trp Glu Arg Asp Leu Thr Phe Val Thr Leu Tyr Glu Ala His

900 905 910

Arg Leu Gly Gly Leu Arg Asn Leu Leu Ser Glu Ser Arg Val Asn Lys

915 920 925

Phe Leu Asp Gly Gln Thr Pro Ala Ala Leu Ser Lys Lys Asn Asn Ala

930 935 940

Glu Glu Asn Gly Met Ile Ser Lys Val Leu Gly Asp Lys Ala Arg Arg

945 950 955 960

Gln Ile Arg Asn Asp Phe Ala His Phe Asn Met Leu Gln Gln Gly Lys

965 970 975

Lys Thr Ile Asn Leu Thr Asp Glu Ile Asn Asn Ala Arg Lys Leu Met

980 985 990

Ala His Asp Arg Lys Leu Lys Asn Ala Ile Thr Arg Ser Val Thr Thr

995 1000 1005

Leu Leu Gln Gln Asp Gly Leu Asp Ile Val Trp Thr Met Asp Ala Ser

1010 1015 1020

His Arg Leu Thr Asp Ala Lys Ile Asp Ser Arg Asn Ala Ile His Leu

1025 1030 1035 1040

His Lys Thr His Asn Arg Ala Asn Ile Arg Glu Pro Leu His Gly Lys

1045 1050 1055

Ser Tyr Cys Arg Trp Val Ala Ala Leu Phe Gly Ala Thr Ser Thr Pro

1060 1065 1070

Ser Ala Thr Lys Lys Ser Asp Lys Ile Arg

1075 1080

<210> 152

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 152

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

1 5 10 15

<210> 153

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 153

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

1 5 10 15

Gly Ser

<210> 154

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 154

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
1 5 10 15

<210> 155

<211> 28

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 155

gccagcuuuc cgggcauugg cuuccauc 28

<210> 156

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 156

gccagctttc cgggcattgg cttccatc 28

<210> 157

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 157

acccaggaat ctcaggaatg tcgacgat 28

<210> 158

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 158

agggttttcc cagtcacgac gttgtaaa 28

<210> 159

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 159

gggcattggc ttccatctct ttgagcacct 30

<210> 160

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 160

gugcagccag cuuuccgggc auuggcuucc 30

<210> 161

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 161

gcagccagcu uuccgggcau uggcuuccau 30

<210> 162

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 162

ggcuuccauc ucuuugagca ccuccagcgg 30

<210> 163

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 163

ggaaugucga cgaucgccuc gccuaugccg 30

<210> 164

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 164

gaaugucgac gaucgccucg ccuaugccgc 30

<210> 165

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 165

gaccugugcg augaacugcu ccaugggcuc 30

<210> 166

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 166

guguggcagc guccugggau gaacuucuuc 30

<210> 167

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 167

guggcagcgu ccugggauga acuucuucau 30

<210> 168

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 168

gcuucuugcc gggcaacuuc ccgcggucag 30

<210> 169

<211> 31

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 169

gcaggguuuu cccagucacg acguuguaaa a 31

<210> 170

<211> 31

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 170

gcaggguuuu cccagucacg acguuguaaa a 31

<210> 171

<211> 28

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 171

acccaggaau cucaggaaug ucgacgau 28

<210> 172

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 172

cagctttccg ggcattggct t 21

<210> 173

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 173

ccgcuggagg ugcucaaaga gauggaagcc 30

<210> 174

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 174

gccagctttc cgggcattgg cttccatc 28

<210> 175

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 175

acccaggaat ctcaggaatg tcgacgat 28

<210> 176

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 176

gggcattggc ttccatctct ttgagcacct 30

<210> 177

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 177

gaaugucgac gaucgccucg ccuaugccgc 30

<210> 178

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 178

caaggcactc ttgcctacgc caccagct 28

<210> 179

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 179

tcatattcgt ccacaaaatg attctgaa 28

<210> 180

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 180

attatttatg gcaaatacac aaagaaag 28

<210> 181

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 181

gaatatcttc aaatgattta gtattatt 28

<210> 182

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 182

accataggta catcttcaga gtccttaa 28

<210> 183

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 183

gtcaaggcac tcttgcctac gccaccagct 30

<210> 184

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 184

gatcatattc gtccacaaaa tgattctgaa 30

<210> 185

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 185

gtattattta tggcaaatac acaaagaaag 30

<210> 186

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 186

gtgaatatct tcaaatgatt tagtattatt 30

<210> 187

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 187

ggaccatagg tacatcttca gagtccttaa 30

<210> 188

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 188

aagagtgcct tgacgataca gcctcgaggc tgtatcgtca aggcactctt 50

<210> 189

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 189

aatcattttg tggacgaata tctcgagata ttcgtccaca aaatgatt 48

<210> 190

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 190

aaataatact aaatcatttg actcgagtca aatgatttag tattattt 48

<210> 191

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 191

aataatacta aatcatttga actcgagttc aaatgattta gtattatt 48

<210> 192

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 192

aaggactctg aagatgtacc tctcgagagg tacatcttca gagtcctt 48

<210> 193

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 193

gagatcaggg caaacagaac tttgactccc 30

<210> 194

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 194

ggatgcagat cagggcaaac agaactttga 30

<210> 195

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 195

gcacagcgat gcagatcagg gcaaacagaa 30

<210> 196

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 196

gctcggccac agcgatgcag atcagggcaa 30

<210> 197

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 197

ggggcttggc ctcggccaca gcgatgcaga 30

<210> 198

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 198

gtgggcttgg cctcggccac agcgatgcag 30

<210> 199

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 199

gtctcggtgg gcttggcctc ggccacagcg 30

<210> 200

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 200

gttcgttgtt ctcggtgggc ttggcctcgg 30

<210> 201

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 201

ggaagtcttc gttgttctcg gtgggcttgg 30

<210> 202

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 202

gatgttgaag tcttcgttgt tctcggtggg 30

<210> 203

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 203

gcggccacga tgttgaagtc ttcgttgttc 30

<210> 204

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 204

gtggccacgg ccacgatgtt gaagtcttcg 30

<210> 205

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 205

ggttgctggc cacggccacg atgttgaagt 30

<210> 206

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 206

gtcgcgaagt tgctggccac ggccacgatg 30

<210> 207

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 207

gccgtggtcg cgaagttgct ggccacggcc 30

<210> 208

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 208

gcgagatccg tggtcgcgaa gttgctggcc 30

<210> 209

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 209

gcagcatcga gatccgtggt cgcgaagttg 30

<210> 210

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 210

gggtcagcat cgagatccgt ggtcgcgaag 30

<210> 211

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 211

gcttcccgcg gtcagcatcg agatccgtgg 30

<210> 212

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 212

ggggcaactt cccgcggtca gcatcgagat 30

<210> 213

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 213

gtcttgccgg gcaacttccc gcggtcagca 30

<210> 214

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 214

ggcagcttct tgccgggcaa cttcccgcgg 30

<210> 215

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 215

gccagcggca gcttcttgcc gggcaacttc 30

<210> 216

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 216

gcacctccag cggcagcttc ttgccgggca 30

<210> 217

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 217

gctttgagca cctccagcgg cagcttcttg 30

<210> 218

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 218

gcatctcttt gagcacctcc agcggcagct 30

<210> 219

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 219

gtccatctct ttgagcacct ccagcggcag 30

<210> 220

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 220

gggcattggc ttccatctct ttgagcacct 30

<210> 221

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 221

gtccgggcat tggcttccat ctctttgagc 30

<210> 222

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 222

ggccagcttt ccgggcattg gcttccatct 30

<210> 223

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 223

gggtgcagcc agctttccgg gcattggctt 30

<210> 224

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 224

gagcccctgg tgcagccagc tttccgggca 30

<210> 225

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 225

gatcagacag cccctggtgc agccagcttt 30

<210> 226

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 226

ggcagatcag acagcccctg gtgcagccag 30

<210> 227

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 227

gacaggcaga tcagacagcc cctggtgcag 30

<210> 228

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 228

gtgatgtggg acaggcagat cagacagccc 30

<210> 229

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 229

gacttgatgt gggacaggca gatcagacag 30

<210> 230

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 230

ggggcgtgca cttgatgtgg gacaggcaga 30

<210> 231

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 231

gcttcatctt gggcgtgcac ttgatgtggg 30

<210> 232

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 232

gtgaacttct tcatcttggg cgtgcacttg 30

<210> 233

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 233

gggatgaact tcttcatctt gggcgtgcac 30

<210> 234

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 234

gtgggatgaa cttcttcatc ttgggcgtgc 30

<210> 235

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 235

gggcagcgtc ctgggatgaa cttcttcatc 30

<210> 236

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 236

gggtgtggca gcgtcctggg atgaacttct 30

<210> 237

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 237

gttcgtaggt gtggcagcgt cctgggatga 30

<210> 238

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 238

gcgccttcgt aggtgtggca gcgtcctggg 30

<210> 239

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 239

gtctttgtcg ccttcgtagg tgtggcagcg 30

<210> 240

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 240

gctttgtcgc cttcgtaggt gtggcagcgt 30

<210> 241

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 241

gtgccgccct gtgcggactc tttgtcgcct 30

<210> 242

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 242

gtatgccgcc ctgtgcggac tctttgtcgc 30

<210> 243

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 243

gcctcgccta tgccgccctg tgcggactct 30

<210> 244

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 244

ggatcgcctc gcctatgccg ccctgtgcgg 30

<210> 245

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 245

gatgtcgacg atcgcctcgc ctatgccgcc 30

<210> 246

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 246

gcaggaatgt cgacgatcgc ctcgcctatg 30

<210> 247

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 247

gaatctcagg aatgtcgacg atcgcctcgc 30

<210> 248

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 248

gcccaggaat ctcaggaatg tcgacgatcg 30

<210> 249

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 249

gccttgaacc caggaatctc aggaatgtcg 30

<210> 250

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 250

gccaagtcct tgaacccagg aatctcagga 30

<210> 251

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 251

gtgggctcca agtccttgaa cccaggaatc 30

<210> 252

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 252

gccatgggct ccaagtcctt gaacccagga 30

<210> 253

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 253

ggaactgctc catgggctcc aagtccttga 30

<210> 254

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 254

gtgcgatgaa ctgctccatg ggctccaagt 30

<210> 255

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 255

ggacctgtgc gatgaactgc tccatgggct 30

<210> 256

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 256

gacagatcga cctgtgcgat gaactgctcc 30

<210> 257

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 257

gacacacaga tcgacctgtg cgatgaactg 30

<210> 258

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 258

gtgcagtcca cacacagatc gacctgtgcg 30

<210> 259

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 259

gccagttgtg cagtccacac acagatcgac 30

<210> 260

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 260

gggcagccag ttgtgcagtc cacacacaga 30

<210> 261

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 261

gtttgaggca gccagttgtg cagtccacac 30

<210> 262

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 262

gaagcccttt gaggcagcca gttgtgcagt 30

<210> 263

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 263

gcacgttggc aagccctttg aggcagccag 30

<210> 264

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 264

gactgcacgt tggcaagccc tttgaggcag 30

<210> 265

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 265

gggtcagaac actgcacgtt ggcaagccct 30

<210> 266

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 266

gcaggtcaga acactgcacg ttggcaagcc 30

<210> 267

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 267

gagcaggtca gaacactgca cgttggcaag 30

<210> 268

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 268

ggccacttct tgagcaggtc agaacactgc 30

<210> 269

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 269

gcggcagcca cttcttgagc aggtcagaac 30

<210> 270

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 270

gtgcggcagc cacttcttga gcaggtcaga 30

<210> 271

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 271

gagcgttgcg gcagccactt cttgagcagg 30

<210> 272

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 272

gaaaggtcgc acagcgttgc ggcagccact 30

<210> 273

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 273

gctggcaaag gtcgcacagc gttgcggcag 30

<210> 274

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 274

gggcaaaggt cgcacagcgt tgcggcagcc 30

<210> 275

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 275

gtggatcttg ctggcaaagg tcgcacagcg 30

<210> 276

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 276

gcacctggcc ctggatcttg ctggcaaagg 30

<210> 277

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 277

gtggccctgg atcttgctgg caaaggtcgc 30

<210> 278

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 278

gtgatcttgt ccacctggcc ctggatcttg 30

<210> 279

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 279

gccccttgat cttgtccacc tggccctgga 30

<210> 280

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 280

gcccttgatc ttgtccacct ggccctggat 30

<210> 281

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 281

gccttgatct tgtccacctg gccctggatc 30

<210> 282

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 282

ggcaaaggtc gcacagcgtt gcggcagcca 30

<210> 283

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 283

gcaaaggtcg cacagcgttg cggcagccac 30

<210> 284

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 284

gaaggtcgca cagcgttgcg gcagccactt 30

<210> 285

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 285

gaggtcgcac agcgttgcgg cagccacttc 30

<210> 286

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 286

ggtaatgcct ggcttgtcga cgcatagtct g 31

<210> 287

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 287

gggaaccttg gccgttataa agtctgacca g 31

<210> 288

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 288

ggagggtgag aatttagaac caagattgtt g 31

<210> 289

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 289

gagtcctggc aatgaacagt ggcgcagtag 30

<210> 290

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 290

gggtgccaca gctgctatca atacattctc 30

<210> 291

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 291

gttacatact gacacattcg gcaacatgtt 30

<210> 292

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 292

gtatgtacca ggttcctgga actggaatct 30

<210> 293

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 293

gccttggttc catccaggtt ctccagggtg 30

<210> 294

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 294

gcagtgatgg gattctcagt agcttgagcg 30

<210> 295

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 295

gagcctggca tctcaacaac agcgatggtg 30

<210> 296

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 296

gtgtctgggg cgattcttac agatcttcct 30

<210> 297

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 297

gctggatctg aagtgaagtc tgtatcttcc 30

<210> 298

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 298

ggcaacgtca tcaggatttc catagagtgg 30

<210> 299

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 299

gaggcgcagg agatggtgta gtagtagaag 30

<210> 300

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 300

gagggaccct ggaattggta tcttgctttg 30

<210> 301

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 301

ggtaagagtc aacattcctg tgtgaaacct 30

<210> 302

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 302

gaccagaatc tgttttccat caacaatgag 30

<210> 303

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 303

gatggctgta gtcagtatgt caccatcttg 30

<210> 304

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 304

gtaccatcga atggatctct aatatgtacg 30

<210> 305

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 305

gagatcacag gctccttcag catcaaaaga 30

<210> 306

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 306

gctttgaccg gcgaagagac tattgcagag 30

<210> 307

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 307

gcccctcagg caatactcgt acatgcatcg 30

<210> 308

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 308

gctggtactt ctagggtgtc tccatgcttt 30

<210> 309

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 309

ggtaatgcct ggcttgtcga cgcatagtct g 31

<210> 310

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 310

gggaaccttg gccgttataa agtctgacca g 31

<210> 311

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 311

ggagggtgag aatttagaac caagattgtt g 31

<210> 312

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 312

gcatcctgcg gcctctactc tgcattcaat t 31

<210> 313

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 313

gaccatcctg cggcctctac tctgcattca a 31

<210> 314

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 314

gaaaccatcc tgcggcctct actctgcatt c 31

<210> 315

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 315

gctaaaccat cctgcggcct ctactctgca t 31

<210> 316

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 316

gttctaaacc atcctgcggc ctctactctg c 31

<210> 317

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 317

gtgttctaaa ccatcctgcg gcctctactc t 31

<210> 318

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 318

gaatgttcta aaccatcctg cggcctctac t 31

<210> 319

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 319

gagaatgttc taaaccatcc tgcggcctct a 31

<210> 320

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 320

gatagaatgt tctaaaccat cctgcggcct c 31

<210> 321

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 321

gccatagaat gttctaaacc atcctgcggc c 31

<210> 322

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 322

gttccataga atgttctaaa ccatcctgcg g 31

<210> 323

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 323

gctttccata gaatgttcta aaccatcctg c 31

<210> 324

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 324

gctctttcca tagaatgttc taaaccatcc t 31

<210> 325

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 325

gatctctttc catagaatgt tctaaaccat c 31

<210> 326

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 326

ggaatctctt tccatagaat gttctaaacc a 31

<210> 327

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 327

gcatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc a 51

<210> 328

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 328

gaccatcctg cggcctctac tctgcattca attacatact gacacattcg g 51

<210> 329

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 329

gaaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt c 51

<210> 330

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 330

gctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca t 51

<210> 331

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 331

gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca c 51

<210> 332

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 332

gtgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga c 51

<210> 333

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 333

gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact g 51

<210> 334

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 334

gagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata c 51

<210> 335

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 335

gatagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca t 51

<210> 336

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 336

gccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta c 51

<210> 337

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 337

gttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat t 51

<210> 338

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 338

gctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca a 51

<210> 339

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 339

gctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt c 51

<210> 340

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 340

gatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca t 51

<210> 341

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 341

ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg c 51

<210> 342

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 342

gtggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc t 51

<210> 343

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 343

gactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac t 51

<210> 344

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 344

ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct a 51

<210> 345

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 345

gtggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct c 51

<210> 346

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 346

gcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc c 51

<210> 347

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 347

gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg g 51

<210> 348

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 348

gggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg c 51

<210> 349

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 349

gcaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc t 51

<210> 350

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 350

gaccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat c 51

<210> 351

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 351

ggtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc a 51

<210> 352

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 352

gcatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc aacatgtttt 60

tcctggttta t 71

<210> 353

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 353

gaccatcctg cggcctctac tctgcattca attacatact gacacattcg gcaacatgtt 60

tttcctggtt t 71

<210> 354

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 354

gaaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt cggcaacatg 60

tttttcctgg t 71

<210> 355

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 355

gctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca ttcggcaaca 60

tgtttttcct g 71

<210> 356

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 356

gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca cattcggcaa 60

catgtttttc c 71

<210> 357

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 357

gtgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc 60

aacatgtttt t 71

<210> 358

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 358

gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact gacacattcg 60

gcaacatgtt t 71

<210> 359

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 359

gagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt 60

cggcaacatg t 71

<210> 360

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 360

gatagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca 60

ttcggcaaca t 71

<210> 361

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 361

gccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca 60

cattcggcaa c 71

<210> 362

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 362

gttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga 60

cacattcggc a 71

<210> 363

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 363

gctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact 60

gacacattcg g 71

<210> 364

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 364

gctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata 60

ctgacacatt c 71

<210> 365

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 365

gatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca 60

tactgacaca t 71

<210> 366

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 366

ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta 60

catactgaca c 71

<210> 367

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 367

gtggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat 60

tacatactga c 71

<210> 368

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 368

gactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca 60

attacatact g 71

<210> 369

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 369

ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt 60

caattacata c 71

<210> 370

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 370

gtggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca 60

ttcaattaca t 71

<210> 371

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 371

gcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg 60

cattcaatta c 71

<210> 372

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 372

gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc 60

tgcattcaat t 71

<210> 373

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 373

gggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac 60

tctgcattca a 71

<210> 374

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 374

gcaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct 60

actctgcatt c 71

<210> 375

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 375

gaccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct 60

ctactctgca t 71

<210> 376

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 376

ggtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc 60

ctctactctg c 71

<210> 377

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 377

gatgtaccag gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg 60

gcctctactc t 71

<210> 378

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 378

ggtatgtacc aggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg 60

cggcctctac t 71

<210> 379

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 379

gacgtatgta ccaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc 60

tgcggcctct a 71

<210> 380

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 380

gacacgtatg taccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat 60

cctgcggcct c 71

<210> 381

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 381

gcaacacgta tgtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc 60

atcctgcggc c 71

<210> 382

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 382

gcccaacacg tatgtaccag gttcctggaa ctggaatctc tttccataga atgttctaaa 60

ccatcctgcg g 71

<210> 383

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 383

ggacccaaca cgtatgtacc aggttcctgg aactggaatc tctttccata gaatgttcta 60

aaccatcctg c 71

<210> 384

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 384

gttgacccaa cacgtatgta ccaggttcct ggaactggaa tctctttcca tagaatgttc 60

taaaccatcc t 71

<210> 385

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 385

gccttgaccc aacacgtatg taccaggttc ctggaactgg aatctctttc catagaatgt 60

tctaaaccat c 71

<210> 386

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 386

gttccttgac ccaacacgta tgtaccaggt tcctggaact ggaatctctt tccatagaat 60

gttctaaacc a 71

<210> 387

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 387

gcatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc aacatgtttt 60

tcctggttta ttttcacaca gtcca 85

<210> 388

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 388

gaccatcctg cggcctctac tctgcattca attacatact gacacattcg gcaacatgtt 60

tttcctggtt tattttcaca cagtc 85

<210> 389

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 389

gaaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt cggcaacatg 60

tttttcctgg tttattttca cacag 85

<210> 390

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 390

gctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca ttcggcaaca 60

tgtttttcct ggtttatttt cacac 85

<210> 391

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 391

gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca cattcggcaa 60

catgtttttc ctggtttatt ttcac 85

<210> 392

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 392

gtgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc 60

aacatgtttt tcctggttta ttttc 85

<210> 393

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 393

gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact gacacattcg 60

gcaacatgtt tttcctggtt tattt 85

<210> 394

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 394

gagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt 60

cggcaacatg tttttcctgg tttat 85

<210> 395

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 395

gatagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca 60

ttcggcaaca tgtttttcct ggttt 85

<210> 396

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 396

gccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca 60

cattcggcaa catgtttttc ctggt 85

<210> 397

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 397

gttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga 60

cacattcggc aacatgtttt tcctg 85

<210> 398

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 398

gctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact 60

gacacattcg gcaacatgtt tttcc 85

<210> 399

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 399

gctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata 60

ctgacacatt cggcaacatg ttttt 85

<210> 400

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 400

gatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca 60

tactgacaca ttcggcaaca tgttt 85

<210> 401

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 401

ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta 60

catactgaca cattcggcaa catgt 85

<210> 402

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 402

gtggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat 60

tacatactga cacattcggc aacat 85

<210> 403

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 403

gactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca 60

attacatact gacacattcg gcaac 85

<210> 404

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 404

ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt 60

caattacata ctgacacatt cggca 85

<210> 405

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 405

gtggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca 60

ttcaattaca tactgacaca ttcgg 85

<210> 406

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 406

gcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg 60

cattcaatta catactgaca cattc 85

<210> 407

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 407

gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc 60

tgcattcaat tacatactga cacat 85

<210> 408

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 408

gggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac 60

tctgcattca attacatact gacac 85

<210> 409

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 409

gcaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct 60

actctgcatt caattacata ctgac 85

<210> 410

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 410

gaccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct 60

ctactctgca ttcaattaca tactg 85

<210> 411

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 411

ggtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc 60

ctctactctg cattcaatta catac 85

<210> 412

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 412

gatgtaccag gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg 60

gcctctactc tgcattcaat tacat 85

<210> 413

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 413

ggtatgtacc aggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg 60

cggcctctac tctgcattca attac 85

<210> 414

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 414

gacgtatgta ccaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc 60

tgcggcctct actctgcatt caatt 85

<210> 415

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 415

gacacgtatg taccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat 60

cctgcggcct ctactctgca ttcaa 85

<210> 416

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 416

gcaacacgta tgtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc 60

atcctgcggc ctctactctg cattc 85

<210> 417

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 417

gcccaacacg tatgtaccag gttcctggaa ctggaatctc tttccataga atgttctaaa 60

ccatcctgcg gcctctactc tgcat 85

<210> 418

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 418

ggacccaaca cgtatgtacc aggttcctgg aactggaatc tctttccata gaatgttcta 60

aaccatcctg cggcctctac tctgc 85

<210> 419

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 419

gttgacccaa cacgtatgta ccaggttcct ggaactggaa tctctttcca tagaatgttc 60

taaaccatcc tgcggcctct actct 85

<210> 420

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 420

gccttgaccc aacacgtatg taccaggttc ctggaactgg aatctctttc catagaatgt 60

tctaaaccat cctgcggcct ctact 85

<210> 421

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 421

gttccttgac ccaacacgta tgtaccaggt tcctggaact ggaatctctt tccatagaat 60

gttctaaacc atcctgcggc ctcta 85

<210> 422

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 422

gggttccttg acccaacacg tatgtaccag gttcctggaa ctggaatctc tttccataga 60

atgttctaaa ccatcctgcg gcctc 85

<210> 423

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 423

gttggttcct tgacccaaca cgtatgtacc aggttcctgg aactggaatc tctttccata 60

gaatgttcta aaccatcctg cggcc 85

<210> 424

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 424

gccttggttc cttgacccaa cacgtatgta ccaggttcct ggaactggaa tctctttcca 60

tagaatgttc taaaccatcc tgcgg 85

<210> 425

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 425

ggcccttggt tccttgaccc aacacgtatg taccaggttc ctggaactgg aatctctttc 60

catagaatgt tctaaaccat cctgc 85

<210> 426

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 426

gccgcccttg gttccttgac ccaacacgta tgtaccaggt tcctggaact ggaatctctt 60

tccatagaat gttctaaacc atcct 85

<210> 427

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 427

gcgccgccct tggttccttg acccaacacg tatgtaccag gttcctggaa ctggaatctc 60

tttccataga atgttctaaa ccatc 85

<210> 428

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 428

ggtcgccgcc cttggttcct tgacccaaca cgtatgtacc aggttcctgg aactggaatc 60

tctttccata gaatgttcta aacca 85

<210> 429

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 429

gtaatgcctg gcttgtcgac gcatagtctg 30

<210> 430

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 430

gaaaacgcag gttcctccag tttcgggagc agcgcacgtc tccctgtagt c 51

<210> 431

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 431

gacgcaggtt cctctagctt cgggagcagc gcacgtctcc ctgtagtcaa g 51

<210> 432

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 432

gtaatgcctg gcttgtcgac gcatagtctg 30

<210> 433

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 433

gatagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca t 51

<210> 434

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 434

gtaatgcctg gcttgtcgac gcatagtctg 30

<210> 435

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 435

gcaaggccac aaaattatcc actgtttttg gaacagtctt tccgaagaga c 51

<210> 436

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 436

gcctgtagcc aaggccacaa aattatccac tgtttttgga acagtctttc c 51

<210> 437

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 437

gctttctctc ctgtagccaa ggccacaaaa ttatccactg tttttggaac a 51

<210> 438

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 438

ggccaaatcc tttctctcct gtagccaagg ccacaaaatt atccactgtt t 51

<210> 439

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 439

gtttttgtag ccaaatcctt tctctcctgt agccaaggcc acaaaattat c 51

<210> 440

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 440

gatttgctgt ttttgtagcc aaatcctttc tctcctgtag ccaaggccac a 51

<210> 441

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 441

gacgatggaa tttgctgttt ttgtagccaa atcctttctc tcctgtagcc a 51

<210> 442

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 442

gatagaatgt tctaaacgat cctgcggcct ctactctgca ttcaattaca t 51

<210> 443

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 443

gatagaatgt tctaaacaat cctgcggcct ctactctgca ttcaattaca t 51

<210> 444

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 444

gatagaatgt tctaaactat cctgcggcct ctactctgca ttcaattaca t 51

<210> 445

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 445

ggtcccacgc ggcccacagc tgcaccagga agaagggtgc ccagcacagc a 51

<210> 446

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 446

ggggtcccac gcggcccaca gctgcaccag gaagaagggt gcccagcaca g 51

<210> 447

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 447

gccgggtccc acgcggccca cagctgcacc aggaagaagg gtgcccagca c 51

<210> 448

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 448

gggtgatgac atcccaggcg atcgtgtggc ctccaggagc ccagagcagg a 51

<210> 449

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 449

gagggtgatg acatcccagg cgatcgtgtg gcctccagga gcccagagca g 51

<210> 450

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 450

gatcagggtg atgacatccc aggcgatcgt gtggcctcca ggagcccaga g 51

<210> 451

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 451

ggtggctcca ttcactccaa tgctgagcac ttccacagag tgggttaaag c 51

<210> 452

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 452

gtttctaata tattttgcca gactgatgga ctattctcaa ttaataatga t 51

<210> 453

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 453

gagatgttgc tgtggatcca gtccacagcc agcccgtcgg gggcctggat g 51

<210> 454

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 454

gcaggccggc ccagctgcca ggtgcacctg ctcggagcat cgggccggat c 51

<210> 455

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 455

gcaaagaacc tctgggtcca agggtagacc accagcagcc tgcccagggc c 51

<210> 456

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 456

gaagagaaac ttagtttcca gggctttggt agagggcaaa ggttgatagc a 51

<210> 457

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 457

gtcagcctag tgcagagcca ctggtagttg gtggttagag tttcaagttc c 51

<210> 458

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 458

ggctcattgt gaacaggcca gtaatgtccg ggatggggcg gcataggcgg g 51

<210> 459

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 459

gtagctaaag aacttgacca agacatatca ggatccacct cagctcctag a 51

<210> 460

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 460

ggggcattgt tctgtgccca gtcctgctgg tagacctgct ccccggtggc t 51

<210> 461

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 461

gagaagtcgt tcatgtgcca ccgtgggagc gtacagtcat cattgatctt g 51

<210> 462

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 462

gggattaatg ctgaacgcac caaagttcat cccaccaccc atattactac c 51

<210> 463

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 463

gctccaaagg ctttcctcca ctgttgcaaa gttattgaat cccaagacac a 51

<210> 464

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 464

gatgaatgaa cgatttccca gaactcccta atcagaacag agtccctggt a 51

<210> 465

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 465

ggagcctctg ccggagccca gagaacccga gagtcagaca gagccagcgc c 51

<210> 466

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 466

ggcttccgtg gagacaccca atcaatttga agagatcttg aagtgatgcc a 51

<210> 467

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 467

gtgggactgc cctcctccca tttgcagatg ccgtcgtaga atcgcagcag g 51

<210> 468

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 468

gcttcttcaa tagttctcca gctacactgg caggcatatg cccgtgttcc t 51

<210> 469

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 469

gattcctttt cttcgtccca attcacctca gtggctagtc gaagaatgaa g 51

<210> 470

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 470

gcagcttcag caccttccag tcagactcct gcttcaagca ctgcagcagg a 51

<210> 471

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 471

gccatttgct tgcagtgcca ctccagagga ttccggattg ccataaatac t 51

<210> 472

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 472

gttcaatagt tttggtccag tatcgtttac agcccttctt ggtagatttc a 51

<210> 473

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 473

ggcaaccgtc ttctgaccaa atggcagaac atttgtcccc aactttccac t 51

<210> 474

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 474

gcgactttcc aatgaaccac tgaagcccag gtatgacaaa gccgatgatc t 51

<210> 475

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 475

gtttacactc atgcttccac agctttaaca gatcatttgg ttccttgatg a 51

<210> 476

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 476

gcttaagctt ccgtgtccag ccttcaggca gggtggggtc atcatacatg g 51

<210> 477

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 477

ggacagctgg gctgatccat gatgtcatcc agaaacactg gggaccctca g 51

<210> 478

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 478

gtctcatctc aactttccat atccgtatca tggaatcata gcatcctgta a 51

<210> 479

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 479

gcatgcagac gcggttccac tcgcagccac agttccagca ccactcgagc c 51

<210> 480

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 480

gttggttagg gtcaacccag tattctccac tcttgagttc aggatggcag a 51

<210> 481

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 481

gctacactgt ccaacaccca ctctcgggtc accacaggtg cctcacacat c 51

<210> 482

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 482

gctgcactgt gtacccccag agctccgtgt tgccgacatc ctggggtggc t 51

<210> 483

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 483

gagcttcctg ccactcccaa caggtttcac agtaagcgcg tatctgttcc a 51

<210> 484

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 484

gacggcaaga gcttacccag tcacttgtgt ggagacttaa atacttgcat a 51

<210> 485

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 485

gcaaggccac aaaattatcc actgtttttg gaacagtctt tccgaagaga c 51

<210> 486

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 486

gcctgtagcc aaggccacaa aattatccac tgtttttgga acagtctttc c 51

<210> 487

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 487

gctttctctc ctgtagccaa ggccacaaaa ttatccactg tttttggaac a 51

<210> 488

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 488

ggccaaatcc tttctctcct gtagccaagg ccacaaaatt atccactgtt t 51

<210> 489

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 489

gtttttgtag ccaaatcctt tctctcctgt agccaaggcc acaaaattat c 51

<210> 490

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 490

gatttgctgt ttttgtagcc aaatcctttc tctcctgtag ccaaggccac a 51

<210> 491

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 491

gacgatggaa tttgctgttt ttgtagccaa atcctttctc tcctgtagcc a 51

<210> 492

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 492

gtaatgcctg gcttgtcgac gcatagtctg 30

<210> 493

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 493

gatagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca t 51

<210> 494

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 494

gcagggtttt cccagtcacg acgttgtaaa gttg 34

<210> 495

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 495

gtcaaggcac tcttgcccac gccaccagct ccaactacca caagtttata t 51

<210> 496

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 496

gcaaagatca cccggcccac atcttcatct ccaattcgta ggtcaaaata c 51

<210> 497

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 497

gcgccaccag ctccaaccac cacaagttta tattcagtca ttttcagcag g 51

<210> 498

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 498

gtttctccat caattaccac ttgcttcctg taggaatcct ctattgttgg a 51

<210> 499

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 499

gctttctctc ctgtagccaa ggccacaaaa ttatccactg tttttggaac a 51

<210> 500

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 500

gtaatgcctg gcttgtcgac gcatagtctg 30

<210> 501

<211> 83

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 501

tctttccata ggccctgaaa aagggcctgt tctaaaccat cctgcggcct ctactcggcc 60

ctgaaaaagg gccattcaat tac 83

<210> 502

<211> 83

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 502

cagctggcga ggccctgaaa aagggccggg gatgtgccgc aaggcgatta agttggggcc 60

ctgaaaaagg gccacgccag ggt 83

<210> 503

<211> 79

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 503

gtggaatagt ataacaatat gctaaatgtt gttatagtat cccactctaa accatcctgc 60

ggggccctct tcagggccc 79

<210> 504

<211> 79

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 504

gtggaatagt ataacaatat gctaaatgtt gttatagtat cccacaccct ggcgttaccc 60

agggccctct tcagggccc 79

<210> 505

<211> 45

<212> RNA

<213> Cypridina luciferase

<400> 505

auugaaugca gaguagaggc cgcaggauag uuuagaacau ucuau 45

<210> 506

<211> 66

<212> RNA

<213> Cypridina luciferase

<400> 506

ucuaaaccau ccugcggccu cuacucugca guuguggaag guccaguuuu gggggcuauu 60

acaaca 66

<210> 507

<211> 105

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 507

tgccgaaugu gucaguaugu aauugaaugc agaguagagg ccgcaggaua guuuagaaca 60

uucuauggaa agagauucca guuccaggaa ccugguacau acgug 105

<210> 508

<211> 65

<212> RNA

<213> Homo sapiens

<400> 508

ucgucuaugu gcugugcugg gcacccuucu uccuggugca gcuguaggcc gcgugggacc 60

cggag 65

<210> 509

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 509

gucccacgcg gcccacagcu gcaccaggaa gaagggugcc cagcacagca 50

<210> 510

<211> 49

<212> RNA

<213> Artificial Sequence

<220

><223> Synthetic

<400> 510

gggucccacg cggcccacag ugcaccagga agaagggugc ccagcacag 49

<210> 511

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 511

ccggguccca cgcggcccac agcugcacca ggaagaaggg ugcccagcac 50

<210> 512

<211> 65

<212> RNA

<213> Homo sapiens

<400> 512

ccucaacuuc cugcucuggg cuccuggagg ccacacgauc gccuaggaug ucaucacccu 60

gauug 65

<210> 513

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 513

ggugaugaca ucccaggcga ucguguggcc uccaggagcc cagagcagga 50

<210> 514

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 514

agggugauga caucccaggc gaucgugugg ccuccaggag cccagagcag 50

<210> 515

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 515

aucaggguga ugacauccca ggcgaucgug uggccuccag gagcccagag 50

<210> 516

<211> 99

<212> RNA

<213> Unknown

<220><223> Synthetic

<400> 516

cccaggugcg ggagagaggc cugcugaaaa ugacugaaua uaaacuugug guaguuggag 60

cugguggcgu aggcaagagu gccuugacga uacagcuaa 99

<210> 517

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 517

caccacaagu uuauauucag ucauuuucag caggccucuc ucccgcaccu 50

<210> 518

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 518

gcuccaacca ccacaaguuu auauucaguc auuuucagca ggccucucuc 50

<210> 519

<211> 49

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 519

cgccaccagc uccaacccca caaguuuaua uucagucauu uucagcagg 49

<210> 520

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 520

cuugccuacg ccaccagcuc caaccaccac aaguuuauau ucagucauuu 50

<210

> 521

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 521

aaggcacucu ugccuacgcc acagcuccaa ccaccacaag uuuuauauuc 50

<210> 522

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 522

guatcgucaa ggcacucuug ccuacgccac caguccaacc accacaagu 49

<210> 523

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 523

aauuagcugu aucgucaagg cacucuugcc uacgccacca gcuccaacca 50

<210> 524

<211> 44

<212> DNA

<213> Cypridinia luciferase

<400> 524

tatgtaattg aatgcagagt agaggccgca ggatagttta gaac 44

<210> 525

<211> 44

<212> DNA

<213> Cypridinia luciferase

<400> 525

tatgtaattg aatgcagagt agaggccgca ggatagttta gaac 44

<210> 526

<211> 38

<212> RNA

<213> Unknown

<220><223> Synthetic

<220><221> misc_feature

<222> (4)..(5)

<223> n is a, c, g, or u

<220

><221> misc_feature

<222> (36)..(38)

<223> n is a, c, g, or u

<400> 526

augnnuugcc accgacgaca ccgugaggau aaauunnn 38

<210> 527

<211> 36

<212> RNA

<213> Homo sapiens

<400> 527

uggaggugcu caaagagaug gaagccaaug cccgga 36

<210> 528

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)

<223> n is a, c, g, or u

<400>

528

nggaggugcu caaagagaug gaagccaaug cccgga 36

<210> 529

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (2)

<223> n is a, c, g, or u

<400> 529

ungaggugcu caaagagaug gaagccaaug cccgga 36

<210> 530

<211> 35

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (3)

<223> n is a, c, g, or u

<400> 530

ugnaggugcu caaagagaug gaagccaaug ccgga 35

<210> 531

<211> 35

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (4)

<223> n is a, c, g, or u

<400> 531

uggnggugcu caaagagaug gaagccaaug ccgga 35

<210> 532

<211> 35

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (18)

<223

> n is a, c, g, or u

<400> 532

uggaggugcu caaagagnug gaagccaaug ccgga 35

<210> 533

<211> 35

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (35)

<223> n is a, c, g, or u

<400> 533

uggaggugcu caaagagaug gaagccaaug ccggn 35

<210> 534

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(2)

<223> n is a, c, g, or u

<400> 534

nngaggugcu caaagagaug gaagccaaug cccgga 36

<210> 535

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)

<223> n is a, c, g, or u

<220><221> misc_feature

<222> (3)

<223> n is a, c, g, or u

<400> 535

ngnaggugcu caaagagaug gaagccaaug cccgga 36

<210> 536

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)

<223> n is a, c, g, or u

<220><221> misc_feature

<222> (4)

<223> n is a, c, g, or u

<400> 536

nggnggugcu caaagagaug gaagccaaug cccgga 36

<210> 537

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)

<223> n is a, c, g, or u

<220><221> misc_feature

<222> (5)

<223> n is a, c, g, or u

<400> 537

nggangugcu caaagagaug gaagccaaug cccgga 36

<210> 538

<211> 37

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (15)

<223> n is a, c, g, or u

<220><221> misc_feature

<222> (19)

<223> n is a, c, g, or u

<400> 538

uggaggugcu caaangagnu ggaagccaau gcccgga 37

<210> 539

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (35)..(36)

<223> n is a, c, g, or u

<400> 539

uggaggugcu caaagagaug gaagccaaug cccgnn 36

<210> 540

<211> 56

<212> RNA

<213> Cypridinia luciferase

<400> 540

aguauguaau ugaaugcaga guagaggccg caggauaguu uagaacauuc uaugga 56

<210> 541

<211> 50

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (34)

<223> n at position 34 is empty

<400> 541

uacauuaacu uacgucucau cuccggcguc cuancaaauc uuguaagaua 50

<210> 542

<211> 59

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 542

auguaauuga augcagagua gaggccgcag gauaguuuag aacauucuau ggaaagaga 59

<210> 543

<211> 83

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 543

ucuuuccaua ggcccugaaa aagggccugu ucuaaaccau ccugcggccu cuacucggcc 60

cugaaaaagg gccauucaau uac 83

<210> 544

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 544

ggccgcagga uaguuuagaa c 21

<210> 545

<211> 54

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 545

aaauguuguu auaguauccc accucuaaac cauccugcgg gcccuagggc ccuu 54

<210> 546

<211> 26

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 546

accgguggaa uaguauaaca auaugc 26