



(24) 등록일자 2025년 05월 21일

- 심사관 : 이형곤

(54) 발명의 명칭 RNA-가이드된 핵산 변형 흡수 및 이의 사용 방법

(57) 요약

본 개시내용은 CasY 단백질, CasY 단백질을 암호화하는 핵산, 및 CasY 단백질 및/또는 이를 암호화하는 핵산을 포함하는 변형된 숙주 세포를 제공한다. 제공된 CasY 단백질은 다양한 적용분야에서 유용하다. 본 개시내용은 CasY 단백질에 결합하고 이에 대한 서열 특이성을 제공하는 CasY 가이드 RNA; CasY 가이드 RNA를 암호화하는 핵산; 및 CasY 가이드 RNA 및/또는 이를 암호화하는 핵산을 포함하는 변형된 숙주 세포를 제공한다. 제공되는 CasY 가이드 RNA는 다양한 적용분야에서 유용하다. 본 개시내용은 CRISPR RNA-가이드된 엔도뉴클레아제를 동정하는 방법을 제공한다.

대표도 - 도1a

[illegible]

(52) CPC특허분류

C12N 15/113 (2013.01)

C12N 15/63 (2013.01)

C07K 2319/09 (2013.01)

C12N 2310/20 (2025.05)

C12N 2830/002 (2013.01)

C12N 2830/008 (2013.01)

(72) 발명자

버스테인 데이비드

미국 캘리포니아 94720 버클리 맥콘 홀 307

해링턴 루카스 벤자민

미국 캘리포니아 94720 버클리 스탠리 홀 731 메인
스트리트#3220

명세서

청구범위

청구항 1

a) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열, 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는 CasY 폴리펩타이드, 또는 CasY 폴리펩타이드를 암호화하는 핵산 분자; 및

b) 서열번호 11 내지 15 중 어느 하나에 제시된 뉴클레오타이드 서열을 포함하는 crRNA를 포함하는 CasY 가이드 RNA, 또는 CasY 가이드 RNA를 암호화하는 하나 이상의 DNA 분자;를 포함하는 조성물로서,

CasY 폴리펩타이드는 절단 활성을 갖거나, 또는 CasY 폴리펩타이드는 촉매적으로 비활성이고, CasY 폴리펩타이드는 CasY 가이드 RNA와 복합체를 형성하고 표적 핵산에 결합하는, 조성물.

청구항 2

제1항에 있어서, DNA 공여자 주형을 더 포함하는 것을 특징으로 하는, 조성물.

청구항 3

제2항에 있어서, DNA 공여자 주형은 표적 핵산에 대해 상동성을 갖는 뉴클레오타이드 서열을 포함하는 것을 특징으로 하는, 조성물.

청구항 4

(a) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하고, 서열번호 11 내지 15 중 어느 하나에 제시된 뉴클레오타이드 서열을 포함하는 crRNA를 포함하는 CasY 가이드 RNA와 복합체를 형성할 때, 표적 핵산에 결합하는 CasY 폴리펩타이드; 및

(b) 이중성 폴리펩타이드

를 포함하는, CasY 융합 폴리펩타이드로서,

CasY 폴리펩타이드는 절단 활성을 갖거나, 또는 CasY 폴리펩타이드는 촉매적으로 비활성인, CasY 융합 폴리펩타이드.

청구항 5

제4항에 있어서, 이중성 폴리펩타이드는:

a) 표적 세포 또는 표적 세포 유형 상의 세포 표면 모이어티에 대한 결합을 제공하는 표적화 폴리펩타이드이거나;

b) 표적 DNA를 변형시키는 효소 활성을 나타내거나;

c) 표적 핵산과 회합된 표적 폴리펩타이드를 변형시키는 효소 활성을 나타내거나;

d) 엔도솜 탈출 폴리펩타이드이거나;

e) 엽록체 수송 펩타이드이거나;

f) 전사를 증가 또는 감소시키는 단백질이거나;

g) 단백질 결합 도메인이거나; 또는

h) 핵 국소화 신호인 것을 특징으로 하는, CasY 융합 폴리펩타이드.

청구항 6

제5항에 있어서, 이중성 폴리펩타이드는:

b) 표적 DNA를 변형시키는 효소 활성을 나타내며, 이중성 폴리펩타이드는 뉴클레아제 활성, 메틸트랜스퍼라제 활성, 데메틸라제 활성, DNA 수선 활성, DNA 손상 활성, 탈아미노화 활성, 디스무타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 재조합효소 활성, 중합효소 활성, 리가제 활성, 헬리카제 활성, 광분해효소 활성 및 글리코실라제 활성으로부터 선택된 하나 이상의 효소 활성을 나타내거나;

c) 표적 핵산과 회합된 표적 폴리펩타이드를 변형시키는 효소 활성을 나타내며, 이중성 폴리펩타이드는 히스톤 변형 활성을 나타내고, 이중성 폴리펩타이드는 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스파타제 활성, 유비퀴틴 리가제 활성, 데유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모일화(SUMOylating) 활성, 탈수모일화(deSUMOylating) 활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성, 탈미리스토일화 활성, 글리코실화 활성 및 탈글리코실화 활성으로부터 선택된 하나 이상의 효소 활성을 나타내거나;

d) GLFXALLXLLXSLWLLXA (서열번호 94), 및 GLFHALLHLLHSLWLLLLHA (서열번호 95)로부터 선택된 아미노산 서열 (여기서 각각의 X는 독립적으로 리신, 히스티딘, 및 아르기닌으로부터 선택된다)을 포함하는 엔도솜 탈출 폴리펩타이드이거나;

e) 엽록체 수송 펩타이드이며, 엽록체 수송 펩타이드는 서열번호 83 내지 93에 제시된 아미노산 서열로부터 선택되거나; 또는

f) 전사 리프레서 도메인 또는 전사 활성화 도메인으로부터 선택된, 전사를 증가 또는 감소시키는 단백질인 것을 특징으로 하는, CasY 융합 폴리펩타이드.

청구항 7

제4항 내지 제6항 중 어느 한 항의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산.

청구항 8

제7항에 있어서, CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 프로모터에 작동 가능하게 연결된 것을 특징으로 하는 핵산.

청구항 9

제1항 내지 제3항 중 어느 한 항에 있어서, CasY 폴리펩타이드는 검출 가능한 표지를 포함하는 것을 특징으로 하는 조성물.

청구항 10

제4항 내지 제6항 중 어느 한 항에 있어서, CasY 폴리펩타이드는 검출 가능한 표지를 포함하는 것을 특징으로 하는 CasY 융합 폴리펩타이드.

청구항 11

(a) 서열번호 11 내지 15 중 어느 하나에 제시된 뉴클레오타이드 서열을 포함하는 crRNA를 포함하는 CasY 가이드 RNA; 및

(b) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는 CasY 폴리펩타이드

를 암호화하는 하나 이상의 핵산 분자로서,

CasY 폴리펩타이드는 절단 활성을 갖거나, 또는 CasY 폴리펩타이드는 촉매적으로 비활성인, 핵산 분자.

청구항 12

a) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는 CasY 폴리펩타이드;

b) CasY 폴리펩타이드를 암호화하는 핵산 분자;

c) i) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는 CasY 폴리펩타이드; 및 ii) 이중성 폴리펩타이드를 포함하는 CasY 융합 폴리펩타이드;

d) CasY 융합 폴리펩타이드를 암호화하는 핵산 분자;

e) 서열번호 11 내지 15 중 어느 하나에 제시된 뉴클레오타이드 서열을 포함하는 crRNA를 포함하는 CasY 가이드 RNA; 및

f) CasY 가이드 RNA를 암호화하는 핵산 분자; 중

a) 내지 d) 중에서 적어도 하나, 그리고 e) 및 f) 중에서 적어도 하나 이상을 포함하는 진핵 세포로서,

CasY 폴리펩타이드는 절단 활성을 갖거나, 또는 CasY 폴리펩타이드는 촉매적으로 비활성이고,

세포는 인간 배아 또는 인간 생식 세포가 아닌, 진핵 세포.

청구항 13

제12항에 있어서, CasY 폴리펩타이드를 암호화하는 핵산 분자를 포함하며, 핵산 분자는 세포의 게놈 DNA 내로 통합되는 것을 특징으로 하는, 진핵 세포.

청구항 14

제13항에 있어서, 진핵 세포는 식물 세포, 포유류 세포, 곤충 세포, 거미류 세포, 진균 세포, 조류 세포, 파충류 세포, 양서류 세포, 무척추동물 세포, 마우스 세포, 래트 세포, 영장류 세포, 비-인간 영장류 세포, 또는 인간 세포인 것을 특징으로 하는, 진핵 세포.

청구항 15

a) i) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는 CasY 폴리펩타이드; 및

ii) 이중성 폴리펩타이드

를 포함하는 CasY 융합 폴리펩타이드; 또는

b) CasY 융합 폴리펩타이드를 암호화하는 핵산 분자를 포함하는 세포로서,

CasY 폴리펩타이드는 절단 활성을 갖거나, 또는 CasY 폴리펩타이드는 촉매적으로 비활성이고,

세포는 인간 배아 또는 인간 생식 세포가 아닌, 세포.

청구항 16

제15항에 있어서, CasY 융합 폴리펩타이드를 암호화하는 핵산 분자는 세포의 게놈 DNA 내로 통합되는 것을 특징으로 하는, 세포.

청구항 17

표적 핵산을 변형시키는 시험관내(*in vitro*) 또는 생체외(*ex vivo*)에서의 방법으로서, 표적 핵산을:

a) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는, 절단 활성을 갖는 CasY 폴리펩타이드; 및

b) 서열번호 11 내지 15 중 어느 하나에 제시된 뉴클레오타이드 서열을 포함하는 crRNA를 포함하는 CasY 가이드 RNA

와 접촉시키는 단계를 포함하며,

접촉 단계는 CasY 폴리펩타이드에 의한 표적 핵산의 변형을 초래하는, 방법.

청구항 18

제17항에 있어서,

- (i) CasY 폴리펩타이드, 또는 CasY 폴리펩타이드를 암호화하는 핵산 분자; 및
- (ii) CasY 가이드 RNA, 또는 CasY 가이드 RNA를 암호화하는 핵산 분자를 세포 내로 도입하는 단계를 포함하는 것을 특징으로 하는, 방법.

청구항 19

제18항에 있어서,

- (iii) DNA 공여자 주형을 세포 내로 도입하는 단계를 포함하며,
- 세포는 인간 유래의 세포를 제외하는 것을 특징으로 하는, 방법.

청구항 20

제19항에 있어서, 공여자 주형은 표적 핵산에 대해 상동성을 갖는 뉴클레오타이드 서열을 포함하는 것을 특징으로 하는, 방법.

청구항 21

제20항에 있어서, 변형은 공여자 주형의 표적 핵산 내로의 통합인 것을 특징으로 하는, 방법.

청구항 22

제18항 내지 제21항 중 어느 한 항에 있어서, 세포는 진핵 세포인 것을 특징으로 하는, 방법.

청구항 23

제22항에 있어서, 진핵 세포는 진균 세포, 포유류 세포, 파충류 세포, 곤충 세포, 조류 세포, 어류 세포, 기생충 세포, 절지동물 세포, 무척추동물의 세포, 척추동물의 세포, 설치류 세포, 마우스 세포, 래트 세포, 영장류 세포, 비-인간 영장류 세포, 및 인간 세포로부터 선택된 것을 특징으로 하는, 방법.

청구항 24

표적 DNA로부터 전사를 조절하거나, 표적 핵산을 변형시키거나, 또는 표적 핵산과 회합된 단백질을 변형시키는 방법으로서, 표적 핵산을:

- a) 이중성 폴리펩타이드에 융합된 CasY 폴리펩타이드를 포함하는 CasY 융합 폴리펩타이드로서, CasY 폴리펩타이드는 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는, CasY 융합 폴리펩타이드; 및
- b) 서열번호 11 내지 15 중 어느 하나에 제시된 뉴클레오타이드 서열을 포함하는 crRNA를 포함하는 CasY 가이드 RNA

와 접촉시키는 단계를 포함하며,

CasY 폴리펩타이드는 절단 활성을 갖거나, 또는 CasY 폴리펩타이드는 촉매적으로 비활성인. 방법.

청구항 25

제24항에 있어서, 접촉 단계는

시험관내(*in vitro*) 또는 생체외(*ex vivo*)에서

- (i) CasY 융합 폴리펩타이드, 또는 CasY 융합 폴리펩타이드를 암호화하는 핵산 분자, 및
 - (ii) CasY 가이드 RNA, 또는 CasY 가이드 RNA를 암호화하는 핵산 분자
- 를 세포 내로 도입하는 단계를 포함하고,
- 세포는 인간 유래의 세포를 제외하는 것을 특징으로 하는, 방법.

청구항 26

a) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는 CasY 폴리펩타이드;

b) i) 서열번호 1 내지 7 중 어느 하나에 제시된 아미노산 서열 및 단일 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인을 포함하는 CasY 폴리펩타이드; 및

ii) 이중성 폴리펩타이드를 포함하고,

서열번호 11 내지 15 중 어느 하나에 제시된 뉴클레오타이드 서열을 포함하는 crRNA를 포함하는 CasY 가이드 RNA와 복합체를 형성할 때, 표적 핵산에 결합하는 CasY 융합 폴리펩타이드; 및

c) 서열번호 11 내지 15 중 어느 하나에 제시된 뉴클레오타이드 서열을 포함하는 crRNA를 포함하는 CasY 가이드 RNA; 중

a)와 b) 중에서 적어도 하나, 및 c)를 암호화하는 뉴클레오타이드 서열을 포함하는 이식유전자가 게놈에 포함된 유전자이식 다세포 비인간 유기체로서,

CasY 폴리펩타이드는 절단 활성을 갖거나, 또는 CasY 폴리펩타이드는 촉매적으로 비활성인, 유전자이식 다세포 비인간 유기체.

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

청구항 76

삭제

청구항 77

삭제

청구항 78

삭제

청구항 79

삭제

청구항 80

삭제

청구항 81

삭제

청구항 82

삭제

청구항 83

삭제

청구항 84

삭제

청구항 85

삭제

청구항 86

삭제

청구항 87

삭제

청구항 88

삭제

청구항 89

삭제

청구항 90

삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

청구항 96

삭제

청구항 97

삭제

청구항 98

삭제

청구항 99

삭제

청구항 100

삭제

청구항 101

삭제

청구항 102

삭제

청구항 103

삭제

청구항 104

삭제

청구항 105

삭제

청구항 106

삭제

청구항 107

삭제

청구항 108

삭제

청구항 109

삭제

청구항 110

삭제

청구항 111

삭제

청구항 112

삭제

청구항 113

삭제

청구항 114

삭제

청구항 115

삭제

청구항 116

삭제

청구항 117

삭제

청구항 118

삭제

청구항 119

삭제

청구항 120

삭제

청구항 121

삭제

청구항 122

삭제

청구항 123

삭제

발명의 설명

기술 분야

[0001] 상호 참조

[0002] 본 출원은 2016년 9월 30일자로 출원된 미국 가출원 특허 제62/402,849호의 유익을 주장하며, 이 기초출원은 그 의 전문이 본 명세서에 참고로 포함된다.

[0003] 텍스트 파일로서 제공된 서열목록의 참고에 의한 편입

[0004] 서열목록은 2017년 9월 28일자에 생성되고, 용량이 244 KB인 텍스트 파일 "BERK-343WO_SeqList_ST25.txt"로서 본 명세서에 제공된다. 텍스트 파일의 내용은 그 전체가 본 명세서에 참고로 편입된다.

배경 기술

[0005] DNA 서열분석 시대 전의 과학에 알려지지 않은 경로의 예인 CRISPR-Cas 시스템은 이제 파지 및 바이러스에 대해 획득된 면역을 박테리아 및 고세균에 부여하는 것으로 이해된다. 과거 십년에 걸쳐 광대한 연구는 이 생화학 시스템을 발견하지 못하였다. CRISPR-Cas 시스템은 외래 DNA 또는 RNA의 획득, 표적화 및 절단에 연루된 Cas 단백질, 및 Cas 단백질을 그들의 표적에 가이드하는 짧은 스페이스 서열에 축적하는 직접 반복부를 포함하는 CRISPR 어레이로 이루어진다. RNA에 결합된 단일 Cas 단백질이 표적화된 서열에 대한 결합 및 절단을 초래하는 클래스 2 CRISPR-Cas는 최신식의 형태이다. 이들 최소 시스템의 프로그램 가능한 특성은 그들의 용도를 게놈 조작 분야의 혁명을 일으키는 다재다능한 기술로서 가능하게 하였다.

[0006] 현재의 CRISPR-Cas 기술은 아직 손대지 않은 단리되지 않은 방대한 유기체를 떠나서 배양 박테리아로부터의 시스템에 기반한다. 지금까지, 소수의 클래스 2 CRISPR/Cas 시스템만이 발견되었다. 추가적인 클래스 2 CRISPR/Cas 시스템(예를 들어, Cas 단백질 + 가이드 RNA 조합)에 대한 당업계의 필요가 있다.

발명의 내용

[0007] 본 발명의 개시내용은 본 명세서에서 "CasY" 폴리펩타이드로서 지칭되는(또한 "CasY 단백질"로서 지칭됨) RNA-가이드 엔도뉴클레아제 폴리펩타이드; CasY 폴리펩타이드를 암호화하는 핵산; 및 CasY 폴리펩타이드를 포함하는 변형된 숙주 세포 및/또는 이를 암호화하는 핵산을 제공한다. 제공되는 CasY 폴리펩타이드는 다양한 적용분야에서 유용하다.

[0008] 본 개시내용은 CasY 단백질에 결합하고 이에 대한 서열 특이성을 제공하는 가이드 RNA(본 명세서에서 "CasY 가이드 RNA"로서 지칭됨); CasY 가이드 RNA를 암호화하는 핵산; 및 CasY 가이드 RNA 및/또는 이를 암호화하는 핵산을 포함하는 변형된 숙주 세포를 제공한다. 제공되는 CasY 가이드 RNA는 다양한 적용분야에서 유용하다.

[0009] 본 개시내용은 CRISPR RNA-가이드된 엔도뉴클레아제를 동정하는 방법을 제공한다.

도면의 간단한 설명

[0010] 도 1은 천연 유래 CasY 단백질 서열의 예를 도시한 도면.

도 2는 천연 유래 CasY 단백질 서열의 정렬을 도시한 도면.

도 3(패널 a 내지 b)은 CasY에 대한 개략적 도메인 제시를 도시한 도면. 또한 CasY의 상동체를 동정하기 위한 다양한 검색 시도로부터의 결과를 나타낸다. 또한 동정된 CasY-함유 CRISPR 좌위의 부분이 도시된다.

도 4는 CasY 및 C2c3 좌위의 개략적 다이어그램을 도시한 도면. 간접 단백질을 녹색으로 나타내고, 획득 단백질을 적색으로 나타낸다. RNA 구조를 이용하여 폴딩된 반복부는 5' 단부에서 강한 헤어핀을 나타내는 우측에 나타나는데, 이는 CRISPR 어레이의 자기 가공을 시사한다.

도 5(패널 a 내지 d)는 CasY에 대한 PAM 서열을 결정하기 위해 수행된 실험(CasY에 의한 PAM 의존적 플라스미드 간접)을 도시한 도면.

도 6(패널 a 내지 b)은 천연 유래 CasY 가이드 RNA 및 표적 DNA에 혼성화하는 예시적 Cas Y 가이드 RNA의 '반복부' 서열을 제시한 도면. (상부 내지 하부, 서열번호 11 내지 15, 및 20)

도 7(패널 a 내지 b)는 비배양 유기체로부터의 신규한 동정된 CRISPR-Cas 시스템을 제시한 도면. a, Hug et al.³²의 데이터에 기반하여 모든 박테리아 및 고세균에서 단리된 대표자가 있는 그리고 대표자가 없는 주된 계통의 비. 결과는 이들 도메인에서 대규모의 아직 거의 연구되지 않은 생물학을 강조한다. 고세균 Cas9 및 신규한 CRISPR-CasY는 단리된 대표자가 없는 계통에서 배타적으로 발견되었다. b, 새로 발견된 CRISPR-Cas 시스템의 국소 조직화.

도 8(패널 a 내지 b)는 ARMAN-1 Cas9 PAM 서열의 ARMAN-1 CRISPR 어레이 다양성 및 동정을 제시한 도면. a, 15개의 상이한 AMD 샘플로부터 재구성한 CRISPR 어레이. 백색 박스는 반복부를 나타내고, 색이 있는 다이아몬드는 스페이서를 나타낸다(동일한 스페이서는 동일하게 색칠하고; 독특한 스페이서는 검정색이다). 어레이의 보존된 영역을 강조한다(우측 상에서). 최근에 획득된 스페이서의 다양성은(좌측 상에서) 시스템이 활성이라는 것을 나타낸다. 또한 판독 데이터로부터의 CRISPR 단편을 포함하는 분석을 도 14에 제시한다. b, AMD 메가게놈 데이터로부터 재구성된 단일 추정 바이러스 콘텍은 ARMAN-1 CRISPR 어레이로부터의 56개 프로토스페이서(적색 수직 막대)를 함유한다. c, 서열분석은 비표적 가닥 상에서 프로토스페이서 하류의 보존된 'NGG' PAM 모티프를 나타내었다.

도 9(패널 a 내지 d)는 이콜라이에서 프로그램 가능한 DNA 간섭을 매개한다는 것을 나타내는 데이터를 제시한 도면. a, CasX 플라스미드 간섭 분석의 다이어그램. 최소 CasX 좌위를 발현시키는 이콜라이는 CRISPR 어레이(표적)에서 서열을 매칭시키는 스페이서를 함유하는 플라스미드 또는 비-매칭 스페이서(비표적)를 함유하는 플라스미드로 형질전환된다. 형질전환된 후에, 배양물을 플레이트팅하고, 콜로니 형성 단위(colony forming unit: cfu)를 정량화하였다. b, 부유균문 CasX 좌위 표적화 스페이서 1(sX.1)을 발현시키고 구체화된 표적(sX1, CasX 스페이서 1; sX2, CasX 스페이서 2; NT, 비-표적)으로 형질전환된 이콜라이의 연속 희석. c, 델타프로테오박테리아 CasX에 의한 플라스미드 간섭. 실험을 3회 중복하여 수행하고 평균 \pm s.d.를 나타낸다. d, 이콜라이에서 발현시킨 부유균문 CasX 좌위에 대한 PAM 고갈 분석. 대조군 라이브러리에 비해 30배 초과로 고갈된 PAM 서열을 사용하여 웹로그(WebLogo)를 생성하였다.

도 10(패널 a 내지 c)은 CasX를 나타내는 데이터가 이중-가이드 CRISPR 복합체라는 것을 제시한 도면. a, 아래쪽 다이어그램의 CasX CRISPR 좌위에 대한 환경 RNA 서열(메타전사체 데이터)의 맵핑(적색 화살표, 추정적 tracrRNA; 백색 박스, 반복부 서열; 녹색 다이아몬드, 스페이서 서열). 삽도는 제1 반복부 및 스페이서의 상세한 관점을 나타낸다. b, CasX 이중-가닥 DNA 간섭의 다이어그램. RNA 가공 부위는 검정색 화살표로 나타낸다. c, CasX 좌위의 추정적 tracrRNA Nick아웃(T, 표적; NT, 비-표적)에 의한 플라스미드 간섭 분석 결과. 실험을 3회 중복하여 수행하고 평균 \pm s.d.를 나타낸다.

도 11(패널 a 내지 c)은 이콜라이에서 CasY 좌위의 발현이 DNA 간섭에 충분하다는 것을 나타내는 데이터를 제시한 도면. **a**, CasY 좌위 및 이웃하는 단백질의 다이어그램. **b**, 대조군 라이브러리에 비해 CasY에 의해 3배 초과로 고갈된 5' PAM 서열의 웹로그. **c**, 이콜라이 발현 CasY.1에 의한 그리고 표시된 PAM을 함유하는 표적으로 형질전환된 플라스미드 간섭. 실험을 3회 중복하여 수행하고 평균 \pm s.d.를 나타낸다.

도 12(패널 a 내지 b)는 공지된 시스템과 관련하여 새로 동정된 CRISPR-Cas를 제시한 도면. **a**, 보편적 Cas1 단백질의 단순화된 계통수. 공지된 시스템의 CRISPR 유형은 켜기와 분지로 나타내고; 새로 기재한 시스템은 볼드체이다. 상술한 Cas1 계통발생을 보조 데이터 2에 제시한다. **b**, II-B형과 II-C형 좌위 사이의 재조합 결과로서 고세균 II형 시스템에 생기는 제안된 진화 시나리오.

도 13은 ARMAN-4로부터의 고세균 Cas9가 축퇴 CRISPR 어레이에 의해 수많은 콘텍 상에서 발견된다는 것을 나타내는 도면. ARMAN-4로부터의 Cas9는 16개의 상이한 콘텍 상에서 진한 적색으로 강조한다. 추정적 도메인 또는 기능을 갖는 단백질을 표시한 반면, 가설 단백질을 표시하지 않는다. 15개의 콘텍은 2개의 축퇴 직접 반복부(1개는 bp 미스매치) 및 단일, 보존된 스페이서를 함유한다. 남아있는 콘텍은 단지 하나의 직접 반복부를 함유한다. ARMAN-1과 달리, 추가적인 Cas 단백질은 ARMAN-4에서 Cas9에 인접하여 발견되지 않는다.

도 14는 ARMAN-1 CRISPR 어레이의 완전한 재구성을 제시한 도면. 기준 조립 서열뿐만 아니라 짧은 DNA 판독으로부터 재구성된 어레이 세그먼트를 포함하는 CRISPR 어레이의 재구성. 녹색 화살표는 반복부를 나타내고, 색이 있는 화살표는 CRISPR 스페이서를 나타낸다(동일한 스페이서는 동일하게 색칠한 반면, 독특한 스페이서는 검정색으로 색칠한다). CRISPR 시스템에서, 스페이서는 전형적으로 일방향성으로 첨가되고, 따라서 좌측 상에서 매우 다양한 스페이서는 최근의 획득에 기인한다.

도 15(패널 a 내지 b)는 ARMAN-1 스페이서가 고세균 군집 구성원의 게놈에 맵핑된다는 것을 도시한 도면. **a**, 동일한 환경으로부터의 나노고세균인 ARMAN-2의 게놈에 대한 ARMAN-1 맵으로부터의 프로토스페이서(적색 화살표). 6개의 프로토스페이서는 2개의 긴-말단 반복부(long-terminal repeat: LTR)에 의해 축적된 게놈의 부분에 독특하게 맵핑되고, 2개의 추가적인 프로토스페이서는 LTR(청색 및 녹색) 내에서 완전하게 매칭된다. 이 영역은 트랜스포존일 가능성이 있는데, ARMAN-1의 CRISPR-Cas 시스템은 이 요소의 이동을 억제하는 역할을 한다는 것을 시사한다. **b**, 프로토스페이서는 또한 ARMAN 유기체와 동일한 샘플에서 발견된 리치몬드 마인 에코시스템(Richmond Mine ecosystem)의 다른 구성원인 써모플라즈마탈레스(*Thermoplasma*) 고세균(I-플라즈마)에 대해 맵핑한다. 프로토스페이서는 짧은, 가설 단백질을 암호화하는 게놈 영역 내에서 클러스터링하는데, 이것이 또한 이동 요소를 나타낼 수 있다는 것을 시사한다.

도 16(패널 a 내지 e)는 ARMAN-1 crRNA 및 tracrRNA의 예측된 2차 구조를 제시한 도면. **a**, CRISPR 반복부 및 tracrRNA 항-반복부를 검정색으로 도시한 반면, 스페이서-유래 서열을 일련의 녹색 N으로 나타낸다. 좌위로부터 분명한 종결 신호를 예측할 수 없고, 따라서 3개의 상이한 tracrRNA 길이는 그들의 2차 구조(각각 적색, 청색 및 핑크색인 69, 104 및 179)에 기반하여 시험하였다. **b**, **a**, **c**에서 이중-가이드에 대응하는 조작된 단일-가이드 RNA, tracrRNA의 3' 말단 상에서 2개의 상이한 헤어핀(75 및 122)을 갖는 ARMAN-4 Cas9에 대한 이중-가이드. **d**, **c**에서의 이중-가이드에 대응하는 조작된 단일-가이드 RNA. **e**, 이콜라이 생체내 표적화 분석에서 시험한 조건.

도 17(패널 a 내지 b)는 시험관내 생화학 연구를 위한 정제 계획을 제시한 도면. **a**, ARMAN-1(AR1) 및 ARMAN-4(AR4) Cas9를 발현시키고 나서, 보충 물질에서 약술할 바와 같이 다양한 조건 하에 정제하였다. 청색 박스로 윤곽을 나타낸 단백질을 시험관내 절단 활성에 대해 시험하였다. **b**, AR1-Cas9 및 AR4-Cas9 정제 분획을 10% SDS-PAGE 겔 상에서 분리하였다.

도 18은 공지된 단백질에 비교하여 새로 동정된 CRISPR-Cas 시스템을 제시한 도면. 다음의 검색에 기반하여 알려진 단백질에 대한 CasX 및 CasY의 유사성: (1) NCBI의 혼하지 않은(non-redundant: NR) 단백질 데이터베이스에 대한 Blast 검색, (2) 모든 공지된 단백질의 HMM 데이터베이스에 대한 히든 마르코프 모델(hidden markov model: HMM) 검색 및 (3) HHpred³⁰을 이용하는 원위 상동성 검색.

도 19(패널 a 내지 d)는 CasX에 의한 프로그래밍된 DNA 간섭과 관련된 데이터를 제시한 도면. **a**, 도 9, 패널 c로부터 계속된 CasX2(부유균문) 및 CasX1(델타프로테오박테리아)에 대한 플라스미드 간섭 분석(sX1, CasX 스페이서 1; sX2, CasX 스페이서 2; NT, 비-표적). 실험을 3회 중복하여 수행하고 평균 \pm s.d.를 나타낸다. **b**, 도 9, 패널 b로부터 계속된, CasX 좌위를 발현시키고 구체화된 표적으로 형질감염된 이콜라이의 연속 희석. **c**, 델타프로테오박테리아 CasX에 대한 PAM 고갈 분석 및 **d**, 이콜라이에서 발현된 부유균문 CasX. 대조군 라이브러리

에 비교하여 나타낸 PAM 고갈 값 역치(PDVT) 초과로 고갈된 PAM 서열을 사용하여 웹로그를 생성하였다.

도 20은 Cas9 상동체의 진화계통수를 제시한 도면. 유형에 기반하여 색칠한 앞서 기재한 시스템을 나타내는 Cas9 단백질의 최대-가능성 계통수: II-A(청색), II-B(녹색) 및 II-C(보라색). 비배양 박테리아로부터의 2개의 새로 기재된 박테리아 Cas9와 함께 II-C형 CRISPR-Cas 시스템에 의한 클러스터인 고세균 Cas9.

도 21은 ARMAN-1 및 ARMAN-4로부터의 Cas9에 대해 분석한 절단 조건 표를 제시한 도면.

발명을 실시하기 위한 구체적인 내용

정의

본 명세서에서 사용되는 바와 같은 "이중성"은 각각 천연 핵산 또는 단백질에서 발견되지 않는 뉴클레오타이드 또는 폴리뉴클레오타이드 서열을 의미한다. 예를 들어, CasY 폴리펩타이드에 대해, 이중성 폴리펩타이드는 CasY 폴리펩타이드 이외의 단백질로부터의 아미노산 서열을 포함한다. 일부 경우에, 하나의 종으로부터의 CasY 단백질 부분은 상이한 종으로부터의 CasY 단백질 부분에 융합된다. 따라서 각각의 종으로부터의 CasY 서열은 서로에 대해 이중성인 것으로 고려될 수 있었다. 다른 예로서, CasY 단백질(예를 들어, dCasY 단백질)은 비-CasY 단백질(예를 들어, 히스톤 데아세틸라제)로부터의 활성 도메인에 융합될 수 있고, 활성 도메인의 서열은 이중성 폴리펩타이드로 고려될 수 있었다(이는 CasY 단백질에 대해 이중성이다).

본 명세서에서 상호 호환적으로 사용되는 용어 "폴리뉴클레오타이드" 및 "핵산"은 리보뉴클레오타이드 또는 데옥시뉴클레오타이드 중 하나의 임의의 길이 뉴클레오타이드의 중합체 형태를 지칭한다. 따라서, 이 용어는 단일-, 이중-, 또는 다중-가닥 DNA 또는 RNA, 게놈 DNA, cDNA, DNA-RNA 혼성체, 또는 퓨린 및 피리미딘 염기 또는 다른 천연, 화학적 또는 생화학적으로 변형된, 비천연 또는 유도체화된 뉴클레오타이드 염기를 포함하는 중합체를 포함하지만, 이들로 제한되지 않는다. 용어 "폴리뉴클레오타이드" 및 "핵산"은 기재 중인 실시형태에 적용 가능한, 단일-가닥(예컨대 센스 또는 안티센스) 및 이중-가닥 폴리뉴클레오타이드를 포함하는 것으로 이해되어야 한다.

용어 "폴리펩타이드", "펩타이드" 및 "단백질"은 본 명세서에서 상호 호환적으로 사용되고, 유전적으로 암호화된 그리고 비유전적으로 암호화된 아미노산, 화학적으로 또는 생화학적으로 변형 또는 유도체화된 아미노산, 및 변형된 펩타이드 골격을 갖는 폴리펩타이드를 포함할 수 있는 임의의 길이의 아미노산의 중합체 형태를 지칭한다. 상기 용어는 N-말단의 메티오닌 잔기가 있거나 또는 없는, 이중성 아미노산 서열과의 융합 단백질, 이중성 및 상동성 리더 서열과의 융합; 면역학적으로 태그된 단백질 등을 포함하지만, 이들로 제한되지 않는 융합 단백질을 포함한다.

핵산, 단백질, 세포 또는 유기체에 적용되는 본 명세서에서 사용되는 용어 "천연 유래"는 천연에서 발견되는 핵산, 세포, 단백질 또는 유기체를 지칭한다.

본 명세서에서 사용되는 용어 "단리된"은 폴리뉴클레오타이드, 폴리펩타이드 또는 세포가 자연적으로 생기는 환경과 상이한 환경에 있는 폴리뉴클레오타이드, 폴리펩타이드 또는 세포를 설명하는 것을 의미한다. 단리된 유전자 변형된 숙주 세포는 유전자 변형된 숙주 세포의 혼합 집단으로 제공될 수 있다.

본 명세서에 사용되는 바와 같은 용어 "외인성 핵산"은 천연에서 주어진 박테리아, 유기체 또는 세포에서 정상적으로 또는 자연적으로 발견되지 않고/않거나 생성되지 않는 핵산을 지칭한다. 본 명세서에 사용되는 바와 같은 용어 "내인성 핵산"은 천연에서 주어진 박테리아, 유기체 또는 세포에서 정상적으로 또는 자연적으로 발견되고/되거나 생성된 핵산을 지칭한다. "내인성 핵산"은 또한 "천연 핵산" 또는 주어진 박테리아, 유기체 또는 세포에 대해 "천연"인 핵산으로서 지칭된다.

본 명세서에서 사용되는 "재조합체"는 특정 핵산(DNA 또는 RNA)이 천연 시스템에서 발견되는 내인성 핵산과 구별 가능한 구조적 암호 또는 비-암호 서열을 갖는 작제물을 초래하는 클로닝, 제한 및/또는 절찰 단계의 다양한 조합의 생성물이라는 것을 의미한다. 일반적으로, 구조적 암호 서열을 암호화하는 DNA 서열은 세포에 또는 무세포 전사 및 번역 시스템에 함유된 재조합 전사 단위로부터 발현될 수 있는 합성 핵산을 제공하기 위해 cDNA 단편 및 짧은 올리고뉴클레오타이드 링커로부터, 또는 일련의 합성 올리고뉴클레오타이드로부터 조립될 수 있다. 이러한 서열은 전형적으로 진행 유전자에 존재하는 내부 비-번역 서열 또는 인트론에 의해 중단되지 않는 오픈 리딩 프레임 형태로 제공될 수 있다. 적절한 서열을 포함하는 게놈 DNA는 또한 재조합 유전자 또는 전사 단위의 형태로 사용될 수 있다. 비번역 DNA의 서열은 오픈 리딩 프레임으로부터의 5' 또는 3'에 존재할 수 있으며, 여기서 이러한 서열은 암호 영역의 조작 또는 발현을 방해하지 않고, 정말로 다양한 메커니즘에 의해 목적으로 하

는 생성물의 생성물을 조작하도록 작용할 수 있다("DNA 조절 서열" 참조, 이하).

- [0019] 따라서, 예를 들어, 용어 "재조합" 폴리뉴클레오타이드 또는 "재조합" 핵산은 천연 유래가 아닌 것, 예를 들어, 인간 개입을 통해 서열의 2개의 다르게 분리된 세그먼트의 인공 조합에 의해 만들어진 것을 지칭한다. 이 인공 조합은 종종 화학적 합성 수단에 의해, 또는 핵산의 단리된 세그먼트의 인공 조작에 의해, 예를 들어, 유전자 조작 기법에 의해 수행된다. 이는 보통 코돈을 동일한 또는 보존적 아미노산을 암호화하는 풍부한 코돈으로 대체하는 한편, 전형적으로 서열 인식 부위를 도입하거나 또는 제거하기 위해 행한다. 대안적으로, 목적으로 하는 기능의 조합을 생성하기 위해 목적으로 하는 기능의 핵산 세그먼트를 함께 결합하도록 수행된다. 이 인공 조합은 종종 화학적 합성 수단에 의해, 또는 핵산의 단리된 세그먼트의 인공 조작에 의해, 예를 들어, 유전자 조작 기법에 의해 수행된다.
- [0020] 유사하게, 용어 "재조합" 폴리펩타이드는 천연 유래가 아니며, 예를 들어, 인간 개입을 통해 아미노 서열의 2개의 다르게 분리된 세그먼트의 인공 조합물에 의해 제조된 폴리펩타이드를 지칭한다. 따라서, 예를 들어, 이중성 아미노산 서열을 포함하는 폴리펩타이드는 재조합이다.
- [0021] "작제물" 또는 "백터"는 특정 뉴클레오타이드 서열(들)의 발현 및/또는 증식의 목적을 위해 생성된, 또는 다른 재조합 뉴클레오타이드 서열의 작제물에서 사용될 재조합 핵산, 일반적으로 재조합 DNA를 의미한다.
- [0022] 본 명세서에서 상호 호환적으로 사용되는 용어 "DNA 조절 서열", "제어 요소" 및 "조절 요소"는 암호 서열의 발현 및/또는 숙주 세포에서 암호화된 폴리펩타이드의 생성을 제공하고/하거나 조절하는 전사 및 번역 제어 서열, 예컨대 프로모터, 인핸서, 폴리아데닐화 신호, 종결자, 단백질 분해 신호 등을 지칭한다.
- [0023] 용어 "형질전환"은 본 명세서에서 "유전자 변형"과 상호 호환적으로 사용되고, 세포 내로 새로운 핵산(예를 들어, 세포에 외인성인 DNA)의 도입 후 세포에 유도된 영구한 또는 일시적 유전자 변화를 지칭한다. 유전자 변형("변형")은 숙주 세포의 게놈 내로 새로운 핵산의 혼입에 의해, 또는 에피솜 요소로서 새로운 핵산의 일시적 또는 안정한 유지에 의해 달성될 수 있다. 세포가 진핵 세포인 경우에, 영구한 유전자 변화는 일반적으로 세포 게놈 내로 새로운 DNA의 도입에 의해 달성된다. 원핵세포에서, 염색체 내로 또는 염색체외 요소, 예컨대 플라스미드 및 발현 백터를 통해 영구적 변화가 도입될 수 있는데, 이는 재조합 숙주 세포에서 그들의 유지에 도움을 주는 하나 이상의 선택 가능한 마커를 함유할 수 있다. 유전자 변형의 적합한 방법은 바이러스 감염, 형질감염, 접합, 원형질체 융합, 전기천공법, 입자총 기법, 인산칼슘 침전, 직접 마이크로주사 등을 포함한다. 방법의 선택은 일반적으로 형질감염 중인 세포의 유형 및 형질감염이 일어나는 환경(즉, 시험관내, 생체외 또는 생체내)에 의존한다. 이들 방법의 일반적 논의는 문헌[Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995]에서 찾을 수 있다.
- [0024] "작동 가능하게 연결된"은 이렇게 기재된 성분이 그들의 의도된 방식으로 작용하도록 허용하는 관계인 병치를 지칭한다. 예를 들어, 프로모터가 그의 전사 또는 발현에 영향을 미친다면 프로모터는 암호 서열에 작동 가능하게 연결된다. 본 명세서에서 사용되는 용어 "이중성 프로모터" 및 "이중성 제어 영역"은 천연에서 특정 핵산과 정상적으로 결합되지 않은 프로모터 및 다른 제어 영역을 지칭한다. 예를 들어, "암호 영역에 대해 이중성인 전사 제어 영역"은 천연에서 암호 영역과 정상적으로 결합되지 않은 전사 제어 영역이다.
- [0025] 본 명세서에서 사용되는 "숙주 세포"는 진핵 또는 원핵 세포가 핵산에 대한 수용자(예를 들어, 발현 백터)로서 사용될 수 있거나 또는 사용되고, 핵산에 의해 유전자 변형된 본래의 세포의 자손을 포함하는, 생체내 또는 시험관내 진핵 세포, 원핵 세포 또는 단세포 독립체로서 배양시킨 다세포 유기체로부터의 세포(예를 들어, 세포주)를 나타낸다. 단일 세포의 자손은 단일 세포의 자손은 천연, 우연한 또는 의도적인 돌연변이에 기인하여, 본래의 모체와 형태가 또는 게놈 또는 총 DNA 상보체와 완전히 동일할 필요는 없다는 것이 이해된다. "재조합 숙주 세포"(또한 "유전자 변형된 숙주 세포"로서 지칭됨)는 이중성 핵산, 예를 들어, 발현 백터가 도입된 숙주 세포이다. 예를 들어, 대상 원핵 숙주 세포는 이중성 핵산, 예를 들어, 원핵 숙주 세포에서 정상적으로 발견되지 않는 원핵 숙주 세포 또는 재조합 핵산에 대해 외래인(그리고 천연에서 정상적으로 발견되지 않는) 외인성 핵산의 적합한 원핵 숙주 세포 내로의 도입에 의한 유전자 변형된 원핵 숙주 세포(예를 들어, 박테리아)이고; 그리고 대상 진핵 숙주 세포는 이중성 핵산, 예를 들어, 진핵 숙주 세포에 정상적으로 발견되지 않는 진핵 숙주 세포, 또는 재조합 핵산에 대해 외래인 외인성 핵산의 적합한 진핵 숙주 세포 내로 도입에 의한 유전자 변형된 진핵 숙주 세포이다.
- [0026] 용어 "보존적 아미노산 치환"은 유사한 측쇄를 갖는 아미노산 잔기의 단백질에서의 상호 호환성을 지칭한다. 예를 들어, 지방족 측쇄를 갖는 아미노산의 그룹은 글리신, 알라닌, 발린, 류신, 및 아이소류신으로 이루어지며;

지방족-하이드록실 측쇄를 갖는 아미노산의 그룹은 세린 및 트레오닌으로 이루어지고; 아마이드-함유 측쇄를 갖는 아미노산의 그룹은 아스파라긴 및 글루타민으로 이루어지며; 방향족 측쇄를 갖는 아미노산 그룹은 페닐알라닌, 타이로신 및 트립토판으로 이루어지고; 염기성 측쇄를 갖는 아미노산의 그룹은 라이신, 알기닌 및 히스티딘으로 이루어지며; 그리고 황-함유 측쇄를 갖는 아미노기의 그룹은 시스테인 및 메티오닌으로 이루어진다. 예시적인 보존적 아미노산 치환 그룹은 발린-류신-아이소류신, 페닐알라닌-타이로신, 라이신-알기닌, 알라닌-발린, 및 아스파라긴-글루타민이다.

[0027] 폴리뉴클레오타이드 또는 폴리펩타이드는 다른 폴리뉴클레오타이드 또는 폴리펩타이드에 대한 특정 "서열 동일성" 백분율을 갖는데, 이는 정렬될 때, 염기 또는 아미노산의 백분율이 동일하고, 두 서열을 비교할 때 동일한 상대적 위치에 있다는 것을 의미한다. 서열 유사성은 다수의 상이한 방식으로 결정될 수 있다. 서열 동일성을 결정하기 위해, 서열은 ncbi.nlm.nih.gov/BLAST에서 월드 와이드 웹에 대해 이용 가능한 BLAST를 포함하는 방법 및 컴퓨터 프로그램을 이용하여 정렬될 수 있다. 예를 들어, 문헌[Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-10] 참조. 다른 정렬 알고리즘은 옥스퍼드 몰레큘러 그룹, 인코포레이티드(Oxford Molecular Group, Inc.)의 전액 출자된 자회사인 미국 위스콘신주 메디슨에 소재한 제네틱스 컴퓨팅 그룹(Genetics Computing Group: GCG) 패키지에서 이용 가능한 FASTA이다. 정렬을 위한 다른 기법은 문헌[Methods in Enzymology, vol. 266: Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA]에 기재되어 있다. 특정 관심 대상 중에서 서열 내 겹을 허용하는 정렬 프로그램이다. 스미스-워터만(Smith-Waterman)은 서열 정렬에서 겹을 허용하는 알고리즘 중 하나의 유형이다. 문헌[Meth. Mol. Biol. 70: 173-187 (1997)] 참조. 또한, 니들만(Needleman) 및 분쉬(Wunsch) 정렬 방법을 이용하는 GAP 프로그램은 서열을 정렬하기 위해 이용될 수 있다. 문헌[J. Mol. Biol. 48: 443-453 (1970)] 참조.

[0028] 본 명세서에서 사용되는 용어 "치료", "치료하는" 등은 목적으로 하는 약학적 및/또는 생리적 효과를 얻는 것을 지칭한다. 효과는 질환 또는 이의 증상을 완전히 또는 부분적으로 방지하는 것에 관해 예방적일 수 있고/있거나 질환 및/또는 질환에 기인하는 유해한 효과에 대한 부분적 또는 완전한 치유에 관해 치유적일 수 있다. 본 명세서에서 사용되는 "치료"는 포유류에서, 예를 들어, 인간에서의 질환의 임의의 치료를 아우르며, (a) 질환의 성향이 있을 수 있지만, 그것을 갖는 것으로 아직 진단되지 않은 대상체에서 질환이 생기는 것을 예방하는 것; (b) 질환을 저해하는 것, 즉, 그의 발생을 저지하는 것; 및 (c) 질환을 완화시키는 것, 즉, 질환의 퇴행을 야기하는 것을 포함한다.

[0029] 본 명세서에서 상호 호환적으로 사용되는 용어 "개체", "대상체", "숙주" 및 "환자"는 무린, 유인원, 인간, 포유류 농장 동물, 포유류 스포츠 동물 및 포유류 반려동물을 포함하지만, 이들로 제한되지 않는 개개 유기체, 예를 들어, 포유류를 지칭한다.

[0030] 본 발명을 추가로 기재하기 전에, 본 발명은 기재한 특정 실시형태로 제한되지 않으며, 그렇게 해서 물론 다를 수 있다는 것이 이해되어야 한다. 또한 본 명세서에 사용되는 용어는 단지 특정 실시형태를 기재할 목적을 위한 것이며, 제한하는 것으로 의도되지 않는데, 본 발명의 범주는 첨부된 청구범위에 의해서만 제한될 것이기 때문이다.

[0031] 수치의 범위가 제공되는 경우에, 문맥에서 달리 분명하게 나타내지 않는 한, 해당 범위 및 해당 언급된 범위에서의 임의의 다른 언급된 또는 개재된 값의 상한과 하한 사이에서 하한 단위의 1/10까지의 각각의 사이의 값은 본 발명 내에 포함된다는 것이 이해된다. 이들 더 작은 범위의 상한 및 하한은 더 작은 범위에 독립적으로 포함될 수 있고, 또한 본 발명 내에 포함되며, 언급된 범위 내에서 임의의 구체적으로 제외된 제한이 적용된다. 언급된 범위가 제한 중 하나 또는 둘 다를 포함하는 경우에, 포함된 제한 중 하나 또는 둘 다를 제외한 범위가 또한 본 발명에 포함된다.

[0032] 달리 정의되지 않는 한, 본 명세서에서 사용되는 모든 기술적 및 과학적 용어는 본 발명이 속하는 기술분야의 당업자에 의해 통상적으로 이해되는 것과 동일한 의미를 가진다. 본 명세서에 기재된 것과 유사하거나 또는 동일한 임의의 방법 및 물질이 또한 본 발명의 실행 또는 시험에서 사용될 수 있지만, 바람직한 방법 및 물질이 이제 기재된다. 본 명세서에서 언급된 모든 간행물은 간행물과 함께 인용된 방법 및/또는 물질을 개시하고 기재하기 위해 본 명세서에 참고로 편입된다.

[0033] 본 명세서에 사용되고 첨부하는 청구범위에 사용되는 바와 같은 단수 형태는 달리 분명하게 나타내지 않는 한 복수의 대상을 포함하는 것임을 주목하여야 한다. 따라서, 예를 들어, "CasY 폴리펩타이드"에 대한 언급은 복수의 이러한 폴리펩타이드를 포함하고, "가이드 RNA"에 대한 언급은 당업자에게 공지된 하나 이상의 가이드 RNA

및 이의 동등물 등에 대한 언급을 포함한다. 청구범위는 임의의 선택적 요소를 제외하도록 작성될 수 있다는 것을 추가로 주목한다. 이렇게 해서, 이러한 언급은 청구범위 요소의 열거와 관련하여 "유일하게", "단지" 등의 사용 또는 "음성" 제한의 사용을 위한 선행 기준으로서 작용하는 것으로 의도된다.

[0034] 명확함을 위해 별개의 실시형태와 관련하여 기재된 본 발명의 특정 특징은 또한 단일 실시형태에서 조합하여 제공될 수 있다는 것이 인식된다. 대조적으로, 간결함을 위해 단일 실시형태와 관련하여 기재된 본 발명의 다양한 특징은 또한 별개로 또는 임의의 적합한 하위-조합으로 제공될 수 있다. 본 발명에 관한 실시형태의 모든 조합은 본 발명에 의해 구체적으로 포함되고, 각각의 그리고 모든 조합이 개별적으로 그리고 명확하게 개시되는 것과 같이 본 명세서에 개시된다. 추가로, 다양한 실시형태의 모든 하위 조합 및 이의 요소는 또한 본 발명에 의해 구체적으로 포함되고, 각각의 그리고 모든 이러한 하위 조합이 개별적으로 그리고 명확하게 본 명세서에 개시되는 것과 같이 본 명세서에 개시된다.

[0035] 본 명세서에서 논의되는 간행물은 본 출원의 출원일 전의 그들의 개시내용에 대해서만 제공된다. 본 명세서의 어떤 것도 본 발명이 선행 발명 때문에 이러한 간행물보다 선행한다는 자격이 부여되지 않는다는 용인으로서 해석되어서는 안 된다. 추가로, 제공된 간행물의 날짜는 독립적으로 확인될 필요가 있을 수 있는 실제 간행물과 상이할 수 있다.

[0036] **상세한 설명**

[0037] 본 발명의 개시내용은 본 명세서에서 "CasY" 폴리펩타이드로서 지칭되는(또한 "CasY 단백질"로서 지칭됨) RNA-가이드 엔도뉴클레아제 폴리펩타이드; CasY 폴리펩타이드를 암호화하는 핵산; 및 CasY 폴리펩타이드를 포함하는 변형된 숙주 세포 및/또는 이를 암호화하는 핵산을 제공한다. 제공되는 CasY 폴리펩타이드는 다양한 적용분야에서 유용하다.

[0038] 본 개시내용은 CasY 단백질에 결합하고 이에 대한 서열 특이성을 제공하는 가이드 RNA(본 명세서에서 "CasY 가이드 RNA"로서 지칭됨); CasY 가이드 RNA를 암호화하는 핵산; 및 CasY 가이드 RNA 및/또는 이를 암호화하는 핵산을 포함하는 변형된 숙주 세포를 제공한다. 제공되는 CasY 가이드 RNA는 다양한 적용분야에서 유용하다.

[0039] 본 개시내용은 CRISPR RNA-가이드된 엔도뉴클레아제를 동정하는 방법을 제공한다.

[0040] **조성물**

[0041] **CRISPR/CasY 단백질 및 가이드 RNA**

[0042] CRISPR/Cas 엔도뉴클레아제(예를 들어, CasY 단백질)는 대응하는 가이드 RNA(예를 들어, CasY 가이드 RNA)와 상호작용(결합)하여 표적 핵산 분자 내에서 가이드 RNA와 표적 서열 사이의 염기쌍을 통해 표적 핵산 내 특정 부위로 표적화된 리보뉴클레오단백질(RNP) 복합체를 형성한다. 가이드 RNA는 표적 핵산의 서열(표적 부위)에 상보적인 뉴클레오타이드 서열(가이드 서열)을 포함한다. 따라서, CasY 단백질은 CasY 가이드 RNA와 복합체를 형성하고, 가이드 RNA는 가이드 서열을 통해 RNP 복합체에 서열 특이성을 제공한다. 복합체의 CasY 단백질은 부위-특이적 활성을 제공한다. 다시 말해서, CasY 단백질은 가이드 RNA와 그의 결합 덕분에 표적 핵산 서열(예를 들어, 염색체 서열 또는 염색체의 서열, 예를 들어, 에피솜 서열, 소형 고리형 서열, 미토콘드리아 서열, 엽록체 서열 등) 내에서 표적 부위(예를 들어, 표적 부위에서 안정화됨)로 가이드된다.

[0043] 본 개시내용은 CasY 폴리펩타이드(및/또는 CasY 폴리펩타이드를 암호화하는 핵산)(예를 들어, CasY 폴리펩타이드가 자연적으로 존재하는 단백질일 수 있는 경우에, 틱내기효소 CasY 단백질, dCasY 단백질, 키메라 CasY 단백질 등)을 포함하는 조성물을 제공한다. 본 개시내용은 CasY 가이드 RNA(및/또는 CasY 가이드 RNA를 암호화하는 핵산)를 포함하는 조성물을 제공한다. 본 개시내용은 (a) CasY 폴리펩타이드(및/또는 CasY 폴리펩타이드를 암호화하는 핵산)(예를 들어, CasY 폴리펩타이드가 자연적으로 존재하는 단백질일 수 있는 경우에, 틱내기효소 CasY 단백질, dCasY 단백질, 키메라 CasY 단백질 등) 및 (b) CasY 가이드 RNA(및/또는 CasY 가이드 RNA를 암호화하는 핵산)를 포함하는 조성물을 제공한다. 본 개시내용은 (a) 본 개시내용의 CasY 폴리펩타이드(예를 들어, CasY 폴리펩타이드가 자연적으로 존재하는 단백질일 수 있는 경우에, 틱내기효소 CasY 단백질, dCasY 단백질, 키메라 CasY 단백질 등); 및 (b) CasY 가이드 RNA를 포함하는, 핵산/단백질 복합체(RNP 복합체)를 제공한다.

[0044] **CasY 단백질**

[0045] CasY 폴리펩타이드(이 용어는 용어 "CasY 단백질"과 상호 호환적으로 사용됨)는 표적 핵산과 결합된 표적 핵산 및/또는 폴리펩타이드에 결합하고/하거나 변형할 수 있다(예를 들어, 절단, 틱내기, 메틸화, 데메틸화 등)(예를 들어, 히스톤 꼬리의 메틸화 또는 아세틸화)(예를 들어, 일부 경우에, CasY 단백질은 활성을 갖는 융합 상대를

포함하고, 일부 경우에 CasY 단백질은 뉴클레아제 활성을 제공한다). 일부 경우에, CasY 단백질은 천연 유래 단백질이다(예를 들어, 원핵 세포에서 자연적으로 생긴다). 다른 경우에, CasY 단백질은 천연 유래 폴리펩타이드가 아니다(예를 들어, CasY 단백질은 변이체 CasY 단백질, 키메라 단백질 등이다).

[0046] 주어진 단백질이 CasY 가이드 RNA와 상호작용하는지의 여부를 결정하는 분석은 단백질과 핵산 사이의 결합에 대해 시험하는 임의의 편리한 결합 분석일 수 있다. 적합한 결합 분석(예를 들어, 겔 이동 분석)은 당업자에게 공지될 것이다(예를 들어, CasY 가이드 RNA 및 표적 핵산에 대한 단백질의 첨가를 포함하는 분석). 단백질이 활성을 갖는지의 여부를 결정하는(예를 들어, 단백질이 표적 핵산을 절단하는 뉴클레아제 활성 및/또는 일부 이중성 활성을 갖는지의 여부를 결정하는) 분석은 임의의 편리한 분석(예를 들어, 핵산 절단에 대해 시험하는 임의의 편리한 핵산 절단 분석)일 수 있다. 적합한 분석(예를 들어, 절단 분석)은 당업자에게 공지될 것이다.

[0047] 천연 유래 CasY 단백질은 표적화된 이중 가닥 DNA(dsDNA) 내 특정 서열에서 이중 가닥 파손을 촉매하는 엔도뉴클레아제로서 작용한다. 서열 특이성은 표적 DNA 내에서 표적 서열에 혼성화하는 회합된 가이드 RNA에 의해 제공된다. 천연 유래 CasY 가이드 RNA는 crRNA이며, 여기서 crRNA는 (i) 표적 DNA에서 표적 서열에 혼성화하는 가이드 서열 및 (ii) CasY 단백질에 결합하는 줄기-루프(헤어핀 - dsRNA 듀플렉스)를 포함하는 단백질 결합 세그먼트를 포함한다.

[0048] 일부 실시형태에서, 대상 방법 및/또는 조성물의 CasY 단백질은 천연 유래(야생형) 단백질이다(또는 이로부터 유래된다). 천연 유래 CasY 단백질의 예는 도 1에 도시하고 서열번호 1 내지 7에 제시한다. 천연 유래 CasY 단백질의 예는 도 1에 도시하고 서열번호 1 내지 8에 제시한다. 천연 유래 CasY 단백질의 정렬 예는 도 2에 제시한다(단백질을 "Y1.", "Y2.", "Y3." 등으로서 표지함). 7개의 천연 유래 CasY CRISPR 좌위(서열분석 데이터로부터 조립)의 부분적 DNA 스캐폴드를 서열번호 21 내지 27로서 제시한다. 이 새로 발견된 단백질(CasY)은 이전에 동정된 CRISPR-Cas 엔도뉴클레아제에 비해 짧고, 따라서 대안으로서 이 단백질의 사용은 단백질을 암호화하는 뉴클레오타이드가 상대적으로 짧다는 이점을 제공한다는 것을 주목하는 것은 중요하다. 이는, 예를 들어, 연구 및/또는 임상 적용을 위한 세포, 예컨대 진핵 세포(예를 들어, 시험관내, 생체외, 생체내 포유류 세포, 인간 세포, 마우스 세포)에 대한 전달을 위해 CasY 단백질을 암호화하는 핵산이 바람직한 경우에, 예를 들어, 바이러스 벡터(예를 들어, AAV 벡터)를 사용하는 경우에 유용하다. 또한 CasY CRISPR 좌위를 보유하는 박테리아는 저온(예를 들어, 10 내지 17°C)에서 수집된 환경 샘플에 존재한다는 것이 본 명세서에서 주목된다. 따라서, CasY는 저온(예를 들어, 10 내지 14°C, 10 내지 17°C, 10 내지 20°C)에서 제대로 작용할 수 있는 것으로 예상된다(예를 들어, 지금까지 발견된 다른 Cas 엔도뉴클레아제보다 더 양호함).

[0049] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100% 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100% 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 단백질의 천연 유래 촉매 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)(예를 들어, 이하에 기재하는 아미노산 위치)을 포함한다는 것을 제외하고, 서열번호 1에 제시된 CasY 단백질 서열을 갖는 아미노산 서열을 포함한다.

[0050] 일부 경우에, CasY 단백질은 서열번호 2에 제시된 CasY 단백질 서열과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100% 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 2에 제시된 CasY 단백질 서열과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100% 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 2에 제시된 CasY 단백질 서열과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상, 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 2에 제시된 CasY 단백질 서열과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을

한다. 일부 경우에, CasY 단백질은 단백질의 천연 유래 촉매 활성을 감소시키는 아미노산 치환(예를 들어, 1, 2 또는 3개의 아미노산 치환)(예를 들어, 이하에 기재하는 아미노산 위치)을 포함한다는 것을 제외하고, 서열번호 1 내지 8 중 임의의 하나에 제시된 CasY 단백질 서열을 갖는 아미노산 서열을 포함한다.

[0062] **CasY 단백질 도메인**

[0063] CasY 단백질의 도메인을 도 3에 도시한다. 도 3의 개략적 표현에 알 수 있는 바와 같이(아미노산은 CasY1 단백질에 기반하여 넘버링됨(서열번호 1)), CasY 단백질은 길이가 거의 800 내지 1000개의 아미노산인 N-말단의 도메인(예를 들어, CasY1에 대해 약 815 및 CasY5에 대해 약 980), 및 CasY 단백질의 주요 아미노산 서열에 대해 인접하지 않지만, 일단 단백질이 생성되고 폴딩된다면 RuvC 도메인을 형성하는 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III, 또한 본 명세서에서 서브도메인으로서 지칭됨)을 포함하는 C-말단의 도메인을 포함한다. 따라서, 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 N-말단 도메인(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음)을 갖는 아미노산 서열을 포함한다. 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III)에 대해 N-말단인 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음) 길이를 갖는 아미노산 서열을 포함한다.

[0064] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열의 N-말단 도메인(예를 들어, 도 3의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인, 패널 a)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열의 N-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 1 내지 812로서 도시된 도메인, 패널 a)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열의 N-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 1 내지 812로서 도시된 도메인, 패널 a)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열의 N-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 1 내지 812로서 도시된 도메인, 패널 a)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열의 아미노산 1 내지 812를 갖는 아미노산 서열을 포함한다.

[0065] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 4 중 임의의 하나에 제시된 CasY 단백질 서열의 N-말단 도메인(예를 들어, 도 3의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인, 패널 a)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 서열번호 1 내지 4 중 임의의 하나에 제시된 CasY 단백질 서열의 N-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 1 내지 812로서 도시된 도메인, 패널 a)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 4 중 임의의 하나에 제시된 CasY 단백질 서열의 N-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 1 내지 812로서 도시된 도메인, 패널 a)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 4 중 임의의 하나에 제시된 CasY 단백질 서열의 N-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 1 내지 812로서 도시된 도메인, 패널 a)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열의 아미노산 1 내지 812에 대응하는 서열번호

서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 서열번호 1 내지 7에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 7에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 7에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1로서 제시된 CasY 단백질 서열의 아미노산 1 내지 812에 대응하는 아미노산 서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II, 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다.

[0072] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 8에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 서열번호 1 내지 8에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 8에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1로서 제시된 CasY 단백질 서열의 아미노산 1 내지 812에 대응하는 아미노산 서열; 및 분할 Ruv C 도메인(예를 들어, 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II, 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다.

[0073] 일부 실시형태에서, (대상 조성물 및/또는 방법의) CasY 단백질의 분할 RuvC 도메인은 RuvC-III 서브도메인보다 더 큰 RuvC-II와 RuvC-III 서브도메인 사이의 영역을 포함한다. 예를 들어, 일부 경우에, RuvC-III 서브도메인의 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1.1 이상(예를 들어, 1.2)이다. 일부 경우에, RuvC-III 서브도메인의 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이다. 일부 경우에, RuvC-III 서브도메인의 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 또는 1 내지 1.2)이다.

[0074] (대상 조성물 및/또는 방법의 CasY 단백질에 대해) 일부 실시형태에서, RuvC-III 서브도메인의 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4

이하)이다. 예를 들어, 일부 경우에, RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하)이다. 일부 실시형태에서, RuvC-III 서브도메인의 길이에 대해 RuvC-II 서브도메인의 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4)의 범위이다.

[0075] (대상 조성물 및/또는 방법의 CasY 단백질에 대해)일부 경우에, RuvC-III 서브도메인의 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이다. 일부 경우에, RuvC-III 서브도메인의 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.3(예를 들어, 1 내지 1.2)이다.

[0076] (대상 조성물 및/또는 방법의 CasY 단백질에 대해) 일부 경우에, RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65, 68 또는 70개의 아미노산)이다. 일부 경우에, RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90 개의 아미노산의 범위)이다.

[0077] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 4에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III-을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다 - 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하)이며; (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0078] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 4로서 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단의 도메인(예를 들어, 도 3, 패널 a의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)을 갖는 75% 이상의 서열 동일성(예를 들어, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단의 제2 아미노산 서열을 포함하며, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-

III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95 개의 아미노산 범위이다.

[0079]

일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 4로서 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단의 도메인(예를 들어, 도 3, 패널 a의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)을 갖는 85% 이상의 서열 동일성(예를 들어, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단의 제2 아미노산 서열을 포함하며, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0080]

일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 5에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III-을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다 - 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산 이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65

내지 95개의 아미노산 범위이다.

[0081] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 5로서 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단의 도메인(예를 들어, 도 3, 패널 a의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)을 갖는 75% 이상의 서열 동일성(예를 들어, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단의 제2 아미노산 서열을 포함하며, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0082] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 5로서 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단의 도메인(예를 들어, 도 3, 패널 a의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)을 갖는 85% 이상의 서열 동일성(예를 들어, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단의 제2 아미노산 서열을 포함하며, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0083] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 7에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III-을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다

- 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산 이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0084] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단의 도메인(예를 들어, 도 3, 패널 a의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)을 갖는 75% 이상의 서열 동일성(예를 들어, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단의 제2 아미노산 서열을 포함하며, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0085] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단의 도메인(예를 들어, 도 3, 패널 a의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)을 갖는 85% 이상의 서열 동일성(예를 들어, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단의 제2 아미노산 서열을 포함하며, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-

III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0086] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 8에 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산 서열; 및 3개의 부분적 RuvC 도메인 - RuvC-I, RuvC-II 및 RuvC-III-을 포함하는 제1 아미노산 서열에 대해 C-말단인 제2 아미노산 서열을 포함한다 - 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산 이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0087] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단의 도메인(예를 들어, 도 3, 패널 a의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)을 갖는 75% 이상의 서열 동일성(예를 들어, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단의 제2 아미노산 서열을 포함하며, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도

70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0088]

일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 N-말단의 도메인(예를 들어, 도 3, 패널 a의 CasY1에 대해 아미노산 1 내지 812로서 도시된 도메인)을 갖는 85% 이상의 서열 동일성(예를 들어, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 제1 아미노산; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 포함하는 제1 아미노산 서열에 대해 C-말단의 제2 아미노산 서열을 포함하며, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0089]

일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 N-말단 도메인(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음)을 갖는 제1 아미노산 서열; 및 3개의 부분적 RuvC 도메인(RuvC-I, RuvC-II 및 RuvC-III)을 갖는 분할 RuvC 도메인을 갖는 제2 아미노산 서열(제1 아미노산 서열에 대해 C-말단)을 포함한다, 여기서: (i) RuvC-III 서브도메인의 길이에 대해 RuvC-II 대 RuvC-III 서브도메인 사이의 영역 길이는 1.1 이상(예를 들어, 1.2)이고; (ii) RuvC-III 서브도메인의 길이에 대해 RuvC-II 내지 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이고; (iii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.4, 1 내지 1.3, 1 내지 1.2)이며; (iv) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 2 이하(예를 들어, 1.8 이하, 1.7 이하, 1.6 이하, 1.5 이하 또는 1.4 이하)이고; (v) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1.5 이하(예를 들어, 1.4 이하); (vi) RuvC-III 서브도메인 길이에 대해 RuvC-II 서브도메인 길이는 1 내지 2(예를 들어, 1.1 내지 2, 1.2 내지 2, 1 내지 1.8, 1.1 내지 1.8, 1.2 내지 1.8, 1 내지 1.6, 1.1 내지 1.6, 1.2 내지 1.6, 1 내지 1.4, 1.1 내지 1.4, 또는 1.2 내지 1.4) 범위이고; (vii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과이며; (viii) RuvC-III 서브도메인 길이에 대해 RuvC-II와 RuvC-III 서브도메인 사이의 영역의 길이는 1 초과 및 1 내지 1.5(예를 들어, 1 내지 1.2) 이고; (ix) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 60개의 아미노산(예를 들어, 길이가 적어도 65 또는 적어도 70개의 아미노산)이며; (x) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 적어도 65개의 아미노산이고; (xi) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 60 내지 110개의 아미노산 범위(예를 들어, 60 내지 105, 60 내지 100, 60 내지 95, 60 내지 90, 65 내지 110, 65 내지 105, 65 내지 100, 65 내지 95, 또는 65 내지 90개의 아미노산 범위)이거나; 또는 (xii) RuvC-II와 RuvC-III 서브도메인 사이의 영역은 길이가 65 내지 95개의 아미노산 범위이다.

[0093] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3의 CasY1에 대해 아미노산 812 내지 1125로서 도시된 도메인, 패널 a)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인, 패널 a)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 7 중 임의의 하나에 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인, 패널 a)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인, 패널 a)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열의 아미노산 812 내지 1125에 대응하는 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 아미노산 서열의 단편을 포함한다.

[0094] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3의 CasY1에 대해 아미노산 812 내지 1125로서 도시된 도메인, 패널 a)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인, 패널 a)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 8 중 임의의 하나에 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인, 패널 a)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3의 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인, 패널 a)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 서열번호 1에 제시된 CasY 단백질 서열의 아미노산 812 내지 1125에 대응하는 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 아미노산 서열의 단편을 포함한다.

[0095] 일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 4로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 4로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에,

CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 4로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 4로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1에 제시된 CasY 단백질 서열의 아미노산 812 내지 1125에 대응하는 서열번호 1 내지 4로서 제시된 CasY 단백질 중 임의의 하나의 아미노산 서열의 단편을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다.

[0096]

일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 5로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 5로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 5로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 5로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97%

이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1에 제시된 CasY 단백질 서열의 아미노산 812 내지 1125에 대응하는 서열번호 1 내지 5로서 제시된 CasY 단백질 중 임의의 하나의 아미노산 서열의 단편을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다.

[0097]

일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 7로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1에 제시된 CasY 단백질 서열의 아미노산 812 내지 1125에 대응하는 서열번호 1 내지 7로서 제시된 CasY 단백질 중 임의의 하나의 아미노산 서열의 단편을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다.

[0098]

일부 경우에, (대상 조성물 및/또는 방법의) CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3,

패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 20% 이상의 서열 동일성(예를 들어, 30% 이상, 40% 이상, 50% 이상, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 예를 들어, 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 50% 이상의 서열 동일성(예를 들어, 60% 이상, 70% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 80% 이상의 서열 동일성(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1 내지 8로서 제시된 CasY 단백질 서열 중 임의의 하나의 C-말단 도메인(예를 들어, 도 3, 패널 a에서 CasY1에 대한 아미노산 812 내지 1125로서 도시된 도메인)과 90% 이상의 서열 동일성(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%의 서열 동일성)을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다. 일부 경우에, CasY 단백질은 길이가 750 내지 1050개의 아미노산(예를 들어, 750 내지 1025, 750 내지 1000, 750 내지 950, 775 내지 1050, 775 내지 1025, 775 내지 1000, 775 내지 950, 800 내지 1050, 800 내지 1025, 800 내지 1000, 또는 800 내지 950개의 아미노산) 범위인 제1 아미노산 서열(N-말단 도메인)(예를 들어, 임의의 융합된 이중성 서열, 예컨대 NLS 및/또는 촉매 활성을 갖는 도메인을 포함하지 않음); 및 서열번호 1에 제시된 CasY 단백질 서열의 아미노산 812 내지 1125에 대응하는 서열번호 1 내지 8로서 제시된 CasY 단백질 중 임의의 하나의 아미노산 서열의 단편을 갖는, 제1 아미노산 서열에 대해 C-말단에 위치한 제2 아미노산 서열을 포함한다.

[0099] CasY 변이체

[0100] 변이체 CasY 단백질은 대응하는 야생형 CasY 단백질의 아미노산 서열에 비교할 때 적어도 하나의 아미노산만큼 상이한 아미노산 서열을 가진다(예를 들어, 결실, 삽입, 치환, 융합을 가진다). 이중 가닥 표적 핵산 중 한 가닥을 절단하지만, 다른 한 가닥은 절단하지 않는 CasY 단백질은 본 명세서에서 "틈내기효소"(예를 들어, "틈내기효소 CasY"로서 지칭된다. 뉴클레아제 활성이 실질적으로 없는 CasY 단백질은 본 명세서에서 사멸 CasY 단백질("dCasY"로서 지칭된다(뉴클레아제 활성이 이하에 상세하게 기재되는 키메라 CasY 단백질의 경우에 이중성 폴리펩타이드(융합 상대)에 의해 제공될 수 있다는 경고가 있음). 본 명세서에 기재된 임의의 CasY 변이체 단백질(예를 들어, 틈내기효소 CasY, dCasY, 키메라 CasY)에 대해, CasY 변이체는 상기 기재된 것과 동일한 매개변수(예를 들어, 존재하는 도메인, 동일성 백분율 등)를 갖는 CasY 단백질 서열을 포함할 수 있다.

[0101] 변이체 - 촉매 활성

[0102] 일부 경우에, CasY 단백질은, 예를 들어, 천연 유래 촉매적으로 활성인 서열에 대해 돌연변이된 변이체 CasY 단백질이고, 대응하는 천연 유래 서열에 비교할 때 감소된 절단 활성을 나타낸다(예를 들어, 90%, 이하, 80% 이하, 70% 이하, 60% 이하, 50% 이하, 40% 이하 또는 30% 이하의 절단 활성을 나타낸다). 일부 경우에, 이러한 변이체 CasY 단백질은 촉매적으로 '사멸' 단백질이고(절단 활성이 실질적으로 없음) 'dCasY'로서 지칭될 수 있다. 일부 경우에, 변이체 CasY 단백질은 틈내기효소이다(이중 가닥 표적 핵산, 예를 들어, 이중 가닥 표적 DNA의 단지 하나의 가닥을 절단함). 본 명세서에서 더 상세하게 기재되는 바와 같이, 일부 경우에, CasY 단백질(일부 경우에, 야생형 절단 활성을 갖는 CasY 단백질 및 일부 경우에 감소된 절단 활성을 갖는 변이체 CasY, 예를

들어, dCasY 또는 틱내기효소 CasY)은 융합 단백질(키메라 CasY 단백질)을 형성하기 위해 관심 대상의 활성화(예를 들어, 관심 대상의 촉매 활성화)를 갖는 이중성 폴리펩타이드에 융합된다(접합된다).

[0103] CasY의 촉매 잔기는 CasY1에 따라 넘버링될 때 D828, E914, D1074(서열번호 1)를 포함한다(이들 잔기는 서열번호 1에 대해 도 1에 밑줄 표시되어 있다). (예를 들어, 도 2, 패널 a 및 b의 정렬).

[0104] 따라서, 일부 경우에, CasY 단백질은 감소된 활성을 가지고, 상기 기재된 아미노산 중 하나 이상(또는 임의의 CasY 단백질의 하나 이상의 대응하는 아미노산)이 돌연변이된다(예를 들어, 알라닌으로 치환된다). 일부 경우에, 변이체 CasY 단백질은 촉매적으로 '사멸' 단백질이고(촉매적으로 비활성임), 'dCasY'로서 지칭된다. dCasY 단백질은 활성을 제공하는 융합 상대에 융합될 수 있고, 그리고 일부 경우에, dCasY(예를 들어, 촉매 활성을 제공하지만, 진핵 세포에서 발현될 때 NLS를 가질 수 있는 융합 상대가 없는 하나)는 표적 DNA에 결합할 수 있고, 표적 DNA로부터의 번역으로부터 RNA 중합효소를 차단할 수 있다. 일부 경우에, 변이체 CasY 단백질은 틱내기효소이다(이중 가닥 표적 핵산, 예를 들어, 이중 가닥 표적 DNA의 단지 하나의 가닥을 절단함).

[0105] **변이체 - 키메라 CasY(즉, 융합 단백질)**

[0106] 상기 언급한 바와 같이, 일부 경우에, CasY 단백질(일부 경우에, 야생형 절단 활성을 갖는 CasY 단백질 및 일부 경우에 감소된 절단 활성을 갖는 변이체 CasY, 예를 들어, dCasY 또는 틱내기효소 CasY)은 융합 단백질(키메라 CasY 단백질)을 형성하기 위해 관심 대상의 활성화(예를 들어, 관심 대상의 촉매 활성화)를 갖는 이중성 폴리펩타이드에 융합된다(접합된다). CasY 단백질이 융합될 수 있는 이중성 폴리펩타이드는 본 명세서에서 '융합 상대'로서 지칭된다.

[0107] 일부 경우에, 융합 상대는 표적 DNA의 전사를 조절할 수 있다(예를 들어, 전사를 저해, 전사를 증가시킬 수 있다). 예를 들어, 일부 경우에 융합 상대는 전사를 저해하는 단백질(또는 단백질로부터의 도메인)(예를 들어, 전사 리프레서, 전사 저해제 단백질의 보충을 통해 작용하는 단백질, 표적 DNA의 변형, 예컨대 메틸화, DNA 변형제의 보충, 표적 DNA와 회합된 히스톤의 조절, 히스톤 변형제, 예컨대 히스톤의 아세틸화 및/또는 메틸화를 변형시키는 것의 보충 등)이다. 일부 경우에 융합 상대는 전사를 증가시키는 단백질(또는 단백질로부터의 도메인)(예를 들어, 전사 활성화제, 전사 활성화제 단백질의 보충을 통해 작용하는 단백질, 표적 DNA의 변형, 예컨대 탈메틸화의 변형, DNA 변형제의 보충, 표적 DNA와 회합된 히스톤의 조절, 히스톤 변형제, 예컨대 히스톤의 아세틸화 및/또는 메틸화를 변형시키는 것의 보충 등)이다.

[0108] 일부 경우에, 키메라 CasY 단백질은 표적 핵산을 변형시키는 효소 활성화(예를 들어, 뉴클레아제 활성화, 메틸트랜스퍼라제 활성화, 데메틸라제 활성화, DNA 수선 활성화, DNA 손상 활성화, 탈아미노화 활성화, 디스무타제 활성화, 알킬화 활성화, 탈퓨린화 활성화, 산화 활성화, 피리미딘 이량체 형성 활성화, 인테그라제 활성화, 트랜스포사제 활성화, 재조합효소 활성화, 중합효소 활성화, 리가제 활성화, 헬리카제 활성화, 광분해효소 활성화 또는 글리코실라제 활성화)을 갖는 이중성 폴리펩타이드를 포함한다.

[0109] 일부 경우에, 키메라 CasY 단백질은 표적 핵산과 회합된 폴리펩타이드(예를 들어, 히스톤)를 변형하는 효소 활성화(예를 들어, 메틸트랜스퍼라제 활성화, 데메틸라제 활성화, 아세틸트랜스퍼라제 활성화, 데아세틸라제 활성화, 키나제 활성화, 포스파타제 활성화, 유비퀴틴 리가제 활성화, 데유비퀴틴화 활성화, 아데닐화 활성화, 탈아데닐화 활성화, 수모일화(SUMOylating) 활성화, 탈수모일화(deSUMOylating) 활성화, 리보실화 활성화, 탈리보실화 활성화, 미리스토일화 활성화 또는 탈미리스토일화 활성화)을 갖는 이중성 폴리펩타이드를 포함한다.

[0110] 전사 증가에서 사용될 수 있는 단백질(또는 이의 단편)의 예는 전사 활성화제, 예컨대 VP16, VP64, VP48, VP160, p65 서브도메인(예를 들어, NFkB로부터 유래), 및 EDLL의 활성화 도메인 및/또는 TAL 활성화 도메인(예를 들어, 식물의 활성화에 대해); 히스톤 라이신 메틸트랜스퍼라제, 예컨대 SET1A, SET1B, MLL1 내지 5, ASH1, SYMD2, NSD1 등; 히스톤 라이신 데메틸라제, 예컨대 JHDM2a/b, UTX, JMJD3 등; 히스톤 아세틸트랜스퍼라제, 예컨대 GCN5, PCAF, CBP, p300, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, SRC1, ACTR, P160, CLOCK 등; 및 DNA 데메틸라제, 예컨대 염색체 10번과 11번 사이의 전좌(Ten-Eleven Translocation: TET) 다이옥시게나제 1(TET1CD), TET1, DME, DML1, DML2, ROS1 등을 포함하지만, 이들로 제한되지 않는다.

[0111] 전사 감소에서 사용될 수 있는 단백질(또는 이의 단편)의 예는 전사 리프레서, 예컨대 쿠루펠(Kruppel) 관련 박스(KRAB 또는 SKD); KOX1 억제 도메인; Mad mSIN3 상호작용 도메인(SID); ERF 리프레서 도메인(ERD), SRDX 억제 도메인(예를 들어, 식물에서의 억제를 위해) 등; 히스톤 라이신 메틸트랜스퍼라제, 예컨대 Pr-SET7/8, SUV4-20H1, RIZ1 등; 히스톤 라이신 데메틸라제, 예컨대 JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY 등; 히스톤 라이신 데아세틸라제, 예컨대 HDAC1,

HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11 등; DNA 메틸라제, 예컨대 HhaI DNA m5c-메틸트랜스퍼라제(M.HhaI), DNA 메틸트랜스퍼라제 1(DNMT1), DNA 메틸트랜스퍼라제 3a(DNMT3a), DNA 메틸트랜스퍼라제 3b(DNMT3b), MET1, DRM3(식물), ZMET2, CMT1, CMT2(식물) 등; 및 주변 보충 요소, 예컨대 라민(Lamin) A, 라민 B 등을 포함하지만, 이들로 제한되지 않는다.

[0112] 일부 경우에, 융합 상대는 표적 핵산(예를 들어, ssRNA, dsRNA, ssDNA, dsDNA)을 변형시키는 효소 활성을 가진다. 융합 상대에 의해 제공될 수 있는 효소 활성의 예는 뉴클레아제 활성, 예컨대 제한 효소(예를 들어, FokI 뉴클레아제)에 의해 제공되는 것, 메틸트랜스퍼라제 활성, 예컨대 메틸트랜스퍼라제(예를 들어, HhaI DNA m5c-메틸트랜스퍼라제(M.HhaI), DNA 메틸트랜스퍼라제 1(DNMT1), DNA 메틸트랜스퍼라제 3a(DNMT3a), DNA 메틸트랜스퍼라제 3b(DNMT3b), MET1, DRM3(식물), ZMET2, CMT1, CMT2(식물) 등)에 의해 제공되는 것; 데메틸라제 활성, 예컨대 데메틸라제(예를 들어, 염색체 10번과 11번 사이의 전좌(TET) 다이옥시게나제 1(TET1CD), TET1, DME, DML1, DML2, ROS1 등)에 의해 제공되는 것, DNA 수선 활성, DNA 손상 활성, 탈아미노화 활성, 예컨대 데아미나제(예를 들어, 사이토신 데아미나제 효소, 예컨대 래트 APOBEC1)에 의해 제공되는 것, 디스무타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 예컨대 인테그라제 및/또는 레슬바제(예를 들어, Gin 인버타제, 예컨대 Gin 인버타제의 과활성 돌연변이체, GinH106Y; 인간 면역결핍 바이러스 1형 인테그라제(IN); Tn3 레슬바제; 등)에 의해 제공되는 것, 트랜스포사제 활성, 재조합효소 활성, 예컨대 재조합효소(예를 들어, Gin 재조합효소의 촉매 도메인)에 의해 제공되는 것, 중합효소 활성, 리가제 활성, 헬리카제 활성, 광분해효소 활성, 및 글리코실라제 활성)를 포함하지만, 이들로 제한되지 않는다.

[0113] 일부 경우에, 융합 상대는 표적 핵산(예를 들어, ssRNA, dsRNA, ssDNA, dsDNA)과 회합되는 단백질(예를 들어, 히스톤, RNA 결합 단백질, DNA 결합 단백질 등)을 변형시키는 효소 활성을 가진다. 융합 상대에 의해 제공될 수 있는 (표적 핵산과 회합된 단백질을 변형시키는) 효소 활성의 예는 메틸트랜스퍼라제 활성, 예컨대 히스톤 메틸트랜스퍼라제(HMT)에 의해 제공되는 것(예를 들어, 얼룩 3-9 상동체 1의 억제자(SUV39H1, 또한 KMT1A로서 알려짐), 진정염색성 히스톤 라이신 메틸트랜스퍼라제 2(G9A, 또한 KMT1C 및 EHMT2로서 알려짐), SUV39H2, ESET/SETDB1 등, SET1A, SET1B, MLL1 내지 5, ASH1, SYMD2, NSD1, DOT1L, Pr-SET7/8, SUV4-20H1, EZH2, RIZ1), 데메틸라제 활성, 예컨대 히스톤 데메틸라제(예를 들어, 라이신 데메틸라제 1A(또한 LSD1로서 알려진 KDM1A), JHDM2a/b, JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, UTX, JMJD3 등)에 의해 제공되는 것, 아세틸트랜스퍼라제 활성, 예컨대 히스톤 아세틸라제 트랜스퍼라제에 의해 제공되는 것(예를 들어, 촉매 코어/인간 아세틸트랜스퍼라제 p300, GCN5, PCAF, CBP, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, HBO1/MYST2, HMOF/MYST1, SRC1, ACTR, P160, CLOCK 등의 단편), 데아세틸라제 활성, 예컨대 히스톤 데아세틸라제(예를 들어, HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11 등)에 의해 제공되는 것, 키나제 활성, 포스포타제 활성, 유비퀴틴 리가제 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모일화 활성, 탈수모일화 활성, 리보실화 활성, 데리보실화 활성, 미리스토일화 활성, 및 데미리스토일화 활성을 포함하지만, 이들로 제한되지 않는다.

[0114] 적합한 융합 상대의 추가적인 예는 다이하이드로엽산 환원효소(DHFR) 탈안정화 도메인(예를 들어, 화학적으로 제어 가능한 키메라 키메라 CasY 단백질을 생성), 및 엽록체 수송 펩타이드이다. 적합한 엽록체 수송 펩타이드는 하기를 포함하지만, 이들로 제한되지 않는다:

MASMISSSAVTTVSRRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRV
KCMQVWPPIGKKKFETLSYLPPLTRDSRA (서열번호 83);
MASMISSSAVTTVSRRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS (SEQ ID
NO:84);
MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQVWPPIEK
KKFETLSYLPDLTDSGGRVNC (서열번호 85);
MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRL
KVMSSVSTAC (서열번호 86);
MAQVSRICNGVWNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRL
KVMSSVSTAC (서열번호 87);
MAQINMAQGIQTLNPNFHKPQVPKSSFLVFGSKKLKNSANSMLVLKKDSIFMQLFCSFRIS
ASVATAC (서열번호 88);
MAALVTSQLATSGTVLSVTDRFRPPGFQGLRPRNPADAALGMRTVGASAAPKQSRKPHRFDRR
CLSMVV (서열번호 89);
MAALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQQRSVQ
RGSRRFPSVVVC (서열번호 90);
MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDTIASNGGRVQC (서열번호
91);
MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVISRSAAAA
(서열번호 92); 및
MGAAATSMQSLKFSNRLVPPSRRLSPVPNNVTCNNLPKSAAPVRTVKCCASSWNSTINGAAATT
NGASAASS (서열번호 93).

[0115]

[0116]

일부 경우에, 본 개시내용의 CasY 융합 폴리펩타이드는 a) 본 개시내용의 CasY 폴리펩타이드; 및 b) 엽록체 수송 펩타이드를 포함한다. 따라서, 예를 들어, CRISPR-CasY 복합체는 엽록체에 표적화될 수 있다. 일부 경우에, 이 표적화는 엽록체 수송 펩타이드(CTP) 또는 색소체 수송 펩타이드로 불리는 N-말단 연장 존재에 의해 달성될 수 있다. 발현된 폴리펩타이드가 식물 색소체(예를 들어, 엽록체)에서 구획화된다면, 박테리아 공급원으로부터의 엽록체 이식유전자는 발현된 폴리펩타이드를 암호화하는 서열에 융합된 CTP 서열을 암호화하는 서열을 가져야 한다. 따라서, 엽록체에 대한 외인성 폴리펩타이드의 국소화는 종종 외인성 폴리펩타이드를 암호화하는 폴리뉴클레오타이드의 5' 영역에 CTP 서열을 암호화하는 폴리뉴클레오타이드 서열을 작동 가능하게 연결함으로써 달성된다. CTP는 색소체 내로 전파 동안 가공 단계에서 제거된다. 그러나, 가공 효율은 펩타이드의 CTP의 서열 그리고 NH 2 말단에서 근처의 서열에 의해 영향받을 수 있다. 기재된 엽록체에 대한 표적화를 위한 다른 선택은 마이스(maize) cab-m7 신호 서열(미국 특허 제7,022,896호, WO 97/41228) 완두콩 글루타티온 환원효소 신호 서열(WO 97/41228) 및 미국 특허 제2009029861호에 기재된 CTP이다.

[0117]

일부 경우에, 본 개시내용의 CasY 융합 폴리펩타이드는 a) 본 개시내용의 CasY 폴리펩타이드; 및 b) 엔도솜 탈출 펩타이드를 포함할 수 있다. 일부 경우에, 엔도솜 탈출 폴리펩타이드는 아미노산 서열 GLFXALLXLLXSLWLLXLA(서열번호 94)를 포함하되, 각각의 X는 라이신, 히스티딘 및 알기닌으로부터 독립적으로 선택된다. 일부 경우에, 엔도솜 탈출 폴리펩타이드는 아미노산 서열 GLFHALLHLLHSLWLLHHA(서열번호 95)를 포함한다.

[0118]

(부위 특이적 표적 변형, 전사의 조절 및/또는 표적 단백질 변형, 예를 들어, 히스톤 변형에 대해) Cas9, 아연 핑거 및/또는 TALE 단백질과의 융합과 관련하여 사용되는 상기 융합 상대의 일부(그리고 더 많은) 예는, 예를 들어, 문헌[Nomura et al., J Am Chem Soc. 2007 Jul 18;129(28):8676-7; Rivenbark et al., Epigenetics. 2012 Apr;7(4):350-60; Nucleic Acids Res. 2016 Jul 8;44(12):5615-28; Gilbert et al., Cell. 2013 Jul 18;154(2):442-51; Kearns et al., Nat Methods. 2015 May;12(5):401-3; Mendenhall et al., Nat Biotechnol. 2013 Dec;31(12):1133-6; Hilton et al., Nat Biotechnol. 2015 May;33(5):510-7; Gordley et al., Proc Natl Acad Sci U S A. 2009 Mar 31;106(13):5053-8; Akopian et al., Proc Natl Acad Sci U S A. 2003 Jul 22;100(15):8688-91; Tan et al., J Virol. 2006 Feb;80(4):1939-48; Tan et al., Proc Natl Acad Sci U S A. 2003 Oct 14;100(21):11997-2002; Papworth et al., Proc Natl Acad Sci U S A. 2003 Feb 18;100(4):1621-6; Sanjana et al., Nat Protoc. 2012 Jan 5;7(1):171-92; Beerli et al., Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):14628-33; Snowden et al., Curr Biol. 2002 Dec 23;12(24):2159-66; Xu

et.al., Xu et al., Cell Discov. 2016 May 3;2:16009; Komor et al., Nature. 2016 Apr 20;533(7603):420-4; Chaikind et al., Nucleic Acids Res. 2016 Aug 11; Choudhury et al., Oncotarget. 2016 Jun 23; Du et al., Cold Spring Harb Protoc. 2016 Jan 4; Pham et al., Methods Mol Biol. 2016;1358:43-57; Balboa et al., Stem Cell Reports. 2015 Sep 8;5(3):448-59; Hara et al., Sci Rep. 2015 Jun 9;5:11221; Piatek et al., Plant Biotechnol J. 2015 May;13(4):578-89; Hu et al., Nucleic Acids Res. 2014 Apr;42(7):4375-90; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; 및 Maeder et al., Nat Methods. 2013 Oct;10(10):977-9] 참조.

[0119] 추가적인 적합한 이중성 폴리펩타이드는 표적 핵산의 증가된 전사 및/또는 번역을 직접적으로 그리고/또는 간접적으로 제공하는 폴리펩타이드(예를 들어, 전사 활성체 또는 이의 단편, 전사 활성체를 보충하는 단백질 또는 이의 단편, 소분자/약물-반응성 전사 및/또는 번역 조절제, 번역-조절 단백질 등)를 포함하지만, 이들로 제한되지 않는다. 증가된 또는 감소된 전사를 달성하기 위한 이중성 폴리펩타이드의 비제한적 예는 전사 활성체 및 전사 리프레이저 도메인을 포함한다. 일부 이러한 경우에, 키메라 CasY 폴리펩타이드는 표적 핵산에서 특정 위치(즉, 서열)에 대해 가이드 핵산(가이드 RNA)에 의해 표적화되고, (예를 들어, 표적 핵산을 변형시키거나 또는 표적 핵산과 회합된 폴리펩타이드를 변형시키는 융합 서열이 사용될 때) 프로모터(전사 활성체 기능을 선택적으로 저해)에 대한 RNA 중합효소 결합을 차단하고/하거나 국소 염색질 상태를 변형시키는 것과 같은 좌위-특이적 조절을 발휘한다. 일부 경우에, 변화는 일시적이다(예를 들어, 전사 억제 또는 활성화). 일부 경우에, 변화는 유전성이다(예를 들어, 표적 핵산에 대해 또는 표적 핵산과 회합된 단백질, 예를 들어, 뉴클레오솜 히스톤에 대해 후성적 변형이 만들어질 때).

[0120] ssRNA 표적 핵산을 표적화할 때 사용하기 위한 이중성 폴리펩타이드의 비제한적 예는 스플라이싱 인자(예를 들어, RS 도메인); 단백질 번역 성분(예를 들어, 번역 개시, 신장 및/또는 방출 인자; 예를 들어, eIF4G); RNA 메틸라제; RNA 편집 효소(예를 들어, RNA 탈아미나제, 예를 들어, RNA 상에서 작용하는 아데노신 데아미나제 (ADAR)(A 내지 I 및/또는 C 내지 U 편집 효소를 포함)); 헬리카제; RNA-결합 단백질 등을 포함한다(그러나 이들로 제한되지 않음). 이중성 폴리펩타이드는 전체 단백질을 포함할 수 있거나 또는 일부 경우에 단백질(예를 들어, 기능성 도메인)의 단편을 포함할 수 있다는 것이 이해된다.

[0121] 대상 키메라 CasY 폴리펩타이드의 이중성 폴리펩타이드는 일시적이든 또는 비가역적이든, 직접적이든 또는 간접적이든; 엔도뉴클레아제(예를 들어, SMG5 및 SMG6과 같은 단백질로부터의 RNase III, CRR22 DYW 도메인, Dicer, 및 PIN(PilT N-말단) 도메인); RNA 절단의 자극을 초래하는 단백질 및 단백질 도메인(예를 들어, CPSF, CstF, CFIm 및 CFIIIm); 엑소뉴클레아제(예를 들어, XRN-1 또는 엑소뉴클레아제 T); 데아데닐라제(예를 들어, HNT3); 넌센스 매개된 RNA 붕괴를 초래하는 단백질 및 단백질 도메인(예를 들어, UPF1, UPF2, UPF3, UPF3b, RNP S1, Y14, DEK, REF2, 및 SRm160); RNA 안정화를 초래하는 단백질 및 단백질 도메인(예를 들어, PABP); 번역 억제를 초래하는 단백질 및 단백질 도메인(예를 들어, Ago2 및 Ago4); 번역 자극을 초래하는 단백질 및 단백질 도메인(예를 들어, Staufen); 번역 조절을 초래하는(예를 들어, 초래할 수 있는) 단백질 및 단백질 도메인(예를 들어, 번역 인자, 예컨대 개시 인자, 신장 인자, 방출 인자 등, 예를 들어, eIF4G); RNA의 폴리아데닐화를 초래하는 단백질 및 단백질 도메인(예를 들어, PAP1, GLD-2, 및 Star-PAP); RNA의 폴리우리딘일화를 초래하는 단백질 및 단백질 도메인(예를 들어, CI D1 및 말단의 우리딜레이트 트랜스퍼라제); RNA 국소화를 초래하는 단백질 및 단백질 도메인(예를 들어, IMP1, ZBP1, She2p, She3p 및 Bicaudal-D로부터의); RNA의 핵 보유를 초래하는 단백질 및 단백질 도메인(예를 들어, Rrp6); RNA의 핵 수송을 초래하는 단백질 및 단백질 도메인(예를 들어, TAP, NXF1, THO, TREX, REF 및 Aly); RNA 스플라이싱의 억제를 초래하는 단백질 및 단백질 도메인(예를 들어, PTB, Sam68 및 hnRNP A1); RNA 스플라이싱의 자극을 초래하는 단백질 및 단백질 도메인(예를 들어, 세린/알기닌-풍부(SR) 도메인); 전사 효율 감소를 초래하는 단백질 및 단백질 도메인(예를 들어, FUS (TLS)); 및 전사 자극을 초래하는 단백질 및 단백질 도메인(예를 들어, CDK7 및 HIV Tat)을 포함하는 군으로부터 선택된 효과기 도메인을 포함하지만, 이들로 제한되지 않는, ssRNA와 상호작용할 수 있는 임의의 도메인(본 개시내용의 목적을 위해, 분자내 및/또는 분자간 2차 구조, 예를 들어, 이중-가닥 RNA 듀플렉스, 예컨대 헤어핀, 줄기-루프 등을 포함함)일 수 있다. 대안적으로, 효과기 도메인은 엔도뉴클레아제; RNA 절단을 자극할 수 있는 단백질 및 단백질 도메인; 엑소뉴클레아제; 데아데닐라제; 넌센스 매개된 RNA 붕괴 활성을 갖는 단백질 및 단백질 도메인; RNA를 안정화시킬 수 있는 단백질 및 단백질 도메인; 번역을 억제할 수 있는 단백질 및 단백질 도메인; 번역을 자극할 수 있는 단백질 및 단백질 도메인; 번역을 조절할 수 있는 단백질 및 단백질 도메인(예를 들어, 번역 인자, 예컨대 개시 인자, 신장 인자, 방출 인자 등, 예를 들어, eIF4G); RNA의 폴리아데닐화를 가능하게 하는 단백질 및 단백질 도메인; RNA의 폴리우리딘일화를 가능하게 하는 단백질 및 단백질 도메인; RNA 국소화 활성을 갖는 단백질 및 단백질 도메인; RNA의 핵 보유를 가능하게 하는 단백질 및 단백질 도메인; RNA 핵 수송 활성을

갖는 단백질 및 단백질 도메인; RNA 스플라이싱의 억제를 가능하게 하는 단백질 및 단백질 도메인; RNA 스플라이싱의 자극을 가능하게 하는 단백질 및 단백질 도메인; 전사 효율을 감소시킬 수 있는 단백질 및 단백질 도메인; 및 전사를 자극할 수 있는 단백질 및 단백질 도메인을 포함하는 군으로부터 선택될 수 있다. 다른 적합한 이중성 폴리펩타이드는 본 명세서에 전문이 참고로 편입된 W02012068627에서 더 상세하게 기재되는 PUF RNA-결합 도메인이다.

[0122] 키메라 CasY 폴리펩타이드에 대한 이중성 폴리펩타이드로서 사용될 수 있는(전체로 또는 이의 단편으로서) 일부 RNA 스플라이싱 인자는 별개의 서열-특이적 RNA 결합 분자 및 스플라이싱 효과기 도메인과 함께 모듈 조직화를 가진다. 예를 들어, 세린/알라닌-풍부(SR) 단백질 패밀리의 구성원은 엑손 포함을 촉진하는 프레-mRNA 및 C-말단의 RS 도메인에서 엑손 스플라이싱 인핸서(ESE)에 결합하는 N-말단의 RNA 인식 모티프(RRM)를 함유한다. 다른 예로서, hnRNP 단백질 hnRNP A1은 그의 RRM 도메인을 통해 엑손 스플라이싱 사일런서(ESS)에 결합하고 C-말단의 글리신-풍부 도메인을 통해 엑손 포함을 저해한다. 일부 스플라이싱 인자는 두 대안의 부위 사이의 조절 서열에 대한 결합에 의해 스플라이스 부위(ss)의 대안의 사용을 조절할 수 있다. 예를 들어, ASF/SF2는 ESE를 인식하고 인트론 근위 부위의 사용을 촉진시킬 수 있는 반면, hnRNP ESS에 결합하고 A1은 인트론 원위 부위의 사용에 대해 스플라이싱을 이동시킬 수 있다. 이러한 인자의 한 가지 적용은 내인성 유전자, 특히 질환 관련 유전자의 대안의 스플라이싱을 조절하는 ESF를 생성하는 것이다. 예를 들어, Bcl-x pre-mRNA는 반대 기능의 단백질을 암호화하기 위해 2개의 대안의 5' 스플라이스 부위를 갖는 2개의 스플라이싱 아이소폼을 생성한다. 긴 스플라이싱 아이소폼 Bcl-xL은 장기간 생존된 유사분열후(postmitotic) 세포에서 발현된 강력한 세포자멸사 저해제이고, 다수의 암세포, 세포자멸사 신호에 대한 보호 세포에서 상향조절된다. 짧은 아이소폼 Bcl-xS는 세포자멸사전 아이소폼이고, 높음 전환율을 갖는 세포(예를 들어, 발생 중인 림프구)에서 고수준으로 발현된다. 2개의 Bcl-x 스플라이싱 아이소폼의 비는 코어 엑손 영역 또는 엑손 연장 영역 중 하나에(즉, 두 대안의 5' 스플라이스 부위 사이에) 위치된 다중 cω'-요소에 의해 조절된다. 더 많은 예에 대해, 본 명세서에 전문이 참고로 편입된 W02010075303 참조.

[0123] 추가적인 적합한 융합 상대는 경계 요소인 단백질(또는 이의 단편)(예를 들어, CTCF), 주변 보충을 제공하는 단백질 및 이의 단편(예를 들어, 라민 A, 라민 B 등), 단백질 도킹 요소(예를 들어, FKBP/FRB, Pil1/Aby1 등)를 포함하지만, 이들로 제한되지 않는다.

[0124] 대상 키메라 CasY 폴리펩타이드에 대한 다양한 추가적인 적합한 이중성 폴리펩타이드(또는 이의 단편)의 예는 다음의 요소에 기재된 것을 포함하지만, 이들로 제한되지 않는다(이들 간행물은 다른 CRISPR 엔도뉴클레아제, 예컨대 Cas9와 관련되지만, 기재된 융합 상대는 또한 CasY 대신 사용될 수 있다): PCT 특허 출원: W02010075303, W02012068627, 및 W02013155555, 그리고, 예를 들어 미국 특허 및 특허 출원에서 찾을 수 있다: 8,906,616; 8,895,308; 8,889,418; 8,889,356; 8,871,445; 8,865,406; 8,795,965; 8,771,945; 8,697,359; 20140068797; 20140170753; 20140179006; 20140179770; 20140186843; 20140186919; 20140186958; 20140189896; 20140227787; 20140234972; 20140242664; 20140242699; 20140242700; 20140242702; 20140248702; 20140256046; 20140273037; 20140273226; 20140273230; 20140273231; 20140273232; 20140273233; 20140273234; 20140273235; 20140287938; 20140295556; 20140295557; 20140298547; 20140304853; 20140309487; 20140310828; 20140310830; 20140315985; 20140335063; 20140335620; 20140342456; 20140342457; 20140342458; 20140349400; 20140349405; 20140356867; 20140356956; 20140356958; 20140356959; 20140357523; 20140357530; 20140364333; 및 20140377868; 이들 모두는 본 명세서에 그들의 전문이 참고로 편입된다.

[0125] 일부 경우에, 이중성 폴리펩타이드(융합 상대)는 세포하 국소화를 제공하고, 즉, 이중성 폴리펩타이드는 핵에 대한 표적화를 위한 세포하 국소화 서열(예를 들어, 핵에 대한 국소화를 위한 핵 국소화 신호(NLS), 핵 밖에서 융합 단백질을 유지하기 위한 서열, 예를 들어, 핵 수송 서열(NES), 세포질에 보유된 융합 단백질을 유지하기 위한 서열, 미토콘드리아에 대한 표적화를 위한 미토콘드리아 국소화 신호, 엽록체, ER 체류 신호 등에 대한 표적화를 위한 엽록체 국소화 신호)을 함유한다. 일부 실시형태에서, CasY 융합 폴리펩타이드는 NLS를 포함하지 않으며, 따라서 단백질은 핵에 대해 표적화되지 않는다(이는, 예를 들어, 표적 핵산이 사이토졸에 존재하는 RNA 일 때 유리할 수 있다). 일부 실시형태에서, 이중성 폴리펩타이드는 수송 및/또는 정제의 용이함을 위한 태그(즉, 이중성 폴리펩타이드는 검출 가능한 표지임)(예를 들어, 형광 단백질, 예를 들어, 녹색 형광 단백질(GFP), YFP, RFP, CFP, mCherry, tdTomato 등; 히스티딘 태그, 예를 들어, 6XHis 태그; 혈구응집소(HA) 태그; FLAG 태그; Myc 태그; 등)을 제공할 수 있다.

[0126] 일부 경우에 CasY 단백질(예를 들어, 야생형 CasY 단백질, 변이체 CasY 단백질, 키메라 CasY 단백질, dCasY 단

백질, 키메라 CasY 단백질, 여기서 CasY 부분은 감소된 뉴클레아제 활성을 가진 - 예컨대 융합상대에 융합된 dCasY 단백질 등)은 핵 국소화 신호(NLS)(예를 들어, 일부 경우에 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)를 포함한다(이에 융합된다). 따라서, 일부 경우에, CasY 폴리펩타이드는 1개 이상의 NLS(예를 들어, 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)를 포함한다. 일부 경우에, 1개 이상의 NLS(2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)는 N-말단 및/또는 C-말단에 또는 근처에(예를 들어, 이의 50개 아미노산 내에) 위치된다. 일부 경우에, 1개 이상의 NLS(2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)는 N-말단에 또는 근처에(예를 들어, 이의 50개 아미노산 내에) 위치된다. 일부 경우에, 1개 이상의 NLS(2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)는 C-말단에 또는 근처에(예를 들어, 이의 50개 아미노산 내에) 위치된다. 일부 경우에, 1개 이상의 NLS(3개 이상, 4개 이상 또는 5개 이상의 NLS)는 N-말단과 C-말단 둘 다에 또는 근처에(예를 들어, 이의 50개 아미노산 내에) 위치된다. 일부 경우에, NLS는 N-말단에 위치되고, NLS는 C-말단에 위치된다.

[0127] 일부 경우에 CasY 단백질(예를 들어, 야생형 CasY 단백질, 변이체 CasY 단백질, 키메라 CasY 단백질, dCasY 단백질, 키메라 CasY 단백질, 여기서 CasY 부분은 감소된 뉴클레아제 활성 - 예컨대 융합 상대에 융합된 dCasY 단백질 등을 가짐)은 1 내지 10개의 NLS(예를 들어, 1-9, 1-8, 1-7, 1-6, 1-5, 2-10, 2-9, 2-8, 2-7, 2-6 또는 2-5개의 NLS)를 포함한다(이에 융합된다). 일부 경우에 CasY 단백질(예를 들어, 야생형 CasY 단백질, 변이체 CasY 단백질, 키메라 CasY 단백질, dCasY 단백질, 키메라 CasY 단백질, 여기서 CasY 부분은 감소된 뉴클레아제 활성 - 예컨대 융합 상대에 융합된 dCasY 단백질 등을 가짐)은 2 내지 5개의 NLS(예를 들어, 2-4 또는 2-3개의 NLS)를 포함한다(이에 융합된다).

[0128] NLS의 비제한적 예는 아미노산 서열 PKKKRKV(서열번호 96)를 갖는 SV40 바이러스 거대 T-항원의 NLS; 뉴클레오플라스민으로부터의 NLS(예를 들어, 서열 KRPAATKKAGQAKKKK(서열번호 97)을 갖는 뉴클레오플라스민 이분(bipartite) NLS)로부터 유래된 NLS 서열; 아미노산 서열 PAAKRVKLD(서열번호 98) 또는 RQRRELKRRSP(서열번호 99)를 갖는 c-myc NLS; 서열 NQSSNFGPMKGGNFGGRSSGPYGGGGYFAKPRNQGGY(서열번호 100)을 갖는 hRNP A1 M9 NLS; 임포틴-알파로부터의 IBB 도메인의 서열 RMRIZFKNGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV(서열번호 101); 근중 T 단백질의 서열 VSRKRPRP(서열번호 102) 및 PPKKARED(서열번호 103); 인간 p53의 서열 PQPKKKPL(서열번호 104); 마우스 c-abl IV의 서열 SALIKKKKKMAP(서열번호 105); 인플루엔자 바이러스 NS1의 서열 DRLRR(서열번호 106) 및 PKQKKRK(서열번호 107); 간염 바이러스 델타 항원의 서열 RKLKKIKKL(서열번호 108); 마우스 Mx1 단백질의 서열 REKKKFLKRR(서열번호 109); 인간 폴리(ADP-리보스) 중합효소의 서열 KRKGDEVDGVDEVAKKSKK(서열번호 110); 및 스테로이드 호르몬 수용체(인간) 글루코코르티코이드의 서열 RKCLQAGMNEARKTKK(서열번호 111)을 포함한다. 일반적으로, NLS(또는 다중 NLS)는 진핵 세포의 핵에서 검출 가능한 양으로 CasY 단백질의 축적을 유도하기에 충분한 강도를 가진다. 핵에서 축적의 검출은 임의의 적합한 기법에 의해 수행될 수 있다. 예를 들어, 검출 가능한 마커는 세포 내의 위치가 시각화될 수 있도록 CasY 단백질에 융합될 수 있다. 세포 핵은 또한 세포로부터 분리될 수 있고, 이어서, 이의 내용은 단백질을 검출하기 위한 임의의 적합한 방법, 예컨대 면역조직화학, 웨스턴 블롯, 또는 효소 활성 분석에 의해 분석될 수 있다. 핵에서의 축적은 또한 간접적으로 결정될 수 있다.

[0129] 일부 경우에, CasY 융합 폴리펩타이드는 "단백질 형질도입 도메인" 또는 PTD(또한 CPP - 세포 침투성 펩타이드로서 알려짐)를 포함하는데, 이는 지질 이중층, 마이셀, 세포막, 세포소기관 막 또는 소수포 막을 가로지르는 것을 용이하게 하는 폴리펩타이드, 폴리뉴클레오타이드, 탄수화물 또는 유기 또는 무기 화합물을 지칭한다. 거대 분자 및/또는 나노분자의 범위일 수 있는 다른 분자에 부착된 PTD는 분자가 막을 가로지르는 것, 예를 들어, 세포의 공간으로부터 세포내 공간 또는 사이토솔을 세포소기관 내로 이동하는 것을 용이하게 한다. 일부 실시형태에서, PTD는 아미노 말단 폴리펩타이드에 공유적으로 연결된다(예를 들어, 융합 단백질을 생성하기 위해 야생형 CasY에 연결되거나, 또는 융합 단백질을 생성하기 위해 변이체 CasY 단백질, 예컨대 dCasY, 틸대기효소 CasY, 또는 키메라 CasY 단백질에 연결됨). 일부 실시형태에서, PTD는 폴리펩타이드의 카복실 말단에 공유적으로 연결된다(예를 들어, 융합 단백질을 생성하기 위해 야생형 CasY에 연결되거나, 또는 융합 단백질을 생성하기 위해 변이체 CasY 단백질, 예컨대 dCasY, 틸대기효소 CasY, 또는 키메라 CasY 단백질에 연결됨). 일부 경우에, PTD는 적합한 삽입 부위에서 CasY 융합 폴리펩타이드에서 내부로 삽입된다(즉, CasY 융합 폴리펩타이드의 N- 또는 C-말단에서가 아님). 일부 경우에, 대상 CasY 융합 폴리펩타이드는 1개 이상의 PTD(예를 들어, 2개 이상, 3개 이상, 4개 이상의 PTD)를 포함한다(이에 접합되고, 융합된다). 일부 경우에, PTD는 핵 국소화 신호(NLS)(예를 들어, 일부 경우에 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)를 포함한다. 따라서, 일부 경우에, CasY 융합 폴리펩타이드는 1개 이상의 NLS(예를 들어, 2개 이상, 3개 이상, 4개 이상 또는 5개 이상의 NLS)를 포함한다. 일부 실시형태에서, PTD는 핵산(예를 들어, CasY 가이드 핵산, CasY 가이드 핵산을 암호화하는 폴리뉴클레오타이드, CasY 융합 폴리펩타이드를 암호화하는 폴리뉴클레오타이드, 공여자 폴리뉴클레오타이드 등)에

공유적으로 연결된다. PTD의 예는 최소 운테카펩타이드 단백질 형질도입 도메인(YGRKKRRQRRR; 서열번호 112를 포함하는 HIV-1 TAT의 잔기 47 내지 57에 대응); 세포 내로 유입을 지시하기에 충분한 다수의 알기닌을 포함하는 폴리알기닌 서열(예를 들어, 3, 4, 5, 6, 7, 8, 9, 10 또는 10 내지 50개의 알기닌); VP22 도메인(Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); 초파리 안테나페디아(Antennapedia) 단백질 형질도입 도메인(Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); 절단된 인간 칼시토닌 펩타이드(Trehin et al. (2004) *Pharm. Research* 21:1248-1256); 폴리라이신(Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR(서열번호 113); 트랜스포탄 GWTLSAGYLLGKINLKALAALAKKIL(서열번호 114); KALAEAKLAKALAKALAKHLAKALAKALAKCEA(서열번호 115); 및 RQIKIWQNRRMKWKK(서열번호 116)을 포함하지만, 이들로 제한되지 않는다. 예시적인 PTD는 YGRKKRRQRRR(서열번호 117), RKKRRQRRR(서열번호 118); 3개의 알기닌 잔기 내지 50개의 알기닌 잔기의 알기닌 동중중합체를 포함하지만, 이들로 제한되지 않고; 예시적인 PTD 도메인 아미노산 서열은 다음 중 어떤 것을 포함하지만, 이들로 제한되지 않는다: YGRKKRRQRRR(서열번호 119); RKKRRQRR(서열번호 120); YARAAARQARA(서열번호 121); THRLPRRRRRR(서열번호 122); 및 GRRRARRRRR(서열번호 123). 일부 실시형태에서, PTD는 활성화 가능한 CPP(ACPP)이다(Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPP는 매칭 절단 가능한 링커를 통해 다음이온(예를 들어, Glu9 또는 "E9")에 연결된 다양이온 CPP(예를 들어, Arg9 또는 "R9")를 포함하는데, 이는 순전하를 거의 0으로 감소시키고, 이에 의해 세포에 대한 접착 및 흡수를 저해한다. 링커의 절단 시, 다음이온이 방출되며, 다알기닌 및 그의 고유한 부착성을 국소로 가리며, 따라서 막을 가로지르도록 ACPP를 "활성화시킨다".

[0130] **링커(예를 들어, 융합 상대에 대해)**

[0131] 일부 실시형태에서, 대상 CasY 단백질은 링커 폴리펩타이드(예를 들어, 하나 이상의 링커 폴리펩타이드)를 통해 융합 상대에 융합될 수 있다. 링커 폴리펩타이드는 임의의 다양한 아미노산 서열을 가질 수 있다. 다른 화학적 결합이 제외되지는 않지만, 단백질은 일반적으로 가요성 특성의 스페이스 펩타이드에 의해 결합될 수 있다. 적합한 링커는 길이가 4개의 아미노산 내지 40개의 아미노산, 또는 길이가 4개의 아미노산 내지 25개의 아미노산인 폴리펩타이드를 포함한다. 이들 링커는 단백질을 결합하기 위해 합성, 링커-암호화 올리고뉴클레오타이드를 이용함으로써 생성될 수 있거나, 또는 융합 단백질을 암호화하는 핵산 서열에 의해 암호화될 수 있다. 일정 정도의 가요성이 있는 펩타이드 링커가 사용될 수 있다. 연결 펩타이드는 사실상 임의의 아미노산 서열을 가질 수 있고, 바람직한 링커는 일반적으로 가요성 펩타이드를 초대하는 서열을 가질 것임을 유념한다. 작은 아미노산, 예컨대 글리신 및 알라닌의 사용은 가요성 펩타이드를 생성하는 용도를 가진다. 이러한 서열의 생성은 당업자에게 일상적이다. 다양한 상이한 링커가 상업적으로 입수 가능하고, 사용에 적합한 것으로 고려된다.

[0132] 링커 폴리펩타이드의 예는 글리신 중합체(G)_n, 글리신-세린 중합체(예를 들어, (GS)_n, GSGGS_n(서열번호 124), GGS_n(서열번호 125) 및 GGS_n(서열번호 126)을 포함, 여기서 n은 적어도 하나의 정수임), 글리신-알라닌 중합체, 알라닌-세린 중합체를 포함한다. 예시적인 링커는 GSGG(서열번호 127), GSGG(서열번호 128), GSGG(서열번호 129), GSGG(서열번호 130), GSGG(서열번호 131), GSSG(서열번호 132) 등을 포함하지만, 이들로 제한되지 않는 아미노산 서열을 포함할 수 있다. 당업자는 링커가 가요성 링커뿐만 아니라 덜 가요성인 구조를 부여하는 하나 이상의 부분을 포함할 수 있도록 임의의 목적으로 하는 요소에 접합된 펩타이드의 설계가 모두 또는 부분적으로 가요성인 링커를 포함할 수 있다는 것을 인식할 것이다.

[0133] **검출 가능한 표지**

[0134] 일부 경우에, 본 개시내용의 CasY 폴리펩타이드는 검출 가능한 표지를 포함한다. 적합한 검출 가능한 표지 및/또는 검출 가능한 신호를 제공할 수 있는 모이어티는 효소, 방사성 동위 원소, 특정 결합쌍의 구성원; 형광단; 형광 단백질; 양자점 등을 포함할 수 있지만, 이들로 제한되지 않는다.

[0135] 적합한 형광 단백질은 녹색 형광 단백질(GFP) 또는 이의 변이체, GFP의 청색 형광 변이체(BFP), GFP의 사이안 형광 변이체(CFP), GFP의 황색 형광 변이체(YFP), 증강된 GFP(EGFP), 증강된 CFP(ECFP), 증강된 YFP(EYFP), GFPs65T, 에메랄드, 토파즈(TYFP), 비너스, 시트린, 엠시트린(mCitrine), GFPuv, 불안정 EGFP(dEGFP), 불안정 ECFP(dECFP), 불안정 EYFP (dEYFP), mCFPm, 세룰리언, T-사파이어, CyPet, YPet, mKO, HcRed, t-HcRed, DsRed, DsRed2, DsRed-단량체, J-Red, 이량체2, t-이량체2(12), mRFP1, 포실로포린, 레닐라 GFP, 몬스터 GFP, paGFP, Kaede 단백질 및 킨들링 단백질, 피코빌리단백질 및 피코빌리단백질 접합체(B-피코에리트린, R-피코에리트린 및 알로피코사이아닌을 포함)를 포함하지만, 이들로 제한되지 않는다. 형광 단백질의 다른 예는 mHoneydew, mBanana, mOrange, dTomato, tdTomato, mTangerine, mStrawberry, mCherry, mGrape1, mRaspberry, mGrape2, mPlum(Shaner et al. (2005) *Nat. Methods* 2:905-909) 등을 포함한다. 예를 들어, 문헌[Matz et al. (1999)

Nature Biotechnol. 17:969-973]에 기재된 바와 같은 화충류(Anthozoan) 종으로부터의 임의의 다양한 형광 및 착색 단백질이 사용에 적합하다.

[0136] 적합한 효소는 겨자무과산화효소(HRP), 알칼리성 포스파타제(AP), 베타-갈락토시다제(GAL), 글루코스-6-인산 탈수소효소, 베타-N-아세틸글루코사미니다제, β -글루쿠로니다제, 인버타제, 잔틴 옥시다제, 반딧불이 루시페라제, 글루코스 옥시다제(GO) 등을 포함하지만, 이들로 제한되지 않는다.

[0137] **프로토스페이서 인접 모티프(PAM)**

[0138] CasY 단백질은 DNA-표적화 RNA와 표적 DNA 사이의 상보성 영역에 의해 정해지는 표적 서열에서 표적 DNA에 결합한다. 다수의 CRISPR 엔도뉴클레아제에 대한 경우에서와 같이, 이중 가닥 표적 DNA의 부위-특이적 결합(및/또는 절단)은 (i) 가이드 RNA와 표적 DNA 사이의 염기쌍 상보성; 그리고 (ii) 표적 DNA 내 짧은 모티프[프로토스페이서 인접 모티프(PAM)로서 지칭됨]에 의해 결정된 위치에서 생긴다.

[0139] 일부 실시형태에서, CasY 단백질에 대한 PAM은 표적 DNA의 비상보성 가닥의 표적 서열의 5' 바로 옆이다(상보성 가닥은 가이드 RNA의 가이드 서열에 혼성화되는 한편, 비-상보성 가닥은 가이드 RNA에 직접 혼성화하지 않으며, 비-상보성 가닥의 역상보체이다). 일부 실시형태에서(예를 들어, 본 명세서에 기재된 바와 같은 CasY1이 사용될 때), 비-상보성 가닥의 PAM 서열은 5'-TA-3'이다(그리고 일부 경우에 XTA, 여기서 X는 C, A 또는 T임). 예로서, 도 5 및 도 7(이때 X가 C, A 또는 T인 경우 PAM이 XTA인 것으로 고려한다면 PAM은 TA 또는 CTA임) 참조. 일부 실시형태에서(예를 들어, 본 명세서에 기재된 바와 같은 CasY1이 사용될 때), 비-상보성 가닥의 PAM 서열은 5'-TA-3'이다(그리고 일부 경우에 HTA, 여기서 H는 C, A 또는 T임). 예로서, 도 5 및 도 7(이때 H가 C, A 또는 T인 경우 PAM이 HTA인 것으로 고려한다면 PAM은 TA 또는 CTA임) 참조. 일부 경우에(예를 들어, 본 명세서에 기재된 바와 같은 CasY2가 사용될 때), 비-상보성 가닥의 PAM 서열은 표적의 5'-YR-3' 측접 서열 5'이다(여기서, Y는 T 또는 C이고, R은 A 또는 G이다). 일부 경우에(예를 들어, 본 명세서에 기재된 바와 같은 CasY2가 사용될 때), 비-상보성 가닥의 PAM 서열은 5'-TR-3'(예를 들어, 5'-DTR-3')이다(여기서, R은 A 또는 G이고, D는 A, G 또는 T임). 예로서, 도 5D 참조.

[0140] 일부 경우에, 상이한 CasY 단백질(즉, 다양한 종으로부터의 CasY 단백질)은 상이한 CasY 단백질의 다양한 효소 특징을 활용하기 위해(예를 들어, 상이한 PAM 서열 선호도에 대해; 증가된 또는 감소된 효소 활성에 대해; 세포 독성의 증가된 또는 감소된 수준에 대해; NHEJ, 상동성-직접 수선, 단일 가닥 파손, 이중 가닥 파손 등 사이의 균형 변화를 위해; 짧은 총 서열의 이점을 취하기 위해; 등) 다양한 제공된 방법으로 사용하는 것이 유리할 수 있다. 상이한 종으로부터의 CasY 단백질은 표적 DNA에서 상이한 PAM 서열을 필요로 할 수 있다. 따라서, 특정 선택 CasY 단백질에 대해, PAM 서열 필요는 상기 기재한 5'-TA-3'(또는 XTA, HTA) 서열과 상이할 수 있다. 적절한 PAM 서열의 동정을 위한 다양한 방법(인실리코 및/또는 습식 실험실 방법을 포함)은 당업계에 공지되어 있고, 일상적이며, 임의의 편리한 방법이 사용될 수 있다. 본 명세서에 기재된 TA(XTA, HTA) PAM 서열은 PAM 고갈 분석(예를 들어, 이하의 작업 실시예의 도 5 참조)을 이용하여 동정되었다.

[0141] **CasY 가이드 RNA**

[0142] 리보뉴클레오타이드 복합체(RNP)를 형성하고, 표적 핵산(예를 들어, 표적 DNA) 내의 특정 위치에 대해 복합체를 표적화하는 CasY 단백질에 결합하는 핵산은 본 명세서에서 "CasY 가이드 RNA"로서 또는 단순히 "가이드 RNA"로서 지칭된다. 일부 경우에, CasY 가이드 RNA가 RNA 염기에 추가로 DNA 염기를 포함하도록 혼성 DNA/RNA가 만들어질 수 있지만, 용어 "CasY 가이드 RNA"는 여전히 본 명세서의 이러한 분자를 포함하기 위해 사용된다는 것이 이해되어야 한다.

[0143] CasY 가이드 RNA는 2개의 세그먼트, 표적화 세그먼트 및 단백질-결합 세그먼트를 포함하는 것으로 언급될 수 있다. CasY 가이드 RNA의 표적화 세그먼트는 표적 핵산(예를 들어, 표적 ssRNA, 표적 ssDNA, 이중 가닥 표적 DNA의 상보성 가닥 등) 내의 특정 서열(표적 부위)에 상보성인(그리고 따라서 이에 혼성화하는) 뉴클레오타이드 서열(가이드 서열)을 포함한다. 단백질-결합 세그먼트(또는 "단백질-결합 서열")는 CasY 폴리펩타이드에 상호작용한다(이에 결합한다). 대상 CasY 가이드 RNA의 단백질의 단백질-결합 세그먼트는 이중 가닥 RNA 듀플렉스(dsRNA 듀플렉스)를 형성하기 위해 서로 혼성화하는 뉴클레오타이드의 2개의 상보적 신장부를 포함한다. 표적 핵산(예를 들어, 게놈 DNA)의 부위-특이적 결합 및/또는 절단은 CasY 가이드 RNA(CasY 가이드 RNA의 가이드 서열)와 표적 핵산 사이의 염기쌍 상보성에 의해 결정되는 위치(예를 들어, 표적 좌위의 표적 서열)에서 일어날 수 있다.

[0144] CasY 가이드 RNA 및 CasY 단백질, 예를 들어, 융합 CasY 폴리펩타이드는 복합체를 형성한다(예를 들어, 비공유 상호작용을 통해 결합한다). CasY 가이드 RNA는 가이드 서열(표적 핵산의 서열에 상보성인 뉴클레오타이드

서열)을 포함하는 표적화 세그먼트를 포함함으로써 복합체에 표적 특이성을 제공한다. 복합체의 CasY 단백질은 부위-특이적 활성(예를 들어, CasY 단백질에 의해 제공된 절단 활성 및/또는 키메라 CasY 단백질의 경우에 융합 상대에 의해 제공된 활성)을 제공한다. 다시 말해서, CasY 단백질은 CasY 가이드 RNA와 그의 회합 때문에 표적 핵산 서열(예를 들어, 표적 서열)로 가이드된다.

[0145] CasY 가이드 RNA의 "표적화 서열"로서도 지칭되는 "가이드 서열"은 PAM 서열을 고려할 수 있다는 것을 제외하고 (예를 들어, 본 명세서에 기재된 바와 같이), CasY 가이드 RNA가 임의의 목적으로 하는 표적 핵산의 임의의 목적으로 하는 서열에 대해 CasY 단백질(예를 들어, 천연 유래 CasY 단백질, 융합 CasY 폴리펩타이드(키메라 CasY) 등)를 표적화할 수 있도록 변형될 수 있다. 따라서, 예를 들어, CasY 가이드 RNA는 진핵 세포 내 핵산, 예를 들어, 바이러스 핵산, 진핵 핵산(예를 들어, 진핵 염색체, 염색체 서열, 진핵 RNA 등) 등에서의 서열에 대해 상보성인(예를 들어, 혼성화할 수 있는) 가이드 서열을 가질 수 있다.

[0146] **CasY 가이드 RNA의 가이드 서열**

[0147] 대상 CasY 가이드 RNA는 표적 핵산 내 서열(표적 부위)에 대해 상보성은 뉴클레오타이드 서열인 가이드 서열(즉, 표적화 서열)을 포함한다. 다시 말해서, CasY 가이드 RNA의 가이드 서열은 혼성화(즉, 염기쌍)를 통해 서열 특이적 방식으로 표적 핵산(예를 들어, 이중 가닥 DNA(dsDNA), 단일 가닥 DNA(ssDNA), 단일 가닥 RNA(ssRNA) 또는 이중 가닥 RNA(dsRNA))과 상호작용할 수 있다. CasY 가이드 RNA의 가이드 서열은 표적 핵산(예를 들어, 진핵 표적 핵산, 예컨대 게놈 DNA) 내에서 임의의 목적으로 하는 표적 서열(예를 들어, PAM을 고려하는 동안, 예를 들어, dsDNA 표적을 표적화할 때)에 혼성화하도록 변형(예를 들어, 유전자 조작에 의해)/설계될 수 있다.

[0148] 일부 실시형태에서, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 60% 이상(예를 들어, 65% 이상, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 80% 이상(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 90% 이상(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 100% 이상이다.

[0149] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 표적 핵산의 7개의 인접한 가장 3'의 뉴클레오타이드에 대해 100%이다.

[0150] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17개 이상의(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 60% 이상(예를 들어, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17개 이상의(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 80% 이상(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17개 이상의(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 90% 이상(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17개 이상의(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상, 22개 이상)의 인접한 뉴클레오타이드에 대해 100%이다.

[0151] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19개 이상의(예를 들어, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 60% 이상(예를 들어, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19개 이상의(예를 들어, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 80% 이상(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19개 이상의(예를 들어, 20개 이상, 21개 이상, 22개 이상) 인접한 뉴클레오타이드에 대해 90% 이상(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19개 이상의(예를 들어, 20개 이상, 21개 이상, 22개 이상)의 인접한 뉴클레오타이드에 대해 100%이다.

[0152] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17 내지 25개의 인접한 뉴클레오타

이드에 대해 60% 이상(예를 들어, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17 내지 25개의 인접한 뉴클레오타이드에 대해 80% 이상(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17 내지 25개의 인접한 뉴클레오타이드에 대해 90% 이상(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 17 내지 25개의 인접한 뉴클레오타이드에 대해 100% 이상이다.

[0153] 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19 내지 25개의 인접한 뉴클레오타이드에 대해 60% 이상(예를 들어, 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19 내지 25개의 인접한 뉴클레오타이드에 대해 80% 이상(예를 들어, 85% 이상, 90% 이상, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19 내지 25개의 인접한 뉴클레오타이드에 대해 90% 이상(예를 들어, 95% 이상, 97% 이상, 98% 이상, 99% 이상 또는 100%)이다. 일부 경우에, 표적 핵산의 가이드 서열과 표적 부위 사이의 상보성 백분율은 19 내지 25개의 인접한 뉴클레오타이드에 대해 100% 이상이다.

[0154] 일부 경우에, 가이드 서열은 길이가 17 내지 30개의 뉴클레오타이드(nt)(예를 들어, 17-25, 17-22, 17-20, 19-30, 19-25, 19-22, 19-20, 20-30, 20-25 또는 20-22 nt) 범위이다. 일부 경우에, 가이드 서열은 길이가 17 내지 25개의 뉴클레오타이드(nt)(예를 들어, 17-22, 17-20, 19-25, 19-22, 19-20, 20-25 또는 20-22 nt) 범위이다. 일부 경우에, 가이드 서열은 길이가 17개 이상의 nt(예를 들어, 18개 이상, 19개 이상, 20개 이상, 21개 이상 또는 22개 이상의 nt; 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt 등)를 가진다. 일부 경우에, 가이드 서열은 길이가 19개 이상의 nt(예를 들어, 20개 이상, 21개 이상 또는 22개 이상의 nt; 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt 등)를 가진다. 일부 경우에, 가이드 서열은 길이가 17 nt이다. 일부 경우에, 가이드 서열은 길이가 18 nt이다. 일부 경우에, 가이드 서열은 길이가 19 nt이다. 일부 경우에, 가이드 서열은 길이가 20 nt이다. 일부 경우에, 가이드 서열은 길이가 21 nt이다. 일부 경우에, 가이드 서열은 길이가 22 nt이다. 일부 경우에, 가이드 서열은 길이가 23 nt이다.

[0155] **CasY 가이드 RNA의 단백질-결합 세그먼트**

[0156] 대상 CasY 가이드 RNA의 단백질-결합 세그먼트는 CasY 단백질과 상호작용한다. CasY 가이드 RNA는 결합된 CasY 단백질을 상기 언급한 가이드 서열을 통해 표적 핵산 내의 특정 뉴클레오타이드 서열로 가이드한다. CasY 가이드 RNA의 단백질-결합 세그먼트는 서로에 대해 상보성이고 이중 가닥 RNA 듀플렉스(dsRNA 듀플렉스)를 형성하도록 혼성화되는 뉴클레오타이드의 2개의 신장부를 포함한다. 따라서, 단백질-결합 세그먼트는 dsRNA 듀플렉스를 포함한다.

[0157] 일부 경우에, dsRNA 듀플렉스 영역은 5-25개의 염기쌍(bp)(예를 들어, 5-22, 5-20, 5-18, 5-15, 5-12, 5-10, 5-8, 8-25, 8-22, 8-18, 8-15, 8-12, 12-25, 12-22, 12-18, 12-15, 13-25, 13-22, 13-18, 13-15, 14-25, 14-22, 14-18, 14-15, 15-25, 15-22, 15-18, 17-25, 17-22 또는 17-18 bp, 예를 들어, 5 bp, 6 bp, 7 bp, 8 bp, 9 bp, 10 bp 등)의 범위를 포함한다. 일부 경우에, dsRNA 듀플렉스 영역은 6 내지 15개의 염기쌍(bp)(예를 들어, 6-12, 6-10 또는 6-8 bp, 예를 들어, 6 bp, 7 bp, 8 bp, 9 bp, 10 bp 등)의 범위를 포함한다. 일부 경우에, 듀플렉스 영역은 5개 이상의 bp(예를 들어, 6개 이상, 7개 이상 또는 8개 이상의 bp)를 포함한다. 일부 경우에, 듀플렉스 영역은 6개 이상의 bp(예를 들어, 7개 이상 또는 8개 이상의 bp)를 포함한다. 일부 경우에, 듀플렉스 영역의 모든 뉴클레오타이드가 짝지어지지 않으며, 따라서 듀플렉스 형성 영역은 벌지(bulge)를 포함할 수 있다. 본 명세서의 용어 "벌지"는 이중 가닥 듀플렉스에 기여하지 않지만, 기여하는 뉴클레오타이드에 의해 5' 및 3'을 둘러싸는 뉴클레오타이드의 신장부(하나의 뉴클레오타이드일 수 있음)를 의미하기 위해 사용되며, 이러한 벌지는 듀플렉스 영역의 부분으로 고려된다. 일부 경우에, dsRNA는 1개 이상의 벌지(예를 들어, 2개 이상, 3개 이상, 4개 이상의 벌지)를 포함한다. 일부 경우에, dsRNA 듀플렉스는 2개 이상의 벌지(예를 들어, 3개 이상, 4개 이상의 벌지)를 포함한다. 일부 경우에, dsRNA 듀플렉스는 1 내지 5개의 벌지(예를 들어, 1-4, 1-3, 2-5, 2-4 또는 2-3개의 벌지)를 포함한다.

[0158] 따라서, 일부 경우에, dsRNA 듀플렉스를 형성하기 위해 서로 혼성화되는 뉴클레오타이드의 신장부는 서로 70% 내지 100% 상보성(예를 들어, 75%-100%, 80%-100%, 85%-100%, 90%-100%, 95%-100% 상보성)을 가진다. 일부 경우에, dsRNA 듀플렉스를 형성하기 위해 서로 혼성화되는 뉴클레오타이드의 신장부는 서로 70% 내지 100% 상보성

(예를 들어, 75%-100%, 80%-10%, 85%-100%, 90%-100%, 95%-100% 상보성)을 가진다. 일부 경우에, dsRNA 듀플렉스를 형성하기 위해 서로 혼성화되는 뉴클레오타이드의 신장부는 서로 85% 내지 100% 상보성(예를 들어, 90%-100%, 95%-100% 상보성)을 가진다. 일부 경우에, dsRNA 듀플렉스를 형성하기 위해 서로 혼성화되는 뉴클레오타이드의 신장부는 서로 70% 내지 95% 상보성(예를 들어, 75%-95%, 80%-95%, 85%-95%, 90%-95% 상보성)을 가진다.

[0159] 다시 말해서, 일부 실시형태에서, dsRNA 듀플렉스는 서로 70%-100% 상보성(예를 들어, 75%-100%, 80%-10%, 85%-100%, 90%-100%, 95%-100% 상보성)을 갖는 뉴클레오타이드의 2개 신장부를 포함한다. 일부 경우에, dsRNA 듀플렉스는 서로 85%-100% 상보성(예를 들어, 90%-100%, 95%-100% 상보성)을 갖는 뉴클레오타이드의 2개 신장부를 포함한다. 일부 경우에, dsRNA 듀플렉스는 서로 70%-95% 상보성(예를 들어, 75%-95%, 80%-95%, 85%-95%, 90%-95% 상보성)을 갖는 뉴클레오타이드의 2개 신장부를 포함한다.

[0160] 대상 CasY 가이드 RNA의 이중가닥 영역은 천연 유래 이중가닥 영역에 대해 1개 이상의 (1, 2, 3, 4, 5개 등) 돌연변이를 포함할 수 있다. 예를 들어, 일부 경우에 염기쌍은 유지될 수 있는 한편 각각의 세그먼트로부터의 염기쌍에 기여하는 뉴클레오타이드는 상이할 수 있다. 일부 경우에, 대상 CasY 가이드 RNA의 이중가닥 영역은 (천연 유래 CasY 가이드 RNA의) 천연 유래 이중가닥 영역에 비교할 때 더 짝지어진 염기, 덜 짝지어진 염기, 더 작은 벌지, 더 큰 벌지, 소수의 벌지, 더 많은 벌지, 또는 이들의 임의의 편리한 조합을 포함한다.

[0161] 다양한 Cas9 가이드 RNA의 예는 당업계에서 찾을 수 있고, 그리고 일부 경우에 Cas9 가이드 RNA 내로 도입된 것과 유사한 변형이 또한 본 개시내용의 CasY 가이드 RNA 내로 도입될 수 있다(예를 들어, dsRNA 듀플렉스 영역에 대한 돌연변이, 다른 단백질과의 상호작용을 제공하기 위해 추가된 안정성을 위한 5' 또는 3' 단부의 연장 등). 예를 들어, 문헌[Jinek et al., Science. 2012 Aug 17;337(6096):816-21; Chylinski et al., RNA Biol. 2013 May;10(5):726-37; Ma et al., Biomed Res Int. 2013;2013:270805; Hou et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9; Jinek et al., Elife. 2013;2:e00471; Pattanayak et al., Nat Biotechnol. 2013 Sep;31(9):839-43; Qi et al., Cell. 2013 Feb 28;152(5):1173-83; Wang et al., Cell. 2013 May 9;153(4):910-8; Auer et al., Genome Res. 2013 Oct 31; Chen et al., Nucleic Acids Res. 2013 Nov 1;41(20):e19; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; Cho et al., Genetics. 2013 Nov;195(3):1177-80; DiCarlo et al., Nucleic Acids Res. 2013 Apr;41(7):4336-43; Dickinson et al., Nat Methods. 2013 Oct;10(10):1028-34; Ebina et al., Sci Rep. 2013;3:2510; Fujii et al., Nucleic Acids Res. 2013 Nov 1;41(20):e187; Hu et al., Cell Res. 2013 Nov;23(11):1322-5; Jiang et al., Nucleic Acids Res. 2013 Nov 1;41(20):e188; Larson et al., Nat Protoc. 2013 Nov;8(11):2180-96; Mali et al., Nat Methods. 2013 Oct;10(10):957-63; Nakayama et al., Genesis. 2013 Dec;51(12):835-43; Ran et al., Nat Protoc. 2013 Nov;8(11):2281-308; Ran et al., Cell. 2013 Sep 12;154(6):1380-9; Upadhyay et al., G3 (Bethesda). 2013 Dec 9;3(12):2233-8; Walsh et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15514-5; Xie et al., Mol Plant. 2013 Oct 9; Yang et al., Cell. 2013 Sep 12;154(6):1370-9; Briner et al., Mol Cell. 2014 Oct 23;56(2):333-9]; 및 미국 특허 및 특허 출원: 8,906,616; 8,895,308; 8,889,418; 8,889,356; 8,871,445; 8,865,406; 8,795,965; 8,771,945; 8,697,359; 20140068797; 20140170753; 20140179006; 20140179770; 20140186843; 20140186919; 20140186958; 20140189896; 20140227787; 20140234972; 20140242664; 20140242699; 20140242700; 20140242702; 20140248702; 20140256046; 20140273037; 20140273226; 20140273230; 20140273231; 20140273232; 20140273233; 20140273234; 20140273235; 20140287938; 20140295556; 20140295557; 20140298547; 20140304853; 20140309487; 20140310828; 20140310830; 20140315985; 20140335063; 20140335620; 20140342456; 20140342457; 20140342458; 20140349400; 20140349405; 20140356867; 20140356956; 20140356958; 20140356959; 20140357523; 20140357530; 20140364333; 및 20140377868을 참조하며; 이들 모두는 본 명세서에 그들의 전문이 참고로 편입된다.

[0162] CasY 가이드 RNA는 dsRNA 듀플렉스를 형성하기 위해 단백질-결합 세그먼트에 혼성화하는 뉴클레오타이드의 가이드 서열과 2개의 신장부("듀플렉스-형성 세그먼트")를 둘 다 포함한다. 주어진 CasY 가이드 RNA의 특정 서열은 crRNA가 발견된 종의 특징일 수 있다. 적합한 CasY 가이드 RNA의 예는 본 명세서에 제공된다.

[0163] **가이드 RNA 서열의 예**

[0164] 도 6(패널 a 및 b)에 도시된 반복부 서열(예시적 CasY 가이드 RNA의 비-가이드 서열 부분)은 CasY1-Y5에 대한 천연 좌위로부터 유래된다. 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) crRNA 서

열 CTCCGAAAGTATCGGGGATAAAGGC(서열번호 31)를 포함한다[RNA는 CUCCGAAAGUAUCGGGGAUAAAGGC(서열번호 11)임](예를 들어, 도 6 참조). 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CTCCGAAAGTATCGGGGATAAAGGC(서열번호 31)와 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CUCCGAAAGUAUCGGGGAUAAAGGC(서열번호 11)이다]. 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CTCCGAAAGTATCGGGGATAAAGGC(서열번호 31)와 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CUCCGAAAGUAUCGGGGAUAAAGGC(서열번호 11)이다].

[0165] 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) crRNA 서열 CACCGAAATTTGGAGAGGATAAGGC(서열번호 32)를 포함한다[RNA는 CACCGAAAUUUGGAGAGGAUAAAGGC(서열번호 12)임](예를 들어, 도 6 참조). 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CACCGAAATTTGGAGAGGATAAGGC(서열번호 32)와 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CACCGAAAUUUGGAGAGGAUAAAGGC(서열번호 12)이다]. 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CACCGAAATTTGGAGAGGATAAGGC(서열번호 32)와 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CACCGAAAUUUGGAGAGGAUAAAGGC(서열번호 12)이다].

[0166] 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) crRNA 서열 CTCCGAATTATCGGGAGGATAAGGC(서열번호 33)를 포함한다[RNA는 CUCCGAAUUAUCGGGAGGAUAAAGGC(서열번호 13)임](예를 들어, 도 6 참조). 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CTCCGAATTATCGGGAGGATAAGGC(서열번호 33)와 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CUCCGAAUUAUCGGGAGGAUAAAGGC(서열번호 13)이다]. 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CTCCGAATTATCGGGAGGATAAGGC(서열번호 33)와 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CUCCGAAUUAUCGGGAGGAUAAAGGC(서열번호 13)이다].

[0167] 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) crRNA 서열 CCCCGAATATAGGGGACAAAAAGGC(서열번호 34)를 포함한다[RNA는 CCCCGAAUUAUCGGGACAAAAAGGC(서열번호 14)임](예를 들어, 도 6 참조). 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CCCCGAATATAGGGGACAAAAAGGC(서열번호 34)와 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CCCCGAAUUAUCGGGACAAAAAGGC(서열번호 14)이다]. 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CCCCGAATATAGGGGACAAAAAGGC(서열번호 34)와 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CCCCGAAUUAUCGGGACAAAAAGGC(서열번호 14)이다].

[0168] 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) crRNA 서열 GTCTAGACATACAGGTGGAAGGTGAGAGTAAAGAC(서열번호 35)를 포함한다[RNA는 GUCUAGACAUACAGGUGGAAAGGTGAGAGTAAAGAC(서열번호 15)임](예를 들어, 도 6 참조). 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 GTCTAGACATACAGGTGGAAGGTGAGAGTAAAGAC(서열번호 35)와 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 GUCUAGACAUACAGGUGGAAAGGTGAGAGTAAAGAC(서열번호 15)이다]. 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 GTCTAGACATACAGGTGGAAGGTGAGAGTAAAGAC(서열번호 35)와 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 GUCUAGACAUACAGGUGGAAAGGTGAGAGTAAAGAC(서열번호 15)이다].

[0169] 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) 서열번호 11 내지 15 중 임의의 하나에 제시된 crRNA 서열을 포함한다. 일부 경우에, 대상 CasY 가이드 RNA는 서열번호 11 내지 15 중 임의의 하나에 제시된 crRNA 서열 crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasY 가이드 RNA는 서열번호 11 내지 15 중 임의의 하나에 제시된 crRNA 서열 crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0170] 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) 서열번호 11 내지 14 중 임의의 하나에 제시된 crRNA 서열을 포함한다. 일부 경우에, 대상 CasY 가이드 RNA는 서열번호 11 내지 14 중 임의의 하나

에 제시된 crRNA 서열 crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasY 가이드 RNA는 서열번호 11 내지 14 중 임의의 하나에 제시된 crRNA 서열 crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0171] CasY18에 대한 천연 좌위로부터의 반복부 서열(예시적 CasY 가이드 RNA의 비-가이드 서열 부분)은 CTCCGTGAATACGTGGGGTAAAGGC(서열번호 36)이다[RNA는 CUCCGUGAAUACGUGGGGUAAGGC(서열번호 16)이다]. 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) crRNA 서열 CTCCGTGAATACGTGGGGTAAAGGC(서열번호 36)를 포함한다[RNA는 CUCCGUGAAUACGUGGGGUAAGGC(서열번호 16)임]. 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CTCCGTGAATACGTGGGGTAAAGGC(서열번호 36)와 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CUCCGUGAAUACGUGGGGUAAGGC(서열번호 16)이다]. 일부 경우에, 대상 CasY 가이드 RNA는 crRNA 서열 CTCCGTGAATACGTGGGGTAAAGGC(서열번호 36)과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다[RNA는 CUCCGUGAAUACGUGGGGUAAGGC(서열번호 16)이다].

[0172] 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) 서열번호 11 내지 16 중 임의의 하나에 제시된 crRNA 서열을 포함한다. 일부 경우에, 대상 CasY 가이드 RNA는 (예를 들어, 가이드 서열에 추가로) 서열번호 11 내지 16 중 임의의 하나에 제시된 crRNA 서열 crRNA 서열과 80% 이상의 동일성(예를 들어, 85% 이상, 90% 이상, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 대상 CasY 가이드 RNA는 서열번호 11 내지 16 중 임의의 하나에 제시된 crRNA 서열 crRNA 서열과 90% 이상의 동일성(예를 들어, 93% 이상, 95% 이상, 97% 이상, 98% 이상, 또는 100%의 동일성)을 갖는 뉴클레오타이드 서열을 포함한다.

[0173] CasY 시스템

[0174] 본 개시내용은 CasY 시스템을 개시한다. 본 개시내용의 CasY 시스템은 a) 본 개시내용의 CasY 폴리펩타이드 및 CasY 가이드 RNA; b) 본 개시내용의 CasY 폴리펩타이드, CasY 가이드 RNA, 및 공여자 주형 핵산; c) 본 개시내용의 CasY 융합 폴리펩타이드 및 CasY 가이드 RNA; d) 본 개시내용의 CasY 융합 폴리펩타이드, CasY 가이드 RNA, 및 공여자 주형 핵산; e) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 mRNA; 및 CasY 가이드 RNA; f) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 공여자 주형 핵산; g) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 mRNA; 및 CasY 가이드 RNA; h) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 공여자 주형 핵산; i) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; j) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 주형 핵산을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; k) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; l) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 주형 핵산을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; m) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; n) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; 및 공여자 주형 핵산; o) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; p) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; 및 공여자 주형 핵산; q) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, 제1 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 제2 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; 또는 r) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, 제1 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 제2 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; 또는 (a) 내지 (r) 중 하나의 일부 변형을 포함할 수 있다.

[0175] **핵산**

[0176] 본 개시내용은 공여자 폴리뉴클레오타이드 서열, CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열(예를 들어, 야생형 CasY 단백질, 틸내기효소 CasY 단백질, dCasY 단백질, 키메라 CasY 단백질 등), CasY 가이드 RNA, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열 중 하나 이상을 포함하는 하나 이상의 핵산을 제공한다. 본 개시내용은 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 제공한다. 본 개시내용은 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터를 제공한다. 본 개시내용은 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터를 제공한다. 본 개시내용은 a) CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; 및 b) CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열(들)을 포함하는, 재조합 발현 벡터를 제공한다. 본 개시내용은 a) CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; 및 b) CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열(들)을 포함하는, 재조합 발현 벡터를 제공한다. 일부 경우에, CasY 단백질을 암호화하는 뉴클레오타이드 서열 및/또는 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열은 선택 세포 유형(예를 들어, 원핵 세포, 진핵 세포, 식물 세포, 동물 세포, 포유류 세포, 영장류 세포, 설치류 세포, 인간 세포 등)에서 작동 가능한 프로모터에 작동 가능하게 연결된다.

[0177] 일부 경우에, 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 코돈 최적화된다. 이런 최적화 유형은 의도된 숙주 유기체 또는 세포의 코돈 선호도를 모방하는 한편, 동일한 단백질을 암호화하는 CasY-암호화 뉴클레오타이드 서열의 돌연변이를 수반한다. 따라서, 코돈은 변할 수 있지만, 암호화된 단백질은 변하지 않고 남아있다. 예를 들어, 의도된 표적 세포가 인간 세포라면, 인간 코돈 최적화된 CasY-암호화 뉴클레오타이드 서열이 사용될 수 있었다. 다른 비제한적 예로서, 의도된 숙주 세포가 마우스 세포였다면, 마우스 코돈-최적화된 CasY-암호화 뉴클레오타이드 서열이 생성될 수 있었다. 다른 비제한적 예로서, 의도된 숙주 세포가 식물 세포였다면, 식물 코돈-최적화된 CasY-암호화 뉴클레오타이드 서열이 생성될 수 있었다. 다른 비제한적 예로서, 의도된 숙주 세포가 곤충 세포였다면, 곤충 코돈-최적화된 CasY-암호화 뉴클레오타이드 서열이 생성될 수 있었다.

[0178] 본 개시내용은 (일부 경우에 상이한 재조합 발현 벡터에서, 그리고 일부 경우에 동일한 재조합 발현 벡터에서): (i) 공여자 주형 핵산의 뉴클레오타이드 서열(공여자 주형이 표적 핵산(예를 들어, 표적 게놈)의 표적 서열에 대해 상동성을 갖는 뉴클레오타이드 서열을 포함하는 경우); (ii) 표적화된 게놈의 표적 좌위의 표적 서열에 혼성화되는 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열(예를 들어, 표적 세포, 예컨대 진핵 세포에서 작동 가능한 프로모터에 작동 가능하게 연결됨); 및 (iii) CasY 단백질을 암호화하는 뉴클레오타이드 서열(예를 들어, 표적 세포, 예컨대 진핵 세포에서 작동 가능한 프로모터에 작동 가능하게 연결됨)을 포함하는 하나 이상의 재조합 발현 벡터를 제공한다. 본 개시내용은 (일부 경우에 상이한 재조합 발현 벡터에서, 그리고 일부 경우에 동일한 재조합 발현 벡터에서): (i) 공여자 주형 핵산의 뉴클레오타이드 서열(공여자 주형이 표적 핵산(예를 들어, 표적 게놈)의 표적 서열에 대해 상동성을 갖는 뉴클레오타이드 서열을 포함하는 경우); 및 (ii) 표적화된 게놈의 표적 좌위의 표적 서열에 혼성화되는 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열(예를 들어, 표적 세포, 예컨대 진핵 세포에서 작동 가능한 프로모터에 작동 가능하게 연결됨)을 포함하는 하나 이상의 재조합 발현 벡터를 제공한다. 본 개시내용은 (일부 경우에 상이한 재조합 발현 벡터에서, 그리고 일부 경우에 동일한 재조합 발현 벡터에서): (i) 표적화된 게놈의 표적 좌위의 표적 서열에 혼성화되는 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열(예를 들어, 표적 세포, 예컨대 진핵 세포에서 작동 가능한 프로모터에 작동 가능하게 연결됨); 및 (ii) CasY 단백질을 암호화하는 뉴클레오타이드 서열(예를 들어, 표적 세포, 예컨대 진핵 세포에서 작동 가능한 프로모터에 작동 가능하게 연결됨)을 포함하는 하나 이상의 재조합 발현 벡터를 제공한다.

[0179] 적합한 발현 벡터는 바이러스 발현 벡터(예를 들어, 백시니아 바이러스; 폴리오 바이러스; 아데노바이러스에 기반한 바이러스 벡터(예를 들어, 문헌[Li et al., Invest Ophthalmol Vis Sci 35:2543 2549, 1994; Borrás et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., Hum Gene Ther 5:1088 1097, 1999]; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 및 WO 95/00655); 아데노-연관 바이러스(AAV)(예를 들어, 문헌[Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Ophthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; 및 Flotte et al., PNAS (1993) 90:10613-10617] 참조); SV40; 단순포진 바이러스; 인간 면역결핍 바이러스(예를 들어, 문헌[Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et

al., J Virol 73:7812-7816, 1999]); 레트로바이러스 벡터(예를 들어, 뮤린 백혈병 바이러스, 비장 괴사 바이러스, 및 레트로바이러스, 예컨대 라우스 육종 바이러스, 하비 육종 바이러스, 조류 백혈병 바이러스, 렌티바이러스, 인간 면역결핍 바이러스, 골수증식성 육종 바이러스 및 유방암 바이러스로부터 유래된 벡터); 등을 포함한다. 일부 경우에, 본 개시내용의 재조합 발현 벡터는 재조합 아데노-연관 바이러스(AAV) 벡터이다. 일부 경우에, 본 개시내용의 재조합 발현 벡터는 재조합 렌티바이러스 벡터이다. 일부 경우에, 본 개시내용의 재조합 발현 벡터는 재조합 레트로바이러스 벡터이다.

[0180] 이용되는 숙주/벡터 시스템에 따라서, 다수의 적합한 전사 및 번역 제어 요소(구성적 및 유도성 프로모터, 전사 인핸서 요소, 전사 종결자 등을 포함)가 발현 벡터에서 사용될 수 있다.

[0181] 일부 실시형태에서, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열은 제어 요소, 예를 들어, 전사 제어 요소, 예컨대 프로모터에 작동 가능하게 연결된다. 일부 실시형태에서, CasY 단백질 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 제어 요소, 예를 들어, 전사 제어 요소, 예컨대 프로모터에 작동 가능하게 연결된다.

[0182] 전사 제어 요소는 프로모터일 수 있다. 일부 경우에, 프로모터는 구성적으로 활성인 프로모터이다. 일부 경우에, 프로모터는 조절 가능한 프로모터이다. 일부 경우에, 프로모터는 유도성 프로모터이다. 일부 경우에, 프로모터는 조직-특이적 프로모터이다. 일부 경우에, 프로모터는 세포 유형-특이적 프로모터이다. 일부 경우에, 전사 제어 요소(예를 들어, 프로모터)는 표적화된 세포 유형 또는 표적화된 세포 집단에서 기능성이다. 예를 들어, 일부 경우에, 전사 제어 요소는 진행 세포, 예를 들어, 조혈모세포(예를 들어, 이동 말초 혈액(mobilized peripheral blood: mPB) CD34(+) 세포, 골수(BM) CD34(+) 세포 등)에서 기능성일 수 있다.

[0183] 진행 프로모터(진행 세포에서 프로모터 기능성)인 비제한적 예는 EF1 α , 거대세포바이러스(CMV) 급초기, 단순포진 바이러스(HSV) 티미딘 키나제, 초기 및 후기 SV40, 레트로바이러스로부터의 긴 말단 반복부(LTR) 및 마우스 메탈로티오네인-I로부터의 프로모터를 포함한다. 적절한 벡터 및 프로모터의 선택은 당업자 수준 내에서 매우 용이하다. 발현 벡터는 또한 번역 개시 및 전사 종결자에 대한 리보솜 결합 부위를 함유할 수 있다. 발현 벡터는 또한 발현을 증폭시키기 위한 적절한 서열을 포함할 수 있다. 발현 벡터는 또한 CasY 단백질에 융합되어, 키메라 CasY 폴리펩타이드를 생성할 수 있는 단백질 태그(예를 들어, 6xHis 태그, 혈구응집소 태그, 형광 단백질 등)를 암호화하는 뉴클레오타이드 서열을 포함할 수 있다.

[0184] 일부 실시형태에서, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열 및/또는 CasY 융합 폴리펩타이드는 유도성 프로모터에 작동 가능하게 연결된다. 일부 실시형태에서, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열 및/또는 CasY 융합 단백질은 구성적 프로모터에 작동 가능하게 연결된다.

[0185] 프로모터는 구성적으로 활성인 프로모터(즉, 구성적으로 활성인/"온(ON)" 상태인 프로모터)일 수 있고, 유도성 프로모터(즉, 상태가 활성/"온" 또는 비활성/"오프"인 프로모터, 외부 자극, 예를 들어, 특정 온도, 화합물 또는 단백질의 존재에 의해 제어됨)일 수 있으며, 공간적으로 제한된 프로모터(즉, 전사 제어 요소, 인핸서 등)(예를 들어, 조직 특이적 프로모터, 세포 유형 특이적 프로모터 등)일 수 있고, 그리고 일시적으로 제한된 프로모터일 수 있다(즉, 프로모터는 배아 발생의 특정 단계 동안에 또는 생물학적 과정(예를 들어, 마우스에서의 모낭 주기)의 특정 단계 동안에 "온" 상태 또는 "오프" 상태이다).

[0186] 적합한 프로모터는 바이러스로부터 유래될 수 있고, 따라서 바이러스 프로모터로서 지칭될 수 있거나, 또는 그들은 원핵 또는 진핵 유기체를 포함하는 임의의 유기체로부터 유래될 수 있다. 적합한 프로모터는 임의의 RNA 중합효소(예를 들어, pol I, pol II, pol III)에 의해 발현을 유도하는 데 사용될 수 있다. 예시적인 프로모터는 SV40 초기 프로모터, 마우스 유방 종양 바이러스 긴 말단 반복부(LTR) 프로모터; 아데노바이러스 주요 후기 프로모터(Ad MLP); 단순포진 바이러스(HSV) 프로모터, 거대세포바이러스(CMV) 프로모터, 예컨대 CMV 급초기 프로모터 영역(CMVIE), 라우스 육종 바이러스(RSV) 프로모터, 인간 U6 소핵 프로모터(U6)(Miyagishi et al., Nature Biotechnology 20, 497 - 500 (2002)), 증강된 U6 프로모터(예를 들어, Xia et al., Nucleic Acids Res. 2003 Sep 1;31(17)), 인간 H1 프로모터(H1) 등을 포함하지만, 이들로 제한되지 않는다.

[0187] 일부 경우에, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열은 진행 세포에서 작동 가능한 프로모터(예를 들어, U6 프로모터, 증강된 U6 프로모터, H1 프로모터 등)(의 제어 하에서) 작동 가능하게 연결된다. 당업자에 의해 이해될 바와 같이, U6 프로모터(예를 들어, 진행 세포에서), 또는 다른 Pol III 프로모터를 이용하여 핵산(예를 들어, 발현 벡터)로부터 RNA(예를 들어, 가이드 RNA)를 발현시킬 때, 행에 몇 개의 T가 있다면(RNA에서 U에 대한 암호) RNA는 돌연변이될 필요가 있을 수 있다. 이는 DNA 내 T(예를 들어, 5개의 T)의 문자열이 중합효

소 III(PolIII)에 대한 종결자로서 작용할 수 있기 때문이다. 따라서, 진핵 세포 내 가이드 RNA의 전사를 보장하기 위해, 때때로 T의 실행을 제거하는 가이드 RNA를 암호화하는 서열을 변형할 필요가 있을 수 있다. 일부 경우에, CasY 단백질(예를 들어, 야생형 CasY 단백질, 틸내기효소 CasY 단백질, dCasY 단백질, 키메라 CasY 단백질 등)을 암호화하는 뉴클레오타이드 서열은 진핵 세포에서 작동 가능한 프로모터(예를 들어, CMV 프로모터, EF1 α 프로모터, 에스트로겐 수용체-조절 프로모터 등)에 작동 가능하게 연결된다.

[0188] 유도성 프로모터의 예는 T7 RNA 중합효소 프로모터, T3 RNA 중합효소 프로모터, 아이소프로필-베타-D-티오갈락토피라노사이드(IPTG)-조절된 프로모터, 락토스 유도된 프로모터, 열 충격 프로모터, 테트라사이클린-조절된 프로모터, 스테로이드-조절된 프로모터, 금속-조절된 프로모터, 에스트로겐 수용체-조절된 프로모터 등을 포함하지만, 이들로 제한되지 않는다. 따라서 유도성 프로모터는 독시사이클린; 에스트로겐 및/또는 에스트로겐 유사체; IPTG 등을 포함하지만, 이들로 제한되지 않는 분자에 의해 조절될 수 있다.

[0189] 사용에 적합한 유도성 프로모터는 본 명세서에 기재되거나 또는 당업자에게 공지된 임의의 유도성 프로모터를 포함한다. 유도성 프로모터의 예는 화학적으로/생화학적으로-조절된 그리고 물리적으로 조절된 프로모터, 예컨대 알코올-조절된 프로모터, 테트라사이클린-조절된 프로모터 (예를 들어, 테트라사이클린 리프레서 단백질(tetR), 테트라사이클린 오퍼레이터 서열(tetO) 및 테트라사이클린 트랜스작용인자 융합 단백질(tTA)을 포함하는, 안하이드로테트라사이클린(aTc)-반응성 프로모터 및 다른 테트라사이클린-반응성 프로모터 시스템), 스테로이드-조절된 프로모터(예를 들어, 래트 글루코코르티코이드 수용체에 기반한 프로모터, 인간 에스트로겐 수용체, 나방 엑디손 수용체, 및 스테로이드/레티노이드/갑상선 수용체 슈퍼패밀리로부터의 프로모터), 금속-조절된 프로모터(예를 들어, 효모, 마우스 및 인간으로부터의 메탈로티오네인으로부터 유래된 프로모터(금속 이온에 결합하고 격리시키는 단백질) 유전자), 발병-조절된 프로모터(예를 들어, 살리실산, 에틸렌 또는 벤조티아다리아졸(BTH)에 의해 유도), 온도/열-유도성 프로모터(예를 들어, 열 충격 프로모터), 및 광-조절된 프로모터(예를 들어, 식물 세포로부터의 광 반응성 프로모터)를 포함하지만, 이들로 제한되지 않는다.

[0190] 일부 경우에, 다세포 유기체에서, 프로모터가 특정 세포의 서브세트에서 활성화(즉, "온")이 되도록 프로모터는 공간적으로 제한된 프로모터(즉, 세포 유형 특이적 프로모터, 조직 특이적 프로모터 등)이다. 공간적으로 제한된 프로모터는 또한 인핸서, 전사 제어 요소, 제어 서열 등으로서 지칭될 수 있다. 프로모터가 표적화된 숙주 세포(예를 들어, 진핵 세포; 원핵 세포)에서 기능성이라면, 임의의 편리한 공간적으로 제한된 프로모터가 사용될 수 있다.

[0191] 일부 경우에, 프로모터는 가역적 프로모터이다. 가역적 유도성 프로모터를 포함하는 적합한 가역적 프로모터는 당업계에 공지되어 있다. 이러한 가역적 프로모터는 다수의 유기체, 예컨대 진핵생물 및 원핵생물로부터 단리 및 유래될 수 있다. 제2 유기체, 예를 들어, 제1 원핵생물 및 제2 진핵생물, 제1 진핵생물 및 제2 원핵생물에서 사용하기 위한 제1 유기체로부터 유래된 가역적 프로모터의 변형은 당업계에 잘 공지되어 있다. 이러한 가역적 프로모터, 및 이러한 가역적 프로모터에 기반할 뿐만 아니라 추가적인 제어 단백질을 포함하는 시스템은 알코올 조절된 프로모터(예를 들어, 알코올 탈수소효소 I(alcA) 유전자 프로모터, 알코올 트랜스작용인자 단백질(AlcR)에 반응성인 프로모터 등), 테트라사이클린 조절된 프로모터, (예를 들어, 테트라사이클린활성제(TetActivator), 테트라사이클린온(TetON), 테트라사이클린오프(TetOFF) 등을 포함하는 프로모터 시스템), 스테로이드 조절된 프로모터(예를 들어, 래트 글루코코르티코이드 수용체 프로모터 시스템, 인간 에스트로겐 수용체 프로모터 시스템, 레티노이드 프로모터 시스템, 갑상선 프로모터 시스템, 엑디손 프로모터 시스템, 미페프리스톤 프로모터 시스템 등), 금속 조절된 프로모터(예를 들어, 메탈로티오네인 프로모터 시스템 등), 발병-관련 조절된 프로모터(예를 들어, 살리실산 조절된 프로모터, 에틸렌 조절된 프로모터, 벤조티아다리아졸 조절된 프로모터 등), 온도 조절된 프로모터(예를 들어, 열 충격 유도성 프로모터(예를 들어, HSP-70, HSP-90, 대두 열 충격 프로모터 등), 광 조절된 프로모터, 합성 유도성 프로모터 등을 포함하지만, 이들로 제한되지 않는다.

[0192] 핵산(예를 들어, 공여자 폴리뉴클레오타이드 서열을 포함하는 핵산, CasY 단백질 및/또는 CasY 가이드 RNA를 암호화하는 하나 이상의 핵산 등)을 숙주 세포 내로 도입하는 방법은 당업계에 공지되어 있고, 핵산(예를 들어, 발현 작제물)을 세포 내로 도입하기 위해 임의의 편리한 방법이 사용될 수 있다. 적합한 방법은, 예를 들어, 바이러스 감염, 형질감염, 리포펙션, 전기천공법, 인산칼슘 침전, 폴리에틸렌이민(PEI)-매개 형질감염, DEAE-덱스트란 매개 형질감염, 리포좀-매개 형질감염, 입자총 기법, 인산칼슘 침전, 직접 미량주사법, 나노입자-매개 핵산 전달 등을 포함한다.

[0193] 재조합 발현 벡터를 세포 내로 도입하는 것은 임의의 배양 배지에서 그리고 세포 생존을 촉진시키는 임의의 배양 조건 하에 생길 수 있다. 표적 세포 내로 재조합 발현 벡터를 도입하는 것은 생체내 또는 생체외에서 수행될

수 있다. 표적 세포 내로 재조합 발현 벡터를 도입하는 것은 시험관외에서 수행될 수 있다.

- [0194] 일부 실시형태에서, CasY 단백질은 RNA로서 제공될 수 있다. RNA는 직접 화학 합성에 의해 제공될 수 있거나 또는 DNA(예를 들어, CasY 단백질을 암호화)로부터 시험관외에서 전사될 수 있다. 일단 합성되면, RNA는 핵산을 세포 내로 도입하기 위한 임의의 잘 공지된 기법(예를 들어, 미량주사법, 전기천공법, 형질감염 등)에 의해 세포 내로 도입될 수 있다.
- [0195] 핵산은 잘 개발된 형질감염 기법(예를 들어, 문헌[Angel and Yanik (2010) PLoS ONE 5(7): e11756] 참조), 및 퀴아젠사(Qiagen)로부터의 상업적으로 입수 가능한 TransMessenger(등록상표) 시약, 스템젠트사(Stemgent)로부터의 Stemfect(상표명) RNA 형질감염 키트, 및 미루스 바이오 엘엘씨(Mirus Bio LLC)로부터의 TransIT(등록상표)-mRNA 형질감염 키트를 이용하여 세포에 제공될 수 있다. 또한 문헌[Beumer et al. (2008) PNAS 105(50):19821-19826] 참조.
- [0196] 벡터는 표적 숙주 세포에 직접 제공될 수 있다. 다시 말해서, 벡터가 세포에 의해 취해지도록, 세포는 대상 핵산을 포함하는 벡터(예를 들어, 공여자 주형 서열을 갖고 CasY 가이드 RNA를 암호화하는 재조합 발현 벡터; CasY 단백질을 암호화하는 재조합 발현 벡터 등)와 접촉된다. 세포를 플라스미드인 핵산 벡터와 접촉시키는 방법은 당업계에 잘 공지되어 있는 전기천공법, 열화칼슘 형질감염, 미량주사법, 및 리포펙션을 포함한다. 바이러스 벡터 전달을 위해, 세포는 대상 바이러스 발현 벡터를 포함하는 바이러스 입자와 접촉될 수 있다.
- [0197] 레트로바이러스, 예를 들어, 렌티바이러스는 본 개시내용의 방법에서 사용하기에 적합하다. 통상적으로 사용되는 레트로바이러스 벡터는 "결합성"이며, 즉, 증식성 감염에 필요한 바이러스 단백질을 생성할 수 없다. 오히려, 벡터의 복제는 패키징 세포주에서의 성장을 필요로 한다. 관심 대상을 포함하는 바이러스 입자를 생성하기 위해, 핵산을 포함하는 레트로바이러스 핵산은 패키징 세포주에 의해 바이러스 캡시드 내로 패키징된다. 상이한 패키징 세포주는캡시드 내로 혼입될 상이한 외피 단백질(동종숙주성(ecotropic), 양쪽성 또는 이종향성(xenotropic))을 제공하는데, 이 외피 단백질은 세포에 대한 바이러스 입자의 특이성(뮤린 및 래트에 대한 동종숙주; 인간, 개 및 마우스를 포함하는 대부분의 포유류 세포 유형에 대해 양쪽성; 및 뮤린 세포를 제외한 대부분의 포유류 세포에 대해 이종향성)을 결정한다. 적절한 패키징 세포주를 사용하여 세포가 패키징된 바이러스 입자에 의해 표적화되는 것을 보장할 수 있다. 대상 벡터 발현 벡터를 패키징 세포주에 도입하고 패키징 세포주에 의해 생성된 바이러스 입자를 수집하는 방법은 당업계에 잘 공지되어 있다. 핵산은 또한 직접 미량주사법(예를 들어, RNA의 주사)에 의해 도입될 수 있다.
- [0198] CasY 가이드 RNA 및/또는 CasY 폴리펩타이드를 암호화하는 핵산을 표적 숙주 세포에 제공하기 위해 사용되는 벡터는 발현을 유도하기 위한, 즉, 관심 대상의 핵산의 전사 활성화를 위한 적합한 프로모터를 포함할 수 있다. 다시 말해서, 일부 경우에, 관심 대상의 핵산은 프로모터에 작동 가능하게 연결될 것이다. 이는 편재하는 작용 프로모터, 예를 들어, CMV- β -액틴 프로모터, 또는 유도성 프로모터, 예컨대 특정 세포 집단에서 활성이거나 또는 테트라사이클린과 같은 약물의 존재에 반응하는 프로모터를 포함할 수 있다. 전사 활성화에 의해, 전사는 10 배, 100배만큼, 더 보통으로는 1000배만큼 표적 세포에서 기준 수준 초과로 증가될 것이 의도된다. 추가로, CasY 가이드 RNA 및/또는 CasY 단백질을 암호화하는 핵산을 세포에 제공하기 위해 사용되는 벡터는 CasY 가이드 RNA 및/또는 CasY 단백질을 취한 세포를 동정하기 위해 표적 세포에서 선택 가능한 마커를 암호화하는 핵산 서열을 포함할 수 있다.
- [0199] CasY 폴리펩타이드, 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산은 일부 경우에 RNA이다. 따라서, CasY 융합 단백질은 RNA로서 세포 내로 도입될 수 있다. 세포 내로 RNA를 도입하는 방법은 당업계에 공지되어 있고, 예를 들어, 직접 미량주사법, 형질감염, 또는 DNA의 도입을 위해 사용되는 임의의 다른 방법을 포함할 수 있다. CasY 단백질은 대신에 폴리펩타이드로서 세포에 제공될 수 있다. 이러한 폴리펩타이드는 생성물의 용해도를 증가시키는 폴리펩타이드에 선택적으로 융합될 수 있다. 도메인은 정해진 프로테아제 절단 부위, 예를 들어, TEV 프로테아제에 의해 절단되는 TEV 서열을 통해 폴리펩타이드에 연결될 수 있다. 링커는 또한 하나 이상의 가요성 서열, 예를 들어, 1 내지 10개의 글리신 잔기를 포함할 수 있다. 일부 실시형태에서, 융합 단백질의 절단은 생성물의 용해도를 유지하는 완충제에서, 예를 들어, 0.5 내지 2 M 유레아의 존재 하에, 용해도 등을 증가시키는 폴리펩타이드 및/또는 폴리뉴클레오타이드의 존재 하에 수행된다. 관심 대상의 도메인은 엔도솜 용해성 도메인, 예를 들어, 인플루엔자 HA 도메인; 및 생성에 도움을 주는 다른 폴리펩타이드, 예를 들어, IF2 도메인, GST 도메인, GRPE 도메인 등을 포함한다. 폴리펩타이드는 개선된 안정성을 위해 제형화될 수 있다. 예를 들어, 펩타이드는 폐길화될 수 있으며(PEGylated), 여기서 폴리에틸렌옥시기는 혈류에서 증강된 증강된 수명을 제공한다.

- [0200] 추가적으로 또는 대안적으로, 본 개시내용의 CasY 폴리펩타이드는 세포에 의한 흡수를 촉진하기 위해 폴리펩타이드 침투 도메인에 융합될 수 있다. 다수의 침투 도메인이 당업계에 공지되어 있고, 펩타이드, 펩타이드 모방체 및 비-펩타이드 담체를 포함하는 본 개시내용의 비통합 폴리펩타이드에서 사용될 수 있다. 예를 들어, 침투 펩타이드는 아미노산 서열 RQIKIWFQNRRMKWKK(서열번호 133)를 포함하는, 페너트라틴(*penetratin*)으로서 지칭되는 노랑초파리 전사 인자 안테나피디어(*Antennapedia*)의 세 번째 알파 나선으로부터 유래될 수 있다. 다른 예로서, 침투 펩타이드는, 예를 들어, 천연 유래 tat 단백질의 아미노산 49 내지 57을 포함할 수 있는 HIV-1 tat 염기성 영역 아미노산 서열을 포함한다. 다른 침투 도메인은 폴리-알기닌 모티프, 예를 들어, HIV-1 rev 단백질, 노나-알기닌, 옥타-알기닌 등의 아미노산 34 내지 56의 영역을 포함한다. (예를 들어, 전좌 펩타이드 및 펩토이드의 교시에 대해 본 명세서에 구체적으로 참고로 편입된 문헌[Futaki et al. (2003) Curr Protein Pept Sci. 2003 Apr; 4(2): 87-9 및 446; 및 Wender et al. (2000) Proc. Natl. Acad. Sci. U.S.A 2000 Nov. 21; 97(24):13003-8]; 공개된 미국 특허 출원 제20030220334호; 제20030083256호; 제20030032593호; 및 제20030022831호를 참조). 노나-알기닌(R9) 서열은 특성규명된 더 효율적인 PTD 중 하나이다(Wender et al. 2000; Uemura et al. 2002). 융합이 이루어지는 부위는 폴리펩타이드의 생물학적 활성, 분비 또는 결합 특징을 최적화하기 위해 선택될 수 있다. 최적의 부위는 일상적인 실험에 의해 결정될 것이다.
- [0201] 본 개시내용의 CasY 폴리펩타이드는 시험관내에서 또는 진핵세포에 의해 또는 원핵 세포에 의해 생성될 수 있고, 비폴딩, 예를 들어, 열 변성, 다이티오트레이틀 환원 등에 의해 추가로 가공될 수 있고, 당업계에 공지된 방법을 이용하여 추가로 재폴딩될 수 있다.
- [0202] 1차 서열을 변경시키지 않는 관심 대상의 변형은 폴리펩타이드의 화학적 유도체화, 예를 들어, 아실화, 아세틸화, 카복실화, 아마이드화 등을 포함한다. 또한 글리코실화의 변형, 예를 들어, 합성 및 가공 동안 또는 추가 가공 단계에서 폴리펩타이드의 글리코실화 패턴을 변형시킴으로써; 예를 들어, 글리코실화, 예컨대 포유류 글리코실화 또는 탈글리코실화 효소에 영향을 미치는 효소에 폴리펩타이드를 노출시킴으로써 만들어지는 것이 포함된다. 또한 포스포틸화된 아미노산 잔기, 예를 들어, 포스포타이로신, 포스포세린 또는 포스포트레오닌을 갖는 서열이 포함된다.
- [0203] 또한 단백질 분해에 대한 그들의 내성을 개선시키기 위해, 표적 서열 특이성을 변화시키기 위해, 용해도 특성을 최적화하기 위해, 단백질 활성(예를 들어, 전사 조절 활성, 효소 활성 등)을 변경하기 위해 또는 그들을 더 적합하게 제공하기 위해 보통의 분자 생물학 기법 및 합성 화학을 이용하여 변형된 핵산(예를 들어, CasY 가이드 RNA를 암호화, CasY 융합 단백질을 암호화 등) 및 단백질(예를 들어, 야생형 단백질 또는 변이체 단백질로부터 유래된 CasY 융합 단백질)이 본 개시내용의 실시형태에 포함하기에 적합하다. 이러한 폴리펩타이드의 유사체는 천연 유래 L-아미노산, 예를 들어, D-아미노산 또는 비천연 유래 합성 아미노산 이외의 잔기를 함유하는 것을 포함한다. D-아미노산은 일부 또는 모든 아미노산 잔기로 치환될 수 있다.
- [0204] 본 개시내용의 CasY 폴리펩타이드는 당업계에 공지된 통상적인 방법을 이용하여 시험관내 합성에 의해 제조될 수 있다. 다양한 상업적 합성 장치, 예를 들어, 어플라이드 바이오시스템즈, 인코포레이티드(Applied Biosystems, Inc.), 베크만(Beckman) 등에 의한 자동화된 합성기를 이용 가능하다. 합성기를 이용함으로써, 천연 유래 아미노산은 비천연 아미노산으로 치환될 수 있다. 특정 서열 및 제조 방식은 편리함, 경제성, 필요한 순도 등에 의해 결정될 것이다.
- [0205] 원한다면, 합성 동안 또는 발현 동안에 다양한 기가 펩타이드 내로 도입되는데, 이는 다른 분자에 또는 표면에 대한 연결을 허용한다. 따라서 티오에터, 금속 이온 복합체에 대한 연결을 위한 히스티딘, 아마이드 또는 에스터를 형성하기 위한 카복실기 또는 아마이드를 형성하기 위한 아미노기 등을 만들기 위해 시스템이 사용될 수 있다.
- [0206] 본 개시내용의 CasY 폴리펩타이드는 또한 분리될 수 있고, 재조합 합성의 통상적인 방법에 따라 정제될 수 있다. 발현 숙주의 용해물이 제조될 수 있고, 용해물은 고성능 액체 크로마토그래피(HPLC), 배제 크로마토그래피, 겔 전기영동, 친화도 크로마토그래피 또는 다른 정제 기법을 이용하여 정제된다. 대부분에 대해, 사용되는 조성물은 생성물의 제조 및 그의 정제 방법과 관련된 오염물질에 대해 20중량% 이상의 목적으로 하는 생성물, 더 보통으로는 75중량% 이상, 바람직하게는 95중량% 이상, 그리고 치료적 목적을 위해, 보통 99.5중량% 이상을 포함할 것이다. 보통, 백분율은 총 단백질을 기준으로 할 것이다. 따라서, 일부 경우에, 본 개시내용의 CasY 폴리펩타이드, 또는 CasY 융합 폴리펩타이드는 적어도 80% 순수, 적어도 85% 순수, 적어도 90% 순수, 적어도 95% 순수, 적어도 98% 순수 또는 적어도 99% 순수(예를 들어, 오염물질, 비-CasY 단백질 또는 다른 거대분자 등이 없음)이다.

- [0207] 표적 핵산(예를 들어, 게놈 DNA)에 대한 절단 또는 임의의 목적으로 하는 변형, 또는 표적 핵산, 본 개시내용의 CasY 가이드 RNA 및/또는 CasY 폴리펩타이드 및/또는 공여자 주형 서열과 관련된 폴리펩타이드에 대한 임의의 목적으로 하는 변형을 유도하기 위해, 그들이 핵산으로서 도입되는 또는 폴리펩타이드로서 도입되는, 약 30분 내지 약 24시간, 예를 들어, 1시간, 1.5시간, 2시간, 2.5시간, 3시간, 3.5시간 4시간, 5시간, 6시간, 7시간, 8시간, 12시간, 16시간, 18시간, 20시간, 또는 약 30분 내지 약 24시간의 임의의 다른 기간 동안 세포에 제공되는데, 이는 대략 매일 내지 약 4일마다, 예를 들어, 1.5일마다, 2일마다, 3일마다의 빈도로, 또는 대략 매일 내지 약 4일마다의 임의의 다른 빈도로 반복될 수 있다. 체제(들)는 대상 세포에 1회 이상, 예를 들어, 1회, 2회, 3회, 3회 초과로 제공될 수 있고, 세포는 각각의 접촉 사건 후 일정한 시간 동안, 예를 들어, 16 내지 24시간 동안 체제(들)와 함께 인큐베이션되도록 허용되며, 이 시간 후에 배지는 신선한 배지로 대체되고, 세포는 추가로 배양된다.
- [0208] 2 이상의 상이한 표적화 복합체가 세포(예를 들어, 동일 또는 상이한 표적 핵산 내에서 상이한 서열에 상보적인 2개의 상이한 CasY 가이드 RNA)에 제공되는 경우에, 복합체는 동시에(예를 들어, 2개의 폴리펩타이드 및/또는 핵산으로서) 제공될 수 있거나, 또는 동시에 전달될 수 있다. 대안적으로, 그들은 연속적으로, 예를 들어, 제1의 제공된 표적화 복합체, 다음에 제2 표적화 복합체 등 또는 그 반대로 제공될 수 있다.
- [0209] DNA 벡터의 표적 세포의 전달을 개선시키기 위해, DNA는 손상으로부터 보호될 수 있고, 세포 내로 그의 유입은, 예를 들어, 리포플렉스(lipoplex) 및 폴리플렉스(polyplex)를 이용함으로써 용이하게 된다. 따라서, 일부 경우에, 본 개시내용의 핵산(예를 들어, 본 개시내용의 재조합 발현 벡터)는 마이셀 또는 리포솜과 같이 조직화된 구조에서 지질로 뒤덮일 수 있다. 조직화된 구조가 DNA와 함께 복합체화될 때, 이는 리포플렉스로 불린다. 3가지 유형의 지질, 음이온(음으로 하전됨), 중성 또는 양이온성(양으로 하전됨)이 있다. 양이온성 지질을 이용하는 리포플렉스는 유전자 전달을 위한 증명된 효용을 가진다. 양전하 지질은 그들의 양전하에 기인하여 음으로 하전된 DNA와 자연적으로 복합체화된다. 또한 그들의 하전의 결과로서, 그들은 세포막과 상호작용한다. 이어서, 리포플렉스의 내포작용이 일어나고, DNA는 세포질 내로 방출된다. 양이온성 지질은 또한 세포에 의해 DNA의 분해에 대해 보호한다.
- [0210] DNA와 중합체의 복합체는 폴리플렉스로 불린다. 대부분의 폴리플렉스는 양이온성 중합체로 이루어지고, 그들의 생성은 이온 상호작용에 의해 조절된다. 폴리플렉스와 리포플렉스의 작용 방법 간의 한 가지 큰 차이는 폴리플렉스는 세포질 내로 그들의 DNA 부하를 방출할 수 없고, 이를 위하여, 엔도솜-용해제(내포작용 동안 만들어진 엔도솜을 용해시키기 위함), 예컨대 비활성화 아데노바이러스와의 공동형질감염이 일어나야 한다는 것이다. 그러나, 이는 항상 그런 것은 아니며; 중합체, 예컨대 폴리에틸렌이민은 키토산 및 트라이메틸키토산과 같이 엔도솜 붕괴의 그 자체의 방법을 가진다.
- [0211] 구체 형태를 갖는 상당히 분지형인 거대분자인 덴드리머는 또한 줄기 세포를 유전적으로 변형하기 위해 사용될 수 있다. 덴드리머 입자의 표면은 그의 특성을 변경시키도록 작용화될 수 있다. 특히, 양이온성 덴드리머(즉, 양성 표면 전하를 갖는 것)를 작제하는 것이 가능하다. 유전자 물질, 예컨대 DNA 플라스미드의 존재 하에 있을 때, 전하 상보성은 양이온성 덴드리머와 핵산의 일시적 회합을 야기한다. 그의 목적에 도달되었을 때, 덴드리머-핵산 복합체는 내포작용에 의해 세포 내로 흡수될 수 있다.
- [0212] 일부 경우에, 개시내용의 핵산(예를 들어, 발현 벡터)은 관심 대상의 가이드 서열에 대한 삽입 부위를 포함한다. 예를 들어, 핵산은 관심 대상의 가이드 서열에 대한 삽입 부위를 포함할 수 있으며, 여기서 삽입 부위는 가이드 서열이 목적으로 하는 표적 서열(예를 들어, 가이드 RNA의 CasY 결합 양상에 기여하는 서열, 예를 들어, CasY 가이드 RNA의 dsRNA 듀플렉스(들)에 기여하는 서열 - 가이드 RNA의 이 부분은 또한 가이드 RNA의 '스캐폴드' 또는 '불변 영역'으로서 지칭될 수 있음)에 혼성화하도록 변할 때 변하지 않는 CasY 가이드 RNA의 일부를 암호화하는 뉴클레오타이드 서열에 바로 인접한다. 따라서, 일부 경우에, 가이드 RNA의 가이드 서열을 암호화하는 부분이 삽입 서열(삽입 부위)라는 것을 제외하고, 대상 핵산(예를 들어, 발현 벡터)는 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함한다. 삽입 부위는 목적으로 하는 서열의 삽입을 위해 사용되는 임의의 뉴클레오타이드 서열이다. 다양한 기법과 함께 사용하기 위한 "삽입 부위"는 당업자에게 공지되어 있고, 임의의 편리한 삽입 부위가 사용될 수 있다. 삽입 부위는 핵산 서열을 조작하기 위한 임의의 방법에 대한 것일 수 있다. 예를 들어, 일부 경우에 삽입 부위는 다중 클로닝 부위(MCS)(예를 들어, 하나 이상의 제한 효소 인식 서열을 포함하는 부위), 절찰 독립적 클로닝을 위한 부위, 재조합 기바나 클로닝을 위한 부위(예를 들어, att 부위에 기반한 재조합), CRISPR/Cas(예를 들어, Cas9) 기반 기법에 의해 인식되는 뉴클레오타이드 서열 등이다.
- [0213] 삽입 부위는 임의의 바람직한 길이일 수 있고, 삽입 부위 유형에 의존할 수 있다(예를 들어, 부위가 하나 이상

의 제한 효소 인식 서열을 포함하는지의 여부(및 얼마나 많이 포함하는지), 부위가 CRISPR/Cas 단백질에 대한 표적 부위를 포함하는지의 여부 등에 의존할 수 있다). 일부 경우에, 대상 핵산의 삽입 부위는 길이가 3개 이상의 뉴클레오타이드(nt)이다(예를 들어, 길이가 5개 이상, 8개 이상, 10개 이상, 15개 이상, 17개 이상, 18개 이상, 19개 이상, 20개 이상 또는 25개 이상 또는 30개 이상의 nt). 일부 경우에, 대상 핵산의 삽입 부위 길이는 길이가 2 내지 50개의 뉴클레오타이드(nt)(예를 들어, 2 내지 40 nt, 2 내지 30개의 nt, 2 내지 25개의 nt, 2 내지 20개의 nt, 5 내지 50개의 nt, 5 내지 40개의 nt, 5 내지 30개의 nt, 5 내지 25개의 nt, 5 내지 20개의 nt, 10 내지 50개의 nt, 10 내지 40개의 nt, 10 내지 30개의 nt, 10 내지 25개의 nt, 10 내지 20개의 nt, 17 내지 50개의 nt, 17 내지 40개의 nt, 17 내지 30개의 nt, 17 내지 25 nt) 범위이다. 일부 경우에, 대상 핵산의 삽입 부위 길이는 길이가 5 내지 40 nt이다.

[0214] **핵산 변형**

[0215] 일부 실시형태에서, 새로운 또는 증강된 특징(예를 들어, 개선된 안정성)을 갖는 핵산을 제공하기 위해 대상 핵산(예를 들어, CasY 가이드 RNA)는 하나 이상의 변형, 예를 들어, 염기 변형, 골격 변형 등을 가진다. 뉴클레오사이드는 염기-당 조합이다. 뉴클레오사이드의 염기 부분은 정상적으로 복소환식 염기이다. 이러한 복소환식 염기의 2가지 가장 통상적인 부류는 퓨린 및 피리미딘이다. 뉴클레오타이드는 뉴클레오사이드의 당 부분에 공유적으로 연결된 인산염기를 추가로 포함하는 뉴클레오사이드이다. 펜토피라노실 당을 포함하는 해당 뉴클레오사이드에 대해, 인산염기는 당의 2', 3' 또는 5' 하이드록실 모이어티에 연결될 수 있다. 올리고뉴클레오타이드를 형성함에 있어서, 인산염기는 서로 인접한 뉴클레오사이드에 공유적으로 연결되어 선형 중합체 화합물을 형성한다. 결국, 이 선형 중합체 화합물의 각각의 단부는 추가로 결합되어 원형 화합물을 형성할 수 있지만, 그러나, 선형 화합물이 적합하다. 추가로, 선형 화합물은 내부 뉴클레오타이드 염기 상보성을 가질 수 있으며, 따라서 완전히 또는 부분적으로 이중-가닥 화합물을 생성하기 위한 방식으로 폴딩될 수 있다. 올리고뉴클레오타이드 내에서, 인산염기는 통상적으로 올리고뉴클레오타이드의 뉴클레오사이드간 골격을 형성하는 것으로 지칭된다. RNA 및 DNA의 정상 결합 또는 골격은 3' 내지 5' 포스포다이에스터 결합이다.

[0216] 적합한 핵산 변형은 하기를 포함하지만, 이들로 제한되지 않는다: 2'-O-메틸 변형된 뉴클레오타이드, 2' 플루오로 변형된 뉴클레오타이드, 잠금 핵산(locked nucleic acid: LNA) 변형된 뉴클레오타이드, 펩타이드 핵산(peptide nucleic acid: PNA) 변형된 뉴클레오타이드, 포스포로티오에이트 결합을 갖는 뉴클레오타이드 및 5' 캡(예를 들어, 7-메틸구아닐레이트 캡(m7G)). 추가적인 상세한 설명 및 추가적인 변형을 이하에 기재한다.

[0217] 2'-O-메틸 변형된 뉴클레오타이드(또한 2'-O-메틸 RNA로서 지칭됨)는 tRNA에서 발견되는 RNA 및 전사후 변형으로서 생기는 다른 작은 RNA의 천연 유래 변형이다. 2'-O-메틸 RNA를 함유하는 올리고뉴클레오타이드가 직접적으로 합성될 수 있다. 이 변형은 RNA:RNA 듀플렉스의 T_m을 증가시키지만 RNA:DNA 안정성의 작은 변화만을 초래한다. 이는 단일-가닥 리보뉴클레아제에 의한 공격에 대해 안정하며 전형적으로 DNA보다 DNase에 대해 5 내지 10 배 덜 민감하다. 이는 표적 전령에 대한 안정성 및 결합 친화도를 증가시키기 위한 수단으로서 안티센스 올리고에서 통상적으로 사용된다.

[0218] 2' 플루오로 변형된 뉴클레오타이드(예를 들어, 2' 플루오로 염기)는 결합 친화도(T_m)를 증가시키는 플루오린 변형된 리보스를 가지며, 또한 천연 RNA에 비해 일부 상대적 뉴클레아제 내성을 부여한다. 이들 변형은 혈청 또는 다른 생물학적 유체에서의 안정성을 개선시키기 위해 리보자임 및 siRNA에서 통상적으로 사용된다.

[0219] LNA 염기는 RNA A-형 나선 듀플렉스 기하학을 선호하는 C3'-엔도 위치에서 염기를 잠그는 리보스 골격에 대한 변형을 가진다. 이 변형은 T_m을 상당히 증가시키며, 또한 매우 뉴클레아제 내성이다. 다중 LNA 삽입은 3'-단부를 제외한 임의의 위치에서 올리고에 위치될 수 있다. 적용은 혼성화 프로브에 대한 안티센스 올리고로부터 SNP 검출 및 대립유전자 특이적 PCR까지의 범위로 기재되었다. LNA에 의해 부여되는 T_m의 큰 증가에 기인하여, 그들은 또한 프라이머 이량체 형성뿐만 아니라 자기-헤어핀 형성의 증가를 야기할 수 있다. 일부 경우에, 단일 올리고 내로 혼입된 LNA의 수는 10개 이하의 염기이다.

[0220] 포스포로티오에이트(PS) 결합(즉, 포스포로티오에이트 결합)은 핵산(예를 들어, 올리고)의 인산염 골격에서 황 원자를 비-브리징 산소로 치환한다. 이 변형은 뉴클레아제 분해에 내성이 있는 뉴클레오타이드간 결합을 제공한다. 포스포로티오에이트 결합은 엑소뉴클레아제 분해를 저해하기 위해 올리고의 5'- 또는 3'-단부에서 마지막 3 내지 5개 뉴클레오타이드 사이에 도입될 수 있다. 올리고 내에서(예를 들어, 전체 올리고를 전체적으로) 포스포로티오에이트 결합을 포함하는 것은 마찬가지로 엔도뉴클레아제에 의한 공격을 감소시킬 수 있다.

[0221] 일부 실시형태에서, 대상 핵산은 2'-O-메틸 변형된 뉴클레오타이드인 하나 이상의 뉴클레오타이드이다. 일부 실

시형태에서, 대상 핵산(예를 들어, dsRNA, siNA 등)은 하나 이상의 2' 플루오로 변형된 뉴클레오타이드를 가진다. 일부 실시형태에서, 대상 핵산(예를 들어, dsRNA, siNA 등)은 하나 이상의 LNA 염기를 가진다. 일부 실시형태에서, 대상 핵산(예를 들어, dsRNA, siNA 등)은 포스포로티오에이트 결합에 의해 연결된 하나 이상의 뉴클레오타이드를 가진다(즉, 대상 핵산은 하나 이상의 포스포로티오에이트 결합을 가진다). 일부 실시형태에서, 대상 핵산(예를 들어, dsRNA, siNA 등)은 5' 캡(예를 들어, 7-메틸구아닐레이트 캡(m7G))을 가진다. 일부 실시형태에서, 대상 핵산(예를 들어, dsRNA, siNA 등)은 변형된 뉴클레오타이드의 조합을 가진다. 예를 들어, 대상 핵산(예를 들어, dsRNA, siNA 등)은 다른 변형을 갖는 하나 이상의 뉴클레오타이드(예를 들어, 2'-O-메틸 뉴클레오타이드 및/또는 2' 플루오로 변형된 뉴클레오타이드 및/또는 LNA 염기 및/또는 포스포로티오에이트 결합)를 갖는 것에 추가로 5' 캡(예를 들어, 7-메틸구아닐레이트 캡(m7G))을 가질 수 있다.

[0222] **변형된 골격 및 변형된 뉴클레오사이드간 결합**

[0223] 변형을 함유하는 적합한 핵산(예를 들어, CasY 가이드 RNA)의 예는 변형된 골격 또는 비천연 뉴클레오사이드간 결합을 함유하는 핵산을 포함한다. 변형된 골격을 갖는 핵산은 골격에 인 원자를 보유하는 것 및 골격에 인 원자를 갖지 않는 것을 포함한다.

[0224] 인 원자를 함유하는 적합한 변형된 올리고뉴클레오타이드 골격은, 예를 들어, 포스포로티오에이트, 카이랄 포스포로티오에이트, 포스포로다이티오에이트, 포스포트라이에스터, 아미노알킬포스포트라이에스터, 메틸 및 다른 알킬 포스포네이트(3'-알킬렌 포스포네이트, 5'-알킬렌 포스포네이트 및 카이랄 포스포네이트를 포함), 포스포네이트, 포스포아미데이트(3'-아미노 포스포아미데이트 및 아미노알킬포스포아미데이트, 포스포로다이아미데이트, 티오노포스포아미데이트를 포함), 티오노알킬포스포네이트, 티오노알킬포스포트라이에스터, 셀레노포스포이트 및 정상 3'-5' 결합을 갖는 보라노포스포이트, 이들의 2'-5' 연결된 유사체, 및 반전된 극성을 갖는 것을 포함하되, 하나 이상의 뉴클레오타이드간 결합은 3' 내지 3', 5' 내지 5' 또는 2' 내지 2' 결합이다. 반전된 극성을 갖는 적합한 올리고뉴클레오타이드는 가장 3'의 뉴클레오타이드간 결합에서 단일 3' 내지 3' 결합, 즉, 염기성일 수 있는 단일 반전 뉴클레오사이드 잔기를 포함한다(뉴클레오염기는 그 대신에 상실되거나 하이드록실기를 가짐). 다양한 염(예컨대, 칼륨 또는 나트륨), 혼합된 염 및 유리 산 형태가 또한 포함된다.

[0225] 일부 실시형태에서, 대상 핵산은 하나 이상의 포스포로티오에이트 및/또는 헤테로원자 뉴클레오사이드간 결합, 특히 $-\text{CH}_2-\text{NH}-\text{O}-\text{CH}_2-$, $-\text{CH}_2-\text{N}(\text{CH}_3)-\text{O}-\text{CH}_2-$ (메틸렌(메틸이미노) 또는 MMI 골격으로서 알려짐), $-\text{CH}_2-\text{O}-\text{N}(\text{CH}_3)-\text{CH}_2-$, $-\text{CH}_2-\text{N}(\text{CH}_3)-\text{N}(\text{CH}_3)-\text{CH}_2-$ 및 $-\text{O}-\text{N}(\text{CH}_3)-\text{CH}_2-\text{CH}_2-$ 를 포함한다(천연 포스포다이에스터 뉴클레오타이드간 결합은 $-\text{O}-\text{P}(=\text{O})(\text{OH})-\text{O}-\text{CH}_2-$ 로서 나타냄). MMI 유형 뉴클레오사이드간 결합은 상기 언급된 미국 특허 제5,489,677호에 개시되어 있으며, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입된다. 적합한 아마이드 뉴클레오사이드간 결합은 미국 특허 제5,602,240호에 개시되어 있으며, 이의 개시내용은 전문이 본 명세서에 참고로 편입된다.

[0226] 또한, 예를 들어, 미국 특허 제5,034,506호에 기재된 바와 같은 몰폴리노 골격 구조를 갖는 핵산이 적합하다. 예를 들어, 일부 실시형태에서, 대상 핵산은 리보스 고리 대신에 6-원 몰폴리노 고리를 포함한다. 일부 이들 실시형태에서, 포스포로다이아미데이트 또는 다른 비-포스포다이에스터 뉴클레오사이드간 결합은 포스포다이에스터 결합을 대체한다.

[0227] 인 원자를 포함하지 않는 적합한 변형된 폴리뉴클레오타이드 골격은 짧은 쇠알킬 또는 사이클로알킬 뉴클레오사이드간 결합, 혼합된 헤테로원자 및 알킬 또는 사이클로알킬 뉴클레오사이드간 결합 또는 하나 이상의 짧은 쇠헤테로원자 또는 복소환식 뉴클레오사이드간 결합에 의해 형성된 골격을 가진다. 이들은 몰폴리노 결합을 갖는 것(뉴클레오사이드의 당 부분으로부터 부분적으로 형성됨); 실록산 골격; 설펜사이드, 설펜사이드 및 설펜 골격; 폼아세틸 및 티오폴아세틸 골격; 메틸렌 폼아세틸 및 티오폴아세틸 골격; 리보아세틸 골격; 알켄 함유 골격; 설펜사이드 골격; 메틸렌이미노 및 메틸렌하이드라지노 골격; 설펜사이드 및 설펜아마이드 골격; 아마이드 골격; 및 혼합된 N, O, S 및 CH_2 성분 부분을 갖는 다른 것을 포함한다.

[0228] **모방체**

[0229] 대상 핵산은 핵산 모방체일 수 있다. 폴리뉴클레오타이드에 적용되는 용어 "모방체"는 폴리뉴클레오타이드를 포함하는 것으로 의도되되, 푸라노스 고리만 또는 푸라노스 고리와 뉴클레오타이드가나 결합 둘 다는 비-푸라노스기로 대체되며, 푸라노스 고리만의 대체는 또한 본 명세서에서 당 대용물로서 지칭된다. 복소환식 염기 모이어티 또는 변형된 복소환식 염기 모이어티는 적절한 표적 핵산과의 혼성화를 위해 유지된다. 우수한 혼성화 특성을 갖는 것을 나타낸 하나의 이러한 핵산, 폴리뉴클레오타이드 모방체는 펩타이드 핵산(PNA)으로서 지칭된다.

PNA에서, 폴리뉴클레오타이드의 당-골격은 골격, 특히 아미노에틸글리신 골격을 함유하는 아마이드로 대체된다. 뉴클레오타이드가 보유되며, 골격의 아마이드 부분의 아자 질소 원자에 직접 또는 간접적으로 결합된다.

[0230] 우수한 혼성화 특성을 갖는 것으로 보고된 하나의 폴리뉴클레오타이드 모방체는 펩타이드 핵산(PNA)이다. PNA 화합물의 골격은 PNA에 아마이드 함유 골격을 제공하는 2 이상의 연결된 아미노에틸 글리신 단위이다. 복소환식 염기 모이어티는 골격의 아마이드 부분의 아자 질소 원자에 직접 또는 간접적으로 결합된다. PNA 화합물의 제조를 기재하는 대표적인 미국 특허는 미국 특허 제5,539,082호; 제5,714,331호; 및 제5,719,262호를 포함하지만, 이들로 제한되지 않으며, 이의 개시내용은 본 명세서에 전문이 참고로 포함된다.

[0231] 연구된 다른 부류의 폴리뉴클레오타이드 모방체는 몰폴리노 고리에 부착된 복소환식 염기를 갖는 연결된 몰폴리노 단위(몰폴리노 핵산)에 기반한다. 몰폴리노 핵산에서 몰폴리노 단량체 단위를 연결하는 다수의 연결기가 보고되었다. 비이온성 올리고머 화합물을 제공하기 위한 연결기의 한 가지 부류가 선택되었다. 비-이온성 몰폴리노계 올리고머 화합물은 세포 단백질과의 원치않는 상호작용을 가질 가능성이 더 적다. 몰폴리노계 폴리뉴클레오타이드는 세포 단백질과의 원치않는 상호작용을 형성할 가능성이 더 적은 올리고뉴클레오타이드의 비이온성 모방체이다(Dwaine A. Braasch and David R. Corey, *Biochemistry*, 2002, 41(14), 4503-4510). 몰폴리노계 폴리뉴클레오타이드는 미국 특허 제5,034,506호에 개시되어 있고, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입된다. 단량체 서브유닛을 결합하는 다양한 상이한 연결기를 갖는 폴리뉴클레오타이드의 몰폴리노 부류 내의 다양한 화합물이 제조되었다.

[0232] 추가적인 부류의 폴리뉴클레오타이드 모방체는 사이클로헥센일 핵산(CeNA)으로서 지칭된다. DNA/RNA 분자에 정량적으로 존재하는 푸라노스 고리는 사이클로헥센일 고리로 대체된다. CeNA DMT 보호된 포스포르아미다이트 단량체가 제조되었고, 고전적 포스포르아미다이트 화학 다음의 올리고머 화합물 합성을 위해 사용되었다. CeNA로 변형된 특정 위치를 갖는 완전히 변형된 CeNA 올리고머 화합물 및 올리고뉴클레오타이드가 제조되고, 연구되었다(문헌[Wang et al., *J. Am. Chem. Soc.*, 2000, 122, 8595-8602]을 참조하며, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨). 일반적으로 DNA 쇄에 대한 CeNA 단량체의 편입은 DNA/RNA 혼성체의 그의 안정성을 증가시킨다. CeNA 올리고아데닐레이트는 천연 복합체와 유사한 안정성을 갖는 RNA 및 DNA 상보체와의 복합체를 형성하였다. 천연 핵산 구조에 CeNA 구조를 혼입하는 연구는 용이한 입체배와 적용에 의해 진행되는 NMR 및 원이색법에 의해 나타낸다.

[0233] 추가적인 변형은 잠금 핵산(LNA)을 포함하며, 이때 2'-하이드록실기는 당 고리의 4' 탄소 원자에 연결됨으로써, 2'-C,4'-C-옥시메틸렌 결합을 형성하고, 이에 의해 이환식 당 모이어티를 형성한다. 결합은 2' 산소 원자와 4' 탄소 원자를 브리징하는 기인 메틸렌(-CH₂-)일 수 있되, n은 1 또는 2이다(문헌[Singh et al., *Chem. Commun.*, 1998, 4, 455-456], 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨). LNA 및 LNA 유사체는 상보성 DNA 및 RNA에 의한 매우 높은 듀플렉스 열 안정성(T_m=+3 내지 +10°C), 3'-핵산외부 분해 및 양호한 용해도 특성에 대한 안정성을 나타낸다. LNA를 함유하는 강력한 그리고 비독성의 안티센스 올리고뉴클레오타이드가 기재되었다(예를 들어, 문헌[Wahlestedt et al., *Proc. Natl. Acad. Sci. U.S.A.*, 2000, 97, 5633-5638] 참조, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨).

[0234] LNA 단량체 아데닌, 사이토신, 구아닌, 5-메틸-사이토신, 티민 및 유라실의 합성 및 제조는, 그들의 올리고머, 및 핵산 인식 특성과 함께 기재되었다(예를 들어, 문헌[Koshkin et al., *Tetrahedron*, 1998, 54, 3607-3630] 참조, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨). LNA 및 그의 제조는 또한 WO 98/39352 및 WO 99/14226뿐만 아니라 미국 특허 출원 제20120165514호, 제20100216983호, 제20090041809호, 제20060117410호, 제20040014959호, 제20020094555호 및 제20020086998호에 기재되어 있고, 이의 개시내용은 그들의 전문이 본 명세서에 참고로 포함된다.

[0235] 변형된 당 모이어티

[0236] 대상 핵산은 또한 하나 이상의 치환된 당 모이어티를 포함할 수 있다. 적합한 폴리뉴클레오타이드는 OH; F; O-, S-, 또는 N-알킬; O-, S- 또는 N-알켄일; O-, S- 또는 N-알킨일; 또는 O-알킬-O-알킬로부터 선택된 당 치환체기를 포함하되, 알킬, 알켄일 및 알킨일은 치환된 또는 비치환된 C₁ 내지 C₁₀ 알킬 또는 C₂ 내지 C₁₀ 알켄일 및 알킨일일 수 있다. O((CH₂)_nO)_mCH₃, O(CH₂)_nOCH₃, O(CH₂)_nNH₂, O(CH₂)_nCH₃, O(CH₂)_nONH₂ 및 O(CH₂)_nON((CH₂)_nCH₃)₂가 특히 적합하며, n 및 m은 1 내지 약 10이다. 다른 적합한 폴리뉴클레오타이드는 C₁ 내지 C₁₀ 저급 알킬, 치환된 저급 알킬, 알켄일, 알킨일, 알카릴, 아르알킬, O-알카릴 또는 O-아르알킬, SH, SCH₃, OCN, Cl, Br, CN, CF₃,

OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂, 헤테로사이클로알킬, 헤테로사이클로알카릴, 아미노알킬아미노, 폴리알킬아미노, 치환된 실릴, RNA 절단기, 리포터기, 삽입제(intercalator), 올리고뉴클레오타이드의 약물동력학 특성을 개선시키기 위한 기, 또는 올리고뉴클레오타이드의 약물동력학 특성을 개선시키기 위한 기 및 유사한 특성을 갖는 다른 치환체로부터 선택된 당 치환체기를 포함한다. 적합한 변형은 2'-메톡시에톡시(2'-O-CH₂CH₂OCH₃, 또한 2'-O-(2-메톡시에틸) 또는 2'-MOE로서 알려짐)(Martin et al., *Helv. Chim. Acta*, 1995, 78, 486-504, 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨) 즉, 알콕시알콕시기를 포함한다. 추가적인 적합한 변형은 본 명세서의 이하의 실시예에 기재된 바와 같은 2'-DMAOE로서도 알려진 2'-다이메틸아미노옥시에톡시, 즉, O(CH₂)₂ON(CH₃)₂ 기, 및 2'-다이메틸아미노에톡시에톡시(또한 당업계에서 2'-O-다이메틸-아미노-에톡시-에틸 또는 2'-DMAEOE로서 알려짐), 즉, 2'-O-CH₂-O-CH₂-N(CH₃)₂를 포함한다.

[0237]

다른 적합한 당 치환체기는 메톡시(-O-CH₃), 아미노프로폭시(-O CH₂ CH₂ CH₂NH₂), 알릴(-CH₂-CH=CH₂), -O-알릴(-O-CH₂-CH=CH₂) 및 플루오로(F)를 포함한다. 2'-당 치환체기는 아라비노(위) 위치 또는 리보(아래) 위치일 수 있다. 적합한 2'-아라비노 변형은 2'-F이다. 유사한 변형은 또한 올리고머 화합물 상의 다른 위치에서, 특히 3' 말단의 뉴클레오사이드 상의 당의 3' 위치에서 또는 2'-5' 연결된 올리고뉴클레오타이드 및 5' 말단 뉴클레오타이드의 5' 위치에서 만들어질 수 있다. 올리고머 화합물은 또한 펜토피라노실 당 대신 당 모방체, 예컨대 사이클로뷰틸 모이어티를 가질 수 있다.

[0238]

염기 변형 및 치환

[0239]

대상 핵산은 또한 핵염기(종종 본 명세서에서 단순히 "염기"로서 지칭됨) 변형 또는 치환을 포함할 수 있다. 본 명세서에서 사용되는 바와 같은, "비변형" 또는 "천연" 핵염기는 퓨린 염기 아데닌(A) 및 구아닌(G), 및 피리미딘 염기 티민(T), 사이토신(C) 및 유라실(U)을 포함한다. 변형된 핵염기는 다른 합성 및 천연 핵염기, 예컨대 5-메틸사이토신(5-me-C), 5-하이드록시메틸 사이토신, 잔틴, 하이드로잔틴, 2-아미노아데닌, 아데닌 및 구아닌의 6-메틸 및 기타 알킬 유도체, 아데닌 및 구아닌의 2-프로필 및 기타 알킬 유도체, 2-티오유라실, 2-티오티민 및 2-티오사이토신, 5-할로유라실 및 사이토신, 5-프로필(-C≡C-CH₃) 유라실 및 피리미딘 염기의 사이토신 및 다른 알킬인 유도체, 6-아조 유라실, 사이토신 및 티민, 5-유라실 (슈도유라실), 4-티오유라실, 8-할로, 8-아미노, 8-티올, 8-티오알킬, 8-하이드록실 및 다른 8-치환된 아데닌 및 구아닌, 5-할로, 특히 5-브로모, 5-트라이플루오로메틸 및 다른 5-치환된 유라실 및 사이토신, 7-메틸구아닌 및 7-메틸아데닌, 2-F-아데닌, 2-아미노-아데닌, 8-아자구아닌 및 8-아자아데닌, 7-데아자구아닌 및 7-데아자아데닌 및 3-데아자구아닌 및 3-데아자아데닌을 포함한다. 추가로 변형된 핵염기는 삼환식 피리미딘, 예컨대 페녹사진 사이티딘(1H-피리미도(5,4-b)(1,4)벤조사진-2(3H)-온), 페노티아진 사이티딘 (1H-피리미도(5,4-b)(1,4)벤조티아진-2(3H)-온), G-클램프, 예컨대 치환된 페녹사진 사이티딘(예를 들어, 9-(2-아미노에톡시)-H-피리미도(5,4-b)(1,4)벤조사진-2(3H)-온), 카바졸 사이티딘(2H-피리미도(4,5-b)인돌-2-온), 피리도인돌 사이티딘(H-피리도(3',2':4,5)피콜로(2,3-d)피리미딘-2-온)을 포함한다.

[0240]

복소환식 염기 모이어티는 또한 퓨린 또는 피리미딘 염기가 다른 헤테로사이클, 예를 들어, 7-데아자-아데닌, 7-데아자구아노신, 2-아미노피리딘 및 2-피리돈으로 대체되는 것을 포함할 수 있다. 추가로 핵염기는 미국 특허 제3,687,808호에 개시된 것, 문헌[*The Concise Encyclopedia Of Polymer Science And Engineering*, 페이지 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990]에 개시된 것, 문헌[Englisch et al., *Angewandte Chemie, International Edition*, 1991, 30, 613]에 의해 개시된 것, 및 문헌[Sanghvi, Y. S., *Chapter 15, Antisense Research and Applications*, 페이지 289-302, Crooke, S. T. and Lebleu, B., ed., CRC Press, 1993]에 의해 개시된 것을 포함하며; 이의 개시내용은 본 명세서에 그들의 전문이 참고로 편입된다. 특정 이들 핵염기는 올리고머 화합물의 결합 친화도를 증가시키는 데 유용하다. 이들은 5-치환된 피리미딘, 6-아자피리미딘 및 N-2, N-6 및 O-6 치환된 퓨린(2-아미노프로필아데닌, 5-프로필유라실 및 5-프로필사이토신을 포함)을 포함한다. 5-메틸사이토신 치환은 0.6 내지 1.2°C만큼 핵산 듀플렉스 안정성을 증가시키는 것으로 나타났고(Sanghvi et al., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, pp. 276-278; 이의 개시내용은 본 명세서에 그의 전문이 참고로 편입됨), 예를 들어, 2'-O-메톡시에틸 당 변형과 조합될 때, 적합한 염기 치환이다.

[0241]

접합체

[0242]

대상 핵산의 다른 가능한 변형은 올리고뉴클레오타이드의 활성, 세포 분포 또는 세포 흡수를 향상시키는 하나

이상의 모이어티 또는 접합체를 폴리뉴클레오타이드에 화학적으로 연결하는 것을 수반한다. 이들 모이어티 또는 접합체는 작용기, 예컨대 1차 또는 2차 하이드록실기에 공유 결합된 접합체기를 포함할 수 있다. 접합체기는 삽입제, 리포터 분자, 폴리아민, 폴리아마이드, 폴리에틸렌 글리콜, 폴리에터, 올리고머의 약력학적 특성을 향상시키는 기, 및 올리고머의 약물동력학적 특성을 향상시키는 기를 포함하지만, 이들로 제한되지 않는다. 적합한 접합기는 콜레스테롤, 지질, 인지질, 바이오틴, 페나진, 염산, 페나트리딘, 안트라퀴논, 아크리딘, 플루오레세인, 로다민, 쿠마린 및 염료를 포함하지만, 이들로 제한되지 않는다. 약력학적 특성을 향상시키는 기는 흡수를 개선시키고/시키거나, 분해에 대한 내성을 향상시키고/시키거나 표적 핵산과의 서열-특이적 혼성화를 강화시키는 기를 포함한다. 약물동태학적 특성을 향상시키는 기는 대상 핵산의 흡수, 분포, 대사 또는 배설을 개선시키는 기를 포함한다.

[0243] 접합 모이어티는 지질 모이어티, 예컨대 콜레스테롤 모이어티(Letsinger et al., *Proc. Natl. Acad. Sci. USA*, 1989, 86, 6553-6556), 콜산(Manoharan et al., *Bioorg. Med. Chem. Lett.*, 1994, 4, 1053-1060), 티오에터, 예를 들어, 핵실-S-트라이틸티올(Manoharan et al., *Ann. N.Y. Acad. Sci.*, 1992, 660, 306-309; Manoharan et al., *Bioorg. Med. Chem. Lett.*, 1993, 3, 2765-2770), 티오콜레스테롤(Oberhauser et al., *Nucl. Acids Res.*, 1992, 20, 533-538), 지방족 쇠, 예를 들어, 도데칸다이올 또는 운데실 잔기(Saison-Behmoaras et al., *EMBO J.*, 1991, 10, 1111-1118; Kabanov et al., *FEBS Lett.*, 1990, 259, 327-330; Svinarchuk et al., *Biochimie*, 1993, 75, 49-54), 인지질, 예를 들어, 다이-핵사데실-rac-글리세롤 또는 트라이에틸암모늄 1,2-다이-O-핵사데실-rac-글리세로-3-H-포스포네이트(Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654; Shea et al., *Nucl. Acids Res.*, 1990, 18, 3777-3783), 폴리아민 또는 폴리에틸렌 글리콜 쇠(Manoharan et al., *Nucleosides & Nucleotides*, 1995, 14, 969-973), 또는 아다만탄 아세트산(Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654), 팔미틸 모이어티(Mishra et al., *Biochim. Biophys. Acta*, 1995, 1264, 229-237), 또는 옥타데실아민 또는 핵실아미노-카보닐-옥시콜레스테롤 모이어티(Crooke et al., *J. Pharmacol. Exp. Ther.*, 1996, 277, 923-937)를 포함하지만, 이들로 제한되지 않는다.

[0244] 접합체는 "단백질 형질도입 도메인" 또는 PTD(또한 CPP - 세포 침투성 펩타이드로서 알려짐)를 포함하는데, 이는 지질 이중층, 마이셀, 세포막, 세포소기관 막 또는 소수포 막을 가로지르는 것을 용이하게 하는 폴리펩타이드, 폴리뉴클레오타이드, 탄수화물 또는 유기 또는 무기 화합물을 지칭할 수 있다. 거대 분자 및/또는 나노분자의 범위일 수 있는 다른 분자에 부착된 PTD는 분자가 막을 가로지르는 것, 예를 들어, 세포외 공간으로부터 세포내 공간 또는 사이토졸을 세포소기관(예를 들어, 핵) 내로 이동하는 것을 용이하게 한다. 일부 실시형태에서, PTD는 외인성 폴리뉴클레오타이드의 3' 단부에 공유적으로 연결된다. 일부 실시형태에서, PTD는 외인성 폴리뉴클레오타이드의 5' 단부에 공유적으로 연결된다. 예시적인 PTD는 최소 운데카펩타이드 단백질 형질도입 도메인(YGRKKRRQRRR; 서열번호 112를 포함하는 HIV-1 TAT의 잔기 47 내지 57에 대응); 세포 내로 유입을 지시하기에 충분한 다수의 알기닌을 포함하는 폴리알기닌 서열(예를 들어, 3, 4, 5, 6, 7, 8, 9, 10 또는 10-50 알기닌); VP22 도메인(Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); 초파리 안테나페디아(Antennapedia) 단백질 형질도입 도메인(Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); 절단된 인간 칼시토닌 펩타이드(Trehin et al. (2004) *Pharm. Research* 21:1248-1256); 폴리라이신(Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR 서열번호 113); 트랜스포탄 GWTLNSAGYLLGKINLKALAALAKKIL 서열번호 114); KALAWKALAKALAKALAKHLAKALAKALKEA 서열번호 115); 및 RQIKIWFQNRRMKWKK 서열번호 116)을 포함하지만, 이들로 제한되지 않는다. 예시적인 PTD는 YGRKKRRQRRR 서열번호 117), RKKRRQRRR 서열번호 118); 3개의 알기닌 잔기 내지 50개의 알기닌 잔기의 알기닌 동중중합체를 포함하지만, 이들로 제한되지 않고; 예시적인 PTD 도메인 아미노산 서열은 다음 중 어떤 것을 포함하지만, 이들로 제한되지 않는다: YGRKKRRQRRR 서열번호 119); RKKRRQRRR 서열번호 120); YARAAARQARA 서열번호 121); THRLPRRRRRR 서열번호 122); 및 GGRRARRRRRRR 서열번호 123). 일부 실시형태에서, PTD는 활성화 가능한 CPP(ACPP)이다(Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPP는 매칭 절단 가능한 링커를 통해 다음이온(예를 들어, Glu9 또는 "E9")에 연결된 다양이온 CPP(예를 들어, Arg9 또는 "R9")를 포함하는데, 이는 순전하를 거의 0으로 감소시키고, 이에 의해 세포에 대한 집착 및 흡수를 저해한다. 링커의 절단 시, 다음이온이 방출되며, 다알기닌 및 그의 고유한 부착성을 국소로 가리며, 따라서 막을 가로지르도록 ACPP를 "활성화시킨다".

[0245] 표적 세포 내로의 성분의 도입

[0246] 본 개시내용의 CasY 가이드 RNA(또는 이를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산) 및/또는 CasY 폴리펩타이드(또는 이를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산) 및/또는 본 개시내용의 CasY 융합 폴리펩타이드(또는 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산)

및/또는 공여자 폴리뉴클레오타이드(공여자 주형)는 임의의 다양한 잘 공지된 방법에 의해 숙주 세포 내로 도입될 수 있다.

[0247] 임의의 다양한 화합물 및 방법은 본 개시내용의 CasY 시스템을 표적 세포에 전달하기 위해 사용될 수 있다(예를 들어, CasY 시스템은 a) 본 개시내용의 CasY 폴리펩타이드 및 CasY 가이드 RNA; b) 본 개시내용의 CasY 폴리펩타이드, CasY 가이드 RNA, 및 공여자 주형 핵산; c) 본 개시내용의 CasY 융합 폴리펩타이드 및 CasY 가이드 RNA; d) 본 개시내용의 CasY 융합 폴리펩타이드, CasY 가이드 RNA, 및 공여자 주형 핵산; e) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 mRNA; 및 CasY 가이드 RNA; f) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 공여자 주형 핵산; g) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 mRNA; 및 CasY 가이드 RNA; h) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 공여자 주형 핵산; i) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; j) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 주형 핵산을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; k) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; l) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 주형 핵산을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; m) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; n) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; 및 공여자 주형 핵산; o) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; p) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; 및 공여자 주형 핵산; q) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, 제1 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 제2 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; 또는 r) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, 제1 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 제2 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; 또는 (a) 내지 (r) 중 하나의 일부 변형을 포함한다. 비제한적 예로서, 본 개시내용의 CasY 시스템은 지질과 조합될 수 있다. 다른 비제한적 예로서, 본 개시내용의 CasY 시스템은 입자와 조합될 수 있거나 또는 입자 내로 제형화될 수 있다.

[0248] 숙주 세포 내로 핵산을 도입하는 방법은 당업계에 공지되어 있고, 임의의 편리한 방법은 대상 핵산(예를 들어, 발현 작제물/벡터)을 표적 세포(예를 들어, 원핵 세포, 진핵 세포, 식물 세포, 동물 세포, 포유류 세포, 인간 세포 등) 내로 도입하는 데 사용될 수 있다. 적합한 방법은, 예를 들어, 바이러스 감염, 형질감염, 접합, 프로토플라스트 융합, 리포펙션, 전기천공법, 인산칼슘 침전, 폴리에틸렌이민(PEI)-매개 형질감염, DEAE-텍스트란 매개 형질감염, 리포좀-매개 형질감염, 입자총 기법, 인산칼슘 침전, 직접 미량주사법, 나노입자-매개 핵산 전달(예를 들어, 문헌[Panyam et., al Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023] 참조)등을 포함한다.

[0249] 일부 경우에, 본 개시내용의 CasY 폴리펩타이드는 CasY 폴리펩타이드를 암호화하는 핵산(예를 들어, mRNA, DNA, 플라스미드, 발현 벡터, 바이러스 벡터 등)으로서 제공된다. 일부 경우에, 본 개시내용의 CasY 폴리펩타이드는 단백질로서(예를 들어, 관련된 가이드 RNA 없이 또는 회합 가이드 RNA와 함께, 즉, 리보핵단백질로서) 직접 제공된다. 본 개시내용의 CasY 폴리펩타이드는 임의의 편리한 방법에 의해 세포 내로 도입될 수 있으며(세포에 제공되며); 이러한 방법은 당업자에게 공지되어 있다. 예시적 예로서, 본 개시내용의 CasY 폴리펩타이드는 (예를 들어, CasY 가이드 RNA 또는 CasY 가이드 RNA를 암호화하는 핵산과 함께 또는 이들 없이, 그리고 공여자 폴리뉴클레오타이드와 함께 또는 이것 없이) 세포 내로 직접 주사될 수 있다. 다른 예로서, 본 개시내용의 CasY 폴리펩타이드 및 CasY 가이드 RNA(RNP)의 수행된 복합체는 (예를 들어, 주사를 통해, 뉴클레오펙션을 통해; 하나 이상의 성분에 접합된, 예를 들어, CasY 단백질에 접합된, 가이드 RNA 등에 접합된, 본 개시내용의 CasY 폴리펩타이드 및 가이드 RNA에 접합된 단백질 형질도입 도메인(PTD)을 통해) 세포(예를 들어, 진핵 세포) 내로 도입될 수 있다.

[0250] 일부 경우에, 본 개시내용의 CasY 융합 폴리펩타이드(예를 들어, 융합 상대에 융합된 dCasY, 융합 상대에 융합

된 틈내기효소 CasY 등)는 CasY 폴리펩타이드를 암호화하는 핵산(예를 들어, mRNA, DNA, 플라스미드, 발현 벡터, 바이러스 벡터 등)으로서 제공된다. 일부 경우에, 본 개시내용의 CasY 융합 폴리펩타이드는 단백질로서(예를 들어, 관련된 가이드 RNA 없이 또는 회합 가이드 RNA와 함께, 즉, 리보핵단백질로서) 직접 제공된다. 본 개시내용의 CasY 융합 폴리펩타이드 임의의 편리한 방법에 의해 세포 내로 도입될 수 있으며(세포에 제공되며); 이러한 방법은 당업자에게 공지되어 있다. 예시적 예로서, 본 개시내용의 CasY 융합 폴리펩타이드는(예를 들어, CasY 가이드 RNA를 암호화하는 핵산과 함께 또는 이들 없이, 그리고 공여자 폴리뉴클레오타이드와 함께 또는 이것 없이) 세포 내로 직접 주사될 수 있다. 다른 예로서, 본 개시내용의 CasY 융합 폴리펩타이드 및 CasY 가이드 RNA(RNP)의 수행된 복합체는(예를 들어, 주사를 통해, 뉴클레오펙션을 통해; 하나 이상의 성분에 접합된, 예를 들어, CasY 융합 단백질에 접합된, 가이드 RNA 등에 접합된, 본 개시내용의 CasY 융합 폴리펩타이드 및 가이드 RNA에 접합된 단백질 형질도입 도메인(PTD)을 통해) 세포 내로 도입될 수 있다.

[0251] 일부 경우에, 핵산(예를 들어, CasY 가이드 RNA; 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산; 등)은 입자 내 또는 입자와 회합된 세포(예를 들어, 표적 숙주 세포) 및/또는 폴리펩타이드(예를 들어, CasY 폴리펩타이드; CasY 융합 폴리펩타이드)에 전달된다. 일부 경우에, 본 개시내용의 CasY 시스템은 입자 내 세포에 전달되거나 또는 입자와 회합된다. 용어 "입자" 및 "나노입자"는 적절하다면, 상호 호환적으로 사용될 수 있다. 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및/또는 CasY 가이드 RNA, 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 mRNA, 및 가이드 RNA를 포함하는 재조합 발현 벡터는 입자 또는 지질 외피를 이용하여 동시에 전달될 수 있고; 예를 들어, CasY 폴리펩타이드 및 예를 들어, 복합체(예를 들어, 리보뉴클레오타이드(RNP) 복합체)로서 CasY 가이드 RNA가 입자, 예를 들어, 지질 또는 리피도이드 및 친수성 중합체, 예를 들어, 양이온성 지질 및 친수성 중합체를 포함하는 전달 입자를 통해 전달될 수 있되, 예를 들어, 양이온성 지질은 1,2-다이올레오일-3-트라이메틸암모늄-프로판(DOTAP) 또는 1,2-다이테트라테칸오일-sn-글리세로-3-포스포콜린(DMPC)을 포함하고/하거나 친수성 중합체는 에틸렌 글리콜 또는 폴리에틸렌 글리콜(PEG)을 포함하고; 그리고/또는 입자는 추가로 콜레스테롤을 포함한다(예를 들어, 제형 1로부터의 입자=DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0; 제형 번호 2=DOTAP 90, DMPC 0, PEG 10, 콜레스테롤 0; 제형 번호 3=DOTAP 90, DMPC 0, PEG 5, 콜레스테롤 5). 예를 들어, 입자는 CasY 폴리펩타이드 및 CasY 가이드RNA가, 예를 들어, 1:1 몰비에서, 예를 들어, 실온에서, 예를 들어, 30분 동안, 예를 들어, 멸균, 무 뉴클레아제 1 x 인산염-완충 식염수(PBS) 중에서 함께 혼합되고; 그리고 별개로, 제형에 대해 적용 가능하다면 DOTAP, DMPC, PEG 및 콜레스테롤은 알코올, 예를 들어, 100% 에탄올 중에서 용해되고; 그리고, 두 용액은 함께 혼합되어 복합체를 함유하는 입자를 형성하는 다단계 공정을 이용하여 형성될 수 있다.

[0252] 본 개시내용의 CasY 폴리펩타이드(또는 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 mRNA; 또는 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터) 및/또는 CasY 가이드 RNA(또는 CasY 가이드 RNA를 암호화하는 하나 이상의 발현 벡터와 같은 핵산)는 입자 또는 지질 외피를 이용하여 동시에 전달될 수 있다. 예를 들어, 인지질 이중층 껍질로 뒤덮인 폴리(β -아미노 에스터)(PBAE) 코어를 갖는 생분해성 코어 껍질 구조의 나노입자를 사용할 수 있다. 일부 경우에, 자가 조립 생접착성 중합체에 기반한 입자/나노입자가 사용되고; 이러한 입자/나노입자는, 예를 들어, 뇌에 대해, 펩타이드의 경구 전달, 펩타이드의 정맥내 전달 및 펩타이드의 비강 전달에 적용될 수 있다. 다른 실시형태, 예컨대 경구 흡수 및 소수성 약물의 눈 전달이 또한 상정된다. 보호되고 질환 부위로 전달되는 조작된 중합체를 수반하는 분자 외피 기법이 사용될 수 있다. 약 5mg/kg의 용량은 다양한 인자, 예를 들어, 표적 조직에 따라서 단일 또는 다회 용량으로 사용될 수 있다.

[0253] 리피도이드 화합물(예를 들어, 미국 특허 출원 제20110293703호에 기재된 바와 같음)은 또한 폴리뉴클레오타이드의 투여에 유용하고, 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 전달하기 위해 사용될 수 있다(예를 들어, CasY 시스템은 a) 본 개시내용의 CasY 폴리펩타이드 및 CasY 가이드 RNA; b) 본 개시내용의 CasY 폴리펩타이드, CasY 가이드 RNA, 및 공여자 주형 핵산; c) 본 개시내용의 CasY 융합 폴리펩타이드 및 CasY 가이드 RNA; d) 본 개시내용의 CasY 융합 폴리펩타이드, CasY 가이드 RNA, 및 공여자 주형 핵산; e) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 mRNA; 및 CasY 가이드 RNA; f) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 공여자 주형 핵산; g) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 mRNA; 및 CasY 가이드 RNA; h) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 공여자 주형 핵산; i) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; j) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 주형 핵산을 암호화하는 뉴클레오타이드

드 서열을 포함하는 재조합 발현 벡터; k) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; l) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 주형 핵산을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; m) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; n) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; 및 공여자 주형 핵산; o) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; p) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; 및 공여자 주형 핵산; q) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, 제1 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 제2 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; 또는 r) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, 제1 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 제2 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; 또는 (a) 내지 (r) 중 하나의 일부 변형을 포함한다. 일 양상에서, 아미노알코올 리피도이드 화합물은 세포 또는 대상에 전달될 제제와 조합되어 마이크로입자, 나노입자, 리포솜 또는 마이셀을 형성한다. 아미노알코올 리피도이드 화합물은 다른 아미노알코올 리피도이드 화합물, 중합체(합성 또는 천연), 계면활성제, 콜레스테롤, 탄수화물, 단백질, 지질 등과 조합되어 입자를 형성할 수 있다. 이어서, 이들 입자는 선택적으로 약제학적 부형제와 조합되어 약제학적 조성물을 형성할 수 있다.

[0254] 폴리(베타-아미노 알코올)(PBAA)은 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 미국 특허 공개 제20130302401호는 조합적 중합을 이용하여 제조된 폴리(베타-아미노 알코올)(PBAA) 부류에 관한 것이다.

[0255] 당-기반 입자는, 예를 들어, 참고로 W02014118272(본 명세서에 참고로 편입됨) 및 문헌[Nair, J K et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961]에 기재된 바와 같은 GalNAc이 사용될 수 있고, 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하기 위해 사용될 수 있다.

[0256] 일부 경우에, 지질 나노입자(LNP)는 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 음으로 하전된 중합체, 예컨대 RNA는 낮은 pH 값(예를 들어, pH 4)에서 LNP에 부하될 수 있고, 여기서 이온화 가능한 지질은 양전하를 나타낸다. 그러나, 생리적 pH 값에서, LNP는 더 긴 순환 시간에 적합한 낮은 표면 전하를 나타낸다. 이온화 가능한 양이온성 지질의 4가지 중, 즉, 1,2-다이리놀레일-3-다이메틸암모늄-프로판(DLinDAP), 1,2-다이리놀레일옥시-3-N,N-다이메틸아미노프로판(DLinDMA), 1,2-다이리놀레일옥시-keto-N,N-다이메틸-3-아미노프로판(DLinKDMA), 및 1,2-다이리놀레일-4-(2-다이메틸아미노에틸)-[1,3]-다이옥솔란(DLinKC2-DMA)에 중점을 두었다. LNP의 제조는, 예를 들어, 문헌[Rosin et al. (2011) Molecular Therapy 19:1286-2200]에 기재되어 있다. 양이온성 지질 1,2-다이리놀레일-3-다이메틸암모늄-프로판(DLinDAP), 1,2-다이리놀레일옥시-3-N,N-다이메틸아미노프로판(DLinDMA), 1,2-다이리놀레일옥시케토-N,N-다이메틸-3-아미노프로판(DLinKDMA), 1,2-다이리놀레일-4-(2-다이메틸아미노에틸)-[1,3]-다이옥솔란(DLinKC2-DMA), (3-o-[2']-(메톡시폴리에틸렌글리콜 2000) 숙신오일]-1,2-다이미리스토일-sn-글리콜 (PEG-S-DMG), 및 R-3-[(.오메가.-메톡시-폴리(에틸렌 글리콜)2000) 카바모일]-1,2-다이미리스틸옥실프로필-3-아민(PEG-C-DOMG)이 사용될 수 있다. 핵산(예를 들어, CasY 가이드 RNA; 본 개시내용의 핵산 등)은 DLinDAP, DLinDMA, DLink-DMA 및 DLinKC2-DMA(40:10:40:10 몰비로 양이온성 지질:DSPC:CHOL: PEGS-DMG 또는 PEG-C-DOMG)를 함유하는 LNP에서 캡슐화될 수 있다. 일부 경우에, 0.2% SP-DiOC18이 혼입된다.

[0257] 구체 핵산(SNA(상표명)) 작제물 및 다른 나노입자(특히 금 나노입자)는 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 예를 들어, 문헌[Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler

et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19): 7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152(2013) 및 Mirkin, et al., Small, 10:186-192] 참조.

- [0258] RNA를 갖는 자기-조립 나노입자는 폴리에틸렌 글리콜(PEG)의 원위 단부에 부착된 Arg-Gly-Asp(RGD)펩타이드 리간드로 폐길화된 폴리에틸렌이민(PEI)에 의해 작체될 수 있다.
- [0259] 일반적으로, "나노입자"는 직경이 1000nm 미만인 입자의 입자를 지칭한다. 일부 경우에, 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산, 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 직경이 500nm 이하, 예를 들어, 25nm 내지 35nm, 35nm 내지 50nm, 50nm 내지 75nm, 75nm 내지 100nm, 100nm 내지 150nm, 150nm 내지 200nm, 200nm 내지 300nm, 300nm 내지 400nm, 또는 400nm 내지 500nm이다. 일부 경우에, 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 직경이 25nm 내지 200nm이다. 일부 경우에, 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산, 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 직경이 100nm 이하이고, 일부 경우에, 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산, 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 직경이 35nm 내지 60nm이다.
- [0260] 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산, 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용하기에 적합한 나노입자는 상이한 형태, 예를 들어, 고체 나노입자(예를 들어, 금속, 예컨대 은, 금, 철, 티타늄), 비-금속, 지질계 고체, 중합체), 나노입자의 현탁액 또는 이들의 조합물로 제공될 수 있다. 금속, 유전체 및 반도체 나노입자뿐만 아니라 혼성 구조(예를 들어, 코어-껍질 나노입자)가 제조될 수 있다. 반도체 물질로 만들어진 나노입자는 또한, 그들이 전자 에너지 준위의 양자화가 일어나기에 충분히 작다면(전형적으로 10nm 미만), 표지된 양자점일 수 있다. 이러한 나노규모 입자는 약물 담체 또는 영상화제로서의 생의학적 적용분야에서 사용되고, 본 개시내용에서 유사한 목적을 위해 채택될 수 있다.
- [0261] 반고체 및 연질 나노입자는 또한 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용하기에 적합하다. 반고체 특성의 원형 나노입자는 리포솜이다.
- [0262] 일부 경우에, 엑소솜은 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 엑소솜은 RNA 및 단백질을 수송하고, RNA를 뇌 및 다른 표적 기관에 전달할 수 있는 내인성 나노소포체이다.
- [0263] 일부 경우에, 리포솜은 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 리포솜은 내부 수성 구획을 둘러싸는 단층 또는 다층 지질 이중층 및 상대적으로 불침투성의 외부 친유성 인지질 이중층으로 구성된 구획 소수포 구조이다. 리포솜은 몇몇 상이한 유형의 지질로부터 이루어질 수 있지만, 그러나 인지질은 리포솜을 생성하는 데 가장 통상적으로 사용된다. 지질막이 수성 용액과 혼합될 때 리포솜 제형은 자발적이지만, 이는 또한 균질기, 초음파분해기 또는 압출 장치를 이용함으로써 진탕 형태로 힘을 적용함으로써 촉진될 수 있다. 이 구조 및 특성을 변형시키기 위해 몇몇 다른 첨가제가 리포솜에 첨가될 수 있다. 예를 들어, 리포솜 구조를 안정화시키기 위해 그리고 리포솜 내부 내용물의 누출을 방지하기 위해 콜레스테롤 또는 스펅고마이엘린 중 하나가 리포솜 혼합물에 첨가될 수 있다. 리포솜 제형은 주로 천연 인지질 및 지질, 예컨대 1,2-다이스테아릴-sn-글리세로-3-포스파티딜 콜린(DSPC), 스펅고마이엘린, 난황 포스파티딜콜린 및 모노시알로강글리오사이드를 포함할 수 있다.
- [0264] 안정한 핵산-지질 입자(SNALP)는 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용될 수 있다. SNALP 제형은 2:40:10:48 몰 백분율 비로 지질 3-N-[(메톡시폴리(에틸렌 글리콜) 2000) 카바모일]-1,2-다이미리스틸옥시-프로필아민(PEG-C-DMA), 1,2-다이리놀레일옥시-N,N-다이메틸-3-아미노프로판(DLinDMA), 1,2-다

이스테아로일-sn-글리세로-3-포스포콜린(DSPC) 및 콜레스테롤을 함유할 수 있다. SNALP 리포솜은 콜레스테롤/D-Lin-DMA/DSPC/PEG-C-DMA의 25:1 지질/siRNA 비 및 48/40/10/2 몰비를 이용하여 다이스테아로일포스파티딜콜린(DSPC), 콜레스테롤 및 siRNA와 함께 D-Lin-DMA 및 PEG-C-DMA를 제형화함으로써 제조될 수 있다. 얻어진 SNALP 리포솜은 크기가 약 80 내지 100nm일 수 있다. SNALP는 합성 콜레스테롤(시그마-알드리치(Sigma-Aldrich), 미국 미주리주 세인트루이스에 소재), 다이팔미토일포스파티딜콜린(아반티 폴라 인코포레이티드(Avanti Polar Lipids), 미국 앨리버마주 알라바스터에 소재), 3-N-[(w-메톡시 폴리(에틸렌 글리콜)2000)카바모일]-1,2-다이미리스틸옥시프로필아민 및 양이온성 1,2-다이리놀레일옥시-3-N,N-다이메틸아미노프로판올을 포함할 수 있다. SNALP는 합성 콜레스테롤(시그마-알드리치), 1,2-다이스테아로일-sn-글리세로-3-포스포콜린(DSPC; 아반티 폴라 리피드 인코포레이티드), PEG-cDMA, 및 1,2-다이리놀레일옥시-3-(N,N-다이메틸)아미노프로판(DLinDMA)을 포함할 수 있다.

[0265] 다른 양이온성 지질, 예컨대 아미노 지질 2,2-다이리놀레일-4-다이메틸아미노에틸-[1,3]-다이옥솔란(DLin-KC2-DMA)은 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산, 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하기 위해 사용될 수 있다. 다음의 지질 조성으로 수행된 소수포가 상정될 수 있다: 각각 몰비 40/10/40/10로 아미노 지질, 다이스테아로일포스파티딜콜린(DSPC), 콜레스테롤 및 (R)-2,3-비스(옥타데실옥시) 프로필-1-(메톡시 폴리(에틸렌 글리콜)2000)프로필카바메이트(PEG-지질), 및 대략 0.05(w/w)의 FVII siRNA/총 지질비. 70 내지 90nm 범위의 좁은 입자 크기 분포 및 0.11+.0.04 (n=56)의 낮은 다분산지수를 보장하기 위해, 입자는 가이드 RNA를 첨가하기 전에 80nm 막을 통해 3회까지 압출될 수 있다. 상당히 강한 아미노 지질 16을 함유하는 입자가 사용될 수 있으며, 이때 4가지 지질 성분 16, DSPC, 콜레스테롤 및 PEG-지질(50/10/38.5/1.5)의 몰비는 생체내 활성을 향상시키도록 추가로 최적화될 수 있다.

[0266] 지질은 본 개시내용의 CasY 시스템 또는 이의 성분(들) 또는 이를 암호화하는 핵산으로 제형화되어 지질 나노입자(LNP)를 형성할 수 있다. 적합한 지질은 DLin-KC2-DMA4, C12-200 및 코지질(colipid) 다이스테로일포스파티딜 콜린, 콜레스테롤 및 PEG-DMG를 포함하지만, 이들로 제한되지 않으며, 자발적 소수포 제형화 절차를 이용하여 본 개시내용의 CasY 시스템, 또는 이의 성분으로 제형화될 수 있다. 성분 몰비는 약 50/10/38.5/1.5(DLin-KC2-DMA 또는 C12-200/다이스테로일포스파티딜 콜린/콜레스테롤/PEG-DMG)일 수 있다.

[0267] 미국 특허 공개 제20130252281호 및 제20130245107호 및 제20130244279호에서 추가로 기재되는 것과 같이 PLGA 마이크로스피어에서 캡슐화된 본 개시내용의 CasY 시스템, 또는 이의 성분이 전달될 수 있다.

[0268] 과급된 단백질이 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용될 수 있다. 과급된 단백질은 흔히 않게 높은 양성 또는 음성 이론적 순전하를 갖는 조작된 또는 천연 유래 단백질의 부류이다. 초음성과 초양성으로 하전된 단백질은 둘 다 열적으로 또는 화학적으로 유도된 응집을 견뎌내는 능력을 나타낸다. 초양성으로 하전된 단백질은 또한 포유류 세포를 침투할 수 있다. 이들 단백질, 예컨대 플라스미드 DNA, RNA 또는 다른 단백질과 화물의 회합은 시험관내 그리고 생체내 둘 다에서 포유류 세포 내로 이들 거대분자의 기능적 전달을 가능하게 할 수 있다.

[0269] 세포 침투성 펩타이드(Cell Penetrating Peptide: CPP)가 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산 또는 본 개시내용의 CasY 시스템을 표적 세포에 전달하는 데 사용될 수 있다. CPP는 전형적으로 양으로 하전된 아미노산, 예컨대 라이신 또는 알기닌의 높은 상대적 존재비를 함유하거나 또는 극성/하전된 아미노산 및 비극성, 소수성 아미노산의 교번 패턴을 함유하는 서열을 갖는 아미노산 조성물을 가진다.

[0270] 이식 가능한 장치는 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산(예를 들어, CasY 가이드 RNA, CasY 가이드 RNA를 암호화하는 핵산, CasY 폴리펩타이드를 암호화하는 핵산, 공여자 주형 등), 또는 본 개시내용의 CasY 시스템을 표적 세포(예를 들어, 생체내 표적 세포, 여기서 표적 세포는 순환에서의 표적 세포, 조직 내 표적 세포, 기관 내 표적 세포임)에 전달하는 데 사용될 수 있다. 본 개시내용의 CasY 폴리펩타이드, 본 개시내용의 CasY 융합 폴리펩타이드, 본 개시내용의 RNP, 본 개시내용의 핵산, 또는 본 개시내용의 CasY 시스템을 표적 세포(예를 들어, 생체내 표적 세포, 여기서 표적 세포는 순환에서의 표적 세포, 조직 내 표적 세포, 기관 내 표적 세포 등임)에 전달하는 데 사용하기에 적합한 이식 가능한 장치는 CasY 폴리펩타이드, CasY 융합 폴리펩타이드, RNP, 또는 CasY 시스템(또는 이의 성분, 예를 들어, 본 개시내용의 핵산)을 포함하는 용기(예를 들어, 저장소, 매트릭스 등)를 포함할 수 있다.

- [0271] 적합한 이식 가능한 장치는 중합체 기질, 예컨대 매트릭스(예를 들어, 장치 본체로서 사용됨), 및 일부 경우에 추가적인 스캐폴드 물질, 예컨대 금속 또는 추가적인 중합체, 및 가시성 및 이미징을 향상시키기 위한 물질을 포함할 수 있다. 이식 가능한 전달 장치는 국소로 그리고 장기간에 걸쳐 방출을 제공하는 데 유리할 수 있고, 여기서 전달될 폴리펩타이드 및/또는 핵산은 표적 부위, 예를 들어, 세포외 기질(ECM), 종양을 둘러싸는 혈관, 질병 조직 등으로 직접 방출된다. 적합한 이식 가능한 전달 장치는 복강과 같은 내강에 대한 전달 및/또는 임의의 다른 유형의 투여에서 사용하기에 적합한 장치를 포함하며, 이때 생체 안정성 및/또는 분해성 및/또는 생체 흡수성 중합체 기질(예를 들어, 선택적으로 매트릭스일 수 있음)을 포함하는 약물 전달 시스템은 고정 또는 부착되지 않는다. 일부 경우에, 적합한 이식 가능한 약물 전달 장치는 분해성 중합체를 포함하되, 주요 방출 메커니즘은 벌크 침식(bulk erosion)이다. 일부 경우에, 적합한 이식 가능한 약물 전달 장치는 비분해성 또는 느리게 분해되는 중합체를 포함하되, 주요 방출 메커니즘은 벌크 침식보다는 확산성이며, 따라서 외부 부분은 막으로서 작용하고, 그의 내부 부분은 약물 저장소로서 작용하는데, 이는 사실상 장기간 동안(예를 들어, 약 1주 내지 약 몇 개월) 주변에 의해 영향받지 않는다. 상이한 중합체와 상이한 방출 메커니즘의 조합이 또한 선택적으로 사용될 수 있다. 농도 구배는 총 방출 기간의 상당한 기간 동안 효과적으로 유지될 수 있고, 따라서 확산 속도는 효과적으로 일정하다("제로 방식(zero mode)" 확산으로 칭해짐). 용어 "일정한"은 치료적 유효성의 하부역치 초과로 유지되지만, 여전히 선택적으로 초기 버스트 확산 속도를 특징으로 할 수 있고/있거나, 예를 들어, 특정 정도로 증가 및 감소하면서 변동할 수 있는 확산 속도를 의미한다. 확산 속도는 장기간 동안 그렇게 유지될 수 있고, 이는 치료적으로 유효한 기간, 예를 들어, 유효한 침묵 기간을 최적화하기 위해 특정 수준에 대해 일정하게 고려될 수 있다.
- [0272] 일부 경우에, 이식 가능한 전달 시스템은, 천연에서 화학물질이든 또는 대상체의 신체에서 효소 및 다른 인자로부터의 공격에 기인하든 뉴클레오타이드계 치료제가 분해되는 것을 막도록 설계된다.
- [0273] 장치의 이식을 위한 부위 또는 표적 부위는 최대 치료 효능을 위해 선택될 수 있다. 예를 들어, 전달 장치는 종양 환경, 또는 종양과 관련된 혈액 공급 내에서 또는 이의 근위에 이식될 수 있다. 표적 위치는, 예를 들어 하기일 수 있다: 1) 기저핵, 백질 및 회백질에서 파킨슨 또는 알츠하이머병에서와 같이 퇴행성 부위에서의 뇌; 2) 근위축성측색경화증(ALS)의 경우에 척추; 3) 자궁경부; 4) 활성 및 만성 염증성 관절염; 5) 건선의 경우에 피부; 7) 알레르기 효과에 대한 교감신경성 및 감각 신경 부위; 7) 뼈; 8) 급성 또는 만성 감염 부위; 9) 질내; 10) 내이--청각계, 내이의 미로, 전정계; 11) 기관내; 12) 심장내; 관상동맥, 심외막; 13) 비뇨기관 또는 방광; 14) 담도계; 15) 신장, 표, 비장을 포함하지만, 이들로 제한되지 않는 실질세포; 16) 림프절; 17) 침샘; 18) 턱탈검(dental gum); 19) 관절강내(관절내); 20) 안구내; 21) 뇌조직; 22) 뇌실; 23) 복강을 포함하는 강(예를 들어, 제한되는 것은 아니지만, 난소암에 대해); 24) 식도내; 및 25) 직장내; 및 26) 혈관구조내.
- [0274] 삽입, 예컨대 이식 방법은 선택적으로 변형 없이 또는 대안적으로 선택적으로 이러한 방법에서 중요하지 않은 변형만으로 다른 유형의 조직 이식을 위해 그리고/또는 삽입을 위해 그리고/또는 조직 샘플링을 위해 선택적으로 이미 사용될 수 있다. 이러한 방법은 선택적으로 브라키테라피 방법, 생검, 초음파와 함께 그리고/또는 초음파 없이 내시경검사, 예컨대 뇌 조직 내로 주축성 방법, 관절, 복부 기관, 방광벽 및 체강 내로 복강경의 이식을 포함하는 복강경검사를 포함하지만, 이들로 제한되지 않는다.
- [0275] **변형된 숙주 세포**
- [0276] 본 개시내용은 본 개시내용의 CasY 폴리펩타이드를 포함하는 변형된 세포 및/또는 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 제공한다. 본 개시내용은 본 개시내용의 CasY 폴리펩타이드를 포함하는 변형된 세포를 제공하며, 여기서 변형된 세포는 본 개시내용의 CasY 폴리펩타이드를 정상적으로 포함하지 않는 세포이다. 본 개시내용은 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 포함하는 변형된 세포(예를 들어, 유전자 변형된 세포)를 제공한다. 본 개시내용은 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 mRNA로 유전자 변형된 유전자 변형 세포를 제공한다. 본 개시내용은 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 mRNA로 유전자 변형된 유전자 변형 세포를 제공한다. 본 개시내용은: a) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; 및 b) 본 개시내용의 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제조합 발현 벡터로 유전자 변형된 유전자 변형 세포를 제공한다. 본 개시내용은: a) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; b) 본 개시내용의 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열; 및 c) 공여자 주형을 암호화하는 뉴클레오타이드 서열을 포함하는 제조합 발현 벡터로 유전자 변형된 유전자 변형 세포를 제공한다.

- [0277] 본 개시내용의 CasY 폴리펩타이드 및/또는 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 및/또는 본 개시내용의 CasY 가이드 RNA에 대한 수용자로서 작용하는 세포는, 예를 들어, 시험관내 세포; 생체내 세포; 생체의 세포; 1차 세포; 암 세포; 동물 세포; 식물 세포; 조류 세포; 진균 세포 등을 포함하는 임의의 다양한 세포일 수 있다. 본 개시내용의 CasY 폴리펩타이드 및/또는 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 및/또는 본 개시내용의 CasY 가이드 RNA에 대한 수용자로서 작용하는 세포는 "숙주 세포" 또는 "표적 세포"로서 지칭된다. 숙주 세포 또는 표적 세포는 본 개시내용의 CasY 시스템의 수용자일 수 있다. 숙주 세포 또는 표적 세포는 본 개시내용의 CasY RNP의 수용자일 수 있다. 숙주 세포 또는 표적 세포는 본 개시내용의 CasY 시스템의 단일 성분의 수용자일 수 있다.
- [0278] 세포(표적 세포)의 비제한적 예는 원핵 세포, 진핵 세포, 박테리아 세포, 고세균 세포, 단세포 진핵 유기체의 세포, 원생동물 세포, 식물로부터의 세포(예를 들어, 식물 작물, 과일, 채소, 곡물, 대두, 옥수수, 마이스, 밀, 종자, 토마토, 벼, 카사바, 사탕수수, 호박, 건초, 감자, 목화, 대마초, 담배, 하룻, 구과 식물, 겉씨식물, 속씨식물, 양치식물, 석송, 뿌리끼리, 우산이끼류, 선대식물, 쌍떡잎식물, 외떡잎식물 등), 조류 세포, (예를 들어, 보트리오크쿠스 브라우니(*Botryococcus braunii*), 칼라마이도모나스 레인하르티(*Chlamydomonas reinhardtii*), 난노클로롭시스 가디타나(*Nannochloropsis gaditana*), 클로렐라 피레노이도사(*Chlorella pyrenoidosa*), 모자반속(*Sargassum patens*, *C. agardh*) 등), 해초(예를 들어, 켈프) 진균 세포(예를 들어, 효모 세포, 버섯으로부터의 세포), 동물 세포, 무척추동물(예를 들어, 과일파리, 강장동물, 극피동물, 선충 등)으로부터의 세포, 척추동물(예를 들어, 어류, 양서류, 파충류, 조류, 포유류)로부터의 세포, 포유류(예를 들어, 유제류(예를 들어, 돼지, 소, 염소, 양)로부터의 세포; 설치류(예를 들어, 래트, 마우스); 비-인간 영장류; 인간; 고양이과(예를 들어, 고양이); 개과(예를 들어, 개); 등) 등을 포함한다. 일부 경우에, 세포는 천연 유기체로부터 유래되지 않은 세포이다(예를 들어, 세포는 합성으로 만들어진 세포일 수 있고; 또한 인공 세포로서 지칭된다).
- [0279] 세포는 시험관내 세포일 수 있다(예를 들어, 확립된 배양 세포주). 세포는 생체의 세포일 수 있다(개체로부터의 배양 세포). 세포는 생체내 세포(예를 들어, 개체 내 세포)일 수 있다. 세포는 단리된 세포일 수 있다. 세포는 유기체 내부의 세포일 수 있다. 세포는 유기체일 수 있다. 세포는 세포 배양물 내 세포일 수 있다(예를 들어, 시험관내 세포 배양물). 세포는 세포 수집물 중 하나일 수 있다. 세포는 원핵 세포이거나 또는 원핵 세포로부터 유래될 수 있다. 세포는 박테리아 세포일 수 있거나 또는 박테리아 세포로부터 유래될 수 있다. 세포는 고세균 세포일 수 있거나 또는 고세균 세포로부터 유래될 수 있다. 세포는 진핵 세포일 수 있거나 또는 진핵 세포로부터 유래될 수 있다. 세포는 식물 세포일 수 있거나 또는 식물 세포로부터 유래될 수 있다. 세포는 동물 세포일 수 있거나 또는 동물 세포로부터 유래될 수 있다. 세포는 무척추동물 세포일 수 있거나 또는 무척추동물 세포로부터 유래될 수 있다. 세포는 척추동물 세포일 수 있거나 또는 척추동물 세포로부터 유래될 수 있다. 세포는 포유류 세포일 수 있거나 또는 포유류 세포로부터 유래될 수 있다. 세포는 설치류 세포일 수 있거나 또는 설치류 세포로부터 유래될 수 있다. 세포는 인간 세포일 수 있거나 또는 인간 세포로부터 유래될 수 있다. 세포는 미생물 세포일 수 있거나 또는 미생물 세포로부터 유래될 수 있다. 세포는 진균 세포일 수 있거나 또는 진균 세포로부터 유래될 수 있다. 세포는 곤충 세포일 수 있다. 세포는 절지동물 세포일 수 있다. 세포는 원생동물 세포일 수 있다. 세포는 유충 세포일 수 있다.
- [0280] 적합한 세포는 줄기 세포(예를 들어, 배아 줄기(ES) 세포, 유도 만능 줄기(iPS) 세포; 생식세포(예를 들어, 난모세포, 정자, 난원세포, 정원세포 등); 체세포, 예를 들어, 섬유아세포, 희소돌기신경교, 교세포, 조혈모세포, 뉴런, 근육 세포, 골세포, 간세포, 췌장 세포 등을 포함한다.
- [0281] 적합한 세포는 인간 배아 줄기 세포, 태아 심장 근육 세포, 근육섬유아세포, 간충직 줄기 세포, 자가이식 확장된 심장 근육 세포, 지방세포, 전분화 세포, 만능 세포, 혈액 줄기 세포, 근아세포, 성체 줄기 세포, 골수 세포, 간충직 세포, 배아 줄기 세포, 실질 세포, 상피세포, 내피세포, 중피세포, 섬유아세포, 조골세포, 연골세포, 외인성 세포, 내인성 세포, 줄기 세포, 조혈모세포, 골수 유래 전구체 세포, 심근세포, 골격 세포, 태아 세포, 미분화 세포, 다능 전구체 세포, 단분화능 전구체 세포, 단핵구, 심장 근아세포, 골격 근아세포, 대식세포, 모세혈관 내피세포, 이종 세포, 동종 세포 및 출산후 줄기 세포를 포함한다.
- [0282] 일부 경우에, 세포는 면역세포, 뉴런, 상피세포 및 내피세포 또는 줄기 세포이다. 일부 경우에, 면역세포는 T 세포, B 세포, 단핵구, 자연살해 세포, 수지상 세포 또는 대식세포이다. 일부 경우에, 면역세포는 세포독성 T 세포이다. 일부 경우에, 면역세포는 헬퍼 T 세포이다. 일부 경우에, 면역세포는 조절 T 세포(Treg)이다.
- [0283] 일부 경우에, 세포는 줄기 세포이다. 줄기 세포는 성체 줄기 세포를 포함한다. 성체 줄기 세포는 또한 체세포

줄기 세포로서 지칭된다.

- [0284] 성체 줄기 세포는 분화된 조직 내에 존재하지만, 자기 재생 특성 및 다중 세포 유형, 보통 줄기 세포가 발견된 전형적인 조직의 세포 유형이 생기게 하는 능력을 보유한다. 체세포 줄기 세포의 수많은 예는 당업자에게 공지되어 있으며, 근육 줄기 세포; 조혈모세포; 상피 줄기 세포; 신경 줄기 세포; 간충직 줄기 세포; 유방 줄기 세포; 장 줄기 세포; 중배엽 줄기 세포; 내피 줄기 세포; 후각 줄기 세포; 신경 능선 줄기 세포 등을 포함한다.
- [0285] 관심 대상의 줄기 세포는 포유류 줄기 세포를 포함하며, 용어 "포유류"는 인간; 비-인간 영장류; 가축 및 농장 동물; 및 동물원, 실험실, 운동 또는 반려 동물, 예컨대 개, 말, 고양이, 소, 마우스, 래트, 토끼 등을 포함하는 포유류로서 분류된 임의의 동물을 지칭한다. 일부 경우에, 줄기 세포는 인간 줄기 세포이다. 일부 경우에, 줄기 세포는 설치류(예를 들어, 마우스; 래트) 줄기 세포이다. 일부 경우에, 줄기 세포는 비-인간 영장류 줄기 세포이다.
- [0286] 줄기 세포는 하나 이상의 줄기 세포 마커, 예를 들어, SOX9, KRT19, KRT7, LGR5, CA9, FXYD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17 및 PPARGC1A를 발현시킬 수 있다.
- [0287] 일부 실시형태에서, 줄기 세포는 조혈모세포(HSC)이다. HSC는 골수, 혈액, 제대혈, 태아간 및 난황낭으로부터 분리될 수 있는 중배엽-유래 세포이다. HSC는 $CD34^{+}$ 및 $CD3^{-}$ 으로서 특성규명된다. HSC는 생체내 적혈구, 호중성-대식세포, 거핵구 및 림프계 조혈세포 계통을 재생시킬 수 있다. 시험관내에서, HSC는 적어도 일부의 자기-재생 세포 분열을 겪도록 유도될 수 있고, 생체내에서 알 수 있는 바와 같이 동일한 계통으로 분화되도록 유도될 수 있다. 이렇게 해서, HSC는 적혈구 세포, 거핵구, 호중구, 대식세포 및 림프구 세포 중 하나 이상으로 분화되도록 유도될 수 있다.
- [0288] 다른 실시형태에서, 줄기 세포는 신경 줄기 세포(NSC)이다. 신경 줄기 세포(NSC)는 뉴런, 및 교세포(희소돌기신경교 및 성상교세포를 포함)로 분화할 수 있다. 신경 줄기 세포는 다회 분열할 수 있는 다능성 줄기 세포이고, 특정 조건 하에서 신경아세포 또는 아교모세포, 예를 들어, 각각 뉴런 및 교세포 중 하나 이상의 유형이 되도록 기여하는 세포일 수 있는 신경 줄기 세포 또는 신경 전구체 세포인 딸 세포를 생성할 수 있다. NSC를 얻는 방법은 당업계에 공지되어 있다.
- [0289] 다른 실시형태에서, 줄기 세포는 간충직 줄기 세포(MSC)이다. MSC는 본래 배아 중배엽으로부터 유래되고, 성체 골수로부터 분리되며, 근육, 뼈, 연골, 지방, 골수 기질 및 힘줄을 형성하도록 분화될 수 있다. MSC를 분리시키는 방법은 당업계에 공지되어 있으며; MSC를 얻기 위한 임의의 공지된 방법이 사용될 수 있다. 예를 들어, 인간 MSC의 단리를 기재하는 미국 특허 제5,736,396호 참조.
- [0290] 세포는 일부 경우에 식물 세포이다. 식물 세포는 외떡잎식물의 세포일 수 있다. 세포는 쌍떡잎식물의 세포일 수 있다.
- [0291] 일부 경우에, 세포는 식물 세포이다. 예를 들어, 세포는 주요 농작물, 예를 들어, 보리, 콩류(건조 식용), 카놀라, 옥수수, 목화(피마), 목화(업랜드), 아마씨, 건초(알팔파), 건초(비-알팔파), 귀리, 땅콩, 벼, 수수, 대두, 사탕무, 사탕수수, 해바라기(오일), 해바라기(비-오일), 고구마, 담배(벌리), 담배(연관처리(Flue-cured)), 토마토, 밀(듀럼), 밀(봄), 밀(겨울) 등의 세포일 수 있다. 다른 예로서, 세포는, 예를 들어, 알팔파, 알로에 잎, 에로루트, 벚꽃, 아티초크, 아스파라거스, 죽순, 바나나꽃, 콩나물, 콩류, 비트트릅, 근대, 비타멜론, 청경채, 브로콜리, 브로콜리 라베(엽경채), 방울다다기 양배추, 양배추, 방울 양배추, 가재발 선인장(노팔), 호리병박 열매, 카르둔, 당근, 콜리플라워, 셀러리, 차요테, 중국 아티초크(두루미냉이), 중국 양배추, 중국 셀러리, 중국 차이브, 채심, 국화잎(통호), 콜라드 그린, 옥수수대, 단옥수수, 오이, 대권, 단테리온 그린, 타로토란, 다우 무에(dau mue)(피 팁(pea tip)), 동과(겨울멜론), 가지, 엔다이브, 에스카롤, 선수 장식 양치식물, 갯류 식물, 치커리, 갯 (중국 겨자), 가일론(gailon), 갈랑가(시암, 태국 생강), 마늘, 생강 뿌리, 우엉, 나물, 하노버 셀러드 나물, 후아우존틀, 예루살렘 아티초크, 히카마, 케일 나물, 콜라비, 명아주(젤리페), 상추(비브), 상추(보스턴), 상추(보스턴 레드), 상추(녹색잎), 상추(아이스버그), 상추(콜라로사), 상추(오크리프 - 녹색), 상추(오크리프 - 적색), 상추(가공처리), 상추(적색잎), 상추(로메인), 상추(루비 로메인), 상추(러시안 레드 겨자), 린콕, 로 북, 긴콩류, 연근, 마하, 용설란 (아가베) 잎, 말랑가, 메스폴랭 믹스, 겨자와, 모압(수세미오이), 무(moo), 모쿠아(moqua)(동과), 버섯, 겨자, 참마, 오크라, 웅초이, 그린 어니언, 오포(opo)(콩 스쿼시), 장식용 옥수수, 관상용 호박, 파슬리, 파스닙, 완두콩, 후추(벨형), 후추, 호박, 라디키오, 무순, 무, 레이프 그린(rape greens), 레이프 그린, 대황, 로메인(베이비 레드), 루타베가, 함초(씨빈), 신쿠아(앵글드/줄기가 선 수세미), 시금치, 스쿼시, 짬 더미, 사탕수수, 고구마, 근대, 타마린드, 타로, 타로잎, 타로

슈트, 텃소이, 테페구아헤(tepeguaje)(구아헤), 빔바, 토마틸로, 토마토, 토마토(체리), 토마토(포도형), 토마토(자두형), 강황, 순무, 어린잎 나물, 순무, 마름, 양파, 양스(네임스), 유초이, 유카(카사바) 등을 포함하지만, 이들로 제한되지 않는 채소 작물의 세포이다.

[0292] 세포는 일부 경우에 절지동물 세포이다. 예를 들어, 세포는, 예를 들어, 협각류, 다지류, 헥시포디아(*Hexipodia*), 거미강, 곤충강, 돌좀목, 좀류, 구시아강, 하루살이목, 잠자리목, 잠자리하목, 실잠자리아목, 신시류, 외시류, 강도래목, 흰개미붙이목, 메뚜기, 민벌레목, 집게벌레목, 바퀴목, 귀뚜라미붙이목, 갈르와벌레과, 만토파스마토데과, 대벌레목, 바퀴목, 흰개미목, 사마귀목, 책좀, 다듬이벌레목, 총채벌레목, 이목, 반시목, 내시류 또는 완전변태류, 막시류, 초시류, 부채벌레목, 약대벌레목, 뱀잠자리목, 풀잠자리목, 밀드리목, 벼룩목, 파리목, 날도래류 또는 인시목의 아목, 과, 아과, 군, 하위군, 또는 종의 세포일 수 있다.

[0293] 세포는 일부 경우에 곤충 세포이다. 예를 들어, 일부 경우에, 세포는 모기, 메뚜기, 노린재류, 파리, 벼룩, 벌, 말벌, 개미, 이, 나방 또는 딱정벌레의 세포이다.

[0294] **키트**

[0295] 본 개시내용은 본 개시내용의 CasY 시스템, 또는 본 개시내용의 CasY 시스템의 성분을 포함하는 키트를 제공한다.

[0296] 본 개시내용의 키트는 a) 본 개시내용의 CasY 폴리펩타이드 및 CasY 가이드 RNA; b) 본 개시내용의 CasY 폴리펩타이드, CasY 가이드 RNA, 및 공여자 주형 핵산; c) 본 개시내용의 CasY 융합 폴리펩타이드 및 CasY 가이드 RNA; d) 본 개시내용의 CasY 융합 폴리펩타이드, CasY 가이드 RNA, 및 공여자 주형 핵산; e) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 mRNA; 및 CasY 가이드 RNA; f) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 공여자 주형 핵산; g) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 mRNA; 및 CasY 가이드 RNA; h) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 공여자 주형 핵산; i) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; j) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 주형 핵산을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; k) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; l) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 공여자 주형 핵산을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; m) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; n) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; 및 공여자 주형 핵산; o) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; p) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 제1 재조합 발현 벡터, 및 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 제2 재조합 발현 벡터; 및 공여자 주형 핵산; q) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, 제1 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 제2 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; 또는 r) 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열, 제1 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열, 및 제2 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터; 또는 (a) 내지 (r) 중 하나의 일부 변형을 포함할 수 있다.

[0297] 본 개시내용의 키트는 a) 본 개시내용의 CasY 시스템, 또는 본 개시내용의 CasY 시스템의 상기 기재한 바와 같은 성분; 및 b) 1종 이상의 추가적인 시약, 예를 들어, i) 완충제; ii) 프로테아제 저해제; iii) 뉴클레아제 저해제; iv) 검출 가능한 표지를 개발하거나 또는 시각화하는 데 필요한 시약; v) 양성 및/또는 음성 대조군 표적 DNA; vi) 양성 및/또는 음성 대조군 CasY 가이드 RNA 등을 포함할 수 있다. 본 개시내용의 키트는 a) 상기 기재한 바와 같은 본 개시내용의 CasY 시스템의 성분, 또는 본 개시내용의 CasY 시스템; 및 b) 치료제를 포함할 수 있다.

[0298] 본 개시내용의 키트는 a) 표적 핵산에서 표적 뉴클레오타이드 서열에 혼성화하는 CasY 가이드 RNA의 일부를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 삽입하기 위한 삽입 부위; 및 b) CasY 가이드 RNA의 CasY-결

합 부분을 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터를 포함할 수 있다. 본 개시내용의 키트는 a) 표적 핵산에서 표적 뉴클레오타이드 서열에 혼성화하는 CasY 가이드 RNA의 일부를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 삽입하기 위한 삽입 부위; b) CasY 가이드 RNA의 CasY-결합 부분을 암호화하는 뉴클레오타이드 서열 및 c) 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 재조합 발현 벡터를 포함할 수 있다.

[0299] **효용**

[0300] 본 개시내용의 CasY 폴리펩타이드, 또는 본 개시내용의 CasY 융합 폴리펩타이드는 다양한 방법에서의(예를 들어, CasY 가이드 RNA와 조합하여 그리고 일부 경우에 추가로 공여자 주형과 조합하여) 용도를 발견한다. 예를 들어, 본 개시내용의 CasY 폴리펩타이드는 (i) 표적 핵산(DNA 또는 RNA; 단일 가닥 또는 이중 가닥)을 변형(예를 들어, 절단, 예를 들어, 틸내기; 메틸화 등); (ii) 표적 핵산의 전사를 조절; (iii) 표적 핵산을 표지; (iv) (예를 들어, 단리, 표지, 영상화, 추적 등의 목적을 위해) 표적 핵산에 결합; (v) 표적 핵산과 회합된 폴리펩타이드(예를 들어, 히스톤)를 변형 등을 위해 사용될 수 있다. 따라서, 본 개시내용은 표적 핵산을 변형시키는 방법을 제공한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을: a) 본 개시내용의 CasY 폴리펩타이드; 및 b) 하나 이상의(예를 들어, 2개의) CasY 가이드 RNA와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 a) 본 개시내용의 CasY 폴리펩타이드; b) CasY 가이드 RNA; 및 c) 공여자 핵산(예를 들어, 공여자 주형)과 접촉시키는 단계를 포함한다. 일부 경우에, 접촉시키는 단계는 시험관내 세포에서 수행된다. 일부 경우에, 접촉시키는 단계는 생체내 세포에서 수행된다. 일부 경우에, 접촉시키는 단계는 생체의 세포에서 수행된다.

[0301] CasY 폴리펩타이드를 사용하는 방법은 (회합된 CasY 가이드 RNA에 의해 거기서 표적화한 덕분에) 표적 핵산 내 특정 영역에 대한 CasY 폴리펩타이드의 결합을 포함하기 때문에, 상기 방법은 일반적으로 본 명세서에서 결합 방법(예를 들어, 표적 핵산의 결합 방법)으로서 지칭된다. 그러나, 일부 경우에, 결합 방법이 표적 핵산의 결합에 불과할 수 있지만, 다른 경우에, 상기 방법은 상이한 최종 결과를 가질 수 있다는 것이 이해되어야 한다(예를 들어, 상기 방법은 표적 핵산의 변형, 예를 들어, 절단/메틸화/등, 표적 핵산으로부터의 전사의 조절; 표적 핵산의 번역의 조절; 게놈 편집; 표적 핵산과 회합된 단백질의 조절; 표적 핵산의 단리 등을 초래할 수 있다).

[0302] 적합한 방법의 예에 대해, 예를 들어, 문헌[Jinek et al., Science. 2012 Aug 17;337(6096):816-21; Chylinski et al., RNA Biol. 2013 May;10(5):726-37; Ma et al., Biomed Res Int. 2013;2013:270805; Hou et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9; Jinek et al., Elife. 2013;2:e00471; Pattanayak et al., Nat Biotechnol. 2013 Sep;31(9):839-43; Qi et al., Cell. 2013 Feb 28;152(5):1173-83; Wang et al., Cell. 2013 May 9;153(4):910-8; Auer et al., Genome Res. 2013 Oct 31; Chen et al., Nucleic Acids Res. 2013 Nov 1;41(20):e19; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; Cho et al., Genetics. 2013 Nov;195(3):1177-80; DiCarlo et al., Nucleic Acids Res. 2013 Apr;41(7):4336-43; Dickinson et al., Nat Methods. 2013 Oct;10(10):1028-34; Ebina et al., Sci Rep. 2013;3:2510; Fujii et al., Nucleic Acids Res. 2013 Nov 1;41(20):e187; Hu et al., Cell Res. 2013 Nov;23(11):1322-5; Jiang et al., Nucleic Acids Res. 2013 Nov 1;41(20):e188; Larson et al., Nat Protoc. 2013 Nov;8(11):2180-96; Mali et al., Nat Methods. 2013 Oct;10(10):957-63; Nakayama et al., Genesis. 2013 Dec;51(12):835-43; Ran et al., Nat Protoc. 2013 Nov;8(11):2281-308; Ran et al., Cell. 2013 Sep 12;154(6):1380-9; Upadhyay et al., G3 (Bethesda). 2013 Dec 9;3(12):2233-8; Walsh et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15514-5; Xie et al., Mol Plant. 2013 Oct 9; Yang et al., Cell. 2013 Sep 12;154(6):1370-9]; 및 미국 특허 및 특허 출원 8,906,616; 8,895,308; 8,889,418; 8,889,356; 8,871,445; 8,865,406; 8,795,965; 8,771,945; 8,697,359; 20140068797; 20140170753; 20140179006; 20140179770; 20140186843; 20140186919; 20140186958; 20140189896; 20140227787; 20140234972; 20140242664; 20140242699; 20140242700; 20140242702; 20140248702; 20140256046; 20140273037; 20140273226; 20140273230; 20140273231; 20140273232; 20140273233; 20140273234; 20140273235; 20140287938; 20140295556; 20140295557; 20140298547; 20140304853; 20140309487; 20140310828; 20140310830; 20140315985; 20140335063; 20140335620; 20140342456; 20140342457; 20140342458; 20140349400; 20140349405; 20140356867; 20140356956; 20140356958; 20140356959; 20140357523; 20140357530; 20140364333; 및 20140377868을 참조하며; 이들 각각은 본 명세서에 그들의 전문이 참고로 포함된다.

[0303] 예를 들어, 본 개시내용은 표적 핵산의 절단 방법; 표적 핵산의 편집 방법; 표적 핵산으로부터 전사를 조절하는 방법; 표적 핵산을 단리시키는 방법, 표적 핵산을 결합시키는 방법, 표적 핵산의 영상화 방법, 표적 핵산을 변

형시키는 방법 등을 제공한다(그러나 이들로 제한되지 않는다).

[0304] 본 명세서에서 사용되는 바와 같은, 예를 들어, CasY 폴리펩타이드와 또는 CasY 융합 폴리펩타이드 등과 "표적 핵산을 접촉시키다" 및 "표적 핵산을 접촉시키는"이라는 용어/어구는 표적 핵산을 접촉시키기 위한 모든 방법을 포함한다. 예를 들어, CasY 폴리펩타이드는 단백질, RNA(CasY 폴리펩타이드를 암호화) 또는 DNA(CasY 폴리펩타이드를 암호화)로서 세포에 제공될 수 있는 반면; CasY 가이드 RNA는 가이드 RNA로서 또는 가이드 RNA를 암호화하는 핵산으로서 제공될 수 있다. 이렇게 해서, 예를 들어, 세포에서(예를 들어, 시험관내 세포 내에서, 생체내 세포 내에서, 생체의 세포 내에서) 방법을 수행할 때, 표적 핵산을 접촉시키는 단계를 포함하는 방법은 그들의 활성/최종 상태에서(예를 들어, CasY 폴리펩타이드에 대한 단백질(들)의 형태로; CasY 융합 폴리펩타이드에 대한 단백질의 형태로; RNA 일부 경우에 가이드 RNA에 대한 RNA의 형태로) 임의의 또는 모든 성분의 세포 내로의 도입을 포함하고, 그리고 또한 하나 이상의 성분을 암호화하는 하나 이상의 핵산(예를 들어, CasY 폴리펩타이드 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열(들)을 포함하는 핵산(들), 가이드 RNA(들)를 암호화하는 뉴클레오타이드 서열(들)을 포함하는 핵산(들), 공여자 주형을 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 등)의 세포 내로의 도입을 포함한다. 상기 방법은 또한 세포밖 시험관내에서 수행될 수 있기 때문에, 표적 핵산을 접촉시키는 단계를 포함하는 방법은, (달리 구체화되지 않는 한) 시험관내 세포 밖에서, 시험관내 세포 내에서, 생체내 세포 내에서, 생체의 세포 내 등에서 접촉시키는 단계를 포함한다.

[0305] 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 세포 내로 CasY 좌위, 예를 들어, CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열뿐만 아니라 CasY 좌위를 포함하는 세포로부터의 CasY-암호화 뉴클레오타이드 서열을 둘러싸는 길이가 약 1 킬로베이스(kb) 내지 5kb인 뉴클레오타이드 서열을 포함하는 핵산을 도입하는 단계를 포함하되(예를 들어, 일부 경우에 천연 상태(천연에서 생기는 상태)에서 세포는 CasY 좌위를 포함함), 여기서 표적 세포는 정상적으로는(그의 천연 상태에서는) CasY 좌위를 포함하지 않는다. 그러나, 암호화된 crRNA(들)에 대한 가이드 서열을 암호화하는 하나 이상의 스페이서 서열은 관심 대상의 하나 이상의 표적 서열이 표적화되도록 변형될 수 있다. 따라서, 예를 들어, 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 세포 내로 CasY 좌위, 예를 들어, 공급원 세포로부터 얻은 핵산을 도입하는 단계를 포함하되(예를 들어, 일부 경우에 천연 상태(천연에서 생기는 상태)의 세포는 CasY 좌위를 포함함), 여기서 핵산은 길이가 100개의 뉴클레오타이드(nt) 내지 5kb(예를 들어, 길이가 100 nt 내지 500 nt, 500 nt 내지 1kb, 1kb 내지 1.5kb, 1.5kb 내지 2kb, 2kb 내지 2.5kb, 2.5kb 내지 3kb, 3kb 내지 3.5kb, 3.5kb 내지 4kb, 또는 4kb 내지 5kb)의 길이를 가지며, CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함한다. 상기 언급한 바와 같이, 일부 이러한 경우에, 암호화된 crRNA(들)에 대한 가이드 서열을 암호화하는 하나 이상의 스페이서 서열은 관심 대상의 하나 이상의 표적 서열이 표적화되도록 변형될 수 있다. 일부 경우에, 상기 방법은 표적 세포 내로: i) CasY 좌위; 및 ii) 공여자 DNA 주형을 도입하는 단계를 포함한다. 일부 경우에, 표적 핵산은 시험관내 무세포 조성물이다. 일부 경우에, 표적 핵산은 표적 세포에 존재한다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 원핵 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 진핵 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 포유류 세포이다. 일부 경우에, 표적 핵산은 표적 세포에 존재하며, 여기서 표적 세포는 식물 세포이다.

[0306] 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 본 개시내용의 CasY 폴리펩타이드와, 또는 본 개시내용의 CasY 융합 폴리펩타이드와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 CasY 폴리펩타이드 및 CasY 가이드 RNA와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 CasY 폴리펩타이드, 제1 CasY 가이드 RNA, 및 제2 CasY 가이드 RNA와 접촉시키는 단계를 포함한다. 일부 경우에, 표적 핵산을 변형시키기 위한 본 개시내용의 방법은 표적 핵산을 본 개시내용의 CasY 폴리펩타이드 및 CasY 가이드 RNA 및 공여자 DNA 주형과 접촉시키는 단계를 포함한다.

[0307] **관심 대상의 표적 핵산 및 표적 세포**

[0308] CasY 가이드 RNA에 결합될 때, 본 개시내용의 CasY 폴리펩타이드, 또는 본 개시내용의 CasY 융합 폴리펩타이드는 표적 핵산에 결합할 수 있고, 일부 경우에, 표적 핵산에 결합하고 이를 변형시킬 수 있다. 표적 핵산은 임의의 핵산(예를 들어, DNA, RNA)일 수 있고, 이중 가닥 또는 단일 가닥일 수 있으며, 임의의 유형의 핵산(예를 들어, 염색체(게놈 DNA), 염색체로부터 유래, 염색체 DNA, 플라스미드, 바이러스, 세포외, 세포내, 미토콘드리아, 엽록체, 선형, 원형 등)일 수 있고, (예를 들어, 표적 핵산이 표적화될 수 있도록 CasY 가이드 RNA가 표적 핵산 내 표적 서열에 혼성화하는 뉴클레오타이드 서열을 포함하는 한) 임의의 유기체로부터 유래될 수 있다.

- [0309] 표적 핵산은 DNA 또는 RNA일 수 있다. 표적 핵산은 이중 가닥(예를 들어, dsDNA, dsRNA) 또는 단일 가닥(예를 들어, ssRNA, ssDNA)일 수 있다. 일부 경우에, 표적 핵산은 단일 가닥이다. 일부 경우에, 표적 핵산은 단일 가닥 RNA(ssRNA)이다. 일부 경우에, 표적 ssRNA(예를 들어, 표적 세포 ssRNA, 바이러스 ssRNA 등)은 mRNA, rRNA, tRNA, 비-암호 RNA(ncRNA), 긴 비-암호 RNA(lncRNA) 및 마이크로RNA(miRNA)로부터 선택된다. 일부 경우에, 표적 핵산은 단일 가닥 DNA(ssDNA)(예를 들어, 바이러스 DNA)이다. 상기 언급한 바와 같이, 일부 경우에, 표적 핵산은 단일 가닥이다.
- [0310] 표적 핵산은 어디에서나, 예를 들어, 시험관내 세포 밖에서, 시험관내 세포 내에서, 생체내 세포 내에서, 생체 외 세포 내 등에 위치될 수 있다. 적합한 표적 세포(표적 핵산, 예컨대 게놈 DNA를 포함할 수 있음)는 박테리아 세포; 고세균 세포; 단세포 진핵 유기체의 세포; 식물 세포; 조류 세포, 예를 들어, 보트리오크쿠스 브라우니, 칼라마이도모나스 레인하르티, 난노클로롭시스 가디타나, 클로렐라 피레노이도사, 모자반속 등; 진균 세포(예를 들어, 효모 세포); 동물 세포; 무척추동물(예를 들어, 과일파리, 강장동물, 극피동물, 선충 등)로부터의 세포; 곤충(예를 들어, 모기; 벌; 해충 등)의 세포; 거미류(예를 들어, 거미; 진드기 등)의 세포; 척추동물(예를 들어, 어류, 양서류, 파충류, 조류, 포유류)로부터의 세포; 포유류(예를 들어, 설치류로부터의 세포; 인간으로부터의 세포; 비-인간 포유류의 세포; 설치류(예를 들어, 마우스, 래트)의 세포; 토끼목(예를 들어, 토끼)의 세포; 유제류(예를 들어, 소, 말, 낙타, 라마, 비쿠냐, 양, 염소 등)의 세포; 해양 포유류(예를 들어, 고래, 바다표범, 바다 코끼리, 돌고래, 바다사자 등) 등을 포함하지만, 이들로 제한되지 않는다. 임의의 유형의 세포에 관심이 있을 수 있다(예를 들어, 줄기 세포, 예를 들어, 배아 줄기(ES) 세포, 유도 만능 줄기(iPS) 세포, 생식세포(예를 들어, 난모세포, 정자, 난원세포, 정원세포 등), 성체 줄기 세포, 체세포, 예를 들어, 섬유아세포, 조혈모세포, 뉴런, 근육 세포, 골세포, 간세포, 췌장 세포; 임의의 단계에서 배아의 시험관내 또는 생체내 배아 세포, 예를 들어, 1-세포, 2-세포, 4-세포, 8-세포 등 단계 제브라피쉬 배아 등).
- [0311] 세포는 확립된 세포주로부터 유래될 수 있거나 또는 그들은 1차 세포일 수 있고, 여기서 "1차 세포", "1차 세포주" 및 "1차 배양물"은 대상으로부터 유래되고 배양물의 제한된 수의 계대, 즉, 분할 동안 시험관내에서 성장이 허용되는 세포 및 세포 배양물을 지칭하기 위해 본 명세서에서 상호 호환적으로 사용된다. 예를 들어, 1차 배양물은 0회, 1회, 2회, 4회, 5회, 10회 또는 15회(그러나, 위기의 단계를 통과하기에 충분한 시간은 아님) 계대될 수 있는 배양물이다. 전형적으로, 1차 세포주는 시험관내에서 10회보다 더 적은 계대 동안 유지된다. 표적 세포는 단세포 유기체일 수 있고/있거나 배양물에서 성장될 수 있다. 세포가 1차 세포라면, 그들은 임의의 편리한 방법에 의해 개체로부터 채취될 수 있다. 예를 들어, 백혈구는 분리반출법, 백혈구분리반출법, 밀도 구배 분리 등에 의해 편리하게 채취될 수 있는 한편, 조직, 예컨대 피부, 근육, 골수, 비장, 간, 췌장, 폐, 장, 위 등으로부터의 세포는 생검에 의해 편리하게 채취될 수 있다.
- [0312] 일부 상기 적용에서, 대상 방법은 (예를 들어, 표적화된 mRNA에 의해 암호화된 단백질의 생성을 파괴하기 위해, 표적 DNA를 절단하거나 또는 달리 변형시키기 위해, 표적 세포 등을 유전적으로 변형시키기 위해) 생체내 및/또는 생체외 및/또는 시험관내 유사분열 또는 유사분열후 세포에서 표적 핵산 절단, 표적 핵산 변형을 유도하기 위해, 그리고/또는 표적 핵산에 결합하기 위해(예를 들어, 수집 및/또는 분석 등을 위한 시각화를 위해) 사용될 수 있다. 가이드 RNA는 표적 핵산에 혼성화함으로써 특이성을 제공하기 때문에, 개시된 방법에서 관심 대상의 유사분열 및/또는 유사분열후 세포는 임의의 유기체로부터의 세포(예를 들어, 박테리아 세포, 고세균 세포, 단일 세포 진핵 유기체의 세포, 식물 세포, 조류 세포, 예를 들어, 보트리오크쿠스 브라우니, 칼라마이도모나스 레인하르티, 난노클로롭시스 가디타나, 클로렐라 피레노이도사, 모자반속 등, 진균 세포(예를 들어, 효모 세포), 동물 세포, 무척추동물(예를 들어, 과일파리, 강장동물, 극피동물, 선충 등)으로부터의 세포, 척추동물(예를 들어, 어류, 양서류, 파충류, 조류, 포유류)로부터의 세포, 포유류로부터의 세포, 설치류로부터의 세포, 인간으로부터의 세포 등)을 포함할 수 있다. 일부 경우에, 대상 CasY 단백질(및/또는 단백질을 암호화하는 핵산, 예컨대 DNA 및/또는 RNA), 및/또는 CasY 가이드 RNA(및/또는 가이드 RNA를 암호화하는 DNA), 및/또는 공여자 주형, 및/또는 RNP는 개체에게 도입될 수 있다(즉, 표적 세포는 생체내일 수 있다)(예를 들어, 포유류, 래트, 마우스, 돼지, 영장류, 비-인간 영장류, 인간 등). 일부 경우에, 이러한 투여는, 예를 들어, 표적화된 세포의 게놈을 편집함으로써 질환을 치료하고/하거나 예방하는 목적을 위한 것일 수 있다.
- [0313] 식물 세포는 외떡잎식물의 세포, 및 쌍떡잎식물의 세포를 포함한다. 세포는 뿌리 세포, 잎 세포, 물관부의 세포, 체관부의 세포, 형성층의 세포, 정단분열조직 세포, 격벽 유세포, 후각조직, 후막조직 세포 등일 수 있다. 식물 세포는 농작물, 예컨대 밀, 옥수수, 벼, 수수, 조, 대두 등의 세포를 포함한다. 식물 세포는 농업 과일 및 견과 식물, 예를 들어, 살구, 오렌지, 레몬, 사과, 자두, 배, 아몬드 등을 생산하는 식물을 포함한다.
- [0314] 표적 세포의 추가적인 예는 상기 표제 "변형 세포" 부문에서 열거된다. 세포(표적 세포)의 비제한적 예는 원핵

세포, 진핵 세포, 박테리아 세포, 고세균 세포, 단세포 진핵 유기체의 세포, 원생동물 세포, 식물로부터의 세포 (예를 들어, 식물 작물, 과일, 채소, 곡물, 대두, 옥수수, 마이스, 밀, 종자, 토마토, 벼, 카사바, 사탕수수, 호박, 건초, 감자, 목화, 대마초, 담배, 하藕, 구과 식물, 겉씨식물, 속씨식물, 양치식물, 석송, 뿔이끼류, 우산이끼류, 선대식물, 쌍떡잎식물, 외떡잎식물 등), 조류 세포, (예를 들어, 보트리오크쿠스 브라우니 (*Botryococcus braunii*), 칼라마이도모나스 레인하르티(*Chlamydomonas reinhardtii*), 난노클로롭시스 가디타나 (*Nannochloropsis gaditana*), 클로렐라 피레노이도사(*Chlorella pyrenoidosa*), 모자반속(*Sargassum patens*, *C. agardh*) 등), 해초(예를 들어, 켈프) 진균 세포(예를 들어, 효모 세포, 버섯으로부터의 세포), 동물 세포, 무척추동물(예를 들어, 과신파리, 강장동물, 극피동물, 선충 등)으로부터의 세포, 척추동물(예를 들어, 어류, 양서류, 파충류, 조류, 포유류)로부터의 세포, 포유류(예를 들어, 유체류(예를 들어, 돼지, 소, 염소, 양)로부터의 세포; 설치류(예를 들어, 래트, 마우스); 비-인간 영장류; 인간; 고양이과(예를 들어, 고양이); 개과(예를 들어, 개); 등) 등을 포함한다. 일부 경우에, 세포는 천연 유기체로부터 유래되지 않은 세포이다(예를 들어, 세포는 합성으로 만들어진 세포일 수 있고; 또한 인공 세포로서 지칭된다).

[0315] 세포는 시험관내 세포일 수 있다(예를 들어, 확립된 배양 세포주). 세포는 생체의 세포일 수 있다(개체로부터의 배양 세포). 세포는 생체내 세포(예를 들어, 개체 내 세포)일 수 있다. 세포는 단리된 세포일 수 있다. 세포는 유기체 내부의 세포일 수 있다. 세포는 유기체일 수 있다. 세포는 세포 배양물 내 세포일 수 있다(예를 들어, 시험관내 세포 배양물). 세포는 세포 수집물 중 하나일 수 있다. 세포는 원핵 세포이거나 또는 원핵 세포로부터 유래될 수 있다. 세포는 박테리아 세포일 수 있거나 또는 박테리아 세포로부터 유래될 수 있다. 세포는 고세균 세포일 수 있거나 또는 고세균 세포로부터 유래될 수 있다. 세포는 진핵 세포일 수 있거나 또는 진핵 세포로부터 유래될 수 있다. 세포는 식물 세포일 수 있거나 또는 식물 세포로부터 유래될 수 있다. 세포는 동물 세포일 수 있거나 또는 동물 세포로부터 유래될 수 있다. 세포는 무척추동물 세포일 수 있거나 또는 무척추동물 세포로부터 유래될 수 있다. 세포는 척추동물 세포일 수 있거나 또는 척추동물 세포로부터 유래될 수 있다. 세포는 포유류 세포일 수 있거나 또는 포유류 세포로부터 유래될 수 있다. 세포는 설치류 세포일 수 있거나 또는 설치류 세포로부터 유래될 수 있다. 세포는 인간 세포일 수 있거나 또는 인간 세포로부터 유래될 수 있다. 세포는 미생물 세포일 수 있거나 또는 미생물 세포로부터 유래될 수 있다. 세포는 진균 세포일 수 있거나 또는 진균 세포로부터 유래될 수 있다. 세포는 곤충 세포일 수 있다. 세포는 절지동물 세포일 수 있다. 세포는 원생동물 세포일 수 있다. 세포는 유충 세포일 수 있다.

[0316] 적합한 세포는 줄기 세포(예를 들어, 배아 줄기(ES) 세포, 유도 만능 줄기(iPS) 세포; 생식세포(예를 들어, 난모세포, 정자, 난원세포, 정원세포 등); 체세포, 예를 들어, 섬유아세포, 희소돌기신경교, 교세포, 조혈모세포, 뉴런, 근육 세포, 골세포, 간세포, 췌장 세포 등을 포함한다.

[0317] 적합한 세포는 인간 배아 줄기 세포, 태아 심장 근육 세포, 근육섬유아세포, 간충직 줄기 세포, 자가이식 확장된 심장 근육 세포, 지방세포, 전분화 세포, 만능 세포, 혈액 줄기 세포, 근아세포, 성체 줄기 세포, 골수 세포, 간충직 세포, 배아 줄기 세포, 실질 세포, 상피세포, 내피세포, 중피세포, 섬유아세포, 조골세포, 연골세포, 외인성 세포, 내인성 세포, 줄기 세포, 조혈모세포, 골수 유래 전구체 세포, 심근세포, 골격 근아세포, 대식세포, 모세혈관 내피세포, 이종 세포, 동종 세포 및 출산후 줄기 세포를 포함한다.

[0318] 일부 경우에, 세포는 면역세포, 뉴런, 상피세포 및 내피세포 또는 줄기 세포이다. 일부 경우에, 면역세포는 T 세포, B 세포, 단핵구, 자연살해 세포, 수지상 세포 또는 대식세포이다. 일부 경우에, 면역세포는 세포독성 T 세포이다. 일부 경우에, 면역세포는 헬퍼 T 세포이다. 일부 경우에, 면역세포는 조절 T 세포(Treg)이다.

[0319] 일부 경우에, 세포는 줄기 세포이다. 줄기 세포는 성체 줄기 세포를 포함한다. 성체 줄기 세포는 또한 체세포 줄기 세포로서 지칭된다.

[0320] 성체 줄기 세포는 분화된 조직 내에 존재하지만, 자기 재생 특성 및 다중 세포 유형, 보통 줄기 세포가 발견된 전형적인 조직의 세포 유형이 생기게 하는 능력을 보유한다. 체세포 줄기 세포의 수많은 예는 당업자에게 공지되어 있으며, 근육 줄기 세포; 조혈모세포; 상피 줄기 세포; 신경 줄기 세포; 간충직 줄기 세포; 유방 줄기 세포; 장 줄기 세포; 중배엽 줄기 세포; 내피 줄기 세포; 후각 줄기 세포; 신경 능선 줄기 세포 등을 포함한다.

[0321] 관심 대상의 줄기 세포는 포유류 줄기 세포를 포함하며, 용어 "포유류"는 인간; 비-인간 영장류; 가축 및 농장 동물; 및 동물원, 실험실, 운동 또는 반려 동물, 예컨대 개, 말, 고양이, 소, 마우스, 래트, 토끼 등을 포함하는 포유류로서 분류된 임의의 동물을 지칭한다. 일부 경우에, 줄기 세포는 인간 줄기 세포이다. 일부 경우에, 줄기 세포는 설치류(예를 들어, 마우스; 래트) 줄기 세포이다. 일부 경우에, 줄기 세포는 비-인간 영장류 줄기

세포이다.

- [0322] 줄기 세포는 하나 이상의 줄기 세포 마커, 예를 들어, SOX9, KRT19, KRT7, LGR5, CA9, FXYD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17 및 PPARGC1A를 발현시킬 수 있다.
- [0323] 일부 실시형태에서, 줄기 세포는 조혈모세포(HSC)이다. HSC는 골수, 혈액, 제대혈, 태아간 및 난황낭으로부터 단리될 수 있는 중배엽-유래 세포이다. HSC는 $CD34^{+}$ 및 $CD3^{-}$ 로서 특성규명된다. HSC는 생체내 적혈구, 호중성-대식세포, 거핵구 및 림프계 조혈세포 계통을 재생시킬 수 있다. 시험관내에서, HSC는 적어도 일부의 자기-재생 세포 분열을 겪도록 유도될 수 있고, 생체내에서 알 수 있는 바와 같이 동일한 계통으로 분화되도록 유도될 수 있다. 이렇게 해서, HSC는 적혈구 세포, 거핵구, 호중구, 대식세포 및 림프구 세포 중 하나 이상으로 분화되도록 유도될 수 있다.
- [0324] 다른 실시형태에서, 줄기 세포는 신경 줄기 세포(NSC)이다. 신경 줄기 세포(NSC)는 뉴런, 및 교세포(회소돌기신경교 및 성상교세포를 포함)로 분화할 수 있다. 신경 줄기 세포는 다회 분열할 수 있는 다능성 줄기 세포이고, 특정 조건 하에서 신경아세포 또는 아교모세포, 예를 들어, 각각 뉴런 및 교세포 중 하나 이상의 유형이 되도록 기여하는 세포일 수 있는 신경 줄기 세포 또는 신경 전구체 세포인 딸 세포를 생성할 수 있다. NSC를 얻는 방법은 당업계에 공지되어 있다.
- [0325] 다른 실시형태에서, 줄기 세포는 간충직 줄기 세포(MSC)이다. MSC는 본래 배아 중배엽으로부터 유래되고, 성체 골수로부터 단리되며, 근육, 뼈, 연골, 지방, 골수 기질 및 힘줄을 형성하도록 분화될 수 있다. MSC를 단리시키는 방법은 당업계에 공지되어 있으며; MSC를 얻기 위한 임의의 공지된 방법이 사용될 수 있다. 예를 들어, 인간 MSC의 단리를 기재하는 미국 특허 제5,736,396호 참조.
- [0326] 세포는 일부 경우에 식물 세포이다. 식물 세포는 외떡잎식물의 세포일 수 있다. 세포는 쌍떡잎식물의 세포일 수 있다.
- [0327] 일부 경우에, 세포는 식물 세포이다. 예를 들어, 세포는 주요 농작물, 예를 들어, 보리, 콩류(건조 식용), 카놀라, 옥수수, 목화(피마), 목화(엽편드), 아마씨, 건초(알팔파), 건초(비-알팔파), 귀리, 땅콩, 벼, 수수, 대두, 사탕무, 사탕수수, 해바라기(오일), 해바라기(비-오일), 고구마, 담배(벌리), 담배(연관처리(Flue-cured)), 토마토, 밀(듀럼), 밀(봄), 밀(겨울) 등의 세포일 수 있다. 다른 예로서, 세포는, 예를 들어, 알팔파, 알로에 잎, 에로루트, 벚꽃, 아티초크, 아스파라거스, 죽순, 바나나꽃, 콩나물, 콩류, 비이트릅, 근대, 비타멜론, 청경채, 브로콜리, 브로콜리 라베(엽경채), 방울다다기 양배추, 양배추, 방울 양배추, 가재발 선인장(노팔), 호리병박 열매, 카르둔, 당근, 콜리플라워, 셀러리, 차요테, 중국 아티초크(두루미냉이), 중국 양배추, 중국 셀러리, 중국 차이브, 채심, 국화잎(통호), 콜라드 그린, 옥수수대, 단옥수수, 오이, 대견, 단데리온 그린, 타로토란, 다우 무에(피 팁), 동과(겨울멜론), 가지, 엔다이브, 에스카롤, 선수 장식 양치식물, 갯류 식물, 치커리, 갯(중국 겨자), 가일론, 갈랑가(시암, 태국 생강), 마늘, 생강 뿌리, 우엉, 나물, 하노버 셀러드 나물, 후아우존틀, 예루살렘 아티초크, 히카마, 케일 나물, 콜라비, 명아주(켈리페), 상추(비브), 상추(보스틴), 상추(보스틴 레드), 상추(녹색잎), 상추(아이스버그), 상추(콜라로사), 상추(오크리프 - 녹색), 상추(오크리프 - 적색), 상추(가공처리), 상추(적색잎), 상추(로메인), 상추(루비 로메인), 상추(러시안 레드 겨자), 린콕, 로 북, 긴콩류, 연근, 마하, 용설란 (아가베) 잎, 말랑가, 메스콜랭 믹스, 겨자과, 모압(수세미오이), 무, 모쿠아(동과), 버섯, 겨자, 참마, 오크라, 웅초이, 그린 어니언, 오포(롱 스쿼시), 장식용 옥수수, 관상용 호박, 파슬리, 파스닙, 완두콩, 후추(벨형), 후추, 호박, 라디키오, 무순, 무, 레이프 그린, 레이프 그린, 대황, 로메인(베이비 레드), 루타베가, 함초(씨빈), 신쿠아(앵글드/줄기가 선 수세미), 시금치, 스쿼시, 쥘 더미, 사탕수수, 고구마, 근대, 타마린드, 타로, 타로잎, 타로 슈트, 탕소이, 테페구아헤(구아헤), 빔바, 토마틸로, 토마토, 토마토(체리), 토마토(포도형), 토마토(자두형), 강황, 순무 어린잎 나물, 순무, 마름, 양파, 양스(네임스), 유초이, 유카(카사바) 등을 포함하지만, 이들로 제한되지 않는 채소 작물의 세포이다.
- [0328] 세포는 일부 경우에 절지동물 세포이다. 예를 들어, 세포는, 예를 들어, 협각류, 다지류, 헥시포디아(*Hexipodia*), 거미강, 곤충강, 돌جوم목, 줄류, 구시아강, 하루살이목, 잠자리목, 잠자리하목, 실잠자리아목, 신시류, 외시류, 강도래목, 흰개미붙이목, 메뚜기, 민벌레목, 집게벌레목, 바퀴목, 귀뚜라미붙이목, 갈르와벌레과, 만토파스마토데과, 대벌레목, 바퀴목, 흰개미목, 사마귀목, 책좀, 다듬이벌레목, 총채벌레목, 이목, 반시목, 내시류 또는 완전변태류, 막시류, 초시류, 부채벌레목, 약대벌레목, 뱀잠자리목, 풀잠자리목, 밀드리목, 벼룩목, 파리목, 날도래류 또는 인시목의 아목, 과, 아과, 군, 하위군, 또는 종의 세포일 수 있다.
- [0329] 세포는 일부 경우에 곤충 세포이다. 예를 들어, 일부 경우에, 세포는 모기, 메뚜기, 노린재류, 파리, 벼룩, 벌,

말벌, 개미, 이, 나방 또는 딱정벌레의 세포이다.

[0330] **표적 세포 내로의 성분의 도입**

[0331] Cas9 가이드 RNA(또는 이를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산), 및/또는 Cas9 융합 폴리펩타이드(또는 이를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산) 및/또는 공여자 폴리뉴클레오타이드는 임의의 다양한 잘 공지된 방법에 의해 숙주 세포 내로 도입될 수 있다.

[0332] 핵산을 세포 내로 도입하는 방법은 당업계에 공지되어 있고, 임의의 편리한 방법을 사용하여 핵산(예를 들어, 발현 작제물)을 표적 세포(예를 들어, 진핵 세포, 인간 세포, 줄기 세포, 전구체 세포 등) 내로 도입할 수 있다. 적합한 방법은 본 명세서의 다른 곳에 더 상세하게 기재되어 있으며, 예를 들어, 바이러스 또는 박테리오파지 감염, 형질감염, 접합, 프로토플라스트 융합, 리포펙션, 전기천공법, 인산칼슘 침전, 폴리에틸렌이민(PEI)-매개 형질감염, DEAE-텍스트란 매개 형질감염, 리포좀-매개 형질감염, 입자총 기법, 인산칼슘 침전, 직접 미량주사법, 나노입자-매개 핵산 전달(예를 들어, 문헌[Panyam et., al Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023] 참조)등을 포함한다. 임의의 또는 모든 성분이 공지된 방법, 예를 들어, 뉴클레오펙션을 이용하여 조성물(예를 들어, CasY 폴리펩타이드, CasY 가이드 RNA, 공여자 폴리뉴클레오타이드 등의 임의의 편리한 조합을 포함)로서 세포 내로 도입될 수 있다.

[0333] **공여자 폴리뉴클레오타이드(공여자 주형)**

[0334] CasY 가이드 RNA에 의해 가이드된, CasY 단백질은 일부 경우에 비-상동성 말단 접합(NHEJ) 또는 상동성-관련 재조합(HDR) 중 하나에 의해 수선된 이중-가닥 DNA(dsDNA) 표적 핵산 내에서 부위-특이적 이중 가닥 파손(double strand break: DSB) 또는 단일 가닥 파손(SSB)(예를 들어, CasY 단백질이 틈내기효소 변이체일 때)을 생성한다.

[0335] 일부 경우에, 표적 DNA를 (CasY 단백질 및 CasY 가이드 RNA와) 접촉시키는 것은 비상동성 말단 결합 또는 상동 직접 수선을 허용하는 조건 하에서 일어난다. 따라서, 일부 경우에, 대상 방법은 표적 DNA를 공여자 폴리뉴클레오타이드와 (예를 들어, 공여자 폴리뉴클레오타이드를 세포 내로 도입함으로써) 접촉시키는 단계를 포함하되, 공여자 폴리뉴클레오타이드, 공여자 폴리뉴클레오타이드의 일부, 공여자 폴리뉴클레오타이드의 복제물, 또는 공여자 폴리뉴클레오타이드 복제물의 일부가 표적 DNA 내로 통합된다. 일부 경우에, 상기 방법은 세포를 공여자 폴리뉴클레오타이드와 접촉시키는 단계를 포함하지 않으며, 표적 DNA 내의 뉴클레오타이드가 결실되도록 표적 DNA는 변형된다.

[0336] 일부 경우에, CasY 가이드 RNA(또는 이를 암호화하는 DNA) 및 CasY 단백질(또는 이를 암호화하는 핵산, 예컨대 RNA 또는 DNA, 예를 들어, 하나 이상의 발현 벡터)는 적어도 표적 DNA 서열에 상동성인 세그먼트를 포함하는 공여자 폴리뉴클레오타이드 서열과 함께 공동투여되며(예를 들어, 표적 핵산과 접촉되고, 세포에 투여됨 등), 대상 방법은 핵산 물질을 표적 DNA 서열에 첨가하기 위해, 즉, 삽입 또는 대체하기 위해(예를 들어, 핵산, 예를 들어, 단백질을 암호화하는 것, siRNA, miRNA 등을 "닉인"하기 위해), 태그(예를 들어, 6xHis, 형광 단백질(예를 들어, 녹색 형광 단백질; 황색형광 단백질 등), 혈구응집소(HA), FLAG 등)를 첨가하기 위해, 조절 서열을 유전자(예를 들어, 프로모터, 폴리아데닐화 신호, 내부 리보솜 유입 서열(IRES), 2A 펩타이드, 개시 코돈, 정지 코돈, 스프라이스 신호, 국소화 신호 등)를 첨가하기 위해, 핵산 서열(예를 들어, 돌연변이를 도입하기 위해, 정확한 서열을 도입함으로써 돌연변이를 야기하는 질환을 제거하기 위해) 등을 변형시키기 위해 사용될 수 있다. 이렇게 해서, CasY 가이드 RNA 및 CasY 단백질을 포함하는 복합체는 부위-특이적, 즉, "표적화된" 방법으로 DNA를 변형시키는 것이 바람직한 임의의 시험관내 또는 생체내 적용, 예를 들어, 유전자 요법에서 사용되는 바와 같은, 예를 들어, 유전자 닉인, 유전자 편집, 유전자 태깅 등에서, 예를 들어, 질환을 치료하거나 또는 항바이러스, 항병원성 또는 항암 치료제로서, 농업에서 유전자 변형된 유기체의 생산, 치료적, 진단적 또는 연구적 목적을 위한 세포에 의한 단백질의 생성, iPS 세포의 유도, 생물학적 연구, 검출 또는 대체 등을 위한 병원균 유전자의 표적화에서 유용하다.

[0337] 폴리뉴클레오타이드 서열을 표적 서열이 절단되는 게놈 내로 삽입하는 것이 바람직한 적용에서, 공여자 폴리뉴클레오타이드(공여자 서열을 포함하는 핵산)는 또한 세포에 제공될 수 있다. "공여자 서열" 또는 "공여자 폴리뉴클레오타이드" 또는 "공여자 주형"은 (예를 들어, dsDNA 절단 후에, 표적 DNA의 틈내기 후에, 표적 DNA의 이중 틈내기 후에 등) CasY 단백질에 의해 절단된 부위에 삽입될 핵산 서열을 의미한다. 공여자 폴리뉴클레오타이드는 그것과 상동성을 보유하는 게놈 서열 사이의 상동 직접 수선을 뒷받침하기 위해, 예를 들어, 표적 부위의 약 50개 이하의 염기 이내, 예를 들어, 약 30개의 염기 이내, 약 15개의 염기 이내, 약 10개의 염기 이내, 약 5개의 염기 이내, 또는 표적 부위에 바로 축적하여 표적 부위에서 게놈 서열에 대한 충분한 상동성, 예를 들어,

표적 부위에 측접하는 뉴클레오타이드 서열과 70%, 80%, 85%, 90%, 95% 또는 100% 상동성을 함유할 수 있다. 공여자와 게놈 서열 사이의 서열 상동성의 대략 25, 50, 100 또는 200개 이상의 뉴클레오타이드, 또는 200개 초과 의 뉴클레오타이드(또는 10 내지 200개 이상의 뉴클레오타이드의 임의의 정수값)는 상동 직접 수선을 뒷받침할 수 있다. 공여자 폴리뉴클레오타이드는 임의의 길이, 예를 들어, 10개 이상의 뉴클레오타이드, 50개 이상의 뉴클레오타이드, 100개 이상의 뉴클레오타이드, 250개 이상의 뉴클레오타이드, 500개 이상의 뉴클레오타이드, 1000개 이상의 뉴클레오타이드, 5000개 이상의 뉴클레오타이드 등을 가질 수 있다.

[0338] 공여자 서열은 전형적으로 그것이 대체하는 게놈 서열과 동일하지 않다. 오히려, 상동 직접 수선을 뒷받침하기 위해(예를 들어, 유전자 보정을 위해, 예를 들어, 질환 원인 염기쌍을 비질환 원인 염기쌍으로 전환시키기 위해) 충분한 상동성이 존재한다면 공여자 서열은 게놈 서열에 대해 적어도 하나 이상의 단일 염기 변화, 삽입, 결실, 역위 또는 재배열을 함유할 수 있다. 일부 실시형태에서, 표적 DNA 영역과 두 측접 서열 사이의 상동 직접 수선이 표적 영역에서 비상동성 서열의 삽입을 초래하도록 공여자 서열은 상동체의 두 영역에 의해 측접되는 비상동성 서열을 포함한다. 공여자 서열은 또한 관심 대상의 DNA 영역에 상동성이 아니고 관심 대상의 DNA 영역 내로의 삽입에 대해 의도되지 않는 서열을 함유하는 벡터 골격을 포함할 수 있다. 일반적으로, 공여자 서열의 상동성 영역(들)은 재조합이 요망되는 게놈 서열에 대해 적어도 50%의 서열 동일성을 가질 것이다. 특정 실시형태에서, 60%, 70%, 80%, 90%, 95%, 98%, 99% 또는 99.9%의 서열 동일성이 존재한다. 공여자 폴리뉴클레오타이드의 길이에 따라서 1% 내지 100%의 서열 동일성의 임의의 값이 존재할 수 있다.

[0339] 공여자 서열은 게놈 서열, 예를 들어, 제한 부위, 뉴클레오타이드 다형성, 선택 가능한 마커(예를 들어, 약물 내성 유전자, 형광 단백질, 효소 등) 등에 비교하여 특정 서열 차이를 포함할 수 있는데, 이는 절단 부위에서 공여자 서열의 성공적인 삽입에 대한 평가를 위해 사용될 수 있거나 또는 일부 경우에 다른 목적을 위해(예를 들어, 표적화된 게놈 좌위에서 발현을 나타내기 위해) 사용될 수 있다. 일부 경우에, 암호 영역에 위치된다면, 이러한 뉴클레오타이드 서열 차이는 아미노산 서열을 변화시키지 않을 것이고, 또는 침묵 아미노산 변화(즉, 단백질의 구조 또는 기능에 영향을 미치지 않는 변화)를 만들 것이다. 대안적으로, 이들 서열 차이는 측접하는 재조합 서열, 예컨대 FLP, loxP 서열 등을 포함할 수 있는데, 이는 마커 서열의 제거를 위해 이후의 시간에 활성화될 수 있다.

[0340] 일부 경우에, 공여자 서열은 단일-가닥 DNA로서 세포에 제공된다. 일부 경우에, 공여자 서열은 이중-가닥 DNA로서 세포에 제공된다. 이는 선형 또는 원형 형태로 세포에 도입될 수 있다. 선형 형태로 도입된다면, 공여자 서열의 말단은 임의의 편리한 방법에 의해 (예를 들어, 핵산외부분해로부터) 보호될 수 있고, 이러한 방법은 당업자에게 공지되어 있다. 예를 들어, 하나 이상의 다이데옥시뉴클레오타이드 잔기가 선형 분자의 3' 말단에 첨가될 수 있고/있거나 자기-상보성 올리고뉴클레오타이드가 단부 중 하나 또는 둘 다에 결합될 수 있다. 예를 들어, 문헌[Chang et al. (1987) Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889] 참조. 분해로부터 외인성 폴리뉴클레오타이드를 보호하기 위한 추가적인 방법은 말단 아미노기(들)의 첨가 및 변형된 뉴클레오타이드간 결합, 예를 들어, 포스포로티오에이트, 포스포르아미데이트 및 0-메틸 리보스 또는 데옥시리보스 잔기의 사용을 포함하지만, 이들로 제한되지 않는다. 선형 공여자 서열 말단 보호에 대한 대안으로서, 서열의 추가적인 길이가 재조합에 영향을 미치는 일 없이 분해될 수 있는 영역 외부에 포함될 수 있다. 공여자 서열은 추가적인 서열, 예를 들어, 복제 기점, 프로모터 및 항생제 내성을 암호화하는 유전자를 갖는 벡터의 부분으로서 세포 내로 도입될 수 있다. 게다가, 공여자 서열은 네이키드 핵산으로서, 리포솜 또는 폴록사머와 같은 제제와 복합체화된 핵산으로서 도입될 수 있거나, 또는 CasY 가이드 RNA 및/또는 CasY 융합 폴리펩타이드 및/또는 공여자 폴리뉴클레오타이드를 암호화하는 핵산에 대해 본 명세서의 다른 곳에 기재된 바와 같이, 바이러스(예를 들어, 아테노바이러스, AAV)에 의해 전달될 수 있다.

[0341] **유전자이식, 비인간 유기체**

[0342] 상기 기재한 바와 같이, 일부 경우에, 본 개시내용(예를 들어, 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산; 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 등)의 핵산(예를 들어, 재조합 발현 벡터)은 본 개시내용의 CasY 폴리펩타이드 또는 CasY 융합 폴리펩타이드를 생성하는 유전자이식 비-인간 유기체를 생성하기 위해 이식유전자로서 사용된다. 본 개시내용은 본 개시내용의 CasY 폴리펩타이드, 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 유전자이식 비인간 유기체를 제공한다.

[0343] **유전자이식, 비-인간 동물**

[0344] 본 개시내용은 유전자이식 비-인간 동물을 제공하는데, 이 동물은 CasY 폴리펩타이드 또는 CasY 융합 폴리펩타

이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산을 포함하는 이식유전자를 포함한다. 일부 실시형태에서, 유전자이식 비-인간 동물의 게놈은 본 개시내용의 CasY 폴리펩타이드 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함한다. 일부 경우에, 유전자이식 비-인간 동물은 유전자 변형에 대해 동형접합적이다. 일부 경우에, 유전자이식 비-인간 동물은 유전자 변형에 대해 이형접합적이다. 일부 실시형태에서, 유전자이식 비-인간 동물은 척추동물, 예를 들어, 어류(예를 들어, 연어, 송어, 제브라피쉬, 금붕어, 복어, 동굴어 등), 양서류(개구리, 영원, 도롱뇽 등), 조류(예를 들어, 닭, 칠면조 등), 파충류(예를 들어, 뱀, 도마뱀 등), 비-인간 포유류(예를 들어, 유제류, 예를 들어, 돼지, 소, 염소, 양 등; 토끼목(예를 들어, 토끼); 설치류(예를 들어, 래트, 마우스); 비-인간 영장류 등) 등이다. 일부 경우에, 유전자이식 비-인간 동물은 무척추동물이다. 일부 경우에, 유전자이식 비-인간 동물은 곤충(예를 들어, 모기; 해충 등)이다. 일부 경우에, 유전자이식 비-인간 동물은 거미류이다.

[0345] 본 개시내용의 CasY 폴리펩타이드, 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 미공지 프로모터(예를 들어, 핵산이 숙주 세포 게놈 내로 무작위로 통합될 때) 제어 하에 있을 수 있거나(즉, 이에 작동 가능하게 연결됨) 또는 알려진 프로모터의 제어 하에 있을 수 있다(즉, 이에 작동 가능하게 연결됨). 적합한 알려진 프로모터는 임의의 알려진 프로모터일 수 있고, 구성적으로 활성인 프로모터(예를 들어, CMV 프로모터), 유도성 프로모터(예를 들어, 열 충격 프로모터, 테트라사이클린-조절된 프로모터, 스테로이드-조절된 프로모터, 금속-조절된 프로모터, 에스트로겐 수용체-조절 프로모터 등), 공간적으로 제한된 그리고/또는 일시적으로 제한된 프로모터(예를 들어, 조직 특이적 프로모터, 세포 유형 특이적 프로모터 등) 등을 포함한다.

[0346] **유전자이식 식물**

[0347] 상기 기재한 바와 같이, 일부 경우에, 본 개시내용(예를 들어, 본 개시내용의 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산; 본 개시내용의 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산 등)의 핵산(예를 들어, 재조합 발현 벡터)은 본 개시내용의 CasY 폴리펩타이드 또는 CasY 융합 폴리펩타이드를 생성하는 유전자이식 식물을 생성하기 위해 이식유전자로서 사용된다. 본 개시내용은 본 개시내용의 CasY 폴리펩타이드, 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 유전자이식 식물을 제공한다. 일부 실시형태에서, 유전자이식 식물의 게놈은 대상 핵산을 포함한다. 일부 실시형태에서, 유전자이식 식물은 유전자 변형에 대해 동형접합적이다. 일부 실시형태에서, 유전자이식 식물은 유전자 변형에 대해 이형접합적이다.

[0348] 식물 세포 내로 외인성 핵산을 도입하는 방법은 당업계에 잘 공지되어 있다. 이러한 식물 세포는 상기 정의한 바와 같이 "형질전환되는" 것으로 고려된다. 적합한 방법은 바이러스 감염(예컨대 이중 가닥 DNA 바이러스), 형질감염, 접합, 원형질체 융합, 전기천공법, 입자충법, 인산칼슘 침전, 직접 미량주사법, 탄화규소 위스커 기법, 아그로박테륨-매개 형질전환 등을 포함한다. 방법의 선택은 일반적으로 형질감염 중인 세포의 유형 및 형질감염이 일어나는 환경(즉, 시험관내, 생체외 또는 생체내)에 의존한다.

[0349] 토양 박테리아 아그로박테륨 투메파시엔스(*Agrobacterium tumefaciens*)에 기반한 형질전환 방법은 외인성 핵산 분자를 관다발 식물 내로 도입하는 데 특히 유용하다. 아그로박테륨의 야생형 형태는 숙주 식물 상에서 종양형 성성 뿌리혹병 성장의 생산을 지시하는 Ti(종양-유도성) 플라스미드를 함유한다. 식물 게놈에 대한 Ti 플라스미드의 종양 유도성 T-DNA 영역의 전달은 Ti 플라스미드-암호화된 독력 유전자뿐만 아니라 T-DNA 경계를 필요로 하는데, 이는 전달될 영역을 설명하는 직접 DNA 반복부의 세트이다. 아그로박테리움계 벡터는 Ti 플라스미드의 변형된 형태인데, 이때 종양 유도성 작용은 식물 숙주 내로 도입될 관심 대상의 핵산 서열에 의해 대체된다.

[0350] 아그로박테륨-매개 형질전환은 일반적으로 합체 벡터 또는 바이너리 벡터 시스템을 사용하는데, 이때 Ti 플라스미드의 성분은 아그로박테륨 숙주에서 영구적으로 체류하고 독력 유전자를 운반하는 헬퍼 벡터와 T-DNA 서열에 의해 결합되는 관심 대상의 유전자를 함유하는 서틀 벡터 사이에서 분할된다. 다양한 바이너리 벡터는 당업계에 잘 공지되어 있고, 예를 들어, 클론테크(Clontech)(캘리포니아주 팔로알토에 소재)로부터 상업적으로 입수 가능하다. 배양된 식물 세포 또는 상처 조직, 예컨대 잎 조직, 뿌리 외식편, 배축, 줄기 조각 또는 덩이줄기와 함께 아그로박테륨을 배양시키는 방법은, 예를 들어, 또한 당업계에 잘 공지되어 있다. 예를 들어, 문헌[Glick and Thompson, (eds.), *Methods in Plant Molecular Biology and Biotechnology*, Boca Raton, Fla.: CRC Press (1993)] 참조.

[0351] 미세투사물-매개 형질전환이 또한 대상 유전자이식 식물을 생산하는 데 사용될 수 있다. 문헌[Klein et al. (*Nature* 327:70-73 (1987))]에 의해 처음 기재된 이 방법은 염화칼슘, 스퍼미딘 또는 폴리에틸렌 글리콜에 의한 침전에 의해 목적으로 하는 핵산 분자로 코팅된 금 또는 텅스텐과 같은 미세투사물에 의존한다. 미세투사물 입

자는 BIOLISTIC PD-1000(바이오래드(Biorad); 캘리포니아주 허큘레스에 소재)와 같은 장치를 이용하여 속씨식물 내로 고속으로 가속화된다.

[0352] 핵산이, 예를 들어 생체내 또는 생체의 프로토콜을 통해 식물 세포(들) 내로 들어갈 수 있는 방식으로 본 개시 내용의 핵산(예를 들어, 본 개시내용의 CasY 폴리펩타이드 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는 핵산(예를 들어, 재조합 발현 벡터))은 식물 내로 도입될 수 있다. "생체내"는 핵산이 식물의 살아있는 몸체에 투여되는 것, 예를 들어, 침윤을 의미한다. "생체외"는 세포 또는 외식편이 식물 외부에서 변형되고, 이어서, 이러한 세포 또는 기관이 식물에 대해 재생된다는 것을 의미한다. 문헌[Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology Academic Press, 및 Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers]에 기재된 것을 포함하는 식물 세포의 안정한 형질전환 또는 유전자이식 식물의 확립에 적합한 다수의 벡터가 기재되었다. 구체적 예는 아그로박테륨 투메파시엔스의 Ti 플라스미드로부터 유래된 것뿐만 아니라 문헌[Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technolo 3: 637-642]에 의해 개시된 것을 포함한다. 대안적으로, 비-Ti 벡터는 유리 DNA 전달 기법을 이용함으로써 식물 및 세포에 대한 DNA의 전달을 위해 사용될 수 있다. 이들 방법을 이용함으로써, 유전자이식 식물, 예컨대, 밀, 벼(Christou (1991) Bio/Technology 9:957-9 및 4462) 및 옥수수(Gordon-Kamm (1990) Plant Cell 2: 603-618)가 생산될 수 있다. 미성숙 배아는 또한 입자충(문헌[Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technolo 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48]을 이용함으로써 직접 DNA 전달 기법을 위한 그리고 아그로박테륨-매개 DNA 전달(Ishida et al. (1996) Nature Biotech 14: 745-750)을 위한 외떡잎식물에 대한 양호한 표적 조직일 수 있다. 엽록체 내로 DNA의 도입을 위한 예시적인 방법은 유전자충격, 프로토플라스트의 폴리에틸렌 글리콜 형질전환, 및 미량주사법이다(문헌[Danieli et al Nat. Biotechnol 16:345-348, 1998; Staub et al Nat. Biotechnol 18: 333-338, 2000; O'Neill et al Plant J. 3:729-738, 1993; Knoblauch et al Nat. Biotechnol 17: 906-909]; 미국 특허 제5,451,513호, 제5,545,817호, 제5,545,818호 및 제5,576,198호; 국제 특허 출원 WO 95/16783; 및 문헌[Boynton et al., Methods in Enzymology 217: 510-536 (1993), Svab et al., Proc. Natl. Acad. Sci. USA 90: 913-917 (1993), 및 McBride et al., Proc. Natl. Acad. Sci. USA 91: 7301-7305 (1994)]). 유전자충격, 프로토플라스트의 폴리에틸렌 글리콜 형질전환 및 미량주사법의 방법에 적합한 임의의 벡터는 엽록체 형질전환을 위한 표적화 벡터로서 적합할 것이다. 임의의 이중 가닥 DNA 벡터는 형질전환 벡터, 특히 도입 방법이 아그로박테륨을 이용하지 않을 때 사용될 수 있다.

[0353] 유전자 변형될 수 있는 식물은 곡물, 사료용 작물, 과일, 채소, 오일시드 작물, 약자나무, 삼림 및 덩굴식물을 포함한다. 변형될 수 있는 식물의 구체적 예는 다음에 따를 수 있다: 마이스, 바나나, 땅콩, 경협종완두, 해바라기, 토마토, 카놀라, 담배, 밀, 보리, 귀리, 감자, 대두, 목화, 카네이션, 수수, 루핀 및 벼.

[0354] 본 개시내용은 형질전환된 식물 세포, 조직, 식물 및 형질전환 식물 세포를 함유하는 생성물을 제공한다. 대상 형질전환 세포, 및 조직 및 이를 포함하는 생성물의 특징은 게놈 내로 통합되는 대상 핵산의 존재, 및 본 개시내용의 CasY 폴리펩타이드, 또는 CasY 융합 폴리펩타이드의 식물 세포에 의한 생성이다. 본 발명의 재조합 식물 세포는 재조합 세포의 집단으로서, 또는 조직, 종자, 전체 식물, 줄기, 열매, 잎, 뿌리, 꽃, 줄기, 덩이줄기, 곡물, 동물 사료, 식물의 발 등으로서 유용하다.

[0355] 본 개시내용의 CasY 폴리펩타이드, 또는 CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 미공지 프로모터(예를 들어, 핵산이 숙주 세포 게놈 내로 무작위로 통합될 때) 제어 하에 있을 수 있거나(즉, 이에 작동 가능하게 연결됨) 또는 공지된 프로모터의 제어 하에 있을 수 있다(즉, 이에 작동 가능하게 연결됨). 적합한 공지된 프로모터는 임의의 공지된 프로모터일 수 있고 구성적으로 활성화된 프로모터, 유도성 프로모터, 공간적으로 제한된 그리고/또는 일시적으로 제한된 프로모터 등을 포함한다.

[0356] **CRISPR RNA-가이드 엔도뉴클레아제의 동정 방법**

[0357] CRISPR RNA-가이드 엔도뉴클레아제의 동정 방법이 제공된다. 예를 들어, 일부 실시형태에서, 이러한 방법은 복수의 메타게놈 뉴클레오타이드 서열에서 Cas1 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 검출하는 단계를 포함한다. Cas1 단백질을 당업계에 공지되어 있고, 엔도뉴클레아제로서 작용하고 적절하게 작용하기 위해 단백질의 복합체와의 상호작용을 필요로 하지 않는 해당 CRISPR 시스템인 클래스 2 CRISPR 시스템의 CRISPR 좌위의 인근에 존재한다. Cas 1 단백질 그 자체는 CRISPR 좌위 내로의 새로운 표적 서열의 획득에 연루되고, 따라서 이 방법에 의한 동정을 위해 목적으로 하는 효과기 단백질이 아니지만, CRISPR 좌위 근처에서 Cas1 단백질의 존재는 좌위 근처에 존재하는 다른 Cas 단백질 중 적어도 하나가 효과기 단백질(RNA-가이드 엔도뉴클레아제)일 수

있다는 표시이다.

[0358] 본 명세서에서 사용되는 바와 같은, 용어 "메타게노믹스"는 샘플, 예를 들어 환경적 샘플, 예컨대 알려지지 않은 양의 원핵생물(박테리아/고세균)을 함유하고 결코 이전에 발견되고/되거나 특정규명되지 않은 원핵생물을 함유하는 샘플 내 다중 미생물(예를 들어, 박테리아, 고세균 등)로부터 회수되는 핵산의 병행 분석을 의미한다. 핵산은 임의의 편리한 방법에 의해 이러한 샘플로부터 회수될 수 있고, 임의의 주어진 핵산 분자가 유래된 미생물이 알려지지 않도록, 일반적으로 핵산은 전체 샘플로부터 함께 회수된다. 일부 실시형태에서, 샘플은 알려지지 않은 혼합물 및/또는 양의 미생물을 함유한다. 이어서, 핵산은 복수의 메타게놈 서열을 생성하도록 서열분석될 수 있다. 일부 경우에, CRISPR RNA-가이드 엔도뉴클레아제를 동정하는 대상 방법은 샘플(예를 들어, 환경 샘플)을 단리시키는 단계를 포함한다. 일부 경우에, CRISPR RNA-가이드 엔도뉴클레아제를 동정하는 대상 방법은 샘플로부터 핵산을 단리시키고/시키거나 샘플로부터 생성된 복수의 메타게놈 뉴클레오타이드 서열에 대해 샘플을 분석하는 단계를 포함한다.

[0359] 일단 Cas1 단백질이 동정되면, CRISPR RNA-가이드 엔도뉴클레아제를 동정하는 대상 방법은 Cas1-암호화 뉴클레오타이드 서열 근처에서 CRISPR 어레이(반복-스페이서-반복 어레이)를 검출하는 단계를 포함할 수 있다. 이어서, 상기 방법은 재조합 CRISPR 좌위 발현 벡터를 생성하기 위해 검출된 CRISPR 어레이를 포함하는 CRISPR 좌위를 (예를 들어, 복수의 메타게놈 뉴클레오타이드 서열로부터 유래된 핵산 샘플로부터) 발현 벡터 내로 클로닝하는 단계를 포함할 수 있다. 이어서, CRISPR 좌위는 표적 핵산을 절단하는 능력에 대해 재조합 CRISPR 좌위 발현 벡터를 분석함으로써 기능에 대해 시험될 수 있다. 임의의 편리한 분석이 사용될 수 있다. 일부 실시형태에서 분석하는 단계는 재조합 CRISPR 좌위 발현 벡터 및 표적 핵산을 세포, 예를 들어, 이중성 숙주 세포, 예컨대 이콜라이 세포 내로 도입하는 것을 포함한다. 예를 들어, 이하의 작업 실시예의 PAM 고갈 분석을 언급한다(도 5). 일부 경우에, 분석하는 단계는 숙주 세포(예를 들어, 이콜라이 세포) 집단 내로 플라스미드 라이브러리를 도입하는 단계를 포함하되, 라이브러리의 각각의 플라스미드는 표적 서열의 4 내지 10개의(예를 들어, 5 내지 10, 5 내지 8, 6 내지 10, 6 내지 8, 5, 6, 7, 8개의) 뉴클레오타이드 무작위화 5' 및/또는 3'을 가진다. 숙주 세포는 시험될 재조합 CRISPR 좌위 발현 벡터를 이미 함유할 수 있고, 또는 재조합 CRISPR 좌위 발현 벡터는 라이브러리 다음에 도입될 수 있다. 기능성이고, 따라서 기능성 CRISPR RNA-가이드 엔도뉴클레아제를 포함하는 시험 CRISPR 좌위만이 표적 서열을 갖는 플라스미드를 절단하는 능력을 초래할 것이다. 표적 서열의 무작위화 서열 5' 및 3'을 포함한 이유는 실험의 착수 시 목적으로 하는 엔도뉴클레아제를 필요로 하는 PAM 서열을 알지 못할 수도 있다는 것이다.

[0360] 발현 벡터가 표적 핵산(예를 들어, 적절한 표적 서열 및 PAM, 예컨대 CRISPR 어레이의 적어도 하나의 스페이서에 매칭되는 표적 서열을 갖는 것)을 절단할 수 있다면, CRISPR 좌위는 후보 CRISPR RNA-가이드 엔도뉴클레아제를 암호화하는 뉴클레오타이드 서열을 포함한다. 따라서, 이어서, CRISPR RNA-가이드 엔도뉴클레아제를 암호화하는 CRISPR 좌위로부터 오픈 리딩 프레임을 동정할 수 있다. 일부 경우에, 이전에 미공지 CRISPR RNA-가이드 엔도뉴클레아제, 그리고 따라서 일부 경우에, 공지된 CRISPR RNA-가이드 엔도뉴클레아제 폴리펩타이드의 아미노산 서열에 대해 20% 미만의 아미노산 서열 동정성(예를 들어, 15% 미만, 10% 미만, 5% 미만의 아미노산 서열 동일성)을 갖는 동정된 폴리펩타이드를 동정하는 것이 바람직하다.

[0361] 개시내용의 비제한적 양상의 예

[0362] 상기 기재한 본 대상의 실시형태를 포함하는 양상은 단독으로 또는 하나 이상의 다른 양상 또는 실시형태와 조합하여 유리하게 될 수 있다. 앞서 언급한 설명을 제한하는 일 없이, 1 내지 123으로 넘버링된 개시내용의 특정 비제한적 양상을 이하에 제공한다. 본 개시내용을 읽을 때 당업자에게 명백할 바와 같이, 각각의 개별적으로 넘버링된 양상은 임의의 앞서 선행하는 또는 뒤에 오는 개개로 넘버링된 양상과 함께 사용되거나 또는 조합될 수 있다. 이는 모든 이러한 조합에 대한 근거를 제공하는 것으로 의도되며, 이하에 명확하게 제공되는 양상의 조합으로 제한되지 않는다:

[0363] 양상

[0364] 1. 조성물로서,

[0365] a) CasY 폴리펩타이드, 또는 상기 CasY 폴리펩타이드를 암호화하는 핵산 분자; 및

[0366] b) CasY 가이드 RNA, 또는 상기 CasY 가이드 RNA를 암호화하는 하나 이상의 DNA 분자를 포함하는, 조성물.

[0367] 2. 1에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 50% 이상의 동일성을 갖는 아미노산 서열을 포

합하는, 조성물.

- [0368] 3. 1 또는 2에 있어서, 상기 CasY 가이드 RNA는 서열번호 11 내지 15 중 임의의 하나에 제시된 crRNA 서열과 80% 이상의 동일성을 갖는 뉴클레오타이드 서열을 포함하는, 조성물.
- [0369] 4. 1 또는 2에 있어서, 상기 CasY 폴리펩타이드는 NLS 서열에 융합된, 조성물.
- [0370] 5. 1 내지 4 중 어느 하나에 있어서, 상기 조성물은 지질을 포함하는, 조성물.
- [0371] 6. 1 내지 4 중 어느 하나에 있어서, a) 및 b)는 리포솜 내에 있는, 조성물.
- [0372] 7. 1 내지 4 중 어느 하나에 있어서, a) 및 b)는 입자 내에 있는, 조성물.
- [0373] 8. 1 내지 7 중 어느 하나에 있어서, 완충제, 뉴클레아제 저해제 및 프로테아제 저해제 중 하나 이상을 포함하는, 조성물.
- [0374] 9. 1 내지 8 중 어느 하나에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 85% 이상의 동일성을 갖는 아미노산 서열을 포함하는, 조성물.
- [0375] 10. 1 내지 9 중 어느 하나에 있어서, 상기 CasY 폴리펩타이드는 이중-가닥 표적 핵산 분자의 하나의 가닥만을 절단할 수 있는 틱내기효소인, 조성물.
- [0376] 11. 1 내지 9 중 어느 하나에 있어서, 상기 CasY 폴리펩타이드는 촉매적으로 비활성인 CasY 폴리펩타이드(dCasY)인, 조성물.
- [0377] 12. 10 또는 11에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1의 D672, E769 및 D935로부터 선택된 것에 대응하는 위치에서 하나 이상의 돌연변이를 포함하는, 조성물.
- [0378] 13. 1 내지 12 중 어느 하나에 있어서, DNA 공여자 주형을 더 포함하는, 조성물.
- [0379] 14. 이중성 폴리펩타이드에 융합된 CasY 폴리펩타이드를 포함하는, CasY 융합 폴리펩타이드.
- [0380] 15. 14에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 50% 이상의 동일성을 갖는 아미노산 서열을 포함하는, CasY 융합 폴리펩타이드.
- [0381] 16. 14에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 85% 이상의 동일성을 갖는 아미노산 서열을 포함하는, CasY 융합 폴리펩타이드.
- [0382] 17. 14 내지 16 중 어느 하나에 있어서, 상기 CasY 폴리펩타이드는 이중-가닥 표적 핵산 분자의 하나의 가닥만을 절단할 수 있는 틱내기효소인, CasY 융합 폴리펩타이드.
- [0383] 18. 14 내지 17 중 어느 하나에 있어서, 상기 CasY 폴리펩타이드는 촉매적으로 비활성인 CasY 폴리펩타이드(dCasY)인, CasY 융합 폴리펩타이드.
- [0384] 19. 17 또는 18에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1의 D672, E769 및 D935로부터 선택된 것에 대응하는 위치에서 하나 이상의 돌연변이를 포함하는, CasY 융합 폴리펩타이드.
- [0385] 20. 14 내지 19 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 상기 CasY 폴리펩타이드의 N-말단 및/또는 C-말단에 융합된, CasY 융합 폴리펩타이드.
- [0386] 21. 14 내지 20 중 어느 하나에 있어서, NLS를 포함하는, CasY 융합 폴리펩타이드.
- [0387] 22. 14 내지 21 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 표적 세포 또는 표적 세포 유형 상의 세포 표면 모이어티에 대한 결합을 제공하는 표적화 폴리펩타이드인, CasY 융합 폴리펩타이드.
- [0388] 23. 14 내지 21 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 표적 DNA를 변형시키는 효소 활성을 나타내는, CasY 융합 폴리펩타이드.
- [0389] 24. 23에 있어서, 상기 이중성 폴리펩타이드는 뉴클레아제 활성, 메틸트랜스퍼라제 활성, 데메틸라제 활성, DNA 수선 활성, DNA 손상 활성, 탈아미노화 활성, 디스무타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 재조합효소 활성, 중합효소 활성, 리가제 활성

성, 헬리카제 활성, 광분해효소 활성 및 글리코실라제 활성으로부터 선택된 하나 이상의 효소 활성을 나타내는, CasY 융합 폴리펩타이드.

- [0390] 25. 24에 있어서, 상기 이중성 폴리펩타이드는 뉴클레아제 활성, 메틸트랜스퍼라제 활성, 데메틸라제 활성, 탈아미노화 활성, 탈퓨린화 활성, 인테그라제 활성, 트랜스포사제 활성 및 재조합효소 활성으로부터 선택된 하나 이상의 효소 활성을 나타내는, CasY 융합 폴리펩타이드.
- [0391] 26. 14 내지 21 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 표적 핵산과 회합된 표적 폴리펩타이드를 변형시키는 효소 활성을 나타내는, CasY 융합 폴리펩타이드.
- [0392] 27. 26에 있어서, 상기 이중성 폴리펩타이드는 히스톤 변형 활성을 나타내는, CasY 융합 폴리펩타이드.
- [0393] 28. 26 또는 27에 있어서, 상기 이중성 폴리펩타이드는 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스포타제 활성, 유비퀴틴 리가제 활성, 데유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모일화 활성, 탈수모일화활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성, 탈미리스토일화 활성, 글리코실화 활성(예를 들어, O-GlcNAc 트랜스퍼라제) 및 탈글리코실화 활성으로부터 선택된 하나 이상의 효소 활성을 나타내는, CasY 융합 폴리펩타이드.
- [0394] 29. 28에 있어서, 상기 이중성 폴리펩타이드는 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성 및 데아세틸라제 활성으로부터 선택된 하나 이상의 효소 활성을 나타내는, CasY 융합 폴리펩타이드.
- [0395] 30. 14 내지 21 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 엔도솜 탈출 폴리펩타이드인, CasY 융합 폴리펩타이드.
- [0396] 31. 30에 있어서, 상기 엔도솜 탈출 폴리펩타이드는 하기로부터 선택된 아미노산 서열을 포함하는, CasY 융합 폴리펩타이드: GLFXALLXLLXSLWXLXLA(서열번호 94), 및 GLFHALLHLLHSLWHLLXLA(서열번호 95), 여기서, 각각의 X는 라이신, 히스티딘 및 알기닌으로부터 독립적으로 선택된다.
- [0397] 32. 14 내지 21 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 엽록체 수송 펩타이드인, CasY 융합 폴리펩타이드.
- [0398] 33. 32에 있어서, 상기 엽록체 수송 펩타이드는
- MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRV
KCMQVWPPIGKKKFETLSYLPPLTRDSRA (서열번호 83);
MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS (SEQ ID
NO:84);
MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQVWPPIEK
KKFETLSYLPDLTDSGGRVNC (서열번호 85);
MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRLPL
KVMSSVSTAC (서열번호 86);
MAQVSRICNGVWNPISLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRLPL
KVMSSVSTAC (서열번호 87);
MAQINNMAQGIQTLNPNISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRLPL
KVMSSVSTAC (서열번호 88);
MAALVTSQVLTSGTVLSTDRFRPQGLRPNPADAALGMRTVGASAAPKQSRKPHRFDRR
CLSMVV (서열번호 89);
MAALTTSQVLTSGTVLSTDRFRPQGLRPNPADAALGMRTVGASAAPKQSRKPHRFDRR
CLSMVV (서열번호 90);
MASSVLSAASAVATRSNVAQANMVAPFTGLKSAASFVSRKQNLDTISIASNGGRVQC (서열번호
91);
MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTVKCSAAVTPQASPVISRSAAAA
(서열번호 92); 및
MGAAATSMQSLKFSNRLVPPSRRLSPVNNVTCNNLPKSAAPVRTVKCCASSWNSTINGAAATT
NGASAASS (서열번호 93)
- 으로부터 선택된 아미노산 서열을 포함하는, CasY 융합 폴리펩타이드.

- [0399] 34. 14 내지 21 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 전사를 증가 또는 감소시키는 단백질인, CasY 융합 폴리펩타이드.
- [0400] 35. 34에 있어서, 상기 이중성 폴리펩타이드는 전사 리프레서 도메인인, CasY 융합 폴리펩타이드.
- [0401] 36. 34에 있어서, 상기 이중성 폴리펩타이드는 전사 활성화 도메인인, CasY 융합 폴리펩타이드.
- [0402] 37. 14 내지 21 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 단백질 결합 도메인인, CasY 융합 폴리펩타이드.
- [0403] 38. 14 내지 37 중 어느 하나의 CasY 융합 폴리펩타이드를 암호화하는 핵산 분자.
- [0404] 39. 38에 있어서, 상기 CasY 융합 폴리펩타이드를 암호화하는 상기 뉴클레오타이드 서열은 프로모터에 작동 가능하게 연결된, 핵산 분자.
- [0405] 40. 39에 있어서, 상기 프로모터는 진핵 세포에서 기능성인, 핵산 분자.
- [0406] 41. 40에 있어서, 상기 프로모터는 식물 세포, 진균 세포, 동물 세포, 무척추동물의 세포, 파리 세포, 척추동물의 세포, 포유류 세포, 영장류 세포, 비-인간 영장류 세포 및 인간 세포 중 하나 이상에서 기능성인, 핵산 분자.
- [0407] 42. 39 내지 41 중 어느 하나에 있어서, 상기 프로모터는 구성적 프로모터, 유도성 프로모터, 세포 유형 특이적 프로모터 및 조직-특이적 프로모터 중 하나 이상인, 핵산 분자.
- [0408] 43. 38 내지 42 중 어느 하나에 있어서, 상기 DNA 분자는 재조합 발현 벡터인, 핵산 분자.
- [0409] 44. 43에 있어서, 상기 재조합 발현 벡터는 재조합 아데노연관 바이러스 벡터, 재조합 레트로바이러스 벡터 또는 재조합 렌티바이러스 벡터인, 핵산 분자.
- [0410] 45. 39에 있어서, 상기 프로모터는 원핵 세포에서 기능성인, 핵산 분자.
- [0411] 46. 38에 있어서, 상기 핵산 분자는 mRNA인, 핵산 분자.
- [0412] 47. 하나 이상의 핵산 분자로서,
- [0413] (a) CasY 가이드 RNA; 및
- [0414] (b) CasY 폴리펩타이드를 암호화하는, 하나 이상의 핵산 분자.
- [0415] 48. 47에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 50% 이상의 동일성을 갖는 아미노산 서열을 포함하는, 하나 이상의 핵산 분자.
- [0416] 49. 47에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 85% 이상의 동일성을 갖는 아미노산 서열을 포함하는, 하나 이상의 핵산 분자.
- [0417] 50. 47 내지 49 중 어느 하나에 있어서, 상기 CasY 가이드 RNA는 서열번호 11 내지 15 중 임의의 하나에 제시된 crRNA 서열과 80% 이상의 동일성을 갖는 뉴클레오타이드 서열을 포함하는, 하나 이상의 핵산 분자.
- [0418] 51. 47 내지 50 중 어느 하나에 있어서, 상기 CasY 폴리펩타이드는 NLS 서열에 융합된, 하나 이상의 핵산 분자.
- [0419] 52. 47 내지 51 중 어느 하나에 있어서, 상기 하나 이상의 핵산 분자는 프로모터에 작동 가능하게 연결된 상기 CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는, 하나 이상의 핵산 분자.
- [0420] 53. 47 내지 52 중 어느 하나에 있어서, 상기 하나 이상의 핵산 분자는 프로모터에 작동 가능하게 연결된 상기 CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 포함하는, 하나 이상의 핵산 분자.
- [0421] 54. 52 또는 53에 있어서, 상기 CasY 가이드 RNA를 암호화하는 상기 뉴클레오타이드 서열에 작동 가능하게 연결된 상기 프로모터, 및/또는 상기 CasY 폴리펩타이드를 암호화하는 상기 뉴클레오타이드 서열에 작동 가능하게 연결된 상기 프로모터는 진핵 세포에서 기능성인, 하나 이상의 핵산 분자.
- [0422] 55. 54에 있어서, 상기 프로모터는 식물 세포, 진균 세포, 동물 세포, 무척추동물의 세포, 파리 세포, 척추

동물의 세포, 포유류 세포, 영장류 세포, 비-인간 영장류 세포 및 인간 세포 중 하나 이상에서 기능성인, 하나 이상의 핵산 분자.

- [0423] 56. 53 내지 55 중 어느 하나에 있어서, 상기 프로모터는 구성적 프로모터, 유도성 프로모터, 세포 유형 특이적 프로모터 및 조직-특이적 프로모터 중 하나 이상인, 하나 이상의 핵산 분자.
- [0424] 57. 47 내지 56 중 어느 하나에 있어서, 상기 하나 이상의 핵산 분자는 하나 이상의 재조합 발현 벡터인, 하나 이상의 핵산 분자.
- [0425] 58. 57에 있어서, 상기 하나 이상의 재조합 발현 벡터는 하나 이상의 아테노연관 바이러스 벡터, 하나 이상의 재조합 레트로바이러스 벡터, 또는 하나 이상의 재조합 렌티바이러스 벡터로부터 선택된, 하나 이상의 핵산 분자.
- [0426] 59. 53에 있어서, 상기 프로모터는 원핵 세포에서 기능성인, 하나 이상의 핵산 분자.
- [0427] 60. 진핵 세포로서,
- [0428] a) CasY 폴리펩타이드, 또는 상기 CasY 폴리펩타이드를 암호화하는 핵산 분자,
- [0429] b) CasY 융합 폴리펩타이드, 또는 상기 CasY 융합 폴리펩타이드를 암호화하는 핵산 분자, 및
- [0430] c) CasY 가이드 RNA, 또는 상기 CasY 가이드 RNA를 암호화하는 핵산 분자
- [0431] 중 하나 이상을 포함하는, 진핵 세포.
- [0432] 61. 60에 있어서, 상기 CasY 폴리펩타이드를 암호화하는 핵산 분자를 포함하되, 상기 핵산 분자는 상기 세포의 상기 게놈 DNA 내로 통합된, 진핵 세포.
- [0433] 62. 60 또는 61에 있어서, 상기 진핵 세포는 식물 세포, 포유류 세포, 곤충 세포, 거미류 세포, 진균 세포, 조류 세포, 파충류 세포, 양서류 세포, 무척추동물 세포, 마우스 세포, 래트 세포, 영장류 세포, 비-인간 영장류 세포, 또는 인간 세포인, 진핵 세포.
- [0434] 63. CasY 융합 폴리펩타이드, 또는 상기 CasY 융합 폴리펩타이드를 암호화하는 핵산 분자를 포함하는 세포.
- [0435] 64. 63에 있어서, 상기 세포는 원핵 세포인, 세포.
- [0436] 65. 63 또는 64에 있어서, 상기 CasY 융합 폴리펩타이드를 암호화하는 핵산 분자를 포함하되, 상기 핵산 분자는 상기 세포의 상기 게놈 DNA 내로 통합된, 세포.
- [0437] 66. 표적 핵산의 변형 방법으로서, 상기 표적 핵산을:
- [0438] a) CasY 폴리펩타이드; 및
- [0439] b) 상기 표적 핵산의 표적 서열에 혼성화하는 가이드 서열을 포함하는 CasY 가이드 RNA와 접촉시키는 단계를 포함하되,
- [0440] 상기 접촉시키는 단계는 상기 CasY 폴리펩타이드에 의해 상기 표적 핵산의 변형을 초래하는, 표적 핵산의 변형 방법.
- [0441] 67. 66에 있어서, 상기 변형은 상기 표적 핵산의 절단인, 방법.
- [0442] 68. 66 또는 67에 있어서, 상기 표적 핵산은 이중 가닥 DNA, 단일 가닥 DNA, RNA, 게놈 DNA, 및 염색체의 DNA로부터 선택된, 방법.
- [0443] 69. 66 내지 68 중 어느 하나에 있어서, 상기 접촉시키는 단계는 세포 외부의 시험관내에서 일어나는, 방법.
- [0444] 70. 66 내지 68 중 어느 하나에 있어서, 상기 접촉시키는 단계는 배양물 내 세포의 내부에서 일어나는, 방법.
- [0445] 71. 66 내지 68 중 어느 하나에 있어서, 상기 접촉시키는 단계는 생체내 세포의 내부에서 일어나는, 방법.
- [0446] 72. 70 또는 71에 있어서, 상기 세포는 진핵 세포인, 방법.
- [0447] 73. 72에 있어서, 상기 세포는 식물 세포, 진균 세포, 포유류 세포, 파충류 세포, 곤충 세포, 조류 세포,

어류 세포, 기생충 세포, 절지동물 세포, 무척추동물의 세포, 척추동물의 세포, 설치류 세포, 마우스 세포, 래트 세포, 영장류 세포, 비-인간 영장류 세포, 및 인간 세포로부터 선택된, 방법.

- [0448] 74. 70 또는 71에 있어서, 상기 세포는 원핵 세포인, 방법.
- [0449] 75. 66 내지 74 중 어느 하나에 있어서, 상기 접촉시키는 단계는 게놈 편집을 초래하는, 방법.
- [0450] 76. 66 내지 75 중 어느 하나에 있어서, 상기 접촉시키는 단계는 세포 내로: (a) 상기 CasY 폴리펩타이드, 또는 상기 CasY 폴리펩타이드를 암호화하는 핵산 분자, 및 (b) 상기 CasY 가이드 RNA, 또는 상기 CasY 가이드 RNA를 암호화하는 핵산 분자를 도입하는 것을 포함하는, 방법.
- [0451] 77. 76에 있어서, 상기 접촉시키는 단계는 DNA 공여자 주형을 상기 세포 내로 도입하는 것을 더 포함하는, 방법.
- [0452] 78. 66 내지 77 중 어느 하나에 있어서, 상기 CasY 가이드 RNA는 서열번호 11 내지 15 중 임의의 하나에 제시된 crRNA 서열과 80% 이상의 동일성을 갖는 뉴클레오타이드 서열을 포함하는, 방법.
- [0453] 79. 66 내지 78 중 어느 하나에 있어서, 상기 CasY 폴리펩타이드는 NLS 서열에 융합된, 방법.
- [0454] 80. 표적 DNA로부터 전사를 조절하거나, 표적 핵산을 변형시키거나, 또는 표적 핵산과 회합된 단백질을 변형시키는 방법으로서, 상기 표적 핵산을,
- [0455] a) 이중성 폴리펩타이드에 융합된 CasY 폴리펩타이드를 포함하는 CasY 융합 폴리펩타이드; 및
- [0456] b) 상기 표적 핵산의 표적 서열에 혼성화하는 가이드 서열을 포함하는 CasY 가이드 RNA
- [0457] 와 접촉시키는 단계를 포함하는, 방법.
- [0458] 81. 80에 있어서, 상기 CasY 가이드 RNA는 서열번호 11 내지 15 중 임의의 하나에 제시된 crRNA 서열과 80% 이상의 동일성을 갖는 뉴클레오타이드 서열을 포함하는, 방법.
- [0459] 82. 80 또는 81에 있어서, 상기 CasY 융합 폴리펩타이드는 NLS 서열을 포함하는, 방법.
- [0460] 83. 80 내지 82 중 어느 하나에 있어서, 상기 변형은 상기 표적 핵산의 절단이 아닌, 방법.
- [0461] 84. 80 또는 83 중 어느 하나에 있어서, 상기 표적 핵산은 이중 가닥 DNA, 단일 가닥 DNA, RNA, 게놈 DNA, 및 염색체의 DNA로부터 선택된, 방법.
- [0462] 85. 80 내지 84 중 어느 하나에 있어서, 상기 접촉시키는 단계는 세포 외부의 시험관내에서 일어나는, 방법.
- [0463] 86. 80 내지 84 중 어느 하나에 있어서, 상기 접촉시키는 단계는 배양물 내 세포의 내부에서 일어나는, 방법.
- [0464] 87. 80 내지 84 중 어느 하나에 있어서, 상기 접촉시키는 단계는 생체내 세포의 내부에서 일어나는, 방법.
- [0465] 88. 86 또는 87에 있어서, 상기 세포는 진핵 세포인, 방법.
- [0466] 89. 88에 있어서, 상기 세포는 식물 세포, 진균 세포, 포유류 세포, 파충류 세포, 곤충 세포, 조류 세포, 어류 세포, 기생충 세포, 절지동물 세포, 무척추동물의 세포, 척추동물의 세포, 설치류 세포, 마우스 세포, 래트 세포, 영장류 세포, 비-인간 영장류 세포, 및 인간 세포로부터 선택된, 방법.
- [0467] 90. 86 또는 87에 있어서, 상기 세포는 원핵 세포인, 방법.
- [0468] 91. 80 내지 90 중 어느 하나에 있어서, 상기 접촉시키는 단계는 세포 내로: (a) 상기 CasY 융합 폴리펩타이드, 또는 상기 CasY 융합 폴리펩타이드를 암호화하는 핵산 분자, 및 (b) 상기 CasY 가이드 RNA, 또는 상기 CasY 가이드 RNA를 암호화하는 핵산 분자를 도입하는 것을 포함하는, 방법.
- [0469] 92. 80 내지 91 중 어느 하나에 있어서, 상기 CasY 폴리펩타이드는 촉매적으로 비활성인 CasY 폴리펩타이드 (dCasY)인, 방법.
- [0470] 93. 80 또는 92에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1의 D672, E769 및 D935로부터 선택된 것에 대응하는 위치에서 하나 이상의 돌연변이를 포함하는, 방법.
- [0471] 94. 80 내지 93 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 표적 DNA를 변형시키는 효소 활성을

나타내는, 방법.

- [0472] 95. 94에 있어서, 상기 이중성 폴리펩타이드는 뉴클레아제 활성, 메틸트랜스퍼라제 활성, 데메틸라제 활성, DNA 수선 활성, DNA 손상 활성, 탈아미노화 활성, 디스무타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 재조합효소 활성, 중합효소 활성, 리가제 활성, 헬리카제 활성, 광분해효소 활성 및 글리코실라제 활성으로부터 선택된 하나 이상의 효소 활성을 나타내는, 방법.
- [0473] 96. 95에 있어서, 상기 이중성 폴리펩타이드는 뉴클레아제 활성, 메틸트랜스퍼라제 활성, 데메틸라제 활성, 탈아미노화 활성, 탈퓨린화 활성, 인테그라제 활성, 트랜스포사제 활성 및 재조합효소 활성으로부터 선택된 하나 이상의 효소 활성을 나타내는, 방법.
- [0474] 97. 80 내지 93 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 표적 핵산과 회합된 표적 폴리펩타이드를 변형시키는 효소 활성을 나타내는, 방법.
- [0475] 98. 97에 있어서, 상기 이중성 폴리펩타이드는 히스톤 변형 활성을 나타내는, 방법.
- [0476] 99. 97 또는 98에 있어서, 상기 이중성 폴리펩타이드는 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스포타제 활성, 유비퀴틴 리가제 활성, 데유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, 수모일화 활성, 탈수모일화활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성, 탈미리스토일화 활성, 글리코실화 활성(예를 들어, O-GlcNAc 트랜스퍼라제) 및 탈글리코실화 활성으로부터 선택된 하나 이상의 효소 활성을 나타내는, 방법.
- [0477] 100. 99에 있어서, 상기 이중성 폴리펩타이드는 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성 및 데아세틸라제 활성으로부터 선택된 하나 이상의 효소 활성을 나타내는, 방법.
- [0478] 101. 80 내지 93 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 전사를 증가 또는 감소시키는 단백질인, 방법.
- [0479] 102. 101에 있어서, 상기 이중성 폴리펩타이드는 전사 리프레서 도메인인, 방법.
- [0480] 103. 101에 있어서, 상기 이중성 폴리펩타이드는 전사 활성화 도메인인, 방법.
- [0481] 104. 80 내지 93 중 어느 하나에 있어서, 상기 이중성 폴리펩타이드는 단백질 결합 도메인인, 방법.
- [0482] 105. 유전자이식, 다세포, 비인간 유기체로서, 계놈이,
- [0483] a) CasY 폴리펩타이드,
- [0484] b) CasY 융합 폴리펩타이드 및
- [0485] c) CasY 가이드 RNA
- [0486] 중 하나 이상을 암호화하는 뉴클레오타이드 서열을 포함하는 이식유전자를 포함하는, 유전자이식, 다세포, 비인간 유기체.
- [0487] 106. 105에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 50% 이상의 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는, 유전자이식, 다세포, 비인간 유기체.
- [0488] 107. 105에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 85% 이상의 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는, 유전자이식, 다세포, 비인간 유기체.
- [0489] 108. 105 내지 107 중 어느 하나에 있어서, 상기 유기체는 식물, 외떡잎식물, 쌍떡잎식물, 무척추동물, 곤충, 절지동물, 거미류, 기생충, 벌레, 강장동물, 척추동물, 어류, 파충류, 양서류, 유체류, 조류, 돼지, 말, 양, 설치류, 마우스, 랫트 또는 비 인간 영장류인, 유전자이식, 다세포, 비인간 유기체.
- [0490] 109. 시스템으로서,
- [0491] a) CasY 폴리펩타이드 및 CasY 가이드 RNA;
- [0492] b) CasY 폴리펩타이드, CasY 가이드 RNA, 및 DNA 공여자 주형;

- [0493] c) CasY 융합 폴리펩타이드 및 CasY 가이드 RNA;
- [0494] d) CasY 융합 폴리펩타이드, CasY 가이드 RNA, 및 DNA 공여자 주형;
- [0495] e) CasY 폴리펩타이드를 암호화하는 mRNA, 및 CasY 가이드 RNA;
- [0496] f) CasY 폴리펩타이드를 암호화하는 mRNA; CasY 가이드 RNA, 및 DNA 공여자 주형;
- [0497] g) CasY 융합 폴리펩타이드를 암호화하는 mRNA, 및 CasY 가이드 RNA;
- [0498] h) CasY 융합 폴리펩타이드를 암호화하는 mRNA, CasY 가이드 RNA, 및 DNA 공여자 주형;
- [0499] i) i) CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; 및 ii) CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 하나 이상의 재조합 발현 벡터;
- [0500] j) i) CasY 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; ii) CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열; 및 iii) DNA 공여자 주형을 포함하는 하나 이상의 재조합 발현 벡터;
- [0501] k) i) CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; 및 ii) CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열을 포함하는 하나 이상의 재조합 발현 벡터; 및
- [0502] l) i) CasY 융합 폴리펩타이드를 암호화하는 뉴클레오타이드 서열; ii) CasY 가이드 RNA를 암호화하는 뉴클레오타이드 서열; 및 DNA 공여자 주형을 포함하는 하나 이상의 재조합 발현 벡터
- [0503] 를 포함하는, 시스템.
- [0504] 110. 109에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 50% 이상의 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는, CasY 시스템.
- [0505] 111. 109에 있어서, 상기 CasY 폴리펩타이드는 서열번호 1 또는 서열번호 2에 제시된 아미노산 서열(또는 서열번호 1 내지 8 중 임의의 하나에 제시된 아미노산 서열)에 대해 85% 이상의 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는, CasY 시스템.
- [0506] 112. 109 내지 111 중 어느 하나에 있어서, 상기 공여자 주형 핵산은 길이가 8개의 뉴클레오타이드 내지 1000개의 뉴클레오타이드인, CasY 시스템.
- [0507] 113. 109 내지 111 중 어느 하나에 있어서, 상기 공여자 주형 핵산은 길이가 25개의 뉴클레오타이드 내지 500개의 뉴클레오타이드인, CasY 시스템.
- [0508] 114. 109 내지 113 중 어느 하나의 CasY 시스템을 포함하는 키트.
- [0509] 115. 114에 있어서, 상기 키트의 부품은 동일한 용기에 있는, 키트.
- [0510] 116. 114에 있어서, 상기 키트의 부품은 별개의 용기에 있는, 키트.
- [0511] 117. 109 내지 116 중 어느 하나의 CasY 시스템을 포함하는 멸균 용기.
- [0512] 118. 117에 있어서, 상기 용기는 주사기인, 멸균 용기.
- [0513] 119. 109 내지 116 중 어느 하나의 CasY 시스템을 포함하는 이식 가능한 장치.
- [0514] 120. 119에 있어서, 상기 CasY 시스템은 기질 내인, 이식 가능한 장치.
- [0515] 121. 119에 있어서, 상기 CasY 시스템은 저장소 내인, 이식 가능한 장치.
- [0516] 122. CRISPR RNA-가이드 엔도뉴클레아제의 동정 방법으로서,
- [0517] 복수의 메타게놈 뉴클레오타이드 서열에서, Cas1 폴리펩타이드를 암호화하는 뉴클레오타이드 서열을 검출하는 단계;
- [0518] 상기 Cas1-암호화 뉴클레오타이드 서열 근처에서 CRISPR 어레이를 검출하는 단계;
- [0519] 재조합 CRISPR 좌위 발현 벡터를 생성하기 위해 상기 검출된 CRISPR 어레이를 포함하는 CRISPR 좌위를, 복수의 메타게놈 뉴클레오타이드 서열이 유래된 핵산 샘플로부터 발현 벡터 내로 클로닝하는 단계;

- [0520] 상기 재조합 CRISPR 좌위 발현 벡터를 표적 핵산을 절단하는 능력에 대해 분석하는 단계로서, 표적 핵산을 절단하는 능력을 갖는 CRISPR 좌위는 CRISPR RNA-가이드 엔도뉴클레아제를 암호화하는 뉴클레오타이드 서열을 포함하는, 상기 분석하는 단계;
- [0521] 상기 CRISPR 좌위에서, 공지된 CRISPR RNA-가이드 엔도뉴클레아제 폴리펩타이드의 상기 아미노산 서열에 대해 20% 미만의 아미노산 서열 동일성을 갖는 폴리펩타이드를 암호화하는 오픈 리딩 프레임을 동정하는 단계를 포함하는, 방법.
- [0522] 123. 122에 있어서, 상기 분석하는 단계는 상기 재조합 CRISPR 좌위 발현 벡터 및 표적 핵산을 세포 내로 도입하는 것을 포함하는, 방법.
- [0523] 실시예
- [0524] 다음의 실시예는 당업자에게 본 발명의 제조하고 사용하는 방법의 완전한 개시내용 및 설명을 제공하기 위해 제시하며, 본 발명자들이 본 발명으로서 간주하는 것의 범주를 제한하는 것으로 의도되지도 또는 이하의 실험이 모든 또는 수행되는 실험만을 나타내는 것으로 의도하지도 않는다. 사용한 숫자(예를 들어, 양, 온도 등)에 관한 정확성을 보장하기 위한 노력을 하였지만, 일부 실험적 오차 및 편차를 고려하여야 한다. 달리 나타내지 않는 한, 부분은 중량부이고, 분자량은 중량 평균 분자량이며, 온도는 섭씨이고, 압력은 대기압에서 또는 대기압 근처이다. 표준 약어, 예를 들어, bp, 염기쌍(들); kb, 킬로베이스(들); pl, 피코리터(들); s 또는 sec, 초(들); min, 분(들); h 또는 hr, 시간(들); aa, 아미노산(들); kb, 킬로베이스(들); bp, 염기쌍(들); nt, 뉴클레오타이드(들); i.m., 근육내(로); i.p., 복강내(로); s.c., 피하(로); 등.
- [0525] 실시예 1
- [0526] 본 명세서에 기재된 작업은 지하수, 퇴적물 및 산성 광산 배수로부터의 미생물 군집의 메타게놈 샘플 분석을 포함한다. 배양 유기체 중에서 나타나지 않은 뉴 클래스 2 CRISPR-Cas 시스템을 동정하였다.
- [0527] **도 3. CasY 도메인 및 유사성 검색.** (패널 a) HHpred를 이용하여 AcCpf1이 있는 원위 상동체 정렬로부터 추론되는 CasY에 대한 개략적 도메인 표현. 보존된 촉매적 잔기를 단백질 위의 적색 막대로 표시한다. CasY는 C-말단 영역에서 RuvC 분할 도메인(RuvC-I, RuvC-II 및 RuvC-III), 및 거대 신규한 N-말단 도메인을 함유한다. 이하의 개략도는 다음의 검색에 기반하여 상부 히트를 나타낸다: (1) NCBI(모델 및 환경 단백질을 포함하는 NR 데이터베이스)에서 모든 단백질에 대한 BLAST 검색. (2) 문헌[Makarova et al. Nat Rev Microbiol. 2015 Nov;13(11):722-36, 및 Shmakov et al. Mol Cell. 2015 Nov 5;60(3):385-97)]에서 기재한 모든 Cas 단백질을 이용하여 구성된 모델에 기반한 프로파일 은닉 마르코프 모델(hidden markov model: HMM) 검색. (3) HHpred에 기반한 원위 상동체 검색. 히트는 그들의 유의도에 기반하여 색-암호화하고, 히트 범위 및 E-값을 제공한다. 특히, CasY는 국소 히트만을 가졌다. CasY의 812 N-말단 아미노산은 하나의 매우 작은 부분적 히트만을 가졌다. 종합하면, 이들 발견은 CasY가 새로운 Cas 단백질이라는 것을 나타낸다. (패널 b) 상이한 CasY-함유 CRISPR 좌위 스캐폴드를 서열 데이터로부터 구성하였다.
- [0528] 실시예 2
- [0529] **도 4. CasY 및 C2c3 좌위의 다이어그램의 개략적 다이어그램.** 간접 단백질을 녹색으로 나타내고, 획득 단백질을 적색으로 나타낸다. RNA 구조를 이용하여 폴딩된 반복부는 5' 단부에서 강한 헤어핀을 나타내는 우측에 나타나는데, 이는 CRISPR 어레이의 자기 가공을 시사한다.
- [0530] **도 5 (패널 a 내지 d) CasY에 의한 PAM 의존적 플라스미드 간접.** (패널 a) CasY를 이용하여 PAM 고갈 분석을 수행하였다. CasY CRISPR 좌위를 함유하는 이콜라이를 표적 서열의 7개 뉴클레오타이드 무작위화된 5' 또는 3'을 갖는 플라스미드 라이브러리로 형질전환시켰다. 표적 플라스미드를 선택하고 형질전환체를 모았다. 무작위화된 영역을 증폭시키고 나서, 심층 서열분석을 위해 제조하였다. 고갈된 서열을 동정하고 나서, PAM 로고를 생성하는 데 사용하였다. (패널 b) CasY.1에 대해 생성된 PAM 로고는 표적의 5'-TA-3' 축적 서열 5'을 함유하는 서열에 대한 강한 선호도를 나타내었다. 3' PAM은 검출되지 않았다. (패널 c) PAM 고갈 분석으로부터 결정된 PAM을 확인하기 위해 4개의 상이한 PAM을 직접 분석하였다. (패널 d) CasY.2에 대해 생성된 PAM 로고는 표적의 5'-YR-3' 및/또는 5'-TR-3'(예를 들어, 5'-DTR-3')(각각 더 낮은 역치 및 더 높은 역치) 축적 서열 5'을 함유하는 서열에 대한 선호도를 나타내었다(여기서, Y는 T 또는 C이고; R은 A 또는 G이며; D는 A, G 또는 T임). 3' PAM은 검출되지 않았다.
- [0531] **도 6. (패널 a) 천연 유래 CasY 가이드 RNA로부터의 '반복' 서열(CasY 좌위 Y1-Y6에 대해).** (패널 b) CasY

RNA 가이드 DNA 절단의 다이어그램. CasY 단백질은 반복 서열(검정색, 반복부; 적색, 스페이서)에서 crRNA(CasY 가이드 RNA)에 결합한다. 정확한 프로토스페이서 인접 모티프(PAM)를 함유하는 표적 서열(청색)에 대한 가이드 RNA의 가이드 서열의 염기쌍은 표적 DNA의 이중 가닥 절단을 초래한다.

[0532] 실시예 3: 비배양 미생물로부터의 새로운 CRISPR-Cas 시스템

[0533] CRISPR-Cas 적응 면역계는 부위-특이적 DNA 절단을 할 수 있는 프로그램 가능한 효소를 제공함으로써 게놈에 혁신을 일으켰다. 그러나, 현재의 CRISPR-Cas 기술은 아직 손대지 않은 단리되지 않은 유기체로부터의 방대한 효소를 떠나서 배양 박테리아로부터의 시스템에만 기반한다. 본 명세서의 데이터는 배양-독립적 게놈-해결 메타게노믹스를 이용하여, 고세균의 생활 도메인에서 처음 보고된 Cas9를 포함하는 새로운 CRISPR-Cas 시스템의 동정을 나타낸다. 이런 분기하는 Cas9 효소는 활성 CRISPR-Cas 시스템의 부분으로서 거의 연구되지 않은 나노고세균에서 발견되었다. 박테리아에서, 2개의 이전에 알려지지 않은 시스템, 즉, 아직 동정되지 않은 가장 최신의 시스템 중에서 CRISPR-CasX 및 CRISPR-CasY가 발견되었다. 특히, 모든 필요한 기능성 성분을 메타게노믹스에 의해 동정하였는데, 이는 이콜라이에서 강한 RNA-가이드 DNA 간섭의 확인을 허용하였다. 본 명세서의 데이터는, 살아 있는 세포에서의 실험과 조합된 환경적 미생물 군집의 의문이 내용이 미생물계 생물기법의 레퍼토리를 확장하는 게놈의 전례없는 다양성에 대한 접근을 허용한다는 것을 나타낸다.

[0534] **결과**

[0535] 지하수, 퇴적물, 및 산성 광산 배수 미생물 군집으로부터의 테라베이스-규모의 메타게놈 데이터셋을 분석하였고, 배양 유기체 중에서 나타내지 않은 클래스 2 CRISPR-Cas 시스템을 탐색하였다. 도메인 고세균에서 제1 Cas9 단백질을 동정하였고, 비배양 박테리아에서 2개의 새로운 CRISPR-Cas 시스템, 즉, CRISPR-CasX 및 CRISPR-CasY를 발견하였다(도 7). 특히, 고세균 Cas9와 CasY는 둘 다 알려진 단리된 대표물이 없는 계통으로부터의 유기체 게놈에서 배타적으로 암호화되었다.

[0536] **고세균 Cas9의 제1 동정**

[0537] CRISPR-Cas9의 특징 중 하나는 박테리아 도메인에서만 그의 추정적 존재였다. 따라서 산성 광산 배수(AMD) 메타게놈 데이터셋에서 나노고세균 ARMAN-1(칸디다투스 마이크라카에움 아시디필름(*Candidatus* Micrarchaeum acidiphilum) ARMAN-1) 및 ARMAN-4(칸디다투스 파르바카에움 아시디필름(*Candidatus* Parvarchaeum acidiphilum) ARMAN-4) 게놈에서 암호화된 Cas9 단백질을 발견한 것은 놀라웠다. 이들 발견은 Cas9-함유 CRISPR 시스템의 발생을 다른 생활 도메인으로 확장시킨다.

[0538] ARMAN-4 *cas9* 유전자는 동일한 게놈 내용의 16개의 상이한 샘플에서 발견되었지만, 다른 인접한 *cas* 유전자는 없었고(25 kbp 초과로 몇몇 DNA 서열 콘텍에서 중심에 위치됨에도 불구하고), 하나의 인접한 CRISPR 반복-스페이서 단위만을 가진다(도 13). 보편적인 CRISPR 인테그라제를 암호화하는 전형적인 CRISPR 어레이 및 *cas1*의 결여는 새로운 스페이서를 획득하는 능력이 없는 시스템을 시사한다. 표적은 스페이서 서열에 대해 동정될 수 없었지만, 몇년에 걸쳐 수집한 샘플 내 좌위의 보존을 제공하였으며, "단일-표적" CRISPR-Cas 시스템에서 그의 기능은 이때에 제외할 수 없다.

[0539] 대조적으로, 15개의 상이한 샘플로부터 회수한 ARMAN-1에서 CRISPR-Cas 좌위는 *cas1*, *cas2*, *cas4* 및 *cas9* 유전자에 인접한 거대 CRISPR 어레이를 포함한다. 대체로 보존된 말단(가장 오래된 스페이서를 포함할 가능성이 있음) 및 다수의 별개의 스페이서가 혼입된 가변 영역을 갖는 수많은 대안의 ARMAN-1 CRISPR 어레이를 재구성하였다(도 8A 및 도 14). 스페이서 내용에서 이 초가변성에 기반하여, 이들 데이터는 ARMAN-1 CRISPR-Cas9 시스템이 샘플링한 집단에서 활성이라는 것을 나타낸다.

[0540] 현저하게는, ARMAN-1 CRISPR-Cas9 시스템의 추정적 스페이서 표적(프로토스페이서) 중 56개를, 그것이 고밀도의 짧은 가설 단백질을 암호화한다는 것을 고려하여 ARMAN-1 바이러스일 가능성이 있는 단일 10kbp 게놈 단편에 위치시켰다(도 8b). 사실, 극저온 전자 단층 촬영 재구성은 종종 ARMAN 세포에 부착된 바이러스 입자를 동정하였다. ARMAN-1 프로토스페이서는 또한 동일한 생태계로부터의 I-플라즈마를 포함하는, ARMAN-2(다른 나노고세균)의 게놈 내에서 추정적 트랜스포존 및 써모플라스마탈레스(*Thermoplasmatales*) 고세균의 게놈 내 추정적 이동 요소로부터 유래되었다(도 15). ARMAN과 써모플라스마탈레스 세포 사이에 직접 세포질 "브리지"가 관찰되었는데, 이는 그들 사이의 밀접한 관계를 나타낸다. 따라서 ARMAN-1 CRISPR-Cas9는 진핵 생식 계통에서 piRNA-매개 방어를 연상시키는 역할인 이들 유기체 사이의 트랜스포존 증식에 대해 방어될 수 있다.

[0541] 활성 DNA-표적화 CRISPR-Cas 시스템은 자기 대 비자기 차별에 대해 표적 서열 다음에 위치한 2 내지 4개의 bp 프로토스페이서-인접 모티프(PAM)를 사용한다. 게놈 표적 서열에 인접한 서열의 시험은 사실 ARMAN-1에서의 강

한 'NGG' PAM 선호도를 나타내었다(도 8c). Cas9는 또한 RNA-가이드 DNA 절단을 위해 2개의 별개의 전사체, 즉, CRISPR RNA(crRNA) 및 전사 촉진 CRISPR RNA(tracrRNA)를 사용한다. ARMAN-1과 ARMAN-4 CRISPR-Cas9 시스템 둘 다의 근처에서 추정적 tracrRNA를 동정하였다(도 16). 이전에, 고세균에서는 숙주 인자인 RNase III의 결여에 기인하여 II형 CRISPR 시스템이 없고, 이는 crRNA-tracrRNA 가이드 복합체 성숙을 초래하였다는 것을 시사하였다. 특히, RNase III 상동체는 ARMAN-1 게놈(95% 완전한 것으로 추정됨)에서 동정되지 않았고, 내부 프로모터는 CRISPR 어레이에 대해 예측되지 않았는데, 이는 가이드 RNA 생산의 아직 결정되지 않은 메커니즘을 시사한다. 이콜라이와 효모 둘 다로부터 정제된 ARMAN-1 및 ARMAN-4 Cas9 단백질의 절단 활성을 시험하기 위한 생화학적 실험 및 생체내 이콜라이 표적화 분석은 임의의 검출 가능한 활성을 나타내지 않았다(도 21 및 도 17 참조).

[0542] **CRISPR-CasX는 새로운 이중-RNA-가이드 CRISPR 시스템이다**

[0543] Cas9에 추가로, 클래스 2 Cas 효과기 단백질의 3가지 패밀리를 발견하였고, 실험적으로 확인하였다: Cpf1, C2c1 및 C2c2. 작은 DNA 단편 상에서만 동정된 다른 유전자인 c2c3은 또한 이러한 단백질 패밀리를 암호화하는 것으로 시사되었다. 새로운 유형의 클래스 2 CRISPR-Cas 시스템은 지하수 및 퇴적물 샘플로부터 반복적으로 회수한 2종의 박테리아의 게놈에서 발견되었다. 상이한 문에 속하는 2종의 유기체, 즉, 델타프로테오박테리아 및 부유균문에서 이 시스템의 높은 보존은 최근의 문 간의 전달을 시사한다. 이 새로 기재된 시스템은 본 명세서에서 CasX로서 지칭되는 Cas1, Cas2, Cas4 및 비특정규명된 대략 980개의 aa 단백질을 포함한다. 각각의 CasX와 회합된 CRISPR 어레이는 37개의 염기쌍의 고도로 유사한 반복부, 33 내지 34개 염기쌍의 스페이서, 및 Cas 오프론과 CRISPR 어레이 사이의 추정적 tracrRNA를 가졌다(도 7b). BLAST 검색은 CasX C-말단의 특정 영역으로 제한된 유사성으로, 트랜스포사제에 대해 약한 유사성만을 나타내었다($e\text{-값} > 1 \times 10^{-4}$). 원위 상동성 검출 및 단백질 모델링은 V형 CRISPR-Cas 시스템에서 발견된 조직화를 연상시키는 CasX C-말단 단부 근처의 RuvC 도메인을 동정하였다(도 18). CasX 단백질의 나머지(630 N-말단의 아미노산)는 임의의 공지된 단백질에 대해 검출 가능한 유사성을 나타내지 않았는데, 이는 신규한 클래스 2 효과기라는 것을 시사한다. tracrRNA와 별개의 Cas1, Cas2 및 Cas4 단백질의 조합은 V형 시스템 중에서 독특하다. 추가로, CasX는 임의의 공지된 V 단백질보다 상당히 더 작다: Cpf1, C2c1 및 C2c3에 대해 980 aa은 1,200 aa보다 더 큰 전형적인 크기와 비교됨.

[0544] 다음에, 그의 작은 크기 및 비정규 좌위 내용에도 불구하고, CasX는 Cas9 및 Cpf1 효소와 유사한 RNA-가이드 DNA 표적화를 할 수 있는지의 여부를 생각하였다. 이 가능성을 시험하기 위해, casX, 짧은 반복부-스페이서 어레이 및 간접 비암호 영역을 포함하는 최소 CRISPR-CasX 좌위를 암호화하는 플라스미드를 합성하였다. 이콜라이에서 발현될 때, 이런 최소 좌위는 메타게놈 분석에 의해 동정된 플라스미드 보유 표적 서열에 의한 형질전환을 차단하였다(도 9a 내지 도 9c, 도 19). 더 나아가, 최소-좌위에서 스페이서 서열이 플라스미드 표적 내 프로토스페이서 서열에 매칭될 때에만 형질전환에 의한 간섭이 일어났다. CasX에 대한 PAM 서열을 동정하기 위해, 표적 부위에 인접한 5' 또는 3' 무작위화 서열 중 하나를 함유하는 플라스미드를 사용하여 형질전환 분석을 이콜라이에서 반복하였다. 이 분석은 프로토스페이서 서열의 5' 바로 옆에 위치한 서열 'TTCN'에 대한 엄격한 선호도를 나타내었다(도 9d). 3' PAM 선호도는 관찰되지 않았다(도 19). 이 발견과 일치되게, 'TTCA'는 환경 샘플에서 동정한 추정적 델타프로테오박테리아 CRISPR-CasX 프로토스페이서 상류에서 발견된 서열이었다. 특히, CRISPR-CasX 좌위는 동일한 PAM 서열을 공유하는데, 이는 그들의 높은 정도의 CasX 단백질 상동성과 연결된다.

[0545] 단일-RNA와 이중-RNA 가이드 시스템 둘 다의 예는 V형 CRISPR 좌위에 존재한다. 실험적 메타-전사체 데이터를 사용하여 CasX가 DNA 표적화 활성에 대해 tracrRNA를 필요로 하는지의 여부를 결정하였다. 이 분석은 Cas2 오픈리딩 프레임과 CRISPR 어레이 사이에 암호화된 CRISPR 반복부에 상보적인 서열을 갖는 비-암호 RNA 전사체를 나타내었다(도 10a). 전사체 맵핑은 추가로 CRISPR RNA(crRNA)가 CRISPR-Cas9 시스템에서 생긴 crRNA 가공과 유사하게 반복부의 22 nt 및 인접한 스페이서의 20nt를 포함하도록 가공된다는 것을 시사한다(도 10a). 더 나아가, 2-nt 3' 돌출부를 동정하였는데, 이는 crRNA-tracrRNA 듀플렉스의 RNase III-매개 가공과 일치된다(도 10b). 추정적 tracrRNA에 대한 CasX 활성의 의존도를 결정하기 위해, 이 영역을 상기 기재한 최소 CRISPR-CasX 좌위로부터 결실시키고, 플라스미드 간섭 분석을 반복하였다. CasX 플라스미드로부터의 추정적 tracrRNA-암호화 서열의 결실은 그의 존재에서 관찰된 강한 형질전환 간섭이 없었다(도 10c). 종합하면, 이들 결과는 새로운 기능성 DNA-표적화, 이중-RNA 가이드 CRISPR 효소로서 CasX를 확립한다.

[0546] **단리물이 없는 박테리아 계통에서 배타적으로 발견된 시스템인 CRISPR-CasY**

[0547] 특정 후보 문 방사선(candidate phyla radiation: CPR) 박테리아의 게놈에서 암호화된 다른 새로운 클래스 2 Cas 단백질을 동정하였다. 이들 박테리아는 전형적으로 작은 세포 크기(크라이오-TEM 데이터 및 여과를 통한 농축에 기반함), 매우 작은 게놈 및 제한된 생합성 능력을 갖는데, 이는 그들이 공생자일 가능성이 있다는 것을

나타낸다. 본 명세서에서 CasY로서 지칭되는 새로운 대략 1,200 aa Cas 단백질은 기껏해야, Cas1 및 CRISPR 어레이를 포함하는 최소 CRISPR-Cas 시스템의 부분인 것으로 나타났다(도 11a). 대부분의 CRISPR 어레이는 보통 17 내지 19 nt의 짧은 스페이서를 갖지만, Cas1을 결여하는 하나의 시스템(CasY.5)은 더 긴 스페이서(27 내지 29 nt)를 가진다. 동정된 CasY 단백질의 6가지 예는 공공의 데이터베이스에서 임의의 단백질에 대해 상당한 서열 유사성이 없었다. 공개된 Cas 단백질^{3,4}로부터 구성된 프로파일 모델(HMM)을 이용하는 민감한 검색은 6개의 CasY 단백질 중 4개는 RuvC 도메인과 중복되는 C-말단 영역 및 N-말단의 작은 영역(대략 45aa)에서 C2c3에 대해 국소 유사성(e -값 4×10^{-11} 내지 3×10^{-18})을 가진다는 것을 나타내었다(도 18 참조). C2c3은 분류학상의 소속 없이 짧은 콘택에 대해 동정된 추정적 V형 Cas 효과기이며, 실험적으로 입증되지 않았다. CasY와 같이, C2c3는 짧은 스페이서 및 Cas1를 갖지만, 다른 Cas 단백질이 없는 어레이 옆에서 발견되었다. 특히, 현재의 연구에서 동정된 CasY 단백질 중 둘은 다른 CasY 단백질과 상당한 서열 유사성(최고 Blast 히트: e -값 6×10^{-85} , 7×10^{-75})을 공유함에도 불구하고, C2c3에 대해 상당한 유사성이 없었다.

[0548] 임의의 실험적으로 확인된 CRISPR 좌위에 대한 CRISPR-CasY의 낮은 상동성을 고려하여, 다음에 이 시스템이 RNA-가이드 DNA 간섭을 부여하는지의 여부를 생각하였지만, 짧은 스페이서 길이 때문에 이러한 활성화에 필요할 수 있는 가능한 PAM 모티프에 관해 신뢰 가능한 정보는 존재하지 않았다. 이것에 착수하기 위해, 단축된 CRISPR 어레이로 전체 CRISPR-CasY.1 좌위를 합성하고 나서, 플라스미드 벡터 상에서 이콜라이에 도입하였다. 이어서, 가능한 PAM을 동정하기 위해 어레이에서 스페이서 서열에 매칭되고 인접한 무작위화 5' 또는 3'영역을 함유하는 서열을 갖는 플라스미드를 이용하여 형질전환 분석에서 이들 세포에 도전하였다. 형질전환체의 분석은 표적화된 서열에 직접적으로 인접한 5' TA를 함유하는 서열의 고갈을 나타내었다(도 11b). 이 동정된 PAM 서열을 이용하여, CasY.1 좌위를 단일 PAM을 함유하는 플라스미드에 대해 시험하였다. 플라스미드 간섭은 동정된 5' TA PAM 서열을 함유하는 표적의 존재에서만 입증되었다(도 11c). 따라서, 이들 데이터는 CRISPR-CasY가 DNA 간섭 활성을 가진다는 것을 나타낸다.

[0549] 논의

[0550] 비배양 박테리아 및 고세균으로부터의 게놈에서 새로운 클래스 2 CRISPR-Cas 적응 면역계를 동정하였고 특성구명하였다. 활성화 CRISPR 좌위에 대해 보편적인 Cas1의 진화 분석(도 12a)은 본 명세서에 기재된 고세균 Cas9 시스템이 임의의 존재하는 II형 서브타입에 명확하게 속하지 않았다는 것을 시사하였다. Cas1 계통발생(뿐만 아니라 *cas4*의 존재)을 야생형 II-B 시스템과 함께 클러스터링시켰고, 아직 Cas9의 서열은 II-C형 단백질과 더 유사하였다(도 20). 따라서, 고세균 II형 시스템은 II-C형 및 II-B형 시스템의 융합으로서 생길 수 있다(도 12b). 마찬가지로, Cas1 계통발생 분석은 CRISPR-CasX 시스템으로부터의 Cas1인 임의의 다른 공지된 V형 시스템으로부터 원위에 있다는 것을 나타내었다. V형 시스템은 조상 I형 시스템으로부터의 적응 모듈(Cas1 내지 Cas2)과 트랜스포존의 융합 결과가 되는 것으로 시사되었다. 따라서 CRISPR-CasX 시스템이 이전에 기재된 V형 시스템에 생기는 것과 상이한 융합 사건 후에 나타났다는 가설을 세운다. 현재하게, CRISPR-CasY와 추정적 C2c3 시스템은 둘 다 CRISPR 좌위 내로 DNA를 통합하는 데 필수적인 것으로 생각되는 단백질인 Cas2를 결여하는 것으로 보인다. 모든 CRISPR-Cas 시스템이 Cas1과 Cas2를 둘 다 함유한 조상 I형 시스템의 후속인 것으로 생각되는 것을 고려하면, CRISPR-CasY 및 C2c3 시스템은 CRISPR-Cas 시스템의 나머지화 상이한 혈통을 가질 수 있거나, 또는 대안적으로, Cas2는 그들의 진화 이력 동안 상실될 수 있었다.

[0551] 고세균에서 Cas9의 본 명세서에 기재된 발견 및 박테리아에서 2가지의 이전에 미공지 CRISPR-Cas 시스템은 복잡한 천연 미생물 군집으로부터 얻은 광대한 DNA 및 RNA 서열을 사용하였다. CasX 및 CasY의 경우에, 게놈 내용은 결집되지 않은 서열 정보로부터 명백하지 않은 기능의 예측에 대해 중요하였다. 추가로, 추정적 tracrRNA뿐만 아니라 표적화된 바이러스 서열의 동정은 메타게놈 데이터 가이드된 기능성 시험의 분석을 통해 다루지 않았다. 흥미롭게도, 지금까지 동정된 가장 뻘뻘한 CRISPR-Cas 좌위의 일부는 매우 작은 게놈을 갖는 유기체에서 발견되었다. 작은 게놈 크기의 결과는 기본적 대사 필요를 위해 다른 군집 구성원에 의존할 가능성이 있으며, 따라서 그들은 대체로 전통적 배양-기반 방법의 범주 밖에서 남아있었다. 간섭에 필요한 제한된 수의 단백질은 이들 최소 시스템이 새로운 게놈 편집 도구의 개발에 특히 가치 있게 만든다. 중요하게는, CRISPR-Cas 시스템과 관련된 메타게놈 발견이 인실리코 관찰로 제한되지 않지만, 그들의 기능이 시험될 수 있는 실험 환경에 도입될 수 있다는 것을 본 명세서에 나타낸다. 생명이 존재하는 사실상 모든 환경이 현재 게놈-해결 메타게놈 방법에 의해 프로빙될 수 있다는 것을 고려하면, 본 명세서에 기재된 조합된 컴퓨터-실험 접근은 공지된 CRISPR-Cas 시스템의 다양성을 크게 확장시켜, 생물학적 연구 및 임상 용도를 위한 새로운 기법을 제공한다는 것이 예상된다.

[0552]

방법

[0553]

메타게놈 및 메타전사체분석

[0554]

3개의 상이한 부위로부터의 메타게놈 샘플을 분석하였다: (1) 캘리포니아주 아이언 마운틴에 소재한 리치몬드 광산으로부터 2006년과 2010년 사이에 수집한 산성 광산 배수(AMD) 샘플 (2) 콜로라도주 라이플 근처의 콜로라도강에 인접한 라이플 인터그레이티드 필드 리서치(Rifle Integrated Field Research: IFRC) 현장으로부터의 2007년과 2013년 사이에 수집한 지하수 및 퇴적물 샘플. (3) 유타주 콜로라도 고원에서 크리스탈 가이저(Crystal Geyser)로부터 2009년과 2014년에 수집한 지하수, 차가운, CO₂-유도 간헐천.

[0555]

AMD 데이터에 대해, 문헌[Denef and Banfield (2012) 및 Miller et al. (2011)]에 의해 DNA 추출 방법 및 짧은 판독 서열분석을 보고하였다. 라이플 데이터에 대해, DNA 및 RNA 추출뿐만 아니라 서열분석, 조립체 및 재구성된 게놈을 문헌[Anantharaman et al. (2016) 및 Brown et al. (2015)]에 의해 기재하였다. 크리스탈 가이저로부터의 샘플에 대해, 방법은 문헌[Probst et al (2016) 및 Emerson et al. (2015)]에 의해 기재된 것에 따른다. 간략하게, 파워소일(PowerSoil) DNA 단리 키트(미국 캘리포니아주 칼스베드에 소재한 모바이오 래버러토리즈 인코포레이티드(MoBio Laboratories Inc.))를 이용하여 DNA를 샘플로부터 추출하였다. 문헌[Brown et al. (2015)]에 의해 기재된 바와 같은 6개의 2011 라이플 지하수로부터 수집된 0.2μm 필터로부터 RNA를 추출하였다. DNA를 일루미나(Illumina) HiSeq2000 플랫폼 상에서, 그리고 메타전사체 cDNA를 5500XL SOLiD 플랫폼 상에서 서열분석하였다. 새로 보고된 크리스탈 가이저 데이터 및 AMD 데이터의 재분석을 위해, IDBA-UD를 이용하여 서열을 조립하였다. Bowtie2를 이용하여 서열분석 범위 및 유전자 발현을 각각 결정하기 위해 사용한 DNA 및 RNA(cDNA) 판독-맵핑을 수행하였다. 프로디갈(Prodigal)을 이용하여 조립 스캐폴드에 대해 오픈 리딩 프레임(Open reading frame: ORF)을 예측하였다. 출현 자기-조직화 맵(Emergent Self-Organizing Map: ESOM)을 이용하여 ABAWACA, ABAWACA2(<https://github.com/CK7>) Maxbin2, 및 테트라뉴클레오타이드 빈도의 조합을 이용하는 차별적 범위 존재비 패턴에 기반하여 크리스탈 가이저 데이터셋으로부터의 스캐폴드를 버렸다. GC 함량%, 분류학상의 소속 및 게놈 완전성을 이용하여 게놈을 수동으로 만들었다. ra2.py를 이용하여 스캐폴딩 오차를 수정하였다(<https://github.com/christophertbrown>).

[0556]

CRISPR-Cas 계산 분석

[0557]

문헌[Makarova et al. 및 Shmakov et al.]으로부터의 정렬에 기반하여 허머(HMMer) 스위트를 이용하여 구성된 은닉 마르코프 모델(HMM) 프로파일을 이용하여 공지된 Cas 단백질에 대해 다양한 샘플로부터의 조립 콘택을 스캐닝하였다. CrisprFinder 소프트웨어의 국소 형태를 이용하여 CRISPR 어레이를 동정하였다. *cas1* 유전자에 인접한 10개의 ORF 중 하나가 800 aa보다 더 큰 비특정규명 단백질에 대해 암호화되었다면 Cas1과 CRISPR 어레이를 둘 다 함유한 좌위를 추가로 분석하였고, 동일한 콘택 상에서 *cas* 간접 유전자는 동정되지 않았다. 이들 거대 단백질을 잠재적 클래스 2 Cas 효과기로서 추가로 분석하였다. MCL을 이용하여 서열 유사성에 기반하여 단백질 패밀리에 대해 잠재적 효과기를 클러스터링하였다. 각각의 이들 패밀리를 나타내는 HMM을 구성함으로써, 그리고 유사한 Cas 단백질에 대해 메타게놈 데이터셋을 검색하기 위해 그들을 이용함으로써 이들 단백질 패밀리를 확장시켰다. 단백질 패밀리가 정말로 새롭게 되도록 하기 위해, NCBI의 혼치않은(nr) 및 메타게놈(env_nr) 단백질 데이터베이스에 대한 BLAST뿐만 아니라 UniProt KnowledgeBase에 대한 HMM 검색을 이용하여 공지된 상동체를 검색하였다. 전장 히트가(단백질 길이의 25% 초과) 없는 단백질만을 신규한 단백질로 고려하였다. HH-스위트로부터의 HHpred를 이용하여 추정적 Cas 단백질의 원위 상동성 검색을 수행하였다. 해결된 결정 구조, 및 JPred4에 의해 예측된 2차 구조와의 비교에 기반하여 도메인 구조를 추론하기 위해 고스코어링 HHpred 히트를 사용하였다. 새로 발견된 Cas 단백질을 포함하는 HMM 데이터베이스를 보조 데이터 1에서 이용 가능하다.

[0558]

CrisprFinder를 이용하여 조립 데이터로부터 스페이스 서열을 결정하였다. 적절한 샘플의 짧은 DNA 판독에서 추가적인 스페이스를 위치시키기 위해 CRASS를 사용하였다. 스페이스에 대해 미스매치가 1 이하인 히트에 대한 적절한 메타게놈 조립체에 대해 BLAST 검색에 의해("-task blastn-short"를 이용) 스페이스 표적(프로토스페이스)을 동정하였다. (프로토스페이스로서 CRISPR 어레이 동정을 피하기 위해) 관련된 반복부를 함유한 콘택에 속하는 히트를 여과시켰다. 프로토스페이스에 축적하는 영역을 정렬함으로써 그리고 웹로고(WebLogo)를 이용하여 시각화함으로써 프로토스페이스 인접 모티프(PAM)를 동정하였다. mFold를 이용하여 RNA 구조를 예측하였다. 조립 데이터로부터 스페이스, 반복부 및 축적 서열을 수동으로 정렬함으로써 CRISPR 어레이 다양성을 분석하였다. 제네이오스(Geneious) 9.1을 이용하여 수동 정렬 및 콘택 시각화를 수행하였다.

[0559]

새로 동정된 시스템의 Cas1 및 Cas9 단백질의 계통발생학적 분석을 위해 문헌[Makarova et al. 및 Shmakov et

al.]으로부터의 단백질과 함께 사용하였다. CD-HIT를 이용하여 90% 이상의 동일성을 갖는 단백질을 함께 클러스터링함으로써 비중복 세트에 따랐다. MAFFT를 이용하여 정렬을 생성하였고, 그리고 치환 모델로서 PROTGAMMALG를 이용하는 RAXML 및 100 부트스트랩(bootstrap) 샘플링을 이용하여 최대-가능성 계통발생을 작성하였다. 카스포존(casposon)을 야기하는 분지를 이용하여 Cas1 나무의 뿌리를 고정시켰다. FigTree 1.4.1 (<http://tree.bio.ed.ac.uk/software/figtree/>) 및 iTOL v3을 이용하여 나무를 시각화하였다.

[0560] **이중성 플라스미드의 생성**

[0561] CasX에 대한 획득과 관련된 단백질을 제거함으로써 그리고 CasX와 CasY 둘 다에 대해 CRISPR 어레이 크기를 감소시킴으로써 메타게놈 콘텍을 최소 CRISPR 간섭 플라스미드로 만들었다. 최소 좌위를 Gblocks(인테그레이티드 DNA 테크놀로지)로서 합성하고 나서, 깃슨 어셈블리(Gibson Assembly)를 이용하여 조립하였다.

[0562] **PAM 고갈 분석**

[0563] 변형에 의해 앞서 기재한 바와 같이 PAM 고갈 분석을 수행하였다. 무작위화된 PAM 서열을 함유하는 플라스미드 라이브러리를 프라이머를 이용하여 7 nt 무작위화된 PAM 영역을 함유하는 DNA 올리고뉴클레오타이드를 어닐링함으로써 조립하고 나서, 클레노브단편(Klenow Fragment: NEB)으로 연장시켰다. EcoRI 및 NcoI로 이중 가닥 DNA를 분해시키고 나서, pUC19 골격에 결합시켰다. 결합된 라이브러리를 DH5α로 형질전환시키고, 10⁸개 초과 세포를 채취하고 나서, 플라스미드를 추출하고, 정제하였다. 200ng의 풀링 라이브러리를 CRISPR 좌위 또는 좌위가 없는 대조군 플라스미드를 보유하는 전기천공적격 이콜라이로 형질전환시켰다. 형질전환 세포를 30시간 동안 25℃에서 카베니실린(100mg ℓ⁻¹) 및 클로람페니콜(30mg ℓ⁻¹)을 함유하는 선택 배지 상에서 플레이팅하였다. 플라스미드 DNA를 추출하고 나서, 일루미나 서열분석을 위한 어댑터를 이용하여 PAM 서열을 증폭시켰다. 7 nt PAM 영역을 추출하고 나서, 각각의 7nt 서열에 대해 PAM 빈도를 계산하였다. 명시된 역치 초과에서 고갈된 PAM 서열을 사용하여 웹로그를 생성하였다.

[0564] **플라스미드 간섭**

[0565] 메타게놈 서열 분석 또는 PAM 고갈 분석으로부터 동정된 추정적 표적을 pUC19 플라스미드 내로 클로닝시켰다. 10 ng의 표적 플라스미드를 CRISPR 좌위 플라스미드를 함유하는 전기천공적격 이콜라이(NEB 안정성)로 형질전환시켰다. 세포를 2시간 동안 25℃에서 회수하고 나서, 선택 배지 상에서 적절한 희석물을 플레이팅하였다. 플레이트를 25℃에서 인큐베이션하고 나서, 콜로니 형성 단위를 계수화하였다. 모든 플라스미드 간섭 실험을 3회 수행하고 나서, 전기천공적격 세포를 각각의 복제물에 대해 독립적으로 제조하였다.

[0566] **ARMAN-Cas9 단백질 발현 및 정제**

[0567] ARMAN-1(AR1) 및 ARMAN-4(AR4)로부터의 Cas9에 대한 발현 작제물을 이콜라이에 대해 코돈 최적화된 gBlocks(인테그레이티드 DNA 테크놀로지)로부터 조립하였다. N-말단 His₆-MBP 또는 His₆ 융합 단백질로서 pET-기반 발현 벡터에 조립 유전자를 클로닝하였다. 발현 벡터를 BL21(DE3) 이콜라이 세포로 형질전환시키고 나서, 37℃에서 LB 브로스에서 성장시켰다. 단백질 발현을 위해, 0.4mM IPTG(아이소프로필 β-D-1-티오갈락토피라노사이드)를 이용하여 중간-대수기 동안 세포를 유도하고 나서, 16℃에서 밤새 인큐베이션시켰다. 모든 후속적 단계를 4℃에서 수행하였다. 세포 펠릿을 용해 완충제(50mM 트리스(Tris)-HCl pH 8, 500mM NaCl, 1mM TCEP, 10mM 이미다졸) 0.5% 트리톤 X-100 중에서 재현탁시키고, 초음파처리에 의한 용해 전에 적격 프로테아제 저해제 혼합물(로슈(Roche))로 보충하였다. 용해물을 15000g에서 40분 동안 원심분리에 의해 정화시키고 나서, 배취에서 슈퍼플로우(Superflow) Ni-NTA 아가로스(퀴아젠(Qiagen))에 적용하였다. 세척 완충제 A(50mM 트리스-HCl pH 8, 500mM NaCl, 1mM TCEP, 10mM 이미다졸)를 이용하여 수지를 광범위하게 세척한 후에, 5개 칼럼 용적의 세척 완충제 B(50mM 트리스-HCl pH 8, 1M NaCl, 1mM TCEP, 10mM 이미다졸)로 세척하였다. 용해 완충제(50mM 트리스-HCl pH 8, 500mM NaCl, 1mM TCEP, 300mM 이미다졸)를 이용하여 Ni-NTA 수지의 단백질을 용리시켰다. 세척 완충제 A에 대해 밤새 투석 동안 TEV 프로테아제에 의해 His₆-MBP 태그를 제거하였다. 제2 Ni-NTA 아가로스 칼럼을 통해 친화도 태그로부터 절단된 Cas9를 제거하였다. 5mL 헤파린 HiTrap 칼럼(GE 라이프 사이언시스(GE Life Sciences))에 대한 적용 전에 IEX 완충제 A(50mM 트리스-HCl pH 7.5, 300mM NaCl, 1mM TCEP, 5% 글리세롤)에 단백질을 투석하였다. 선형 NaCl (0.3 내지 1.5M) 구배로 Cas9를 용리시켰다. 분획을 풀링하고 나서, 30kDa 스핀 농축기(써모 피셔(Thermo Fisher))를 이용하여 농축시켰다. 적용 가능할 때, 슈퍼덱스(Superdex) 200 pg 칼럼(GE 라이프 사이언시스) 상에서 크기-배제 크로마토그래피를 통해 Cas9를 추가로 정제하였고, 후속적 절단 분석을 위해 IEX 완충제 A에서 저장하였다. 효모 발현을 위해, AR1-Cas9를 Gal1/10 His6-MBP TEV Ura 사카로마이세스 세

레비시애(*S. cerevisiae*) 발현 벡터(애드진 플라스미드(Addgene plasmid) # 48305)에 클로닝시켰다. 벡터를 BY4741 URA3 균주에 형질전환시키고 나서, 배양물을 30℃로 배지에서 성장시켰다. 대략 0.6의 OD600에서, 2% w/v 갈락토스를 이용하여 단백질 발현을 유도하고 나서, 16℃에서 밤새 인큐베이션시켰다. 상기와 같이 단백질 정제를 수행하였다.

[0568] RNA 시험관내 전사 및 올리고뉴클레오타이드 정제

[0569] T7 프로모터 서열을 함유하는 합성 DNA 주형을 이용하여 앞서 기재한 바와 같이 시험관내 전사 반응을 수행하였다.⁶⁵ 변성 PAGE를 통해 모든 시험관내 전사된 가이드 RNA 및 표적 RNA 또는 DNA를 정제하였다. 95℃에서 1분 동안 인큐베이션에 의해 20mM 트리스 HCl pH 7.5 및 100mM NaCl 중에서 이중-가닥 표적 RNA 및 DNA를 혼성화시킨 후에, 실온으로 서서히 냉각시켰다. 천연 PAGE에 의해 혼성체를 정제하였다.

[0570] 시험관내 절단 분석

[0571] 1x PNK 완충제 중에서 37℃에서 30분 동안 T4 폴리뉴클레오타이드 키나제(NEB) 및 [γ -32]ATP(퍼킨-엘머)를 이용하여 정제된 DNA 및 RNA 올리고뉴클레오타이드를 방사성표지하였다. 65℃에서 20분 동안 PNK를 열 불활성화시키고 나서, 일러스트라 마이크로스핀(illustra Microspin) G-25 칼럼(GE 라이프 사이언시즈)을 이용하여 표지 반응으로부터 유리 ATP를 제거하였다. CrRNA와 tracrRNA를 1x 재폴딩 완충제(50mM 트리스 HCl pH 7.5, 300mM NaCl, 1mM TCEP, 5% 글리세롤) 중에서 등몰량으로 혼합하고 나서, 70℃에서 5분 동안 인큐베이션시키고, 이어서, 실온으로 서서히 냉각시켰다. 반응물을 1mM 최종 금속 농도로 보충하고, 후속적으로 50℃에서 5분 동안 가열하였다. 실온으로 서서히 냉각시킨 후에, 재폴딩된 가이드를 얼음에 위치시켰다. 완충제, 염 농도에 대해 언급되지 않는 한, Cas9를 37℃에서 10분 동안 1x 절단 완충제(50mM 트리스 HCl pH 7.5, 300mM NaCl, 1mM TCEP, 5% 글리세롤, 5mM 2가 금속) 중에서 등몰량의 가이드로 재구성하였다. 37℃ 또는 표시된 온도에서 방사성 표지 표적에 대해 10x 과량의 Cas9-가이드 복합체를 갖는 1x 절단 완충제에서 절단 반응을 수행하였다. 50mM EDTA로 보충한 동일 용적의 겔 부하 완충제에서 반응을 쿨링시켰다. 10% 변성 PAGE 상에서 절단 생성물을 분해하고 나서, 포스포리미징에 의해 시각화하였다.

[0572] 생체내 이콜라이 간섭 분석

[0573] AR1- 및 AR4-Cas9에 대한 이콜라이 형질전환 분석을 앞서 공개된 바와 같이 수행하였다⁶⁶. 간략하게, 가이드 RNA에 의해 형질전환된 이콜라이는 전기천공적격으로 만들어졌다. 이어서, 세포를 야생형 또는 촉매적으로 비활성인 Cas9(dCas9)를 암호화하는 플라스미드의 9fmol로 형질전환시켰다. 회수한 세포의 연속 희석물을 선택적 항생제와 함께 LB 플레이트 상에 플레이팅하였다. 37℃에서 16시간 후에 콜로니를 계수하였다.

표 1

[0574]

CRISPR-Cas 시스템을 동정한 유기체 및 게놈 위치에 관한 상세한 설명뿐만 아니라 재구성된 스페이서의 수 및 평균 길이 및 반복부 길이에 대한 정보(NA, 이용 가능하지 않음). ARMAN-1 스페이서를 16개의 샘플로부터 재구성하였다.						
분류학적 그룹	Cas 효과기	NCBI 접근	좌표	반복부 길이	스페이서 수	스페이서 평균 길이
ARMAN-1	Cas9	MOEG01000017	1827..7130	36	271	34.5
ARMAN-4	Cas9	KY040241	11779..14900	36	1	36
델타프로테오박테리아	CasX	MGPG01000094	4319..9866	37	5	33.6
부유균문	CasX	MHYZ01000150	1..5586	37	7	32.3
칸디다투스 칸타노박테리아(Candidatus Katanobacteria)	CasY.1	MOEH01000029	459..5716	26	14	17.1
칸디다투스 보겔박테리아(Candidatus Vogelbacteria)	CasY.2	MOEJ01000028	7322..13087	26	18	17.3
칸디다투스 보겔박테리아	CasY.3	MOEK01000006	1..4657	26	12	17.3
칸디다투스 파르쿠박테리아(Candidatus Parcubacteria)	CasY.4	KY040242	1..5193	25	13	18.4
칸디다투스 코메일리박테리아(Candidatus Komeilibacteria)	CasY.5	MOEI01000022	2802..7242	36	8	26

칸디다투스 케르펠드박테리아(Candidatus Kerfeldbacteria)	CasY.6	MHKD01000036	11503..1536 6	NA	NA	NA
--	--------	--------------	------------------	----	----	----

[0575]

본 발명을 이의 구체적 실시형태를 참고로 하여 기재하였지만, 본 발명의 진정한 정신과 범주로부터 벗어나는 일 없이 다양한 변화가 이루어질 수 있고 동등물이 치환될 수 있다는 것이 이해되어야 한다. 추가로, 본 발명의 객관적 정신 및 범주에 대해 특정 상황, 물질, 대상 조성물, 공정, 공정 단계 또는 단계들을 적용하여 다수의 변형이 이루어질 수 있다. 모든 이러한 변형은 본 명세서에 첨부하는 청구범위의 범주 내인 것으로 의도된다.

도면

도면1a

```
>CasY1
MRKKLFKGYILHNKRLVYTGKAAIRSIKYPLVAPNKTALNNLSEKIIYDYEHLFGPLNVA
SYARNSNRYSLVDFWIDSLRAGVIWQSKSTSLIDLISKLEGSKSPSEKIFEQIDFELKNK
LDKEQFKDIILLNTGIRSSSNVRSLRGRFLKCFKEEFRDTEEVACVDKWSKDLIVEGKS
ILVSKQFLYWEEFFGIKIFPHFKDNHDLPKLTFFVEPSLEFSPHLPLANCLERLKKFDIS
RESLLGLDNNFSAFSNYFNELFNLLSRGEIKKIVTAVLAVSKSWENEPELEKRLHFLSEK
AKLLGYPKLTSSWADYRMIIGGKIKSWHSNYTEQLIKVREDLKKHQIALDKLQEDLKKVV
DSSLREQIEAQREALLPLDMLKEKDFSDDELYRFLSDFKSLNGSYQRYIQTEER
KEDRDVTKKYKDLYSNLRNIPRFFGESKKEQFNKFINKSLPTIDVGLKILEDIRNALETV
SVRKPPSITEEYVTKQLEKLSRKYKINAFNSNRFKQITEQVLRKYNNGELPKISEVFYRY
PRESHVAIRILPVKISNPRKDISYLLDKYQISPDWKNNSNPGEVVDLIEIYKLTGLWLLSC
NKDFSMDFFSSYDLKLFPEAASLIKNGFSGCLSGYYLSKMIFNCITSEIKGMITLYTRDKFV
VRYVTQMIGSNQKFPLLCLVGEKQTKNFSRNWGVLEEKGDLGEEKNQEKCLIFKDKTDF
AKAKEVEIFKNNIWRIRTSKYQIQFLNRLFKKTKEWDLNMLVLEPSLVLEEEWGVSWDK
DKLLPLLKKEKSCEERLYSLPLNLVPATDYKEQSAEIEQRNTYLGIDVGEFGVAYAVVR
IVRDRIELLSWGLFKDPALRKIRERVQDMKKKQVMAVFSSSSTAVARVREMAIHSLRNQI
HSIALAYKAKIYISISINPFTGGNRMAKIYRSIKVSDVYRESGADTLVSEMIWGKKNKQ
MGNHISSYATSYTCCNCARTPFELVIDNDKEYEKGDEFIFNVGDEKKVRGFLQKSLLGK
TIKKGKVLKSIKEYARPPIREVLLLEGEDVEQLLKRRGNSYIYRCPCFGYKTDADIQAALN
IACRGYISDNAKDAVKEGERKLDYILEVRKLWEKNGAVLRSKFL (서열번호 1)
```

```
>CasY2
MQKVRKTLSEVHKPNPYGTVKVRNAKTGYSLOIERLSYTGKEGMRSEFKIPIENKNKEVFDEF
VKKIRNDYISQVGLNLSDWYEHYQEKQEHYSLADFWLDSLRAGVIFAHKETEIKNLISK
IRGDKSIVDKFNASIKKKHADLYALVDIKALYDFTSDARRGLKTEEEFFNSKRNTLFPK
FRKKDNKAVDLWVKKFIFGLDNKDKLNFTKKFIFGDPNPQIKYDHTFFFHQDINFDLERIT
TPKELISTYKKFLGKNKDLGSDETTEDQLKMVLGFHNNHGAFSKYFNASLEAFRGRDNS
LVEQIINNSPYWNNSHRKELEKRIIFLQVQSKKIKETELGKPHEYLASFSGGKFESWVSNYL
RQEEEVKRQLFGYEEENKKGQKKFIVGNKQELDKIIRGTDEYEIKAISKETIGLTQKCLKL
LEQLKDSVDDYTLISLYRQLLIVELRIRLNVEFQETYPELIGKSEKDKEKDAKNKRADKRYP
QIFKDIKLIPNFLGETKQMVYKKFIRSADILYEGINFIDQIDKQITQNLPCFKNDKERI
EFTEKQFETLRKYILMNSSRFHHVIEGIINNRKLIEMKKRENSSELKTFSDSKFVLSKLF
LKKGGKYENEVYTFYINPKARDQRRKIVLDINGNNSVGILQDLVQKLKPKWDDIICKN
DMGELIDAIEIEKVRLGILIALYCEHKFKIKKELLSLDLFASAYQYLELEDDPEELSGTN
LGRFLQSLVCSEIKGAINKISRTEYIERITVQPMNTEKNYPLLINKEGKATWHIAAKDDL
SKKKGGGTVAMNQKIGKNFFGQDYKTVFMLQDKRFDLLTSKYHLQFLSKTLDTGGGSWW
KNKNIDLNLSSYSFIFEQKVKVEWDLTNLDHPKIKIPSENSDDRRLFVSIPTVIKPKQTK
RKDLQTRVNYMGIDIGEYGLAWTIINIDLKNKKINKISKQGFIEPLTHKVRDYVATIKD
NQVRGTFGMPDTPKLARLRENAITSLRNQVHDIAMRYDAKPVYEFELSNFETGSKNVKVIY
DSVKRADIGRGQNTEADNTEVNLVWGKTSKQFGSQIGAYATSYICSFCGYSPYYEFENS
KSGDEEGARDNLYQMKKLSRPSLEDFLQGNFVYKTRFRDFDKYKNDQRLQKTGDKDGWKT
HRGNTAIYACQKCRHISDADIQASYWIALKQVVRDFYKDKEMDGDLIQGDNDKDRKVNEL
NRLIGVHKDVPIINKNLITSLDINLL (서열번호 2)
```


도면1b

>CasY3

MKAKKSFYNQKRKFGKRGYRLHDERIAYS GGIGSMRSIKYELKDSYGIAGLRNR
IADATISDNKWLYGNINLNDYLEWRSSKTDKQIEDGDRESSLLGFWLEALRLGFVFSKQS
HAPNDFNETALQDLFETLDDDLKHVLD RKKWCDFIKIGTPKTNDQGRLKKQIKNLLKGNK
REEIEKTLNESDDELKEKINRIADVFAKNKSDKYTIFKLDKPNTKEYPRINDVQVAFFCH
PDFEEITERDRTKTLDLIINRFNKRYEITENKKDDKTSNRMALYSLNQGYIPRVLNDLFL
FVKDNEDDFSQFLSDLENFFSFSNEQIKI IKERLKKLKKYAEPIPGKPLADKWDDYASD
FGGKLESWYSNRIEKLKKIPESVSDLRNNLEKIRNVLKKQNNASKILELSQKIIEYIRDY
GVSEFEKPEIIFSWINKTKDGQKKVYFVAKMADREFIEKLDLWMADLRSQLEYNQDNKV
SFKKKGKKIEELGVLDFAFNKAKKNKSTKNENGWQQLSESIQSAPLFFGEGNRVRNEEV
YNLKDLLFSEIKNVENILMSSEAEDLKNIKIEYKEDGAKKGNVNLVNLARFYARFNEGDY
GGWNKVKTVLENIAREAGTDFSKYGNMNNRNAGRFYLNGRERQVFTLIKFEKSITVEKIL
ELVKLPSSLDEAYRDLVNENKNHKL RDVIQLSKTIMALVLSHSDKEKQIGGNYIHSKLSG
YNALISKRDFISRYSVQTTNGTQCKLAIGKGSKKGNEIDRYFYAFQFFKNDDSKINLKV
IKNNSHKNI DFNENKINALQVYSSNYQIQFLDWFFEKHQGGKTSLEVGGSTIAEKS
LTI DWSGSNPRVGFKRSDTEEKRVFVSQPF TLIPDDEDKERRKERMIKTKNRFIGIDIGEY
GLAWSLIEVDNGDKNNRGIRQLES GFITDNQQQVLKKNVKSQRQNIQIRQTFTSPDTKIAR
LRESLIGSYKNQLES LMAKKANLSFEYEVSGFEVGGKRVAKIYDSIKRGSVRKKDNNSQ
NDQSWGKKGIN EWSFETTAAGTSQFCTHCKRWSSLAIVDIEEYELKDYNDNLFKVKINDG
EVRLLGKKGWRSGEKIKGKELFGPVKDAMPNV DGLGMKIVKRKYLKLDLRDWVSRYGNM
AIFICPYVDCHHISHADKQAAFNIAV (서열번호 3)

>CasY4

MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYV
GLYGLSNFDDLYNAEK RNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYEL
TKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCENKL
ADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNR
NRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITEK
KAMMDITDAWRGQEQEEELEKRLRILAAALTIKLREP KFDNHWGGRSDINGKLSSWLQNY
INQTVKIKEDLKGHKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKP
DIPAIAIYRRFLSDGRLTLNRVQREDVQEALIKERLEAEKKKKPKKRKKKSDAEDEKET
IDFKELFPHLAKPLKLVNPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSSSLKN
SFFD TDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQS
RSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKT
ALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRLGLA
GLMSRKEFITRS AIQTMNGKQAELLYIPHEFQSAKITTPKEMSR AFLDLAPAEFATSLEP
ESLSEKSL LKLMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKT
LGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTV
ALEPVSGSERVFSVQPF TIFPEKSAEEEGQRYL GIDIGEYGIAYTALEITGDSAKILDQN
FISDPQLKTLREEVKGLKLDQRRGT FAMPSTKIAR IRESLVHSLRNRIHHLALKHKAKIV
YELEVS RFEEGKQIKIKVYATLKKADVSEIDADKNLQTTVWGKLAVASEISASYTSQFC
GACKKLWRAEMQVDETITTTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDT PFPKYRDF
CDKHHISKMRGNSCLFICPFCRANADADIQASQTIALRYVKEEKKVEDYFERFRKLKN
IKVLGQMKKI (서열번호 4)

도면1c

>CasY5

MAESKQMQRCKGASMKYEVI GLGKKSCRYMCPDCGNHTSARKIQNKKKRD
 KKYGSASKAQSORIAVAGALYPDKKVQTIKTYKYPADLNGEVHDSGVAEKIAQAIQEDEI
 GLLGPSSEYACWIASQKQSEPYSVVDFWFDVAVCAGGVFAYSGARLLSTVLQLSGEESVLR
 AALASSPFVDDINLAQAEKFLAVSRRTGQDKLGKRIGECFAEGRLEALGIKDRMREFVQA
 IDVAQTAGQRFAAKLKIFGISQMPEAKQWNND SGLTVCILPDYYVPEENRADQLVVLLRR
 LREIAYCMGIEDEAGFEHLGIDPGALSNFSGNGNPKRGFLGRLLNNDIIALANNMSAMTPY
 WEGRKGE LIERLAWLKHRAEGLYLKEPHFGNSWADHRSRIFSRIAGWLSGCAGKLIKIAKD
 QISGVRTDLFLLKRLLDVAVPQSAPSPDFIASISALDRFLEAAESSQDPAEQVRALYAFHL
 NAPA VRSIANKAVQRSDSQEWLIKELDAVDHLEFNKAFPPFFSDTGKKKKKGANSNGAPSE
 E EYTETESIQQPEDAEQEVNGQEGNGASKNQKKFQRI PRFFGEGSRSEYRILTEAPQYFD
 MFCNNMRAIFMQLESQPRKAPRDFKCF LQNR LQKLYKQTF L NARSNKCRALLESVLI SWG
 EFYTYGAN EKKFRLRHEASERSSDPDYVVQQALEIARRLFLFGFEWRDCSAGERVDLVEI
 HKKAISFLLAITQAEVSVGSYNWLGNS TVSR YLSVAGTDTLYGTQLEEF LNATVLSQMRG
 LAIRLSSQELKDGFDVQLESSCQDNLQHLLVYRASRDLAACKRATCPAELDPKILVLPVG
 AFIASVMKMIERGDEPLAGAYLRHRPHSFGWQIRVRGVAEVMGMDQGTALAFQKPTSEFF
 KIKPFS AQYGPVLWLNSSSYSSQSYLDGFLSQPKNWSMRVLPQAGSVRVEQRVALIWNLQ
 AGKMRLERSGARAFFMPVPF SFRPSGSGDEAVLAPNRYLGLFPHSGGIEYAVVDVLD SAG
 FKILERG TI AVNGFSQKRGERQEEAHREKQRRGISDIGRKKPVQAEVDAANELHRKYTDV
 ATRLGCRIVVQWAPQPKPGTAPT AQTVYARAVRTEAPRSGNQEDHARMKSSWGYTWGTYW
 EKRKPEDILGISTQVYWTGGIGESCPAVAVALLGHIRATSTQTEWEKEEVVFGRLKKFFP
 S (서열번호 6)

>CasY6

MAESKQMQRCKGASMKYEVI GLGKKSCRYMCPDCGNHTSARKIQNKKKRD
 KKYGSASKAQSORIAVAGALYPDKKVQTIKTYKYPADLNGEVHDSGVAEKIEQAIQEDEI
 GLLGPSSEYACWIASQKQSEPYSVVDFWFDVAVCAGGVFAYSGARLLSTVLQLSGEESVLR
 AALASSPFVDDINLAQAEKFLAVSRRTGQDKLGKRIGECFAEGRLEALGIKDRMREFVQA
 IDVAQTAGQRFAAKLKIFGISQMPEAKQWNND SGLTVCILPDYYVPEENRADQLVVLLRR
 LREIAYCMGIEDEAGFEHLGIDPGALSNFSGNGNPKRGFLGRLLNNDIIALANNMSAMTPY
 WEGRKGE LIERLAWLKHRAEGLYLKEPHFGNSWADHRSRIFSRIAGWLSGCAGKLIKIAKD
 QISGVRTDLFLLKRLLDVAVPQSAPSPDFIASISALDRFLEAAESSQDPAEQVRALYAFHL
 NAPA VRSIANKAVQRSDSQEWLIKELDAVDHLEFNKAFPPFFSDTGKKKKKGANSNGAPSE
 E EYTETESIQQPEDAEQEVNGQEGNGASKNQKKFQRI PRFFGEGSRSEYRILTEAPQYFD
 MFCNNMRAIFMQLESQPRKAPRDFKCF LQNR LQKLYKQTF L NARSNKCRALLESVLI SWG
 EFYTYGAN EKKFRLRHEASERSSDPDYVVQQALEIARRLFLFGFEWRDCSAGERVDLVEI
 HKKAISFLLAITQAEVSVGSYNWLGNS TVSR YLSVAGTDTLYGTQLEEF LNATVLSQMRG
 LAIRLSSQELKDGFDVQLESSCQDNLQHLLVYRASRDLAACKRATCPAELDPKILVLPAG
 AFIASVMKMIERGDEPLAGAYLRHRPHSFGWQIRVRGVAEVMGMDQGTALAFQKPTSEFF
 KIKPFS AQYGPVLWLNSSSYSSQSYLDGFLSQPKNWSMRVLPQAGSVRVEQRVALIWNLQ
 AGKMRLERSGARAFFMPVPF SFRPSGSGDEAVLAPNRYLGLFPHSGGIEYAVVDVLD SAG
 FKILERG TI AVNGFSQKRGERQEEAHREKQRRGISDIGRKKPVQAEVDAANELHRKYTDV
 ATRLGCRIVVQWAPQPKPGTAPT AQTVYARAVRTEAPRSGNQEDHARMKSSWGYTWSTYW
 EKRKPEDILGISTQVYWTGGIGESCPAVAVALLGHIRATSTQTEWEKEEVVFGRLKKFFP
 S (서열번호 7)

도면1d

>CasY7
 MKRIILNSLKVAAALRLLFRGKGSELVKTVMKYPLVSPVQGAVEELAEAIRHDNLHLFGQKEI
 VDLMEKDEGTQVYSVVDLWDLTLRLGMFFSPSANALKITLGKFNSDQVSPFRKVLEQSPF
 FLAGRLKVEPAERILSVEIRKIGKREN RVENYAADVETCFIGQLSSDEKQSIQKLANDIW
 DSKDHEEQRM LKADFFAIPLIKDPKAVTEEDPENETAGKQKPLELCVCLVPELYTRGFGS
 IADFLVQRLTLLRDKMSTDTAEDCLEYVGIEEEKGNMGMSLLGTFLKNLQGDGFEQIFQF
 MLGSYVWGQKEDVLRERLDLLAEKVRLPKPKFAGEWSGHRMFLHGLKSWSSNFFRLF
 NETRELLESIKSDIQHATMLISYVEEKGGYHPQLLSQYRKLMEQLPALRTKVLDPEIEMT
 HMSEAVRSYIMIHKS VAGFLPDLLSLDRDKDREFLLSIFPRIPKIDKKTKEIVAWELPG
 EP EEGYLFTANNLFRNFLENPKHVPRFMAERIPEDWTRLRSAPVWFDGMVKQWQKVVNQL
 VESPGALYQFNESFLRQLQAMLT VYKRD LQTEKFLKLLADVCRPLVDFFGLGGNDII FK
 SCQDPRKQWQTVIPLSVADVYTACEGLAIRLRETIGFEWKNLKGHEREDFLRLHQLLGN
 LLFWIRDAKLVVKLEDWMNNPCVQ EYVEARKAIDLPLEIFGFEVPIFLNGYLFSELQLE
 LLLRRKSVMTSYSVKTGSPNRLFQLVYLP LNPSPDEPKNSNNFQERLDTPTGLSRRFLD
 LTLD AFAAGLLTDPVTQELKTMAGFYDHLFGFKLPCKLAAMSNHPGSSSKMVVLAKPKKG
 VASNIGFEPIPDPAHPVFRVRSSWPELKYLEGLLYLPEDTPTIELAETS VSCQSVSSVA
 FDLKNLTITILGRVGEFRVTADQPFKLTPIIPEKEESFIGKTYLGLDAGERSGVGFAIVTV
 DGDGYEVQRLGVHEDTQLMALQCVASKSLKEPVFQPLRKGTFRQQERIRKSLRGCYWNFY
 HALMIKYRAKVVEESVGS SGLVGQWLRAFQKDLKKADVLPKKKGKNGVDKKKRESSAQD
 TLWGGAFSKKEEQIAFEVQAAGSSQFCLKCGWWFQLGMREVNRVQESGVVLDWNRSIVT
 FLIESSGEKVYGFSPQOLEKGFPRDIETFKKMVRDFMRPPMFDRKGRPAAAYERFVLGRR
 HRRYRFDKVFEERFGRSALFICPRVGCNGFDHSSEQS AVVLALIGYIADKEGMSGKKLVY
 VRLAELMAEWKLLKLLERSRVEEQSSAQ (서열번호 5)

>CasY18
 MKRIAKFRHDKPVKREAWSKGYRVHKNRIINKVTRSIKYPLVVKDEWKKRLIDDAHDYRWLVG
 PINYSWDCRDPNQYSILEFWIDFLCVGGVFQSSHSNICRLAIQLSGG SVFEQEWKDLSPFVRAN
 LIQGIKPAEFIGFLTAEFRSSSNPKNFISKFFEGSNEDLES LTNEFASIVDFIKAKDISLLRKS
 LPSCCKIAPNLWEKAVGSHSTNELKLLTKYTRVMLVAEP SHSDRVFSQTVLQSNQDDPELTG
 PLPSHKVGKASYLFIPEFIREVNLDKISKLDLSAKSKLAVEQVKLSELTSDFKQIENQSEAYF
 GLSTSFNELSNFLGILIRTLRNAPEAILKDQIALCAPLDDKILKITLDWLCDRAQALPENPRFE
 TNWAEYRSYLGKIKSWFSNYENFFEIPQAASSQONNNREKKLGNRSAIRALNLKEAFEKARE
 TFKGDKGTLEKIDLAYRLLGSISPEVLQCDEGLKLYQQFNDELLVLNETINQKFQDAKRDIKAK
 KEKESFEKLQRNLSSPLPRIPEFFGERAKKGYQKARVSPKLARHLLLECLNDWLARFAKVEESAF
 SEKEFQRILDWLRTSDFLPVFIRKSKDPSPWLRYIARVATGKYFFWVSEYSRKRQI IDKP I AQ
 NPLKELISWFLNLDKDAFSRDNE LFKGLSSKMVT LARIMAGILRDRGEGLKELQAMTSKLDNIGL
 LHPSFSVPVTD SLKDAAFYRAFFSELEGLLNIGRSRLTIERITLQSQQSKNKKTRRPLMPEPFI
 NEDKEVFLAPPKFETKNKVKGTRVVYN SPDEVNLLSPIRSSKGQLSFMFRCLSEDAKIMTTSG
 GCSYIVEFKLLLEAQEEVLSIHDCDIIPRAFVSIPFTLERESEETKPDWKPNRFMGVDIGEYAV
 AYC VIEKGTDSIEILD CGIVRNGAHRVLKEKVDRLKRRQRSMTFGAMDT SIAAARESLVGNRYN
 RLHAIALKHGAKLVY EYEVSAFESGGNRIKKVYETLKKSDCTGETEADKNARKHIWGETNAVGD
 QIGAGWTSQTCAKCGRSFGADLKAGNFGVAVPVPEKVEDSKGHYAYHEFPFEDGLKVRGFLKPN
 KIISDQKELAKAVHAYMRPPLVALGKRKL PKNARYRRGNSSLFRCPFSDCGFTADADIQAAYNI
 AVKQLYKPKKGYPKERKWQDFVILKPKEPSKLFDKQFYRPN (서열번호 8)

도면2ac

	237	244	251	261	265	275
1. CasY1	LKKFDISRES	-----LGLDN-NFSAPSNYFNE	----	LE--NLISRGETIKKIIVTAVLA		
2. CasY2	LYGSDETTED	-QLKMLVLCFHN-NHGAPSKYFNA	----	SLEAFGRDMSIVE	----	QITIN
3. CasY3	YEITENKKDD	-KTSNRMATYSLNQGYIPRVLND	----	LFLFVKDNEDFSQFLSDLEN		
4. CasY4	AQKLDKNEGS	LEMWEYIGIGN-SGTAFSNILGE	----	GFL-GRLENKITEIKKAMMD		
5. CasY5	AYCMGIEDFA	-GFEHLGL--DPGALSNFSNGNPKRGFL		GRLLNNDIIALANNMSA		
6. CasY6	AYCMGIEDFA	-GFEHLGL--DPGALSNFSNGNPKRGFL		GRLLNNDIIALANNMSA		
7. CasY7	RDKMSTDTAE	-DCLEYVGLTEEEKGNVNSLLGT	----	FL-KNLQGDGFEQIFQFMLG		
	285	292	302	310	320	330
1. CasY1	VSKSTW	---ENEPLEKRLHFLSEKAKLL	--GYPKLTSSWADYRVI	IGGKTKSWHSNY		
2. CasY2	NSPYW	---NSHRKELEKRIIFLQVQSKKT	--KTEELGKP-HEVLAISFGKFE	SWVSNY		
3. CasY3	FFSFS	---NEQIKIKERLKKKKYAFEP	P-GKPOLADKTDYASDFGKLE	SWVSNR		
4. CasY4	ITDAWRGQEO	EELEKRLRIILAALTIKD	--REPKFDNHWGGRSDINCKLS	SWLQNY		
5. CasY5	MTPYW	---EGRKGELEIRLAWLKHRAEG	YLKEPHFGNSWADHRSRIFSR	LAGWLSGC		
6. CasY6	MTPYW	---EGRKGELEIRLAWLKHRAEG	YLKEPHFGNSWADHRSRIFSR	LAGWLSGC		
7. CasY7	SYVGV	---QGKEEDVLRERLDLAEKVKR	--PKPKFAGEWSGHRMFLHGQL	KSWSNF		
	340	343	353	363	373	
1. CasY1	TEQLIKVREDL	-----KKHQIAQDKLQEDLKKKIVDS	SLREQIEAQREALPLLDTM			
2. CasY2	LRQEEEMKRO	LFGYEENKKGQKKFIVGNQQLDKIIRGT	DEYEIKAISKETIGLTQKC			
3. CasY3	IEELKKIPESV	-----SDLNNLEKIRNVILKKQNN	-----SKIQLSQKT			
4. CasY4	INQTVKIKEDL	-----KGHKKDKKAKEMINRFGES	DTKE--EAVVSSLDESIEKT			
5. CasY5	AGLKIAKDQI	-----SGVETDLFLKRLDAVPOSAP	-----SPDFIASISAL			
6. CasY6	AGLKIAKDQI	-----SGVETDLFLKRLDAVPOSAP	-----SPDFIASISAL			
7. CasY7	FRLFNEIRELL	-----ESIQSITQHATMLISYVEEKGGYH	-----PQLSQYRKT			

도면2ad

	383	392	395	396	398
1. CasY1	LKEKDFSD	-----LELV	-----RFLISDFTCSL		
2. CasY2	LKLEQLKDSV	DDYT--LSTV	-----RLIVEIRIR		
3. CasY3	IEYIRDYGV	SFKPE-LIKFSWINKTKDGQKKV	FYAKMADREFIEKLDLWMADIRSQ		
4. CasY4	MPDSDADEK	PDIPA--IATV	-----RRFLISDGLT		
5. CasY5	DRFLEAAESS	QDPAEQVRALYAF	-----HLNAPAVRSI		
6. CasY6	DRFLEAAESS	QDPAEQVRALYAF	-----HLNAPAVRSI		
7. CasY7	MEQLPALRT	KVLDP--IEMTHMSEAVRSYIMIHKS	V-----AGFLPDLLES		
	412	413	419	424	
1. CasY1	LNGSYQR	-----YIQTHE	-----ERKED	---RDVTKK	-----
2. CasY2	LNVFQF	---TYPEELIGK	-----SEK	-----DKKDAKNKRADKR	-----
3. CasY3	LNEYND	---NKVSFKKK	-----GKKIEELGVLD	FALNKA	KKNKSTKNENGW
4. CasY4	LNRVORE	DVQEAIIKE	-----RLTAKK	-----KKKPKKRKKKSDAE	DEKETI
5. CasY5	ANKAVORS	SDSQEWLIKELDAVDHLEFNK	-AFPFFSDTGKKK	KKKGANSNGAPSEEEYT	
6. CasY6	ANKAVORS	SDSQEWLIKELDAVDHLEFNK	-AFPFFSDTGKKK	KKKGANSNGAPSEEEYT	
7. CasY7	LDRDKDRE	-FLLSIFPR	-----IPKIDK	-----KTKEIVAWELPGEPEEGYLF	
	430	433	443	453	463
1. CasY1	-----	YKDYLSNLRNTPREFGE	SKKEQFNK	FINKSLPTIDV	
2. CasY2	-----	YPQIFKDIKLTFNFLGEIKOMV	VKKFIR	SAD-ILYE	
3. CasY3	-----	QKLSESIQSAPLEFFEGE	GNVRNEEVYN	LKD-LLFS	
4. CasY4	DFKEL	-----FPHLAKPLKLVPEVYGD	SKREL	VKKYKNAI	-YTDA
5. CasY5	ETESIQQP	EDAEQEVNGQEGNGASKNQK	FORTPREFEGESSE	VRILTEAPQ	-YFDM
6. CasY6	ETESIQQP	EDAEQEVNGQEGNGASKNQK	FORTPREFEGESSE	VRILTEAPQ	-YFDM
7. CasY7	TANNL	-----FRNFLNPKHVPRFMAERIPED	WTRLSAPV	-WFDG	

도면2ae

	473	483	490	497	507	
1. CasY1	GLKILEDTRNALETVSVERK---PPSITETEVTKQ---LEKISRKYKINAFNS-----					
2. CasY2	GINFTDQIDKQITQNL LPC-FKNDKERIEETEKQ---FETLRKKYIL--MNS-----					
3. CasY3	FIKNVENILMSSEAEDLKN--IKIEYKEDGAKKGNVYVINVLAFFIAR--FNEDGYGGW					
4. CasY4	LWKAVEKITYKSAFSSSLKNSFFDTDFDKDFFIKR---LQKIFSVYRR--FNT-----					
5. CasY5	FCNNMRAIFMQLESQPRKA---PRDFKCFELQNR---LQKLYKOTFLN-ARS-----					
6. CasY6	FCNNMRAIFMQLESQPRKA---PRDFKCFELQNR---LQKLYKOTFLN-ARS-----					
7. CasY7	MVKQWQKVVNQLVESPGAL---YQFNESFLROR---LQAMLTVMKRD-LQT-----					
	512	523	522	525	535	
1. CasY1	NRFKQITEQV-----LRKYNNGELPKISEVFYR---YPRFESHVAIR					
2. CasY2	SRFHVVIEGTINNRKLIEMKKRENSELKTSISKFVLSKFLKKGKKYENEVYTYFY					
3. CasY3	NKVKTILENT-----AREAGTDFSKYGNNNNRNAGRFLN---GTFRQVFTL					
4. CasY4	NKWKPIIVKNS-----FAPYCDIVSLAENEVLVK---PKQSRSRKS					
5. CasY5	NKCRALLESV-----LISWGEFYTYGANEKKFR---LRHEASERS					
6. CasY6	NKCRALLESV-----LISWGEFYTYGANEKKFR---LRHEASERS					
7. CasY7	EKFIKILLADV-----CRPLVDFFGLGGNDIIFKSCQDPKQWQTVI					
	550	551	561	571	577	587
1. CasY1	I-----LPVKISNPRKIDTSYLLDKYQISPDW---KNSNPGEVMDLIEIYKL					
2. CasY2	INPKARDORRIKIVLDINGNNSVGITQDLVQKLKPKWDDIICKNDMGELEIDATEIEKV					
3. CasY3	IKF---EK-----SITVEKILELVKLPSSLDEAYRDLVNENKNHKLRLDVILSKIT					
4. CasY4	AAI---DKNRVRLPSTENIAKAGTATARELSVAGRDKDLLKKEEHEEYIDHIELHKIT					
5. CasY5	SDP---D-----YVVQALELARRLFLFGFEW---RDCSAGERMDLVEIHKK					
6. CasY6	SDP---D-----YVVQALELARRLFLFGFEW---RDCSAGERMDLVEIHKK					
7. CasY7	P-----LSVPADVYTACEGLATRLRET LGFEW---KNLKGHREDFLRHLQL					

도면2af

	597	607	615	625	630	640
1. CasY1	TLGWLLSCNKDFSMDFSSY---DLKLPPEAASLTKNF---GSCLSGYYSKMTFNCCI					
2. CasY2	RTGILLIAYCEHKFKTKKELLSLDLFASAYQVLELE---DDPEETSCTNLTGRFLQSLV					
3. CasY3	IMALVLSHS---DKKQKQIGG-----NYII					
4. CasY4	ALALILAVT---ETOLDTSAL---DFVENGTVKDFMKTR---DGNLVLEGRFLEMFSQSIV					
5. CasY5	ALSFIILAIT---QAEMSVGSY---NWLGNSTVSRYL SVA---GTDTLYGTOLEEFINATV					
6. CasY6	ALSFIILAIT---QAEMSVGSY---NWLGNSTVSRYL SVA---GTDTLYGTOLEEFINATV					
7. CasY7	LGNLDFWTR---DAKLVVVKLE---DWMNNPCVQEVVEARKAIDLPLETFGFEVPIFLNGYL					
	650	660	678			
1. CasY1	TSEIKGMITLYTRDKFVVRVVTQMIGSINQKFPLLC-----					
2. CasY2	CSEIKGAINKISRTEYTERVTVMPMNTEKNYPLLII-----					
3. CasY3	HSKLSGYNALISKRFISRYSVQTTNGTQCKLAIG-----					
4. CasY4	FSEIRGLAGLMSRKEFITRSATQTMNGKQAEALLYTPHEFQS-----AKIT					
5. CasY5	LSQMRGLAIRLSSQELKDGEDVQLESSECQDNLQHLVYRASRDIAA-----CKRAT					
6. CasY6	LSQMRGLAIRLSSQELKDGEDVQLESSECQDNLQHLVYRASRDIAA-----CKRAT					
7. CasY7	FSEIRQLELLLRKISVMTSVSVRTTGSIPNRLFQIVYLP LNPSDPEKKNSNNFQERLD					
	679	687	697			
1. CasY1	-----LVGEKQTK-NFSRNNQVLTTEEKG---TL					
2. CasY2	-----NKEGKATW-----HTAAKD---TL					
3. CasY3	-----KGKSKGN-EIDRYEYAFQFFKN---LD					
4. CasY4	TPKEMSRAFLDLAPAEFATSLEPESLSEKSLKLLKQMR-YYPHYEGYETTRTGQGI DG					
5. CasY5	CPAELDPKILVLPVGAFIASVMKMIERGDEPLAGAYLR-HRPHSEGWQTRVKGVA-EV					
6. CasY6	CPAELDPKILVLPVGAFIASVMKMIERGDEPLAGAYLR-HRPHSEGWQTRVKGVA-EV					
7. CasY7	TPTGLSRRFLDLTLDAFAGKL---LTD PVTQELKTMAGFYDHLTFGFKTPCKLAAMSN					

도면2ag

	704	711	721	729	737	747			
1. CasY1	GEEK	---	NQEKCLIFKDKTDFAK	---	AKVEVEIFKNNT	---	WRIRTSKYQIQFLNRLFKKT		
2. CasY2	SKKKG	---	GTAMNQKIGKNFFGKQDYK	---	TVFMLQDKR	---	FDLLTSKYHLQFLSKTLDTG		
3. CasY3	SKIN	---	---	---	---	---	LEVIKNNSHK	---	NIDFNDNENKTNALQVYSSNYQIQFLDWFFEKH
4. CasY4	GVAE	---	NALRLKSPVKREIKCKQ	---	YKTLGRGQNK	---	VLYVRSSYYQIQFLDWFFELHRIP		
5. CasY5	GMDQ	---	GTALAFQKPTSEPFK	---	IKPFSAQYGPV	---	LMLNSSYSQSQYLDGFLISQIP		
6. CasY6	GMDQ	---	GTALAFQKPTSEPFK	---	IKPFSAQYGPV	---	LMLNSSYSQSQYLDGFLISQIP		
7. CasY7	HPGSS	---	SRMVLAKPKKGVASNIG	---	FEPIPDPAHPV	---	FRVRSSWPELKYLEGLLYLP		

	754	762	772	782	789	799				
1. CasY1	---	KEWDLNNLV	ISEPSILVLE	EEWGVSWDKDKLLPL	---	KEKSCEERYIYSLP				
2. CasY2	GGSWW	KNKN	---	TDNLSSYSIFIFKQKVKVEWDL	---	TNLDHPTIKIPSEN	---	DDRRIFTMSIP		
3. CasY3	---	QGKK	---	TSLEVIGGSFTIAEKSL	---	TDWSGSNPRVGF	---	RSDEEKRVFMSQP		
4. CasY4	---	KNVQ	---	TDVAVSGSFILDEK	---	KKVKTWNNDALTVALE	---	PVSGSERVETVSQP		
5. CasY5	---	KNWS	---	MRVLPQAGSVRVE	---	ORMALIWNLOAGKMRL	---	RSGARAFFMPVP		
6. CasY6	---	KNWS	---	MRVLPQAGSVRVE	---	ORMALIWNLOAGKMRL	---	RSGARAFFMPVP		
7. CasY7	---	EDTP	---	LITIEELAETS	---	VSCQSVSSVAFDLKNLT	---	TITLG	---	RVGEFRVTADQP

	809	818	827	836	842	851					
1. CasY1	LNLV	PATDYKEQSAE	---	IEQRNTYLGIDVGE	---	FGVAVAVVR	---	VRDRIEILLSWG			
2. CasY2	EVIKPK	---	QTKRK	---	LIQTRVNYMGIDIGE	---	YGLAWTTINIDL	---	KNKKINKLSKQG		
3. CasY3	ETLIP	DDDEKERRK	---	ERMIKTKNRFIGIDIGE	---	YGLAWSLIEVDNGDK	---	NNRGIRQLIESG			
4. CasY4	ETIFPE	---	KSATFE	---	---	GQRYLGIDIGE	---	YGLAYTALEI	---	TGDSAKILIDQN	
5. CasY5	ESFRPS	---	SGGDEA	---	VLAPNRYLGLFP	---	HS	---	GGIEYAVVDVL	---	DSAGFKILERG
6. CasY6	ESFRPS	---	SGGDEA	---	VLAPNRYLGLFP	---	HS	---	GGIEYAVVDVL	---	DSAGFKILERG
7. CasY7	EKLTP	---	IPEKEE	---	SFIGKTYLGLDAGERS	---	GVGFALIVTV	---	---	---	DGDGYEVORLG

도면2ah

	861	870	876	886	896
1. CasY1	F L K D P A L R K I R E - R V Q D M K - - - - K K Q V M A V F S S S S T A V A E V R E M A I H S L R M C I H S H I A L Y				
2. CasY2	F L Y E P L T H K M R D - Y V A T I K - - - - D N Q V R G T E G M P D T K L A R I L R E N A I T S L R M C V H D T I A M				
3. CasY3	F L T D N Q Q V L K K - N V K S W R - - - - D N Q I R Q T E T S P D T K I A R I L R E S L I G S Y K N C L E S I L V V				
4. CasY4	F I S D P Q L K T I R E - E V K G L K - - - - L D Q R E G T A M P S T K L A R I L R E S L V H S L R M R I H I A L Y				
5. CasY5	T L A V N G F S Q K R G - E R Q E E A H - - R E K Q R R G I S D I G R K K P V Q A E V D A A N E L H R K Y T D V A T				
6. CasY6	T L A V N G F S Q K R G - E R Q E E A H - - R E K Q R R G I S D I G R K K P V Q A E V D A A N E L H R K Y T D V A T				
7. CasY7	V H E D T Q L M A L Q Q V A S K S L K E P V F Q P L R K G T I - - - - R Q E R I R K S L R G C Y W T F Y H A L V I T				

	906	916	926	936	942	947
1. CasY1	A Y K A K I L Y E I S I S N F E T G G N R M A K T Y R S I T K V S I X V Y R - - - - - E S G A D T L V S E M I					
2. CasY2	R Y D A K P V Y E F E T S N F E T G S N K K V I T Y D S V K R A D T I G R G Q - - - - - N N T E A D N T E V N L V					
3. CasY3	A K K A N I S F E Y E V S G F E V G G K R V A K I Y D S I T K R G S M R - - - - - A D N N S Q N D Q S					
4. CasY4	R H K A K I V Y E L E M S R F E E G K Q I K K V Y A T L K K A D V Y S - - - - - E I D A D K N L Q T T V					
5. CasY5	R L G C R I V V Q W A P Q P K P G T A P T A Q T V Y A R A V R T E A P R S - - - - - G N Q E D H A R M K S S					
6. CasY6	R L G C R I V V Q W A P Q P K P G T A P T A Q T V Y A R A V R T E A P R S - - - - - G N Q E D H A R M K S S					
7. CasY7	K Y R A K V V H E S V G S S G L V G Q W L R A F Q K D L K K A D V L P K K G G K N G V D K K R E S S A Q D T L					

	956	967	977	986	996
1. CasY1	W G - - - - - K N K N Q M G N H I S I S Y A T S Y T C N Q A R - T P F E L V I D N D K E Y E K G - - - - G				
2. CasY2	W G - - - - - K T S K Q F G S Q I G A Y A T S Y I C S F G Y - S P Y Y E F E N S K S G D E E G - - - - A				
3. CasY3	W G - - - - - K K G I N E W S F E T T A G T S Q F C T H C K R - W S S L A I V D I E E Y E L K D - - - - Y				
4. CasY4	W G - - - - - K L A V A S E I S A S Y T S Q F C G A C K K L W R A E M Q V D E T I T T - - - - -				
5. CasY5	W G Y T W G T Y W E K R K P E D I L G I S T Q V Y W T G G I G E S C P - - - - -				
6. CasY6	W G Y T W S T Y W E K R K P E D I L G I S T Q V Y W T G G I G E S C P - - - - -				
7. CasY7	W G - - - - - G A F S - K K E E Q Q I A F E V Q A C S I S Q F C L K C G W - W F O L G M R E V N R V Q E S G V L D D				

도면2a

	1,002	1,009	1,019	1,028	1,042
1. CasY1	DEFIFNVGDIIEKKVRGFLQKSL	-GKT	IKGKEVLKSIKE	YARPP	IREV-----
2. CasY2	RDNIYQMK-----	KLSRPSLE--	DFLQGNPVYKTF	DFDKYKNDQR--	-----
3. CasY3	NDNIFKVKIN--	DGEVRLLGKKGWSRSGEKIKGKELFGPV	KDAMKPNVDGL	-GMKIVKR	
4. CasY4	MELIGTVR-----	-----	VIKGGTLIDAIKDFMR	PIFDENDTPFPKY	
5. CasY5	-----	-----	-----	-----	-----
6. CasY6	-----	-----	-----	-----	-----
7. CasY7	NRSITVTFLIESSGEKVYGFSPQQL	EKGFRPDIETFKKMVR	DFMR	PPMFD	RKGRPAAAY

	1,043	1,046	1,055	1,062	1,070	1,080
1. CasY1	---LLEGEDVEQLLK---	RRGNSYIYRC	PF--	CGYKTDADIQAALN	ITACRGYISD	
2. CasY2	---LQKTGDKDGEWKT---	HRGNTAIYACOK--	CRHISDADIDIASY	WTAL	KQVVRD	
3. CasY3	KYL---KDLDRDWVS---	RYGNMAIFICPYVD	CHHISHADKQA	FNIAV		
4. CasY4	RDF----CDKHHISK---	MRCNSCLFICPF--	CRANADADIDIAS	QTIALIRYVKE		
5. CasY5	-----	-----	-----	-----	AVAVALLGHIRA	
6. CasY6	-----	-----	-----	-----	AVAVALLGHIRA	
7. CasY7	ERFVLGRRHRRYRFDKVF	EERFERSALFIC	PRVGCIGNFDH	SEESAVV	LALIGYIAD	

	1,090	1,100	1,110	1,114	1,122	1,125
1. CasY1	NAKDAVKGERKLDYILEVRK	LWE---	KNGA	VILRS	KFL	
2. CasY2	FYKDKEMGD	LIQGD	NKDKRKVNELNRL	IGVHKDVP	TIINK	NLITS
3. CasY3						LL
4. CasY4	EKKVE-----	DYFERFRK	---	KNIKV	LIGOV	KKI
5. CasY5	TSTQTEW	-----	-----	KEEV	VFGRL	KKFFPS
6. CasY6	TSTQTEW	-----	-----	KEEV	VFGRL	KKFFPS
7. CasY7	KEGMSGKK---	LVIYVRLAELMAE	-----	WKLK	LIESR	VVEEQSI-SAQ

도면2ba

	818	827	836	844	850
1. CasY1	IIEQRNTYLG	LIVGEF	-GVAYAVVR	II--VRDR---	IELLS-WGFLK
2. CasY2	IQTRVNYMG	TDIGEY	-GLAWTIIIN	IDLK	NKK---INK
3. CasY3	IKTKNRF	IGTDIGEY	-GLAWSIIEM	D-NGDK	NNRGIRQLE-SGFI
4. CasY4	GQRV	LGTDIGEY	-GIAYTALET	-TGDS---	AKIID-QNFI
5. CasY5	VLAPNRV	LGLFPHSG	-GIEYAVVDV	--LDSA---	GFKILE-RGTIAVNG
6. CasY6	VLAPNRV	LGLFPHSG	-GIEYAVVDV	--LDSA---	GFKILE-RGTIAVNG
7. CasY7	GKT	VLGLDAGERS	GVGFAITV	---	DGDG---YEVQR-LGVHEDTQ
8. AsCpf1_5843	EHPETPI	IGIDRGER	-NLIYITV	-IDSTGKI	----LEQRS-----
7. LbCpf1_SID6	HDDNPYV	IGIDRGER	-NLLYIVV	-VDGKGN	I---VEQYSLNEIINN

	859	868	874	884	889
1. CasY1	LRKIRE	-RVQDMK---	KKQVMAVFSS	SST---	AVARVREMAIHS
2. CasY2	THKVRD	-YVATIK---	DNOV	RGTEGMPDT---	KIARLR
3. CasY3	QQVLRK	-NVKSWR---	QNOIR	QTFTSPDT---	KIARLR
4. CasY4	LKTIRE	-EVKGLK---	LDOR	RGTEAMPST---	KIARLR
5. CasY5	FSQKRG	-EROEEAH--	REKOR	RGISDGRK---	KPVQAEVDAANE
6. CasY6	FSQKRG	-EROEEAH--	REKOR	RGISDGRK---	KPVQAEVDAANE
7. CasY7	LMALQ	QVASISLKEPV	FQPLRGTER	-----	QOERIRKSLRGCIW
8. AsCpf1_5843	LNTIQ	QFDYQKKL---	DNRE	KERVAARQAWSV	VTIKDKLKGYSQVI
7. LbCpf1_SID6	GIRIK	T-DYHSL	L----DKKEKER	FEARQNWTSIEN	IKELKAGYISQVI

도면2bb

	899	909	919	928	937
1. CasY1	NQTHSIAIAAYKAKITVEISISNFTTGGNRM--AKTYRSIRVSQVYR----				
2. CasY2	NOVHDIAMRYDAKPNVEFEISNFTTGSNKV--KVTYDSVKRADTIGRG---				
3. CasY3	NQLESIMVAKKANLSFEYEVSGFEVGGKRV--AKTYDSIRKGSVRK----				
4. CasY4	NRTHHLAALKHKAKITVELEVSRFEEGKQKIT--KKVYATLKADVIYS----				
5. CasY5	RKYTDVATRLGCRIVVQWAPQPKPGTAPTA--QTVYARAVRTAPR----				
6. CasY6	RKYTDVATRLGCRIVVQWAPQPKPGTAPTA--QTVYARAVRTAPR----				
7. CasY7	NFYHALMIKYRAKIVHEESVGSGLVGQWL--RAFQKDLKADVLPPKGG				
8. AsCpf1_5843	HEIVDLMITHYQAVVVLLENLNFGEFKSKRTGTAEKAVYQQFEKMLIDK----				
7. LbCpf1_SID6	HKICEIVKDYDAVTALEDLNSGFKNRVRKV--EKQVYQKFEKMLIDK----				
	942	955	956	960	
1. CasY1	-----ESGADTLVSEMIWQ-----KKN-KQMGNHISISY				
2. CasY2	---QNNTEADNTEVNLVWQ-----KTS-KQFGSQIGAY				
3. CasY3	-----KDNNSQNDQS WQ-----KGINESWSETTA				
4. CasY4	-----EIDAKNLQTTVWQ-----KLA---VASEISAS				
5. CasY5	---SGNQEDHARMKSSWQ--YTWG-----TYWEKRPEDILGISTQVY				
6. CasY6	---SGNQEDHARMKSSWQ--YTWS-----TYWEKRPEDILGISTQVY				
7. CasY7	KNGVDKKRQISSAQDTLWQ-----GAFSKKEE-QQIAFETQQA				
8. AsCpf1_5843	----LNCVLKDYPAEKVQGVLPNPYQLTDQFTSFAKMGTSQSGFLFYVPA				
7. LbCpf1_SID6	----LNYMVKKSNPCATGALKGYQITNKFESFKSMSTQNGFIFYTPAW				

도면2bc

	970	980	994	995
1. CasY1	ATSYTCNCARTPFELVIDNDKEYEK-----GGDFFIFN			
2. CasY2	ATSYICSFCGYSPYEFENSKSGDEE-----GARDNIIYO			
3. CasY3	GTISQFCTHCKRWSSLAIVDIEEYELK-----DYNINLFK			
4. CasY4	YTSQFCGACKKLWRAEMQVDETITTO-----ELIGT			
5. CasY5	WTGGIGESC PAVAVALLGHIRATSTQ-----			
6. CasY6	WTGGIGESC PAVAVALLGHIRATSTQ-----			
7. CasY7	GSSQFCLKCGWWFQLGMREVNRVQESGVVL-----DWNRSIVT			
8. AsCpf1_5843	YTSKIDPLTGFVDPFVWVKTIKNHESRKHFLEGDFLHY--DVKTGDFILH			
7. LbCpf1_SID6	LTSKIDPSTGFVNLLKTKYTSIADSKK-FISSFDRIMYVPEEDLFFALD			
	1,004	1,005	1,012	1,016
1. CasY1	VG-----DEKKVRGF-----LQKSLLGKTIKG			
2. CasY2	MKK-----LSRPSLEDFLQG			
3. CasY3	VKI-----NDGEVRLLGK-----KGWRSGEKIKG			
4. CasY4	VRV-----IKG			
5. CasY5	-----TEWEKE			
6. CasY6	-----TEWEKE			
7. CasY7	FLIE-----SSGEKVYGFSP-----QLEKGFDPDI			
8. AsCpf1_5843	FKMNRNLSFQRLPGFMPAWDIVFEKNETQFDAQGTPFIAGKRIVPVIEN			
7. LbCpf1_SID6	YK-----NFSRTDADYIKKWKLYSYGNRIRIFAA-----AKKNNVFAWE			

도면2bd

	1,026	1,036	1,042	1,043	1,052
1. CasY1	KEVILKSITKEVYARIPPI-----REV-----	-----	-----	-----	-----
2. CasY2	NPVYKTFIRDIDKYKN-----DOR-----	-----	-----	-----	-----
3. CasY3	KEIFGPMKIDAMRPNV-----DGLGMKIVKRKYLK-----	-----	-----	-----	-----
4. CasY4	GTLIDATIKDFMRDPI-----FDENDTPFPKYRDFC-----	-----	-----	-----	-----
5. CasY5	EVVFGRLKKKFF-ES				
6. CasY6	EVVFGRLKKKFF-ES				
7. CasY7	ETFKKMVRDIFMRPPM-----FDRKGRPAAAYERFV-----	-----	-----	-----	-----
8. AsCpf1_5843	HRFTGRYRDLY-PANELIALLEEKGIVFRDGSNILPKLEND-SHAIDTM				
7. LbCpf1_SID6	VCITISAYKEFL-----NKYGINYQQG-DIRALICEQS-KAFYSSE				

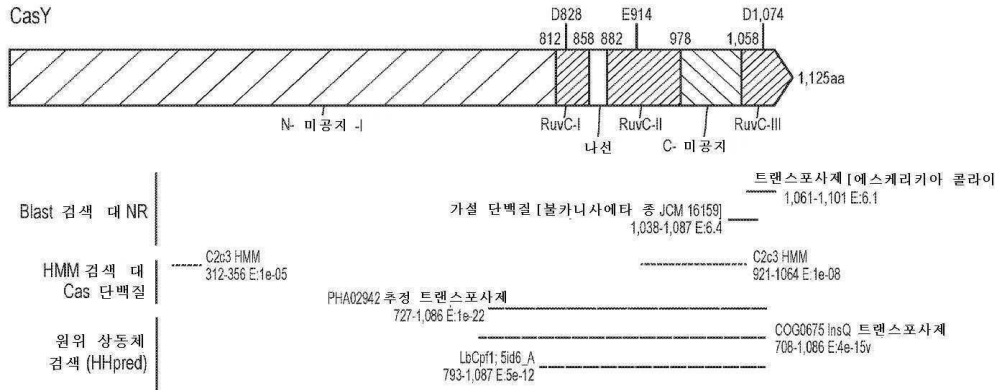
	1,053	1,061	1,067
1. CasY1	-----LKRRCNSYIYRCPF-----	-----	-----
2. CasY2	-----THRGNTAIVACQK-----	-----	-----
3. CasY3	-----LDLRDWVSRYGNMALEICPYV-----	-----	-----
4. CasY4	-----KMRGNSCLFICPF-----	-----	-----
5. CasY5			
6. CasY6			
7. CasY7	FE-----ERFGRSALFICPRV-----	-----	-----
8. AsCpf1_5843	VALIRSVLQMRNS-NAATGEDYINS	PVRDLNGVCFDSRFQNP	----WPM
7. LbCpf1_SID6	MALMSLMLQMRNSITGRTDVDFLIS	PVKNSDGIIFYDSRNYEAQENAILPK	

도면2be

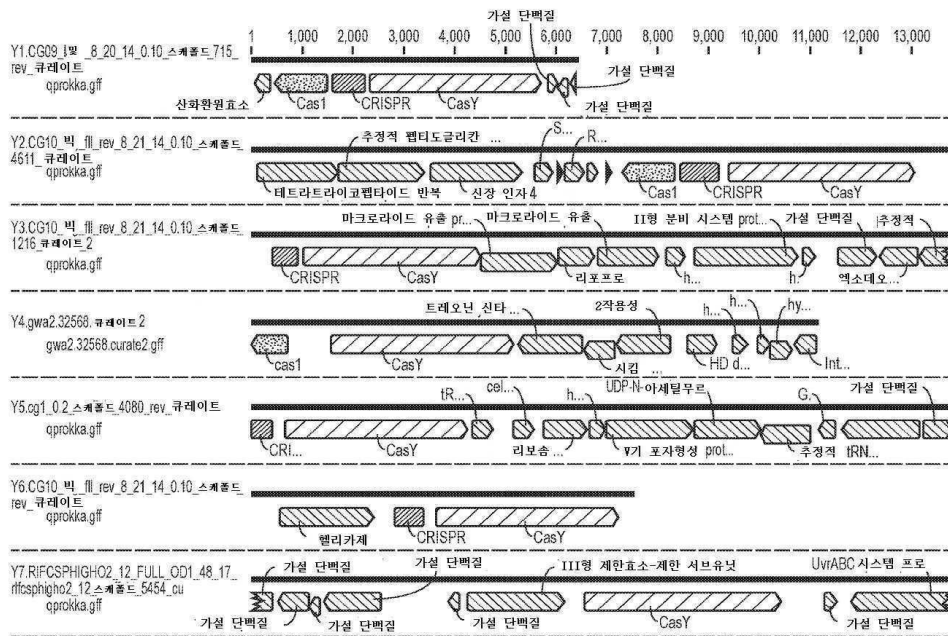
	1,072	1,083	1,093	1,103	1,113
1. CasY1	DADIQALNLTACRGYTS	DNAKDAVKEGERKLDYI	ILEVFKIWEKNGAV---		
2. CasY2	DADIQAISYWLALKQVVR	DFYKDKEMDGLIQGDN	KDKKVNELNRLIGVH		
3. CasY3	HADKQAFNLTAV				
4. CasY4	DADIQAISQTIALLRYVKEEKK	-----VEDYFERFRK	-----		
5. CasY5					
6. CasY6					
7. CasY7	HSEQSAVVVLALLIGYTADKEG	---MSGK-KLVYVRLAELMAEWKLKK---			
8. AsCpf1_5843	DADANGAYHTALKGQT	-----LINHLKESKDKLQN-----			
7. LbCpf1_SID6	NADANGAYNTARKVVLW	-----AIIQGFKAEDEKLDKVKI---			

	1,119	1,125
1. CasY1	---LRSKFT	
2. CasY2	KDVPINKNLTITSLDINLL	
3. CasY3		
4. CasY4	---LKNIKVIG--QMKKI	
5. CasY5		
6. CasY6		
7. CasY7	---LERSRVEE--QSIAQ	
8. AsCpf1_5843	---GISNQDWLTAYIQELRN	
7. LbCpf1_SID6	---AISNKEWLEYAQT	SVK

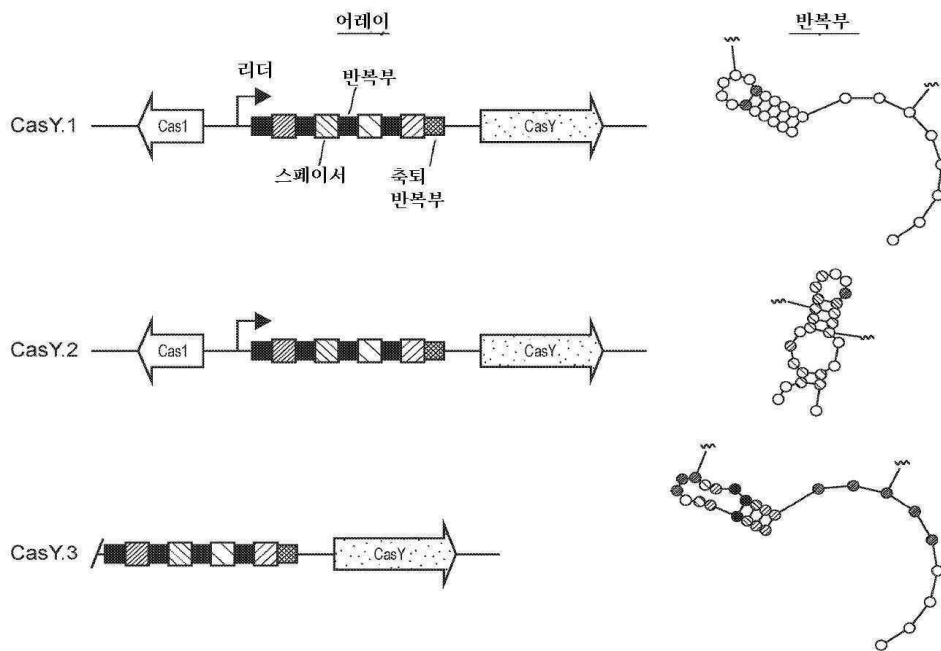
도면3a



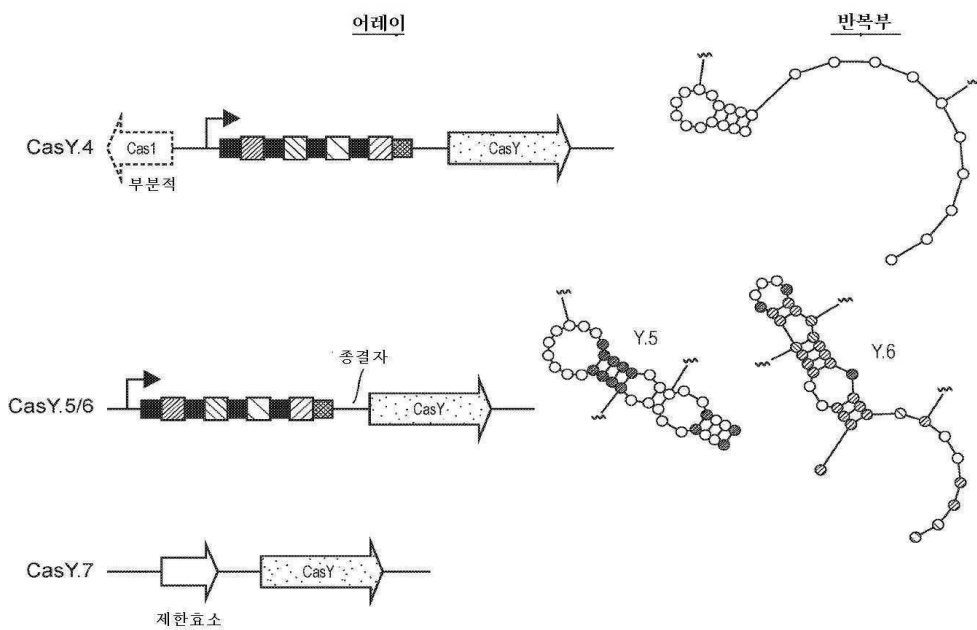
도면3b



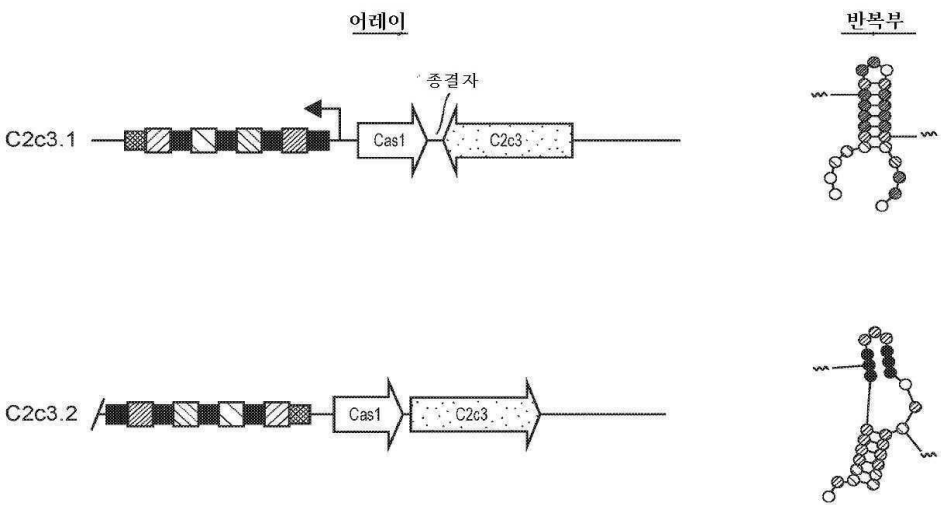
도면4a



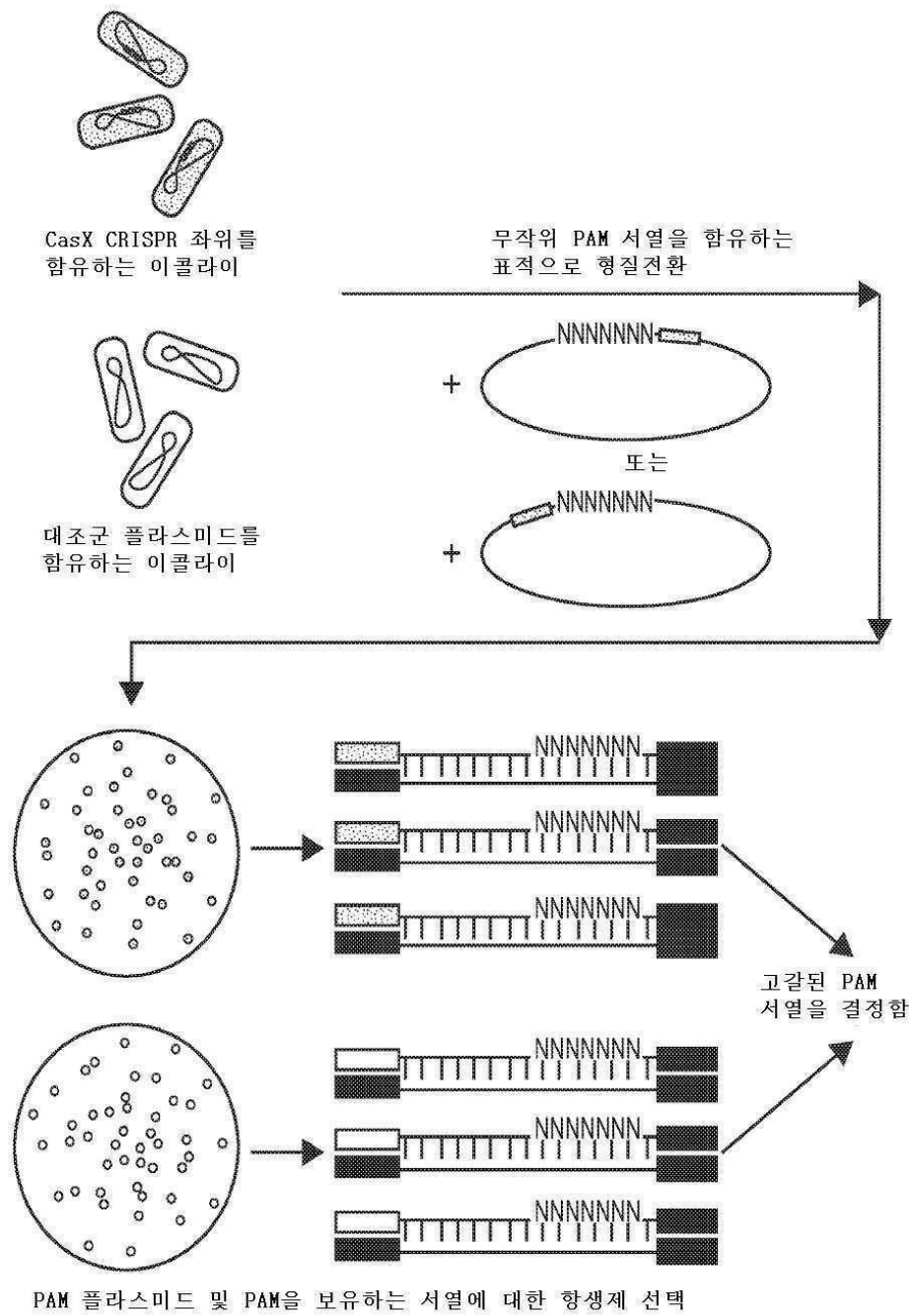
도면4b



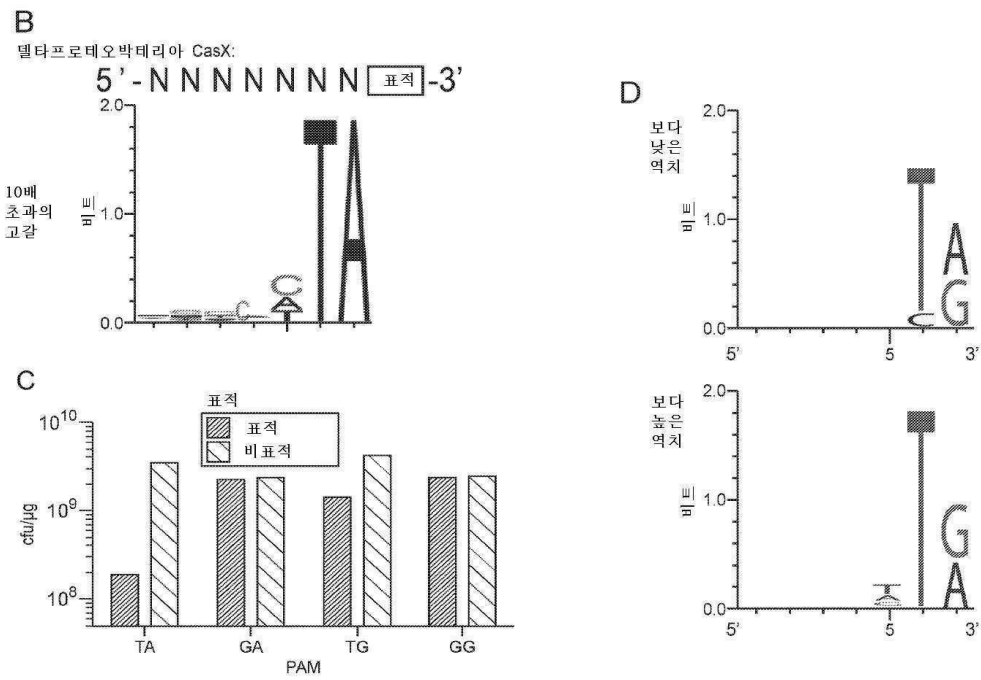
도면4c



도면5a



도면5bd



도면6

A

반복부

CasY1 CTCCGAAAGTATCGGGGATAAAGGC
 CasY2 CACCGAAATTTGGAGAGGATAAGGC
 CasY3 CTCCGAATTATCGGGAGGATAAGGC
 CasY4 CCCCCAATATAGGGGACAAAAAGGC

 CasY5 GTCTAGACATACAGGTGGAAAGGTGAGAGTAAAGAC
 (및 Y6)

B

PAM

5' GACATGATCGCTAATCAATACCAAACCTCTGGACCGAATTC
 |||||
 3' CTGTACTAGCGATTAGTTATGGTTTGAGACCTGGCTTAAG

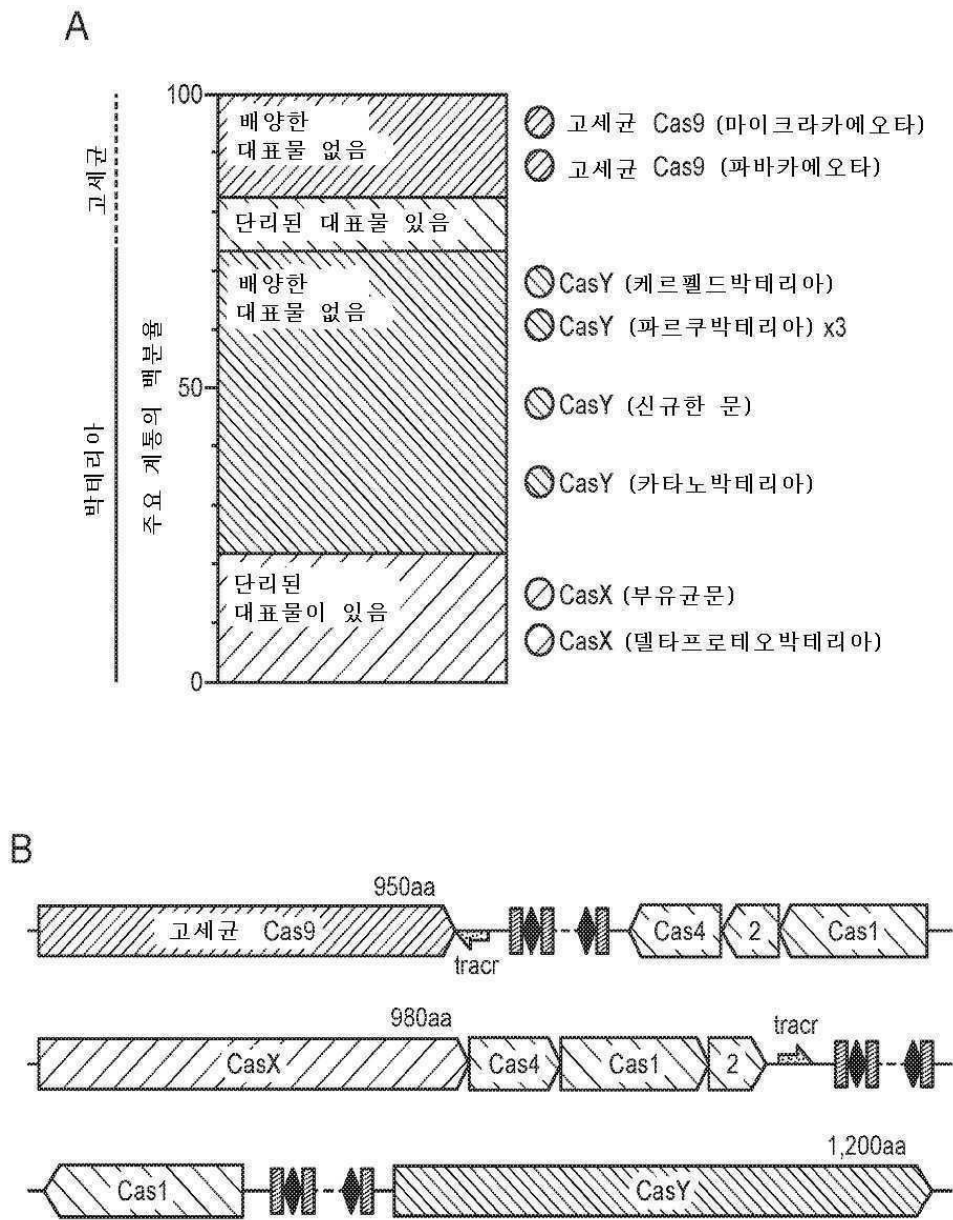
T A TCGGGGATAAAGG C ATCAATACCAAACCTCTGG 3'

G |||||
 AAAGCCTC 5'

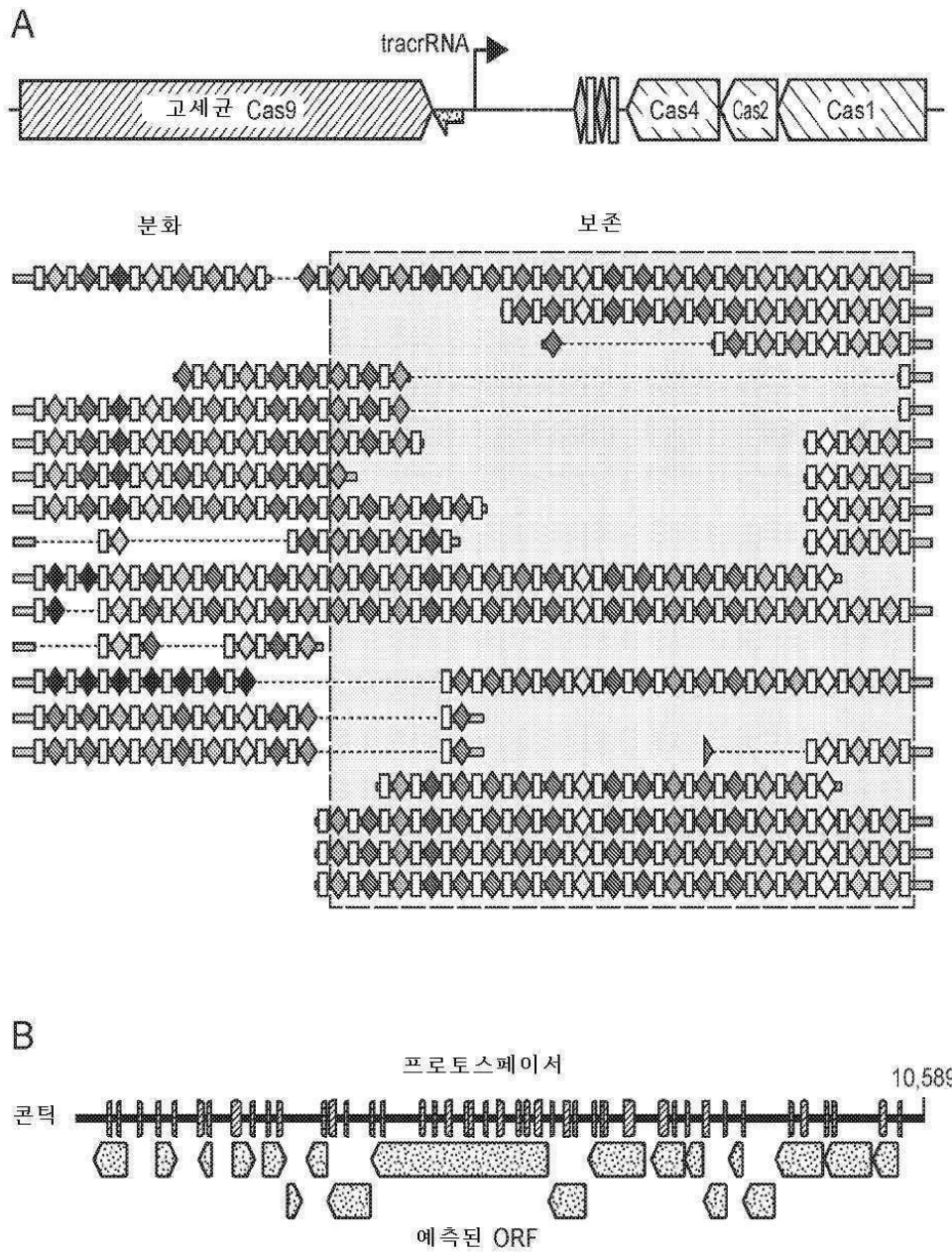
반복부

가이드 서열

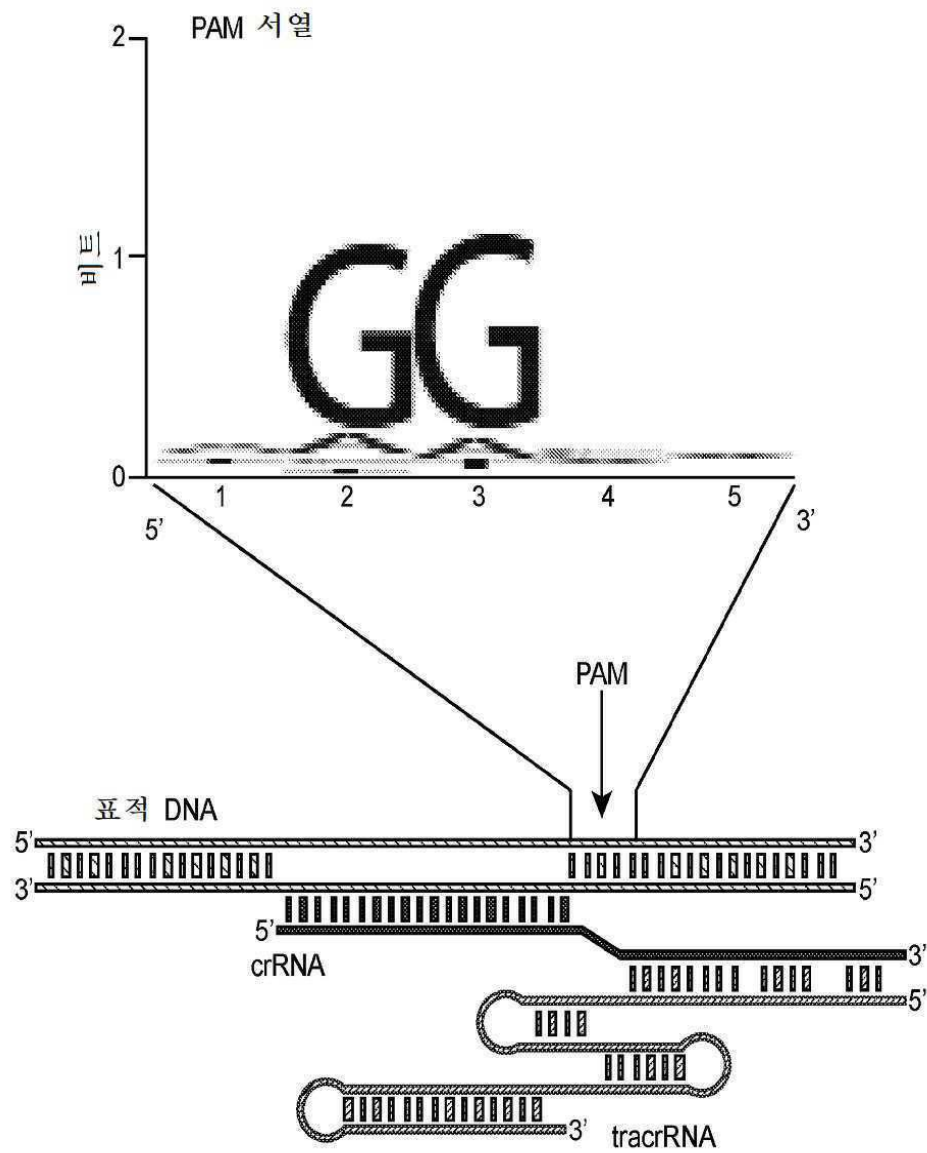
도면7



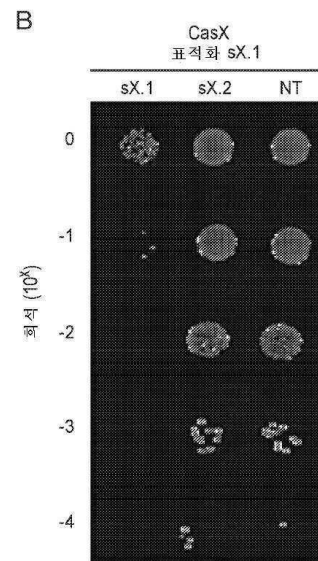
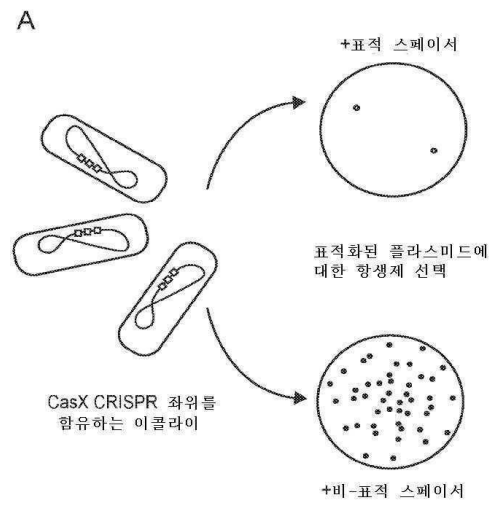
도면 8ab



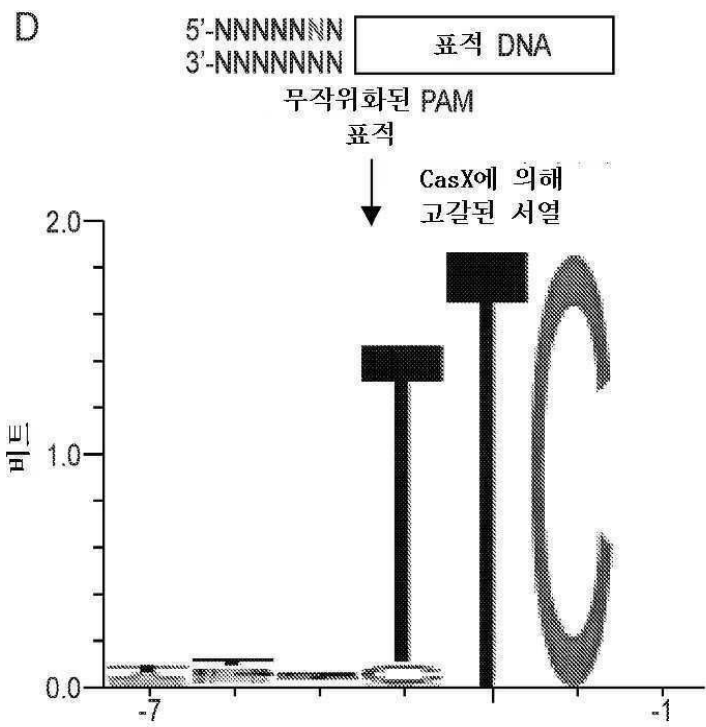
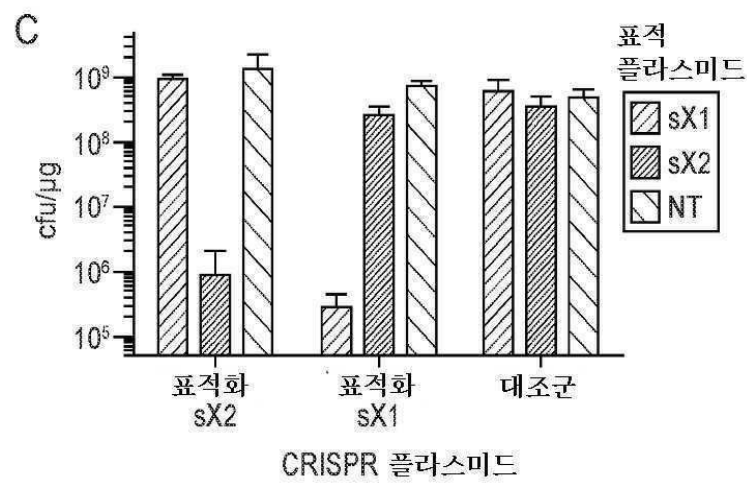
도면8c



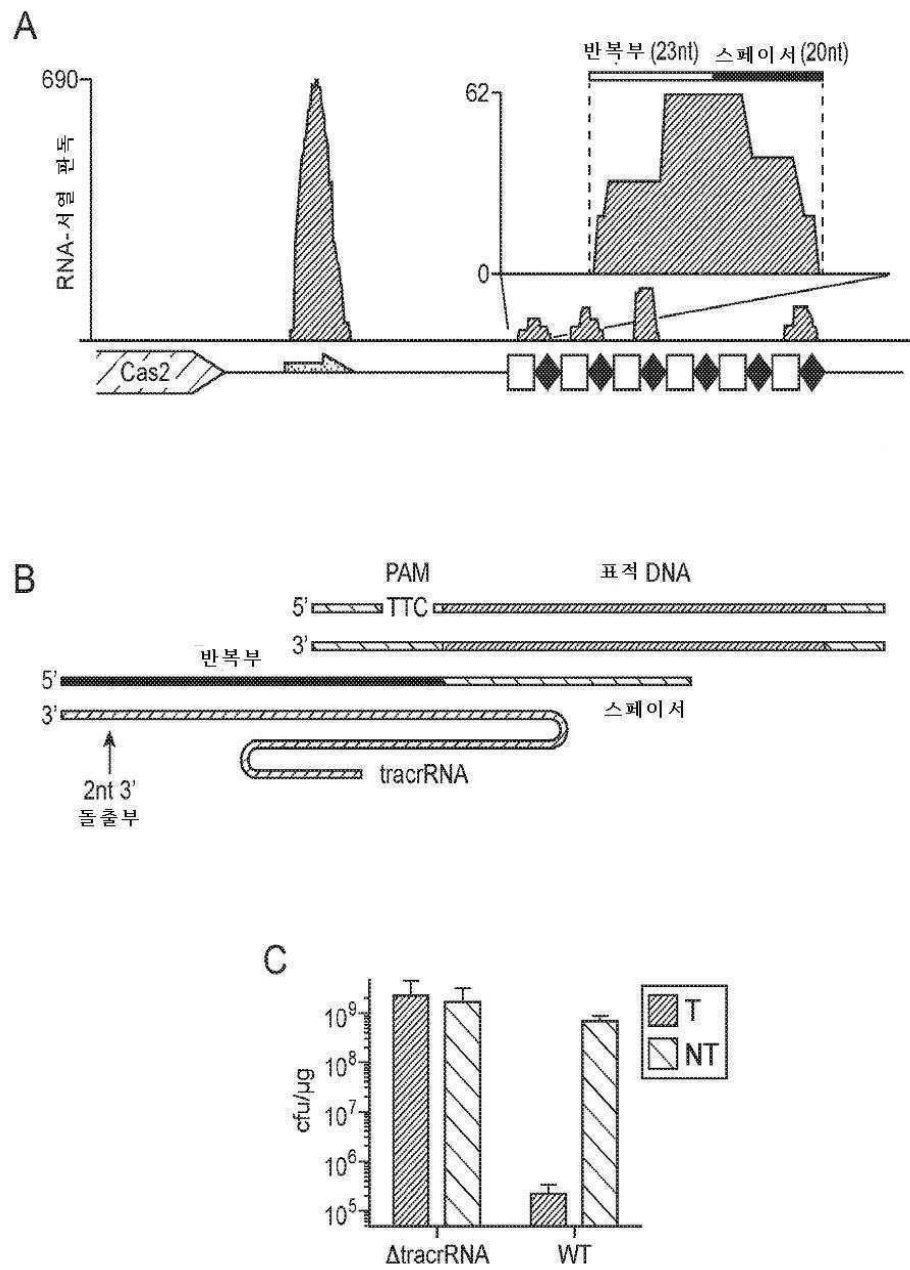
도면9ab



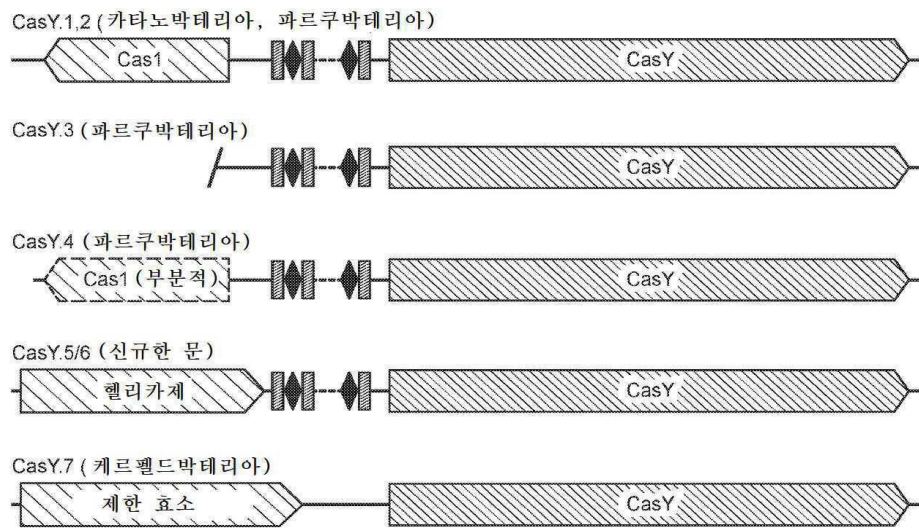
도면9cd



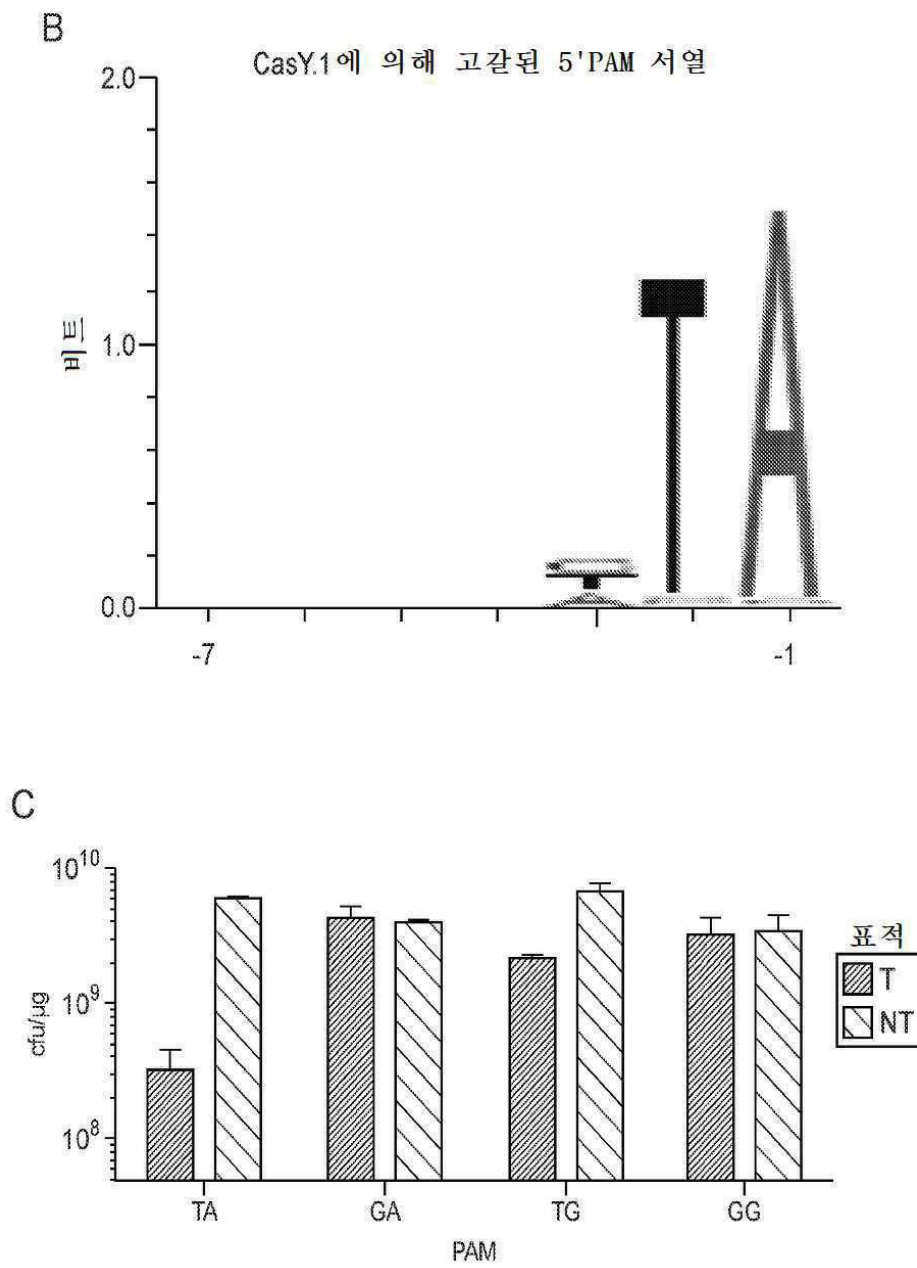
도면10



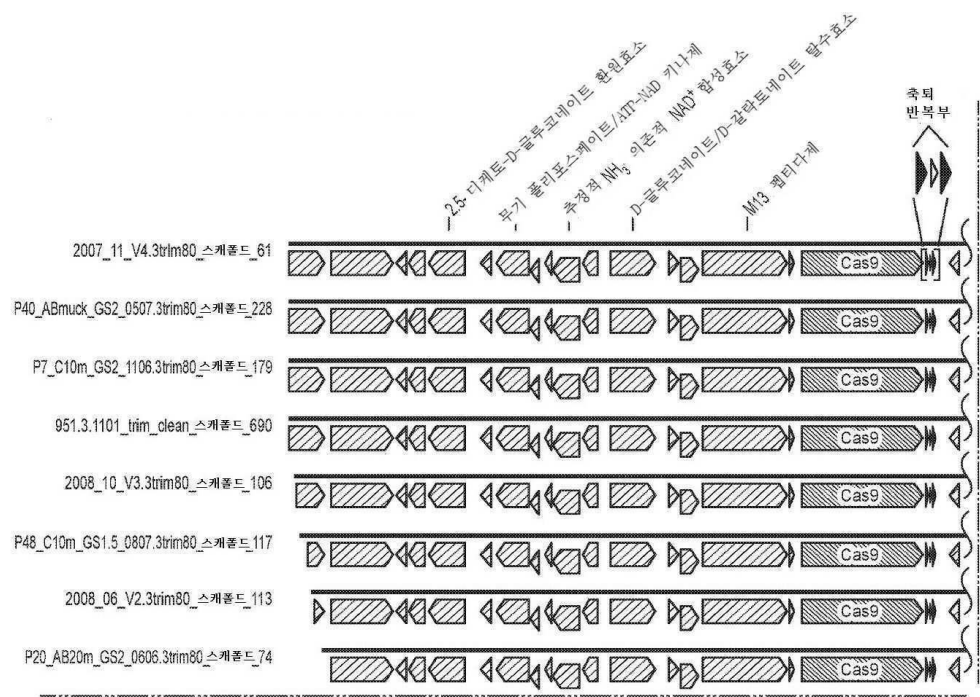
도면11a



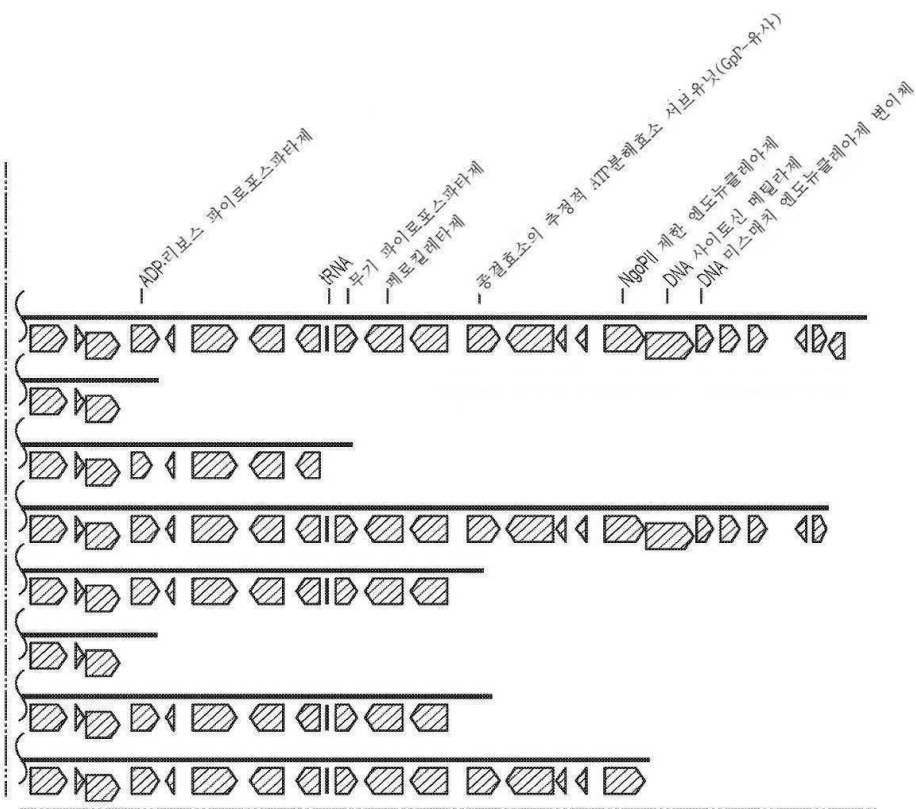
도면11bc



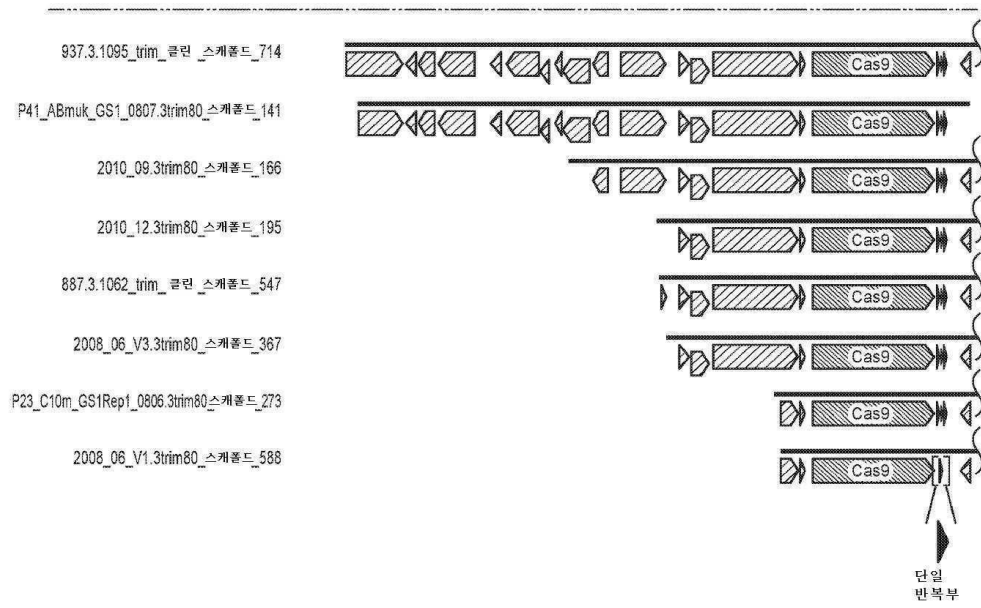
도면13a



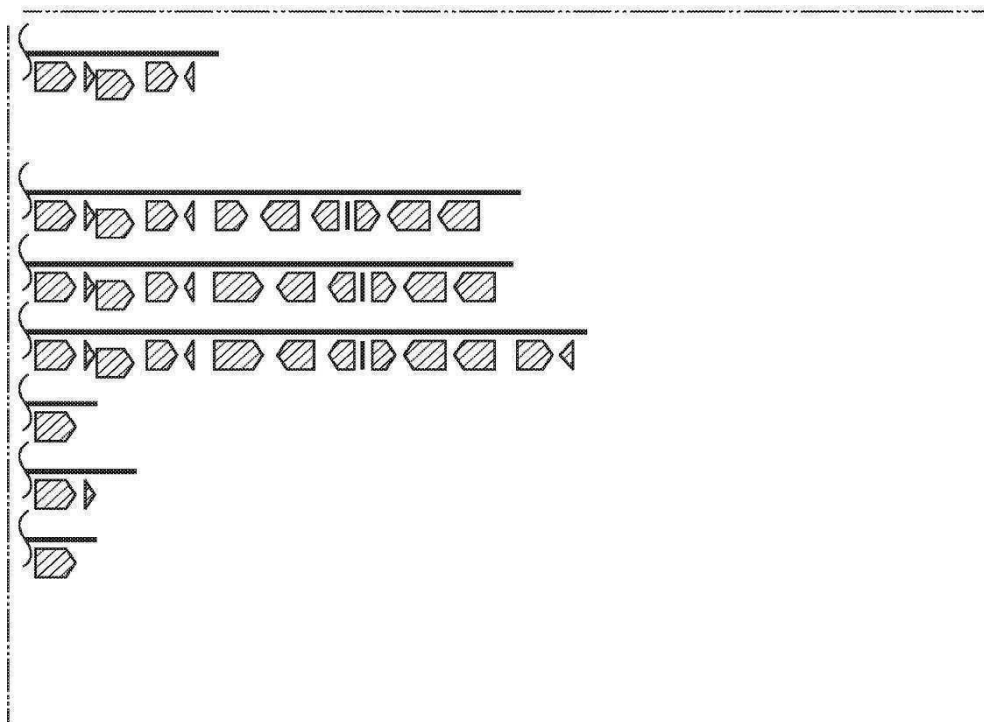
도면13b



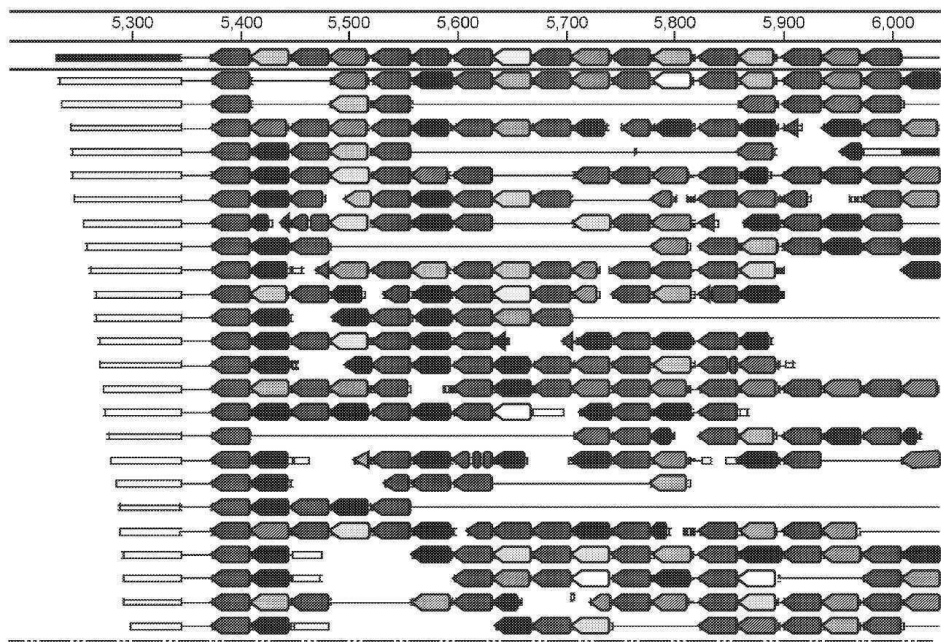
도면13c



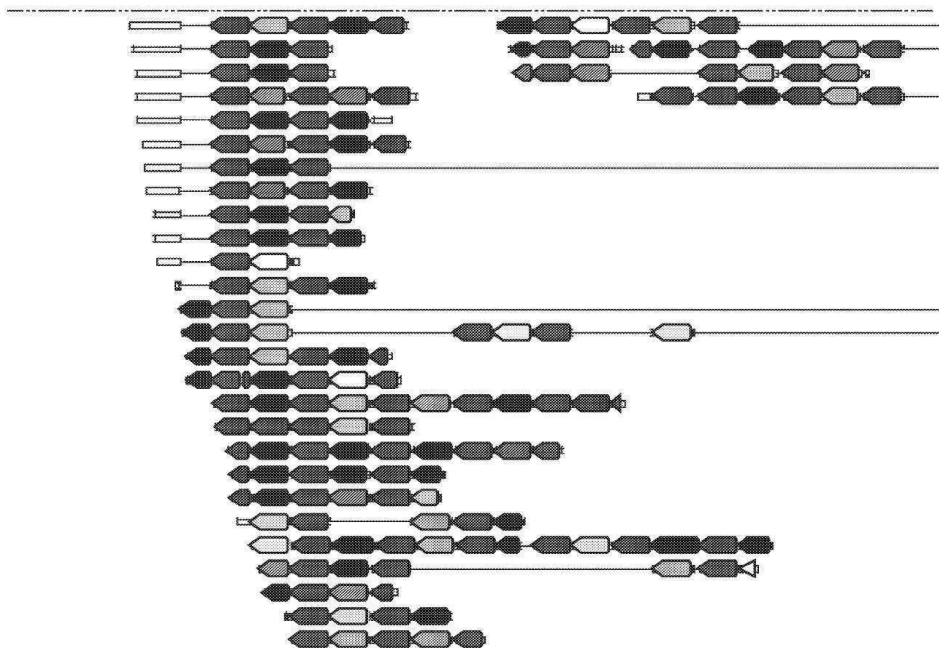
도면13d



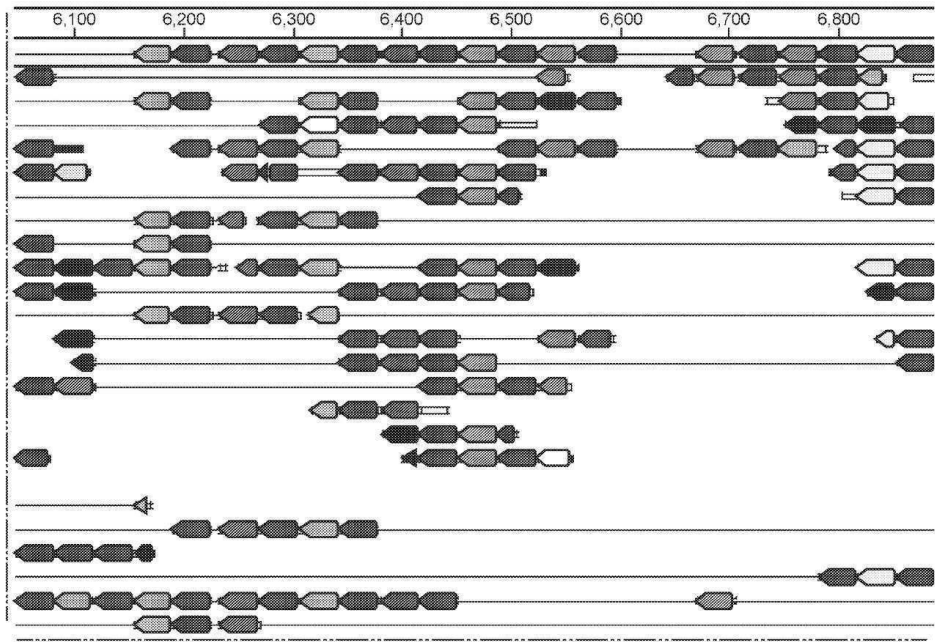
도면14a



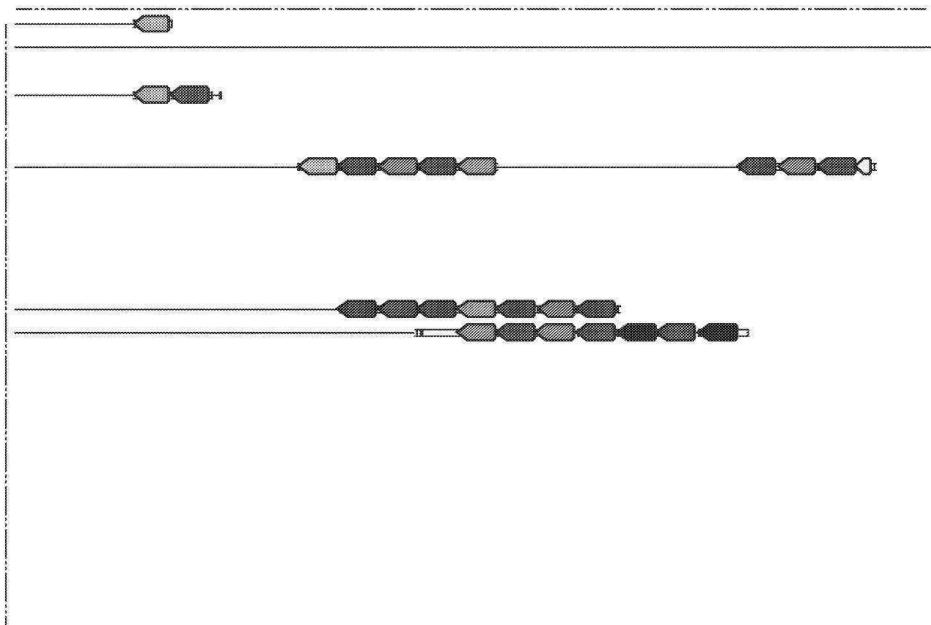
도면14b



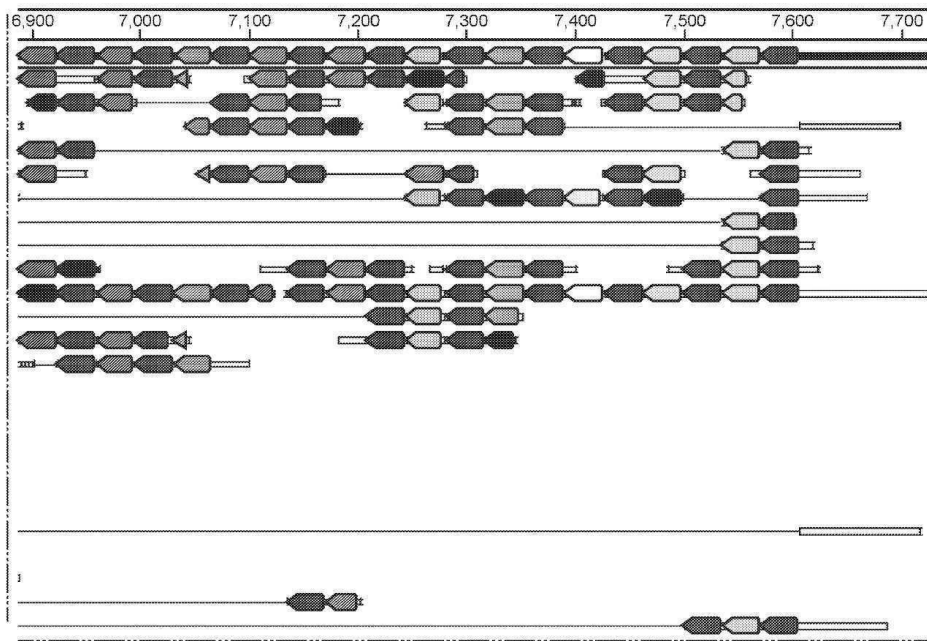
도면14c



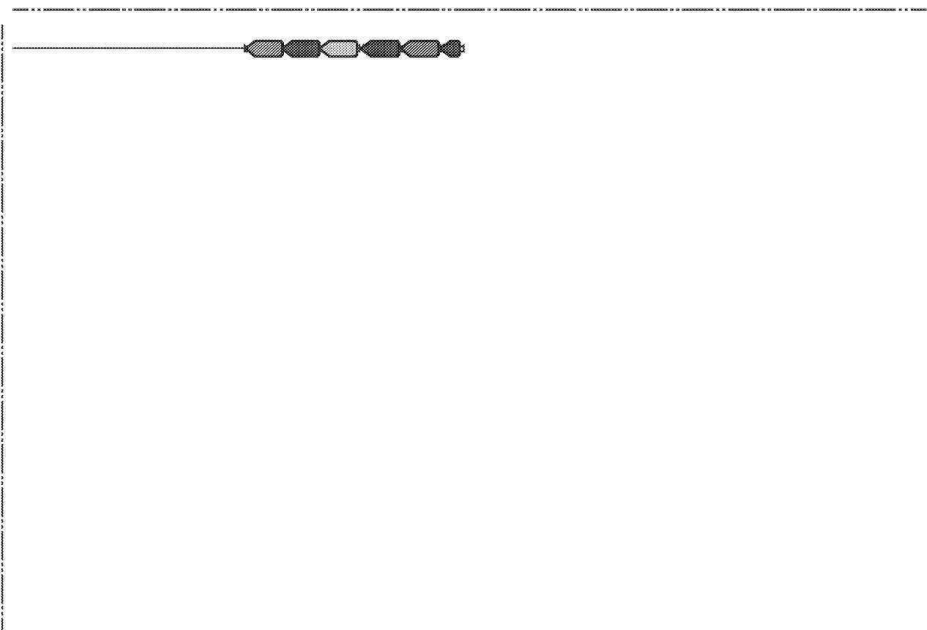
도면14d



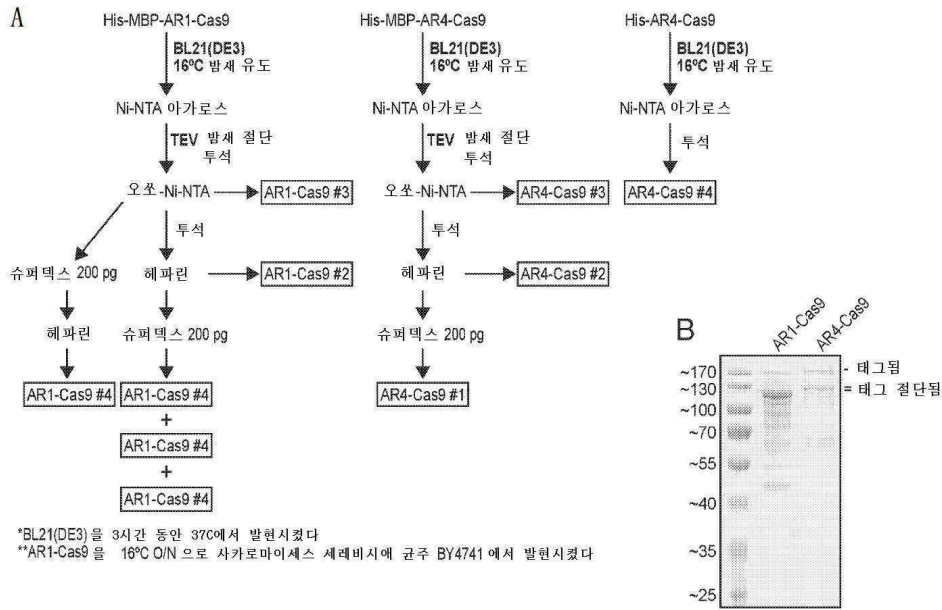
도면14e



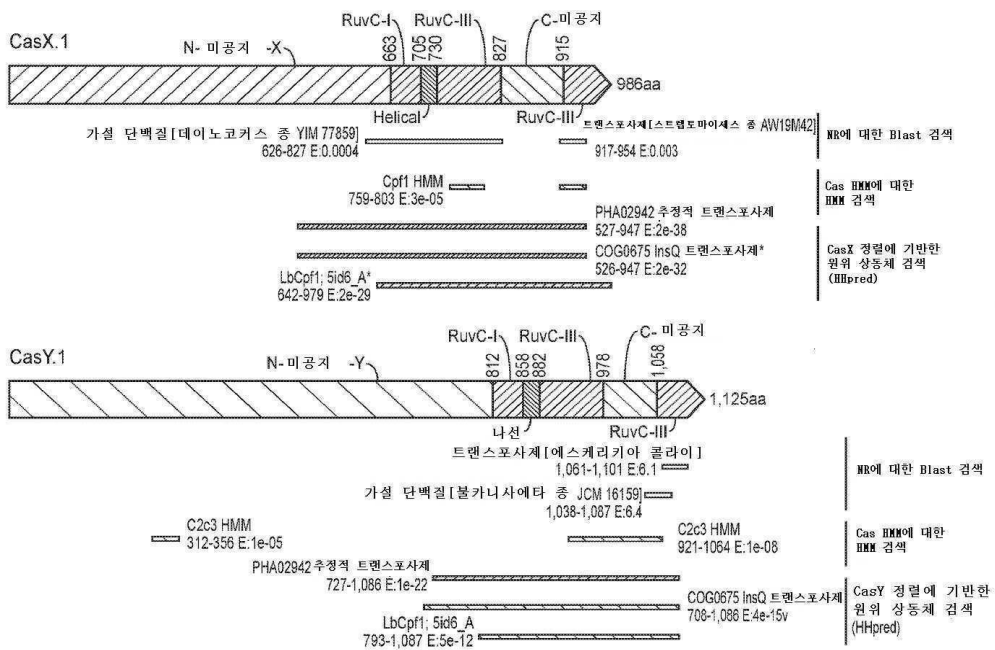
도면14f



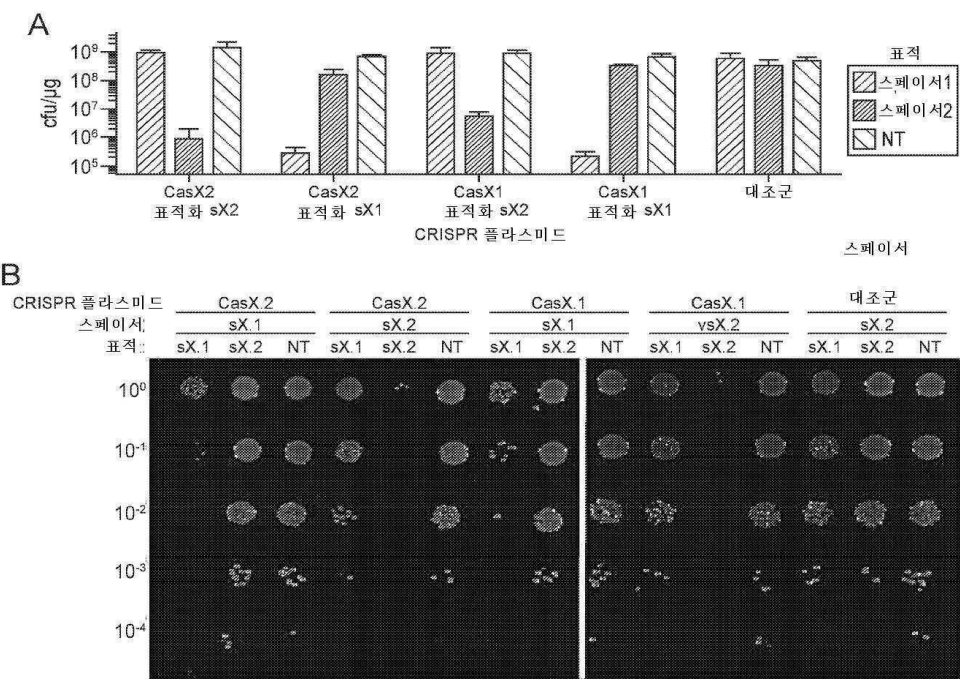
도면17



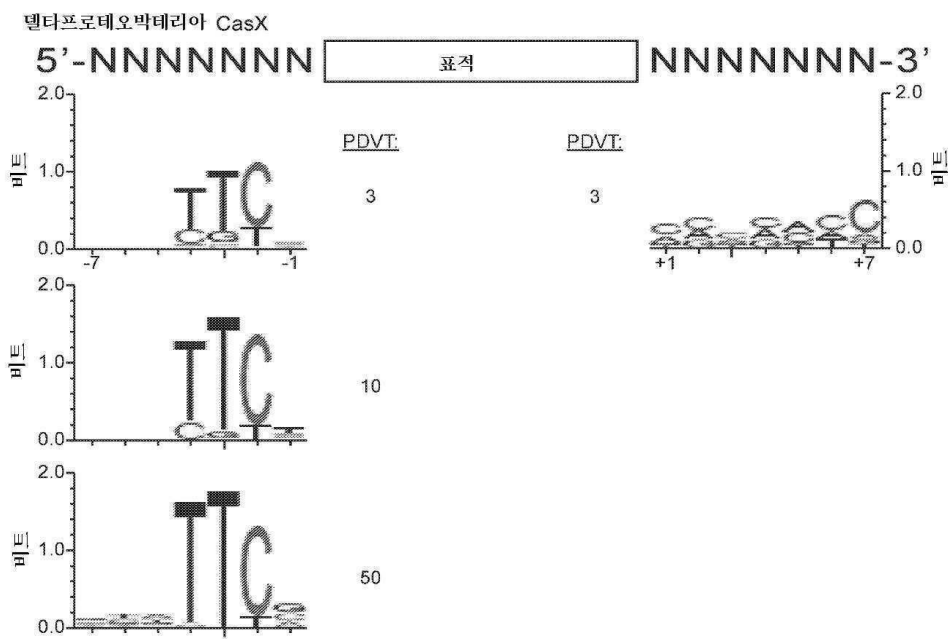
도면18



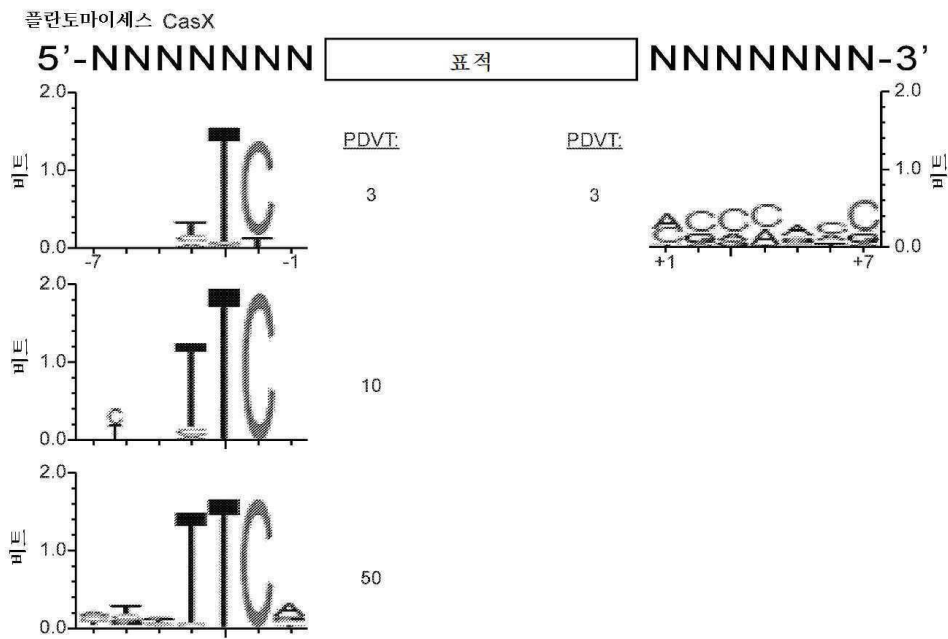
도면19ab



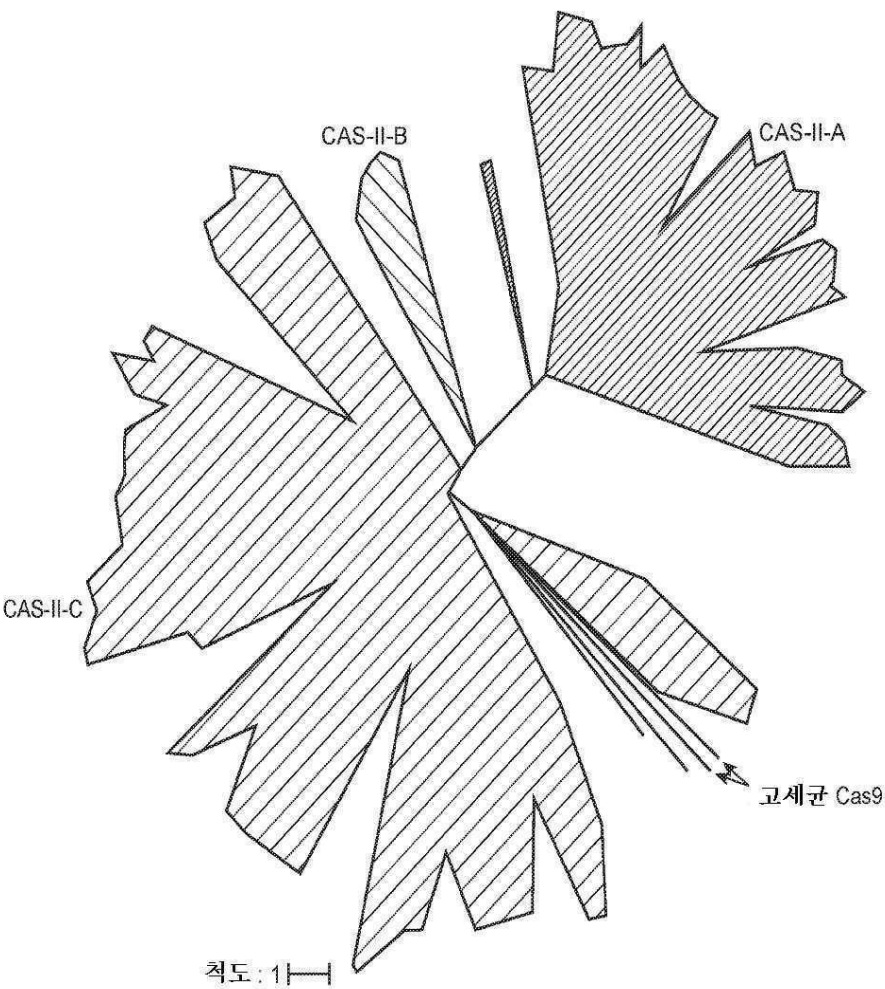
도면19c



도면19d



도면20



도면21

ARMAN-1로부터 Cas9에 대해 분석한 시험관내 절단 조건

단백질 정제	완충제	염 (mM)	금속	가이드	표적	온도
AR1-Cas9 #1	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	crRNA cr:69 cr:69 cr:69	dsDNA ssDNA DNA 버블 ssRNA dsDNA	37
AR1-Cas9 #1	트리스 ph 7.5	100-500	Mg ²⁺	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #1	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	cr:69 cr:104 cr:179	dsDNA	30-48
AR1-Cas9 #1	MOPS: pH 6 pH 6.5 pH 7.0 pH 7.5	300	Mg ²⁺	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #1	시트레이트 pH 5 pH 5.5 pH 6	300	Mg ²⁺	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #1	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	cr:69 cr:104 cr:179	플라스미드	37-50
AR1-Cas9 #2	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #3	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #4	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #5	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #6	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	cr:69 cr:104 cr:179	ssDNA dsDNA	37
AR4-Cas9 #1	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	sgRNA- 122	dsDNA	37
AR4-Cas9 #2	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	sgRNA- 122	dsDNA	37
AR4-Cas9 #3	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	sgRNA- 122	dsDNA	37
AR4-Cas9 #4	트리스 ph 7.5	300	Mg ²⁺ Mn ²⁺ Zn ²⁺	sgRNA- 122	dsDNA	37

서열 목록

SEQUENCE LISTING

<110> THE REGENTS OF THE UNIVERSITY OF CALIFORNIA

<120> RNA-GUIDED NUCLEIC ACID MODIFYING ENZYMES AND METHODS OF USE
THEREOF

<130> WO/2018/064352

<140> PCT/US2017/054047

<141> 2017-09-28

<150> US 62/402,849

<151> 2016-09-30

<160> 134

<170> PatentIn version 3.5

<210> 1

<211> 1125

<212> PRT

<213> Unknown

<220><223> Synthetic sequence

<400> 1

Met Arg Lys Lys Leu Phe Lys Gly Tyr Ile Leu His Asn Lys Arg Leu

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

Ala Gly Val Ile Trp Gln Ser Lys Ser Thr Ser Leu Ile Asp Leu Ile

85 90 95

Ser Lys Leu Glu Gly Ser Lys Ser Pro Ser Glu Lys Ile Phe Glu Gln

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

Glu Glu Val Ile Ala Cys Val Asp Lys Trp Ser Lys Asp Leu Ile Val

165	170	175	
Glu Gly Lys Ser Ile Leu Val Ser Lys Gln Phe Leu Tyr Trp Glu Glu			
180	185	190	
Glu Phe Gly Ile Lys Ile Phe Pro His Phe Lys Asp Asn His Asp Leu			
195	200	205	
Pro Lys Leu Thr Phe Phe Val Glu Pro Ser Leu Glu Phe Ser Pro His			
210	215	220	
Leu Pro Leu Ala Asn Cys Leu Glu Arg Leu Lys Lys Phe Asp Ile Ser			
225	230	235	240
Arg Glu Ser Leu Leu Gly Leu Asp Asn Asn Phe Ser Ala Phe Ser Asn			
245	250	255	
Tyr Phe Asn Glu Leu Phe Asn Leu Leu Ser Arg Gly Glu Ile Lys Lys			
260	265	270	
Ile Val Thr Ala Val Leu Ala Val Ser Lys Ser Trp Glu Asn Glu Pro			
275	280	285	
Glu Leu Glu Lys Arg Leu His Phe Leu Ser Glu Lys Ala Lys Leu Leu			
290	295	300	
Gly Tyr Pro Lys Leu Thr Ser Ser Trp Ala Asp Tyr Arg Met Ile Ile			
305	310	315	320
Gly Gly Lys Ile Lys Ser Trp His Ser Asn Tyr Thr Glu Gln Leu Ile			
325	330	335	
Lys Val Arg Glu Asp Leu Lys Lys His Gln Ile Ala Leu Asp Lys Leu			
340	345	350	
Gln Glu Asp Leu Lys Lys Val Val Asp Ser Ser Leu Arg Glu Gln Ile			
355	360	365	
Glu Ala Gln Arg Glu Ala Leu Leu Pro Leu Leu Asp Thr Met Leu Lys			
370	375	380	
Glu Lys Asp Phe Ser Asp Asp Leu Glu Leu Tyr Arg Phe Ile Leu Ser			
385	390	395	400
Asp Phe Lys Ser Leu Leu Asn Gly Ser Tyr Gln Arg Tyr Ile Gln Thr			
405	410	415	

Glu Glu Glu Arg Lys Glu Asp Arg Asp Val Thr Lys Lys Tyr Lys Asp
 420 425 430
 Leu Tyr Ser Asn Leu Arg Asn Ile Pro Arg Phe Phe Gly Glu Ser Lys
 435 440 445
 Lys Glu Gln Phe Asn Lys Phe Ile Asn Lys Ser Leu Pro Thr Ile Asp
 450 455 460

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

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

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

 Asp Lys Phe Val Val Arg Tyr Val Thr Gln Met Ile Gly Ser Asn Gln

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

Tyr Glu Ile Ser Ile Ser Asn Phe Glu Thr Gly Gly Asn Arg Met Ala

915 920 925

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

930 935 940

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

945 950 955 960

Met Gly Asn His Ile Ser Ser Tyr Ala Thr Ser Tyr Thr Cys Cys Asn

965 970 975

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

980 985 990

Glu Lys Gly Gly Asp Glu Phe Ile Phe Asn Val Gly Asp Glu Lys Lys

995 1000 1005

Val Arg Gly Phe Leu Gln Lys Ser Leu Leu Gly Lys Thr Ile Lys

1010 1015 1020

Gly Lys Glu Val Leu Lys Ser Ile Lys Glu Tyr Ala Arg Pro Pro

1025 1030 1035

Ile Arg Glu Val Leu Leu Glu Gly Glu Asp Val Glu Gln Leu Leu

1040 1045 1050

Lys Arg Arg Gly Asn Ser Tyr Ile Tyr Arg Cys Pro Phe Cys Gly

1055 1060 1065

Tyr Lys Thr Asp Ala Asp Ile Gln Ala Ala Leu Asn Ile Ala Cys

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

Lys Asn Gly Ala Val Leu Arg Ser Ala Lys Phe Leu

1115 1120 1125

<210> 2

<211> 1226

<212> PRT

<213> Unknown

<220><223> Synthetic sequence

<400> 2

Met Gln Lys Val Arg Lys Thr Leu Ser Glu Val His Lys Asn Pro Tyr

1 5 10 15

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

20 25 30

Arg Leu Ser Tyr Thr Gly Lys Glu Gly Met Arg Ser Phe Lys Ile Pro

35 40 45

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

50 55 60

Arg Asn Asp Tyr Ile Ser Gln Val Gly Leu Leu Asn Leu Ser Asp Trp

65 70 75 80

Tyr Glu His Tyr Gln Glu Lys Gln Glu His Tyr Ser Leu Ala Asp Phe

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

Leu Val Asp Ile Lys Ala Leu Tyr Asp Phe Leu Thr Ser Asp Ala Arg

145 150 155 160

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

165 170 175

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

180 185 190

Val Lys Lys Phe Ile Gly Leu Asp Asn Lys Asp Lys Leu Asn Phe Thr

195 200 205

Lys Lys Phe Ile Gly Phe Asp Pro Asn Pro Gln Ile Lys Tyr Asp His

210 215 220

Thr Phe Phe Phe His Gln Asp Ile Asn Phe Asp Leu Glu Arg Ile Thr

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

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

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

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

 465 470 475 480

Gln Ile Phe Lys Asp Ile Lys Leu Ile Pro Asn Phe Leu Gly Glu Thr
 485 490 495
 Lys Gln Met Val Tyr Lys Lys Phe Ile Arg Ser Ala Asp Ile Leu Tyr
 500 505 510
 Glu Gly Ile Asn Phe Ile Asp Gln Ile Asp Lys Gln Ile Thr Gln Asn
 515 520 525
 Leu Leu Pro Cys Phe Lys Asn Asp Lys Glu Arg Ile Glu Phe Thr Glu
 530 535 540
 Lys Gln Phe Glu Thr Leu Arg Arg Lys Tyr Tyr Leu Met Asn Ser Ser
 545 550 555 560
 Arg Phe His His Val Ile Glu Gly Ile Ile Asn Asn Arg Lys Leu Ile
 565 570 575
 Glu Met Lys Lys Arg Glu Asn Ser Glu Leu Lys Thr Phe Ser Asp Ser
 580 585 590
 Lys Phe Val Leu Ser Lys Leu Phe Leu Lys Lys Gly Lys Lys Tyr Glu
 595 600 605
 Asn Glu Val Tyr Tyr Thr Phe Tyr Ile Asn Pro Lys Ala Arg Asp Gln
 610 615 620
 Arg Arg Ile Lys Ile Val Leu Asp Ile Asn Gly Asn Asn Ser Val Gly
 625 630 635 640
 Ile Leu Gln Asp Leu Val Gln Lys Leu Lys Pro Lys Trp Asp Asp Ile
 645 650 655
 Ile Lys Lys Asn Asp Met Gly Glu Leu Ile Asp Ala Ile Glu Ile Glu
 660 665 670
 Lys Val Arg Leu Gly Ile Leu Ile Ala Leu Tyr Cys Glu His Lys Phe
 675 680 685
 Lys Ile Lys Lys Glu Leu Leu Ser Leu Asp Leu Phe Ala Ser Ala Tyr
 690 695 700
 Gln Tyr Leu Glu Leu Glu Asp Asp Pro Glu Glu Leu Ser Gly Thr Asn
 705 710 715 720
 Leu Gly Arg Phe Leu Gln Ser Leu Val Cys Ser Glu Ile Lys Gly Ala

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

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

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

1205 1210 1215
Thr Ser Leu Asp Ile Asn Leu Leu
1220 1225

<210> 3
<211> 1160
<212> PRT
<213> Unknown
<220><223> Synthetic sequence
<400> 3

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

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

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

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

180 185 190

Ile Asn Arg Ile Ala Asp Val Phe Ala Lys Asn Lys Ser Asp Lys Tyr

195 200 205

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

210 215 220

Asn Asp Val Gln Val Ala Phe Phe Cys His Pro Asp Phe Glu Glu Ile

225 230 235 240

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

245 250 255

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

260 265 270

Asn Arg Met Ala Leu Tyr Ser Leu Asn Gln Gly Tyr Ile Pro Arg Val

275 280 285

Leu Asn Asp Leu Phe Leu Phe Val Lys Asp Asn Glu Asp Asp Phe Ser

290 295 300

Gln Phe Leu Ser Asp Leu Glu Asn Phe Phe Ser Phe Ser Asn Glu Gln

305 310 315 320

Ile Lys Ile Ile Lys Glu Arg Leu Lys Lys Leu Lys Lys Tyr Ala Glu

325 330 335

Pro Ile Pro Gly Lys Pro Gln Leu Ala Asp Lys Trp Asp Asp Tyr Ala

340 345 350

Ser Asp Phe Gly Gly Lys Leu Glu Ser Trp Tyr Ser Asn Arg Ile Glu

355 360 365

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

370 375 380

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

385 390 395 400

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

405 410 415

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

420 425 430
Lys Asp Gly Gln Lys Lys Val Phe Tyr Val Ala Lys Met Ala Asp Arg

435 440 445
Glu Phe Ile Glu Lys Leu Asp Leu Trp Met Ala Asp Leu Arg Ser Gln

450 455 460
Leu Asn Glu Tyr Asn Gln Asp Asn Lys Val Ser Phe Lys Lys Lys Gly

465 470 475 480
Lys Lys Ile Glu Glu Leu Gly Val Leu Asp Phe Ala Leu Asn Lys Ala

485 490 495
Lys Lys Asn Lys Ser Thr Lys Asn Glu Asn Gly Trp Gln Gln Lys Leu

500 505 510
Ser Glu Ser Ile Gln Ser Ala Pro Leu Phe Phe Gly Glu Gly Asn Arg

515 520 525
Val Arg Asn Glu Glu Val Tyr Asn Leu Lys Asp Leu Leu Phe Ser Glu

530 535 540
Ile Lys Asn Val Glu Asn Ile Leu Met Ser Ser Glu Ala Glu Asp Leu

545 550 555 560
Lys Asn Ile Lys Ile Glu Tyr Lys Glu Asp Gly Ala Lys Lys Gly Asn

565 570 575
Tyr Val Leu Asn Val Leu Ala Arg Phe Tyr Ala Arg Phe Asn Glu Asp

580 585 590
Gly Tyr Gly Gly Trp Asn Lys Val Lys Thr Val Leu Glu Asn Ile Ala

595 600 605
Arg Glu Ala Gly Thr Asp Phe Ser Lys Tyr Gly Asn Asn Asn Asn Arg

610 615 620
Asn Ala Gly Arg Phe Tyr Leu Asn Gly Arg Glu Arg Gln Val Phe Thr

625 630 635 640
Leu Ile Lys Phe Glu Lys Ser Ile Thr Val Glu Lys Ile Leu Glu Leu

645 650 655
Val Lys Leu Pro Ser Leu Leu Asp Glu Ala Tyr Arg Asp Leu Val Asn

660 665 670

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

915	920	925	
Leu Lys Lys Asn Val Lys Ser Trp Arg Gln Asn Gln Ile Arg Gln Thr			
930	935	940	
Phe Thr Ser Pro Asp Thr Lys Ile Ala Arg Leu Arg Glu Ser Leu Ile			
945	950	955	960
Gly Ser Tyr Lys Asn Gln Leu Glu Ser Leu Met Val Ala Lys Lys Ala			
965	970	975	
Asn Leu Ser Phe Glu Tyr Glu Val Ser Gly Phe Glu Val Gly Gly Lys			
980	985	990	
Arg Val Ala Lys Ile Tyr Asp Ser Ile Lys Arg Gly Ser Val Arg Lys			
995	1000	1005	
Lys Asp Asn Asn Ser Gln Asn Asp Gln Ser Trp Gly Lys Lys Gly			
1010	1015	1020	
Ile Asn Glu Trp Ser Phe Glu Thr Thr Ala Ala Gly Thr Ser Gln			
1025	1030	1035	
Phe Cys Thr His Cys Lys Arg Trp Ser Ser Leu Ala Ile Val Asp			
1040	1045	1050	
Ile Glu Glu Tyr Glu Leu Lys Asp Tyr Asn Asp Asn Leu Phe Lys			
1055	1060	1065	
Val Lys Ile Asn Asp Gly Glu Val Arg Leu Leu Gly Lys Lys Gly			
1070	1075	1080	
Trp Arg Ser Gly Glu Lys Ile Lys Gly Lys Glu Leu Phe Gly Pro			
1085	1090	1095	
Val Lys Asp Ala Met Arg Pro Asn Val Asp Gly Leu Gly Met Lys			
1100	1105	1110	
Ile Val Lys Arg Lys Tyr Leu Lys Leu Asp Leu Arg Asp Trp Val			
1115	1120	1125	
Ser Arg Tyr Gly Asn Met Ala Ile Phe Ile Cys Pro Tyr Val Asp			
1130	1135	1140	
Cys His His Ile Ser His Ala Asp Lys Gln Ala Ala Phe Asn Ile			
1145	1150	1155	

Ala Val

1160

<210> 4

<211> 1210

<212> PRT

<213> Unknown

<220><223> Synthetic sequence

<400> 4

Met Ser Lys Arg His Pro Arg Ile Ser Gly Val Lys Gly Tyr Arg Leu

1 5 10 15

His Ala Gln Arg Leu Glu Tyr Thr Gly Lys Ser Gly Ala Met Arg Thr

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

Lys Val Tyr Ser Val Leu Asp Phe Trp Tyr Asp Cys Val Gln Tyr Gly

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

Ile Ile Asp Cys Phe Lys Ala Glu Tyr Arg Glu Arg His Lys Asp Gln

165 170 175

Cys Asn Lys Leu Ala Asp Asp Ile Lys Asn Ala Lys Lys Asp Ala Gly

180 185 190

Ala Ser Leu Gly Glu Arg Gln Lys Lys Leu Phe Arg Asp Phe Phe Gly
195 200 205

Ile Ser Glu Gln Ser Glu Asn Asp Lys Pro Ser Phe Thr Asn Pro Leu
210 215 220

Asn Leu Thr Cys Cys Leu Leu Pro Phe Asp Thr Val Asn Asn Asn Arg
225 230 235 240

Asn Arg Gly Glu Val Leu Phe Asn Lys Leu Lys Glu Tyr Ala Gln Lys
245 250 255

Leu Asp Lys Asn Glu Gly Ser Leu Glu Met Trp Glu Tyr Ile Gly Ile
260 265 270

Gly Asn Ser Gly Thr Ala Phe Ser Asn Phe Leu Gly Glu Gly Phe Leu
275 280 285

Gly Arg Leu Arg Glu Asn Lys Ile Thr Glu Leu Lys Lys Ala Met Met
290 295 300

Asp Ile Thr Asp Ala Trp Arg Gly Gln Glu Gln Glu Glu Leu Glu
305 310 315 320

Lys Arg Leu Arg Ile Leu Ala Ala Leu Thr Ile Lys Leu Arg Glu Pro
325 330 335

Lys Phe Asp Asn His Trp Gly Gly Tyr Arg Ser Asp Ile Asn Gly Lys
340 345 350

Leu Ser Ser Trp Leu Gln Asn Tyr Ile Asn Gln Thr Val Lys Ile Lys
355 360 365

Glu Asp Leu Lys Gly His Lys Lys Asp Leu Lys Lys Ala Lys Glu Met
370 375 380

Ile Asn Arg Phe Gly Glu Ser Asp Thr Lys Glu Glu Ala Val Val Ser
385 390 395 400

Ser Leu Leu Glu Ser Ile Glu Lys Ile Val Pro Asp Asp Ser Ala Asp
405 410 415

Asp Glu Lys Pro Asp Ile Pro Ala Ile Ala Ile Tyr Arg Arg Phe Leu
420 425 430

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

435 440 445
 Gln Glu Ala Leu Ile Lys Glu Arg Leu Glu Ala Glu Lys Lys Lys Lys
 450 455 460
 Pro Lys Lys Arg Lys Lys Lys Ser Asp Ala Glu Asp Glu Lys Glu Thr

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

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

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

 660 665 670
 Leu Asp Ile Ser Ala Leu Asp Phe Val Glu Asn Gly Thr Val Lys Asp
 675 680 685

Phe Met Lys Thr Arg Asp Gly Asn Leu Val Leu Glu Gly Arg Phe Leu
 690 695 700
 Glu Met Phe Ser Gln Ser Ile Val Phe Ser Glu Leu Arg Gly Leu Ala
 705 710 715 720
 Gly Leu Met Ser Arg Lys Glu Phe Ile Thr Arg Ser Ala Ile Gln Thr
 725 730 735
 Met Asn Gly Lys Gln Ala Glu Leu Leu Tyr Ile Pro His Glu Phe Gln
 740 745 750
 Ser Ala Lys Ile Thr Thr Pro Lys Glu Met Ser Arg Ala Phe Leu Asp
 755 760 765
 Leu Ala Pro Ala Glu Phe Ala Thr Ser Leu Glu Pro Glu Ser Leu Ser
 770 775 780
 Glu Lys Ser Leu Leu Lys Leu Lys Gln Met Arg Tyr Tyr Pro His Tyr
 785 790 795 800
 Phe Gly Tyr Glu Leu Thr Arg Thr Gly Gln Gly Ile Asp Gly Gly Val
 805 810 815
 Ala Glu Asn Ala Leu Arg Leu Glu Lys Ser Pro Val Lys Lys Arg Glu
 820 825 830
 Ile Lys Cys Lys Gln Tyr Lys Thr Leu Gly Arg Gly Gln Asn Lys Ile
 835 840 845
 Val Leu Tyr Val Arg Ser Ser Tyr Tyr Gln Thr Gln Phe Leu Glu Trp
 850 855 860
 Phe Leu His Arg Pro Lys Asn Val Gln Thr Asp Val Ala Val Ser Gly
 865 870 875 880
 Ser Phe Leu Ile Asp Glu Lys Lys Val Lys Thr Arg Trp Asn Tyr Asp
 885 890 895
 Ala Leu Thr Val Ala Leu Glu Pro Val Ser Gly Ser Glu Arg Val Phe
 900 905 910
 Val Ser Gln Pro Phe Thr Ile Phe Pro Glu Lys Ser Ala Glu Glu Glu
 915 920 925
 Gly Gln Arg Tyr Leu Gly Ile Asp Ile Gly Glu Tyr Gly Ile Ala Tyr

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

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

1040 1045 1050
 Asp Lys Asn Leu Gln Thr Thr Val Trp Gly Lys Leu Ala Val Ala
 1055 1060 1065
 Ser Glu Ile Ser Ala Ser Tyr Thr Ser Gln Phe Cys Gly Ala Cys
 1070 1075 1080
 Lys Lys Leu Trp Arg Ala Glu Met Gln Val Asp Glu Thr Ile Thr
 1085 1090 1095
 Thr Gln Glu Leu Ile Gly Thr Val Arg Val Ile Lys Gly Gly Thr

1100 1105 1110
 Leu Ile Asp Ala Ile Lys Asp Phe Met Arg Pro Pro Ile Phe Asp
 1115 1120 1125
 Glu Asn Asp Thr Pro Phe Pro Lys Tyr Arg Asp Phe Cys Asp Lys
 1130 1135 1140
 His His Ile Ser Lys Lys Met Arg Gly Asn Ser Cys Leu Phe Ile
 1145 1150 1155
 Cys Pro Phe Cys Arg Ala Asn Ala Asp Ala Asp Ile Gln Ala Ser

1160 1165 1170

Gln Thr Ile Ala Leu Leu Arg Tyr Val Lys Glu Glu Lys Lys Val

1175 1180 1185

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

1190 1195 1200

Leu Gly Gln Met Lys Lys Ile

1205 1210

<210> 5

<211> 1287

<212> PRT

<213> Unknown

<220><223> Synthetic sequence

<400> 5

Met Lys Arg Ile Leu Asn Ser Leu Lys Val Ala Ala Leu Arg Leu Leu

1 5 10 15

Phe Arg Gly Lys Gly Ser Glu Leu Val Lys Thr Val Lys Tyr Pro Leu

20 25 30

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

35 40 45

His Asp Asn Leu His Leu Phe Gly Gln Lys Glu Ile Val Asp Leu Met

50 55 60

Glu Lys Asp Glu Gly Thr Gln Val Tyr Ser Val Val Asp Phe Trp Leu

65 70 75 80

Asp Thr Leu Arg Leu Gly Met Phe Phe Ser Pro Ser Ala Asn Ala Leu

85 90 95

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

100 105 110

Lys Val Leu Glu Gln Ser Pro Phe Phe Leu Ala Gly Arg Leu Lys Val

115 120 125

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

130 135 140

Arg Glu Asn Arg Val Glu Asn Tyr Ala Ala Asp Val Glu Thr Cys Phe

145 150 155 160

Ile Gly Gln Leu Ser Ser Asp Glu Lys Gln Ser Ile Gln Lys Leu Ala
165 170 175
Asn Asp Ile Trp Asp Ser Lys Asp His Glu Glu Gln Arg Met Leu Lys
180 185 190
Ala Asp Phe Phe Ala Ile Pro Leu Ile Lys Asp Pro Lys Ala Val Thr
195 200 205
Glu Glu Asp Pro Glu Asn Glu Thr Ala Gly Lys Gln Lys Pro Leu Glu
210 215 220
Leu Cys Val Cys Leu Val Pro Glu Leu Tyr Thr Arg Gly Phe Gly Ser
225 230 235 240
Ile Ala Asp Phe Leu Val Gln Arg Leu Thr Leu Leu Arg Asp Lys Met
245 250 255
Ser Thr Asp Thr Ala Glu Asp Cys Leu Glu Tyr Val Gly Ile Glu Glu
260 265 270
Glu Lys Gly Asn Gly Met Asn Ser Leu Leu Gly Thr Phe Leu Lys Asn
275 280 285
Leu Gln Gly Asp Gly Phe Glu Gln Ile Phe Gln Phe Met Leu Gly Ser
290 295 300
Tyr Val Gly Trp Gln Gly Lys Glu Asp Val Leu Arg Glu Arg Leu Asp
305 310 315 320
Leu Leu Ala Glu Lys Val Lys Arg Leu Pro Lys Pro Lys Phe Ala Gly
325 330 335
Glu Trp Ser Gly His Arg Met Phe Leu His Gly Gln Leu Lys Ser Trp
340 345 350
Ser Ser Asn Phe Phe Arg Leu Phe Asn Glu Thr Arg Glu Leu Leu Glu
355 360 365
Ser Ile Lys Ser Asp Ile Gln His Ala Thr Met Leu Ile Ser Tyr Val
370 375 380
Glu Glu Lys Gly Gly Tyr His Pro Gln Leu Leu Ser Gln Tyr Arg Lys
385 390 395 400
Leu Met Glu Gln Leu Pro Ala Leu Arg Thr Lys Val Leu Asp Pro Glu

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

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

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

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

 645 650 655

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

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

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

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

900	905	910	
Val Gly Glu Phe Arg Val Thr Ala Asp Gln Pro Phe Lys Leu Thr Pro			
915	920	925	
Ile Ile Pro Glu Lys Glu Glu Ser Phe Ile Gly Lys Thr Tyr Leu Gly			
930	935	940	
Leu Asp Ala Gly Glu Arg Ser Gly Val Gly Phe Ala Ile Val Thr Val			
945	950	955	960
Asp Gly Asp Gly Tyr Glu Val Gln Arg Leu Gly Val His Glu Asp Thr			
965	970	975	
Gln Leu Met Ala Leu Gln Gln Val Ala Ser Lys Ser Leu Lys Glu Pro			
980	985	990	
Val Phe Gln Pro Leu Arg Lys Gly Thr Phe Arg Gln Gln Glu Arg Ile			
995	1000	1005	
Arg Lys Ser Leu Arg Gly Cys Tyr Trp Asn Phe Tyr His Ala Leu			
1010	1015	1020	
Met Ile Lys Tyr Arg Ala Lys Val Val His Glu Glu Ser Val Gly			
1025	1030	1035	
Ser Ser Gly Leu Val Gly Gln Trp Leu Arg Ala Phe Gln Lys Asp			
1040	1045	1050	
Leu Lys Lys Ala Asp Val Leu Pro Lys Lys Gly Gly Lys Asn Gly			
1055	1060	1065	
Val Asp Lys Lys Lys Arg Glu Ser Ser Ala Gln Asp Thr Leu Trp			
1070	1075	1080	
Gly Gly Ala Phe Ser Lys Lys Glu Glu Gln Gln Ile Ala Phe Glu			
1085	1090	1095	
Val Gln Ala Ala Gly Ser Ser Gln Phe Cys Leu Lys Cys Gly Trp			
1100	1105	1110	
Trp Phe Gln Leu Gly Met Arg Glu Val Asn Arg Val Gln Glu Ser			
1115	1120	1125	
Gly Val Val Leu Asp Trp Asn Arg Ser Ile Val Thr Phe Leu Ile			
1130	1135	1140	

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

1145 1150 1155

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

1160 1165 1170

Arg Asp Phe Met Arg Pro Pro Met Phe Asp Arg Lys Gly Arg Pro

1175 1180 1185

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

1190 1195 1200

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

1205 1210 1215

Leu Phe Ile Cys Pro Arg Val Gly Cys Gly Asn Phe Asp His Ser

1220 1225 1230

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

1235 1240 1245

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

1250 1255 1260

Ala Glu Leu Met Ala Glu Trp Lys Leu Lys Lys Leu Glu Arg Ser

1265 1270 1275

Arg Val Glu Glu Gln Ser Ser Ala Gln

1280 1285

<210> 6

<211> 1192

<212> PRT

<213> Unknown

<220><223> Synthetic sequence

<400> 6

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

1 5 10 15

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

20 25 30

Pro Asp Cys Gly Asn His Thr Ser Ala Arg Lys Ile Gln Asn Lys Lys

35	40	45	
Lys Arg Asp Lys Lys Tyr Gly Ser Ala Ser Lys Ala Gln Ser Gln Arg			
50	55	60	
Ile Ala Val Ala Gly Ala Leu Tyr Pro Asp Lys Lys Val Gln Thr Ile			
65	70	75	80
Lys Thr Tyr Lys Tyr Pro Ala Asp Leu Asn Gly Glu Val His Asp Ser			
85	90	95	
Gly Val Ala Glu Lys Ile Ala Gln Ala Ile Gln Glu Asp Glu Ile Gly			
100	105	110	
Leu Leu Gly Pro Ser Ser Glu Tyr Ala Cys Trp Ile Ala Ser Gln Lys			
115	120	125	
Gln Ser Glu Pro Tyr Ser Val Val Asp Phe Trp Phe Asp Ala Val Cys			
130	135	140	
Ala Gly Gly Val Phe Ala Tyr Ser Gly Ala Arg Leu Leu Ser Thr Val			
145	150	155	160
Leu Gln Leu Ser Gly Glu Glu Ser Val Leu Arg Ala Ala Leu Ala Ser			
165	170	175	
Ser Pro Phe Val Asp Asp Ile Asn Leu Ala Gln Ala Glu Lys Phe Leu			
180	185	190	
Ala Val Ser Arg Arg Thr Gly Gln Asp Lys Leu Gly Lys Arg Ile Gly			
195	200	205	
Glu Cys Phe Ala Glu Gly Arg Leu Glu Ala Leu Gly Ile Lys Asp Arg			
210	215	220	
Met Arg Glu Phe Val Gln Ala Ile Asp Val Ala Gln Thr Ala Gly Gln			
225	230	235	240
Arg Phe Ala Ala Lys Leu Lys Ile Phe Gly Ile Ser Gln Met Pro Glu			
245	250	255	
Ala Lys Gln Trp Asn Asn Asp Ser Gly Leu Thr Val Cys Ile Leu Pro			
260	265	270	
Asp Tyr Tyr Val Pro Glu Glu Asn Arg Ala Asp Gln Leu Val Val Leu			
275	280	285	

Leu Arg Arg Leu Arg Glu Ile Ala Tyr Cys Met Gly Ile Glu Asp Glu

290

295

300

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

305

310

315

320

Ser Asn Gly Asn Pro Lys Arg Gly Phe Leu Gly Arg Leu Leu Asn Asn

325

330

335

Asp Ile Ile Ala Leu Ala Asn Asn Met Ser Ala Met Thr Pro Tyr Trp

340

345

350

Glu Gly Arg Lys Gly Glu Leu Ile Glu Arg Leu Ala Trp Leu Lys His

355

360

365

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

370

375

380

Ala Asp His Arg Ser Arg Ile Phe Ser Arg Ile Ala Gly Trp Leu Ser

385

390

395

400

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

405

410

415

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

420

425

430

Ser Ala Pro Ser Pro Asp Phe Ile Ala Ser Ile Ser Ala Leu Asp Arg

435

440

445

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

450

455

460

Ala Leu Tyr Ala Phe His Leu Asn Ala Pro Ala Val Arg Ser Ile Ala

465

470

475

480

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

485

490

495

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

500

505

510

Ser Asp Thr Gly Lys Lys Lys Lys Lys Gly Ala Asn Ser Asn Gly Ala

515

520

525

Pro Ser Glu Glu Glu Tyr Thr Glu Thr Glu Ser Ile Gln Gln Pro Glu

530 535 540
 Asp Ala Glu Gln Glu Val Asn Gly Gln Glu Gly Asn Gly Ala Ser Lys

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

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

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

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

Phe Asp Val Gln Leu Glu Ser Ser Cys Gln Asp Asn Leu Gln His Leu
 785 790 795 800
 Leu Val Tyr Arg Ala Ser Arg Asp Leu Ala Ala Cys Lys Arg Ala Thr

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

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

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

 995 1000 1005
 Ser Ala Gly Phe Lys Ile Leu Glu Arg Gly Thr Ile Ala Val Asn
 1010 1015 1020
 Gly Phe Ser Gln Lys Arg Gly Glu Arg Gln Glu Glu Ala His Arg

1025 1030 1035
 Glu Lys Gln Arg Arg Gly Ile Ser Asp Ile Gly Arg Lys Lys Pro
 1040 1045 1050
 Val Gln Ala Glu Val Asp Ala Ala Asn Glu Leu His Arg Lys Tyr

1055 1060 1065
 Thr Asp Val Ala Thr Arg Leu Gly Cys Arg Ile Val Val Gln Trp
 1070 1075 1080
 Ala Pro Gln Pro Lys Pro Gly Thr Ala Pro Thr Ala Gln Thr Val
 1085 1090 1095
 Tyr Ala Arg Ala Val Arg Thr Glu Ala Pro Arg Ser Gly Asn Gln
 1100 1105 1110
 Glu Asp His Ala Arg Met Lys Ser Ser Trp Gly Tyr Thr Trp Gly

1115 1120 1125
 Thr Tyr Trp Glu Lys Arg Lys Pro Glu Asp Ile Leu Gly Ile Ser
 1130 1135 1140
 Thr Gln Val Tyr Trp Thr Gly Gly Ile Gly Glu Ser Cys Pro Ala
 1145 1150 1155
 Val Ala Val Ala Leu Leu Gly His Ile Arg Ala Thr Ser Thr Gln
 1160 1165 1170
 Thr Glu Trp Glu Lys Glu Glu Val Val Phe Gly Arg Leu Lys Lys

1175 1180 1185
 Phe Phe Pro Ser
 1190
 <210> 7
 <211> 1192
 <212> PRT
 <213> Unknown
 <220><223> Synthetic sequence
 <400> 7
 Met Ala Glu Ser Lys Gln Met Gln Cys Arg Lys Cys Gly Ala Ser Met
 1 5 10 15
 Lys Tyr Glu Val Ile Gly Leu Gly Lys Lys Ser Cys Arg Tyr Met Cys

20 25 30
 Pro Asp Cys Gly Asn His Thr Ser Ala Arg Lys Ile Gln Asn Lys Lys

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

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

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

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

Asp Tyr Tyr Val Pro Glu Glu Asn Arg Ala Asp Gln Leu Val Val Leu
 275 280 285
 Leu Arg Arg Leu Arg Glu Ile Ala Tyr Cys Met Gly Ile Glu Asp Glu
 290 295 300
 Ala Gly Phe Glu His Leu Gly Ile Asp Pro Gly Ala Leu Ser Asn Phe
 305 310 315 320
 Ser Asn Gly Asn Pro Lys Arg Gly Phe Leu Gly Arg Leu Leu Asn Asn
 325 330 335
 Asp Ile Ile Ala Leu Ala Asn Asn Met Ser Ala Met Thr Pro Tyr Trp
 340 345 350
 Glu Gly Arg Lys Gly Glu Leu Ile Glu Arg Leu Ala Trp Leu Lys His
 355 360 365
 Arg Ala Glu Gly Leu Tyr Leu Lys Glu Pro His Phe Gly Asn Ser Trp
 370 375 380
 Ala Asp His Arg Ser Arg Ile Phe Ser Arg Ile Ala Gly Trp Leu Ser
 385 390 395 400
 Gly Cys Ala Gly Lys Leu Lys Ile Ala Lys Asp Gln Ile Ser Gly Val
 405 410 415
 Arg Thr Asp Leu Phe Leu Leu Lys Arg Leu Leu Asp Ala Val Pro Gln
 420 425 430
 Ser Ala Pro Ser Pro Asp Phe Ile Ala Ser Ile Ser Ala Leu Asp Arg
 435 440 445
 Phe Leu Glu Ala Ala Glu Ser Ser Gln Asp Pro Ala Glu Gln Val Arg
 450 455 460
 Ala Leu Tyr Ala Phe His Leu Asn Ala Pro Ala Val Arg Ser Ile Ala
 465 470 475 480
 Asn Lys Ala Val Gln Arg Ser Asp Ser Gln Glu Trp Leu Ile Lys Glu
 485 490 495
 Leu Asp Ala Val Asp His Leu Glu Phe Asn Lys Ala Phe Pro Phe Phe
 500 505 510
 Ser Asp Thr Gly Lys Lys Lys Lys Lys Gly Ala Asn Ser Asn Gly Ala

515 520 525
 Pro Ser Glu Glu Glu Tyr Thr Glu Thr Glu Ser Ile Gln Gln Pro Glu
 530 535 540
 Asp Ala Glu Gln Glu Val Asn Gly Gln Glu Gly Asn Gly Ala Ser Lys

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

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

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

 740 745 750
 Tyr Gly Thr Gln Leu Glu Glu Phe Leu Asn Ala Thr Val Leu Ser Gln
 755 760 765

Met Arg Gly Leu Ala Ile Arg Leu Ser Ser Gln Glu Leu Lys Asp Gly
770 775 780

Phe Asp Val Gln Leu Glu Ser Ser Cys Gln Asp Asn Leu Gln His Leu
785 790 795 800

Leu Val Tyr Arg Ala Ser Arg Asp Leu Ala Ala Cys Lys Arg Ala Thr
805 810 815

Cys Pro Ala Glu Leu Asp Pro Lys Ile Leu Val Leu Pro Ala Gly Ala
820 825 830

Phe Ile Ala Ser Val Met Lys Met Ile Glu Arg Gly Asp Glu Pro Leu
835 840 845

Ala Gly Ala Tyr Leu Arg His Arg Pro His Ser Phe Gly Trp Gln Ile
850 855 860

Arg Val Arg Gly Val Ala Glu Val Gly Met Asp Gln Gly Thr Ala Leu
865 870 875 880

Ala Phe Gln Lys Pro Thr Glu Ser Glu Pro Phe Lys Ile Lys Pro Phe
885 890 895

Ser Ala Gln Tyr Gly Pro Val Leu Trp Leu Asn Ser Ser Ser Tyr Ser
900 905 910

Gln Ser Gln Tyr Leu Asp Gly Phe Leu Ser Gln Pro Lys Asn Trp Ser
915 920 925

Met Arg Val Leu Pro Gln Ala Gly Ser Val Arg Val Glu Gln Arg Val
930 935 940

Ala Leu Ile Trp Asn Leu Gln Ala Gly Lys Met Arg Leu Glu Arg Ser
945 950 955 960

Gly Ala Arg Ala Phe Phe Met Pro Val Pro Phe Ser Phe Arg Pro Ser
965 970 975

Gly Ser Gly Asp Glu Ala Val Leu Ala Pro Asn Arg Tyr Leu Gly Leu
980 985 990

Phe Pro His Ser Gly Gly Ile Glu Tyr Ala Val Val Asp Val Leu Asp
995 1000 1005

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

1010	1015	1020
Gly Phe Ser Gln Lys Arg	Gly Glu Arg Gln Glu Glu	Ala His Arg
1025	1030	1035
Glu Lys Gln Arg Arg Gly Ile	Ser Asp Ile Gly Arg	Lys Lys Pro
1040	1045	1050
Val Gln Ala Glu Val Asp Ala	Ala Asn Glu Leu His	Arg Lys Tyr

1055	1060	1065
Thr Asp Val Ala Thr Arg Leu	Gly Cys Arg Ile Val	Val Gln Trp
1070	1075	1080
Ala Pro Gln Pro Lys Pro Gly	Thr Ala Pro Thr Ala	Gln Thr Val
1085	1090	1095
Tyr Ala Arg Ala Val Arg Thr	Glu Ala Pro Arg Ser	Gly Asn Gln
1100	1105	1110
Glu Asp His Ala Arg Met Lys	Ser Ser Trp Gly Tyr	Thr Trp Ser

1115	1120	1125
Thr Tyr Trp Glu Lys Arg Lys	Pro Glu Asp Ile Leu	Gly Ile Ser
1130	1135	1140
Thr Gln Val Tyr Trp Thr Gly	Gly Ile Gly Glu Ser	Cys Pro Ala
1145	1150	1155
Val Ala Val Ala Leu Leu Gly	His Ile Arg Ala Thr	Ser Thr Gln
1160	1165	1170
Thr Glu Trp Glu Lys Glu Glu	Val Val Phe Gly Arg	Leu Lys Lys

1175	1180	1185
Phe Phe Pro Ser		

1190

<210> 8

<211> 1193

<212> PRT

<213> Unknown

<220><223> Synthetic sequence

<400> 8

Met Lys Arg Ile Ala Lys Phe Arg His Asp Lys Pro Val Lys Arg Glu

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

 35 40 45
 Lys Arg Leu Ile Asp Asp Ala Ala His Asp Tyr Arg Trp Leu Val Gly
 50 55 60
 Pro Ile Asn Tyr Ser Asp Trp Cys Arg Asp Pro Asn Gln Tyr Ser Ile
 65 70 75 80
 Leu Glu Phe Trp Ile Asp Phe Leu Cys Val Gly Gly Val Phe Gln Ser
 85 90 95
 Ser His Ser Asn Ile Cys Arg Leu Ala Ile Gln Leu Ser Gly Gly Ser

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

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

 225 230 235 240
 Gln Thr Val Leu Gln Ser Asn Asp Gln Asp Asp Pro Glu Leu Thr Gly
 245 250 255

Pro Leu Pro Ser His Lys Val Gly Lys Ala Ser Tyr Leu Phe Ile Pro
260 265 270

Glu Phe Ile Arg Glu Val Asn Leu Asp Lys Ile Ser Lys Leu Asp Leu
275 280 285

Ser Ala Lys Ser Lys Leu Ala Val Glu Gln Val Lys Lys Leu Ser Glu
290 295 300

Leu Thr Ser Asp Phe Lys Gln Ile Glu Asn Gln Ser Glu Ala Tyr Phe
305 310 315 320

Gly Leu Ser Thr Ser Phe Asn Glu Leu Ser Asn Phe Leu Gly Ile Leu
325 330 335

Ile Arg Thr Leu Arg Asn Ala Pro Glu Ala Ile Leu Lys Asp Gln Ile
340 345 350

Ala Leu Cys Ala Pro Leu Asp Lys Asp Ile Leu Lys Ile Thr Leu Asp
355 360 365

Trp Leu Cys Asp Arg Ala Gln Ala Leu Pro Glu Asn Pro Arg Phe Glu
370 375 380

Thr Asn Trp Ala Glu Tyr Arg Ser Tyr Leu Gly Gly Lys Ile Lys Ser
385 390 395 400

Trp Phe Ser Asn Tyr Glu Asn Phe Phe Glu Ile Pro Gln Ala Ala Ser
405 410 415

Ser Gln Gln Asn Asn Asn Arg Glu Lys Lys Leu Gly Asn Arg Ser Ala
420 425 430

Ile Arg Ala Leu Asn Leu Lys Lys Glu Ala Phe Glu Lys Ala Arg Glu
435 440 445

Thr Phe Lys Gly Asp Lys Gly Thr Leu Glu Lys Ile Asp Leu Ala Tyr
450 455 460

Arg Leu Leu Gly Ser Ile Ser Pro Glu Val Leu Gln Cys Asp Glu Gly
465 470 475 480

Leu Lys Leu Tyr Gln Gln Phe Asn Asp Glu Leu Leu Val Leu Asn Glu
485 490 495

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

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

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

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

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

 740 745 750

Ser Lys Asn Lys Lys Thr Arg Arg Pro Leu Met Pro Glu Pro Phe Ile
 755 760 765
 Asn Glu Asp Lys Glu Val Phe Leu Ala Phe Pro Lys Phe Glu Thr Lys
 770 775 780
 Asn Lys Val Lys Gly Thr Arg Val Val Tyr Asn Ser Pro Asp Glu Val
 785 790 795 800
 Asn Trp Leu Leu Ser Pro Ile Arg Ser Ser Lys Gly Gln Leu Ser Phe
 805 810 815
 Met Phe Arg Cys Leu Ser Glu Asp Ala Lys Ile Met Thr Thr Ser Gly
 820 825 830
 Gly Cys Ser Tyr Ile Val Glu Phe Lys Lys Leu Leu Glu Ala Gln Glu
 835 840 845
 Glu Val Leu Ser Ile His Asp Cys Asp Ile Ile Pro Arg Ala Phe Val
 850 855 860
 Ser Ile Pro Phe Thr Leu Glu Arg Glu Ser Glu Glu Thr Lys Pro Asp
 865 870 875 880
 Trp Lys Pro Asn Arg Phe Met Gly Val Asp Ile Gly Glu Tyr Ala Val
 885 890 895
 Ala Tyr Cys Val Ile Glu Lys Gly Thr Asp Ser Ile Glu Ile Leu Asp
 900 905 910
 Cys Gly Ile Val Arg Asn Gly Ala His Arg Val Leu Lys Glu Lys Val
 915 920 925
 Asp Arg Leu Lys Arg Arg Gln Arg Ser Met Thr Phe Gly Ala Met Asp
 930 935 940
 Thr Ser Ile Ala Ala Ala Arg Glu Ser Leu Val Gly Asn Tyr Arg Asn
 945 950 955 960
 Arg Leu His Ala Ile Ala Leu Lys His Gly Ala Lys Leu Val Tyr Glu
 965 970 975
 Tyr Glu Val Ser Ala Phe Glu Ser Gly Gly Asn Arg Ile Lys Lys Val
 980 985 990
 Tyr Glu Thr Leu Lys Lys Ser Asp Cys Thr Gly Glu Thr Glu Ala Asp

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

1055	1060	1065
Ala Tyr His Glu Phe Pro Phe	Glu Asp Gly Leu Lys	Val Arg Gly
1070	1075	1080
Phe Leu Lys Pro Asn Lys Ile	Ile Ser Asp Gln Lys	Glu Leu Ala
1085	1090	1095
Lys Ala Val His Ala Tyr Met	Arg Pro Pro Leu Val	Ala Leu Gly
1100	1105	1110
Lys Arg Lys Leu Pro Lys Asn	Ala Arg Tyr Arg Arg	Gly Asn Ser

1115	1120	1125
Ser Leu Phe Arg Cys Pro Phe	Ser Asp Cys Gly Phe	Thr Ala Asp
1130	1135	1140
Ala Asp Ile Gln Ala Ala Tyr	Asn Ile Ala Val Lys	Gln Leu Tyr
1145	1150	1155
Lys Pro Lys Lys Gly Tyr Pro	Lys Glu Arg Lys Trp	Gln Asp Phe
1160	1165	1170
Val Ile Leu Lys Pro Lys Glu	Pro Ser Lys Leu Phe	Asp Lys Gln

1175	1180	1185
Phe Tyr Arg Pro Asn		
1190		

<210> 9

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 9

Ala Ala Ala Ala

1

<210> 10

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 10

Ala Ala Ala Ala

1

<210> 11

<211> 25

<212> RNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 11

cuccgaaagu aucggggaua aaggc

25

<210> 12

<211> 25

<212> RNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 12

caccgaaauu uggagaggau aaggc

25

<210> 13

<211> 25

<212> RNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 13

cuccgaauua ucgggaggau aaggc

25

<210> 14

<211> 25

<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 14	
ccccgaauau aggggacaaa aaggc	25
<210> 15	
<211> 36	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 15	
gucuagacau acagguggaa aggugagagu aaagac	36
<210> 16	
<211> 25	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 16	
cuccgugaau acguggggua aaggc	25
<210> 17	
<211>	
> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 17	
aaaaaaaaaa	10
<210> 18	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 18	
aaaaaaaaaa	10

<210> 19	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 19	
aaaaaaaaaa	10
<210> 20	
<211> 43	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 20	
cuccgaaagu aucggggaua aaggcaucaa uaccaaacuc ugg	43
<210> 21	
<211> 6430	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 21	
ttaaaaggac agtttctaata agcatataat cattatagca ttacatacgg aaaactactt	60
caaatttgcg gcagatcgga ttttgctggc ccagagatat attttccttc ttgttataaa	120
gcggatttat ggcaagggca gagccagttt ttatttttat cttccattc aacgatgcat	180
ccaagatgtg ggcaattgg agagagtttt aaaatttctc ctttttcatt ttgtatacgc	240
gcaactttct ttccttctat ctcaacaatt tttcctgtgt tgttttttaa attgtctaaa	300
gtaccgaag ttttcataaa gcgccccttc ataaaaagat aaggaaaaag aaatatattgt	360
tttaataatg ttaacatata gcttgttgaa ttataacatt tatccgagag gtggtctaac	420
ttatgcaact tattgattct tacttttagga gaatagtctt actctaggcg tatagagaac	480
ttttgttgaa aggtttttgc aatatctcta ctttctggcc aaaaatcggg ttttcccgcg	540
aatctgccgt atagtttgta tcttcttcta acaggtctgc ctccgctagg ttttcccggg	600
aaaggtacta taaatctctt atttcctaag agataagagc gcaaaccgag aattaagcca	660
tgatagagtt cctgaaaagt agcagtttgg cgagttgctg caacataaat ttctgtatcc	720

atgaaatcct ttaggttttc cattgtatag ggaagtgttt tactttcatc cccaccgttt	780
tcttgatatct cttttattgt attaaaggcg actccgtcga taaaacctct atatggttcc	840
atcaaactgt agattagaga ggggtaatct gaaggtgtgt ggggtgatcc gtgaaaagga	900
ctaaaatgat ggtaaaccac ccaacgcaag ataataccgc taacaaatit tgaagaagca	960
tctaaaacat tacagataaa attacctttt gatcgtcgcc tatcttttagg atatcccaaa	1020
gactttaga aatgttccca atatcttttg gcatgccacg attccactcc aactatagac	1080
tccacggacg ataagccctg cagttcctgc gttggggctg ggattaacca ttccatggat	1140
ttgaatttag cgtaaatcaa tcttttcgtt atatatgcgc gtttcttttc attttgtctg	1200
aatagaatct gttttgtag taaatcttct ctattagatg ttgtagaagg aacgatccaa	1260
acaccgctgg gcatatttgc tcgatgtatt gttaaaggaa tgccccaagc actgcatttt	1320
tctagaaatt cttgttctag cggacaaacg ctaccataaa acatgataga gtgaatctct	1380
ggaaaggaca aatccagctc accacctttg taagagaatt taacactctt tcccgataag	1440
tctatggatt ttacataggg taaccagata aattgtttac gcttggcgaa atatctctc	1500
atttcgtatt ggatatatgt ctcaaattat gctatattta aggtacattt tcaagcggtt	1560
tttagctcgt ttacatttta atatcaacaa aatcggggag aagtcctcca aagtatcggg	1620
gataaaggca tcaataccaa actctggctc cgaaagtatc ggggataaag gcattcccaa	1680
tatctcatta ctccgaaagt atcggggata aaggctctc cegtatctgt caactccgaa	1740
agtatcgggg ataaaggctt aaaaaggaa accccactcc gaaagtatcg gggataaagg	1800
cttgtactcc acatccgcta ctccgaaagt atcggggata aaggcactga aacttgaatt	1860
gtactccgaa agtatcgggg ataaaggcat ctgcgactt tctcttctcc gaaagtatcg	1920
gggataaagg ctcttcggtt ggtacgggtt ctccgaaagt atcggggata aaggcttatg	1980
gcagtatcgc atactccgaa agtatcgggg ataaaggctt cataagtacg cctaaactcc	2040
gaaagtatcg gggataaagg cagatgaggc tatacttaac tccgaaagta tcggggataa	2100
aggcacaaac ataaaggga aactccgaaa gtatcgggga taaaggcata aatctggtga	2160
acttactccg aaagtatcgg ggataaaggc tactgttatt gttgtacact ccgaaagtat	2220
cggggataaa ggcataacta gcgttcccat tctccgaaag tatcaaaata aaaagggttt	2280
ccagttttta actaaacttt agccttccac cctttcctga tttgttgat aattaataat	2340
gcgcaaaaaa ttgtttaagg gttacatttt acataataag aggcttgtat atacaggtaa	2400
agctgcaata cgttctatta aatatccatt agtcgctcca aataaaacag ccttaacaa	2460
tttatcagaa aagataatit atgattatga gcattttatc ggacctttaa atgtggctag	2520
ctatgcaaga aattcaaaca ggtacagcct tgtggatttt tggatagata gcttgcgagc	2580

agggtgaatt tggcaaagca aaagtacttc gctaattgat ttgataagta agctagaagg 2640
atctaaatcc ccatcagaaa agatatattga acaaatagat tttagctaa aaaataagtt 2700

ggataaagag caattcaaag atattattct tcttaataca ggaattcgtt ctagcagtaa 2760
tgttcgcagt ttgagggggc gctttctaaa gtgtttttaa gaggaattta gagataccga 2820
agaggttatt gcctgtgtag ataaatggag caaggacctt atcgttagagg gtaaaagtat 2880
actagttagt aaacagtttc tttattggga agaagagttt ggtattaaaa tttttcctca 2940
ttttaagat aatcacgatt taccaaaact aacttttttt gtggagcctt ccttgggaatt 3000
tagtccgcac ctcccttag ccaactgtct tgagcgtttg aaaaaattcg atatttcgag 3060
tgaaagtttg ctcggttag acaataattt ttcggccttt tctaattatt tcaatgagct 3120

ttttaactta ttgtccaggg gggagattaa aaagattgta acagctgtcc ttgctgtttc 3180
taaactgtgg gagaatgagc cagaattgga aaagcgctta cattttttga gtgagaaggc 3240
aaagttatta gggtagccta agcttacttc ttcgtgggag gattatagaa tgattattgg 3300
cggaaaaatt aaatcttggc attctaacta tacccaacaa ttaataaaaag ttagagagga 3360
cttaaagaaa catcaaatcg cctttgataa attacaggaa gatttaaaaa aagtagtaga 3420
tagctcttta agagaacaaa tagaagctca acgagaagct ttgcttcctt tgcttgatac 3480
catgttaaaa gaaaaagatt tttccgatga tttagagctt tacagattta tcttgtcaga 3540

ttttaagagt ttgttaaatg ggtcttatca aagatatatt caaacagaag aggagagaaa 3600
ggaggacaga gatgttacca aaaaatataa agatttatat agtaatttgc gcaacatacc 3660
tagatttttt ggggaaagta aaaaggaaca attcaataaa tttataaata aatctctccc 3720
gaccatagat gtgtgtttta aaatacttga ggatattcgt aatgctctag aaactgtaag 3780
tgttcgcaaa ccccttcaa taacagaaga gtatgtaaca aagcaacttg agaagttaag 3840
tagaaagtac aaaattaacg cctttaattc aaacagattt aaacaaataa ctgaacaggt 3900
gctcagaaaa tataataacg gagaactacc aaagatctcg gaggtttttt atagatacce 3960

gagagaatct catgtggcta taagaatatt acctgttaaa ataagcaatc caagaaagga 4020
tatatcttat ctctcgcaca aatatcaaat tagccccgac tggaaaaaca gtaaccagg 4080
agaagttgta gatttgatag agatatataa attgacattg ggttggctct tgagttgtaa 4140
caaggatttt tcgatggatt tttcatcgta tgacttgaaa ctcttccag aagccgcttc 4200
cctcataaaa aattttggct cttgcttgag tggttactat ttaagcaaaa tgatatttaa 4260
ttgcataacc agtgaaataa aggggatgat tactttatat actagagaca agtttgttgt 4320

tagatatgtt acacaaatga taggtagcaa tcagaaatth cctttgttat gtttggtggg	4380
agagaaacag actaaaaact tttctcgcaa ctggggtgta ttgatagaag agaagggaga	4440
tttgggggag gaaaaaaacc aggaaaaatg tttgatattt aaggataaaa cagattttgc	4500
taaagctaaa gaagtagaaa tttttaaaaa taatatttgg cgtatcagaa cctctaagta	4560
ccaaatccaa tttttgaata ggctttttta gaaaaccaa gaatgggatt taatgaatct	4620
tgtattgagc gagcctagct tagtattgga ggaggatgg ggtgtttcgt gggataaaga	4680
taaactttta cctttactga agaaagaaaa atcttgcgaa gaaagattat attactcact	4740
tccccttaac ttggtgcctg ccacagatta taaggagcaa tctgcagaaa tagagcaaag	4800
gaatacatat ttgggtttgg atgttggaga atttgggtgt gcctatgcag tggtaagaat	4860
agtaagggac agaataagagc ttctgtcctg gggatttcctt aaggaccag ctcttcgaaa	4920
aataagagag cgtgtacagg atatgaagaa aaagcaggta atggcagtat tttctagctc	4980
ttccacagct gtgcgcgag tacgagaaat ggctatacac tctttaagaa atcaaattca	5040
tagcattgct ttggcgata aagcaaagat aatttatgag atatctataa gcaattttga	5100
gacagtggtt aatagaatgg ctaaaatata ccgatctata aaggtttcag atgtttatag	5160
ggagagtgtt gcggataccc tagtttcaga gatgatctgg ggcaaaaaga ataagcaaat	5220
gggaaacat atatcttctt atgcgacaag ttacacttgt tgcaattgtg caagaacccc	5280
ttttgaactt gttatagata atgacaagga atatgaaaag ggaggcgacg aatttattht	5340
taatgttggc gatgaaaaga aggttaagggg gtttttacia aagagtctgt taggaaaaac	5400
aattaaaggg aaggaagtgt tgaagtctat aaaagagtac gcaaggccgc ctataaggga	5460
agtcttgctt gaaggagaag atgtagagca gttgttgaag aggagaggaa atagctatat	5520
ttatagatgc cctttttgtg gatataaaac tgatcggat attcaagcg cgttgaatat	5580
agctttagg ggatatattt cggataacgc aaaggatgct gtgaaggaag gagaaagaaa	5640
attagattac attttgaag ttagaaaatt gtgggagaag aatggagctg ttttgagaag	5700
cgccaaatth ttatagttat attggatata tcttttcaaa aaatctgaat tggcttagga	5760
ccgcggaatc ctatggtaat ttctacgtcc agaatgtagc gccatgccat tagaccagtc	5820
cccgaattaa acatcgccga acttcttggg gatgttatgg caaagagaat gcgacagcgc	5880
ctattcattg agcaagatat ggaaagtatt cctccagggc aaacaatggg tttgaatatg	5940
ggggagcctg ttgtgggaac ggaatttaca catcggcgga atattaatgg gaaagagtgc	6000
gttttattht ttgcagtga actttttaaa gacgacagcg cgtagtcagt acatcttcgg	6060

cccatcttaa tcttccattg gggttattaa gactgcccac tttagcagca agatttttaa 6120
 ggtgactcct taattctttc tcgtgcggag ttagatctat ttttcaaaa tctttatccg 6180
 catggttttag gaatatttgt atagagtcta ggggaatttc cttaccgatg tcccccgctg 6240
 cggtaacaac tcigtaaaga tccatcttta ttgaatttaa tataaactgt ctgtcttttt 6300
 tcatatttct aaatgctttt ttgttaattc aaataaccta cccctcacat tcttatcgta 6360
 tatctcatat gtatacttac ctatgacagg ttgttaattt ctcatagcca tatattcaac 6420
 ttcttttgaa 6430

<210> 22

<211> 13819

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 22

ctctttttct tgactatggt catcgcttag ctggcgggg acgtttgatc ttgcttcta 60
 gttaaatcct ttttctgtcc ttgttgtttt taatgaatta cccctcaatt tggggtttat 120
 tagcttttag tttattggct ttagtgattc taacttgggt gaaaaaggct tggactaaat 180
 ggttgtagt cccactgata atttttctgc tggttgacac tctagcgatt ttgcttcaa 240
 aacctatfff agctaaacca atttttgatc taaatcaaag ttgaaaatt aatagttttg 300
 attcgcgacc taatttagat agcactgctc aagtactaa agccagtttg aaagctcatc 360

cctttttagg ttttggcca aatcgttttt ggcgagcttg gactctttat aagccaaaat 420
 tatttaatca atcagtaatc tggtcagttg attatcgtct ggcttatggt tttattccaa 480
 caatgttagt aactcaagggt ggctcgggtt ttctggcttg gtttaattctg ataatttcta 540
 gttttattta tctttatcat ttattcaaac aaagttcagt agaagatttt tccacgataa 600
 ttttatttag tctaagtttt atttatctct ggtaaattt actcattctt aatcctaatt 660
 ttgttatect ctctctggct tttgggtgct tgggtgggtt gttagttttt aatcataaaa 720
 tttctaataca gctttcttgg cacattaaat tagatacgtt tctaaaaagt ttagtggcaa 780

aactaggtct tagtattatt ttgggttttt tatttttaat cattattttg tcaactgctta 840
 attatagttc ttgatctta ttcatcggg gtctttcatc tttggatcgg ggggattttt 900
 ccgccaccga aaaaaattgg cgtttagcta gtcgtttgag tcctcagaca gtttataatc 960
 gttctttggc tgatcttaaa ctgcgtcaga ttaatcaact tctgacgact cctaattctg 1020
 attctcaaaa aacttttagcc gagttttccc gtttttatgg tgagtcaatt ggatttggct 1080
 tgactgctcg tgaccaagat ccttttgatt atttaaatg gtttaattta ggtcaagttt 1140

atgaagctgg gattccgctt aaaattaaag gggccgatat tcaagctcgg aaaatttatac	1200
aagaagtgc tagattaaac ccggtttggc cagtcatttg gctaaatttg gctcgagtgg	1260
aattaggctc tgatcacctt gatttagcgc gagaagattt acttaaagct ttggaattaa	1320
aagccgatta ttccgatgct ctgttagctt tagccgaatt agattatagt caaggtcgat	1380
tatcaaaagc tttagcggga gctaagggtg cagttctgaa agaaccaaat aatttgggag	1440
cttggttttc ctttggtttt ttccagtatc aaattggaca ttatgatgaa gctgtcattt	1500
ctttagaaaa agtcttaacc tttaatcaaa attcagctga tactaaatat tttcttgggt	1560
taagtttagc tgaacttgat cgaacgactg aggcgattga cctatttcaa tctttagttc	1620
gggctaatac cgacaatcaa gagcttaaaa atattttaac taatctcaa gctggtcgaa	1680
cagcttttagc gccaccagag accaaaacca aaacaaaata ataattcatg gtgtctaaaa	1740
ttactcgctt acttcaaaaa gaatttacca atcttcacca agcagctttt ttgttggcta	1800
cttcggcctt gcgtctcaa tttttgggtt tgtggcggga tcgtttatta gcctctgggt	1860
ttggagctag tcatcaatta gatatttatt atacggcttt tcgcttaccg gatttaattt	1920
acgtttcggt ggcttctttt gtttcgatca cggtcctat tcttttgatt attaataaga	1980
tggaaactgg tggtaaaccg gcggtggaaa aattttctca ttcagtgtcg acagtttttt	2040
taattgggat ggtttcagtt tccgcgttat tatttatatt tatgccctgg ttatcgaaaa	2100
ttaccgctcc cgggttttct tcagttgatc aacaaacctt agtcacctta tctcgaattt	2160
tattgttgtc tcccttattg ttgggtttgt ctaatctctt gggaggagcc actcaagctt	2220
ttcgtaaatt tgccgctat gccttttagt ctatttttta taatttggga attatatttg	2280
ggattttctt ttctatcct ttgcttgggt tgccgggctt agtctgggga gtaattctcg	2340
gtgcagctctt acatttatca attcaattgc cagttttaag tcaattaggt ttacgtcttc	2400
gtttatcgag attaatfaat tggccggaaa tgagaaaagt gatgtcata tccctaccgc	2460
gaactattac ctatcggtt aatcaactat ctttattagt tttagtggct ttagcttcgt	2520
ttttgcccaa aggttcaatt tcggttttta atttttcgct caatcttcaa tcagtcccc	2580
tgctgattat cggagtttct tattcgggtg cggtttttcc cgtcttggcc aaattttttg	2640
tcgctggtca acacaaagaa ttgtctgttg aaattatcgc cgccattcga catattattt	2700
tttggctcgc tccagtgtgc gttttgttta ttgttttacg agctcaaac gtccgggtga	2760
ttttaggctc aggacgtttt gattggctcg ccaactcgatt gacggcagct tgtttggcga	2820
ttttttctgt gtcagtgatt gctcaaagtt tgattttagt tttagtccga gcttactatg	2880

ccgctgggga aaccaaact cccttgatca ttaattcctt atcatctttg ggaacaatta 2940
 ttttgctttt aattttatgg caactgttca aagtttggcc ggcctttcat ctgattttgg 3000
 aacaaattct aagattgaaa gatttaccag ggacaattat tttagtctta cctctcgttt 3060
 tttcgattgg agcgattatc aatgtttttg ttttatggtg ggccttcgaa cgacgctttg 3120
 ctatcggaat ttggcgcaat ttagaggtag ttagtcttca gtccttagtc gcttctttat 3180
 ttggtggctt tgtggcctat aacttactaa atgtcttttag tctgtattat aaattagata 3240
 ctttttggtc aatctttgag cagggatttt tagccggtat tttgggctta attgcctgga 3300

 tttcggtctt aattcttttg aaaagtgaag aattggctga attgggacgt tctctgtcag 3360
 cccgagtctg gaaagtgtc cctattgtcc cagaacgaga agaactgtag gatgggaaag 3420
 tctttatatg gatttaaac actatcgtaa ttttctatt attgctcacc ccagtagaac 3480
 agccaagctg tctacggggc aagtattgat cataaattag tcttatggat ttaaaacact 3540
 atcgtaattt ttctattatt gcccatatag atcatgggaa gagiactttg tctgatcggc 3600
 ttttagattt gacagggaca attgaaaagc gaaaaatgcg agaacaagtc cttgattcga 3660
 tggagttaga acgtgaacga ggaataacca tcaaatgca accagtccga atgaattata 3720

 aattggctgg tgaagattat attctgaatc taattgatac tccgggtcat attgattttt 3780
 cttatgaagt gtctcgttcg cttcaagcag tggaaggggt cttgcttttg gttgacgcca 3840
 ctcaaggggt ccaagctcaa acttttactg ttttagcgat ggctcaagaa ttgggtttaa 3900
 cgattattcc cgttttaaac aaaattgatt taccaattgc tcgaacagct gaagcaaac 3960
 aagagattgt taatctatta aaatgtcagc ccgaagatat tatggcggtt tctggcaaaa 4020
 ccggtgaagg agtagataaa ttattaattg agattattaa aaaaattcct agtccaattt 4080
 cagaaataaa agttgttaaa ccttgccgag cgctggtatt tgattttgaa tattctattc 4140

 ataaaggagt ggtggtctat gttcgagttt tagatggcga aattactccc gctgatcaac 4200
 taaactttgt cgcttctggt gaaaaatttt cggttttaga attaggttat tttcgacctc 4260
 aagctgaacc acaaaaaaaaa ttacaggcgg gtgacattgg ttatttagtc actggaatta 4320
 aaaaaccagg caatgctaaa gtgggggata cgattaccac tttagttagt cctcttccag 4380
 ctgtaccggg ctatatgact cctcgaccgg tggctctgggc ttctctttat ccagctagcc 4440
 aagatgattt tgccttactc aagcaatccc tcgaacgatt aaatcttcaa gatgccgctc 4500
 tgtcttttga agaggaaagc tcgggtgctt tgggacgagg ttttagagct ggttttctgg 4560

 gaatgcttca ttggaaatc attagcgaac gattgaagcg agaattttct ttaaatttaa 4620
 ttgtgacgac accgagtatt agttatcgtc taattaatac tcggacacaa gaagaagtca 4680
 ggattttctc tctcacctt ttccacttg aaatcaagga ttatgaaatt tacgaatctt 4740

gggtagcggg tagaattatt agtcccgcg attatcttag tccgattatt caattacttc 4800
atgaacacga agcgaagta atgactatgg aaacttttag ttctagtcgc accgctttgt 4860
ctatcctcat gcctttacga gaattgatgc gtaatTTTT tgatagtta aaaagtgtct 4920
cttctggctt tgcttctttt tcttatgaat tagccgaaga acgtctcgct gatgtctctc 4980

gcttgatat ttttaattaat ggtgaataa ttcgggcttt ttcggaatt gtttcgcgtc 5040
gacgaatcga aaaggatgct tcggaatgg ctgaacgttt agagggtttg attcccaaac 5100
aattgattac gattaaaac caagttcaag gtttagggcg aattttggcg gcgcgttcaa 5160
tttccgctct acgaaaagat gtcactgact atctctatgg cggcgatatt actcgaaaaa 5220
tgaaattacg agaaaagcag aaaaaaggca agaaaaaaat gcaacagctg ggtaaggtaa 5280
atatcccca agaagttttt ctaaagatga tgcgaaatgc ggactagcgc ggactggacg 5340
cagactaatg cgaatttacc ctatggagta gcttgctata ctccataggg taaacgcaga 5400

tagtcacaaa caagacactg atcagatcag cgttttttta gcattgatcg gcgttttattc 5460
taaacaagaa ggggagagag taaagggcga ccatacttaa aataacaaga ataccaactg 5520
tcgctgagat gatttgaag atttttttgt gtttgctctg aaataacatt agttgtagta 5580
taaggctgtg accagatttt atcaagtcga aaaacatttt aagtggctaa atgttctctt 5640
tcttattgtc actttaatct tgggtatttt ttgggctcga ggggtttggc gagtttataa 5700
tcagagtctg ttgctaatt ctaattatct ttgactaaa gatcgtctta ctaaattaga 5760
agacagacaa aaacaaatta ctgatcgtct agaaaaatta tcaaccgatc gtggtttaga 5820

agaagaattt agaaataatt tticagtcgt gcgaccaggg gaaaaaatga ttttaattgt 5880
cgatagtatt gaaacagcta ctgatacagc cactactagt gaggctagtc ttggggggac 5940
tttaaaagcc ttattattaa gtcgttaatt aaaaaagcga gattggttca gcttgccctc 6000
ttaaatttct tgtgcaaata tgcgggtatg gtttagtttg ccttttaaaa tttttgtcc 6060
gaacatgcga gtatggttta gtggtagaat gcgaccttcc caaggttgag acgcgagttc 6120
gattctcgct actgcacaa aaaacttttt aggggtgaata gaatgcgacc cccgaagaac 6180
agcaaagctg tctacggggc aggtttccca agcataagac gctggttcga tttccgcatt 6240

tcgcacaatt ggccgattaa aatagtattt tttttttta tgtcctccac ctttaaacga 6300
actatcgaaa attttacttg tgctcattgt ggagcggagg tgattggtaa tggttatact 6360
gatcactgtc ctaaagcct ctggggcctc catgtagatg atttcccggg agatcgagct 6420
aatccttggt tgggcttaat gaagccgatt ggagtggatt tagcgaaggg agattatact 6480
ttaagctatc aatgtgaaaa atgtcacatc attaaaaacta ataaaactgc tccggacgat 6540
gaacttaaca agtacttgac cggtagttta taattgttaa ataagttaaa tttaaaatat 6600

aaaatgaaga aagttacat ttattccact cccacttggt gttattgtaa aattgctaaa	6660
caattcttta aagataaggg aattgatttt acagagattg atgtcactac tgatttagct	6720
gggcgacagg ccttagaaca aaaaattggc cgaattacgg gtgtgccagt aattacgatt	6780
gacgaagaag ccgtcgtggg ttttgatcaa gctcatattg cgaagatgtt agggatttaa	6840
actagtaca atttaccg ccttctgcca gccggttagag gatgggtttt ttggtaatt	6900
tgctaacaac aaacaaggag tctattatga agattaagtt ttgcctctg tagttccgc	6960
cataatcctt aaataaattt aggattatgg cgggcgggaa acaagccggt taacgctctc	7020
atagttcaaa ggatagaact gtctcgtcct aagagaccaa tctccgttcg agtcggagt	7080
agagcacaga ttaaaaaaca ttgactagag tctacttgc cagcctaaga ttgctttag	7140
taaagttttg gcgggaggga aagatgtagg ttcgattcct accagaggca caattcgtaa	7200
cttggtcaaa tcatittcaa aacaaatgat accacacaca gaggagagga tatggggcac	7260
agccttcgtc agtttgataa ctcaaggaaa caaatctaaa aataaaactt caccgatgtt	7320
atcatttga agccatctgc ttttcgtgcg tatttgaac atttttggca acactccaga	7380
aatcggtagg gccggccgtc ctccatata gtttgaacc aacttttact ggtcgccac	7440
cgtttgggtt tcttgaatt ggtacgatga attgtcgga tttccctgt aagtatgatc	7500
gtaacgataa tacagaccg tgtaataatt ctgaaaagt tactatctga cgagtagtgt	7560
tggataaag atttgatcg agataatctt caacggcgat tatacacctt gccagaaaat	7620
cctttcttc aaccttttc gatttagctt gctgtattgt attgaataca atttttcaa	7680
tatttcctct ataaggttcc attagatcgt aaactaaaga aggataatct gtcggaatgt	7740
ggagaaatcc atgatagggg ctcatcgggt ggtaaattat ataacgcagt gtaataccgc	7800
ttattaattt tgaaccgcg tccaaaatag attttatcgt atttcccct cctctgcgtg	7860
aatatccact gtatccgagt attttataat atttttcca atatacctt gcattgctgtg	7920
cctcaatgtt taccatttgt ttaatagaat atcttttacc gtcaataac attggatatg	7980
aaaccagcca actcatactt ttgaatttgg ctgtataaat cttcttggca atgtggacac	8040
gtttcttctc attatttcta aatgagattt gcttacttaa gatatcatct ttggccgagg	8100
tttttacgt tggagtaac catacggcat tactcattgt tcttcggtgt agacatatag	8160
gaacgcata ttttgcgcaa agttgtaaaa aattttcact taaatcaca gtccaccat	8220
aaagcataat cgaagaatg tttttaatgt ttgcggtata ttgccacct ttatattgaa	8280
aagttacaat atttttctt acttctattt ggaagggtga gggtagccat aagggtatct	8340

ttttattctt gctaatagac atgttttttg atattattac cctagaaaga gttaggtttt	8400
gaatacaaaa tctaacttat attttgtatt ttgtcaagta aaataaagag aaaagagaga	8460
acctcaccga aatttggaga ggataaggca agacaacaca catcttgac cgaaatttgg	8520
agaggataag gcataccgct ctggctttga acaccgaaat ttggagagga taaggcaata	8580
ttcaaaatat ctaccaccga aatttggaga ggataaggct caatcttttt atagcctaca	8640
ccgaaatttg gagaggataa ggcaactcaa cataaagggt gcaccgaaat ttggagagga	8700
taaggcggat cgagataagt cgaacaccga aatttggaga ggataaggcg ctaacaaaat	8760
taccaccac cgaaatttgg agaggataag gcaaaccagc agggacttca caccgaaatt	8820
ttggagaggat aaggcacaat tgtcatgttt attcaccgaa atttggagag gataaggctc	8880
gtttatgtta gcgaccacac cgaaatttgg agaggataag gcaagaaaca ataaccgcag	8940
aacaccgaaa tttggagagg ataaggccaa ttataatata gcctgcaccg aaatttggag	9000
aggataaggc aagatactgt tccaataaca ccgaaatttg gagaggataa ggcaaatat	9060
cataatccat tcaccgaaat ttggagagga taaggcatgg cttgtttttg taatcaccga	9120
aatttggaga ggataaggca caggagagaaa ttgcgaacac cgaaatttgg agaggataag	9180
gcgtttggca ataagtctcg caccgaaatt ttggagaggat aaggcatggg tcaatccaac	9240
ccgtcaccga aatttggaga ggatgatggg ttgggttcaa aaattctaag aatctgcttt	9300
attttcttca cttcacctac acggctcttc gtctcgttcc ttctagtaac acgagacctc	9360
gcctttccga ccgttctctt tgtctcttta tttatctga cagaatatgc aaaaagtaag	9420
aaaaacttta tcagaggtag ataaaaatcc ttatggtaca aaagtccgta atgcaaagac	9480
ttggctactca ctacagatag agaggctttc gtatactgga aaagagggga tgagaagttt	9540
taagattcca ctcgaaaata aaaataaaga agtttttgat gaattcgtaa aaaagatcag	9600
gaatgattat atcagtcagg ttgggttgc caatctttct gattggtatg aacattatca	9660
ggagaaacaa gaacattatt ctttggcgga tttttggtta gatagtttga gggccggagt	9720
gatttttgcg cacaagaaa ctgagataaa gaatcttata tctaagatac gtggtgataa	9780
atcgattgtt gataaattta atgcaagtat aaagaaaaaa cacgccgac tttatgccct	9840
tgtcgatata aaagctctct acgattttct tacctccgac gcaagaaggg gattaaagac	9900
cgaagaagaa ttttttaact caaaaaggaa taccttgttt ccgaaattta gaaaaaaga	9960
taacaaagcc gtcgaccttt gggtaaaaa atttattggg ctggataata aagacaaatt	10020
aaattttacc aaaaagtta tcggtttcga tccaaatcct cagattaaat atgaccatac	10080
tttctctttt catcaagaca ttaattttga tctagagaga atcacgactc cgaaggaact	10140
tatttcgact tataagaaat tcttaggaaa aaataaggat ctatacggtt ctgatgaaac	10200

aacggaagat caacttaaaa tggatattagg ttttcataat aatcacggcg ctttttctaa	10260
gtatttcaac gcgagcttgg aagcttttag ggggagagac aactccttgg ttgaacaaat	10320
aattaataat tctccttact ggaatagcca tcggaaagaa ttggaaaaga gaatcatttt	10380
tttgcaagtt cagtctaaaa aaataaaaga gaccgaactg ggaaagcctc acgagtatct	10440
tgcgagtttt ggcggaagt ttgaatcttg ggtttcaaac tatttacgtc aggaagaaga	10500
ggtcaaacgt caactttttg gttatgagga gaataaaaaa ggccagaaaa aatttatcgt	10560
gggcaacaaa caagagctag ataaaatcat cagagggaca gatgagtatg agattaaagc	10620
gatttctaag gaaaccattg gacttactca gaaatgttta aaattacttg aacaactaaa	10680
agatagtgtc gatgattata cacttagcct atatcggcaa ctcatagtcg aattgagaat	10740
cagactgaat gtigaattcc aagaactta tccggaatta atcggtaaga gtgagaaaga	10800
taaagaaaaa gatgcgaaaa ataaacgggc agacaagcgt taccgcgaaa tttttaagga	10860
tataaaatta atccccaatt ttctcgggtga aacgaaacaa atggtatata agaaatttat	10920
tcgttccgct gacatccttt atgaaggaat aaattttatc gaccagatcg ataaacagat	10980
tactcaaaat ttgttgcctt gttttaagaa cgacaaggaa cggattgaat ttaccgaaaa	11040
acaatttgaa actttacggc gaaaatactt tctgatgaat agttcccgtt ttcacatgt	11100
tattgaagga ataatacaata ataggaaact tattgaaatg aaaaagagag aaaatagcga	11160
gttgaaaact ttctccgata gtaagtttgt tttatctaag ctttttctta aaaaaggcaa	11220
aaaatatgaa aatgaggctt attatacttt ttatataaat ccgaaagctc gtgaccagcg	11280
acggataaaa attgttcttg atataaatgg gaacaattca gtcggaattt tacaagatct	11340
tgtccaaaag ttgaaaccaa aatgggacga catcataaag aaaaatgata tgggagaatt	11400
aatcgatgca atcgagattg agaaagtccg gtcggcatc ttgatagcgt tatactgtga	11460
gcataaattc aaaattaaaa aagaactctt gtcattagat ttgtttgcca gtgcctatca	11520
atatctagaa ttggaagatg accctgaaga actttctggg acaaacctag gtcggttttt	11580
acaatccttg gtctgtccg aaattaaagg tgcgattaat aaaataagca ggacagaata	11640
tatagagcgg tatactgtcc agccgatgaa tacggagaaa aactatcctt tactcatcaa	11700
taaggaggga aaagccactt ggcatattgc tgctaaggat gacttgtcca agaagaagg	11760
tgggggcact gtcgctatga atcaaaaaat cgcaagaat ttttttgga aacaagatta	11820
taaaactgtg ttatgtctc aggataagcg gtttgatcta ctaacctcaa agtatcactt	11880
gcagttttta tctaaaactc ttgatactgg tggagggtct tgggtgaaaa aaaaaatat	11940
tgatttaaat ttaagctctt attctttcat tttcgaacaa aaagtaaaag tcgaatggga	12000
tttaaccaat ctigaccatc ctataaagat taagcctagc gagaacagtg atgatagaag	12060

gcttttcgta tccattcctt ttgttattaa accgaaacag acaaaaagaa aggatttgca	12120
aactcgagtc aattatatgg ggattgatat cggagaatat ggtttggctt ggacaattat	12180
taatattgat ttaaagaata aaaaaataaa taagatttca aaacaagggt tcatctatga	12240
gccgttgaca cataaagtc gcgattatgt tgctaccatt aaagataatc aggttagagg	12300
aacttttggc atgcctgata cgaaactagc cagattgcga gaaaatgccca ttaccagctt	12360
gcgcaatcaa gtgcatgata ttgctatgcg ctatgacgcc aaaccggtat atgaatttga	12420
aattttccaat ttigaaacgg ggtctaataa agtgaaagta atttatgatt cggttaagcg	12480
agctgatatc ggccgaggcc agaataatac cgaagcagac aatactgagg ttaatcttgt	12540
ctgggggaag acaagcaaac aatttggcag tcaaatcggc gcttatgcga caagttacat	12600
ctgttcattt tgtggttatt ctccatatta tgaatttgaa aattctaagt cgggagatga	12660
agaaggggct agagataatc tatatcagat gaagaaattg agtcgcccct ctcttgaaga	12720
tttcctccaa ggaaatccgg ttataagac atttagggat ttgataagt ataaaaacga	12780
tcaacggttg caaaagacgg gtgataaaga tggatgaatgg aaaacacaca gaggggaatac	12840
tgcaatatac gcctgtcaaa agtgtagaca tatctctgat gcggatatcc aagcatcata	12900
ttggattgct ttgaagcaag ttgtaagaga tttttataaa gacaaagaga tggatggtga	12960
tttgattcaa ggagataata aagacaagag aaaagtaaac gagcttaata gacttattgg	13020
agtacataaa gatgtgccta taataaataa aaatttaata acatcactcg acataaactt	13080
actatagagt tctcttcatt ggattgaaaa tagatccgat tcctaccaga gacaccaa	13140
aaatttaaaa ttaaaaatta cctgccaaaa ttctgttcaa cgaaacttaa gcaggcaaga	13200
aaatttaaaa ttaaatccgc tggtagggcg ataaagtc aaattgaaaa tatattaaat	13260
tgacaatatg ttctttatta gattgcgatg ttgtaatacc tcggggcttc gaatcagtag	13320
attcgtggct tggccataaa tccacaggta ttcaaacacg cgatgtgttt tgtatggcgg	13380
ggtgggcat acctattcta acaaaacaac catggtgttt ggctgacctc atacctcatc	13440
ggctctgccg tgaggatagg acacgcaact tgttttatta tgatataatg aaaggtagaa	13500
attgtcattt tgtaatggaa cagtaaaaaa gaggtgccgg tgatgaacaa aagagtgact	13560
aaaggagaca tcaggattta cctgatgatg tggaagggtg ctattatgac cgtctgtgtc	13620
gcgagtctgg ttggcatcat ccttgggtcca gtctatcttt tgatcatttt tccgttgaag	13680
aaaatgatca gaaggatct gatcgatttt tcggatttgc tcaaaggctt ttgatgactt	13740
ttaggcaaga agattgtttg ttagctctct accgcaagga ggagggtttt ttcttttttt	13800

taaattaatt tacctttca 13819

<210> 23

<211> 34045

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<220><221> misc_feature

<222> (29562)..(29573)

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

<400> 23

atgttcctc ttcttttctg tgcctctgaa taagatttgc ttactcaaga tatcttcctt 60

agaagacgtc tttatgcttg gggtaatcca gatcgcggtc ctcatcggtc tgcgatggat 120

gcaaacagga acactatatt tagtgcatag ttgcaagaaa tcctccttta aatcacaggt 180

gccgccataa agcattatcg ataagatgtt ttgacgtca gcagaataga cacctccttt 240

glaatggaaa gttatcttat cttttttcac ctctattgcg gaagtataag ggaaccatag 300

ggggattctt ctgttggtat ttttcatgtt ttgatatata attacactag atatgggcac 360

atttcaggag taaaatctaa cccatttttt gtattttgtc aaataaaaata aaggtaaagg 420

agagaacctc tccgaattat cgggaggata aggcagcgtc tgataattct tcctccgaat 480

tatcgggagg ataaggcaag actggtaaac tctagctccg aattatcggg aggataaggc 540

acagtaacaa catacgggct ccgaattatc gggaggataa ggcaaaactaa ccgttgctct 600

actccgaatt atcgggagga taaggcaaag cgtttaaagc cgacactccg aattatcggg 660

aggataaggc aaacgccta taacgaatc tccgaattat cgggaggata aggcgtagtt 720

agtggataat ttactccgaa ttatcgggag gataaggcga cgctgacgat aaactgctcc 780

gaattatcgg gaggataagg cacaacatt tcctcgacat ctccgaatta tcgggaggat 840

aaggcataat tactcgctcg acactccgaa ttatcgggag gataaggcaa aatcatatcg 900

ttcttgctcc gaattatcgg gaggataagg caccgccgaca aaattaagcc tccgaattat 960

cgggaggata agtatggata tttccacaat ctfgaaagaa agatttgta gcctttaatc 1020

cattctcctt tccttttatt ttatctgaca acatatgaaa gctaaaaaaa gtttttataa 1080

tcaaaagcgg aagttcggta aaagaggtta tcgtcttcac gatgaacgta tcgcgtattc 1140

aggagggatt ggatcgatgc gatctattaa atatgaattg aaggattcgt atggaattgc 1200

tgggcttcgt aatcgaatcg ctgacgcaac tatttctgat aataagtggc tgtacgggaa 1260

tataaatcta aatgattatt tagagtggcg atcttcaaag actgacaaac agattgaaga 1320

cggagaccga gaatcatcac tcctgggttt ttggctggaa gcgttacgac tgggattcgt 1380
gttttcaaaa caatctcatg ctccgaatga ttttaacgag accgctctac aagatttggt 1440
tgaaactctt gatgatgatt tgaaacatgt tcttgatagg aaaaaatggt gtgactttat 1500

caagatagga acacctaaga caaatgacca aggtcgttta aaaaaacaaa tcaagaattt 1560
gttaaaagga aacaagagag aggaaattga aaaaactctc aatgaatcag acgatgaatt 1620
gaaagagaaa ataacagaa ttgccgatgt ttttgcaaaa aataagtctg ataaatacac 1680
aattttcaaa ttagataaac ccaatacgga aaaatacccc agaatcaacg atgttcaggt 1740
ggcgtttttt tgtcatcccg attttgagga aattacagaa cgagatagaa caaagactct 1800
agatctgac attaatcggg ttaataagag atatgaaatt accgaaaata aaaaagatga 1860
caaaacttca aacaggatgg ccttgatttc cttgaaccag ggctatattc ctgcgctcct 1920

gaatgattta ttcttgtttg tcaaagacaa tgaggatgat tttagtcagt ttttatctga 1980
tttgagaat ttcttctctt tttccaacga acaaatfaaa ataataaagg aaaggttaaa 2040
aaaacttaaa aaatatgctg aaccaattcc cggaaaagccg caacttgctg ataaatggga 2100
cgattatgct tctgatttgg gcggttaaatt ggaaagctgg tactccaatc gaatagagaa 2160
attaaagaag attccggaag gcgtttccga tctgcggaat aatttggaag agatacgcaa 2220
tgttttaaaa aaacaaaata atgcatctaa aatcctggag ttatctcaaa agatcattga 2280
atacatcaga gattatggag tttcttttga aaagccggag ataattaagt tcagctggat 2340

aaataagacg aaggatgggc agaaaaaagt tttctatggt gcgaaaatgg cggatagaga 2400
attcatagaa aagcttgatt tatggatggc tgatttacgc agtcaattaa atgaatacaa 2460
tcaagataat aaagtttctt tcaaaaagaa aggtaaaaaa atagaagagc tcggtgtctt 2520
ggattttgct cttataaag cgaaaaaaa taaaagtaca aaaaatgaaa atggctggca 2580
acaaaaattg tcagaatcta ttcaatctgc cccgttatit tttggcgaag ggaatcgtgt 2640
acgaaatgaa gaagtttata atttgaagga ctttctgttt tcagaaatca agaatttga 2700
aaatatttta atgagctcgg aagcgggaaga cttaaaaaat ataaaaattg aatataaaga 2760

agatggcgcg aaaaaaggga actatgtctt gaatgtcttg gctagatttt acgcgagatt 2820
caatgaggat ggctatggtg gttggaacaa agtaaaaacc gttttggaag atattgcccg 2880
agaggcgggg actgattttt caaaatatgg aaataataac aatagaaatg ccggcagatt 2940
ttatctaaac ggccgcgaac gacaagtttt tactctaate aagtttgaaa aaagtatcac 3000
ggtggaaaaa atacttgaat tggtaaaatt acctagccta cttgatgaag cgtatagaga 3060
tttagtcaac gaaaataaaa atcataaatt acgcgacgta attcaattga gcaagacaat 3120

tatggctctg gttttatctc attctgataa agaaaaacaa attggaggaa attatatcca	3180
tagtaaatg agcggataca atgcgcttat ttcaaagcga gattttatct cgcggtatag	3240
cgtgcaaacg accaacggaa ctcaatgtaa attagccata ggaaaaggca aaagcaaaaa	3300
aggtaatgaa attgacaggt atttctacgc ttttcaattt ttttaagaatg acgacagcaa	3360
aattaattta aaggtaatca aaaataattc gcataaaaac atcgatttca acgacaatga	3420
aaataaaatt aacgcattgc aagtgtattc atcaaactat cagattcaat tcttagactg	3480
gttttttgaa aaacatcaag ggaagaaaac atcgctcgag gtcggcggat cttttacat	3540
cgccgaaaag agtttgacaa tagactggtc ggggagtaat ccgagagtcg gttttaaaag	3600
aagcgacacg gaagaaaaga gggtttttgt ctcgcaacca ttacattaa taccagacga	3660
tgaagacaaa gagcgtcgta aagaaagaat gataaagacg aaaaaccgtt ttatcggat	3720
cgatatcggt gaatatggtc tggcttggag tctaactgaa gtggacaatg gagataaaaa	3780
taatagagga attagacaac ttgagagcgg ttttattaca gacaatcagc agcaagtctt	3840
aaagaaaaac gtaaaatcct ggaggcaaaa ccaaatcgt caaacgttta cttcaccaga	3900
cacaaaaatt gctcgtcttc gtgaaagttt gatcggaagt tacaaaaatc aactggaaag	3960
tctgatggtt gctaaaaag caaatcttag tttgaatac gaagtttccg ggtttgaagt	4020
tgggggaaag agggttgcaa aaatatacga tagtataaag cgtgggtcgg tgcgtaaaaa	4080
ggataataac tcacaaatg atcaaagttg gggtaaaaag ggaattaatg agtggtcatt	4140
cgagacgacg gctgccgga catcgcaatt ttgtactcat tgcaagcggg ggagcagttt	4200
agcgatagta gatattgaag aatatgaatt aaaagattac aacgataatt tatittaaggt	4260
aaaaattaat gatggtagaag ttctgtctct tggtaaagaa ggttggagat ccggcgaaaa	4320
gatcaaggg aaagaattat ttgggtccgt caaagacgca atgcgccaa atgttgacgg	4380
actagggatg aaaattgtaa aaagaaaata tctaaaactt gatctccgag attgggtttc	4440
aagatatggg aatatggcta ttttcatctg tccttatgtc gattgccacc atatctctca	4500
tgcggataaa caagctgctt ttaatattgc cgtgcgaggg tatttgaaaa gcgttaatcc	4560
tgacagagca ataaaacacg gagataaagg tttgtctagg gactttttgt gccagaaga	4620
gggtaagctt aattttgaac aaatagggtt attatgaatc taaaaatagt cgtgatcaac	4680
aaactcaatc atttgaaaa tttttatcgt cgccatccaa agaaaatcct ttggttgggg	4740
gtgccattgc tattgcttat cgggttgggg gcttgggctt atactcgag gactcaaccc	4800
gagttcgaaa cagaggtggt gaagttgggc gaggtggccg atgtggtgag cgatactggt	4860

ttggtgacgg ccgagaatga tctcactctc tcgttcgaga cgggcgggggt cgttcgcacg	4920
gttaaggtta ccgaagggtga cgcgggtttat cgaggacaga cgttagtctc gctggatgcc	4980
agtttgaagg cggcggaagt ggcgagcgcg cgcgccacgt tggccgctca agaagccaaa	5040
ttggctgaac tggtagcggg cccgaccaag ctgatttag cttcgccaa gacgaaactc	5100
gagaacgccc gcaagacctt gctgaccgcc gacctgcaag cgtacttcgc cggctcttca	5160
gccgattatg cggcttcttc attcatttat acggcgccga cggttttggg gacttacaat	5220
tccgatcaag agggcgaata cgtgcttgag ttatatcaat caggcgcgcc gtcgggctac	5280
tcggtggagt actccggttt ggagacgggg attatggagg gcgccgaagg acgagccgag	5340
cccttgggcc ggcgcggtct ctatctccaa ttcccgaga attcattcg ggccgagag	5400
gtaatttggc gcgtgcctat cccaacacc aagtccgctt cttatgtac taaccggcgc	5460
gcctacgaac aggtcaagc cgattacgac ctgaaagtgg ctggcactcg cgccgaacaa	5520
attgtcgccg ccgaagccca agcgcgccaa gcccgcgcca ccctccaatc ggcgaggcc	5580
tcgtgtcca agctctccct tacggcgccg gtggccggtt tggtagatc cgttcggtt	5640
accgtagggg agacggttac cgttggttca ccagctgtgg cgttggtctc ggatcataat	5700
tattacgtga ccctctatgt gccggaggct gagatggcca acttgacggt cggcgacttg	5760
gccgagatcc ggctcaaggc cttccccgat cgcgtcttc gcgccaccgt ggggagtgtg	5820
gccccggcgg ccgaagatcg tgatggcgtg gcttcgttta aagttaaatt atatttcaa	5880
gaatccgac ccaaatag agtggggatg tcggctgacg tcgacctga ggcgcttaag	5940
aagaccgacg tcatggtggt gcccggcggg gcggtggtgc gctctaattg gcgaatcttt	6000
gtccgggttt ggagcaataa gaccgtcgag gaacgctcgg tggagattgg tctgctggc	6060
tctgatggct cggtaggat tgtctcgga ctctcggtgg gcgaagaggt gattactttt	6120
atccgtgacg aggagttaga tcgcttggcg gactaatcc ctttcggcgt ttatggcttt	6180
acttgaactc gaccaagtta ctaaatttta ttatagcgc gatctacca ctgagatctt	6240
gcgcgggatt tcgtttacca ttaatgaagg cgaattcgtc tcgattatgg gccgtccgg	6300
ttcgggcaaa tcaacctct tgcacgttct cggattcttg gctgatcgca ccgccgtac	6360
ttaccgttc aacggcaagc aatttgccga acataccgat gaggagatcg gcgggtacg	6420
caatgaagaa atggggttcg tcttcagac ttcaactta cttggtcgta ataccgtctt	6480
cgaaaatgtg cgttgccgc tcatctactc gcgcgtgcc gaaggagagt ggccgcctt	6540
ggttgatcag gctatcgccc aagttaagct tgatcatcgg cgcgactatg cctgctcaa	6600
gctctccggc ggcgagcaac aacgcgtcgc catcgctcgc gccttggta accgaccaa	6660
cgtcctcttc gccgacgaac cgaccggcaa cttagactcc gcttcggggg gagcggtgat	6720

ggatacttta caacacttgc atgaagattc tggtcagacg gtgatcttaa tcactcacga 6780
 gacctatacc gccgagcatg ctcagcggat catcaagatt ttggatggcc gggtcgaagc 6840
 cgatttcaga cttgagacca gacgacgcgc cagcaggggt tatcataagt agttcgattt 6900
 aatttatcct gagggtaatc gaaggactca ccacaagtaa aatgcaacgt tacaaattta 6960

 gcttcctttc ggccttggag gcgatcaaaa ccaatcgtac gcgctctatc ctcaccactt 7020
 tggggatcgt tattgggggtg gcggcgatca ttgtgattat gtcgttgggc gccggcgccc 7080
 agagttaaat tttaaatgag atcaatcaga tggggggcga gacggtcacg gtgttgccgg 7140
 gtgagatcac tgatgccgcg gcggttttct cggactcact gacgcaacgt gacctggccg 7200
 cggatgaaggt taagtccaat gtgcccaatt tggcgcgcg gcgcccggcg gtcacgtcc 7260
 caggcaagac cacttataga ggtacgactt atacccccgc catgattatc ggcactgaag 7320
 cggaattctt cggtagaggtt tttaatattt accctaaggt gggcacaatc tatgatcaag 7380

 atgatatcga gacagcggcg cgggtggcga ttattggcga caaggttaag accgagcttt 7440
 ttggcgcttc tgacgcgggtg ggcgagcgga tcgatatcaa gggcaagcaa ttccgcgtgg 7500
 tgggggtgta tccaacgacg gggcaaaaag gacctttcga tatcgacggc ttggtgatga 7560
 ttccgcacac caccgccag acttatctct taggcactaa ctattatcat cgccttatga 7620
 ctcaagccga cagttcggac aatgtcgaga aattggcaca cgacatcacc gcgacctgc 7680
 gggagactca tggcttttat cctggtgatg acgacgactt ctcggtggta actcaacaag 7740
 cgctggtgga tcaaatttcg atcattatca acattctcac ggccttcttg gcggccgtgg 7800

 tggcgatctc cttggtggtg ggcggtatcg gcgtgatgaa tattatgctc gtgtcgggta 7860
 ccgaacgcac taaagagatt ggtttgcgca aggcgctcgg ggcgacccgc tcggccatta 7920
 tgacgcaatt tctctttgag gcgattgcgc tgaccttggt tgggggcgtg ctggggatca 7980
 tgatcggcgc ctgctctcg ctcgtgctct cggggattct cacttacgcc gtggggctca 8040
 attggtcctt ccacttcccc attagcgccg cgatgctcgg ggtcacggtc tcggcgccgg 8100
 tcggactggt gtttggcctc tatccggcgc gtcgcgccgc cgccaagac ccaatcgaag 8160
 cgttcgggta tgaatagaac cggggaggtt tgacgtgact attgattagt gttagactat 8220

 tgaaggaagt taatttgatt ttttgttcga aacaaagaaa aaaagaagga ggttaccatg 8280
 tcggataaaa tcgtgagatt gcctcacctt aaagtttggc aacgagatcg gtgttggtgg 8340
 ggacaattac tcttactga tcgctcgatg agcgaagagt tcaacggcaa gttcttggcc 8400
 ttggtcgctc tgcttgaagc ccaagagcga aaaagtgttg ttaatgaaga catcctcgat 8460
 ctacttgatc agattgggaa atccccattg tcggagacag attgtcttcg gctacgacgt 8520
 gacggtcatg ataaggtaga tgtggttctg gttaaaatta tgagaaattg ggtccgcgac 8580

tcggctcaaa atgagcgacg tgaatttgag ctcgtaagtt ttaaaaccac cattatgtcc	8640
aaacaggcgg cgaaagccac cttcaactga aatTTTTctc gcctgcgaat ctccaagcag	8700
accggtcgga gcacgtgttg ctcgggccct ttatTTTTaa taaatatttg cccgaggatt	8760
gttttctcaa attctctttt ttctttaagt cggggTTTT ggactgaaac ggaagagttg	8820
taatctagaa actcactttt ttggatggt ttttcaacaa atagctgtta caatagaaga	8880
gtggaaaaat aaaatgagtt gttttaaac aggtacgggt aatcgaaagc tcagcacaat	8940
tccgggtttt accttgattg aaatcttggg ggtggttgcc attatcggtt tttgtcggg	9000
aataatttcg aataatttaa ggggtgctaa aattaaagcc cgagaagcct cggcccttca	9060
aaatgcgcgg caattagatt tggcggatc gctttttgaa atagataaag gttattatcc	9120
gggaaccctg ggggttgaga caaatcaaga tgaccaaagc actggttgga aagaaggacc	9180
aggaaccctg cagcagatc tggttcccaa atatatttct aaattacca cgagtgatga	9240
gataaagttt atttatcttg ccatgaacc atgtccaac gaccagacga aacctgtcg	9300
agctaagata gttatcgata ctgaccaa atgtcgatggt gacggaggga cccccacc	9360
ccccaccga cccccacc cagctaaggt gattgttccg gacttggtta ataaaaccga	9420
agccgaagcc ctcggggcca tctcgcggc taatttagca gtaggcttca atgatgatgg	9480
gtgtagtgat atggtttctt ctggttatgt tttttctcaa tcgttgacgg ccggtgctag	9540
tgttgatgaa ggtacggcga ttaatattgt tgtttctgcc ggagggtgta tttctccgc	9600
accggtcggg tcgattccta tctcaagttg tggcacaata ataactcaac ctggagatta	9660
ccatctggcc ctggtggagg agaccgagtt gaatcaact aattccggga tctgtattta	9720
tgttaacaat gtigataatg ttaatttaga ctgtcagaat ataagataa agggtagcga	9780
taccacagag tcatcgaac aatatggcgt aattgtcggg aattcgtctg gggtagccgt	9840
taaaaattgt ctgattgaaa acgtcggcac cggaattagg gtatattcgt ctgataacat	9900
ctcgattgaa aacaatcgac tgtcaactt aggcaggga gggatgtatc ttaaagataa	9960
ttcagatgtg attattcgaa ataatcagct gaccaacgcc ggtgcaagag cgattgctat	10020
ttatcgagaa tggcgagtc ttatttccgg ttacgtgtt gataataaca ccatcaaggg	10080
ggggtcctat ggtattacgt tcgggcatct gttaccgac agtcgtctc ccggtgagat	10140
taaagagatc gttataaac gcaataattt atatgatatt gtcactacgg ctctatcctt	10200
aaatttagtc gagaacctct caatcattaa taattacatt tatgacccga aaatattcct	10260
ccaaatagac gattctaaaa atttactcat agacaacaac ttcggccaaa atatcacctg	10320

ggacatgttt atcggttatt cagataatgt aaccttttct aacaataagc ttaagagcgc	10380
ttcggcgact aaatcggtgg ttttagtttg gatgtttagg gttaataact tagatttctc	10440
tcgcaacgaa attgaaggct acaatcgtaa tttgttaaaa cttgacgata gttatgattt	10500
ctcgatcaaa aataatatit tcaatagccg ggttgggtgt tatgaagggg tgattttggg	10560
taaaggtttt ctcggtgtat ctggtgaagt ttctgaaaat gattttttacg gcggtggcga	10620
gggcgtctct ttagcttttag atatttatca taattcggcc aaccgtctgg cgatctttaa	10680
taataatitit attgattatt tgggggcgtc gttaagatat gattctagtt ttttggattt	10740
aggagctaata tattatggta caaccgactg tgccttattg cgggcgacaa ctggcccga	10800
ctgggtgata ataccacctt ctctcggttt acccagtcct ttgctttact tggattcgtt	10860
ttggcctaaa gggaacgttc aaacttgcaa ttaatttagg ctaaactgcg agtgagggtg	10920
ttttcttgat atttagatta aaaagtata taagtataaa agagaaagga ggttctgatg	10980
tctcaaatgg gtattgccca cgcgtctttt tacaagcgag gggattgtct ccaagctcgg	11040
atcgttttcg gcgacggctg cttgagcgaa gagttcagct cccgtctcga agggatggag	11100
attctgacaa aatctcgta ggataagctc atttctcacc aagagatgac ctctctggcg	11160
ttggaatttg cggaatcgac ttgcccggcg agaactccgt cggcggaaat tgttgacggc	11220
cttctgatgg cgatgaagct tgacctttga aagctttatc aaaaccgctc tccggctgat	11280
ctcggggcgg tttttttgtt taaatttaaa gggatggagt tatttcgagc gggggatgcg	11340
atgcttctga tgagtgaagt tggcgttgaa gtttgacttg aagttttgat tgttcggccc	11400
gccccatttc tgaacttga agactgacgg ggtgcggaaa ccggcgttgc ccggttgttg	11460
ctgtttgttt tgtttgtcc gattgggtgt ttccatatcc ttttaattata aatcgaagtt	11520
ggattatggc aagcagtaag ataaacgtcc taattgtgac gtgattgaca gaaaagataa	11580
aacaatgtag gatagatttc ggatcctgaa ccttcaactc tectcaacag aatcaacaga	11640
aaggaagaca gaatgaagaa gatgcttgtc ttgttgctcg cgtttgtctt gaccatcgcc	11700
gagctggctt cgcccgatc gttctctgac ccgttcgatg cccttgatc ggcttgggtg	11760
accgatcggt tcgagccggc cggattctcc agcgtcgtct tcgacggcga caatcggttg	11820
gagattgcga ttagcgcgac cgactcggag gctaactgct cggccgggtt cactagtggg	11880
ttttataaca cgcaaggccg tcaacgagat gccttgatgg cggaaccttg ggtcatctcc	11940
ggcgatcttt acitgtcgtt ggatatgctc ttgggcgaca atttgcgcgg gactgatctc	12000
tgggcgcgaa cttcggacgg tccggaggct aatgcgcaat acccgattat cgggatgcgt	12060
cggtttgacc cgtttgatcc ctccaaccg ctggcgggtg atattgcctc aacttggcga	12120
gtctgggatt cggacacggt cgacggttgg gtcaatttgg ccacccgat ggtggctggt	12180

tggaacacgc tticgattga gagtgcggt ctatcatatc tctatcggat caacggggtt	12240
gaggtctatg aggacctcac catcagcgct ttgcgcaccg atctgaccac ggtctttctc	12300
caaggttata acttcggcgg tgactacgaa gtctattggg acaatgtctc tgccgccacc	12360
ttggctccgg tgcccagacc ggccacgac ttgcttttaa tgcgggggc cggcgtggtg	12420
gcgattcgtc gtcatttcgc gaaacaacaa taactaactt gagaggttag ggtccgcaa	12480
cccgttcgtc gtcgcgagcg ggttttttta ttggcgagaa gttaaggggt gatgtttagt	12540
tgaccaaggt aatagcgaag ggtgtagagc caatcctcgt cttcttcgcc ggcttccagt	12600
ttttgtttca gaagccattc gagataaccg cgatcggtct tggccacttc ggcgagcgtt	12660
cggcttttat gcttgccaaa accgaatttt ttgaagagtg acggacgaga cgagatctca	12720
atcattttgg cgagcgtttc ttctgcggag agttcgcgcg aaccaagag cgagccgtcg	12780
ccggctttca atttttgcca taaccgatta aacagcgctt cggtcaccaa aacatcgccc	12840
acggcgtcat gagcggtgcc atcaagatcc aagtcgagat aataacgcaa gaattgcaga	12900
ttgtattccg gaatcacccc ttccgtatcc agttcgcgag ccaagcgcag ggtgcagata	12960
tattgcggca ctttgactcc ttccggcgcc aagatagcga tgtcgaattt ggcattgtgc	13020
gccaccaaca cgtgatcagc gagaaggggt tccagctcgc gacggaaggc gctctcggcg	13080
aagggttctt tgcggccac cagcttattg gtgatgtgag tgatactcat cgacttaacc	13140
gagatgggga ctggcggctt gaagtaggcg gtgcgagtgg tggttttggt ttgttagcag	13200
acctgacaaa ggcgatcttt ggtcacgtcg ttgccggtgg ttccggtatc taagaataag	13260
atttccatgg tcggttaagc ggccggttgg tcggtcgaat caaccttaac gttttggata	13320
attacgggcg tgacggggcg atcgttttgg tcagtggcga cttggccgat ctggtttaca	13380
atttcttgtc caacagttac ccgaccgaag atggtgtagt tattgggtag cggataatct	13440
tcgagcatga taaagaattg actgccgttg gtattgggac cggcgttggc catcgccaac	13500
acgccttgcc ggtagccggc ctggtatgac ggagtggccg gatcgagctc gtcggcgaat	13560
tggttaaccg ggccgccgtt accgcagggg ccggtggcgg ggactttgac ggattcaggt	13620
gaacagttcg ggtcgccgcc ttggatcata aaaccttga tcaactgatg gaaggtgaga	13680
ccgttgtaat aaccggctcg ggccagcttg ataaagtgg caaccgtgtt gggggcgtct	13740
ttttcgtaga gaacaggggt aatctcgcca agattggttt gcaaggtgat ttggttaggc	13800
atagttgagg tggtcagtc cgagcttgct cgcggtgagt tcgtcgaatc cgtcgaggtg	13860
gcttgagatt gataaatgtt acttgttaaa tcggcaggat tgggcgctct ctgatttaac	13920
ttttgccaac caaaaagtcc agccaggccg agtaaaataa taagaactaa aatcacctgt	13980
ttgttcatgg gaattgagaa acgggttaaa gatgggctga taattgtgaa ttataacaat	14040

aaccgttaga gtaaggcaat gaagagtga gaaccggaag attatcggct aggttggcgg	14100
cccttcttgg gttgccaagt ggatctctct cagcgaccgt tgattccgcg cgaggagacg	14160
gaattctggg ttgatcaagc aatcaaggaa cttaaaccag aatcaaccgc cggcaaacaa	14220
gtcttggact tgtttgcggg ttccggttgc atcggcttgg cggtgcttga gactgtccg	14280
ggcgtggcgg tgactttcgg cgaaagggag gaaaaatfff gtgggcagat tcggaagaac	14340
ctcaagttaa acccgccagc cagatttgat ttcccgcag accttcgggc ggcctctcaa	14400
ggtctggcgg gtggaaggac catggcctct caaggtctgg cgggtgaaag gaccatggcc	14460
tctcaaggtc tggcggggcg aattagagtc gagtctcgg gaaaggttgt ccaaaccgac	14520
atTTTTTcca aaatcaaagg gcagtttgat tttatTTTcg ccaaccgcc ttatgtcgcg	14580
accagaagaa gtcgggttca agcctcgggt gcgactggg agccggccgg agcgtctttt	14640
gcccggcccc acggtttggc ggtgattcga ccgtttttgg ttgaagcgaa aaaacgtttg	14700
cacccgggtg gccggattta tttggaattc ggttacggcc aaaaaggcgc tctggaagag	14760
ttattgcggc aaaacggata taaaggttgg tcgtttcggc gcgaccagtt tggccgctgg	14820
cgttgggtcg tgatacaata gcggtatcaa aagttaatff tttaattcta aaatTTtatg	14880
acagacaaaa acaaagcttt cattctctgg ttcaatgatt tgacaattgg cgacgtcgg	14940
ttggttggcg gcaagaacgc cgctttgggc gaaatggtca acaacctggt tccgcttga	15000
gttaatgtgc cgaatggttt cgcgattacg gcgcacgctt acgcctactt cttagacaag	15060
acaggcttaa aacagaggat taaggaaatt ttgaccgatc tcaatactca caatatcaac	15120
gatttgcaaa aacgcggcgc ccaagtccgc gccgcgatta ttaaagaaga attgccggaa	15180
gaactgcaag tggagattat caacgcttat cgcaagctta gcgccaacta tcacagccag	15240
gccgtggatg tggcgggtcg gtcttccgcc acggccgagg atttccccg ggcctcgttt	15300
gccggtcaac aagaaactta tcttaatgtc gccagcgaaa aggagttgat gttgtcgg	15360
cgcaagtgtc tcgcctcgt cttaccaat cgcccatct cttatcgggt tgataagggt	15420
ttctcaatgt ttgatgtttt gctttcggtc ggggtacaga agatggtgcg cagcgatttg	15480
gccgcggccg gcgtgatgtt ttcggtcgac accgaaaccg gtttcgataa ggtggtggtg	15540
atcaacggtg cctacggttt gggcgagatg gtggtcttgg gcaaagtcac tcccgatgaa	15600
ttcgtggtct tcaagccgc gctggagcgc ggttatcagg cgattctctc caagacgctt	15660
ggtcgcaagg acgtgaagtt ggtttacggc gccaaaggca ccaaacaggt gtcgggtccg	15720
gccaaagagg tgaaccgttt ttgtctcaaa gacgaggagg tttccaaact ggccgcttgg	15780

ggcctgacca ttgagaaata tttttccggc aaacacaatc gctatcaacc gatggatatg 15840
 gagtgggcca aggacggcaa gaccggcgaa ctctttattg ttcaagctcg ccccagacg 15900
 gtccacgccg aagccgacaa gaatgtttac gaagagcata ttttgaaaga gaaaggcaag 15960
 gagtgtgttc gtggcaacgc catcggcgcc aagatcactg ccggcaaagt gcgcctgatc 16020
 aagagcgcca accagatgaa cacttcaag ccggcgaga tcttggttac cgagatcacc 16080
 gatccggatt gggaaaccgat tatgaagatc gcggcgcgga ttatcaccga gaaggcgagg 16140
 cggaccagtc atcgggccat tgtctcgct gagcttggag tgcctccat cgtgggcacg 16200

 ggcaacgcca ccaaggtgct aaaaaacggc cagctggtga ccgtggattg ttcctccggc 16260
 aaagaaggag tggtttacga aggcaagcct gcctttgaga aaaaagaaca tcgtctaacc 16320
 gctaccgcca agacgcgcac caagtaatg gtcaatatcg gttcacccga cgatgccttc 16380
 cgcaatttct atttgccgt ttcgggggtc ggtttaggtc ggttgaatt tatcattaat 16440
 tcttacatca aggttaccac caacgcgctc ttggattaca aagagcttaa ggccagtcgc 16500
 gatccgcgcg ccaagaaggc ggttaaggcg attgatgagt tgacggttga atacaaaaac 16560
 aagaccgatt attacgtcgg cgaattggcc gaagggttg ccaaaatcgc ggccaccttc 16620

 taccgcacg acgtgattat ccgtttctcc gatttcaaga ccaacgagta ccgactctg 16680
 atcggcgcg atctctacga gccgaagag gagaaccga tgatcggttg gcgcggcgt 16740
 tcggttatt atgateccaa tttccgtcgc gctttcgct tggaatgtcg cgctctctac 16800
 caagtgcgta gcgagatggc ctttccaac gtgatccga tgattccct ctgtcgacg 16860
 gcggaagaag gccggcaagt ggtggagatt atgaccgaag ccggtctgga ccgtcaggct 16920
 gacccttcgc tcaagattta tgtgatgtc gagattcctt ccaacgtggt ggaggccgat 16980
 gccttttttg aagtcttca cgggatgtcg atcggttcca acgacctgac ccagctgatg 17040

 ctgtgttttg atcgcatc caactgatc agccatatcg ccaacgagaa tcatccggcc 17100
 gtcaagaaga tgattaggt ggcgattaaa gcttgcggg ccaagggcaa gtatatcggc 17160
 atttgcggtc aggcgccgc cgattatccg gattttgccg attttttgt gcagaacggg 17220
 atcgggagca tctcgtcaa tccgattcg gtgattaaga cttaccctg gattgaggcg 17280
 gccgaagaga agtatccca aagataataa aaatatgaaa atcgcttttt ttgaattgga 17340
 gacttgggaa aaaaaatact tgcaagagcg aactctgccc ggcgaggtcg tttttatcga 17400
 cggaccgttg gatgagacca agttgccgga gcaaaacgat ttcgacgcca tttcggtttt 17460

 tgttaattcc attgtcggcg acaaagtgtt gggacatttt cccaatctcc agttgattgc 17520
 caccgcctcg accggttatg atcattttga cctgccaact tgcgccgctc ggggggtcaa 17580
 ggtggccaac gtgccgagtt acggcgaaga taccgtggcc gattacgcct tcgccttaat 17640

gtcactctc tcgcgaaga ttgagagag ttatgagcgt attcgcgaga cggcagttt	17700
cgatctcacc ggctgcgcg gctttgatct gaagggcaag accttggggg tgatcgccac	17760
tggctcgatc ggcaaaaacg cgatcgagat cgcgcggggc ttcaatatga atatcgtcgc	17820
ttacgacaaa ttccccgacc cggtttatgc cgaagaatg ggctatcgtt atctgtctct	17880
ggacgaggtg ctggccacgg ccgatatctt gaccttgac gtgccctacc tgccgagaa	17940
tcatcatttg atcaatgccg aaacgtggc caaatgaag tcgggggctt acctgatcaa	18000
caccgctcgc ggtggcttga ttgacaccgc ggctctgctc gtggcgctta agtcggggca	18060
aattgccga gccggtttgg acgtgctcga agaggagggc gtaatcaaag atgaggtcaa	18120
tttcttaacc aacggtcgtt tggatcaagg cgatctgaag acggtgctcg gcaatcatat	18180
tttgattgat ttgccaacg tgatcattac tccgcataat gccttcaaca cttgggaggc	18240
gctgaagcgc attttagaca ccaccgtggc gaatctggtg gcttttgaag ctggaatgcc	18300
gcaaaatttg atcagtggcg attaaggcgg ttatttgacg tttaccttg ataacggtac	18360
aataaggtea gattccgttc ggggtgagtg gaaaaacgtc ggttctagac aacggaagga	18420
gattttatgg ccagaaagtc tgccactgaa attgtttgag ctgctctgctc tgcgtgaccg	18480
acgagcttgt gttttgttta aataaaaaga tggctgaatt caatttcaa atcgaaaaga	18540
aaattgccgg ccgtctcggc cgagcgggaa caataatgac gcctcacgga gacatctcca	18600
ctccggcgtt tatcaccgtg gggaccaagg ccaccgtcaa ggcgctctcg ccggagcaag	18660
taatggcctc cggttcacgg gcggcgttgg ccaatactta ccacctctc ttggagccgg	18720
gcgcggaagc ggtggcgcg gctggcggtt tgcacgcta tatgaattgg ccggggccgc	18780
tgattaccga ttccggcggc ttccaggctt tctcgctcgg cgcggcttat gacgaggcg	18840
ggatcaataa attcctcaag ccgggcctac cctcgcgac cgcaccgaag cgaccttcgg	18900
aagaaggtec gcgggagccg aagccggcca agattgacga agacggagt acgtttcgtt	18960
cgcttttga tggcgccgaa caccgcctga cgccggagag ctcgattcaa attcaacatc	19020
aacttggcgc cgatattatt ttgtcttcg acgaatgcac ggcgcccac gccgattacg	19080
tttatcagaa ggaagccatg aatcgactc accgtgggc cgagcggagt ttggctgaac	19140
acgagcggct aaccaggct aagactcggg aaaatgcttc taaaaagtc ctcggtcctc	19200
ttcaggcttc gcttgaggcc agactttttg ataagcattt tcccagctct tattcggcct	19260
tgttcggcat cgtccaagc ggccgcttc aagacttgag ggaggcgagc gccaaattta	19320
ttgccagctt gcctttcgcc ggttttggga ttggcggttc cttcgataag accgatatgg	19380
gcacggcggc cgggtgggtc aatgcgatct tgccgaccga caaacgcgc cacctgctgg	19440
ggattggcga accggaggat atgtttgagg cgttggcgca aggggcccac actttcgatt	19500

gtgtcactcc aacgcgcttg gcgcgccatg ccactttatt gacggcgacc ggccggctca 19560

atattttgaa tgccgctcac cgtgacgac cgacatcgat cgaagccgat tgtgactgtt 19620

acgcctgccaa aaattattcg cgcgcttact tggctcacct tttccgcgcc ggtgagattt 19680

ttggcgccac ttggccacg attcacaatt tgcgctttat gaatcgtctg tcggagcaaa 19740

tgcgcgccgc gattttggcc gagcgatttt tggagttaa ggccgagtgg ctagccaaat 19800

atcaaagatg aagaaacccc cctcaacccc aaaacttttt cgtttgaaa ggccttctgc 19860

gccggcgccg gatcaaccgg cagcgattaa ggcgctgacc gaaggtcttg cagcaatct 19920

tcgtcatcaa acctgtttgg gggtagccgg ttggggcaaa acttttacca tggcgggagt 19980

gattgccgct tacaacaagc cgaccttggg gattgcccat aataaaactt tggcgccca 20040

attggcgag gagtaccgaa gttttttccc cgaccacgcg gtgcattact ttgtttctta 20100

ttacgattat tatcaaccgg aggcttacgt ggccggccagc gacacttata tcgagaaaga 20160

cgccagcatc aacgaagaga tcgaacggct tcgtcacgcc tctaccgaag cgcttctgac 20220

gcccgcgcac gtgatcattg tcgcttcggt gtctgtgac tacggtttgg gcagtccgga 20280

ggaatagcc aaaagtttta tcaattttta tcttggcggg aaaattgaac gccaaacctt 20340

gattgagaaa ctggtcagtc tttattatga gcgaatcaac gccgatctct cggccggcac 20400

ctttcgccc atcgccaatt ctgtggagat tatgccccc ggtcaacgag agatcatcaa 20460

tctcaagttg accggggacc accttgccga aattttgatc gttgacgctg tttcgcgccg 20520

agtgtgaac cagccgggag agatttcaat ttatccggt aagcacttta tcaccagcgc 20580

cgacgaacgc cagcgccca tcgctttgat taagaccgag ttggtgaga ggttgaaaga 20640

gttggttgc gccggcaaga atctggaggc cgaacgcctg aagcgccga ccaattacga 20700

tttggcgatg atcaaagaaa tcggtactg caatggcatt gagaattatt cagccacct 20760

ctcggggcgg gcggcgggcg aggcgccggc caccttgctt gattattttc ctaagacttc 20820

tttcggtcgg ccgattttt tgaccatcat tgatgagtct cacgtaacgg tgccgcagct 20880

tggcgggatg tttccggcg acgagaaccg gaagaaaaat ttggtggcct atggttttcg 20940

tctgccagc gctctggaca atcgccgct caagtttccc gagtttgaag ccgaattgg 21000

tcccactatc tataccagcg ccacccggg caaatacgag cttgaagcca gtaatccca 21060

aaaaggcggg cagatcatcg aacagattat ccggcccacc ggcttggtgg atccggcaat 21120

tgaaattaaa ccgatcgtt cgaccgcgcg ctatctcggg caaatccagg attttatcgc 21180

cgaggtgaaa aaagaaattg ctcaaggctg gcgggctatc gccacgacct taaccaaacg 21240

gatggccgaa gatttgagcg agtatattgaa aggtgagggg attaaggccg aatatttgca	21300
cagcgagatc aaaacgttgg agcggatcaa aatcctcacc gacttcgcc gcggcgagtt	21360
cgactgcttg gtcggcggtta atctcttgcg cgaaggtttg gatctgcccg aagtgtcgct	21420
gatcggcatt ttggatgctg ataaggaggg ctctcttgcg tcggaagtgg cgttgatcca	21480
gaccattggc cgggcggcgc gcaatttggc cggccgggtg attctctacg cggagacgat	21540
aaccgactcg atgaagcggg cgatggatga gacggcgcgc cggcggacca aacaactggc	21600
ttacaatcag caacatggca ttacgccggt ttcaatcgtg aagaagatta aagacatcac	21660
cgacagtttg gctaaagatc ggcaacaatc ggttaccgct ctcttggcaa tagatgaaga	21720
gctttatggt aaaaacaaga aaaaattaat cagggagaag gtcaagcaaa tgagcgaagc	21780
ggtcaagaac ctcgatttgc aaaccgccgc tctctccgc gacgaaatca agatcttga	21840
aaacgtcaag actaaggcca aatgatatcg gaggatgatg ttggcgtgac atcccccca	21900
caatttttat cccaattcat acacgaccgt gcacggatag ggatgattag gaagtctgag	21960
gcaggttgaa aaattttctc aaccaacgat ctttttcgat ttgggtgact tccagatata	22020
aaatttcatt tccgattcgg taattggctt taatcatcgc gacaatttcg cggcaatcat	22080
aaggcgaaac ccagacgctg ttttgcaatc tgactaagcc aaggtgggtg aaccaacgac	22140
gaagtttgtc tcgggtgctt cgcttcatt ccttaatc aaagatgatg attcgatatt	22200
tcgggtccca tttagcaggt ttttttatgg tcaacttctt taactggtat tctcttaatc	22260
tcgcttgacc ttttttagtt aaacgaacaa ttttttgatt ttgatgattg gtttgaatct	22320
caagcaacc ttggttcttc attttctcta ttaccgtatt ggtgtaatat tttttcttg	22380
attgttgtcc gggcaaatat tttagcagtt gaacgcagtt gggggccaac aaggtaaaag	22440
caatcacccc ggtgataccg atgatactta aaataagctc ttgataatcc gctttgtcta	22500
ttcgtgacat ataccttatt ataaacggtc gtataagata agggaagata gaaaagatag	22560
gaaaagaggg aatccctcaa agcttttttg ttgggtcgg atgtgttata atcgctaggt	22620
tcctatggg cgggccacg gggggtttcg gcgtcatccg gaataagatt aagaaatttt	22680
tatggatcag aaacatcagg ataaaaatcaa aatcaaaggg gcgcggacgc acaacctgaa	22740
gaatatcagt ttggagattc cgcgcatca actcacggtg attaccggtt tatcgggctc	22800
gggcaagtct agcttggctt tcgacactat ttttgccgaa ggccagcgac gctatattga	22860
gtcactttca gcttacgcgc gccaattttt gaaacaatta cccaaaccgg aggtggacga	22920
gatctctggt ctctgccgg cgattgccat tgaccagaaa tcgcttcgc acaatccgcg	22980
ctcaaccgtg gcgaccgtga ccgagatcta cgattatctg cgctgtctct acgcgcggat	23040
cggccggccg cactgtccgg tgtgtggagt ggcgattgag aaactctcgc tggaggaaat	23100

cgtgaatttc gccaaagaga aaattgccgt cagtcacggt ggtaaaaaaa atctcaagat	23160
ttcaattacc gcgcccttgg tgcgcggacg gaaaggggag tattatcagc tcctctacga	23220
tttactggac aagggttacc tcgaagtgtt ggtggacggt caaacttatc aactgcgcga	23280
acgcacgta atgaccaaga ccaagaagca tgatattgac gccgtggtcg acatgattga	23340
ttggagcgat cagggcgagg ttgtcgcggc cgccacggt ttggccgagg cgggtgaacg	23400
ggcgctcaaa gattcggacg gtctagttaa gattgtgatt gataacgaga acttctctgt	23460
ttcttccaaa ttttcttggc ccaacgatgg ctctctttt cccgagattg aaccgcgact	23520
cttctcttc aattgcctt acggcgcttg tccacttgt cacggtattg gcaccaagca	23580
cctcttcggt ggccaacctt gcgatacttg ccaaggggct cgcttgcgtc gggaggcctt	23640
ggaggtgaga attggcgga aaaacattat ggaagcgggtg tcgtctcaa ttgccgacgc	23700
ggccagcttt ttgcacaagc tgaagttagc ccgaaagag aaaacaattt ccgaggtgct	23760
gtggcgcgag atcaaggcgc gattgaagt tttgctcgat gtgggttttg attacgtgga	23820
gttgaatcg cgcccgaca cgctctcggg cggtagaggc caacgcatcc gcctggcttc	23880
gcagtgggg tcggtttgg tcggcacgt ctactgtctt gatgaacca cgattggttt	23940
gcatgtctgc gataacgcca aactgattaa gactttgctt gatttgcgcg atttgggcaa	24000
caccattgtg gtggtggagc acgacgaaga cacaattttt gcctctgatt atttgttga	24060
tatcgccctt ggggcggggg tgcacggggg caagtggtg gccgccggtc caaccgagaa	24120
atttttaacc agcaagaaga acgattataa ttctttgacg attgattacc ttcggggcga	24180
caagactatc gctttgccgg aaaaacggcg aggaaccag aaggcgcgcg tgaatttcg	24240
cgggggcaaa atttttaaca tcaagaatct caatgtggac ctgccgtct cgcgcttgg	24300
ggcgattacc ggcgtgtcgg gttcgggcaa atctctttc gtctacgaaa ttctttataa	24360
aaatttgcag gccaaactgg agcgtcgtta tcgcaccaac acctgttta attgtcgga	24420
atttggcgga acggaatact tgagccgag gtcttagtg gatcagtcac cgatcggtcg	24480
gaccccgcg tcgaatccgg ccacttatac cgcgcccttc acctcatcc gggaactttt	24540
tgcggttcg gtcttgccc gggcgcgcg ctggaagccg tctcgcttct cttcaacgt	24600
ggctggcggc cgggtcgagg cctgccaagg taacggcgaa gtggcgggtg agatgcattt	24660
cttacctacc atctttgttc cttgcgatgt ttgcggcggc aaacgctacg agaaggaaac	24720
tctggaagcg ctctataaag gaaaaaatat ttacgaagtg ttgcagatga cgggtgaaga	24780
agcctttagt ttttcgaag atattccggc catcttcgac cggctcaaaa cgttgaacga	24840
agtcggtttg ggttatgttg aattgggtca atcgccacc acctctcgg gaggcgagc	24900
ccaacgggtc aaaatctcca ctgaacttta tcggccgttt accgaacga cgatttatat	24960

cttggacgaa ccaacggtcg gattgcatta cgaagatgtt aaaaaccta acgaaat	25020
gcaaaaattg gtgaccaaag gcaataccgt ggtgggtgatt gagcataatt tggaagtgg	25080
caagagcgcc gattacgtga ttgatctcgg gcccgccggc ggcaaagacg gcggcgagtt	25140
ggtggcggtc ggaacgccgg aagaattggc ctacgctcct ggctcccata cggggaata	25200
tctcaagcgt ctgttgaac aacaataatt aaagttagaa gatggaaagc cgggagctta	25260
aaaaatatca attgccgat gggcccgggg tctacttctt caagcagggc cggcgaatcc	25320
tttatgtggg caaagccacg tcgctcaagg atcgggtgcg cagttat	25380
tgggcgaaac gcgcggacca aaaattgagc ggatgcttga gttggccaac cgcgtggact	25440
ggcaaaccac ggactcgggtg ttggaagcgc tcttgctgga gtcggccttg atcaagaa	25500
atcaaccgcc ctataacacc agagaaaaag atgacaagag ctactggttc gtggtgatta	25560
ctcacgaacc ttttcccgga gtattgttgt gtcggggccg gcaattgtcg aacggttcat	25620
tctctcttgc gcttaaaatc aaaaaat	25680
aggccgcctt gctcgtgac cgaaaaat	25740
tggccggcgg acctgtttt aatcgtcagc tcggactctg ccccggggtg tgcaccggcg	25800
aaattaacca aaccgattat cggcggctga ttgccaacat tgaacgcttg tttgccgggc	25860
gtaaaagga attgctcgtt cgtctggaac gcgccatgaa acgagcggcc agaactcaac	25920
gtttcgaagc ggcggttcaa attcgcaatc aaattttcgc cctcaaacat attcaagatt	25980
tggcgttgtt gaaatcaagc cccaaccgcc tcaagggaat atccgttcgg atcgaggctt	26040
acgatgtggc tcattggcaa ggcgaggccg cgggtgggagc catggcggtt tggcaagacg	26100
gagagttgga tcgaagtcag ttccccaat tcaaacttcg ggcgacaacg cggggggacg	26160
atttgccggg gttgcgcgaa atcttgactc gacgtctggg tcatcgggag tggcccgagc	26220
cctctctggt ggtggtggat ggagaccagc gacaggtcgc cacggcccaa gtcgcattgg	26280
ctcgtcaagg tcttgactgg ccggtagtgc gactgaccaa agaccgtcat caccgcgccg	26340
tcgctttggc gggcaatctt gaggcagaga gttttgaccg tcaagccgtg attgaagtca	26400
acgacgcggc tcatcgcgtg gccattgtc atcatcggc acgtttgcgt ttgggtcgg	26460
aaggtcaggc cttatccctt ggagcgtct tccgaaat atgtaaaata aggtcggata	26520
atcaacttta tgttttggtc tgacttagtc gcaaagtgc ccaccgagcc ctcggtttgg	26580
attgcccggt tgggtttgtt tgggttcgcc tttttccttg gttat	26640
tcgaccagga cgagatggca ggtcaagcag gagatgttga agaaccagca gattattgaa	26700

ctggaaaaag tcaaccagaa cttggcggcc aaaaatcgtg aactctatgc caaagaattg	26760
gagctgacca tcgccaacaa acatctccaa gcgctggaag cagccaaatc caaatattac	26820
gccgtgacca ctcaccaatt gcgcacgccg ctctcggtg tgaagtggac gctggatttg	26880
gcggccaaag gtcaattggg caaggctgac gaagagcaaa aaagtittctt aaacaaaggc	26940
ttgattagt tcaaccgggt tattgccatc gtgaacgaac tcttgcgcgt ggactcgggtg	27000
gagaccgatc aagtcgtcta ttgtttccaa cccgtcaatt ttatcaagct gttcgacgaa	27060
gtgttgtttg aattcgaagt gcaggccaag agcaaagggg tgaaactctc ggtgcgtcgg	27120
ccggagactg acctgctcc aattgatttg gatgaaacca agattaaaat ggtgatggaa	27180
aatcttttcg acaacgccat taaatacacg ccggtgggcg gtctgggtga agtggttgtc	27240
tccgacaagc gtctcaaccg cgccgaaggg gcgattgagg tgacgggtgcg cgattccggc	27300
atcggtatcc cgagcgagga aaagaacaac attttccaaa aatttttcg cgcgaccaac	27360
gcgatcaagg ccgagcccga cggttccggt ctcggtctct ttatcgctca cgatattgtg	27420
actcggcata atggctcaat gtggtttgag ccggccgcgg gcggaggcac gattttacc	27480
ttcactttac cgattcatca gaagacgcta taattttaaa gactcttacc aatttaattc	27540
taaaagacaa tggacaagaa aaaaatccia atcgtggagg acgacgagtt cctccgttcc	27600
ctcaacgcca agaagctgga gagcgagggt tatgccgtta gtgtgtcgcc cgacgggacc	27660
agcgcatcg aattgattcc tgaagaattg ccgacttgg tgtttctgga tcttctgttg	27720
ccgggcggca aagacggttt cgatgtttta acggcgatca aggccgacga aaaaaccaag	27780
aatatccgg tcgtggtttt ctccaatctc ggccaagccg aggatatcaa gaaggctaag	27840
gacttgggcg cgattgactt ttgatcaaa gccaaactta cccttgacga cgtggtgacg	27900
aaaattaaag aaattttgaa ataaaacaaa tcaatggcgc ccattcgagt cggatatctg	27960
cgcggtggca tcggatccga gtatgaagtt tcgcttcgaa ccggcgccgg tgttttgcgc	28020
cacttgccgg gcgacaagta tcagccggtg gatattttgc tgtctcgaga cggggcgtgg	28080
tatgccggcg gtttgcgcgc caccgccgag cgggcggtac ggggagtcga tgtgatcttc	28140
aacgccttgc acggcgagtt cggcgaagac ggtcaagcgc aacaactgct tgattatctg	28200
ttcaagccct atactggttc cggcgcggtc gccagcgtc tggggatgga taagcctcga	28260
gccaaagagc tcttcggca ggctggtctg cgggtgccc aacggcgcggt gcttcggcga	28320
gcggatcgtc ccgaggaaac cgatgccgag gcggtggctt acgatgtctt caaaaaatt	28380
ccgccgcctt ggatcgtgaa gccggccagc ggtggctcct cggtgatct ccggctggcg	28440
cgccattacc ccgagttagt ggcgcggtg gccgccggcc ttaagcagaa cgatcgaatc	28500
ttggttgagg aatcgtgcg cggtaagaa gccacggtgg gggtcgtcga tcgtctgcgc	28560

ggccgcgac attatccgtt gttgccggtt gagattgtca cgctgccaga caaggctttg	28620
tttgattacg aagcgaagta cggcggccaa accaaagaaa ttggccccgg ccgctttcgg	28680
ccggaagaca agcttgagtt ggaacgtcaa gccgttttga ttcatacaaa attaggcctg	28740
cgctactatt ctcgttccga ttttatcatc tcgcctcgcg gtatctacgt gctggaagtc	28800
aacactttgc cggcctgac cgaagagtct ctggtgcccc aggcgctggc cgctgccggc	28860
atcgcttacc cgcagttttt ggatcacttg gtgaccttgg cgttagaacg acgctgaatt	28920
tgaaggacaa aaaagccccg cgagagaaga tgcagtgtac tcaagggggc aagaggaggg	28980
gatgaaaggt atgaaggaac taccaatgaa ggggatggaa ctgggacaaa agaacaatt	29040
aggtggcaga gccttcagt ccactcgaaa gctctgccgg ttaggggtga aaggtcgagc	29100
gagcgaccta tcctcaggtt atcataaggt gtgatttttt gcaagggcgg agggattatc	29160
ttggtggtgt tattataata gcatttgctc gaacttattt tcaagacaaa atgaaggact	29220
gaacgccccg ccaccgcct cgcgacttg gcggacacca gaaacaaaa attttcttaa	29280
cattttccga ttggcgcca ggaagaattt ctcttaaatg gaaaagaaaa tttgtttct	29340
ggtgtttctg cctcaagtc tcgggcagtt ggcggggctt cagaaattcg gacagaaaat	29400
taaaaagtgt catcccccc aaacccaac cactttttta tttctgatt cctacaatgt	29460
ttcgtttggg ggtgttattt tagcatttgc tcgaacttat ttccaagaaa aaatgaaaga	29520
ctagcgttcc ccgcgcgtg aagcgctct gtgcaaagca cnnnnnnnn nnnggggatt	29580
ttgaattttg tccgcgcgga ggcagggtct gggagggaat ccgcgcgggc tttatttttt	29640
tgaatttttt tggcgtagag cttgtataaa atacaattat atggtataaa aatagtaaga	29700
gaaagtcac gtggctttct caaaaccgct cattgacaac taaaaagga ggatccaatg	29760
attatttcat tcagtgggcc ctccggtatc ggtaagggtc tcatcaaaga acgactatta	29820
cagctttatc cagacatcca agaattggtg tggatataca ctcgcacctt gcgacaaaac	29880
gaacaagggt caaacagaat tcaagtttca cttccgagt ttaaccagtc ggttgaactt	29940
ggcaagctta ctttagtgca agatcttttt ggtcatcggt atggtctaaa aaaagaagat	30000
ctcgtaacga gttcgggtat caagttgact gagttgcac cagcaaatct agtgaagca	30060
ctcaaaatca acccgaagat ttttgcaatt ggtctttaa cttctgattt atcactactt	30120
cgtaaaagac ttactgttgt gagaaagacg gaaagcgaag cagagataga gaaaagagtt	30180
acgaaagcta aaagcgagat cgagataatt ctacaacaca ggtcttttta tgcttccgtg	30240
attgaaatta cagaagctga agaagatcaa gtgttcaaca aggttcatgc aatattgcaa	30300
tcacaaatca aaccgaaagg aggaaaaaat gaaactagaa acacaagttg gtagtctgaa	30360
gttgcacaca ccgttgttcg tggcttcagg ttacattacc gaaacaccag agttctttct	30420

gagagctcaa ccctacggct gttcgggtat cgttaccga tcaactaaac aaaatgttcc	30480
agcggaaacga tcacggatta catctccacg ctatgcagtc tttgtaatg acagcatgct	30540
taactgcgag tggggaaatg aaagaccgtg gacggattgg cgagatcatg gagtgaaca	30600
ggtaaagca attggttgc taatcatcat ttctcttctg gggcgagatt tggatagctg	30660
ttgtaatttg attcgtgcac tgcataagat cgggtgtgat gcctacgaaa tcaacatctc	30720
atgttcgcat tctggagcac tgcattggaa tctgaatgtt gatgtgcttc acctagaaca	30780
actgatgaaa agagtgcgta acattacgac gactccaatc tggatcaagt tgcgtattc	30840
aaacctgctg ttctcaatgg caaaacaagc cgaagagttt agagcagatg cgatagtgtg	30900
cacaaatagc atcgggtccag gaatgttgat cgacacaaa accgctaaac cgaaactcgg	30960
aatcaagggc ggaggcgggt gaatgacggg aaaagcaatt ttcccgatcg ctctatggtg	31020
tgtgcatcag ctttcaaaaa ccgtgagtat cctgttgtc ggttgtgtg gaattttcac	31080
cgcagacgat gtaattcaaa tgctcatggc aggtgctagt gcagttcaac tctacacagc	31140
tcctgcgctg aaaggtccta cggctcttag acgagtaaag gctggactac aaaggtttct	31200
cgatgagaat ccgaagtatg ctctcagtaa agacctcgtt ggacttacgc tcgacaaaac	31260
aggtgagcat aagttttctt caccctgtcc agtcgtgatt gaagaaaagt gcacaggatg	31320
tggaatctgt attcaatcct gtgcatttga cgcctgtca atggttcgta gtgctgatag	31380
caaagcactg gcggtcattg ccgataactg catctcatgc aacgcttgcg ttggagtatg	31440
tcctccgaaa ttcgacgcta tcaaagcatc attctaggag gtaatacaga aatgaaaaaa	31500
aacacataca tcatcgcggt tcaatgcaat gcgtgtcgaa ccctactgta tcgttacaaa	31560
aaagaaggtg gtggacatct cctcaagtgt tatgccgaca tgataatgtc ggattacact	31620
aaaggcgatc taaggtgtcc ttcttgcggt caagagtttg ctcgacatgc aatcatccac	31680
aatcgtcag cacataagat aatccgaggg agagtctttg tgaagggtca tcatggataa	31740
catcatcaca acgggtggtt tgattcaatc agaccacccg ttattttttt attttagttc	31800
aaatctgttt ttgaaataat tagatgtata gtttttataa tcaaaaatct cattagattc	31860
tttatttagt ttttctacat attcaaaaaa ttgtttttta tcaaaaatat caagactaag	31920
ttctttacaa acatttgcaa ttcttttaac caattcatcg ccattttcat taccagaggc	31980
cattttttct gcttcgtaat aataactatg tcccggtact tctaccaatg caaattcaat	32040
atccttatat tcatatactt tactttttct aacacaaaagc atacctttac caaatccaag	32100
ttcgccgaaa atttcaacca acgtgtcaaa atcgcttgt ttagtgaaaa ccgagagctc	32160

tttacgttgc tcatttcctc cccattcgcc aattttaaga ataatttcag gaattccatt	32220
ggtcactcgc aatcgatat ctttttttct atgttctacc cctccctcta gaaaagtga	32280
ataatcaatc aatactctat ttttctctga tttctttttt ccactactgt caaaaaattt	32340
taccagattc tcaaattctc cttttgataa aggtcctcgt atttcaattt ctatattttc	32400
atccatattt attgattttt taggtttata aatagttgct ttattatcat ggtcgcataa	32460
ctaccagtag gtaagtaaaa ggaaagtga attttcatth ttttttatg aagatcgta	32520
gactctaaat catgagcata catattagtg gcgaccaaga gatttctctt gttcagtttt	32580
ggttttgcta aaaaattttc tggaattagt tcaaaccctc cagcttcaca aatatgtgga	32640
cattgaaata cagcatttgt tggcaaatat aatttgcaa catttttgaa tataaatttt	32700
ttacttttag ttttttctc tatcaacaaa gatgcctgtg tattccacag aaactatta	32760
tatgcggaca caaaaaaga aacttttttt ggattcatga catcaaaaac cttttttag	32820
tctgagatat cttttgcttt tagttcagct ccttgcgtaa tattatttgt aatttttagt	32880
tgttcataag cctgtttcca attatcttct actattgcct taccaatcag atgagtatta	32940
tagggggcac caggcattcc aaatctttga ttgtcatagt aatttataaa ataaagtgt	33000
ttgtgattgt ggacataatt tgaaagatta tctgcaatcg tagaatttaa atttcttacc	33060
actattttta aagcatttcc gtgtaaagcc ctttctttta ttggtttttc cccatgacct	33120
attacaaact taatttttga aaattgattt ttaaatttgt gtttcttgtt aaatactatg	33180
atatcttttt ctttcaagat ttttttgatg gaaataagtt gttcggtaat agcatcctca	33240
tcttttaatc cttggctaca tacatcctca aatgaaagtt taaaaaatag ctttatttgt	33300
tctaaggctt caaatgttgt aaatccagat tttttagacc aaatataagt aaacttacgt	33360
ttaccttttg atataaatga tggcataaga gagacctccg tcatctgaaa gtcttcgttt	33420
atgtgtttta ttttataatc ctcatattta tccataatat aaataattta acataaataa	33480
ccttatttgt aaataattcg ccaaaaaatc ccaaaaaaca aaagcccgcg cggattccct	33540
cccagacctt gcctccgcgc ggacaaaatt caaaatcccc gccgaatttc aaaaacatta	33600
gtctcggttt tgcgaacctt tctccagaa aatagttttt gcaaaaccga gtccatattt	33660
gcatttctgc acctcgctc attctcccag attattagtg gcgaggggca gggcgtttcc	33720
ccgcacttct gtttcagcag aagctctgtg ctttgacag aggcgcttca gcgcgcgggg	33780
aacgctagtc tttcatthtt tcttggaat aagttcgagc aaatgctaaa ataacaccac	33840
caaacgaaac ttgttcggaa ttaagaaagc ggagcgattt tgcgggagcc aaaatcgcg	33900
tatcatthtt ttcaaaacct tttccgcta cggcggaagc ggtgaattcc caaagttccc	33960
ccaattgaa atcatgaaag acctcaaacc aaaatattht ctctacgca ggaaatcaac	34020

agaggatgat gaccacaaa taatg 34045

<210> 24

<211> 11142

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<220><221> misc_feature

<222> (6655)..(6659)

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

<400> 24

catcttcatt tgtatgcgta tcagagagat caaaaactat gttatcaatg atggcgcggt 60

atggttcaat gagatcgaag gcgagagcgg ggtaatcagt cgtctcgtgc aggaacccat 120

ggaagggaga gaggtgatgg tagtgaatcc atcggagaaa aattcctatt aaaaattttg 180

acatgcgatt cagcgcgttg ctggccgggt ttttaccgcg cctcatgaag gctgaatgtc 240

cgagcttctt gaaatatgcg ctccaatagc gctggcgttg gagcgcttcg tgattgcgca 300

gttcttgaat ggtcatggtg cgggagagtt tttttgcagg aggaacgata agccacgcca 360

tgctgttgaa ttttgcgttt agaatttgcc gcgcgatata ctttttgatg cgcaaatcag 420

agcgcctttac aagtgtgtga gaaagcaggt catttccatc tgcacggtta ctggcggtaa 480

tccagactgt attggttaaa tttctcctat gaataatgat aggaatttta tgacgcgctg 540

tgaattcaag tgtgctcggg gctaaggag ggctatctcc gtaaatcatg atggagagga 600

gcttggcagg attgcaggtt acttcgcctc ctttatactt aatgtgaata tttttccctt 660

tgacttcaaa tgtttcaaca taaggcgcgc aaagaggtat tttttgcgag tatgttttca 720

tgttatagaa taaagtgagt attgaaatat aaaactttat atggtaatgt aagacacata 780

attttgcaag atgtgttgca aaaaagcgat tttttgaggg gtcgccccga atatagggga 840

caaaaaggct agcatacttt tttggaacct cgaatatagg ggacaaaaag gcttatgagc 900

tgaaaaagat ccccgaaat aggggacaaa aaggcacgcc gctttcgcgt tcaaccccga 960

atatagggga caaaaaggca attaccgcat aaatcatccc cgaatatagg ggacaaaaag 1020

gcaacatgac ccacctcct ccccgaaat aggggacaaa aaggtatga gactttctgaa 1080

atcccccgga atatagggga caaaaaggct taagcccat gctttctccc cgaatatagg 1140

ggacaaaaag gctgaagtac gcaatctgca acccgaata taggggacaa aaagcatgc 1200

tgtttgtatc ttcaccccga atatagggga caaaaaggca aggatattca agcgcacccc 1260

ccgaatatag gggacaaaaa ggcttaccac acaacttatt gaccccgaat ataggggaca 1320

aaaaggctgt gagcgatgta aaccaccccg aatataggag acaaaaaggc gcgtgggtcaa 1380

tgctcgtgcc ccgaatatag gggacaaaaa ggccttttagc ttcatttaag attttaggta 1440

tttccggaca gcggcttgac cgcatcgtcc tcgcttttc ctaaaatcgc ccctcttaaa 1500

tcgcttgccct tacagacgca tgtataaaga tattttgaag attaatgtat cgcatacttt 1560

atgagtaagc gacatcctag aattagcggc gtaaaagggt accgtttgca tgcgcaacgg 1620

ctggaatata ccggcaaaag tggggcaatg cgaacgatta aatatcctct ttattcatct 1680

ccgagcgggtg gaagaacggt tccgcgcgag atagtttcag caatcaatga tgattatgta 1740

gggctgtacg gttttagtaa ttttgacgat ctgtataatg cggaaaagcg caacgaagaa 1800

aaggcttact cggttttaga tttttggtac gactgcgtcc aatacggcgc ggttttttcg 1860

tatacagcgc cgggtctttt gaaaaatgtt gccgaagtgc gcgggggaag ctacgaactt 1920

acaaaaacgc ttaaaggagc ccatttataat gatgaattgc aaattgataa agtaattaaa 1980

tttttgaata aaaaagaaat ttcgcgagca aacggatcgc ttgataaact gaagaaagac 2040

atcattgatt gcttcaaagc agaatacggg gaacgacata aagatcaatg caataaactg 2100

gctgatgata ttaaaaaatgc aaaaaaagac gcgggagctt ctttagggga gcgtcaaaaa 2160

aaattatttc gcgatttttt tggaaattca gagcagctctg aaaatgataa accgtctttt 2220

actaatccgc taaacttaac ctgctgttta ttgccttttg acacagtga taacaacaga 2280

aaccgcggcg aagttttgtt taacaagctc aaggaatatg ctcaaaaatt ggataaaaac 2340

gaagggtcgc ttgaaatgtg ggaatatatt ggcatcggga acagcggcac tgccttttct 2400

aatttttttag gagaagggtt tttgggcaga ttgcgcgaga ataaaattac agagctgaaa 2460

aaagccatga tggatattac agatgcattg cgtgggcagg aacaggaaga agagttagaa 2520

aaacgtctgc ggatacttgc cgccttacc ataaaattgc gcgagccgaa atttgacaac 2580

cactggggag ggtatcgagc tgatataaac ggcaaattat ctagctggct tcagaattac 2640

ataaatcaaa cagtcaaaat caaagaggac ttaaaggagc acaaaaagga cctgaaaaaa 2700

gcgaaagaga tgataaatag gtttggggaa agcgacacaa aggaagaggc ggttgtttca 2760

tctttgcttg aaagcattga aaaaattgtt cctgatgata gcgctgatga cgagaaaccc 2820

gatattccag ctattgctat ctatcgccgc tttctttcgg atggacgatt aacattgaat 2880

cgctttgtcc aaagagaaga tgtgcaagag gcgctgataa aagaaagatt ggaagcggag 2940

aaaaagaaaa aaccgaaaaa gcgaaaaaag aaaagtgcg ctgaagatga aaaagaaaca 3000

attgacttca aggagttatt tcctcatctt gccaaacat taaaattggg gccaaacttt	3060
tacggcgaca gtaagcgtga gctgtacaag aatataga acgccgtat ttatacagat	3120
gctctgtgga aagcagtgga aaaaatatac aaaagcgct tctcgtcgtc tctaaaaat	3180
tcattttttg atacagattt tgataaagat ttttttatta agcggcttca gaaaattttt	3240
tcggtttatc gtcggtttta tacagacaaa tggaaaccga ttgtgaaaaa ctctttcgcg	3300
ccctattgcg acatcgtctc acttcgggag aatgaagttt tgtataaacc gaaacagtcg	3360
cgcagtagaa aatctgccgc gattgataaa aacagagtgc gtctcccttc cactgaaaat	3420
atcgcaaaag ctggcattgc cctcgcgcgg gagctttcag tcgcaggatt tgactggaaa	3480
gatttgtaa aaaaagagga gcatgaagaa tacattgatc tcatagaatt gcacaaaacc	3540
gcgcttgccg ttcttcttgc cgtaacagaa acacagcttg acataagcgc gttggatttt	3600
gtagaaaatg ggacggtcaa ggattttatg aaaacgcggg acggaatct ggttttggaa	3660
gggcgtttcc ttgaaatgtt ctgcagtcga attgtgtttt cagaattgcg cgggcttgcg	3720
ggtttaatga gccgcaagga atttatcact cgctccgca ttcaactat gaacggcaaa	3780
caggcggagc ttctctacat tccgcatgaa ttccaatcgg caaaaattac aacgccaaag	3840
gaaatgagca gggcgtttct tgaccttgcg cccgcggaat ttgctacatc gcttgagcca	3900
gaatcgcttt cggagaagtc attattgaaa ttgaagcaga tgcggtacta tccgcattat	3960
tttgatgatg agcttacgcg aacaggacag gggattgatg gtggagtcgc ggaaaatgcg	4020
ttacgacttg agaagtcgcc agtaaaaaaa cgagagataa aatgcaaaca gtataaaact	4080
ttgggacgcg gacaaaataa aatagtgtta tatgtccgca gttcttatta tcagacgcaa	4140
tttttggaaat ggtttttgca tcggccgaaa aacgttcaaa ccgatgttgc ggttagcgg	4200
tcgtttctta tcgacgaaaa gaaagtaaaa actcgtgga attatgacgc gtttacagtc	4260
gcgcttgaac cagtttcggg aagcgagcgg gtctttgtct cacagccgtt tactattttt	4320
ccggaaaaaa gcgcagagga agaaggacag aggtatcttg gcatagacat cggcgaatac	4380
ggcattgcgt atactgcgt tgagataact ggcgacagtg caaagattct tgatcaaaat	4440
tttatttcag acccccagct taaaactctg cgcgaggagg tcaaaggatt aaaacttgac	4500
caaaggcgcg ggacatttgc catgccaagc acgaaaatcg cccgcatccg cgaaagcctt	4560
gtgcatagtt tgcggaaccg catacatcat ctgctgttaa agcacaaagc aaagattgtg	4620
tatgaattgg aagtgtcgcg ttttgaagag ggaaagcaaa aaattaagaa agtctacgt	4680
acgttaaaaa aagcggatgt gtattcagaa attgacgcgg ataaaaattt acaaacgaca	4740
gtatggggaa aattggccgt tgcaagcgaa atcagcgcaa gctatacaag ccagttttgt	4800
ggtgctgtga aaaaattgtg gcgggcggaa atgcaggttg acgaaacaat tacaacccaa	4860

gaactaatcg gcacagtttag agtcataaaa gggggcactc ttattgacgc gataaaggat 4920
 tttatgcgcc cgccgatttt tgacgaaaat gacactccat ttccaaaata tagagacttt 4980
 tgcgacaagc atcacatttc caaaaaaatg cgtggaaaca gctgtttgtt catttgtcca 5040
 ttctgcccgcg caaacgcgga tgctgatatt caagcaagcc aaacaattgc gcttttaagg 5100

tatgttaagg aagagaaaaa ggtagaggac tactttgaac gatttagaaa gctaaaaaac 5160
 attaaagtgc tcggacagat gaagaaaata tgatagacgt tgtttttaca ccatcgctat 5220
 tgactaggtg atctttacgt cagaacccca tcagaaattc cttaaaactcc taaaacttgt 5280
 ttgaaagcgg gagaacctgt ttttgtttgt gtagaagctt tttgagatca gcggggagag 5340
 gtattttttt gccgatgagt gggtccacta ttgcgttgaa tttcactgga tgcgcggtct 5400
 caagaaaaat gccgagagta tttttctttt tattttgagc acaatatttt ttgaggccta 5460
 aataggcaac cgcgccgtgc ggatctgcac tatagccaca gcggttatac agttcagaaa 5520

ttgccccgcg cgtttcagcg tcagtaaacg atgcgccgaa aatatctttt tgcatttcag 5580
 cgcgttcatc atgatacaga gtgcgcatac gcgcgaagtt actcggattt ccgatatcca 5640
 tggcatttga aattgttcgt attgacgggt ttggaatgaa cggtcacccg cataaatatc 5700
 gcgggacgac atcattgctg tttgtggcgg cgatgaattg tctcacagga agccccattt 5760
 tttttgcaat gagccctgcg gtgaggttgc caaaatttcc gcacggcact gaaaatacaa 5820
 gcggcgggca tacagcgaaac gagcgagctt gcgcttgggc atacgcgtaa aaataataga 5880
 atgtctcgca aataagccgc gcgatattga ttgaatttgc agaggcaagg cgcaatgttc 5940

gggcaagctc ccgatcgga aatgcttgtt ttacgagggt ttggcagtcg tcaaactgac 6000
 cgtttatctc aagcgccgtg atgtttttgc ctaagccagt aatctgtttt tcctgaatag 6060
 cacttactcc gtcttttggg tatagaatta taatgtgcac gcgctcactt tgaaaaaagc 6120
 tgtgcgccac tgccgcgccg gtgtctccgc ttgttgcggc aagaatggtt aaacatctgt 6180
 cgtcattttc caaaaaataa cacatcaatt ccgccatgaa tcgcgcgcca aaatctttaa 6240
 acgagagtgt ttggccgtga aaaagttcaa gtacagcgag cgtttcattt aaaaacacaa 6300
 gaggcgcgtc aaatgtgaga gatttttcaa taatgcgggt gatgtcttgt tttggaattt 6360

tagggaacca caactcgctt gtttcccgcg caatatcttt gagggatttt ttggcaatgc 6420
 ttttgaaaaa tgatgaagag agccgggggaa tttcaagcgg catgaacagg ccgcatccg 6480
 gcgcgagcgg ggaagagaga ccatgtttta aggaaaaaat tttattgttt ctatttgtgc 6540
 ttttaagctt catggcaggt ttgtataaaa ttctctgctg aaaattcggg cgaccgtagt 6600
 ctgtgatagg ggatggttgc gtgcgcgtat tgtttatagc gattggtgag atagnnnnnc 6660
 agttttgggt aacatcgcgc gagcgagag cgattgtttt cgttattccg cttttcaaac 6720

atattccccc acagcacggg ctttggatcg cgaaggtact gttcaaacat ttccttgcgt 6780

acttttgccg gcgtgtataa atataccaca cgcgtatatt ttttgagcag attgcataat 6840

gcggggtcaa cataaataac actccctgtc gtgtcaataa ctgtgcgaca atcaagtttt 6900

cttttttgta ttaaaccgat aatttttcgt ataacgtac gctcgcaacg caaataatgg 6960

ctttgattcg cgttgtattg ggactcgtat ggctggccaa gccatcgca tacatcttga 7020

atgcccttat agccgtgctt ttttaagcaag gaagcaagct ttttttcaat taaatcgta 7080

cagcagatat gcgcgtaccc aaagcgcgca agctgttgcg cccagtatga ttttcccgcg 7140

cctgacatgc cgataagcgc gattggtttt tcttgcacac tataatgtt cataaacgca 7200

ctgccttaaa aatatctgaa aaaactccig cggatgtcac ctctgcgcct gtcctttgc 7260

ctcgtacgat aagcgggtgt tcatggtaat gatcgggtgt aaatgaaaat atattgtcgc 7320

tcccgcggag cccggcaaac ggatgattag aggcaacttc ttttaagaaac atttttgcct 7380

tgccattttc tatttcagca acaaagcgaa gcactgcgcc gcgtgcgac gcgcgttgtt 7440

tttttgcttc aaattggcg tcgtaccgtt caagtgttt taaaaattct ttaacggttt 7500

ccttttttct gccttcgga atgagctgtt ctatttcaac atccgcgcat tccatgggga 7560

gagcgcactc tcttgcaaca atcaccaatt ttcgcgcgcg gtccatgccg ttttaagtcgt 7620

ttcgcggatc tggttccgtg taaccgagct tctgcgcctc gcgcaccgct ttgctcaatg 7680

ttgtatttcc ctcaaatgag ttaaagatat agcttagcgt tccagaaacg attgctgaaa 7740

ttttttctac gcggtcgccg cagagcatga aatctcgtat ggtggaaagc acaggaagcc 7800

ctgccccgac ggttgtttca tataaaaacc gcgtatggtt ttgagaggcg agtagtttta 7860

aatttttata gaatttaaaa ttggatgaaa ggctttttt attcggcggtt acaatggcaa 7920

tgcgctctgc aagtatggtg ttatagaggg cggaatttc ttcgctcgcg gtgcagtcca 7980

caaacacggc gtttgaagg cgcatgtcct tcatgccggc gacaaattga gcaagatcag 8040

ctttttgtcc gcgcgtgtta agctcttctt tccagccaga aagcgtgccg aggtgttccc 8100

caagaacat tttcttggtg ttgacgatgc ctgcaacttt gagcgcaata cctctcttg 8160

ccaaaagccg ctctctttga gcatigattt tcgtaagaag cgcagatccg ataagcccgc 8220

ttcccgcgag aaacacgtga atgttttgtg gtccatagg tataaaaaa ccgctccaga 8280

catgtgggta atgtccggag cggaagaagt tataatgcgc cttgttttta tttttaactc 8340

ttcacaacca aacatcaccc gccttttgcg gtaatatgtg tgatgatggt agtgatgcta 8400

ttttgacgca taagaatttt tttgactctc atagtatagc acaagtaaaa ttttttgcgc 8460

aaggttttgg tgagttgata gagttttgag gttgatatct aattgtcaag aaacggggat	8520
aatgtgcaca cattatcaca acagattgaa tatatgcggg ttttgtgaaa taatggcatt	8580
atatactttg atgaacctca ccaaacctgc caattttttc tttgaacttg gcatgatgaa	8640
acgggaaaag catcagggtt ttgctattgc gggcgtgcat cacgacatgg ggtctttagc	8700
ggatcatacg tgtcgcgcgg ctttaattgg cgcaatttta gcggaatgg aaggcgcgga	8760
cgtgaataaa gttgccatga tgggtctttt gcacgatata ccggaacgc gcattgggga	8820
tcatcataaa gttgcggcgc ggtattttga tacgaaaaa gtggaacgcg ctattttttt	8880
agaacaaatt cagtttctgc ctgacctttt gcaaaaaaaaa tggctcgcgc tctacgacga	8940
aaaagcaaag agaagcacta aagagggtat tgtcgcaaaa gacgcggact ggcttgaact	9000
ggcgattttc gcgcgtgaat acatacacat cggtataaaa gatttgcagt tgtgggttga	9060
taatgttcgg agcgcgcttg aaactgaatc cgccaaaaa cttcttgcag aaatagaaaa	9120
acaaggcacc tacgactggg cccgcggttt agaaaagatg acatatcaga aattatcgtg	9180
atctgcaatt ttttctata attataaaaa agtttcattc caacatctaa cgcaacattg	9240
aggaaaaact tcaatgcaat gatgagtatt gtgaaaaagt tgggaccagc tctctttccc	9300
attttgcagg atatgcgtct ctcgatcag gtgcatggaa aggagtaaaa aaatacacgc	9360
cgcttgcaaa tttagaagac gtacggaata gagccgttgc gattagaaaa gaagcagaca	9420
aagaaaaacc agatagtta gagattgatc gtatttttaac ggattttatg aatgcggagc	9480
taaaggaatt atggaatacc atagataaac gtattgttga tgcggcgaaa aagtttatac	9540
aaaacttcaa agatcatccc gaagacgcga ggagagcgaa ggtggagagt tggggactag	9600
aagaatggaa aagagattta gaacggatag tcaaaacccc aattaatcaa atgatggggg	9660
acgcatcatt tgtgattaac agaggagtgg atcagtatcg tgcgcgcgat atggcgaaaa	9720
ttatgggtaa gataagtgtt ttttatcaac ccttgtgtg ggagaaggcg tcataacce	9780
tgagaattat cacaaaattc tctgttcat atacaccatc gctccgtaaa gccccgagga	9840
atcgagagc tttgattttt gaatcggcgg aaaggacggg aacaggggtt gatttgattt	9900
cttgacacgc tgtgagttgg gcagtagagt agtaagaaag taatatTTTT ttatattcat	9960
gaacactaag ataatacaaa aagctacatc tcgggggaaa attacgttc caggacagtg	10020
gcgtaaaaag tttcctacga accaatatct tgttgaagtg gaagaagatt tgcttaagat	10080
taagcctttt gaagtggaca cggcggggca attagaagaa caagtaaaag tgttgaattg	10140
tgtcaataga tttgagggac ttgcgataaa aggaagaaaa tttgctaaaa agagaggaat	10200
taaaatggac gatgttttaa aagatgatta aagcagtact tgatacgaat attttaattt	10260
ccgcactttt ttggaaaggc accccatata ttattgtgca ggatggatta gaggggtgtgt	10320

ttgaaatggt tacttcaaaa gcaataatga gtgaaacgaa agagaagttg attcaaaaat	10380
ttgaattttc tgttgaagat actctaagat acttggaaact cttggtttgt aagtcgttcg	10440
ttgtatcacc gatggtacag cataatgtgg tgaaaaatga tagtactgat aataaaattc	10500
ttgagtgtgc ggtaagcgcc aacgcagatt atattgtgac aggagataaa catctactaa	10560
atatcaagca ttatcaaggg atcactattc tactgcacg cagatttgat gagatacttg	10620
aaaatgaacg gagtagaatg agaagaaata agcgataggg acagaataac ttggatccaa	10680
ccttctaacg caacagcgtt aagaatgaat taattgattg aaaacctcgt atgggtgttg	10740
aaagtcgagt gtttttctcg gtcggccatt caggagatgt tgcgctcgtt tcacttcgta	10800
ccgcgatacc ttggtaaagt tggttccttt cggaaaaaat tgtctgatga gtccattggg	10860
gttttcgttc gtgcctcgtt cccatggact ccggggatgg gcgaagtaga ctttgactcc	10920
ggtcagattc gtgaataatt tgtggctggc catttcccgc ccttggctcg atgtcatcgt	10980
cagtctcatt tgtttcggca attttttcac ttccttggca aacgctttgg ccacatcttc	11040
ggcagatttg cttttcacgg ggataaggat agtcgtgcgg gtcgtgcgct caaccagagt	11100
gccaaagacc gaacgattgt tctttccaac aatgagatcg cc	11142
<210> 25	
<211> 13879	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 25	
tttccaccgc cgctcaatca gtctagacat acaggtggaa aggtgagagt aaagacgtga	60
caaccttctc atcctcttca aagtctagac atacaggtgg aaagtgaga gtaaagacaa	120
accgtgccac actaaaccga tgagtctaga catacaggtg gaaaggtgag agtaaagact	180
caagtaacta cctgttcttt cacaagtcta gacctgcagg tggtaaagtg agagtaaaga	240
cttttatect cctctctatg cttctgagtc tagacattta ggtggaaagg tgagagtaaa	300
gacttgtgga gatccatgaa cttcggcagc ctagacctgc aggtggaaag gtgagagtaa	360
agacgtcctt cacacgatct tctctgttta gtctaggcct gcaggtggaa aggtgagagt	420
aaagacgcat aagcgtaatt gaagctctct ccggtccaga ccttgtcgcg cttgtgttgc	480
gacaaaggcg gagtccgcaa taagttcttt ttacaatgtt ttttcataa aaccgataca	540
atcaagtatc ggttttgctt tttttatgaa aatatgttat gctatgtgct caaataaaaa	600
tatcaataaa atagcgtttt ttgtataatt tatcgctaaa attatacata atcacgcaac	660

attgccattc tcacacagga gaaaagtcac gccagaaagc aagcagatgc aatgccgcaa 720
gtgcggcgca agcatgaagt atgaagtaat tggattgggc aagaagtcac gcagatatat 780
gtgccagat tgcggcaatc acaccagcgc gcgcaagatt cagaacaaga aaaagcgca 840
caaaaagtat ggatccgcaa gcaaagcgca gagccagagg atagctgtgg ctggcgcgt 900
ttatccagac aaaaaagtc agaccataaa gacctacaaa taccagcgg atcttaatgg 960
cgaagtcat gacagcggcg tcgcagagaa gattgcgcag gcgattcagg aagatgagat 1020

cggcctgctt ggcccgtcca gcgaatacgc ttgctggatt gcttcacaaa aacagagcga 1080
gccgtattca gtttagatatt tttggtttga cgcggtgtgc gcaggcggag tattcgcgta 1140
ttctggcgcg cgctgcttt ccacagtcct ccagttgagt ggcgaggaaa gcgttttgcg 1200
cgctgcttta gcatctagcc cgtttgtaga tgacattaat ttggcgcaag cggaaaagtt 1260
cctagccgtt agccggcgca caggccaaga taagctaggc aagcgattg gagaatgttt 1320
tgcggaaggc cggcttgaag cgcttggcat caaagatcgc atgcgcgaat tcgtgcaagc 1380
gattgatgtg gcccaaaccg cgggccagcg gttcgcggcc aagctaaaga tattcggcat 1440

cagtcagatg cctgaagcca agcaatggaa caatgattcc gggtcactg tatgtatatt 1500
gccggattat taigtcccgg aagaaaaccg cgcggaccag ctggttggtt tgcttcggcg 1560
cttacgcgag atcgcgtatt gcatgggaat tgaggatgaa gcaggatttg agcatctagg 1620
cattgacctt gggtctcttt ccaatttttc caatggcaat ccaaagcgag gattttctcg 1680
ccgcctgctc aataatgaca ttatagcgtt ggcaaacac atgtcagcca tgacgccgta 1740
ttgggaaggc agaaaaggcg agttgattga gcgccttgca tggcttaaac atcgcgctga 1800
aggattgtat ttgaaagagc cacatttcgg caactcctgg gcagaccacc gcagcaggat 1860

tttcagtcgc attgcgggct ggctttccgg atgcgcgggc aagctcaaga ttgccaagga 1920
tcagatttca ggcgtgcgta cggatttggt tctgctcaag cgcttcttg atgcggtacc 1980
gcaaagcgcg ccgtcgccgg actttattgc ttccatcagc gcgctggatc ggtttttgga 2040
agcggcagaa agcagccagg atccggcaga acaggtacgc gctttgtacg cgtttcatct 2100
gaacgcgcct gcggtccgat ccatcgccaa caaggcggta cagaggtctg attcccagga 2160
gtggcttacc aaggaactgg atgctgtaga tcacctgaa ttcaacaaag catttccgtt 2220
tttttcggat acaggaaaga aaaagaagaa aggagcgaat agcaacggag cgccttctga 2280

agaagaatac acggaacag aatccattca acaaccagaa gatgcagagc aggaagtga 2340
tggtcaagaa ggaatggcg cttcaaagaa ccagaaaaag tttcagcgca ttcctcgatt 2400
tttcggggaa ggttcaagga gtgagtatcg aattttaaca gaagcgccgc aatattttga 2460
catgttctgc aataatatgc gcgcgatctt tatgcagcta gagagtcagc cgcgcaaggc 2520

gcctcgtgat ttcaaatgct ttctgcagaa tcgtttgcag aagctttaca agcaaacctt 2580
tctcaatgct cgcagtaata aatgccgcgc gcttctggaa tccgtcctta tttcatgggg 2640
agaattttat acttatggcg cgaatgaaaa gaagtttcgt ctgcgccatg aagcgagcga 2700

gcgcagctcg gatccggact atgtggttca gcaggcattg gaaatcgcgc gccggctttt 2760
cttgttcggg tttgagtggc gcgattgctc tgctggagag cgcgtggatt tggttgaaat 2820
ccacaaaaaa gcaatctcat ttttgcttgc aatcactcag gccgaggttt cagttggttc 2880
ctataactgg ctigggaata gcaccgtgag ccggtatctt tcggttgctg gcacagacac 2940
attgtacggc actcaactgg aggagttttt gaacgccaca gtgctttcac agatgcgtgg 3000
gtcggcgatt cggctttcat ctcaggagtt aaaagacgga tttgatgttc agttggagag 3060
ttcgtgccag gacaatctcc agcatctgct ggtgtatcgc gcttcgcgcg acttggtctg 3120

gtgcaaacgc gctacatgcc cggctgaatt ggatccgaaa attcttgttc tgccggttgg 3180
tgcgtttatc gcgagcgtaa tgaaaatgat tgagcgtggc gatgaacat tagcaggcgc 3240
gtatttgcgt catcggccgc attcattcgg ctggcagata cgggttcgtg gagtggcgga 3300
agtaggcatg gatcagggca cagcgtagc attccagaag ccgactgaat cagagccgtt 3360
taaaataaag ccgttttccg ctcaatacgg cccagtactt tggtttaatt ctcatccta 3420
tagccagagc cagtatctgg atggattttt aagccagcca aagaattggt ctatgcgggt 3480
gctacctcaa gccggatcag tgcgcgtgga acagcgcgtt gctctgatat ggaatttgca 3540

ggcaggcaag atgcggctgg agcgtctctg agcgcgcgcg tttttcatgc cagtgccatt 3600
cagcttcagg ccgtctggtt caggagatga agcagtattg gcgccgaatc ggtacttggg 3660
actttttccg cattccggag gaatagaata cgcggtggtg gatgtattag attccgcggg 3720
tttcaaaatt cttagcgcg gtacgattgc ggtaaaatggc ttttccaga agcgcggcga 3780
acgccaagag gaggcacaca gagaaaaaca gagacgcgga attttgata taggccgcaa 3840
gaagccggtg caagctgaag ttgacgcagc caatgaattg caccgcaaat acaccgatgt 3900
tgccactcgt ttagggtgca gaattgtggt tcagtgggcg cccagccaa agccgggcac 3960

agcgcgcgacc gcgcaaacag tatacgcgcg cgcagtgcgg accgaagcgc cgcgatctgg 4020
aaatcaagag gatcatgctc gtatgaaatc ctcttgggga tatacttggg gcacctattg 4080
ggagaagcgc aaaccagagg atattttggg catctcaacc caagtatact ggaccggcgg 4140
tatagcgag tcatgtcccg cagtcgcggt tgcgcttttg gggcacatta gggcaacatc 4200
cactcaaact gaatgggaaa aagaggaggt tgtattcggt cgactgaaga agttctttcc 4260
aagctagacg atctttttaa aaactgggct gctggctatc gtatggtcag tagctcttat 4320

ttttttactt gatatatggt attatctcaa taatatgcat ctcttcatag atacaacaga	4380
aaaagaatca tttgatattg ctttgattga tgaatgagcgc gttatcaaaa agaagcgaat	4440
caaatcaatc cgccaacatt cggaagagct tttgaaatca attgacgagc ttttgttgtc	4500
cgcaaaatca tctctgaaag atatacaagg catcatcgcg gtaaaaggcc ctgggtcatt	4560
tacctcattg cgcattggaa tcgcgacagc caacgcgttg gcattcgctt tgggagtggg	4620
gattgctgga gttgacaaaa cagatgagtg gagtaagatt gtttcttcag cagatttgat	4680
ctttaaaaag caaaaaaga acttaaatat cgtcatacc gaatacggca gagagccgga	4740
cattacctaa ataggagggt ttagaaatgt tatgtctcat tttgattctc acaatagttt	4800
tgagcatcat tcttttgtgc ttttgcgctt ttattctctg cataatcaca gaagatggca	4860
gggaaatgct tttgatgttt ggaataggca aatgccactt gaattattaa agtggctttt	4920
ttatttgtac aaaaacagtg tcagagcgcc gattcggcgc tctgacactg ttttacaac	4980
cctcacccca accctctccc gaatacagga gaggaattt ttatctgtg cataacttgt	5040
gcgcaaatag tgcctagata aggggtgcgt aaaattacaa gagtgggtga taatatcatc	5100
atagtggtag ggagtgggga taagtggtag agaacctcat caataataga taccaatgtt	5160
cataggagaa tacaacata ctattgatac caaaggaaga atggcaatc ctgccaatt	5220
tcggcaggat ttgaaaaagg gcgcaatcgt aacaaaagga ttggataatt gccttttgt	5280
atacactcaa gatgaatgga aaaaactcgt ggacaagcta tctaattctc caatctcaca	5340
gcagaaaagc cgggcatttg ccagattaat gctagcagga gcaatggacg tgcaaatga	5400
ctccaaggc agaattctta taccagaata tcttcgcaa ttcgcgtcaa tcaagaaaga	5460
caccataata gcagggtttt acagtcggct tgaatatgg gattcaaaag aatgggaaaa	5520
atacaaatca gccactgaaa agataagcac aaaaatagct gaagagctca cgctctagc	5580
caaaaacaaa aataaaattc aaaacaatca cgagatcctt cgactccgag agtacgcttc	5640
gctcagagcc tgccccgagt attccgaggg gatgacggtt gaaattcgga tggcataata	5700
attttatatt tggagctggt ctttttagtag ctccattttt tatcccatga gcaaatcaga	5760
acacatacca gtattattaa acgaagtaat tgaaggctt gacttgctct ctaatgatac	5820
agtaatagac gccacagtag gcggagcagg acacgcgcaa gctattttag aaaaaaccgc	5880
gccatcagcg aagcttcttg gaattgattg ggacgcgaaa gcaatcgagc gcgcgcgaga	5940
acatctaaaa agatttagca accgaattat attaaaaaca ggaaattaca cagatataaa	6000
acaacttctc tatgaatcag gaattaataa ggttaatgct atattattgg acttgggctt	6060

atctcttgat caactcaaag attcctctag aggatttagc ttccaatctg aaggaccatt 6120
ggacatgagg ttttctgacc agatggacac aacagctttt gatattgtga acacctggcc 6180
agagaatgat ctggtacaaa tctttcaaga atacggtgaa gagaggcgcg ctgcacgtgc 6240
agcacgcaat atcgccactg cgcgcgagtca cgcgccaatc aacaccgcaa aagatctggc 6300
agaattagtt atgcgcgggg ccggaaggcg aggcaaggtt catcccgcta cccgcatatt 6360
ccaggccctg cgcatgtcta caaatcatga attagacaat gtcaaacaag cattgcctaa 6420
tatgattgat atgcttttctt cagaaggaag attagcagtt atcacattcc attccttaga 6480

agaccgcatt gtgaagcagt atttcaagcc attggctaaa gaggaaaatc cgcgcattaa 6540
gtcatcaat aagaaagtaa taaagccaag ccgagaggag caagtgaaaa atccagcatc 6600
cagaagcgcg aaattgagaa tcgtggaaaa gatttaatca ttccaaaaac aaaaatagca 6660
tcacatgaca acatattcgc acaaaaaaac gccgtatctg tggcacgcat tttcaatatt 6720
gtgatttta gtattagtgg ttacttattt agtacagata aacagccaag cagaaacatc 6780
ttactctatt aaaggattag aagaaaaaaa gcaagaattg aatagtatta tagaagataa 6840
agaactigaa gcagtttcag cgcgatcttt aaatggaatc gcgcttaagg caaaagaaat 6900

gaatttgag gatccaaagg atgttacatt cataaaaaata ggattaagca cagttgccgt 6960
gagcgaagag ctttctccat aacatgactt catattcatc atcaaaaaag agcaattcag 7020
ctacgcgcgc gaaattcata attggcgcggt tttttatttt tggcgttatt ttgatttacc 7080
gcttagctga ttacagctt atcaatactc aagaaattca ggcatctgcc gcgcgccagc 7140
agtcaacagt gcgcatcctt ccagctgaac gaggaagat tttttacaag gagagaatag 7200
gtgatgaaga atttccagtc gcgactaata gatcatataa ccaggtattc attattccaa 7260
aagacataca ggatccaatc aaagccgcgg aaaagctatt gcctttggtt gagccatag 7320

ggcttgatga agaaacatta ttattccgat taagcaagca aaatgacatt tacgagccat 7380
tagcgcataa attaacagat gaagagcttg agccatttat tgggcttgat ttaattgggc 7440
ttgaatcaga agatgaaaaa gctaggtttt acccgacgc tgatttgctc gcgcatataa 7500
ctgggtttgt cggggtttca gaacaaggca aggttgggtca atatgggctt gagggatttt 7560
ttgaaaatga gctcaaagga aaggacgggc ttattgaggg caaaacagat atatttggca 7620
ggcttataca aacaggaact ttaaaacgca cccaaggcga gccaggagat gatttattat 7680
taaccataca gcgcacttg caggcataatg tgtgcagaaa attagatgaa aaaattgagc 7740

aaataagagc tgctggcgga tcagtaataa ttgtgaacc agatactggc gctattctcg 7800
cgatgtgctc ttaccatca tttgatccga ataattataa tcaagttgaa gatattagcg 7860
tatacatgaa tccagcagtg agctcaagct atgagccagg atcaattttc aagccattta 7920

caatggccgc ggcaattaat gagaaagcag ttactagcga tacaacatat attgatgagg	7980
gagtgaaga gatcggcaaa taaaaatcc gcaattctga caacaaagcg cacggggaag	8040
ttaatatggt aactgtttta gatgaatcat tgaatactgg cgcgattttt gtccagcgtc	8100
agattgaaa tgagaagttc aaagattatg ttgaaaaatt cggatttggc agaacaacag	8160
atattgaatt aggaaatgag gtttctggaa atatttcttc attgtataag gatggagata	8220
tttacgggc aactggctcg ttggccaag gaattactgt tacgcctatt cagatggtaa	8280
tggcatatgc ggcgattgct aatggaggaa aattaatgca gccatatctt attgctcagc	8340
gacaaagaca ggataaaact attgtactg agccagtcca aattgatgag ccgatttcag	8400
tgcaggcctc aactattata tctggaatgt tggtagcgt ggtgcgtgct gggcacgcta	8460
tatctgctgg agtgaagga tattatattg ccggcaaac tggaaccgcg caggtcgcg	8520
aaggcggagg gtatggaagc aagaccattc attcatttgc cgggtttggg cctgttgatg	8580
agccagtgtt tgcaatgctt gtgaaattag attatcctca atacggcgca tgggcagcta	8640
atactgcggc tcctttgttt ggcgaaatag ccaaatttat actacaatac tatgaaatac	8700
ctcctgatga ggcgataata ataaaaatg aaaaaataa taattacaat ttacaaaact	8760
ctggccaaaa gagtatttta caaatataag ccaaagtgg tggctattac tggctcagtc	8820
ggaaaaaccg cgactaagga ggcagtgtt gctgtattga ataagaaatt gcaagtgcgc	8880
aagaatgaag gcaattttta cacggaatc gggttgcctt tgacaatcat tggcttgcaa	8940
aatcaccag gcaaaaatcc attcaatgg ctgacagtgt acgcgcgcgc tattggcctt	9000
ttaatcttta ggattgatta tccaaaagt ttggttcttg aaatgggcgc tgataagcca	9060
ggagatattg ctgaattaat aagtattgct aagccagaca ttggcataat taccgcgatt	9120
agcgtgttc atacagagca gtttaatagt attgctggcg ttgtcgtga aaaaggaaag	9180
ctctttcgcg ttgttgaata ggatggttgg attatcgtga ataacgacg atctgaagtt	9240
tatgatatcg cgaaaagtg cgacgcgaaa aaagtatata ttgggcagtg cgctgaatta	9300
tctgataaca ccccttttcc agtatgcgcg tccgagattt cagtgagcat gtcagaagct	9360
caagaaaccg gcattgctgg cacttcattt aagcttcata ctgatgaaa ggttattccg	9420
gttttgatga aaggaattat tggggagcat tggacatata ctgcatgta cgcggcagct	9480
gttgcgcga ttcttgggtt tcatatggtt gatgttactg agggtttgcg cgagattaat	9540
cctcaatcag gaaggatgcg agttttagct ggcattaaaa aaacaatttt aattgatgat	9600
acttataatt cttcgccaaa cgcggctaag agcgcggttg atactttagc gttattgcgt	9660
attggaaggg agaaatatg cgtgtttggg gatatgttgg agcttggttc tatatctgaa	9720
gaagagcatc aaaaattagg catgcttgc gcgcgcgagg ggattgatta tctgatttgc	9780

gttggcgagc gcgcgcgcga cattgcgcgg ggcgctataa aagcaaagat gccgaaggat	9840
catgtgtttg aatttgataa tactaaagat gctgggctct ttatccaaaa gcgtttggag	9900
caaggggata tggttctgat taaaggttcg caaggcgtgc gcatggagcg cgtgacccaaa	9960
gagattatgg cgcatccgga aaaatcaaaa gaacttcttg tgcggcaaag taaagaatgg	10020
ttgagtaagg cctagtgcgt atttttgata atttctcca cttcttccgc attttctgca	10080
tccatcaatt tcacgcgcaa ttgctttgcc ccatcccagc cagaaacata ggccttgaaa	10140
tgttttttca ttacagcgaa tgatttgtgt ttgataagtt tttcgtagag tttggcgtgc	10200
tctattaaaa cgcgcaattt gttatctttg ctgggataga aaacggagaa aacggtgtca	10260
agagtctgtt tcigtaaaaa acgactccig acaccgtttt cttigaagaa ccacggattg	10320
ccgaaaattg cgcggccgat cataacgcca tcaacaccgg tctcccgggc tttttgatgc	10380
gcatcgtcta aatacgaaac atctccattc ccgataataa gcgtcttggg cgcgattttg	10440
tctcgcattc gaataacgtt tttagccaaa tgccatttag caggaacgcg ggacatttct	10500
tttctagtgc gccagtgaat cgtcaaagcc gcaatgtctg tcttcagaag aataggaatc	10560
caggatcaaa tttcattttt cgtatateca atgcgcgttt taacagaaat tggcaatttt	10620
ggcgcgcctt ttttggctgc agcaatcaaa gcgcgcgcta aatcagggtt tttcatcaaa	10680
ccagccccag cgccttgctt ttcaactttc cgttcggggc atcccatgtt aatatcta	10740
ccatcaaaac ccaaatcctg aattatgcga gctgtttttt tcatattatc tggatttgct	10800
glaaatactt gcgcgacaat aggccgctct ttcgcggaaa atttaagatt tttagaatt	10860
tcatctttgt cgccaagagc aatgccatcc gcggacacga attcagtcca cattacattc	10920
ggcttgccat actttgcgat aatccgccta aaagccgcgt ctgtcacgtc agacatagga	10980
gccaacaga agaattggtt tttgagttgt tgccaaaaat tattcatgtc atcttgcgct	11040
tatttgcatt cccgaggtt aattatatat ttttagaaaa taggatgtgg taaacggatt	11100
atataagtgt aatagtaatg ccacacaagc cgagaggatc tcgtctttaa gagctcgaga	11160
tgacaataca aggcgagaga atctcgcgac taataactat gcttattatc aaataaatcc	11220
ttccaatcag aattgaattt gtttataagc aacaccttat ttcgtggct tagttttttt	11280
agcttctttt cgcgtcaat agcatacgag atattgtcaa agtgttcata atacaccagt	11340
ttatcagtat tgtattttga agtaaacctt ggtatttttt tatttttatg ttcccaaatt	11400
cttctggata atgaattgca tactccggtt taaaataccg tatgtcgtat gtttgttgtt	11460
atatacatat aaaagttata ttgattttgt cttggcatgt tttgtttca taagatcctc	11520

tcggcctgca aggatTTTTG ttttggactc catgattcgt ttaccacata ttcgatatta	11580
tgtagtattg taaggtctcg ggatgacagg taaaaggcat gggaatggca tctaaatctc	11640
ctcctttttc tcatgcacat aattcatcca ttcctcaatc acttttataa acgccttgaa	11700
cggagcctct ataataaaat ccaacgcaaa aatgaaaatg ttaatttgcg cgaaccgcgt	11760
ggacatccat ttgccagcat gcagaatcgg aattgtaaaa aacgcccata agccccggat	11820
aaatccttgc tttgggggca ggacaatcat ttcctgattt gactggcgga tgcggtagc	11880
gaataaggaa acaaacgaga ggaacaagag aaagataaaa atgccgataa acgtgaaatt	11940
cagcgcgac ccaatataaa tcatcaaacc gaacgaaatg ccaaatagca ttcctgataa	12000
caaagtaaac accgcgcgca ggaaaaagct acgcttgcta gatttgcgca tctgaataat	12060
ttcgccttga ttttggataa tatgatttat gccacttata atttgattgg tgttttcttc	12120
atcaggcagt ttagttgaga gtgcgataag cgcgagcaag gcaggcgga aaattaaatt	12180
aatagccaaa ggcatataat caattttgtg aatcaataaa taatcaacag gaatttcag	12240
aaccacggct aataaaaatt tagtaattac caataaata atacttcgct taatgcctcg	12300
gtgtaaagaa gcgcgggatt tttcgtactg cttttggcag atggcgcgca ccctttgctc	12360
aaattcatgc ccggtgttca tatcagacca ggcttttctt ggatcctgcg caatcgcgtc	12420
ttgcaaaata gtgaaatata caacgtattt cctgaacaaa ggagcgagtt tttcttttat	12480
aggcgagttt aaatcttgcg ttattgtaga gtgtatttca ttcaaatgct ctcctatttc	12540
ccgtataaga tcgtgatttg cgcgcgtcca tttcggataa taggtcaata gcaaatgata	12600
tccaatagtg tcattgtcgt ttttatatag aattcggcta gtggctatat aaatctgctt	12660
caaacgttct cgatcattaa tttcatcttc aattctaacg cgctcctgaa gatattcata	12720
catggcattg attgacgcgt gcattacata tggtagcata aggaattcgt caatttcgt	12780
tgctgctatg ccagagagcc aaaatgaaag agaagaagaa tcattgatat cttttatagg	12840
agcgtgacca agcaaggtag aatatatttc aaataataa tcaagtctt ttatttttcg	12900
ttcagggata gtattgttcg gaaggtaccg cgcgtggata agttctgaaa ttagattttc	12960
tgagatatta tttttatggc ctgatgaaat cattctacgc aaaatgcgt caatcgcgtt	13020
tctgcggatt aaatgttctt ctttatattc aaccgcgttg cgcatgcgt cgtatataaa	13080
agttgcctgt ccagcgcggg tggttatgga gatTTTTGGT tccgtcgggt ctgtatcttt	13140
tgagcgcgt tcttccctgg ccgcgcgcac gattcgtgg attgtttctg gtatttgcatt	13200
ttctttatac tagctgattt tgcttgtttt ttcaattgtt ttataaaaaa agtgcccga	13260
atgcaaatg cgcatccgg gcttggggag acagggcagg ggatgccctg tttggggctt	13320
actgccggtc ggtcagatca cgggctacta ccgccgaat cctcgccacc gccaggcag	13380

taacgagacg acitctttttt tacctgattg acgaccgtac cgtcgagcag gacgttatcg 13440
ccgagcagat tcgtgtatt gatgtccgta gccgcggtag ccgcgatagt cgtggtcgtc 13500
gtcgtggttt ccgtagtggc tgtgccgacc gcgctgtttt cgccgccctc ttttgtcatc 13560
cgaatgacat catcgccatt cagagtcgtt tcctcgctga ccgggttggtt ggtcccgcag 13620

ccgatcattc cgatcagggc gaccagcgcg atacagaaga aaatcatgaa atacttcac 13680
gggtgctcct ttttatgagg tttttgaaa acgatatcac gctttgtatt attcacctcc 13740
cttccaaagc aagcgcaata tcggtctttt ttactatttt aagaacggac gagcatctta 13800
tactatttta aaaataatgt caagagtgtt aacaaatata aaaaattgac tcatataaaa 13860
acggtgtcag gagtctgtt 13879

<210> 26

<211> 7532

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<220><221> misc_feature

<222> (2669)..(2692)

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

<400> 26

tacctaatcc tgggcgtctt tgggtgtatta tgcacttgcg gttagaatac acccgaacat 60
aattgacaaa gaccataaaa tgtcttatta tccttttaga aaaatcgtgt tcatttataa 120
tatatacata cccaattcc aaggatttct tgactggcag cgggcttggt atcctgcgaa 180
acacagccag tttgggaaac ctgggtcttt atttttaag acacaggaat tcccgcgtct 240
tttgccttgg aacaccaacc acctattgcg cttttttct cattttagca aaagtggctg 300
tctagacctt caggtggaaa ggtgagagta aagacattgg gcctgcacga ttcattgggcc 360

ggtctagacc ttcaggtgga aaggtagagag taaagactct accgcgtcca gcactatctt 420
ggtccgtcta gacatttaga tggaaaggcg agagtaaaga tgcgcgaaag acggctacat 480
tgttccacaa ggcagaaagg attagccgcc tactgcttga acatccgcag tatttaaccc 540
attttccaa aggaggaaaa tcatgggtac gcagattatc aagcggatag accttgactg 600
gcagtcaagt tttccgcacg ccaagatgct ggtgaatcag gaagcatcat ttaaccacat 660
tgcagagtcc ggactcacgg cgctcataga agcgcgcgacc ggatccggaa aaaccgcgac 720
tggtataacc tttctttcgg ccatagccct tcgcgcgcgc aagagtccgc aatttaaggg 780

ccggctcgtg taigtgtctc cgaataaagc attagtcggg caggtgcaga acatgcatcc 840
 agatgtgaaa gtcgcgcttg gtcgcaacga gcatacatgc tcgtattacg atggaattca 900
 tcaagcagac gaagtgccgt gtctgttttt ggttcgctcc ggccgggtgtg gccactatgt 960
 gaatcaagaa accggcgcaa cacttgaatt tggagctgaa ccatgttcgt attatcagca 1020
 aatctatgag gcaaagcgcg gcatcggaat tctggcatgc actgacgcgt tttggctgtt 1080
 cagcatttg tttaatccaa agcagtggcc tcagcccatg ggtttggtat tggacgaggt 1140
 tgaccgcttg gcigatatgt ttcgcagggt cttgtcatac gaaatttctg attggcgcat 1200

 tgagcgcgcc attaatttgc ttgaaaaagt cggttcagtt caggtgcagt atctctcgtc 1260
 ttttttgcgc accttgaatc ggggtggtatc aaaaaagccg gccctggagc ccattttgct 1320
 ggatgatgag gagattcgcc aactgtttga aaaagtgggg cgcatcagcg cggatgtcat 1380
 caaatccgat ttggacgccg cgattgagcg caacaagggt gacctatgg ctgagcgca 1440
 aatccttaag cagatagaaa cactttgctt tgacatcagc cggatgtgc ggagtttggg 1500
 atacgcgctt ccgaatcgca gaggcaaggg tgatgaacgc aagcgcatg ctctctttc 1560
 gtacgcgtac gcgtatcata aatccgagcg cgacgctggg gcgcatgtgc agaacaaagt 1620

 tgtggtgtgt tcctattggg tgcgcctct tatccgcaag ctctttggaa agaacacgt 1680
 cgcgtattca gcgtttgtcg gggataaaac gattttggat tatgaggctg gagttgattt 1740
 tccattaate tctctgcgtt cccaatttcc ggcgagcaat gcgcgattgt atgtgccgag 1800
 cgattctcca aatttggcat ataagagca ggatgtcggg gacatggcta agactttgcg 1860
 ccatattgcc atatcaactc ggcggtttgc cgagcgcggc tttcgttctc tcttctgac 1920
 tgtttcaaat agagagcgtg aattgctgta cgtcgcgtgc gcggaactga aagggtgga 1980
 tgctataagt tatggcagtg gcgttactgc gcgcgcggcc gcgtagatg tcaaagaagg 2040

 agaaggggac gctcttattg gcgttttgc gcattatggc actgggctgg atttgccagg 2100
 caagattgct aacattgttt ttctctgcg gccgaatttt cctccaccaa aagatcctat 2160
 ggacagttt gagattcgcc gggccgagcg catcaaaaag tcgcattggc ccgtgtggtg 2220
 ctggcgcgcg taccgagagg ctctgaatgc ccagggacgc ccgatacgaa gcgccgatga 2280
 caaaggggtc gcgttcttta tctcccagca attcaagaag cgtttattca acattttgcc 2340
 ggagcatctt gagagcgcat atcgagccg cctcacatgg gaccagtgcg agaaagacgc 2400
 gctgaaactg tttgaggaat aggggtattt tttcgttgtt tttatggccc ggatggtgtt 2460

 ttttatacat catccgggtt tttatgttga ttgatgcga taatcatgat ttttgcgtgg 2520
 tattgacaaa cattataaaa aacgtatta tccgctaca aaacctataa atcgttcatt 2580
 tataatatat acatacccca attccaagga tttcttgact ggacgcgggc ttggtatcct 2640

gcgaaacaca gccagtttgg gaaacctgnn nnnnnnnnnn nnnnnnnnnn nngccagttt 2700

gggaaacctg ggtcttttatt tttaaagaca caggaattcc cgcgtctttt gccttggaaac 2760

accaaccacc tattgcgtct ttttcgtca ttttagcaaa agtggctgtc tagacataca 2820

ggtggaaagg tgagagtaaa gacatggcct gaatagcgtc ctgcctctcg tctagacata 2880

caggtggaaa ggtgagagta aagaccggag cactcatcct ctactctat tttgtctaga 2940

catacagggtg gaaaggtag agtaaagaca aaccgtgcc aactaaaccg atgagtctag 3000

acatacagggt ggaaggtag gagtaaagac tcaagtaact acctgttctt tcacaagtct 3060

agacatacag gtggaaagggt gagagtaaa actcaagtaa ctacctgttc tttcacaagt 3120

ctagacctgc aggtggtaag gtgagagtaa agactcaagt aactacctgt tctttcaca 3180

gtctagacct gcaggtagta aggtgagagt aaagactttt atcctcctct ctatgcttct 3240

gagtctagac atttaggtgg aaaggtaga gtaaagactt gtggagatcc atgaacttcg 3300

gcagtctaga cctgcagggtg gaaaggtag agtaaagacg tccttcacac gatcttctc 3360

tgtagtcta ggcctgcagg tggaaagggt agagtaaaga cgcataagcg taattgaagc 3420

tctctccggt ccagaccttg tcgcgttgt gttgcgacaa aggcggagtc cgcaataagt 3480

tctttttaca atgttttttc cataaaaccg atacaatcaa gtatcggttt tgcttttttt 3540

atgaaaatat gttatgctat gtgctcaaataaaaaatca ataaaatagc gtttttttga 3600

taatttatcg ctaaaattat acataatcac gcaacattgc cattctcaca caggagaaaa 3660

gtcatggcag aaagcaagca gatgcaatgc cgcaagtgcg gcgcaagcat gaagtatgaa 3720

gtaattggat tgggcaagaa gtcatgcaga tatatgtgcc cagattgcgg caatcacacc 3780

agcgcgcgca agattcagaa caagaaaaag cgcgacaaa agtatggatc cgcaagcaaa 3840

gcgcagagcc agaggatagc tgtggctggc gcgctttatc cagacaaaaa agtcagacc 3900

ataaagacct acaaatacc agcggtatctg aatggcgaag ttcattgacag aggcgtcgca 3960

gagaagattg agcaggcgat tcaggaagat gagatcggcc tgcttggccc gtccagcgaa 4020

tacgcttctt ggattgcttc acaaaaacaa agcgagccgt attcagttgt agatttttg 4080

tttgacgagg tgtgcgcagg cggagtattc gcgtattctg gcgcgcgctt gctttccaca 4140

gtctccagt tgagtggcga ggaaagcgtt ttgcgcgtg ctttagcatc tagcccgttt 4200

gtagatgaca ttaatttggc gcaagcggaa aagttcctag ccgttagccg gcgcacaggc 4260

caagataagc taggcaagcg cattggagaa tgtttcgcgg aaggccggt tgaagcgctt 4320

ggcatcaaag atcgcatgcg cgaattcgtg caagcgattg atgtggcca aaccgcgggc 4380

cagcggttcg cgccaagct aaagatatc ggcatcagtc agatgcctga agccaagcaa 4440

tggaacaatg attccgggct cactgtatgt attttgcgg attattatgt cccggaagaa 4500

aaccgcgcgg accagctggt tgttttgctt cggcgcttac gcgagatcgc gtattgcatg	4560
ggaattgagg atgaagcagg atttgagcat ctaggcattg accctggcgc tctttccaat	4620
ttttccaatg gcaatccaaa gcgaggattt ctcggccgcc tgcacaataa tgacattata	4680
gcgctggcaa acaacatgtc agccatgacg ccgtattggg aaggcagaaa aggcgagttg	4740
attgagcgcc ttgcatggct taaacatcgc gctgaaggat tgtatttgaa agagccacat	4800
ttcggcaact cctgggcaga ccaccgcgc aggattttca gtcgcattgc gggctggctt	4860
tccggatcgc cgggcaagct caagattgcc aaggatcaga tttcaggcgt gcgtacggat	4920
ttgtttctgc tcaagcgctt tctggatgcg gtaccgcaaa gcgcgccgtc gccggacttt	4980
attgcttcca tcagcgcgtt ggatcggttt ttggaagcgg cagaaagcag ccaggatccg	5040
gcagaacagg tacgcgcttt gtacgcgttt catctgaacg cgcctgcggt ccgatccatc	5100
gccaacaagg cggtagacag gtctgattcc caggagtggc ttatcaagga actggatgct	5160
gtagatcacc ttgaattcaa caaagcattt ccgttttttt cggatcacagg aaagaaaaag	5220
aagaaaggag cgaatagcaa cggagcgctt tctgaagaag aatacacgga aacagaatcc	5280
attcaacaac cagaagatgc agagcaggaa gtgaatggtc aagaaggaaa tggcgcttca	5340
aagaaccaga aaaagtttca gcgcattcct cgatttttcg gggaagggtc aaggagttag	5400
tatcgaattt taacagaagc gccgcaatat ttgacatgt tctgcaataa tatgcgcgcg	5460
atctttatgc agctagagag tcagccgcgc aaggcgcttc gtgatttcaa atgctttctg	5520
cagaatcgtt tgcagaagct ttacaagcaa acctttctca atgctcgag taataaatgc	5580
cgcgcgcttc tggaaatccgt cttattttca tggggagaat ttatattta tggcggaat	5640
gaaaagaagt ttcgtctgcg ccatgaagcg agcagcgca gctcgatcc ggactatgtg	5700
gttcagcagg cattggaat cgcgcgccgg cttttcttgt tcggatttga gtggcgcgat	5760
tgctctgctg gagagcgctt ggatttgggt gaaatccaca aaaaagcaat ctcatttttg	5820
cttgcaatca ctcaggccga ggtttcagtt ggttcctata actggcttgg gaatagcacc	5880
gtgagccggt atctttcgtt tgcctggcaca gacacattgt acggcactca actggaggag	5940
tttttgaacg ccacagtgtt ttacagatg cgtgggctgg cgattcggct ttcattctag	6000
gagttaaaag acggatttga tgttcagttg gagagtctgt gccaggacaa tctccagcat	6060
ctgctgggtg atcgcgcttc gcgcgacttg gctgcgtgca aacgcgctac atgcccggt	6120
gaattggatc cgaataattct tgttctgccg gctgggtcgt ttatcgcgag cgtaatgaaa	6180
atgattgagc gtggcgatga accattagca ggcgcgctatt tgcgtcatcg gccgcattca	6240

ttcggctggc agatacgggt tcgtggagtgc gcggaagtag gcatggatca gggcacagcg	6300
ctagcattcc agaagccgac tgaatcagag ccgtttaaaa taaagccgtt ttccgctcaa	6360
tacggcccag tactttggct taattcttca tcctatagcc agagccagta tctggatgga	6420
ttttaagcc agccaaagaa ttgggtctatg cgggtgctac ctcaagccgg atcagtgcg	6480
gtggaacagc gcgttgctct gatatggaat ttgcaggcag gcaagatgcg gctggagcgc	6540
tctggagcgc gcgcgttttt catgccagtgc ccattcagct tcaggccgtc tggttcagga	6600
gatgaagcag tattggcgcc gaatcgggtac ttgggacttt ttccgcattc cggaggaata	6660
gaatacgcgg tggtaggatgt attagattcc gcgggtttca aaattcttga gcgcggtacg	6720
attgcggtaa atggcttttc ccagaagcgc ggccaacgcc aagaggaggc acacagagaa	6780
aaacagagac gcggaatttc tgatataggc cgcaagaagc cggtgcaagc tgaagttgac	6840
gcagccaatg aattgcaccg caaatcacc gatgttgcca ctcgtttagg gtgcagaatt	6900
gtggttcagt gggcgcccca gccaaagccg ggcacagcgc cgaccgcgca aacagtatac	6960
gcgcgcgcag tgcggaccga agcgcgcgca tctggaaatc aagaggatca tgctcgtatg	7020
aaatcctctt ggggatafac ctggagcacc tatitgggaga agcgcaaacc agaggatatt	7080
ttgggcatct caacccaagt atactggacc ggcggtatag gcgagtcattg tcccgcagtc	7140
gcggttgccg ttttggggca cattagggca acatccactc aaactgaatg ggaaaaagag	7200
gaggttgatc tcggtcgact gaagaagttc tttccaagct agacgatctt tttaaaaact	7260
gggctgctgg ctatcgtatg gtcagtagct cttatttttt tacttgatat atggtattat	7320
ctcaataata tgcattctct catagataca acagaaaaag aatcatttga tattgctttg	7380
attgatgatg agcgcgttat caaaaagaag cgaatcaaat caatccgcca acattcggaa	7440
aagcttttga aatcaattga cgcgcttttg ttgtccgcaa aatcatctct gaaagatata	7500
caaggcatca tcgcggtaaa aggccttggg tc	7532
<210> 27	
<211> 16262	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 27	
cggaaaggcg gccagaaac gggttgacca aattttgtgt tcagtggatga tgatggcgat	60
gycgatgtcg ctgcttacgc gggcgttgtg caggccgatg gattcggaat tcagaatggc	120
ctggacgtgg gggagggtgg ccagccagcg caggtaatga tgccgtttgc gcagtttgg	180
ttcgggtgagg ccgtagcggg ccaggcgcag ggggacgagg tgggagcggg tttttaggtg	240

ataaaagcca tcggtgtgag tgatgtgtgg gtgagtggcg agggcggcag tgagttcggc	300
gggcgtgggtg gtgtgttgcc acagccagcg ctggagtcca ccggcgggtca gcggaaattc	360
catgagggtca aagtagctca tgggtggcgggt gatggcgtgc tcgagttggg ggcgacaagc	420
gggtttcatg ctctattat agcagatttt cagagttgga tttttgctgt tttttcttgg	480
ccggagtacc cgttttttta ttgtttgaaa aatcagggtt taaaaatttt aggtgagagt	540
ctttttgcta tatccaagaa gaaattttgc catatttttt ggtcaatttt tattttcatt	600
cttggtaggt cttttaattc ggtcactttt aatagttggc ttcccatgtt tactgggtcg	660
atgtgccagt caaattttat cttggccttt ttatcagat catcgaatgt ccattctttt	720
ttttggagaa tacaatacag gtctataaaa tcccgtgaac gtggtttttg atacatggta	780
aatactttgt tgacggcaat gtctaacagg ctgtcaattt tcagaccgtt tgttttcaag	840
cccttttgta taatcggaaa ggggtagtag gtaaatccg ttttgataac atccttgctg	900
atatggataa aaaacagatt ccggttgaag ctctgtgaa aatctatctt tttaaatttt	960
acctttttct gtattttttt gagtatagta aaaatatccg tagaatcgaa ttctttttcc	1020
gaaaagaaat ccaaatcttc ggataaccga tgatgcagat aaaattctgc gagagcggtg	1080
ccaccggaaa gataaaattt ttcccggatg agtttttctt gtgatagctg ctgaaggaga	1140
gcgcttttgg ttggtgtcag gattgttggc cccataggag aaaagataaa aattttttct	1200
tacctgggtc gaggtccagc ctatcccagt actttttcag ttgacttcgc ttgatttttt	1260
ttccacccaa gccaaaattt accatctgtt cgagtttcca gatagtgtat ttttctttat	1320
ttttttttag ctctgtgagg tcaatattcc aattgtacat ggctgtatit tagcatatag	1380
cagcttaaat ttcaatttta ttttagccaa aatagtagaa tgggtggcgggt gttagatgaa	1440
tatttcgtag ttgtcttttg atatcacctg gaattttgcg tcttggtagg catcgtgaa	1500
tgcttttggc gtcgggctg attttttccc ccatttgaat tcaaatgccc tgagttttcc	1560
atttttttct tccaagtaat caatttctgc ctttttgggtgc gtgcgccaga aatatctgtt	1620
taccgaattt tcagtatttt ccaatttttt cattctttct acgaggagaa aattttccca	1680
gagccccccg acatcttcgc gtaaggagag aggattgaga ttattaatga gtgcgttgcg	1740
aatgccgaga tcatagaagt aaattttccg cagttttttg agttcgttgc gaatgtttcg	1800
actgtatggt ttcaaagtaa aaacaatgaa agccttctca agaatgccta tataattagc	1860
cacggttttt tgatcgatat tgagcagggt ggacagttcc cggtaggaaa cttctttacc	1920
tatctggagt gccagcgctt gcaggagttt gtcgagtact tcaggattgc ggatgttctg	1980

aaatgccaga atgtctttat ataaataact tctggtgata ttgcgcagca attcctcagc	2040
ttccgatgat ttgaggacaa tttccggata cataccgaaa atcattcttt gttccagtgt	2100
tcttttttct tctgtatat tctgtatctg cctgagttct tccagtgaag agggatagag	2160
gataaattca tattttcttc ctgtgagcgg ctcaacgata tgattagcga gatcaaaaga	2220
agatgatccg gtggcgataa tctgcatttc cggaaagtgt tcaacaagta gtttcagtgt	2280
cagtccgata ttttttactc tttgcgcctc gtcaaggaag ataagtgttg catctcccag	2340
ataggccttg agttcggctg aagttttgtc ggtaagagcg gtgcgaacgt ccggttcac	2400
acagttgagg tagcgagagg tgtggctcgc aagcttttcc tcaagggtt tgaggatagt	2460
tgtcttacct acctgtctgg cgccatagat aataataacc tttttttga aaagggtgtt	2520
ttcaataata ggctgaaggt ctctgctaata ccgcatagta tataatgattt agatgattat	2580
actcctctca ttatatatta aaatgcggat ttagtcaatg cattctacta taaatgcttt	2640
atattagcca aaatgtcaga aaattgatat ttttgacat ttttactata tttcggacac	2700
cttatttttg ttctcgattc atgtatcact gcccgctgta ggttgccggc caatttttaa	2760
aggagaattt tatgatgcct gttgtgctct ttataaaatc gtttttttg atttccata	2820
gttctctctt gtagggactt gaaataaaat gttttttata ctactatagg cctagttcct	2880
taacaatatt ttgcttactt taaagcgaaa ataggtaagg cacacctata ccataaggat	2940
ttaaagactc ttggcgaca gctttccacc gacctgagt agttaagac tgacgtatca	3000
tgtcataaca ccaacatttc tagatataaa gacgcgacag ctttcaggcg ataccgacgt	3060
ttctagacat aaagacttg gataaaccat aatgcaccga cgttcctcga tataaagacc	3120
cgtttgggtc ccaaaattca ccgacatttt aagaggtaaa gacaagtga cctgagtcgc	3180
tgcaccgaca ttcccgatat aaagactgtc gctcaacccc aaaacaccga cattcccgat	3240
ataaagactc gccctagatc ttcttgacc gactgtataa ggaataaaga cgtccgacca	3300
cgtgcaccac accgactcgt gtgaacctaa agactcaccg ccgcactacc ctaccgact	3360
atatcaaacc taaagattgg taacttggtt gtctgacacc gactgtatca gagataaaga	3420
ctgttttcca tgcgttgccg accgacgttc ctatagataa agactatcat tatcggggaa	3480
accgccgact gtactagata taaagaccg tcgctttgtt tgaacgccga cattcttaag	3540
aaataaagac gtggtaagag tagtgtttca ccgacattcc tttatgtaa gacaatgaat	3600
agtctttttc acaccgactg tgaatgtatg aaatctaaag acctgaaagt gcaatgcaaa	3660
tgctgacagt gttagtctaa agacaaagta ggaatcagga tccgccgact aaataaaact	3720
taaagacaag ccagatatcc aggccacacc gacgtcccta gatgtaaaga ctagtgactc	3780
catgctatgc accgacattc cgaggcctaa agacagagag gctaacattt gtgcaccgac	3840

ccttcaagag gtaaagacat aggggaacacg ctgaatcacc gacgttccta ggtatacaga 3900
 cgaaatgcaa tgaaaaacgt caccgacatt tcaagacgta aagacccaag aatctttgcc 3960
 cgtccccgac attccaagac gtaaagacta gccaaaacct ccagaccccc gacattccaa 4020
 gacataaaga caagcgctcc aacatgtgtc accgacatta ttccgcccc gcatcgatca 4080

 ttttgacttg gaaagagaca ttcttctttc caagttttta ttttgagcaa aatttgactt 4140
 tttattgggt atcctttatt actatgggtg cttagtgcac cgaaagggtg gctaagcaca 4200
 acaaaagtgt tctttttatc ttaaaactga ggttttagac ctcatcaacc caaaaagggt 4260
 gtaacatcat gaaacatcag aaacatcaag aaaatgcagt ctctgacgaa acatctaacc 4320
 ctcccgccga gccatggatt ttgtattttg agaaatgggt gccctacgat acgtatccca 4380
 ccatgcatca taatcaatcc gaggttttca aattaattcg aagtgtccta cggaaagaag 4440
 gtgtgggtaa aaccatcctt gaacttcta ccggtatctg gaagacggtc attgggatcg 4500

 tgtatctcct tactttgcat cacaagatgc aggaaggcga gattcctaca gctccgctgt 4560
 ttacatcgt gcctaataag gcgctggtaa agcaggtgtg tgaaatgtc ccagatatca 4620
 cctttgggtg gtatggccgg aatgaatatg attgtctgta ttaccagccg aaagaaacgt 4680
 ttacagccga tcagattccc tgtttgggtc taccatgcaa gcatcgggtg aaccaggatg 4740
 atggaactac gcaagaatct ggtgtcgtg catgtccgta ttatttggtg aagtataagg 4800
 cgaagcagct gactcagaag gctcgaatca ttgtctgtac cgcttctttt tatcttttca 4860
 ctcaactcat tcatgagtgg ccgctgcctg gaggactggg tattgacgaa acggatgagc 4920

 tggctgaaat ttttcggcgg gcgctctcca cgaaagtcag tgattggcac ctgagtcagt 4980
 gcgtcacgat gatgcggcaa agtgggatgg atggtgaagc ggatctcatg cagaaatttt 5040
 atgacgccgt ggttagaatt gtcggagtca agtctcctca aaagcctacg cttttgaaga 5100
 aacacgaaat cagtgagctc ctcgaggtag ttctctcagtt cgacaccaga aaactgaaaa 5160
 ggcgtataaa tgcctcatc aaagacggaa agattgatgc agagaattcg cgtgaagtgc 5220
 tgaatcagct gactgtggtt gccaatgac tgaacgata cgccgtttcg ctgacctatg 5280
 ccttgacctga gggtagacct agggccctta attacctta tgcattatga gaaggaccgg 5340

 atgatcttcc agggaagaag aaagtctggt gtgtcattaa tatctgcaat tggtagatgc 5400
 cgctctcat taggcggatt ctctcgctc ggacctggc atatacagcc actatcggtg 5460
 agtatagtga cttgcctac gataccggaa ttgaaggttc gttttatacc atgaactctg 5520
 attttccggt ggagaactcc cgtatcttca tgcccgatga cgttgccaac ttggctgtga 5580
 aatcggtcaa accaggcgac aaagatcgga tgatgcgtct gattgctaag tcagctcgtg 5640
 aatttgcgga tcaaggatcat cggagtctgg tgggtggtcat ttccaatgag gagcgttcaa 5700

ggtttctgga aattgttgaa gaatacagtc tcaaaatgct cacctatgga aatgggtgtt 5760

cggcgcgcgga ggctattgca aggtttcagg ctggtgaagg ggaggtgttt gtgggaacgg 5820

cagccaactg ttctcatggc ctgaacttcg ataagcagac tgctccggtg attttttttc 5880

tgcggcctgg ttatccggtg caggagatc cactcgcaga tttcgaagaa gagcggatgg 5940

gaaataagag gtggggtgtt tggacctggc gggttatgcg gcagttactt cagggtcgtg 6000

gccggaatat ccgcagtcgg gaggttttgg gagttatttt cctgatgtca ggccagtta 6060

aacgtttcgc aggggaaggcg attccggggg ggcttatcaa agcctatatc tccggcaaga 6120

aattcagggc ctgtgtgtca gaggccaaaa agctcctgaa aaagtcttaa ttaagccaaa 6180

aaaattgttt ttttgtctct gtccttgaca atataattga actttgctaa gttagggtcc 6240

cctgttagag gaaacagcag caaagggaag tctgagcgcg agaggcctta gtcttttagag 6300

ttcttaataa gaacttttct gggcccaaaag tgcgttttag tctttattcc ctgagctctg 6360

tctactttga tggggccttt ttttattcaa atttttttat tttcgctacg tcttgacaaa 6420

aatatagatg tatactatat ttgccccgag gtaataaaga aaatagcggg aaagctataa 6480

gattttatta tttcatttat aagaactttg aaaaccgaca ttatcaaaaa ccatgcaaag 6540

cccttttagat gagggcagga ggttgaaaaa atgaagagaa tttcgaacag tctgaaagtt 6600

gctgccttga gacttctgtt tgcaggcaaa ggttctgaat tagtgaagac agtcaaatat 6660

ccattggttt ccccggttca aggcgcggtt gaagaacttg ctgaagcaat tcggcacgac 6720

aacctgcacc tttttggcca gaaggaaata gtggatctta tggagaaaga cgaaggaacc 6780

cagggtgatt cggttgtgga tttttggttg gatacctgc gtttagggat gtttttctca 6840

ccatcagcga atgcgttgaa aatcacgctg ggaaaattca attctgatca ggtttcacct 6900

tttcgtaagg ttttgagca gtcacctttt tttcttgccg gtcgcttgaa ggttgaacct 6960

gcggaaagga tactttctgt tgaaatcaga aagattggta aaagagaaaa cagagttgag 7020

aactatgccg ccgatgtgga gacatgcttc attggtcagc tttcttcaga tgagaaacag 7080

agtatccaga agctggcaaa tgatatctgg gatagcaagg atcatgagga acagagaatg 7140

ttgaaggcgg atttttttgc tatacctctt ataaaagacc ccaaagctgt cacagaagaa 7200

gatcctgaaa atgaaacggc gggaaaacag aaaccgcttg aattatgtgt ttgtcttgtt 7260

cctgagttgt ataccgagg tttcggtcc attgctgatt tttcgttca gcgacttacc 7320

ttgctgcgtg acaaaatgag taccgacacg gcggaagatt gcctcgagta tgttggcatt 7380

gaggaagaaa aaggcaatgg aatgaattcc ttgctcggca cttttttgaa gaacctgcag 7440

ggtgatggtt ttgaacagat ttttcagttt atgcttgggt cttatgttgg ctggcagggg	7500
aaggaagatg tactgcgcga acgattggat ttgctggccg aaaaagtcaa aagattacca	7560
aagccaaaat ttgccggaga atggagtggg catcgatatg ttctccatgg tcagctgaaa	7620
agctggtcgt cgaatttcct ccgtcttttt aatgagacgc gggaacttct ggaaagtatc	7680
aagagtata ttcaacatgc caccatgctc attagctatg tggaagagaa aggagctat	7740
catccacagc tgttagtca gtatcggaag ttaatggaac aattaccggc gttgcggact	7800
aaggttttgg atcctgagat tgagatgacg catatgtccg aggctgttcg aagttacatt	7860
atgatacaca agtctgtagc gggatttctg ccggattttac tcgagtcttt ggatcgagat	7920
aaggataggg aatttttgc tttcatcttt cctcgatttc caaagataga taagaagacg	7980
aaagagatcg ttgcatggga gctaccgggc gagccagagg aaggctatct gttcacagca	8040
aacaaccttt tccggaattt tcttgagaat ccgaaacatg tgccacgatt tatggcagag	8100
aggattcccc aggattggac gcgtttgcgc tcggcccttg tgtggtttga tgggatggtg	8160
aagcaatggc agaaggtggg gaatcagttg gttgaatctc caggcgccct ttatcagttc	8220
aatgaaagtt ttttgcgtca aagactgcaa gcaatgctta cggctctataa gcgggatctc	8280
cagactgaga agtttctgaa gctgtctggc gatgtctgtc gtccactcgt tgattttttc	8340
ggacttggag gaaatgatat tatcttcaag tcatgtcagg atccaagaaa gcaatggcag	8400
actgttattc cactcagtg cccagcggat gtttatacag catgtgaagg cttggctatt	8460
cgtctccgcg aaactcttgg attcgaatgg aaaaatctga aaggacacga gcgggaagat	8520
tttttacggc tgcatcagtt gctgggaaat ctgctgttct ggatcaggga tgcgaaactt	8580
gtcgtgaagc tggaagactg gatgaacaat ccttgtgttc aggagtatgt ggaagcacga	8640
aaagccattg atcttcctt ggagattttc ggatttgagg tgccgatttt tctcaatggc	8700
tatctctttt cggaactgcg ccagctggaa ttgttgcgtg ggcgttaagtc ggtgatgacg	8760
tcttacacgc taaaacgac aggctcgcca aataggctct tccagttggg ttacctacct	8820
ctaaaccttt cagatccgga aaagaaaaat tccaacaact ttcaggagcg cctcgataca	8880
cctaccggtt tgtcgcgtcg ttttctggat cttacgtgg atgcatttgc tggcaaactc	8940
ttgacggatc cggtlaactca ggaactgaag acgatggccg gtttttacga tcatctttt	9000
ggcttcaagt tgccgtgtaa actggcggcg atgagtaacc atccaggatc ctcttccaaa	9060
atggtgggtc tggcaaaacc aaagaagggt gttgctagta acatcggctt tgaacctatt	9120
cccgatcctg ctcatcctgt gttccgggtg agaagttcct ggccggagtt gaagtacctg	9180
gaggggttgt tgtatcttcc cgaagataca ccatgacca ttgaactggc ggaaacgtcg	9240
gtcagttgtc agtctgtgag ttcagtcgct ttcgatttga agaactgtac gactatcttg	9300

ggtcgtgttg gtgaattcag ggtgacggca gatcaacctt tcaagctgac gcccattatt	9360
cctgagaaag aggaatcctt catcggaag acctacctcg gtcttgatgc tggagagcga	9420
tctggcgttg gtttcgcgat tgtgacggtt gacggcgatg ggtatgaggt gcagaggttg	9480
ggtgtgcatg aagatactca gcttatggcg cttcagcaag tcgccagcaa gtctcttaag	9540
gagccggttt tccagccact ccgtaagggc acatttcgtc agcaggagcg cattcgcaaa	9600
agcctccgcg gttgctactg gaatttctat catgcattga tgatcaagta ccgagctaaa	9660
gttgtgcatg aggaatcggg gggttcatcc ggtctggtgg ggcatggct gcgtgcattt	9720
cagaaggatc tcaaaaaggc tgatgttctg cccaagaagg gtggaaaaaa tgggtgtagac	9780
aaaaaaaaaga gagaaagcag cgctcaggat accttatggg gaggagcttt ctggaagaag	9840
gaagagcagc agatagcctt tgaggttcag gcagctggat caagccagtt ttgtctgaag	9900
tgtggttggg ggtttcagtt ggggatgcgg gaagtaaata gtgtgcagga gagtggcgtg	9960
gtgctggact ggaaccggtc cattgtaacc ttctcatcg aatcctcagg agaaaaggta	10020
tatggtttca gtcctcagca actggaaaaa ggctttcgtc ctgacatcga aacgttcaaa	10080
aaaatggtaa gggattttat gagaccccc atgtttgatc gcaaaggctg gccggccgcg	10140
gcgtatgaaa gattcgtact gggacgtcgt caccgtcgtt atcgctttga taaagttttt	10200
gaagagagat ttggtcgcag tgctcttttc atctgccgc gggtcgggtg tgggaatttc	10260
gatcactcca gtgagcagtc agccgttgtc cttgccctta ttggttacat tgctgataag	10320
gaaggatga gtggtaagaa gcttgtttat gtgaggctgg ctgaacttat ggctgagtgg	10380
aagctgaaga aactggagag atcaaggggtg gaagaacaga gtcggcaca ataatttgag	10440
aagtaaaata gttttttaga ttacgtttcg caaaggaggt gatttgggtc tttgaagaga	10500
ggtgtcatta tatgtggcat ctcttttcat tttagagat tttttctaaa aataaaaactt	10560
ggaaagaaat agttctttcc aagtcaaaat gatcgatttt aaggaatgtc ggtgaagtga	10620
tttatgaaca aatgtcttta tattcatat ggtcgggtga agtacgaatg cgagttgcct	10680
ttaggttttt accgtcggta atccacatta ttacttggg ctttaggctt catagcgtcg	10740
gtattctttt tatatatgca agtctttaca ttgaggaacg tcgatgttca aaccagatgt	10800
gtttgtcttt atacctcgga atgtcgggtga agtgatttat gaacaaagtc ttttaatttt	10860
acacagtcgg tggctttccg agcaagagta gtctttatat ttagaacagt cggcgtcggc	10920
agtgcctttt ataagtcttt gtatctcatg tagtcgggtc attgtctttg caactgggtc	10980
tttatctctt aatatggctg gtggaaactc ttgtgggaat ctttatctca agaaaagtcg	11040
gtgtgcctg aaagctgtcg cgtctttagg tctcatgcag tcggtgtcgg tcaaaagctc	11100
gcttgtcttt atattttata cagtcgggtg aaaggtgagc tggctgagtc tttatccctc	11160

ttaaagtcgg tgcaagaagt atggcgggtat gtcctttactt gtcgttaggt cgggtgttcat	11220
ccgtctcttag gggtgtcttta tctttatgaa tgcgggtgta ggtccaaacg atgtatgtct	11280
tacatcagga attcaggaat gtcgggggta ctaatatgca atggagtctt tatgtctggg	11340
aacgtcgtta ttttactctt gcgagattgt ctttactcag gaagtcggag ctcgattgat	11400
tgacattgcg tcttttagat accatactgt cgggtgtggac ggctcgcctg atgggtcttta	11460
ccttttatac ggtcgggtggg ttgctgggcg cttcagtcctt tacgtttcat gcggtcgggtg	11520
tcattctcat gccctacgtc tttatctcta agaatgtcgg tggagcgact taggtgcact	11580
gggtctttatg tttagaaatg tcgggtgtgat tacaggatc aaatgtcttt agctctggga	11640
aggtcgggtat cgatccaaag atccgggggt ttaaattgtt gtcaatgaac taggcacata	11700
gtaatatataa aaacatttta ttacaagccc cctcctttt tgtttggcgc ccaacaaaaa	11760
aaatcgccca aaagagcagc ttttcgggcg cggcgccctc atatatagcg caccaaacta	11820
tttcaacgcc ctggccaaat acctccccgt gtgactcttt tttaccttgg ccacatcacg	11880
cggcgctacct tcggccacca gcaaaccacc gtgattgcca ccttccggac ccagatcaat	11940
caccagtcg gaagatttaa taacttccaa attgtgttca ataataata gactgttgcc	12000
cttatccacc agcttgctca gcacgtgcag caaccgttc acatcatcaa aatgcaaacc	12060
cgctcgtcggc tcatccaaaa tatacaacgt ctttcccgtc gagcgccgtg acaattccgt	12120
cggcagcttc acacgtcgcg cttcaccacc actcagcgtc gtcgcattct gtcccagctg	12180
aatatagccc aaaccactt caaacagcgt cttcaacttt tcatgaataa tcggaatatt	12240
gttgaaaaat ttctgctcat cttcgaccgt catgttcagt acctcgaaa tatttttccc	12300
cttgtaatga atttccaaag cctgctcgtt gtagcggcgg cctttgcatt cgtcgcaatc	12360
cacatacacg tccggcagga agtgcatttc aattttggtc acaccatgc cctgacaggc	12420
ttcgcagcgg ccacccttca cattgaaact gaaacgcccg gccttgtagc cgcgcattct	12480
cgcttccggc acctgcgtga acagatcgcg aatgtaggta aacacgccgg ttaggtggc	12540
ggcgttggag cggggagtag ggccgatcgg cgactgatca atatcaatca ccttatcgag	12600
atattccagt ccgcgcagct ctttgtgttt gccgggaata tccttggcat tatgaaaatg	12660
ttgtgacaac gcgcgggcca gaatatcggc catcaacgtc gatttgccgc tgccggaaac	12720
gccggtgatg cacactaatt tteccagcgg aatgcgcacg ttgatatttt gtaggttgtg	12780
ggcgggtggca ccgcggattt caatatattt gccgttgccg cggcgggtact tgtcggcgcc	12840
ttcaatgaat tttttgccgc tcagatattg accggtcaat gacgttttat ttttaataat	12900

ttcctgaggt gtgccaaggg caacaatttc gccaccgtgt ttgccggcac caggccccac	12960
gtcaataaca taatcagcgg agcgaatcgt ttcttcatcg tgctcgacga cgatcacggt	13020
attgcctaata tcgcgcagcg ctttgagtgt gtctatgagt ttggagtgtg cgcgttggtg	13080
caagccaatg ctgggttcac cgaggatata gataacgccg accaaagatg aaccgatttg	13140
cgtggccaga cgaatgcgtt gcgcttcacc gccgcttaaa gtcgaagcag cgcgatctaa	13200
agtcaataa tccagacctt cattatgtaa aaaagtcagg cgttcgcgga ttcttttcat	13260
gatctgatgc gaaattttgg cttcgcgtac ggacatgacg tagacattat tttttgcat	13320
gctgttgccg cgggagtgg caccaccttt gccggccgcg tttttggcgc cagcaccctt	13380
cgcgccagca cccgcaccac caaccacaaa cccctcaaaa aatgcctgcg cttcttcaat	13440
gtcaacccc gtctgtcag aaatggattt gccgcgaatc gttacggcca gtgcaatttt	13500
gttcaaccgt ttcccgtag acgtcggaca atcaaagacg cgcatgtacg gttcgatttc	13560
cgagcggata tattccgact cggtttcttt gtagcgccgt tccaaattcg gtatcacgcc	13620
ttcatacgtc gtcacaaatt cacggatttt ggatgtcgag ttcatgccgc tgttgacgtc	13680
gaaagattct tcgccggtgc cgtaaaacac cagcttcagt tgcgcggcgg tcattttttt	13740
caccggttcg tccaaagaaa aaccgtattt ggccgccact gtcgccagaa tccgcagcat	13800
ccagccctga ttggaagacg tgcgtgacca gggctctgat gcaccctgat tgatgtcaa	13860
atttttattg ggaatgatca gttcagcgtc gacttcgagc ttggtgccca atccagtga	13920
ttccacgcag gcgccgtgcg ggctgttaaa cgaaaacagg cgcggttcaa tttccggcag	13980
gttgatgccg cagcgcggac aggcaagtg ctgactgaac agctgatctt tttcgtcgtt	14040
actgtcgtgc acaatcacca taccatcacc caaatccaag gcggtttcca gagattcgtg	14100
caagcggctg cggtttttgc gcagctcttt gtcaacaacc aagcgatcta caacaacatc	14160
aatggtatgt ttctttttct tatcaggac gagatcgagt gcttcttca tgctcatcat	14220
attcccgttg acgcgcacgc gcacaaaacc ggctttgcgc gtttcttcaa agacgtgttt	14280
gtgttcacct tttttgtcgc ggataatttg cgcgatgagc ataaatttcg tatccgcttt	14340
caggcgcaga atttgttcga ggatttggtt ggtggtttgt ttgctgactt tatcaccgca	14400
gttggggcag tgtggttggc cgatgcgggc gtagagcaaa cgcaggtaat cgtaaatttc	14460
ggtgacggtg ccgacggtgg atcggggatt gtgggatgtg gttttttgat cgatggagat	14520
ggcgggcgag aggccttcaa tgctgtcgac gtcaggcttg tccatcaggc cgaggaattg	14580
gcgggcgtag gaagacaggc ttctgacgta gcggcgctga ccttcggcat agatcgtatc	14640
aaaagccagg gaagattttc ccgagccgga caggccggtg atgacgacga gctggtcacg	14700
ggggatgtcc aggtgatata ttttcaggtt gtggacgcgg gcgcctttga tgatgatcga	14760

attttcacct gccataattg atcgttatga gacaacaaaa atttttagag caaagcccgt 14820
 aacctgcttt cgaggcagaa ttttcaaat actgccgagg cgaaggaaaa aattttgagg 14880
 aatactgtta gtatttcgag aaatTTTTta caagccgcag gcggattttg aaaattatga 14940
 tccggaatga ggttgcgggt tttactctag acgaacttcc gccagtctac tacttttttt 15000

 tgcgtaagtc aaccgtttgt gggcggggct gattcggttt tgtggtggtt tcgggagcag 15060
 catagatgta gcggaaaatt caaaaaactg gtataatatt gctacaacct atacaaacaa 15120
 aagcgtaaaa atcatgcatt tttcacgttt cggattttat ttccgtaacc gacgcatggt 15180
 agaacgtttc ttcgttctat tttgtgctat tttttctgct gtcttggttt tgtcgcttgt 15240
 tgccctggtg ctggtggctg acaaaattaa tatcaatccc attgtgcaca tcttgtttcg 15300
 ttttttcag cgacctttg tcagtgcgct gattctgtct ttttcgtca caaccttct 15360
 ttacgccgtt tttgttctgg tgcattcagt gcagcatcat accgtgtatt ggcagcgta 15420

 ttcgcagcga tatcatattc gcaagaaatc ccatattcac cgcagattgc gtcacattcc 15480
 cgcgagaca tcacataagc tgttggcgct cagttcactt tttgttgtgg ttaaaattgt 15540
 ttttgcagt tttgcctccg gttttttacc gcatgatgtt ttggcacaga cgtttgatec 15600
 gagcgagcag aaaagtcagt cgggtttggt ggcggcggtt tatgtccagg tgcttgattc 15660
 cgatgatttg tatatttga tttttatgtt gggccttttg ccgtggcggt ttctgatttt 15720
 tttcatggtt tttcgttcgc atatttttcc gcataagaat tttcattatg agagcgacaa 15780
 tctggatacg aatattgtca cttttgcggc ccggaagaag gcggagcagc ggcgcaaaaa 15840

 gccatcacct cgggcgggta ttgtaccttt gcatgatgca taacctatga attctgtttt 15900
 gcagaaaaaa ttagctggtc tgccgcatca acccggcgtc tatgtgtata aagacgcacg 15960
 ggggtgatgtt ttgtacgtgg ggaaggccaa agatttggcg aagcgctgc gatcgtattg 16020
 gcagtccggt cgctcgctgg tgccggacaa agctttgatg gtgagtcagg cggctgatat 16080
 cgatatcacg gtggtgagtt cggaacgga agcttttttg ctggaagcga gtttcattaa 16140
 aaaataccgg ccgcggttta atatttttt gaaagatgat aaaagttttt cgtatattaa 16200
 ggtgacgttg cgggaagaat ttccgagggt gctggtggtg cggcgctga cgcgcatgg 16260

 ca 16262
 <210> 28
 <211> 10
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic sequence

<400> 28	
aaaaaaaaaa	10
<210> 29	
<211> 10	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 29	
aaaaaaaaaa	10
<210> 30	
<211> 10	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 30	
aaaaaaaaaa	10
<210> 31	
<211> 25	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 31	
ctccgaaagt atcggggata aaggc	25
<210> 32	
<211> 25	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 32	
caccgaaatt tggagaggat aaggc	25
<210> 33	
<211>	
> 25	
<212> DNA	

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 33

ctccgaatta tcgggaggat aaggc 25

<210> 34

<211> 25

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 34

ccccgaatat aggggacaaa aaggc 25

<210> 35

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 35

gtctagacat acaggtggaa aggtgagagt aaagac 36

<210> 36

<211> 25

<212> DNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 36

ctccgtgaat acgtggggta aaggc 25

<210> 37

<211> 10

<212> RNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 37

aaaaaaaaaa 10

<210> 38

<211> 10

<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 38	
aaaaaaaaaa	10
<210> 39	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 39	
aaaaaaaaaa	10
<210> 40	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 40	
aaaaaaaaaa	10
<210> 41	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 41	
aaaaaaaaaa	10
<210> 42	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 42	
aaaaaaaaaa	10

<210> 43	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 43	
aaaaaaaaaa	10
<210> 44	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 44	
aaaaaaaaaa	10
<210> 45	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 45	
aaaaaaaaaa	10
<210> 46	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 46	
aaaaaaaaaa	10
<210> 47	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 47	

aaaaaaaaaa	10
<210> 48	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 48	
aaaaaaaaaa	10
<210> 49	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 49	
aaaaaaaaaa	10
<210> 50	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 50	
aaaaaaaaaa	10
<210> 51	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 51	
aaaaaaaaaa	10
<210> 52	
<211> 10	
<212> RNA	
<213> Artificial sequence	

<220><223> Synthetic sequence

<400> 52

aaaaaaaaaa

10

<210> 53

<211> 10

<212> RNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 53

aaaaaaaaaa

10

<210> 54

<211> 10

<212> RNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 54

aaaaaaaaaa

10

<210> 55

<211> 10

<212> RNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 55

aaaaaaaaaa

10

<210> 56

<211> 10

<212> RNA

<213> Artificial sequence

<220><223> Synthetic sequence

<400> 56

aaaaaaaaaa

10

<210> 57

<211> 10

<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 57	
aaaaaaaaaa	10
<210> 58	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 58	
aaaaaaaaaa	10
<210> 59	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 59	
aaaaaaaaaa	10
<210> 60	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 60	
aaaaaaaaaa	10
<210> 61	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 61	
aaaaaaaaaa	10
<210> 62	

<211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 62
 aaaaaaaaaa 10

<210> 63
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 63
 aaaaaaaaaa 10

<210> 64
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 64
 aaaaaaaaaa 10

<210> 65
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 65
 aaaaaaaaaa 10

<210> 66
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 66

aaaaaaaaaa	10
<210> 67	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 67	
aaaaaaaaaa	10
<210> 68	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 68	
aaaaaaaaaa	10
<210> 69	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 69	
aaaaaaaaaa	10
<210> 70	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 70	
aaaaaaaaaa	10
<210> 71	
<211> 10	
<212> RNA	
<213> Unknown	
<220><223> Synthetic sequence	

<400> 71	
aaaaaaaaaa	10
<210> 72	
<211> 10	
<212> RNA	
<213> Unknown	
<220><223> Synthetic sequence	
<400> 72	
aaaaaaaaaa	10
<210> 73	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 73	
aaaaaaaaaa	10
<210> 74	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 74	
aaaaaaaaaa	10
<210> 75	
<211> 10	
<212> RNA	
<213> Artificial sequence	
<220><223> Synthetic sequence	
<400> 75	
aaaaaaaaaa	10
<210> 76	
<211> 10	
<212> RNA	

<213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 76
 aaaaaaaaaa 10
 <210> 77
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 77
 aaaaaaaaaa 10

 <210> 78
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 78
 aaaaaaaaaa 10
 <210> 79
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 79
 aaaaaaaaaa 10
 <210> 80
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 80
 aaaaaaaaaa 10

 <210> 81

<211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 81
 aaaaaaaaaa 10
 <210> 82
 <211> 10
 <212> RNA
 <213> Artificial sequence
 <220><223> Synthetic sequence
 <400> 82
 aaaaaaaaaa 10
 <210> 83
 <211> 84
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic sequence
 <400> 83
 Met Ala Ser Met Ile Ser Ser Ser Ala Val Thr Thr Val Ser Arg Ala

 1 5 10 15
 Ser Arg Gly Gln Ser Ala Ala Met Ala Pro Phe Gly Gly Leu Lys Ser
 20 25 30
 Met Thr Gly Phe Pro Val Arg Lys Val Asn Thr Asp Ile Thr Ser Ile
 35 40 45
 Thr Ser Asn Gly Gly Arg Val Lys Cys Met Gln Val Trp Pro Pro Ile
 50 55 60
 Gly Lys Lys Lys Phe Glu Thr Leu Ser Tyr Leu Pro Pro Leu Thr Arg

 65 70 75 80
 Asp Ser Arg Ala

 <210> 84
 <211> 57

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 84

Met Ala Ser Met Ile Ser Ser Ser Ala Val Thr Thr Val Ser Arg Ala

1 5 10 15

Ser Arg Gly Gln Ser Ala Ala Met Ala Pro Phe Gly Gly Leu Lys Ser

20 25 30

Met Thr Gly Phe Pro Val Arg Lys Val Asn Thr Asp Ile Thr Ser Ile

35 40 45

Thr Ser Asn Gly Gly Arg Val Lys Ser

50 55

<210> 85

<211> 85

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 85

Met Ala Ser Ser Met Leu Ser Ser Ala Thr Met Val Ala Ser Pro Ala

1 5 10 15

Gln Ala Thr Met Val Ala Pro Phe Asn Gly Leu Lys Ser Ser Ala Ala

20 25 30

Phe Pro Ala Thr Arg Lys Ala Asn Asn Asp Ile Thr Ser Ile Thr Ser

35 40 45

Asn Gly Gly Arg Val Asn Cys Met Gln Val Trp Pro Pro Ile Glu Lys

50 55 60

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

65 70 75 80

Gly Arg Val Asn Cys

85

<210> 86

<211> 76

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic sequence

<400> 86

Met Ala Gln Val Ser Arg Ile Cys Asn Gly Val Gln Asn Pro Ser Leu

1 5 10 15

Ile Ser Asn Leu Ser Lys Ser Ser Gln Arg Lys Ser Pro Leu Ser Val

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75

<210> 87

<211> 76

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 87

Met Ala Gln Val Ser Arg Ile Cys Asn Gly Val Trp Asn Pro Ser Leu

1 5 10 15

Ile Ser Asn Leu Ser Lys Ser Ser Gln Arg Lys Ser Pro Leu Ser Val

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75

<210> 88

<211> 72

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 88

Met Ala Gln Ile Asn Asn Met Ala Gln Gly Ile Gln Thr Leu Asn Pro

1 5 10 15

Asn Ser Asn Phe His Lys Pro Gln Val Pro Lys Ser Ser Ser Phe Leu

20 25 30

Val Phe Gly Ser Lys Lys Leu Lys Asn Ser Ala Asn Ser Met Leu Val

35 40 45

Leu Lys Lys Asp Ser Ile Phe Met Gln Leu Phe Cys Ser Phe Arg Ile

50 55 60

Ser Ala Ser Val Ala Thr Ala Cys

65 70

<210> 89

<211> 69

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 89

Met Ala Ala Leu Val Thr Ser Gln Leu Ala Thr Ser Gly Thr Val Leu

1 5 10 15

Ser Val Thr Asp Arg Phe Arg Arg Pro Gly Phe Gln Gly Leu Arg Pro

20 25 30

Arg Asn Pro Ala Asp Ala Ala Leu Gly Met Arg Thr Val Gly Ala Ser

35 40 45

Ala Ala Pro Lys Gln Ser Arg Lys Pro His Arg Phe Asp Arg Arg Cys

50 55 60

Leu Ser Met Val Val

65

<210> 90

<211> 77

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 90

Met Ala Ala Leu Thr Thr Ser Gln Leu Ala Thr Ser Ala Thr Gly Phe

1 5 10 15

Gly Ile Ala Asp Arg Ser Ala Pro Ser Ser Leu Leu Arg His Gly Phe

20 25 30

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

35 40 45

Ser Val Thr Thr Ser Ala Arg Ala Thr Pro Lys Gln Gln Arg Ser Val

50 55 60

Gln Arg Gly Ser Arg Arg Phe Pro Ser Val Val Val Cys

65 70 75

<210> 91

<211> 57

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 91

Met Ala Ser Ser Val Leu Ser Ser Ala Ala Val Ala Thr Arg Ser Asn

1 5 10 15

Val Ala Gln Ala Asn Met Val Ala Pro Phe Thr Gly Leu Lys Ser Ala

20 25 30

Ala Ser Phe Pro Val Ser Arg Lys Gln Asn Leu Asp Ile Thr Ser Ile

35 40 45

Ala Ser Asn Gly Gly Arg Val Gln Cys

50 55

<210> 92

<211> 65

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 92

Met Glu Ser Leu Ala Ala Thr Ser Val Phe Ala Pro Ser Arg Val Ala
1 5 10 15

Val Pro Ala Ala Arg Ala Leu Val Arg Ala Gly Thr Val Val Pro Thr
20 25 30

Arg Arg Thr Ser Ser Thr Ser Gly Thr Ser Gly Val Lys Cys Ser Ala
35 40 45

Ala Val Thr Pro Gln Ala Ser Pro Val Ile Ser Arg Ser Ala Ala Ala
50 55 60

Ala

65

<210> 93

<211> 72

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400

> 93

Met Gly Ala Ala Ala Thr Ser Met Gln Ser Leu Lys Phe Ser Asn Arg
1 5 10 15

Leu Val Pro Pro Ser Arg Arg Leu Ser Pro Val Pro Asn Asn Val Thr
20 25 30

Cys Asn Asn Leu Pro Lys Ser Ala Ala Pro Val Arg Thr Val Lys Cys
35 40 45

Cys Ala Ser Ser Trp Asn Ser Thr Ile Asn Gly Ala Ala Ala Thr Thr
50 55 60

Asn Gly Ala Ser Ala Ala Ser Ser

65 70

<210> 94

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<220><221> misc_feature

<222> (4)..(4)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (8)..(8)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (11)..(11)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (15)..(15)

<223> Xaa can be any naturally occurring amino acid

<220><221> misc_feature

<222> (19)..(19)

<223> Xaa can be any naturally occurring amino acid

<400> 94

Gly Leu Phe Xaa Ala Leu Leu Xaa Leu Leu Xaa Ser Leu Trp Xaa Leu

1 5 10 15

Leu Leu Xaa Ala

20

<210> 95

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 95

Gly Leu Phe His Ala Leu Leu His Leu Leu His Ser Leu Trp His Leu

1 5 10 15

Leu Leu His Ala

20

<210> 96

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 96

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 97

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 97

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 98

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 98

Pro Ala Ala Lys Arg Val Lys Leu Asp

1 5

<210> 99

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 99

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro

1 5 10

<210> 100

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 100

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly

1 5 10 15
Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro

20 25 30
Arg Asn Gln Gly Gly Tyr

35

<210> 101

<211> 42

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 101

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu

1 5 10 15

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

20 25 30
Asp Glu Gln Ile Leu Lys Arg Arg Asn Val

35 40

<210> 102

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 102

Val Ser Arg Lys Arg Pro Arg Pro

1 5

<210> 103

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 103

Pro Pro Lys Lys Ala Arg Glu Asp

1 5

<210> 104

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 104

Pro Gln Pro Lys Lys Lys Pro Leu

1 5

<210> 105

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 105

Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro

1 5 10

<210> 106

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 106

Asp Arg Leu Arg Arg

1 5

<210> 107

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 107

Pro Lys Gln Lys Lys Arg Lys

1 5

<210> 108

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 108

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu

1 5 10

<210> 109

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 109

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg

1 5 10

<210> 110

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 110

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys

1 5 10 15

Lys Ser Lys Lys

20

<210> 111

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 111

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

1 5 10 15

Lys

<210> 112

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 112

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5 10

<210> 113

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 113

Arg Arg Gln Arg Arg Thr Ser Lys Leu Met Lys Arg

1 5 10

<210> 114

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 114

Gly Trp Thr Leu Asn Ser Ala Gly Tyr Leu Leu Gly Lys Ile Asn Leu

1 5 10 15

Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu

20 25

<210> 115

<211> 33

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 115

Lys Ala Leu Ala Trp Glu Ala Lys Leu Ala Lys Ala Leu Ala Lys Ala

1 5 10 15

Leu Ala Lys His Leu Ala Lys Ala Leu Ala Lys Ala Leu Lys Cys Glu

20 25 30

Ala

<210> 116

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 116

Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys

1 5 10 15

<210> 117

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 117

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5 10

<210> 118

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 118

Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5

<210> 119

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 119

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5 10

<210> 120

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 120

Arg Lys Lys Arg Arg Gln Arg Arg

1 5

<210> 121

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 121

Tyr Ala Arg Ala Ala Ala Arg Gln Ala Arg Ala

1 5 10

<210> 122

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 122

Thr His Arg Leu Pro Arg Arg Arg Arg Arg

1 5 10

<210> 123

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 123

Gly Gly Arg Arg Ala Arg Arg Arg Arg Arg Arg

1 5 10

<210> 124

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 124

Gly Ser Gly Gly Ser

1 5

<210> 125

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 125

Gly Gly Ser Gly Gly Ser

1 5

<210> 126

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 126

Gly Gly Gly Ser

1

<210> 127

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 127

Gly Gly Ser Gly

1

<210> 128

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 128

Gly Gly Ser Gly Gly

1 5

<210> 129

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 129

Gly Ser Gly Ser Gly

1 5

<210> 130

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 130

Gly Ser Gly Gly Gly

1 5

<210> 131

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 131

Gly Gly Gly Ser Gly

1 5

<210> 132

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 132

Gly Ser Ser Ser Gly

1 5

<210> 133

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 133

Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys

1 5 10 15

<210> 134

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 134

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg

1 5 10