

**(19) 대한민국특허청(KR)**
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0062289

(43) 공개일자 2022년05월16일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) **C12N 15/10** (2017.01)
C12N 15/113 (2010.01) **C12Q 1/6818** (2018.01)

(52) CPC특허분류
C12N 9/22 (2013.01)
C12N 15/102 (2013.01)

(21) 출원번호 10-2022-7008114
 (22) 출원일자(국제) 2020년08월11일
 심사청구일자 없음

(85) 번역문제출일자 2022년03월11일
 (86) 국제출원번호 PCT/US2020/045759
 (87) 국제공개번호 WO 2021/030344
 국제공개일자 2021년02월18일

(30) 우선권주장
 62/885,483 2019년08월12일 미국(US)
 (뒷면에 계속)

(71) 출원인
라이프에디트 테라퓨틱스, 인크.
 미국 27560 노스캐롤라이나주 모리스빌 에어포트
 불러바드 507 스위트 101

(72) 발명자
보웬, 타이슨 디.
 미국 27560 노스캐롤라이나주 모리스빌 윌튼스 크
 락 로드 137
크롤리, 알렉산드라 브리너
 미국 27511 노스캐롤라이나주 캐리 스콧 플레이스
 1220
 (뒷면에 계속)

(74) 대리인
양영준, 이상남

전체 청구항 수 : 총 34 항

(54) 발명의 명칭 **RNA-가이드된 뉴클레아제 및 그의 활성 단편 및 변이체 및 사용 방법****(57) 요약**

관심 표적 서열에 결합하기 위한 조성물 및 방법이 제공된다. 조성물은 관심 표적 서열을 절단 또는 변형시키고, 관심 표적 서열을 가시화하고, 관심 서열의 발현을 변형시키는데 사용된다. 조성물은 RNA-가이드된 뉴클레아제 폴리펩티드, CRISPR RNA, 트랜스-활성화 CRISPR RNA, 가이드 RNA, 및 그를 코딩하는 핵산 분자를 포함한다. 핵산 분자를 포함하는 벡터 및 숙주 세포가 또한 제공된다. RNA-가이드된 뉴클레아제 폴리펩티드 및 하나 이상의 가이드 RNA를 포함하는, 관심 표적 서열에 결합하기 위한 CRISPR 시스템이 추가로 제공된다. 표적 DNA 서열을 검출하기 위한 방법 및 키트가 또한 제공된다.

(52) CPC특허분류

C12N 15/113 (2013.01)

C12Q 1/6818 (2018.05)

C12N 2310/20 (2017.05)

(72) 발명자

일리치, 테드 디.

미국 27705 노스캐롤라이나주 더럼 알룸우드 플레
이스 5

코일, 마이클

미국 27514 노스캐롤라이나주 채플 힐 엔. 엘리엇
로드 407

(30) 우선권주장

62/901,875 2019년09월18일 미국(US)

63/030,088 2020년05월26일 미국(US)

명세서

청구범위

청구항 1

RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드를 코딩하는 폴리뉴클레오티드를 포함하는 핵산 분자이며, 여기서 상기 폴리뉴클레오티드는

a) 서열식별번호: 117, 30, 75, 1, 9, 16, 23, 38, 46, 61, 69, 82, 89, 95, 103 또는 110에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 RGN 폴리펩티드; 및

b) 서열식별번호: 54로 제시된 아미노산 서열을 포함하는 RGN 폴리펩티드

로 이루어진 군으로부터 선택되는 RGN 폴리펩티드를 코딩하는 뉴클레오티드 서열을 포함하고;

여기서, 상기 RGN 폴리펩티드는 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA (gRNA)에 결합될 때 RNA-가이드 서열 특이적 방식으로 DNA 분자의 상기 표적 DNA 서열에 결합할 수 있고,

여기서, 상기 RGN 폴리펩티드를 코딩하는 폴리뉴클레오티드는 상기 폴리뉴클레오티드에 대해 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.

청구항 2

제1항에 있어서, 상기 RGN 폴리펩티드가 뉴클레아제 불활성이거나 또는 니카제로서 기능할 수 있는 것인 핵산 분자.

청구항 3

제1항 또는 제2항에 있어서, RGN 폴리펩티드가 염기-편집 폴리펩티드에 작동가능하게 융합된 것인 핵산 분자.

청구항 4

제1항 내지 제3항 중 어느 한 항의 핵산 분자를 포함하는 벡터.

청구항 5

제4항에 있어서, 벡터가 상기 표적 DNA 서열에 혼성화할 수 있는 상기 gRNA를 코딩하는 적어도 하나의 뉴클레오티드 서열을 추가로 포함하고, 가이드 RNA가 서열식별번호: 118, 31, 76, 2, 10, 17, 24, 39, 47, 55, 62, 70, 83, 90, 96, 104 또는 111에 대해 적어도 95% 서열 동일성을 갖는 CRISPR 반복 서열을 포함하는 CRISPR RNA를 포함하는 것인 벡터.

청구항 6

제5항에 있어서, 상기 gRNA가 서열식별번호: 119, 32, 77, 3, 11, 18, 25, 40, 48, 56, 63, 71, 84, 91, 97, 105 또는 112에 대해 적어도 95% 서열 동일성을 갖는 tracrRNA를 포함하는 것인 벡터.

청구항 7

제1항 내지 제3항 중 어느 한 항의 핵산 분자 또는 제4항 내지 제6항 중 어느 한 항의 벡터를 포함하는 세포.

청구항 8

CRISPR RNA (crRNA)를 코딩하는 폴리뉴클레오티드를 포함하는 핵산 분자이며, 여기서 상기 crRNA는 스페이서 서열 및 CRISPR 반복 서열을 포함하고, 상기 CRISPR 반복 서열은 서열식별번호: 118, 31, 76, 2, 10, 17, 24, 39, 47, 55, 62, 70, 83, 90, 96, 104, 또는 111에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하고;

여기서,

a) 상기 crRNA; 및 임의로

b) 상기 crRNA의 상기 CRISPR 반복 서열에 혼성화할 수 있는 트랜스-활성화 CRISPR RNA (tracrRNA)

를 포함하는 가이드 RNA는

상기 가이드 RNA가 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드에 결합될 때 상기 crRNA의 스페이서 서열을 통해 서열 특이적 방식으로 DNA 분자의 표적 DNA 서열에 혼성화할 수 있고,

여기서, 상기 crRNA를 코딩하는 폴리뉴클레오티드는 상기 폴리뉴클레오티드에 대해 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.

청구항 9

제8항의 핵산 분자를 포함하는 벡터.

청구항 10

제9항에 있어서, 벡터가 상기 tracrRNA를 코딩하는 폴리뉴클레오티드를 추가로 포함하고, 상기 tracrRNA가 서열 식별번호: 119, 32, 77, 3, 11, 18, 25, 40, 48, 56, 63, 71, 84, 91, 97, 105 또는 112에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 벡터.

청구항 11

서열식별번호: 119, 32, 77, 3, 11, 18, 25, 40, 48, 56, 63, 71, 84, 91, 97, 105 또는 112에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 트랜스-활성화 CRISPR RNA (tracrRNA)를 코딩하는 폴리뉴클레오티드를 포함하는 핵산 분자이며;

여기서,

a) 상기 tracrRNA; 및

b) 스페이서 서열 및 CRISPR 반복 서열을 포함하는 crRNA를 포함하고, 여기서 상기 tracrRNA는 상기 crRNA의 상기 CRISPR 반복 서열과 혼성화할 수 있는 것인 가이드 RNA는

상기 가이드 RNA가 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드에 결합될 때 상기 crRNA의 스페이서 서열을 통해 서열 특이적 방식으로 DNA 분자의 표적 DNA 서열에 혼성화할 수 있고,

여기서, 상기 tracrRNA를 코딩하는 폴리뉴클레오티드는 상기 폴리뉴클레오티드에 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.

청구항 12

제11항의 핵산 분자를 포함하는 벡터.

청구항 13

제9항에 있어서, 상기 벡터가 상기 crRNA를 코딩하는 폴리뉴클레오티드를 추가로 포함하고, 상기 crRNA의 CRISPR 반복 서열이 서열식별번호: 118, 31, 76, 2, 10, 17, 24, 39, 47, 55, 62, 70, 83, 90, 96, 104 또는 111에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 벡터.

청구항 14

DNA 분자의 표적 DNA 서열에 결합하기 위한 시스템이며, 상기 시스템은

a) 상기 표적 DNA 서열에 혼성화할 수 있는 하나 이상의 가이드 RNA, 또는 상기 하나 이상의 가이드 RNA (gRNA)를 코딩하는 뉴클레오티드 서열을 포함하는 하나 이상의 폴리뉴클레오티드; 및

b) i) 서열식별번호: 117, 30, 75, 1, 9, 16, 23, 38, 46, 61, 69, 82, 89, 95, 103 또는 110에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열; 및

ii) 서열식별번호: 54로 제시된 아미노산 서열

로 이루어진 군으로부터 선택되는 아미노산 서열을 포함하는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드;

또는 상기 RGN 폴리펩티드를 코딩하는 뉴클레오티드 서열을 포함하는 폴리뉴클레오티드를 포함하고;

여기서, 상기 하나 이상의 가이드 RNA를 코딩하는 서열 및 상기 RGN 폴리펩티드를 코딩하는 뉴클레오티드 서열은 각각 상기 뉴클레오티드 서열에 이중성인 프로모터에 작동가능하게 연결되고;

여기서, 하나 이상의 가이드 RNA는 RGN 폴리펩티드와 복합체를 형성하여 상기 RGN 폴리펩티드가 DNA 분자의 상기 표적 DNA 서열에 결합하도록 지시할 수 있는 것인 시스템.

청구항 15

제14항에 있어서, 표적 DNA 서열이 세포 내에 있는 것인 시스템.

청구항 16

제14항 또는 제15항에 있어서, 상기 RGN 폴리펩티드가 뉴클레아제 불활성이거나 또는 니카제로서 기능할 수 있고, RGN 폴리펩티드가 염기-편집 폴리펩티드에 작동가능하게 연결된 것인 시스템.

청구항 17

제14항 또는 제15항에 있어서, 하나 이상의 공여자 폴리뉴클레오티드, 또는 하나 이상의 공여자 폴리뉴클레오티드를 코딩하는 하나 이상의 뉴클레오티드 서열을 추가로 포함하는 시스템.

청구항 18

제14항 내지 제17항 중 어느 한 항에 따른 시스템을 표적 DNA 서열 또는 표적 DNA 서열을 포함하는 세포에 전달하는 것을 포함하는, DNA 분자의 상기 표적 DNA 서열에 결합시키는 방법.

청구항 19

제14항 내지 제17항 중 어느 한 항에 따른 시스템을 표적 DNA 서열 또는 DNA 분자를 포함하는 세포에 전달하는 것을 포함하며, 상기 표적 DNA 서열의 절단 또는 변형이 일어나는 것인, DNA 분자의 상기 표적 DNA 서열을 절단 또는 변형시키는 방법.

청구항 20

DNA 분자의 표적 DNA 서열에 결합시키는 방법이며, 상기 방법은

a) RNA-가이드된 뉴클레아제 (RGN) 리보뉴클레오티드 복합체의 형성에 적합한 조건 하에

i) 표적 DNA 서열에 혼성화할 수 있는 하나 이상의 가이드 RNA; 및

ii) A) 서열식별번호: 117, 30, 75, 1, 9, 16, 23, 38, 46, 61, 69, 82, 89, 95, 103 또는 110에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열; 및

B) 서열식별번호: 54로 제시된 아미노산 서열

로 이루어지는 군으로부터 선택되는 아미노산 서열을 포함하는 RGN 폴리펩티드

를 조합하여 시험관내에서 RGN 리보뉴클레오티드 복합체를 조립하는 단계; 및

b) 상기 표적 DNA 서열 또는 상기 표적 DNA 서열을 포함하는 세포를 시험관내-조립된 RGN 리보뉴클레오티드 복합체와 접촉시키는 단계

를 포함하고;

여기서, 하나 이상의 가이드 RNA는 표적 DNA 서열에 혼성화하여, 상기 RGN 폴리펩티드가 상기 표적 DNA 서열에 결합하도록 지시하는 것인 방법.

청구항 21

DNA 분자의 표적 DNA 서열을 절단 및/또는 변형시키는 방법이며, 상기 방법은 DNA 분자를

a) i) 서열식별번호: 117, 30, 75, 1, 9, 16, 23, 38, 46, 61, 69, 82, 89, 95, 103 또는 110에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열; 및

ii) 서열식별번호: 54로 제시된 아미노산 서열

로 이루어진 군으로부터 선택되는 아미노산 서열을 포함하는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드;

b) (a)의 RGN을 표적 DNA 서열에 표적화할 수 있는 하나 이상의 가이드 RNA

와 접촉시키는 것을 포함하고;

여기서, 하나 이상의 가이드 RNA는 표적 DNA 서열에 혼성화하여, 상기 RGN 폴리펩티드가 상기 표적 DNA 서열에 결합하도록 지시하고, 상기 표적 DNA 서열의 절단 및/또는 변형이 일어나는 것인 방법.

청구항 22

제21항에 있어서, 상기 변형된 표적 DNA 서열이 표적 DNA 서열 내의 적어도 하나의 뉴클레오타이드의 결실 또는 돌연변이를 포함하는 것인 방법.

청구항 23

제21항 또는 제22항에 있어서, 상기 RGN 폴리펩티드가 뉴클레아제 불활성이거나 또는 니카제로서 기능하고, RGN 폴리펩티드가 염기-편집 폴리펩티드에 작동가능하게 연결된 것인 방법.

청구항 24

제21항에 있어서, 상기 변형된 표적 DNA 서열이 표적 DNA 서열 내로의 이중성 DNA의 삽입을 포함하는 것인 방법.

청구항 25

제21항 내지 제24항 중 어느 한 항에 있어서, 표적 DNA 서열이 세포 내에 있는 것인 방법.

청구항 26

제25항에 있어서, RGN 폴리펩티드가 발현되고 표적 DNA 서열을 절단하여 변형된 DNA 서열을 포함하는 DNA 분자를 생성하는 조건 하에 세포를 배양하고; 상기 변형된 표적 DNA 서열을 포함하는 세포를 선택하는 것을 추가로 포함하는 방법.

청구항 27

제26항의 방법에 따라 변형된 표적 DNA 서열을 포함하는 세포.

청구항 28

DNA 분자의 표적 DNA 서열에 결합시키기 위한 시스템이며, 상기 시스템은

a) 상기 표적 DNA 서열에 혼성화할 수 있는 하나 이상의 가이드 RNA, 또는 하나 이상의 가이드 RNA (gRNA)를 코딩하는 하나 이상의 뉴클레오타이드 서열을 포함하는 하나 이상의 폴리뉴클레오타이드; 및

b) i) 서열식별번호: 117, 30, 75, 1, 9, 16, 23, 38, 46, 61, 69, 82, 89, 95, 103 또는 110에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열; 및

ii) 서열식별번호: 54로 제시된 아미노산 서열

로 이루어진 군으로부터 선택되는 아미노산 서열을 포함하는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드를 포함하고;

여기서, 하나 이상의 가이드 RNA는 표적 DNA 서열에 혼성화할 수 있고,

여기서, 하나 이상의 가이드 RNA는 RGN 폴리펩티드와 복합체를 형성하여 상기 RGN 폴리펩티드가 DNA 분자의 상기 표적 DNA 서열에 결합하도록 지시할 수 있는 것인 시스템.

청구항 29

샘플에서 DNA 분자의 표적 DNA 서열을 검출하는 방법이며, 상기 방법은

a) 샘플을

i) 서열식별번호: 54 또는 137에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 상기 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA에 결합될 때 RNA-가이드된 서열 특이적 방식으로 DNA 분자의 상기 표적 DNA 서열에 결합할 수 있는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드;

ii) 상기 가이드 RNA; 및

iii) 가이드 RNA와 혼성화하지 않는 검출 단일-가닥 DNA (ssDNA)

와 접촉시키는 단계; 및

b) RGN에 의한 검출 ssDNA의 절단에 의해 생성된 검출가능한 신호를 측정하여 표적 DNA를 검출하는 단계를 포함하는 방법.

청구항 30

샘플에서 DNA 분자의 표적 DNA 서열을 검출하기 위한 키트이며, 상기 키트는

a) 서열식별번호: 54 또는 137에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 상기 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA에 결합될 때 RNA-가이드된 서열 특이적 방식으로 DNA 분자의 상기 표적 DNA 서열에 결합할 수 있는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드;

b) 상기 가이드 RNA; 및

c) 가이드 RNA와 혼성화하지 않는 검출 단일-가닥 DNA (ssDNA)

를 포함하는 키트.

청구항 31

제29항 또는 제30항에 있어서, 상기 검출 ssDNA가 형광단/켄처 쌍 또는 형광 공명 에너지 전달 (FRET) 쌍을 포함하는 것인 방법 또는 키트.

청구항 32

단일-가닥 DNA를 절단하는 방법이며, 상기 방법은 표적 DNA 서열 및 복수의 비-표적 ssDNA를 포함하는 DNA 분자를 포함하는 핵산의 집단을

a) 서열식별번호: 54 또는 137에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 상기 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA에 결합될 때 상기 표적 DNA 서열에 RNA-가이드된 서열 특이적 방식으로 결합할 수 있는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드; 및

b) 상기 가이드 RNA

와 접촉시키는 것을 포함하고;

여기서, RGN 폴리펩티드가 상기 복수의 비-표적 ssDNA를 절단하는 것인 방법.

청구항 33

CRISPR RNA (crRNA)를 코딩하는 폴리뉴클레오티드를 포함하는 핵산 분자이며, 여기서 상기 crRNA는 스페이서 서열 및 CRISPR 반복 서열을 포함하고, 상기 CRISPR 반복 서열은 서열식별번호: 240에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하고;

여기서,

a) 상기 crRNA; 및 임의로

b) 상기 crRNA의 상기 CRISPR 반복 서열에 혼성화할 수 있는 트랜스-활성화 CRISPR RNA (tracrRNA)

를 포함하는 가이드 RNA는

상기 가이드 RNA가 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드에 결합될 때 상기 crRNA의 스페이서 서열을 통해 서열 특이적 방식으로 DNA 분자의 표적 DNA 서열에 혼성화할 수 있고,

여기서, 상기 crRNA를 코딩하는 폴리뉴클레오티드는 상기 폴리뉴클레오티드에 대해 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.

청구항 34

서열식별번호: 241에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 트랜스-활성화 CRISPR RNA (tracrRNA)를 코딩하는 폴리뉴클레오티드를 포함하는 핵산 분자이며;

여기서,

a) 상기 tracrRNA; 및

b) 스페이서 서열 및 CRISPR 반복 서열을 포함하는 crRNA를 포함하고, 여기서 상기 tracrRNA는 상기 crRNA의 상기 CRISPR 반복 서열과 혼성화할 수 있는 것인 가이드 RNA는

상기 가이드 RNA가 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드에 결합될 때 상기 crRNA의 스페이서 서열을 통해 서열 특이적 방식으로 DNA 분자의 표적 DNA 서열에 혼성화할 수 있고,

여기서, 상기 tracrRNA를 코딩하는 폴리뉴클레오티드는 상기 폴리뉴클레오티드에 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.

발명의 설명

기술 분야

[0001] 본 발명은 분자 생물학 및 유전자 편집 분야에 관한 것이다.

[0002] <EFS-웹을 통한 텍스트 파일로서 제출된 서열 목록에 대한 참조>

[0003] 본 출원은 EFS-웹을 통해 ASCII 포맷으로 제출된 서열 목록을 함유하며, 이는 그 전문이 본원에 참조로 포함된다. 2020년 8월 11일에 생성된 상기 ASCII 카피는 파일명이 L103438_1170W0_0049_2_Seq_List.txt이고, 크기가 489,326 바이트이다.

배경 기술

[0004] 표적화된 게놈 편집 또는 변형은 기초 및 응용 연구를 위한 중요한 도구가 되고 있다. 초기 방법은 뉴클레아제, 예컨대 메가뉴클레아제, 아연 핑거 융합 단백질 또는 TALEN의 조작을 수반하였고, 각각의 특정한 표적 서열에 대해 특이적인 조작된, 프로그래밍가능한, 서열-특이적인 DNA-결합 도메인을 갖는 키메라 뉴클레아제의 생성을 필요로 한다. RNA-가이드된 뉴클레아제, 예컨대 CRISPR-Cas 박테리아 시스템의 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR)-연관된 (Cas) 단백질은 특정한 표적 서열과 특이적으로 혼성화하는 가이드 RNA와 뉴클레아제를 복합체화시킴으로써 특이적인 서열의 표적화를 가능하게 한다. 표적-특이적인 가이드 RNA의 생성은 각각의 표적 서열에 대한 키메라 뉴클레아제를 생성하는 것에 비해 비용이 적게 들고 더 효율적이다. 이러한 RNA-가이드된 뉴클레아제를 사용하여, 임의로 오류가 발생하기 쉬운 비상동성 말단-연결 (NHEJ)을 통해 복구되는 서열-특이적인 이중-가닥 파손에 의해 게놈을 편집하여, 특이적인 게놈 위치에 돌연변이를 도입시킬 수 있다. 대안적으로, 이중성 DNA는 상동성-지정된 복구를 통해 게놈 부위에 도입될 수 있다.

발명의 내용

[0005] 관심 표적 서열에 결합시키기 위한 조성물 및 방법이 제공된다. 조성물은 관심 표적 서열의 절단 또는 변형, 관심 표적 서열의 검출, 및 관심 서열의 발현의 변형에 사용될 수 있다. 조성물은 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드, CRISPR RNA (crRNA), 트랜스-활성화 CRISPR RNA (tracrRNA), 가이드 RNA (grRNA), 이를 코딩하는 핵산 분자, 및 핵산 분자를 포함하는 벡터 및 숙주 세포를 포함한다. 관심 표적 서열과 결합하기 위한 CRISPR 시스템 또한 제공되며, CRISPR 시스템은 RNA-가이드된 뉴클레아제 폴리펩티드 및 하나 이상의 가이드 RNA를 포함한다. 따라서, 본원에 개시된 방법은 관심 표적 서열의 결합, 일부 실시양태에서, 관심 표적 서열의

절단 또는 변형을 위해 도출된다. 관심 표적 서열은 예를 들어 도입된 공여자 서열에 의한 비상동성 말단 결합 또는 상동성-지정된 복구의 결과로서 변형될 수 있다. 검출 단일-가닥 DNA를 사용하여 DNA 분자의 표적 DNA 서열을 검출하기 위한 방법 및 키트가 추가로 제공된다.

발명을 실시하기 위한 구체적인 내용

- [0006] 본원에 제시된 본 발명의 여러 변형 및 다른 실시양태는 관련 기술분야의 기술자에 의해 이들 발명이 상기 설명 및 관련된 도면에 제공된 교시내용의 이익을 갖는 것으로 고려될 것이다. 따라서, 본 발명이 개시된 구체적인 실시양태로 제한되지 않고, 변형 및 다른 실시양태가 첨부된 실시양태의 범위 내에 포함되는 것으로 의도됨을 이해해야 한다. 특정한 용어들이 본원에서 사용되지만, 이들은 단지 일반적인 설명적인 의미로만 사용되고, 제한하려는 목적이 아니다.
- [0007] I. 개요
- [0008] RNA-가이드된 뉴클레아제 (RGN)는 게놈 내의 단일 부위의 표적화된 조작을 가능하게 하고, 치료 및 연구 적용을 위해 유전자 표적화의 맥락에서 유용하다. 포유동물을 비롯한 다양한 유기체에서, RNA-가이드된 뉴클레아제는 예를 들어 비상동성 말단 결합 및 상동성 재조합을 자극함으로써 게놈 조작에 사용되었다. 본원에 기재된 조성물 및 방법은 폴리뉴클레오티드에서 단일- 또는 이중-가닥 파손을 생성하거나, 폴리뉴클레오티드를 변형시키거나, 폴리뉴클레오티드 내의 특정한 부위를 검출하거나, 또는 특정한 유전자의 발현을 변형시키는데 유용하다.
- [0009] 본원에 개시된 RNA-가이드된 뉴클레아제는 표적 서열을 변형시킴으로써 유전자 발현을 변경시킬 수 있다. 구체적인 실시양태에서, RNA-가이드된 뉴클레아제는 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR) RNA-가이드된 뉴클레아제 시스템의 일부로서 가이드 RNA (gRNA)에 의해 표적 서열로 지시된다. RGN은 "RNA-가이드된" 것으로 간주되는데, 이는 가이드 RNA가 RNA-가이드된 뉴클레아제와 복합체를 형성하여, RNA-가이드된 뉴클레아제가 표적 서열에 결합하고, 일부 실시양태에서는 표적 서열에 단일-가닥 또는 이중-가닥 절단을 도입하도록 지시하기 때문이다. 표적 서열이 절단된 후에, 파손은 표적 서열의 DNA 서열이 복구 과정 동안에 변형되도록 복구될 수 있다. 따라서, 숙주 세포의 DNA에서 표적 서열을 변형시키기 위해 RNA-가이드된 뉴클레아제를 사용하는 방법이 본원에 제공된다. 예를 들어, RNA-가이드된 뉴클레아제를 사용하여 진핵 세포 또는 원핵 세포의 게놈 유전자좌에서 표적 서열을 변형시킬 수 있다.
- [0010] II. RNA-가이드된 뉴클레아제
- [0011] RNA-가이드된 뉴클레아제가 본원에 제공된다. 용어 RNA-가이드된 뉴클레아제 (RGN)는 서열-특이적인 방식으로 특정한 표적 뉴클레오티드 서열에 결합하며, 폴리펩티드와 복합체화하고 표적 서열과 혼성화하는 가이드 RNA 분자에 의해 표적 뉴클레오티드 서열로 가이드되는 폴리펩티드를 지칭한다. RNA-가이드된 뉴클레아제가 결합시 표적 서열을 절단시킬 수 있지만, 용어 RNA-가이드된 뉴클레아제는 또한 표적 서열을 절단하지 않고 그에 결합할 수 있는 뉴클레아제-데드(nuclease-dead) RNA-가이드된 뉴클레아제를 포함한다. RNA-가이드된 뉴클레아제에 의한 표적 서열의 절단은 단일- 또는 이중-가닥 파손을 일으킬 수 있다. 이중-가닥 핵산 분자의 단일 가닥만을 절단할 수 있는 RNA-가이드된 뉴클레아제는 본원에서 닉카제로 지칭된다.
- [0012] 본원에 개시된 RNA-가이드된 뉴클레아제는 APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG05712.1, APG01658.1, APG06498.1, APG09106.1, APG09882.1, APG02675.1, APG01405.1, APG06250.1, APG06877.1, APG09053.1, APG04293.1, APG01308.1, APG06646.1, APG09748, 및 APG07433.1 RNA-가이드된 뉴클레아제 (이들의 아미노산 서열은 각각 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 또는 235로 제시됨), 및 RNA-가이드된 서열-특이적 방식으로 표적 뉴클레오티드 서열에 결합하는 능력을 보유하는 그의 활성 단편 또는 변이체를 포함한다. 이들 실시양태 중 일부에서, APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG05712.1, APG01658.1, APG06498.1, APG09106.1, APG09882.1, APG02675.1, APG01405.1, APG06250.1, APG06877.1, APG09053.1, APG04293.1, APG01308.1, APG06646.1, APG09748, 또는 APG07433.1 RGN의 활성 단편 또는 변이체는 단일- 또는 이중-가닥 표적 서열을 절단할 수 있다. 일부 실시양태에서, APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG05712.1, APG01658.1, APG06498.1, APG09106.1, APG09882.1, APG02675.1, APG01405.1, APG06250.1, APG06877.1, APG09053.1, APG04293.1, APG01308.1, APG06646.1, APG09748, 또는 APG07433.1 RGN의 활성 변이체는 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 또는 235로 제시된 아미노산 서열에 대해 적어도 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과 서열 동일성을 갖는 아미노산 서열을 포함한다. 특정 실시양태에서, APG05733.1, APG06207.1, APG01647.1,

APG08032.1, APG05712.1, APG01658.1, APG06498.1, APG09106.1, APG09882.1, APG02675.1, APG01405.1, APG06250.1, APG06877.1, APG09053.1, APG04293.1, APG01308.1, APG06646.1, APG09748, 또는 APG07433.1 RGN의 활성 단편은 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137, 또는 235로 제시된 아미노산 서열의 적어도 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050개 또는 그 초과인 접 아미노산 잔기를 포함한다. 본원에 제공된 RNA-가이드된 뉴클레아제는 적어도 하나의 뉴클레아제 도메인 (예를 들어, DNase, RNase 도메인), 및 가이드 RNA와 상호작용하기 위한 적어도 하나의 RNA 인식 및/또는 RNA 결합 도메인을 포함할 수 있다. 추가로, 본원에 제공된 RNA-가이드된 뉴클레아제에서 확인될 수 있는 도메인에는 다음이 포함되나 이로 제한되지 않는다: DNA 결합 도메인, 헬리카제 도메인, 단백질-단백질 상호작용 도메인, 및 이합체화 도메인. 구체적인 실시양태에서, 본원에 제공된 RNA-가이드된 뉴클레아제는 DNA 결합 도메인, 헬리카제 도메인, 단백질-단백질 상호작용 도메인, 및 이합체화 도메인 중 하나 이상에 대해 적어도 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%를 포함할 수 있다.

[0013] 표적 뉴클레오티드 서열은 본원에 제공된 RNA-가이드된 뉴클레아제에 의해 결합되고, RNA-가이드된 뉴클레아제와 회합된 가이드 RNA와 혼성화한다. 이어서, 폴리펩티드가 뉴클레아제 활성을 갖는 경우에, 표적 서열은 후속적으로 RNA-가이드된 뉴클레아제에 의해 절단될 수 있다. 용어 "절단하다" 또는 "절단"은 표적 서열 내에 단일-가닥 또는 이중-가닥 파손을 생성할 수 있는, 표적 뉴클레오티드 서열의 백본 내에서 적어도 하나의 포스포디에스테르 결합의 가수분해를 지칭한다. 본원에 개시된 RGN은 엔도뉴클레아제로서 기능하여 폴리뉴클레오티드 내의 뉴클레오티드를 절단할 수 있거나, 또는 폴리뉴클레오티드의 말단 (5' 및/또는 3' 말단)으로부터 연속적인 뉴클레오티드를 제거하는 엑소뉴클레아제일 수 있다. 다른 실시양태에서, 개시된 RGN은 폴리뉴클레오티드의 임의의 위치 내에서 표적 서열의 뉴클레오티드를 절단할 수 있고, 따라서 엔도뉴클레아제 및 엑소뉴클레아제 둘 다로 기능할 수 있다. 본원에 개시된 RGN의 표적 폴리뉴클레오티드의 절단은 엇갈린 파손 또는 평활 말단을 생성할 수 있다.

[0014] 본원에 개시된 RNA-가이드된 뉴클레아제는 박테리아 또는 고세균 종으로부터 유래된 야생형 서열일 수 있다. 대안적으로, RNA-가이드된 뉴클레아제는 야생형 폴리펩티드의 변이체 또는 단편일 수 있다. 야생형 RGN은 예를 들어 뉴클레아제 활성을 변경시키도록 또는 PAM 특이성을 변경시키도록 변형될 수 있다. 일부 실시양태에서, RNA-가이드된 뉴클레아제는 자연 발생이 아니다.

[0015] 특정한 실시양태에서, RNA-가이드된 뉴클레아제는 Nickase로서 기능하여, 표적 뉴클레오티드 서열의 단일 가닥만을 절단한다. 이러한 RNA-가이드된 뉴클레아제는 단일 기능 뉴클레아제 도메인을 갖는다. 이들 일부 실시양태에서, 추가의 뉴클레아제 도메인은 뉴클레아제 활성이 감소되거나 또는 제거되도록 돌연변이되었다.

[0016] 다른 실시양태에서, RNA-가이드된 뉴클레아제는 뉴클레아제 활성이 완전히 결여되거나 또는 감소된 뉴클레아제 활성을 나타내고, 본원에서 뉴클레아제-데드 또는 뉴클레아제 불활성으로 지칭된다. 돌연변이를 아미노산 서열에 도입시키기 위한 관련 기술분야에 공지된 임의의 방법, 예컨대 PCR-매개된 돌연변이유발 및 부위-지정된 돌연변이유발은 Nickase 또는 뉴클레아제-데드 RGN을 생성하기 위해 이용될 수 있다. 예를 들어, 미국 공개 번호 2014/0068797 및 미국 특허 번호 9,790,490을 참고하며; 이들 각각은 그의 전문이 참조로 포함된다.

[0017] 뉴클레아제 활성이 결여된 RNA-가이드된 뉴클레아제는 융합된 폴리펩티드, 폴리뉴클레오티드, 또는 소분자 페이로드를 특정한 게놈 위치로 전달하기 위해 사용될 수 있다. 이들 일부 실시양태에서, RGN 폴리펩티드 또는 가이드 RNA는 특정한 서열의 검출이 가능하도록 검출가능한 표지에 융합될 수 있다. 비제한적인 예로서, 뉴클레아제-데드 RGN은 검출가능한 표지 (예를 들어, 형광 단백질)에 융합될 수 있고, 질환-연관된 서열의 검출이 가능하도록 질환과 연관된 특정한 서열에 대해 표적화될 수 있다.

[0018] 대안적으로, 뉴클레아제-데드 RGN은 원하는 서열의 발현을 변경시키도록 특정한 게놈 위치에 대해 표적화될 수 있다. 일부 실시양태에서, 뉴클레아제-데드 RNA-가이드된 뉴클레아제와 표적 서열의 결합은 표적화된 게놈 영역 내에서 RNA 폴리머라제 또는 전사 인자의 결합을 방해함으로써 표적 서열, 또는 표적 서열에 의한 전사 조절하에 있는 유전자의 발현을 억제한다. 다른 실시양태에서, RGN (예를 들어, 뉴클레아제-데드 RGN) 또는 그의 복합체화된 가이드 RNA는 표적 서열과 결합시 표적 서열, 또는 표적 서열에 의한 전사 조절하에 있는 유전자의 발현을 억제하거나 또는 활성화시키는 작용을 하는 발현 조절인자를 추가로 포함한다. 이들 일부 실시양태에서, 발현 조절인자는 후성적인 메커니즘을 통해 표적 서열 또는 조절된 유전자의 발현을 조절한다.

[0019] 다른 실시양태에서, 뉴클레아제-데드 RGN 또는 Nickase 활성만을 갖는 RGN은, 염기-편집 폴리펩티드, 예를 들어 뉴클레오티드 염기를 탈아미노화시키는 데아미나제 폴리펩티드 또는 그의 활성 변이체 또는 단편과의 융합을 통

해 표적 폴리뉴클레오티드의 서열을 변형시켜, 하나의 뉴클레오티드 염기를 또 다른 염기로 전환시키도록, 특정한 게놈 위치에 대해 표적화될 수 있다. 염기-편집 폴리펩티드는 그의 N-말단 또는 C-말단에서 RGN에 융합될 수 있다. 추가로, 염기-편집 폴리펩티드는 펩티드 링커를 통해 RGN에 융합될 수 있다. 이러한 조성물 및 방법에 유용한 데아미나제 폴리펩티드의 비제한적 예는 시티딘 데아미나제 또는 아데노신 데아미나제 (예컨대, [Gaudelli et al. (2017) Nature 551:464-471], 미국 공개 번호 2017/0121693 및 2018/0073012, 및 국제 공개 번호 WO/2018/027078에 기재된 아데노신 데아미나제 염기 편집제, 또는 국제 출원 번호 PCT/US2019/068079에 개시된 임의의 데아미나제, 이들 각각은 그 전문이 본원에 참조로 포함됨)를 포함한다. 추가로, RGN과 염기-편집 효소 사이의 특정 융합 단백질은 또한 데아미나제에 의해 핵산 분자에서 시티딘, 데옥시티딘 또는 시토신의 티미딘, 데옥시티미딘 또는 티민으로의 돌연변이를 증가시키는 적어도 하나의 우라실 안정화 폴리펩티드를 포함할 수 있는 것으로 관련 기술분야에 공지되어 있다. 우라실 안정화 폴리펩티드의 비제한적 예는 2020년 7월 15일에 출원된 미국 가출원 번호 63/052,175에 개시된 것들, 및 염기 편집 효율을 증가시킬 수 있는 우라실 글리코실라제 억제제 (UGI) 도메인 (서열식별번호: 261) (미국 특허 번호 10,167,547, 본원에 참조로 포함됨)을 포함한다. 따라서, 융합 단백질은 본원에 기재된 RGN 또는 그의 변이체, 데아미나제, 및 임의로 적어도 하나의 우라실 안정화 폴리펩티드, 예컨대 UGI를 포함할 수 있다.

[0020] 폴리펩티드 또는 도메인에 융합된 RNA-가이드된 뉴클레아제는 링커에 의해 분리되거나 또는 결합될 수 있다. 본원에서 사용된 바와 같이, 용어 "링커"는 2개의 분자 또는 모이어티, 예를 들어 뉴클레아제의 결합 도메인 및 절단 도메인을 연결하는 화학적 기 또는 분자를 지칭한다. 일부 실시양태에서, 링커는 RNA 가이드된 뉴클레아제의 gRNA 결합 도메인 및 염기-편집 폴리펩티드, 예컨대 데아미나제를 결합시킨다. 일부 실시양태에서, 링커는 뉴클레아제-데드 RGN 및 데아미나제를 결합시킨다. 전형적으로, 링커는 2개의 기, 분자 또는 다른 모이어티 사이에 위치하거나 또는 그에 의해 플랭킹되고, 공유 결합을 통해 서로 연결되어 이들 둘을 연결시킨다. 일부 실시양태에서, 링커는 아미노산 또는 여러개의 아미노산 (예를 들어, 펩티드 또는 단백질)이다. 일부 실시양태에서, 링커는 유기 분자, 기, 중합체, 또는 화학적 모이어티이다. 일부 실시양태에서, 링커는 5-100개 아미노산 길이, 예를 들어 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150 또는 150-200개 아미노산 길이이다. 더 길거나 또는 더 짧은 링커 또한 고려된다.

[0021] 본원에 개시된 RNA-가이드된 뉴클레아제는 세포의 핵으로 RGN의 수송을 증강시키기 위해 적어도 하나의 핵 국제화 신호 (NLS)를 포함할 수 있다. 핵 국제화 신호는 관련 기술분야에 공지되어 있고, 일반적으로 염기성 아미노산의 스트레치를 포함한다 (예를 들어, [Lange et al., J. Biol. Chem. (2007) 282:5101-5105] 참고). 특정한 실시양태에서, RGN은 2, 3, 4, 5, 6개 또는 그 초과개의 핵 국제화 신호를 포함한다. 핵 국제화 신호(들)은 이중성 NLS일 수 있다. 본원에 개시된 RGN에 대해 유용한 핵 국제화 신호의 비제한적인 예는 SV40 거대 T-항원, 뉴클레오파스민, 및 c-Myc의 핵 국제화 신호이다 (예를 들어, [Ray et al. (2015) Bioconjug Chem 26(6):1004-7] 참고). 특정한 실시양태에서, RGN은 서열식별번호: 125 또는 127로 제시된 NLS 서열을 포함한다. RGN은 그의 N-말단, C-말단, 또는 N-말단 및 C-말단 둘 다에 하나 이상의 NLS 서열을 포함할 수 있다. 예를 들어, RGN은 N-말단 영역에 2개의 NLS 서열 및 C-말단 영역에서 4개의 NLS 서열을 포함할 수 있다.

[0022] 폴리펩티드를 특정한 세포하 위치(들)에 국제화시키는 관련 기술분야에 공지된 다른 국제화 신호 서열, 예컨대 비제한적으로 색소체 국제화 서열, 미토콘드리아 국제화 서열, 및 색소체 및 미토콘드리아 둘 다를 표적화하는 이중-표적화 신호 서열을 또한 사용하여, RGN을 표적화할 수 있다 (예를 들어, [Nassoury and Morse (2005) Biochim Biophys Acta 1743:5-19; Kunze and Berger (2015) Front Physiol dx.doi.org/10.3389/fphys.2015.00259; Herrmann and Neupert (2003) IUBMB Life 55:219-225; Soll (2002) Curr Opin Plant Biol 5:529-535; Carrie and Small (2013) Biochim Biophys Acta 1833:253-259; Carrie et al. (2009) FEBS J 276:1187-1195; Silva-Filho (2003) Curr Opin Plant Biol 6:589-595; Peeters and Small (2001) Biochim Biophys Acta 1541:54-63; Murcha et al. (2014) J Exp Bot 65:6301-6335; Mackenzie (2005) Trends Cell Biol 15:548-554; Glaser et al. (1998) Plant Mol Biol 38:311-338] 참고).

[0023] 특정한 실시양태에서, 본원에 개시된 RNA-가이드된 뉴클레아제는 RGN의 세포 흡수를 용이하게 하는 적어도 하나의 세포-투과 도메인을 포함한다. 세포-투과 도메인은 관련 기술분야에 공지되어 있고, 일반적으로 양으로 하전된 아미노산 잔기 (즉, 다중양이온성 세포-투과 도메인), 교대하는 극성 아미노산 잔기 및 비-극성 아미노산 잔기 (즉, 양친매성 세포-투과 도메인), 또는 소수성 아미노산 잔기 (즉, 소수성 세포-투과 도메인)의 스트레치를 포함한다 (예를 들어, [Milletti F. (2012) Drug Discov Today 17:850-860] 참고). 세포-투과 도메인의 비제한적인 예는 인간 면역결핍 바이러스 1로부터의 트랜스-활성화 전사 활성화제 (TAT)이다.

- [0024] 핵 국제화 신호, 색소체 국제화 신호, 미토콘드리아 국제화 신호, 이중-표적화 국제화 신호, 및/또는 세포-투과 도메인은 RNA-가이드된 뉴클레아제의 아미노-말단 (N-말단), 카르복실-말단 (C-말단), 또는 내부 위치에 위치할 수 있다.
- [0025] 본원에 개시된 RGN은 이펙터 도메인, 예컨대 절단 도메인, 데아미나제 도메인, 또는 발현 조절인자 도메인에 직접적으로 또는 링커 펩티드를 통해 간접적으로 융합될 수 있다. 이러한 도메인은 RNA-가이드된 뉴클레아제의 N-말단, C-말단, 또는 내부 위치에 위치할 수 있다. 이들 일부 실시양태에서, 융합 단백질의 RGN 성분은 뉴클레아제-데드 RGN이다.
- [0026] 일부 실시양태에서, RGN 융합 단백질은 절단 도메인을 포함하며, 이는 폴리뉴클레오티드 (즉, RNA, DNA, 또는 RNA/DNA 혼성체)를 절단할 수 있는 임의의 도메인이고, 제한 엔도뉴클레아제 및 귀소 엔도뉴클레아제, 예컨대 유형 IIS 엔도뉴클레아제 (예를 들어, FokI)가 포함되나 이로 제한되지 않는다 (예를 들어, [Belfort et al. (1997) *Nucleic Acids Res.* 25:3379-3388; Linn et al. (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993] 참고).
- [0027] 다른 실시양태에서, RGN 융합 단백질은 뉴클레오티드 염기를 탈아미노화시켜 하나의 뉴클레오티드 염기를 또 다른 염기로 전환시키는 데아미나제 도메인을 포함하며, 시티딘 데아미나제 또는 아데노신 데아미나제 염기 편집기가 포함되나 이로 제한되지 않는다 (예를 들어, [Gaudelli et al. (2017) *Nature* 551:464-471], 미국 공개 번호 2017/0121693 및 2018/0073012, 미국 특허 번호 9,840,699, 및 국제 공개 번호 WO/2018/027078 참고).
- [0028] 일부 실시양태에서, RGN 융합 단백질의 이펙터 도메인은 발현 조절인자 도메인일 수 있고, 이는 전사를 상향조절하거나 또는 하향조절하는 도메인이다. 발현 조절인자 도메인은 후성적 변형 도메인, 전사 리프레이서 도메인 또는 전사 활성화 도메인일 수 있다.
- [0029] 이들 일부 실시양태에서, RGN 융합 단백질의 발현 조절인자는 DNA 서열을 변경시키지 않고 히스톤 구조 및/또는 염색체 구조를 변경시키도록 DNA 또는 히스톤 단백질을 공유적으로 변형시켜, 유전자 발현에서의 변화 (즉, 상향조절 또는 하향조절)를 유도하는 후성적 변형 도메인을 포함한다. 후성적 변형의 비제한적인 예에는 DNA에서 리신 잔기의 아세틸화 또는 메틸화, 아르기닌 메틸화, 세린 및 트레오닌 인산화, 및 히스톤 단백질의 리신 유비퀴틴화 및 SUMO화, 및 시토신 잔기의 메틸화 및 히드록시메틸화가 포함된다. 후성적 변형 도메인의 비제한적인 예에는 히스톤 아세틸트랜스퍼라제 도메인, 히스톤 데아세틸라제 도메인, 히스톤 메틸트랜스퍼라제 도메인, 히스톤 데메틸라제 도메인, DNA 메틸트랜스퍼라제 도메인, 및 DNA 데메틸라제 도메인이 포함된다.
- [0030] 다른 실시양태에서, 융합 단백질의 발현 조절인자는 적어도 하나의 유전자의 전사를 감소시키거나 또는 종료시키기 위해, 전사 조절 요소 및/또는 전사 조절 단백질, 예컨대 RNA 폴리머라제 및 전사 인자와 상호작용하는 전사 리프레이서 도메인을 포함한다. 전사 리프레이서 도메인은 관련 기술분야에 공지되어 있고, Sp1-유사 리프레이서, IκB, 및 크루펠(Kruppel) 연관된 박스 (KRAB) 도메인이 포함되나 이로 제한되지 않는다.
- [0031] 여전히 다른 실시양태에서, 융합 단백질의 발현 조절인자는, 적어도 하나의 유전자의 전사를 증가시키거나 또는 활성화시키기 위해, 전사 조절 요소 및/또는 전사 조절 단백질, 예컨대 RNA 폴리머라제 및 전사 인자와 상호작용하는 전사 활성화 도메인을 포함한다. 전사 활성화 도메인은 관련 기술분야에 공지되어 있고, 단순 포진 바이러스 VP16 활성화 도메인 및 NFAT 활성화 도메인이 포함되나 이로 제한되지 않는다.
- [0032] 본원에 개시된 RGN 폴리펩티드는 검출가능한 표지 또는 정제 태그를 포함할 수 있다. 검출가능한 표지 또는 정제 태그는 RNA-가이드된 뉴클레아제의 N-말단, C-말단 또는 내부 위치에 직접적으로 또는 링커 펩티드를 통해 간접적으로 위치할 수 있다. 이들 일부 실시양태에서, 융합 단백질의 RGN 성분은 뉴클레아제-데드 RGN이다. 다른 실시양태에서, 융합 단백질의 RGN 성분은 니카제 활성을 갖는 RGN이다.
- [0033] 검출가능한 표지는 가시화될 수 있거나 또는 달리 관찰될 수 있는 분자이다. 검출가능한 표지는 융합 단백질 (예를 들어, 형광 단백질)로서 RGN에 융합될 수 있거나, 또는 시각적으로 또는 다른 수단에 의해 검출될 수 있는 RGN 폴리펩티드에 접합된 소분자일 수 있다. 본원에 개시된 RGN에 융합 단백질로서 융합될 수 있는 검출가능한 표지에는 임의의 검출가능한 단백질 도메인, 예컨대 비제한적으로 특이적인 항체에 의해 검출될 수 있는 형광 단백질 또는 단백질 도메인이 포함된다. 형광 단백질의 비제한적인 예에는 녹색 형광 단백질 (예를 들어, GFP, EGFP, ZsGreen1) 및 황색 형광 단백질 (예를 들어, YFP, EYFP, ZsYellow1)이 포함된다. 소분자 검출가능한 표지의 비제한적인 예에는 방사선 활성 표지, 예컨대 ³H 및 ³⁵S가 포함된다.
- [0034] 본 발명의 RGN 폴리펩티드는 또한 정제 태그를 포함할 수 있고, 이는 혼합물 (예를 들어, 생물학적 샘플, 배양

배지)로부터 단백질을 또는 융합된 단백질을 분리하기 위해 사용될 수 있는 임의의 분자이다. 정제 태그의 비제한적인 예에는 비오틴, myc, 말토스 결합 단백질 (MBP), 및 글루타티온-S-트랜스퍼라제 (GST)가 포함된다.

[0035] II. 가이드 RNA

[0036] 본 개시내용은 가이드 RNA, 및 이를 코딩하는 폴리뉴클레오타이드를 제공한다. 용어 "가이드 RNA"는, 표적 서열과 혼성화하도록 표적 뉴클레오타이드 서열과 충분한 상보성을 가지며, 연관된 RNA-가이드된 뉴클레아제와 표적 뉴클레오타이드 서열의 서열-특이적인 결합을 지시하는 뉴클레오타이드 서열을 지칭한다. 따라서, RGN의 각각의 가이드 RNA는 RGN에 결합하여 RGN이 특정한 표적 뉴클레오타이드 서열과 결합하도록 가이드하는 하나 이상의 RNA 분자 (일반적으로, 1개 또는 2개)이며, 이들 예에서 RGN은 니카제 또는 뉴클레아제 활성을 갖고, 또한 표적 뉴클레오타이드 서열을 절단한다. 일반적으로, 가이드 RNA는 CRISPR RNA (crRNA) 및 일부 실시양태에서 트랜스-활성화 CRISPR RNA (tracrRNA)를 포함한다. crRNA 및 tracrRNA 둘 다를 포함하는 본래의 가이드 RNA는 일반적으로 crRNA의 반복 서열 및 tracrRNA의 안티-반복 서열을 통해 서로 혼성화하는 2가지 별개의 RNA 분자를 포함한다.

[0037] CRISPR 어레이 내의 본래의 직접적인 반복 서열은 일반적으로 28 내지 37개 염기쌍의 길이 범위를 갖지만, 상기 길이는 약 23 bp 내지 약 55 bp로 달라질 수 있다. CRISPR 어레이 내의 스페이서 서열은 일반적으로 약 32 내지 약 38 bp의 길이 범위를 갖지만, 상기 길이는 약 21 bp 내지 약 72 bp일 수 있다. 각각의 CRISPR 어레이는 일반적으로 CRISPR 반복-스페이서 서열의 50개 미만의 단위를 포함한다. CRISPR은 일차 CRISPR 전사체로 명명되는 긴 전사체의 일부로서 전사되며, 이는 CRISPR 어레이의 대부분을 포함한다. 일차 CRISPR 전사체는 Cas 단백질에 의해 절단되어, crRNA를 생성하거나 또는 일부 경우에는 추가의 Cas 단백질에 의해 성숙 crRNA로 추가로 가공되는 프리-crRNA를 생성한다. 성숙 crRNA는 스페이서 서열 및 CRISPR 반복 서열을 포함한다. 프리-crRNA가 성숙 (또는 가공된) crRNA로 가공되는 일부 실시양태에서, 성숙은 약 1 내지 약 6개 또는 그 초과 5', 3', 또는 5' 및 3' 뉴클레오타이드의 제거를 수반한다. 특정한 관심 표적 뉴클레오타이드 서열의 게놈 편집 또는 표적화의 목적을 위해, 프리-crRNA 분자의 성숙 동안에 제거된 이들 뉴클레오타이드가 반드시 가이드 RNA를 생성하거나 또는 설계할 필요는 없다.

[0038] CRISPR RNA (crRNA)는 스페이서 서열 및 CRISPR 반복 서열을 포함한다. "스페이서 서열"은 관심 표적 뉴클레오타이드 서열과 직접적으로 혼성화하는 뉴클레오타이드 서열이다. 스페이서 서열은 관심 표적 서열에 의해 완전히 또는 부분적으로 상보성하도록 조작된다. 다양한 실시양태에서, 스페이서 서열은 약 8개 뉴클레오타이드 내지 약 30개 뉴클레오타이드, 또는 그 초과를 포함할 수 있다. 예를 들어, 스페이서 서열은 약 8, 약 9, 약 10, 약 11, 약 12, 약 13, 약 14, 약 15, 약 16, 약 17, 약 18, 약 19, 약 20, 약 21, 약 22, 약 23, 약 24, 약 25, 약 26, 약 27, 약 28, 약 29, 약 30개, 또는 그 초과 뉴클레오타이드 길이일 수 있다. 일부 실시양태에서, 스페이서 서열은 약 10 내지 약 26개 뉴클레오타이드 길이, 또는 약 12 내지 약 30개 뉴클레오타이드 길이이다. 특정한 실시양태에서, 스페이서 서열은 약 30개 뉴클레오타이드 길이이다. 일부 실시양태에서, 적합한 정렬 알고리즘을 이용하여 최적으로 정렬하였을 때 스페이서 서열과 그의 상응하는 표적 서열 사이의 상보성 정도는 약 50%, 약 60%, 약 70%, 약 75%, 약 80%, 약 81%, 약 82%, 약 83%, 약 84%, 약 85%, 약 86%, 약 87%, 약 88%, 약 89%, 약 90%, 약 91%, 약 92%, 약 93%, 약 94%, 약 95%, 약 96%, 약 97%, 약 98%, 약 99% 또는 그 초과이거나, 또는 대략 상기 값 초과이다. 특정한 실시양태에서, 스페이서 서열은 이차 구조를 함유하지 않으며, 이는 관련 기술분야에 공지된 임의의 적합한 폴리뉴클레오타이드 폴딩 알고리즘, 예컨대 비제한적으로 mFold (예를 들어, [Zuker and Stiegler (1981) Nucleic Acids Res. 9:133-148] 참고) 및 RNAfold (예를 들어, [Gruber et al. (2008) Cell 106(1):23-24] 참고)를 사용하여 예측될 수 있다.

[0039] CRISPR RNA 반복 서열은 tracrRNA에 혼성화하도록 충분한 상보성을 갖는 영역을 포함하는 뉴클레오타이드 서열을 포함한다. 다양한 실시양태에서, CRISPR RNA 반복 서열은 약 8개 뉴클레오타이드 내지 약 30개 뉴클레오타이드, 또는 그 초과를 포함할 수 있다. 예를 들어, CRISPR 반복 서열은 약 8, 약 9, 약 10, 약 11, 약 12, 약 13, 약 14, 약 15, 약 16, 약 17, 약 18, 약 19, 약 20, 약 21, 약 22, 약 23, 약 24, 약 25, 약 26, 약 27, 약 28, 약 29, 약 30개, 또는 그 초과 뉴클레오타이드 길이일 수 있다. 일부 실시양태에서, CRISPR 반복 서열은 약 21개 뉴클레오타이드 길이이다. 일부 실시양태에서, 적합한 정렬 알고리즘을 이용하여 최적으로 정렬될 때 CRISPR 반복 서열과 그의 상응하는 tracrRNA 서열 사이의 상보성 정도는 약 50%, 약 60%, 약 70%, 약 75%, 약 80%, 약 81%, 약 82%, 약 83%, 약 84%, 약 85%, 약 86%, 약 87%, 약 88%, 약 89%, 약 90%, 약 91%, 약 92%, 약 93%, 약 94%, 약 95%, 약 96%, 약 97%, 약 98%, 약 99% 또는 그 초과, 또는 대략 상기 값 초과이다. 특정한 실시양태에서, CRISPR 반복 서열은, 가이드 RNA 내에 포함될 때 본원에 제공된 연관된 RNA-가이드된 뉴클레아제와 관심 표적 서열의 서열-특이적인 결합을 지시할 수 있는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287의 뉴클레오타이드 서열, 또는 그의 활성 변이체 또는 단편

을 포함한다. 특정한 실시양태에서, 야생형 서열의 활성 CRISPR 반복 서열 변이체는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287로 제시된 뉴클레오타이드 서열에 대해 적어도 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과와 서열 동일성을 갖는 뉴클레오타이드 서열을 포함한다. 특정한 실시양태에서, 야생형 서열의 활성 CRISPR 반복 서열 단편은 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287로 제시된 뉴클레오타이드 서열의 적어도 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 연속 뉴클레오타이드를 포함한다.

[0040] 특정한 실시양태에서, crRNA는 자연 발생이 아니다. 이들 일부 실시양태에서, 특이적인 CRISPR 반복 서열은 본래 조작된 스페이스 서열에 연결되지 않고, CRISPR 반복 서열은 스페이스 서열에 대해 이중성인 것으로 고려된다. 특정한 실시양태에서, 스페이스 서열은 자연 발생이 아닌 조작된 서열이다.

[0041] 트랜스-활성화 CRISPR RNA 또는 tracrRNA 분자는 crRNA의 CRISPR 반복 서열에 혼성화하도록 충분한 상보성을 갖는 영역을 포함하는 뉴클레오타이드 서열을 포함하며, 이는 본원에서 안티-반복 영역으로 지칭된다. 일부 실시양태에서, tracrRNA 분자는 이차 구조 (예를 들어, 스템-루프)를 갖는 영역을 추가로 포함하거나, 또는 그의 상응하는 crRNA와 혼성화시 이차 구조를 형성한다. 특정한 실시양태에서, CRISPR 반복 서열과 완전히 또는 부분적으로 상보성인 tracrRNA의 영역은 분자의 5' 말단에 있고, tracrRNA의 3' 말단은 이차 구조를 포함한다. 이러한 이차 구조 영역은 일반적으로 안티-반복 서열과 인접하여 발견되는 몇몇 헤어핀 구조, 예컨대 넥서스 헤어핀을 포함한다. 넥서스 헤어핀은 종종 tracrRNA의 많은 넥서스 헤어핀에서 발견되는 모티프 UNANNA, CNANNG, CNANNU, UNANNG, UNANNC, 또는 CNANNU (각각 서열식별번호: 8, 37, 45, 53, 68, 및 102)와 함께 헤어핀 스템의 염기에서 보존된 뉴클레오타이드 서열을 갖는다. 종종 tracrRNA의 3' 말단에 구조 및 개수가 다양할 수 있지만 종종 GC-풍부 Rho-비의존성 전사 종결인자 헤어핀에 이어서 3' 말단에 U의 문자열을 포함하는 말단 헤어핀이 있다. 예를 들어, [Briner et al. (2014) Molecular Cell 56:333-339, Briner and Barrangou (2016) Cold Spring Harb Protoc; doi: 10.1101/pdb.top090902], 및 미국 공개 번호 2017/0275648을 참고하며, 이들 각각은 그의 전문이 본원에 참조로 포함된다.

[0042] 다양한 실시양태에서, CRISPR 반복 서열에 대해 완전히 또는 부분적으로 상보성인 tracrRNA의 안티-반복 영역은 약 8개 뉴클레오타이드 내지 약 30개 뉴클레오타이드, 또는 그 초과를 포함한다. 예를 들어, tracrRNA 안티-반복 서열과 CRISPR 반복 서열 사이의 염기쌍 영역은 약 8, 약 9, 약 10, 약 11, 약 12, 약 13, 약 14, 약 15, 약 16, 약 17, 약 18, 약 19, 약 20, 약 21, 약 22, 약 23, 약 24, 약 25, 약 26, 약 27, 약 28, 약 29, 약 30개, 또는 그 초과와 뉴클레오타이드 길이일 수 있다. 특정한 실시양태에서, CRISPR 반복 서열에 대해 완전히 또는 부분적으로 상보성인 tracrRNA의 안티-반복 영역은 약 20개 뉴클레오타이드 길이이다. 일부 실시양태에서, 적합한 정렬 알고리즘을 이용하여 최적으로 정렬될 때 CRISPR 반복 서열과 그의 상응하는 tracrRNA 안티-반복 서열 사이의 상보성 정도는 약 50%, 약 60%, 약 70%, 약 75%, 약 80%, 약 81%, 약 82%, 약 83%, 약 84%, 약 85%, 약 86%, 약 87%, 약 88%, 약 89%, 약 90%, 약 91%, 약 92%, 약 93%, 약 94%, 약 95%, 약 96%, 약 97%, 약 98%, 약 99% 또는 그 초과, 또는 대략 상기 값 초과이다.

[0043] 다양한 실시양태에서, 전체 tracrRNA는 약 60개 뉴클레오타이드 내지 약 140개 초과와 뉴클레오타이드를 포함할 수 있다. 예를 들어, tracrRNA는 약 60, 약 65, 약 70, 약 75, 약 80, 약 85, 약 90, 약 95, 약 100, 약 105, 약 110, 약 115, 약 120, 약 125, 약 130, 약 135, 약 140개 또는 그 초과와 뉴클레오타이드 길이일 수 있다. 특정한 실시양태에서, tracrRNA는 약 80 내지 약 90개 뉴클레오타이드 길이, 예컨대 약 80, 약 81, 약 82, 약 83, 약 84, 약 85, 약 86, 약 87, 약 88, 약 89, 및 약 90개 뉴클레오타이드 길이이다. 특정한 실시양태에서, tracrRNA는 약 85개 뉴클레오타이드 길이이다.

[0044] 특정한 실시양태에서, tracrRNA는, 가이드 RNA 내에 포함될 때 본원에 제공된 연관된 RNA-가이드된 뉴클레아제와 관심 표적 서열의 서열-특이적인 결합을 지시할 수 있는 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 또는 119의 뉴클레오타이드 서열, 또는 그의 활성 변이체 또는 단편을 포함한다. 특정한 실시양태에서, 야생형 서열의 활성 tracrRNA 서열 변이체는 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274, 또는 286으로 제시된 뉴클레오타이드 서열에 대해 적어도 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과와 서열 동일성을 갖는 뉴클레오타이드 서열을 포함한다. 특정한 실시양태에서, 야생형 서열의 활성 tracrRNA 서열 단편은 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274, 또는 286으로 제시된 뉴클레오타이드 서열의 적어도 5, 10, 15, 20, 25, 30, 35, 40,

45, 50, 55, 60, 65, 70, 75, 80개 또는 그 초과인 연속 뉴클레오타이드를 포함한다.

[0045] 2개의 폴리뉴클레오타이드 서열은 2개의 서열이 엄격도 조건 하에 서로 혼성화할 때 실질적으로 상보성인 것으로 고려된다. 마찬가지로, RGN은 RGN에 결합된 가이드 RNA가 엄격도 조건 하에 표적 서열에 결합하는 경우 서열-특이적인 방식 내에서 특정한 표적 서열에 결합하는 것으로 고려된다. "엄격도 조건" 또는 "엄격도 혼성화 조건"은 2개의 폴리뉴클레오타이드 서열이 다른 서열에 비해 검출가능하게 더 큰 정도로 (예를 들어, 백그라운드에 비해 적어도 2배) 서로 혼성화하도록 하는 의도된 조건이다. 엄격도 조건은 서열-의존성이고, 상이한 환경에서 상이할 것이다. 전형적으로, 엄격도 조건은 염 농도가 pH 7.0 내지 8.3에서 약 1.5 M 미만의 Na 이온, 전형적으로 약 0.01 내지 1.0 M Na 이온 농도 (또는 다른 염)이고 온도가 짧은 서열 (예를 들어, 10 내지 50개 뉴클레오타이드)의 경우 적어도 약 30℃ 및 긴 서열 (예를 들어, 50개 초과인 뉴클레오타이드)의 경우 적어도 약 60℃인 조건일 것이다. 엄격도 조건은 또한 탈안정화제, 예컨대 포름아미드의 첨가에 의해 달성될 수 있다. 예시적인 저엄격도 조건에는 37℃에서 30 내지 35% 포름아미드, 1 M NaCl, 1% SDS (나트륨 도데실 황산염)의 완충제 용액에 의한 혼성화, 및 50 내지 55℃에서 1X 내지 2X SSC (20X SSC = 3.0 M NaCl/0.3 M 시트르산나트륨) 중에서 세척이 포함된다. 예시적인 중간 엄격도 조건에는 37℃에서 40 내지 45% 포름아미드, 1.0 M NaCl, 1% SDS 중에서 혼성화, 및 55 내지 60℃에서 0.5X 내지 1X SSC 중에서 세척이 포함된다. 예시적인 고엄격도 조건에는 37℃에서 50% 포름아미드, 1 M NaCl, 1% SDS 중에서 혼성화, 및 60 내지 65℃에서 0.1X SSC 중에서 세척이 포함된다. 임의적으로, 세척 완충제는 약 0.1% 내지 약 1% SDS를 포함할 수 있다. 혼성화 지속시간은 일반적으로 약 24 시간 미만, 보통 약 4 내지 약 12 시간이다. 세척 지속시간은 적어도 평형에 도달하기에 충분한 시간일 것이다.

[0046] T_m은 (정의된 이온 강도 및 pH하에) 상보성 표적 서열의 50%가 매칭된 서열에 완벽하게 혼성화하는 온도이다. DNA-DNA 혼성체의 경우, T_m은 [Meinkoth and Wahl (1984) Anal. Biochem. 138:267-284]의 방정식으로부터 계산될 수 있으며: $T_m = 81.5^{\circ}\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$; 여기서 M은 일가 양이온의 몰 농도이고, %GC는 DNA에서 구아노신 및 시토신 뉴클레오타이드의 백분율이고, % form은 혼성화 용액 중에서 포름아미드의 백분율이고, L은 염기쌍에서 혼성체의 길이이다. 일반적으로, 엄격도 조건은 정의된 이온 강도 및 pH에서 특이적인 서열 및 그의 상보체에 대한 열 용점 (T_m)보다 약 5℃ 낮도록 선택된다. 그러나, 매우 엄격도 조건은 열 용점 (T_m)보다 1, 2, 3 또는 4℃ 낮은 온도에서 혼성화 및/또는 세척을 이용할 수 있고; 중간 엄격도 조건은 열 용점 (T_m)보다 6, 7, 8, 9 또는 10℃ 낮은 온도에서 혼성화 및/또는 세척을 이용할 수 있고; 저엄격도 조건은 열 용점 (T_m)보다 11, 12, 13, 14, 15 또는 20℃ 낮은 온도에서 혼성화 및/또는 세척을 이용할 수 있다. 상기 방정식, 혼성화 및 세척 조성물, 및 원하는 T_m을 이용하여, 통상의 기술자는 혼성화 및/또는 세척 용액의 엄격도에서의 변동이 본질적으로 기재되어 있음을 이해할 것이다. 핵산의 혼성화에 대한 광범위한 안내는 [Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, Part I, Chapter 2 (Elsevier, New York); and Ausubel et al., eds. (1995) Current Protocols in Molecular Biology, Chapter 2 (Greene Publishing and Wiley-Interscience, New York)]에서 확인한다. [Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York)]을 참고한다.

[0047] 가이드 RNA는 단일 가이드 RNA 또는 이중-가이드 RNA 시스템일 수 있다. 단일 가이드 RNA는 RNA의 단일 분자 상에 crRNA 및 tracrRNA를 포함하는 반면에, 이중-가이드 RNA 시스템은 2가지 별개의 RNA 분자 상에 존재하는 crRNA 및 tracrRNA를 포함하며, 이들은 crRNA의 CRISPR 반복 서열의 적어도 일부분, 및 crRNA의 CRISPR 반복 서열에 대해 완전히 또는 부분적으로 상보성일 수 있는 tracrRNA의 적어도 일부분을 통해 서로 혼성화된다. 가이드 RNA가 단일 가이드 RNA인 이들 일부 실시양태에서, crRNA 및 tracrRNA는 링커 뉴클레오타이드 서열에 의해 분리된다. 일반적으로, 링커 뉴클레오타이드 서열은 링커 뉴클레오타이드 서열의 뉴클레오타이드 내에서 또는 이를 포함하는 이차 구조의 형성을 피하기 위해 상보성 염기를 포함하지 않는 것이다. 일부 실시양태에서, crRNA와 tracrRNA 사이의 링커 뉴클레오타이드 서열은 적어도 3, 적어도 4, 적어도 5, 적어도 6, 적어도 7, 적어도 8, 적어도 9, 적어도 10, 적어도 11, 적어도 12개 또는 그 초과인 뉴클레오타이드 길이이다. 특정한 실시양태에서, 단일 가이드 RNA의 링커 뉴클레오타이드 서열은 적어도 4개 뉴클레오타이드 길이이다. 특정한 실시양태에서, 링커 뉴클레오타이드 서열은 서열식별번호: 123으로 제시된 뉴클레오타이드 서열이다.

[0048] 단일 가이드 RNA 또는 이중-가이드 RNA는 화학적으로 또는 시험관내 전사를 통해 합성될 수 있다. RGN과 가이드 RNA 사이의 서열-특이적인 결합을 결정하기 위한 검정은 관련 기술분야에 공지되어 있고, 발현된 RGN과 가이드 RNA 사이의 시험관내 결합 검정이 포함되나 이로 제한되지 않으며, 이는 검출가능한 표지 (예를 들어, 비오틴)에 의해 태그 부착될 수 있고, 검출가능한 표지를 통해 (예를 들어, 스트렙타비딘 비드에 의해) 가이드

RNA:RGN 복합체를 포획하는 풀다운(pull-down) 검출 검정에서 이용될 수 있다. 가이드 RNA에 대해 관련이 없는 서열 또는 구조를 갖는 대조군 가이드 RNA는 RGN과 RNA의 비-특이적인 결합에 대한 음성 대조군으로서 사용될 수 있다. 특정한 실시양태에서, 가이드 RNA는 서열식별번호: 4, 12, 19, 26, 33, 41, 49, 57, 64, 72, 78, 85, 92, 98, 106, 113, 또는 120이고, 스페이서 서열은 임의의 서열일 수 있고 폴리-N 서열로 표시된다.

[0049] 특정한 실시양태에서, 가이드 RNA는 RNA 분자로서 표적 세포, 소기관 또는 배아에 도입될 수 있다. 가이드 RNA는 시험관내에서 전사될 수 있거나 또는 화학적으로 합성될 수 있다. 다른 실시양태에서, 가이드 RNA를 코딩하는 뉴클레오타이드 서열은 세포, 소기관 또는 배아에 도입된다. 이들 일부 실시양태에서, 가이드 RNA를 코딩하는 뉴클레오타이드 서열은 프로모터 (예를 들어, RNA 폴리머라제 III 프로모터)에 작동가능하게 연결된다. 프로모터는 본래의 프로모터일 수 있거나 또는 가이드 RNA-코딩 뉴클레오타이드 서열에 대해 이중성일 수 있다.

[0050] 다양한 실시양태에서, 가이드 RNA는 본원에 기재된 바와 같이 리보핵단백질 복합체로서 표적 세포, 소기관 또는 배아에 도입될 수 있으며, 가이드 RNA는 RNA-가이드된 뉴클레아제 폴리펩티드에 결합된다.

[0051] 가이드 RNA는 가이드 RNA와 표적 뉴클레오타이드 서열의 혼성화를 통해 연관된 RNA-가이드된 뉴클레아제를 특정한 관심 표적 뉴클레오타이드 서열로 지시한다. 표적 뉴클레오타이드 서열은 DNA, RNA, 또는 이들 둘의 조합물을 포함할 수 있고, 단일-가닥 또는 이중-가닥일 수 있다. 표적 뉴클레오타이드 서열은 게놈 DNA (즉, 염색체 DNA), 플라스미드 DNA, 또는 RNA 분자 (예를 들어, 메신저 RNA, 리보솜 RNA, 전달 RNA, 마이크로 RNA, 소형 간섭 RNA) 일 수 있다. 표적 뉴클레오타이드 서열은 시험관내에서 또는 세포에서 RNA-가이드된 뉴클레아제에 의해 결합될 수 있다 (그리고 일부 실시양태에서 절단될 수 있다). RGN에 의해 표적화된 염색체 서열은 핵, 색소체 또는 미토콘드리아 염색체 서열일 수 있다. 일부 실시양태에서, 표적 뉴클레오타이드 서열은 표적 게놈에서 고유하다.

[0052] 표적 뉴클레오타이드 서열은 프로토스페이서 인접한 모티프 (PAM)에 인접해 있다. 프로토스페이서 인접한 모티프는 일반적으로 표적 뉴클레오타이드 서열로부터 약 1 내지 약 10개 뉴클레오타이드 이내에, 예컨대 표적 뉴클레오타이드 서열로부터 약 1, 약 2, 약 3, 약 4, 약 5, 약 6, 약 7, 약 8, 약 9 또는 약 10개 뉴클레오타이드 이내에 있다. PAM은 표적 서열의 5' 또는 3'일 수 있다. 일부 실시양태에서, PAM은 본원에 개시된 RGN에 대한 표적 서열의 3'이다. 일반적으로, PAM은 약 3-4개 뉴클레오타이드의 컨센서스 서열이지만, 특정한 실시양태에서 2, 3, 4, 5, 6, 7, 8, 9개 또는 그 초과 뉴클레오타이드 길이일 수 있다. 다양한 실시양태에서, 본원에 개시된 RGN에 의해 인식되는 PAM 서열은 서열식별번호: 7, 15, 22, 29, 36, 44, 52, 60, 67, 81, 88, 101, 109, 또는 116으로 제시된 컨센서스 서열을 포함한다.

[0053] 특정 실시양태에서, 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117을 갖는 RNA-가이드된 뉴클레아제 또는 그의 활성 변이체 또는 단편은 각각 서열식별번호: 7, 15, 22, 29, 36, 44, 52, 60, 67, 81, 88, 101, 109 또는 116으로 제시된 PAM 서열에 인접한 표적 뉴클레오타이드 서열에 결합한다. 일부 실시양태에서, 서열식별번호: 54 또는 137을 갖는 RNA-가이드된 뉴클레아제 또는 그의 활성 변이체 또는 단편은 서열식별번호: 147로 기재된 PAM 서열에 인접한 표적 뉴클레오타이드 서열에 결합한다. 이들 실시양태 중 일부에서, RGN은 각각 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111 또는 118에 제시된 CRISPR 반복 서열 또는 그의 활성 변이체 또는 단편, 및 각각 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 또는 119에 제시된 tracrRNA 서열 또는 그의 활성 변이체 또는 단편을 포함하는 가이드 서열에 결합한다. RGN 시스템은 본 명세서의 실시예 1 및 표 1 및 2에 추가로 기재된다.

[0054] 주어진 뉴클레아제 효소에 대한 PAM 서열 특이성이 효소 농도에 의해 영향을 받는다는 것은 관련 기술분야에 널리 공지되어 있고 (예를 들어, [Karvelis et al. (2015) Genome Biol 16:253] 참고), 이는 RGN을 발현하기 위해 사용된 프로모터, 또는 세포, 소기관 또는 배아에 전달된 리보핵단백질 복합체의 양을 변경시킴으로써 변형될 수 있다.

[0055] 그의 상응하는 PAM 서열을 인식할 때, RGN은 특이적인 절단 부위에서 표적 뉴클레오타이드 서열을 절단할 수 있다. 본원에서 사용된 바와 같이, 절단 부위는 표적 뉴클레오타이드 서열 내의 2개의 특정한 뉴클레오타이드로 구성되며, 이들 사이에서 뉴클레오타이드 서열이 RGN에 의해 절단된다. 절단 부위는 5' 또는 3' 방향으로 PAM으로부터 1번째 및 2번째, 2번째 및 3번째, 3번째 및 4번째, 4번째 및 5번째, 5번째 및 6번째, 6번째 및 7번째, 7번째 및 8번째, 또는 8번째 및 9번째 뉴클레오타이드를 포함할 수 있다. 일부 실시양태에서, 절단 부위는 5' 또는 3' 방향으로 PAM으로부터 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 뉴클레오타이드에 걸쳐 있을 수 있다. 일부 실시양태에서, 절단 부위는 PAM으로부터 4개 뉴클레오타이드 떨어져 있다. 다른 실시양태에서, 절단 부위는 PAM으로부터 적어도 15개 뉴클레오타이드 떨어져 있다. RGN이 표적 뉴클레오타이드 서열을 절단하여 엇갈린 말단을 생성할

수 있기 때문에, 일부 실시양태에서, 절단 부위는 폴리뉴클레오티드의 양성 (+) 가닥 상의 PAM으로부터 2개 뉴클레오티드의 거리 및 폴리뉴클레오티드의 음성 (-) 가닥 상의 PAM으로부터 2개 뉴클레오티드의 거리를 기반으로 하여 정의된다.

[0056] PAM은 유형 II CRISPR 시스템의 RNA-가이드된 뉴클레아제의 특징으로서 간주된다 ([Szczelkun et al., PNAS, 111: 9798-9803, 2014; Sternberg et al., Nature 507: 62-67, 2014]). 흥미롭게도, APG06646.1 및 APG04293.1은 RNA-가이드된 뉴클레아제로서 기능하고 유형 II CRISPR Cas9 뉴클레아제와 동일한 도메인 중 다수를 공유하지만, 이들은 각각 전형적인 PAM-상호작용 도메인 (PID; 인터프로: IPR032237; Pfam: PF16595)이 결여되어 있다. 따라서, APG06646.1 및 APG04293.1은 또한 상기 기재된 바와 같은 2-5개 뉴클레오티드의 모티프인 전형적인 PAM 요건을 보유하지 않는다. 대신에, 이들 단백질은 그의 C-말단에 고유한 DNA 인식 도메인 (APG06646.1의 잔기 821-1092 (전장 서열은 서열식별번호: 117임); APG04293.1의 잔기 1064-1401 (전장 서열은 서열식별번호: 103임))을 갖는다. 이들 고유한 DNA 인식 도메인은 뉴클레아제가 게놈 표적 서열 부근의 단일-뉴클레오티드 모티프 (서열식별번호: 109; 표 2 참조)에 기초하여 게놈 표적 부위에서 절단할 수 있게 한다.

[0057] APG04293.1은 또한 그의 N-말단에 근접한 133개 아미노산 잔기 (잔기 144-276)의 고유한 시그너처 도메인을 공유한다. 이러한 도메인의 기능은 유형 II CRISPR Cas9 뉴클레아제 또는 일반적으로 관련 기술분야에 공지되어 있지 않다.

[0058] III. RNA-가이드된 뉴클레아제, CRISPR RNA, 및/또는 tracrRNA를 코딩하는 뉴클레오티드

[0059] 본 개시내용은 본원에 개시된 CRISPR RNA, tracrRNA 및/또는 sgRNA를 포함하는 폴리뉴클레오티드 및 본원에 개시된 RNA-가이드된 뉴클레아제, CRISPR RNA, tracrRNA 및/또는 sgRNA를 코딩하는 뉴클레오티드 서열을 포함하는 폴리뉴클레오티드를 제공한다. 본원에 개시된 폴리뉴클레오티드는, 가이드 RNA 내에 포함되는 경우 연관된 RNA-가이드된 뉴클레아제의 관심 표적 서열에 대한 서열-특이적 결합을 지시할 수 있는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 또는 287의 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열을 포함하거나 코딩하는 폴리뉴클레오티드, 또는 그의 활성 변이체 또는 단편을 포함한다. 또한, 가이드 RNA 내에 포함되는 경우 연관된 RNA-가이드된 뉴클레아제의 관심 표적 서열에 대한 서열-특이적 결합을 지시할 수 있는 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 또는 286의 뉴클레오티드 서열을 포함하는 tracrRNA를 포함하거나 코딩하는 폴리뉴클레오티드, 또는 그의 활성 변이체 또는 단편이 개시된다. RNA-가이드된 서열-특이적 방식으로 표적 뉴클레오티드 서열에 결합하는 능력을 보유하는 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 또는 235로 제시된 아미노산 서열을 포함하는 RNA-가이드된 뉴클레아제를 코딩하는 폴리뉴클레오티드, 및 그의 활성 단편 또는 변이체가 또한 제공된다.

[0060] 용어 "폴리뉴클레오티드"의 사용은 본 개시내용을 DNA를 포함하는 폴리뉴클레오티드로 제한하는 것으로 의도되지 않는다. 관련 기술분야의 통상의 기술자는 폴리뉴클레오티드가 리보뉴클레오티드 (RNA), 및 리보뉴클레오티드 및 데옥시리보뉴클레오티드의 조합물을 포함할 수 있음을 인지할 것이다. 이러한 데옥시리보뉴클레오티드 및 리보뉴클레오티드에는 자연 발생 분자 및 합성 유사체 둘 다 포함된다. 이들에는 펩티드 핵산 (PNA), PNA-DNA 키메라, 록킹된 핵산 (LNA), 및 포스포티오레이트 연결된 서열이 포함된다. 본원에 개시된 폴리뉴클레오티드는 또한 모든 형태의 서열, 예컨대 비제한적으로 단일-가닥 형태, 이중-가닥 형태, DNA-RNA 혼성체, 삼중 구조, 스텝-루프 구조 등을 포괄한다.

[0061] RGN을 코딩하는 핵산 분자는 관심 유기체에서의 발현을 위해 코돈 최적화될 수 있다. "코돈-최적화된" 코딩 서열은 특정한 숙주 세포의 선호된 코돈 사용 빈도 또는 전사 조건을 모방하도록 설계된 그의 코돈 사용 빈도를 갖는 폴리뉴클레오티드 코딩 서열이다. 특정한 숙주 세포 또는 유기체에서의 발현은 번역된 아미노산 서열이 변하지 않도록 핵산 수준에서 하나 이상의 코돈 변경의 결과로서 증강된다. 핵산 분자는 전체적으로 또는 부분적으로 코돈 최적화될 수 있다. 광범위한 유기체에 대한 선호도 정보를 제공하는 코돈 표 및 다른 참고자료가 관련 기술분야에서 이용가능하다 (예를 들어, 식물-선호 코돈 용법의 논의에 대해서는 [Campbell and Gowri (1990) Plant Physiol. 92:1-11] 참고). 식물-선호 유전자를 합성하는 방법은 관련 기술분야에서 이용가능하다. 예를 들어, 미국 특허 번호 5,380,831, 및 5,436,391, 및 [Murray et al. (1989) Nucleic Acids Res. 17:477-498]을 참고하며, 이들은 본원에 참조로 포함된다.

[0062] 본원에 제공된 RGN, crRNA, tracrRNA 및/또는 sgRNA를 코딩하는 폴리뉴클레오티드는 시험관내 발현 또는 관심 세포, 소기관, 배아 또는 유기체에서의 발현을 위해 발현 카세트에 제공될 수 있다. 카세트는 폴리뉴클레오티드의 발현을 가능하게 하는 본원에 제공된 RGN, crRNA, tracrRNA 및/또는 sgRNA를 코딩하는 폴리뉴클레오티드에

작동가능하게 연결된 5' 및 3' 조절 서열을 포함할 것이다. 카세트는 유기체에 동시 형질전환되는 적어도 하나의 추가의 유전자 또는 유전자 요소를 추가로 함유할 수 있다. 추가의 유전자 또는 요소가 포함되는 경우, 성분들은 작동가능하게 연결된다. 용어 "작동가능하게 연결된"은 2가지 이상의 요소들 사이의 기능적 연결을 의미하는 것으로 의도된다. 예를 들어, 프로모터와 관심 코딩 영역 (예를 들어, RGN, crRNA, tracrRNA 및/또는 sgRNA를 코딩하는 영역) 사이의 연결은 관심 코딩 영역의 발현을 가능하게 하는 기능적 연결이다. 작동가능하게 연결된 요소는 연속적 또는 비-연속적일 수 있다. 2가지 단백질 코딩 영역의 결합을 지칭하기 위해 사용되는 경우, 작동가능하게 연결된이란 코딩 영역이 동일한 리딩 프레임 내에 있는 것으로 의도된다. 대안적으로, 추가의 유전자(들) 또는 요소(들)가 다중 발현 카세트 상에 제공될 수 있다. 예를 들어, 본원에 개시된 RGN을 코딩하는 뉴클레오타이드 서열은 하나의 발현 카세트 상에 존재할 수 있는 반면에, crRNA, tracrRNA 또는 완전 가이드 RNA를 코딩하는 뉴클레오타이드 서열은 별도의 발현 카세트 상에 존재할 수 있다. 이러한 발현 카세트에는 조절 영역의 전사 조절하에 있는 폴리뉴클레오타이드의 삽입을 위한 다수개의 제한 부위 및/또는 재조합 부위가 제공된다. 발현 카세트는 선택가능한 마커 유전자를 추가로 함유할 수 있다.

[0063] 발현 카세트는 5'-3' 전사 방향으로 관심 유기체에서 기능적인 전사 (일부 실시양태에서는 번역) 개시 영역 (즉, 프로모터), 본 발명의 RGN-, crRNA-, tracrRNA- 및/또는 sgRNA-코딩 폴리뉴클레오타이드, 및 전사 (일부 실시양태에서는 번역) 종결 영역 (즉, 종결 영역)을 포함할 것이다. 본 발명의 프로모터는 숙주 세포에서 코딩 서열의 발현을 지지하거나 또는 유도할 수 있다. 조절 영역 (예를 들어, 프로모터, 전사 조절 영역, 및 번역 종결 영역)은 숙주 세포에 대해 또는 서로에 대해 내인성 또는 이종성일 수 있다. 본원에서 사용된 바와 같이, 서열과 관련하여 "이종성"은 외래 종으로부터 기원하는 서열이거나, 또는 동일한 종으로부터 기원하는 경우에는 의도적인 인간 개입에 의해 조성 및/또는 게놈 유전자좌에서 그의 본래의 형태로부터 실질적으로 변형된 서열이다. 본원에서 사용된 바와 같이, 키메라 유전자는 코딩 서열에 대해 이종성인 전사 개시 영역에 작동가능하게 연결된 코딩 서열을 포함한다.

[0064] 편리한 종결 영역은 에이. 투메파시엔스(*A. tumefaciens*)의 Ti-플라스미드, 예컨대 옥토펜 신타제 및 노팔린 신타제 종결 영역으로부터 입수가능하다. [Guerineau et al. (1991) Mol. Gen. Genet. 262:141-144; Proudfoot (1991) Cell 64:671-674; Sanfacon et al. (1991) Genes Dev. 5:141-149; Mogen et al. (1990) Plant Cell 2:1261-1272; Munroe et al. (1990) Gene 91:151-158; Ballas et al. (1989) Nucleic Acids Res. 17:7891-7903; and Joshi et al. (1987) Nucleic Acids Res. 15:9627-9639] 또한 참고한다.

[0065] 추가의 조절 신호에는 전사 개시 시작 부위, 오퍼레이터, 활성화제, 인핸서, 다른 조절 요소, 리보솜 결합 부위, 개시 코돈, 종결 신호 등이 포함되나 이로 제한되지 않는다. 예를 들어, 미국 특허 번호 5,039,523 및 4,853,331; EPO 0480762A2; [Sambrook et al. (1992) Molecular Cloning: A Laboratory Manual, ed. Maniatis et al. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.)], 이후 "Sambrook 11"; [Davis et al., eds. (1980) Advanced Bacterial Genetics (Cold Spring Harbor Laboratory Press), Cold Spring Harbor, N.Y.], 및 여기에 인용된 참고문헌을 참고한다.

[0066] 발현 카세트를 제조하기 위해, 적절한 배향으로, 적절한 경우 적절한 리딩 프레임으로 DNA 서열을 제공하도록 다양한 DNA 단편을 조작할 수 있다. 이를 위해, 어댑터 또는 링커를 사용하여 DNA 단편을 결합시킬 수 있거나, 또는 편리한 제한 부위, 불필요한 DNA의 제거, 제한 부위의 제거 등을 제공하기 위해 다른 조작이 수반될 수 있다. 이를 위해, 시험관내 돌연변이유발, 프라이머 복구, 제한, 어닐링, 재치환, 예를 들어 전이 및 전환이 수반될 수 있다.

[0067] 수많은 프로모터가 본 발명의 실시예에 사용될 수 있다. 프로모터는 원하는 결과를 기준으로 선택될 수 있다. 핵산은 구성적, 유도성, 성장 단계-특이적, 세포 유형-특이적, 조직-선택, 조직-특이적, 또는 관심 유기체에서의 발현을 위한 다른 프로모터와 조합될 수 있다. 예를 들어, WO 99/43838, 및 미국 특허 번호 8,575,425; 7,790,846; 8,147,856; 8,586,832; 7,772,369; 7,534,939; 6,072,050; 5,659,026; 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; 및 6,177,611에 제시된 프로모터를 참고하며; 이들은 본원에 참조로 포함된다.

[0068] 식물에서의 발현을 위해, 구성적 프로모터에는 CaMV 35S 프로모터 ([Ode11 et al. (1985) Nature 313:810-812]); 쌀 액틴 ([McElroy et al. (1990) Plant Cell 2:163-171]); 유비퀴틴 ([Christensen et al. (1989) Plant Mol. Biol. 12:619-632] 및 [Christensen et al. (1992) Plant Mol. Biol. 18:675-689]); pEMU ([Last et al. (1991) Theor. Appl. Genet. 81:581-588]); 및 MAS ([Velten et al. (1984) EMBO J. 3:2723-2730]) 또한 포함된다.

- [0069] 유도성 프로모터의 예는 저산소 또는 저온 스트레스에 의해 유도성인 Adh1 프로모터, 열 스트레스에 의해 유도성인 Hsp70 프로모터, 둘 모두 빛에 의해 유도성인 PPKK 프로모터 및 펙카르복실라제 프로모터이다. 화학적으로 유도성인 프로모터, 예컨대 완화제 유도된 In2-2 프로모터 (미국 특허 번호 5,364,780), 옥신 유도되고 용단층 특이적이지만 캘러스에서도 활성인 AxiG1 프로모터 (PCT US01/22169), 스테로이드-반응성 프로모터 (예를 들어, 에스트로겐 유도된 ERE 프로모터, 및 [Schena et al. (1991) Proc. Natl. Acad. Sci. USA 88:10421-10425] 및 [McNeill et al. (1998) Plant J. 14(2):247-257]에서의 글루코코르티코이드-유도성 프로모터 참고) 및 테트라시클린-유도성 및 테트라시클린-억제성 프로모터 (예를 들어, [Gatz et al. (1991) Mol. Gen. Genet. 227:229-237], 및 미국 특허 번호 5,814,618 및 5,789,156 참고) 또한 유용하며, 이들은 본원에 참조로 포함된다.
- [0070] 조직-특이적 또는 조직-선택 프로모터는 특정한 조직 내에서 발현 구축물의 표적 발현을 위해 사용될 수 있다. 특정한 실시양태에서, 조직-특이적 또는 조직-선택 프로모터는 식물 조직에서 활성이다. 식물에서 발달 조절하에 있는 프로모터의 예에는 특정 조직, 예컨대 잎, 뿌리, 과실, 종자 또는 꽃에서 전사를 우선적으로 개시하는 프로모터가 포함된다. "조직 특이적" 프로모터는 특정 조직에서만 전사를 개시하는 프로모터이다. 유전자의 구성적 발현과는 달리, 조직-특이적인 발현은 유전자 조절의 여러 상호작용 수준의 결과이다. 따라서, 상동성인 또는 밀접하게 관련된 식물 중으로부터의 프로모터가 특정한 조직에서 트랜스진의 효율적이고 신뢰할 수 있는 발현을 달성하기 위해 바람직하게 사용될 수 있다. 일부 실시양태에서, 발현은 조직-선택 프로모터를 포함한다. "조직 선택" 프로모터는 특정 조직에서 전사를 우선적으로 개시하지만 반드시 전체적으로 또는 단독으로 개시하는 것은 아닌 프로모터이다.
- [0071] 일부 실시양태에서, RGN, crRNA 및/또는 tracrRNA를 코딩하는 핵산 분자는 세포 유형-특이적인 프로모터를 포함한다. "세포 유형 특이적인" 프로모터는 하나 이상의 장기의 특정 세포 유형에서 주로 발현을 유도하는 프로모터이다. 식물에서 기능성인 세포 유형 특이적인 프로모터가 주로 활성일 수 있는 식물 세포의 일부 예에는 예를 들어 BETL 세포, 뿌리, 잎에서의 맥관 세포, 자루 세포, 및 줄기 세포가 포함된다. 핵산 분자는 또한 세포 유형 선택 프로모터를 포함할 수 있다. "세포 유형 선택" 프로모터는 하나 이상의 장기의 특정 세포 유형에서 주로 대부분의 발현을 유도하지만 반드시 전체적으로 또는 단독으로 유도하는 것은 아닌 프로모터이다. 식물에서 기능성인 세포 유형 선택 프로모터가 우선적으로 활성일 수 있는 식물 세포의 일부 예에는 예를 들어 BETL 세포, 뿌리, 잎에서의 맥관 세포, 자루 세포, 및 줄기 세포가 포함된다.
- [0072] RGN, crRNA, tracrRNA 및/또는 sgRNA를 코딩하는 핵산 서열은 예를 들어 시험관내 mRNA 합성을 위해 파지 RNA 폴리머라제에 의해 인식되는 프로모터 서열에 작동가능하게 연결될 수 있다. 이러한 실시양태에서, 시험관내-전사된 RNA는 본원에 기재된 방법에서 사용하기 위해 정제될 수 있다. 예를 들어, 프로모터 서열은 T7, T3 또는 SP6 프로모터 서열, 또는 T7, T3 또는 SP6 프로모터 서열의 변형일 수 있다. 이러한 실시양태에서, 발현된 단백질 및/또는 RNA는 본원에 기재된 게놈 변형 방법에서 사용하기 위해 정제될 수 있다.
- [0073] 특정한 실시양태에서, RGN, crRNA, tracrRNA 및/또는 sgRNA를 코딩하는 폴리뉴클레오티드는 또한 폴리아데닐화 신호 (예를 들어, SV40 폴리A 신호 및 식물에서 기능성인 다른 신호) 및/또는 적어도 하나의 전사 종결 서열에 연결될 수 있다. 추가로, RGN을 코딩하는 서열은 또한 본원에 기재된 바와 같이 적어도 하나의 핵 국제화 신호, 적어도 하나의 세포-투과 도메인, 및/또는 단백질을 특정한 세포로 위치로 수송할 수 있는 적어도 하나의 신호 펩티드를 코딩하는 서열(들)에 연결될 수 있다.
- [0074] RGN, crRNA, tracrRNA 및/또는 sgRNA를 코딩하는 폴리뉴클레오티드는 벡터 또는 다중 벡터에 존재할 수 있다. "벡터"는 핵산을 숙주 세포로 수송, 전달 또는 도입하기 위한 폴리뉴클레오티드 조성물을 지칭한다. 적합한 벡터에는 플라스미드 벡터, 파지미드, 코스미드, 인공/미니-염색체, 트랜스포손, 및 바이러스 벡터 (예를 들어, 렌티바이러스 벡터, 아데노-연관된 바이러스 벡터, 바쿨로바이러스 벡터)가 포함된다. 벡터는 추가의 발현 조절 서열 (예를 들어, 인핸서 서열, 코작 서열, 폴리아데닐화 서열, 전사 종결 서열), 선택가능한 마커 서열 (예를 들어, 항생제 내성 유전자), 복제 기점 등을 포함할 수 있다. 추가의 정보는 ["Current Protocols in Molecular Biology" Ausubel et al., John Wiley & Sons, New York, 2003] 또는 ["Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3rd edition, 2001]에서 확인할 수 있다.
- [0075] 벡터는 또한 형질전환된 세포의 선택을 위해 선택가능한 마커 유전자를 포함할 수 있다. 선택가능한 마커 유전자는 형질전환된 세포 또는 조직의 선택을 위해 사용된다. 마커 유전자에는 항생제 내성을 코딩하는 유전자, 예컨대 네오마이신 포스포트랜스퍼라제 II (NEO) 및 히그로마이신 포스포트랜스퍼라제 (HPT)를 코딩하는

유전자, 뿐만 아니라 제조성 화합물, 예컨대 글루포시네이트 암모늄, 브로복시닐, 이미다졸리논, 및 2,4-디클로로페녹시아세테이트 (2,4-D)에 대한 내성을 부여하는 유전자가 포함된다.

[0076] 일부 실시양태에서, RGN 폴리펩티드를 코딩하는 서열을 포함하는 발현 카세트 또는 벡터는 crRNA 및/또는 tracrRNA를 코딩하는 서열, 또는 가이드 RNA를 생성하기 위해 조합된 crRNA 및 tracrRNA를 추가로 포함할 수 있다. crRNA 및/또는 tracrRNA를 코딩하는 서열(들)은 관심 유기체 또는 숙주 세포에서 crRNA 및/또는 tracrRNA의 발현을 위해 적어도 하나의 전사 조절 서열에 작동가능하게 연결될 수 있다. 예를 들어, crRNA 및/또는 tracrRNA를 코딩하는 폴리뉴클레오티드는 RNA 폴리머라제 III (Pol III)에 의해 인식되는 프로모터 서열에 작동가능하게 연결될 수 있다. 적합한 Pol III 프로모터의 예에는 포유동물 U6, U3, H1, 및 7SL RNA 프로모터 및 쌀 U6 및 U3 프로모터가 포함되나 이로 제한되지 않는다.

[0077] 나타낸 바와 같이, RGN, crRNA, tracrRNA 및/또는 sgRNA를 코딩하는 뉴클레오티드 서열을 포함하는 발현 구축물을 이용하여 관심 유기체를 형질전환시킬 수 있다. 형질전환 방법은 뉴클레오티드 구축물을 관심 유기체에 도입시키는 것을 수반한다. "도입시키는"은 구축물이 숙주 세포의 내부로 접근하게 하는 방식으로 뉴클레오티드 구축물을 숙주 세포에 도입시키는 것으로 의도된다. 본 발명의 방법은 뉴클레오티드 구축물을 숙주 유기체에 도입시키기 위해 특정한 방법을 필요로 하지 않으며, 단지 뉴클레오티드 구축물이 숙주 유기체의 적어도 하나의 세포의 내부로 접근하기만 하면 된다. 숙주 세포는 진핵생물 또는 원핵 세포일 수 있다. 특정한 실시양태에서, 진핵생물 숙주 세포는 식물 세포, 포유동물 세포, 또는 곤충 세포이다. 안정한 형질전환 방법, 일시적인 형질전환 방법, 및 바이러스-매개된 방법을 비롯하여 이로 제한되지 않는, 뉴클레오티드 구축물을 식물 및 다른 숙주 세포에 도입시키는 방법은 관련 기술분야에 공지되어 있다.

[0078] 상기 방법은 형질전환된 유기체, 예컨대 식물, 예컨대 전체 식물, 뿐만 아니라 식물 기관 (예를 들어, 잎, 줄기, 뿌리 등), 종자, 식물 세포, 주아, 배아 및 그의 자손을 생성한다. 식물 세포는 분화될 수 있거나 또는 분화되지 않을 수 있다 (예를 들어 캘러스, 현탁액 배양 세포, 원형질체, 잎 세포, 뿌리 세포, 체관부 세포, 화분).

[0079] "트랜스제닉 유기체" 또는 "형질전환된 유기체" 또는 "안정하게 형질전환된" 유기체 또는 세포 또는 조직은 본 발명의 RGN, crRNA 및/또는 tracrRNA를 코딩하는 폴리뉴클레오티드에 도입되거나 또는 통합된 유기체를 지칭한다. 다른 외인성 또는 내인성 핵산 서열 또는 DNA 단편 또한 숙주 세포에 혼입될 수 있는 것으로 인지된다. 아그로박테리움(*Agrobacterium*)- 및 유전자총-매개된 형질전환은 식물 세포의 형질전환을 위해 주로 이용되는 2 가지 접근법이다. 그러나, 숙주 세포의 형질전환은 감염, 형질감염, 미량주사, 전기천공, 미량투영, 유전자총 또는 입자 충격, 전기천공, 실리카/탄소 섬유, 초음파 매개, PEG 매개, 인산칼슘 공동 침전, 다가양이온 DMSO 기술, DEAE 텍스트란 절차, 및 바이러스 매개, 리포솜 매개 등에 의해 수행될 수 있다. RGN, crRNA 및/또는 tracrRNA를 코딩하는 폴리뉴클레오티드의 바이러스-매개된 도입에는 레트로바이러스, 렌티바이러스, 아데노바이러스, 및 아데노-연관된 바이러스 매개된 도입 및 발현, 뿐만 아니라 카울리모바이러스, 제미니바이러스, 및 RNA 식물 바이러스가 포함된다.

[0080] 형질전환 프로토콜 뿐만 아니라, 폴리펩티드 또는 폴리뉴클레오티드 서열을 식물에 도입시키기 위한 프로토콜은 형질전환을 위해 표적화된 숙주 세포의 유형 (예를 들어, 외떡잎 또는 쌍떡잎 식물 세포)에 따라 달라질 수 있다. 형질전환 방법은 관련 기술분야에 공지되어 있고, 미국 특허 번호: 8,575,425; 7,692,068; 8,802,934; 7,541,517에 제시된 것들이 포함되며, 이들 각각은 본원에 참조로 포함된다. 또한, [Rakoczy-Trojanowska, M. (2002) Cell Mol Biol Lett. 7:849-858; Jones et al. (2005) Plant Methods 1:5; Rivera et al. (2012) Physics of Life Reviews 9:308-345; Bartlett et al. (2008) Plant Methods 4:1-12; Bates, G.W. (1999) Methods in Molecular Biology 111:359-366; Binns and Thomashow (1988) Annual Reviews in Microbiology 42:575-606; Christou, P. (1992) The Plant Journal 2:275-281; Christou, P. (1995) Euphytica 85:13-27; Tzfira et al. (2004) TRENDS in Genetics 20:375-383; Yao et al. (2006) Journal of Experimental Botany 57:3737-3746; Zupan and Zambryski (1995) Plant Physiology 107:1041-1047; Jones et al. (2005) Plant Methods 1:5]을 참고한다

[0081] 형질전환은 세포로 핵산의 안정한 또는 일시적인 포함을 일으킬 수 있다. "안정한 형질전환"은, 숙주 세포에 도입된 뉴클레오티드 구축물이 숙주 세포의 게놈과 통합하고, 그의 자손에 의해 유전될 수 있는 것을 의미하는 것으로 의도된다. "일시적인 형질전환"은, 폴리뉴클레오티드가 숙주 세포에 도입되고, 숙주 세포의 게놈에 통합되지 않는 것을 의미하는 것으로 의도된다.

[0082] 염록체의 형질전환 방법은 관련 기술분야에 공지되어 있다. 예를 들어, [Svab et al. (1990) Proc. Natl.

Acad. Sci. USA 87:8526-8530; Svab and Maliga (1993) Proc. Natl. Acad. Sci. USA 90:913-917; Svab and Maliga (1993) EMBO J. 12:601-606]을 참고한다. 상기 방법은 선택가능한 마커를 함유하는 DNA의 입자 총 전달, 및 상동성 재조합을 통해 색소체 계층으로 DNA의 표적화에 의존한다. 추가로, 색소체 형질전환은 핵-코딩된 및 색소체-지정된 RNA 폴리머라제의 조직-선택 발현에 의한 침묵 색소체-매개 트랜스진의 전사촉진에 의해 달성될 수 있다. 이러한 시스템은 [McBride et al. (1994) Proc. Natl. Acad. Sci. USA 91:7301-7305]에서 보고되었다.

[0083] 형질전환된 세포는 통상적인 방식에 따라 트랜스제닉 유기체, 예컨대 식물로 성장할 수 있다. 예를 들어, [McCormick et al. (1986) Plant Cell Reports 5:81-84]를 참고한다. 이어서, 이들 식물을 성장시킨 다음, 동일한 형질전환된 균주 또는 상이한 균주와 수분시킬 수 있고, 생성된 혼성체는 확인된 원하는 표현형 특징의 구성적 발현을 가질 수 있다. 원하는 표현형 특징의 발현이 안정하게 유지되고 유전된 다음, 종자가 수확되는 것을 보장하도록 2 세대 이상 성장시켜, 원하는 표현형 특징의 발현이 달성될 수 있도록 보장할 수 있다. 이러한 방식으로, 본 발명은 계층에 안정하게 도입된 본 발명의 뉴클레오티드 구축물, 예를 들어 본 발명의 발현 카세트를 갖는 형질전환된 종자 ("트랜스제닉 종자"로도 지칭됨)를 제공한다.

[0084] 대안적으로, 형질전환된 세포는 유기체에 도입될 수 있다. 이들 세포는 유기체로부터 기원할 수 있고, 세포는 생체의 접근법으로 형질전환된다.

[0085] 본원에 제공된 서열은 임의의 식물 종, 예컨대 비제한적으로, 외떡잎 식물 및 쌍떡잎 식물의 형질전환을 위해 사용될 수 있다. 관심 식물의 예에는 옥수수 (메이즈), 수수, 밀, 해바라기, 토마토, 크루시퍼, 후추, 감자, 목화, 쌀, 대두, 사탕무, 사탕수수, 담배, 보리, 및 유채, 브라시카(Brassica) 종, 알팔파, 호밀, 조, 홍화, 땅콩, 고구마, 카사바, 커피, 코코넛, 파인애플, 감귤 나무, 코코야, 차, 바나나, 아보카도, 무화과, 구아바, 망고, 올리브, 파파야, 캐슈, 마카다미아, 아몬드, 귀리, 채소, 관상용 식물, 및 구과 식물이 포함되나 이로 제한되지 않는다.

[0086] 채소에는 토마토, 상추, 그린 빈, 리마 빈, 완두콩, 및 큐큐미스(Cucumis) 속의 구성원, 예컨대 오이, 칸탈루프, 및 머스크 멜론이 포함되나 이로 제한되지 않는다. 관상용 식물에는 진달래, 수국, 히비스커스, 장미, 튜립, 수선화, 페튜니아, 카네이션, 포인세티아, 및 국화가 포함되나 이로 제한되지 않는다. 바람직하게는, 본 발명의 식물은 농작물 (예를 들어, 옥수수, 수수, 밀, 해바라기, 토마토, 크루시퍼, 후추, 감자, 목화, 쌀, 대두, 사탕무, 사탕수수, 담배, 보리, 유채 등)이다.

[0087] 본원에서 사용된 바와 같이, 용어 식물에는 식물 세포, 식물 원형질체, 식물이 재생될 수 있는 식물 세포 조직 배양물, 식물 캘리(calli), 식물 덩어리, 및 식물 또는 식물의 일부분, 예컨대 배아, 화분, 배주, 종자, 잎, 꽃, 가지, 과실, 커널, 이삭, 콕, 겉껍질, 자루, 뿌리, 근단, 꽃밥 등에서 온전한 식물 세포가 포함된다. 곡물은 종의 재배 또는 번식 이외의 목적으로 상업적인 재배자에 의해 생산된 성숙한 종자를 의미하는 것으로 의도된다. 재생된 식물의 자손, 변이체 및 돌연변이체 또한 본 발명의 범위 내에 포함되나, 단, 이들 부분은 도입된 폴리뉴클레오티드를 포함한다. 본원에 개시된 서열을 보유하는 가공된 식물 생성물 또는 부산물, 예를 들어 대두박이 추가로 제공된다.

[0088] RGN, crRNA 및/또는 tracrRNA를 코딩하는 폴리뉴클레오티드는 또한 임의의 원핵생물 종, 예컨대 비제한적으로, 고세균류 및 박테리아 (예를 들어, 바실루스(*Bacillus*) 종, 클레브시엘라(*Klebsiella*) 종, 스트렙토마이세스(*Streptomyces*) 종, 리조비움(*Rhizobium*) 종, 에스케리키아(*Escherichia*) 종, 슈도모나스(*Pseudomonas*) 종, 살모넬라(*Salmonella*) 종, 시겔라(*Shigella*) 종, 비브리오(*Vibrio*) 종, 예르시니아(*Yersinia*) 종, 마이코플라스마(*Mycoplasma*) 종, 아그로박테리움, 락토바실러스(*Lactobacillus*) 종)을 형질전환시키기 위해 사용될 수 있다.

[0089] RGN, crRNA 및/또는 tracrRNA를 코딩하는 폴리뉴클레오티드는 임의의 진핵생물 종, 예컨대 비제한적으로 동물 (예를 들어, 포유동물, 곤충, 어류, 새, 및 파충류), 진균, 아메바, 조류 및 효모를 형질전환시키기 위해 사용될 수 있다.

[0090] 통상적인 바이러스 및 비-바이러스 기반 유전자 전달 방법을 이용하여 포유동물 세포 또는 표적 조직에서 핵산을 도입시킬 수 있다. 이러한 방법을 이용하여 CRISPR 시스템의 성분을 코딩하는 핵산을 배양물 중의 또는 숙주 유기체 중의 세포에 투여할 수 있다. 비-바이러스 벡터 전달 시스템에는 DNA 플라스미드, RNA (예를 들어, 본원에 기재된 벡터의 전사체), 네이키드 핵산, 및 리포솜과 같은 전달 비히클과 복합체화된 핵산이 포함된다. 바이러스 벡터 전달 시스템에는 세포로 전달 후에 에피솜 또는 통합 계층을 갖는 DNA 및 RNA 바이러스가 포함된다. 유전자 요법 절차의 검토를 위해, [Anderson, Science 256: 808- 813 (1992); Nabel & Feigner, 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 Bohm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994)]를 참고한다.

[0091] 핵산의 비-바이러스 전달 방법에는 리포펙션, 뉴클레오펙션, 미량주사, 유전자총, 비로솜, 리포솜, 이뮤노리포솜, 다가양이온 또는 지질: 핵산 집합체, 네이키드 DNA, 인공 비리온, 및 DNA의 작용제-증강된 흡수가 포함된다. 리포펙션은 예를 들어 미국 특허 번호 5,049,386, 4,946,787; 및 4,897,355)에 기재되어 있고, 리포펙션 시약은 상업적으로 판매된다 (예를 들어, 트랜스펙탐(Transfectam)TM 및 리포펙틴(Lipofectin)TM). 폴리뉴클레오티드의 효율적인 수용체-인식 리포펙션에 적합한 양이온성 및 중성 지질에는 프라이그너(Feigner), WO 91/17424; WO 91/16024의 것들이 포함된다. 전달은 세포로 (예를 들어 시험관내 또는 생체의 투여) 또는 표적 조직으로 (예를 들어 생체내 투여)일 수 있다. 지질:핵산 복합체, 예컨대 표적화된 리포솜, 예컨대 면역지질 복합체의 제조는 관련 기술분야의 기술자에게 널리 공지되어 있다 (예를 들어, [Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992)]; 미국 특허 번호 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 및 4,946,787 참고).

[0092] 핵산의 전달을 위한 RNA 또는 DNA 바이러스 기반 시스템의 사용은 바이러스를 신체의 특이적인 세포로 표적화하고 바이러스 페이로드를 핵으로 수송하는 고도로 진화된 과정을 활용한다. 바이러스 벡터는 환자에게 직접적으로 투여될 수 있거나 (생체내) 또는 이들을 이용하여 시험관내에서 세포를 치료할 수 있고, 임의적으로 변형된 세포가 환자에게 투여될 수 있다 (생체외). 통상적인 바이러스 기반 시스템에는 유전자 전달을 위한 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관된 및 단순 포진 바이러스 벡터가 포함될 수 있다. 숙주 계통에서 통합은 레트로바이러스, 렌티바이러스, 및 아데노-연관된 바이러스 유전자 전달 방법에 의해 가능하며, 이는 종종 삽입된 트랜스진의 장기간 발현을 일으킨다. 추가로, 높은 형질도입 효율이 여러 상이한 세포 유형 및 표적 조직에서 관찰되었다.

[0093] 레트로바이러스의 형성은 외래 외피 단백질을 도입하고, 표적 세포의 잠재적인 표적 집단을 확장시킴으로써 변경될 수 있다. 렌티바이러스 벡터는 비-분열 세포를 형질도입하거나 또는 감염시킬 수 있고, 전형적으로 높은 바이러스 역가를 생성할 수 있는 레트로바이러스 벡터이다. 따라서, 레트로바이러스 유전자 전달 시스템의 선택은 표적 조직에 따라 좌우될 것이다. 레트로바이러스 벡터는 6-10 kb 이하의 외래 서열에 대해 패키징 능력을 갖는 시스-작용 긴 말단 반복부로 구성된다. 최소의 시스-작용 LTR은 벡터의 복제 및 패키징에 충분하며, 이를 사용하여 치료 유전자를 표적 세포에 통합시켜 영구적인 트랜스진 발현을 제공한다. 광범위하게 사용되는 레트로바이러스 벡터에는 뮤린 백혈병 바이러스 (MuLV), 긴팔 원숭이 백혈병 바이러스 (GaLV), 유인원 면역결핍 바이러스 (SIV), 인간 면역결핍 바이러스 (HIV), 및 이들의 조합물을 기반으로 하는 것들이 포함된다 (예를 들어, [Buchscher et al., J. Viral. 66:2731-2739 (1992); Johann et al., J. Viral. 66:1635-1640 (1992); Sommerfelt et al., Viral. 176:58-59 (1990); Wilson et al., J. Viral. 63:2374-2378 (1989); Miller et al., 1. Viral. 65:2220-2224 (1991)]; PCT/US94/05700 참고).

[0094] 일시적인 발현이 바람직한 경우에, 아데노바이러스 기반 시스템이 사용될 수 있다. 아데노바이러스 기반 벡터는 여러 세포 유형에서 매우 높은 형질도입 효율을 가질 수 있고, 세포 분열을 필요로 하지 않는다. 이러한 벡터를 사용하여, 높은 역가 및 수준의 발현이 수득되었다. 이 벡터는 비교적 간단한 시스템으로 다량 생산될 수 있다. 또한, 아데노-연관된 바이러스 ("AAV") 벡터를 사용하여, 예를 들어 핵산 및 펩티드의 시험관내 생성에서 및 생체내 및 생체의 유전자 요법 절차를 위해 세포를 표적 핵산으로 형질도입시킬 수 있다 (예를 들어, [West et al., Virology 160:38-47 (1987)]; 미국 특허 번호 4,797,368; WO 93/24641; [Katin, Human Gene Therapy 5:793-801 (1994); Muzyczka, 1. Clin. Invest. 94:1351 (1994)] 참고). 제조용 AAV 벡터의 구축은 수많은 공보, 예컨대 미국 특허 번호 5,173,414; [Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., 1. Viral. 63:03822-3828 (1989)]에 기재되어 있다. 전형적으로 패키징 세포를 사용하여 숙주 세포를 감염시킬 수 있는 바이러스 입자를 형성한다. 이러한 세포에는 아데노바이러스를 패키징하는 293 세포, 및 레트로바이러스를 패키징하는 ψ J2 세포 또는 PA317 세포가 포함된다.

[0095] 유전자 요법에서 사용되는 바이러스 벡터는 일반적으로 핵산 벡터를 바이러스 입자로 패키징하는 세포주를 생성

함으로써 생성된다. 벡터는 전형적으로 패키징 및 숙주로의 후속적인 통합을 위해 필요한 최소 바이러스 서열을 함유하며, 다른 바이러스 서열은 발현될 폴리뉴클레오티드(들)에 대한 발현 카세트에 의해 대체된다. 잃어버린 바이러스 기능은 전형적으로 패키징 세포주에 의해 트랜스로 제공된다. 예를 들어, 유전자 요법에서 사용되는 AAV 벡터는 전형적으로 패키징 및 숙주 게놈으로의 통합을 위해 필요한 AAV 게놈으로부터의 ITR 서열만을 갖는다. 바이러스 DNA는 다른 AAV 유전자, 즉, rep 및 cap을 코딩하는 헬퍼 플라스미드를 함유하지만 ITR 서열은 결여된 세포주에서 패키징된다.

[0096] 세포주는 또한 헬퍼로서 아데노바이러스에 의해 감염될 수 있다. 헬퍼 바이러스는 헬퍼 플라스미드로부터 AAV 벡터의 복제 및 AAV 유전자의 발현을 촉진시킨다. 헬퍼 플라스미드는 ITR 서열의 결여로 인해 유의한 양으로 패키징되지 않는다. 아데노바이러스에 의한 오염은 예를 들어 AAV에 비해 아데노바이러스가 덜 민감한 열 처리에 의해 감소될 수 있다. 세포로 핵산의 추가의 전달 방법은 관련 기술분야의 기술자에게 공지되어 있다. 예를 들어, 본원에 참조로 포함된 US20030087817을 참고한다.

[0097] 일부 실시양태에서, 숙주 세포는 본원에 기재된 하나 이상의 벡터에 의해 일시적으로 또는 비-일시적으로 형질감염된다. 일부 실시양태에서, 세포는 대상체에서 자연적으로 발생하는 것처럼 형질감염된다. 일부 실시양태에서, 형질감염된 세포를 대상체로부터 취한다. 일부 실시양태에서, 세포는 대상체로부터 취한 세포, 예컨대 세포주로부터 유래된다. 조직 배양을 위한 매우 다양한 세포주가 관련 기술분야에 공지되어 있다. 세포주의 예에는 C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLaS3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TF1, CTLL-2, CIR, Rat6, CVI, RPTE, A10, T24, 182, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, lurkat, 145.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 원숭이 신장 상피, BALB/3T3 마우스 배아 섬유모세포, 3T3 Swiss, 3T3-L1, 132-d5 인간 태아 섬유모세포; 10.1 마우스 섬유모세포, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-I 세포, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr-/-, COR-L23, COR-L23/CPR, COR-L235010, CORL23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalclc7, HL-60, HMEC, HT-29, lurkat, 1Y 세포, K562 세포, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Me1 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCKII, MDCKII, MOR/ 0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT 세포주, Peer, PNT-1A/ PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 세포, Sf-9, SkBr3, T2, T-47D, T84, THP1 세포주, U373, U87, U937, VCaP, Vero 세포, WM39, WT-49, X63, YAC-1, YAR, 및 그의 트랜스제닉 변이체가 포함되나 이로 제한되지 않는다. 세포주는 관련 기술분야의 기술자에게 공지된 다양한 공급처로부터 입수가 가능하다 (예를 들어, 아메리칸 타입 컬처 콜렉션(American Type Culture Collection, ATCC) (버지니아주 마나사스) 참고).

[0098] 일부 실시양태에서, 본원에 기재된 하나 이상의 벡터로 형질감염된 세포를 사용하여, 하나 이상의 벡터-유래된 서열을 포함하는 새로운 세포주를 확립한다. 일부 실시양태에서, 본원에 기재된 CRISPR 시스템의 성분으로 일시적으로 형질감염되고 (예컨대 하나 이상의 벡터의 일시적인 형질감염, 또는 RNA로의 형질감염에 의해), CRISPR 복합체의 활성을 통해 변형된 세포를 사용하여, 변형을 함유하지만 임의의 다른 외인성 서열이 결여된 세포를 포함하는 새로운 세포주를 확립한다. 일부 실시양태에서, 본원에 기재된 하나 이상의 벡터에 의해 일시적으로 또는 비-일시적으로 형질감염된 세포, 또는 이러한 세포로부터 유래된 세포주는 하나 이상의 시험 화합물을 평가하는데 사용된다.

[0099] 일부 실시양태에서, 본원에 기재된 하나 이상의 벡터를 사용하여, 비-인간 트랜스제닉 동물 또는 트랜스제닉 식물을 생성한다. 일부 실시양태에서, 트랜스제닉 동물은 포유동물, 예컨대 마우스, 래트 또는 토끼이다.

[0100] IV. 폴리펩티드 및 폴리뉴클레오티드의 변이체 및 단편

[0101] 본 개시내용은 그의 아미노산 서열이 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 또는 235로 제시된 자연-발생 (즉, 야생형) RNA-가이드된 뉴클레아제의 활성 변이체 및 단편, 뿐만 아니라 자연-발생 CRISPR 반복부, 예컨대 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273 또는 287로 제시된 서열의 활성 변이체 및 단편, 및 자연-발생 tracrRNA, 예컨대 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274 또는 286으로 제시된 서열의 활성 변이체 및 단편, 및 이를 코딩하는 폴리뉴클레오티드를 제공한다.

- [0102] 변이체 또는 단편의 활성이 관심 폴리뉴클레오타이드 또는 폴리펩티드에 비해 변경될 수 있지만, 변이체 및 단편은 관심 폴리뉴클레오타이드 또는 폴리펩티드의 기능을 보유해야 한다. 예를 들어, 변이체 또는 단편은 관심 폴리뉴클레오타이드 또는 폴리펩티드와 비교하여 증가된 활성, 감소된 활성, 상이한 스펙트럼의 활성 또는 활성에서의 임의의 다른 변경을 가질 수 있다.
- [0103] 본원에 개시된 것과 같은 자연 발생 RGN 폴리펩티드의 단편 및 변이체는 서열-특이적인, RNA-가이드된 DNA-결합 활성을 보유할 것이다. 특정한 실시양태에서, 본원에 개시된 것과 같은 자연 발생 RGN 폴리펩티드의 단편 및 변이체는 뉴클레아제 활성 (단일-가닥 또는 이중-가닥)을 보유할 것이다.
- [0104] 본원에 개시된 것과 같은 자연 발생 CRISPR 반복부의 단편 및 변이체는 (tracrRNA를 포함하는) 가이드 RNA의 일부일 때 (가이드 RNA와 복합체화된) RNA-가이드된 뉴클레아제에 결합하여 이를 서열-특이적인 방식으로 표적 뉴클레오타이드 서열로 가이드하는 능력을 보유할 것이다.
- [0105] 본원에 개시된 것과 같은 자연 발생 tracrRNA의 단편 및 변이체는 (CRISPR RNA를 포함하는) 가이드 RNA의 일부일 때 (가이드 RNA와 복합체화된) RNA-가이드된 뉴클레아제를 서열-특이적인 방식으로 표적 뉴클레오타이드 서열로 가이드하는 능력을 보유할 것이다.
- [0106] 용어 "단편"은 본 발명의 폴리뉴클레오타이드 또는 폴리펩티드 서열의 일부분을 지칭한다. "단편" 또는 "생물학적으로 활성인 부분"에는 생물학적 활성을 보유하도록 (즉, 가이드RNA 내에 포함될 때 서열-특이적인 방식으로 RGN과 결합하여 이를 표적 뉴클레오타이드 서열로 지시하도록) 충분한 개수의 연속 뉴클레오타이드를 포함하는 폴리뉴클레오타이드가 포함된다. "단편" 또는 "생물학적으로 활성인 부분"에는 생물학적 활성을 보유하도록 (즉, 가이드 RNA와 복합체화될 때 서열-특이적인 방식으로 표적 뉴클레오타이드 서열에 결합하도록) 충분한 개수의 연속 아미노산 잔기를 포함하는 폴리펩티드가 포함된다. RGN 단백질의 단편에는 대안적인 하류 시작 부위의 사용으로 인해 전장 서열보다 더 짧은 것들이 포함된다. RGN 단백질의 생물학적으로 활성인 부분은 예를 들어 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137, 또는 235의 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050개 또는 그 초과인 연속 아미노산 잔기를 포함하는 폴리펩티드일 수 있다. 이러한 생물학적으로 활성인 부분은 재조합 기술에 의해 제조되고 서열-특이적인 RNA-가이드된 DNA-결합 활성에 대해 평가될 수 있다. CRISPR 반복 서열의 생물학적으로 활성인 단편은 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287의 적어도 8개 연속 아미노산을 포함할 수 있다. CRISPR 반복 서열의 생물학적으로 활성인 부분은 예를 들어 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287의 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20개 연속 뉴클레오타이드를 포함하는 폴리뉴클레오타이드일 수 있다. tracrRNA의 생물학적으로 활성인 부분은 예를 들어 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274, 또는 286의 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, 55, 60, 65, 70, 75, 80개 또는 그 초과인 연속 뉴클레오타이드를 포함하는 폴리뉴클레오타이드일 수 있다.
- [0107] 일반적으로, "변이체"는 실질적으로 유사한 서열을 의미하는 것으로 의도된다. 폴리뉴클레오타이드의 경우, 변이체는 본래의 폴리뉴클레오타이드 내의 하나 이상의 내부 부위에서 하나 이상의 뉴클레오타이드의 결실 및/또는 부가, 및/또는 본래의 폴리뉴클레오타이드에서 하나 이상 부위에서 하나 이상의 뉴클레오타이드의 치환을 포함한다. 본원에서 사용된 바와 같이, "본래의" 또는 "야생형" 폴리뉴클레오타이드 또는 폴리펩티드는 각각 자연 발생 뉴클레오타이드 서열 또는 아미노산 서열을 포함한다. 폴리뉴클레오타이드의 경우, 보존적 변이체에는 유전자 코드의 축퇴성 때문에 관심 유전자의 본래의 아미노산 서열을 코딩하는 이들 서열이 포함된다. 이들과 같은 자연 발생 대립유전자 변이체는 널리 공지된 분자 생물학 기술, 예를 들어 하기에서 설명되는 폴리머라제 연쇄 반응 (PCR) 및 혼성화 기술을 사용하여 확인될 수 있다. 변이체 폴리뉴클레오타이드에는 또한 합성에 의해 유래된 폴리뉴클레오타이드, 예를 들어 부위-지정된 돌연변이유발을 이용하여 생성되었지만 관심 폴리펩티드 또는 폴리뉴클레오타이드를 여전히 코딩하는 것들이 포함된다. 일반적으로, 본원에 개시된 특정한 폴리뉴클레오타이드의 변이체는 본원에 기재된 서열 정렬 프로그램 및 파라미터에 의해 결정시 특정한 해당 폴리뉴클레오타이드와 적어도 약 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과인 서열 동일성을 가질 것이다.
- [0108] 본원에 개시된 특정한 폴리뉴클레오타이드 (즉, 기준 폴리뉴클레오타이드)의 변이체는 또한 변이체 폴리뉴클레오타이드에 의해 코딩된 폴리펩티드 및 기준 폴리뉴클레오타이드에 의해 코딩된 폴리펩티드 사이의 퍼센트 서열 동일성을 비교함으로써 평가될 수 있다. 임의의 두 폴리펩티드 사이의 퍼센트 서열 동일성은 본원에 기재된 서열 정

렬 프로그램 및 파라미터를 사용하여 계산될 수 있다. 본원에 개시된 폴리뉴클레오타이드의 임의의 주어진 쌍을 그들이 코딩하는 두 폴리펩티드에 의해 공유된 퍼센트 서열 동일성을 비교함으로써 평가하는 경우, 코딩된 두 폴리펩티드 사이의 퍼센트 서열 동일성은 적어도 약 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과인 서열 동일성이다.

[0109] 특정 실시양태에서, 본원에 개시된 폴리뉴클레오타이드는 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137 또는 235의 아미노산 서열에 대해 적어도 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과인 동일성을 갖는 아미노산 서열을 포함하는 RNA-가이드된 뉴클레아제 폴리펩티드를 코딩한다.

[0110] 본 발명의 RGN 폴리펩티드의 생물학적으로 활성인 변이체는 적게는 약 1-15개 아미노산 잔기, 적게는 약 1-10개, 예컨대 약 6-10개, 적게는 5개, 적게는 4개, 적게는 3개, 적게는 2개 또는 적게는 1개 아미노산 잔기만 큼 상이할 수 있다. 구체적인 실시양태에서, 폴리펩티드는 N-말단 또는 C-말단 절단을 포함할 수 있으며, 이는 폴리펩티드의 N 또는 C 말단으로부터 적어도 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050개 아미노산 또는 그 초과인 결실을 포함할 수 있다.

[0111] 특정한 실시양태에서, 본원에 개시된 폴리뉴클레오타이드는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287로 제시된 뉴클레오타이드 서열에 대해 적어도 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과인 동일성을 갖는 뉴클레오타이드 서열을 포함하는 CRISPR 반복부를 포함하거나 또는 코딩한다.

[0112] 본원에 개시된 폴리뉴클레오타이드는 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274, 또는 286으로 제시된 뉴클레오타이드 서열에 대해 적어도 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과인 동일성을 갖는 뉴클레오타이드 서열을 포함하는 tracrRNA를 포함하거나 또는 코딩할 수 있다.

[0113] 본 발명의 CRISPR 반복부 또는 tracrRNA의 생물학적으로 활성인 변이체는 적게는 약 1-15개 뉴클레오타이드, 적게는 약 1-10개, 예컨대 약 6-10개, 적게는 5개, 적게는 4개, 적게는 3개, 적게는 2개 또는 적게는 1개 뉴클레오타이드만큼 상이할 수 있다. 특정 실시양태에서, 폴리뉴클레오타이드는 5' 또는 3' 말단 절단을 포함할 수 있고, 이는 폴리뉴클레오타이드의 5' 또는 3' 말단으로부터 적어도 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80개 뉴클레오타이드 또는 그 초과인 결실을 포함할 수 있다.

[0114] 변이체 단백질 및 폴리뉴클레오타이드를 생성하는 본원에 제공된 RGN 폴리펩티드, CRISPR 반복부 및 tracrRNA에 대해 변형이 이루어질 수 있음을 인지한다. 인위적으로 설계된 변화는 부위-지정된 돌연변이유발 기술의 적용을 통해 도입될 수 있다. 대안적으로, 본원에 개시된 서열과 구조적으로 및/또는 기능적으로 관련이 있는 아직 공지되지 않거나 또는 아직 확인되지 않은 본래의 폴리뉴클레오타이드 및/또는 폴리펩티드 또한 본 발명의 범위 내에 속하는 것으로 확인될 수 있다. 보존적 아미노산 치환이 RGN 단백질의 기능을 변경시키지 않는 비보존된 영역에서 이루어질 수 있다. 대안적으로, RGN 또는 데아미나제의 활성을 개선시키는 변형이 이루어질 수 있다.

[0115] 변이체 폴리뉴클레오타이드 및 단백질은 또한 돌연변이유발 및 재조합 유발 절차, 예컨대 DNA 서플링으로부터 유래된 서열 및 단백질을 포함한다. 이러한 절차에 의해, 본원에 개시된 하나 이상의 상이한 RGN 또는 데아미나제 단백질 (예를 들어, 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137, 또는 235)은 원하는 성질을 갖는 새로운 RGN 단백질을 생성하도록 조작된다. 이러한 방식으로, 실질적인 서열 동일성을 갖는 서열 영역을 포함하는 관련된 서열 폴리뉴클레오타이드의 집단으로부터 재조합 폴리뉴클레오타이드의 라이브러리가 생성되고, 시험관내 또는 생체내에서 상동성 재조합될 수 있다. 예를 들어, 이 접근법을 이용하여, 관심 도메인을 코딩하는 서열 모티프가 본원에 제공된 RGN 서열 및 다른 공지된 RGN 유전자 사이에서 서플링되어, 개선된 관심 성질, 예컨대 효소의 경우 증가된 K_m 을 갖는 단백질을 코딩하는 새로운 유전자를 수득할 수 있다. 이러한 DNA 서플링을 위한 전략은 관련 기술분야에 공지되어 있다. 예를 들어, [Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751; Stemmer (1994) Nature 370:389-391; Cramer et al. (1997) Nature Biotech. 15:436-438; Moore et al. (1997) J. Mol. Biol. 272:336-347; Zhang et al. (1997) Proc. Natl. Acad. Sci. USA 94:4504-4509; Cramer et al. (1998) Nature 391:288-291]; 및 미국 특허 번호

5,605,793 및 5,837,458을 참고한다. "서플링된" 핵산은 서플링 절차, 예컨대 본원에 제시된 임의의 서플링 절차에 의해 생성되는 핵산이다. 서플링된 핵산은 예를 들어 인공적인, 임의적으로 재귀적인 방식으로 2가지 이상의 핵산 (또는 문자열)을 (물리적으로 또는 가상으로) 재조합함으로써 생성된다. 일반적으로, 하나 이상의 스크리닝 단계를 서플링 과정에서 이용하여 관심 핵산을 확인하고; 이 스크리닝 단계는 임의의 재조합 단계 이전에 또는 이후에 수행될 수 있다. 일부 (그러나 전부는 아님) 서플링 실시양태에서, 스크리닝될 풀의 다양성을 증가시키기 위해 선택 이전에 다중 라운드의 재조합을 수행하는 것이 바람직하다. 재조합 및 선택의 전체 과정을 임의적으로 재귀적으로 반복한다. 문맥에 따라, 서플링은 재조합 및 선택의 전체 과정을 지칭할 수 있거나, 또는 대안적으로 전체 과정 중 재조합 부분을 간단히 지칭할 수 있다.

[0116] 본원에서 사용된 바와 같이, 두 폴리뉴클레오타이드 또는 폴리펩티드 서열과 관련하여 "서열 동일성" 또는 "동일성"은 명시된 비교 범위에 걸쳐 최대 대응을 위해 정렬될 때 동일한 두 서열의 잔기를 지칭한다. 서열 동일성의 백분율이 단백질과 관련하여 사용될 때, 동일하지 않은 잔기 위치는 종종 보존적 아미노산 치환에 의해 상이한 것으로 인지되며, 여기서 아미노산 잔기는 유사한 화학적 성질 (예를 들어, 전하 또는 소수성)을 갖는 다른 아미노산 잔기로 치환되고, 따라서 분자의 기능적 성질이 변하지 않는다. 서열이 보존적 치환에서 상이할 때, 퍼센트 서열 동일성은 치환의 보존적 성질을 보정하기 위해 상향 조정될 수 있다. 이러한 보존적 치환에 의해 상이한 서열은 "서열 유사성" 또는 "유사성"을 갖는다고 한다. 이러한 조정을 위한 수단은 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다. 전형적으로 이는 전체가 아니라 부분 미스매치로서 보존적 치환을 평점하여, 서열 동일성 백분율을 증가시킨다. 따라서, 예를 들어, 동일한 아미노산에 1의 점수가 주어지고, 비-보존적 치환에 0의 점수가 주어지는 경우, 보존적 치환에는 0 내지 1의 점수가 주어진다. 보존적 치환의 평점은 예를 들어 프로그램 PC/GENE (인텔리제네틱스(Intelligenetics), 캘리포니아주 마운틴뷰)에서 구현되는 바와 같이 계산된다.

[0117] 본원에서 사용된 바와 같이, "서열 동일성 백분율"은 비교 범위에 걸쳐 최적으로 정렬된 두 서열을 비교함으로써 결정된 값을 의미하며, 비교 범위에서 폴리뉴클레오타이드 서열의 일부분은 두 서열의 최적의 정렬을 위해 기준 서열 (부가 또는 결실을 포함하지 않음)과 비교하여 부가 또는 결실 (즉, 갭)을 포함할 수 있다. 백분율은 매칭된 위치의 개수를 획득하기 위해 두 서열에서 발생하는 동일한 핵산 염기 또는 아미노산 잔기의 위치의 개수를 결정하고, 매칭된 위치의 개수를 비교 범위의 총 위치의 개수로 나누고, 그 결과에 100을 곱하여 서열 동일성 백분율을 획득함으로써 계산된다.

[0118] 달리 명시되지 않는다면, 본원에 제공된 서열 동일성/유사성 값은 하기 파라미터를 사용하여 GAP 버전 10을 사용하여 획득한 값을 지칭한다: 50의 GAP 가중치 및 3의 길이 가중치, 및 nwsgapdna.cmp 평점 행렬을 사용하는 뉴클레오타이드 서열에 대한 % 동일성 및 % 유사성; 8의 GAP 가중치 및 2의 길이 가중치, 및 BLOSUM62 평점 행렬을 사용하는 아미노산 서열에 대한 % 동일성 및 % 유사성; 또는 그의 임의의 등가의 프로그램. "등가의 프로그램"은, GAP 버전 10에 의해 생성된 상응하는 정렬과 비교할 때, 해당하는 임의의 두 서열에 대해, 동일한 뉴클레오타이드 또는 아미노산 잔기 매치 및 동일한 퍼센트 서열 동일성을 갖는 정렬을 생성하는 임의의 서열 비교 프로그램을 의미한다.

[0119] 해당 서열 쌍에 대해 가능한 가장 높은 점수에 도달하도록 정의된 아미노산 치환 행렬 (예를 들어, BLOSUM62), 갭 존재 패널티 및 갭 확장 패널티를 이용하는 유사성 평점을 위해 두 서열이 정렬될 때, 이들은 "최적으로 정렬된" 것이다. 아미노산 치환 행렬, 및 두 서열 사이의 유사성을 정량화하는데 있어서 그들의 사용은 관련 기술분야에 널리 공지되어 있고, 예를 들어 [Dayhoff et al. (1978) "A model of evolutionary change in proteins." In "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M. O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, D.C. and Henikoff et al. (1992) Proc. Natl. Acad. Sci. USA 89:10915-10919]에 기재되어 있다. BLOSUM62 행렬은 종종 서열 정렬 프로토콜에서 디폴트 평점 치환 행렬로서 사용된다. 갭 존재 패널티는 정렬된 서열 중 하나에서 단일 아미노산 갭의 도입에 대해 부과되고, 갭 확장 패널티는 이미 개방된 갭에 삽입된 각각의 추가의 빈 아미노산 위치에 대해 부과된다. 정렬은 정렬이 시작되고 끝나는 각각의 서열의 아미노산 위치에 의해, 임의적으로 가능한 가장 높은 점수에 도달하도록 하나의 서열 또는 두 서열 모두에서 하나의 갭 또는 다중 갭의 삽입에 의해 정의된다. 최적의 정렬 및 평점은 수동으로 달성될 수 있지만, 상기 과정은 컴퓨터-실행된 정렬 알고리즘, 예를 들어 [Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402]에 기재되고 생명공학 정보 웹사이트 국제 센터(National Center for Biotechnology Information Website (www.ncbi.nlm.nih.gov))에서 공개적으로 입수가능한 갭이 있는 BLAST 2.0의 사용에 의해 용이해진다. 다중 정렬을 비롯하여 최적의 정렬은 예를 들어 www.ncbi.nlm.nih.gov로부터 입수가능하고 [Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402]에 기재된 PSI-BLAST를 이용하여 제조될 수 있다.

- [0120] 기준 서열에 대해 최적으로 정렬된 아미노산 서열과 관련하여, 아미노산 잔기는 정렬에서 잔기가 쌍을 형성한 기준 서열의 위치에 "상응한다". "위치"는 N-말단에 대한 그의 위치를 기반으로 하여 기준 서열에서 각각의 아미노산을 순차적으로 식별하는 번호로 표시된다. 최적의 정렬을 결정할 때 고려되어야 하는 결실, 삽입, 말단 절단, 융합 등으로 인해, 일반적으로 N-말단으로부터 간단히 카운팅함으로써 결정되는 시험 서열에서의 아미노산 잔기 개수는 기준 서열에서 그의 상응하는 위치의 번호와 반드시 동일하지는 않을 것이다. 예를 들어, 정렬된 시험 서열에서 결실이 있는 경우, 결실 부위에서는 기준 서열에서의 위치에 상응하는 아미노산이 없을 것이다. 정렬된 기준 서열에서 삽입이 있는 경우, 해당 삽입은 기준 서열에서 임의의 아미노산 위치에 상응하지 않을 것이다. 말단 절단 또는 융합의 경우, 상응하는 서열에서 임의의 아미노산에 상응하지 않는 기준 또는 정렬된 서열에서의 아미노산 스트레치가 있을 수 있다.
- [0121] V. 항체
- [0122] 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 또는 117로 제시된 아미노산 서열을 갖는 것들을 비롯하여, RGN 폴리펩티드; 또는 본 발명의 RGN 폴리펩티드를 포함하는 리보핵단백질에 대한 항체, 또는 그의 활성 변이체 또는 단편이 또한 포함된다. 항체의 생성 방법 또한 관련 기술분야에 널리 공지되어 있다 (예를 들어, [Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.]; 및 미국 특허 번호 4,196,265 참고). 이들 항체는 RGN 폴리펩티드 또는 리보핵단백질의 검출 및 단리를 위한 키트에서 사용될 수 있다. 따라서, 본 개시내용은 본원에 기재된 폴리펩티드 또는 리보핵단백질, 예를 들어 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 또는 117의 서열을 갖는 폴리펩티드에 특이적으로 결합하는 항체를 포함하는 키트를 제공한다.
- [0123] VI. 관심 표적 서열에 결합하기 위한 시스템 및 리보핵단백질 복합체, 및 이를 제조하는 방법
- [0124] 본 개시내용은 관심 표적 서열에 결합하기 위한 시스템을 제공하고, 상기 시스템은 적어도 하나의 가이드 RNA 또는 이를 코딩하는 뉴클레오티드 서열, 및 적어도 하나의 RNA-가이드된 뉴클레아제 또는 이를 코딩하는 뉴클레오티드 서열을 포함한다. 가이드 RNA는 관심 표적 서열에 혼성화하고, 또한 RGN 폴리펩티드와 복합체를 형성하여, RGN 폴리펩티드가 표적 서열에 결합하도록 지시한다. 이들 일부 실시양태에서, RGN은 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137, 또는 235의 아미노산 서열, 또는 그의 활성 변이체 또는 단편을 포함한다. 다양한 실시양태에서, 가이드 RNA는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287의 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열, 또는 그의 활성 변이체 또는 단편을 포함한다. 특정한 실시양태에서, 가이드 RNA는 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274, 또는 286의 뉴클레오티드 서열을 포함하는 tracrRNA, 또는 그의 활성 변이체 또는 단편을 포함한다. 상기 시스템의 가이드 RNA는 단일 가이드 RNA 또는 이중-가이드 RNA일 수 있다. 특정한 실시양태에서, 상기 시스템은 가이드RNA에 대해 이중성인 RNA-가이드된 뉴클레아제를 포함하고, RGN 및 가이드RNA는 원래 자연적으로 복합체화되지 않는다.
- [0125] 본원에 제공된 관심 표적 서열에 결합하기 위한 시스템은 적어도 하나의 단백질에 결합된 RNA의 적어도 하나의 분자인 리보핵단백질 복합체일 수 있다. 본원에 제공된 리보핵단백질 복합체는 RNA 성분으로서 적어도 하나의 가이드 RNA 및 단백질 성분으로서 RNA-가이드된 뉴클레아제를 포함한다. 이러한 리보핵단백질 복합체는 RGN 폴리펩티드를 자연적으로 발현하고 관심 표적 서열에 대해 특이적인 특정한 가이드 RNA를 발현하도록 조작된 세포 또는 유기체로부터 정제될 수 있다. 대안적으로, 리보핵단백질 복합체는, RGN 폴리펩티드 및 가이드 RNA를 코딩하는 폴리뉴클레오티드로 형질전환되고 RGN 폴리펩티드 및 가이드 RNA의 발현을 허용하는 조건 하에 배양된 세포 또는 유기체로부터 정제될 수 있다. 따라서, RGN 폴리펩티드 또는 RGN 리보핵단백질 복합체의 제조 방법이 제공된다. 이러한 방법은 RGN 폴리펩티드를 코딩하는 뉴클레오티드 서열, 및 일부 실시양태에서 가이드 RNA를 코딩하는 뉴클레오티드 서열을 포함하는 세포를 RGN 폴리펩티드 (및 일부 실시양태에서 가이드 RNA)가 발현되는 조건 하에 배양하는 것을 포함한다. 이어서, RGN 폴리펩티드 또는 RGN 리보핵단백질을 배양된 세포의 용해물로부터 정제할 수 있다.
- [0126] 생물학적 샘플의 용해물로부터 RGN 폴리펩티드 또는 RGN 리보핵단백질 복합체를 정제하는 방법은 관련 기술분야에 공지되어 있다 (예를 들어, 크기 배제 및/또는 친화도 크로마토그래피, 2D-PAGE, HPLC, 역상 크로마토그래피, 면역침전). 특정한 방법에서, RGN 폴리펩티드는 재조합적으로 생성되고, 그의 정제에 도움이 되도록 정제 태그, 예컨대 비제한적으로 글루타티온-S-트랜스퍼라제 (GST), 키틴 결합 단백질 (CBP), 말토스 결합 단백질, 티오레독신 (TRX), 폴리(NANP), 직렬식 친화도 정제 (TAP) 태그, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis,

10xHis, 비오틴 카르복실 담체 단백질 (BCCP), 및 칼모듈린을 포함한다. 일반적으로, 태그 부착된 RGN 폴리펩티드 또는 RGN 리보핵단백질 복합체는 고정된 금속 친화도 크로마토그래피를 이용하여 정제된다. 관련 기술분야에 공지된 다른 유사한 방법, 예컨대 다른 형태의 크로마토그래피 또는 예를 들어 면역침전을 단독으로 또는 조합하여 이용할 있음을 이해할 것이다.

[0127] "단리된" 또는 "정제된" 폴리펩티드, 또는 그의 생물학적으로 활성인 부분은 그의 자연 발생 환경에서 발견되는 폴리펩티드에 일반적으로 동반되거나 또는 그와 상호작용하는 성분을 실질적으로 또는 본질적으로 함유하지 않는다. 따라서, 단리된 또는 정제된 폴리펩티드는 다른 세포 물질, 또는 재조합 기술에 의해 생성되는 경우 배양 배지를 실질적으로 함유하지 않거나, 또는 화학적으로 합성되는 경우 화학적 전구체 또는 다른 화학물질을 실질적으로 함유하지 않는다. 세포 물질을 실질적으로 함유하지 않는 단백질에는 약 30%, 20%, 10%, 5% 또는 1% (건조 중량 기준) 미만의 오염 단백질을 갖는 단백질 제제가 포함된다. 본 발명의 단백질 또는 그의 생물학적으로 활성인 부분이 재조합적으로 생성될 때, 최적으로 배양 배지는 약 30%, 20%, 10%, 5% 또는 1% (건조 중량 기준) 미만의 화학적 전구체 또는 관심 단백질이 아닌 화학물질을 나타낸다.

[0128] 관심 표적 서열의 결합 및/또는 절단을 위해 본원에 제공된 특정한 방법은 시험관내 조립된 RGN 리보핵단백질 복합체의 사용을 수반한다. RGN 리보핵단백질 복합체의 시험관내 조립체는, RGN 폴리펩티드와 가이드 RNA의 결합을 가능하게 하는 조건 하에서 RGN 폴리펩티드를 가이드 RNA와 접촉시키는 관련 기술분야에 공지된 임의의 방법을 이용하여 수행될 수 있다. 본원에서 사용된 바와 같이, "접촉하다", "접촉하는", "접촉된"은 원하는 반응을 수행하는데 적합한 조건 하에서 원하는 반응의 성분들을 함께 배치시키는 것을 지칭한다. RGN 폴리펩티드는 시험관내 번역을 통해 생성되거나 또는 화학적으로 합성된 생물학적 샘플, 세포 용해물 또는 배양 배지로부터 정제될 수 있다. 가이드 RNA는 시험관내에서 전사되거나 또는 화학적으로 합성된 생물학적 샘플, 세포 용해물 또는 배양 배지로부터 정제될 수 있다. RGN 폴리펩티드 및 가이드 RNA는 RGN 리보핵단백질 복합체의 시험관내 조립을 가능하게 하도록 용액 (예를 들어, 완충된 식염수 용액)과 접촉할 수 있다.

[0129] VII. 표적 서열의 결합, 절단 또는 변형 방법

[0130] 본 개시내용은 관심 표적 뉴클레오타이드 서열을 결합, 절단 및/또는 변형시키는 방법을 제공한다. 상기 방법은 적어도 하나의 가이드 RNA 또는 그를 코딩하는 폴리뉴클레오타이드, 및 적어도 하나의 RGN 폴리펩티드 또는 그를 코딩하는 폴리뉴클레오타이드를 포함하는 시스템을 표적 서열, 또는 표적 서열을 포함하는 세포, 소기관 또는 배아에 전달하는 것을 포함한다. 이들 일부 실시양태에서, RGN은 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137, 또는 235의 아미노산 서열, 또는 그의 활성 변이체 또는 단편을 포함한다. 다양한 실시양태에서, 가이드 RNA는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287의 뉴클레오타이드 서열을 포함하는 CRISPR 반복 서열, 또는 그의 활성 변이체 또는 단편을 포함한다. 특정한 실시양태에서, 가이드 RNA는 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274, 또는 286의 뉴클레오타이드 서열을 포함하는 tracrRNA, 또는 그의 활성 변이체 또는 단편을 포함한다. 시스템의 가이드 RNA는 단일 가이드 RNA 또는 이중-가이드 RNA일 수 있다. 시스템의 RGN은 뉴클레아제 데드 RGN일 수 있거나, Nickase 활성을 가질 수 있거나, 또는 융합 폴리펩티드일 수 있다. 일부 실시양태에서, 융합 폴리펩티드는 염기-편집 폴리펩티드, 예를 들어 시티딘 데아미나제 또는 아데노신 데아미나제를 포함한다. 다른 실시양태에서, RGN 융합 단백질은 리버스 트랜스크립타제를 포함한다. 다른 실시양태에서, RGN 융합 단백질은, 2020년 1월 27일에 출원된 미국 가출원 번호 62/966,203 (그 전문이 참조로 포함됨)에 기재된 바와 같이, 기능적 핵산 복구 복합체의 구성원, 예컨대 뉴클레오타이드 절제 복구 (NER) 또는 전사 커플링된-뉴클레오타이드 절제 복구 (TC-NER) 경로의 구성원을 동원하는 폴리펩티드를 포함한다 ([Wei et al., 2015, PNAS USA 112(27):E3495-504 ; Troelstra et al., 1992, Cell 71:939-953; Marnef et al., 2017, J Mol Biol 429(9):1277-1288]). 일부 실시양태에서, RGN 융합 단백질은 CSB ([van den Boom et al., 2004, J Cell Biol 166(1):27-36; van Gool et al., 1997, EMBO J 16(19):5955-65]; 그의 예는 서열식별번호: 268로 제시됨)를 포함하고, 이는 TC-NER (뉴클레오타이드 절제 복구) 경로의 구성원이며 다른 구성원의 동원에서 기능한다. 추가 실시양태에서, RGNABP 융합 단백질은 CSB의 활성 도메인, 예컨대 서열식별번호: 268의 아미노산 잔기 356-394를 포함하는 CSB의 산성 도메인을 포함한다 ([Teng et al., 2018, Nat Commun 9(1):4115]).

[0131] 특정한 실시양태에서, RGN 및/또는 가이드 RNA는, RGN 및/또는 가이드 RNA (또는 RGN 및 가이드 RNA 중 적어도 하나를 코딩하는 폴리뉴클레오타이드(들))가 도입되는 세포, 소기관 또는 배아에 대해 이중성이다.

[0132] 상기 방법이 가이드 RNA 및/또는 RGN 폴리펩티드를 코딩하는 폴리뉴클레오타이드를 전달하는 것을 포함하는 것인

이들 실시양태에서, 세포 또는 배아는 가이드 RNA 및/또는 RGN 폴리펩티드가 발현되는 조건 하에 배양될 수 있다. 다양한 실시양태에서, 상기 방법은 표적 서열을 RGN 리보핵단백질 복합체와 접촉시키는 것을 포함한다. RGN 리보핵단백질 복합체는 뉴클레아제 테드이거나 또는 니카제 활성을 갖는 RGN을 포함할 수 있다. 일부 실시양태에서, 리보핵단백질 복합체의 RGN은 염기-편집 폴리펩티드를 포함하는 융합 폴리펩티드이다. 특정한 실시양태에서, 상기 방법은 표적 서열을 포함하는 세포, 소기관 또는 배아에 RGN 리보핵단백질 복합체를 전달하는 것을 포함한다. RGN 리보핵단백질 복합체는 본원에 기재된 바와 같이 생물학적 샘플로부터 정제되고, 재조합적으로 생성되고, 후속적으로 정제되거나, 또는 시험관내에서 조립된 것일 수 있다. 표적 서열 또는 세포 소기관 또는 배아와 접촉하는 RGN 리보핵단백질 복합체가 시험관내에서 조립된 것인 이들 실시양태에서, 상기 방법은 표적 서열, 세포, 소기관 또는 배아와 접촉하기 전에 복합체의 시험관내 조립체를 추가로 포함할 수 있다.

[0133] 정제된 또는 시험관내 조립된 RGN 리보핵단백질 복합체를 관련 기술분야에 공지된 임의의 방법, 예컨대 비제한적으로 전기천공을 이용하여 세포, 소기관 또는 배아에 도입시킬 수 있다. 대안적으로, RGN 폴리펩티드, 및/또는 가이드 RNA를 코딩하거나 또는 포함하는 폴리뉴클레오티드를 관련 기술분야에 공지된 임의의 방법 (예를 들어, 전기천공)을 이용하여 세포, 소기관 또는 배아에 도입시킬 수 있다.

[0134] 표적 서열, 또는 표적 서열을 포함하는 세포, 소기관 또는 배아에 전달하거나 또는 그와 접촉시, 가이드 RNA는 RGN이 서열-특이적인 방식으로 표적 서열과 결합하도록 지시한다. RGN이 뉴클레아제 활성을 갖는 것인 이들 실시양태에서, RGN 폴리펩티드는 결합시 관심 표적 서열을 절단한다. 표적 서열은 제공된 공여자 폴리뉴클레오티드에 의해 내인성 복구 메커니즘, 예컨대 비상동성 말단 결합, 또는 상동성-지정된 복구를 통해 후속적으로 변형될 수 있다.

[0135] RGN 폴리펩티드와 표적 서열의 결합을 측정하는 방법은 관련 기술분야에 공지되어 있고, 염색질 면역침전 검정, 겔 이동성 변화 검정, DNA 풀다운 검정, 리포터 검정, 마이크로플레이트 포획 및 검출 검정이 포함된다. 마찬가지로, 표적 서열의 절단 또는 변형을 측정하는 방법은 관련 기술분야에 공지되어 있고, 시험관내 또는 생체내 절단 검정이 포함되며, 절단은 분해 생성물의 검출을 용이하게 하기 위해 표적 서열에 적절한 표지 (예를 들어, 방사성 동위원소, 형광 기질)를 부착시키거나 또는 부착시키지 않고 PCR, 서열분석, 또는 겔 전기영동에 의해 확인된다. 대안적으로, 니킹 촉발된 지수적 증폭 반응 (NTEXPAR) 검정을 이용할 수 있다 (예를 들어, [Zhang et al. (2016) Chem. Sci. 7:4951-4957] 참고). 생체내 절단은 서베이어(Surveyor) 검정을 이용하여 평가될 수 있다 (Guschin et al. (2010) Methods Mol Biol 649:247-256).

[0136] 일부 실시양태에서, 상기 방법은 1개 초과 가이드 RNA와 복합체화된 단일 유형의 RGN의 사용을 수반한다. 1개 초과 가이드 RNA는 단일 유전자의 상이한 영역을 표적화할 수 있거나 또는 다중 유전자를 표적화할 수 있다.

[0137] 공여자 폴리뉴클레오티드가 제공되지 않는 것인 이들 실시양태에서, RGN 폴리펩티드에 의해 도입된 이중-가닥 파손은 비상동성 말단-연결 (NHEJ) 복구 과정에 의해 복구될 수 있다. NHEJ의 오류가 발생하기 쉬운 성질 때문에, 이중-가닥 파손의 복구는 표적 서열에 대해 변형을 일으킬 수 있다. 본원에서 사용된 바와 같이, 핵산 분자와 관련하여 "변형"은 핵산 분자의 뉴클레오티드 서열에서의 변화를 지칭하고, 이는 하나 이상의 뉴클레오티드의 결실, 삽입 또는 치환, 또는 이들의 조합일 수 있다. 표적 서열의 변형은 변경된 단백질 생성물의 발현 또는 코딩 서열의 불활성화를 일으킬 수 있다.

[0138] 공여자 폴리뉴클레오티드가 존재하는 이들 실시양태에서, 공여자 폴리뉴클레오티드의 공여자 서열은 도입된 이중-가닥 파손의 복구 과정 동안에 표적 뉴클레오티드 서열에 통합되거나 또는 그와 교환되어, 외인성 공여자 서열의 도입을 일으킬 수 있다. 따라서, 공여자 폴리뉴클레오티드는 관심 표적 서열에 도입되는 것을 원하는 공여자 서열을 포함한다. 일부 실시양태에서, 공여자 서열은 새로 통합된 공여자 서열이 RGN에 의해 인식 및 절단되지 않도록 원래의 표적 뉴클레오티드 서열을 변경시킨다. 공여자 서열의 통합은 표적 뉴클레오티드 서열을 플랭킹하는 서열과 실질적인 서열 동일성을 갖는 플랭킹 서열을 갖는 공여자 폴리뉴클레오티드 내에 포함시킴으로써 증강될 수 있으며, 이는 상동성-지정된 복구 과정을 가능하게 한다. RGN 폴리펩티드가 이중-가닥 엇갈린 파손을 도입하는 것인 이들 실시양태에서, 공여자 폴리뉴클레오티드는 상동성 오버행에 의해 플랭킹된 공여자 서열을 포함할 수 있으며, 이는 이중-가닥 파손의 복구 동안에 비상동성 복구 과정에 의해 오버행을 포함하는 절단된 표적 뉴클레오티드 서열로의 공여자 서열의 직접적인 라이게이션을 가능하게 한다.

[0139] 상기 방법이 니카제인 (즉, 이중-가닥 폴리뉴클레오티드의 단일 가닥만을 절단하는) RGN의 사용을 수반하는 것인 이들 실시양태에서, 상기 방법은 동일한 또는 중복되는 표적 서열을 표적화하고 폴리뉴클레오티드의 상이한 가닥을 절단하는 2개의 RGN 니카제를 도입하는 것을 포함할 수 있다. 예를 들어, 이중-가닥 폴리뉴클레오티드

의 양성 (+) 가닥만을 절단하는 RGN Nickase는 이중-가닥 폴리뉴클레오타이드의 음성 (-) 가닥만을 절단하는 제2 RGN Nickase와 함께 도입될 수 있다.

[0140] 다양한 실시양태에서, 표적 뉴클레오타이드 서열과 결합시키고, 표적 서열을 검출하는 방법이 제공되며, 상기 방법은 세포, 소기관 또는 배아에 적어도 하나의 가이드 RNA 또는 이를 코딩하는 폴리뉴클레오타이드, 및 적어도 하나의 RGN 폴리펩티드 또는 이를 코딩하는 폴리뉴클레오타이드를 도입시키고, 가이드 RNA 및/또는 RGN 폴리펩티드 (코딩 서열이 도입된 경우)를 발현시키는 것을 포함하고, RGN 폴리펩티드는 뉴클레아제-데드 RGN이고 검출가능한 표지를 추가로 포함하며, 상기 방법은 검출가능한 표지를 검출하는 것을 추가로 포함한다. 검출가능한 표지는 융합 단백질 (예를 들어, 형광 단백질)로서 RGN에 융합될 수 있거나, 또는 시각적으로 또는 다른 수단에 의해 검출될 수 있는 RGN 폴리펩티드에 접합되거나 또는 그 안에 혼입된 소분자일 수 있다.

[0141] 표적 서열의 조절하에 표적 서열 또는 관심 유전자의 발현을 조절하는 방법 또한 본원에 제공된다. 상기 방법은 세포, 소기관 또는 배아에 적어도 하나의 가이드 RNA 또는 이를 코딩하는 폴리뉴클레오타이드, 및 적어도 하나의 RGN 폴리펩티드 또는 이를 코딩하는 폴리뉴클레오타이드를 도입시키고, 가이드 RNA 및/또는 RGN 폴리펩티드 (코딩 서열이 도입된 경우)를 발현시키는 것을 포함하고, RGN 폴리펩티드는 뉴클레아제-데드 RGN이다. 이들 일부 실시양태에서, 뉴클레아제-데드 RGN은 본원에 기재된 발현 조절인자 도메인 (즉, 후성적 변형 도메인, 전사 활성화 도메인 또는 전사 리프레서 도메인)을 포함하는 융합 단백질이다.

[0142] 본 개시내용은 또한 관심 표적 뉴클레오타이드 서열의 결합 및/또는 변형을 위한 방법을 제공한다. 상기 방법은 적어도 하나의 가이드 RNA 또는 이를 코딩하는 폴리뉴클레오타이드, 및 본 발명의 RGN 및 염기-편집 폴리펩티드, 예를 들어 시티딘 데아미나제 또는 아데노신 데아미나제를 포함하는 적어도 하나의 융합 폴리펩티드, 또는 융합 폴리펩티드를 코딩하는 폴리뉴클레오타이드를 포함하는 시스템을 표적 서열, 또는 표적 서열을 포함하는 세포, 소기관 또는 배아에 전달하는 것을 포함한다.

[0143] 관련 기술분야의 통상의 기술자는 본원에 개시된 임의의 방법을 이용하여 단일 표적 서열 또는 다중 표적 서열을 표적화할 수 있음을 이해할 것이다. 따라서, 상기 방법은 단일 유전자 내의 다중의 별개의 서열 및/또는 다중 유전자를 표적화할 수 있는 다중의 별개의 가이드 RNA와 조합하여 단일 RGN 폴리펩티드를 사용하는 것을 포함한다. 다중의 별개의 가이드 RNA를 다중의 별개의 RGN 폴리펩티드와 조합하여 도입하는 방법 또한 본원에 포함된다. 이들 가이드 RNA 및 가이드 RNA/RGN 폴리펩티드 시스템은 단일 유전자 내의 다중의 별개의 서열 및/또는 다중 유전자를 표적화할 수 있다.

[0144] 한 측면에서, 본 발명은 상기 방법 및 조성물에 개시된 임의의 하나 이상의 요소를 함유하는 키트를 제공한다. 일부 실시양태에서, 키트는 벡터 시스템, 및 키트의 사용에 대한 지침서를 포함한다. 일부 실시양태에서, 벡터 시스템은 (a) crRNA 서열을 코딩하는 DNA 서열 및 코딩된 crRNA 서열의 상류에 가이드 서열을 삽입하기 위한 1 개 이상의 삽입 부위에 작동가능하게 연결된 제1 조절 요소 (여기서, 발현되는 경우, 가이드 서열은 진핵 세포에서 표적 서열에 대한 CRISPR 복합체의 서열-특이적 결합을 지시하고, CRISPR 복합체는 가이드 RNA 폴리뉴클레오타이드와 복합체화된 CRISPR 효소를 포함함); 및/또는 (b) 핵 국제화 서열을 포함하는 상기 CRISPR 효소를 코딩하는 효소 코딩 서열에 작동가능하게 연결된 제2 조절 요소를 포함한다. 요소들은 개별적으로 또는 조합으로 제공될 수 있고, 임의의 적합한 용기, 예컨대 바이알, 병 또는 튜브에 제공될 수 있다.

[0145] 일부 실시양태에서, 키트는 1가지 이상의 언어로 된 지침서를 포함한다. 일부 실시양태에서, 키트는 본원에 기재된 요소 중 하나 이상을 이용하는 과정에서 사용하기 위한 하나 이상의 시약을 포함한다. 시약은 임의의 적합한 용기에 제공될 수 있다. 예를 들어, 키트는 하나 이상의 반응 또는 보관 완충제를 제공할 수 있다. 시약은 특정한 검정에서 이용가능한 형태로, 또는 사용하기 전에 하나 이상의 다른 성분의 첨가를 필요로 하는 형태로 (예를 들어 농축물 또는 동결건조된 형태로) 제공될 수 있다. 완충제는 임의의 완충제, 예컨대 비제한적으로 탄산나트륨 완충제, 중탄산나트륨 완충제, 붕산염 완충제, Tris 완충제, MOPS 완충제, HEPES 완충제, 및 이들의 조합물일 수 있다. 일부 실시양태에서, 완충제는 알칼리성이다. 일부 실시양태에서, 완충제는 약 7 내지 약 10의 pH를 갖는다.

[0146] 일부 실시양태에서, 키트는 가이드 서열 및 조절 요소와 작동가능하게 연결하기 위해 벡터에 삽입하기 위한 가이드 서열에 상응하는 하나 이상의 올리고뉴클레오타이드를 포함한다. 일부 실시양태에서, 키트는 상동성 재조합 주형 폴리뉴클레오타이드를 포함한다. 한 측면에서, 본 발명은 CRISPR 시스템의 하나 이상의 요소를 사용하는 방법을 제공한다. 본 발명의 CRISPR 복합체는 표적 폴리뉴클레오타이드를 변형시키는데 효과적인 수단을 제공한다. 본 발명의 CRISPR 복합체는 다양한 세포 유형에서 표적 폴리뉴클레오타이드를 변형시키는 (예를 들어, 결실, 삽입, 전좌, 불활성화, 활성화, 염기 편집) 것을 비롯하여 다양한 유용성을 갖는다. 따라서, 본 발명의 CRISPR

복합체는 예를 들어 유전자 요법, 약물 스크리닝, 질환 진단, 및 예후에서 광범위한 적용을 갖는다. 예시적인 CRISPR 복합체는 표적 폴리뉴클레오티드 내의 표적 서열에 혼성화된 가이드 서열과 복합체화된 CRISPR 효소를 포함한다.

[0147] VIII. 표적 폴리뉴클레오티드

[0148] 한 측면에서, 본 발명은 진핵 세포에서 표적 폴리뉴클레오티드를 변형시키는 방법을 제공하며, 이는 생체내, 생체외 또는 시험관내일 수 있다. 일부 실시양태에서, 상기 방법은 인간 또는 비-인간 동물 또는 식물 (예컨대 미세조류)로부터 세포 또는 세포 집단을 샘플링하고, 세포 또는 세포들을 변형시키는 것을 포함한다. 배양은 임의의 생체의 단계에서 일어날 수 있다. 세포 또는 세포들은 심지어 비-인간 동물 또는 식물 (예컨대 미세조류)에 재도입될 수 있다.

[0149] 자연적인 가변성을 이용하여, 식물 재배자는 바람직한 특성, 예컨대 수율, 품질, 균일성, 강도 및 해충에 대한 내성에 대해 가장 유용한 유전자들을 조합한다. 이들 바람직한 특성에는 성장, 광주기 선호도, 온도 요건, 꽃 또는 번식 발달의 개시일, 지방산 함량, 곤충 내성, 질환 내성, 선충 내성, 진균 내성, 제초제 내성, 다양한 환경 인자, 예컨대 가뭄, 열, 습기, 추위, 바람, 및 고염도를 비롯한 불리한 토양 조건에 대한 내성이 포함된다. 이들 유용한 유전자의 공급원에는 본래 또는 외래 품종, 가보(heirloom) 품종, 야생 식물 동족, 및 유도된 돌연변이, 예를 들어 돌연변이유발원으로 처리된 식물 물질이 포함된다. 본 발명을 이용하여, 식물 재배자는 돌연변이를 유도하는 새로운 도구를 제공한다. 따라서, 관련 기술분야의 기술자는 유용한 유전자의 공급원에 대한 계능을 분석할 수 있고, 원하는 특징 또는 특성을 갖는 품종에서 본 발명을 이용하여 이전의 돌연변이유발원에 비해 더욱 정밀하게 유용한 유전자의 증가를 유도할 수 있고, 따라서 식물 재배 프로그램을 가속화 및 개선시킬 수 있다.

[0150] RGN 시스템의 표적 폴리뉴클레오티드는 진핵 세포에 대해 내인성 또는 외인성인 임의의 폴리뉴클레오티드일 수 있다. 예를 들어, 표적 폴리뉴클레오티드는 진핵 세포의 핵에 존재하는 폴리뉴클레오티드일 수 있다. 표적 폴리뉴클레오티드는 유전자 생성물 (예를 들어, 단백질)을 코딩하는 서열 또는 비-코딩 서열 (예를 들어, 조절 폴리뉴클레오티드 또는 정크 DNA)일 수 있다. 이론에 구애되지 않지만, 표적 서열이 PAM (프로토스페이스인서 인접한 모티프); 즉, CRISPR 복합체에 의해 인식되는 짧은 서열과 회합되어야 하는 것으로 믿어진다. PAM에 대한 정확한 서열 및 길이 요건은 사용된 CRISPR 효소에 따라 다르지만, PAM은 전형적으로 프로토스페이스 (즉, 표적 서열)에 인접한 2-5개 염기쌍 서열이다.

[0151] CRISPR 복합체의 표적 폴리뉴클레오티드는 수많은 질환-연관된 유전자 및 폴리뉴클레오티드, 뿐만 아니라 신호 전달 생화학적 경로-연관된 유전자 및 폴리뉴클레오티드를 포함할 수 있다. 표적 폴리뉴클레오티드의 예에는 신호전달 생화학적 경로와 연관된 서열, 예를 들어 신호전달 생화학적 경로-연관된 유전자 또는 폴리뉴클레오티드가 포함된다. 표적 폴리뉴클레오티드의 예에는 질환 연관된 유전자 또는 폴리뉴클레오티드가 포함된다. "질환-연관된" 유전자 또는 폴리뉴클레오티드는, 비-질환 대조군의 조직 또는 세포와 비교하여 질환-발병 조직으로부터 유래된 세포에서 전사 또는 번역 생성물을 비정상적인 수준으로 또는 비정상적인 형태로 생성하는 임의의 유전자 또는 폴리뉴클레오티드를 지칭한다. 이는 비정상적으로 높은 수준으로 발현되는 유전자일 수 있고; 이는 비정상적으로 낮은 수준으로 발현되는 유전자일 수 있고, 변경된 발현은 질환의 발병 및/또는 진행과 상관관계가 있다. 질환-연관된 유전자는 또한 질환의 병인 (예를 들어, 원인 돌연변이)의 원인이 되는 유전자(들)에 직접적인 책임이 있거나 또는 그들과 연결 불균형 상태에 있는 돌연변이(들)를 보유하는 유전자 또는 유전자 변이를 지칭한다. 전사된 또는 번역된 생성물은 공지되어 있거나 또는 공지되어 있지 않을 수 있고, 추가로 정상 또는 비정상 수준일 수 있다. 질환-연관된 유전자 및 폴리뉴클레오티드의 예는 월드 와이드 웹(World Wide Web)으로 이용가능한 맥쿠식-나탄 인스티튜트 오브 제네틱 메디신, 존스 홉킨스 유니버시티(McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, 메릴랜드주 볼티모어) 및 내셔널 센터 포 바이오 테크놀로지 인포메이션, 내셔널 라이브러리 오브 메디신(National Center for Biotechnology Information, National Library of Medicine, 메릴랜드주 베데스다)으로부터 입수할 수 있다.

[0152] CRISPR 시스템이 관심 계능 서열로의 표적화에서 비교적 용이하다는 점에서 특히 유용하지만, RGN이 원인 돌연변이를 해결할 수 있는 것에 대한 문제가 여전히 남아 있다. 한 접근법은 RGN (바람직하게는 RGN의 불활성 또는 Nickase 변이체)과 염기-편집 효소 또는 염기 편집 효소의 활성 도메인, 예컨대 시티딘 데아미나제 또는 아데노신 데아미나제 염기 편집기 사이의 융합 단백질을 생성하는 것이다 (미국 특허 번호 9,840, 699, 본원에 참조로 포함됨). 일부 실시양태에서, 상기 방법은 DNA 분자를 (a) 본 발명의 RGN 및 염기-편집 폴리펩티드, 예컨대 데아미나제를 포함하는 융합 단백질을; 및 (b) (a)의 융합 단백질을 DNA 가닥의 표적 뉴클레오티드 서열로 표적화

하는 gRNA와 접촉시키는 것을 포함하며; 뉴클레오티드 염기를 탈아미노화시키는데 효과적인 양으로 및 그에 적합한 조건 하에 DNA 분자를 융합 단백질 및 gRNA와 접촉시킨다. 일부 실시양태에서, 표적 DNA 서열은 질환 또는 장애와 연관된 서열을 포함하고, 뉴클레오티드 염기의 탈아미노화는 질환 또는 장애와 연관된 서열을 생성한다. 일부 실시양태에서, 표적 DNA 서열은 농작물의 대립유전자에 존재하고, 관심 특성을 갖는 특정한 대립유전자는 더 적은 농경학적 가치를 갖는 식물을 생성한다. 뉴클레오티드 염기의 탈아미노화는 식물의 특성을 개선시키고 농경학적 가치를 증가시키는 대립유전자를 생성한다.

[0153] 일부 실시양태에서, DNA 서열은 질환 또는 장애와 연관된 T→C 또는 A→G 점 돌연변이를 포함하며, 돌연변이형 C 또는 G 염기의 탈아미노화는 질환 또는 장애와 연관된 서열을 생성한다. 일부 실시양태에서, 탈아미노화는 질환 또는 장애와 연관된 서열에서 점 돌연변이를 수정한다.

[0154] 일부 실시양태에서, 질환 또는 장애와 연관된 서열은 단백질을 코딩하고, 탈아미노화는 질환 또는 장애와 연관된 서열에 정지 코돈을 도입시켜, 코딩된 단백질의 말단 절단을 일으킨다. 일부 실시양태에서, 접촉은 질환 또는 장애를 갖기 쉬운, 갖고 있는 또는 가진 것으로 진단된 대상체의 생체 내에서 수행된다. 일부 실시양태에서, 질환 또는 장애는 게놈에서 점 돌연변이 또는 단일-염기 돌연변이와 연관된 질환이다. 일부 실시양태에서, 질환은 유전 질환, 암, 대사 질환, 또는 리소좀 저장 질환이다.

[0155] 염기-편집을 사용한 원인 돌연변이의 변형

[0156] 본 발명의 RGN-염기 편집기 융합 단백질에 의존하는 접근법을 이용하여 수정될 수 있는 유전적으로 유전된 질환의 예는 혈허 증후군이다. MPS-1로도 공지된 혈허 증후군은 α-L-이두로니다제 (IDUA)의 결핍으로 인한 것이며, 리소좀에서 데르마탄 술페이트 및 헤파란 술페이트가 분자 수준으로 축적되는 것을 특징으로 하는 리소좀 저장 질환을 일으킨다. 이 질환은 일반적으로 α-L-이두로니다제를 코딩하는 IDUA 유전자에서 돌연변이에 의해 유발된 유전적으로 유전된 장애이다. 일반적인 IDUA 돌연변이는 W402X 및 Q70X이며, 이들 둘 다 번역의 조기 종결을 일으키는 넌센스 돌연변이이다. 예를 들어 염기-편집 접근법에 의한 단일 뉴클레오티드의 복귀가 야생형 코딩 서열을 복구하고, 유전자좌의 내인성 조절 메커니즘에 의해 조절되는 단백질 발현을 일으키기 때문에, 이러한 돌연변이는 정밀한 게놈 편집 (PGE) 접근법에 의해 잘 해결된다. 추가로, 이형접합체가 무증상인 것으로 공지되어 있기 때문에, 돌연변이된 대립유전자 중 하나만을 수정하면 되므로, 이들 돌연변이 중 하나를 표적화하는 PGE 요법은 이 질환을 가진 많은 환자에게 유용할 것이다 ([Bunge et al. (1994) Hum. Mol. Genet. 3(6): 861-866], 본원에 참조로 포함됨).

[0157] 혈허 증후군에 대한 현재의 치료에는 효소 대체 요법 및 골수 이식이 포함된다 ([Vellodi et al. (1997) Arch. Dis. Child. 76(2): 92-99; Peters et al. (1998) Blood 91(7): 2601-2608] 본원에 참조로 포함됨). 효소 대체 요법이 혈허 증후군 환자의 생존 및 삶의 질에 극적인 영향을 미쳤지만, 이 접근법은 고가이며 시간 소모적인 매주 주입을 필요로 한다. 추가의 접근법에는 발현 백터 상의 IDUA 유전자의 전달 또는 혈청 알부민과 같이 고도로 발현된 유전자좌로의 유전자의 삽입이 포함된다 (미국 특허 번호 9,956,247, 본원에 참조로 포함됨). 그러나, 이들 접근법은 원래의 IDUA 유전자좌를 수정된 코딩 서열로 복구하지 않는다. 게놈-편집 전략은 수많은 이점을 가지며, 가장 주목할만하게는, 유전자 발현의 조절이 건강한 개체에서 존재하는 자연적인 메커니즘에 의해 조절된다는 것이다. 추가로, 염기 편집의 이용은 종양 억제 메커니즘의 방해에 의해 큰 염색체 재배열, 세포 사멸, 또는 종양 발생을 초래할 수 있는 이중 가닥 DNA 파손을 일으킬 필요가 없다. 일반적인 전략은 인간 게놈에서 특정 질환-유발 돌연변이를 표적화하고 수정하기 위해 본 발명의 RGN-염기 편집기 융합 단백질을 사용하는 것일 수 있다. 유전자 편집에 의해 수정될 수 있는 표적 질환에 대한 유사한 접근법 또한 추구될 수 있음을 이해할 것이다. 추가로, 다른 종, 특히 일반적인 가정 애완동물 또는 가축에서 질환-유발 돌연변이를 표적화하는 유사한 접근법 또한 본 발명의 RGN을 사용하여 이용될 수 있음을 이해할 것이다. 일반적인 가정 애완동물 및 가축에는 개, 고양이, 말, 돼지, 소, 양, 닭, 당나귀, 뱀, 페럿, 및 어류, 예전대 연어 및 새우가 포함된다.

[0158] 표적화된 결실에 의한 원인 돌연변이의 변형

[0159] 본 발명의 RGN은 또한 원인 돌연변이가 더욱 복잡한 인간 치료 접근법에서 유용할 수 있다. 예를 들어, 프리드라이히 운동실조 및 헌팅턴 질환과 같은 일부 질환은 유전자의 특정한 영역에서 3개 뉴클레오티드 모티프의 반복부에서의 유의한 증가의 결과이며, 이는 발현된 단백질이 기능하고 발현되는 능력에 영향을 미친다. 프리드라이히 운동실조 (FRDA)는 척수에서 신경 조직의 진행성 퇴행을 일으키는 상염색체성 열성 질환이다. 미토콘드리아에서 프라탁신 (FXN) 단백질의 감소된 수준은 세포 수준에서 산화성 손상 및 철 결핍을 초래한다. 감소된 FXN 발현은 체세포 및 생식선 FXN 유전자에서 인트론 1 내의 GAA 삼중항 확장과 관련이 있었다. FRDA

환자에서, GAA 반복부는 종종 70개 초과, 때때로 심지어 1000개 초과 (가장 흔하게는 600-900개)의 삼중항으로 구성되는 반면에, 발병하지 않은 개체는 약 40개 이하의 반복부를 갖는다 ([Pandolfo et al. (2012) Handbook of Clinical Neurology 103: 275-294; Campuzano et al. (1996) Science 271: 1423-1427; Pandolfo (2002) Adv. Exp. Med. Biol. 516: 99-118]; 모두 본원에 참조로 포함됨).

[0160] 프리드라이히 운동실조 (FRDA)를 초래하는 트리뉴클레오티드 반복 서열의 확장은 FRDA 불안정 영역으로 지칭되는 FXN 유전자 내의 정의된 유전자좌에서 발생한다. RNA 가이드된 뉴클레아제 (RGN)는 FRDA 환자 세포에서 불안정 영역을 절제하기 위해 사용될 수 있다. 이 접근법은 하기를 필요로 한다: 1) 인간 게놈에서 대립유전자를 표적화하도록 프로그래밍될 수 있는 RGN 및 가이드 RNA 서열; 및 2) RGN 및 가이드 서열에 대한 전달 접근법. 게놈 편집을 위해 사용되는 여러 뉴클레아제, 예컨대 흔히 사용되는 에스. 피로게네스(*S. pyogenes*)로부터의 Cas9 뉴클레아제 (SpCas9)는 본질적으로 기능적 발현 카세트를 위해 필요한 다른 유전자 요소 외에도 SpCas9 유전자 및 가이드 RNA의 길이를 고려할 때 너무 커서 아테노-연관된 바이러스 (AAV) 벡터에 패키징되지 않는다. 이는 SpCas9를 사용하는 접근법을 더 어렵게 만든다.

[0161] 본 발명의 특정 RNA 가이드된 뉴클레아제는 가이드 RNA와 함께 AAV 벡터 내로의 패키징에 매우 적합하다. 2개의 가이드 RNA를 패키징하는 것은 제2 벡터를 필요로 할 수도 있지만, 이러한 접근법은 두 벡터 사이의 단백질 서열을 분할하는데 요구될 수 있는 더 큰 뉴클레아제, 예컨대 SpCas9에서 요구되는 것보다 여전히 유리하게 비교된다. 본 발명은 게놈 불안정 영역이 제거되는 본 발명의 RGN을 사용하는 전략을 포함한다. 이러한 전략은 유사한 유전적 근거를 갖는 다른 질환 및 장애, 예컨대 헌팅턴 질환에 적용가능하다. 본 발명의 RGN을 이용하는 유사한 전략은 농경학적으로 또는 경제적으로 중요한 비-인간 동물, 예컨대 개, 고양이, 말, 돼지, 소, 양, 닭, 당나귀, 뱀, 페럿, 및 어류, 예컨대 연어 및 새우에서 유사한 질환 및 장애에 또한 적용가능하다.

[0162] 표적화된 돌연변이유발에 의한 원인 돌연변이의 변형

[0163] 본 발명의 RGN은 또한 유익한 효과를 일으킬 수 있는 파괴적 돌연변이를 도입하기 위한 것일 수 있다. 헤모글로빈을 코딩하는 유전자, 특히 베타 글로빈 쇠 (HBB 유전자)에서의 유전적 결함은 겸상 적혈구 빈혈 및 지중해 빈혈을 비롯하여 혈액소병증으로 공지된 수많은 질환의 원인이 될 수 있다.

[0164] 성인 인간에서, 헤모글로빈은 2개의 알파 (α)-유사 글로빈 쇠 및 2개의 베타 (β)-유사 글로빈 쇠 및 4개의 헴기를 포함하는 이종사합체이다. 성인에서, $\alpha_2\beta_2$ 사합체는 헤모글로빈 A (HbA) 또는 성인 헤모글로빈으로 지칭된다. 전형적으로, 알파 및 베타 글로빈 쇠는 대략 1:1 비로 합성되고, 이 비는 헤모글로빈 및 적혈구 (RBC) 안정화의 측면에서 중요한 것으로 보인다. 발달 중인 태아에서는, 산소가 모체의 혈류를 통해 아기의 시스템으로 전달될 수 있도록 헤모글로빈 A에 비해 산소에 대해 더 높은 결합 친화도를 갖는 상이한 형태의 헤모글로빈인 태아 헤모글로빈 (HbF)이 생성된다. 태아 헤모글로빈 또한 2개의 α 글로빈 쇠를 함유하지만, 성인 β -글로빈 쇠 대신에 이는 2개의 태아 감마 (γ)-글로빈 쇠를 갖는다 (즉, 태아 헤모글로빈은 $\alpha_2\gamma_2$ 임). 감마-글로빈에서 베타-글로빈 생성으로의 전환의 조절은 매우 복잡하며, 베타 글로빈 전사의 상향조절과 동시에 감마 글로빈 전사의 하향조절을 주로 수반한다. 대략 임신 30 주에, 태아에서 베타 글로빈의 생성이 증가하면서 감마 글로빈의 합성이 감소하기 시작한다. 일부 HbF는 성인기까지 지속되지만 (총 헤모글로빈의 대략 1-3%), 대략 10 개월째에 신생아의 헤모글로빈은 거의 모두 $\alpha_2\beta_2$ 이다. 혈액소병증을 가진 대부분의 환자에서는, 감마 글로빈을 코딩하는 유전자가 여전히 존재하지만, 상기 기재된 바와 같이 분만시 발생하는 정상 유전자 억제에 의해 발현은 비교적 낮다.

[0165] 겸상 적혈구 질환은 β 글로빈 유전자 (HBB)에서 V6E 돌연변이 (DNA 수준에서 GAG에서 GTG로)에 의해 유발되며, 생성된 헤모글로빈은 "헤모글로빈S" 또는 "HbS"로 지칭된다. 보다 낮은 산소 조건 하에서는, HbS 분자가 응집하고 섬유성 침전물을 형성한다. 이들 응집물은 RBC의 이상 또는 '겸상화'를 유발하여, 세포의 유연성을 상실한다. 겸상 RBC는 더이상 모세혈관계로 넣어질 수 없으며, 겸상 적혈구 환자에서 혈관 폐색 위기를 일으킬 수 있다. 또한, 겸상 RBC는 정상 RBC에 비해 더 약하고, 용혈의 경향을 가지며, 결국 환자에서 빈혈을 초래한다.

[0166] 겸상 적혈구 환자의 치료 및 관리는 항생제 치료, 통증 관리, 및 급성 에피소드 동안 수혈을 수반하는 평생의 일이다. 한 접근법은 히드록시우레아의 사용이며, 이는 감마 글로빈의 생성을 증가시킴으로써 부분적으로 그의 효과를 발휘한다. 만성 히드록시우레아 요법의 장기간 부작용은 아직 알려져 있지 않지만, 상기 치료는 원치않는 부작용을 일으키고, 환자마다 다양한 효능을 가질 수 있다. 겸상 적혈구 치료의 효능에서의 증가에도 불구하고, 환자의 기대 수명은 여전히 50대 중후반에 불과하며, 상기 질환과 연관된 이환율은 환자의 삶의 질에 큰 영향을 미친다.

- [0167] 지중해 빈혈 (알파 지중해 빈혈 및 베타 지중해 빈혈) 또한 헤모글로빈과 관련된 질환이며, 전형적으로 감소된 글로빈 쇠 발현을 수반한다. 이는 유전자의 조절 영역의 돌연변이를 통해, 또는 감소된 발현 또는 감소된 수준 또는 기능적 글로빈 단백질을 유발하는 글로빈 코딩 서열에서의 돌연변이로부터 발생할 수 있다. 지중해 빈혈의 치료에는 일반적으로 수혈 및 철 킬레이트 요법이 수반된다. 적절한 공여자가 확인될 수 있는 경우에는 골수 이식 또한 중증 지중해 빈혈을 가진 사람의 치료를 위해 이용되고 있지만, 이 절차는 상당한 위험을 가질 수 있다.
- [0168] SCD 및 베타 지중해 빈혈 둘 다의 치료에 대해 제안된 한 접근법은, HbF가 비정상적인 성인 헤모글로빈을 기능적으로 대체하도록 감마 글로빈의 발현을 증가시키는 것이다. 상기 언급된 바와 같이, 히드록시우레아에 의한 SCD 환자의 치료는 감마 글로빈 발현을 증가시키는 그의 효과로 인해 부분적으로 성공적인 것으로 여겨진다 ([DeSimone (1982) Proc Nat'l Acad Sci USA 79(14):4428-31; Ley, et al., (1982) N. Engl. J. Medicine, 307: 1469-1475; Ley, et al., (1983) Blood 62: 370-380; Constantoulakis et al., (1988) Blood 72(6):1961-1967], 모두 본원에 참조로 포함됨). HbF의 발현 증가는 그의 생성물이 감마 글로빈 발현을 조절하는 역할을 하는 유전자의 확인을 수반한다. 이러한 한 유전자는 BCL11A이다. BCL11A는 성인 적혈구 전구체 세포에서 발현되는 아연 핑거 단백질을 코딩하고, 그의 발현의 하향조절은 감마 글로빈 발현을 증가시킨다 ([Sankaran et al. (2008) Science 322: 1839], 본원에 참조로 포함됨). BCL11A 유전자에 대해 표적화된 억제성 RNA의 사용이 제안되었지만 (예를 들어, 미국 특허 공보 2011/0182867, 본원에 참조로 포함됨), 이 기술은 몇가지 잠재적인 단점을 가지며, 예를 들어 완전한 녹다운이 달성되지 않을 수 있고, 이러한 RNA의 전달이 문제가 될 수 있으며, RNA가 지속적으로 존재해야 하므로, 평생 여러 치료를 필요로 한다.
- [0169] 본 발명의 RGN을 이용하여 BCL11A 인핸서 영역을 표적화하여, BCL11A의 발현을 방해할 수 있고, 이로써 감마 글로빈 발현이 증가된다. 이러한 표적화된 방해는 비상동성 말단 결합 (NHEJ)에 의해 달성될 수 있으며, 이에 의해 본 발명의 RGN이 BCL11A 인핸서 영역 내의 특정한 서열에 대해 표적화되고, 이중-가닥 파손이 형성되고, 전형적으로 유해한 돌연변이를 동시에 도입시키면서 세포 기구가 파손을 복구한다. 다른 질환 표적에 대해 기재된 것과 유사하게, 본 발명의 RGN은 생체내 전달을 위해 RGN 및 그의 가이드 RNA에 대한 발현 카세트를 단일 AAV 벡터에 패키징할 수 있게 하는 그의 비교적 작은 크기 때문에 다른 공지된 RGN에 비해 이점을 가질 수 있다. 본 발명의 RGN을 이용하는 유사한 전략은 또한 인간 및 농경학적으로 또는 경제적으로 중요한 비-인간 동물 모두에서 유사한 질환 및 장애에 적용될 수 있다.
- [0170] IX. 폴리뉴클레오티드 유전자 변형을 포함하는 세포
- [0171] 본원에 기재된 RGN, crRNA, 및/또는 tracrRNA에 의해 매개되는 과정을 이용하여 변형된 관심 표적 서열을 포함하는 세포 및 유기체가 본원에 제공된다. 이들 일부 실시양태에서, RGN은 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117, 137, 또는 235의 아미노산 서열, 또는 그의 활성 변이체 또는 단편을 포함한다. 다양한 실시양태에서, 가이드 RNA는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118, 240, 273, 또는 287의 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열, 또는 그의 활성 변이체 또는 단편을 포함한다. 특정한 실시양태에서, 가이드 RNA는 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119, 241, 274, 또는 286의 뉴클레오티드 서열을 포함하는 tracrRNA, 또는 그의 활성 변이체 또는 단편을 포함한다. 상기 시스템의 가이드 RNA는 단일 가이드 RNA 또는 이중-가이드 RNA일 수 있다.
- [0172] 변형된 세포는 진핵생물 (예를 들어, 포유동물, 식물, 곤충 세포) 또는 원핵생물일 수 있다. 본원에 기재된 RGN, crRNA 및/또는 tracrRNA를 이용하는 과정에 의해 변형된 적어도 하나의 뉴클레오티드 서열을 포함하는 소기관 및 배아 또한 제공된다. 유전자 변형된 세포, 유기체, 소기관 및 배아는 변형된 뉴클레오티드 서열에 대해 이형접합성 또는 동형접합성일 수 있다.
- [0173] 세포, 유기체, 소기관 또는 배아의 염색체 변형은 변경된 발현 (상향조절 또는 하향조절), 불활성화, 또는 변경된 단백질 생성물 또는 통합된 서열의 발현을 일으킬 수 있다. 염색체 변형이 유전자의 불활성화 또는 비기능성 단백질 생성물의 발현을 일으키는 것인 이들 예에서, 유전자 변형된 세포, 유기체, 소기관 또는 배아는 "녹아웃"으로 지칭된다. 녹아웃된 표현형은 결실 돌연변이 (즉, 적어도 하나의 뉴클레오티드의 결실), 삽입 돌연변이 (즉, 적어도 하나의 뉴클레오티드의 삽입), 또는 넌센스 돌연변이 (즉, 정지 코돈이 도입되도록 하는 적어도 하나의 뉴클레오티드의 치환)의 결과일 수 있다.
- [0174] 대안적으로, 세포, 유기체, 소기관 또는 배아의 염색체 변형은 "녹인"을 생성할 수 있으며, 이는 단백질을 코딩하는 뉴클레오티드 서열의 염색체 통합으로부터 발생한다. 이들 일부 실시양태에서, 코딩 서열은 야생형 단백

질을 코딩하는 염색체 서열이 불활성화되지만 외인성으로 도입된 단백질이 발현되도록 염색체에 통합된다.

[0175] 다른 실시양태에서, 염색체 변형은 변이체 단백질 생성물을 생성한다. 발현된 변이체 단백질 생성물은 적어도 하나의 아미노산 치환 및/또는 적어도 하나의 아미노산의 부가 또는 결실을 가질 수 있다. 변경된 염색체 서열에 의해 코딩되는 변이체 단백질 생성물은 야생형 단백질, 예컨대 비제한적으로 변경된 효소 활성 또는 기질 특이성과 비교하여 변형된 특징 또는 활성을 나타낼 수 있다.

[0176] 여전히 다른 실시양태에서, 염색체 변형은 단백질의 변경된 발현 패턴을 일으킬 수 있다. 비제한적인 예로서, 단백질 생성물의 발현을 조절하는 조절 영역에서의 염색체 변경은 단백질 생성물의 과발현 또는 하향조절, 또는 변경된 조직 또는 일시적인 발현 패턴을 일으킬 수 있다.

[0177] X. 표적 DNA 또는 단일-가닥 DNA를 검출하기 위한 키트 및 방법

[0178] 일부 RGN (예를 들어, 서열식별번호: 54 및 137로 제시된 APG09106.1 및 APG09748)은 표적 DNA의 검출에 의해 활성화되면 무차별적으로 비-표적화 단일-가닥 DNA (ssDNA)를 절단할 수 있다. 따라서, 샘플 내 표적 DNA (이중-가닥 또는 단일-가닥)를 검출하기 위한 조성물 및 방법이 본원에 제공된다. 일부 실시양태에서, 목적하는 표적은 RNA, 예컨대 RNA 바이러스, 예컨대 예를 들어 코로나바이러스의 게놈 또는 게놈의 일부로서 존재할 수 있다. 일부 실시양태에서, 코로나바이러스는 SARS-유사 코로나바이러스일 수 있다. 추가 실시양태에서, 코로나바이러스는 SARS-CoV-2, SARS-CoV, 또는 박쥐 SARS-유사 코로나바이러스, 예컨대 박쥐-SL-CoVZC45 (수탁 MG772933)일 수 있다. 표적이 RNA로서 존재하는 실시양태에서, 표적은 RGN에 의해 효과적으로 표적화될 수 있는 DNA 분자로 역전사시킬 수 있다. 역전사에 이어 증폭 단계, 예컨대 열순환을 수반하는 관련 기술분야에 공지된 RT-PCR 방법이 이어질 수 있거나, 또는 등은 방법, 예컨대 RT-LAMP (역전사 루프-매개 등은 증폭)이 이어질 수 있다 ([Notomi et al., Nucleic Acids Res 28: E63, (2000)]).

[0179] 이들 조성물 및 방법은 가이드RNA와 혼성화하지 않고 비-표적 ssDNA인 검출 ssDNA의 사용을 수반한다. 일부 실시양태에서, 검출 ssDNA는 검출 ssDNA의 절단 후에 검출가능한 신호를 제공하는 검출가능한 표지를 포함한다. 비제한적 예는 형광단/켄처 쌍을 포함하는 검출 ssDNA이며, 여기서 형광단은 그의 신호가 극히 근접한 켄처의 존재에 의해 억제되기 때문에 검출 ssDNA가 전체 (즉, 비절단)인 경우에 형광을 발하지 않는다. 검출 ssDNA의 절단은 켄처의 제거를 유발하고, 이어서 형광 표지를 검출할 수 있다. 형광 표지 또는 염료의 비제한적 예는 Cy5, 플루오레세인 (예를 들어, FAM, 6 FAM, 5(6) FAM, FITC), Cy3, 알렉사 플루오르(Alexa Fluor)® 염료, 및 텍사스 레드를 포함한다. 켄처의 비제한적 예는 아이오와 블랙(Iowa Black)®FQ, 아이오와 블랙® RQ, Qx1 켄처, ATTO 켄처 및 QSY 염료를 포함한다. 일부 실시양태에서, 검출 ssDNA는 제2 켄처, 예컨대 ZEN™, TAO™, 및 블랙 홀 켄처(Black Hole Quencher)®와 같은 내부 켄처를 포함하며, 이는 배경을 낮추고 신호 검출을 증가시킬 수 있다.

[0180] 다른 실시양태에서, 검출 ssDNA는 검출 ssDNA의 절단 전에 검출가능한 신호를 제공하는 검출가능한 표지를 포함하고, ssDNA의 절단은 신호의 검출을 억제하거나 방지한다. 이러한 시나리오의 비제한적 예는 형광 공명 에너지 전달 (FRET) 쌍을 포함하는 검출 ssDNA이다. FRET는 무방사선 에너지 전달이 극히 근접하여 제1 (공여자) 형광단의 여기 상태에서부터 제2 (수용자) 형광단으로 일어나는 과정이다. 공여자 형광단의 방출 스펙트럼은 수용자 형광단의 여기 스펙트럼과 중첩된다. 따라서, 수용자 형광단은 검출 ssDNA가 전체 (즉, 비절단)일 때 형광을 발할 것이고, 수용자 형광단은 검출 ssDNA가 절단될 때 더 이상 형광을 발하지 않을 것인데, 이는 공여자 및 수용자 형광단이 더 이상 서로 극히 근접해 있지 않을 것이기 때문이다. FRET 공여자 및 수용자 형광단은 관련 기술분야에 공지되어 있고, 시안 형광 단백질 (CFP)/녹색 형광 단백질 (GFP), Cy3/Cy5, 및 GFP/황색 형광 단백질 (YFP)을 포함하나 이에 제한되지는 않는다.

[0181] 일부 실시양태에서, 검출 ssDNA는 약 2개 뉴클레오타이드 내지 약 30개 뉴클레오타이드, 예컨대 비제한적으로 약 2, 약 3, 약 4, 약 5, 약 6, 약 7, 약 8, 약 9, 약 10, 약 11, 약 12, 약 13, 약 14, 약 15, 약 16, 약 17, 약 18, 약 19, 약 20, 약 21, 약 22, 약 23, 약 24, 약 25개 뉴클레오타이드, 약 26개 뉴클레오타이드, 약 27개 뉴클레오타이드, 약 28개 뉴클레오타이드, 약 29개 뉴클레오타이드, 및 약 30개 뉴클레오타이드 길이를 갖는다.

[0182] DNA 분자의 표적 DNA를 검출하는 방법은 샘플을, RGN, RGN 및 DNA 분자 내의 표적 DNA 서열과 혼성화할 수 있는 가이드 RNA, 및 가이드 RNA와 혼성화하지 않는 검출 단일-가닥 DNA (ssDNA)와 접촉시킨 후, RGN에 의한 ssDNA의 절단에 의해 생성된 검출가능한 신호를 측정함으로써 DNA 분자의 표적 DNA 서열을 검출하는 것을 포함한다. 일부 실시양태에서, 방법은 RGN 및 가이드RNA와의 접촉 전에 또는 그와 동시에 샘플 내의 핵산 분자를 증폭시키는 단계를 포함할 수 있다. 이들 실시양태 중 일부에서, 가이드 RNA가 혼성화할 특정 서열이 검출 방법의 감도를

증가시키기 위해 증폭될 수 있다.

- [0183] 표적 DNA가 검출 ssDNA를 포함하는 이들 조성물 및 방법을 이용하여 검출될 수 있는 샘플은 핵산 (예를 들어, DNA 또는 RNA 분자)을 포함하거나 또는 이를 포함하는 것으로 여겨지는 임의의 샘플을 포함한다. 샘플은 정제된 핵산의 합성 조합, 또는 생물학적 샘플, 예컨대 호흡기 스왑 (예를 들어, 비인두 스왑) 추출물, 세포 용해물, 환자 샘플, 세포, 조직, 타액, 혈액, 혈청, 혈장, 소변, 흡인물, 생검 샘플, 뇌 척수액, 또는 유기체 (예를 들어, 박테리아, 바이러스)를 포함하는 임의의 공급원으로부터 유래될 수 있다.
- [0184] 샘플을 RGN, 가이드 RNA 및 검출 ssDNA와 접촉시키는 것은 시험관내, 생체외 또는 생체내에서 접촉시키는 것을 포함할 수 있다. 일부 실시양태에서, 검출 ssDNA 및/또는 RGN 및/또는 가이드 RNA는, 예를 들어 측방 유동 장치 상에 고정화되며, 여기서 샘플은 고정화된 검출 ssDNA 및/또는 RGN 및/또는 가이드 RNA와 접촉한다. 일부 실시양태에서, 검출 ssDNA 상의 항원 모이어티에 대한 항체는 절단된 검출 ssDNA를 무손상 검출 ssDNA와 차별화할 수 있는 방식으로, 예를 들어 측방 유동 장치 상에 고정된다.
- [0185] 일부 실시양태에서, 방법은 샘플에 존재하는 표적 DNA의 양을 측정하는 것을 추가로 포함할 수 있다. 시험 샘플 내 검출가능한 신호의 측정은 참조 측정 (예를 들어, 기지의 양의 표적 DNA를 포함하는 참조 샘플 또는 그의 시리즈의 측정)과 비교될 수 있다.
- [0186] 조성물 및 방법의 비제한적 적용에는 단일-뉴클레오타이드 다형성 (SNP) 검출, 암 스크리닝, 박테리아 감염의 검출, 항생제 내성의 검출, 및 바이러스 감염의 검출을 포함한다.
- [0187] RGN에 의한 ssDNA의 절단에 의해 생성된 검출가능한 신호는 형광 신호의 측정, 겔 상의 밴드의 시각적 분석, 비색 변화, 및 전기 신호의 존재 또는 부재를 포함하나 이에 제한되지는 않는 관련 기술분야에 공지된 임의의 적합한 방법을 사용하여 측정될 수 있다.
- [0188] 본 발명은 RGN 폴리펩티드, RGN 및 DNA 분자 내의 표적 DNA 서열과 혼성화할 수 있는 가이드 RNA, 및 가이드 RNA와 혼성화하지 않는 검출 ssDNA를 포함하는, 샘플 내 DNA 분자의 표적 DNA를 검출하기 위한 키트를 제공한다.
- [0189] 또한, DNA 분자의 표적 DNA 서열 및 복수의 비-표적 ssDNA를 포함하는 핵산 집단을 RGN, 및 RGN 및 표적 DNA 서열과 혼성화할 수 있는 가이드 RNA와 접촉시킴으로써 단일-가닥 DNA를 절단하는 방법이 본원에 제공된다.
- [0190] 단수 형태는 1개 또는 1개 초과 (즉, 적어도 하나의) 문법적 대상을 지칭하기 위해 본원에서 사용된다. 예를 들어, "폴리펩티드"는 하나 이상의 폴리펩티드들을 의미한다.
- [0191] 본 명세서에 언급된 모든 공보 및 특허 출원은 이 개시내용이 속하는 기술분야의 기술자의 수준을 나타낸다. 모든 공보 및 특허 출원은 각각의 개별 공보 또는 특허 출원이 구체적이고 개별적으로 참조로 포함되는 것으로 나타낸 것과 동일한 정도로 본원에 참조로 포함된다.
- [0192] 상기 발명이 이해의 명료함의 목적으로 설명 및 예시를 위해 일부 상세하게 기재되었지만, 특정 변화 및 변형이 첨부된 실시양태의 범위 내에서 실시될 수 있음이 명백할 것이다.
- [0193] 비제한적인 실시양태에는 하기가 포함된다:
- [0194] 1. RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드를 코딩하는 폴리뉴클레오타이드를 포함하는 핵산 분자이며, 여기서 상기 폴리뉴클레오타이드는 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 RGN 폴리펩티드를 코딩하는 뉴클레오타이드 서열을 포함하고;
- [0195] 여기서, 상기 RGN 폴리펩티드는 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA (gRNA)에 결합될 때 RNA-가이드 서열 특이적 방식으로 DNA 분자의 상기 표적 DNA 서열에 결합할 수 있고,
- [0196] 여기서, RGN 폴리펩티드를 코딩하는 상기 폴리뉴클레오타이드는 상기 폴리뉴클레오타이드에 대해 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.
- [0197] 2. 실시양태 1에 있어서, 상기 RGN 폴리펩티드가 결합시 상기 표적 DNA 서열을 절단할 수 있는 것인 핵산 분자.
- [0198] 3. 실시양태 2에 있어서, 상기 RGN 폴리펩티드가 이중-가닥 절단을 생성할 수 있는 것인 핵산 분자.
- [0199] 4. 실시양태 2에 있어서, 상기 RGN 폴리펩티드가 단일-가닥 절단을 생성할 수 있는 것인 핵산 분자.

- [0200] 5. 실시양태 2에 있어서, 상기 RGN 폴리펩티드가 뉴클레아제 불활성인 핵산 분자.
- [0201] 6. 실시양태 1-5 중 어느 하나에 있어서, RGN 폴리펩티드가 염기-편집 폴리펩티드에 작동가능하게 융합된 것인 핵산 분자.
- [0202] 7. 실시양태 6에 있어서, 염기-편집 폴리펩티드가 데아미나제인 핵산 분자.
- [0203] 8. 실시양태 7에 있어서, 데아미나제가 시티딘 데아미나제 또는 아데노신 데아미나제인 핵산 분자.
- [0204] 9. 실시양태 1-8 중 어느 하나에 있어서, RGN 폴리펩티드가 하나 이상의 핵 국제화 신호를 포함하는 것인 핵산 분자.
- [0205] 10. 실시양태 1-9 중 어느 하나에 있어서, RGN 폴리펩티드가 진핵 세포에서의 발현을 위해 코돈 최적화된 것인 핵산 분자.
- [0206] 11. 실시양태 1-10 중 어느 하나에 있어서, 상기 표적 DNA 서열이 프로토스페이서 인접한 모티프 (PAM)에 인접하게 위치하는 것인 핵산 분자.
- [0207] 12. 실시양태 1-11 중 어느 하나의 핵산 분자를 포함하는 벡터.
- [0208] 13. 실시양태 12에 있어서, 상기 표적 DNA 서열에 혼성화할 수 있는 상기 gRNA를 코딩하는 적어도 하나의 뉴클레오티드 서열을 추가로 포함하는 벡터.
- [0209] 14. 실시양태 13에 있어서, 가이드 RNA가 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 CRISPR 반복 서열을 포함하는 CRISPR RNA를 포함하는 것인 벡터.
- [0210] 15. 실시양태 13 또는 14에 있어서, 상기 gRNA가 tracrRNA를 포함하는 것인 벡터.
- [0211] 16. 실시양태 15에 있어서, tracrRNA가 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 또는 119에 대해 적어도 95% 서열 동일성을 갖는 것인 벡터.
- [0212] 17. 실시양태 15 또는 16에 있어서, 상기 gRNA가 단일 가이드 RNA인 벡터.
- [0213] 18. 실시양태 15 또는 16에 있어서, 상기 gRNA가 이중-가이드 RNA인 벡터.
- [0214] 19. 실시양태 1-11 중 어느 하나의 핵산 분자 또는 실시양태 12-18 중 어느 하나의 벡터를 포함하는 세포.
- [0215] 20. RGN 폴리펩티드가 발현되는 조건 하에 실시양태 18의 세포를 배양하는 것을 포함하는, RGN 폴리펩티드를 제조하는 방법.
- [0216] 21. 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드를 코딩하는 뉴클레오티드 서열을 포함하는 이중성 핵산 분자를 세포에 도입시키고;
- [0217] 여기서, 상기 RGN 폴리펩티드는 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA (gRNA)에 결합될 때 RNA-가이드 서열 특이적 방식으로 DNA 분자의 표적 DNA 서열에 결합할 수 있고;
- [0218] RGN 폴리펩티드가 발현되는 조건 하에 상기 세포를 배양하는 것
- [0219] 을 포함하는, RGN 폴리펩티드를 제조하는 방법.
- [0220] 22. 실시양태 20 또는 21에 있어서, 상기 RGN 폴리펩티드를 정제하는 것을 추가로 포함하는 방법.
- [0221] 23. 실시양태 20 또는 21에 있어서, 상기 세포가 상기 RGN 폴리펩티드에 결합하여 RGN 리보핵단백질 복합체를 형성하는 하나 이상의 가이드 RNA를 추가로 발현하는 것인 방법.
- [0222] 24. 실시양태 23에 있어서, 상기 RGN 리보핵단백질 복합체를 정제하는 것을 추가로 포함하는 방법.
- [0223] 25. CRISPR RNA (crRNA)를 코딩하는 폴리뉴클레오티드를 포함하는 핵산 분자이며, 여기서 상기 crRNA는 스페이서 서열 및 CRISPR 반복 서열을 포함하고, 상기 CRISPR 반복 서열은 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하고;

- [0224] 여기서,
- [0225] a) 상기 crRNA; 및 임의로
- [0226] b) 상기 crRNA의 상기 CRISPR 반복 서열에 혼성화할 수 있는 트랜스-활성화 CRISPR RNA (tracrRNA)
- [0227] 를 포함하는 가이드 RNA는
- [0228] 상기 가이드 RNA가 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드와 결합할 때 상기 crRNA의 스페이서 서열을 통해 서열 특이적인 방식으로 DNA 분자의 표적 DNA 서열에 혼성화할 수 있고,
- [0229] 여기서, 상기 crRNA를 코딩하는 폴리뉴클레오티드는 상기 폴리뉴클레오티드에 대해 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.
- [0230] 26. 실시양태 25의 핵산 분자를 포함하는 벡터.
- [0231] 27. 실시양태 26에 있어서, 상기 벡터가 상기 tracrRNA를 코딩하는 폴리뉴클레오티드를 추가로 포함하는 것인 벡터.
- [0232] 28. 실시양태 27에 있어서, 상기 tracrRNA가 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 또는 119에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 벡터.
- [0233] 29. 실시양태 27 또는 28에 있어서, 상기 crRNA를 코딩하는 상기 폴리뉴클레오티드 및 상기 tracrRNA를 코딩하는 상기 폴리뉴클레오티드가 동일한 프로모터에 작동가능하게 연결되고, 단일 가이드 RNA로서 코딩되는 것인 벡터.
- [0234] 30. 실시양태 27 또는 28에 있어서, 상기 crRNA를 코딩하는 상기 폴리뉴클레오티드 및 상기 tracrRNA를 코딩하는 상기 폴리뉴클레오티드가 별도의 프로모터에 작동가능하게 연결된 것인 벡터.
- [0235] 31. 실시양태 26-30 중 어느 하나에 있어서, 상기 벡터가 상기 RGN 폴리펩티드를 코딩하는 폴리뉴클레오티드를 추가로 포함하고, 상기 RGN 폴리펩티드가 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 벡터.
- [0236] 32. 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 또는 119에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 트랜스-활성화 CRISPR RNA (tracrRNA)를 코딩하는 폴리뉴클레오티드를 포함하는 핵산 분자이며;
- [0237] 여기서,
- [0238] a) 상기 tracrRNA; 및
- [0239] b) 스페이서 서열 및 CRISPR 반복 서열을 포함하는 crRNA를 포함하고, 여기서 상기 tracrRNA는 상기 crRNA의 상기 CRISPR 반복 서열과 혼성화할 수 있는 것인 가이드 RNA는
- [0240] 상기 가이드 RNA가 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드에 결합할 때 상기 crRNA의 스페이서 서열을 통해 서열 특이적인 방식으로 DNA 분자의 표적 DNA 서열에 혼성화할 수 있고,
- [0241] 여기서, 상기 tracrRNA를 코딩하는 폴리뉴클레오티드는 상기 폴리뉴클레오티드에 대해 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.
- [0242] 33. 실시양태 32의 핵산 분자를 포함하는 벡터.
- [0243] 34. 실시양태 33에 있어서, 상기 벡터가 상기 crRNA를 코딩하는 폴리뉴클레오티드를 추가로 포함하는 것인 벡터.
- [0244] 35. 실시양태 34에 있어서, 상기 crRNA의 CRISPR 반복 서열이 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 벡터.
- [0245] 36. 실시양태 34 또는 35에 있어서, 상기 crRNA를 코딩하는 상기 폴리뉴클레오티드 및 상기 tracrRNA를 코딩하는 상기 폴리뉴클레오티드가 동일한 프로모터에 작동가능하게 연결되고, 단일 가이드 RNA로서 코딩되는 것인 벡터.

- [0246] 37. 실시양태 34 또는 35에 있어서, 상기 crRNA를 코딩하는 상기 폴리뉴클레오타이드 및 상기 tracrRNA를 코딩하는 상기 폴리뉴클레오타이드가 별도의 프로모터에 작동가능하게 연결된 것인 벡터.
- [0247] 38. 실시양태 33-37 중 어느 하나에 있어서, 상기 벡터가 상기 RGN 폴리펩티드를 코딩하는 폴리뉴클레오타이드를 추가로 포함하고, 상기 RGN 폴리펩티드가 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 벡터.
- [0248] 39. DNA 분자의 표적 DNA 서열에 결합하기 위한 시스템이며, 상기 시스템은
- [0249] a) 상기 표적 DNA 서열에 혼성화할 수 있는 하나 이상의 가이드 RNA, 또는 상기 하나 이상의 가이드 RNA (gRNA)를 코딩하는 뉴클레오타이드 서열을 포함하는 하나 이상의 폴리뉴클레오타이드; 및
- [0250] b) 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드; 또는 상기 RGN 폴리펩티드를 코딩하는 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드
- [0251] 를 포함하고;
- [0252] 여기서, 상기 하나 이상의 가이드 RNA를 코딩하는 서열 및 상기 RGN 폴리펩티드를 코딩하는 뉴클레오타이드 서열은 각각 상기 뉴클레오타이드 서열에 이중성인 프로모터에 작동가능하게 연결되고;
- [0253] 여기서, 하나 이상의 가이드 RNA는 RGN 폴리펩티드와 복합체를 형성하여 상기 RGN 폴리펩티드가 DNA 분자의 상기 표적 DNA 서열에 결합하도록 지시할 수 있는 것인 시스템.
- [0254] 40. 실시양태 39에 있어서, 상기 gRNA가 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 CRISPR 반복 서열을 포함하는 것인 시스템.
- [0255] 41. 실시양태 39 또는 40에 있어서, 상기 gRNA가 tracrRNA를 포함하는 것인 시스템.
- [0256] 42. 실시양태 41에 있어서, 상기 tracrRNA가 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 또는 119에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 것인 시스템.
- [0257] 43. 실시양태 41 또는 42에 있어서, 상기 gRNA가 단일 가이드 RNA (sgRNA)인 시스템.
- [0258] 44. 실시양태 41 또는 42에 있어서, 상기 gRNA가 이중-가이드 RNA인 시스템.
- [0259] 45. 실시양태 39-44 중 어느 하나에 있어서, 상기 표적 DNA 서열이 프로토스페이서 인접 모티프 (PAM)에 인접하여 위치하는 것인 시스템.
- [0260] 46. 실시양태 39-45 중 어느 하나에 있어서, 표적 DNA 서열이 세포 내에 있는 것인 시스템.
- [0261] 47. 실시양태 46에 있어서, 세포가 진핵 세포인 시스템.
- [0262] 48. 실시양태 47에 있어서, 진핵 세포가 식물 세포인 시스템.
- [0263] 49. 실시양태 47에 있어서, 진핵 세포가 포유동물 세포인 시스템.
- [0264] 50. 실시양태 47에 있어서, 진핵 세포가 곤충 세포인 시스템.
- [0265] 51. 실시양태 46에 있어서, 세포가 원핵 세포인 시스템.
- [0266] 52. 실시양태 39-51 중 어느 하나에 있어서, 전사될 때 하나 이상의 가이드 RNA가 표적 DNA 서열에 혼성화할 수 있고, 가이드 RNA가 RGN 폴리펩티드와 복합체를 형성하여 표적 DNA 서열의 절단을 지시할 수 있는 것인 시스템.
- [0267] 53. 실시양태 52에 있어서, 상기 RGN 폴리펩티드가 이중-가닥 파손을 생성할 수 있는 것인 시스템.
- [0268] 54. 실시양태 52에 있어서, 상기 RGN 폴리펩티드가 단일-가닥 파손을 생성할 수 있는 것인 시스템.
- [0269] 55. 실시양태 52에 있어서, 상기 RGN 폴리펩티드가 뉴클레아제 불활성인 시스템.
- [0270] 56. 실시양태 39-55 중 어느 하나에 있어서, RGN 폴리펩티드가 염기-편집 폴리펩티드에 작동가능하게 연결된 것인 시스템.
- [0271] 57. 실시양태 56에 있어서, 염기-편집 폴리펩티드가 데아미나제인 시스템.

- [0272] 58. 실시양태 57에 있어서, 데아미나제가 시티딘 데아미나제 또는 아데노신 데아미나제인 시스템.
- [0273] 59. 실시양태 39-58 중 어느 하나에 있어서, RGN 폴리펩티드가 하나 이상의 핵 국제화 신호를 포함하는 것인 시스템.
- [0274] 60. 실시양태 39-59 중 어느 하나에 있어서, RGN 폴리펩티드가 진행 세포에서의 발현을 위해 코돈 최적화된 것인 시스템.
- [0275] 61. 실시양태 39-60 중 어느 하나에 있어서, 하나 이상의 가이드 RNA를 코딩하는 뉴클레오티드 서열을 포함하는 폴리뉴클레오티드 및 RGN 폴리펩티드를 코딩하는 뉴클레오티드 서열을 포함하는 폴리뉴클레오티드가 하나의 벡터 상에 위치하는 것인 시스템.
- [0276] 62. 실시양태 39-61 중 어느 하나에 있어서, 상기 시스템이 하나 이상의 공여자 폴리뉴클레오티드, 또는 하나 이상의 공여자 폴리뉴클레오티드를 코딩하는 하나 이상의 뉴클레오티드 서열을 추가로 포함하는 것인 시스템.
- [0277] 63. 실시양태 39-62 중 어느 하나에 따른 시스템을 표적 DNA 서열, 또는 표적 DNA 서열을 포함하는 세포에 전달하는 것을 포함하는, DNA 분자의 표적 DNA 서열에 결합시키는 방법.
- [0278] 64. 실시양태 63에 있어서, 상기 RGN 폴리펩티드 또는 상기 가이드 RNA가 검출가능한 표지를 추가로 포함하여, 상기 표적 DNA 서열의 검출을 가능하게 하는 것인 방법.
- [0279] 65. 실시양태 63에 있어서, 상기 가이드 RNA 또는 상기 RGN 폴리펩티드가 발현 조절인자를 추가로 포함하여, 상기 표적 DNA 서열, 또는 상기 표적 DNA 서열에 의한 전사 조절하에 있는 유전자의 발현을 조절하는 것인 방법.
- [0280] 66. 실시양태 39-62 중 어느 하나에 따른 시스템을 표적 DNA 서열 또는 DNA 분자를 포함하는 세포에 전달하는 것을 포함하며, 상기 표적 DNA 서열의 절단 또는 변형이 일어나는 것인, DNA 분자의 표적 DNA 서열을 절단 또는 변형시키는 방법.
- [0281] 67. 실시양태 66에 있어서, 상기 변형된 표적 DNA 서열이 표적 DNA 서열로의 이중성 DNA의 삽입을 포함하는 것인 방법.
- [0282] 68. 실시양태 66에 있어서, 상기 변형된 표적 DNA 서열이 표적 DNA 서열로부터 적어도 하나의 뉴클레오티드의 결실을 포함하는 것인 방법.
- [0283] 69. 실시양태 66에 있어서, 상기 변형된 표적 DNA 서열이 표적 DNA 서열에서 적어도 하나의 뉴클레오티드의 돌연변이를 포함하는 것인 방법.
- [0284] 70. DNA 분자의 표적 DNA 서열에 결합시키는 방법이며, 상기 방법은
- [0285] a) RNA-가이드된 뉴클레아제 (RGN) 리보뉴클레오티드 복합체의 형성에 적합한 조건 하에
- [0286] i) 표적 DNA 서열에 혼성화할 수 있는 하나 이상의 가이드 RNA; 및
- [0287] ii) 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산을 포함하는 RGN 폴리펩티드
- [0288] 를 조합하여 시험관내에서 RGN 리보뉴클레오티드 복합체를 조립하는 단계; 및
- [0289] b) 상기 표적 DNA 서열 또는 상기 표적 DNA 서열을 포함하는 세포를 시험관내-조립된 RGN 리보뉴클레오티드 복합체와 접촉시키는 단계
- [0290] 를 포함하고:
- [0291] 여기서, 하나 이상의 가이드 RNA는 표적 DNA 서열에 혼성화하여, 상기 RGN 폴리펩티드가 상기 표적 DNA 서열에 결합하도록 지시하는 것인 방법.
- [0292] 71. 실시양태 70에 있어서, 상기 RGN 폴리펩티드 또는 상기 가이드 RNA가 검출가능한 표지를 추가로 포함하여, 상기 표적 DNA 서열의 검출을 가능하게 하는 것인 방법.
- [0293] 72. 실시양태 70에 있어서, 상기 가이드 RNA 또는 상기 RGN 폴리펩티드가 발현 조절인자를 추가로 포함하여, 상기 표적 DNA 서열의 발현의 조절을 가능하게 하는 것인 방법.
- [0294] 73. DNA 분자의 표적 DNA 서열을 절단 및/또는 변형시키는 방법이며, 상기 방법은 DNA 분자를

- [0295] a) 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산을 포함하는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드; 및
- [0296] b) (a)의 RGN을 표적 DNA 서열에 표적화할 수 있는 하나 이상의 가이드 RNA
- [0297] 와 접촉시키는 것을 포함하고;
- [0298] 여기서, 하나 이상의 가이드 RNA는 표적 DNA 서열에 혼성화하여, 상기 RGN 폴리펩티드가 상기 표적 DNA 서열에 결합하도록 지시하고, 상기 표적 DNA 서열의 절단 및/또는 변형이 일어나는 것인 방법.
- [0299] 74. 실시양태 73에 있어서, 상기 RGN 폴리펩티드에 의한 절단이 이중-가닥 파손을 생성하는 것인 방법.
- [0300] 75. 실시양태 73에 있어서, 상기 RGN 폴리펩티드에 의한 절단이 단일-가닥 파손을 생성하는 것인 방법.
- [0301] 76. 실시양태 73에 있어서, 상기 RGN 폴리펩티드가 뉴클레아제 불활성인 방법.
- [0302] 77. 실시양태 73-76 중 어느 하나에 있어서, 상기 RGN 폴리펩티드가 염기-편집 폴리펩티드에 작동가능하게 연결된 것인 방법.
- [0303] 78. 실시양태 77에 있어서, 상기 염기-편집 폴리펩티드가 데아미나제를 포함하는 것인 방법.
- [0304] 79. 실시양태 78에 있어서, 상기 데아미나제가 시티딘 데아미나제 또는 아데노신 데아미나제인 방법.
- [0305] 80. 실시양태 73-79 중 어느 하나에 있어서, 상기 변형된 표적 DNA 서열이 표적 DNA 서열로의 이중성 DNA의 삽입을 포함하는 것인 방법.
- [0306] 81. 실시양태 73 내지 79 중 어느 하나에 있어서, 상기 변형된 표적 DNA 서열이 표적 DNA 서열로부터의 1개 이상의 뉴클레오타이드의 결실을 포함하는 것인 방법.
- [0307] 82. 실시양태 73-79 중 어느 하나에 있어서, 상기 변형된 표적 DNA 서열이 표적 DNA 서열에서 적어도 하나의 뉴클레오타이드의 돌연변이를 포함하는 것인 방법.
- [0308] 83. 실시양태 70-82 중 어느 하나에 있어서, 상기 gRNA가 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 CRISPR 반복 서열을 포함하는 것인 방법.
- [0309] 84. 실시양태 70-83 중 어느 하나에 있어서, 상기 gRNA가 tracrRNA를 포함하는 것인 방법.
- [0310] 85. 실시양태 84에 있어서, 상기 tracrRNA가 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 또는 119에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 것인 방법.
- [0311] 86. 실시양태 84 또는 85에 있어서, 상기 gRNA가 단일 가이드 RNA (sgRNA)인 방법.
- [0312] 87. 실시양태 84 또는 85에 있어서, 상기 gRNA가 이중-가이드 RNA인 방법.
- [0313] 88. 실시양태 63-87 중 어느 하나에 있어서, 상기 표적 DNA 서열이 프로토스페이서 인접한 모티프 (PAM)에 인접하게 위치하는 것인 방법.
- [0314] 89. 실시양태 63-88 중 어느 하나에 있어서, 표적 DNA 서열이 세포 내에 위치하는 것인 방법.
- [0315] 90. 실시양태 89에 있어서, 세포가 진핵 세포인 방법.
- [0316] 91. 실시양태 90에 있어서, 진핵 세포가 식물 세포인 방법.
- [0317] 92. 실시양태 90에 있어서, 진핵 세포가 포유동물 세포인 방법.
- [0318] 93. 실시양태 90에 있어서, 진핵 세포가 곤충 세포인 방법.
- [0319] 94. 실시양태 89에 있어서, 세포가 원핵 세포인 방법.
- [0320] 95. 실시양태 89-94 중 어느 하나에 있어서, RGN 폴리펩티드가 발현되는 조건 하에 세포를 배양하여, 표적 DNA 서열을 절단하여 변형된 DNA 서열을 포함하는 DNA 분자를 생성하고; 상기 변형된 표적 DNA 서열을 포함하는 세포를 선택하는 것을 추가로 포함하는 방법.
- [0321] 96. 실시양태 95의 방법에 따른 변형된 표적 DNA 서열을 포함하는 세포.

- [0322] 97. 실시양태 96에 있어서, 세포가 진핵 세포인 세포.
- [0323] 98. 실시양태 97에 있어서, 진핵 세포가 식물 세포인 세포.
- [0324] 99. 실시양태 98의 세포를 포함하는 식물.
- [0325] 100. 실시양태 98의 세포를 포함하는 종자.
- [0326] 101. 실시양태 97에 있어서, 진핵 세포가 포유동물 세포인 세포.
- [0327] 102. 실시양태 98에 있어서, 진핵 세포가 곤충 세포인 세포.
- [0328] 103. 실시양태 96에 있어서, 세포가 원핵 세포인 세포.
- [0329] 104. 유전적으로 유전된 질환에 대한 원인 돌연변이에서의 수정에 의해 유전자 변형된 세포를 생성하는 방법이며, 상기 방법은 세포에
- [0330] a) RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드, 또는 상기 RGN 폴리펩티드를 코딩하는 폴리뉴클레오티드 - 여기서 RGN 폴리펩티드는 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하고, RGN 폴리펩티드를 코딩하는 상기 폴리뉴클레오티드는 세포에서 RGN 폴리펩티드의 발현이 가능하도록 프로모터에 작동가능하게 연결됨; 및
- [0331] b) 가이드 RNA (gRNA), 또는 상기 gRNA를 코딩하는 폴리뉴클레오티드 - 여기서 gRNA는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열을 포함하고, gRNA를 코딩하는 상기 폴리뉴클레오티드는 세포에서 gRNA의 발현이 가능하도록 프로모터에 작동가능하게 연결됨
- [0332] 를 도입시키는 것을 포함하고;
- [0333] 이로써 RGN 및 gRNA는 원인 돌연변이의 게놈 위치에 대해 표적화되고, 원인 돌연변이를 제거하도록 게놈 서열을 변형시키는 것인 방법.
- [0334] 105. 실시양태 104에 있어서, RGN이 염기-편집 폴리펩티드에 작동가능하게 연결된 것인 방법.
- [0335] 106. 실시양태 105에 있어서, 염기-편집 폴리펩티드가 데아미나제인 방법.
- [0336] 107. 실시양태 106에 있어서, 데아미나제가 시딘인 데아미나제 또는 아데노신 데아미나제인 방법.
- [0337] 108. 실시양태 104-107 중 어느 하나에 있어서, 세포가 동물 세포인 방법.
- [0338] 109. 실시양태 104-107 중 어느 하나에 있어서, 동물 세포가 포유동물 세포인 방법.
- [0339] 110. 실시양태 108에 있어서, 세포가 개, 고양이, 마우스, 래트, 토끼, 말, 소, 돼지, 또는 인간으로부터 유래되는 것인 방법.
- [0340] 111. 실시양태 108에 있어서, 유전적으로 유전된 질환이 단일 뉴클레오티드 다형성에 의해 유발되는 것인 방법.
- [0341] 112. 실시양태 111에 있어서, 유전적으로 유전된 질환이 헐러 증후군인 방법.
- [0342] 113. 실시양태 112에 있어서, gRNA가 원인 단일 뉴클레오티드 다형성에 근접한 영역을 표적화하는 스페이서 서열을 추가로 포함하는 것인 방법.
- [0343] 114. 질환을 유발하는 불안정 게놈 영역에서 결실을 갖는 유전자 변형된 세포를 생성하는 방법이며, 상기 방법은 세포에
- [0344] a) RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드, 또는 상기 RGN 폴리펩티드를 코딩하는 폴리뉴클레오티드 - 여기서 RGN 폴리펩티드는 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하고, RGN 폴리펩티드를 코딩하는 상기 폴리뉴클레오티드는 세포에서 RGN 폴리펩티드의 발현이 가능하도록 프로모터에 작동가능하게 연결됨; 및
- [0345] b) 가이드 RNA (gRNA), 또는 상기 gRNA를 코딩하는 폴리뉴클레오티드 - 여기서 gRNA는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열을 포함하고, gRNA를 코딩하는 상기 폴리뉴클레오티드는 세포에서 gRNA의 발현이 가능하도록 프로모터에 작동가능하게 연결되고, 추가로 gRNA는 불안정 게놈 영역의 5' 플랭크

를 표적화하는 스페이서 서열을 포함함; 및

- [0346] c) 제2 가이드 RNA (gRNA), 또는 상기 gRNA를 코딩하는 폴리뉴클레오티드 - 여기서 gRNA는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열을 포함하고, gRNA를 코딩하는 상기 폴리뉴클레오티드는 세포에서 제2 gRNA의 발현이 가능하도록 프로모터에 작동가능하게 연결되고, 추가로 상기 제2 gRNA는 불안정 게놈 영역의 3' 플랭크를 표적화하는 스페이서 서열을 포함함
- [0347] 를 도입하는 것을 포함하고;
- [0348] 이로써 RGN 및 2개의 gRNA는 불안정 게놈 영역을 표적화하고, 불안정 게놈 영역의 적어도 일부분이 제거되는 것인 방법.
- [0349] 115. 실시양태 114에 있어서, 세포가 동물 세포인 방법.
- [0350] 116. 실시양태 114에 있어서, 세포가 포유동물 세포인 방법.
- [0351] 117. 실시양태 115에 있어서, 세포가 개, 고양이, 마우스, 래트, 토끼, 말, 소, 돼지, 또는 인간으로부터 유래되는 것인 방법.
- [0352] 118. 실시양태 115에 있어서, 유전적으로 유전된 질환이 프리드라이히 운동실조 또는 헌팅턴 질환인 방법.
- [0353] 119. 실시양태 118에 있어서, 제1 gRNA가 불안정 게놈 영역 내의 또는 그에 근접한 영역을 표적화하는 스페이서 서열을 추가로 포함하는 것인 방법.
- [0354] 120. 실시양태 118에 있어서, 제2 gRNA가 불안정 게놈 영역 내의 또는 그에 근접한 영역을 표적화하는 스페이서 서열을 추가로 포함하는 것인 방법.
- [0355] 121. 감소된 BCL11A mRNA 및 단백질 발현을 갖는 유전자 변형된 포유동물의 조혈 전구 세포를 생성하는 방법이며, 상기 방법은 단리된 인간 조혈 전구 세포에
- [0356] a) RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드, 또는 상기 RGN 폴리펩티드를 코딩하는 폴리뉴클레오티드 - 여기서 RGN 폴리펩티드는 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하고, RGN 폴리펩티드를 코딩하는 상기 폴리뉴클레오티드는 세포에서 RGN 폴리펩티드의 발현이 가능하도록 프로모터에 작동가능하게 연결됨; 및
- [0357] b) 가이드 RNA (gRNA), 또는 상기 gRNA를 코딩하는 폴리뉴클레오티드 - 여기서 gRNA는 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 또는 118에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열을 포함하고, gRNA를 코딩하는 상기 폴리뉴클레오티드는 세포에서 gRNA의 발현이 가능하도록 프로모터에 작동가능하게 연결됨
- [0358] 를 도입하는 것을 포함하고,
- [0359] 이로써 RGN 및 gRNA는 세포에서 발현되고, BCL11A 인핸서 영역에서 절단하여, 인간 조혈 전구 세포의 유전자 변형을 일으키고, BCL11A의 mRNA 및/또는 단백질 발현을 감소시키는 것인 방법.
- [0360] 122. 실시양태 121에 있어서, gRNA가 BCL11A 인핸서 영역 내의 또는 그에 근접한 영역을 표적화하는 스페이서 서열을 추가로 포함하는 것인 방법.
- [0361] 123. 실시양태 104-122 중 어느 하나에 있어서, 가이드 RNA가 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112 또는 119에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 tracrRNA를 포함하는 것인 방법.
- [0362] 124. DNA 분자의 표적 DNA 서열에 결합하기 위한 시스템이며, 상기 시스템은
- [0363] a) 상기 표적 DNA 서열에 혼성화할 수 있는 하나 이상의 가이드 RNA, 또는 하나 이상의 가이드 RNA (gRNA)를 코딩하는 하나 이상의 뉴클레오티드 서열을 포함하는 하나 이상의 폴리뉴클레오티드; 및
- [0364] b) 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110 또는 117에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드
- [0365] 를 포함하고;

- [0366] 여기서, 하나 이상의 가이드 RNA는 표적 DNA 서열에 혼성화할 수 있고,
- [0367] 여기서, 하나 이상의 가이드 RNA는 RGN 폴리펩티드와 복합체를 형성하여 상기 RGN 폴리펩티드가 DNA 분자의 상기 표적 DNA 서열에 결합하도록 지시할 수 있는 것인 시스템.
- [0368] 125. 실시양태 124에 있어서, 상기 RGN 폴리펩티드가 뉴클레아제 불활성이거나 또는 니카제로서 기능할 수 있는 것인 시스템.
- [0369] 126. 실시양태 124 또는 125에 있어서, 상기 RGN 폴리펩티드가 염기-편집 폴리펩티드에 작동가능하게 융합된 것인 시스템.
- [0370] 127. 실시양태 126에 있어서, 염기-편집 폴리펩티드가 데아미나제인 시스템.
- [0371] 128. 실시양태 127에 있어서, 데아미나제가 시티딘 데아미나제 또는 아데노신 데아미나제인 시스템.
- [0372] 129. 샘플에서 DNA 분자의 표적 DNA 서열을 검출하는 방법이며, 상기 방법은
- [0373] a) 샘플을
- [0374] i) 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117 또는 137에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 상기 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA에 결합될 때 RNA-가이드된 서열 특이적 방식으로 DNA 분자의 상기 표적 DNA 서열에 결합할 수 있는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드;
- [0375] ii) 상기 가이드 RNA; 및
- [0376] iii) 가이드 RNA와 혼성화하지 않는 검출 단일-가닥 DNA (ssDNA)
- [0377] 와 접촉시키는 단계; 및
- [0378] b) RGN에 의한 검출 ssDNA의 절단에 의해 생성된 검출가능한 신호를 측정하여 표적 DNA를 검출하는 단계를 포함하는 방법.
- [0379] 130. 실시양태 129에 있어서, 상기 샘플이 세포 용해물로부터의 DNA 분자를 포함하는 것인 방법.
- [0381] 131. 실시양태 129에 있어서, 상기 샘플이 세포를 포함하는 것인 방법.
- [0382] 132. 실시양태 131에 있어서, 상기 세포가 진핵 세포인 방법.
- [0383] 133. 실시양태 129에 있어서, 표적 DNA 서열을 포함하는 DNA 분자가 RNA를 포함하는 샘플에 존재하는 RNA 주형 분자의 역전사에 의해 생성된 것인 방법.
- [0384] 134. 실시양태 133에 있어서, RNA 주형 분자가 RNA 바이러스인 방법.
- [0385] 135. 실시양태 134에 있어서, RNA 바이러스가 코로나바이러스인 방법.
- [0386] 136. 실시양태 135에 있어서, 코로나바이러스가 박쥐 SARS-유사 코로나바이러스, SARS-CoV, 또는 SARS-CoV-2인 방법.
- [0387] 137. 실시양태 133-136 중 어느 하나에 있어서, RNA를 포함하는 샘플이 세포를 포함하는 샘플로부터 유래된 것인 방법.
- [0388] 138. 실시양태 129-137 중 어느 하나에 있어서, 상기 검출 ssDNA가 형광단/켄처 쌍을 포함하는 것인 방법.
- [0389] 139. 실시양태 129-137 중 어느 하나에 있어서, 상기 검출 ssDNA가 형광 공명 에너지 전달 (FRET) 쌍을 포함하는 것인 방법.
- [0390] 140. 실시양태 129-139 중 어느 하나에 있어서, 상기 가이드 RNA가 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118 또는 273에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열을 포함하는 것인 방법.
- [0391] 141. 실시양태 129-140 중 어느 하나에 있어서, 상기 가이드 RNA가 tracrRNA를 포함하는 것인 방법.
- [0392] 142. 실시양태 141에 있어서, 상기 tracrRNA가 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119 또는 274에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것

인 방법.

- [0393] 143. 실시양태 141 또는 142에 있어서, 상기 가이드 RNA가 단일 가이드 RNA인 방법.
- [0394] 144. 실시양태 141 또는 142에 있어서, 상기 가이드 RNA가 이중-가이드 RNA인 방법.
- [0395] 145. 실시양태 129-144 중 어느 하나에 있어서, 단계 a)의 접촉 전에 또는 접촉과 함께 샘플에서 핵산을 증폭시키는 것을 추가로 포함하는 방법.
- [0396] 146. 실시양태 145에 있어서, 증폭이 RNA 분자의 역전사에 의해 생성된 DNA 분자의 증폭인 방법.
- [0397] 147. 샘플에서 DNA 분자의 표적 DNA 서열을 검출하기 위한 키트이며, 상기 키트는
 - a) 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117 또는 137에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 상기 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA에 결합될 때 RNA-가이드된 서열 특이적 방식으로 DNA 분자의 상기 표적 DNA 서열에 결합할 수 있는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드;
 - b) 상기 가이드 RNA; 및
 - c) 가이드 RNA와 혼성화하지 않는 검출 단일-가닥 DNA (ssDNA)
- [0401] 를 포함하는 키트.
- [0402] 148. 실시양태 147에 있어서, 상기 검출 ssDNA가 형광단/켄처 쌍을 포함하는 것인 키트.
- [0403] 149. 실시양태 147에 있어서, 상기 검출 ssDNA가 형광 공명 에너지 전달 (FRET) 쌍을 포함하는 것인 키트.
- [0404] 150. 실시양태 147-149 중 어느 하나에 있어서, 상기 가이드 RNA가 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118 또는 273에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열을 포함하는 것인 키트.
- [0405] 151. 실시양태 147-150 중 어느 하나에 있어서, 상기 가이드 RNA가 tracrRNA를 포함하는 것인 키트.
- [0406] 152. 실시양태 151에 있어서, 상기 tracrRNA가 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119 또는 274에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 것인 키트.
- [0407] 153. 실시양태 151 또는 152에 있어서, 상기 가이드 RNA가 단일 가이드 RNA인 키트.
- [0408] 154. 실시양태 151 또는 152에 있어서, 상기 가이드 RNA가 이중-가이드 RNA인 키트.
- [0409] 155. 단일-가닥 DNA를 절단하는 방법이며, 상기 방법은 표적 DNA 서열 및 복수의 비-표적 ssDNA를 포함하는 DNA 분자를 포함하는 핵산의 집단을
 - a) 서열식별번호: 1, 9, 16, 23, 30, 38, 46, 54, 61, 69, 75, 82, 89, 95, 103, 110, 117 또는 137에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 상기 표적 DNA 서열에 혼성화할 수 있는 가이드 RNA에 결합될 때 RNA-가이드된 서열 특이적 방식으로 상기 표적 DNA 서열에 결합할 수 있는 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드; 및
 - b) 상기 가이드 RNA
- [0412] 와 접촉시키는 것을 포함하고;
- [0413] 여기서, RGN 폴리펩티드가 상기 복수의 비-표적 ssDNA를 절단하는 것인 방법.
- [0414] 156. 실시양태 155에 있어서, 상기 핵산 집단이 세포 용해물 내에 있는 것인 방법.
- [0415] 157. 실시양태 155 또는 156에 있어서, 상기 가이드 RNA가 서열식별번호: 2, 10, 17, 24, 31, 39, 47, 55, 62, 70, 76, 83, 90, 96, 104, 111, 118 또는 273에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하는 CRISPR 반복 서열을 포함하는 것인 방법.
- [0416] 158. 실시양태 155-157 중 어느 하나에 있어서, 상기 가이드 RNA가 tracrRNA를 포함하는 것인 방법.

- [0417] 159. 실시양태 158에 있어서, 상기 tracrRNA가 서열식별번호: 3, 11, 18, 25, 32, 40, 48, 56, 63, 71, 77, 84, 91, 97, 105, 112, 119 또는 274에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 것인 방법.
- [0418] 160. 실시양태 158 또는 159에 있어서, 상기 가이드 RNA가 단일 가이드 RNA인 방법.
- [0419] 161. 실시양태 158 또는 159에 있어서, 상기 가이드 RNA가 이중-가이드 RNA인 방법.
- [0420] 162. CRISPR RNA (crRNA)를 코딩하는 폴리뉴클레오타이드를 포함하는 핵산 분자이며, 여기서 상기 crRNA는 스페이서 서열 및 CRISPR 반복 서열을 포함하고, 상기 CRISPR 반복 서열은 서열식별번호: 240에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하고;
- [0421] 여기서,
- [0422] a) 상기 crRNA; 및 임의로
- [0423] b) 상기 crRNA의 상기 CRISPR 반복 서열에 혼성화할 수 있는 트랜스-활성화 CRISPR RNA (tracrRNA)
- [0424] 를 포함하는 가이드 RNA는
- [0425] 상기 가이드 RNA가 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드에 결합될 때 상기 crRNA의 스페이서 서열을 통해 서열 특이적 방식으로 DNA 분자의 표적 DNA 서열에 혼성화할 수 있고,
- [0426] 여기서, 상기 crRNA를 코딩하는 폴리뉴클레오타이드는 상기 폴리뉴클레오타이드에 대해 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.
- [0427] 163. 실시양태 162의 핵산 분자를 포함하는 벡터.
- [0428] 164. 실시양태 163에 있어서, 상기 tracrRNA를 코딩하는 폴리뉴클레오타이드를 추가로 포함하는 벡터.
- [0429] 165. 실시양태 164에 있어서, 상기 tracrRNA가 서열식별번호: 241에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 것인 벡터.
- [0430] 166. 실시양태 164 또는 165에 있어서, 상기 crRNA를 코딩하는 상기 폴리뉴클레오타이드 및 상기 tracrRNA를 코딩하는 상기 폴리뉴클레오타이드가 동일한 프로모터에 작동가능하게 연결되고, 단일 가이드 RNA로서 코딩되는 것인 벡터.
- [0431] 167. 실시양태 164 또는 165에 있어서, 상기 crRNA를 코딩하는 상기 폴리뉴클레오타이드 및 상기 tracrRNA를 코딩하는 상기 폴리뉴클레오타이드가 별개의 프로모터에 작동가능하게 연결된 것인 벡터.
- [0432] 168. 실시양태 163-167 중 어느 하나에 있어서, 상기 RGN 폴리펩티드를 코딩하는 폴리뉴클레오타이드를 추가로 포함하고, 상기 RGN 폴리펩티드가 서열식별번호: 235에 대해 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 벡터.
- [0433] 169. 서열식별번호: 241에 대해 적어도 95% 서열 동일성을 갖는 뉴클레오타이드 서열을 포함하는 트랜스-활성화 CRISPR RNA (tracrRNA)를 코딩하는 폴리뉴클레오타이드를 포함하는 핵산 분자이며;
- [0434] 여기서,
- [0435] a) 상기 tracrRNA; 및
- [0436] b) 스페이서 서열 및 CRISPR 반복 서열을 포함하는 crRNA를 포함하고, 여기서 상기 tracrRNA는 상기 crRNA의 상기 CRISPR 반복 서열과 혼성화할 수 있는 것인 가이드 RNA는
- [0437] 상기 가이드 RNA가 RNA-가이드된 뉴클레아제 (RGN) 폴리펩티드에 결합될 때 상기 crRNA의 스페이서 서열을 통해 서열 특이적 방식으로 DNA 분자의 표적 DNA 서열에 혼성화할 수 있고,
- [0438] 여기서, 상기 tracrRNA를 코딩하는 폴리뉴클레오타이드는 상기 폴리뉴클레오타이드에 이중성인 프로모터에 작동가능하게 연결된 것인 핵산 분자.
- [0439] 170. 실시양태 169의 핵산 분자를 포함하는 벡터.
- [0440] 171. 실시양태 170에 있어서, 상기 crRNA를 코딩하는 폴리뉴클레오타이드를 추가로 포함하는 벡터.
- [0441] 172. 실시양태 171에 있어서, 상기 crRNA의 CRISPR 반복 서열이 서열식별번호: 240에 대해 적어도 95% 서열 동

일성을 갖는 뉴클레오티드 서열을 포함하는 것인 벡터.

[0442] 173. 실시양태 171 또는 172에 있어서, 상기 crRNA를 코딩하는 상기 폴리뉴클레오티드 및 상기 tracrRNA를 코딩하는 상기 폴리뉴클레오티드가 동일한 프로모터에 작동가능하게 연결되고, 단일 가이드 RNA로서 코딩되는 것인 벡터.

[0443] 174. 실시양태 171 또는 172에 있어서, 상기 crRNA를 코딩하는 상기 폴리뉴클레오티드 및 상기 tracrRNA를 코딩하는 상기 폴리뉴클레오티드가 별개의 프로모터에 작동가능하게 연결된 것인 벡터.

[0444] 175. 실시양태 170-174 중 어느 하나에 있어서, 상기 RGN 폴리펩티드를 코딩하는 폴리뉴클레오티드를 추가로 포함하고, 상기 RGN 폴리펩티드가 서열식별번호: 235에 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하는 것인 벡터.

[0445] 하기 실시예는 예시로서 제한 없이 제공된다.

[0446] 실시예

[0447] 실시예 1. RNA-가이드된 뉴클레아제의 확인

[0448] 17개의 별개의 CRISPR-연관 RNA-가이드된 뉴클레아제 (RGN)를 확인하였고, 하기 표 1에 기재하였다. 표 1은 각각의 RGN의 명칭, 그의 아미노산 서열, 그가 유래된 공급원, 및 가공된 crRNA 및 tracrRNA 서열을 제공한다 (확인 방법에 대해서는 실시예 2 참조). 표 1은 포괄적인 단일 가이드 RNA (sgRNA) 서열을 추가로 제공하며, 여기서 폴리-N은 sgRNA의 핵산 표적 서열을 결정하는 스페이서 서열의 위치를 나타낸다. RGN 시스템 APG05733.1, APG06207.1, APG01647.1, APG08032.1, APG02675.1, APG01405.1, APG06250.1, APG04293.1, 및 APG01308.1은 UNANNA (서열식별번호: 8)의 tracrRNA의 헤어핀 스템의 염기에서 보존된 서열을 가졌다. APG05712.1의 경우, 동일한 위치의 서열은 CNANNG (서열식별번호: 37)이다. APG01658.1의 경우, 동일한 위치의 서열은 CNANNU (서열식별번호: 45)이다. RGN 시스템 APG06498.1 및 APG06877.1의 경우, tracrRNA의 헤어핀 스템의 염기에서 보존된 서열은 UNANNG (서열식별번호: 53)이다. APG09882.1 및 APG06646.1의 경우, 동일한 위치에서의 서열은 UNANNC (서열식별번호: 68)이다. APG09053.1의 경우, 동일한 위치의 서열은 CNANNU (서열식별번호: 102)이다.

[0449] 표 1: 서열식별번호 및 CRISPR 연관된 시스템의 개요

RGN ID	서열식별 번호	공급원	crRNA 반복 서열 (서열식별번호)	tracrRNA (서열식별 번호)	sgRNA (서열식별 번호)
APG05733.1	1	바실루스 종	2	3	4
APG06207.1	9	크리세오박테리움 종	10	11	12
APG01647.1	16	스핑크고박테리움 종	17	18	19
APG08032.1	23	크리세오박테리움 종	24	25	26
APG05712.1	30	브루셀라 종	31	32	33
APG01658.1	38	스타필로코쿠스 종	39	40	41
APG06498.1	46	바실루스 종	47	48	49
APG09106.1	54	브레비바실루스 종	55	56	57
APG09882.1	61	엔테로코쿠스 종	62	63	64
APG02675.1	69	스핑크고박테리움 종	70	71	72
APG01405.1	75	크리세오박테리움 종	76	77	78
APG06250.1	82	크리세오박테리움 종	83	84	85
APG06877.1	89	바실루스 종	90	91	92
APG09053.1	95	리조비움 종	96	97	98
APG04293.1	103	미로이테스 종	104	105	106
APG01308.1	110	크리세오박테리움 종	111	112	113
APG06646.1	117	바실루스 종	118	119	120

[0450]

[0451] 실시예 2: 가이드 RNA 확인 및 sgRNA 구축

[0452] 연구 중인 RNA-가이드된 뉴클레아제 시스템을 본래 발현하는 박테리아의 배양물을 중간-로그 단계 (~0.600의 OD600)로 성장시키고, 펠렛화하고, 급속 냉동하였다. mirVANA miRNA 단리 키트 (라이프 테크놀로지스(Life Technologies), 캘리포니아주 칼스바드)를 사용하여 RNA를 펠렛으로부터 단리하고, 단리된 RNA로부터 NEBNext 소형 RNA 라이브러리 제조 키트 (NEB, 메사추세츠주 베벌리)를 사용하여 서열분석 라이브러리를 제조하였다. 라이브러리 제조물을 6% 폴리아크릴아미드 겔 상에서 분획화하여 200nt 미만의 RNA 종을 포획함으로써 각각

crRNA 및 tracrRNA를 검출하였다. 심층 서열분석 (75 bp 쌍-말단)을 Next Seq 500 (하이 아웃풋 키트(High Output kit)) 상에서 서비스 제공자 (모진(MoGene), 미주리주 세인트 루이스)에 의해 수행하였다. 판독을 컷어댑트(Cutadapt)를 사용하여 품질을 다듬고, 보우타이(Bowtie)2를 이용하여 기준 게놈에 맵핑하였다. 커스텀 RNAseq 파이프라인을 파이썬에 기록하여 crRNA 및 tracrRNA 전사체를 검출하였다. 가공된 crRNA 경계를 본래의 반복 스페이스 어레이의 서열 커버리지에 의해 결정하였다. tracrRNA의 안티-반복 부분은 허용적 BLASTn 파라미터를 사용하여 확인하였다. RNA 서열분석 깊이는 안티-반복부를 함유하는 전사체를 확인함으로써 가공된 tracrRNA의 경계를 확인시켰다. RNA의 수동 큐레이션을 RNA 폴딩 소프트웨어인 NUPACK에 의해 이차 구조 예측을 이용하여 수행하였다. sgRNA 카세트를 DNA 합성에 의해 제조하고, 일반적으로 다음과 같이 설계하였다 (5'->3'): 20-30 bp 스페이스 서열, 그의 3' 말단에서 작동가능하게 연결된 crRNA의 가공된 반복 부분, 작동가능하게 연결된 4 bp 비상보적 링커 (AAAG; 서열식별번호: 123), 그의 3' 말단에서 작동가능하게 연결된 가공된 tracrRNA. 다른 4 bp 비상보적 링커가 또한 사용될 수 있다.

[0453] 시험관내 검정의 경우, 진아트(GeneArt)TM 정밀 gRNA 합성 키트 (써모피셔(ThermoFisher))를 사용하여 sgRNA 카세트의 시험관내 전사에 의해 sgRNA를 합성하였다. 각각의 RGN 폴리펩티드에 대한 가공된 crRNA 및 tracrRNA 서열을 확인하였고, 표 1에 제시되어 있다. PAM 라이브러리 1 및 2에 대해 구축된 sgRNA에 대해서는 하기를 참고한다.

[0454] 실시예 3: 각각의 RGN에 대한 PAM 요건의 결정

[0455] 각각의 RGN에 대한 PAM 요건을 [Kleinstiver et al. (2015) Nature 523:481-485] 및 [Zetsche et al. (2015) Cell 163:759-771]로부터 본질적으로 채택된 PAM 고갈 검정을 사용하여 결정하였다. 간략히, 2개의 플라스미드 라이브러리 (L1 및 L2)를 pUC18 백본 (ampR)에서 생성하였고, 이들 각각은 8개의 무작위 뉴클레오티드 (즉, PAM 영역)에 의해 플랜킹된 별개의 30bp 프로토스페이스 (표적) 서열을 함유하였다. 각각의 RGN에 대한 라이브러리 1 및 라이브러리 2의 표적 서열 및 플랜킹 PAM 영역은 표 2에 제시되어 있다.

[0456] L1 또는 L2에서 프로토스페이스에 상응하는 스페이스 서열을 함유하는 동족 sgRNA와 함께, 본 발명의 RGN을 함유하는 pRSF-1b 발현 벡터 (이. 콜라이(*E. coli*)에 대해 코돈 최적화됨)를 보유하는 이. 콜라이 BL21(DE3) 세포에 라이브러리를 별도로 전기전공하였다. 충분한 라이브러리 플라스미드를 형질전환 반응에서 사용하여, >10⁶ CFU를 수득하였다. pRSF-1b 백본에서 RGN 및 sgRNA 둘 다 T7 프로모터의 조절하에 있었다. 형질전환 반응을 1 시간 동안 회복시킨 후, 이를 카르베니실린 및 카나마이신을 함유하는 LB 배지로 희석하고, 밤새 성장시켰다. 다음 날, 혼합물을 자체-유도 오버나이트 익스프레스(Overnight Express)TM 즉석 TB 배지 (밀리포어 시그마 (Millipore Sigma))로 희석하여 RGN 및 sgRNA를 발현시키고, 추가 4 시간 또는 20 시간 동안 성장시켰고, 그 후에 세포를 스핀 다운하고, 플라스미드 DNA를 미니-프렙 키트(Mini-prep kit) (키아젠(Qiagen), 메릴랜드주 저먼타운)에 의해 단리하였다. 적절한 sgRNA의 존재하에, RGN에 의해 인식가능한 PAM을 함유하는 플라스미드가 절단되어, 집단으로부터 제거될 것이다. RGN에 의해 인식되지 않거나 또는 적절한 sgRNA를 함유하지 않는 박테리아로 형질전환된 PAM을 함유하는 플라스미드는 생존하고 복제할 것이다. 절단되지 않은 플라스미드의 PAM 및 프로토스페이스 영역을 PCR-증폭시키고, 공개된 프로토콜 (16s-메가게놈 라이브러리 제조 가이드 15044223B, 일루미나(Illumina), 캘리포니아주 샌디에고)에 따라 서열분석을 위해 준비하였다. 심층 서열분석 (75bp 단일 말단 판독)은 서비스 제공자 (모진, 미주리주 세인트 루이스)에 의해 MiSeq (일루미나) 상에서 수행되었다. 전형적으로, 앰플리콘당 1-4 M 판독이 수득되었다. PAM 영역을 추출하고, 카운팅하고, 각각의 샘플의 총 판독에 대해 정규화하였다. 플라스미드 절단을 유도하는 PAM은 대조군과 비교할 때 (즉, 라이브러리가 RGN을 함유하지만 적절한 sgRNA가 결여된 이. 콜라이로 형질전환되었을 때) 적게 표시되는 것으로 확인되었다. 신규한 RGN에 대한 PAM 요건을 표시하기 위해, 해당 영역의 모든 서열에 대한 고갈률 (샘플에서 빈도/대조군에서 빈도)을 -log 염기 2 형질전환에 의해 농축 값으로 전환시켰다. 충분한 PAM은 농축 값 >2.3 (고갈률 < ~0.2에 상응함)을 가진 것들로 정의되었다. 두 라이브러리 모두에서 이 역치보다 높은 PAM을 수집하였고, 웹 로그를 생성하기 위해 사용하였으며, 이는 예를 들어 "웹로그(weblogo)"로 공지된 인터넷 상의 웹-기반 서비스를 이용하여 생성될 수 있다. 상위 농축 PAM에서 일정한 패턴이 있을 때 PAM 서열을 확인하고 보고하였다. 각각의 RGN에 대한 컨센서스 PAM (농축 지수 (EF) >2.3을 가짐)이 표 2에 제공된다. PAM 배향을 또한 표 2에 나타낸다. 본 출원의 다른 곳에서 언급된 바와 같이, APG06646.1 및 APG04293.1 뉴클레아제는 PAM-상호작용 도메인을 보유하지 않는다. 표 2의 결과는 이들이 또한 2-5개 뉴클레오티드인 전형적인 PAM 요건을 보유하지 않는다는 것을 나타낸다. APG06646.1 및 APG04293.1은 절단을 위해 단일 뉴클레오티드 요건을 갖는 것으로 나타났다.

[0457] 표 2: PAM 또는 PAM-유사 결정

RGD ID	sgRNA L1 (서열식별번호)	sgRNA L2 (서열식별번호)	PAM (서열식별번호)	PAM 배향
APG05733.1	5	6	7	5'-표적-PAM-3'
APG06207.1	13	14	15	5'-표적-PAM-3'
APG01647.1	20	21	22	5'-표적-PAM-3'
APG08032.1	27	28	29	5'-표적-PAM-3'
APG05712.1	34	35	36	5'-표적-PAM-3'
APG01658.1	42	43	44	5'-표적-PAM-3'
APG06498.1	50	51	52	5'-표적-PAM-3'
APG09106.1	58	59	60	5'-PAM-표적-3'
APG09882.1	65	66	67	5'-표적-PAM-3'
APG02675.1	73	74	15	5'-표적-PAM-3'
APG01405.1	79	80	81	5'-표적-PAM-3'
APG06250.1	86	87	88	5'-표적-PAM-3'
APG06877.1	93	94	7	5'-표적-PAM-3'
APG09053.1	99	100	101	5'-표적-PAM-3'
APG04293.1	107	108	109	5'-표적-PAM-3'
APG01308.1	114	115	116	5'-표적-PAM-3'
APG06646.1	121	122	109	5'-표적-PAM-3'

[0458]

[0459] 실시예 4: 뉴클레아제 활성을 증가시키기 위한 가이드 RNA의 조작

[0460] 4.1 RGN APG09748 및 APG09106.1

[0461] 매우 높은 서열 동일성을 갖고, 동일한 PAM을 갖는 RGN APG09748 (서열식별번호: 137로 제시되고, APG09748 crRNA 반복 서열, tracrRNA 서열, 및 일반 sgRNA 서열은 각각 서열식별번호: 273, 274, 및 275로 제시되며; 모든 서열은 국제 출원 번호 PCT/US2019/068079 (그 전문이 참조로 포함됨)에 기재되어 있음) 및 APG09106.1에 대해, RNA 폴딩 예측을 이용하여 뉴클레아제 활성을 최적화하도록 변경될 수 있는 가이드 RNA 내의 영역을 결정하였다. 반복:안티-반복 영역에서 crRNA:tracrRNA 염기쌍 형성의 안정성은 반복:안티-반복 영역을 단축시키고, G-C 염기쌍을 부가하고, G-U 동요(wobble) 쌍을 제거함으로써 증가되었다. "최적화된" 가이드 변이체를 시험관 내 절단 검정에서 RGN APG09748을 이용하여 시험하고 야생형 gRNA와 비교하였다.

[0462] RNP 형성을 위한 RGN을 생성하기 위해, C-말단 His6 (서열식별번호: 276) 또는 His10 (서열식별번호: 277) 태그에 융합된 RGN을 함유하는 발현 플라스미드를 구축하고, 이. 콜라이의 BL21 (DE3) 균주로 형질전환시켰다. 발현은 50 µg/mL 카나마이신으로 보충된 매직 미디어(Magic Media) (써모 피셔)를 사용하여 수행하였다. 용해 및 정화 이후, 단백질을 고정된 금속 친화도 크로마토그래피에 의해 정제하고, 큐비트(Qubit) 단백질 정량화 키트 (써모 피셔)를 사용하여 또는 계산된 소광 계수를 이용하는 UV-vis에 의해 정량화하였다.

[0463] 리보핵단백질 (RNP)은 정제된 RGN을 sgRNA와 함께 ~2:1 비로 실온에서 20 분 동안 인큐베이션함으로써 제조하였다. 시험관내 절단 반응을 위해, RNP를 바람직한 PAM 서열에 의해 플랭킹된 표적화된 프로토스페이서를 함유하는 플라스미드 또는 선형 dsDNA와 함께 실온에서 >30 분 동안 인큐베이션하였다. TRAC 유전자좌 내의 2가지 표적 핵산 서열, TRAC11 (서열식별번호: 278) 및 TRAC14 (서열식별번호: 279)를 시험하였다. gRNA를 정확한 표적 핵산 서열을 갖는 (예를 들어, gRNA는 TRAC11 스페이서 서열을 갖고, 검정된 표적은 TRAC11임) 및 정확한 표적 핵산 서열을 갖지 않는 (예를 들어, gRNA는 TRAC11 스페이서 서열을 갖고, 검정된 표적이 TRAC14임) 표적화된 활성 둘 다에 대해 검정하였다. 플라스미드 절단에 의해 결정된 활성을 아가로스 겔 전기영동에 의해 평가하였다. 결과는 표 3에 나타내었다. 가이드 변이체는 서열식별번호: 280-283으로서 나열되고, 스페이서 서열과 함께 제공된다. 이들 가이드 서열은 AAAA의 비상보성 뉴클레오티드 링커 (서열식별번호: 284)를 사용한다. 반복:안티-반복 결합이 증가된 최적화된 gRNA (서열식별번호: 285; 폴리-N은 스페이서 서열의 위치를 나타냄)는 최적화된 tracrRNA (서열식별번호: 286) 및 최적화된 crRNA (서열식별번호: 287) 성분을 갖는다. 최적화된 가이드 변이체는 이전에 야생형 가이드 RNA를 사용하여 절단이 검출되지 않은 2가지 유전자좌를 절단할 수 있다. 반복:안티-반복 영역에서 혼성화의 최적화를 통해, TRAC 유전자좌에서 다중 표적에 대해 APG09748의 시험관내 절단이 0% 절단에서 100% 절단으로 증가하였다.

표 3: 조작된 가이드 변이체에 의한 APG09748의 편집 효율

gRNA 변이체 (서열식별번호)	가이드 설계	검정된 표적	젤 1 - 2 μ L 로드		젤 2 - 1 μ L 로드	
			% 무손상	% 절단	% 무손상	% 절단
280	최적화	TRAC11	68	32	57	43
280	최적화	TRAC14	100	0	100	0
281	최적화	TRAC11	100	0	100	0
281	최적화	TRAC14	70	30	69	31
282	WT	TRAC11	100	0	100	0
282	WT	TRAC14	100	0	100	0
283	WT	TRAC11	100	0	100	0
283	WT	TRAC14	100	0	100	0
없음		TRAC11	100	0	100	0
없음		TRAC14	100	0	100	0

추가로, 최적화된 gRNA 변이체를 설계하고 검정하였다. 추가로, 스페이서 길이가 절단 효율에 어떻게 영향을 미치는지 결정하기 위해 상이한 길이의 스페이서 서열 또한 시험하였다. 스페이서 서열 외부의 sgRNA는 이 검정에서 "백본"으로 지칭된다. 표 4에서, 이들은 "WT" (서열식별번호: 288, 야생형 서열), 및 3가지 최적화된 sgRNA: V1 (서열식별번호: 289), V2 (서열식별번호: 290) 및 V3 (서열식별번호: 291)으로 표시된다. 이들 서열 모두 스페이서 서열의 위치를 나타내기 위해 폴리-N을 갖는다. 가이드는 시험관내 전사 (IVT)에 의해 sgRNA로 발현되었다. 야생형 sgRNA 백본과 비교하여, V1은 87.8% 동일하고, V2는 92.4% 동일하고, V3은 85.5% 동일하다. 이중-가이드 RNA를 나타내지만 그 외에는 상기 인용된 야생형 및 최적화된 sgRNA와 유사한 합성 tracrRNA:crRNA 듀플렉스 ("합성") 또한 생성하여 시험하였다.

이 검정 세트의 경우, RGN APG09106.1이 사용되었고; 그 외에는 시험관내 절단 반응을 위한 방법이 상기 기재된 것과 유사하였다. 표적화된 핵산 서열은 표적 1 (서열식별번호: 292) 및 표적 2 (서열식별번호: 293)이었다. 결과는 표 4에 나타내었다.

표 4: 조작된 가이드 변이체에 의한 APG09106.1의 편집 효율

RNA 공급원	표적	스페이서 길이	백본	스페이서 서열식별번호	절단 %
합성	2	18	WT	294	12.3
합성	1	20	WT	295	0
합성	2	20	WT	296	55.0
합성	1	25	WT	297	0
합성	2	25	WT	298	61.4
IVT	2	25	V1	299	1.1
IVT	2	25	V2	300	0.9
IVT	2	25	V3	301	0.7
IVT	2	20	V3	302	21.0
IVT	1	25	V3	303	2.0

4.2 RGN APG07433.1

RNA 폴딩 예측을 사용하여 APG07433.1 (서열식별번호: 235로 제시되고, 미국 출원 공개 번호 2019/0367949 및 국제 출원 공개 번호 WO 2019/236566 (이들 각각은 그 전문이 본원에 참조로 포함됨)에 기재됨)에 대해 뉴클레아제 활성을 최적화하고 바이러스 벡터 내로의 패키징을 위한 가이드 RNA 단축을 위해 변경될 수 있는 가이드 RNA 내의 영역을 결정하였다. 2개의 영역을 변경을 만들기 위한 잠재적 위치로서 확인하였다: crRNA 및 tracrRNA 영역 사이의 반복:안티-반복 쌍형성 및 tracrRNA 내의 말단 헤어핀. 반복:안티-반복 영역을 7개 (APG07433.1-7bp, 서열식별번호: 238 및 239), 13개 (APG07433.1-13bp, 서열식별번호: 250 및 251), 및 15개 염기 쌍 길이 (APG07433.1-15bp, 서열식별번호: 242 및 243)로 말단절단하였다. 추가로, 반복:안티-반복 영역의 서열을 변경하고 쌍형성을 11개 염기쌍 길이로 감소시킨 제4 변이체 (APG07433.1-11bp-syn, 서열식별번호: 240 및 241)를 야생형 가이드에 비해 보다 작은 RNA 벌지의 도입과 함께 시험하였다. 또한, 천연 서열을 40개 (APG07433.1-40ntTHP, 서열식별번호: 254 및 255) 및 42개 (APG07433.1-42ntTHP, 서열식별번호: 244 및 245) 뉴클레오타이드 길이로 감소시키는 것을 포함하여, tracrRNA 내의 말단 헤어핀에 대해 말단절단을 수행하였다. 변경된 스템-루프를 35개 (APG07433.1-35ntTHP-syn, 서열식별번호: 248 및 249) 및 39개 (APG07433.1-39ntTHP-

syn, 서열식별번호: 246 및 247) 뉴클레오티드로 단축되도록 설계하였다. 하나의 가이드는 13-염기쌍 단축된 반복:안티-반복 영역 및 42-뉴클레오티드 단축된 말단 헤어핀 (APG07433.1-13bp42ntTH, 서열식별번호: 252 및 253)을 조합하였다. 절단에 대한 스페이서 길이의 효과를 또한 시험하였다. 야생형 가이드 RNA 백본 상의 25개 (APG07433.1-천연 [25], 서열식별번호: 236 및 237; 스페이서 서열은 서열식별번호: 271로 제시됨) 및 18개 (APG07433.1-천연 [18], 서열식별번호: 256 및 257; 스페이서 서열은 서열식별번호: 272로 제시됨) 뉴클레오티드의 스페이서 길이를 또한 시험하였다.

[0472] 하기 이중 가이드 RNA를, 어닐링 완충제 (신테고(Synthego)) 중 20 μ M crRNA 및 10 μ M tracrRNA를 함유하는 용액을 제조하여 crRNA 및 tracrRNA를 어닐링함으로써 제조한 다음, 78°C로 10 분 동안 가열하고, 이어서 37°C에서 30 분 동안 가열하였다. 포스페이트 완충 염수 중 0.5 μ M의 정제된 APG07433.1 단백질 및 1 μ M의 이중 가이드 RNA와 함께 20 분 동안 인큐베이션함으로써 RNP를 형성하였다. 본 실험에 사용된 가이드 RNA를 표 5에 나타낸다.

[0473] 표 5. 이중 가이드 RNA

이중 gRNA 명칭	crRNA 서열식별번호	tracrRNA 서열식별번호
APG07433.1-천연[25]	236	237
APG07433.1-7bp	238	239
APG07433.1-11bp-syn	240	241
APG07433.1-15bp	242	243
APG07433.1-42ntTHP	244	245
APG07433.1-39ntTHP-syn	246	247
APG07433.1-35ntTHP-syn	248	249
APG07433.1-13bp	250	251
APG07433.1-13bp42ntTH	252	253
APG07433.1-40ntTHP	254	255
APG07433.1-천연[18]	256	257
APG07433.1cr-13bp tr-35ntTHP	258	259

[0474]

[0475] 이들 RNP를, 절단 생성물의 정확하고 민감한 정량화를 용이하게 하기 위해 FAM 및 Cy3 표지된 프라이머를 사용한 증폭에 의해 생성된 적절한 표적 서열 (서열식별번호: 260)을 함유하는 PCR 생성물과 함께 인큐베이션하였다. 반응물은 1X 컷스마트 완충제 (뉴 잉글랜드 바이오랩스) 중 상기와 같이 생성된 250 nM RNP, 및 150 nM의 FAM 및 Cy3 표지된 PCR 생성물을 함유하였다. 반응을 37°C에서 15 분 동안 진행되도록 하고, RNase A를 0.1 mg/mL까지 및 EDTA를 45 mM까지 첨가함으로써 종결시켰다. 켄칭된 반응물을 50°C에서 30 분 동안 및 95°C에서 5 분 동안 가열하였다. 이어서, 샘플을 5% TBE 겔 (바이오-라드) 상에서 천연 아크릴아미드 겔 전기영동을 사용하여 분석하였다. 이들을 케미독(ChemiDoc) MP 이미지 상에서 영상화하였다. 2종의 염료 각각을 정량화에 사용할 수 있고, 각각의 샘플을 이중으로 수행하였다. 상기 실험의 결과를 표 6에 나타냈다.

[0476] 표 6. 절단 결과

RNP	% 절단				전체 평균
	Cy3 채널		FAM 채널		
	Rep 1	Rep 2	Rep 1	Rep 2	
뉴클레아제 단독 (가이드 없음)	0	0	0	0	0
APG07433.1-천연[25]	66.41	61.16	68.81	56.85	63.3
APG07433.1-7bp	0	3.38	2.83	0	1.6
APG07433.1-11bp-syn	66.33	76.38	66.73	68	69.4
APG07433.1-15bp	50.7	65.61	54.62	59.95	57.7
APG07433.1-42ntTHP	44.81	50.59	48.84	51.77	49.0
APG07433.1-39ntTHP-syn	19.16	32.88	24.91	33.43	27.6
APG07433.1-35ntTHP-syn	26.77	38.61	32.09	40.2	34.4
APG07433.1-13bp	44.36	33.03	42.93	35.22	38.9
APG07433.1-13bp42ntTHP	35.91	36.74	33.84	39.2	36.4
APG07433.1-40ntTHP	44.52	57.14	48.34	57.84	52.0
APG07433.1-천연[18]	11.1	9.76	12.93	13.24	11.8
APG07433.1cr-13bp tr-35ntTHP	10.66	12.25	12.65	13.54	12.3

[0477]

[0478] 상기 분석은 천연 백본이 대부분의 단축된 변이체, 및 단축된 표적 서열을 함유하는 변이체 (APG07433.1-천연 [18], 서열식별번호: 256 및 257)를 능가한다는 것을 입증한다. 단축된 백본 서열 중에서 가장 높은 수준의 절

단을 갖는 것은, 25 nt 표적화 서열 (스페이서; 서열식별번호: 271), crRNA에 대한 서열식별번호: 240 및 tracrRNA로서의 서열식별번호: 241을 포함하는 APG07433.1-11bp-syn으로 명명된 서열이다. 이 가이드 변이체는 스템에 조작된 별지를 갖는 변경된 반복:안티-반복 스템 루프를 포함하였다.

[0479] 실시예 5: 포유동물 세포에서 유전자 편집 활성의 입증

[0480] RGN 발현 카세트를 생성하고, 포유동물 발현을 위해 벡터 내로 도입하였다. RGN APG09748, APG09106.1, APG05712.1, APG01658.1, APG05733.1, APG06498.1, APG06646.1, APG09882.1, APG01405.1, 및 APG01308.1을 각각 포유동물 발현을 위해 코돈-최적화하고 (각각 서열식별번호: 304, 305, 및 411-418), 발현된 단백질을 N-말단에서 SV40 핵 국제화 서열 (NLS; 서열식별번호: 125) 및 3xFLAG 태그 (서열식별번호: 126)에 작동가능하게 융합시키고, C-말단에서 뉴클레오플라스민 NLS 서열 (서열식별번호: 127)에 작동가능하게 융합시켰다. 직렬식으로 작동가능하게 융합된 NLS 서열의 2개의 카피를 사용하였다. 각각의 발현 카세트는 시토메갈로바이러스 (CMV) 프로모터 (서열식별번호: 306)의 조절하에 있었다. CMV 전사 인핸서 (서열식별번호: 307) 또한 CMV 프로모터를 포함하는 구축물에 포함될 수 있음이 관련 기술분야에 공지되어 있다. 인간 RNA 폴리머라제 III U6 프로모터 (서열식별번호: 308)의 조절하에 단일 gRNA 각각을 코딩하는 가이드 RNA 발현 구축물을 생성하여 발현 벡터 내로 도입하였다. 가이드는 표 7에 나타난 바와 같이, RelA, AurkB, GAPDH, LINC01509, HBB, CFTR, HPRT1, TRA, EMX1 및 VEGFA를 포함한 선택된 유전자의 영역을 표적화하였다. RNA-가이드된 뉴클레아제 APG09106.1의 경우, 특정 잔기를 돌연변이시켜 단백질의 뉴클레아제 활성을 증가시켰고, 구체적으로 APG09106의 T849 잔기를 아르기닌으로 돌연변이시켰다 (서열식별번호: 309). 이러한 점 돌연변이는 포유동물 세포에서 편집률을 증가시켰다.

[0481] 상기 기재된 구축물을 포유동물 세포에 도입하였다. 형질감염 하루 전날에, 1×10^5 HEK293T 세포 (시그마)을 돌베코 변형된 이글 배지 (DMEM) 플러스 10% (vol/vol) 태아 소 혈청 (김코(Gibco)) 및 1% 페니실린-스트렙토마이신 (김코) 중에서 24-웰 디쉬에 플레이팅하였다. 세포가 50-60% 전면성장한 다음 날, 500 ng의 RGN 발현 플라스미드 플러스 500 ng의 단일 gRNA 발현 플라스미드를 제조업체의 지침에 따라 웰당 1.5 μ L의 리포펙타민 3000 (써모 사이언티픽)을 사용하여 공동 형질감염시켰다. 48 시간 성장시킨 후, 총 게놈 DNA를 제조업체의 지침에 따라 게놈 DNA 단리 키트 (마세리-나겔(Machery-Nagel))를 사용하여 수확하였다.

[0482] 이어서, 총 게놈 DNA를 분석하여, 표적화된 유전자에서 편집률을 결정하였다. 올리고뉴클레오티드를 생성하여, PCR 증폭, 및 증폭된 게놈 표적 부위 (서열식별번호: 310 및 311)의 후속적인 분석을 위해 사용하였다. 모든 PCR 반응은 0.5 μ M의 각각의 프라이머를 포함하는 20 μ L 반응 중에서 10 μ L의 2X 마스터 믹스 퓨전 하이-피델리티(Master Mix Phusion High-Fidelity) DNA 폴리머라제 (써모 사이언티픽)를 사용하여 수행하였다. 먼저, 각각의 표적 유전자를 포함하는 큰 게놈 영역을 PCR#1 프라이머 (서열식별번호: 310 및 311)를 사용하고 하기 프로그램을 이용하여 증폭시켰다: 98 $^{\circ}$ C, 1 분; [98 $^{\circ}$ C, 10 초; 62 $^{\circ}$ C, 15 초; 72 $^{\circ}$ C, 5 분]의 30 주기; 72 $^{\circ}$ C, 5 분; 12 $^{\circ}$ C, 계속.

[0483] 이어서, 1 마이크로리터의 이 PCR 반응을 각각의 가이드에 대해 특이적인 프라이머 (PCR#2 프라이머; 서열식별번호: 365-370)를 사용하고 하기 프로그램을 이용하여 추가로 증폭시켰다: 98 $^{\circ}$ C, 1 분; [98 $^{\circ}$ C, 10 초; 67 $^{\circ}$ C, 15 초; 72 $^{\circ}$ C, 30 초]의 35 주기; 72 $^{\circ}$ C, 5 분; 12 $^{\circ}$ C, 계속. PCR#2에 대한 프라이머에는 일루미나 서열분석을 위한 넥스테라(Nextera) 판독 1 및 판독 2 트랜스포사제 어댑터 오버행 서열이 포함된다.

[0484] 제2 PCR 증폭 이후, DNA를 제조업체의 지침에 따라 PCR 클린업 키트 (자이모(Zymo))를 사용하여 세척하고, 물에 용리시켰다. 200-500ng의 정제된 PCR#2 생성물을 20 μ L 반응 중에서 2 μ L의 10X NEB 완충제 2 및 물과 합하고, 하기 프로그램을 이용하여 어닐링하여 헤테로듀플렉스 DNA를 생성하였다: 95 $^{\circ}$ C, 5 분; 95-85 $^{\circ}$ C, 2 $^{\circ}$ C/초의 속도로 냉각; 85-25 $^{\circ}$ C, 0.1 $^{\circ}$ C/초의 속도로 냉각; 12 $^{\circ}$ C, 계속. 어닐링 후에, 5 μ L의 DNA를 비 효소 대조군으로서 제거하였고, 1 μ L의 T7 엔도뉴클레아제 I (NEB)을 첨가하고, 반응을 37 $^{\circ}$ C에서 1 시간 동안 인큐베이션하였다. 인큐베이션 이후, 5x 플래쉬겔(FlashGel) 로딩 염료 (론자(Lonza))를 첨가하고, 겔 전기영동을 이용하여 2.2% 아가로스 플래쉬겔 (론자)에 의해 5 μ L의 각각의 반응 및 대조군을 분석하였다. 겔의 가시화 이후에, 비상동성 말단 결합 (NHEJ)의 백분율은 하기 방정식을 이용하여 결정하였고: %NHEJ 사건 = $100 \times [1 - (1 - \text{절단된 분획})^{1/2}]$, 여기서 (절단된 분획)은 다음과 같이 정의된다: (소화된 생성물의 밀도)/(소화된 생성물의 밀도 + 소화되지 않은 모 밴드).

[0485] 일부 샘플의 경우, 서베이어®를 이용하여 포유동물 세포에서의 발현 이후의 결과를 분석하였다. 형질감염 이후 세포를 37 $^{\circ}$ C에서 72 시간 동안 인큐베이션한 다음, 게놈 DNA를 추출하였다. 게놈 DNA는 제조업체의 프로토콜에 따라 퀵익스트랙트 DNA 추출 용액 (에피센터)을 사용하여 추출하였다. RGN 표적 부위를 플랭킹하는 게놈

영역을 PCR 증폭시키고, 생성물을 제조업체의 프로토콜에 따라 쿼릭 스핀 컬럼 (쿼아젠)을 사용하여 정제하였다. 200-500 ng의 총 정제된 PCR 생성물을 10 μ l의 최종 부피로 1 μ l의 10x Taq DNA 폴리머라제 PCR 완충제 (엔자이마틱스) 및 초순수와 혼합하고, 재어닐링 과정에 적용하여 헤테로듀플렉스를 형성하였다: 10 분 동안 95 $^{\circ}$ C, -2 $^{\circ}$ C/s 속도로 95 $^{\circ}$ C에서 85 $^{\circ}$ C로, -0.25 $^{\circ}$ C/s 속도로 85 $^{\circ}$ C에서 25 $^{\circ}$ C로, 및 1 분 동안 25 $^{\circ}$ C 유지.

[0486] 재어닐링 이후, 생성물을 제조업체의 권고된 프로토콜에 따라 서베이어® 뉴클레아제 및 서베이어® 인헨서 S (인테그레이티드 디엔에이 테크놀로지스)로 처리하고, 4-20% 노벡스(Novex) TBE 폴리아크릴아미드 겔 (라이프 테크놀로지스) 상에서 분석하였다. 겔을 SYBR 골드 DNA 염색 (라이프 테크놀로지스)으로 10 분 동안 염색하고, 겔 독(Gel Doc) 겔 영상화 시스템 (바이오-라드)으로 영상화하였다. 정량화는 상대적인 밴드 강도를 기반으로 하였다. Indel 백분율은 하기 식에 의해 결정하고: $100 \times (1 - (1 - (b+c)/(a+b+c))^{\frac{1}{2}})$, 여기서 a는 소화되지 않은 PCR 생성물의 통합 강도이고, b 및 c는 각각의 절단 생성물의 통합 강도이다.

[0487] 추가로, 일루미나 오버행 서열을 함유하는 PCR#2로부터의 생성물은 일루미나 16S 메가게놈 시퀀싱 라이브러리 (Metagenomic Sequencing Library) 프로토콜에 따라 라이브러리 제조를 거쳤다. 심층 서열분석은 서비스 제공자 (모진)에 의해 일루미나 Mi-Seq 플랫폼 상에서 수행되었다. 전형적으로 앰플리콘당 200,000개의 250 bp 쌍-말단 판독 (2 x 100,000 판독)이 생성된다. 판독을 CRISPResso ([Pinello, et al. 2016 Nature Biotech, 34:695-697])를 이용하여 분석하여, 편집률을 계산하였다. 출력 정렬을 수동-큐레이션하여, 삽입 및 결실 부위를 확인할 뿐만 아니라, 재조합 부위에서 미세상동성 부위를 확인하였다. 각각의 샘플에서의 전체 편집률, 뿐만 아니라 결실률 및 삽입률을 표 7에 나타낸다. 모든 실험은 인간 세포에서 수행하였다. "표적 서열"은 유전자 표적 내의 표적화된 서열이다. 각각의 표적 서열의 경우, 가이드 RNA는 사용된 RGN에 따라 상보성 RNA 스페이서 서열 및 적절한 sgRNA를 포함하였다. 가이드 RNA에 의한 선택된 실험 분석은 표 8.1 및 8.2에 나타내었다.

[0488] 표 7: 유전자 표적에 대한 전체 편집률

RGN	유전자 표적	가이드 RNA (서열식별번호)	표적 (서열식별번호)	전체 편집률	결실률	삽입률
APG09106.1	AurkB	316	314	0.55%	100%	
APG09106.1	AurkB	312	315	0.60%	54%	46%
APG09106.1 T849R	AurkB	316	314	2.97%	98%	2.00%
APG09106.1 T849R	AurkB	312	315	2.36%	100%	
APG05712.1	RelA	419	482	0.36%	72.33%	27.67%
APG05712.1	AurkB	420	483	1.39%	25.94%	75.12%
APG05712.1	RelA	421	484	1.04%	32.10%	70.33%
APG05712.1	RelA	422	485	0.28%	64.13%	35.87%
APG01658.1	RelA	423	486	10.03%	70.66%	29.33%
APG01658.1	GAPDH	424	487	5.60%	76.96%	25.97%
APG01658.1	LINC01509	425	488	9.87%	75.10%	26.61%
APG01658.1	HBB	426	489	0.42%	63.86%	36.13%
APG01658.1	CFTR	427	490	6.06%	75.42%	27.86%
APG01658.1	AurkB	428	491	13.07%	88.02%	13.14%
APG05733.1	HPRT1	429	492	0.36%		100%
APG05733.1	AurkB	430	493	5.01%	86.40%	13.60%
APG05733.1	RelA	431	494	0.76%		100%
APG05733.1	HPRT1	432	495	0.17%	38.42%	61.58%
APG05733.1	RelA	433	496	0.96%	77.52%	22.48%
APG06498.1	TRA	434	497	35.35%	92.25%	8.29%
APG06498.1	TRA	435	498	21.04%	84.87%	16.34%
APG06498.1	TRA	436	499	10.16%	89.44%	10.69%
APG06498.1	TRA	437	500	13.21%	87.24%	15.00%
APG06498.1	TRA	438	501	13.56%	86.73%	13.66%
APG06498.1	TRA	439	502	9.62%	95.85%	19.90%
APG06498.1	TRA	440	503	1.51%	75.89%	24%
APG06498.1	TRA	441	504	2.63%	88.81%	11.76%
APG06498.1	TRA	442	505	4.27%	52.09%	47.90%
APG06646.1	TRA	443	497	0.33%	0%	100%
APG06646.1	VEGFA	444	506	4.52%	40.83%	63.79%
APG06646.1	AurkB	445	507	0.66%	69.38%	30.62%
APG06646.1	AurkB	446	508	6.55%	79.97%	20.37%
APG06646.1	TRA	447	509	2.47%	73.40%	40.81%
APG06646.1	TRA	448	510	0.14%	34.46%	65.55%
APG06646.1	TRA	449	511	0.61%	0%	100%
APG06646.1	AurkB	450	493	7.94%	67.95%	32.40%
APG06646.1	TRA	451	512	7.03%	86.95%	14.03%
APG06646.1	TRA	452	513	0.50%	21.80%	78.10%

[0489]

RGN	유전자 표적	가이드 RNA (서열식별번호)	표적 (서열식별번호)	전체 편집률	결실률	삽입률
APG06646.1	TRA	453	514	3.53%	86.42%	16.28%
APG06646.1	RelA	454	515	6.22%	80.39%	45.34%
APG06646.1	TRA	455	516	0.38%	0%	100%
APG06646.1	TRA	456	517	0.34%	66.76%	48.22%
APG06646.1	RelA	457	518	3.78%	75.25%	67.90%
APG06646.1	TRA	458	519	0.42%	69.61%	30.39%
APG06646.1	TRA	459	520	0.08%	100.00%	0.00%
APG06646.1	AurkB	460	521	1.09%	30.18%	69.81%
APG06646.1	AurkB	461	522	0.28%	86.60%	21.78%
APG06646.1	AurkB	462	523	10.92%	78.87%	22.05%
APG06646.1	TRA	463	504	0.08%	0%	100%
APG06646.1	TRA	464	505	2.84%	38.26%	61.47%
APG06646.1	TRA	465	524	0.43%	71.67%	28.33%
APG06646.1	TRA	466	525	0.03%	0.00%	100.00%
APG06646.1	TRA	467	526	0.08%	100%	0%
APG09882.1	RelA	468	482	0.32%	100.00%	0.00%
APG09882.1	EMX1	469	527	12.84%	82.46%	17.79%
APG09882.1	VEGFA	470	528	14.76%	92.56%	7.77%
APG09882.1	TRA	471	529	14.66%	94.19%	7.33%
APG09882.1	TRA	472	530	7.84%	95.07%	5.75%
APG09882.1	VEGFA	473	531	24.45%	88.96%	11.57%
APG09882.1	TRA	474	532	14.43%	90.24%	10.96%
APG01405.1	RelA	475	482	0.06%	0%	100%
APG01405.1	AurkB	476	533	6.81%	80.58%	20.27%
APG01405.1	AurkB	477	534	0.54%	33.83%	66.16%
APG01405.1	HPRT1	478	535	1.07%	42.14%	57.87%
APG01405.1	RelA	479	484	0.55%	2.07%	97.93%
APG01308.1	HPRT1	480	536	1.19%	80.57%	19.43%
APG01308.1	AurkB	481	537	0.90%	94.97%	6.46%

[0490]

[0491]

각각의 가이드에 대한 특이적인 삽입 및 결실을 표 8.1 및 8.2에 나타내었다. 이들 표에서, 표적 서열은 굵은 대문자로 표시되었다. 8합체 PAM 영역은 이중 밑줄로 표시되었고, 인식된 주요 뉴클레오티드는 굵게 표시되었다. 삽입은 소문자로 표시되었다. 결실은 점선 (---)으로 나타내었다. INDEL 위치는 표적 서열의 PAM 근위 엡지로부터 계산되고, 엡지는 위치 0이다. 위치가 엡지의 표적 쪽에 있는 경우, 위치는 양수 (+)이고; 위치가 엡지의 PAM 쪽에 있는 경우, 위치는 음수 (-)이다.

[0492]

표 8.1: RGN APG09106.1을 사용한 가이드 831에 대한 특이적인 삽입 및 결실

편집된 표적 서열	# 판독	% 판독	% INDEL	유형	INDEL 위치	크기
GTCTGATTGCCTGTCGTTGCCCTCCAGATCAT GGAGGAGTTGGCAGA (야생형; 서열식별번호: 316)	92294	99.40				
GTCTGATTGCCTGTCGTTGCCCTCCCA ----- --AGGAGTTGGCAGA (서열식별번호: 317)	263	0.28	54.22	결실	+19	8
GTCTGATTGCCTGTCGTTGCC Ctaagtgtatta agcattgtctcagagattttGGAGGAGTTGGCAG A (서열식별번호: 318)	222	0.24	45.77	삽입	+13	20

[0493]

상화 시스템 (바이오-라드)에 의해 영상화한다. 정량화는 상대적인 밴드 강도를 기반으로 한다. Indel 백분율은 하기 식에 의해 결정하고: $100 \times (1 - (1 - (b+c)/(a+b+c))^{\frac{1}{2}})$, 여기서 a는 소화되지 않은 PCR 생성물의 통합 강도이고, b 및 c는 각각의 절단 생성물의 통합 강도이다.

[0499]

실시예 7: 질환 표적의 확인

[0500]

임상 변이체의 데이터베이스를 NCBI ClinVar 웹사이트에서 월드 와이드 웹을 통해 입수가 가능한 NCBI ClinVar 데이터베이스로부터 획득하였다. 병원성 단일 뉴클레오타이드 다형성 (SNP)을 이 목록으로부터 확인하였다. 게놈 유전자와 정보를 이용하여, 각각의 SNP와 중복되고 그를 둘러싸는 영역에서 CRISPR 표적을 확인하였다. 원인 돌연변이 ("Cas1 Mut.")를 표적화기 위해 APG09748 또는 APG09106.1을 포함하는 RGN 시스템과 조합하여 염기 편집을 이용하여 교정될 수 있는 SNP의 선택이 표 9.1에 열거된다. APG06646.1 또는 APG01658.1을 포함하는 RGN 시스템과 조합하여 염기 편집을 사용해 원인 돌연변이 ("Cas1 Mut.")를 표적화도록 수정될 수 있는 SNP의 선택이 표 9.2에 열거된다. 표 9.1 및 9.2 둘 다에서, 각각의 질환의 하나의 가명만이 열거된다. "RS#"은 NCBI 웹사이트에서 SNP 데이터베이스를 통한 RS 수탁 번호에 상응한다. 염색체 수탁 번호는 NCBI 웹사이트를 통해 발견된 수탁 참조 정보를 제공한다. 표 9.1 및 9.2는 또한 각각의 질환에 대해 RGN 시스템 APG09748 또는 APG09106.1 (표 9.1) 또는 APG06646.1 또는 APG01658.1 (표 9.2)에 적합한 게놈 표적 서열 정보를 제공한다. 표적 서열 정보는 또한 본 발명의 상응하는 RGN에 필요한 sgRNA의 생성을 위한 프로토스페이서 서열을 제공한다.

표 9.1: APG09748 및 APG09106.1에 대한 질환 표적

질환	RS#	CasI Mut.	염색체 수탁	유전자 기호	표적 (서열식별번호)
스타르가르트병 1	1800728	A>G	NC_000001.10,NC_000001.11	ABCA4	313
글리코젠 축적 질환 유형 1A	1801175	C>T	NC_000017.10,NC_000017.11	G6PC	335
중증 복합 면역결핍 질환	3218716	C>T	NC_000014.8,NC_000014.9	MYH7	336
페닐케톤뇨증	5030858	G>A	NC_000012.11,NC_000012.12	PAH	337
고페닐알라닌혈증	5030860	T>C	NC_000012.11,NC_000012.12	PAH	338
알파-1-항트립신 결핍증	28929474	C>T	NC_000014.8,NC_000014.9	SERPINA1	339
MECP2-관련 장애	28934906	G>A	NC_000023.10,NC_000023.11	MECP2	340
봉입체 근병증 2	28937594	A>G	NC_000009.11,NC_000009.12	GNE	341
폰 빌레브란트 질환	41276738	C>T	NC_000012.11,NC_000012.12	VWF	342
유방 땀 또는 난소암	41293455	G>A	NC_000017.10,NC_000017.11	BRCA1	343
MECP2-관련 장애	61750240	G>A	NC_000023.10,NC_000023.11	MECP2	344
MEFV-관련 장애	61752717	T>C	NC_000016.9,NC_000016.10	MEFV	345
유방 땀 또는 난소암	62625307	G>A	NC_000017.10,NC_000017.11	BRCA1	346
유방 땀 또는 난소암	62625308	G>A	NC_000017.10,NC_000017.11	BRCA1	347
유전성 암-소인 증후군	63749795	C>T	NC_000003.11,NC_000003.12	MLH1	348
유전성 암-소인 증후군	63749849	C>T	NC_000002.11,NC_000002.12	MSH2	349
유전성 암-소인 증후군	63750636	C>T	NC_000002.11,NC_000002.12	MSH2	350
카르티린 칼미트릴트랜스퍼라제 II 결핍증	74315294	C>T	NC_000001.10,NC_000001.11	CPT2	351
낭성 섬유증	74597325	C>T	NC_000007.13,NC_000007.14	CFTR	352
UDP 글루코스-헥소스-1-포스페이트 우리딜트랜스퍼라제의 결핍증	75391579	A>G	NC_000009.11,NC_000009.12	GALT	353
낭성 섬유증	75527207	G>A	NC_000007.13,NC_000007.14	CFTR	354
아밀로이드생성 트랜스페레틴 아밀로이드증	76992529	G>A	NC_000018.9,NC_000018.10	TTR	355

결환	RS#	Cost Mut.	염색체 수탁	유전자 기호	표적 (서열식별번호)
남성 섬유증	77010898	G>A	NC_000007.13,NC_000007.14	CFTR	356
이염성 백질이영양증	80338815	C>T	NC_000022.10,NC_000022.11	ARSA	357
코우텐 증후군 3	80338844	C>T	NC_000011.9,NC_000011.10	SDHD	358
스미스-렐리-오퍼츠 증후군	80338853	G>A	NC_000011.9,NC_000011.10	DHCR7	359
유방 빛/또는 난소암	80356962	C>T	NC_000017.10,NC_000017.11	BRCAl	360
유방 빛/또는 난소암	80357123	G>A	NC_000017.10,NC_000017.11	BRCAl	361
선천성 유전 결환	80358259	A>G	NC_000018.9,NC_000018.10	NPC1	362
유방 빛/또는 난소암	80359212	C>T	NC_000013.10,NC_000013.11	BRCAl	363
판코니 빈혈	104886457	G>A	NC_000009.11,NC_000009.12	FANCC	364
SLC26A2-관련 장애	104893915	C>T	NC_000005.9,NC_000005.10	SLC26A2	365
심근병증	104894368	C>T	NC_000012.11,NC_000012.12	MYL2	366
단청, X-연관	104894396	C>T	NC_000013.10,NC_000013.11	GB2	367
선천성 유전 결환	104894635	C>T	NC_000017.10,NC_000017.11	SGSH	368
거북성 지중해열	104895097	C>T	NC_000016.9,NC_000016.10	MEFV	369
거북성 지중해열	111033171	A>G	NC_000009.11,NC_000009.12	ELP1	370
슈와크만 증후군	11393993	A>G	NC_000007.13,NC_000007.14	SBDS	371
RYR1-관련 장애	118192172	C>T	NC_000019.9,NC_000019.10	RYR1	372
세로이드 리포푸신증 뉴런 2	11945955	G>A	NC_000011.9,NC_000011.10	TPPI	373
중간쇄 아일-조효소 A 메히드로게나제 결핍증	12134274	G>A	NC_000001.10,NC_000001.11	ACADM	374
원발성 고옥살산뇨증	121908529	G>A	NC_000002.11,NC_000002.12	AGXT	375
심근병증	121908987	C>T	NC_000007.13,NC_000007.14	PRKAG2	376
코우텐 증후군	121909219	C>T	NC_000010.10,NC_000010.11	PTEN	377
심근병증	121913628	C>T	NC_000014.8,NC_000014.9	MYH7	378
선천성 유전 결환	121918243	G>A	NC_000001.10,NC_000001.11	MMACHC	379
PTPN11-관련 장애	121918457	C>T	NC_000012.11,NC_000012.12	PTPN11	380

질환	RS#	CasI Mut.	염색체 수탁	유전자 기호	표적 (서열식별번호)
소아 골수단핵구백혈병	121918462	C>T	NC_000012.11,NC_000012.12	PTPN11	381
소아 골수단핵구백혈병	121918466	A>G	NC_000012.11,NC_000012.12	PTPN11	382
뮤코폴리사카라이드증 유형 I	121965020	C>T	NC_000004.11,NC_000004.12	IDUA	383
세로이드 리포푸신증 뉴런 I	137852700	G>A	NC_000001.10,NC_000001.11	PPT1	384
CHEK2-관련 암 감수성	137853007	G>A	NC_000022.10,NC_000022.11	CHEK2	385
결장직장암	137854568	C>T	NC_000005.9,NC_000005.10	APC	386
가족성 고콜레스테롤혈증	137929307	G>A	NC_000019.9,NC_000019.10	LDLR	387
심장-얼굴-피부 증후	180177035	T>C	NC_000007.13,NC_000007.14	BRAF	388
가족성 유방암	180177083	G>A	NC_000016.10,NC_000016.9	PALB2	389
MYBPC3-관련 장애	200411226	C>T	NC_000011.9,NC_000011.10	MYBPC3	390
RYR1-관련 장애	200563280	C>T	NC_000019.9,NC_000019.10	RYR1	391
MYBPC3-관련 장애	387907267	G>A	NC_000011.9,NC_000011.10	MYBPC3	392
테스토이드 질환, 유전성	397515734	C>T	NC_000005.9,NC_000005.10	APC	393
마르팡 증후군/로이스네즈 증후군/가족성 흉부 대동맥류 및 박리	397515757	C>T	NC_000015.9,NC_000015.10	FBN1	394
면역결핍 14	397518423	G>A	NC_000001.10,NC_000001.11	PIK3CD	395
선천성 유전 결핍	398123009	C>T	NC_000011.9,NC_000011.10	PACSL	396
B 림프모구성 백혈병 림프종, ICD-O 하위유형 없음	529008617	G>A	NC_000001.10,NC_000001.11	MUTYH	397
가족성 유방암	587780021	G>A	NC_000002.11,NC_000002.12	BARD1	398
가족성 고콜레스테롤혈증	746118995	C>T	NC_000019.9,NC_000019.10	LDLR	399
가족성 고콜레스테롤혈증	769370816	G>A	NC_000019.10,NC_000019.9	LDLR	400

표 9.2: APG6646.1 또는 APG01658.1에 대한 질환 표적

질환	RGN	RS#	CasI Mut.	염색체 수박	유전자 기호	표적 (서열식별번호)
스타르가르트병 1	APG01658.1	1800553	C>T	NC_000001.10,NC_000001.11	ABCA4	538
스타르가르트병 1	APG06646.1	1800553	C>T	NC_000001.10,NC_000001.11	ABCA4	539
유전성 안-소인 증후군	APG01658.1	5030804	A>G	NC_000003.11,NC_000003.12	VHL	540
유전성 안-소인 증후군	APG06646.1	5030804	A>G	NC_000003.11,NC_000003.12	VHL	541
페닐케톤뇨증	APG01658.1	5030851	G>A	NC_000012.11,NC_000012.12	PAH	542
페닐케톤뇨증	APG06646.1	5030851	G>A	NC_000012.11,NC_000012.12	PAH	543
호색증 및 갈수증	APG01658.1	11547328	G>A	NC_000012.11,NC_000012.12	CDR4	544
호색증 및 갈수증	APG06646.1	11547328	G>A	NC_000012.11,NC_000012.12	CDR4	545
가족성 고콜레스테롤혈증	APG01658.1	11547917	C>T	NC_000019.9,NC_000019.10	LDLR	546
가족성 고콜레스테롤혈증	APG06646.1	11547917	C>T	NC_000019.9,NC_000019.10	LDLR	547
샤르코-마리-투스병	APG01658.1	61444459	G>A	NC_000001.10,NC_000001.11	LMNA	548
샤르코-마리-투스병	APG06646.1	61444459	G>A	NC_000001.10,NC_000001.11	LMNA	549
안젤만 증후군	APG01658.1	61751362	G>A	NC_000023.10,NC_000023.11	MECP2	550
안젤만 증후군	APG06646.1	61751362	G>A	NC_000023.10,NC_000023.11	MECP2	551
페닐케톤뇨증	APG01658.1	62514895	C>T	NC_000012.11,NC_000012.12	PAH	552
페닐케톤뇨증	APG06646.1	62514895	C>T	NC_000012.11,NC_000012.12	PAH	553
페닐케톤뇨증	APG01658.1	62516152	C>T	NC_000012.11,NC_000012.12	PAH	554
페닐케톤뇨증	APG06646.1	62516152	C>T	NC_000012.11,NC_000012.12	PAH	555
린치 증후군	APG01658.1	63749795	C>T	NC_000003.11,NC_000003.12	MLH1	556
린치 증후군	APG06646.1	63749795	C>T	NC_000003.11,NC_000003.12	MLH1	557
유전성 암	APG01658.1	63749843	C>T	NC_000002.11,NC_000002.12	MSH6	558
유전성 암	APG06646.1	63749843	C>T	NC_000002.11,NC_000002.12	MSH6	559

결환	RCN	RS#	Cast Mut.	임상세 수탁	유전자 기호	포적 (서열식별번호)
유전성 암-소인 중후군	APG01638.1	63749849	C>T	NC_000002.11,NC_000002.12	MSH2	560
유전성 암-소인 중후군	APG06646.1	63749849	C>T	NC_000002.11,NC_000002.12	MSH2	561
유전성 암-소인 중후군	APG01638.1	63750308	C>T	NC_000002.11,NC_000002.12	MSH2	562
유전성 암-소인 중후군	APG06646.1	63750308	C>T	NC_000002.11,NC_000002.12	MSH2	563
유전성 암-소인 중후군	APG01638.1	63750636	C>T	NC_000002.11,NC_000002.12	MSH2	564
유전성 암-소인 중후군	APG06646.1	63750636	C>T	NC_000002.11,NC_000002.12	MSH2	565
유전성 암-소인 중후군	APG06646.1	63750871	G>A	NC_000007.13,NC_000007.14	PMS2	566
린치 증후군	APG01638.1	63751194	C>T	NC_000003.11,NC_000003.12	MLH1	567
린치 증후군	APG06646.1	63751194	C>T	NC_000003.11,NC_000003.12	MLH1	568
린치 증후군	APG01638.1	63751711	G>A	NC_000003.11,NC_000003.12	MLH1	569
린치 증후군	APG06646.1	63751711	G>A	NC_000003.11,NC_000003.12	MLH1	570
낭성 섬유증	APG01638.1	75039782	C>T	NC_000007.13,NC_000007.14	CFTR	571
낭성 섬유증	APG06646.1	75039782	C>T	NC_000007.13,NC_000007.14	CFTR	572
낭성	APG01638.1	76434661	C>T	NC_000013.10,NC_000013.11	GJB2	573
낭성	APG06646.1	76434661	C>T	NC_000013.10,NC_000013.11	GJB2	574
낭성 섬유증	APG01638.1	76713772	G>A	NC_000007.13,NC_000007.14	CFTR	575
낭성 섬유증	APG06646.1	76992529	G>A	NC_000018.9,NC_000018.10	TTR	576
낭성 섬유증	APG01638.1	76992529	G>A	NC_000018.9,NC_000018.10	TTR	577
낭성 섬유증	APG06646.1	77010898	G>A	NC_000007.13,NC_000007.14	CFTR	579
낭성 섬유증	APG01638.1	80224560	G>A	NC_000007.13,NC_000007.14	CFTR	580
낭성 섬유증	APG06646.1	80224560	G>A	NC_000007.13,NC_000007.14	CFTR	581
낭성 섬유증	APG01638.1	80338844	C>T	NC_000011.9,NC_000011.10	SDHD	583
낭성 섬유증	APG06646.1	80338844	C>T	NC_000011.9,NC_000011.10	SDHD	584

질환	RCN	RS#	CasI Mut.	염색체 수태	유전자 기호	표제 (서열식별번호)
SLC26A4-관련 장애	APG01658.1	80338849	G>A	NC_000007.13,NC_000007.14	SLC26A4	585
SLC26A4-관련 장애	APG06646.1	80338849	G>A	NC_000007.13,NC_000007.14	SLC26A4	586
유광-난소암	APG01658.1	80356885	C>T	NC_000017.10,NC_000017.11	BRC A1	587
유광-난소암	APG06646.1	80356885	C>T	NC_000017.10,NC_000017.11	BRC A1	588
유광-난소암	APG01658.1	80358557	C>T	NC_000013.10,NC_000013.11	BRC A2	589
유광-난소암	APG06646.1	80358557	C>T	NC_000013.10,NC_000013.11	BRC A2	590
유광-난소암	APG01658.1	80358650	G>A	NC_000013.10,NC_000013.11	BRC A2	591
유광-난소암	APG06646.1	80358650	G>A	NC_000013.10,NC_000013.11	BRC A2	592
유광-난소암	APG01658.1	80358659	C>T	NC_000013.10,NC_000013.11	BRC A2	593
유광-난소암	APG06646.1	80358659	C>T	NC_000013.10,NC_000013.11	BRC A2	594
유광-난소암	APG06646.1	80358663	C>T	NC_000013.10,NC_000013.11	BRC A2	595
유광-난소암	APG01658.1	80358871	G>A	NC_000013.10,NC_000013.11	BRC A2	596
유광-난소암	APG06646.1	80358871	G>A	NC_000013.10,NC_000013.11	BRC A2	597
유광-난소암	APG01658.1	80358920	C>T	NC_000013.10,NC_000013.11	BRC A2	598
유광-난소암	APG06646.1	80358920	C>T	NC_000013.10,NC_000013.11	BRC A2	599
유광-난소암	APG01658.1	80358972	C>T	NC_000013.10,NC_000013.11	BRC A2	600
유광-난소암	APG06646.1	80358972	C>T	NC_000013.10,NC_000013.11	BRC A2	601
유광-난소암	APG01658.1	80358981	C>T	NC_000013.10,NC_000013.11	BRC A2	602
유광-난소암	APG06646.1	80358981	C>T	NC_000013.10,NC_000013.11	BRC A2	603
유광-난소암	APG01658.1	80359003	G>A	NC_000013.10,NC_000013.11	BRC A2	604
유광-난소암	APG06646.1	80359003	G>A	NC_000013.10,NC_000013.11	BRC A2	605
유광-난소암	APG01658.1	80359004	G>A	NC_000013.10,NC_000013.11	BRC A2	606
유광-난소암	APG06646.1	80359004	G>A	NC_000013.10,NC_000013.11	BRC A2	607
유광-난소암	APG01658.1	80359013	G>A	NC_000013.10,NC_000013.11	BRC A2	608
유광-난소암	APG06646.1	80359013	G>A	NC_000013.10,NC_000013.11	BRC A2	609

결판	RCN	RS#	Cast Mut.	염색체 수탁	유전자 기호	표적 (서열식별번호)
유방-난소암	APG01638.1	80359027	G>A	NC_000013.10,NC_000013.11	BRC A2	610
유방-난소암	APG06646.1	80359027	G>A	NC_000013.10,NC_000013.11	BRC A2	611
유방-난소암	APG01638.1	80359071	G>A	NC_000013.10,NC_000013.11	BRC A2	612
유방-난소암	APG06646.1	80359071	G>A	NC_000013.10,NC_000013.11	BRC A2	613
유방-난소암	APG01638.1	80359115	C>T	NC_000013.10,NC_000013.11	BRC A2	614
유방-난소암	APG06646.1	80359115	C>T	NC_000013.10,NC_000013.11	BRC A2	615
유방-난소암	APG01638.1	80359140	C>T	NC_000013.10,NC_000013.11	BRC A2	616
유방-난소암	APG06646.1	80359140	C>T	NC_000013.10,NC_000013.11	BRC A2	617
유방-난소암	APG06646.1	80359159	C>T	NC_000013.10,NC_000013.11	BRC A2	618
유방-난소암	APG01638.1	80359180	C>T	NC_000013.10,NC_000013.11	BRC A2	619
유방-난소암	APG06646.1	80359180	C>T	NC_000013.10,NC_000013.11	BRC A2	620
유방-난소암	APG01638.1	81002846	G>A	NC_000013.10,NC_000013.11	BRC A2	621
유방-난소암	APG06646.1	81002846	G>A	NC_000013.10,NC_000013.11	BRC A2	622
유방-난소암	APG06646.1	81002862	A>G	NC_000013.10,NC_000013.11	BRC A2	623
관코니 빈혈	APG01638.1	104886457	G>A	NC_000009.11,NC_000009.12	FANCC	624
관코니 빈혈	APG06646.1	104886457	G>A	NC_000009.11,NC_000009.12	FANCC	625
단골무발생증	APG01638.1	104893915	C>T	NC_000005.9,NC_000005.10	SLC26A2	626
단골무발생증	APG06646.1	104893915	C>T	NC_000005.9,NC_000005.10	SLC26A2	627
RYR1-관련 장애	APG01638.1	118192172	C>T	NC_000019.9,NC_000019.10	RYR1	628
RYR1-관련 장애	APG06646.1	118192172	C>T	NC_000019.9,NC_000019.10	RYR1	629
CHEK2-관련 암 감수성	APG01638.1	137853007	G>A	NC_000022.10,NC_000022.11	CHEK2	630
CHEK2-관련 암 감수성	APG06646.1	137853007	G>A	NC_000022.10,NC_000022.11	CHEK2	631
신경섬유종증	APG06646.1	137854550	A>G	NC_000017.10,NC_000017.11	NF1	632
신경섬유종증	APG01638.1	137854556	G>A	NC_000017.10,NC_000017.11	NF1	633
신경섬유종증	APG06646.1	137854556	G>A	NC_000017.10,NC_000017.11	NF1	634

질환	RCN	RS#	Cas# Mut.	염색체 수태	유전자 기호	표적 (서열식별번호)
근병증	APG01658.1	193922390	C>T	NC_000014.8,NC_000014.9	MYH7	635
근병증	APG06646.1	193922390	C>T	NC_000014.8,NC_000014.9	MYH7	636
긴 QT 증후군	APG01658.1	199473428	C>T	NC_000007.13,NC_000007.14	KCNH2	637
긴 QT 증후군	APG06646.1	199473428	C>T	NC_000007.13,NC_000007.14	KCNH2	638
페닐케톤뇨증	APG01658.1	199475575	G>A	NC_000012.11,NC_000012.12	PAH	639
페닐케톤뇨증	APG06646.1	199475575	G>A	NC_000012.11,NC_000012.12	PAH	640
페닐케톤뇨증	APG01658.1	199475679	C>T	NC_000012.11,NC_000012.12	PAH	641
페닐케톤뇨증	APG06646.1	199475679	C>T	NC_000012.11,NC_000012.12	PAH	642
MYBPC3-관련 장애	APG01658.1	200411226	C>T	NC_000011.9,NC_000011.10	MYBPC3	643
MYBPC3-관련 장애	APG06646.1	200411226	C>T	NC_000011.9,NC_000011.10	MYBPC3	644
근병증	APG06646.1	267606908	T>C	NC_000014.8,NC_000014.9	MYH7	645
심근병증	APG01658.1	267607003	C>T	NC_000010.10,NC_000010.11	RBM20	646
심근병증	APG06646.1	267607003	C>T	NC_000010.10,NC_000010.11	RBM20	647
심근병증	APG01658.1	267607048	A>G	NC_000010.10,NC_000010.11	SHOC2	648
심근병증	APG06646.1	267607048	A>G	NC_000010.10,NC_000010.11	SHOC2	649
근단 증후군	APG01658.1	267607213	G>A	NC_000019.9,NC_000019.10	LDLR	650
근단 증후군	APG06646.1	267607213	G>A	NC_000019.9,NC_000019.10	LDLR	651
가족성 고콜레스테롤혈증	APG01658.1	267607789	G>A	NC_000003.11,NC_000003.12	MLH1	652
가족성 고콜레스테롤혈증	APG06646.1	267607789	G>A	NC_000003.11,NC_000003.12	MLH1	653
린치 증후군	APG01658.1	267607845	G>A	NC_000003.11,NC_000003.12	MLH1	654
린치 증후군	APG06646.1	267607845	G>A	NC_000003.11,NC_000003.12	MLH1	655
유전성 암-소인 증후군	APG01658.1	267608098	A>G	NC_000002.11,NC_000002.12	MSH6	656
유전성 암-소인 증후군	APG06646.1	267608098	A>G	NC_000002.11,NC_000002.12	MSH6	657
MYBPC3-관련 장애	APG01658.1	387906397	A>G	NC_000011.9,NC_000011.10	MYBPC3	658
MYBPC3-관련 장애	APG06646.1	387906397	A>G	NC_000011.9,NC_000011.10	MYBPC3	659

질환	RCN	RS#	Casl Mut.	염색체 수탁	유전자 기호	표적 (서열식별번호)
유방-난소암	APG01638.1	387906843	G>A	NC_000017.10,NC_000017.11	RAD51D	660
유방-난소암	APG06646.1	387906843	G>A	NC_000017.10,NC_000017.11	RAD51D	661
MYBPC3-관련 장애	APG01638.1	387907267	G>A	NC_000011.9,NC_000011.10	MYBPC3	662
MYBPC3-관련 장애	APG06646.1	387907267	G>A	NC_000011.9,NC_000011.10	MYBPC3	663
유방-난소암	APG01638.1	397507389	G>A	NC_000013.10,NC_000013.11	BRC42	664
유방-난소암	APG06646.1	397507389	G>A	NC_000013.10,NC_000013.11	BRC42	665
유방-난소암	APG01638.1	397507404	G>A	NC_000013.10,NC_000013.11	BRC42	666
유방-난소암	APG06646.1	397507404	G>A	NC_000013.10,NC_000013.11	BRC42	667
백혈병	APG01638.1	397507545	G>A	NC_000012.11,NC_000012.12	PTPN11	668
백혈병	APG06646.1	397507545	G>A	NC_000012.11,NC_000012.12	PTPN11	669
백혈병	APG01638.1	397507547	A>G	NC_000012.11,NC_000012.12	PTPN11	670
백혈병	APG06646.1	397507547	A>G	NC_000012.11,NC_000012.12	PTPN11	671
유방-난소암	APG01638.1	397507922	G>A	NC_000013.10,NC_000013.11	BRC42	672
유방-난소암	APG06646.1	397507922	G>A	NC_000013.10,NC_000013.11	BRC42	673
마르판 증후군	APG01638.1	397515757	C>T	NC_000015.9,NC_000015.10	FBN1	674
마르판 증후군	APG06646.1	397515757	C>T	NC_000015.9,NC_000015.10	FBN1	675
MYBPC3-관련 장애	APG01638.1	397516074	C>T	NC_000011.9,NC_000011.10	MYBPC3	676
MYBPC3-관련 장애	APG06646.1	397516074	C>T	NC_000011.9,NC_000011.10	MYBPC3	677
심근병증	APG01638.1	397516354	C>T	NC_000019.9,NC_000019.10	TNNI3	678
심근병증	APG06646.1	397516354	C>T	NC_000019.9,NC_000019.10	TNNI3	679
면역결핍	APG01638.1	397518423	G>A	NC_000001.10,NC_000001.11	PRK3CD	680
면역결핍	APG06646.1	397518423	G>A	NC_000001.10,NC_000001.11	PRK3CD	681
B 림프모구성 백혈병 림프종	APG01638.1	529008617	G>A	NC_000001.10,NC_000001.11	MUTYH	682
B 림프모구성 백혈병 림프종	APG06646.1	529008617	G>A	NC_000001.10,NC_000001.11	MUTYH	683
가족성 고콜레스테롤혈증	APG01638.1	570942190	C>T	NC_000019.9,NC_000019.10	LDLR	684

질환	RCN	RS#	Cast Mut.	염색체 수탁	유전자 기호	표적 (서열식별번호)
거족성 고콜레스테롤혈증	APG06646.1	570942190	C>T	NC_000019.9,NC_000019.10	LDLR	685
운동실조-모세혈관확장증 증후군	APG01638.1	587776551	G>A	NC_000011.10,NC_000011.9	ATM	686
운동실조-모세혈관확장증 증후군	APG06646.1	587776551	G>A	NC_000011.10,NC_000011.9	ATM	687
선양 낭성 암종	APG01638.1	587778720	C>T	NC_000017.10,NC_000017.11	TP53	688
선양 낭성 암종	APG06646.1	587778720	C>T	NC_000017.10,NC_000017.11	TP53	689
유전성 암-소인 증후군	APG01638.1	587779075	C>T	NC_000002.11,NC_000002.12	MSH2	690
유전성 암-소인 증후군	APG06646.1	587779075	C>T	NC_000002.11,NC_000002.12	MSH2	691
유전성 암-소인 증후군	APG06646.1	587779333	T>C	NC_000007.13,NC_000007.14	PMS2	692
운동실조-모세혈관확장증 증후군	APG01638.1	587779866	A>G	NC_000011.10,NC_000011.9	ATM	693
운동실조-모세혈관확장증 증후군	APG06646.1	587779866	A>G	NC_000011.10,NC_000011.9	ATM	694
유방암	APG01638.1	587780021	G>A	NC_000002.11,NC_000002.12	BARD1	695
유방암	APG06646.1	587780021	G>A	NC_000002.11,NC_000002.12	BARD1	696
유방-난소암	APG01638.1	587781629	G>A	NC_000013.10,NC_000013.11	BRC A2	697
유방-난소암	APG06646.1	587781629	G>A	NC_000013.10,NC_000013.11	BRC A2	698
유전성 암-소인 증후군	APG01638.1	587782144	C>T	NC_000017.10,NC_000017.11	TP53	699
유전성 암-소인 증후군	APG06646.1	587782144	C>T	NC_000017.10,NC_000017.11	TP53	700
마르팡 증후군	APG01638.1	727503054	A>G	NC_000015.9,NC_000015.10	FBN1	701
마르팡 증후군	APG06646.1	727503054	A>G	NC_000015.9,NC_000015.10	FBN1	702
난관 증후군	APG01638.1	727503110	T>C	NC_000012.11,NC_000012.12	KRAS	703
난관 증후군	APG06646.1	727503110	T>C	NC_000012.11,NC_000012.12	KRAS	704
거족성 고콜레스테롤혈증	APG01638.1	746118995	C>T	NC_000019.9,NC_000019.10	LDLR	705
거족성 고콜레스테롤혈증	APG06646.1	746118995	C>T	NC_000019.9,NC_000019.10	LDLR	706
거족성 고콜레스테롤혈증	APG01638.1	748944640	G>A	NC_000019.9,NC_000019.10	LDLR	707
거족성 고콜레스테롤혈증	APG06646.1	748944640	G>A	NC_000019.9,NC_000019.10	LDLR	708
거족성 고콜레스테롤혈증	APG01638.1	765696008	G>A	NC_000019.10,NC_000019.9	LDLR	709

질환	RGN	RS#	Cas9 Mut.	염색체 수박	유전자 기호	표적 (서열식별번호)
가족성 고콜레스테롤혈증	APG06646.1	765696008	G>A	NC_000019.10,NC_000019.9	LDLR	710
가족성 고콜레스테롤혈증	APG01638.1	769370816	G>A	NC_000019.10,NC_000019.9	LDLR	711
가족성 고콜레스테롤혈증	APG06646.1	769370816	G>A	NC_000019.10,NC_000019.9	LDLR	712
가족성 고콜레스테롤혈증	APG01638.1	769737896	C>T	NC_000019.10,NC_000019.9	LDLR	713
가족성 고콜레스테롤혈증	APG06646.1	769737896	C>T	NC_000019.10,NC_000019.9	LDLR	714
가족성 고콜레스테롤혈증	APG01638.1	771019366	A>G	NC_000019.10,NC_000019.9	LDLR	715
가족성 고콜레스테롤혈증	APG06646.1	771019366	A>G	NC_000019.10,NC_000019.9	LDLR	716
가족성 고콜레스테롤혈증	APG01638.1	772551056	C>T	NC_000001.11,NC_000001.10	SDHB	717
가족성 고콜레스테롤혈증	APG06646.1	772551056	C>T	NC_000001.11,NC_000001.10	SDHB	718
가족성 고콜레스테롤혈증	APG01638.1	786201042	C>T	NC_000002.12,NC_000002.11	MSH6	719
가족성 고콜레스테롤혈증	APG06646.1	786201042	C>T	NC_000002.12,NC_000002.11	MSH6	720
가족성 고콜레스테롤혈증	APG01638.1	875989907	G>A	NC_000019.9,NC_000019.10	LDLR	721
가족성 고콜레스테롤혈증	APG06646.1	875989907	G>A	NC_000019.9,NC_000019.10	LDLR	722
가족성 고콜레스테롤혈증	APG01638.1	879254797	G>A	NC_000019.10,NC_000019.9	LDLR	723
가족성 고콜레스테롤혈증	APG06646.1	879254797	G>A	NC_000019.10,NC_000019.9	LDLR	724
가족성 고콜레스테롤혈증	APG01638.1	879254871	C>T	NC_000019.10,NC_000019.9	LDLR	725
가족성 고콜레스테롤혈증	APG06646.1	879254871	C>T	NC_000019.10,NC_000019.9	LDLR	726
가족성 고콜레스테롤혈증	APG01638.1	879255000	T>C	NC_000019.10,NC_000019.9	LDLR	727
가족성 고콜레스테롤혈증	APG06646.1	879255000	T>C	NC_000019.10,NC_000019.9	LDLR	728

실시예 8: 프리드라이히 운동실조의 원인이 되는 표적화 돌연변이

프리드라이히 운동실조 (FRDA)를 초래하는 트리뉴클레오타이드 반복 서열의 확장은 FRDA 불안정 영역으로 지칭되는 FXN 유전자 내의 정의된 유전자좌에서 발생한다. RNA 가이드된 뉴클레아제 (RGN)는 FRDA 환자 세포에서 불안정 영역을 절제하기 위해 사용될 수 있다. 이 접근법은 다음을 필요로 한다: 1) 인간 게놈에서 대립유전자를 표적화도록 프로그래밍될 수 있는 RGN 및 가이드 RNA 서열; 및 2) RGN 및 가이드 서열에 대한 전달 접근법. 게놈 편집을 위해 사용되는 여러 뉴클레아제, 예컨대 흔히 사용되는 에스. 피로케네스로부터의 Cas9 뉴클레아제 (SpCas9)는 본질적으로 기능적 발현 카세트를 위해 필요한 다른 유전자 요소 외에도 SpCas9 유전자 및 가이드 RNA의 길이를 고려할 때 너무 커서 아데노-연관된 바이러스 (AAV) 벡터에 패키징되지 않는다. 이는 SpCas9를 사용하는 실행가능한 접근법을 가능하지 않게 만든다.

본 발명의 조밀한 RNA 가이드된 뉴클레아제, 예컨대 APG09748, APG09106.1, 및 APG06646.1은 FRDA 불안정 영역의 절제에 매우 적합하다. 각각의 RGN은 FRDA 불안정 영역 근처에 있는 PAM 요건을 갖는다. 추가로, 이들 RGN 각각은 가이드 RNA와 함께 AAV 벡터에 패키징될 수 있다. 2개의 가이드 RNA의 패키징은 제2 벡터를 필요로 할 수도 있지만, 이 접근법은 두 벡터 사이의 단백질 서열을 분할할 수 있는 더 큰 뉴클레아제, 예컨대 SpCas9에서 필요로 되는 것보다 여전히 유리하게 비교된다.

표 10은 FRDA 불안정 영역의 5' 및 3' 플랭킹으로 APG09748, APG09106.1, 또는 APG06646.1을 표적화하는데 적합한 게놈 표적 서열의 위치, 뿐만 아니라 게놈 표적에 대한 sgRNA의 서열을 나타낸다. 유전자좌에서, RGN은 FA 불안정 영역을 절제할 것이다. 상기 영역의 절제는 유전자좌의 일루미나 서열분석에 의해 검증될 수 있다.

[0518] 표 10: RGN 시스템에 대한 게놈 표적 서열

FRDA 불안정 영역에 대한 위치	APG09748 또는 APG09106.1에 대한 게놈 표적 서열 (서열식별번호)	APG09748 또는 APG09106.1에 대한 sgRNA (서열식별번호)	APG06646.1에 대한 게놈 표적 서열 (서열식별번호)	APG06646.1에 대한 sgRNA (서열식별번호)
5'	401	405	729	733
5'	402	406	730	734
3'	403	407	731	735
3'	404	408	732	736

[0519] 실시예 9: 겸상 적혈구 질환의 원인이 되는 표적화 돌연변이

[0520]

[0521] BCL11A 인핸서 영역 (서열식별번호: 220) 내의 표적화 서열은 겸상 적혈구 질환의 증상을 치유하거나 또는 완화시키기 위해 태아 헤모글로빈 (HbF)을 증가시키는 메커니즘을 제공할 수 있다. 예를 들어, 게놈 차원의 연관성 연구에 의해 증가된 HbF 수준과 연관된 BCL11A에서의 유전자 변이의 집합을 확인하였다. 이들 변이는 단계-특이적인 계통-제한된 인핸서 영역으로서 기능하는 BCL11A의 비-코딩 영역에서 발견되는 SNP의 집합이다. 추가의 연구를 통해, 이 BCL11A 인핸서가 BCL11A 발현을 위해 적혈구 세포에서 필요하다는 것이 밝혀졌다 ([Bauer et al., (2013) Science 343:253-257], 본원에 참조로 포함됨). 인핸서 영역은 BCL11A 유전자의 인트론 2 내에서 발견되었고, 인트론 2에서 DNaseI 과민성 (종종 조절 능력과 연관된 염색질 상태를 나타냄)인 3개의 영역이 확인되었다. 이들 3개의 영역은 BCL11A의 전사 시작 부위로부터의 거리 (킬로염기)에 따라 "+62", "+58" 및 "+55"로 확인되었다. 이들 인핸서 영역은 대략 350 (+55); 550 (+58); 및 350 (+62) 뉴클레오티드 길이이다 (Bauer et al., 2013).

[0522] 실시예 9.1: 바람직한 RGN 시스템의 확인

[0523] 여기서 BCL11A가 성인 헤모글로빈에서 베타-글로빈을 제조하는 것을 담당하는 유전자인 HBB 유전자좌 내의 그의 결합 부위로 결합하는 것을 방해하는 RGN 시스템을 사용하는 베타-혈색소병증의 잠재적인 치료가 기재된다. 이 접근법은 포유동물 세포에서 더욱 효율적인 NHEJ를 사용한다. 또한, 이 접근법은 생체내 전달을 위해 단일 AAV 벡터로 패키징될 수 있는 충분히 작은 크기를 갖는 뉴클레아제를 사용한다.

[0524] 인간 BCL11A 인핸서 영역 (서열식별번호: 220)에서 GATA1 인핸서 모티프는 성인 인간 적혈구에서 HbF의 동시 재발현과 함께 BCL11A 발현을 감소시키기 위해 RNA 가이드된 뉴클레아제 (RGN)를 사용하여 방해하기 위한 이상적인 표적이다 ([Wu et al. (2019) Nat Med 387:2554]). APG09748 또는 APG09106.1과 상용성인 몇몇 PAM 서열은 이 GATA1 부위를 둘러싸는 유전자좌에서 용이하게 명백하다. 이들 뉴클레아제는 5'-DTTN-3'의 PAM 서열 (서열식별번호: 60)을 갖고, 소형의 크기를 가지며, 이는 잠재적으로 단일 AAV 또는 아데노바이러스 벡터에서 적절한 가이드 RNA와 함께 그들의 전달을 가능하게 한다. 그의 크기에 추가로, APG06646.1은 최소 PAM 요건 (서열식별번호: 109)을 가지며, 이는 이를 상기 접근법에 매우 적합하게 만든다. 이 전달 접근법은 다른 것들에 비해 조절 줄기 세포로의 접근, 및 잘 확립된 안전성 프로파일 및 제작 기술과 같은 다양한 이점을 제공한다.

[0525] 흔히 사용되는 에스. 피로게네스로부터의 Cas9 뉴클레아제 (SpyCas9)는 5'-NGG-3'의 PAM 서열 (서열식별번호: 101)을 필요로 하며, 이 중에서 몇몇은 GATA1 모티프 근처에 존재한다. 그러나, SpyCas9의 크기는 단일 AAV 또는 아데노바이러스 벡터로 패키징되는 것을 방지하며, 따라서 상기 언급된 이 접근법의 이점을 포기한다. 이중 전달 전략을 이용할 수 있지만, 이는 상당한 제작 복잡성 및 비용을 부가할 것이다. 추가로, 주어진 세포에서 성공적으로 편집되려면 두 벡터 모두로의 감염을 필요로 하기 때문에, 이중 바이러스 벡터 전달은 유전자 수정의 효율을 상당히 감소시킨다.

[0526] 인간 코돈 최적화된 APG09748 (서열식별번호: 409), APG09106.1 (서열식별번호: 410), 또는 APG06646.1 (서열식별번호: 415)을 코딩하는 발현 카세트는 실시예 5에 기재된 것과 유사하게 생성된다. RGN APG09748, APG09106.1, 또는 APG06646.1에 대한 가이드 RNA를 발현하는 발현 카세트 또한 생성된다. 이들 가이드 RNA는 다음을 포함한다: 1) BCL11A 인핸서 유전자좌 (표적 서열) 내의 비-코딩 또는 코딩 DNA 가닥에 대해 상보적인 프로토스페이서 서열 및 2) 가이드 RNA와 RGN의 회합을 위해 필요한 RNA 서열. 각각의 RGN에 의해 표적화하기 위한 몇몇 잠재적인 PAM 서열은 BCL11A GATA1 인핸서 모티프를 둘러싸고 있기 때문에, BCL11A GATA1 인핸서 서열의 강건한 절단 및 NHEJ 매개된 방해를 제공하는 최상의 프로토스페이서 서열을 결정하기 위해 몇몇 잠재적인 가이드 RNA 구축물이 생성된다. 표 11에서 표적 게놈 서열은 표 11에 제공된 sgRNA를 이용하여 평가된다.

표 11: APG06646.1을 사용한 BCL11A GATA1 인핸서 유전자좌에 대한 표적 서열

APG09748 또는 APG0916.1에 대한 표적 계놈 서열 (서열식별번호)	APG09748 또는 APG0916.1에 대한 sgRNA (서열식별번호)	APG06646.1에 대한 표적 계놈 서열 (서열식별번호)	APG06646.1에 대한 sgRNA (서열식별번호)
221	224	737	740
222	269	738	741
223	270	739	742

APG09748, APG09106.1, 또는 APG06646.1이 BCL11A 인핸서 영역을 방해하는 삽입 또는 결실을 생성하는 효율을 평가하기 위해, 인간 세포주, 예컨대 인간 배아 신장 세포 (HEK 세포)를 사용한다. RGN 발현 카세트를 포함하는 DNA 벡터 (예를 들어, 실시예 5에 기재됨)를 생성한다. 표 11의 가이드 RNA 서열에 대한 코딩 서열을 포함하는 발현 카세트를 포함하는 별도의 벡터를 또한 생성한다. 이러한 발현 카세트는 실시예 5에 기재된 바와 같이 인간 RNA 폴리머라제 III U6 프로모터 (서열식별번호: 308)를 추가로 포함할 수 있다. 대안적으로, RGN 및 가이드 RNA 둘 다의 발현 카세트를 포함하는 단일 벡터가 사용될 수 있다. 벡터를 실시예 5에 기재된 것과 같은 표준 기술을 이용하여 HEK 세포에 도입시키고, 세포를 1-3 일 동안 배양한다. 이 배양 기간 이후에, 계놈 DNA를 단리하고, 삽입 또는 결실의 빈도를 실시예 5에 기재된 바와 같이 T7 엔도뉴클레아제 I 소화 및/또는 직접적인 DNA 서열분석을 이용하여 결정한다.

표적 BCL11A 영역을 포함하는 DNA 영역을 일루미나 넥스테라 XT 오버행 서열을 함유하는 프라이머를 사용하여 PCR에 의해 증폭시킨다. 이들 PCR 앰플리콘은 T7 엔도뉴클레아제 I 소화를 이용하여 NHEJ 형성에 대해 시험하거나, 또는 일루미나 16S 메타게놈 시퀀싱 라이브러리 프로토콜 또는 유사한 차세대 서열분석 (NGS) 라이브러리 제조에 따라 라이브러리 제조를 거친다. 심층 서열분석 이후에, 생성된 판독을 CRISPResso에 의해 분석하여, 편집률을 계산한다. 출력 정렬을 수동-규레이션하여, 삽입 및 결실 부위를 확인한다. 이 분석을 통해 바람직한 RGN 및 상응하는 바람직한 가이드 RNA (sgRNA)가 확인된다. 상기 분석은 APG09748, APG09106.1, 또는 APG06646.1이 동등하게 바람직하게 하거나, 하나의 RGN이 가장 바람직하게 할 수 있다. 추가로, 상기 분석은 1 개 초과인 바람직한 가이드 RNA가 있거나, 또는 표 17의 모든 표적 계놈 서열이 동등하게 바람직함을 결정할 수 있다.

실시예 9.2: 태아 헤모글로빈의 발현에 대한 검정

이 실시예에서, BCL11A 인핸서 영역을 방해하는 APG09748, APG09106.1, 또는 APG06646.1 생성된 삽입 또는 결실을 태아 헤모글로빈의 발현에 대해 검정한다. 건강한 인간 공여자 CD34⁺ 조혈 줄기 세포 (HSC)를 사용한다. 이들 HSC를 배양하고, 바람직한 RGN 및 바람직한 sgRNA의 코딩 영역을 포함하는 발현 카세트를 포함하는 벡터 (들)을 실시예 5에 기재된 것과 유사한 방법을 이용하여 도입시킨다. 대안적으로, 전기천공이 사용될 수 있다. 전기천공 이후에, 확립된 프로토콜을 이용하여 시험관내에서 이들 세포를 적혈구로 분화시킨다 (예를 들어, [Giarratana et al. (2004) Nat Biotechnology 23:69-74], 본원에 참조로 포함됨). 이어서, HbF의 발현을 항-인간 HbF 항체에 의해 웨스턴 블롯팅을 이용하여 측정하거나, 또는 고성능 액체 크로마토그래피 (HPLC)를 이용하여 정량화한다. BCL11A 인핸서 유전자좌의 성공적인 방해가 가이드 없이 RGN에 의해서만 전기천공된 HSC와 비교할 때 HbF 생성의 증가를 유도할 것으로 예상된다.

실시예 9.3: 감소된 겸상 적혈구 형성에 대한 검정

이 실시예에서, BCL11A 인핸서 영역을 방해하는 APG09748, APG09106.1, 또는 APG06646.1 생성된 삽입 또는 결실을 감소된 겸상 적혈구 형성에 대해 검정한다. 겸상 적혈구 질환에 걸린 환자로부터의 공여자 CD34⁺ 조혈 줄기 세포 (HSC)를 사용한다. 이들 HSC를 배양하고, 바람직한 RGN 및 바람직한 sgRNA의 코딩 영역을 포함하는 발현 카세트를 포함하는 벡터 (들)을 실시예 5에 기재된 것과 유사한 방법을 이용하여 도입시킨다. 대안적으로, 전기천공이 사용될 수 있다. 전기천공 이후에, 확립된 프로토콜을 이용하여 시험관내에서 이들 세포를 적혈구로 분화시킨다 (예를 들어, [Giarratana et al. (2004) Nat Biotechnology 23:69-74], 본원에 참조로 포함됨). 이어서, HbF의 발현을 항-인간 HbF 항체에 의해 웨스턴 블롯팅을 이용하여 측정하거나, 또는 고성능 액체 크로마토그래피 (HPLC)를 이용하여 정량화한다. BCL11A 인핸서 유전자좌의 성공적인 방해가 가이드 없이 RGN에 의해서만 전기천공된 HSC와 비교할 때 HbF 생성의 증가를 유도할 것으로 예상된다.

겸상 적혈구 형성을 메타비슬파이트의 첨가에 의해 이들 분화된 적혈구에서 유도한다. 현미경을 이용하여 겸상

대 정상 적혈구의 개수를 카운팅한다. 처리되지 않거나 또는 RGN 단독으로 처리된 세포에 비해 APG09748, APG09106.1, 또는 APG06646.1 플러스 sgRNA로 처리된 세포에서 겸상 적혈구의 개수가 더 적을 것으로 예상된다.

[0536] 실시예 9.4: 뮌헨 모델에서 질환 치료 검증

[0537] BCL11A 유전자좌의 APG09748, APG09106.1, 또는 APG06646.1 방해를 이용하는 효능을 평가하기 위해, 겸상 적혈구 빈혈의 적합한 인간화 마우스 모델을 이용한다. 바람직한 RGN 및 바람직한 sgRNA를 코딩하는 발현 카세트를 AAV 벡터 또는 아데노바이러스 벡터에 패키징한다. 특히, 아데노바이러스 유형 Ad5/35가 HSC를 표적화하는데 효과적이다. 겸상 적혈구 대립유전자를 갖는 인간화 HBB 유전자좌를 함유하는 적합한 마우스 모델, 예컨대 B6;FVB-Tg(LCR-HBA2,LCR-HBB*E26K)53Hhb/J 또는 B6.Cg-Hbattm1Paz Hbattm1Tow Tg(HBA-HBBs)41Paz/HhbJ를 선택한다. 이들 마우스를 과립구 콜로니-자극 인자 단독으로 또는 플렉사포어와 조합하여 처리하여, HSC를 순환으로 동원시킨다. 이어서, RGN 및 가이드 플라스미드를 갖는 AAV 또는 아데노바이러스를 정맥내로 주사하고, 마우스를 1 주일 동안 회복시킨다. 이들 마우스로부터 수득한 혈액을 메타비술과이트를 사용하여 시험관내 겸상화 검정에서 시험하고, 마우스를 사망률 및 조혈 기능을 모니터링하기 위해 종적으로 추적한다. 두 발현 카세트 모두 걸여된 바이러스 또는 RGN 발현 카세트만을 갖는 바이러스로 처리된 마우스와 비교할 때 RGN 및 가이드 RNA를 갖는 AAV 또는 아데노바이러스로의 처리가 겸상화, 사망률을 감소시키고, 조혈 기능을 개선시킬 것으로 예상된다.

[0538] 실시예 10: 포유동물 세포에서의 염기 편집 활성

[0539] 시티딘 데아미나제-RGN 융합 단백질을 생성하는 발현 카세트를 하기와 같이 구축하였다. RGN (본원에 기재된 APG06646.1, 및 미국 출원 공개 번호 2019/0367949 및 국제 출원 공개 번호 WO 2019/236566 (이들 각각은 그 전문이 본원에 참조로 포함됨)에 기재된 APG08290.1)에 대한 코딩 서열은 포유동물 발현을 위해 코돈 최적화되고, 니카제로서 기능하도록 돌연변이되었다 (각각 서열식별번호: 128 및 262). 이러한 코딩 서열을, N-말단에서 NLS (서열식별번호: 125), 그의 C-말단에서 작동가능하게 연결된 3xFLAG 태그 (서열식별번호: 126), 그의 C-말단에서 작동가능하게 연결된 시티딘 데아미나제 (서열식별번호: 129-132) (이들 모두는 국제 출원 번호 PCT/US2019/068079 (이는 그 전문이 본원에 참조로 포함됨)에 개시되어 있음), 그의 C-말단에서 작동가능하게 연결된 아미노산 링커 (서열식별번호: 133), 그의 C-말단에서 작동가능하게 연결된 RGN 니카제, 그의 C-말단에서 작동가능하게 연결된 제2 NLS (서열식별번호: 127)를 포함하는 융합 단백질을 생성하는 발현 카세트 내로 도입하였다. 이들 발현 카세트를 각각 포유동물 세포에서 융합 단백질의 발현을 유도할 수 있는 pTwist CMV 벡터 (트위스트 바이오사이언스) 내로 도입하였다. 또한, 포유동물 세포에서 인간 U6 프로모터의 조절하에 가이드 RNA를 발현하는 별개의 벡터를 생산하였다. 이들 가이드 RNA (nAPG06646.1의 경우 서열식별번호: 134-136 및 nAPG08290.1의 경우 서열식별번호: 263-265)는 데아미나제-nRGN 융합 단백질 또는 RGN 그 자체를 각각 염기 편집 또는 유전자 편집을 위한 표적화된 게놈 서열로 가이드할 수 있다.

[0540] 500 ng의 시티딘 데아미나제-RGN 발현 플라스미드 또는 표준 RGN 발현 플라스미드, 및 500 ng의 가이드 RNA 발현 플라스미드를 리포펙타민 2000 시약 (라이프 테크놀로지스)을 사용하여 24-웰 플레이트에서 75-90% 전면생장으로 HEK293FT 세포 내로 공동-형질감염시켰다. 이어서, 세포를 37°C에서 72 시간 동안 인큐베이션하였다. 이어서, 게놈 DNA를 제조업체의 프로토콜에 따라 뉴클레오스핀 96 티슈(NucleoSpin 96 Tissue) (마슈레-나겔 (Macherey-Nagel))을 사용하여 추출하였다. 가이드RNA 표적 부위에 플랭킹된 게놈 영역을 PCR 증폭시키고, 생성물을 제조업체의 프로토콜에 따라 ZR-96 DNA 클린 앤드 콘센트레이터(ZR-96 DNA Clean and Concentrator) (자이모 리서치(Zymo Research))를 사용하여 정제하였다. 이어서, 정제된 PCR 생성물을 차세대 서열분석을 위해 일루미나 MiSeq (2x250) 상에 보냈다. 결과를 indel 형성 또는 특이적 시토신 돌연변이에 대해 분석하였다.

[0541] 하기 표 12는 각각의 구축물 및 가이드 조합의 시티딘 염기 편집 및 indel 형성을 나타낸다. 흥미롭게도, 동일한 가이드 RNA를 사용하여 RGN APG06646.1의 활성을 시티딘 데아미나제-nAPG06646.1의 활성과 비교할 때, 유전자 편집보다 최대 20x 더 높은 시티딘 염기 편집이 관찰되었다. 이들 결과는 특정 표적 부위에서 낮은 뉴클레아제 활성을 갖는 RGN이 여전히 그 부위에서 효율적인 염기 편집체일 수 있음을 나타낸다. 또한, 염기 편집 적용에서의 indel 형성은 종종 원치 않는 결과이기 때문에, 부위에서 낮은 뉴클레아제 활성을 갖는 RGN이 염기 편집 적용에 바람직할 수 있다.

[0542] 표 12: RGN 및 시티딘 데아미나제-nRGN의 시티딘 염기 편집 및 유전자 편집

RGN 또는 데아미나제-RGN	가이드 RNA	시티딘 염기 편집을 갖는 % 총 판독	INDEL 형성을 갖는 % 총 판독
ARM05-nAPG06646.1	SGN000775	8.02	0.82
ARM06CTD-nAPG06646.1	SGN000775	0	0
ARM08-nAPG06646.1	SGN000775	5.15	0.46
ARM11-nAPG06646.1	SGN000775	0.66	0
APG06646.1	SGN000775	0	0.47
ARM05-nAPG06646.1	SGN000777	2.5	0.08
ARM06CTD-nAPG06646.1	SGN000777	0.53	0
ARM08-nAPG06646.1	SGN000777	8.12	0.25
ARM11-nAPG06646.1	SGN000777	3.52	1.04
APG06646.1	SGN000777	0	1.44
ARM05-nAPG06646.1	SGN000781	13.51	0.27
ARM06CTD-nAPG06646.1	SGN000781	2.69	0.13
ARM08-nAPG06646.1	SGN000781	12.97	0.12
ARM11-nAPG06646.1	SGN000781	4.7	0.14
APG06646.1	SGN000781	0	0.65
ARM05-nAPG08290.1	SGN000143	1.35	7.04
ARM06CTD-nAPG08290.1	SGN000143	1.36	3.17
ARM08-nAPG08290.1	SGN000143	2.59	10.98
ARM11-nAPG08290.1	SGN000143	4.78	5.77
APG08290.1	SGN000143	0	6.21
ARM05-nAPG08290.1	SGN000169	4.26	12.87
ARM06CTD-nAPG08290.1	SGN000169	1.63	8.29
ARM08-nAPG08290.1	SGN000169	6.99	14.27
ARM11-nAPG08290.1	SGN000169	4.99	6.39
APG08290.1	SGN000169	0	11.85
ARM05-nAPG08290.1	SGN000173	3.26	5.51
ARM06CTD-nAPG08290.1	SGN000173	0.72	0.96
ARM08-nAPG08290.1	SGN000173	6.6	6.06
ARM11-nAPG08290.1	SGN000173	2.83	4.82
APG08290.1	SGN000173	0	2.58

[0543]

[0544] 실시예 11: 트랜스 ssDNA 절단

[0545] 11.1 트랜스 DNA 절단에 대한 검정 조건의 결정

[0546] 정제된 APG09748 (서열식별번호: 137로 제시되고, 국제 출원 번호 PCT/US2019/068079 (그 전문이 참조로 포함됨)에 기재됨)을 컷스마트 완충제 (뉴 잉글랜드 바이오랩스 B7204S) 중 단일 가이드 RNA (sgRNA)와 함께 10 분 동안 50 nM 뉴클레아제 및 100 nM sgRNA 또는 200 nM 뉴클레아제 및 400 nM sgRNA의 최종 농도로 인큐베이션하였다. 이들 RNP 용액을 1.5X 컷스마트 완충제 (뉴 잉글랜드 바이오랩스 B7204S) 중 10 nM의 최종 농도로 ssDNA - 표적 또는 미스매칭된 음성 대조군 ssDNA - 및 250 nM의 최종 농도로 리포터 프로브의 용액에 첨가하였다. 리포터 프로브 (TB0125 및 TB0089, 각각 서열식별번호: 138 및 139로 제시됨)는 5' 말단에 형광 염료 (TB0125의 경우 56-FAM 및 TB0089의 경우 Cy5), 3' 말단에 켄처 (TB0125의 경우 3IABkFQ 및 TB0089의 경우 3IABRQSp), 및 임의로 내부 켄처 (내부 켄처 ZEN은 TB0125 상에만 존재함)를 함유한다. 리포터 프로브의 절단은 형광 염료의 탈켄칭 및 이에 따른 형광 신호의 증가를 유발한다. 형광 강도를 모니터링하기 위해, 10 μ l의 각각의 반응물을 30°C에서 마이크로플레이트 판독기 (클라리오스타 플러스(CLARIOstar Plus)) 내 코닝 저부피 384-웰 마이크로플레이트에서 인큐베이션하였다.

[0547] 본 검정에 적합한 파라미터를 결정하기 위해 다수의 조건을 조사하였다. 켄칭된 프로브 설계 또는 형광단 특징의 효과가 있는지를 측정하기 위해, 2종의 이러한 리포터를 각 반응에 혼합물로서 포함시켰다. 이들은 임의의 주어진 반응에서 서로 동일한 농도였다. 모든 경우, 대조군 또는 표적 ssDNA 농도 (LE201 또는 LE205, 각각 서열식별번호: 140 및 141로 제시됨)는 10 nM이었다. RNP 명칭은 하기 표 13에 나타난 바와 같이 뉴클레아제 및 표적을 의미한다.

[0548] 표 13. 리보핵단백질 복합체

RNP 명칭	뉴클레아제	sgRNA	의도된 표적
APG09748.1	APG09748	27sg.1 (서열식별번호: 144)	LE201 (서열식별번호: 140)
APG09748.2	APG09748	27sg.2 (서열식별번호: 145)	LE205 (서열식별번호: 141)

[0549]

[0550] 결과를 하기 표 14에 나타낸다.

[0551] 표 14. 트랜스 DNA 절단 검정의 결과

RNP 명칭	[RNP] (nM)	[리포터] (nM)	표적	기울기 (RFU/분)	
				Cy5 채널	FAM 채널
APG09748.1	25 nM	0	LE201	-0.30	-2.40
APG09748.1	25 nM	50	LE201	420.00	46.20
APG09748.1	25 nM	250	LE201	1040.40	47.40
APG09748.1	25 nM	500	LE201	1090.20	57.00
APG09748.1	25 nM	750	LE201	996.60	36.60
APG09748.1	25 nM	1000	LE201	770.40	36.00
APG09748.1	25 nM	0	LE205	0.60	-1.80
APG09748.1	25 nM	50	LE205	6.00	-1.20
APG09748.1	25 nM	250	LE205	16.80	-3.00
APG09748.1	25 nM	500	LE205	33.00	1.80
APG09748.1	25 nM	750	LE205	18.00	3.00
APG09748.1	25 nM	1000	LE205	-21.60	-51.60
APG09748.2	25 nM	0	LE201	-0.18	-1.80
APG09748.2	25 nM	50	LE201	4.20	1.20
APG09748.2	25 nM	250	LE201	42.60	3.00
APG09748.2	25 nM	500	LE201	70.80	10.20
APG09748.2	25 nM	750	LE201	67.80	10.80
APG09748.2	25 nM	1000	LE201	55.80	7.80
APG09748.2	25 nM	0	LE205	0.06	-2.40
APG09748.2	25 nM	50	LE205	321.00	64.20
APG09748.2	25 nM	250	LE205	669.60	63.00
APG09748.2	25 nM	500	LE205	706.80	52.20
APG09748.2	25 nM	750	LE205	627.00	55.20
APG09748.2	25 nM	1000	LE205	534.60	46.80
APG09748.1	100 nM	0	LE201	-0.42	-1.80
APG09748.1	100 nM	50	LE201	1063.80	155.40
APG09748.1	100 nM	250	LE201	2386.20	144.00
APG09748.1	100 nM	500	LE201	3058.80	108.60
APG09748.1	100 nM	750	LE201	2974.20	99.00
APG09748.1	100 nM	1000	LE201	3282.00	113.40
APG09748.1	100 nM	0	LE205	1.20	-0.36
APG09748.1	100 nM	50	LE205	6.60	0.18
APG09748.1	100 nM	250	LE205	27.00	1.80
APG09748.1	100 nM	500	LE205	52.80	12.60
APG09748.1	100 nM	750	LE205	57.60	9.00
APG09748.1	100 nM	1000	LE205	73.80	16.20
APG09748.2	100 nM	0	LE201	0.06	-0.48
APG09748.2	100 nM	50	LE201	25.80	9.00
APG09748.2	100 nM	250	LE201	198.00	15.00

[0552]

RNP 명칭	[RNP] (nM)	[리포터] (nM)	표적	기울기 (RFU/분)	
				Cy5 채널	FAM 채널
APG09748.2	100 nM	500	LE201	265.80	16.20
APG09748.2	100 nM	750	LE201	277.20	19.20
APG09748.2	100 nM	1000	LE201	258.00	7.80
APG09748.2	100 nM	0	LE205	0.12	0.00
APG09748.2	100 nM	50	LE205	844.80	257.40
APG09748.2	100 nM	250	LE205	2091.60	270.60
APG09748.2	100 nM	500	LE205	2551.80	196.80
APG09748.2	100 nM	750	LE205	2490.60	186.00
APG09748.2	100 nM	1000	LE205	2511.00	138.00

[0553]

[0554]

이 실험으로부터, 100 nM 농도의 RNP가 일반적으로 25 nM RNP 농도보다 더 높은 리포터 프로브의 절단율을 유발하는 것으로 결론지었다. 일반적으로, 리포터 절단율은 250 내지 500 nM 리포터 농도까지는 리포터 올리고뉴클레오타이드의 농도가 높을수록 더 높으며, 리포터 농도의 추가의 증가로부터는 이익이 거의 관찰되지 않는다. 특히, TB0089 리포터 (Cy5 채널에서 검출됨)의 경우, 특히 250 nM 초과 리포터 농도에서 표적 분해를 방해하는 실질적으로 더 높은 수준의 배경 활성이 존재한다. 따라서, 250 nM 초과 리포터 농도는 유익하지 않을 것으로 결론지었다. 두 프로브는 RNP에 의한 표적 인식의 관찰에 적합한 것으로 보인다.

[0555]

11.2 APG09748 트랜스 DNA 절단 및 비-특이적 활성화에 대한 정제의 효과

[0556]

정제된 APG09748을 1X 컷스마트 완충제 (뉴 잉글랜드 바이오랩스 B7204S) 중 단일 가이드 RNA (sgRNA)와 함께 37°C에서 10 분 동안 200 nM 뉴클레아제 및 400 nM sgRNA의 최종 농도로 인큐베이션하였다. 이어서, 이들 RNP 용액을 1.5X 컷스마트 완충제 (뉴 잉글랜드 바이오랩스 B7204S) 중 10 nM의 최종 농도의 ssDNA - 표적 또는 미스매칭된 음성 대조군 ssDNA - 및 250 nM의 최종 농도의 리포터 프로브 (TP0003)의 용액에 첨가하였다. 리포터 프로브는 5' 말단에 형광 염료 및 3' 말단에 켄처를 함유한다. 리포터 프로브의 절단은 형광 염료의 탈켄칭 및 이에 따른 형광 신호의 증가를 유발한다. 형광 강도를 모니터링하기 위해, 10 μ l의 각각의 반응물을 37°C에서 마이크로플레이트 판독기 (클라리오스타 플러스) 내 코닝 저부피 384-웰 마이크로플레이트에서 인큐베이션하였다.

[0557]

표적 서열과의 인큐베이션은 음성 대조군에 비해 시간의 함수로서 형광 강도의 실질적인 증가를 유발하였다. 절단율은 표 15에 나타난 바와 같이 형광 대 시간 함수의 선형 부분의 기울기로서 요약된다.

[0558]

표 15. 트랜스 DNA 절단 검정의 결과

ssDNA 기질 서열	기울기 (RFU/분)
미스매치 (LE111; 서열식별번호: 142)	549
매치 (LE113; 서열식별번호: 143)	7724

[0559]

이들 데이터는 RNP에 의한 음성 대조군과 표적 서열의 차별화를 보여주며, 이는 절단된 리포터 프로브를 사용하여 명백하게 검출가능하다.

[0561]

11.3 병렬 플라스미드 DNA 라이브러리를 사용한 PAM 결정

[0562]

5 nt 축중성 PAM 서열 (NNNNN)이 선행하는 표적 서열을 함유하는 올리고뉴클레오타이드 (각각 서열식별번호: 266 및 267로 제시된 LE00680 및 LE00688)를 어닐링하고, NEBuilder HiFi DNA 어셈블리 마스터 믹스 (뉴 잉글랜드 바이오랩스)를 사용하여 이중 소화된 pUC19 플라스미드 내로 클로닝하였다. 상기 반응물의 형질전환으로부터 생성된 각각의 콜로니는 클론 플라스미드 DNA 서열에 상응하여, 단일 콜로니로부터 유래된 배양물로부터의 플라스미드 DNA의 제제가 원래의 라이브러리로부터 샘플링된 고유의 플라스미드 제제가 되게 한다. 96개 콜로니의 샘플링으로부터 플라스미드 제제를 수득하였다. 이들 제제를 개별적으로 생어(Sanger) 서열분석하여 그의 PAM 서열을 확인하였다.

[0563]

정제된 APG09748을 1X 컷스마트 완충제 (뉴 잉글랜드 바이오랩스 B7204S) 중 sgRNA (27sg.2)와 함께 실온에서 20 분 동안 200 nM 뉴클레아제 및 400 nM sgRNA의 최종 농도로 인큐베이션하였다.

[0564]

이들 RNP 용액을 1.5X 컷스마트 완충제 (뉴 잉글랜드 바이오랩스 B7204S) 중 8.3 nM의 최종 농도의 플라스미드 DNA 표적, 및 각각 250 nM 및 50 nM의 TB0125 및 TB0089 리포터 (각각 서열식별번호: 138 및 139로 제시됨)의 용액에 100 nM의 최종 농도로 첨가하였다. 형광 강도를 모니터링하기 위해, 10 μ l의 각각의 반응물을 37°C에

서 마이크로플레이트 판독기 (클라리오스타 플러스) 내 코닝 저부피 384-웰 마이크로플레이트에서 인큐베이션하였다.

[0565] 소수의 샘플에서 플라스미드 농도가 농도 정규화의 목표보다 낮거나, 또는 용액의 부피가 의도된 목표량을 반응 웰에 전달하기에 불충분하였다. 이들 샘플은 분석으로부터 배제되지 않았고, 따라서 이중실험 간 불일치를 통해 오류에 기여할 것으로 예상된다. 생어 서열분석에 의한 평가시, 표적 서열에서 교차 오염 (PAM 영역에서 명백한 다중 트레이스) 또는 변경을 갖는 것으로 결정된 플라스미드를 사용한 샘플을 분석으로부터 제거하고, 그 결과는 하기에 나타나지 않는다. 분석 결과를 FAM 채널에서의 기울기에 의해 내림차순으로 하기 표 16에 나타낸다.

[0566] 표 16. PAM 서열

PAM	서열식별번호	기울기 (RFU/분)	
		Cy5 채널	FAM 채널
ATATG	147	111.52	1728.13
ATATG	147	102.42	1612.86
TATTG	148	137.15	846.12
TATTG	148	148.61	799.82
CGTTC	149	151.06	744.18
CGTTC	149	134.61	735.35
TTTTT	150	167.71	731.11
TTTTT	150	149.83	711.78
CAATC	151	120.84	675.71
AATCT	152	134.67	622.52
AATCT	152	134.1	620.2
CAATC	151	96.94	608.58
AAATT	153	152.6	607.24
AAATT	153	150.57	591.77
CATCC	154	97.34	588.7
AGATA	155	214.69	579.56
TATTC	156	124.22	570.81
ATTTT	157	134.72	551.91
AGATA	155	198.45	544.81
TATTC	156	108.57	538.18
TCTTT	158	141.66	532.87
ATTTT	157	107.9	526.75
AGATC	159	106.93	497.71
AATAA	160	78.94	484.67
TCTTT	158	117.19	484.13
GATCC	161	112.85	479.91
GATCC	161	105.04	474.94
AGATC	159	106.5	473.98
ATTCA	162	115.03	473.24
AATAA	160	77.43	465.75
AAAAT	163	78.3	458.23
TTATC	164	102.14	456.73
TATCA	165	106.45	448.78
TCACT	166	167.02	429.36
CTTAT	167	82.9	418.38
TATTG	148	116.08	417.94
TCACT	166	160.29	415.64
AAAAT	163	56.43	411.26
TGATA	168	89.02	406.14

[0567]

PAM	서열식별번호	기울기 (RFU/분)	
		Cy5 채널	FAM 채널
TGATA	168	85.99	395.57
TACTA	169	90.61	391.06
ATTTT	157	113.18	382.59
TACTA	169	74.33	364.56
ATTTT	157	106.72	364.02
TATTG	148	93.95	341.36
CTTAT	167	76.12	336.8
ATATA	170	83.39	326.86
ATATA	170	72.98	318.2
TATCA	165	65.41	317.53
GACTC	171	56.56	302.2
GCTCA	172	22.1	293.68
TAACC	173	56.08	283.99
CGTGT	174	66.57	268.89
TAACC	173	56.55	265.45
TTTCG	175	59.55	262.91
CGTGT	174	50.71	262.37
TTTCG	175	61.42	261.65
CAGTA	176	51.81	240.09
GCTCA	172	19.08	228.08
CAGTA	176	51.09	220.25
CCCTT	177	17.83	136.69
CCTAT	178	40.4	125.81
TCCCT	179	39.93	124.6
AGGTT	180	40.96	117.18
AGGTT	180	40.81	114.79
TAACG	181	32.53	94.68
AACAG	182	8.19	91.72
CCTAT	178	34.51	87.57
TGAAC	183	10.22	86.38
TAAGT	184	19.35	85.12
TCCCT	179	26.67	83.02
TAACG	181	27.62	79.17
TAACG	181	28.89	78.55
TAACG	181	22.23	78.36
CTGTA	185	23.27	77.06
AGGGG	186	30.8	74.4
GCTGT	187	28.3	71.54
ATCTA	188	18.73	69.15
GCTGT	187	25.61	69.06
TTATC	164	22.8	68.13

[0568]

PAM	서열식별번호	기울기 (RFU/분)	
		Cy5 채널	FAM 채널
AACAG	182	23.98	67.27
TCAGG	189	17.25	67.04
ATGTC	190	13.91	65.93
TGCTA	191	26.33	65.73
TAAAA	192	17.32	65.43
ATCTA	188	16.88	65.3
CTGTA	185	21.81	64.15
GACTC	171	24.05	64.02
ATTCA	162	25.04	63.87
TGAAC	183	24.53	63.64
CCCTT	177	21.35	61.95
ATGTC	190	12.99	60.6
ATCAG	193	22.32	59.63
ATCAG	193	18.55	56.58
CGGGG	194	27.86	56.28
TAAAA	192	17.72	56.03
TTCTC	195	25.7	55.83
GTCAT	196	23.07	55.76
GCGTG	197	17.36	55.12
GCGTG	197	22.02	55.02
GCAAT	198	15.47	54.91
TTCTC	199	21.89	54.89
GTCAT	196	17.22	54.34
GCAAT	198	17.36	53.32
AGCAT	200	22.47	52.23
GTCAA	201	14.14	52.12
AGAGC	202	19.62	51.79
AGAGC	202	15.99	50.37
GTGGG	203	22.38	50.1
CTGAA	204	21.24	49.45
CTGAT	205	12.05	49.32
ATGAA	206	17.93	49.15
ACCTT	207	19.66	49.12
ACCTT	207	22.74	49.12
GTCAA	201	13	48.21
AACCA	208	17.81	48.04
AACCA	208	15.75	47.87
GTAAA	209	16.3	47.85
TAGGA	210	18.68	47.73
CCGAA	211	20.24	47.58
AGCAT	200	22.95	47.52

[0569]

PAM	서열식별번호	기울기 (RFU/분)	
		Cy5 채널	FAM 채널
CGGGG	194	20.94	47.41
GTCAT	196	15.17	47.24
CTGAA	204	20.4	46.36
GTGGG	203	19.98	46.25
TACGA	212	16.79	46.18
TGGCT	213	14.81	45.75
GTCAT	196	15.55	45.51
ATGAA	206	19.11	45.09
TAGGA	210	21.65	44.94
TTCTC	199	21.44	44.76
CCAAG	214	16.98	44.61
GTAAA	209	15.35	44.59
TGCTT	215	15.79	44.54
CTAAA	216	12.72	44.28
TGCTA	191	16	44.14
TACGA	212	16.07	44.06
TCGAT	217	22.28	43.96
TTCTC	199	15.94	42.55
CTGAT	205	16.13	42.27
TCAGG	189	11.38	41.9
CCAAG	214	11.28	41.45
AGGGG	186	16.04	41.44
TCGAT	217	18.68	40.5
CCGAA	211	15.39	39.09
CTAAA	216	10.72	30.37
GTAAC	218	4.36	9.51
TGCTT	215	-0.32	-0.7
TAAGT	184	-0.57	-1.3
CATCC	154	0.47	-1.95
TGGCT	213	-0.03	-2.2
GTAAC	219	-0.46	-2.3

[0570]

[0571]

가장 높은 기울기를 갖는 서열은 국제 출원 번호 PCT/US2019/068079 (그 전문이 참조로 포함됨)에 이전에 기재된 플라스미드 고갈 검정에 의해 결정된 예측된 PAM (서열식별번호: 60으로 제시된 DTTN)을 따르는 것으로 보인다. 특히, 본 발명자들은 표적의 2개의 뉴클레오타이드 5' 위치에서 "T"에 대한 강한 선호를 인식한다. 놀랍게도, 상기 검정에서 관찰된 가장 활성인 PAM 부위 (ATATG, 서열식별번호: 147)는 DTTN (서열식별번호: 60)의 컨센서스와 정확하게 매칭되지는 않았고, 이는 상기 인식 부위에서 일부 수준의 유연성을 시사한다.

[0572]

11.4. PCR 증폭된 DNA의 존재 하에 트랜스 ssDNA 절단 활성

[0573]

PCR 증폭된 표적을 적절한 프라이머를 사용하여 게놈 DNA (각각 서열식별번호: 225 및 226으로 제시된 TRAC 및 VEGF 앰플리콘)로부터 생성하였다. VEGF 표적 (서열식별번호: 227로 제시됨)을 서열 LE573 및 LE578 (각각 서열식별번호: 228 및 229로 제시됨)을 갖는 프라이머에 의해 HEK293T 세포 게놈 DNA로부터 PCR 증폭시켰다. TRAC 표적 (서열식별번호: 230으로 제시됨)을 LE257 및 LE258 (각각 서열식별번호: 231 및 232로 제시됨)에 의해 PCR 증폭시켰다.

[0574]

1X NE버퍼(NEBuffer) 2 (뉴 잉글랜드 바이오랩스) 중 각각 0.5 μ M 및 1 μ M의 본원에 기재된 APG09106.1 뉴클레아제 및 sgRNA의 인큐베이션에 의해 RNP를 형성하고, 실온에서 20 분 동안 인큐베이션하였다.

[0575]

표 17. 리보핵단백질 복합체

RNP	뉴클레아제	가이드 RNA	의도된 표적
APG09106.2	APG09106.1	27sg.2 (서열식별번호: 145)	LET126 앰플리콘 무작위화 PAM 앰플리콘
APG09106.836	APG09106.1	27sg.836 (서열식별번호: 233)	VEGF 앰플리콘
APG09106.838	APG09106.1	27sg.838 (서열식별번호: 234)	TRAC 앰플리콘

[0576]

[0577]

5' TEX615 표지 및 3' 아이오와 블랙 FQ 켄처를 갖는 1.5 μ M ssDNA 올리고뉴클레오타이드 리포터, 및 100 nM의

각각의 PCR 생성물과 함께 1.5X NE버퍼 2 중에서 절단 반응을 수행하였다. 리포터 프로브의 절단은 형광 염료의 탈켄칭 및 이에 따른 형광 신호의 증가를 유발한다. 형광 강도를 모니터링하기 위해, 10 μ l의 각각의 반응물을 37°C에서 마이크로플레이트 판독기 (클라리오스타 플러스) 내 코닝 저부피 384-웰 마이크로플레이트에서 인큐베이션하였다. 동역학적 분석의 결과를 표 18에 나타낸다.

표 18. 트랜스 DNA 절단 검정의 결과

RNP	표적 농도 (nM)	표적	기울기 (RFU/분)
28.836	30	TRAC 앰플리콘	1204
28.838	30	TRAC 앰플리콘	8436
28.838	30	VEGF 앰플리콘	766
28.836	30	VEGF 앰플리콘	3484

이들 결과는 다양한 공급원으로부터의 PCR 증폭된 DNA의 존재 하에 트랜스 ssDNA 절단 활성의 특이적 활성화를 나타낸다. 활성은 PCR 증폭된 기질의 농도에 의존한다.

11.5 진단제로서의 ssDNA 절단의 용도

표적 DNA 서열의 존재 하에 광학적으로 검출가능한 신호를 생성하는 이들 뉴클레아제의 능력으로 인해, 이들은 유전 질환 또는 감염성 질환의 요인, 예컨대 박테리아, 바이러스 또는 진균 검출용 진단 장치로의 구현을 위한 유용성이 기대된다.

진단 절차는 시험될 샘플로부터 핵산의 단리 또는 증폭을 포함할 수 있다. 또한, 증폭 (예컨대, PCR) 없이 검출가능하기에 충분히 높은 양으로, 또는 검출 또는 신호 생성을 방해하는 물질 없이 샘플에 존재할 수 있는 경우에 핵산의 임의의 단리 또는 정제를 수행하지 않으면서 일부 샘플을 사용하는 것이 적합할 수 있다.

다른 실시예에 기재된 바와 같이 형성된 RNP는 이어서 리포터, 예컨대 이전 실시예에서 사용된 형광단 및 켄처 변형된 ssDNA 올리고뉴클레오타이드, 또는 절단될 때 가시적이거나 또는 달리 용이하게 검출가능한 신호를 생성하는 일부 다른 종류의 ssDNA 기질과 함께 샘플 (또는 이전 단락에 기재된 바와 같은 가공된 샘플)에 노출될 수 있다. 형광단-켄처 접합된 DNA 올리고뉴클레오타이드 (이전에 기재된 실시예에서와 같음)를 사용하는 경우, 이들은 이전 실시예에 기재된 바와 같은 형광계를 사용하여 검출될 수 있다. 검출을 단순화하기 위해, 상기 기재된 동역학적 검정 대신 종점 검정을 수행할 수 있으며, 이는 검정이 양성 및 음성 대조군에 비해 고정된 시간 동안 수행되고 이 경과 시간의 종료 시에 판독될 수 있음을 의미한다.

이들 시약은 또한 매우 적은 기기로 주어진 병원체 또는 특이적 핵산 서열 (예컨대, 개체에서의 이환 대립유전자)의 검출을 가능하게 하는 측방 유동 시험 장치에 통합될 수 있다. 이 검정에서, ssDNA 리포터는 항체 또는 친화성 시약 포획에 적합한 다중 분자, 예컨대 플루오레세인, 비오틴 및/또는 디곡시게닌에 접합될 것이다.

서열 목록

SEQUENCE LISTING

<110> LIFEEDIT, INC.

<120> RNA-GUIDED NUCLEASES AND ACTIVE FRAGMENTS AND VARIANTS THEREOF
AND METHODS OF USE

<130> L103438 1170WO (0049.2)

<140><141><150> 63/030,088

<151> 2020-05-26

<150> 62/901,875

<151> 2019-09-18

<150> 62/885,483

<151> 2019-08-12

<160> 742

<170> PatentIn version 3.5

<210> 1

<211> 1071

<212> PRT

<213> Bacillus sp.

<400> 1

Met Arg Glu Leu Asp Tyr Arg Ile Gly Leu Asp Ile Gly Thr Asn Ser

1 5 10 15

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

20 25 30

Tyr Glu Lys Thr Arg Ile Val Asp Lys Gly Val Arg Met Phe Asp Lys

35 40 45

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

50 55 60

Ala Arg Ser Ser Arg Arg Arg Leu Asn Arg Lys Ser Gln Arg Lys Lys

65 70 75 80

Asp Ile Arg Asn Leu Leu Val Gln His Glu Ile Ile Ser Gln Lys Glu

85 90 95

Leu Thr Ser Leu Tyr Pro Leu Ser Lys Ser Ser Met Asp Ile Trp Asp

100 105 110

Ile Arg Leu Asp Gly Leu Asp Arg Leu Leu Asp Arg Phe Glu Trp Ala

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

Trp Met Lys Asn Ala Asp Cys Ser Lys Tyr Gly Lys Arg Arg Asn Ser

180 185 190

Pro Asn Glu Tyr Val Phe Ser Val Ser Arg Ala Asp Leu Glu Lys Glu

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

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

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

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

Glu Gly Arg Thr Tyr Gln Glu Ala Ile His Gln Leu Gly Tyr Asp Thr
450 455 460

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

Asp Glu Ile Thr Asn Pro Ile Val Lys Arg Ala Leu Thr Gln Ala Arg
485 490 495

Lys Val Val Asn Ala Ile Ile Arg Arg Tyr Gly Ser Pro Asn Ser Val
500 505 510

His Ile Glu Leu Ala Arg Glu Leu Ser Lys Ser His Asp Glu Arg Lys
515 520 525

Lys Ile Met Thr Ala His Asp Glu Asn Tyr Lys Lys Asn Lys Gly Ala
530 535 540

Ile Ser Ile Leu Ile Glu Asn Gly Ile Leu Asn Pro Thr Gly Tyr Asp
545 550 555 560

Ile Val Arg Tyr Lys Leu Trp Lys Glu Gln Gly Glu Arg Cys Ala Tyr
565 570 575

Ser Leu Lys Lys Ile Pro Ala Asn Thr Phe Phe Asn Glu Leu Lys Lys
580 585 590

Glu Arg Ser Gly Pro Pro Val Leu Glu Val Asp His Ile Leu Pro Tyr
595 600 605

Ser Gln Ser Phe Ile Asp Ser Tyr His Asn Lys Val Leu Val Tyr Gly
610 615 620

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

Glu Glu Asp Lys Glu Trp Glu Ser Phe Glu Ser Tyr Val Arg Ser Asn
645 650 655

Ser Phe Phe Ser Lys Lys Lys Arg Gly Tyr Leu Leu Lys Lys Ala Tyr
660 665 670

Leu Pro Arg Glu Ser Asn Leu Ile Lys Glu Arg His Leu Asn Asp Thr
675 680 685

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

690 695 700
 Phe Lys Glu Ala Val Gly Ile Thr Arg Lys Lys Tyr Val Gln Thr Val
 705 710 715 720

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

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

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

 Asn Leu Ile Lys Lys Ile Lys Val Glu Val Gln Thr Lys Ser Phe Val
 915 920 925
 Arg Glu Ile Asn Gly Gly Val Ala Gln Asn Gly Asp Leu Val Arg Val
 930 935 940

Asp Leu Phe Glu Lys Asp Asn Arg Tyr Tyr Met Val Pro Ile Tyr Val
 945 950 955 960
 Val Asp Thr Val Arg Ser Glu Leu Pro Asn Lys Ala Val Thr Ser Ser
 965 970 975

Lys Gly Tyr Glu Gln Trp Leu Ser Ile Asp Asn Ser Phe Thr Phe Lys
 980 985 990
 Phe Ser Leu Tyr Pro Tyr Asp Leu Val Arg Leu Val Lys Gly Asp Glu
 995 1000 1005
 Asp Arg Phe Leu Tyr Phe Ser Thr Leu Asp Ile Asn Ser Asp Arg
 1010 1015 1020
 Leu Asn Phe Lys Asp Val Asn Lys Pro Ser Lys Gln Ala Glu Tyr
 1025 1030 1035

Arg Tyr Ser Leu Lys Thr Ile Glu Asn Leu Glu Lys Tyr Glu Ile
 1040 1045 1050
 Gly Val Leu Gly Asp Leu Arg Leu Val Arg Gln Glu Thr Arg Lys
 1055 1060 1065
 Ile Phe Lys
 1070

<210> 2

<211> 34

<212> RNA

<213> Bacillus sp.

<400> 2

gucauaguuc cauuaaagcc augcuguuu uaug 34

<210> 3

<211> 148

<212> RNA

<213> Bacillus sp.

<400> 3

caugauacag caauggcuuu gauguuucua ugauaagggc uucggcccgu ggcguugggg 60

aucgccugcc cauuuuaaug ggcucuccc caucuauuuu augagaaau uacaaccuug 120

gcuaucuua aauagcuaag guuuuuuu 148

<210> 4

<211> 215

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 4

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng ucaguagucc auuaaagcca uugcuguuuu	60
augaaagcau gauacagcaa ugguuuugau guuucuauga uaaggguuuc ggcccugggc	120
guuggggauc gccugcccau uuuaaugggc uucuccccau cuuuuuuau agaaaauuac	180
aaccuuggcu auucuuuuu agcuaagguu uuuuuu	215

<210> 5

<211> 216

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 5

gagcggacag cagcuuccua uaucucguac gucauaguuc cauuaaagcc auugcuguuu	60
uaugaaagca ugauacagca augguuuuga uguuuuauug auaaggguu cgcccuggg	120
cguuggggau cgccugccca uuuaauggg cuucucccca ucuuuuuau gagaaaauua	180
caaccuuggc uauucuuuuu uagcuaaggu uuuuuu	216

<210> 6

<211> 216

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 6

ccaugauaua gacguugugg cuguuguagu gucauaguuc cauuaaagcc auugcuguuu 60

uaugaaagca ugauacagca auggcuuuga uguuucuaug auaagggcuu cggcccgugg 120

cguuggggau cgccugccca uuuuaauggg cuucucccca ucuauuuauu gagaaauua 180

caaccuuggc uauucuuaaa uagcuaaggu uuuuuu 216

<210> 7

<211> 6

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(4)

<223> a, c, t, g, unknown or other

<400> 7

nnnncc 6

<210> 8

<211> 6

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (2)..(2)

<223> a, c, u, g, unknown or other

<220><221> modified_base

<222> (4)..(5)

<223> a, c, u, g, unknown or other

<400> 8

unanna 6

<210> 9

<211> 1403

<212> PRT

<213> Chryseobacterium sp.

<400> 9

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

1 5 10 15

Trp Ala Leu Val Glu Glu Asp Val Glu Asn Pro Glu Asn Asn Lys Ile

20 25 30

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

35 40 45

Asn Phe Glu Lys Gly Lys Ser Ile Thr Thr Asn Ala Gly Arg Thr Ser

50 55 60

Ala Arg Ser Ala Arg Arg Asn Leu Gln Arg Phe Lys Leu Arg Arg Lys

65 70 75 80

Ser Leu Trp Glu Val Leu Ile Asn Tyr Lys Ile Ile Lys Asn Asp Thr

85 90 95

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

100 105 110

Arg Ala Lys Ala Ala Lys Glu Lys Ile Glu Leu Asp Glu Leu Ala Arg

115 120 125

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

130 135 140

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

145 150 155 160

Lys Glu Leu Tyr Glu Lys Asn Tyr Thr Pro Gly Gln Tyr Val Tyr Asp

165 170 175

Leu Leu Lys Asn Gly Lys Lys Gln Ile Pro Asp Phe Tyr Arg Ser Asp

180 185 190

Leu Gln Asn Glu Leu Asp Arg Val Trp Glu Phe Gln Lys Gln Tyr His

195 200 205

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

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

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

Asn Glu Lys Leu Tyr Asp Ala Phe Leu Lys Ile Leu Glu Leu Lys Gly
485 490 495

Tyr Asn Val Phe Glu Leu Leu Lys Ile Lys Glu Asp Arg Asp Glu Ala
500 505 510

Lys Leu Ser Glu Leu Lys Ala Ser Ala Asp Glu Ile Lys Arg Met Val
515 520 525

Lys Glu Ile Phe Asn Ala Asn Leu Ile Asn Thr Ser Ile Leu Asp Phe
530 535 540

Asn Ala Glu Leu Glu Gly Lys Asp Phe Glu Asn Gln Ala Ser Tyr Gln
545 550 555 560

Phe Trp His Leu Ile Tyr Ser Tyr Glu Gly Asp Asn Ser Pro Ser Gly
565 570 575

Asn Glu Lys Leu Tyr Glu Leu Leu Glu Asn Lys Phe Gly Phe Lys Lys
580 585 590

Glu His Ser Lys Ile Leu Ser Thr Ile Val Phe Pro Gln Glu Tyr Gly
595 600 605

Ser Leu Ser Ala Lys Ala Met Arg Arg Ile Tyr Pro Tyr Ile Lys Asp
610 615 620

Asn Lys Tyr Ser Glu Ala Cys Leu Leu Ala Gly Tyr Asn His Ser Lys
625 630 635 640

Gln Ser Leu Thr Lys Glu Glu Leu Glu Asn Arg Lys Leu Lys Glu Gln
645 650 655

Leu Glu Val Leu Pro Lys Asn Ser Leu Arg Asn Pro Val Val Glu Lys
660 665 670

Ile Leu Asn Gln Met Ile Asn Leu Val Asn Ala Ile Ile Ser Glu His
675 680 685

Gly Lys Leu Asp Gly Val Arg Ile Glu Leu Ala Arg Glu Leu Lys Lys
690 695 700

Ser Ala Gln Glu Arg Glu Asp Met Thr Lys Ser Ile Asn Glu Ala Thr

705 710 715 720
 Ile Leu His Gln Lys Tyr Ala Thr Val Leu Arg Gln Glu Phe Gly Val
 725 730 735

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

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

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

 Glu Glu Lys Glu Val Ile Pro Asp Trp Thr Lys Arg Asn Asp His Arg
 930 935 940
 His His Ala Met Asp Ala Leu Thr Val Ala Phe Thr Thr Arg Asn His
 945 950 955 960

Ile Gln Tyr Leu Asn Tyr Leu Asn Ala Arg Lys Asp Glu Lys His Lys
965 970 975

Glu His Gln Asn Ile Tyr Ala Ile Glu Asn Ile Ile Thr Glu Ile Ile
980 985 990

Glu Lys Lys Asn Gly Ser Lys Ser Arg Arg Phe Lys Thr Pro Met Asn
995 1000 1005

Arg Phe Arg His Glu Ala Lys Gln His Leu Lys Glu Val Leu Val
1010 1015 1020

Ser His Lys Thr Lys Asn Lys Val Val Thr Ser Asn Ile Asn Lys
1025 1030 1035

Thr Lys Lys Lys Arg Gly Ile His Lys Lys Ala Glu Leu Thr Pro
1040 1045 1050

Arg Gly Gln Leu His Lys Glu Thr Val Tyr Gly Ser Lys Lys Phe
1055 1060 1065

Leu Gln Ser Arg Glu Glu Lys Val Ser Ala Lys Phe Asp Tyr Glu
1070 1075 1080

Thr Ile Leu Met Val Val Asn Pro Ile His Arg Asn Ala Leu Leu
1085 1090 1095

Asn Arg Leu Lys Glu Tyr Gly Asn Asp Pro Lys Lys Ala Phe Ser
1100 1105 1110

Gly Lys Asn Ala Val Asn Lys Thr Pro Val Tyr Leu Ser Asp Ser
1115 1120 1125

Lys Ile Glu Val Leu Pro Glu Lys Val Thr Leu Ser Trp Phe Glu
1130 1135 1140

Thr Gly Tyr Thr Ile Arg Lys Ala Val Thr Pro Asp Asn Phe Lys
1145 1150 1155

Asp Tyr Lys Asn Leu Glu Lys Ile Thr Asp Met Gly Ile Lys Lys
1160 1165 1170

Ile Leu Lys Glu Arg Leu Asp Gln Phe Lys Gly Asn Ala Lys Glu
1175 1180 1185

Ala Phe Ser Asn Leu Asp Lys Glu Pro Ile Trp Leu Asn Glu Glu

1190	1195	1200
Lys Gly Ile Ala Ile Lys Thr	Val Thr Ile Thr Gly	Val Ser Asn
1205	1210	1215
Ala Glu Ser Leu His Phe Lys	Lys Asp His Leu Gly	Lys Glu Ile
1220	1225	1230
Leu Asp Glu Asn Gly Asn Lys	Ile Pro Val Asp Phe	Val Ser Thr
1235	1240	1245
Gly Asn Asn His His Leu Ala	Ile Tyr Leu Asn Glu	Glu Glu Lys
1250	1255	1260
Leu Asp Asp Lys Met Val Thr	Phe Tyr Glu Ala Val	Leu Arg Val
1265	1270	1275
Asn Gln Gly Leu Pro Val Ile	Asp Lys Asn Tyr Asn	Arg Glu Lys
1280	1285	1290
Gly Tyr Lys Phe Leu Met Thr	Leu Lys Gln Asn Glu	Met Phe Val
1295	1300	1305
Phe Pro Asn Glu Glu Phe Asp	Pro Ser Glu Ile Asp	Leu Leu Asp
1310	1315	1320
Glu Lys Asn Leu Glu Gln Ile	Ser Gln Asn Leu Phe	Arg Val Gln
1325	1330	1335
Lys Ile Ser Lys Val Gly Tyr	Gly Asn Ser Phe Ile	Arg Asp Phe
1340	1345	1350
Val Phe Arg His His Leu Glu	Thr Thr Val Glu Asp	Arg Lys Glu
1355	1360	1365
Leu Arg Asn Thr Thr Tyr Ile	Gln Leu Lys Ser Leu	Glu Gly Leu
1370	1375	1380
Arg Asn Ile Val Lys Val Arg	Leu Asn His Leu Gly	Lys Ile Val
1385	1390	1395
Gln Ile Gly Glu Tyr		
1400		
<210> 10		
<211> 24		
<212> RNA		

<213> Chryseobacterium sp.

<400> 10

guugugaaau gcuuucacaa cuua

24

<210> 11

<211> 78

<212> RNA

<213> Chryseobacterium sp.

<400> 11

auaaaauuug aaagcacuuc acaauaagga uuauuccguu gugaaaacau ucaaggcggg

60

gcaacucgcc uuuuuuuu

78

<210> 12

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 12

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugugaaaug cuuucacaaac uuaaaagaua

60

aaaaauugaaa gcacuucaca auaaggauua uuccguugug aaaacauuca aggcggggca

120

acucgccuuu uuuuuu

135

<210> 13

<211> 136

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 13

gagcggacag cagcuuccua uaucucguac guugugaaau gcuuucaaaa cuuuuagau 60

aaaauuugaa agcacuucac aaauaggauu auuccguugu gaaaacauuc aaggcggggc 120

aacucgccuu uuuuuu 136

<210> 14

<211> 136

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 14

ccaugauuaa gacguugugg cuguuguagu guugugaaau gcuuucaaaa cuuuuagau 60

aaaauuugaa agcacuucac aaauaggauu auuccguugu gaaaacauuc aaggcggggc 120

aacucgccuu uuuuuu 136

<210> 15

<211> 6

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(1)

<223> a, c, t, g, unknown or other

<400> 15

nrrttt 6

<210> 16

<211> 1404

<212> PRT

<213> Sphingobacterium sp.

<400> 16

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

1 5 10 15

 Trp Ala Tyr Val Leu Glu Ala Gln Asp Glu Asn Glu Ile Ser Gln Ile
 20 25 30
 Lys Lys Leu Gly Val Arg Val Asn Pro Leu Thr Thr Asp Glu Gln Leu
 35 40 45
 Asn Phe Glu Lys Gly Lys Pro Ile Thr Thr Asn Ala Gly Arg Thr Leu
 50 55 60
 Ala Arg Ser Ala Arg Arg Asn Leu Gln Arg Tyr Lys Leu Arg Arg Glu
 65 70 75 80

 Asn Leu Ile Glu Val Leu Lys Arg Glu Lys Trp Ile Asp Asp Lys Ser
 85 90 95
 Ile Leu Ala Glu Asn Gly Asn Lys Ser Thr Phe Glu Thr Tyr Ala Leu
 100 105 110
 Arg Ala Ala Ala Ala Ser Lys Glu Ile Lys Leu Glu Glu Leu Ala Arg
 115 120 125
 Val Leu Leu Met Ile Asn Lys Lys Arg Gly Tyr Lys Ser Ser Arg Lys
 130 135 140

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

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

Trp Arg Lys Asp Gly Leu Thr Lys Gln Leu Glu Gln Glu Glu Val Ala
260 265 270

Tyr Ile Ile Ala Glu Ile Asn Gly Gln Leu Ala Asn Ser Ser Gly Tyr
275 280 285

Leu Gly Ala Ile Ser Asp Arg Ser Lys Glu Leu Tyr Phe Asn Lys Gln
290 295 300

Thr Val Gly Gln Phe Leu Phe Asn Gln Leu Gln His Asn Thr His Ala
305 310 315 320

Arg Val Lys Asn Gln Val Phe Tyr Arg Gln Asp Tyr Leu Asp Glu Phe
325 330 335

Glu Lys Ile Trp Ser Val Gln Ser Gln Tyr His Gln Gly Leu Thr Glu
340 345 350

Gln Leu Lys Glu Glu Ile Arg Asp Ile Ile Ile Phe Tyr Gln Arg Lys
355 360 365

Leu Lys Ser Gln Lys Gly Leu Ile Ser Phe Cys Glu Phe Glu Gln His
370 375 380

Glu Val Val Ile Asn Gly Lys Asn Lys Ile Val Gly Leu Arg Val Ala
385 390 395 400

Pro Lys Ser Ser Pro Ile Phe Gln Glu Phe Lys Ile Trp Gln Gln Ile
405 410 415

Asn Asn Val Lys Leu Arg Asn Lys Ile Thr Lys Glu Val Tyr Asn Leu
420 425 430

Ala Glu Glu Gln Lys Arg Ser Leu Phe Glu Thr Leu Asn Leu Lys Gly
435 440 445

Lys Leu Ser Ala Tyr Gln Val Leu Asn Leu Ile Glu Val Lys Pro Lys
450 455 460

Glu Trp Glu Leu Asn Tyr Thr Glu Leu Glu Gly Asn Ser Thr Asn Leu
465 470 475 480

Ala Leu Tyr Asn Ala Tyr Leu Asp Ile Leu Asp Ile Glu Gly Tyr Asp
485 490 495

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

500 505 510
 Ser Asp Leu Asp Ile Pro Val Ser Glu Ile Lys Gly Met Ile Arg Arg
 515 520 525

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

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

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

 Glu Leu Thr Lys Ala Ile Asn Thr Ser Lys Ser Glu His Glu Lys Ile
 725 730 735
 Ile Lys Ile Leu Gln Thr Glu Asp Gly Ile Lys Asn Pro Thr Arg Asn
 740 745 750

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

755 760 765

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

770 775 780

Asn Val Tyr Asp Ile Asp His Ile Ile Pro Gln Ser Arg Leu Phe Asp

785 790 795 800

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

805 810 815

Lys Gly Asn Gln Thr Ala Phe Asp Tyr Val His Ala Lys Phe Gly Glu

820 825 830

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

835 840 845

Lys Glu Glu Gly Val Ser Arg Ser Lys Tyr Lys Lys Leu Leu Met Arg

850 855 860

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

865 870 875 880

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

885 890 895

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

900 905 910

Trp Gly Leu Val Ser Val Met Lys Glu Leu Asn Leu Pro Lys Phe Lys

915 920 925

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

930 935 940

Glu Val Ile Lys Asp Trp Ser Lys Arg Asn Asp His Arg His His Ala

945 950 955 960

Met Asp Ala Leu Thr Val Ala Phe Thr Lys His Ser His Ile His Tyr

965 970 975

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

980 985 990

Thr Ile Lys Ala Ile Glu Ile Lys Glu Thr Tyr Val Glu Arg Asp Asp

995	1000	1005
Leu Gly Asn Arg Lys Arg Leu Phe Lys Glu Pro Ile Pro His Phe		
1010	1015	1020
Arg Ser Ile Ala Lys Glu His Leu Glu Ser Val Leu Val Ser His		
1025	1030	1035
Lys Ala Lys Asn Lys Val Val Thr Lys Asn Arg Asn Lys Ile Asp		
1040	1045	1050
Gly Lys Lys Gln Ala Gln Glu Val Leu Thr Pro Arg Gly Gln Leu		
1055	1060	1065
His Lys Glu Ser Ile Tyr Gly Lys Ile Leu Gln Tyr Ala Ser Lys		
1070	1075	1080
Glu Glu Lys Ile Ser Ala Lys Phe Asp Ala Ala Thr Ile Ala Met		
1085	1090	1095
Val Ser Asn Pro Arg Tyr Arg Ser Ala Leu Leu Asp Arg Leu Ser		
1100	1105	1110
Ala Tyr Gly Asn Asp Pro Lys Lys Ala Phe Thr Gly Lys Asn Ser		
1115	1120	1125
Pro Asn Lys Ser Pro Ile Tyr Leu Asp Gln Glu Lys Gln Leu Ala		
1130	1135	1140
Val Pro Glu Lys Val Lys Leu Val Trp Leu Glu Glu Asp Tyr Thr		
1145	1150	1155
Ile Arg Lys Glu Ile Gly Pro Asp Leu Lys Ile Glu Lys Val Ile		
1160	1165	1170
Asp Gln Gly Val Lys Arg Ile Leu Glu Gln Arg Leu Thr Glu Tyr		
1175	1180	1185
Gly Gly Asp Ser Lys Lys Ala Phe Ser Asp Leu Asp Asn Asn Pro		
1190	1195	1200
Ile Trp Leu Asn Lys Glu Lys Gly Ile Ala Ile Lys Arg Val Thr		
1205	1210	1215
Ile Ser Gly Val Lys Asn Ala Glu Ala Leu His Ile Lys Lys Asp		
1220	1225	1230

His Leu Gly Gln Val Leu Lys Asp Lys Asp Gly Lys Pro Met Pro
1235 1240 1245

Val Asp Phe Val Ser Thr Gly Asn Asn His His Val Ala Ile Tyr
1250 1255 1260

Glu Asp Asp Asn Gly Ala Leu Gln Glu Arg Ile Val Pro Phe Tyr
1265 1270 1275

Glu Ala Val Ala Arg Val Asn Asn Gly Phe Pro Ile Ile Asp Lys
1280 1285 1290

His Tyr Asn Gln His Leu Gly Trp Lys Phe Leu Phe Ser Met Lys
1295 1300 1305

Gln Asn Glu Met Phe Leu Phe Pro Ser Glu Glu Phe Asp Pro Lys
1310 1315 1320

Gln Ile Asn Leu Leu Asp Arg Ser Asn Tyr Arg Leu Ile Ser Lys
1325 1330 1335

Asn Leu Phe Arg Val Gln Lys Phe Gly Glu Leu Ser Lys Ser Gly
1340 1345 1350

Phe Trp Phe Arg His His Leu Glu Thr Ser Val Asp Val Asn Lys
1355 1360 1365

Ser Leu Arg Ser Ile Ser Tyr Phe Asp Phe Tyr Ser Lys Asp Phe
1370 1375 1380

Met Lys Asn Ile Ala Lys Val Arg Ile Asn His Leu Gly Glu Ile
1385 1390 1395

Ile His Val Gly Glu Tyr
1400

<210> 17

<211> 18

<212> RNA

<213> Sphingobacterium sp.

<400> 17

guugugaauu gcuuucuu

<210> 18

<211> 72

<212> RNA

<213> Sphingobacterium sp.

<400> 18

augaaagcaa uucacaauaa ggauuauuuc cguuguguaa acauuuagcg ccucgucuaau 60

cuacggggca uu 72

<210> 19

<211> 123

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 19

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugugaaug cuuucuaaaa gaugaaagca 60

auucacaaua aggauuauuu ccguugugua aacauuuagc gccucgucua ucuacggggc 120

auu 123

<210> 20

<211> 124

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 20

gagcggacag cagcuuccua uaucucguac guugugaaau gcuuucuaaa agauggaaagc 60

aaucacaaua aaggauuauu uccguugugu aaacauuuag cgccucgucu aucuacgggg 120

cauu 124

<210> 21

<211> 124

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 21

ccaugauaau gacguugugg cuguuguagu guugugaaau gcuuucauaa agauggaaagc 60

aaucacaaau aaggauuaau uccguugugu aaacauuuag cgccucgucu aucuacgggg 120

cauu 124

<210> 22

<211> 7

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(2)

<223> a, c, t, g, unknown or other

<220><221> modified_base

<222> (5)..(5)

<223> a, c, t, g, unknown or other

<400> 22

nnrrngg 7

<210> 23

<211> 1453

<212> PRT

<213> Chryseobacterium sp.

<400> 23

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

1 5 10 15

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

20 25 30

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

35 40 45
 Glu Phe Gly Ser Gly Asn Ser Val Ser Gln Thr Ala Asp Arg Thr Lys

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

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

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

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

Lys Ser His Lys Thr Val Gly Thr Tyr Ile Tyr Glu Thr Ile Leu His
 290 295 300
 Asn Pro Lys Gln Lys Ile Lys Gly Lys Leu Val Arg Thr Ile Glu Arg

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

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

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

 500 505 510
 Val Glu Asn Ile Ser Glu Asn Phe Leu Thr Ser Glu Ile Glu His Lys
 515 520 525
 Ile Trp His Ile Ile Tyr Ser Val Asn Asp Lys Ile Glu Tyr Glu Lys

530 535 540
 Ala Leu Lys Ser Phe Ala Asn Lys Asn Asn Leu Asp Glu Val Ser Phe
 545 550 555 560
 Phe Glu Ala Phe Lys Lys Phe Pro Pro Phe Lys Asn Glu Tyr Gly Ser

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

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

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

 755 760 765
 Leu Ala Glu Met Met Asn Asp Ser Ser Val Glu Asn Val Arg Pro Phe
 770 775 780

Ser Pro Met Gln Gln Glu Ile Leu Lys Ile Tyr Glu Asp Gly Val Leu
 785 790 795 800
 Asn Ser Gly Ile Glu Ile Glu Asp Glu Tyr Leu Lys Ile Ser Lys Thr
 805 810 815
 Ala Gln Pro Ser Ser Ser Asp Leu Lys Arg Tyr Lys Leu Trp Leu Glu
 820 825 830
 Gln Lys Tyr Lys Ser Pro Tyr Thr Gly Gln Ile Ile Pro Leu Asn Lys
 835 840 845
 Leu Phe Thr Pro Glu Tyr Glu Ile Glu His Ile Ile Pro Lys Ser Arg
 850 855 860
 Tyr Phe Asp Asp Ser Phe Ser Asn Lys Ile Ile Cys Glu Ser Ala Val
 865 870 875 880
 Asn Lys Leu Lys Asp Asn Tyr Ile Gly Leu Gly Phe Ile Lys Gln Phe
 885 890 895
 Ala Gly Ala Ile Val Glu Leu Gly Phe Gly Lys Asn Val Lys Val Phe
 900 905 910
 Glu Ile Glu Glu Tyr Glu Asp Phe Val Lys Lys His Tyr Ala Asn Asn
 915 920 925
 Arg Gly Lys Arg Asn Lys Leu Leu Leu Glu Glu Ile Pro Glu Lys Met
 930 935 940
 Ile Glu Arg Gln Leu Asn Asp Thr Arg His Ile Ser Lys Tyr Ile Ser
 945 950 955 960
 Ser Val Leu Ser Asn Ile Val Arg Val Glu Asp Gly Thr Asp Glu Gly
 965 970 975
 Val Asn Ser Lys Asn Ile Val Pro Gly Asn Gly Lys Ile Thr Ile Gln
 980 985 990
 Leu Lys Gln Asp Trp Gly Leu Asn Asp Val Trp Asn Asp Leu Ile Leu
 995 1000 1005
 Pro Arg Phe Glu Arg Met Asn Gln Leu Thr Asn Ser Thr Asp Phe
 1010 1015 1020
 Thr Ala Trp Asn Lys Asn His Gln Lys Tyr Leu Pro Thr Val Pro

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

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

1130	1135	1140
Val Ile Asn Lys Ala Thr Asn	His Tyr Glu Lys Tyr	Val Glu Lys
1145	1150	1155
Glu Gly Val Lys Met Lys Glu	Arg Val Glu Gln Thr	Gly Thr Asn
1160	1165	1170
Trp Ala Ile Arg Lys Ser Leu	His Glu Glu Thr Val	Ser Gly Lys
1175	1180	1185
Ile Thr Leu Ser Trp Val Glu	Thr Ser Lys Gly Glu	Phe Ile Thr

1190	1195	1200
Ala Thr Arg Lys Pro Leu Asp	Ser Ser Phe Thr Phe	Glu Lys Ile
1205	1210	1215
Asn Lys Ile Thr Asp Thr Gly	Ile Gln Lys Ile Leu	Asn Asn Tyr
1220	1225	1230
Leu Glu Ala Lys Asp Asn Asn	Pro Glu Leu Ala Phe	Ser Ala Glu
1235	1240	1245
Gly Ile Glu Asp Leu Asn Lys	Asn Ile Glu Lys Tyr	Asn Asp Gly

1250	1255	1260

Lys Ser His Gln Pro Ile Ile Lys Ile Arg Leu Tyr Glu Lys Gly
 1265 1270 1275
 Lys Gly Arg Phe Val Leu Gly Gln Thr Gly Asn Lys Ile Asn Lys
 1280 1285 1290
 Tyr Val Gln Gly Ser Pro Asn Leu Phe Phe Ala Ile Tyr Lys Asp
 1295 1300 1305
 Glu Asn Gly Lys Lys Ile Phe Glu Ser Ile Arg Leu Asp Ile Val
 1310 1315 1320
 Ile Glu Arg Leu Lys Gln Gly Leu Gln Ala Ile Pro Glu Thr Asn
 1325 1330 1335
 Gln Asn Gly Val Ser Leu Tyr Gln Ser Leu Ser Pro Leu Asp Leu
 1340 1345 1350
 Val Tyr Ile Pro Thr Glu Tyr Glu Leu Glu Ser Pro His Ile Leu
 1355 1360 1365
 Asp Phe Ser Lys Leu Asn Lys Thr Gln Ile Thr Arg Leu Tyr Asn
 1370 1375 1380
 Thr Asn Asp Phe Ser Gly Val Thr Ala Tyr Phe Ser Gln Asn Ser
 1385 1390 1395
 Phe Ala Lys His Ile Tyr Pro Lys Glu Met Asp Leu Ser Trp Asn
 1400 1405 1410
 Glu Lys Lys Gln Lys Leu Ser Gly Ser Phe Asp Ser Lys Thr Ala
 1415 1420 1425
 Ser Tyr Asn Asn Thr Ser Ile Lys Asp Ile Phe Ile Lys Val Lys
 1430 1435 1440
 Val Asp Arg Leu Gly Asn Ile Ser Lys Ala
 1445 1450
 <210> 24
 <211> 29
 <212> RNA
 <213> Chryseobacterium sp.
 <400> 24
 guugugaauu gcuuucaaaa auuauuauuc

<210> 25
 <211> 83
 <212> RNA
 <213> Chryseobacterium sp.
 <400> 25
 uaauuuuuu gaaagcauu cacaauaagg auuauuccgu ugugaaaaca uucaaggcgg 60
 ggcaacucgc cuuuuuucgu uuu 83

<210> 26
 <211> 145
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <220><221> modified_base
 <222> (1)..(29)
 <223> a, c, u, g, unknown or other
 <400> 26
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugugaauug cuuucaaaaa uuauuuauca 60
 aguaauuuu uugaaagcaa uucacaauaa ggauuuaucc guugugaaaa cauucaaggc 120
 ggggcaacuc gccuuuuuuc guuuu 145

<210> 27
 <211> 146
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 27
 gagcggacag cagcuuccua uaucucguac guugugaauu gcuuucaaaa auuauuuauca 60
 aaguaauauu uuugaaagca auucacaaua aggauuuauuc cguugugaaa acauucaagg 120
 cggggcaacu cgccuuuuuu cguuuu 146

<210> 28

<211> 146

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 28

ccaugauaua gacguugugg cuguuguagu guugugaaau gcuuucaaaa auuauuauca 60

aaguaauauu uuugaaagca auucacaaaua aggauuauuc cguugugaaa acauucaagg 120

cggggcaacu cgccuuuuuu cguuuu 146

<210> 29

<211> 6

<212> DNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<220><221> modified_base

<222> (1)..(2)

<223> a, c, t, g, unknown or other

<400> 29

nnraar 6

<210> 30

<211> 1093

<212> PRT

<213> Brucella sp.

<400> 30

Met Lys Thr Leu Gly Ile Asp Leu Gly Thr Ala Ser Ile Gly Trp Ala

1 5 10 15

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

20 25 30

Gln Ser Glu Met Ala Gly Arg Asp Pro Gln Ser Lys Ala Ser Leu Ala

35 40 45
 Val Ala Arg Arg Asp Ala Arg Gly Ala Arg Arg Arg Arg Asp Arg Tyr
 50 55 60
 Leu Lys Arg Arg Arg Arg Leu Leu Asp Leu Leu Thr Glu His Gly Leu
 65 70 75 80
 Ala Pro Ala Asp Glu Lys Ser Arg Lys Ala Leu Val Arg Glu Tyr Glu
 85 90 95
 Asp Gly Lys Asp Gly Asp Leu Ser Asn Ser Val Tyr Ala Leu Arg Ala
 100 105 110

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

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

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

Val Lys Leu Thr Pro Ala Gln Arg Asp Val Leu Thr Leu Ala Leu Arg
 290 295 300

Thr Gly Leu Thr Lys Gln Gly Arg Leu Pro Phe Ser Lys Leu Arg Ser
 305 310 315 320

Ile Leu Lys Leu Gly Lys Glu Val Arg Phe Asn Lys Glu Lys Asp Asn
 325 330 335

Arg Thr Asp Leu Glu Gly Asp Val Ile His Phe Arg Val Ser Arg Pro
 340 345 350

Asp Cys Phe Gly Asn Arg Trp Ala Ala Met Pro Val Glu Glu Gln Ala
 355 360 365

Ala Val Thr Glu Lys Leu Arg Thr Glu Pro Asp Tyr Ser Ala Leu Leu
 370 375 380

Asp Trp Leu Lys Asn Glu Ala Gly Leu Asp Glu Ala His Ala Arg Ala
 385 390 395 400

Val Ala Asp Thr Pro Val Pro Asp Gly Phe Gly Arg Met Gly Pro Ser
 405 410 415

Ala Leu Ser Ala Leu Ala Asp Ala Met Glu His Glu Ile Asp Ala Gln
 420 425 430

Gly Phe Val Ile Thr Glu Ala Glu Ala Ala Lys Arg Val Tyr Gly Arg
 435 440 445

Thr Asn Ser Glu Ala Asp Pro Gly Arg Lys Gly Val Asp Gln Leu Pro
 450 455 460

Lys Tyr Gln Glu Val Leu Gln Arg His Ile Pro Pro Gly Thr Gly Glu
 465 470 475 480

Pro Asp Asp Pro Tyr Asp Glu Tyr Met Gly Arg Ile Thr Asn Pro Thr
 485 490 495

Val His Ile Ala Leu Asn Gln Leu Arg Arg Leu Val Asn Ala Leu Ile
 500 505 510

Arg Lys Tyr Gly Lys Pro Asn Lys Ile Ala Ile Glu Val Gly Arg Glu
 515 520 525

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

530 535 540
 Gln Asn Thr Arg Ala Ala Met Ala Arg Gly Gln Gln Leu Val Glu Ile
 545 550 555 560

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

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

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

 Asp Gly Ala Lys Glu Lys Asn Arg Leu Asp His Arg His His Ala Ile
 755 760 765
 Asp Ala Ile Val Ile Ala Cys Thr Ser Gln Ser Leu Ile Asn Arg Leu
 770 775 780

Ser Thr Ala Ala Gly Glu Ala Glu Glu Arg Gly Ala Ala Arg Val Val
785 790 795 800
Glu Arg Ile Asp Pro Pro Trp Pro Ser Phe Arg Glu Asp Val Arg Glu
805 810 815

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

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

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

Met Ile Ala Tyr Asp Asp Pro Asp Ser Gly Lys Arg Val Ile Ala
1010 1015 1020
Ile Val Arg Lys Phe Asp Gln Arg Asn Lys Gln Leu Tyr Leu Asp

1025 1030 1035
 Pro His Asn Glu Ala Gly Asn Leu Asp Gln Arg Glu Lys Glu Lys
 1040 1045 1050
 Thr Tyr Lys Pro Leu Arg Pro Met Pro Asn Pro Leu Lys Lys Tyr
 1055 1060 1065

Arg Pro Arg Gln Val Arg Val Asp Glu Ile Gly Gln Val Phe Asp
 1070 1075 1080
 Pro Gly Pro Trp Trp Glu Lys Arg Ser Asp
 1085 1090

<210> 31

<211> 27

<212> RNA

<213> Brucella sp.

<400> 31

guugcggguug gccugcgauu ucugaac

27

<210> 32

<211> 81

<212> RNA

<213> Brucella sp.

<400> 32

guucagaaau cgaguccag ccguuaacaa gcugagauau gcaccaaaua aggcgcucgc

60

uucggcgggc gcuuuuucgu u

81

<210> 33

<211> 141

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 33

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugcgguugg ccugcgauuu cugaacaaag 60
 guucagaaa cgcaguccag ccguuaacaa gcugagauau gcaccaaaua aggcgcucgc 120
 uucggcgggc gcuuuuucgu u 141
 <210> 34
 <211> 142

<212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 34

gagcggacag cagcuuccua uaucucguac guugcgguug gccugcgauu ucugaacaaa 60
 gguucagaaa ugcagucca gccguuaaca agcugagaua ugcaccaaau aaggcgcucg 120
 cuucggcggg cgcuuuuucg uu 142
 <210> 35
 <211> 142
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic

polynucleotide"
 <400> 35
 ccaugauaua gacguugugg cuguuguagu guugcgguug gccugcgauu ucugaacaaa 60
 gguucagaaa ugcagucca gccguuaaca agcugagaua ugcaccaaau aaggcgcucg 120
 cuucggcggg cgcuuuuucg uu 142
 <210> 36
 <211> 6
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <220><221> modified_base

<222> (1)..(2)

<223> a, c, t, g, unknown or other

<220><221> modified_base

<222> (4)..(4)

<223> a, c, t, g, unknown or other

<400> 36

nnrnat

6

<210> 37

<211> 6

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (2)..(2)

<223> a, c, u, g, unknown or other

<220><221> modified_base

<222> (4)..(5)

<223> a, c, u, g, unknown or other

<400> 37

cnanng

6

<210> 38

<211> 1054

<212> PRT

<213> Staphylococcus sp.

<400> 38

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

1 5 10 15

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

20 25 30

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

35 40 45

Ser Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Arg Ile His Arg Leu

50

55

60

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

65

70

75

80

Gln Ile Pro Thr Ser Asn Asn Pro Tyr Gln Ile Arg Val Lys Gly Leu

85

90

95

Ser Glu Ile Leu Ser Lys Asp Glu Leu Ala Ile Ala Leu Leu His Leu

100

105

110

Ala Lys Arg Arg Gly Ile His Asn Ile Asn Val Ser Ser Glu Asp Glu

115

120

125

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

130

135

140

Lys Leu Leu Lys Asp Lys Tyr Val Cys Glu Val Gln Leu Gln Arg Leu

145

150

155

160

Lys Glu Gly Gln Ile Arg Gly Glu Lys Asn Arg Phe Lys Thr Thr Asp

165

170

175

Ile Leu Lys Glu Ile Asp Gln Leu Leu Lys Val Gln Lys Asp Tyr His

180

185

190

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

195

200

205

Arg Arg Glu Tyr Phe Glu Gly Pro Gly Gln Gly Ser Pro Phe Gly Trp

210

215

220

Asn Gly Asp Leu Lys Lys Trp Tyr Glu Met Leu Met Gly His Cys Thr

225

230

235

240

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

245

250

255

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

260

265

270

Asn Ser Glu Lys Leu Glu Tyr His Glu Lys Tyr His Ile Ile Glu Asn

275

280

285

Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu

290 295 300
 Ile Gly Val Asn Pro Glu Asp Ile Lys Gly Tyr Arg Ile Thr Lys Ser

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

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

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

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

Leu Met Asp Leu Leu Asn Asn Pro Gln Asn Tyr Glu Val Asp His Ile
 545 550 555 560
 Ile Pro Arg Ser Val Ala Phe Asp Asn Ser Ile His Asn Lys Val Leu
 565 570 575
 Val Lys Gln Ile Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Tyr Gln
 580 585 590
 Tyr Leu Asn Ser Ser Asp Ala Lys Leu Ser Tyr Asn Gln Phe Lys Gln
 595 600 605
 His Ile Leu Asn Leu Ser Lys Ser Lys Asp Arg Ile Ser Lys Lys Lys
 610 615 620
 Lys Asp Tyr Leu Leu Glu Glu Arg Asp Ile Asn Lys Phe Glu Val Gln
 625 630 635 640
 Lys Glu Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg
 645 650 655
 Glu Leu Thr Ser Tyr Leu Lys Ala Tyr Phe Ser Ala Asn Asn Met Asp
 660 665 670
 Val Lys Val Lys Thr Ile Asn Gly Ser Phe Thr Asn His Leu Arg Lys
 675 680 685
 Val Trp Arg Phe Asp Lys Tyr Arg Asn His Gly Tyr Lys His His Ala
 690 695 700
 Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Leu Phe Lys Glu Asn
 705 710 715 720
 Lys Lys Leu Gln Asn Ala Asn Lys Ile Leu Glu Lys Pro Thr Ile Glu
 725 730 735
 Asn Asn Thr Lys Lys Val Thr Val Glu Lys Glu Glu Asp Tyr Asn Asn
 740 745 750
 Val Phe Glu Thr Pro Lys Leu Val Glu Asp Ile Lys Gln Tyr Arg Asp
 755 760 765
 Tyr Lys Phe Ser His Arg Val Asp Lys Lys Pro Asn Arg Gln Leu Ile
 770 775 780
 Asn Asp Thr Leu Tyr Ser Thr Arg Met Lys Asp Glu His Asp Tyr Ile

785 790 795 800
 Val Gln Thr Ile Thr Asp Ile Tyr Gly Lys Asp Asn Thr Asn Leu Lys
 805 810 815
 Lys Gln Phe Asn Lys Asn Pro Glu Lys Phe Leu Met Tyr Gln Asn Asp

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

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

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

 1010 1015 1020
 Glu Ser Ile Glu Lys Phe Thr Thr Asp Val Leu Gly Asn Leu Tyr
 1025 1030 1035

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

1040

1045

1050

Leu

<210> 39

<211> 28

<212> RNA

<213> Staphylococcus sp.

<400> 39

guuuuaguac ucuguaauuu uagguaug

28

<210> 40

<211> 85

<

212> RNA

<213> Staphylococcus sp.

<400> 40

ucauaccuaa aaauacagaa ucuacugaaa caagacuaua ugucguguuu auccacuaa

60

uuuuuuagug ggauuuuuuu guuuu

85

<210> 41

<211> 146

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 41

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uuuuaguacu cuguaauuuu agguaugaaa

60

gucauaccua aaauacaga aucuacugaa acaagacuau augucguguu uauccacua

120

auuuuuuagu gggauuuuuu uguuuu

146

<210> 42

<211> 147

<212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 42
 gagcggacag cagcuuccua uaucucguac guuuuaguac ucuguaauuu uagguaugaa 60
 agucauaccu aaaaauacag aaucucuga aacaagacua uaugucgugu uuaucccacu 120
 aauuuauuag uggaauuuuu uguuuuu 147

<210> 43
 <211> 147
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 43
 ccaugauuaa gacguugugg cuguuguagu guuuuaguac ucuguaauuu uagguaugaa 60
 agucauaccu aaaaauacag aaucucuga aacaagacua uaugucgugu uuaucccacu 120
 aauuuauuag uggaauuuuu uguuuuu 147

<210> 44
 <211> 4
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic

oligonucleotide"
 <220><221> modified_base
 <222> (1)..(2)
 <223> a, c, t, g, unknown or other
 <400> 44
 nngg 4
 <210> 45

<211> 6
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <220><221> modified_base
 <222> (2)..(2)
 <223> a, c, u, g, unknown or other
 <220><221> modified_base
 <222> (4)..(5)
 <223> a, c, u, g, unknown or other

<400> 45

cnannu

6

<210> 46

<211> 1076

<212> PRT

<213> Bacillus sp.

<400> 46

Met Ser Glu Leu Asp Tyr Arg Ile Gly Leu Asp Ile Gly Thr Asn Ser

1 5 10 15

Ile Gly Trp Gly Val Ile Glu Leu Phe Trp Asn Lys Asp Arg Glu Arg

20 25 30

Tyr Glu Lys Val Arg Ile Val Asp Lys Gly Val Arg Met Phe Asp Lys

35 40 45

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

50 55 60

Ala Arg Ser Ser Arg Arg Arg Leu Asn Arg Lys Ser Gln Arg Lys Lys

65 70 75 80

Glu Ile Arg Asn Leu Leu Val Gln His Gly Val Ile Thr Gln Glu Glu

85 90 95

Leu Asp Leu Leu Tyr Pro Leu Ser Lys Lys Ser Met Asp Ile Trp Asp

100	105	110	
Ile Arg Leu Asp Gly Leu Asp Arg Leu Leu Asn His Leu Glu Trp Thr			
115	120	125	
Arg Leu Leu Ile His Leu Ala Gln Arg Arg Gly Phe Lys Ser Asn Arg			
130	135	140	
Lys Ser Glu Leu Lys Asp Ala Glu Thr Gly Lys Val Leu Ser Ser Ile			
145	150	155	160
Gln Val Asn Glu Lys Arg Leu Ser Leu Tyr Arg Thr Val Gly Glu Met			
165	170	175	
Trp Ile Lys Asp Ala Glu Phe Ser Lys Tyr Asp Arg Arg Arg Asn Ser			
180	185	190	
Pro Asn Glu Tyr Val Phe Ser Val Ser Arg Ala Asp Leu Glu Lys Glu			
195	200	205	
Ile Val Thr Leu Phe Glu Ala Gln Arg Lys Phe Gln Ser Ser Tyr Ser			
210	215	220	
Ser Lys Asn Leu Gln Glu Thr Tyr Leu Gln Ile Trp Ala His Gln Leu			
225	230	235	240
Pro Phe Ala Ser Gly Asn Ala Ile Leu Asn Lys Val Gly Tyr Cys Ser			
245	250	255	
Leu Leu Lys Gly Lys Glu Arg Arg Ile Pro Lys Ala Thr Tyr Thr Phe			
260	265	270	
Gln Tyr Phe Ser Ala Leu Asp Gln Val Asn Arg Thr Arg Leu Gly Pro			
275	280	285	
Asp Phe Gln Pro Phe Thr Gln Glu Gln Lys Glu Ile Ile Leu Asp Lys			
290	295	300	
Met Phe Gln Arg Thr Asp Tyr Tyr Lys Lys Lys Thr Ile Pro Glu Val			
305	310	315	320
Ser Tyr Tyr Asp Ile Arg Lys Trp Leu Glu Leu Asp Glu Thr Ile Gln			
325	330	335	
Phe Lys Gly Leu Asn Tyr Asp Pro Asn Glu Glu Leu Lys Lys Ile Glu			
340	345	350	

Lys Lys Pro Phe Ile Asn Leu Lys Ala Phe Tyr Glu Ile Lys Lys Val
 355 360 365

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

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

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

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

595 600 605
 Ser Gln Ser Phe Ile Asp Ser Tyr His Asn Lys Val Leu Val Tyr Ser
 610 615 620

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

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

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

 Asp Tyr Ile Phe Val Ser Arg Met Pro Lys Arg Ser Val Thr Gly Ala
 820 825 830
 Ala His Asn Gln Met Ile Arg Arg Lys Gly Gly Ile Asp Lys Lys Gly
 835 840 845

Lys Thr Ile Ile Ile Lys Arg Val Arg Leu Lys Asp Ile Lys Phe Asp
850 855 860

Glu Asn Gly Asp Phe Lys Met Val Gly Lys Glu Gln Asp Leu Ala Thr
865 870 875 880

Tyr Glu Ala Ile Lys Gln Arg Tyr Leu Asp His Gly Lys Asn Leu Lys
885 890 895

Lys Ala Phe Glu Thr Pro Leu Tyr Lys Pro Ser Lys Lys Gly Thr Gly
900 905 910

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

Arg Glu Val Asn Gly Gly Val Ala Gln Asn Ser Asp Leu Val Arg Val
930 935 940

Asp Leu Phe Glu Lys Asp Asp Lys Tyr Tyr Met Val Pro Ile Tyr Val
945 950 955 960

Pro Asp Thr Val Cys Ser Glu Leu Pro Lys Lys Val Val Lys Ser Gly
965 970 975

Lys Gly Tyr Glu Gln Trp Leu Thr Leu Asp Asn Ser Phe Thr Phe Lys
980 985 990

Phe Ser Leu Tyr Pro Tyr Asp Leu Val Arg Leu Val Lys Gly Asp Glu
995 1000 1005

Asp Arg Phe Leu Tyr Phe Gly Thr Leu Asp Ile Asp Ser Asp Arg
1010 1015 1020

Leu Asn Phe Lys Asp Val Asn Lys Pro Ser Lys Gln Asn Glu Tyr
1025 1030 1035

Arg Tyr Ser Leu Lys Thr Ile Glu Asp Leu Glu Lys Tyr Glu Val
1040 1045 1050

Gly Val Leu Gly Asp Leu Arg Leu Val Arg Lys Glu Thr Arg Arg
1055 1060 1065

Asn Phe His Glu Ile Lys Ile Asn
1070 1075

<210> 47

<211> 20

<212> RNA

<213> Bacillus sp.

<400> 47

gucauaguuc cauuaaagcc 20

<210> 48

<211> 84

<212> RNA

<213> Bacillus sp.

<400> 48

ggcuuugaug uuucuaugau aaggguuucg gcccguggcg ucggggaucg ccugcccauu 60

ccgaugggcu ucuccecauu uauu 84

<210> 49

<211> 137

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 49

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng ucuauguucc auuaaagcca aagggcuuug 60

auguuucuaugauu gaaagggguu ucggcccgug gcgucgggga ucgccugccc auuccgaugg 120

gcuucucuccc auuuauu 137

<210> 50

<211> 138

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 50

gagcggacag cagcuuccua uaucucguac gucauaguuc cauuaaagcc aaagggcuuu 60

gauguuucua ugauaagggu uucggcccgu ggcgucgggg aucgccugcc cauuccgaug 120

ggcuucucucc cauuuauu 138

<210> 51

<211> 138

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 51

cgaugauua gagcuugugg cuguuguagu gucauaguuc cauuaaagcc aaagggcuuu 60

gauguuucua ugauaagggu uucggcccgu ggcgucgggg aucgccugcc cauuccgaug 120

ggcuucucucc cauuuauu 138

<210> 52

<211> 6

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(4)

<223> a, c, t, g, unknown or other

<400> 52

nnnncc 6

<210> 53

<211> 6

<212>

> RNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<220><221> modified_base

<222> (2)..(2)

<223> a, c, u, g, unknown or other

<220><221> modified_base

<222> (4)..(5)

<223> a, c, u, g, unknown or other

<400> 53

unannng

<210> 54

<211> 1108

<212> PRT

<213> Brevibacillus sp.

<400> 54

Met Ala Ile Arg Ser Ile Lys Leu Lys Leu Lys Thr Arg Thr Gly Pro

1 5 10 15

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

20 25 30

Asn Glu Gly Val Ala Tyr Tyr Met Lys Met Leu Leu Leu Phe Arg Gln

35 40 45

Glu Ser Thr Gly Gly Gln Thr Lys Lys Glu Leu Gln Glu Glu Leu Val

50 55 60

Arg His Ile Arg Glu Gln Gln Gln Lys Asn Arg Ala Asp Lys Asn Thr

65 70 75 80

Gln Ala Leu Pro Leu Asp Lys Ala Phe Ala Ala Leu Arg Gln Leu Tyr

85 90 95

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

100 105 110

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

115 120 125

Gly Lys Gly Thr Ser Lys Ala Gly Ala Lys Pro Thr Trp Gln Lys Lys

130 135 140

6

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

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

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

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

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

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

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

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

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

645 650 655

Gln Asn Arg Glu Arg Ala Asp Glu Leu Asp His Leu Ser Gln Ala Leu

660 665 670

Glu Lys Gln Lys Ser Leu Leu Asp Gln Thr Asp Arg Thr Phe Trp Asn

675 680 685

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

690 695 700

Trp Glu Gln Ala Val Val Gln Ile His Arg Lys Ala Glu Glu His Val

705 710 715 720

Gly Lys Val Val Gln Ala Trp Arg Lys Arg Phe Asp Ala Asp Glu Arg

725 730 735

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

740 745 750

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

755 760 765

Glu Ile Asn Arg Phe Glu Gln Gly His Thr Ser His Gln Arg Leu Leu

770 775 780

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

785 790 795 800

Ala Ile Val Met Thr Ala Leu Gly Tyr Val Tyr Asp Glu Lys Lys Leu

805 810 815

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

820 825 830

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

835 840 845

Thr Leu Met Lys Trp Ala His Arg Ser Ile Pro Lys Tyr Val His Met

850 855 860

Gln Ala Glu Pro Tyr Gly Ile Gln Ile Gly Asp Val Arg Ala Glu Tyr

865 870 875 880

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

885 890 895
Met Val Lys Gly Gln Glu Leu Gln Gly Lys Arg Phe Glu Asn Leu Gln

900 905 910
Lys Arg Leu Val Ser Glu Gln Phe Leu Thr Glu Glu Gln Val Lys Gln

915 920 925
Leu Arg Pro Gly Asp Ile Val Pro Asp Asp Ser Gly Glu Trp Phe Met

930 935 940
Thr Leu Ser Asp Gly Ser Glu Gly Lys Glu Val Val Phe Leu Gln Ala

945 950 955 960
Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Gln Arg Tyr

965 970 975
Asn Glu Leu Phe Lys Val Ser Cys Arg Val Leu Ile Arg Gly Glu Glu

980 985 990
Glu Tyr Leu Ile Pro Lys Ala Lys Ser Val Gln Ala Lys Leu Gly Lys

995 1000 1005
Gly Leu Phe Val Lys Lys Thr Asp Thr Val Met Lys Asp Val Tyr

1010 1015 1020
Val Trp Asp Ser Gln Ala Lys Leu Lys Gly Lys Thr Thr Phe Thr

1025 1030 1035
Glu Glu Ser Glu Ser Pro Glu Gln Leu Glu Asp Phe Gln Glu Ile

1040 1045 1050
Ile Glu Glu Ala Glu Glu Ala Lys Gly Thr Tyr Arg Thr Leu Phe

1055 1060 1065
Arg Asp Pro Ser Gly Val Phe Phe Pro Glu Phe Val Trp Asn Thr

1070 1075 1080
Gln Lys Asp Phe Trp Ser Glu Val Lys Arg Arg Leu Tyr Gly Lys

1085 1090 1095
Leu Arg Glu Arg Phe Leu Met Lys Thr Arg

1100 1105
<210> 55
<211> 22

<212> RNA

<213> Brevibacillus sp.

<400> 55

uggaaagcuu cgagguuagc ac 22

<210> 56

<211> 105

<212> RNA

<213> Brevibacillus sp.

<400> 56

cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cggcuaccac gaauagucac gaaguuccac ugagu 105

<210> 57

<211> 160

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (132)..(160)

<223> a, c, u, g, unknown or other

<400> 57

cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc 120

gagguuagca cnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 160

<210> 58

<211> 161

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 58

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60
cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc 120
gagguuagca cguacgagau auaggaagcu gcuguccgcu c 161

<210> 59

<211> 161

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 59

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60
cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc 120
gagguuagca cacuacaaca gccacaacgu cuauaucaug g 161

<210> 60

<211> 4

<212> DNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<220><221> modified_base

<222> (4)..(4)

<223> a, c, t, g, unknown or other

<400> 60

dttn

4

<210> 61

<211> 1150

<212> PRT

<213> Enterococcus sp.

<400> 61

Met Tyr Ser Ile Gly Leu Asp Leu Gly Ile Ser Ser Val Gly Trp Ser

1

5

10

15

Val Ile Asp Glu Arg Thr Gly Asn Val Ile Asp Leu Gly Ile Arg Leu
20 25 30

Phe Ser Ala Lys Asn Ser Glu Lys Asn Leu Glu Arg Arg Thr Asn Arg
35 40 45

Gly Gly Arg Arg Leu Ile Arg Arg Lys Thr Asn Arg Leu Lys Asp Ala
50 55 60

Lys Lys Ile Leu Ala Ala Val Gly Phe Tyr Glu Asp Lys Ser Leu Lys
65 70 75 80

Asn Ser Cys Pro Tyr Gln Leu Arg Val Lys Gly Leu Thr Glu Pro Leu
85 90 95

Ser Lys Gly Glu Ile Tyr Lys Val Thr Leu His Ile Leu Lys Lys Arg
100 105 110

Gly Ile Ser Tyr Leu Asp Glu Asp Asp Thr Glu Ala Ala Lys Glu Ser
115 120 125

Gln Asp Tyr Lys Glu Gln Val Arg Glu Asn Ala Gln Leu Leu Thr Lys
130 135 140

Tyr Thr Pro Gly Gln Ile Gln Leu Gln Arg Leu Lys Glu Asn Asn Arg
145 150 155 160

Val Lys Thr Gly Ile Asn Ala Gln Gly Asn Tyr Gln Leu Asn Val Phe
165 170 175

Lys Val Ser Ala Tyr Ala Asn Glu Leu Ala Thr Ile Leu Lys Thr Gln
180 185 190

Gln Ala Phe Tyr Pro Asn Glu Leu Thr Asp Asp Trp Ile Ala Leu Phe
195 200 205

Val Gln Pro Gly Ile Ala Glu Glu Ala Gly Leu Ile Tyr Arg Lys Arg
210 215 220

Pro Tyr Tyr His Gly Pro Gly Asn Glu Ala Asn Asn Ser Pro Tyr Gly
225 230 235 240

Arg Trp Ser Asp Phe Lys Lys Thr Gly Gln Pro Ala Thr Asn Ile Phe
245 250 255

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

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

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

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

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

 Ala Leu Leu Val Lys Tyr Gly Lys Glu Gln Ile Arg Tyr Ile Thr Ile
 500 505 510

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

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

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

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

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

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

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

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

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

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

Ser Gln Cys Tyr Asp Val Arg Phe Tyr Asn Phe Gln Ser Ala Asn
 1070 1075 1080
 Ser Ile Asn Phe Lys Gly Leu Glu Gln Glu Met Met Pro Ala Glu
 1085 1090 1095
 Asn Gln Phe Lys Gln Lys Pro Tyr Asn Asn Gly Ala Ile Asn Leu
 1100 1105 1110
 Asn Ile Ala Lys Tyr Gly Lys Glu Gly Lys Lys Leu Arg Lys Phe
 1115 1120 1125

Asn Thr Asp Ile Leu Gly Lys Lys His Tyr Leu Tyr Tyr Glu Lys
 1130 1135 1140
 Glu Pro Lys Asn Ile Ile Lys
 1145 1150

<210> 62

<211> 19

<212> RNA

<213> Enterococcus sp.

<400> 62

guuuuuguac ucucaauaa

19

<210> 63

<211> 73

<212> RNA

<213> Enterococcus sp.

<400> 63

uuauugagaa ucuacaaaaa uaaggcauuu ugccgaauuu accgcccuaa auauguaggg

60

cgguuuuuuu auu

73

<210> 64

<211> 125

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 64

nnnnnnnnnn nnnnnnnnnn nnnnnnnng uuuuuguacu cucauaaaa aguuauugag 60

aaucuacaaa aaauaggcau uuugccgaau uuaccgcccu acauauguag ggcggguuuu 120

uuauu 125

<210> 65

<211> 126

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 65

gagcggacag cagcuuccua uaucucguac guuuuuguac ucucauaaaa aaguauuga 60

gaauacucaa aaauaaggca uuugccgaa uuuaccgcc uacauaugua ggcggguuuu 120

uuuauu 126

<210> 66

<211> 126

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 66

ccaugauuaa gacguugugg cuguuguagu guuuuuguac ucuaauaaa aaguuaauuga 60

gaaucuacaa aaauaaggca uuugccgaa uuuaccgccc uacauaugua gggcgguuuu 120

uuuauu 126

<210> 67

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(4)

<223> a, c, t, g, unknown or other

<400> 67

nnnna 5

<210> 68

<211> 6

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (2)..(2)

<223> a, c, u, g, unknown or other

<220><221> modified_base

<222> (4)..(5)

<223> a, c, u, g, unknown or other

<400> 68

unannc 6

<210> 69

<211> 1408

<212> PRT

<213> Sphingobacterium sp.

<400> 69

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

Arg Ser Ala Arg Arg Asn Leu Gln Arg Phe Lys Leu Arg Arg Asn Asn

65 70 75 80

Leu Ile Glu Val Leu Lys Arg Glu Lys Trp Ile Asp Asp Lys Ser Ile

85 90 95

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

100 105 110

Ala Ala Ala Ala Ser Lys Glu Ile Lys Leu Glu Glu Leu Ala Arg Val

115 120 125

Leu Leu Met Ile Asn Lys Lys Arg Gly Tyr Lys Ser Ser Arg Lys Val

130 135 140

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

145 150 155 160

Lys Leu Tyr His Glu Asn Leu Thr Pro Gly Glu Phe Val Cys Arg Leu

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

465 470 475 480
 Leu Tyr Asn Ala Tyr Leu Asn Ile Leu Asp Ile Glu Gly Tyr Asp Val
 485 490 495
 Arg Ser Glu Leu Lys Ile Lys Leu Asn Lys Asp Glu Ile Thr Leu Ser

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

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

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

 690 695 700
 Ile Glu Leu Ala Arg Glu Leu Lys Lys Asn Ala Lys Glu Arg Glu Glu
 705 710 715 720

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

965	970	975
Asn His Leu Asn Ala Arg Lys Asn Glu Lys Ser Glu Phe Phe Ser Ser		
980	985	990
Ile Lys Ala Ile Glu Ile Lys Glu Thr Tyr Val Glu Arg Asp Asp Leu		
995	1000	1005
Gly Asn Arg Lys Arg Leu Phe Lys Glu Pro Ile Ala His Phe Arg		
1010	1015	1020
Thr Ile Val Lys Glu His Leu Glu Arg Val Leu Val Ser His Lys		
1025	1030	1035
Ala Lys Asn Lys Val Val Thr Lys Asn Arg Asn Lys Ile Asp Gly		
1040	1045	1050
Lys Lys Glu Ala Gln Glu Val Leu Thr Pro Arg Gly Gln Leu His		
1055	1060	1065
Lys Glu Thr Ile Tyr Gly Lys Ile Leu Gln Tyr Ala Ser Lys Glu		
1070	1075	1080
Glu Lys Ile Ser Ala Lys Phe Asp Ala Ala Thr Ile Ala Met Val		
1085	1090	1095
Ser Asn Pro Arg Tyr Arg Ser Ala Leu Leu Asp Arg Leu Ile Glu		
1100	1105	1110
Tyr Gly Asn Asp Pro Lys Lys Ala Phe Thr Gly Lys Asn Ser Pro		
1115	1120	1125
Ser Lys Ser Pro Ile Tyr Leu Asp Gln Gln Arg Gln Leu Ala Val		
1130	1135	1140
Pro Glu Lys Val Lys Leu Val Trp Leu Glu Ala Asp Tyr Thr Ile		
1145	1150	1155
Arg Lys Glu Ile Gly Pro Asp Leu Lys Ile Glu Lys Val Ile Asp		
1160	1165	1170
Gln Gly Val Lys Arg Ile Leu Lys Gln Arg Leu Ser Val Tyr Asp		
1175	1180	1185
Gly Asp Thr Lys Lys Ala Phe Ser Asp Leu Asp Asn Asn Pro Ile		
1190	1195	1200

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

<210> 70

<211> 16

<212> RNA

<213> Sphingobacterium sp.

<400> 70	
guugugaaau gcuuuc	16
<210> 71	
<211> 72	
<212> RNA	
<213> Sphingobacterium sp.	
<400> 71	
gaaagcaauu cacaauaagg auuauuuccg uuguguaaac auuuagcgcc ucgucuaucu	60
acggggcauu uu	72
<210> 72	
<211> 121	
<212> RNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"	
<220><221> modified_base	
<222> (1)..(29)	
<223> a, c, u, g, unknown or other	
<400> 72	
nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugugaaug cuuucaaagg aaagcaauuc	60
acaauaagga uuauuuccgu uguguaaaca uuuaagcgccu cgucuaucua cggggcauuu	120
u	121
<210> 73	
<211> 122	
<212> RNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"	
<400> 73	
gagcggacag cagcuuccua uaucucguac guugugaaau gcuuucaaag gaaagcaauu	60
cacaauaagg auuauuuccg uuguguaaac auuuagcgcc ucgucuaucu acggggcauu	120

uu 122

<210> 74

<211> 122

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 74

ccaugauua gacguuggg cuguuguagu guugugaaau gcuuucaaag gaaagcaauu 60

cacaauaagg auuauuuccg uuguguaaac auuuagcgcc ucgucuaucu acggggcauu 120

uu 122

<210> 75

<211> 1429

<212> PRT

<213> Chryseobacterium sp.

<400> 75

Met Met Ile Lys His Ile Leu Gly Leu Asp Leu Gly Thr Asn Ser Ile

1 5 10 15

Gly Trp Ala Leu Ile Lys Gln Asn Phe Glu Asn Lys Tyr Gly Glu Ile

20 25 30

Leu Gly Met Gly Ser Arg Ile Ile Pro Met Ser Gln Asp Ile Leu Gly

35 40 45

Glu Phe Gly Lys Gly Asn Ser Val Ser Gln Thr Ala Ala Arg Thr Asp

50 55 60

Tyr Arg Gly Ile Arg Arg Leu Arg Glu Arg Phe Leu Leu Arg Arg Glu

65 70 75 80

Arg Leu His Arg Ile Leu Asn Val Leu Asn Phe Leu Pro Glu His Tyr

85 90 95

Ala Ser Gln Ile Asp Phe Asp Lys Arg Phe Gly Lys Phe Lys Val Glu

100 105 110

Thr Glu Pro Lys Leu Ala Trp Lys Asn Ser Asp Gly Lys Phe Ser Phe

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

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

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

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

 340 345 350
 Ser Asp Asp Leu Tyr Asp Asp Cys Ile Arg Glu Leu Tyr Arg Asn Asn
 355 360 365

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

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

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

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

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

 850 855 860

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

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

1160	1165	1170
Glu Thr Val Ser Gly Lys Ile	Asp Leu Pro Trp Val	Lys Val Pro
1175	1180	1185
Lys Gly Lys Ile Leu Thr Ala	Thr Arg Lys Ser Leu	Asp Thr Ser
1190	1195	1200
Phe Asp Leu Lys Ala Ile Ala	Ser Ile Thr Asp Thr	Gly Ile Gln
1205	1210	1215
Lys Ile Leu Lys Asn Tyr Leu	Glu Phe Lys Glu Ser	Pro Glu Leu

1220	1225	1230
Ala Phe Ser Pro Glu Gly Ile	Glu Asp Met Asn Lys	Asn Ile Lys
1235	1240	1245
Lys Tyr Asn Gly Gly Lys Pro	His Gln Pro Ile Ser	Lys Val Arg
1250	1255	1260
Val Phe Glu Leu Gly Ser Lys	Phe Gln Val Gly Gln	Thr Gly Asn
1265	1270	1275
Lys Lys Asp Lys Tyr Val Glu	Ala Ala Lys Gly Thr	Asn Leu Phe

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

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

1340 1345 1350

Ser Ile Glu Ile Lys Asn Ile Thr Lys Glu Lys Ser Glu Arg Phe

1355 1360 1365

Tyr Lys Val Val Ser Phe Ser Gly Ser Gln Ile Phe Phe Val Arg

1370 1375 1380

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

1385 1390 1395

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

1400 1405 1410

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

1415 1420 1425

Ala

<210> 76

<211> 19

<212> RNA

<213> Chryseobacterium sp.

<400> 76

guugugaaau gcuucaaa

19

<210> 77

<211> 75

<212> RNA

<213> Chryseobacterium sp.

<400> 77

uuugaaaagc aaucacaa aaggauuau ccguugugaa aacauucaag gcggggcaac

60

ucgccuuuuu ucguu

75

<210> 78

<211> 127

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 78

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugugaaug cuuucacaaa aguuugaaaa 60

gcaauucaca auaaggauua uuccguugug aaaacauuca aggcggggca acucgccuuu 120

uuucguu 127

<210> 79

<211> 128

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 79

gagcggacag cagcuuccua uaucucguac guugugaaau gcuuucacaaa aaguugaaaa 60

agcaauucac auaaggauu auuccguugu gaaaacauuc aaggcggggc aacucgccuu 120

uuuucguu 128

<210> 80

<211> 128

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 80

ccaugauuaa gacguugugg cuguuguagu guugugaaau gcuuucacaaa aaguugaaaa 60

agcaauucac auaaggauu auuccguugu gaaaacauuc aaggcggggc aacucgccuu 120

uuuucguu 128

<210> 81

<211> 6

<212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <220><221> modified_base
 <222> (1)..(2)
 <223> a, c, t, g, unknown or other

<400> 81

nnraat

6

<210> 82

<211> 1440

<212> PRT

<213> Chryseobacterium sp.

<400> 82

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

1 5 10 15

Ala Leu Val Lys Gln Asp Phe Lys Asn Lys Gln Gly Glu Ile Leu Gly

20 25 30

Met Gly Thr Arg Ile Ile Pro Met Ser Gln Asp Ile Leu Gly Asp Phe

35 40 45

Gly Lys Gly Asn Ser Val Ser Gln Thr Ala Glu Arg Thr Lys Tyr Arg

50 55 60

Ser Ala Arg Arg Leu Arg Glu Arg Phe Leu Leu Arg Arg Glu Arg Leu

65 70 75 80

His Arg Val Leu Asn Ile Leu Asn Phe Leu Pro Glu His Tyr Ala Ser

85 90 95

Gln Ile Asp Phe Glu Lys Arg Phe Gly Lys Phe Lys Val Glu Thr Glu

100 105 110

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

115 120 125

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

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

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

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

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

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

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

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

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

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

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

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

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

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

 Ser Glu Thr Lys Arg Tyr Asp Leu Lys Lys Lys Leu Met Lys Phe
 1070 1075 1080
 Glu Arg Val Val Tyr Leu His Thr Gln Thr Gly Glu Lys Ile Glu
 1085 1090 1095
 Arg Glu Val Pro Lys His Phe Leu Lys Pro Trp Glu Asn Phe Thr
 1100 1105 1110
 Val Asp Val Lys His Asn Leu Asp Thr Val Ile Val Ser Phe Lys

1115	1120	1125
Gln Asn Leu Arg Val Ile Asn Lys Ala Thr Asn Tyr Tyr Glu Lys		
1130	1135	1140
Tyr Ala Glu Asn Gly Gly Ala Arg Asn Lys Ala Arg Val Glu Gln		
1145	1150	1155
Lys Gly Ile Asn Trp Ala Ile Arg Lys Pro Met His Lys Asp Thr		
1160	1165	1170
Val Ser Gly Lys Val Asp Leu Pro Trp Val Lys Val Pro Lys Gly		
1175	1180	1185
Lys Ile Leu Thr Ala Thr Arg Lys Ser Leu Asp Thr Ser Phe Asp		
1190	1195	1200
Leu Lys Ser Ile Gly Ser Ile Thr Asp Thr Gly Ile Gln Lys Ile		
1205	1210	1215
Leu Lys Asn Tyr Leu Met Leu Lys Asp Gly Asn Pro Glu Leu Ala		
1220	1225	1230
Phe Ser Pro Glu Gly Ile Glu Asp Leu Asn Lys Ser Ile Glu Lys		
1235	1240	1245
Tyr Asn Asp Gly Lys Pro His Gln Pro Ile Asn Lys Val Arg Val		
1250	1255	1260
Phe Glu Leu Gly Ser Lys Phe Gln Ile Gly Gln Ala Gly Asn Lys		
1265	1270	1275
Lys Asp Lys Tyr Val Glu Ala Ala Lys Gly Thr Asn Leu Phe Phe		
1280	1285	1290
Ala Val Tyr Glu Asp Lys Asn Gly Lys Arg Ser Tyr Glu Thr Ile		
1295	1300	1305
Pro Leu Asn Glu Val Ile Glu Arg Gln Lys Gln Gly Leu Pro Val		
1310	1315	1320
Val Asp Leu Arg Ser Thr Asn Asp Phe Tyr Leu Cys Pro Asn Asp		
1325	1330	1335
Leu Val Tyr Ile Leu Ser Asp Asp Glu Arg Gly Asn Met Thr Asn		
1340	1345	1350

Asn Asp Phe Glu Asn Leu Ser Asp Glu Gln Val Lys Arg Ile Tyr
1355 1360 1365

Lys Phe Val Ser Cys Thr Gly Gly Glu Gly His Phe Ile Pro Tyr
1370 1375 1380

Ala Asn Ala Thr Glu Ile Ile Lys Asn Glu Asn Gly Thr Asn Ser
1385 1390 1395

Lys Ser Glu Arg Met Gln Asn Phe Tyr Asp Gly Ser Cys Met Leu
1400 1405 1410

Asp Lys Asn Ser Lys Pro Ile Met Ile Lys Glu Asn Cys Val Lys
1415 1420 1425

Leu Lys Val Asp Arg Leu Gly Asn Ile Ser Lys Ile
1430 1435 1440

<210> 83

<211> 20

<212> RNA

<213> Chryseobacterium sp.

<400> 83

guugugaaau gcuucaaaa 20

<210> 84

<211> 79

<212> RNA

<213> Chryseobacterium sp.

<400> 84

uuuugaaagc aaucacaa aaggauuuu ccguugugaa aacauuuaga gccucgacua 60

ccuucggggc auuuuuuuu 79

<210> 85

<211> 132

<212>

> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 85

nnnnnnnnnn nnnnnnnnnn nnnnnnnng uugugaauug cuuucaaaaa aaguuuugaa 60

agcaauucac aaauaggauu auuccguugu gaaaacauu agagccucga cuaccuucgg 120

ggcauuuuua uu 132

<210> 86

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 86

gagcggacag cagcuuccua uaucucguac guugugaauu gcuucaaaaa aaaguuuuga 60

aagcaauuca caauaggau uauuccguug ugaacaacau uagagccucg acuaccuucg 120

gggcauuuuu auu 133

<210> 87

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 87

ccaugauua gacguugugg cuguuguagu guugugaauu gcuucaaaaa aaaguuuuga 60

aagcaauuca caauaggau uauuccguug ugaacaacau uagagccucg acuaccuucg 120

gggcauuuuu auu 133

<210> 88

<211> 6

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(2)

<223> a, c, t, g, unknown or other

<220><221> modified_base

<222> (5)..(5)

<223> a, c, t, g, unknown or other

<400> 88

nnrrng

6

<210> 89

<211> 1071

<212> PRT

<213> Bacillus sp.

<400> 89

Met Arg Glu Leu Asp Tyr Arg Ile Gly Leu Asp Ile Gly Thr Asn Ser

1 5 10 15

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

20 25 30

Tyr Glu Lys Thr Arg Ile Val Asp Lys Gly Val Arg Met Phe Asp Lys

35 40 45

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

50 55 60

Ala Arg Ser Ser Arg Arg Arg Leu Asn Arg Lys Ser Gln Arg Lys Lys

65 70 75 80

Asp Ile Arg Asn Leu Leu Val Gln His Glu Ile Ile Ser Gln Lys Glu

85 90 95

Leu Thr Ser Leu Tyr Pro Leu Ser Lys Ser Ser Met Asp Ile Trp Asp

100 105 110

Ile Arg Leu Asp Gly Leu Asp Arg Leu Leu Asp Arg Phe Glu Trp Ala

115 120 125

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

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

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

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

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

 610 615 620

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

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

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

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

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

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

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

 1055 1060 1065
Ile Phe Lys
 1070
<210> 90
<211> 24
<212> RNA
<213> Bacillus sp.
<400> 90

gucauaguuc cauuaaagcc auug	24
<210> 91	
<211> 138	
<212> RNA	
<213> Bacillus sp.	
<400> 91	
caauggcuuu gauguuucua ugauaagggc uucggcccgu ggcguugggg aucgccugcc	60
cauuuuuauug ggcucucucc caucuauuuu augagaauuu uacaaccuug gcuaauucua	120
aauagcuaag guuuuuuu	138
<210> 92	
<211> 195	
<212> RNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"	
<220><221> modified_base	
<222> (1)..(29)	
<223> a, c, u, g, unknown or other	
<400> 92	
nnnnnnnnnn nnnnnnnnnn nnnnnnnnng ucuauguucc auuaaagcca uugaaagcaa	60
uggcuuugau guuuuauaga uaagggcuuc ggcccguggc guuggggauc gccugcccau	120
uuuaaugggc uucuccccau cuauuuuauug agaauuuuac aaccuuggcu auucuuaau	180
agcuaagguu uuuuu	195
<210> 93	
<211> 196	
<212> RNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"	
<400> 93	
gagcggacag cagcuuccua uaucucguac gucauaguuc cauuaaagcc auugaaagca	60

auggcuuuga uguuucuaug auaagggcuu cggcccgugg cguuggggau cgccugccca 120
uuuuuauggg cuucucccca ucuauuuauu gagaauuuua caaccuuggc uauucuuaaa 180
uagcuaaggu uuuuuu 196

<210> 94

<211> 196

<212

> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 94

ccaugauua gacguuggg cuguuguagu gucauaguuc cauuaaagcc auugaaagca 60
auggcuuuga uguuucuaug auaagggcuu cggcccgugg cguuggggau cgccugccca 120
uuuuuauggg cuucucccca ucuauuuauu gagaauuuua caaccuuggc uauucuuaaa 180
uagcuaaggu uuuuuu 196

<210> 95

<211> 1091

<212> PRT

<213> Rhizobium sp.

<400> 95

Met Asp Ile Glu Thr Leu Phe Ala Phe Asp Ile Gly Thr Asn Ser Ile

1 5 10 15

Gly Phe Cys Val Phe Ala Leu Asp Glu Asn Gly Glu Pro Tyr Arg Ile

20 25 30

Ile Asp Ile Gly Thr Arg Ile Tyr Ala Asp Gly Arg Asp Pro Gln Ser

35 40 45

Lys Ala Ser Leu Ala Val Ala Arg Arg Glu Ala Arg Ala Met Ser Arg

50 55 60

Arg Arg Asp Arg Tyr Val Asn Arg Arg Lys Ala Val Leu Arg Thr Leu

65 70 75 80

Thr Glu Tyr Gly Leu Met Pro Ala Glu Ala Ser Ala Arg Gln Ala Leu

85 90 95

Ile Ala Glu Thr Ser Asp Arg Asn Gly Ala Ala Gly Glu Ala Ser Ala
100 105 110

Asn Pro Tyr Ala Leu Arg Ala Met Ala Leu Asp Glu Lys Leu Pro Leu
115 120 125

Tyr Trp Ile Gly Arg Val Leu Phe His Leu Asn Gln Arg Arg Gly Phe
130 135 140

Lys Ser Asn Arg Lys Thr Asp Arg Lys Asp Asn Glu Lys Gly Val Ile
145 150 155 160

Ala Leu Gly Ile Gly Glu Leu Arg Ala Ala Met His Lys Thr Lys Ala
165 170 175

Arg Thr Tyr Gly Glu Trp Leu Ala Ala Arg Arg Glu Asp Gly His Val
180 185 190

Thr Arg Leu Arg Ala Gly Ser Asp Ala Phe Glu Gly Asp Gly Tyr Ala
195 200 205

Phe Tyr Pro Glu Arg Ser Leu Leu Glu Ala Glu Phe Arg Glu Ile Trp
210 215 220

Ala Arg Gln Ser Ser Phe Tyr Pro Asp Val Leu Thr Glu Ala Arg Gly
225 230 235 240

Asp His Leu Phe Gln Val Met Phe Tyr Gln Arg Pro Leu Lys Lys Pro
245 250 255

Arg Val Gly Lys Cys Ala Phe Asn Pro Leu Glu Glu Arg Leu Ala Lys
260 265 270

Ala His Pro Leu Phe Gln Glu Phe Arg Leu Tyr Lys Glu Val Asn Glu
275 280 285

Leu Glu Leu Val Leu Pro Asp His Ser His Lys Lys Leu Asp Ile Glu
290 295 300

Gln Arg Asn Ala Leu Val Ile Leu Leu Arg Ala Asn Arg Glu Val Thr
305 310 315 320

Phe Lys Lys Leu Arg Thr Ala Leu Lys Leu Ala Pro Asp Ile Val Phe
325 330 335

Asn Lys Glu Ser Glu Ser Arg Glu Lys Leu Arg Gly Asp Glu Val His

340 345 350
 Ser Val Leu Ala Asp Lys Lys Arg Phe Gly Pro Gln Trp Gly Ala Leu
 355 360 365
 Ser Arg Glu Arg Gln Trp Gln Ile Ile Gln Thr Leu Lys Asp Glu Glu
 370 375 380

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

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

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

 Glu Leu Asn Pro Ser Lys Pro Glu Asp Arg Val Cys Ile Tyr Ser Gly
 580 585 590

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

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

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

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

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

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

Ser Lys Ser Thr Arg Lys Thr Asn

1085

1090

<210> 96

<211> 29

<212> RNA

<213> Rhizobium sp.

<400> 96

guugcggcug gaccgcguuu ucugaucug

29

<210> 97

<211> 78

<212> RNA

<213> Rhizobium sp.

<400> 97

cagaucagaa aacgcggucu ggccguuaac aagcuagaag caccaauaa ggccguuccu

60

ucgggagcgg cuuuuucu

78

<210> 98

<211> 140

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 98

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugcggcugg accgcguuuu cugaucugaa

60

agcagaucag aaaacgcggu cuggccguua acaagcuaga agcaccaau aaggccguuc

120

cuucgggagc ggcuuuuucu

140

<210> 99

<211> 141

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 99

gagcggacag cagcuuccua uaucucguac guugcggcug gaccgcguuu ucugaucuga 60

aagcagauca gaaaacgcgg ucuggccguu aacaagcuag aagcaccaaa uaaggccguu 120

ccuucgggag cggcuuuuuc u 141

<210> 100

<211> 141

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 100

ccaugauua gacguuggg cuguuguagu guugcggcug gaccgcguuu ucugaucuga 60

aagcagauca gaaaacgcgg ucuggccguu aacaagcuag aagcaccaaa uaaggccguu 120

ccuucgggag cggcuuuuuc u 141

<210> 101

<211> 3

<212> DNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<220><221> modified_base

<222> (1)..(1)

<223> a, c, t, g, unknown or other

<400> 101

ngg 3

<210> 102

<211> 6

<212> RNA

<213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"

<220><221> modified_base

<222> (2)..(2)

<223> a, c, u, g, unknown or other

<220><221> modified_base

<222> (4)..(5)

<223> a, c, u, g, unknown or other

<400> 102

cnannu

6

<210> 103

<211> 1401

<212> PRT

<213> Myroides sp.

<400> 103

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

1 5 10 15

Ala Tyr Val His Glu Ala Glu Asn Glu Asn Glu Ile Ser Ser Ile Glu

20 25 30

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

35 40 45

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

50 55 60

Arg Ser Ala Arg Arg Asn Leu Gln Arg Phe Lys Leu Arg Arg Lys His

65 70 75 80

Leu Ile Glu Ile Leu Lys His Asn Asn Trp Ile Ser Lys Gln Thr Leu

85 90 95

Leu Thr Glu Asn Gly Asn His Thr Thr Phe Glu Thr Ile Ser Leu Arg

100 105 110

Asp Lys Ala Ala Thr Gln Lys Val Thr Leu Glu Glu Leu Ala Arg Val

115 120 125

Leu Leu Met Ile Asn Lys Lys Arg Gly Tyr Lys Ser Ser Arg Lys Val

130 135 140

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

145 150 155 160

Lys Leu Tyr His Glu Gln Leu Thr Pro Gly Gln Leu Ser Leu Gln Leu

165 170 175

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

180 185 190

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

195 200 205

His Ile Leu Thr Ser Glu Phe Lys Glu Leu Leu Lys Gly Lys Gly Leu

210 215 220

Arg Ala Thr Ser Asn Gln Phe Trp Thr Asn Tyr Asn Phe Asn Thr Ala

225 230 235 240

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

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Val Gly Gln Tyr Leu His Arg Gln Leu Ile Ala Asn Pro His Thr Arg

305 310 315 320

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

325 330 335

Lys Ile Trp Glu Thr Gln Ser Lys Phe Tyr Pro Lys Glu Leu Thr Thr

340 345 350

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

355 360 365

Leu Lys Ser Gln Lys Gly Leu Ile Ser Phe Cys Glu Phe Glu Gln Glu

370 375 380
 Gln Lys Ile Ile Asn Gly Lys Thr Lys Thr Ile Gly His Arg Val Ile

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

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

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

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

Tyr Ser Thr Ala Cys Glu Leu Ala Gly Tyr Lys His Ser Lys His Ser
 625 630 635 640
 Leu Thr Lys Glu Glu Asn Glu Lys Arg Glu Leu Lys Ser Arg Leu Thr

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

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

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

 835 840 845
 Gly Lys Asp Glu Gly Ile Ser Lys Ser Lys Tyr Lys Lys Leu Leu Met
 850 855 860
 Gln Glu Ser Glu Ile Gly Lys Gly Phe Ile Glu Arg Asp Leu Arg Glu

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

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

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

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

 1085 1090 1095
 Leu Val Ser Asn Pro Thr Tyr Arg Asp Leu Leu Leu Lys Arg Leu
 1100 1105 1110

Gln Glu Asn Asn Asn Asp Pro Lys Lys Ala Phe Gly Gly Lys Asn		
1115	1120	1125
Ala Leu Ser Lys Ser Pro Ile Tyr Ile Asn Leu Glu Lys Asn Ile		
1130	1135	1140
Ile Val Pro Glu Val Val Lys Leu Val Trp Leu Glu Asp Asp Tyr		
1145	1150	1155
Thr Ile Arg Lys Glu Ile Ser Pro Glu Leu Lys Ile Asp Lys Val		
1160	1165	1170
Val Asp Glu Gly Val Lys Arg Ile Leu Gln Asn Arg Leu Ala Leu		
1175	1180	1185
Tyr Asn Gly Asp Ala Lys Ile Ala Phe Ser Asp Leu Asp Asn Asn		
1190	1195	1200
Pro Ile Trp Leu Ser Lys Asp Lys Gly Ile Thr Ile Lys Arg Val		
1205	1210	1215
Thr Leu Ser Gly Val Lys Asn Ala Thr Ala Leu His Thr Lys Lys		
1220	1225	1230
Asp His Leu Gly Gln Pro Ile Leu Asp His Lys Gly Gln Glu Ile		
1235	1240	1245
Pro Val Asp Phe Ile Ser Thr Gly Asn Asn His His Val Ala Ile		
1250	1255	1260
Tyr Glu Asp Lys Asn Gly Lys Leu Gln Glu Ser Val Val Ser His		
1265	1270	1275
Phe Glu Ala Val Glu Arg Val Asn Gln Gln Leu Pro Ile Ile Asp		
1280	1285	1290
Lys Thr Phe Asn Gln His Leu Gly Trp Lys Phe Leu Phe Thr Met		
1295	1300	1305
Lys Gln Asn Glu Leu Phe Leu Phe Pro Ser Asp Asp Phe Asn Pro		
1310	1315	1320
Asn Glu Ile Asp Leu Phe Asp Thr Arg Asn Tyr Ala Val Ile Ser		
1325	1330	1335
Lys His Leu Phe Arg Ala Gln Lys Leu Ser Thr Lys Asp Tyr Ser		

1340 1345 1350
Phe Arg His His Leu Glu Thr Asn Val Glu Asp Ile Lys Asp Leu
1355 1360 1365
Lys Thr Ile Thr Trp Arg Arg Glu Gly Leu Asn Gly Ile Gln Asn
1370 1375 1380
Ile Val Lys Ile Arg Thr Asn His Leu Gly Glu Ile Val His Ile

1385 1390 1395
Gly Glu Tyr
1400
<210> 104
<211> 21
<212> RNA
<213> Myroides sp.
<400> 104
guugugaauu gcuucaaac a 21
<210> 105
<211> 105
<212> RNA
<213> Myroides sp.
<400> 105
ugaaagcaau ucacaauaag gauuauuccg uugugaaaac aucuaguccu cgacuaauuu 60
cggggaauaa agccuuaacu cuguuaaggc uuuuuuuagu uuuuu 105
<210> 106
<211> 159
<212> RNA
<213> Artificial Sequence

<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"
<220><221> modified_base
<222> (1)..(29)
<223> a, c, u, g, unknown or other
<400> 106

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugugaaug cuuuccaaaca aaagugaaag 60
caauucacaa uaaggauuau uccguuguga aaacaucuag uccucgacua auuucgggga 120
auaaaagccuu aacucuguua aggcuuuuuu uaguuuuuu 159

<210> 107

<211> 160

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 107

gagcggacag cagcuuccua uaucucguac guugugaaau gcuuuccaaac aaaagugaaa 60
gcaauucaca auaaggauua uuccguugug aaaacaucua guccucgacu aauuucgggg 120
aauaaagccu uaacucuguu aaggcuuuuu uuaguuuuuu 160

<210> 108

<211> 160

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 108

ccaugauaua gacguugugg cuguuguagu guugugaaau gcuuuccaaac aaaagugaaa 60

gcaauucaca auaaggauua uuccguugug aaaacaucua guccucgacu aauuucgggg 120
aauaaagccu uaacucuguu aaggcuuuuu uuaguuuuuu 160

<210> 109

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(4)

<223> a, c, t, g, unknown or other

<400> 109

nnnnc

5

<210> 110

<211> 1432

<212> PRT

<213> Chryseobacterium sp.

<400> 110

Met Ala Lys Asn Ile Leu Gly Leu Asp Leu Gly Thr Asn Ser Ile Gly

1 5 10 15

Trp Ala Leu Ile Asn Gln Asp Phe Glu Asn Lys Gln Gly Lys Ile Leu

20 25 30

Gly Met Gly Ser Arg Ile Ile Pro Met Ser Gln Asp Ile Leu Gly Asp

35 40 45

Phe Gly Lys Gly Asn Ser Val Ser Gln Thr Ala Ala Arg Thr Gly Tyr

50 55 60

Arg Gly Val Arg Arg Leu Arg Glu Arg Phe Leu Leu Arg Arg Glu Arg

65 70 75 80

Leu His Arg Val Leu Asn Ile Ile Asn Phe Leu Pro Glu His Tyr Ala

85 90 95

Ser Gln Ile Asp Phe Glu Lys Arg Phe Gly Lys Phe Lys Glu Glu Thr

100 105 110

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

115 120 125

Phe Glu Glu Met Leu Ala Asp Phe Lys Asn Tyr Gln Pro Gln Leu Leu

130 135 140

Glu Asn Asp Lys Lys Ile Pro Tyr Asp Trp Thr Ile Tyr Tyr Leu Arg

145 150 155 160

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

165 170 175

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

180 185 190

Asp Phe Glu Glu Glu Lys Asp Lys Met Phe Val Arg Leu Lys Val Glu

195 200 205

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

210 215 220

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

225 230 235 240

Glu Asp Trp Ile Glu Arg Val Lys Glu Phe Ile Val Thr Glu Ser Phe

245 250 255

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

260 265 270

Lys Asp Trp Ile Ala Ile Lys Thr Lys Thr Glu Gln Glu Ile Asp Lys

275 280 285

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

290 295 300

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

305 310 315 320

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

325 330 335

His Gln Glu Leu Lys Asn Asp Asp Leu Tyr Asn Asp Cys Val Arg Glu

340 345 350

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

355 360 365

Phe Val His Leu Leu Met Glu Asp Ile Ile Phe Tyr Gln Arg Pro Leu

370 375 380

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

385 390 395 400

Tyr Lys Asp Glu Asn Gly Val Glu His Ile Gln Tyr Leu Lys Ala Val

405 410 415

Pro Lys Ser Asn Pro Tyr Tyr Gln Glu Phe Arg Ile Trp Gln Trp Ile

420 425 430
 Phe Asn Leu Asn Ile Tyr Lys Arg Asp Asp Glu Gln Asn Pro Val Thr

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

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

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

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

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

915	920	925	
Gln Leu Asn Asp Thr Arg His Ile Ser Lys Tyr Ile Ser Gly Val Leu			
930	935	940	
Ser Asn Ile Val Arg Val Glu Asp Gly Ser Asp Glu Gly Val Asn Ser			
945	950	955	960
Lys Asn Ile Val Pro Gly Asn Gly Lys Ile Thr Thr Gln Leu Lys Gln			
965	970	975	
Asp Trp Gly Leu Asn Asp Val Trp Asn Asp Leu Ile Leu Pro Arg Phe			
980	985	990	
Glu Arg Met Asn Gln Leu Thr Asn Ser Thr Asp Phe Thr Ala Trp Asn			
995	1000	1005	
Glu Asn His Gln Lys Phe Leu Pro Thr Val Pro Ile Glu Phe Ser			
1010	1015	1020	
Lys Gly Phe Ser Lys Lys Arg Ile Asp His Arg His His Ala Leu			
1025	1030	1035	
Asp Ala Leu Val Ile Ala Cys Ala Thr Lys Asp His Ile Asn Leu			
1040	1045	1050	
Leu Asn Asn Gln Ser Ala Lys Ser Asp Thr Lys Arg Tyr Asp Leu			
1055	1060	1065	
Lys Lys Lys Leu Met Lys Phe Glu Lys Gly Val Tyr Asn His Pro			
1070	1075	1080	
Gln Thr Gly Glu Arg Ile Gln Arg Asp Val Pro Lys Gln Phe Leu			
1085	1090	1095	
Lys Pro Trp Glu Ser Phe Thr Ile Asp Ala Lys Asn Asn Leu Asp			
1100	1105	1110	
Lys Ile Ile Ile Ser Phe Lys Gln Asn Leu Arg Val Ile Asn Lys			
1115	1120	1125	
Ala Thr Asn Tyr Tyr Glu Lys Tyr Val Glu Lys Asn Gly Ile Lys			
1130	1135	1140	
Thr Lys Glu Arg Val Glu Gln Thr Gly Thr Asn Trp Ala Ile Arg			
1145	1150	1155	

Lys Pro Met His Lys Glu Thr	Val Ser Gly Ile Val	Asn Leu Pro
1160	1165	1170
Trp Val Lys Val Pro Lys Gly	Lys Ile Leu Thr Ala	Thr Arg Lys
1175	1180	1185
Ser Leu Asp Thr Thr Phe Asp	Leu Lys Ser Ile Asn	Ser Ile Thr
1190	1195	1200
Asp Thr Gly Ile Gln Lys Ile	Leu Arg Asn Tyr Leu	Glu Phe Lys
1205	1210	1215
Gly Ser Pro Glu Leu Ala Phe	Ser Pro Glu Gly Ile	Glu Asp Met
1220	1225	1230
Asn Lys Asn Ile Glu Lys Tyr	Asn Asp Gly Lys Leu	His Gln Pro
1235	1240	1245
Ile Asn Lys Val Arg Val Phe	Glu Leu Gly Ser Lys	Phe Gln Val
1250	1255	1260
Gly Gln Thr Gly Asn Lys Lys	Asp Lys Tyr Val Glu	Ala Ala Lys
1265	1270	1275
Gly Thr Asn Leu Phe Phe Ala	Val Tyr Glu Asp Lys	Asn Gly Lys
1280	1285	1290
Arg Asn Tyr Glu Thr Ile Pro	Leu Asn Ile Val Ile	Glu Arg Gln
1295	1300	1305
Lys Gln Gly Leu Leu Ala Cys	Pro Glu Ser Asn Glu	Lys Gly Glu
1310	1315	1320
Lys Leu Leu Phe Gln Leu Ser	Pro Asn Asp Phe Val	Tyr Leu Thr
1325	1330	1335
Thr Glu Glu Glu Asn Asp Asn	Ala Thr Pro Ile Asn	Phe Ser Leu
1340	1345	1350
Leu Ser Lys Glu Gln Ile Asn	Asn Leu Tyr Lys Ile	Val Ser Phe
1355	1360	1365
Thr Gly Asn Arg Leu Tyr Gly	Ile Pro Ile Cys Val	Ala Thr Thr
1370	1375	1380
Ile Val Asn Lys Ala Glu Tyr	Thr Gln Leu Asn Lys	Ile Glu Phe

1385 1390 1395
 Thr Lys Glu Lys Asp Leu Leu Leu Lys Leu Asn Val Asn Arg Leu
 1400 1405 1410
 Gly Asp Val Lys Thr Phe Thr Ala Asn Asp Ile Arg Lys Ile Phe
 1415 1420 1425
 Asn Arg Gln Glu
 1430
 <210> 111
 <211
 > 16
 <212> RNA
 <213> Chryseobacterium sp.
 <400> 111
 guugugaaau gcuuuc 16
 <210> 112
 <211> 73
 <212> RNA
 <213> Chryseobacterium sp.
 <400> 112
 gaaagcaauu cacaauaagg auuauuccgu ugugaaaaca uuuagcgccu cgacuaucuu 60
 cggggcauuu uuu 73
 <210> 113
 <211> 122
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <220><221> modified_base
 <222> (1)..(29)
 <223> a, c, u, g, unknown or other
 <400> 113
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnng uugugaaauug cuuucaaagg aaagcaauuc 60
 acaauaagga uuauuccguu gugaaaacau uuagcgccuc gacuaucuuu ggggcauuuu 120

uu 122

<210> 114

<211> 123

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 114

gagcggacag cagcuuccua uaucucguac guugugaaau gcuucaaag gaaagcaauu 60

cacaauaagg auuauuccgu ugugaaaaca uuugcgccu cgacuaucuu cggggcauuu 120

uuu 123

<210> 115

<211> 123

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 115

ccaugauua gacguugugg cuguuguagu guugugaaau gcuucaaag gaaagcaauu 60

cacaauaagg auuauuccgu ugugaaaaca uuugcgccu cgacuaucuu cggggcauuu 120

uuu 123

<210> 116

<211> 6

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(5)

<223> a, c, t, g, unknown or other

<400> 116

nnnnnc

6

<210> 117

<211> 1092

<212> PRT

<213> Bacillus sp.

<400> 117

Met Arg Tyr Ser Ile Gly Leu Asp Ile Gly Thr Thr Ser Ile Gly Asn

1 5 10 15

Ala Val Ile Asn Lys Asp Leu Gln Arg Phe Glu His Leu Gly Val Arg

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

Asn Glu Phe Ala Arg Val Leu Ile His Phe Ser Lys Asn Arg Gly Phe

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

Lys Arg Asn Lys Gly Gly Asp Tyr Ser His Val Val Ala Arg Ser Asp

180 185 190

Ile Glu Asn Glu Ile Cys Leu Leu Phe Gln Lys Gln Arg Glu Tyr Gly

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

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

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

 Asn Met Ala Asn Arg Glu Tyr Asn Leu Glu Leu Ile Glu Glu Leu Leu
 405 410 415
 Gly Leu Ser Phe Ala Lys Phe Gly His Leu Ser Leu Lys Ala Leu Asn
 420 425 430
 Asn Ile Leu Pro Tyr Met Glu Glu Gly Glu Pro Tyr His Ile Ala Cys
 435 440 445

Glu Met Ala Ser Tyr Gln Phe Ser Gln Arg Leu Ser Lys Glu Lys His
450 455 460

Arg Leu Leu Pro Pro Ile Pro Val Asp Glu Ile Pro Asn Pro Val Val
465 470 475 480

Val Arg Ala Leu Thr Gln Val Arg Lys Val Leu Asn Ser Ile Ile Lys
485 490 495

Lys Tyr Gly Pro Pro Ser Asp Ile Tyr Ile Glu Leu Ala Arg Glu Met
500 505 510

Ser Lys Pro Phe Lys Glu Arg Lys Ser Leu Glu Arg Glu Phe Asn Glu
515 520 525

Asn Arg Gln Ile Asn Glu Lys Ala Lys Ala His Ile Ser Glu Leu Tyr
530 535 540

Arg Ile Pro Asn Asp Pro Arg Pro His Asp Ile Leu Lys Phe Lys Leu
545 550 555 560

Trp Asn Glu Gln Asn Gly Ile Cys Pro Tyr Ser Leu Lys Pro Ile Ser
565 570 575

Ile Glu Tyr Leu Phe Asn Ile Gly Tyr Ala Glu Val Asp His Ile Ile
580 585 590

Pro Tyr Ser Arg Ser Phe Asp Asp Ser Asn Gly Asn Lys Val Leu Val
595 600 605

Leu Thr Arg Glu Asn Gln Asn Lys Leu Asn Arg Thr Pro Tyr Glu Trp
610 615 620

Phe Gly His Glu Glu Asn Arg Trp Glu Asp Phe Val Ser Phe Ile Arg
625 630 635 640

Thr Met Lys Val Gly Lys Lys Lys Lys Asn Met Leu Leu Lys Lys Asn
645 650 655

Phe Asp Glu Glu Gln Glu Glu Gln Ile Leu Ser Arg Asn Leu Asn Asp
660 665 670

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

Glu Phe Arg Thr Glu Glu Asn Lys Glu Gln Tyr Val His Thr Val Asn

690					695					700									
Gly	Ala	Tyr	Thr	Ser	Leu	Met	Arg	Lys	Arg	Trp	Gly	Leu	Asn	Lys	Asp				
705					710					715					720				
Arg	Arg	Gly	Asn	Asp	Leu	His	His	Ala	Val	Asp	Ala	Ala	Ile	Ile	Ala				
725					730					735									
Val	Ser	Leu	Pro	Phe	Lys	Asn	Lys	Val	Asn	Ala	Tyr	Phe	Lys	Arg	Gln				
740					745					750									
Glu	Thr	Gly	Leu	Ser	Lys	Leu	Leu	Asn	Asn	Lys	Lys	Asp	Ile	Phe	Pro				
755					760					765									
Glu	Pro	Trp	Arg	Asn	Phe	Ile	Lys	Glu	Leu	Glu	Ala	Arg	Met	Ile	Gln				
770					775					780									
Asp	Pro	Glu	Lys	Met	Lys	Arg	Ala	Leu	Glu	Ser	Leu	Glu	Leu	Glu	Thr				
785					790					795					800				
Tyr	Gly	Glu	Ile	Phe	Leu	Asn	Lys	Leu	Lys	Pro	Ile	Phe	Val	Ser	Arg				
805					810					815									
Met	Pro	Lys	His	Ser	Ile	Lys	Gly	Gln	Ile	His	Glu	Glu	Thr	Ile	Arg				
820					825					830									
Arg	Val	Arg	Gly	Phe	Thr	Glu	Glu	Gly	Phe	Leu	Val	Thr	Val	Lys	Lys				
835					840					845									
Thr	Arg	Leu	Asp	Gln	Ile	Pro	Phe	Asp	Lys	Asn	Gly	Asp	Phe	Pro	Met				
850					855					860									
Tyr	Gly	Lys	Glu	Thr	Asp	Ile	Lys	Thr	Tyr	Met	Ala	Ile	Lys	Gln	Arg				
865					870					875					880				
Tyr	Leu	Glu	Tyr	Gly	Gln	Asp	Lys	Gln	Lys	Ala	Phe	Ala	Val	Pro	Leu				
885					890					895									
Arg	Lys	Pro	Ser	Lys	Asn	Pro	Lys	Asn	Ala	Pro	Ile	Val	Arg	Ser	Val				
900					905					910									
Lys	Ile	Glu	Gly	Lys	Ala	Asn	Arg	Val	Val	Met	Leu	Asp	Asp	Lys	Ala				
915					920					925									
Ala	Ala	Asp	Asn	Ala	Ser	Ile	Val	Arg	Thr	Glu	Val	Phe	Arg	His	Lys				
930					935					940									

Lys Thr Gly Glu Tyr Tyr Leu Thr Pro Val Tyr Val Ala Asp Ile Leu
 945 950 955 960
 Ser Asn Lys Ile Pro Asp Arg Leu Ile Thr Ile Lys Lys Ser Tyr Ser
 965 970 975

Asp Trp Asp Arg Ile Thr Asp Glu His Glu Tyr Leu Phe Ser Leu Tyr
 980 985 990
 Asn Asn Asp Leu Val Lys Ile Ile Leu Pro Lys Glu Lys Glu Thr Lys
 995 1000 1005
 Lys Tyr Thr Gly Gly Asn His Leu Trp Gln Glu Gly Phe Phe Tyr
 1010 1015 1020
 Phe Lys Gly Val Asp Ser Ser Asn Ala Gly Ile Lys Ile Ile Asn
 1025 1030 1035

His Leu Asn Ser Phe Glu Ala Arg Ile Gly Thr Lys Arg Leu Ile
 1040 1045 1050
 Ala Phe Glu Lys Tyr Gln Val Asn Pro Leu Gly Glu Ile Asn Lys
 1055 1060 1065
 Val His Gly Glu Lys Arg Pro Gly Glu Leu Leu Asn Lys Glu Glu
 1070 1075 1080
 Ile Lys Glu Asn Arg Lys Asn Ile Ser
 1085 1090

<210> 118

<211> 23

<212> RNA

<213>

> Bacillus sp.

<400> 118

gccauaaauuc cucuguaaaa cuu

23

<210> 119

<211> 112

<212> RNA

<213> Bacillus sp.

<400> 119

aagguuuuaua gaguuuuuau gguaaggcaa uaugccgugg cguuggggau cgccuauguc

60

cgguuuuacc ggaucucccu aaaggugacu aacuuugguu agucaccuuu uu 112

<210> 120

<211> 168

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (1)..(29)

<223> a, c, u, g, unknown or other

<400> 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnng ccuaauucc ucuguaaaac uuaaagaagg 60

uuuauagagu uauuauggua aggcaauaug ccguggcguu ggggaucgcc uauguccggu 120

uuuaccggau cucccuaaag gugacuaacu uugguuaguc accuuuuu 168

<210> 121

<211> 169

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 121

gagcggacag cagcuuccua uaucucguac gccauaauc cucuguaaaa cuuaaagaag 60

guuuauagag uuauuauggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg 120

uuuuaccgga ucuccuaaa ggugacuaac uuugguuagu caccuuuuu 169

<210> 122

<211> 169

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 122
 ccaugauaua gacguugugg cuguuguagu gccauaaauuc cucuguaaaa cuuaaagaag 60
 guuuauagag uuauuauggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg 120
 uuuuaccgga ucucccuaaa ggugacuaac uuugguuagu caccuuuuu 169

<210> 123

<211> 4

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 123

aaag 4

<210> 124

<211> 13

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 124

gccgccrcca tgg 13

<210> 125

<211> 7

<212> PRT

<213> Artificial Sequence

<220><221> source

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

<400> 125

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 126

<211> 22

<212> PRT

<213> Artificial Sequence

<220><221> source

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

<400> 126

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

1 5 10 15

Lys Asp Asp Asp Asp Lys

20

<210> 127

<211> 16

<212> PRT

<213> Artificial Sequence

<220><221> source

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

<400> 127

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 128

<211> 1092

<212> PRT

<213> Artificial Sequence

<220><221> source

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

<400> 128

Met Arg Tyr Ser Ile Gly Leu Ala Ile Gly Thr Thr Ser Ile Gly Asn

1 5 10 15

Ala Val Ile Asn Lys Asp Leu Gln Arg Phe Glu His Leu Gly Val Arg

20 25 30

Ile Phe Asp Ala Ala Glu Asn Pro Lys Asp Gly Ser Ser Leu Ser Ala
35 40 45

Pro Arg Arg Leu Ala Arg Ser Ser Arg Arg Arg Leu Arg Arg Arg Lys
50 55 60

His Arg Val Glu Arg Thr Lys Gln Leu Leu Ile Asn Lys Gly Leu Leu
65 70 75 80

Thr Lys Thr Gln Val Lys Asn Leu Tyr Asn Ser Lys Asn Ile Asn Leu
85 90 95

Asp Ile Trp Asp Ile Arg Val Ser Gly Ile Asp Arg Lys Leu Phe Asn
100 105 110

Asn Glu Phe Ala Arg Val Leu Ile His Phe Ser Lys Asn Arg Gly Phe
115 120 125

Lys Ser Asn Arg Lys Ser Glu Leu Lys Glu Asp Asp Asn Gly Ala Ile
130 135 140

Leu Ser Ala Val Lys Glu Asn Arg Glu Leu Met Asp Glu Lys Gly Tyr
145 150 155 160

Arg Thr Ile Ala Glu Met Leu Val Ser Asp Glu Lys Tyr Glu Gly Thr
165 170 175

Lys Arg Asn Lys Gly Gly Asp Tyr Ser His Val Val Ala Arg Ser Asp
180 185 190

Ile Glu Asn Glu Ile Cys Leu Leu Phe Gln Lys Gln Arg Glu Tyr Gly
195 200 205

His Pro Phe Ala Thr Glu Glu Asn Glu Glu Ala Phe Leu Ser Ile Trp
210 215 220

Ser Ser Gln Arg Pro Phe Ser Thr Lys Asp Asp Ile Val Lys Lys Ile
225 230 235 240

Gly Asn Cys Thr Leu Glu Pro Lys Glu Lys Arg Ala Pro Lys Ser Thr
245 250 255

Tyr Thr Phe Glu Arg Phe Arg Ala Leu Asp Lys Leu Asn Arg Leu Arg
260 265 270

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

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

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

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

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

 Ser Lys Pro Phe Lys Glu Arg Lys Ser Leu Glu Arg Glu Phe Asn Glu
 515 520 525

Asn Arg Gln Ile Asn Glu Lys Ala Lys Ala His Ile Ser Glu Leu Tyr
 530 535 540
 Arg Ile Pro Asn Asp Pro Arg Pro His Asp Ile Leu Lys Phe Lys Leu
 545 550 555 560
 Trp Asn Glu Gln Asn Gly Ile Cys Pro Tyr Ser Leu Lys Pro Ile Ser
 565 570 575

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

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

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

 Glu Pro Trp Arg Asn Phe Ile Lys Glu Leu Glu Ala Arg Met Ile Gln

770	775	780	
Asp Pro Glu Lys Met Lys Arg Ala Leu Glu Ser Leu Glu Leu Glu Thr			
785	790	795	800
Tyr Gly Glu Ile Phe Leu Asn Lys Leu Lys Pro Ile Phe Val Ser Arg			
	805	810	815
Met Pro Lys His Ser Ile Lys Gly Gln Ile His Glu Glu Thr Ile Arg			
	820	825	830
Arg Val Arg Gly Phe Thr Glu Glu Gly Phe Leu Val Thr Val Lys Lys			
	835	840	845
Thr Arg Leu Asp Gln Ile Pro Phe Asp Lys Asn Gly Asp Phe Pro Met			
	850	855	860
Tyr Gly Lys Glu Thr Asp Ile Lys Thr Tyr Met Ala Ile Lys Gln Arg			
865	870	875	880
Tyr Leu Glu Tyr Gly Gln Asp Lys Gln Lys Ala Phe Ala Val Pro Leu			
	885	890	895
Arg Lys Pro Ser Lys Asn Pro Lys Asn Ala Pro Ile Val Arg Ser Val			
	900	905	910
Lys Ile Glu Gly Lys Ala Asn Arg Val Val Met Leu Asp Asp Lys Ala			
	915	920	925
Ala Ala Asp Asn Ala Ser Ile Val Arg Thr Glu Val Phe Arg His Lys			
	930	935	940
Lys Thr Gly Glu Tyr Tyr Leu Thr Pro Val Tyr Val Ala Asp Ile Leu			
945	950	955	960
Ser Asn Lys Ile Pro Asp Arg Leu Ile Thr Ile Lys Lys Ser Tyr Ser			
	965	970	975
Asp Trp Asp Arg Ile Thr Asp Glu His Glu Tyr Leu Phe Ser Leu Tyr			
	980	985	990
Asn Asn Asp Leu Val Lys Ile Ile Leu Pro Lys Glu Lys Glu Thr Lys			
	995	1000	1005
Lys Tyr Thr Gly Gly Asn His Leu Trp Gln Glu Gly Phe Phe Tyr			
1010	1015	1020	

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

Ile Lys Glu Asn Arg Lys Asn Ile Ser
 1085 1090

<210> 129

<211> 230

<212> PRT

<213> Unknown

<220><221> source

<223> /note="Description of Unknown:

APG09980 deaminase sequence"

<400> 129

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

Asn Arg His Val Asn Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Trp
115 120 125
Tyr Gln Gly Tyr Glu Ala Gly Leu Arg Ala Leu Lys Ala Ala Gly Ala
130 135 140
Glu Val Ala Met Met Thr Leu His Glu Phe Glu Tyr Cys Trp Asn Asn
145 150 155 160
Phe Val Asp His Gln Gln Asp Glu Asp Thr Pro Phe Pro Pro Trp Asp
165 170 175
Asn Leu Val Ala Arg Ser Glu Glu Leu Ser Gln Arg Leu Glu Gly Ile
180 185 190
Leu Gln Pro Ser Val Leu Val Phe Cys Trp Pro Ser Gln Val Ser Val
195 200 205
Thr Ala Ala His Ser Asp Ile Met Ser Gln Ala Ser Arg Ala Trp Glu
210 215 220
Lys Arg Arg Asp Pro Pro
225 230

<210> 130

<211> 182

<212> PRT

<213> Unknown

<220><221> source

<223> /note="Description of Unknown:

APG07386 CTD sequence"

<400> 130

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

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

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

Leu Gln Glu Ile Leu Gln
180

<210> 131

<211> 201

<212> PRT

<213> Unknown

<220><221> source

<223> /note="Description of Unknown:

APG05840 deaminase sequence"

<400> 131

Met Glu Ala Ser Pro Ala Ser Arg Pro Arg Pro Leu Met Asp Pro His
1 5 10 15
Met Phe Thr Gly Asn Phe Thr Asn Asn Pro Arg Val Phe Gly Leu His
20 25 30
Gln Thr Tyr Leu Cys Tyr Glu Val Lys Arg Gln Gly Pro Asp Gly Thr
35 40 45
Arg Asp Leu Met Asn Glu Gln Arg Asp Phe Leu Cys Asn Gln Ala Lys
50 55 60

Asn His Phe Ser Gly Ser Glu Asp His His Ala Glu Arg Cys Phe Leu
 65 70 75 80
 Asp Arg Ile Pro Ser Trp Gln Leu Asp Pro Ala Gln Thr Tyr Arg Val
 85 90 95
 Thr Cys Phe Ile Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Val

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

 165 170 175
 Pro Trp Glu Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu
 180 185 190
 Gln Ala Ile Leu Gln Asn Gln Gly Asn
 195 200
 <210> 132
 <211> 192
 <212> PRT
 <213> Unknown
 <220><221> source
 <223> /note="Description of Unknown:
 APG00868 deaminase sequence"
 <400> 132
 Met Glu Pro Trp Arg Pro Ser Pro Arg Asn Pro Met Asp Arg Ile Asp
 1 5 10 15

 Pro Asn Thr Phe Arg Phe His Phe Pro Asn Leu Leu Tyr Ala Ser Gly
 20 25 30
 Arg Lys Leu Cys Tyr Leu Cys Phe Gln Val Glu Thr Gly Asp Tyr Phe
 35 40 45

Ser Cys Asp Asp Ser Asp Arg Gly Val Phe Arg Asn Lys Val His Pro
50 55 60

Trp Ala Arg Cys His Ala Glu Gln Cys Phe Leu Ser Trp Phe Arg Asp
65 70 75 80

Gln Tyr Pro Cys Arg Asp Glu Tyr Tyr Asn Val Thr Trp Phe Leu Ser
85 90 95

Trp Ser Pro Cys Pro Thr Cys Ala Glu Glu Val Val Glu Phe Leu Glu
100 105 110

Glu Tyr Arg Asn Leu Thr Leu Ser Ile Phe Thr Ser Arg Leu Tyr Tyr
115 120 125

Phe Tyr His Pro Asn Tyr Gln Gln Gly Leu Arg Lys Leu Trp Asp Ala
130 135 140

Gly Val Gln Leu Asp Ile Met Ser Cys Asp Asp Phe Glu His Cys Trp
145 150 155 160

Asp Asn Phe Val Asp His Lys Gly Met Arg Phe Gln Arg Arg Asn Leu
165 170 175

Leu Lys Asp Tyr Asp Phe Leu Ala Ala Glu Leu Gln Glu Ile Leu Arg
180 185 190

<210> 133

<211> 32

<212> PRT

<213> Artificial Sequence

<220><221> source

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

polypeptide"

<400> 133

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30

<210> 134

<211> 159

<212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 134
 ggccattaa aacctctcca gccataattc ctctgtaaaa cttaaagaag gtttatagag 60
 ttattatggt aaggcaatat gccgtggcgt tggggatcgc ctatgtccgg ttttaccgga 120
 tctccctaaa ggtgactaac tttggttagt caccttttt 159
 <210> 135
 <211> 159
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 135
 catggcagta cattagagca gccataattc ctctgtaaaa cttaaagaag gtttatagag 60
 ttattatggt aaggcaatat gccgtggcgt tggggatcgc ctatgtccgg ttttaccgga 120
 tctccctaaa ggtgactaac tttggttagt caccttttt 159
 <210> 136
 <211> 164
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 136
 agctgttttg gaggtcagaa atagggccat aattcctctg taaaacttaa agaaggttta 60
 tagagttatt atggttaaggc aatatgccgt ggcgttgggg atgcctatg tccggtttta 120
 ccgatctcc ctaaaggatga ctaacttttg ttagtcacct tttt 164
 <210> 137
 <211> 1151

<212> PRT

<213> Brevibacillus sp.

<400> 137

Met Ala Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys

1 5 10 15
Lys Lys Ala Ile Arg Ser Ile Lys Leu Lys Leu Lys Thr Arg Thr Gly

20 25 30
Pro Glu Ala Gln Asn Leu Arg Lys Gly Ile Trp Arg Thr His Arg Leu

35 40 45
Leu Asn Glu Gly Val Ala Tyr Tyr Met Lys Met Leu Leu Leu Phe Arg

50 55 60
Gln Glu Ser Thr Gly Gly Gln Thr Lys Lys Glu Leu Gln Glu Glu Leu

65 70 75 80
Val Arg His Ile Arg Glu Gln Gln Gln Lys Asn Arg Ala Asp Lys Asn

85 90 95
Thr Gln Ala Leu Pro Leu Asp Lys Ala Phe Ala Ala Leu Arg Gln Leu

100 105 110
Tyr Glu Leu Leu Val Pro Ser Ser Ile Gly Gln Ser Gly Asp Ala Gln

115 120 125
Ile Ile Ser Arg Lys Phe Leu Ser Pro Leu Val Asp Pro Asn Ser Glu

130 135 140
Gly Gly Lys Gly Thr Ser Lys Ala Gly Ala Lys Pro Thr Trp Gln Lys

145 150 155 160
Lys Lys Glu Ala Asn Asp Pro Thr Trp Glu Gln Asp Tyr Glu Lys Trp

165 170 175
Lys Lys Arg Arg Glu Glu Asp Pro Thr Ala Ser Val Ile Thr Thr Leu

180 185 190
Glu Glu Tyr Gly Ile Arg Pro Ile Phe Pro Leu Tyr Thr Asn Thr Val

195 200 205
Ala Asp Ile Ala Trp Leu Pro Leu Gln Ser Asn Gln Phe Val Arg Thr

210 215 220

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

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

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

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

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

 705 710 715 720

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

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

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

1085 1090 1095
Ser Thr Gln Lys Asp Phe Trp Ser Glu Val Lys Arg Arg Leu Tyr
1100 1105 1110
Gly Lys Leu Arg Glu Arg Phe Leu Met Lys Thr Arg Pro Lys Lys
1115 1120 1125
Lys Arg Lys Val Glu Asn Leu Tyr Phe Gln Ser Gly Ser His His
1130 1135 1140
His His His His His His His
1145 1150

<210> 138

<211> 11

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 138
aaaggaatcc c 11
<210> 139
<211> 22
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 139
ttcgattccg gagaggagc ct 22
<210> 140
<211> 50
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 140
aggagcggac agcagcttcc tatatctcgt acagggaact caacatggtt 50
<210> 141
<211> 50
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 141
agccatgata tagacgttgt ggctgttgta gtagggaact caacatggtt 50
<210> 142
<211> 50
<212> DNA
<213> Artificial Sequence

<220><221> source

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

<400> 142

aggagcggac agcagcttcc tatactctgt acaggaccgt caacatggtt 50

<210> 143

<211> 50

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 143

agccatgata tagacgttgt ggctgttgta gtaggaccgt caacatggtt 50

<210> 144

<211> 142

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 144

ggcgacgggu agaggccgua ugucgauuug cuuuuuuuuc gugcgugugc auugcgucc 60

uccauuacag ggcggcuacc acgaauagcc acgaaguaaa agcuucgugg cuagcacgua 120

cgagauauag gaagcugcug uc 142

<210> 145

<211> 137

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 145

ggcgacgguu agaggccgua ugucgauuug cuuuuuuuuc gugcgugugc auugucgucc 60
uccauuacag ggcggcuacc acgaauagcc acgaaguaaa agcuucgugg cuagcacacu 120
acaacagcca caacguc 137

<210> 146

<211> 4

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 146

cttc

4

<210> 147

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 147

atatg

5

<210> 148

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 148

tattg

5

<210> 149

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 149

cgttc 5

<210> 150

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 150

ttttt 5

<210> 151

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 151

caatc 5

<210> 152

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 152

aatct 5

<210> 153
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 153
 aaatt 5
 <210> 154
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 154
 catcc 5

 <210> 155
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 155
 agata 5
 <210> 156
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"

<400> 156

tattc

5

<210> 157

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 157

atddd

5

<210> 158

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 158

tcttt

5

<210> 159

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 159

agatc

5

<210> 160

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 160

aataa 5

<210> 161

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 161

gatcc 5

<210> 162

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 162

attca 5

<210> 163

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 163

aaaat 5

<210> 164

<211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 164
 ttatc 5

<210> 165
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 165
 tatca 5

<210> 166
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 166
 tcact 5

<210> 167
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"

<400> 167
 cttat 5
 <210> 168
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 168
 tgata 5
 <210> 169
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 169
 tacta 5
 <210> 170
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 170
 atata 5
 <210> 171
 <211> 5
 <212> DNA
 <213> Artificial Sequence

<220><221> source

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

<400> 171

gactc 5

<210> 172

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 172

gctca 5

<210> 173

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 173

taacc 5

<210> 174

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 174

cgtgt 5

<210> 175

<211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 175
 ttctcg 5
 <210> 176
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 176
 cagta 5
 <210> 177
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 177
 ccctt 5
 <210> 178
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 178

cctat	5
<210> 179	
<211> 5	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
<400> 179	
tcctt	5
<210> 180	
<211> 5	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
<400> 180	
aggtt	5
<210> 181	
<211> 5	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
<400> 181	
taacg	5
<210> 182	
<211> 5	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	

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

<400> 182

aacag

5

<210> 183

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 183

tgaac

5

<210> 184

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 184

taagt

5

<210> 185

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 185

ctgta

5

<210> 186

<211> 5

<212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 186
 agggg

5

<210> 187
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 187
 gctgt

5

<210> 188
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 188
 atcta

5

<210> 189
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 189

tcagg	5
<210> 190	
<211> 5	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
<400> 190	
atgtc	5
<210> 191	
<211> 5	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
<400> 191	
tgcta	5
<210> 192	
<211> 5	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
<400> 192	
taaaa	5
<210> 193	
<211> 5	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	

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

<400> 193

atcag

5

<210> 194

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 194

cgggg

5

<210> 195

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 195

ttctc

5

<210> 196

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 196

gtcat

5

<210> 197

<211> 5

<212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"

<400> 197

gcgtg

5

<210> 198

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 198

gcaat

5

<210> 199

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 199

ttctc

5

<210> 200

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 200

agcat 5

<210> 201
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 201

gtcaa 5

<210> 202
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 202

agagc 5

<210> 203
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 203

gtggg 5

<210> 204
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source

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

<400> 204

ctgaa

5

<210> 205

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 205

ctgat

5

<210> 206

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 206

atgaa

5

<210> 207

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 207

acctt

5

<210> 208

<211> 5

<212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 208
 aacca

5

<210> 209
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 209
 gtaaa

5

<210> 210
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 210
 tagga

5

<210> 211
 <211> 5
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 211

ccgaa 5

<210> 212

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 212

tacga 5

<210> 213

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 213

tggt 5

<210> 214

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 214

ccaag 5

<210> 215

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 215

tgctt

5

<210> 216

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 216

ctaaa

5

<210> 217

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 217

tcgat

5

<210> 218

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 218

gtaac

5

<210> 219

<211> 5

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 219

gtaac 5

<210> 220

<211> 500

<212> DNA

<213> Homo sapiens

<400> 220

aaaaatggac aattatgagg aggggagagt gcagacaggg gaagcttcac ctctttaca 60

atTTTgggag tccacacggc atggcataca aattatttca ttccattga gaaataaaat 120

ccaattctcc atcaccaaga gagccttccg aaagaggccc ccctgggcaa acggccaccg 180

atggagaggt ctgccagtc ttttctaccc caccacgcc cccaccctaa tcagaggcca 240

aacccttctt ggagcctgtg ataaaagcaa ctgttagctt gactagact agcttcaaag 300

ttgtattgac cctgggtgtg tatgtctaag agtagatgcc atatctcttt tctggcctat 360

gttattacct gtatggactt tgcactggaa tcagctatct gctcttactt atgcacacct 420

ggggcataga gccagcctg tatcgctttt cagccatctc actacagata actcccaagt 480

cctgtctagc tgccttcctt 500

<210> 221

<211> 20

<212> DNA

<213> Homo sapiens

<400> 221

tattgaccct ggtgtgttat 20

<210> 222

<211> 20

<212> DNA

<213> Homo sapiens

<400> 222

gcttgcacta gactagcttc 20

<210> 223

<211> 20

<212> DNA

<213> Homo sapiens

<400> 223

cttttatcac aggctccagg 20

<210> 224

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 224

cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cggcuaccac gaauagccac gaaguaaaag cuucguggcu agcacuauug 120

accugggugu guuau 135

<210> 225

<211> 1977

<212> DNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 225

caagcatgag gagggaaagta ggcagaatcc tctggagcca aagctctgga tgtctctccc 60

ctctgacat ggagcccacc cctgtctccac tgctccaggg acagccctat gctgcaggca 120

gtctctcccc cactcagcat cccaggggct gatttctttg gttttggatc cagctggatg 180

tctgcattgc cgaggccacc agggctggct cagcaactgt cggggaatca ccagggtctg 240

agaaatcttg tgcgcatgtg aggggctgtg ggagcagaga accactgggt gggaaattct 300

aateccccacc ctgctggaat ctctctgggt ggccccaaca tgctaattct cgggcaaacc 360

tctgtttcct cctcaaaagg caggaggtcg gaaagaataa acaatgagag tcacattaaa 420

aacacaaaat cctacggaaa tactgaagaa tgagtctcag cactaaggaa aagcctccag 480

 cagctcctgc tttctgaggg tgaaggatag acgctgtggc tctgcatgac tcactagcac 540
 tctatcacgg ccatattctg gcagggtcag tggctccaac taacatttgt ttggtacttt 600
 acagtttatt aaatagatgt ttatatggag aagctctcat ttttttctca gaagagcctg 660
 gctaggaagg tggatgaggg accatattca ttttcaggt gaaattcctg agatgtaagg 720
 agctgctgtg acttgctcaa ggccttataat cgagtaaagc gtagtgctgg ggcttagacg 780
 caggtgttct gatattatgt tcaaaacctc tatcaatgag agagcaatct cctggtaatg 840
 tgatagattt cccaacttaa tgccaacata ccataaacct cccattctgc taatgcccag 900

 cctaagttgg ggagaccact ccagattcca agatgtacag tttgctttgc tgggcctttt 960
 tccccgcct gcctttactc tgccagagtt atattgctgg ggttttgaag aagatcctat 1020
 taaataaaag aataagcagt attattaagt agccctgcat ttcaggtttc cttgagtggc 1080
 aggccaggcc tggccgtgaa cgttcactga aatcatggcc tcttggccaa gattgatagc 1140
 ttgtgcctgt cctgagtc cagtccatca cgagcagctg gtttctaaga tgctatttcc 1200
 cgtataaagc atgagaccgt gacttgccag cccacagag ccccgccctt gtccatcact 1260
 ggcatctgga ctccagcctg ggttggggca aagagggaaa tgagatcatg tcctaaccct 1320

 gatcctcttg tcccacagat atccagaacc ctgaccctgc cgtgtaccag ctgagagact 1380
 ctaaatccag tgacaagtct gtctgcctat tcaccgattt tgattctcaa acaaagtgtg 1440
 cacaaagtaa ggattctgat gtgtatatca cagacaaaac tgtgctagac atgaggtcta 1500
 tggacttcaa gagcaacagt gctgtggcct ggagcaacaa atctgacttt gcatgtgcaa 1560
 acgccttcaa caacagcatt attccagaag acaccttctt cccagccca ggtaagggca 1620
 gctttggtgc cttgcagggc tgtttccttg cttcaggaat ggccaggttc tgcccagagc 1680
 tctggtaaat gatgtctaaa actcctctga ttggtggtct cggccttata cattgccacc 1740

 aaaaccctct tttactaag aaacagtgag ccttgttctg gcagtcaga gaatgacacg 1800
 ggaaaaaagc agatgaagag aaggtggcag gagagggcac gtggcccagc ctgagtctct 1860
 ccaactgagt tctgcctgc ctgcctttgc tcagactggt tgccccttac tgctcttcta 1920
 ggccctattc taagcccctt ctccaagttg cctctcctta tttctcctg tctgcca 1977

 <210> 226
 <211> 566
 <212> DNA
 <213> Artificial Sequence
 <220><221> source

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

<400> 226

tcgtcggcag cgtcagatgt gtataagaga cagtttgggt tttgccagac tccacagtgc 60

atacgtgggc tccaacaggt cctcttcct cccagtcact gactaaccac ggaaccacac 120

agcttcccgt tctcagctcc acaaacttgg tgccaaattc ttctcccctg ggaagcatec 180

ctggacactt ccaaaggac cccagtcact ccagcctgtt ggctgccgct cactttgatg 240

tctgcaggcc agatgagggc tccagatggc acattgtcag agggacacac tgtggcccct 300

gtgccagcc ctgggtcttc tgtacatgaa gcaactccag tcccaaatat gtagctgttt 360

gggaggtcag aaataggggg tccaggagca aactccccc acccccttcc caaagcccat 420

tcctcttta gccagagccg ggggtgtgcag acggcagtca ctagggggcg ctcgccacc 480

acagggaagc tgggtgaatg gagcgagcag cgtcttcgag agtgaggacg tgctgtctct 540

tatacacatc tccgagccca cgagac 566

<210> 227

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 227

gaaagggggt ggggggagtt tgctc 25

<210> 228

<211> 53

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 228

tcgtcggcag cgtcagatgt gtataagaga cagtttgggt tttgccagac tcc 53

<210> 229

<211> 54

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 229

gtctcgtggg ctcgagatg tgtataagag acagcctctg acaatgtgcc atct 54

<210> 230

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 230

cccgtataaa gcatgagacc 20

<210> 231

<211> 21

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 231

caagcatgag gaggaagta g 21

<210> 232

<211> 21

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 232

tggcagacag ggagaaataa g

21

<210> 233

<211> 137

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 233

ggcgacgguu agaggccgua ugucgauuug cuuuuuuuuc gugcgugugc auugucgucc

60

uccauuacag ggcggcuacc acgaauagcc acgaaguaaa agcuucgugg cuagcacgaa

120

agggggguggg gggaguu

137

<210> 234

<211> 137

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 234

ggcgacgguu agaggccgua ugucgauuug cuuuuuuuuc gugcgugugc auugucgucc

60

uccauuacag ggcggcuacc acgaauagcc acgaaguaaa agcuucgugg cuagcacccc

120

guauaaagca ugagacc

137

<210> 235

<211> 1071

<212> PRT

<213> Bacillus sp.

<400> 235

Met Arg Glu Leu Asp Tyr Arg Ile Gly Leu Asp Ile Gly Thr Asn Ser

1

5

10

15

Ile Gly Trp Gly Val Ile Glu Leu Ser Trp Asn Lys Asp Arg Glu Arg

20

25

30

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

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

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

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

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

 515 520 525

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

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

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

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

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

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

1025

1030

1035

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

1040

1045

1050

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

1055

1060

1065

Asn Phe His

1070

<210> 236

<211> 46

<212> RNA

<213> Bacillus sp.

<400> 236

gacagcagcu uccuauaucu cguacgucau aguuccauua aagcca

46

<210> 237

<211> 85

<212> RNA

<213> Bacillus sp.

<400> 237

uggcuuugau guuucuauga uaaggguuuc gacccguggc gucggggauc gccugcccau

60

ugaaaugggc uucuccccau uuauu

85

<210> 238

<211> 32

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 238

gacagcagcu uccuauaucu cguacgucau ag

32

<210> 239

<211> 71

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 239

cuaugauaag gguuucgacc cgugcgucg gggaucgccu gcccaugaa augggcuucu 60

ccccauuuau u 71

<210> 240

<211> 36

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 240

gacagcagcu uccuauaucu cguacgucau agccau 36

<210> 241

<211> 75

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 241

augucuauga uaaggguuuc gacccguggc gucggggauc gccugcccau ugaaaugggc 60

uucuccccau uuauu 75

<210> 242

<211> 40

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 242

gacagcagcu uccuauaucu cguacgucau aguuccauua

40

<210> 243

<211> 79

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 243

ugauguuucu augauaaggg uuucgacccg uggcgucggg gaucgccugc ccuugaaaau

60

gggcuucucc ccuuuuuuu

79

<210> 244

<211> 46

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 244

gacagcagcu uccuauaucu cguacgucau aguuccauua aagcca

46

<210> 245

<211> 78

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 245

uggcuuugau guuucuauga uaaggguuuc gacccguggc gucggggauc gccugcccau

60

ugaaaugggc uucucccc

78

<210> 246

<211> 46

<212> RNA

<213> Artificial Sequence

<

220><221> source

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

<400> 246

gacagcagcu uccuauaucu cguacgucau aguuccauua aagcca 46

<210> 247

<211> 75

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 247

uggcuuugau guuucuauga uaaggguuuc gacccguggc gucggggacc gcccauugaa 60

augggccucc ccuuu 75

<210> 248

<211> 46

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 248

gacagcagcu uccuauaucu cguacgucau aguuccauua aagcca 46

<210> 249

<211> 71

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 249
 uggcuuugau guuucuauga uaaggguuuc gacccguggc gucgggaccg ccuugaaaau 60

ggccuccuu u 71

<210> 250
 <211> 38
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 250
 gacagcagcu uccuauaucu cguacgucau aguuccau 38

<210> 251
 <211> 77
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 251

auguuucua u gauaaggguu ucgacccgug gcgucgggga ucgccugccc auugaaaugg 60
 gcuucucccc auuuauu 77

<210> 252
 <211> 38
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 252
 gacagcagcu uccuauaucu cguacgucau aguuccau 38

<210> 253
 <211> 70

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 253

auguuucuau gauaagggguu ucgacccgug gcgucgggga ucgccugccc auugaaaugg 60

gcuucucucc 70

<210> 254

<211> 46

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 254

gacagcagcu uccuauaucu cguacgucau aguuccauua aagcca 46

<210> 255

<211> 76

<212> RNA

<213> Artificial Sequence

<

220><221> source

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

<400> 255

uggcuuugau guuucuauga uaagggguuc gacccguggc gucggggauc gccugcccau 60

ugaaaugggc uucucc 76

<210> 256

<211> 39

<212> RNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<400> 256

gcuuccuaua ucucguacgu cauaguucca uuaaagcca 39

<210> 257

<211> 85

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 257

uggcuuugau guuucuauga uaaggguuuc gacccguggc gucggggauc gccugcccau 60

ugaaaugggc uucucccca uuuuu 85

<210> 258

<211> 38

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 258

gacagcagcu uccuauaucu cguacgucau aguuccau 38

<210> 259

<211> 71

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 259

uggcuuugau guuucuauga uaaggguuuc gacccguggc gucgggacgc ccauugaaaau 60

ggccucccuu u 71

<210> 260

<211> 682

<212> DNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 260

ttgtctcatg agcgataca tatttgaatg tatthagaaa aataaacaac taggggttcc	60
gcgcacattt ccccgaaaag tgccagatac ctgaaacaaa acccatcgta cggccaagga	120
agtctccaat aactgtgac caccacaagc gccagggttt tcccagtcac gacgttgtaa	180
aacgacggcc agtcatgcat aatccgcacg catctggaat aaggaagtgc cattccgcct	240
gacctgaaga ccaggctcga aagccttcac accctcagta agagctttct caatgtcgat	300
tgtgccacac accctgtcct cagtagcacc tagcggtaaa tcaaccatgt tgacggctcct	360
gtacgagata taggaagctg ctgtccgctc ctcaatttgt cacggacttc tgcactcatt	420

acttcttggg catctggac tgaattgaac ggatcaccag aaataacttt gatttcagga	480
agtaaactca ccaagatct aaccgtgggt gacttcccgg ttcctctatc acccattatc	540
atcacaggct aggtggaggg tcagtatga taagtctgcg atggtggatg catgtgtcat	600
ggtcatagct gtttctgtg tgaattgtt atccgctcag agggcacaat cctattccgc	660
gctatccgac aatctcaag ac	682

<210> 261

<211> 82

<212> PRT

<213> Artificial Sequence

<220><221> source

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

polypeptide"

<400> 261

Thr	Asn	Leu	Ser	Asp	His	Glu	Lys	Glu	Thr	Gly	Lys	Gln	Leu	Val	Ile
1						5				10					15
Gln	Glu	Ser	Ile	Leu	Met	Leu	Pro	Glu	Glu	Val	Glu	Glu	Val	Ile	Gly
						20				25					30
Asn	Lys	Pro	Glu	Ser	Asp	Ile	Leu	Val	His	Thr	Ala	Tyr	Asp	Glu	Ser

35 40 45
Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys
50 55 60

Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys
65 70 75 80
Met Leu

<210> 262

<211> 1072

<212> PRT

<213> Artificial Sequence

<220><221> source

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

<400> 262

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

20 25 30
Tyr Glu Lys Val Arg Ile Val Asp Lys Gly Val Arg Met Phe Asp Lys
35 40 45

Ala Glu Ile Pro Asn Lys Gly Ala Ser Leu Ala Glu Pro Arg Arg Ile
50 55 60

Ala Arg Ser Ser Arg Arg Arg Leu Asn Arg Lys Ser Gln Arg Lys Lys
65 70 75 80
Glu Ile Arg Asn Leu Leu Val Gln His Gly Met Ile Thr Gln Glu Glu

85 90 95
Leu Asp Leu Leu Tyr Pro Leu Ser Lys Lys Ser Ile Asp Ile Trp Asp
100 105 110

Ile Arg Leu Asp Gly Leu Asp Arg Leu Leu Asn His Leu Glu Trp Ala
115 120 125

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

130 135 140
 Lys Ser Glu Leu Lys Asp Ala Glu Thr Gly Lys Val Leu Ser Ser Ile

 145 150 155 160
 Gln Val Asn Glu Lys Arg Leu Phe Leu Tyr Arg Thr Val Gly Glu Met

 165 170 175
 Trp Ile Lys Asp Ala Glu Phe Ser Lys Tyr Asp Arg Arg Arg Asn Ser

 180 185 190
 Pro Asn Glu Tyr Val Phe Ser Val Ser Arg Ala Asp Leu Glu Lys Glu

 195 200 205
 Ile Val Thr Leu Phe Glu Ala Gln Arg Lys Phe Gln Ser Ser Tyr Ala

 210 215 220
 Ser Lys Asn Leu Gln Glu Thr Tyr Leu Gln Ile Trp Ala His Gln Leu

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

 245 250 255
 Leu Leu Lys Gly Lys Glu Arg Arg Ile Pro Lys Ala Thr Tyr Thr Phe

 260 265 270
 Gln Tyr Phe Ser Ala Leu Asp Gln Val Asn Arg Thr Arg Leu Gly Pro

 275 280 285
 Asp Phe Gln Pro Phe Thr Gln Glu Gln Lys Glu Ile Ile Leu Asp Lys

 290 295 300
 Met Phe Gln Arg Thr Asp Tyr Tyr Lys Lys Lys Thr Ile Pro Glu Val

 305 310 315 320
 Ser Tyr Tyr Asp Ile Arg Lys Trp Leu Glu Leu Asp Glu Thr Ile Gln

 325 330 335
 Phe Lys Gly Leu Asn Tyr Asp Pro Asn Glu Glu Leu Lys Lys Ile Glu

 340 345 350
 Lys Lys Pro Phe Ile Asn Leu Lys Ala Phe Tyr Glu Ile Lys Lys Val

 355 360 365
 Val Ala Asn Tyr Ala Glu Arg Thr Asn Glu Ala Phe Ser Thr Leu Asp

 370 375 380

Tyr Asp Ala Ile Ala Tyr Ala Leu Thr Val Tyr Lys Thr Asp Lys Asp
 385 390 395 400
 Ile Arg Ser Tyr Leu Lys Lys Ser Asn Asn Leu Ser Lys Arg Cys Tyr

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

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

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

 595 600 605
 Ser Gln Ser Phe Ile Asp Ser Tyr His Asn Lys Val Leu Val Tyr Ser
 610 615 620
 Asp Glu Asn Arg Asn Lys Gly Asn Arg Ile Pro Tyr Thr Tyr Phe Leu

625 630 635 640
 Glu Thr Asn Lys Asp Trp Glu Ala Phe Glu Arg Tyr Val Arg Ser Asn
 645 650 655
 Lys Leu Phe Ser Lys Lys Lys Arg Glu Tyr Leu Leu Lys Lys Thr Tyr

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

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

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

 850 855 860
 Glu Asn Gly Asp Phe Lys Met Val Gly Lys Glu Gln Asp Leu Ala Thr
 865 870 875 880

Tyr Glu Ala Ile Lys Gln Arg Tyr Leu Glu His Arg Lys Asn Ser Lys
 885 890 895
 Lys Ala Phe Glu Thr Pro Leu Tyr Lys Pro Ser Lys Lys Gly Thr Gly
 900 905 910
 Asn Leu Ile Lys Arg Val Lys Ile Glu Gly Gln Thr Lys Ala Phe Val
 915 920 925
 Arg Glu Val Asn Gly Gly Val Ala Gln Asn Ser Asp Leu Val Arg Val
 930 935 940
 Asp Leu Phe Glu Lys Asp Asp Lys Tyr Tyr Met Val Pro Ile Tyr Val
 945 950 955 960
 Pro Asp Thr Val Cys Ser Glu Leu Pro Lys Lys Val Val Lys Ser Gly
 965 970 975
 Lys Gly Tyr Glu Gln Trp Leu Thr Leu Asp Asn Ser Phe Thr Phe Lys
 980 985 990
 Ser Ser Leu Tyr Pro Tyr Asp Leu Val Arg Leu Val Lys Gly Asn Glu
 995 1000 1005
 Asp Arg Phe Leu Tyr Phe Gly Thr Leu Asp Ile Asp Ser Asp Arg
 1010 1015 1020
 Leu Asn Phe Lys Asp Val Asn Lys Pro Ser Lys Gln Asn Glu Tyr
 1025 1030 1035
 Arg Tyr Ser Leu Lys Thr Ile Glu Asn Leu Glu Lys Tyr Glu Val
 1040 1045 1050
 Gly Val Leu Gly Asp Leu Arg Leu Val Lys Gln Glu Thr Arg Arg
 1055 1060 1065
 Ile Phe Asn Arg
 1070
 <210> 263
 <211> 130
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic

polynucleotide"

<400> 263

catggcagta cattagagca gtcatagttc cattaaagcc aaaagtggct ttgatgtttc 60

tatgataagg gtttcgacct gtggcgctcg ggatcgctg cccattgaaa tgggcttctc 120

cccatttatt 130

<210> 264

<211> 130

<212> DNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 264

cacatctcga gcaagacgtt gtcatagttc cattaaagcc aaaagtggct ttgatgtttc 60

tatgataagg gtttcgacct gtggcgctcg ggatcgctg cccattgaaa tgggcttctc 120

cccatttatt 130

<210> 265

<211> 130

<212> DNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 265

cttctatagc ctcttcccc gtcatagttc cattaaagcc aaaagtggct ttgatgtttc 60

tatgataagg gtttcgacct gtggcgctcg ggatcgctg cccattgaaa tgggcttctc 120

cccatttatt 130

<210> 266

<211> 50

<212> DNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<400> 266

caggtcgact ctagaggatc ccatgatata gacgttgtgg ctgtttagt 50

<210> 267

<211> 53

<212> DNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<220><221> modified_base

<222> (19)..(23)

<223> a, c, t, g, unknown or other

<400> 267

aacgacggcc agtgaattnn nnnactacaa cagccacaac gtctatatca tgg 53

<210> 268

<211> 1493

<212> PRT

<213> Homo sapiens

<400> 268

Met Pro Asn Glu Gly Ile Pro His Ser Ser Gln Thr Gln Glu Gln Asp

1 5 10 15

Cys Leu Gln Ser Gln Pro Val Ser Asn Asn Glu Glu Met Ala Ile Lys

20 25 30

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

35 40 45

Ser Val Gly Asp Gly Leu Ser Thr Ser Ala Val Gly Cys Ala Ser Ala

50 55 60

Ala Pro Arg Arg Gly Pro Ala Leu Leu His Ile Asp Arg His Gln Ile

65 70 75 80

Gln Ala Val Glu Pro Ser Ala Gln Ala Leu Glu Leu Gln Gly Leu Gly

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

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

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

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

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

 Val Met His Gln Trp Val Lys Glu Phe His Thr Trp Trp Pro Pro Phe

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

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

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

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

Gly Tyr Trp Lys Arg Ser Gly Lys Met Ile Val Val Glu Ser Leu Leu
835 840 845

Lys Ile Trp His Lys Gln Gly Gln Arg Val Leu Leu Phe Ser Gln Ser
850 855 860

Arg Gln Met Leu Asp Ile Leu Glu Val Phe Leu Arg Ala Gln Lys Tyr
865 870 875 880

Thr Tyr Leu Lys Met Asp Gly Thr Thr Thr Ile Ala Ser Arg Gln Pro
885 890 895

Leu Ile Thr Arg Tyr Asn Glu Asp Thr Ser Ile Phe Val Phe Leu Leu
900 905 910

Thr Thr Arg Val Gly Gly Leu Gly Val Asn Leu Thr Gly Ala Asn Arg
915 920 925

Val Val Ile Tyr Asp Pro Asp Trp Asn Pro Ser Thr Asp Thr Gln Ala
930 935 940

Arg Glu Arg Ala Trp Arg Ile Gly Gln Lys Lys Gln Val Thr Val Tyr
945 950 955 960

Arg Leu Leu Thr Ala Gly Thr Ile Glu Glu Lys Ile Tyr His Arg Gln
965 970 975

Ile Phe Lys Gln Phe Leu Thr Asn Arg Val Leu Lys Asp Pro Lys Gln
980 985 990

Arg Arg Phe Phe Lys Ser Asn Asp Leu Tyr Glu Leu Phe Thr Leu Thr
995 1000 1005

Ser Pro Asp Ala Ser Gln Ser Thr Glu Thr Ser Ala Ile Phe Ala
1010 1015 1020

Gly Thr Gly Ser Asp Val Gln Thr Pro Lys Cys His Leu Lys Arg
1025 1030 1035

Arg Ile Gln Pro Ala Phe Gly Ala Asp His Asp Val Pro Lys Arg
1040 1045 1050

Lys Lys Phe Pro Ala Ser Asn Ile Ser Val Asn Asp Ala Thr Ser
1055 1060 1065

Ser Glu Glu Lys Ser Glu Ala Lys Gly Ala Glu Val Asn Ala Val

1070	1075	1080
Thr Ser Asn Arg Ser Asp Pro	Leu Lys Asp Asp Pro	His Met Ser
1085	1090	1095
Ser Asn Val Thr Ser Asn Asp	Arg Leu Gly Glu Glu	Thr Asn Ala
1100	1105	1110
Val Ser Gly Pro Glu Glu Leu	Ser Val Ile Ser Gly	Asn Gly Glu
1115	1120	1125
Cys Ser Asn Ser Ser Gly Thr	Gly Lys Thr Ser Met	Pro Ser Gly
1130	1135	1140
Asp Glu Ser Ile Asp Glu Lys	Leu Gly Leu Ser Tyr	Lys Arg Glu
1145	1150	1155
Arg Pro Ser Gln Ala Gln Thr	Glu Ala Phe Trp Glu	Asn Lys Gln
1160	1165	1170
Met Glu Asn Asn Phe Tyr Lys	His Lys Ser Lys Thr	Lys His His
1175	1180	1185
Ser Val Ala Glu Glu Glu Thr	Leu Glu Lys His Leu	Arg Pro Lys
1190	1195	1200
Gln Lys Pro Lys Asn Ser Lys	His Cys Arg Asp Ala	Lys Phe Glu
1205	1210	1215
Gly Thr Arg Ile Pro His Leu	Val Lys Lys Arg Arg	Tyr Gln Lys
1220	1225	1230
Gln Asp Ser Glu Asn Lys Ser	Glu Ala Lys Glu Gln	Ser Asn Asp
1235	1240	1245
Asp Tyr Val Leu Glu Lys Leu	Phe Lys Lys Ser Val	Gly Val His
1250	1255	1260
Ser Val Met Lys His Asp Ala	Ile Met Asp Gly Ala	Ser Pro Asp
1265	1270	1275
Tyr Val Leu Val Glu Ala Glu	Ala Asn Arg Val Ala	Gln Asp Ala
1280	1285	1290
Leu Lys Ala Leu Arg Leu Ser	Arg Gln Arg Cys Leu	Gly Ala Val
1295	1300	1305

Ser Gly Val Pro Thr Trp Thr Gly His Arg Gly Ile Ser Gly Ala
1310 1315 1320

Pro Ala Gly Lys Lys Ser Arg Phe Gly Lys Lys Arg Asn Ser Asn
1325 1330 1335

Phe Ser Val Gln His Pro Ser Ser Thr Ser Pro Thr Glu Lys Cys
1340 1345 1350

Gln Asp Gly Ile Met Lys Lys Glu Gly Lys Asp Asn Val Pro Glu
1355 1360 1365

His Phe Ser Gly Arg Ala Glu Asp Ala Asp Ser Ser Ser Gly Pro
1370 1375 1380

Leu Ala Ser Ser Ser Leu Leu Ala Lys Met Arg Ala Arg Asn His
1385 1390 1395

Leu Ile Leu Pro Glu Arg Leu Glu Ser Glu Ser Gly His Leu Gln
1400 1405 1410

Glu Ala Ser Ala Leu Leu Pro Thr Thr Glu His Asp Asp Leu Leu
1415 1420 1425

Val Glu Met Arg Asn Phe Ile Ala Phe Gln Ala His Thr Asp Gly
1430 1435 1440

Gln Ala Ser Thr Arg Glu Ile Leu Gln Glu Phe Glu Ser Lys Leu
1445 1450 1455

Ser Ala Ser Gln Ser Cys Val Phe Arg Glu Leu Leu Arg Asn Leu
1460 1465 1470

Cys Thr Phe His Arg Thr Ser Gly Gly Glu Gly Ile Trp Lys Leu
1475 1480 1485

Lys Pro Glu Tyr Cys
1490

<210> 269

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221

> source

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

<400> 269

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagccac gaaguuuag cuucguggcu agcacgcuug	120
cacuagacua gcuuc	135

<210> 270

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400>

270

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagccac gaaguuuag cuucguggcu agcaccuuuu	120
aucacaggcu ccagg	135

<210> 271

<211> 25

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 271

gacagcagcu uccuauaucu cguac	25

<210> 272

<211>

18

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 272	
gcuuccuaua ucucguac	18
<210> 273	
<211> 22	
<212> RNA	
<213> Brevibacillus sp.	
<400> 273	
uggaaagcuu cgagguuagc ac	22
<210> 274	
<211> 95	
<212> RNA	
<213> Brevibacillus sp.	
<400> 274	
cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cgguaccac gaauagccac gaagu	95
<210> 275	
<211> 151	
<212> RNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"	
<220><221> modified_base	
<222> (132)..(151)	
<223> a, c, u, g, unknown or other	
<400> 275	
cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cgguaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc	120
gagguuagca cnnnnnnnnn nnnnnnnnnn n	151
<210> 276	
<211> 6	
<212> PRT	
<213> Artificial Sequence	

<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
6xHis tag"

<400> 276

His His His His His His

1 5

<210> 277

<211> 10

<212> PRT

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
10xHis tag"

<400> 277

His His His His His His His His His His

1 5 10

<210> 278

<211> 25

<212>

> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 278

cccgtataaa gcatgagacc gtgac

25

<210> 279

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 279

ctcagactgt ttgccctta ctgct

25

<210> 280

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 280

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagccac gaaguuuag cuucguggcu agcaccgccu	120
auaaagcaug agacc	135

<210> 281

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 281

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagccac gaaguuuag cuucguggcu agcaccucag	120
acuguuugcc ccuaa	135

<210> 282

<211> 156

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 282

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguuuagu ggaaagcuuc	120

gagguuagca ccccguaaua agcaugagac cgugac 156

<210> 283

<211> 156

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 283

cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cgguaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc 120

gagguuagca ccucagacug uuugcccuu acugcu 156

<210> 284

<211> 4

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 284

aaaa 4

<210> 285

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (116)..(135)

<223> a, c, u, g, unknown or other

<400>

> 285

cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cggcuaccac gaauagccac gaaguaaaag cuucguggcu agcacnnnnn 120
 nnnnnnnnnn nnnnn 135
 <210> 286
 <211> 95
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 286
 cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60
 cauuacaggg cggcuaccac gaauagccac gaagu 95
 <210> 287
 <211> 16
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 287
 gcuucguggc uagcac 16
 <210> 288
 <211> 151
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <220><221> modified_base
 <222> (132)..(151)
 <223> a, c, u, g, unknown or other
 <
 400> 288
 cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cgguaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc 120

gagguuagca cnnnnnnnnn nnnnnnnnnn n 151

<210> 289

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (116)..(135)

<223> a, c, u, g, unknown or other

<400> 289

cgacgguuag aggccguaug ucgaauugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cgguaccac gaauagucac gaaguaaaag cuucgagguu agcacnnnnn 120

nnnnnnnnnn nnnnn 135

<210> 290

<211> 144

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (125)..(144)

<223> a, c, u, g, unknown or other

<400> 290

cgacgguuag aggccguaug ucgaauugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cgguaccac gaauagccac gaaguuccaa aauggaaagc uucguggcua 120

gcacnnnnnn nnnnnnnnnn nnnn 144

<210> 291

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<220><221> modified_base

<222> (116)..(135)

<223> a, c, u, g, unknown or other

<400> 291

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cgguaccac gaauagccac gaaguaaaag cuucguggcu agcacnnnnn 120

nnnnnnnnnn nnnnn 135

<210> 292

<211> 30

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 292

gtacgagata taggaagctg ctgtccgctc 30

<210> 293

<211> 29

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 293

actacaacag ccacaacgtc tatatcatg 29

<210> 294

<211> 149

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 294

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc	120
gagguuagca cacuacaaca gccacaacg	149

<210> 295

<211> 151

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 295

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc	120
gagguuagca cguacgagau auaggaagcu g	151

<210> 296

<211> 151

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 296

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc	120
gagguuagca cacuacaaca gccacaacgu c	151

<210> 297

<211> 156

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 297

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc	120
gagguuagca cguacgagau auaggaagcu gcuguc	156

<210> 298

<211> 156

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 298

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagucac gaaguuccac ugaguaaagu ggaaagcuuc	120
gagguuagca cacuacaaca gccacaacgu cuauau	156

<210> 299

<211> 140

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 299

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagucac gaaguaaaag cuucgagguu agcacacuac	120
aacagccaca acgucuauau	140

<210> 300

<211> 149

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 300

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cggcuaccac gaauagccac gaaguuccaa augggaaagc uucguggcua 120

gcacacuaca acagccacaa cgucuauau 149

<210> 301

<211> 140

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 301

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cggcuaccac gaauagccac gaaguaaaag cuucguggcu agcacacuac 120

aacagccaca acgucuauau 140

<210> 302

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 302

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cggcuaccac gaauagccac gaaguaaaag cuucguggcu agcacacuac 120

aacagccaca acguc 135

<210> 303

<211> 140

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 303

cgacgguaug aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60
cauuacaggg cgguuaccac gaauagccac gaaguaaaag cuucguggcu agcacguacg 120
agauauagga agcugcuguc 140

<210> 304

<211> 3324

<212> DNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 304

atggccattc ggagcatcaa gctcaagctg aagaccagaa ccggtcctga ggctcagaac 60

ctgcggaagg gcatctggag aaccacccgg ctgctgaacg agggcgtggc ctactacatg 120
aagatgctgc tgctcttccg gcaggaatcc actggcggcc agacaaagaa ggaactgcag 180
gaggaactgg tgcggcacat ccgggagcag cagcagaaaa acagagccga caagaatacc 240
caggccctgc ctctggataa ggccttcgcc gctctgagac aactgtacga gctgctcgtg 300
ccatctagca tcggccagag cggcgatgcc cagatcatct ctagaaaatt cctgtctcct 360
ttggtggacc ctaacagcga gggcggcaag ggactagca aggccggagc caagcccaca 420
tggcagaaaa agaaggaggc caatgaccct acatgggagc aagattacga gaagtggaaa 480

aagcgtagag aggaagatcc taccgccagt gtgataacaa ccctggaaga atatggaatt 540
agacctatct ttccactgta caccaacacc gtggccgata tcgcttggt gcctctgcag 600
tctaaccagt ttgtccggac atgggatcgg gacatgctgc agcaggccat cgagaggcta 660
ctctcttggg agtcttggaa caagcgggtg caagaggaat acagcaagct gcaggagaag 720
atgacgcaac tgaacgagca actggagggc ggacaggagt ggatcagcct gctggaacaa 780
tacgaggaac agagagagca ggagctgata gaaaacatga ccgccgaaa cgataaatac 840
cgcatcacca agcggcagat gaagggtggt aatgagctgt atgagcagt gtctaccgtg 900

ctgcccaatg ccagccacga gcagtacaga gaggcactga aaagagtga acagcggtg 960

aggggcagat tcggcgacgc tcatttcttc cagtatctga tgaaagaaga acaccacctg 1020
 atctggaagg ggaacctca gagaatccac tacttcgtgg ccagaaacga gctgaagaaa 1080
 agactggaag aagccaagca gaacgccacc atgacctgc ctgatgccag aaagcacctt 1140
 ctgtgggtca gatttgacgc cagaggcggc aacctgcaag actactacct gacggccgaa 1200
 gccgacaacc ccagaagcag aaggttcgtg accttcagcc aactgatctg gcctaacgag 1260
 agcggctgga tgaaaaaaca ggacgtggaa gttgaactgg ccctgagcaa gcagttctac 1320

 cagcaagtga cactgcagaa aaacgacaag gggaagcagg aaatcgagtt caaagacaag 1380
 ggcatggca gcaccttcag cggccacctg ggccgggcca agctgcaact ggagagagga 1440
 gatctggaga aggaagaaaa agactttgaa ggccggcgaaa tcggcagcgt gtacctgaac 1500
 atcgtgatcg acttcgagcc actgcaggag gtgaagaacg gccgactgca atctccttac 1560
 ggccaggtgc tgcaactggt cagaaggcct aatgagttcc ccaaggtgac cacctacaag 1620
 tctgaggaac tggtcgagtg gatcaagagc agcaccaagg acagcgccgg cgtggagtcc 1680
 ttagagagcg gttttagagt gatgagcatc gacctgggac tgcggaccgc cgccgcaaca 1740

 agcatcttct ctgtagagga atccaacgac gccaacgccg cgggtttcag ctactggatc 1800
 gagggaaact ctctggttgc cgttcataag cggtcataata tgctgaaact gcccggagag 1860
 caagtcgaaa agcaggtgag agagaagcgg gacgagcggc aggatcagca gagaagagt 1920
 aggttccaaa tcagaatcct gagccaggtg atccggatgg ccaagaagca gaaccgggag 1980
 cgggctgatg agctggacca cctgtcccag gccttgagga agcaaaaatc tctgctggac 2040
 cagaccgatc ggaccttttg gaacggcatc gtgtgcgacc tgacagacgc tctgagagag 2100
 aaagaggcgc gatgggaaca ggccgtggtc cagatccaca ggaaggccga ggagcacgtg 2160

 ggcaaggtgg tgcaagcctg gcggaacgg ttcgacgccg atgaacgcaa gggcatcgcc 2220
 ggctgtctat tgtggtctat cgaggagctg gacagcctgc ggaagctgct gatctcttgg 2280
 agcagaagaa caagaaaccc cagagaaatc aactgcttcg agcagggcca caccagccac 2340
 cagcggtgac tgacacacat ccagaacgtg aaggaggacc ggctgaagca actgagccac 2400
 gccattgtga tgacagcctt gggctacgtg tacgacgaga agaaattgga atggtttgcc 2460
 aagtaccctg ctgttcaggt gatcctgttc gagaacctgt ccagtagcgc gtccaacatg 2520
 gacagaagca ccaagagaa tagcaccctg atgaaatggg cccacaggag catccctaag 2580

 tacgtgcaca tgcaggccga gccttacggc atccagatcg gcgatgtgag gcccgagtac 2640
 tccagcagat tccagccaa gacaggcaca cctggcatcc ggtgcaagat ggtgtccgga 2700
 cagcacctgc aaggcaggcg ctctgagaac ctgcagaagc ggttaatctc tgaacagttc 2760
 ctgacagagg agcaagtga gacgtcaga cccggcgaca tcgtgcccga cgactccggc 2820

gagtgtttca tgacctgag cgacggcagc gaaggcaaag aggttgtgtt cctccaagcc 2880
gacatcaacg ccgccccaaa tctgcaaaag agattctggc agcggtagaa cgagctgttt 2940
aaggtctcct gcagagtgt gatccgagga gaagaggaat acctgatccc caagacaaag 3000

tccgtgcaag ccaagctggg caaaggcctg ttcgtgaaaa aaaccgacac cgtgatgaag 3060
gacgtgtacg tgtgggacag ccaggctaag ctgaaggga aaaccacatt caccgaggag 3120
tccgaaagcc ctgagcaact ggaggatfff caggagatca tcgaagaagc cgaagaagct 3180
aagggcacat acagaacact gtttagagac ccagcggag tgttcttccc tgagtctgtg 3240
tgggtccccc agaaagattt ctggtccgag gtgaagagac ggctgtacgg caagctgaga 3300
gagcggttcc tgatgaagac cagg 3324

<210> 305

<211> 3324

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 305

atggccatca gatctatcaa gctgaagctc aagaccagaa caggctctga ggcccagaac 60
ctgcggaagg gcatttggcg gacacaccgc ctctcaacg agggcgctcg ctattatatg 120
aaaatgctgc tgctgttcag acaggagagc acaggcggcc agaccaagaa agagctgcag 180
gaagaactgg tccggcacat cagagagcag cagcagaaga atcgggcccga caagaacacc 240
caggccctgc ctctggacaa ggcttttgcc gctctcggc agctctacga gctgctggtg 300
ccttcttcta tcggccagag cggagacgcg cagatcatca gcagaaagtt cctgtcccct 360

ctcgtggacc ccaacagcga gggcggcaag ggcacatcta aagctggcgc caagcctaca 420
tggcagaaaa agaaggaggc caacgacccc acctgggaac aggattacga gaaatggaag 480
aagcggagag aggaggaccc taccgccagc gtgattacca ccctggaaga atacggcatc 540
agaccaatct tccactgta caccaacacc gtggccgaca tcgcttggt gcctctgcag 600
tccaaccagt tctgctggac atgggacaga gacatgctgc agcaggctat cgagagactg 660
ctgagctggg aaagctggaa caagagagt caagaggaat acagcaagct gcaagagaag 720
atgacacagc tcaacgagca actggagggc ggccaggagt ggatcagcct gctggaacag 780

tacgaggagc aacgggagca ggagctgac gagaacatga ccgccgctaa cgacaaatat 840

agaataacaa agcggcagat gaagggtctg aacgagctgt acgagcagtg gagcaccgtg 900
ctgccaatg cctctcacga gcagtaccgg gaagccctta agcgggtcca gcaaagactg 960
cggggcagat tcggcgacgc tcattttctt cagtatctga tgaaggaaga gcaccacctg 1020
atttgaagg gcaaccccca gagaatccac tactttgtgg ccagaaacga gctgaagaaa 1080
agactggaag aggccaagca gaacgccact atgacctgc cagacgccag aaagcaccac 1140
ctgtgggtgc ggttcgacgc cagaggcgga aatctgcagg actactacct gaccgccgag 1200

gccgataacc ccagatctag aagattcgtt accttttagcc agttgatctg gcctaacgag 1260
tccggctgga tggaaaagca ggacgtggaa gtggaactcg ccctgagcaa gcagttctac 1320
cagcaagtga ccctgcagaa gaacgataag gggaaacagg agatcgagtt taaggacaag 1380
ggctccggat ctacgttcag tggccatctg ggcggggcta agctgcaact ggagcgaggc 1440
gacctggaga aagaagagaa ggactttgag ggcggagaaa tcggaagcgt gtacctgaac 1500
atcgtgatcg acttcgagcc ctgagcaggag gtgaaaaacg gcagactgca gagccatac 1560
ggccaggtgc tgcagctcgt tcggagacct aatgagttcc ctaaggtgac cacatacaag 1620

tctgaagaac ttgttgagtg gatgaaggcc agccagaatc acagcagcgg cgtggagtct 1680
ctggagtcgg gcttcagagt gatgagcatc gatctgggac tgaggacagc cgctgccacc 1740
agcattttct ctgtggaaga aagcaacgat gccaacgccg ctggcttcag ctactggatc 1800
gagggcaccc ctctggtcgc cgtgcacaag agaagctaca tgctgaagct gccaggcgaa 1860
caagtggaaa aacaggtcgc ggaaaagaga gatgagagac aagaccagca gaggcgcgtc 1920
agatttcaga tcagaatcct gagccaggtg atcaggatgg ccaagaaaca aaacagagaa 1980
agagctgacg aactggacca cctgagccag gcactggaga agcagaagtc cctgctcgat 2040

cagaccgata gaaccttctg gaacggcatc gtttgtgacc tgaccgatgc gctgcgcgaa 2100
aaggaggag gctgggagca agccgtggtc caaatccaca gaaaggccga ggaacacgtg 2160
ggcaaggtgg tgcaagcctg gcggaaaaga tttagcgcg atgagcggaa gggcatcgcc 2220
ggcctgagca tgtggtccat agaagagctg gacagcctcc ggaagctgct cattagctgg 2280
agcagaagga caagaaaccc tcaggagatc aacagattcg agcagggccca cacctctcac 2340
cagcggtcgc tgacacatat ccagaacgtg aaggaagata gactgaagca actgagccac 2400
gccatctga tgaccgcctt gggctacgtg tacgacgaga agaagctgga gtggttcgcc 2460

aaataccccc cctgccaggt gatcctgttc gagaatctgt ctcagtacag aagccacatg 2520
gacagatcca cgaaggaaaa tagcaccctg atgaaatggg cccacagaag catccctaag 2580
tacgtccaca tgcaggccga gccttacgga atccagatcg gagatgtgag agccgaatat 2640
agcagcaggt tccacgcaa gacagggaca cctggcatcc gttgcaagat ggtgaaggga 2700

caagaactgc aaggcaagcg attcgagaac ctgcaaaaga gactggtgtc cgaacagttt 2760
ctgaccgagg aacaggtgaa gcagcttcgg cctggagata tcgtgccaga tgacagcgga 2820
gagtggttca tgaccctgag tgatggcagc gaaggcaagg aagtgggtgtt cctgcaagcc 2880

gacatcaacg ccgcccagaa ccttcagaaa cgattctggc agagatacaa cgaactgttc 2940
aaggtgtcat gcagagtgtc gatcagaggc gaggaagagt acctgatccc caaggccaag 3000
agcgtgcaag ccaaactcgg caagggactg ttcgtgaaga aaaccgacac cgtgatgaag 3060
gacgtgtacg tgtgggatag ccaggccaaa ctgaaggga aaacaacctt caccgaggaa 3120
agcgagagcc ctgagcaact ggaggacttc caggagatca tcgaagaagc cgaggaagcc 3180
aagggcacct acagaacact gtttagagat cctagcggcg ttttcttccc cgagtctgtg 3240
tggaatacc agaaagactt ctggtccgag gtgaagagaa ggctgtacgg caagctgcgc 3300

gagagattcc tgatgaagac ccgg 3324

<210> 306

<211> 203

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 306

tgatgcggtt ttggcagtac atcaatgggc gtggatagcg gtttgactca cggggatttc 60
caagtctcca cccattgac gtcaatggga gtttgtttg gcacaaaaat caacgggact 120
ttccaaaatg tcgtaacaac tccgccccat tgacgcaaat gggcggtagg cgtgtacggt 180
gggaggtcta tataagcaga gct 203

<210> 307

<211> 304

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 307

cgttacataa cttacggtaa atggcccgcc tggtgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca	120
atgggtggag tatttacgtt aaactgcca cttggcagta catcaagtgt atcatatgcc	180
aagtacgcc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta	240
catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac	300
catg	304
<210> 308	
<211> 318	
<212> DNA	
<213> Homo sapiens	
<400> 308	
tgtacaaaa agcaggcttt aaaggaacca attcagtcga ctggatccgg taccaaggct	60
gggcaggaag agggcctatt tcccatgatt ccttcatatt tgcatatacg atacaaggct	120
gttagagaga taattagaat taatttgact gtaaacacaa agatattagt acaaaatag	180
tgacgtagaa agtaataatt tcttgggtag ttgcagttt taaaattatg ttttaaatg	240
gactatcata tgcttaccgt aacttgaaag tatttcgatt tcttggcttt atatatcttg	300
tggaaaggac gaaacacc	318
<210> 309	
<211> 3324	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"	
<400> 309	
atggccatca gatctatcaa gctgaagctc aagaccagaa caggctctga ggcccagaac	60
ctgcggaagg gcatttggcg gacacaccgc ctctcaacg agggcgctcg ctattatatg	120
aaaatgctgc tgcgtttcag acaggagagc acaggcggcc agaccaagaa agagctgcag	180
gaagaactgg tccggcacat cagagagcag cagcagaaga atcgggccga caagaacacc	240
caggccctgc ctctggacaa ggcttttggc gctctgcggc agctctacga gctgctggtg	300
ccttcttcta tcggccagag cggagacgag cagatcatca gcagaaagtt cctgtccct	360
ctcgtggacc ccaacagcga gggcggcaag ggcacatcta aagctggcgc caagcctaca	420
tggcagaaaa agaaggagc caacgacccc acctgggaac aggattacga gaaatggaag	480

aagcggagag aggaggaccc taccgccagc gtgattacca ccctggaaga atacggcatc	540
agaccaatct tcccactgta caccaacacc gtggccgaca tcgcctggct gcctctgcag	600
tccaaccagt tcgtgcggac atgggacaga gacatgctgc agcaggctat cgagagactg	660
ctgagctggg aaagctggaa caagagagt caagaggaat acagcaagct gcaagagaag	720
atgacacagc tcaacgagca actggagggc ggccaggagt ggatcagcct gctggaacag	780
tacgaggagc aacgggagca ggagctgac gagaacatga ccgccgctaa cgacaaatat	840
agaataacaa agcggcagat gaagggtggt aacgagctgt acgagcagt ggcaccctg	900
ctgccaatg cctctcacga gcagtaccgg gaagccctta agcgggtcca gcaaagactg	960
cggggcagat tcggcgacgc tcatttcttc cagtatctga tgaaggaaga gcaccacctg	1020
atttgaagg gcaaccccc gagaatccac tactttgtgg ccagaaacga gctgaagaaa	1080
agactggaag aggccaagca gaacgccact atgacctgc cagacgccag aaagcacccc	1140
ctgtgggtgc ggttcgacgc cagaggcgga aatctgcagg actactacct gaccgccgag	1200
gccgataacc ccagatctag aagattcgtt acctttagcc agttgatctg gcctaacgag	1260
tccggctgga tgaaaaagca ggacgtggaa gtggaactcg ccctgagcaa gcagttctac	1320
cagcaagtga ccctgcagaa gaacgataag gggaaacagg agatcgagtt taaggacaag	1380
ggctccggat ctacgttcag tggccatctg ggccgggcta agctgcaact ggagcgaggc	1440
gacctggaga aagaagagaa ggactttgag ggccggagaaa tcggaagcgt gtacctgaac	1500
atcgtgatcg acttcgagcc cttgcaggag gtgaaaaacg gcagactgca gagccatac	1560
ggccaggatgc tgcagctcgt tcggagacct aatgagttcc ctaagggtac cacatacaag	1620
tctgaagaac ttgttgagt gatgaaggcc agccagaatc acagcagcgg cgtggagtct	1680
ctggagtcgg gcttcagagt gatgagcatc gatctgggac tgaggacagc cgctgccacc	1740
agcattttct ctgtggaaga aagcaacgat gccaacgccg ctggcttcag ctactggatc	1800
gagggcaccc ctctggtcgc cgtgcacaag agaagctaca tgctgaagct gccaggcgaa	1860
caagtggaaa aacaggtcgc ggaaaagaga gatgagagac aagaccagca gaggcgcgtc	1920
agatttcaga tcagaatcct gagccagggtg atcaggatgg ccaagaaaca aaacagagaa	1980
agagctgacg aactggacca cctgagccag gactggaga agcagaagtc cctgctcgat	2040
cagaccgata gaaccttctg gaacggcatc gtttgtgacc tgaccgatgc gctgcgcgaa	2100
aaggaggag gctgggagca agccgtggtc caaatccaca gaaaggccga ggaacacgtg	2160
ggcaagggtg tgcaagcctg gcggaaaaga tttagcgcg atgagcggaa gggcatcgcc	2220
ggcctgagca tgtgttccat agaagagctg gacagcctcc ggaagctgct cattagctgg	2280

agcagaagga caagaaaccc tcaggagatc aacagattcg agcagggcca cacctctcac 2340

cagcggtgc tgacacatat ccagaacgtg aaggaagata gactgaagca actgagccac 2400

gccatcgtga tgaccgcctt gggctacgtg tacgacgaga agaagctgga gtggttcgcc 2460

aaataccccc cctgccaggt gatcctgttc gagaatctgt ctcagtacag aagccacatg 2520

gacagatcca cgaaggaaaa tagccgcctg atgaaatggg cccacagaag catccctaag 2580

tacgtccaca tgcaggccga gccttacgga atccagatcg gagatgtgag agccgaatat 2640

agcagcaggt tccacgcaa gacagggaca cctggcatcc gttgcaagat ggtgaaggga 2700

caagaactgc aaggcaagcg attcgagaac ctgcaaaaga gactggtgtc cgaacagttt 2760

ctgaccgagg aacaggtgaa gcagcttcgg cctggagata tcgtgccaga tgacagcgga 2820

gagtgtttca tgacctgag tgatggcagc gaaggcaagg aagtgtgttt cctgcaagcc 2880

gacatcaacg ccgccagaa ccttcagaaa cgattctggc agagatacaa cgaactgttc 2940

aaggtgtcat gcagagtgt gatcagaggc gaggaagagt acctgatccc caaggccaag 3000

agcgtgcaag ccaaactcgg caagggactg ttcgtgaaga aaaccgacac cgtgatgaag 3060

gacgtgtacg tgtgggatag ccaggccaaa ctgaagggca aaacaacctt caccgaggaa 3120

agcgagagcc ctgagcaact ggaggacttc caggagatca tcgaagaagc cgaggaagcc 3180

aagggcacct acagaacact gtttagagat cctagcggcg ttttcttccc cgagttcgtg 3240

tggaataccc agaaagactt ctggccgag gtgaagagaa ggctgtacgg caagctgcgc 3300

gagagattcc tgatgaagac ccgg 3324

<210> 310

<211> 21

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 310

cccagcccta ggttgtttat t 21

<210> 311

<211> 23

<

212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 311

ctggctacat cttccttgac tac 23

<210> 312

<211> 140

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 312

cgacggguag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc 60

cauuacaggg cggcuaccac gaauagccac gaaguaaaag cuucguggcu agcacccuug 120

gcuagucuug aaauccuagg 140

<210> 313

<211> 25

<212> DNA

<213> Homo sapiens

<400> 313

ccaacagtcc tacttcctg ttcca 25

<210> 314

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 314

gagatgtccc cagtgaactc caaat 25

<210> 315

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 315

cctgtcgttg cccctcccag atcat

25

<210> 316

<211> 49

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 316

gucugauugc cugucguugc ccucuccaga ucauggagga guuggcaga

49

<210> 317

<211> 41

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 317

gtctgattgc ctgtcgttgc ccctcccaag gatttggcag a

41

<210> 318

<211> 69

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 318

gtctgattgc ctgtcgttgc ccctaagtgt attaagcatt gtctcagaga ttttgagga 60
gttggcaga 69

<210> 319

<211> 39

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 319

gtctgattgc ctgtcgttgc ccctggagga gttggcaga 39

<210> 320

<211> 47

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 320

gtctgattgc ctgtcgttgc ccctcccaga tcggaggagt tggcaga 47

<210> 321

<211> 44

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 321

gtctgattgc ctgtcgttgc ccctcccaga taggagttgg caga 44

<210> 322

<211> 39

<212> DNA

<213> Artificial Sequence

<220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 322
 gtctgattgc ctgtcgttgc ccactctggga gttggcaga 39

<210> 323
 <211> 41
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 323
 gtctgattgc ctgtcgttgc ccctctggag gatttggcag a 41

<210> 324
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 324
 gtctgattgg aggagttggc aga 23

<210> 325
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 325
 gtctgattgc ctgtcgttgc ccctcggagg agttggcaga 40
 <210> 326

<211> 42
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 326
 gtctgattgc ctgtcgttgc ccctccagga ggagttggca ga 42

<210> 327
 <211> 34
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 327
 gtctgattgc ctgtcgttgc ccggagttgg caga 34

<210> 328
 <211> 42
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 328
 gtctgattgc ctgtcgttgc ccatcatgga ggagttggca ga 42

<210> 329
 <211> 43
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"

<400> 329
gtctgattgc ctgtcgttgc ccctccatgg aggagttggc aga 43
<210> 330
<211> 33
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 330
gtctgattgc ctgtcgttgc ccctcccagt act 33
<210> 331
<211> 44
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 331
gtctgattgc ctgtcgttgc cccaatcttg gaggagttgg caga 44
<210> 332
<211> 43
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 332
gtctgattgc ctgtcgttgc cctgggatgg aggagttggc aga 43
<210> 333
<211> 35
<212> DNA
<213> Artificial Sequence

<220><221> source

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

<400> 333

gtctgattgc ctgtcgttgc ccctcagttg gcaga 35

<210> 334

<211> 42

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 334

gtctgattgc ctgtcgttgc ccctcatgga ggagttggca ga 42

<210> 335

<211> 25

<212> DNA

<213> Homo sapiens

<400> 335

tcctttggaca gtgtccatac tgggtg 25

<210> 336

<211> 25

<212> DNA

<213> Homo sapiens

<400> 336

tggtgagcac acctcgggac tgggc 25

<210> 337

<211> 25

<212> DNA

<213> Homo sapiens

<400> 337

ctgccacaat accttggccc ttctc 25

<210> 338

<211> 25

<212> DNA
 <213> Homo sapiens
 <400> 338
 gctgcgaccc atacacccaa aggat 25

<210> 339
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 339
 ttgtcgatgg tcagcacagc cttat 25

<210> 340
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 340
 tgacttcatg gtaactggga gaggg 25

<210> 341
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 341
 tgtagtccag aaccgtgctg gcagc 25

<210> 342
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 342
 cacttcaggt cctgacagac actag 25

<210> 343
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 343

cacaggtcct caagggcaga agagt	25
<210> 344	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 344	
ccaagaaacg gggctgaaag ccggg	25
<210> 345	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 345	
tccttcacca ttatcaccac ccagt	25
<210> 346	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 346	
ggcttagggt taccgaagag gggcc	25
<210> 347	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 347	
ggctcagggt tactgaagag gggcc	25
<210> 348	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 348	
cctaaaggaa atgactgcag	20
<210> 349	
<211> 20	
<212> DNA	

<213> Homo sapiens
 <400> 349
 acttcgtcta ttcccagatc 20
 <210> 350
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 350
 tggactgcat cttagcccta 20

 <210> 351
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 351
 tgggtaggta ggctgggctg tgggt 25
 <210> 352
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 352
 ttgctcattg acctccactc agtgt 25
 <210> 353
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 353
 taacccccac cccactgcc gggta 25
 <210> 354
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 354
 ttgctcgttg atctccactc agtgt 25

<210> 355
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 355
 caccacggct gtcgacacca 20
 <210> 356
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 356
 caacagtgga ggaaagcctt 20
 <210> 357
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 357
 ctatctggtc gtgggagtac gggat 25
 <210> 358
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 358
 aaataagcag caggaagcag 20
 <210> 359
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 359
 ctcgatatag gtggagtctt 20
 <210> 360
 <211> 25
 <212> DNA
 <213> Homo sapiens

<400> 360
 tggttgtgca gccagatgcc tagac 25
 <210> 361
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 361
 tcttgctcac ttggacctt ggtgg 25
 <210> 362
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 362
 ctggctgtaa gtcgggcttt ctcca 25
 <210> 363
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 363
 ctgcaagcaa cctccagtgg tgacc 25
 <210> 364
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 364
 aggagctctc gagccagttt 20
 <210> 365
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 365
 gtcagaggaa ggttgagccc 20
 <210> 366
 <211> 25

<212> DNA
 <213> Homo sapiens
 <400> 366
 gggtctgttt gaacatggag aacac 25

<210> 367
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 367
 gaaagatcta gctcaccgtc ctctt 25

<210> 368
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 368
 tgcaagccca ccttggtcca tgttg 25

<210> 369
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 369
 ttccctccat catgtgtccc agggc 25

<210> 370
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 370
 ggaagtgggt ggacaagtaa gcgcc 25

<210> 371
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 371

cccgctgct tacagatttc agttt 25

<210> 372

<211> 20

<212> DNA

<213> Homo sapiens

<400> 372

gagagtacag ccacaccatt 20

<210> 373

<211> 25

<212> DNA

<213> Homo sapiens

<400> 373

tatcacttac ggatcacaga ggggg 25

<210> 374

<211> 25

<212> DNA

<213> Homo sapiens

<400> 374

gtgacagagc tggtttcaaa gttgc 25

<210> 375

<211> 25

<212> DNA

<213> Homo sapiens

<400> 375

cctgaagcca tcaaggggct gcagc 25

<210> 376

<211> 25

<212> DNA

<213> Homo sapiens

<400> 376

tttactctcc cacagtggcg ctgct 25

<210> 377

<211> 25

<212> DNA

<213> Homo sapiens
 <400> 377
 ctccaattca ggacccacat gacgg 25
 <210> 378
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 378
 atctctctct catctccag cctct 25

 <210> 379
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 379
 gggggctggt ttgccatcca agggg 25
 <210> 380
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 380
 gccggacagg gatgttcatt gtgat 25
 <210> 381
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 381
 gccattttgg ctgagttggt ccagt 25
 <210> 382
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 382
 gccactttgg ctgagttggt ccggt 25

<210> 383
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 383
 agctgctagt cccagctgag gacgt 25
 <210> 384
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 384
 ttggactccc ttgatgccca ggaga 25
 <210> 385
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 385
 cctaccctga aaatccgaaa 20
 <210> 386
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 386
 tagtagcaca cactctgcac cttga 25
 <210> 387
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 387
 ccctcgttga catcgatgct tgaga 25
 <210> 388
 <211> 25
 <212> DNA
 <213> Homo sapiens

<400> 388	
gacagcggaa accccggaaa agcag	25
<210> 389	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 389	
acctgagtgt tttagctacg gtgag	25
<210> 390	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 390	
gaaggtctcc tccctggtca	20
<210> 391	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 391	
ctgctgaatc agccggcaga accag	25
<210> 392	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 392	
tgtgcccga ctcgaaaaag	20
<210> 393	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 393	
cttgggactg taaaagctgt	20
<210> 394	
<211> 20	

<212> DNA
 <213> Homo sapiens
 <400> 394
 gtaagtgatc catcctaggt 20

<210> 395
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 395
 aacgaagccc tccgtgagag 20

<210> 396
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 396
 caagaatctg accatcttgg 20

<210> 397
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 397
 tcgtccgtga cctgggagcc 20

<210> 398
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 398
 acagatgaac actggtcagc 20

<210> 399
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 399

tcagtgcmaa cctcctcaca 20

<210> 400

<211> 20

<212> DNA

<213> Homo sapiens

<400> 400

ttgatcttgt cgggagttcc 20

<210> 401

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 401

aagactaacc tggccaacat 20

<210> 402

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 402

gtcttgaact ccggacctca 20

<210> 403

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 403

gccgggcgtg gtgtcgcgcg	20
<210> 404	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
<400> 404	
tcctgccgca gcctctggag	20
<210> 405	
<211> 135	
<212> RNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"	
<400> 405	
cgacggguag aggccguaug ucgaauugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagccac gaaguuuag cuucguggcu agcacaagac	120
uaaccuggcc aacau	135
<210> 406	
<211> 135	
<212> RNA	
<213> Artificial Sequence	
<220><221> source	
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"	
<400> 406	
cgacggguag aggccguaug ucgaauugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagccac gaaguuuag cuucguggcu agcagucuu	120
gaacuccgga ccuca	135

<210> 407

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 407

cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagccac gaaguaaaag cuucguggcu agcacgccgg	120
gcgugguguc gcgcg	135

<210> 408

<211> 135

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 408

cgacgguuag aggccguaug ucgauuugcu uuaauuucgu gcgugugcau ugucguccuc	60
cauuacaggg cggcuaccac gaauagccac gaaguaaaag cuucguggcu agcacuccug	120
ccgcagccuc uggag	135

<210> 409

<211> 3324

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 409

atggccattc ggagcatcaa gctcaagctg aagaccagaa ccggtcctga ggctcagaac	60
ctgcggaagg gcatctggag aaccaccgg ctgctgaacg agggcgtggc ctactacatg	120
aagatgctgc tgctcttccg gcaggaatcc actggcggcc agacaaagaa ggaactgcag	180

gaggaactgg tgcggcacat ccgggagcag cagcagaaaa acagagccga caagaatacc	240
caggccctgc ctctggataa ggccttcgcc gctctgagac aactgtacga gctgctcgtg	300
ccatctagca tcggccagag cggcgatgcc cagatcatct ctagaaaatt cctgtctcct	360
ttggtggacc ctaacagcga gggcggcaag ggcactagca aggcgggagc caagcccaca	420
tggcagaaaa agaaggaggc caatgaccct acatgggagc aagattacga gaagtggaaa	480
aagcgtagag aggaagatcc taccgccagt gtgataacaa ccctggaaga atatggaatt	540
agacctatct ttccactgta caccaacacc gtggccgata tcgcctggct gcctctgcag	600
tctaaccagt ttgtccggac atgggatcgg gacatgctgc agcaggccat cgagaggcta	660
ctctcttggg agtcttgaa caagcgggtg caagaggaat acagcaagct gcaggagaag	720
atgacgcaac tgaacgagca actggagggc ggacaggagt ggatcagcct gctggaacaa	780
tacgaggaac agagagagca ggagctgac gaaaacatga ccgccgcaaa cgataaatac	840
cgcataacca agcggcagat gaagggtggt aatgagctgt atgagcagtg gtctaccgtg	900
ctgccaatg ccagccacga gcagtacaga gaggcactga aaagagtga acagcggctg	960
aggggcagat tcggcgacgc tcatctcttc cagtatctga tgaaagaaga acaccacctg	1020
atctggaagg ggaaccctca gagaatccac tacttcgtgg ccagaaacga gctgaagaaa	1080
agactggaag aagccaagca gaacgccacc atgacctgc ctgatgccag aaagcacctt	1140
ctgtgggtca gatttgacgc cagaggcggc aacctgcaag actactacct gacggccgaa	1200
gccgacaacc ccagaagcag aaggttcgtg accttcagcc aactgatctg gcctaacgag	1260
agcggctgga tggaaaaaca ggacgtggaa gttgaactgg ccctgagcaa gcagttctac	1320
cagcaagtga cactgcagaa aaacgacaag gggaagcagg aaatcgagtt caaagacaag	1380
ggcagtgga gcaccttcag cggccacctg ggcggggcca agctgcaact ggagagagga	1440
gatctggaga aggaagaaaa agactttgaa ggcggcgaaa tcggcagcgt gtacctgaac	1500
atcgtgatcg acttcagacc actgcaggag gtgaagaacg gccgactgca atctccttac	1560
ggccagggtg tgcaactggt cagaaggcct aatgagttcc ccaagggtgac cacctacaag	1620
tctgaggaac tggtcgagtg gatcaagagc agcaccaagg acagcgccgg cgtggagtcc	1680
ttagagagcg gttttagagt gatgagcatc gacctgggac tgcggaccgc cgccgcaaca	1740
agcatcttct ctgtagagga atccaacgac gccaacgccg cgggtttcag ctactggatc	1800
gagggaaactc ctctggttgc cgttcataag cggatcatata tgctgaaact gcccggagag	1860
caagtcgaaa agcagggtgc agagaagcgg gacgagcggc aggatcagca gagaagagtg	1920
aggttccaaa tcagaatcct gagccagggt atccggatgg ccaagaagca gaaccgggag	1980
cgggctgatg agctggacca cctgtcccag gccttgaga agcaaaaatc tctgctggac	2040

cagaccgatac ggaccttttg gaacggcatic gtgtgacgacc tgacagacgc tctgagagag 2100
aaagagggcg gatgggaaca ggccgtgggc cagatccaca ggaaggccga ggagcacgtg 2160

ggcaaggtgg tgcaagcctg gcggaacggg ttgacgccc atgaacgcaa gggcatcgcc 2220
ggcctgtcta tgggtctat cgaggagctg gacagcctgc ggaagctgct gatctcttgg 2280
agcagaagaa caagaaaccc cagagaaatc aactgcttcg agcagggcca caccagccac 2340
cagcggtgc tgacacacat ccagaacgtg aaggaggacc ggctgaagca actgagccac 2400
gccattgtga tgacagcctt gggctacgtg tacgacgaga agaaattgga atggtttgcc 2460
aagtaccctg cttgtcaggt gatcctgttc gagaacctgt cccagtaccg gtccaacatg 2520
gacagaagca ccaaagagaa tagcaccttg atgaaatggg cccacaggag catccctaag 2580

tacgtgcaca tgcaggccga gccttacggc atccagatcg gcgatgtgcg ggccgagtac 2640
tccagcagat tccagccaa gacaggcaca cctggcatcc ggtgcaagat ggtgtccgga 2700
cacgacctgc aaggcaggcg cttcgagaac ctgcagaagc ggttaatctc tgaacagttc 2760
ctgacagagg agcaagtga gacgtcaga cccggcgaca tcgtgcccga cgactccggc 2820
gagtggttca tgacctgag gcagggcagc gaaggcaaag aggttgtgtt cctccaagcc 2880
gacatcaacg ccgccccaaa tctgcaaaag agattctggc agcggtacaa cgagctgttt 2940
aaggtctcct gcagagtgt gatccgagga gaagaggaat acctgatccc caagacaaag 3000

tccgtgcaag ccaagctggg caaaggcctg ttcgtgaaaa aaaccgacac cgtgatgaag 3060
gacgtgtacg tgtgggacac ccagcctaag ctgaagggca aaaccacatt caccgaggag 3120
tccgaaagcc ctgagcaact ggaggatfff caggagatca tcgaagaagc cgaagaagct 3180
aagggcacat acagaacact gtttagagac cccagcggag tgttcttccc tgagtctgtg 3240
tggtccacc agaaagattt ctggtccgag gtgaagagac ggctgtacgg caagctgaga 3300
gagcggttcc tgatgaagac cagg 3324

<210> 410
<211> 3324
<212> DNA
<213> Artificial Sequence

<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"
<400> 410
atggccatca gatctatcaa gctgaagctc aagaccagaa caggctccta gggccagaac 60

ctgcggaagg gcatTTggcg gacacaccgc ctctcaacg agggcgTcgC ctattatatg	120
aaaatgctgc tgctgttCag acaggagagc acaggcgGCC agaccaagaa agagTgcag	180
gaagaactgg tccggcacat cagagagcag cagcagaaga atcgggccga caagaacacc	240
caggccctgc ctctggacaa ggcttttgcc gctctgcggc agctctacga gctgctggTg	300
cctttcttcta tgggccagag cggagacgcg cagatcatca gcagaaagtT cctgtccct	360
ctcgtggacc ccaacagcga gggcgGcaag ggcacatcta aagctggcgC caagcctaca	420
tggcagaaaa agaaggaggc caacgacccc acctgggaac aggattacga gaaatggaag	480
aagcggagag aggaggaccc taccgccagc gtgattacca ccctggaaga atacggcatc	540
agaccaatct tcccactgta caccaacacc gtggccgaca tcgcttggt gcctctgcag	600
tccaaccagt tcgtgcggac atgggacaga gacatgctgc agcaggctat cgagagactg	660
ctgagctggg aaagctggaa caagagagtg caagaggaat acagcaagct gcaagagaag	720
atgacacagc tcaacgagca actggagggc ggccaggagt ggatcagcct gctggaacag	780
tacgaggagc aacgggagca ggagctgatC gagaacatga ccgccgctaa cgacaaatat	840
agaataacaa agcggcagat gaagggttgG aacgagctgt acgagcagtg gagcacctgT	900
ctgcccaatg cctctcacga gcagtaccgg gaagccctta agcgggtcca gcaaagactg	960
cggggcagat tggcgacgc tcatTTcttc cagtatctga tgaagggaaga gcaccacctg	1020
atttgaagg gcaaccccca gagaatccac tactttgtgg ccagaaacga gctgaagaaa	1080
agactggaag aggccaagca gaacgccact atgacctgc cagacgccag aaagacccc	1140
ctgtgggtgc ggttcgacgc cagaggcgga aatctgcagg actactacct gaccgccgag	1200
gccgataacc ccagatctag aagattcgTt acctttagcc agttgatctg gcctaacgag	1260
tccggttgga tggaaaagca ggacgtggaa gtggaactcg ccctgagcaa gcagtTctac	1320
cagcaagtga ccctgcagaa gaacgataag gggaaacagg agatcgagtT taaggacaag	1380
ggctccggat ctacgttcag tggccatctg ggcggggcta agctgcaact ggagcgaggc	1440
gacctggaga aagaagagaa ggactttgag ggcggagaaa tcggaagcgt gtacctgaac	1500
atcgtgatcg acttcgagcc cttgcaggag gtgaaaaacg gcagactgca gagccatac	1560
ggccaggtgc tgcagctcgt tcggagacct aatgagTtcc ctaaggtgac cacatacaag	1620
tctgaagaac ttgttgagtg gatgaaggcc agccagaatc acagcagcgg cgtggagtct	1680
ctggagtcgG gcttcagagt gatgagcatc gatctgggac tgaggacagc cgctgccacc	1740
agcattttct ctgtggaaga aagcaacgat gccaacgccg ctggcttcag ctactggatc	1800
gagggcaccc ctctggtcgc cgtgcacaag agaagctaca tgctgaagct gccaggcgaa	1860
caagtggaaa aacaggtgcg ggaaaagaga gatgagagac aagaccagca gaggcgcgtc	1920

agatttcaga tcagaatcct gagccagggtg atcaggatgg ccaagaaaca aaacagagaa 1980

agagctgacg aactggacca cctgagccag gcaactggaga agcagaagtc cctgctcgat 2040

cagaccgata gaaccttctg gaacggcatc gtttgtgacc tgaccgatgc gctgcgcgaa 2100

aaggaggagg gctgggagca agccgtgggtc caaatccaca gaaaggccga ggaacacgtg 2160

ggcaaggtgg tgcaagcctg gcggaaaaga ttgacgccg atgagcggaa gggcatcgcc 2220

ggcctgagca tgttgtccat agaagagctg gacagcctcc ggaagctgct cattagctgg 2280

agcagaagga caagaaacct tcaggagatc aacagattcg agcagggccca cacctctcac 2340

cagcggtgctg tgacacatat ccagaacgtg aaggaagata gactgaagca actgagccac 2400

gccatcgtga tgaccgcctt gggctacgtg tacgacgaga agaagctgga gtggttcgcc 2460

aaataccccg cctgccaggt gatcctgttc gagaatctgt ctcatgacag aagccacatg 2520

gacagatcca cgaaggaata tagcaccctg atgaaatggg cccacagaag catccctaag 2580

tacgtccaca tgcaggccga gccttacgga atccagatcg gagatgtgag agccgaatat 2640

agcagcaggt tccacgcaa gacagggaca cctggcatcc gttgcaagat ggtgaaggga 2700

caagaactgc aaggcaagcg attcgagaac ctgcaaaaga gactgggtgc cgaacagttt 2760

ctgaccgagg aacaggtgaa gcagcttcgg cctggagata tcgtgccaga tgacagcggg 2820

gagtgggtca tgaccctgag tgatggcagc gaaggcaagg aagtgggtgtt cctgcaagcc 2880

gacatcaacg ccgcccagaa ccttcagaaa cgattctggc agagatacaa cgaactgttc 2940

aaggtgtcat gcagagtgt gatcagaggc gaggaagagt acctgatccc caaggccaag 3000

agcgtgcaag ccaaaactcg caagggactg ttcgtgaaga aaaccgacac cgtgatgaag 3060

gacgtgtacg tgtgggatag ccaggccaaa ctgaaggga aaacaacctt caccgaggaa 3120

agcgagagcc ctgagcaact ggaggacttc caggagatca tcgaagaagc cgaggaagcc 3180

aagggcacct acagaacct gtttagagat cctagcggcg ttttcttccc cgagttcgtg 3240

tggaatacc agaaagactt ctggccgag gtgaagagaa ggctgtacgg caagctgcgc 3300

gagagattcc tgatgaagac ccgg 3324

<210> 411

<211> 3279

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 411

atgaaaaccc tggggatcga tctgggaacc gccagcatcg gctgggctat cctggacgag	60
gccagcatca tgcctccgg cgtgcggatc tttagccaga gcgagatggc tggcagagat	120
ccccagagca aggccagcct ggccgtcgcc cggagggacg ccagaggagc tagaaggagg	180
agagacagat acctgaaaag acggaggaga ctgctggacc tgctgacaga gcacggcctg	240
gctccagccg atgagaaaag ccggaaggcc ctggtgcggg aatacgagga tggcaaagat	300
ggagatctga gcaacagcgt gtacgccctt agagctagag ccttggatga agccctgacc	360
ccttacgaga tggggcgggc catcttccaa ctgaaccagc ggcggggctt taagtctaac	420
agaaagaccg acagcaacga ccccgagcag ggaataatcg ccacagcaat ccacgtgctg	480
gatgccaaaa tggacgagga taaggcccgg accttcggag agtggctgca catgagaaga	540
ctcaaaggcc tgagcgtcg cgctcggatg accgccgacg gcgattctta cgatttctac	600
cctagccgtg ccgccctgga aagagagttc gacagactga tggccagcca gaagagattc	660
caccctgacc tgctgaactc aagcgtgatc gacgacatca ggaaggtggt gtttcaccag	720
cggcctctga agcctgtgca acctggcaag ttagctaca accacagaga atctcggctg	780
cccaaggcac accccttgtt tcagaaattc cggctgctga aagaggtgaa tgagctggaa	840
atcgtcggcg aagatcagag atacgtgaag ctgaccctg cccagcgcga tgtgctgaca	900
ctggccctca gaacaggcct gaccaagcaa ggtagactgc cgttctctaa gctgcggagc	960
attctgaagc tgggcaagga ggttcggttt aacaaggaaa aggacaaccg gaccgacctg	1020
gaaggcgatg tgatccactt cagagtgtcc aggcccgact gcttcggcaa tagatgggcc	1080
gctatgcctg tggaagagca agccgtgtg accgagaagc tgagaaccga gccagattac	1140
agcgccctgc tcgattgggt gaagaacgag gctggcttgg atgaggctca tgcccagacc	1200
gtggcgcgaca cacctgtgcc cgacggcttc ggcaaatgg gaccagcgc cttgtctgcc	1260
ctggctgatg ctatggaaca cgagattgat gcacagggtc tcgtgatcac cgaggccgag	1320
gccgccaaga gagtgtacgg cagaaccaat agcgaggctg acccggccg gaaaggagtg	1380
gaccaactgc ctaagtacca ggaggtcctg cagaggcata ttctctctgg aacaggcgaa	1440
cctgatgacc cctacgacga gtacatgggc agaatacca accccaccgt gcacatcgcc	1500
ctgaaccaac tgagaagact ggtcaacgca ctgatccgga agtacggcaa gccaacaaa	1560
atcgccatcg aagtgggccg ggagctgaaa ctgaacgaaa agcagagaaa cgaggccaac	1620
cgagaaatcg gcaaaaatac cagagccgt atggccagag gccagcaact ggtggaatc	1680
ttcaagcagc caaacaccgg atacaaccgg ctgagactgg aactgtggga agatctgaac	1740
cgggaacagc ctctgaaacg gctctgtacc tactcggca aggccatcgc tgcccacatg	1800

ctgttcaacg gcgagacaga tatcgatcac atcctgccat attccaagac actggatgac 1860
tccaaggcca acaggctgtt gtgctgcaca ccttgcaaca gagagaagaa gaactacgcc 1920

cttgccaacg tgctgcagtg gcgggaccac tacggcgaga tcttgccag agccaccgcc 1980
ctccctaaga acaagcagtg gagattcgcc gaggatgcta tgaccagata cgaggccgag 2040
ggcggtttcc tggcccgga gctcacagat atgcagtaca tcagcagact cgccttgaca 2100
tacctggccc acctgtacga ctacaggaa cccgacctgg acggcgtgta caagagacac 2160
gacagagtga gagccctccc tgggcgcagt accgagatgc tgagacggca gtgggacctg 2220
aacgaactgc tccacggcca caacctggct ggtggcgacg gcgccaagga aaagaacaga 2280
ctggatcaca gacaccacgc cattgatgct atcgtgatcg cttgtaccag ccagtctctt 2340

attaacagac tgtccaccgc cgccggcgag gccgaagaac ggggcgccgc cagagtggtc 2400
gagagaatcg accctccttg gcctagcttc cggaagatg tgagagaagc tgttaatgcc 2460
attgtggttt ctcaaacgc tgaccacggc accgctagcc ggagcggcta cgacaagggc 2520
agaggccaga ccgccggcaa gctgcacaac gacactgcct atggcgagac aggcgagaaa 2580
gaccacaacg gcaacaacct ggtggtgcgg cggatcgcca tcagcgacat caagcggagc 2640
gccgacatca tgaagatcag aacaacgcc caccggccaca gcgagctgag agacagactg 2700
tacgaagcca ctcgggatct ggaaggcaag gctttcgagc agcccgtagc agcttttgtg 2760

aagcacgacg ccaagttcaa ggggatcaga cacgtgcgag tgaccgaagt tcagaacccc 2820
gtctggatca cccacggcgg aggcaaatac aagaagggt accctgcctgg aggcaatgat 2880
cggttcgacg tgtgggagct gcctgatggt aaatgggacg ccgaggtggt gacgaccttt 2940
gatgcccatc ggctgactt caccctaga atgcgtagc agcaccacaa cgctcggaag 3000
atcatgtctc tgaagaaagg cgacatgatc gcctacgacg atcctgacag cgggaagcgc 3060
gtgattgcca tcgtgcggaa attcgaccag cggaacaagc aactgtacct ggacctcac 3120
aatgaggcgc gcaacctgga ccagagagaa aaagaaaaa catacaagcc cctaagacct 3180

atgcccaacc ctctgaaaaa gtacagacc agacaggtgc gggtagacga gattggacaa 3240
gtgttcgacc ctggcccttg gtgggagaag cgctctgat 3279

<210> 412
<211> 3162
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic

polynucleotide"

<400> 412

atgaaagaga agtatatcct gggcctagac ctgggcatta catctgtggg atacggcatc	60
atcaacttcg agacaaagaa aatcattgac gccggcgtgc ggctgtttcc tgaagccaac	120
gtggacaaca atgaggggaag acggagcaag agaggcagca gaaggctgaa gagaagaaga	180
attcacagac tggaagagt gaagctgctg ctgaccgagt acgacctgat taacaaggaa	240
cagatcccta caagcaacaa cccctaccag atcagagtga aaggactgag cgagatactg	300
agcaaggatg aactggccat cgctcttctg cacctggcta aaagacgggg aatccacaac	360
atcaacgtca gcagcgaaga tgaggatgct agcaacgagc tgtccaccaa ggaacagatt	420
aaccggaaca acaagctgct caaggacaaa tacgtgtgcg aggtgcaact gcagcggctg	480
aaggaggggc agatccgggg cgagaagaac agattcaaga caaccgacat cctgaaagaa	540
atcgaccaac tgctgaaggt gcaaaaagat taccacaacc tggacatcga ctatcatcaat	600
caatacaagg agatcgtgga aacacggaga gactacttcg agggccccgg ccagggtctt	660
ccattcggct ggaacgggga tctgaagaaa tggatgaga tgctgatggg cactgcacc	720
tacttcccc aggagctgag aagcgtgaag tacgcctaca gcgccgactt gtttaacgcc	780
ctcaatgac tgaacaatct gatcattcag agagacaact ccgagaaact ggaataccac	840
gagaagtacc acataattga aaatgtgttc aaacagaaga agaaaccac cctgaagcag	900
attgctaag agatcggcgt gaacctgag gacatcaagg gctacagaat caccaagagc	960
ggcaccacc agttaccga gttcaagtta taccacgac tgaagagcat cgtctttgac	1020
aagtccatcc tggaaaacga ggccatctc gatcagatcg ccgagatcct gaccatctac	1080
caagatgaac agagcatcaa ggaagagctg aacaaactgc ctgagattct gaatgagcag	1140
gacaaggccg agattgcca actcattggc tacaacggta cacacagact gagcctgaag	1200
tgtatccacc tgatcaatga ggaactgtgg cagacaagta gaaaccagat ggaaatcttc	1260
aactacctga acatcaagcc caacaagggt gacctgtctg aacagaacaa gatccccaag	1320
gacatggtga acgacttcat tctgagccct gtggtgaaaa gaaccttcat ccagagcatc	1380
aatgtgatca ataaggtgat cgagaagtac ggcatcccag aggacatcat cattgagctg	1440
gccagagaga acaacagcga cgaccggaag aagttcatca acaacctgca gaagaagaac	1500
gaggccacaa gaaagcggat taacgagatc atcgccaga ccggcaacca gaatgcaaaa	1560
cgcatcgtgg aaaagatcag actgcacgac cagcaggagg gcaagtcct gtacagcctg	1620
gaaagcattg ctctgatgga cctgtgaac aatcctcaga attacgaggt ggaccacatc	1680
atccccagaa gcgtggcctt cgacaacagc atccacaaca aggtgctggt gaagcagatc	1740

gagaatagca agaagggcaa ccggaccctt taccaatacc tgaattctag cgatgccaaag 1800
ctgtcttaca accagttcaa gcaacacatc ctgaacctgt ccaaaagcaa ggatagaatc 1860

agcaaaaaga agaaggacta cctgctggaa gagagagata tcaacaagtt cgagggtgcaa 1920
aaggagttca tcaaccgaaa cctgggtggac accagatagc ccaccagaga actgacgtct 1980
tatctgaagg cctacttttc tgccaacaac atggatgtga aagtcaagac catcaacggc 2040
tccttcacca accacctgcg gaaagtgtgg cggttcgaca agtaccgcaa ccacggctac 2100
aagcaccacg ccgaggatgc cctgatcatc gaaaacgccg atttctgtt taaggaaaac 2160
aaaaagctgc agaacgctaa caagattctg gaaaagccca caatcgagaa caataccaag 2220
aaggtgaccg tggaaaaaga ggaagattat aataatgtgt tcgagactcc taagctggtc 2280

gaggatatca aacagtaccg ggactacaag ttcagccacc gggtggataa gaagcctaac 2340
agacaactga tcaacgacac tctgtattcc acaagaatga aggacgagca cgactatata 2400
gttcagacca tcaccgacat ctacggcaag gataacacca atctgaaaaa gcagttcaac 2460
aagaaccccg aaaaattcct gatgtaccag aacgataccta agaccttcga gaagctgtcc 2520
atcatcatga aacagtacag cgacgagaag aacctctgg ccaagtatta tgaggaaacc 2580
ggagagtatc tgacaaagta ctcaagaag aataacggcc caatcgtgaa gaagatcaag 2640
ctgctgggca acaaggtggg caatcacctg gatgtacca acaaatcga gaacagcacg 2700

aaaaagctgg tgaagctgag tatcaagaac tacagattcg acgtctacct gacagagaag 2760
ggctacaaat tcgtgacct cgcttactta aatgtgttca agaaggacaa ctactactac 2820
attcctaagg acaagtacca ggagttgaag gaaaaaaga agatcaaaga taccgaccag 2880
tttatcgctt ctttttaca gaatgacctg atcaagctga acggcgacct ttataagatc 2940
atcggcgtga actccgatga cagaaatata atcgagctgg actactacga cataaaatac 3000
aaggactact gcgaaatcaa caacatcaag ggagagccta gaatcaagaa aaccatcgga 3060
aaaaagacag agagcatcga gaagttcacc acagatgtgc tgggaaatct gtacctgcat 3120

tctaccgaaa aggccctca actgatcttc aagagagggc tg 3162

<210> 413
<211> 3213
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 413

atgaggagc tggactacag gatcgccctg gacatcggca ccaacagcat cggctggggc	60
gtgatcgagc tgagctggaa caaggacagg gagcagtacg agaagaccag gatcgtggac	120
aagggcgtga ggatgttcga caaggccgag atccccaaga ccggcgccag cctggccgag	180
cccaggagga tcgccaggag cagcaggagg aggctgaaca ggaagagcca gaggaagaag	240
gacatcagga acctgctggt gcagcacgag atcatcagcc agaaggagct gaccagcctg	300
tacccctga gcaagagcag catggacatc tgggacatca ggctggacgg cctggacagg	360
ctgctggaca ggttcgagtg ggccaggctg ctgatccacc tggcccagag gaggggcttc	420
aagagcaaca ggaagagcga gctgaaggac gtggagaccg gcaaggtgct gagcagcatc	480
caggtgaacg agaagaggct gagcctgtac aggaccgtgg gcgagatgtg gatgaagaac	540
gccgactgca gcaagtacgg caaggaggagg aacagcccca acgagtacgt gttcagcgtg	600
agcagggccg acctggagaa ggagatcgtg acctgttcg aggcccagag gaagtccac	660
agcagctacg ccagcgtgga cctgcagaag acctacatcc agatctgggc ccaccagctg	720
cccttcgcca gcggcaacgc catcgtgaac aaggtgggct actgcagcct gctgaagggc	780
aaggagaaga gggtagccaa ggccacctac accttcagct acttcaaac cctggaccag	840
atcaacagga ccaggctggg cccaacttc cagcccttca ccaaggagca gagggacatc	900
atcctggaca agatgttcca gaggaccgac tactacaaga agaagaccat ccccagggtg	960
acctactacg acatcaggaa gtggctggcc ctggacgaga ccatccagtt caagggcctg	1020
acctacgacc ccaacgagga gctgaagaag atcgagatga agcccttcat caacctgaag	1080
cccttctacg agatcaagaa ggtggtgacc aactacgcca agaagaccaa cgagggttc	1140
agcgccctgg actacgacac cgtggcctac gcctgaccg tgtacaagac cgacaaggac	1200
atcaggagct acctgaagag gagcaacaac ctgagcaaga ggtgctacga cgaccagctg	1260
atcgaggagc tgctgacct gagctacacc aagttcggcc acctgagctt caaggccatc	1320
aaccacgtgc tgcccatcat gcaggagggc aggacctacc aggaggccat ccaccagctg	1380
ggctacgaca ccaccaacct gaagaaggag aacaggagca tgttcctgcc catcatcccc	1440
gacgagatca ccaaccccat cgtgaagagg gcctgacct aggccaggaa ggtggtgaac	1500
gccatcatca ggaggtacgg cagccccaac agcgtgcaca tcgagctggc caggagctg	1560
agcaagagcc acgacgagag gaagaagatc atgaccgcc acgacgagaa ctacaagaag	1620
aacaagggcg ccatcagcat cctgatcgag aacggcatcc tgaacccac cggtacgac	1680
atcgtgaggt acaagctgtg gaaggagcag ggcgagaggt gcgcctacag cctgaagaag	1740
atccccgcca acaccttctt caacgagctg aagaaggaga ggagcggccc cccgtgctg	1800

gaggtggacc acatcctgcc ctacagccag agcttcatcg acagctacca caacaaggtg 1860
ctggtgtacg gcgacgagaa ccagaagaag ggcaacagga tcccctacac cttcttcagc 1920

gaggaggaca aggagtggga gagcttcgag agctacgtga ggagcaacag cttcttcagc 1980
aagaagaaga ggggctacct gctgaagaag gcctacctgc ccaggagag caacctgatc 2040
aaggagaggc acctgaacga caccaggtac gccagcagct acctgaagaa cttcatcgag 2100
aagaacctga agttcaagga ggccgtgggc atcaccagga agaagtacgt gcagaccgtg 2160
aacggcgtga tcaccgcccc cctgaggaag aggtggggcc tggagaagga caggcaggag 2220
acctacctgc accacgccat ggacgccatc atcgtggcct gcaccgacca ccacatggtg 2280
accaaggtga ccgagtacta ccagatcaag gagggcaaca agagcatcaa gaagccctac 2340

ttccccctgc cctggatggg cttcaggag gagatcctga gccacctgga gagccagccc 2400
atgccagga agatcagcga ggagctgaag atcggtacc agagcagcga ctacatcctg 2460
gtgagcagga tgccaagag gagcgtgacc ggagcagccc acgaccagac cgtgatgaag 2520
aaggcgcca tcgacaagaa gggcaagacc atcatcatca agagggtgca cctgaaggac 2580
atcaagtctg acgagaacgg cgacttcaag atggtgggca aggagcagga cctggccacc 2640
tacgaggcca tcaagcagag gtacctggag tacaggaaga agagcaagaa ggccttcgag 2700
acccccctgt acaagcccag caagaagggc aagggaacc tgatcaagaa gatcaaggtg 2760

gaggtgcaga ccaagagctt cgtgaggag atcaacggcg gcgtggccca gaacggcgac 2820
ctggtgaggg tggacctgtt cgagaaggac aacaggtact acatggtgcc catctacgtg 2880
gtggacaccg tgaggagcga gctgccaac aaggccgtga ccagcagcaa gggctacgag 2940
cagtggctga gcatcgaaa cagcttcacc ttcaagtca gcctgtacc ctacacctg 3000
gtgagcgtgg tgaaggcgca cgaggacagg ttctgtact tcagcacct ggacatcaac 3060
agcgacaggc tgaacttcaa ggacgtgaac aagcccagca agcaggccga gtacaggtac 3120
agcctgaaga ccatcgagaa cctggagaag tacgagatcg gcgtgctggg cgacctgagg 3180

ctggtgaggc aggagaccag gaagatcttc aag 3213

<210> 414
<211> 3228
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 414

atgagcgagc tggactacag aatcgccctg gacataggaa caaacagcat cggctggggc	60
gttatagaac tgttttggaa caaggacaga gaaagatagc agaaggtgag gatagtgagc	120
aagggcgtgc ggatgtttga caaggctgaa atccccaaga ccggagcctc tctggccgaa	180
cctagaagaa tcgccagaag cagcagaagg agacttaata gaaagagcca gagaaaaaag	240
gagatcagaa acctgctggt gcaacacggc gtgatcaccc aggaagagct ggacctgctg	300
tacccctgt ctaagaagag catggacatt tgggacatcc gcctggacgg cctcgacaga	360
ctgctgaacc acctggaatg gacacggctg ctgattcacc tggctcagag gcgtggattt	420
aagtctaacc ggaagtctga gctgaaagac gccgaaacag gcaaagtctt gagcagcatc	480
caggtgaacg agaaaagact gagcctgtat agaaccgtgg gcgaaatgtg gatcaaggac	540
gccgaattct caaaatacga tagacggaga aatagcccta acgaatacgt ctttagcgtg	600
tccagagccg acctggagaa ggaaatcgtt accctcttcg aggccagcg gaaattccag	660
agcagctaca gcagcaaaaa cctccaggag acctacctgc aaatctgggc ccaccaactg	720
cctttcgcca gcggcaacgc catcttaaac aaggtgggct actgcagcct cctgaaaggc	780
aaggagcgga gaatccctaa ggctacatac accttcaggt acttcagcgc cctggaccaa	840
gtgaaccgga ccagactcgg tctgacttc caaccttta cacaggagca gaaagaaatc	900
atcctggaca agatgttcca gcggaccgac tactacaaga agaagaccat tcccaggtg	960
tcctattatg atatcagaaa gtggctggaa ctcgacgaaa ctatccaatt taagggcctg	1020
aactacgac ctaacgagga actgaaaaag atcgagaaga agcctttcat caacctgaag	1080
gccttctacg agatcaaaaa ggttgttgcc aactacgccg agagaaccaa cgaggccttc	1140
tccacctgg actacgacgc catcgcttac gctctgacag tgtacaagac agacaaggat	1200
attagaagtt atctgaagaa aagcaacaac ctgagcaaac ggtgctacga cgaccaactg	1260
atcgaagaac tcttcacct gagctatacc aagttcggcc acctgagctt caaggccatc	1320
aacagagtgc tgcctatcat gcaggagggc agaacctacc aggaggccat ccagcaactg	1380
ggctacgaca ccaccaacct gaagaaagag aatagaagca tgttcctgcc tetgatcccc	1440
gatgagatca ccaatcctat cgtgaagcgg gcctgaccc aggccagaaa agtggatgaac	1500
gccatcatcc gccgctacgg atctccaaac tccgtgcaca tcgagctggc tagagagctg	1560
tcaaagagcc acgatgagag aaagaagatc atgacagccc acgacgaaaa ctacaagaag	1620
aataaggcg ccatcagcat cctgatcgag aacggcatcc tgaacccac aggtacgat	1680
atcgtcgggt acaaactgtg gaaggagcag ggcgagcggg gcgcctactc cctgaaggaa	1740
attccacctg acactttttt caacgagctg aaaaaagaac gcaacggcag tagcatcctg	1800

gaagtggacc acatttctccc ctactcccag tcttttatcg atagctacca caacaaagtg 1860
ctggtctaca gcgacgaaaa tcggaacaag ggcaacagaa tcccttacac ctacttcctg 1920

gaaacaaaca aagattggga agccttcgag agatacgtgc ggagcaataa gctcttcagc 1980
aagaaaaaga gagagtacct gctgaagaag gcttacctgc cacgggagtc tgaactgatc 2040
aaggaaagac acctgaacga caccagatac gccagcacct tctgaagaa cttcatcgag 2100
cagaacctgc agttcaagga agtgggaagt aacctgagaa aaaagcgggt gcaaaccgtg 2160
aacggcgtga tcaccgccca cctcagaaag cggtagggcc tggaaaagaa ccggcaggag 2220
acctacctgc accacgttat ggacgttatt atcgtggctt gtacagacca tcacatggtc 2280
accagaatca cagagtacta ccagatcaag gagagcaaca agtccgtgaa gaaaccttat 2340

ttcccatgc cctgggaggg cttcagagat gagctgctga gccatctggc tagccagcct 2400
atgcccaaga agattttctga agagctgaaa gccggatacc agagcagcga ttacatcttc 2460
gtgagcagaa tgcctaagcg gtccgtgacc ggccgccccc ataactcagat gatccggaga 2520
aaggcgagaa tcgacaaaaa gggcaagacg atcatcatta agagagttag actgaaagat 2580
atcaaattcg atgagaacgg cgatttcaag atggtgggaa aagaacaaga cctggccaca 2640
tacgaggcca ttaaacagag atacctggat cacggcaaga acctcaaaaa ggcctttgaa 2700
accctctgt acaagccttc taaaaaaggc accggaacc tgatcaagcg ggtgaagatc 2760

gaggacaga ccaaggcatt cgtgagagag gtgaacggcg gaggggcca aaattctgac 2820
ctcgtgcggg tggacctttt cgagaaggat gacaagtact acatggtgcc catctacgtg 2880
cctgatacag tgtgtagcga gctgcctaag aaagtgtca aaagtggcaa gggctatgag 2940
cagtggctga cctggacaa cagcttcacc ttcaagttct cctgtaccc ttacgactta 3000
gtgagactgg tgaaggcgga cgaagataga ttctgtact ttggcacact ggatatcgac 3060
agcgaccggc tgaacttcaa ggacgtgaat aaaccttcta agcagaacga gtacagatat 3120
agcctgaaga ccatcgagga cctcgaagaa tacgaggtgg gcgtgctggg cgacctgcgg 3180

ctggtgcgaa aagaaactcg gcggaatttc cacgagatca aaattaac 3228

<210> 415
<211> 3276
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 415

atgcggtaca gcatcggact ggatatcgga accacatcta tcggaaatgc cgtgatcaac	60
aaggatctgc agaggtttga gcacctgggc gtgagaatct tcgatgccgc agagaacccg	120
aaagacggct ctctctgag cgccccctaga agactggccc ggagcagcag acggcggtg	180
cggagaagaa agcaccgggt cgagcggacc aaacaactgc tgatcaacaa gggtctgctg	240
accaagaccc aggtgaagaa cctgtacaac agcaaaaaca tcaacctgga catctgggac	300
atccgggtgt caggcatcga tagaaagctg ttcaacaacg agttcccgag ggtgctgac	360
cacttcagca aaaaccgagg attcaagtcc aacaggaaaa gcgagctgaa agaagatgac	420
aatggcgcta tcttgagcgc cgtcaaggaa aacagagagc tgatggacga gaagggtac	480
agaacaatcg ccgagatgct ggtgtccgat gagaagtacg agggcaccaa gagaacaag	540
ggcggcgatt acagccacgt ggtggcccga agcgacatcg agaacgaaat ctgcctgctg	600
ttccagaagc agagggagta cggccacccc ttccgacag aggagaacga ggaagccttc	660
ctgagcatct ggagcagcca gagacctttt agcaccaaag acgacatcgt caagaagatc	720
ggcaactgca ccttgaacc taaggaaaag agggccccta agtccactta caccttcgag	780
cggttcagag ccttgacaa gctcaacaga ctgagaatcc tgtctaccac cggcccttct	840
cggccctga ctaacgaaga aagaaagagc atcctgagca gcctgttctc caagaaggaa	900
gtgaaataca aggaactcg gaagctctc aaactgacag acgaccagcg gttcaacgag	960
atctactact ctccagatga gacaatcgag aaaaccgaga atagaacatt cctgagcctt	1020
gagtcccagt acaagatcaa gaagatcatc gagaagaccg aaagcaagaa tatgcagagc	1080
agctaccacc ctattgacta cgacaccatc ggctatgccc tgaccgtgtt caaagacgac	1140
aaggacatcc aacactacct gcagaacagc tacatcgaca gcaagggcaa ggccatcccc	1200
aacatggcca acagagagta caacctggaa ctgatcgagg agctgcttgg cctgagcttt	1260
gccaaattcg gccacctgag cctcaaggct ctgaacaaca tcctgcctta tatggaagag	1320
ggcgagccct atcacatcg ttgtgaaatg gctagctacc agttcagcca gcggtgtcc	1380
aaagaaaagc acagactgct gcctccaatc ccagtggacg agatcccca ccccggtgtc	1440
gtgcggggcc tgacgcaggt gcgcaaagtg ctgaacagca tcatcaagaa gtacggccct	1500
cctagcgaca tctacatcga gctagccaga gaaatgagca agcccttcaa ggagagaaaa	1560
tctctgaaa gagaattcaa cgagaaccgg cagatcaacg agaaggcaa ggcccatc	1620
agcgagctgt atagaatccc caatgacccc agacccacg acatcctgaa gttcaagctg	1680
tggaacgagc aaaacggcat ctgccctac tcctgaagc ctatttctat cgagtacctg	1740
ttcaacatcg ggtacgccga ggtggaccac attatcccat acagccggag cttcgacgac	1800

agcaacggaa acaaggtgct ggtcctgacc agagaaaacc agaacaaact gaaccggacc 1860
ccttatgagt ggttcggcca tgaggaaaac cgggtgggagg atttcgtgtc gtttatcagg 1920

acaatgaagg tgggcaagaa gaaaaaaaaat atgctgctga aaaagaattt cgacgaagag 1980
caagaggagc agatcctgtc tagaaatctg aatgacacca gatacatcac aagatacatc 2040
aagagcttca tcgaagataa cctggagtgc cggacagaag aaaacaagga acagtacgtg 2100
cacaccgtga atggggccta tacaagcctg atgagaaaga gatggggcct gaacaaggac 2160
cggagaggaa atgatctgca ccacgccgtg gacgccgcca tcattgccgt cagcctccct 2220
ttcaagaaca aggtgaacgc ctactttaag agacaggaga ccggcctgag caagctgctg 2280
aacaacaaga aggacatctt ccccgagcct tggcgggaact tcatcaagga actggaagcc 2340

agaatgatcc aggaccctga gaagatgaag agagccctgg agagcctgga gctggaacaa 2400
tacggagaaa ttttctgaa caagctgaag cccatctttg tgtccagaat gcctaagcat 2460
tctatcaaag gccagatcca cgaggaaaca attagacggg tgcggggcct taccgaggaa 2520
ggctttcttg tgaccgtgaa gaagaccaga ctggatcaga tccctttcga caagaacggc 2580
gacttccta tgtacggcaa ggaaaccgac atcaagacct acatggccat caagcagcga 2640
tacctggaat acggccagga caaacaaaag gcttttctg ttcctctgag aaagcctagc 2700
aaaaacccca aaaatgcccc tatcgtgcgg agcgtgaaga tcgagggcaa ggccaataga 2760

gtggtgatgc tggacgataa ggccgctgcc gacaacgcca gcattgtgcg gaccgaagtg 2820
tttagacaca agaagaccgg cgagtactac ctgacacctg tgtacgtggc cgatatcctg 2880
tccaataaaa tcccagacag actgattacc atcaagaaga gctacagcga ctgggataga 2940
atcacagacg aacacgagta tctgttttct ctgtacaaca acgacctcgt gaaaattatc 3000
ctgcccaagg agaaagagac aaagaagtac accggaggca atcacctctg gcaggagggt 3060
ttcttctact tcaagggagt tgatagctcc aacgccggca tcaagataat caaccacctg 3120
aattctttcg aggccaggat cggcaccaag cggctgattg catttgagaa gtatcaggtg 3180

aaccctctgg gcgaaatcaa caaagtgcac ggcgagaaac ggcttggcga gttgctgaat 3240
aaagaagaga tcaaagaaaa tagaaagaat atctct 3276

<210> 416
<211> 3450
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic

polynucleotide"

<400> 416

atgtacagca tcggcctgga cctggggatc agcagcgtgg gctggagtgt tatagatgag	60
agaaccggca acgtgatcga cctcggcatc agactgtttt cagccaagaa cagcgaaaag	120
aatctggaac ggcgaccaa caggggcgga cgacggctga tccggcgaaa gaccaaccgg	180
ctgaaggacg ccaagaagat cctcgccgcc gtgggcttct acgaggacaa gaggcctcaaa	240
aactcttgcc cttatcaact gagagtgaag ggcttgacag agcctctgtc taagggggaa	300
atctacaagg tgacctcca taccctgaag aagcggggca tcagctacct ggatgaagat	360
gatacagagg ccgctaagga gagccaggac tacaagagc aagtgcggga aaacgcccc	420
ctgctgacaa agtacacccc tggacaaatc caactgcagc ggctgaaaga aaacaacaga	480
gtgaagactg gcatcaacgc ccagggaat taccaactga acgtgttcaa ggtgtccgt	540
tacgccaacg agctggccac tattctgaag acacagcagg cttctaccc caacgagctg	600
accgatgact ggatcgccct gtctgtgcaa cctggcatcg ccgaggagc cggcctgac	660
tatagaaaac ggcttacta ccacggccct ggcaacgaag ctaacaatag cccctacggc	720
agatggagcg attttaagaa aacaggacag cctgctacca acatcttcga caagctgac	780
ggcaaagact tccaaggaga actcagagcc agcggactca gcctgagcgc tcaacagtac	840
aacctgctga acgaccttac aaacctgaag atcgacggcg agattcccct gagccctgag	900
cagaaggagt acatcctggc cgagctgatg accaaggagt tcaccagatt cggcgtgaac	960
gacgtggtga agctgctggg agtgaagaag gaacggctga gcggctggcg gctggataag	1020
aagggaagc ctgagatcca caccctgaaa ggctacagaa actggagaaa gatcttcgcc	1080
gagagcggga tcgacctgc tacactgcct acggagacca tcgactgcct ggccaaggtg	1140
ctgacctga aactgagag agaaggcatc gagaacaccc ttgcctttga gctgtctgaa	1200
ctggccgaga gcgtcaagct gctggtgctg gaccggtaca aggaactgag ccagagcgtg	1260
agcaccagc cctggcacag attttctctg aaaacctgc acctgctgat cctgagctg	1320
atgaacgcca ccagcgagca gaacacctt ctggaacagt ttcaactgaa gtccgatgtg	1380
cggaaaagat acagcgagta caagaagctg cctacaaagg atgtcctgac cgaaatctac	1440
aacctaccg tgaacaaaac cgtagccaa gccttcaagg tgatcgatgc cctgctctg	1500
aagtacggaa aagagcagat cagatacatt accatcgaga tgcccagaga tgacaacgag	1560
gaggatgaga agaaaagaat taaggaactg cagccaaaa acagccagag aaaaaacgac	1620
tcacagagct acttcatgca gaaaagcgga tgggtccaag agaagttcca gacaaccatc	1680
cagaagaaca gaagatttct ggctaagctg ctgtactact acgagcagga tggcatctgc	1740

gcctacaccg gcctgtctat ctctccagag ctgctgggtg cgcactccac cgagatcgac 1800
cacattatcc ctatcagcat cagcctggac gacagtatca acaacaaggt gctgggtgctg 1860

tccaaggcca accaggtgaa aggccagcag accccttacg acgcctggat ggacggatct 1920
ttcaagaaga tcaacggcaa gttagcaac tgggacgatt accagaagtg ggtcgaaagc 1980
tgtcacttca gccacaagaa agagaacaac ctgctggaaa ccagaaacat ctttgactcc 2040
gagcaagtcg aaaaatttct ggctagaaac ctgaacgaca ccagatacgc ttctagactg 2100
gtgctgaaca ccctgcagag cttcttcgct aaccaggaga caaaagtgcg ggtgggtgaat 2160
ggcagcttca cccacacctt gcggaaaaag tggggcgccg acctcgacaa gacaagagag 2220
acacaccacc accatgccgt tgatgccacc ctgtgcgccg tgaccctttt cgtgaaggtg 2280

tctcggatc actacgcggt gaaagaagaa acaggcgaga aggtgatgag agagatcgac 2340
ttcgagaccg gcgagatcgt ggacgaaatg tcttatagag aattcaagaa gagcaagaag 2400
tacgagcgga agacgtacca ggtgaagtgg cctaacttca gagagcaact gaagcccgctg 2460
aacctgcac ctagaatcaa gttagtcac caggtggaca gaaaagccaa tagaaaactg 2520
agcgacgcca cgatctatag cgtgcgggaa aagaccgagg tgaagacct caagttagga 2580
aaacagaaaa taacaacaga cgagtacacc atcggcaaga tcaaggacat ctacaccgtg 2640
gatgggtggg aggccttcaa gaaaaagcag gacaagctgc tgatgaaaga cctggacgag 2700

aagacctacg aaagactgct gtccatcgcc gaaaccaccc cgcacttcca ggaggtggaa 2760
gagaagaacg gcaaggtgaa gcgggtgaag aggtcccat tcgccgtgta ctgcgaggaa 2820
aatgacatcc ccctatccg gaaatcgcc aagaaaaaca acggccccct gatcagatct 2880
ctgaagtatt acgacggcaa actgaacaag cacatcaaca ttacaaaaga tagtcaaggc 2940
agacctgtgg aaaaaaccaa gaatggcaga aaggtgacac tgcagagcct gaaaccctac 3000
agatatgata tctaccaaga cctggaaacc aaggcctact atacagtga actgtactac 3060
agcgacctgc gggtcgtgga aggcaagtac ggcataacag agaaagaata catgaagaag 3120

gtggccgagc agaccaaggg ccaagtgtt agattctgct tcagcctgca aaagaacgat 3180
ggactggaaa tcgagtggaa ggacagccaa tgttacgacg tgcgatttta caacttcag 3240
tctgctaatt ctatcaactt caaggcctg gaacaggaga tgatgccagc cgaaaatcag 3300
ttcaagcaaa agccataata caacggggcc atcaacctga atatcgctaa gtacggcaag 3360
gaaggtgaaga aactgagaaa gttcaacacc gacatcctgg gcaaaaagca ctacctgtac 3420
tacgagaagg aaccaagaa tatcatcaaa 3450

<210> 417

<211> 4287

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 417

atgatgatca aacacattct gggcctggat ctgggcacga acagcatcgg atgggcctc	60
atcaagcaaa atttcgagaa caaatatgga gagatcctgg gaatgggcag cagaatcata	120
cctatgagcc aagatattct cggcgagttc ggcaaggga atagcgtgtc acagaccgcc	180
gctagaaccg actaccgggg catccgccgg ctgagagaaa ggtttctgct gcggagagaa	240
cggctgcaca gaatcctgaa cgtgctgaac ttctgcctg aacactacgc cagccagatc	300
gacttcgaca aacgcttcgg caagtttaaa gtcgagacag aacctaagct ggcctggaaa	360
aacagcgacg gcaaattctc attcctgttc cagaccagct tcaacgagat gctggaggac	420
ttcaaggccc acggccagga cctgaagggt ccctatgact ggaccatcta ttacctgcgg	480
aaaaaagctc tgagccagaa aatcgaaaa gaggaactcg cctggattct gctgaacttc	540
aaccagaaga gaggtacta ccaactgaga ggagaagagg aagaagaaaa tcccaataaa	600
ctgatcgagt tctacagcct gaaaatcatc gacgtgctgg ccgacgagcc tcagaagggc	660
aatctgata tctggtacag ccttgcctt gagaacggct ggatctacag aagatcctcc	720
aaaacaagcc tgctggactg gaaggataag atcagagatt ttatcgtcac caccgatctt	780
aacgacaacg gctccgtgaa gacagacaag gagggaaatg aaaagcggag cttcagagcc	840
cctggcgaga acgactggac acttgtaag accaaaaccg agcaggagat agaccgtcc	900
agaaagaccg tcggcacgta catctacgag aactgtctgc agaaccceaa gcagaagatc	960
aagggaagc tcgtgagaac aatcgagaga aaattctaca aggaggagct gaaacagatc	1020
ctggaaaagc aaaaggaatt ccactacgag ctgcagtcgg acgatctgta cgacgactgc	1080
atccgggagc tgtacagaaa caacgaggcc catcaactga cctcagcaa gaaagacttc	1140
gtgcacctgt tcatggaaga tataattttc tatcaaagac ctctgaagtc ccagaaaagc	1200
agcatatcta actgcacct ggagttcaga aagtacaaag acgagaacgg cgtggagcac	1260
accagttcc tgaaggccat tcctaagagc aaccctatt atcaggaatt cagaatttgg	1320
cagtggattt tcaatctgaa catctacaag aaggacaacg atgacaacgt gaccaaggaa	1380
ttctgagca ccaccgaaga cttcgagaac ctgttcgagt ttctgaataa tcggaaagag	1440
atcgatcaga aggtctttct gaaacacttc aagctcaatg agaaaactca cagatggaag	1500

tacgtggaag ataagaagta tccttgcaac gagacaaaaa ccatgatttc cgagagattg	1560
aaaaaggtgg aaaatatctc cgacgacttt ctgacaagag gcatggaaca gaagatctgg	1620
cacatcatct actccgtgaa tgacaaaacc gaatacgaga aagctctgaa gagctttgcc	1680
gaaaagaaca atctggatga gaatagcttc ttcgaggcct ttcggaaatt tccacctttc	1740
aagtccgaat acggctcttt ctccgaaaag gccatcaaga agctgctccc actgatgaga	1800
ctgggcaagt attggagcta cgccaacatc gacctctact caaagaacag aatccagaaa	1860
atcatcaccg gcgaattcga cgagaacatc aaggatagag tgcgggagaa agcaatccac	1920
ctgaccgccg agaatgactt ccagggcctg caactgtggc tggcccagta catcgtgtac	1980
gggagacata gcgaggcaac catgatcggc aagtggaaact ctgccgatga tctggaggaa	2040
tttctgaagg agtttaagca gcacagcctg cggaacccaa tcgtggagca ggtgattaca	2100
gaaacccctgc gtgtggtgaa ggacatttgg ctgaaatacg gaaacggcgc caaggacttc	2160
tttaacgagg tgcacatcga gctgggaaga gagatgaagc agacaaagga cgaacgggct	2220
aatgccacaa agacaatcac cgaaaaatgaa aacaccaacc tgagaatcaa agccctgctg	2280
gctgagatga tgaacgacca cagcgtggaa aacgtgcggc cttattctcc tatgcagcag	2340
gagatcctga aaatctacga ggacggcatc ctgaaatcgg atatcgagat cgatgacgac	2400
atcctgaaga tcagcaaaac cgcccaacct agcagcagcg atctgaaaag atacaaactg	2460
tggctggagc agaaatacaa gtccctttac accggacaga tcatcccct gaataagctc	2520
ttcaccctg agtacgaaat cgaacacatc atccctcagt ccagatactt cgacgacagc	2580
ctgagcaaca aaattatctg cgagagcgcc gtgaacaagc tgaaagataa ctacatcggc	2640
ctgggtttca tcaagcaatt cggcggcaca atcatcgagt gcggcttcgg caagagagtg	2700
aagggtttca aggccgagga atacaggagaa ttcgtgaaga agcactacgc caacaaccgg	2760
ggcaagagaa acaagctgct gctggaagac atcccagaga agatgatcga gcggcagatg	2820
aacgacacc ggcatatctc taagtacatc agcggcattc tgtcccatc cgtccgggtg	2880
gaagacggca ccgacgagg cgtaacagc aaaaatgtga tccccggcaa cggcaaaatc	2940
accacacaac tgagacagga ctggggactg aatgatgtgt ggaacgagct tgtgctgccg	3000
cgattcgaga gaatgaacca actgaccaac agcacagatt ttacaagctg gaacgagaac	3060
caccagaagt acctgcctac agtgcctgtg gagtttagca agggattctc taagaaaaga	3120
atcgaccacc ggcaccacgc cctggacgcc ctggtgatcg cctgcgccac caaggatcat	3180
gtgaacctgc tgaacaatca gagcgccaag agcgacacga agcgatacga tctgaagaag	3240

aagctgatga agttcgagaa gacagtgtac aaggaccccc agaccgagaa gcgaatcgag 3300

cgagaagtgc caaagtactt tctgaagcct tgggagactt ttaccgtgga cgccaagaac 3360

aagctggaaa ccatcatcgt gtccttcaaa cagaacctca gagtcatcaa caaggccaca 3420

aactactacg agaagtacgt ggacaaggac ggcgtgaaga cgaaggagcg ggttgagcag 3480

actggaacaa actgggccat cagaaaacct atgcacaagg aaaccgtgtc cggcaaaatt 3540

gacctgcctt gggtgaaagt gcccagggc aagatcctga ccgccaccag aaagagcctg 3600

gacaccagct tcgacctcaa ggctatcgcc agcatcacgg acaccggcat ccagaagatc 3660

ctgaagaact acctggagtt caaggaatcc cctgagctgg ccttcagccc tgagggcac 3720

gaggacatga acaagaacat caaaaagtac aacggcggaa agccccacca gcctatctct 3780

aaggtgcgcg tctttgagct gggcagcaag ttccaagtgg gccagaccgg caataagaag 3840

gacaagtacg tggaagccgc caagggaacc aacctgttct tcgccatcta tgaggacaga 3900

aagggcaagc gctcctacga gaccatccct ctgaacgagg tgatcgagag acagaagcag 3960

ggtctgtctg tgggtggacct gaagaacatt aacgacttct tcctgtgtcc taacgacctg 4020

gtgtacattc ccagcggcga cgagttggaa aacggcggct ccatagagat caaaaacatc 4080

accaaagaaa agtccgaaag attttataag gtggtctctt tttctggcag ccaaatttc 4140

ttcgtcgggc acgacattgc cgtgtccatc gtgaacaagg gggaattctc caccctaaac 4200

aagatggaac gcgccatcga tggctctatg gtgaaggaat cttgtatcaa gctgaagatc 4260

gacagactgg gcaacgtgct gaaagcg 4287

<210> 418

<211> 4296

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 418

atggccaaga atatcctcgg cctggacctg ggccaacaac gcatcggtcg ggccttatac 60

aaccaagact tcgagaacaa gcagggttaag atcctcggaa tgggcagccg gatcattcca 120

atgagccaag atatcctggg cgacttcggc aagggaact ctgtgtcca gaccgccgt 180

agaaccggct accggggcgt gcggagactc agagagcgg tttctgtgag aagagagaga 240

ctgcaccggg tgctgaatat aatcaacttc ctgccgaac attacgcctc tcagatcgac 300

tttgagaaaa gattcggcaa gttcaaggaa gaaaccgagc ctaagctggc ctacaacaaa	360
gacggcttcg tgitcaagga tagctttgag gaaatgctgg ccgattttaa gaactaccag	420
cctcaactgc tggaaaacga caagaaaatt ccctacgact ggaccatcta ctacctgaga	480
aagaaggccc tgitccagaa aatcgagaag gaagagtgg cctggattct gctcaatttc	540
aaccagaaga gaggttatta ccaactgagg ggcggaagact tcgaggaaga gaaggacaaa	600
atgttcgtcc gcctgaaggt ggagaagatc atcgatagcg gcgacaacgt gaagggttaag	660
attctgtacg acgtctactt cgagaacggc tggaaatagc acaaacagat cgttaagacc	720
gaggactgga tcgaaagagt gaaggaattc atcgtcaccg aatctttcct gaagaatggc	780
gacatcaaga gaacatttaa ggccgtggac tccgagaagg attggatcgc catcaaaaca	840
aaaacagaac aagagattga caagtccac aagaccgtgg gcgtgtacat ctacgaaaca	900
ctgctccaga accctaagca gaagattaag ggcaagctgg tgagaacaat cgagcgcaag	960
ttttacaagg acgaactgaa gcagatcttg gagaagcaga aagagttcca ccaggagctg	1020
aagaacgatg acctgtataa cgactgcgtg agagaactgt acagaaaca tgaagctcat	1080
caactgaccc tctctaagaa ggagtttgtg cacctgtga tggaagatat cattttctac	1140
caaagacccc tgaggtccca gaagagcagc atcagcaatt gcagcctgga attcagaaag	1200
tacaaggacg agaacggcgt ggagcatatc caatatctga aagccgtgcc taagagcaat	1260
ccttactacc aggagtttcg gatctggcag tggatcttta acctgaacat ctacaagcgg	1320
gacgatgagc aaaacccgtg gaccacagag ttctgaaca ccaccaccga tattgaaaac	1380
ctcttcgagt tctcaacaa ccggaaggaa gtggatcaga aagccctgct gaaacatttc	1440
aagctgaacg agaagaccca cagatggaag tacgtggagg ataagaaata cccctgcaac	1500
gaaaccaaga gcatgatttc cgagcggctg aagaaggtgg agaacatagc taacgacttc	1560
ctgaccagag agattgagca gaaaatctgg catatcatct attctgtgaa cgacaaaatc	1620
gaatacgaaa aagccctgaa gagttttgcc aagaagaaca acctggacga gaactccttt	1680
ttcgaggctt tcaagaaatt ccccccttc aaaagcgagt acggcagttt cagcgaaaag	1740
gccatcaaga agctgctgcc tctgatgcgg cttggcaagt actggaactg cgagaacatc	1800
agcgacgaca gcaaggaaag aatccagaag attatcaacg gcgagtacga cgagaatatc	1860
aaggatagag tgcgcgagaa agccattcac ctgaccagcg aaaacaactt ccagggcctg	1920
cagctctggc tggcccagta cgtggtgtac gacagacatt cggaagccag catgatcggc	1980
aagtggaata gcgccgacga cctcgaagaa ttctgaagg agttcaaaca gcacagcctt	2040
agaaacccta ttgtggagca ggtgatcaca gagacactgc ggggtggtgaa ggacatctgg	2100
ctgaagtacg gcaatggagc caaagatttc ttcaatgaga ttcacatcga actgggaaga	2160

gagatgaagc aaaccaaaga cgagcgctc gatgctacga agcggatcac cgagaacgag 2220

aataccaatc tgagaatcaa ggctctgctg gctgagatga tgaatgataa cagcgtggaa 2280

aacgtccggc cttacagccc tatgcagcag gagatcctga agatctacga ggatggcgtg 2340

ctgaatagtg gaattgagat tgaagatgaa tacctgaaga tttctaagac cgcccagcct 2400

agcccttctg acctcaagg gtacaagctg tggctggaac agagatacaa gagcccatc 2460

accggccaga tcattccct gaacaagctg ttaccctg aatacgagat tgagcacatc 2520

atccctcaga gcagatactt cgacgacagc ttacgaaca agatcatctg tgaaagcgcc 2580

gtgaacaagc tgaaggacaa ctatatcgga ctgggcttca tcaagcagtt cggcggaacc 2640

atcgtggagt gcggcctggg aaaaaacgtg aaggtgttg aagtgaacga gtacaggac 2700

ttcgtgaaga aacactacgc caacaacaga ggaaagagga ataagctgct cctggaggaa 2760

atccctgaga aaatgatcga gaggaactg aacgacacca ggcatatcag caagtacatc 2820

tccggcgtgc tgagcaacat cgtgcgggtg gaagacggct ctgacgagg agtgaacagc 2880

aaaaacatcg tgccaggcaa tggcaagatc accaccaac tgaagcagga ttggggactg 2940

aacgatgtgt ggaacgacct gatcctgcct agattcgaga gaatgaacca actgaccaat 3000

tcaacagact tcaccgctg gaacgaaaat caccagaagt ttctccctac agtgcccatc 3060

gaatttagca agggatttag caagaaaaga atcgaccaca gacaccacgc cctggacgcc 3120

ctcgtgatcg cctgtgccac caaggaccac atcaatctgc tgaacaatca gtctgcaaaa 3180

agcgacacca aacggtacga cctgaagaaa aagctcatga aattcgagaa gggcgtgtac 3240

aaccaccctc agaccggcga gagaatccag cgggacgtgc cgaagcaatt tctgaagcct 3300

tgggagagct tcaccatcga cgctaagaac aacctggaca agatcatcat cagcttcaag 3360

cagaacttga gggatgatcaa caaagccaca aactactacg agaagtatgt ggaaaagaac 3420

ggcatcaaga caaagagag agtcgaacag acaggcaca actgggcat cgggaagccc 3480

atgcacaagg aaaccgtgtc tggcatcgtt aacctgcctt gggtgaaagt gcctaaaggc 3540

aaaatcctga cggccaccg gaaaagcctg gacaccacat ttgacctgaa aagcatcaac 3600

tccataacag ataccggcat tcagaagatc ttgagaaact acctggaatt caaaggcagt 3660

cctgaactgg ccttcagccc agagggcac gaggacatga acaagaacat cgagaaatac 3720

aacgatggca agctgcacca acctataaac aaagtgagag ttttcgagct gggcagcaaa 3780

ttccaagtgg gacagaccgg caacaagaag gacaagtacg tggaggccgc caagggcacc 3840

aacctgttct tcgccgtcta cgaggacaag aacggcaaaa gaaactacga aacaatccct 3900

ctgaacatag tgatcgagag acagaagcag ggctctctgg cctgccccga gagtaatgaa 3960
aaaggcgaga agctgctgtt ccaactgagc ccaaacgact tcgtgtatct cagcacagag 4020
gaagaaaatg ataacgccac acctatcaat ttctctctgc tgagcaagga acagattaac 4080
aacctgtaca aaatcgtgtc ctttaccggc aacagactgt acggcatccc tatctgcgtg 4140
gccaccacaa tcgtgaacaa agccgagtag acccaactga acaagatcga gttcaccaaa 4200
gagaaagatc tgctgctgaa gctcaatgtc aatagactgg gcgacgtgaa gacctttaca 4260
gccaatgaca tcagaaagat tttcaaccgg caggag 4296

<210> 419

<211> 137

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 419

aagugggaga guacugaaug ggcacguugc gguuggccug cgauuucuga acaaagguuc 60
agaaaucgca guccagccgu uaacaagcug agauaugcac caaauaaggc gcucgcuucg 120
gcgggcgcuu uuucguu 137

<210> 420

<211> 137

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 420

accuucucc cauggcagua cauuguugc gguuggccug cgauuucuga acaaagguuc 60
agaaaucgca guccagccgu uaacaagcug agauaugcac caaauaaggc gcucgcuucg 120
gcgggcgcuu uuucguu 137

<210> 421

<211> 137

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 421

gcccacaacu gcgaaauga cucuaguugc gguuggccug cgauuucuga acaaagguuc 60

agaaaucgca guccagccgu uaacaagcug agauaugcac caaauaaggc gcucgcuucg 120

gcggggcgcuu uuucguu 137

<210> 422

<211> 137

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 422

ugggggcgcu caguuccag aaccuguugc gguuggccug cgauuucuga acaaagguuc 60

agaaaucgca guccagccgu uaacaagcug agauaugcac caaauaaggc gcucgcuucg 120

gcggggcgcuu uuucguu 137

<210> 423

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 423

acagugcgca ucuccuggu caccguuuua guacucugua auuuuaggua ugaaaguc 60

accuaaaauu acagaaucua cugaaacaag acuauauguc guguuuaucc cacuaauuuu 120

uuagugggau uuu 133

<210> 424

<211> 131

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 424

auuuacagcc uggccuuugg ggguuuuagu acucuguaau uuuagguaug aaagucauac 60

cuaaaauuac agaauacuac gaaacaagac uauaugucgu guuuauccca cuaauuuuuu 120

agugggauuu u 131

<210> 425

<211> 134

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 425

ccuuggggcc cagacugagc acgugguuuu aguacucugu aauuuuaggu augaaaguca 60

uaccuaaaau uacagaaucu acugaaacaa gacuauaugu cguguuuuac ccacuaauuu 120

auuaguggga uuuu 134

<210> 426

<211> 134

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 426

cuucuccuca ggagucagau gcaccguuuu aguacucugu aauuuuaggu augaaaguca 60

uaccuaaaau uacagaaucu acugaaacaa gacuauaugu cguguuuuac ccacuaauuu 120

auuaguggga uuuu 134

<210> 427

<211> 132

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 427

gaugauuuu ucuuuuaggg ugcguuuuag uacucuguaa uuuuagguau gaaagucuaa	60
ccuaaaaaua cagaauacuac ugaaacaaga cuauaugucg uguuuauccc acuaauuuau	120
uagugggauu uu	132

<210> 428

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 428

ggcuggggcca uuaaaaccuc uccaguuuua guacucugua auuuuaggua ugaaagucau	60
accuaaaaau acagaauucia cugaaacaag acuauauguc guguuuaucc cacuaauuuu	120
uuagugggau uuu	133

<210> 429

<211> 206

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 429

cacaucucga gcaagacguu gucauaguuc cauuaagcc auugcuguuu uaugaaagca	60
ugauacagca auggcuuuga uguuucuaug auaagggcuu cggcccggugg cguuggggau	120
cgccugccca uuuuaaggg cuucucccca ucuauuuuau gagaaauuuu caaccuuggc	180
uauucuuaaa uagcuaaggu uuuuuu	206

<210> 430

<211> 206

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 430

cauggcagua cauuagagca gucauaguuc cauuaaagcc auugcuguuu uaugaaagca 60
ugauacagca auggcuuuga uguuucuaug auaagggcuu cggcccugg cguuggggau 120

cgccugccca uuuuaauggg cuucucccca ucuauuuau gagaaauua caaccuuggc 180
uauucuuaaa uagcuaaggu uuuuuu 206

<210> 431

<211> 206

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 431

ccugaaugcu gugcggcucu gucauaguuc cauuaaagcc auugcuguuu uaugaaagca 60
ugauacagca auggcuuuga uguuucuaug auaagggcuu cggcccugg cguuggggau 120
cgccugccca uuuuaauggg cuucucccca ucuauuuau gagaaauua caaccuuggc 180

uauucuuaaa uagcuaaggu uuuuuu 206

<210> 432

<211> 206

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 432

cuucuauagc cuccuucccc gucauaguuc cauuaaagcc auugcuguuu uaugaaagca 60
ugauacagca auggcuuuga uguuucuaug auaagggcuu cggcccugg cguuggggau 120
cgccugccca uuuuaauggg cuucucccca ucuauuuau gagaaauua caaccuuggc 180

uauucuuaaa uagcuaaggu uuuuuu

206

<210> 433

<211> 206

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 433

ggacagugcg caucuccug gucauaguuc cauuaagcc auugcuguuu uaugaaagca 60

ugauacagca auggcuuuga uguuucuaug auaaggcguu cggcccugg cguuggggau 120

cgccugccca uuuuaauggg cuucuccca ucuuuuaau gagaaauua caaccuuggc 180

uauucuuaaa uagcuaaggu uuuuuu 206

<210> 434

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 434

aagacauuac acucauaua caccgucgu aguccauua aagccaaagg gcuuugaugu 60

uucuaugaua aggguuucgg cccgugcggu cggggauccg cugcccauuc cgaugggcuu 120

cucccauuu auu 133

<210> 435

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 435

agagccauca ccaucacauc ccuaagucan aguuccauua aagccaaagg gcuuugaugu 60
uucuaugaua aggguuucgg cccguggcgu cggggaucgc cugcccauuc cgaugggcuu 120
cuccccauuu auu 133

<210> 436

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 436

aucuggaggg aacuuacagc auauggucau aguuccauua aagccaaagg gcuuugaugu 60

uucuaugaua aggguuucgg cccguggcgu cggggaucgc cugcccauuc cgaugggcuu 120
cuccccauuu auu 133

<210> 437

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 437

ccaccagcgc ugcucugccu ugagggucau aguuccauua aagccaaagg gcuuugaugu 60
uucuaugaua aggguuucgg cccguggcgu cggggaucgc cugcccauuc cgaugggcuu 120
cuccccauuu auu 133

<210> 438

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 438

gaacaacuca aauggaaaug aauaugucau aguuccauua aagccaaagg gcuuugaugu 60
uucuaugaua aggguuucgg cccguggcgu cggggaucgc cugcccauuc cgaugggcuu 120
cuccccauuu auu 133

<210> 439

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 439

gcuucucac uc uggcuuaca acccagucac aguuccauua aagccaaagg gcuuugaugu 60
uucuaugaua aggguuucgg cccguggcgu cggggaucgc cugcccauuc cgaugggcuu 120
cuccccauuu auu 133

<210> 440

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 440

ggccaaauc cagcugccuu ccuuggucau aguuccauua aagccaaagg gcuuugaugu 60

uucuaugaua aggguuucgg cccguggcgu cggggaucgc cugcccauuc cgaugggcuu 120

cuccccauuu auu 133

<210> 441

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 441

uaaaaugauu uggaagagca gagacgucau aguuccauua aagccaaagg gcuuugaugu 60
uucuaugaua aggguuucgg cccguggcgu cggggaucgc cugcccauuc cgaugggcuu 120
cuccccauuu auu 133

<210> 442

<211> 133

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 442

uccuguucca ucaccaucaa aaaaagucan aguuccauua aagccaaagg gcuuugaugu 60
uucuaugaua aggguuucgg cccguggcgu cggggaucgc cugcccauuc cgaugggcuu 120
cuccccauuu auu 133

<210> 443

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 443

aagacauuac acucauaua caccgcgau aauuccucug uaaaacuuaa agaagguuaa 60
uagaguuaau augguaaggc aaauagccgu ggcguugggg aucgccuauug uccgguuuua 120
ccggauccu cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 444

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 444

agcuguuugg gaggucagaa auagggccau aaauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaauagccgu ggcguugggg aucgccuaug uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 445

<211> 159

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 445

agguuuuaau ggcccagccu gccauaauc cucuguaaaa cuuaaagaag guuuauagag 60

uuuuuaggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg uuuuaccgga 120

ucucccuaaa ggugacuaac uuugguuagu caccuuuuu 159

<210> 446

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 446

auguacugcc augggaagaa ggugagccau aaauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaauagccgu ggcguugggg aucgccuaug uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 447

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 447

cagaagucag gcuggacacu gaggagccau aaauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaauagccgu ggcgugggg aucgccuug uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 448

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 448

cagggccucg auaaugagau aaauugccau aaauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaauagccgu ggcgugggg aucgccuug uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 449

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 449

caucuacuaa uuucuccca ccccgccau aaauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaauagccgu ggcgugggg aucgccuug uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 450

<211> 159

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 450
cauggcagua cauuagagca gccauaauc cucuguaaaa cuuaaagaag guuuauagag 60
uuuuuaggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg uuuuaccgga 120
ucucccuaaa ggugacuaac uuugguuagu caccuuuuu 159

<210> 451

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 451
cccuggcugu gcacauuccc uccuggccau aaauccucug uaaaacuuaa agaagguuaa 60
uagaguuaau augguaaggc aaauagccgu ggcgugggg aucgccuaug uccgguuuua 120
ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 452

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 452
ccgcuguuuc cauuccuau cauugccau aaauccucug uaaaacuuaa agaagguuaa 60
uagaguuaau augguaaggc aaauagccgu ggcgugggg aucgccuaug uccgguuuua 120
ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 453

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 453

ccuagagccc aagagaaccc acugagccau aaauccucug uaaaacuuaa agaagguuaa	60
uagaguuaau augguaaggc aaauagccgu ggcgugggg aucgccuaug uccgguuuua	120
ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu	164

<210> 454

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 454

cgcacagcau ucaggucgua gucccgccau aaauccucug uaaaacuuaa agaagguuaa	60
uagaguuaau augguaaggc aaauagccgu ggcgugggg aucgccuaug uccgguuuua	120
ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu	164

<210> 455

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 455

cggcacuauc auugacucag cugcugccau aaauccucug uaaaacuuaa agaagguuaa	60
uagaguuaau augguaaggc aaauagccgu ggcgugggg aucgccuaug uccgguuuua	120
ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu	164

<210> 456

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 456

cucaacauuc cagggcucaa gcgaugccau aaauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaauugccgu ggcuuggggg aucgccuau uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 457

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 457

gaaugcugug cggcucugcu uccaggccau aaauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaauugccgu ggcuuggggg aucgccuau uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 458

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 458

gcuaccaccu cugugccccc ccggcgccau aaauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaauugccgu ggcuuggggg aucgccuau uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 459

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 459

ggaggccugc agcuccugca ccuccgccau aaauccucug uaaaacuuua agaagguuaa 60

uagaguuaau augguaaggc aaauagccgu ggcguugggg aucgccuaug uccgguuuaa 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 460

<211> 159

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 460

gggccauuaa aaccucucca gccauaaauuc cucuguaaaa cuuaaagaag guuuauagag 60

uuuuuaggu aaggcaauau gccguggcgu ugaggauccg cuauugccgg uuuuaccgga 120

ucuccuuaa ggugacuaac uuugguuagu caccuuuuu 159

<210> 461

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 461

gggccauuaa aaccucucca ggggcgccau aaauccucug uaaaacuuua agaagguuaa 60

uagaguuaau augguaaggc aaauagccgu ggcguugggg aucgccuaug uccgguuuaa 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 462

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 462

guuuuaaugg cccagccuca caccgccau aauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaugccgu gccguugggg aucgccuaug uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 463

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 463

uaaaaugauu uggaagagca gagacccau aauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaugccgu gccguugggg aucgccuaug uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 464

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 464

uccguuucca ucaccauca aaaaagccau aauccucug uaaaacuuaa agaagguuaa 60

uagaguuaau augguaaggc aaugccgu gccguugggg aucgccuaug uccgguuuua 120

ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu 164

<210> 465

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 465

ugucucuuuu auaaaaaaaaa auaaggccau aaauccucug uaaaacuuua agaagguuuu	60
uagaguuuuu augguaaggc aaauagccgu ggcuugggg aucgccuaug uccgguuuua	120
ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu	164

<210> 466

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 466

uguugguuac cuccugcca ccaccgccau aaauccucug uaaaacuuua agaagguuuu	60
uagaguuuuu augguaaggc aaauagccgu ggcuugggg aucgccuaug uccgguuuua	120
ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu	164

<210> 467

<211> 164

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 467

uuuagguucg uaucuguaaa accaagccau aaauccucug uaaaacuuua agaagguuuu	60
uagaguuuuu augguaaggc aaauagccgu ggcuugggg aucgccuaug uccgguuuua	120
ccggaucucc cuaaagguga cuaacuugg uuagucaccu uuuu	164

<210> 468

<211> 118

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 468

aagugggaga guacugaaug ggcacguuuu uguacucuca aaaaaaaguu auugagaauac 60

uacaaaaua aggcauuuug ccgaauuuac cgcccuacau auguagggcg guuuuuuu 118

<210> 469

<211> 113

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 469

accggaggac aaaguacaaa guuuuuguac ucucaauaaa aaguuaauuga gaauacuacaa 60

aaauaaggca uuuugccgaa uuuaaccgcc uacauaugua gggcgguuuu uuu 113

<210> 470

<211> 113

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 470

ccacggcucc uccgaagcga guuuuuguac ucucaauaaa aaguuaauuga gaauacuacaa 60

aaauaaggca uuuugccgaa uuuaaccgcc uacauaugua gggcgguuuu uuu 113

<210> 471

<211> 113

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 471

cccugauccu cuugucccac guuuuuguac ucucaauaaa aaguuaauuga gaauacuacaa 60

aaauaaggca uuuugccgaa uuuaccgccc uacauaugua gggcgguuuu uu 113

<210> 472

<211> 113

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 472

cuggagcaac aaucugacu guuuuuguac ucucaaaaa aaguuuuuga gaaucaaca 60

aaauaaggca uuuugccgaa uuuaccgccc uacauaugua gggcgguuuu uu 113

<210> 473

<211> 113

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 473

gggggucggg gcucgcggcg guuuuuguac ucucaaaaa aaguuuuuga gaaucaaca 60

aaauaaggca uuuugccgaa uuuaccgccc uacauaugua gggcgguuuu uu 113

<210> 474

<211> 113

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 474

uggauuuuga gucucucagc guuuuuguac ucucaaaaa aaguuuuuga gaaucaaca 60

aaauaaggca uuuugccgaa uuuaccgccc uacauaugua gggcgguuuu uu 113

<210> 475

<211> 123

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 475

aagugggaga guacugaaug ggcacguugu gaauugcuuu caaaaaaguu ugaaaagcaa	60
uucacaauaa ggauuauucc guugugaaaa cauucaaggc ggggcaacuc gccuuuuuuc	120
guu	123

<210> 476

<211> 123

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 476

aggugauuca cagagacaua aagccguugu gaauugcuuu caaaaaaguu ugaaaagcaa	60
uucacaauaa ggauuauucc guugugaaaa cauucaaggc ggggcaacuc gccuuuuuuc	120
guu	123

<210> 477

<211> 123

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 477

auuuagcucu uaguccuggc aagccguugu gaauugcuuu caaaaaaguu ugaaaagcaa	60
uucacaauaa ggauuauucc guugugaaaa cauucaaggc ggggcaacuc gccuuuuuuc	120
guu	123

<210> 478
 <211> 123
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 478
 cugggagguc aagacugcaa ugagaguugu gaauugcuuu caaaaaaguu ugaaaagcaa 60
 uucacaauaa ggauuauucc guugugaaaa cauucaaggc ggggcaacuc gccuuuuuuc 120
 guu 123

<210> 479
 <211> 123
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 479
 gcccacaacu gcgauaauga cucuaguugu gaauugcuuu caaaaaaguu ugaaaagcaa 60
 uucacaauaa ggauuauucc guugugaaaa cauucaaggc ggggcaacuc gccuuuuuuc 120
 guu 123

<210> 480
 <211> 118
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"
 <400> 480
 aagcaacucu ccugucucag ccuccguugu gaauugcuuu caaaggaaag caauucacaa 60
 uaaggauuau uccguuguga aaacauuuag cgccucgacu aucuucgggg cauuuuuu 118
 <210> 481

<211> 118

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 481

gggaagaagg ugauucacag agacaguugu gaauugcuuu caaaggaaag caauucacaa 60

uaaggauuau uccguuguga aaacauuuag cgccucgacu aucuucgggg cauuuuuu 118

<210> 482

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 482

aagtgggaga gtactgaatg ggcac 25

<210> 483

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 483

accttcttcc catggcagta catta 25

<210> 484

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

oligonucleotide"

<400> 484

gcccacaact gcgataatga ctcta 25

<210> 485

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 485

tgggggcgct cagtttccag aacct 25

<210> 486

<211> 24

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 486

acagtgcgca tctccctggc cacc 24

<210> 487

<211> 22

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 487

atttacagcc tggcctttgg gg 22

<210> 488

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 488

ccttggggcc cagactgagc acgtg 25

<210> 489

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 489

cttctcctca ggagtcagat gcacc 25

<210> 490

<211> 23

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 490

gatgatattt tctttaatgg tgc 23

<210> 491

<211> 24

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 491

ggctgggccca ttaaacctc tcca 24

<210> 492
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 492
 cacatctcga gcaagacgtt 20
 <210> 493
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 493
 catggcagta cattagagca 20
 <210> 494
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 494
 cctgaatgct gtgcggctct 20
 <210> 495
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"

<400> 495

cttctatagc ctccttcccc

20

<210> 496

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 496

ggacagtgcg catctccctg

20

<210> 497

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 497

aagacattac actcatatta caccc

25

<210> 498

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 498

agagccatca ccatcacatc cctaa

25

<210> 499

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 499

atctggaggg aacttacagc atatg 25

<210> 500

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 500

ccaccagcc tgctctgcct tgggg 25

<210> 501

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 501

gaacaactca aatggaaatg aatat 25

<210> 502

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 502

gcttctactc ttggcttaca accca 25

<210> 503

<211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 503
 ggccaaaatc cagctgcctt ccttg 25

<210> 504
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 504
 taaaatgatt tggaagagca gagac 25

<210> 505
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 505
 tcctgttcca tcaccatcaa aaaaa 25

<210> 506
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"

<400> 506
agctgtttgg gaggtcagaa atagg 25

<210> 507
<211> 20
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 507
aggttttaat ggcccagcct 20

<210> 508
<211> 25
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 508
atgtactgcc atgggaagaa ggtga 25

<210> 509
<211> 25
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 509
cagaagtcag gctggacact gagga 25

<210> 510
<211> 25
<212> DNA
<213> Artificial Sequence

<220><221> source

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

<400> 510

cagggcctcg ataagagat aattt 25

<210> 511

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 511

catctactaa tttcttccca ctccc 25

<210> 512

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 512

ccctggctgt gcacattccc tcctg 25

<210> 513

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 513

ccgctgtttc cattcctcat cattc 25

<210> 514

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 514

cctagagccc aagagaaccc actga 25

<210> 515

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 515

cgcacagcat tcaggtcgta gtcce 25

<210> 516

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 516

cggcactatc attgactcag ctgct 25

<210> 517

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 517

ctcaacattc cagggctcaa gcgat

25

<210> 518

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 518

gaatgctgtg cggctctgct tccag

25

<210> 519

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 519

gctaccacct ctgtgcccc cgggc

25

<210> 520

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 520

ggaggcctgc agctcctgca cctcc

25

<210> 521

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 521
gggccattaa aacctctcca 20

<210> 522
<211> 25
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 522
gggccattaa aacctctcca ggggc 25
<210> 523
<211> 25
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 523
gttttaatgg cccagcctca cacc 25

<210> 524
<211> 25
<212> DNA
<213> Artificial Sequence
<220><221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 524
tgtctcaa atataaaaaa ataag 25
<210> 525
<211> 25

<212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 525
 tgtttggttac ctccctgccca ccacc 25

<210> 526
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 526
 ttttaggttcg tatctgtaaa accaa 25

<210> 527
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 527
 accggaggac aaagtacaaa 20

<210> 528
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 528

ccacggctcc tccgaagcga 20

<210> 529

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 529

ccctgacct cttgtccac 20

<210> 530

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 530

ctggagcaac aaatctgact 20

<210> 531

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 531

gggggtcggg gctcgcggcg 20

<210> 532

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 532

tggatttaga gtctctcagc 20

<210> 533

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 533

aggtgattca cagagacata aagcc 25

<210> 534

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 534

atttagctct tagtcctggc aagcc 25

<210> 535

<211> 25

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 535

ctgggagggtc aagactgcaa tgaga 25

<210> 536

<211> 25

<212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 536
 aagcaactct cctgtctcag cctcc 25
 <210> 537
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"
 <400> 537
 gggaagaagg tgattcacag agaca 25
 <210> 538
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 538
 gacaggtgt gtgtcagagt tcgcc 25
 <210> 539
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 539
 agtggacagg ctgtgtgtca gagtt 25
 <210> 540
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 540
 ctgcgattgc agaagatgac ctggg 25

<210> 541
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 541
 acgcgcggac tgcgattgca gaaga 25

<210> 542
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 542
 agtactcacg attcgggggt ataca 25

<210> 543
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 543
 agtactcacg attcgggggt ataca 25

<210> 544
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 544
 atcacgtgcc ttgtacactg tccca 25

<210> 545
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 545
 gtggggatca cgtgccttgt acact 25

<210> 546
 <211> 25
 <212> DNA

<213> Homo sapiens
 <400> 546
 ctttgatggg ttcacccgac cagtc 25
 <210> 547
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 547
 tgggttcac cgaccagtc cggca 25
 <210> 548
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 548
 ctgggacctg gttccatgac ccac 25
 <210> 549
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 549
 accaggaagt ggccatgcgc aagct 25

 <210> 550
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 550
 cacagatcga atagaagact ccttc 25
 <210> 551
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 551
 cacagatcga atagaagact ccttc 25
 <210> 552

<211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 552
 ccctaactga gcagctcagg ctgcc 25
 <210> 553
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 553
 caggctgccg tggtcacct gtcca 25

 <210> 554
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 554
 cttccatgaa gataacattc cccag 25
 <210> 555
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 555
 attccccagc tggaagaagt ttctc 25
 <210> 556
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 556
 ggtacaagct gcagtcattt ccttt 25
 <210> 557
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 557

gtcatttcct ttagggaatc atctt	25
<210> 558	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 558	
catcaccccc tagactatag ttagc	25
<210> 559	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 559	
tcacccccta gactatagtt agcca	25
<210> 560	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 560	
acttcttggc aagtcggtta agatc	25
<210> 561	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 561	
aagatctggg aatagacgaa gtaaa	25
<210> 562	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 562	
aaaatggctg gtagtacata tggaa	25
<210> 563	
<211> 25	

<212> DNA
 <213> Homo sapiens
 <400> 563
 tggctggtag tacatatgga acagg 25
 <210> 564
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 564
 caattgactg tcaccagccc ctact 25
 <210> 565
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 565
 agcccctact agggctaaga tgcag 25

 <210> 566
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 566
 catcagtcga gttccaacct tcgcc 25
 <210> 567
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 567
 atacactgtt tctatggctt ttctc 25
 <210> 568
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 568
 aggaagttga ttctaccaga tgatc 25

<210> 569
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 569
 accgtggaca atattcgctc catct 25

<210> 570
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 570
 ggaaatgctg ttagtcgtta tgctg 25

<210> 571
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 571
 taccatgaat agaacatttc ctttc 25

<210> 572
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 572
 gtgtcttact agccatttta atact 25

<210> 573
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 573
 ggacctacac aagcagcatc ttctt 25

<210> 574
 <211> 25
 <212> DNA

<213> Homo sapiens
 <400> 574
 tgggtggacct acacaagcag catct 25
 <210> 575
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 575
 aaagtgactc tctaattttc tatatt 25
 <210> 576
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 576
 ttttttggtaa tagaacatct ccaag 25
 <210> 577
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 577
 caccacggct gtcgacacca atccc 25

 <210> 578
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 578
 tattccacca cggctgtcga cacca 25
 <210> 579
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 579
 actttgcaac agtggaggaa agcct 25
 <210> 580

<211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 580
 aacagtggag gaaagccttt ggagt 25
 <210> 581
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 581
 aggtggctgc ttctttggtt gtgct 25

 <210> 582
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 582
 tggctccttg gaaagtgaga attcc 25
 <210> 583
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 583
 tcaaataagc agcaggaagc agacc 25
 <210> 584
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 584
 aaataagcag caggaagcag accca 25
 <210> 585
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 585

atgctggcat tgttaaatcc atccc 25

<210> 586

<211> 25

<212> DNA

<213> Homo sapiens

<400> 586

atccatccca agggggtgag tgtgg 25

<210> 587

<211> 25

<212> DNA

<213> Homo sapiens

<400> 587

gataggtagt acatgcacag ttgct 25

<210> 588

<211> 25

<212> DNA

<213> Homo sapiens

<400> 588

ccatcattag atgatagta gtaca 25

<210> 589

<211> 25

<212> DNA

<213> Homo sapiens

<400> 589

acaatttcaa cacaagctaa actag 25

<210> 590

<211> 25

<212> DNA

<213> Homo sapiens

<400> 590

taaactagta ggatatagtt cttca 25

<210> 591

<211> 25

<212> DNA
 <213> Homo sapiens
 <400> 591
 aatttcaaaa aatgttgcc tctct 25
 <210> 592
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 592
 tccaatagtc atttttacct acgat 25
 <210> 593
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 593
 tctgagtgtt tccctccttc ataaa 25
 <210> 594
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 594
 ccttcataaa caggccagat aattt 25
 <210> 595
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 595
 cagttaacag ttctttatgt gaagt 25
 <210> 596
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 596
 tgtgtattta caagaaagag cagat 25

<210> 597
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 597
 agttctactg tgctcataga taata 25

<210> 598
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 598
 acaatcttag atcttttatt gtgcc 25

<210> 599
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 599
 cataaacaat cttagatctt ttatt 25

<210> 600
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 600
 ctttaattagc atatcctgta tatct 25

<210> 601
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 601
 tcttcttaat tagcatatcc tgtat 25

<210> 602
 <211> 25
 <212> DNA

<213> Homo sapiens
 <400> 602
 tgctttcaga gagattagag gcaga 25
 <210> 603
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 603
 cctactgctg ctttcagaga gatta 25
 <210> 604
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 604
 ggtggatgac tcataccctc caatg 25
 <210> 605
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 605
 gttggctgat ggtggatgac tcata 25
 <210> 606
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 606
 ggtggatgga tcataccctc caatg 25
 <210> 607
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 607
 gttggctgat ggtggatgga tcata 25
 <210> 608

<211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 608
 tcactataga tggtcatat ggaaa 25
 <210> 609
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 609
 ataataccta tagatggctc atatg 25
 <210> 610
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 610
 aaagcattac attacgtaat catat 25
 <210> 611
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 611
 ctcaactaa aatacagcca agttt 25
 <210> 612
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 612
 acagttggac agaagattat tcttc 25
 <210> 613
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 613

gcagactgac agttggacag aagat 25

<210> 614

<211> 25

<212> DNA

<213> Homo sapiens

<400> 614

gaataaggct tctagtctct ttagt 25

<210> 615

<211> 25

<212> DNA

<213> Homo sapiens

<400> 615

ggcttctagt ctctttagtt gggcc 25

<210> 616

<211> 22

<212> DNA

<213> Homo sapiens

<400> 616

gcttacaata cgcaacttcc ac 22

<210> 617

<211> 25

<212> DNA

<213> Homo sapiens

<400> 617

ccttgataaa cctagttcct tttgt 25

<210> 618

<211> 25

<212> DNA

<213> Homo sapiens

<400> 618

tgtegetgct aacagtatgt tagct 25

<210> 619

<211> 25

<212> DNA
 <213> Homo sapiens
 <400> 619
 acctcagaac aagatggcag aaagt 25
 <210> 620
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 620
 ctcagaacaa gatggcagaa agtct 25
 <210> 621
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 621
 acaaatcaaa gagaagctgc aagtc 25
 <210> 622
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 622
 agtcattggca agtcctctgt ttagt 25
 <210> 623
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 623
 atttcatctg aaacatacag aaaac 25
 <210> 624
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 624
 gctctcgagc cagtttttct gatct 25

<210> 625
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 625
 taaggagctc tcgagccagt ttttc 25

<210> 626
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 626
 gtgagccac accattagtc agagg 25

<210> 627
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 627
 cacaccatta gtcagaggaa ggttg 25

<210> 628
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 628
 caagttctca gtaataagat cttgg 25

<210> 629
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 629
 tggttggaga gtacagccac accat 25

<210> 630
 <211> 20
 <212> DNA

<213> Homo sapiens
 <400> 630
 taccagctct cctagatata 20
 <210> 631
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 631
 ctaccctgaa aatccgaaag tgttt 25
 <210> 632
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 632
 aattcaccgt tgacattaac ttcaa 25
 <210> 633
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 633
 gcagactctc ttccgaggca acagc 25

 <210> 634
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 634
 gcagactctc ttccgaggca acagc 25
 <210> 635
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 635
 agagcgggtcc cggaagctgg cggag 25
 <210> 636

<211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 636
 tgattgagac tagtgagagg gtgca 25
 <210> 637
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 637
 ctcacgcatt ggctggctgc acaac 25
 <210> 638
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 638
 cacacatgga ctcacgcatt ggctg 25
 <210> 639
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 639
 tcctccatgt attccactcg aggga 25
 <210> 640
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 640
 tgtattccac tcgagggatg ggctg 25
 <210> 641
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 641

cgccagtaag tctgccttgc ttggt 25

<210> 642

<211> 25

<212> DNA

<213> Homo sapiens

<400> 642

gacattgcct acaactaccg ccagt 25

<210> 643

<211> 25

<212> DNA

<213> Homo sapiens

<400> 643

gccacccacc caccggcta ggctg 25

<210> 644

<211> 25

<212> DNA

<213> Homo sapiens

<400> 644

ggacggggtg gagctgacca gggag 25

<210> 645

<211> 25

<212> DNA

<213> Homo sapiens

<400> 645

gatcagctga tcacagcgt cctca 25

<210> 646

<211> 25

<212> DNA

<213> Homo sapiens

<400> 646

gaccggctca ctggactacg agacc 25

<210> 647

<211> 25

<212> DNA
 <213> Homo sapiens
 <400> 647
 accggtcac tggactacga gaccg 25
 <210> 648
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 648
 aactactact catggtgact caagc 25
 <210> 649
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 649
 cctaaactac tactcatggt gactc 25

 <210> 650
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 650
 tgcattctcct acaagtgagt ctgcg 25
 <210> 651
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 651
 tgcattctcct acaagtgagt ctgcg 25
 <210> 652
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 652
 tcaaccgtaa gttaaaaaga accac 25

<210> 653
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 653
 actcttcac t aaccgtaagt taaaa 25

<210> 654
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 654
 atactcattt atgattttgc caatt 25

<210> 655
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 655
 ttcattttta ttttcagtga agaac 25

<210> 656
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 656
 agccagcatg tgaatgaggg aggaa 25

<210> 657
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 657
 catctgggcc attacagcta ataag 25

<210> 658
 <211> 25
 <212> DNA

<213> Homo sapiens
 <400> 658
 ccctgggctc agcatggtct tcttg 25
 <210> 659
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 659
 ccctgggctc agcatggtct tcttg 25
 <210> 660
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 660
 ggggcagaac agcaggctca cctgc 25
 <210> 661
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 661
 acagtgcctc gaagctcctg cagca 25
 <210> 662
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 662
 gtgtgcccgc actcgaaaaa gcagc 25
 <210> 663
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 663
 gcccgcactc gaaaaagcag ccggg 25
 <210> 664

<211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 664
 gttttattttg ttagatgta tacgt 25
 <210> 665
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 665
 gttttattttg ttagatgta tacgt 25

 <210> 666
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 666
 ttacagtgga tggagaagac atcat 25
 <210> 667
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 667
 taacacatta ttacagtgga tggag 25
 <210> 668
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 668
 tattgacgtt cccaaaacca tccag 25
 <210> 669
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 669

gcggtctcag aggtcagaga tggtc 25

<210> 670

<211> 25

<212> DNA

<213> Homo sapiens

<400> 670

agataaatcg gtactgtgct tctgt 25

<210> 671

<211> 25

<212> DNA

<213> Homo sapiens

<400> 671

gtctggaccc tccctgacct ctgag 25

<210> 672

<211> 25

<212> DNA

<213> Homo sapiens

<400> 672

agttccctct gcgtgttctc ataaa 25

<210> 673

<211> 25

<212> DNA

<213> Homo sapiens

<400> 673

cgtgttctca taaacaggca tgtgt 25

<210> 674

<211> 25

<212> DNA

<213> Homo sapiens

<400> 674

gagtgtattg gtaagtgatc catcc 25

<210> 675

<211> 25

<212> DNA
 <213> Homo sapiens
 <400> 675
 cgtggggagt gtattggtaa gtgat 25
 <210> 676
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 676
 acttcaatct cactgtccaa ggtga 25
 <210> 677
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 677
 cttcaatctc actgtccaag gtgag 25
 <210> 678
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 678
 gcaggcgctg ctgggggccc gggct 25
 <210> 679
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 679
 ggcgctgctg ggggcccggg ctaag 25
 <210> 680
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 680
 agtttaacga agccctccgt gagag 25

<210> 681
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 681
 gtgaagttaa acgaagccct ccgtg 25

<210> 682
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 682
 aagctgctct gagggctccc aggtc 25

<210> 683
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 683
 gtcacggacg agaactccca cagtc 25

<210> 684
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 684
 tggggatgag gctgggttac tcgct 25

<210> 685
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 685
 tgtactcgct caggtccagc gtcac 25

<210> 686
 <211> 25
 <212> DNA

<213> Homo sapiens
 <400> 686
 acctcacctt gtgaaaaagt tatat 25
 <210> 687
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 687
 gtaaactctgt gaaagagaat ggatt 25
 <210> 688
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 688
 tgacagaaac actttttgac atagt 25
 <210> 689
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 689
 gaaacacttt ttgacatagt gtggt 25

 <210> 690
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 690
 acttcttggc aagtaggtta agatc 25
 <210> 691
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 691
 tggcaagtag gttaagatct gggaa 25
 <210> 692

<211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 692
 tcgctccata gatgcaacac ccgat 25
 <210> 693
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 693
 cttacagcta atctctagaa ttcca 25

 <210> 694
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 694
 tctttttctta cagctaattct ctaga 25
 <210> 695
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 695
 gacagcccac tgcctataag taaa 25
 <210> 696
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 696
 ggtccatccc tacgctgacc agtgt 25
 <210> 697
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 697

ccacttacat ttgcaaagtc tgatt 25

<210> 698

<211> 25

<212> DNA

<213> Homo sapiens

<400> 698

aaatgctttt gaagcaccac ttaca 25

<210> 699

<211> 25

<212> DNA

<213> Homo sapiens

<400> 699

ccgccccggc acccgctct gcgcc 25

<210> 700

<211> 25

<212> DNA

<213> Homo sapiens

<400> 700

ccgccccggc acccgctct gcgcc 25

<210> 701

<211> 25

<212> DNA

<213> Homo sapiens

<400> 701

gctgccagaa catccttggg ggcta 25

<210> 702

<211> 25

<212> DNA

<213> Homo sapiens

<400> 702

catggetgcc agaacatcct tgggg 25

<210> 703

<211> 25

<212> DNA
 <213> Homo sapiens
 <400> 703
 atgattctga attagctata tcgtc 25
 <210> 704
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 704
 gaattagcta tatcgtcaag gcact 25
 <210> 705
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 705
 acctgtgagg aggttggcac tgaaa 25
 <210> 706
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 706
 acctgtgagg aggttggcac tgaaa 25
 <210> 707
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 707
 ggacggctca gacgtggcct cctgc 25
 <210> 708
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 708
 tgcttggacg gctcagacgt ggcct 25

<210> 709
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 709
 cccctgacct cgtccccga acccc 25

<210> 710
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 710
 cccctgacct cgtccccga acccc 25

<210> 711
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 711
 tggggaactc ccgacaagat caaga 25

<210> 712
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 712
 ggaactcccg acaagatcaa gaaag 25

<210> 713
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 713
 cttegcattct tagctgggcc accag 25

<210> 714
 <211> 25
 <212> DNA

<213> Homo sapiens
 <400> 714
 cgcatcttag ctgggccacc agctg 25
 <210> 715
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 715
 acttggccgt cgcacctcca gaact 25
 <210> 716
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 716
 cagtcactt ggccgtcgca cctcc 25
 <210> 717
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 717
 tgccatctat tgatgggacc cagac 25

 <210> 718
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 718
 caagaaattt gccatctatt gatgg 25
 <210> 719
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 719
 gtgctcagtc gcgacatacc gacag 25
 <210> 720

<211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 720
 aggggtgctca gtcgcgacat accga 25
 <210> 721
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 721
 gactgcaagg acatgagcgt tgaag 25
 <210> 722
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 722
 gactgcaagg acatgagcgt tgaag 25
 <210> 723
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 723
 ggagggtggc tacaagtgcc agtgt 25
 <210> 724
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 724
 tgcgtgaacc tggagggtgg ctaca 25
 <210> 725
 <211> 25
 <212> DNA
 <213> Homo sapiens
 <400> 725

ctcacctgca gatcattctc tggga 25

<210> 726

<211> 25

<212> DNA

<213> Homo sapiens

<400> 726

tgcagatcat tctctgggac aggtc 25

<210> 727

<211> 25

<212> DNA

<213> Homo sapiens

<400> 727

ccacttgtgt gtctagatct cctca 25

<210> 728

<211> 25

<212> DNA

<213> Homo sapiens

<400> 728

ctcagtggcc gcctctactg ggttg 25

<210> 729

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

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

<400> 729

tttcaccatg ttggccaggt 20

<210> 730

<211> 20

<212> DNA

<213> Artificial Sequence

<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 730
tactgggttt caccatgttg 20
<210> 731
<211> 20
<212> DNA
<213> Artificial Sequence
<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 731
agaaaagtta gccgggcgtg 20
<210> 732
<211> 20
<212> DNA
<213> Artificial Sequence
<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"
<400> 732
acaggcgcgc gacaccacgc 20
<210> 733
<211> 159
<212> RNA
<213> Artificial Sequence
<220><221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"
<400> 733
uuucaccaug ugggccaggu gccauaaauuc cucuguaaaa cuuaaagaag guuuauagag 60
uuauuauggu aaggcaauau gccguggcgu uggggaucgc cuauguccgg uuuuaccgga 120
ucucccuaaa ggugacuaac uuugguuagu caccuuuuu 159

<210> 734

<211> 159

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 734

uacuggguuu caccauguug gccauaauc cucuguaaaa cuuaaagaag guuuauagag 60

uuauuauggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg uuuuaccgga 120

ucucccuaaa ggugacuaac uuugguuagu caccuuuuu 159

<210> 735

<211> 159

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 735

agaaaaguua gccgggcgug gccauaauc cucuguaaaa cuuaaagaag guuuauagag 60

uuauuauggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg uuuuaccgga 120

ucucccuaaa ggugacuaac uuugguuagu caccuuuuu 159

<210> 736

<211> 159

<212> RNA

<213> Artificial Sequence

<220><221> source

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

<400> 736

acaggcgcgc gacaccacgc gccauaauc cucuguaaaa cuuaaagaag guuuauagag 60

uuauuauggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg uuuuaccgga 120

ucucccuaaa ggugacuaac uuugguuagu caccuuuuu 159

<210> 737
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic

oligonucleotide"

<400> 737
 tcctggagcc tgtgataaaa 20

<210> 738
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic

oligonucleotide"

<400> 738
 gctaacagtt gcttttatca 20

<210> 739
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic

oligonucleotide"

<400> 739
 taacagttgc ttttatcaca 20

<210> 740
 <211> 159
 <212> RNA
 <213> Artificial Sequence
 <220><221> source
 <223> /note="Description of Artificial Sequence: Synthetic

polynucleotide"

<400> 740

uccuggagcc ugugauaaaa gccauaauc cucuguaaaa cuuaaagaag guuuauagag	60
uuuuuauuggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg uuuuaccgga	120
ucucccuaaa ggugacuaac uuugguuagu caccuuuuu	159

<210> 741

<211> 159

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 741

gcuaacaguu gcuuuuauca gccauaauc cucuguaaaa cuuaaagaag guuuauagag	60
uuuuuauuggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg uuuuaccgga	120
ucucccuaaa ggugacuaac uuugguuagu caccuuuuu	159

<210> 742

<211> 159

<212> RNA

<213> Artificial Sequence

<220><221> source

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

polynucleotide"

<400> 742

uaacaguugc uuuuauca gccauaauc cucuguaaaa cuuaaagaag guuuauagag	60
uuuuuauuggu aaggcaauau gccguggcgu ugagggaucgc cuauguccgg uuuuaccgga	120
ucucccuaaa ggugacuaac uuugguuagu caccuuuuu	159