(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0056329
(43) 공개일자 2021년05월18일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) *C12N 15/10* (2017.01)
C12N 15/113 (2010.01) *C12N 15/63* (2006.01)
C12Q 1/686 (2018.01)
(52) CPC특허분류
C12N 9/22 (2013.01)
C12N 15/102 (2013.01)
(21) 출원번호 10-2021-7004081
(22) 출원일자(국제) 2019년08월07일
심사청구일자 없음
(85) 번역문제출일자 2021년02월09일
(86) 국제출원번호 PCT/US2019/045582
(87) 국제공개번호 WO 2020/033601
국제공개일자 2020년02월13일
(30) 우선권주장
62/715,640 2018년08월07일 미국(US)
(뒷면에 계속)

(71) 출원인
더 브로드 인스티튜트, 인코퍼레이티드
미국 02142 매사 추세츠주 캠브리지 메인 스트리트 415
매사추세츠 인스티튜트 오브 테크놀로지
미국 02139 매사추세츠주 캠브리지 매사추세츠 애비뉴 77
(72) 발명자
장 펑
미국 02142 매사추세츠주 캠브리지 메인 스트리트 415 씨/오
스트레커 조나단
미국 02142 매사추세츠주 캠브리지 메인 스트리트 415 씨/오
(뒷면에 계속)
(74) 대리인
특허법인코리아나

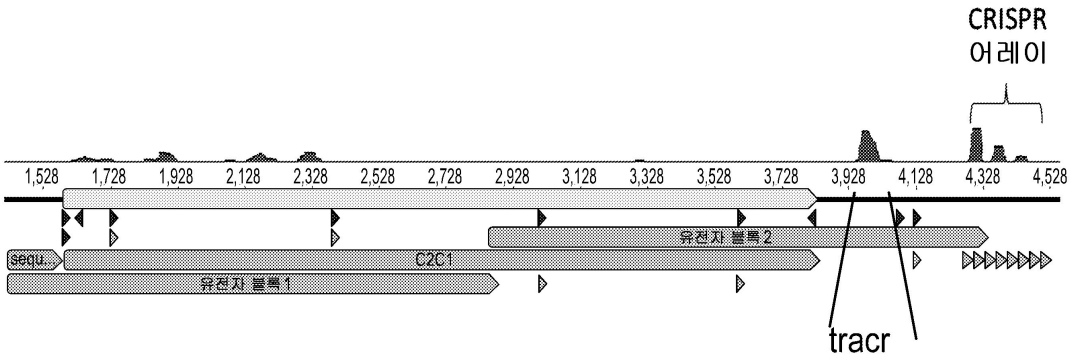
전체 청구항 수 : 총 85 항

(54) 발명의 명칭 **신규 CAS12B 효소 및 시스템**

(57) 요 약

본 개시는 핵산을 표적화하기 위한 시스템, 방법, 및 조성물을 제공한다. 특히, 본 발명은 신규 RNA-표적화 Cas12b 이펙터 단백질 및 적어도 하나의 표적화 핵산 성분 예컨대 가이드 가이드 RNA 또는 crRNA를 포함하는 비천연 발생 또는 조작된 RNA-표적화 시스템을 제공한다.

대 표 도



(52) CPC특허분류
C12N 15/113 (2013.01)
C12N 15/63 (2013.01)
C12Q 1/686 (2018.05)
C07K 2319/09 (2013.01)
C12N 2310/20 (2017.05)
C12Q 2527/101 (2019.08)

(72) 발명자
슬레이메이커 이안
미국 02142 매사추세츠주 캠브리지 메인 스트리트 415 씨/오
존스 사라
미국 02142 매사추세츠주 캠브리지 메인 스트리트 415 씨/오

(30) 우선권주장
62/744,080 2018년10월10일 미국(US)
62/751,196 2018년10월26일 미국(US)
62/794,929 2019년01월21일 미국(US)
62/831,028 2019년04월08일 미국(US)

명 세 서

청구범위

청구항 1

i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질, 및

ii) 표적 서열에 하이브리드화할 수 있는 가이드 서열을 포함하는 가이드

를 포함하는, 비천연 발생 또는 조작된 시스템.

청구항 2

제1항에 있어서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스 (*Alicyclobacillus kakegawensis*), 바실러스 (*Bacillus*) *sp.* V3-13, 바실러스 히사시이 (*Bacillus hisashii*), 렌티스파에리아 박테리움 (*Lentisphaeria bacterium*), 및 라세이엘라 세디미니스 (*Laceyella sediminis*)로 이루어진 군으로부터 선택되는 박테리아로부터 기원하는 것인 시스템.

청구항 3

제1항에 있어서, tracr RNA는 직접 반복부 서열의 5' 말단에서 crRNA에 융합되는 것인 시스템.

청구항 4

제1항에 있어서, 2종의 상이한 표적 서열 또는 동일한 표적 서열의 상이한 영역에 하이브리드화할 수 있는 둘 이상의 가이드 서열을 포함하는 것인 시스템.

청구항 5

제1항에 있어서, 가이드 서열은 원핵생물 세포 내 하나 이상의 표적 서열에 하이브리드화하는 것인 시스템.

청구항 6

제1항에 있어서, 가이드 서열은 진핵생물 세포 내 하나 이상의 표적 서열에 하이브리드화하는 것인 시스템.

청구항 7

제1항에 있어서, Cas12b 이펙터 단백질은 하나 이상의 핵 국재화 신호 (NLS)를 포함하는 것인 시스템.

청구항 8

제1항에 있어서, Cas12b 이펙터 단백질은 촉매적으로 불활성인 시스템.

청구항 9

제1항에 있어서, Cas12b 이펙터 단백질은 하나 이상의 기능성 도메인과 회합되는 것인 시스템.

청구항 10

제9항에 있어서, 하나 이상의 기능성 도메인은 하나 이상의 표적 DNA 서열을 절단하는 것인 시스템.

청구항 11

제10항에 있어서, 기능성 도메인은 하나 이상의 표적 서열의 전사 또는 번역을 변형시키는 것인 시스템.

청구항 12

제1항에 있어서, Cas12b 이펙터 단백질은 하나 이상의 기능성 도메인과 회합되고; Cas12b 이펙터 단백질은 RuvC 및/또는 Nuc 도메인 내에 하나 이상의 돌연변이를 포함하여서, 형성된 CRISPR 복합체는 표적 서열로 또는 인접하여 후성적 변형자 또는 전사 또는 번역 활성화 또는 억제 신호를 전달할 수 있는 것인 시스템.

청구항 13

제1항에 있어서, Cas12b 이펙터 단백질은 아데노신 디아미나제 또는 시티딘 디아미나제와 회합되는 것인 시스템.

청구항 14

제1항에 있어서, 재조합 주형을 더 포함하는 것인 시스템.

청구항 15

제14항에 있어서, 재조합 주형은 상동성-지정 복구 (HDR)에 의해 삽입되는 것인 시스템.

청구항 16

제1항에 있어서, tracr RNA를 더 포함하는 것인 시스템.

청구항 17

표 1 또는 2로부터의 Cas12b 이펙터 단백질을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제1 조절 엘리먼트, 및

i) a) 가이드 서열을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트, 및

b) tracr RNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제3 조절 엘리먼트; 또는

ii) 가이드 서열 및 tracr RNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트

를 포함하는, 하나 이상의 벡터를 포함하는, Cas12b 벡터 시스템.

청구항 18

제17항에 있어서, Cas12b 이펙터 단백질을 코딩하는 뉴클레오티드 서열은 진핵생물 세포에서의 발현을 위해 코돈 최적화되는 것인 벡터 시스템.

청구항 19

제17항 또는 제18항에 있어서, 단일 벡터에 포함되는 것인 벡터 시스템.

청구항 20

제17항 내지 제19항 중 어느 한 항에 있어서, 하나 이상의 벡터는 바이러스 벡터를 포함하는 것인 벡터 시스템.

청구항 21

제17항 내지 제20항 중 어느 한 항에 있어서, 하나 이상의 벡터는 하나 이상의 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 또는 헤르페스 심플렉스 바이러스 벡터를 포함하는 것인 벡터 시스템.

청구항 22

비천연 발생 또는 조작된 조성물의 하나 이상의 핵산 성분 및 Cas12b 이펙터 단백질을 전달하도록 구성된 전달 시스템으로서,

i) 표 1 또는 2로부터 선택된 Cas12b 이펙터 단백질,

ii) 하나 이상의 표적 서열에 하이브리드화할 수 있는 가이드 서열, 및

iii) tracr RNA

를 포함하는, 전달 시스템.

청구항 23

제22항에 있어서, 하나 이상의 벡터, 또는 하나 이상의 폴리뉴클레오티드 분자를 포함하고, 하나 이상의 벡터

또는 폴리뉴클레오티드 분자는 비천연 발생 또는 조작된 조성물의 하나 이상의 핵산 성분 및 Cas12b 이펙터 단백질을 코딩하는 하나 이상의 폴리뉴클레오티드 분자를 포함하는 것인 전달 시스템.

청구항 24

제22항 또는 제23항에 있어서, 리포솜(들), 입자(들), 엑소솜(들), 미세소포(들), 유전자총, 또는 바이러스 벡터(들)를 포함하는 전달 비히클을 포함하는 것인 전달 시스템.

청구항 25

치료적 치료 방법에서 사용을 위한, 제1항 내지 제16항의 비천연 발생 또는 조작된 시스템, 제17항 내지 제21항의 벡터 시스템, 또는 제22항 내지 제24항의 전달 시스템.

청구항 26

하나 이상의 관심 표적 서열을 변형시키는 방법으로서, 방법은 하나 이상의 표적 서열을

i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질,

ii) 하나 이상의 표적 서열에 하이브리드화할 수 있는 가이드 서열, 및

iii) tracr RNA

를 포함하는 하나 이상의 비천연 발생 또는 조작된 조성물과 접촉시켜서,

crRNA 및 tracr RNA과 복합체 형성된 Cas12b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성되는 것인 단계를 포함하고,

가이드 서열은 세포 내 하나 이상의 표적 서열과 서열-특이적 결합을 유도하여, 하나 이상의 표적 서열의 발현이 변형되는 것인, 변형 방법.

청구항 27

제26항에 있어서, 하나 이상의 표적 서열의 변형은 하나 이상의 표적 서열의 절단을 포함하는 것인 변형 방법.

청구항 28

제26항 또는 제27항에 있어서, 하나 이상의 표적 서열의 변형은 하나 이상의 표적 서열의 발현의 증가 또는 감소를 포함하는 것인 변형 방법.

청구항 29

제28항에 있어서, 조성물은 재조합 주형을 더 포함하고, 하나 이상의 표적 서열의 변형은 재조합 주형 또는 이의 부분의 삽입을 포함하는 것인 변형 방법.

청구항 30

제26항 내지 제29항 중 어느 한 항에 있어서, 하나 이상의 표적 서열은 원핵생물 세포에 존재하는 것인 변형 방법.

청구항 31

제26항 내지 제30항 중 어느 한 항에 있어서, 하나 이상의 표적 서열은 진핵생물 세포에 존재하는 것인 변형 방법.

청구항 32

하나 이상의 변형된 표적 서열을 포함하는 세포 또는 이의 자손으로서, 하나 이상의 표적 서열은 제23항 내지 제29항 중 어느 한 항의 방법에 따라 변형되었고, 임의로 치료적 T 세포 또는 항체-생성 B-세포이거나 또는 상기 세포는 식물 세포인 세포 또는 이의 자손.

청구항 33

제32항에 있어서, 세포는 원핵생물 세포인 세포.

청구항 34

제32항에 있어서, 세포는 진핵생물 세포인 세포.

청구항 35

제32항 내지 제34항 중 어느 한 항에 있어서, 하나 이상의 표적 서열의 변형은

적어도 하나의 유전자 산물의 변경된 발현을 포함하는 세포;

적어도 하나의 유전자 산물의 변경된 발현을 포함하는 것으로서, 적어도 하나의 유전자 산물의 발현이 증가된 것인 세포;

적어도 하나의 유전자 산물의 변경된 발현을 포함하는 것으로서, 적어도 하나의 유전자 산물의 발현이 감소된 것인 세포; 또는

내생성 또는 비내생성 생물학적 산물 또는 화학적 화합물을 생산 및/또는 분비하는 세포 또는 개체군

을 야기시키는 것인 세포.

청구항 36

제32항 또는 제35항에 있어서, 세포는 포유동물 세포 또는 인간 세포인 진핵생물 세포.

청구항 37

제32항 내지 제36항 중 어느 한 항에 따른 세포 또는 이의 자손을 포함하는 또는 이의 세포주.

청구항 38

제32항 내지 제36항 중 어느 한 항에 따른 하나 이상의 세포를 포함하는 다세포 유기체.

청구항 39

제32항 내지 제36항 중 어느 한 항에 따른 하나 이상의 세포를 포함하는 식물 또는 동물 모델.

청구항 40

제32항 내지 제36항 중 어느 한 항에 따른 세포, 또는 제37항의 세포주, 또는 제38항의 유기체, 또는 제39항의 식물 또는 동물 모델로부터의 유전자 산물.

청구항 41

제40항에 있어서, 발현되는 유전자 산물의 양은 변경된 발현을 갖지 않는 세포로부터의 유전자 산물의 양을 초과하거나 또는 그 미만인 유전자 산물.

청구항 42

표1 또는 2로부터의 단리된 Cas12b 이펙터 단백질.

청구항 43

제42항의 Cas12b 이펙터 단백질을 코딩하는 단리된 핵산.

청구항 44

제43항에 있어서, DNA이고 crRNA 및 tracr RNA를 코딩하는 서열을 더 포함하는 것인 단리된 핵산.

청구항 45

제43항 또는 제44항에 따른 핵산 또는 제42항의 Cas12b를 포함하는 단리된 진핵생물 세포.

청구항 46

i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질을 코딩하는 mRNA,

ii) 가이드 서열, 및

iii) tracr RNA

를 포함하는, 비천연 발생 또는 조작된 시스템.

청구항 47

제46항에 있어서, tracr RNA는 직접 반복부의 5' 말단에서 crRNA에 융합되는 것인 비천연 발생 또는 조작된 시스템.

청구항 48

표적화 도메인 및 아데노신 디아미나제, 시티딘 디아미나제, 또는 이의 촉매 도메인을 포함하는 부위 지정 염기 편집을 위한 조작된 조성물로서, 표적화 도메인은 Cas12b 이펙터 단백질, 또는 올리고뉴클레오티드-결합 활성을 유지하는 이의 단편, 및 가이드 분자를 포함하는 것인 조작된 조성물.

청구항 49

제48항에 있어서, Cas12b 이펙터 단백질은 촉매적으로 불활성인 조성물.

청구항 50

제48항에 있어서, Cas12b 이펙터 단백질은 표 1 또는 2로부터 선택되는 것인 조성물.

청구항 51

제50항에 있어서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스 (*Alicyclobacillus kakegawensis*), 바실러스 (*Bacillus*) *sp.* V3-13, 바실러스 히사시이 (*Bacillus hisashii*), 렌티스파에리아 박테리움 (*Lentisphaeria bacterium*), 및 라세이엘라 세디미니스 (*Laceyella sediminis*)로 이루어진 군으로부터 선택되는 박테리아로부터 기원하는 것인 조성물.

청구항 52

하나 이상의 관심 표적 올리고뉴클레오티드 내 아데노신 또는 시티딘을 변형시키는 방법으로서, 상기 하나 이상의 표적 올리고뉴클레오티드에, 제48항 내지 제51항 중 어느 한 항에 따른 조성물을 전달하는 단계를 포함하는 것인, 변형 방법.

청구항 53

제52항에 있어서, 병원성 T→C 또는 A→G 점 돌연변이를 함유하는 전사물에 의해 초래된 질환의 치료 또는 예방에서 사용을 위한 것인 변형 방법.

청구항 54

제48항 또는 제49항 중 어느 한 항의 방법으로 수득되고/되거나 제48항 내지 제51항 중 어느 한 항의 조성물을 포함하는 단리된 세포.

청구항 55

제54항에 있어서, 상기 진핵생물 세포, 바람직하게 인간 또는 비인간 동물 세포, 임의로 치료적 T 세포 또는 항체-생성 B-세포이거나 또는 상기 세포는 식물 세포인 세포 또는 이의 자손.

청구항 56

제50항 또는 제51항의 상기 변형된 세포 또는 이의 자손을 포함하는 비인간 동물.

청구항 57

제56항의 상기 변형된 세포를 포함하는 식물.

청구항 58

제56항 또는 제57항에 있어서, 요법, 바람직하게 세포 요법에서 사용을 위한 것인 변형된 세포.

청구항 59

표적 올리고뉴클레오티드 내 아데닌 또는 시토신을 변형시키는 방법으로서, 상기 표적 올리고뉴클레오티드에:

(a) 촉매적 불활성 Cas12b 단백질;

(b) 직접 반복부에 연결된 가이드 서열을 포함하는 가이드 분자; 및

(c) 아데노신 또는 시티딘 디아미나제 단백질 또는 이의 촉매 도메인

을 전달하는 단계를 포함하고;

상기 아데노신 또는 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 상기 촉매적 불활성 Cas12b 단백질에 공유적으로 또는 비공유적으로 연결되거나 또는 상기 가이드 분자는 전달 후 이에 연결되거나 적합화되고;

상기 가이드 분자는 상기 촉매적 불활성 Cas12b와 복합체를 형성하고 상기 복합체가 상기 표적 올리고뉴클레오티드에 결합하도록 유도하며, 상기 가이드 서열은 상기 표적 올리고뉴클레오티드 내 표적 서열과 하이브리드화할 수 있어서 올리고뉴클레오티드 듀플렉스를 형성하는 것인, 변형 방법.

청구항 60

제59항에 있어서, (A) 상기 시토신은 상기 올리고뉴클레오티드 듀플렉스를 형성하는 상기 표적 서열 밖에 존재하고, 상기 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 상기 올리고뉴클레오티드 듀플렉스 밖의 상기 시토신을 탈아미드화시키거나, 또는 (B) 상기 시토신은 상기 올리고뉴클레오티드 듀플렉스를 형성하는 상기 표적 서열 내에 존재하고, 상기 가이드 서열은 상기 시토신에 상응하는 위치에 비쌍형성 아데닌 또는 우라실을 포함하여서 상기 올리고뉴클레오티드 듀플렉스에 C-A 또는 C-U 미스매치를 야기시키고, 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 비쌍형성 아데닌 또는 우라실의 반대쪽 올리고뉴클레오티드 듀플렉스 내 시토신을 탈아미드화시키는 것인 변형 방법.

청구항 61

제59항에 있어서, 상기 아데노신 디아미나제 단백질 또는 이의 촉매 도메인은 올리고뉴클레오티드 듀플렉스 내 상기 아데닌 또는 시토신을 탈아미드화시키는 것인 변형 방법.

청구항 62

제59항에 있어서, Cas12b 단백질은 표 1 또는 2로부터 선택되는 것인 변형 방법.

청구항 63

제62항에 있어서, Cas12b 단백질은 알리시클로바실러스 카케가웬시스 (*Alicyclobacillus kakegawensis*), 바실러스 (*Bacillus*) *sp.* V3-13, 바실러스 히사시이 (*Bacillus hisashii*), 렌티스파에리아 박테리움 (*Lentisphaeria bacterium*), 및 라세이엘라 세디미니스 (*Laceyella sediminis*)로 이루어진 군으로부터 선택되는 박테리아로부터 기원하는 것인 변형 방법.

청구항 64

하나 이상의 시험관내 샘플에서 하나 이상의 표적 서열의 존재를 검출하기 위한 시스템으로서,

Cas12b 단백질;

하나 이상의 표적 서열과 일정 정도의 상보성을 갖도록 설계된 가이드 서열을 포함하고, Cas12b 단백질과 복합체를 형성하도록 설계된 적어도 하나의 가이드 폴리뉴클레오티드; 및

비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체

를 포함하고,

Cas12b 단백질은 부차적 뉴클레아제 활성을 나타내고 하나 이상의 표적 서열에 의해 활성화되면 올리고뉴클레오티드 기반 차폐성 구성체의 비표적 서열을 절단하는 것인, 시스템.

청구항 65

하나 이상의 시험관내 샘플에서 표적 폴리펩티드의 존재를 검출하기 위한 시스템으로서,

Cas12b 단백질;

각각이 하나 이상의 표적 폴리펩티드 중 하나에 결합하도록 설계된, 하나 이상의 검출 압타머로서, 각각의 검출 압타머는 차폐된 프로모터 결합 부위 또는 차폐된 프라이머 결합 부위 및 기폭제 서열 주형을 포함하는 것인 검출 압타머; 및

비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체

를 포함하는 것인 시스템.

청구항 66

제64항 또는 제65항에 있어서, 표적 서열 또는 기폭제 서열을 증폭시키기 위한 핵산 증폭 시약을 더 포함하는 것인 시스템.

청구항 67

제66항에 있어서, 핵산 증폭 시약은 등온 증폭 시약인 시스템.

청구항 68

제65항 내지 제67항 중 어느 한 항에 있어서, Cas12b 단백질은 표 1 또는 2로부터 선택되는 것인 시스템.

청구항 69

제68항에 있어서, Cas12b 단백질은 알리시클로바실러스 카케가웬시스 (*Alicyclobacillus kakegawensis*), 바실러스 (*Bacillus*) *sp.* V3-13, 바실러스 히사시이 (*Bacillus hisashii*), 렌티스파에리아 박테리움 (*Lentisphaeria bacterium*), 및 라세이엘라 세디미니스 (*Laceyella sediminis*)로 이루어진 군으로부터 선택되는 박테리아로부터 기원하는 것인 시스템.

청구항 70

하나 이상의 시험관내 샘플에서 하나 이상의 표적 서열을 검출하기 위한 방법으로서,

하나 이상의 샘플을

i) Cas12b 이펙터 단백질

ii) 하나 이상의 표적 서열과 일정 정도의 상보성을 갖도록 설계된 가이드 서열을 포함하고, Cas12b 이펙터 단백질과 복합체를 형성하도록 설계된 적어도 하나의 가이드 폴리뉴클레오티드; 및

iii) 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체

와 접촉시키는 단계를 포함하고,

상기 Cas12 이펙터 단백질은 부차적 뉴클레아제 활성을 나타내고 올리고뉴클레오티드-기반 차폐성 구성체의 비표적 서열을 절단하는 것인, 검출 방법.

청구항 71

제70항에 있어서, Cas12b 이펙터 단백질은 표 1 또는 2로부터 선택되는 것인 검출 방법.

청구항 72

제71항에 있어서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스 (*Alicyclobacillus kakegawensis*), 바실러스 (*Bacillus*) *sp.* V3-13, 바실러스 히사시이 (*Bacillus hisashii*), 렌티스파에리아 박테리움 (*Lentisphaeria bacterium*), 및 라세이엘라 세디미니스 (*Laceyella sediminis*)로 이루어진 군으로부터 선택되는 박테리아로부터 기원하는 것인 검출 방법.

청구항 73

효소 또는 리포터 모이어티의 불활성 제1 부분에 연결된 Cas12b 단백질을 포함하는 비천연 발생 또는 조작된 조성물로서, 효소 또는 리포터 모이어티는 효소 또는 리포터 모이어티의 상보성 부분과 접촉할 때 재구성되는 것인 조성물.

청구항 74

제73항에 있어서, 효소 또는 리포터 모이어티는 단백질가수분해 효소를 포함하는 것인 조성물.

청구항 75

제73항 또는 제74항에 있어서, Cas12b 단백질은 효소 또는 리포터 모이어티의 상보성 부분에 연결된 제1 Cas12b 단백질 및 제2 Cas12b 단백질을 포함하는 것인 조성물.

청구항 76

제73항에 있어서,

i) 제1 Cas12b 단백질과 복합체를 형성할 수 있고 표적 핵산의 제1 표적 서열에 하이브리드화할 수 있는 제1 가이드; 및

ii) 제2 Cas12b 단백질과 복합체를 형성할 수 있고, 표적 핵산의 제2 표적 서열에 하이브리드화할 수 있는 제2 가이드

를 더 포함하는 것인 조성물.

청구항 77

제73항 내지 제76항 중 어느 한 항에 있어서, 효소는 캐스파제를 포함하는 것인 조성물.

청구항 78

제73항 내지 제77항 중 어느 한 항에 있어서, 효소는 담배 식각 바이러스 (TEV)를 포함하는 것인 조성물.

청구항 79

표적 올리고뉴클레오티드를 함유하는 세포에서 단백질가수분해 활성을 제공하는 방법으로서,

a) 세포 또는 세포 개체군을

i) 단백질가수분해 효소의 불활성 부분에 연결된 제1 Cas12b 이펙터 단백질;

ii) 단백질가수분해 효소의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 단백질가수분해 효소의 단백질가수분해 활성은 단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 제2 Cas12b 이펙터 단백질;

iii) 제1 Cas12b 이펙터 단백질에 결합하고 표적 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드; 및

iv) 제2 Cas12b 이펙터 단백질에 결합하고 표적 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드

와 접촉시켜서,

단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉하고 단백질가수분해 효소의 단백질 가수분해 활성이 재구성되는 것인 단계를 포함하는, 제공 방법.

청구항 80

제79항에 있어서, 효소는 캐스파제인 제공 방법.

청구항 81

제80항에 있어서, 단백질가수분해 효소는 TEV 프로테아제이고, TEV 프로테아제의 단백질가수분해 활성은 재구성되어서, TEV 기질이 절단되고 활성화되는 것인 제공 방법.

청구항 82

제81항에 있어서, TEV 기질은 TEV 표적 서열을 함유하도록 조작된 프로캐스파제여서 TEV 프로테아제에 의한 절단은 프로캐스파제를 활성화시키는 것인 제공 방법.

청구항 83

관심 올리고뉴클레오티드를 함유하는 세포를 확인하기 위한 방법으로서, 방법은 세포 내 올리고뉴클레오티드를,

i) 단백질가수분해 효소의 불활성 제1 부분에 연결된 제1 Cas12b 이펙터 단백질;

ii) 단백질가수분해 효소의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 단백질가수분해 효소의 활성은 단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 제2 Cas12b 이펙터 단백질;

iii) 제1 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드;

iv) 제2 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드; 및

v) 검출가능하게 절단되는 리포터

를 포함하는 조성물과 접촉시키는 단계를 포함하고,

단백질가수분해 효소의 제1 부분 및 상보성 부분은 관심 올리고뉴클레오티드가 세포에 존재할 때 접촉되어서, 단백질가수분해 효소의 활성은 재구성되고 리포터를 검출가능하게 절단하는 것인, 확인 방법.

청구항 84

관심 올리고뉴클레오티드를 함유하는 세포를 확인하는 방법으로서, 방법은 세포 내 올리고뉴클레오티드를

i) 리포터의 불활성 제1 부분에 연결된 제1 Cas12b 이펙터 단백질;

ii) 리포터의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 리포터의 활성은 리포터의 제1 부분 및 상보성 부분이 접촉할 때 재구성되는 것인 제2 Cas12b 이펙터 단백질;

iii) 제1 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드;

iv) 제2 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드; 및

v) 리포터

를 포함하는 조성물과 접촉시키는 단계를 포함하고,

리포터의 제1 부분 및 상보성 부분은 관심 올리고뉴클레오티드가 세포에 존재할 때 접촉되어서, 리포터의 활성이 재구성되는 것인, 확인 방법.

청구항 85

제83항 또는 제84항에 있어서, 리포터는 형광 단백질 또는 발광 단백질인 확인 방법.

발명의 설명

기 술 분 야

[0001] **관련 출원의 교차 참조**

[0002] 본 출원은 2018년 8월 7일 출원된 미국 가출원 제62/715,640호, 2018년 10월 10일 출원된 미국 가출원 제62/744,080호, 2018년 10월 26일 출원된 미국 가출원 제62/751,196호, 2019년 1월 21일 출원된 미국 가출원 제62/794,929호, 및 2019년 4월 8일 출원된 미국 가출원 제62/831,028호의 우선권을 청구한다. 상기 확인된 출원의 전체 내용은 참조로 본 명세서에 완전히 편입된다.

[0003] **주정부 지원 연구에 관한 진술**

[0004] 본 발명은 미국 국립 보건원 (National Institutes of Health)이 수여하는 보조금 번호 MH110049, 및 HL141201 하의 정부 지원으로 만들어졌다. 정부는 본 발명에 대해 일정 권리를 갖는다.

[0005] **전자 서열 목록에 관한 참조**

[0006] 전자 서열 목록 ("BROD-2670_ST25.txt"; 크기는 879,558바이트이고 2019년 7월 25일에 생성되었음)의 내용은 그 전문이 참조로 본 명세서에 편입된다.

[0007] **기술 분야**

[0008] 본 명세서에 개시된 주제는 일반적으로, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), 및 이의 성분과 관련된 시스템, 방법, 및 조성물에 관한 것이다. 본 발명은 또한 일반적으로 대형 페이로드의 전달에 관한 것이고 대형 페이로드, 예컨대 CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), CRISPR 단백질 (예를 들어, Cas, C2c1), CRISPR-Cas 또는 CRISPR 시스템 또는 CRISPR-Cas 복합체, 이의 성분, 핵산 분자, 예를 들어, 이들을 포함하는 벡터를 전달하는데 적합한, 특히 지질 및 바이러스 입자를 사용한 신규 전달 입자, 및 또한 신규 바이러스 캡시드, 및 다른 양상들 중에서도, 모든 전술한 것의 용도를 포함한다. 추가로, 본 발명은 CRISPR-Cas 시스템 기반 요법 또는 치료법을 개발하거나 또는 설계하기 위한 방법에 관한 것이다.

배 경 기 술

[0009] 최근 게놈 시퀀싱 기법 및 분석 방법의 진보는 다양한 범위의 생물학적 기능 및 질환과 연관된 유전적 요인을 목록화하고 맵핑하는 능력을 상당히 가속화시켰다. 정밀한 게놈 표적화 기술은 합성 생물학, 생명공학 및 의학적 적용 분야를 진보시키기 위해서 뿐만 아니라, 개별 유전자 요소의 선택적 교란을 허용하여 인과적인 유전자 변이의 체계적인 역조작을 가능하게 하기 위해 필요하다. 게놈-편집 기술 예컨대 디자이너 징크 핑거, 전사 활성인자-유사 이펙터 (TALE; transcription activator-like effector), 또는 귀소성 메가뉴클레아제가 표적화된 게놈 교란을 일으키기 위해 이용가능하지만, 새로운 전략 및 분자 기전을 적용하고, 감당할만하고, 셋업이 용이하고, 규모확장가능하며, 진핵생물 게놈 내 복수개 위치의 표적화를 처리할 수 있는 신규한 게놈 조작 기술에 대한 필요성이 여전히 남아있다. 이것은 게놈 조작 및 생명공학의 새로운 적용분야에 주요한 자원을 제공하게 될 것이다.

[0010] 박테리아 및 고세균의 적응 면역의 CRISPR-Cas 시스템은 단백질 조성 및 게놈 유전자좌 구조의 극단적인 다양성을 나타낸다. CRISPR-Cas 시스템 유전자좌는 50 개 초과의 유전자 패밀리를 가지며, 유전자좌 구조의 빠른 진화 및 극단적인 다양성을 나타내는 엄격한 보편적 유전자는 존재하지 않는다. 지금까지는, 다면적인 접근법을 채택하여, 93 개의 Cas 단백질에 대해 약 395 개의 프로파일의 포괄적인 cas 유전자 식별이 있다. 분류는 특징적인 유전자 프로파일 및 유전자좌 구성의 특징을 포함한다. CRISPR-Cas 시스템의 새로운 분류가 제안되어 있으며, 여기서 이들 시스템은 광범위하게 2 개의 부류, 즉 다수 서브유닛 이펙터 복합체를 가지는 클래스 1 및 Cas9 단백질로 예시되는 단일-서브유닛 이펙터 모듈을 가지는 클래스 2로 나뉜다. 클래스 2 CRISPR-Cas 시스템과 연관된 신규 이펙터 단백질은 강력한 게놈 조작 도구로서 개발될 수 있으며, 추정상의 신규 이펙터 단백질 및 이의 조작 및 최적화의 예측이 중요하다. 신규 Cas12b 오솔로그 및 이의 용도가 바람직하다.

[0011] 본 출원에서 임의의 문헌의 인용 또는 확인은 상기 문헌이 본 출원에 대한 선행 기술로서 이용가능하다는 인정이 아니다.

발명의 내용

[0012] 일 양상에서, 본 개시는 i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질, 및 ii) 표적 서열과 하이브리드화할 수 있는 가이드 서열을 포함하는 가이드를 포함하는 비천연 발생 또는 조작된 시스템을 제공한다. 일부 구현예에서, 시스템은 tracr RNA를 더 포함한다.

[0013] 일부 구현예에서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스 (Alicyclobacillus kakegawensis), 바실러스 (Bacillus) sp. V3-13, 바실러스 히사시이 (Bacillus hisashii), 렌티스파에리아 박테리움 (Lentisphaeria bacterium), 및 라세이엘라 세디미니스 (Laceyella sediminis)로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다. 일부 구현예에서, tracr RNA는 직접 반복부의 5' 말단에서 crRNA와 융합된다. 일부 구현예에서, 시스템은 둘 이상의 crRNA를 포함한다. 일부 구현예에서, 가이드 서열은 원핵생물 세포 내 하나 이상의 표적 서열과 하이브리드화한다. 일부 구현예에서, 가이드 서열은 진핵생물 세포 내 하나 이상의 표적 서열과 하이브리드화한다. 일부 구현예에서, Cas12b 이펙터 단백질은 하나 이상의 핵 국재화 신호 (NLS)를 포함한다. 일부 구현예에서, Cas12b 이펙터 단백질은 촉매적으로 불활성이다. 일부 구현예에서, Cas12b 이펙터 단백질은 하나 이상의 기능성 도메인과 회합된다. 일부 구현예에서, 하나 이상의 기능성 도메인은 하나 이상의 표적 서열을 절단한다. 일부 구현예에서, 기능성 도메인은 하나 이상의 표적 서열의 전사 또는 번역을 변형시킨다. 일부 구현예에서, Cas12b 이펙터 단백질은 하나 이상의 기능성 도메인과 회합되고, Cas12b 이펙터 단백질은 RuvC 및/또는 Nuc 도메인 내에 하나 이상의 돌연변이를 함유하여서, 형성된 CRISPR 복합체가 표적 서열에서 또는 그에 인접하여 후생유전적 변형인자 또는 전사 또는 번역 활성화 또는 억제 신호를 전달할 수 있다. 일부 구현예에서, Cas12b 이펙터 단백질은 아데노신 디아미나제 또는 시티딘 디아미나제와 회합된다. 일부 구현예에서, 시스템은 재조합 주형을 더 포함한다. 일부 구현예에서, 재조합 주형은 상동성-지정 복구 (homology-directed repair) (HDR)에 의해 삽입된다.

[0014] 다른 양상에서, 본 개시는 표 1 또는 2로부터의 Cas12b 이펙터 단백질을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제1 조절 엘리먼트, 및 i) a) 가이드 서열을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트, 및 b) tracr RNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제3 조절 엘리먼트, 또는 ii) 가이드 서열 및 tracr RNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함하는 하나 이상의 벡터를 포함하는, Cas12b 벡터 시스템을 제공한다.

[0015] 일부 구현예에서, Cas12b 이펙터 단백질을 코딩하는 뉴클레오티드 서열은 진핵생물 세포에서 발현을 위해 코돈-최적화된다. 일부 구현예에서, 시스템은 단일 벡터에 포함된다. 일부 구현예에서, 하나 이상의 벡터는 바이러스 벡터를 포함한다. 일부 구현예에서, 하나 이상의 벡터는 하나 이상의 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연합 또는 헤르페스 심플렉스 바이러스 벡터를 포함한다.

[0016] 다른 양상에서, 본 개시는 Cas12b 이펙터 단백질 및 i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질, ii) 하나 이상의 표적 서열과 하이브리드화할 수 있는 3' 가이드 서열, 및 iii) tracr RNA를 포함하는 비천연 발생 또는 조작된 조성물의 하나 이상의 핵산 성분을 전달하도록 구성된 전달 시스템을 제공한다.

[0017] 일부 구현예에서, 본 발명의 벡터 시스템은 하나 이상의 벡터, 또는 하나 이상의 폴리뉴클레오티드 분자를 포함하고, 하나 이상의 벡터 또는 폴리뉴클레오티드 분자는 Cas12b 이펙터 단백질 및 비-천연 발생 또는 조작된 조성물의 하나 이상의 핵산 성분을 코딩하는 하나 이상의 폴리뉴클레오티드 분자를 포함한다. 일부 구현예에서, 전달 시스템은 리포솜(들), 입자(들), 엑소솜(들), 미세소포(들), 유전자총, 또는 바이러스 벡터 (들)를 포함하는 전달 비히클을 포함한다.

[0018] 다른 양상에서, 본 개시는 치료적 치료 방법에서 사용을 위한, 본 명세서의 비천연 발생 또는 조작된 시스템, 본 명세서의 벡터 시스템, 또는 본 명세서의 전달 시스템을 제공한다.

[0019] 다른 양상에서, 본 개시는 하나 이상의 관심 표적 서열을 변형시키는 방법을 제공하고, 이 방법은 하나 이상의 표적 서열을 i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질, ii) 표적 DNA 서열과 하이브리드화할 수 있는 3' 가이드 서열, 및 iii) tracr RNA를 포함하여서, crRNA 및 tracr RNA와 복합체형성된 Cas12b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성된 것인 하나 이상의 비천연 발생 또는 조작된 조성물과 접촉시키는 단계를 포함하고, 여기서 가이드 서열이 세포 내 하나 이상의 표적 서열과의 서열-특이적 결합을 유도하여, 하나 이상의 표적 서열의 발현이 변형된다. 일부 구현예에서, 표적 유전자의 발현의 변형은 하나 이상의 표적 서열의 절단을 포함한다. 일부 구현예에서, 표적 유전자의 발현의 변형은 하나 이상의 표적 서열의 발현의 증가 또는 감소를 포함한다. 일부 구현예에서, 조성물은 재조합 주형을 더 포함하고, 하나 이상의 표적 서열의 변형은 재조합 주형 또는 이의 일부의 삽입을 포함한다. 일부 구현예예서, 하나 이상의 표적 서열은 원핵생물 세포

내에 존재한다. 일부 구현예에서, 하나 이상의 표적 서열은 진핵생물 세포 내에 존재한다.

[0020] 다른 양상에서, 본 개시는 하나 이상의 변형된 표적 서열을 포함하는 세포 또는 이의 자손, 임의로 치료적 T 세포 또는 항체-생산 B-세포를 제공하고, 여기서 하나 이상의 표적 서열은 본 명세서에 개시된 방법에 따라 변형되었거나 또는 상기 세포는 식물 세포이다. 일부 구현예에서, 세포는 원핵생물 세포이다. 일부 구현예에서, 세포는 진핵생물 세포이다. 일부 구현예에서, 하나 이상의 표적 서열의 변형은 하기의 결과를 야기시킨다: 적어도 하나의 유전자 산물의 변경된 발현을 포함하는 세포; 적어도 하나의 유전자 산물의 발현이 증가된 것인, 적어도 하나의 유전자 산물의 변경된 발현을 포함하는 세포; 적어도 하나의 유전자 산물의 발현이 감소된 것인, 적어도 하나의 유전자 산물의 변경된 발현을 포함하는 세포; 내생성 또는 비내생성 생물학적 산물 또는 화학적 화합물을 생산 및/또는 분비하는 세포 또는 개체군. 일부 구현예에서, 세포는 포유동물 세포 또는 인간 세포이다. 다른 양상에서, 본 개시는 본 명세서의 세포를 포함하는 세포주, 또는 이의 자손을 제공한다.

[0021] 다른 양상에서, 본 개시는 본 명세서의 하나 이상의 세포를 포함하는 다세포 유기체를 제공한다.

[0022] 다른 양상에서, 본 개시는 본 명세서의 하나 이상의 세포를 포함하는 식물 또는 동물 모델을 제공한다.

[0023] 다른 양상에서, 본 개시는 본 명세서의 세포, 세포주, 유기체, 또는 식물, 또는 동물 모델로부터의 유전자 산물을 제공한다. 일부 구현예에서, 발현되는 유전자 산물의 양은 변경된 발현을 갖지 않는 세포로부터의 유전자 산물의 양을 초과하거나 또는 그 미만이다.

[0024] 다른 양상에서, 본 개시는 표 1 또는 2로부터의 단리된 Cas12b 이펙터 단백질을 제공한다.

[0025] 다른 양상에서, 본 개시는 Cas12b 이펙터 단백질을 코딩하는 단리된 핵산을 제공한다. 일부 구현예에서, 단리된 핵산은 DNA이고 crRNA 및 tracr RNA를 코딩하는 서열을 더 포함한다.

[0026] 다른 양상에서, 본 개시는 본 명세서의 핵산 또는 Cas12b 단백질을 포함하는 단리된 진핵생물 세포를 제공한다.

[0027] 다른 양상에서, 본 개시는 i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질을 코딩하는 mRNA, ii) 가이드 서열, 및 iii) tracr RNA를 포함하는 비천연 발생 또는 조작된 시스템을 제공한다. 일부 구현예에서, tracr RNA는 직접 반복부의 5' 말단에서 crRNA에 융합된다.

[0028] 다른 양상에서, 본 개시는 표적화 도메인 및 아데노신 디아미나제, 시티딘 디아미나제, 도는 이의 촉매 도메인을 포함하는 부위 지정 염기 편집을 위한 조작된 시스템을 제공하고, 여기서 표적화 도메인은 Cas12b 이펙터 단백질, 또는 올리고뉴클레오티드-결합 활성을 유지하는 이의 단편 및 가이드 분자를 포함한다. 일부 구현예에서, Cas12b 이펙터 단백질은 촉매적으로 불활성이다. 일부 구현예에서, Cas12b 이펙터 단백질은 하기 표 1 또는 2로부터 선택된다. 일부 구현예에서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp. V3-13, 바실러스 히사시이, 렌티스파에리아 박테리움, 및 라세이엘라 세디미니스로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다.

[0029] 다른 양상에서, 본 개시는 하나 이상의 관심 표적 올리고뉴클레오티드 내 아데노신 또는 시티딘을 변형시키는 방법을 제공하고, 이 방법은 상기 하나 이상의 표적 올리고뉴클레오티드로 본 명세서의 조성물을 전달하는 단계를 포함한다. 일부 구현예서, 병원성 T→C 또는 A→G 점 돌연변이를 함유하는 전사물에 의해 초래된 질환의 치료 또는 예방에서 사용을 위한다. 다른 양상에서, 본 개시는 본 명세서의 방법으로부터 수득되고/되거나 본 명세서의 조성물을 포함하는 단리된 세포를 제공한다. 일부 구현예에서, 상기 진핵생물 세포, 바람직하게 인간 또는 비인간 동물 세포, 임의로 치료적 T 세포 또는 항체-생산 B-세포 또는 여기서 상기 세포는 식물 세포이다.

[0030] 다른 양상에서, 본 개시는 상기 변형된 세포 또는 이의 자손을 포함하는 비인간 동물을 제공한다.

[0031] 다른 양상에서, 본 개시는 본 명세서의 변형된 세포를 포함하는 식물을 제공한다.

[0032] 다른 양상에서, 본 개시는 요법, 바람직하게 세포 요법에서 사용을 위한 변형된 세포를 제공한다.

[0033] 다른 양상에서, 본 개시는 표적 올리고뉴클레오티드 내 아데닌 또는 시토신을 변형시키는 방법을 제공하고, 방법은 상기 표적 올리고뉴클레오티드로, 촉매적 불활성 Cas12b 단백질; 직접 반복부에 연결된 가이드 서열을 포함하는 가이드 분자; 및 아데노신 또는 시티딘 디아미나제 단백질 또는 이의 촉매 도메인을 전달하는 단계를 포함하고, 상기 아데노신 또는 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 전달 이후에 상기 촉매적 불활성 Cas12b 단백질에 공유적으로 또는 비공유적으로 연결되거나 또는 상기 가이드 분자는 그에 연결되도록 조정되고, 상기 가이드 분자는 상기 촉매적 불활성 Cas12b와 복합체를 형성하고 상기 복합체가 상기 표적 올리고뉴

클레오티드에 결합하도록 유도하며, 상기 가이드 서열은 상기 표적 올리고뉴클레오티드 서열 내 표적 서열과 하이브리드화할 수 있어서 올리고뉴클레오티드 듀플렉스를 형성한다.

[0034] 일부 구현예에서, (A) 상기 시토신은 상기 올리고뉴클레오티드 듀플렉스를 형성하는 상기 표적 서열 밖에 존재하고, 상기 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 상기 RNA 듀플렉스 밖의 상기 시토신을 탈아민화시키거나, 또는 (B) 상기 시토신은 상기 RNA 듀플렉스를 형성하는 상기 표적 서열 내에 있고, 상기 가이드 서열은 상기 시토신에 상응하는 위치에 비-쌍형성 아데닌 또는 우라실을 포함하여 그 결과로 상기 올리고뉴클레오티드 듀플렉스에 C-A 또는 C-U 미스매치를 야기시키고, 시토신 디아미나제 단백질 또는 이의 촉매 도메인은 비-쌍형성 아데닌 또는 우라실 반대쪽 올리고뉴클레오티드 듀플렉스 내 시토신을 탈아민화시킨다. 일부 구현예에서, 상기 아데노신 디아미나제 단백질 또는 이의 촉매 도메인은 올리고뉴클레오티드 듀플렉스 내 상기 아데닌 또는 시토신을 탈아민화시킨다. 일부 구현예에서, Cas12b 이펙터 단백질은 하기 표 1 또는 2로부터 선택된다. 일부 구현예에서, Cas12b 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp. V3-13, 바실러스 히사시이, 렌티스파에리아 박테리움, 및 라세이엘라 세디미니스로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다.

[0035] 다른 양상에서, 본 개시는 Cas12b 단백질; 표적 서열과 일정 정도의 상보성을 갖도록 설계된 가이드 서열을 포함하고, Cas12b와 복합체를 형성하도록 설계된 적어도 하나의 가이드 폴리뉴클레오티드; 및 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체를 포함하는, 하나 이상의 시험관내 샘플에서 핵산 표적 서열의 존재를 검출하기 위한 시스템을 제공하고, 여기서 Cas12b는 부차적 뉴클레아제 활성을 나타내고 표적 서열에 의해 활성화될 때 올리고뉴클레오티드 기반 차폐성 구성체의 비표적 서열을 절단한다.

[0036] 다른 양상에서, 본 개시는 Cas12b 단백질; 각각이 하나 이상의 표적 폴리펩티드 중 하나에 결합하도록 설계된 하나 이상의 검출 압타머로서, 각각은 차폐된 프로모터 결합 부위 또는 차폐된 프라이머 결합 부위 및 기폭제 서열 주형을 포함하는 것인 검출 압타머; 및 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체를 포함하는, 하나 이상의 시험관내 샘플에서 하나 이상의 표적 폴리펩티드의 존재를 검출하기 위한 시스템을 제공한다.

[0037] 일부 구현예에서, 시스템은 표적 서열 또는 기폭제 서열을 증폭시키기 위한 핵산 증폭 시약을 더 포함한다. 일부 구현예에서, 핵산 증폭 시약은 등온 증폭 시약이다. 일부 구현예에서, Cas12b 단백질은 하기 표 1 또는 2로부터 선택된다. 일부 구현예에서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp. V3-13, 바실러스 히사시이, 렌티스파에리아 박테리움, 및 라세이엘라 세디미니스로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다.

[0038] 다른 양상에서, 본 개시는 하나 이상의 시험관내 샘플에서 핵산 서열을 검출하기 위한 방법을 제공하고, 방법은 하나 이상의 샘플을 i) Cas12b 단백질, ii) 표적 서열과 일정 정도의 상보성을 갖도록 설계된 가이드 서열을 포함하고, Cas12b 단백질과 복합체를 형성하도록 설계된 적어도 하나의 가이드 폴리뉴클레오티드; 및 iii) 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체와 접촉시키는 단계를 포함하고, 상기 Cas12b 단백질은 부차적 뉴클레아제 활성을 나타내고 올리고뉴클레오티드 기반 차폐성 구성체의 비표적 서열을 절단한다.

[0039] 일부 구현예에서, Cas12b 단백질은 하기 표 1 또는 2로부터 선택된다. 일부 구현예에서, Cas12b 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp. V3-13, 바실러스 히사시이, 렌티스파에리아 박테리움, 및 라세이엘라 세디미니스로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다. 다른 양상에서, 본 개시는 효소 또는 리포터 모이어티의 불활성 제1 부분에 연결된 Cas12b 단백질을 포함하는 비천연 발생 또는 조작된 조성물을 제공하고, 효소 또는 리포터 모이어티는 효소 또는 리포터 모이어티의 상보성 부분과 접촉될 때 재구성된다. 일부 구현예에서, 효소 또는 리포터 모이어티는 단백질가수분해 효소를 포함한다. 일부 구현예에서, Cas12 단백질은 제1 Cas12b 단백질 및 효소 또는 리포터 모이어티의 상보성 부분에 연결된 제2 Cas12b 단백질을 포함한다. 일부 구현예에서, 조성물은 i) 제1 Cas12b 단백질과 복합체를 형성할 수 있고 표적 핵산의 제1 표적 서열과 하이브리드화할 수 있는 제1 가이드; 및 ii) 제2 Cas12b 단백질과 복합체를 형성할 수 있고, 표적 핵산 상의 제2 표적 서열과 하이브리드화할 수 있는 제2 가이드를 더 포함한다. 일부 구현예에서, 단백질가수분해 효소는 캐스파제를 포함한다. 일부 구현예에서, 단백질가수분해 효소는 담배 식각 바이러스 (TEV)를 포함한다.

[0040] 일부 양상에서, 본 개시는 표적 올리고뉴클레오티드를 함유하는 세포에서 단백질가수분해 활성을 제공하는 방법을 제공하고, 방법은 a) 세포 또는 세포 개체군을, i) 단백질가수분해 효소의 불활성 부분에 연결된 제1 Cas12b 이펙터 단백질; ii) 단백질가수분해 효소의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 단백질가수분

해 효소의 단백질가수분해 활성은 단백질 가수분해 효소의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 이펙터 단백질; iii) 제1 Cas12b 이펙터 단백질에 결합하고 표적 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드; 및 iv) 제2 Cas12b 이펙터 단백질에 결합하고 표적 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드와 접촉시켜서, 단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉하고 단백질가수분해 효소의 단백질가수분해 활성이 재구성되는 것인 단계를 포함한다.

[0041] 일부 구현예에서, 단백질가수분해 효소는 캐스파제를 포함한다. 일부 구현예에서, 단백질가수분해 효소는 TEV 프로테아제이고, TEV 프로테아제의 단백질가수분해 활성은 재구성되어서, TEV 기질이 절단되고 활성화된다.

일부 구현예에서, TEV 기질은 TEV 표적 서열을 함유하도록 조작된 프로캐스파제이고 그리하여 TEV 프로테아제에 의한 절단은 프로캐스파제를 활성화시킨다.

[0042] 다른 양상에서, 본 개시는 관심 올리고뉴클레오티드를 함유하는 세포를 확인하는 방법을 제공하고, 방법은 세포 내 올리고뉴클레오티드를, i) 단백질가수분해 효소의 불활성 제1 부분에 연결된 제1 Cas12b 이펙터 단백질; ii) 단백질가수분해 효소의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 단백질가수분해 효소의 활성은 단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉할 때 재구성되는 것인 이펙터 단백질; iii) 제1 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드; iv) 제2 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드; 및 v) 검출가능하게 절단되는 리포터를 포함하는 조성물과 접촉시키는 단계를 포함하고, 관심 올리고뉴클레오티드가 세포에 존재할 때 단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉하고, 그리하여 단백질가수분해 효소의 활성이 재구성되고 리포터를 검출가능하게 절단한다.

[0043] 다른 양상에서, 본 개시는 관심 올리고뉴클레오티드를 함유하는 세포를 확인하는 방법을 제공하고, 방법은 세포 내 올리고뉴클레오티드를, i) 리포터의 불활성 제1 부분에 연결된 제1 Cas12b 이펙터 단백질; ii) 리포터의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 리포터의 활성은 리포터의 제1 부분 및 상보성 부분이 접촉할 때 재구성되는 것인 이펙터 단백질; iii) 제1 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드; iv) 제2 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드; 및 v) 리포터를 포함하는 조성물과 접촉시키는 단계를 포함하고, 관심 올리고뉴클레오티드가 세포에 존재할 때 리포터의 제1 부분 및 상보성 부분이 접촉하고, 그리하여 리포터의 활성이 재구성된다. 일부 구현예에서, 리포터는 형광성 단백질 또는 발광성 단백질이다.

[0044] 예시 구현예의 이들 및 다른 양상, 목적, 특징 및 이점은 예시된 예시 구현예의 하기 상세한 설명을 고려하면 당업자에게 명백해 질 것이다.

도면의 간단한 설명

[0045] 본 발명의 특성 및 장점의 이해는 본 발명의 원리를 이용할 수 있는, 실례적인 구현예를 제시하는 하기 상세한 설명, 및 첨부된 도면을 참조로 하여 수득될 것이다:

도 1은 피시스파에라에 박테리움 (Phycisphaerae bacterium) CRISPR-C2c1 유전자좌를 도시한다. 소형 RNAseq는 tracrRNA의 위치 및 성숙한 crRNA의 아키텍처를 밝혀주었다.

도 2A-2C 는 예상 tracrRNA (**도 2A**) (SEQ ID NO:1-11) 및 Tracer#1 (**도 2B**) 및 Tracer #5 (**도 2C**) (SEQ ID NO:12, 656, 및 13)에 대한 직접 반복부 (적색)와 tracer (녹색)의 듀플렉스의 예상 배수를 도시한다.

도 3A 는 Seqlogos에대한 PAM 스크린의 결과가 가장 느슨한 예상 PAM에 대해 제공된 것을 도시하고 **도 3B** 는 가장 엄격한 예상 PAM을 도시한다.

도 4 는 TTH (H = A, T 또는 C)로서 PhbC2c1 PAM의 생체내 확인을 도시한다. 세포는 인식가능한 프로토스페이서의 5'에 위치된 상이한 PAM 서열을 코딩하는 플라스미드 DNA로 형질전환되었다.

도 5 는 Cpf1 닉카제를 사용한 서열 특이적 닉카제 증폭을 도시한다.

도 6 은 압타머 발색을 예시한다.

도 7 은 플란크토마이세테스 (Planctomycetes) CRISPR-C2c1 유전자좌를 도시한다. 소형 RNAseq는 tracrRNA의 위치 및 성숙한 crRNA의 아키텍처를 밝혀주었다.

도 8A 는 Seqlogos에대한 PAM 결과가 가장 느슨한 예상 PAM에 대해 제공된 것을 도시하고 **도 8B** 는 가장 엄격한 예상 PAM을 도시한다 (B). 스크린은 플란크토리세테스 (Planctolycetes)에 대한 PAM가 TTR (R = G 또는 A)이라

는 것을 도시한다.

도 9 는 TTR (R = G 또는 A)로서 C2c1 PAM의 생체내 확인을 도시한다. 세포는 인식가능한 프로토스페이서의 5'에 위치된 상이한 PAM 서열을 코딩하는 플라스미드 DNA로 형질전환되었다.

도 10 은 crRNA-tracrRNA 복합체와 C2c1의 단리를 위한 플라스미드의 예를 도시한다. 플라스미드는 PhyciC2c1 및/또는 tracrRNA 및/또는 CRISPR 어레이를 함유한다. 프로세스된 crRNA 및 tracrRNA는 C2c1과 복합체 형성하게 될 것이고 C2c1 단백질과 함께 동시-정제될 수 있다 (C2c1-RNA 복합체).

도 11A 는 단백질 풀다운 어세이에서 PhyciC2c1 및 PlancC2c1의 밴드를 도시한다. RNase 및 DNase 분해 실험을 수행하였고, **도 11B** 에서 RNA가 PhysiC2c1 단백질에 존재한다는 것이 입증되었다 (PhyC2c1 단백질은 DNase 분해가 아니라 RNase 분해에 감수성이었음). PhysiC2c1 단백질 중 RNA의 존재는 **도 11C** 에서 더욱 확인되었다. 동시 정제된 RNA의 크기는 crRNA와 일치하지만 118 nt 예상 tracrRNA에 비해 더 크게 보인다.

도 12 는 시험관내 절단 실험의 조건 및 결과를 제공하며, 여기서 PhysiC2c1-RNA 복합체가 CRISPR 어세이의 제1 가이드와 매칭되는 프로토스페이서를 함유하는 DNA를 절단할 수 있다는 것이 입증되었다.

도 13 은 상이한 sgRNA를 도시한다. 이. 콜라이 (E.coli)에서 발현된 BhCas12b 유전자좌로부터의 소형 RNA-seq는 tracrRNA 및 crRNA를 밝혀주었다. sgRNA 변이체를 형성하기 위한 tracrRNA 및 crRNA의 융합의 다이아그램 (SEQ ID NO:14-29).

도 14 는 상이한 표적 부위에 대한, 플라스미드 형질감염 이후, 도 13의 상이한 sgRNA로 수득된 indel 백분율을 도시한다. 사용된 Cas12b는 바실리서 히사시이 (Bacillus hisashii) 균주 C4로부터의 것이었다. HEK293 세포에서 BhCas12b 및 sgRNA의 발현은 다중 게놈 부위에서 indel 돌연변이를 발생시킨다.

도 15A-15C 는 각각 Ls, Ak, 및 Bv로부터의 Cas12b 오솔로그를 사용한 정제된 단백질 및 RNA에 의한 PAM 발굴, 시험관내 절단을 도시한다 (도 15A - SEQ ID NO:30 및 657; 도 15B - SEQ ID NO:31 및 658; 도 15C - SEQ ID NO:32 및 659). **도 15D-15E** 는 각각 Phyci 및 Planc로부터의 Cas12B 오솔로그를 사용한 정제된 단백질 RNA에 의한 시험관내 절단을 도시한다.

도 16 은 정제된 AmCas12b (AmC2C1) 단백질 및 소형 RNAseq로부터의 상이한 예상 tracr RNA에 의한 시험관내 절단 어세이를 도시한다.

도 17A-17E 는 AmC2C1에 대한 sgRNA 디자인을 도시한다. (**도 17A** - SEQ ID NO:33 및 660; **도 17B** - SEQ ID NO:34 및 661; **도 17C** - SEQ ID NO:35; **도 17D** - SEQ ID NO:36; **도 17E** - SEQ ID NO:37).

도 18 은 sgRNA 효율의 비교를 위한 AmC2C1에 의한 시험관내 절단을 도시한다.

도 19 는 AmC2C1 RuvC 돌연변이체의 활성을 도시한다.

도 20 은 시험관내 PAM 스크린에 의한 Cas12b 오솔로그에 대한 PAM의 결정을 도시한다.

도 21A 는 소형 RNAseq tracr 예상을 도시한다. **도 21B** 는 시험관내 스크리닌으로부터 BhC2C1 (바실러스 히사시이 Cas12b) PAM을 도시한다. **도 21C** 는 BhC2C1 단백질 정제를 도시한다. **도 21D** 는 각각 37℃ 및 48℃에서 BhC2C1 단백질 및 예상 tracr RNA에 의한 시험관내 절단을 도시한다.

도 22A-22D 는 BhC2C1에 대한 sgRNA 디자인을 도시한다. (**도 22A** - SEQ ID NO:38 및 662; **도 22B** - SEQ ID NO:39; **도 22C** - SEQ ID NO:40; **도 22D** - SEQ ID NO:41).

도 23 은 BhC2C1을 함유하는 예시적인 구성체의 플라스미드 맵을 도시한다.

도 24 는 표 12의 상이한 표적 부위에 대한, 플라스미드 형질감염 이후, 표 12의 상이한 sgRNA로 수득된 indel 백분율을 도시한다. 사용된 Cas12b는 BvCas12b였다 (SEQ ID NO:42-47).

도 25 는 BvCas12b를 함유하는 예시적인 구성체의 플라스미드 맵을 도시한다.

도 26 은 BhCas12b를 함유하는 예시적인 구성체의 플라스미드 맵을 도시한다.

도 27 은 EbCas12b를 함유하는 예시적인 구성체의 플라스미드 맵을 도시한다.

도. 28 는 AkCas12b를 함유하는 예시적인 구성체의 플라스미드 맵을 도시한다.

도. 29 는 PhyciCas12b를 함유하는 예시적인 구성체의 플라스미드 맵을 도시한다.

도 30 는 PlancCas12b를 함유하는 예시적인 구성체의 플라스미드 맵을 도시한다.

도 31 는 BvCas12b를 함유하는 예시적인 구성체 pZ143-pcDNA3-BvCas12b의 플라스미드 맵을 도시한다.

도 32 는 BvCas12b sgRNA 스캐폴드를 함유하는 예시적인 구성체 pZ147-BvCas12b-sgRNA-scaffold의 플라스미드 맵을 도시한다.

도 33 는 BhCas12b sgRNA 스캐폴드를 함유하는 예시적인 구성체 pZ148-BhCas12b-sgRNA-scaffold의 플라스미드 맵을 도시한다.

도 34 는 S893, K846, 및 E836에 돌연변이를 갖는 BhCas12b를 함유하는 예시적인 구성체 pZ149-BhCas12b-S893R-K846R-E836G의 플라스미드 맵을 도시한다.

도 35 는 S893, K846, 및 E836에 돌연변이를 갖는 BhCas12b를 함유하는 예시적인 구성체 pZ150-pCDNA3-BhCas12b-S893R-K846R-E836K의 플라스미드 맵을 도시한다.

도 36 은 다양한 조건 하에서 BhCas12b에 대한 PAM 발굴 결과를 도시한다.

도 37 은 다양한 조건 하에서 BvCas12b에 대한 PAM 발굴 결과를 도시한다.

도 38 은 상이한 결합 부위에서 BhCas12b 변이체의 indel 백분율을 도시한다.

도 39 은 상이한 결합 부위에서 추가의 BhCas12b 변이체의 indel 백분율을 도시한다.

도 40A 는 DNMT1-1에서 BhCas12b (실시예 20의 변이체 4) 및 BvCas12b에 의한 절단에 따른 HDR을 도시한다 (SEQ ID NO:48-51). **도 40B** 는 BhCas12b (실시예 20의 변이체 4) 및 BvCas12bat VEGFA-2 (SEQ ID No:52-55)에 의한 절단에 따른 HDR을 도시한다.

도 41A 는 TTTV PAM에서 AsCas12a 및 BhCas12b 변이체 4 및 BvCas12b ATTN PAMS의 indel 백분율의 비교를 도시한다. **도 41B** 는 상이한 PAM 서열에서 BhCas12b 변이체 4 및 BvCas12b 활성의 붕괴를 도시한다.

도 42A 는 ssDNA 도너 (SEQ ID NO:56-59)와 함께 도입시키려는 바람직한 변화를 포함한 VEGFA 표적의 개략도를 도시한다. **도 42B** 는 VEGFA 표적 부위에서 각 뉴클레아제의 indel 활성을 도시한다. **도 42C** 는 VEGFA 부위에서 바람직한 편집 (2개 뉴클레오티드 치환)을 함유하는 세포의 백분율을 도시한다. **도 42D** 는 ssDNA 도너 (SEQ ID No:60-63)와 함께 도입시키려는 바람직한 변화를 포함하는 DNMT1 표적의 개략도를 도시한다. **도 42E** 는 DNMT1 표적 부위에서 각 뉴클레아제의 indel 활성을 도시한다. **도 42F** 는 DNMT1 부위에서 바람직한 편집 (2개 뉴클레오티드 치환)을 함유하는 세포의 백분율을 도시한다.

도 43 - 좌측 패널은 CXCR4의 표적화된 엑손 및 각각 BhCas12b (v4) 및 BvCas12b (SEQ ID No:64-77)에 의해 표적화된 CXCR4 서열을 도시한다. 우측 패널은 2 도너로부터의 T 세포에서 CXCR4에 대한 BhCas12b (v4) 및 BvCas12b의 효과를 도시한 indel 백분율을 도시한다.

도 44A-44E. 중온성 Cas12b 뉴클레아제의 확인. **도 44A**) Cas9, Cas12a, 및 Cas12b 뉴클레아제 간 편차를 강조한 유전자좌 개략도 및 단백질 도메인 구조. SpCas9 (PDB:4oo8), AsCas12a (PDB:5b43), 및 AacCas12b (PDB:5u30)의 결정 구조. **도 44B**) 정제된 Cas12b 단백질 및 합성된 crRNA 및 RNA-Seq를 통해 확인된 tracrRNA를 갖는 Cas12b 시스템의 시험관내 재구성. 반응은 90분 동안 표시된 온도 및 250 nM Cas12b 단백질에서 수행되었다. **도 44C, 도 44D**) 6종 sgRNA 변이체를 갖는 293T 세포에서 AkCas12b 및 BhCas12b indel 활성. 오차 막대는 n=4 복제물로부터의 s.d.로 표시된다. sgRNA 서열의 경우 **도 50B** 및 **50C** 를 참조한다. **도 44E**) BhCas12b sgRNA 구조 및 시험된 변이체 (SEQ ID No:78)의 위치의 개략도.

도 45A-45H. BhCas12b의 합리적 조작. **도 45A**) 차등적으로 표지된 DNA 가닥과 시험관내 Cas12b 반응. 보다 느리게 이동하는 산물은 천연 PAGE 분리 동안 관찰되고 변성 PAGE에 의한 분리는 저온에서 비표적 가닥을 절단하기 위해 AkCas12b 및 BhCas12b에 대한 선호도를 밝혀준다. **도 45B**) 표적 가닥 및 RuvC 활성 부위 (보라색) 사이 포켓 내 12개 시험된 잔기 중 10개의 위치. BhCas12b 잔기는 고도로 유사한 BthCas12b (PDB: 5wti)의 구조에 강조되어 있다. **도 45C**) 야생형 (회색 기호)에 대해 정규화된 DNMT1 표적 4 및 VEGFA 표적 2에서 268 BhCas12b 돌연변이의 Indel 활성. 오차 막대는 n=2 복제물로부터의 s.d.로 표시된다. **도 45D**) 글리신으로 돌연변이된 표면 노출된 잔기의 위치. **도 45E**) 야생형 (회색 기호)에 대해 정규화된 IDNMT1 표적 4 및 VEGFA 표적 2에서 66 BhCas12b 돌연변이의 Indel 활성. 오차 막대는 n=2 복제물로부터의

s.d.로 표시된다. **도 45F)** BhCas12b 초활성 변이체의 요약. **도 45G**) 4 표적 부위에서 BhCas12b 변이체의 Indel 활성. 오차 막대는 n= 3- 6 복제물로부터의 s.d.로 표시된다. **도 45H**) BhCas12b WT 및 v4 변이체의 농도 증가에 따른 시험관내 절단. 겔은 n=2 실험으로부터의 대표적인 영상이다.

도 46A-46G. BhCas12b v4 및 BvCas12b는 인간 세포주에서 게놈 편집을 매개한다. **도 46A**) 28 TTTV 표적에서 AsCpf1, 33 ATTN 표적에서 BhCas12b v4, 및 37 ATTN 표적에서 BvCas12b의 293T 세포에서의 Indel 활성. 각각의 도트는 n=4 복제물로부터 평균낸 단일 표적 부위를 나타낸다. **도 46B**) 30 활성 가이드로부터 평균낸 Cas12b 게놈 편집으로부터의 평균 indel 길이. **도 46C**) SpCas9 및 Cas12a/b 뉴클레아제에 의해 표적화가능한 DNMT1 표적 부위 및 TG에서 CA로의 돌연변이 및 PAM 파괴 돌연변이를 함유하는 120 nt ssODN 도너 (SEQ ID No:79-83)의 개략도. **도 46D**) 유전자좌에서 각 뉴클레아제의 Indel 활성. 오차 막대는 n=8 복제물로부터의 s.d.로 표시된다. **도 46E**) 표적 가닥 (T) 또는 비표적 가닥 (NT) 도너를 사용한 상동성-지정 복구 (HDR)의 빈도. 회색 막대는 TG에서 CA로의 돌연변이의 빈도를 나타내는 한편, 붉은색 막대는 돌연변이가 없는 패널 c의 HDR 서열을 함유하는 완벽한 편집을 의미한다. 오차 막대는 n=6 복제물로부터의 s.d.로 표시된다. **도 46F**) 30 활성 BhCas12b 가이드, 45 활성 AsCas12a 가이드, 및 39 활성 SpCas9 가이드에 의한 게놈 편집 동안 평균 indel 길이. **도 46G**) BhCas12b v4 RNP 전달 후 CD4+ 인간 T 세포에서 Indel 활성. 각각의 도트는 개별 전기영동 (n=2)을 나타낸다. 소스 데이터는 소스 데이터 파일로서 제공된다.

도 47A-47B. BhCas12b v4 및 BvCas12b는 고도로 특이적인 뉴클레아제이다. **도 47A**) Guide-Seq 분석을 위해 선택된 9 표적 부위에서 293T 세포에서 Indel 활성. 오차 막대는 n=4 복제물로부터의 s.d.로 표시된다. **도 47B**) 각 뉴클레아제에 대한 검출된 절단 부위의 수 및 상대적 비율을 도시한 Guide-Seq 분석. 오프-표적은 연회색 웨지로 표시되는 반면 온-표적은 하기 표시된 온-표적 판독 분율과 함께 파란색으로 강조된다. 오프-표적은 오직 SpCas9에 의해 검출되었으며, 완전 분석의 경우 **도 55** 를 참조한다.

도 48A-48E Cas12b 오솔로그의 PAM 발굴. **도 48A**) Cas12b 오솔로그의 정렬. **도 48B**) 정렬 기반의 V-B 아형 이펙터 Cas12b 단백질의 계통수. 서열은 Genbank 단백질 수탁 번호 및 종 명칭으로 표시된다. 이 작업에서 실험적으로 연구된 단백질은 굵게 표시된다. 37C에서 강력한 편집 활성을 보여주고 상세하게 연구된 4개 단백질은 밑줄표시되어 있다. **도 48C**) 이. 콜라이에서 PAM 발굴 어세이의 개략도. **도 48D**) 고갈된 PAM은 이. 콜라이에서 12 Cas12b 시스템 중 오직 4종에서 검출되었다. 고갈 한계치는 2.32로 한계치 설정값을 갖는 CbCas12b를 제외하고 3.32의 $-\log_2$ 비율 (점선)로 설정되었다. 고갈된 PAM은 서열 정보를 나타내는 제1 5' 염기에 대한 휠의 중간에서 시작하는 PAM 휠[22] 을 비롯하여 서열 모티프로서 도시된다. **도 48E)** V-B 아형 이펙터 Cas12b 단백질의 계통수. 서열은 Genbank 단백질 수탁 번호 및 종 명칭으로 표시된다. 이 작업에서 실험적으로 연구된 단백질은 파란색으로 표시된다.

도 49A-49F. Cas12b RNA-Seq 및 시험관내 재구성. **도 49A-49D**) AkCas12b, BhCas12b, EbCas12b, 및 LsCas12b에 대한 소형 RNA-Seq 판독치의 정렬. 절단 반응에서 사용된 tracrRNA의 위치는 노란색으로 강조되어 있다. **도 49E**) 이 연구에서 사용된 정제된 Cas12b 단백질의 쿠마시 염색 SDS-PAGE 겔 및 상업적으로 제조된 AsCas12a (IDT). **도 49F**) tracrRNA 및 crRNA를 v1 sgRNA 스캐폴드와 비교하는 AkCas12b 및 BhCas12b에 의한 시험관내 절단 반응.

도 50A-50E. 포유동물 세포에서 Cas12b sgRNA 최적화. **도 50A**) 포유동물 세포에서 indel 활성에 대한 발현 구성체 및 어세이의 개략도. **도 50B**) AkCas12b sgRNA 변이체 (SEQ ID NO:84-89). **도 50C**) BhCas12b sgRNA 변이체 (SEQ ID NO:90-95). **도 50D**) AkCas12b sgRNA 구조 및 시험된 변이체 (SEQ ID No:96)의 위치의 개략도. **도 50E**) BhCas12b 및 다양한 스페이서 길이에 따른 293T 세포에서의 Indel 활성. 오차 막대는 n=2 복제물로부터의 s.d.로 표시된다.

도 51A-51J. BhCas12b의 합리적 조작. **도 51A**) 293T 세포에서 BhCas12b 및 고도로 유사한 BthCas12b 간 indel 활성의 비교. 오차 막대는 n=2 복제물로부터의 s.d.로 표시된다. **도 51B- 도. 51E**) DNMT1 표적 4 및 VEGFA 표적 2에서 BhCas12b 돌연변이체 조합의 Indel 활성. 오차 막대는 n=2 복제물로부터의 s.d.로 표시된다. **도 51F**) Pymol (Schrodinger)을 사용한 BthCas12b의 구조로 모델링된 BhCas12b v4 돌연변이. **도 51G**) 정제된 BhCas12b WT 및 v4 단백질의 쿠마시 염색된 SDS-PAGE 겔. **도 51H**) BhCas12b WT 및 v4 변이체에 의한 시험관내 절단 시간-과정. 겔은 n=3 실험으로부터의 대표적인 영상이다. **도 51I, 도. 51J**) 패널 h에 도시된 반응으로부터의 dsDNA 절단 산물 (**도 51I**) 및 상부 닉킹된 산물 (**도 51J**)의 정량. 오차 막대는 n=3 실험으로부터의 s.d.로 표시된다.

도 52A-52J. BvCas12b의 특징규명. **도 52A**) **도 48C** 및 **48D**에 기술된 바와 같은 PAM 발굴. **도 52B**) BvCas12b에 대한 소형 RNA-Seq 판독의 정렬. 절단 반응에서 사용된 tracrRNA의 위치는 노란색으로 강조되어 있다. **도 52C-52D**) 정제된 단백질 및 합성된 RNA 반응에 의한 BvCas12b의 시험관내 재구성은 90분 동안 표시된 온도 및 250 nM BvCas12b 단백질에서 수행되었다. **도. 52E**) 정제된 BvCas12b의 쿠마시 염색된 SDS-PAGE 겔. **도 52F**) BvCas12b sgRNA 변이체 (SEQ ID NO:97-102). **도 52G**) BvCas12b sgRNA 구조 및 시험된 변이체 (SEQ ID No:103)의 위치의 개략도. **도 52H**) sgRNA 변이체를 갖는 293T 세포에서 BvCas12b indel 활성. 오차 막대는 n=4 복제물로부터의 s.d.로 표시된다. **도 52I**) 57 표적에서 293T 세포에서 BvCas12b indel 활성. 각각의 도트는 n=4 복제물로부터 평균낸 단일 표적 부위를 나타낸다. **도 52J**) 매치된 표적 부위에서 BhCas12b v4 및 BvCas12b 활성의 상관성. 소스 데이터는 소스 데이터 파일로서 제공된다.

도 53A-53E. BvCas12b의 돌연변이유발. **도 53A**) 강조된 위치에서 확인된 표적-가닥 중 BhCas12b 위치 및 BvCas12b의 그들 상응하는 아미노산의 정렬. **도 53B**) 도 **45A** 에 기재된 바와 같이 차등적으로 표지된 DNA 가닥과 시험관내 BvCas12b 반응. **도 53C**) 잔기 Q635, D748, R849, H896, T909, I914 및 I919를 표적화하는 79 BvCas12b 돌연변이의 Indel 활성. Indel은 야생형 (회색 기호)에 대해 정규화된 DNMT1 표적 6 및 VEGFA 표적 5에서 측정되었다. 오차 막대는 n=2 복제물로부터의 s.d.로 표시된다. **도 53D- 53E**) DNMT1 표적 6 및 VEGFA 표적 5에서 BhCas12b 돌연변이의 Indel 활성. 오차 막대는 n=2 복제물로부터의 s.d.로 표시된다.

도 54A-54F. BhCas12b v4 및 BvCas12b는 인간 세포주에서 게놈 편집을 매개하였다. **도 54A**) 56 표적에서 BhCas12b v4, 및 57 표적에서 BvCas12b의 293T 세포에서의 Indel 활성. 각각의 도트는 n=4 복제물로부터 평균낸 단일 표적 부위를 나타낸다. **도 54B**) 매치된 표적 부위에서 BhCas12b v4 및 BvCas12b 활성의 상관성. **도 54C**) 클래스 2 CRISPR-Cas 뉴클레아제에 대한 PAM 우세도의 분석. 비차폐된 인간 코딩 서열 내 각 염기로부터 가장 가까운 Cas9 또는 Cas12 절단 부위까지의 거리에 대한 확률 질량 함수. **도 54D**) SpCas9 및 Cas12b 뉴클레아제에 의해 표적화가능한 VEGFA 표적 부위 및 TC에서 CA로의 돌연변이 및 PAM 파괴 돌연변이를 함유하는 120 nt ssODN 도너 (SEQ ID No:104-108)의 개략도. **도 54E**) 유전자좌에서 각 뉴클레아제의 Indel 활성. 오차 막대는 n=3 복제물로부터의 s.d.로 표시된다. **도 54F**) 표적 가닥 (T) 또는 비표적 가닥 (NT) 도너를 사용한 상동성-지정 복구 (HDR)의 빈도. 회색 막대는 TC에서 CA로의 돌연변이의 빈도를 나타내는 한편, 파란색 막대는 돌연변이가 없는 패널 d의 HDR 서열을 함유하는 완벽한 편집을 의미한다. 오차 막대는 n=3 복제물로부터의 s.d.로 표시된다.

도 55A-55C. BhCas12b v4 및 BvCas12b 미스매치 내성 및 특이성. **도 56A**) 각 뉴클레아제에 대한 검출된 절단 부위의 수 및 상대적 비율을 도시한 비매치된 표적의 Guide-Seq 분석. 오프-표적은 연회색 웨지로 표시되는 반면 온-표적은 하기 표시된 온-표적 판독 분율과 함께 파란색으로 강조된다. 전체 분석에 대해 **도. 57** 을 참조한다. **도 55B-55C**) 가이드 sgRNA 및 표적 DNA 사이에 미스매치가 존재할 때 293T 세포에서 Cas12b indel 활성. 미스매치는 표적 가닥을 매치시키기 위해서 sgRNA에 삽입되었다 (즉, C에서 G, A에서 T). BhCas12b v4는 DNMT1 표적 6 및 VEGFA 표적 2에서 시험되었고, 반면 BvCas12b는 DNMT1 표적 6 및 VEGFA 표적 5에서 시험되었다. 오차 막대는 n=4 복제물로부터의 s.d.로 표시된다.

도 56. 매치된 CRISPR-Cas 뉴클레아제 표적의 특이성 분석. **도 47B**의 검출된 오프-표적의 전체 Guide-Seq 분석. 검출된 절단 부위 (표적 당 최대 20)의 목록이 작은 박스로 표시된 온-표적 부위와 함께 각 뉴클레아제에 대해 제시되어 있다. 가이드 서열에 대한 미스매치가 강조되어 있다. 표적 1:EMX1 (SEQ ID NO:109-130); 표적 2:EMX1 (SEQ ID NO:131-152); 표적 3:DNMT1 (SEQ ID NO:153-174); 표적 4:CXCR4 (SEQ ID NO:175-176); 표적 5:CXCR4 (SEQ ID NO:178-181); 표적 6:CXCR4 (SEQ ID NO:182-186); 표적 7:VEGFA (SEQ ID NO:187-209); 표적 8:GRIN2B (SEQ ID NO:210-215); 표적 9:CXCR4 (SEQ ID NO:216-221); 표적 10:HPRT1 (SEQ ID NO:222-225).

도 57. 비미스매치된 CRISPR-Cas 뉴클레아제 표적의 특이성 분석. **도 56** 의 검출된 오프-표적의 전체 Guide-Seq 분석. 검출된 절단 부위 (표적 당 최대 20)의 목록이 작은 박스로 표시된 온-표적 부위와 함께 각 뉴클레아제에 대해 제시되어 있다. 가이드 서열에 대한 미스매치가 강조되어 있다. SpCas9 비매치 1:DNMT1 (SEQ ID NO:226); SpCas9 비매치 2:EMX1 (SEQ ID NO:227-246); SpCas9 비매치 3:VEGFA (SEQ ID NO:247-248); SpCas9 비매치 4:VEGFA (SEQ ID NO:249-268); SpCas9 비매치 5:VEGFA (SEQ ID NO:269-288); SpCas9 비매치 6:GRIN2B (SEQ ID NO:289-290); AsCas12a 비매치 1:DNMT1 (SEQ ID NO:291); AsCas12a 비매치 2:VEGFA (SEQ ID NO:292-293); AsCas12a 비매치 2:EMX1 (SEQ ID NO:294); AsCas12a 비매치 2:EMX1 (SEQ ID

NO:295); SpCas9 비매치 7:VEGFA (SEQ ID NO:296-311); SpCas9 비매치 8:EMX1 (SEQ ID NO:312-320); SpCas9 비매치 9:GRIN2B (SEQ ID NO:321-322); SpCas9 비매치 10:TUBB (SEQ ID NO:323-334); BhCas12b v4 비매치 1:DNMT1-BvCas12b 비매치 8:DNMT1 (SEQ ID NO:335-353); BhCas12b v4 비매치 9:CXCR4-BvCas12b 비매치 14:VEGFA (SEQ ID NO:354-367).

도 58은 Cas12 내 구조적으로 예상되는 ssDNA 경로 (PDB 구조 5U30 기반)를 도시한다.

도 59 은 RESCUE 돌연변이체의 용량 반응이 T 모티프에 대해 시험되었다는 것을 도시한다.

도 60 은 RESCUE 돌연변이체의 용량 반응이 C 및 G 모티프에 대해 시험되었다는 것을 도시한다.

도 61 및 **62** 는 RESCUE v3, v6, v7, 및 v8에 의한 내생성 표적화를 도시한다.

도 63 은 RESCUE v9에 대한 돌연변이에 대한 스크리닝을 수행되었음을 도시한다.

도 64 는 RESCUEv9에 대한 잠재적 돌연변이가 확인되었음을 도시한다.

도 65 는 염기 플립 및 모티프 시험이 수행되었음을 도시한다.

도 66 은 RESCUEv9의 효과가 상이한 모티프 플립에 대해 시험되었음을 도시한다.

도 67 은 50 bp 가이드를 갖는 RESCUE v1 및 v8에 의한 B6 및 B12 간 비교를 도시한다.

도 68 은 30 bp 가이드를 갖는 RESCUE v1 및 v8에 의한 B6 및 B12 간 비교를 도시한다.

도 69 은 스크리닝된 RESCUE 돌연변이의 요약을 도시한다.

도 70 은 더 나은 베타 카테닌 돌연변이체가 선택된 실험의 결과를 예시하는 그래프이다.

도 71 은 RESCUE 라운드 12의 결과를 예시하는 그래프를 도시한다.

도 72 는 베타 카테닌 이동 어세이를 예시하는 개략도이다.

도 73 은 베타 카테닌에 의해 유도된 세포 이동 어세이의 결과를 도시하는 그래프이다.

도 74 는 특이성 돌연변이가 A-I 오프-표적을 제거한 것을 예시한 그래프를 도시한다.

도 75 는 Stat1/3 인산화 부위 표적화가 신호전달을 감소시킨다는 것을 예시하는 그래프를 도시한다.

도 76 은 STAT1 비처리에 대한 결과를 도시한 도. 64A 및 STAT1 IFN 처리에 대한 결과를 도시한 도 64B와 함께, Stat1/3 인산화 부위 표적화가 신호전달을 감소시킨다는 것을 예시하는 그래프를 도시한다.

도 77 은 STAT3 IL6 활성화에 대한 결과를 도시한 도 65A 및 STAT3 비처리에 대한 결과를 도시한 도 65B와 함께, Stat1/3 인산화 부위 표적화가 신호전달을 감소시킨다는 것을 예시하는 그래프를 도시한다.

도 78 은 RESCUE 라운드 12의 결과를 예시하는 그래프를 도시한다.

도 79 은 RESCUE 라운드 13의 결과를 예시하는 그래프를 도시한다.

도 80 은 베타 카테닌에 의해 유도된 세포 이동 어세이의 결과를 도시하는 그래프이다.

도 81 - C에서 T 염기 편집 능력을 갖는 Bhv4 절두. 촉매적 불활성 Bhv4의 C-말단 142 아미노산 (dBhv4Δ143 - 불활성화 돌연변이 D574A, 새로운 크기 966 아미노산)의 제거 및 링커 및 래트 Apobec 도메인과 C-말단부의 융합 후에, C에서 T 염기 편집은 비표적 가닥 상의 가이드 염기쌍 위치에서 최대 10.95%의 빈도로 관찰된다. 6.97% 편집 효율이 가이드 위치 15에서 검출된다. 이 활성은 가이드 의존적이다. 존재하는 구성체와의 융합 또는 자유 발현을 통해서, 우라실-DNA 글리코실라제 억제제 (UGI) 도메인의 첨가는 이러한 C에서 T로의 전환을 증가시킬 것으로 예측된다. 열거된 가이드 서열 (대문자)은 HEK 293T 세포에서 GRIN2B 내부 영역을 표적으로 한다 (SEQ ID NO:368).

도 82A-82C- 도 82A) Guide-Seq 분석을 위해 선택된 9 표적 부위에서 293T 세포에서 Cas9, Cas12b, 및 Cas12a indel 활성의 비교 (3 TTTV PAM 부위에서만 시험된, Cas12a 제외). 오차 막대는 n=4 복제물로부터의 s.d.로 표시된다. **도 82B**) 각 뉴클레아제에 대한 검출된 절단 부위의 수 및 상대적 비율을 도시한 Guide-Seq 분석. 오프-표적은 연회색 웨지로 표시되는 반면 온-표적은 하기 표시된 온-표적 판독 분율과 함께 보라색 (SpCas9 경우), 진한 파란색 (BhCas12b v4 경우), 또는 연한 파란색 (AsCas12a 경우)으로 강조된다. 오프-표적은 SpCas9에 의해서만 검출되었다. n.t., 미시험됨. 도 **82C**) 가이드 sgRNA 및 표적 DNA 사이에 미스매치가 존

재할 때 293T 세포에서 BhCas12b indel 활성. 미스매치는 표적 가닥을 매치시키기 위해서 sgRNA에 삽입되었다 (즉, C에서 G, A에서 T). 오차 막대는 n=4 복제물로부터의 s.d.로 표시된다.

도 83 - Cas12 절두 및 APOBEC과 N-말단 및 C-말단 융합 및 이의 염기 편집 활성의 개략도를 제공한다.

도 84 - 일정 예의 구현예에 따른 Cas12 염기 편집 데이터를 제공한다 (SEQ ID NO:369-375).

도 85 - 일정 예의 구현예에 따른 Cas12 염기 편집 데이터를 제공한다.

도 86 - 일정 예의 구현예에 따른 가이드 상의 Cas12 염기 편집 데이터를 제공한다 (SEQ ID NO:376-377).

도 87 은 전체 길이 BhCas12b (SEQ ID No:378)를 사용한 예시적인 염기 편집 접근법을 도시한다.

도 88A-88C- 도 88A 는 BhCas12b v4 및 다른 오솔로그 AaCas12b의 indel 활성 간 비교를 도시한다. **도 88B 및 88C** 는 BhCas12b v4 또는 BhCas12b를 발현하는 AAV1/2로 뉴런의 형질도입을 입증한다.

도 89A-89B- 도 89A 는 px602-bh-optimize-AAV의 맵을 도시한다. **도 89A** 는 px602-bh-optimize-AAV의 맵을 도시한다.

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

발명을 실시하기 위한 구체적인 내용

[0046] **일반 정의**

[0047] 달리 정의하지 않으면, 본 명세서에서 사용되는 기술 및 과학 용어는 본 개시가 속하는 분야의 당업자가 통상적으로 이해하는 바와 동일한 의미를 갖는다. 분자 생물학의 통상의 용어의 정의 및 기술은 다음의 문헌들에서 확인할 수 있다: Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0048] 본 명세서에서 사용되는 바와 같이, 단수형 표현은 문맥에서 달리 명확하게 명시하지 않는 한, 단수형 및 복수형 대상 둘 모두를 포함한다.

[0049] 용어 "임의의" 또는 "임의로는"은 후술되는 사건, 상황 또는 치환기가 존재하지 않을 수도 있거나 또는 존재할 수도 있고, 그 설명은 사건 또는 상황이 일어나는 예 및 일어나지 않는 예를 포함한다는 것을 의미한다.

[0050] 종료점에 의한 수치 범위의 설명은 언급된 종료점을 비롯하여, 각 범위 내에 포함된 모든 수 및 분수를 포함한다.

[0051] 측정가능한 값 예컨대 매개변수, 양, 시간적 지속기간 등을 언급할 때 본원에서 사용되는 용어 "약" 또는 "대략" 은 명시된 값과 그로부터의 변동, 예컨대 그러한 변동이 개시된 발명에서 수행하기에 적절하다면, 명시된 값과 그로부터의 +/-10% 이하, +/-5% 이하, +/-1% 이하, 및 +/-0.1% 이하의 변동을 포괄한다는 것을 의미한다. 수식어 "약" 또는 "대략" 이 언급되는 값은 그 자체로 또한 특별히, 그리고 바람직하게 개시된다는 것이 이해될 것이다.

[0052] 용어 "예시적인" 은 "예, 실례, 또는 예시의 역할을 하는 것" 을 의미하도록 본 명세서에서 사용된다. "예

시적인" 것으로서 본 명세서에서 설명된 임의의 양태 또는 설계는 다른 양태들 또는 설계들에 비해 반드시 선호되거나 유리한 것으로서 해석되지는 않아야 한다.

[0053] 본 명세서에서 사용되는 "생물학적 샘플"은 전체 세포 및/또는 생존 세포 및/또는 세포 찌꺼기를 함유할 수 있다. 생물학적 샘플은 "체액" 을 함유할 수 있다 (또는 그로부터 유래할 수 있다). 본 발명은 체액이 양수, 수양액, 유리체액, 담즙, 혈액 혈청, 유액, 뇌척수액, 이구 (귀지), 유미, 미즙, 내림프, 외림프, 삼출액, 대변, 여성 사출액, 위산, 위액, 림프, 점액 (비강 배액 및 점액질 포함), 심낭액, 복막액, 흉수, 고름, 점막분비물, 타액, 피지 (피부 유분), 정액, 객담, 활액, 땀, 눈물, 소변, 질 분비물, 토사물, 및 이의 하나 이상의 혼합물로부터 선택되는 구체예를 포괄한다. 생물학적 샘플은 세포 배양물, 체액, 체액 유래 세포 배양물을 포함한다. 체액은 예를 들어 천공, 또는 기타 수집 또는 샘플링 절차에 의해 포유동물 유기체로부터 수득할 수 있다.

[0054] 용어 "대상체", "개체", 및 "환자"는 척추동물, 바람직하게 포유동물, 보다 바람직하게 인간을 의미하고자 본 명세서에서 상호교환적으로 사용된다. 포유동물은 제한없이 쥣과동물, 유인원, 인간, 농장 동물, 스포츠 동물, 및 반려 동물을 포함한다. 생체 내에서 수득되거나 또는 시험관 내에서 배양된 생물학적 독립체의 조직, 세포 및 그들 자손이 또한 포괄된다.

[0055] 이하 다양한 구현예를 기재한다. 특별한 구현예는 본 명세서에 논의되는 보다 넓은 양상에 대한 제한으로서 또는 완전한 설명으로서 의도되는 것이 아님을 유의해야 한다. 특정 구현예와 함께 기재되는 하나의 양태는 반드시 그 구현예로 제한되지 않으며 임의의 다른 구현예(들) 로 실시될 수 있다. 본 명세서 전반에서 "하나의 구현예", "한 구현예", "예시적 구현예" 에 대한 언급은 구현예와 함께 기재된 특정한 특성, 구조 또는 특징이 본 발명의 적어도 하나의 구현예에 포함된다는 것을 의미한다. 따라서, 본 명세서 전반의 다양한 위치에서 어구 "하나의 구현예에서", "한 구현예에서", 또는 "예시적 구현예에서" 의 출현은 반드시 모두 동일한 구현예를 언급하지 않지만, 그럴 수도 있다. 더 나아가서, 특정한 특성, 구조 또는 특징은 하나 이상의 구현예에서, 본 개시로부터 당업자에게 자명하게 되는 바와 같이, 임의의 적합한 방식으로 조합될 수 있다. 더 나아가서, 본 명세서에 기술된 일부 구현예가 다른 구현예에 포함된 다른 특성이 아닌 일부를 포함하지만, 상이한 구현예의 특성의 조합이 본 발명의 범주 내에 있다는 것을 의미한다. 예를 들어, 첨부된 청구항에서, 임의의 청구된 구현예들은 임의의 조합으로 사용될 수 있다.

[0056] 본원에서 인용되는 모든 출판물, 공개 특허 문서, 및 특허 출원은 각각의 개별 출판물, 공개 특허 문서, 또는 특허 출원이 참조로 포함되는 것으로서 특별히 개별적으로 표시한 바와 동일한 정도로 참조로 본원에 포함된다.

[0057] **개요**

[0058] 일 양상에서, 본 명세서에 개시된 구현예는 조작되거나 또는 단리된 CRISPR-Cas 이펙터 단백질 및 오솔로그에 관한 것이다. 특히, 본 발명은 Cas12b 이펙터 단백질 및 오솔로그에 관한 것이다. 본 명세서에서 사용되는, 용어 Cas12b는 C2c1과 상호교환적으로 사용된다. 본 발명은 또한 이러한 오솔로그를 포함하는 CRISPR-Cas 시스템을 비롯하여, 이러한 오솔로그 또는 시스템을 코딩하는 폴리뉴클레오티드 서열 및 이를 포함하는 벡터 또는 벡터 시스템 및 이를 포함하는 전달 시스템에 관한 것이다. 본 발명은 또한 이러한 Cas12b 단백질, CRISPR-Cas 시스템, 폴리핵산 서열, 벡터, 벡터 시스템, 전달 시스템을 포함하는 세포 또는 세포주 또는 유기체에 관한 것이다. 본 발명은 또한 이러한 단백질, CRISPR-Cas 시스템, 폴리핵산 서열, 벡터, 벡터 시스템, 전달 시스템, 세포, 세포주 등의 의학 및 비의학 용도에 관한 것이다. 다른 양상에서, 본 명세서에 개시된 구현예는 야생형에 비해서 결합 부위에 대한 CRISPR 복합체의 결합을 증강시키고/시키거나 편집 선호도를 변경시키는 비변형 CRISPR-Cas 이펙터 단백질과 비교해 적어도 하나의 변형을 포함하는 조작된 CRISPR-Cas 이펙터 단백질에 관한 것이다. 일정 구현예에서, CRISPR-Cas 이펙터 단백질은 V형 이펙터 단백질, 바람직하게 V-B형이다. 일정한 다른 예의 구현예에서, V-B형 이펙터 단백질은 C2c1 이다. 본 명세서에 개시된 구현예에서 사용을 위해 적합한 예의 C2c1 단백질은 이하에 더욱 상세하게 논의된다. 다른 양상에서, 개시된 구현예는 조작된 가이드를 포함하는 조작된 CRISPR-Cas 시스템에 관한 것이다. 본 명세서에서 사용되는 용어 CRISPR 이펙터 또는 CRISPR 단백질 또는 Cas (단백질 또는 이펙터)는 Cas12b 단백질 또는 이펙터와 상호교환적으로 사용되고 돌연변이 (예컨대 점 돌연변이(들) 및/또는 절두 포함)될 수 있거나 또는 야생형 단백질일 수 있다.

[0059] 일부 예에서, 본 개시는 i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질, ii) a) 하나 이상의 표적 서열, 일정 구현예에서, 하나 이상의 표적 DNA 서열에 하이브리드화할 수 있는 3' 가이드 서열, 및 b) 5' 직접 반복부 서열을 포함하는 crRNA, 및 iii) tracr RNA를 포함하여서, crRNA 및 tracr RNA와 복합체 형성된 Cas12b 이펙터 단

백질을 포함하는 CRISPR 복합체가 형성된 것인 비천연 발생 또는 조작된 시스템을 제공한다.

[0060] 일부 예에서, 본 개시는 i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질, 및 ii) 표적 서열과 하이브리드화할 수 있는 가이드 서열을 포함하는 가이드를 포함하는 비천연 또는 조작된 시스템을 제공한다. 일부 구현예에서, 시스템은 tracrRNA를 더 포함한다.

[0061] 다른 양상에서, 본 명세서에 개시된 구현예는 C2c1을 포함하는, CRISPR-Cas 이펙터 단백질의 전달을 위한 벡터에 관한 것이다. 일정 예의 구현예에서, 벡터는 단일 벡터 내에서 CRISPR-Cas 이펙터 단백질의 패키징을 허용하도록 설계된다. 또한 표적화 전달 및 조직-특이성을 위해 더 큰 이식유전자를 패키징하고 따라서 발현시키기 위한 치밀한 프로모터의 디자인에 대한 관심이 증가하고 있다. 따라서, 다른 양상에서 본 명세서에 개시된 일정 구현예는 전신 전달을 위해 더 큰 유전자를 전달하는 전달 벡터, 구성체, 및 방법에 관한 것이다.

[0062] 다른 양상에서, 본 발명은 CRISPR-Cas 시스템을 개발 또는 설계하기 위한 방법에 관한 것이다. 다른 양상에서, 본 발명은 제한없이, 치료제 개발, 생물생성, 및 식물 및 농업 적용분야를 포함한, 광범위한 적용의, 최적화된 CRISPR-Cas 시스템을 개발하거나 또는 설계하기 위한 방법에 관한 것이다. 일정 기반 요법 또는 치료제에서. 본 발명은 특히 CRISPR-Cas 시스템, 예컨대 CRISPR-Cas 시스템 기반 요법 또는 치료제를 개선시키기 위한 방법에 관한 것이다. 성공적인 CRISPR-Cas 시스템, 예컨대 CRISPR-Cas 시스템 기반 요법 또는 치료제의 핵심 특징은 고특이성, 고효율성, 및 고안전성을 포함한다. 고특이성 및 고안전성은 특히 오프-표적 효과의 감소에 의해 달성될 수 있다. 유사하게 개선된 특이성 및 효율성은 식물 및 생물생성에서의 적용분야를 개선시키는데 사용될 수 있다.

[0063] 따라서, 일 양상에서, 본 발명은 CRISPR-Cas 시스템, 예컨대 CRISPR-Cas 시스템 기반 요법 또는 치료제의 특이성을 증가시키기 위한 방법에 관한 것이다. 추가 양상에서, 본 발명은 CRISPR-Cas 시스템, 예컨대 CRISPR-Cas 시스템 기반 요법 또는 치료제의 효율성을 증가시키기 위한 방법에 관한 것이다. 추가 양상에서, 본 발명은 CRISPR-Cas 시스템, 예컨대 CRISPR-Cas 시스템 기반 요법 또는 치료제의 안전성을 증가시키기 위한 방법에 관한 것이다. 추가 양상에서, 본 발명은 CRISPR-Cas 시스템, 예컨대 CRISPR-Cas 시스템 기반 요법 또는 치료제의 특이성, 효율성, 및/또는 안전성, 바람직하게 전부를 증가시키기 위한 방법에 관한 것이다.

[0064] 일정 구현예에서, CRISPR-Cas 시스템은 본 명세서의 다른 곳에 정의된 바와 같은 CRISPR 이펙터를 포함한다.

[0065] 본 발명의 방법은 본 명세서의 다른 곳에 더욱 기술된 바와 같이, CRISPR-Cas 시스템과 연관된 선택된 매개변수 또는 변수 및/또는 이의 기능성의 최적화를 포함한다. 본 명세서에 기술된 바와 같은 방법에서 CRISPR-Cas 시스템의 최적화는 표적(들), 예컨대 치료 표적 또는 치료 표적들, CRISPR-Cas 시스템 조절 방식 또는 유형, 예컨대 CRISPR-Cas 시스템 기반 치료 표적(들) 조절, 변형, 또는 조작을 비롯하여, CRISPR-Cas 시스템 성분의 전달에 따라 좌우될 수 있다. 하나 이상의 표적은 유전자형 분석 및/또는 표현형 분석 결과에 따라서, 선택될 수 있다. 예를 들어, 하나 이상의 치료 표적은 (유전자) 질환 병인론 또는 바람직한 치료 결과에 따라서 선택될 수 있다. (치료) 표적(들)은 단일 유전자, 유전자좌, 또는 다른 게놈 부위일 수 있거나, 또는 다수 유전자, 유전자좌, 또는 다른 게놈 부위일 수 있다. 당분야에 공지된 바와 같이, 단일 유전자, 유전자좌, 또는 다른 게놈 부위는 예컨대 다수 gRNA의 사용을 통해서, 1회 초과로 표적화될 수 있다.

[0066] CRISPR-Cas 시스템 활성, 예컨대 CRISPR-Cas 시스템 디자인은 표적 파괴, 예컨대 표적 돌연변이, 예컨대 유전자 녹아웃 초래를 포함할 수 있다. CRISPR-Cas 시스템 활성, 예컨대 CRISPR-Cas 시스템 디자인은 특정 표적 부위의 치환, 예컨대 표적 교정의 초래를 포함할 수 있다. CISPR-Cas 시스템 디자인은 특정 표적 부위의 제거, 예컨대 표적 결실의 초래를 포함할 수 있다. CRISPR-Cas 시스템 활성은 표적 부위 기능성, 예컨대 예를 들어 (전사 및/또는 후생유전적) 유전자 또는 게놈 영역 활성화 또는 유전자 또는 게놈 영역 침묵화를 초래하는, 표적 부위 활성 또는 접근성의 조절을 포함할 수 있다. 당업자는 본 명세서의 다른 곳에 기술된 바와 같이, 표적 부위 기능성의 조절이 CRISPR 이펙터 돌연변이 (예컨대 예를 들어 촉매적 불활성 CRISPR 이펙터의 발생) 및/또는 기능화 (예컨대 예를 들어 이종성 기능성 도메인, 예컨대 전사 활성인자 또는 억제인자와 CRISPR 이펙터의 융합)를 포함할 수 있다는 것을 이해할 것이다. 따라서, 다른 양상에서, 본 발명은 변형된 CRISPR 이펙터 단백질 및 기능성 도메인(들)을 포함하는 부위 지정 염기 편집을 위한 조작된 조성물에 관한 것이다. 본 발명의 일 구현예에서, RNA 염기-편집이 존재한다. 본 발명의 일 구현예에서, DNA 염기-편집이 존재한다. 일정 구현예에서, 기능성 도메인은 시티딘 및 아데노신 디아미나제를 포함한, 디아미나제 또는 이의 촉매 도메인을 포함한다. 본 명세서에 개시된 구현예에서 사용을 위해 적합한 예의 기능성 도메인은 이하에 더욱 상세하게 논의된다.

[0067] 일정 예의 구현예에서, CRISPR 복합체를 형성하도록 가이드 서열을 포함하는 핵산과 복합체를 형성하는 조작된 CRISPR-Cas 이펙터 단백질로서, CRISPR 복합체에서 핵산 분자는 하나 이상의 폴리뉴클레오티드 유전자좌를 표적으로 하고, 단백질은 야생형과 비교하여 편집 선호도를 변경시키고/시키거나 결합 부위에 대한 CRISPR 복합체의 결합을 증강시키는 비변형된 단백질과 비교하여 적어도 하나의 변형을 포함한다. 편집 선호도는 indel 형성과 관련될 수 있다. 일정 예의 구현예에서, 적어도 하나의 변형은 표적 유전자좌에서 하나 이상의 특이적 indel의 형성을 증가시킬 수 있다. CRISPR-Cas 이펙터 단백질은 V형 CRISPR-Cas 이펙터 단백질일 수 있다. 일정 예의 구현예에서, CRISPR-Cas 단백질은 Cas12b로도 공지된, C2c1, 또는 이의 오솔로그이다.

[0068] 본 발명은 관심 표적 유전자좌와 연관된 또는 상기 표적 유전자좌에서의 서열을 변형시키거나 게놈 편집하는 방법을 제공하며, 상기 방법은 임의의 원하는 세포 유형, 원핵 또는 진핵생물 세포로 C2c1 이펙터 단백질 복합체를 도입하여 C2c1 이펙터 단백질 복합체가 효과적으로 진핵생물 또는 원핵생물 세포의 게놈 내로 DNA 삽입물을 통합시키는 작용을 하는 단계를 포함한다. 바람직한 구현예에서, 세포는 진핵생물 세포이고 게놈은 포유동물 게놈이다. 바람직한 구현예에서, DNA 삽입물의 통합은 비-상동성 말단 접합(non-homology end joining; NHEJ)-기반 유전자 삽입 메커니즘에 의해 촉진된다. 바람직한 구현예에서, DNA 삽입물은 외생성으로 도입된 DNA 주형 또는 복구 주형이다. 바람직한 일 구현예에서, 외생성으로 도입된 DNA 주형 또는 복구 주형은 복합체의 성분의 발현을 위하여 C2c1 이펙터 단백질 복합체 또는 하나의 성분 또는 폴리뉴클레오티드 벡터와 함께 전달된다. 더 바람직한 구현예에서, 진핵생물 세포는 비분할 세포(예를 들어 HDR을 통한 게놈 편집이 특히 어려운 비분열 세포)이다.

[0069] 본 발명은 또한 관심 표적 유전자좌를 변형시키는 방법을 제공하고, 방법은 상기 유전자좌에 C2c1 유전자좌 이펙터 단백질 및 하나 이상의 핵산 성분을 포함하는 비-천연 발생 또는 조작된 조성물을 전달하는 단계를 포함하고, C2c1 이펙터 단백질은 하나 이상의 핵산 성분과 복합체를 형성하고 상기 복합체가 관심 유전자좌에 결합시 이펙터 단백질은 관심 표적 유전자좌의 변형을 유도한다. 일 구현예에서, 변형은 가닥 파손의 도입이다. 가닥 파손은 비상동성 말단 접합이 후속될 수 있다. 다른 구현예에서, 복구 주형이 제공되고, 파손은 상동성 재조합이 후속된다.

[0070] 본 발명에 따라서, 핵산을 변형시키는 효소가 제공된다. 이러한 일 구현예에서, DNA의 염기 편집이 존재한다. 다른 이러한 일 구현예에서, RNA의 염기 편집이 존재한다. 보다 특히, 본 발명은 세포에서 핵염기를 변형시킬 수 있는 디아미나제 및 디아미나제 변이체를 제공한다. 일 구현예에서, 디아미나제는 DNA/RNA 듀플렉스 내 미스매치를 표적으로 하고 표적의 미스매치된 DNA 염기를 편집한다. 다른 구현예에서, 디아미나제는 RNA/RNA 듀플렉스 내 미스매치를 표적으로 하고 표적 RNA 염기를 편집한다.

[0071] 이러한 방법에서 관심 표적 유전자좌는 세포 내의 핵산 분자에 포함될 수 있다. 세포는 원핵생물 세포 또는 진핵생물 세포일 수 있다. 세포는 포유동물 세포일 수 있다. 포유류 세포는 비-인간 영장류, 소, 돼지, 설치류 또는 마우스 세포일 수 있다. 세포는 비-포유류 진핵생물 세포 예컨대 가금류, 어류 또는 새우일 수 있다. 세포는 식물 세포일 수도 있다. 식물 세포는 작물 식물 예컨대 카사바, 옥수수, 수수, 밀 또는 쌀의 세포일 수 있다. 식물 세포는 또한 조류, 수목 또는 채소의 세포일 수 있다. 본 발명에 의해 세포에 도입된 변형은, 세포 및 세포의 자손이 생물학적 생성물, 예컨대 항체, 전분, 알코올 또는 기타 다른 원하는 세포 산출물의 개선된 생성을 위하여 변경되도록 하는 것일 수 있다. 본 발명에 의해 세포에 도입된 변형은 세포 및 세포의 자손이 생성된 생물학적 생성물을 변화시키는 변경을 포함하도록 하는 것일 수 있다.

[0072] 기재된 방법 중 임의의 것에서, 관심 표적 유전자좌는 관심 게놈 또는 후생유전체 유전자좌일 수 있다. 기재된 방법 중 임의의 것에서, 복합체는 다중화된 사용을 위하여 다수의 가이드와 함께 전달될 수 있다. 기재된 방법 중 임의의 것에서 하나 초과의 단백질(들)이 사용될 수 있다.

[0073] **CRISPR-Cas 시스템**

[0074] 일반적으로, CRISPR 시스템은 전술한 문헌, 예컨대 WO 2014/093622 (PCT/US2013/074667)에서 사용된 대로 일 수 있고 Cas 유전자, 특히 C2c1 유전자를 코딩하는 서열, tracr (trans-activating CRISPR) 서열 (예를 들어, tracrRNA 또는 활성 부분 tracrRNA), tracr-메이트 서열 (내생성 CRISPR 시스템의 경우에 "직접 반복부" 및 tracrRNA-프로세싱된 부분 직접 반복부 포함), 가이드 서열 (내생성 CRISPR 시스템의 경우에 "스페이서"라고도 함), 또는 이 용어가 본 명세서에서 사용되는 대로의 "RNA(들)" (예를 들어, 가이드 C2c1에 대한 RNA(들), 예를 들어, CRISPR RNA 및 트랜스활성화 (tracr) RNA 또는 단일 가이드 RNA (sgRNA) (키메라 RNA)) 또는 CRISPR 유전자좌 유래의 다른 서열 및 전사물을 포함하는, CRISPR-연관 ("Cas") 유전자의 발현 또는 그의 활성 유도에 관

여하는 전사물 및 다른 엘리먼트를 집합적으로 의미한다.

[0075] 일반적으로, CRISPR 시스템은 표적 서열 (내생성 CRISPR 시스템의 경우 프로토스페이서로도 지칭) 의 위치에서 CRISPR 복합체의 형성을 촉진하는 요소를 특징으로 한다. CRISPR 복합체 형성의 맥락에서, "표적 서열" 은 가이드 서열이 상보성을 갖도록 설계된 서열을 지칭하며, 여기서, 표적 서열과 가이드 서열 간의 하이브리드화는 CRISPR 복합체의 형성을 촉진시킨다. Cas12b 단백질을 포함하는 구현예에서 형성된 CRISPR 복합체는 본 명세서의 다른 곳에 기술된, crRNA 및 tracrRNA와의 복합체를 포함할 수 있다. 표적 서열에 대한 상보성이 절단 활성에 중요한 가이드 서열의 부분은 본 명세서에서 씨드 서열이라고 한다. 표적 서열은 임의의 폴리뉴클레오티드, 예컨대 DNA 또는 RNA 폴리뉴클레오티드를 포함할 수 있다. 일부 구현예에서, 표적 서열은 세포의 핵 또는 세포질에 위치되고, 미토콘드리아, 세포기관, 소포, 리포솜 또는 세포 내에 존재하는 입자에서의 또는 이들로부터의 핵산을 포함할 수 있다. 일부 구현예에서, 특히 비-핵 용도를 위해, NLS 는 바람직하지 않다. 일부 구현예에서, CRISPR 시스템은 하나 이상의 핵 이출 신호 (NES)를 포함한다. 일부 구현예에서, CRISPR 시스템은 하나 이상의 NLS 및 하나 이상의 NES 를 포함한다. 일부 구현예에서, 직접 반복부는 임의의 또는 전부의 하기 기준을 충족하는 반복 모티프에 대한 검색을 통해서 인 실리코에서 확인할 수 있다. 1. II형 CRISPR 유전자좌에 측접한 게놈 서열의 2 Kb 창에서 발견; 2. 20 내지 50 bp 범위; 및 3. 20 내지 50 bp 간격. 일부 구현예에서, 이들 기준 중 2 가지, 예를 들어 1 과 2, 2 와 3, 또는 1 과 3 이 사용될 수 있다. 일부 구현예에서, 3 가지 기준 모두가 사용될 수 있다.

[0076] 일반적으로, CRISPR 시스템은 표적 서열 부위에서 CRISPR 복합체의 형성을 촉진하는 구성요소를 특징으로 한다. CRISPR 복합체의 형성 경우에, "표적 서열"은 가이드 서열이 상보성을 갖도록 디자인된 서열을 의미하고, 여기서 표적 DNA 서열과 가이드 서열 간 하이브리드화가 CRISPR 복합체의 형성을 촉진한다.

[0077] 용어 "가이드 분자", "가이드 RNA" 및 "가이드"는 제한없이 CRISPR-Cas 단백질과 복합체를 형성할 수 있고 표적 핵산 서열과 충분한 상보성을 갖는 가이드 서열을 포함하여 표적 핵산 서열과 하이브리드화하고 표적 핵산 서열과 복합체의 서열-특이적 결합을 유도하는 RNA-기반 분자를 포함하여, 핵산-기반 분자를 의미하고자 본 명세서에서 상호교환적으로 사용된다. 가이드 분자 또는 가이드 RNA는 특별히 본 명세서에 기술된 바와 같이, (예를 들어, 2개의 리보뉴클레오티드를 화학적으로 연결하거나 또는 하나 이상의 리보뉴클레오티드를 하나 이상의 데옥시리보뉴클레오티드로 치환시켜서) 하나 이상의 화학적 변형을 갖는 RNA-기반 분자를 포괄한다.

[0078] 일정 구현예에서, 표적 서열은 PAM (protospacer adjacent motif) 또는 PFS (protospacer flanking sequence 또는 site), 다시 말해서, CRISPR 복합체에 의해 인식되는 짧은 서열과 회합되어야 한다. CRISPR-Cas 단백질의 성질에 따라서, 표적 서열은 DNA 듀플렉스 내 이의 상보성 서열 (본 명세서에서 비-표적 서열이라고도 함)이 PAM의 상류 또는 하류에 존재하도록 선택되어야 한다. CRISPR-Cas 단백질이 C2c1 단백질인 본 발명의 구현예에서, a 내 표적 서열의 상보성 서열은 PAM의 하류 또는 3'에 존재한다. PAM을 위한 정확한 서열 및 길이 요건은 사용되는 C2c1 단백질에 따라 상이하지만, PAM은 전형적으로 프로토스페이서 (즉, 표적 서열)에 인접한 25 염기쌍 서열이다. 상이한 C2c1 오솔로그에 대한 천연 PAM 서열의 예는 하기 본 명세서에서 제공하고 당업자는 소정 C2c1 단백질과 사용을 위한 추가 PAM 서열을 확인할 수 있을 것이다.

[0079] 시스템은 하나 이상의 표적 서열의 변형을 위해 사용될 수 있다 (예를 들어, 세포 또는 세포 개체군에서). 변형은 적어도 하나의 유전자 산물의 변경된 발현을 야기시킬 수 있다. 일부 예들에서, 적어도 하나의 유전자 산물의 발현은 증가될 수 있다. 일부 예들에서, 적어도 하나의 유전자 산물의 발현은 감소될 수 있다.

[0080] 일부 예에서, 변형은 세포 또는 세포의 개체군에서 만들 수 있고, 변형은 내생성 또는 비내생성 생물학적 산물 또는 화학적 화합물을 생산하고/하거나 분비하는 세포 또는 개체군을 야기시킬 수 있다. 화학적 화합물 또는 생물학적 산물은 저분자량 화합물을 포함할 수 있을 뿐만 아니라, 또한 거대 화합물, 또는 변형 및 비변형 핵산 예컨대 안티센스 핵산, RNAi, 예컨대 siRNA 또는 shRNA, CRISPR-Cas 시스템, 펩티드, 펩티드모방체, 수용체, 리간드, 및 항체, 압타머, 폴리펩티드, 핵산 유사체 또는 이의 변이체를 포함하여, 소정 상황에서 효과적인 임의의 유기 또는 무기 분자일 수도 있다. 예에는 제한없이 단백질, 올리고뉴클레오티드, 리보자임, DNAzyme, 당단백질, siRNA, 지단백질, 압타머, 및 이의 변형 및 조합을 포함한, 핵산, 아미노산, 또는 탄수화물의 올리고머를 포함한다. 작용제는 화학제; 소형 분자; 핵산 서열; 핵산 유사체; 단백질; 펩티드; 압타머; 항체; 또는 이의 단편을 포함한 군으로부터 선택될 수 있다. 핵산 서열은 RNA 또는 DNA일 수 있고, 단일 또는 이중 가닥일 수 있으며, 관심 단백질을 코딩하는 핵산, 올리고뉴클레오티드, 핵산 유사체, 예를 들어 펩티드 -핵산 (PNA), 슈도-상보성 PNA (pc-PNA), 잠김 핵산 (LNA), 변형된 RNA (mod-RNA), 단일 가이드 RNA 등을 포함하는 군으로부터 선택될 수 있다. 이러한 핵산 서열은 예를 들어, 제한없이, 예를 들어 전사 억제인자

로서 작용하는 단백질을 코딩하는 핵산 서열, 안티센스 분자, 리보자임, 소형 억제성 핵산 서열, 예를 들어 제한없이, RNAi, shRNAi, siRNA, 마이크로 RNAi (mRNAi), 안티센스 올리고뉴클레오티드, 예를 들어 특이적 DNA 표적 서열로 CRISPR 효소를 표적화하는 CRISPR 가이드 RNA 등을 포함한다. 단백질 및/또는 펩티드 또는 이의 단편은 임의의 관심 단백질, 예를 들어, 제한없이, 돌연변이된 단백질; 치료 단백질 및 절두 단백질일 수 있고, 여기서 단백질은 세포 내에서 정상적으로 부재하거나 또는 저 수준으로 발현된다. 단백질은 또한 돌연변이된 단백질, 유전자 조작된 단백질, 펩티드, 합성 펩티드, 재조합 단백질, 키메라 단백질, 항체, 미니바디, 인간화 단백질, 인간화 항체, 키메라 항체, 변형된 단백질 및 이의 단편을 포함하는 군으로부터 선택될 수 있다. 대안적으로, 작용제는 세포로 핵산 서열의 도입 결과로서 세포 내에서 세포내일 수 있고 이의 전사는 세포 내에서 유전자의 단백질 조절인자 및/또는 핵산의 생산을 일으킬 수 있다. 일부 구현예에서, 작용제는 제한없이 합성 및 천연 발생 비단백질성 독립체를 포함한, 임의의 화학물, 독립체 또는 모이어티이다. 일정 구현예에서, 작용제는 화학 모이어티를 갖는 소형 분자이다. 작용제는 바람직한 활성 및/또는 속성을 갖는 것으로 알려진 것일 수 있거나, 또는 다양한 화합물의 라이브러리로부터 선택될 수 있다.

[0081] **PAM의 결정**

[0082] 출원인은 PAM과 내성 유전자 둘 모두를 함유하는 플라스미드를 이종성 이. 콜라이 내로 도입하고, 이어서, 상응하는 항생제 상에서 플레이팅한다. 플라스미드의 DNA 절단이 있다면, 출원인은 생존 가능한 콜로니를 관찰하지 못한다. 추가 상세한 설명에서, 어세이는 DNA 표적에 대해 다음과 같다. 2 종의 이. 콜라이 균주가 본 어세이에서 사용된다. 하나는 박테리아 균주로부터 내생성 이펙터 단백질 유전자좌를 코딩하는 플라스미드를 운반한다. 다른 균주는 빈 플라스미드 (예를 들어, pACYC184, 대조군 균주)를 운반한다. 모든 가능한 7 또는 8 bp PAM 서열은 항생체 내성 플라스미드 (암피실린 내성 유전자를 갖는 pUC19) 상에 존재한다. PAM은 프로토-스페이서 1 (내생성 이펙터 단백질 유전자좌에서 제1 스페이서에 대한 DNA 표적)의 서열 다음에 위치된다. 2개의 PAM 라이브러리를 클로닝하였다. 하나는 프로토-스페이서의 5'에 8개의 무작위 bp를 갖는다 (예를 들어, 총 65536개의 상이한 PAM 서열 = 복잡성). 다른 라이브러리는 프로토-스페이서의 3'에 7개의 무작위 bp를 갖는다(예를 들어, 총 복잡성은 16384 상이한 PAM임). 라이브러리는 둘 모두 가능한 PAM 당 평균 500개의 플라스미드를 갖도록 클로닝되었다. 시험 균주 및 대조군 균주는 별개의 형질전환에서 5'PAM 및 3'PAM 라이브러리로 형질전환되었고, 형질전환된 세포는 암피실린 플레이트 상에서 별개로 플레이팅되었다. 플라스미드에 의한 인식 및 후속적 절단/간섭은 암피실린에 취약한 세포를 제공하며, 성장을 방지한다. 형질전환의 대략 12시간 후에, 채취한 시험 및 대조군 균주에 의해 형성되는 모든 콜로니 및 플라스미드 DNA가 단리되었다. 플라스미드 DNA는 PCR 증폭 및 후속적 심층 시퀀싱을 위한 주형으로서 사용되었다. 비형질전환 라이브러리에서 모든 PAM의 제시는 형질전환 세포에서 PAM의 예상되는 표현을 나타내었다. 대조군 균주에서 발견되는 모든 PAM의 표현은 실제 표현을 보였다. 시험 균주에서 모든 PAM의 표현은 PAM이 효소에 의해 인식되지 않으며 대조군 균주에 대한 비교가 감손된 PAM 서열의 추출을 허용한다는 것을 나타낸다.

[0083] 지금까지 확인된 C2c1 오솔로그 경우에, 하기 PAM이 확인되었다: 알리시클로바실러스 악시도테레스트리스 (Alicyclobacillus acidoterrestris) ATCC 49025 C2c1p (AacC2c1)는 5' TTN PAM (여기서 N은 A, C, G, 또는 T이고, 보다 바람직하게 N은 A, G, 또는 T 임)이 선행하는 표적 부위를 절단할 수 있고; 바실러스 써모아밀로보란스 (Bacillus thermoamylovorans) 균주 B4166 C2c1p (BthC2c1)는 ATTN (여기서 N은 A/C/G 또는 T임)가 선행되는 부위를 절단할 수 있다.

[0084] ***코돈 최적화된 핵산 서열***

[0085] 이펙터 단백질이 핵산으로서 투여되는 경우에, 본 출원은 코돈-최적화된 CRISPR-Cas V형 단백질, 보다 특히 C2c1-코딩 핵산 서열 (및 임의로는 단백질 서열)의 사용을 예상한다. 코돈-최적화된 서열의 예로서 예를 들어 진핵생물, 예를 들어 인간 (즉, 인간에서 발현을 위해 최적화됨), 또는 다른 진핵생물의 경우에, 본 명세서에 논의된 바와 같은 동물 또는 포유동물에서의 발현을 위해 최적화된 서열이고; 예를 들어 WO 2014/093622 (PCT/US2013/074667)의 SaCas9 인간 코돈-최적화된 서열을 코돈 최적화된 서열의 예로서 참조한다 (당업계 및 본 개시의 지식으로부터, 특히 이펙터 단백질 (예를 들어, C2c1)에 대한, 코돈 최적화 코딩 핵산 분자(들)는 당업자의 영역 내임). 이것이 바람직하지만, 다른 예가 가능하며, 인간 이외의 숙주 종에 대한 코돈 최적화 또는 특정 장기에 대한 코돈 최적화가 공지되어 있다는 것을 이해할 것이다. 일부 구현예에서, DNA/RNA-표적화 Cas 단백질을 코딩하는 효소 코딩 서열은 특정 세포, 예컨대, 진핵생물 세포에서 발현에 대해 코돈 최적화된다. 진핵생물 세포는 특정 유기체, 예컨대, 제한없이 인간, 또는 비-인간 진핵생물 또는 본 명세서에서 논의되는 동물 또는 포유동물, 예를 들어, 마우스, 래트, 토끼, 개, 가축 또는 인간이외의 포유동물 또는 영장류를 포함하는, 포유동물 또는 식물의 것일 수 있거나 또는 그로부터 유래될 수 있다. 일부 구현예에서, 인

간의 생식 계통 유전적 동일성을 변형시키기 위한 과정 및/또는 사람 또는 동물, 및 이러한 과정으로부터 초래된 동물에 임의의 실질적인 의학적 이점 없이 그들에게 고통을 야기할 가능성이 있는 동물의 유전적 동일성을 변형시키는 과정은 배제될 수 있다. 일반적으로, 코돈 최적화는 천연 서열의 적어도 하나의 코돈 (예를 들어, 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 이상의 코돈)을 해당 숙주 세포의 유전자에서 더 빈번하게 또는 가장 빈번하게 사용되는 코돈으로 대체함으로써 관심 숙주 세포에서 증강된 발현을 위한 핵산 서열을 변형시키는 한편, 천연 아미노산 서열을 유지시키는 과정을 의미한다. 다양한 종은 특정 아미노산의 소정의 코돈에 대한 특정 편향성을 나타낸다. 코돈 편향성 (유기체 사이의 코돈 용법 차이)은 종종 메신저 RNA (mRNA)의 번역 효율과 상관관계가 있는데, 이는 결국 특히 번역 중인 코돈의 특성 및 특정 전달 RNA(tRNA) 분자의 이용 가능성에 의존하는 것으로 여겨진다. 세포에서 선택된 tRNA의 우세성은 일반적으로 펩티드 합성에서 가장 빈번하게 사용되는 코돈의 반영이다. 따라서, 유전자는 코돈 최적화에 기반하여 주어진 유기체에서 최적의 유전자 발현을 위해 맞춤될 수 있다. 코돈 용법 표는, 예를 들어, www.kazusa.orjp/codon/에서 이용 가능한 "코돈 용법 데이터베이스"에서 용이하게 이용 가능하며, 이들 표는 다수의 방법에서 적합하게 될 수 있다. 참조: Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases:status for the year 2000" Nucl. Acids Res. 28:292 (2000). 특정 숙주 세포에서의 발현을 위해 특정 서열을 코돈 최적화시키는 컴퓨터 알고리즘이 또한 이용 가능하며, 예를 들어 Gene Forge (Aptagen; Jacobus, PA) 이 또한 이용가능하다. 일부 구현예에서, DNA/RNA-표적화 Cas 단백질을 코딩하는 서열에서 하나 이상의 코돈 (예를 들어, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 이상, 또는 모든 코돈)은 특정 아미노산에 대해 가장 빈번하게 사용되는 코돈에 대응한다 효모에서의 코돈 용법에 관해서, http://www.yeastgenome.org/community/codon_usage.shtml에서 입수가능한 온라인 Yeast Genome 데이타베이스 또는 [Codon selection in yeast, Bennetzen and Hall, J Biol Chem. 1982 Mar 25;257(6):3026- 31]를 참조한다. 조류를 포함한 식물에서 코돈 용법에 관해서, [Codon usage in higher plants, green algae, and cyanobacteria, Campbell and Gowri, Plant Physiol. 1990 Jan; 92(1): 1-11.]를 비롯하여, [Codon usage in plant genes, Murray et al, Nucleic Acids Res. 1989 Jan 25;17(2):477-98] 또는 [Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages, Morton BR, J Mol Evol. 1998 Apr;46(4):449-59]를 참조한다.

[0086] **가이드 분자**

[0087] 본 명세서에서 사용되는 용어 V형 또는 VI 형 CRISPR-Cas 유전자좌 이펙터 단백질의 "crRNA" 또는 "가이드 RNA" 또는 "단일 가이드 RNA" 또는 "sgRNA" 또는 "하나 이상의 핵산 성분"은 적합한 정렬 알고리즘을 사용해 최적으로 정렬 시, 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% 이상의 상보성 정도로, 표적 핵산 서열과 하이브리드화하고, 표적 핵산 서열과 핵산-표적화 복합체의 서열-특이적 결합을 유도하기에 충분한 상보성을 갖는 임의의 폴리뉴클레오티드 서열을 포함한다. 최적 정렬은 서열을 정렬하기 위한 임의의 적합한 알고리즘을 사용하여 결정할 수 있고, 이의 비제한적인 예는 스미스-워터만 (Smith-Waterman) 알고리즘, 니들만-분취 (Needleman-Wunsch) 알고리즘, 버로우스-윌러스 (Burrows-Wheeler) 전환 기반 알고리즘 (예를 들어, Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; www.novocraft.com에서 입수가능), ELAND (Illumina, San Diego, CA), SOAP (soap.genomics.org.cn에서 입수가능), 및 Maq (maq.sourceforge.net에서 입수가능)을 포함한다. 표적 핵산 서열에 대한 핵산-표적화 복합체의 서열-특이적 결합을 유도하는 (핵산-표적화 가이드 RNA 내의) 가이드 서열의 능력은 임의의 적합한 어세이에 의해 평가될 수 있다. 예를 들어, 시험될 가이드 서열을 비롯한 핵산-표적화 복합체를 형성하는데 충분한 핵산-표적화 CRISPR 시스템의 성분은 상응하는 표적 핵산 서열을 갖는 숙주 세포에, 예컨대, 핵산-표적화 복합체 성분을 코딩하는 벡터에 의한 형질감염, 이후 표적 핵산 서열 내의 우선적인 표적화 (예를 들어, 절단) 평가에 의해, 예컨대, 본원에서 기재된 바와 같은 Surveyor 어세이에 의해 제공될 수 있다. 유사하게는, 표적 핵산 서열의 절단은 시험될 가이드 서열 및 시험 가이드 서열과 상이한 대조군 가이드 서열을 비롯한, 핵산-표적화 복합체 성분인 표적 핵산 서열을 제공함으로써, 그리고 시험 가이드 서열과 대조군 가이드 서열 반응 사이의 표적 서열에서의 결합 또는 절단 비율을 비교함으로써 시험 튜브에서 평가될 수 있다. 다른 어세이가 가능하며, 당업자에게 떠오를 것이다. 임의의 표적 핵산 서열을 표적화하기 위해 가이드 서열, 및 그에 따른 핵산-표적화 가이드가 선택될 수 있다. 표적 서열은 DNA 일 수 있다. 표적 서열은 임의의 RNA 서열일 수 있다. 일부 구현예에서, 표적 서열은 메신저 RNA (mRNA), 프리-mRNA, 리보솜 RNA (rRNA), 트랜스퍼 RNA (tRNA), 마이크로-RNA (miRNA), 소형 간섭 RNA (siRNA), 소형 핵 RNA (snRNA), 소형 핵소체 RNA (snoRNA), 이중 가닥 RNA (dsRNA), 비-코딩 RNA (ncRNA), 긴 비-코딩 RNA (lncRNA), 및 소형 세포질 RNA (scRNA) 로 이루어지는 군에서 선택되는 RNA 분자 내의 서열일 수 있다. 일부 바람직한 구현예에서, 표적 서열은 mRNA, 전-mRNA 및 rRNA로

이루어진 군으로부터 선택된 RNA 분자 내의 서열일 수 있다. 일부 바람직한 구현예에서, 표적 서열은 ncRNA 및 lncRNA로 이루어진 군으로부터 선택된 RNA 분자 내의 서열일 수 있다. 일부 더 바람직한 구현예에서, 표적 서열은 mRNA 분자 또는 전-mRNA 분자 내의 서열일 수 있다. 본 발명에서, 디아미나제는 표적 핵산 서열을 접합시키거나 또는 표적 서열은 본 명세서에서 "표적 아데노신"이라고도 하는 탈아미노화되는 표적 아데노신을 포함하는 서열이다. 일부 구현예에서, 상기 본 명세서에 기술된 상보성은 의도된 미스매치, 예컨대 본 명세서에 기술된 dA-C 미스매치는 배제한다. 가이드 서열은 원핵생물 세포 내 표적 DNA 서열과 하이브리드화할 수 있다. 가이드 서열은 진핵생물 세포 내 표적 DNA 서열과 하이브리드화할 수 있다.

[0088] 일부 구현예에서, 핵산-표적화 가이드는 핵산-표적화 가이드 내에서 2차 구조 정도를 감소시키도록 선택된다. 일부 구현예에서, 최적으로 폴딩될 때 핵산-표적화 가이드의 뉴클레오티드의 약 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1% 이하가 자기-상보성 염기 짝짓기에 참여한다. 최적의 폴딩은 임의의 적합한 폴리뉴클레오티드 폴딩 알고리즘에 의해 결정될 수 있다. 일부 프로그램은 깁스 (Gibbs) 자유 에너지의 계산을 기반으로 한다. 이러한 알고리즘의 예는 [Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148)]에 기술된 바와 같은, mFold이다. 또 다른 폴딩 알고리즘의 예는 중심 구조 예측 알고리즘을 사용하여, 비엔나 대학의 이론 화학 연구소 (Institute for Theoretical Chemistry at the University of Vienna) 에서 개발한 온라인 웹서버 RNAfold 이다 (예를 들어, [A.R. Gruber et al., 2008, Cell 106(1): 23-24]; 및 [PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62] 참조).

[0089] 일정 구현예에서, 가이드 RNA 또는 crRNA 는 직접 반복 (DR) 서열 및 가이드 서열 또는 스페이서 서열을 포함할 수 있거나, 그로 본질적으로 이루어질 수 있거나, 그로 이루어질 수 있다. 일정 구현예에서, 가이드 RNA 또는 crRNA 는 가이드 서열 또는 스페이서 서열에 융합되거나 연결된 직접 반복 서열을 포함할 수 있거나, 본질적으로 그로 이루어질 수 있거나, 그로 이루어질 수 있다. 특정 구현예에서, 직접 반복 서열은 가이드 서열 또는 스페이서 서열의 상류 (즉, 5') 에 위치될 수 있다. 다른 구현예에서, 직접 반복 서열은 가이드 서열 또는 스페이서 서열의 하류 (즉, 3') 에 위치될 수 있다.

[0090] 일부 구현예에서, 가이드 서열과 표적 서열 사이에 형성된 헤테로듀플렉스가 표적 서열 상에서 탈아미노화를 위해 표적 A에 마주보는 가이드 서열에서 짝짓지 않은 C를 포함하도록, 가이드 분자는 표적 서열과 적어도 하나의 미스매치를 갖도록 설계된 가이드 서열을 포함한다. 일부 구현예에서, 이런 A-C 미스매치 이외에, 적합한 정렬 알고리즘을 이용하여 최적으로 정렬될 때, 상보성 정도는 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99% 이상이다.

[0091] 일정 구현예에서, 가이드 분자의 가이드 서열 또는 스페이서 길이는 10 nt 내지 50 nt, 보다 특히 15 내지 35 nt 이다. 일정 구현예에서, 가이드 RNA의 스페이서 길이는 적어도 15 뉴클레오티드이다. 일정 구현예에서, 스페이서 길이는 10 내지 15 nt, 예를 들어 10, 11, 12, 13, 14, 14, 15 내지 17 nt, 예를 들어, 15, 16, 또는 17 nt, 17 내지 20 nt, 예를 들어 17, 18, 19, 또는 20 nt, 20 내지 24 nt, 예를 들어, 20, 21, 22, 23, 또는 24 nt, 23 내지 25 nt, 예를 들어, 23, 24, 또는 25 nt, 24 내지 27 nt, 예를 들어, 24, 25, 26, 또는 27 nt, 27-30 nt, 예를 들어, 27, 28, 29, 또는 30 nt, 30-35 nt, 예를 들어, 30, 31, 32, 33, 34, 또는 35 nt, 또는 35 nt 또는 그 이상의 길이이다. 일정 예의 구현예에서, 가이드 서열은 15, 16, 17,18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 또는 100 nt이다.

[0092] CRISPR-Cas 시스템의 일부 구현예에서, 가이드 서열과 이에 상응하는 표적 서열 사이의 상보성 정도는 약 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, 또는 100% 이상일 수 있고; 가이드 또는 RNA 또는 sgRNA 는 약 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 개 이상의 뉴클레오티드 길이일 수 있거나; 가이드 또는 RNA 또는 sgRNA 는 약 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 개 이하의 뉴클레오티드 길이일 수 있고; 유리하게는 tracr RNA 는 30 또는 50 개 뉴클레오티드 길이이다. 그러나, 본 발명의 양태는 오프-표적 상호작용을 감소시키는 것이고, 예를 들어 낮은 상보성을 갖는 표적 서열과 상호작용하는 가이드를 감소시키는 것이다. 실제로, 예에서, 80% 초과 내지 약 95% 상보성, 예를 들어, 83%-84% 또는 88-89% 또는 94-95% 상보성을 갖는 오프-표적 서열과 표적 서열을 구별할 수 있는 CRISPR-Cas 시스템을 생성시키는 돌연변이를 포함한다는 것을 보여준다 (예를 들어, 18 개 뉴클레오티드를 갖는 표적과 1, 2 또는 3 개 미스매치를 갖는 18 개 뉴클레오티드의 오프-표적을 구별함). 따라서, 본 발명의 상황에서, 가이드 서열과 이의 상응하는 표적 서열 사이의 상보성 정도는 94.5% 또는 95% 또는 95.5% 또는 96% 또

는 96.5% 또는 97% 또는 97.5% 또는 98% 또는 98.5% 또는 99% 또는 99.5% 또는 99.9%, 또는 100% 초과이다. 오프 표적은 서열과 가이드 사이에, 100% 또는 99.9% 또는 99.5% 또는 99% 또는 99% 또는 98.5% 또는 98% 또는 97.5% 또는 97% 또는 96.5% 또는 96% 또는 95.5% 또는 95% 또는 94.5% 또는 94% 또는 93% 또는 92% 또는 91% 또는 90% 또는 89% 또는 88% 또는 87% 또는 86% 또는 85% 또는 84% 또는 83% 또는 82% 또는 81% 또는 80% 미만의 상보성이며, 오프 표적이 서열과 가이드 사이에, 100% 또는 99.9% 또는 99.5% 또는 99% 또는 99% 또는 98.5% 또는 98% 또는 97.5% 또는 97% 또는 96.5% 또는 96% 또는 95.5% 또는 95% 또는 94.5% 의 상보성인 것이 유리하다.

[0093] 본 발명에 따른 특히 바람직한 구현예에서, 가이드 RNA (Cas를 표적 유전자좌로 가이드할 수 있음)는 (1) 진핵 생물 세포에서 게놈 표적 유전자좌와 하이브리드화할 수 있는 가이드 서열; (2) tracr 서열; 및 (3) tracr 메이트 서열을 포함할 수 있다. (1) 내지 (3)은 모두 단일 RNA, 즉, sgRNA (5'에서 3' 배향으로 배열됨)에 존재할 수 있거나, 또는 tracr RNA는 가이드 및 tracr 서열을 함유하는 RNA와 상이한 RNA일 수 있다. tracr은 tracr 메이트 서열에 하이브리드화하고, CRISPR/Cas 복합체를 표적 서열로 유도한다. tracr RNA가 가이드 및 tracr 서열을 함유하는 RNA와 상이한 RNA 상에 있는 경우에, 각각의 RNA의 길이는 그들의 각각의 천연 길이로부터 짧아지기에 최적화될 수 있고, 각각은 독립적으로 세포의 RNase에 의해 분해로부터 보호하도록 화학적으로 변형되거나 또는 달리 안정성을 증가시킬 수 있다.

[0094] "tracrRNA" 서열 또는 유사체라는 용어는 하이브리드화를 위해 crRNA 서열과 충분한 상보성을 갖는 임의의 폴리뉴클레오티드 서열을 포함한다. 일부 구현예에서, 최적으로 정렬될 때 둘 중 더 짧은 것의 길이를 따라서 tracrRNA 서열과 crRNA 서열 사이의 상보성 정도는 약 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99% 이상이다. 일부 구현예에서, tracr 서열은 길이가 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50 이상의 뉴클레오티드이다. 일부 구현예에서, tracr 서열 및 crRNA 서열은, 둘 사이의 하이브리드화가 2차 구조, 예컨대 헤어핀을 갖는 전사물을 생성하도록 단일 전사물 내에 함유된다. 본 발명의 구현예에서, 전사물 또는 전사된 폴리뉴클레오티드 서열은 적어도 2개 이상의 헤어핀을 가진다. 바람직한 구현예에서, 전사물은 2, 3, 4 또는 5개의 헤어핀을 가진다. 본 발명의 추가적인 구현예에서, 전사물은 최대 5개의 헤어핀을 가진다. 헤어핀 구조에서, 최종 "N"의 서열 5' 부분 및 루프의 상류는 tracr 메이트 서열에 대응하고, 루프의 서열 3'의 부분은 tracr 서열에 대응한다. 일부 구현예에서, 시스템은 하나 이상의 crRNA를 포함한다. 예를 들어, 시스템은 둘 이상의 crRNA를 포함할 수 있다.

[0095] 일반적으로, 상보성 정도는 두 서열 중 더 짧은 것의 길이를 따라서, 가이드 서열 및 tracr 서열의 최적의 정렬에 관한 것이다. 최적 정렬은 임의의 적합한 정렬 알고리즘에 의해 결정될 수 있고, 이차 구조, 예컨대 sca 서열 또는 tracr 서열 내 자가-상보성에 대해 더욱 설명될 수 있다. 일부 구현예에서, 최적으로 정렬될 때 둘 중 더 짧은 것의 길이를 따라서 tracrRNA 서열과 crRNA 서열 사이의 상보성 정도는 약 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99% 이상이다.

[0096] 본 발명의 한 측면에서, 가이드는 5'-핸들 및 씨드 영역 및 3'-말단을 추가로 포함하는 가이드 분절을 갖는, C2c1에 대한 변형된 crRNA를 포함한다. 일부 구현예에서, 변형된 가이드는 표 1 및 2에 열거된 오솔로그 중 어느 하나의 C2c1과 함께 사용될 수 있다.

[0097] ***변형된 가이드***

[0098] 일정 구현예에서, 본 발명의 가이드는 비-천연 발생 핵산 및/또는 비-천연 발생 뉴클레오티드 및/또는 뉴클레오티드 유사체, 및/또는 화학적 변형을 포함한다. 비-천연 발생 핵산은 예를 들어 천연 및 비-천연 뉴클레오티드의 혼합물을 포함할 수 있다. 비천연 발생 뉴클레오티드 및/또는 뉴클레오티드 유사체는 리보스, 포스페이트, 및/또는 염기 모이어티에서 변형될 수 있다. 본 발명의 일 구현예에서, 가이드 핵산은 리보뉴클레오티드 및 비-리보뉴클레오티드를 포함한다. 이러한 일 구현예에서, 가이드는 하나 이상의 리보뉴클레오티드 및 하나 이상의 데옥시리보뉴클레오티드를 포함한다. 본 발명의 일 구현예에서, 가이드는 하나 이상의 비-천연 발생 뉴클레오티드 또는 뉴클레오티드 유사체, 예컨대 포스포로티오에이트 결합, 보라노포스페이트 결합이 있는 뉴클레오티드, 리보스 고리의 2' 및 4' 탄소 사이에 메틸렌 브릿지를 포함하는 잠금 핵산(LNA) 뉴클레오티드, 펩티드 핵산 (PNA), 또는 브릿지된 핵산(BNA)을 포함한다. 변형된 뉴클레오티드의 다른 예는 2'-O-메틸 유사체, 2'-데옥시 유사체, 2-티오우리딘 유사체, N6-메틸아데노신 유사체, 또는 2'-플루오로 유사체를 포함한다. 변형된 뉴클레오티드의 추가적인 예는 펩티드, 핵 국재화 서열(NLS), 펩티드 핵산(PNA), 폴리에틸렌 글리콜(PEG), 트라이에틸렌 글리콜 또는 테트라에틸렌글리콜(TEG)을 포함하지만, 이들로 제한되지 않는 2' 위치에서 화학적 모이어티의 결합을 포함한다. 변형된 염기의 추가의 예에는 2-아미노퓨린, 5-브로모-우리

딘, 슈도우리딘 (Ψ), N1-메틸슈도우리딘 (me1Ψ), 5-메톡시우리딘(5moU), 이노신, 7-메틸구아노신이 포함되나 이들에 한정되지 않는다. 가이드 RNA 화학적 변형의 예는 하나 이상의 말단 뉴클레오티드에서 2'-O-메틸(M), 2'-O-메틸 3'-포스포로티오에이트(MS), 포스포로티오에이트(PS), S-제약형 에틸 (cEt), 2'-O-메틸 3'-티오PACE(MSP) 또는 2'-O-메틸-3'-포스포노아세테이트 (MP)의 도입을 포함하지만, 이들로 제한되지 않는다. 이러한 화학적으로 변형된 가이드는, 온 (on)-표적 대 오프 (off)-표적 특이성이 예측가능하지 않지만, 비변형 가이드에 비해 증가된 안정성 및 증가된 활성을 포함할 수 있다. (2015년 6월 29일자로 온라인 공개된, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290; Ragdarm et al., 0215, PNAS, E7110-E7111; Allerson et al., J. Med. Chem. 2005, 48:901-904; Bramsen et al., Front. Genet., 2012, 3:154; Deng et al., PNAS, 2015, 112:11870-11875; Sharma et al., MedChemComm., 2014, 5:1454-1471; Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Li et al., Nature Biomedical Engineering, 2017, 1, 0066 DOI:10.1038/s41551-017-0066; Ryan et al., Nucleic Acids Res. (2018) 46(2): 792-803 참조).

[0099] 일부 구현예에서, 가이드에 대한 변형은 화학적 변형, 삽입, 결실 또는 분할이다. 일부 구현예에서, 화학적 변형은 제한없이 2'-O-메틸 (M) 유사체, 2'-데옥시 유사체, 2-티오우리딘 유사체, N6-메틸아데노신 유사체, 2'-플루오로 유사체, 2-아미노푸린, 5-브로모-우리딘, 슈도우리딘 (Ψ), N1-메틸슈도우리딘 (me1Ψ), 5-메톡시우리딘 (5moU), 이노신, 7-메틸구아노신, 2'-O-메틸 -3'포스포로티오에이트 (MS), S-제약형 에틸(cEt), 포스포로티오에이트 (PS), 또는 2'-O-메틸 -3'-티오PACE (MSP), 또는 2'-O-메틸-3'-포스포노아세테이트 (MP)의 도입을 포함한다. 일부 구현예에서, 가이드는 하나 이상의 포스포로티오에이트 변형을 포함한다. 특정 구현예에서, 가이드의 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 또는 25 개 뉴클레오티드가 화학적으로 변형된다. 일부 구현예에서, 모든 뉴클레오티드는 화학적으로 변형된다. 일정 구현예에서, 씨드 영역 내 하나 이상의 뉴클레오티드가 화학적으로 변형된다. 일정 구현예에서, 3'-말단 내의 하나 이상의 뉴클레오티드가 화학적으로 변형된다. 일정 구현예에서, 5'-핸들에서 어떠한 뉴클레오티드 서열도 화학적으로 변형되지 않는다. 일부 구현예에서, 씨드 영역에서의 화학적 변형은 작은 변형, 예컨대, 2'-플루오로 유사체의 도입이다. 특정 구현예에서, 씨드 영역의 하나의 뉴클레오티드는 2'-플루오로 유사체로 대체된다. 일부 구현예에서, 3'-말단에서 5 또는 10개 뉴클레오티드가 화학적으로 변형된다. Cpf1 CrRNA의 3'-말단에서의 이러한 화학적 변형은 유전자 절단 효율을 향상시킨다 (참조, Li, et al., Nature Biomedical Engineering, 2017, 1:0066). 특정 구현예에서, 3'-말단 내의 5개의 뉴클레오티드는 2'-플루오로 유사체로 대체된다. 특정 구현예에서, 3'-말단 내의 10개의 뉴클레오티드는 2'-플루오로 유사체로 대체된다. 특이적 구현예에서, 3'-말단 내의 5 개 뉴클레오티드는 2'-O-메틸 (M) 유사체로 대체된다. 일부 구현예에서, 3' 및 5' 말단부 각각에서 3개의 뉴클레오티드는 화학적으로 변형된다. 구체적 구현예에서, 변형은 2'-O-메틸 또는 포스포로티오에이트 유사체를 포함한다. 구체적 구현예에서, 테트라루프에서 12개의 뉴클레오티드 및 줄기-루프 영역에서 16개의 뉴클레오티드는 2'-O-메틸 유사체로 대체된다. 이러한 화학적 변형은 생체내 편집 및 안정성을 개선시킨다(문헌[Finn et al., Cell Reports (2018), 22:2227-2235] 참조).

[0100] 일부 구현예에서, 가이드 RNA 의 5' 및/또는 3' 말단은 형광 염료, 폴리에틸렌 글리콜, 콜레스테롤, 단백질 또는 검출 표지를 비롯한, 다양한 기능적 모이어티에 의해 변형된다 (Kelly et al., 2016, J. Biotech. 233:74-83). 특정 구현예에서, 가이드는 표적 DNA 에 결합하는 영역 내에 리보뉴클레오티드 및 Cas9, Cpf1, 또는 C2c1 에 결합하는 영역 내에 하나 이상의 데옥시리보뉴클레오티드 및/또는 뉴클레오티드 유사체를 포함한다. 본 발명의 한 구현예에서, 데옥시리보뉴클레오티드 및/또는 뉴클레오티드 유사체는 조작된 가이드 구조, 예컨대, 제한 없이 5' 및/또는 3' 말단, 스템-루프 영역, 및 씨드 영역에 혼입된다. 특정 구현예에서, 변형은 스템-루프 영역의 5'-핸들에 있지 않다. 가이드의 스템-루프 영역의 5'-핸들에서의 화학적 변형은 그 기능을 소멸시킬 수 있다 (참조: Li, et al., Nature Biomedical Engineering, 2017, 1:0066). 특정 구현예에서, 가이드의 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 또는 75개의 뉴클레오티드가 화학적으로 변형된다. 일부 구현예에서, 가이드의 3' 또는 5' 말단에서의 3-5 개 뉴클레오티드가 화학적으로 변형된다. 일부 구현예에서, 2'-F 변형과 같은 작은 변형 만이 씨드 영역에 도입된다. 일부 구현예에서, 2'-F 변형은 가이드의 3 '말단에 도입된다. 일정 구현예에서, 가이드의 5' 및/또는 3' 말단부에서 3 내지 5개의 뉴클레오티드는 2'-O-메틸(M), 2'-O-메틸-3'-포스포로티오에이트(MS), S-제약 에틸(cEt), 2'-O-메틸-3'-티오PACE(MSP) 또는 2'-O-메틸-3'-포스포노아세테이트(MP)로 화학적으로 변형된다. 이러한 변형은 게놈 편집 효율을 증강시킬 수 있다 (Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Ryan et al., Nucleic Acids Res. (2018) 46(2): 792-803 참조). 일정 구현예에서, 가이드의 모든 포스포디에스테르 결합은 유전자 파괴 수준을 향상시키기 위해 포스포로티오에이트 (PS)로 치환된다. 일정 구현예에서, 가이드의 5' 및/또는 3' 말단에서 5개 초과의

뉴클레오티드가 2'-O-Me, 2'-F 또는 S-제약형 에틸(cEt)로 화학적으로 치환된다. 이러한 화학적으로 변형된 가이드는 유전자 파괴의 증강된 수준을 매개할 수 있다 (Ragdarm et al., 0215, PNAS, E7110-E7111 참조). 본 발명의 일 구현예에서, 가이드는 이의 3' 및/또는 5' 말단에서 화학적 모이어티를 포함하도록 변형된다. 이러한 모이어티는 아민, 아자이드, 알카인, 티오, 다이벤조사이클로옥틴(DBCO), 로다민, 펩티드, 핵 국재화 서열(NLS), 펩티드 핵산(PNA), 폴리에틸렌 글리콜(PEG), 트라이에틸렌 글리콜 또는 테트라에틸렌글리콜(TEG)을 포함하지만, 이들로 제한되지 않는다. 일정 구현예에서, 화학적 모이어티는 링커, 예컨대, 알킬쇄에 의해 가이드에 접합된다. 일정 구현예에서, 변형된 가이드의 화학적 모이어티는 다른 분자, 예컨대, DNA, RNA, 단백질 또는 나노입자에 가이드를 부착시키는 데 사용될 수 있다. 이러한 화학적으로 변형된 가이드는 CRISPR 시스템에 의해 유전자 편집된 세포를 동정하거나 또는 농축시키는 데 사용될 수 있다(문헌[Lee et al., eLife, 2017, 6:e25312, DOI:10.7554] 참조). 일부 구현예에서, 3' 및 5' 말단 각각에서 3개의 뉴클레오티드는 화학적으로 변형된다. 구체적 구현예에서, 변형은 2'-O-메틸 또는 포스포로티오에이트 유사체를 포함한다. 구체적 구현예에서, 테트라루프에서 12개의 뉴클레오티드 및 줄기-루프 영역에서 16개의 뉴클레오티드는 2'-O-메틸 유사체로 대체된다. 이러한 화학적 변형은 생체내 편집 및 안정성을 개선시킨다(문헌[Finn et al., Cell Reports (2018), 22:2227-2235] 참조). 일부 구현예에서, 가이드의 60 또는 70개 초과의 뉴클레오티드는 화학적으로 변형된다. 일부 구현예에서, 이 변형은 뉴클레오티드의 2'-O-메틸 또는 2'-플루오로 뉴클레오티드 유사체로의 대체 또는 포스포로다이에스터 결합의 포스포로티오에이트(PS) 변형을 포함한다.

일부 구현예에서, 화학적 변형은 CRISPR 복합체가 형성될 때 뉴클레아제 단백질 외부로 연장되는 가이드 뉴클레오티드의 2'-O-메틸 또는 2'-플루오로 변형 또는 가이드의 3'-말단의 20 내지 30개 이상의 뉴클레오티드의 PS 변형을 포함한다. 특정 구현예에서, 화학적 변형은 씨드 및 꼬리 영역에서 가이드 또는 2'-플루오로 유사체의 5' 말단부에서 2'-O-메틸 유사체를 추가로 포함한다. 이러한 화학적 변형은 뉴클레아제 분해에 대한 안정성을 개선시키고, 게놈-편집 활성 또는 효율을 유지하거나 향상시키지만, 모든 뉴클레아제의 변형은 가이드의 작용을 없앨 수 있다 (Yin et al., Nat. Biotech. (2018), 35(12):1179- 1187). 이러한 화학적 변형은 제한된 수의 뉴클레아제 및 RNA 2'-OH 상호작용의 지식을 비롯한, CRISPR 복합체 구조의 지식에 의해 가이드될 수 있다 (Yin et al., Nat. Biotech. (2018), 35(12):1179-1187). 일부 구현예에서, 하나 이상의 가이드 RNA 뉴클레오티드는 DNA 뉴클레오티드로 대체될 수 있다. 일부 구현예에서, 5'-말단부 꼬리/씨드 가이드 영역의 2, 4, 6, 8, 10 또는 12개까지의 RNA 뉴클레오티드는 DNA 뉴클레오티드로 대체된다. 일정 구현예에서, 3' 말단부에서 대다수의 가이드 RNA 뉴클레오티드는 DNA 뉴클레오티드로 대체된다. 특정 구현예에서, 3' 말단부에서 16개의 가이드 RNA 뉴클레오티드는 DNA 뉴클레오티드로 대체된다. 특정 구현예에서, 5'-말단부 꼬리/씨드 영역의 8개의 가이드 RNA 뉴클레오티드 및 3' 말단부에서 16개의 RNA 뉴클레오티드는 DNA 뉴클레오티드로 대체된다. 특정 구현예에서, CRISPR 복합체가 형성될 때 뉴클레아제 단백질 외부로 연장되는 가이드 RNA 뉴클레오티드는 DNA 뉴클레오티드로 대체된다. 다중 RNA 뉴클레오타이드의 DNA 뉴클레오타이드로의 이러한 대체는 감소된 오프-표적 활성이지만, 비변형 가이드에 비해 유사한 온-표적 활성을 야기하며; 그러나, 3' 말단에서 모든 RNA 뉴클레오타이드의 대체는 가이드 작용을 없앨 수 있다 (Yin et al., Nat. Chem. Biol. (2018) 14, 311-316). 이러한 변형은 제한된 수의 뉴클레아제 및 RNA 2'-OH 상호작용의 지식을 비롯한, CRISPR 복합체의 구조 지식에 의해 가이드될 수 있다 (Yin et al., Nat. Chem. Biol. (2018) 14, 311-316).

[0101] 가이드 서열, 및 따라서 핵산-표적화 가이드 RNA는 임의의 표적 핵산 서열을 표적화하도록 선택될 수 있다. 표적 서열은 DNA일 수 있다. 표적 서열은 게놈 DNA일 수 있다. 표적 서열은 미토콘드리아 DNA일 수 있다. 클래스 2 V형 CRISPR-Cas 단백질의 가이드 분자 또는 가이드 RNA는 tracr-메이트 서열(내생성 CRISPR 시스템과 관련하여 "직접 반복부(direct repeat)"를 포함함) 및 가이드 서열(또한 내생성 CRISPR 시스템과 관련하여 "스페이서"로서 지칭됨)을 포함한다. 천연 Cas12b CRISPR-Cas 시스템은 tracr 서열을 적용한다.

[0102] 일정 구현예에서, (표적 유전자좌로 C2c1을 가이드할 수 있는) 가이드 분자는 (1) 표적 유전자좌와 하이브리드화할 수 있는 가이드 서열 및 (2) 직접 반복부 서열이 가이드 서열의 상류 (즉, 5')에 위치되는 것인 직접 반복부 서열 또는 tracr 메이트를 포함한다. 특정한 구현예에서, C2c1 가이드 서열의 씨드 서열 (즉, 표적 유전자좌에서 서열을 인식하고/하거나 하이브리드화하는데 결정적인 필수 서열)은 대략 가이드 서열의 처음 10개 뉴클레오티드 내에 존재한다. 특정 구현예에서, 씨드 서열은 가이드 서열의 5' 말단 상의 대략 처음 5 nt 내에 존재한다.

[0103] 일부 구현예에서, 가이드의 5'-핸들의 루프가 변형된다. 일부 구현예에서, 가이드의 5'-핸들의 루프는 결실, 삽입, 분할 또는 화학적 변형을 갖도록 변형된다. 일정 구현예에서, 변형된 루프는 3, 4 또는 5개의 뉴클레오티드를 포함한다. 일정 구현예에서, 루프는 UCUU, UUUU, UAUU, 또는 UGUU의 서열을 포함한다. 일부 구현예에서, 가이드 분자는 DNA 또는 RNA일 수 있는 별개의 비공유적으로 연결된 서열을 갖는 스템루프를

형성한다.

[0104] ***스템 루프 & 헤어핀***

[0105] 핵산-표적화 복합체 또는 시스템과 관련하여 바람직하게, crRNA 서열 및 키메라 가이드 서열은 하나 이상의 스템 루프 또는 헤어핀을 포함할 수 있다. 압타머-변형된 가이드의 사용은 가이드에 어댑터-함유 단백질의 결합을 허용한다. 어댑터는 임의의 기능성 도메인에 융합될 수 있어서, 가이드에 기능성 도메인의 부착을 제공한다. 2종의 상이한 압타머의 사용은 2종 가이드에 의해 별도 표적화를 허용한다. 이러한 변형된 핵산-표적화 가이드 RNA의 다수, 예를 들어 10 개 또는 20 개 또는 30 개 등이 모두 동시에 사용될 수 있는 한편, 비교적 적은 수의 com 단백질 분자가 다수의 변형된 가이드와 함께 사용될 수 있으므로, 오직 하나 (또는 적어도 최소수)의 이펙터 단백질 분자가 전달되는 것이 필요하다. 어댑터 단백질과 기능성 도메인 예컨대 활성인자 또는 억제인자 간 융합은 링커를 포함할 수 있다. 예를 들어, GlySer 링커 GGGS가 사용될 수 있다. 그들은 필요에 따라, 적합한 길이를 제공하기 위해서, 3 $(GGGGS)_3$ (SEQ ID NO:393) 또는 6 (SEQ ID NO:394), 9 (SEQ ID NO:395), 또는 심지어 12 (SEQ ID NO:396) 또는 그 이상의 반복부로 사용될 수 있다. 링커는 가이드 RNA와 기능적 도메인(활성화제 또는 억제제) 사이, 또는 핵산-표적화 Cas 단백질(Cas)과 기능적 도메인(활성화제 또는 억제제) 사이에서 사용될 수 있다. 사용자는 링커를 "기계적 유연성"의 적절한 양으로 조작한다.

[0106] 특정 구현예에서, 스템은 상보성 X 및 Y 서열을 포함하는 적어도 약 4 bp를 포함하지만, 더 많거나, 예를 들어, 5, 6, 7, 8, 9, 10, 11 또는 12개, 또는 더 적은, 예를 들어, 3, 2개 염기쌍의 스템이 또한 고려된다. 따라서, 예를 들어 X2-10 및 Y2-10 (여기서 X 및 Y는 뉴클레오티드의 임의의 상보성 세트를 나타냄)이 고려될 수 있다. 일 양상에서, 루프와 함께, X 및 Y 뉴클레오티드로 만들어진 스템은 전체 2차 구조에서 완전한 헤어핀을 형성하게 될 것이고, 이것은 유리할 수 있으며, 염기쌍의 양은 완전한 헤어핀을 형성하는 임의의 양일 수 있다. 일 양상에서, 임의의 상보성 X:Y 염기쌍 서열 (예를 들어, 길이에 대해)은 전체 가이드 분자의 2차 구조가 보존되는 한, 허용된다. 일 양상에서, X:Y 염기쌍으로 만들어진 스템을 연결시키는 루프는 가이드 분자의 전체 2차 구조를 파괴하지 않는 동일 길이 (예를 들어, 4 또는 5개 뉴클레오티드)이거나 또는 더 긴 임의 서열일 수 있다. 일 양상에서, 스템루프는 예를 들어, MS2 압타머를 더 포함할 수 있다. 일 양상에서, 스템은 상보성 X 및 Y 서열을 포함하는 약 5-7 bp를 포함하지만, 그 이상 또는 그 이하의 염기쌍의 스템도 고려된다. 일 양상에서, 달리 이러한 쌍형성이 일반적으로 그 위치에서 스템루프의 아키텍처를 보존하는 경우에, 비-왓슨 크릭 염기쌍이 고려된다.

[0107] 특정 구현예에서, 가이드 분자의 천연 헤어핀 또는 스템-루프 구조는 연장되거나 또는 연장된 스템-루프에 의해 대체된다. 일부 경우에 스템의 연장은 CRISPR-Cas 단백질과 가이드 분자의 조립을 증강시킬 수 있다는 것이 입증되었다 (Chen et al. Cell. (2013); 155(7): 1479- 1491). 특정 구현예에서 스템루프의 스템은 적어도 1, 2, 3, 4, 5 또는 그 이상의 상보성 염기쌍 (즉, 가이드 분자 내 2, 4, 6, 8, 10개 또는 그 이상의 뉴클레오티드의 첨가에 상응)만큼 연장된다. 특정 구현예에서 이들은 스템루프의 루프에 인접하는, 스템의 말단에 위치된다.

[0108] 일부 구현예에서, 가이드 분자는 DNA 또는 RNA일 수 있는 별개의 비공유적으로 연결된 서열과 스템 루프를 형성한다. 특정 실시형태에서, 가이드를 형성하는 서열은 표준 포스포르아미다이트 합성 프로토콜을 사용해 먼저 합성된다 (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). 일부 구현예에서, 이들 서열은 당업자에게 공지된 표준 프로토콜을 사용하여 결찰을 위한 적절한 작용기를 함유하도록 작용화될 수 있다 (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). 작용기의 예는 제한없이 히드록실, 아민, 카르복실산, 카르복실산 할라이드, 카르복실산 활성 에스테르, 알데히드, 카르보닐, 클로로카르보닐, 이미다졸릴카르보닐, 히드로지드, 세미카르바지드, 티오 세미카르바지드, 티올, 말레이미드, 할로알킬, 술포닐, 알릴, 프로파르길, 디엔, 및 아지드를 포함한다. 이러한 서열이 작용화되면, 공유적 화학 결합 또는 연결부가 이러한 서열과 직접 반복부 서열 사이에 형성될 수 있다. 화학적 결합의 예는 제한없이 카바메이트, 에테르, 에스테르, 아미드, 이민, 아미딘, 아미노트리진, 히드로존, 디술피드, 티오에테르, 티오에스테르, 포스포로티오에이트, 포스포로디티오에이트, 술폰아미드, 술포네이트, 풀폰, 술폭시드, 우레아, 티오우레아, 히드라지드, 옥심, 트리아졸, 광불안정 연결부, C-C 결합 형성기 예컨대 딜스-알더 (Diels-Alder) 고리-부가 쌍 또는 고리-폐쇄 복분해 쌍, 및 마이클 (Michael) 반응 쌍을 기반으로 하는 것을 포함한다.

[0109] 일부 구현예에서, 이들 스템-루프 형성 서열이 화학적으로 합성될 수 있다. 일부 구현예에서, 화학 합성은

2'-아세톡시에틸 오르쏘에스테르 (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) 또는 2'-티오노카바메이트 (2'-TC) 화학 (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989)을 사용하는 자동화, 고체상 올리고뉴클레오티드 합성 기계를 사용한다.

[0110] ***감소된 RNase 감수성***

[0111] 일부 구현예에서, RNA 절단, 예컨대 Cas12b에 의한 절단에 대한 가이드 분자의 감수성을 감소시키는 것에 관심을 갖는다. 따라서, 특정 구현예에서, 가이드 분자는 Cas12b 또는 다른 RNA-절단 효소에 의한 절단을 피하도록 조정된다.

[0112] 특정 구현예에서, RNAse 또는 감소된 발현에 대한 가이드 분자의 감수성은 이의 기능에 영향을 미치지 않는 가이드 분자의 서열의 약간의 변형에 의해 감소될 수 있다. 예를 들어, 특정 구현예에서, 전사물의 미성숙 종결, 예컨대 U6 Pol-III의 미성숙 전사는 가이드 분자 서열 내 추정의 Pol-III 종결자 (4개의 연속 U)를 변형시켜 제거될 수 있다. 이러한 서열 변형이 가이드 분자의 스템루프에서 필요한 경우에, 바람직하게는 염기쌍 플립에 의해 보장된다.

[0113] ***감소된 2차 구조***

[0114] 일부 구현예에서, 가이드 분자 (직접 반복부 및/또는 스페이서)의 서열은 가이드 분자 내 2차 구조 정도를 감소시키도록 선택된다. 일부 구현예에서, 핵산-표적화 가이드 RNA의 뉴클레오티드의 약 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1% 이하가 최적으로 폴딩되었을 때 자기-상보성 염기쌍에 참여한다. 최적 폴딩은 임의의 적합한 폴리뉴클레오티드 폴딩 알고리즘을 통해 결정될 수 있다. 일부 프로그램은 최소 깁스 (Gibbs) 자유 에너지의 계산을 기반으로 한다. 이러한 알고리즘의 예는 [Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148)]에 기술된 바와 같은, mFold이다. 또 다른 폴딩 알고리즘의 예는 중심 구조 예측 알고리즘을 사용하여, 비엔나 대학의 이론 화학 연구소 (Institute for Theoretical Chemistry at the University of Vienna) 에서 개발한 온라인 웹서버 RNAfold 이다 (예를 들어, [A.R. Gruber et al., 2008, Cell 106(1): 23-24]; 및 [PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62] 참조).

[0115] ***접합된 tracr 서열***

[0116] 일부 구현예에서, 가이드 분자는 비-포스포다이에스터 결합을 통해 화학적으로 연결되거나 또는 접합된 tracr 서열 및 tracr 메이트 서열을 포함한다. 일 양상에서, 가이드는 비-뉴클레오티드 루프를 통해 화학적으로 연결되거나 또는 접합된 tracr 서열 및 tracr 메이트 서열을 포함한다. 일부 구현예에서, tracr 및 tracr 메이트 서열은 비-포스포다이에스터 공유 링커를 통해 결합된다. 공유 링커의 예는 카바메이트, 에터, 에스터, 아마이드, 이민, 아미딘, 아미노트라이진, 하이드로존, 이황화물, 티오에터, 티오에스터, 포스포로티오에이트, 포스포로다이티오에이트, 설폰아마이드, 설포네이트, 풀폰, 설폭사이드, 유레아, 티오유레아, 하이드라자이드, 옥심, 트리아졸, 광 분해성 결합, C-C 결합 형성기, 예컨대, 딜스-알더(Diels-Alder) 고리 첨가 쌍 또는 고리-폐쇄 복분해 쌍, 및 마이클 반응(Michael reaction) 쌍에 기반한 것을 포함하지만, 이들로 제한되지 않는다.

[0117] 일부 실시형태에서, tracr 및 tracr 메이트 서열은 표준 포스포르아미다이트 합성 프로토콜을 이용하여 처음 합성된다(Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis:Methods and Applications, Humana Press, New Jersey (2012)). 일부 구현예에서, tracr 및 tracr 메이트 서열은 당업계에 공지된 표준 프로토콜을 이용하여 결찰을 위한 적절한 작용기를 함유하도록 작용기화될 수 있다 (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). 작용기의 예는 제한 없이, 히드록실, 아민, 카르복실산, 카르복실산 할라이드, 카르복실산 활성 에스테르, 알데히드, 카르보닐, 클로로카르보닐, 이미다졸릴카르보닐, 히드로지드, 세미카르바지드, 티오 세미카르바지드, 티올, 말레이미드, 할로알킬, 수포닐, 알릴, 프로파르길, 디엔, 알킨 및 아지드를 포함한다. 일단 tracr 및 tracr 메이트 서열이 작용기화되면, 두 올리고뉴클레오티드 사이에 공유 화학 결합 또는 연결이 형성될 수 있다. 화학적 결합의 예는 제한 없이, 카르바메이트, 에테르, 에스테르, 아미드, 이민, 아미드, 아미노트리진, 히드로존, 디설피드, 티오에테르, 티오에스테르, 포스포로티오에이트, 포스포로디티오에이트, 설폰아미드, 설포네이트, 풀폰, 설폭시드, 우레아, 티오우레아, 히드라지드, 옥심, 트리아졸, 광불안정성 결합, 딜스-알더 시클로-부가 쌍 또는 고리-폐쇄 복분해 쌍, 및 마이클 반응 쌍과 같은 C-C 결합 형성기를 기본으로 하는 것을 포함한다.

[0118] 일부 구현예에서, tracr 및 tracr 메이트 서열은 화학적으로 합성될 수 있다. 일부 구현예에서, 화학 합성은 2'-아세톡시에틸 오르쏘에스테르 (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821;

Scaringe, Methods Enzymol. (2000) 317: 3-18) 또는 2'-티오노카바메이트 (2'-TC) 화학 (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989)을 사용하는 자동화, 고체상 올리고뉴클레오티드 합성 기계를 사용한다.

[0119] 일부 구현예에서, tracr 및 tracr 메이트 서열은 다양한 생체접합 반응, 루프, 브리지, 및 당, 뉴클레오티드간 포스포다이에스터 결합, 퓨린 및 피리미딘 잔기의 변형을 통한 비-뉴클레오티드 연결을 이용하여 공유적으로 연결될 수 있다. Sletten et al., Angew. Chem. Int. Ed. (2009) 48:6974-6998; Manoharan, M. Curr. Opin. Chem. Biol. (2004) 8: 570-9; Behlke et al., Oligonucleotides (2008) 18: 305-19; Watts, et al., Drug. Discov. Today (2008) 13: 842-55; Shukla, et al., ChemMedChem (2010) 5: 328-49.

[0120] 일부 구현예에서, tracr 및 tracr 메이트 서열은 클릭 화학을 이용하여 공유적으로 연결될 수 있다. 일부 구현예에서, tracr 및 tracr 메이트 서열은 트리아졸 링커를 이용하여 공유적으로 연결될 수 있다. 일부 실시형태에서, tracr 및 tracr 메이트 서열은 알카인 및 아자이드를 수반하는 후이스겐(Huisgen) 1,3-쌍극자 첨가환화 반응을 이용하여 공유적으로 연결되어 고도로 안정한 트리아졸 링커를 수득할 수 있다 (He et al., ChemBioChem (2015) 17:1809- 1812; WO 2016/186745). 일부 구현예에서, tracr 및 tracr 메이트 서열은 5'-헥신 tracrRNA 및 3'-아자이드 crRNA를 결찰시킴으로써 공유적으로 연결된다. 일부 실시형태에서, 5'-헥신 tracrRNA 및 3'-아자이드 crRNA 중 하나 또는 둘 다 다르마콘(Dharmacon) 프로토콜을 이용하여 후속적으로 제거될 수 있는 2'-아세톡시에틸 오쏘에스터(2'-ACE) 기로 보호될 수 있다 (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3- 18).

[0121] 일부 구현예에서, tracr 및 tracr 메이트 서열은 스페이서, 부착, 생체접착, 발색단, 리포터 기, 염료 표지된 RNA, 및 비천연유래 뉴클레오티드 유사체와 같은 모이어티를 포함하는, 링커(예를 들어, 비-뉴클레오티드 루프)를 통해 공유적으로 연결될 수 있다. 더 구체적으로는, 본 발명의 목적을 위한 적합한 스페이서는 폴리에터(예를 들어, 폴리에틸렌 글리콜, 폴리알코올, 폴리프로필렌 글리콜 또는 에틸렌과 프로필렌 글리콜의 혼합물), 폴리아민기(예를 들어, 스페닌, 스퍼미딘 및 이들의 중합체 유도체), 폴리에스터(예를 들어, 폴리(에틸 아크릴레이트)), 폴리포스포다이에스터, 알킬렌 및 이들의 조합물을 포함하지만, 이들로 제한되지 않는다. 적합한 부착은 링커에 추가적인 특성, 예컨대 이하로 제한되는 것은 아니지만, 형광 표지를 더하기 위해 링커에 부가되는 임의의 모이어티를 포함한다. 적합한 생체접합체는 펩티드, 글리코사이드, 지질, 콜레스테롤, 인지질, 다이아실 글리세롤 및 다이알킬 글리세롤, 지방산, 탄화수소, 효소 기질, 스테로이드, 바이오틴, 디곡시게닌, 탄수화물, 다당류를 포함하지만, 이들로 제한되지 않는다. 적합한 발색단, 리포터 기 및 염료 표지된 RNA는 형광 염료, 예컨대 플루오레세인 및 로다민, 화학발광, 전자화학발광 및 생체발광 마커 화합물을 포함하지만, 이들로 제한되지 않는다. 2개의 RNA 성분을 접합하는 예시적 링커의 설계는 또한 WO 2004/015075에 기재되어 있다.

[0122] 링커(예를 들어, 비-뉴클레오티드 루프)는 임의의 길이를 가질 수 있다. 일부 구현예에서, 링커는 약 0-16개 뉴클레오티드와 동등한 길이를 갖는다. 일부 구현예에서, 링커는 약 0-8개 뉴클레오티드와 동등한 길이를 갖는다. 일부 구현예에서, 링커는 약 0-4개 뉴클레오티드와 동등한 길이를 갖는다. 일부 구현예에서, 링커는 약 2개 뉴클레오티드와 동등한 길이를 갖는다. 링커 디자인의 예는 또한 WO2011/008730에 기재되어 있다.

[0123] 전형적인 Cas9 sgRNA는 (5'에서 3' 방향으로):가이드 서열, 폴리 U 트랙, 제1 상보성 스트레치 ("반복부"), 루프 (테트라루프), 제2 상보성 스트레치 ("안티-반복부"는 반복부에 상보성임), 스템 및 추가 스템 루프 및 스템 및 폴리 A(RNA에서 종종 폴리 U) 꼬리부 (종결자)를 포함한다. 전형적인 Cas12b sgRNA는 유사한 성분을 포함하지만, 반대 배향으로, 즉 3'에서 5' 방향으로 포함한다. 직접 반복부 (DR)는 tracrRNA와 하이브리드화하여서 crRNA:tracrRNA 듀플렉스를 형성하고, 이것은 Cas12b 상에 적재되어 DNA 인식 및 절단을 가이드한다. Cas12b는 DNA 간섭을 매개하도록 프로토스페이서 서열의 5' 말단에서 T-풍부 PAM을 인식한다. 일정 구현예에서, tracr 의 5' 말단은 스템-루프를 형성한다. 일정 구현예에서, tracrRNA 및 5' DR의 뉴클레오티드는 반복부;안티-반복부 듀플렉스를 형성한다. 일정 구현예에서, sgRNA 아키텍처는 [Shmakov et al., 2015, Molecular Cell 60, 385-397]가 예측한 구조와 일치한다. 일정 구현예에서, sgRNA 아키텍처는 [Liu et al., 2017, Molecular Cell 65, 310-322]가 예측한 구조와 일치한다. 바람직한 구현예에서, 가이드 아키텍처의 소정의 양상은 보유되고, 가이드 아키텍처의 소정의 양상은, 예를 들어, 특징의 부가, 차감 또는 치환에 의해 변형될 수 있는 반면, 가이드 구조의 소정의 다른 양상은 유지된다. 삽입, 결실 및 치환을 포함하지만, 이들로 제한되지 않는 조작된 sgRNA 변형에 대한 바람직한 위치는 CRISPR 단백질 및/또는 표적, 예를 들어, 테트라루프 및/또는 루프 2와 복합체를 형성할 때, 노출되는 sgRNA의 가이드 말단 및 영역을

포함한다.

[0124] 일정 구현예에서, 본 발명의 가이드는 (예를 들어, 융합 단백질을 통해) 하나 이상의 기능성 도메인을 포함할 수 있는 어댑터 단백질에 대한 특정 결합 부위(예를 들어, 어댑터)를 포함한다. 이러한 가이드가 CRISPR 복합체(즉, 가이드 및 표적에 결합하는 CRISPR 효소)를 형성할 때, 어댑터 단백질이 결합하며, 어댑터 단백질과 결합된 기능성 도메인은 결과된 기능이 유효하게 되는데 유리한 공간적 배향으로 위치된다. 예를 들어, 기능성 도메인이 전사 활성인자 (예를 들어, VP64 또는 p65)라면, 전사 활성인자는 표적의 전사에 영향을 미치도록 허용하는 공간적 배향으로 위치된다. 마찬가지로, 전사 억제인자는 표적의 전사에 영향을 미치도록 유리하게 위치될 것이며, 뉴클레아제(예를 들어, Fok1)는 표적을 절단하거나 또는 부분적으로 절단하도록 유리하게 위치될 것이다.

[0125] 당업자는 어댑터 + 기능성 도메인의 결합을 가능하게 하는 가이드에 대한 변형을 이해할 것이지만, 어댑터 + 기능성 도메인의 적절하지 않은 위치화(예를 들어, CRISPR 복합체의 3차원 구조 내의 입체장애에 기인)는 의도되지 않은 변형이다. 하나 이상의 변형된 가이드는 본 명세서에 기재된 바와 같은 테트라 루프, 스템 루프 1, 스템 루프 2, 또는 스템 루프 3에서, 바람직하게는 테트라 루프 또는 스템 루프 2에서, 가장 바람직하게는 테트라 루프와 스템 루프 2 둘 모두에서 변형될 수 있다.

[0126] 반복부:안티 반복부 이중가닥은 sgRNA의 2차 구조로부터 분명하게 될 것이다. 전형적인 Cas9 sgRNA에서, (5'에서 3' 방향으로) 폴리U 트랙 이후 및 테트라루프 앞에는 전형적으로 제1 상보성 스트레치일 것이고, (5'에서 3' 방향으로) 테트라루프 이후 및 폴리A 트랙 전에는 제2 상보성 스트레치일 것이다. 제1 상보성 스트레치("반복부")은 제2 상보성 스트레치("안티-반복부")에 상보성이다. 일정 구현예에서, Cas12b sgRNA의 아키텍처는 [Shmakov et al., 2015, Molecular Cell 60, 385-397]이 예측한 구조와 일치한다. 일정 구현예에서, Cas12b sgRNA 아키텍처는 [Liu et al., 2017, Molecular Cell 65, 310-322]이 예측한 구조와 일치한다. 이와 같이, 그들 sRNA는 다른 것 상에서 역으로 폴딩될 때 dsRNA의 듀플렉스를 형성하도록 왓슨-크릭 염기쌍을 포함한다. 이와 같이, 안티-반복부 서열은 A-U 또는 C-G 염기쌍이라는 관점에서 뿐만 아니라, 안티-반복부가 스템-루프 또는 다른 아키텍처 특성에 기인하여 역배향이라는 사실의 관점에서, 안티-반복부 서열은 반복부의 상보성 서열이다.

[0127] 본 발명의 구현예에서, 가이드 아키텍처의 변형은 스템루프 2에서 염기를 대체하는 단계를 포함한다. 예를 들어, 일부 구현예에서, 스템루프2에서 "actt"(RNA에서 "acuu") 및 "aagt"(RNA에서 "aagu") 염기는 "cgcc" 및 "gcgg"로 대체된다. 일부 구현예에서, 스템루프 2에서 "actt" 및 "aagt" 염기는 4개 뉴클레오티드의 상보성 GC-풍부 영역으로 대체된다. 일부 구현예에서, 4개 뉴클레오티드의 상보성 GC-풍부 영역은 "cgcc" 및 "gcgg"(둘 다 5'에서 3' 방향으로)이다. 일부 구현예에서, 4개의 뉴클레오티드의 상보성 GC-풍부 영역은 "gcgg" 및 "cgcc"(둘 다 5'에서 3' 방향으로)이다. 4개의 뉴클레오티드의 상보성 GC-풍부 영역에서 CCCC 및 GGGG를 비롯한 C와 G의 다른 조합이 자명해 질 것이다.

[0128] 일 양상에서, 스템루프 2, 예를 들어, "ACTTgtttAAGT" (SEQ ID NO:397)는 임의의 "XXXXgtttYYYY" (SEQ ID NO:398)로 대체될 수 있으며, 예를 들어, 여기서 XXXX 및 YYYY는 줄기를 생성하기 위해 서로 함께 염기쌍이 되는 뉴클레오티드의 임의의 상보성 세트를 나타낸다.

[0129] 일 양상에서, 상기 스템은, 더 많은, 예를 들어, 5, 6, 7, 8, 9, 10, 11 또는 12개 또는 더 적은 수, 예를 들어, 3, 2개의 염기쌍의 스템이 또한 상정되지만, 상보성 X 및 Y 서열을 포함하는 적어도 약 4bp를 포함한다. 따라서, 예를 들어 X2-12 및 Y2-12 (여기서 X 및 Y는 뉴클레오티드의 임의의 상보성 세트를 나타냄)이 고려될 수 있다. 일 양상에서, "gttt"와 함께, X 및 Y 뉴클레오티드로 만들어진 스템은 전체 2차 구조에서 완전한 헤어핀을 형성하게 될 것이고, 이것은 유리할 수 있으며, 염기쌍의 양은 완전한 헤어핀을 형성하는 임의의 양일 수 있다. 일 양상에서, 전체 sgRNA의 2차 구조가 보존된다면, 임의의 상보성 X:Y 염기쌍 서열이 (예를 들어, 길이에 대해) 용인된다. 일 양상에서, 스템은 DR:tracr 이중가닥, 및 3 스템루프를 가진다는 점에서 전체 sgRNA의 2차 구조를 방해하지 않는 X:Y 염기쌍의 형태일 수 있다. 일 양상에서, ACTT 및 AAGT(또는 X:Y 염기쌍으로 만들어진 임의의 대안의 줄기)를 연결하는 "gttt" 테트라루프는 sgRAN 분자의 전반적 2차 구조를 방해하지 않는 동일한 길이(예를 들어, 4개의 염기쌍) 또는 더 긴 임의의 서열일 수 있다. 일 양상에서, 스템루프는 스템루프2를 추가로 연장시키는 것일 수 있고, 예를 들어, MS2 압타머일 수 있다. 일 양상에서, 스템루프3 "GGCACCGagtCGGTGC" (SEQ ID NO:399)는 유사하게 "XXXXXXXagtYYYYYYY" (SEQ ID NO:400) 형태를 취할 수 있으며, 예를 들어, 여기서 X7 및 Y7은 스템을 생성하도록 서로 함께 염기쌍을 형성하게 되는 뉴클레오티드의 임의의 상보성 세트를 나타낸다. 일 양상에서, 스템은 더 많은 또는 더 적은 염기쌍의 스템이 또한 상정

된다고 해도, 상기 스템은 상보성 X 및 Y 서열을 포함하는 약 7 bp를 포함한다. 일 양상에서, "agt"와 함께 X 및 Y 뉴클레오티드로 이루어진 스템은 전반적인 2차 구조에서 완전한 헤어핀을 형성할 것이다. 일 양상에서, 전체 sgRNA의 2차 구조가 보존된다면, 임의의 상보성 X:Y 염기쌍 서열이 용인된다. 일 양상에서, 스템은 DR:tracr 이중가닥, 및 3개의 스템루프를 가진다는 점에서 전체 sgRNA의 2차 구조를 붕괴시키지 않는 X:Y 염기쌍의 형태일 수 있다. 일 양상에서, 스템루프 3의 "agt" 서열은 연장되거나 또는 압타머, 예를 들어, 스템루프3의 아키텍처를 달리 일반적으로 보존하는 MS2 압타머 또는 서열로 대체될 수 있다. 대안의 스템루프 2 및/또는 3에 대한 일 양상에서, 각각의 X 및 Y 쌍은 임의의 염기쌍을 지칭할 수 있다. 일 양상에서, 달리 이러한 쌍형성이 일반적으로 그 위치에서 스템루프의 아키텍처를 보존하는 경우에, 비-왓슨 크릭 염기쌍이 고려된다.

[0130] 일 양상에서, DR:tracrRNA 이중가닥은 형태:gYYYYag(N)NNNNxxxxNNNN(AAN)uuRRRRu (SEQ ID NO:401) (뉴클레오티드에 대한 표준 IUPAC 명명법을 이용)로 대체될 수 있으며, (N) 및 (AAN)은 이중가닥에서의 벌지 부분을 나타내고, "xxxx"는 링커 서열을 나타낸다. tracrRNA의 대응하는 NNNN 부분과 염기쌍을 이룬다면, 직접 반복부 상의 NNNN은 임의의 것일 수 있다. 일 양상에서, DR:tracrRNA 이중가닥은 전반적 구조를 변경시키지 않는다면, 임의의 길이의 링커(xxxx...), 임의의 염기 조성물에 의해 연결될 수 있다.

[0131] 일 양상에서, sgRNA 구조적 필요는 이중가닥 및 3개의 스템루프를 갖는 것이다. 대부분의 양상에서, 다수의 특정 염기 필요에 대한 실제 서열 필요는 lax이고, 즉, DR:tracrRNA 이중가닥의 아키텍처는 보존되어야 하지만, 아키텍처, 즉, 줄기, 루프, 벌지 등을 생성하는 서열은 변경될 수 있다.

[0132] 제1 압타머/RNA-결합 단백질 쌍을 갖는 하나의 가이드는 활성인자에 연결되거나 또는 융합될 수 있는 반면, 제2 압타머/RNA-결합 단백질 쌍을 갖는 제2 가이드는 억제인자에 연결된거나 또는 융합될 수 있다. 가이드는 상이한 표적(유전자좌)을 위한 것이며, 따라서 이는 하나의 유전자가 활성화되고 하나는 억제되게 한다. 예를 들어, 다음의 도식은 이러한 접근을 나타낸다:

[0133] 가이드 1 - MS2 압타머-------MS2 RNA-결합 단백질-------VP64 활성인자; 및

[0134] 가이드 2 - PP7 압타머-------PP7 RNA-결합 단백질-------SID4x 억제인자.

[0135] 본 발명은 또한 직교 PP7/MS2 유전자 표적화에 관한 것이다. 이런 예에서, 상이한 유전자좌를 표적화하는 sgRNA는 그들의 표적 유전자좌를 각각 활성화시키고 억제하는 MS2-VP64 또는 PP7-SID4X를 보충하기 위해 별개의 RNA 루프로 변형된다. PP7은 박테리오파지 슈도모나스(Pseudomonas)의 RNA-결합 외피 단백질이다. MS2처럼, 이것은 특별한 RNA 서열 및 2차 구조에 결합된다. PP7 RNA-인식 모티프는 MS2와 별개이다. 결과적으로, PP7 및 MS2는 상이한 게놈 유전자좌에서 동시에 별개의 효과를 매개하도록 다중복합체화될 수 있다. 예를 들어, sgRNA 표적화 유전자좌 A는 MS2 루프로 변형되어, MS2-VP64 활성인자를 보충할 수 있는 반면, 다른 sgRNA 표적화 유전자좌 B는 PP7 루프로 변형되어, PP7-SID4X 억제인자 도메인을 보충할 수 있다. 동일한 세포에서, dC2c1은 직교, 유전자좌-특이적 변형을 매개할 수 있다. 이 원칙은 다른 직교 RNA-결합 단백질, 예컨대 Q-베타를 혼입하도록 연장될 수 있다.

[0136] 직교 억제를 위한 대안의 선택은 (가이드 내에 통합된 MS2/PP7 루프와 유사한 위치에서 또는 가이드의 3' 말단에서) 가이드 내에 전사활성 억제 기능을 갖는 비암호 RNA를 혼입시키는 것을 포함한다. 예를 들어, 가이드는 비암호 (그러나 억제성인 것으로 알려짐) RNA 루프 (예를 들어, 포유류 세포에서 RNA 중합효소 II를 방해하는 (RNA에서의) Alu 억제인자를 이용)로 설계된다. Alu RNA 서열은 본 명세서에서 사용된 바와 같은 MS2 RNA 서열 대신에 (예를 들어, 테트라루프 및/또는 줄기 루프 2에서); 및/또는 가이드의 3' 말단에 위치되었다. 이는 테트라루프 및/또는 스템루프 2 위치에서 MS2, PP7 또는 Alu의 가능한 조합뿐만 아니라, 선택적으로, (링커와 함께 또는 링커 없이) 가이드의 3' 말단부에서 Alu의 첨가를 제공한다.

[0137] 두 상이한 압타머 (별개의 RNA)의 사용은 활성인자-어댑터 단백질 융합, 및 상이한 가이드를 갖는 사용될 억제인자-어댑터 단백질 융합을 가능하게 하여, 하나의 유전자의 발현을 활성화시키는 한편 다른 것은 억제한다. 상이한 가이드와 함께 그들은 다중복합 접근으로 함께 또는 실질적으로 함께 투여될 수 있다. 매우 다수의 이러한 변형된 가이드는 모두 동시에, 예를 들어, 10 또는 20 또는 30개 등이 사용될 수 있는 한편, 비교적 소수의 C2c1이 매우 다수의 변형된 가이드와 함께 사용됨에 따라, 오직 하나 (또는 적어도 최소 수)의 C2c1이 전달된다. 어댑터 단백질은 하나 이상의 활성인자 또는 하나 이상의 억제인자에 결합될 수 있다(바람직하게는 이에 연결되거나 또는 융합될 수 있다). 예를 들어, 어댑터 단백질은 제1 활성인자 및 제2 활성인자와 결합될 수 있다. 제1 활성인자 및 제2 활성인자는 동일할 수 있지만, 그들은 바람직하게는 상이한 활성인자

이다. 예를 들어, 이들이 단지 예시적이고, 다른 전사 활성인자가 예상되지만, 하나는 VP64일 수 있는 반면, 다른 것은 p65일 수 있다. 3가지 이상 또는 심지어 4가지 이상의 활성인자(또는 억제인자)가 사용될 수 있지만, 패키지 크기는 숫자를 5개의 상이한 기능성 도메인보다 더 크게 제한할 수 있다. 링커는 바람직하게는 어댑터 단백질에 대한 직접 융합 이상으로 사용되며, 여기서 2 이상의 기능성 도메인은 어댑터 단백질과 결합된다. 적합한 링커는 GlySer 링커를 포함할 수 있다.

[0138] 또한 전체로서 효소-가이드 복합체는 2 이상의 기능성 도메인과 결합될 수 있다는 것이 예상된다. 예를 들어, 효소와 결합된 2 이상의 기능성 도메인이 있을 수 있거나 또는 (하나 이상의 어댑터 단백질을 통해) 가이드와 결합된 2 이상의 기능성 도메인이 있을 수 있거나, 또는 효소와 결합된 하나 이상의 기능성 도메인 및 (하나 이상의 어댑터 단백질을 통해) 가이드와 결합된 하나 이상의 기능성 도메인이 있을 수 있다.

[0139] 어댑터 단백질과 활성인자 또는 억제인자 사이의 융합은 링커를 포함할 수 있다. 예를 들어, GlySer 링커 GGGS가 사용될 수 있다. 이들은 3($(GGGGS)_3$) 또는 6, 9 또는 심지어는 12 이상의 반복부로 사용되어, 필요에 따라 적합한 길이를 제공할 수 있다. 링커는 RNA-결합 단백질과 기능성 도메인(활성인자 또는 억제인자) 사이에, 또는 CRISPR 효소(C2c1)와 기능성 도메인(활성인자 또는 억제인자) 사이에 사용될 수 있다. 사용자는 링커를 "기계적 유연성"의 적절한 양으로 조작한다.

[0140] ***에스코트 & 유도성 가이드***

[0141] 바람직한 구현예에서, 직접 반복부는 변형되어 하나 이상의 단백질-결합 RNA 압타머를 포함할 수 있다. 특정한 구현예에서, 하나 이상의 압타머는 예컨대 최적 2차 구조의 일부분을 포함할 수 있다. 이러한 압타머는 본 명세서에서 더욱 상술되는 바와 같은 박테리오파지 외피 단백질에 결합할 수도 있다.

[0142] 특정 구현예에서, 가이드는 에스코트 가이드이다. "에스코트"란 Cas12b CRISPR-Cas 시스템 또는 복합체 또는 가이드가 세포 내에서 선택된 시간 또는 위치로 전달되어서, Cas12b CRISPR-Cas 시스템 또는 복합체 또는 가이드의 활성이 공간적으로 또는 시간적으로 제어되는 것을 의미한다. 예를 들어, Cas12b CRISPR-Cas 시스템 또는 복합체 또는 가이드의 활성 및 목적은 압타머 리간드, 예컨대 세포 표면 단백질 또는 다른 국재화된 세포 성분에 대해 결합 친화성을 갖는 에스코트 RNA 압타머 서열에 의해 제어될 수 있다. 대안적으로, 에스코트 압타머는 예를 들어 세포 상 또는 세포 내 압타머 이펙터, 예컨대 일시적 이펙터, 예컨대 특정한 시기에 세포에 적용되는 외부 에너지원에 반응성일 수 있다.

[0143] 에스코트된 Cas12b CRISPR-Cas 시스템 또는 복합체는 가이드 분자 구조, 아키텍처, 안정성, 유전자 발현, 또는 이의 임의 조합을 개선시키도록 디자인된 기능성 구조를 갖는 가이드 분자를 갖는다. 이러한 구조는 압타머를 포함할 수 있다.

[0144] 압타머는 예를 들어 기하급수적 농축을 통한 리간드의 체계적 진화 (systematic evolution of ligands by exponential enrichment)라고 불리는 기술을 사용하여, 다른 리간드에 단단하게 결합되도록 디자인되거나 또는 선택된 생물분자이다 (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science 1990, 249:505- 510). 핵산 압타머는 예를 들어 생물의학적으로 관련된 광범위한 표적에 대해 높은 결합 친화성 및 특이성을 갖는, 무작위-서열 올리고뉴클레오티드의 풀로부터 선택할 수 있어서, 압타머에 대한 광범위한 치료적 활용성을 시사한다 (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." Nature Reviews Drug Discovery 9.7 (2010): 537- 550). 이들 특징은 또한 약물 전달 비히클로서 압타머에 대한 광범위한 용도를 시사한다 (Levy-Nissenbaum, Etgar, et al. "Nanotechnology and aptamers: applications in drug delivery." Trends in biotechnology 26.8 (2008): 442-449; and, Hicke BJ, Stephens AW. "Escort aptamers: a delivery service for diagnosis and therapy." J Clin Invest 2000, 106:923-928.). 압타머는 또한 녹색 형광 단백질의 활성을 모방하도록 형광단에 결합하는 RNA 압타머와 같이, 속성을 변화시켜 큐 (que)에 반응하는, 분자 스위치로서 기능하게 구축될 수 있다 (Paige, Jeremy S., Karen Y. Wu, and Samie R. Jaffrey. "RNA mimics of green fluorescent protein." Science 333.6042 (2011): 642- 646). 압타머는 예를 들어 세포 표면 단백질을 표적화하는, 표적화된 siRNA 치료제 전달 시스템의 성분으로서 사용될 수 있다는 것이 또한 제안되었다 (Zhou, Jiehua, and John J. Rossi. "Aptamer-targeted cell-specific RNA interference." Silence 1.1 (2010): 4).

[0145] 따라서, 특정 구현예에서, 가이드 분자는 예를 들어 세포막을 통해서, 세포내 구획, 또는 핵으로의 전달을 포함하여, 가이드 분자 전달을 개선시키도록 디자인된 하나 이상의 압타머(들)를 통해 변형된다. 이러한 구조는

가이드 분자를 선택된 이펙터에 대해 전달가능하거나, 유도가능하거나 또는 반응성이도록 만들기 위해서, 하나 이상의 압타머(들)의 첨가 또는 이러한 하나 이상의 압타머(들), 모이어티(들)없이 포함될 수 있다. 따라서 본 발명은 제한없이 pH, 저산소, O_2 농도, 온도, 단백질 농도, 효소 농도, 지질 구조, 광노출, 기계적 파괴 (예를 들어, 초음파), 자기장, 전기장 또는 전자기 방사선을 포함한 정상 또는 병적 생리학적 조건에 반응하는 가이드 분자를 이해한다.

[0146] 유도성 시스템의 광 반응성은 크립토크롬-2 및 CIB1의 활성화 및 결합을 통해 획득될 수 있다. 파란색 빛 자극은 크립토크롬-2의 활성화 입체형태 변화를 유도하여서, 이의 결합 파트너 CIB1의 동원을 야기시킨다. 이러한 결합은 빠르고 가역적이어서, 펄스된 자극 후에 <15초의 포화를 획득하고 자극 종료 후 <15분에 기준치로 복귀된다. 이들 신속한 결합 동역학은 유도제의 흡수 및 청소보다는, 전사/번역 및 전사물/단백질 분해의 속도에 의해서만 시간적으로 제한되는 시스템을 야기시킨다. 크립토크롬-2 활성화는 또한 고도로 민감하여서, 낮은 및 강도 자극의 사용을 허용하고 광독성 위험성을 완화시킨다. 또한, 예컨대 온전한 포유동물 뇌의 경우에, 다양한 빛 광도를 사용하여 자극 영역의 크기를 제어하여서, 벡터 전달 단독으로 제공될 수 있는 것보다 더 큰 정밀도를 허용한다.

[0147] 본 발명은 가이드를 유도시키기 위해서 에너지원, 예컨대 전자기 방사선, 소리 에너지 또는 열 에너지를 고려한다. 유리하게, 전자기 방사선은 가시광선의 성분이다. 바람직한 구현예에서, 빛은 약 450 내지 약 495 nm 파장의 파란색 빛이다. 특히 바람직한 구현예에서, 파장은 약 488 nm이다. 다른 바람직한 구현예에서, 광 자극은 펄스를 통한다. 광력은 약 0-9 mW/cm^2의 범위일 수 있다. 바람직한 구현예에서, 15초마다 0.25초 만큼 낮은 자극 파라다임이 최대 활성화를 야기시켜야 한다.

[0148] 화학적 또는 에너지 민감성 가이드는 가이드로서 작용하고 C2c1 CRISPR-Cas 시스템 또는 복합체 기능을 갖도록 허용하는 에너지 또는 화학적 소스에 의한 결합을 통한 유도 시에 입체형태적 변화를 겪을 수 있다. 본 발명은 가이드 기능 및 C2c1 CRISPR-Cas 시스템 또는 복합체 기능을 갖도록 화학적 소스 또는 에너지를 인가하는 단계; 및 임의로는 게놈 유전자좌의 발현이 변경된다는 것을 결정하는 추가 단계를 포함할 수 있다.

[0149] 이 화학적 유도성 시스템의 몇몇 상이한 설계가 있다: 1. 아브시스산(ABA)에 의해 유도가능한 ABI-PYL 기반 시스템(예를 들어, stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2 참조), 2. 라파마이신에 의해 유도가능한 FKBP-FRB 기반 시스템 (또는 라파마이신을 기반으로 한 관련 화학 물질)(예를 들어, 웹사이트 nature.com/nmeth/journal/v2/n6/full/nmeth763.html 참조), 3. 지베렐린(GA)에 의해 유도가능한 GID1-GAI 기반 시스템(예를 들어, www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html 참조).

[0150] 화학적 유도성 시스템은 4-히드록시타목시펜 (4OHT) (예를 들어,www.pnas.org/content/104/3/1027.abstract 참조)에 의해 유도가능한 에스트로겐 수용체 (ER) 기반 시스템일 수 있다. ERT2라고 하는 에스트로겐 수용체의 돌연변이된 리간드-결합 도메인은 4-히드록시타목시펜의 결합 시 세포의 핵으로 전위된다. 본 발명의 추가 구현예에서, 임의의 핵 수용체, 갑상선 호르몬 수용체, 레티노산 수용체, 에스트로겐 수용체, 에스트로겐-관련 수용체, 글루코코티코이드 수용체, 프로게스테론 수용체, 안드로겐 수용체의 임의의 천연 유래 또는 조작된 유도체는 ER 기반 유도성 시스템에 유사한 유도성 시스템에서 사용될 수 있다.

[0151] 다른 유도성 시스템은 에너지, 열 또는 전파에 의해 유도성인 일시적 수용체 전위(일시적 수용체 potential:TRP) 이온 통로 기반 시스템을 이용하는 설계에 기반한다(예를 들어, www.sciencemag.org/content/336/6081/604). 이들 TRP 패밀리 단백질은 빛과 열을 포함한, 상이한 자극에 반응한다. 이러한 단백질은 빛 또는 열에 의해 활성화될 때, 이온 채널이 열리게 되어 칼슘 이온과 같은 이온의 형질막으로의 진입을 허용한다. 이러한 이온 유입은 가이드 및 C2c1 CRISPR-Cas 복합체 또는 시스템의 다른 성분을 포함한 폴리펩티드에 연결된 세포내 이온 상호작용 파트너에 결합하게 될 것이고 이러한 결합은 폴리펩티드의 세포하 국재화의 변화를 유도하여, 전체 폴리펩티드가 세포의 핵으로 진입하게 한다. 핵 내부에 존재하게 되면, 가이드 단백질 및 C2c1 CRISPR-Cas 복합체의 다른 성분이 활성화되어 세포에서 표적 유전자 발현을 조절하게 될 것이다.

[0152] 빛 활성화가 유리한 구현예일 수 있지만, 때때로 빛이 피부 또는 다른 장기를 침투할 수 없는 생체내 적용에서는 특히 불리할 수 있다. 이러한 예에서, 에너지 활성화의 다른 방법, 특히 유사한 효과를 갖는 전기장 에너지 및/또는 초음파가 고려된다.

[0153] 전기장 에너지는 바람직하게는 생체내 조건 하에서 약 1 Volt/cm 내지 약 10 kVolts/cm의 하나 이상의 전기 펄스를 사용하여, 당분야에 기술된 바와 같이, 실질적으로 투여된다. 펄스 대신에 또는 그이외에도, 전기장은

연속적인 방식으로 전달될 수 있다. 전기 펄스는 1 μs 내지 500 밀리초, 바람직하게 1 μs 내지 100 밀리초 동안 인가될 수 있다. 전기장은 약 5분 동안 펄스식 방식으로 또는 연속적으로 인가될 수 있다.

[0154] 본 명세서에서 사용되는 "전기장 에너지"는 세포가 노출되는 전기 에너지이다. 바람직하게, 전기장은 생체내 조건 하에서 약 1 Volt/cm 내지 약 10 kVolts/cm 이상의 강도를 갖는다 (WO97/49450 참조).

[0155] 본 명세서에서 사용되는 용어 "전기장"은 가변 용량 및 전압에서 하나 이상의 펄스를 포함하고, 지수파 및/또는 사각파 및/또는 변조파 및/또는 변조 사각파 형태를 포함한다. 전기장 및 전기장은 세포 환경에서 전기적 전위차의 존재에 대한 참조를 포함하도록 언급되어야 한다. 이러한 환경은 당업계에 공지된 바와 같은 정전기, 교류 (AC), 직류 (DC)에 의해 셋업될 수 있다. 전기장은 균일, 비균일 등일 수 있으며, 시간 의존적 방식으로 강도 및/또는 방향이 다를 수 있다.

[0156] 전기장의 단회 또는 다회 인가뿐만 아니라 초음파의 단회 또는 다회 인가는 임의의 순서로 그리고 임의의 조합으로 가능하다. 초음파 및/또는 전기장은 단회 또는 다회 연속 인가로서, 또는 펄스 (박동성 전달)로서 전달될 수 있다.

[0157] 외래 물질을 살아있는 세포 내로 도입하기 위해 시험관내 절차와 생체내 절차 둘 모두에서 전기천공법이 사용되어 왔다. 시험관내 적용에 의해, 생존 세포의 샘플은 먼저 관심의 작용제와 혼합되고, 전극, 예컨대, 평행판 사이에 위치된다. 이어서, 전극은 전기장을 세포/이식 혼합물에 인가된다. 시험관내 전기천공법을 수행하는 시스템의 예는 Electro Cell Manipulator ECM600 제품, 및 Electro Square Porator T820를 포함하며, 이들은 Genetronics, Inc의 BTX Division에서 제조되었다 (미국 특허 제5,869,326호 참조).

[0158] 공지된 전기천공법 기법은 (시험관내 및 생체내) 처리 영역 주변에 위치된 전극에 짧은 고전압 펄스를 인가함으로써 작용한다. 전극 사이에서 발생된 전기장은 세포막을 일시적으로 다공성이 되게 하여서, 그때에 관심 작용제의 분자가 세포로 진입한다. 공지된 전기천공법 적용에서, 전기장은 약 100 μs 지속기간의, 1000V/cm 규모로 단일 사각파 펄스를 포함한다. 이러한 펄스는, 예를 들어, 일렉트로 스퀘어 포레이터 T820의 공지된 적용에서 생성될 수 있다.

[0159] 바람직하게는, 전기장은 시험관내 조건 하에서 약 1V/cm 내지 약 10kV/cm의 강도를 가진다. 따라서, 전계는 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100 V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm 이상의 강도를 가질 수 있다. 보다 바람직하게 시험관내 조건 하에서 약 0.5 kV/cm 내지 약 4.0 kV/cm. 바람직하게는 전기장은 생체내 조건 하에서 약 1 V/cm 내지 약 10 kV/cm의 강도를 갖는다. 그러나, 표적 부위에 전달된 펄스 수가 증가되는 경우에 전기장 강도는 낮아질 수 있다. 따라서, 더 낮은 전기장 강도에서 전기장의 박동성 전달이 계획된다.

[0160] 바람직하게는 전기장의 인가는 동일한 강도 및 용량의 이중 펄스 또는 다양한 강도 및/또는 용량의 순차적 펄스와 같은 다중 펄스 형태이다. 본 명세서에서 사용되는 용어 "펄스"는 가변 용량 및 전압에서 하나 이상의 전기적 펄스를 포함하고, 지수 및/또는 사각파 및/또는 변조파/사각파 형태를 포함한다.

[0161] 바람직하게는 전기적 펄스는 지수 파형, 사각 파형, 변조 파형으로부터 선택된 파형으로서 전달된다.

[0162] 바람직한 구현예는 저전압에서 직류를 사용한다. 따라서, 출원인은 1V/cm 내지 20V/cm의 전기장 강도에서, 100 밀리초 이상, 바람직하게는 15분 이상의 기간 동안 세포, 조직 또는 조직 덩어리에 인가되는 전기장의 사용을 개시한다.

[0163] 초음파는 유리하게는 약 0.05W/㎠ 내지 약 100W/㎠의 전력 수준으로 투여된다. 진단 또는 치료 초음파, 또는 이들의 조합이 사용될 수 있다.

[0164] 본 명세서에서 사용되는 용어 "초음파"는 기계적 진동으로 이루어지고, 이의 주파수는 너무 높아서 인간 청력 범위 이상인 에너지의 형태를 의미한다. 초음파 스펙트럼의 하한 주파수는 일반적으로 약 20 kHz로서 취해질 수 있다. 초음파의 대부분의 진단적 적용은 1 내지 15 ㎒' 범위의 주파수를 사용한다 (Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[0165] 초음파는 진단 및 치료 적용분야 둘 모두에서 사용되었다. 진단 도구로서 사용될 때 ("진단적 초음파"), 최대 750 mW/㎠의 에너지 밀도가 사용되었지만, 초음파는 전형적으로 최대 약 100 mW/㎠ (FDA 권장사항)의 에너지 밀도 범위에서 사용된다. 물리치료에서, 초음파는 전형적으로 약 3 내지 4 W/㎠ (WHO 권장사항) 범위에서

에너지 공급원으로서 사용된다. 다른 치료적 적용분야에서, 더 높은 강도의 초음파는 단기간 동안, 예를 들어, 100 W/㎝ 내지 1 kW/㎠ (또는 훨씬 높은)에서의 HIFU가 적용될 수 있다. 본 명세서에서 사용되는 용어 "초음파"는 진단, 치료 및 집속 초음파를 포함하는 것으로 의도된다.

[0166] 집속 초음파 (Focused ultrasound: FUS)는 비침습 프로브 없이 열 에너지를 전달할 수 있게 한다 ([Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142] 참조). 집속 초음파의 다른 형태는 [Moussatov et al in Ultrasonics (1998) Vol.36, No.8, pp.893-900] 및 [TranHuuHue et al in Acustica (1997) Vol.83, No.6, pp.1103-1106]에서 고찰된 고강도 집속 초음파 (high intensity focused ultrasound: HIFU)이다.

[0167] 바람직하게는, 진단 초음파와 치료 초음파의 조합이 사용된다. 그러나, 이 조합은 제한하려는 의도가 아니고, 당업자는 초음파의 임의의 다양한 조합이 사용될 수 있다는 것을 인식할 것이다. 추가적으로, 에너지 밀도, 초음파 주파수 및 노출 시간은 변할 수 있다.

[0168] 바람직하게는 초음파 에너지원에 대한 노출은 약 0.05 내지 약 100Wcm^{-2}의 전력 밀도에서이다. 훨씬 더 바람직하게는, 초음파 에너지원에 대한 노출은 약 1 내지 약 15Wcm^{-2}의 전력 밀도에서이다.

[0169] 바람직하게는 초음파 에너지원에 대한 노출은 약 0.015 내지 약 10.0 ㎒의 주파수에서이다. 더 바람직하게는 초음파 에너지원에 대한 노출은 약 0.02 내지 약 5.0 ㎒ 또는 약 6.0 ㎒의 주파수에서이다. 가장 바람직하게는, 초음파는 3 ㎒의 주파수에서 인가된다.

[0170] 바람직하게는 노출은 약 10 밀리초 내지 약 60분의 기간 동안이다. 바람직하게는 노출은 약 1초 내지 약 5분의 기간 동안이다. 더 바람직하게는, 초음파는 약 2분 동안 인가된다. 그러나, 붕괴하려는 특정 표적 세포에 따라서, 노출은 더 긴 지속기간, 예를 들어, 15분 동안일 수 있다.

[0171] 유리하게, 표적 조직이 약 0.0 5Wcm^{-2} 내지 약 10 Wcm^{-2}의 음향 출력 밀도에서 약 0.015 내지 약 10㎒ 범위의 주파수로 초음파 에너지원에 노출된다 (WO 98/52609 참조). 그러나, 대안, 예를 들어, 100Wcm^{-2} 초과의 음향 출력 밀도에서, 그러나 감소된 시간 기간 동안, 예를 들어, 밀리초 범위 이하의 기간 동안 1000Wcm^{-2}에서 초음파 에너지에 대한 노출이 또한 가능하다.

[0172] 바람직하게는 초음파의 인가는 다중 펄스 형태이고, 따라서, 지속파와 펄스파 (초음파 맥동성 전달)는 둘 모두 임의의 조합으로 사용될 수 있다. 예를 들어, 지속파 초음파, 다음에 펄스파 초음파 또는 그 반대로 적용될 수 있다. 이는 임의의 횟수, 임의의 순서 및 조합으로 반복될 수 있다. 펄스파 초음파는 지속파 초음파의 배경에 대해 적용될 수 있고, 임의의 펄스 수가 임의의 그룹 수에서 사용될 수 있다.

[0173] 바람직하게는, 초음파는 펄스파 초음파를 포함할 수 있다. 고도로 바람직한 구현예에서, 초음파는 지속파로서 0.7Wcm^{-2} 또는 1.25Wcm^{-2}의 출력 밀도로 적용된다. 펄스파 초음파가 사용된다면, 더 높은 출력 밀도가 사용될 수 있다.

[0174] 초음파의 사용은 표적 상에서 정확하게 집속시킬 수 있기 때문에 빛만큼 유리하다. 게다가, 초음파는 빛과 달리 조직에 더 깊게 집속시킬 수 있기 때문에 유리하다. 따라서 전체 조직 침투 (예컨대, 이에 제한없이, 간엽) 또는 전체 장기 (예컨대, 제한없이, 전체 간 또는 전체 근육, 예컨대, 심장) 요법에 더 적합하게 된다.
다른 중요한 이점은 초음파가 매우 다양한 진단적 및 치료적 적용분야에서 사용되는 비침습성 자극이라는 것이다. 예로서, 초음파는 의학적 영상화 기법에서, 추가적으로 정형외과적 요법에서 잘 공지되어 있다. 더 나아가, 대상 척추동물에 대한 초음파의 적용에 적합한 기기가 널리 이용 가능하며, 그들의 용도는 당업계에 충분히 공지되어 있다.

[0175] 본 발명의 빠른 전사 반응 및 내생성 표적화는 전사 역학 연구를 위한 이상적인 시스템에 도움이 된다. 예를 들어, 본 발명은 표적 유전자의 유도된 발현 시 변이체 생성 역학을 연구하는 데 사용될 수 있다. 전사 주기의 다른 마지막에, mRNA 분해 연구는 종종 과잉 유전자의 발현 수준 변화를 야기하는 강한 세포외 자극에 반응하여 수행된다. 본 발명은 내생성 표적의 전사를 역으로 유도하는 데 이용될 수 있으며, 이 시점 후에 자극은 중단될 수 있고, 독특한 표적의 분해 역학이 추적될 수 있다.

[0176] 본 발명의 일시적 정확성은 실험적 개입과 관련하여 시간 유전적 조절에 대한 검증력을 제공할 수 있다. 예를 들어, 장기 상승 작용(long-term potentiation:LTP)에 연루되는 것으로 의심되는 표적은 세포의 정상 발생

방해를 피하기 위해, 기관형(organotypic) 또는 해리된 뉴런 배양물에서, 그러나 LTP를 유도하기 위한 자극 동안에만 조절될 수 있다. 유사하게, 질환 표현형을 나타내는 세포 모델에서, 특정 요법의 유효성에 연루되는 것으로 의심되는 표적은 치료 동안에만 조절될 수 있다. 대조적으로, 유전자 표적은 병원성 자극 동안에만 조절될 수 있다. 외부 실험 자극에 대한 유전적 신호 시기가 적절성을 갖는 다수의 실험은 본 발명의 효용이 잠재적으로 유리할 수 있다.

[0177] 생체내 상황은 유전자 발현을 제어하기 위해 본 발명에 대한 동등하게 풍부한 기회를 제공한다. 광유도 능력은 공간적 정확성에 대한 잠재력을 제공한다. 광단자 기술 개발을 기회로 하여, 자극성 광섬유 리드는 뇌 영역에 위치될 수 있다. 이어서, 자극 영역 크기는 광 강도에 의해 조율될 수 있다. 이는 본 발명의 C2c1 CRISPR-Cas 시스템 또는 복합체의 전달과 함께 행해질 수 있거나, 또는 유전자이식 C2c1 동물의 경우에, 본 발명의 가이드 RNA는 전달될 수 있고, 광단자 기술은 정확한 뇌 영역에서 유전자 발현을 조절을 가능하게 할 수 있다. 투과적 C2c1 발현 유기체는 그에 투여되는 본 발명의 가이드 RNA를 가질 수 있고, 이어서, 극도로 정확한 레이저 유도된 국소 유전자 발현 변화가 있을 수 있다.

[0178] 숙주 세포의 배양을 위한 배양 배지는 조직 배양을 위해 일반적으로 사용되는 매질, 예컨대 특히 M199-얼(earle) 염기, 이글(Eagle) MEM(E-MEM), 둘베코 MEM(DMEM), SC-UCM102, UP-SFM(GIBCO BRL), EXCELL302(Nichirei), EX-CELL293-S(Nichirei), TFBM-01(Nichirei), ASF104를 포함한다. 특정 세포 유형에 대한 적합한 배양 배지는 미국 미생물 보존센터(ATCC) 또는 유럽 세포 배양물 보존센터(European Collection of Cell Cultures:ECACC)에서 찾을 수 있다. 배양 배지는 아미노산, 예컨대 L-글루타민, 염, 항진균제 또는 항균제 예컨대 Fungizone®, 페니실린-스트렙토마이신, 동물 혈청 등으로 보충될 수 있다. 세포 배양 배지는 선택적으로 무혈청일 수 있다.

[0179] 본 발명은 또한 생체 내에서 소중한 시간적 정확도를 또한 제공할 수 있다. 본 발명은 특정 발생 단계 동안 유전자 발현을 변경시키기 위해 사용될 수 있다. 본 발명은 특정 실험창에 대한 유전적 신호의 시간 측정을 위해 사용될 수 있다. 예를 들어, 학습에 연루된 유전자는 무손상 설치류 또는 영장류 뇌의 정확한 영역에서 학습 자극 동안에만 과발현되거나 또는 억제될 수 있다. 추가로, 본 발명은 질환 발생의 특정 단계 동안에만 유전자 발현 변화를 유도하는 데 사용될 수 있다. 예를 들어, 일단 종양이 특정 크기 또는 전이 단계에 도달된다면, 종양유전자는 단지 과발현될 수 있다. 역으로, 알츠하이머의 발생시 의심되는 단백질은 동물의 생애에서 규정된 시점에서, 특정 뇌 영역 내에서만 무너질 수 있다. 이들 예는 본 발명의 잠재적 적용을 철저하게 열거하지는 않지만, 그들은 본 발명이 강력한 기술일 수 있는 일부 영역을 강조한다.

[0180] ***보호된 가이드***

[0181] 특정 구현예에서, 가이드 분자는 CRISPR-Cas 시스템의 특이성을 증가시키기 위해 2차 구조에 의해 변형되고, 그 2차 구조는 엑소뉴클레아제 활성에 대해 보호될 수 있고, 본 명세서에서 또한 보호된 가이드 분자라고도 언급되는 가이드 서열에 대한 5' 첨가를 허용한다.

[0182] 일 양상에서, 본 발명은 가이드 분자의 서열과 "보호자 RNA" 를 하이브리드화하기 위해 제공되고, 여기서 "보호자 RNA" 는 가이드 분자의 3' 말단에 상보성이어서 부분적으로 이중-가닥인 가이드 RNA를 생성시키는 RNA 가닥이다. 본 발명의 일 구현예에서, 완벽하게 상보성인 보호자 서열로 미스매치된 염기 (즉, 가이드 서열의 일부를 형성하지 않는 가이드 분자의 염기)를 보호하는 것은 3' 말단에서 미스매치된 염기쌍과 표적 DNA 결합 가능성을 감소시킨다. 본 발명의 특정 구현예에서, 연장된 길이를 포함하는 추가 서열은 가이드가 가이드 분자 내에 보호자 서열을 포함하도록 가이드 분자 내에 존재할 수 있다. 이러한 "보호자 서열" 은 가이드 분자가 "노출된 서열" (표적 서열과 하이브리드화되는 가이드 서열의 일부 포함) 이외에도 "보호된 서열" 을 포함하는 것을 보장한다. 특정 구현예에서, 가이드 분자는 2차 구조 예컨대 헤어핀을 포함하도록 보호자 가이드의 존재에 의해 변형된다. 유리하게, 보호된 서열, 가이드 서열 또는 둘 모두에 대해 상보성을 갖는 3 또는 4 내지 30 이상, 예를 들어 약 10 이상의 인접한 염기쌍이 존재한다. 보호된 부분은 이의 표적과 상호작용하는 CRISPR-Cas 시스템의 열역학을 방해하지 않는 것이 유리하다. 부분적으로 이중 가닥인 가이드 분자를 포함한 이러한 연장부를 제공하여, 가이드 분자가 보호되는 것을 고려하고 그 결과로 비활성을 유지하면서, CRISPR-Cas 복합체의 개선된 비결합을 야기시킨다.

[0183] 게놈 표적에 매칭되는 가이드 RNA(gRNA) 연장은 gRNA 보호를 제공하고, 특이성을 향상시킨다. 개개 게놈 표적에 대한 스페이서 씨드의 말단부에 대해 원위인 매칭 서열에 의한 gRNA의 연장은 향상된 특이성을 제공하는 것으로 예상된다. 특이성을 향상시키는 매칭 gRNA 연장은 절단 없이 세포에서 관찰되었다. 이들 안정한 길이 연장을 수반하는 gRNA 구조의 예측은 안정한 형태가 보호 상태로부터 생긴다는 것을 나타내며, 여기서, 연

장은 스페이서 연장 및 스페이서 종자에서의 상보성 서열에 기인하여 gRNA 종자를 갖는 폐쇄 루프를 형성한다.

이들 결과는 보호된 가이드 개념이 또한 20량체 스페이서-결합 영역의 게놈 표적 서열 원위에 매칭되는 서열을 포함한다는 것을 입증한다. 열역학적 예측은 보호된 gRNA 상태를 초래하는 완전히 매칭되거나 또는 부분적으로 매칭되는 가이드 연장을 예측하기 위해 사용될 수 있다. 이는 보호된 gRNA의 개념을 X와 Z 사이의 상호작용까지 연장시키며, 여기서 X는 일반적으로 길이가 17 내지 20nt이고, Z는 길이가 1 내지 30nt일 것이다.

열역학적 예측은 X 및 Z 간 보호된 입체형태의 형성을 촉진하기 위해서 Z에 적은 수의 미스매치를 잠재적으로 도입하는, Z에 대한 최적 연장 상태를 결정하는데 사용될 수 있다. 본 출원 전반에서, 용어 " X" 및 씨드 길이 (SL)는 결합하는 표적 DNA에 대해 이용가능한 뉴클레오티드의 수를 의미하는 용어 노출된 길이 (EpL)과 상호교환적으로 사용되고; 용어 " Y" 및 보포자 길이 (PL)은 보호자의 길이를 나타내기 위해 상호교환적으로 사용되며; 용어 "Z", "E", " E" 및 "EL" 은 표적 서열이 연장된 뉴클레오티드의 수를 나타내는 용어 연장된 길이 (ExL)에 상응하게 상호교환적으로 사용된다.

[0184] 연장된 길이 (ExL)에 상응하는 연장 서열은 보호된 가이드 서열의 3' 말단에서 가이드 서열에 선택적으로 직접 부착될 수 있다. 연장 서열은 길이가 2 내지 12개의 뉴클레오티드일 수 있다. 바람직하게는 ExL은 길이가 0, 2, 4, 6, 8, 10 또는 12개의 뉴클레오티드로서 나타낼 수 있다. 바람직한 구현예에서, ExL은 길이가 0 또는 4개의 뉴클레오티드로서 나타낼 수 있다. 더 바람직한 구현예에서, ExL은 길이가 4개의 뉴클레오티드로서 나타낼 수 있다. 연장 서열은 표적 서열에 대해 상보성일 수도 있고 상보성이 아닐 수도 있다.

[0185] 연장 서열은 보호된 가이드 서열의 5' 말단에서 가이드 서열에 또한 및 보호 서열의 3' 말단에 선택적으로 추가적으로 직접 부착될 수 있다. 그 결과, 연장 서열은 보호된 서열과 보호 서열 사이의 연결 서열로서 작용한다. 이론에 의해 구속되는 일 없이, 이러한 연결은 보호되니 서열에 대한 보호 서열의 개선된 결합을 위해 보호된 서열 근처에 보호 서열을 위치시킬 수 있다. 씨드, 보호자 및 연장의 상기 기재된 관계는 가이드의 원위 단부(즉, 표적화 단부)가 5' 말단부이고, 예를 들어, 작용하는 가이드가 Cas 시스템인 경우에 적용된다는 것이 이해될 것이다. 구현예에서, 가이드의 원위 단부는 3' 말단부이고, 관계는 반전될 것이다. 이러한 구현예에서, 본 발명은 가이드 서열에 "보호자 RNA" 를 하이브리드화시키도록 제공함으로써, 부분적으로 이중-가닥 gRNA를 생성하되, "보호자 RNA" 는 가이드 RNA(gRNA)의 3' 말단부에 상보성인 RNA 가닥이다.

[0186] gRNA의 원위 단부에 대한 gRNA 미스매치의 첨가는 향상된 특이성을 입증할 수 있다. Y에서 보호되지 않은 원위 미스매치의 도입 또는 원위 미스매치(Z)에 의한 gRNA의 연장은 향상된 특이성을 입증할 수 있다. 언급한 바와 같은 이런 개념은 보호된 gRNA에서 사용되는 X, Y 및 Z 성분에 묶여 있다. 보호되지 않은 미스매치 개념은 보호된 가이드 RNA에 대해 기재된 X, Y 및 Z의 개념에 대해 추가로 일반화될 수 있다.

[0187] *tru-가이드*

[0188] 특정 구현예에서, 절단된 가이드 (tru-가이드), 즉 정규 가이드 서열 길이에 비해서 길이가 절단된 가이드 서열을 포함하는 가이드 분자가 이용된다. Nowak 등 (Nucliec Acids Res (2016) 44 (20): 9555- 9564)이 기술한 바와 같이, 이러한 가이드는 표적 DNA를 절단하지 않고 촉매적 활성 CRISPR-Cas 효소가 이의 표적에 결합할 수 있게 한다. 특정 구현예에서, 절단된 가이드는 CRISPR-Cas 효소의 닉카제 활성만을 보유하면서 표적의 결합을 허용하는 것이 사용된다.

[0189] 특정 구현예에서 가이드 분자는 직접 반복부 서열에 연결된 가이드 서열, 또는 직접 반복부 서열 및 tracr 서열에 연결된 가이드 서열을 포함하고, 여기서 직접 반복부 서열, crRNA, 및/또는 tracr 서열은 하나 이상의 스템 루프 또는 최적화된 2차 구조를 포함한다. 특정 구현예에서, 직접 반복부는 16 nt의 최소 길이 및 단일 스템 루프를 갖는다. 추가 구현예에서, 직접 반복부는 16 nt 초과, 바람직하게 17 nt 초과의 길이를 가지고, 하나 초과의 스템 루프 또는 최적화된 2차 구조를 갖는다. 특정 구현예에서 가이드 분자는 천연 직접 반복부 서열의 전체 또는 일부에 연결된 가이드 서열을 포함하거나 또는 그로 이루어진다. 전형적인 V-B형 C2c1/Cas12b 가이드 분자는 (3'에서 5' 방향으로) 가이드 서열 및 tracr의 3' 말단에 상보성인, 상보성 스트레치 ("반복부")를 포함한다. 반복부 및 tracr은 제2 상보성 스트레치 반복부에 상보성인 tracr의 "안티-반복부") 및 폴리 A (RNA에서는 종종 폴리U) 꼬리 (종결자)를 포함한, 스템-루프 (루프는 전형적으로 4 또는 5 뉴클레오티드 길이)를 형성하도록 설계된 영역을 포함한 키메라 가이드로 결합될 수 있다. 특정 구현예에서, 가이드 아키텍처의 일정 양상은 예를 들어 특성의 첨가, 차감, 또는 치환을 통해 변형될 수 있는 반면, 가이드 아키텍처의 일정한 다른 양상은 유지된다. 제한없이 삽입, 결실, 및 치환을 포함하는, 조작된 가이드 분자 변형을 위한 바람직한 위치는 C2c1 단백질 및/또는 표적과 복합체를 형성했을 때, 예를 들어 직접 반복부 서열

의 스템루프를 노출시키는 가이드 분자의 영역 및 가이드 말단을 포함한다.

[0190] *키메라 가이드*

[0191] 본 발명은 다양한 Cas12b 시스템 가이드를 제공한다. 일정 구현예에서, 가이드는 2개의 하이브리드화가능한 부분을 포함하고, 제1 부분의 3' 말단은 제2 부분의 5' 말단에 적어도 부분적으로 상보성이고 그와 하이브리드화할 수 있다. 일정 구현예에서, 2개 부분은 결합된다. 즉, 단일 가이드 ("키메라 가이드")는 Cas12b tracr 서열에 상응하는 3' 말단에서 제2 절편에 결합된, 천연 Cas12b 가이드의 직접 반복부 및 가이드 서열에 상응하는 5' 말단에 제1 절편을 포함하는 것이 적용될 수 있다. 2개 절편은 제1 절편의 3' 말단의 상보성 서열 및 제2 절편의 5' 말단이 예를 들어 스템-루프 구조로 하이브리드화할 수 있도록 결합된다.

[0192] *데드 가이드*

[0193] 일 양상에서, 본 발명은 CRISPR 복합체의 형성 및 표적에 대한 성공적인 결합을 가능하게 하는 한편, 동시에 성공적인 뉴클레아제 활성을 가능하게 하지 않는 방식으로(즉, 뉴클레아제 활성 없이/indel 활성 없이) 변형된 가이드 서열을 제공한다. 설명을 위하여, 이러한 변형된 가이드 서열은 "데드 가이드" 또는 "데드 가이드 서열" 로 지칭된다. 이들 데드 가이드 또는 데드 가이드 서열은 뉴클레아제 활성에 관해 촉매적으로 비활성이거나 또는 입체배좌적으로 비활성인 것으로 생각될 수 있다. 뉴클레아제 활성은 당업계에서 통상적으로 사용되는 바와 같은 서베이어 분석 또는 심층 서열분석, 바람직하게는 서베이어 분석을 이용하여 측정될 수 있다. 유사하게는, 데드 가이드 서열은 촉매적 활성을 촉진시키거나 또는 온-표적 및 오프-표적 결합 활성을 구별하는 능력에 관해 생산적 염기 짝짓기에서 충분하게 관여하지 않을 수도 있다. 간략하게, 서베이어 분석은 유전자에 대해 CRISPR 표적 부위를 정제하고 증폭시키는 것 및 CRISPR 표적 부위를 증폭시키는 프라이머와의 헤테로듀플렉스을 형성하는 것을 수반한다. 재-어닐링 후에, 생성물을 제조처의 권고된 프로토콜에 따라 SURVEYOR 뉴클레아제 및 SURVEYOR 인핸서 S(Transgenomics)로 처리하고, 겔 상에서 분석하고, 상대적 밴드 세기에 기초하여 정량화시킨다.

[0194] 따라서, 관련된 양상에서, 본 발명은 본 명세서에 기재된 기능성 Cas12b 및 가이드 RNA(gRNA)를 포함하는 비천연 발생 또는 조작된 조성물 C2c1 CRISPR-Cas 시스템을 제공하되, gRNA는 데드 가이드 서열을 포함함으로써, Cas12b CRISPR-Cas 시스템이 서베이어 분석에 의해 검출된 바와 같은 시스템의 비돌연변이체 Cas12b 효소의 뉴클레아제 활성으로부터 초래된 검출 가능한 indel 활성 없이 세포에서 관심 대상의 게놈 유전자좌로 향하도록, gRNA는 표적 서열에 하이브리드화될 수 있다. 간략함의 목적을 위해, Cas12b CRISPR-Cas 시스템이 서베이어 분석에 의해 검출되는 바와 같은 시스템의 비돌연변이체 Cas12b 효소의 뉴클레아제 활성으로부터 초래되는 검출 가능한 indel 활성 없이 세포에서 관심 대상의 게놈 유전자좌로 향하도록, gRNA가 표적 서열에 하이브리드화할 수 있는, 데드 가이드 서열을 포함하는 gRNA는 "데드 gRNA" 로 지칭된다. 본 명세서의 다른 곳에 기재된 바와 같은 본 발명에 따른 임의의 gRNA가 본 명세서에서 이하에 기재되는 바와 같은 데드 gRNA/데드 가이드 서열을 포함하는 gRNA로서 사용될 있다는 것이 이해되어야 한다. 본 명세서의 다른 곳에 기재된 바와 같은 임의의 방법, 생성물, 조성물 및 용도는 이하에 추가로 상술하는 바와 같은 데드 gRNA/데드 가이드 서열을 포함하는 gRNA와 동일하게 적용 가능하다. 추가적인 가이드에 의해, 다음의 특정 양상 및 구현예가 제공된다.

[0195] 표적 서열에 대한 CRISPR 복합체의 서열-특이적 결합을 지시하는 데드 가이드 서열의 능력은 임의의 적합한 분석에 의해 평가될 수 있다. 예를 들어, 시험될 데드 가이드 서열을 포함하는, CRISPR 복합체를 형성하는 데 충분한 CRISPR 시스템의 성분은 대응하는 표적 서열을 갖는 숙주 세포에, 예컨대 CRISPR 서열 성분을 코딩하는 벡터에 의한 형질감염에 의해, 그 다음에 표적 서열 내의 우선적인 절단 평가, 예컨대 본 명세서에 기재된 바와 같은 서베이어 분석에 의해 제공될 수 있다. 유사하게는, 표적 폴리뉴클레오티드 서열의 절단은 시험될 데드 가이드 서열 및 시험 데드 가이드 서열과 상이한 대조군 가이드 서열을 포함하는, CRISPR 복합체의 성분인 표적 서열을 제공함으로써, 그리고 시험 가이드 서열 반응과 대조군 가이드 서열 반응 사이의 표적 서열에서 결합 또는 절단율을 비교함으로써, 시험관에서 평가될 수 있다. 다른 분석이 가능하며, 당업자에게 일어날 것이다. 데드 가이드 서열은 임의의 표적 서열을 표적화하도록 선택될 수 있다. 일부 구현예에서, 표적 서열은 세포 게놈 내의 서열이다.

[0196] 본 명세서에 추가로 설명하는 바와 같이, 몇몇 구조적 매개변수는 이러한 데드 가이드에서 적절한 프레임워크가 도달하는 것을 가능하게 한다. 데드 가이드 서열은 활성 Cas12b-특이적 indel 형성을 초래하는 각각의 가이드 서열보다 더 짧다. 데드 가이드는 동일한 Cas12b로 향하는 각각의 가이드보다 5%, 10%, 20%, 30%, 40%, 50% 더 짧아서 활성 Cas12b-특이적 indel 형성을 야기한다.

[0197] 하기에 설명되고, 해당 분야에 알려져 있는 바와 같이, gRNA - C2c1 특이성의 일 양태는 직접 반복부 서열이며, 이는 적절하게 이러한 가이드에 연결될 것이다. 특히, 이것은 직접 반복부 서열이 C2c1의 기원에 따라 설계되는 것을 암시한다. 따라서, 입증된 데드 가이드 서열에 이용가능한 구조적 데이타는 C2c1 특이적 균등물을 디자인하는데 사용될 수 있다. 예를 들어, 둘 이상의 C2c1 이펙터 단백질의 오솔로그 뉴클레아제 도메인 RuvC의 구조적 유사성을 사용하여 등가의 설계 데드 가이드를 전달할 수 있다. 따라서, 본원의 데드 가이드는 길이 및 서열을 적절하게 변형시켜, 이러한 C2c1 특이적 등가물을 반영하여, CRISPR 복합체의 형성 및 표적으로의 성공적인 결합을 가능하게 함과 동시에, 성공적인 뉴클레아제 활성을 허용하지 않을 수 있다.

[0198] 본 명세서뿐만 아니라 당업계의 언급과 관련하여 데드 가이드의 사용은 시험관내, 생체외 및 생체내 적용에서 네트워크 생물학 및/또는 시스템 생물학을 위한 놀랍고도 예상되지 않은 플랫폼을 제공하여, 다중복합 유전자 표적화, 및 특히 양방향 다중복합 유전자 표적화를 가능하게 한다. 데드 가이드의 사용 전에, 다중 표적의 처리, 예를 들어, 유전자 활성의 활성화, 억제 및/또는 침묵은 도전되어 왔으며, 일부 경우에 가능하지 않다.
데드 가이드의 사용에 의해, 다중 표적 및 그에 따른 다중 활성은, 예를 들어, 동일한 세포에서, 동일한 동물에서, 또는 동일한 환자에서 처리될 수 있다. 이러한 다중복합은 동시에 일어나거나 또는 목적하는 시간틀 동안 시차를 둘 수 있다.

[0199] 예를 들어, 데드 가이드는 이제 뉴클레아제 활성 결과 없이 유전자 표적화를 위한 수단으로서 gRNA를 사용하는 첫 시간을 가능하게 하는 한편, 동시에 활성화 또는 억제를 위해 지시 수단을 제공한다. 데드 가이드를 포함하는 가이드 RNA는 유전자 활성의 활성화 또는 억제를 허용하는 방식으로 요소, 특히, 본원의 다른 곳에 기재된 바와 같은 단백질 어댑터(예를 들어, 압타머)를 추가로 포함하여, 유전자 이펙터(예를 들어, 유전자 활성의 활성인자 또는 억제인자)의 기능적 배치를 허용하도록 변형될 수 있다. 일 예는 본 명세서에 설명되는 바와 같이 그리고 당업계의 상태에서, 압타머의 혼입이다. 데드 가이드를 포함하는 gRNA를 조작하여, 단백질-상호작용 압타머(문헌[Konermann et al., "Genome-scale transcription 활성화by an engineered CRISPR-Cas9 complex," doi:10.1038/nature14136, incorporated herein by reference])를 혼입시킴으로써, 다수의 별개의 이펙터 도메인으로 이루어진 합성 전사 활성화 복합체를 조립할 수 있다. 이렇게 해서 천연 전사 활성화 과정 후에 모델링될 수 있다. 예를 들어, 이펙터 (예를 들어, 활성인자 또는 억제인자; 활성인자 또는 억제인자와의 융합 단백질로서 이량체화된 MS2 박테리오파지 코트 단백질)에 선택적으로 결합하는 압타머, 또는 그 자체가 이펙터 (예를 들어, 활성인자 또는 억제인자)에 결합하는 단백질은 데드 gRNA 테트라루프 및/또는 스템-루프 2에 첨부될 수 있다. MS2의 경우에, 융합 단백질 MS2-VP64는 테트라루프 및/또는 줄기-루프 2에 결합하고, 결국, 예를 들어, Neurog2에 대해 전사 상향조절을 매개한다. 다른 전사 활성인자는 예를 들어, VP64. P65, HSF1 및 MyoD1이다. 단지 이 개념의 예로서, MS2 줄기-루프의 PP7-상호작용 줄기-루프로의 대체는 억제성 요소를 보충하는 데 사용될 수 있다.

[0200] 따라서, 일 양상은 데드 가이드를 포함하는 본 발명의 gRNA이되, gRNA는 본 명세서에 기재된 바와 같이, 유전자 활성화 또는 억제를 제공하는 변형을 추가로 포함한다. 데드 gRNA는 하나 이상의 압타머를 포함할 수 있다.
압타머는 유전자 이펙터, 유전자 활성인자 또는 유전자 억제인자에 특이적일 수 있다. 대안적으로, 압타머는 단백질에 특이적일 수 있으며, 이는 차례로 특정 유전자 이펙터, 유전자 활성 인자 또는 유전자 억제인자에 특이적이고, 이를 동원/이에 결합한다. 활성인자 또는 억제인자 보충을 위한 다중 부위가 있다면, 부위는 활성인자 또는 억제인자 중 하나에 특이적인 것이 바람직하다. 활성인자 또는 억제인자 결합에 대한 다중 부위가 있다면, 부위는 동일 활성인자 또는 동일 억제인자에 특이적일 수 있다. 부위는 또한 상이한 활성인자 또는 상이한 억제인자에 특이적일 수 있다. 유전자 이펙터, 유전자 활성인자, 유전자 억제인자는 융합 단백질의 형태로 존재할 수 있다.

[0201] 구현예에서, 본 명세서에 기재된 바와 같은 데드 gRNA 또는 본 명세서에 기재된 바와 같은 C2c1 CRISPR-Cas 복합체는 2 이상의 어댑터 단백질을 포함하는 천연 유래 또는 조작된 조성물을 포함하되, 각각의 단백질은 하나 이상의 기능성 도메인과 결합되고, 어댑터 단백질은 데드 gRNA의 적어도 하나의 루프에 삽입된 별개의 RNA 서열(들)에 결합한다.

[0202] 그리하여, 일 양상은 세포에서 관심 게놈 유전자좌의 표적 서열에 하이브리드화할 수 있는 데드 가이드 서열을 포함하는 가이드 RNA (gRNA)를 포함하는 비천연 발생 또는 조작된 조성물을 제공하고, 여기서 데드 가이드 서열은 본 명세서에 정의된 바와 같이, 적어도 하나 이상의 핵 국재화 서열을 포함하는 C2c1이고, C2c1은 데드 gRNA의 적어도 하나의 루프가 하나 이상의 어댑터 단백질에 결합하는 별개 RNA 서열(들)의 삽입에 의해 변형되는 것인 적어도 하나의 돌연변이를 임의로 포함하고, 어댑터 단백질은 하나 이상의 기능성 도메인과 회합되거나; 또는 데드 gRNA는 적어도 하나의 비코딩 기능성 루프를 갖도록 변형되고, 조성물은 둘 이상의 어댑터 단백질을 포

함하고, 각각의 단백질은 하나 이상의 기능성 도메인과 회합된다.

[0203] 일정 구현예에서, 어댑터 단백질은 기능성 도메인을 포함하는 융합 단백질이고, 상기 융합 단백질은 선택적으로 어댑터 단백질과 기능성 도메인 사이에 링커를 포함하며, 링커는 선택적으로 GlySer 링커를 포함한다.

[0204] 일정 구현예에서, 데드 gRNA의 적어도 하나의 루프는 2 이상의 어댑터 단백질에 결합하는 별개의 RNA 서열(들)의 삽입에 의해 변형되지 않는다.

[0205] 일정 구현예에서, 어댑터 단백질에 결합된 하나 이상의 기능성 도메인은 전사 활성화 도메인이다.

[0206] 일정 구현예에서, 어댑터 단백질에 결합된 하나 이상의 기능성 도메인은 VP64, p65, MyoD1, HSF1, RTA 또는 SET7/9를 포함하는 전사 활성화 도메인이다.

[0207] 일정 구현예에서, 어댑터 단백질에 결합된 하나 이상의 기능성 도메인은 전사 억제인자 도메인이다.

[0208] 일정 구현예에서, 전사 억제인자 도메인은 KRAB 도메인이다.

[0209] 일정 구현예에서, 전사 억제인자 도메인은 NuE 도메인, NcoR 도메인, SID 도메인 또는 SID4X 도메인이다.

[0210] 일정 구현예에서, 어댑터 단백질에 결합된 하나 이상의 기능성 도메인 중 적어도 하나는 메틸라제 활성, 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, DNA 통합 활성 RNA 절단 활성, DNA 절단 활성 또는 핵산 결합 활성을 포함하는 한 가지 이상의 활성을 가진다.

[0211] 일정 구현예에서, DNA 절단 활성은 Fok1 뉴클레아제에 기인한다.

[0212] 일정 구현예에서, 데드 gRNA는 데드 gRNA가 어댑터 단백질에 결합하고 추가로 C2c1 및 표적에 결합하도록 변형되고, 기능성 도메인이 그의 기인하는 기능에서 작동하는 것을 가능하게 하는 공간적 배향이다.

[0213] 일정 구현예에서, 데드 gRNA의 적어도 하나의 루프는 테트라 루프 및/또는 루프2이다. 일정 구현예에서, 데드 gRNA의 테트라 루프 및 루프 2는 별개의 RNA 서열(들)의 삽입에 의해 변형된다.

[0214] 일정 구현예에서, 하나 이상의 어댑터 단백질에 결합하는 별개의 RNA 서열(들)의 삽입은 압타머 서열이다. 일정 구현예에서, 압타머 서열은 동일한 어댑터 단백질에 특이적인 2 이상의 압타머 서열이다. 일정 구현예에서, 압타머 서열은 상이한 어댑터 단백질에 특이적인 2 이상의 압타머 서열이다.

[0215] 일정 구현예에서, 어댑터 단백질은 MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ΦCb5, ΦCb8r, ΦCb12r, ΦCb23r, 7s, PRR1을 포함한다.

[0216] 일정 구현예에서, 세포는 진핵생물 세포이다. 일정 구현예에서, 진핵생물 세포는 포유류 세포, 임의로 마우스 세포이다. 일정 구현예에서, 포유류 세포는 인간 세포이다.

[0217] 일정 구현예에서, 제1 어댑터 단백질은 p65 도메인과 결합되고, 제2 어댑터 단백질은 HSF1 도메인과 결합된다.

[0218] 일정 구현예에서, 조성물은 적어도 3개의 기능성 도메인을 갖는 C2c1 CRISPR-Cas 복합체를 포함하며, 이 중 적어도 하나는 C2c1과 결합되고, 이 중 적어도 둘은 데드 gRNA와 결합된다.

[0219] 일정 구현예에서, 조성물은 제2 gRNA를 추가로 포함하되, 제2 C2c1 CRISPR-Cas 시스템이 시스템의 C2c1 효소의 뉴클레아제 활성으로부터 초래된 제2 게놈 유전자좌에서 검출 가능한 indel 활성을 갖는 세포에서 관심 대상의 제2 게놈 유전자좌로 향하도록, 제2 gRNA는 제2 표적 서열에 하이브리드화할 수 있는 생 gRNA이다.

[0220] 일정 구현예에서, 조성물은 복수의 데드 gRNA 및/또는 복수의 생 gRNA를 추가로 포함한다.

[0221] 본 발명의 일 양태는 gRNA 스캐폴드의 모듈성 및 맞춤성을 이용하여, 직교 방식으로 별개의 유형의 이펙터를 동원하기 위한 상이한 결합 부위(특히 압타머)를 갖는 일련의 gRNA 스캐폴드를 확립하는 것이다. 다시, 더 넓은 개념의 예시 및 실례를 위하여, PP7-상호작용 스템-루프로의 MS2 스템-루프의 대체를 사용하여, 억제 요소를 결합/동원하여, 다중화된 양방향성 전사 제어를 가능하게 할 수 있다. 따라서, 일반적으로, 데드 가이드를 포함하는 gRNA는 다중복합 전사 제어 및 바람직한 양방향 전사 제어를 제공하는 데 사용될 수 있다. 이 전사 제어는 유전자에서 가장 바람직하다. 예를 들어, 데드 가이드(들)를 포함하는 하나 이상의 gRNA는 하나 이상의 표적 유전자의 활성화를 표적화하는 데 사용될 수 있다. 동시에, 데드 가이드(들)를 포함하는 하나 이상의 gRNA는 하나 이상의 표적 유전자의 억제를 표적화하는 데 사용될 수 있다. 이러한 서열은 다양한 상이한 조합으로 적용될 수 있으며, 예를 들어, 표적 유전자는 처음 억제되고, 이어서, 적절한 기간에 다른 표적

이 활성화되거나, 또는 선택 유전자가 활성화되는 것과 동시에 선택 유전자가 억제되며, 추가 활성화 및/또는 억제가 이어진다. 그 결과, 하나 이상의 생물학적 시스템의 다중 성분은 유리하게 함께 처리될 수 있다.

[0222] 양상에서, 본 발명은 데드 gRNA를 코딩하는 핵산 분자(들) 또는 C2c1 CRISPR-Cas 복합체 또는 본 명세서에 기재된 바와 같은 조성물을 제공한다.

[0223] 양상에서, 본 발명은 본 명세서에 정의된 바와 같은 데드 가이드 RNA를 코딩하는 핵산 분자를 포함하는 벡터 시스템을 제공한다. 일정 구현예에서, 벡터 시스템은 C2c1을 코딩하는 핵산 분자(들)를 추가로 포함한다. 일정 구현예에서, 벡터 시스템은 (생) gRNA를 코딩하는 핵산 분자(들)를 추가로 포함한다. 일정 구현예에서, 핵산 분자 또는 벡터는 가이드 서열(gRNA)을 코딩하는 핵산 분자 및/또는 C1c1을 코딩하는 핵산 분자 및/또는 선택적 핵 국재화 서열(들)에 작동적으로 연결된 진핵생물 세포에서 작동 가능한 조절 서열(들)을 추가로 포함한다.

[0224] 또 다른 양태에서, 또한, 구조적 분석을 사용하여 DNA 결합을 가능하게 하지만 DNA 절단을 가능하게 하지 않는 데드 가이드와 활성 C2c1 뉴클레아제 간의 상호작용을 연구할 수 있다. 이러한 방식으로 C2c1의 뉴클레아제 활성에 중요한 아미노산이 결정된다. 이러한 아미노산의 변형은 유전자 편집을 위해 사용되는 C2c1 효소의 개선을 허용한다.

[0225] 추가적인 양상은 본 명세서에 설명될 뿐만 아니라 당업계에 공지된 바와 같은 CRISPR의 다른 적용과 본 명세서에 설명된 바와 같은 데드 가이드의 사용을 조합하는 것이다. 예를 들어, 표적화된 다중복합 유전자 활성화 또는 억제 또는 표적화된 다중복합 양방향 유전자 활성화/억제를 위해 데드 가이드(들)를 포함하는 gRNA는 본 명세서에 설명된 바와 같은 뉴클레아제 활성을 유지하는 가이드를 포함하는 gRNA와 조합될 수 있다. 뉴클레아제 활성을 유지하는 가이드를 포함하는 이러한 gRNA는 유전자 활성의 억제를 가능하게 하는 변형(예를 들어, 압타머)을 추가로 포함할 수 있거나 또는 포함하지 않을 수도 있다. 뉴클레아제 활성을 유지하는 가이드를 포함하는 이러한 gRNA는 유전자 활성의 활성화를 가능하게 하는 변형(예를 들어, 압타머)을 추가로 포함할 수 있거나 또는 포함하지 않을 수도 있다. 이러한 방식으로, 다중복합 유전자 제어를 위한 추가적인 수단이 도입된다(예를 들어, 뉴클레아제 활성 없이/indel 활성 없이, 다중복합 유전자 표적화된 활성화는 뉴클레아제 활성에 의한 유전자 표적화된 억제와 동시에 또는 이와 조합하여 제공될 수 있다).

[0226] 예를 들어, 1) 하나 이상의 유전자를 표적화하는 데드 가이드(들)를 포함하고 유전자 활성인자의 동원을 위해 적절한 어댑터로 더욱 변형된 하나 이상의 gRNA (예를 들어, 1-50, 1-40, 1-30, 1-20, 바람직하게 1-10, 보다 바람직하게 1-5)의 사용; 2) 하나 이상의 유전자를 표적화하는 데드 가이드(들)를 포함하고 유전자 억제인자의 동원에 적절한 압타머로 더욱 변형된 하나 이상의 gRNA (예를 들어, 1-50, 1-40, 1-30, 1-20, 바람직하게 1-10, 보다 바람직하게 1-5)와 조합될 수 있다. 이어서, 1) 및/또는 2)는 3) 하나 이상의 유전자에 표적화된 하나 이상의 gRNA(예를 들어, 1 내지 50, 1 내지 40, 1 내지 30, 1 내지 20, 바람직하게는 1 내지 10, 더 바람직하게는 1 내지 5)와 조합될 수 있다. 이어서, 이 조합은 결국 1) + 2) + 3)과 함께 4) 하나 이상의 유전자에 표적화되고 유전자 억제인자의 동원을 위해 적절한 압타머로 추가로 변형된 하나 이상의 gRNA (예를 들어, 1 내지 50, 1 내지 40, 1 내지 30, 1 내지 20, 바람직하게는 1 내지 10, 더 바람직하게는 1 내지 5)로 수행될 수 있다. 이어서, 이 조합은 결국 1) + 2) + 3) + 4)와 함께 5) 하나 이상의 유전자에 표적화되고 유전자 억제인자의 동원을 위해 적절한 압타머로 추가로 변형된 하나 이상의 gRNA(예를 들어, 1 내지 50, 1 내지 40, 1 내지 30, 1 내지 20, 바람직하게는 1 내지 10, 더 바람직하게는 1 내지 5)로 수행될 수 있다. 그 결과, 다양한 용도 및 조합이 본 발명에 포함된다. 예를 들어, 조합 1) + 2); 조합 1) + 3); 조합 2) + 3); 조합 1) + 2) + 3); 조합 1) + 2) +3) +4); 조합 1) + 3) + 4); 조합 2) + 3) +4); 조합 1) + 2) + 4); 조합 1) + 2) +3) +4) + 5); 조합 1) + 3) + 4) +5); 조합 2) + 3) +4) +5); 조합 1) + 2) + 4) +5); 조합 1) + 2) +3) + 5); 조합 1) + 3) +5); 조합 2) + 3) +5); 조합 1) + 2) +5).

[0227] 일 양상에서, 본 발명은 표적 유전자 좌위에 C2c1 CRISPR-Cas 시스템을 가이드하기 위해 데드 가이드 RNA 표적화 서열(데드 가이드 서열)을 설계하거나, 평가하거나 또는 선택하기 위한 알고리즘을 제공한다. 특히, 데드 가이드 RNA 특이성은 i) GC 함량 및 ii) 표적화 서열 길이에 관한 것이며, 이들을 변화시킴으로써 최적화될 수 있다는 것이 결정되었다. 양상에서, 본 발명은 데드 가이드 RNA의 비표적 결합 또는 상호작용을 최소화하는 데드 가이드 RNA 표적화 서열을 설계하거나 또는 평가하기 위한 알고리즘을 제공한다. 본 발명의 일 구현예에서, 유기체 내 유전자 좌위로 CRISPR 시스템을 유도하기 위한 데드 가이드 RNA 표적화 서열을 선택하기 위한 알고리즘은 a) 유전자 좌위 내 하나 이상의 CRISPR 모티프를 위치시키는 단계, i) 서열의 GC 함량을 결정

하는 단계; ii) 유기체의 게놈에서 CRISPR 모티프에 가장 가까운 15 하류 뉴클레오티드의 오프-표적 매치가 존재하는지 여부를 결정하는 단계에 의해 각각의 CRISPR 모티프의 하류 20 nt 서열을 분석하는 단계, 및 c) 서열의 GC 함량이 70% 이하이고 오프-표적 매치가 확인되지 않으면 데드 가이드 RNA에서 사용을 위한 15 뉴클레오티드 서열을 선택하는 단계를 포함한다. 구현예에서, GC 함량이 60% 이하라면, 서열은 표적화 서열에 대해 선택된다. 일정 구현예에서, GC 함량이 55% 이하, 50% 이하, 45% 이하, 40% 이하, 35% 이하 또는 30% 이하라면, 서열은 표적화 서열을 위해 선택된다. 구현예에서, 유전자 좌위의 2 이상의 서열이 분석되며, 가장 낮은 GC 함량, 다음의 낮은 GC 함량 또는 다음의 낮은 GC 함량을 갖는 서열이 선택된다. 구현예에서, 오프-표적 매치가 유기체 게놈에서 동정되지 않는다면, 서열은 표적화 서열에 대해 선택된다. 구현예에서, 오프-표적 매치가 게놈의 조절 서열에서 확인되지 않는다면, 표적화 서열이 선택된다.

[0228] 본 발명의 일 양상에서, 본 발명은 유기체 내 유전자 좌위로 기능화된 CRISPR 시스템을 유도하기 위한 데드 가이드 RNA 표적화 서열을 선택하기 위한 방법을 제공하고, a) 유전자 좌위 내 하나 이상의 CRISPR 모티프를 위치시키는 단계, b) i) 서열의 GC 함량을 결정하는 단계; ii) 유기체의 게놈에서 서열의 제1 15 nt의 오프-표적 매치가 존재하는지 여부를 결정하는 단계에 의해 각각의 CRISPR 모티프의 하류 20 nt 서열을 분석하는 단계, 및 c) 서열의 GC 함량이 70% 이하이고 오프-표적 매치가 확인되지 않으면 가이드 RNA에서 사용을 위한 서열을 선택하는 단계를 포함한다. 구현예에서, GC 함량이 50% 이하라면, 서열은 선택된다. 구현예에서, GC 함량이 40% 이하라면, 서열은 선택된다. 구현예에서, GC 함량이 30% 이하라면, 서열은 선택된다. 구현예에서, 2 이상의 서열이 분석되고, 가장 낮은 GC 함량을 갖는 서열이 선택된다. 구현예에서, 오프-표적 매치가 유기체의 조절 서열에서 결정된다. 구현예에서, 유전자 좌위는 조절 영역이다. 양상은 앞서 언급한 방법에 따라 선택된 표적화 서열을 포함하는 데드 가이드 RNA를 제공한다.

[0229] 양상에서, 본 발명은 유기체에서 작용화된 CRISPR 시스템을 유전자 좌위에 표적화하기 위한 데드 가이드 RNA를 제공한다. 본 발명의 구현예에서, 데드 가이드 RNA는 표적화 서열을 포함하되, 표적 서열의 GC 함량은 70% 이하이고, 표적화 서열의 처음 15 nt는 유기체에서 다른 유전자 좌위의 조절 서열 내 CRISPR 모티프로부터의 하류에 오프-표적 서열을 매칭하지 않는다. 일정 구현예에서, 표적화 서열의 GC 함량은 60% 이하, 55% 이하, 50% 이하, 45% 이하, 40% 이하, 35% 이하 또는 30% 이하이다. 일정 구현예에서, 표적화 서열의 GC 함량은 70% 내지 60% 또는 60% 내지 50% 또는 50% 내지 40% 또는 40% 내지 30%이다. 구현예에서, 표적화 서열은 유전자좌의 잠재적 표적화 서열 중에 가장 낮은 GC 함량을 가진다.

[0230] 본 발명의 구현예에서, 데드 가이드의 처음 15 nt는 표적 서열에 매칭된다. 다른 구현예에서, 데드 가이드의 처음 14 nt는 표적 서열에 매칭된다. 다른 구현예에서, 데드 가이드의 처음 13 nt는 표적 서열에 매칭된다. 다른 구현예에서, 데드 가이드의 처음 12 nt는 표적 서열에 매칭된다. 다른 구현예에서, 데드 가이드의 처음 11 nt는 표적 서열에 매칭된다. 다른 구현예에서, 데드 가이드의 처음 10 nt는 표적 서열에 매칭된다. 다른 구현예에서, 데드 가이드의 처음 15 nt는 다른 유전자 좌위의 조절 영역에서 CRISPR 모티프로부터의 하류의 오프-표적 서열에 매칭되지 않는다. 다른 구현예에서, 데드 가이드의 처음 14 nt, 또는 처음 13 nt, 또는 가이드의 처음 12 nt 또는 데드 가이드의 처음 11 nt 또는 데드 가이드의 처음 10 nt는 다른 유전자 좌위의 조절 영역에서 CRISPR 모티프로부터의 하류의 오프-표적 서열에 매칭되지 않는다. 다른 구현예에서, 데드 가이드의 처음 15 nt, 또는 14 nt, 또는 13 nt, 또는 12 nt, 또는 11 nt는 게놈에서 CRISPR 모티프로부터의 하류의 오프-표적 서열에 매칭되지 않는다.

[0231] 일정 구현예에서, 데드 가이드 RNA는 표적 서열과 일치하지 않는 3' 말단에 추가의 뉴클레오티드를 포함한다. 따라서, CRISPR 모티프의 하류의 처음 15 nt 또는 14 nt 또는 13 nt 또는 12 nt 또는 11 nt를 포함하는 데드 가이드 RNA는 3' 말단에서 길이가 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt 이상까지 연장될 수 있다.

[0232] 본 발명은 데드 C2c1 (dC2c1) 또는 기능화된 C2c1 시스템(기능화된 C2c1 또는 기능화된 가이드를 포함할 수 있음)을 포함하지만, 이들로 제한되지 않는 C2c1 CRISPR-Cas 시스템을 유전자 좌위로 보내는 방법을 제공한다. 양상에서, 본 발명은 데드 가이드 RNA 표적화 서열을 선택하고, 작용화된 CRISPR 시스템을 유기체에서 유전자 좌위로 보내기 위한 방법을 제공한다. 양상에서, 본 발명은 데드 가이드 RNA 표적화 서열을 선택하고 기능화된 C2c1 CRISPR-Cas 시스템에 의해 표적 유전자 좌위의 유전자 조절을 달성하는 방법을 제공한다. 일정 구현예에서, 상기 방법은 표적 유전자 조절을 달성하는 한편, 오프-표적 효과를 최소화하기 위해 사용된다. 양상에서, 본 발명은 2 이상의 데드 가이드 RNA 표적화 서열을 선택하고 기능화된 C2c1 CRISPR-Cas 시스템에 의해 2 이상의 표적 유전자 좌위의 유전자 조절을 달성하는 방법을 제공한다. 일정 구현예에서, 상기 방법

은 2 이상의 표적 유전자 좌위의 조절을 달성하는 한편, 오프-표적 효과를 최소화하기 위해 사용된다.

[0233] 양상에서, 본 발명은 유기체의 유전자 좌위로 기능화된 C2c1을 유도시키기 위한 데드 가이드 RNA 표적화 서열을 선택하는 방법을 제공하고, 방법은 a) 유전자 좌위에 하나 이상의 CRISPR 모티프를 위치시키는 단계; b) i) CRISPR 모티프에 인접한 10 내지 15 nt를 선택하는 단계, ii) 서열의 GC 함량을 결정하는 단계에 의해 각각의 CRISPR 모티프의 하류 서열을 분석하는 단계; 및 c) 서열의 GC 함량이 40% 이상이면 가이드 RNA에서 사용을 위한 표적화 서열로서 10 내지 15 nt 서열을 선택하는 단계를 포함한다. 구현예에서, GC 함량이 50% 이상이라면 서열은 선택된다. 구현예에서, GC 함량이 60% 이상이라면 서열은 선택된다. 구현예에서, GC 함량이 70% 이상이라면 서열은 선택된다. 구현예에서, 2 이상의 서열이 분석되고, 가장 높은 GC 함량을 갖는 서열이 선택된다. 일 실시형태에서, 방법은 CRISPR 모티프의 하류 서열과 매치되지 않는 선택된 서열의 3' 말단에 뉴클레오티드를 첨가하는 단계를 더 포함한다. 양상은 앞서 언급한 방법에 따라 선택된 표적화 서열을 포함하는 데드 가이드 RNA를 제공한다.

[0234] 양상에서, 본 발명은 유기체에서 작용화된 CRISPR 시스템을 유전자 좌위로 보내기 위한 데드 가이드 RNA를 제공하되, 데드 가이드 RNA의 표적화 서열은 유전자 좌위의 CRISPR 모티프에 인접한 10 내지 15개의 뉴클레오티드로 이루어지고, 표적 서열의 CG 함량은 50% 이상이다. 일정 실시형태에서, 데드 가이드 RNA는 유전자 좌위의 CRISPR 모티프의 하류 서열과 매치되지 않는 표적화 서열의 3' 말단에 첨가된 뉴클레오티드를 더 포함한다.

[0235] 양상에서, 본 발명은 하나 이상, 또는 둘 이상의 유전자 좌위로 유도되는 단일 이펙터를 제공한다. 일정 구현예에서, 이펙터는 C2c1과 회합되고, 하나 이상, 또는 둘 이상의 선택된 데드 가이드 RNA는 C2c1-회합된 이펙터를 하나 이상, 또는 둘 이상의 선택된 표적 유전자 좌위로 유도시키는데 사용된다. 일정 구현예에서, 이펙터는 하나 이상, 또는 둘 이상의 선택된 데드 가이드 RNA와 결합되고, C2c1 효소와 복합체화될 때 각각의 선택된 데드 가이드 RNA는 이의 회합된 이펙터가 데드 가이드 RNA 표적에 국재화되게 야기시킨다. 이러한 CRISPR 시스템의 한 가지 비제한적 예는 동일한 전사 인자에 의한 조절에 대해 하나 이상 또는 둘 이상의 유전자 좌위 대상의 활성을 조절한다.

[0236] 양상에서, 본 발명은 하나 이상의 유전자 좌위로 유도되는 둘 이상의 이펙터를 제공한다. 일정 구현예에서, 둘 이상의 데드 가이드 RNA가 사용되며, 둘 이상의 이펙터 각각은 선택된 데드 가이드 RNA와 결합되고, 둘 이상의 이펙터 각각은 그의 데드 가이드 RNA의 선택된 표적으로 국소화된다. 이러한 CRISPR 시스템의 한 가지 비제한적 예는 상이한 전사 인자에 의한 조절에 대해 하나 이상, 또는 둘 이상의 유전자 좌위 대상을 조절한다. 따라서, 하나의 비제한적 구현예에서, 둘 이상의 전사 인자는 단일 유전자의 상이한 조절 서열로 국재화된다. 다른 비제한적 구현예에서, 둘 이상의 전사 인자가 상이한 유전자의 상이한 조절 서열로 국재화된다. 일정 구현예에서, 하나의 전사 인자는 활성인자이다. 일정 구현예에서, 하나의 전사 인자는 억제제이다. 일정 구현예에서, 하나의 전사 인자는 활성인자이고, 다른 전사 인자는 억제제이다. 일정 구현예에서, 동일한 조절 경로의 상이한 성분을 발현시키는 유전자 좌위가 조절된다. 일정 구현예에서, 상이한 조절 경로의 성분을 발현시키는 유전자 좌위가 조절된다.

[0237] 양상에서, 본 발명은 또한 표적 DNA 절단 또는 표적 결합 및 활성 C2c1 CRISPR-Cas 시스템에 의해 매개되는 유전자 조절에 특이적인 데드 가이드 RNA를 설계하고 선택하기 위한 방법 및 알고리즘을 제공한다. 일정 구현예에서, C2c1 CRISPR-Cas 시스템은 하나의 유전자 좌위에서 표적 DNA를 절단하는 한편, 동시에 다른 유전자 좌위에 결합하고, 이의 조절을 촉진시키는 활성 C2c1을 이용하여 직교 유전자 제어를 제공한다.

[0238] 양상에서, 본 발명은 절단 없이, 유기체에서 유전자 좌위로 기능화된 Cas12b를 유도시키기 위한 데드 가이드 RNA 표적화 서열을 선택하는 방법을 제공한다,. 일정 구현예에서, 방법은 a) 유전자 좌위에 하나 이상의 CRISPR 모티프를 위치시키는 단계; b) i) CRISPR 모티프에 인접한 10 내지 15 nt를 선택하는 단계, ii) 서열의 GC 함량을 결정하는 단계에 의해 각각의 CRISPR 모티프 하류의 서열을 분석하는 단계, 및 c) 서열의 GC 함량이 30% 이상, 40% 이상이면 데드 가이드 RNA에서 사용하기 위한 표적화 서열로서 10 내지 15 nt 서열을 선택하는 단계를 포함한다. 일정 구현예에서, 표적화 서열의 GC 함량은 35% 이상, 40% 이상, 45% 이상, 50% 이상, 55% 이상, 60% 이상, 65% 이상, 또는 70% 이상이다. 일정 구현예에서, 표적화 서열의 GC 함량은 30% 내지 40% 또는 40% 내지 50% 또는 50% 내지 60% 또는 60% 내지 70%이다. 본 발명의 구현예에서, 유전자 좌위에서 둘 이상의 서열이 분석되고, 가장 높은 GC 함량을 갖는 서열이 선택된다.

[0239] 본 발명의 구현예에서, GC 함량이 평가되는 표적화 서열의 일부는 PAM에 가장 가까운 15개의 표적 뉴클레오티드의 10 내지 15개의 인접한 뉴클레오티드이다. 본 발명의 구현예에서, GC 함량이 고려되는 가이드의 일부는 PAM에 가장 가까운 15개의 뉴클레오티드 중 10 내지 11개의 뉴클레오티드 또는 11 내지 12개의 뉴클레오티드 또

는 12 내지 13개의 뉴클레오티드 또는 13, 또는 14, 또는 15개의 인접한 뉴클레오티드이다.

[0240] 일 양태에서, 본 발명은 기능적 활성화 또는 억제를 피하면서, CRISPR 시스템 유전자좌 절단을 촉진시키는 데드 가이드 RNA를 확인하기 위한 알고리즘을 추가로 제공한다. 16 내지 20개의 뉴클레오티드의 데드 가이드 RNA에서 증가된 GC 함량은 증가된 DNA 절단 및 감소된 기능성 활성화와 동시에 일어난다는 것이 관찰된다.

[0241] 기능화된 Cas12b의 효율은 CRISPR 모티프 하류의 표적 서열과 매치되지 않는 가이드 RNA의 3' 말단부에 뉴클레오타이드의 첨가에 의해 증가될 수 있다. 예를 들어, 길이가 11 내지 15 nt인 데드 가이드 RNA 중에서, 가이드가 짧을수록 표적 절단을 촉진시킬 가능성은 적을 수 있지만, 이는 또한 CRISPR 시스템 결합 및 기능성 제어의 촉진에서 덜 효율적이다. 특정 구현예에서, 데드 가이드 RNA의 3' 말단으로의 표적 서열과 매치되지 않는 뉴클레오티드의 첨가는 활성화 효율을 증가시키는 한편, 요망되지 않는 표적 절단을 증가시키지 않는다. 양상에서, 본 발명은 또한 DNA 결합 및 유전자 조절에서 CRISPRP 시스템을 효과적으로 촉진시키는 한편, DNA 절단을 촉진시키지 않는 개선된 데드 가이드 RNA를 확인하기 위한 방법 및 알고리즘을 제공한다. 따라서, 특정 구현예에서, 본 발명은 CRISPR 모티프의 하류의 처음 15 nt 또는 14 nt 또는 13 nt 또는 12 nt 또는 11 nt를 포함하며, 3' 말단에서 표적과 미스매치되는 뉴클레오티드에 의해 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt 이상으로 길이가 연장되는 데드 가이드 RNA를 제공한다.

[0242] 양상에서, 본 발명은 선택적 직교 유전자 제어를 달성하기 위한 방법을 제공한다. 본 명세서의 개시내용으로부터 인식될 바와 같이, 가이드 길이 및 GC 함량을 고려하는 본 발명에 따른 데드 가이드 선택은, 예를 들어, 활성화 또는 저해에 의해 유전자 좌위의 전사를 조절하고 오프-표적 효과를 최소화하기 위해, 기능성 Cas12b CRISPR-Cas 시스템에 의한 효과적이고 선택적인 전사 제어를 제공한다. 따라서, 개개 표적 유전자좌의 효과적인 조절을 제공함으로써, 본 발명은 또한 둘 이상의 표적 유전자좌의 효과적인 직교 조절을 제공한다.

[0243] 일정 구현예에서, 직교 유전자 제어는 둘 이상의 표적 유전자좌의 활성화 또는 저해에 의한다. 일정 구현예에서, 직교 유전자 제어는 하나 이상의 표적 유전자좌의 활성화 또는 저해 및 하나 이상의 표적 유전자좌의 절단에 의한다.

[0244] 일 양상에서, 본 발명은 본 명세서에 기재된 방법 또는 알고리즘에 따라 개시되거나 또는 이루어진 하나 이상의 데드 가이드 RNA를 포함하는 비천연 발생 Cas12b CRISPR-Cas 시스템을 포함하는 세포를 제공하되, 하나 이상의 유전자 산물의 발현은 변경되었다. 본 발명의 구현예에서, 둘 이상의 유전자 산물의 세포에서 발현이 변경되었다. 본 발명은 또한 이러한 세포로부터의 세포주를 제공한다.

[0245] 일 양상에서, 본 발명은 본 명세서에 기재된 방법 또는 알고리즘에 따라 개시되거나 또는 이루어진 하나 이상의 데드 가이드 RNA를 포함하는 비천연 발생 Cas12b CRISPR-Cas 시스템을 포함하는 하나 이상의 세포를 포함하는 다세포 유기체를 제공한다. 일 양상에서, 본 발명은 본 명세서에 기재된 방법 또는 알고리즘에 따라 개시되거나 또는 이루어진 하나 이상의 데드 가이드 RNA를 포함하는 비천연 발생 Cas12b CRISPR-Cas 시스템을 포함하는 세포, 세포주 또는 다세포 유기체로부터의 산물을 제공한다.

[0246] 본 발명의 추가적인 양상은, 선택적으로 본 명세서에 기재된 바와 같은 또는 기술 상태에서 가이드(들)를 포함하는 gRNA와 조합하여, Cas12b의 과발현 또는 바람직하게는 녹인 Cas12b에 대해 조작된 시스템, 예를 들어, 세포, 유전자이식 동물, 유전자이식 마우스, 유도성 유전자이식 동물, 유도성 유전자이식 마우스와 조합하여, 본 명세서에 기재된 바와 같은 데드 가이드(들)를 포함하는 gRNA의 용도이다. 결과로서, 단일 시스템(예를 들어, 유전자이식 동물, 세포)는 시스템/네트워크 생물학에서 다중복합 유전자 변형에 대한 기준으로서 작용할 수 있다. 데드 가이드 때문에, 이는 이제 시험관내, 생체외와 생체내에서 가능하다.

[0247] 예를 들어, 일단 Cas12b이 제공되면, 하나 이상의 데드 gRNA는 다중복합 유전자 조절, 바람직하게는 다중복합 양방향 유전자 조절을 보내도록 제공될 수 있다. 필요하거나 요망된다면 하나 이상의 데드 gRNA는 공간적으로 그리고 일시적으로 적절한 방식(예를 들어, Cas12b 발현의 조직 특이적 유도)으로 제공될 수 있다. 관심 대상의 세포, 조직, 동물에서 유전자이식/유도성 Cas12b가 제공되기 때문에(예를 들어, 발현됨) 데드 가이드를 포함하는 gRNA와 가이드를 포함하는 gRNA는 동일하게 효과적이다. 동일한 방식으로, 본 발명의 추가적인 양상은, 선택적으로 본 명세서에 기재된 바와 같은 또는 기술 상태에서 가이드(들)를 포함하는 gRNA와 조합하여, 녹아웃 Cas12b CRISPR-Cas를 위해 조작된 시스템(예를 들어, 세포, 유전자이식 동물, 유전자이식 마우스, 유도성 유전자이식 동물, 유도성 유전자이식 마우스)과 조합하여, 본 명세서에 기재된 바와 같은 데드 가이드(들)를 포함하는 gRNA의 용도이다.

[0248] 그 결과, 본 명세서에 기재된 CRISPR 적용 및 당업계에 공지된 CRISPR 적용과 함께 본 명세서에 기재된 바와 같

은 데드 가이드의 조합은 시스템의 다중복합 스크리닝을 위한 고도로 효율적이고 정확한 수단(예를 들어, 네트워크 생물학)을 초래한다. 이러한 스크리닝은, 예를 들어, 질환(예를 들어, 온/오프 조합), 특히 유전자 관련 질환을 초래하는 유전자를 동정하기 위한 유전자 활성의 구체적 조합의 확인을 허용한다. 이러한 스크리닝의 바람직한 적용은 암이다. 동일한 방식에서, 이러한 질환에 대한 치료를 위한 스크리닝이 본 발명에 포함된다. 세포 또는 동물은 비정상 병태에 노출되어 질환 또는 질환 유사 효과를 초래할 수 있다. 후보 조성물이 제공될 수 있으며, 목적하는 다중복합 환경에서의 효과에 대해 스크리닝될 수 있다. 예를 들어, 유전자 조합이 그들이 사망하는 것을 야기하는 환자의 암 세포가 스크리닝될 수 있으며, 이어서, 적절한 요법을 확립하기 위해 이 정보를 사용할 수 있다.

[0249] 일 양상에서, 본 발명은 본 명세서에 기재된 성분 중 하나 이상을 포함하는 키트를 제공한다. 키트는 본 명세서에 기재된 바와 같은 가이드를 갖거나 또는 가이드가 없는 본 명세서에 기재된 바와 같은 데드 가이드를 포함할 수 있다.

[0250] 본 명세서에 제공된 구조적 정보는 표적 DNA 및 Cas12b와 데드 gRNA 상호작용의 질의를 가능하게 하여, 전체 Cas12b CRISPR-Cas 시스템의 작용성을 최적화하도록 데드 gRNA 구조의 조작 또는 변경을 허용한다. 예를 들어, 데드 gRNA의 루프는 RNA에 결합할 수 있는 어댑터 단백질의 삽입에 의한 Cas12b 단백질과의 충돌 없이 연장될 수 있다. 이들 어댑터 단백질은 하나 이상의 기능성 도메인을 포함하는 이펙터 단백질 또는 융합을 추가로 보충할 수 있다.

[0251] 일부 바람직한 구현예에서, 기능성 도메인은 전사 활성화 도메인, 바람직하게는 VP64이다. 일부 구현예에서, 기능성 도메인은 전사 억제 도메인, 바람직하게는 KRAB이다. 일부 구현예에서, 전사 억제 도메인은 SID, 또는 SID의 콘카타머(concatemer)(예를 들어, SID4X)이다. 일부 구현예에서, 후성적 변형 효소가 제공되도록, 기능성 도메인은 후성적 변형 도메인이다. 일부 구현예에서, 기능성 도메인은 P65 활성화 도메인일 수 있는 활성화 도메인이다. 일부 구현예에서, Cas12b 이펙터 단백질은 하나 이상의 기능성 도메인과 회합되고; Cas12b 이펙터 단백질은 RuvC 및/또는 Nuc 도메인 내에 하나 이상의 돌연변이를 함유하여서, 형성된 CRISPR 복합체가 후성적 변형자 또는 전사 또는 번역 활성화 또는 억제 신호를 전달할 수 있다.

[0252] 본 발명의 양상은 상기 요소가 단일 조성물에 포함되거나 또는 개개 조성물에 포함된다는 것이다. 이들 조성물은 게놈 수준에 대해 기능성 효과를 유발하기 위해 숙주에 유리하게 적용될 수 있다.

[0253] 일반적으로, 데드 gRNA는 (예를 들어, 융합 단백질을 통해) 하나 이상의 기능성 도메인을 포함하는 어댑터 단백질이 결합하는 특정 결합 부위(예를 들어, 압타머)를 제공하는 방식으로 변형된다. 일단 데드 gRNA가 CRISPR 복합체(즉, 데드 gRNA 및 표적에 결합하는 Cas12b)를 형성하도록 변형된 데드 gRNA는 변형되고, 어댑터 단백질이 결합하며, 어댑터 단백질 상의 기능성 도메인은 속성 작용이 유효하게 되는 것이 유리한 공간적 배향으로 위치된다. 예를 들어, 기능성 도메인이 전사 활성인자(예를 들어, VP64 또는 p65)라면, 전사 활성인자는 표적의 전사에 영향을 미치도록 허용하는 공간적 배향으로 위치된다. 마찬가지로, 전사 억제인자는 표적의 전사에 영향을 미치도록 유리하게 위치될 것이며, 뉴클레아제(예를 들어, Fok1)는 표적을 절단하거나 또는 부분적으로 절단하도록 유리하게 위치될 것이다.

[0254] 당업자는 어댑터 + 기능성 도메인의 결합을 가능하게 하지만 어댑터 + 기능성 도메인의 적절한 위치화를 가능하게 하지 않는(예를 들어, CRISPR 복합체의 3차원 구조 내의 입체 장애에 기인) 데드 gRNA에 대한 변형은 의도되지 않은 변형이라는 것을 이해할 것이다.

[0255] 본 명세서에서 설명된 바와 같이, 기능성 도메인은, 예를 들어, 메틸라제 활성, 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, RNA 절단 활성, DNA 절단 활성, 핵산 결합 활성, 및 분자 스위치(예를 들어, 광 유도성)로 이루어진 군으로부터의 하나 이상의 도메인일 수 있다. 일부 경우에, 추가적으로 적어도 하나의 NLS가 제공되는 것이 유리하다. 일부 예에서, N 말단에 NLS를 위치시키는 것이 유리하다. 하나 초과의 기능성 도메인이 포함될 때, 기능성 도메인은 동일하거나 또는 상이할 수 있다.

[0256] 데드 gRNA는 동일 또는 상이한 어댑터 단백질에 특이적인 다중 결합 인식 부위(예를 들어, 압타머)를 포함하도록 설계될 수 있다. 데드 gRNA는 전사 개시 부위(즉, TSS) 상류의 프로모터 영역 -1000 - +1 핵산, 바람직하게는 -200 핵산에 결합하도록 설계될 수 있다. 이 위치화는 유전자 활성화(예를 들어, 전사 활성인자) 또는 유전자 저해(예를 들어, 전사 억제인자)에 영향을 미치는 기능성 도메인을 개선시킨다. 변형된 데드 gRNA는 조성물에 포함된 하나 이상의 표적 유전자좌(예를 들어, 적어도 1 gRNA, 적어도 2 gRNA, 적어도 5 gRNA,

적어도 10 gRNA, 적어도 20 gRNA, 적어도 30 gRNA, 적어도 50 gRNA)에 표적화된 하나 이상의 변형된 데드 gRNA일 수 있다.

[0257] 일단 데드 gRNA가 CRISPR 복합체에 도입되었다면, 어댑터 단백질은 변형된 데드 gRNA에 도입되는 압타머 또는 인식 부위에 결합되고, 하나 이상의 기능성 도메인의 적절한 위치화를 가능하게 하여 속성 작용을 갖는 표적에 영향을 미치는 다수의 단백질일 수 있다. 본 출원에서 상세히 설명된 바와 같이, 이는 외피 단백질, 바람직하게는 박테리오파지 외피 단백질일 수 있다. (예를 들어, 융합 단백질의 형태로) 이러한 어댑터 단백질과 결합된 기능성 도메인은, 예를 들어, 메틸라제 활성, 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, RNA 절단 활성, DNA 절단 활성, 핵산 결합 활성 및 분자 스위치(예를 들어, 광 유도성)로 이루어진 군으로부터의 하나 이상의 도메인을 포함할 수 있다. 바람직한 도메인은 Fok1, VP64, P65, HSF1, MyoD1이다. 기능성 도메인이 전사 활성인자 또는 전사 억제인자인 사건에서, 추가적으로 적어도 NLS가 바람직하게는 N 말단에서 제공되는 것이 유리하다. 하나 초과의 기능성 도메인이 포함될 때, 기능성 도메인은 동일 또는 상이할 수 있다. 어댑터 단백질은 이러한 기능성 도메인에 부착되는 공지된 링커를 이용할 수 있다.

[0258] 따라서, 변형된 데드 gRNA, (비활성화된) Cas12b(기능성 도메인이 있거나 또는 없음), 및 하나 이상의 기능성 도메인을 갖는 결합 단백질은 각각 개개로 조성물에 포함될 수 있고, 숙주에 개개로 또는 총괄적으로 투여될 수 있다. 대안적으로, 이들 성분은 숙주에 대한 투여를 위해 단일 조성물에서 제공될 수 있다. 숙주에 대한 투여는 숙주에 전달을 위해 당업자에게 공지되거나 또는 본 명세서에 기재된 바이러스 벡터(예를 들어, 렌티바이러스 벡터, 아데노바이러스 벡터, AAV 벡터)를 통해 수행될 수 있다. 본 명세서에 설명된 바와 같이, 상이한 선택 마커(예를 들어, 렌티바이러스 gRNA 선택을 위함) 및 gRNA의 농도(예를 들어, 다중 gRNA가 사용되는지의 여부에 의존함)의 사용은 개선된 효과를 유발하는 데 유리할 수 있다.

[0259] 이런 개념을 기초로, DNA 절단, 유전자 활성화 또는 유전자 탈활성화를 비롯한, 게놈 유전자좌 사건을 유발하는 것이 적절하다. 제공된 조성물을 이용하여, 당업자는 하나 이상의 게놈 유전자좌 사건을 유발하기 위해 동일 또는 상이한 기능성 도메인을 갖는 단일 또는 다중 유전자좌를 유리하게 그리고 특이적으로 표적화할 수 있다. 조성물은 세포 내 라이브러리에서의 스크리닝 및 생체내 기능성 모델링을 위한 매우 다양한 방법에 적용될 수 있다(예를 들어, lincRNA의 유전자 활성화 및 기능의 확인; 기능획득 모델링; 기능상실 모델링; 최적화 및 스크리닝 목적을 위해 세포주 및 유전자이식 동물을 확립하는 본 발명의 조성물의 용도).

[0260] 본 발명은 본 발명 또는 적용 전에 믿어지지 않은 조건적 또는 유도성 CRISPR 유전자이식 세포/동물을 확립하고 이용하기 위한 본 발명의 조성물의 용도를 이해한다. 예를 들어, 표적 세포는 조건적으로 또는 유도성으로 (예를 들어, Cre 의존적 구성체의 형태로) Cas12b 및/또는 조건적으로 또는 유도성으로 어댑터 단백질을 포함하고, 표적 세포에 도입된 벡터의 발현에 대해, 벡터는 표적 세포에서 Cas12b 발현 및/또는 어댑터 발현의 조건을 유도하거나 또는 일으키는 것을 발현시킨다. CRISPR 복합체를 생성하는 공지된 방법으로 본 발명의 교시 및 조성물을 적용함으로써, 기능성 도메인에 의해 영향받은 유도성 게놈 사건은 또한 본 발명의 양상이다. 이의 일례는 CRISPR 녹인/조건적 유전자이식 동물(예를 들어, Lox-정지-폴리A-Lox(LSL) 카세트를 포함하는 마우스)의 생성 및 본 명세서에 기재된 바와 같이 하나 이상의 변형된 데드 gRNA(예를 들어, 유전자 활성화 목적을 위한 관심 대상의 표적 유전자의 TSS에 대한 -200개의 뉴클레오티드)(예를 들어, 외피 단백질에 의해 인식되는 하나 이상의 압타머, 예를 들어, MS2를 갖는 변형된 데드 gRNA)를 제공하는 하나 이상의 조성물의 후속적 전달, 본 명세서에 기재되는 바와 같은 하나 이상의 압타머 단백질(하나 이상의 VP64에 연결된 MS2 결합 단백질) 및 조건적 동물을 유도하기 위한 수단(예를 들어, Cas12b 발현 유도성을 제공하기 위한 Cre 재조합효소)이다. 대안적으로, 어댑터 단백질은 스크리닝 목적을 위한 유효 모델을 제공하기 위해 조건적 또는 유도성 Cas12b를 갖는 조건적 또는 유도성 요소로서 제공될 수 있는데, 이는 유리하게는 단지 최소의 설계 및 다수의 적용을 위한 특정 데드 gRNA의 투여를 필요로 한다.

[0261] 다른 양상에서, 데드 가이드는 특이성을 개선시키기 위해 추가로 변형된다. 보호된 데드 가이드가 합성될 수 있으며, 데드 가이드의 3' 말단부에 2차 구조가 도입되어 그의 특이성을 개선시킨다. 보호된 가이드 RNA(pgRNA)는 세포 및 보호자 가닥 내 관심 대상의 게놈 유전자좌에서 표적 서열에 하이브리드화할 수 있는 가이드 서열을 포함하되, 보호자 가닥은 선택적으로 가이드 서열에 대해 상보성이고, 가이드 서열은 보호자 가닥에 일부분 하이브리드화가능할 수 있다. pgRNA는 선택적으로 연장 서열을 포함한다. pgRNA-표적 DNA 하이브리드화의 열역학은 가이드 RNA와 표적 DNA 사이에서 상보성인 염기의 수에 의해 결정된다. '열역학 보호'를 사용함으로써, 데드 gRNA의 특이성은 보호자 서열을 더하는 것에 의해 개선될 수 있다. 예를 들어, 한 방법은 데드 gRNA 내에서 가이드 서열의 3' 말단부에 다양한 길이의 상보성 보호자 가닥을 더한다. 그

결과, 보호자 가닥은 데드 gRNA의 적어도 일부에 결합되고, 보호된 gRNA(pgRNA)를 제공한다. 결국, 본 명세서의 데드 gRNA 언급은 기재된 구현예를 이태 중 하나일 수 있다. 보호자 가닥은 별개의 RNA 전사체 또는 가닥 또는 데드 gRNA 가이드 서열의 3' 말단부에 결합된 키메라 형일 수 있다.

[0262] 본 발명자들은 본 명세서에 정의된 바와 같은 CRISPR 효소가 활성을 상실하는 일 없이 하나 초과의 RNA 가이드를 사용할 수 있다는 것을 나타내었다. 이는 본 명세서에 정의된 바와 같은 단일 효소, 시스템 또는 복합체로 다중 DNA 표적, 유전자 또는 유전자 좌위를 표적화하기 위해 본 명세서에 정의된 바와 같은 CRISPR 효소, 시스템 또는 복합체의 사용을 가능하게 한다. 가이드 RNA는 직렬로 배열되고, 선택적으로 뉴클레오티드 서열, 예컨대 본 명세서에 정의된 바와 같은 직접 반복부에 의해 분리될 수 있다. 상이한 가이드 RNA의 위치는 활성에 영향을 미치지 않는 직렬이다.

[0263] **다중복합 CRISPR-Cas 시스템**

[0264] 일 양상에서, 본 발명은 비천연 발생 또는 조작된 CRISPR 효소, 바람직하게는 클래스 2 CRISPR 효소, 바람직하게는 본 명세서에 기재된 바와 같은 V 또는 VI형 CRISPR 효소, 예컨대 제한 없이, 직렬 또는 다중복합 표적화를 위해 사용되는 본 명세서의 다른 곳에 기재된 바와 같은 Cas12b를 제공한다. 본 명세서의 다른 곳에 기재된 바와 같은 본 발명에 따른 CRISPR(또는 CRISPR-Cas 또는 Cas) 효소, 복합체 또는 시스템 중 어느 것이 이러한 접근에서 사용될 수 있다는 것이 이해되어야 한다. 본 명세서의 다른 곳에 기재된 바와 같은 방법, 생성물, 조성물 및 용도 중 어느 것은 이하에 추가로 상술하는 다중복합 또는 직렬 표적화 접근과 동일하게 적용 가능하다. 추가적인 가이드에 의해, 다음의 특정 양상 및 구현예가 제공된다.

[0265] 일 양상에서, 본 발명은 다중 유전자 좌위를 표적화하기 위해 본 명세서에 정의된 바와 같은 Cas12b 효소, 복합체 또는 시스템의 용도를 제공한다. 일 구현예에서, 이는 다중(직렬 또는 다중복합) 가이드 RNA (gRNA) 서열을 이용함으로써 확립될 수 있다.

[0266] 일 양상에서, 본 발명은 직렬 또는 다중복합 표적화를 위해 본 명세서에 정의된 바와 같은 Cas12b 효소, 복합체 또는 시스템의 하나 이상의 요소를 이용하는 방법을 제공하되, 상기 CRISP 시스템은 다중 가이드 RNA 서열을 포함한다. 바람직하게는, 상기 gRNA 서열은 뉴클레오티드 서열, 예컨대 본 명세서의 다른 곳에 정의된 바와 같은 직접 반복부에 의해 분리된다.

[0267] 본 명세서에 정의된 바와 같은 Cas12b 효소, 시스템 또는 복합체는 다중 표적 폴리뉴클레오티드를 변형시키기 위한 효과적인 수단을 제공한다. 본 명세서에 정의된 바와 같은 Cas12b 효소, 시스템 또는 복합체는 세포 유형의 다중도에서 하나 이상의 표적 폴리뉴클레오티드를 변형시키는 것(예를 들어, 결실, 삽입, 전좌, 비활성화, 활성화)을 포함하는 매우 다양한 유용성을 가진다. 이렇게 해서, 본 발명의 본 명세서에 정의된 바와 같은 Cas12b 효소, 시스템 또는 복합체는 단일 CRISPR 시스템 내의 다중 유전자 좌위 표적화를 포함하는 광범위한 적용, 예를 들어, 유전자 요법, 약물 스크리닝, 질환 진단, 및 예후를 가진다.

[0268] 일 양상에서, 본 발명은 본 명세서에 정의된 바와 같은 Cas12b 효소, 시스템 또는 복합체, 즉, 이와 연관된 적어도 하나의 탈안정화 도메인을 갖는 Cas12b 단백질을 갖는 Cas12b CRISPR-Cas 복합체, 및 다중 핵산 분자, 예컨대 DNA 분자를 표적화하는 다중 가이드 RNA를 제공하고, 이에 의해 상기 다중 가이드 RNA의 각각은 그의 대응하는 핵산 분자, 예를 들어, DNA 분자를 특이적으로 표적화한다. 각각의 핵산 분자 표적, 예를 들어, DNA 분자는 유전자 산물을 암호화하거나 또는 유전자 좌위를 포함할 수 있다. 따라서 다중 가이드 RNA를 이용하는 것은 다중 유전자 좌위 또는 다중 유전자의 표적화를 가능하게 한다. 일부 구현예에서 Cas12b 효소는 그 유전자 생성물을 코딩하는 DNA 분자를 절단할 수 있다. 일부 구현예에서, 유전자 산물의 발현은 변경된다. Cas12b 단백질 및 가이드 RNA는 자연적으로 함께 생기지 않는다. 본 발명은 직렬 배열된 가이드 서열을 포함하는 가이드 RNA를 이해한다. 본 발명은 진핵생물 세포에서 발현을 위해 코돈 최적화된 Cas12b 단백질에 대한 암호 서열을 이해한다. 바람직한 구현예에서, 진핵생물 세포는 포유류 세포, 식물 세포 또는 효모 세포이고, 더 바람직한 구현예에서, 포유류 세포는 인간 세포이다. 유전자 산물의 발현은 감소될 수 있다. Cas12b 효소는 일련의 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 25, 25, 30 또는 30개 초과의 가이드 서열을 포함하는 직렬 배열된 가이드 RNA(gRNA)를 추가로 포함하는 CRISPR 시스템 또는 복합체의 일부를 형성할 수 있으며, 각각 세포 내 관심 대상의 게놈 유전자좌에서 표적 서열에 특이적으로 하이브리드화할 수 있다. 일부 구현예에서, 기능성 Cas12b CRISPR 시스템 또는 복합체는 다중 표적 서열에 결합한다. 일부 구현예에서, 기능성 CRISPR 시스템 또는 복합체는 다중 표적 서열을 편집할 수 있으며, 예를 들어, 표적 서열은 게놈 유전자좌를 포함할 수 있고, 일부 구현예에서, 유전자 발현의 변경이 있을 수 있다. 일부 구현예에서, 기능성 CRISPR 시스템 또는 복합체는 추가적인 기능성 도메인을 포함할 수 있다. 일부 구현예에서, 본 발명은 다중 유전자

산물을 변경시키거나 또는 변형시키는 방법을 제공한다. 방법은 상기 표적 핵산, 예를 들어 DNA 분자를 함유하거나, 또는 표적 핵산, 예를 들어 DNA 분자를 함유하고 발현하는 세포로 도입시키는 단계를 포함할 수 있고; 예를 들어, 표적 핵산은 유전자 산물을 코딩할 수 있거나 또는 유전자 산물의 발현을 위해 제공될 수 있다 (예를 들어, 조절 서열).

[0269] 바람직한 구현예에서, 다중복합 표적화를 위해 사용되는 CRISPR 효소는 Cas12b이거나, 또는 CRISPR 시스템 또는 복합체는 Cas12b를 포함한다. 일부 구현예에서, 다중복합 표적화에 사용된 Cas12b 효소는 DNA의 가닥 모두를 절단하여 이중 가닥 파손(DSB)을 생성한다. 일부 구현예에서, 다중복합 표적화를 위해 사용되는 CRISPR 효소는 닉카제이다. 일부 구현예에서, 멀티플렉스 표적화에 사용된 Cas12b 효소는 이중 닉카제이다. 일부 구현예에서, 다중복합 표적화를 위해 사용되는 Cas12b 효소는 본 명세서의 다른 곳에 정의된 바와 같은 DD Cas12b 효소와 같은 Cas12b 효소이다.

[0270] 일부 일반적 구현예에서, 다중복합 표적화를 위해 사용되는 Cas12b 효소는 하나 이상의 기능성 도메인과 결합된다. 일부 더 구체적인 구현예에서, 다중복합 표적화를 위해 사용되는 CRISPR 효소는 본 명세서의 다른 곳에 정의된 바와 같은 데드Cas12b이다.

[0271] 양상에서, 본 발명은 본 명세서에 정의된 바와 같은 다중 표적화 또는 본 명세서에 정의된 폴리뉴클레오티드에서 사용하기 위한 Cas12b 효소, 시스템 또는 복합체를 전달하기 위한 수단을 제공한다. 이러한 전달 수단의 비제한적 예는, 예를 들어, 복합체의 성분(들)을 전달하는 입자(들), 본 명세서에 논의된 폴리뉴클레오티드(들)를 포함하는(예를 들어, CRISPR 효소를 암호화하고, CRISPR 복합체를 코딩하는 뉴클레오티드를 제공하는) 벡터(들)이다. 일부 구현예에서, 벡터는 플라스미드 또는 바이러스 벡터, 예컨대 AAV 또는 렌티바이러스일 수 있다. 플라스미드에 의한, 예를 들어, HEK 세포 내로의 일시적 형질감염은 유리하며, 특히 AAV의 크기 제한을 제공할 수 있는 반면, Cas12b는 AAV에 꼭 들어맞고, 추가적인 가이드 RNA에 의해 상한에 도달할 수 있다.

[0272] 또한 다중복합 표적화에서 사용하기 위한 본 명세서에 사용된 바와 같은 Cas12b 효소, 복합체 또는 시스템을 구성적으로 발현시키는 모델이 제공된다. 유기체는 유전자이식일 수 있고, 본 벡터로 형질감염될 수 있거나 또는 이렇게 형질감염된 유기체의 자손일 수 있다. 추가적인 양상에서, 본 발명은 본 명세서에 정의된 바와 같은 CRISPR 효소, 시스템 및 복합체 또는 본 명세서에 기재된 폴리뉴클레오티드 또는 벡터를 포함하는 조성물을 제공한다. 또한 Cas12b CRISPR 시스템 또는 다중 가이드 RNA를 포함하는 복합체가, 바람직하게는 직렬 배열된 형식으로 제공된다. 상기 상이한 가이드 RNA는 뉴클레오티드 서열, 예컨대 직접 반복부에 의해 분리될 수 있다.

[0273] 대상체, 예를 들어 치료를 필요로 하는 대상체를 치료하는 방법으로서, 이는 대상체를 Cas12b CRISPR 시스템 또는 복합체를 코딩하는 폴리뉴클레오티드 또는 본 명세서에 설명된 임의의 폴리뉴클레오티드 또는 벡터로 형질전환시킴으로써 유전자 편집을 유도하고, 그를 대상체로 투여하는 것을 포함하는 방법이 또한 제공된다. 적합한 복구 주형이 또한 제공될 수 있으며, 예를 들어, 상기 복구 주형을 포함하는 벡터에 의해 전달된다. 또한 대상체를 본 명세서에 기재된 폴리뉴클레오티드 또는 벡터로 형질전환시킴으로써 다중 표적 유전자 좌위의 전사 활성화 또는 억제를 유도하는 단계를 포함하는, 대상체, 예를 들어, 치료가 필요한 대상체의 치료 방법이 제공되되, 상기 폴리뉴클레오티드 또는 벡터는, 바람직하게는 직렬 배열된 다중 가이드 RNA를 포함하는 Cas12b 효소, 복합체 또는 시스템을 암호화하거나 또는 포함한다. 임의의 치료가 생체 외, 예를 들어 세포 배양물에서 일어나는 경우, 이는 용어 '대상체'가 '세포 또는 세포 배양물'이라는 구절로 대체될 수 있음을 이해해야 할 것이다.

[0274] 본 명세서의 다른 곳에 정의된 바와 같은 치료 방법에서 사용하기 위한, 바람직하게는 직렬 배열된 다중 가이드 RNA를 포함하는 Cas12b 효소, 복합체 또는 시스템을 포함하는 조성물, 또는 바람직하게는 직렬 배열된 다중 가이드 RNA를 포함하는 상기 Cas12b 효소, 복합체 또는 시스템을 암호화하거나 또는 포함하는 폴리뉴클레오티드 또는 벡터가 또한 제공된다. 이러한 조성물을 포함하는 부분의 키트가 제공될 수 있다. 이러한 치료 방법을 위한 의약의 제조에서 상기 조성물의 용도가 또한 제공된다. 스크리닝에서 Cas12b CRISPR 시스템의 용도, 예를 들어, 기능 획득 스크리닝이 또한 본 발명에 의해 제공된다. 유전자를 과발현시키도록 인공적으로 힘이 가해진 세포는, 예를 들어, 음성 피드백 루프에 의해 시간에 따라 유전자를 하향조절할 수 있다(평형상태 재확립). 시간에 따라, 비조절 유전자가 다시 감소될 수 있도록 스크리닝을 시작한다. 유도성 Cas12b 활성 인자를 이용하는 것은 스크리닝 바로 전에 전사를 유도하도록 하며, 따라서, 위음성 히트(false negative hit)의 기회를 최소화한다. 따라서, 스크리닝, 예를 들어, 기능 획득 스크리닝에서 본 발명의 사용에 의해, 위

음성 기회 결과는 최소화될 수 있다.

[0275] 일 양상에서, 본 발명은 Cas12b 단백질 및 각각이 세포 내 유전자 산물을 코딩하는 DNA 분자를 특이적으로 표적화하는 다중 가이드 RNA를 포함하는 조작된, 비천연 발생 CRISPR 시스템을 제공하고, 이에 의해 다중 가이드 RNA는 각각이 유전자 산물을 코딩하는 그들의 특이적 DNA 분자를 표적화하고 Cas12b 단백질은 유전자 산물을 코딩하는 표적 DNA 분자를 절단하여서, 유전자 산물의 발현이 변경되고, CRISPR 단백질 및 가이드 RNA는 함께 천연적으로 발생되지 않는다. 본 발명은 바람직하게는 뉴클레오티드 서열, 예컨대 직접 반복부에 의해 분리되고 선택적으로 tracr 서열에 융합된 다중 가이드 서열을 포함하는 다중 가이드 RNA를 이해한다. 본 발명의 구현예에서, CRISPR 단백질은 V형 또는 VI CRISPR-Cas 단백질이고, 더 바람직한 구현예에서 CRISPR 단백질은 Cas12b 단백질이다. 본 발명은 추가로 진핵생물 세포에서 발현을 위해 코돈 최적화된 Cas12b 단백질을 이해한다. 바람직한 구현예에서, 진핵생물 세포는 포유류 세포이고, 더 바람직한 구현예에서, 포유류 세포는 인간 세포이다. 본 발명의 추가적인 구현예에서, 유전자 산물의 발현은 감소된다.

[0276] **표적 서열의 변형**

[0277] 일정 구현예에서, 관심 유전자좌는 주형 DNA 서열을 삽입, 또는 " 녹-인" 시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 특정 구현예에서, DNA 삽입은 적절한 배향으로 게놈에 통합되도록 설계된다. 바람직한 구현예에서, 관심 유전자좌는 비분열 세포에서 CRISPR-C2c1 시스템에 의해 변형되고, 여기서 상동성 지정 복구 (HDR) 기전을 통한 게놈 편집이 특히 도전적이다 (Chan et al., Nucleic acids research. 2011;39:5955-5966). Maresca 등 (Genome Res. 2013 Mar; 23(3): 539-546)은 징크 핑거 뉴클레아제 (ZFN) 및 Tale 뉴클레아제 (TALEN)를 사용해 적용가능한 부위 지정된, 정밀 삽입 방법을 기술하는데, 여기서 5' 오버행을 갖는 짧은, 이중-가닥 DNA가 상보성 말단에 결찰되어, 인간 세포주에서 한정된 유전자좌에서 15-kb 외생성 발현 카세트의 정밀한 삽입을 가능하게 하였다. He 등 (Nucleic Acids Res. 2016 May 19; 44(9))은 described CRISPR/Cas9-induced site-specific knock-in of a 4.6 kb 무프로모터 ires-eGFP 단편을 GAPDH 유전자좌에 CRISPR/Cas9-유도 부위-특이적 녹-인으로 체세포 LO2 세포에서 최대 20% GFP+ 세포를 산출하였고, NHEJ 경로에 의해 매개되는 인간 배아 줄기 세포에서 1.70% GFP+ 세포를 산출하였다고 기술하였으며, 또한 NHEJ-기반 녹-인이 조사된 모든 인간 세포 유형에서 HDR-매개 유전자 표적화에 비해 더 효율적이라고 보고하였다. C2c1이 5' 오버행을 갖는 스태거드 절단을 생성시키기 때문에, 당업자는 본 명세서에 개시된 CRISPR-C2c1 시스템을 사용해 관심 유전자좌에서 외생성 DNA 삽입을 발생시키기 위해서 Meresca 등 및 He 등에 기술된 것과 유사한 방법을 사용할 수 있다.

[0278] 일정 구현예에서, 관심 유전자좌는 PAM 서열의 원위 말단에서 CRISPR-C2c1 시스템에 의해 먼저 변형되고, PAM 서열 근처에서 CRISPR-C2c1 시스템에 의해 더욱 변형되며 HDR을 통해 복구된다. 일정 구현예에서, 관심 유전자좌는 HDR을 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ를 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 바람직한 구현예에서, 외생성 DNA는 3' 및 5' 말단 둘 모두 상에서 단일 가이드 DNA-PAM 서열이 측접된다. 바람직한 구현예에서, 외생성 DNA는 CRISPR-C2c1 절단 이후에 방출된다. 참조: Zhang et al., Genome Biology201718:35; He et al., Nucleic Acids Research, 44: 9, 2016.

[0279] ***주형***

[0280] 일부 구현예에서, 재조합 주형이 또한 제공된다. 재조합 주형은 본 명세서에 기재된 바와 같은 또 다른 벡터의 성분일 수 있으며, 별개의 벡터 내에 포함되거나, 별개의 폴리뉴클레오티드로서 제공될 수 있다. 일부 구현예에서, 재조합 주형은 상동성 재조합에서, 예컨대 핵산-표적화 복합체의 일부로서 핵산-표적화 이펙터 단백질에 의해 닉킹되거나 절단된 표적 서열 내에 또는 그 가까이에서 주형으로서의 역할을 하도록 설계된다. 일부 예에서, 시스템은 재조합 주형을 포함한다. 재조합 주형은 상동성-지정 복구 (HDR)에 의해 삽입될 수 있다.

[0281] 일 구현예에서, 주형 핵산은 표적 위치의 서열을 변경시킨다. 일 구현예에서, 주형 핵산은 변형되거나 비-천연 발생 염기의 표적 핵산 내로의 혼입을 초래한다.

[0282] 주형 서열은 표적 서열과의 파손 매개된 또는 촉매된 재조합 처리될 수 있다. 일 구현예에서, 주형 핵산은 C2c1 매개된 절단 사건에 의해 절단된 표적 서열 상에서의 부위에 상응하는 서열을 포함할 수 있다. 일 구현예에서, 주형 핵산은 제1 C2c1 매개된 사건에서 절단된 표적 서열 상의 제1 부위, 및 제2 C2c1 매개된 사건에

서 절단된 표적 서열 상의 제2 부위, 모두에 상응하는 서열을 포함할 수 있다.

[0283] 특정 구현예에서, 주형 핵산은 번역된 서열의 코딩 서열 내에서의 변경을 초래하는 서열을 포함할 수 있으며, 예를 들어 단백질 생성물 내에서 하나의 아미노산을 또 다른 것으로 치환, 예를 들어 돌연변이체 대립형질을 야생형 대립형질로 형질전환, 야생형 대립형질을 돌연변이체 대립형질로 형질전환, 및/또는 정지 코돈의 도입, 아미노산 잔기의 삽입, 아미노산 잔기의 결실, 또는 논센스 돌연변이를 초래하는 것을 포함할 수 있다. 특정 구현예에서, 주형 핵산은 비-코딩 서열에서의 변경, 예를 들어 엑손에서의 변경 또는 5' 또는 3' 비-번역된 또는 비-전사된 영역을 초래하는 서열을 포함할 수 있다. 이러한 변경은 제어 요소, 예를 들어 프로모터, 인핸서, 및 시스-작용 또는 트랜스-작용 제어 요소에서의 변경을 포함한다.

[0284] 표적 유전자에서 표적 위치와의 상동성을 갖는 주형 핵산은 표적 서열의 구조를 변경시키는 데 사용될 수 있다. 주형 서열은 원하지 않는 구조, 예를 들어 원하지 않는 또는 돌연변이체 뉴클레오티드를 변경시키는 데 사용될 수 있다. 주형 핵산은 통합될 때, 양성 제어 엘리먼트의 활성 감소; 양성 제어 엘리먼트의 활성 증가; 음성 제어 엘리먼트의 활성 감소; 음성 제어 엘리먼트의 활성 증가; 유전자의 발현 감소; 유전자의 발현 증가; 질병 또는 질환에 대한 내성 증가; 바이러스 진입에 대한 내성 증가;돌연변이 교정 또는 유전자 산물의 생물학적 속성을 부여, 증가, 폐기 또는 감소시키는 원치 않는 아미노산 잔기의 변경, 예를 들어 효소의 효소 활성 증가, 또는 다른 분자와 상호작용하는 유전자 산물의 능력 증가를 야기시키는 서열을 포함할 수 있다.

[0285] 주형 핵산은 표적 서열의 1 개, 2 개, 3 개, 4 개, 5 개, 6 개, 7 개, 8 개, 9 개, 10 개, 11 개, 12 개 이상의 뉴클레오티드 서열에서의 변화를 초래하는 서열을 포함할 수 있다.

[0286] 주형 폴리뉴클레오티드는 임의의 적합한 길이를 가질 수 있으며, 예컨대 길이가 약 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000 개 이상의 뉴클레오티드, 또는 이를 초과할 수 있다. 일 구현예에서, 주형 핵산은 길이가 20+/- 10, 30+/- 10, 40+/- 10, 50+/- 10, 60+/- 10, 70+/- 10, 80+/- 10, 90+/- 10, 100+/- 10, 110+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 180+/- 10, 190+/- 10, 200+/- 10, 210+/-10, 또는 220+/- 10 개의 뉴클레오티드일 수 있다. 일 구현예에서, 주형 핵산은 길이가 30+/-20, 40+/-20, 50+/-20, 60+/-20, 70+/-20, 80+/-20, 90+/-20, 100+/-20, 110+/-20, 120+/-20, 130+/-20, 140+/-20, 150+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/-20, 200+/-20, 210+/-20, 또는 220+/-20 개의 뉴클레오티드일 수 있다. 일 구현예에서, 주형 핵산은 길이가 10 내지 1,000, 20 내지 900, 30 내지 800, 40 내지 700, 50 내지 600, 50 내지 500, 50 내지 400, 50 내지 300, 50 내지 200, 또는 50 내지 100 개의 뉴클레오티드이다.

[0287] 일부 구현예에서, 주형 폴리뉴클레오티드는 표적 서열을 포함하는 폴리뉴클레오티드의 일부에 상보적이다. 최적으로 정렬된 경우, 주형 폴리뉴클레오티드는 표적 서열의 하나 이상의 뉴클레오티드(예를 들어, 약 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 개 이상 또는 이를 초과하는 뉴클레오티드)와 중첩될 수 있다. 일부 구현예에서, 표적 서열 및 표적 서열을 포함하는 폴리뉴클레오티드의 경우, 표적 서열로부터 주형 폴리뉴클레오티드의 최근접 뉴클레오티드는 약 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000 개 이상의 뉴클레오티드 내에 속한다.

[0288] 외생의 폴리뉴클레오티드 주형은 통합될 서열(예를 들어, 돌연변이된 유전자)을 포함한다. 통합을 위한 서열은 세포에 대해 내생 또는 외생성 서열일 수 있다. 통합될 서열의 예로는 단백질 또는 비-코딩 RNA(예를 들어, 마이크로RNA)를 코딩하는 폴리뉴클레오티드를 포함한다. 따라서, 통합을 위한 서열은 적절한 제어 서열 또는 서열들에 작동적으로 연결될 수 있다. 대안적으로, 통합될 서열은 조절 작용을 제공할 수 있다.

[0289] 상류 또는 하류 서열은 약 20 bp 내지 약 2500 bp, 예를 들어, 약 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 또는 2500 bp를 포함할 수 있다. 일부 방법에서, 예시적인 상류 또는 하류 서열은 약 200 bp 내지 약 2000 bp, 약 600 bp 내지 약 1000 bp, 또는 더욱 구체적으로는 약 700 bp 내지 약 1000 bp를 갖는다.

[0290] 상류 또는 하류 서열은 약 20 bp 내지 약 2500 bp, 예를 들어, 약 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 또는 2500 bp를 포함할 수 있다. 일부 방법에서, 예시적인 상류 또는 하류 서열은 약 200 bp 내지 약 2000 bp, 약 600 bp 내지 약 1000 bp, 또는 더욱 구체적으로는 약 700 bp 내지 약 1000 bp를 갖는다.

[0291] 일정 구현예에서, 하나 또는 2 개의 상동성 암 (arm)은 단축되어 특정 서열 반복 요소의 포함을 회피할 수 있다. 예를 들어, 5' 상동성 암은 단축되어 서열 반복 요소를 회피할 수 있다. 다른 구현예에서, 3' 상

동성 암은 단축되어 서열 반복 요소를 회피할 수 있다. 일부 구현예에서, 5' 및 3' 개의 상동성 암 모두는 특정 서열 반복 요소 포함을 회피하기 위하여 단축될 수 있다.

[0292] 일부 방법에서, 외생성 폴리뉴클레오티드 주형은 마커를 추가로 포함할 수 있다. 그러한 마커는 표적된 통합을 스크리닝하기 쉽게 만들 수 있다. 적합한 마커의 예로는 제한 부위, 형광 단백질 또는 선택성 마커를 포함한다. 본 발명의 외생성 RNA 주형은 재조합 기술을 이용하여 구축될 수 있다(예를 들어, 문헌 [Sambrook et al., 2001 및 Ausubel et al., 1996] 참조)

[0293] 일정 구현예에서, 돌연변이 교정을 위한 주형 핵산은 단일-가닥 올리고뉴클레오티드로서 사용을 위해 설계될 수 있다. 단일-가닥 올리고뉴클레오티드를 사용할 때, 5' 및 3' 상동성 암은 최대 약 200 염기쌍 (bp) 길이, 예를 들어 적어도 25, 50, 75, 100, 125, 150, 175, 또는 200 bp 길이 범위일 수 있다.

[0294] Suzuki 등은 CRISPR/Cas9 매개 상동성-독립적 표적화 통합을 통한 생체내 게놈 편집을 기술한다 (2016, Nature 540:144-149).

[0295] 따라서, 본 명세서에서 CRISPR 시스템을 언급할 때, 일부 양상 또는 구현예에서, CRISPR 시스템은 (i) CRISPR 단백질 또는 CRISPR 이펙터 단백질을 코딩하는 폴리뉴클레오티드, 및 (ii) CRISPR 복합체를 형성하도록 CRISPR 단백질과 복합체를 형성하고, 표적 서열과 복합체를 형성하도록 조작된 하나 이상의 폴리뉴클레오티드를 포함한다.

[0296] 일부 구현예에서, 치료제는 생체내 또는 생체외에서 진핵생물 세포에 전달 (또는 도포 또는 투여)을 위한다.

[0297] 일부 구현예에서, CRISPR 단백질은 표적 서열의 위치에서 한쪽 또는 양쪽 가닥의 절단을 유도하는 뉴클레아제이거나, 또는 CRISPR 단백질은 표적 서열의 위치에서 절단을 유도하는 닉카제이다.

[0298] 일부 구현예에서, CRISPR 단백질은 CRISPR-Cas 시스템 RNA 폴리뉴클레오티드 서열과 복합체 형성된 C2c1 단백질이고, 여기서 폴리뉴클레오티드 서열은 a) 표적 HBV 서열과 하이브리드화할 수 있는 가이드 RNA 폴리뉴클레오티드; 및 (b) 직접 반복부 RNA 폴리뉴클레오티드를 포함한다.

[0299] 일부 구현예에서, CRISPR 단백질은 C2c1이고, 시스템은 I. CRISPR-Cas 시스템 RNA 폴리뉴클레오티드 서열로서, (a) 표적 서열과 하이브리드화할 수 있는 가이드 RNA 폴리뉴클레오티드, 및 (b) 직접 반복부 RNA 폴리뉴클레오티드를 포함하는 것인 폴리뉴클레오티드 서열, 및 II. 임의로 적어도 하나 이상의 핵 국재화 서열을 포함하는, C2c1을 코딩하는 폴리뉴클레오티드 서열을 포함하고, 여기서 직접 반복부 서열은 가이드 서열과 하이브리드화하여 표적 서열로 CRISPR 복합체의 서열-특이적 결합을 유도시키고, CRISPR 복합체는 (1) 표적 서열과 하이브리드화하거나 또는 하이브리드화가능한 가이드 서열, 및 (2) 직접 반복부 서열과 복합체를 형성하는 CRISPR 단백질을 포함하고, CRISPR 단백질을 코딩하는 폴리뉴클레오티드 서열은 DNA 또는 RNA이다.

[0300] 본 발명은 또한 임의의 본 명세서에 기술된 조작된 CRISPR 효소 (예를 들어, 조작된 Cas 이펙터 모듈), 조성물, 또는 임의의 본 명세서에 기술된 시스템 또는 벡터 시스템과 세포를 접촉시키는 단계를 포함하는 세포에서 관심 유전자좌를 변형시키는 방법을 제공하거나, 또는 세포는 세포 내에 존재하는 임의의 본 명세서에 기술된 CRISPR 복합체를 포함한다. 이러한 방법에서 세포는 원핵생물 또는 진핵생물 세포, 바람직하게 진핵생물 세포일 수 있다. 이러한 방법에서, 유기체는 세포를 포함할 수 있다. 이러한 방법에서, 유기체는 인간 또는 다른 동물이 아닐 수 있다. 일정 구현예에서, 세포는 A/T 풍부 게놈을 포함할 수 있다. 일부 구현예에서, 세포 게놈은 T-풍부 PAM을 포함한다. 특정 구현예에서, PAM은 5'-TTN-3' 또는 5'-ATTN-3' 이다. 특정 구현예에서, PAM은 5'-TTG-3' 이다. 특정 구현예에서, 세포는 플라스모듐 팔시파럼 (Plasmodium falciparum) 세포이다.

[0301] 일부 구현예에서, CRISPR 이펙터 단백질은 C2c1 단백질이다. C2c1은 Cas9에 의해 생성되는 PAM의 근위 말단에서의 절단과 대조적으로, PAM의 원위 말단에서 이중 가닥 파손을 생성시킨다 (Jinek et al., 2012; Cong et al., 2013). Cpf1 돌연변이된 표적 서열은 단일 gRNA에 의한 반복된 절단에 감수성일 수 있어서, HDR 매개 게놈 편집에서 Cpf1의 적용을 촉진한다고 제안된다 (Front Plant Sci. 2016 Nov 14;7:1683). Cpf1 및 C2c1은 둘 모두가 구조 유사성을 공유하는 V형 CRISPR-Cas 단백질이다. PAM의 원위 말단에서 블런트 (blunt) 절단부를 생성시키는 Cas9와 달리, Cpf1 및 C2c1은 PAM의 원위 말단에서 스태거드 절단부를 생성시킨다. 따라서, 일정 구현에에서, 관심 유전자좌는 상동성 지정 복구 (HR 또는 HDR)을 통해 CRISPR-C2c1 복합체에 의해 변형된다.
일정 구현예에서, 관심 유전자좌는 HR 독립적인 CRISPR-C2c1 복합체에 의해 변형된다. 일정 구현에에서, 관심 유전자좌는 비상동성 말단 결합 (NHEJ)을 통해 CRISPR-C2c1 복합체에 의해 변형된다.

[0302] C2c1는 Cas9에 의해 생성되는 블런트 말단부와 대조적으로, 5' 오버행을 갖는 스태거드 절단부를 생성시킨다 (Garneau et al., Nature. 2010;468:67-71; Gasiunas et al., Proc Natl Acad Sci U S A. 2012;109:E2579-2586). 절단 산물의 이러한 구조는 포유동물 게놈으로 비상동성 말단 결합 (NHEJ)-기반 유전자 삽입을 촉진하기 위해 특이 유리할 수 있다 (Maresca et al., Genome research. 2013;23:539-546).

[0303] 일정 구현예에서, 관심 유전자좌는 주형 DNA 서열을 삽입, 또는 " 녹-인" 시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 특정 구현예에서, DNA 삽입은 적절한 배향으로 게놈에 통합되도록 설계된다. 바람직한 구현예에서, 관심 유전자좌는 비분열 세포에서 CRISPR-C2c1 시스템에 의해 변형되고, 여기서는 상동성-지정 복구 (HDR) 기전을 통한 게놈 편집이 특히 도전이다 (Chan et al., Nucleic acids research. 2011;39:5955-5966). Maresca 등 (Genome Res. 2013 Mar; 23(3): 539-546)은 징크 핑거 뉴클레아제 (ZFN) 및 Tale 뉴클레아제 (TALEN)를 사용해 적용가능한 부위 지정된, 정밀 삽입 방법을 기술하는데, 여기서 5' 오버행을 갖는 짧은, 이중-가닥 DNA가 상보성 말단에 결찰되어, 인간 세포주에서 한정된 유전자좌에서 15-kb 외생성 발현 카세트의 정밀한 삽입을 가능하게 하였다. He 등 (Nucleic Acids Res. 2016 May 19; 44(9))은 4.6 kb 무프로모터 ires-eGFP 단편을 GAPDH 유전자좌에 CRISPR/Cas9-유도 부위-특이적 녹-인으로 체세포 LO2 세포에서 최대 20% GFP+ 세포를 산출하였고, NHEJ 경로에 의해 매개되는 인간 배아 줄기 세포에서 1.70% GFP+ 세포를 산출하였다고 기술하였으며, 또한 NHEJ-기반 녹-인이 조사된 모든 인간 세포 유형에서 HDR-매개 유전자 표적화에 비해 더 효율적이라고 보고하였다. C2c1이 5' 오버행을 갖는 스태거드 절단을 생성시키기 때문에, 당업자는 본 명세서에 개시된 CRISPR-C2c1 시스템을 사용해 관심 유전자좌에서 외생성 DNA 삽입을 발생시키기 위해서 Meresca 등 및 He 등에 기술된 거소가 유사한 방법을 사용할 수 있다.

[0304] 일정 구현예에서, 관심 유전자좌는 PAM 서열의 원위 말단에서 CRISPR-C2c1 시스템에 의해 먼저 변형되고, PAM 서열 근처에서 CRISPR-C2c1 시스템에 의해 더욱 변형되며 HDR을 통해 복구된다. 일정 구현예에서, 관심 유전자좌는 HDR을 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ를 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 바람직한 구현예에서, 외생성 DNA는 3' 및 5' 말단 둘 모두 상에서 단일 가이드 DNA(sgDNA)-PAM 서열이 측접된다. 바람직한 구현예에서, 외생성 DNA는 CRISPR-C2c1 절단 이후에 방출된다. 참조: Zhang et al., Genome Biology201718:35; He et al., Nucleic Acids Research, 44: 9, 2016.

[0305] 일부 구현예에서, CRISPR 단백질은 알리시클로바실러스 악시도테레스트리스 (Alicyclobacillus acidoterrestris) ATCC 49025 또는 바실러스 써모아밀로보란스 (Bacillus thermoamylovorans) 균주 B4166 유래의 C2c1이다.

[0306] 본 발명은 또한 본 명세서에 기재된 임의의 방법 또는 조성물에서 진핵생물 또는 진핵생물 세포에서의 발현을 위해 코돈 최적화된 이펙터 단백질을 코딩하는 뉴클레오티드 서열을 제공한다. 본 발명의 구현예에서, 코돈 최적화된 이펙터 단백질은 본 명세서에 기술된 임의의 C2c1이고 진핵생물 세포 또는 유기체, 예를 들어 본 명세서의 다른 곳에서 언급된 이러한 세포 또는 유기체, 예를 들어, 제한없이, 마우스 세포, 래트 세포, 및 인간 세포 또는 인간이외 진핵생물 유기체, 예를 들어 식물을 포함하는, 효모 세포, 또는 포유동물 세포 또는 유기체에서 작동능을 위해 코돈 최적화된다.

[0307] 일부 구현예에서, CRISPR 단백질은 유기체의 세포의 핵에서 검출가능한 양으로 CRISPR 단백질의 축적을 구동시킬 수 있는 하나 이상의 핵 국재화 신호 (NLS)를 더 포함한다.

[0308] 본 발명의 일정 구현예에서, 적어도 하나의 핵 국재화 신호(nuclear localization signal; NLS)가 C2c1 이펙터 단백질을 코딩하는 핵산 서열에 부착된다. 바람직한 구현예에서, 적어도 하나 이상의 C-말단 또는 N-말단 NLS 가 부착된다 (그리고 이에 따라 C2c1 이펙터 단백질을 코딩하는 핵산 분자(들)는, 발현된 산물이 부착된 또는 연결된 NLS(들)를 갖도록 NLS(들)에 대한 코딩을 포함할 수 있다). 바람직한 구현예에서, C-말단 NLS는 진핵생물 세포, 바람직하게는 인간 세포에서 최적 발현 및 핵 표적화를 위해 부착된다. 바람직한 구현예에서, 코돈 최적화된 이펙터 단백질은 C2c1 이고, 가이드 RNA 의 스페이서 길이는 15 내지 35 nt 이다. 특정 구현예에서, 가이드 RNA의 스페이서 길이는 적어도 16 개의 뉴클레오티드, 예컨대 적어도 17 개의 뉴클레오티드이다. 일정 구현예에서, 스페이서 길이는 15 내지 17 nt, 17 내지 20 nt, 20 내지 24 nt, 예를 들어 20, 21, 22, 23, 또는 24 nt, 23 내지 25 nt, 예를 들어, 23, 24, 또는 25 nt, 24 내지 27 nt, 27-30 nt, 30-35 nt, 또는 35 nt 또는 그 이상이다. 본 발명의 일정 구현예에서, 코돈 최적화된 이펙터 단백질은 C2c1 이고, 가이드 RNA 의 직접 반복부 길이는 적어도 16 뉴클레오티드이다. 일정 구현예에서, 코돈 최적화된 이

펙터 단백질은 C2c1 이고, 가이드 RNA 의 직접 반복부 길이는 16 내지 20 nt, 예를 들어, 16, 17, 18, 19, 또는 20 개 뉴클레오티드이다. 일정 바람직한 구현예에서, 가이드 RNA 의 직접 반복부 길이는 19 개 뉴클레오티드이다.

[0309] 일부 구현예에서, CRISPR 단백질은 하나 이상의 돌연변이를 포함한다.

[0310] 일부 구현예에서, CRISPR 단백질은 촉매 도메인에 하나 이상의 돌연변이를 가지며, 단백질은 하나 이상의 기능성 도메인을 더 포함한다.

[0311] 일부 구현예에서, CRISPR 시스템은 전달 시스템 내에, 임의로, 하나 이상의 벡터를 포함하는 벡터 시스템을 포함하고, 임의로 벡터는 하나 이상의 바이러스 벡터를 포함하며, 임의로 하나 이상의 바이러스 벡터는 하나 이상의 렌티바이러스, 아데노바이러스, 또는 아데노-연합 바이러스 (AAV) 벡터, 또는 입자 또는 지질 입자를 포함하고, 임의로 CRISPR 단백질은 폴리뉴클레오티드와 복합체를 형성하여 CRISPR 복합체를 형성한다.

[0312] 일부 구현예에서, 시스템, 복합체, 또는 단백질은 관심 게놈 유전자좌의 표적 서열의 조작에 의해 유기체 또는 비인간 유기체를 변형시키는 방법에서 사용을 위한다.

[0313] 일부 구현예에서, CRISPR 시스템을 코딩하거나 또는 제공하는 서열을 코딩하는 폴리뉴클레오티드는 리포솜, 입자, 세포 침투 펩티드, 엑소솜, 미세소포, 또는 유전자-총을 통해 전달된다. 일부 구현예에서, 전달 시스템이 포함된다. 일부 구현예에서, CRISPR 시스템은 조작된 폴리뉴클레오티드 및 CRISPR 단백질을 코딩하는 폴리뉴클레오티드를 포함하는 하나 이상의 벡터를 포함하는 벡터 시스템을 포함하고, 임의로 벡터는 하나 이상의 바이러스 벡터를 포함하며, 임의로 하나 이상의 바이러스 벡터는 CRISPR 시스템 또는 CRISPR 복합체를 함유하는, 하나 이상의 렌티바이러스, 아데노바이러스, 또는 아데노-연합 바이러스 (AAV) 벡터, 또는 입자 또는 지질 입자를 포함한다.

[0314] 일부 구현예에서, 재조합/복구 주형이 제공된다.

[0315] 본 명세서에 기재된 바와 같은 본 발명에 따른 방법은 본 명세서에 논의되는 바와 같은 벡터에 세포를 전달하는 단계를 포함하는 본 명세서에 논의된 바와 같은 진핵생물 세포에서(시험관내, 즉, 단리된 진핵생물 세포에서) 하나 이상의 돌연변이를 유도하는 것이 이해된다. 돌연변이(들)는 가이드(들) RNA(들) 또는 sgRNA(들)를 통해 각각의 표적 서열에서 하나 이상의 뉴클레오티드의 도입, 결실 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들)를 통해 상기 세포(들)의 각각의 표적 서열에서 1 내지 75개의 뉴클레오티드의 도입, 결실 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들)를 통해 상기 세포(들)의 각각의 표적 서열에서 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50 또는 75개의 뉴클레오티드의 도입, 결실 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들)를 통해 상기 세포(들)의 각각의 표적 서열에서 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50 또는 75개의 뉴클레오티드의 도입, 결실 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들)를 통해 상기 세포(들)의 각각의 표적 서열에서 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50 또는 75개의 뉴클레오티드의 도입, 결실 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들)를 통해 상기 세포(들)의 각각의 표적 서열에서 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50 또는 75개의 뉴클레오티드의 도입, 결실 또는 치환을 포함할 수 있다. 돌연변이는 가이드(들) RNA(들) 또는 sgRNA(들)를 통해 상기 세포(들)의 각각의 표적 서열에서 40, 45, 50, 75, 100, 200, 300, 400 또는 500개의 뉴클레오티드의 도입, 결실 또는 치환을 포함할 수 있다.

[0316] 독성 및 오프-표적 효과의 최소화를 위해, 전달된 Cas mRNA 및 가이드 RNA의 농도를 제어하는 것이 중요할 수 있다. Cas mRNA 및 가이드 RNA의 최적의 농도는 세포의 또는 비인간 진핵동물 모델에서 상이한 농도를 시험함으로써 그리고 잠재적 오프-표적 게놈 유전자좌에서 변형 정도를 분석하는 심층 서열분석을 이용함으로써 결정될 수 있다. 대안적으로, 독성 및 오프-표적 효과 수준을 최소화하기 위해, Cas 닉카제 mRNA(예를 들어, D10A 돌연변이를 갖는 스트렙토코커스 피오게네스(S. pyogenes) Cas9)는 관심 대상 부위를 표적화하는 가이드 RNAn 쌍으로 전달될 수 있다. 독성 및 표적외 효과를 최소화하기 위한 가이드 서열 및 전략 표적은 WO 2014/093622 (PCT/US2013/074667)에 기술된 바와 같이, 또는 본 명세서에서와 같은 돌연변이를 통해서일 수 있다.

[0317] 전형적으로, 내생성 CRISPR 시스템과 관련하여, CRISPR 복합체(표적 서열에 하이브리드화되고 하나 이상의 Cas

단백질과 복합체화된 가이드 서열을 포함)의 형성은 표적 서열에서 또는 근처에서 (예를 들어, 표적 서열로부터 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 이상의 염기 내에서) 가닥 중 하나 또는 둘 다의 절단을 초래한다. 이론에 의해 구속되는 일 없이, 야생형 tracr 서열의 모두 또는 일부(예를 들어, 야생형 tracr 서열의 약 20, 26, 32, 45, 48, 54, 63, 67, 85개 이상의 뉴클레오티드)를 포함하거나 또는 이들로 이루어질 수 있는 tracr 서열은 또한, 예컨대 tracr 서열의 적어도 일부를 따라서 가이드 서열에 작동적으로 연결된 tracr 메이트 서열의 모두 또는 일부에 대한 하이브리드화에 의해 CRISPR 복합체의 부분을 형성할 수 있다.

[0318] **조작된 CRISPR-Cas 시스템**

[0319] 일반적으로, SPIDR(스페이서 산재된 직접 반복부)로도 알려진 CRISPR(클러스터링되고 규칙적으로 산재된 짧은 팔린드로믹 반복부)은 통상 특정 박테리아 종에 특이적인 DNA 유전자좌의 패밀리를 구성한다. CRISPR 유전자좌는 대장균에서 인지되는 구별되는 계열의 산재된 짧은 서열 반복부(SSR)(Ishino et al., J. Bacteriol., 169:54295433 [1987]; 및 Nakata et al., J. Bacteriol., 171:35533556 [1989]) 및 관련 유전자를 포함한다. 유사한 산재된 SSR이 할로페락스 메디테라네이 (Haloferax mediterranei), 스트렙토코커스 피오게네스 (Streptococcus pyogenes), 아나바에나 (Anabaena) 및 마이코박테리움 튜베르큘로시스 (Mycobacterium tuberculosis)에서 확인되었다(문헌[Groenen et al., Mol. Infect. Biophys. Microbiol., 17:8593 [1995]] 참조). CRISPR 유전자좌는 통상적으로 SRSR(규칙적으로 산재된 짧은 반복부(short regularly spaced repeat))로 명명된 반복부의 구조가 다른 SSR과 상이하다 (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; 및 Mojica et al., Mol. Microbiol., 36:244-246 [2000]). 일반적으로, 반복부는 실질적으로 고정된 길이를 갖는 독특한 개재 서열에 의해 규칙적으로 산재된 클러스터에 존재하는 짧은 요소이다(상기 문헌 [Mojica et al., [2000]]). 반복 서열이 균주들 간에 고도로 보존되어 있지만, 산재된 반복부의 수와 스페이서 영역의 서열은 통상적으로 균주마다 상이하다 (van Embden et al., J. Bacteriol., 182:23932401 [2000]). CRISPR 유전자좌는 아에로피룸(Aeropyrum), 피로바쿨룸(Pyrobaculum), 술폴로부스(Sulfolobus), 아캐오글로부스(Archaeoglobus), 할로카르쿨라(Halocarcula), 메타노박테리움(Methanobacterium), 메타노코커스(Methanococcus), 메타노사르시나(Methanosarcina), 메타노피러스(Methanopyrus), 피로코커스(Pyrococcus), 피크로필러스(Picrophilus), 써모플라스마(Thermoplasma), 코리네박테리움(Corynebacterium), 마이코박테리움(Mycobacterium), 스트렙토마이세스(Streptomyces), 아퀴펙스(Aquifex), 포르피로모나스(Porphyromonas), 클로로비움(Chlorobium), 써머스(Thermus), 바실러스(Bacillus), 리스테리아(Listeria), 스타필로코커스(Staphylococcus), 클로스트리디움(Clostridium), 써모아나에로박터(Thermoanaerobacter), 마이코플라스마(Mycoplasma), 푸소박테리움(Fusobacterium), 아자쿠스(Azarcus), 크로모박테리움(Chromobacterium), 네이세리아(Neisseria), 니트로소모나스(Nitrosomonas), 데설포비브리오(Desulfovibrio), 게오박터(Geobacter), 믹소코커스(Myxococcus), 캄필로박터, 볼리넬라(Wolinella), 아시네토박터(Acinetobacter), 에르위니아(Erwinia), 에스케리키아, 레지오넬라(Legionella), 메틸로코커스(메틸ococcus), 파스퇴렐라(Pasteurella), 포토박테리움(Photobacterium), 살모넬라(Salmonella), 잔토모나스(Xanthomonas), 예르시니아(Yersinia), 트레포네마(Treponema) 및 써모토가(Thermotoga)를 포함하나 이들에 한정되지 않는 40개 초과의 원핵생물에서 확인되었다 (참조, 예를 들어, Jansen et al., Mol. Microbiol., 43:1565- 1575 [2002]; 및 Mojica et al., [2005]).

[0320] **부차적 활성**

[0321] Cas12 효소는 부차적 활성을 보유할 수 있는데, 즉 일정 환경에서 활성화된 Cas12 효소는 표적 서열과의 결합 이후에 활성인 채로 남아서 비표적 올리고뉴클레오티드를 비특이적으로 계속 절단하게 된다. 이러한 가이드 분자-프로그램된 부차적 절단 활성은 핀독치로서 제공될 수 있는 생체내 프로그램된 세포 사멸 또는 시험관내 비특이적 RNA 분해를 촉발시키는 특이적 표적 올리고뉴클레오티드의 존재를 검출하기 위해 Cas12b 시스템을 사용하는 능력을 제공한다 (Abudayyeh et al. 2016; East-Seletsky et al, 2016).

[0322] RNA-가이드된 C2c1 의 프로그램 가능성, 특이성 및 부차적 활성은 또한 핵산의 비특이적 절단을 위한 이상적인 전환가능 뉴클레아제를 가능하게 한다. 일 구현예에서, C2c1 시스템은 ssDNA 같은, 핵산의 부차적인 비특이적 절단을 제공하고 이용하도록 조작된다. 또 다른 구현예에서, C2c1 시스템은 ssDNA 의 부차적인 비특이적 절단을 제공하고 이용하도록 조작된다. 따라서, 조작된 C2c1 시스템은 핵산 검출 및 전사체 조작, 및 세포 사멸 유도를 위한 플랫폼을 제공한다. C2c1 은 포유동물 전사물 녹다운 및 결합 도구로서 사용하기 위해 개발된다. C2c1 은 서열 특이적 표적화된 DNA 결합에 의해 활성화될 때 RNA 및 ssDNA 의 강건한 부차적인 절단을 가능하게 한다.

[0323] 일정 구현예에서, C2c1은 시험관 내 시스템 또는 세포에서 일시적으로 또는 안정적으로 제공되거나 발현되고,

비-특이적으로 절단성 세포 핵산을 표적화하거나 또는 촉발시킨다. 일 구현예에서, C2c1은 ssDNA, 예를 들어 바이러스 ssDNA를 녹다운시키도록 조작된다. 다른 구현예에서, C2c1은 RNA를 녹 다운시키도록 조작된다. 녹다운이 세포 또는 시험관 내 시스템에 존재하는 표적 DNA에 의존하거나, 시스템 또는 세포에 표적 핵산의 첨가에 의해 촉발되도록 시스템을 고안할 수 있다.

[0324] 한 구현예에서, C2c1 시스템은 비정상 DNA 서열의 존재에 의해 구별될 수 있는 세포의 서브세트에서 RNA 를 비-특이적으로 절단하도록, 예를 들어 비정상 DNA 의 절단이 불완전하거나 비효과적일 수 있도록 조작된다. 하나의 비-제한적인 예에서, 암 세포에 존재하고 세포 형질전환을 유도하는 DNA 전좌가 표적화된다. 염색체 DNA를 겪고 복구되는 세포의 소집단이 생존할 수 있는 반면, 비-특이적 부차적인 리보뉴클레아제 활성은 유리하게는 잠재적 생존자의 세포 사멸을 초래한다.

[0325] 부차적인 활성은 많은 임상 진단에 유용한 SHERLOCK 으로 불리는 매우 민감하고 특이적인 핵산 검출 플랫폼에 최근 활용되었다 (Gootenberg, J. S. et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356, 438- 442 (2017)).

[0326] 본 발명에 따르면, 조작된 C2c1 시스템은 DNA 또는 RNA 엔도뉴클레아제 활성에 최적화되어 있으며 포유동물 세포에서 발현될 수 있고 세포에서 리포터 분자 또는 전사체를 효과적으로 녹다운시키는 것을 목표로 한다.

[0327] 등온 증폭에 의한 조작된 Cas13b의 부차적 효과는 높은 민감성 및 단일-염기 미스매치 특이성을 갖는 신속한 DNA 또는 RNA 검출을 제공하는 CRISPR-기반 진단을 제공한다. C2c1-기반 분자 검출 플랫폼은 바이러스의 특정 균주를 검출하고, 병원성 박테리아, 유전자형 인간 DNA를 구별하고, 무-세포 종양 DNA 돌연변이를 식별하는데 사용된다. 또한, 반응 시약은 저온-사슬 독립 및 장기 저장을 위해 동결건조될 수 있고, 현장 적용을 위해 서류상 쉽게 재구성될 수 있다.

[0328] 휴대용 플랫폼에서 높은 민감성 및 단일-염기 미스매치 특이성을 갖는 핵산을 신속하게 검출하는 능력은 질병 진단 및 모니터링, 역학 및 일반적인 실험실 작업에 도움이 될 수 있다. 핵산을 검출하기 위한 방법이 존재하지만, 민감성, 특이성, 단순성, 비용 및 속도간에 절충점이 있다.

[0329] 미생물 CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) 및 CRISPR-연관 (CRISPR-Cas) 적응 면역 시스템은 CRISPR-기반 진단 (CRISPR-Dx)에 활용할 수 있는 프로그램가능한 엔도뉴클레아제를 포함한다. C2c1 (Cas12b라고도 알려짐)은 특이적 DNA 감지를 위한 플랫폼을 제공하기 위해 CRISPR RNA (crRNAs)로 재프로그램될 수 있다. DNA 표적의 인식에 따라, 활성화된 C2c1은 근처의 비-표적 핵산 (즉, RNA 및/또는 ssDNA)의 " 부차적" 절단에 관여한다. 이러한 crRNA-프로그램된 부차적 절단 활성은 C2c1이 프로그램된 세포 사멸을 유발하거나 표지된 RNA 또는 ssDNA의 비특이적 분해에 의해 생체 내에서 특정 DNA의 존재를 검출할 수 있게 한다. 여기에 표적의 실시간 검출을 가능하게 하는, 상업적인 리포터 RNA의 핵산 증폭 및 C2c1-매개된 부차적 절단에 기초하여 높은 민감성을 갖는 시험관 내 핵산 검출 플랫폼이 기재되어 있다.

[0330] 일정 예의 구현예에서, 본 명세서에 개시된 오솔로그는 진단 조성물 및 어세이에서 단독으로, 또는 다른 Cas12 또는 Cas13 오솔로그와 조합하여 사용될 수 있다. 예를 들어, 본 명세서에 개시된 Cas12b 오솔로그는 표적 서열을 검출하기 위한 다중복합 어세이에서 사용될 수 있고, 올리고뉴클레오티드-기반 리포터의 비특이적 절단을 통해서, 검출가능한 신호를 발생시킨다.

[0331] **리포터/차폐성 구성체**

[0332] 본 명세서에서 사용되는 "차폐성 구성체" 는 본 명세서에 기술된 활성화된 CRISPR 시스템 이펙터 단백질에 의해 절단될 수 있거나 또는 달리 탈활성화될 수 있는 분자를 의미한다. 용어 "차폐성 구성체" 는 또한 "검출 구성체" 로서 대체하여 언급될 수도 있다. CRISPR 이펙터 단백질의 뉴클레아제 활성에 의존하여, 차폐성 구성체는 RNA-기반 차폐성 구성체 또는 DNA-기반 차폐성 구성체일 수 있다. 핵산-기반 차폐성 구성체는 CRISPR 이펙터 단백질에 의해 절단가능한 핵산 엘리먼트를 포함한다. 핵산 엘리먼트의 절단은 작용제를 방출하거나 또는 검출가능한 신호를 생성시킬 수 있게 하는 입체형태 변화를 일으킨다. 핵산 엘리먼트가 어떻게 사용되어서 검출가능한 신호의 발생을 방지 또는 차폐할 수 있는지를 입증하는 예로서의 구성체는 하기에 기술되고 본 발명의 구현예는 이의 변이체를 포함한다. 절단 이전에, 또는 차폐성 구성체가 '활성' 상태일 때, 차폐성 구성체는 양성의 검출가능한 신호의 발생 또는 검출을 차단한다. 일정 예의 구현예에서 최소 배경 신호가 활성 차폐성 구성체의 존재 하에서 생성될 수 있다는 것을 이해할 것이다. 양성 검출가능한 신호는 광학, 형광, 화학발광, 전기화학 또는 다른 당분야에 공지된 검출 방법을 사용해 검출될 수 있는 임의 신호일 수 있다. 용어 "양성의 검출가능한 신호" 는 차폐성 구성체의 존재 하에서 검출가능할 수 있는 다른 검출가능

한 신호와 구별하기 위해 사용된다. 예를 들어, 일정 구현예에서 제1 신호 (즉, 음성의 검출가능한 신호)는 차폐제가 존재할 때 검출될 수 있을 것이고, 이것은 이후 표적 분자의 검출 및 활성화된 CRISPR 이펙터 단백질에 의한 차폐제의 절단 또는 탈활성화 시에 제2 신호 (예를 들어, 양성의 검출가능한 신호)로 전환된다.

[0333] 일정한 예의 구현예에서, 차폐성 구성체는 유전자 생성물의 발생을 억제할 수 있다. 유전자 생성물은 샘플에 첨가되는 리포터 구성체에 의해 코딩될 수 있다. 차폐성 구성체는 RNA 간섭 경로에 관여되는 간섭 RNA, 예컨대 짧은 헤어핀 RNA (shRNA) 또는 소형 간섭 RNA (siRNA)일 수 있다. 차폐성 구성체는 또한 마이크로 RNA (miRNA)를 포함할 수 있다. 존재하지만, 차폐성 구성체는 유전자 생성물의 발현을 억제한다. 유전자 생성물은 형광 단백질 또는 다른 RNA 전사물일 수 있거나 또는 달리 표지된 프로브, 압타머, 또는 항체에 의해 검출가능하지만 차폐성 구성체의 존재를 위한 단백질일 수 있다. 이펙터 단백질의 활성화 시에 차폐성 구성체는 절단되거나 또는 달리 침묵화되어서 양성 검출가능한 신호로서 유전자 생성물의 발현 및 검출이 가능하게 된다.

[0334] 일정한 예의 구현예에서, 차폐성 구성체는 차폐성 구성체로부터 하나 이상의 시약의 방출이 검출가능한 양성 신호의 발생을 일으키도록 검출가능한 양성 신호를 발생시키는데 필요한 하나 이상의 시약을 격리시킬 수 있다. 하나 이상의 시약은 비색 신호, 화학 발광 신호, 형광 신호, 또는 임의의 다른 검출가능한 신호를 생성시키도록 조합될 수 있고 이러한 목적에 적합한 것으로 공지된 임의 시약을 포함할 수 있다. 일정한 예의 구현예에서, 하나 이상의 시약은 하나 이상의 시약에 결합하는 RNA 압타머에 의해 격리된다. 하나 이상의 시약은 표적 분자의 검출 시 이펙터 단백질이 활성화되고 RNA 또는 DNA 압타머가 분해될 때 방출된다.

[0335] 일정한 예의 구현예에서, 차폐성 구성체는 개별 이산 부피 (하기에 더욱 정의됨) 내에 고형 기재 상에 고정화될 수 있고 단일 시약을 격리시킬 수 있다. 예를 들어, 시약은 염료를 포함하는 비드일 수 있다. 고정화 시약에 의해 격리될 때, 개별 비드는 너무 확산되어 검출가능한 신호를 발생시키지 못하지만, 차폐성 구성체로부터 방출 시에 예를 들어 응집에 의해서 또는 용액 농도의 단순 증가에 의해서 검출가능한 신호를 발생시킬 수 있다. 일정 예의 구현예에서, 고정된 차폐제는 표적 분자의 검출 시 활성화된 이펙터 단백질에 의해 절단될 수 있는 RNA- 또는 DNA-기반 압타머이다.

[0336] 일정한 다른 예의 구현예에서, 차폐성 구성체는 용액 중 고정화 시약에 결합하여서 용액 중에 유리된 별개의 표지된 결합 파트너에 결합하는 시약의 능력을 차단한다. 따라서, 샘플에 세척 단계의 적용 시, 표지된 결합 파트너는 표적 분자의 부재 하에서 샘플을 세척해 낼 수 있다. 그러나, 이펙터 단백질이 활성화되면, 차폐성 구성체는 시약에 결합하는 차폐성 구성체의 능력을 방해하기에 충분한 정도로 절단되어서 표지된 결합 파트너가 고정된 시약에 결합할 수 있게 한다. 따라서, 표지된 결합 파트너는 세척 단계 후에 남아서 샘플 중에 표적 분자의 존재를 의미한다. 일정한 양상에서, 고정화된 시약에 결합하는 차폐성 구성체는 DNA 또는 RNA 압타머이다. 고정된 시약은 단백질일 수 있고 표지된 결합 파트너는 표지된 항체일 수 있다. 대안적으로, 고정된 시약은 스트렙타비딘일 수 있고 표지된 결합 파트너는 표지된 바이오틴일 수 있다. 상기 구현예에서 사용되는 결합 파트너 상의 표지는 당분야에 공지된 임의의 검출가능한 표지일 수 있다. 또한, 다른 공지된 결합 파트너가 본 명세서에 기술된 전체 설계에 따라서 사용될 수 있다.

[0337] 일정한 예의 구현예에서, 차폐성 구성체는 리보자임을 포함할 수 있다. 리보자임은 촉매적 특성을 갖는 RNA 분자이다. 천연 및 조작 리보자임은 본 명세서에 개시된 이펙터 단백질에 의해 표적화될 수 있는, RNA를 포함하거나, 또는 그로 이루어진다. 리보자임은 음성 검출가능한 신호를 발생시키거나 또는 양성 대조군 신호의 발생을 방지하는 반응을 촉매하도록 선택될 수 있거나 또는 조작될 수 있다. 활성화된 이펙터 단백질에 의한 리보자임의 탈활성화 시 음성의 대조군 신호를 발생시키거나, 또는 양성의 검출가능한 신호의 발생을 방지하는 반응은 제거되고 그리하여 양성의 검출가능한 신호가 발생될 수 있게 한다. 일례의 구현예에서, 리보자임은 용액이 제1 색상을 나타내게 하는 비색 반응을 촉매할 수 있다. 리보자임이 탈활성화될 때 그러면 용액이 제2 색상으로 변하게 되며, 제2 색상은 검출가능한 양성 신호이다. 리보자임이 비색 반응을 촉매하기 위해 어떻게 사용되는가의 예는 [Zhao et al. "Signal amplification of glucosamine-6-포스페이트 based on ribozyme glmS," Biosens Bioelectron. 2014; 16:337-42]에 기술되어 있고, 이러한 시스템이 본 명세서에 개시된 구현예의 맥락에서 작용하도록 어떻게 변형될 수 있는가의 예를 제공한다. 대안적으로, 리보자임은 존재하는 경우에, 예를 들어 RNA 전사물의 절단 생성물을 발생시킬 수 있다. 따라서, 양성 검출가능한 신호의 검출은 오직 리보자임의 부재 하에서만 발생되는 비절단된 RNA 전사물의 검출을 포함할 수 있다.

[0338] 일정한 예의 구현예에서, 하나 이상의 시약은 단백질이 단백질에 하나 이상의 DNA 또는 RNA 압타머의 결합에 의해 검출가능한 신호를 발생시킬 수 없도록 억제되거나 또는 격리되는, 검출가능한 신호, 예컨대 비색, 화학발광

또는 형광발광 신호의 발생을 촉진할 수 있는, 단백질, 예컨대 효소이다. 본 명세서에 개시된 이펙터 단백질의 활성화 시, DNA 또는 RNA 압타머는 그들이 더 이상 검출가능한 신호를 발생시키는 단백질의 능력을 억제하지 않는 정도까지 절단 또는 분해된다. 일정한 예의 구현예에서, 압타머는 트롬빈 억제제 압타머이다. 일정한 예의 구현예에서, 트롬빈 억제제 압타머는 GGGAACAAAGCUGAAGUACUUACCC (SEQ ID NO: 439)의 서열을 갖는다. 이러한 압타머가 절단될 때, 트롬빈은 활성화될 것이고 펩티드 비색 또는 형광 기질을 절단할 것이다.

일정한 예의 구현예에서, 비색 기질은 트롬빈에 대한 펩티드 기질에 공유적으로 연결된 파라-니트로아닐리드(pNA)이다. 트롬빈에 의해 절단 시, pNA가 방출되고 노란 색상이 되어 쉽게 육안으로 볼 수 있다. 일정 예의 구현예에서, 형광 기질은 7-아미노-4-메틸쿠마린으로서 형광 검출기를 사용해 검출할 수 있는 파란색 형광단이다. 억제성 압타머가 또한 홀스래디쉬 퍼옥시다제 (HRP), 베타-갈락토시다제, 또는 송아지 알칼리 포스파타제 (CAP)에 대해 사용될 수 있고 상기 제시된 일반 원리에 속한다.

[0339] 일정한 구현예에서, RNAse 또는 DNAse 활성은 효소-억제성 압타머의 절단을 통해 비색적으로 검출된다. DNase 또는 RNase 활성을 비색 신호로 전환시키는 하나의 잠재적인 방식은 비색 출력을 생성시킬 수 있는 효소의 재활성화와 DNA 또는 RNA 압타머의 절단을 커플링시키는 것이다. RNA 또는 DNA 절단의 부재 하에서, 온전한 압타머는 효소 표적에 결합하여 이의 활성을 억제하게 될 것이다. 이러한 판독 시스템의 장점은 효소가 추가 증폭 단계를 제공한다는 것이다: 부차적 활성 (예를 들어, C2c1 부차적 활성)을 통해서 압타머로부터 유리되면, 비색 효소는 비색 생성물을 계속 생성시켜서, 신호의 배가를 일으키게 될 것이다.

[0340] 일정한 구현예에서, 비색 판독의 효소를 억제하는 현존 압타머가 사용된다. 비색 판독되는 몇몇 압타머/효소 쌍에는 예컨대 트롬빈, 단백질 C, 호중구 엘라스타제 및 서브스틸리신이 존재한다. 이들 프로테아제는 pNA를 기반으로 비색 기질을 가지며 상업적으로 입수가능하다. 일정한 구현예에서, 일반적인 비색 효소를 표적화하는 신규한 압타머가 사용된다. 일반적인 강건한 효소, 예컨대 베타-갈락토시다제, 홀스래디쉬 퍼옥시다제 또는 송아지 장 알칼리 포스파타제는 선택 전략 예컨대 SELEX에 의해 디자인된 조작된 압타머에 의해 표적화될 수 있다. 이러한 전략은 나노몰 결합 효율로 압타머의 신속한 선택을 가능하게 하고 비색 판독을 위한 추가의 효소/압타머 쌍의 개발에 사용될 수 있다.

[0341] 일정한 구현예에서, RNase 또는 DNase 활성은 RNA-테더드 억제제의 절단을 통해서 비색으로 검출된다. 많은 일반 비색 효소는 경쟁적, 가역적 억제제를 가지며, 예를 들어 베타-갈락토시다제는 갈락토스에 의해 억제될 수 있다. 많은 이들 억제제는 약하지만, 그들의 효과는 국소 농도를 증가시켜서 증가될 수 있다. 억제제의 국소 농도를 DNase 및/또는 RNAse 활성과 연결시킴으로써, 비색 효소 및 억제제 쌍은 DNase 및 RNAse 센서에 조작될 수 있다. 소형-분자 억제제를 기반으로 하는 비색 DNase 또는 RNAse 센서는 3종의 성분을 포함하는데, 비색 효소, 억제제, 및 억제제를 효소에 속박시키는, 억제제와 효소 둘 모두에 공유적으로 연결된 브릿징 RNA 또는 DNA이다. 미절단된 구성에서, 효소는 소형 분자의 증가된 국소 농도에 의해 억제되고, DNA 또는 RNA가 (예를 들어, Cas13 또는 Cas12 부차적 절단에 의해) 절단될 때, 억제제가 방출될 것이고 비색 효소가 활성화될 것이다.

[0342] 일정한 구현예에서, RNAse 또는 Dnase 활성은 G-사중체의 형성 및/또는 활성화를 통해 비색적으로 검출된다. DNA의 G 사중체는 헴 (철 (III)-프로토포르피린 IX)과 복합체를 형성하여 퍼옥시다제 활성과 DNAzyme을 형성할 수 있다. 퍼옥시다제 기질 (예를 들어, ABTS: (2,2'-아지노비스 [3-에틸벤조티아졸린-6-술폰산]-디암모늄 염))이 공급될 때, 과산화수소의 존재 하에서 G-사중체-헴 복합체가 기질의 산화를 야기하고, 그 다음에 용액 중에서 녹색을 형성시킨다. G-사중체 형성 DNA 서열의 예는 GGGTAGGGCGGGTTGGGA (SEQ ID NO: 440)이다.

DNA 압타머와, 본 명세서에서 " 스테이플 (staple)" 이라고 하는, 추가의 DNA 또는 RNA 서열의 하이브리드화에 의해서, G-사중체 구조의 형성은 제한될 것이다. 부차적 활성화 시, 스테이플은 절단되어 G 사중체를 형성할 수 있게 하고 헴이 결합할 수 있게 할 것이다. 이러한 전략은 부차적 활성화 이후에 추가 증폭이 존재한다는 것을 의미하는, 색상 형성이 효소적이기 때문에 특히 매력적이다.

[0343] 일정한 예의 구현예에서, 차폐성 구성체는 개별 이산 부피 (하기에 더욱 정의됨) 내에 고형 기재 상에 고정화될 수 있고 단일 시약을 격리시킬 수 있다. 예를 들어, 시약은 염료를 포함하는 비드일 수 있다. 고정화 시약에 의해 격리될 때, 개별 비드는 너무 확산되어 검출가능한 신호를 발생시키지 못하지만, 차폐성 구성체로부터 방출 시에 예를 들어 응집에 의해서 또는 용액 농도의 단순 증가에 의해서 검출가능한 신호를 발생시킬 수 있다. 일정 예의 구현예에서, 고정된 차폐제는 표적 분자의 검출 시 활성화된 이펙터 단백질에 의해 절단될 수 있는 DNA- 또는 RNA-기반 압타머이다.

[0344] 일례의 구현예에서, 차폐성 구성체는 검출제가 응집되는지 또는 용액에 분산되는지 여부에 따라서 생상을 변화

시키는 검출제를 포함한다. 예를 들어, 일정한 나노입자, 예컨대 콜로이드 금은 그들이 응집물로부터 분산된 입자로 이동하면서 가시적인 보라색에서 붉은 색으로 색상 이동을 겪는다. 따라서, 일정한 예의 구현예에서, 이러한 검출제는 하나 이상의 브릿지 분자에 의해 응집물로 유지될 수 있다. 브릿지 분자의 적어도 일부분은 RNA 또는 DNA를 포함한다. 본 명세서에 개시된 이펙터 단백질의 활성화 시에, 브릿지 분자의 RNA 또는 DNA 부분은 절단되어서 검출제가 분산될 수 있게 하고 색상의 상응하는 변화를 일으킬 수 있다. 일정한 예의 구현예에서, 검출제는 콜로이드 금속이다. 콜로이드 금속 재료는 액체, 히드로졸 또는 금속 졸에 분산된 수불용성 금속 입자 또는 금속 화합물을 포함할 수 있다. 콜로이드 금속은 주기율표의 그룹 IA, IB, IIB 및 IIIB의 금속을 비롯하여, 전이 금속, 특히 그룹 VIII의 것으로부터 선택될 수 있다. 바람직한 금속은 금, 은, 알루미늄, 루테늄, 아연, 철, 니켈 및 칼슘을 포함한다. 다른 적합한 금속은 또한 모든 그들의 다양한 산화 상태의 하기의 것들을 포함한다: 리튬, 소듐, 마그네슘, 포타슘, 스칸듐, 티타늄, 바나듐, 크롬, 망간, 코발트, 구리, 갈륨, 스트론튬, 니오븀, 몰리브데늄, 팔라듐, 인듐, 주석, 텅스텐, 레늄, 플래티늄, 및 가돌리늄. 금속은 바람직하게 적절한 금속 화합물로부터 유래된 이온 형태, 예를 들어 A13+, Ru3+, Zn2+, Fe3+, Ni2+ 및 Ca2+ 이온으로 제공된다.

[0345] RNA 또는 DNA 브릿지가 활성화된 CRISPR 이펙터에 의해 절단될 때, 상기 언급된 색상 이동이 관찰된다. 일정 예의 구현예에서, 입자는 콜로이드 금속이다. 다른 일정 예의 구현예에서, 콜로이드 금속은 콜로이드 금이다. 일정 예의 구현예에서, 콜로이드 나노입자는 15 nm 금 나노입자 (AuNP)이다. 콜로이드 금 나노입자의 고유한 표면 성질 덕분에, 용액에서 완전히 분산되고 육안으로 붉은 색상이 나타날 때 520 nm에서 최대 흡광도가 관찰된다. AuNP의 응집 시, 그들은 최대 흡광도에서 붉은색-이동을 나타내고 색상이 더 진하게 나타나며, 궁극적으로 용액으로부터 진한 보라색 응집체로 침전된다. 일정 예의 구현예에서, 나노입자는 나노입자의 표면으로부터 연장된 DNA 링커를 포함하도록 변형된다. 개별 입자는 각 말단 상에서 DNA 링커의 적어도 일부분에 하이브리드화하는 단일 가닥 RNA (ssRNA) 또는 단일 가닥 DNA 브릿지에 의해 함께 연결된다. 따라서, 나노입자는 연결된 입자의 망을 형성하게 되고 응집하게 되어, 진한 침전물로서 나타나게 될 것이다.

본 명세서에 개시된 CRISPR 이펙터의 활성화 시, ssRNA 또는 ssDNA 브릿지는 절단될 것이고, 연결된 메시로부터 AU NPS를 방출하여 가시적인 붉은 색상을 생성시키게 된다. 예시적인 DNA 링커 및 브릿지 서열은 하기에 열거된다. DNA 링커의 말단 상에 티올 링커는 AuNPS와 표면 접합을 위해 사용될 수 있다. 다른 형태의 접합이 사용될 수도 있다. 일정 예의 구현예에서, 각 DNA 링커에 대해 하나씩, 2개 집단의 AuNP가 발생될 수 있다. 이것은 적절한 배향으로 ssRNA 가교의 적절한 결합을 촉진하도록 돕게 될 것이다. 일정 예의 구현예에서, 제1 DNA 링커는 3' 말단으로 접합되는 반면 제2 DNA 링커는 5' 말단으로 접합된다.

[0346]

표 5: DNA 링커 및 브릿지 서열	
C2c2 c 비색 DNA1	TTATAACTATTCCTAAAAAAAAAAA/3ThioMC3-D/ (SEQ ID NO:441)
C2c2 c 비색 DNA2	/5ThioMC6-D/AAAAAAAAAACTCCCCTAATAACAAT (SEQ ID NO:442)
C2c2 c 비색 브릿지	GGGUAGGAAUAGUUAUAAUUUCCCUUUCCCAUUGUUAUUAGGGAG (SEQ ID NO:443)

[0347] 일정한 다른 예의 구현예에서, 차폐성 구성체는 검출가능한 표지 및 그 검출가능한 표지의 차폐제가 부착되는 RNA 또는 DNA 올리고뉴클레오티드를 포함할 수 있다. 이러한 검출가능한 표지/차폐제 쌍의 예는 형광단 및 형광단의 소광제가 있다. 형광단의 소광은 형광단 및 다른 형광단 또는 비형광 분자 간 비형광성 복합체의 형성 결과로서 일어날 수 있다. 이러한 기전은 바닥-상태 복합체 형성, 정적 소광, 또는 접촉 소광으로서 알려져 있다. 따라서, RNA 또는 DNA 올리고뉴클레오티드는 형광단 및 소광제가 접촉 소광이 일어나도록 충분히 근접하도록 디자인될 수 있다. 형광단 및 그들의 동족 소광제는 당분야에 공지되어 있고, 당업자에 의해서 이러한 목적을 위해 선택될 수 있다. 특정한 형광단/소광제 쌍은 본 발명의 상황에서 핵심적이지 않고, 오직 형광단/소광제 쌍의 선택이 형광단의 차폐를 보장한다. 본 명세서에 개시된 이펙터 단백질의 활성화 시에, RNA 또는 DNA 올리고뉴클레오티드는 절단되고 그리하여 접촉 소광 효과를 유지하는데 필요한 형광단 및 소광제 간 근접성을 잘라낸다. 따라서, 형광단의 검출은 샘플 중 표적 분자의 존재를 결정하는데 사용될 수 있다.

[0348] 일정한 다른 예의 구현예에서, 차폐성 구성체는 하나 이상의 금속 나노입자, 예컨대 금 나노입자가 부착되는 하나 이상의 RNA 올리고뉴클레오티드를 포함할 수 있다. 일부 구현예에서, 차폐성 구성체는 닫힌 루프를 형성

하는 다수의 RNA 또는 DNA 올리고뉴클레오티드에 의해 가교된 다수의 금속 나노입자를 포함한다. 일 구현예에서, 차폐성 구성체는 닫힌 루프를 형성하는 3개의 RNA 또는 DNA 올리고뉴클레오티드에 의해 교차된 3개의 금 나노입자를 포함한다. 일부 구현예에서, CRISPR 이펙터 단백질에 의한 RNA 또는 DNA 올리고뉴클레오티드의 절단은 금속 나노입자에 의해 생성되는 검출가능한 신호를 야기시킨다.

[0349] 일정한 다른 예의 구현예에서, 차폐성 구성체는 하나 이상의 퀀텀 도트가 부착되는 하나 이상의 RNA 또는 DNA 올리고뉴클레오티드를 포함할 수 있다. 일부 구현예에서, CRISPR 이펙터 단백질에 의한 RNA 또는 DNA 올리고뉴클레오티드의 절단은 퀀텀 도트에 의해 생성되는 검출가능한 신호를 야기시킨다.

[0350] 일례의 구현예에서, 차폐성 구성체는 퀀텀 도트를 포함할 수 있다. 퀀텀 도트는 표면에 부착되는 다수의 링커 분자를 가질 수 있다. 링커 분자의 적어도 일부분은 RNA 또는 DNA를 포함한다. 링커 분자는 한쪽 말단에서 퀀텀 도트에 부착되고 링커의 길이를 따라서 또는 말단부에서 하나 이상의 소광제에 부착되어서 소광제가 퀀텀 도트의 소광이 일어나도록 충분히 근접하게 유지된다. 링커는 분지될 수 있다. 상기처럼, 퀀텀 도트/소광제 쌍은 핵심적이지 않고, 오직 퀀텀 도트/소광제 쌍의 선택이 형광단의 차폐를 보장한다. 퀀텀 도트 및 그들의 동족 소광제는 당분야에 공지되어 있고 당업자에 의해서 이러한 목적을 위해 선택될 수 있다. 본 명세서에 개시된 이펙터 단백질의 활성화 시, 링커 분자의 RNA 또는 DNA 부분은 절단되어서 소광 효과를 유지하는데 필요한 퀀텀 도트와 하나 이상의 소광제 간 근접성을 제거시킨다. 일정 예의 구현예에서 퀀텀 도트는 스트렙타비딘 접합된다. RNA 또는 DNA는 바이오틴 링커를 통해서 부착되고 서열 /5Biosg/UCUCGUACGUUC/3IAbRQSp/ (SEQ ID NO: 444) 또는 /5Biosg/UCUCGUACGUUCUCUCGUACGUUC/3IAbRQSp/ (SEQ ID NO: 445)을 갖는 소광 분자를 동원하며, 여기서 /5Biosg/는 바이오틴 태그이고 /3IAbRQSp/는 아이오와 블랙 소광제이다. 절단 시, 본 명세서에 개시된 활성화된 이펙터에 의해서 퀀텀 도트는 가시적으로 형광발광하게 될 것이다.

[0351] 유사한 방식으로, 형광 에너지 전달 (FRET)은 검출가능한 양성 신호를 발생시키기 위해 사용될 수 있다. FRET는 에너지 여기된 형광단 (즉, "도너 형광단")으로부터의 광자가 다른 분자 (즉, "억셉터") 내 전자의 에너지 상태를 더 높은 진동 수준의 여기된 단일항 상태로 상승시키는 비복사 과정이다. 도너 형광단은 그 형광단의 특징적인 형광을 방출하지 않고 바닥 상태로 복귀된다. 억셉터는 다른 형광단일 수 있거나 또는 비형광성 분자일 수 있다. 억셉터가 형광단이면, 전달된 에너지는 그 형광단의 특징적인 형광으로서 발광된다. 억셉터가 비형광성 분자이면 흡수된 에너지는 열로서 소실된다. 따라서, 본 명세서에 개시된 구현예의 상황에서, 형광단/소광제 쌍은 올리고뉴클레오티드 분자에 부착된 도너 형광단/억셉터 쌍으로 교체된다. 온전할 때, 차폐성 구성체는 억셉터로부터 방출되는 열 또는 형광에 의해 검출되는 제1 신호 (음성 검출가능한 신호)를 발생시킨다. 본 명세서에 개시된 이펙터 단백질의 활성화 시 RNA 올리고뉴클레오티드는 절단되고 FRET은 파괴되어서 도너 형광단의 형광이 이제 검출된다 (양성 검출가능한 신호).

[0352] 일정한 예의 구현예에서, 차폐성 구성체는 짧은 뉴클레오티드로 긴 RNA 또는 DNA의 절단에 대응하여 그들 흡광도를 변화시키는 인터컬레이팅 염료의 사용을 포함한다. 몇몇 이러한 염료가 존재한다. 예를 들어, 파이로닌-Y는 RNA와 복합체를 형성하게 될 것이고, 572 nm에서 흡광도를 갖는 복합체를 형성하게 될 것이다. RNA의 절단은 그 결과로 흡광도의 소실 및 색상 변화를 일으킨다. 메틸렌 블루는 유사한 방식으로 사용될 수 있고, RNA 절단 시 688 nm에서의 흡광도가 변화한다. 따라서, 일정 예의 구현예에서, 차폐성 구성체는 본 명세서에 개시된 이펙터 단백질에 의한 RNA의 절단 시 흡광도를 변화시키는 RNA 및 인터컬레이팅 염료 복합체를 포함한다.

[0353] 일정 예의 구현예에서, 차폐성 구성체는 HCR 반응을 위한 개시제를 포함할 수 있다. 참조: Dirks and Pierce. PNAS 101, 15275-15728 (2004). HCR 반응은 2가지 헤어핀 종에서 위치 에너지를 이용한다. 헤어핀 중 하나의 상응하는 영역에 상보성인 부분을 갖는 단일 가닥 개시제가 이전에 안정된 혼합물로 방출될 때, 이것이 한 종의 헤어핀을 개방시킨다. 이어서 이 과정은 다른 종의 헤어핀을 개방시키는 단일 가닥 영역을 노출시킨다. 다음으로 이 과정은 본래 개시제와 동일한 단일 가닥 영역을 노출시킨다. 최종 연쇄 반응은 헤어핀 공급이 고갈될 때까지 성장하는 닉킹된 이중 나선의 형성을 일으킬 수 있다. 최종 생성물의 검출은 겔 상에서 또는 비색적으로 수행될 수 있다. 예로서 비색 검출 방법은 예를 들어, 하기 문헌들에 개시된 것들을 포함한다: Lu et al. " Ultra-sensitive colorimetric assay system based on the hybridization chain reaction-triggered enzyme cascade amplification ACS Appl Mater Interfaces, 2017, 9(1):167-175, Wang et al. " An enzyme-free colorimetric assay using hybridization chain reaction amplification and split aptamers" Analyst 2015, 150, 7657-7662, 및 Song et al. " Non covalent fluorescent labeling of hairpin DNA probe coupled with hybridization chain reaction for sensitive DNA detection." Applied

Spectroscopy, 70(4): 686-694 (2016).

[0354] 일정 예의 구현예에서, 차폐성 구성체는 HCR 개시제 서열 및 개시제가 HCR 반응을 개시시키는 것을 방지하는 절단가능한 구조적 엘리먼트, 예컨대 루프 또는 헤어핀을 포함할 수 있다. 활성화된 CRISPR 이펙터 단백질에 의한 구조적 엘리먼트의 절단 시, 개시제는 방출되어 HCR 반응을 촉발시키고, 이의 검출은 샘플 중 하나 이상의 표적의 존재를 의미한다. 일정 예의 구현예에서, 차폐성 구성체는 RNA 루프와 헤어핀을 포함한다. 활성화된 CRISPR 이펙터 단백질이 RNA 루프를 절단할 때, 개시제는 방출되어 HCR 반응을 촉발시킬 수 있다.

[0355] **표적 올리고뉴클레오티드의 증폭**

[0356] 일정 예의 구현예에서, 표적 RNA 및/또는 DNA는 CRISPR 이펙터 단백질을 활성화시키기 전에 증폭될 수 있다. 임의의 적합한 RNA 또는 DNA 증폭 기술이 사용될 수 있다. 특정 예시적 구현예에서, RNA 또는 DNA 증폭은 등온 증폭이다. 특정 예시적 구현예에서, 등온 증폭은 핵산 서열-기반 증폭 (NASBA), 리콤비나아제 중합효소 증폭 (RPA), 루프-매개 등온 증폭 (LAMP), 가닥 전위 증폭 (SDA), 헬리카아제-의존적 증폭 (HDA), 또는 닉킹 효소 증폭 반응 (NEAR) 일 수 있다. 특정 예시적 구현예에서, 비-등온 증폭 방법은 PCR, 다중 전위 증폭 (MDA), 롤링 써클 증폭 (RCA), 리가아제 연쇄 반응 (LCR), 또는 세분화 증폭 방법 (RAM) 을 비제한적으로 포함하는 것이 사용될 수 있다.

[0357] 특정 예시적 구현예에서, RNA 또는 DNA 증폭은 RNA/DNA 듀플렉스를 형성하기 위해 서열-특이적 역방향 프라이머에 의한 표적 RNA의 역전사에 의해 개시되는, NASBA 이다. 그 다음으로 RNase H는 RNA 주형을 분해하는데 사용되어서, 프로모터, 예컨대 T7 프로모터를 함유하는 전방향 프라이머가 결합되어 상보성 가닥의 연장을 개시할 수 있고, 이중-가닥 DNA 생성물이 생성된다. DNA 주형의 RNA 중합효소 프로모터-매개 전사가 표적 RNA 서열의 카피를 생성시킨다. 중요한 것은, 신규 표적 RNA 각각이 가이드 RNA 에 의해 검출될 수 있고 그리하여 어세이의 감도를 더 증강시킬 수 있다는 것이다. 그런 다음, 가이드 RNA 에 의한 표적 RNA 의 결합이 CRISPR 이펙터 단백질의 활성화를 야기시키고 방법은 상기 약술된 대로 진행된다. NASBA 반응은 예를 들어 대략 41℃ 의 중간 등온 조건 하에서 진행될 수 있다는 추가의 장점을 가져서, 임상 실험실로부터 멀리 떨어진 현장에서 조기 및 직접 검출을 위해 배치된 시스템 및 장치에 적합하다.

[0358] 일정한 다른 일례의 구현예에서, 리콤비나제 중합효소 증폭 (RPA) 반응은 표적 핵산을 증폭시키는데 사용될 수 있다. RPA 반응은 듀플렉스 DNA 의 상동성 서열과 서열-특이적 프라이머의 쌍을 형성할 수 있는 리콤비나아제를 이용한다. 표적 DNA 가 존재하면, DNA 증폭이 개시되고 다른 샘플 조작 예컨대 열 사이클 또는 화학적 용융이 필요하지 않다. 전체 RPA 증폭 시스템은 건조된 제제로서 안정하고 냉동 없이 안전하게 수송될 수 있다. RPA 반응은 또한 37-42℃ 의 최적 반응 온도로 등온 온도에서 실행될 수 있다. 서열 특이적 프라이머는 검출할 표적 핵산 서열을 포함하는 서열이 증폭되도록 설계된다. 특정 예시적 구현예에서, RNA 중합효소 프로모터, 예컨대 T7 프로모터는 프라이머 중 하나에 첨가된다. 그 결과로 표적 서열 및 RNA 중합효소 프로모터를 포함하는 증폭된 이중-가닥 DNA 생성물이 얻어진다. RPA 반응 이후, 또는 그 동안, RNA 중합효소가 첨가되어 이중-가닥 DNA 주형으로부터 RNA 를 생성시키게 될 것이다. 이어서, 증폭된 표적 RNA 가 그 다음으로 CRISPR 이펙터 시스템에 의해 검출될 수 있다. 이러한 방식으로 표적 DNA 는 본원에 개시된 구현예를 사용하여 검출될 수 있다. RPA 반응은 또한 표적 RNA 를 증폭시키는데 사용될 수 있다. 표적 RNA 는 먼저 역전사효소를 사용해 cDNA 로 전환되고, 그 다음으로 제 2 가닥 DNA 합성이 후속되며, 이 시점에 RPA 반응은 상기 약술된 대로 진행된다.

[0359] 본 발명의 일 구현예에서, 니킹 효소는 CRISPR 단백질이다. 따라서, dsDNA로 닉의 도입은 프로그램가능할 수 있고 서열-특이적일 수 있다. 도 5는 dsDNA 표적의 반대 가닥을 표적화하도록 설계된 2개 가이드에서 시작되는 본 발명의 일 구현예를 도시한다. 본 발명에 따라서, 닉카제는 Cpf1, C˚C와 함께 사용되는 C2c1 또는 C2c1일 수 있다. 다른 구현예에서, 등온 증폭의 온도는 상이한 온도에서 작동가능한 중합효소 (예를 들어, Bsu, Bst, Phi29, 클레노우 단편 등)를 선택하여 선택될 수 있다.

[0360] 따라서, 닉킹 등온 증폭 기술은 (예를 들어, 닉킹 효소 증폭 반응 또는 NEAR에서) 고정된 서열 선호도를 갖는 닉킹 효소를 사용하여, 표적의 말단에 닉킹 기질을 첨가하는 프라이머의 어닐링 및 연장을 허용하기 위해 본래 dsDNA 표적의 변성을 필요로 하지만, 닉킹 부위가 가이드 RNA를 통해서 프로그램될 수 있는 것인 CRISPR 닉카제의 사용은 변성 단계를 필요로 하지 않아서, 전체 반응이 진짜로 등온성이게 된다는 것을 의미한다. 이것은 또한 닉킹 기질을 첨가하는 이들 프라이머가 이후 반응에서 사용되는 프라이머와 상이하기 때문에 반응을 단순화시키며, 이것은 NEAR이 2개 프라이머 세트 (즉, 4개 프라이머)를 필요로 하지만 C2c1 닉킹 증폭은 오직 하나의 프라이머 세트 (즉, 2개 프라이머)를 필요로 한다는 것을 의미한다. 이것은 변성 및 이후 등온 온도로

냉각을 수행하기 위해 복잡한 장비없이 작업하도록 닉킹 C2c1 증폭을 훨씬 더 단순하고 쉽게 만든다.

[0361] 따라서, 일정한 예의 구현예에서, 본 명세서에 개시된 시스템은 증폭 시약을 포함할 수 있다. 핵산의 증폭에 유용한 상이한 성분 또는 시약은 본 명세서에 기술되어 있다. 예를 들어, 본 명세서에 기술된 바와 같은 증폭 시약은 완충제, 예컨대 Tris 완충제를 포함할 수 있다. Tris 완충제는 예를 들어, 제한없이, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11 mM, 12 mM, 13 mM, 14 mM, 15 mM, 25 mM, 50 mM, 75 mM, 1 M 등의 농도를 포함하여, 바람직한 적용 또는 용도에 적절한 임의 농도로 사용될 수 있다. 당업자는 완충제 예컨대 본 발명에서 사용을 위한 Tris의 적절한 농도를 결정할 수 있을 것이다.

[0362] 염, 예컨대 마그네슘 클로라이드 (MgCl2), 포타슘 클로라이드 (KCl), 또는 소듐 클로라이드 (NaCl) 가 핵산 단편의 증폭을 개선시키기 위해서, 증폭 반응, 예컨대 PCR 에 포함될 수 있다. 염 농도가 특정한 반응 및 적용분야에 의존적일 것이지만, 일부 구현예에서, 특정한 크기의 핵산 단편은 특정한 염 농도에서 최적 결과를 생성시킬 수 있을 것이다. 바람직한 결과를 생성시키기 위해, 더 큰 생성물은 변경된 염 농도, 전형적으로 더 낮은 염을 요구할 수 있는 한편, 더 작은 생성물의 증폭은 더 높은 염 농도에서 보다 양호한 결과를 생성시킬 수 있다. 당업자는 염 농도의 변경과 함께, 염의 존재 및/또는 농도가 생물학적 또는 화학적 반응의 엄격도를 변경시킬 수 있고, 그러므로 본원에서 기재된 바와 같이 본 발명의 반응을 위해 적절한 조건을 제공하는 임의의 염을 사용할 수 있다는 것을 이해하게 될 것이다.

[0363] 생물학적 또는 화학적 반응의 다른 성분은 세포 안의 물질의 분석을 위해 세포를 파쇄하거나 용해시키기 위해 세포 용해 성분을 포함할 수 있다. 세포 용해 성분은 세제, 상기 기재된 바와 같은 염, 예컨대 NaCl, KCl, 암모늄 술페이트 [(NH4)2SO4], 또는 다른 것들을 포함할 수 있으나 이에 제한되는 것은 아니다. 본 발명에 적절할 수 있는 세제는 Triton X-100, 소듐 도데실 술페이트 (SDS), CHAPS (3-[(3-콜아미도프로필)디메틸암모니오]-1-프로판술포네이트), 에틸 트리메틸 암모늄 브로마이드, 노닐 페녹시폴리에톡실에탄올 (NP-40)을 포함할 수 있다. 세제의 농도는 특정 적용분야에 의존적일 수 있고, 일부 경우에서 반응에 특이적일 수 있다. 증폭 반응은 예컨대 100 nM, 150 nM, 200 nM, 250 nM, 300 nM, 350 nM, 400 nM, 450 nM, 500 nM, 550 nM, 600 nM, 650 nM, 700 nM, 750 nM, 800 nM, 850 nM, 900 nM, 950 nM, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 20 mM, 30 mM, 40 mM, 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 150 mM, 200 mM, 250 mM, 300 mM, 350 mM, 400 mM, 450 mM, 500 mM 등의 농도를 비제한적으로 포함하여, 본 발명에 적절한 임의 농도로 사용되는 dNTP 및 핵산 프라이머를 포함할 수 있다. 마찬가지로, 본 발명에 따라서 유용한 폴리머라아제는 Taq 폴리머라아제, Q5 폴리머라아제 등을 포함하는, 당분야에 공지되어 있고 본 발명에서 유용한 임의의 특이적이거나 일반적인 폴리머라아제일 수 있다.

[0364] 일부 구현예에서, 본원에서 기재된 바와 같은 증폭 시약은 핫-스타트 증폭에서 사용하기에 적절할 수 있다. 핫 스타트 증폭은 어댑터 분자 또는 올리고의 이량체화를 감소시키거나 제거시키기 위해서, 또는 달리 원치않는 증폭 생성물 또는 인공물을 방지하고 바람직한 생성물의 최적 증폭을 수득하기 위해서 일부 구현예에서 유리할 수 있다. 증폭에 사용하기 위한 본원에 기재된 많은 성분이 또한 핫-스타트 증폭에서 사용될 수 있다.

일부 구현예에서, 핫-스타트 증폭에서 사용하기에 적절한 시약 또는 성분은 적절하다면 조성물 성분 중 하나 이상 대신에 사용될 수 있다. 예를 들어, 중합효소 또는 다른 시약은 특정한 온도 또는 다른 반응 조건에서 바람직한 활성을 나타내는 것이 사용될 수 있다. 일부 구현예에서, 시약은 핫-스타트 증폭에서 사용하기 위해 디자인되거나 최적화된 것을 사용할 수 있으며, 예를 들어, 중합효소는 전위 이후 또는 특정한 온도에 도달 이후에 활성화될 수 있다. 이러한 중합효소는 항체-기반 또는 압타머-기반일 수 있다. 본 명세서에 기술된 바와 같은 중합효소는 당분야에 공지되어 있다. 이러한 시약의 예는 제한없이, 핫-스타트 중합효소, 핫-스타트 dNTP, 및 광-케이징된 dNTP를 포함할 수 있다. 이러한 시약은 공지되어 있고 당분야에서 입수가능하다. 당업자는 개별 시약에 적절하게 최적 온도를 결정할 수 있을 것이다.

[0365] 핵산의 증폭은 특별한 열 사이클 기계 또는 장비를 사용하여 수행될 수 있고, 단일 반응으로 또는 대량으로 수행될 수 있어, 임의의 바람직한 횟수의 반응이 동시에 수행될 수 있다. 일부 구현예에서, 증폭은 미세유체 또는 로봇식 장치를 사용하여 수행될 수 있거나, 바람직한 증폭을 달성하도록 온도의 수동 변경을 사용하여 수행될 수도 있다. 일부 구현예에서, 특정한 적용분야 또는 물질에 대한 최적 반응 조건을 수득하기 위해 최적화가 수행될 수 있다. 당업자는 충분한 증폭이 수득되도록 반응 조건을 이해하게 될 것이고 최적화시킬 수 있을 것이다.

[0366] 특정 구현예에서, 본 발명의 방법 또는 시스템에 의한 DNA 의 검출은 검출 전에 (증폭된) DNA 를 RNA 로 전사시키는 것을 필요로 한다.

[0367] 본 발명의 검출 방법이 다양한 조합으로 핵산 증폭 및 검출 절차를 포함할 수 있다는 것은 자명할 것이다. 검출하려는 핵산은 검출할 수 있는 중간 생성물을 제공하기 위한 임의의 적합한 방법에 의해 증폭될 수 있는, DNA 및 RNA 를 제한없이 포함하는, 임의의 천연 발생 또는 합성 핵산일 수 있다. 중간 생성물의 검출은 직접 또는 부차적 활성에 의해 검출가능한 신호 모이어티를 생성시키는 CRISPR 단백질의 결합 및 활성화를 제한없이 포함하는, 임의의 적합한 방법에 의한 것일 수 있다.

[0368] 본 명세서에 개시된 시스템, 장치, 및 방법은 또한 특이적으로 구성된 폴리펩티드 검출 압타머의 도입을 통해서 핵산의 검출 이외에도 폴리펩티드 (또는 다른 분자)의 검출에 적합화될 수 있다. 폴리펩티드 검출 압타머는 상기 기술된 차폐성 구성체 압타머와 별개이다. 첫번째로, 압타머는 하나 이상의 표적 분자에 특이적으로 결합하도록 디자인된다. 일례의 구현예에서 표적 분자는 표적 폴리펩티드이다. 다른 일례의 구현예에서 표적 분자는 표적 화학적 화합물, 예컨대 표적 치료 분자이다. 소정 표적에 대한 특이성으로 압타머를 디자인하고 선택하기 위한 방법, 예컨대 SELEX는 당분야에 공지되어 있다. 소정 표적에 대한 특이성이외에도, 압타머는 R중합효소 프로모터 결합 부위를 도입시키도록 더욱 설계된다. 일정한 예의 구현예에서, 중합효소 프로모터는 T7 프로모터이다. 표적에 압타머의 결합 전에, 중합효소 부위는 중합효소가 접근가능하거나, 또는 달리 인식가능하지 않다. 그러나, 압타머는 표적의 결합 시 압타머의 구조가 입체형태 변화를 겪어서 중합효소 프로모터가 노출되도록 구성된다. 중합효소 프로모터의 하류의 압타머 서열은 RNA 또는 DNA 중합효소에 의한 기폭제 올리고뉴클레오티드의 생성을 위한 주형으로서 작용한다. 따라서, 압타머의 주형 부분은 소정 압타머 및 이의 표적을 식별하는 바코드 또는 다른 식별 서열을 더 도입시킬 수 있다. 상기 기술된 바와 같은 가이드 RNA는 이들 특이적 기폭제 올리고뉴클레오티드 서열을 인식하도록 디자인될 수 있다. 기폭제 올리고뉴클레오티드에 가이드 RNA의 결합은 CRISPR 이펙터 단백질을 활성화시키고, 이전에 기술된 바와 같이 차폐성 구성체를 탈활성화시켜서 양성 검출가능한 신호를 발생시키도록 진행된다.

[0369] 따라서, 일정한 예의 구현예에서, 본 명세서에 개시된 방법은 샘플 또는 샘플의 세트를 개별 이산 부피의 세트에 분배시키는 단계로서, 각각의 개별 이산 부피는 펩티드 검출 압타머, CRISPR 이펙터 단백질, 하나 이상의 가이드 RNA, 차폐성 구성체를 포함하는 것인 단계, 및 샘플 또는 샘플의 세트를 하나 이상의 표적 분자와 펩티드 검출 압타머의 결합을 허용하기 위해 충분한 조건 하에서 인큐베이션시키는 단계로서, 상응하는 표적에 대한 압타머의 결합은 중합효소 프로모터 결합 부위를 노출시켜서 RNA 중합효소 프로모터 결합 부위에 RNA 중합효소의 결합을 통해 기폭제 올리고뉴클레오티드의 합성이 개시되는 것인 단계의 추가 단계를 포함한다.

[0370] 다른 일례의 구현예에서 압타머의 결합은 표적 폴리펩티드에 압타머의 결합 시 프라이머 결합 부위를 노출시킬 수 있다. 예를 들어, 압타머는 RPA 프라이머 결합 부위를 노출시킬 수 있다. 따라서, 프라이머의 첨가 또는 포함은 증폭 반응, 예컨대 상기 약술된 바와 같은 RPA 반응에 공급되어 질 것이다.

[0371] 일정 예의 구현예에서, 압타머는 관심 표적에 결합 시, 2차 구조를 변화시켜서 단일-가닥 DNA의 새로운 영역을 노출시킬 수 있는, 입체형태-전환 압타머일 수 있다. 일정 예의 구현예에서, 이들 단일-가닥 DNA의 새로운 영역은 결찰을 위한 기질로서 사용될 수 있어서, 압타머를 연장하고 더 긴 ssDNA 분자를 생성시켜서 이것이 본 명세서에 개시된 구현예를 사용해 특이적으로 검출될 수 있다. 압타머 디자인은 글루코스와 같은, 저-에피토프 표적의 검출을 위한 3원 복합체와 더 조합될 수 있다 (Yang et al. 2015: pubs.acs.org/doi/abs/10.1021/acs.analchem.5b01634). 예시적인 입체형태 이동 압타머 및 상응하는 가이드 RNA (crRNA)는 하기 표에 표시되어 있다.

트롬빈 압타머	tgtggttggt gtggttggtt catggtcata ttggtttttt ttttttttc caaccacagtctctgt (SEQ ID NO:446)
트롬빈 결찰 프로브	ggttggtagt ctcgaattgc tctctttcac tggcc (SEQ ID NO:447)
트롬빈 RPA 전방향 1 프라이머	gaaattaata cgactcacta taggggggttg gttcatggtc atattggt (SEQ ID NO:448)
트롬빈 RPA 전방향 2 프라이머	gaaattaata cgactcacta taggggggttg gtgtggttgg ttcatggtca tattggt (SEQ ID NO:449)
트롬빈 RPA 역방향 1 프라이머	ggccagtgaa agagagcaat tcgagactac c (SEQ ID NO:450)
트롬빈 crRNA 1	gauuuagacu accccaaaaa cgaaggggac uaaaacccag ugaaagagag caauucgaga cuac (SEQ ID NO:451)
트롬빈 crRNA 2	gauuuagacu accccaaaaa cgaaggggac uaaaacaaag agagcaauuc gagacuacca acca (SEQ ID NO:452)
트롬빈 crRNA 3	gauuuagacu accccaaaaa cgaaggggac uaaaacagac uaccaaccac agagacugug guug (SEQ ID NO:453)
PTK7 전체 길이 앰플리콘 대조군	gttagatcgc aagcatatca ttgcgcttgc gatctaactg ctgcgccgcc gggaaaatac tgtacggtta gatcgcatag tctcgaattg ctctctttca ctggcc (SEQ ID NO:454)
PTK7 압타머	gttagatcgc aagcatatca ttgcgcttgc gatctaactg ctgcgccgcc gggaaaatac tgtacggtta g (SEQ ID NO:455)
PTK7 결찰 프로브	atcgcatagt ctcgaattgc tctctttcac tggcc (SEQ ID NO:456)
PTK7 RPA 전방향 1	gaaattaata cgactcacta tagggatcgc aagcatatca ttgcgcttgc

[0372]

프라이머	(SEQ ID NO:457)
PTK7 RPA 역방향 1 프라이머	ggccagtgaa agagagcaat tcgagactat g (SEQ ID NO:458)
PTK7 crRNA 1	gauuuagacu accccaaaaa cgaaggggac uaaaacccag ugaaagagag caauucgaga cuau (SEQ ID NO:459)
PTK7 crRNA 2	gauuuagacu accccaaaaa cgaaggggac uaaaacagag caauucgaga cuaugcgauc uaac (SEQ ID NO:460)
PTK7 crRNA 3	gauuuagacu accccaaaaa cgaaggggac uaaaacacua ugcgaucuaa ccguacagua uuuu (SEQ ID NO:461)

[0373]

[0374] **CRISPR 시스템의 사용 방법에 대한 일반 설명**

[0375] 특정 구현예에서, 본 명세서에 기술된 방법은 하나 이상의 관심 폴리뉴클레오티드 표적을 표적화하는 단계를 포함할 수 있다. 관심 폴리뉴클레오티드 표적은 특별한 질환 또는 이의 치료와 관련되거나, 소정 관심 특성의 생성과 관련되거나, 또는 관심 분자의 생성과 관련된 표적일 수 있다. "폴리뉴클레오티드 표적" 의 표적화를 언급할 때, 이것은 코딩 영역, 인트론, 프로모터, 및 임의의 다른 5' 또는 3' 조절 영역 예컨대 종결 영역, 리보솜 결합 부위, 인핸서, 사일렌서 등 중 하나 이상을 표적화하는 것을 포함할 수 있다. 유전자는 임의의 관심 단백질 또는 RNA를 코딩할 수 있다. 따라서, 표적은 mRNA, tRNA, 또는 rRNA로 전사될 수 있는 코딩 영역일 수 있거나, 또한 이의 복제, 전사, 및 조절에 관여되는 단백질에 대한 인식 부위일 수 있다.

[0376] 특정 구현예에서, 본 명세서에 기술된 방법은 하나 이상의 관심 유전자를 표적화하는 단계를 포함할 수 있고, 적어도 하나의 관심 유전자는 긴 비코딩 RNA (lncRNA)를 코딩한다. 한편 lncRNA는 세포 기능화에 결정적인

것으로 확인되었다. 필수적인 lncRNA가 각 세포 유형에서 상이한 것으로 확인되었으므로 (C.P. Fulco et al., 2016, Science, doi:10.1126/science.aag2445; N.E. Sanjana et al., 2016, Science, doi:10.1126/science.aaf8325), 본 명세서에 제공되는 방법은 관심 세포를 위한 세포 기능과 관련된 lncRNA를 결정하는 단계를 포함할 수 있다.

[0377] 외생의 폴리뉴클레오티드 주형을 통합함으로써 표적 폴리뉴클레오티드를 변형하기 위한 예시적인 방법에서, 이중 가닥 파손은 CRISPR 복합체에 의해 게놈 서열 내로 도입되고, 그 파손은 그 주형이 게놈 내로 통합되도록 하는 외생성 폴리뉴클레오티드 주형을 이용한 상동성 재조합을 통해 복구된다. 이중가닥 파손의 존재는 주형의 통합을 촉진한다.

[0378] 다른 구현예에서, 본 발명은 진핵생물 세포 내에서 폴리뉴클레오티드의 발현을 변형하는 방법을 제공한다. 방법은 폴리뉴클레오티드와 결합하는 CRISPR 복합체를 사용하여 표적 폴리뉴클레오티드의 발현을 증가시키거나 감소시키는 단계를 포함한다.

[0379] 일부 방법에서, 표적 폴리뉴클레오티드는 비활성화되어 세포 내에서 발현의 변형을 일으킬 수 있다. 예를 들어, CRISPR 복합체의 세포 내의 표적 서열로의 결합시, 표적 폴리뉴클레오티드는 비활성화되어, 서열이 전사되지 않거나, 코딩된 단백질이 생산되지 않거나, 서열이 야생형 서열이 작용하는 것처럼 작용하지 않게 된다. 예를 들어, 단백질 또는 마이크로RNA 코딩 서열은 비활성화되어 단백질이 생산되지 않게 된다.

[0380] 일부 방법에서, 제어 서열이 더 이상 조절 서열로서 기능하지 않도록 비활성화될 수 있다. 본 명세서에서 사용된 바와 같은, "제어 서열" 은 핵산 서열의 전사, 번역, 또는 접근성에 영향을 미치는 임의의 핵산 서열을 지칭한다. 제어 서열의 예들은 프로모터, 전사 종결자, 및 인핸서를 포함하며 이들이 제어 서열이다. 비활성화된 표적 서열은 결실 돌연변이(즉, 하나 이상의 뉴클레오티드의 결실), 삽입 돌연변이(즉, 하나 이상의 뉴클레오티드의 삽입), 또는 논센스 돌연변이(즉, 정지 코돈이 도입되도록 하는, 하나의 뉴클레오티드의 또 다른 뉴클레오티드로의 치환)를 포함할 수 있다. 일부 방법에서, 표적 서열의 비활성화는 표적 서열의 "녹아웃" 을 초래한다.

[0381] 본 명세서는 또한 복수의 조합 교란을 도입시켜 세포 상호작용을 확인하는 단계 및 관찰된 게놈, 유전자, 단백질체, 후성유전, 및/또는 표현형 효과를 " perturb-seq" 라고도 하는, 단일 세포에서 관찰된 교란과 상관짓는 단계를 포함하는 기능적 유전체학 방법을 제공한다. 일 구현예에서, 이들 방법은 combine 단일-세포 RNA 시퀀싱 (RNA-seq) 및 CRISPR (clustered regularly interspaced short palindromic repeats)-기반 교란을 조합한다 (Dixit et al. 2016, Cell 167, 1853-1866; Adamson et al. 2016, Cell 167, 1867-1882). 일반적으로, 이들 방법은 세포 개체군 중 다수 세포로 다수의 조합 교란을 도입시키는 단계로서, 다수 세포의 각각의 세포는 적어도 1 교란을 받는 것인 단계, 임의의 교란을 받지 않은 하나 이상의 세포와 비교하여 단일 세포에서 게놈, 유전자, 단백질체, 후성유전, 및/또는 표현형 편차를 검출하는 단계, 및 단일 세포에서 교란(들)을 검출하는 단계; 및 측정된 편차에 대해 공변량을 처리하는 모델을 적용하여 교란과 관련된 측정된 편차를 결정하여, 세포간 및/또는 세포내 네트워크 또는 회로를 추론하는 것인 단계를 포함한다. 보다 특히, 단일 세포 시퀀싱은 세포 바코드를 포함하여서, 각 RNA의 기원 세포를 기록한다. 보다 특히, 단일 세포 시퀀싱은 고유 분자 식별자 (UMI)를 포함하여서, 단일 세포내에서, 전사물 카피수 또는 프로브 결합 사건같은, 측정된 신호의 포획률을 결정한다.

[0382] 이들 방법은 세포 회로의 조합적 탐지, 세포 회로 해체, 분자 경로의 설명, 및/또는 치료제 개발을 위한 관련 표적의 확인을 위해 사용될 수 있다. 보다 특히, 이들 방법은 그들 분자 프로파일링을 기반으로 세포군을 확인하기 위해 사용될 수 있다. 유기적 (예를 들어, 질환) 및 (예를 들어, 소형 분자에 의한) 유도적 상태 간 유전자-발현 프로파일의 유사성은 임상적으로 효과적인 요법을 확인할 수 있다.

[0383] 따라서, 특정 구현예에서, 본 명세서에서 제공되는 치료 방법은 대상체로부터 단리된 세포 개체군에 대해서, 상기 기술된 바와 같은 perturb-seq를 사용하여, 최적 치료 표적 및/또는 치료제를 결정하는 단계를 포함한다.

[0384] 특정 구현에에서, 본 명세서의 다른 곳에 언급된 바와 같은 perturb-seq 방법은 단리된 세포 또는 세포주에서, 관심 분자의 생성에 영향을 미칠 수 있는 세포 회로를 결정하는데 사용된다.

[0385] **추가의 CRISPR-Cas 개발 및 사용 고려사항**

[0386] 본 발명은 하기 문헌들에서 설명된 바와 같이, 특히 세포 및 생물에서 CRISPR 단백질 복합체의 전달 및 RNA 가이드된 엔도뉴클레아제의 이용에 관한 것과 같이, CRISPR-Cas9 개발 및 이용의 양태를 기반으로 추가로 예시 및

연장될 수 있다:

- Multiplex genome engineering using CRISPR/Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. Science Feb 15;339(6121):819-23 (2013);
- RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013);

[0387]

- One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM ., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013);
- Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Nature. Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23 (2013);
- Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. Cell Aug 28. pii: S0092-8674(13)01015-5 (2013-A);
- DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013);
- Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. Nature Protocols Nov;8(11):2281-308 (2013-B);
- Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013). [Epub ahead of print];
- Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Cell Feb 27, 156(5):935-49 (2014);
- Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. Nat Biotechnol. Apr 20. doi: 10.1038/nbt.2889 (2014);
- CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F. Cell 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014);
- Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu PD, Lander ES, Zhang F., Cell. Jun 5;157(6):1262-78 (2014).

[0388]

- Genetic screens in human cells using the CRISPR/Cas9 system, Wang T, Wei JJ, Sabatini DM, Lander ES., Science. January 3; 343(6166): 80–84. doi:10.1126/science.1246981 (2014);
- Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation, Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE., (published online 3 September 2014) Nat Biotechnol. Dec;32(12):1262-7 (2014);
- *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F., (published online 19 October 2014) Nat Biotechnol. Jan;33(1):102-6 (2015);
- Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015).
- A split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (published online 02 February 2015) Nat Biotechnol. Feb;33(2):139-42 (2015);
- Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246–1260, March 12, 2015 (multiplex screen in mouse), and
- In *vivo* genome editing using *Staphylococcus aureus* Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (published online 01 April 2015), Nature. Apr 9;520(7546):186-91 (2015).
- Shalem et al., "High-throughput functional genomics using CRISPR-Cas9," Nature Reviews Genetics 16, 299-311 (May 2015).
- Xu et al., "Sequence determinants of improved CRISPR sgRNA design," Genome Research 25, 1147-1157 (August 2015).
- Parnas et al., "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks," Cell 162, 675-686 (July 30, 2015).
- Ramanan et al., CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus," Scientific Reports 5:10833. doi: 10.1038/srep10833 (June 2, 2015)

[0389]

- Nishimasu et al., Crystal Structure of Staphylococcus aureus Cas9," Cell 162, 1113-1126 (Aug. 27, 2015)
- BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis, Canver et al., Nature 527(7577):192-7 (Nov. 12, 2015) doi: 10.1038/nature15521. Epub 2015 Sep 16.
- *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System*, Zetsche et al., Cell 163, 759-71 (Sep 25, 2015).
- *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems*, Shmakov et al., Molecular Cell, 60(3), 385–397 doi: 10.1016/j.molcel.2015.10.008 Epub October 22, 2015.
- *Rationally engineered Cas9 nucleases with improved specificity*, Slaymaker et al., Science 2016 Jan 1 351(6268): 84-88 doi: 10.1126/science.aad5227. Epub 2015 Dec 1. [Epub ahead of print].
- Gao *et al*, "Engineered Cpf1 Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: dx.doi.org/10.1101/091611 (Dec. 4, 2016)

[0390]

[0391] 참조로 본 명세서에 편입되는 이들 각각은 본 발명의 실시에서 고려될 수 있고, 하기에 간략하게 설명한다:

[0392] ▷문헌 [Cong et al.]은 스트렙토코커스 써모필러스 (Streptococcus thermophilus) Cas9와 또한 스트렙토코커스 피오게네스 (Streptococcus pyogenes) Cas9 둘 모두에 기반하여 진핵생물 세포에서 사용하기 위한 II형 CRISPR-Cas 시스템을 조작하였고, Cas9 뉴클레아제가 짧은 RNA에 의해 인간 및 마우스 세포에서 DNA의 정확한 절단을 유도하도록 지정될 수 있다는 것을 입증하였다. 이들의 연구는 닉킹 효소로 전환됨에 따라 Cas9 가 최소의 돌연변이유발 활성을 갖는 진핵생물 세포에서 상동성-유도된 복구을 촉진하는데 이용될 수 있음을 추가로 보여준다. 추가적으로, 그들의 연구는 포유류 게놈 내의 내생성 게놈 유전자좌 부위에서 몇몇의 자발적 편집을 가능하게 하기 위해 다중 가이드 서열이 단일 CRISPR 어레이로 암호화될 수 있다는 것을 입증하였는데, 이는 RNA-가이드된 뉴클레아제 기술의 용이한 프로그램 가능성(programmability) 및 넓은 적용 가능성을 입증하였다. 세포에서 서열 특이적 DNA 절단을 프로그램하기 위해 RNA를 사용하는 이런 능력은 새로운 부류의 게놈 조작 도구를 정하였다. 이들 연구는 다른 CRISPR 유전자좌가 포유동물 세포에 이식 가능하게 될 가능성이 있고, 또한 포유동물 게놈 절단을 매개할 수 있다는 것을 추가로 보여주었다. 중요하게는, CRISPR-Cas 시스템의 몇몇 양상은 그의 효율 및 융통성을 증가시키기 위해 더 개선될 수 있다는 것을 상정할 수 있다.

[0393] ▷문헌 [Jiang et al.]은 스트렙토코커스 뉴모니아 (Streptococcus pneumoniae) 및 에스케리치아 콜라이 (Escherichia coli)의 게놈에서 정확한 돌연변이를 도입하기 위해 이중-RNA와 복합체화된 주기적 간격으로 분포하는, 짧은 회문구조 반복부 (CRISPR)-연관 Cas9 엔도뉴클레아제를 사용하였다. 상기 접근법은 비돌연변이 세포를 사멸시키고 선택 가능한 마커 또는 반대-선택 시스템에 대한 필요를 피하기 위해 표적화된 게놈 부위에서 이중-RNA:Cas9-지정 절단에 의존한다. 상기 연구는 편집 주형 상에서 수행되는 단일- 및 다중뉴클레오티드 변화를 생성하기 위해 짧은 CRISPR RNA(crRNA) 서열을 변화시킴으로써 재프로그래밍 이중-RNA:Cas9 특이성을 보고하였다. 본 연구는 2개의 crRNA의 동시 사용이 다중복합 돌연변이유발을 가능하게 하였다는 것을 보여주었다. 또한, 상기 접근법이 리컴비니어링 (recombineering) 과 병용하여 이용된 경우, 스트렙토코커스 뉴모니애에서 기재된 접근법을 이용하여 회수된 세포의 거의 100% 가 요망되는 돌연변이를 함유했고, 에스케리키아 콜라이에서, 회수된 65%가 돌연변이를 함유했다.

[0394] ▷문헌 [Wang et al (2013)]은 배아 줄기 세포에서의 순차적 재조합 및/또는 단일 돌연변이를 갖는 마우스의 시간-소모적 상호교잡 (intercrossing)에 의해 다중 단계에서 전통적으로 생성된 다중 유전자에서의 돌연변이를 운반하는 마우스의 1단계 생성을 위한 CRISPR-Cas 시스템을 사용하였다. CRISPR-Cas 시스템은 기능적으로 불필요한 유전자의 그리고 상위 유전자 상호작용의 생체내 연구를 크게 가속화시킬 것이다. CRISPR-Cas 시스템은 기능적으로 중복되는 유전자 및 상위 유전자 상호작용의 생체내 연구를 크게 가속화시킬 것이다.

[0395] ▷문헌 [Konermann et al. (2013)]은 CRISPR Cas9 효소 및 또한 전사 활성체 유사 이펙터에 기반하여 DNA-결합 도메인의 광학적 그리고 화학적 조절을 가능하게 하는 다재다능하고 강한 기술에 대한 당업계의 요구를 해결하였다.

[0396] ▷문헌 [Ran et al. (2013-A)]은 표적화된 이중-가닥 파괴를 도입하기 위해 짝지어진 가이드 RNA와 Cas9 닉카제 돌연변이체를 조합한 접근을 기재하였다. 이는 가이드 서열에 의해 특정 게놈 유전자좌로 표적화되는 미생물 CRISPR-Cas 시스템으로부터의 Cas9 뉴클레아제의 문제를 해결하는데, 이는 DNA 표적에 대한 소정의 미스매치를 용인하고 이에 의해 원치않는 오프-표적 돌연변이유발을 촉진시킬 수 있다. 게놈에서 개개의 틈은 높은 충실도로 복구되기 때문에, 적절하게 오프셋 가이드 RNA를 통한 동시 닉킹은 이중-가닥 파괴에 필요하고, 표적 절단을 위해 특이적으로 인식되는 염기의 수를 연장시킨다. 저자는 짝지어진 닉킹의 사용이 세포주에서 50배 내지 1,500배 만큼 오프-표적 활성을 감소시킬 수 있고 온-표적의 절단 효율을 희생시키지 않고 마우스 접합체에서 유전자 녹아웃을 용이하게 한다는 것을 입증하였다. 이 다재다능한 전략은 높은 특이성을 필요로 하는 매우 다양한 게놈 편집 적용을 가능하게 한다.

[0397] ▷문헌 [Hsu et al. (2013)]은 표적 부위의 선택을 알아내기 위해 그리고 오프-표적 효과를 피하기 위해 인간 세포에서 SpCas9 표적화 특이성을 특징규명하였다. 연구는 293T 및 293FT 세포에서의 100개 초과의 예측된 게놈 오프-표적 유전자좌에서 700개 초과의 가이드 RNA 변이체 및 SpCas9-유도된 indel 돌연변이 수준을 평가하였다. 상기 저자는 SpCas9 가 미스매치의 수, 위치 및 분포에 민감한 서열-의존성 방식으로 상이한 위치에서 가이드 RNA 와 표적 DNA 사이의 미스매치를 용인하는 것을 보고했다. 상기 저자는 SpCas9-매개 절단이 DNA 메틸화에 의해 영향을 받지 않고, SpCas9 및 gRNA의 용량을 적정하여 오프-표적 변형을 최소화시킬 수 있음을 추가로 나타내었다. 또한, 포유류 게놈 조작 응용을 용이하게 하기 위해, 상기 저자는 표적 서열의 선택 및 확인 뿐만 아니라 오프-표적 분석을 안내하기 위해 웹-기반 소프트웨어 도구를 제공한 것을 보고했다.

[0398] ▷문헌 [Ran et al. (2013-B)]은 포유동물 세포에서 비-상동성 말단 결합 (NHEJ) 또는 상동성-지정 복구 (HDR)뿐만 아니라 하류의 기능성 연구를 위해 변형된 세포주의 생성을 통한 Cas9-매개 게놈 편집을 위한 도구의 세트를 기재하였다. 오프-표적 절단을 최소화하기 위해, 저자들은 짝지워진 가이드 RNA와 함께 Cas9 닉카제 돌연변이체를 이용하는 이중-닉킹 전략을 추가로 기재하였다. 저자에 의해 제공되는 프로토콜은 표적 부위의 선택, 절단 효율의 평가 및 오프-표적 활성의 분석을 위한 가이드라인을 실험적으로 유도하였다. 상기 연구들은 표적 디자인에서 시작하여, 유전자 변형이 1 내지 2주 내에 달성될 수 있고, 변형된 클론 세포주가 2 내지 3주 내에 유도될 수 있다는 것을 보여주었다.

[0399] ▷문헌 [Shalem et al.]은 게놈-와이드 규모에 대한 유전자 기능의 정보를 얻기 위한 새로운 방법을 기재하였다. 그들의 연구는 64,751개의 독특한 가이드 서열에 의해 18,080개의 유전자를 표적화한 게놈-규모 CRISPR-Cas9 녹아웃 (GeCKO) 라이브러리의 전달이 인간 세포에서 음성 스크리닝 및 양성 스크리닝 스크리닝 둘 모두를 가능하게 하였다는 것을 보여주었다. 처음에, 저자는 암 및 다능성 줄기 세포에서 세포 생존도에 필수적인 유전자를 식별하기 위한 GeCKO 라이브러리의 사용을 보여주었다. 다음으로, 흑색종 모델에서, 상기 저자는 유전자의 소실이 돌연변이 단백질 키나제 BRAF 를 억제하는 치료제인 베무라페닙 (vemurafenib) 에 대한 내성과 관련된 유전자를 스크리닝했다. 그들의 연구는 가장 높은 순위의 후보물질이 이전에 확인된 유전자 NF1 및 MED12 뿐만 아니라 새로운 히트 (hit) NF2, CUL3, TADA2B, 및 TADA1 을 포함한 것을 나타내었다.

저자들은 동일한 유전자를 표적화하는 독립적인 가이드 RNA 및 높은 비율의 히트 확증 간 높은 수준의 일관성을 관찰했고, 따라서, Cas9 에 의한 게놈 규모 스크리닝의 유망성을 입증했다.

[0400] ▷문헌 [Nishimasu et al.]은 2.5 A° 분해능으로 sgRNA 및 그의 표적 DNA와의 복합체에서 스트렙토코커스 피오게네스 Cas9의 결정 구조를 보고하였다. 상기 구조는 표적 인식 및 뉴클레아제 로브(lobe)로 구성된 바이로브(bilobed) 구조를 나타내었는데, 이는 그들의 계면에서 양으로 하전된 그루브에 sgRNA:DNA 헤테로듀플렉스를 수용한다. 인식 로브는 sgRNA 및 DNA의 결합에 필수적인 반면, 뉴클레아제 로브는 HNH 및 RuvC 뉴클레아제 도메인을 함유하는데, 이는 각각 표적 DNA의 상보성 및 비상보성 가닥의 절단을 위해 적절하게 위치된다. 뉴클레아제 로브는 또한 프로토스페이서 인접 모티프 (PAM)와의 상호작용을 초래하는 카복실-말단의 도메인을 함유한다. 이러한 고-해상도 구조 및 수반하는 기능 분석은 Cas9 에 의한 RNA-안내된 DNA 표적화의 분자 메카니즘을 나타내었고, 이에 따라 새로운 다능성 게놈-편집 기술의 합리적 디자인을 용이하게 한다.

[0401] ▷ Wu 등은 마우스 배아 줄기 세포 (mESC) 에서 단일 가이드 RNA (sgRNA) 가 로딩된 스트렙토코커스 피오게네스 유래의 촉매적 비활성 Cas9 (dCas9) 의 게놈-와이드 결합 자리를 맵핑했다. 저자들은 시험한 4개의 shRNA의 각각이 dCas9를 sgRNA 및 NGG 프로토스페이서 인접 모티프 (PAM)에서 5-뉴클레오티드 씨드 영역을 빈번하게 특징으로 하는, 수십 내지 수천개의 게놈 부위로 표적화한다는 것을 나타내었다. 염색질 비접근성은 일치하는 시드 서열과 다른 자리에 대한 dCas9 결합을 감소시키고, 따라서 70% 의 표적외 자리가 유전자와 연관되었다.

상기 저자는 촉매적 활성 Cas9 로 트랜스펙션된 mESC 내의 295 개의 dCas9 결합 자리의 표적화된 시퀀싱에 의해, 백그라운드 수준 초과로 돌연변이된 단지 하나의 자리가 확인된 것을 나타내었다. 상기 저자는 씨드

매치가 결합을 촉발하나, 표적 DNA 와의 광범위한 쌍형성이 절단에 필요한 Cas9 결합 및 절단을 위한 2-상태 모델을 제안했다.

[0402] ▷문헌 [Platt et al.]은 Cre-의존적 Cas9 녹인 (knock in) 마우스를 확립하였다. 저자들은 뉴런, 면역 세포 및 내피 세포에서 가이드 RNA의 아데노-연합 바이러스(AAV)-, 렌티바이러스-, 또는 입자-매개 전달을 이용하는 생체내 및 생체외 게놈 편집을 입증하였다.

[0403] ▷문헌 [Hsu et al. (2014)]은 세포의 유전자 스크리닝을 포함하여, 요거트에서부터 게놈 편집까지 CRISPR-Cas9 역사를 일반적으로 논의하는 리뷰 논문이다.

[0404] ▷문헌 [Wang et al. (2014)]은 게놈-규모 렌티바이러스 단일 가이드 RNA (sgRNA) 라이브러리를 사용하는 양성 선택 및 음성 선별 둘 모두에 적합한 풀링된, 기능상실 유전자 스크리닝 접근법에 관한 것이다.

[0405] ▷문헌 [Doench et al.]은 6개의 내생성 마우스 및 3개의 내생성 인간 유전자 패널의 모든 가능한 표적 부위에 걸쳐 타일링된 sgRNA의 풀을 생성하였고, 항체 염색 및 유세포분석에 의해 그들의 표적 유전자의 비대립유전자를 생성하는 그들의 능력을 정량적으로 평가하였다. 저자들은 PAM 개선된 활성의 최적화를 보였으며, sgRNA의 디자인을 위한 온라인 도구를 또한 제공했다.

[0406] ▷문헌 [Swiech et al.]은 AAV-매개 SpCas9 게놈 편집이 뇌에서 유전자 기능의 역 유전자 연구를 가능하게 할 수 있다는 것을 입증한다.

[0407] ▷문헌 [Konermann et al. (2015)]은 링커 존재 및 부재로 스템 또는 테트라루프와 같은 가이드 상의 적절한 위치에서 다중 이펙터 도메인, 예를 들어, 전사 활성체, 기능성 및 후생적 조절자에 부착하는 능력을 논의한다.

[0408] ▷문헌 [Zetsche et al.]은 Cas9 효소가 둘로 분할될 수 있고, 따라서 활성화를 위한 Cas9의 조립체가 제어될 수 있다는 것을 입증한다.

[0409] ▷문헌 [Chen et al.]은 마우스에서의 게놈-와이드 생체내 CRISPR-Cas9 스크린이 폐 전이를 조절하는 유전자를 밝혀준다는 것을 입증한 다중복합 스크리닝에 관한 것이다.

[0410] ▷문헌 [Ran et al. (2015)]은 SaCas9 및 게놈을 편집하는 그의 능력에 관한 것이며, 생화학적 분석으로부터 추론할 수 없다는 것을 입증한다.

[0411] ▷문헌 [Shalem et al. (2015)]은 촉매적 불활성인 Cas9 (dCas9) 융합체를 합성적으로 발현을 억제 (CRISPRi) 또는 활성화 (CRISPRa)시키는데 사용하여, 전사 활성을 조절하는 전략 및 게놈 유전자좌를 불활성화시키는 정렬되고 풀링된 스크린, 녹아웃 접근법을 포함하여, 게놈-규모 스크린에 Cas9를 사용하는 장점을 보여주는 방식을 기술한다.

[0412] ▷문헌 [Xu et al. (2015)]은 CRISPR-기반 스크리닝에서 단일 가이드 RNA (sgRNA) 효율에 기여하는 DNA 서열 특징을 평가하였다. 저자들은 CRISPR/Cas9 녹아웃 및 절단 자리에서 뉴클레오티드 선호성을 탐구했다. 저자들은 또한 CRISPRi/a 에 대한 서열 선호성이 CRISPR/Cas9 녹아웃에 대한 선호와 실질적으로 상이함을 발견했다.

[0413] ▷문헌 [Parnas et al. (2015)]은 박테리아성 지질다당체 (LPS)에 의한 종양 괴사 인자 (Tnf)의 유도를 제어하는 유정자를 확인하기 위해 게놈-와이드 풀링된 CRISPR-Cas9 라이브러리를 수지상 세포 (DC) 내로 도입하였다. Tlr4 신호전달의 알려진 조절자 및 이전에 알려지지 않은 후보를 동정하였고, LPS에 대한 정규 반응에 대해 별개의 효과를 갖는 3개의 기능성 모듈로 분류하였다.

[0414] ▷문헌 [Ramanan et al (2015)]은 감염 세포에서 바이러스 에피솜 DNA (cccDNA)의 절단을 입증하였다. HBV 게놈은 HBV 생활사에서 중요한 성분인 공유결합폐환형 DNA(covalently closed circular DNA:cccDNA)로 불리는 3.2kb의 이중-가닥 에피솜 DNA로서, 감염된 간세포의 핵에 존재하며, 이의 복제는 현재의 요법에 의해 저해되지 않는다. HBV 의 고도로 보존된 영역을 특이적으로 표적화하는 sgRNA 가 바이러스 복제 및 감손된 cccDNA 를 저해함을 보여주었다.

[0415] ▷문헌 [Nishimasu et al. (2015)]은 5'-TTGAAT-3' PAM 및 5'-TTGGGT-3' PAM을 함유하는 단일 가이드 RNA (sgRNA) 및 그의 이중-가닥 DNA와의 복합체에서 SaCas9의 결정 구조를 보고하였다. SpCas9 와 SaCas9 의 구조 비교는 구조적 보존 및 다양성 모두에서 강조표시하였으며, 이는 그들의 구별되는 PAM 특이성 및 오솔로그 sgRNA 인식을 설명한다.

[0416] ▷문헌 [Canver et al. (2015)]은 넌코딩 게놈 구성요소의 CRISPR-Cas9-기반 기능성 연구를 입증하였다. 저자들은 인핸서의 중요한 특징을 나타낸 인간과 마우스 BCL11A 인핸서의 인시추 포화 돌연변이유발을 수행하기 위해 풀링된 CRISPR-Cas9 가이드 RNA 라이브러리를 개발하였다.

[0417] ▷문헌 [Zetsche et al. (2015)]은 Cas9와 별개의 특징을 갖는 프란시셀라 노비시다 (Francisella novicida) U112로부터의 클래스 2 CRISPR 뉴클레아제인 Cas13의 특징규명을 보고하였다. Cpf1은 tracrRNA를 결여하는 단일 RNA-가이드 엔도뉴클레아제이고, T-풍부 프로토스페이서-인접 모티프를 이용하고, 스태거드 DNA 이중-가닥 파괴를 통해 DNA를 절단한다.

[0418] ▷문헌 [Shmakov et al. (2015)]은 3종의 별개 클래스 2 CRISPR-Cas 시스템을 보고하였다. 2종의 시스템 CRISPR 효소 (C2c1 및 C2c3)는 Cpf1과 멀리 관련된 RuvC-유사 엔도뉴클레아제 도메인을 함유한다. Cpf1과 달리, C2c1은 DNA 절단을 위해 crRNA와 tracrRNA 둘 다에 의존한다. 제3 효소(C2c2)는 2개의 예측된 HEPN RNase 도메인을 함유하고, tracrRNA 독립적이다.

[0419] ▷문헌[Slaymaker et al (2016)]은 스트렙토코커스 피오게네스 Cas9(SpCas9)의 특이성을 개선시키기 위한 구조-가이드 조작의 사용을 보고하였다. 저자들은 오프-표적 효과가 감소된 강한 표적 상의 절단이 유지된 " 향상된 특이성" SpCas9(eSpCas9) 변이체를 개발하였다.

[0420] 본 명세서에 제공된 방법 및 도구는 tracrRNA를 이용하지 않는 II 형 뉴클레아제, C2c1을 예시한다. C2c1의 오솔로그는 본 명세서에 기재된 바와 같은 상이한 박테리아 종에서 동정되었다. 유사한 특성을 갖는 추가적인 II형 뉴클레아제는 당업계에 기재된 방법을 이용하여 동정될 수 있다 (Shmakov et al. 2015, 60:385- 397; Abudayeh et al. 2016, Science, 5;353(6299)). 특정 구현예에서, 신규한 CRISPR 이펙터 단백질을 동정하기 위한 이러한 방법은 CRISPR Cas 유전자좌의 존재를 동정하는 씨드를 코딩하는 데이터베이스로부터 서열을 선택하는 단계, 선택된 서열에서 오픈 리딩 프레임 (ORF)을 포함하는 10kb의 씨드 내에 위치된 유전자좌를 동정하는 단계, 이로부터, 단일 ORF만이 700개 초과의 아미노산 및 공지된 CRISPR 이펙터에 대해 90% 이하의 상동성을 갖는 신규한 CRISPR 이펙터를 코딩하는 ORF를 포함하는 유전자좌를 선택하는 단계를 포함할 수 있다. 특정 구현예에서, 씨드는 CRISPR-Cas 시스템, 예컨대, Cas1에 대해 통상적인 단백질이다. 추가 구현예에서, CRISPR 어레이는 새로운 이펙터 단백질을 동정하기 위한 씨드로서 사용된다.

[0421] C2c1 및 crRNA를 포함하는 사전 조립된 재조합 CRISPR-C2c1 복합체는, 예를 들어 전기천공으로 형질감염되어, 높은 돌연변이율 및 검출 가능한 오프-표적 돌연변이의 부재를 초래한다. Hur, J.K. et al, Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596. [Epub ahead of print]. Cpf1을 사용하는 효율적인 다중복합 시스템은 본 발명의 tRNA를 함유하는 어레이로부터 가공된 gRNA를 사용하여 초파리에서 입증되었다. Port, F. et al, Expansion of the CRISPR toolbox in an animal with tRNA-flanked Cas9 and Cpf1 gRNAs. doi: dx.doi.org/10.1101/046417.

Cpf1 및 C2c1은 둘 모두가 구조 유사성을 공유하는 V형 CRISPR 단백질이다. C2c1 처럼, Cpf1은 (PAM의 말단에 블런트 절단부를 생성하는 Cas9와 대조적으로) PAM의 원위 말단에 스태거드 이중 가닥 파손을 생성시킨다. 따라서, C2c1을 적용한 유사한 복합다중 시스템이 고려된다.

[0422] 또한, ["Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome edit", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014)]은 연장된 서열을 인식하고 인간 세포에서 고효율로 내생성 유전자를 편집할 수 있는 이량체 RNA-가이드 FokI 뉴클레아제에 관한 것이다.

[0423] 양 및 제제에 대한 것을 포함하여, 방법, 재료, 전달 비히클, 벡터, 입자, AAB, 및 이의 제조 및 사용을 포함한, CRISPR-Cas 시스템, 이의 성분, 및 이러한 성분의 전달에 대한 일반 정보와 관련하여, 본 발명의 실시에 유용한 모든 것은 다음을 참조한다: US 특허 출원 번호 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418, 8,895,308, 8,906,616, 8,932,814, 8,945,839, 8,993,233 및 8,999,641; US 공개 특허 출원 US 2014-0310830 (US 특허 출원 일련 번호 14/105,031), US 2014-0287938 A1 (US 특허 출원 일련 번호 14/213,991), US 2014-0273234 A1 (US 특허 출원 일련 번호 14/293,674), US2014-0273232 A1 (US 특허 출원 일련 번호 14/290,575), US 2014-0273231 (US 특허 출원 일련 번호 14/259,420), US 2014-0256046 A1 (US 특허 출원 일련 번호 14/226,274), US 2014-0248702 A1 (US 특허 출원 일련 번호 14/258,458), US 2014-0242700 A1 (US 특허 출원 일련 번호 14/222,930), US 2014-0242699 A1 (US 특허 출원 일련 번호 14/183,512), US 2014-0242664 A1 (US 특허 출원 일련 번호 14/104,990), US 2014-0234972 A1 (US 특허 출원 일련 번호 14/183,471), US 2014-0227787 A1 (US 특허 출원 일련 번호 14/256,912), US 2014-0189896 A1 (US 특허 출원

일련 번호 14/105,035), US 2014-0186958 (US 특허 출원 일련 번호 14/105,017), US 2014-0186919 A1 (US 특허 출원 일련 번호 14/104,977), US 2014-0186843 A1 (US 특허 출원 일련 번호 14/104,900), US 2014-0179770 A1 (US 특허 출원 일련 번호 14/104,837) 및 US 2014-0179006 A1 (US 특허 출원 일련 번호 14/183,486), US 2014-0170753 (특허 출원 일련 번호 14/183,429); US 2015-0184139 (US 특허 출원 일련 번호 14/324,960); 14/054,414 유럽 특허 출원 EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), 및 EP 2 784 162 (EP14170383.5); 및 PCT 공개 번호 WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO 2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO 2014/093701 (PCT/US2013/074800), WO 2014/018423 (PCT/US2013/051418), WO 2014/204723 (PCT/US2014/041790), WO 2014/204724 (PCT/US2014/041800), WO 2014/204725 (PCT/US2014/041803), WO 2014/204726 (PCT/US2014/041804), WO 2014/204727 (PCT/US2014/041806), WO 2014/204728 (PCT/US2014/041808), WO 2014/204729 (PCT/US2014/041809), WO 2015/089351 (PCT/US2014/069897), WO 2015/089354 (PCT/US2014/069902), WO 2015/089364 (PCT/US2014/069925), WO 2015/089427 (PCT/US2014/070068), WO 2015/089462 (PCT/US2014/070127), WO 2015/089419 (PCT/US2014/070057), WO 2015/089465 (PCT/US2014/070135), WO 2015/089486 (PCT/US2014/070175), PCT/US2015/051691, PCT/US2015/051830. 또한 하기를 참조한다: 미국 가출원 61/758,468; 61/802,174; 61/806,375; 61/814,263; 61/819,803 및 61/828,130 (각각, 2013 년 1 월 30 일; 2013 년 3 월 15 일; 2013 년 3 월 28 일; 2013 년 4 월 20 일; 2013 년 5 월 6 일 및 2013 년 5 월 28 일에 출원됨). 또한 하기를 참조한다: 미국 가출원 61/836,123 (2013 년 6 월 17 일에 출원됨). 또한, 각각 2013년 6월 17일에 출원된 미국 가출원 제61/835,931호, 제61/835,936호, 제61/835,973호, 제61/836,080호, 제61/836,101호, 및 제61/836,127호에 대하여 참조가 이루어진다. 또한 하기를 참조한다: 미국 가출원 61/862,468 및 61/862,355 (2013 년 8 월 5 일에 출원됨); 61/871,301 (2013 년 8 월 28 일에 출원됨); 61/960,777 (2013 년 9 월 25 일에 출원됨) 및 61/961,980 (2013 년 10 월 28 일에 출원됨). 또한 다음을 더욱 참조한다: PCT/US2014/62558 (2014년 10월 10일 출원), 및 US 가출원 일련 번호: 61/915,148, 61/915,150, 61/915,153, 61/915,203, 61/915,251, 61/915,301, 61/915,267, 61/915,260, 및 61/915,397 (각각 2013년 12월 12일 출원); 61/757,972 및 61/768,959 (2013년 1월 29일 및 2013년 2월 25일 출원); 62/010,888 및 62/010,879 (둘 모두 2014년 6월 11일 출원); 62/010,329, 62/010,439 및 62/010,441 (각각 2014년 6월 10일 출원); 61/939,228 및 61/939,242 (각각 2014년 2월 12일 출원); 61/980,012 (2014년 4월 15일 출원); 62/038,358 (2014년 8월 17일 출원); 62/055,484, 62/055,460 및 62/055,487 (각각 2014년 9월 25일 출원); 및 62/069,243 (2014년 10월 27일 출원). 또한 하기를 참조한다: PCT 출원, 특히, 미국 출원 번호 PCT/US14/41806 (2014 년 6 월 10 일에 출원됨). 하기를 참조한다: 미국 가출원 61/930,214 (2014 년 1 월 22 일에 출원됨). 또한 하기를 참조한다: PCT 출원, 특히, 미국 출원 번호 PCT/US14/41806 (2014 년 6 월 10 일에 출원됨).

[0424] 또한 하기가 언급된다: 미국 출원 62/180,709, 17-Jun-15, PROTECTED GUIDE RNAS (PGRNAS); 미국 출원 62/091,455, 출원일 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); 미국 출원 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); 미국 출원 62/091,462, 12-Dec-14, 62/096,324, 23-Dec-14, 62/180,681, 17-Jun-2015, 및 62/237,496, 5-Oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; 미국 출원 62/091,456, 12-Dec-14 및 62/180,692, 17-Jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; 미국 출원 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); 미국 출원 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; 미국 출원 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; 미국 출원 62/098,059, 30-Dec-14, 62/181,641, 18-Jun-2015, 및 62/181,667, 18-Jun-2015, RNA-TARGETING SYSTEM; 미국 출원 62/096,656, 24-Dec-14 및 62/181,151, 17-Jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; 미국 출원 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; 미국 출원 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; 미국 출원 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; 미국 출원 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; 미국 출원 61/939,154, 12-FEB-14, SYSTEMS, METHODS AND

COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; 미국 출원 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; 미국 출원 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; 미국 출원 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; 미국 출원 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; 미국 출원 62/054,675, 24-Sep-14 및 62/181,002, 17-Jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; 미국 출원 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; 미국 출원 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); 미국 출원 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; 미국 출원 62/087,475, 4-Dec-14 및 62/181,690, 18-Jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; 미국 출원 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; 미국 출원 62/087,546, 4-Dec-14 및 62/181,687, 18-Jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; 및 미국 출원 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[0425] 2015 년 6 월 18 일에 출원된 미국 출원 번호 62/181,659 및 2015 년 8 월 19 일에 출원된 번호 62/207,318 (ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION) 가 언급된다. 2015 년 6 월 18 일에 출원된 미국 출원 번호 62/181,659 및 2015 년 8 월 19 일에 출원된 번호 62/207,318 (ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION) 가 언급된다. 미국 출원 62/181,663, 18-Jun-2015 및 62/245,264, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, 미국 출원 62/181,675, 18-Jun-2015, 62/285,349, 22-Oct-2015, 62/296,522, 17-Feb-2016, 및 62/320,231, 8-Apr-2016, NOVEL CRISPR ENZYMES AND SYSTEMS, 미국 출원 62/232,067, 24-Sep-2015, 미국 출원 14/975,085, 18-Dec-2015, 유럽 출원 번호 16150428.7, 미국 출원 62/205,733, 16-Aug-2015, 미국 출원 62/201,542, 5-Aug-2015, 미국 출원 62/193,507, 16-Jul-2015, 및 미국 출원 62/181,739, 18-Jun-2015 (발명의 명칭: NOVEL CRISPR ENZYMES AND SYSTEMS) 및 미국 출원 62/245,270, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS 이 언급된다. 미국 출원 61/939,256, 12-Feb-2014, 및 WO 2015/089473 (PCT/US2014/070152), 12-Dec-2014 (발명의 명칭: ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION) 이 또한 언급된다. PCT/US2015/045504, 15-Aug-2015, 미국 출원 62/180,699, 17-Jun-2015, 및 미국 출원 62/038,358, 17-Aug-2014 (발명의 명칭: GENOME EDITING USING CAS9 NICKASES) 이 또한 언급된다.

[0426] 또한, 계면활성제, 인지질, 생분해성 중합체, 지단백질 및 알콜을 포함하거나 또는 그로 본질적으로 이루어지거나 또는 그로 이루어진 혼합물과 sgRNA 및 Cpf1 단백질 (및 임의로 HDR 주형)을 포함하는 혼합물을 혼합하는 단계를 포함하는 sgRNA-및-Cpf1 단백질을 제조하는 방법 및 이러한 방법에 의한 입자에 대해서, 다음의 문헌을 언급하고, 참조로 본 명세서에 편입된다: PCT 출원 PCT/US14/70057 (대리인 참조번호 47627.99.2060 및 BI-2013/107) (발명의 명칭: "DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS) (다음의 US 가출원 중 하나 이상 또는 전부의 우선권을 주장함: 62/054,490 (2014년 9월 24일 출원); 62/010,441 (2014년 6월 10일 출원); 및 61/915,118, 61/915,215 및 61/915,148 (각각 2013년 12월 12일 출원) ("the Particle Delivery PCT"), 및 PCT 출원 PCT/US14/70127 (대리인 참조 번호 47627.99.2091 및 Bi-2013/101)(발명의 명칭: "DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING") (다음의 US 가출원 중 하나 이상 또는 전부의 우선권을 주장함: 61/915,176; 61/915,192; 61/915,215; 61/915,107, 61/915,145; 61/915,148; 및 61/915,153 (각각 2013년 12월 12일 출원) ("the Eye PCT"). 예를 들어, Cpf1 단백질 및 sgRNA 는 적합하게, 예를 들어 3:1 내지 1:3 또는 2:1 내지 1:2 또는 1:1 의 몰비율로, 적합한 온도, 예를 들어 15-30℃, 예를 들어 20-25℃, 예를 들어 실온에서, 적합한 시간, 예를 들어 15 내지 45분, 예컨대 39 분간, 유리하게 멸균된, 뉴클레아제 무함유 완충액, 예를 들어 1×PBS 에서 함께

혼합되었다. 별도로, 하기와 같은 또는 하기를 포함하는 입자 성분: 계면활성제, 예를 들어, 양이온성 지질, 예를 들어, 1,2-디올레오일-3-트리메틸암모늄-프로판 (DOTAP); 인지질, 예를 들어, 디미리스토일포스파티딜콜린 (DMPC); 생분해성 중합체, 예컨대 에틸렌-글리콜 중합체 또는 PEG, 및 지단백질, 예컨대 저밀도 지단백질, 예를 들어 콜레스테롤과 같은 또는 이를 포함하는 입자 성분을 알콜, 유리하게 C1-6 알킬 알콜, 예컨대 메탄올, 에탄올, 이소프로판올, 예를 들어 100% 에탄올에 용해시켰다. 2 가지 용액을 함께 혼합하여 Cas9-sgRNA 복합체를 함유하는 입자를 형성시켰다. 따라서, sgRNA 는 입자에 전체 복합체를 제제화시키기 전에 Cpf1 단백질과 예비 복합체화될 수 있다. 제제는 세포로 핵산의 전달을 촉진하는 것으로 알려진 상이한 성분의 상이한 몰비로 제조될 수 있다 (예를 들어, 1,2-디올레오일-3-트리메틸암모늄프로판 (DOTAP), 1,2-디테트라데카노일-sn-글리세로-3-포스포콜린 (DMPC), 폴리에틸렌 글리콜 (PEG), 및 콜레스테롤). 예를 들어, DOTAP : DMPC : PEG : 콜레스테롤 몰비율은 DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0; 또는 DOTAP 90, DMPC 0, PEG 10, 콜레스테롤 0; 또는 DOTAP 90, DMPC 0, PEG 5, 콜레스테롤 5일 수 있다. DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0. 따라서 이 출원은 sgRNA, Cpf1 단백질 및 입자를 형성하는 성분을 혼합하는 것 뿐 아니라; 이러한 혼합으로부터의 입자로 이해된다. 본 발명의 측면은 예를 들어, 예로서 본 발명에서와 같은 sgRNA 및/또는 Cpf1 를 포함하는 혼합물 및 입자를 형성하는 성분을 혼합하여, 예를 들어 입자 전달 PCT에서 처럼, 입자를 형성하기 위한, 입자 전달 PCT의 것과 유사한 방법을 사용한 입자 및 그러한 혼합 유래의 입자 (또는 물론, 본 발명 에서와 같은 sgRNA 및/또는 Cpf1 를 포함하는 다른 입자) 를 포함할 수 있다. Cpf1 및 C2c1은 둘 모두가 구조 유사성을 공유하는 V형 CRISPR-Cas 단백질이다. PAM의 원위 말단에서 블런트 절단부를 생성시키는 Cas9와 달리, Cpf1 및 C2c1은 PAM의 원위 말단에서 스태거드 절단부를 생성시킨다. 따라서, C2c1을 사용한 유사한 시스템이 고려될 수 있다.

[0427] 대상 발명은 결과 또는 데이터의 전송이 존재하는 것인 검색 프로그램의 일부로서 사용될 수 있다. 컴퓨터 시스템 (또는 디지털 장치)는 결과를 수신, 전송, 디스플레이 및/또는 저장하고, 데이터 및/또는 결과를 분석하고/하거나, 결과 및/또는 데이터 및/또는 분석의 보고서를 생성시키는데 사용될 수 있다. 컴퓨터 시스템은 임의로 고정 매체를 구비한 서버에 접속될 수 있는, 매체 (예를 들어, 소프트웨어) 및/또는 네트워크 포트 (예를 들어, 인터넷 유래)로부터의 명령어를 판독할 수 있는 논리적인 장비로서 이해될 수 있다. 컴퓨터 시스템은 하나 이상의 CPU, 디스크 드라이브, 입력 장치 예컨대 키보드 및/또는 마우스, 및 디스플레이 (예를 들어, 모니터)를 포함할 수 있다. 데이터 통신, 예컨대 명령어 또는 보고서의 전송은 로컬 또는 원격 위치에 서버로 통신 매체를 통해서 달성될 수 있다. 통신 매체는 전송 및/또는 수신 데이터의 임의 수단을 포함할 수 있다. 예를 들어, 통신 매체는 네트워크 접속, 무선 접속, 또는 인터넷 접속일 수 있다. 이러한 접속은 월드 와이드 웹 상에서 통신을 위해 제공된다. 본 발명과 관련된 데이터는 수신자에 의한 접수 및/또는 고찰을 위해서 이러한 네트워크 또는 접속 (또는 제한없이 인쇄물같은 실물 보고서의 발송을 포함하여, 정보 전송을 위한 임의의 다른 적합한 수단)으로 전송할 수 있다는 것을 고려한다. 수신자는 개체, 또는 전자 시스템 (예를 들어, 하나 이상의 컴퓨터, 및/또는 하나 이상의 서버)일 수 있지만 이에 제한되지 않는다. 일부 구현예에서, 컴퓨터 시스템은 하나 이상의 프로세서를 포함한다. 프로세스는 하나 이상의 컨트롤러, 계산 유닛, 및/또는 컴퓨터 시스템의 다른 유닛과 결합될 수 있거나, 또는 바람직하다면 펌웨어로 삽입될 수 있다. 소프트웨어로 구현되면, 루틴은 임의의 컴퓨터 판독가능 메모리 예컨대 RAM, ROM, 플래시 메모리, 자기 디스크, 레이저 디스크, 또는 다른 적합한 저장 매체에 저장될 수 있다. 유사하게, 이러한 소프트웨어는 예를 들어 통신 채널 예컨대 전화선, 인터넷, 무선 접속 등에서, 또는 운반가능 매체, 예컨대 컴퓨터 판독가능 디스크, 플래시 드라이브 등을 통하는 것을 포함하여, 임의의 기지 전달 방법을 통해 컴퓨터 장치로 전달될 수 있다. 다양한 단계들이 하드웨어, 펌웨어, 소프트웨어, 또는 하드웨어, 펌웨어 및/또는 소프트웨어의 이므이 조합에서 구현될 수 있는, 다양한 블록, 운영, 도구, 모듈, 및 기술로서 구현될 수 있다. 하드웨어로 구현될 때, 일부 또는 모든 블록, 운영, 기술 등은 예를 들어 주문 IC (integrated circuit), ASIC (application specific integrated circuit), FPGA (field programmable logic array), PLA (programmable logic array) 등에서 구현될 수 있다. 고객-서버, 관련 데이터베이스 아키텍처는 본 발명이 구현예에서 사용될 수 있다. 고객 서버 아키텍처는 네트워크 상의 각 컴퓨터 또는 프로세스가 고객 도는 서버인 네트워크 아키텍처이다. 서버 컴퓨터는 전형적으로 디스크 드라이브 (파일 서버), 프린터 (프린트 서버), 또는 네트워크 트래픽 (네트워크 서버)를 관리하도록 전용된 강력한 컴퓨터이다. 고객 컴퓨터는 사용자가 어플리케이션을 비롯하여, 본 명세서에 개시된 바와 같은 예시의 출력 장치를 실행하는 PC (개인 컴퓨터) 또는 단말기를 포함한다. 고객 컴퓨터는 리소스, 예컨대 파일, 장치, 및 심지어 프로세싱 파워를 위한 서버 컴퓨터에 의존한다. 본 발명의 일부 구현예에서, 서버 컴퓨터는 모든 데이터베이스 기능성을 다룬다. 고객 컴퓨터는 모든 프론트-엔드 데이터 관리를 취급하고 또한 사용자로부터 데이터 입력을 수신할 수 있는 소프트웨어를 구비할 수 있다.

컴퓨터-실행가능한 코드를 포함한 기계 판독가능 매체는 제한없이, 유형 저장 매체, 반송파 매체, 또는 물리적 전송 매체를 포함하여, 많은 형태를 취할 수 있다. 비휘발성 저장 매체는 예를 들어 광학 또는 자기 디스크, 예컨대 임의 컴퓨터(들) 등의 임의의 저장 장치를 포함하고, 예컨대 도면에 도시된 데이터베이스 등을 구현하는데 사용될 수 있다. 휘발성 저장 매체는 이러한 컴퓨터 플랫폼의 주요 메모리같은, 동적 메모리를 포함한다. 유형 전송 매체는 컴퓨터 시스템 내에 버스를 포함하는 와이어를 포함하여, 동축 케이블; 구리 와이어 및 광섬유를 포함한다. 반송파 전송 매체는 전기 또는 전자기 신호, 또는 음향 또는 광파 예컨대 라디오 주파수 (RF) 또는 적외선 (IR) 데이터 통신 동안 발생되는 것의 형태를 취할 수 있다. 그러므로 컴퓨터-판독가능 매체의 통상의 형태는 예를 들어 플로피 디스크, 플렉시블 디스크, 하드 디스크, 자기 테이프, 임의의 다른 자기 매체, CD-ROM, DVD 또는 DVD-ROM, 임의의 다른 광학 매체, 펀치 카드 페이퍼 테이프, 홀 패턴을 갖는 임의의 다른 물리적 저장 매체, RAM, ROM, PROM 및 EPROM, FLASH-EPROM, 임의의 다른 메모리 칩 또는 카트리지, 반송파 전송 데이터 또는 명령, 케이블 또는 링크 전송 예컨대 판송파, 또는 컴퓨터가 프로그램된 코드 및/또는 데이터를 판독할 수 있는 임의의 다른 매체를 포함한다. 많은 이들 형태의 컴퓨터 판독가능 매체는 실행을 위해 프로세서로 하나 이상의 명령의 하나 이상의 배열을 운반하는 것을 포함할 수 있다. 따라서, 본 발명은 본 명세서에 논의된 임의 방법을 수행하고 그로부터의 데이터 및/또는 결과 및/또는 이의 분석을 비롯하여, 중간체를 포함하여 본 명세서에 논의된 임의 방법을 수행한 산물을 저장 및/또는 전송하는 것을 포괄한다.

[0428] Cas12b (C2c1)

[0429] 본 발명은 C2c1 (V-B형; Cas12b) 이펙터 단백질 및 오솔로그를 제공한다. 용어 "오솔로그 (orthologue)" (본원에서 "오솔로그 (ortholog)" 로도 나타냄) 및 "상동체 (homologue)" (본원에서 "상동체 (homolog)" 로도 나타냄) 는 당업계에 널리 공지되어 있다. 추가 지침에 의해서, 본원에서 사용되는 단백질의 "상동체"는 이의 상동체인 단백질과 동일하거나 유사한 기능을 수행하는 동일 종의 단백질이다. 상동성 단백질은 구조적으로 관련될 필요가 없거나, 오직 부분적으로 구조적으로 관련된다. 본원에서 사용되는 바와 같은 단백질의 "오솔로그" 는 이의 오솔로그인 단백질과 동일하거나 유사한 기능을 수행하는 상이한 종의 단백질이다.
오솔로그성 단백질은 구조적으로 관련될 필요가 없거나, 오직 부분적으로 구조적으로 관련된다. 상동체 및 오솔로그는 상동체 모델링을 통해 확인할 수 있다 (참조: 예를 들어 Greer, Science vol. 228 (1985) 1055, 및 Blundell et al. Eur J Biochem vol 172 (1988), 513) 또는 "structural BLAST" (Dey F, Cliff Zhang Q, Petrey D, Honig B. Toward a "structural BLAST" : using structural relationships to infer function. Protein Sci. 2013 Apr;22(4):359-66. doi: 10.1002/pro.2225.). 또한, CRISPR-Cas 유전자좌 분야에서 적용을 위한 [Shmakov et al. (2015)]를 참조한다. 그러나 상동성 단백질은 구조적으로 관련될 필요가 없거나, 또는 오직 부분적으로 구조적으로 관련된다.

[0430] C2c1 유전자는 여러 다양한 박테리아 게놈에서, 전형적으로는 cas1, cas2 및 cas4 유전자 및 CRISPR 카세트를 갖는 동일한 유전자좌에서 발견된다. 따라서, 이러한 추정 신규 CRISPR-Cas 시스템의 레이아웃은 II-B형의 경우와 유사한 것으로 나타난다. 또한 Cas9 와 유사하게, C2c1 단백질은 활성 RuvC-유사 뉴클레아제, 아르기닌-풍부 영역, 및 Zn 핑거 (Cas9 에 부재함) 를 함유한다.

[0431] 본 발명은 V-A 아형으로서 나타내는 C2c1 유전자좌에서 유래하는, C2c1 (Cas12b) 이펙터 단백질의 용도를 포함한다. 여기서 이러한 이펙터 단백질은 또한 " C2c1p", 예를 들어 C2c1 단백질로서 지칭된다 (그리고 이러한 이펙터 단백질 또는 C2c1 단백질 또는 C2c1 유전자좌에서 유래하는 단백질은 또한 "CRISPR 효소" 로 불린다).
현재, V-B 아형 유전자좌는 cas1-Cas4 융합, cas2, C2c1 로 나타낸 별개의 유전자 및 CRISPR 어레이를 포함한다. C2c1 (CRISPR-연관 단백질 C2c1) 은 Cas9 의 특징적인 아르기닌-풍부 클러스터에 대응하는 부분과 함께 Cas9 의 해당하는 도메인에 상응하는 RuvC-유사 뉴클레아제 도메인을 함유하는 거대 단백질 (약 1100 - 1300 개 아미노산) 이다. 그러나, C2c1 은 모든 Cas9 단백질에 존재하는 HNH 뉴클레아제 도메인이 결핍되어 있고, RuvC-유사 도메인은 HNH 도메인을 포함하는 긴 삽입물을 함유하는 Cas9 와 달리 C2c1 서열에서 연속적이다. 따라서, 특정 구현예에서, CRISPR-Cas 효소는 RuvC-유사 뉴클레아제 도메인만을 포함한다.

[0432] C2c1 (Cas12b 로도 공지됨) 단백질은 RNA 가이드 뉴클레아제이다. 이의 절단은 tracr RNA 에 의존하여, 가이드 서열 및 직접 반복부를 포함하는 가이드 RNA 를 모집하는데, 여기서 가이드 서열은 표적 뉴클레오티드 서열과 하이브리드화하여 DNA/RNA 헤테로듀플렉스를 형성한다. 현재 연구를 기반으로 하여, C2c1 뉴클레아제 활성은 또한 PAM 서열의 인식에 의존을 요구한다. C2c1 PAM 서열은 T-풍부 서열일 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 특정 구현예에서, PAM 서열은 5' TTC 3' 이다. 특정 구현예에서, PAM 은 플라스모듐 팔시파럼 (Plasmodium

falciparum) 의 서열 내에 있다.

[0433] C2c1 은 표적 유전자좌에서, 5' 오버행을 갖는 스태거드 (staggered) 절단부, 또는 표적 서열의 PAM 원위 측에서 " 스티키 말단 (sticky end)" 을 생성한다. 일부 구현예에서, 5' 오버행은 7 nt 이다. 참조: Lewis and Ke, Mol Cell. 2017 Feb 2;65(3):377-379.

[0434] 본 발명은 또한 C2c1 이펙터 단백질의 사용을 포괄하는 CRISPR-C2c1 시스템을 제공한다. 일부 구현예에서, 시스템은 I. CRISPR-Cas 시스템 RNA 폴리뉴클레오티드 서열로서, (a) 직접 반복부 폴리뉴클레오티드 및 (b) 표적 서열과 하이브리드화할 수 있는 가이드 서열 폴리뉴클레오티드를 포함하는 것인 폴리뉴클레오티드 서열; II. Tracr RNA 폴리뉴클레오티드; 및 III. 임의로 적어도 하나 이상의 핵 국재화 서열을 포함하는, C2c1을 코딩하는 폴리뉴클레오티드 서열을 포함하고, 여기서 직접 반복부 서열은 가이드 서열과 하이브리드화하여 표적 서열로 CRISPR 복합체의 서열-특이적 결합을 유도시키고, CRISPR 복합체는 (1) 표적 서열과 하이브리드화하거나 또는 하이브리드화가능한 가이드 서열, 및 (2) 직접 반복부 서열과 복합체를 형성하는 CRISPR 단백질을 포함하고, CRISPR 단백질을 코딩하는 폴리뉴클레오티드 서열은 DNA 또는 RNA이다. Tracr은 crRNA에 융합될 수 있다. 예를 들어, tracr RNA는 직접 반복부의 5' 말단에서 crRNA와 융합될 수 있다. 본 명세서에서 사용되는 용어 crRNA는 CRISPR RNA라고 하고 본 명세서에서 용어 gRNA 또는 가이드 RNA와 상호교환적으로 사용할 수 있다. Tracr은 gRNA의 crRNA에 융합될 때, 이는 단일 가이드 RNA 또는 합성 가이드 RNA (sgRNA)라고 할 수 있다.

[0435] C2c1은 Cas9에 의해 생성되는 PAM의 근위 말단에서의 절단과 대조적으로, PAM의 원위 말단에서 이중 가닥 파손을 생성시킨다 (Jinek et al., 2012; Cong et al., 2013). Cpf1 돌연변이된 표적 서열은 단일 gRNA에 의한 반복된 절단에 감수성일 수 있어서, HDR 매개 게놈 편집에서 Cpf1의 적용을 촉진한다고 제안된다 (Front Plant Sci. 2016 Nov 14;7:1683). Cpf1 및 C2c1은 둘 모두가 구조 유사성을 공유하는 V형 CRISPR Cas 단백질이다. C2c1 처럼, Cpf1은 (PAM의 근위 말단에 블런트 절단부를 생성하는 Cas9와 대조적으로) PAM의 원위 말단에 스태거드 이중 가닥 파손을 생성시키지만, Cpf1과 달리, C2c1 시스템은 tracrRNA를 적용한다. 따라서, 일정 구현에에서, 관심 유전자좌는 상동성 지정 복구 (HR 또는 HDR)을 통해 CRISPR-C2c1 복합체에 의해 변형된다. 일정 구현예에서, 관심 유전자좌는 HR 독립적인 CRISPR-C2c1 복합체에 의해 변형된다. 일정 구현에에서, 관심 유전자좌는 비상동성 말단 결합 (NHEJ)을 통해 CRISPR-C2c1 복합체에 의해 변형된다.

[0436] C2c1는 Cas9에 의해 생성되는 블런트 말단부와 대조적으로, 5' 오버행을 갖는 스태거드 절단부를 생성시킨다 (Garneau et al., Nature. 2010;468:67-71; Gasiunas et al., Proc Natl Acad Sci U S A. 2012;109:E2579-2586). 절단 산물의 이러한 구조는 포유동물 게놈으로 비상동성 말단 결합 (NHEJ)-기반 유전자 삽입을 촉진하기 위해 특히 유리할 수 있다 (Maresca et al., Genome research. 2013;23:539-546).

[0437] 특정 구현예에서, 이펙터 단백질은 알리시클로바실러스 (Alicyclobacillus), 데술포비브리오 (Desulfovibrio), 데술포나트로늄 (Desulfonatronum), 오피투타세 (Opitutaceae), 투베리바실러스 (Tuberibacillus), 바실러스 (Bacillus), 브레비바실러스 (Brevibacillus), 칸디다투스 (Candidatus), 데술파티랍디움 (Desulfatirhabdium), 시트로박터 (Citrobacter), 엘루시마이크로비아 (Elusimicrobia), 메틸로박테리움 (Methylobacterium), 옴니트로피카 (Omnitrophica), 피시스페레 (Phycisphaerae), 플랑크토마이세테스 (Planctomycetes), 스피로카에테스 (Spirochaetes), 및 베루코마이크로비아세아에 (Verrucomicrobiaceae), 렌티스파에리아 (Lentisphaeria), 라세이엘라 (Laceyella) 를 포함하는 속으로부터의 유기체 유래 또는 그로부터 기원하는 C2c1 이펙터 단백질이다.

[0438] 추가 특정 구현예에서, C2c1 이펙터 단백질은 알리시클로바실러스 악시도테레스트리스 (Alicyclobacillus acidoterrestris) (예를 들어, ATCC 49025), 알리시클로바실러스 콘타미난스 (Alicyclobacillus contaminans) (예를 들어, DSM 17975), 알리시클로바실러스 마크로스포란지두스 (Alicyclobacillus macrosporangiidus) (예를 들어, DSM 17980), 바실러스 히사시이 (Bacillus hisashii) 균주 C4, 칸디다투스 린도우박테리아 박테리움 (Candidatus Lindowbacteria bacterium) RIFCSPLOWO2, 데술포비브리오 이노피나투스 (Desulfovibrio inopinatus) (예를 들어, DSM 10711), 데술포나트로늄 티오디스무탄스 (Desulfonatronum thiodismutans) (예를 들어, 균주 MLF-1 또는 genbank 수탁 번호 WP_031386437), 엘루시미크로비아 박테리움 (Elusimicrobia bacterium) RIFOXYA12, 옴니트로피카 WOR_2 박테리움 (Omnitrophica WOR_2 bacterium) RIFCSPHIGHO2, 오피투타세아에 박테리움 (Opitutaceae bacterium) TAV5 또는 genbank 수탁 번호 WP_009513281, 피시스파에라에 박테리움 (Phycisphaerae bacterium) ST-NAGAB-D1, 플란크토마이세테스 박테리움 (Planctomycetes bacterium) RBG_13_46_10, 스피로카에테스 박테리움 (Spirochaetes bacterium) GWB1_27_13, 베루코미크로비아세아에 박테

리움 (Verrucomicrobiaceae bacterium) UBA2429, 투베리바실러스 칼리두스 (Tuberibacillus calidus) (예를 들어, DSM 17572), 바실러스 써모아밀로보란스 (Bacillus thermoamylovorans) (예를 들어, 균주 B4166), 브레비바실러스 sp. (Brevibacillus sp.) CF112, 바실러스 (Bacillus) sp. NSP2.1, 데술파티랍디움 부티라티보란스 (Desulfatirhabdium butyrativorans) (예를 들어, DSM 18734 또는 genbanck 수탁 번호 WP_028326052), 알리시클로바실러스 허바리우스 (Alicyclobacillus herbarius) (예를 들어, DSM 13609), 시트로박터 프레운디 (Citrobacter freundii) (예를 들어, ATCC 8090), 브레비바실러스 아그리 (Brevibacillus agri) (예를 들어, BAB-2500), 메틸로박테리움 노둘란스 (Methylobacterium nodulans) (예를 들어, ORS 2060 또는 genbank 수탁 번호 WP_043747912), 알리시클로바실러스 카케가웬시스 (Alicyclobacillus kakegawensis) (예를 들어, genbank 수탁 번호 WP_067936067), 바실러스 (Bacillus) sp. V3-13 (예를 들어, genbank 수탁 번호 WP_101661451), 렌티스파에리아 박테리움 (Lentisphaeria bacterium) (예를 들어, DCFZ01000012 유래), 라세이엘라_ 세디미니스 (Laceyella_sediminis) (예를 들어, genbank 수탁 번호 WP_106341859)에서 선택되는 종으로부터 유래하거나 또는 기원한다.

[0439] 일정 구현예에서, C2c1 이펙터 단백질은 알리시클로바실러스, 바실러스, 데술파티랍디움, 데술포나트로늄, 렌티스파에리아, 라세이엘라, 메틸로박테리움, 또는 오피투타세아에 속으로부터 선택되는 종으로부터 유래되거나 또는 기원한다.

[0440] 일정 구현예에서, C2c1 이펙터 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp._V3-13, 데술파티랍디움 부티라티보란스, 데술포나트로늄 티오디스뮤탄스, 렌티스파에리아 박테리움, 라세이엘라 세디미니스, 메틸로박테리움 노둘란스, 또는 오피투타세아에 박테리움으로부터 선택된 종으로부터 유래하거나 또는 기원한다.

[0441] 일정 구현예에서, C2c1 이펙터 단백질은 야생형 서열이 WP_067936067의 서열에 상응하는 알리시클로바실러스 카케가웬시스, 야생형 서열이 Wp_101661451의 서열에 상응하는 바실러스 sp._V3-13, 야생형 서열이 Wp_028326052의 서열에 상응하는 데술파티랍디움 부티라티보란스, 야생형 서열이 Wp_031386437의 서열에 상응하는 데술포나트로늄 티오디스뮤탄스, 야생형 서열이 DCFZ01000012의 서열에 상응하는 렌티스파에리아 박테리움, 야생형 서열이 Wp_106341859의 서열에 상응하는 라세이엘라 세디미니스, 야생형 서열이 Wp_043747912의 서열에 상응하는 메틸로박테리움 노둘란스, 또는 야생형 서열이 Wp_009513281의 서열에 상응하는 오피투타세아에 박테리움으로부터 선택되는 종으로부터 유래하거나 또는 기원한다.

[0442] 일정 구현예에서, C2c1 이펙터 단백질은 표 1로부터 선택되는 종으로부터 유래하거나 또는 기원하고 표 1에 표시된 바와 같은 야생형 서열을 갖는다. 본 명세서의 다른 곳에 기술된 바와 같은 돌연변이되거나 또는 절두된 Cas12b 단백질은 표시된 서열로부터 벗어날 수 있다는 것을 이해할 것이다.

표 1 - Cas12b 오솔로그

종	서열
알리시클로바실러스 카케가웬시스 (SEQ ID NO:379)	MAVKSIKVKLRLSECPDILAGMWQLHRATNAGVRYYTEWVSLMRQ EILYSRGPDGGQQCYMTAEDCQRELLRRLRNRQLHNGRQDQPGTD ADLLAISRRLYEILVLQSIGKRGDAQQIASSFLSPLVDPNSKGGRGEA KSGRKPAWQKMRDQGDPRWVAAREKYEQRKAVDPSKEILNSLDAL GLRPLFAVFTETYRSGVDWKPLGKSQGVRTWDRDMFQQALERLMS WESWNRRVGEEYARLFQQKMKFEQEHFAEQSHLVKLARALEADM RAASQGFEAKRGTAHQITRRALRGADRVFEIWKSIPEEALFSQYDEVI RQVQAEKRRDFGSHDLFAKLAEPKYQPLWRADETFLTRYALYNGV LRDLEKARQFATFTLPDACVNPIWTRFESSQGSNLHKYEFLFDHLGP GRHAVRFQRLLVVESEGAKERDSVVVPVAPSGQLDKLVLREEEKSS VALHLHDTARPDGFMAEWAGAKLQYERSTLARKARRDKQGMRSW RRQPSMLMSAAQMLEDAKQAGDVYLNISVRVKSPSEVRGQRRPPY AALFRIDDKQRRVTVNYNKLSAYLEEHPDKQIPGAPGLLSGLRVMS VDLGLRTSASISVFRVAKKEEVEALGDGRPPHYYPIHGTDDLVAVHE RSHLIQMPGETETKQLRKLREERQAVLRPLFAQLALLRLLVRCGAAD ERIRTRSWQRLTKQGREFTKRLTPSWREALELELTRLEAYCGRVPDD EWSRIVDRTVIALWRRMGKQVRDWRKQVKSGAKVKVKGYQLDVV GGNSLAQIDYLEQQYKFLRRWSFFARASGLVVRADRESHFAVALRQ HIENAKRDRLKKLADRILMEALGYVYEASGPREGQWTAQHPPCQLII LEELSAYRFSDDRPPSENSKLMAWGHRGILEELVNQAQVHDVLVGT VYAAFSSRFDARTGAPGVRCRRVPARFVGATVDDSLPLWLTEFLDK HRLDKNLLRPDDVIPTGEGEFLVSPCGEEAARVRQVHADINAAQNL QRRLWQNFDITELRLRCDVKMGGEGTVLVPRVNNARAKQLFGKKV LVSQDGVTFFERSQTGGKPHSEKQTDLTDKELELIAEADEARAKSVV LFRDPSGHIGKGHWIRQREFWSLVKQRIESHTAERIRVRGVGSSLD
바실러스 sp._V3-13 (SEQ ID NO:380)	MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYR QEAIGDKTKEAYQAELINIIRNQQRNNGSSEEHGSDQEILALLRQLYE LIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNAGRKPRWKRL KEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQ

[0443]

	KDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQ LKEKTESYYKEHLTGGEEWIEKIRKFEKERNMELEKNAFAPNDGYFI TSRQIRGWDRVYEKWSKLPESASPEELWKVVAEQQNKMSEGFGDP KVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTLP DAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWI EKENIEIPLAPSIQFNRQIKLKQHVKGKQEISFSDYSSRISLDGVLGGS RIQFNRKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIG KALKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLLEGMRVM SIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKRSFLLN LPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTPDER KKAIHKLMEIVQSYDSWTASQKEVWEKELNLLTNMAAFNDEIWKE SLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRR LLISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLII MTALGFKYDKEEKDRYKRWKETYPACQIILFENLNRYLFNLDRSRR ENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAP GIRCHALTEEDLKAGSNTLKRLIEDGFINESELAYLKKGDIIPSQGGEL FVTLSKRYKKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPC QLARMGEDKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEK MKIKTDTTFDLQDLDGFEDISKTIELAQEQQKKYLTMFRDPSGYFFN NETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL
데술파틸하브듐 부티라티보란스 (SEQ ID NO:381)	MPLSNNPPVTQRAYTLRLRGADPSDLSWREALWHTHEAVNKGAKV FGDWLLTLRGGLDHTLADTKVKGGKGKPDRDPTPEERKARRILLAL SWLSVESKLGAPSSYIVASGDEPAKDRNDNVVSALEEILQSRKVAKS EIDDWKRDCSASLSAAIRDDAVWVNRSKVFDEAVKSVGSSLTREEA WDMLERFFGSRDAYLTPMKDPEDKSSETEQEDKAKDLVQKAGQWL SSRYGTSEGADFCRMSDIYGKIAAWADNASQGGSSTVDDLVSELRQ HFDTKESKATNGLDWIIGLSSYTGHTPNPVHELLRQNTSLNKSHLDD LKKKANTRAESCKSKIGSKGQRPYSDAILNDVESVCGFTYRVDKDG QPVSVADYSKYDVDYKWGTARHYIFAVMLDHAARRISLAHKWIKR AEAERHKFEEDAKRIANVPARAREWLDSFCKERSVTSGAVEPYRIRR RAVDGWKEVVAAWSKSDCKSTEDRIAAARALQDDSEIDKFGDIQLF EALAEDDALCVWHKDGEATNEPDFQPLIDYSLAIEAEFKKRQFKVP AYRHPDELLHPVFCDFGKSRWKINYDVHKNVQAPFYRGLCLTLWT GSEIKPVPLCWQSKRLTRDLALGNNHRNDAASAVTRADRLGRAASN VTKSDMVNITGLFEQADWNGRLQAPRQQLEAIAVVRDNPRLSEQER NLRMCGMIEHIRWLVTFSVKLQPQGPWCAYAEQHGLNTNPQYWPH ADTNRDRKVHARLILPRLPGLRVLSVDLGHRYAAACAVWEAVNTE TVKEACQNVGRDMPKEHDLYLHIKVKKQGIGKQTEVDKTTIYRRIG ADTLPDGRPHPAPWARLDRQFLIKLQGEEKDAREASNEEIWALHQM ECKLDRTKPLIDRLIASGWGLLKRQMARLDALKELGWIPAPDSSENL SREDGEAKDYRESLAVDDLMFSAVRTLRLALQRHGNRARIAYYLIS EVKIRPGGIQEKLDENGRIDLLQDALALWHELFSSPGWRDEAAKQL WDSRIATLAGYKAPEENGDNVSDVAYRKKQQVYREQLRNVAKTLS GDVITCKELSDAWKERWEDEDQRWKKLLRWFKDWVLPSGTQANN ATIRNVGGLSLSRLATITEFRRKVQVGFFTRLRPDGTRHEIGEQFGQK TLDALELLREQRVKQLASRIAEAALGIGSEGGKGWDGGKRPRQRIN

[0444]

	DSRFAPCHAVVIENLANYRPDETRTRLENRRLMTWSASKVHKYLSE ACQLNGLYLCTVSAWYTSRQDSRTGAPGIRCQDVSVREFMQSPFWR KQVKQAEAKHDENKGDARERFLCELNKTWKAKTPAEWKKAGFVRI PLRGGEIFVSADSKSPSAKGIHADLNAAANIGLRALTDPDWPGKWW YVPCDPVSFESKMDYVKGCAAVKVGQPLRQPAQTNADGAASKIRK GKKNRTAGTSKEKVYLWRDISAFPLESNEIGEWKETSAYQNDVQYR VIRMLKEHIKSLDNRTGDNVEG
테술포나트로뉴 티오디스뮤탄스 (SEQ ID NO:382)	MVLGRKDDTAELRRALWTTHEHVNLAVAEVERVLLRCRGRSYWTL DRRGDPVHVPESQVAEDALAMAREAQRRNGWPVVGEDEEILLALR YLYEQIVPSCLLDDLGKPLKGDAQKIGTNYAGPLFDSDTCRRDEGKD VACCGPFHEVAGKYLGALPEWATPISKQEFDGKDASHLRFKATGGD DAFFRVSIEKANAWYEDPANQDALKNKAYNKDDWKKEKDKGISS WAVKYIQKQLQLGQDPRTEVRRKLWLELGLLPLFIPVFDKTMVGNL WNRLAVRLALAHLLSWESWNHRAVQDQALARAKRDELAALFLGM EDGFAGLREYELRRNESIKQHAFEPVDRPYVVSGRALRSWTRVREE WLRHGDTQESRKNICNRLQDRLRGKFGDPDVFHWLAEDGQEALWK ERDCVTSFSLLNDADGLLEKRKGYALMTFADARLHPRWAMYEAPG GSNLRTYQIRKTENGLWADVVLLSPRNESAAVEEKTFNVRLAPSGQ LSNVSFDQIQKGSKMVGRCRYQSANQQFEGLLGGAEILFDRKRIANE QHGATDLASKPGHVWFKLTLDVRPQAPQGWLDGKGRPALPPEAKH FKTALSNKSKFADQVRPGLRVLSVDLGVRSFAACSVFELVRGGPDQ GTYFPAADGRTVDDPEKLWAKHERSFKITLPGENPSRKEEIARRAAM EELRSLNGDIRRLKAILRLSVLQEDDPRTEHLRLFMEAIVDDPAKSAL NAELFKGFGDDRFRSTPDLWKQHCHFFHDKAEKVVAERFSRWRTET RPKSSSWQDWRERRGYAGGKSYWAVTYLEAVRGLILRWNMRGRT YGEVNRQDKKQFGTVASALLHHINQLKEDRIKTGADMIIQAARGFV PRKNGAGWVQVHEPCRLILFEDLARYRFRTDRSRRENSRLMRWSHR EIVNEVGMQGELYGLHVDTTEAGFSSRYLASSGAPGVRCRHLVEED FHDGLPGMHLVGELDWLLPKDKDRTANEARRLLGGMVRPGMLVP WDGGELFATLNAASQLHVIHADINAAQNLQRRFWGRCGEAIRIVCN QLSVDGSTRYEMAKAPKARLLGALQQLKNGDAPFHLTSIPNSQKPE NSYVMTPTNAGKKYRAGPGEKSSGEEDELALDIVEQAEELAQGRKT FFRDPSGVFFAPDRWLPSEIYWSRIRRRIWQVTLERNSSGRQERAEM DEMPY
렌티스파에리아 박테리움 (SEQ ID NO:383)	MAVELNRIYQGRVNHVYIFDENQNQVSVDNGDDLLFVHHELYQDAI NYYLVALAAMALDSKDSLFGKFKMQIRAVWNDFYRNGQLRPGLKH SLIRSLGHAAELNTSNGADIAMNLILEDGGIPSEILNAALEHLAEKCT GDVSQLGKTFFPRFCDTAYHGNWDVDAKSFSEKKGRQRLVDALYS LHPVQAVQELAPEIEIGWGGVKTQTGKFFTGDEAKASLKKAISYFLQ DTGKNSPELQEYFSVAGKQPLEQYLGKIDTFPEISFGRISSHQNINISN AMWILKFFPDQYSVDLIKNLIPNKKYEIGIAPQWGDDPVKLSRGKRG YTFRAFTDLAMWEKNWKVFDRAAFSDALKTINQFRNKTQERNDQL KRYCAALNWMDGESSDKKPPVEPADADAVDEAATSVLPILAGDKR WNALLQLQKELGICNDFTENELMDYGLSLRTIRGYQKLRSMMLEKE EKMRAKTADDEEISQALQEIIIKFQSSHRDTIGSVSLFLKLAEPKYFCV WHDADKNQNFASVDMVADAVRYYSYQEEKARLEEPIQITPADARY

[0445]

	SRRVSDLYALVYKNAKECKTGYGLRPDGNFVFEIAQKNAKGYAPA KVVLAFSAPRLKRDGLIDKEFSAYYPPVLQAFLREEEAPKQSFKTTA VILMPDWDKNGKRRILLNFPIKLDVSAIHQKTDHRFENQFYFANNTN TCLLWPSYQYKKPVTWYQGKKPFDVVAVDLGQRSAGAVSRITVST EKREHSVAIGEAGGTQWYAYRKFSGLLRLPGEDATVIRDGQRTEELS GNAGRLSTEEETVQACVLCKMLIGDATLLGGSDEKTIRSFPKQNDKL LIAFRRATGRMKQLQRWLWMLNENGLCDKAKTEISNSDWLVNKNI DNVLKEEKQHREMLPAILLQIADRVLPLRGRKWDWVLNPQSNSFVL QQTAHGSGDPHKKICGQRGLSFARIEQLESLRMRCQALNRILMRKTG EKPATLAEMRNNPIPDCCPDILMRLDAMKEQRINQTANLILAQALGL RHCLHSESATKRKENGMHGEYEKIPGVEPAAFVVLEDLSRYRFSQD RSSYENSRLMKWSHRKILEKLALLCEVFNVPILQVGAAYSSKFSANA IPGFRAEECSIDQLSFYPWRELKDSREKALVEQIRKIGHRLLTFDAKA TIIMPRNGGPVFIPFVPSDSKDTLIQADINASFNIGLRGVADATNLLCN NRVSCDRKKDCWQVKRSSNFSKMVYPEKLSLSFDPIKKQEGAGGNF FVLGCSERILTGTSEKSPVFTSSEMAKKYPNLMFGSALWRNEILKLER CCKINQSRLDKFIAKKEVQNEL
라세이엘라 세디미니스 (SEQ ID NO:384)	MSIRSFKLKIKTKSGVNAEELRRGLWRTHQLINDGIAYYMNWLVLL RQEDLFIRNEETNEIEKRSKEEIQGELLERVHKQQQRNQWSGEVDDQ TLLQTLRHLYEEIVPSVIGKSGNASLKARFFLGPLVDPNNKTTKDVSK SGPTPKWKKMKDAGDPNWVQEYEKYMAERQTLVRLEEMGLIPLFP MYTDEVGDIHWLPQASGYTRTWDRDMFQQAIERLLSWESWNRRVR ERRAQFEKKTHDFASRFSESDVQWMNKLREYEAQQEKSLEENAFAP NEPYALTKKALRGWERVYHSWMRLDSAASEEAYWQEVATCQTAM RGEFGDPAIYQFLAQKENHDIWRGYPERVIDFAELNHLQRELRRAKE DATFTLPDSVDHPLWVRYEAPGGTNIHGYDLVQDTKRNLTLILDKFI LPDENGSWHEVKKVPFSLAKSKQFHRQVWLQEEQKQKKREVVFYD YSTNLPHLGTLAGAKLQWDRNFLNKRTQQQIEETGEIGKVFFNISVD VRPAVEVKNGRLQNGLGKALTVLTHPDGTKIVTGWKAEQLEKWVG ESGRVSSLGLDSLSEGLRVMSIDLGQRTSATVSVFEITKEAPDNPYKF FYQLEGTELFAVHQRSFLLALPGENPPQKIKQMREIRWKERNRIKQQ VDQLSAILRLHKKVNEDERIQAIDKLLQKVASWQLNEEIATAWNQA LSQLYSKAKENDLQWNQAIKNAHHQLEPVVGKQISLWRKDLSTGR QGIAGLSLWSIEELEATKKLLTRWSKRSREPGVVKRIERFETFAKQIQ HHINQVKENRLKQLANLIVMTALGYKYDQEQKKWIEVYPACQVVL FENLRSYRFSYERSRRENKKLMEWSHRSIPKLVQMQGELFGLQVAD VYAAYSSRYHGRTGAPGIRCHALTEADLRNETNIIHELIEAGFIKEEH RPYLQQGDLVPWSGGELFATLQKPYDNPRILTLHADINAAQNIQKRF WHPSMWFRVNCESVMEGEIVTYVPKNKTVHKKQGKTFRFVKVEGS DVYEWAKWSKNRNKNTFSSITERKPPSSMILFRDPSGTFFKEQEWVE QKTFWGKVQSMIQAYMKKTIVQRMEE
메틸로박테리움 노둘란스 (장형) (SEQ ID NO:385)	MYEAIVLADDANAQLANAFLGPLTDPNSAGFLEAFNKVDRPAPSWL DQVPASDPIDPAVLAEANAWLDTDAGRAWLVDTGAPPRWRSLAAK QDPIWPREFARKLGELRKEAASGTSAIIKALKRDFGVLPLFQPSLAPRI LGSRSSLTPWDRLAFRLAVGHLLSWESWCTRARDEHTARVQRLEQF SSAHLKGDLATKVSTLREYERARKEQIAQLGLPMGERDFLITVRMTR

[0446]

	GWDDLREKWRRSGDKGQEALHAIIATEQTRKRGRFGDPDLFRWLA RPENHHVWADGHADAVGVLARVNAMERLVERSRDTALMTLPDPV AHPRSAQWEAEGGSNLRNYQLEAVGGELQITLPLLKAADDGRCIDT PLSFSLAPSDQLQGVVLTKQDKQQKITYCTNMNEVFEAKLGSADLL LNWDHLRGRIRDRVDAGDIGSAFLKLALDVAHVLPDGVDDQLARA AFHFQSAKGAKSKHADSVQAGLRVLSIDLGVRSFATCSVFELKDTAP TTGVAFPLAEFRLWAVHERSFTLELPGENVGAAGQQWRAQADAEL RQLRGGLNRHRQLLRAATVQKGERDAYLTDLREAWSAKELWPFEA SLLSELERCSTVADPLWQDTCKRAARLYRTEFGAVVSEWRSRTRSR EDRKYAGKSMWSVQHLTDVRRFLQSWSLAGRASGDIRRLDRERGG VFAKDLLDHIDALKDDRLKTGADLIVQAARGFQRNEFGYWVQKHA PCHVILFEDLSRYRMRTDRPRRENSQLMQWAHRGVPDMVGMQGEI YGIQDRRDPDSARKHARQPLAAFCLDTPAAFSSRYHASTMTPGIRCH PLRKREFEDQGFLELLKRENEGLDLNGYKPGDLVPLPGGEVFVCLN ANGLSRIHADINAAQNLQRRFWTQHGDAFRLPCGKSAVQGQIRWAP LSMGKRQAGALGGFGYLEPTGHDSGSCQWRKTTEAEWRRLSGAQK DRDEAAAAEDEELQGLEEELLERSGERVVFFRDPSGVVLPTDLWFPS AAFWSIVRAKTVGRLRSHLDAQAEASYAVAAGL
오피투타세아에 박테리움 (SEQ ID NO:386)	MSLNRIYQGRVAAVETGTALAKGNVEWMPAAGGDEVLWQHHELF QAAINYYLVALLALADKNNPVLGPLISQMDNPQSPYHVWGSFRRQG RQRTGLSQAVAPYITPGNNAPTLDEVFRSILAGNPTDRATLDAALMQ LLKACDGAGAIQQEGRSYWPKFCDPDSTANFAGDPAMLRREQHRLL LPQVLHDPAITHDSPALGSFDTYSIATPDTRTPQLTGPKARARLEQAI TLWRVRLPESAADFDRLASSLKKIPDDDSRLNLQGYVGSSAKGEVQ ARLFALLLFRHLERSSFTLGLLRSATPPPKNAETPPPAGVPLPAASAA DPVRIARGKRSFVFRAFTSLPCWHGGDNIHPTWKSFDIAAFKYALTV INQIEEKTKERQKECAELETDFDYMHGRLAKIPVKYTTGEAEPPPILA NDLRIPLLRELLQNIKVDTALTDGEAVSYGLQRRTIRGFRELRRIWRG HAPAGTVFSSELKEKLAGELRQFQTDNSTTIGSVQLFNELIQNPKYW PIWQAPDVETARQWADAGFADDPLAALVQEAELQEDIDALKAPVK LTPADPEYSRRQYDFNAVSKFGAGSRSANRHEPGQTERGHNTFTTEI AARNAADGNRWRATHVRIHYSAPRLLRDGLRRPDTDGNEALEAVP WLQPMMEALAPLPTLPQDLTGMPVFLMPDVTLSGERRILLNLPVTLE PAALVEQLGNAGRWQNQFFGSREDPFALRWPADGAVKTAKGKTHI PWHQDRDHFTVLGVDLGTRDAGALALLNVTAQKPAKPVHRIIGEAD GRTWYASLADARMIRLPGEDARLFVRGKLVQEPYGERGRNASLLE WEDARNIILRLGQNPDELLGADPRRHSYPEINDKLLVALRRAQARLA RLQNRSWRLRDLAESDKALDEIHAERAGEKPSPLPPLARDDAIKSTD EALLSQRDIIRRSFVQIANLILPLRGRRWEWRPHVEVPDCHILAQSDP GTDDTKRLVAGQRGISHERIEQIEELRRRCQSLNRALRHKPGERPVL GRPAKGEEIADPCPALLEKINRLRDQRVDQTAHAILAAALGVRLRAP SKDRAERRHRDIHGEYERFRAPADFVVIENLSRYLSSQDRARSENTR LMQWCHRQIVQKLRQLCETYGIPVLAVPAAYSSRFSSRDGSAGFRA VHLTPDHRHRMPWSRILARLKAHEEDGKRLEKTVLDEARAVRGLFD RLDRFNAGHVPGKPWRTLLAPLPGGPVFVPLGDATPMQADLNAAIN IALRGIAAPDRHDIHHRLRAENKKRILSLRLGTQREKARWPGGAPAV

[0447]

	TLSTPNNGASPEDSDALPERVSNLFVDIAGVANFERVTIEGVSQKFAT GRGLWASVKQRAWNRVARLNETVTDNNRNEEEDDIPM
바실러스 sp. NSP2.1 (SEQ ID NO:387)	MAIRSIKLKLKTHTGPEAQNLRKGIWRTHRLLNEGVAYYMKMLLLF RQESTGERPKEELQEELICHIREQQQRNQADKNTQALPLDKALEALR QLYELLVPSSVGQSGDAQIISRKFLSPLVDPNSEGGKGTSKAGAKPT WQKKKEANDPTWEQDYEKWKKRREEDPTASVITTLEEYGIRPIFPLY TNTVTDIAWLPLQSNQFVRTWDRDMLQQAIERLLSWESWNKRVQE EYAKLKEKMAQLNEQLEGGQEWISLLEQYEENRERELRENMTAAN DKYRITKRQMKGWNELYELWSTFPASASHEQYKEALKRVQQRLRG RFGDAHFFQYLMEEKNRLIWKGNPQRIHYFVARNELTKRLEEAKQS ATMTLPNARKHPLWVRFDARGGNLQDYYLTAEADKPRSRRFVTFS QLIWPSESGWMEKKDVEVELALSRQFYQQVKLLKNDKGKQKIEFK DKGSGSTFNGHLGGAKLQLERGDLEKEEKNFEDGEIGSVYLNVVIDF EPLQEVKNGRVQAPYGQVLQLIRRPNEFPKVTTYKSEQLVEWIKASP QHSAGVESLASGFRVMSIDLGLRAAAATSIFSVEESSDKNAADFSYW IEGTPLVAVHQRSYMLRLPGEQVEKQVMEKRDERFQLHQRVKFQIR VLAQIMRMANKQYGDRWDELDSLKQAVEQKKSPLDQTDRTFWEGI VCDLTKVLPRNEADWEQAVVQIHRKAEEYVGKAVQAWRKRFAAD ERKGIAGLSMWNIEELEGLRKLLISWSRRTRNPQEVNRFERGHTSHQ RLLTHIQNVKEDRLKQLSHAIVMTALGYVYDERKQEWCAEYPACQ VILFENLSQYRSNLDRSTKENSTLMKWAHRSIPKYVHMQAEPYGIQI GDVRAEYSSRFYAKTGTPGIRCKKVRGQDLQGRRFENLQKRLVNEQ FLTEEQVKQLRPGDIVPDDSGELFMTLTDGSGSKEVVFLQADINAAH NLQKRFWQRYNELFKVSCRVIVRDEEEYLVPKTKSVQAKLGKGLFV KKSDTAWKDVYVWDSQAKLKGKTTFTEESESPEQLEDFQEIIEEAEE AKGTYRTLFRDPSGVFFPESVWYPQKDFWGEVKRKLYGKLRERFLT KAR
메틸로박테리움 노둘란스 (단형) (SEQ ID NO:388)	MLTKQDKQQKITYCTNMNEVFEAKLGSADLLLNWDHLRGRIRDRV DAGDIGSAFLKLALDVAHVLPDGVDDQLARAAFHFQSAKGAKSKH ADSVQAGLRVLSIDLGVRSFATCSVFELKDTAPTTGVAFPLAEFRLW AVHERSFTLELPGENVGAAGQQWRAQADAELRQLRGGLNRHRQLL RAATVQKGERDAYLTDLREAWSAKELWPFEASLLSELERCSTVADP LWQDTCKRAARLYRTEFGAVVSEWRSRTRSREDRKYAGKSMWSV QHLTDVRRFLQSWSLAGRASGDIRRLDRERGGVFAKDLLDHIDALK DDRLKTGADLIVQAARGFQRNEFGYWVQKHAPCHVILFEDLSRYR MRTDRPRRENSQLMQWAHRGVPDMVGMQGEIYGIQDRRDPDSAR KHARQPLAAFCLDTPAAFSSRYHASTMTPGIRCHPLRKREFEDQGFL ELLKRENEGLDLNGYKPGDLVPLPGGEVFVCLNANGLSRIHADINAA QNLQRRFWTQHGDAFRLPCGKSAVQGQIRWAPLSMGKRQAGALGG FGYLEPTGHDSGSCQWRKTTEAEWRRLSGAQKDRDEAAAAEDEEL QGLEEELLERSGERVVFFRDPSGVVLPTDLWFPSAAFWSIVRAKTVG RLRSHLDAQAEASYAVAAGL

[0448]

[0449] 일정 구현예에서, C2c1 이펙터 단백질은 렌티스파에리아 또는 라세이엘라 속으로부터 선택되는 종으로부터 유래하거나 또는 기원한다.

[0450] 일정 구현예에서, C2c1 이펙터 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp._V3-13, 렌티스파에리아 박테리움, 또는 라세이엘라 세디미니스로부터 선택되는 종으로부터 유래하거나 또는 기원한다.

[0451] 일정 구현예에서, C2c1 이펙터 단백질은 야생형 서열이 WP_067936067의 서열에 상응하는 알리시클로바실러스 카케가웬시스, 야생형 서열이 Wp_101661451의 서열에 상응하는 바실러스 sp._V3-13, 야생형 서열이 DCFZ01000012의 서열에 상응하는 렌티스파에리아 박테리움, 또는 야생형 서열이 WP_106341859의 서열에 상응하는 라세이엘라 세디미니스로부터 선택되는 종으로부터 유래하거나 또는 기원한다.

[0452] 일정 구현예에서, C2c1 이펙터 단백질은 표 2로부터 선택되는 종으로부터 유래하거나 또는 기원하고 표 2에 표시된 바와 같은 야생형 서열을 갖는다. 본 명세서의 다른 곳에 기술된 바와 같은 돌연변이되거나 또는 절두된 Cas12b 단백질은 표시된 서열로부터 벗어날 수 있다는 것을 이해할 것이다.

[0453]

표 2 - Cas12b 오솔로그	
종	서열
알리시클로바실러스 카케가웬시스 (SEQ ID NO:389)	MAVKSIKVKLRLSECPDILAGMWQLHRATNAGVRYYTEWVSLMRQ EILYSRGPDGGQQCYMTAEDCQRELLRRLRNRQLHNGRQDQPGTD ADLLAISRRLYEILVLQSIGKRGDAQQIASSFLSPLVDPNSKGGRGEA KSGRKPAWQKMRDQGDPRWVAAREKYEQRKAVDPSKEILNSLDAL GLRPLFAVFTETYRSGVDWKPLGKSQGVRTWDRDMFQQALERLMS WESWNRRVGEEYARLFQQKMKFEQEHFAEQSHLVKLARALEADM RAASQGFEAKRGTAHQITRRALRGADRVFEIWKSIPEEALFSQYDEVI RQVQAEKRRDFGSHDLFAKLAEPKYQPLWRADETFLTRYALYNGV LRDLEKARQFATFTLPDACVNPIWTRFESSQGSNLHKYEFLFDHLGP GRHAVRFQRLLVVESEGAKERDSVVVPVAPSGQLDKLVLREEEKSS VALHLHDTARPDGFMAEWAGAKLQYERSTLARKARRDKQGMRSW RRQPSMLMSAAQMLEDAKQAGDVYLNISVRVKSPSEVRGQRRPPY AALFRIDDKQRRVTVNYNKLSAYLEEHPDKQIPGAPGLLSGLRVMS VDLGLRTSASISVFRVAKKEEVEALGDGRPPHYYPIHGTDDLVAVHE RSHLIQMPGETETKQLRKLREERQAVLRPLFAQLALLRLLVRCGAAD ERIRTRSWQRLTKQGREFTKRLTPSWREALELELTRLEAYCGRVPDD EWSRIVDRTVIALWRRMGKQVRDWRKQVKSGAKVKVKGYQLDVV GGNSLAQIDYLEQQYKFLRRWSFFARASGLVVRADRESHFAVALRQ HIENAKRDRLKKLADRILMEALGYVYEASGPREGQWTAQHPPCQLII LEELSAYRFSDDRPPSENSKLMAWGHRGILEELVNQAQVHDVLVGT VYAAFSSRFDARTGAPGVRCRRVPARFVGATVDDSLPLWLTEFLDK HRLDKNLLRPDDVIPTGEGEFLVSPCGEEAARVRQVHADINAAQNL QRRLWQNFDITELRLRCDVKMGGEGTVLVPRVNNARAKQLFGKKV LVSQDGVTFFERSQTGGKPHSEKQTDLTDKELELIAEADEARAKSVV LFRDPSGHIGKGHWIRQREFWSLVKQRIESHTAERIRVRGVGSSLD
바실러스 sp._V3-13 (SEQ ID NO:390)	MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTLYR QEAIGDKTKEAYQAELINIIRNQQRNNGSSEEHGSDQEILALLRQLYE LIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNAGRKPRWKRL KEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLPLFPLFTNIQ KDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWNRRVADEYKQ LKEKTESYYKEHLTGGEEWIEKIRKFEKERNMELEKNAFAPNDGYFI TSRQIRGWDRVYEKWSKLPESASPEELWKVVAEQQNKMSEGFGDP KVFSFLANRENRDIWRGHSERIYHIAAYNGLQKKLSRTKEQATFTLP DAIEHPLWIRYESPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWI EKENIEIPLAPSIQFNRQIKLKQHVKGKQEISFSDYSSRISLDGVLGGS RIQFNRKYIKNHKELLGEGDIGPVFFNLVVDVAPLQETRNGRLQSPIG KALKVISSDFSKVIDYKPKELMDWMNTGSASNSFGVASLLEGMRVM SIDMGQRTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKRSFLLN LPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRLETKKTPDER KKAIHKLMEIVQSYDSWTASQKEVWEKELNLLTNMAAFNDEIWKE SLVELHHRIEPYVGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRR LLISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLII MTALGFKYDKEEKDRYKRWKETYPACQIILFENLNRYLFNLDRSRR ENSRLMKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAP

	GIRCHALTEEDLKAGSNTLKRLIEDGFINESELAYLKKGDIIPSQGGEL FVTLSKRYKKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPC QLARMGEDKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKSEK MKIKTDTTFDLQDLDGFEDISKTIELAQEQQKKYLTMFRDPSGYFFN NETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL
렌티스파에리아 박테리움 (SEQ ID NO:391)	MAVELNRIYQGRVNHVYIFDENQNQVSVDNGDDLLFVHHELYQDAI NYYLVALAAMALDSKDSLFGKFKMQIRAVWNDFYRNGQLRPGLKH SLIRSLGHAAELNTSNGADIAMNLILEDGGIPSEILNAALEHLAEKCT GDVSQLGKTFFPRFCDTAYHGNWDVDAKSFSEKKGRQRLVDALYS LHPVQAVQELAPEIEIGWGGVKTQTGKFFTGDEAKASLKKAISYFLQ DTGKNSPELQEYFSVAGKQPLEQYLGKIDTFPEISFGRISSHQNINISN AMWILKFFPDQYSVDLIKNLIPNKKYEIGIAPQWGDDPVKLSRGKRG YTFRAFTDLAMWEKNWKVFDRAAFSDALKTINQFRNKTQERNDQL KRYCAALNWMDGESSDKKPPVEPADADAVDEAATSVLPILAGDKR WNALLQLQKELGICNDFTENELMDYGLSLRTIRGYQKLRSMMLEKE EKMRAKTADDEEISQALQEIIIKFQSSHRDTIGSVSLFLKLAEPKYFCV WHDADKNQNFASVDMVADAVRYYSYQEEKARLEEPIQITPADARY SRRVSDLYALVYKNAKECKTGYGLRPDGNFVFEIAQKNAKGYAPA KVVLAFSAPRLKRDGLIDKEFSAYYPPVLQAFLREEEAPKQSFKTTA VILMPDWDKNGKRRILLNFPIKLDVSAIHQKTDHRFENQFYFANNTN TCLLWPSYQYKKPVTWYQGKKPFDVVAVDLGQRSAGAVSRITVST EKREHSVAIGEAGGTQWYAYRKFSGLLRLPGEDATVIRDGQRTEELS GNAGRLSTEEETVQACVLCKMLIGDATLLGGSDEKTIRSFPKQNDKL LIAFRRATGRMKQLQRWLWMLNENGLCDKAKTEISNSDWLVNKNI DNVLKEEKQHREMLPAILLQIADRVLPLRGRKWDWVLNPQSNSFVL QQTAHGSGDPHKKICGQRGLSFARIEQLESLRMRCQALNRILMRKTG EKPATLAEMRNNPIPDCCPDILMRLDAMKEQRINQTANLILAQALGL RHCLHSESATKRKENGMHGEYEKIPGVEPAAFVVLEDLSRYRFSQD RSSYENSRLMKWSHRKILEKLALLCEVFNVPILQVGAAYSSKFSANA IPGFRAEECSIDQLSFYPWRELKDSREKALVEQIRKIGHRLLTFDAKA TIIMPRNGGPVFIPFVPSDSKDTLIQADINASFNIGLRGVADATNLLCN NRVSCDRKKDCWQVKRSSNFSKMVYPEKLSLSFDPIKKQEGAGGNF FVLGCSERILTGTSEKSPVFTSSEMAKKYPNLMFGSALWRNEILKLER CCKINQSRLDKFIAKKEVQNEL
라세이엘라 세디미니스 (SEQ ID NO:392)	MSIRSFKLKIKTKSGVNAEELRRGLWRTHQLINDGIAYYMNWLVLL RQEDLFIRNEETNEIEKRSKEEIQGELLERVHKQQQRNQWSGEVDDQ TLLQTLRHLYEEIVPSVIGKSGNASLKARFFLGPLVDPNNKTTKDVSK SGPTPKWKKMKDAGDPNWVQEYEKYMAERQTLVRLEEMGLIPLFP MYTDEVGDIHWLPQASGYTRTWDRDMFQQAIERLLSWESWNRRVR ERRAQFEKKTHDFASRFSESDVQWMNKLREYEAQQEKSLEENAFAP NEPYALTKKALRGWERVYHSWMRLDSAASEEAYWQEVATCQTAM RGEFGDPAIYQFLAQKENHDIWRGYPERVIDFAELNHLQRELRRAKE DATFTLPDSVDHPLWVRYEAPGGTNIHGYDLVQDTKRNLTLILDKFI LPDENGSWHEVKKVPFSLAKSKQFHRQVWLQEEQKQKKREVVFYD YSTNLPHLGTLAGAKLQWDRNFLNKRTQQQIEETGEIGKVFFNISVD VRPAVEVKNGRLQNGLGKALTVLTHPDGTKIVTGWKAEQLEKWVG

[0454]

	ESGRVSSLGLDSLSEGLRVMSIDLGQRTSATVSVFEITKEAPDNPYKF FYQLEGTELFAVHQRSFLLALPGENPPQKIKQMREIRWKERNRIKQQ VDQLSAILRLHKKVNEDERIQAIDKLLQKVASWQLNEEIATAWNQA LSQLYSKAKENDLQWNQAIKNAHHQLEPVVGKQISLWRKDLSTGR QGIAGLSLWSIEELEATKKLLTRWSKRSREPGVVKRIERFETFAKQIQ HHINQVKENRLKQLANLIVMTALGYKYDQEQKKWIEVYPACQVVL FENLRSYRFSYERSRRENKKLMEWSHRSIPKLVQMQGELFGLQVAD VYAAYSSRYHGRTGAPGIRCHALTEADLRNETNIIHELIEAGFIKEEH RPYLQQGDLVPWSGGELFATLQKPYDNPRILTLHADINAAQNIQKRF WHPSMWFRVNCESVMEGEIVTYVPKNKTVHKKQGKTFRFVKVEGS DVYEWAKWSKNRNKNTFSSITERKPPSSMILFRDPSGTFFKEQEWVE QKTFWGKVQSMIQAYMKKTIVQRMEE

[0455]

[0456] 이펙터 단백질은 제1 이펙터 단백질 (예를 들어, C2c1) 오솔로그로부터의 제1 단편 및 제2 이펙터 (예를 들어, C2c1) 단백질 오솔로그로부터의 제2 단편을 포함하는 키메라 이펙터 단백질을 포함할 수 있고, 여기서 제1 이펙터 단백질 오솔로그와 제2 이펙터 단백질 오솔로그는 상이하다. 제1 및 제2 이펙터 단백질 (예를 들어, C2c1) 오솔로그 중 적어도 하나는 알리시클로바실러스 (Alicyclobacillus), 데술포비브리오 (Desulfovibrio), 데술포나트로늄 (Desulfonatronum), 오피투타세아에 (Opitutaceae), 투베리바실러스 (Tuberibacillus), 바실러스 (Bacillus), 브레비바실러스 (Brevibacillus), 칸디다투스 (Candidatus), 데술파티랍디움 (Desulfatirhabdium), 엘루시미크로비아 (Elusimicrobia), 시트로박터 (Citrobacter), 메틸로박테리움 (Methylobacterium), 옴니트로피카이 (Omnitrophicai), 피시스파에라에 (Phycisphaerae), 플란크토마이세테스

(Planctomycetes), 스피로카에테스 (Spirochaetes), 베루코미크로비아세아에 (Verrucomicrobiaceae), 렌티스파에리아 (Lentisphaeria) 또는 라세이엘라 (Laceyella)를 포함하는 유기체로부터 유래하거나 또는 기원하는 이펙터 단백질 (예를 들어, C2c1)을 포함할 수 있고; 예를 들어, 키메라 이펙터 단백질은 제1 단편 및 제2 단편을 포함하며 여기서 각각의 제1 및 제2 단편은 알리시클로바실러스 (Alicyclobacillus), 데술포비브리오 (Desulfovibrio), 데술포나트로늄 (Desulfonatronum), 오피투타세아에 (Opitutaceae), 투베리바실러스 (Tuberibacillus), 바실러스 (Bacillus), 브레비바실러스 (Brevibacillus), 칸디다투스 (Candidatus), 데술파티랍디움 (Desulfatirhabdium), 엘루시미크로비아 (Elusimicrobia), 시트로박터 (Citrobacter), 메틸로박테리움 (Methylobacterium), 옴니트로피카이 (Omnitrophicai), 피시스파에라에 (Phycisphaerae), 플란크토마이세테스 (Planctomycetes), 스피로카에테스 (Spirochaetes), 및 베루코미크로비아세아에 (Verrucomicrobiaceae), 렌티스파에리아 (Lentisphaeria) 또는 라세이엘라 (Laceyella)를 포함하는 유기체의 C2c1로부터 선택되고, 제1 단편 및 제2 단편은 동일한 박테리아로부터의 것이 아니고; 예를 들어 키메라 이펙터 단백질은 제1 단편 및 제2 단편을 포함하며, 여기서 각각의 제1 및 제2 단편은 알리시클로바실러스 악시도테레스트리스 (Alicyclobacillus acidoterrestris) (예를 들어, ATCC 49025), 알리시클로바실러스 콘타미난스 (Alicyclobacillus contaminans) (예를 들어, DSM 17975), 알리시클로바실러스 마크로스포란지두스 (Alicyclobacillus macrosporangiidus) (예를 들어, DSM 17980), 바실러스 히사시이 (Bacillus hisashii) 균주 C4, 칸디다투스 린도우박테리아 박테리움 (Candidatus Lindowbacteria bacterium) RIFCSPLOWO2, 데술포비브리오 이노피나투스 (Desulfovibrio inopinatus) (예를 들어, DSM 10711), 데술포나트로늄 티오디스무탄스 (Desulfonatronum thiodismutans) (예를 들어, 균주 MLF-1 또는 genbank 수탁 번호 WP_031386437), 엘루시미크로비아 박테리움 (Elusimicrobia bacterium) RIFOXYA12, 옴니트로피카 WOR_2 박테리움 (Omnitrophica WOR_2 bacterium) RIFCSPHIGHO2, 오피투타세아에 박테리움 (Opitutaceae bacterium) TAV5 또는 genbank 수탁 번호 WP_009513281, 피시스파에라에 박테리움 (Phycisphaerae bacterium) ST-NAGAB-D1, 플란크토마이세테스 박테리움 (Planctomycetes bacterium) RBG_13_46_10, 스피로카에테스 박테리움 (Spirochaetes bacterium) GWB1_27_13, 베루코미크로비아세아에 박테리움 (Verrucomicrobiaceae bacterium) UBA2429, 투베리바실러스 칼리두스 (Tuberibacillus calidus) (예를 들어, DSM 17572), 바실러스 써모아밀로보란스 (Bacillus thermoamylovorans) (예를 들어, 균주 B4166), 브레비바실러스 sp.(Brevibacillus sp.)CF112, 바실러스 sp.(Bacillus sp.)NSP2.1, 데술파티랍디움 부티라티보란스 (Desulfatirhabdium butyrativorans) (예를 들어, DSM 18734 또는 genbank 수탁 번호 WP_028326052), 알리시클로바실러스 허바리우스 (Alicyclobacillus herbarius) (예를 들어, DSM 13609), 시트로박터 프레운디 (Citrobacter freundii) (예를 들어, ATCC 8090), 브레비바실러스 아그리 (Brevibacillus agri) (예를 들어, BAB-2500), 메틸로박테리움 노둘란스 (Methylobacterium nodulans) (예를 들어, ORS 2060 또는 genbank 수탁 번호 WP_043747912), 알리시클로바실러스 카케가웬시스 (Alicyclobacillus kakegawensis) (예를 들어, genbank 수탁 번호 WP_067936067), 바실러스 (Bacillus) sp. V3-13 (예를 들어, genbank 수탁 뻔호 WP_101661451), 렌티스파에리아 박테리움 (Lentisphaeria bacterium) (예를 들어, DCFZ01000012 유래), 라세이엘라_세디미니스 (Laceyella_sediminis) (예를 들어, genbank 수탁 번호 WP_106341859)의 C2c1로부터 선택되고, 여기서 제1 및 제2 단편은 동일한 박테리아로부터의 것이 아니다. 본 명세서에서 사용시, Cas12 단백질 (예를 들어, Cas12b)가 종으로부터 기원할 때, 종의 야생형 Cas12 단백질, 또는 종의 야생형 Cas12 단백질의 상동체일 수 있다. 종의 야생형 Cas12 단백질의 상동체인 Cas12 단백질은 야생형 Cas12 단백질의 하나 이상의 변이 (예를 들어, 돌연변이, 절두 등)을 포함할 수 있다.

[0457] 보다 바람직한 구현예에서, C2c1b 이펙터 단백질은 알리시클로바실러스 악시도테레스트리스 (Alicyclobacillus acidoterrestris) (예를 들어, ATCC 49025), 알리시클로바실러스 콘타미난스 (Alicyclobacillus contaminans) (예를 들어, DSM 17975), 알리시클로바실러스 마크로스포란지두스 (Alicyclobacillus macrosporangiidus) (예를 들어, DSM 17980), 바실러스 히사시이 (Bacillus hisashii) 균주 C4, 칸디다투스 린도우박테리아 박테리움 (Candidatus Lindowbacteria bacterium) RIFCSPLOWO2, 데술포비브리오 이노피나투스 (Desulfovibrio inopinatus) (예를 들어, DSM 10711), 데술포나트로눔 티오디스무탄스 (Desulfonatronum thiodismutans) (예를 들어, 균주 MLF-1 또는 genbank 수탁 번호WP_031386437), 엘루시미크로비아 박테리움 (Elusimicrobia bacterium) RIFOXYA12, 옴니트로피카 WOR_2 박테리움 (Omnitrophica WOR_2 bacterium) RIFCSPHIGHO2, 오피투타세아에 박테리움 (Opitutaceae bacterium) TAV5 또는 genbank 수탁 번호 WP_009513281, 피시스파에라에 박테리움 (Phycisphaerae bacterium) ST-NAGAB-D1, 플란크토마이세테스 박테리움 (Planctomycetes bacterium) RBG_13_46_10, 스피로카에테스 박테리움 (Spirochaetes bacterium) GWB1_27_13, 베루코미크로비아세아에 박테리움 (Verrucomicrobiaceae bacterium) UBA2429, 투베리바실러스 칼리두스 (Tuberibacillus calidus) (예를 들어, DSM 17572), 바실러스 써모아밀로보란스 (Bacillus thermoamylovorans) (예를 들어, 균주 B4166), 브레비

바실러스 sp. (Brevibacillus sp.) CF112, 바실러스 (Bacillus) sp. NSP2.1, 데술파티랍디움 부티라티보란스 (Desulfatirhabdium butyrativorans) (예를 들어, DSM 18734 또는 genbanck 수탁 번호 WP_028326052), 알리시클로바실러스 허바리우스 (Alicyclobacillus herbarius) (예를 들어, DSM 13609), 시트로박터 프레운디 (Citrobacter freundii) (예를 들어, ATCC 8090), 브레비바실러스 아그리 (Brevibacillus agri) (예를 들어, BAB-2500), 메틸로박테리움 노둘란스 (Methylobacterium nodulans) (예를 들어, ORS 2060 또는 genbank 수탁 번호 WP_043747912), 알리시클로바실러스 카케가웬시스 (Alicyclobacillus kakegawensis) (예를 들어, genbank 수탁 번호 WP_067936067), 바실러스 (Bacillus) sp. V3-13 (예를 들어, genbank 수탁 번호 WP_101661451), 렌티스파에리아 박테리움 (Lentisphaeria bacterium) (예를 들어, DCFZ01000012 유래), 라세이엘라_ 세디미니스 (Laceyella_sediminis) (예를 들어, genbank 수탁 번호 WP_106341859)에서 선택되는 박테리아 종으로부터 유래하거나 또는 기원한다. 특정 구현예에서, C2c1p 는 알리시클로바실러스 악시도테레스트리스 (Alicyclobacillus acidoterrestris) (예를 들어, ATCC 49025), 알리시클로바실러스 콘타미난스 (Alicyclobacillus contaminans) (예를 들어, DSM 17975)에서 선택되는 박테리아 종에서 유래한다.

[0458] 특정 구현예에서, 본원에서 나타내는 바와 같은 C2c1 의 상동체 또는 오솔로그는 C2c1 과 적어도 80%, 보다 바람직하게는 적어도 85%, 보다 더 바람직하게는 적어도 90%, 예를 들어 적어도 95% 의 서열 상동성 또는 동일성을 갖는다. 추가 구현예에서, 본원에서 나타내는 바와 같은 C2c1 의 상동체 또는 오솔로그는 야생형 C2c1 과 적어도 80%, 보다 바람직하게는 적어도 85%, 보다 더 바람직하게는 적어도 90%, 예를 들어 적어도 95% 의 서열 동일성을 갖는다. C2c1 이 하나 이상의 돌연변이를 갖는 경우 (돌연변이됨), 본원에서 나타내는 바와 같은 상기 C2c1 의 상동체 또는 오솔로그는 돌연변이된 C2c1 과 적어도 80%, 보다 바람직하게는 적어도 85%, 보다 더 바람직하게는 적어도 90%, 예를 들어, 적어도 95% 의 서열 동일성을 갖는다.

[0459] 일 구현예에서, C2c1 단백질은 제한없이 알리시클로바실러스, 데술포비브리오, 데술포나트로늄, 오피투타세아에, 투베리바실러스, 바실러스, 브레비바실러스, 칸디다투스, 데술파티랍디움, 엘루시미크로비아, 시트로박터, 메틸로박테리움, 옴니트로피카이, 피시스파에라에, 플란크토마이세테스, 스피로카에테스, 베루코미크로비아세아에, 렌티스파에리아 또는 라세이엘라를 포함하는 속의 유기체의 오솔로그일 수 있고; 특정 구현예에서, V형 Cas 단백질은 제한없이, 알리시클로바실러스 아시도테레스트리스 (예를 들어, ATCC 49025), 알리시클로바실러스 콘타미난스 (예를 들어, DSM 17975), 알리시클로바실러스 마크로스포란지이더스 (예를 들어, DSM 17980), 바실러스 히사시이 균주 C4, 칸디다투스 린도우박테리아 박테리움 RIFCSPLOWO2, 데술포비브리오 이노피나투스 (예를 들어, DSM 10711), 데술포나트로늄 티오디스뮤탄스 (예를 들어, 균주 MLF-1 또는 genbank 수탁 번호 WP_031386437), 엘루시미크로비아 박테리움 RIFOXYA12, 옴니트로피카 WOR_2 박테리움 RIFCSPHIGHO2, 오피투타세아에 박테리움 TAV5 또는 genbank 수탁 번호 WP_009513281, 피시스파에라에 박테리움 ST-NAGAB-D1, 플란크토마이세테스 박테리움 RBG_13_46_10, 스피로카에테스 박테리움 GWB1_27_13, 베루코미크로비아세아에 박테리움 UBA2429, 투베리바실러스 칼리두스 (예를 들어, DSM 17572), 바실러스 써모아밀로보란스 (예를 들어, 균주 B4166), 브레비바실러스 sp. CF112, 바실러스 sp. NSP2.1, 데술파티랍디움 부티라티보란스 (예를 들어, DSM 18734 또는 genbank 수탁 번호 WP_028326052), 알리시클로바실러스 허바리우스 (예를 들어, DSM 13609), 시트로박터 프루운디이 (예를 들어, ATCC 8090), 브레비바실러스 아그리 (예를 들어, BAB-2500), 메틸로박테리움 노둘란스 (예를 들어, ORS 2060 또는 genbank 수탁 번호 WP_043747912), 알리시클로바실러스 카케가웬시스 (예를 들어, genbank 수탁 번호 WP_067936067), 바실러스 sp. V3-13 (예를 들어, genbank 수탁 번호 WP_101661451), 렌티스파에리아 박테리움 (예를 들어, DCFZ01000012 유래), 라세이엘라_세디미니스 (예를 들어, Genbank 수탁 번호 WP_106341859), 바실러스 sp. V3-13 (예를 들어, GenBank 수탁 번호 Wp_101661451)를 포함한 종의 유기체의 오솔로그일 수 있다. 특정 구현예에서, 본 명세서에서 언급되는 C2c1의 상동체 또는 오솔로그는 본 명세서에 개시된 C2c1 서열 중 하나 이상과 적어도 80%, 더 바람직하게는 적어도 85%, 훨씬 더 바람직하게는 적어도 90%, 예를 들어, 적어도 95%의 서열 상동성 또는 동일성을 갖는다. 추가 구현예에서, 본원에서 나타내는 바와 같은 C2c1 의 상동체 또는 오솔로그는 야생형 AacC2c1 또는 BthC2c1 과 적어도 80%, 보다 바람직하게는 적어도 85%, 보다 더 바람직하게는 적어도 90%, 예를 들어 적어도 95% 의 서열 동일성을 갖는다.

[0460] 특정 구현예에서, 본 발명의 C2c1 단백질은 AacC2c1 또는 BthC2c1 과 적어도 60%, 보다 특히 적어도 70, 예컨대 적어도 80%, 보다 바람직하게는 적어도 85%, 보다 더 바람직하게는 적어도 90%, 예를 들어 적어도 95% 의 서열 상동성 또는 동일성을 갖는다. 추가 구현예에서, 본원에서 나타내는 바와 같은 C2c1 단백질은 야생형 AacC2c1 과 적어도 60%, 예컨대 적어도 70%, 보다 특히 적어도 80%, 보다 바람직하게는 적어도 85%, 보다 더 바람직하게는 적어도 90%, 예를 들어 적어도 95% 의 서열 동일성을 갖는다. 특정 구현예에서, 본 발명의 C2c1 단백질은 AacC2c1 과 60% 미만의 서열 동일성을 갖는다. 당업자는 이것이 C2c1 단백질의 절두 형태를 포함

하고, 이에 의해 절두 형태의 길이에 대해 서열 동일성이 결정된다는 것을 이해할 것이다.

[0461] 일정 예의 구현예에서, Cas12b 오솔로그는 예를 들어 약 25℃, 약 26℃, 약 27℃, 약 28℃, 약 29℃, 약 30℃, 약 31℃, 약 32℃, 약 33℃, 약 34℃, 약 35℃, 약 36℃, 약 37℃, 약 38℃, 약 39℃, 약 40℃, 약 41℃, 약 42℃, 약 43℃, 약 44℃, 약 45℃, 약 46℃, 약 47℃, 약 48℃, 약 49℃, 또는 약 50° C의 온도에서 활성 (예를 들어, 핵산 (예컨대 RNA 또는 DNA) 절단 활성)을 가질 수 있다. 소정 Cas12b 오솔로그는 예를 들어 30℃ 내지 50℃, 30℃ 내지 48℃, 37℃ 내지 42℃, 또는 37℃ 내지 48℃의 온도 범위에서 이의 최적 활성을 가질 수 있다 일부 예에서, BvCas12b는 약 37℃에서 활성을 가질 수 있다. 일부 예에서, BhCas12b (예를 들어, 본 명세서에 개시된 변이체 4)는 약 37℃에서 활성을 가질 수 있다. 일부 예에서, AkCas12b는 약 48℃에서 활성을 가질 수 있다. 활성은 진핵생물 세포에서 Cas12b 오솔로그의 활성일 수 있다. 대안적으로 또는 추가적으로, 활성은 원핵생물 세포에서 오솔로그의 활성일 수 있다. 일부 경우에서, 이러한 활성은 최적 활성일 수 있다.

[0462] **변형된 C2c1 효소**

[0463] 특정 구현예에서, 본 명세서에 정의한 바와 같은 조작된 C2c1 단백질, 예컨대, C2c1을 사용하게 하는 데 관심이 있고, 여기서 단백질은 RNA를 포함하는 핵산 분자와 복합체를 형성하여 CRISPR 복합체를 형성하고, CRISPR 복합체에서, 핵산 분자가 하나 이상의 표적 폴리뉴클레오티드 유전자좌를 표적화할 때, 단백질은 비변형 C2c1 단백질에 비해 적어도 하나의 변형을 포함하고, 변형된 단백질을 포함하는 CRISPR 복합체는 비변형 C2c1 단백질을 포함하는 복합체에 비해 변경된 활성을 갖는다. 본 명세서에서 CRISPR "단백질"을 말할 경우, C2c1 단백질은 바람직하게, 예컨대 C2c1(이로 제한되지 않음) 변형된 CRISPR 효소(예를 들어, 증가되거나 감소된 효소 활성을 가짐(또는 효소 활성을 가지지 않음))이다. 용어 "CRISPR 단백질"은, 야생형 CRISPR 단백질에 비하여, CRISPR 단백질이 변경된 효소 활성, 예컨대 증가되거나 감소된 효소 활성(또는 효소 활성을 가지지 않음)을 가지는지 여부와 무관하게, "CRISPR 효소" 와 상호교환 가능하게 사용될 수 있다.

[0464] 상기 기술된 돌연변이이외에도, CRISPR-Cas 단백질은 추가적으로 변형될 수 있다. 본 명세서에서 사용되는, CRISPR-Cas 단백질에 대한 용어 "변형된" 은 일반적으로 그것이 유래된 야생형 Cas 단백질과 비교하여 하나 이상의 변형 또는 돌연변이 (점 돌연변이, 절단, 삽입, 결실, 키메라, 융합 단백질 등 포함)를 갖는 CRISPR-Cas 단백질을 의미한다. 유래된 이란 유래된 효소가 대체로 높은 서열 상동성 정도를 갖는다는 의미에서, 야생형 효소를 기반으로 하지만, 당분야에 공지되거나 또는 본 명세서에 기술된 바와 같이 일부 방식으로 돌연변이(변형)되었다는 것을 의미한다.

[0465] CRISPR-Cas 단백질의 추가적인 변형은 변경된 작용성을 야기할 수도 있고 야기하지 않을 수도 있다. 예로서 그리고 특히 CRISPR-Cas 단백질에 관해, 변경된 작용성을 야기하지 않는 변형은, 예를 들어, 특정 숙주 내로의 발현을 위한 코돈 최적화, 또는 (예를 들어, 시각화를 위해) 특정 마커를 갖는 뉴클레아제를 제공하는 것을 포함한다. 변경된 작용기를 초래할 수 있는 변형은 또한 점 돌연변이, 삽입, 결실, 절단(분할 뉴클레아제를 포함) 등을 포함하는 돌연변이를 포함할 수 있다. 융합 단백질은, 예를 들어, 이종성 도메인 또는 기능성 도메인(예를 들어, 국소화 신호, 촉매적 도메인 등)과의 융합을 포함할 수 있지만, 이들로 제한되지 않는다.
일정 구현예에서, 다양한 상이한 변형은 조합될 수 있다(예를 들어, 촉매적으로 비활성이고, 추가로 기능성 도메인에 융합되고, 예컨대, 예를 들어, DNA 메틸화 또는, 예컨대, 파손(예를 들어, 상이한 뉴클레아제(도메인)에 의함), 돌연변이, 결실, 삽입, 대체, 결찰, 분해, 파손 또는 재조합을 포함하지만, 이들로 제한되지 않는 다른 핵산 변형을 유도하는, 돌연변이된 뉴클레아제). 본 명세서에서 사용되는 "변경된 기능성" 은 제한없이 변경된 특이성 (예를 들어, 변경된 표적 인식, 증가 (예를 들어, "증강된" Cas 단백질) 또는 감소된 특이성, 또는 변경된 PAM 인식), 변경된 활성 (예를 들어, 촉매적으로 불활성인 뉴클레아제 또는 닉카제를 포함하는 증가 또는 감소된 촉매 활성), 및/또는 변경된 안정성 (예를 들어, 탈안정화 도메인과 융합)을 포함한다. 적합한 이종성 도메인은 제한없이 뉴클레아제, 리가제, 복구 단백질, 메틸트랜스퍼라제, (바이러스) 인테그라제, 리콤비나제, 트랜스포사제, 아르고노트, 시티딘 디아미나제, 레트론, 그룹 II 인트론, 포스파타제, 포스포릴라제, 술프퓨릴라제, 키나제, 폴리머라제, 엑소뉴클레아제 등을 포함한다. 모든 이들 변형의 예는 당업계에 공지되어 있다. 본 명세서에서 언급되는 "변형된" 뉴클레아제, 특히 "변형된" Cas 또는 "변형된" CRISPR-Cas 시스템 또는 복합체는 바람직하게는 (예를 들어, 가이드 분자와 복합체인) 폴리핵산과 상호작용하거나 또는 그와 결합하는 능력을 여전히 갖는다는 것을 이해하게 될 것이다. 이러한 변형된 Cas 단백질은 본 명세서에 기술된 바와 같은 디아미나제 단백질 또는 이의 활성 도메인과 조합될 수 있다.

[0466] 일정 실시형태에서, CRISPR-Cas 단백질은 예컨대 표적화 또는 비표적화 가닥을 안정화시키는 돌연변이된 잔기를

포함하여, 증강된 활성 및/또는 특이성을 야기시키는 하나 이상의 변형을 포함할 수 있다 (예를 들어, 그 전문이 참조로 본 명세서에 편입되는, [eCas9; "Rationally engineered Cas9 nucleases with improved specificity" , Slaymaker et al. (2016), Science, 351(6268):84-88] 참조). 일정 구현예에서, 조작된 CRISPR 단백질의 변경 또는 변형 활성은 증가된 표적화 효율 또는 감소된 오프-표적 결합을 포함한다. 일정 구현예에서, 조작된 CRISPR 단백질의 변경된 활성은 변형된 절단 활성을 포함한다. 일정 구현예에서, 변경된 활성은 표적 폴리뉴클레오티드 유전자좌에 대해 증가된 변형된 절단 활성을 포함한다. 일정 구현예에서, 변경된 활성은 표적 폴리뉴클레오티드 유전자좌에 대해 감소된 변형된 절단 활성을 포함한다. 일정 구현예에서, 변경된 활성은 오프-표적 폴리뉴클레오티드 유전자좌에 대해 감소된 변형된 절단 활성을 포함한다. 일정 구현예에서, 변형된 뉴클레아제의 변경 또는 변형된 활성은 변경된 헬리카제 동역학을 포함한다. 일정 구현예에서, 변형된 뉴클레아제는 RNA (Cas 단백질의 경우)를 포함하는 핵산 분자, 또는 표적 폴리뉴클레오티드 유전자좌의 가닥, 또는 오프-표적 폴리뉴클레오티드 유전자좌의 가닥과 단백질의 회합을 변경시키는 변형을 포함한다. 본 발명의 일 양상에서, 조작된 CRISPR 단백질은 CRISPR 복합체의 형성을 변경시키는 변형을 포함한다. 일정 구현예에서, 변경된 활성은 오프-표적 폴리뉴클레오티드 유전자좌에 대해 증가된 변형된 절단 활성을 포함한다. 따라서, 일정 구현예에서, 오프-표적 폴리뉴클레오티드 유전자좌와 비교하여 표적 폴리뉴클레오티드 유전자좌에 대한 증가된 특이성이 존재한다. 다른 구현예에서, 오프-표적 폴리뉴클레오티드 유전자좌와 비교하여 표적 폴리뉴클레오티드 유전자좌에 대해 감소된 특이성이 존재한다. 일정 구현예에서, 돌연변이는 예컨대 Cas 단백질이 예를 들어 표적과 가이드 RNA 간 미스매치에 대해 낮아진 내성이 야기된 경우에, 감소된 오프-표적 효과 (예를 들어, 절단 또는 결합 속성, 활성, 또는 동역학)를 야기한다. 다른 돌연변이는 증가된 오프-표적 효과 (예를 들어, 절단 또는 결합 성질, 활성 또는 동역학)를 야기시킬 수 있다. 다른 돌연변이는 증가되거나 또는 감소된 온-표적 효과 (예를 들어, 절단 또는 결합 성질, 활성 또는 동역학)를 야기시킬 수 있다. 일정 구현예에서, 돌연변이는 기능성 뉴클레아제 복합체 (예를 들어, CRISPR-Cas 복합체)의 변경 (예를 들어, 증가 또는 감소)된 헬리카제 활성, 회합 또는 형성을 야기시킨다. 일정 구현예에서, 상기 기술된 바와 같이, 돌연변이는 변경된 PAM 인식을 야기시키며, 즉 상이한 PAM이, 비변형된 Cas 단백질과 비교하여, (추가로 또는 대안적으로) 인식될 수 있다. 특히 바람직한 돌연변이는 특이성을 증강시키기 위해서 양으로 하전된 잔기 및/또는 (진화적으로) 보존된 잔기, 예컨대 보존된 양으로 하전된 잔기를 포함한다. 일정 구현예에서, 이러한 잔기는 비하전된 잔기, 예컨대 알라닌으로 돌연변이될 수 있다.

[0467] C2c1의 결정 구조는 다른 V형 Cas 단백질, Cpf1 (Cas12a라고도 알려짐)과 유사성을 밝혀준다. C2c1 및 Cpf1 둘 모두는 α-나선형 인식 로브 (REC) 및 뉴클레아제 로브 (NUC)로 이루어진다. NUC 로브는 온전한 3D C2c1 구조를 형성하기 위한 구조적 셔플링 및 폴딩으로, 올리고뉴클레오티드-결합 (WED/OBD) 도메인, RuvC 도메인, Nuc 도메인, 및 브릿지 헬릭스 (Bh)를 더 함유한다 (Liu et al. Mol. Cell 65, 310-322). Nuc 도메인 내 일정 돌연변이 (예를 들어, AsCpf1의 R1226A, BvCas12b의 R894A)는 Cpf1을 비표적 가닥 절단을 위한 닉카제가 되게 한다. RuvC 도메인의 촉매 잔기의 돌연변이 (예를 들어, AsCpf1의 D908, E933, D1263에서의 돌연변이)는 뉴클레아제로서 Cpf1의 촉매 활성을 없앤다. 또한, Cpf1의 PAM 상호작용 (PI) 도메인의 돌연변이 (예를 들어, AsCpf1의 S542, K548, N522, 및 K607의 돌연변이)는 Cpf1 특이성을 변경시켜서, 잠재적으로 오프-표적 절단을 증가 또는 감소시키는 것으로 확인되었다 (참조: Gao et al. Cell Research (2016) 26, 901-913 (2016); Gao et al. Nature Biotechnology 35, 789-792 (2017)). 또한 C2c1의 결정 구조는 C2c1이 식별가능한 PI 도메인이 결여된 것을 밝혀주었는데, 정확히 말해서, C2c1이 PAM 인식 및 R-루프 형성을 위해 PAM 근위 이중 가닥 DNDA의 결합을 수용하도록 입체형태 조정을 겪는다는 것을 시사하고; C2c1은 아마도 큰홈 및 작은홈 측면 둘 모두로부터 PAM 듀플렉스를 인식하기 위해서 알파 헬릭스 도메인 및 WED/OBD를 관여시키는 듯 하다 (Yang et al, Cell 167, 1814-1828 (2016)).

[0468] 본 발명에 따라서, 효소의 불활성화를 야기하거나 또는 닉카제 활성으로 이중 가닥 뉴클레아제를 변형시키거나, 또는 C2c1의 PAM 인식 특이성을 변경시키는 돌연변이체가 생성될 수 있다. 일정 구현예에서, 이러한 정보는 오프-표적 효과가 감소된 효소를 개발하는데 사용된다..

[0469] 일정 예의 구현예에서, 편집 선호도는 표적 영역 내 특이적 삽입 또는 결실을 위한 것이다. 일정 예의 구현예에서, 적어도 하나의 변형은 하나 이상의 특이적 indel의 형성을 증가시킨다. 일정 예의 구현예에서, 적어도 하나의 변형은 C-말단 RuvC 유사 도메인, NUC 도메인, N-말단 알파-헬릭스 영역, 혼합 알파 및 베타 영역, 또는 이의 조합에 존재한다. 일정 예의 구현예에서, 변경된 편집 선호도는 indel 형성이다. 일정 예의 구현예에서, 적어도 하나의 변형은 하나 이상의 특이적 삽입의 형성을 증가시킨다.

[0470] 일정 예의 구현예에서, 적어도 하나의 변형은 하나 이상의 특이적 삽입의 형성을 증가시킨다. 일정 예의 구

현예에서, 적어도 하나의 변형은 표적 영역 내 A, T, G, 또는 C에 인접한 A의 삽입을 야기시킨다. 일정 예의 구현예에서, 적어도 하나의 변형은 표적 영역 내 A, T, G, 또는 C에 인접한 T의 삽입을 야기시킨다. 일정 예의 구현예에서, 적어도 하나의 변형은 표적 영역 내 A, T, G, 또는 C에 인접한 G의 삽입을 야기시킨다.

일정 예의 구현예에서, 적어도 하나의 변형은 표적 영역 내 A, T, G, 또는 C에 인접한 C의 삽입을 야기시킨다. 삽입은 인접한 뉴클레오티드에 대해 5' 또는 3' 일 수 있다. 일례의 구현예에서, 하나 이상의 변형은 존재하는 T에 인접한 T의 삽입을 유도한다. 일정 예의 구현예에서, 존재하는 T는 가이드 서열의 결합 영역 내 4번째 위치에 상응한다. 일정 예의 구현예에서, 하나 이상의 변형은 상기 기술된 것과 같은, 보다 정확한 1-염기 삽입 또는 결실을 보장하는 효소를 야기시킨다. 보다 특히, 하나 이상의 변형은 효소에 의한 다른 유형의 indel의 형성을 감소시킬 수 있다. 1-염기 삽입 또는 결실을 생성시키는 능력은 다수의 적용분야, 예컨대 작은 결실에 의해 초래되는 질환에서 유전자 돌연변이체의 교정, 보다 특히 HDR이 불가능한 경우에 흥미로울 수 있다. 예를 들어, 낭성 섬유증의 가장 일반적인 유전자형인, 3개 T의 삽입을 유도하는 3개 sRNA의 전달을 통한 CFTR의 F508del 돌연변이의 교정, 또는 뇌의 CDKL5에서 Alia Jafar의 단일 뉴클레오티드 결실의 교정. 편집 방법이 오직 NHEJ만을 요구하므로, 편집은 뇌같은 유사분열 후 세포에서 가능하다.

1 염기쌍 삽입/결실을 생성시키는 능력은 또한 게놈-와이드 CRISPR-Cas 음성 선별 스크린에서 유용할 수 있다. 일정예의 구현예에서, 적어도 하나의 변형은 돌연변이이다. 일정한 다른 예의 구현예에서, 하나 이상의 변형은 결합 특이성을 증가시키고/시키거나 오프-표적 효과를 감소시키기 위해서 변형을 포함하는 하기 기술된 돌연변이 또는 하나 이상의 추가 변형과 조합될 수 있다.

[0471] 일정 예의 구현에예서, 야생형과 비교하여 편집 선호도를 변경시키는 적어도 하나의 변형을 포함하는 조작된 CRISPR-cas 이펙터는 RNA 또는 표적 폴리펩티드 유전자좌를 포함하는 핵산 분자에 대한 결합 속성을 변경시키거나, 핵산 분자 또는 표적 분자 또는 표적 폴리뉴클레오티드에 대한 결합 동역학을 변경시키거나, 또는 핵산 분자에 대한 결합 특이성을 변경시키는 하나 이상의 추가 변형을 더 포함할 수 있다. 이러한 변형의 예는 하기 문단에서 요약한다. 상기 정보에 기반하여, 효소의 불활성화를 야기하거나 또는 닉카제 활성으로 이중 가닥 뉴클레아제를 변형시키는 돌연변이체가 생성될 수 있다. 대안적인 구현예에서, 이러한 정보는 오프-표적 효과가 감소된 효소를 개발하는데 사용된다.

[0472] **변형된 닉카제**

[0473] 돌연변이는 뉴클레아제 활성에 참여하는 아미노산의 이웃 잔기에서 만들어 질 수 있다. 일부 구현예에서, 오직 RuvC 도메인만이 불활성화되고, 다른 구현예에서는, 다른 추정의 뉴클레아제 도메인이 불활성화되며, 여기서 이펙터 단백질 복합체는 닉카제로 기능하고 오직 하나의 DNA 가닥만을 절단한다. 일부 구현예에서, 2 가지 C2c1 변이체(각각 상이한 닉카제)는 특이성을 증가시키는 데 사용되고, 2 가지 닉카제 변이체는 표적(여기서 2 가지 닉카제는 모두, 1 개의 DNA 가닥만이 절단되고 이어서 복구되는 오프-표적 변형을 최소화하거나 제거하면서, DNA 가닥을 절단함)에서 DNA를 절단하는 데 사용된다. 바람직한 구현예에서, C2c1 이펙터 단백질은 2 가지 C2c1 이펙터 단백질 분자를 포함하는 동형이량체로서 관심 표적 유전자좌와 연관되거나 관심 표적 유전자좌에서 서열을 절단한다. 바람직한 구현예에서, 동종이량체는 그들 개별 RuvC 도메인에 상이한 돌연변이를 포함하는 2종의 C2c1 이펙터 단백질 분자를 포함할 수 있다.

[0474] 본 발명은 둘 이상의 닉카제를 사용하는 방법, 특히 듀얼 또는 이중 닉카제 접근법을 고려한다. 일부 양상 및 구현예에서, 단일 유형의 C2c1 닉카제, 예를 들어 변형된 C2c1 또는 본 명세서에 기재된 바와 같은 변형된 C2c1이 전달될 수 있다. 이는 표적 DNA가 2 개의 C2c1 닉카제에 의해 결합되게 한다. 추가적으로 또한, 상이한 오솔로그, 예를 들어 DNA의 하나의 가닥(예를 들어, 코딩 가닥) 상에 C2c1 닉카제, 그리고 비-코딩 또는 반대 DNA 가닥 상에 오솔로그가 사용될 수 있다고 예상된다. 오솔로그는 제한없이 Cas9 닉카제 예컨대 SaCas9 닉카제 또는 SpCas9 닉카제일 수 있다. 상이한 PAM을 요구하고 또한 상이한 가이드 요건을 가져서, 사용자에게 훨씬 더 많은 제어를 허용하는 2종의 상이한 오솔로그를 사용하는 것이 유리할 수 있다. 일정 구현예에서, DNA 절단은 적어도 4 가지 유형의 닉카제를 수반할 것이며, 여기서 각각의 유형은 상이한 서열의 표적 DNA로 가이드되고, 각각의 쌍은 하나의 DNA 가닥 내로 제1 닉을 도입하고 둘째는 제2 DNA 가닥 내로 닉을 도입한다. 이러한 방법에서, 단일 가닥 파손의 적어도 2 개의 쌍이 표적 DNA 내로 도입되고, 여기서 단일-가닥 파손의 제1 및 제2 쌍의 도입 시, 단일-가닥 파손의 제1 및 제2 쌍 사이의 표적 서열은 절단된다. 특정 구현예에서, 오솔로그의 하나 또는 둘 모두는 제어가능하며, 즉 유도성이다.

[0475] 본 발명에 따른 일정 방법에서, CRISPR-Cas 단백질은 바람직하게는 상응하는 야생형 효소에 대해 돌연변이되어서 돌연변이된 CRISPR-Cas 단백질이 표적 서열을 함유하는 표적 유전자좌의 한쪽 또는 양쪽 DNA 가닥을 절단하는 능력이 결여된다. 특정 구현예에서, C2c1 단백질의 하나 이상의 촉매 도메인은 표적 서열의 오직 하나의

DNA 가닥을 절단하는 돌연변이된 Cas 단백질을 생성시키도록 돌연변이된다.

[0476] 본 명세서에서 제공되는 방법의 일정 구현예에서, CRISPR-Cas 단백질은 오직 하나의 DNA 가닥을 절단하는 돌연변이된 CRISPR-Cas 단백질, 즉 닉카제이다. 보다 특히, 본 발명에서, 닉카제는 비-표적 서열, 즉 표적 서열의 반대쪽 DNA 가닥 상에 있고 PAM 서열의 3'에 있는 서열 내에서 절단을 보장한다. 추가 지침으로서, 제한없이, 알리시클로바실러스 악시도테레스트리스 (Alicyclobacillus acidoterrestris) 유래 C2c1의 Nuc 도메인에서 아르기닌 대 아르기닌 (arginine-to-alanine) 치환 (R911A)은 C2C1을 양쪽 가닥을 절단하는 뉴클레아제에서 (단일 가닥을 절단하는) 닉카제로 전환시킨다. 효소가 AacC2c1이 아닌 경우에, 돌연변이는 상응하는 위치 내 잔기에서 만들어질 수 있다는 것을 당업자는 이해하게 될 것이다.

[0477] 일정 구현예에서, C2c1 단백질은 Nuc 도메인에 돌연변이를 포함하는 C2c1 닉카제이다. 일부 구현예에서, C2c1 닉카제는 알리시클로바실러스 악시도테레스트리스 C2c1의 아미노산 위치 R911, R1000, 또는 R1015에 상응하는 돌연변이를 포함한다. 일부 구현예에서, C2c1 닉카제는 알리시클로바실러스 악시도테레스트리스 C2c1의 R911A, R1000A, 또는 R1015A에 상응하는 돌연변이를 포함한다. 일부 실시형태에서, C2c1 닉카제는 바실러스 sp. V3-13 C2c1 의 R894A에 상응하는 돌연변이를 포함한다. 일정 구현예에서, C2c1 단백질은 단백질의 비돌연변이 또는 비변형된 형태와 비교하여 증가되거나 또는 감소된 특이성으로 PAM을 인식한다. 일부 구현예에서, C2c1 단백질은 단백질의 비돌연변이 또는 비변형된 형태와 비교하여 변경된 PAM을 인식한다.

[0478] **탈활성화/불활성화 C2c1 단백질**

[0479] C2c1 단백질이 뉴클레아제 활성을 갖는 경우에, 이 단백질은 야생형 효소와 비교하여 감소된 뉴클레아제 활성, 예를 들어 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 97% 또는 100%의 뉴클레아제 불활성화를 갖도록 변형될 수 있거나, 또는 다른 방식으로 C2c1 효소는 유리하게 비돌연변이 또는 야생형 C2c1 효소 또는 CRISPR 효소의 뉴클레아제 활성의 약 0%를 갖거나, 또는 비돌연변이 또는 야생형 C2c1 효소의 뉴클레아제 활성의 약 3% 또는 약 5% 또는 약 10%를 넘지 않는다. 일부 구현예에서, CRISPR-Cas 단백질은 돌연변이된 효소의 DNA 절단 활성이 효소의 비돌연변이된 DNA 절단 활성의 약 25%, 10%, 5%, 1%, 0.1%, 0.01% 이하일 때 모든 DNA 절단 활성이 실질적으로 결여된 것으로 간주되고, 예로는 돌연변이된 형태의 DNA 절단 활성이 비돌연변이된 형태와 비교하여 무시할만하거나 또는 전무한 경우일 수 있다. 이들 구현예에서, CRISPR-Cas 단백질은 포괄적인 DNA 결합 단백질로서 사용된다. 이것은 돌연변이를 C2c1 및 이의 오솔로그의 뉴클레아제 도메인 내로 도입시켜 가능하다.

[0480] 특정 구현예에서, CRISPR 효소는 조작되어 뉴클레아제 활성을 감소시키거나 제거하는 하나 이상의 돌연변이를 포함할 수 있다.

[0481] 일정 구현예에서, C2c1 단백질은 RuvC 도메인에 돌연변이를 포함하는 촉매적 불활성 C2c1이다. 일부 구현예에서, 촉매적으로 불활성인 C2c1 단백질은 알리시클로바실러스 악시도테레스트리스 C2c1 내 아미노산 위치 D570, E848, 또는 D977에 상응하는 돌연변이를 포함한다. 일부 구현예에서, 촉매적으로 불활성인 C2c1 단백질은 알리시클로바실러스 악시도테레스트리스 C2c1 내 D570A, E848A, 또는 D977A에 상응하는 돌연변이를 포함한다.

[0482] 일부 구현예에서, 촉매적 불활성 C2c1 단백질은 바실러스 히사시이 C2c1의 아미노산 위치 D574, E828, 또는 D952에 상응하는 돌연변이를 포함한다. 일부 구현예에서, 촉매적으로 불활성인 C2c1 단백질은 바실러스 히사시이 C2c1 내 D574A, E828A, 또는 D952A에 상응하는 돌연변이를 포함한다.

[0483] 일부 구현예에서, 촉매적 불활성 C2c1 단백질은 바실러스 sp. V3-13 C2c1의 아미노산 위치 D567, E831, 또는 D963에 상응하는 돌연변이를 포함한다. 일부 구현예에서, 촉매적으로 불활성인 C2c1 단백질은 바실러스 sp. V3-13 C2c1 내 D567A, E831A, 또는 D963A에 상응하는 돌연변이를 포함한다.

[0484] 일정 구현예에서, C2c1 단백질은 RuvC 도메인에 돌연변이를 포함하는 촉매적 불활성 C2c1이다. 일부 구현예에서, 촉매적으로 불활성인 C2c1 단백질은 알리시클로바실러스 악시도테레스트리스 C2c1 내 아미노산 위치 D570, E848, 또는 D977에 상응하는 돌연변이를 포함한다. 일부 구현예에서, 촉매적으로 불활성인 C2c1 단백질은 알리시클로바실러스 악시도테레스트리스 C2c1 내 D570A, E848A, 또는 D977A에 상응하는 돌연변이를 포함한다.

[0485] 일부 구현예에서, 촉매적 불활성 C2c1 단백질은 바실러스 히사시이 C2c1의 아미노산 위치 D574, E828, 또는 D977에 상응하는 돌연변이를 포함한다. 일부 구현예에서, 촉매적으로 불활성인 C2c1 단백질은 바실러스 히

사시이 C2c1 내 D574A, E828A, 또는 D977A에 상응하는 돌연변이를 포함한다.

[0486] 일부 구현예에서, 촉매적 불활성 C2c1 단백질은 바실러스 sp. V3-13 C2c1의 아미노산 위치 D567, E831, 또는 D963에 상응하는 돌연변이를 포함한다. 일부 구현예에서, 촉매적으로 불활성인 C2c1 단백질은 바실러스 sp. V3-13 C2c1 내 D567A, E831A, 또는 D963A에 상응하는 돌연변이를 포함한다.

[0487] 일정 구현예에서, C2c1 단백질은 Nuc 도메인에 돌연변이를 포함하는 C2c1 닉카제이다. 일부 구현예에서, C2c1 닉카제는 알리시클로바실러스 악시도테레스트리스 C2c1의 아미노산 위치 R911, R1000, 또는 R1015에 상응하는 돌연변이를 포함한다. 일부 구현예에서, C2c1 닉카제는 알리시클로바실러스 악시도테레스트리스 C2c1의 R911A, R1000A, 또는 R1015A에 상응하는 돌연변이를 포함한다. 일부 실시형태에서, C2c1 닉카제는 바실러스 sp. V3-13 C2c1에 상응하는 돌연변이 R894A를 포함한다. 일정 구현예에서, C2c1 단백질은 단백질의 비돌연변이 또는 비변형된 형태와 비교하여 증가되거나 또는 감소된 특이성으로 PAM을 인식한다. 일부 구현예에서, C2c1 단백질은 단백질의 비돌연변이 또는 비변형된 형태와 비교하여 변경된 PAM을 인식한다.

[0488] 일부 구현예에서, CRISPR-Cas 단백질은 돌연변이된 효소의 DNA 절단 활성이 효소의 비돌연변이된 DNA 절단 활성의 약 25%, 10%, 5%, 1%, 0.1%, 0.01% 이하일 때 모든 DNA 절단 활성이 실질적으로 결여된 것으로 간주되고, 예로는 돌연변이된 형태의 DNA 절단 활성이 비돌연변이된 형태와 비교하여 무시할만하거나 또는 전무한 경우일 수 있다. 이들 구현예에서, CRISPR-Cas 단백질은 포괄적인 DNA 결합 단백질로서 사용된다. 돌연변이는 인공적으로 도입된 돌연변이일 수 있거나 또는 기능 획득 또는 기능 상실 돌연변이일 수 있다.

[0489] 상기 기재한 돌연변이에 추가로, CRISPR-Cas 단백질은 추가적으로 변형될 수 있다. 본 명세서에서 사용되는, CRISPR-Cas 단백질에 대한 용어 " 변형된" 은 일반적으로 그것이 유래된 야생형 Cas 단백질과 비교하여 하나 이상의 변형 또는 돌연변이 (점 돌연변이, 절단, 삽입, 결실, 키메라, 융합 단백질 등 포함)를 갖는 CRISPR-Cas 단백질을 의미한다. 유래된 이란 유래된 효소가 대체로 높은 서열 상동성 정도를 갖는다는 의미에서, 야생형 효소를 기반으로 하지만, 당분야에 공지되거나 또는 본 명세서에 기술된 바와 같이 일부 방식으로 돌연변이 (변형)되었다는 것을 의미한다.

[0490] 비활성화된 C2c1 CRISPR 효소는, 예를 들어 디아미나제 활성 메틸라제 활성, 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, RNA 절단 활성, DNA 절단 활성, 핵산 결합 활성 및 분자 스위치 (예를 들어, 광 유도성)를 포함하거나, 또는 이로 본질적으로 이루어지거나, 또는 이로 이루어진 군으로부터의 하나 이상의 도메인을 포함하여, (예를 들어, 융합 단백질을 통해) 회합된 하나 이상의 기능적 도메인을 가질 수 있다. 본 발명의 방법에서 사용을 위해 적합한 링커는 당업자에게 충분히 공지되어 있고 제한없이, 직쇄 또는 분지쇄 탄소 링커, 복소환 탄소 링커, 또는 펩티드 링커를 포함한다. 그러나, 본 명세서에서 사용되는 링커는 또한 공유 결합 (탄소-탄소 결합 또는 탄소-이종원자 결합)일 수 있다. 특정 구현예에서, 각각의 단백질이 그의 필요한 기능적 특성을 보유하는 것을 보장하는 데 충분한 거리만큼 표적화 도메인 및 아데노신 디아미나제를 분리시키기 위해 링커가 사용된다. 바람직한 펩티드 링커 서열은 가요성의 연장된 입체배좌를 채택하고, 정돈된 2차 구조를 발생시키는 경향을 나타내지 않는다. 일정 구현예에서, 링커는 단량체, 이량체, 다량체 또는 중합체일 수 있는 화학적 모이어티일 수 있다. 바람직하게는, 링커는 아미노산을 포함한다. 가요성 링커 내 전형적인 아미노산은 Gly, Asn 및 Ser을 포함한다. 따라서, 특정 구현예에서, 링커는 Gly, Asn 및 Ser 아미노산 중 하나 이상의 조합물을 포함한다. 다른 근처의 중성 아미노산, 예컨대, Thr 및 Ala은 또한 링커 서열에서 사용될 수 있다. 예시적인 링커는 [Maratea et al. (1985), Gene 40: 39-46]; [Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62]; 미국 특허 출원 번호 4,935,233; 및 미국 특허 출원 번호 4,751,180에 개시된다. 예를 들어, GlySer 링커 GGS, GGGS (SEQ ID NO:402) 또는 GSG가 사용될 수 있다. GGS, GSG, GGGS 또는 GGGGS (SEQ ID NO:403) 링커는 적합한 길이를 제공하기 위해서, 3 (예컨대 (GGS)3 (SEQ ID NO:404), (GGGGS)3 (SEQ ID NO:393) 또는 5 (SEQ ID NO:405), 6 (SEQ ID NO:394), 7 (SEQ ID NO:406), 9 (SEQ ID NO:395) 또는 심지어 12 (SEQ ID NO:396) 또는 그 이상의 반복부로 사용될 수 있다. 특정한 구현예에서, (GGGGS)3과 같은 링커가 바람직하게 본 명세서에서 사용된다. (GGGGS)6 (GGGGS)9 또는 (GGGGS)12가 바람직하게는 대안으로서 사용될 수 있다. 다른 바람직한 대안은 (GGGGS)1 (SEQ ID NO:403), (GGGGS)2 (SEQ ID NO:407), (GGGGS)4 (SEQ ID NO:408), (GGGGS)5 (SEQ ID NO:405), (GGGGS)7 (SEQ ID NO:406), (GGGGS)8 (SEQ ID NO:409), (GGGGS)10 (SEQ ID NO:410), 또는 (GGGGS)11 (SEQ ID NO:411)이다. 또한 추가 구현예에서, LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO:412)은 링커로서 사용된다. 또한 추가적인 구현예에서, 링커는 XTEN 링커이다. 추가로, N- 및 C-말단의 NLS는 또한 링커(예를 들어, PKKKRKVEASSPKKRKVEAS (SEQ ID NO:413)로서 작용할 수 있다.

[0491] 링커의 예는 하기 표에 표시된다.

GGS	GGTGGTAGT (SEQ ID NO:414)
GGSx3 (9)	GGTGGTAGTGGAGGGAGCGGCGGTTCA (SEQ ID NO:415)
GGSx7 (21)	ggtggaggaggctctggtggaggcggtagcggaggcggagggtcgGGTGGTAGTGGAGG GAGCGGCGGTTCA (SEQ ID NO:416)
XTEN	TCGGGATCTGAGACGCCTGGGACCTCGGAATCGGCTACGCCCGAA AGT (SEQ ID NO:417)
Z-EGFR_Short	Gtggataacaaatttaacaaagaaatgtgggcggcgtgggaagaaattcgtaacctgccgaacctgaacgg ctggcagatgaccgcgtttattgcgagcctggtggatgatccgagccagagcgcgaacctgctggcggaag cgaaaaaactgaacgatgcgcaggcgccgaaaaccggcggtggttctggt (SEQ ID NO:418)
GSAT	Ggtggttctgccggtggctccggttctggctccagcggtggcagctctggtgcgtccggcacgggtactgc gggtggcactggcagcggttccggtactggctctggc (SEQ ID NO:419)

[0492]

[0493] 예시적인 기능성 도메인은 아데노신 디아미나제 도메인 함유 (ADAD) 패밀리 멤버, Fok1, VP64, P65, HSF1, MyoD1이다. 디아미나제가 제공되는 사건에서, 가이드 서열은 가이드 서열과 표적 서열 사이에 형성된 RNA 듀플렉스 또는 RNA/DNA 헤테로듀플렉스에 하나 이상의 미스매치를 도입시키도록 설계되는 것이 유리하다. 특정 구현예에서, 가이드 서열과 표적 서열 사이의 듀플렉스는 A-C 미스매치를 포함한다. Flk1이 제공되는 사건에서, 다수의 Fok1 기능성 도메인이 제공되어 기능성 이량체를 허용하고 gRNA는 [Tsai et al. Nature Biotechnology, Vol. 32, Number 6, June 2014]에 특별히 기술된 바와 같이 기능적 사용 (Fok1)을 위한 적절한 공간을 제공하도록 설계되는 것이 유리하다. 어댑터 단백질은 이러한 기능성 도메인에 부착되는 공지된 링커를 이용할 수 있다. 일부 경우에, 추가적으로 적어도 하나의 NLS가 제공되는 것이 유리하다. 일부 예에서, N 말단에 NLS를 위치시키는 것이 유리하다. 하나 초과의 기능성 도메인이 포함될 때, 기능성 도메인은 동일하거나 또는 상이할 수 있다.

[0494] 일반적으로, 비활성화된 C2c1 효소 상에의 하나 이상의 기능적 도메인의 위치 선정은 기능적 도메인이 귀속된 작용 효과로 표적에 작용하도록 공간 배향을 수정하는 것을 가능하게 하는 것이다. 예를 들어, 기능적 도메인이 전사 활성화제(예를 들어, VP64 또는 p65)인 경우, 전사 활성화제는, 표적의 전사에 작용하도록 하는 공간 배향으로 위치된다. 마찬가지로, 전사 억제인자는 표적의 전사에 영향을 미치도록 유리하게 위치될 것이며, 뉴클레아제(예를 들어, Fok1)는 표적을 절단하거나 또는 부분적으로 절단하도록 유리하게 위치될 것이다. 이는 CRISPR 효소의 N-말단/C-말단 외의 위치를 포함할 수 있다. 기능성 도메인은 표적 DNA 서열의 전사 또는 번역을 변형시킨다.

[0495] 일부 구현예에서, Cas12b 이펙터 단백질은 하나 이상의 기능성 도메인과 회합되고; Cas12b 이펙터 단백질은 RuvC 및/또는 Nuc 도메인 내에 하나 이상의 돌연변이를 함유하여서, 형성된 CRISPR 복합체가 후성적 변형자 또는 전사 또는 번역 활성화 또는 억제 신호를 전달할 수 있다.

[0496] 일정 구현예에서, 본 명세서에 개시된 CRISPR-Cas 시스템은 자가-불활성화 시스템이고 Cas 이펙터 단백질은 일시적으로 발현된다. 일부 구현예에서, 자가-불활성화 시스템은 바이러스 벡터 예컨대 AAV 벡터를 포함한다. 일부 구현예에서, 자가-불활성화 시스템은 내생성 표적 서열과 80%, 81%, 82%, 83%, 84%, 85%, 86%, 97%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%의 동일성을 공유하는 DNA 서열을 포함한다. 일부 구현예에서, 자가-불활성화 시스템은 둘 이상의 벡터 시스템을 포함한다. 일부 구현예에서, 자가-불활성화 시스템은 단일 벡터를 포함한다. 일부 구현예에서, 자가-불활성화 시스템은 내생성 DNA 표적 서열 및 Cas 이펙터 단백질을 코딩하는 벡터 서열을 동시에 표적화하는 Cas 이펙터 단백질을 포함한다. 일부 구현예에서, 자가-불활성화 시스템은 내생성 DNA 표적 서열 및 Cas 이펙터 단백질을 코딩하는 벡터 서열을 순차적으로 표적화하는 Cas 이펙터 단백질을 포함한다. 일부 구현예에서, Cas 이펙터를 코딩하는 뉴클레오티드 및 가이드 서열은 단일 벡터 상의 개별 조절 엘리먼트에 작동적으로 연결된다. 일부 구현예에서, Cas 이펙터를 코딩하는 뉴클레오티드 및 가이드 서열은 개별 벡터 상에서 개별 조절 엘리먼트에 작동적으로 연결된다. 일부 실시형태들에서, 조절 엘리먼트는 항상성이다. 일부 실시형태들에서, 조절 엘리먼트는 유도성이다.

[0497] **탈안정화된 C2c1**

[0498] 일 구현예에서, 본원에 기재된 바와 같은 본 발명에 따른 이펙터 단백질 (CRISPR 효소; C2c1)은 탈안정화 도메인 (DD)과 회합되거나 또는 융합된다. 일부 구현예에서, DD은 ER50이다. 이 DD에 대한 상응하는 안정화 리간드는, 일부 구현예에서, 4HT이다. 이와 같이, 일부 구현예에서, 적어도 하나의 DD 중 하나는 ER50이고

이의 안정화 리간드는 4HT 또는 CMP8이다. 일부 실시형태에서, DD는 DHFR50이다. 이 DD에 대한 상응하는 안정화 리간드는, 일부 구현예에서, TMP이다. 이와 같이, 일부 구현예에서, 적어도 하나의 DD 중 하나는 DHFR50이고, 이의 안정화 리간드는 TMP이다. 일부 실시형태에서, DD는 ER50이다. 이 DD에 대한 상응하는 안정화 리간드는, 일부 구현예에서, CMP8이다. 그러므로, CMP8은 ER50 시스템에서 4HT에 대한 대안적인 안정화 리간드일 수 있다. CMP8 및 4HT가 경쟁적 방식으로 사용될 수 있고/사용되어야 하는 것이 가능할 수 있지만, 일부 세포 유형은 이들 2개 리간드 중 하나 또는 다른 하나에 더 감수성일 수 있으며, 본 개시 및 당분야의 지식으로부터 당업자는 CMP8 및/또는 4HT를 사용할 수 있다.

[0499] 일부 구현예에서, 하나 또는 2개의 DD는 CRISPR 효소의 N-말단에 융합될 수 있고, 하나 또는 2개의 DD는 CRISPR 효소의 C-말단에 융합된다. 일부 구현예에서, 적어도 2개의 DD는 CRISPR 효소와 회합되고 DD는 동일한 DD이고, 즉 DD는 상동성이다. 따라서, DD 둘 모두 (또는 둘 이상)은 ER50 DD일 수 있다. 이는 일부 구현예에서 바람직하다. 대안적으로, DD 둘 모두 (또는 둘 이상)은 DHFR50 DD일 수 있다. 이는 또한 일부 구현예에서 바람직하다. 일부 구현예에서, 적어도 2개의 DD는 CRISPR 효소와 회합되고 DD는 상이한 DD이며, 즉 DD는 이종성이다. 따라서, DD 중 하나는 ER50일 수 있는 한편 DD 중 하나 이상 또는 임의의 다른 DD는 DHFR50일 수 있다. 더 높은 수준의 분해 제어도를 제공하므로 이종성인 2 이상의 DD를 갖는 것이 유리할 수 있다. N 또는 C-말단에서 하나 초과의 DD의 탠덤 융합은 분해를 향상시킬 수 있으며; 이러한 탠덤 융합은 예를 들어 ER50-ER50-C2c1 일 수 있다. 이는, 어느 하나의 안정화 리간드 부재시 높은 수준의 분해가 발생할 것이며, 하나의 안정화 리간드 부재 및 기타 다른(또는 다른) 안정화 리간드의 존재시 중간 수준의 분해가 발생할 것인 한편, 2 개(또는 둘 이상)의 모든 안정화 리간드의 존재 하에서 낮은 수준의 분해가 발생할 것임이 생각된다. 또한 N-말단 ER50 DD 및 C-말단 DHFR50 DD를 가져서 제어성이 부여될 수 있다.

[0500] 일부 구현예에서, DD와 CRISPR 효소의 융합은 DD와 CRISPR 효소 사이에 링커를 포함한다. 일부 구현예에서, 링커는 GlySer 링커이다. 일부 구현예에서, DD-CRISPR 효소는 적어도 하나의 핵 이출 신호 (NES)를 더 포함한다. 일부 구현예에서, DD-CRISPR 효소는 둘 이상의 NES를 포함한다. 일부 구현예에서, DD-CRISPR 효소는 적어도 하나의 핵 국재화 신호(NLS)를 포함한다. 이는 NES에 추가적일 수 있다. 일부 구현예에서, CRISPR 효소는 CRISPR 효소 및 DD 사이의 링커로서, 또는 링커의 일부로서 국재화 (핵 이입 또는 이출) 신호를 포함하거나 또는 그로 본질적으로 이루어지거나 또는 그로 이루어진다. HA 또는 Flag 태그는 또한 링커로서 본 발명의 영역 내에 있다. 출원인은 링커로서 NLS 및/또는 NES를 사용하고, 또한 GS 내지 (GGGGS)3 정도로 짧은 글리신 세린 링커를 사용한다.

[0501] 탈안정화 도메인은 광범위한 단백질에 불안정성을 부여하는데 일반적인 유용성을 가지며, 예를 들어 참조로 본 명세서에 편입되는 [Miyazaki, J Am Chem Soc. Mar 7, 2012; 134(9): 3942-3945]을 참조한다. CMP8 또는 4-히드록시탐옥시펜은 탈안정화 도메인일 수 있다. 보다 일반적으로, N-말단 규칙에 의한 탈안정화 잔기, 포유동물 DHFR의 온도-감수성 돌연변이체 (DHFRts)는 허용되는 온도에서는 안정하지만, 37℃에서는 불안정한 것으로 확인되었다. DHFRts를 발현하는 세포에, 포유동물 DHFR에 대한 고-친화성 리간드, 메토트렉세이트의 첨가는 단백질의 분해를 부분적으로 억제하였다. 이는 소형 분자 리간드가, 그렇지 않으면 세포 내에서 분해를 위해 표적화될 단백질을 안정화할 수 있다는 중요한 설명이었다. 라파마이신 유도체를 사용하여 mTOR (FRB*)의 FRB 도메인의 불안정 돌연변이체를 안정화시켰고 융합된 키나제, GSK-3β.6,7의 기능을 복원시켰다.

이 시스템은 리간드-의존적 안정성이 복잡한 생물학적 환경에서 특이적 단백질의 기능을 조절하기 위한 매력적인 전략을 의미하였다는 것을 입증하였다. 단백질 활성을 제어하기 위한 시스템은 유비퀴틴 상보성이, FK506-결합 단백질 및 FKBP12의 라파마이신 유도된 이량체에 의해 발생하는 경우, 기능적이게 되는 DD를 수반할 수 있다. 인간 FKBP12 또는 ecDHFR 단백질의 돌연변이체는 이의 고친화성 리간드, 실드(Shield)-1 또는 트리메토프림(TMP) 각각의 부재 하에서 대사적으로 불안정하도록 조작될 수 있다. 이들 돌연변이체는 본 발명의 실시에 유용한 가능한 탈안정화 도메인(DD)의 일부이며, CRISPR 효소와의 융합물로서 DD의 불안정성이 프로테아솜에 의한 전체 융합 단백질의 CRISPR 단백질 분해에 부여된다. 실드-1 및 TMP는 용량-의존적 방식으로 DD에 결합하고 이를 안정화시킨다. 에스트로겐 수용체 리간드 결합 도메인 (ERLBD, ERS1의 잔기 305-549)는 또한 탈안정화 도메인으로서 조작될 수 있다. 에스트로겐 수용체 신호전달 경로가 다양한 질환 예컨대 유방암에 관여되므로, 그 경로는 폭넓게 연구되었고 에스트로겐 수용체의 수많은 효현제 및 길항제가 개발되었다.

따라서, ERLBD와 약물의 상용성 쌍이 알려져 있다. 돌연변이체에 결합하지만 ERLBD의 야생 형태에는 결합하지 않는 리간드가 존재한다. 3 개의 돌연변이를 코딩하는 이들 돌연변이체 도메인들 (L384M, M421G, G521R)12 중 하나를 사용함으로써, 내생성 에스트로겐-민감성 네트워크를 동요시키지 않는 리간드를 사용하여 ERLBD-유도된 DD의 안정성을 조절하는 것이 가능하다. 추가적인 돌연변이(Y537S)가 도입되어 ERLBD를 추가로 불안정화시키고, 잠재적 DD 후보물로서 이를 구성할 수 있다. 이러한 테트라-돌연변이체는 유리한 DD 개

발이다. 돌연변이체 ERLBD는 CRISPR 효소에 융합될 수 있고, 이의 안정성은 리간드를 사용하여 조절 또는 동요될 수 있으며, 이로 인해 CRISPR 효소는 DD를 갖는다. 다른 DD는 쉴드-1 리간드에 의해 안정화된 돌연변이된 FKBP 단백질을 기반으로 12-kDa (107-아미노산) 태그일 수 있으며, 예를 들어, 문헌 [Nature Methods 5, (2008)]을 참조한다. 예를 들어, DD는 합성된, 생물학적으로 불활성인 소형 분자, 쉴드-1에 결합하여 그에 의해 가역적으로 안정화되는 변형된 FK506 결합 단백질 12 (FKBP12)일 수 있다; 예를 들어, 참조: Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell. 2006;126:9951004; Banaszynski LA, Sellmyer MA, Contag CH, Wandless TJ, Thorne SH. Chemical control of protein stability and function in living mice. Nat Med. 2008;14:11231127; Maynard-Smith LA, Chen LC, Banaszynski LA, Ooi AG, Wandless TJ. A directed approach for engineering conditional protein stability using biologically silent small molecules. The Journal of biological chemistry. 2007;282:2486624872; 및 Rodriguez, Chem Biol. 391398 - 이들 모두는 본 명세서에 참고로서 포함되어 있으며, 본 발명의 실시에서 CRISPR 효소와 회합시키기 위해 선택된 DD에서 본 발명의 실시에 이용될 수 있다. 확인할 수 있는 바와 같이, 당분야의 지식은 수많은 DD를 포함하고, DD는 유리하게 링커와 함께, CRISPR 효소와 회합, 예를 들어 융합되고, 그리하여 DD는 리간드의 존재 하에서 안정화될 수 있고 이의 부재시에 DD는 탈안정화될 수 있고, 그리하여 CRISPR 효소가 전체적으로 탈안정화되거나, 또는 DD가 리간드의 부재 하에서 안정화될 수 있고 리간드가 존재하는 경우 DD는 탈안정화될 수 있으며, DD는 CRISPR 효소 및 그에 따라 CRISPR-Cas 복합체 또는 시스템이 조절되거나 또는 제어될 수 있어서, 말하자면 켜지거나 또는 꺼질 수 있어서, 그렇게 함으로써 예를 들어 생체내 또는 시험관 환경에서 시스템의 조절 또는 제어를 위한 수단을 제공할 수 있다. 예를 들어, 관심 단백질이 DD 태그를 이용하여 융합물로서 발현된 경우, 이는 세포 내에서, 예를 들어 프로테아솜에 의해 탈안정화되고 신속히 분해된다. 따라서, 안정화 리간드의 부재는 분해되는 D 연합된 Cas를 초래한다. 새로운 DD가 관심 단백질에 융합된 경우, 이의 불안정성이 관심 단백질에 부여되어, 전체 융합 단백질의 신속한 분해를 초래한다.

Cas에 대한 피크 활성도는 때때로 표적외 효과를 감소시키는 데 유익하다. 따라서, 높은 활성의 짧은 파열이 바람직하다. 본 발명은 이러한 피크를 제공할 수 있다. 일부 의미에서, 시스템은 유도성이다. 일부 기타 다른 의미에서, 안정화 리간드의 부재 하에서 시스템은 억제되고, 안정화 리간드의 존재 하에서 탈-억제된다.

[0502] **분할 디자인**

[0503] C2c1은 또한 강력한 핵산 검출을 할 수 있게 한다. 일정 구현예에서, C2c1은 이의 뉴클레아제 활성의 불활성화를 통해서 핵산 결합 단백질 ("데드 C2c1; dC2c1)로 전환된다. 핵산 결합 단백질로 전환될 때, C2c1은 서열 의존적 방식으로 표적 핵산으로 다른 기능성 성분을 국재화시키는데 유용하다. 성분은 천연일 수 있거나 또는 합성일 수 있다. 본 발명에 따라서, dC2c1은 (i) 대규모 스크리닝, 합성 조절 회로의 구축 및 다른 목적을 위해 사용할 수 있는, 기능 또는 전사를 조절하기 위해 특이적 핵산으로 이펙터 모듈을 가져하고; (ii) 그들 수송 및/또는 국재화를 가시화시키기 위해 특이적 핵산을 형광적으로 태그화하고; (iii) 특이적 세포하 구획에 대한 친화성을 갖는 도메인을 통해 핵산 국재화를 변경시키고; (iv) RNA 및 단백질을 포함한, 근위 분자 파트너에 대해 농축시키기 위해 특이적 핵산을 (dC2c2의 직접 풀다운 또는 바이오틴 리가제 활성 국재화를 위한 dC2c2의 사용을 통해) 포획하는데 사용된다. dC2c1은 i) 세포의 성분을 조직화시키고, ii) 세포의 성분 또는 활성을 켜거나 또는 끄고, iii) 세포에 존재하는 특이적 전사물의 존재 또는 양을 기반으로 세포 상태를 제어하는데 사용될 수 있다. 예시적인 구현예에서, 본 발명은 분할 효소 및 리포터 분자를 제공하고, 이들의 일부는 핵산-결합 CRISPR 이펙터, 예컨대 제한없이 C2c1을 포함하는 하이브리드 분자로 제공된다. 세포 내 핵산의 존재에 가까워질 때, 분할 리포터 또는 효소의 활성은 재구성되고 활성을 측정할 수 있게 된다. 이러한 방식으로 재구성된 분할 효소는 제한없이 내생성 성분 또는 경로, 또는 외생성 성분 또는 경로를 포함하여, 세포 성분 및/또는 경로에 대해 검출가능하게 작용할 수 있다. 이러한 방식으로 재구성된 분할 리포터는 검출가능한 신호, 예컨대 제한없이 형광 또는 다른 검출가능한 모이어티를 제공할 수 있다. 일정 구현예에서, 분할 단백질가수분해 효소는 검출가능한 방식으로 하나 이상의 성분 (내생성 또는 외생성)에 작용할 때 제공된다. 예시적인 일 구현예에서, 세포 내 핵산 종의 검출 시 프로그램된 세포 사멸을 유도하는 방법이 제공된다. 이러한 방법이 어떻게 예를 들어 세포 내 바이러스의 존재를 기반으로, 세포 개체군을 제거하는데 사용될 수 있는가는 자명해질 것이다.

[0504] 본 발명에 따라서, 관심 핵산을 함유하는 세포에서 세포 사멸을 유도하는 방법을 제공하고, 방법은 세포 내 핵산을, 세포 사멸을 유도할 수 있는 단백질가수분해 효소의 불활성 제1 부분에 연결된 제1 CRISPR 단백질, 효소의 상보성 부분에 연결된 제2 CRISPR 단백질로서, 단백질 가수분해 효소의 효소 활성은 단백질의 제1 부분 및

상보성 부분이 접촉될 때 재구성되는 것인 단백질, 및 제1 CRISPR 단백질에 결합하고 핵산의 제1 표적 서열에 하이브리드화하는 제1 가이드, 및 제2 CRISPR 단백질에 결합하고 핵산의 제2 표적 서열에 하이브리드화하는 제2 가이드를 포함하는 조성물과 접촉시키는 단계를 포함한다. 관심 표적 핵산이 존재할 때, 단백질가수분해 효소의 제1 및 제2 부분이 접촉되고 효소의 단백질가수분해 활성이 재구성되어 세포 사멸을 유도한다. 본 발명의 일 구현예에서, 단백질가수분해 효소는 캐스파제이다. 다른 이러한 구현예에서, 단백질가수분해 효소는 TEV 프로테아제이고, TEV 프로테아제의 단백질가수분해 활성이 재구성될 때, TEV 프로테아제 기질이 절단되고/되거나 활성화된다. 본 발명의 일 구현예에서, TEV 프로테아제 기질은 TEV 프로테아제가 재구성될 때, 프로캐스파제가 절단되고 활성화되어서, 아폽토시스가 발생되게 하는 조작된 프로캐스파제이다. 본 발명의 일 구현예에서, 단백질가수분해적으로 절단가능한 전사 인자는 " 전사-커플링된" 리포터 시스템을 생산하기 위한 임의의 선택된 하류 리포터 유전자와 조합될 수 있다. 일 구현예에서, 분할 프로테아제는 검출가능한 기질로부터 데그론을 절단하거나 또는 노출시키는데 사용된다.

[0505] 본 발명에 따라서, 관심 핵산을 함유하는 세포를 마킹하거나 또는 확인하는 방법을 제공하고, 방법은 세포 내 핵산을, 단백질가수분해 효소의 불활성 제1 부분에 연결된 제1 CRISPR 단백질, 효소의 상보성 부분에 연결된 제2 CRISPR 단백질로서, 단백질 가수분해 효소의 효소 활성은 단백질의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 단백질, 및 제1 CRISPR 단백질에 결합하고 핵산의 제1 표적 서열에 하이브리드화하는 제1 가이드, 제2 CRISPR 단백질에 결합하고 핵산의 제2 표적 서열에 하이브리드화하는 제2 가이드, 및 재구성된 단백질가수분해 효소에 의해 검출가능하게 절단되는 지시자를 포함하는 조성물과 접촉시키는 단계를 포함한다. 단백질가수분해 효소의 제1 및 제2 부분은 관심 핵산이 세포에 존재할 때 접촉되고, 그리하여 단백질가수분해 효소의 활성이 재구성되고 지시자가 검출가능하게 절단된다. 이러한 일 구현예에서, 검출가능한 지시자는 형광 단백질, 예컨대, 제한없이 녹색 형광 단백질이다. 이러한 다른 구현예에서, 검출가능한 지시자는 발광 단백질, 예컨대, 제한없이 루시퍼라제이다. 일 구현예에서, 분할 리포터는 레닐라 레니포르미스 (Renilla reniformis) 루시퍼라제 (Rluc)의 분할 단편의 재구성을 기반으로 한다. 본 발명의 일 구현예에서, 분할 리포터는 노란색 형광 단백질 (YFP)의 2개 비형광 단편 간 상보성을 기반으로 한다.

[0506] **전사 및 조절**

[0507] 일 양상에서, 본 발명은 세포 또는 조직 내 특정 핵산의 존재 또는 수준을 기반으로 세포 또는 조직의 상태를 확인, 측정, 및/또는 조절하는 방법을 제공한다. 일 구현예에서, 본 발명은 선택된 관심 핵산 종의 존재를 기반으로, 천연 또는 합성 시스템 또는 성분일 수 있는, 세포 시스템 또는 성분의 존재 및/또는 활성을 조절하도록 설계된 CRISPR-기반 제어 시스템을 제공한다. 일반적으로, 제어 시스템은 선택된 관심 핵산 종이 존재할 때 재구성되는 불활성화된 단백질, 효소, 또는 활성을 특징으로 한다. 본 발명의 일 구현예에서, 불활성화된 단백질, 효소, 또는 활성의 재구성은 불활성 성분을 함께 집합시켜 활성 복합체를 조립시키는 것을 포함한다.

[0508] **분할 아폽토시스 구성체**

[0509] 특별한 세포 유형의 역할을 연구하기 위한 기초 생물학 적용분야 또는 암 또는 감염된 세포 제거같은 치료적 적용분야를 위해서, 이상 내생성 또는 외래 DNA의 존재를 기반으로 세포를 고갈 또는 사멸시키는 것이 종종 바람직하다 (Baker, D.J., Childs, B.G., Durik, M., Wijers, M.E., Sieben, C.J., Zhong, J., Saltness, R.A., Jeganathan, K.B., Verzosa, G.C., Pezeshki, A., et al. (2016). Naturally occurring p16(Ink4a)-positive cells shorten healthy lifespan. Nature 530, 184-189.). 이러한 표적화된 세포 사멸은 DNA와의 결합시 표적화된 유전자 또는 유전자 세트를 특이적으로 발현하는 세포의 사멸을 초래하는, C2c1 단백질에 분할 아폽토시스 도메인의 융합을 통해서 달성될 수 있다. 일정 구현예에서, 아폽토시스 도메인은 분할 캐스파제 3일 수 있다 (Chelur, D.S., and Chalfie, M. (2007). Targeted cell killing by reconstituted caspases. Proc. Natl. Acad. Sci. U. S. A. 104, 2283-2288.). 다른 가능성은 2개 캐스파제 8 (Pajvani, U.B., Trujillo, M.E., Combs, T.P., Iyengar, P., Jelicks, L., Roth, K.A., Kitsis, R.N., and Scherer, P.E. (2005). Fat apoptosis through targeted activation of caspase 8: a new mouse model of inducible and reversible lipoatrophy. Nat. Med. 11, 797-803.) 또는 캐스파제 9 (Straathof, K.C., Pul, M.A., Yotnda, P., Dotti, G., Vanin, E.F., Brenner, M.K., Heslop, H.E., Spencer, D.M., and Rooney, C.M. (2005). An inducible caspase 9 safety switch for T-cell therapy. Blood 105, 4247-4254.) 이펙터를 C2c1 결합을 통해 가깝게 오게하는, 캐스파제의 어셈블리이다. 전사물 상에서 C2c1 결합을 통해서 분할 TEV를 재구성시키는 것도 가능하다 (Gray, D.C., Mahrus, S., and Wells, J.A. (2010). activation of specific apoptotic caspases with an engineered small-molecule-activation protease. Cell 142, 637-646.). 이러한 분할 TEV는 발광 및

형광 판독을 포함한, 다양한 판독에서 사용될 수 있다 (Wehr, M.C., Laage, R., Bolz, U., Fischer, T.M., Grunewald, S., Scheek, S., Bach, A., Nave, K.-A., and Rossner, M.J. (2006). Monitoring regulated protein-protein interactions using split TEV. Nat. Methods 3, 985-993.). 일 구현예는 세포 사멸을 야기하는, 변형된 프로-캐스파제 3 또는 프로-캐스파제 7을 절단하기 위한 이러한 분할 TEV의 재구성을 포함한다 (Gray, D.C., Mahrus, S., and Wells, J.A. (2010). activation of specific apoptotic caspases with an engineered small-molecule-activation protease. Cell 142, 637-646).

[0510] 유도성 아폽토시스. 본 발명에 따라서, 가이드는 아폽토시스를 유도하기 위해 기능성 도메인을 보유하는 C2c1 복합체를 위치시키는데 사용될 수 있다. C2c1은 임의의 오솔로그일 수 있다. 한 구현예에서 기능성 도메인은 단백질의 C-말단에서 융합된다. C2c1은 예를 들어 뉴클레아제 활성을 녹아웃시키는 돌연변이를 통해 촉매적으로 불활성이다. 시스템의 적응성은 표적 핵산을 따라 가이드 공간의 최적화, 및 캐스파제 활성화의 다양한 방법을 적용하여 입증할 수 있다. 캐스파제 8 및 캐스파제 9 ("개시제" 캐스파제라고도 함) 활성은 C2c1과 회합된 캐스파제 8 또는 캐스파제 9 효소를 집합시키도록 C2c1 복합체 형성을 사용해 유도될 수 있다.

대안적으로, 캐스파제 3 및 캐스파제 7 ("이펙터" 캐스파제라고도 함) 활성은 담배 식각 바이러스 (TEV) N-말단 및 C-말단 부분 ("스니퍼")을 보유하는 C2c1 복합체가 가깝게 유지될 때 유도될 수 있어서, TEV 프로테아제 활성을 활성화시키고 캐스파제 3 또는 캐스파제 7 프로-단백질의 절단 및 활성화를 야기시킨다. 시스템은 분할 캐스파제 3을 적용할 수 있는데, 표적 핵산에 결합된 C2c1 복합체에 부착에 의해 캐스파제 3 부분이 이종이량체화된다. 예시적인 아폽토시스 성분은 하기 표 3에 기재된다.

표 3 – 아폽토시스 성분

iCasp9 (SEQ ID NO:420)	GFGDVGALESLRGNADLAYILSMEPCGHCLIINNVNF CRESGLRTRTGSNIDCEKLRRRFSSLHFMVEVKGDLT AKKMVLALLELARQDHGALDCCVVVILSHGCQASH LQFPGAVYGTDGCPVSVEKIVNIFNGTSCPSLGGKPK LFFIQACGGEQKDHGFEVASTSPEDESPGSNPEPDATP FQEGLRTFDQLDAISSLPTPSDIFVSYSTFPGFVSWRD PKSGSWYVETLDDIFEQWAHSEDLQSLLLRVANAVS VKGIYKQMPGCFNFLRKKLFFKTSVD	Straathof, K.C., et al. (2005) Blood 105, 4247–4254
캐스파제 8 (SEQ ID NO:421)	SESQTLDKVYQMKSKPRGYCLIINNHNFAKAREKVP KLHSIRDRNGTHLDAGALTTTFEELHFEIKPHDDCTV EQIYEILKIYQLMDHSNMDCFICCILSHGDKGIIYGTD GQEAPIYELTSQFTGLKCPSLAGKPKVFFIQACQGDN YQKGIPVETDSEEQPYLEMDLSSPQTRYIPDEADFLLG MATVNNCVSYRNPAEGTWYIQSLCQSLRERCPRGDD ILTILTEVNYEVSNKDDKKNMGKQMPQPTFTLRKKL VFPSD	Pajvani, U.B., et al. (2005). Nat. Med. 11, 797–803
분할 캐스파제 3 (p12) (SEQ ID NO:663)	SGVDDDMACHKIPVEADFLYAYSTAPGYYSWRNSK DGSWFIQSLCAMLKQYADKLEFMHILTRVNRKVATE FESFSFDATFHAKKQIPCIVSMLTKELYFYH	Chelur, D.S., and Chalfie, M. (2007). Proc. Natl. Acad. Sci. U. S. A. 104, 2283–2288
분할 캐스파제 3 (p17) (SEQ ID NO:422)	SGISLDNSYKMDYPEMGLCIIINNKNFHKSTGMTSRS GTDVDAANLRETFRNLKYEVRNKNDLTREEIVELMR DVSKEDHSKRSSFVCVLLSHGEEGIIFGTNGPVDLKKI TNFFRGDRCRSLTGKPKLFIIQACRGTELDCGIETD	Chelur, D.S., and Chalfie, M. (2007). Proc. Natl. Acad. Sci. U. S. A. 104, 2283–2288
스니퍼 N-TEV (SEQ ID NO:423)	GESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGP FIITNKHLFRRNNGTLLVQSLHGVFKVKNTTTLQQHL IDGRDMIIIRMPKDFPPFPQKLKFREPQREERICLVTTN FQT	Gray, D.C., et al. (2010) Cell 142, 637–646
스니퍼 C-TEV (SEQ ID NO:424)	KSMSSMVSDTSCTFPSSDGIFWKHWIQTKDGQCGSPL VSTRDGFIVGIHSASNFTNTNNYFTSVPKNFMELLTN QEAQQWVSGWRLNADSVLWGGHKVFMV	Gray, D.C., et al. (2010) Cell 142, 637–646
스니퍼 캐스파제 7 (SEQ ID NO:425)	ATGGCAGATGATCAGGGCTGTATTGAAGAGCAGG GGGTTGAGGATTCAGCAAATGAAGATTCAGTGGAA AATCTCTACTTCCAGGCTAAGCCAGACCGGTCCTC GTTTGTACCGTCCCTCTTCAGTAAGAAGAAGAAAA ATGTCACCATGCGATCCATCAAGACCACCCGGGAC CGAGTGCCTACATATCAGTACAACATGAATTTTGA AAAGCTGGGCAAATGCATCATAATAAACAACAAG AACTTTGATAAAGTGACAGGTATGGGCGTTCGAAA CGGAACAGACAAAGATGCCGAGGCGCTCTTCAAGT	Gray, D.C., et al. (2010) Cell 142, 637–646

[0511]

	GCTTCCGAAGCCTGGGTTTTGACGTGATTGTCTATA ATGACTGCTCTTGTGCCAAGATGCAAGATCTGCTT AAAAAAGCTTCTGAAGAGGACCATACAAATGCCG CCTGCTTCGCCTGCATCCTCTTAAGCCATGGAGAA GAAAATGTAATTTATGGGAAAGATGGTGTCACACC AATAAAGGATTTGACAGCCCACTTTAGGGGGGATA GATGCAAAACCCTTTTAGAGAAACCCAAACTCTTC TTCATTCAGGCTTGCCGAGGGACCGAGCTTGATGA TGGCATCCAGGCCGAAAATCTCTACTTCCAGTCGG GGCCCATCAATGACACAGATGCTAATCCTCGATAC AAGATCCCAGTGGAAGCTGACTTCCTCTTCGCCTA TTCCACGGTTCCAGGCTATTACTCGTGGAGGAGCC CAGGAAGAGGCTCCTGGTTTGTGCAAGCCCTCTGC TCCATCCTGGAGGAGCACGGAAAAGACCTGGAAA TCATGCAGATCCTCACCAGGGTGAATGACAGAGTT GCCAGGCACTTTGAGTCTCAGTCTGATGACCCACA CTTCCATGAGAAGAAGCAGATCCCCTGTGTGGTCT CCATGCTCACCAAGGAACTCTACTTCAGTCAA	
스니퍼 캐스파제 3 (SEQ ID NO:426)	ATGGAGAACACTGAAAACTCAGTGGATTCAAAATC CATTAAAAATTTGGAACCAAAGATCATACATGGAA GCGAATCAATGGAAAATCTCTACTTCCAGTCTGGA ATATCCCTGGACAACAGTTATAAAATGGATTATCC TGAGATGGGTTTATGTATAATAATTAATAATAAGA ATTTTCATAAAAGCACTGGAATGACATCTCGGTCT GGTACAGATGTCGATGCAGCAAACCTCAGGGAAA CATTCAGAAACTTGAAATATGAAGTCAGGAATAAA AATGATCTTACACGTGAAGAAATTGTGGAATTGAT GCGTGATGTTTCTAAAGAAGATCACAGCAAAAGGA GCAGTTTTGTTTGTGTGCTTCTGAGCCATGGTGAAG AAGGAATAATTTTTGGAACAAATGGACCTGTTGAC CTGAAAAAAATAACAAACTTTTTCAGAGGGGATCG TTGTAGAAGTCTAACTGGAAAACCCAAACTTTTCA TTATTCAGGCCTGCCGTGGTACAGAACTGGACTGT GGCATTGAGACAGAAAATCTCTACTTCCAGAGTGG TGTTGATGATGACATGGCGTGTCATAAAATACCAG TGGAGGCCGACTTCTTGTATGCATACTCCACAGCA CCTGGTTATTATTCTTGGCGAAATTCAAAGGATGG CTCCTGGTTCATCCAGTCGCTTTGTGCCATGCTGAA ACAGTATGCCGACAAGCTTGAATTTATGCACATTC TTACCCGGGTTAACCGAAAGGTGGCAACAGAATTT GAGTCCTTTTCCTTTGACGCTACTTTTCATGCAAAG AAACAGATTCCATGTATTGTTTCCATGCTCACAAA AGAACTCTATTTTTATCAC	Gray, D.C., et al. (2010) Cell 142, 637–646

[0512]

[0513] *분할-검출 구성체*

[0514] 본 발명의 시스템은 세포 또는 조직에 존재할 수 있는 관심 전사물 상에서 연결된 효소 부분을 갖는 CRISPR 단백질을 국재화시키기 위한 가이드를 더 포함한다. 따라서, 시스템은 제1 CRISPR 단백질에 결합하고 관심 전사물에 하이브리드화하는 제1 가이드 및 제2 CRISPR 단백질에 결합하고 관심 핵산에 하이브리드화하는 제2 가이드를 포함한다. 대부분의 구현예에서, 제1 및 제2 가이드는 인접한 위치에서 관심 핵산에 하이브리드화하는 것이 바람직하다. 위치는 직접 인접할 수 있거나 또는 소수 뉴클레오티드에 의해 이격되고, 예컨대 1 nt, 2 nt, 3 nt, 4 nt, 5 nt, 6 nt, 7 nt, 8 nt, 9 nt, 10 nt, 11 nt, 12 nt 이상 만큼 이격될 수 있다. 일정 구현예에서, 제1 및 제2 가이드는 예상 스템 루프에 의해 핵산 상에서 이격된 위치에 결합될 수 있다. 선형 핵산을 따라 이격되지만, 핵산은 가이드 표적 서열이 가깝게 근접되게 하는 2차 구조를 취할 수 있다.

[0515] 본 발명의 일 구현예에서, 단백질가수분해 효소는 캐스파제를 포함한다. 본 발명의 일 구현예에서, 단백질가수분해 효소는 개시제 캐스파제, 예컨대 제한없이 캐스파제 8 또는 캐스파제 9를 포함한다. 개시제 캐스파제는 일반적으로 단량체로서 불활성이고 동종이량체화시 활성을 획득한다. 본 발명의 일 구현예에서, 단백질가수분해 효소는 이펙터 캐스파제, 예컨대 제한없이 캐스파제 3 또는 캐스파제 7를 포함한다. 이러한 개시제 캐스파제는 단편으로 절단될 때까지 정상적으로 불활성이다. 절단되면 단편은 회합되어 활성 효소를 형성한다. 예시적인 일 구현예에서, 단백질가수분해 효소의 제1 부분은 캐스파제 3 p12를 포함하고 단백질가수분해 효소의 상보성 부분은 캐스파제 3 p17을 포함한다.

[0516] 본 발명의 일 구현예에서, 단백질가수분해 효소는 특정 아미노산 서열을 표적화하도록 선택되고 그에 따라 기질

이 선택되거나 또는 조작된다. 이러한 프로테아제의 비제한적인 예는 담배 식각 바이러스 (TEV) 프로테아제이다. 따라서, 일부 구현예에서, 절단가능하게 조작된 TEV 프로테아제에 의해 절단가능한 기질은 프로테아제에 의해 작용되는 시스템 성분으로서 제공된다. 일 구현예에서, NEV 프로테아제 기질은 프로캐스파제 및 하나 이상의 TEV 절단 부위를 포함한다. 프로캐스파제는 예를 들어 재구성된 TEV 프로테아제에 의해 절단가능하도록 조작된 캐스파제 3 또는 캐스파제 7일 수 있다. 절단되면, 프로캐스파제 단편은 활성 입체형태를 취하도록 유리된다.

[0517] 본 발명의 일 구현예에서, TEV 기질은 형광 단백질 및 TEV 절단 부위를 포함한다. 다른 구현예에서, TEV 기질은 발광 단백질 및 TEV 절단 부위를 포함한다. 일정 구현예에서, TEV 절단 부위는 기질 단백질의 형광 또는 발광 속성이 절단 시에 상실되도록 기질의 절단을 제공한다. 일정 구현예에서, 형공 또는 발광 단백질은 예를 들어 TEV 프로테아제가 재구성될 때 절단되는 형광 또는 발광을 방해하는 모이어티를 첨부하여, 변형될 수 있다.

[0518] 본 발명에 따라서, 관심 핵산을 함유하는 세포에서 단백질가수분해 활성을 제공하는 방법을 제공하고, 방법은 세포 내 핵산을, 단백질가수분해 효소의 불활성 제1 부분에 연결된 제1 CRISPR 단백질, 및 단백질가수분해 효소의 상보성 부분에 연결된 제2 CRISPR 단백질로서, 단백질 가수분해 효소의 활성은 단백질의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 단백질, 및 제1 CRISPR 단백질에 결합하고 핵산의 제1 표적 서열에 하이브리드화하는 제1 가이드, 및 제2 CRISPR 단백질에 결합하고 핵산의 제2 표적 서열에 하이브리드화하는 제2 가이드를 포함하는 조성물과 접촉시키는 단계를 포함한다. 관심 표적 핵산이 존재할 때, 단백질가수분해 효소의 제1 및 제2 부분이 접촉되고 효소의 단백질가수분해 활성이 재구성되어 효소의 기질이 절단된다.

[0519] 분할-형광단 구성체는 전사물과 2개 C2c1 단백질의 결합 시 분할 형광단의 재구성을 통해 배경이 감소된 영상화에 유용하다. 이들 분할 단백질은 iSplit (Filonov, G.S., and Verkhusha, V.V. (2013). A near-infrared BiFC reporter for in vivo imaging of protein-protein interactions. Chem. Biol. 20, 1078-1086.), Split Venus (Wu, B., Chen, J., and Singer, R.H. (2014). Background free imaging of single mRNAs in live cells using split fluorescent proteins. Sci. Rep. 4, 3615.), 및 분할 초양성 GFP (Blakeley, B.D., Chapman, A.M., and McNaughton, B.R. (2012). Split-superpositive GFP reassembly is a fast, efficient, and robust method for detecting protein-protein interactions in vivo. Mol. Biosyst. 8, 2036-2040.)를 포함한다. 이러한 단백질은 하기 표 4 에 기재된다.

표 4 – 영상화 성분		
iRFP 의 iSplit PAS 도메인 (N-말단) (SEQ ID NO:427)	MAEGSVARQPDLLTCDDEPIHIPGAIQPHG LLLALAADMTIVAGSDNLPELTGLAIGALI GRSAADVFDSETHNRLTIALAEPGAAVGA PITVGFTMRKDAGFIGSWHRHDQLIFLELE PPQRGGSEVSALEKEVSALEKEVSALEKE VSALEKEVSALEKGGS*	Filonov, G.S., and Verkhusha, V.V. (2013). Chem. Biol. 20, 1078–1086
iRFP 의 iSplit GAFm 도메인 (C-말단) (SEQ ID NO:428)	MGGSKVSALKEKVSALKEKVSALKEKVS ALKEKVSALKEGGSPPQRDVAEPQAFFRR TNSAIRRLQAAETLESACAAAAQEVRKIT GYDRVMIYRFASDFSGEVIAEDRCAEVES KLGLHYPASTVPAQARRLYTINPVRIIPDIN YRPVPVTPYLNPVTGRPIDLSFAILRSVSPV HLEFMRNIGMHGTMSISILRGERLWGLIVC HHRTPYYVDLDGRQACELVAQVLARQIG VMEE*	Filonov, G.S., and Verkhusha, V.V. (2013). Chem. Biol. 20, 1078–1086
분할 비너스 N-말단 (SEQ ID NO:429)	MVSKGEELFTGVVPILVELDGDVNGHKFS VSGEGEGDATYGKLTLKLICTTGKLPVPW PTLVTTLGYGLQCFARYPDHMKQHDFFKS AMPEGYVQERTIFFKDDGNYKTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYN YNSHNVYIT*	Wu, B., Chen, J., and Singer, R.H. (2014). Sci. Rep. 4, 3615.
분할 비너스 C-말단 (SEQ ID NO:430)	ADKQKNGIKANFKIRHNIEDGGVQLADHY QQNTPIGDGPVLLPDNHYLSYQSALSKDP NEKRDHMVLLEFVTAAGITLGMDELYK	Wu, B., Chen, J., and Singer, R.H. (2014). Sci. Rep. 4, 3615.
분할 초양성 GFP N-말단 (SEQ ID NO:431)	SKGERLFRGKVPILVELKGDVNGHKFSVR GEGKGDATRGKLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPKHMKRHDFFKSA MPKGYVQERTISFKKDGKYKTRAEVKFE GRTLVNRIKLKGRDFKEKGNILGHKLRYN FNSHKVYITADKR	Blakeley, B.D., Chapman, A.M., and McNaughton, B.R. (2012). Mol. Biosyst. 8, 2036–2040.
분할 초양성 GFP C-말단 (SEQ ID NO:432)	KNGIKAKFKIRHNVKDGSVQLADHYQQN TPIGRGPVLLPRNHYLSTRSKLSKDPKEKR DHMVLLEFVTAAGIKHGRDERYK	Blakeley, B.D., Chapman, A.M., and McNaughton, B.R. (2012). Mol. Biosyst. 8, 2036–2040.

[0520]

[0521] **dCas로 표적 농축**

[0522] 일정한 예의 구현예에서, 표적 RNA 또는 DNA는 먼저 표적 RNA 또는 DNA의 검출 또는 증폭 전에 농축될 수 있다. 일정한 예의 구현예에서, 이러한 농축은 CRISPR 이펙터 시스템에 의한 표적 핵산의 결합에 의해 획득될 수 있다.

[0523] 현재의 표적-특이적 농축 프로토콜은 프로브와의 하이브리드화 이전에 단일-가닥 핵산을 필요로 한다. 다양한 장점 중에서, 본 발명의 구현예는 이러한 단계를 생략할 수 있고 이중 가닥 (부분 또는 완전 이중 가닥) DNA를 직접 표적화할 수 있다. 또한, 본 명세서에서 개시된 구현예는 등온 농축을 허용하는 보다 빠른 운동학 및 보다 용이한 작업흐름을 제공하는 효소-구동 표적화 방법이다. 일정 예의 구현예에서, 농축은 20-37℃ 정도로 낮은 온도에서 일어날 수 있다. 일정한 예의 구현예에서, 상이한 표적 핵산에 대한 가이드 RNA의 세트가 단일 어세이에서 사용되어, 다수의 표적 및/또는 단일 표적의 다수 변이체의 검출을 가능하게 한다.

[0524] 일정한 예의 구현예에서, 데드 CRISPR 이펙터 단백질은 용액 중에서 표적 핵산에 결합할 수 있고 그 이후에 상기 용액으로부터 단리될 수 있다. 예를 들어, 표적 핵산에 결합된 데드 CRISPR 이펙터 단백질은 데드 CRISPR 이펙터 단백질에 특이적으로 결합하는, 항체 또는 다른 분자, 예컨대 압타머를 사용해 용액으로부터 단리될 수 있다.

[0525] 다른 일례의 구현예에서, 데드 CRISPR 이펙터 단백질은 고형 기재에 결합될 수 있다. 고정된 기재는 폴리펩티드 또는 폴리뉴클레오티드의 부착에 적절하거나 또는 적절하도록 변형될 수 있는 임의 재료를 의미할 수 있다. 가능한 기재는 제한없이, 유리 및 변형된 기능성 유리, 플라스틱 (아크릴, 폴리스티렌 및 스티렌과 다른 재료의 공중합체, 폴리프로필렌, 폴리에틸렌, 폴리부틸렌, 폴리우레탄, Teflon™ 등 포함), 다당류, 나일론 또는 니트로셀룰로스, 세라믹, 레진, 규소 및 변형 규소를 포함하는 실리카 또는 실리카-기반 재료, 탄소, 금속, 무기 유리, 플라스틱, 광학 섬유 번들 및 다양한 다른 중합체를 포함한다. 일부 구현예에서, 고형 지

지체는 규칙적인 패턴으로 분자의 고정에 적합한 패턴화된 표면을 포함한다. 일정 구현예에서, 패턴화된 표면은 고형 지지체의 노출된 층 내 또는 그 위의 상이한 영역의 배열을 의미한다. 일부 구현예에서, 고형 지지체는 표면의 웰 또는 오목부의 어레이를 포함한다. 고형 지지체의 조성 및 기하학적 구조는 이의 용도에 따라 다양할 수 있다. 일부 구현예에서, 고형 지지체는 평면 구조 예컨대 슬라이드, 칩, 마이크로칩 및/또는 어레이이다. 이와 같이, 기재의 표면은 평면층의 형태일 수 있다. 일부 구현예에서, 고형 지지체는 플로우셀의 하나 이상의 표면을 포함한다. 본 명세서에서 사용되는 용어 " 플로우셀" 은 하나 이상의 유체 시약이 흐를 수 있는 고형 표면을 포함하는 챔버를 의미한다. 본 개시의 방법에서 쉽게 사용할 수 있는 예시적인 플루우셀 및 관련 유체 시스템 및 검출 플랫폼은 예를 들어, [Bentley et al. Nature 456:53-59 (2008)], WO 04/0918497, U.S. 7,057,026; WO 91/06678; WO 07/123744; US 7,329,492; US 7,211,414; US 7,315,019; U.S. 7,405,281, 및 US 2008/0108082에 기술되어 있다. 일부 구현예에서, 고형 지지체 또는 이의 표면은 비평면, 예컨대 튜브 또는 용기의 내부 또는 외부 표면이다. 일부 구현예에서, 고형 지지체는 미세구 또는 비드를 포함한다. "미세구", "비드", "입자" 는 제한없이, 플라스틱, 세라믹, 유리 및 폴리스티렌을 포함하는 다양한 재료로 만들어진 소형 개별 입자를 의미하는 것으로 고형 기재의 상황 내에서 의미하고자 한다. 일정 구현예에서, 미세구는 자성 미세구 또는 비드이다. 대안적으로 또는 추가적으로, 비드는 다공성일 수 있다. 비드 크기는 나노미터, 예를 들어 100 nm에서 밀리미터, 예를 들어 1 mm의 범위이다.

[0526] 다음으로, 표적 핵산을 함유하거나, 또는 함유하는 것으로 의심되는 샘플을 기재에 노출시켜서 결합된 데드 CRISPR 이펙터 단백질에 표적 핵산이 결합할 수 있게 할 수 있다. 비-표적 분자는 세척해 버릴 수 있다. 일정한 예의 구현예에서, 이후에, 표적 핵산은 본 명세서에 개시된 방법을 사용하여 추가 검출을 위해 CRISPR 이펙터 단백질/가이드 RNA 복합체로부터 방출될 수 있다. 일정한 예의 구현예에서, 표적 핵산은 본 명세서에 기술된 바와 같이 먼저 증폭될 수 있다.

[0527] 일정한 예의 구현예에서, CRISPR 이펙터는 결합 태그로 표지될 수 있다. 일정한 예의 구현예에서, CRISPR 이펙터는 화학적으로 태그화될 수 있다. 예를 들어, CRISPR 이펙터는 화학적으로 바이오틴화될 수 있다. 다른 일례의 구현예에서 CRISPR 이펙터에 융합체를 코딩하는 추가 서열을 첨가하여 융합체를 생성시킬 수 있다. 이러한 융합체의 일례는 고유한 15개 아미노산 펩티드 태그 상에 단일 바이오틴의 고도로 표적화된 효소 접합을 적용하는 AviTag™ 이다. 일정 구현예에서, CRISPR 이펙터는 포획 태그 예컨대, 제한없이, GST, Myc, 헤마글루티닌 (HA), 녹색 형광성 단백질 (GFP), flag, His 태그, TAP 태그, 및 Fc 태그로 표지될 수 있다. 결합 태그는 융합체, 화학적 태그, 또는 포획 태그인지 여부와 무관하게, 표적 핵산에 결합되면 CRISPR 이펙터 시스템을 풀다운하거나 또는 고형 기재 상에 CRISPR 이펙터 시스템을 고정시키는데 사용될 수 있다.

[0528] 일정한 예의 구현예에서, 가이드 RNA는 결합 태그로 표지될 수 있다. 일정한 예의 구현예에서, 전체 가이드 RNA는 하나 이상의 바이오틴화된 뉴클레오티드, 예컨대, 바이오틴화된 우라실을 도입시킨 시험관내 전사 (IVT)를 사용하여 표지될 수 있다. 일부 구현예에서, 바이오틴은 화학적으로 또는 효소적으로 가이드 RNA에 첨가될 수 있고, 예컨대 가이드 RNA의 3' 말단에 하나 이상의 바이오틴 기의 첨가일 수 있다. 결합 태그는 예를 들어, 가이드 RNA/표적 핵산을 스트렙타비딘 코팅된 고형 기재에 노출시켜서, 결합이 발생된 이후 가이드 RNA/표적 핵산 복합체를 풀다운하는데 사용될 수 있다.

[0529] **절두**

[0530] 일정 예의 구현예에서, Cas12 단백질은 절두될 수 있다. 일정 예의 구현예에서, 절두형태는 탈활성화 또는 데드 Cas12 단백질일 수 있다. Cas12 단백질은 N-말단, C-말단, 또는 둘 모두 상에서 변형될 수 있다. 일례의 구현예에서, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150개 아미노산이 N-말단, C-말단, 또는 이의 조합으로부터 제거된다. 다른 예의 구현예에서, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69,

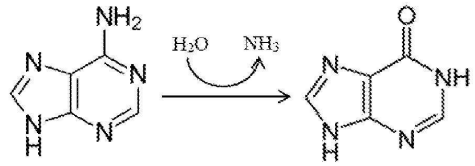
70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150개 아미노산이 C-말단으로부터 제거된다. 일정 예의 구현예에서, 1-10, 1-20, 1-30, 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, 1-100, 1-110, 1-120, 1-130, 1-140, 1-150, 1-160, 1-170, 1-180, 1-190, 1-200, 1-220, 1-230, 1-240, 1-250, 200-250, 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 10-100, 20-100, 30-100, 40-100, 50-100, 60-100, 70-100, 80-100, 90-100, 또는 150-250개 아미노산은 N-말단, C-말단, 또는 이의 조합으로부터 제거된다. 일정 예의 구현예에서, 아미노산 위치는 BhCas12의 것 또는 이에 상응하는 오솔로그의 아미노산이다. 일정 예의 구현예에서, 절두체는 뉴클레오티드 디아미나제에 융합되거나 또는 달리 부착될 수 있고 이하 더욱 상세히 개시하는 염기 편집 구현예에서 사용될 수 있다.

[0531] **염기 편집**

[0532] 일정 예의 구현예에서, Cas12b, 예를 들어 dCas12는 염기 편집 목적을 위해 아데노신 디아미나제 또는 시티딘 디아미나제와 융합될 수 있다.

[0533] ***아데노신 디아미나제***

[0534] 본 명세서에서 사용되는 용어 "아데노신 디아미나제" 또는 "아데노신 디아미나제 단백질"은 이하에 표시된 바와 같이, 아데닌 (또는 분자의 아데닌 모이어티)을 하이포잔틴 (또는 분자의 하이포잔틴 모이어티)으로 전환시키는 가수분해 탈아미노화 반응을 촉매할 수 있는 단백질, 폴리펩티드, 또는 단백질 또는 폴리펩티드의 하나 이상의 기능성 도메인(들)에 관한 것이다. 일부 구현예에서, 아데닌-함유 분자는 아데노신(A)이고, 하이포잔틴-함유 분자는 이노신(I)이다. 아데닌-함유 분자는 데옥시리보핵산(DNA) 또는 리보핵산(RNA)일 수 있다.



아데닌 하이포잔틴

[0535]

[0536] 본 공개에 따르면, 본 공개와 관련하여 사용될 수 있는 아데노신 디아미나제는 RNA (ADAR) 에 대해 작용하는 아데노신 디아미나제로서 알려진 효소 패밀리의 일원, tRNA (ADAT) 에 대해 작용하는 아데노신 디아미나제로서 알려진 효소 패밀리의 일원, 및 다른 아데노신 디아미나제 도메인-함유 (ADAD) 패밀리 일원을 포함하나 그에 제한되지 않는다. 본 공개에 따르면, 아데노신 디아미나제는 RNA/DNA 및 RNA 듀플렉스에서 아데닌을 표적화할 수 있다. 실제로, Zheng 등 (Nucleic Acids Res. 2017, 45(6): 3369- 3377)은 ADAR이 RNA/DNA 및 RNA/RNA 듀플렉스 상에서 아데노신의 이노신 편집 반응을 수행할 수 있다는 것을 입증한다. 특정 구현예에서, 아데노신 디아미나제는 본 명세서에서 이하에 상술하는 바와 같이 RNA 듀플렉스의 RNA/DNA 헤테로듀플렉스에서 DNA를 편집하는 그의 능력을 증가시키도록 변형되었다.

[0537] 일부 구현예에서, 아데노신 디아미나제는 포유동물, 조류, 개구리, 오징어, 어류, 파리 및 벌레를 포함하나 그에 제한되지 않는 하나 이상의 후생동물 종에서 유래한다. 일부 구현예에서, 아데노신 디아미나제는 인간, 오징어 또는 초파리 아데노신 디아미나제이다.

[0538] 일부 구현예에서, 아데노신 디아미나제는 hADAR1, hADAR2, hADAR3 을 포함하는 인간 ADAR 이다. 일부 구현예에서, 아데노신 디아미나제는 ADR-1 및 ADR-2 를 포함하는 예쁜꼬마선충 (Caenorhabditis elegans) ADAR 단백질이다. 일부 구현예에서, 아데노신 디아미나제는 dAdar 를 포함하는 초파리 ADAR 단백질이다. 일부 구현예에서, 아데노신 디아미나제는 sqADAR2a 및 sqADAR2b 를 포함하는 오징어 롤리고 페알레이 (Loligo pealeii) ADAR 단백질이다. 일부 구현예에서, 아데노신 디아미나제는 인간 ADAT 단백질이다. 일부 구현예에서, 아데노신 디아미나제는 초파리 ADAT 단백질이다. 일부 구현예에서, 아데노신 디아미나제는 TENR (hADAD1) 및 TENRL (hADAD2) 을 포함하는 인간 ADAD 단백질이다.

[0539] 일부 구현예에서, 아데노신 디아미나제는 TadA 단백질, 예컨대, 이.콜라이 (E. coli) TadA이다. 참조: Kim et al., Biochemistry 45:64076416 (2006); Wolf et al., EMBO J. 21:38413851(2002). 일부 구현예에서,

아데노신 디아미나제는 마우스 ADA이다. 참조: Grunebaum et al., Curr. Opin. Allergy Clin. Immunol. 13:630-638 (2013). 일부 구현예에서, 아데노신 디아미나제는 인간 ADAT2이다. 참조: Fukui et al., J. Nucleic Acids 2010:260512 (2010). 일부 구현예에서, 디아미나제 (예를 들어, 아데노신 또는 시티딘 디아미나제)는 하기 문헌들에 기술된 것들 중 하나 이상이다: Cox et al., Science. 2017, November 24; 358(6366): 1019-1027; Komore et al., Nature. 2016 May 19;533(7603):420-4; 및 Gaudelli et al., Nature. 2017 Nov 23;551(7681):464-471.

[0540] 일부 구현예에서, 아데노신 디아미나제 단백질은 이중-가닥 핵산 기질에서 하나 이상의 표적 아데노신 잔기(들)을 인지하고 이노신 잔기(들)로 전환시킨다. 일부 구현예에서, 이중-가닥 핵산 기질은 RNA-DNA 하이브리드 듀플렉스이다. 일부 구현예에서, 아데노신 디아미나제 단백질은 이중-가닥 기질 상에서 결합창을 인식한다. 일부 구현예에서, 결합창은 적어도 하나의 표적 아데노신 잔기(들)를 함유한다. 일부 구현예에서, 결합창은 약 3 bp 내지 약 100 bp 범위이다. 일부 구현예에서, 결합창은 약 5 bp 내지 약 50 bp 범위이다. 일부 구현예에서, 결합창은 약 10 bp 내지 약 30 bp 범위이다. 일부 구현예에서, 결합창은 약 1 bp, 2 bp, 3 bp, 5 bp, 7 bp, 10 bp, 15 bp, 20 bp, 25 bp, 30 bp, 40 bp, 45 bp, 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 85 bp, 90 bp, 95 bp, 또는 100 bp 이다.

[0541] 일부 구현예에서, 아데노신 디아미나제 단백질은 하나 이상의 디아미나제 도메인을 포함한다. 특정 이론에 의해 구속되는 것으로 의도하지는 않지만, 디아미나제 도메인은 이중가닥 핵산 기질에 함유된 하나 이상의 표적 아데노신(A) 잔기(들)을 인식하고, 이를 이노신(I) 잔기(들)로 전환시키는 작용을 하는 것으로 상정된다. 일부 구현예에서, 디아미나제 도메인은 활성 중심을 포함한다. 일부 구현예에서, 활성 중심은 아연 이온을 포함한다. 일부 구현예에서, A>I 편집 과정 동안, 표적 아데노신 잔기에서 염기쌍이 붕괴되며, 표적 아데노신 잔기는 아데노신 디아미나제에 의해 접근 가능하게 되도록 이중 나선 밖으로 " 플립(flipped)" 된다. 일부 구현예에서, 활성 중심 내 또는 근처의 아미노산 잔기는 표적 아데노신 잔기에 대해 하나 이상의 뉴클레오티드(들) 5'과 상호작용한다. 일부 구현예에서, 활성 중심 내 또는 근처의 아미노산 잔기는 표적 아데노신 잔기에 대해 하나 이상의 뉴클레오티드(들) 3'과 상호작용한다. 일부 구현예에서, 활성 중심에서의 또는 그 근처의 아미노산 잔기는 또한 맞은편 가닥의 표적 아데노신 잔기에 상대적인 뉴클레오티드와 상호작용한다. 일부 구현예에서, 아미노산 잔기는 뉴클레오티드의 2' 히드록실 기와 수소 결합을 형성한다.

[0542] 일부 구현예에서, 아데노신 디아미나제는 인간 ADAR2 전체 단백질 (hADAR2) 또는 그의 디아미나제 도메인 (hADAR2-D) 을 포함한다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2 또는 hADAR2-D 에 상동성인 ADAR 패밀리 일원이다.

[0543] 특히, 일부 구현예에서, 상동성 ADAR 단백질은 인간 ADAR1 (hADAR1) 또는 그의 디아미나제 도메인 (hADAR1-D) 이다. 일부 구현예에서, hADAR1-D의 글리신 1007은 글리신487 hADAR2-D에 상응하고, hADAR1-D의 글루탐산 1008은 hADAR2-D의 글루탐산488에 상응한다.

[0544] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 야생형 아미노산 서열을 포함한다. 일부 구현예에서, hADAR2-D의 편집 효율, 및/또는 기질 편집 선호도가 특정 필요에 따라 변화되도록, 아데노신 데아미나제는 hADAR2-D 서열에서 하나 이상의 돌연변이를 포함한다.

[0545] hADAR1 및 hADAR2 단백질의 소정의 돌연변이는 하기 문헌들에 기술되었으며, 그 각각은 참조로 그 전문이 본 명세서에 편입된다: Kuttan et al., Proc Natl Acad Sci U S A. (2012) 109(48):E3295-304; Want et al. ACS Chem Biol. (2015) 10(11):2512-9; and Zheng et al. Nucleic Acids Res. (2017) 45(6):3369-337.

[0546] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 글리신336 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 336번 위치의 글리신 잔기는 아스파르트산 잔기로 치환된다 (G336D).

[0547] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 글리신487, 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 487번 위치에서 글리신 잔기는 비교적 소형 측쇄를 갖는 비극성 아미노산 잔기로 치환된다. 예를 들어, 일부 구현예에서, 487번 위치에서 글리신 잔기는 알라닌 잔기로 치환된다 (G487A). 일부 구현예에서, 487번 위치에서 글리신 잔기는 발린 잔기로 치환된다 (G487V). 일부 구현예에서, 487번 위치에서 글리신 잔기는 비교적 큰 측쇄를 갖는 아미노산 잔기로 치환된다. 일부 구현예에서, 487번 위치에서 글리신 잔기는 아르기닌 잔기로 치환된다 (G487R). 일부 구현예에서, 487번 위치에서 글리신 잔기는 리신 잔기로 치환된다 (G487K). 일부 구현예에서, 487번 위치에서 글

리신 잔기는 트립토판 잔기로 치환된다 (G487W). 일부 구현예에서, 487번 위치에서 글리신 잔기는 티로신 잔기로 치환된다 (G487Y).

[0548] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 글루탐산488, 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 글루타민 잔기로 치환된다(E488Q). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 히스티딘 잔기로 치환된다(E488H). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 알기닌 잔기로 치환된다(E488R). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 라이신 잔기로 치환된다(E488K). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 아스파라긴 잔기로 치환된다(E488N). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 알라닌 잔기로 치환된다(E488A). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 메티오닌 잔기로 치환된다(E488M). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 세린 잔기로 치환된다(E488S). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 페닐알라닌 잔기로 치환된다(E488F). 일부 구현예에서, 488번 위치에서 글루탐산 잔기는 라이신 잔기로 치환된다(E488L). 일부 구현예에서, 위치 488 에서의 글루탐산 잔기는 트립토판 잔기로 치환된다 (E488W).

[0549] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 트레오닌490, 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 시스테인 잔기로 치환된다 (T490C). 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 세린 잔기로 치환된다(T490S). 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 알라닌 잔기로 치환된다(T490A). 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 페닐알라닌 잔기로 치환된다(T490F). 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 타이로신 잔기로 치환된다(T490Y). 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 세린 잔기로 치환된다(T490R). 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 알라닌 잔기로 치환된다 (T490K). 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 페닐알라닌 잔기로 치환된다(T490P). 일부 구현예에서, 490번 위치에서 트레오닌 잔기는 타이로신 잔기로 치환된다(T490E).

[0550] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 발린493, 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 493번 위치에서 발린 잔기는 알라닌 잔기로 치환된다 (V493A). 일부 구현예에서, 493번 위치에서 발린 잔기는 세린 잔기로 치환된다 (V493S). 일부 구현예에서, 위치 493 에서의 발린 잔기는 트레오닌 잔기로 치환된다 (V493T). 일부 구현예에서, 위치 493 에서의 발린 잔기는 아르기닌 잔기로 치환된다 (V493R). 일부 구현예에서, 위치 493 에서의 발린 잔기는 아스파르트산 잔기로 치환된다 (V493D). 일부 구현예에서, 위치 493 에서의 발린 잔기는 프롤린 잔기로 치환된다 (V493P). 일부 구현예에서, 위치 493 에서의 발린 잔기는 글라이신 잔기로 치환된다 (V493G).

[0551] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 알라닌589, 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 589번 위치에서 알라닌 잔기는 발린 잔기로 치환된다 (A589V).

[0552] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 아스파라긴597, 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 597번 위치에서 아스파라긴 잔기는 리신 잔기로 치환된다 (N597K). 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 597번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 597번 위치에서 아스파라긴 잔기는 아르기닌 잔기로 치환된다 (N597R). 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 597번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 597번 위치에서 아스파라긴 잔기는 알라닌 잔기로 치환된다 (N597A). 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 597번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 597번 위치에서 아스파라긴 잔기는 글루탐산 잔기로 치환된다 (N597E). 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 597번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 597번 위치에서 아스파라긴 잔기는 히스티딘 잔기로 치환된다 (N597H). 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 597번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 위치 597 에서의 아스파라긴 잔기는 글라이신 잔기로 치환된다 (N597G). 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 597번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 위치 597 에서의 아스파라긴 잔기는 티로신 잔기로 치환된다 (N597Y). 일부 구현예에서, 위치 597 에서의 아스파라긴 잔기는 페닐알라닌 잔기로 치환된다 (N597F). 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597I를 포함한다. 일부 구현예에서, 아데노신 디아미나

제는 돌연변이 N597L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597V를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597M를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597C를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597P를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597T를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597W를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N597D를 포함한다. 일정 예시적 구현예에서, 상기 기재한 N597에서의 돌연변이는 E488Q 배경 상황에서 추가로 만들어 진다.

[0553] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 세린599, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 599번 위치에서 세린 잔기는 트레오닌 잔기로 치환된다 (S599T).

[0554] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 아스파라긴613, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 613번 위치에서 아스파라긴 잔기는 리신 잔기로 치환된다 (N613K). 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 613번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 613번 위치에서 아스파라긴 잔기는 아르기닌 잔기로 치환된다 (N613R). 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 613번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 613번 위치에서 아스파라긴 잔기는 알라닌 잔기로 치환된다(N613A) 일부 구현예에서, 아데노신 디아미나제는 야생형 서열에서 아스파라긴 잔기를 갖는 아미노산 서열의 613번 위치에서 돌연변이를 포함한다. 일부 구현예에서, 613번 위치에서 아스파라긴 잔기는 글루탐산 잔기로 치환된다 (N613E). 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613I를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613V를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613F를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613M를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613C를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613P를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613T를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613Y를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613W를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613H를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N613D를 포함한다. 일부 구현예에서, 상기 기재한 N613에서의 돌연변이는 E488Q 돌연변이와 조합하여 추가로 생성된다.

[0555] 일부 실시형태에서, 편집 효율을 개선하기 위해서, 아데노신 데아미나아제는 하기 돌연변이 중 하나 이상을 포함할 수 있다: G336D, G487A, G487V, E488Q, E488H, E488R, E488N, E488A, E488S, E488M, T490C, T490S, V493T, V493S, V493A, V493R, V493D, V493P, V493G, N597K, N597R, N597A, N597E, N597H, N597G, N597Y, A589V, S599T, N613K, N613R, N613A, N613E (hADAR2-D 의 아미노산 서열 위치 기반), 및 상기에 상응하는 상동성 ADAR 단백질에서의 돌연변이.

[0556] 일부 실시형태에서, 편집 효율을 감소시키기 위해서, 아데노신 데아미나아제는 하기 돌연변이 중 하나 이상을 포함할 수 있다: E488F, E488L, E488W, T490A, T490F, T490Y, T490R, T490K, T490P, T490E, N597F (hADAR2-D 의 아미노산 서열 위치 기반), 및 상기에 상응하는 상동성 ADAR 단백질에서의 돌연변이. 특정한 구현예에서, 오프-표적 효과를 감소시키기 위해서 효율이 감소된 아데노신 디아미나제 효소를 사용하는 것이 흥미로울 수 있다.

[0557] 일부 실시형태에서, 오프-표적 효과를 감소시키기 위해서, 아데노신 데아미나아제는 하기 돌연변이 중 하나 이상을 포함할 수 있다: R348, V351, T375, K376, E396, C451, R455, N473, R474, K475, R477, R481, S486, E488, T490, S495, R510 (hADAR2-D 의 아미노산 서열 위치에 기초함) 에서의 돌연변이, 및 상기에 상응하는 상동성 ADAR 단백질에서의 돌연변이. 일부 구현예에서, 아데노신 디아미나제는 E488 에서 및 R348, V351, T375, K376, E396, C451, R455, N473, R474, K475, R477, R481, S486, T490, S495, R510 로부터 선택되는 하나 이상의 부가적 위치에서 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 T375, 및 임의로는 하나 이상의 추가 위치에 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 N473, 및 임의로는 하나 이상의 추가 위치에 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 V351, 및 임의로

는 하나 이상의 추가 위치에 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 E488 및 T375, 및 임의로는 하나 이상의 추가 위치에 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 E488 및 N473, 및 임의로는 하나 이상의 추가 위치에 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 E488 및 V351 에서, 및 임의로 하나 이상의 부가적 위치에서 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 E488 에서 및 T375, N473, 및 V351 중 하나 이상에서 돌연변이를 포함한다.

[0558] 일부 구현예에서, 오프-표적 효과를 감소시키기 위해서, 아데노신 데아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로, R348E, V351L, T375G, T375S, R455G, R455S, R455E, N473D, R474E, K475Q, R477E, R481E, S486T, E488Q, T490A, T490S, S495T, 및 R510E로부터 선택되는 돌연변이, 및 상기에 상응하는 상동성 ADAR 단백질 내 돌연변이 중 하나 이상을 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 E488Q, 및 R348E, V351L, T375G, T375S, R455G, R455S, R455E, N473D, R474E, K475Q, R477E, R481E, S486T, T490A, T490S, S495T 및 R510E로부터 선택되는 하나 이상의 추가 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375G 또는 T375S, 및 임의로 하나 이상의 추가 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N473D, 및 임의로는 하나 이상의 추가 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351L, 및 임의로는 하나 이상의 추가 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 E488Q, 및 T375G 또는 T375G, 및 임의로 하나 이상의 추가 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 E488Q 및 N473D, 및 임의로 하나 이상의 추가 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 E488Q 및 V351L, 및 임의로 하나 이상의 추가 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 E488Q 및 T375G/S, N473D 및 V351L 중 하나 이상을 포함한다.

[0559] 일정 예에서, 아데노신 디아미나제 단백질 또는 이의 촉매 도메인은 hADAR2-D 아미노산의 E488에 돌연변이, 바람직하게 E488Q, 또는 상동성 ADAR 단백질의 상응하는 위치에 돌연변이를 포함하도록 변형되었고/되었거나, 아데노신 디아미나제 단백질 또는 이의 촉매 도메인은 hDAR2-D 아미노산 서열의 T375에 돌연변이, 바람직하게 T375G, 또는 상동성 ADAR 단백질의 상응하는 위치에 돌연변이를 포함한다. 일정 예에서, 아데노신 디아미나제 단백질 또는 이의 촉매 도메인은 hADAR1d 아미노산 서열의 E1008에 돌연변이, 바람직하게 E1008Q, 또는 상동성 ADAR 단백질의 상응하는 위치에 돌연변이를 포함하도록 변형되었다.

[0560] 듀플렉스 RNA에 결합된 인간 ADAR2 디아미나제 도메인의 결정 구조는 변형 부위의 5' 측면 상에서 RNA에 결합하는 단백질 루프를 밝혀준다. 이러한 5' 결합 루프는 ADAR 패밀리 멤버 간 기질 특이성 편차에 대한 하나의 원인 인자이다. [Wang et al., Nucleic Acids Res., 44(20):98729880 (2016)]을 참조하고, 이의 내용은 그 전문이 참조로 본 명세서에 편입된다. 또한, ADAR2-특이적RNA-결합 루프가 효소 활성 부위 근처에서 확인되었다. [Mathews et al., Nat. Struct. Mol. Biol., 23(5):426- 33 (2016)]을 참조하고, 이의 내용은 그 전문이 참조로 본 명세서에 편입된다. 일부 구현예에서, 아데노신 디아미나제는 편집 특이성 및/또는 효율을 개선시키도록 RNA 결합 루프 내에 하나 이상의 돌연변이를 포함한다.

[0561] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 알라닌454, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 454의 알라닌은 세린 잔기로 치환된다 (A454S). 일부 구현예에서, 위치 454의 알라닌 잔기는 시스테인 잔기로 치환된다 (A454C). 일부 구현예에서, 위치 454의 알라닌 잔기는 아스파르트산 잔기로 치환된다 (A454D).

[0562] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 아르기닌455, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 455의 아르기닌 잔기는 알라닌 잔기로 치환된다 (R455A). 일부 구현예에서, 위치 455의 아르기닌 잔기는 발린 잔기로 치환된다 (R455V). 일부 구현예에서, 위치 455의 아르기닌 잔기는 히스티딘 잔기로 치환된다 (R455H). 일부 구현예에서, 위치 455의 아르기닌 잔기는 글리신 잔기로 치환된다 (R455G). 일부 구현예에서, 위치 455의 아르기닌 잔기는 세린 잔기로 치환된다 (R455S). 일부 구현예에서, 위치 455의 아르기닌 잔기는 글루탐산 잔기로 치환된다 (R455E). 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455C를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455I를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455K를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455M를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455N를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455F를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455W를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455P를 포함한다. 일부 구현예

에서, 아데노신 디아미나제는 돌연변이 R455Y를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R455D를 포함한다. 일부 구현예에서, 상기 기재한 R455에서의 돌연변이는 E488Q 돌연변이와 조합하여 추가로 생성된다.

[0563] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 이소류신456, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 456의 이소류신 잔기는 발린 잔기로 치환된다 (I456V). 일부 구현예에서, 위치 456의 이소류신 잔기는 류신 잔기로 치환된다 (I456L). 일부 구현예에서, 위치 456의 이소류신 잔기는 아스파르트산 잔기로 치환된다 (I456D).

[0564] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 페닐알라닌457, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 457번 위치에서 페닐알라닌 잔기는 타이로신 잔기로 치환된다(F457Y). 일부 구현예에서, 457번 위치에서 페닐알라닌 잔기는 알기닌 잔기로 치환된다 (F457R). 일부 구현예에서, 457번 위치에서 페닐알라닌 잔기는 글루탐산 잔기로 치환된다(F457E).

[0565] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 세린458, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 458의 세린 잔기는 발린 잔기로 치환된다 (S458V). 일부 구현예에서, 위치 458의 세린 잔기는 페닐알라닌 잔기로 치환된다 (S458F). 일부 구현예에서, 위치 458의 세린 잔기는 프롤린 잔기로 치환된다 (S458P). 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458I를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458M를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458C를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458A를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458T를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458Y를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458W를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458N를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458H를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458K를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458R를 포함한다. 일부 구현예에서, 상기 기재한 S458에서의 돌연변이는 E488Q 돌연변이와 조합하여 추가로 생성된다.

[0566] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 프롤린459, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 459의 프롤린 잔기는 시스테인 잔기로 치환된다 (P459C). 일부 구현예에서, 위치 459의 프롤린 잔기는 히스티딘 잔기로 치환된다 (P459H). 일부 구현예에서, 위치 459의 프롤린 잔기는 트립토판 잔기로 치환된다 (P459W).

[0567] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 히스티딘460, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 460의 히스티딘 잔기는 아르기닌 잔기로 치환된다 (H460R). 일부 구현예에서, 위치 460의 히스티딘 잔기는 이소류신 잔기로 치환된다 (H460I). 일부 구현예에서, 위치 460의 히스티딘 잔기는 프롤린 잔기로 치환된다 (H460P). 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460V를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460F를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460M를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460C를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460A를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460T를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460Y를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460W를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460N를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 H460K를 포함한다. 일부 구현예에서, 상기 기재한 H460에서의 돌연변이는 E488Q 돌연변이와 조합하여 추가로 생성된다.

[0568] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 프롤린462, 또는 상동성 ADAR 단백질 내 상

응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 462의 프롤린 잔기는 세린 잔기로 치환된다 (P462S). 일부 구현예에서, 위치 462의 프롤린 잔기는 트립토판 잔기로 치환된다 (P462W). 일부 구현예에서, 위치 462의 프롤린 잔기는 글루탐산 잔기로 치환된다 (P462E).

[0569] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 아스파르트산469, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 469의 아스파르트산 잔기는 글루타민 잔기로 치환된다 (D469Q). 일부 구현예에서, 위치 469의 아스파르트산 잔기는 세린 잔기로 치환된다 (D469S). 일부 구현예에서, 위치 469의 아스파르트산 잔기는 티로신 잔기로 치환된다 (D469Y).

[0570] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 아르기닌470, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 470의 아르기닌 잔기는 알라닌 잔기로 치환된다 (R470A). 일부 구현예에서, 470번 위치에서 알기닌 잔기는 아이소류신 잔기로 치환된다 (R470I). 일부 구현예에서, 470번 위치에서 알기닌 잔기는 아스파르트산 잔기로 치환된다 (R470D).

[0571] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 히스티딘471, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 471의 히스티딘 잔기는 리신 잔기로 치환된다 (H471K). 일부 구현예에서, 위치 471의 히스티딘 잔기는 트레오닌 잔기로 치환된다 (H471T). 일부 구현예에서, 위치 471의 히스티딘 잔기는 발린 잔기로 치환된다 (H471V).

[0572] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 프롤린472, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 472의 프롤린 잔기는 리신 잔기로 치환된다 (P472K). 일부 구현예에서, 위치 472의 프롤린 잔기는 트레오닌 잔기로 치환된다 (P472T). 일부 구현예에서, 위치 472의 프롤린 잔기는 아스파르트산 잔기로 치환된다 (P472D).

[0573] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 아스파라긴473, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 473의 아스파라긴 잔기는 아르기닌 잔기로 치환된다 (N473R). 일부 구현예에서, 위치 473의 아스파라긴 잔기는 트립토판 잔기로 치환된다 (N473W). 일부 구현예에서, 위치 473의 아스파라긴 잔기는 프롤린 잔기로 치환된다 (N473P). 일부 구현예에서, 위치 473의 아스파라긴 잔기는 아스파르트산 잔기로 치환된다 (N473D).

[0574] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 아르기닌 474, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 474의 아르기닌 잔기는 리신 잔기로 치환된다 (R474K). 일부 구현예에서, 위치 474의 아르기닌 잔기는 글리신 잔기로 치환된다 (R474G). 일부 구현예에서, 위치 474의 알기닌 잔기는 아스파르트산 잔기로 치환된다(R474D). 일부 구현예에서, 위치 474의 아르기닌 잔기는 글루탐산 잔기로 치환된다 (R474E).

[0575] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 리신475, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 475의 리신 잔기는 글루타민 잔기로 치환된다 (K475Q). 일부 구현예에서, 위치 475의 리신 잔기는 아스파라긴 잔기로 치환된다 (K475N). 일부 구현예에서, 위치 475의 리신 잔기는 아스파르트산 잔기로 치환된다 (K475D).

[0576] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 알라닌476, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 476의 알라닌은 세린 잔기로 치환된다 (A476S). 일부 구현예에서, 위치 476의 알라닌 잔기는 아르기닌 잔기로 치환된다 (A476R). 일부 구현예에서, 위치 476의 알라닌 잔기는 글루탐산 잔기로 치환된다 (A476E).

[0577] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 아르기닌477, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 477의 아르기닌 잔기는 리신 잔기로 치환된다 (R477K). 일부 구현예에서, 위치 477의 아르기닌 잔기는 트레오닌 잔기로 치환된다 (R477T). 일부 구현예에서, 위치 477의 아르기닌 잔기는 페닐알라닌 잔기로 치환된다 (R477F). 일부 구현예에서, 위치 474의 아르기닌 잔기는 글루탐산 잔기로 치환된다 (R477E).

[0578] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 글리신478, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 478의 글리신 잔기는 알라닌 잔기로 치환된다 (G478A). 일부 구현예에서, 위치 478의 글리신 잔기는 아르기닌 잔기로 치환된다 (G478R). 일부 구현예에서, 478번 위치에서 글리신 잔기는 티로신 잔기로 치환된다 (G478Y). 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478I를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478L를 포함한다.

일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478V를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478F를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478M를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478C를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478P를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478T를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478W를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478N를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478H를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478K를 포함한다. 일부 구현예에서, 상기 기재한 G478에서의 돌연변이는 E488Q 돌연변이와 조합하여 추가로 생성된다.

[0579] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 글루타민479에서, 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 위치 479의 글루타민 잔기는 아스파라긴 잔기로 치환된다 (Q479N). 일부 구현예에서, 위치 479의 글루타민 잔기는 세린 잔기로 치환된다 (Q479S). 일부 구현예에서, 위치 479번의 글루타민 잔기는 프롤린 잔기로 치환된다(Q479P).

[0580] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열에서 아르기닌348, 또는 상동성 ADAR 단백질 내 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 위치 348의 아르기닌 잔기는 알라닌 잔기로 치환된다 (R348A). 일부 구현예에서, 위치 348의 아르기닌 잔기는 글루탐산 잔기로 치환된다 (R348E).

[0581] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 발린351, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 351의 발린 잔기는 류신 잔기로 치환된다 (V351L). 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351Y를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351M를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351T를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351A를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351F를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351I를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351C를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351H를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351P를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351K를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351N를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351W를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 V351R를 포함한다. 일부 구현예에서, 상기 기재한 V351에서의 돌연변이는 E488Q 돌연변이와 조합하여 추가로 생성된다.

[0582] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 트레오닌375에서, 또는 상동성 ADAR 단백질 내 대응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 위치 375의 트레오닌 잔기는 글리신 잔기로 치환된다 (T375G). 일부 구현예에서, 위치 375의 트레오닌 잔기는 세린 잔기로 치환된다 (T375S). 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375H를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375C를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375N를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375M를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375A를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375W를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375V를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375R를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375K를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375F를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375I를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375P를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375Y를 포함한다. 일부 구현예에서, 상기 기재된 T375Y에서의 돌연변이는

추가로 E488Q 돌연변이와 조합하여 생성된다.

[0583] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 Arg481, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 481의 아르기닌 잔기는 글루탐산 잔기로 치환된다 (R481E).

[0584] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 Ser486, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 486의 세린 잔기는 트레오닌 잔기로 치환된다 (S486T).

[0585] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 Thr490, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 490의 트레오닌 잔기는 알라닌 잔기로 치환된다 (T490A). 일부 구현예에서, 위치 490의 트레오닌 잔기는 세린 잔기로 치환된다 (T490S).

[0586] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 Ser495, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 495의 세린 잔기는 트레오닌 잔기로 치환된다 (S495T).

[0587] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 Arg510, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 510의 아르기닌 잔기는 글루타민 잔기로 치환된다 (R510Q). 일부 구현예에서, 위치 510의 아르기닌 잔기는 알라닌 잔기로 치환된다 (R510A). 일부 구현예에서, 위치 510의 아르기닌 잔기는 글루탐산 잔기로 치환된다 (R510E).

[0588] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 Gly593, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 593의 글리신 잔기는 알라닌 잔기로 치환된다 (G593A). 일부 구현예에서, 위치 593의 글리신 잔기는 글루탐산 잔기로 치환된다 (G593E).

[0589] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 Lys594, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 594의 리신 잔기는 알라닌 잔기로 치환된다 (K594A).

[0590] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 위치 A454, R455, I456, F457, S458, P459, H460, P462, D469, R470, H471, P472, N473, R474, K475, A476, R477, G478, Q479, R348, R510, G593, K594, 또는 상동성 ADAR 단백질 내 상응하는 위치 중 임의의 하나 이상에 돌연변이를 포함한다.

[0591] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 돌연변이 A454S, A454C, A454D, R455A, R455V, R455H, I456V, I456L, I456D, F457Y, F457R, F457E, S458V, S458F, S458P, P459C, P459H, P459W, H460R, H460I, H460P, P462S, P462W, P462E, D469Q, D469S, D469Y, R470A, R470I, R470D, H471K, H471T, H471V, P472K, P472T, P472D, N473R, N473W, N473P, R474K, R474G, R474D, K475Q, K475N, K475D, A476S, A476R, A476E, R477K, R477T, R477F, G478A, G478R, G478Y, Q479N, Q479S, Q479P, R348A, R510Q, R510A, G593A, G593E, K594A, 또는 상동성 ADAR 단백질 내 상응하는 위치 내 돌연변이 중 임의의 하나 이상을 포함한다.

[0592] 일정 구현예에서, 아데노신 디아미나제는 활성이 시티딘 디아미나제로 전환되도록 돌연변이된다. 따라서, 일부 구현예에서, 아데노신 디아미나제는 E396, C451, V351, R455, T375, K376? S486, Q488, R510, K594, R348, G593, S397, H443, L444, Y445, F442, E438, T448, A353, V355, T339, P539, T339, P539, V525 I520, P462 및 N579로부터 선택되는 위치에 하나 이상의 돌연변이를 포함한다. 특정한 구현예에서, 아데노신 디아미나제는 V351, L444, V355, V525 및 I520으로부터 선택되는 위치에 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로, E488, V351, S486, T375, S370, P462, N597에서의 돌연변이, 및 상기에 상응하는 상동성 ADAR 단백질 내 돌연변이 중 하나 이상을 포함할 수 있다.

[0593] 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로, 돌연변이: E488Q, V351G, 및 상기에 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, 및 상기 상응하는 상동성 ADAR 단백질의

돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치를 기반으로 돌연변이: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T, 및 상기 상응하는 상동성 ADAR 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 예에서, 본 명세서는 돌연변이된 디아미나제, 예를 들어, 데드 Cas12b 단백질 또는 Cas12 닉카제와 융합된, E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T의 하나 이상의 돌연변이를 포함하는 아데노신 디아미나제를 포함하여, 제공한다.

특정 예에서, 본 발명은 돌연변이된 아데노신 디아미나제, 예를 들어 데드 Cas12b 단백질 또는 Cas12 닉카제와 융합된, E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, 및 S661T를 포함하는 아데노신 디아미나제를 포함하여, 제공한다.

[0594] 일부 구현예에서, 아데노신 디아미나제는 임의로 E488에서의 돌연변이와 조합하여, hADAR2-D 아미노산 서열의 위치 T375, V351, G478, S458, H460, 또는 상동성 ADAR 단백질 내 상응하는 위치 중 어느 하나 이상에서 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 임의로 E488Q와 조합하여, T375G, T375C, T375H, T375Q, V351M, V351T, V351Y, G478R, S458F, H460I로부터 선택된 돌연변이 중 하나 이상을 포함한다.

[0595] 일부 구현예에서, 아데노신 디아미나제는 임의로 E488Q와 조합하여, T375H, T375Q, V351M, V351Y, H460P로부터 선택된 돌연변이 중 하나 이상을 포함한다.

[0596] 일부 구현예에서, 아데노신 디아미나제는 임의로 E488Q와 조합하여, 돌연변이 T375S 및 S458F를 포함한다.

[0597] 일부 구현예에서, 아데노신 디아미나제는 임의로 E488에서의 돌연변이와 조합하여, hADAR2-D 아미노산 서열의

위치 T375, N473, R474, G478, S458, P459, V351, R455, R455, T490, R348, Q479, 또는 상동성 ADAR 단백질의 상응하는 위치 중 둘 이상에서 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 임의로 E488Q와 조합하여, T375G, T375S, N473D, R474E, G478R, S458F, P459W, V351L, R455G, R455S, T490A, R348E, Q479P로부터 선택된 돌연변이 중 둘 이상을 포함한다.

[0598] 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375G 및 V351L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375G 및 R455G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375G 및 R455S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375G 및 T490A를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375G 및 R348E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375S 및 V351L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375S 및 R455G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375S 및 R455S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375S 및 T490A를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 T375S 및 R348E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N473D 및 V351L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N473D 및 R455G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N473D 및 R455S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N473D 및 T490A를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 N473D 및 R348E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R474E 및 V351L를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R474E 및 R455G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R474E 및 R455S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R474E 및 T490A를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 R474E 및 R348E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458F 및 T375G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458F 및 T375S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458F 및 N473D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458F 및 R474E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 S458F 및 G478R를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478R 및 T375G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478R 및 T375S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478R 및 N473D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 G478R 및 R474E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 P459W 및 T375G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 P459W 및 T375S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 P459W 및 N473D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 P459W 및 R474E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 P459W 및 G478R를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 P459W 및 S458F를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 Q479P 및 T375G를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 Q479P 및 T375S를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 Q479P 및 N473D를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 Q479P 및 R474E를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 Q479P 및 G478R를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 Q479P 및 S458F를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 Q479P 및 P459W를 포함한다. 본 단락에 기재된 모든 돌연변이는 또한 추가로 E488Q 돌연변이와 조합하여 생성될 수 있다.

[0599] 일부 구현예에서, 아데노신 디아미나제는 임의로 E488에서의 돌연변이와 조합하여, hADAR2-D 아미노산 서열의 위치 K475, Q479, P459, G478, S458, 또는 상동성 ADAR 단백질 내 대응하는 위치 중 어느 하나 이상에서 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 임의로 E488Q와 조합하여, K475N, Q479N, P459W, G478R, S458P, S458F로부터 선택된 돌연변이 중 하나 이상을 포함한다.

[0600] 일부 구현예에서, 아데노신 디아미나제는 임의로 E488에서의 돌연변이와 조합하여, hADAR2-D 아미노산 서열의 위치 T375, V351, R455, H460, A476, 또는 상동성 ADAR 단백질 내 상응하는 위치 중 어느 하나 이상에서 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 임의로 E488Q와 조합하여, T375G, T375C, T375H, T375Q, V351M, V351T, V351Y, R455H, H460P, H460I, A476E로부터 선택된 돌연변이 중 하나 이상을 포함한다.

[0601] 일정 구현예에서, 편집의 개선 및 오프-표적 변형의 감소는 gRNA의 화학적 변형에 의해 달성된다. 문헌 [Vogel et al. (2014), Angew Chem Int Ed, 53:6267- 6271, doi:10.1002/anie.201402634 (본 명세서에 전문이

참고로 편입됨)]에 예시된 바와 같이 화학적으로 변형된 gRNA는 오프-표적 활성을 감소시키고, 온-표적 효율을 개선시킨다. 2'-O-메틸 및 포스포티오에이트 변형된 가이드 RNA 는 일반적으로 세포에서 편집 효율을 개선한다.

[0602] ADAR 은 편집된 A 의 양측에서 이웃 뉴클레오티드에 대한 선호도를 입증하는 것으로 알려졌다 (www.nature.com/nsmb/journal/v23/n5/full/nsmb.3203.html, Matthews et al. (2017), Nature Structural Mol Biol, 23(5): 426433, 그 전문이 본원에 참조로 포함된다). 따라서, 일정 구현예에서, gRNA, 표적, 및/또는 ADAR 은 모티프 선호도에 대해 최적화되도록 선택된다.

[0603] 의도적 미스매치는 시험관내에서 비-선호 모티프의 편집을 허용한다는 것이 입증되었다 (참조: 참조로 전문이 본 명세서에 편입되는, academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku272; Schneider et al (2014), Nucleic Acid Res, 42(10):e87); Fukuda et al. (2017), Scientific Reports, 7, doi:10.1038/srep41478). 따라서, 특정 구현예에서, 바람직하지 않은 5' 또는 3' 이웃 염기에 대한 RNA 편집 효율을 향상시키기 위해서, 이웃 염기에 의도적 미스매치가 도입된다.

[0604] 일부 구현예에서, 아데노신 디아미나제는 tRNA-특이적 아데노신 디아미나제 또는 이의 변이체일 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: W23L, W23R, R26G, H36L, N37S, P48S, P48T, P48A, I49V, R51L, N72D, L84F, S97C, A106V, D108N, H123Y, G125A, A142N, S146C, D147Y, R152H, R152P, E155V, I156F, K157N, K161T, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: D108N, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, A142N, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, A142N, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, A142N, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, R152P, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 일부 구현예에서, 아데노신 디아미나제는 이. 콜라이 TadA의 아미노산 서열 위치를 기반으로 돌연변이: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, R152P, A142N, 및 상기에 상응하는 상동성 디아미나제 단백질의 돌연변이 중 하나 이상을 포함할 수 있다.

[0605] 결과는 ADAR 디아미나제 도메인의 표적화 창에서 A의 대항 C가 다른 염기보다 우선적으로 편집된다는 것을 시사

한다. 추가로, 표적화된 염기의 소수 염기 내 U와 염기-쌍형성된 A는 Cas12b-ADAR 융합에 의한 낮은 수준의 편집을 보여서, 효소가 다수 A를 편집하는데 유연성이 있음을 시사한다. 이들 2가지 관찰은 Cas12b-ADAR 융합체의 활성창에서 다수 A는 편집하려는 모든 A를 C와 미스매치시켜서 편집을 위해 지정할 수 있다는 것을 시사한다. 따라서, 일정 구현예에서, 활성창의 다수 A:C 미스매치는 다수 A:I 편집을 생성하도록 설계된다. 일정 구현예에서, 활성 창에서 잠재적 비표적 편집을 억제하기 위해, 비-표적 A는 A 또는 G와 짝지어진다.

[0606] 용어 " 편집 특이성" 및 " 편집 선호도" 는 이중가닥 기질 내 특정 아데노신 부위에서 A-대-I 편집 정도를 지칭하기 위해 본 명세서에서 상호 호환 가능하게 사용된다. 일부 구현예에서, 기질 편집 선호도는 표적 아데노신 잔기의 5' 가장 가까운 이웃 및/또는 3' 가장 가까운 이웃에 의해 결정된다. 일부 구현예에서, 아데노신 디아미나제는 U>A>C>G(" >" 는 더 큰 선호도를 나타냄)으로 순위매겨지는, 기질의 5' 가장 가까운 이웃에 대한 선호도를 갖는다. 일부 구현예에서, 아데노신 디아미나제는 G>C~A>U(" >" 는 더 큰 선호도를 나타냄; " ~" 는 유사한 선호도를 나타냄)로서 순위매겨지는, 기질의 3' 가장 가까운 이웃에 대해 선호도를 갖는다. 일부 구현예에서, 아데노신 디아미나제는 G>C>U~A (" >" 는 더 큰 선호도를 나타냄; " ~" 는 유사한 선호도를 나타냄)로서 순위매겨지는, 기질의 3' 가장 가까운 이웃에 대해 선호도를 갖는다. 일부 구현예에서, 아데노신 디아미나제는 G>C>A>U (" >" 는 더 큰 선호도를 나타냄)로서 순위매겨지는, 기질의 3' 가장 가까운 이웃에 대해 선호도를 갖는다. 일부 구현예에서, 아데노신 디아미나제는 C~G~A>U (" >" 는 더 큰 선호도를 나타냄; " ~" 는 유사한 선호도를 나타냄)로서 순위매겨지는, 기질의 3' 가장 가까운 이웃에 대해 선호도를 갖는다. 일부 구현예에서, 아데노신 디아미나제는 TAG>AAG>CAC>AAT>GAA>GAC (" >" 는 더 큰 선호도를 나타냄)(중심 A는 표적 아데노신 잔기임)로서 순위매겨지는, 표적 아데노신 잔기를 함유하는 삼중 서열에 대한 선호도를 갖는다.

[0607] 일부 구현예에서, 아데노신 디아미나제의 기질 편집 선호도는 아데노신 디아미나제 단백질에서의 핵산 결합 도메인의 존재 또는 부재에 의해 영향을 받는다. 일부 구현예에서, 기질 편집 선호도를 조정하기 위해서, 디아미나제 도메인은 이중-가닥 RNA 결합 도메인 (dsRBD) 또는 이중-가닥 RNA 결합 모티프 (dsRBM) 와 연결된다. 일부 구현예에서, dsRBD 또는 dsRBM 은 ADAR 단백질, 예컨대 hADAR1 또는 hADAR2 로부터 유래할 수 있다. 일부 구현예에서, 적어도 하나의 dsRBD 및 디아미나제 도메인을 포함하는 전장 ADAR 단백질이 사용된다. 일부 구현예에서, 하나 이상의 dsRBM 또는 dsRBD 은 디아미나제 도메인의 N-말단에 있다. 다른 구현예에서, 하나 이상의 dsRBM 또는 dsRBD 은 디아미나제 도메인의 C-말단에 있다.

[0608] 일부 구현예에서, 아데노신 디아미나제의 기질 편집 선호도는 효소의 활성 중심에서의 또는 그 근처의 아미노산 잔기에 의해 영향을 받는다. 일부 실시형태에서, 기질 편집 선호도를 조정하기 위해서, 아데노신 데아미나아제는 하기 돌연변이 중 하나 이상을 포함할 수 있다: G336D, G487R, G487K, G487W, G487Y, E488Q, E488N, T490A, V493A, V493T, V493S, N597K, N597R, A589V, S599T, N613K, N613R (hADAR2-D 의 아미노산 서열 위치에 기초함), 및 상기에 상응하는 상동성 ADAR 단백질에서의 돌연변이.

[0609] 특히, 일부 구현예에서, 편집 특이성을 감소시키기 위해서 아데노신 디아미나제는 하기 돌연변이 중 하나 이상을 포함할 수 있다: E488Q, V493A, N597K, N613K (hADAR2-D 의 아미노산 서열 위치에 기초함), 및 상기에 상응하는 상동성 ADAR 단백질에서의 돌연변이. 일부 구현예에서, 편집 특이성을 증가시키기 위해서, 아데노신 디아미나제는 돌연변이 T490A 를 포함할 수 있다.

[0610] 일부 구현예에서, 바로 5' G를 갖는 표적 아데노신(A), 예컨대, 삼중 서열 GAC를 포함하는 기질(중심 A는 표적 아데노신 잔기임)에 대한 편집 선호도를 증가시키기 위해, 아데노신 디아미나제는 hADAR2-D의 아미노산 서열 위치에 기반한 돌연변이 G336D, E488Q, E488N, V493T, V493S, V493A, A589V, N597K, N597R, S599T, N613K, N613R, 및 상기에 대응하는 상동성 ADAR 단백질에서의 돌연변이 중 하나 이상을 포함할 수 있다.

[0611] 특히, 일부 실시형태에서, 하기 삼중 서열을 포함하는 기질을 편집하기 위해 아데노신 데아미나아제는 돌연변이 E488Q 또는 상동성 ADAR 단백질에서의 상응하는 돌연변이를 포함한다: GAC, GAA, GAU, GAG, CAU, AAU, UAC (중심 A 는 표적 아데노신 잔기이다).

[0612] 일부 구현예에서, 아데노신 디아미나제는 hADAR1-D의 야생형 아미노산 서열을 포함한다. 일부 실시형태에서, 아데노신 데아미나아제는 하나 이상의 돌연변이를 hADAR1-D 서열에 포함하여, hADAR1-D 의 편집 효율, 및/또는 기질 편집 선호도는 특별한 필요에 따라 변화된다.

[0613] 일부 구현예에서, 아데노신 디아미나제는 hADAR1-D 아미노산 서열의 글리신1007, 또는 상동성 ADAR 단백질 내 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 1007의 글리신 잔기는 비교적 소형 측쇄를

갖는 비극성 아미노산 잔기로 치환된다. 예를 들어, 일부 구현예에서, 위치 1007의 글리신 잔기는 알라닌 잔기로 치환된다 (G1007A). 일부 구현예에서, 위치 1007의 글리신 잔기는 발린 잔기로 치환된다 (G1007V). 일부 구현예에서, 위치 1007의 글리신 잔기는 비교적 큰 측쇄를 갖는 아미노산 잔기로 치환된다. 일부 구현예에서, 위치 1007의 글리신 잔기는 아르기닌 잔기로 치환된다 (G1007R). 일부 구현예에서, 위치 1007의 글리신 잔기는 리신 잔기로 치환된다 (G1007K). 일부 구현예에서, 위치 1007의 글리신 잔기는 트립토판 잔기로 치환된다 (G1007W). 일부 구현예에서, 위치 1007의 글리신 잔기는 티로신 잔기로 치환된다 (G1007Y). 부가적으로, 다른 구현예에서, 위치 1007의 글리신 잔기는 류신 잔기로 치환된다 (G1007L). 다른 구현예에서, 위치 1007의 글리신 잔기는 트레오닌 잔기로 치환된다 (G1007T). 다른 구현예에서, 위치 1007의 글리신 잔기는 세린 잔기로 치환된다 (G1007S).

[0614] 일부 구현예에서, 아데노신 디아미나제는 hADAR1-D 아미노산 서열의 글루탐산1008, 또는 상동성 ADAR 단백질에서의 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 위치 1008의 글루탐산 잔기는 상대적으로 큰 측쇄를 갖는 극성 아미노산 잔기로 치환된다. 일부 구현예에서, 위치 1008의 글루탐산 잔기는 글루타민 잔기로 치환된다 (E1008Q). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 히스티딘 잔기로 치환된다 (E1008H). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 아르기닌 잔기로 치환된다 (E1008R). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 리신 잔기로 치환된다 (E1008K). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 비극성 또는 소형 극성 아미노산 잔기로 치환된다. 일부 구현예에서, 위치 1008의 글루탐산 잔기는 페닐알라닌 잔기로 치환된다 (E1008F). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 트립토판 잔기로 치환된다 (E1008W). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 글라이신 잔기로 치환된다 (E1008G). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 이소류신 잔기로 치환된다 (E1008I). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 발린 잔기로 치환된다 (E1008V). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 프롤린 잔기로 치환된다 (E1008P). 일부 구현예에서, 위치 1008의 글루탐산 잔기는 세린 잔기로 치환된다 (E1008S). 다른 구현예에서, 위치 1008의 글루탐산 잔기는 아스파라긴 잔기로 치환된다 (E1008N). 다른 구현예에서, 위치 1008의 글루탐산 잔기는 알라닌 잔기로 치환된다 (E1008A). 다른 구현예에서, 위치 1008의 글루탐산 잔기는 메티오닌 잔기로 치환된다 (E1008M). 일부 구현예에서, 위치 1008번의 글루탐산 잔기는 류신 잔기로 치환된다 (E1008L).

[0615] 일부 실시형태에서, 편집 효율을 개선하기 위해서, 아데노신 디아미나아제는 hADAR1-D의 아미노산 서열 위치를 기반으로, 돌연변이: E1007S, E1007A, E1007V, E1008Q, E1008R, E1008H, E1008M, E1008N, E1008K, 및 상기에 상응하는 상동성 ADAR 단백질에서의 돌연변이 중 하나 이상을 포함할 수 있다.

[0616] 일부 실시형태에서, 편집 효율을 개선하기 위해서, 아데노신 디아미나아제는 hADAR1-D의 아미노산 서열 위치를 기반으로, 돌연변이: E1007R, E1007K, E1007Y, E1007L, E1007T, E1008G, E1008I, E1008P, E1008V, E1008F, E1008W, E1008S, E1008N, E1008K, 및 상기에 상응하는 상동성 ADAR 단백질에서의 돌연변이 중 하나 이상을 포함할 수 있다.

[0617] 일부 구현예에서, 아데노신 디아미나제의 기질 편집 선호도, 효율 및/또는 선택성은 효소의 활성 중심에서의 또는 그 근처의 아미노산 잔기에 의해 영향을 받는다. 일부 구현예에서, 아데노신 디아미나제는 hADAR1-D 서열의 글루탐산 1008 위치, 또는 상동성 ADAR 단백질에서의 상응하는 위치에서 돌연변이를 포함한다. 일부 구현예에서, 돌연변이는 E1008R, 또는 상동성 ADAR 단백질에서의 상응하는 돌연변이이다. 일부 구현예에서, E1008R 돌연변이체는 맞은편 가닥에 미스매치되는 G 잔기를 갖는 표적 아데노신 잔기에 대해 증가된 편집 효율을 갖는다.

[0618] 일부 구현예에서, 이중-가닥 핵산 기질을 인지하고 그에 결합하기 위해 아데노신 디아미나제 단백질은 하나 이상의 이중-가닥 RNA (dsRNA) 결합 모티프 (dsRBM) 또는 도메인 (dsRBD) 을 추가로 포함하거나 또는 그에 연결된다. 일부 구현예에서, 아데노신 디아미나제와 이중-가닥 기질 사이의 상호작용은 CRISPR/CAS 단백질 인자를 포함하는 하나 이상의 부가적 단백질 인자(들)에 의해 매개된다. 일부 구현예에서, 아데노신 디아미나제와 이중-가닥 기질 사이의 상호작용은 가이드 RNA 를 포함하는 하나 이상의 핵산 성분(들)에 의해 추가로 매개된다.

[0619] C에서 U로 탈아미노화 활성을 갖는 변형된 아데노신 디아미나제

[0620] 일정한 예의 구현예에서, 지정된 진화는 아데닌의 하이포잔틴으로의 탈아미노화 이외의 추가적인 반응을 촉매할 수 있는 변형된 ADAR 단백질을 설계하기 위해 사용될 수 있다. 예를 들어, 변형된 ADAR 단백질은 시티딘의 우라실로의 탈아미노화를 촉매할 수 있다. 특정 이론으로 구속되지는 않지만, C에서 U로의 활성을 개선시키

는 돌연변이는 더 작은 시티딘 염기를 더 잘 받을 수 있도록 결합 포켓의 형상을 변경시킬 수 있다.

[0621] 일부 구현예에서, C에서 U로의 탈아미노화 활성을 갖는 변형된 아데노신 디아미나제는 hADAR2-D 아미노산 서열의 위치 V351, T375, R455 및 E488, 또는 상동성 ADAR 단백질 내 대응하는 위치 중 하나 이상에서 돌연변이를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 E488Q를 포함한다. 일부 구현예에서, 아데노신 디아미나제는 V351I, V351L, V351F, V351M, V351C, V351A, V351G, V351P, V351T, V351S, V351Y, V351W, V351Q, V351N, V351H, V351E, V351D, V351K, V351R, T375I, T375L, T375V, T375F, T375M, T375C, T375A, T375G, T375P, T375S, T375Y, T375W, T375Q, T375N, T375H, T375E, T375D, T375K, T375R, R455I, R455L, R455V, R455F, R455M, R455C, R455A, R455G, R455P, R455T, R455S, R455Y, R455W, R455Q, R455N, R455H, R455E, R455D, R455K로부터 선택된 돌연변이 중 하나 이상을 포함한다. 일부 구현예에서, 아데노신 디아미나제는 돌연변이 E488Q를 포함하고, 추가로 V351I, V351L, V351F, V351M, V351C, V351A, V351G, V351P, V351T, V351S, V351Y, V351W, V351Q, V351N, V351H, V351E, V351D, V351K, V351R, T375I, T375L, T375V, T375F, T375M, T375C, T375A, T375G, T375P, T375S, T375Y, T375W, T375Q, T375N, T375H, T375E, T375D, T375K, T375R, R455I, R455L, R455V, R455F, R455M, R455C, R455A, R455G, R455P, R455T, R455S, R455Y, R455W, R455Q, R455N, R455H, R455E, R455D, R455K로부터 선택된 돌연변이 중 하나 이상을 포함한다.

[0622] C에서 U로의 탈아미노화 활성을 갖는 앞서 언급한 변형된 ADAR 단백질과 관련하여, 본 명세서에 기재된 발명은 또한 표적 RNA 또는 DNA에 본 명세서에 개시된 AD-작용화된 조성물을 전달하는 단계를 포함하는, 관심 대상의 표적 RNA 서열에서 C를 탈아미노화시키는 방법에 관한 것이다.

[0623] 소정의 예시적 실시형태에서, 표적 RNA 서열에서 C를 탈아미노화시키기 위한 방법으로서, 상기 방법은 상기 표적 RNA로, (a) 촉매적으로 비활성인 (데드) Cas; (b) 직접 반복부 서열에 연결된 가이드 서열을 포함하는 가이드 분자; 및 (c) C-대-U 탈아미노화 활성을 갖는 변형된 ADAR 단백질 또는 이의 촉매적 도메인을 전달하는 단계를 포함하고, 상기 변형된 ADAR 단백질 또는 이의 촉매적 도메인은 상기 데드 Cas 단백질 또는 상기 가이드 분자에 공유적으로 또는 비공유적으로 연결되거나 또는 전달 후에 이에 연결되기에 적합하게 되고; 가이드 분자는 상기 데드 Cas 단백질과 복합체를 형성하고, 상기 복합체가 상기 관심 대상의 표적 RNA 서열에 결합하도록 지시하며; 상기 가이드 서열은 상기 C를 포함하는 표적 서열과 하이브리드화하여 RNA 듀플렉스를 형성할 수 있고; 선택적으로, 상기 가이드 서열은 상기 C에 대응하는 위치에서 비쌍형성 A 또는 U를 포함하고 형성된 RNA 듀플렉스에 미스매치를 초래하며; 상기 변형된 ADAR 단백질 또는 이의 촉매적 도메인은 상기 RNA 듀플렉스의 상기 C를 탈아미노화시킨다.

[0624] C-대-U 탈아미드화 활성을 갖는 상기 언급된 변형된 ADAR 단백질과 관련하여, 본 명세서에 기술된 방법은 또한 (a) 직접 반복부 서열에 연결된 가이드 서열을 포함하는 가이드 분자, 또는 상기 가이드 분자를 코딩하는 뉴클레오타이드 서열; (b) 촉매적으로 불활성인 Cas13 단백질, 또는 상기 촉매적으로 불활성인 Cas13 단백질을 코딩하는 뉴클레오타이드 서열; (c) C-대-U 탈아미드화 활성을 갖는 변형된 ADAR 단백질 또는 이의 촉매 도메인, 또는 상기 변형된 ADAR 단백질 또는 이의 촉매 도메인을 코딩하는 뉴클레오티드를 포함하는, 관심 표적 유전자좌에서 C를 탈아미드화시키는데 적합한 조작된, 비천연 발생 시스템에 관한 것으로서, 상기 변형된 ADAR 단백질 또는 이의 촉매 도메인은 상기 Cas12 단백질 또는 상기 가이드 분자에 공유적으로 또는 비공유적으로 연결되거나 또는 전달 후 이에 연결되도록 조정되며; 상기 가이드 서열은 RNA 듀플렉스를 형성하도록 C를 포함하는 표적 RNA 서열과 하이브리드화할 수 있고, 임의로는, 상기 가이드 서열은 형성된 RNA 듀플렉스에 미스매치를 초래하는 상기 C에 상응하는 위치에 비쌍형성 A 또는 U를 포함하고, 임의로, 시스템은 (a) 상기 가이드 서열을 포함하는 상기 가이드 분자를 코딩하는 뉴 클레오티드 서열에 작동적으로 연결된 제1 조절 엘리먼트, (b) 상기 촉매적 불활성 Cas13 단백질을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트; 및 (c) 상기 1 또는 제2 조절 엘리먼트 하에 있거나 또는 제3 조절 엘리먼트에 작동적으로 연결된 C-대-U 탈아미드화 활성을 갖는 변형된 ADAR 단백질 또는 이의 촉매 도메인을 코딩하는 뉴클레오티드 서열을 포함하는 하나 이상의 벡터를 포함하는 벡터 시스템이고; 변형된 ADAR 단백질 또는 이의 촉매 도메인을 코딩하는 상기 뉴클레오티드 서열이 제3 조절 엘리먼트에 작동적으로 연결되면, 상기 변형된 ADAR 단백질 또는 이의 촉매 도메인은 발현 이후에 상기 가이드 분자 또는 상기 Cas12에 연결되도록 조정되고, 성분 (a), (b), 및 (c)는 시스템의 동일하거나 또는 상이한 벡터 상에 위치하고, 임의로는 상기 제1, 제2, 및/또는 제3 조절 엘리먼트는 유도성 프로모터이다.

[0625] 본 발명의 일 구현예에서, 아데노신 디아미나제의 기질은 이의 DNA 표적에 대한 가이드 분자의 결합 시 형성된 RNA/DNA 헤테로듀플렉스이며, 이어서, CRISPR-Cas 효소와 CRISPR-Cas 복합체를 형성한다. RNA/DNA 또는 DNA/RNA 헤테로듀플렉스는 또한 본 명세서에서 "RNA/DNA 하이브리드", "DNA/RNA 하이브리드" 또는 "이중가닥

기질" 로서 지칭된다.

[0626] 본 발명에 따르면, 아데노신 디아미나제의 기질은 가이드 분자가 그것의 DNA 표적 (이는 그 후 CRISPR-Cas 효소와 CRISPR-Cas 복합체를 형성한다) 에 결합시 형성되는 RNA/DNAn RNA 듀플렉스이다. 아데노신 디아미나제의 기질은 또한 가이드 분자가 그의 DNA 표적에 결합 시 형성된 RNA/RNA 듀플렉스일 수 있고, 이어서, CRISPR-Cas 효소와의 CRISPR-Cas 복합체를 형성한다. RNA/DNA 또는 DNA/RNAn RNA 듀플레스는 또한 본 명세서에서 "RNA/DNA 하이브리드", "DNA/RNA 하이브리드" 또는 "이중-가닥 기질" 로서 지칭된다. 가이드 분자 및 CRISPR-Cas 효소의 특정 특징이 이하에 상술된다.

[0627] 본 명세서에서 사용되는 용어 "편집 선택성" 은 아데노신 디아미나제에 의해 편집된 이중가닥 기질 상의 모든 부위의 분율을 지칭한다. 이론에 의해 구속되는 일 없이, 아데노신 디아미나제의 편집 선택성은 이중가닥 기질의 길이 및 2차 구조, 예컨대, 미스매치된 염기, 벌지(bulge) 및/또는 내부 루프의 존재에 의해 영향받는다는 것이 상정된다.

[0628] 일부 구현예에서, 기질이 50 bp 초과의 완벽히 염기 쌍을 이룬 듀플렉스일 때, 아데노신 디아미나제는 듀플렉스 내의 복수의 아데노신 잔기 (예를 들어, 모든 아데노신 잔기의 50%) 를 탈아미노화할 수 있다. 일부 구현예에서, 기질이 50 bp 보다 짧을 때, 아데노신 디아미나제의 편집 선택성은 표적 아데노신 자리에서의 미스매치의 존재에 의해 영향을 받는다. 특히, 일부 구현예에서, 맞은편 가닥에 미스매치되는 시티딘 (C) 잔기를 갖는 아데노신 (A) 잔기는 높은 효율로 탈아미노화된다. 일부 구현예에서, 맞은편 가닥에 미스매치되는 구아노신 (G) 잔기를 갖는 아데노신 (A) 잔기는 편집 없이 건너뛰어 진다.

[0629] 특정 구현예에서, 아데노신 디아미나제 단백질 또는 이의 촉매 도메인은 세포로 전달되거나 또는 별개 단백질로서 세포 내에서 발현되지만, C2c1 단백질 또는 가이드 분자에 연결될 수 있도록 변형된다. 특정한 구현예에서, 이것은 다양한 박테리오파지 외피 단백질 내에 존재하는 직교성 RNA-결합 단백질 또는 어댑터 단백질/압타머 조합의 사용에 의해 보장된다. 이러한 외피 단백질의 예는 제한없이 MS2, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ΦCb5, ΦCb8r, ΦCb12r, ΦCb23r, 7s 및 PRR1을 포함한다. 압타머는 천연 발생일 수 있거나 또는 특이적 표적에 결합하도록 시험관내 스크리닝 또는 SELEX (systematic evolution of ligands by exponential enrichment)의 반복 라운드를 통해서 조작된 합성 올리고뉴클레오티드일 수 있다.

[0630] 특정 구현에에서, 가이드 분자는 어댑터 단백질을 동원할 수 있는 하나 이상의 별개의 RNA 루프(들) 또는 별개의 서열(들)이 제공된다. 가이드 분자는 별개 RNA 루프(들) 또는 별개 서열(들)에 결합할 수 있는 어댑터 단백질을 동원할 수 있는 별개 RNA 루프(들) 또는 별개 서열(들)의 삽입에 의해 C2c1 단백질과 충돌없이 연장될 수 있다. 변형된 가이드 및 C2c1 복합체로 이펙터 도메인을 동원시키는데서 그들의 용도가 문헌 [Konermann (Nature 2015, 517(7536): 583- 588)]에 제공된다. 특정 구현예에서, 압타머는 포유동물 세포에서 이량체화된 MS2 박테리오파지 외피 단백질에 선택적으로 결합하고, 가이드 분자, 예컨대 스템루프 및/또는 테트라루프 내에 도입되는 최소 헤어핀 압타머이다. 이들 구현예에서, 아데노신 디아미나제 단백질은 MS2에 융합된다. 아데노신 디아미나제 단백질은 CRISPR-Cas 단백질 및 상응하는 가이드 RNA와 함께 공동-전달된다.

[0631] 일부 구현예에서, 본 명세서에 기술된 C2c1-ADAR 염기 편집 시스템은 (a) 촉매적 불활성 또는 닉카제인 C2c1 단백질; (b) 가이드 서열을 포함하는 가이드 분자; 및 (c) 아데노신 디아미나제 단백질 또는 이의 촉매적 도메인을 포함하고, 아데노신 디아미나제 단백질 또는 이의 촉매적 도메인은 C2c1 단백질 또는 가이드 분자에 공유적으로 또는 비공유적으로 연결되거나 또는 전달 후에 이에 연결되도록 적합화되고; 가이드 서열은 표적 서열에 실질적으로 상보성이지만, 탈아미노화를 위해 표적화되는 A에 상응하는 비쌍형성 C를 포함하여, 가이드 서열 및 표적 서열에 의해 형성된 DNA-RNA 또는 RNA-RNA 듀플렉스에 A-C 미스매치를 초래한다. 진핵생물 세포에서 적용을 위해, C2c1 단백질 및/또는 아데노신 디아미나제는 바람직하게 NLS-태그화된다.

[0632] 일부 구현예에서, 성분 (a), (b) 및 (c)는 리보핵단백질 복합체로서 세포로 전달된다. 리보핵단백질 복합체는 하나 이상의 지질 나노입자를 통해 전달될 수 있다.

[0633] 일부 구현예에서, 성분 (a), (b) 및 (c)는 하나 이상의 RNA 분자, 예컨대 하나 이상의 가이드 RNA 및 C2c1 단백질, 아데노신 디아미나제 단백질, 및 임의로는 어댑터 단백질을 코딩하는 하나 이상의 mRNA 분자로서 세포로 전달된다. RNA 분자는 하나 이상의 지질 나노입자를 통해서 전달될 수 있다.

[0634] 일부 구현예에서, 성분 (a), (b) 및 (c)는 하나 이상의 DNA 분자로서 세포에 전달된다. 일부 구현예에서, 하나 이상의 DNA 분자는 하나 이상의 벡터 예컨대 바이러스 벡터 (예를 들어, AAV) 내에 포함된다. 일부 구

현예에서, 하나 이상의 DNA 분자는 C2c1 단백질, 가이드 분자, 및 아데노신 디아미나제 단백질 또는 이의 촉매 도메인을 발현하도록 작동적으로 구성된 하나 이상의 조절 엘리먼트를 포함하고, 임의로는 하나 이상의 조절 엘리먼트는 유도성 프로모터를 포함한다.

[0635] 일부 구현에에서 가이드 분자는 표적 유전자좌에서 제1 DNA 가닥 또는 RNA 가닥 내에 탈아미드화되는 아데닌을 포함하는 표적 서열과 하이브리드화할 수 있어서 상기 아데닌에 대항한 비쌍형성 시토신을 포함하는 DNA-RNA 또는 RNA-RNA 듀플렉스를 형성한다. 듀플렉스 형성 시, 가이드 분자는 C2c1 단백질과 복합체를 형성하고, 관심 표적 유전자좌에서 상기 복합체가 상기 제1 DNA 가닥 또는 상기 RNA 가닥에 결합하도록 유도한다. C2c1-ADAR 염기 편집 시스템의 가이드의 양상에 대한 상세 설명은 하기 본 명세서에서 제공된다.

[0636] 일부 구현에에서, 정규 길이 (예를 들어, AsCas13의 경우 약 20 nt)를 갖는 C2c1 가이드 RNA는 표적 DNA 또는 RNA와 DNA-RNA 또는 RNA-RNA 듀플렉스를 형성시키는데 사용된다. 일부 구현예에서, 정규 길이보다 긴 (예를 들어, AsCas13의 경우 >20 nt) C2c1 가이드 분자는 C2c1-가이드 RNA-표적 DNA 복합체의 외부를 포함하는 표적 DNA 또는 RNA와 DNA-RNA 또는 RNA-RNA 듀플렉스를 형성시키는데 사용된다. 일정 예의 구현예에서, 가이드 서열은 상기 표적 서열과 DNA-RNA 또는 RNA-RNA 듀플렉스를 형성할 수 있는 약 29 내지 53 nt의 길이를 갖는다. 일정 예의 다른 구현예에서, 가이드 서열은 상기 표적 서열과 DNA-RNA 또는 RNA-RNA 듀플렉스를 형성할 수 있는 약 40 내지 50 nt의 길이를 갖는다. 일정 예의 구현예에서, 상기 가이드 서열의 5' 말단과 상기 비-쌍형성 C 사이의 거리는 20-30개 뉴클레오티드이다. 일정 예의 구현예에서, 상기 가이드 서열의 3' 말단과 상기 비-쌍형성 C 사이의 거리는 20-30개 뉴클레오티드이다.

[0637] 적어도 제1 디자인에서, C2c1-ADAR 시스템은 (a) C2c1 단백질에 융합 또는 연결된 아데노신 디아미나제로서, C2c1 단백질은 촉매적으로 불활성이거나 또는 닉카제인 것인 아데노신 디아미나제, 및 (b) 가이드 서열과 표적 서열 간에 형성된 DNA-RNA 또는 RNA-RNA 듀플렉스에 A-C 미스매치를 도입시키도록 설계된 가이드 서열을 포함하는 가이드 분자를 포함한다. 일부 구현예에서, C2c1 단백질 및/또는 아데노신 디아미나제는 N-말단 또는 C-말단 또는 둘 모두 상에서, NLS-태그화된다.

[0638] 적어도 제2 디자인에서, C2c1-ADAR 시스템은 (a) 촉매적으로 불활성이거나 또는 닉카제인 C2c1 단백질, (b) 가이드 서열 및 표적 서열 간에 형성된 DNA-RNA 또는 RNA-RNA 듀플렉스에 A-C 미스매치를 도입시키기 위해 디자인된 가이드 서열, 및 어댑터 단백질 (예를 들어, MS2 외피 단백질 또는 PP7 외피 단백질)에 결합할 수 있는 압타머 서열 (예를 들어, MS2 RNA 모티프 또는 PP7 RNA 모티프)를 포함하는 가이드 분자, 및 (c) 어댑터 단백질에 융합 또는 연결된 아데노신 디아미나제로서, 압타머와 어댑터 단백질의 결합이 A-C 미스매치의 A에서 표적화된 탈아미노화를 위해 가이드 서열과 표적 서열 간에 형성된 DNA-RNA 또는 RNA-RNA 듀플렉스로 아데노신 디아미나제를 동원하는 것인 아데노신 디아미나제를 포함한다. 일부 구현예에서, 어댑터 단백질 및/또는 아데노신 디아미나제는 N-말단 또는 C-말단 또는 둘 모두 상에서, NLS-태그화된다. C2c1 단백질은 또한 NLS-태그화될 수 있다.

[0639] 상이한 압타머 및 상응하는 어댑터 단백질의 사용은 또한 직교성 유전자 편집을 제공할 수 있게 한다. 아데노신 디아미나제를 직교성 유전자 편집/탈아미노화를 위해 시티딘 디아미나제와 조합하여 사용하는 일례에서, 상이한 유전자좌를 표적화하는 sgRNA는 각각 MS2-아데노신 디아미나제 및 PP7-시티딘 디아미나제 (또는 PP7-아데노신 디아미나제 및 MS2-시티딘 디아미나제)를 동원하기 위해 별개의 RNA 루프에 의해 변형되어서, 각각 관심 표적 유전자좌에서 A 또는 C의 직교성 탈아미노화를 일으킨다. PP7은 박테리오파지 슈도모나스(Pseudomonas)의 RNA-결합 외피 단백질이다. MS2 처럼, 이것은 특별한 RNA 서열 및 2차 구조에 결합된다. PP7 RNA-인식 모티프는 MS2와 별개이다. 결과적으로, PP7 및 MS2는 상이한 게놈 유전자좌에서 동시에 별개의 효과를 매개하도록 다중복합체화될 수 있다. 예를 들어, 유전자좌 A를 표적화하는 sgRNA는 MS2 루프에 의해 변형되어서, MS2-아데노신 디아미나제를 동원하는 한편, 유전자좌 B를 표적화하는 다른 sgRNA는 PP7 루프에 의해 변형되어, PP7-시티딘 디아미나제를 동원한다. 따라서 동일한 세포에서, 직교성, 유전자좌-특이적 변형을 인지한다. 이러한 원리는 다른 직교성 RNA-결합 단백질을 도입시키기 위해 확대될 수 있다.

[0640] 적어도 제3 디자인에서, C2c1-ADAR CRISPR 시스템은 (a) C2c1 단백질의 내부 루프 또는 비구조화 영역에 삽입된 아데노신 디아미나제로서, C2c1 단백질은 촉매적으로 불활성이거나 또는 닉카제인 아데노신 디아미나제, 및 (b) 가이드 서열 및 표적 서열 간에 형성된 DNA-RNA 또는 RNA-RNA 듀플렉스에 A-C 미스매치를 도입하도록 설계된 가이드 서열을 포함하는 가이드 분자를 포함한다.

[0641] 아데노신 디아미나제의 삽입에 적합한 C2c1 단백질 분할 부위는 결정 구조의 도움으로 확인될 수 있다. 예를 들어, AacC2c1 돌연변이체의 경우에, 예를 들어 서열 정렬을 위한 상응하는 위치가 무엇인지 쉽게 이해될 것

이다. 다른 C2c1 단백질의 경우에 오솔로그와 의도하는 C2c1 단백질 간에 비교적 높은 정도의 상동성이 존재하면 오솔로그의 결정 구조를 사용할 수 있다.

[0642] 분할 위치는 영역 또는 루프 내에 위치될 수 있다. 바람직하게, 분할 위치는 아미노산 서열의 파괴가 구조적 특성 (예를 들어, 알파-나선 또는 베타-시트)의 부분 또는 완전 파괴를 초래하지 않는 곳에서 발생된다. 비구조화된 영역 (이들 영역이 결정 내에서 "동결" 되기에 충분하게 구조화되지 않기 때문에 결정 구조에서 나타나지 않는 영역)이 종종 바람직한 선택사항이다. C2c1의 표면 상에 노출된 모든 비구조적 영역에서 분할은 본 발명의 실시에서 계획된다. 비구조화된 영역 또는 외부 루프 내 위치는 정확하게 상기 제공된 번호일 필요는 없지만, 분할 위치가 루프 외부의 비구조화된 영역 내에 여전히 위치되는 한, 루프의 크기에 따라서, 상기 제공된 위치의 양쪽 측면에서 예를 들어 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 아미노산에 따라 다양할 수 있다.

[0643] 본 명세서에 기재된 C2c1-ADAR CRISPR 시스템은 탈아미노화를 위해 DNA 서열 내에서 특정 아데닌을 표적화하는데 사용될 수 있다. 예를 들어, 가이드 분자는 C2c1 단백질과 복합체를 형성할 수 있고, 관심 대상의 표적 유전자좌에서 복합체가 표적 서열에 결합하도록 지시한다. 가이드 서열이 비쌍형성 C를 갖도록 설계되기 때문에, 가이드 서열과 표적 서열 사이에 형성된 헤테로듀플렉스은 아데노신 디아미나제가 비쌍형성 C와 마주보는 A와 접촉하여 탈아미노화시키고, 이를 이노신 (I)으로 전환시키도록 지시하는, A-C 미스매치를 포함한다. 이노신 (I) 염기는 C 와 쌍을 이루고 세포 과정에서 G 와 같이 기능하므로, 본원에 기재된 A 의 표적화된 탈아미노화는 원하지 않는 G-A 및 C-T 돌연변이의 보정, 뿐만 아니라 바람직한 A-G 및 T-C 돌연변이의 수득에 유용하다.

[0644] **염기 절제 복구 억제제**

[0645] 일부 구현예에서, AD-작용화된 CRISPR 시스템은 염기 절제 복구 (base excision repair: BER) 억제제를 추가로 포함한다. 임의의 특정 이론에 국한하고 싶지 않지만, I:T 쌍형성의 존재에 대한 세포의 DNA-복구 반응은 세포 내 핵염기 편집 효율의 감소를 초래할 수 있다. 알킬아데닌 DNA 글리코실라제 (또한 DNA-3-메틸아데닌 글리코실라제, 3-알킬아데닌 DNA 글리코실라제, 또는 N-메틸퓨린 DNA 글리코실라제로서 알려짐)는 세포에서 DNA로부터 하이포잔틴의 제거를 촉매하는데, 이는 결과로서 I:T 쌍의 A:T 쌍으로의 반전과 함께 염기 절제 복구를 개시한다.

[0646] 일부 구현예에서, BER 억제제는 알킬아데닌 DNA 글리코실라제의 억제제이다. 일부 구현예에서, BER 억제제는 인간 알킬아데닌 DNA 글리코실라제의 억제제이다. 일부 구현예에서, BER 억제제는 폴리펩티드 억제제이다. 일부 구현예에서, BER 억제제는 하이포잔틴에 결합하는 단백질이다. 일부 구현예에서, BER 억제제는 DNA에서 하이포잔틴에 결합하는 단백질이다. 일부 구현예에서, BER 억제제는 촉매적 불활성인 알킬아데닌 DNA 글리코실라제 단백질 또는 이의 결합 도메인이다. 일부 구현예에서, BER 억제제는 DNA로부터 하이포잔틴을 절단하지 않는 촉매적 불활성인 알킬아데닌 DNA 글리코실라제 단백질 또는 이의 결합 도메인이다. 알킬아데닌 DNA 글리코실라제 염기-절단 복구 효소를 억제할 수 있는 (예를 들어, 입체적으로 차단하는) 다른 단백질은 본 개시내용의 범주 내이다. 추가적으로, 염기-절단 복구를 차단 또는 억제하는 임의의 단백질은 본 개시내용의 범주 내이다.

[0647] 임의의 특정 이론으로 구속되는 일 없이, 염기 절제 복구sms 편집된 가닥에 결합하고/하거나, 편집 염기를 차단하고/하거나, 알킬아데닌 DNA 글리코실라제를 억제하고/하거나, 염기 절제 복구를 억제하고/하거나, 편집 염기를 보호하고/하거나 비편집 가닥의 고정을 촉진시키는 분자에 의해 억제될 수 있다. 본 명세서에 기재된 BER 억제제의 사용은 A의 I로의 변화를 촉매할 수 있는 아데노신 디아미나제의 편집 효율을 증가시킬 수 있는 것으로 여겨진다.

[0648] 따라서, 상기 논의한 AD-작용화된 CRISPR 시스템의 제1 디자인에서, CRISPR-Cas 단백질 또는 아데노신 디아미나제는 BER 억제제(예를 들어, 알킬아데닌 DNA 글리코실라제의 억제제)에 융합되거나 또는 연결될 수 있다. 일부 구현예에서, BER 억제제는 하기 구조 중 하나로 포함될 수 있다 (nC2c1=C2c1 닉카제; dC2c1=데드 C2c1):[AD]-[선택적 링커]-[nC2c1/dC2c1]-[선택적 링커]-[BER 억제제]; [AD]-[선택적 링커]-[BER 억제제]-[선택적 링커]-[nC2c1/dC2c1]; [BER 억제제]-[선택적 링커]-[AD]-[선택적 링커]-[nC2c1/dC2c1]; [BER 억제제]-[선택적 링커]-[nC2c1/dC2c1]-[선택적 링커]-[AD]; [nC2c1/dC2c1]-[선택적 링커]-[AD]-[선택적 링커]-[BER 억제제]; [nC2c1/dC2c1]-[선택적 링커]-[BER 억제제]-[선택적 링커]-[AD].

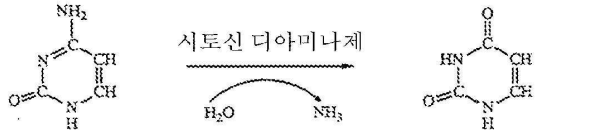
[0649] 유사하게는, 상기 논의한 AD-작용화된 CRISPR 시스템의 제2 디자인에서, CRISPR-Cas 단백질, 아데노신 디아미나

제 또는 어댑터 단백질이 BER 억제제 (예를 들어, 알킬아데닌 DNA 글리코실라제의 억제제)에 융합되거나 또는 연결될 수 있다. 일부 구현예에서, BER 억제제는 하기 구조 중 하나로 포함될 수 있다 (nC2c1=C2c1 닉카제; dC2c1=데드 C2c1):[nC2c1/dC2c1]-[선택적 링커]-[BER 억제제]; [BER 억제제]-[선택적 링커]-[nC2c1/dC2c1]; [AD]-[선택적 링커]-[어댑터]-[선택적 링커]-[BER 억제제]; [AD]-[선택적 링커]-[BER 억제제]-[선택적 링커]-[어댑터]; [BER 억제제]-[선택적 링커]-[AD]-[선택적 링커]-[어댑터]; [BER 억제제]-[선택적 링커]-[어댑터]-[선택적 링커]-[AD]; [어댑터]-[선택적 링커]-[AD]-[선택적 링커]-[BER 억제제]; [어댑터]-[선택적 링커]-[BER 억제제]-[선택적 링커]-[AD].

[0650] 상기 논의한 AD-작용화된 CRISPR 시스템의 제3 디자인에서, BER 억제제는 CRISPR-Cas 단백질의 내부 루프 또는 비구조화된 영역에 삽입될 수 있다.

[0651] ***시티딘 디아미나제***

[0652] 일부 구현예에서, 디아미나제는 시티딘 디아미나제이다. 본 명세서에서 사용되는 용어 "시티딘 디아미나제" 또는 "시티딘 디아미나제 단백질" 은 이하에 표시된 바와 같이, 시토신 (또는 분자의 시토신 모이어티)을 우라실 (또는 분자의 우라실 모이어티)으로 전환시키는 가수분해 탈아미노화 반응을 촉매할 수 있는 단백질, 폴리펩티드, 또는 단백질 또는 폴리펩티드의 하나 이상의 기능성 도메인(들)에 관한 것이다. 일부 구현예에서, 시토신-함유 분자는 시티딘(C)이고, 우라실-함유 분자는 우리딘(U)이다. 시토신-함유 분자는 데옥시리보핵산 (DNA) 또는 리보핵산 (RNA)일 수 있다.



시토신
(4-아미노-2-옥시피리미딘)

우라실
(2,4-디옥소피리미딘)

[0653]

[0654] 본 개시에 따라서, 본 개시와 함께 사용할 수 있는 시티딘 디아미나제는 제한없이 아포리포단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 디아미나제로서 알려진 효소 패밀리의 멤버, 활성화-유도된 디아미나제 (AID), 또는 시티딘 디아미나제 1 (CDA1)을 포함한다. 특정 구현예에서, 디아미나제는 APOBEC1 디아미나제, APOBEC2 디아미나제, APOBEC3A 디아미나제, APOBEC3B 디아미나제, APOBEC3C 디아미나제, 및 APOBEC3D 디아미나제, APOBEC3E 디아미나제, APOBEC3F 디아미나제, APOBEC3G 디아미나제, APOBEC3H 디아미나제, 또는 APOBEC4 디아미나제이다.

[0655] 본 발명의 방법 및 시스템에서, 시티딘 디아미나제는 DNA 단일 가닥의 시토신을 표적화할 수 있다. 일정 예의 구현예에서, 시티딘 디아미나제는 결합 성분, 예를 들어 결합된 Cas13의 외부에 존재하는 단일 가닥 상에서 편집할 수 있다. 다른 예의 구현예에서, 시티딘 디아미나제는 국재된 버블, 예컨대 표적 편집 부위이지만 가이드 서열에서 미스매치에 의해 형성된 국재된 버블에서 편집할 수 있다. 일정 예의 구현예에서 시티딘 디아미나제는 [Kim et al., Nature Biotechnology (2017) 35(4):371-377 (doi:10.1038/nbt.3803]에 개시된 것과 같은 국소 활성을 돕는 돌연변이를 함유할 수 있다.

[0656] 일부 구현예에서, 시티딘 디아미나제는 포유동물, 조류, 개구리, 오징어, 어류, 파리 및 벌레를 포함하지만, 제한없이, 1종 이상의 후생동물 종으로부터 유래된다. 일부 구현예에서, 시티딘 디아미나제는 인간, 영장류, 소, 개, 래트, 또는 마우스 시티딘 디아미나제이다.

[0657] 일부 구현예에서, 시티딘 디아미나제는 hAPOBEC1 또는 hAPOBEC3을 포함한 인간 POBEC이다. 일부 구현예에서, 시티딘 디아미나제는 인간 AID이다.

[0658] 일부 구현예에서, 시티딘 디아미나제 단백질은 RNA 듀플렉스의 단일 가닥 버블의 하나 이상의 표적 시토신 잔기(들)를 인식하여 우라실 잔기(들)로 전환시킨다. 일부 구현예에서, 시티딘 디아미나제 단백질은 RNA 듀플렉스의 단일 가닥 버블 상의 결합창을 인식한다. 일부 구현예에서, 결합창은 적어도 하나의 표적 시토신 잔기(들)를 함유한다. 일부 구현예에서, 결합창은 약 3 bp 내지 약 100 bp 범위이다. 일부 구현예에서, 결합창은 약 5 bp 내지 약 50 bp 범위이다. 일부 구현예에서, 결합창은 약 10 bp 내지 약 30 bp 범위이다. 일부 구현예에서, 결합창은 약 1 bp, 2 bp, 3 bp, 5 bp, 7 bp, 10 bp, 15 bp, 20 bp, 25 bp, 30 bp, 40 bp, 45 bp, 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 85 bp, 90 bp, 95 bp, 또는 100 bp 이다.

[0659] 일부 구현예에서, 시티딘 디아미나제 단백질은 하나 이상의 디아미나제 도메인을 포함한다. 특정 이론에 의

해 구속되는 것으로 의도하지는 않지만, 디아미나제 도메인은 RNA 듀플렉스 단일-가닥 버블에 함유된 하나 이상의 표적 시토신 (C) 잔기(들)를 인식하고, 이를 우라실 (U) 잔기(들)로 전환시키는 작용을 하는 것으로 고려된다. 일부 구현예에서, 디아미나제 도메인은 활성 중심을 포함한다. 일부 구현예에서, 활성 중심은 아연 이온을 포함한다. 일부 구현예에서, 활성 중심 내 또는 근처의 아미노산 잔기는 표적 시토신 잔기에 대해 하나 이상의 뉴클레오티드(들) 5'과 상호작용한다. 일부 구현예에서, 활성 중심 내 또는 근처의 아미노산 잔기는 표적 시토신 잔기에 대해 하나 이상의 뉴클레오티드(들) 3'과 상호작용한다.

[0660] 일부 실시형태에서, 아데노신 디아미나제는 인간 APOBEC1 완전 단백질(hAPOBEC1) 또는 이의 디아미나제 도메인(hAPOBEC1-D) 또는 C-말단 절두된 버전 (hAPOBEC-T)을 포함한다. 일부 구현예에서, 시티딘 디아미나제는 hAPOBEC1, hAPOBEC-D 또는 hAPOBEC-T에 상동성인 APOBEC 패밀리 멤버이다. 일부 실시형태에서, 시티딘 디아미나제는 인간 AID1 완전 단백질(hAID) 또는 이의 디아미나제 도메인(hAID-D) 또는 이의 C-말단 절두 형태(hAID-T)를 포함한다. 일부 구현예에서, 시티딘 디아미나제는 hAID, hAID-D 또는 hAID-T에 상동성인 AID 패밀리 멤버이다. 일부 구현예에서, hAID-T는 약 20개 아미노산이 C-말단 절두된 hAID이다.

[0661] 일부 구현예에서, 시티딘 디아미나제는 시토신 디아미나제의 야생형 아미노산 서열을 포함한다. 일부 구현예에서, 시티딘 디아미나제는 시토신 디아미나제의 편집 효율, 및/또는 기질 편집 선호도가 특정 필요에 따라 변화되도록, 시토신 디아미나제 서열에 하나 이상의 돌연변이를 포함한다.

[0662] APOBEC1 및 APOBEC3 단백질의 일정 돌연변이는 [Kim et al., Nature Biotechnology (2017) 35(4):371-377 (doi:10.1038/nbt.3803)]; 및 [Harris et al. Mol. Cell (2002) 10:1247-1253]에 기술되어 있고, 이들 각각은 그 전문이 참조로 본 명세서에 편입된다.

[0663] 일구 구현예에서, 시티딘 디아미나제는 인간 APOBEC3G에서 W285, R313, D316, D317X, R320, 또는 R326에 상응하는 아미노산 위치에 하나 이상의 돌연변이를 포함하는 래트 APOBEC1, 또는 APOBEC3G 디아미나제의 W90, R118, H121, H122, R126, 또는 R132에 상응하는 아미노산 위치에 하나 이상의 돌연변이를 포함하는 APOBEC1 디아미나제이다.

[0664] 일부 구현예에서, 시티딘 디아미나제는 래트 APOBEC1 아미노산 서열의 트립토판90, 또는 상동성 APOBEC 단백질 내 상응하는 위치, 예컨대 APOBEC3G의 트립토판 285에 돌연변이를 포함한다. 일부 구현예에서, 위치 90의 트립토판 잔기는 티로신 또는 페닐알라닌 잔기 (W90Y 또는 W90F)로 치환된다.

[0665] 일부 구현예에서, 시티딘 디아미나제는 래트 APOBEC1 아미노산 서열의 아르기닌 118, 또는 상동성 APOBEC 단백질의 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 118의 아르기닌 잔기는 알라닌 잔기로 치환된다 (R118A).

[0666] 일부 구현예에서, 시티딘 디아미나제는 래트 APOBEC1 아미노산 서열의 히스티딘121, 또는 상동성 APOBEC 단백질의 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 121의 히스티딘 잔기는 아르기닌 잔기로 치환된다 (H121R).

[0667] 일부 구현예에서, 시티딘 디아미나제는 래트 APOBEC1 아미노산 서열의 히스티딘122, 또는 상동성 APOBEC 단백질의 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 122의 히스티딘 잔기는 아르기닌 잔기로 치환된다 (H122R).

[0668] 일부 구현예에서, 시티딘 디아미나제는 래트 APOBEC1 아미노산 서열의 아르기닌126, 또는 상동성 APOBEC 단백질의 상응하는 위치, 예컨대 APOBEC3G의 아르기닌320에 돌연변이를 포함한다. 일부 구현예에서, 위치 126의 아르기닌 잔기는 알라닌 잔기 (R126A) 또는 글루탐산 (R126E)으로 치환된다.

[0669] 일부 구현예에서, 시티딘 디아미나제는 APOBEC1 아미노산 서열의 아르기닌132, 또는 상동성 APOBEC 단백질의 상응하는 위치에 돌연변이를 포함한다. 일부 구현예에서, 위치 132의 아르기닌 잔기는 글루탐산 잔기로 치환된다 (R132E).

[0670] 일부 구현예에서, 편집창의 너비를 좁히기 위해서, 시티딘 디아미나제는 래트 APOBEC1의 아미노산 서열 위치를 기반으로, 돌연변이: W90Y, W90F, R126E 및 R132E, 및 상기에 상응하는 상동성 APOBEC 단백질의 돌연변이 중 하나 이상을 포함할 수 있다.

[0671] 일부 구현예에서, 편집 효율을 감소시키기 위해서, 시티딘 디아미나제는 래트 APOBEC1의 아미노산 서열 위치를 기반으로, 돌연변이: W90A, R118A, R132E, 및 상기에 상응하는 상동성 APOBEC 단백질의 돌연변이 중 하나 이상을 포함할 수 있다. 특정한 구현예에서, 오프-표적 효과를 감소시키기 위해서 효율이 감소된 시티딘 디아미

나제 효소를 사용하는 것이 흥미로울 수 있다.

[0672] 일부 구현예에서, 시티딘 디아미나제는 야생형 래트 APOBEC1 (rAPOBEC1, 또는 이의 촉매 도메인이다. 일부 구현예에서, rAPOBEC1의 편집 효율, 및/또는 기질 편집 선호도가 특정 필요에 따라 변화되도록, 아데노신 디아미나제는 rAPOBEC1 서열에서 하나 이상의 돌연변이를 포함한다.

[0673] rAPOBEC1:

[0674] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO:433)

[0675] 일부 구현예에서, 시티딘 디아미나제는 야생형 인간 APOBEC1 (hAPOBEC1), 또는 이의 촉매 도메인이다. 일부 구현예에서, hAPOBEC1의 편집 효율, 및/또는 기질 편집 선호도가 특정 필요에 따라 변화되도록, 시티딘 디아미나제는 hAPOBEC1 서열에서 하나 이상의 돌연변이를 포함한다.

[0676] APOBEC1:

[0677] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKN TTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYV ARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQY

[0678] PPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLI HPSVAWR (SEQ ID NO:434)

[0679] 일부 구현예에서, 시티딘 디아미나제는 야생형 인간 APOBEC3G (hAPOBEC3G), 또는 이의 촉매 도메인이다. 일부 구현예에서, hAPOBEC3G의 편집 효율, 및/또는 기질 편집 선호도가 특정 필요에 따라 변화되도록, 시티딘 디아미나제는 hAPOBEC3G 서열에서 하나 이상의 돌연변이를 포함한다.

[0680] hAPOBEC3G:

[0681] MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKV TLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRE LFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERM HNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFT SWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYS EFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO:435)

[0682] 일부 구현예에서, 시티딘 디아미나제는 야생형 페트로마이존 마리너스 (Petromyzon marinus) CDA1 (pmCDA1), 또는 이의 촉매 도메인이다. 일부 구현예에서, pmCDA1의 편집 효율, 및/또는 기질 편집 선호도가 특정 필요에 따라 변화되도록, 시티딘 디아미나제는 pmCDA1 서열에서 하나 이상의 돌연변이를 포함한다.

[0683] pmCDA1:

[0684] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNK PQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRG NGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHN QLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO:436)

[0685] 일부 구현예에서, 시티딘 디아미나제는 야생형 인간 AID (hAID), 또는 이의 촉매 도메인이다. 일부 구현예에서, pmCDA1의 편집 효율, 및/또는 기질 편집 선호도가 특정 필요에 따라 변화되도록, 시티딘 디아미나제는 pmCDA1 서열에서 하나 이상의 돌연변이를 포함한다.

[0686] hAID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPYLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN
SVRLSRQLRRILLPLYEVDDLRDAFRTLGLLD (SEQ ID NO:437)

[0687]

[0688] 일부 구현예에서, 시티딘 디아미나제는 hAID의 절두 형태 (hAID-DC), 또는 이의 촉매 도메인이다. 일부 구현예에서, hADAR2-D의 편집 효율, 및/또는 기질 편집 선호도가 특정 필요에 따라 변화되도록, 아데노신 디아미나제는 hADAR2-D 서열에서 하나 이상의 돌연변이를 포함한다.

[0689] hAID-DC:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC
HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR
LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN
SVRLSRQLRRILL (SEQ ID NO:438)

[0690]

[0691] 시티딘 디아미나제의 추가 구현예는 그 전문이 참조로 본 명세서에 편입되는 발명의 명칭: "Nucleobase Editor and Uses Thereof" 의 WO2017/070632에 개시된다.

[0692] 일부 구현예에서, 시티딘 디아미나제는 탈아미노화 편집에 감수성인 뉴클레오티드를 포함하는 효율적인 탈아미노화 창을 갖는다. 따라서, 일부 구현예에서, "편집 창 폭" 은 시티딘 디아미나제의 편집 효율이 표적 부위에 대하여 반수-최대 값을 초과하는 주어진 표적 부위에서의 뉴클레오티드 위치의 수를 말한다. 일부 구현예에서, 시티딘 디아미나제는 약 1 내지 약 6 개의 범위의 뉴클레오티드의 편집 창 폭을 갖는다. 일부 구현예에서, 시티딘 디아미나제의 편집 창 폭은 1, 2, 3, 4, 5 또는 6 개 뉴클레오티드이다.

[0693] 이론에 국한하려는 의도는 아니나, 일부 구현예에서, 링커 서열의 길이가 편집 창 폭에 영향을 미치는 것으로 고려된다. 일부 구현예에서, 편집 창 폭은 링커 길이가 연장되면서 (예를 들어, 약 3 내지 약 21개 아미노산), 증가된다 (예를 들어, 약 3 내지 약 6개 뉴클레오티드). 비제한적인 예에서, 16-잔기 링커는 약 5개 뉴클레오티드의 효율적인 탈아미노화 창을 제공한다. 일부 구현예에서, 가이드 RNA의 길이는 편집 창 폭에 영향을 미친다. 일부 구현예에서, 가이드 RNA의 단축은 시티딘 디아미나제의 좁은 효율적인 탈아미노화 창을 야기한다.

[0694] 일부 구현예에서, 시티딘 디아미나제에 대한 돌연변이는 편집 창 폭에 영향을 미친다. 일부 구현예에서, CD-기능화된 CRISPR 시스템의 시티딘 디아미나제 성분은 디아미나제가 DNA 결합 사건 당 다수 시티딘의 탈아미노화를 방지하도록, 시티딘 디아미나제의 촉매 효율을 삼소시키는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, APOBEC1의 잔기 90에서의 트립토판(W90) 또는 상동성 서열에서의 상응하는 트립토판 잔기가 돌연변이된다. 일부 구현예에서, 촉매적 불활성화 Cas13은 W90Y 또는 W90F 돌연변이를 포함하는 APOBEC1 돌연변이체에 융합 또는 연결된다. 일부 구현예에서, APOBEC3G의 잔기 285의 트립토판(W285) 또는 상동성 서열에서의 상응하는 트립토판 잔기가 돌연변이된다. 일부 구현예에서, 촉매적 불활성화 Cas13은 W285Y 또는 W285F 돌연변이를 포함하는 APOBEC3G 돌연변이체에 융합 또는 연결된다.

[0695] 일부 구현예에서, Cd-기능화된 CRISPR 시스템의 시티딘 디아미나제 성분은 디아미나제 활성 부위에 시티딘의 비최적 제시에 대한 내성을 감소시키는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, 시티딘 디아미나제는 디아미나제 활성 부위의 기질 결합 활성을 변경시키는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, 시티딘 디아미나제는 인식되고, 디아미나제 활성 부위에 의해 결합될 DNA의 입체형태를 변경시키는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, 시티딘 디아미나제는 디아미나제 활성 부위에 대한 기질 접근 가능성을 변경시키는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, APOBEC1의 잔기 126의 아르기닌(R126) 또는 상동성 서열에서의 상응하는 아르기닌 잔기가 돌연변이된다. 일부 구현예에서, 촉매적 불활성화 Cas13은 R126A 또는 R126E 돌연변이를 포함하는 APOBEC1 돌연변이체에 융합 또는 연결된다. 일부 구현예에서, APOBEC3G의 잔기 320의 트립토판(R320) 또는 상동성 서열에서의 상응하는 아르기닌 잔기가 돌연변이된다. 일부 구현예에서, 촉매적 불활성화 Cas13은 R320A 또는 R320E 돌연변이를 포함하는 APOBEC3G 돌연변이체에 융합 또는 연결된다. 일부 구현예에서, APOBEC1의 잔기 132의 아르기닌(R132) 또는 상동성 서열에서의

상응하는 아르기닌 잔기가 돌연변이된다. 일부 구현예에서, 촉매적 불활성화 Cas13은 R132E 돌연변이를 포함하는 APOBEC1 돌연변이체에 융합 또는 연결된다.

[0696] 일부 구현예에서, CD-기능화된 CRISPR 시스템의 APOBEC1 도메인은 W90Y, W90F, R126A, R126E, 및 R132E로부터 선택된 하나, 둘, 또는 3개 돌연변이를 포함한다. 일부 구현예에서, APOBEC1 도메인은 W90Y 및 R126E의 이중 돌연변이를 포함한다. 일부 구현예에서, APOBEC1 도메인은 W90Y 및 R132E의 이중 돌연변이를 포함한다. 일부 구현예에서, APOBEC1 도메인은 R126E 및 R132E의 이중 돌연변이를 포함한다. 일부 구현예에서, APOBEC1 도메인은 W90Y, R126E 및 R132E의 3개의 돌연변이를 포함한다.

[0697] 일부 구현예에서, 본원에 개시된 바와 같은 시티딘 디아미나제의 하나 이상의 돌연변이는 편집 창 폭을 약 2 개의 뉴클레오티드로 감소시킨다. 일부 구현예에서, 본원에 개시된 바와 같은 시티딘 디아미나제의 하나 이상의 돌연변이는 편집 창 폭을 약 1 개의 뉴클레오티드로 감소시킨다. 일부 구현예에서, 본원에 개시된 바와 같은 시티딘 디아미나제의 하나 이상의 돌연변이는 효소의 편집 효율에 오직 최소로 또는 중등으로 영향을 미치면서, 편집 창 폭을 감소시킨다. 일부 구현예에서, 본원에 개시된 바와 같은 시티딘 디아미나제의 하나 이상의 돌연변이는 효소의 편집 효율을 감소시키지 않고, 편집 창 폭을 감소시킨다. 일부 구현예에서, 본원에 개시된 바와 같은 시티딘 디아미나제의 하나 이상의 돌연변이는 다르게 시티딘 디아미나제에 의하여 유사한 효율로 편집될 이웃 시티딘 뉴클레오티드의 식별을 가능하게 한다.

[0698] 일부 구현예에서, 아데노신 디아미나제 단백질은 이중 가닥 핵산 기질을 인식하고 결합하기 위해서 하나 이상의 이중 가닥 RNA (dsRNA) 결합 모티프 (dsRBM) 또는 도메인 (dsRBD)를 더 포함하거나 또는 그에 연결된다. 일부 구현예에서, 시티딘 디아미나제 및 기질 간 상호작용은 CRISPR/CAS 단백질 인자를 포함하여, 하나 이상의 추가 단백질 인자(들)에 의해 매개된다. 일부 구현예에서, 시티딘 디아미나제 및 기질 간 상호작용은 가이드 RNA를 포함하여, 하나 이상의 핵산 성분(들)에 의해 더욱 매개된다.

[0699] 본 발명에 따라서, 시티딘 디아미나제의 기질은 CRISPR-Cas 효소와 CRISPR-Cas 복합체를 형성하게 되는 이의 DNA 표적과 가이드 분자의 결합 시 시티딘 디아미나제에 접근가능하게 만든, 관심 시토신을 포함하는 RNA 듀플렉스의 DNA 단일 가닥 버블이고, 그리하여 시토신 디아미나제는 CRISPR-Cas 복합체의 하나 이상의 성분, 즉 CRISPR-Cas 효소 및/또는 가이드 분자에 융합되거나 또는 결합될 수 있다. 가이드 분자 및 CRISPR-Cas 효소의 특별한 특색은 아래에 상세히 기재되어 있다.

[0700] **염기 편집 가이드 분자 디자인 고려사항**

[0701] 일부 구현예에서, 가이드 서열은 10 내지 50 nt 길이, 보다 특히 약 20-30 nt, 유리하게 약 20 nt, 23-25 nt 또는 24 nt의 RNA 서열이다. 염기 편집 구현예에서, 가이드 서열은 탈아미노화되는 아데노신을 포함하는 표적 서열과의 하이브리드화를 보장하도록 선택된다. 이것은 하기에 보다 상술된다. 선택은 탈아미노화의 효율 및 특이성을 증가시키는 단계를 더 포괄한다.

[0702] 일부 구현예에서, 가이드 서열은 약 20 nt 내지 약 30 nt 길이이고 표적 DNA 가닥과 하이브리드화하여 표적 아데노신 부위에서 dA-C 미스매치를 갖는 것을 제외하고는, 거의 완벽하게 일치되는 듀플렉스를 형성한다. 특히, 일부 구현예에서, dA-C 미스매치는 표적 서열의 중심 (따라서 표적 서열과 가이드 서열의 하이브리드화 시 듀플렉스의 중심에 가깝게 위치되어서, 아데노신 디아미나제를 협소한 편집창 (예를 들어, 약 4 bp 너비)으로 제한시킨다. 일부 구현예에서, 표적 서열은 탈아미노화하려는 하나 초과의 표적 아데노신을 포함할 수 있다. 추가 구현예에서 표적 서열은 표적 아데노신 부위의 3'에 하나 이상의 dA-C 미스매치를 더 포함할 수 있다. 일부 구현예에서, 표적 서열 내 의도하지 않은 아데닌 부위에서 오프-표적 편집을 피하기 위해서, 가이드 서열은 일정 아데노신 디아미나제 예컨대 ADAR1 및 ADAR2에 대해 촉매적으로 비호의적인, dA-G 미스매치를 도입시키기 위해 상기 의도하지 않은 아데닌에 상응하는 위치에 비-쌍형성 구아닌을 포함하도록 디자인될 수 있다. 본 명세서에 그 전문이 참조로 편입되는 [Wong et al., RNA 7:846858 (2001)]을 참조한다.

[0703] 일부 구현예에서, 정규 길이를 갖는 Cas12b 가이드 서열(예를 들어, AsCpf1에 대해 약 20 nt)을 사용하여 표적 DNA를 갖는 헤테로듀플렉스을 형성한다. 일부 구현예에서, 정규 길이 보다 긴 Cas12b 가이드 분자(예를 들어, AsCpf1에 대해 >20 nt)를 사용하여 Cas12b-가이드 RNA-표적 DNA 복합체 바깥을 포함하는 표적 DNA와 헤테로듀플렉스을 형성한다. 이것은 소정 스트레치의 뉴클레오티드 내 하나 초과의 아데닌의 탈아미노화에 관심있을 때 특히 흥미로울 수 있다. 대안적인 구현예에서, 정규 가이드 서열 길이의 제한성을 유지시키는 것이 흥미롭다. 일부 구현예에서, 가이드 서열은 Cas12b 가이드의 정규 길이 외부에 dA-C 미스매치를 도입시키도록 디자인되어서, 이것은 아데노신 디아미나제 및 dA-C 미스매치 간 접촉 빈도를 증가시키고 Cas13에 의한 입체

방해를 감소시킬 수 있다.

[0704] 일부 구현예에서, 미스매치된 핵염기(예를 들어, 시티딘)의 위치는 PAM이 DNA 표적 상에 있는 경우에 계산된다. 일부 구현예에서, 미스매치된 핵염기는 PAM으로부터 12 내지 21nt, 또는 PAM으로부터 13 내지 21nt, 또는 PAM으로부터 14 내지 21nt, 또는 PAM으로부터 14 내지 20nt, 또는 PAM으로부터 15 내지 20nt, 또는 PAM으로부터 16 내지 20nt, 또는 PAM으로부터 14 내지 19nt, 또는 PAM으로부터 15 내지 19nt, 또는 PAM으로부터 16 내지 19nt, 또는 PAM으로부터 17 내지 19nt, 또는 PAM으로부터 약 20nt, 또는 PAM으로부터 약 19nt, 또는 PAM으로부터 약 18nt, 또는 PAM으로부터 약 17nt, 또는 PAM으로부터 약 16nt, 또는 PAM으로부터 약 15nt, 또는 PAM으로부터 약 14nt에 위치된다. 바람직한 구현예에서, 미스매치된 핵염기는 PAM으로부터 17 내지 19nt 또는 18nt에 위치된다.

[0705] 미스매치 거리는 Cas12b 스페이서의 3' 말단과 미스매치된 핵염기(예를 들어, 시티딘) 사이의 염기 수이되, 미스매치된 염기는 미스매치 거리 계산의 부분으로서 포함된다. 일부 구현예에서, 미스매치 거리는 1 내지 10nt, 또는 1 내지 9nt, 또는 1 내지 8nt, 또는 2 내지 8nt, 또는 2 내지 7nt, 또는 2 내지 6nt, 또는 3 내지 8nt, 또는 3 내지 7nt, 또는 3 내지 6nt, 또는 3 내지 5nt, 또는 약 2nt, 또는 약 3nt, 또는 약 4nt, 또는 약 5nt, 또는 약 6nt, 또는 약 7nt, 또는 약 8 nt이다. 바람직한 구현예에서, 미스매치 거리는 3 내지 5 nt 또는 4 nt이다.

[0706] 일부 구현예에서, 본 명세서에 기재된 Cas12b-ADAR 시스템의 편집창은 PAM으로부터 12 내지 21nt, 또는 PAM으로부터 13 내지 21nt, 또는 PAM으로부터 14 내지 21nt, 또는 PAM으로부터 14 내지 20nt, 또는 PAM으로부터 15 내지 20nt, 또는 PAM으로부터 16 내지 20nt, 또는 PAM으로부터 14 내지 19nt, 또는 PAM으로부터 15 내지 19nt, 또는 PAM으로부터 16 내지 19nt, 또는 PAM으로부터 17 내지 19nt, 또는 PAM으로부터 약 20nt, 또는 PAM으로부터 약 19nt, 또는 PAM으로부터 약 18nt, 또는 PAM으로부터 약 17nt, 또는 PAM으로부터 약 16nt, 또는 PAM으로부터 약 15nt, 또는 PAM으로부터 약 14 nt이다. 일부 구현예에서, 본 명세서에 기재된 Cas12b-ADAR 시스템의 편집창은 Cas12b 스페이서의 3' 말단부로부터 1 내지 10nt, 또는 Cas12b 스페이서의 3' 말단부로부터 1 내지 9nt, 또는 Cas12b 스페이서의 3' 말단부로부터 1 내지 8nt, 또는 Cpf1 스페이서의 3' 말단부로부터 2 내지 8nt, 또는 Cas12b 스페이서의 3' 말단부로부터 2 내지 7nt, 또는 Cas12b 스페이서의 3' 말단부로부터 2 내지 6nt, 또는 Cas12b 스페이서의 3' 말단부로부터 3 내지 8nt, 또는 Cas12b 스페이서의 3' 말단부로부터 3 내지 7nt, 또는 Cas12b 스페이서의 3' 말단부로부터 3 내지 6nt, 또는Cas12b 스페이서의 3' 말단부로부터 3 내지 5nt, 또는 Cas12b 스페이서의 3' 말단부로부터 약 2nt, 또는 Cas12b 스페이서의 3' 말단부로부터 약 3nt, 또는 Cas12b 스페이서의 3' 말단부로부터 약 4nt, 또는 Cas12b 스페이서의 3' 말단부로부터 약 5nt, 또는 Cas12b 스페이서의 3' 말단부로부터 약 6nt, 또는 Cas12b 스페이서의 3' 말단부로부터 약 7nt, 또는 Cas12b 스페이서의 3' 말단부로부터 약 8 nt이다.

[0707] **벡터**

[0708] 일반적으로, 그리고 본원에서, 용어 "벡터" 는 그것이 연결된 다른 핵산을 수송할 수 있는 핵산 분자를 지칭한다. 이는 레플리콘, 예컨대 플라스미드, 파지, 또는 코스미드로, 또 다른 DNA 절편이 그 안에 삽입될 수 있어서, 삽입된 절편의 복제를 유발한다. 일반적으로, 벡터는 적절한 제어 엘리먼트와 관련되었을 때 복제가 가능하다.

[0709] 일부 구현예에서, 본 개시는 CRISPR-Cas 시스템의 하나 이상의 성분을 코딩하는 하나 이상의 폴리뉴클레오티드를 포함하는 벡터 시스템을 제공한다. 일부 구현예에서, 벡터 시스템은 표 1 또는 2로부터의 Cas12b 이펙터 단백질을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제1 조절 엘리먼트, 및 i) a) crRNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트, 및 b) tracr RNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제3 조절 엘리먼트, 또는 ii) 가이드 서열 및 tracr RNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함하는 하나 이상의 벡터를 포함하는, Cas12b 벡터 시스템이다. 일부 경우에서, 벡터 시스템은 단일 벡터를 포함한다. 대안적으로, 벡터 시스템은 복수 벡터를 포함한다. 벡터(들)는 바이러스 벡터(들)일 수 있다.

[0710] 벡터는 제한 없이, 단일 가닥, 이중 가닥 또는 부분 이중 가닥인 핵산 분자; 하나 이상의 자유 말단을 포함하는, 자유 말단을 포함하지 않는 (예 : 환형) 핵산 분자; DNA, RNA 또는 둘 다를 포함하는 핵산 분자; 및 당업계에 공지된 다른 종류의 폴리뉴클레오티드를 포함한다. 하나의 유형의 벡터는 "플라스미드"이며, 이는 추가의 DNA 절편이 예를 들어, 표준 분자 클로닝 기술에 의해 삽입될 수 있는 환형 이중 가닥 DNA 루프를 지칭한다. 벡터의 다른 종류는 바이러스 벡터로서, 바이러스-유래된 DNA 또는 RNA 서열이 바이러스에 봉입되는

벡터에 존재한다(예를 들어, 레트로바이러스, 복제 결함 레트로바이러스, 아데노바이러스, 복제 결함 아데노바이러스, 및 아데노-관련 바이러스(AAV)). 바이러스 벡터는 또한 숙주 세포 내로의 트랜스펙션을 위한, 바이러스가 보유하는 폴리뉴클레오티드를 포함한다. 특정 벡터는 그것이 도입된 숙주 세포에서 자율적 복제가 가능하다(예를 들어, 박테리아 복제 기원을 가진 박테리아 벡터 및 에피솜 포유류 벡터). 다른 벡터(예를 들어, 비-에피솜 포유동물 벡터)는 숙주 세포 내로 도입시 숙주 세포의 게놈에 통합되며, 이에 의해 숙주 게놈과 함께 복제된다. 더욱이, 특정 벡터는 그것이 작동적으로 연결된 유전자의 발현을 지시할 수 있다. 이러한 벡터는 본원에서 " 발현 벡터" 로 지칭된다. 진핵생물 세포에서의 발현을 위한 벡터 및 이를 초래하는 벡터는 본 명세서에서 "진핵세포 발현 벡터" 로 지칭될 수 있다. 재조합 DNA 기술에 유용한 통상적인 발현 벡터는 종종 플라스미드의 형태로 존재한다.

[0711] 재조합 발현 벡터는 숙주 세포에서 핵산의 발현에 적합한 형태로 본 발명의 핵산을 포함할 수 있는데, 이는 재조합 발현 벡터가 하나 이상의 조절 엘리먼트를 포함하는 것을 의미하며, 하나 이상의 조절 엘리먼트는 발현에 사용될 숙주 세포에 기반하여 선택될 수 있고, 발현될 핵산 서열에 작동적으로 연결된다. 재조합 발현 벡터 내에서, "작동적으로 연결된"은 대상 뉴클레오티드 서열이 (예를 들어, 시험관내 전사/번역 시스템 내에서, 또는 벡터가 숙주 세포 내로 도입되는 경우 숙주 세포 내에서) 뉴클레오티드 서열의 발현을 가능하게 하는 방식으로 조절 엘리먼트(들)에 연결된 것을 의미하는 의도이다. 유리한 벡터는 렌티바이러스 및 아데노-연합 바이러스를 포함하고, 이러한 벡터의 유형은 또한 특정 세포 유형을 표적화하도록 선택된다.

[0712] 재조합 및 클로닝 방법에 관해, 미국 특허 제20040171156 A1호로서 2004년 9월 2일자로 공개된 미국 특허 출원 제10/815,730호가 언급되며, 이의 내용은 본 명세서에 그들의 전문이 참고로 편입된다.

[0713] 용어 "조절 엘리먼트" 는 프로모터, 인핸서, 내부 리보솜 진입 부위 (IRES) 및 다른 발현 제어 엘리먼트 (예를 들어, 전사 종결 신호, 예컨대, 폴리아데닐화 신호 및 폴리-U 서열)를 포함하도록 의도된다. 이와 같은 조절 엘리먼트는, 예를 들어 문헌 [Goeddel, GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)]에 기재되어 있다. 조절 엘리먼트는 많은 유형의 숙주 세포에서 뉴클레오티드 서열의 구성적 발현을 지시하는 것들과 특정 숙주 세포에서만 뉴클레오티드 서열의 발현을 지시하는 것들(예를 들어, 조직-특이적 조절 서열)을 포함한다. 조직-특이적 프로모터는 근육, 뉴런, 뼈, 피부, 혈액, 특정 장기(예를 들어, 간, 췌장)와 같은 요망되는 관심 조직, 또는 특정 세포 유형(예를 들어, 림프구)에서 주로 발현을 유도할 수 있다. 조절 엘리먼트는 시간-의존적 방식으로, 예컨대 세포-주기 의존적 또는 발생 단계 의존적 방식으로 발현을 지시할 수 있으며, 이것은 조직 특이적이거나 세포-유형 특이적일 수 있거나, 또한 그렇지 않을 수 있다. 일부 구현예에서, 벡터는 하나 이상의 pol III 프로모터 (예를 들어, 1, 2, 3, 4, 5개 이상의 pol III 프로모터), 하나 이상의 pol II 프로모터 (예를 들어, 1, 2, 3, 4, 5개 이상의 pol II 프로모터), 하나 이상의 pol I 프로모터 (예를 들어, 1, 2, 3, 4, 5개 이상의 pol I 프로모터), 또는 이들의 조합을 포함한다. pol III 프로모터의 예는 U6 및 H1 프로모터를 포함하지만, 이들로 제한되지 않는다. pol II 프로모터의 예는 레트로바이러스 라우스 육종 바이러스 (RSV) LTR 프로모터(임의로 RSV 인핸서를 가짐), 거대세포바이러스 (CMV) 프로모터 (임의로 CMV 인핸서를 가짐)[예를 들어, 문헌 [Boshart et al, Cell, 41:521530 (1985)] 참조], SV40 프로모터, 디히드로폴레이트 리덕타제 프로모터, β-액틴 프로모터, 포스포글리세롤 키나제 (PGK) 프로모터 및 EF1α 프로모터를 포함하지만, 이들로 제한되지 않는다. 또한 용어 "조절 엘리먼트" 는 인핸서 엘리먼트, 예컨대 WPRE; CMV 인핸서; HTLV-I의 LTR의 R-U5' 절편 (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 인핸서; 및 토끼β-글로빈의 엑손 2 및 3 사이의 인트론 서열 (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981)을 포괄한다. 발현 벡터의 디자인은 형질전환시키려는 숙주 세포의 선택, 바람직한 발현 수준 등과 같은 인자에 의존할 수 있다는 것은 당업자가 이해하게 될 것이다. 벡터는 숙주 세포로 도입될 수 있어서, 본 명세서에 기술된 바와 같은 핵산에 의해 코딩되는, 융합 단백질 또는 펩티드를 포함하여, 전사물, 단백질, 또는 펩티드 (예를 들어, CRISPR 전사물, 단백질, 효소, 이의 돌연변이체 형태, 이의 융합 단백질 등)를 생산할 수 있다. 조절 서열에 관해, 미국 특허 출원 제10/491,026호가 언급되며, 이의 내용은 본 명세서에 그의 전문이 참고로 편입된다. 프로모터에 관해, 국제 특허 출원 WO 2011/028929 및 미국 특허 출원 제12/511,940호가 언급되며, 이의 내용은 본 명세서에 그들의 전문이 참고로 편입된다.

[0714] 유리한 벡터는 렌티바이러스 및 아데노-연합 바이러스를 포함하고, 이러한 벡터의 유형은 또한 특정 세포 유형을 표적화하도록 선택된다.

[0715] 특정 구현예에서, 가이드 RNA 및 (임의로 변형 또는 돌연변이된) CRISPR 효소 (예를 들어, C2c1)를 위한 바이시스트론 벡터가 이용된다. 가이드 RNA 및 (임의로 변형 또는 돌연변이된) CRISPR 효소에 대한 바이시스트론 발현 벡터가 바람직하다. 일반적으로 특히 이 구현예에서, (임의로 변형 또는 돌연변이된) CRISPR 효소는

바람직하게 CBh 프로모터에 의해서 구동된다. RNA는 바람직하게는 Pol III 프로모터, 예컨대, U6 프로모터에 의해 유도될 수 있다. 이상적으로는 2개가 조합된다.

[0716] 벡터는 원핵 또는 진핵생물 세포에서 CRISPR 전사물 (예를 들어, 핵산 전사물, 단백질 또는 효소)의 발현을 위해 디자인될 수 있다. 예를 들어, CRISPR 전사물은 박테리아 세포, 예컨대, 에스케리치아 콜라이, 곤충 세포 (배큘로바이러스 발현 벡터를 이용), 효모 세포 또는 포유동물 세포에서 발현될 수 있다. 적합한 숙주 세포는 [Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)] 에 추가로 논의된다. 대안적으로, 재조합 발현 벡터는, 예를 들어 T7 프로모터 조절 서열 및 T7 중합효소를 이용하여 시험관내에서 전사 및 번역될 수 있다.

[0717] 벡터는 원핵생물 또는 원핵 세포에 도입되고 증식될 수 있다. 일부 구현예에서, 원핵생물은 진핵생물 세포 내로 도입될 벡터의 복제물을 증식시키기 위해 또는 진핵생물 세포 내로 도입될 벡터 생성에서의 중간 벡터로서 사용된다 (예를 들어, 바이러스 벡터 패키징 시스템의 부분으로서 플라스미드를 증폭시킴). 일부 구현예에서, 원핵생물은 벡터의 복제물을 증폭시키기 위해 그리고, 예컨대, 숙주 세포 또는 숙주 유기체에 전달을 위한 하나 이상의 단백질 공급원을 제공하기 위해 하나 이상의 핵산을 발현시키기 위해 사용된다. 원핵생물 내 단백질의 발현은 융합 또는 비융합 단백질 중 하나의 발현을 지시하는 구성적 또는 유도성 프로모터를 함유하는 벡터로 에스케리치아 콜라이에서 가장 흔히 수행된다. 융합 벡터는 그 안에서 코딩되는 단백질에, 예컨대, 재조합 단백질의 아미노 말단에 다수의 아미노산을 부가한다. 이러한 융합 벡터는 하나 이상의 목적, 예컨대: (i) 재조합 단백질 발현을 증가시키고; (ii) 재조합 단백질의 용해도를 증가시키고; (iii) 친화도 정제에서 리간드로서 작용함으로써 재조합 단백질의 정제에 도움을 주는 작용을 할 수 있다. 종종, 융합 발현 벡터에서, 단백질 절단 부위는 융합 모이어티 및 재조합 단백질의 접합부에 도입되어 융합 단백질의 후속 정제를 위해 융합 모이어티로부터 재조합 단백질의 분리를 가능하게 한다. 이러한 효소, 및 그들의 동족 인식 서열은 인자 Xa, 트롬빈 및 엔테로키나제를 포함한다. 예시적 융합 발현 벡터는 각각 글루타티온 S-트랜스퍼라제 (GST), 말토스 E 결합 단백질 또는 단백질 A를 표적 재조합 단백질에 융합시키는, pGEX(Pharmacia Biotech Inc; Smith and Johnson, 1988. Gene 67: 31- 40), pMAL (New England Biolabs, Beverly, Mass.) 및 pRIT5 (Pharmacia, Piscataway, N.J.)를 포함한다. 적합한 유도성 비융합 이. 콜라이 발현 벡터의 예는 pTrc (Amrann et al., (1988) Gene 69:301- 315) 및 pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60- 89)를 포함한다. 일부 구현예에서, 벡터는 효모 발현 벡터이다. 효모 사카로마이세스 세레비지아에 (Saccharomyces cerivisae)에서 발현을 위한 벡터의 예는 pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), 및 picZ (InVitrogen Corp, San Diego, Calif.)를 포함한다.

일부 구현예에서, 벡터는 배큘로바이러스 발현 벡터를 이용하여 곤충 세포에서 단백질 발현을 유도한다. 배양된 곤충 세포 (예를 들어, SF9 세포)에서 단백질의 발현에 이용가능한 배큘로바이러스 벡터는 pAc 시리즈 (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) 및 pVL 시리즈 (Lucklow and Summers, 1989. Virology 170: 31-39)를 포함한다.

[0718] 일부 구현예에서, 벡터는 포유동물 발현 벡터를 이용하여 포유동물 세포에서 하나 이상의 서열의 발현을 유도할 수 있다. 포유동물 발현 벡터의 예는 pCDM8 (Seed, 1987. Nature 329: 840) 및 pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195)를 포함한다. 포유동물 세포에서 사용될 때, 발현 벡터의 제어 기능은 전형적으로 하나 이상의 조절 엘리먼트에 의해 제공된다. 예를 들어, 통상적으로 사용되는 프로모터는 폴리오마바이러스, 아데노바이러스 2, 거대세포바이러스, 유인원 바이러스 40, 및 본 명세서에 개시되고 당업계에 공지된 다른 것으로부터 유래된다. 원핵생물 및 진핵생물 세포 둘 모두를 위한 다른 적절한 발현 시스템에 대하여, 예를 들어, 문헌[Sambrook, et al., MOLECULAR CLONING:A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989]의 16 및 17장을 참조한다.

[0719] 일부 구현예에서, 재조합 포유동물 발현 벡터는 특정 세포 유형에서 핵산의 발현을 우선적으로 유도할 수 있다 (예를 들어, 조직-특이적 조절 엘리먼트는 핵산을 발현시키기 위해 사용된다). 조직-특이적 조절 엘리먼트는 당업계에 공지되어 있다. 적절한 조직-특이적 프로모터의 비제한적인 예는 알부민 프로모터 (간-특이적; Pinkert, et al., 1987. Genes Dev. 1: 268-277), 림프구-특이적 프로모터 (Calame and Eaton, 1988. Adv. Immunol. 43: 235-275), 특히 T 세포 수용체 (Winoto and Baltimore, 1989. EMBO J. 8: 729-733) 및 면역글로불린 (Baneiji, et al., 1983. Cell 33: 729-740; Queen and Baltimore, 1983. Cell 33: 741-748)의

프로모터, 뉴론-특이적 프로모터 (예를 들어, 뉴로필라멘트 프로모터; Byrne and Ruddle, 1989. Proc. Natl. Acad. Sci. USA 86: 5473-5477), 췌장-특이적 프로모터 (Edlund, et al., 1985. Science 230: 912-916), 및 유선-특이적 프로모터 (예를 들어, 밀크 웨이 프로모터; U.S. 특허 출원 번호 4,873,316 및 유럽 공개 특허 출원 번호 264,166)를 포함한다. 발생-조절 프로모터, 예를 들어 쥣과 hox 프로모터 (Kessel and Gruss, 1990. Science 249: 374-379) 및 α-페토단백질 프로모터 (Campes and Tilghman, 1989. Genes Dev. 3: 537-546)가 포괄된다. 이들 원핵 및 진핵 벡터에 대해, 미국 특허 제6,750,059호가 언급되며, 이의 내용은 본 명세서에 그들의 전문이 참고로 편입된다. 본 발명의 다른 구현예는 바이러스 벡터의 용도에 관한 것일 수 있으며, 미국 특허 출원 제13/092,085호가 언급되고, 이의 내용은 본 명세서에 전문이 참고로 편입된다. 조직-특이적 조절 엘리먼트는 당업계에 공지되어 있고, 이와 관련하여, 미국 특허 제7,776,321호가 언급되며, 이의 내용은 본 명세서에서 그들의 전문이 참고로 편입된다. 일부 구현예에서, 조절 엘리먼트는 CRISPR 시스템의 하나 이상의 요소의 발현을 유도하기 위해 CRISPR 시스템의 하나 이상의 요소에 작동적으로 연결된다.

[0720] 핵산-표적화 시스템의 하나 이상의 구성요소의 발현을 구동하는 하나 이상의 벡터는 핵산-표적화 시스템의 구성요소의 발현이 하나 이상의 표적 부위에서 핵산-표적화 복합체의 형성을 유도하도록 숙주 세포에 도입된다. 예를 들어, 핵산-표적화 이펙터 효소 및 핵산-표적화 가이드 RNA 및/또는 tracr은 별개의 벡터 상에서 별개의 조절 엘리먼트에 각각 작동적으로 연결될 수 있다. 핵산-표적화 시스템의 RNA(들)는 유전자이식 핵산-표적화 이펙터 단백질 동물 또는 포유동물, 예를 들어, 핵산-표적화 이펙터 단백질을 구성적으로 또는 유도적으로 또는 조건적으로 발현시키는 동물 또는 포유동물; 또는 핵산-표적화 이펙터 단백질을 달리 발현시키거나 또는 핵산-표적화 이펙터 단백질을 함유하는 세포를 갖는 동물 또는 포유동물에, 예컨대, 생체내 핵산-표적화 이펙터 단백질을 코딩하고 이를 발현시키는 벡터 또는 벡터들의 사전 투여에 의해 전달될 수 있다. 대안적으로, 동일하거나 또는 상이한 조절 엘리먼트로부터 발현되는 둘 이상의 엘리먼트는 단일 벡터에 조합될 수 있고, 하나 이상의 추가 벡터가 제1 벡터에 포함되지 않은 핵산-표적화 시스템의 임의 성분을 제공한다. 단일 벡터에 조합되는 핵산-표적화 시스템은 임의의 적합한 배향으로 정렬될 수 있는데, 예컨대 한 엘리먼트는 제2 엘리먼트에 대해 5' ("상류") 또는 3' ("하류")에 위치된다. 한 구성요소의 코딩 서열은 제2 구성요소의 코딩 서열의 동일하거나 또는 반대쪽 가닥 상에 위치되고, 동일 또는 반대 방향으로 배향될 수 있다. 일부 구현예에서, 단일 프로모터는 하나 이상의 인트론 서열 내에 (예를 들어, 상이한 인트론에 각각, 적어도 하나의 인트론에 2 이상, 또는 단일 인트론에 모두) 함입된 핵산-표적화 이펙터 단백질 및 핵산-표적화 가이드 RNA를 코딩하는 전사체의 발현을 유도한다. 일부 구현예에서, 핵산-표적화 이펙터 단백질 및 핵산-표적화 가이드 RNA는 동일한 프로모터에 작동적으로 연결되고 동일한 프로모터로부터 발현될 수 있다. 핵산-표적화 시스템의 하나 이상의 구성요소의 발현을 위한 전달 비히클, 벡터, 입자, 나노입자, 제형 및 이들의 성분은 앞서 언급한 문헌, 예컨대, WO 2014/093622 (PCT/US2013/074667)에서 사용되는 바와 같다. 일부 구현예에서, 벡터는 하나 이상의 삽입 부위, 예컨대, 제한 엔도뉴클레아제 인식 서열 (또한 "클로닝 부위" 라고도 함)을 포함한다. 일부 구현예에서, 하나 이상의 삽입 부위 (예를 들어, 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 삽입 부위)가 하나 이상의 벡터의 하나 이상의 서열 구성요소의 상류 및/또는 하류에 위치된다. 다중의 상이한 가이드 서열이 사용될 때, 세포 내에서 다중의 상이한, 상응하는 표적 서열에 대해 핵산-표적화 활성을 표적화하기 위해 단일 발현 구성체가 사용될 수 있다. 예를 들어, 단일 벡터는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20개 이상의 가이드 서열을 포함할 수 있다. 일부 구현예에서, 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 이러한 가이드-서열-함유 벡터가 제공될 수 있으며, 선택적으로 세포에 전달된다. 일부 구현예에서, 벡터는 핵산-표적화 이펙터 단백질을 코딩하는 효소-코딩 서열에 작동적으로 연결된 조절 엘리먼트를 포함한다.

핵산-표적화 이펙터 단백질 또는 핵산-표적화 가이드 RNA 또는 RNA(들)은 개별적으로 전달될 수 있고, 유리하게 이들 중 적어도 하나는 입자 복합체를 통해 전달된다. RNA-표적화 이펙터 단백질 mRNA는 핵산-표적화 이펙터 단백질이 발현되는 시간을 제공하기 위하여 RNA-표적화 가이드 RNA 이전에 전달될 수 있다. 핵산-표적화 이펙터 단백질 mRNA는 핵산-표적화 가이드 RNA의 투여 전에 1 내지 12시간 (바람직하게는 대략 2 내지 6시간)에 투여될 수 있다. 대안적으로, 핵산-표적화 이펙터 단백질 mRNA 및 핵산-표적화 가이드 RNA는 함께 투여될 수 있다. 유사하게는, 가이드 RNA의 제2 부스터 용량은 핵산-표적화 이펙터 단백질 mRNA + 가이드 RNA의 초기 투여 후 1 내지 12시간 (바람직하게는 대략 2 내지 6시간)에 투여될 수 있다. 핵산-표적화 이펙터 단백질 mRNA 및/또는 가이드 RNA의 추가 투여는 게놈 변형의 가장 효율적인 수준을 달성하는 데 유용할 수 있다.

[0721] 일부 구현예에서, 벡터는 하나 이상의 핵 국재화 서열(NLS), 예를 들어, 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 개 이상의 NLS를 포함하는 CRISPR 효소를 코딩한다. 보다 특히, 벡터는 C2c1 이펙터 단백질에 천연적으로 존재하지 않는 하나 이상의 NLS를 포함한다. 가장 특히, NLS는 C2c1 이펙터 단백질 서열의 벡터 5' 및/또는 3'

에 존재한다. 일부 구현예에서, RNA-표적화 이펙터 단백질은 아미노-말단 또는 그 근처에 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 개 이상, 또는 이를 초과하는 NLS, 카르복시-말단 또는 그 근처에 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 개 이상, 또는 이를 초과하는 NLS, 또는 이들의 조합(예를 들어 아미노-말단에서 0 개 또는 적어도 1 개 이상의 NLS, 그리고 카르복시 말단에서 0 개 또는 적어도 1 개 이상의 NLS)을 포함한다. 1 개 초과의 NLS가 존재하는 경우, 각각은 단일의 NLS가 1 개 초과의 복제물로 존재하고/존재하거나 1 개 이상의 복제물로 존재하는 1 개 이상의 기타 다른 NLS와 조합될 수 있도록, 다른 것들로부터 독립적으로 선택될 수 있다. 일부 구현예에서, NLS의 가장 가까운 아미노산이 N-말단 또는 C-말단으로부터 폴리펩티드를 따라서 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50개 이상의 아미노산 내에 있을 때, NLS는 N-말단 또는 C-말단 근처에서 고려된다. NLS의 비제한적 예는 하기로부터 유래된 NLS 서열을 포함한다: 아미노산 서열 PKKKRKV (SEQ ID No. 462)을 갖는, SV40 바이러스 거대 T-항원의 NLS; 뉴클레오플라스민으로부터의 NLS(예를 들어, 서열 KRPAATKKAGQAKKKK (SEQ ID NO:463)을 갖는 뉴클레오플라스민 이분형 NLS); 아미노산 서열 PAAKRVKLD (SEQ ID NO:464) 또는 RQRRNELKRSP (SEQ ID NO:465)을 갖는 c-myc NLS; 서열 NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:466)을 갖는 hRNPA1 M9 NLS; 임포틴-알파 유래 IBB 도메인의 서열 RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV(SEQ ID NO:467); 근종 T 단백질의 서열 VSRKRPRP (SEQ ID NO:468) 및 PPKKARED (SEQ ID NO:469); 인간 p53의 서열 PQPKKKPL (SEQ ID NO:470); 마우스 c-abl IV의 서열 SALIKKKKKMAP (SEQ ID NO:471); 인플루엔자 바이러스 NS1의 서열 DRLRR (SEQ ID NO:472) 및 PKQKKRK (SEQ ID NO:473); 간염 바이러스 델타 항원의 서열 RKLKKKIKKL (SEQ ID NO:474); 마우스 Mx1 단백질의 서열 REKKKFLKRR (SEQ ID NO:475); 인간 폴리(ADP-리보스) 중합효소의 서열 KRKGDEVDGVDEVAKKKSKK (SEQ ID NO:476); 및 스테로이드 호르몬 수용체(인간) 글루코코티코이드의 서열 RKCLQAGMNLEARKTKK (SEQ ID NO:477). 일반적으로, 하나 이상의 NLS는 진핵생물 세포의 핵에서 검출 가능한 양의 DNA/RNA-표적화 Cas 단백질의 축적을 유도하기에 충분한 세기의 것이다. 일반적으로, 핵 국재화 활성의 세기는 핵산-표적화 이펙터 단백질 내 NLS의 수, 사용되는 특정 NLS(들), 또는 이들 인자의 조합으로부터 유래할 수 있다. 핵에서의 축적의 검출은 임의의 적합한 기법에 의해 실행될 수 있다. 예를 들어, 검출 가능한 마커는, 세포 내의 위치가 시각화될 수 있도록, 예컨대, 핵 위치를 검출하기 위한 수단 (예를 들어, DAPI와 같은 핵에 특이적인 염색)과 조합하여, 핵산-표적화 단백질에 융합될 수 있다. 세포핵은 또한 세포로부터 단리될 수 있으며, 이어서, 이의 함량은 단백질을 검출하기 위한 임의의 적합한 공정, 예컨대, 면역조직화학, 웨스턴 블롯, 또는 효소 활성 분석에 의해 분석될 수 있다. 핵 내 축적은 또한 간접적으로, 예컨대 핵산-표적화 복합체 형성의 효과에 대한 어세이 (예를 들어, 표적 세포에서 DNA 또는 RNA 절단 또는 돌연변이를 위한 어세이, 또는 DNA 또는 RNA-표적화 복합체 형성 및/또는 DNA 또는 RNA-표적화 Cas 단백질 활성에 의해 영향받는 변경된 유전자 발현 활성을 위한 어세이)를 통해서, 핵산-표적화 Cas 단백질 또는 핵산-표적화 복합체에 노출되지 않았거나, 또는 하나 이상의 NLS가 결여된 핵산-표적화 Cas 단백질에 노출된 대조군과 비교하여 결정할 수 있다. 본 명세서에 기재된 C2c1 이펙터 단백질 복합체의 바람직한 구현예에서, 코돈 최적화된 이펙터 단백질은 단백질의 C-말단에 부착된 NLS를 포함한다. 일정 구현예에서, 다른 국재화 태그는 예컨대 제한없이 Cas를 세포의 특정 부위, 예컨대 세포내 소기관, 예컨대 미토콘드리아, 색소체, 엽록체, 소포, 골지 (핵 또는 세포) 막, 리보솜, 핵소체, ER, 세포골격, 액포, 센트로솜, 뉴클레오솜, 과립, 중심립 등에 국재화시키기 위해 Cas 단백질에 융합될 수 있다.

[0722] 본 발명은 또한 치료의 치료적 방법에서 사용을 위한 비천연 발생 또는 조작된 조성물, 또는 상기 조성물의 성분을 코딩하는 하나 이상의 폴리뉴클레오티드, 또는 상기 조성물의 성분을 코딩하는 하나 이상의 폴리뉴클레오티드를 포함하는 벡터 시스템을 제공한다. 치료적 처리 방법은 유전자 또는 게놈 편집, 또는 유전자 요법을 포함할 수 있다.

[0723] 일부 구현예에서, 치료의 치료적 방법은 표적 유기체의 개체군에서 요법 또는 치료제를 기반으로 설계된 가이드 서열을 포함하는 CRISPR-Cas 시스템을 포함한다. 일부 구현예에서, 표적 유기체 개체군은 적어도 1000 개체, 예컨대 적어도 5000 개체, 예컨대 적어도 10000 개체, 예컨대 적어도 50000 개체를 포함한다. 일부 구현예에서, 개체곤 전반에서 최소 서열 변이를 갖는 표적 부위는 개체군의 적어도 99%, 바람직하게 적어도 99.9%, 보다 바람직하게 적어도 99.99%에서 서열 변이의 부재를 특징으로 한다.

[0724] 본 명세서에서 사용되는 용어 일배체형 (반수체 유전자형)은 단일 부모로부터 함께 유전된 유기체에서의 유전자 그룹이다. 본 명세서에서 사용되는 일배체형 빈도 추정 ("단계화" 라고도 함)은 유전자형 데이터로부터 일배체형의 통계 추정 과정을 의미한다. Toshikazu 등 (Am J Hum Genet. 2003 Feb; 72(2): 384-398)은 개시된 본 발명에서 사용할 수 있는 일배체형 빈도의 추정 방법을 기술한다.

[0725] 본 명세서에 기재된 핵산-표적화 시스템, 벡터 시스템, 벡터 및 조성물은 다양한 핵산-표적화 응용, 유전자 산

물, 예컨대 단백질의 합성 변경 또는 변형, 핵산 절단, 핵산 편집, 핵산 스플라이싱; 표적 핵산의 트래피킹, 표적 핵산의 추적, 표적 핵산의 단리, 표적 핵산의 시각화 등에 사용될 수 있다.

[0726] 일반적으로, 본 명세서 전반에서, 용어 "벡터" 는 그것이 연결된 다른 핵산을 수송할 수 있는 핵산 분자를 지칭한다. 벡터는 제한 없이, 단일 가닥, 이중 가닥 또는 부분 이중 가닥인 핵산 분자; 하나 이상의 자유 말단을 포함하는, 자유 말단을 포함하지 않는 (예 : 환형) 핵산 분자; DNA, RNA 또는 둘 다를 포함하는 핵산 분자; 및 당업계에 공지된 다른 종류의 폴리뉴클레오티드를 포함한다. 하나의 유형의 벡터는 "플라스미드"이며, 이는 추가의 DNA 절편이 예를 들어, 표준 분자 클로닝 기술에 의해 삽입될 수 있는 환형 이중 가닥 DNA 루프를 지칭한다. 벡터의 다른 종류는 바이러스 벡터로서, 바이러스-유래된 DNA 또는 RNA 서열이 바이러스에 봉입되는 벡터에 존재한다(예를 들어, 레트로바이러스, 복제 결함 레트로바이러스, 아데노바이러스, 복제 결함 아데노바이러스, 및 아데노-관련 바이러스(AAV)). 바이러스 벡터는 또한 숙주 세포 내로의 형질감염을 위해 바이러스가 운반하는 폴리뉴클레오티드를 포함한다. 특정 벡터는 그것이 도입된 숙주 세포에서 자율적 복제가 가능하다(예를 들어, 박테리아 복제 기원을 가진 박테리아 벡터 및 에피솜 포유류 벡터). 다른 벡터(예를 들어, 비-에피솜 포유동물 벡터)는 숙주 세포 내로 도입시 숙주 세포의 게놈에 통합되며, 이에 의해 숙주 게놈과 함께 복제된다. 더욱이, 특정 벡터는 그것이 작동적으로 연결된 유전자의 발현을 지시할 수 있다. 이러한 벡터는 본원에서 "발현 벡터" 로 지칭된다. 진핵생물 세포에서의 발현을 위한 벡터 및 이를 초래하는 벡터는 본 명세서에서 "진핵생물 발현 벡터" 로 지칭될 수 있다. 재조합 DNA 기술에 유용한 통상적인 발현 벡터는 종종 플라스미드의 형태로 존재한다.

[0727] 일정 구현예에서, 벡터 시스템은 역 방향으로 프로모터-가이드 발현 카세트를 포함한다.

[0728] 재조합 발현 벡터는 숙주 세포에서 핵산의 발현에 적합한 형태로 본 발명의 핵산을 포함할 수 있는데, 이는 재조합 발현 벡터가 하나 이상의 조절 엘리먼트를 포함하는 것을 의미하며, 하나 이상의 조절 엘리먼트는 발현에 사용될 숙주 세포에 기반하여 선택될 수 있고, 발현될 핵산 서열에 작동적으로 연결된다.

[0729] 유리한 벡터는 렌티바이러스 및 아데노-연합 바이러스를 포함하고, 이러한 벡터의 유형은 또한 특정 세포 유형을 표적화하도록 선택된다.

[0730] 핵산-표적화 시스템의 하나 이상의 구성요소의 발현을 구동하는 하나 이상의 벡터는 핵산-표적화 시스템의 구성요소의 발현이 하나 이상의 표적 부위에서 핵산-표적화 복합체의 형성을 유도하도록 숙주 세포에 도입된다. 예를 들어, 핵산-표적화 이펙터 모듈 및 핵산-표적화 가이드 RNA는 별개의 벡터 상에서 별개의 조절 엘리먼트에 각각 작동적으로 연결될 수 있다. 핵산-표적화 시스템의 RNA(들)는 유전자이식 핵산-표적화 이펙터 모듈 동물 또는 포유동물, 예를 들어, 핵산-표적화 이펙터 단백질을 구성적으로 또는 유도적으로 또는 조건적으로 발현시키는 동물 또는 포유동물; 또는 핵산-표적화 이펙터 모듈을 달리 발현시키거나 또는 핵산-표적화 이펙터 모듈을 함유하는 세포를 갖는 동물 또는 포유동물에, 예컨대, 생체내 핵산-표적화 이펙터 모듈을 코딩하고 이를 발현시키는 벡터 또는 벡터들의 사전 투여에 의해 전달될 수 있다. 대안적으로, 동일하거나 또는 상이한 조절 엘리먼트로부터 발현되는 둘 이상의 엘리먼트는 단일 벡터에 조합될 수 있고, 하나 이상의 추가 벡터가 제1 벡터에 포함되지 않은 핵산-표적화 시스템의 임의 성분을 제공한다. 단일 벡터에 조합되는 핵산-표적화 시스템은 임의의 적합한 배향으로 정렬될 수 있는데, 예컨대 한 엘리먼트는 제2 엘리먼트에 대해 5' ("상류") 또는 3' ("하류")에 위치된다. 한 구성요소의 코딩 서열은 제2 구성요소의 코딩 서열의 동일하거나 또는 반대쪽 가닥 상에 위치되고, 동일 또는 반대 방향으로 배향될 수 있다. 일부 구현예에서, 단일 프로모터는 하나 이상의 인트론 서열 (예를 들어, 각각 상이한 인트론, 적어도 한 인트론에 둘 이상, 또는 단일 인트론에 전부)에 내포된, 핵산-표적화 이펙터 모듈 및 핵산-표적화 가이드 RNA를 코딩하는 전사물의 발현을 구동한다. 일부 구현예에서, 핵산-표적화 이펙터 모듈 및 핵산-표적화 가이드 RNA는 동일한 프로모터에 작동적으로 연결되고 동일한 프로모터로부터 발현될 수 있다.

[0731] 본 발명은 또한 다수의 핵산 성분을 전달하기 위한 방법을 포함하며, 여기서 각각의 핵산 성분은 상이한 대상 표적 유전자좌에 특이적이고, 이에 의해 다수의 대상 표적 유전자좌를 변형시킨다. 복합체의 핵산 성분은 하나 이상의 단백질-결합 RNA 압타머를 포함할 수 있다. 하나 이상의 압타머는 박테리오파지 외피 단백질에 결합할 수 있다. 박테리오파지 외피 단백질은 Qβ, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ΦCb5, ΦCb8r, ΦCb12r, ΦCb23r, 7s 및 PRR1을 포함하는 군으로부터 선택될 수 있다. 바람직한 구현예에서, 박테리오파지 외피 단백질은 MS2 이다.
본 발명은 또한 30 개 이상, 40 개 이상 또는 50 개 이상 뉴클레오티드 길이의 복합체의 핵산 성분을 제공한다.

[0732] 일 양상에서, 본 발명은 하나 이상의 벡터를 포함하는 벡터 시스템을 제공하고, 하나 이상의 벡터는 a) 본 명세서에 정의된 바와 같은 조작된 CRISPR 단백질을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제1 조절 엘리먼트; 및 임의로 b) 가이드 서열, 직접 반복부 서열을 포함하는 가이드 RNA를 포함하는 하나 이상의 핵산 분자를 코딩하는 하나 이상의 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함하고, 임의로 성분 (a) 및 (b)는 동일하거나 또는 상이한 벡터에 존재한다.

[0733] 본 발명은 또한 a) 본 발명의 구성체 중 어느 하나의 비천연 발생 CRISPR 효소를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제1 조절 엘리먼트; 및 b) 하나 이상의 가이드 RNA로서, 가이드 서열을 포함하는 것인 가이드 서열, 직접 반복부 서열을 코딩하는 하나 이상의 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함하는 하나 이상의 벡터를 포함하는 조작된, 비천연 발생 CRISPR (Clustered Regularly Interspersed Short Palindromic Repeats)-CRISPR 연관 (Cas 이펙터 모듈) (CRISPR-Cas 이펙터 모듈) 벡터 시스템을 제공하고, 성분 (a) 및 (b)은 동일하거나 또는 상이한 벡터 상에 위치되고, CRISPR 복합체가 형성되며; 가이드 RNA는 표적 폴리뉴클레오티드 유전자좌를 표저고하하고 효소는 폴리뉴클레오티드 유전자좌를 변경시키고, CRISPR 복합체의 효소는 비변형된 효소와 비교하여 하나 이상의 오프-표적 유전자좌를 변경시키는 능력이 감소되고/되거나 그리하여 CRISPR 복합체의 효소는 비변형된 효소와 비교하여 하나 이상의 표적 유전자좌를 변형시키는 능력이 증가되었다.

[0734] 본 명세서에서 사용되는 CRISPR Cas 이펙터 모듈 또는 CRISPR 이펙터 모듈은 C2c1을 포함하지만, 이에 제한되지 않는다. 일부 구현예에서, CRISPR-Cas 이펙터 모듈은 조작될 수 있다.

[0735] 이러한 시스템에서, 성분 (I)은 가이드 서열, 직접 반복부 서열을 포함하는 폴리뉴클레오티드 서열에 작동적으로 연결된 제1 조절 엘리먼트를 포함할 수 있고, 성분 (II)는 CRISPR 효소를 코딩하는 폴리뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함할 수 있다. 이러한 시스템에서, 적용가능한 경우에 가이드 RNA는 키메라 RNA를 포함할 수 있다.

[0736] 이러한 시스템에서, 성분 (I)은 가이드 서열 및 직접 반복부 서열에 작동적으로 연결된 제1 조절 엘리먼트를 포함할 수 있고, 성분 (II)은 CRISPR 효소를 코딩하는 폴리뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함할 수 있다. 이러한 시스템은 하나 초과의 가이드 RNA를 포함할 수 있고, 각각의 가이드 RNA는 상이한 표적을 가져서 다중복합화가 존재한다. 성분 (a) 및 (b)는 동일 벡터 상에 있을 수 있다.

[0737] 벡터를 포함하는 임의의 이러한 시스템에서, 하나 이상의 벡터는 하나 이상의 바이러스 벡터, 예컨대 하나 이상의 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스, 또는 헤르페스 심플렉스 바이러스를 포함할 수 있다.

[0738] 조절 엘리먼트를 포함하는 임의의 이러한 시스템에서, 상기 조절 엘리먼트 중 적어도 하나는 조직-특이적 프로모터를 포함할 수 있다. 조직-특이적 프로모터는 포유동물 혈액 세포, 포유동물 간 세포, 또는 포유동물 눈에서 발현을 유도할 수 있다.

[0739] 임의의 상기 기술된 조성물 또는 시스템에서, 직접 반복부 서열은 하나 이상의 단백질-상호작용 RNA 압타머를 포함할 수 있다. 하나 이상의 압타머는 테트라루프에 위치될 수 있다. 하나 이상의 압타머는 MS2 박테리오파지 외피 단백질에 결합할 수 있다.

[0740] 임의의 상기 기술된 조성물 또는 시스템에서 세포는 진핵생물 세포 또는 원핵생물 세포일 수 있고, CRISPR 복합체는 세포에서 작동가능하고, 그리하여 CRISPR 복합체의 효소는 비변형된 효소와 비교하여 세포의 하나 이상의 오프-표적 유전자좌를 변형시키는 능력이 감소되었고/되었거나 그리하여 CRISPR 복합체의 효소는 비변형된 효소와 비교하여 하나 이상의 표적 유전자좌를 변형시키는 능력이 증가되었다.

[0741] 본 발명은 또한 임의의 상기 기술된 조성물 또는 임의의 상기 기술된 시스템의 CRISPR 복합체를 제공한다.

[0742] 본 발명은 또한 임의의 본 명세서에 기술된 조작된 CRISPR 효소 (예를 들어, 조작된 Cas 이펙터 모듈), 조성물, 또는 임의의 본 명세서에 기술된 시스템 또는 벡터 시스템과 세포를 접촉시키는 단계를 포함하는 세포에서 관심 유전자좌를 변형시키는 방법을 제공하거나, 또는 세포는 세포 내에 존재하는 임의의 본 명세서에 기술된 CRISPR 복합체를 포함한다. 이러한 방법에서 세포는 원핵생물 또는 진핵생물 세포, 바람직하게 진핵생물 세포일 수 있다. 이러한 방법에서, 유기체는 세포를 포함할 수 있다. 이러한 방법에서, 유기체는 인간 또는 다른 동물이 아닐 수 있다.

[0743] 일정 구현예에서, 본 발명은 또한 비천연 발생, 조작된 조성물 (예를 들어, AAV 벡터에 맞춰질 수 있는 C2c1 또

는 임의의 Cas 단백질)을 제공한다. 또한 사용할 수 있는 다른 단백질에 대한 목록 및 지침을 제공하기 위해서 참조로 본 명세서에 편입된 US 8,697,359의 도 19A, 19B, 19C, 19D, 및 20A-F를 참조한다.

[0744] 임의의 이러한 방법은 생체외 또는 시험관내일 수 있다.

[0745] 일정 구현예에서, 적어도 하나의 상기 가이드 RNA 또는 C2c1 이펙터 모듈을 코딩하는 뉴클레오티드 서열은 관심 유전자의 프로모터를 포함하는 조절 엘리먼트와 세포에서 작동적으로 연결되어서, 적어도 하나의 CRISPR-Cas 이펙터 모듈 시스템 성분의 발현이 관심 유전자의 프로모터에 의해 구동된다. "작동적으로 연결된" 은 가이드 RNA 및/또는 Cas 이펙터 모듈을 코딩하는 뉴클레오티드 서열이 역시 본 명세서의 다른 곳에 언급된 바와 같이, 뉴클레오티드 서열의 발현을 허용하는 방식으로 조절 엘리먼트(들)에 연결된 것을 의미하고자 한다. 용어 "조절 엘리먼트" 는 본 명세서의 다른 곳에서 기술된다. 본 발명에 따라서, 조절 엘리먼트느 관심 유전자의 프로모터, 예컨대 바람직하게 관심 내생성 유전자의 프로모터를 포함한다. 일정 구현예에서, 프로모터는 이의 내생성 게놈 위치에 있다. 이러한 구현예에서, CRISPR 및/또는 Cas를 코딩하는 핵산은 이의 천연 게놈 위치에서 관심 유전자의 프로모터의 전사 제어 하에 있다. 일정 다른 구현예에서, 프로모터는 (개별) 핵산 분자, 예컨대 벡터 또는 플라스미드, 또는 다른 염색체외 핵산 상에 제공되고, 다시 말해 프로모터는 이의 천연 게놈 위치에서 제공되지 않는다. 일정 구현예에서, 프로모터는 비천연 게놈 위치에서 게놈에 통합된다.

[0746] 본 발명은 또한 세포를 본 명세서에 기술된 조작된 CRISPR 효소 (예를 들어, 조작된 Cas 이펙터 모듈), 조성물, 시스템 또는 CRISPR 복합체와 접촉시켜서 CRISPR-Cas 이펙터 모듈 (벡터)를 전달하고 CRISPR-Cas 이펙터 모듈 복합체가 형성되어 표적에 결합할 수 있게 하는 단계, 게놈 유전자좌의 발현이 변경, 예컨대 증가 또는 감소된 발현, 또는 유전자 산물의 변경 여부를 결정하는 단계를 포함하는 포유동물 세포에서 관심 게놈 유전자좌의 발현을 변경시키는 방법을 제공한다.

[0747] 본 발명은 본 명세서에 기술된 바와 같은 본 발명에 따라서 Cas 이펙터 모듈 또는 CRISPR 효소의 오솔로그인 돌연변이 또는 변형된 Cas 이펙터 모듈에 돌연변이를 만드는 방법을 제공하고, 방법은 오솔로그가 가깝게 근접할 수 있거나 또는 핵산 분자, 예를 들어 DNA, RNA, gRNA 등을 접촉할 수 있는 아미노산(들), 및/또는 변형 및/또는 돌연변이를 위해서 본 명세서에 기술된 바와 같은 본 발명에 따른 CRISPR 효소의 본 명세서에서 확인된 아미노산(들)과 유사하거나 또는 상응하는 아미노산(들)을 확인하는 단계, 및 변형(들) 및/또는 돌연변이(들)를 포함하거나, 그로 이루어지거나 또는 그로 본질적으로 이루어진 오솔로그를 합성 또는 제조 발현시키거나 또는 본 명세서에 논의된 바와 같이, 중성 아미노산을 하전, 예를 들어 양으로 하전된 아미노산 예를 들어 알라닌으로 돌연변이, 예를 들어 변형, 예를 들어 변화 또는 돌연변이시키는 단계를 포함한다. 이렇게 변형된 오솔로그는 CRISPR-Cas 이펙터 모듈 시스템에서 사용할 수 있고; 이를 발현하는 핵산 분자(들)는 본 명세서에 논의된 바와 같은 CRISPR-Cas 이펙터 모듈 시스템 성분을 코딩하는 분자를 전달하는 벡터 시스템에서 사용할 수 있다.

[0748] 일 양상에서, 본 발명은 본 명세서에 기재된 성분 중 하나 이상을 포함하는 키트를 제공한다. 일부 구현예에서, 키트는 벡터 시스템 및 키트를 이용하기 위한 설명서를 포함한다. 일부 실시형태에서, 벡터 시스템은 (a) 직접 반복부 서열 및 DR 서열의 하류에 하나 이상의 가이드 서열을 삽입하기 위한 하나 이상의 삽입 부위에 작동적으로 연결된 제1 조절 엘리먼트로서, 발현될 때, 가이드 서열이 진핵생물 세포에서 표적 서열에 대한 Cas13 CRISPR 복합체의 서열-특이적 결합을 유도하고, CRISPR-Cas 이펙터 모듈 복합체는 (1) 표적 서열에 하이브리드화된 가이드 서열, (2) DR 서열, 및 (3) tracr 서열과 복합체를 형성한 Cas 이펙터 모듈을 포함하는 것인 제1 조절 엘리먼트; 및/또는 (b) 핵 국재화 서열을 포함하는 상기 Cas 이펙터 모듈을 코딩하는 효소-코딩 서열에에 작동적으로 연결된 제2 조절 엘리먼트를 포함하고, 유리하게 이것은 분할 Cas 이펙터 모듈을 포함한다.

일부 구현예에서, 키트는 시스템의 동일하거나 또는 상이한 벡터 상에 위치된 성분 (a) 및 (b)를 포함한다.

일부 구현예에서, 성분 (a)는 제1 조절 엘리먼트에 작동적으로 연결된 둘 이상의 가이드 서열을 더 포함하고, 발현될 때, 둘 이상의 가이드 서열 각각은 진핵생물 세포의 상이한 표적 서열과 CRISPR-Cas 이펙터 모듈 복합체의 서열 특이적 결합을 유도한다. tracr은 가이드 (스페이서) 및 직접 반복부 서열과 (코딩되는) 동일한 폴리뉴클레오티드 상에서 융합될 수 있거나 또는 융합되지 않을 수 있다.

[0749] 일 양상에서, 본 발명은 진핵생물 세포에서 표적 폴리뉴클레오티드를 변형시키는 방법을 제공한다. 일부 구현예에서, 방법은 CRISPR-Cas 이펙터 모듈 복합체가 표적 폴리뉴클레오티드에 결합할 수 있게 하여서 상기 폴리뉴클레오티드의 절단을 실시하여 표적 폴리뉴클레오티드를 변형시키는 단계를 포함하고, CRISPR-Cas 이펙터 모듈 복합체는 상기 표적 폴리뉴클레오티드 내 표적 서열에 하이브리드화된 가이드 서열과 복합체 형성하는 Cas 이펙터 모듈을 포함하고, 상기 가이드 서열은 직접 반복부 서열에 연결된다. 일부 구현예에서, 상기 절단은 상기 Cas 이펙터 모듈에 의해 표적 서열의 위치에서 1 또는 2개의 가닥을 절단하는 단계를 포함하고, 이것은 분

할 Cas 이펙터 모듈을 포함한다. 일부 구현예에서, 상기 절단은 표적 유전자의 감소된 전사를 초래한다. 일부 구현예에서, 상기 방법은 외생성 주형 폴리뉴클레오티드와의 상동성 재조합에 의해 상기 절단된 표적 폴리뉴클레오티드를 복구하는 단계를 더 포함하며, 상기 복구은 상기 표적 폴리뉴클레오티드의 하나 이상의 뉴클레오티드의 삽입, 결실 또는 치환을 포함하는 돌연변이를 야기한다. 일부 구현예에서, 상기 돌연변이는 표적 서열을 포함하는 유전자로부터 발현된 단백질에서 하나 이상의 아미노산 변화를 초래한다. 일부 구현예에서, 상기 방법은 하나 이상의 벡터를 상기 진핵생물 세포에 전달하는 단계를 추가로 포함하며, 여기서, 하나 이상의 벡터는 Cas 이펙터 모듈, 및 DR 서열에 연결된 가이드 서열 중 하나 이상의 발현을 구동한다. 일부 구현예에서, 상기 벡터는 대상체에서 진핵생물 세포에 전달된다. 일부 구현예에서, 상기 변형은 세포 배양물에서 상기 진핵생물 세포에서 일어난다. 일부 구현예에서, 상기 방법은 상기 변형 전에 대상체로부터 상기 진핵생물 세포를 단리시키는 단계를 추가로 포함한다. 일부 구현예에서, 방법은 상기 진핵생물 세포 및/또는 이로부터 유래된 세포를 상기 대상체에게 복귀시키는 단계를 더 포함한다. 일 양상에서, 본 발명은 진핵생물 세포에서 표적 폴리뉴클레오티드를 변형 또는 편집시키는 방법을 제공한다. 일부 구현예에서, 방법은 CRISPR-Cas 이펙터 모듈 복합체가 표적 폴리뉴클레오티드에 결합할 수 있게 하여서 DNA 염기 편집을 실시하는 단계를 포함하고, CRISPR-Cas 이펙터 모듈 복합체는 상기 표적 폴리뉴클레오티드 내 표적 서열에 하이브리드화된 가이드 서열과 복합체 형성하는 Cas 이펙터 모듈을 포함하고, 상기 가이드 서열은 직접 반복부 서열에 연결된다. 일부 구현예에서, Cas 이펙터 모듈은 촉매적으로 불활성인 CRISPR-Cas 단백질을 포함한다. 일부 구현예에서, 가이드 서열은 표적 서열과 가이드 서열 간에 형성된 DNA/RNA 헤테로듀플렉스에 하나 이상의 미스매치를 도입시키도록 설계된다. 특정 구현예에서, 미스매치는 A-C 미스매치이다. 일부 구현예에서, Cas 이펙터는 하나 이상의 기능성 도메인과 (예를 들어, 융합 단백질 또는 적합한 링커를 통해) 회합될 수 있다. 일부 구현예에서, 이펙터 도메인은 가수분해적 탈아미드화를 통해 내생성 편집을 매개하는 하나 이상의 시티딘 또는 아데노신 디아미나제를 포함한다.

[0750] 일 양상에서, 본 발명은 진핵생물 세포에서 폴리뉴클레오티드의 발현을 변형시키는 방법을 제공한다. 일부 구현예에서, 방법은 CRISPR-Cas 이펙터 모듈 복합체가 폴리뉴클레오티드와 결합될 수 있게 하여서 상기 결합이 상기 폴리뉴클레오티드의 증가되거나 또는 감소된 발현을 초래하게 하는 단계를 포함하고, CRISPR-Cas 이펙터 모듈 복합체는 상기 폴리뉴클레오티드 내의 표적 서열과 하이브리드화된 가이드 서열과 복합체 형성된 Cas 이펙터 모듈을 포함하고, 상기 가이드 서열은 분할 Cas 이펙터 모듈을 포함할 수 있는, 직접 반복부 서열에 연결된다. 일부 구현예에서, 상기 방법은 하나 이상의 벡터를 상기 진핵생물 세포에 전달하는 단계를 추가로 포함하며, 여기서, 하나 이상의 벡터는 DR 서열에 연결된 가이드 서열, 및 Cas 이펙터 모듈 중 하나 이상의 발현을 구동시킨다.

[0751] 일 양상에서, 본 발명은 진핵생물 세포에서 표적 전사물을 변형 또는 편집시키는 방법을 제공한다. 일부 구현예에서, 방법은 CRISPR-Cas 이펙터 모듈 복합체가 표적 폴리뉴클레오티드에 결합할 수 있게 하여서 RNA 염기 편집을 실시하는 단계를 포함하고, CRISPR-Cas 이펙터 모듈 복합체는 상기 표적 폴리뉴클레오티드 내 표적 서열에 하이브리드화된 가이드 서열과 복합체 형성하는 Cas 이펙터 모듈을 포함하고, 상기 가이드 서열은 직접 반복부 서열에 연결된다. 일부 구현예에서, Cas 이펙터 모듈은 촉매적으로 불활성인 CRISPR-Cas 단백질을 포함한다. 일부 구현예에서, 가이드 서열은 표적 서열과 가이드 서열 간에 형성된 RNA/RNA 듀플렉스에 하나 이상의 미스매치를 도입시키도록 설계된다. 특정 구현예에서, 미스매치는 A-C 미스매치이다. 일부 구현예에서, Cas 이펙터는 하나 이상의 기능성 도메인과 (예를 들어, 융합 단백질 또는 적합한 링커를 통해) 회합될 수 있다. 일부 구현예에서, 이펙터 도메인은 가수분해적 탈아미드화를 통해 내생성 편집을 매개하는 하나 이상의 시티딘 또는 아데노신 디아미나제를 포함한다. 특정 구현예에서, 이펙터 도메인은 효소의 RNA (ADAR) 패밀리에 작용하는 아데노신 디아미나제를 포함한다. 특정 구현예에서, 아데노신 디아미나제 단백질 또는 이의 촉매적 도메인은 RNA에서 아데노신 또는 시토신을 탈아미노화시킬 수 있거나 또는 RNA 특이적 아데노신 디아미나제이고/이거나 박테리아, 인간, 두족류 또는 초파리 아데노신 디아미나제 단백질 또는 이의 촉매적 도메인, 바람직하게는 TadA, 더 바람직하게는 ADAR, 임의로 huADAR, 임의로 (hu)ADAR1 또는 (hu)ADAR2, 바람직하게는 huADAR2 또는 이의 촉매적 도메인이다. 일부 구현예에서, 시티딘 디아미나제는 인간, 래트, 또는 램프리 시티딘 디아미나제이다. 일부 구현예에서, 시티딘 디아미나제는 아포리포단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 디아미나제, 활성화-유도된 디아미나제 (AID), 또는 시티딘 디아미나제 1 (CDA1)이다.

[0752] 본 출원은 관심 표적 DNA 서열을 변형시키는 것에 관한 것이다.

[0753] 본 발명의 추가적인 양상은 예방적 또는 치료적 치료에서 사용하기 위해 본 명세서에서 고려되는 바와 같은 방법 및 조성물에 관한 것으로서, 바람직하게 상기 관심 표적 유전자좌는 인간 또는 동물 내에 있는 것인 방법 및

조성물, 및 상기 표적 DNA에, 상기 본 명세서에 기술된 바와 같은 조성물을 전달하는 단계를 포함하는, 관심 표적 DNA 서열의 아데닌 또는 시티딘을 변형시키는 방법에 관한 것이다. 특정 구현예에서, CRISPR 시스템 및 아데노신 디아미나제, 또는 이의 촉매적 도메인은 하나 이상의 폴리뉴클레오티드 분자로서, 리보핵단백질 복합체로서, 임의로 입자, 소포, 또는 하나 이상의 바이러스 벡터를 통해 전달된다. 특정 구현예에서, 조성물은 병원성 G→A 또는 C→T 점 돌연변이를 함유하는 전사체에 의해 야기되는 질환의 치료 또는 예방에서 사용하기 위한 것이다. 특정 구현예에서, 따라서 본 발명은 요법에서 사용하기 위한 조성물을 포함한다. 이는 방법이 생체내, 생체외 또는 시험관내에서 수행될 수 있다는 것을 나타낸다. 특정 구현예에서, 상기 방법은 동물 또는 인간 신체의 치료 방법 또는 인간 세포의 생식 계열 유전적 동일성을 변형시키기 위한 방법이 아니다. 특정 구현예에서, 방법을 수행할 때, 표적 RNA는 인간 또는 동물 세포 내에 포함되지 않는다. 특정한 구현예에서, 표적이 인간 또는 동물 표적일 때, 방법은 생체외 또는 시험관내에서 수행된다.

[0754] 본 발명의 추가적인 양상은 예방적 또는 치료적 치료에서 사용하기 위해 본 명세서에서 예상되는 방법으로서, 바람직하게는 상기 관심 표적은 인간 또는 동물 내에 있는 것인 방법, 및 상기 표적 RNA에 본 명세서에서 상기 기재된 바와 같은 조성물을 전달하는 단계를 포함하는, 관심 표적 DNA 서열의 아데닌 또는 시티딘을 변형시키는 방법에 관한 것이다. 특정 구현예에서, CRISPR 시스템 및 아데노신 디아미나제, 또는 이의 촉매적 도메인은 하나 이상의 폴리뉴클레오티드 분자로서, 리보핵단백질 복합체로서, 임의로 입자, 소포, 또는 하나 이상의 바이러스 벡터를 통해 전달된다. 특정 구현예에서, 조성물은 병원성 G→A 또는 C→T 점 돌연변이를 함유하는 전사체에 의해 야기되는 질환의 치료 또는 예방에서 사용하기 위한 것이다. 특정 구현예에서, 따라서 본 발명은 요법에서 사용하기 위한 조성물을 포함한다. 이는 방법이 생체내, 생체외 또는 시험관내에서 수행될 수 있다는 것을 나타낸다. 특정 구현예에서, 상기 방법은 동물 또는 인간 신체의 치료 방법 또는 인간 세포의 생식 계열 유전적 동일성을 변형시키기 위한 방법이 아니다. 특정 구현예에서, 방법을 수행할 때, 표적 RNA는 인간 또는 동물 세포 내에 포함되지 않는다. 특정한 구현예에서, 표적이 인간 또는 동물 표적일 때, 방법은 생체외 또는 시험관내에서 수행된다.

[0755] 본 발명은 또한 변이체를 야기하는 질환 또는 표적화된 탈아미드화를 통해 질환을 치료 또는 예방하는 방법에 관한 것이다. 예를 들어, A의 탈아미노화는 병원성 G→A 또는 C→T 점 돌연변이를 함유하는 전사체에 의해 야기되는 질환을 치유할 수 있다. 본 발명에 의해 치료되거나 또는 예방될 수 있는 질환의 예는 암, 메이어-고린 증후군, 시클 증후군 4, 주버트 증후군 5, 레버 선천성 흑암시 10; 샤르코마리 투스병, 2형; 샤르코마리 투스병, 2형; 어셔 증후군, 2C형; 유전성 실조증 28; 유전성 실조증 28; 유전성 실조증 28; 긴 QT 증후군 2; 쇼그렌-라르손 증후군; 유전성 과당뇨증; 유전성 과당뇨증; 신경아세포종; 신경아세포종; 칼만 증후군 1; 칼만 증후군 1; 칼만 증후군 1; 이염성 백질디스트로피를 포함한다.

[0756] 일 양상에서, 본 발명은 돌연변이된 질환 유전자를 포함하는 모델 진핵생물 세포를 생성하는 방법을 제공한다. 일부 구현예에서, 질환 유전자는 질환을 갖거나 또는 질환이 발생할 위험의 증가와 연관된 임의의 유전자이다. 일부 실시형태에서, 방법은 (a) 하나 이상의 벡터를 진핵 세포에 도입하는 단계로서, 하나 이상의 벡터는 Cas 이펙터 모듈, 및 직접 반복부 서열에 연결된 가이드 서열 중 하나 이상의 발현을 유도하는 것인 단계; 및 (b) CRISPR-Cas 이펙터 모듈 복합체가 표적 폴리뉴클레오티드에 결합하여 상기 질환 유전자 내의 표적 폴리뉴클레오티드의 절단을 실시할 수 있게 하는 단계로서, CRISPR-Cas 이펙터 모듈 복합체는 (1) 표적 폴리뉴클레오티드 내 표적 서열과 하이브리드화하는 가이드 서열, (2) DR 서열, 및 (3) tracr 서열과 복합체 형성하는 Cas 이펙터 모듈을 포함하여서, 돌연변이된 질환 유전자를 포함하는 모델 진핵생물 세포를 생성시키는 것인 단계를 포함하고, 이것은 분할 Cas 이펙터 모듈을 포함한다. 일부 구현예에서, 상기 절단은 상기 Cas 이펙터 모듈에 의해 표적 서열의 위치에서 1 또는 2개의 가닥을 절단하는 단계를 포함한다. 바람직한 구현예에서, 가닥 파손은 5' 오버행을 갖는 스태거드 절단부이다. 일부 구현예에서, 상기 절단은 표적 유전자의 감소된 전사를 초래한다. 일부 구현예에서, 상기 방법은 외생성 주형 폴리뉴클레오티드와의 상동성 재조합에 의해 상기 절단된 표적 폴리뉴클레오티드를 복구하는 단계를 더 포함하며, 상기 복구은 상기 표적 폴리뉴클레오티드의 하나 이상의 뉴클레오티드의 삽입, 결실 또는 치환을 포함하는 돌연변이를 야기한다. 일부 구현예에서, 상기 돌연변이는 표적 서열을 포함하는 유전자로부터의 단백질 발현에서 하나 이상의 아미노산 변화를 초래한다. 일부 구현예에서, 모델 진핵생물 세포는 돌연변이된 질환 유전자를 포함하고, 돌연변이는 5' 오버행을 갖는 스태거드 이중 가닥 파손에 의해 도입된다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, 모델 진핵생물 세포는 돌연변이된 질환 유전자를 포함하고, 돌연변이는 HDR을 통해서 스태거드 5' 오버행에서 DNA 삽입에 의해 도입된다. 일부 구현예에서, 모델 진핵생물 세포는 돌연변이된 질환 유전자를 포함하고, 돌연변이는 NHEJ를 통해서 스태거드 5' 오버행에서 DNA 삽입에 의해 도입된다. 일부 구현예에서, 모델 진핵생물 세포는 CRISPR-C2c1 시스템에 의해 도입된 외생성 DNA 서열 삽입을 포함한다. 특정 구현예에서,

CRISPR-C2c1 시스템은 5' 및 3' 말단 둘 모두 상에서 가이드 서열이 측접된 외생성 DNA를 포함한다. 일부 구현예에서, 모델 진핵생물 세포는 돌연변이된 질환 유전자를 포함하고, 돌연변이 c는 특정 구현예에서 스태거드 5' 오버행에서 DNA 삽입에 의해 도입되고, Cas 이펙터 모듈은 C2c1 단백질, 또는 이의 촉매 도메인, 및 PAM 서열, T-풍부 서열을 포함한다. 특정 구현예에서, PAM은 5'-TTN 또는 5'-ATTN 이며, 여기서 N 은 임의의 뉴클레오티드이다. 특정 구현예에서, PAM은 5'-TTG 이다. 특정 구현예에서, 모델 진핵생물 세포는 암과 연관된 돌연변이된 유전자를 포함한다. 특정 구현예에서, 모델 진핵생물 세포는 자궁경부 상피내 종양(CIN)에서 인간 파필로마바이러스 (HPV) 구동된 발암과 연관된 돌연변이된 질환 유전자를 포함한다. 다른 특정 구현예에서, 모델 진핵생물 세포는 파킨슨병, 낭성 섬유증, 심근병증, 및 허혈성 심장 질환과 연관된 돌연변이된 질환 유전자를 포함한다.

[0757] 일 양상에서 본 발명은 하나 이상의 세포(들)에서 유전자에 하나 이상의 돌연변이를 도입시켜 하나 이상의 세포(들)를 선택하는 방법을 제공하고, 방법은 하나 이상의 벡터를 세포(들)에 도입시키는 단계로서, 하나 이상의 벡터가 Cas 이펙터 모듈, 직접 반복부 서열에 연결된 가이드 서열, 및 편집 주형 중 하나 이상의 발현을 구동시키고, 편집 주형은 Cas 이펙터 모듈 절단을 없애는 하나 이상의 돌연변이를 포함하는 것인 단계; 세포(들)에서 표적 폴리뉴클레오티드와 편집 주형의 상동성 재조합을 선택할 수 있게 하는 단계; CRISPR-Cas 이펙터 모듈 복합체가 표적 폴리뉴클레오티드에 결합할 수 있게 하여서 상기 유전자 내 표적 폴리뉴클레오티드의 절단을 실시하는 단계로서, CRISPR-Cas 이펙터 모듈 복합체는 (1) 표적 폴리뉴클레오티드 내 표적 서열과 하이브리드화하는 가이드 서열, 및 (2) 직접 반복부 서열과 복합체 형성하는 Cas 이펙터 모듈을 포함하고, Cas 이펙터 모듈 CRISPR-Cas 이펙터 모듈 복합체의 표적 폴리뉴클레오티드와의 결합은 세포 사멸을 유도시켜서 하나 이상의 돌연변이가 도입된 하나 이상의 세포(들)를 선택할 수 있게 하는 단계를 포함하고, 이것은 분할 Cas 이펙터 모듈을 포함한다. 본 발명의 또 다른 바람직한 구현예에서, 선택되는 세포는 진핵생물 세포일 수 있다. 본 발명의 양상은 선택 마커를 필요로 하는 일 없이 특정 세포의 선택 또는 계수기-선택 시스템을 포함할 수 있는 2 단계 공정을 가능하게 한다.

[0758] 일 양상에서, 본 발명은 변형 또는 편집된 유전자를 포함하는 진핵생물 세포를 생성시키는 방법을 제공한다. 일부 구현예에서, 변형 또는 편집된 유전자는 질환 유전자이다. 일부 구현예에서, 방법은 (a) 하나 이상의 벡터를 진핵생물 세포에 도입시키는 단계로서, 하나 이상의 벡터는 Cas 이펙터 모듈, 및 직접 반복부 서열에 연결된 가이드 서열 중 하나 이상의 발현을 구동시키고, Cas 이펙터 모듈은 염기 편집을 매개하는 하나 이상의 이펙터 도메인을 회합하는 것인 단계, 및 (b) CRISPR-Cas 이펙터 모듈 복합체가 표적 폴리뉴클레오티드에 결합할 수 있게 하여 상기 질환 유전자 내 표적 폴리뉴클레오티드의 염기 편집을 실시하는 것인 단계로서, CRISPR-Cas 이펙터 모듈 복합체는 표적 폴리뉴클레오티드 내 표적 서열에 하이브리드화하는 가이드 서열과 복합체 형성하는 Cas 이펙터 모듈을 포함하고, 가이드 서열은 가이드 서열 및 표적 서열 간에 형성된 DNA/RNA 헤테로듀플렉스 또는 RNA/RNA 듀플렉스 사이에 하나 이상의 미스매치를 도입하도록 디자인될 수 있는 것인 단계를 포함한다.

특정 구현예에서, 미스매치는 A-C 미스매치이다. 일부 구현예에서, Cas 이펙터는 하나 이상의 기능성 도메인과 (예를 들어, 융합 단백질 또는 적합한 링커를 통해) 회합될 수 있다. 일부 구현예에서, 이펙터 도메인은 가수분해적 탈아미드화를 통해 내생성 편집을 매개하는 하나 이상의 시티딘 또는 아데노신 디아미나제를 포함한다. 특정 구현예에서, 이펙터 도메인은 효소의 RNA (ADAR) 패밀리에 작용하는 아데노신 디아미나제를 포함한다. 특정 구현예에서, 아데노신 디아미나제 단백질 또는 이의 촉매적 도메인은 RNA에서 아데노신 또는 시토신을 탈아미노화시킬 수 있거나 또는 RNA 특이적 아데노신 디아미나제이고/이거나 박테리아, 인간, 두족류 또는 초파리 아데노신 디아미나제 단백질 또는 이의 촉매적 도메인, 바람직하게는 TadA, 더 바람직하게는 ADAR, 임의로 huADAR, 임의로 (hu)ADAR1 또는 (hu)ADAR2, 바람직하게는 huADAR2 또는 이의 촉매적 도메인이다.

일부 구현예에서, 시티딘 디아미나제는 인간, 래트, 또는 램프리 시티딘 디아미나제이다. 일부 구현예에서, 시티딘 디아미나제는 아포리포단백질 B mRNA-편집 복합체 (APOBEC) 패밀리 디아미나제, 활성화-유도된 디아미나제 (AID), 또는 시티딘 디아미나제 1 (CDA1)이다.

[0759] 추가적인 양상은 상기 기재된 방법으로부터 얻어지거나 또는 얻을 수 있고/있거나 상기 기재된 조성물을 포함하는 단리된 세포 또는 상기 변형된 세포의 자손에 관한 것이되, 바람직하게는 상기 세포는 상기 방법을 실시하지 않은 대응하는 세포에 비해 관심 대상의 상기 표적 RNA에서 상기 아데닌 대신 하이포잔틴 또는 구아닌을 포함한다. 특정 구현예에서, 세포는 진핵생물 세포, 바람직하게는 인간 또는 비-인간 동물 세포, 선택적으로 치료적 T 세포 또는 항체-생성 B-세포이거나 또는 상기 세포는 식물 세포이다. 추가적인 양상은 상기 변형된 세포 또는 이의 자손을 포함하는 비인간 동물 또는 식물을 제공한다. 또한 추가적인 양상은 요법, 바람직하게는 세포 요법에서 사용하기 위해 본 명세서에서 상기 기재한 바와 같은 변형된 세포를 제공한다.

[0760] 일부 구현예에서, 변형된 세포는 치료적 T 세포, 예컨대 CAR-T 요법에 적합한 T 세포이다. 변형은 제한없이, 면역 체크포인트 수용체 (예를 들어, PDA, CTLA4)의 감소된 발현, HLA 단백질 (예를 들어, B2M, HLA-A)의 감소된 발현, 및 내생성 TCR의 감소된 발현을 포함하여, 치료적 T 세포에서 하나 이상의 바람직한 특성을 일으킬 수 있다.

[0761] 본 발명은 또한 본원에 기재된 변형된 세포를 이를 필요로 하는 환자에게 투여하는 것을 포함하는 세포 요법을 위한 방법에 관한 것이며, 변형된 세포의 존재는 환자에서 질환을 치료한다. 하나의 구현예에서, 세포 요법을 위한 변형된 세포는 종양 세포를 인지 및/또는 공격할 수 있는 CAR-T 세포이다. 다른 구현예에서, 세포 요법을 위한 변형된 세포는 줄기 세포, 예컨대 신경 줄기 세포, 중간엽 줄기 세포, 조혈 줄기 세포, 또는 iPSC 세포이다.

[0762] 본 명세서의 다른 곳에 정의된 바와 같은 치료 방법에서 사용하기 위한, 바람직하게는 직렬 배열된 다중 가이드 RNA를 포함하는 Cas 이펙터 모듈, 복합체 또는 시스템을 포함하는 조성물, 또는 바람직하게는 직렬 배열된 다중 가이드 RNA를 포함하는 상기 Cas 이펙터 모듈, 복합체 또는 시스템을 코딩하거나 또는 포함하는 폴리뉴클레오티드 또는 벡터가 또한 제공된다. 이러한 조성물을 포함하는 부분의 키트가 제공될 수 있다. 이러한 치료 방법을 위한 의약의 제조에서 상기 조성물의 용도가 또한 제공된다. 스크리닝에서 Cas 이펙터 모듈 CRISPR 시스템의 용도, 예를 들어, 기능 획득 스크리닝이 또한 본 발명에 의해 제공된다. 유전자를 과발현시키도록 인공적으로 힘이 가해진 세포는, 예를 들어, 음성 피드백 루프에 의해 시간에 따라 유전자를 하향조절할 수 있다(평형상태 재확립). 시간에 따라, 비조절 유전자가 다시 감소될 수 있도록 스크리닝을 시작한다. 유도성 Cas 이펙터 모듈 활성인자를 이용하는 것은 스크리닝 바로 전에 전사를 유도하도록 하며, 따라서, 위음성 히트(false negative hit)의 기회를 최소화한다. 따라서, 스크리닝, 예를 들어, 기능 획득 스크리닝에서 본 발명의 사용에 의해, 위음성 기회 결과는 최소화될 수 있다.

[0763] 다른 양상에서, 본 발명은 유전자 산물을 코딩하는 DNA 분자를 각각 특이적으로 표적화하는 다중 Cas12b CRISPR 시스템 가이드 RNA에 작동적으로 연결된 제1 조절 엘리먼트 및 CRISPR 단백질에 대한 암호에 작동적으로 연결된 제2 조절 엘리먼트를 포함하는 하나 이상의 벡터를 포함하는, 조작된, 비천연 발생 벡터 시스템을 제공한다. 조절 엘리먼트는 둘 다 시스템의 동일한 벡터 상에 또는 상이한 벡터 상에 위치될 수 있다. 다중 가이드 RNA는 세포에서 다중 유전자 산물을 코딩하는 다중 DNA 분자를 표적화하고, CRISPR 단백질은 유전자 산물을 코딩하는 다중 DNA 분자를 절단할 수 있고 (이는 가닥 중 하나 또는 둘 다를 절단할 수 있거나 또는 뉴클레아제 활성이 실질적으로 없을 수 있음), 그리하여 다중 유전자 산물의 발현이 변경되고, CRISPR 단백질 및 다중 가이드 RNA는 천연적으로 함께 존재하지 않는다. 바람직한 구현예에서, CRISPR 단백질은 선택적으로 진핵생물 세포에서 발현을 위해 코돈 최적화된 Cas12b 단백질이다. 바람직한 구현예에서, 진핵생물 세포는 포유류 세포, 식물 세포 또는 효모 세포이고, 더 바람직한 구현예에서, 포유류 세포는 인간 세포이다. 본 발명의 추가적인 구현예에서, 다중 유전자 산물 각각의 발현은 변경되며, 바람직하게는 감소된다.

[0764] 일 양상에서, 본 발명은 하나 이상의 벡터를 포함하는 벡터 시스템을 제공한다. 일부 실시형태에서, 시스템은 (a) 직접 반복 서열 및 직접 반복 서열의 상류 또는 하류에(어느 쪽이든 적용 가능함) 하나 이상의 가이드 서열을 삽입하기 위한 하나 이상의 삽입 부위에 작동적으로 연결된 제1 조절 서열로서, 발현될 때, 하나 이상의 가이드 서열(들)이 진핵 세포에서 하나 이상의 표적 서열(들)에 대한 CRISPR 복합체의 서열-특이적 결합을 지시하고, CRISPR 복합체는 하나 이상의 표적 서열(들)에 혼성화된 하나 이상의 가이드 서열(들)과 복합체화된 Cas12b 효소를 포함하는, 상기 제1 조절 엘리먼트; 및 (b) 바람직하게는 적어도 하나의 핵 국소화 서열 및/또는 적어도 하나의 NES를 포함하는, 상기 Cas12b 효소를 코딩하는 효소-암호화 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함하되; 성분 (a) 및 (b)는 시스템의 동일 또는 상이한 벡터 상에 위치된다. 적절한 경우, tracr 서열이 또한 제공될 수 있다. 일부 구현예에서, 성분 (a)는 제1 조절 엘리먼트에 작동적으로 연결된 2 이상의 가이드 서열을 추가로 포함하되, 발현될 때, 2 이상의 가이드 서열 각각은 진핵생물 세포에서 상이한 표적 서열에 대한 Cas12b CRISPR 복합체의 서열 특이적 결합을 지시한다. 일부 구현예에서, CRISPR 복합체는 진핵생물 세포에서 핵 내에서 또는 핵 밖에서 검출 가능한 양으로 상기 Cas12b CRISPR 복합체의 축적을 유도하는데 충분한 강도의 하나 이상의 핵 국재화 서열 및/또는 하나 이상의 NES를 포함한다. 일부 구현예에서, 제1 조절 엘리먼트는 중합효소 III 프로모터이다. 일부 구현예에서, 제2 조절 엘리먼트는 중합효소 II 프로모터이다. 일부 구현예에서, 가이드 서열의 각각은 길이가 적어도 16, 17, 18, 19, 20, 25개의 뉴클레오티드, 또는 16 내지 30, 또는 16 내지 25 또는 16 내지 20개의 뉴클레오티드이다.

[0765] 재조합 발현 벡터는 숙주 세포에서 핵산의 발현에 적합한 형태로 본 명세서에 정의된 바와 같은 다중 표적화에서 사용하기 위한 Cas12b 효소, 시스템 또는 복합체를 코딩하는 폴리뉴클레오티드를 포함할 수 있는데, 이는 재

조합 발현 벡터가 발현을 위해 사용될 숙주 세포에 기반하여 선택될 수 있는, 즉, 발현될 핵산 서열에 작동적으로 연결되는 하나 이상의 조절 엘리먼트를 포함한다는 것을 의미한다. 재조합 발현 벡터 내에서, "작동적으로 연결된" 은 대상 뉴클레오티드 서열이 (예를 들어, 시험관내 전사/번역 시스템 내에서, 또는 벡터가 숙주 세포 내로 도입되는 경우 숙주 세포 내에서) 뉴클레오티드 서열의 발현을 가능하게 하는 방식으로 조절 엘리먼트(들)에 연결된 것을 의미하는 의도이다.

[0766] 일부 구현예에서, 숙주 세포는 본 명세서에 정의된 바와 같은 다중 표적화에서 사용하기 위한 Cas12b 효소, 시스템 또는 복합체를 코딩하는 폴리뉴클레오티드를 포함하는 하나 이상의 벡터로 일시적으로 또는 비일시적으로 형질감염된다. 일부 구현예에서, 세포는 대상체에서 자연적으로 발생되면서 형질감염된다. 일부 구현예에서, 형질감염된 세포가 대상체로부터 취해진다. 일부 구현예에서, 세포는 대상체로부터 취한 세포, 예컨대 세포주로부터 유래된다. 조직 배양을 위한 매우 다양한 세포주는 당업계에 공지되어 있으며, 본 명세서의 다른 곳에 예시되어 있다. 세포주는 당업자에게 공지된 다양한 공급원으로부터 입수 가능하다(예를 들어, 미국 미생물 보존센터(American Type Culture Collection:ATCC)(버지니아주 매너서스에 소재)). 일부 구현예에서, 본 명세서에 정의된 바와 같은 다중 표적화에서 사용하기 위한 Cas12b 효소, 시스템 또는 복합체를 코딩하는 폴리뉴클레오티드를 포함하는 하나 이상의 하나 이상의 벡터로 형질감염된 세포는 하나 이상의 벡터-유래 서열을 포함하는 새로운 세포주를 확립하기 위해 사용된다. 일부 구현예에서, 본 명세서에 기재된 바와 같은 다중 표적화에서 사용하기 위한 Cas12b CRISPR 시스템 또는 복합체의 성분으로 일시적으로 형질감염된 세포(예컨대 하나 이상의 벡터의 일시적 형질감염, 또는 RNA에 의한 형질감염에 의함), Cas12b CRISPR 시스템 또는 복합체의 활성을 통해 변형된 세포는 변형을 함유하지만, 임의의 다른 외생성 서열을 결여하는 세포를 포함하는 새로운 세포주를 확립하기 위해 사용된다. 일부 구현예에서, 본 명세서에 정의된 바와 같은 다중 표적화에서 사용하기 위한 Cas12b 효소, 시스템 또는 복합체를 코딩하는 폴리뉴클레오티드를 포함하는 하나 이상의 벡터로 일시적으로 또는 비일시적으로 형질감염된 세포, 또는 이러한 세포로부터 유래된 세포주는 하나 이상의 시험 화합물을 평가하는 데 사용된다.

[0767] 용어 "조절 엘리먼트" 는 본 명세서의 다른 부분에서 정의된 바와 같다.

[0768] 유리한 벡터는 렌티바이러스 및 아데노-연합 바이러스를 포함하고, 이러한 벡터의 유형은 또한 특정 세포 유형을 표적화하도록 선택된다.

[0769] 일 양상에서, 본 발명은 (a) 직접 반복부 서열의 (적용가능하면 어느 쪽이든) 상류 또는 하류에 하나 이상의 까이드 RNA를 삽입하기 위한 하나 이상의 삽입 부위 및 직접 반복부 서열에 작동적으로 연결된 제1 조절 엘리먼트로서, 발현될 때, 가이드 서열(들)은 진핵생물 세포의 개별 표적 서열(들)과 Cas12b CRISPR 복합체의 서열-특이적 결합을 유도시키고, Cas12b CRISPR 복합체는 개별 표적 서열(들)과 하이브리드화하는 하나 이상의 가이드 서열(들)과 복합체 형성되는 Cas12b 효소를 포함하는 것인 단계; 및 (b) 바람직하게 적어도 하나의 핵 국재화 서열 및/또는 NES를 포함하는 상기 Cas12b 효소를 코딩하는 효소-코딩 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함하는 진핵생물 숙주 세포를 제공한다. 일부 구현예에서, 숙주 세포는 성분 (a) 및 (b)를 포함한다. 적용가능한 경우에, tracr 서열이 또한 제공될 수 있다. 일부 구현예에서, 성분 (a), 성분 (b), 또는 성분 (a) 및 (b)는 숙주 진핵생물 세포의 게놈 내로 안정하게 통합된다. 일부 구현예에서, 성분 (a)는 제1 조절 엘리먼트에 작동적으로 연결되고, 선택적으로 직접 반복부에 의해 분리되는 2 이상의 가이드 서열을 추가로 포함하되, 발현될 때, 2 이상의 가이드 서열의 각각은 진핵생물 세포 내 상이한 표적 서열에 대한 Cas12b CRISPR 복합체의 서열 특이적 결합을 지시한다. 일부 구현예에서, Cas12b 효소는 진핵생물 세포 핵 내 및/또는 밖에서 검출 가능한 양으로 상기 CRISPR 효소의 축적을 유도하는 데 충분한 강도의 하나 이상의 핵 국재화 서열 및/또는 핵 유출 서열 또는 NES를 포함한다.

[0770] 일부 구현예에서, 가이드 분자는 편집하려는 적어도 하나의 표적 아데노신 잔기를 포함하는 표적 DNA 가닥과 듀플렉스를 형성한다. 표적 DNA 가닥에 대한 가이드 RNA 분자의 하이브리드화 시, 아데노신 디아미나제는 이중가닥에 결합하고, DNA-RNA 이중가닥 내에 포함된 하나 이상의 표적 아데노신 잔기의 탈아미노화를 촉매한다.

[0771] 추가로, PAM 상호작용 (PI) 도메인의 조작은 PAM 특이성의 프로그래밍을 가능하게 할 수 있고, 표적 부위 인식 신뢰성을 개선시킬 수 있으며, 예를 들어 [Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592]의 Cas9에 대해 기술된 바와 같이, CRISPR-Cas 단백질의 다재다능성을 증가시킨다. 본 명세서에서 더욱 상술하는 바와 같이, 당업자는 C2c1 단백질을 유사하게 변형시킬 수 있다는 것을 이해하게 될 것이다.

[0772] 특정 구현예에서, 가이드 서열은 탈아미노화시키려는 아데닌에 대한 디아미나제의 최적 효율을 보장하기 위해

선택된다. C2c1 닉카제의 절단 부위에 대해서 표적 가닥 내 아데닌의 위치를 고려할 수 있다. 특정 구현예에서 닉카제는 비-표적 가닥 상에서, 탈아미노화시키려는 아데닌 부근에서 작용할 것임을 보장하는 것에 관심있다. 예를 들어, 특정 구현예에서, Cas12b 닉카제는 PAM의 하류의 비표적화 가닥을 절단하고 이것은 탈아미드화시키려는 아데닌에 상응하는 시토신이 상응하는 비표적 가닥의 서열 내 닉카제 절단 부위의 상류 또는 하류 10 bp 이내에 가이드 서열에 위치하는 것인 가이드를 설계하는 것이 흥미로울 수 있다.

[0773] **전달**

[0774] 일부 구현예에서, CrISPR-Cas 시스템의 성분은 다양한 형태, 예컨대 DNA/RNA 또는 RNA/RNA 또는 단백질 RNA의 조합으로 전달될 수 있다. 예를 들어, C2c1 단백질은 DNA-코딩 폴리뉴클레오티드 또는 RNA-코딩 폴리뉴클레오티드로서 또는 단백질로서 전달될 수 있다. 가이드는 DNA-코딩 폴리뉴클레오티드 또는 RNA로서 전달될 수 있다. 혼합된 전달 형태를 포함하여, 모든 가능한 조합이 고려된다.

[0775] 일부 양상에서, 본 발명은 하나 이상의 폴리뉴클레오티드, 예컨대 또는 본 명세서에 기술된 바와 같은 하나 이상의 벡터, 이의 하나 이상의 전사물, 및/또는 그로부터 전사되는 하나 이상의 단백질을 숙주 세포에 전달하는 단계를 포함하는 방법을 제공한다.

[0776] **전달 비히클로서 벡터**

[0777] 재조합 발현 벡터는 숙주 세포에서 핵산의 발현에 적합한 형태로 본 발명의 핵산을 포함할 수 있는데, 이는 재조합 발현 벡터가 하나 이상의 조절 엘리먼트를 포함하는 것을 의미하며, 하나 이상의 조절 엘리먼트는 발현에 사용될 숙주 세포에 기반하여 선택될 수 있고, 발현될 핵산 서열에 작동적으로 연결된다. 재조합 발현 벡터 내에서, "작동적으로 연결된" 은 대상 뉴클레오티드 서열이 (예를 들어, 시험관내 전사/번역 시스템 내에서, 또는 벡터가 숙주 세포 내로 도입되는 경우 숙주 세포 내에서) 뉴클레오티드 서열의 발현을 가능하게 하는 방식으로 조절 엘리먼트(들)에 연결된 것을 의미하는 의도이다. 유리한 벡터는 렌티바이러스 및 아데노-연합 바이러스를 포함하고, 이러한 벡터의 유형은 또한 특정 세포 유형을 표적화하도록 선택된다.

[0778] 재조합 및 클로닝 방법에 관해, 미국 특허 제20040171156 A1호로서 2004년 9월 2일자로 공개된 미국 특허 출원 제10/815,730호가 언급되며, 이의 내용은 본 명세서에 그들의 전문이 참고로 편입된다.

[0779] 용어 "조절 엘리먼트" 는 프로모터, 인핸서, 내부 리보솜 진입 부위 (IRES) 및 다른 발현 제어 구성요소 (예를 들어, 전사 종결 신호, 예컨대, 폴리아데닐화 신호 및 폴리-U 서열)를 포함하도록 의도된다. 이와 같은 조절 엘리먼트는, 예를 들어 문헌 [Goeddel, GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)]에 기재되어 있다. 조절 엘리먼트는 많은 유형의 숙주 세포에서 뉴클레오티드 서열의 구성적 발현을 지시하는 것들과 특정 숙주 세포에서만 뉴클레오티드 서열의 발현을 지시하는 것들(예를 들어, 조직-특이적 조절 서열)을 포함한다. 조직-특이적 프로모터는 근육, 뉴런, 뼈, 피부, 혈액, 특정 장기(예를 들어, 간, 췌장)와 같은 요망되는 관심 조직, 또는 특정 세포 유형(예를 들어, 림프구)에서 주로 발현을 유도할 수 있다. 조절 엘리먼트는 시간-의존적 방식으로, 예컨대 세포-주기 의존적 또는 발생 단계 의존적 방식으로 발현을 지시할 수 있으며, 이것은 조직 특이적이거나 세포-유형 특이적일 수 있거나, 또한 그렇지 않을 수 있다. 일부 구현예에서, 벡터는 하나 이상의 pol III 프로모터 (예를 들어, 1, 2, 3, 4, 5개 이상의 pol III 프로모터), 하나 이상의 pol II 프로모터 (예를 들어, 1, 2, 3, 4, 5개 이상의 pol II 프로모터), 하나 이상의 pol I 프로모터 (예를 들어, 1, 2, 3, 4, 5개 이상의 pol I 프로모터), 또는 이들의 조합을 포함한다. pol III 프로모터의 예는 U6 및 H1 프로모터를 포함하지만, 이들로 제한되지 않는다. pol II 프로모터의 예는 레트로바이러스 라우스 육종 바이러스 (RSV) LTR 프로모터(임의로 RSV 인핸서를 가짐), 거대세포바이러스 (CMV) 프로모터 (임의로 CMV 인핸서를 가짐)[예를 들어, 문헌 [Boshart et al, Cell, 41:521530 (1985)] 참조], SV40 프로모터, 디히드로폴레이트 리덕타제 프로모터, β-액틴 프로모터, 포스포글리세롤 키나제 (PGK) 프로모터 및 EF1α 프로모터를 포함하지만, 이들로 제한되지 않는다. 또한 용어 " 조절 엘리먼트" 는 인핸서 엘리먼트, 예컨대 WPRE; CMV 인핸서; HTLV-I의 LTR의 R-U5' 절편 (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 인핸서; 및 토끼β-글로빈의 엑손 2 및 3 사이의 인트론 서열 (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981)을 포괄한다. 발현 벡터의 디자인은 형질전환시키려는 숙주 세포의 선택, 바람직한 발현 수준 등과 같은 인자에 의존할 수 있다는 것은 당업자가 이해하게 될 것이다. 벡터는 숙주 세포로 도입될 수 있어서, 본 명세서에 기술된 바와 같은 핵산에 의해 코딩되는, 융합 단백질 또는 펩티드를 포함하여, 전사물, 단백질, 또는 펩티드 (예를 들어, CRISPR 전사물, 단백질, 효소, 이의 돌연변이체 형태, 이의 융합 단백질 등)를 생산할 수 있다. 조절 서열에 관해, 미국 특허 출원 제10/491,026호가 언급되며, 이의 내용은 본 명세서에 그의 전문이 참고로 편입된다. 프로모터에 관해, 국제 특허 출원 WO 2011/028929

및 미국 특허 출원 제12/511,940호가 언급되며, 이의 내용은 본 명세서에 그들의 전문이 참고로 편입된다.

[0780] 유리한 벡터는 렌티바이러스 및 아데노-연합 바이러스를 포함하고, 이러한 벡터의 유형은 또한 특정 세포 유형을 표적화하도록 선택된다.

[0781] 특정 구현예에서, 가이드 RNA 및 아데노신 데아미나제에 융합된 (임의로 변형 또는 돌연변이된) CRISPR-Cas 단백질에 대해 바이시스트론 벡터가 사용된다. 가이드 RNA 및 아데노신 디아미나제에 융합된 (임의로 변형 또는 돌연변이된) CRISPR-Cas 단백질에 대한 바이시스트론 발현 벡터가 바람직하다. 일반적으로 특히 본 구현예에서, 아데노신 디아미나제에 융합된 (임의로 변형 또는 돌연변이된) CRISPR-Cas 단백질은 바람직하게는 CBh 프로모터에 의해 구동된다. RNA는 바람직하게는 Pol III 프로모터, 예컨대, U6 프로모터에 의해 유도될 수 있다. 이상적으로는 2개가 조합된다.

[0782] 벡터는 원핵 또는 진핵생물 세포에서 CRISPR 전사물 (예를 들어, 핵산 전사물, 단백질 또는 효소)의 발현을 위해 디자인될 수 있다. 예를 들어, CRISPR 전사물은 박테리아 세포, 예컨대, 에스케리치아 콜라이, 곤충 세포 (배큘로바이러스 발현 벡터를 이용), 효모 세포 또는 포유동물 세포에서 발현될 수 있다. 적합한 숙주 세포는 [Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)] 에 추가로 논의된다. 대안적으로, 재조합 발현 벡터는, 예를 들어 T7 프로모터 조절 서열 및 T7 중합효소를 이용하여 시험관내에서 전사 및 번역될 수 있다.

[0783] 벡터는 원핵생물 또는 원핵 세포에 도입되고 증식될 수 있다. 일부 구현예에서, 원핵생물은 진핵생물 세포 내로 도입될 벡터의 복제물을 증식시키기 위해 또는 진핵생물 세포 내로 도입될 벡터 생성에서의 중간 벡터로서 사용된다 (예를 들어, 바이러스 벡터 패키징 시스템의 부분으로서 플라스미드를 증폭). 일부 구현예에서, 원핵생물은 벡터의 복제물을 증폭시키기 위해 그리고, 예컨대, 숙주 세포 또는 숙주 유기체에 전달을 위한 하나 이상의 단백질 공급원을 제공하기 위해 하나 이상의 핵산을 발현시키기 위해 사용된다. 원핵생물 내 단백질의 발현은 융합 또는 비융합 단백질 중 하나의 발현을 지시하는 항상성 또는 유도성 프로모터를 함유하는 벡터로 에스케리치아 콜라이에서 가장 흔히 수행된다. 융합 벡터는 그 안에서 코딩되는 단백질, 예컨대, 재조합 단백질의 아미노 말단에 다수의 아미노산을 부가한다. 이러한 융합 벡터는 하나 이상의 목적, 예컨대: (i) 재조합 단백질 발현을 증가시키고; (ii) 재조합 단백질의 용해도를 증가시키고; (iii) 친화도 정제에서 리간드로서 작용함으로써 재조합 단백질의 정제에 도움을 주는 작용을 할 수 있다. 종종, 융합 발현 벡터에서, 단백질 절단 부위는 융합 모이어티 및 재조합 단백질의 접합부에 도입되어 융합 단백질의 후속 정제를 위해 융합 모이어티로부터 재조합 단백질의 분리를 가능하게 한다. 이러한 효소, 및 그들의 동족 인식 서열은 인자 Xa, 트롬빈 및 엔테로키나제를 포함한다. 예시적 융합 발현 벡터는 각각 글루타티온 S-트랜스퍼라제(GST), 말토스 E 결합 단백질 또는 단백질 A를 표적 재조합 단백질에 융합시키는, pGEX(Pharmacia Biotech Inc; Smith and Johnson, 1988. Gene 67: 31- 40), pMAL (New England Biolabs, Beverly, Mass.) 및 pRIT5 (Pharmacia, Piscataway, N.J.)를 포함한다. 적합한 유도성 비융합 이. 콜라이 발현 벡터의 예는 pTrc (Amrann et al., (1988) Gene 69:301- 315) 및 pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60- 89)를 포함한다. 일부 구현예에서, 벡터는 효모 발현 벡터이다. 효모 사카로마이세스 세레비지아에 (Saccharomyces cerivisae)에서 발현을 위한 벡터의 예는 pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), 및 picZ (InVitrogen Corp, San Diego, Calif.)를 포함한다. 일부 구현예에서, 벡터는 배큘로바이러스 발현 벡터를 이용하여 곤충 세포에서 단백질 발현을 유도한다. 배양된 곤충 세포 (예를 들어, SF9 세포)에서 단백질의 발현에 이용가능한 배큘로바이러스 벡터는 pAc 시리즈 (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) 및 pVL 시리즈 (Lucklow and Summers, 1989. Virology 170: 31-39)를 포함한다.

[0784] 일부 구현예에서, 벡터는 포유동물 발현 벡터를 이용하여 포유동물 세포에서 하나 이상의 서열의 발현을 유도할 수 있다. 포유동물 발현 벡터의 예는 pCDM8 (Seed, 1987. Nature 329: 840) 및 pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195)를 포함한다. 포유동물 세포에서 사용될 때, 발현 벡터의 제어 기능은 전형적으로 하나 이상의 조절 엘리먼트에 의해 제공된다. 예를 들어, 통상적으로 사용되는 프로모터는 폴리오마바이러스, 아데노바이러스 2, 거대세포바이러스, 유인원 바이러스 40, 및 본 명세서에 개시되고 당업계에 공지된 다른 것으로부터 유래된다. 원핵생물 및 진핵생물 세포 둘 모두를 위한 다른 적절한 발현 시스템에 대하여, 예를 들어, 문헌[Sambrook, et al., MOLECULAR CLONING:A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989]의 16 및 17장을 참조한

다.

[0785] 일부 구현예에서, 재조합 포유동물 발현 벡터는 특정 세포 유형에서 핵산의 발현을 우선적으로 유도할 수 있다 (예를 들어, 조직-특이적 조절 엘리먼트는 핵산을 발현시키기 위해 사용된다). 조직-특이적 조절 엘리먼트는 당업계에 공지되어 있다. 적절한 조직-특이적 프로모터의 비제한적인 예는 알부민 프로모터 (간-특이적; Pinkert, et al., 1987. Genes Dev. 1: 268-277), 림프구-특이적 프로모터 (Calame and Eaton, 1988. Adv. Immunol. 43: 235-275), 특히 T 세포 수용체 (Winoto and Baltimore, 1989. EMBO J. 8: 729-733) 및 면역글로불린 (Baneiji, et al., 1983. Cell 33: 729-740; Queen and Baltimore, 1983. Cell 33: 741-748)의 프로모터, 뉴론-특이적 프로모터 (예를 들어, 뉴로필라멘트 프로모터; Byrne and Ruddle, 1989. Proc. Natl. Acad. Sci. USA 86: 5473-5477), 췌장-특이적 프로모터 (Edlund, et al., 1985. Science 230: 912-916), 및 유선-특이적 프로모터 (예를 들어, 밀크 웨이 프로모터; U.S. 특허 출원 번호 4,873,316 및 유럽 공개 특허 출원 번호 264,166)를 포함한다. 발생-조절 프로모터, 예를 들어 쥣과 hox 프로모터 (Kessel and Gruss, 1990. Science 249: 374-379) 및 α-페토단백질 프로모터 (Campes and Tilghman, 1989. Genes Dev. 3: 537-546)가 포괄된다. 이들 원핵 및 진핵 벡터에 대해, 미국 특허 제6,750,059호가 언급되며, 이의 내용은 본 명세서에 그들의 전문이 참고로 편입된다. 본 발명의 다른 구현예는 바이러스 벡터의 용도에 관한 것일 수 있으며, 미국 특허 출원 제13/092,085호가 언급되고, 이의 내용은 본 명세서에 전문이 참고로 편입된다. 조직-특이적 조절 엘리먼트는 당업계에 공지되어 있고, 이와 관련하여, 미국 특허 제7,776,321호가 언급되며, 이의 내용은 본 명세서에서 그들의 전문이 참고로 편입된다. 일부 구현예에서, 조절 엘리먼트는 CRISPR 시스템의 하나 이상의 요소의 발현을 구동하기 위해 CRISPR 시스템의 하나 이상의 요소에 작동적으로 연결된다.

[0786] 핵산-표적화 시스템의 하나 이상의 구성요소의 발현을 구동하는 하나 이상의 벡터는 핵산-표적화 시스템의 구성요소의 발현이 하나 이상의 표적 부위에서 핵산-표적화 복합체의 형성을 유도하도록 숙주 세포에 도입된다. 예를 들어, 핵산-표적화 이펙터 효소 및 핵산-표적화 가이드 RNA는 별개의 벡터 상에서 별개의 조절 엘리먼트에 각각 작동적으로 연결될 수 있다. 핵산-표적화 시스템의 RNA(들)는 유전자이식 핵산-표적화 이펙터 단백질 동물 또는 포유동물, 예를 들어, 핵산-표적화 이펙터 단백질을 구성적으로 또는 유도적으로 또는 조건적으로 발현시키는 동물 또는 포유동물; 또는 핵산-표적화 이펙터 단백질을 달리 발현시키거나 또는 핵산-표적화 이펙터 단백질을 함유하는 세포를 갖는 동물 또는 포유동물에, 예컨대, 생체내 핵산-표적화 이펙터 단백질을 코딩하고 이를 발현시키는 벡터 또는 벡터들의 사전 투여에 의해 전달될 수 있다. 대안적으로, 동일하거나 또는 상이한 조절 엘리먼트로부터 발현되는 둘 이상의 엘리먼트는 단일 벡터에 조합될 수 있고, 하나 이상의 추가 벡터가 제1 벡터에 포함되지 않은 핵산-표적화 시스템의 임의 성분을 제공한다. 단일 벡터에 조합되는 핵산-표적화 시스템은 임의의 적합한 배향으로 정렬될 수 있는데, 예컨대 한 엘리먼트는 제2 엘리먼트에 대해 5' ("상류") 또는 3' ("하류")에 위치된다. 한 구성요소의 코딩 서열은 제2 구성요소의 코딩 서열의 동일하거나 또는 반대쪽 가닥 상에 위치되고, 동일 또는 반대 방향으로 배향될 수 있다. 일부 구현예에서, 단일 프로모터는 하나 이상의 인트론 서열 내에 (예를 들어, 상이한 인트론에 각각, 적어도 하나의 인트론에 2 이상, 또는 단일 인트론에 모두) 함입된 핵산-표적화 이펙터 단백질 및 핵산-표적화 가이드 RNA를 코딩하는 전사체의 발현을 유도한다. 일부 구현예에서, 핵산-표적화 이펙터 단백질 및 핵산-표적화 가이드 RNA는 동일한 프로모터에 작동적으로 연결되고 동일한 프로모터로부터 발현될 수 있다. 핵산-표적화 시스템의 하나 이상의 구성요소의 발현을 위한 전달 비히클, 벡터, 입자, 나노입자, 제형 및 이들의 성분은 앞서 언급한 문헌, 예컨대, WO 2014/093622 (PCT/US2013/074667)에서 사용되는 바와 같다. 일부 구현예에서, 벡터는 하나 이상의 삽입 부위, 예컨대, 제한 엔도뉴클레아제 인식 서열 (또한 "클로닝 부위" 라고도 함)을 포함한다. 일부 구현예에서, 하나 이상의 삽입 부위 (예를 들어, 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 삽입 부위)가 하나 이상의 벡터의 하나 이상의 서열 구성요소의 상류 및/또는 하류에 위치된다. 다중의 상이한 가이드 서열이 사용될 때, 세포 내에서 다중의 상이한, 상응하는 표적 서열에 대해 핵산-표적화 활성을 표적화하기 위해 단일 발현 구성체가 사용될 수 있다. 예를 들어, 단일 벡터는 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20개 이상의 가이드 서열을 포함할 수 있다. 일부 구현예에서, 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 이러한 가이드-서열-함유 벡터가 제공될 수 있으며, 선택적으로 세포에 전달된다. 일부 구현예에서, 벡터는 핵산-표적화 이펙터 단백질을 코딩하는 효소-코딩 서열에 작동적으로 연결된 조절 엘리먼트를 포함한다. 핵산-표적화 이펙터 단백질 또는 핵산-표적화 가이드 RNA 또는 RNA(들)은 개별적으로 전달될 수 있고, 유리하게 이들 중 적어도 하나는 입자 복합체를 통해 전달된다. RNA-표적화 이펙터 단백질 mRNA는 핵산-표적화 이펙터 단백질이 발현되는 시간을 제공하기 위하여 RNA-표적화 가이드 RNA 이전에 전달될 수 있다. 핵산-표적화 이펙터 단백질 mRNA는 핵산-표적화 가이드 RNA의 투여 전에 1 내지 12시간 (바람직하게는 대략 2 내지 6시간)에 투여될 수 있다. 대안적으로, 핵산-표적화 이펙터 단백질 mRNA 및 핵산-표적화 가이드 RNA는 함께 투여될

수 있다. 유사하게는, 가이드 RNA의 제2 부스터 용량은 핵산-표적화 이펙터 단백질 mRNA + 가이드 RNA의 초기 투여 후 1 내지 12시간 (바람직하게는 대략 2 내지 6시간)에 투여될 수 있다. 핵산-표적화 이펙터 단백질 mRNA 및/또는 가이드 RNA의 추가 투여는 게놈 변형의 가장 효율적인 수준을 달성하는 데 유용할 수 있다.

[0787] 통상적인 바이러스 및 비바이러스 기반 유전자 전달 방법은 포유동물 세포 또는 표적 조직에서 핵산을 도입하는 데 사용될 수 있다. 이러한 방법은 배양물에서 또는 숙주 세포 유기체에서 세포에 핵산-표적화 시스템의 성분을 코딩하는 핵산을 투여하는 데 사용될 수 있다. 비-바이러스 벡터 전달 시스템은 DNA 플라스미드, RNA (예를 들어, 본 명세서에 기재된 벡터의 전사체), 나형 핵산, 및 전달 비히클, 예컨대, 리포솜과 복합체화된 핵산을 포함한다. 바이러스 벡터 전달 시스템은 세포에 전달 후에 에피솜 또는 통합 게놈 중 하나를 갖는 DNA 및 RNA 바이러스를 포함한다. 유전자 요법 절차의 검토를 위해, 다음의 문헌들을 참조한다: Anderson, Science 256:808813 (1992); Nabel & Felgner, TIBTECH 11:211217 (1993); Mitani & Caskey, TIBTECH 11:162166 (1993); Dillon, TIBTECH 11:167175 (1993); Miller, Nature 357:455460 (1992); Van Brunt, Biotechnology 6(10):11491154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):3144 (1995); Haddada et al., Current Topics in Microbiology and Immunology, Doerfler and Bohm (eds) (1995); 및 Yu et al., Gene Therapy 1:13-26 (1994).

[0788] 핵산의 비-바이러스 전달 방법은 리포펙션, 뉴클레오펙션, 미세주입, 바이오리스틱스, 비로솜, 리포솜, 면역리포솜, 다가양이온 또는 지질:핵산 콘쥬게이트, 네이키드 DNA, 인공 비리온, 및 DNA의 제제-향상 흡수를 포함한다. 리포펙션은 예를 들어 미국 특허 번호 5,049,386, 4,946,787; 및 4,897,355에 기재되고, 리포펙션 시약은 상업적으로 판매된다 (예를 들어, Transfectam™ 및 Lipofectin™). 폴리뉴클레오티드의 효율적인 수용체-인식 리포펙션에 적절한 양이온 및 중성 지질은 Felgner, WO 91/17424; WO 91/16024의 것을 포함한다. 전달은 세포 (예를 들어, 시험관내 또는 생체외 투여) 또는 표적 조직 (예를 들어, 생체내 투여)에 대한 것일 수 있다.

[0789] 플라스미드 전달은 CRISPR-Cas 단백질 발현 플라스미드 내로 가이드 RNA의 클로닝 및 세포 배양물에서의 DNA 형질감염에 의해 수행된다. 플라스미드 골격은 상업적으로 입수 가능하며, 특정 장비가 필요하지 않다. 그들은 상이한 크기의 CRISPR-Cas 코딩 서열 (더 큰 크기의 단백질을 코딩하는 것을 포함)뿐만 아니라 선택 마커를 운반할 수 있는 모듈의 이점을 갖는다. 플라스미드의 이점은 둘 모두 그들이 일시적이지만, 지속적 발현을 보장할 수 있다는 것이다. 그러나, 플라스미드의 전달은 간단하지 않으므로, 생체내 효율이 종종 낮아진다. 지속 발현은 또한 오프-표적 편집을 증가시킬 수 있다는 점에서 불리할 수 있다. 추가로, CRISPR-Cas 단백질의 과량의 증강은 세포에 대해 독성이 될 수 있다. 마지막으로, 플라스미드는 보다 특히 이중-가닥 파괴가 생성된다는 관점에서 (온 및 오프-표적), 숙주 게놈 내에서 dsDNA의 무작위 통합 위험성을 항상 보유한다.

[0790] 면역지질 복합체 같은 표적화 리포솜을 포함하여, 지질:핵산 복합체의 제조는 당업자에게 충분히 공지되어 있다 (참조: 예를 들어 Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); 미국 특허 출원 번호 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 및 4,946,787). 이는 이하에서 더욱 상세하게 논의된다.

[0791] 핵산의 전달을 위한 RNA 또는 DNA 바이러스 기반 시스템의 사용은 체내 특정 세포에 대해 바이러스를 표적화하고 핵에 바이러스 페이로드를 수송하는 고도로 진화된 과정을 이용한다. 바이러스 벡터는 환자에게 (생체내) 직접적으로 투여될 수 있거나 또는 그들은 시험관내에서 세포를 처리하는 데 사용될 수 있고, 변형된 세포는 선택적으로 환자에게 (생체외) 투여될 수 있다. 통상적인 바이러스 기반 시스템은 유전자 전달을 위해 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 및 단순포진 바이러스 벡터를 포함할 수 있었다. 숙주 게놈 내 통합은 종종 삽입 이식유전자의 장기간 발현을 초래하는 레트로바이러스, 렌티바이러스, 및 아데노-연합 바이러스 유전자 전달 방법에 의해 가능하다. 추가적으로, 다수의 상이한 세포 유형 및 표적 조직에서 높은 형질도입 효율이 관찰되었다.

[0792] 레트로바이러스의 향성은 외래 외피 단백질을 혼입시켜, 표적 세포의 잠재적 표적 집단을 확장시킴으로써 변경될 수 있다. 렌티바이러스 벡터는 비-분화 세포를 형질도입 또는 감염시키고 전형적으로 고 바이러스 역가를 생성하는 레트로바이러스 벡터이다. 따라서 레트로바이러스 유전자 전달 시스템의 선택은 표적 조직에

의존한다. 레트로바이러스 벡터는 최대 6 내지 10 kb의 외래 서열에 대한 패키징 능력을 갖는 시스-작용성 긴 말단 반복부를 포함한다. 최소 시스-작용성 LTR은 벡터의 복제 및 패키징에 충분하며, 이어서, 표적 세포 내로 치료 유전자를 통합시키기 위해 사용되어 영구 이식유전자 발현을 제공한다. 널리 사용된 레트로바이러스 벡터는 쥐과 백혈병 바이러스(MuLV), 기본 유인원(gibbon ape) 백혈병 바이러스(GaLV), 원숭이 면역결핍 바이러스(SIV), 인간 면역결핍 바이러스(HIV), 및 이들의 조합을 기반으로 한 벡터를 포함한다 (참조: 예를 들어, Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

[0793] 일시적 발현이 바람직한 적용에서, 아데노바이러스 기반 시스템이 사용될 수 있다. 아데노바이러스 기반 벡터는 다수의 세포 유형에서 매우 높은 형질도입 효율을 가능하게 하고, 세포 분할을 필요로 하지 않는다. 이러한 벡터에 의해, 높은 발현 역가 및 수준이 얻어졌다. 이 벡터는 상대적으로 단순한 시스템에서 다량으로 생성될 수 있다. 아데노-연관 바이러스(" AAV") 벡터는 또한, 예를 들어 핵산 및 펩티드의 시험관 내 생산에서, 그리고 생체내 및 생체 외 유전자 치료법 과정을 위해 세포를 표적 핵산으로 세포 형질도입시키는 데 사용될 수 있다 (참조: 예를 들어, West et al., Virology 160:38-47 (1987); U.S. 특허 번호 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). 재조합 AAV 벡터의 구축은 다음의 문헌들을 포함한, 다수 공개물에 기술되어 있다: US 특허 번호 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); 및 Samulski et al., J. Virol. 63:03822-3828 (1989).

[0794] 본 발명은 CRISPR 시스템을 코딩하는 외생성 핵산 분자, 예를 들어, 프로모터, CRISPR-연관(Cas) 단백질 (추정 뉴클레아제 또는 헬리카제 단백질), 예를 들어, C2c1를 코딩하는 핵산 분자 및 종결자를 포함하거나 또는 본질적으로 그로 이루어진 제1 카세트, 및 하나 이상의, 유리하게는 벡터의 패키징 크기 한계까지, 예를 들어, 총 (제1 카세트를 포함) 5개의, 프로모터, 가이드 RNA(gRNA)를 코딩하는 핵산 분자 및 종결자를 포함하는 카세트 (예를 들어, 프로모터-gRNA1-종결자, 프로모터-gRNA2-종결자 ... 프로모터-gRNA(N)-종결자로서 개략적으로 나타내는 각각의 카세트, 여기서 N은 벡터의 패키징 크기 제한의 상한에서 삽입될 수 있는 수임)를 포함하거나 또는 이루어진 다수의 카세트를 함유하거나 또는 그로 본질적으로 이루어진 AAV, 또는 각각이 CRISPR 시스템의 하나 또는 하나 초과의 카세트를 함유하는 둘 이상의 개별 rAAV, 예를 들어, 프로모터, Cas, 예를 들어, Cas(C2c1)를 코딩하는 핵산 분자 및 종결자를 포함하거나 또는 본질적으로 이루어진 제1 카세트를 함유하는 제1 rAAV, 프로모터, 가이드 RNA(gRNA)를 코딩하는 핵산 분자 및 종결자를 포함하거나 또는 그로 본질적으로 이루어진 하나 이상의 카세트를 함유하는 제2 rAAV(예를 들어, 각각의 카세트는 프로모터-gRNA1-종결자, 프로모터-gRNA2-종결자 ... 프로모터-gRNA(N)-종결자로서 개략적으로 나타내며, 여기서 N은 벡터의 패키징 크기 한계의 상한으로 삽입될 수 있는 수이다)를 제공한다. 대안적으로, C2c1이 그 자체의 crRNA/gRNA를 프로세싱하기 때문에, 단일 crRNA/gRNA 어레이가 다중복합 유전자 편집에 사용될 수 있다. 그러한 이유로, gRNA를 전달하기 위한 다수 카세트를 포함하는 대신에, rAAV는 프로모터, 다수의 crRNA/gRNA, 및 종결자를 포함하거나 또는 그로 본질적으로 이루어지는 단일 카세트를 함유할 수도 있다 (예를 들어, 개략적으로 프로모터-gRNA1-종결자, 프로모터-gRNA2-종결자...프로모터-gRNA(N)-종결자로 표시되고, 여기서 N은 벡터의 패키징 크기 한계의 상한치에서 삽입될 수 있는 개수임). 그 전문이 참조로 본 명세서에 편입되는, [Zetsche et al Nature Biotechnology 35, 3134 (2017)]를 참조한다. rAAV가 DNA 바이러스이므로, AAV 또는 rAAV에 관해 본 명세서에서 논의되는 핵산 분자는 유리하게는 DNA이다. 프로모터는 일부 구현예에서, 유리하게 인간 시냅신 I 프로모터 (hSyn)이다. 세포에게 핵산의 전달을 위한 추가 방법은 당업자에게 충분히 공지되어 있다. 예를 들어, 참조로 본 명세서에 편입되는 미국 공개 특허 출원 제20030087817호를 참조한다.

[0795] 다른 구현예에서, 코칼 베시쿨로바이러스 (Cocal vesiculovirus) 외피 위형 레트로바이러스 벡터 입자가 고려된다 (참조: 예를 들어, 미국 프레드 허친슨 암 연구 센터(Fred Hutchinson Cancer Research Center)에 양도된 미국 특허 공개 제20120164118호). 코칼 바이러스는 베시쿨로바이러스 속이며, 포유동물에서 수포성 구내염의 원인 물질이다. 코컬 바이러스는 원래 트리니다드에서 진드기로부터 분리되었고(문헌[Jonkers et al., Am. Res. 25:236242 (1964)]), 트리니다드, 브라질 및 아르헨티나에서, 곤충, 소 및 말에서 감염이 확인되었다. 포유동물를 감염시키는 베시쿨로바이러스 중 다수는 자연적으로 감염된 절지동물로부터 단리되었는데, 이는 그들이 매개체 감염이라는 것을 시사한다. 베시쿨로바이러스에 대한 항체는 바이러스가 풍토성이고 실험실 획득된 시골 지역에 살고 있는 사람들 중에서는 일반적이고; 인간에서의 감염은 보통 인플루엔자-유사 증상을 초래한다. 코칼 바이러스 외피 당단백질은 아미노산 수준을 VSV-G 인디애나와 71.5% 동일성으로 공유하며, 베

시쿨로바이러스의 외피 유전자의 계통 발생론적 비교는 코칼 바이러스가 베시쿨로바이러스 중에서도 VSV-G 인디애나 균주와 혈청학적으로 별개이지만, 가장 밀접하게 관련된다는 것을 나타낸다. Jonkers et al., Am. J. Vet. Res. 25:236-242 (1964) 및 Travassos da Rosa et al., Am. J. Tropical Med. & Hygiene 33:999-1006 (1984). 코칼 베시쿨로바이러스 외피 유사형 레트로바이러스 벡터 입자는, 예를 들어, 레트로바이러스 Gag, Pol, 및/또는 하나 이상의 부속 단백질(들) 및 코칼 베시쿨로바이러스 외피 단백질을 포함할 수 있는 렌티바이러스, 알파레트로바이러스, 베타레트로바이러스, 감마레트로바이러스, 델타레트로바이러스, 및 엡실론레트로바이러스 벡터 입자를 포함할 수 있다. 이들 구현예의 소정의 양상 내에서, Gag, Pol, 및 부속 단백질은 렌티바이러스 및/또는 감마레트로바이러스이다.

[0796] 일부 구현예에서, 숙주 세포는 본 명세서에 기재된 하나 이상의 벡터로 일시적으로 또는 비일시적으로 형질감염된다. 일부 구현예에서, 세포는 선택적으로 세포에 재도입될 대상체에서 자연적으로 발생함에 따라 형질감염된다. 일부 구현예에서, 형질감염된 세포가 대상체로부터 취해진다. 일부 구현예에서, 세포는 대상체로부터 취한 세포, 예컨대 세포주로부터 유래된다. 조직 배양을 위한 매우 다양한 세포주는 당업계에 공지되어 있다. 세포주의 예는 C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 원숭이 신장 상피, BALB/ 3T3 마우스 배아 섬유아세포, 3T3 Swiss, 3T3-L1, 132-d5 인간 태아 섬유아세포; 10.1 마우스 섬유아세포, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 세포, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY 세포, K562 세포, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 148, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT 세포주, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 세포, Sf-9, SkBr3, T2, T-47D, T84, THP1 세포주, U373, U87, U937, VCaP, Vero 세포, WM39, WT-49, X63, YAC-1, YAR, 및 이들의 유전자이식 변이체를 포함하지만, 이들로 제한되지 않는다. 세포주는 당업자에게 공지된 다양한 공급원으로부터 입수 가능하다 (예를 들어, 미국 미생물 보존 센터 (American Type Culture Collection)(ATCC)(Manassus, Va.)).

[0797] 특정 구현예에서, CRISPR-C2c1 시스템 성분 중 하나 이상의 일시적 발현 및/또는 존재는, 예컨대, 오프-표적 효과를 감소시키는 데 관심이 있을 수 있다. 일부 구현예에서, 본 명세서에 기재된 하나 이상의 벡터로 형질감염된 세포는 하나 이상의 벡터-유래 서열을 포함하는 새로운 세포주를 확립하는 데 사용된다. 일부 구현예에서, 본 명세서에 기재된 바와 같은 CRISPR-C2c1 시스템의 성분으로 (예컨대, 하나 이상의 벡터의 일시적 형질감염, 또는 RNA에 의한 형질감염에 의해) 일시적으로 형질감염되고 CRISPR 복합체 활성을 통해 변형된 세포는 변형을 함유하지만 임의의 다른 외생성 서열을 결여하는 세포를 포함하는 새로운 세포주를 확립하는 데 사용된다. 일부 구현예에서, 본 명세서에 기재된 하나 이상의 벡터, 또는 이러한 세포주로부터 유래된 세포주로 일시적으로 또는 비일시적으로 형질감염된 세포는 하나 이상의 시험 화합물을 평가하는 데 사용된다.

[0798] 일부 구현예에서, RNA 및/또는 단백질을 숙주 세포에 직접적으로 도입하는 것이 계획된다. 예를 들어, CRISPR-Cas 단백질은 시험관내 전사된 가이드 RNA와 함께 코딩 mRNA로서 전달될 수 있다. 이러한 방법은 CRISPR-Cas 단백질의 효과를 보장하기 위한 시간을 감소시키고, 추가로 CRISPR 시스템 성분의 장기간 발현을 방지할 수 있다.

[0799] 일부 구현예에서, 본 발명의 RNA 분자는 리포솜 또는 리포펙틴 제형 등으로 전달되고, 당업자에게 잘 공지된 방법에 의해 제조될 수 있다. 이러한 방법은 예를 들어, 참조로 본 명세서에 편입되는, 예를 들어, 미국 특허 출원 번호 5,593,972, 5,589,466, 및 5,580,859에 기술되어 있다. 포유동물 세포로 siRNA의 증강되고 개선된 전달을 특별히 목적으로 하는 전달 시스템이 개발되었고 (참조: 예를 들어, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-10; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen. 2002, 32: 107-108 및 Simeoni et al., NAR 2003, 31, 11: 2717-2724), 본 발명에 적용할 수 있다. siRNA는 최근에 영장류에서의 유전자 발현의 억제를 위해 성공적으로 사용되었다 (참조: 예를 들어, Tolentino et al., Retina 24(4):660

(본 발명에 또한 적용할 수 있음).

[0800] 사실, RNA 전달은 유용한 생체내 전달 방법이다. 리포솜 또는 입자를 이용하여 세포 내로 CcC1, 아데노신 디아미나제, 및 가이드 RNA를 전달하는 것이 가능하다. 따라서, CRISPR-Cas 단백질, 예컨대, C2c1의 전달, 아데노신 디아미나제 (CRISPR-Cas 단백질 또는 어댑터 단백질에 융합될 수 있음)의 전달, 및/또는 본 발명의 RNA 전달은 RNA 형태로 그리고 미세소포, 리포솜 또는 입자 또는 나노입자들을 통할 수 있다. 예를 들어, C2c1 mRNA, 아데노신 디아미나제 mRNA, 및 가이드 RNA는 생체내 전달을 위해 리포솜 입자 내로 패키징될 수 있다. 리포솜 형질감염 시약, 예컨대, Life Technologies의 Lipofectamine 및 시판 중인 다른 시약은 RNA 분자를 간 내로 효과적으로 전달할 수 있다.

[0801] 또한 바람직한 RNA의 전달 수단은 입자 (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticle for small interfering RNA delivery to epithelial cells, Advanced Functional Materials, 19: 3112- 3118, 2010) 또는 엑소솜 (Schroeder, A., Levins, C., Cortez, C., Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641)을 통한 RNA의 전달을 포함한다. 사실, 엑소솜은 CRISPR 시스템에 일부 병행하는 시스템인 전달 siRNA에서 특히 유용한 것으로 나타났다. 예를 들어, El-Andaloussi S 등 ("Exosome-mediated delivery of siRNA in vitro and in vivo." Nat Protoc. 2012 Dec;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov 15.)은 어떻게 엑소솜이 상이한 생물학적 장벽을 넘어 약물 전달을 위한 유망한 도구이고 시험관내 또는 생체내에서 siRNA의 전달을 활용될 수 있는가를 기술한다. 그들의 접근은 펩티드 리간드와 융합된 엑소솜 단백질을 포함하는, 발현 벡터의 형질감염을 통해 표적화된 엑소솜을 생성하는 것이다. 이어서, 엑소솜은 형질감염된 세포 상청액을 정제하고 이로부터 특징 규명되며, 이어서, RNA는 엑소솜에 로딩된다. 본 발명에 따른 전달 또는 투여는 엑소솜, 특히, 제한없이, 뇌에 의해 수행될 수 있다. 비타민 E (α-토코페롤)는 CRISPR Cas와 접합될 수 있고, 뇌에 고밀도 리포단백질 (HDL)과 함께, 예를 들어, 뇌에 짧은-간섭 RNA (siRNA)를 전달하기 위해 문헌 [Uno et al. (HUMAN GENE THERAPY 22:711- 719 (June 2011))]에 의해 행해지는 것과 동일한 방식으로 전달될 수 있다. 포스페이트-완충 염수 (PBS) 또는 유리 TocsiBACE 또는 Toc-siBACE/HDL을 추전시키고 뇌 Infusion 키트 3 (Alzet)와 연결된 삼투압 미니펌프 (model 1007D; Alzet, Cupertino, CA)를 통해 마우스에 주입된다. 뇌-주입 캐뉼라는 배측 제3 뇌실 내로 주입을 위해 정중선에서 브레그마 뒤쪽 약 0.5㎜에 놓인다. Uno 등은 HDL과 함께 3 nmol 정도로 적은 Toc-siRNA가 동일한 ICV 주입 방법에 의해 비슷한 정도로 표적 감소를 유도할 수 있었다는 것을 발견하였다. α-토코페롤에 접합되고, 뇌에 표적화된 HDL과 공동 투여되는 CRISPR Cas의 유사한 용량이 본 발명의 인간에 대해 고려될 수 있으며, 예를 들어, 약 3 nmol 내지 약 3 μmol의 뇌에 표적화되는 CRISPR Cas가 고려될 수 있다. Zou 등 ((HUMAN GENE THERAPY 22:465- 475 (April 2011))은 래트의 척수에서 침묵화된 생체내 유전자에 대해 PKCγ를 표적화하는 짧은-헤어핀 RNA의 렌티바이러스-매개 전달 방법을 기재한다. Zou 등은 척수강내 카테터에 의해 1×109개의 형질도입 유닛 (TU)/㎖의 역가를 갖는 약 10 ㎕의 재조합 렌티바이러스를 투여하였다. 뇌에 표적화된 렌티바이러스 벡터에서 발현되는 유사한 투약량의 CRISPR Cas가 본 발명에서 인간에 대해 상정될 수 있으며, 예를 들어, 1×109개의 형질도입 단위(TU)/㎖의 역가를 갖는 렌티바이러스에서 뇌에 표적화된 약 10 내지 50㎖의 CRISPR Cas가 상정될 수 있다.

[0802] **벡터의 용량**

[0803] 일부 구현예에서, 벡터, 예를 들어, 플라스미드 또는 바이러스 벡터는 관심 조직에, 예를 들어, 근육내 주사에 의해 전달되는 한편, 다른 시기에 정맥내, 경피, 비강내, 경구, 점막 또는 다른 전달 방법을 통해 전달된다. 이러한 전달은 단일 용량 또는 다수 용량을 통할 수 있다. 당업자는 본 명세서에서 전달하려는 실제 용량이 다양한 인자, 예컨대, 벡터 선택, 표적 세포, 유기체, 또는 조직, 치료하려는 대상체의 일반적 병태, 추구되는 형질전환/변형 정도, 투여 경로, 투여 방식, 추구되는 형질전환/변형 유형 등에 따라 크게 다를 수 있다는 것을 이해한다.

[0804] 이러한 용량은, 예를 들어, 담체 (물, 식염수, 에탄올, 글리세롤, 락토스, 수크로스, 인산칼슘, 젤라틴, 덱스트란, 한천, 펙틴, 땅콩유, 참깨유 등), 희석제, 약학적으로 허용가능한 담체 (예를 들어, 포스페이트-완충 식염수), 약학적으로 허용가능한 부형제 및/또는 당업계에 공지된 다른 화합물을 추가로 함유할 수 있다. 용량은 하나 이상의 약학적으로 허용가능한 염 예컨대, 예를 들어, 미네랄산 염 예컨대 히드로크롤라이드, 히드로브로마이드, 포스페이트, 술페이트 등; 및 유기산의 염 예컨대 아세테이트, 프로피오네이트, 말로네이트, 벤조에이트 등을 더 함유할 수 있다. 추가로, 보조 물질, 예컨대 습윤제 또는 유화제, pH 완충 물질, 겔 또는 겔화 재료, 풍미제, 착색제, 미세구, 중합체, 현탁제 등이 또한 본 명세서에서 존재할 수 있다. 추가로, 특히

제형이 재구성 가능한 형태라면, 1종 이상의 다른 통상적인 약제 성분, 예컨대, 보존제, 습윤제, 현탁제, 계면활성제, 항산화제, 고결방지제, 충전제, 킬레이트제, 코팅제, 화학적 안정제 등이 또한 제공될 수 있다. 적합한 예시적 성분은 미정질 셀룰로스, 카복시메틸셀룰로스 소듐, 폴리솔베이트 80, 페닐에틸 알콜, 클로로부탄올, 솔브산칼륨, 솔브산, 이산화황, 갈산프로필, 파라벤, 에틸 바닐린, 글리세린, 페놀, 파라클로로페놀, 젤라틴, 알부민 및 이들의 조합물을 포함한다. 약학적으로 허용가능한 부형제의 완전한 논의는 참조로 본 명세서에 편입되는, [REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991)]에서 입수가능하다.

[0805] 본 명세서의 일 구현예에서, 전달은 아데노바이러스를 통한 것으로, 이는 적어도 1×10^{5} 개의 입자(입자 단위, pu로서도 지칭됨)의 아데노바이러스 벡터를 함유하는 단일 부스터 용량일 수 있다. 본 명세서의 구현예에서, 용량은 바람직하게는 적어도 약 1 x 10^{6} 개 입자(예를 들어, 약 1 x 10^{6}-1 x 10^{12} 개 입자), 더 바람직하게는 적어도 약 1 x 10^{7} 개 입자, 더 바람직하게는 적어도 약 1 x 10^{8} 개 입자(예를 들어, 약 1 x 10^{8}-1 x 10^{11} 개 입자 또는 약 1 x 10^{8}-1 x 10^{12} 개 입자), 및 가장 바람직하게는 적어도 약 1 x 100 개 입자(예를 들어, 약 1 x 10^{9}-1 x 10^{10} 개 입자 또는 약 1 x 10^{9}-1 x 10^{12} 개 입자), 또는 심지어 적어도 약 1 x 10^{10} 개 입자(예를 들어, 약 1 x 10^{10}-1 x 10^{12} 개 입자)의 아데노바이러스 벡터이다. 대안적으로, 용량은 1 x 10^{14} 개 이하의 입자, 바람직하게는 약 1 x 10^{13} 개 이하의 입자, 훨씬 더 바람직하게는 약 1 x 10^{12} 개 이하의 입자, 훨씬 더 바람직하게는 약 1 x 10^{11} 개 이하의 입자, 및 가장 바람직하게는 약 1 x 10^{10} 개 이하의 입자(예를 들어, 약 1 x 10^{9} 개 이하의 입자)를 포함한다. 따라서, 용량은, 예를 들어 약 1 x 10^{6} 개 입자 단위(pu), 약 2 x 10^{6} pu, 약 4 x 10^{6} pu, 약 1 x 10^{7} pu, 약 2 x 10^{7} pu, 약 4 x 10^{7} pu, 약 1 x 10^{8} pu, 약 2 x 10^{8} pu, 약 4 x 10^{8} pu, 약 1 x 10^{9} pu, 약 2 x 10^{9} pu, 약 4 x 10^{9} pu, 약 1 x 10^{10} pu, 약 2 x 10^{10} pu, 약 4 x 10^{10} pu, 약 1 x 10^{11} pu, 약 2 x 10^{11} pu, 약 4 x 10^{11} pu, 약 1 x 10^{12} pu, 약 2 x 10^{12} pu, 또는 약 4 x 10^{12} pu의 아데노바이러스 벡터에 의한 단일 용량의 아데노바이러스 벡터를 함유할 수 있다. 예를 들어, 본 명세서에 참고로 편입된 2013년 6월 4일자로 등록된 Nabel 등의 미국 특허 제8,454,972 B2호의 아데노바이러스 벡터; 및 이의 29 칼럼, 36 내지 58 라인의 용량을 참조한다. 본 명세서의 구현예에서, 아데노바이러스는 다수 용량을 통해 전달된다.

[0806] 본 명세서의 구현예에서, 상기 전달은 AAV를 통한다. 인간에 대한 AAV의 생체내 전달을 위한 치료적 유효량은 약 1 x 10^{10} 내지 약 1 x 10^{10} 기능성 AAV/ml 용액을 함유하는 식염수 용액의 약 20 내지 약 50 ml의 범위에 있는 것으로 믿어진다. 투약량은 임의의 부작용에 대해 치료적 이점의 균형을 맞추도록 조절될 수 있다. 본 명세서의 구현예에서, AAV 용량은 일반적으로 약 1 x 10^{5} 내지 1 x 10^{50} 개 게놈 AAV, 약 1 x 10^{8} 개 내지 1 x 10^{20} 개 게놈 AAV, 약 1 x 10^{10} 내지 약 1 x 10^{16} 개 게놈, 또는 약 1 x 10^{11} 내지 약 1 x 10^{16} 개 게놈 AAV의 농도 범위에 있다. 인간 투약량은 약 1 x 10^{13} 개 게놈 AAV일 수 있다. 이러한 농도는 약 0.001 mℓ 내지 약 100 mℓ, 약 0.05 내지 약 50 mℓ, 또는 약 10 내지 약 25 mℓ의 담체 중에 전달될 수 있다. 다른 유효 용량은 용량 반응 곡선을 확립하는 일상적 시행을 통해 당업자에 의해 용이하게 확립될 수 있다. 예를 들어, 2013년 3월 26일자로 등록된 Hajjar 등의 미국 특허 제8,404,658 B2호, 27 칼럼, 45 내지 60행 참조한다.

[0807] 본 명세서의 구현예에서, 전달은 플라스미드를 통한다. 이러한 플라스미드 조성물에서, 용량은 반응을 유발하는 충분한 양의 플라스미드이어야 한다. 예를 들어, 플라스미드 조성물 내 적합한 양의 플라스미드 DNA는 70 kg 개체당 약 0.1 mg 내지 약 2 mg, 또는 약 1 ㎍ 내지 약 10 ㎍일 수 있다. 본 발명의 플라스미드는 일반적으로 (i) 프로모터; (ii) 상기 프로모터에 작동적으로 연결된 CRISPR-Cas 단백질을 코딩하는 서열; (iii) 선택 가능한 마커; (iv) 복제 기원; 그리고 (v) (ii)의 하류에 작동적으로 연결된 전사 종결자를 포함할 것이다. 플라스미드는 또한 CRISPR 복합체의 RNA 성분을 코딩할 수 있지만, 이들 중 하나 이상은 대신에 상이한 벡터 상에서 코딩될 수 있다.

[0808] 본 명세서의 용량은 평균 70 kg의 개체를 기준으로 한다. 투여 빈도는 의학 또는 수의학 종사자(예를 들어, 의사, 수의사) 또는 당업계의 숙련된 과학자의 영역 내이다. 또한 실험에서 사용되는 마우스는 전형적으로 약 20g이고, 마우스 실험으로부터 70 kg까지의 개체로 확대될 수 있다는 것이 주목된다.

[0809] 본 명세서에 제공되는 조성물에 대해 사용되는 용량은 반복된 투여 또는 반복 용량을 위한 용량을 포함한다. 특정 구현예에서, 투여는 몇 주, 몇 개월 또는 몇 년의 기간 내에 반복된다. 적합한 어세이는 최적의 용량 요법을 얻기 위해 수행될 수 있다. 반복 투여는 오프-표적 변형에 긍정적으로 영향을 미칠 수 있는 더 낮은 용량의 사용을 허용할 수 있다.

[0810] **RNA 전달**

[0811] 특정 구현예에서, RNA 기반 전달이 사용된다. 이들 구현예에서, CRISPR-Cas 단백질의 mRNA, 아데노신 디아미나제의 mRNA(CRISPR-Cas 단백질 또는 어댑터에 융합될 수 있음)는 시험관내 전사된 가이드 RNA와 함께 전달된다. Liang 등은 RNA 기반 전달을 이용하는 효율적인 게놈 편집을 기재한다 (단백질 Cell. 2015 May; 6(5): 363- 372). 일부 구현예에서, C2c1 및/또는 아데노신 디아미나제를 코딩하는 mRNA(들)는 화학적으로 변형될 수 있는데, 이는 플라스미드-코딩된 C2c1 및/또는 아데노신 디아미나제에 비해 개선된 활성을 야기할 수 있다. 예를 들어, mRNA(들)에서 우리딘은 슈도우리딘 (Ψ), N1-메틸슈도우리딘 (me1Ψ), 5-메톡시우리딘 (5moU)으로 부분적으로 또는 완전히 치환될 수 있다. 문헌 [Li et al., Nature Biomedical Engineering 1, 0066 DOI:10.1038/s41551-017-0066 (2017)]을 참조하며, 이의 전문은 본 명세서에 참고로 편입된다.

[0812] **전달 접근법의 예**

[0813] <u>RNP</u>

[0814] 특정 구현예에서, 사전-복합체화된 가이드 RNA, CRISPR-Cas 단백질 및 아데노신 디아미나제 (CRISPR-Cas 단백질 또는 어댑터에 융합될 수 있음)는 리보핵단백질 (RNP)로서 전달된다. RNP는 RNA 방법보다 훨씬 더 빠른 편집 효과를 야기한다는 장점을 갖는데, 이 과정이 전사에 대한 필요를 피하기 때문이다. 중요한 이점은 RNP 전달이 둘 모두 일시적이어서, 오프-표적 효과 및 독성 문제를 감소시킨다는 것이다. 상이한 세포 유형에서 효율적인 게놈 편집은 Kim et al. (2014, Genome Res. 24(6):1012-9), Paix et al. (2015, Genetics 204(1):47-54), Chu et al. (2016, BMC Biotechnol. 16:4), 및 Wang et al. (2013, Cell. 9;153(4):910-8)가 관찰하였다.

[0815] 특정 구현예에서, 리보핵단백질은 WO2016161516에 기재된 바와 같은 폴리펩티드-기반 셔틀 작용제의 방법에 의해 전달된다. WO2016161516은 세포 침투성 도메인 (CPD)에, 히스티딘-풍부 도메인 및 CPD에 작동적으로 연결된 엔도솜 누출 도메인 (ELD)을 포함하는 합성 펩티드를 이용하는 폴리펩티드 카고의 효율적인 형질도입을 기재한다. 유사하게, 이들 폴리펩티드는 진핵생물 세포에서 CRISPR-이펙터 기반 RNP의 전달을 위해 사용될 수 있다.

[0816] <u>입자</u>

[0817] 일부 양상 또는 구현예에서, 전달 입자 제형을 포함하는 조성물이 사용될 수 있다. 일부 양상 또는 구현예에서, 제형은 CRISPR 복합체를 포함하며, 상기 복합체는 표적 서열에 대한 CRISPR 복합체의 서열 특이적 결합을 지시하는 가이드 및 CRISPR 단백질을 포함한다. 일부 구현예에서, 전달 입자는 지질계 입자, 선택적으로 지질 나노입자, 또는 양이온성 지질 및 선택적으로 생분해성 중합체를 포함한다. 일부 구현예에서, 양이온성 지질은 1,2-다이올레오일-3-트라이메틸암모늄-프로판 (DOTAP)을 포함한다. 일부 구현예에서, 친수성 중합체는 에틸렌 글리콜 또는 폴리에틸렌 글리콜을 포함한다. 일부 구현예에서, 전달 입자는 지질단백질, 바람직하게는 콜레스테롤을 추가로 포함한다. 일부 구현예에서, 전달 입자는 직경이 500 nm 미만, 선택적으로 직경이 250 nm 미만, 선택적으로 직경이 100 nm 미만, 선택적으로 직경이 약 35 nm 내지 약 60 nm이다.

[0818] 입자 전달 시스템 및/또는 제형의 몇몇 유형은 다양한 범위의 생의학적 적용에서 유용한 것으로 알려져 있다. 일반적으로, 입자는 그의 수송 및 특성에 관해 전체 단위로서 거동하는 소형 물체로서 정의된다. 입자는 추가로 직경에 따라 분류된다. 조악한 입자는 2,500 내지 10,000 나노미터 범위를 포괄한다. 미세한 입자는 크기가 100 내지 2,500 나노미터이다. 초미세 입자, 또는 나노입자는 일반적으로 크기가 1 내지 100 나노미터이다. 100 nm 한계 기준은 벌크 물질을 입자와 구별하는 신규 성질이 전형적으로 100 nm 이하의 임계 길이 규모에서 발생된다는 사실이다.

[0819] 본 명세서에서 사용되는 입자 전달 시스템/제형은 본 발명에 따른 입자를 포함하는 임의의 생물학적 전달 시스템/제형으로서 정의된다. 본 발명에 따른 입자는 100 마이크론 (㎛) 미만의 최고 치수 (즉, 직경)을 갖는 임의의 독립체이다. 일부 구현예에서, 본 발명의 입자는 10 ㎛ 미만의 가장 큰 치수를 갖는다. 일부 구현예에서, 본 발명의 입자는 2000 나노미터 (nm) 미만의 최고 치수를 갖는다. 일부 구현예에서, 본 발명의

입자는 1000 나노미터 (nm) 미만의 최고 치수를 갖는다. 일부 구현예에서, 본 발명의 입자는 900 nm, 800 nm, 700 nm, 600 nm, 500 nm, 400 nm, 300 nm, 200 nm, 또는 100 nm 미만의 최고 치수를 갖는다. 전형적으로, 본 발명의 입자는 500 nm 이하의 최고 치수 (예를 들어, 직경)를 갖는다. 일부 구현예에서, 본 발명의 입자는 250 nm 이하의 최고 치수 (예를 들어, 직경)를 갖는다. 일부 구현예에서, 본 발명의 입자는 200 nm 이하의 최고 치수 (예를 들어, 직경)를 갖는다. 일부 구현예에서, 본 발명의 입자는 150 nm 이하의 최고 치수 (예를 들어, 직경)를 갖는다. 일부 구현예에서, 본 발명의 입자는 100 nm 이하의 최고 치수 (예를 들어, 직경)를 갖는다. 예를 들어, 50 nm 이하의 최고 치수를 갖는 보다 작은 입자가 본 발명의 일부 구현예에서 사용된다. 일부 구현예에서, 본 발명의 입자는 25nm 내지 200nm의 범위의 최고 치수를 갖는다.

[0820] 본 발명에 관해, 나노입자 또는 지질 외피를 이용하여 전달되는 CRISPR 복합체, 예를 들어, CRISPR-Cas 단백질 또는 mRNA, 또는 아데노신 디아미나제 (CRISPR-Cas 단백질 또는 어댑터에 융합될 수 있음) 또는 mRNA, 또는 가이드 RNA 중 하나 이상의 성분을 갖는 것이 바람직하다. 다른 전달 시스템 또는 벡터가 본 발명의 입자 측면과 함께 사용될 수 있다.

[0821] 일반적으로, "나노입자" 는 1000 nm 미만의 직경을 갖는 입의의 입자를 의미한다. 바람직한 일정 구현예에서, 본 발명의 나노입자는 500 nm 이하의 가장 큰 치수(예를 들어, 직경)를 갖는다. 다른 바람직한 구현예에서, 본 발명의 나노입자는 25 nm 내지 200 nm 범위의 가장 큰 치수를 갖는다. 다른 구현예에서, 본 발명의 나노입자는 100 nm 이하의 가장 큰 치수를 갖는다. 다른 구현예에서, 본 발명의 나노입자는 35 nm 내지 60 nm 범위의 가장 큰 치수를 갖는다. 입자 또는 나노입자에 대한 언급은 적절한 경우, 상호 호환 가능할 수 있다는 것이 인식될 것이다.

[0822] 입자의 크기는 로딩 전에 측정되는지 또는 로딩 후에 측정되는지에 따라 상이하다는 것이 이해될 것이다. 따라서, 특정 구현예에서, 용어 " 나노입자" 는 로딩 전 입자에 대해서만 적용될 수 있다.

[0823] 본 발명에 포함된 입자는 상이한 형태, 예를 들어 고체 입자(예를 들어, 금속, 예컨대 은, 금, 철, 티타늄), 비-금속, 지질-기반 고체, 중합체), 입자의 현탁액, 또는 이들의 조합물로 제공될 수 있다. 금속, 유전체, 및 반도체 입자뿐만 아니라 하이브리드 구조(예를 들어, 코어-쉘 입자)가 제조될 수 있다. 반도체 물질로 제조된 입자는 또한 이들이 전자 에너지 수준의 양자화가 발생하기에 충분히 작은(통상적으로, 10 nm 이하) 경우 양자점으로 표지화될 수 있다. 이러한 나노규모 입자는 약물 담체 또는 영상화 작용제로서 생물의학 응용에서 사용되며, 본 발명에서 유사한 목적을 위해 적합하게 될 수 있다.

[0824] 반고체 및 연성 입자가 제조되었고, 이들은 본 발명의 범주 내에 속한다. 반고체 성질의 프로토타입 입자는 리포솜이다. 다양한 유형의 리포솜 입자가 항암 약물 및 백신을 위한 전달 시스템으로서 현재 임상적으로 사용된다. 절반이 친수성이고, 나머지 절반이 소수성인 입자는 야누스 입자로 명명되며, 에멀젼을 안정화시키는데 특히 효과적이다. 이들은 물/오일 계면에서 자가-조립되며, 고체 계면활성제로 작용할 수 있다.

[0825] 입자 특징규명 (예를 들어, 형태, 치수 등의 특징규명을 포함)은 다양한 상이한 기법을 사용하여 행해진다. 통상적인 기법은 전자 현미경 (TEM, SEM), 원자력 현미경 (AFM), 동적광산란 (DLS), X-선 광전자 분광법 (XPS), 분말 X-선 회절 (XRD), 푸리에 변환 적외선 분광법 (FTIR), 매트릭스-보조 레이저 탈착/이온화 시간 비행 질량분석법 (MALDI-TOF), 자외선-가시광선 분광학, 이중 극성화 간섭계법 및 핵자기공명 (NMR)이다. 특징규명 (치수 측정)은 본 발명의 임의의 시험관내 생체외 및/또는 생체내 적용을 위해서 전달을 위한 최적 크기의 입자를 제공하기 위해, 천연 입자 (즉, 로딩 전)로서 또는 카고의 로딩 이후 (여기서 카고는 예를 들어, CRISPR-Cas 시스템 중 하나 이상의 성분, 예를 들어, CRISPR-Cas 단백질 또는 mRNA, 아데노신 디아미나제 (CRISPR-Cas 단백질 또는 어댑터에 융합될 수 있음) 또는 mRNA, 또는 가이드 RNA, 또는 이들의 임의의 조합물을 의미하고, 추가 담체 및/또는 부형제를 포함할 수 있음)에 수행될 수 있다. 바람직한 일정 구현예에서, 입자 치수 (예를 들어, 직경) 특징규명은 동적 레이저 산란 (DLS)을 이용하는 측정에 기반한다. 입자, 그들의 제조 및 사용 방법 및 이의 측정에 관해서, US 특허 번호 8,709,843; US 특허 번호 6,007,845; US 특허 번호 5,855,913; US 특허 번호 5,985,309; US 특허 번호 5,543,158; 및 [James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014)의 공개물 (2014년 5월 11일 온라인 공개), doi:10.1038/nnano.2014.84를 참조한다.

[0826] 본 발명의 범주 내의 입자 전달 시스템은 고체, 반고체, 에멀션 또는 콜로이드 입자를 포함하지만, 이들로 제한되지 않는 임의의 형태로 제공될 수 있다. 이와 같이, 예를 들어, 지질-기반 시스템, 리포솜, 마이셀, 미세소포, 엑소솜, 또는 유전자 총을 포함하지만, 이들로 제한되지 않는 본 명세서에 기재된 임의의 전달 시스템은 본 발명의 범주 내에서 입자 전달 시스템으로서 제공될 수 있다.

[0827] CRISPR-Cas 단백질 mRNA, 아데노신 디아미나제 (CRISPR-Cas 단백질 또는 어댑터에 융합될 수 있음) 또는 mRNA, 및 가이드 RNA는 입자 또는 지질 외피를 이용하여 동시에 전달될 수 있고; 예를 들어, 본 발명의 CRISPR-Cas 단백질 및 RNA는, 예를 들어, 복합체로서, Dahlman 등의 WO2015089419 A2 및 이에 인용된 문헌에서와 같은 입자, 예컨대, 7C1 (예를 들어, 문헌 [James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014)] 참조, 2014년 5월 11일자에 공개, doi:10.1038/nnano.2014.84)을 통해 전달될 수 있고, 예를 들어, 전달 입자는 지질 또는 리피도이드 및 친수성 중합체, 예를 들어, 양이온성 지질 및 친수성 중합체를 포함하고, 예를 들어 여기서 양이온성 지질은 1,2-이올레오일-3-트리메틸암모늄-프로판 (DOTAP) 또는 1,2-디테트라데카노일-sn-글리세로-3-포스포콜린 (DMPC)을 포함하고/하거나 친수성 중합체는 에틸렌 글리콜 또는 폴리에틸렌 글리콜 (PEG)을 포함하고/하거나, 입자는 콜레스테롤 (예를 들어, 제제 1의 입자 = DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0; 제제 번호 2 = DOTAP 90, DMPC 0, PEG 10, 콜레스테롤 0; 제제 번호 3 = DOTAP 90, DMPC 0, PEG 5, 콜레스테롤 5)을 더 포함하고, 입자는 효율적인, 다단계 과정을 사용해 형성되는데, 먼저, 이펙터 단백질 및 RNA는, 예를 들어, 1:1 몰비로, 예를 들어, 실온에서, 예를 들어, 30분 동안, 예를 들어, 멸균, 뉴클레아제 무함유 1X PBS에서 함께 혼합되고, 별개로, 제제에 적용가능하다면 DOTAP, DMPC, PEG, 및 콜레스테롤은 알콜, 예를 들어, 100% 에탄올에 용해되어서, 2개 용액을 함께 혼합하여 복합체를 함유하는 입자를 형성시킨다).

[0828] 핵산-표적화 이펙터 단백질 (예를 들어, V형 단백질, 예컨대, C2c1) mRNA 및 가이드 RNA는 입자 또는 지질 외피를 이용하여 동시에 전달될 수 있다. 적합한 입자의 예는 미국 특허 제9,301,923호에 기재된 것을 포함하지만, 이들로 제한되지 않는다.

[0829] 예를 들어, 문헌[Su X, Fricke J, Kavanagh DG, Irvine DJ (" In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. 2011 Jun 6;8(3):77487. Epub 2011 Apr 1)]은 인지질 이중층 쉘에 의해 둘러싸인 폴리(β-아미노 에스테르)(PBAE) 코어를 지니는 생분해성 코어-쉘 구조의 나노입자를 기재한다. 이들은 생체내 mRNA 전달을 위해 개발되었다. pH-반응성 PBAE 성분은 엔도솜 하괴를 촉진하도록 선택되었고, 반면 지질 표면층은 다가양이온 코어의 독성을 최소화하도록 선택되었다. 그러므로, 이것은 본 발명의 RNA 전달에 바람직하다.

[0830] 일 구현예에서, 모두 뇌에 대한, 펩티드의 경구 전달, 펩티드의 정맥내 전달 및 펩티드의 비강 전달에 적용할 수 있는, 자기 조립형 생체부착 중합체 기반 입자/나노입자가 고려된다. 다른 구현예, 예컨대, 경구 흡수 및 소수성 약물의 안구 전달이 또한 고려된다. 분자 엔벨로프 기술은 보호되고 질환 부위에 전달되는 조작된 중합체를 포함한다 (참조: 예를 들어, Mazza, M. et al. ACSNano, 2013. 7(2): 1016-1026; Siew, A., et al. Mol Pharm, 2012. 9(1):14-28; Lalatsa, A., et al. J Contr Rel, 2012. 161(2):523-36; Lalatsa, A., et al., Mol Pharm, 2012. 9(6):1665-80; Lalatsa, A., et al. Mol Pharm, 2012. 9(6):1764-74; Garrett, N.L., et al. J Biophotonics, 2012. 5(5-6):458-68; Garrett, N.L., et al. J Raman Spect, 2012. 43(5):681-688; Ahmad, S., et al. J Royal Soc Interface 2010. 7:S423-33; Uchegbu, I.F. Expert Opin Drug Deliv, 2006. 3(5):629-40; Qu, X.,et al. Biomacromolecules, 2006. 7(12):3452-9 및 Uchegbu, I.F., et al. Int J Pharm, 2001. 224:185-199). 표적 조직에 따라 단일 또는 다수 용량으로, 약 5mg/kg의 용량이 고려된다.

[0831] 일 구현예에서, MIT의 단 앤더슨 랩(Dan Anderson's lab)에 의해 개발된 종양 성장을 중단시키기 위해 암 세포에 RNA를 전달할 수 있는 입자/나노입자가 사용되고/되거나 본 발명의 AD-작용화된 CRISPR-Cas 시스템에 적합하게 될 수 있다. 특히, 앤더슨 랩은 새로운 생체물질 및 나노제형의 합성, 정제, 특성규명 및 제형화를 위한 완전히 자동화된, 조합 시스템을 개발하였다. 참조, 예를 들어, Alabi et al., Proc Natl Acad Sci U S A. 2013 Aug 6;110(32):12881-6; Zhang et al., Adv Mater. 2013 Sep 6;25(33):4641-5; Jiang et al., Nano Lett. 2013 Mar 13;13(3):1059-64; Karagiannis et al., ACS Nano. 2012 Oct 23;6(10):8484-7; Whitehead et al., ACS Nano. 2012 Aug 28;6(8):6922-9 및 Lee et al., Nat Nanotechnol. 2012 Jun 3;7(6):389-93.

[0832] 미국 공개 특허 출원 제20110293703호는 본 발명의 AD-기능화된 CRISPR-Cas 시스템을 전달하는 데 적용될 수 있는 폴리뉴클레오티드의 투여에 특히 유용한 리피도이드 화합물에 관한 것이다. 일 양상에서, 아미노알콜 리피도이드 화합물은 세포 또는 대상체에 전달하려는 작용제와 조합되어 미세입자, 나노입자, 리포솜 또는 마이셀을 형성한다. 입자, 리포솜 또는 마이셀에 의해 전달하려는 작용제는 기체, 액체 또는 고체의 형태일 수 있고, 작용제는 폴리뉴클레오티드, 단백질, 펩티드 또는 소형 분자일 수 있다. 아미노알콜 리피도이드 화합물은 다른 아미노알콜 리피도이드 화합물, 중합체 (합성 또는 천연), 계면활성제, 콜레스테롤, 탄수화물, 단백질, 지질 등과 조합되어 입자를 형성할 수 있다. 이어서, 이들 입자는 선택적으로 약학 부형제와 조합되어 약학 조성물을 형성할 수 있다.

[0833] 미국 공개 특허 출원 제20110293703호는 또한 아미노알콜 리피도이드 화합물의 제조 방법을 제공한다. 아민의 하나 이상의 등가물은 적합한 조건 하에 에폭시드-종결 화합물의 하나 이상의 등가물과 반응되어 본 발명의 아미노알콜 리피도이드 화합물을 형성한다. 일정 구현예에서, 아민의 모든 아미노기는 에폭시드-종결 화합물과 완전히 반응되어 3차 아민을 형성한다. 다른 구현예에서, 아민의 모든 아미노기는 에폭시드-종결 화합물과 완전히 반응되지 않아서 3차 아민을 형성하고, 이에 의해 아미노알콜 리피도이드 화합물에서 1차 또는 2차 아민을 초래한다. 이들 1차 또는 2차 아민은 그대로 남아있거나 또는 상이한 에폭시드-종결 화합물과 같은 다른 친핵체와 반응될 수 있다. 당업자에 의해 이해되는 바와 같이, 과량 미만의 에폭시드-종결 화합물과 아민의 반응은 다양한 수의 꼬리부를 갖는 복수의 상이한 아미노알콜 리피도이드 화합물을 생성시키게 될 것이다. 일정 아민은 2개의 에폭시드-유래 화합물 꼬리부에 의해 완전히 작용화될 수 있는 반면, 다른 분자는 에폭시드-유래 화합물 꼬리부에 의해 완전히 작용화되지 않을 것이다. 예를 들어, 디아민 또는 폴리아민은 분자의 다양한 아미노 모이어티로부터 1, 2, 3 또는 4개의 에폭시드-유래 화합물 꼬리부를 포함하여, 1차, 2차 및 3차 아민을 초래할 수 있다. 일정 구현예에서, 모든 아미노기는 완전히 작용화되지 않는다. 일정 구현예에서, 2종의 동일한 유형의 에폭시드-종결된 화합물이 사용된다. 다른 구현예에서, 2종 이상의 상이한 에폭시드-종결된 화합물이 사용된다. 아미노알콜 리피도이드 화합물의 합성은 용매와 함께 또는 용매 없이 수행되고, 합성은 30 내지 100℃, 바람직하게는 대략 50 내지 90℃ 범위의 고온에서 수행될 수 있다. 제조된 아미노알콜 리피도이드 화합물은 임의로 정제될 수 있다. 예를 들어, 아미노알콜 리피도이드 화합물의 혼합물은 정제되어 특정 수의 에폭시드-유래 화합물 꼬리부를 갖는 아미노알콜 리피도이드 화합물을 수득할 수 있다. 또는 혼합물은 정제되어 특정 입체- 또는 위치이성질체를 수득할 수 있다. 아미노알콜 리피도이드 화합물은 또한 알킬 할로겐화물 (예를 들어, 메틸 아이오다이드) 또는 다른 알킬화제를 이용하여 알킬화될 수 있고/있거나 그들은 아실화될 수 있다.

[0834] 미국 특허 공개 제20110293703호는 또한 본 발명의 방법에 의아 제조된 아미노알콜 리피도이드 화합물의 라이브러리를 제공한다. 이들 아미노알콜 리피도이드 성분은 액체 취급기, 로봇, 마이크로타이터 플레이트, 컴퓨터 등을 포함한 고속-대용량 기술을 사용해 제조 및/또는 스크리닝될 수 있다. 일정 구현예에서, 아미노알콜 리피도이드 화합물은 세포로 폴리뉴클레오티드 또는 다른 작용제 (예를 들어, 단백질, 펩티드, 소형 분자)를 전달하는 그들 능력에 대해 스크리닝된다.

[0835] 미국 특허 공개 제20130302401호는 조합적 중합화를 사용하여 제조된 폴리(베타-아미노 알코올)(PBAA)의 부류에 관한 것이다. 본 발명의 PBAA는 생체기법 및 생의학적 적용분야에서 코팅 (예컨대, 의학적 장치 또는 이식물에 대한 필름의 코팅 또는 다중층 필름), 첨가제, 물질, 부형제, 비-바이오파울링제 (non-biofouling agent), 미세패턴화제 및 세포의 캡슐화제로서 사용될 수 있다. 표면 코팅제로서 사용될 때, 이들 PBAA는 그들의 화학 구조에 따라서 시험관내와 생체내 둘 모두에서 상이한 염증 수준을 유발하였다. 이 부류의 물질의 더 큰 화학적 다양성은 본 발명자들이 시험관내 거대세포 활성화를 억제하는 중합체 코팅을 확인할 수 있게 하였다. 더 나아가, 이들 코팅은 염증 세포의 동원을 감소시키고, 카복실화된 폴리스티렌 미세입자의 피하 이식 후에 섬유증을 감소시킨다. 이들 중합체는 세포 캡슐화를 위해 다가전해질 복합체 캡슐을 형성하는 데 사용될 수 있다. 본 발명은 또한 다수의 다른 생물학적 적용분야, 예컨대, 항미생물 코팅, DNA 또는 siRNA 전달 및 줄기 세포 조직 조작을 가질 수 있다. 미국 공개 특허 출원 제20130302401호의 교시는 본 발명의 AD-기능화된 CRISPR-Cas 시스템에 적용될 수 있다.

[0836] C2c1, (C2c1 또는 어댑터 단백질에 융합될 수 있는) 아데노신 디아미나제, 및 가이드 RNA를 포함하는 사전 조립된 재조합 CRISPR-Cas 복합체는 예를 들어 전기천공에 의해 형질감염되어, 높은 돌연변이율 및 검출 가능한 오프-표적 돌연변이의 부재를 초래한다. Hur, J.K. et al, Targeted mutagenesis in mice by electroporation of C2c1 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596.

[0837] 뇌로의 국소 전달과 관련하여, 이것은 다양한 방식으로 달성될 수 있다. 예를 들어, 재료는 선조체내로, 예를 들어, 주사를 통해 전달될 수 있다. 주사는 개두술을 통해 정위적으로 수행될 수 있다.

[0838] 일부 구현예에서, 당-기반 입자, 예를 들어 GalNAc은 본 명세서에 기재된 바와 같이 사용될 수 있으며, 특히 전달에 관해 WO2014118272(본 명세서에 참고로 편입됨) 및 문헌 [Nair, JK et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961)] 및 본 명세서의 교시를 참고로 하여 달리 명확하지 않다면, 모든 입자에 적용된다. 이는 당-기반 입자인 것으로 고려될 수 있고, 다른 입자 전달 시스템 및/또는 제형에 대한 추가적인 상세한 설명이 본 명세서에 제공된다. 그러므로, GalNAc는 본 명세서에 기술된 다른 입자의 의미에서 입자라고 간주될 수 있어서, 일반적인 용도 및 다른 고려사항, 예를 들어 상기 입자의 전달이 역시 GalNAc 입자에 적용된다. 5'-헥실아미노 변형된 올리고뉴클레오티드(5'-HA ASO, 분자량 대략 8000Da;

문헌[et al., Bioconjugate Chem., 2015, 26 (8), pp 1451-1455])에 대해 PFP(펜타플루오로페닐) 에스터로서 활성화된, 예를 들어, 3안테나 GalNAc 클러스터(분자량 대략 2000)를 부착하기 위하여 용액상 접합 전략이 사용될 수 있다. 유사하게, 생체내 핵산 전달을 위한 폴리(아크릴레이트) 중합체가 기술된 바 있다 (참조로 본 명세서에 편입되는 WO2013158141 참조). 추가적인 대안의 구현예에서, CRISPR 나노 입자 (또는 단백질 복합체)와 천연 발생 혈청 단백질의 사전 혼합이 전달을 개선시키기 위해 사용될 수 있다 (Akinc A et al, 2010, Molecular Therapy vol. 18 no. 7, 13571364).

[0839] **나노클루 (Nanoclew)**

[0840] 또한, CRISPR-C2c1 시스템은 예를 들어, 하기 문헌 들에 기술된 나노클루를 사용해 전달될 수 있다: Sun W et al, Cocoon-유사 self-degradable DNA nanoclew for anticancer drug delivery., J Am Chem Soc. 2014 Oct 22;136(42):14722-5. doi: 10.1021/ja5088024. Epub 2014 Oct 13.; 또는 Sun W et al, Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing., Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. doi: 10.1002/anie.201506030. Epub 2015 Aug 27.

[0841] **지질 입자**

[0842] 일부 구현예에서, 전달은 지질 입자, 예컨대, LNP에 C2c1 단백질 또는 mRNA 형태의 캡슐화에 의한다. 따라서, 일부 구현예에서, 지질 나노입자 (LNP)가 고려된다. 안티트랜스타이레틴 짧은 간섭 RNA는 지질 나노입자에 캡슐화되었고, 인간에게 전달되며 (예를 들어, 문헌 [Coelho et al., N Engl J Med 2013;369:81929] 참조), 이러한 시스템은 본 발명의 CRISPR Cas 시스템에 적합하고, 적용될 수 있다. 정맥내로 투여되는 kg 체중당 약 0.01 내지 약 1 mg의 용량이 고려된다. 주입-관련 반응 위험을 감소시키기 위한 약물이 고려되고, 예컨대, 덱사메타손, 아세트암피노펜, 다이펜하이드라민 또는 세티리진 및 라니티딘이 고려된다. 5회 용량의 경우 4주마다 킬로그램 당 약 0.3 mg의 다수 용량이 또한 고려된다.

[0843] LNP는 간에 siRNA를 전달하는 데 고도로 효과적인 것으로 확인되었고 (예를 들어, 문헌 [Tabernero et al., Cancer Discovery, April 2013, Vol. 3, No. 4, pages 363-470 참조), 그러므로 CRISPR Cas를 코딩하는 RNA를 간에 전달하는 것이 고려된다. 2주마다 6 mg/kg의 LNP의 약 4회 용량의 용량이 고려될 수 있다. Tabernero 등은 0.7 mg/kg 용량의 LNP의 최초 2회 주기 이후에 종양 퇴행이 관찰되었고, 6회 주기의 종료까지, 환자는 림프절 전이의 완전한 퇴행 및 간 종양의 실질적 축소로 부분 반응을 달성하였다는 것을 입증하였다. 이 환자에서 40 용량 이후에 완전 반응이 수득되었고, 환자는 26개월 동안 용량을 받은 후에 관해 및 완전 치료된 채로 남았다. VEGF 경로 억제제에 의한 사전 요법 후에 진행된 신장, 폐 및 림프절을 포함하여 간외 질환 부위 및 RCC를 갖는 2명 환자는 대략 8개월 내지 12개월 동안 모든 부위에서 안정한 질환을 가졌고, PNET 및 간 전이를 갖는 환자는 안정한 질환을 가진채로 18개월 (36회 용량) 동안 계속 연장 연구되었다.

[0844] 그러나, LNP의 전하가 고려되어야만 한다. 양이온성 지질이 음으로 하전된 지질과 조합되어 세포내 전달을 용이하게 하는 비이중층 구조를 유도하기 때문이다. 하전된 LNP는 정맥내 주사 후 순환으로부터 빠르게 제거되기 때문에, pKa 값이 7 미만인 이온화 가능한 양이온성 지질이 발생되었다 (예를 들어, 문헌 [Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011] 참조). 음으로 하전된 중합체, 예컨대, RNA는 이온화가능한 지질이 양전하를 나타내는 낮은 pH 값 (예를 들어, pH 4)에서 LNP에 로딩될 수 있다. 그러나, 생리적 pH 값에서, LNP는 더 긴 순환 시간과 상용성인 낮은 표전 전하를 나타낸다. 이온화 가능한 양이온성 지질의 4가지 종, 즉, 1,2-디리네오일-3-디메틸암모늄-프로판 (DLinDAP), 1,2-디리놀레일옥시-3-N,N-디메틸아미노프로판 (DLinDMA), 1,2-디리놀레일옥시-케토-N,N-디메틸-3-아미노프로판 (DLinKDMA) 및 1,2-디리놀레일-4-(2-디메틸아미노에틸)-[1,3]-디옥솔란 (DLinKC2-DMA)에 집중하였다. 이들 지질을 함유하는 LNP siRNA 시스템은 생체내 간세포에서 현저하게 상이한 유전자 침묵 특성을 나타내며, 역가는 인자 VII 유전자 침묵 모델을 적용하는 시리즈 DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP에 따라 다양하다 (예를 들어, 문헌 [Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 12862200, Dec. 2011] 참조). 특히 DLinKC2-DMA를 함유하는 제형에 대하여, 1 ㎍/㎖의 투여량의 LNP 또는 LNP 중의 또는 이와 연합된 CRISPR Cas RNA가 고려될 수 있다.

[0845] LNP 및 CRISPR Cas 캡슐화의 제조는 문헌 [Rosin et al, Molecular Therapy, vol. 19, no. 12, p 12862200, Dec. 2011]로부터 사용되고/되거나 개조될 수 있다. 양이온성 지질 1,2-다이리네오일-3-다이메틸암모늄-프로판(DLinDAP), 1,2-다이리놀레일옥시-3-N,N-다이메틸아미노프로판(DLinDMA), 1,2-다이리놀레일옥시케토-N,N-다이메틸-3-아미노프로판(DLinK-DMA), 1,2-다이리놀레일-4-(2-다이메틸아미노에틸)-[1,3]-다이옥솔란(DLinKC2-DMA), (3-o-[2"-(메톡시폴리에틸렌글리콜 2000) 석시노일]-1,2-다이미리스토일-sn-글리콜(PEG-S-DMG) 및 R-3-

[(ω-메톡시-폴리(에틸렌 글리콜)2000) 카바모일]-1,2-다이미리스틸옥시프로필-3-아민 (PEG-C-DOMG)은 Tekmira Pharmaceuticals (캐나다 밴쿠버에 소재)에 의해 제공되거나 또는 합성될 수 있다. 콜레스테롤은 Sigma (St Louis, MO)로부터 구입할 수 있다. 특정 CRISPR Cas RNA는 DLinDAP, DLinDMA, DLinK-DMA, 및 DLinKC2-DMA (40:10:40:10 몰비의 양이온성 지질:DSPC:CHOL: PEGS-DMG 또는 PEG-C-DOMG)을 함유하는 LNP에 캡슐화될 수 있다. 필요하다면, 0.2% SP-DiOC18 (Invitrogen, Burlington, Canada)은 세포 흡수, 세포내 전달 및 생체분포를 평가하기 위해 혼입될 수 있다. 캡슐화는 에탄올에 양이온성 지질:DSPC:콜레스테롤:PEG-c-DOMG (40:10:40:10 몰비)를 포함하는 지질 혼합물을 10 mmol/ℓ 의 최종 지질 농도로 용해시킴으로써 수행될 수 있다.

이러한 지질의 에탄올 용액을 50 mmol/ℓ 시트레이트, pH 4.0에 적가하여 다층판 소포체를 형성하여 30% 에탄올 vol/vol의 최종 농도를 생성할 수 있다. 압출기(캐나다 밴쿠버에 소재한 노던 리피즈(Northern Lipids))를 이용하여 2개의 적층된 80nm 뉴클레포어(Nuclepore) 폴리카보네이트 필터를 통해 다층판 소포의 압출 후 거대 단층판 소포가 형성될 수 있다. 30% 에탄올 vol/vol 적가물 pH 4.0을 함유하는 50mmol/ℓ 시트르산염에서 2mg/mℓ로 용해시킨 RNA를 압출 수행된 거대 단층판 소포에 첨가함으로써 그리고 0.06/1 wt/wt의 최종 RNA/지질 중량 비로 일정하게 혼합하면서 31℃에서 30분 동안 인큐베이션시킴으로써 캡슐화가 달성될 수 있다. Spectra/Por 2 재생된 셀룰로스 투석막을 이용하여 16시간 동안 인산염-완충 식염수(PBS), pH 7.4에 대한 투석에 의해 에탄올의 제거 및 제형 완충제의 중화가 수행되었다. NICOMP 370 입도 분석기, 소포체/강도 모드 및 가우시안 적합도 (Nicomp Particle Sizing, Santa Barbara, CA)를 이용하는 동적광산란에 의해 나노입자 크기 분포를 결정할 수 있다. 모든 3가지의 LNP 시스템에 대한 입자 크기는 ~70 ㎚의 직경일 수 있다. RNA 캡슐화 효율은 투석 이전 및 이후에 수집된 샘플로부터 VivaPureD MiniH 컬럼 (Sartorius Stedim Biotech)을 사용해 유리 RNA를 제거하여 결정할 수 있다. 캡슐화된 RNA는 용리된 입자로부터 추출되고, 260 nm에서 정량화될 수 있다. RNA 대 지질비는 Wako Chemicals USA(Richmond, VA)의 콜레스테롤 E 효소 어세이를 이용하여 소포체 내 콜레스테롤 함량 측정에 의해 결정되었다. 본 명세서에서 논의된 LNP 및 PEG 지질과 함께, PEG화된 리포솜 또는 LNP가 역시 CRISPR-Cas 시스템 또는 이의 성분의 전달에 적합하다.

[0846] 지질 사전혼합 용액 (20.4 mg/mℓ의 총 지질 농도)은 DLinKC2-DMA, DSPC 및 콜레스테롤을 50:10:38.5 몰비로 함유하는 에탄올 중에서 제조될 수 있다. 소듐 아세테이트는 0.75:1 (소듐 아세테이트:DLinKC2-DMA)의 몰비로 지질 사전혼합물에 첨가될 수 있다. 지질은 후속하여 격렬하게 교반시키면서 혼합물을 1.85 부피의 시트레이트 완충제 (10 mmol/ℓ, pH 3.0)와 조합하여 수하될 수 있어서, 그 결과로 35% 에탄올을 함유하는 수성 완충액 중에 자발적 리포솜 형성을 야기시킨다. 리포솜 용액은 입자 크기의 시간-의존적 증가를 가능하게 하기 위해 37℃에서 인큐베이션될 수 있다. 동적광산란 (Zetasizer Nano ZS, Malvern Instruments, Worcestershire, UK)에 의한 리포솜 크기 변화를 조사하기 위해 인큐베이션 동안 다양한 시간에 분취액을 제거할 수 있다. 목적 입자 크기가 얻어지면, 수성 PEG 지질 용액 (스톡 = 35% (vol/vol) 에탄올 중 10 mg/mℓ PEG-DMG)을 리포솜 혼합물에 첨가하여 3.5% 총 지질의 최종 PEG 몰농도가 산출될 수 있다. PEG-지질의 첨가 시, 리포솜은 그들의 크기를 조절하여 추가 성장을 효과적으로 중단시켜야 한다. 이어서, RNA는 대략 1:10의 RNA 대 총 지질 비율 (wt:wt))로 빈 리포솜에 첨가된 후에, 30분 동안 37℃에서 인큐베이션되어 로딩된 LNP가 형성될 수 있다. 혼합물은 후속적으로 PBS에서 밤새 투석될 수 있고, 0.45 ㎛ 시린지 필터로 여과될 수 있다.

[0847] 구형 핵산 (SNA™) 구성체 및 다른 나노 입자 (특히 금 나노입자)는 또한 의도된 표적에 CRISPR-Cas 시스템을 전달하기 위한 수단으로서 고려된다. 유의한 데이터는 핵산-작용화된 금 나노입자에 기반한 AuraSense Therapeutics의 구형 핵산(SNA™) 구성체가 유용하다는 것을 보여준다.

[0848] 본 명세서의 교시와 함께 적용될 수 있는 문헌은 다음을 포함한다: Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) 및 Mirkin, et al., Small, 10:186-192.

[0849] RNA를 갖는 자가 조립 입자는, 폴리에틸렌 글리콜(PEG)의 원위 말단에 부착된 Arg-Gly-Asp(RGD) 펩티드 리간드로 PEG화된 폴리에틸렌이민(PEI)을 이용하여 작제될 수 있다. 이러한 시스템은, 예를 들어 인테그린을 발현하는 종양 신생혈관을 표적화하고, 혈관 내피 성장 인자 수용체-2(VEGF R2) 발현을 억제하는 siRNA를 전달하여, 이에 의해, 종양 혈관형성을 달성하기 위한 수단으로 사용된다(참조, 예를 들어, Schiffelers et al., Nucleic

Acids Research, 2004, Vol. 32, No. 19). 2 내지 6 범위의 순 몰 과량의 이온화가능한 질소 (중합체) 대 포스페이트 (핵산)를 제공하도록 동일 부피의 양이온 중합체 수용액 및 핵산을 혼합하여 나노플렉스를 제조할 수 있다. 양이온성 중합체와 핵산 사이의 정전기적 상호작용은 본 명세서에서 나노플렉스라고 하는 약 100 nm의 평균 입자 크기 분포를 갖는 폴리플렉스의 형성을 야기시킨다. 약 100 내지 200 mg 용량의 CRISPR Cas가 Schiffelers 등의 자가-조립 입자로의 전달을 위해 계획된다.

[0850] Bartlett 등의 나노플렉스 (PNAS, September 25, 2007,vol. 104, no. 39)가 또한 본 발명에 적용될 수 있다. Bartlett 등의 나노플렉스는 동일 부피의 양이온성 중합체 및 핵산의 수용액을 혼합함으로써 제조되어 2 내지 6의 범위에 걸쳐 인산염(핵산)에 대한 순 몰 과량의 이온화 가능한 질소(중합체)를 제공한다. 양이온성 중합체와 핵산 간의 정전기적 상호작용은 약 100 nm의 평균 입자 크기를 지니는 다중복합체 형성을 야기하며, 따라서 본 명세서에서 나노복합체로서 지칭된다. Bartlett 등의 DOTA-siRNA는 다음과 같이 합성되었다: 1,4,7,10-테트라아자사이클로도데칸-1,4,7,10-테트라아세트산 모노(N-하이드록시숙신이미드 에스테르)(DOTA-NHS 에스테르)를 Macrocyclics(미국 텍사스주 달라스 소재)에서 주문하였다. 카보네이트 완충액 (pH 9) 중 100 배 몰 과량의 DOTA-NHS-에스테르로 아민 변형된 RNA 센스 가닥을 미세원심분리 튜브에 첨가하였다. 내용물은 4시간 동안 실온에서 교반시켜 반응시켰다. DOTA-RNA센스 접합체를 에탄올 침전시키고 나서, 물에서 재현탁시킨 후, 비변형 안티센스 가닥에 어닐링시켜서 DOTA-siRNA를 산출하였다. 모든 액체는 Chelex-100(Bio-Rad, Hercules, CA)로 전처리하여 미량의 금속 오염물을 제거하였다. Tf-표적화 및 비표적화 siRNA 입자를 사이클로덱스트린-함유 다가양이온을 사용함으로써 형성할 수 있다. 통상적으로, 입자를 전하비 3(+/-) 및 siRNA 농도 0.5 g/리터로 수 중에서 형성하였다. 표적화된 입자의 표면 상의 아다만탄-PEG 분자의 1%를 Tf로 변형시켰다(아다만탄-PEG-Tf). 입자를 주사용 5%(wt/vol) 글루코스 담체 용액 중에서 현탁화시켰다.

[0851] Davis 등(Nature, Vol 464, 15 April 2010)은 표적화된 나노입자-전달 시스템을 사용하는 RNA 임상 시험 (임상시험 등록 번호 NCT00689065)을 수행한다. 표준 치료 요법에 대해 난치성인 고형암을 갖는 환자에게 21일 주기 중 1일, 3일, 8일 및 10일에 30분 정맥내 주입으로 표적화된 나노입자의 용량을 투여한다. 나노입자는 (1) 선형, 시클로덱스트린계 중합체 (CDP), (2) 암 세포 표면 상에서 TF 수용체(TFR)와 맞물리도록 나노입자 외부에 제시되는 리간드 표적화 인간 트랜스페린 단백질 (TF), (3) 친수성 중합체 (생물학적 유체에서 나노입자 안정성을 촉진시키기 위해 사용되는 폴리에틸렌 글리콜 (PEG)), 및 (4) RRM2의 발현을 감소시키도록 디자인된 siRNA (임상에서 사용되는 서열은 이전에 표시된siR2B+5임)를 함유하는 합성 전달 시스템으로 이루어진다. TFR은 악성 세포에서 상향조절되는 것으로 오랫동안 알려져있었고, RRM2는 확립된 항암 표적이다. 이들 나노입자(CALAA-01로서 정의된 임상 형태)는 비-인간 영장류에서의 다회 투약 연구에서 잘 용인되는 것으로 나타났다. 만성 골수성 백혈병을 갖는 단일 환자에게 리포솜 전달에 의해 siRNA가 투여되었지만, Davis 등의 임상 시험은 표적화된 전달 시스템으로 siRNA를 전신에 전달하고 고형암을 갖는 환자를 치료하기 위한 초기 인간 시험이다. 표적화된 전달 시스템이 인간 종양에 기능성 siRNA의 효과적인 전달을 제공하는지의 여부를 확인하기 위해, Davis 등은 3종의 상이한 투약 코호트로부터 3명의 환자로서, 모두 전이성 흑색종을 갖고 각각 18, 24 및 30 mg m-2 siRNA의 CALAA-01 용량을 받은 환자 A, B 및 C로부터의 생검을 조사하였다. 본 발명의 CRISPR Cas 시스템에 대해 유사한 용량이 또한 고려될 수 있다. 본 발명의 전달은 선형, 사이클로덱스트린-기반 중합체(CDP), 암세포의 표면 상의 TF 수용체(TFR)에 맞물리는 입자 외부에 나타나는 인간 트랜스페린 단백질(TF) 표적화 리간드 및/또는 친수성 중합체(예를 들어, 생물학적 유체 중에서 입자 안정성을 촉진하기 위해 사용되는 폴리에틸렌 글리콜(PEG))를 함유하는 입자에 의해 달성될 수 있다.

[0852] 참고로 본 명세서에 포함되는 미국 특허 8,709,843은 조직, 세포, 및 세포내 구획으로의 치료제-함유 입자의 표적화된 전달을 위한 약물 전달 시스템을 제공한다. 상기 발명은 계면활성제, 친수성 중합체 또는 지질에 접합된 중합체를 포함하는 포함하는 표적화된 입자를 제공한다. 참고로 본 명세서에 포함되는 미국 특허 6,007,845는 다작용성 화합물과 하나 이상의 소수성 중합체 및 하나 이상의 친수성 중합체를 공유적으로 연결시킴으로써 형성된 멀티블록 공중합체의 코어를 갖고, 생물학적으로 활성인 물질을 함유하는 입자를 제공한다. 본 명세서에 참고로 편입된 미국 특허 제5,855,913호는 탭 밀도가 0.4 g/㎤ 미만이고 평균 직경이 5 ㎛ 내지 30 ㎛이며, 폐 시스템에 약물 전달을 위해 표면 상에 계면활성제를 혼입시키는 공기역학적 광 입자를 갖는 미립자 조성물을 제공한다. 참고로 본 명세서에 포함되는 미국 특허 5,985,309는 폐기관계로의 전달을 위한 계면활성제 및/또는 양으로 또는 음으로 하전된 치료제 또는 진단제 및 반대 전하로 하전된 분자의 친수성 또는 소수성 복합체가 혼입된 입자를 제공한다. 참고로 본 명세서에 포함되는 미국 특허 5,543,158은 표면 상에 폴리(알킬렌 글리콜) 모이어티 및 생물학적 활성 물질을 함유하는 생분해성 고체 코어를 갖는 생분해성 주사가능한 입자를 제공한다. 본 명세서에 참고로 편입된 WO2012135025(또한 미국 특허 제20120251560호로서 공개

됨)는 접합된 폴리에틸렌이민(PEI) 중합체 및 접합된 아자-거대고리(총괄적으로 "접합 리포머" (lipomer) 또는 "리포머" 로서 지칭됨)를 기재한다. 일정 구현예에서, 이러한 접합 리포머는 단백질 발현의 조율을 비롯한 유전자 발현을 변형시키는 시험관내, 생체외 및 생체내 게놈 동요를 달성하기 위해 CRISPR-Cas 시스템과 관련하여 사용될 수 있다는 것이 고려될 수 있다.

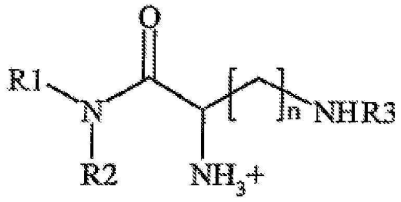
[0853] 일 실시형태에서, 입자는 에폭시드-변형된 지질-중합체, 유리하게는 7C1일 수 있다 (예를 들어, 2014년 5월 11일자로 온라인 상에서 공개된 [James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014)], doi:10.1038/nnano.2014.84] 참조). C71을 14:1의 몰비로 C15 에폭시드-말단 지질과 PEI600을 반응시킴으로써 합성하였고, 적어도 40 일 동안 PBS 용액에서 안정적인 입자(35 nm 내지 60 nm의 직경)를 생성시키기 위해 C14PEG2000과 함께 제형화하였다.

[0854] 에폭시드-변형된 지질-중합체는 폐, 심장 또는 신장 세포에 본 발명의 CRISPR-Cas 시스템을 전달하기 위해 이용될 수 있지만, 당업자는 다른 표적 기관에 전달하도록 시스템을 개조할 수 있다. 약 0.05 내지 약 0.6 mg/kg 범위의 용량이 계획된다. 수 일 또는 수 주 동안 용량이 또한 고려되고 총 용량은 약 2 mg/kg이다.

[0855] 일부 구현예에서, RNA 분자를 전달하기 위한 LNP는 당업계에 공지된 방법, 예컨대, 본 명세서에 참고로 편입된 WO 2005/105152 (PCT/EP2005/004920), WO 2006/069782 (PCT/EP2005/014074), WO 2007/121947 (PCT/EP2007/003496), 및 WO 2015/082080 (PCT/EP2014/003274)에 기재된 것에 의해 제조된다. 포유류 세포내로 siRNA의 향상된 그리고 개선된 전달 시 특이적으로 목적으로 하는 LNP는, 예를 들어, 본 명세서에 참고로 편입되고, 본 기술에 적용될 수 있는, 하기 문헌들에 기술되어 있다: Aleku et al., Cancer Res., 68(23): 9788-98 (Dec. 1, 2008), Strumberg et al., Int. J. Clin. Pharmacol. Ther., 50(1): 76-8 (Jan. 2012), Schultheis et al., J. Clin. Oncol., 32(36): 4141-48 (Dec. 20, 2014), 및 Fehring et al., Mol. Ther., 22(4): 811-20 (Apr. 22, 2014)

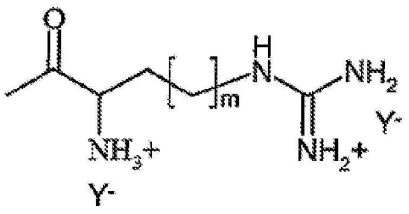
[0856] 일부 구현예에서, LNP는 WO 2005/105152 (PCT/EP2005/004920), WO 2006/069782 (PCT/EP2005/014074), WO 2007/121947 (PCT/EP2007/003496) 및 WO 2015/082080 (PCT/EP2014/003274)에 개시된 임의의 LNP를 포함한다.

[0857] 일부 구현예에서, LNP는 화학식 I의 구조를 갖는 적어도 하나의 지질을 포함한다:



[0858] (화학식 I)

[0859] 식 중, R1 및 R2는 각각 그리고 독립적으로 알킬을 포함하는 기로부터 선택되고, n은 1 내지 4의 임의의 정수이며, R3은 하기 화학식 II에 따른 라이실, 오르니틸, 2,4-다이아미노뷰티릴, 히스티딜 및 아실 모이어티를 포함하는 군으로부터 선택된 아실이고:



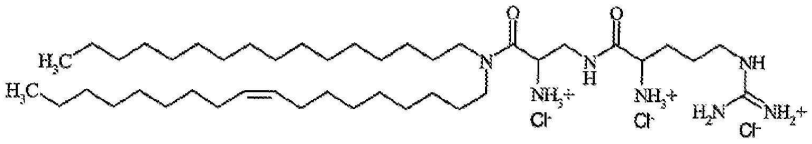
[0860] (화학식 II)

[0861] 식 중, m은 1 내지 3의 임의 정수이고, Y- 는 약학적으로 허용가능한 음이온이다. 일부 구현예에서, 화학식 I에 따른 지질은 적어도 2개의 비대칭 C 원자를 포함한다. 일부 구현예에서, 화학식 I의 거울상 이성질체는 제한없이 R-R; S-S; R-S 및 S-R 거울상 이성질체를 포함한다.

[0862] 일부 구현예에서, R1은 라우릴이고 R2는 미리스틸이다. 다른 구현예에서, R1은 팔미틸이고 R2는 올레일이다. 일부 구현예에서, m은 1 또는 2이다. 일부 구현예에서, Y-는 할로게니드, 아세테이트 또는 트리플루오로아세테이트로부터 선택된다.

[0863] 일부 구현예에서, LNP는 하기로부터 선택되는 하나 이상의 지질을 포함한다:

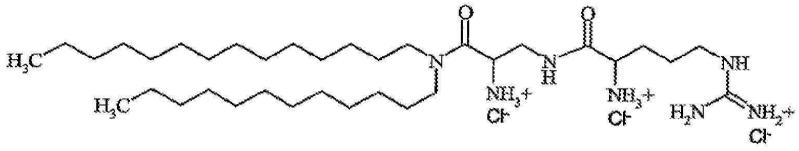
[0864] -아르기닐-2,3-디아미노 프로피온산-N-팔미틸-N-올레일-아미드 트리히드로클로라이드 (화학식 III):



[0865]

[0866] (화학식 III)

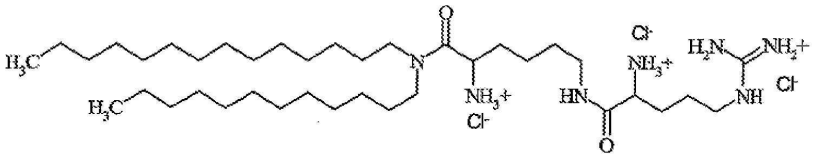
[0867] -아르기닐-2,3-디아미노 프로피온산-N-라우릴-N-미리스틸-아미드 트리히드로클로라이드 (화학식 IV):



[0868]

[0869] (화학식 IV); 및

[0870] -아르기닐-리신-N-라우릴-N-미리스틸-아미드 트리히드로클로라이드 (화학식 V):



[0871] (화학식 V).

[0872] 일부 구현예에서, LNP는 또한 구성성분을 포함한다. 예로서, 제한없이,일부 구현예에서, 구성성분은 펩티드, 단백질, 올리고뉴클레오티드, 폴리뉴클레오티드, 핵산 또는 이들의 조합물로부터 선택된다. 일부 구현예에서, 구성성분은 항체, 예를 들어, 단일클론 항체이다. 일부 구현예에서, 구성성분은, 예를 들어, 리보자임, 압타머, 스피에겔머 (spiegelmer), DNA, RNA, PNA, LNA, 또는 이들의 조합으로부터 선택된 핵산이다. 일부 구현예에서, 핵산은 가이드 RNA 및/또는 mRNA이다.

[0873] 일부 구현예에서, LNP의 구성성분은 CRIPSR-Cas 단백질을 코딩하는 mRNA를 포함한다. 일부 구현예에서, LNP의 구성성분은 II형 또는 V형 CRIPSR-Cas 단백질을 코딩하는 mRNA를 포함한다. 일부 구현예에서, LNP의 구성성분은 아데노신 디아미나제 (CRISPR-Cas 단백질 또는 어댑터 단백질에 융합될 수 있음)를 코딩하는 mRNA를 포함한다.

[0874] 일부 구현예에서, LNP의 구성성분은 추가로 하나 이상의 가이드 RNA를 포함한다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 혈관내피에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 폐 내피에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 간에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 폐에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 심장에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 비장에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 신장에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 췌장에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 뇌에 전달하도록 구성된다. 일부 구현예에서, LNP는 앞서 언급된 mRNA 및 가이드 RNA를 대식세포에 전달하도록 구성된다.

[0875] 일부 구현예에서, LNP는 또한 적어도 하나의 헬퍼 지질을 포함한다. 일부 구현예에서, 헬퍼 지질은 인지질 및 스테로이드로부터 선택된다. 일부 구현예에서, 인지질은 인산의 디에스테르 및/또는 모노에스테르이다. 일부 구현예에서, 인지질은 포스포글리세라이드 및/또는 스핑고지질이다. 일부 구현예에서, 스테로이드는 부분적으로 수소화된 사이클로펜타[a]페난트렌에 기반한 천연 유래 및/또는 합성 화합물이다. 일부 구현예에서, 스테로이드는 21 내지 30개의 C 원자를 함유한다. 일부 구현예에서, 스테로이드는 콜레스테롤이다.

일부 구현예에서, 헬퍼 지질은 1,2-디피타노일-sn-글리세로-3-포스포에탄올아민 (DPhyPE), 세라미드, 및 1,2-디올레일-sn-글리세로-3-포스포에탄올아민(DOPE)으로부터 선택된다.

[0876] 일부 구현예에서, 적어도 하나의 헬퍼 지질은 PEG 모이어티, HEG 모이어티, 폴리히드록시에틸 전분 (폴리HES) 모이어티 및 폴리프로필렌 모이어티를 포함하는 군으로부터 선택된 모이어티를 포함한다. 일부 구현예에서, 모이어티는 분자량이 약 500 내지 10,000 Da 또는 약 2,000 내지 5,000 Da이다. 일부 구현예에서, PEG 모이어티는 1,2-디스테아로일-sn-글리세로-3 포스포에탄올아민, 1,2-디알킬-sn-글리세로-3-포스포에탄올아민, 및 세라미드-PEG로부터 선택된다. 일부 구현예에서, PEG 모이어티는 분자량이 약 500 내지 10,000 Da 또는 약 2,000 내지 5,000 Da이다. 일부 구현예에서, PEG 모이어티는 분자량이 2,000 Da이다.

[0877] 일부 구현예에서, 헬퍼 지질은 조성물의 총 지질 함량의 약 20 mol% 내지 80 mol%이다. 일부 구현예에서, 헬퍼 지질 성분은 LNP의 총 지질 함량의 약 35 mol% 내지 65 mol%이다. 일부 구현예에서, LNP는 LNP의 총 지질 함량의 50 mol%의 지질 및 50 mol%의 헬퍼 지질을 포함한다.

[0878] 일부 구현예에서, LNP는 DPhyPE와 조합하여 -3-아르기닐-2,3-디아미노프로피온산-N-팔미틸-N-올레일-아미드 트리히드로클로라이드, -아르기닐-2,3-디아미노프로피온산-N-라우릴-N-미리스틸-아미드 트리히드로클로라이드 또는 -아르기닐-리신-N-라우릴-N-미리스틸-아미드 트리히드로클로라이드 중 임의의 것을 포함하고, 여기서 DPhyPE의 함량은 LNP의 전체 지질 함량의 약 80 mol%, 65 mol%, 50 mol% 및 35 mol%이다. 일부 구현예에서, LNP는 -아르기닐-2,3-디아미노 프로피온산-N-파니틸-N-올레일-아미드 트리히드로클로라이드 (지질) 및 1,2-디피타노일-sn-글리세로-3-포스포에탄올아민 (헬퍼 지질)을 포함한다. 일부 구현예에서, LNP는 β-아르기닐-2,3-디아미노 프로피온산-N-팔미틸-N-올레일-아미드 트리히드로클로라이드(지질), 1,2-디피타노일-sn-글리세로-3-포스포에탄올아민 (제1 헬퍼 지질), 및 1,2-다이스테로일-sn-글리세로-3-포스포에탄올아민-PEG2000 (제2 헬퍼 지질)을 포함한다.

[0879] 일부 구현예에서, 제2 헬퍼 지질은 총 지질 함량의 약 0.05 mol% 내지 4.9 mol% 또는 약 1 mol% 내지 3 mol%이다. 일부 구현예에서, LNP는 총 지질 함량의 약 45 mol% 내지 50 mol%의 지질, 총 지질 함량의 약 45 mol% 내지 50 mol%의 제1 헬퍼 지질을 포함하며, 단 총 지질 함량 중 약 0.1 mol% 내지 5 mol%, 약 1 mol% 내지 4 mol%, 또는 약 2 mol%의 PEG화된 제2 헬퍼 지질이 존재하고, 지질, 제1 헬퍼 지질 및 제2 헬퍼 지질 함량의 총 합은 총 지질 함량의 100 mol%이고, 제1 헬퍼 지질과 제2 헬퍼 지질의 합은 총 지질 함량의 50 mol%이다. 일부 실시형태에서, LNP는: (a) 50 mol%의 -아르기닐-2,3-디아미노 프로피온산-N-팔미틸-N-올레일-아미드 트리히드로클로라이드, 48 mol%의 1,2-디피타노일-sn-글리세로-3-포스포에탄올아민; 및 2 mol% 1,2-디스테아로일-sn-글리세로-3-포스포에탄올아민-PEG2000; 또는 (b) 50 mol%의 -아르기닐-2,3-디아미노 프로피온산-N-팔미틸-N-올레일-아미드 트리히드로클로라이드, 49 mol% 1,2-디피타노일-sn-글리세로-3-포스포에탄올아민; 및 1 mol% N(카보닐-메톡시폴리에틸렌글리콜-2000)-1,2-디스테아로일-sn-글리세로3-포스포에탄올아민, 또는 이의 소듐 염을 포함한다.

[0880] 일부 실시형태에서, LNP는 핵산을 함유하고, 핵산 골격 포스페이트 대 양이온성 지질 질소 원자의 전하비는 약 1: 1.5 내지 7 또는 약 1:4이다.

[0881] 일부 구현예에서, LNP는 또한 생체내 조건 하에 지질로부터 제거 가능한 차폐 화합물을 포함한다. 일부 구현예에서, 차폐 화합물은 생물학적 불활성 화합물이다. 일부 구현예에서, 차폐 화합물은 그의 표면 상에서 또는 이러한 분자 상에서 임의의 전하를 운반하지 않는다. 일부 구현예에서, 차폐 화합물은 폴리에틸렌글리콜(PEG), 하이드록시에틸글루코스(HEG) 기반 중합체, 폴리하이드록시에틸 전분(폴리HES) 및 폴리프로필렌이다.

일부 구현예에서, PEG, HEG, 폴리HES 및 폴리프로필렌은 약 500 내지 10,000Da 또는 약 2000 내지 5000Da의 무게가 나간다. 일부 구현예에서, 차폐 화합물은은 PEG2000 또는 PEG5000이다.

[0882] 일부 구현예에서, LNP는 생체내 조건 하에 지질로부터 제거 가능한 적어도 하나의 지질, 제1 헬퍼 지질 및 차폐 화합물을 포함한다. 일부 구현예에서, LNP는 또한 제2 헬퍼 지질을 포함한다. 일부 구현예에서, 제1 헬퍼 지질은 세라미드이다. 일부 구현예에서, 제2 헬퍼 지질은 세라미드이다. 일부 구현예에서, 세라미드는 6 내지 10개의 탄소 원자의 적어도 하나의 짧은 탄소쇄 치환체를 포함한다. 일부 구현예에서, 세라미드는 8개의 탄소 원자를 포함한다. 일부 구현예에서, 차폐 화합물은 세라미드에 부착된다. 일부 구현예에서, 차폐 화합물은 세라미드에 부착된다. 일부 구현예에서, 차폐 화합물은 세라미드에 공유 부착된다. 일부 구현예에서, 차폐 화합물은 LNP에서 핵산에 부착된다. 일부 구현예에서, 차폐 화합물은 핵산에 공유 부착된다. 일부 구현예에서, 차폐 화합물은 링커에 의해 핵산에 부착된다. 일부 구현예에서, 링커는 생리적 조건 하에 절단된다. 일부 구현예에서, 링커는 ssRNA, ssDNA, dsRNA, dsDNA, 펩티드, S-S-링커 및 pH

민감성 링커로부터 선택된다. 일부 구현예에서, 링커 모이어티는 핵산의 센스 가닥의 3' 말단에 부착된다.

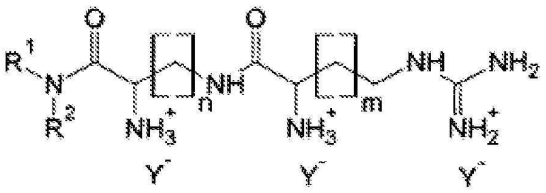
일부 구현예에서, 차폐 화합물은 pH-민감성 링커 또는 pH-민감성 모이어티를 포함한다. 일 구현예에서, pH-민감성 링커 또는 pH-민감성 모이어티는 음이온성 링커 또는 음이온성 모이어티이다. 일부 구현예에서, 음이온성 링커 또는 음이온성 모이어티는 산성 환경에서 덜 음이온성이거나 또는 중성이다. 일부 구현예에서, pH-민감성 링커 또는 pH-민감성 모이어티는 올리고 (글루탐산), 올리고페놀레이트(들) 및 디에틸렌 트리아민 펜타 아세트산으로부터 선택된다.

[0883] 앞 단락의 임의의 LNP 구현예에서, LNP는 약 50 내지 600 밀리오스몰 (mosmole)/kg, 약 250 내지 350 밀리오스몰/kg, 또는 약 280 내지 320 밀리오스몰/kg의 삼투질 농도를 가질 수 있고/있거나, 지질 및/또는 1 또는 2종의 헬퍼 지질 및 차폐 화합물에 의해 형성되는 LNP는 약 20 내지 200 nm, 약 30 내지 100 nm, 또는 약 40 내지 80 nm의 입자 크기를 갖는다.

[0884] 일부 구현예에서, 차폐 화합물은 더 긴 생체내 순환 시간을 제공하고, 핵산 함유 LNP의 더 양호한 생체 분포를 가능하게 한다. 일부 구현예에서, 차폐 화합물은 LNP가 투여되는 다른 체액 또는 세포질 막, 예를 들어, 혈관구조 내피벽의 세포질 막의 혈청 화합물 또는 화합물과 LNP의 즉각적인 상호작용을 방지한다. 추가적으로 또는 대안적으로, 일부 구현예에서, 차폐 화합물은 또한 LNP와의 상호작용 직후로부터 면역계의 요소를 막는다.

추가적으로 또는 대안적으로, 일부 구현예에서, 차폐 화합물은 항-옵소닌 작용 화합물로서 작용한다. 임의의 메커니즘 또는 이론에 의해 구속되는 일 없이, 일부 구현예에서, 차폐 화합물은 그의 환경과의 상호작용을 위해 이용 가능한 LNP의 표면적을 감소시키는 커버 또는 외피를 형성한다. 추가적으로 또는 대안적으로, 일부 구현예에서, 차폐 화합물은 LNP의 전반적인 전하를 차폐한다.

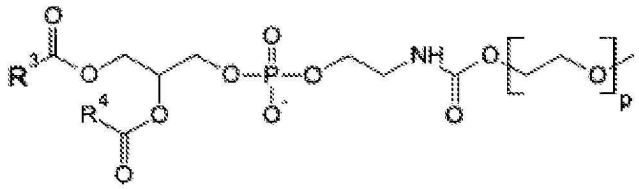
[0885] 다른 구현예에서, LNP는 하기 화학식 VI을 갖는 적어도 하나의 양이온성 지질을 포함한다:



[0886] (화학식 VI)

[0887] 식 중, n은 1, 2, 3, 또는 4이고, m은 1, 2, 또는 3이고, 여기서 Y- 는 양이온이고, 각각의 R^1 및 R^2 는 선형 C12-C18 알킬 및 선형 C12-C18 알케닐, 스테롤 화합물로 이루어진 군으로부터 개별적으로, 독립적으로 선택되고, 여기서 스테롤 화합물은 콜레스테롤 및 스티그마스테롤, 및 PEG화 지질로 이루어진 군으로부터 선택되고, 여기서 PEG화 지질은 PEG 모이어티를 포함하며, 여기서 PEG화 지질은

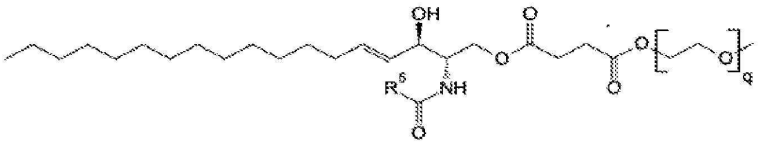
[0888] 하기 화학식 VII의 PEG화 포스포에탄올아민:



[0889] (화학식 VII)

[0890] (식에서, R^3 및 R^4 는 개별적으로, 그리고 독립적으로 선형 C13-C17 알킬이고, p는 15 내지 130의 임의 정수이다.)

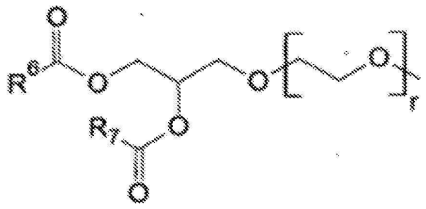
[0891] 하기 화학식 VIII의 PEG화 세라미드:



[0892] (화학식 VIII)

[0893] 식에서, R^5 는 선형 C7-C15 알킬이고, q는 15 내지 130의 임의 수이다.

[0894] 하기 화학식 IX의 PEG화 디아실글리세롤로 이루어진 군으로부터 선택된다:

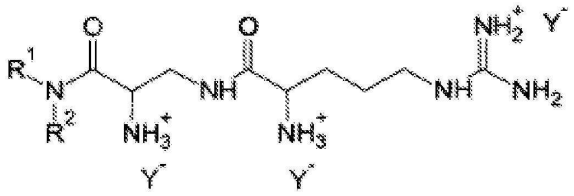


[0895] (화학식 IX)

[0896] 식에서, 각각의 R^6 및 R^7 은 개별적으로 그리고 독립적으로 선형 C11-C17 알킬이고, r은 15 내지 130의 임의 정수이다.

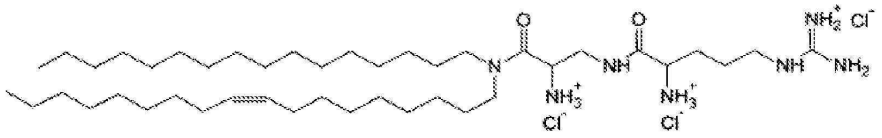
[0897] 일부 구현예에서, R^1 및 R^2 서로 상이하다. 일부 구현예에서, R^1은 팔미틸이고, R^2는 올레일이다. 일부 구현예에서, R^1은 라우릴이고 R^2는 미리스틸이다. 일부 구현예에서, R^1 및 R^2 는 동일하다. 일부 구현예에서, R^1 및 R^2 의 각각은 개개로 그리고 독립적으로 C12 알킬, C14 알킬, C16 알킬, C18 알킬, C12 알케닐, C14 알케닐, C16 알케닐 및 C18 알케닐로 이루어진 군으로부터 선택된다. 일부 구현예에서, C12 알케닐, C14 알케닐, C16 알케닐 및 C18 알케닐의 각각은 1 또는 2개의 이중 결합을 포함한다. 일부 구현예에서, C18 알케닐은 C9와 C10 사이에 하나의 이중 결합과 함께 C18 알케닐이다. 일부 구현예에서, C18 알케닐은 시스-9-옥타데실이다.

[0898] 일부 구현예에서, 양이온성 지질은 하기 화학식 X의 화합물이다:



[0899] (화학식 X)

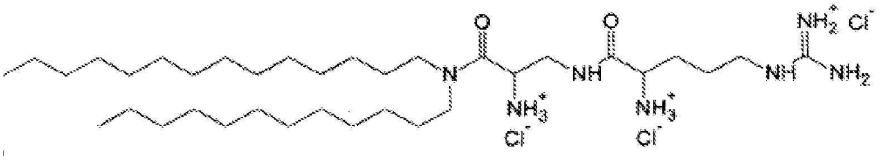
[0900] 일부 구현예에서, Y- 는 할로게니드, 아세테이트 및 트리플루오로아세테이트로부터 선택된다. 일부 구현예에서, 양이온성 지질은 하기 화학식 III의 -아르기닐-2,3-디아미노 프로피온산-N-팔미틸-N-올레일-아미드 트리히드로클로라이드이다:



[0901]

[0902] (화학식 III).

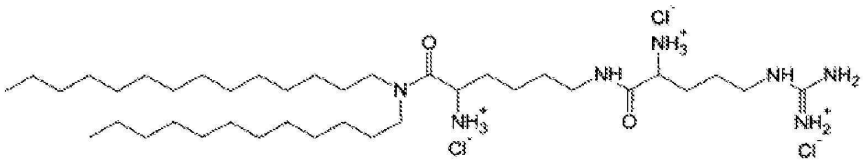
[0903] 일부 구현예에서, 양이온성 지질은 화학식 IV의 -아르기닐-2,3-디아미노 프로피온산-N-라우릴-N-미리스틸-아미드 트리히드로클로라이드이다:



[0904]

[0905] (화학식 IV)

[0906] 일부 구현예에서, 양이온성 지질은 화학식 V의 -아르기닐-리신-N-라우릴-N-미리스틸-아미드 트리히드로클로라이드이다:

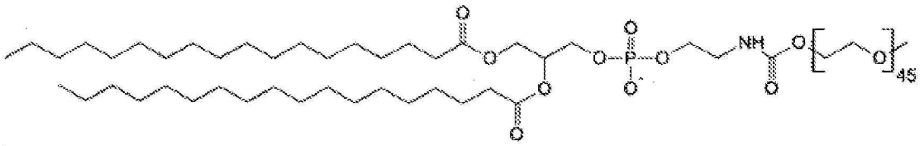


[0907]

[0908] (화학식 V).

[0909] 일부 구현예에서, 스테롤 화합물은 콜레스테롤이다. 일부 구현예에서, 스테롤 화합물은 스티그마스테린이다.

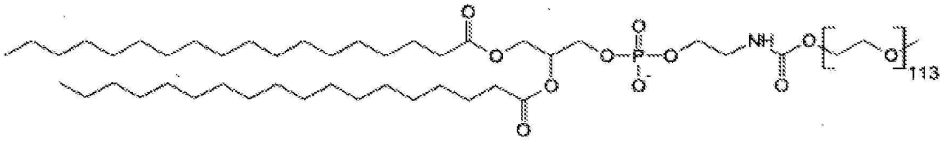
[0910] 일부 구현예에서, PEG화 지질의 PEG 모이어티는 약 800 내지 5,000 Da의 분자량을 갖는다. 일부 구현예에서, PEG화 지질의 PEG 모이어티의 분자량은 약 800 Da이다. 일부 구현예에서, PEG화 지질의 PEG 모이어티의 분자량은 약 2,000 Da이다. 일부 구현예에서, PEG화 지질의 PEG 모이어티의 분자량은 약 5,000 Da이다. 일부 구현예에서, PEG화 지질은 화학식 VII의 PEG화 포스포에탄올아민이고, 여기서 각각의 R^3 및 R^4는 개별적으로, 그리고 독립적으로 선형 C13-C17 알킬이고, p는 18, 19 또는 20, 또는 44, 45 또는 46 또는 113, 114 또는 115의 임의 정수이다. 일부 구현예에서, R^3 및 R^4 는 동일하다. 일부 구현예에서, R^3 및 R^4 는 상이하다. 일부 구현예에서, 각각의 R^3 및 R^4 는 개별적으로 C13 알킬, C15 알킬 및 C17 알킬로 이루어진 군으로부터 독립적으로 선택된다. 일부 구현예에서, 화학식 VII 의 PEG화 포스포에탄올아민은 1,2-디스테아로일-sn-글리세로-3-포스포에탄올아민-N-[메톡시(폴리에틸렌 글리콜)-2000] (암모늄 염)이다:



[0911]

[0912] (화학식 XI).

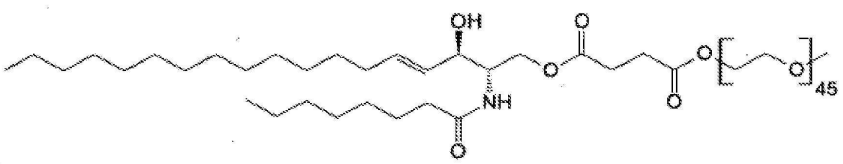
[0913] 일부 구현예에서, 화학식 VII의 PEG화 포스포에탄올아민은 1,2-디스테아로일-sn-글리세로-3-포스포에탄올아민-N-[메톡시(폴리에틸렌 글리콜)-5000] (암모늄 염)이다:



[0914]

[0915] (화학식 XII)

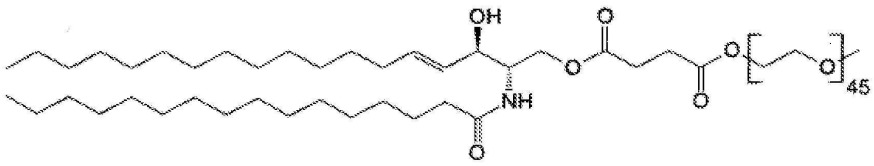
[0916] 일부 구현예에서, PEG화 지질은 화학식 VIII의 PEG화 세라미드이고, 여기서 R^5 는 선형 C7-C15 알킬이고, q는 18, 19 또는 20, 또는 44, 45 또는 46 또는 113, 114 또는 115의 임의 정수이다. 일부 구현예에서, R^5 는 선형 C7 알킬이다. 일부 구현예에서, R5 는 선형 C15 알킬이다. 일부 구현예에서, 화학식 VIII의 PEG화 세라미드는 N-옥타노일-스핀고신-1- {숙시닐[메톡시(폴리에틸렌 글리콜)2000]}이다:



[0917]

[0918] (화학식 VIII)

[0919] 일부 구현예에서, 화학식 VIII의 PEG화 세라미드는 N-팔미토일-스핀고신-1- {숙시닐[메톡시(폴리에틸렌 글리콜)2000]}이다:

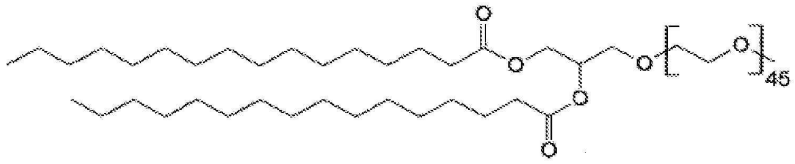


[0920]

[0921] (화학식 XIV)

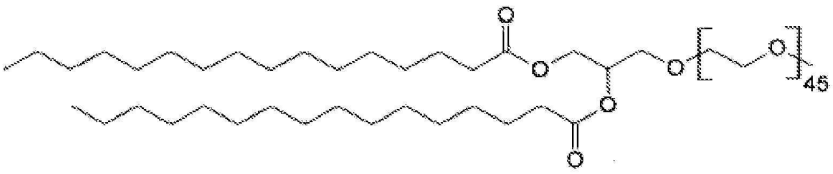
[0922] 일부 구현예에서, PEG화 지질은 화학식 IX의 PEG화 디아실글리세롤이고, 여기서 각각의 R^6 및 R^7 은 개별적으로, 그리고 독립적으로 선형 C11-C17 알킬이고, r은 18, 19 또는 20, 또는 44, 45 또는 46 또는 113, 114 또는 115의 임의 정수이다. 일부 구현예에서, R^6 및 R^7 은 동일하다. 일부 구현예에서, R^6 및 R^7 은 상이하다.

일부 구현예에서, 각각의 R^6 및 R^7 은 개별적으로, 선형 C17 알킬, 선형 C15 알킬 및 선형 C13 알킬로 이루어진 군으로부터 독립적으로 선택된다. 일부 구현예에서, 화학식 IX 의 PEG화 디아실글리세롤은 1,2-디스테아로일-sn-글리세롤 [메톡시(폴리에틸렌 글리콜)2000]이다:



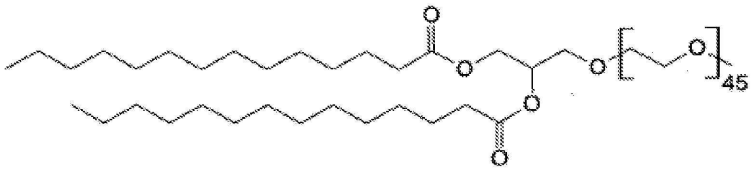
[0923] (화학식 XV)

[0924] 일부 구현예에서, 화학식 IX 의 PEG화 디아실글리세롤은 1,2-디팔미토일-sn-글리세롤 [메톡시(폴리에틸렌 글리콜)2000]이다:



[0925] (화학식 XVI).

[0926] 일부 구현예에서, 화학식 IX의 PEG화 디아실글리세롤은 다음과 같다:



[0927]

[0928] (화학식 XVII)

[0929] 일부 구현예에서, LNP는 화학식 III, IV 및 V로부터 선택된 적어도 하나의 양이온성 지질, 콜레스테롤 및 스티그마스테린으로부터 선택된 적어도 하나의 스테롤 화합물을 포함하되, 페길화된 지질은 화학식 XI 및 XII로부터 선택된 적어도 하나이다. 일부 구현예에서, LNP는 화학식 III, IV 및 V로부터 선택된 적어도 하나의 양이온성 지질, 콜레스테롤 및 스티그마스테린로부터 선택된 적어도 하나의 스테롤 화합물을 포함하고, 여기서 PEG화 지질은 화학식 XIII 및 XIV로부터 선택된 적어도 하나이다. 일부 구현예에서, LNP는 화학식 III, IV 및 V로부터 선택된 적어도 하나의 양이온성 지질, 콜레스테롤 및 스티그마스테린로부터 선택된 적어도 하나의 스테롤 화합물을 포함하고, 여기서 PEG화 지질은 화학식 XV 및 XIV로부터 선택된 적어도 하나이다. 일부 구현예에서, LNP는 화학식 III의 양이온성 지질, 스테롤 화합물로서 콜레스테롤을 포함하고, 여기서 PEG화 지질은 화학식 XI이다.

[0930] 이전 단락에서 임의의 LNP 구현예에서, 양이온성 지질 조성물의 함량은 약 65 mole% 내지 75 mole%이고, 스테롤 화합물의 함량은 약 24 mole% 내지 34 mole%이고, PEG화 지질의 함량은 약 0.5 mole% 내지 1.5 mole%이고, 지질 조성물에 대해 양이온성 지질, 스테롤 화합물 및 PEG화 지질의 함량의 총합은 100 mole%이다. 일부 구현예에서, 양이온성 지질의 함량은 약 70 mole%이고, 스테롤 화합물의 함량은 약 29 mole%이고 PEG화 지질의 함량

은 약 1 mole%이다. 일부 구현예에서, LNP는 70 mole%의 화학식 III, 29 mole%의 콜레스테롤, 및 1 mole%의 화학식 XI이다.

[0931] **엑소솜**

[0932] 엑소솜은 RNA 및 단백질을 수송하고, 뇌 및 다른 표적 기관에 RNA를 전달할 수 있는, 내생성 나노-소포체이다. 면역원성을 감소시키기 위해, 문헌 [Alvarez-Erviti et al. (2011, Nat Biotechnol 29: 341)]은 엑소솜 생산을 위한 자가-유래 수지상 세포를 사용하였다. 뇌에 대한 표적화는 뉴런-특이적 RVG 펩티드에 융합된 엑소솜 막 단백질인 Lamp2b를 발현시키기 위해 수지상 세포를 조작함으로써 달성되었다. 정제된 엑소솜은 전기천공법에 의해 외생성 RNA와 함께 부하되었다. 정맥내로 주사된 RVG-표적화된 엑소솜은 뇌에서 뉴런, 미세아교세포, 희돌기교세포에 GAPDH siRNA를 특이적으로 전달하여, 특이적 유전자 녹다운을 초래한다. RVG 엑소솜에 대한 사전 노출은 녹다운을 약화시키지 않았고, 다른 조직에서의 비-특이적 흡수는 관찰되지 않았다. 엑소솜-매개 siRNA 전달의 치료적 잠재력은 알츠하이머병에서의 치료적 표적인 BACE1의 강한 mRNA(60%) 및 단백질(62%) 녹다운에 의해 입증되었다.

[0933] 면역학적으로 불활성인 엑소솜의 풀을 얻기 위해, Alvarez-Erviti 등은 균질한 주요 조직적합 복합체 (주요 조직적합성 복합체: MHC) 단상형을 갖는 근친교배 C57BL/6 마우스로부터 골수를 채취하였다. 미성숙 수지상 세포는 T-세포 활성체, 예컨대, MHC-II 및 CD86가 결여된 다량의 엑소솜을 생성하는데, Alvarez-Erviti 등은 7일 동안 과립구/대식세포-콜로니 자극 인자(GM-CSF)를 갖는 수지상 세포를 선택하였다. 엑소솜은 충분히 확립된 초원심분리 프로토콜을 사용하여 다음날에 배양 상청액으로부터 정제하였다. 생성된 엑소솜은 물리적으로 균질하였고, 크기 분포는 입자 추적 분석(NTA) 및 전자 현미경에 의해 결정하였을 때, 직경 80 nm에서 최고였다. Alvarez-Erviti 등은 106개의 세포당 6 내지 12 ㎍의 엑소솜을 얻었다 (단백질 농도 기반으로 측정).

[0934] 다음에, Alvarez-Erviti 등은 나노규모 적용분야에 적합한 전기천공법 프로토콜을 이용하여 외생성 카고를 변형된 엑소솜에 로딩하는 가능성을 조사하였다. 나노미터 규모에서 막 입자에 대한 전기천공법은 잘 특징규명되고 있지 않기 때문에, 전기천공법 프로토콜의 경험적 최적화를 위해 비특이적 Cy5-표지 RNA가 사용되었다. 엑소솜의 초원심분리 및 용해 후에 캡슐화된 RNA의 양이 분석되었다.엑소솜의 초원심분리 및 용해 후에 캡슐화된 RNA의 양이 분석되었다. 400 V 및 125 μF에서 전기천공은 RNA의 최고 체류를 초래하였고, 모든 후속 실험을 위해 사용되었다.

[0935] Alvarez-Erviti은 150 ㎍의 RVG 엑소솜에 캡슐화된 150 ㎍의 각각의 BACE1 siRNA를 정상 C57BL/6 마우스에 투여하였고, 녹다운 효율은 4종의 대조군: 비처리 마우스, RVG 엑소솜 단독 주사된 마우스, 생체내 양이온성 리포솜 시약과 복합체화된 BACE1 siRNA를 주사한 마우스 및 RVG-9R에 복합체화된 BACE1 siRNA를 주사한 마우스와 비교하였고, RVG 펩티드는 siRNA에 정전기적으로 결합되는 9 D-아르기닌에 접합된 것이다. 피질 조직 샘플은 투여 후 3일에 분석되었고, siRNA-RVG-9R-처리 마우스와 siRNARVG 엑소솜-처리 마우스 둘 모두에서 유의한 단백질 녹다운 (45%, $P < 0.05$, 대 62%, $P < 0.01$)이 관찰되었는데, 이는 BACE1 mRNA 수준의 상당한 감소를 초래하였다 (각각 66% [+ 또는 -] 15%, $P < 0.001$ 및 61% [+ 또는 -] 13%, $P < 0.01$). 게다가, 출원인들은 RVG-엑소솜-처리 동물에서 알츠하이머 병리에서 아밀로이드 플라크의 주된 성분인 총 [베타]-아밀로이드 142 수준의 유의한 감소 (55%, $P < 0.05$)를 입증하였다. 관찰된 감소는 BACE1 억제제의 심실내 주사 후에 정상 마우스에서 입증된 β-아밀로이드 140 감소보다 더 컸다. Alvarez-Erviti외 다수는 BACE1 절단 생성물에 대해 cDNA 말단의 5'-신속한 증폭(RACE)을 수행하였는데, 이는 siRNA에 의한 RNAi-매개 녹다운의 증거를 제공하였다.

[0936] 최종적으로, Alvarez-Erviti 등은 IL-6, IP-10, TNFα 및 IFN-α 혈청 농도를 평가함으로써 RNA-RVG 엑소솜이 생체내 면역 반응을 유도하는지의 여부를 연구하였다. 엑소솜 처리 후에, 모든 사이토카인의 유의하지 않은 변화는 IL-6 분비를 강하게 자극한 siRNA-RVG-9R와 대조적으로 siRNA-형질감염 시약 처리와 유사하게 나타나서, 엑소솜 처리의 면역학적 불활성 프로파일을 확증하였다. 엑소솜이 20%의 siRNA만을 캡슐화한다는 것을 고려하면, RVG-엑소솜에 의한 전달은 비슷한 mRNA 녹다운으로서 RVG-9R 전달보다 더 효율적인 것으로 나타나고, 상응하는 면역 자극 수준 없이 5배 더 적은 siRNA에 의해 더 큰 단백질 녹다운이 달성되었다. 이 실험은 신경퇴행성 질환과 관련된 유전자의 장기간 침묵에 잠재적으로 적합한 RVG-엑소솜 기술의 치료 가능성을 입증하였다. Alvarez-Erviti 등의 엑소솜 전달 시스템은 본 발명의 AD-기능화된 CRISPR-Cas 시스템을 치료적 표적, 특히 신경퇴행성 질환에 전달하기 위해 적용될 수 있다. 약 100 내지 1000 mg의 RVG 엑소솜에 캡슐화된 약 100 내지 1000mg의 CRISPR Cas의 용량이 본 발명에서 고려될 수 있다.

[0937] El-Andaloussi 등 (Nature Protocols 7,2112-2126(2012))은 배양 세포로부터 유래된 엑소솜이 시험관내 및 생

체내 RNA의 전달을 위해 이용될 수 있는 방법을 개시한다. 이 프로토콜은 먼저 펩티드 리간드에 융합된 엑소좀 단백질을 포함하는, 발현 벡터의 형질감염을 통해 표적화된 엑소좀의 생성을 기재한다. 다음에, El-Andaloussi 등은 형질감염 세포 상청액으로부터 엑소좀을 정제하고 특징규명하는 방법을 설명한다. 다음에, El-Andaloussi 등은 RNA를 엑소좀에 로딩하기 위한 중요한 단계를 상술한다. 최종적으로, El-Andaloussi 등은 마우스 뇌에서 시험관내 및 생체내에서 RNA를 효율적으로 전달하기 위해 엑소좀을 이용하는 방법을 약술한다. 엑소좀-매개 RNA 전달이 기능성 분석 및 영상화에 의해 평가된 예상 결과의 예가 또한 제공된다. 전체 프로토콜은 약 3 주가 걸린다. 본 발명에 따른 전달 또는 투여는 자가-유래 수지상 세포로부터 생성된 엑소좀을 이용하여 수행될 수 있다. 본 명세서의 교시로부터, 이는 본 발명의 실행에서 사용될 수 있다.

[0938] 다른 구현예에서, Wahlgren 등의 혈장 엑소좀 (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130)이 고려된다. 엑소좀은 수지상 세포 (DC), B 세포, T 세포, 비만 세포, 상피 세포 및 종양 세포를 비롯한 다수의 세포 유형에 의해 생성된 나노-크기 소포체 (크기가 30 내지 90 nm)이다. 이들 소포체는 후기 엔도좀의 내향 버딩 (inward budding)에 의해 형성되고, 이어서, 혈장막과의 융합 시 세포외 환경으로 방출된다. 엑소좀은 세포 사이에서 RNA를 자연적으로 운반하기 때문에, 이 특성은 유전자 요법에서 유용할 수 있고, 본 개시내용으로부터 본 발명의 실행에서 사용될 수 있다.

[0939] 혈장을 단리시키기 위해 900g에서 20분 동안 버피코트의 원심분리 다음에 세포 상청액을 채취하고, 세포를 제거하기 위해 300g에서 10분 동안 그리고 16 500g에서 30분 동안 원심분리시킨 후 0.22 ㎜ 필터를 통해 여과시킴으로써 혈장으로부터의 엑소좀을 제조할 수 있다. 엑소좀은 70분 동안 120,000 g에서 초원심분리를 통해 펠렛화된다. 엑소좀으로 siRNA의 화학적 형질감염은 RNAi Human/마우스 Starter Kit (Quiagen, Hilden, Germany)의 제조사 설명서에 따라 수행된다. siRNA는 2 mmol/mL의 최종 농도로 100 mL PBS에 첨가된다. HiPerFect 형질감염 시약을 첨가한 후에, 혼합물은 10분 동안 실온에서 인큐베이션된다. 과량의 마이셀을 제거하기 위해, 엑소좀은 알데히드/설페이트 라텍스 비드를 이용하여 재단리된다. 엑소좀 내로 CRISPR Cas의 화학적 형질감염은 siRNA와 유사하게 수행될 수 있다. 엑소좀은 건강한 공혈자의 말초 혈액으로부터 단리된 단핵구 및 림프구와 함께 공동배양될 수 있다. 따라서, CRISPR Cas를 함유하는 엑소좀이 인간의 단핵구 및 림프구에 도입될 수 있고, 인간에 자가적으로 재도입될 수 있다는 것이 고려될 수 있다. 따라서, 본 발명에 따른 전달 또는 투여는 혈장 엑소좀을 이용하여 수행될 수 있다.

[0940] **리포좀**

[0941] 본 발명에 따른 전달 또는 투여는 리포좀에 의해 수행될 수 있다. 리포좀은 내부 수성 구획을 둘러싸며 친유성의 인지질 이중층 밖에서 상대적으로 불침투성인 단층판 또는 다층판 지질 이중층으로 구성된 구형 소포체 구조이다. 리포좀은 이들이 생체적합성, 비독성이기 때문에 약물 전달 담체로서 상당한 관심을 얻었고, 친수성과 친지성 약물 분자를 둘 다 전달할 수 있으며, 혈장 효소에 의한 분해로부터 그들의 카고를 보호하고, 생물학적 막 및 혈액 뇌 장벽(BBB)을 가로질러 그들의 로드를 전달할 수 있다(예를 들어, 검토를 위해 Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 참조).

[0942] 리포좀은 여러 가지 다른 유형의 지질로 만들 수 있다; 그러나 인지질이 약물 담체로서 리포좀을 생성하는데 가장 일반적으로 사용된다. 지질막이 수용액과 혼합될 때 리포좀 형성은 자발적이지만, 이는 또한 균질기, 초음파분쇄기, 또는 압출 장치를 이용하여 진탕 형태로 힘을 적용함으로써 신속히 처리될 수 있다 (예를 들어, 고찰을 위해 [Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조).

[0943] 몇몇 다른 첨가제가 리포좀에 첨가되어 이의 구조 및 특성을 변형시킬 수 있다. 예를 들어, 콜레스테롤 또는 스핀고미엘린 중 하나가 리포좀 혼합물에 첨가되어 리포좀 구조를 안정화시키고 리포좀 내부의 수송물 누출을 방지할 수 있다. 추가로, 리포좀은 수소화된 에그 포스파티딜콜린 또는 에그 포스파티딜콜린, 콜레스테롤 및 다이세틸 포스페이트로부터 제조되며, 그들의 평균 소포체 크기는 약 50 내지 100 nm로 조절되었다. (예를 들어, 고찰을 위해 문헌 [Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조).

[0944] 리포좀 제제는 주로 천연 인지질 및 지질, 예컨대, 1,2-디스테아로일-sn-글리세로-3-포스파티딜 콜린(DSPC), 스핀고미엘린, 에그 포스파티딜콜린 및 모노시알로강글리오시드로 구성될 수 있다. 이 제제는 인지질만으로 구성되기 때문에, 리포좀 제제는 많은 도전과 접하게 되었는데, 이 중 하나는 혈장에서의 불안정성이다. 이들 도전을 극복하기 위한 몇몇 시도가, 특히 지질막의 조작에서 만들어졌다. 이들 시도 중 하나는 콜레스테

롤의 조작에 중점을 두었다. 통상적인 제제에 대한 콜레스테롤의 첨가는 혈장내로 캡슐화된 생활성 화합물의 빠른 방출을 감소시키거나 또는 1,2-디올레오일-sn-글리세로-3-포스포에탄올아민 (DOPE)은 안정성을 증가시킨다 (예를 들어, 고찰을 위해 문헌 [Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조).

[0945] 특히 유리한 실시형태에서, 트로이 목마 (Trojan Horse) 리포솜 (또한 분자 트로이 목마로도 공지됨)가 바람직하며, 프로토콜은 cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long에서 확인할 수 있다. 이들 입자는 혈관내 주사 후에 전체 뇌에 대한 이식유전자 전달을 허용한다. 이론에 의해 국한하지 않고, 표면에 접합된 특정 항체를 갖는 중성 지질 입자는 세포내이입을 통해 혈액 뇌 장벽을 가로지르는 것으로 여겨진다.

트로이 목마 리포솜은 혈관내 주사를 통해 뇌에 뉴클레아제의 CRISPR 패밀리를 전달하는 데 사용될 수 있는데, 이는 이는 배아 조작에 대한 필요 없이 전체 뇌 유전자이식 동물을 허용한다. 약 1 내지 5g의 DNA 또는 RNA는 리포솜으로 생체내 투여가 고려될 수 있다.

[0946] 다른 구현예에서, AD-기능화된 CRISPR Cas 시스템 또는 이의 성분은 리포솜, 예컨대, 안정한 핵산-지질 입자 (SNALP)로 투여될 수 있다 (예를 들어, 문헌 [Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005] 참조). SNALP에서 표적화된 약 1, 3 또는 5 mg/kg/일의 특정 CRISPR Cas의 1일 정맥내 주사가 고려된다. 1일 처리는 약 3일 동안일 수 있고, 그 다음으로 약 5주 동안 매주일 수 있다. 다른 구현예에서, 약 1 또는 2.5mg/kg 용량으로 정맥내 주사에 의해 투여되는 특정 CRISPR Cas 캡슐화 SNALP가 또한 고려된다 (예를 들어, 문헌 [Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006] 참조). SNALP 제형은 지질 3-N-[(w-메톡시 폴리(에틸렌 글리콜) 2000) 카바모일]-1,2-다이미리스틸옥시-프로필아민 (PEG-C-DMA), 1,2-디리놀레일옥시-N,N-디메틸-3-아미노프로판 (DLinDMA), 1,2-디스테아로일-sn-글리세로-3-포스포콜린(DSPC) 및 콜레스테롤을 2:40:10:48 몰 백분율 비로 함유할 수 있다 (예를 들어, 문헌 [Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006] 참조).

[0947] 다른 구현예에서, 안정한 핵산-지질 입자 (SNALP)는 고도로 혈관화된 HepG2-유래 간 종양에 대한 효과적인 전달 분자가 되지만, 불량하게 혈관화된 HCT-116 유래 간 종양에서는 그렇지 않은 것으로 증명되었다 (예를 들어, 문헌 [Li, Gene Therapy (2012) 19, 775-780] 참조). SNALP 리포솜은 콜레스테롤/D-Lin-DMA/DSPC/PEG-C-DMA의 48/40/10/2 몰비 및 25:1 지질/siRNA 비율을 사용하여 D-Lin-DMA 및 PEG-C-DMA와 디스테아로일포스파티딜콜린 (DSPC), 콜레스테롤 및 siRNA를 함께 제제화하여 제조될 수 있다. 얻어진 SNALP 리포솜은 크기가 약 80 내지 100 nm이다.

[0948] 또 다른 실시형태에서, SNALP는 합성 콜레스테롤 (Sigma-Aldrich, St Louis, MO, USA), 디팔미토일포스파티딜콜린 (Avanti Polar Lipids, Alabaster, AL, USA), 3-N-[(w-메톡시 폴리(에틸렌 글리콜)2000)카바모일]-1,2-디미레스틸옥시프로필아민, 및 양이온성 1,2-디리놀레일옥시-3-N,N-디메틸아미노프로판 (예를 들어, 문헌 [Geisbert et al., Lancet 2010; 375: 1896- 905] 참조)을 포함할 수 있다. 예를 들어, 볼루스 정맥내 주입으로서 투여되는 용량 당 약 2 mg/kg의 총 CRISPR Cas 용량이 고려될 수 있다.

[0949] 또 다른 실시형태에서, SNALP는 합성 콜레스테롤 (Sigma-Aldrich), 1,2-디스테아로일-sn-글리세로-3-포스포콜린 (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA, 및 1,2-디리놀레일옥시-3-(N;N-디메틸)아미노프로판 (DLinDMA)을 포함할 수 있다 (예를 들어, 문헌 [Judge, J. Clin. Invest. 119:661-673 (2009)] 참조). 생체내 연구를 위해 사용되는 제형은 약 9:1의 최종 지질/RNA 질량비를 포함할 수 있다.

[0950] RNAi 나노메디신의 안전성 프로파일은 Alnylam Pharmaceuticals의 Barros 및 Gollob에 의해 검토되었다 (예를 들어, 문헌 [Advanced Drug Delivery Reviews 64 (2012) 1730-1737] 참조). 안정한 핵산 지질 입자 (SNALP)는 4가지의 상이한 지질로 구성된다 - 낮은 pH에서 양이온성인 이온화 가능한 지질 (DLinDMA), 중성 헬퍼 지질, 콜레스테롤, 및 확산성 폴리에틸렌 글리콜(PEG)-지질. 입자는 직경이 대략 80 nm이고, 생리적 pH에서 전하-중성이다. 제제화 동안에, 이온화 가능한 지질은 입자 형성 동안 음이온성 RNA와 지질을 축합시키는 역할을 한다.

점차적으로 산성인 엔도솜 조건 하에서 양으로 하전될 때, 이온화 가능한 지질은 또한 엔도솜 막과 SNALP의 융합을 매개하여 RNA이 세포질 내로 방출되는 것을 가능하게 한다. PEG-지질은 입자를 안정화시키고, 제제화 동안 응집을 감소시키며, 후속적으로 약물동태학적 특성을 개선시키는 외부의 중성 친수성을 제공한다.

[0951] 지금까지, RNA를 갖는 SNALP 제형을 이용하여 두 임상 프로그램이 개시되었다. Tekmira Pharmaceuticals는 LDL 콜레스테롤이 상승된 성인에서 SNALP-ApoB의 I상 단일-용량 연구를 최근에 완료하였다. ApoB는 간 및 공장에서 우세하게 발현되며, VLDL 및 LDL의 조립 및 분비에 필수적이다. 17명의 대상체는 SNALP-ApoB의 단일 용량을 받았다 (7회 용량 수준에 걸쳐 용량 상승). 간 독성의 증거는 없었다 (전임상 연구에 기반하여

잠재적 용량 제한 독성으로서 예상됨). 최고 용량의 (2명 중) 1명 대상체는 면역계 자극과 일관되게 감기-유사 증상을 경험하였고, 시험을 결론내리기 위해 결정되었다.

[0952] Alnylam Pharmaceuticals는 유사하게 진행된 ALN-TTR01을 갖는데, 이는 상기 기재한 SNALP 기술을 사용하며 돌연변이체와 야생형 TTR 둘 모두의 간세포 생성을 표적화하여 TTR 아밀로이드증 (ATTR)을 치료하였다. 3가지 ATTR 증후군이 기재되었다: 가족성 아밀로이드증성 말초신경병증 (FAP) 및 가족성 아밀로이드증성 심장근육병증 (FAC) - (둘 모두 TTR에서의 상염색체 우성 돌연변이에 의해 야기); 및 야생형 TTR에 의해 야기되는 노인성 전신 아밀로이드증 (SSA). ALN-TTR01의 위약-대조, 단일 용량-상승 I상 시험은 ATTR을 갖는 환자에서 최근에 완료되었다. ALN-TTR01은 0.01 내지 1.0 mg/kg (siRNA에 기반)의 용량 범위 내에서 31명의 환자 (23명은 연구 약물이고, 8명은 위약임)에게 15분 IV 주입으로서 투여되었다. 간 기능 검사에서의 유의한 증가 없이 치료는 충분히 내성이 있었다. 주입-관련 반응은 0.4 mg/kg 이상에서 23 명의 환자 중 3 명에서 관찰되었으며; 모두 주입 속도가 느려지는 것에 반응했으며 모두 연구를 계속했다. 2명의 환자에서 혈청 사이토카인 IL-6, IP-10 및 IL-1ra의 최소 및 일시적 상승이 1 mg/kg의 최고 용량에서 확인되었다 (전임상 및 NHP 연구로부터 예상). 혈청 TTR이 저하되면서, ALN-TTR01의 예상되는 약력학적 효과가 1mg/kg에서 관찰되었다.

[0953] 또 다른 구현예에서, 양이온성 지질, DSPC, 콜레스테롤 및 PEG-지질을, 예를 들어, 에탄올에서, 예를 들어, 각각 40:10:40:10의 몰비로 가용화함으로써 SNALP가 생성될 수 있다 (문헌 [Semple et al., Nature Niotechnology, Volume 28 Number 2 February 2010, pp. 172-177] 참조). 지질 혼합물은 최종 에탄올과 30% (vol/vol) 및 6.1 mg/mℓ 농도의 지질에 각각 혼합하여 수성 완충제 (50 mM 시트레이트, pH 4)에 첨가되었고, 압출 전에 22℃에서 2분 동안 평형 상태가 되게 하였다. 동적광산란 분석에 의해 결정하여 70 내지 90 nm의 소포체 직경이 얻어질 때까지, 수화된 지질을 Lipex Extruder (Northern Lipids)를 이용하여 22℃에서 2 적층 80 nm 포어-크기 필터 (Nuclepore)를 통해 압출시켰다. 이는 일반적으로 1 내지 3회의 통과를 필요로 하였다. siRNA (50 mM 시트레이트에서 가용화됨, 30% 에탄올을 함유하는 pH 4 수용액)을 혼합하면서 ~5 mL/분의 속도로 사전 평형화시킨(35℃) 소포체에 첨가하였다. 0.06 (wt/wt)의 최종 표적 siRNA/지질 비에 도달된 후에, 혼합물은 추가 30분 동안 35℃에서 인큐베이션되어 siRNA의 소포체 재조직화 및 캡슐화를 허용한다. 이어서, 에탄올이 제거되고, 외부 완충제는 투석 또는 접선유동 정용여과 중 하나에 의해 PBS (155mM NaCl, 3mM Na2HPO4, 1mM KH2PO4, pH 7.5)로 대체된다. siRNA는 제어된 단계식 희석 방법 과정을 사용해 SNALP에 캡슐화되었다. KC2-SNALP의 지질 성분은 57.1:7.1:34.3:1.4의 몰비로 사용된 DLin-KC2-DMA (양이온 지질), 디팔미토일포스파티딜콜린 (DPPC; Avanti Polar Lipids), 합성 콜레스테롤 (Sigma) 및 PEG-C-DMA였다. 로딩된 입자의 형성시, SNALP를 PBS에 대해 투석하였고, 사용 전 0.2 μm 필터를 통해 멸균여과시켰다. 평균 입자 크기는 75 내지 85 nm이었고, siRNA의 90 내지 95%는 지질 입자 내에서 캡슐화되었다. 생체내 시험을 위해 사용되는 제형에서의 최종 siRNA/지질 비는 ~0.15 (wt/wt)이었다. 인자 VII siRNA를 함유하는 LNP-siRNA 시스템은 사용 직전에 멸균 PBS에서 적절한 농도로 희석되었고, 제형은 10 mL/kg의 총 용적으로 옆쪽 꼬리 정맥을 통해 정맥내로 투여되었다. 이 방법 및 이들 전달 시스템은 본 발명의 AD-기능화된 CRISPR Cas 시스템에 대해 추론될 수 있다.

[0954] **기타 지질**

[0955] 기타 다른 양이온성 지질, 예컨대 아미노 지질 2,2-디리놀레일-4-디메틸아미노에틸-[1,3]-디옥솔란(DLin-KC2-DMA)은 예를 들어, siRNA와 유사한 CRISPR Cas 또는 이의 성분 또는 이를 코딩하는 핵산 분자(들)를 캡슐화하기 위해 이용될 수 있으며 (참조: 예를 들어, Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 -8533), 그러한 이유로 본 발명의 실시에서 적용될 수 있다. 다음의 지질 조성물을 이용하여 사전 형성한 소포가 고려될 수 있다:40/10/40/10의 몰비 및 대략0.05(w/w) FVII siRNA/총 지질비로 각각 아미노지질, 디스테아로일포스파티딜콜린(DSPC), 콜레스테롤 및 (R)-2,3-비스(옥타데실옥시) 프로필-1-(메톡시 폴리(에틸렌 글리콜)2000)프로필카바메이트(PEG-지질). 70 내지 90 nm 범위의 좁은 입자 크기 분포 및 0.11±0.04 (n=56)의 낮은 다분산 지수를 보장하기 위해, 입자는 가이드 RNA 첨가 전에 80 nm 막을 통해 3회까지 압출될 수 있다. 고도로 강한 아미노 지질 16(생체내 활성을 활성화하도록 추가로 최적화될 수 있는 4 가지 지질 구성성분 16, DSPC, 콜레스테롤 및 PEG-지질의 몰비(50/10/38.5/1.5))을 함유하는 입자가 사용될 수 있다.

[0956] Michael S D Kormann 등 (" Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: Nature Biotechnology, Volume:29, Pages: 154- 157 (2011))은 RNA를 전달하기 위한 지질 외피의 사용을 기재한다. 지질 외피의 사용은 또한 본 발명에서 바람직하다.

[0957] 다른 구현예에서, 지질은 본 발명의 AD-기능화된 CRISPR Cas 시스템 또는 이의 성분(들) 또는 이를 코딩하는 핵

산 분자(들)로 제제화되어 지질 나노입자(LNP)를 형성할 수 있다. 지질은 DLin-KC2-DMA4, C12-200 및 코리피드 디스테로일포스파티딜 콜린, 콜레스테롤을 포함하지만, 이들로 제한되지 않으며, PEG-DMG는 자발적 소포체 형성 절차를 이용하여 siRNA 대신에 CRISPR Cas로 제제화될 수 있다 (예를 들어, 문헌 [Novobrantseva, Molecular Therapy-Nucleic Acid (2012) 1, e4; doi:10.1038/mtna.2011.3] 참조). 성분 몰비는 약 50/10/38.5/1.5 (DLin-KC2-DMA 또는 C12-200/디스테로일포스파티딜 콜린/콜레스테롤/PEG-DMG)일 수 있다. 최종 지질:siRNA 중량비는 각각 DLin-KC2-DMA 및 C12-200 지질 나노입자 (LNP)의 경우에 대략 12:1 내지 9:1일 수 있다. 제제는 평균 입자 직경이 ~80 nm이고 포획 효율이 90% 초과일 수 있다. 3 mg/kg 용량이 고려될 수 있다.

[0958] Tekmira는 LNP 및 LNP 제제의 다양한 양태와 관련된 미국 및 해외에서의 대략 95 개의 특허 패밀리의 포트폴리오를 가지며 (참조: 예를 들어, 미국 특허 번호 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 및 7,838,658 및 유럽 특허 번호 1766035; 1519714; 1781593 및 1664316), 이들 모두는 본 발명에 대해 사용되고/사용되거나 적합하게 될 수 있다.

[0959] AD-기능화된 CRISPR Cas 시스템 또는 이의 성분 또는 이를 코딩하는 핵산 분자(들)는 단백질, 단백질 전구체를 코딩하거나, 또는 단백질 또는 단백질 전구체의 부분적으로 또는 완전히 가공된 형태를 코딩할 수 있는 변형된 핵산 분자를 포함하는 조성물의 제형의 양상에 관한 미국 특허 출원 공개 제20130252281호 및 제20130245107호 및 제20130244279호 (Moderna Therapeutics에게 양도)에 추가로 기재되는 것과 같이 PLGA 미세구에서 캡슐화되어 전달될 수 있다. 제제는 몰비 50:10:38.5:1.5 내지 3.0 (양이온성 지질:융합유도성 (fusogenic) 지질:콜레스테롤:PEG 지질)을 가질 수 있다. PEG 지질은 PEG-c-DOMG, PEG-DMG로부터 선택되지만, 이들로 제한되지 않을 수 있다. 융합생성 지질은 DSPC일 수 있다. 또한 스크럼(Schrum) 등의 미국 특허 출원 공개 제20120251618호(발명의 명칭:Delivery and Formulation of Engineered Nucleic Acids) 참조.

[0960] 나노머 기술은 저분자량 소수성 약물, 펩티드 및 핵산 기반 치료제(플라스미드, siRNA, miRNA)를 비롯한 광대한 범위의 치료제에 대한 생체 이용 가능성 도전을 처리한다. 기술이 분명한 이점을 입증한 특정 투여 경로는 경구 경로, 뇌-혈관-장벽을 가로지르는 수송, 고형 종양뿐만 아니라 눈에 대한 전달을 포함한다. 참조: 예를 들어 SMazza et al., 2013, ACS Nano. 2013 Feb 26;7(2):1016-26; Uchegbu and Siew, 2013, J Pharm Sci. 102(2):305-10 및 Lalatsa et al., 2012, J Control Release. 2012 Jul 20; 161(2):523-36.

[0961] 미국 공개 특허 출원 번호 20050019923은 폴리뉴클레오티드 분자, 펩티 및 폴리펩티드 및/또는 약학제)와 같은 생활성 분자를 포유류 신체에 전달하기 위한 양이온성 덴드리머를 기재한다. 덴드리머는, 예를 들어, 간, 비장, 폐, 신장 또는 심장 (또는 심지어 뇌)에 대한 생활성 분자의 전달을 표적화하는 데 적합하다. 덴드리머는 단순 분지형 단량체 단위로부터 단계식-방식으로 제조되는 합성 3-차원 거대분자로서, 이의 성질 및 기능성은 쉽게 제어할 수 있으며 다양할 수 있다. 덴드리머는 다작용성 코어에 대해 (합성에 대한 분기적 접근), 또는 다작용성 코어를 향한 (합성에 대한 수렴적 접근) 빌딩 블록의 반복 첨가로부터 합성되고, 빌딩 블록의 3차원 껍질의 각각의 첨가는 더 높은 세대의 덴드리머의 형성을 야기한다. 폴리프로필렌이민 덴드리머는 디아미노부탄 코어로부터 시작되는데, 이에 대해 1차 아민에 대한 아크릴로나이트릴의 이중 마이클 첨가에 의해 2배의 아민 기 수가 첨가된 후에, 나이트릴의 수소화가 이어진다. 이는 아민 기의 배가를 초래한다. 폴리프로필렌이민 덴드리머는 100% 양성자화 가능한 질소 및 64개까지의 말단 아미노 기 (5 세대, DAB 64)를 함유한다. 양성자화 가능한 기는 보통 중성 pH에서 양성자를 받아들일 수 있는 아민 기이다. 유전자 전달제로서 덴드리머의 사용은 대체로 유전자 전달을 위해 저급 세대 폴리프로필렌이민 덴드리머의 사용에 대한 작업이 보고되지 않은, 각각 접합 유닛으로서 아민/아미드 또는 N--P(O2)S의 혼합물을 갖는 인 함유 화합물, 및 폴리아미도아민의 사용에 집중되었다. 폴리프로필렌이민 덴드리머는 또한 주변 아미노 기에 의해 화학적으로 변형될 때 게스트 분자의 그들 캡슐화 및 약물 절단을 위한 pH 민감성 제어 방출 시스템으로서 연구되었다. 폴리프로필렌이민 덴드리머의 DNA와의 상호작용 및 세포독성을 비롯하여 DAB 64의 형질감염 효능이 또한 연구되었다.

[0962] 미국 공개 특허 출원 번호 20050019923은 보다 앞선 보고와 대조적으로, 양이온성 덴드리머, 예컨대, 폴리프로필렌이민 덴드리머가 생활성 분자, 예컨대, 유전자 물질의 표적화된 전달에서 사용하기 위한 적합한 특성, 예컨대, 특이적 표적화 및 저독성을 나타낸다는 관찰에 기반하였다. 추가로, 양이온성 덴드리머의 유도체는 또한 생활성 분자의 표적화된 전달을 위한 적합한 특성을 나타낸다. 또한 미국 공개 특허 출원 번호 20080267903의 생활성 중합체를 참조하는데, 여기서는 " 양이온성 폴리아민 중합체 및 덴드리머 중합체를 비롯한 다양한 중합체가 항증식 활성을 갖는 것으로 나타나며, 따라서 원치않는 세포 증식을 특징으로 하는 장애,

예컨대, 신생물 및 종양, 염증 장애(자가면역 장애를 포함), 건선 및 죽상동맥경화증의 치료에 유용할 수 있다.

중합체는 단독으로 활성제로서, 또는 다른 치료제, 예컨대, 유전자 요법을 위한 약물 분자 또는 핵산에 대한 전달 비히클로서 사용될 수 있다. 이러한 경우에, 중합체 자체의 고유한 항종양 활성은 전달될 제제의 활성을 보완할 수 있다" 는 것을 개시한다. 이들 특허 공보의 개시내용은 AD-기능화된 CRISPR Cas 시스템(들) 또는 이의 성분(들) 또는 이를 코딩하는 핵산 분자(들)의 전달을 위해 본 명세서에서 함께 사용될 수 있다.

[0963] **초전하 단백질**

[0964] 초전하 단백질은 보통 높은 순 이론적 양전하 또는 음전하를 갖는 조작되거나 또는 천연 발생인 단백질의 부류이고, AD-기능화된 CRISPR Cas 시스템(들) 또는 이의 성분(들) 또는 이를 코딩하는 핵산 분자(들)의 전달에서 사용될 수 있다. 초음전하 단백질과 초양전하 단백질은 둘 모두 열적으로 또는 화학적으로 유도된 응집을 견뎌내는 현저한 능력을 나타낸다. 초양전하 단백질은 또한 포유동물 세포에 침투할 수 있다. 이들 단백질, 예컨대, 플라스미드 DNA, RNA, 또는 다른 단백질과 카고의 회합은 시험관내와 생체내 포유동물 세포 둘 모두에서 이들 거대분자의 기능적 전달을 가능하게 할 수 있다. 초전하 단백질의 생성 및 특징규명은 2007년에 보고되었다 (Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110-10112).

[0965] 포유동물 세포 내로 RNA 및 플라스미드 DNA의 비바이러스 전달은 연구와 치료적 적용 둘 모두에 가치 있다 (Akinc et al., 2010, Nat. Biotech. 26, 561- 569). 정제된 +36 GFP 단백질 (또는 다른 초양전하 단백질)은 적절한 무혈청 배지에서 RNA와 혼합되고, 세포에 첨가 전에 복합체를 형성되게 한다. 이 단계에서 혈청의 포함은 초하전 단백질-RNA 복합체의 형성을 억제하고, 치료 유효성을 감소시킨다. 다양한 세포주에 대해 다음의 프로토콜이 유효한 것으로 확인되었다 (McNaughton et al., 2009, Proc. Natl. Acad. Sci. USA 106, 6111-6116) (그러나, 단백질 및 RNA 용량을 달리한 파일럿 실험은 특정 세포주에 대한 절차를 최적화하기 위해 수행되어야 한다): (1) 치료 1일 전에, 48-웰 플레이트에서 웰 당 1×105개의 세포를 플레이팅한다. (2) 치료일에, 무혈청 배지 내 정제된 +36 GFP 단백질을 최종 농도 200 nM로 희석시킨다. 50 nM의 최종 농도로 RNA를 첨가한다. 혼합물을 교반시키고, 실온에서 10분 동안 인큐베이션시킨다. (3) 인큐베이션 동안, 세포로부터 배지를 흡입하고 PBS로 1회 세척한다. (4) +36 GFP 및 RNA의 인큐베이션 후에, 단백질-RNA 복합체를 세포에 첨가함. (5) 37℃에서 4 시간 동안 복합체와 세포를 인큐베이션시킴. (6) 인큐베이션 후에, 배지를 흡입하고, 20 U/㎖ 헤파린 PBS로 3 회 세척함. 활성에 대한 어세이에 따라 추가 48 시간 또는 그 이상 동안 혈청-함유 배지와 세포를 인큐베이션시킴. (7) 세포를 면역블롯, qPCR, 표현형 분석 또는 기타 다른 적절한 방법에 의해 분석함.

[0966] +36 GFP는 다양한 세포에서 효과적인 플라스미드 전달 시약인 것이 추가로 확인되었다. 플라스미드 DNA는 siRNA보다 더 큰 카고이기 때문에, 비례적으로 더 많은 +36 GFP 단백질이 효과적으로 복잡한 플라스미드에 필요하다. 효과적인 플라스미드 전달을 위해, 출원인들은 인플루엔자 바이러스 혈구응집소 단백질로부터 유래된 공지된 엔도솜-붕괴 펩티드인 C-말단의 HA2 펩티드 태그를 보유하는 +36 GFP의 변이체를 개발하였다. 다음의 프로토콜은 다양한 세포에서 효과적이었지만, 상기와 같이 플라스미드 DNA 및 초하전 단백질 용량이 특정 세포주 및 전달 응용분야를 위해 최적화되는 것이 권장된다: (1) 치료 1일 전에, 48-웰 플레이트에서 웰 당 1×105개의 세포를 플레이팅한다. 2) 처리일에, 무혈청 배지 내 þ36 GFP 단백질을 최종 농도 2 mM로 희석시킨다. 1㎎의 플라스미드 DNA를 첨가한다. 혼합물을 교반시키고, 실온에서 10분 동안 인큐베이션시킨다.

(3) 인큐베이션 동안, 세포로부터 배지를 흡입하고 PBS로 1회 세척한다. (4) þ36 GFP와 플라스미드 DNA의 인큐베이션 후에, 단백질-DNA 복합체를 세포에 부드럽게 첨가한다. (5) 37℃에서 4시간 동안 세포를 복합체와 인큐베이션시킨다. (6) 인큐베이션 후에, 배지를 흡입하고 나서, PBS로 세척한다. 혈청-함유 배지에서 세포를 인큐베이션시키고 나서, 추가 24 내지 48시간 동안 인큐베이션시킨다. (7) 적절하다면, 플라스미드 전달을 (예를 들어, 플라스미드-유도 유전자 발현에 의해) 분석한다.

[0967] 참조: 예를 들어, McNaughton et al., Proc. Natl. Acad. Sci. USA 106, 6111-6116 (2009); Cronican et al., ACS Chemical Biology 5, 747-752 (2010); Cronican et al., Chemistry & Biology 18, 833-838 (2011); Thompson et al., Methods in Enzymology 503, 293-319 (2012); Thompson, D.B., et al., Chemistry & Biology 19 (7), 831-843 (2012). 초전하 단백질 방법이 사용되고/되거나 본 발명의 AD-기능화된 CRISPR Cas 시스템의 전달에 적합화될 수 있다. 본 명세서의 교시와 함께 이들 시스템은 AD-기능화된 CRISPR Cas 시스템(들) 또는 이의 성분(들) 또는 이를 코딩하는 핵산 분자(들)의 전달에서 사용될 수 있다.

[0968] **세포 침투성 펩티드(CPP)**

[0969] 또 다른 구현예에서, AD-기능화된 CRISPR Cas 시스템의 전달을 위한 세포 침투성 펩티드(CPP)가 고려된다.

CPP는 다양한 분자 수송물(나노크기 입자로부터 작은 화학적 분자 및 DNA의 거대 단편까지)의 세포 흡수를 용이하게 하는 짧은 펩티드이다. 본 명세서에서 사용되는 용어 "카고" 는 치료제, 진단 프로브, 펩티드, 핵산, 안티센스 올리고뉴클레오티드, 플라스미드, 단백질, 입자(나노입자를 포함), 리포솜, 발색단, 소형 분자 및 방사성 물질로 이루어진 군을 포함하지만, 이들로 제한되지 않는다. 본 발명의 양상에서, 카고는 또한 AD-기능화된 CRISPR Cas 시스템 또는 전체 AD-기능화된 기능성 CRISPR Cas 시스템의 임의의 성분을 포함할 수 있다.

본 발명의 양상은 하기 단계들을 포함하는, 목적하는 카고를 대상체에 전달하는 방법을 추가로 제공한다: (a) 본 발명의 세포 침투성 펩티드 및 목적하는 카고를 포함하는 복합체를 준비하는 단계, 및 (b) 대상체에 복합체를 경구로, 관절내로, 복강내로, 초내로, 동맥내로, 비강내로, 실질세포내로, 피하로, 근육내로, 정맥내로, 피내로, 직장내로, 또는 국소로 투여하는 단계. 카고는 공유 결합을 통한 화학적 결합 또는 비-공유 상호작용을 통해 펩티드와 연합된다.

[0970] CPP의 기능은 카고를 세포에 전달하는 것으로서, 살아있는 포유동물 세포의 엔도솜에 전달되는 카고에 의한 세포내 이입을 통해 통상적으로 일어나는 과정이다. 세포-침투성 펩티드는 상이한 크기, 아미노산 서열 및 전하를 가지지만, 모든 CPP는 한가지 구별되는 특징을 갖는데, 그것은 혈장막을 전좌하고 세포질 또는 유기체에 다양한 분자 카고의 전달을 용이하게 하는 능력이다. CPP 전좌는 3가지 주요 유입 메커니즘으로 분류될 수 있다: 막 내 직접 침투, 세포내 이입-매개 유입 및 전이 구조의 형성을 통한 전좌. CPP는 암 및 바이러스 억제제를 비롯한 상이한 질환의 치료에서의 약물 전달제뿐만 아니라 세포 표지를 위한 조영제로서 의학에서의 수많은 용도가 발견되었다. 후자의 예는 GFP에 대한 담체, MRI 조영제, 또는 양자점으로서의 작용을 포함한다.

CPP는 연구 및 의학에서 사용하기 위한 시험관내 및 생체내 전달 벡터로서 큰 가능성을 보유한다. CPP는 전형적으로 라이신 또는 알기닌과 같은 양으로 하전된 아미노산의 높은 상대적 존재비를 함유하거나 또는 극성/하전 아미노산 및 비극성의 소수성 아미노산의 교번의 패턴을 함유하는 서열을 갖는, 아미노산 조성물을 가진다. 이들 두 유형의 구조는 각각 다양이온 또는 양쪽성으로서 지칭된다. CPP의 제3 부류는 낮은 순전하를 갖는 비극성 잔기만을 함유하는 소수성 펩티드이거나 또는 세포 흡수에 중요한 소수성 아미노산 기를 가진다. 발견된 초기 CPP 중 하나는 배양물 내 수많은 세포 유형에 의해 주변 배지로부터 효율적으로 취해지는 것으로 발견된 인간 면역결핍 바이러스 1(HIV-1)로부터의 전사 촉진성 전사 활성인자(Tat)였다. 이후로, 공지된 CPP의 수는 상당히 확장되며 더 효과적인 단백질 형질도입 특성을 갖는 소분자 합성 유사체가 생성되었다. CPP는 페네트라틴 (Penetratin), Tat (48-60), 트랜스포탄 (Transportan) 및 (R-AhX-R4)(Ahx=아미노헥사노일)을 포함하지만, 이들로 제한되지 않는다.

[0971] 미국 특허 제8,372,951호는 고도의 세포 침투성 효율 및 낮은 독성을 나타내는 호산구 양이온성 단백질 (ECP)로부터 유래된 CPP를 제공한다. CPP와 이의 카고를 척추동물 대상체에 전달하는 양태가 또한 제공된다. CPP 및 그들의 전달의 추가적인 양태는 미국 특허 제8,575,305호; 제8;614,194호 및 제8,044,019호에 기재되어 있다. CPP는 AD-기능화된 CRISPR-Cas 시스템 또는 이의 성분을 전달하는 데 사용될 수 있다. CPP가 CRISPR-Cas 시스템 또는 이의 성분을 전달하기 위해 이용될 수 있는 것은 또한 본원에 전체가 참고로 포함된, Suresh Ramakrishna, Abu-Bonsrah Kwaku Dad, Jagadish Beloor 등에 의한 ["Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA", Genome Res. 2014 Apr 2.]의 원고에서 제공되며, 여기에서 CPP-접합된 재조합 Cas9 단백질 및 CPP-복합체화 가이드 RNA를 이용한 치료가 인간 세포주에서의 내생성 유전자 파괴를 이끌어냄이 입증되었다. 이 논문에서 Cas9 단백질은 티오에스테르 결합을 통해 CPP에 콘쥬게이트되는 반면, 가이드 RNA는 CPP와 복합체화되어, 축합된 양으로 하전된 입자를 형성한다.

변형된 Cas9 및 가이드 RNA에 의한, 배아 줄기 세포, 진피 섬유아세포, HEK293T 세포, HeLa 세포 및 배아 암종 세포를 포함하는 인간 세포의 동시 및 후속적 처리는 플라스미드 형질감염에 비해 감소된 오프-표적 돌연변이를 갖는 효율적인 유전자 붕괴를 야기하는 것으로 나타났다.

[0972] **에어로졸 전달**

[0973] 폐 질환에 대해 치료된 대상체는, 예를 들어 자발적으로 호흡하면서 폐에 대해 기관지내삽관으로 전달되는 약학적 유효량의 에어로졸화된 AAV 벡터 시스템을 받을 수 있다. 이와 같이, 일반적으로 AAV 전달에 에어로졸화된 전달이 바람직하다. 전달을 위해 아데노바이러스 또는 AAV 입자가 사용될 수 있다. 각각 하나 이상의 조절 서열에 작동적으로 연결된 적합한 유전자 구성체는 전달 벡터에 클로닝될 수 있다.

[0974] **패키징 및 프로모터**

[0975] CRISPR-Cas 단백질 및 임의로 기능성 도메인 (예를 들어, 아데노신 데아미나제) 코딩 핵산 분자 발현을 구동시키기 위해 사용되는 프로모터는 AAV ITR을 포함할 수 있고, 이것이 프로모터로서 작용할 수 있다. 이는 추

가 프로모터 구성요소 (벡터 내 공간을 차지할 수 있음)의 필요를 제거하는데 유리하다. 추가의 공간 개방은 추가 구성요소 (gRNA 등)의 발현을 구동하는데 사용될 수 있다. 또한, ITR 활성은 상대적으로 더 약하며, 따라서 Cas13의 과발현에 기인하여 잠재적 독성을 감소시키는 데 사용될 수 있다.

[0976] 편재성 발현을 위해서, 사용할 수 있는 프로모터는 CMV, CAG, CBh, PGK, SV40, 페리틴 중쇄 또는 경쇄 등을 포함한다.뇌 또는 다른 CNS 발현을 위해서, 시냅신I이 모든 뉴런에 대해 사용될 수 있고, CaMKII알파는 흥분성 뉴런에 대해 사용될 수 있고, GAD67 또는 GAD65 또는 VGAT는 GABA성 뉴런에 사용될 수 있다. 간 발현에 대해, 알부민 프로모터가 사용될 수 있다. 폐 발현에 대해, SP-B가 사용될 수 있다. 내피 세포에 대해, ICAM이 사용될 수 있다. 조혈 세포에 대해, IFN베타 또는 CD45가 사용될 수 있다. 골아세포에 대해, OG-2가 사용될 수 있다.

[0977] 가이드 RNA를 구동시키는데 사용되는 프로모터는 Pol III 프로모터, 예컨대, U6 또는 H1을 비롯하여, Pol II 프로모터의 사용 및 가이드 RNA의 발현을 위한 인트론 카세트를 포함할 수 있다.

[0978] 일정 구현예에서, CRISPR-Cas 시스템은 아데노 연관 바이러스 (AAV), 백혈병 바이러스 (MuMLV), 렌티바이러스, 아데노바이러스 또는 다른 플라스미드 또는 바이러스 벡터 유형을 사용해 전달된다.

[0979] **아데노 연관 바이러스 (AAV)**

[0980] CRISPR-Cas 단백질, 아데노신 디아미나제, 및 하나 이상의 가이드 RNA는 아데노 연관 바이러스 (AAV), 렌티바이러스, 아데노바이러스 또는 다른 플라스미드 또는 바이러스 벡터 유형을 이용하여, 특히, 예를 들어, 미국 특허 제8,454,972호 (제제, 아데노바이러스에 대한 용량), 제8,404,658호 (제제, AAV에 대한 용량) 및 제5,846,946호 (제제, DNA 플라스미드에 대한 용량)으로부터의 용량, 및 임상 시험 및 렌티바이러스, AAV 및 아데노바이러스를 수반하는 임상 시험에 관한 간행물로부터의 용량을 이용하여, 전달될 수 있다. 예를 들어, AAV에 대해, 투여 경로, 제형 및 용량은 미국 특허 제8,454,972호와 같고, AAV를 수반하는 임상 시험과 같을 수 있다. 아데노바이러스에 대해, 투여 경로, 제형 및 용량은 미국 특허 제8,404,658호와 같고 아데노바이러스를 수반하는 임상 시험과 같을 수 있다. 플라스미드 전달을 위해, 투여 경로, 제형 및 용량은 미국 특허 제5,846,946호에서와 같고 플라스미드를 수반하는 임상 연구에서와 같을 수 있다. 용량은 평균 70 kg 개체(예를 들어, 남성 성인 인간)에 기반하거나 또는 이에 대해 추론될 수 있고, 상이한 체중 및 종의 환자, 대상체, 포유동물에 대해 조절될 수 있다. 투여 빈도는 환자 또는 대상체의 연령, 성별, 일반적 건강상태, 다른 병태 및 처리될 특정 병태 또는 증상을 비롯한 보통의 인자에 따라서, 의학적 또는 수의학적 실행자 (예를 들어, 의사, 수의사)의 영역 내이다. 바이러스 벡터는 관심 조직 내로 주사될 수 있다. 세포 유형 특이적 게놈 변형을 위해, C2c1 및 아데노신 디아미나제의 발현은 세포 유형 특이적 프로모터에 의해 구동될 수 있다. 예를 들어, 간-특이적 발현은 알부민 프로모터를 사용하고, 뉴런-특이적 발현은 (예를 들어, CNS 장애 표적화를 위해) 시냅신 I 프로모터를 사용할 것이다.

[0981] 생체 전달과 관련하여, AAV는 몇가지 이유때문에 다른 바이러스 벡터에 비해 유리하다: 낮은 독성 (이는 면역반응을 활성화시킬 수 있는 세포 입자의 초원심분리를 필요로 하지 않는 정제 방법에 기인할 수 있음); 및 숙주 게놈 내로 통합되지 않기 때문에 삽입 돌연변이유발을 야기할 낮은 가능성.

[0982] AAV는 4.5 또는 4.75 Kb의 패키징 한계를 갖는다. 이것은 Cas13뿐만 아니라 프로모터 및 전사 종결자가 동일한 바이러스 벡터에 모두 적합하게 될 것임 의미한다. 4.5 또는 4.75 Kb보다 더 큰 구성체는 상당히 감소된 바이러스 생성을 야기할 것이다. SpCas9는 상당히 크며, 유전자 그 자체는 4.1 Kb인데, 이는 AAV에 패키징되는 것을 어렵게 만든다. 따라서, 본 발명의 구현예는 더 짧은 Cas13의 상동체를 이용하는 것을 포함한다. 일부 구현예에서, 바이러스 캡시드는 VP1, VP2, VP3 캡시드 단백질 중 하나 이상을 포함한다.

[0983] AAV에 대해, AAV는 AAV1, AAV2, AAV5 또는 이들의 임의의 조합일 수 있다. 표적화하려는 세포에 대한 AAV의 AAV를 선택할 수 있으며; 예를 들어, 뇌 또는 뉴런 세포를 표적화하기 위해 AAV 혈청형 1, 2, 5 또는 하이브리드 캡시드 AAV1, AAV2, AAV5 또는 이들의 임의의 조합을 선택할 수 있고; 심장 조직을 표적화하기 위해 AAV4를 선택할 수 있다. AAV8은 간에 대한 전달에 유용하다. 본 명세서의 프로모터 및 벡터는 개개로 바람직하다. 이들 세포에 대한 소정의 AAV 혈청형의 표는 다음과 같다 (참조: Grimm, D. et al, J. Virol. 82:5887- 5911(2008)):

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

[0984]

[0985] **렌티바이러스**

[0986] 렌티바이러스는 유사분열 세포와 유사분열 후 세포 둘 모두에서 그들의 유전자를 감염시키고 발현시키는 능력을 갖는 복잡한 레트로바이러스이다. 가장 통상적으로 알려진 렌티바이러스는 인간 면역결핍 바이러스(HIV)인데, 이는 매우 다양한 세포 유형을 표적화하기 위해 다른 바이러스의 외피 당단백질을 사용한다.

[0987] 렌티바이러스는 다음과 같이 제조될 수 있다. pCasES10(렌티바이러스 전달 플라스미드 골격을 함유함)을 클로닝시킨 후에, 10% 소 태아 혈청이 있고 항생제가 없는 DMEM에서의 형질감염 전달에 낮은 계대로(p=5) HEK293FT를 T-75 플라스크에서 50% 합류점(confluence)으로 파종시켰다. 20시간 후에, 배지를 옵티멤(OptiMEM)(무혈청) 배지로 바꾸고 나서, 4시간 후에 형질감염을 행하였다. 세포를 10 ㎍의 렌티바이러스 전달 플라스미드(pCasES10) 및 다음의 패키징 플라스미드: 5 ㎍의 pMD2.G (VSV-g 위형) 및 7.5 ㎍의 psPAX2(gag/pol/rev/tat)로 형질감염시켰다. 4㎖ OptiMEM에서 양이온성 지질 전달제(50uL 피포펙타민 2000 및 100㎕ 플러스 시약)를 이용하여 형질감염을 행하였다. 6시간 후에, 배지를 10% 소 태아 혈청이 있는 무항생제 DMEM으로 바꾸었다. 이들 방법은 세포 배양 동안 혈청을 사용하지만, 무혈청 방법이 바람직하다.

[0988] 렌티바이러스는 다음과 같이 정제될 수 있다. 48시간 후에 바이러스 상청액이 채취되었다. 상청액은 처음에 파편이 클리어런스되고, 0.45㎛ 저 단백질 결합(PVDF) 필터를 통해 여과된다. 이어서, 24,000 rpm에서 2시간 동안 초원심분리로 교반되었다. 바이러스 펠렛은 밤새 4℃에서 50ul의 DMEM 중에서 재현탁된다. 이어서, 그들은 분취되어, 즉시 -80℃에 냉동되었다.

[0989] 다른 실시형태에서, 말 전염성 빈혈 바이러스 (EIAV)에 기반한 최소 비영장류 렌티바이러스 벡터가 또한 특히 안구 유전자 요법에 대해 고려된다 (참조: 예를 들어, Balagaan, J Gene Med 2006; 8: 275 -285). 다른 구현예에서, 노인성 황반 변성의 그물망 형태 치료를 위해 망막하 주사를 통해 전달되는 혈관억제 단백질 엔도스타틴 및 앤지오스타틴을 발현시키는 말 전염성 빈혈 바이러스-기반 렌티바이러스 유전자 요법 벡터, RetinoStat®가 또한 고려되며 (예를 들어, 문헌 [Binley et al., HUMAN GENE THERAPY 23:980991(September 2012)] 참조), 이 벡터는 본 발명의 AD-기능화된 CRISPR-Cas 시스템에 대해 변형될 수 있다.

[0990] 또 다른 구현예에서, HIV tat/rev에 의해 공유된 일반 엑손을 표적화하는 siRNA, 핵소체-국소화 TAR 유인체(decoy), 및 항-CCR5-특이적 망치머리형 리보자임을 갖는 자가-비활성화 렌티바이러스성 벡터 (예를 들어, 문헌 [DiGiusto et al. (2010) Sci Transl Med 2:36ra43] 참조)가 사용되고/거나 본 발명의 CRISPR-Cas 시스템에 대해 조정될 수 있다. 최소 2.5 × 10^6개의 CD34+ 세포/킬로그램 환자 체중이 수집되고, 2 μmol/L-글루타민, 줄기 세포 인자(100ng/㎖), Flt-3 리간드(Flt-3L) (100ng/㎖), 및 트롬보포이에틴(10ng/㎖)(셀제닉스(CellGenix))을 2 × 106개의 세포/㎖의 밀도로 함유하는 X-VIVO 15 배지(론자(Lonza)) 배지에서 16 내지 20시간 동안 사전자극하였다. 사전 자극된 세포는 피브로넥틴 (25 ㎎/㎠)(RetroNectin, Takara Bio Inc.)으로 코팅한 75-㎠ 조직 배양 플라스크에서 16 내지 24시간 동안 5의 감염 다중도로 렌티바이러스를 형질도입시킬 수 있다.

[0991] 렌티바이러스 벡터는 파킨슨병에 대한 치료에 대해 개시되어 있고, 예를 들어, 미국 공개 특허 출원 제

20120295960호 및 미국 특허 제7303910호 및 7351585호를 참조한다. 렌티바이러스 벡터는 또한 안질환의 치료용으로 개시되었으며, 예를 들어 미국 특허 공개 20060281180, 20090007284, US20110117189; US20090017543; US20070054961, US20100317109 참고. 렌티바이러스 벡터는 또한 뇌로의 전달을 위해 개시되어 있다, 참조, 예를 들어, 미국 특허 공개번호 US20110293571; US20110293571, US20040013648, US20070025970, US20090111106 및 미국 특허 US7259015.

[0992] **중합체-기반 입자**

[0993] 본 명세서의 시스템 및 조성물은 중합체-기반 입자 (예를 들어, 나노입자)를 사용해 전달될 수 있다. 일부 구현예에서, 중합체-기반 입자는 막 융합의 바이러스 기전을 모방할 수 있다. 중합체-기반 입자는 인플루엔자 바이러스 기구의 합성 카피일 수 있고 산성 구획의 형성을 포함하는 과정인, 엔도시토시스 경로를 통해 세포가 흡수하는 다양한 유형의 핵산 (siRNA, miRNA, 플라스미드 DNA, 또는 shRNA, mRNA)과 형질감염 복합체를 형성한다. 후기 엔도솜의 저 pH는 화학 스위치로서 작용하여 입자 표면을 소수성으로 만들어서 막 크로싱을 촉진한다. 시토졸로 들어가면, 입자는 세포 작용을 위해 이의 페이로드를 방출한다. 이러한 활성 엔도솜 탈출 기술은 안전하고 천연 흡수 경로를 사용하므로 형질감염 효율을 최대화시킨다. 일부 구현예에서, 중합체-기반 입자는 알킬화 및 카르복시알킬화 분지형 폴리에틸렌이민을 포함할 수 있다. 일부 예에서, 중합체-기반 입자는 VIROMER, 예를 들어, VIROMER RNAi, VIROMER RED, VIROMER mRNA, VIROMER CRISPR이다. 본 명세서에서 시스템 및 조성물을 전달하는 예시적인 방법은 하기 문헌에 기술된 것들을 포함한다; Bawage SS et al., Synthetic mRNA expressed Cas13a mitigates RNA virus infections, www.biorxiv.org/content/10.1101/370460v1.full doi: doi.org/10.1101/370460, Viromer® RED, a powerful tool for transfection of 케라틴ocytes. doi: 10.13140/RG.2.2.16993.61281, Viromer® Transfection - Factbook 2018: technology, product overview, users' data., doi:10.13140/RG.2.2.23912.16642.

[0994] **일반 적용분야**

[0995] 본 개시는 본 명세서의 성분 및 시스템으로 표적 핵산 (예를 들어, DNA), 또는 하나 이상의 표적 핵산의 발현을 변형시키는 방법을 제공한다. 일부 구현예에서, 방법은 본 명세서의 하나 이상의 비천연 발생 또는 조작된 조성물과 표적 핵산을 접촉시키는 단계를 포함한다. 예를 들어, 본 개시는 관심 표적 서열을 변형시키는 방법을 제공하고, 이 방법은 표적 DNA를, i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질, ii) a) 표적 DNA 서열과 하이브리드화할 수 있는 3' 가이드 서열, 및 b) 5' 직접 반복부 서열을 포함하는 crRNA, 및 iii) tracr RNA를 포함하여서, crRNA 및 tracr RNA와 복합체형성된 Cas12b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성되는 것인 하나 이상의 비천연 발생 또는 조작된 조성물과 접촉시키는 단계를 포함하고, 여기서 가이드 서열이 세포 내 표적 DNA와의 서열-특이적 결합을 유도하여, 관심 표적 유전자좌의 발현이 변형된다.

[0996] 방법은 표적 유전자의 발현을 변형시키기 위해 사용될 수 있다. 변형은 시스템 또는 조성물에 의한 처리없이 또는 그 전에 표적 유전자의 발현과 비교하여 표적 유전자의 발현을 변경시킬 수 있다. 변형은 시스템 또는 조성물에 의한 처리없이 또는 그 전에 표적 유전자의 발현과 비교하여 표적 유전자의 발현을 증가시킬 수 있다. 변형은 시스템 또는 조성물에 의한 처리없이 또는 그 전에 표적 유전자의 발현과 비교하여 표적 유전자의 발현을 감소시킬 수 있다.

[0997] 일부 구현예에서, 방법은 표적 올리고뉴클레오티드의 하나 이상의 염기 (예를 들어, 아데닌 또는 시토신)을 변형시키는 단계를 포함할 수 있다. 이러한 방법은 표적 올리고뉴클레오티드로 본 명세서의 염기 편집자의 하나 이상의 성분을 전달하는 단계를 포함한다. 일부 예에서, 본 개시는 상기 표적 올리고뉴클레오티드에, 촉매적 불활성 Cas12b 단백질; 직접 반복부에 연결된 가이드 서열을 포함하는 가이드 분자; 및 아데노신 또는 시티딘 디아미나제 단백질 또는 이의 촉매 도메인을 전달하는 단계를 포함하고, 상기 아데노신 또는 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 전달 이후에 상기 촉매적 불활성 Cas12b 단백질에 공유적으로 또는 비공유적으로 연결되거나 또는 상기 가이드 분자는 그에 연결되도록 조정되고, 상기 가이드 분자는 상기 촉매적 불활성 Cas12b와 복합체를 형성하고 상기 복합체가 상기 표적 올리고뉴클레오티드에 결합하도록 유도하며, 상기 가이드 서열은 상기 표적 올리고뉴클레오티드 서열 내 표적 서열과 하이브리드화할 수 있어서 올리고뉴클레오티드 듀플렉스를 형성한다. 일부 구현예에서, 시토신은 상기 올리고뉴클레오티드 듀플렉스를 형성하는 상기 표적 서열 밖에 존재하고, 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 RNA 듀플렉스 밖의 시토신을 탈아민화시키거나, 또는 (B) 시토신은 RNA 듀플렉스를 형성하는 표적 서열 내에 있고, 가이드 서열은 시토신에 상응하는 위치에 비-쌍형성 아데닌 또는 우라실을 포함하여 그 결과로 RNA 듀플렉스에 C-A 또는 C-U 미스매치를 야기시키고, 시토신 디아미나제 단백질 또는 이의 촉매 도메인은 비-쌍형성 아데닌 또는 우라실 반대쪽 RNA 듀

플렉스 내 시토신을 탈아민화시킨다. 가이드 분자는 상기 CRISPR 이펙터 단백질과 복합체를 형성하고 상기 복합체가 관심 표적 올리고뉴클레오티드 서열에 결합하도록 유도하고, 가이드 서열은 아데닌 또는 시토신을 포함하는 표적 서열과 하이브리드화할 수 있어 RNA 듀플렉스를 형성하고, 아데닌 디아미나제 단백질 또는 이의 촉매 도메인은 RNA 듀플렉스의 아데닌 또는 시토신을 탈아미드화시킨다.

[0998] 일부 구현예에서, 방법 및 시스템은 하나 이상의 샘플 중 핵산 표적 서열의 존재를 검출하는데 사용될 수 있다.

일부 구현예에서, 하나 이상의 시험관내 샘플에서 핵산 표적 서열의 존재를 검출하기 위한 시스템은 Cas12b 단백질; 표적 서열과 일정 정도의 상보성을 갖도록 설계된 가이드 서열을 포함하고, Cas12b와 복합체를 형성하도록 설계된 적어도 하나의 가이드 폴리뉴클레오티드; 및 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체를 포함하고, Cas12b는 부차적 뉴클레아제 활성을 나타내고 표적 서열에 의해 활성화될 때 올리고뉴클레오티드 기반 차폐성 구성체의 비표적 서열을 절단한다. 일정 구현예에서, Cas12b 단백질; 각각이 하나 이상의 표적 폴리펩티드 중 하나에 결합하도록 설계된 하나 이상의 검출 압타머로서, 각각은 차폐된 프로모터 프롬프터 결합 부위 또는 차폐된 프라이머 결합 부위 및 기폭제 서열 주형을 포함하는 것인 검출 압타머; 및 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체를 포함하는, 하나 이상의 시험관내 샘플에서 하나 이상의 표적 폴리펩티드의 존재를 검출하기 위한 시스템을 제공한다. 하나 이상의 시험관내 샘플에서 핵산 서열을 검출하기 위한 방법은 하나 이상의 샘플을 i) Cas12b 이펙터 단백질, ii) 표적 서열과 일정 정도의 상보성을 갖도록 설계된 가이드 서열을 포함하고, Cas12b 이펙터 단백질과 복합체를 형성하도록 설계된 적어도 하나의 가이드 폴리뉴클레오티드; 및 iii) 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체와 접촉시키는 단계를 포함할 수 있고, 상기 Cas12b 이펙터 단백질은 부차적 뉴클레아제 활성을 나타내고 올리고뉴클레오티드 기반 차폐성 구성체의 비표적 서열을 절단한다.

[0999] 다른 양상에서, 본 개시는 표적 올리고뉴클레오티드를 함유하는 세포에서 효소적 (예를 들어, 단백질가수분해적) 활성을 제공하기 위한 방법을 제공한다. 방법은 효소의 불활성 부분에 연결된 제1 Cas 단백질, 및 효소의 상보성 부분에 연결된 제2 Cas 단백질을 접촉시키는 단계를 포함할 수 있다. 효소의 활성은 효소의 불활성 부분 및 상보성 부분이 접촉될 때 재구성된다. 일부 구현예에서, 표적 올리고뉴클레오티드를 함유하는 세포에서 단백질가수분해 활성을 제공하는 방법은 a) 세포 또는 세포 개체군을, i) 단백질가수분해 효소의 불활성 부분에 연결된 제1 Cas12b 이펙터 단백질; ii) 단백질가수분해 효소의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 단백질가수분해 효소의 단백질가수분해 활성은 단백질 가수분해 효소의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 이펙터 단백질; iii) 제1 Cas12b 이펙터 단백질에 결합하고 RNA의 제1 표적 서열에 하이브리드화하는 제1 가이드; 및 iv) 제2 Cas12b 이펙터 단백질에 결합하고 RNA의 제2 표적 서열에 하이브리드화하는 제2 가이드와 접촉시켜서, 단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉하고 단백질가수분해 효소의 단백질가수분해 활성이 재구성되는 것인 단계를 포함한다.

[1000] 다른 양상에서, 본 개시는 관심 올리고뉴클레오티드를 함유하는 세포를 확인하기 위한 방법을 제공한다. 방법은 리포터의 불활성 부분에 연결된 제1 Cas 단백질, 및 리포터의 상보성 부분에 연결된 제2 Cas 단백질을 접촉시키는 단계를 포함할 수 있다. 리포터의 활성은 리포터의 불활성 부분 및 상보성 부분이 접촉될 때 재구성된다. 일부 구현예에서, 관심 올리고뉴클레오티드를 함유하는 세포를 확인하는 방법을 제공하고, 방법은 세포 내 올리고뉴클레오티드를, i) 리포터의 불활성 제1 부분에 연결된 제1 Cas12b 이펙터 단백질; ii) 리포터의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 리포터의 활성은 리포터의 제1 부분 및 상보성 부분이 접촉할 때 재구성되는 것인 이펙터 단백질; iii) 제1 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드; iv) 제2 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드; 및 v) 리포터를 포함하는 조성물과 접촉시키는 단계를 포함하고, 관심 올리고뉴클레오티드가 세포에 존재할 때 리포터의 제1 부분 및 제2 부분이 접촉하고, 그리하여 리포터의 활성이 재구성된다. 일부 예에서, 리포터는 형광성 단백질 또는 발광성 단백질이다

[1001] **비동물 유기체에서의 적용**

[1002] **C2c1-CRISPR 시스템의 식물 및 효모에의 적용**

[1003] 일반적으로, 용어 " 식물" 은 세포 분할에 의해 특징적으로 성장하고, 엽록체를 함유하며 셀룰로스로 구성된 세포벽을 갖는, 식물계 (kingdom Plantae)의 임의의 다양한 광합성, 진핵, 단세포 또는 다세포 유기체에 관한 것이다. 용어 식물은 단자엽 및 쌍자엽 식물을 포함한다. 구체적으로, 식물은 제한없이 속씨식물 및 겉씨식물, 예컨대, 아카시아, 알팔파, 아마란스, 사과, 살구, 아티초크, 물푸레나무, 아스파라거스, 아보카도, 바나나, 보리, 콩, 비트, 자작나무, 너도밤나무, 블랙베리, 블루베리, 브로콜리, 양배추, 캐비지, 카놀라,

칸탈루프, 당근, 카사바, 콜리플라워, 삼나무, 곡류, 셀러리, 밤나무, 체리, 차이니즈 캐비지, 감귤류, 귤, 클로버, 커피, 옥수수, 목화, 동부, 오이, 사이프러스, 가지, 느릅나무, 엔다이브, 유칼립투스, 펜넬, 무화과, 전나무, 제라늄, 포도, 자몽, 땅콩, 꽈리, 솔송나무, 히코리, 케일, 키위, 콜라비, 낙엽송, 상추, 리크, 레몬, 라임, 로커스트, 소나무, 공작고사리, 메이즈, 망고, 단풍나무, 멜론, 낱알 곡물, 버섯, 겨자, 견과, 오크, 귀리, 기름야자, 오크라, 양파, 오렌지, 관상용 식물 또는 꽃 또는 나무, 파파야, 야자나무, 파슬리, 파스닙, 완두콩, 복숭아, 땅콩, 배, 피트(peat), 후추, 감나무, 피젼피, 소나무, 파인애플, 플랜틴, 자두, 석류, 감자, 호박, 라디치오, 무, 유채, 라즈베리, 벼, 호밀, 수수, 잇꽃, 갯버들, 대두, 시금치, 가문비나무, 스쿼시, 딸기, 사탕무, 사탕수수, 해바라기, 고구마, 단옥수수, 탄저린, 찻잎, 담배, 토마토, 나무, 라이밀, 잔디풀, 순무, 덩굴, 호두, 물냉이, 수박, 밀, 참마, 주목나무, 및 주키니를 포함하고자 한다. 용어 식물은 또한 뿌리, 잎 및 보다 고등 식물의 특징이 되는 다 기관의 결여에 의해 일차적으로 통합되는 광독립영양생물인 조류를 포함한다.

[1004] 본 명세서에 설명된 바와 같은 C2c1 시스템을 이용하는 게놈 편집 방법은 바람직한 특성을 본질적으로 임의의 식물에 수여하는데 사용될 수 있다. 광범위한 종류의 식물 및 식물 세포 시스템은, 본 개시 내용의 핵산 구조물과 상기 언급된 각종 형질전환 방법을 이용하여 본 명세서에 설명된 바람직한 생리학적 및 작물학적 특징들에 대해, 조작될 수 있다. 바람직한 구현예에서, 조작을 위한 표적 식물 및 식물 세포는 단자엽 및 쌍자엽 식물, 예컨대, 곡물 (예를 들어, 밀, 메이즈, 벼, 낱알 곡물, 보리), 과일 작물 (예를 들어, 토마토, 사과, 배, 딸기, 오렌지), 사료 작물 (예를 들어, 알팔파), 뿌리 식물 작물 (예를 들어, 당근, 감자, 사탕무, 참마), 잎줄기채소 작물 (예를 들어, 상추, 시금치); 개화 식물 (예를 들어, 페튜니아, 장미, 국화), 송백류 및 소나무 (예를 들어, 소나무 전나무, 가문비나무); 식물 환경 정화에서 사용되는 식물 (예를 들어, 중금속 축적 식물); 기름 작물(예를 들어, 해바라기, 평지씨) 및 실험 목적으로 사용되는 식물 (예를 들어, 애기장대)를 포함하는 작물을 포함하지만, 이들로 제한되지 않는다. 따라서, 방법 및 CRISPR-Cas 시스템은 매우 다양한 식물, 예컨대, 마그니오랄레스(Magniolales), 일리시알레스(Illiciales), 라우랄레 스(Laurales), 피페랄레스(Piperales), 아리스토치알레스(Aristochiales), 님파에알레스(Nymphaeales), 라눈쿠 랄레스(Ranunculales), 파페베랄레스(Papeverales), 사라세니아케아이(Sarraceniaceae), 트로코덴드랄레스(Trochodendrales), 하마멜리달레스(Hamamelidales), 유코미알레스(Eucomiales), 레이트네리알레스(Leitneriales), 미리칼레스(Myricales), 파갈레스(Fagales), 카수아리날레스(Casuarinales), 카리오필랄레스(Caryophyllales), 바탈레스(Batales), 폴리고날레스(Polygonales), 플룸바지날레스(Plumbaginales), 딜레니알 레스(Dilleniales), 테알레스(Theales), 말발레스(Malvales), 우르티칼레스(Urticales), 레시티달레스(Lecythidales), 비올랄레스(Violales), 살리칼레스(Salicales), 카파랄레스(Capparales), 에리칼레스(Ericales), 디아펜살레스(Diapensales), 에베날레스(Ebenales), 프리무랄레스(Primulales), 로살레스(Rosales), 파발레스(Fabales), 포도스테말레스(Podostemales), 할로라갈레스(Haloragales), 미르탈레스(Myrtales), 코르날레스(Cornales), 프로테알레스(Proteales), 산탈레스(Santales), 라플레시알레스(Rafflesiales), 셀라스트랄레스(Celastrales), 유포르비알레스(Euphorbiales), 람날레스(Rhamnales), 사핀달 레스(Sapindales), 유글란달레스(Juglandales), 게라니알레스(Geraniales), 폴리갈랄레스(Polygalales), 움벨랄레스(Umbellales), 겐티아날레스(Gentianales), 폴레모니알레스(Polemoniales), 라미알레스(Lamiales), 플란타지날레스(Plantaginales), 스크로풀라리알레스(Scrophulariales), 캄파누랄레스(Campanulales), 루비알레스(Rubiales), 딥사칼레스(Dipsacales) 및 아스테랄레스(Asterales) 목에 속하는 쌍자엽 식물에 대해 사용될 수 있고; 상기 방법 및 CRISPR-Cas 시스템은 알리스마탈레스(Alismatales), 히드로차리탈레스(Hydrocharitales), 나자달레스(Najadales), 트리우리달레스(Triuridales), 콤멜리날레스(Commelinales), 에리오카우랄레스(Eriocaulales), 레스티오날레스(Restionales), 포알레스(Poales), 준칼레스(Juncales), 시페랄레스(Cyperales), 티팔레스(Typhales), 브로멜리알레스(Bromeliales), 진기베랄레스(Zingiberales), 아레칼레스(Arecales), 시클란탈레스(Cyclanthales), 판다날레스(Pandanales), 아랄레스(Arales), 릴리알레스(Lilliales) 및 오르치달레스(Orchidales) 목에 속하는 것과 같은 단자엽식물, 또는 겉씨식물문, 예를 들어 피날레스(Pinales), 징코알레스(Ginkgoales), 시카달레스(Cycadales) 및 네탈레스(Gnetales)에 사용될 수 있다.

[1005] 본 명세서에 기재된 CRISPR-C2c1 시스템 및 사용 방법은 이하의 쌍자엽, 단자엽 또는 겉씨씩물 속의 비제한적 열거에 포함되는 광범위한 식물 종에 걸쳐 사용될 수 있다: 아트로파 (Atropa), 알세오다프네 (Alseodaphne), 아나카르디움 (Anacardium), 아라치스 (Arachis), 벨리쉬 미에디아 (Beilschmiedia), 브라시카 (Brassica), 카르타무스 (Carthamus), 코쿠루스 (Cocculus), 크로톤 (Croton), 쿠쿠미스 (Cucumis), 시트러스 (Citrus), 시트룰루스 (Citrullus), 캅시쿰 (Capsicum), 카타란투스 (Catharanthus), 코코스 (Cocos), 코페아 (Coffea), 쿠쿠르비타 (Cucurbita), 다우쿠스 (Daucus), 두구에티아 (Duguetia), 에스크스 콜지아 (Eschscholzia), 피쿠스

(Ficus), 프라가리아 (Fragaria), 글라우시움 (Glaucium), 글리신 (글리신), 고시 피움 (Gossypium), 헬리안투스 (Helianthus), 헤베아 (Hevea), 히오시아무스 (Hyoscyamus), 락투카 (Lactuca), 란돌피아 (Landolphia), 리눔 (Linum), 리트세아 (Litsea), 리코페르시콘 (Lycopersicon), 루피누스 (Lupinus), 마니호트 (Manihot), 마조라나 (Majorana), 말루스 (Malus), 메디카고 (Medicago), 니코티아나 (Nicotiana), 올레아 (Olea), 파르테니움 (Parthenium), 파파베르 (Papaver), 페르세아 (Persea), 파세올루스 (Phaseolus), 피스타치아 (Pistacia), 피숨 (Pisum), 피루스 (Pyrus), 프루누스 (Prunus), 라파누스 (Raphanus), 리치누스 (Ricinus), 세네치오 (Senecio), 시노메니움 (Sinomenium), 스테파니아 (Stephania), 시나피스 (Sinapis), 솔라눔 (Solanum), 테오브로마 (Theobroma), 트리폴리움 (Trifolium), 트리고넬라 (Trigonella), 비키아 (Vicia), 빈카 (Vinca), 비티스 (Vitis) 및 비그나 (Vigna); 및 알리움 (Allium), 안드로포곤 (Andropogon), 아라그로스티스 (Aragrostis), 아스파라거스 (Asparagus), 아베나 (Avena), 시노돈 (Cynodon), 엘라에리스 (Elaeis), 페스투카 (Festuca), 페스투로리움 (Festulolium), 헤테로칼리스 (Heterocalis), 호르데움 (Hordeum), 렘나 (Lemna), 롤리움 (Lolium), 무사 (Musa), 오리자 (Oryza), 파니쿰 (Panicum), 판네세툼 (Pannesetum), 플레움 (Phleum), 포아 (Poa), 세칼레 (Secale), 소르굼 (Sorghum), 트리티쿰 (Triticum) 및 제아(Zea) 속 외떡잎식물; 또는 아비에스 (Abies), 쿤닝하미아 (Cunninghamia), 피세아 (Picea), 피누스 (Pinus) 및 프세우도추가 (Pseudotsuga).

[1006] CRISPR-C2c1 시스템 및 사용 방법은 또한, 예를 들어, 홍조식물 (Rhodophyta) (홍조류), 녹색식물문 (Chlorophyta) (녹조류), 갈조식물문 (Phaeophyta) (갈조류), 규조문 (Bacillariophyta) (규조류), 유스티그마토피타 (Eustigmatophyta) 및 와편모류 (dinoglagellates) 를 포함하는 몇몇 진핵생물문뿐만 아니라 원핵생물문 남세균 (cyanobacteria) (청-녹조류)로부터 선택된 조류를 비롯한, 광범위한 "조류" 또는 "조류 세포" 에 대해 사용될 수 있다. 용어 "조류"는 예를 들어, 암포라 (Amphora), 아나바나 (Anabaena), 아니크스트로데스미스 (Anikstrodesmis), 보트리오코커스 (Botryococcus), 차에토세로스 (Chaetoceros), 클라미도모나스 (Chlamydomonas), 클로렐라 (Chlorella), 클로로코쿰 (Chlorococcum), 시클로텔라 (Cyclotella), 실린드로테카 (Cylindrotheca), 두날리엘라 (Dunaliella), 에밀리아나 (Emiliana), 유글레나 (Euglena), 헤마토코커스 (Hematococcus), 이소크리시스 (Isochrysis), 모노크리시스 (Monochrysis), 모노라피디움 (Monoraphidium), 나노클로리스 (Nannochloris), 나노클로롭시스 (Nannnochloropsis), 나비쿨라 (Navicula), 네프로클로리스 (Nephrochloris), 네프로셀미스 (Nephroselmis), 니츠키아 (Nitzschia), 노둘라리아 (Nodularia), 노스톡 (Nostoc), 오크로모나스 (Oochromonas), 우시스티스 (Oocystis), 오실라르토리아 (Oscillartoria), 파블로바 (Pavlova), 파에오닥틸룸 (Phaeodactylum), 플라이트모나스 (Playtmonas), 플레우로크리시스 (Pleurochrysis), 포르히라 (Porhyra), 슈도아나바나 (Pseudoanabaena), 피라미모나스 (Pyramimonas), 스티코코커스 (Stichococcus), 시네코코커스 (Synechococcus), 시네코시스티스 (Synechocystis), 테트라셀미스 (Tetraselmis), 탈라시오시라 (Thalassiosira) 및 트리코데스뮴 (Trichodesmium)으로부터 선택된 조류를 포함한다.

[1007] 식물의 부분, 즉 "식물 조직" 은 본 발명의 방법에 따라 처리되어 개선된 식물을 생산할 수 있다. 식물 조직은 또한 식물 세포를 포함한다.본 명세서에서 사용되는 용어 "식물 세포" 는 무손상 전체 세포에서 또는 시험관내 조직 배양으로부터 성장된 단리 형태로, 배지 또는 한천 상에서, 성장 배지 내 현탁액 또는 완충제에서 또는 더 고등의 조직 단위의 일부, 예컨대, 식물 조직, 식물 기관 또는 전체 식물로서의, 살아있는 식물의 개개 단위를 의미한다.

[1008] "원형질" 이라는 용어는, 예를 들어 기계적 또는 효소적 수단을 이용하여 그의 보호성 세포 벽이 완전히 또는 부분적으로 제거된 식물 세포를 지칭하며, 그 결과 살아있는 식물의 온전한 생화학 컴피턴트 단위가 결과로서 초래되며, 이는 적절한 성장 조건 하에서 그의 세포 벽을 재형성하고, 번식 및 재생하여 전체 식물로 성장할 수 있다.

[1009] 용어 "형질전환" 은 식물 숙주가 아그로박테리아에 의해서 또는 다양한 화학적 또는 물리적 방법 중 하나에 의해서 DNA의 도입에 의해 유전자 변형되는 과정을 광범위하게 의미한다. 본 명세서에서 사용되는 용어 " 식물 숙주" 는 식물의 임의의 세포, 조직, 기관 또는 자손을 포함하는 식물을 의미한다. 다수의 적합한 식물 조직 또는 식물 세포는 형질전환될 수 있고, 원형질체, 체성 배아, 화분, 잎, 묘목, 줄기, 캘러스, 주근, 미세괴경 및 순을 포함하지만, 이들로 제한되지 않는다. 식물 조직은 또한 이러한 식물, 종자, 자손, 유성 또는 무성 생식이건 간에 번식체, 및 이들 중 어느 것의 후손, 예컨대, 꺾꽂이 순 또는 종자의 임의의 클론을 의미한다.

[1010] 본 명세서에서 사용되는 용어 "형질전환된" 은 외래 DNA 분자, 예컨대, 구성체가 도입되는 세포, 조직 또는 유기체를 의미한다. 도입된 DNA 분자가 후속 자손으로 전달되도록, 도입된 DNA 분자는 수용 세포, 조직, 기관

또는 유기체의 게놈 DNA에 통합될 수 있다. 이들 구현예에서, "형질전환된" 또는 "유전자이식" 세포 또는 식물은 또한 교배에서 모체로서 이러한 형질전환된 식물을 사용하고 도입된 DNA 분자의 존재로부터 초래되는 변경된 표현형을 나타내는 육종 프로그램으로부터 생성된 자손 또는 세포 또는 식물의 자손을 포함할 수 있다. 바람직하게는, 유전자이식 식물은 생식력이 있으며, 도입된 DNA가 성적 번식을 통해 자손에게 전달될 수 있다.

[1011] 용어 "자손", 예컨대, 유전자이식 식물의 자손은 식물 또는 유전자이식 식물로부터 기인하거나, 야기되거나 또는 유래되는 것이다. 도입된 DNA 분자는 또한 도입된 DNA 분자가 후속 자손에 의해 유전되지 않고, 따라서 "유전자이식" 으로 고려되지 않도록 수용 세포 내로 일시적으로 도입될 수 있다. 따라서, 본 명세서에서 사용되는, "비-유전자이식" 식물 또는 식물 세포는 게놈 내로 안정하게 통합된 외래 DNA를 함유하지 않는 식물이다.

[1012] 본 명세서에서 사용되는 용어 "식물 프로모터" 는 식물 유래이건 그렇지 않건 간에 식물 세포에서 전사를 개시할 수 있는 프로모터이다. 예시적인 적합한 식물 프로모터는 식물 세포에서 발현되는 유전자를 포함하는 식물, 식물 바이러스 및 박테리아, 예컨대, 아그로박테리움 또는 리조비움 (Rhizobium)으로부터 얻어지는 것을 포함하지만, 이들로 제한되지 않는다.

[1013] 본 명세서에서 사용되는 "진균 세포" 는 진균류 내의 임의의 유형의 진행 세포를 의미한다. 진균류 내의 문은 아스코미코타 (Ascomycota), 바시디오마이코타 (Basidiomycota), 블라스토클라디오마이코타 (Blastocladiomycota), 키트리디오마이코타 (Chytridiomycota), 글로메로마이코타 (Glomeromycota), 마이크로스포리디아 (Microsporidia), 및 네오칼리마스티고마이코타 (Neocallimastigomycota)를 포함한다. 진균 세포는 효모, 곰팡이 및 사상성진균을 포함할 수 있다. 일부 구현예에서, 진균 세포는 효모 세포이다.

[1014] 본 명세서에서 사용되는 용어 "효모 세포" 는 아스코미코타 및 바시디오마이코타 문 내의 임의의 진균 세포를 의미한다. 효모 세포는 출아 효모 세포, 분열 효모 세포 및 곰팡이 세포를 포함할 수 있다. 이들 유기체로 제한되지 않지만, 실험실 및 산업 환경에서 사용되는 다수 유형의 효모는 아스코미코타 문의 일부이다. 일부 구현예에서, 효모 세포는 에스. 세레비지아에 (S. cerervisiae), 클루이베로마이세스 마르시아누스 (Kluyveromyces marxianus), 또는 아사켄키아 오리엔탈리스 (Issatchenkia orientalis) 세포이다. 다른 효모 세포는 칸디다 종 (Candida spp.)(예를 들어, 칸디다 알비칸스 (Candida albicans)), 야로위아 종 (Yarrowia spp.) (예를 들어, 야로위아 리폴리티카 (Yarrowia lipolytica)), 피키아 종 (Pichia spp.)(예를 들어, 피키아 파스토리스 (Phichia pastoris), 클루이베로마이세스 종 (Kluyveromyces spp.) (예를 들어, 클루이베로마이세스 락티스 (Kluyveromyces lactis) 및 클루이베로마이세스 마르시아누스), 뉴로스포라 종 (Neurospora spp.) (예를 들어, 뉴로스포라 크라사 (Neurospora crassa)), 푸사리움 종 (Fusarium spp.) (예를 들어, 푸사리움 옥시스포룸 (Fusarium oxysporum)), 및 이사켄키아 종 (Issatchenkia spp.) (예를 들어, 아사켄키아 오리엔탈리스 (Issatchenkia orientalis) (피키아 쿠드리아베제비 (Pichia kudriavzevii)라고도 함) 및 칸디다 아시도써모필룸 (Candida acidothermophilum))을 포함할 수 있지만, 이들로 제한되지 않는다. 일부 구현예에서, 진균 세포는 사상 진균 세포이다. 본 명세서에서 사용되는 용어 "사상 진균 세포" 는 필라멘트, 즉, 균사 또는 균사체로 성장하는 임의의 유형의 진균 세포를 지칭한다. 섬유성 균류 세포의 예로는 이로 제한되지 않지만, 누룩곰팡이 종 (Aspergillus spp.)(예를 들어, 아스페르길루스 나이거(Aspergillus niger)), 트리코더마 종(Trichoderma spp.)(예를 들어, 트리코더마 리세이(Trichoderma reesei)), 리조푸스 종 (Rhizopus spp.)(예를 들어, 리조푸스 오리재(Rhizopus oryzae)), 및 모르티에렐라 종(Mortierella spp.)(예를 들어, 모르티에렐라 이사벨리나)가 포함될 수 있다.

[1015] 일부 구현예에서, 균류 세포는 산업용 균주이다. 본 명세서에서 사용된 바와 같은 "산업용 균주" 는 산업적 공정, 예를 들어 상업적 또는 산업적 규모에서의 생성물의 생산으로부터 사용 또는 분리된 균주 세포의 임의의 균주를 지칭한다. 산업용 균주는 산업적 공정에서 통상적으로 사용되는 균류 종을 지칭할 수 있거나, 비-산업적 목적에도 사용될 수 있는(예를 들어, 실험실 연구) 균류 종의 분리물을 지칭할 수 있다. 산업적 공정의 예는 (예를 들어, 식품 또는 음료 제품의 생산에서) 발효, 증류, 생물연료 생산, 화합물의 생산 및 폴리펩티드의 생산을 포함할 수 있다. 산업용 균주의 예는 제한 없이, JAY270 및 ATCC4124를 포함할 수 있다.

[1016] 일부 구현예에서, 균류 세포는 배수체 세포이다. 본 명세서에서 사용된 바와 같은 "배수체" 세포는 그의 게놈이 하나 초과의 복사물로 존재하는 임의의 세포를 지칭할 수 있다. 배수체 세포는 자연에서 배수체 상태로 발견되는 세포의 유형을 지칭할 수 있거나, 배수체 상태로 존재하도록 유도된 세포를 지칭할 수 있다(예를 들어, 특이적 조절, 변경, 비활성화, 활성화, 또는 감수분열의 변형, 세포질 분열, 또는 DNA 복제를 통해).

배수체 세포는 그의 전체 게놈이 배수체인 세포를 지칭할 수 있거나, 특정한 관심 게놈 유전자좌 내 배수체인 세포를 지칭할 수 있다. 이론에 한정하고자 하지는 않지만, 가이드 RNA의 존재비는 반수체 세포보다 배수체 세포의 게놈 조작에서 종종 더욱 속도-제한적인 성분일 수 있고, 따라서 본 명세서에 기술된 C2c1 CRISPR 시스템을 사용하는 방법은 일정 진균류 세포 유형을 사용하는 것을 활용할 수 있다.

[1017] 일부 구현예에서, 균류 세포는 이배체 세포이다. 본 명세서에서 사용된 바와 같은 "이배체" 세포는 그의 게놈이 두 개의 복사물로 존재하는 임의의 세포를 지칭할 수 있다. 이배체 세포는 이배체 상태에서 자연적으로 발견되는 세포 유형을 지칭할 수 있거나, 또는 (예를 들어, 특정 조절, 변경, 불활성화, 활성화 또는 감수분열, 세포질분열 또는 DNA 복제의 변형을 통해) 이배체 상태로 존재하도록 유도된 세포를 지칭할 수 있다. 예를 들어, 사카로마이세스 세레비지아 균주 S228C는 반수체 또는 이배체 상태에서 유지될 수 있다. 이배체 세포는 전체 게놈이 이배체인 세포를 지칭할 수 있거나, 또는 관심 특정 게놈 유전자좌에서 이배체인 세포를 지칭할 수 있다. 일부 구현예에서, 균류 세포는 반수체 세포이다. 본 명세서에서 사용된 바와 같은 "반수체" 세포는 그의 게놈이 하나의 복사물로 존재하는 임의의 세포를 지칭할 수 있다. 반수체 세포는 자연에서 반수체 상태로 발견되는 세포의 유형을 지칭할 수 있거나, 반수체 상태로 존재하도록 유도된 세포를 지칭할 수 있다 (예를 들어, 특이적 조절, 변경, 비활성화, 활성화, 또는 감수분열의 변형, 세포질 분열, 또는 DNA 복제를 통해). 예를 들어, 사카로마이세스 세레비시에 균주 S228C는 반수체 또는 이배체 상태로 유지될 수 있다. 반수체 세포는 그의 전체 게놈이 반수체인 세포를 지칭할 수 있거나, 특정한 관심 게놈 유전자좌 내 반수체인 세포를 지칭할 수 있다.

[1018] 본 명세서에서 사용된 바와 같은 "효모 발현 벡터" 는 RNA 및/또는 폴리펩티드를 코딩하는 하나 이상의 서열을 함유하는 핵산을 지칭하며, 핵산(들)의 발현을 제어하는 임의의 바람직한 요소뿐만 아니라, 효모 세포 내부의 발현 벡터의 복제 및 유지를 가능하게 하는 임의의 요소를 추가로 함유할 수 있다. 다수의 적합한 효모 발현 벡터 및 이의 특징은 당업계에 공지되어 있으며; 예를 들어, 다양한 벡터 및 기법은 문헌 [Yeast Protocols, 2nd edition, Xiao, W., ed. (Humana Press, 뉴욕, 2007)] 및 [Buckholz, R.G. and Gleeson, M.A. (1991) Biotechnology (NY) 9(11): 1067- 72]에 예시되어 있다. 효모 벡터는 제한 없이, 동원체 (CEN) 서열, 자율 증식 서열 (ARS), 프로모터, 예컨대, 관심 서열 또는 유전자에 작동적으로 연결된, RNA 폴리머라제 III 프로모터, RNA 폴라머라제 III 종결자와 같은 종결자, 복제 기원, 및 마커 유전자 (예를 들어, 영양요구성, 항생성, 또는 다른 선별 마커)를 함유할 수 있다. 효모에서의 이용을 위한 발현 벡터의 예로는, 플라스미드, 효모 인공 염색체, 2μ 플라스미드, 효모 편입형 플라스미드, 효모 자가증식형 플라스미드, 셔틀 벡터, 및 에피솜 유사 플라스미드를 포함할 수 있다.

[1019] **식물 및 식물 세포의 게놈 내 CRISPR-C2c1 시스템 성분의 안정한 통합**

[1020] 특정 구현예에서, CRISPR-C2c1 시스템의 성분을 코딩하는 폴리뉴클레오티드가 식물 세포의 게놈 내로 안정한 통합을 위해 도입되는 것이 고려된다. 이들 구현예에서, 형질전환 벡터 또는 발현 시스템의 설계는, 가이드 RNA 및/또는 C2c1 유전자가 언제, 어디에서 어떤 조건에서 발현되는지에 따라 조절될 수 있다.

[1021] 특정 구현예에서, 식물 세포의 게놈 DNA 내로 CRISPR-C2c1 시스템의 성분을 안정하게 도입하는 것이 계획된다. 추가적으로 또는 대안적으로, 식물 세포기관, 예컨대, 제한없이, 색소체, 미토콘드리아 또는 엽록체의 DNA 내로 안정한 통합을 위해 CRISPR-C2c1 시스템 성분을 도입하는 것이 고려된다.

[1022] 식물 세포 게놈 내로 안정한 통합을 위한 발현 시스템은 다음의 구성요소 중 하나 이상을 함유할 수 있다: 식물 세포에서 RNA 및/또는 C2c1 효소를 발현시키기 위해 사용될 수 있는 프로모터 구성요소; 발현을 증강시키기 위한 5' 비번역 영역; 소정의 세포, 예컨대, 단자엽 세포에서 발현을 추가로 증강시키는 인트론 구성요소; 가이드 RNA 및/또는 융합 단백질 및 C2c1 유전자 서열 및 다른 목적하는 요소를 삽입하기 위한 편리한 제한 부위를 제공하기 위한 다중 클로닝 부위; 및 발현된 전사체의 효율적인 종결을 제공하기 위한 3' 비번역 영역.

[1023] 발현 시스템의 요소는 플라스미드 또는 형질전환 벡터와 같은 환형, 또는 선형 이중 가닥 DNA와 같은 비-환형 중 어느 하나인 하나 이상의 발현 구축물 상에 있을 수 있다.

[1024] 특정 구현예에서, C2c1 CRISPR 발현 시스템은 적어도, 식물에서 표적 서열과 하이브리드화하는 가이드 RNA(gRNA)를 코딩하는 뉴클레오티드 서열로서, 가이드 RNA는 가이드 서열 및 직접 반복부 서열을 포함하는 서열, tracr RNA를 코딩하는 뉴클레오티드 서열, 및 C2c1 단백질을 코딩하는 뉴클레오티드 서열을 포함하고, 여기서 성분 (a) 또는 성분 (b) 또는 (c)는 동일하거나 상이한 구축물 상에 위치되며, 이로 인해 상이한 뉴클레오티드 서열은 식물 세포 내에서 동일하거나 상이한 조절 엘리먼트의 제어 하에 있을 수 있다. Tracr은 가이

드 RNA에 융합될 수 있다.

[1025] CRISPR-C2c1 시스템의 성분, 및, 적용 가능한 경우, 주형 서열을 함유하는 DNA 구성체(들)는 다양한 통상적인 기법에 의해 식물, 식물 부분 또는 식물 세포 게놈 내로 도입될 수 있다. 프로세스는 일반적으로 적합한 숙주 세포 또는 숙주 조직을 선택하는 단계, 구축물(들)을 숙주 세포 또는 숙주 조직 내로 도입하는 단계, 식물 세포 또는 그로부터의 식물을 재생하는 단계를 포함한다.

[1026] 특정 구현예에서, DNA 구축물은, 이로 제한되지는 않지만 예컨대 식물 세포 원형질체의 전기천공, 미세주입, 에어로졸 빔 주사 기술을 이용하여 식물 세포 내로 도입될 수 있거나, DNA 구축물은 DNA 유전자 총(particle bombardment)과 같은 바이오리스틱 방법을 이용하여 식물 조직에 직접 도입될 수 있다(참조: Fu et al., Transgenic Res. 2000 Feb;9(1):11-9) 유전자 총의 기준은 입자에 의한 원형질의 침투 및 전형적으로 게놈 내로의 안정한 통합을 초래하는, 세포에 대해 관심 유전자로 코팅된 입자의 가속도이다. (참조: Klein et al, Nature (1987), Klein et ah, Bio/Technology (1992), Casas et ah, Proc. Natl. Acad. Sci. USA (1993).).

[1027] 특정 구현예에서, CRISPR-C2c1 시스템의 성분을 함유하는 DNA 구성체는 아그로박테리움-매개 형질전환에 의해 식물 내로 도입될 수 있다. DNA 구성체는 적합한 T-DNA 측접 영역과 합쳐지고, 통상적인 아그로박테리움 투메파시엔스 (Agrobacterium tumefaciens) 숙주 벡터 내로 도입될 수 있다. 외래 DNA는 식물을 감염시킴으로써 또는 하나 이상의 Ti (종양-유도성) 플라스미드를 함유하는 식물 원형질체를 아그로박테리움 박테리아와 함께 인큐베이션시킴으로써 게놈 내로 혼입될 수 있다. (참조: 예를 들어, Fraley et al., (1985), Rogers et al., (1987), 및 미국 특허 번호 5,563,055).

[1028] **식물 프로모터**

[1029] 식물 세포에서 적절한 발현을 보장하기 위해, 본 명세서에 기재된 CRISPR-C2c1 시스템의 성분은 전형적으로 식물 프로모터, 즉, 식물 세포에서 작동가능한 프로모터의 제어 하에 위치된다. 상이한 유형의 프로모터들의 이용이 고려된다.

[1030] 항상성 식물 프로모터는 식물의 모든 또는 거의 모든 발생 단계 동안 모든 또는 거의 모든 식물 조직에서 제어되는 오픈 리딩 프레임 (ORF)을 발현 ("항상성 발현" 이라함)시킬 수 있는 프로모터이다. 항상성 프로모터의 한 가지 비제한적 예는 콜리플라워 모자이크 바이러스 35S 프로모터이다. "조절된 프로모터" 는 항성성이 아니지만, 시간적으로 및/또는 공간적으로 조절된 방식으로 유전자 발현을 지시하는 프로모터를 지칭하고, 조직-특이적, 조직-선호성 및 유도성 프로모터를 포함한다. 상이한 프로모터는 상이한 조직 또는 세포 유형에서 또는 상이한 발생 단계 또는 상이한 환경 조건에 반응하여 유전자 발현을 지정할 수 있다. 특정 구현예에서, 하나 이상의 C2c1 CRISPR 성분은 구조적 프로모터, 예컨대 콜리플라워 모자이크 바이러스 35S 프로모터의 제어 하에 발현된다. 조직-선호된 프로모터는 특정 식물 조직 내에서 소정의 세포 유형 내, 예를 들어, 잎 또는 뿌리 내의 맥관 세포 또는 씨드의 특정 세포 내 증진된 발현을 표적화하는데 사용될 수 있다. CRISPR-C2c1 시스템에서 사용하기 위한 특정 프로모터의 예는 문헌 [Kawamata et al., (1997) Plant Cell Physiol 38:792803]; [Yamamoto et al., (1997) Plant J 12:25565]; [Hire et al, (1992) Plant Mol Biol 20:20718], [Kuster et al, (1995) Plant Mol Biol 29:75972], 및 [Capana et al., (1994) Plant Mol Biol 25:681 -91]에서 확인된다.

[1031] 유도성이며, 유전자 편집 또는 유전자 발현의 시공적 제어를 가능하게 하는 프로모터의 예는 에너지 형태를 사용할 수 있다. 에너지 형태는 소리 에너지, 전자기 복사, 화학적 에너지 및/또는 열 에너지를 포함할 수 있지만, 이들로 제한되지 않는다. 유도성 시스템의 예는 테트라사이클린 유도성 프로모터 (Tet-온 또는 Tet-오프), 소형 분자 2-하이브리드 전사 활성화 시스템 (FKBP, ABA 등), 또는 광 유도성 시스템 (피토크롬, LOV 도메인 또는 크립토크롬), 예컨대, 서열-특이적 방식으로 전사 활성의 변화를 지시하는 광 유도성 전사 이펙터 (LITE)를 포함한다. 광 유도성 시스템의 성분은 C2c1 CRISPR 효소, 광-반응성 시토크롬 이형이량체(예를 들어, 애기장대로부터), 및 전사 활성화/억제 도메인을 포함할 수 있다. 유도성 DNA 결합 단백질 및 이들의 사용 방법의 추가적인 예는 본 명세서에 전문이 참고로 편입된 US 61/736465 및 US 61/721,283에서 제공된다.

[1032] 특정 구현예에서, 일시적 또는 유도성 발현은, 예를 들어, 화학적-조절 프로모터를 이용함으로써 달성될 수 있고, 즉, 이에 의해 외생성 화학물질의 적용은 유전자 발현을 유도한다. 유전자 발현의 조절은 또한 화학물질-억제 프로모터에 의해 얻어질 수 있으며, 화학물질의 적용은 유전자 발현을 억제한다. 화학물질-유도성 프로모터는 벤젠 설폰아미드 제초제 약해 경감제에 의해 활성화되는 메이즈 ln2-2 프로모터 (De Veylder et

al., (1997) Plant Cell Physiol 38:56877), 발아전 제초제로서 사용되는 소수성 친전자성 화합물에 의해 활성화되는 메이즈 GST 프로모터 (GST-11-27, WO93/01294), 및 살리실산에 의해 활성화되는 담배 PR-1 프로모터 (Ono et al., (2004) Biosci Biotechnol Biochem 68:8037)를 포함하지만, 이들로 제한되지 않는다. 항생제에 의해 조절되는 프로모터, 예컨대, 테트라사이클린-유도성 및 테트라사이클린-억제성 프로모터 (문헌 [Gatz et al., (1991) Mol Gen Genet 227:22937]; 미국 특허 제5,814,618호 및 제5,789,156호 참조)가 또한 본 명세서에서 사용될 수 있다.

[1033] **특별한 식물 세포소기관에서 전좌 및/또는 발현**

[1034] 발현 시스템은 특별한 식물 세포소기관으로의 전좌 및/또는 그에서의 발현을 위한 구성요소를 포함할 수 있다.

[1035] 엽록체 표적화 특정 구현예에서, CRISPR-C2c1 시스템은 엽록체 유전자를 특이적으로 변형시키기 위해 또는 엽록체에서의 발현을 보장하기 위해 사용되는 것이 고려된다. 이러한 목적을 위해, 엽록체 형질전환 방법 또는 C2c1 CRISPR 시스템 성분의 엽록체로의 구획화를 이용한다. 예를 들어, 색소체 게놈 내 유전적 변형의 도입은 화분을 통한 유전자 흐름과 같은 생물안전성 문제를 감소시킬 수 있다.

[1036] 엽록체 형질전환 방법은 해당 기술 분야에 알려져 있으며, 유전자 총법, PEG 처리, 및 미세주입을 포함한다. 추가적으로, 핵 게놈으로부터 색소체로의 형질전환 카세트 전좌를 수반하는 방법은 WO2010061186에 기재된 바와 같이 사용될 수 있다.

[1037] 대안적으로, 하나 이상의 C2c1 CRISPR 시스템 성분을 식물 엽록체로 표적화하는 것이 고려된다. 이는 발현 구성체 내로, C2c1 단백질을 코딩하는 서열의 5' 영역에 작동적으로 연결된, 엽록체 이동 펩티드 (CTP) 또는 색소체 이동 펩티드를 코딩하는 서열을 통합시킴으로써 달성된다. CTP는 엽록체 내로의 전좌 동안 가공 단계에서 제거된다. 발현된 단백질의 엽록체 표적화는 숙련자에게 공지이다(참조: 예를 들어, Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology,Vol. 61: 157-180). 이러한 구현예에서, 또한 식물 엽록체에 대해 가이드 RNA를 표적화하는 것이 바람직하다. 엽록체 국소화 서열에 의해 엽록체에 가이드 RNA를 전좌시키기 위해 사용될 수 있는 방법 및 구성체는, 예를 들어, 본 명세서에 참고로 편입된 US 20040142476에 기재되어 있다. 구성체의 그러한 변이는 본 발명의 발현 시스템 내로 통합되어 C2c1-가이드 RNA를 효율적으로 전위시킬 수 있다.

[1038] **CRISPR-C2c1 시스템을 코딩하는 폴리뉴클레오티드의 조류 세포 내 도입**

[1039] 유전자이식 조류(또는 유채와 같은 다른 식물)는 채소유 또는 생물연료 예컨대 알코올(특히 메탄올 및 에탄올) 또는 다른 생성물의 생산에서 특히 유용할 수 있다. 이들은 오일 또는 생체연료 산업에서 사용을 위한 고수준의 오일 또는 알코올을 발현 또는 과발현시키도록 유전자조작될 수 있다.

[1040] US 8945839호는 Cas9를 이용하여 미세조류(클라미도모나스 레인하티(Chlamydomonas reinhardtii) 세포) 종을 조작하는 방법을 설명한다. 유사한 도구를 이용하여, 본 명세서에 기재된 CRISPR-C2c1 시스템의 방법은 클라미도모나스 (Chlamydomonas) 종 및 다른 조류에 적용될 수 있다. 특정 구현예에서, C2c1 및 가이드 RNA는, Hsp70A-Rbc S2 또는 베타2 -튜불린과 같은 구조 프로모터의 제어 하에서 C2c1을 발현하는 벡터를 사용하여 발현된 조류 내에 도입된다. 가이드 RNA는 T7 프로모터를 함유하는 벡터를 사용하여 선택적으로 전달된다. 대안적으로, C2c1 mRNA 및 시험관내 전사된 가이드 RNA는 조류 세포에 전달될 수 있다. 전기천공법 프로토콜, 예컨대, GeneArt Chlamydomonas Engineering kit로부터의 표준 권장 프로토콜은 당업자가 이용 가능하다.

[1041] 특정 구현예에서, 본 명세서에서 사용된 엔도뉴클레아제는 분할 C2c1 효소이다. 분할 C2c1 효소는 WO 2015086795에 Cas9에 대하여 기재된 바와 같은 표적된 게놈 변형을 위한 조류 내에서 우선적으로 사용된다. C2c1 분할 시스템의 이용은 게놈 표적화의 유도 방법에 특히 적합하며, 조류 세포 내 C2c1 과잉발현의 잠재적인 독성 효과를 회피한다. 특정 구현예에서, 상기 C2c1 분할 도메인(RuvC 및 HNH 도메인)은 동시에 또는 순차적으로 세포 내로 도입될 수 있어서, 상기 분할 C2c1 도메인(들)은 조류 세포 내에서 표적 핵산 서열을 가공한다. 야생형 C2c1에 비하여 분할 C2c1의 감소된 크기는 CRISPR 시스템의 세포로 전달하는 다른 방법들, 예컨대 본 명세서에 설명된 바와 같은 세포 침투 펩티드의 이용을 가능하게 한다. 이 방법은 유전적으로 변형된 조류를 생성하는데 특히 유익하다.

[1042] **효모 세포 내에서 C2c1 성분을 코딩하는 폴리뉴클레오티드의 도입**

[1043] 특정 구현예에서, 본 발명은 효모 세포의 게놈 편집을 위한 CRISPR-C2c1 시스템의 용도에 관한 것이다.

CRISPR-C2c1 시스템 성분을 코딩하는 폴리뉴클레오티드를 도입하는데 사용될 수 있는 효모 세포를 형질전환하는 방법은 당업자에게 공지이며, 다음의 문헌에서 검토된다: Kawai et al., 2010, Bioeng Bugs. 2010 Nov-Dec; 1(6): 395-403. 비제한적 예는 리튬 아세테이트 처리 (담체 DNA 및 PEG 처리를 추가로 포함할 수 있음), 유전자총 또는 전기천공법에 의한 효모 세포의 형질전환을 포함한다.

[1044] **식물 및 식물 세포 내에서 C2c1 CRISP 시스템 성분의 일시적 발현**

[1045] 특정 구현예에서, 가이드 RNA 및/또는 C2c1 유전자는 식물 세포 내에서 일시 발현된다고 생각된다. 이들 구현예에서, CRISPR-C2c1 시스템은, 가이드 RNA 및 C2c1 단백질 모두가 세포 내에 존재하는 때에만 표적 유전자의 변형을 보장할 수 있어서, 게놈 변형은 추가로 제어될 수 있다. C2c1 효소의 발현이 일시적임에 따라, 그러한 식물 세포로부터 생성된 식물은 통상적으로 외래 DNA를 함유하지 않는다. 특정 구현예에서 C2c1 효소는 식물 세포에 의해 안정하게 발현되고, 가이드 서열이 일시적으로 발현된다.

[1046] 특정 실시형태에서, CRISPR-기능화된 CRISPR 시스템 성분은 식물 바이러스 벡터를 이용하여 식물 세포에 도입될 수 있다 (Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299- 323). 추가의 특정 구현예에서, 상기 바이러스 벡터는 DNA 바이러스로부터의 벡터이다. 예를 들어, 제미니바이러스 (예를 들어, 캐비지 잎 말림 바이러스, 콩 누른오갈병 바이러스, 밀 오갈병 바이러스, 토마토 잎 말림 바이러스, 메이즈 도말 바이러스, 담배 잎 말림 바이러스, 또는 토마토 골든 모자이크 바이러스) 또는 나노바이러스 (예를 들어, 잠두 괴사성 황화 바이러스). 다른 특정 구현예에서, 상기 바이러스 벡터는 RNA 바이러스 유래 벡터이다. 예를 들어, 토브라 바이러스 (예를 들어, 담배 얼룩 바이러스 (tobacco rattle virus), 담배 모자이크 바이러스), 포텍스바이러스 (예를 들어, 감자 바이러스 X) 또는 호르데이바이러스 (예를 들어, 보리 줄무늬병 모자이크 바이러스). 식물 바이러스의 복제 게놈은 비통합 벡터이다.

[1047] 특정 구현예에서, C2c1 CRISPR 구성체의 일시적 발현을 위해 사용되는 벡터는, 예를 들어, 원형질체에서 아그로박테리움-매개 일시적 발현에 대해 맞춤된 pEAQ 벡터이다 (Sainsbury F. et al., Plant Biotechnol J. 2009 Sep;7(7):682-93). CRISPR 효소를 발현시키는 안정한 유전자이식 식물에서 gRNA를 발현시키기 위해 변형된 양배추 잎 말림 바이러스(CaLCuV) 벡터를 이용하는 게놈 위치의 정확한 표적화가 입증되었다 (Scientific Reports 5, Article number: 14926 (2015), doi:10.1038/srep14926).

[1048] 특정 구현예에서, 가이드 RNA 및/또는 C2c1 유전자를 코딩하는 이중-가닥 DNA 단편은 식물 세포 내로 일시적으로 도입될 수 있다. 이러한 구현예에서, 도입된 이중-가닥 DNA 단편은 세포를 변형시키기 위한 충분량으로 제공되지만, 통과된 시간의 고려되는 기간 후에 또는 1회 이상의 세포 분열 후에 지속되지 않는다. 식물에서 직접 DNA 전달 방법은 당업자에게 알려져 있다(참조: 예를 들어, Davey et al. Plant Mol Biol. 1989 Sep;13(3):273-85.)

[1049] 다른 구현예에서, C2c1단백질을 코딩하는 RNA 폴리뉴클레오티드가 식물 세포 내로 도입되고, 이는 이후 번역되고 숙주 세포에 의해 가공되어(하나 이상의 가이드 RNA 존재 하에서) 세포를 변형시키기에는 충분한 양이지만 고려된 기간이 지나거나 일 회 이상의 세포 분열 후에는 지속되지 않는 양의 단백질을 생성한다. 일시적 발현을 위해 식물 원형질체에 mRNA를 도입하는 방법은 당업자에게 공지되어 있다 (참조: 예를 들어, Gallie, Plant Cell Reports (1993), 13;119122).

[1050] 상기 설명된 상이한 방법들의 조합이 또한 고려된다.

[1051] **C2c1 CRISPR 성분의 식물 세포로의 전달**

[1052] 특정 구현예에서, CRISPR-C2c1 시스템의 하나 이상의 성분을 직접적으로 식물 세포에 전달하는 데 관심이 있다. 이는 특히, 비-유전자이식 식물의 생성에서 흥미롭다 (이하 참조). 특정 구현예에서, 하나 이상의 C2c1 성분이 식물 또는 식물 세포 외부에서 제조되고 세포로 전달된다. 예를 들어 특정 구현예에서, C2c1 단백질은 식물 세포로의 도입 전에 시험관 내에서 제조된다. C2c1 단백질은 당업자에게 알려진 각종 방법에 의해 제조될 수 있으며, 재조합 생산을 포함한다. 발현 후, C2c1 단백질이 분리되고, 필요에 따라 재접힘되고, 정제 및 선택적으로 처리되어 임의의 정제 태그, 예컨대 His-태그를 제거할수 있다. 조질의, 부분적으로 정제된, 또는 더욱 완전하게 정제된 C2c1 단백질이 일단 수득되면, 단백질은 식물 세포로 도입될 수 있다.

[1053] 특정 구현예에서, C2c1 단백질은 관심 유전자를 표적화하는 가이드 RNA와 혼합되어 사전-조립된 리보뉴클레오단백질을 형성한다.

[1054] 개별적인 성분 또는 사전-조립된 리보뉴클레오단백질은 전기 천공을 통해, C2c1-연관된 유전자 생성물 코팅된

입자를 이용한 충격에 의해, 화학적 트랜스펙션에 의해, 또는 세포막을 지나 운송하기 위한 일부 다른 수단에 의해 식물 세포 내로 도입될 수 있다. 예를 들어, 사전-조립된 CRISPR 리보뉴클레오단백질을 이용한 식물 원형질체의 트랜스펙션은, 식물 게놈이 표적된 변형을 보장하는 것으로 증명되었다(Woo et al. Nature Biotechnology, 2015; DOI: 10.1038/nbt.3389에 의해 기술된 바와 같음).

[1055] 특정 구현예에서, CRISPR-C2c1 시스템 성분은 입자를 이용하여 식물 세포 내로 도입된다. 단백질 또는 핵산으로서, 또는 이의 조합 중 어느 하나인 성분은 입자 위에 업로드되거나, 입자 내에 패키지되고, 식물에 적용될 수 있다(예를 들어, WO 2008042156 및 US 20130185823에 설명된 것과 같음). 특히, 본 발명의 구현예는, WO2015089419에 설명된 바와 같이, C2c1 단백질을 코딩하는 DNA 분자(들), 가이드 RNA 및/또는 분리된 가이드 RNA를 코딩하는 DNA 분자로 업로드된 또는 그로 패킹된 입자를 포함한다.

[1056] 식물 세포에 CRISPR-C2c1 시스템의 하나 이상의 성분을 도입하는 추가적인 수단은 세포 침투성 펩티드 (CPP)를 이용하는 것에 의한다. 따라서, 특히 본 발명에서 구현예는 C2c1 단백질에 연결된 세포 침투 펩티드를 포함하는 조성물을 포함한다. 본 발명의 특정 구현예에서, C2c1 단백질 및/또는 가이드 RNA는 식물 원형질체 내부로 그들을 효과적으로 수송하기 위해 하나 이상의 CPP와 커플링되며, 또한 다음의 문헌을 참조한다: Ramakrishna (20140 Genome Res. 2014 Jun; 24(6):1020-7 for Cas9 in human cells). 다른 구현예에서, C2c1 유전자 및/또는 가이드 RNA는 식물 원형질체 전달을 위한 하나 이상의 CPP에 커플링되는 하나 이상의 원형 또는 비원형 DNA 분자(들)에 의해 코딩된다. 식물 원형질체는 이후 식물 세포로, 그리고 추가로 식물에 재생된다. CPP는, 단백질로부터 유도된 또는 수용체 독립적인 방식으로 세포막을 지나 생물분자를 운송할 수 있는 키메라 서열로부터 유도된 어느 하나인, 일반적으로 35 개 미만의 아미노산의 짧은 펩티드로서 설명된다. CPP는 양이온성 펩티드, 소수성 서열을 갖는 펩티드, 양극성 펩티드, 프롤린-풍부 및 항-미생물 서열을 갖는 펩티드, 및 키메라 또는 이분 펩티드일 수 있다 (Pooga and Langel 2005). CPP는 생물학적 막에 침투할 수 있고, 이렇게 해서, 세포막을 가로질러서 세포질 내로 다양한 생체분자의 이동을 촉발시키고, 그들의 세포내 경로를 개선시켜서, 생체분자와 표적과의 상호작용을 용이하게 한다. CPP의 예는 특히 하기를 포함한다: HIV 타입 1 에 의한 바이러스 복제에 필요한 핵 전사 활성화제 단백질인 Tat, 페네트라틴, 카포시 섬유아세포 성장 인자 (FGF) 신호 펩티드 서열, 인테그린 β3 신호 펩티드 서열; 폴리아르기닌 펩티드 Args 서열, 구아닌 풍부-분자 수송체, 스위트 화살표 펩티드 등.

[1057] **유전자 변형된 비유전자이식 식물의 제조를 위한 CRISPR-C2c1 시스템의 용도**

[1058] 특정 구현예에서, 본 명세서에 기재된 방법은 식물 게놈에서 외래 DNA의 존재를 피하기 위해 CRISPR 성분을 코딩하는 것을 포함하는, 임의의 외래 유전자의 식물 게놈 내로의 영구한 도입 없이 내생성 유전자를 변형시키거나 또는 그들의 발현을 변형시키기 위해 사용된다. 이것은 비유전자이식 식물에 대한 조절 요건이 덜 엄격하기 때문에 흥미로울 수 있다.

[1059] 특정 구현예에서, 이는 C2c1 CRISPR 성분의 일시적 발현에 의해 보장된다. 특정 구현예에서 하나 이상의 CRISPR 성분은, 본 명세서에서 설명된 방법에 따른 관심 유전자의 변형을 일관되고 꾸준히 보장하기에 충분한 C2c1 단백질 및 가이드 RNA를 생산하는 하나 이상의 바이러스 벡터 상에서 발현된다.

[1060] 특정 구현예에서, C2c1 CRISPR 구성체의 일시 발현은 식물 원형질체 내에서 보장되며, 이에 따라 게놈 내로 통합되지는 않는다. 제한된 발현 창은 CRISPR-C2c1 시스템이 본 명세서에 기재된 바와 같은 표적 유전자의 변형을 보장하도록 하기에 충분할 수 있다.

[1061] 특정 구현예에서, CRISPR-C2c1 시스템의 상이한 성분은 별개로 또는 혼합물로, 본 명세서에서 상기 기재한 바와 같은 입자 또는 CPP 분자와 같은 미립자 전달 분자의 도움에 의해, 식물 세포, 원형질체 또는 식물 조직에 도입된다.

[1062] C2c1 CRISPR 성분의 발현은, C2c1 뉴클레아제 및 선택적으로 주형 DNA의 도입의 직접 활성에 의해 또는 본 명세서에서 설명된 바와 같은 CRISPR-C2c1 시스템을 이용하여 표적된 유전자의 변형 중 어느 하나에 의해 게놈의 표적된 변형을 유도할 수 있다. 본 명세서에 상기 설명된 상이한 전략은, C2c1 CRISPR 성분의 식물 게놈 내로의 도입을 필요로 하지 않으면서 C2c1-매개된 표적된 게놈 편집을 가능하게 한다. 식물 세포 내로 일시적으로 도입된 성분은 전형적으로 교배 시 제거된다.

[1063] **식물 게놈-선별 마커에서의 변형 검출**

[1064] 특정 구현예에서, 방법이 식물 게놈의 내생성 표적 유전자의 변형을 수반하는 경우, 식물, 식물 일부 또는 식물 세포가 CRISPR-C2c1 시스템으로 감염 또는 트랜스펙션된 후, 유전자 표적화 또는 표적된 돌연변이 유발이 표적

부위에서 발생했는지의 여부를 확인하는 임의의 적합한 방법이 사용된다. 방법이 이식유전자의 도입을 수반하는 경우, 형질전환된 식물 세포, 캘러스(callus), 조직 또는 식물은, 이식유전자의 존재에 대해 또는 이식유전자에 의해 코딩된 형질에 대해 조작된 식물 재료를 선택 또는 스크리닝함으로써 확인 및 분리될 수 있다. 물리적 및 생화학적 방법을, 삽입된 유전자 구성체 또는 내생성 DNA 변형을 함유하는 식물 또는 식물 세포 형질전환체를 확인하는데 사용할 수 있다. 이들 방법은 제한없이, 1) 재조합 DNA 삽입물 또는 변형된 내생성 유전자의 구조를 검출 및 결정하기 위한 써던 분석 또는 PCR 증폭; 2) 유전자 구성체의 RNA 전사물을 검출 및 조사하기 위한 노던 블롯, S1 Rnase 보호, 프라이머-연장 또는 역전하효소-PCR 증폭; 3) 효소 또는 리보자임 활성 검출을 위한 효소 어세이로서, 이러한 유전자 산물은 유전자 구성체에 의해 코딩되거나 또는 발현은 유전자 변형에 의해 영향받는 것인, 검출; 4) 유전자 구성체 또는 내생성 유전자 산물이 단백질인 경우, 웨스턴 블롯 기술, 면역침전, 또는 효소-연결 면역어세이를 포함한다. 추가의 기술, 예컨대 현장 하이브리드화, 효소 염색, 및 면역염색 또한 재조합 구성체의 발현이 존재 또는 발현을 검출 또는 특정 식물 기관 및 조직 내 내생성 유전자의 변형을 검출하는데 사용될 수 있다. 이들 모든 분석의 실시 방법은 당업자에게 공지이다.

[1065] 추가적으로(또는 대안적으로), C2c1 CRISPR 성분을 코딩하는 발현 시스템은, 초기 단계에서 및 대규모에서, CRISPR-C2c1 시스템을 함유 및/또는 그에 의해 변형된 세포를 분리 또는 효율적으로 선택하는 수단을 제공하는, 통상적으로 하나 이상의 선택가능한 또는 검출가능한 마커를 포함하도록 설계된다.

[1066] 아그로박테리움-매개된 형질전환의 경우, 마커 카세트는 측부배치되는 T-DNA 경계들에 인접하거나 그 사이에 있을 수 있으며, 바이너리 벡터 내에 함유될 수 있다. 또 다른 구현예에서, 마커 카세트는 T-DNA의 외부에 있을 수 있다. 선별 마커 카세트는 발현 카세트로서 동일한 T-DNA 경계 내에 또는 그에 인접할 수 있거나, 바이너리 벡터 상의 제2 T-DNA 내 어느 부분엔가 있을 수 있다(예를 들어, 2 T-DNA 시스템).

[1067] 유전자 총법의 경우 또는 원형질체 형질전환과 함께, 발현 시스템은 하나 이상의 분리된 선형 단편을 포함할 수 있거나, 세균 복제 요소, 세균 선별 마커 또는 다른 검출가능한 요소를 함유할 수 있는 더욱 큰 구성체의 일부일 수 있다. 가이드 및/또는 C2c1을 코딩하는 폴리뉴클레오티드를 포함하는 발현 카세트(들)은 마커 카세트에 물리적으로 연결될 수 있거나, 마커 카세트를 코딩하는 제2 핵산 분자와 함께 혼합될 수 있다. 마커 카세트는 형질전환된 세포의 효율적인 선발을 가능하게 하는 검출가능한 또는 선별 마커를 발현하는데 필수적인 요소들로 구성된다.

[1068] 선별 마커에 기초한 세포에 대한 선택 절차는 마커 유전자의 성질에 따라 달라질 것이다. 특정 구현예에서, 선별 마커, 즉 마커의 발현을 기초로 한 세포의 직접 선택을 가능하게 하는 마커가 이용된다. 선별 마커는 양성 또는 음성 선택을 부여할 수 있으며, 외부 기재의 존재에 대해 조건적 또는 비-조건적이다(Miki et al. 2004, 107(3): 193-232). 가장 흔하게는, 항생제 또는 제초제 저항성 유전자가 마커로서 사용되며, 이로 인해 선택은 마커 유전자가 그에 대한 저항성을 부여하는 항생제 또는 제초제의 억제량을 함유하는 매질 상에서 조작된 식물 재료를 성장시킴으로서 수행된다. 그러한 유전자의 예로는, 하이그로마이신(hpt) 및 카나마이신(nptII)와 같은 항생제에 대한 저항성을 부여하는 유전자, 및 포스피노트리신(bar) 및 클로로설푸론(als)과 같은 제초제에 대한 저항성을 부여하는 유전자이다.

[1069] 형질전환된 식물 및 식물 세포는 또한 가시성 마커, 통상적으로 착색된 기재 (예를 들어, β-글루쿠로니다제, 루시퍼라제, B 또는 C1 유전자)를 가공할 수 있는 효소의 활성을 스크리닝함으로써 확인될 수 있다. 그러한 선택 및 스크리닝 방법은 당업자에게 공지이다.

[1070] **식물 배양 및 재생**

[1071] 특정 구현예에서, 변형된 게놈을 갖고 본 명세서에 기재된 임의의 방법에 의해 생성되거나 또는 얻어지는 식물 세포는 형질전환되거나 또는 변형된 유전자형을 갖는 전체 식물 및 그에 따라 목적하는 표현형을 재생하기 위해 배양될 수 있다. 통상적인 재생 기법은 당업자에게 잘 공지되어 있다. 이러한 재생 기법의 특정 예는 조직 배양 성장 배지에서 소정의 식물 호르몬의 조작에 의존하며, 전형적으로 목적하는 뉴클레오티드 서열과 함께 도입된 항생제 및/또는 제초제 마커에 의존한다. 추가의 특정 구현예에서, 식물 재생은 배양된 원형질체, 식물 캘러스, 외식편(explant), 기관, 화분, 배아, 또는 그의 일부로부터 수득된다 (참조: 예를 들어, (Evans et al. (1983), Handbook of Plant Cell Culture, Klee et al (1987) Ann. Rev. of Plant Phys.).

[1072] 특정 구현예에서, 본 명세서에 기재된 형질전환되거나 또는 개선된 식물은 본 발명의 동형접합적 개선 식물(DNA 변형에 대해 동형접합적)에 대한 종자를 제공하도록 자가수분되거나 또는 이형접합적 식물에 대한 종자를

제공하도록 비유전자이식 식물 또는 상이한 개선 식물과 교배될 수 있다. 재조합 DNA가 식물 세포 내로 도입되는 경우에, 이러한 교배로 얻어진 식물은 재조합 DNA 분자에 이형접합적인 식물이다. 개선된 식물로부터의 교배 그리고 유전자 변형의 포함 (재조합 DNA일 수 있음)에 의해 얻어지는 이러한 동형접합적 식물과 이형접합적 식물은 둘 모두 본 명세서에서 "자손" 로서으로서 지칭된다. 자손 식물은 본래의 유전자이식 식물로부터의 자손이 되고 본 명세서에 제공된 방법에 의해 도입되는 게놈 변형 또는 재조합 DNA 분자를 함유하는 식물이다. 대안적으로, 유전적으로 변형된 식물은 Cpf1 효소를 이용하여 앞서 설명된 방법들 중 하나에 의해 수득될 수 있으며, 이에 의해 외래의 DNA가 게놈 내로 통합되지 않는다. 추가적인 육종에 의해 얻어지는 이러한 식물의 자손은 또한 유전자 변형을 함유할 수 있다.. 육종은 상이한 작물에 대해 통상적으로 사용되는 임의의 육종 방법에 의해 수행된다 (예를 들어, Allard, Principles of Plant Breeding, John Wiley & Sons, NY, U. of CA, Davis, CA, 5098 (1960)).

[1073] **증강된 작물학적 형질을 갖는 식물의 생성**

[1074] 본 명세서에 제공된 C2c1 기반 CRISPR 시스템은, 표적된 이중-가닥 또는 단일-가닥 파손을 도입 및/또는 유전자 활성화제 및/또는 억제제 시스템을 도입하는데 사용될 수 있고, 제한없이 유전자 표적화, 유전자 대체, 표적된 돌연변이유발, 표적된 결실 또는 삽입, 표적된 역전 및/또는 표적된 전위에 사용될 수 있다. 단일 세포에서 다중 변형을 달성하도록 지시된 다중 표적화 RNA의 공동발현에 의해, 다중복합 게놈 변형이 보장될 수 있다. 이 기술은 증강된 영양 품질, 질환에 대해 증가된 내성 및 항생제 및 비항생제 스트레스에 대한 내성, 및 상업적으로 가치있는 식물 제품 또는 이종성 화합물의 증가된 생산을 비롯한 개선된 형질을 갖는 식물의 높은 정확도 조작에 사용될 수 있다.

[1075] 특정 구현예에서, 본 명세서에서 설명된 바와 같은 CRISPR-C2c1 시스템은 내생성 DNA 서열에서 표적된 이중-가닥 파손(DSB)을 도입하는데 사용된다. DSB는 세포 DNA 복구 경로를 활성화하고, 이는 파손 부위 가까이에서 바람직한 DNA 서열을 달성하도록 매일 수 있다. 내생성 유전자의 비활성화가 바람직한 형질을 부여 또는 그에 기여할 수 있는 경우 이는 유익하다. 특정 구현예에서, 주형 서열을 갖는 상동성 재조합은 관심 유전자를 도입하기 위하여, DSB의 부위에서 촉진된다. 일부 구현예에서, HR-독립적 재조합은 스태거드 DSB에서 관심 서열 또는 유전자를 도입시키기 위해서 DSB의 부위에서 촉진된다. 특정 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 스태거드 DSB를 생성시킨다. 일정한 특정 구현예에서, CRISPR-C2c1 시스템은 가이드 서열에 삽입 주형 서열을 포함하고 스태거드 DSB에 특이적 DNA 삽입을 도입시킨다.

[1076] 특정 구현예에서, CRISPR-C2c1 시스템은, 내생성 식물 유전자의 활성화 및/또는 억제를 위한 작용 도메인에 융합 또는 그에 작동적으로 연결되는, 일반적인 핵산 결합 단백질로서 사용될 수 있다. 예시적인 기능성 도메인은 제한없이 RNA 또는 DNA 디아미나제, 번역 개시인자, 번역 활성인자, 번역 억제인자, 뉴클레아제, 특히 리보뉴클레아제, 스플라이시오솜, 비드, 광 유도성/제어성 도메인 또는 화학 유도성/제어성 도메인을 포함할 수 있다. 통상적으로 이들 구현예에서, C2c1 단백질은 적어도 하나의 돌연변이를 포함하여, 적어도 하나의 돌연변이를 갖지 않는 C2c1 단백질의 5% 이하의 활성을 갖게 하고, 가이드 RNA는 표적 서열에 하이브리드화할 수 있는 가이드 서열을 포함한다.

[1077] 본 명세서에 기재된 방법은 일반적으로 그들이 야생형 식물에 비교되는 하나 이상의 바람직한 형질을 갖는다는 점에서 "개선된 식물" 의 생성을 초래한다. 특정 구현예에서, 얻어진 식물, 식물 세포 또는 식물 부분은 식물 세포의 모두 또는 일부에 혼입된 외생성 DNA 서열을 포함하는 유전자이식 식물이다. 특정 구현예에서, 비-유전자이식 유전자 변형 식물, 식물 일부 또는 세포가 얻어지며, 즉, 외생성 DNA 서열은 식물의 식물 세포 중 어느 것의 게놈에 혼입된다. 이러한 구현예에서, 개선된 식물은 비-유전자이식이다. 내생성 유전자의 변형만이 보장되고, 외래 유전자는 식물 게놈에서 도입 또는 유지되지 않는 경우에, 얻어진 유전자 변형된 작물은 외래 유전자를 함유하지 않으며, 따라서 기본적으로 비유전자이식으로 간주될 수 있다. 식물 게놈 편집을 위한 CRISPR-C2c1 시스템의 상이한 응용이 하기에서 더욱 상세하게 설명된다:

[1078] **농업적 관심 형질을 부여하는 하나 이상의 외래 유전자의 도입**

[1079] 본 발명은 게놈 편집 또는 관심의 표적 유전자좌에서 또는 그에 연관된 서열을 변형하는 방법을 제공하며, 여기서 방법은 C2c1 이펙터 단백질 복합체를 식물 세포 내로 도입하는 단계를 포함하며, 이로 인해 C2c1 이펙터 단백질 복합체는 효과적으로 작용하여, 예를 들어 외래 관심 유전자를 코딩하는, DNA 삽입물을 식물 세포의 게놈 내로 통합시킨다. 일부 구현예에서, DNA 삽입물의 통합은 외생적으로 도입된 DNA 주형 또는 복구 주형을 갖는 HR에 의해 촉진된다. 일부 바람직한 구현예에서, DNA 삽입물의 통합은 HR-독립적 통합 (예를 들어, NHEJ)에 의해 촉진된다. 통상적으로 외생적으로 도입된 DNA 주형 또는 복구 주형은, C2c1 이펙터 단백질

복합체 또는 일 성분 또는 복합체의 성분의 발현을 위한 폴리뉴클레오티드 벡터와 함께 전달된다.

[1080] 본 명세서에서 제공된 CRISPR-C2c1 시스템은 표적화된 유전자 전달을 가능하게 한다. 관심 유전자 발현의 효율은 게놈 내로의 통합의 위치에 의해 큰 정도로 결정된다는 것이 더욱 명백하게 되었다. 본 방법은 외래 유전자의 게놈 내 바람직한 위치로의 표적된 통합을 가능하게 한다. 위치는 이전에 생성된 이벤트의 정보에 기초하여 선택될 수 있거나, 본 명세서에 어느 부분에서 개시된 방법에 의해 선택될 수 있다.

[1081] 특정 구현예에서, 본 명세서에 제공된 방법은 (a) 직접 반복부 및 가이드 서열을 포함하는 가이드 RNA를 포함하는 C2c1 CRISPR 복합체를 세포에 도입시키는 단계로서, 가이드 서열은 식물 세포에 내생성인 표적 서열에 하이브리드화하는 것인 단계; (b) 가이드 서열이 표적 서열과 하이브리드화하고 가이드 서열이 표적화되는 서열 또는 그 근처에 이중 가닥 파손을 유도할 때 가이드 RNA와 복합체 형성하는 C2c1 이펙터 분자를 식물 세포에 도입시키는 단계; 및 (c) 관심 유전자를 코딩하고 HDR의 결과로서 DS 파손의 위치로 도입되는 HDR 복구 주형을 코딩하는 뉴클레오티드 서열을 세포에 도입시키는 단계를 포함한다. 특정 구현예에서, 도입 단계는 C2c1 이펙터 단백질, 가이드 RNA 및 복구 주형을 코딩하는 하나 이상의 폴리뉴클레오티드를 식물 세포로 전달하는 것을 포함할 수 있다. 특정 구현예에서, 폴리뉴클레오티드는 DNA 바이러스 (예를 들어, 제미니바이러스) 또는 RNA 바이러스 (예를 들어, 토브라바이러스)에 의해 세포 내로 전달된다. 특정 구현예에서, 도입 단계는 복구 주형, 가이드 RNA 및 C2c1 이펙터 단백질을 코딩하는 하나 이상의 폴리뉴클레오티드 서열을 함유하는 T-DNA를 식물 세포로 전달하는 것을 포함하며, 여기서 전달은 아그로박테리움을 통한 것이다. C2c1 이펙터 단백질을 코딩하는 핵산 서열은 프로모터, 예컨대 구조적 프로모터(예를 들어, 콜리플라워 모자이크 바이러스 35S 프로모터) 또는 세포 특이적 또는 유도성 프로모터에 작동적으로 연결될 수 있다. 특정 구현예에서, 폴리뉴클레오티드는 미세투사 유전자 총에 의해 도입된다. 특정 구현예에서, 본 방법은 도입 단계 후, 본 방법은 복구 주형, 즉 관심 유전자가 도입되었는지의 여부를 결정하기 위하여, 식물 세포를 스크리닝하는 것을 추가로 포함한다. 특정 구현예에서, 상기 방법은 식물 세포로부터 식물을 재생시키는 단계를 포함한다. 추가 구현예에서, 상기 방법은 유전적으로 요망되는 식물 계통을 얻기 위해 식물을 교배육종시키는 단계를 포함한다. 관심 형질을 코딩하는 외래 유전자의 예가 하기 열거된다.

[1082] **농업적 관심 형질 부여하는 내생성 유전자의 편집**

[1083] 본 발명은 게놈 편집 또는 관심의 표적 유전자좌에서 또는 그에 연관된 서열을 변형하는 방법을 제공하며, 여기서 방법은 C2c1 이펙터 단백질 복합체를 식물 세포 내로 도입하는 단계를 포함하며, 이로 인해 C2c1 복합체는 식물의 내생성 유전자의 발현을 변형시킨다. 이는 상이한 방식으로 전달될 수 있으며, 특정 구현예에서, 내생성 유전자의 발현의 제거가 바람직하고, 유전자 발현을 변형시키기 위하여 C2c1 CRISPR 복합체는 내생성 유전자를 표적화하고 절단하는데 사용된다. 이들 구현예에서, 본 명세서에서 제공되는 방법은 (a) 직접 반복부 및 가이드 서열을 포함하는 가이드 RNA를 포함하는 C2c1 CRISPR 복합체를 세포에 도입시키는 단계로서, 가이드 서열은 식물 세포의 게놈의 관심 유전자 내 표적 서열에 하이브리드화하는 것인 단계; (b) 표적 서열에 하이브리드화하는 가이드 서열을 포함하는 가이드 RNA에 결합 시, 가이드 서열이 표적화되는 서열 또는 그 근처에 이중 가닥 파손을 보장하는 C2c1 이펙터 단백질을 세포에 도입시키는 단계를 포함하고; 특정 구현예에서, 도입 단계는 C2c1 이펙터 단백질 및 가이드 RNA를 코딩하는 하나 이상의 폴리뉴클레오티드를 식물 세포에 전달하는 단계를 포함한다.

[1084] 특정 구현예에서, 폴리뉴클레오티드는 DNA 바이러스 (예를 들어, 제미니바이러스) 또는 RNA 바이러스 (예를 들어, 토브라바이러스)에 의해 세포 내로 전달된다. 특정 구현예에서, 도입 단계는 가이드 RNA 및 C2c1 이펙터 단백질을 코딩하는 하나 이상의 폴리뉴클레오티드 서열을 함유하는 T-DNA를 식물 세포로 전달하는 것을 포함하며, 여기서 전달은 아그로박테리움을 통한 것이다. CRISPR-C2c1 시스템의 성분을 코딩하는 폴리뉴클레오티드 서열은 프로모터, 예컨대, 항상성 프로모터 (예를 들어, 콜리플라워 모자이크 바이러스 35S 프로모터), 또는 세포 특이적 또는 유도성 프로모터에 작동적으로 연결될 수 있다. 특정 구현예에서, 폴리뉴클레오티드는 미세투사 유전자 총에 의해 도입된다. 특정 구현예에서, 상기 방법은 관심 대상 유전자의 발현이 변형되었는지의 여부를 결정하기 위한 단계들을 도입한 후에 식물 세포를 스크리닝하는 단계를 추가로 포함한다. 특정 구현예에서, 상기 방법은 식물 세포로부터 식물을 재생시키는 단계를 포함한다. 추가 구현예에서, 상기 방법은 유전적으로 요망되는 식물 계통을 얻기 위해 식물을 교배육종시키는 단계를 포함한다.

[1085] 상기 기재된 방법의 특정 구현예에서, 질환 감수성 유전자 또는 식물 방어 유전자의 조절자를 코딩하는 유전자(예를 들어, Mlo 유전자)의 표적화된 돌연변이에 의해 질환 내성 작물이 얻어진다. 특정 구현예에서, 아세토락테이트 신타제(ALS) 및 프로토포르피리노겐 옥시다제(PPO)를 코딩하는 것과 같은 식물 유전자에서 특정 뉴

클레오티드의 표적화된 치환에 의해 제초제-내성 작물이 생성된다. 특정 구현예에서, 항생제 스트레스 내성의 음성 조절자를 코딩하는 유전자의 표적화된 돌연변이에 의한 가뭄 및 염 내성 작물, Waxy 유전자의 표적화된 돌연변이에 의한 저 아밀로스 곡물, 벼 또는 호분층에서의 주요 리파제 유전자의 표적화된 돌연변이에 의해 산패가 감소된 다른 곡물 등. 특정 구현예에서. 관심 특성을 코딩하는 내생성 유전자의 보다 광범위한 목록이 하기에 열거된다.

[1086] **농업적 관심 특성을 부여하는 CRISPR-C2c1 시스템에 의한 내생성 유전자의 조절**

[1087] 본 명세서는 또한, 본 명세서에서 제공된 C2c1 단백질을 이용하여 내생성 유전자 발현을 조절(즉, 활성화 또는 억제)하는 방법을 제공한다. 그러한 방법은 C2c1 복합체에 의해 식물 게놈으로 표적되는 구별된 RNA 서열(들)을 이용한다. 더욱 특히, 별개 RNA 서열(들)은 둘 이상의 어댑터 단백질(예를 들어, 압타머)에 결합하고, 이에 의해 각각의 어댑터 단백질은 하나 이상의 작용 도메인과 회합되고, 어댑터 단백질과 회합된 하나 이상의 기능성 도메인 중 적어도 하나는 디아미나제 활성, 메틸라제 활성, 디메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, DNA 통합 활성 RNA 절단 활성, DNA 절단 활성 또는 핵산 결합 활성을 포함하는 하나 이상의 활성을 가지고; 기능성 도메인은 바람직한 특성을 수득하도록 내생성 식물 유전자의 발현을 조절하는데 사용된다. 통상적으로, 이들 구현예에서, C2c1 이펙터 단백질은 하나 이상의 돌연변이를 가져서, 하나 이상의 돌연변이를 갖지 않는 C2c1 이펙터 단백질의 뉴클레아제 활성의 5% 미만을 갖는다.

[1088] 특정 구현예에서, 본 명세서에 제공된 방법은 (a) tracr RNA, 직접 반복부 및 가이드 서열을 포함하는 가이드 RNA를 포함하는 C2c1 CRISPR 복합체를 세포에 도입시키는 단계로서, 가이드 서열은 식물 세포에 내생성인 표적 서열에 하이브리드화하는 것인 단계; (b) 가이드 서열이 표적 서열과 하이브리드화할 때 가이드 RNA와 복합체를 형성하는 C2c1 이펙터 분자를 식물 세포에 도입시키는 단계를 포함하고, 가이드 RNA는 기능성 도메인에 결합하는 별개 RNA 서열 (압타머)를 포함하도록 변형되고/되거나 C2c1 이펙터 단백질은 기능성 도메인에 연결되게 변형된다. 특정 구현예에서, 도입 단계는 (변형된) 가이드 RNA 및 (변형된) C2c1 이펙터 단백질을 코딩하는 하나 이상의 폴리뉴클레오티드를 식물 세포로 전달하는 것을 포함할 수 있다. 이들 방법에서의 이용을 위한 CRISPR-C2c1 시스템의 성분의 상세한 내용은 본 명세서 어느 부분에서 설명되어 있다.

[1089] 특정 구현예에서, 폴리뉴클레오티드는 DNA 바이러스 (예를 들어, 제미니바이러스) 또는 RNA 바이러스 (예를 들어, 토브라바이러스)에 의해 세포 내로 전달된다. 특정 구현예에서, 도입 단계는 가이드 RNA 및 C2c1 이펙터 단백질을 코딩하는 하나 이상의 폴리뉴클레오티드 서열을 함유하는 T-DNA를 식물 세포로 전달하는 것을 포함하며, 여기서 전달은 아그로박테리움을 통한 것이다. CRISPR-C2c1 시스템의 하나 이상의 성분을 코딩하는 핵산 서열은 프로모터, 예컨대 구조적 프로모터(예를 들어, 콜리플라워 모자이크 바이러스 35S 프로모터) 또는 세포 특이적 또는 유도성 프로모터에 작동적으로 연결될 수 있다. 특정 구현예에서, 폴리뉴클레오티드는 미세투사 유전자 총에 의해 도입된다. 특정 구현예에서, 상기 방법은 관심 대상 유전자의 발현이 변형되었는지의 여부를 결정하기 위한 단계들을 도입한 후에 식물 세포를 스크리닝하는 단계를 추가로 포함한다. 특정 구현예에서, 상기 방법은 식물 세포로부터 식물을 재생시키는 단계를 포함한다. 추가 구현예에서, 상기 방법은 유전적으로 요망되는 식물 계통을 얻기 위해 식물을 교배육종시키는 단계를 포함한다. 관심 특성을 코딩하는 내생성 유전자의 보다 광범위한 목록이 하기에 열거된다.

[1090] **배수체 식물을 변형하기 위한 C2c1의 용도**

[1091] 다수의 식물은 배수체인데, 이는 그들이 게놈 복제물의 2배 (때때로 밀에서와 같이 6배만큼)를 운반한다는 것을 의미한다. C2c1 CRISPR 이펙터 단백질을 이용하는, 본 발명에 따른 방법은 "다중화된" 것이어서 유전자의 모든 카피에 영향을 미치거나 수십 개의 유전자를 한번에 표적할 수 있다. 예를 들어, 특정 구현예에서, 본 발명의 방법은 질환에 대해 방어를 억제하는 것을 초래하는 상이한 유전자에서 기능 상실 돌연변이를 동시에 보장하기 위해 사용된다. 특정 구현예에서, 본 발명의 방법은 밀 식물이 백분병균에 내성이라는 것을 보장하기 위해 밀 식물 세포에서 TaMLO-A1, TaMLO-B1 및 TaMLO-D1 핵산 서열의 발현을 동시에 억제하고 이로부터 밀 식물을 재생하는 데 사용된다 (또한 WO2015109752 참조).

[1092] **작물학적 형질을 부여하는 예시적인 유전자**

[1093] 본 명세서에서 상기 설명된 것과 같이 특정 구현예에서, 본 발명은, 하나 이상의 식물 발현성 유전자(들)을 포함하는 관심 DNA의 삽입을 위해 본 명세서에서 설명된 것과 같은 CRISPR-C2c1 시스템의 이용을 포괄한다. 추가의 특정 구현예에서, 본 발명은 본 명세서에서 설명된 바와 같은 C2c1 시스템을 하나 이상의 식물 발현된

유전자(들)의 부분적 또는 완전한 결실을 위해 사용하는 도구 및 방법을 포함한다. 다른 추가의 구현예에서, 본 발명은 하나 이상의 뉴클레오티드의 돌연변이, 치환, 삽입에 의해 하나 이상의 식물 발현된 유전자들의 변형을 보장하기 위하여 본 명세서에서 설명된 바와 같은 C2c1 시스템을 사용하는 도구 및 방법을 포함한다. 다른 특정 구현예에서, 본 발명은 식물 발현된 유전자의 발현을 유도하는 하나 이상의 조절 엘리먼트의 특정 변형에 의해 하나 이상의 상기 유전자의 발현의 변형을 보장하기 위하여 본 명세서에서 설명된 것과 같은 CRISPR-C2c1 시스템의 이용을 포함한다.

[1094] 특정 구현예에서, 본 발명은 외생성 유전자 및/또는 내생성 유전자의 표적화 및 예컨대 하기 열거되는 이들의 조절 엘리먼트의 도입을 수반하는 방법을 포함한다:

[1095] ***1. 해충 또는 질환에 내성을 부여하는 유전자:***

[1096] 식물 질환 내성 유전자. 식물은 특정 병원성 균주에 대해 내성이 있는 식물을 조작하기 위해 클로닝된 내성 유전자로 형질전환될 수 있다. 예를 들어, 문헌 [Jones et al., Science 266:789 (1994)](클라도스포륨 풀붐(Cladosporium fulvum)에 대한 내성을 위한 토마토 Cf-9 유전자의 클로닝); 문헌 [Martin et al., Science 262:1432 (1993)](슈도모나스 시링가에 (Pseudomonas syringae) pv에 대한 내성을 위한 토마토 Pto 유전자. 토마토는 단백질 키나제를 코딩함); 문헌 [Mindrinos et al., Cell 78:1089 (1994)](아라비돕스는 슈도모나스 시링가에 대한 내성을 위한 RSP2 유전자일 수 있음). 병원균 감염 동안 상향 또는 하향 조절되는 식물 유전자는 병원균 내성을 위해 조작될 수 있다. 참조: 예를 들어, Thomazella et al., bioRxiv 064824; doi: doi.org/10.1101/064824 Epub. July 23, 2016 (병원체 감염 동안 정상적으로 상향조절되는 SIDMR6-1에 결실을 갖는 토마토 식물).

[1097] 해충, 예컨대, 대두 시스트 선충에 대한 내성을 부여하는 유전자. 예를 들어, 국제 특허 출원 WO 96/30517; 국제 특허 출원 WO 93/19181 참조.

[1098] 바실러스 투링기엔시스 (Bacillus thuringiensis) 단백질, 예를 들어, 문헌 [Geiser et al., Gene 48:109 (1986)]을 참조한다.

[1099] 렉틴, 예를 들어, Van Damme et al., Plant Molec. Biol. 24:25 (1994. 참조.

[1100] 비타민-결합 단백질, 예컨대, 아비딘, 국제 특허 출원 US93/06487을 참조하고, 해충에 대한 살유충제로서 아비딘 및 아비딘 상동체의 용도를 교시한다.

[1101] 효소 억제제, 예컨대, 프로테아제 또는 프로테이나제 억제제 또는 아밀라제 억제제. 참조, 예를 들어 Abe et al., J. Biol. Chem. 262:16793 (1987), Huub et al., Plant Molec. Biol. 21:985 (1993)), Sumitani et al., Biosci. Biotech. Biochem. 57:1243 (1993) 및 US 특허 번호 5,494,813.

[1102] 곤충-특이적 호르몬 또는 페로몬, 예컨대, 에크디스테로이드 또는 유약 호르몬, 이의 변이체, 이에 기반한 모방체 또는 이들의 길항제 또는 효현제. 예를 들어, 문헌[Hammock et al., Nature 344:458 (1990)] 참조.

[1103] 발현 시 영향받는 해충의 생리를 붕괴시키는 곤충-특이적 펩티드 또는 신경펩티드. 예를 들어, Regan, J. Biol. Chem. 269:9 (1994) 및 Pratt et al., Biochem. Biophys. Res. Comm. 163:1243 (1989). 또한, 미국 특허 번호 5,266,317을 참조한다.

[1104] 뱀, 말벌 또는 임의의 다른 유기체에 의해 자연에서 생성되는 곤충-특이적 독. 예를 들어, [Pang et al., Gene 116: 165 (1992)]를 참조한다.

[1105] 모노터펜, 세스퀴터펜, 스테로이드, 하이드록삼산, 페닐프로파노이드 유도체 또는 살충 활성을 갖는 다른 비단백질 분자의 과축적을 초래하는 효소.

[1106] 생물학적 활성 분자; 예를 들어, 천연 또는 합성이건, 해당 효소, 단백질 분해 효소, 지방분해 효소, 뉴클레아제, 사이클라제, 트랜스아미나제, 에스터라제, 가수분해효소, 포스파타제, 키나제, 포스포릴라제, 중합효소, 엘라스타제, 키티나제 및 글루카나제의 번역후 변형을 포함하는, 변형에 관여되는 효소. 국제 특허 출원 WO93/02197, 문헌[Kramer et al., Insect Biochem.. Biol. 23:691(1993)] 및 [Kawalleck et al., Plant Molec. Biol. 21 :673 (1993)]을 참조한다.

[1107] 신호 전달을 자극하는 분자. 예를 들어, 문헌 [Lamb et al.,. Biol.:757 (1994)] 및 [Toubart et al., Plant J.:1467 (1994)]을 참조한다.

[1108] 바이러스-침입성 단백질 또는 그로부터 유도된 복합 독소. 문헌[Beachy et al., Ann. rev. Phytopathol. 28:451 (1990)을 참조한다.

[1109] 병원균 또는 기생충에 의해 자연에서 생성된 발생-저지 단백질. 문헌 [Lamb et al., Bio/Technology 10:1436 (1992)] 및 [Toubart et al., Plant J. 2:367 (1992)]을 참조한다.

[1110] 식물에 의해 자연계에서 생성되는 발생-저지 단백질. 예를 들어, 문헌 [Logemann et al., Bio/Technology 10:305 (1992)]을 참조한다.

[1111] 식물에서, 병원균은 종종 숙주-특이적이다. 예를 들어, 일부 푸사리움 종은 토마토 시들음을 야기할 것이지만, 토마토만을 공격하고, 다른 푸사리움 종은 밀만을 공격한다. 식물은 대부분의 병원균에 저항하는 기존의 유도된 방어성을 갖는다. 식물 생성에 걸친 돌연변이 및 재조합 사건은 감수성을 일으키는 유전자 가변성을 야기하는데, 특히 병원균이 식물보다 더 빈번하게 재생되기 때문이다. 식물에서, 비숙주 내성일 수 있고, 예를 들어, 숙주 및 병원균은 비적합하거나 또는 전형적으로 다수의 유전자에 의해 제어되는 모든 품종의 병원균에 대해 부분적 내성일 수 있고/있거나 일부 품종의 병원균에 대해서는 완전한 내성이지만 다른 품종에 대해서는 그렇지 않을 수 있다. 이러한 내성은 전형적으로 소수의 유전자에 의해 제어된다. CRISP-C2c1 시스템의 성분 및 방법을 이용하여, 새로운 도구가 이제 존재하여 여기서 예측되는 특정 돌연변이를 유도한다.

따라서, 내성 유전자의 공급원 게놈을 분석할 수 있고, 목적하는 특징 또는 형질을 갖는 식물에서, 내성 유전자 상승을 유도하기 위해 CRISPR-C2c1 시스템의 방법 및 성분을 사용할 수 있다. 본 시스템은 이전의 돌연변이유발제보다 더 정확성을 가질 수 있고, 따라서 식물 육종 프로그램을 가속화시키고 개선시킨다.

[1112] ***2. 식물 질환 관여 유전자***

[1113] 식물 질환 관여 유전자 식물은 특이적 병원체 균주에 내성인 식물을 조작하기 위해 질환 감수성 또는 관련 유전자를 변형시키는 CRISPR-C2c1 시스템으로 형질전환될 수 있다. 예를 들어, 추정 당 수송체를 코딩하는 식물 SWEET 유전자는 벼-병원성 잔토모나스 오리자에 (Xanthomonas oryzae) 유래의 TAL 이펙터에 의해 유도되어서, 병원체 감염의 증강된 확산을 초래하는 것으로 알려져 있다. 참조: Streubel et al, New Phytologist, 2013 Nov;200(3):808-19. doi: 10.1111/nph.12411. Epub 2013 Jul 24. 시트러스의 CsLOB는 시트러스 질환, 예컨대 시트러스 동고병에 관여하는 TAL 이펙터에 의해 유도되는 것으로 알려져 있다.

[1114] 본 발명은 또한 임의의 본 명세서에 기술된 조작된 CRISPR 효소 (예를 들어, 조작된 Cas 이펙터 모듈), 조성물, 또는 임의의 본 명세서에 기술된 시스템 또는 벡터 시스템과 식물 세포를 접촉시키는 단계를 포함하는 세포에서 관심 유전자좌를 변형시키는 방법을 제공하거나, 또는 세포는 세포 내에 존재하는 임의의 본 명세서에 기술된 CRISPR 복합체를 포함한다. 일정 구현예에서, 식물 세포는 A/T 풍부 게놈을 포함할 수 있다. 일부 구현예에서, 세포 게놈은 T-풍부 PAM을 포함한다. 특정 구현예에서, PAM은 5'-TTN-3' 또는 5'-ATTN-3' 이다.

특정 구현예에서, 변형된 유전자좌는 식물 질환과 관련된다. 특정 구현예에서, 식물 질환은 병원체 감수성과 관련된다. 특정 구현예에서, 변형된 유전자좌는 SWEET 유전자좌 또는 CsLOB 유전자좌를 포함한다.

특정 구현예에서, 식물 질환은 시트러스 동고병 또는 벼 병충해 질환이다. 일부 구현예에서, 세포 게놈은 T-풍부 PAM을 포함한다. 특정 구현예에서, PAM은 5'-TTN-3' 또는 5'-ATTN-3' 이다. 일부 실시형태에서, CRISPR-Cas 시스템은 CRISPR- C2c1 시스템이다. 일부 구현예에서, 관심 유전자좌는 5' 오버행을 갖는 스태거드 절단부를 도입시켜 C2c1 이펙터 단백질에 의해 변형된다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, 관심 유전자좌는 단일 뉴클레오티드 결실 또는 돌연변이에 의해 변형된다.

일부 구현예에서, 관심 유전자좌는 50 nt 미만의 결실 또는 돌연변이에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 HDR로 CRISPR-C2c1 시스템을 도입하여 5' 오버행을 갖는 스태거드 절단부에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ로 CRISPR-C2c1 시스템을 도입하여 5' 오버행을 갖는 스태거드 절단부에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 PAM의 원위 말단에서 CRISPR-C2c1 시스템을 도입한 후, HDR에 의해 복구하여, 5' 오버행을 갖는 스태거드 절단부에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 HDR로 5' 오버행에서 CRISPR-C2c1 시스템을 도입하여 외생성 Dna 서열의 삽입에 의해 변형된다. 바람직한 구현예에서, 관심 유전자좌는 HDR로 5' 오버행에서 CRISPR-C2c1 시스템을 도입하여 외생성 Dna 서열의 삽입에 의해 변형된다.

[1115] 식물 질환에 관여하는 유전자, 예컨대 WO 2013046247에 열거된 것들:

[1116] 벼 병해: 마그나포르테 그리세아 (Magnaporthe grisea), 코클리오볼루스 미야베아누스 (Cochliobolus miyabeanus), 리족토니아 솔라니 (Rhizoctonia solani) 및 지베렐라 푸지쿠로이 (Gibberella fujikuroi); 밀의

병해: 에리시페 그라미니스 (Erysiphe graminis), 푸사리움 그라미네아룸 (Fusarium graminearum), 푸사리움 아베나세움 (F. avenaceum), 푸사리움 쿨모룸 (F. culmorum), 마이크로도키움 니발레 (Microdochium nivale), 푸치니아 스트리이포르미스 (Puccinia striiformis), 푸치니아 그라미니스 (P. graminis), 푸치니아 레콘디타 (P. recondita), 마이크로넥트리엘라 니발레 (Micronectriella nivale), 티풀라 종 (Typhula sp.), 우스틸라고 트리티시 (Ustilago tritici), 틸레티아 카리에스 (Tilletia caries), 슈도세르코스포렐라 헤르포트리코이데스 (Pseudocercosporella herpotrichoides), 미코스파에렐라 그라미니콜라 (Mycosphaerella graminicola), 스타고노스포라 노도룸 (Stagonospora nodorum), 피레노포라 트리티 시레펜티스 (Pyrenophora tritici-repentis); 보리의 병해: 에리시페 그라미니스 (Erysiphe graminis), 푸사리움 그라미네아룸 (Fusarium graminearum), 푸사리움 아베나세움 (F. avenaceum), 푸사리움 쿨모룸 (F. culmorum), 마이크로도키움 니발레 (Microdochium nivale), 푸치니아 스트리이포르미스 (Puccinia striiformis), 피. 그라미니스 (P. graminis), 피. 호르데이 (P. hordei), 우스틸라고 누다 (Ustilago nuda), 린코스포륨 세칼리스(Rhynchosporium secalis), 피레노포라 테레스(Pyrenophora teres), 코킬로볼루스 사비투스 (Cochliobolus sativus), 피레노포라 그라라미네아 (Pyrenophora graminea), 및 리족토니아 솔라니 (Rhizoctonia solani); 메이즈 병해: 우스틸라고 메이디스 (Ustilago maydis), 코킬로볼루스 헤테로스트로푸스 (Cochliobolus heterostrophus), 글로에오세르코스포라 소르기 (Gloeocercospora sorghi), 푸치니아 폴리소라 (Puccinia polysora), 세르코스포라 자에-메이디스 (Cercospora zeae-maydis), 리족토니아 솔라니 (Rhizoctonia solani);

[1117] 감귤 병해: 디아포르테 시트리 (Diaporthe citri), 엘시노에 파우세티 (Elsinoe fawcetti), 페니실리움 디지타툼 (Penicillium digitatum), 피. 이탈리쿰 (P. italicum); 피토프토라 파라시티카 (Phytophthora parasitica), 피토프토라 시트로프토라 (Phytophthora citrophthora); 사과 병해: 모닐리니아 말리 (Monilinia mali), 발사 세라토스페르마 (Valsa ceratosperma), 포도스파에라 류코트리차 (Podosphaera leucotricha), 알테르나리아 알테르나타 사과 병원형 (Alternaria alternata apple pathotype), 벤투리아 이나에쿠 알리스 (Venturia inaequalis), 콜레토트리쿰 아쿠타툼 (Colletotrichum acutatum) 및 피토프토라 칵토룸 (Phytophtora cactorum);

[1118] 배 병해: 벤투리아 나시콜라 (Venturia nashicola), 벤투리아 피리나 (V. pirina), 알테르나리아 알테르나타 일본 배 병원형 (Alternaria alternata Japanese pear pathotype), 짐노스포란지움 하라에아눔 (Gymnosporangium haraeanum) 및 피토프토라 칵토룸 (Phytophtora cactorum);

[1119] 복숭아 병해: 모닐리니아 프룩티콜라 (Monilinia fructicola), 클라도스포리움 카르포 필룸 (Cladosporium carpophilum) 및 포몹시스 종 (Phomopsis sp.);

[1120] 포도 병해: 엘시노에 암펠리나 (Elsinoe ampelina), 글로메렐라 신구라타 (Glomerella cingulata), 운시눌라 네카토르 (Uncinula necator), 파콥소라 암펠롭시디스 (Phakopsora ampelopsidis), 구이그나르디아 비드웰리이 (Guignardia bidwellii), 및 플라스모파라 비티콜라 (Plasmopara viticola);

[1121] 감 병해: 글로에오스포리움 카키 (Gloeosporium kaki) 및 세르코스포라 카키 (Cercospora kaki), 미코스파에렐라 나웨 (Mycosphaerella nawae);

[1122] 박 병해: 콜레토트리쿰 라제나리움 (Colletotrichum lagenarium), 스파에로테카 풀리지네아 (Sphaerotheca fuliginea), 미코스파에렐라 멜로니스 (Mycosphaerella melonis), 푸사리움 옥시스포룸 (Fusarium oxysporum), 슈도페로노스포라 쿠벤시스 (Pseudoperonospora cubensis), 피토프토라 종 (Phytophthora sp.) 및 피티움 종 (Pythium sp.);

[1123] 토마토 병해: 알테르나리아 솔라니 (Alternaria solani), 클라도스포리움 풀붐 (Cladosporium fulvum) 및 피토프토라 인페스탄스 (Phytophthora infestans); 슈도모나스 시링가에 pv. 토마토; 피토프토라 카프시시 (Phytophthora capsici); 잔토모나스 (Xanthomonas)

[1124] 가지 병해; 포모프시스 벡산스 (Phomopsis vexans), 에리시페 시코라세아럼 (Erysiphe cichoracearum); 겨자과 식물 병해: 알테르나리아 자포니카 (Alternaria japonica), 세르코스포렐라 브라시카에 (Cercosporella brassicae), 플라스모디오포라 브라시카에 (Plasmodiophora brassicae), 페로노스포라 파라시티카 (Peronospora parasitica);

[1125] 웰시 양파 병해: 푸치니아 알리이 (Puccinia allii) 및 페로노스포라 데스트럭터 (Peronospora destructor);

[1126] 대두 병해: 세르코스포라 키쿠치이 (Cercospora kikuchii), 엘시노에 글리시네스 (Elsinoe 글리신s), 디아포르테 파세오로룸 변이체 소자에 (Diaporthe phaseolorum var. sojae), 셉토리아 글리신스 (Septoria 글리신s),

세르코스포라 소지나 (Cercospora sojina), 파콥소라 파치리지 (Phakopsora pachyrhizi), 피토프토라 소자에 (Phytophthora sojae), 리족토니아 솔라니 (Rhizoctonia solani), 코르니스포라 카시콜라 (Corynespora cassiicola), 및 스클레로니티아 스클레오티오룸 (Sclerotinia sclerotiorum);

[1127] 강낭콩 병해: 콜레토트리쿰 린뎀티아눔 (Colletotrichum lindemthianum);

[1128] 땅콩 병해: 세르코스포라 페르소나타 (Cercospora personata), 세르코스포라 아라키디콜라 (Cercospora arachidicola) 및 스클레로티움 롤프시 (Sclerotium rolfsii);

[1129] 완두 병해: 에리시페 피시 (Erysiphe pisi);

[1130] 감자 병해: 알테르나리아 솔라니 (Alternaria solani) 및 피토프토라 인페스탄스 (Phytophthora infestans), 피토프토라 에리스로셉티카 (), 및 스폰고스포라 서브터란네안 피토프토라 종 서브터라네안 f. 종 서브터라네안 (Spongospora subterranean f. sp. Subterranean);

[1131] 딸기 병해: 스파에로테카 후물리 (Sphaerotheca humuli) 및 글로메렐라 신구라타 (Glomerella cingulata);

[1132] 차 병해: 엑소바시디움 레티쿨라툼 (Exobasidium reticulatum), 엘시노에 류코스필라 (Elsinoe leucospila), 페스탈로티옵시스 종 (Pestalotiopsis sp) 및 콜레토트리쿰 테아에시넨시스 (Colletotrichum theaesinensis);

[1133] 담배 병해: 알테르나리아 론지페스 (Alternaria longipes), 에리시페 시초라세아룸 (Erysiphe cichoracearum), 콜레토트리쿰 타바쿰 (Colletotrichum tabacum), 페로노스포라 타바시나 (Peronospora tabacina) 및 피토프토라 니코티아나에 (Phytophthora nicotianae);

[1134] 유채 병해: 스클레로티니아 스클레로티오룸 (Sclerotinia sclerotiorum), 및 리족토니아 솔라니 (Rhizoctonia solani);

[1135] 면화 병해: 리족토니아 솔라니 (Rhizoctonia solani);

[1136] 사탕무 병해: 세르코스포라 베티콜라 (Cercospora beticola), 타나테포루스 쿠쿠메리스 (Thanatephorus cucumeris), 타나테포루스 쿠쿠메리스 (Thanatephorus cucumeris) 및 아파니데르마툼 코클리오이데스 (Aphanidermatum cochlioides);

[1137] 장미 병해: 디플로카르폰 로사에 (Diplocarpon rosae), 스파에로테카 판노사 (Sphaerotheca pannosa) 및 페로노스포라 스파르사 (Peronospora sparsa);

[1138] 국화 및 국화과 병해: 브레미아 락투카에 (Bremia lactucae), 셉토리아 크리산테미-인디시 (Septoria chrysanthemi-indici) 및 푸치니아 호리아나 (Puccinia horiana);

[1139] 다양한 식물의 병해: 피티움 아파니데르마툼 (Pythium aphanidermatum), 피티움 데바리아눔 (Pythium debarianum), 피티움 그라미니콜라 (Pythium graminicola), 피티움 이레귤라레 (Pythium irregulare), 및 피티움 울티뭄 (Pythium ultimum), 보트리티스 시네레아 (Botrytis cinerea), 및 스클레로티니아 스클레로티오룸 (Sclerotinia sclerotiorum);

[1140] 왜무 병해: 알테르나리아 브라시키콜라 (Alternaria brassicicola);

[1141] 잔디 병해: 스클레로티니아 호메오카르파 (Sclerotinia homeocarpa), 리족토니아 솔라니 (Rhizoctonia solani);

[1142] 바나나 병해: 미코스파에렐라 피지엔시스 (Mycosphaerella fijiensis), 미코스파에렐라 무시콜라 (Mycosphaerella musicola);

[1143] 해바라기 병해: 플라스모파라 할스테디 (Plasmopara halstedii);

[1144] 아스퍼질러스 종, 페니실륨 종 (Penicillium spp.), 푸사리움 종, 지베렐라 종 (Gibberella spp.), 트리코더마 종, 티엘라비옵시스 종 (Thielaviopsis spp.), 리조푸스 종, 무코르 종 (Mucor spp.), 코르티시움 종 (Corticium spp.), 로마 종 (Rhoma spp.), 리족토니아 종 (Rhizoctonia spp.) 또는 디플로디아 종 (Diplodia spp.)에 의해 초래되는 다양한 식물의 초기 성장 단계에서의 병해 또는 종자 병해;

[1145] 폴릭시마 종 (Polymixa spp.) 또는 올피디움 종 (Olpidium spp.) 등에 의해 매개되는 다양한 식물의 바이러스병 등.

[1146] *3. 제초제 내성을 부여하는 유전자의 예*:

[1147] 성장점 또는 분열조직을 저해하는 제초제에 대한 내성, 예컨대, 이미다졸리논 또는 설포닐유리아, 예를 들어, 각각 문헌[Lee et al., EMBO J. 7:1241(1988), 및 Miki et al., Theor. Appl. Genet. 80:449 (1990)].

[1148] 글리포세이트 내성 (예를 들어, 스트렙토마이세스 하이그로스코피쿠스 (Streptomyces hygroscopicus) 및 스트렙토마이세스 비리디크로모게네스 (Streptomyces viridichromogenes)를 비롯한 스트렙토마이세스 종으로부터의 각각 돌연변이체 5-엔올파이루빌시키메이트-3-포스페이트 신타제(EPSP) 유전자, aroA 유전자 및 글리포세이트 아세틸 트랜스퍼라제(GAT) 유전자에 의해 부여되는 내성), 또는 다른 포스포노 화합물, 예컨대, 글루포시네이트 (포스피노트리신 아세틸 트랜스퍼라제(PAT) 유전자에 대한 내성), 및 ACCase 억제제-코딩 유전자에 의한 피리딘옥시 또는 페녹시 프로프리온산 및 시클로헥손에 대한 내성. 참조: 예를 들어 U.S. 특허 번호 4,940,835 및 U.S. 특허 6,248,876 , U.S. 특허 번호 4,769,061 , EP 번호 0 333 033 및 U.S. 특허 번호 4,975,374. 참조: 또한 EP 번호 0242246, DeGreef et al., Bio/Technology 7:61 (1989), Marshall et al., Theor. Appl. Genet. 83:435 (1992), WO 2005012515 to Castle et. al. 및 WO 2005107437.

[1149] 광합성을 억제하는 제조체에 대한 내성, 예컨대, 문헌 [Przibila et al., Plant Cell 3:169 (1991)], 미국 특허 번호 4,810,648, 및 문헌 [Hayes et al., Biochem. J. 285: 173 (1992)]의 트라이아진(psbA 및 gs+ 유전자) 또는 벤조니트릴 (니트릴라제 유전자), 및 글루타티온 S-트랜스퍼라제.

[1150] 제초제를 해독하는 효소 또는 저해에 내성이 있는 돌연변이체 글루타민 합성효소를 암호화하는 유전자, 예를 들어, 미국 특허 출원 제11/760,602호. 또는 해독 효소는 포스피노트리신 아세틸트랜스퍼라제를 코딩하는 효소 (예컨대, 스트렙토마이세스 종으로부터의 bar 또는 pat 단백질)이다. 포스피노트리신 아세틸트랜스퍼라제는 예를 들어 U.S. 특허 번호 5,561,236; 5,648,477; 5,646,024; 5,273,894; 5,637,489; 5,276,268; 5,739,082; 5,908,810 및 7,112,665에 기술된다.

[1151] 히드록시페닐피루베이트 디옥시게나제 (HPPD) 억제제, 천연 발생 HPPD 내성 유전자, 또는 WO 96/38567, WO 99/24585, 및 WO 99/24586, WO 2009/144079, WO 2002/046387, 또는 미국 특허 제6,768,044호에 기재된 바와 같은 돌연변이 또는 키메라 HPPD 효소를 코딩하는 유전자.

[1152] *4. 항생제 스트레스 내성에 관여된 유전자의 예*:

[1153] WO 00/04173 또는 WO/2006/045633에 기재된 바와 같은 식물 세포 또는 식물에서의 폴리(ADP-리보스) 중합효소 (PARP) 유전자의 발현 및/또는 활성을 감소시킬 수 있는 이식유전자.

[1154] WO 2004/090140에 기재된 바와 같은, 식물 또는 식물 세포의 유전자를 코딩하는 PARG의 발현 및/또는 활성을 감소시킬 수 있는 이식유전자.

[1155] 예를 들어, 유럽 특허 제04077624.7호, WO 2006/133827, PCT/EP07/002,433, 유럽 특허 제1999263호 또는 WO 2007/107326에 기재된 바와 같은 니코틴아미다제, 니코티네이트 포스포리보실트랜스퍼라제, 니코틴산 모노뉴클레오티드 아데닐 트랜스퍼라제, 니코틴아마이드 아데닌 다이뉴클레오티드 합성효소 또는 니코틴 아마이드 포스포리보실트랜스퍼라제를 비롯한 니코틴아마이드 아데닌 다이뉴클레오티드 샐비지 합성(salvage synthesis) 경로의 식물-기능성 효소를 코딩하는 이식유전자.

[1156] 탄수화물 합성에 관여되는 효소는 예를 들어 하기 문헌에 기술된 것들을 포함한다: EP 0571427, WO 95/04826, EP 0719338, WO 96/15248, WO 96/19581, WO 96/27674, WO 97/11188, WO 97/26362, WO 97/32985, WO 97/42328, WO 97/44472, WO 97/45545, WO 98/27212, WO 98/40503, WO99/58688, WO 99/58690, WO 99/58654, WO 00/08184, WO 00/08185, WO 00/08175, WO 00/28052, WO 00/77229, WO 01/12782, WO 01/12826, WO 02/1059, WO 03/071860, WO 2004/056999, WO 2005/030942, WO 2005/030941, WO 2005/095632, WO 2005/095617, WO 2005/095619, WO 2005/095618, WO 2005/123927, WO 2006/018319, WO 2006/103107, WO 2006/108702, WO 2007/009823, WO 00/22140, WO 2006/063862, WO 2006/072603, WO 02/034923, EP 06090134.5, EP 06090228.5, EP 06090227.7, EP 07090007.1, EP 07090009.7, WO 01/14569, WO 02/79410, WO 03/33540, WO 2004/078983, WO 01/19975, WO 95/26407, WO 96/34968, WO 98/20145, WO 99/12950, WO 99/66050, WO 99/53072, U.S. 특허 번호 6,734,341, WO 00/11192, WO 98/22604, WO 98/32326, WO 01/98509, WO 01/98509, WO 2005/002359, U.S. 특허 번호 5,824,790, U.S. 특허 번호 6,013,861, WO 94/04693, WO 94/09144, WO 94/11520, WO 95/35026 또는 WO 97/20936 또는 EP 0663956, WO 96/01904, WO 96/21023, WO 98/39460, 및 WO 99/24593에 개시된 바ㅣ와 같은 폴리프룩토스, 특히 인슐린 및 레반-유형의 생산에 관여하는 효소, WO 95/31553, US 2002031826, U.S. 특허 번호

6,284,479, U.S. 특허 번호 5,712,107, WO 97/47806, WO 97/47807, WO 97/47808 및 WO 00/14249에 개시된 알파-1,4-글루칸의 생산, WO 00/73422에 개시된 바와 같은, 알파-1,6 분지형 알파-1,4-글루칸의 생산, 예를 들어 WO 00/47727, WO 00/73422, EP 06077301.7, U.S. 특허 번호 5,908,975 및 EP 0728213에 개시된 바와 같은 알테르난의 생산, 예를 들어 WO 2006/032538, WO 2007/039314, WO 2007/039315, WO 2007/039316, JP 2006304779, 및 WO 2005/012529에 개시된 바와 같은 히알루로난의 생산.

[1157] 가뭄 내성을 개선시키는 유전자. 예를 들어, WO 2013122472는 기능성 유비퀴틴 단백질 리가제 단백질(UPL) 단백질, 보다 특히, UPL3의 부재 또는 감소된 수준이 물에 대한 감소된 필요 또는 상기 식물의 가뭄에 대한 개선된 내성을 야기한다는 것을 개시한다. 증가된 가뭄 내성을 갖는 유전자이식 식물의 다른 예는, 예를 들어, 미국 특허 제2009/0144850호, 미국 특허 제2007/0266453호 및 WO 2002/083911에 개시되어 있다. 미국 특허 제2009/0144850호는 DR02 핵산의 변경된 발현에 기인하여 가뭄 내성 표현형을 나타낸 식물을 기재한다. 미국 특허 제2007/0266453호는 DR03 핵산의 변경된 발현에 기인하여 가뭄 내성 표현형을 나타내는 식물을 기재하고, WO 2002/083911호는 가드 세포에서 발현되는 ABC 수송체의 감소된 활성에 기인하여 가뭄 스트레스에 대해 증가된 내성을 갖는 식물을 기재한다. 다른 예는 유전자이식 식물에서 DREB1 A를 코딩하는 cDNA의 과발현이 정상 성장 조건 하에 많은 스트레스 내성 유전자의 발현을 활성화시키고 가뭄, 염 로딩 및 결빙에 대한 개선된 내성을 초래한다는 것을 기재하는 Kasuga와 그의 공저자 (1999)에 의한 작업이다. 그러나, DREB1A의 발현은 또한 정상 성장 조건 하에 몇몇 성장 지연을 초래하였다 (Kasuga (1999) Nat Biotechnol 17(3) 287291).

[1158] 추가 특정 구현예에서, 작물 식물은 특정 식물 형질에 영향을 미침으로써 개선될 수 있다. 예를 들어, 살충제-내성 식물의 개발, 식물에서 질환 내성의 개선, 식물 곤충 및 선충 개성의 개선, 기생충 잡초에 대한 식물 내성의 개선, 식물 가뭄 내성의 개선, 식물 영양가의 개선, 식물 스트레스 내성, 자기-수분 회피, 식물 사료 소화흡수율 생물량, 곡물 수확률 등의 개선에 의함. 소수의 특정 비제한적 예를 본 명세서에서 이하에 제공한다.

[1159] 단일한 유전자의 표적된 돌연변이에 추가하여, 다수의 유전자의 표적된 돌연변이, 염색체 단편의 결실, 이식유전자의 부위-특이적 통합, 생체 내 부위-유도된 돌연변이 및 식물에서 정밀 유전자 치환 또는 대립 형질 교체를 가능하게 하도록 C2c1 CrISPR 복합체가 설계될 수 있다. 따라서, 본 명세서에 기재된 방법은 유전자 발견 및 확인, 돌연변이 및 동종 유전자이식(cisgenic) 육종, 및 잡종 육종에서 광범위하게 적용된다. 이들 출원은 다양한 개선된 작물학적 형질, 예컨대, 제초제 내성, 질환 내성, 항생제 스트레스 내성, 고수율 및 우수한 품질을 갖는 유전자 변형된 작물의 새로운 생성의 생산을 용이하게 한다.

[1160] **웅성불임 식물의 창출을 위한 C2c1 유전자의 이용**

[1161] 잡종 식물은 전형적으로 근친교배 식물에 비해 유리한 작물학적 형질을 갖는다. 그러나, 자기-수분 식물에 대해, 잡종의 생성은 도전 중일 수 있다. 상이한 식물 유형에서, 식물 생식력, 보다 특히, 수컷 생식력에 중요한 유전자가 동정되었다. 예를 들어, 메이즈에서, 생식력에 중요한 둘 이상의 유전자가 확인되었다 (Amitabh Mohanty International Conference on New Plant Breeding Molecular Technologies Technology Development And Regulation, Oct 9-10, 2014, Jaipur, India; Svitashev et al. Plant Physiol. 2015 Oct;169(2):931-45; Djukanovic et al. Plant J. 2013 Dec;76(5):888-99). 본 명세서에 제공된 방법은, 쉽게 교배하여 잡종을 생성할 수 있는 웅성 불임 식물을 생산하기 위하여 웅성 생식력에 요구되는 유전자를 표적화하는데 사용될 수 있다. 특정 구현예에서, 본 명세서에 제공된 CRISPR-C2c1 시스템은 사이토크롬 P450-유사 유전자 (MS26) 또는 메가뉴클레아제 유전자 (MS45)의 표적화된 돌연변이유발을 위해 사용되며, 이에 의해 메이즈 식물에 대한 웅성 불임성을 부여한다. 이렇게 해서 유전자가 변경된 메이즈 식물은 잡종 육종 프로그램에서 사용될 수 있다.

[1162] **식물에서 수정 단계의 증가**

[1163] 특정 구현예에서, 본 명세서에서 제공된 방법은, 벼와 같은 식물의 생식력을 연장시키는데 사용된다. 예를 들어, 벼 생식 단계 유전자, 예컨대, Ehd3은 유전자에서 돌연변이를 생성하기 위해 표적화될 수 있고, 소식물체는 연장된 재생 식물 생식 단계를 위해 선택될 수 있다 (CN 104004782에 기재된 바와 같음).

[1164] **관심 작물에서 유전적 변이를 생성하기 위한 C2c1의 용도**

[1165] 작물 식물에서 야생 생식질 및 유전자 변형의 이용 가능성은 작물 개선 프로그램에서 중요하지만, 작물 식물로부터의 생식질에서 이용 가능한 다양성은 제한되어 있다. 본 발명은 관심 생식질에서 유전자 변형의 다양성을 생성하기 위한 방법을 계획한다. CRISPR-C2c1 시스템의 이러한 적용에서, 식물 게놈에서의 상이한 위치를 표적화하는 가이드 RNA의 라이브러리가 제공되고, 이는 C2c1 이펙터 단백질과 함께 식물 세포 내로

도입된다. 이 방법에서, 게놈-규모 점 돌연변이 및 유전자 녹아웃의 컬렉션이 생성될 수 있다. 특정 구현예에서, 상기 방법은 이렇게 얻어진 세포로부터 식물 부분 또는 식물을 생성하는 단계 및 관심 형질에 대해 세포를 스크리닝하는 단계를 포함한다. 표적 유전자는 코딩 영역과 넌코딩 영역을 둘 모두 포함할 수 있다.

특정 구현예에서, 형질은 스트레스 내성이며, 상기 방법은 스트레스-내성 작물 변종의 생성을 위한 방법이다.

[1166] **과일-숙성에 영향을 미치는 C2c1의 용도**

[1167] 숙성은 과일 및 채소의 성숙 과정에서의 정상적 단계이다. 이를 시작하고 며칠만에 과일 또는 채소는 먹을 수 없게 된다. 이 과정은 농부와 소비자 둘 모두에게 상당한 손실을 가져온다. 특정 구현예에서, 본 발명의 방법은 에틸렌 생성을 감소시키는 데 사용된다. 이는 다음 중 하나 이상을 보장함으로써 보장된다: a. ACC 신타제 유전자 발현의 억제. ACC(1-아미노시클로프로판-1-카복실산) 신타제는 S-아데노실메티오닌(SAM)의 ACC로의 전환을 담당하는 효소이고; 에틸렌 생합성의 두번째 단계에서 마지막 단계이다. 효소 발현은 안티센스(" 거울상") 또는 신타제 유전자의 절단된 복제물이 식물 게놈에 삽입될 때 입체방해된다; b. ACC 디아미나제 유전자의 삽입. 효소를 코딩하는 유전자는 공통 비병원성 토양 박테리아인 슈도모나스 클로로라피스(Pseudomonas chlororaphis)로부터 얻어진다. 이것은 ACC를 상이한 화합물로 전환시킴으로써 에틸렌 생성에 이용 가능한 ACC의 양을 감소시킨다; c. SAM 가수분해효소 유전자의 삽입. 이 접근은 ACC 디아미나제와 유사한데, 여기서 에틸렌 생성은 이의 전구체 대사물질의 양이 감소될 때 방해되며; 이 경우에 SAM이 호모세린으로 전환된다. 효소를 코딩하는 유전자는 이. 콜라이 T3 박테리오파지로부터 얻어진다. 그리고 d. ACC 옥시다제 유전자 발현의 억제. ACC 옥시다제는 에틸렌 생합성 경로에서의 마지막 단계인 ACC의 에틸렌으로의 산화를 촉매하는 효소이다. 본 명세서에 기재된 방법을 이용하여, ACC 옥시다제 유전자의 하향 조절은 에틸렌 생성의 억제를 초래하고, 이에 의해 과일 숙성을 지연시킨다. 특정 구현예에서, 상기 기재한 변형에 추가적으로 또는 대안적으로, 본 명세서에 기재된 방법은 과일에 의해 얻어지는 에틸렌 신호를 방해하기 위해 에틸렌 수용체를 변형시키는 데 사용된다. 특정 구현예에서, 에틸렌 결합 단백질을 코딩하는 ETR1 유전자의 발현은 변형되며, 보다 특히, 억제된다. 특정 구현예에서, 상기 기재한 변형에 추가적으로 또는 대안적으로, 본 명세서에 기재된 방법은 식물 세포벽의 완전성을 유지하는 물질인 펙틴의 분해를 초래하는 효소인 폴리갈락투로나제 (PG)를 코딩하는 유전자의 발현을 변형시키기 위해 사용된다. 펙틴 분해는 과일의 연화를 야기하는 숙성 과정의 시작 시 일어난다. 따라서, 특정 구현예에서, 본 명세서에 기재된 방법은 생성된 PG 효소의 양을 감소시킴으로써 펙틴 분해를 지연시키기 위해 PG 유전자에 돌연변이를 도입하거나 또는 PG 유전자의 활성화를 억제하기 위해 사용된다.

[1168] 따라서 특정 구현예에서, 상기 방법은 상기 기재한 바와 같은 식물 세포 게놈의 하나 이상의 변형을 보장하고 이로부터 식물을 재생하기 위한 CRISPR-C2c1 시스템의 용도를 포함한다. 특정 구현예에서, 식물은 토마토 식물이다.

[1169] **식물 저장 수명의 증가**

[1170] 특정 구현예에서, 본 발명의 방법은 식물 또는 식물 부분의 저장 수명에 영향을 미치는 화합물의 생성에 관여되는 유전자를 변형시키기 위해 사용된다. 보다 특히, 변형은 감자 괴경에서 환원당 축적을 방지하는 유전자에서의 변형이다. 고고온 처리 시, 이들 환원당은 유리 아미노산과 반응하여 갈색의, 쓴 맛이 나는 생성물 및 잠재적 발암 물질인 아크릴아미드의 상승된 수준을 초래한다. 특정 구현예에서, 본 명세서에서 제공된 방법은, 수크로오스를 글루코오스와 프룩토오스로 분해시키는 단백질을 코딩하는, 액포성 인버타제 유전자(VInv)의 발현을 감소 또는 저해시키는데 사용된다(Clasen et al. DOI: 10.1111/pbi.12370).

[1171] **가치 부가 형질을 보장하기 위한 CRISPR-C2c1 시스템의 용도**

[1172] 특정 구현예에서, CRISPR-C2c1 시스템은 영양적으로 개선된 농업 작물을 생산하는데 사용된다. 특정 구현예에서, 본 명세서에 제공된 방법은 " 기능성 식품" , 즉, 그것이 함유하는 전통적 영양소 이상으로 건강상 이득을 제공할 수 있는 변형 식품 또는 식품 성분 및 또는 " 뉴트라슈티컬 (nutraceutical)" , 즉, 식품 또는 식품의 부분으로 고려되고 질환의 예방 및 치료를 비롯한 건강상 이익을 제공할 수 있는 물질을 생산하는 데 적합하다. 특정 구현예에서, 뉴트라슈티컬은 암, 당뇨병, 심혈관 질환 및 고혈압 중 하나 이상의 예방 및/또는 치료에서 유용하다.

[1173] 영양적으로 개선된 작물의 예는 하기를 포함한다 (Newell-McGloughlin, Plant Physiology, July 2008, Vol. 147, pp. 939-953):

[1174] 변형된 단백질 품질, 함량 및/또는 아미노산 조성이 예컨대 바히아그라스(Bahiagrass) (문헌 [Luciani et al. 2005, Florida Genetics Conference Poster]), 카놀라 (Roesler et al., 1997, Plant Physiol 113 75?81), 메이즈 (문헌 [Cromwell et al, 1967, 1969 J Anim Sci 26 1325-1331, O'Quin et al. 2000 J Anim Sci 78 21442149, Yang et al. 2002, Transgenic Res 11 1120, Young et al. 2004, Plant J 38 910922]), 감자 (문헌 [Yu J and Ao, 1997 Acta Bot Sin 39 329334; (2000) Proc Natl Acad Sci USA 97 37243729, Enfissi et al. 2001, Chin Sci Bull 46 482484]), 쌀 (문헌 [Katsube et al. 1999, Plant Physiol 120 10631074]), 대두 (문헌 [Dinkins et al. 2001, Rapp 2002, In Vitro Cell Dev Biol Plant 37 742747]), 고구마 (문헌 [Egnin and Prakash 1997, In Vitro Cell Dev Biol 33 52A])에 대해 설명되었다.

[1175] 필수 아미노산 함량이, 예컨대, 카놀라 (문헌 [Falco et al. 1995, Bio/Technology 13 577582]), 루핀(Lupin) (문헌 [White et al. 2001, J Sci Food Agric 81 147154]), 메이즈 (문헌 [Lai and Messing, 2002, Agbios 2008 GM crop database (March 11, 2008)]), 감자 (문헌 [Zeh et al. 2001, Plant Physiol 127 792802]), 수수 (문헌 [Zhao et al. 2003, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 413416]), 대두 (문헌 [Falco et al. 1995 Bio/Technology 13 577-582; Galili et al. 2002 Crit Rev Plant Sci 21 167204]) 에 대해 설명되었다.

[1176] 오일 및 지방산 예컨대 카놀라 (문헌 [Dehesh et al. (1996) Plant J 9 167-172 [PubMed] ; Del Vecchio (1996) INFORM International News on Fats, Oils and Related Materials 7 230-243; Roesler et al. (1997) Plant Physiol 113 75-81 [PMC free article] [PubMed]; Froman and Ursin (2002, 2003) Abstracts of Papers of the American Chemical Society 223 U35; James et al. (2003) Am J Clin Nutr 77 1140-1145 [PubMed]; Agbios (2008, above); cotoan (Chapman et al. (2001) . J Am Oil Chem Soc 78 941-947; Liu et al. (2002) J Am Coll Nutr 21 205S-211S [PubMed]; O'Neill (2007) Australian Life Scientist. www.biotechnews.com.au/index.php/id;866694817;fp;4;fpid;2 (June 17, 2008), 아마인 (Abbadi et al., 2004, Plant Cell 16: 2734-2748), 메이즈 (Young et al., 2004, Plant J 38 910-922), 팜유 (Jalani et al. 1997, J Am Oil Chem Soc 74 1451-1455; Parveez, 2003, AgBiotechNet 113 1-8), 벼 (Anai et al., 2003, Plant Cell Rep 21 988-992), 대두 (Reddy and Thomas, 1996, Nat Biotechnol 14 639-642; Kinney and Kwolton, 1998, Blackie Academic and Professional, London, pp 193-213), 해바라기 (Arcadia, Biosciences 2008)

[1177] 탄수화물, 예컨대 치커리에 대해 설명된 프룩탄(Fructans) (문헌 [Smeekens (1997) Trends Plant Sci 2 286287, Sprenger et al.(1997) FEBS Lett 400 355358, Sevenier et al.(1998) Nat Biotechnol 16 843846]), 메이즈 (문헌 [Caimi et al.(1996) Plant Physiol 110 355363]), 감자 (문헌 [Hellwege et al.,1997 Plant J 12 10571065]), 사탕무 (문헌 [Smeekens et al. 1997, 상기]), 감자에 대해 설명된 것과 같은, 이눌린 (문헌 [Hellewege et al. 2000, Proc Natl Acad Sci USA 97 86998704]), 쌀에 대해 설명된 것과 같은, 전분 (문헌 [Schwall et al.(2000) Nat Biotechnol 18 551554, Chiang et al.(2005) Mol Breed 15 125143]),

[1178] 카놀라 (문헌 [Shintani and DellaPenna (1998) Science 282 20982100]), 메이즈 (문헌 [Rocheford et al.(2002). J Am Coll Nutr 21 191S-198S, Cahoon et al.(2003) Nat Biotechnol 21 10821087, Chen et al.(2003) Proc Natl Acad Sci USA 100 35253530]), 겨자씨 (문헌 [Shewmaker et al.(1999) Plant J 20 401412]), 감자 (문헌 [Ducreux et al., 2005, J Exp Bot 56 8189]), 쌀 (문헌 [Ye et al.(2000) Science 287 303305]), 딸기 (문헌 [Agius et al.(2003), Nat Biotechnol 21 177181]), 토마토 (문헌 [Rosati et al.(2000) Plant J 24 413419, Fraser et al.(2001) J Sci Food Agric 81 822827, Mehta et al.(2002) Nat Biotechnol 20 613618, Diaz de la Garza et al.(2004) Proc Natl Acad Sci USA 101 1372013725, Enfissi et al.(2005) Plant Biotechnol J 3 1727, DellaPenna (2007) Proc Natl Acad Sci USA 104 36753676])에 대해 설명된 것과 같은, 비타민 및 카로티노이드.

[1179] 사과 (스틸벤(stilbene), 문헌 [Szankowski et al. (2003) Plant Cell Rep 22: 141-149), 알팔파 (레스베라트롤, Hipskind and Paiva (2000) Mol Plant Microbe Interact 13 551-562), 키위 (resveratrol, Kobayashi et al.(2000) Plant Cell Rep 19 904910]), 메이즈 및 대두 (플라보노이드, Yu et al.(2000) Plant Physiol 124 781794), 감자 (안토시아닌 및 알칼로이드 글리코시드, 문헌 [Lukaszewicz et al.(2004) J Agric Food Chem 52 15261533]), 쌀 (플라보노이드 & 레스베라트롤, 문헌 [Stark-Lorenzen et al.(1997) Plant Cell Rep 16 668673, Shin et al.(2006) Plant Biotechnol J 4 303315]), 토마토 (+레스베라트롤, 클로로겐산, 플라보노이드, 스틸벤; Rosati et al.(2000) 상기, Muir et al.(2001) Nature 19 470474, Niggeweg et al.(2004) Nat Biotechnol 22 746754, Giovinazzo et al.(2005) Plant Biotechnol J 3 5769]), 밀 (카페인산 및 페룰산, 레스

베라트롤;United Press International (2002)); 및

[1180] 알팔파 (피타제(phytase), 문헌 [Austin-Phillips et al. (1999) www.molecularfarming.com/nonmedical.html), 상추 (iron, Goto et al.(2000) Theor Appl Genet 100 658664]), 쌀 (철, 문헌 [Lucca et al.(2002) J Am Coll Nutr 21 184S-190S]), 메이즈, 대두 및 밀 (피타제, [Drakakaki et al.(2005) Plant Mol Biol 59 869880, Denbow et al.(1998) Poult Sci 77 878881, Brinch-Pedersen et al.(2000) Mol Breed 6 195206])에 설명된 것과 같은 무기질 이용성.

[1181] 특정 구현예에서, 가치-부가된 형질은 식물에 존재하는 화합물의 생각되는 건강상의 이점에 관한 것이다. 예를 들어, 특정 구현예에서, 가치-부가된 작물은 다음의 화합물 중 하나 이상의 변형을 보장하거나 또는 다음의 화합물 중 하나 이상의 합성을 유도/증가시키기 위해 본 발명의 방법을 적용함으로써 얻어진다:

[1182] 카로테노이드, 예컨대, 세포에 대한 손상을 야기할 수 있는 유리 라디칼을 중화시키는 당근에 존재하는 α-카로텐 또는 유리 라디칼을 중화시키는 다양한 과일 및 채소에 존재하는 β-카로텐.

[1183] 건강한 시력의 유지에 기여하는 녹색 채소에 존재하는 루테인.

[1184] 전립선 암의 위험을 감소시키는 것으로 여겨지는, 토마토 및 토마토 생성물에 존재하는 라이코펜.

[1185] 건강한 시력의 유지에 기여하는 감귤류 및 메이즈에 존재하는 제아잔틴.

[1186] 유방 및/또는 결장암의 위험을 감소시킬 수 있는 밀겨에 존재하는 불용성 섬유질과 같은 식이섬유 및 귀리에 존재하는 β-글루칸, 심혈관 질환(CVD)의 위험을 감소시키는 금불초 및 전체 곡류에 존재하는 가용성 섬유질.

[1187] 지방산, 예컨대, CVD 위험을 감소시키고 정신 및 시각 기능을 개선시키는 ω-3 지방산, 신체 조성을 개선시킬 수 있고, 특정 암의 위험을 감소시킬 수 있는 접합된 리놀렌산 및 암 및 CVD의 염증 위험을 감소시킬 수 있고, 신체 조성을 개선시킬 수 있는 GLA.

[1188] 플라보노이드, 예컨대, 항산화제-유사 활성을 갖고, 퇴행성 질환의 위험을 감소시킬 수 있는, 밀에 존재하는 하이드록시신나메이트, 자유 라디칼을 중화시키고 암 위험을 감소시킬 수 있는 과일 및 채소에 존재하는 플라보놀, 카테신.

[1189] 글루코시놀레이트, 인돌, 아이소티오시아네이트, 예컨대, 자유 라디칼을 중화시키고, 암 위험을 감소시킬 수 있는, 십자화과 채소(브로콜리, 케일), 겨자무에 존재하는 설포라판.

[1190] 페놀 수지류, 예컨대, 퇴행성 질환, 심장병 및 암 위험을 감소시킬 수 있고, 장기간의 효과를 가질 수 있는 포도에 존재하는 스틸벤 및 항산화제-유사 활성을 갖고, 퇴행성 질환, 심장병, 및 눈병 위험을 감소시킬 수 있는 채소 및 감귤류에 존재하는 카페인산 및 페룰산, 및 항산화제-유사 활성을 갖고, 퇴행성 질환 및 심장병의 위험을 감소시킬 수 있는 카카오에 존재하는 에피카테킨.

[1191] 혈액 콜레스테롤 수준을 낮춤으로써 관상동맥 심장병의 위험을 감소시키는 메이즈, 콩, 밀 및 목재 오일에 존재하는 식물 스탄올/스테롤.

[1192] 위장 건강을 개선시킬 수 있는 예루살렘 아티초크, 샬롯, 양파 분말에 존재하는 프럭탄, 이눌린, 프럭토-올리고당.

[1193] LDL 콜레스테롤을 낮출 수 있는 대두에 존재하는 사포닌.

[1194] 심장병 위험을 감소시킬 수 있는 대두에 존재하는 대두 단백질.

[1195] 식물성 에스트로겐, 예컨대, 폐경기 증상, 예컨대, 일과성 열감을 감소시킬 수 있고, 골다공증 및 CVD를 감소시킬 수 있는 대두에 존재하는 아이소플라본, 및 심장병 및 일부 암에 대해 보호할 수 있고, LDL 콜레스테롤, 총 콜레스테롤을 낮출 수 있는 아마, 호밀 및 채소에 존재하는 리그난.

[1196] 설파이드 및 티올, 예컨대, 양파, 마늘, 올리브, 리크 및 스캘런에 존재하는 다이알릴 설파이드 및 LDL 콜레스테롤을 낮출 수 있고 건강한 면역계를 유지하도록 하는 십자화과 채소에 존재하는 알릴 메틸 트라이설파이드, 다이티올티온.

[1197] 탄닌, 예컨대, 요로 건강상태를 개선시킬 수 있고, CVD 및 고혈압 위험을 감소시킬 수 있는 체키, 코코아에 존재하는 프로안토시아니딘.

[1198] 추가로, 본 발명의 방법은 또한 단백질/전분 작용성, 저장 수명, 맛/미관, 섬유 품질, 및 알레르겐, 항영양소

(antinutrient) 및 독소 감소 형질의 변형을 계획한다.

[1199] 따라서, 본 발명은 영양적 가치가 더해진 식물을 생산하는 방법을 포함하고, 상기 방법은 본 명세서에 기재된 바와 같은 CRISPR-C2c1 시스템을 이용하여 영양적 가치가 더해진 성분의 생산에 관여되는 유전자를 코딩하는 유전자를 식물 세포 내로 도입하는 단계 및 상기 식물 세포로부터 식물을 재생하는 단계를 포함하고, 상기 식물은 영양적 가치가 더해진 상기 성분의 발현 증가를 특징으로 한다. 특정 구현예에서, CRISPR-C2c1 시스템은, 예를 들어, 본 화합물의 대사를 제어하는 하나 이상의 전사 인자를 변형시킴으로써, 이들 화합물의 내생성 합성을 간접적으로 변형시키는 데 사용된다. 식물 세포 내로 관심 유전자를 도입하고/하거나 CRISPR-C2c1 시스템을 이용하여 내생성 유전자를 변형시키는 방법은 본 명세서에서 상기에 기재되어 있다.

[1200] 가치-부가된 형질을 부여하도록 변형된 식물에서 변형의 일부 특정 예는 식물의 스테아르산 함량을 증가시키도록 스테아릴-ACP 불포화효소의 안티센스 유전자로 식물을 형질전환시켜서 지방산 대사작용이 변형된 식물이다. 문헌 [Knultzon et al., Proc. Natl. Acad. Sci. U.S.A. 89:2624 (1992)]을 참조한다. 다른 예는 예를 들어 저수준의 피트산을 특징으로 하는 메이즈 돌연변이체의 원인일 수 있는 단일 대립유전자와 연관된 DNA를 클로닝하여 재도입시킴으로써, 피테이트 함량을 감소시키는 것을 포함한다. 문헌 [Raboy et al, Maydica 35:383 (1990)] 참조한다.

[1201] 유사하게, 강력 프로모터의 제어 하에서 메이즈 호분층 내 플라보노이드의 생산을 조절하는 메이즈 (제아 메이스 (Zea mays)) Tfs C1 및 R의 발현은 아마도 전체 경로를 활성화시킴으로써, 애기장대 (아라비돕시스 탈리아나)에서 안토시아닌의 높은 축적률을 일으켰다 (Bruce et al., 2000, Plant Cell 12:6580). DellaPenna (Welsch et al., 2007 Annu Rev Plant Biol 57: 711- 738)는 Tf RAP2.2 및 그의 상호작용 상대 SINAT2가 애기장대 잎에서 카로텐 발생을 증가시켰다는 것을 발견하였다. Tf Dof1의 발현은 탄소 골격 생성을 위한 효소를 코딩하는 유전자의 상향조절, 아미노산 함량의 현저한 증가 및 유전자이식 애기장대에서 Glc 수준의 감소 (Yanagisawa, 2004 Plant Cell Physiol 45: 386- 391), 및 애기장대에서의 글루코시놀레이트 생합성 경로의 DOF Tf AtDof1.1(OBP2) 상향조절된 모든 단계를 유도하였다 (Skirycz et al., 2006 Plant J 47: 10-24).

[1202] **식물에서 알레르겐의 감소**

[1203] 특정 구현예에서, 본 명세서에 제공되는 방법은 알레르겐 수준이 감소된 식물을 생성하여, 소비자에 대해 그들을 더 안전하게 만들기 위해 사용된다. 특정 구현예에서, 상기 방법은 식물 알레르겐을 생성을 초래하는 하나 이상의 유전자 발현을 변형시키는 단계를 포함한다. 예를 들어, 특정 실시형태에서, 방법은 식물 세포, 예컨대, 독보리 식물 세포에서 Lol p5 유전자의 발현을 하향 조절하기 위해 CRISPR-C2c1 시스템을 전달하는 단계 및 상기 식물의 화분의 알레르기 항원성을 감소시키기 위해 그들로부터 식물을 재생시키는 단계를 포함한다 (Bhalla et al. 1999, Proc. Natl. Acad. Sci. USA Vol. 96: 11676- 11680). 특정 구현예에서, CRISPR-C2c1 시스템은 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 Lol p5 유전자에 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 Lol p5를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 Lol p5 유전자에 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 Lol p5 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1204] 일반적으로 땅콩 알레르기 및 콩과 식물에 대한 알레르기는 실제로 심각한 건강상의 문제이다. 본 발명의 C2c1 이펙터 단백질 시스템은 이러한 콩과의 알레르겐성 단백질을 코딩하는 유전자를 확인 및 이후 편집 또는 침묵화하는 데 사용될 수 있다. 이러한 유전자 및 단백질에 관해 제한없이, Nicolaou 등은 땅콩, 대두, 렌틸콩, 완두콩, 루핀, 그린 빈스, 및 녹두에서의 알레르기 단백질을 동정하였다. 문헌 [Nicolaou et al., Current Opinion in Allergy and Clinical Immunology 2011;11(3):222)]을 참조한다.

[1205] **관심 내생성 유전자에 대한 스크리닝 방법**

[1206] 본 명세서에 제공된 방법은 식물 종, 식물 문 및 식물계 전체에서, 영양적 부가 가치 성분의 생산에 관여되는 효소를 코딩하는 가치있는 유전자 또는 일반적으로 관심 작물적 형질에 영향을 미치는 유전자의 동정을 더욱 허용한다. 예를 들어, 본 명세서에 기재된 바와 같은 CRISPR-C2c1 시스템을 이용하여 식물에서 대사 경로의 효소를 코딩하는 유전자를 선택적으로 표적화함으로써, 식물의 소정의 영양학적 양상을 초래하는 유전자가 동정될 수 있다. 유사하게는, 바람직한 작물학적 형질에 영향을 미치는 유전자를 선택적으로 표적화함으로써, 적절한 유전자가 동정될 수 있다. 따라서, 본 발명은 특정 영양학적 가치 및/또는 작물학적 형질을 갖는 화합물의 생성에 관여되는 효소를 코딩하는 유전자에 대한 스크리닝 방법을 포함한다.

[1207] **생물연료 생산에서 CRISPR-C2c1 시스템의 용도**

[1208] 본 명세서에서 설명된 것과 같은 용어 "생물연료(biofuel)" 는 식물 및 식물-유도된 자원으로부터 제조된 대안적인 연료이다. 재생성 생물연료는 에너지가 탄소 고정 방법을 통해 얻어지거나 또는 생물량의 사용 또는 전환을 통해 제조되는 유기 물질로부터 추출될 수 있다. 이러한 생물량은 생물연료에 대해 직접적으로 사용될 수 있거나 또는 열전환, 화학적 전환 및 생화학적 전환에 의해 물질을 함유하는 편리한 에너지로 전환될 수 있다. 이 생물량 전환은 고체, 액체 또는 기체 형태로 연료를 생성할 수 있다. 2가지 유형의 생물연료가 있다: 바이오에탄올 및 바이오디젤. 바이오에탄올은 대부분 메이즈 및 사탕수수로부터 유래된 셀룰로스 (전분)의 당 발효 공정에 의해 주로 생산된다. 반면에 바이오디젤은 주로 유채, 야자나무 및 대두와 같은 기름 작물로부터 생산된다. 생물연료는 주로 수송을 위해 사용된다.

[1209] **생물연료 생산을 위한 식물 속성 증강**

[1210] 특정 구현예에서, 본 명세서에 기재된 바와 같은 CRISPR-C2c1 시스템을 이용하는 방법은 발효를 위한 당의 더 효율적인 방출을 위해 중요한 가수분해제에 의한 접근을 용이하게 하기 위하여 세포벽의 특성을 변경시키는 데 사용된다. 특정 구현예에서, 셀룰로오스 및/또는 리그닌의 생합성이 변형된다. 셀룰로오스는 세포벽의 주 성분이다. 셀룰로오스 및 리그닌의 생합성은 공동조절된다. 식물에서 리그닌의 비를 감소시킴으로써 셀룰로오스의 비가 증가될 수 있다. 특정 구현예에서, 본 명세서에서 설명된 방법은, 발효가능한 탄수화물을 증가시키기 위하여 식물에서 리그닌 생합성을 하향조절하는데 사용된다. 보다 특히, 본 명세서에 기재된 방법은 WO 2008064289 A2에 개시된 바와 같이 4-쿠마레이트 3-히드록실라제 (C3H), 페닐알라닌 암모니아-리아제 (PAL), 신남에이트 4-히드록실라제 (C4H), 히드록시신남오일 트랜스퍼라제 (HCT), 카페인산 O-메틸트랜스퍼라제 (COMT), 카페오일 CoA 3-O-메틸트랜스퍼라제 (CCoAOMT), 페룰레이트 5-히드록실라제 (F5H), 신남일 알콜 디히드로게나제 (CAD), 신남오일 CoA-리덕타제 (CCR), 4-쿠마레이트-CoA 리가제 (4CL), 모노리그놀-리그닌-특이적 글리코실트랜스퍼라제, 및 알데히드 디히드로게나제 (ALDH)로 이루어진 군으로부터 선택되는 적어도 첫 번째의 리그닌 생합성 유전자를 하향조절하는 데 사용된다.

[1211] 특정 구현예에서, 본 명세서에 기재된 방법은 발효 동안 더 낮은 아세트산 수준을 생성하는 식물 매스 (mass)를 생산하는 데 사용된다 (또한 WO 2010096488 참조). 보다 특히, 본 명세서에 개시된 방법은 다당 아세틸화를 감소시키기 위해 CaslL에 대한 상동체에서 돌연변이를 생성하는 데 사용된다.

[1212] **생물연료 생산을 위한 효모 변형**

[1213] 특정 구현예에서, 본 명세서에서 제공된 C2c1 효소는 재조합 미생물에 의한 바이오에탄올 생산에 사용된다. 예를 들어, C2c1은 미생물, 예컨대 효모를 조작하여 발효성 당으로부터의 생물연료 또는 생체고분자를 생성하고, 선택적으로 발효성 당 공급원으로서 농업 폐기물로부터 유도된 식물-유래 리그노셀룰로오스를 분해시키는데 사용될 수 있다. 더욱 구체적으로는, 본 발명은, C2c1 CRISPR 복합체가 생물연료 생산에 요구되는 외래 유전자를 미생물 내로 도입 및/또는 생물연료 합성을 방해할 수 있는 원인의 내생성 유전자를 변형시키는데 사용되는 방법을 제공한다. 더욱 구체적으로, 본 방법은 피루베이트의 에탄올 또는 또 다른 관심 생성물로의 전환에 관련된 효소를 코딩하는 하나 이상의 뉴클레오티드 서열을 효모와 같은 미생물 내로 도입하는 것을 포함한다. 특정 구현예에서 본 방법은, 미생물이 셀룰로오스를 분해하는 것을 허용하는 하나 이상의 효소, 예컨대 셀룰라아제의 도입을 보장하도록 한다. 추가의 구현예에서, C2c1 CRISPR 복합체는 생물연료 생산 경로와 경쟁하는 내생성 대사 경로를 변형하는데 사용된다.

[1214] 따라서, 더욱 구체적인 구현예에서, 본 명세서에서 설명된 방법은 다음과 같이 미생물을 변형시키는데 사용된다:

[1215] 상기 미생물이 상기 핵산을 발현, 상기 식물 세포벽 분해 효소를 생산 및 분해할 수 있도록, 하나 이상의 이종성 핵산을 도입 또는 식물 세포벽 분해 효소를 코딩하는 하나 이상의 내생성 핵산의 발현을 증가;

[1216] 상기 숙주 세포가 상기 핵산을 발현할 수 있도록 하는, 하나 이상의 이종성 핵산을 도입 또는 아세트알데히드를 에탄올로 전환시키는 효소를 코딩하는 하나 이상의 이종성 핵산과 선택적으로 조합된, 피루베이트를 아세트알데히드로 전환하는 효소를 코딩하는 하나 이상의 내생성 핵산의 발현을 증가; 및/또는

[1217] 상기 숙주 세포 내 대사 경로에서 효소를 코딩하는 하나 이상의 핵산을 변형시키고, 여기서 상기 경로는 피루베이트로부터 아세트알데히드 또는 아세트알데히드로부터 에탄올 외의 대사산물을 생산하고, 여기서 상기 변형은 상기 대사산물의 감소된 생산을 결과로서 초래하거나, 또는 상기 효소의 억제제를 코딩하는 하나 이상의 뉴클레

오티드를 도입.

[1218] **식물성 오일 또는 생물연료의 생산을 위한 조류 및 식물의 변형**

[1219] 유전자이식 조류 또는 다른 식물, 예컨대, 유채는, 예를 들어, 식물성 오일 또는 생물연료, 예컨대, 알콜(특히 메탄올 및 에탄올)의 생산에 특히 유용할 수 있다. 이들은 오일 또는 생체연료 산업에서 사용을 위한 고수준의 오일 또는 알코올을 발현 또는 과발현시키도록 유전자조작될 수 있다.

[1220] 본 발명의 특정 구현예에 따르면, CRISPR-C2c1 시스템은 생물연료 생산에서 유용한 지질-풍부 규조류를 생성하는 데 사용된다.

[1221] 특정 구현예에서, 식물에 의해 생산되는 생물량 생산과 관련된 유전자를 특별히 변형시키는 것이 고려된다. 특정 구현예에서, CRISPR-C2c1 시스템은 테오신 분지된 (tb) 유전자 또는 이의 상동체를 표적화하여 높은 생물량 식물을 생성시키는데 사용된다. 일정 구현예에서, CRISPR-C2c1 시스템은 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 tb 유전자에 도입시킨다. 특정 구현예에서, CRISPR-C2c1 시스템은 tb 유전자에 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 tb 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다. 일정한 특정 구현예에서, CRISPR-C2c1 시스템은 tb 유전자 또는 이의 상동체에 단일 뉴클레오티드 돌연변이를 도입시키기 위해서 (예를 들어, 융합 단백질 또는 적합한 링커를 통해) 아데노신 또는 시티딘 디아미나제같은 기능성 도메인에 회합된 촉매적으로 불활성화된 C2c1 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 tb1a 및 tb1b 유전자를 표적화하고 단일 뉴클레오티드 돌연변이를 도입시켜 높은 생물량 지팽이풀 식물을 생성시키는데 사용된다. 참조: Liu et. al, Plant Biotechnology Journal (doi:10.1111/pbi.12778).

[1222] 특정 구현예에서, 조류 세포에 의해 생성된 지질의 양 및/또는 지질의 질의 변형에 연루된 유전자를 특이적으로 변형시키는 것으로 예상된다. 지방산 합성 경로에 연루된 유전자 코딩 효소의 예는, 예를 들어, 아세틸-CoA 카복실라제, 지방산 신타제, 3-케토아실_아실- 캐리어 단백질 신타제 III, 글리세롤-3-인산 디히드로게나제 (G3PDH), 엔오일-아실 캐리어 단백질 리덕타제 (엔오일-ACP-리덕타제), 글리세롤-3-포스페이트 아실트랜스퍼라제, 리소포스파티드 아실 트랜스퍼라제 또는 다이아실글리세롤 아실트랜스퍼라제, 인지질:다이아실글리세롤 아실트랜스퍼라제, 포스파티데이트 포스파타제, 지방산 티오에스터라제, 예컨대, 팔미토일 단백질 티오에스터라제, 또는 말산 효소 활성을 갖는 단백질을 코딩할 수 있다. 추가 구현예에서, 지질 축적이 증가된 규조류를 생성시키는 것이 계획된다. 이는 지질 이화작용을 감소시키는 유전자를 표적화시킴으로써 달성될 수 있다. 특히, 트리아실글리세롤 및 유리 지방산 둘 모두의 활성화에 관여되는 유전자뿐만 아니라 지방산, 예컨대, 아실-CoA 신타제, 3-케토아실-CoA 티올라제, 아실-CoA 옥시다제 활성 및 포스포글루코뮤타제의 β-산화에 직접적으로 관여되는 유전자에 본 발명의 방법에서의 사용에 대한 관심이 있다. 본 명세서에 기재된 CRISPR-C2c1 시스템 및 방법은 그들의 지질 함량을 증가시키기 위해 규조류에서 이러한 유전자를 특이적으로 활성화시키는 데 사용될 수 있다.

[1223] 유기체, 예컨대, 미세조류는 합성 생물학을 위해 널리 사용된다. Stovicek 등 (Metab. Eng. Comm., 2015; 2:13]은 산업적 생산을 위해 강한 균주를 효율적으로 생산하기 위한 산업적 효모, 예를 들어 사카로마이세스 세레비시애의 게놈 편집을 기재한다. Stovicek은 내생성 유전자의 대립유전자와 이종성 유전자에서의 녹아웃을 둘 모두를 동시에 붕괴시키는 효모에 대해 코돈-최적화된 CRISPR-Cas9 시스템을 사용하였다. Cas9 및 gRNA는 게놈 또는 에피좀 2μ-기반 벡터 위치로부터 발현되었따. 저자는 또한, 유전자 파괴 효율이 Cas9 및 gRNA 발현 수준의 최적화에 의해 개선될 수 있음을 나타내었다. Hlavovaa 등 (Biotechnol. Adv. 2015)은 삽입 돌연변이유발 및 스크리닝을 위해 핵 및 엽록체 유전자를 표적화하는 CRISPR와 같은 기법을 사용하여 미세조류의 종 또는 균주의 발생을 논의한다. Stovicek 및 Hlavovaa의 방법은 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다. CRISPR-C2c1 시스템과 관련하여, 일부 구현에에서, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'의 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 Lol p5 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1224] US 8,945,839호는 Cas9를 이용하여 미세조류 (클라미도모나스 레인하티(Chlamydomonas reinhardtii) 세포) 종을 조작하는 방법을 설명한다. 유사한 도구를 이용하여, 본 명세서에 기재된 CRISPR-C2c1 시스템의 방법은

클라미도모나스 (Chlamydomonas) 종 및 다른 조류에 적용될 수 있다. 특정 구현예에서, C2c1 및 가이드 RNA는, Hsp70A-Rbc S2 또는 베타2 -튜불린과 같은 구조 프로모터의 제어 하에서 C2c1을 발현하는 벡터를 사용하여 발현된 조류 내에 도입된다. 가이드 RNA는 T7 프로모터를 함유하는 벡터를 사용하여 전달될 것이다. 대안적으로, C2c1 mRNA 및 시험관내 전사된 가이드 RNA는 조류 세포에 전달될 수 있다. 전기천공법 프로토콜은 GeneArt 클라미도모나스 조작 키트로부터의 표준 권고 프로토콜을 따른다. US 8,945,839의 방법은 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다. CRISPR-C2c1 시스템과 관련하여, 일부 구현에에서, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'의 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1225] **지방산 생산이 가능한 미생물의 생성에서 C2c1의 용도**

[1226] 특정 구현예에서, 본 발명의 방법은 지방 에스테르, 예컨대, 지방산 메틸 에스테르 ("FAME") 및 지방산 에틸 에스테르 ("FAEE")를 생산할 수 있는 유전자 조작된 미생물의 생성을 위해 사용된다.

[1227] 전형적으로, 숙주 세포는 티오에스터라제를 코딩하는 유전자, 아실-CoA 신타제를 코딩하는 유전자, 및 에스테르 신타제를 코딩하는 유전자의 발현 또는 과발현에 의해 배지에 존재하는 탄소 공급원, 예컨대, 알콜로부터의 지방 에스터를 생산하도록 조작될 수 있다. 따라서, 본 명세서에 제공된 방법은 티오에스터라제 유전자, 아실-CoA 신타제를 코딩하는 유전자, 및 에스테르 신타제를 코딩하는 유전자를 과발현시키거나 또는 도입하기 위해 미생물을 변형하는 데 사용된다. 특정 구현예에서, 티오에스터라제 유전자는 tesA, 'tesA, tesB, fatB, fatB2, fatB3, fatA1 또는 fatA로부터 선택된다. 특정 구현예에서, 아실-CoA 신타제를 코딩하는 유전자는 fadDJadK, BH3103, pfl-4354, EAV15023, fadD1, fadD2, RPC_4074, fadDD35, fadDD22, faa39, 또는 동일한 특성을 갖는 효소를 코딩하는 동정된 유전자로부터 선택된다. 특정 구현예에서, 에스테르 신타제를 코딩하는 유전자는 심몬드시아 키넨시스 (Simmondsia chinensis), 아시네토박터 종 (Acinetobacter sp.) ADP, 알카니보락스 보르쿠멘시스 (Alcanivorax borkumensis), 슈도모나스 애루지노사 (Pseudomonas aeruginosa), 펀디박터 자덴시스 (Fundibacter jadensis), 아라비돕시스 탈리아나 또는 알칼리게네스 유트로푸스 (Alkaligenes eutrophus) 유래의 신타제/아실-CoA:다이아실글리세릴 아실트랜스퍼라제 또는 이의 변이체를 코딩하는 유전자이다. 추가적으로 또는 대안적으로, 본 명세서에 제공된 방법은 상기 미생물에서 아실-CoA 탈수소효소를 코딩하는 유전자, 외막 단백질 수용체를 코딩하는 유전자, 및 지방산 생합성의 전사 조절제를 코딩하는 유전자 중 적어도 하나의 발현을 감소시키도록 사용된다. 특정 구현예에서, 이들 유전자 중 하나 이상은, 예컨대, 돌연변이의 도입에 의해 비활성화된다. 특정 구현예에서, 아실-CoA 탈수소효소를 코딩하는 유전자는 fadE이다. 특정 구현예에서, 지방산 생합성의 전사 조절자를 코딩하는 유전자는 DNA 전사 억제인자, 예를 들어, fabR을 암호화한다.

[1228] 추가적으로 또는 대안적으로, 상기 미생물은 파이루베이트 포메이트 리아제를 코딩하는 유전자, 락테이트 탈수소효소를 코딩하는 유전자, 또는 둘 다 중 적어도 하나의 발현을 감소시키도록 변형된다. 특정 구현예에서, 파이루베이트 포메이트 리아제를 코딩하는 유전자는 pflB이다. 특정 구현예에서, 락테이트 탈수소효소를 코딩하는 유전자는 IdhA이다. 특정 구현예에서, 이들 유전자 중 하나 이상은, 예컨대, 그에 대한 돌연변이의 도입에 의해 비활성화된다. 일부 구현예에서, CRISPR-C2c1 시스템은 5'-오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1229] 특정 구현예에서, 미생물은 에스케리키아(Escherichia), 바실러스, 락토바실러스, 로도코커스(Rhodococcus), 시네코코커스(Synechococcus), 시네코이스티스(Synechoystis), 슈도모나스, 아스퍼질러스(Aspergillus), 트리코더마(Trichoderma), 뉴로스포라(Neurospora), 푸사리움(Fusarium), 휴미콜라(Humicola), 리조무코르(Rhizomucor), 클루이베로마이세스(Kluyveromyces), 피키아(Pichia), 무코르(Mucor), 마이실리오프토라(Myceliophtora), 페니실리움(Penicillium), 파네로카에테(Phanerochaete), 플레우로투스(Pleurotus), 트라메

테스(Trametes), 크리소스포리움(Chrysosporium), 사카로마이세스(Saccharomyces), 스테노트로파모나스(Stenotrophamonas), 쉬조사카로마이세스(Schizosaccharomyces), 야로위아(Yarrowia) 또는 스트렙토마이세스(Streptomyces) 속으로부터 선택된다.

[1230] **유기산 생산을 할 수 있는 미생물의 생성에서 C2c1의 용도**

[1231] 본 명세서에 제공된 방법은 유기산을, 보다 특히, 펜토스 또는 헥소스 당으로부터 생산할 수 있는 미생물을 조작하기 위해 추가로 사용된다. 특정 구현예에서, 상기 방법은 미생물에 외생성 LDH 유전자를 도입하는 단계를 포함한다. 특정 구현예에서, 상기 미생물에서의 유기산 생산은 관심 유기산 이외의 대사산물을 생성하는 내생성 대사 경로에 관여된 단백질을 코딩하는 내생성 유전자를 불활성화함으로써 추가적으로 또는 대안적으로 증가되고/되거나, 여기서 내생성 대사 경로는 유기산을 소비한다. 특정 구현예에서, 변형은 관심 유기산 이외의 대사물질의 생성이 감소된다는 것을 보장한다. 특정 구현예에 따르면, 상기 방법은 적어도 하나의 조작된 유전자 결실 및/또는 유기산이 소비되는 내생성 경로의 불활성화 또는 관심 유기산 이외의 대사물질을 생성하는 내생성 경로에 관여된 생성물을 코딩하는 유전자를 도입하기 위해 사용된다. 특정 구현예에서, 적어도 하나의 조작된 유전자 결실 또는 불활성화는 피루베이트 데카복실라제 (pdc), 푸마레이트 리덕타제, 알콜 디히드로게나제 (adh), 아세트알데히드 디히드로게나제, 포스포엔올피루베이트 카복실라제(ppc), D-락테이트 디히드로게나제 (d-ldh), L-락테이트 디히드로게나제 (l-ldh), 락테이트 2-모노옥시게나제로 이루어진 군으로부터 선택된 효소를 코딩하는 하나 이상의 유전자에 존재한다. 추가 구현예에서, 적어도 하나의 조작된 유전자 결실 및/또는 불활성화는 피루베이트 데카복실라제(pdc)를 코딩하는 내생성 유전자에 존재한다.

[1232] 추가 구현예에서, 미생물은 락트산을 생성하도록 조작되고, 적어도 하나의 조작된 유전자 결실 및/또는 불활성화는 락테이트 디히드로게나제를 코딩하는 내생성 유전자에 존재한다. 추가적으로 또는 대안적으로, 미생물은 시토크롬-의존적 락테이트 디히드로게나제, 예컨대, 시토크롬 B2-의존적 L-락테이트 디히드로게나제를 코딩하는 내생성 유전자의 적어도 하나의 조작된 유전자 결실 또는 불활성화를 포함한다.

[1233] **효모 균주를 활용하여 개선된 자일로스 또는 셀로비오스의 재생에서 C2c1의 용도**

[1234] 특정 구현예에서, CRISPR-C2c1 시스템은 효모 균주를 이용하는 개선된 자일로스 또는 셀로비오스를 선택하기 위해 적용될 수 있다. 오류 유발 PCR은 자일로스 이용 또는 셀로비오스 이용 경로에 관여되는 하나 (이상의) 유전자를 증폭시키는 데 사용될 수 있다. 자일로스 활용 경로 및 셀로비오스 활용 경로에 관련된 유전자들의 예로는, 이로 제한됨이 없이, 문헌 [Ha, S.J., et al. (2011) Proc. Natl. Acad. Sci. USA 108(2):5049 및 Galazka, J.M., et al. (2010) Science 330(6000):846]에 기재된 것들을 포함한다. 이러한 선택 유전자 내에 무작위 돌연변이를 각각 포함하는, 이중 가닥 DNA 분자의 최종 라이브러리는 효모 균주 (예를 들어, S288C)를 CRISPR-C2c1 시스템의 성분과 공동 형질전환시킬 수 있고, WO2015138855에 기재된 바와 같은 증강된 자일로스 또는 셀로비오스 이용 능력을 갖는 균주가 선택될 수 있다.

[1235] **이소프레노이드 생합성에서의 이용을 위한 개선된 효모 균주의 생성에서 C2c1의 용도**

[1236] Tadas Jakociunas 등은 빵 효모인 사카로마이세스 세레비시에에서 하나의 형질전환 단계에서 5 개에 달하는 상이한 게놈 유전자좌의 게놈 조적을 위한 멀티플렉스 CRISPR/Cas9 시스템의 성공적인 적용을 설명하였으며 (문헌 [Metabolic Engineering Volume 28, March 2015, Pages 213-222]), 이는 산업적으로 중요한 이소프레노이드 생합성 경로를 위한 열쇠인, 높은 메발로네이트 생산을 갖는 균주를 결과로서 초래한다. 특정 구현예에서, CRISPR-C2c1 시스템은 아이소프레노이드 합성에서 사용하기 위한 추가적인 높은 생성 효모 균주를 동정하기 위해 본 명세서에 기재된 바와 같은 다중복합 게놈 조작 방법에 적용될 수 있다. C2c1 단백질과 관련하여, 일부 구현에에서, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 7-nt 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1237] **효모 균주를 생산하는 락트의 생성에서 C2c1의 용도**

[1238] 다른 구현예에서, 복합다중 CRISPR-C2c1 시스템의 성공적인 적용이 포함된다. Vratislav Stovicek 등 (Metabolic Engineering Communications, Volume 2, December 2015, Pages 13-22)과 유사하게, 개선된 락트산-

생산 균주가 설계될 수 있고, 단일 형질전환 이벤트에서 수득될 수 있다. 특정 구현예에서, CRISPR-C2c1 시스템은 이종성 락테이트 탈수소효소 유전자를 삽입하는 동시에 두 개의 내생성 유전자 PDC1 및 PDC5 유전자 붕괴하는데 사용된다. C2c1 단백질과 관련하여, 일부 구현에에서, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 PDC1 또는 PD5 유전자에 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 PDC1 또는 PDC5 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1239] **식물에서 CRISPR-C2c1 시스템의 추가 적용**

[1240] 특정 구현예에서, CRISPR 시스템, 및 바람직하게는 본 명세서에서 설명된 바와 같은 CRISPR-C2c1 시스템은 유전적 요소 동력학의 시각화에 사용될 수 있다. 예를 들어, CRISPR 이미징은 반복적 또는 비반복적 게놈 서열을 시각화, 텔로미어 길이 변화 및 텔로미어 이동의 보고, 및 세포 주기에 걸친 유전자좌의 동력학을 감독할 수 있다(Chen et al., Cell, 2013). 이들 방법은 또한 식물에 적용될 수 있다.

[1241] CRISPR 시스템, 및 바람직하게는 본 명세서에서 설명된 바와 같은 CRISPR-C2c1 시스템의 기타 응용은 시험관 내 및 생체 내 표적된 유전자 붕괴 양성-선택 스크리닝이다(Malina et al., Genes and Development, 2013). 이들 방법은 또한 식물에 적용될 수 있다.

[1242] 특정 구현예에서, 비활성 C2c1 엔도뉴클레아제의 히스톤-변형 효소와의 융합은 복합체 후생유전자에서의 사용자 변경을 도입할 수 있다 (Rusk et al., Nature Methods, 2014). 이들 방법은 또한 식물에 적용될 수 있다.

[1243] 특정 구현예에서, CRISPR 시스템, 및 바람직하게는 본 명세서에서 설명된 바와 같은 CRISPR-C2c1 시스템은 염색질의 특이적 부분을 정제하고, 연관된 단백질을 확인하기 위하여 사용될 수 있으며, 이에 따라 전사에서 이들의 조절 역할을 해명한다(Waldrip et al., Epigenetics, 2014). 이들 방법은 또한 식물에 적용될 수 있다.

[1244] 특정 구현예에서, 본 발명은, 바이러스 DNA 및 RNA를 모두 절단할 수 있음에 따라, 식물체에서 바이러스 제거를 위한 치료법으로서 사용될 수 있다. 인간 시스템에서 이전 연구들은 단일 가닥 RNA 바이러스, C형 간염 바이러스 (A. Price, et al., Proc. Natl. Acad. Sci, 2015)를 비롯하여, 이중 가닥 DNA 바이러스, B형 간염 바이러스 (V. Ramanan, et al., Sci. Rep, 2015)를 표적화하는데서 CRISPR을 이용한 성공을 입증하였다. 이들 방법은 또한 식물에서 CRISPR-C2c1 시스템을 이용하여 조정될 수 있다.

[1245] 특정 구현예에서, 본 발명은 게놈 복잡성을 변경하는데 사용될 수 있다. 추가의 특정 구현예에서, CRISPR 시스템, 및 바람직하게는 본 명세서에서 설명된 바와 같은 CRISPR-C2c1 시스템은 염색체 수를 붕괴 또는 변경시키는데 사용될 수 있으며, 하나의 부모로부터의 염색체만을 함유할 수 있는 반수체 식물을 재생할 수 있다. 그러한 식물은 염색체 이중복제되도록 유도되어, 단지 동형접합성 대립형질만을 함유하는 이배체 식물로 전환될 수 있다 (Karimi-Ashtiyani et al., PNAS, 2015; Anton et al., Nucleus, 2014). 이들 방법은 또한 식물에 적용될 수 있다.

[1246] 특정 구현예에서, 본 명세서에서 설명된 바와 같은 CRISPR-C2c1 시스템은 자가-절단에 사용될 수 있다. 이러한 구현예에서, C2c1 효소 및 gRNA의 프로모터는 구조적 프로모터일 수 있으며, 제2 gRNA는 동일한 형질전환 카세트 내로 도입되지만, 유도성 프로모터에 의해 제어된다. 이러한 제2 gRNA는 비기능성 C2c1을 생성하기 위하여 C2c1 유전자에서 부위-특이적 절단을 유도하도록 지정될 수 있다. 추가의 특정 구현예에서, 제2 gRNA는 형질전환 카세트의 양 말단 상에서의 절단을 유도하며, 숙주 게놈으로부터의 카세트의 제거를 결과로서 초래한다. 이 시스템은 Cas 효소에 대한 세포 노출의 제어된 기간을 제공하며, 추가로 표적외 편집을 최소화한다. 나아가, CRISPR/Cas 카세트의 양 말단의 절단은 2-대립유전자 돌연변이를 갖는 이식유전자-무함유 T0 식물을 생성하는데 사용될 수 있다 (Cas9에 대하여 기재된 바와 같음, 예를 들어, Moore et al., Nucleic Acids Research, 2014; Schaeffer et al., Plant Science, 2015). Moore 등의 방법은 본 명세서에서 설명된 CRISPR-C2c1 시스템에 적용될 수 있다.

[1247] Sugano 등 (Plant Cell Physiol. 2014 Mar;55(3):475-81. doi: 10.1093/pcp/pcu014. Epub 2014 Jan 18)은 육

상 식물 진화를 연구하기 위한 모델 종으로서 떠오른, 우산이끼 마르찬티아 폴리모파 L(Marchantia polymorpha L.)에서의 표적화된 돌연변이유발에 CRISPR-Cas9의 적용을 보고한다. 마르찬티아 폴리모파의 U6 프로모터가 확인되고, 클로닝되어 gRNA를 발현하였다. gRNA의 표적 서열은 마르찬티아 폴리모파에서 옥신 반응 인자 1(ARF1)을 코딩하는 유전자를 파괴하도록 설계되었다. 아그로박테리움-매개 형질전환을 사용하여, Sugano외 다수는 마르찬티아 폴리모파의 배우체 생성에서 안정적인 돌연변이체를 분리하였다. 생체내에서 CRISPR-Cas9-기반 부위-특이적 돌연변이유발은, 콜리플라워 모자이크 바이러스 35S 또는 마르찬티아 폴리모파 EF1α 프로모터 중 어느 하나를 사용하여 Cas9를 발현하도록 달성되었다. 옥신-내성 표현형을 나타내는 분리된 돌연변이체 개체는 키메라가 아니었다. 나아가, 안정적인 돌연변이체는 T1 식물의 무성 생식에 의해 생산되었다. 복수 arf1 대립유전자는 CRIPSR-Cas9-기반의 표적화된 돌연변이유발을 사용하여 용이하게 수립되었다. Sugano외 다수의 방법은 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다.

[1248] Kabadi 등 (Nucleic Acids Res. 2014 Oct 29;42(19):e147. doi: 10.1093/nar/gku749. Epub 2014 Aug 13)은 편리한 골든 게이트 클로닝 방법에 의하여 벡터 내로 혼입된 독립된 RNA 중합효소 III 프로모터로부터, Cas9 변이체, 리포터 유전자 및 최대 4 개의 sgRNA를 발현시키기 위하여 단일 렌티바이러스 시스템을 개발하였다. 각각의 sgRNA는 효율적으로 발현되었고, 불사화된 일차 인간 세포에서 멀티플렉스 유전자 편집 및 지속된 전사 활성화를 매개할 수 있다. Kabadi 등의 방법은 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다.

[1249] Ling 등 (BMC Plant Biology 2014, 14:327)은 gRNA 뿐만 아니라, pGreen 또는 pCAMBIA 골격을 기반으로 CRISPR-Cas9 이원 벡터를 개발하였다. 이러한 툴키트는 한번의 클로닝 단계로 높은 효율로 하나 이상의 gRNA 및 메이즈-코돈 최적화된 Cas9를 보유한 최종 구성체를 생성시키기 위해 BsaI 이외에 제한 효소를 요구하지 않는다. 툴키트는 메이즈 원형질, 유전자이식 메이즈 계통, 및 유전자이식 아라비돕시스 계통을 사용하여 검증되었으며, 높은 효율 및 특이성을 보이는 것으로 나타났다. 더욱 중요하게는, 이러한 툴키트를 사용하여, 3 개의 아라비돕시스 유전자의 표적화된 돌연변이가 T1 세대의 유전자이식 묘목에서 검출되었다. 또한, 다중-유전자 돌연변이가 그 다음 세대로 유전될 수 있었다. 식물에서 멀티플렉스 게놈 편집을 위한 툴키트로서, (가이드 RNA) 모듈 벡터 세트. Lin 등의 툴박스가 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다.

[1250] CRISPR-C2c1을 통한 표적화된 식물 게놈 편집의 프로토콜은 또한 [Methods in Molecular Biology pp 239-255 10 February 2015] 시리즈의 1284권에 CRISPR-Cas9 시스템에 대해 개시된 것을 기반으로 이용가능하다. 모델 세포 시스템으로서 아라비돕시스 탈리아나 (Arabidopsis thaliana) 및 니코티아나 벤타미아나 (Nicotiana benthamiana) 원형질체를 사용하여 식물 코돈 최적화된 Cas9 (pcoCas9) 매개 게놈 편집을 위한 이중 gRNA의 디자인, 구성, 및 평가를 위한 상세한 절차를 기술한다. 전체 식물에서의 표적화된 게놈 변형의 생성에 CRISPR-Cas9 시스템을 적용하는 전략 또한 논의된다. 그 장에 기재된 프로토콜이 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다.

[1251] 상기 언급된 방법 및 프로토콜에서 C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1252] Ma 등 (Mol Plant. 2015 Aug 3;8(8):1274-84. doi: 10.1016/j.molp.2015.04.007) 단자엽 및 쌍자엽 식물에서 편리하며 고효율 멀티플렉스 게놈 편집을 위해, 식물 코돈 최적화된 Cas9 유전자를 이용한, 강력한 CRISPR-Cas9 벡터 시스템을 보고한다. Ma 등은 골든 게이트 결찰 또는 깁슨 조립에 의한 일 회전의 클로닝에서 바이너리 CRISPR-Cas9 벡터 내로 조립될 수 있는, 다중 sgRNA 발현 카세트를 신속히 생성하기 위한 PCR-기반 절차를 설계하였다. 시스템을 이용하여, Ma외 다수는 평균 돌연변이율 85.4%로, 대개 이중대립유전자 및 동형접합성 상태로, 벼에서 46 개의 표적 부위를 편집하였다. Ma 등은 유전자 패밀리의 다수의(최대 8 개) 멤버, 생합성 경로에서의 다수의 유전자, 또는 단일 유전자에서 다수의 부위의 동시 표적화에 의해 T0 쌀 및 T1 아라비돕시스 식물에서 기능 손실 유전자 돌연변이의 예를 제공한다. Ma 등의 방법은 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM

서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1253] Lowder 등 (Plant Physiol. 2015 Aug 21. pii: pp.00636.2015)은 또한, 멀티플렉스 게놈 편집 및 식물에서 발현된, 침묵화된, 또는 비-코딩 유전자의 전사 조절을 가능하게 하는 CRISPR-Cas9 툴박스를 개발하였다. 이 툴박스는 골든 게이트 및 게이트웨이 클로닝 방법을 사용하여 단자엽 및 쌍자엽에 대한 빠르고 효율적인 조립 작용 CRISPR-Cas9 T-DNA 구성체에 대한 프로토콜 및 시약을 연구자에게 제공한다. 이는 다중화된 유전자 편집 및 식물 내생성 유전자의 전사 활성화 또는 억제를 포함하는, 능력들의 완전한 세트가 함께 제공된다. T-DNA 기반 형질전환 기술은 현대 식물 생명공학, 유전학, 분자 생물학 및 생리학의 기초가 된다. 이와 같이, C2c1 (WT, 닉카제 또는 dC2c1) 및 gRNA(들)는 관심 T-DNA 목적지-벡터 내로의 조립 방법을 개발하였다. 조립 방법은 골든 게이트 조립 및 멀티사이트 게이트웨이 재조합 모두를 기초로 한다. 3 개의 모듈이 조립에 요구된다. 제1 모듈은 C2c1 진입 벡터로, 이는 attL1 및 attR5 부위에 측접된 무-프로모터 C2c1 또는 이의 유도체 유전자를 함유한다. 제2 모듈은 gRNA 진입 벡터로, 이는 attL5 및 attL2 부위에 측접된 진입 gRNA 발현 카세트를 함유한다. 제3 모듈은 C2c1 발현을 위해 선택된 프로모터를 제공하는 attR1-attR2-함유 목적지 T-DNA 벡터를 포함한다. Lower외 다수의 도구 박스는 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1254] Wang 등 (bioRxiv 051342; doi: doi.org/10.1101/051342; Epub. May 12, 2016)은 단일 프로모터의 제어 하에서 몇개 gRNA-tRNA 유닛을 갖는 다중복합 유전자 편집 구성체를 사용해 육배체 밀에서 중요한 농경 형질에 영향을 미치는 4개 유전자의 동종 카피의 편집을 입증하였다.

[1255] 유리한 구현예에서, 식물은 나무일 수 있다. 본 발명은 또한 본 명세서에서 개시된 CRISPR Cas 시스템을 초본경에 이용할 수 있다(예를 들어, 문헌[Belhaj et al., Plant Methods 9:39] 및 [Harrison et al., Genes & Development 28:18591872] 참조). 특히 유리한 구현예에서, 본 발명의 CRISPR Cas 시스템은 나무에서 단일 뉴클레오티드 다형성(SNP)을 표적화할 수 있다(예를 들어, 문헌[Zhou et al., New Phytologist, Volume 208, Issue 2, pages 298301, October 2015] 참조). Zhou 등의 실험에서, 저자는 사례 연구로서 4-쿠마레이트:CoA 리가제 (4CL) 유전자 패밀리를 사용하여 목재질 다년생 포풀러(Populus)에서 CRISPR Cas 시스템을 적용하여 표적화된 2개 4CL 유전자에 대해 100% 돌연변이 효율을 달성하였으며, 조사된 모든 형질전환체가 이중대립유전자 변형을 보유하였다. Zhou 등의 연구에서, CRISPR-Cas9 시스템은 단일 뉴클레오티드 다형성(SNP)에 대해 고도로 민감하였으며, 제3의 4CL 유전자에 대한 절단은 표적 서열 내 SNP로 인해 삭제되었다. 이러한 방법은 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1256] Zhou 등의 방법(문헌[New Phytologist, Volume 208, Issue 2, pages 298301, October 2015])은 본 발명에 다음과 같이 적용될 수 있다. 리그닌 및 플라보노이드 생합성에 연관된, 2 개의 4CL 유전자, 즉 4CL1 및 4CL2는

각각 CRISPR-Cas9 편집에 대해 표적화된다. 형질전환에 일상적으로 사용되는 포플러스 트레뮬라 (Populus tremula) × alba 클론 717-1B4는 게놈-시퀀싱된 포퓰러스 트리코카르파(Populus trichocarpa)로부터 분기된다. 따라서, 기준 게놈으로부터 설계된 4CL1 및 4CL2 gRNA는 인-하우스 717 RNA-Seq 데이터로 질의되어 Cas 효율을 제한할 수 있는 SNP의 부재를 보장한다. 4CL1의 게놈 복제물인, 4CL5에 대해 설계된 제3 gRNA 또한 포함된다. 상응하는 717 서열은 PAM 가까이/내의 각각의 대립형질에서 하나의 SNP를 포함하며, 이들 모두는 4CL5-gRNA에 의한 표적화를 제거하는 것으로 예상된다. 세 개의 gRNA 표적 부위 모두는 제1 엑손 내에 위치된다. 717 형질전환의 경우, gRNA는 바이너리 벡터 내 CaMV 35S 프로모터의 제어 하에, 인간 코돈-최적화된 Cas와 함께 개자리(Medicago) U6.6 프로모터로부터 발현된다. Cas-만의 벡터를 이용한 형질전환은 대조군으로서 제공될 수 있다. 무작위로 선택된 4CL1 및 4CL2 계통은 앰플리콘-시퀀싱 처리된다. 그 데이터는 이후 가공되고, 이중대립형질 돌연변이는 모든 경우에서 확인된다. 이러한 방법은 본 발명의 C2c1 이펙터 단백질 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1257] 식물에서, 병원균은 종종 숙주-특이적이다. 예를 들어, 푸사리움 옥시스포럼 (Fusarium oxysporum) f. sp. 리코페르시시 (lycopersici)는 토마토 시듬을 초래하지만 오직 토마토만을 공격하고, 에프. 옥시스포럼 (F. oxysporum) f. 디안티이 푸시니아 그라미니스 (dianthii Puccinia graminis) f. sp. 트리티시 (tritici)는 오직 밀만을 공격한다. 식물은 대부분의 병원균에 저항하는 기존의 유도된 방어성을 갖는다. 식물 생성에 걸친 돌연변이 및 재조합 사건은 감수성을 일으키는 유전자 가변성을 야기하는데, 특히 병원균이 식물보다 더 빈번하게 재생되기 때문이다. 식물에서는, 비-숙주 저항성이 존재할 수 있으며, 예를 들어 숙주 및 병원균은 비양립성이다. 통상적으로 많은 유전자에 의해 제어되는, 수평 저항성, 예를 들어 병원균의 모든 종족에 대하여 부분 저항성, 및 통상적으로 몇몇 유전자에 의해 제어되는, 수직 저항성, 예를 들어 다른 종족은 아니지만 일부 종족의 병원균에 대해 완전한 저항성이 또한 존재할 수 있다. 유전자 대 유전자 수준에서, 식물 및 병원균은 함께 진화하며, 하나의 균형에서의 유전적 변화는 다른 하나에서 변화를 일으킨다. 따라서, 자연적 변동성을 사용하여, 품종 개량자는, 산출량, 품질, 균일성, 강인성(hardiness), 저항성에 대해 가장 유용한 유전자를 조합한다. 저항성 유전자 공급원은 천연 또는 외래의 변종들, 토종(Heirloom Varieties), 야생 근연 식물, 및 유도된 돌연변이, 예를 들어 돌연변이원성 제제를 이용한 식물 재료 처리를 포함한다. 본 발명을 사용하여, 식물 품종 개량자는 돌연변이를 유도하는 새로운 도구를 제공받는다. 따라서, 당업자는 저항성 유전자의 게놈을 분석할 수 있고, 원하는 특징 또는 특성을 갖는 변종들에서, 이전의 돌연변이원성 제제보다 더욱 정확하고, 이에 따라 식물 품종 개량 프로그램을 가속화하고 개선하는 저항성 유전자의 증가를 유도하기 위하여 본 발명을 이용할 수 있다.

[1258] 하기 표는 CRISPR-Cas 복합체, 변형된 이펙터 단백질, 시스템, 및 최적화 방법이 생물생산을 개선시키는데 사용될 수 있는 관련 분야 및 추가 참조를 제공한다. 일부 구현예에서, CRISPR-Cas 복합체는 tracrRNA, 직접 반복부에 연결된 가이드 서열을 포함하는 가이드 RNA와 복합체를 형성하는 C2c1 단백질 또는 이의 촉매 도메인을 포함하고, 가이드 서열은 표적 서열과 하이브리드화된다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

2-17-2014	PCT/US15/63434 (WO2016/099887)	원형 폴리뉴클레오티드 변형 주형과 조합하여 가이드 RNA/Cas 엔도뉴클레아제 시스템을 사용하는 이. 콜라이에서 효율적인 유전자 편집을 위한 조성물 및 방법.
08-13-2014	PCT/US15/41256 (WO2016/025131)	RNA-가이드된 엔도뉴클레아제를 사용하는 비-통상적 효모에서의 유전자 표적화.
11-06-2014	PCT/US15/58760 (WO2016/073433)	세포로의 RNA-가이드된 엔도뉴클레아제의 펩티드-매개 전달.
10-12-2015	PCT/US16/56404 (WO2017/066175)	유전자 변형 및 세포 내 증가된 상동성 재조합을 위한 보호된 DNA 주형 및 사용 방법.
12-11-2015	PCT/US16/65070 (WO2017/100158)	증강된 뉴클레아제-매개 게놈 변형 및 감소된 오프-표적 부위 효과를 위한 방법 및 조성물.
12-18-2015	PCT/US16/65537 (WO 2017/105991)	T-RNA 기반 가이드 RNA 발현을 위한 방법 및 조성물.
12-18-2015	PCT/US16/66772 (WO2017/106414)	중합효소 II (Pol-II) 기반 가이드 RNA 발현을 위한 방법 및 조성물.
12-16-2014	PCT/US15/65693 (WO2016/100272)	진균 게놈 변형 시스템 및 사용 방법.
12-16-2014	PCT/US15/66195 (WO2016/100571)	진균 게놈 변형 시스템 및 사용 방법.
12-16-2014	PCT/US15/66192 (WO 2016/100568)	진균 게놈 변형 시스템 및 사용 방법.
12-16-2014	PCT/US15/66178 (WO 2016/100562)	트리코더마 레에세이 (*Trichoderma reesei*)에서 표적화 DNA 카세트의 상동성 통합을 개선시키기 위한 침묵화된 NHEJ 를 갖는 헬퍼 균주의 용도.
07-28-2015	PCT/US16/44489 (WO 2017/019867)	게놈 편집 시스템 및 사용 방법.

[1259]

[1260] **개선된 식물 및 효모 세포**

[1261] 본 발명은 또한 본 명세서에 제공된 방법에 의해 얻을 수 있고 얻어진 식물 및 효모 세포를 제공한다. 본 명세서에 기재된 방법에 의해 얻어진 개선된 식물은, 예를 들어, 식물 해충, 제초제, 가뭄 또는 저온 또는 고온, 과량의 물 등에 대한 내성을 보장하는 유전자의 발현을 통해 식품 또는 사료 생산에서 유용할 수 있다.

[1262] 본 명세서에 기재된 방법에 의해 수득되는 개선된 식물은, 특히 작물 및 조류이고, 예를 들어, 야생형에서 정상적으로 보이는 것보다 더 높은 단백질, 탄수화물, 영양소 또는 비타민 수준의 발현을 통해 식품 또는 사료 생산에서 유용할 수 있다. 이와 관련하여, 개선된 식물, 특히 두류 및 괴경이 바람직하다.

[1263] 개선된 조류 또는 다른 식물, 예컨대, 유채는, 예를 들어, 식물성 오일 또는 생물연료, 예컨대, 알콜 (특히 메탄올 및 에탄올)의 생산에서 특히 유용할 수 있다. 이들은 오일 또는 생체연료 산업에서 사용을 위한 고수준의 오일 또는 알코올을 발현 또는 과발현시키도록 유전자조작될 수 있다.

[1264] 본 발명은 또한 식물의 개선된 부분을 위해 제공된다. 식물 부분은 이로 제한되지는 않지만, 잎, 줄기, 뿌리, 괴경, 종자, 배젖, 배주, 및 화분을 포함한다. 여기서 고려되는 식물 부분은 자생성, 비자생성, 재생성, 및/또는 비재생성일 수 있다.

[1265] 일 구현예에서, Soyk 등 (Nat Genet. 2017 Jan;49(1):162-168)은 조기 수확 토마토를 생산하기 위해서 토마토의 개화 억제인자 SP5G를 표적화하는 CRISPR-Cas9 매개된 돌연변이를 사용하였고, 본 발명에 개시된 바와 같은 CRISPR-Cas 시스템을 위해 변형될 수 있다. 일부 구현예에서, CRISPR 단백질은 C2c1이고, 시스템은 I. CRISPR-Cas 시스템 RNA 폴리뉴클레오티드 서열로서, (a) 표적 서열과 하이브리드화할 수 있는 가이드 RNA 폴리뉴클레오티드, 및 (b) 직접 반복부 RNA 폴리뉴클레오티드를 포함하는 것인 폴리뉴클레오티드 서열, 및 II. 임의로 적어도 하나 이상의 핵 국재화 서열을 포함하는, C2c1을 코딩하는 폴리뉴클레오티드 서열을 포함하고, 여기서 직접 반복부 서열은 가이드 서열과 하이브리드화하여 표적 서열로 CRISPR 복합체의 서열-특이적 결합을 유도시키고, CRISPR 복합체는 (1) 표적 서열과 하이브리드화하거나 또는 하이브리드화가능한 가이드 서열, 및 (2) 직접 반복부 서열과 복합체를 형성하는 CRISPR 단백질을 포함하고, CRISPR 단백질을 코딩하는 폴리뉴클레오티드

서열은 DNA 또는 RNA이다. 일부 구현예에서, 식물 세포 게놈은 T-풍부 PAM을 포함한다. 특정 구현예에서, PAM은 5'-TTN-3' 또는 5'-ATTN-3' 이다. 특정 구현예에서, PAM은 5'-TTG-3' 이다. 일부 구현예에서, CRISPR 이펙터 단백질은 C2c1 단백질이다. C2c1은 Cas9에 의해 생성되는 PAM의 근위 말단에서의 절단과 대조적으로, PAM의 원위 말단에서 이중 가닥 파손을 생성시킨다 (Jinek et al., 2012; Cong et al., 2013). Cpf1 돌연변이된 표적 서열은 단일 gRNA에 의한 반복된 절단에 감수성일 수 있어서, HDR 매개 게놈 편집에서 Cpf1의 적용을 촉진한다고 제안된다 (Front Plant Sci. 2016 Nov 14;7:1683). Cpf1 및 C2c1은 둘 모두가 구조 유사성을 공유하는 V형 CRISPR 단백질이다. C2c1 처럼, Cpf1은 (PAM의 근위 말단에 블런트 절단부를 생성하는 Cas9와 대조적으로) PAM의 원위 말단에 스태거드 이중 가닥 파손을 생성시킨다. 따라서, 일정 구현에에서, 관심 유전자좌는 상동성 지정 복구 (HR 또는 HDR)을 통해 CRISPR-C2c1 복합체에 의해 변형된다. 일정 구현예에서, 관심 유전자좌는 HR 독립적인 CRISPR-C2c1 복합체에 의해 변형된다. 일정 구현에에서, 관심 유전자좌는 비상동성 말단 결합 (NHEJ)을 통해 CRISPR-C2c1 복합체에 의해 변형된다.

[1266] 또한 본 발명의 방법에 따라 생성된 식물 세포 및 식물을 제공하는 것으로 본 명세서에 포함된다. 전통적인 육종 방법에 의해 생성되는 유전자 변형을 포함하는 식물의 배우체, 종자, 배아, 접합체 또는 체세포 중 하나, 자손 또는 잡종은 본 발명의 범주 내에 포함된다. 이러한 식물은 표적 서열에 또는 표적 서열 대신에 삽입되는 이종성 또는 외래 DNA 서열을 함유할 수 있다. 대안적으로, 이러한 식물은 하나 이상의 뉴클레오티드에서의 변경(돌연변이, 결실, 삽입, 치환)만을 함유할 수 있다. 이렇게 해서, 이러한 식물은 단지 특정 변형의 존재에 의해 그들의 조상 식물과 다를 것이다.

[1267] 따라서, 본 발명은 본 방법 또는 이의 자손에 의해 생성되는 식물, 동물 또는 세포를 제공한다. 자손은 생성된 식물 또는 동물의 클론일 수 있거나 또는 그들의 새끼에 추가적인 바람직한 형질을 이용하기 위해 동일 종의 다른 개체와 교배시킴으로써 유성생식으로부터 초래될 수 있다. 세포는 다세포 유기체, 특히 동물 또는 식물의 경우에 생체내 또는 생체외일 수 있다.

[1268] 매우 다양한 식물, 조류, 진균, 효모 등 및 식물 조류, 진균, 효모 세포 또는 조직 시스템이 본 개시의 핵산 구성체 및 상기 언급한 다양한 형질전환 방법을 사용하여 본 명세서에 기술된 바람직한 생리적 및 작물학적 특징을 위해 조작될 수 있다. 매우 다양한 식물, 조류, 진균, 효모 등 및 식물 조류, 진균, 효모 세포 또는 조직 시스템이 본 개시의 핵산 구성체 및 상기 언급한 다양한 형질전환 방법을 사용하여 본 명세서에 기술된 바람직한 생리적 및 작물학적 특징을 위해 조작될 수 있다.

[1269] 특정 구현예에서, 본 명세서에 기재된 방법은 식물 게놈 내 외래 DNA의 존재를 피하기 위해, CRISPR 성분을 코딩하는 것을 포함하는, 임의의 외래 유전자의 식물 게놈, 조류, 진균, 효모 등으로의 영구적인 도입 없이, 내생성 유전자를 변형시키거나 또는 그들의 발현을 변형시키기 위해 사용된다. 이것은 비유전자이식 식물에 대한 조절 요건이 덜 엄격하기 때문에 흥미로울 수 있다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, 기능성 도메인은 디아미나제, 바람직하게 아데노신 디아미나제를 포함한다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1270] 본 명세서에 제공된 CRISPR 시스템은, 표적된 이중-가닥 또는 단일-가닥 파손을 도입 및/또는 유전자 활성화제 및/또는 억제제 시스템을 도입하는데 사용될 수 있고, 제한없이 유전자 표적화, 유전자 대체, 표적된 돌연변이 유발, 표적된 결실 또는 삽입, 표적된 역전 및/또는 표적된 전위에 사용될 수 있다. 단일 세포에서 다중 변형을 달성하도록 지시된 다중 표적화 RNA의 공동발현에 의해, 다중복합 게놈 변형이 보장될 수 있다. 이 기술은 증강된 영양 품질, 질환에 대해 증가된 내성 및 항생제 및 비항생제 스트레스에 대한 내성, 및 상업적으로 가치있는 식물 제품 또는 이종성 화합물의 증가된 생산을 비롯한 개선된 형질을 갖는 식물의 높은 정확도 조작에 사용될 수 있다.

[1271] 본 명세서에 기재된 방법은 일반적으로 야생형 식물에 비해 하나 이상의 바람직한 형질을 가진다는 점에서 "개선된 식물, 조류, 진균, 효모 등" 의 생성을 초래한다. 특정 구현예에서, 얻어진 식물, 조류, 진균, 효모 등, 세포 또는 부분은 식물 세포의 전부 또는 일부의 게놈에 혼합된 외생성 DNA 서열을 포함하는 유전자이식 식물이다. 특정 구현예에서, 비-유전자이식 유전자 변형 식물, 조류, 진균, 효모 등, 부분 또는 세포가 얻어지며, 즉, 외생성 DNA 서열 중 어떤 것도 식물의 임의 세포의 게놈에 혼입되지 않는다. 이러한 구현예에서, 개선된 식물, 조류, 진균, 효모 등은 비-유전자이식이다. 내생성 유전자의 변형만이 보장되고, 외래 유전자 중 어떤 것도 식물, 조류, 진균, 효모 등의 게놈에 도입되거나 또는 유지되지 않는 경우에, 얻어진 유전자 변형 작물은 외래 유전자를 함유하지 않으며, 따라서 기본적으로 비유전자이식으로 간주될 수 있다. 식물, 조류,

진균, 효모 등 게놈 편집을 위한 CRISPR-C21 시스템의 상이한 적용은 제한없이, 관심 농업적 형질을 부여하는 하나 이상의 외래 유전자의 도입; 관심 농업적 형질을 부여하는 내생성 유전자의 편집; 관심 농업적 형질을 부여하도록 CRISPR-C2c1 시스템에 의한 내생성 유전자의 조절을 포함한다. C2c1 단백질이 표적 부위에 스태거드 이중 가닥 파손 (DSB)를 생성시키기 때문에, 외생성 DNA 서열은 상동성 지정 복구 (HR) (예를 들어, NHEJ를 통함)에 의해 또는 없이 도입되거나, 또는 녹-인될 수 있다. 작물학적 형질을 부여하는 예시적인 유전자는 해충 또는 병해에 내성을 부여하는 유전자; 식물 질환에 관여되는 유전자, 예컨대, WO 2013046247에 열거된 것; 제초제, 살진균제 등에 대한 내성을 부여하는 유전자; (항생제) 스트레스 내성에 관여되는 유전자를 포함하지만, 이들로 제한되지 않는다. CCRISPR-Cas 시스템의 사용의 다른 양상은 (웅성) 불임 식물의 생성; 식물/조류 등에서 생식 단계의 증가; 관심 작물에서의 유전자 변형의 생성; 과일-숙성에 대한 영향; 식물/조류 등의 저장 수명 증가; 식물/조류 등에서 알레르겐의 감소; 형질(예를 들어, 영양적 개선)이 더해진 가치의 보장; 관심 내생성 유전자에 대한 스크리닝 방법; 생물연료, 지방산, 유기산 등의 생성을 포함하지만, 이들로 제한되지 않는다.

[1272] C2c1 이펙터 단백질 복합체는 비동물 유기체, 예컨디 식물, 조류, 진균, 효모 등에서 사용될 수 있다.

[1273] 본 명세서에 기술된 바와 같은 C2c1 시스템을 사용한 게놈 편집을 위한 방법은 본질적으로 임의의 식물, 조류, 진균, 효모 등에 바람직한 형질을 부여하기 위해 사용될 수 있다. 매우 다양한 식물, 조류, 진균, 효모 등 및 식물 조류, 진균, 효모 세포 또는 조직 시스템이 본 개시의 핵산 구성체 및 상기 언급한 다양한 형질전환 방법을 사용하여 본 명세서에 기술된 바람직한 생리적 및 작물학적 특징을 위해 조작될 수 있다.

[1274] 갈변 방지 흰단추 버섯 (아리쿠스 비스포러스 (Aaricus bisporus)) 품종은 PEG 형질전환을 통해 버섯 세포로 전달된, 가이드 RNA 및 Cas9 단백질을 포함하는 CRISPR-Cas9 시스템으로 폴리페놀 옥시다제 유전자에 1-14 nt 결실을 도입하여 개발되었다. 참조: Yang et al. (news.psu.edu/story/432734/2016/10/19/academics/penn-state-developer-gene-edited-mushroom-wins-best-whats-new). 본 명세서에 기술된 바와 같은 CRISPR-C2c1 시스템은 Yang 등의 방법으로 사용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부의 PAM 서열을 인식한다. 일부 구현예에서, PAM은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, C2c1 이펙터 단백질은 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, C2c1 이펙터 단백질은 닉카제이다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 관심 유전자좌를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, 본 명세서에 기재된 방법은 식물 게놈 내 외래 DNA의 존재를 피하기 위해, CRISPR 성분을 코딩하는 것을 포함하는, 임의의 외래 유전자의 식물 게놈, 조류, 진균, 효모 등으로의 영구적인 도입 없이, 내생성 유전자를 변형시키거나 또는 그들의 발현을 변형시키기 위해 사용된다. 이것은 비유전자이식 식물에 대한 조절 요건이 덜 엄격하기 때문에 흥미로울 수 있다.

[1275] 본 명세서에 제공된 CRISPR 시스템은, 표적된 이중-가닥 또는 단일-가닥 파손을 도입 및/또는 유전자 활성화제 및/또는 억제제 시스템을 도입하는데 사용될 수 있고, 제한없이 유전자 표적화, 유전자 대체, 표적된 돌연변이 유발, 표적된 결실 또는 삽입, 표적된 역전 및/또는 표적된 전위에 사용될 수 있다. 단일 세포에서 다중 변형을 달성하도록 지시된 다중 표적화 RNA의 공동발현에 의해, 다중복합 게놈 변형이 보장될 수 있다. 이 기술은 증강된 영양 품질, 질환에 대해 증가된 내성 및 항생제 및 비항생제 스트레스에 대한 내성, 및 상업적으로 가치있는 식물 제품 또는 이종성 화합물의 증가된 생산을 비롯한 개선된 형질을 갖는 식물의 높은 정확도 조작에 사용될 수 있다.

[1276] 본 명세서에 기재된 방법은 일반적으로 야생형 식물에 비해 하나 이상의 바람직한 형질을 가진다는 점에서 "개선된 식물, 조류, 진균, 효모 등" 의 생성을 초래한다. 특정 구현예에서, 얻어진 식물, 조류, 진균, 효모 등, 세포 또는 부분은 식물 세포의 전부 또는 일부의 게놈에 혼합된 외생성 DNA 서열을 포함하는 유전자이식 식물이다. 특정 구현예에서, 비-유전자이식 유전자 변형 식물, 조류, 진균, 효모 등, 부분 또는 세포가 얻어지며, 즉, 외생성 DNA 서열 중 어떤 것도 식물의 임의 세포의 게놈에 혼입되지 않는다. 이러한 구현예에서, 개선된 식물, 조류, 진균, 효모 등은 비-유전자이식이다. 내생성 유전자의 변형만이 보장되고, 외래 유전자 중 어떤 것도 식물, 조류, 진균, 효모 등의 게놈에 도입되거나 또는 유지되지 않는 경우에, 얻어진 유전자 변형 작물은 외래 유전자를 함유하지 않으며, 따라서 기본적으로 비유전자이식으로 간주될 수 있다. 식물, 조류, 진균, 효모 등 게놈 편집을 위한 CRISPR-C21 시스템의 상이한 적용은 제한없이, 관심 농업적 형질을 부여하는

하나 이상의 외래 유전자의 도입; 관심 농업적 형질을 부여하는 내생성 유전자의 편집; 관심 농업적 형질을 부여하도록 CRISPR-C2c1 시스템에 의한 내생성 유전자의 조절을 포함한다. 작물학적 형질을 부여하는 예시적인 유전자는 해충 또는 병해에 내성을 부여하는 유전자; 식물 질환에 관여되는 유전자, 예컨대, WO 2013046247에 열거된 것; 제초제, 살진균제 등에 대한 내성을 부여하는 유전자; (항생제) 스트레스 내성에 관여되는 유전자를 포함하지만, 이들로 제한되지 않는다. CCRISPR-Cas 시스템의 사용의 다른 양상은 (웅성) 불임 식물의 생성; 식물/조류 등에서 생식 단계의 증가; 관심 작물에서의 유전자 변형의 생성; 과일-숙성에 대한 영향; 식물/조류 등의 저장 수명 증가; 식물/조류 등에서 알레르겐의 감소; 형질(예를 들어, 영양적 개선)이 더해진 가치의 보장; 관심 내생성 유전자에 대한 스크리닝 방법; 생물연료, 지방산, 유기산 등의 생성을 포함하지만, 이들로 제한되지 않는다.

[1277] **비인간 동물에서 적용**

[1278] 일 양상에서, 본 발명은 임의의 기재된 구현예에 따라 진핵생물 숙주 세포를 포함하는, 비-인간 진핵생물 유기체; 바람직하게는 다세포 진핵생물 유기체를 제공한다. 다른 양상에서, 본 발명은진핵 유기체; 바람직하게는 기재된 구현예 중 어느 것에 따른 진핵 숙주 세포를 포함하는 다세포 진핵 유기체를 제공한다. 이들 양상의 일부 구현예에서 유기체는 동물; 예를 들어, 포유류일 수 있다. 또한, 유기체는 절지동물, 예컨대 곤충일 수 있다. 본 발명은 또한 다른 농업 적용분야, 예를 들어, 농장 및 생산 동물로 확대될 수 있다. 예를 들어, 돼지는 그들을 생의학적 모델, 특히 재생 의학으로서 매력있는 많은 특징을 갖는다. 특히, 중증 합병성 면역결핍 (SCID)을 갖는 돼지는 재생의학, 이종장기이식 (또한 본 명세서의 다른 곳에서 논의됨) 및 종양 발생에 유용한 모델을 제공할 수 있으며, 인간 SCID 환자를 위한 요법 개발에 도움을 줄 수 있다. Lee 등 (Proc Natl Acad Sci U S A. 2014 May 20;111(20):72605)은 대립유전자 둘 모두에 영향을 미치는 일부를 포함하여, 체세포에서 고효율로 재조합 활성 유전자 (RAG) 2의 표적화된 변형을 생성시키기 위해 리포터-가이드된 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN) 시스템을 이용하였다. C2c1 이펙터 단백질은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1279] Lee 등 (Proc Natl Acad Sci U S A. 2014 May 20;111(20):72605)의 방법은 다음과 유사하게 본 발명에 적용될 수 있다. 돌연변이된 돼지는 태아 섬유아세포에서 RAG2의 표적화 변형에 후속하여 SCNT 및 배아 전달을 통해 생산된다. CRISPR Cas 및 리포터를 코딩하는 구성체는 태아-유래 섬유아세포에 전기천공된다. 48시간 후에, 녹색 형광 단백질을 발현시키는 형질감염 세포는 웰당 단일 세포의 추정 희석으로 96-웰 플레이트의 개개 웰로 분류된다. RAG2의 표적화 변형은 임의의 CRISPR Cas 절단 부위가 측접된 게놈 DNA 단편을 증폭시킨 후에 PCR 산물을 시퀀싱하여 스크리닝된다. 스크리닝 및 오프-부위 (off-site) 돌연변이 확인 후, RAG2의 표적화 변형을 보유한 세포를 SCNT에 사용한다. 난모세포의 인접한 세포질의 일부분과 함께, 아마도 중기 II 플레이트를 함유하는 극체는 제거되고, 도너 세포가 위란강에 위치된다. 재구축된 배아는 이후에 전기적으로 천공되어 도너 세포를 난모세포와 융합시키고, 화학적으로 활성화된다. 활성화된 배아는 14시간 내지 16시간 동안 0.5 μM 스크립타이드 (S7817; Sigma-Aldrich)가 존재하는 돼지 접합자 배지 3 (PZM3)에서 인큐베이션된다. 그 다음으로 배아는 스크립타이드를 제거하기 위해 세척되고 그들이 대리모 돼지의 난관에 전달될 때까지 PZM3에서 배양된다. C2c1 이펙터 단백질은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1280] 본 발명은 동물, 일부 구현예에서 포유류, 일부 구현예에서 인간의 질환 또는 장애를 모델링하기 위한 플랫폼을

생성하는 데 사용된다. 일정 구현예에서, 이러한 모델 및 플랫폼은, 비제한적 예에서 래트 또는 마우스에 기반한 설치류이다. 이러한 모델 및 플랫폼은 근친교배 설치류 균주 중에서의 차이 그리고 이들 간의 비교를 이용할 수 있다. 일정 구현예에서, 이러한 모델 및 플랫폼은 영장류, 말, 소, 양, 염소, 돼지, 개, 고양이 또는 조류를, 예를 들어, 이러한 동물의 질환 및 장애를 직접적으로 모델링하기 위해 또는 이러한 동물의 변형되고/되거나 개선된 계통을 생성하기 위해 이용한다. 유사하게는, 일정 구현예에서, 동물 기반 플랫폼 또는 모델은 인간 질환 또는 장애를 모방하기 위해 생성된다. 예를 들어, 돼지와 인간의 유사성은 돼지를 인간 질환을 모델링하기 위한 이상적인 플랫폼으로 만든다. 설치류 모델에 비해, 돼지 모델의 개발은 비용이 들며 시간 집약적이었다. 반면에, 돼지 및 다른 동물은 유전적으로, 해부학적으로, 생리적으로 그리고 병리생리학적으로 인간과 훨씬 더 유사하다. 본 발명은 이러한 동물 플랫폼 및 모델에서 사용될 표적화된 유전자 및 게놈 편집, 유전자 및 게놈 변형 및 유전자 및 게놈 조절을 위한 고효율 플랫폼을 제공한다. 윤리적 표준이 인간 모델 그리고 다수의 사례에서 비인간 영장류에 기반한 모델의 개발을 차단하지만, 본 발명은 세포 배양 시스템, 3차원 모델 및 시스템, 및 모방을 위한 오가노이드(organoids)를 포함하지만, 이들로 제한되지 않는 시험관내 시스템과 함께 사용되고, 인간의 구조, 기관 및 시스템의 유전자, 해부학, 생리학 및 병리생리학을 모델링하고 연구한다. 플랫폼 및 모델은 단일 및 다중 표적의 조작을 제공한다.

[1281] 소정의 실시형태에서, 본 발명은 Schomberg 등 (FASEB Journal, April 2016; 30(1):Suppl 571.1)]과 유사한 질환 모델에 적용 가능하다. 유전성 질환 섬유종증 1형(NF-1) 숌버그(Schomberg)의 모델링은 돼지 배아 내로 CRISPR/Cas9 성분의 사이토졸 미량주사법에 의해 돼지 뉴로피브로민 1 유전자에서 돌연변이를 도입하기 위해 CRISPR-Cas9를 사용하였다. Cas9에 의한 표적화된 절단을 위해 유전자 내의 엑손의 상류와 하류 둘 다의 부위를 표적화하는 영역에 대해 CRISPR 가이드 RNA(gRNA)가 생성되었고, 2500 bp 결실을 도입하기 위해 특정 단일-가닥 올리고데옥시 뉴클레오티드(ssODN) 주형에 의해 수선이 매개되었다. CRISPR-Cas 시스템은 또한 특정 NF-1 돌연변이 또는 돌연변이의 클러스터를 갖는 돼지를 조작하기 위해 사용되었고, 주어진 인간 개체에 특이적이거나 또는 주어진 인간 개체를 나타내는 돌연변이를 조작하는데 사용될 수 있다. 본 발명은 인간 다유전자 질환의 돼지 모델을 포함하지만, 이들로 제한되지 않는 동물 모델을 개발하기 위해 유사하게 사용된다. 본 발명에 따르면, 하나의 유전자에서 또는 다중 유전자에서의 다중 유전자 좌위는 다중복합 가이드 및 선택적으로 하나 또는 다중 주형을 이용하여 동시에 표적화된다.

[1282] 본 발명은 또한 다른 동물, 예컨대, 소의 SNP를 변형시키는 것에 적용 가능하다. Tan 등 (Proc Natl Acad Sci U S A. 2013 Oct 8; 110(41): 16526- 16531)은 플라스미드, rAAV 및 올리고뉴클레오티드 주형을 이용하여 전사 활성인자-유사(TAL) 이펙터 뉴클레아제(TALEN)- 및 주기적 간격으로 분포하는, 짧은 회문구조 반복부(CRISPR)/Cas9-자극된 상동성-지정 복구 (HDR)를 포함하도록 가축 유전자 편집 툴박스 (toolbox)를 확장시켰다.

유전자 특이적 gRNA 서열은 그들 방법에 따라서 Church lab gRNA 벡터 (Addgene ID: 41824)에 클로닝하였다 (Mali P, et al. (2013) RNA-Guided Human Genome Engineering via Cas9. Science 339(6121):823-826). Cas9 뉴클레아제는 RCIScript-hCas9로부터 합성된 mRNAh 또는 Cas9 플라스미드 (Addgene ID: 41815)의 공동형질감염에 의해 제공되었다. 이러한 RCIScript-hCas9는 hCas9 플라스미드 (hCas9 cDNA를 포함)로부터의 XbaI-AgeI 단편을 RCIScript 플라스미드로 서브클로닝하여 구축되었다. C2c1 이펙터 단백질은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 일부 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물의 SNP 위치에서 단일 뉴클레오티드 변형을 도입시킨다.

[1283] Heo 등 (Stem Cells Dev. 2015 Feb 1;24(3):393-402. doi: 10.1089/scd.2014.0278. Epub 2014 Nov 3)은 소 다능성 세포 및 주기적 간격으로 분포하는, 짧은 회문구조 반복부(CRISPR)/Cas9 뉴클레아제를 이용하는 소 게놈에서의 고도로 효율적인 유전자 표적화를 보고하였다. 첫째로, Heo 등은 야마나카 인자의 이소성 발현 및 GSK3β 및 MEK 억제제(2i) 처리에 의해 소 체세포 섬유아세포로부터 유도만능줄기세포(iPSC)를 생성한다. Heo 등은 이들 소 iPSC가 기형종에서의 유전자 발현 및 발생 잠재력에 관해 미경험 다능성 줄기 세포와 고도로 유사하다는 것을 관찰하였다. 게다가, 소 NANOG 유전자좌에 특이적인 CRISPR-Cas9 뉴클레아제는 소 iPSC 및 배아에서 소 게놈의 고도로 효율적인 편집을 나타내었다. C2c1 이펙터 단백질은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수

있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 NANOG 유전자좌 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 NANOG 유전자좌를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 NANOG 유전자좌에 상응하는 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1284] Igenity®는 소와 같은 동물의 프로파일 분석을 제공하여, 경제적으로 중요한 경제적 특성들의 특성, 예컨대 도체 조성, 도체 품질, 모계 및 번식 특성, 및 평균 일일 체중증가를 수행 및 전송한다. 종합적 Igenity® 프로파일의 분석은 DNA 마커 (가장 흔하게는 단일 뉴클레오티드 다형성 또는 SNP)의 발굴에서 시작된다. Igenity® 프로파일 배후의 모든 마커는 대학교, 연구 기관 및 USDA와 같은 정부 기관을 포함한, 연구 시설에서 독립 과학자에 의해 발굴되었다. 이어서, 마커는 검증 개체군에서 Igenity®로 분석된다. Igenity®는, 흔하게 이용가능하지 않은 표현형을 수집하기 위해 종축, 소-송아지, 가축사육장 및/또는 소고기 산업 포장 부문의 산업 파트너와 종종 함께 작업하여, 다양한 생산 환경 및 생물학적 유형을 대표하는 다수 자원 집단을 이용한다. 소 게놈 데이터베이스는 널리 이용 가능하며, 예를 들어, NAGRP 소 게놈 협응 프로그램 (Cattle Genome Coordination Program) (www.animalgenome.org/cattle/maps/db.html)을 참조한다. 따라서, 본 발명은 소 SNP를 표적화하는 데 적용될 수 있다. 당업자는 SNP를 표적화하기 위해 상기 프로토콜을 이용할 수 있으며, 그들에, 예를 들어, Tan 등 또는 Heo 등이 기재한 바와 같이 소 SNP를 적용할 수 있다.

[1285] Qingjian Zou 등 (Journal of Molecular Cell Biology Advance Access published October 12, 2015)은 개 미오스타틴 (MSTN) 유전자 (골격근량의 음성 조절자)의 제1 엑손을 표적화함으로써 개에서의 증가된 근육량을 입증하였다. 먼저, sgRNA의 효율은 개 배아 섬유아세포 (CEF)로 Cas9 벡터와 MSTN 표적화 sgRNA의 공동형질감염을 이용하여, 입증되었다. 이후에, Cas9 mRNA와 MSTN sgRNA 혼합물에 의한 정상 형태를 갖는 배아의 미량주사 및 동일한 암컷 개의 수란관 내로 접합체의 자가이식에 의해 MSTN KO 개가 생성되었다. 녹아웃 강아지는 야생형 한배새끼 자매에 비해 넓적다리에 대해 분명한 근육 표현형을 나타내었다. 이는 또한 본 명세서에 제공된 CRISPR-C2c1 시스템을 이용하여 수행될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1286] **가축**

[1287] 가축에서 바이러스 표적은, 일부 구현예에서, 예를 들어 돼지 마크로파지상의 돼지 CD163을 포함할 수 있다. CD163은 PRRSv (돼지 생식기 호흡기 증후군 바이러스 (Porcine Reproductive and Respiratory Syndrome virus, arterivirus)에 의한 감염과 관련된다 (바이러스 세포 진입을 통한 것으로 여겨짐). 특히 돼지 폐포 마크로파지 (폐에 존재)의 PRRSv에 의한 감염은 사육 돼지에서의 번식 장애, 체중 손실 및 높은 폐사율을 포함하여, 고통을 초래하는 사전 치유불가의 돼지 증후군 ("미확인 돼지 질환" 또는 "청색귀 질환")을 야기한다. 기회감염, 예컨대, 유행성 폐렴, 뇌수막염 및 귀 부종은 종종 마크로파지 활성 상실을 통한 면역 결핍에 기인하는 것으로 보인다. 또한 증가된 항생제 사용 및 재정 손실 (연간 $660m로 추정)로 인해 상당한 경제적 및 환경적 영향을 받는다.

[1288] 미주리 유니버시티에서 그리고 제너스 plc(Genus Plc)과의 공동작업으로 크리스틴 엠 휘트워스(Kristin M Whitworth) 및 랜달 프라터 박사(Dr Randall Prather) 등(Nature Biotech 3434 published online 07 December 2015)에 의해 보고된 바와 같이, CD163은 CRISPR-Cas9를 이용하여 표적화되었고 편집된 돼지의 새끼는 PRRSv에 노출되었을 때 내성이 있었다. 모두 CD163의 엑손 7에서 돌연변이를 갖는 한 마리의 파운더 (founder) 수컷과 한 마리의 파운더 암컷을 새끼를 생산하도록 육종되었다. 파운더 수컷은 하나의 대립유전자 상의 엑손 7에서 11-bp 결실을 가졌는데, 이는 프레임 시프트 돌연변이 및 도메인 5 내 아미노산 45에서의 미스센스 번역 및 아미노산 64에서 후속적인 조기 정지 코돈을 초래한다. 다른 대립유전자는 선행 인트론에서의 엑손 7 및 377-bp 결실에서 2bp 첨가를 갖는데, 이는 도메인 5의 처음 49개 아미노산의 발현을 초래하고, 그 다음에 아미

노산 85에서 조기 정지 코돈을 초래하는 것으로 예측되었다. 암퇘지는 번역되었을 때 도메인 5의 처음 48개 아미노산, 아미노산 70에서 조기 정지 코돈을 발현시키는 하나의 대립유전자에서 7 bp 첨가를 가진다. 암퇘지의 다른 대립유전자는 증식 가능하지 않다. 선택된 새끼는 비대립(null) 동물(CD163-/-), 즉, CD163 녹아웃인 것으로 예측되었다.

[1289] 따라서, 일부 구현예에서, 돼지 파리 대식세포는 CRISPR 단백질에 의해 표적화될 수 있다. 일부 구현예에서, 돼지 CD163은 CRISPR 단백질에 의해 표적화될 수 있다. 일부 구현예에서, 돼지 CD163은 DSB의 유도를 통해, 또는 삽입 또는 결실, 예를 들어, 상기 기재한 것 중 하나 이상을 포함하는, 엑손 7의 표적화 결실 또는 변형, 또는 유전자의 다른 영역, 예를 들어 엑손 5의 결실 또는 변형을 통해 녹아웃될 수 있다.

[1290] 편집된 돼지 및 그의 자손, 예를 들어 CD163 녹아웃 돼지가 또한 고려된다. 이는 가축, 육종 또는 모델링 목적 (즉, 돼지 모델)을 위한 것일 수 있다. 유전자 녹아웃을 포함하는 정액이 또한 제공된다.

[1291] CD163은 스캐빈저 수용체 시스테인-풍부 (SRCR) 슈퍼패밀리의 멤버이다. 단백질의 SRCR 도메인 5 시험관내 연구는 바이러스 게놈의 패키징 해체 (unpackagin) 및 방출을 초래하는 도메인에 기반한다. 이와 같이, SRCR 슈퍼패밀리의 다른 멤버은 또한 다른 바이러스에 대한 내성을 평가하기 위해 표적화될 수 있다. PRRSV는 또한 또한 쥐과 락테이트 디히드로게나제-상승 바이러스, 유인원 출혈열 바이러스 및 말 동맥염 바이러스를 포함하는 포유동물 아테리바이러스 그룹의 멤버이다. 아테리바이러스는 마크로파지 향성 및 중증 질환 및 지속적 감염을 유발하는 능력을 포함하는, 중요한 병원성 특성을 공유한다. 따라서, 아테리바이러스, 및 특히 쥐과 락테이트 디히드로게나제-상승 바이러스, 유인원 출혈열 바이러스 및 말 동맥염 바이러스는, 예를 들어, 돼지 CD163 또는 다른 종의 이의 상동체를 통해 표적화될 수 있고, 및 쥐과, 유인원 및 말 모델 및 녹아웃이 또한 제공된다.

[1292] 사실, 이 접근은 인간에 전염될 수 있는 다른 가축 질환을 야기하는 바이러스 또는 박테리아, 예컨대, 인플루엔자 C 및 H1N1, H1N2, H2N1, H3N1, H3N2 및 H2N3으로서 알려진 인플루엔자 A의 아형을 포함하는 돼지 인플루엔자 바이러스(SIV) 균주뿐만 아니라 상기 언급된 폐렴, 뇌수막염 및 부종에까지 확대될 수 있다.

[1293] C2c1 이펙터 단백질은 상기 기술된 바와 같은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부의 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 CD163 유전자좌에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 CD163을 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 가축의 게놈을 변형시키지 않고 CD163 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1294] UCP1 (언커플링 단백질 1)은 내부 미토콘드리아 막 상에 위치하고 내막을 가로지르는 양자 수송으로부터 ATP 합성을 탈결합시켜서 열을 발생시킨다. UCP1은 비오한 열생산의 핵심 요소이고 신체 지방과다의 조절에서 아마도 중요한 듯 하다. 기능성 UCP1 유전자가 결여된 돼지 (Artiodactyl family Suidae)는 부실한 체온조절을 겪어서 추위에 취약하다. 돼지의 지방 축적은 또한 그들의 UCP1 결여와 연관될 수 있으므로, 돼지 생산 효율에 영향을 미칠 수 있다. Zheng 등은 돼지 내생성 UCP1 유 전자좌로 마우스 아디포넥틴-UCP1을 효율적으로 삽입시키기 위해 CRISPR/Cas9-매개, 상동성 재조합 (HR)-독립적 접근법의 적용을 보고하였다

[1295] UCP1 녹-인 (KI) 돼지는 극심한 저온 노출 동안 체온을 유지하는 개선된 능력을 보였지만, 신체 활동 수준 또는 총 일일 에너지 소비 (DEE)에 변경은 없었다. 백색 지방 조직 (WAT)에서 이소성 UCP1 발현은 지방 축적을 4.89% ($P < 0.01$) 까지 극적으로 감소시켜서, 결과적으로 도체 살코기 비율이 증가된다 (CLP; $P < 0.05$). 기전 연구는 WAT에서 UCP1 활성화 시 지방 손실이 상승된 지방분해와 연관되었다는 것을 의미하였다. 본 발명에서 개시된 CRISPR-C2c1 시스템은 Zheng 등이 기술한 바와 같은 유사한 시스템에 적용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 도입시킨다. 일부 구현예에서, 5' 오버행은 7 nt 이다. 특정 구현예에서, CRISPR-C2c1 시스템은 HR 또는 HR 독립 기전, 예컨대 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시키는데 사용될 수 있다.

[1296] Niu 등 (DOI: 10.1126/science.aan4187)은 CRISPR-Cas9 시스템으로 체세포 핵 전달을 통해 돼지 내생성 레트로바이러스 (PERV) 불활성화된 돼지 가축을 보고하였다. 이종이식은 인간 이식을 위한 장기 부족을 완화시키기 위한 유망한 전략이다. 돼지에게 무해한 돼지 내생성 레트로바이러스 (PERV)의 종간 전달의 주요 위험 하나는 인간에게 치명적일 수 있다는 것이다. CRISPR-Cas9를 사용한 불멸화 돼지 세포주에서 PERV 활성의 불활성화 및 체세포 핵 전달을 통한 PERV-불활성화된 돼지의 생성이 기술되어 있다. Wu 등 (Scientific Reports 7, Article number: 10487 (2017) doi:10.1038/s41598-017-08596-5)은 키메라-적능 인간 다능성 줄기세포와 조합했을 때 돼지의 장기 및 인간 조직의 이종-생성을 위한 적합한 플랫폼으로서 제공될 수 있는, PDX1 유전자를 표적화하는 이중 sgRNA 및 Cas9 mRNA의 접합체 동시-전달을 통한 돼지 배아에서 췌장형성의 효율적인 불능화를 보고하였다. Zhou 등 (Hum Mutat 37:110-118, 2016)은 80%로 높은 효율로 주형으로서 단일 가닥 DNA를 사용하여 돼지 접합자에서 CRISPR-Cas9 유도된 HDR을 통한 정밀한 오솔로그 인간 돌연변이 (Sox 10 c.A325>T)를 보유한 유전자-변형된 돼지를 보고하였다. 본 명세서에 개시된 바와 같은 CRISPR-C2c1 시스템은 돼지 가축을 생산하기 위해서 Niu 등, Wu 등, Zhou 등이 기술한 바와 유사한 시스템에 적용될 수 있다. 일정 구현예에서, CRISPR-C2c1 시스템은 바이러스 내성 관련 유전자를 변형시킨다. 부정 구현예에서, CRISPR-C2c1 시스템은 질환 관련 유전자를 변형시킨다. 일정 구현예에서, CRISPR-C2c1 시스템은 가축 생물량 관련 유전자를 변형시킨다. 일정 구현예에서, CRISPR-C2c1 시스템은 가축 형질 관련 유전자를 변형시킨다. 특정 구현예에서, 형질 관련 유전자는 지방 과다의 조절에 관여된다. 일부 구현예에서, 형질 관련 유전자는 특별한 단백질의 발현 조절에 관여되고, 이러한 단백질은 음식 알레르기와 관련된다. 특정 구현예에서, CRISPR-C2c1 시스템은 UCP1 유전자좌를 변형시킨다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부의 PAM 서열을 인식한다. 일부 구현예에서, PAM은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 관심 유전자좌를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 가축의 게놈을 변형시키지 않고 관심 표적 유전자좌의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1297] Gao 등 (Genome Biology 201718:13, doi:10.1186/s13059-016-1144-4)은 소의 선택된 유전자좌에서 유전자 삽입을 유도하기 위해 단일 CRISPR-Cas9 닉카제 (Cas9n)을 사용한 결핵 (TB) 내성 소 가축을 보고하였다. 촉매적 불활성 Cas9 단백질의 주요 결합 부위는 염색질 면역침전 시퀀싱 (ChIP-seq)를 사용하여 소 태아 섬유아세포 (BFF)에서 결정하였다. 후속하여, CRISPR-Cas9n-유도된 단일-가닥 파손을 사용하여 자연 내성-연관 마크로파지 단백질-1 (NRAMP1) 유전자의 삽입을 자극시켰다. TB 내성 가축은 체세포 핵 전달을 통해 수득되었다.

Carlson 등 (Nat Biotechnol. 2016 May 6;34(5):479-81. doi: 10.1038/nbt.3560)은 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN)을 사용해 소 배아 섬유아세포의 게놈에 POLLED 유전자의 대립유전자를 삽입시키고 나서 체세포 핵 전달을 후속하여 유전자 조작된 세포주를 클로닝하고 수령체 소에게 배아를 이식한 무뿔 젖소 가축을 보고하였다. 본 명세서에 개시된 바와 같은 CRISPR-C2c1 시스템은 소 가축 생산에서 Gao 등 및 Carlson 등에서 기술된 바와 유사한 시스템에 적용될 수 있다. 일정 구현예에서, CRISPR-C2c1 시스템은 바이러스 내성 관련 유전자를 변형시킨다. 부정 구현예에서, CRISPR-C2c1 시스템은 질환 관련 유전자를 변형시킨다. 일정 구현예에서, CRISPR-C2c1 시스템은 가축 생물량 관련 유전자를 변형시킨다. 일정 구현예에서, CRISPR-C2c1 시스템은 가축 형질 관련 유전자를 변형시킨다. 일부 구현예에서, 형질 관련 유전자는 특별한 단백질의 발현 조절에 관여되고, 이러한 단백질은 음식 알레르기와 관련된다.특정 구현예에서, 형질 관련 유전자는 지방 과다의 조절에 관여된다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부의 PAM 서열을 인식한다. 일부 구현예에서, PAM은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, C2c1 이펙터 단백질은 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, C2c1 이펙터 단백질은 닉카제이다.

일부 구현예에서, CRISPR-C2c1 시스템은 표적 관심 유전자좌를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 가축의 게놈을 변형시키지 않고 관심 표적 유전자좌의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1298] 2종의 닭 유전자 오브알부민 (OVA) 및 오보뮤코이드 (OVM)은 계란 흰자 알레르기와 관련된 것으로 확인되었다. OVA 및 OVM의 유전자 파괴는 계란에서 낮은 알레르기원성을 생성시킬 잠재성을 가져서, 계란 흰자-함유 식제품 및 백신같은 항목에 민감한 개체에서 면역 반응을 감소시킨다. Oishi 등 (Scientific Reports 6, Article number: 23980 (2016) doi:10.1038/srep23980)은 닭에서 CRISPR/Cas9-매개 유전자 표적화를 보고하였다. 2종의 계란 흰자 유전자, 오브알부민 및 오보뮤코이드는 Cas9, 단일 가이드 RNA, 및 약물 내성을 코딩하는 유전자를 코딩하는 원형 플라스미드의 형질감염에 이어서 일시적 항생제 선별에 의해 배양된 닭 원시 생식세포 (PGC)에서 효율적으로 (> 90%) 돌연변이유발되었다. CRISPR-유도된 돌연변이체-오보뮤코이드 PGC는 수령체 닭 배아로 이식되었고 3종의 배선 키메라 수탉 (G0)을 확립하였다. 모든 수탁은 도너-유래 돌연변이체-오보뮤코이드 정자를 가졌고 높은 전파율의 도너-유래 생식체를 갖는 둘은 다음 세대 (G1)에서 그들 도너-유래 자손의 약 절반만큼 이형접합 돌연변이체 오보뮤코이드 닭을 생산하였다. 오보뮤코이드 동형접합 돌연변이체 자손 (G2)는 G1 돌연변이체 닭을 교배하여 생성시켰다.

[1299] 전통적인 조류의 트랜스제네시스 방법은 배반엽의 레트로바이러스 감염 또는 원시 생식 세포 (PGC)의 생체외 조작에 이어 수령체 배아로 세포의 역 주입을 포함한다. 포유동물 시스템과 달리, 조류 배아 PGC는 그들이 정자 또는 난자 생산 세포가 되는 생식샘으로의 그들 경로 상의 혈관구조를 통한 이동을 겪는다. Tyack 등 (Transgenic Res. 2013 Dec;22(6):1257-64. doi: 10.1007/s11248-013-9727-2)은 트랜스포존에 운반되는 리포터 유전자를 발현시키는 생체내 생성 유전자이식 자손에서 PGC를 안정하게 형질전환시키기 위해 Tol2 트랜스포존 및 트랜스포사제 플라스미드와 복합체 형성된 리포펙타민 2000을 사용한 PGC의 형질전환 방법을 기술하였다. 본 명세서에 개시된 바와 같은 CRISPR-C2c1 시스템은 가금류 가축 생산에서 Oishi 등에서 기술된 바와 유사한 시스템에 적용될 수 있다. 일정 구현예에서, CRISPR-C2c1 시스템은 바이러스 내성 관련 유전자를 변형시킨다. 부정 구현예에서, CRISPR-C2c1 시스템은 질환 관련 유전자를 변형시킨다. 일정 구현예에서, CRISPR-C2c1 시스템은 가축 생물량 관련 유전자를 변형시킨다. 일정 구현예에서, CRISPR-C2c1 시스템은 가축 형질 관련 유전자를 변형시킨다. 일부 구현예에서, 형질 관련 유전자는 특별한 단백질의 발현 조절에 관여되고, 이러한 단백질은 음식 알레르기와 관련된다. 특정 구현예에서, 형질 관련 유전자는 지방 과다의 조절에 관여된다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부의 PAM 서열을 인식한다. 일부 구현예에서, PAM은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, C2c1 이펙터 단백질은 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, C2c1 이펙터 단백질은 닉카제이다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 관심 유전자좌를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 가축의 게놈을 변형시키지 않고 관심 표적 유전자좌의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1300] **동물 모델**

[1301] 본 발명은 생체내, 생체외 및 시험관내에서 동물 모델 및 세포 모델을 개발하는데 사용할 수 있는 CRISPR-Cas 시스템을 제공한다.

[1302] Niu 등 (Cell. 2014 Feb 13;156(4):836-43. doi: 10.1016/j.cell.2014.01.027)은 인간 질환을 연구하고 치료 전략을 개발하기 위해서 중요한 모델 종으로서 제공될 수 있는 원숭이 모델을 개발하였지만, 생물의학 연구에서 원숭이의 적용은 CRISPR/Cas9 시스템을 적용하여 바람직한 표적 부위에서 유전자 변형된 동물을 생성시키는데 어려움으로 인해 상당히 방해받았다. 시스템은 1 단계에서 2개 표적 유전자 (Ppar-γ 및 Rag1)의 동시 파괴를 가능하게 하고, 오프-표적 돌연변이유발이 종합 분석을 통해 검출되지 않았다.

[1303] Wang 등 (Cell. 2013;153(4):910-8)은 Cas9 mRNA 및 sgRNA의 수정된 접합체로의 직접 주사를 사용하여 단일 (95%) 또는 이중 돌연변이체 (70-80%) 마우스를 생산하기 위해 고효율로 배아 줄기 세포 (ESC) 형질감염 모델의 생산을 기술하였다. 마우스 접합체 세포에서 다양한 마우스 모델은 하기 문헌에 기술되었다: yen et al, Dev Biol. 2014;393(1):3-9, Aida et al. Biol. 2015;16(1):87, Inui et al., Sci Rep. 2014;4:5396, Yang et al. Cell. 2013;154(6):1370-9. 줄기 세포의 생체외 변형을 포함하는 이식 질환 모델은 예를 들어, Eμ-Myc 림프종에서 p53을 표적화하는 sgRNA를 사용해, 배선 돌연변이를 생성시키기 위한 대안을 제공한다. 참조: Heckl et al, Nat Biotechnol. 2014;32(9):941-946, Chen et al. Cell. 2015;160(6):1246-1260.

[1304] 일 양상에서, 본 발명은 특히 1형 신경섬유종증 (NF1) 신경유전 병태에 의해 유도되는 중추신경계의 종양의 치료를 제공한다. NF1을 갖는 개체는 NF1 유전자에 배선 돌연변이를 갖고 태어나지만, 자폐증 및 주의력 결핍부터 뇌 및 말초 신경초 종양의 범위에 걸쳐서, 수많은 개별 신경학적 문제가 발생될 수 있다. 본 발명은 환자-특이적 질환 모델을 개발하고 동질유전자 배경에서 유도 만능 줄기 세포 (iPSC)-유래 질환 관련 세포를 연구하기 위해 사용될 수 있다. 유도 만능 줄기 세포 또는 iPSC로도 알려진, 배아 줄기 세포 (ESC)-유사 세포는 성인 환자의 피부 또는 혈액 세포로부터 생성될 수 있다. 최근의 연구 노력은 NF1 환자에 영향을 미치는 중추 및 말초 신경계 (CNS 및 PNS)에서 다양한 세포 유형으로 iPSC를 분화시키는 배양 프로토콜을 개발하기 위해 시작되었다. 본 발명의 CRISPR C2c1 시스템은 현존 돌연변이체 유전자를 복구하거나 또는 새로운 돌연변이를 창출하여 특이적 질환 유전자를 유전자 편집하는데 사용될 수 있다. NF1 연구의 최전선에 자리잡기 위해서, 국립 아동 의료 센터 (Children's National Medical Center)의 길버트 가족 신경섬유종증 연구소 (Gilbert Family Neurofibromatosis Institute) (GFNI)는 이들 최근의 흥미로운 연구 개발을 조사하고, 환자-특이적 인간 NF1 질환 모델을 체계적으로 개발하고, 개별 NF 환자에 대한 약물 스크리닝 및 평가를 위한 도구를 제공하는 것이 중요할 것이다.

[1305] 일 양상에서, 본 발명은 유도가능한 질화나 모델을 개발하는 방바ㅓㅂ을 제공한다.

[1306] Platt 등 (Cell. 2014;159(2):440-55.) 은 Cre-의존적 CAGs-LSL-Cas9 녹인 이식유전자를 개발하였지만, '올-인-원', 독소시클린 (dox)-유도성 구성체가 동물의 배선에서 sgRNA 및 Cas9 둘 모두를 제공하기 위해 생성되었다. Cre-의존적 모델은 현행 Cre-구동 시스템으로 CRISPR-매개 표적화의 단순 도입을 가능하게 하고, 강력한 CAG 프로모터의 하류에서 강력하고 광범위한 Cas9 발현을 제공한다. dox-유도성 모델은 개체 또는 다수 조직에서 표적화를 가능하게 하고, 외생성 sgRNA를 전달하는 능력에 의해 제한되지 않으며, 유전자 변형 후 Cas9 발현을 없애는 수단을 제공한다. 양쪽 접근법은 다수 조직에서 단일 또는 다중 유전자 변형의 극도로 높은 효율을 보이고, 전통적인 유전자 녹아웃에서 보이느 표현형 결과를 요약한다. 각각은 외생적으로 또는 동물의 배선을 통해서 sgRNA의 전달에 적합하지만, 중요하게, 게놈 내 Cas9의 안정한 통합은 크기 제한적 바이러스 카세트로 거대 Cas9 cDNA 패키징의 복잡성을 피한다.

[1307] CRISPR/Cas9 게놈 편집에 의한 인간 질환의 단일-뉴클레오티드 다형성 (SNP) 모델의 창출은 이제 설치류에서 일반적이다. 이들 모델은 인간 유전학에 기능적 통찰력을 이끌어주고 잠재적인 새로운 요법의 개발을 허용한다. 예를 들어, 인간 GWAS는 인간에서 혈소판 분비의 조절인자인, STXBP5 유전자에서 잠재적 병리성 SNP (rs1039084 A;>G)를 확인하였다. 이 돌연변이를 거의 동일한 혈전증 표현형을 갖는 마우스에서 CRISPR을 통해 재현하여 인간에서 이 SNP의 인과관계를 확인하였다 (Zhu et al. Arterioscler Thromb Vasc Biol. 2017;37:264-270). 유사하게, 전체-게놈 시퀀싱을 사용하여 에스토니아의 개체군-기반 바이오뱅크에서 GWAS를 수행하였다. 수많은 잠재적 인과적 변이체 및 기초 기전을 확인하였다. 그들 중 하나는 호염기구 생산에 필수적인 조절 요소로서, 전사 인자 CEBPA의 발현을 조절하는 이러한 과정 동안 특이적으로 작용한다. 이러한 인핸서는 조혈 줄기 및 전구 세포에서 CRISPR/Cas9에 의해 교란되어서 호염기구 분화 동안 CEBPA 발현을 특이적으로 조절한다는 것을 입증한다 (Guo et al. Proc Natl Acad Sci. 2016;114:E327-E336. doi: 10.1073/pnas.1619052114).

[1308] 본 명세서에 개시된 CRISPR-C2c1 시스템은 Zhu 등, Niu 등, 및 Wang 등에 기술된 바와 같은 교란 및 파괴 시스템에 기술된 방법과 사용될 수 있다. 일부 실시형태에서, 동물 모델은 비-인간 진핵생물 세포이다. 일부 실시형태에서, 동물 모델은 비인간 포유동물 세포를 포함한다. 일부 실시형태에서, 동물 모델은 영장류 세포를 포함한다. 일정 구현예예서, 동물 모델은 어류, 제브라 피쉬, 유인원, 침팬지, 마카크 원숭이, 마우스, 토끼, 래트, 개, 소, 양, 염소 또는 돼지 세포를 포함한다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부의 PAM 서열을 인식한다. 일부 구현예에서, PAM은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, C2c1 이펙터 단백질은 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, C2c1 이펙터 단백질은 닉카제이다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 관심 유전자좌를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 가축의 게놈을 변형시키지 않고 관심 표적 유전자좌의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1309] **치료적 응용**

[1310] 명백할 것이지만, 본 시스템은 임의의 관심 폴리뉴클레오티드 서열을 표적화하는데 사용될 수 있다. 본 발명은 비-천연 발생 또는 조작된 조성물, 또는 상기 조성물의 성분을 코딩하는 하나 이상의 폴리뉴클레오티드, 또는 생체내, 생체외 또는 시험관내 표적 세포를 변형하는 데 사용하기 위한 상기 조성물 성분을 코딩하는 하나 이상의 폴리뉴클레오티드의 벡터 또는 전달 시스템을 제공하고, 일단 변형되면, CRISPR 변형된 세포의 자손 또는 세포주는 변형된 표현형을 보유하도록 세포를 변경시키는 방식으로 수행될 수 있다. 변형된 세포 및 자손은 다중 세포 유기체, 예컨대, 목적하는 세포 유형에 대한 CRISPR 시스템의 생체외 또는 생체내 적용을 갖는 식물 또는 동물의 일부분일 수 있다. CRISPR 발명은 치료적 처치 방법일 수 있다. 치료적 처리 방법은 유전자 또는 게놈 편집, 또는 유전자 치료법을 포함할 수 있다.

[1311] **박테리아, 균류 및 기생 병원균과 같은 병원균의 치료**

[1312] 본 발명은 박테리아, 균류 및 기생 병원균을 처리하는데 적용될 수도 있다. 대부분의 연구 노력은 새로운 항생제의 개발에 집중하며, 일단 개발되면, 그럼에도 불구하고 이는 약물 저항성의 동일한 문제의 대상이 될 것이다. 본 발명은 그러한 어려움들을 극복하는 신규 CRISPR-기반 대안을 제공한다. 나아가, 기존의 항생제와 달리, CRISPR-기반 치료는 병원균 특이적으로 제조될 수 있어서, 유익한 박테리아는 회피하면서 표적 병원균의 박테리아 세포 사멸을 유도한다.

[1313] Jiang외 다수(문헌["RNA-guided editing of bacterial genomes using CRISPR-Cas systems", Nature Biotechnology vol. 31, p. 2339, March 2013])는, CRISPR-Cas9 시스템을 사용하여 S. 뉴모니아에 및 에스케리키아 콜라이를 돌연변이 또는 사멸시킨다. 정확한 돌연변이를 게놈 내로 도입한 이 작업은 돌연변이되지 않은 세포를 사멸시키기 위하여 표적화된 게놈 부위에서의 이중-RNA:Cas9-유도된 절단에 의존하였으며, 선별 마커 또는 대항-선택 시스템에 대한 필요를 피하였다. CRISPR 시스템은 항생제 저항성을 역전하고, 균주들 간의 저항성의 전달을 제거하는 데 사용되었다. Bickard 등은, 독성 유전자를 표적하도록 재프로그램화된 Cas9가 독성인 S. 아우레우스를 사멸시키지만, 비독성 S. 아우레우스는 사멸시키지 않았음을 보였다. 항생제 저항성 유전자를 표적화하는 뉴클레아제의 재프로그램화는, 항생제 저항성 유전자를 품은 스태필로코커스 플라스미드를 파괴하고, 플라스미드-유래 저항성 유전자의 확산에 대해 면역화되었다. (참조: Bikard et al., "Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials," Nature Biotechnology vol. 32, 1146-1150, doi:10.1038/nbt.3043, published online 05 October 2014.) Bikardss는, CRISPR-Cas9 항균제가 마우스 피부 콜로니화 모델에서 S. 아우레우스를 사멸시키도록 생체 내에서 작용함을 보였다. 유사하게, Yosef외 다수는 CRISPR 시스템을 사용하여 β-락탐 항생제에 대한 저항성을 수여하는 효소를 인코딩하는 유전자를 표적화하였다 (see Yousef et al., "Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria," Proc. Natl. Acad. Sci. USA, vol. 112, p. 7267-7272, doi: 10.1073/pnas.1500107112 published online May 18, 2015).

[1314] CRISPR 시스템은 기타 다른 유전학적 접근에 대해 저항성인 기생충의 게놈을 편집하는 데 사용될 수 있다. 예를 들어, CRISPR-Cas9 시스템은 플라스모듐 요엘리(Plasmodium yoelii) 게놈 내로 이중-가닥 파손을 도입하는 것을 나타내었다(참조: Zhang et al., "Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System," mBio. vol. 5, e01414-14, Jul-Aug 2014). Ghorbal 등 (" 'Genome editing in the human malaria parasite Plasmodium falciparumusing the CRISPR-Cas9 system," Nature Biotechnology, vol. 32, p. 819-821, doi: 10.1038/nbt.2925, published online June 1, 2014)은 각각 아르테미시닌에 대한 내성 출현 및 유전자 침묵화에서 추정 역할을 갖는, 2개 유전자, orc1 및 kelch13의 서열을 변형시켰다. 적절한 부위에서 변경된 기생충은 변형에 대한 직접 선택은 없음에도 불구하고 매우 높은 효율로 회수되어, 중립성 또는 심지어는 유해한 돌연변이가 이 시스템을 사용하여 생성될 수 있음을 나타낸다. CRISPR-Cas9는 또한 톡소플라스마 곤디를 포함한, 다른 병원성 기생충의 게놈을 변형시키는데 사용된다 (참조: Shen et al., "Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9," mBio vol. 5:e01114-14, 2014; and Sidik et al., "Efficient Genome Engineering of Toxoplasma gondii Using CRISPR/Cas9," PLoS One vol. 9, e100450, doi: 10.1371/journal.pone.0100450, published online June 27, 2014).

[1315] Vyas 등 ("A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families," Science Advances, vol. 1, e1500248, DOI: 10.1126/sciadv.1500248, April 3, 2015) 은 CRISPR 시스템을 이용하여 C. 알비칸스에서의 유전적 조작에 대한 오랜 장애를 극복하였으며, 몇몇 상이한 유전자의 두

개의 카피 모두를 단일 실험에서 효율적으로 돌연변이시켰다. 몇몇 메카니즘이 약물 저항성에 기여하는 유기체에서, Vyas는 부모의 임상적 분리물 Can90에 의해 나타나는 플루코나졸 또는 시클로헥시미드에 대한 초저항성을 더 이상 나타내지 않는 동형접합성 이중 돌연변이체를 생산하였다. Vyas는 조건적 대립형질을 만듬으로써 C. 알비칸스의 필수 유전자에서의 동형접합성 기능 손실 돌연변이를 또한 수득하였다. 리보솜 RNA 가공에 대해 요구되는 DCR1의 널 대립형질은 저온에서 치명적이지만 고온에서는 생존성이다. Vyas는 논센스 돌연변이를 도입하고, 16℃에서 성장하는 데 실패한 dcr1/dcr1 돌연변이체를 분리한 복구 주형을 사용하였다.

[1316] 염색체 유전자좌의 파괴에 의해 피.팔시파럼에서 사용을 위한 본 발명의 CRISPR 시스템. Ghorbal 등 (문헌 ["Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system" , Nature Biotechnology, 32, 819821 (2014), DOI:10.1038/nbt.2925, 2014년 6월 1일)은 CRISPR 시스템을 사용하여 특정 유전자 녹아웃 및 단일-뉴클레오티드 치환을 말라리아 게놈에서 도입하였다. 피. 팔시파럼에 CRISPR-Cas9 시스템을 적합화시키기 위해서, Ghorbal 등은 DSM1에 대한 내성을 부여하는, 약물-선별 마커 ydhodh, 피. 팔시파럼 디히드로오로테이트 디히드로게나제 (PfDHODH) 억제제를 보유하고, sgRNA의 전사를 위해, 동일한 플라스미드, pL7 상에서 상동성 재조합 복구를 위한 도너 DNA 주형 및 가이드 RNA를 위치시킨 피. 팔시파럼 U6 소형 핵 (sn)RNA 조절 엘리먼트를 보유하는 pUF1-Cas9 에피솜의 플라스모듐 조절 엘리먼트의 조절 하의 발현 벡터를 생성시킨다. 참조: Zhang C. et al. ("Efficient editing of malaria parasite genome using the CRISPR/Cas9 system", MBio, 2014 Jul 1; 5(4):E01414-14, doi: 10.1128/MbIO.01414-14) 및 Wagner et al. ("Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum, Nature Methods 11, 915-918 (2014), DOI: 10.1038/nmeth.3063).

[1317] 일 양상에서, 본 발명은 A/T 풍부 게놈을 갖는 유기체 예컨대 피. 팔시파럼의 염색체 유전자좌를 파괴하는 방법을 제공한다. 일부 구현예에서, 본 발명의 CRISPR 시스템은 CRISPR-C2c1 시스템을 포함하고, C2c1 단백질은 표적 부위에서 7-nt 스태거드 절단부를 생성시키고, PAM 서열은 T-풍부 서열이다 (Gardner et al., Nature. 2002;419:531-534). 당업자는 Jiang 등, Bikard 등, Yosef 등, Vyas 등, Ghober 등, Zhang 등 및 Wagner 등에 기술된 방법을 A/T 풍부 게놈에서 서열 파괴를 도입시키기 위해 본 명세서에 개시된 CRISPR-C2c1 시스템과 사용할 수 있다.

[1318] 일정 구현예에서, 관심 유전자좌는 주형 DNA 서열을 삽입, 또는 "녹-인" 시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 특정 구현예에서, DNA 삽입은 적절한 배향으로 게놈에 통합되도록 설계된다. 바람직한 구현예에서, 관심 유전자좌는 비분열 세포에서 CRISPR-C2c1 시스템에 의해 변형되고, 여기서 상동성 지정 복구 (HDR) 기전을 통한 게놈 편집이 특히 도전적이다 (Chan et al., Nucleic acids research. 2011;39:5955-5966). Maresca 등 (Genome Res. 2013 Mar; 23(3): 539-546)은 징크 핑거 뉴클레아제 (ZFN) 및 Tale 뉴클레아제 (TALEN)를 사용해 적용가능한 부위 지정된, 정밀 삽입 방법을 기술하는데, 여기서 5' 오버행을 갖는 짧은, 이중-가닥 DNA가 상보성 말단에 결찰되어, 인간 세포주에서 한정된 유전자좌에서 15-kb 외생성 발현 카세트의 정밀한 삽입을 가능하게 하였다. He 등 (Nucleic Acids Res. 2016 May 19; 44(9))은 4.6 kb 무프로모터 ires-eGFP 단편을 GAPDH 유전자좌에 CRISPR/Cas9-유도 부위-특이적 녹-인으로 체세포 LO2 세포에서 최대 20% GFP+ 세포를 산출하였고, NHEJ 경로에 의해 매개되는 인간 배아 줄기 세포에서 1.70% GFP+ 세포를 산출하였다고 기술하였으며, 또한 NHEJ-기반 녹-인이 조사된 모든 인간 세포 유형에서 HDR-매개 유전자 표적화에 비해 더 효율적이라고 보고하였다. C2c1이 5' 오버행을 갖는 스태거드 절단을 생성시키기 때문에, 당업자는 본 명세서에 개시된 CRISPR-C2c1 시스템을 사용해 관심 유전자좌에서 외생성 DNA 삽입을 발생시키기 위해서 Meresca 등 및 He 등에 기술된 거소가 유사한 방법을 사용할 수 있다.

[1319] 일정 구현예에서, 관심 유전자좌는 PAM 서열의 원위 말단에서 CRISPR-C2c1 시스템에 의해 먼저 변형되고, PAM 서열 근처에서 CRISPR-C2c1 시스템에 의해 더욱 변형되며 HDR을 통해 복구된다. 일정 구현예에서, 관심 유전자좌는 HDR을 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ를 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 바람직한 구현예에서, 외생성 DNA는 3' 및 5' 말단 둘 모두 상에서 단일 가이드 DNA(sgDNA)-PAM 서열이 측접된다. 바람직한 구현예에서, 외생성 DNA는 CRISPR-C2c1 절단 이후에 방출된다.

[1320] **HIV와 같은 바이러스 병원균같은 병원균의 치료**

[1321] Cas-매개된 게놈 편집은 체세포 조직 내로 보호성 돌연변이를 도입하여 비유전적 또는 복합 질병과 싸우는데 사

용될 수 있다. 예를 들어, 림프구에서 CCR5 수용체의 NHEJ-매개 불활성화 (Lombardo et al., Nat Biotechnol. 2007 Nov; 25(11):1298-306)는 HIV 감염을 회피하기 위한 실용적인 전략일 수 있는 한편, PCSK9의 결필 (Cohen et al., Nat Genet. 2005 Feb; 37(2):161-5) 오란지오포이어틴 (Musunuru et al., N Engl J Med. 2010 Dec 2; 363(23):2220-7)은 스타틴 내성 고콜레스테롤혈증 도는 고지혈증에 대한 치료 효과를 제공할 수 있다. 이들 표적은 siRNA-매개된 단백질 녹아웃을 이용하여 다루어질 수도 있으며, NHEJ-매개된 유전자 비활성화의 독특한 장점은 치료를 지속할 필요 없이 영구 치료 이익을 달성하는 능력이다. 모든 유전자 치료에서처럼, 각각의 제안된 치료적 이용이 양호한 이익-위험 비율을 갖도록 수립되는 것이 물론 중요할 것이다.

[1322] 복구 주형과 함께 가이드 RNA 및 Cas9를 코딩하는 플라스미드 DNA의 티로신혈증의 성체 마우스 모델의 간 내로의 유체역학적 전달은 돌연변이체 Fah 유전자를 수정할 수 있고, 250 개의 세포 중 약 1 개의 야생형 Fah 단백질의 발현을 구제할 수 있는 것으로 나타났다 (문헌 [Nat Biotechnol. 2014 Jun;32(6):5513] 참조). 추가적으로, 임상 시험은 CCR5 수용체의 생체 외 녹아웃에 의해 HIV 감염과 싸우기 위해 ZF 뉴클레아제를 성공적으로 사용하였다. 모든 환자에서, HIV DNA 수준은 감소하였으며, 4 명의 환자 중 한 명에서 HIV RNA는 검출가능하지 않게 되었다 (문헌 [Tebas et al., N Engl J Med. 2014 Mar 6;370(10):90110]에서 이용 가능한 온라인 효모 게놈 데이터베이스가 언급될 수 있다. 이들 둘 모두의 결과는 신규 치료 플랫폼으로서 프로그램가능한 뉴클레아제의 가능성을 증명하였다. C2c1 이펙터 단백질은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1323] 또 다른 구현예에서, HIV tat/rev에 의해 공유된 일반 엑손을 표적화하는 siRNA, 핵소체-국소화 TAR 유인체 (decoy), 및 항-CCR5-특이적 망치머리형 리보자임을 갖는 자가-비활성화 렌티바이러스성 벡터 (예를 들어, 문헌 [DiGiusto et al. (2010) Sci Transl Med 2:36ra43] 참조)가 사용되고/거나 본 발명의 CRISPR-Cas 시스템에 대해 조정될 수 있다. 최소 2.5 × 10^6개의 CD34+ 세포/킬로그램 환자 체중이 수집되고, 2 μmol/L-글루타민, 줄기 세포 인자(100ng/mℓ), Flt-3 리간드(Flt-3L) (100ng/mℓ), 및 트롬보포이에틴(10ng/mℓ)(CellGenix)을 2 × 10^6개의 세포/mℓ의 밀도로 함유하는 X-VIVO 15 배지(Lonza) 배지에서 16 내지 20시간 동안 사전자극하였다. 사전 자극된 세포는 피브로넥틴(25mg/㎠)(레트로넥틴(RetroNectin), 타카라 바이오 인코포레이티드(Takara Bio Inc.))로 코팅한 75-㎠ 조직 배양 플라스크에서 16 내지 24시간 동안 5의 감염 다중도로 렌티바이러스로 형질도입될 수 있다. C2c1 이펙터 단백질은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1324] 해당 기술 분야의 지식 및 본 개시 내용의 교시를 이용하여, 숙련자는 CCR5를 표적화하고 녹아웃하는 CRISPR-C2c1 시스템과 HSC를 접촉시키는 것을 포함하는 HIV/AIDS와 같은 면역결핍 상태에 관련된 HSC를 수정할 수 있다. CCR5-및 C2c1 단백질 함유 입자를 표적화하고 녹아웃시키는 가이드 RNA (및 유리하게 이중 가이드 접근법, 예를 들어 상이한 가이드 RNA의 쌍; 예를 들어, 초대 인간 CD4+ T 세포 및 CD34+ 조혈 줄기 및 선조 세포 (HSSPC)에서, 두 개의 임상적으로 관련된 유전자인 B2M 및 CCR5의 가이드 RNA 표적화)는 HSC와 접촉된다. 이렇게 접촉된 세포를 투여할 수 있고; 임의로 치료 /확장된다; cf. Cartier. 참조: Kiem, "Hematopoietic stem cell-based gene therapy for HIV disease, "Cell Stem Cell. Feb 3, 2012; 10(2): 137-147 (이의 인용 문헌들과 함께 참조로 본 명세서에 편입됨); Mandal et al, "Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9," Cell Stem Cell, Volume 15, Issue 5, p643-652, 6 November 2014 (이의 인용 문헌들과 함께 참조로 본 명세서에 편입됨). CRISPR-C2c1 시스템을 이용하여 HIV/AIDS와 싸우기 위한 또 다른 수단으로서, 문헌 ([Ebina, "CRISPR/Cas9 system to suppress HIV-1

expression by editing HIV-1 integrated proviral DNA" SCIENTIFIC REPORTS | 3 :2510 | DOI:10.1038/srep02510, 그에 언급된 문헌과 함께 본 명세서에 참고로서 포함됨])이 또한 언급될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1325] HIV 치료를 위한 게놈 편집에 대한 근거는, 바이러스에 대한 세포성 공동-수용체인, CCR5에서 기능 손실 돌연변이에 대해 동형접합성인 개인들이 감염에 대해 고도로 저항성이며, 그렇지 않으면 건강하다는 관찰로부터 유래되며, 이는 게놈 편집을 이용한 이러한 돌연변이의 모방은 안전하며 효과적인 치료 전략일 수 있음을 제안한다 (문헌 [Liu, R., et al. Cell 86, 367377 (1996)]). 이러한 아이디어는, HIV 감염된 환자가 CCR5 돌연변이의 기능 손실에 대해 동형접합성인 공여자로부터 동종이계성(allogeneic) 골수 이식을 받은 경우, 검출가능하지 않은 수준의 HIV 및 정상 CD4 T-세포 수의 회복을 결과로서 초래한다는 것을 임상적으로 입증하였다 (문헌 [Hutter, G., et al. The New England journal of medicine 360, 692698 (2009)]). 비용 및 잠재적인 이식편 대 숙주 질병으로 인해, 골수 이식은 대부분의 HIV 환자에게 현실적인 치료 전략은 아니지만, 환자 자신의 T-세포를 CCR5로 전환하는 HIV 치료는 바람직하다.

[1326] HIV의 인간화 마우스 모델에서 CCR5를 녹아웃하기 위한 ZFNs 및 NHEJ를 이용하는 초기 연구는, CCR5 편집된 CD4 T세포의 장기이식이 개선된 바이러스 부하 및 CD4 T-세포 수를 개선하였음을 보였다 (문헌 [Perez, E.E., et al. Nature biotechnology 26, 808816 (2008)]). 중요하게는, 이들 모델은 또한 HIV 감염이 CCR5 널 세포의 선택하는 결과를 초래하였음을 나타내어, 편집이 적합성 장점을 수여함을 제안하며, 적은 수의 편집된 세포들이 치료 효과를 내는 것을 잠재적으로 가능하게 한다.

[1327] 이러한 및 다른 유망한 임상전 연구의 결과로서, 환자 T 세포 내에서 CCR5를 녹아웃하는 게놈 편집 치료는 이제 인간에서 시험되었다 (문헌 [Holt, N., et al. Nature biotechnology 28, 839847 (2010), Li, L., et al. Molecular therapy :the journal of the American Society of Gene Therapy 21, 12591269 (2013)]). 최근 임상 시험 제 I기에서, HIV를 갖는 환자로부터의 CD4+ T 세포는 제거되고, CCR5 유전자를 녹아웃하도록 설계된 ZFN을 이용하여 편집되어, 환자 내로 다시 자가이식되었다 (문헌 [Tebas, P., et al. The New England journal of medicine 370, 901910 (2014)]).

[1328] 또 다른 연구 (문헌 [Mandal et al., Cell Stem Cell, Volume 15, Issue 5, p643-652, 6 November 2014])에서, CRISPR-Cas9는, 인간 CD4+ T 세포 및 CD34+ 조혈성 줄기 및 선조 세포 (HSPCs)에서 두 개의 임상 관련 유전자 B2M 및 CCR5를 표적하였다. 단일 RNA 가이드의 이용은 HSPC에서 매우 효율적인 돌연변이 유발을 이끌었으나 T 세포에서는 그렇지 않았다. 이중 가이드 시도는 두 세포 유형 모두에서 개선된 유전자 결실 효능을 시도하였다. CRISPR-Cas9를 이용하여 게놈 편집 처리된 HSPC는 다혈통(multilineage) 능을 보유하였다. 예측된 표적상 및 표적외 돌연변이는 HSPC에서 표적 포착 시퀀싱을 통해 검사되었으며, 낮은 수준의 표적외 돌연변이유발이 단지 하나의 부위에서 관찰되었다. 이들은 CRISPR-Cas9가 최소의 표적외 돌연변이유발을 이용하여 HSPC에서 유전자를 효율적으로 제거할 수 있음을 증명하며, 이는 조혈 세포-기반 치료에 대한 넓은 적용성을 갖는다.

[1329] Wang et al. (PLoS One. 2014 Dec 26;9(12):e115987 doi: doi:10.1371/journal.pone.0115987])은 Cas9 및 CCR5 가이드 RNA를 발현하는 렌티바이러스 벡터를 이용하여, CRISPR 연관 단백질 9 (Cas9) 및 단일 가이드 RNA (가이드 RNA)를 통해 CCR5를 침묵화하였다. Wang 등은 Cas9 및 CCR5 가이드 RNA를 발현하는 렌티바이러스 벡터의 HIV-1 민감성 인간 CD4+ 세포 내로의 일 회전 형질도입은 높은 빈도의 CCR5 유전자 붕괴를 산출함으로 나타내었다. CCR5 유전자-붕괴된 세포는 전달된/파운더 (T/F) HIV-1 분리물을 포함하는, R5-지향성 HIV-1에 저항성일 뿐만 아니라, R5-지향성 HIV-1 감염 동안 CCR5 유전자-붕괴되지 않은 세포에 대해 선택적인 장점을 또한 갖는다. 형질도입 후 84일이 지나서도 안정하게 형질도입된 세포 내에서 이들 CCR5 가이드 RNA에 대해 고도로 상동성인 잠재적인 표적외 부위에서의 게놈 돌연변이는 T7 엔조뉴클레아제 I 분석에 의해 검출되지 않았다.

[1330] Fine 등 (Sci Rep. 2015 Jul 1;5:10777 doi. doi:10.1038/srep10777])은, 세포 내에서 함께 스플라이스되어 부

위-특이적 DNA 절단이 가능한 작용성 단백질을 형성하는 스트렙토코커스 피오게네스 Cas9 (SpCas9)의 일부를 발현하는 2-카세트 시스템을 확인하였다. 특정 CRISPR 가이드 가닥을 이용하여, Fine 등은 단일 Cas9로서 및 Cas9 닉카제의 쌍으로서 인간 HEK-293T 세포에서 HBB와 CCR5 유전자를 절단하는데 있어서 본 시스템의 효능을 증명하였다. 트랜스-스플라이스된 SpCas9 (tsSpCas9)는 표준 트랜스펙션 도스에서 야생형 SpCas9 (wtSpCas9)와 비교시 뉴클레아제 활성의 약 35%를 나타내었지만, 더 낮은 수준에서 실질적으로 더 감소된 활성을 가졌다. wtSpCas9에 비하여 tsSpCas9의 크게 감소된 오픈 리딩 프레임 길이는 잠재적으로 더욱 복잡하고 긴 유전 요소가, 조직-특이적 프로모터, 다중화된 가이드 RNA 발현, 및 SpCas9에 대한 이펙터 도메인 융합물을 포함하는 AAV 벡터 내로 패키지되는 것을 가능하게 한다.

[1331] Li 등 (J Gen Virol. 2015 Aug;96(8):2381-93. doi: 10.1099/vir.0.000139. Epub 2015 Apr 8])은 CRISPR-Cas9가 세포주 내에서 CCR5 유전자좌의 편집을 효율적으로 매개할 수 있어서, 세포 표면 상에서 CCR5의 녹아웃을 결과로서 초래함을 증명하였다. 다음세대 시퀀싱은 각종 돌연변이가 CCR5의 예측된 절단 부위 주위에 도입되었음을 밝혀주었다. 세 개의 가장 효과적인 가이드 RNA 각각의 경우, 현저한 표적외 효과는 15 개의 최고-점수 가능 부위에서 검출되지 않았다. CRISPR-Cas9 성분을 운반하는 키메라 Ad5F35 아데노바이러스를 구축함으로써, Li 등은 일차적인 CD4+ T-림프구를 효율적으로 형질도입하였으며, CCR 발현을 붕괴시키고, 양성으로 형질도입된 세포는 HIV-1 저항성을 받았다.

[1332] 당업자는 본 발명의 CRISPR Cas 시스템으로 CCR5를 표적하기 위하여, 예를 들어, 문헌 [Holt, N., et al. Nature biotechnology 28, 839847 (2010), Li, L., et al. Molecular therapy :the journal of the American Society of Gene Therapy 21, 12591269 (2013), Mandal et al., Cell Stem Cell, Volume 15, Issue 5, p643-652, 6 November 2014, Wang et al. (PLoS One. 2014 Dec 26;9(12):e115987 doi: 10.1371/journal.pone.0115987), Fine et al. (Sci Rep. 2015 Jul 1;5:10777 doi. 10.1038/srep10777) 및 Li et al. (J Gen Virol. 2015 Aug;96(8):2381-93. doi: 10.1099/vir.0.000139. Epub 2015 Apr 8)]의 상기 연구들을 활용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 특히, T-풍부 PAM은 비분열 세포 및 조직에서 본 발명의 적용을 허용한다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다.

일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1333] **바이러스성 병원균, 예컨대 HBV와 같은 병원균의 치료**

[1334] 본 발명은 또한 간염 B 바이러스 (HBV)를 치료하기 위하여 적용될 수 있다. 그러나, CRISPR Cas 시스템은 RNAi의 결점, 예컨대 내생성 작은 RNA 경로를 예를 들어 용량 및 서열을 최적화함으로써 과포화하는 (oversatring) 위험을 회피하도록 조정되어야만 한다 (예를 들어, 문헌 [Grimm et al., Nature vol. 441, 26 May 2006) 참조). 예를 들어, 인간에 대해서 약 110 x 1014 입자와 같은 적은 용량이 고찰된다. 다른 구현예에서, CRISPR Cas 시스템 또는 이의 성분은 리포솜, 예컨대, 안정한 핵산-지질 입자(SNALP)로 투여될 수 있다 (예를 들어, 문헌 [Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005] 참조). SNALP 내 HBV RNA로 표적되는 CRISPR Cas의 약 1, 3 또는 5 mg/kg/일의 일일 정맥내 투여가 고려된다. 1일 처리는 약 3일 동안일 수 있고, 그 다음으로 약 5주 동안 매주일 수 있다. 또 다른 구현예에서, Chen 등의 시스템 (문헌 [Gene Therapy (2007) 14, 1119])은 본 발명의 CRISPR Cas 시스템에 이용 및/또는 조정될 수 있다. Chen 등은 shRNA를 전달하기 위해 이중-가닥 아데노연관 바이러스 8-위형(pseudotyped) 벡터 (dsAAV2/8)를 이용한다. HBV-특이적 shRNA를 갖는, dsAAV2/8 벡터의 단일 투여 (마우스 당 1×1012 벡터 게놈)는 HBV 유전자이식 마우스의 간에서 HBV 단백질, mRNA 및 복제형 DNA의 꾸준한 수준을 효과적으로 저해하여, 순환에서 HBV 로드에서의 23 log10 이하에 달하는 감소를 초래한다. 현저한 HBV 저해는 벡터 투여 후 적어도 120일 동안 유지되었다. shRNA의 치료 효과는 표적 서열 의존성이었으며, 인터페론의 활성화를 수반하지 않았다. 본 발명의 경우, HBV를 향한 CRISPR Cas 시스템은 AAV 벡터, 예컨대 dsAAV2/8 내로 클로닝되고, 예를 들어 인간 당 약 1×1015 벡터 게놈 내지 약 1×1016 벡터 게놈의 용량으로 인간에게 투여될 수 있다. 다른 구현예에서, Wooddell 등의 방법 (Molecular Therapy vol. 21 no. 5, 973-985 May 2013)은 본 발명의 CRISPR Cas 시시템에 사용될 수 있고/조정될 수 있다. Woodell 등은 간세포-표적화된, N-아세틸갈락토사민-콘쥬게이트 멜리틴-유사 펩티드 (NAG-MLP)와 응고인자 VII (F7)을 표적화하는 간-향성 콜레스테롤-콘쥬게이트 siRNA (chol-

siRNA) 의 간단한 동시주사가 임상 화학의 변화나 사이토카인의 유도 없이 마우스 및 비-인간 영장류에서 효과적인 F7 녹다운을 가져온다는 것을 나타낸다. HBV 감염의 일시 및 이식유전자 마우스 모델을 이용하여, Wooddell 등은 강한 chol-siRNA 표적화 보존된 HBV 서열과 NAG-MLP의 단일 공동 주입이 긴 기간의 효과를 갖는 바이러스 RNA, 단백질, 및 바이러스 DNA의 멀티로그(multilog) 억제를 결과로서 초래하였음을 보여준다. 예를 들어, 약 6 mg/kg의 NAG-MLP 및 6 mg/kg의 HBV 특이적 CRISPR Cas의 정맥내 공동주입이 본 발명을 위해 고려될 수 있다. 대안적으로, 약 3 mg/kg의 NAG-MLP 및 3 mg/kg의 HBV 특이적 CRISPR Cas가 1일에 전달될 수 있고, 이어서 2주 후 약 약 2-3 mg/kg의 NAG-MLP 및 2-3 mg/kg의 HBV 특이적 CRISPR Cas의 투여가 뒤따른다.

[1335] 일부 구현예에서, 표적 서열은 HBV 서열이다. 일부 구현예에서, 표적 서열은 에피솜 바이러스 핵산 분자를 조작하기 위해서 유기체의 게놈으로 통합되지 않은 에피솜 바이러스 핵산 분자에 포함된다. 일부 구현예에서, 에피솜 핵산 분자는 이중 가닥 DNA 폴리뉴클레오티드 분자이거나 또는 공유적으로 폐쇄된 원형 DNA (cccDNA)이다. 일부 구현예에서, CRISPR 복합체는 복합체 제공이 부재하는 유기체의 세포에서 에피솜 바이러스 핵산 분자의 양과 비교하여 유기체의 세포에서 에피솜 바이러스 핵산 분자의 양을 감소시킬 수 있거나, 또는 에피솜 핵산 분자의 분해를 촉진하기 위해서 에피솜 바이러스 핵산 분자를 조작할 수 있다. 일부 구현예에서, 표적 HBV 서열은 유기체의 게놈으로 통합된다. 일부 구현예에서, 세포 내에서 형성될 때, CRISPR 복합체는 유기체의 게놈으로부터 표적 HBV 핵산의 전체 또는 일부의 절제를 촉진하기 위해 통합된 핵산을 조작할 수 있다. 일부 구현예에서, 상기 적어도 하나의 표적 HBV 핵산은 이중 가닥 DNA 폴리뉴클레오티드 cccDNA 분자 및/또는 유기체의 게놈으로 통합된 바이러스 DNA에 포함되고, CRISPR 복합체는 바이러스 cccDNA 및/또는 통합된 바이러스 DNA를 절단하기 위해서 적어도 하나의 표적 HBV 핵산을 조작한다. 일부 구현예에서, 상기 절단은 바이러스 cccDNA 및/또는 통합된 바이러스 DNA에 도입된 하나 이상의 이중 가닥 파손(들), 임의로 적어도 두개의 이중 가닥 파손(들)을 포함한다. 일부 구현예에서, 상기 절단은 바이러스 cccDNA 및/또는 통합된 바이러스 DNA에 도입된 하나 이상의 단일 가닥 파손(들), 임의로 적어도 두개의 단일 가닥 파손(들)을 포함한다. 일부 구현예에서, 상기 하나 이상의 이중 가닥 파손(들) 또는 상기 하나 이상의 단일 가닥 파손(들)은 바이러스 cccDNA 서열 및/또는 통합된 바이러스 DNA 서열에 하나 이상의 삽입 또는 결실 돌연변이(INDEL)의 형성을 초래한다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1336] Lin 등 (Mol Ther Nucleic Acids. 2014 Aug 19;3:e186. doi: 10.1038/mtna.2014.38)은 유전자형 A의 HBV에 대한 8종 gRNA를 설계하였다. HBV-특이적 gRNA를 이용하여, CRISPR-Cas9 시스템은 HBV-발현 벡터로 트랜스펙션된 Huh-7 세포에서 HBV 코어 및 표면 단백질의 생산을 현저히 감소시켰다. 8 개의 스크리닝된 gRNA들 중, 두 개의 효과적인 것들을 확인하였다. 보존된 HBV 서열을 표적화하는 하나의 gRNA는 상이한 유전형에 대하여 작용하였다. 유체역학-HBV 지속 마우스 모델을 이용하여, Lin 등은 이 시스템이 간 내 HBV 게놈-함유 플라스미드를 절단하고, 그의 생체 내 소제(clearance)를 촉진하여, 혈청 표면 항원 수준의 감소를 결과로서 초래함을 추가로 증명하였다. 이들 데이터는, CRISPR-Cas9 시스템이 시험관 내 생체 내 모두에서 HBV-발현 주형을 붕괴시킬 수 있음을 제안하며, 지속적인 HBV 감염을 근절하는 그의 능력을 표시한다.

[1337] Dong 등 (Antiviral Res. 2015 Jun;118:110-7. doi: 10.1016/j.antiviral.2015.03.015. Epub 2015 Apr 3)은 HBV 게놈을 표적화하고 HBV 감염을 효율적으로 억제하기 위해서 CRISPR-Cas9 시스템을 사용하였다. Dong 등은 HBV의 보존된 영역을 표적화하는 4종의 단일 가이드 RNA (가이드 RNA)를 합성하였다. Cas9와 이들 가이드 RNA의 발현은 Huh7 세포를 비롯하여 HBV-복제 세포 HepG2.2.15에서 바이러스 생산을 감소시켰다. Dong 등은 CRISPR-Cas9가 절단을 유도하고 형질감염된 세포의 HBV cccDNA에서 절단-매개 돌연변이유발이 발생되었다는 것을 더욱 입증하였다. HBV cccDNA를 보유하는 마우스 모델에서, 신속한 꼬리 정맥을 통한 가이드 RNA-Cas9 플라스미드의 주사는 낮은 수준의 cccDNA 및 HBV 단백질을 초래하였다.

[1338] Liu 등 (J Gen Virol. 2015 Aug;96(8):2252-61. doi: 10.1099/vir.0.000159. Epub 2015 Apr 22])은 상이한 HBV 유전형의 보존된 영역을 표적한 8 개의 가이드 RNA(gRNA)를 설계하였으며, 이는 시험관 내 및 생체 내 모두에서 HBV 복제를 현저히 저해시켜 CRISPR-Cas9 시스템을 이용하여 HBV DNA 주형을 붕괴시키는 가능성을 조사할

수 있었다. HBV-특이적 gRNA/C2c1 시스템은 세포 내에서 상이한 유전형의 HBV의 복제를 저해할 수 있을 것이며, 바이러스성 DNA는 단일 gRNA/C2c1 시스템에 의해 현저히 감소되었으며, 상이한 gRNA/C2c1 시스템의 조합에 의해 소제되었다.

[1339] Wang 등 (World J Gastroenterol. 2015 Aug 28;21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554)은 유전자형 A-D의 HBV에 대해 15 gRNA를 설계하였다. HBV의 조절 영역을 포괄하는 2종의 상기 gRNA (이중-gRNA)의 11 조합이 선택되었다. HBV (유전형 A-D) 복제의 저해에 대한 각각의 gRNA 및 11 개의 이중-gRNA의 효율은 배양물 상등액 내 HBV 표면 항원 (HBsAg) 또는 e 항원 (HBeAg)의 측정에 의해 검사하였다. HBV-발현 벡터의 파괴는 이중-gRNA 및 HBV-발현 벡터로 공동-트랜스펙션된 HuH7 세포 내에서 중합효소 연쇄 반응 (PCR) 및 시퀀싱 방법을 이용하여 검사되었으며, cccDNA의 파괴는 KCl 제조, 플라스미드-안전 ATP-의존성 DNase (PSAD) 소화, 회전환 증폭 및 정량 PCR 조합된 방법을 이용하여 HepAD38 세포에서 검사되었다. 이들 gRNA의 세포독성은 미토콘드리아 테트라졸륨 분석에 의해 평가되었다. 모든 gRNA는 배양물 상등액 내 HBsAg 또는 HBeAg 생산을 현저히 감소시킬 수 있었으며, 이는 gRNA가 맞서는 영역에 따라 좌우되었다. 이중 gRNA는 모두 유전형 A-D의 HBV에 대해 HBsAg 및/또는 HBeAg 생산을 효율적으로 억제하였으며, HBsAG 및/또는 HBeAg 생산 저해에 있어 이중 gRNAdml 효능은, 단일 gRNA를 단독으로 사용한 것에 비하여 현저히 증가되었다. 추가로, PCR 직접 시퀀싱에 의해, 출원인은 이들 이중 gRNA가 gRNA를 사용한 두 개의 절단 부위 사이의 단편을 제거함으로써 HBV 발현 주형을 특이적으로 파괴할 수 있었음을 확인하였다. 가장 중요하게는, gRNA-5 및 gRNA-12 조합은 HBsAg 및/또는 HBeAg 생산을 효율적으로 저해할 수 있었을 뿐 아니라, HepAD38 세포 내에서 cccDNA 저장소를 파괴할 수 있었다.

[1340] Karimova 등 (Sci Rep. 2015 Sep 3;5:13734. doi: doi:10.1038/srep13734])은 Cas9 닉카제에 의한 특이적이고 효과적인 절단에 대해 표적된 HBV 게놈의 S 및 X 영역에서 교차-유전형 보존된 HBV 서열을 확인하였다. 이러한 시도는 리포터 세포주에서 에피좀성 cccDNA 및 염색체로 통합된 HBV 표적 부위 뿐만 아니라, 만성적으로 및 신규 감염된 간암 세포주 내 HBV 복제를 방해하였다.

[1341] 당업자는 본 발명의 CRISPR Cas 시스템을 사용해 HBV를 표적화하기 위해서 예를 들어, 다음의 연구들을 이용할 수 있다: Lin et al. (Mol Ther Nucleic Acids. 2014 Aug 19;3:e186. doi: 10.1038/mtna.2014.38), Dong et al. (Antiviral Res. 2015 Jun;118:110-7. doi: 10.1016/j.antiviral.2015.03.015. Epub 2015 Apr 3), Liu et al. (J Gen Virol. 2015 Aug;96(8):2252-61. doi: 10.1099/vir.0.000159. Epub 2015 Apr 22), Wang et al. (World J Gastroenterol. 2015 Aug 28;21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554) 및 Karimova et al. (Sci Rep. 2015 Sep 3;5:13734. doi: 10.1038/srep13734). C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1342] 만성 B형 간염 바이러스 (HBV) 감염은 우세하게, 치명적이고, 감염된 세포에서 바이러스 에피솜 DNA (cccDNA)의 지속성으로 인해 거의 치유되지 않는다. Ramanan 등 (Ramanan V, Shlomai A, Cox DB, Schwartz RE, Michailidis E, Bhatta A, Scott DA, Zhang F, Rice CM, Bhatia SN, .Sci Rep. 2015 Jun 2;5:10833. doi: 10.1038/srep10833, published online 2nd June 2015.)은 showed that the CRISPR/Cas9 시스템이 HBV 게놈 내 보존된 영역을 특이적으로 표적화하여 절단할 수 있어서, 바이러스 유전자 발현 및 복제의 강력한 억제를 야기시킨다는 것을 보여주었다. Cas9 및 적절하게 선택된 가이드 RNA의 지속적인 발현 시, 그들은 Cas9에 의한 cccDNA의 절단 및 cccDNA 및 바이러스 유전자 발현과 복제의 다른 매개변수의 극적인 감소를 입증하였다. 따라서, 그들은 바이러스 에피솜 DNA의 직접 표적화가 바이러스를 제어하고 아마도 환자를 치유하는 신규 치료 접근법이라는 것을 보여주었다. 이것은 또한 WO2015089465 A1 (The Broad Institute 등의 출원인)에 기술되어 있고, 이의 내용을 참조로 본 명세서에 편입시킨다.

[1343] 이와 같이 HBV의 바이러스 에피솜 DNA의 표적화가 일부 구현예에서 바람직하다.

[1344] 본 발명은 또한 병원균, 예를 들어 박테리아, 균류 및 기생 병원균을 처리하는 데 적용될 수 있다. 대부분의 연구 노력은 새로운 항생제의 개발에 집중하며, 일단 개발되면, 그럼에도 불구하고 이는 약물 저항성의 동일

한 문제의 대상이 될 것이다. 본 발명은 그러한 어려움들을 극복하는 신규 CRISPR-기반 대안을 제공한다.

나아가, 기존의 항생제와 달리, CRISPR-기반 치료는 병원균 특이적으로 제조될 수 있어서, 유익한 박테리아는 회피하면서 표적 병원균의 박테리아 세포 사멸을 유도한다.

[1345] 본 발명은 또한 C형 간염 바이러스(HCV)를 치료하기 위해서 적용될 수도 있다. Roelvinki 등의 방법(Molecular Therapy vol. 20 no. 9, 1737-1749 Sep 2012)은 CRISPR Cas 시스템에 적용될 수 있다. 예를 들어, AAV8과 같은 AAV 벡터가 고찰된 벡터일 수 있으며, 예를 들어 킬로그램 체중 당 약 1.25 × 1011 내지 1.25 × 1013 벡터 게놈 (vg/kg)의 용량이 고찰될 수 있다. 본 발명은 또한 병원균, 예를 들어 박테리아, 균류 및 기생 병원균을 처리하는 데 적용될 수 있다. 대부분의 연구 노력은 새로운 항생제의 개발에 집중하며, 일단 개발되면, 그럼에도 불구하고 이는 약물 저항성의 동일한 문제의 대상이 될 것이다. 본 발명은 그러한 어려움들을 극복하는 신규 CRISPR-기반 대안을 제공한다. 나아가, 기존의 항생제와 달리, CRISPR-기반 치료는 병원균 특이적으로 제조될 수 있어서, 유익한 박테리아는 회피하면서 표적 병원균의 박테리아 세포 사멸을 유도한다. 일부 구현예에서, CRISPR-C21 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1346] Jiang 등 (문헌[" RNA-guided editing of bacterial genomes using CRISPR-Cas systems", Nature Biotechnology vol. 31, p. 2339, March 2013])은, CRISPR-Cas9 시스템을 사용하여 S. 뉴모니아에 및 에스케리키아 콜라이를 돌연변이 또는 사멸시킨다. 정확한 돌연변이를 게놈 내로 도입한 이 작업은 돌연변이되지 않은 세포를 사멸시키기 위하여 표적화된 게놈 부위에서의 이중-RNA:Cas9-유도된 절단에 의존하였으며, 선별 마커 또는 대항-선택 시스템에 대한 필요를 피하였다. CRISPR 시스템은 항생제 저항성을 역전하고, 균주들 간의 저항성의 전달을 제거하는 데 사용되었다. Bickard 등은, 독성 유전자를 표적하도록 재프로그램화된 Cas9가 독성인 S. 아우레우스를 사멸시키지만, 비독성 S. 아우레우스는 사멸시키지 않았음을 보였다. 항생제 저항성 유전자를 표적화하는 뉴클레아제의 재프로그램화는, 항생제 저항성 유전자를 품은 스태필로코커스 플라스미드를 파괴하고, 플라스미드-유래 저항성 유전자의 확산에 대해 면역화되었다. (참조: Bikard et al., " Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials," Nature Biotechnology vol. 32, 1146-1150, doi:10.1038/nbt.3043, published online 05 October 2014.) Bikardss는, CRISPR-Cas9 항균제가 마우스 피부 콜로니화 모델에서 S. 아우레우스를 사멸시키도록 생체 내에서 작용함을 보였다. 유사하게, Yosef외 다수는 CRISPR 시스템을 사용하여 β-락탐 항생제에 대한 저항성을 수여하는 효소를 인코딩하는 유전자를 표적화하였다 (see Yousef et al., "Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria," Proc. Natl. Acad. Sci. USA, vol. 112, p. 7267-7272, doi: 10.1073/pnas.1500107112 published online May 18, 2015).

[1347] 본 발명은 노로바이러스 감염에 대한 치료 개발에 적용될 수도 있다. 노로바이러스는 안전하지 않은 음식물로 인간 설사 질환의 원인이 되는 가장 일반적인 병원체 중 하나이다. 또한 음식 매개 감염에서 어린이와 성인 치사율의 주요 원인이다. 노로바이러스는 단지 식중독만이 아니다. 최근의 메타 분석에서 노로 바이러스는 산발적 및 유행병 상황의 급성 위장염의 모든 원인 (개인간 전파 포함)의 거의 5분의 1을 차지하고 모든 연령군에 영향을 미친다. 분명하게, 노로바이러스는 개발도상국 및 선진국 모두에서 중요한 공중 보건 문제이다. 노로바이러스 병리학을 더욱 잘 이해하기 위한 연구 노력이 표적화 중재술을 위해 필요할 것이다. 중동 호흡기 증후군 코로나바이러스부터 지카 바이러스까지, 바이러스 감염을 매개하는데 중요한 숙주 인자를 동정하기 위한 노력은 항상 연구 우선순위였다. 이러한 정보는 항바이러스 중재술에서 잠재적 치료 표적에 빛을 밝혀줄 것이다. 노로바이러스 바이러스-숙주 상호작용 연구는 지난 20여년간 강력한 세포 배양 모델의 결여로 인해 방해되었다. 2016년에, 노로바이러스는 결국 엔테로이드 또는 미니-내장이라고 하는 줄기 세포-유래 3차원 인간 내장-유사 구조에서 성공적으로 배양되었다. Chan 등은 참가자로부터 채취한 십이지장 생검에서 단리된 장 줄기 세포를 사용하였고, 미니-내장에서 분화시켰다. 녹아웃 CRISPR 및 기능 획득 CRISPR SAM은 노로바이러스 감염에 관여된 유전자의 최종 후보를 동정하는데 사용되었다. 본 발명에서 개시된 C2c1- CRISPR 시스템은 Chan 등이 기술한 바와 같은 유사한 시스템에 적용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열

은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1348] 본 발명은 또한 인간 파필로마바이러스 (HPV) 관련 악성 신생물 및 HVP 유도 자궁경부암을 치료하는데 적용될 수 있다. 자궁경부암은 전세계적으로 여성에서 2번째로 가장 흔한 암이다. 고위험 인간 파필로마바이러스 (HR-HPV), 특히 HPV16 및 HPV18이 자궁경부암의 주요 유발원으로 여겨진다. 종양유전자 E6 및 E7은 HVP 감염의 초기 단계에 발현되고, 그들 기능은 정상 세포 주기를 파괴하고 형질전환된 악성 표현형을 유지시키는 것이다. 예를 들어, C7 단백질은 쿨린 2 유비퀴틴 리가제 복합체에 결합하여 레티노블라스토마 (pRb) 종양 억제인자의 유비퀴틴화 및 분해를 초래한다.

[1349] Hu 등 (Biomed Res Int. 2014;2014:612823. doi: 10.1155/2014/61283)은 CRISPR-Cas9 시스템을 사용하여 HPV 양성 세포주에서 HPV16-E7 DNA를 표적화하였고 showed that HPV16-E7 단일-가이드 RNA (sgRNA) 가이드된 CRISPR/Cas 시스템이 특이적 부위에서 HPV16-E7 DNA를 파괴할 수 있어서, HPV 양성 SiHa 및 Caski 세포에서 아폽토시스 및 성장 억제를 유도하지만, HPV 음성 C33A 및 HEK293 세포에서는 그렇지 않았다는 것을 보여주었다.

게다가, E7 DNA의 파괴는 E7 단백질의 하향 조절 및 종양 억제인자 단백질 pRb의 상향조절을 직접적으로 유도한다. HPV16-E7을 표적화하는 gRNA는 Mali 등의 프로토콜에 따라 설계되었고 그들은 Genewiz Company (China)에서 합성되었다. SSA 루시퍼라제 리포터 pSSA Rep3-1은 CRISPR 시스템의 전달의 리포팅 시스템으로서 사용되었다. 세포는 23-웰 플레이트 중에서 0.8 μg 의 Cas9 플라스미드 및 0.2 μg 의 gRNA 플라스미드와 함께 공동 형질감염시켰다. 형질감염 후 48시간에, 그들을 회수하였고, 제조사 설명서에 따라서 아넥신 V-FITC 아폽토시스 검출 키트 (KeyGen BioTech)를 사용해 플루오레세인 이소티오시아테이트- (FITC-) 접합된 아넥신 V (아넥신 V-FITC) 및 프로피듐 아이오다이드 (PI)로 이중 염색하였다. 모든 4종 CRISPR/Cas 시스템 처리된 세포주의 아폽토시스는 유도된 세포 사멸을 계산하기 위해서 FACS Calibur (BD Bioscience)를 사용해 분석하였다. 데이터는 BD Cell Quest 소프트웨어를 사용해 분석하였다. 시험관내 세포 증식은 Cell Counting Kit-8 (CCK-8; Beyotime)을 사용해 결정하였다. gRNA-4/Cas9 플라스미드로 1 × 104 세포/웰을 형질감염시켰고, 형질감염 후 24시간에 세포를 트립신처리하고 96웰 플레이트 상에 파종하였다. 96웰 플레이트 상에 파종한 후 0시간, 24시간, 48시간, 72시간, 및 96시간에, 10μL CCK-8 용액을 각 웰에 첨가하고 37℃에서 2.5시간 인큐베이션하였다. 본 발명에서 개시된 CRISPR- C2c1 시스템은 시스템에서 사용할 수 있다.

C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 5' TTN 3' 또는 5' ATTN 3'인 PAM 서열을 인식할 수 있고, 여기서 N은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다.

일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다.

다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1350] CRISPR 시스템은 기타 다른 유전학적 접근에 대해 저항성인 기생충의 게놈을 편집하는 데 사용될 수 있다. 예를 들어, CRISPR-Cas9 시스템은 플라스모듐 요엘리(Plasmodium yoelii) 게놈 내로 이중-가닥 파손을 도입하는 것을 나타내었다(참조: Zhang et al., "Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System," mBio. vol. 5, e01414-14, Jul-Aug 2014). Ghorbal 등 (" 'Genome editing in the human malaria parasite Plasmodium falciparumusing the CRISPR-Cas9 system," Nature Biotechnology, vol. 32, p. 819-821, doi: 10.1038/nbt.2925, published online June 1, 2014)은 각각 아르테미시닌에 대한 내성 출현 및 유전자 침묵화에서 추정 역할을 갖는, 2개 유전자, orc1 및 kelch13의 서열을 변형시켰다. 적절한 부위에서 변경된 기생충은 변형에 대한 직접 선택은 없음에도 불구하고 매우 높은 효율로 회수되어, 중립성 또는 심지어는 유해한 돌연변이가 이 시스템을 사용하여 생성될 수 있음을 나타낸다. CRISPR-Cas9은 또한 톡소플라스마 곤디 (Toxoplasma gondii)를 포함한 다른 병원성 기생충의 게놈을 변형시키는데 사용될 수 있다 (참조: Shen et al., "Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9," mBio vol. 5:e01114-14, 2014; 및 Sidik et al., "Efficient Genome Engineering of Toxoplasma gondii Using CRISPR/Cas9," PLoS One vol. 9, e100450, doi: 10.1371/journal.pone.0100450,

published online June 27, 2014). C2c1 이펙터 단백질은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1351] Vyas 등 ("A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families," Science Advances, vol. 1, e1500248, DOI: 10.1126/sciadv.1500248, April 3, 2015) 은 CRISPR 시스템을 이용하여 C. 알비칸스에서의 유전적 조작에 대한 오랜 장애를 극복하였으며, 몇몇 상이한 유전자의 두 개의 카피 모두를 단일 실험에서 효율적으로 돌연변이시켰다. 몇몇 메카니즘이 약물 저항성에 기여하는 유기체서, Vyas는 부모의 임상적 분리물 Can90에 의해 나타나는 플루코나졸 또는 시클로헥시미드에 대한 초저항성을 더 이상 나타내지 않는 동형접합성 이중 돌연변이체를 생산하였다. Vyas는 조건적 대립형질을 만듬으로써 C. 알비칸스의 필수 유전자에서의 동형접합성 기능 손실 돌연변이를 또한 수득하였다. 리보솜 RNA 가공에 대해 요구되는 DCR1의 무효 대립형질은 저온에서 치명적이지만 고온에서는 생존성이다. Vyas는 논센스 돌연변이를 도입하고, 16℃에서 성장하는 데 실패한 dcr1/dcr1 돌연변이체를 분리한 복구 주형을 사용하였다.
C2c1 이펙터 단백질은 유사한 시스템에 적용될 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 전사물에 단일 뉴클레오티드 변형을 DCR1의 전사물에 도입시킨다. 일부 구현예에서, 복구 주형은 PAM 서열을 포함하지 않는다.

[1352] **유전적 또는 후생적 양태를 갖는 질병의 치료**

[1353] 본 발명의 CRISPR-Cas 시스템은 TALEN 및 ZFN를 사용해 제한적인 성공으로 이전에 시도되었던 유전자 돌연변이를 교정하는데 사용될 수 있고, Gluckmann 등의 WO 2015/048577 CRISPR-RELATED METHODS AND COMPOSITIONS; Glucksmann 등의 WO 2015/070083 CRISPR-RELATED METHODS AND COMPOSITIONS WITH GOVERNING gRNAS를 포함하여, 유전자 요법으로 질환을 치료적으로 해결하기 위해 표적 유전자좌로 Cas9 시스템을 이용하기 위한 방법을 설명하는 Editas Medicine의 공개 출원을 포함하여, Cas9 시스템에 대한 잠재적 표적으로서 동정되었다; 일부 구현예에서, 원발성 개방각 녹내장 (POAG)의 치료, 예방 또는 진단이 제공된다. 표적은 바람직하게 MYOC 유전자이다. 이것은 Wo2015153780에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다.

[1354] Maeder 등의 WO2015/134812 CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING USHER SYNDROME AND 망막 색소변성증를 참조한다. 본 명세서의 교시를 통해서 본 발명은 본 명세서의 교시와 함께 적용되는 이들 문헌의 재료 및 방법을 이해한다. 안구 및 청각 유전자 요법의 양상에서, 어셔 증후군 및 망막 색소 변성증을 치료하기 위한 방법 및 조성물은 본 발명의 CRISPR-Cas 시스템에 대해 적합화될 수 있다 (참조: 예를 들어 WO 2015/134812). 일 구현예에서, WO 2015/134812는 USH2A 유전자의 위치 2299에서 구아닌 결실을 교정하기 위해 (예를 들어, USH2A 유전자의 위치 2299에서 결실된 구아닌 잔기 치환) 예를 들어 CRISPR-Cas9 매개 방법을 사용하는, 유전자 편집에 의해 IIA형 어셔 증후군 (USH2A, USH11A) 및 망막 색소변성증 39 (Rp39)의 개시 또는 진행의 치료 또는 지연을 포함한다.유사한 효과가 C2c1로 획득될 수 있다. 관련 양상에서, 예를 들어, 점 돌연변이 (예를 들어, 단일 뉴클레오티드, 예를 들어 구아닌, 결실)을 교정하는 도너 주형으로 HDR을 유도하기 위해서, 하나 이상의 뉴클레아제, 하나 인상의 닉카제, 또는 이의 조합으로 절단을 통해 돌연변이를 표적화한다. 돌연변이체 USH2A 유전자의 변경 또는 교정은 임의 기전으로 매개될 수 있다. 돌연변이체 HSH2A 유전자의 변경 (예를 들어, 교정)과 연관될 수 있는 예시적인 기전은 제한없이 비상동성 말단 연결, 미세상동성-매개 말단 연결 (MMEJ), 상동성-지정 복구 (예를 들어, 내생성 도너 주형 매개), SDSA (합성 의존적 가닥 어닐링), 단일 가닥 어닐링, 또는 단일 가닥 침입을 포함한다. 일 구현예에서, 어셔 증후군 및 망막 색

소변성증을 치료하기 위해 사용되는 방법은 예를 들어 USH2A 유전자의 적절한 부분을 시퀀싱하여, 대상체가 보유하는 돌연변이의 지식을 획득하는 것을 포함할 수 있다.

[1355] 따라서, 일부 구현예에서, 망막 색소변성증의 치료, 예방 또는 진단이 제공된다. RP1, RP2 등과 같은 다수의 상이한 유전자가 망막 색소변성증과 연관되거나 또는 그를 초래하는 것으로 알려져 있다. 이들 유전자는 일부 구현예에서 표적화되고 적합한 주형의 제공을 통해 녹아웃되거나 또는 복구된다. 일부 구현예에서, 눈에 전달은 주사에 의한다.

[1356] 하나 이상의 망막 색소변성증 유전자는 일부 구현예에서, 하기로부터 선택될 수 있다: RP1 (망막 색소변성증-1), RP2 (망막 색소변성증-2), RPGR (망막 색소변성증-3), PRPH2 (망막 색소변성증-7), RP9 (망막 색소변성증-9), IMPDH1 (망막 색소변성증-10), PRPF31 (망막 색소변성증-11), CRB1 (망막 색소변성증-12, 상염색체 열성), PRPF8 (망막 색소변성증-13), TULP1 (망막 색소변성증-14), CA4 (망막 색소변성증-17), HPRPF3 (망막 색소변성증-18), ABCA4 (망막 색소변성증-19), EYS (망막 색소변성증-25), CERKL (망막 색소변성증-26), FSCN2 (망막 색소변성증-30), TOPORS (망막 색소변성증-31), SNRNP200 (망막 색소변성증 33), SEMA4A (망막 색소변성증-35), PRCD (망막 색소변성증-36), NR2E3 (망막 색소변성증-37), MERTK (망막 색소변성증-38), USH2A (망막 색소변성증-39), PROM1 (망막 색소변성증-41), KLHL7 (망막 색소변성증-42), CNGB1 (망막 색소변성증-45), BEST1 (망막 색소변성증-50), TTC8 (망막 색소변성증 51), C2orf71 (망막 색소변성증 54), ARL6 (망막 색소변성증 55), ZNF513 (망막 색소변성증 58), DHDDS (망막 색소변성증 59), BEST1 (망막 색소변성증, 동심), PRPH2 (망막 색소변성증, 이유전자), LRAT (망막 색소변성증, 소아), SPATA7 (망막 색소변성증, 소아, 상염색체 열성), CRX (망막 색소변성증, 후기 발병 우성), 및/또는 RPGR (망막 색소변성증, X-연관, 및 부비동 감염, 난청 동반 또는 비동반).

[1357] 일부 구현예에서, 망막 색소변성증 유전자는 MERTK (망막 색소변성증-38) 또는 USH2A (망막 색소변성증-39)이다.

[1358] 또한 WO 2015/138510을 언급하며 본 명세서의 교시를 통해서 본 발명 (CRISPR-Cas9 시스템 사용)은 레버 선천성 흑암시 10 (LCA 10)의 치료 또는 개시 또는 진행의 지연을 제공함을 이해한다. LCA 10은 CEP290 유전자 내 돌연변이, 예를 들어 인트론 26에 크립틱 슬라이스 부위를 발생시키는 CEP290 유전자내, 아데닌에서 구아닌으로의 돌연변이, a c.2991+1655에 의해 초래된다. 이것은 CEP290의 인트론 26의 뉴클레오티드 1655에서의 돌연변이, 예를 들어 A에서 G로의 돌연변이이다. CEP290은 CT87; MKS4; POC3; rd16; BBS14; JBTS5; LCAJO; NPHP6; SLSN6; 및 3H11Ag로도 공지되어 있다 (참조: 예를 들어 WO 2015/138510). 유전자 요법의 양상에서, 본 발명은 CEP290 유전자의 적어도 하나의 대립유전자에서 LCA 표적 위치 (예를 들어, c.2991 + 1655; A에서 G)의 부위 근처에 하나 이상의 파손을 도입시키는 것을 포함한다. LCA10 표적 위치의 변경은 refers to (1) 밀접하게 근접하여 indel의 파손-유도 도입 (본 명세서에서는 또한 indel의 NHEJ-매개 도입이라고도 함) 또는 LCA10 표적 위치 (예를 들어, c.2991+1655A to G) 도입, 또는 (2) LCA10 표적 위치 (예를 들어, c.2991+1655A에서 G)에 돌연변이를 포함하는 게놈 서열의 파손-유도 결실 (본 명세서에서는 또한 NHEJ-매개 결실이라고도 함)이라고 하낟. 양쪽 접근법은 LCA 10 표적 위치에 돌연변이로 인한 크립틱 스플라이스 부위의 소실 또는 파괴를 일으킨다. 따라서, LCA의 치료에서 C2c1의 사용이 특히 고려된다.

[1359] 연구자들은 유전자 치료를 광범위한 질병을 치료하는데 사용할 수 있는지의 여부를 숙고하고 있다. C2c1 이펙터 단백질을 기반으로 한 본 발명의 CRISPR 시스템은 추가의 예시된 표적된 영역 및 하기 전달 방법들에 제한되지만 이들을 포함하는, 그러한 치료적 이용을 고려한다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3'이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다. 본 시스템을 이용하여 유용하게 치료될 수 있는 증상 또는 질병의 일부 예는 본 명세서에 포함된 유전자 및 참고문헌의 예에 포함되고, 또한 그러한 증상들과 현재 연관된 것들도 또한 거기 제공된다. 예시된 유전자 및 증상은 빠짐없이 완전한 것은 아니다.

[1360] **순환계 질병 치료**

[1361] 본 발명은 또한 본 명세서에 설명된 신규 CRISPR-Cas 시스템, 구체적으로 CRISPR 이펙터 단백질 시스템을 혈액 또는 조혈 줄기 세포로 전달하는 것을 고려한다. Wahlgren 등 (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130)의 혈장 엑소솜은 이전에 기술되었고 혈액으로 CRISPR Cas 시스템을 전달하는데 이용될 수 있다.

본 발명의 핵산-표적화 시스템은 지중해빈혈 및 겸상 세포 질병과 같은 이상헤모글로빈증의 치료를 또한 고려한다. 예를 들어, 본 발명의 CRISPR Cas 시스템에 의해 표적될 수 있는 잠재적 표적에 대한, 국제 특허 공개 WO 2013/126794호를 참조한다.

[1362] Drakopoulou의 문헌 ["Review Article, The Ongoing Challenge of Hematopoietic Stem Cell-Based Gene Therapy for β-Thalassemia," Stem Cells International, Volume 2011, Article ID 987980, 10 pages, doi:10.4061/2011/987980]은 전체를 명시된 것 같은 그가 언급하는 문헌들과 함께 본 명세서에 참고문헌으로 통합되며, 이는 β-글로빈 또는 γ-글로빈에 대한 유전자를 전달하는 렌티바이러스를 이용하여 HSC를 변형시키는 것을 논의한다. 렌티바이러스 이용에 대조적으로, 해당 분야의 지식 및 본 개시 내용에서의 교시를 이용하여, 숙련자는 β-지중해빈혈에 대하여 돌연변이를 표적화하고 수정하는 CRISPR-Cas 시스템을 이용하여 HSC를 교정할 수 있고 (예를 들어, β-글로빈 또는 γ-글로빈, 유리하게는 비-겸상적혈구화 β-글로빈 또는 γ-글로빈에 대한 코딩 서열을 전달하는 적합한 HDR 주형을 이용); 특히, 가이드 RNA는 β-지중해빈혈을 야기시키는 돌연변이를 표적화할 수 있고, HDR은 β-글로빈 또는 γ-글로빈의 적절한 발현을 위한 코딩을 제공할 수 있다. 돌연변이-및-Cas 단백질 함유 입자를 표적화하는 가이드 RNA는 돌연변이를 운반하는 HSC와 접촉된다. 또한 입자는 β-글로빈 또는 γ-글로빈의 적절한 발현을 위해 돌연변이를 교정하기 위해 적합한 HDR 주형을 함유할 수 있거나; 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다. 이렇게 접촉된 세포를 투여할 수 있고; 임의로 치료 /확장된다; cf. Cartier. 이와 관련하여 다음을 언급할 수 있다: In this regard mention is made of: Cavazzana, "Outcomes of Gene Therapy for β-Thalassemia Major via Transplantation of Autologous Hematopoietic Stem Cells Transduced Ex Vivo with a Lentiviral βA-T87Q-Globin Vector." tif2014.org/abstractFiles/Jean%20Antoine%20Ribeil_Abstract.pdf; Cavazzana-Calvo, "Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia", Nature 467, 318-322 (16 September 2010) doi:10.1038/nature09328; Nienhuis, "Development of Gene Therapy for Thalassemia, Cold Spring Harbor Perpsectives in Medicine, doi: 10.1101/cshperspect.a011833 (2012), LentiGlobin BB305, a lentiviral vector containing an engineered β-globin gene (βA-T87Q); 및 Xie et al., "Seamless gene correction of β-thalassaemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyback" Genome Research gr.173427.114 (2014) www.genome.org/cgi/doi/10.1101/gr.173427.114 (Cold Spring Harbor Laboratory Press); 이것은 인간 β-지중해빈혈을 포함한 Cavazzana 작업의 주제이고, Xie 작업의 주제이며, 이에 인용되거나 이와 연관된 모든 문헌들과 함께, 모두 참조로 본 명세서에 편입된다. 본 발명에서, HDR 주형은 HSC가 조작된 β-글로빈 유전자 (예를 들어, βA-T87Q), 또는 Xie에서와 같이 β-글로빈을 발현하도록 제공될 수 있다.

[1363] Xu 등 (Sci Rep. 2015 Jul 9;5:12065 doi. doi:10.1038/srep12065])은 TALEN 및 CRISPR-Cas9를 설계하여 글로빈 유전자 내에서 인트론2 돌연변이 부위인 IVS2-654를 직접 표적한다. Xu 등은 IVS2-654 유전자좌에서 TALEN 및 CRISPR-Cas9를 이용하여 이중-가닥 파손 (DSB)의 상이한 빈도를 관찰하였으며, TALEN은 돼지Bac 트랜스포손 공여자와 조합된 경우 CRISPR-Cas9에 비교하여 더욱 높은 상동성 유전자 표적화 효율을 매개하였다. 추가적으로, 더욱 명백한 표적외 이벤트는 TALEN에 비하여 CRISPR-Cas9의 경우에 관찰되었다. 최종적으로, TALEN-수정된 iPSC 클론이 OP9 공동-배양 시스템을 이용하여 적혈모세포 분화를 위해 선택되었고, 미수정된 세포 외에 HBB의 상대적으로 더욱 높은 전사를 검출하였다.

[1364] Song 등 (Stem Cells Dev. 2015 May 1;24(9):1053-65. doi: 10.1089/scd.2014.0347. Epub 2015 Feb 5)은 CRISPR/Cas9를 사용해 β-Thal iPSC를 교정하였고; 유전자-교정 세포는 정상 핵형을 나타내며 인간 배아 줄기 세포 (hESC)로서 완전한 다능성은 오프-표적화 효과를 보이지 않았다. 그 다음, Song 등은 유전자-수정된 β-Thal iPSC의 분화 효율을 평가하였다. Song 등은 조혈성 분화 동안, 유전자-수정된 β-Thal iPSC가 증가된 배양체 비율 및 각종 조혈성 선조 세포 퍼센트를 나타내었음을 발견하였다. 더욱 중요하게는, 유전자-수정된 β-Thal iPSC 계통이 HBB 발현을 복구하고, 수정되지 않은 기에 비하여 반응성 산소 종을 감소시켰다. Song 등의 연구는 β-Thal iPSC의 조혈성 분화가, CRISPR-Cas9 시스템에 의해 일단 수정되면, 크게 개선됨을 제안하였다. 여기 설명된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질을 포함하는 시스템을 활용하여 유사한 방법들이 수행될 수 있다.

[1365] 겸상 적혈구성 빈혈은 적혈 세포가 낫-모양이 되는 상염색체 열성 유전 질병이다. 이는 β-글로빈 유전자에

서 염색체 11의 단완에 위치하는 단일 염기 치환에 의해 생성한다. 결과로서, 발린은 겸상 헤모글로빈(HbS)의 생산을 발생시키는 글루탐산 대신 발린이 생산된다. 이는 왜곡된 형태의 적혈구의 형성을 야기한다. 비정상적인 모양에 의해, 작은 혈관이 차단되어, 뼈, 비장 및 피부 조직에 심각한 손상을 발생시킨다. 이는 통증의 에피소드, 빈번한 감염 및 손-발 증후군 또는 심지어는 다수 기관 부전을 이끌어낼 수 있다. 왜곡된 적혈구는 또한 용혈에 더 민감하여, 심각한 빈혈증을 이끌어낸다. β-지중해빈혈의 경우에서와 같이, 겸상 적혈구 빈혈증은 CRISPR-Cas 시스템으로 HSC를 변형시켜 편집될 수 있다. 시스템은 이의 DNA를 절단한 후 이를 자체 복구하도록 하여 세포의 게놈의 특이적 편집을 가능하게 한다. Cas 단백질은 돌연변이 점으로 RNA 가이드에 의해 삽입되고 유도된 후 이 지점에서 DNA를 절단한다. 동시에, 건강한 버전의 서열이 삽입된다. 이 서열은 세포 자신의 복구 시스템에 의해 사용되어 유도된 절단을 고친다. 이 방식에서, CRISPR-Cas는 이전에 수득된 줄기 세포에서 돌연변이의 편집을 가능하게 한다. 당 분야의 지식 및 이 내용에서의 기재를 이용하여, 숙련자는 (예를 들어, β-글로빈, 유리하게는 비-겸상 β-글로빈에 대한 코딩 서열을 전달하는 적합한 HDR 주형을 가지고) 돌연변이를 표적화하고 편집하는 CRISPR-Cas 시스템을 사용하여 겸상 적혈구 빈혈증에 관한 HSC를 편집할 수 있으며; 돌연변이-및-Cas 단백질 함유 입자를 표적화하는 가이드 RNA는 돌연변이를 운반하는 HSC와 접촉된다. 또한 입자는 β-글로빈 또는 글로빈의 적절한 발현을 위해 돌연변이를 교정하기 위해 적합한 HDR 주형을 함유할 수 있거나; 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다. 이렇게 접촉된 세포를 투여할 수 있고; 임의로 치료 /확장된다; cf. Cartier. HDR 주형은 HSC가 조작된 β-글로빈 유전자 (예를 들어, βA-T87Q), 또는 Xie에서와 같이 β-글로빈을 발현하도록 제공할 수 있다.

[1366] 전체로 기재된 대로, 인용된 문헌들과 함께 참조로 본 명세서에 편입되는, [Williams, "Broadening the Indications for Hematopoietic Stem Cell Genetic Therapies," Cell Stem Cell 13:263-264 (2013)]은 리소솜 저장 질환 이염성 백질 이영양증 (MLD), 신경 탈수초를 초래하는 아릴술파타제 A 결핍 (ARSA)으로 야기된 유전병 환자로부터의 HSC/P 세포로 렌티바이러스-매개 유전자 전달; 및 위스코트-알드리치 증후군 (WAS) 환자 (혈액 세포 계통에서 세포골격 기능을 조절하는 소형 GTPase CDC42의 이펙터인 WAS 단백질이 결핍된 환자로서, 재발성 감염 동반 면역 결핍, 자가면역 증후군, 및 과다 출혈을 초래하는 비정상적으로 작고 이상기능인 혈소판 존재로 백혈병 및 림프종의 위험성이 증가되는 혈소판 감소증을 앓게됨)의 HSC로 렌티바이러스-매개 유전자 전달을 보고한다. 렌티바이러스의 사용과 달리, 당 분야에서의 지식 및 본원 상세한 설명에서의 개시를 가지고, 숙련자는 (예를 들어, ARSA에 대한 코딩 서열을 전달하는 적절한 HDR 주형으로) 돌연변이 (아실술파타제 A (ARSA)의 결핍)를 표적으로 하고 편집하는 CRISPR-Cas 시스템을 사용하여 MLD (아실술파타제 A (ARSA)의 결핍)에 관하여 HSC를 편집할 수 있다; 특별히, 가이드 RNA는 MLD (결핍 ARSA)를 발생시킨 돌연변이를 표적으로 할 수 있고 ARSA의 적절한 발현을 위한 코딩을 제공할 수 있다. 돌연변이-및-Cas 단백질 함유 입자를 표적화하는 가이드 RNA는 돌연변이를 운반하는 HSC와 접촉된다. 또한 입자는 ARSA 글로빈 또는 글로빈의 적절한 발현을 위해 돌연변이를 교정하기 위해 적합한 HDR 주형을 함유할 수 있거나; 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다. 이렇게 접촉된 세포를 투여할 수 있고; 임의로 치료 /확장된다; cf. Cartier. 렌티바이러스의 사용과 달리, 당 분야에서의 지식 및 본원 상세한 설명에서의 개시를 가지고, 숙련자는 (예를 들어, WAS 단백질에 대한 코딩 서열을 전달하는 적절한 HDR 주형으로) 돌연변이 (WAS 단백질 결핍)를 표적으로 하고 교정하는 CRISPR-Cas 시스템을 사용하여 WAS에 관하여 HSC를 교정할 수 있고; 특히, 가이드 RNA는 WAS (결핍 WAS 단백질)를 발생시킨 돌연변이를 표적으로 할 수 있고 WAS 단백질의 적절한 발현을 위한 코딩을 제공할 수 있다. 돌연변이-및-C2c1 단백질 함유 입자를 표적화하는 가이드 RNA는 돌연변이를 운반하는 HSC와 접촉된다. 또한 입자는 WAS 글로빈 또는 글로빈의 적절한 발현을 위해 돌연변이를 교정하기 위해 적합한 HDR 주형을 함유할 수 있거나; 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다. 이렇게 접촉된 세포를 투여할 수 있고; 임의로 치료 /확장된다; cf. Cartier.

[1367] 완전히 기재된 바와 같이, 인용된 문헌과 함께 참조로 본 명세서에 편입되는, [Watts, "Hematopoietic Stem Cell Expansion and Gene Therapy" Cytotherapy 13(10):1164-1171.doi:10.3109/14653249.2011.620748 (2011)]은 혈액학적 질환, HIV/AIDS를 포함하는 면역결핍, 및 SCID-X1, ADA-SCID, β-지중해빈혈, X-관련 CGD, 위스코트-알드리치 증후군, 판코니 빈혈증, 부신백질이영양증 (ALD), 및 이염성 백질 이영양증 (MLD)을 포함하는, 리소솜 축적병과 같은 그외 유전 장애를 포함하는 많은 장애에 대한 매우 매력적인 치료 옵션으로서 조혈 줄기 세포 (HSC) 유전자 치료, 예를 들어, 바이러스-매개 조혈 줄기 세포 (HSC) 유전자 치료법을 논의한다.

[1368] Cellectis로 양도된 미국 특허 공개 제20110225664호, 제20110091441호, 제20100229252호, 제20090271881호 및 제20090222937호는 CREI 변이체에 관한 것으로, 여기서 2 개의 I-CreI 단량체 중 적어도 하나는 적어도 2 개의

치환을 가지고 있으며, LAGLIDADG (SEQ ID NO: 26) 코어 도메인의 2 개의 작용성 서브도메인 각각에서의 하나는 I-CreI의 위치 26 내지 40 및 44 내지 77에 각각 위치하며, 상기 변이체는 공동 사이토카인 수용체 감마 사슬 유전자 또는 감마 C 유전자로도 명명된 인간 인터루킨-2 수용체 감마 사슬(IL2RG) 유전자으로부터 DNA 표적 서열을 절단할 수 있다. 표적 서열 미국 특허 공개 제20110225664호, 제20110091441호, 제20100229252호, 제20090271881호 및 제20090222937호에서 확인된 표적 서열이 본 발명의 핵산-표적화 시스템에서 사용될 수 있다.

[1369] 중증 복합형 면역 결핍증 (SCID)은 림프구 B에서의 작용성 결함과 항상 관련되는 림프구 T 성숙에서의 결함으로부터 야기한다 (문헌 [Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). 전체 발생은 75 000명의 출생 중 1명으로 평가된다. 치료되지 않은 SCID를 갖는 환자는 다수의 기회적 미생물 감염이 걸릴수 있으며, 일반적으로 1년 이후 생존하지 않는다. SCID는 가족 기증자로부터 동종 이계 조혈 줄기 세포 이송에 의해 치료될 수 있다. 기증자와의 조직적합성은 광범위하게 변할 수 있다. SCID 형태 중 하나인 아데노신 디아미나제(Deaminase:ADA) 결핍증의 경우에, 환자는 재조합 아데노신 디아미나제 효소의 주사에 의해 치료될 수 있다.

[1370] ADA 유전자가 SCID 환자에서 돌연변이 된 것으로 보여진 이후로 (문헌 [Giblett et al., Lancet, 1972, 2, 10671069]), SCID와 관련된 여러 다른 유전자가 확인되었다 (문헌 [Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). SCID에 대한 4 개의 주요 원인이 있다:(i) SCID의 가장 빈번한 형태인, SCID-X1 (X-관련 SCID 또는 X-SCID)은 IL2RG 유전자에서의 돌연변이에 의해 발생하며, 이는 성숙 T 림프구 및 NK 세포의 부재를 야기한다. IL2RG는 적어도 5 개 인터루킨 수용체 복합체의 공통 성분인 감마 C 단백질을 코딩한다 (문헌 [Noguchi, et al., Cell, 1993, 73, 147157]). 이들 수용체는 JAK3 키나제를 통해 몇몇 표적을 활성화시키는데 (Macchi et al., Nature, 1995, 377, 65-68), 이의 불활성화는 감마 C 불활성화와 동일한 증후군을 초래하고; (ii) ADA 유전자의 돌연변이는 림프종 전구체에 치명적이어서, 이후 B, T, 및 NK 세포의 유사 부재를 야기시키는 푸린 물질대사의 결함을 야기하고; (iii) V(D)J 재조합은 면역글로불린 및 T 림프구 수용체 (TCR)의 성숙화에서 필수 단계이다. 이러한 과정에 관여되는 3개 유전자, 재조합 활성화 유전자 1 및 2 (RAG1 및 RAG2) 및 Artemis 내 돌연변이는 성숙한 T 및 B 림프구의 부재를 야기하고; (iv) T 세포 특이적 신호전달에 관여하는, CD45 같은 다른 유전자 내 돌연변이가 또하나 보고되었지만, 그들은 소수 사례를 대표한다 (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). 그들의 유전적 기반이 확인된 때로부터, 상이한 SCID 형태는 두가지 주요 이유로 인해 유전자 치료 시도에 대한 통상적인 예가 된다 (문헌 [Fischer et al., Immunol. Rev., 2005, 203, 98-109). 첫번째로는, 모든 혈액 질환에서와 같이, 생체 외 치료가 고려될 수 있다. 조혈성 줄기 세포 (HSC)는 골수로부터 회복될 수 있으며, 그의 다능성 특성을 수 회의 세포 분열 동안 유지한다. 따라서, 이들은 시험관 내에서 처리될 수 있으며, 이후 환자에게 다시 주사될 수 있어서, 여기서 이들은 골수를 다시 채운다. 두번째로, 림프구의 성숙은 SCID 환자에서 손상되기 때문에, 편집된 세포는 선택적 장점을 갖는다. 따라서, 적은 수의 편집된 세포가 작용성 면역계를 회복할 수 있다. 이러한 가정은 (i) SCID 환자에서 돌연변이 반전과 연관된 면역 기능의 부분적 복원 (Hirschhorn et al., Nat. Genet., 1996, 13, 290-295; Stephan et al., N. Engl. J. Med., 1996, 335, 1563-1567; Bousso et al., Proc. Natl., Acad. Sci. USA, 2000, 97, 274-278; Wada et al., Proc. Natl. Acad. Sci. USA, 2001, 98, 8697-8702; Nishikomori et al., Blood, 2004, 103, 4565-4572), (ii) 조혈 세포에서 시험관 내 SCID-X1 결핍의 교정 (Candotti et al., Blood, 1996, 87, 3097-3102; Cavazzana-Calvo et al., Blood, 1996, Blood, 88, 3901-3909; Taylor et al., Blood, 1996, 87, 3103-3107; Hacein-Bey et al., Blood, 1998, 92, 4090-4097), (iii) SCID-X1 (Soudais et al., Blood, 2000, 95, 3071-3077; Tsai et al., Blood, 2002, 100, 72-79), JAK-3 (Bunting et al., Nat. Med., 1998, 4, 58-64; Bunting et al., Hum. Gene Ther., 2000, 11, 2353-2364) 및 RAG2 (Yates et al., Blood, 2002, 100, 3942-3949) 결핍의 동물 모델에서 생체내 교정 및 (iv) 유전자 요법 임상 시험의 결과 (Cavazzana-Calvo et al., Science, 2000, 288, 669-672; Aiuti et al., Nat. Med., 2002; 8, 423-425; Gaspar et al., Lancet, 2004, 364, 2181-2187)에 의해 수차례 검증되었다.

[1371] Children's Medical Center Corporation 및 President and Fellows of Harvard College에 의해 양도된 미국 특허 공개 제20110182867호는 RNAi 및 항체와 같은 BCL11A 발현 또는 활성의 억제제를 통해 조혈 전구체 세포에서 태아 헤모글로빈 발현(HbF)을 조절하는 방법 및 용도에 관한 것이다. BCL11a와 같은 미국 특허 공개 제20110182867호에 개시된 표적은 태아 헤모글로빈 발현을 조절하기 위한 본 발명의 CRISPR Cas 시스템에 의해 표적될 수 있다. 또한 추가의 BCL11A 표적에 대해 다음을 참조한다: Bauer et al. (Science 11 October 2013: Vol. 342 no. 6155 pp. 253-257) 및 Xu et al. (Science 18 November 2011: Vol. 334 no. 6058 pp.

993-996).

[1372] 해당 분야의 지식 및 본 개시 내용의 교시를 이용하여, 당업자는 본 명세서에 개시된 C2c1-CRISPR 시스템 및 상기 기술된 바와 같은 CrISPR-Cas 시스템을 사용해 유전적 조혈 장애, 예를 들어, β-지중해빈혈, 혈우병, 또는 유전적 리소좀 축적병에 대해 HSC를 편집할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1373] **HSC - 조혈 줄기 세포로의 전달 및 이의 편집; 및 특정 병태**

[1374] 용어 "조혈 줄기 세포" 또는 " HSC" 는 HSC, 예를 들어 모든 다른 혈액 세포를 발생시키고 중간엽으로부터 유래되고; 대부분의 뼈 중심에 함유된 적색 골수에 위치하는, 혈액 세포로 간주되는 광범위한 세포를 포함하는 것을 의미한다. 본 발명의 HSC는 소형 크기, 계통 결여 (lin) 마커, 및 CD34, CD38, CD90, CD133, CD105, CD45, 및 또한 c-kit, - 줄기 세포 인자의 수용체와 같은, 분화 시리즈의 클러스터에 속하는 마커로 확인되는, 조혈 줄기 세포의 표현형을 갖는 세포를 포함한다. 조혈 줄기 세포는 계통 관련성의 검출에 사용되는 마커에 음성이고, 따라서, Lin- 라고 하며; FACS에 의한 그들 정제 동안, 다수의 최대 14개의 상이한 성숙한 혈액-계통 마커, 예를 들어 인간의 경우에, 골수에 대한 CD13 & CD33, 적혈구에 대한 CD71, B 세포에 대한 CD19, 거핵세포에 대한 CD61; 및 B 세포에 대한 B220 (쥐과 CD45), 단핵구에 대한 Mac-1 (CD11b/CD18), 과립구에 대한 Gr-1, 적혈 세포에 대한 Ter119, T 세포에 대한 I17Ra, CD3, CD4, CD5, CD8 등, 마우스 HSC 마커: CD34lo/-, SCA-1+, Thy1.1+/lo, CD38+, C-kit+, lin-, 및 인간 HSC 마커: CD34+, CD59+, Thy1/CD90+, CD38lo/-, C-kit/CD117+, 및 lin- 가 있다. HSC는 마커를 통해 확인된다. 그러므로 본 명세서에 논의된 구현예에서, HSC는 CD34+ 세포일 수 있다. HSC는 또한 CD34-/CD38- 인 조혈 줄기 세포이다. HSC로서 작용하는 것으로 간주되는 세포 표면 상의 c-kit가 결여될 수 있는 줄기 세포는 본 발명의 영역 내에 있을뿐만 아니라, 유사하게 CD133+ 세포도 당분야에서 HSC로 간주한다.

[1375] CRISPR-Cas (eg C2c1) 시스템은 HSC에서 유전자 좌위 또는 좌위들을 표적으로 삼도록 조작될 수 있다. 유리하게, 진핵생물 세포 및 특히 포유동물 세포, 예를 들어 인간 세포, 예를 들어, HSC에 대해 코돈-최적화된 Cas (eg C2c1) 단백질, 및 HSC의 유전자좌 또는 유전자좌들, 예를 들어, 유전자 EMX1을 표적화하는 sgRNA가 제조될 수 있다. 이들은 입자를 통해 전달될 수 있다. 입자는 Cas (eg C2c1) 단백질 및 gRNA를 혼합하여 형성될 수 있다. gRNA 및 Cas (eg C2c1) 단백질 혼합물은 예를 들어 계면활성제, 인지질, 생분해성 중합체, 지단백질, 및 알콜을 포함하거나 또는 그로 본질적으로 이루어지거나 또는 이루어진 혼합물과 혼합될 수 있고, 그리하여 gRNA 및 Cas (eg C2c1) 단백질을 함유하는 입자가 형성될 수 있다. 본 발명은 이렇게 만든 입자 및 이러한 방법에 의한 입자를 비롯하여 이의 용도를 이해한다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1376] 보다 일반적으로, 입자는 효과적인 방법을 사용해 형성될 수 있다. 첫째로, Cas (eg C2c1) 단백질 및 유전자 EMX1을 표적화하는 gRNA 또는 대조군 유전자 EMX1 또는 대조군 유전자 LacZ는 적절하게, 예를 들어 3:1 내지 1:3 또는 2:1 내지 1:2 또는 1:1 몰비율로, 적절한 온도, 예를 들어 15-30C, 예를 들어, 20-25C, 예를 들어, 실온에서, 적절한 시간, 예를 들어 5-45, 예컨대 30분 동안, 유리하게 멸균, 뉴클레아제 무함유 완충액, 예를 들어 1X PBS 중에서 함께 혼합될 수 있다. 별도로, 하기와 같은 또는 하기를 포함하는 입자 성분: 계면활성제, 예를 들어, 양이온성 지질, 예를 들어, 1,2-디올레오일-3-트리메틸암모늄-프로판 (DOTAP); 인지질, 예를 들

어, 디미리스토일포스파티딜콜린 (DMPC); 생분해성 중합체, 예컨대 에틸렌-글리콜 중합체 또는 PEG, 및 지단백질, 예컨대 저밀도 지단백질, 예를 들어 콜레스테롤과 같은 또는 이를 포함하는 입자 성분을 알콜, 유리하게 C1-6 알킬 알콜, 예컨대 메탄올, 에탄올, 이소프로판올, 예를 들어 100% 에탄올에 용해시켰다. 2 가지 용액을 함께 혼합하여 Cas (예를 들어, C2c1)-gRNA 복합체를 함유하는 입자를 형성시킬 수 있다. 일정 구현예에서, 입자는 HDR 주형을 함유할 수 있다. 그것은 gRNA+Cas (eg C2c1) 단백질-함유 입자와 동시-투여되는 입자일 수 있거나, 또는 다시 말해서, HSC를 gRNA+Cas (eg C2c1) 단백질-함유 입자와 접촉시키는 것 이외에도, HSC를 HDR 주형을 함유하는 입자와 접촉시키거나, 또는 HSC는 모든 gRNA, Cas (eg C2c1) 및 HDR 주형을 함유하는 입자와 접촉된다. HDR 주형은 개별 벡터로 투여될 수 있는데, 그리하여 제1 예에서 입자는 HSC 세포르 투과하고 개별 벡터가 또한 세포를 투과하며, 여기서 HSC 게놈은 gRNA+Cas (eg C2c1)에 의해 변형되고 HDR 주형이 또한 존재하며, 그리하여 게놈 유전자좌는 HDR에 의해 변형되며; 예를 들어 이러한 결과로 돌연변이가 교정될 수 있다.

[1377] 입자 형성 후에, 96웰 플레이트 내 HSC는 웰 당 15 ug Cas (eg C2c1) 단백질이 형질감염될 수 있다. 형질감염 후 3일에, HSC를 수확할 수 있고, EMX1 유전자좌에서 삽입 및 결실 (indel)의 수를 정량할 수 있다.

[1378] 이것은 HSC가 HSC 내 관심 게놈 유전자좌 또는 유전자좌들을 표적으로 하는 CRISPR-Cas (eg C2c1)을 사용해 어떻게 변형될 수 있는가를 예시한다. 변형시키려는 HSC는 생체내일 수 있고, 즉 유기체, 예를 들어 인간 또는 비인간 진핵생물, 예를 들어 동물, 예컨대 어류, 예를 들어 제브라 피쉬, 포유동물, 예를 들어 영장류, 예를 들어, 유인원, 침팬지, 마카크 원숭이, 설치류, 예를 들어 마우스, 토끼, 래트, 개, 개과 또는 개, 가축 (소/소과, 양/양과, 염소 또는 돼지), 새 또는 가금류, 예를 들어 닭일 수 있다. 변형시키려는 HSC는 시험관내, 즉 이러한 유기체 밖일 수 있다. 그리고, 변형된 HSC는 생체외에서 사용할 수 있고, 즉 이러한 유기체의 하나 이상의 HSC는 유기체로부터 수득 또는 단리될 수 있고, 임의로 HSC(들)는 확장될 수 있으며, HSC(들)는 예를 들어, HSC(들)를 조성물과 접촉시켜서, HSC 내 유전자 좌위 또는 좌위들을 표적으로 하는 CRISPR-Cas (eg C2c1)을 포함하는 조성물에 의해 변형되고, 예를 들어, 여기서 조성물은 CRISPR 효소 및 HSC의 유전자 좌위 또는 좌위들을 표적으로 하는 하나 이상의 gRNA를 함유하는 입자를 포함하고, 이러한 입자는 계면활성제, 인지질, 생분해성 중합체, 지단백질, 및 알콜을 포함하거나 또는 그로 본질적으로 이루어지거나 또는 그로 이루어진 혼합물과 gRNA (하나 이상의 gRNA는 HSC의 유전자 좌위 또는 유전자좌를 표적으로 함) 및 Cas (eg C2c1) 단백질 혼합물과 혼합하여 수득되거나 또는 수득가능하고, 임의로 최종 변형된 HSC를 확장시키고 유기체에게 최종 변형된 HSC를 투여한다. 일부 예에서, 단리 또는 수득된 HSC는 제1 유기체, 예컨대 제2 유기체와 동일한 종으로부터의 유기체 유래일 수 있고, 제2 유기체는 최종 변형된 HSC가 투여되는 유기체일 수 있으며, 예를 들어 제1 유기체는 제2 유기체에 대한 도너 (예컨대 부모 또는 형제같은 친척)일 수 있다. 변형된 HSC는 개체 또는 대상체 또는 환자의 질환 증상 또는 병태 상태를 해결하거나 또는 완화시키거나 또는 감소시키기 위해 유전자 변형을 가질 수 있다. 예를 들어, 제2 유기체에 대한 제1 유기체 도너의 예에서, 변형된 HSC는 HSC가 하나 이상의 단백질, 예를 들어 제2 유기체의 것과 같은 표면 마커 또는 단백질을 갖도록 유전자 변형을 가질 수 있다.

변형된 HSC는 개체 또는 대상체 또는 환자의 질환 또는 병태 상태를 모의하도록 유전자 변형을 가질 수 있고 동물 모델을 준비하기 위해 비인간 유기체에게 재투여될 수 있다. HSC의 확장은 본 개시 및 당업자의 지식을 통해 당업자의 영역 내이며, 예를 들어, 다음을 참조한다: Lee, "Improved ex vivo expansion of adult hematopoietic stem cells by overcoming CUL4-mediated degradation of HOXB4." Blood. 2013 May 16;121(20):4082-9. doi: 10.1182/blood-2012-09-455204. Epub 2013 Mar 21.

[1379] 활성을 개선시키기 위해 표시된 바와 같이, sgRNA 는 입자에 전체 복합체를 제제화시키기 전에, Cas (eg C2c1) 단백질과 사전 복합체 형성될 수 있다. 제제는 세포로 핵산의 전달을 촉진하는 것으로 알려진 상이한 성분의 상이한 몰비로 제조될 수 있다 (예를 들어, 1,2-디올레오일-3-트리메틸암모늄프로판 (DOTAP), 1,2-디테트라데카노일-sn-글리세로-3-포스포콜린 (DMPC), 폴리에틸렌 글리콜 (PEG), 및 콜레스테롤). 예를 들어, DOTAP : DMPC : PEG : 콜레스테롤 몰비율은 DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0; 또는 DOTAP 90, DMPC 0, PEG 10, 콜레스테롤 0; 또는 DOTAP 90, DMPC 0, PEG 5, 콜레스테롤 5. DOTAP 100, DMPC 0, PEG 0, 콜레스테롤 0일 수 있다. 따라서 본 발명은 gRNA, Cas (eg C2c1) 단백질 및 입자를 형성하는 성분을 혼합하는 것 뿐 아니라; 이러한 혼합으로부터의 입자로 이해된다.

[1380] 바람직한 구현예에서, Cas (eg C2c1)-gRNA 복합체를 함유하는 입자는 Cas (eg C2c1) 단백질 및 하나 이상의 gRNA를 함께, 바람직하게 1:1 몰비율의 효소: 가이드 RNA로 혼합하여 형성될 수 있다. 개별적으로, 핵산의 전달을 촉진한다고 알려진 상이한 성분 (예를 들어, DOTAP, DMPC, PEG, 및 콜레스테롤)은 바람직하게 에탄올에 용해된다. 2 가지 용액을 함께 혼합하여 Cas (예를 들어, C2c1)-gRNA 복합체를 함유하는 입자를

형성시킨다. 입자가 형성된 후에, Cas (eg C2c1)-gRNA 복합체가 세포 (예를 들어, HSC)가 세포를 형질감염시킬 수 있다. 바코드가 적용될 수 있다. 입자, Cas 및/또는 gRNA는 바코드화될 수 있다.

[1381] 본 발명은 gRNA 및 Cas (eg C2c1) 단백질 혼합물을 계면활성제, 인지질, 생분해성 중합체, 지단백질, 및 알콜을 포함하거나 또는 그로 본질적으로 이루어지거나 또는 그로이루어진 혼합물과 혼합하는 단계를 포함하는 gRNA-및-Cas (eg C2c1) 단백질 함유 입자를 제조하는 방법을 포괄한다. 구현예는 방법으로부터의 gRNA-및-Cas (eg C2c1) 단백질 함유 입자를 포괄한다. 일 구현예에서 본 발명은 gRNA가 관심 게놈 유전자좌를 표적으로 하는 것인, 입자와 관심 게놈 유전자좌를 함유하는 세포를 접촉시키는 단계를 포함하는, 관심 게놈 유전자좌에서 표적 서열의 조작에 의한, 관심 게놈 유전자좌, 또는 유기체 또는 비인간 유기체를 변형시키는 방법; 또는 gRNA가 관심 게놈 유전자좌를 표적화하는 것인, 입자와 관심 게놈 유전자좌를 함유하는 세포를 접촉시키는 단계를 포함하는, 관심 게놈 유전자좌의 표적 서열의 조작에 의한 관심 게놈 유전자좌, 또는 유기체 또는 비인간 유기체의 변형 방법에서, 입자의 용도를 포괄한다. 이들 구현예에서, 관심 게놈 유전자좌는 유리하게 HSC 내 게놈 유전자좌이다.

[1382] 치료적 적용에서 고려사항: 게놈 편집 요법에서 고려사항은 서열-특이적 뉴클레아제, 예컨대 C2c1 뉴클레아제의 선택이다. 각각의 뉴클레아제 변이체는 그 자신의 고유한 세트의 강점 및 약점을 보유할 수 있고, 많은 이들은 치료적 이득을 극대화하기 위해 치료 상황에서 균형을 맞춰야만 한다. 따라서, 뉴클레아제를 사용하는 2개의 치료 편집 접근법은 유전자 파괴 및 유전자 교정이라는 상당한 유망성을 보여주었다. 유전자 파괴는 유전자 엘리먼트에서 표적화된 indel을 생성시키도록 NHEJ의 자극을 포함하고, 종종 환자에게 유리한 기능 상실 돌연변이를 생성시킨다. 대조적으로, 유전자 교정은 교정된 엘리먼트의 생리적 조절을 보존하면서 질환 초래 돌연변이를 직접적으로 반전시켜 기능을 복구하기 위해 HDR을 사용한다. HDR은 또한 잃어버린 유전자 기능을 회복하기 위해서 게놈 내 정해진 '세이프 하버' 유전자좌로 치료적 이식유전자를 삽입시키는데 사용될 수 있다. 특별한 편집 요법이 효과적이기 위해서, 충분하게 높은 수준의 변형이 질환 증상을 반전시키기 위해 표적 세포 개체군에서 달성되어야만 한다. 이러한 치료적 변형 '한계치'는 치료 후 편집된 세포의 피트니스 및 증상을 반전시키는데 필요한 유전자 산물의 양에 의해 결정된다. 피트니스와 관련하여, 편집은 그들의 미편집 대응부에 대해서 치료된 세포에 대해 3가지 잠재적인 결과: 증가, 중화, 또는 감소 피트니스를 일으킨다. 증가된 적합화 경우에, 예를 들어 SCID-X1의 치료에서, 변형된 조혈 전구체 세포는 그들의 미편집된 대응부에 비해서 선택적으로 확장된다. SCID-X1는 그 기능이 조혈 림프구 계통의 적절한 발생에 필요한, IL2RG 유전자 내 돌연변이로인해 초래된 질환이다 [Leonard, W.J., et al. Immunological reviews 138, 61-86 (1994); Kaushansky, K. & Williams, W.J. Williams hematology, (McGraw-Hill Medical, 뉴욕, 2010)]. SCID-X1, 및 SCID-X1 돌연변이의 자발적 교정의 희귀 예에 대한 바이러스 유전자 요법을 받은 환자의 임상 시험에서, 교정된 조혈 전구체 세포는 이러한 발생 차단을 극복할 수 있고 요법을 매개하도록 그들 질환 대응물에 대해 확장된다 [Bousso, P., et al. Proceedings of the National Academy of Sciences of the United States of America 97, 274-278 (2000); Hacein-Bey-Abina, S., et al. The New England journal of medicine 346, 1185-1193 (2002); Gaspar, H.B., et al. Lancet 364, 2181-2187 (2004)]. 이러한 경우에, 편집된 세포가 선택적 장점을 보유하는 경우에, 비록 작은 수의 편집된 세포가 확장을 통해 증폭될 수 있어도, 환자에게 치료적 이득을 제공한다. 대조적으로, 만성 육아종성 장애 (CGD) 같은 다른 조혈 질환을 위한 편집은 편집된 조혈 전구체 세포에 대한 적합화의 변화를 유도하지 않아서, 치료적 변형 한계치를 증가시킨다. CGD는 병원체를 사멸시키는 반응성 산소종을 발생시키기 위해 호중구에 의해 정상적으로 사용되는, 대식세포 옥시다제 단백질을 코딩하는 유전자 내 돌연변이로 인해 초래된다 [Mukherjee, S. & Thrasher, A.J. Gene 525, 174-181 (2013)]. 이들 유전자의 이상기능이 조혈 전구체 세포 피트니스 또는 발생에 영향을 미치지 않지만, 오직 감염과 싸우는 성숙한 조혈 세포 유형의 능력에 영향을 미치므로, 이러한 질환에서 편집된 세포의 우선적 확장은 아마도 없을 것이다. 실제로, CGD에서 유전자 교정된 세포에 대한 선택적 장점이 유전자 요법 시험에서 관찰되지 않았고, 장기간 세포 생착의 어려움을 야기시켰다 [Malech, H.L., et al. Proceedings of the National Academy of Sciences of the United States of America 94, 12133-12138 (1997); Kang, H.J., et al. Molecular therapy : the journal of the American Society of Gene Therapy 19, 2092-2101 (2011)]. 이와 같이, 유의하게 더 높은 수준의 편집이 CGD같은 질환을 치료하는데 요구되고, 여기서 편집은 편집이 표적 세포에 대해 증가된 피트니스를 생성시키는 질환에 비해서, 중성 피트니스 장점을 생성시킨다. 편집이 피트니스 단점을 부여하면, 암 세포에서 종양 억제인자 유전자에 기능을 복원시키기 위한 경우에서 처럼, 변형된 세포는 그들 질환 대응물이 능가하여, 편집 비율에 비해 치료 이득이 낮아지게 된다. 이러한 후자 부류의 질환은 게놈 편집 요법으로 치료하는데 특히 어렵다.

[1383] 세포 피트니스 이외에도, 질환을 치료하는데 필요한 유전자 산물의 양이 또한 증상을 반전시키기 위해 획득되어

야 하는 치료적 게놈 편집의 최소 수준에 영향을 미친다. 혈우병 B는 유전자 산물 수준의 작은 변화가 임상 결과에 상당한 변화를 일으킬 수 있는 질환 중 하나이다. 이러한 질환은 응고 캐스캐이드의 성분으로서 기능하고, 정상적으로 간에 의해 혈액으로 분배되는 단백질인 인자 IX를 코딩하는 유전자 내 돌연변이로 인해 초래된다. 혈우병 B의 임상 중증도는 인자 IX 활성량과 관련된다. 중증 질환이 1% 미만의 정상 활성과 연관되는 반면, 경증 형태의 질환은 1% 초과의 인자 IX 활성과 연관된다 [Kaushansky, K. & Williams, W.J. Williams hematology, (McGraw-Hill Medical, 뉴욕, 2010); Lofqvist, T., et al. Journal of internal medicine 241, 395-400 (1997)]. 이것은 적은 비율의 간세포로라도 인자 IX 발현을 복원할 수 있는 편집 요법은 임상 결과에 큰 영향을 미칠 수 있다는 것을 의미한다. 탄생 직후 혈우병 마우스 모델을 교정하기 위해 ZFN을 사용한 연구는 3-7% 교정이 질환 증상을 반전시키는데 충분하였고, 이러한 가설을 위한 전임상 증거를 제공하였음을 입증하였다 [Li, H., et al. Nature 475, 217-221 (2011)].

[1384] 유전자 산물 수준의 작은 변화가 임상 결과에 영향을 미치는 장애 및 편집된 세포에 피트니스 장점이 존재하는 질환은 게놈 편집 요법을 위한 이상적인 표적인데, 치료 변형 한계치가 현행 기술을 고려하여 높은 성공 기회를 허용하기에 충분히 낮기 때문이다. 이들 질환의 표적화는 이제 전임상 수준 및 I상 임상 시험에서 편집 요법에 의한 성공을 이끌었다. 편집된 세포에 대해 중성 피트니스 장점을 갖는 질환, 또는 많은 양의 유 전자 산물이 치료에 필요한 경우까지 이러한 유망한 결과를 확장시키는데 DSB 복구 경로 조작 및 뉴클레아제 전달의 개선이 필요하다. 하기 표 6은 치료 모델에 대한 게놈 편집 적용의 일부 예를 보여주고, 하기 표 6의 참조 및 그들 참조에 인용된 문헌들은 완전히 기재된 바와 같이 참조로 본 명세서에 편입된다.

질환 유형	적용된 뉴클레아제 플랫폼	치료 전략	참조 문헌
혈우병 B	ZFN	교정 유전자 서열의 HDR-매개 삽입	Li, H., et al. Nature 475, 217-221 (2011)
SCID	ZFN	교정 유전자 서열의 HDR-매개 삽입	Genovese, P., et al. Nature 510, 235-240 (2014)
유전성 티로신혈증	CRISPR	간에서 돌연변이의 HDR-매개 교정	Yin, H., et al. Nature Biotechnology 32, 551-553 (2014)

[1385]

[1386] 유리하게 본 명세서에서와 같은 전달 시스템, 예를 들어 입자 전달 시스템을 통한, 돌연변이의 HDR-매개 교정, 또는 교정된 유전자 서열의 HDR-매개 삽입에 의해 표적에 대해 CRISPR-Cas (eg C2c1) 시스템을 사용하여, 전술한 표의 각 병태의 해결은 본 시 및 당분야의 지식을 통해 당업자의 영역 내이다. 따라서, 일 구현예는 혈우병 B, SCID (예를 들어, SCID-X1, ADA-SCID) 또는 유전적 티로신혈증 돌연변이-보유 HSC와 혈우병 B, SCID (예를 들어, SCID-X1, ADA-SCID) 또는 유전적 티로신혈증에 대한 관심 게놈 유전자좌를 표적화하는 입자를 함유하는 gRNA-및-Cas (eg C2c1) 함유 입자를 접촉시키는 단계를 포괄한다 (Li, Genovese or Yin). 또한 입자는 돌연변이를 교정하기 위해 적합한 HDR 주형을 함유할 수 있거나; 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다. 이와 관련하여, 혈우병 B는 응고 캐스캐이드의 핵심 성분인, 인자 Ix를 코딩하는 유전자 내 기능 상실 돌연변이로 인해 초래되는 X-연결 열성 장애라는 것을 언급한다. 중증 발병 개체에서 이의 수준의 1% 이상까지 인자 IX 활성의 회복은 질환을 상당히 더 경증의 형태로 전환시킬 수 있는데, 이러한 수준을 달성하기 위해 어린 나이부터 예방적으로 이러한 환자에게 재조합 인자 Ix의 주입이 대체로 임상 합병증을 완화시키기 때문이다. 당분야의 지식 및 본 개시의 교시를 통해서, 당업자는 돌연변이를 표적화하고 교정하는 CRISPR-Cas (eg C2c1) 시스템을 사용하여 혈우병 B (인자 IX를 코딩하는 유전자 내 기능 상실 돌연변이로 초래된 X-연관 열성 장애)에 대해서 (예를 들어, 인자 Ix에 대한 코디이 서열을 전달하는 적합한 HDR 주형을 사용해) HSC를 교정할 수 있고; 특히 gRNA는 혈우병 B을 발생시키는 돌연변이를 표적화할 수 있고, HDR은 인자 XI의 적절한 발현을 위한 코딩을 제공할 수 있다. 돌연변이-및-Cas (eg C2c1) 단백질 함유 입자를 표적화하는 가이드 RNA는 돌연변이를 운반하는 HSC와 접촉된다. 또한 입자는 인자 의 적절한 발현을 위해 돌연변이를 교정하기 위해 적합한 HDR 주형을 함유할 수 있거나; 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다. 이렇게 접촉된 세포를 투여할 수 있고; 임의로 치료/확장된다; 본 명세서에 논의된, Cartier 참조.

[1387] 전체로 기재된 바와 같이, 인용하는 문헌과 함께 참조로 본 명세서에 편입되는, [Cartier, "MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-Linked Adrenoleukodystrophy," Brain Pathology 20 (2010) 857-862]에서, 동종이계 조혈

줄기 세포 이식 (HSCT)이 후를러 질환 환자의 뇌로 정상 리소솜 효소를 전달하기 위해 이용되었다는 인식, 및 ALD를 치료하기 위한 HSC 유전자 요법의 논의가 존재한다. 2명 환자에서, 말초 CD34+ 세포를 과립구-콜로니 자극 인자 (G-CSF) 고정 후 수집하였고 골수증식성 육종 바이러스 인핸서, 음성 대조군 영역 결실, d1587rev 프라이머 결합 부위 치환된 (MND)-ALD 렌티바이러스 벡터로 형질도입시켰다. 환자로부터의 CD34+ 세포는 저농도로 사이토카인의 존재 하에서 16시간 동안 MND-ALD 벡터로 형질도입시켰다. 형질도입된 CD34+ 세포는 특히 3회 복제-적격 렌티바이러스 (RCL) 어세이에 포함된 다양한 안전성 시험을 5%의 세포에 대해 수행하기 위해 형질도입 후 냉동시켰다. CD34+ 세포의 형질도입 효율은 35% 내지 50% 범위이고 평균 렌티바이러스 통합 카피수는 0.65 내지 0.70이었다. 형질도입된 CD34+ 세포의 해동 후, 환자는 부설판 및 사이클로포스파미드로 완전 골수절제 후 4.106 초과의 CD34+ 세포/kg가 재주입되었다. 환자의 HSC는 유전자-교정된 HSC의 생착에 유리하도록 절제되었다. 2명 환자의 경우에 혈액학적 회복은 13일 내지 15일에 발생되었다. 거의 완전한 면역학적 회복이 제1 환자의 경우에 12개월에 일어났고 제2 환자는 9개월에 일어났다. 렌티바이러스 사용과 대조적으로, 당분야의 지식 및 본 개시의 교시에 따라서, 당업자는 (예를 들어, 적합한 HDR 주형을 사용해) 돌연변이를 표적화 및 교정하는 CRISPR-Cas (C2c1) 시스템을 사용해 ALD에 대해 HSC를 교정할 수 있고; 특히 gRNA는 퍼옥시솜 막 수송체 단백질, ALD를 코딩하는 X 염색체 상에 존재하는 유전자인, ABCD1 내 돌연변이를 표적화할 수 있고, HDR은 단백질의 적절한 발현을 위한 코딩을 제공한다. 돌연변이-및-Cas (C2c1) 단백질 함유 입자를 표적화하는 가이드 RNA는 Cartier에서 처럼, HSC, 예를 들어 돌연변이를 운반하는 CD34+ 세포와 접촉된다. 또한 입자는 퍼옥시솜 막 수송체 단백질의 발현을 위해 돌연변이를 교정하도록 적합한 HDR 주형을 함유할 수 있거나; 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다.
이렇게 접촉된 세포는 임의로 Cartier에서 처럼 처치될 수 있다. 이렇게 접촉된 세포는 임의로 Cartier에서 처럼 투여될 수 있다.

[1388] WO 2015/148860을 언급하며, 본 명세서의 교시를 통해서 본 발명은 본 명세서의 교시와 함께 적용되는 이들 문헌의 재료 및 방법을 이해한다. 혈액-관련 질환의 양상에서, 베타 지중해 빈혈을 치료하기 위한, 유전자 요법, 방법, 및 조성물은 본 발명의 CRISPR-Cas 시스템에 대해 적합화될 수 있다 (참조: 예를 들어 WO 2015/148860). 일 구현예에서, WO 2015/148860은 예를 들어 B-세포 CLL/림프종11A (BCL11A)에 대한 유전자를 변경시킴으로써, 베타 지중해 빈혈 또는 이의 증상의 치료 또는 예방을 포함한다. BCL11A 유전자는 또한 B-세포 CLL/림프종 11A, BCL11A -L, BCL11A -S, BCL11AXL, CTIP 1, HBFQTL5 및 ZNF로서 알려져 있다. BCL11A는 글로빈 유전자 발현의 조절에 관여하는 아연-핑거 단백질을 코딩한다. BCL11A 유전자 (예를 들어, BCL11A 유전자의 한쪽 또는 양쪽 대립유전자)를 변경시켜서, 감마 글로빈의 수준을 증가시킬 수 있다. 감마 글로빈은 헤모글로빈 복합체의 베타 글로빈을 치환시킬 수 있고 효과적으로 산소를 조직으로 운반하여서, 베타 지중해 빈혈 질환 표현형을 완화시킨다.

[1389] WO 2015/148863을 언급하며, 본 명세서의 교시를 통해서 본 발명은 본 명세서의 CRISPR-Cas 시스템에 적합화될 수 있는 이들 문헌의 방법 및 재료를 포괄한다. 유전적 혈액학적 질환인, 겸상 적혈구 질환의 치료 및 예방의 일 양상에서, WO 2015/148863은 BCL11A 유전자의 변경을 포괄한다. BCL11A 유전자 (예를 들어, BCL11A 유전자의 한쪽 또는 양쪽 대립유전자)를 변경시켜서, 감마 글로빈의 수준을 증가시킬 수 있다. 감마 글로빈은 헤모글로빈 복합체의 베타 글로빈을 치환시킬 수 있고 효과적으로 산소를 조직으로 운반하여서, 겸상 적혈구 질환 표현형을 완화시킨다.

[1390] 본 발명의 일 양상에서, 표적 핵산 서열의 편집, 또는 표적 핵산 서열의 발현 조절을 포함한 방법 및 조성물, 및 암 면역요법과 관련하여 이의 적용이 본 발명의 CRISPR-Cas 시스템을 적합화하여 포괄된다. 하나 이상의 T-세포 발현된 유전자, 예를 들어 FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC 및/또는 TRBC 유전자 중 하나 이상을 변경시켜서, T-세포 증식, 생존, 및/또는 기능에 영향을 미치는데 사용될 수 있는 방법 및 조성물을 포함하는 WO 2015/161276에서의 유전자 요법의 적용을 참조한다. 관련 양상에서, T-세포 증식은 하나 이상의 T-세포 발현된 유전자, 예를 들어 CBLB 및/또는 PTPN6 유전자, FAS 및/또는 BID 유전자, CTLA4 및/또는 PDCDI 및/또는 TRAC 및/또는 TRBC 유전자를 변경시켜 영향받을 수 있다.

[1391] 키메라 항원 수용체 (CAR)19 T-세포는 환자 악성종에서 항-백혈병 효과를 나타낸다. 그러나, 백혈병 환자는 종종 수집하기에 충분한 T-세포를 갖지 않는데, 치료가 도너로부터의 변형된 T 세포를 포함시켜야한다는 것을 의미한다. 따라서, 도너 T-세포 은행의 확립에 관심이 존재한다. Qasim 등 ("First Clinical Application of Talen Engineered Universal CAR19 T Cells in B-ALL" ASH 57th Annual Meeting and Exposition, Dec. 5-8, 2015, Abstract 2046 (ash.confex.com/ash/2015/webprogram/Paper81653.html published online November 2015)은 T-세포 수용체 발현의 파괴 및 CD52 표적화를 통해 이식편 대 숙주 질환의

위험성을 제거하기 위한 CAR19 T 세포의 변형을 논의한다. 더 나아가서, CD52 세포는 알렘투주맙에 불감성이 되도록 표적화되었고, 따라서 알렘투주맙이 인간 백혈구 항원 (HLA) 미스매치된 CAR19 T-세포의 숙주-매개 거부를 예방할 수 있게 한다. 연구자는 RQR8에 연결된 4g7 CAR19 (CD19 scFv-4-1BB-CD3ζ)를 코딩하는 3세대 자가-불활성화 렌티바이러스 벡터를 사용하였고, 그 다음으로 T-세포 수용체 (TCR) 알파 불변 사슬 유전자좌 및 CD52 유전자좌 둘 모두에 대한 다중 표적화를 위해 2쌍의 TALEN mRNA로 세포를 전기천공시켰다. 생체외 확장 이후에도 여전히 TCR을 발현하는 세포는 CliniMacs α/β TCR 고갈을 사용해 고갈시켜서, < 1% TCR 발현의 T-세포 산물 (UCART19)을 산출하였는데, 이의 85%는 CAR19를 발현하였고, 64%는 CD52 음성이 되었다. 변형된 CAR19 T 세포는 환자의 재발성 급성 림프아구성 백혈병을 치료하기 위해 투여되었다. 본 명세서에서 제공되는 교시는 제한없이 T 세포, B 세포, 단핵구, 마크로파지, 호중구, 호염기구, 호산구, 적혈구, 수지상 세포, 및 거핵세포 또는 혈소판, 자연 살해 세포 및 그들 전구체 및 선구체를 포함하여, 골수 및 림프 계통의 혈액 세포를 포함해, 변형된 조혈 줄기 세포 및 이의 자손을 제공하는 효과적인 방법을 제공한다. 이러한 세포는 상기 기술된 바와 같이, 예를 들어 CD52, 및 다른 표적, 예컨대, 제한없이, CXCR4 및 Pd-1을 제거 또는 조절하기 위해서, 표적을 녹아웃, 녹인, 또는 달리 조절하여 변형될 수 있다. 따라서 본 발명의 조성물, 세포, 및 방법은 면역 반응을 조절하고, 제한없이, 환자에게 T 세포 또는 다른 세포의 투여의 변형과 함께, 악성종, 바이러스 감염, 및 면역 장애를 치료하기 위해 사용될 수 있다.

[1392] WO 2015/148670을 언급하며, 본 명세서의 교시를 통해서 본 발명은 본 명세서의 교시와 함께 적용되는 이들 문헌의 재료 및 방법을 이해한다. 면역 요법의 양상에서 인간 면역결핍 바이러스 (HIV) 및 후천성 면역결핍 증후군 (AIDS)와 관련하여 또는 그와 함께 표적 서열의 편집을 위한 방법 및 조성물이 포괄된다. 관련 양상에서, 본 명세서에 기술된 방법은 C-C 케모카인 수용체 5형 (CCR5)에 대한 유전자에 하나 이상의 돌연변이의 도입에 의한, HIV 감염 및 AIDS의 예방 및 치료를 포괄한다. CCR5 유전자는 또한 CKR5, CCR-5, CD195, CKR-5, CCCKR5, CMKBR5, IDDM22, 및 CC-CKR-5로도 공지되어 있다. 추가 양상에서, 본 명세서에 기술되니 방법은 HIV 감염의 예방 또는 감소 및/또는 예를 들어, 이미 감염된 대상체의 숙주 세포로 진입하는 HIV의 능력의 예방 및/또는 감소를 위해 제공된다. HIV에 대한 예시적인 숙주 세포는 제한없이, CD4 세포, T 세포, 내장 연관 림프성 조직 (GALT), 마크로파지, 수지상 세포, 골수 전구 세포, 및 미세아교세포를 포함한다. 숙주 세포로 바이러스 진입은 바이러스 당단백질 gp41 및 gp12과 CD4 수용체 및 공수용체, 예를 들어 CCR5 둘 모두와의 상호작용을 필요로 한다. 공수용체, 예를 들어 CCR5가 숙주 세포의 표면에 존재하지 않으면, 바이러스는 결합할 수 없고 숙주 세포로 들어갈 수 없다. 따라서 질환의 진행이 방해된다. 예를 들어, 보호적 돌연변이 (예컨대 CCR5 델타 32 돌연변이)를 도입시켜, 숙주 세포의 CCR5를 녹아웃 또는 녹다운시켜서, 숙주 세포로 HIV 바이러스의 진입을 예방한다.

[1393] X-연관 만성 육아종성 질환 (CGD)는 대식세포 NADPH 옥시다제의 부재 또는 감소된 활성으로 인한 숙주 방어의 유전적 장애이다. (예를 들어, 대식세포 NADPH 옥시다제에 대한 코딩 서열을 전달하는 적합한 HDR 주형을 사용하여) 돌연변이 (대식세포 NADPH 옥시다제의 부재 또는 감소된 활성)를 표적화하고 교정하는 CRISPR-Cas (C2c1)를 사용하고; 특히 gRNA는 CGD (대식세포 NADPH 옥시다제 결핍)를 발생시키는 돌연변이를 표적화하고, HDR은 대식세포 NADPH 옥시다제의 적절한 발현을 위한 코딩을 제공할 수 있다. 돌연변이-및-Cas (C2c1) 단백질 함유 입자를 표적화하는 가이드 RNA는 돌연변이를 운반하는 HSC와 접촉된다. 또한 입자는 대식세포 NADPH 옥시다제의 적절한 발현을 위해 돌연변이를 교정하기 위해 적합한 HDR 주형을 함유할 수 있거나; 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다. 이렇게 접촉된 세포를 투여할 수 있고; 임의로 치료 /확장된다; cf. Cartier.

[1394] 판코니 빈혈: 적어도 15개 유전자의 돌연변이 (FANCA, FANCB, FANCC, FANCD1/BRCA2, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ/BACH1/BRIP1, FANCL/PHF9/POG, FANCM, FANCN/PALB2, FANCO/Rad51C, 및 FANCP/SLX4/BTBD12)가 판코니 빈혈을 초래할 수 있다. 이들 유전자로부터 생산된 단백질은 FA 경로로 알려진 세포 과정에 관여된다. FA 경로는 DNA 복제라고 하는 새로운 DNA 카피를 만드는 과정이 DNA 손상으로 인해 차단될 때 켜진다 (활성화된다). FA 경로는 손상 영역으로 소정 단백질을 보내서, DNA 복구가 촉발되어서 DNA 복제가 계속될 수 있다. FA 경로는 가닥간 교차-연결 (ICL)로 알려진 일정 유형의 DNA 손상에 특히 반응한다. ICL은 DNA의 반대 가닥 상의 2개 DNA 빌딩 블록 (뉴클레오티드)가 비정상적으로 함께 부착 또는 연결되어 DNA 복자 과정이 중지될 때 발생된다. ICL은 일정 암 요법 약물에 의한 치료 또는 체내에서 생산되는 독성 물질의 축적으로 초래될 수 있다. 8개 단백질이 함께 판코니 빈혈 그룹과 연관되어서 FA 코어 복합체로 알려진 복합체를 형성한다. FA 코어 복합체는 FANCD2 및 FANCI라고 하는 2개 단백질을 활성화시킨다. 이들 2개 단백질의 활성화는 DNA 복구 단백질을 ICL의 영역으로 가져가서 교차-연결을 제거할 수 있고 DNA 복제가 계속될 수 있다. FA 코어 복합체. 보다 특히, FA 코어 복합체는 FANCA, FANCB, FANCC, FANCE,

FANCF, FANCG, FANCL, 및 FANCM로 이루어진 핵 다중단백질 복합체이고, E3 유비퀴틴 리가제로서 기능하며, FANCD2 및 FANCI로 구성된 이종이량체인, ID 복합체의 활성화를 매개한다. 모노유비퀴틴화되면, FANCD1/BRCA2, FANCN/PALB2, FANCJ/BRIP1, 및 FANCO/Rad51C를 포함한 FA 경로의 하류에 고전적인 종양 억제인자와 상호작용하여서, 상동성 재조합 (Hr)을 통한 DNA 복구에 기여한다. FA 사려의 80 내지 90%가 FANCA, FANCC, 및 FANCG의 3개 유전자 중 하나의 돌연변이로 인한 것이다. 이들 유전자는 FA 코어 복합체의 성분을 생산하기 위한 명령을 제공한다. FA 코어 복합체와 연관된 이러한 유전자의 돌연변이는 복합체를 비기능성이게 만들고 전체 FA 경로를 파괴하게 될 것이다. 그 결과로, DNA 손상은 효율적으로 복구되지 않고 시간 경과에 따라 ICL이 축적된다. [Geiselhart, " Review Article, Disrupted Signaling through the Fanconi Anemia 경로 Leads to Dysfunctional Hematopoietic Stem Cell Biology: Underlying Mechanisms and Potential Therapeutic Strategies," Anemia Volume 2012 (2012), Article ID 265790, dx.doi.org/10.1155/2012/265790]는 생체내에서 HSC의 교정을 야기시킨 FANCC 유전자를 코딩하는 렌티바이러스의 대퇴부내 주사를 포함한 동물 실험 및 FA를 고찰한다. FA와 연관된 하나 이상의 돌연변이를 표적화하는 CRISPR-Cas (C2c1) 시스템, 예를 들어, gRNA(들) 및 HDR 주형(들)을 갖는 CRISPR-Cas (C2c1) 시스템을 사용하여, 각각 FA를 발생시키는 FANCA, FANCC, 또는 FANCG의 하나 이상의 돌연변이를 표적화하고 FANCA, FANCC 또는 FANCG 중 하나 이상의 교정된 발현을 제공하고, 예를 들어 gRNA는 FANCC에 대한 돌연변이를 표적화하고, HDR은 FANCC의 적절한 발현을 위한 코딩을 제공할 수 있다. 돌연변이(들)(예를 들어, Fa에 관여하는 하나 이상, 예컨대 FANCA, FANCC 또는 FANCG 중 어느 하나 이상에 대한 돌연변이(들))를 표적화하는 gRNA-및-Cas (C2c1) 단백질 함유 입자는 돌연변이(들)를 보유하는 HSC와 접촉한다. 또한 입자는 Fa에 관여하는 하나 이상의 단백질, 예컨대 FANCA, FANCC, 또는 FANCG 중 어느 하나 이상의 적절한 발현을 위해 돌연변이를 교정하도록 적합한 HDR 주형(들)을 함유할 수 있거나, 또는 HSC는 HDR 주형을 함유하거나 또는 전달하는 제2 입자 또는 벡터와 접촉할 수 있다. 이렇게 접촉된 세포를 투여할 수 있고; 임의로 치료 /확장된다; cf. Cartier.

[1395] (예를 들어, gRNA(들) 및 Cas(C2c1), 임의로 HDR 주형(들), 또는 HDR 주형(들)을 포함하는 것에 관하여; 예를 들어, 혈우병 B, SCID, SCID-X1, ADA-SCID, 유전성 티로신혈증, β-지중해빈혈, X-연관 CGD, 위스코트-알드리치 증후군, 판코니 빈혈, 부신백질이영양증 (ALD), 이염성 백질이영양증 (MLD), HIV/AIDS, 면역결핍 장애, 혈액병태, 또는 유전성 리소솜 저장 질환에 대해) 본 명세서에 논의된 입자는 유리하게 gRNA(들) 및 Cas(C2c1) 단백질 혼합물 (임의로 HDR 주형(들)을 함유하거나 또는 이러한 혼합물은 주형(들)에 대해 별개 입자가 바람직할 때 HDR 주형(들)만을 함유)을, 계면활성제, 인지질, 생분해성 중합체, 지단백질 및 알콜을 포함하거나 또는 그로 본질적으로 이루어지거나 또는 그로 이루어진 혼합물과 혼합하여 수득하거나 또는 수득가능하다 (여기서 하나 이상의 gRNA는 HSC에서 유전자 좌위 또는 좌위들을 표적화함).

[1396] 실제로, 본 발명은 특히 게놈 편집으로 조혈 유전 장애, 및 면역결핍 장애, 예컨대 유전적 면역결핍 장애를 특히 본 명세서에 논의된 입자의 사용을 통해 치료하는데 적합하다. 유전성 면역결핍은 본 발명의 게놈 편집 중재술이 성공적일 수 있는 질환이다. 그 이유는 면역 세포의 서브세트인 조혈 세포가 치료적으로 접근하다는 것을 포함한다. 그들은 신체로부터 제거될 수 있고 자가 또는 동종으로 이식될 수 있다. 또한, 일정 유전성 면역결핍 예를 들어, 중증 복합 면역결핍 (SCID)는 면역 세포의 증식성 단점을 생성시킨다. 희귀한, 자발적 '역' 돌연변이에 의해 SCID를 초래하는 유전자 병변의 교정은 심지어 하나의 림프구 전구체의 교정이라도 환자에서 면역 기능을 회복하는데 충분할 수 있음을 의미한다.../../../Users/t_kowalski/AppData/Local/Microsoft/Windows/Temporary Internet Files/Content.Outlook/GA8VY8LK/Treating SCID for Ellen.docx - _ENREF_1 참조: Bousso, P., et al. Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor. Proceedings of the National Academy of Sciences of the United States of America 97, 274-278 (2000). 편집된 세포에 대한 선택적 장점은 낮은 수준의 편집에서도 치료적 효과를 일으키게 한다. 본 발명의 이러한 효과는 헤모글로빈 결핍이 적혈구 전구체의 피트니스에 부정적으로 영향을 미치는 알파- 및 베타-지중해 빈혈같은 다른 조혈 장애를 포함하여, 본 명세서에 언급된 SCID, 위스코트-알드리치 증후군 및 다른 병태에서 확인할 수 있다.

[1397] NHEJ 및 HDR DSB 복구의 활성은 세포 유형 및 세포 상태에 따라 유의하게 다양하다. NHEJ는 세포 주기에 의해 고도로 조절되지 않고 세포 유형에 걸쳐 효율적이어서, 접근가능한 표적 세포 개체군에서 높은 수준의 유전자 파괴를 허용한다. 대조적으로, HDR은 S/G2 시기 동안 주로 작용하고, 그러므로 활동적으로 분열하는 세포에 제한되어서, 정확한 게놈 변형을 요구하는 치료를 유사분열 세포에 국한시킨다 [Ciccia, A. & Elledge, S.J. Molecular cell 40, 179-204 (2010); Chapman, J.R., et al. Molecular cell 47, 497-510 (2012)]. 특히, C2c1 단백질을 포함하는 CRISPR-C2c1 시스템은 표적 부위에서 스태거드 절단부를 생성시킨다. 그러

므로, 본 발명에서 표적 서열의 절단, 변형, 및/또는 복구는 HDR 의존적이거나 또는 독립적일 수 있다. 특정 구현예에서, CRISPR-C2c1 시스템은 NHEJ를 통해 스태거드 DSB 복구를 도입시킨다. 일정한 특정 구현예에서, 본 발명의 CRISPR-C2c1 시스템은 뉴런 같은 비분열 세포에서 NHEJ를 통해 스태거드 DSB 복구를 도입시킨다.

[1398] HDR을 통한 교정의 효율은 후생적 상태 또는 표적화된 유전자좌의 서열, 또는 사용되는 특이적 복구 주형 구성(단일 대 이중 가닥, 장형 대 단형 상동성 팔부)에 의해 제어될 수 있다[Hacein-Bey-Abina, S., et al. The New England journal of medicine 346, 1185-1193 (2002); Gaspar, H.B., et al. Lancet 364, 2181-2187 (2004); Beumer, K.J., et al. G3 (2013)]. 표적 세포에서 NHEJ 및 HDR 기구의 상대적 활성이 또한 이들 경로가 DSB를 분해하기 위해 경쟁할 수 있으므로, 유전자 교정 효율에 영향을 미칠 수 있다 [Beumer, K.J., et al. Proceedings of the National Academy of Sciences of the United States of America 105, 19821-19826 (2008)]. HDR은 또한 뉴클레아제 및 복구 주형의 동시 전달을 요구하므로, NHEJ 전략에서 보이지 않은 전달 도전을 부여한다. 실제로, 이들 제한은 치료적으로 관련된 세포 유형에서 낮은 수준의 HDR을 이끌었다. 그러므로 임상적 해석은 질환을 치료하기 위해 대체로 NHEJ 전략에 집중되었지만, 개념 증명 전임상 HDR 치료가 이제 혈우병 B 및 유전성 티로신혈증의 마우스 모델에 대해 설명되었다 [Li, H., et al. Nature 475, 217-221 (2011); Yin, H., et al. Nature biotechnology 32, 551-553 (2014)].

[1399] 임의의 소정 게놈 편집 적용은 단백질, 소형 RNA 분자, 및/또는 복구 주형의 조합을 포함할 수 있고, 이들 다수 부분의 전달을 소형 분자 치료제에 비해 실질적으로 더욱 도전적이게 만든다. 게놈 편집 도구의 전달을 위한 주요 2가지 전략, 생체외 및 생체내가 개발되었다. 생체외 치료에서, 질환 세포는 신체에서 제거되어, 편집된 후에 다시 환자에게 이식된다. 생체외 편집은 표적 세포 개체군을 충분히 한정되게 하고 특별한 용량의 치료 분자가 특정된 세포에게 전달되게 하는 장점을 갖는다. 후자의 고려사항은 뉴클레아제의 양 적정이 이러한 돌연변이를 감소시킬 수 있으므로, 오프-표적 변형이 문제일 때 특히 중요하다 (Hsu et al., 2013). 생체외 접근법의 다른 장점은 전형적으로 연구 및 유전자 요법 적용을 위해 배양 세포로 단백질 및 핵산을 위한 효율적인 전달 시스템의 개발로 인해, 달성될 수 있는 것보다 높은 편집율이다.

[1400] 소수 질환에 적용에 제한된다는 것이 생체외 접근법의 단점일 수 있다. 예를 들어, 표적 세포는 체외 조작에서 생존할 수 있어야 한다. 뇌와 같은 많은 조직의 경우, 신체 밖에서 세포의 배양은 주요한 도전인데, 세포가 생존에 실패하거나, 또는 생체내 그들 기능에 필요한 성질을 상실하기 때문이다. 따라서, 본 개시 및 당분야의 지식 관점에서, 생체외 배양 및 조작을이 가능한 성체 줄기 세포 개체군의 조직에 대한, CRISPR-Cas (C2c1)에 의한 생체외 요법이 이용가능하다. [Bunn, H.F. & Aster, J. Pathophysiology of blood disorders, (McGraw-Hill, 뉴욕, 2011)]

[1401] 생체내 게놈 편집은 그들 천연 조직 내 세포 유형에게 편집 시스템의 직접 전달을 포함한다. 생체내 편집은 질환 세포 개체군이 생체외 조작으로 처리할 수 없는 질환을 치료할 수 있게 한다. 더 나아가서, 인시츄로 세포에 뉴클레아제 전달은 다수 조직 및 세포 유형의 치료를 허용한다. 이들 성질은 아마도 생체외 요법에 비해 더 광범위한 질환에 생체내 치료를 적용할 수 있게 한다.

[1402] 지금가지, 생체내 편집은 대체로 한정된, 조직-특이적 향성을 갖는 바이러스 벡터의 사용을 통해 달성되었다. 이러한 벡터는 현재 카고 운반 능력 및 향성 관점에서 제한적이어서, 임상적으로 유용한 벡터에 의한 형질도입이 효율적인 장기 시스템, 예컨대 간, 근육, 및 눈으로 이러한 방식의 요법을 제한한다 [Kotterman, M.A. & Schaffer, D.V. Nature reviews. Genetics 15, 445-451 (2014); Nguyen, T.H. & Ferry, N. Gene therapy 11 Suppl 1, S76-84 (2004); Boye, S.E., et al. Molecular therapy : the journal of the American Society of Gene Therapy 21, 509-519 (2013)].

[1403] 생체내 전달을 위한 잠재적인 장벽은 치료에 필요한 대량의 바이러스에 반응하여 생성될 수 있는 면역 반응이지만, 이러한 현상은 게놈 편집에만 고유한 것이 아니고 다른 바이러스 기반 유전자 요법에서도 관찰된다 [Bessis, N., et al. Gene therapy 11 Suppl 1, S10-17 (2004)]. 또한 편집 뉴클레아제로부터의 펩티드는 그 자체가 MHC 클래스 I 분자 상에 제시되어 면역 만응을 자극시키는 것이 가능하지만, 전임상 수준에서 이러한 해프닝을 뒷받침할 증거가 적다. 이러한 방식의 요법에서 다른 주요한 어려움은 예측이 어려울 수 있는 오프-표적 돌연변이 프로파일을 야기시키는, 분포 및 결과적으로 생체내 게놈 편집 뉴클레아제의 용량의 제어이다.

그러나, 암의 치료에서 사용되는 바이러스- 및 입자-기반 요법의 사용을 포함하여, 당분야의 지식 및 본 개시의 관점에서, 예를 들어 입자 또는 바이러스에 의한 전달에 의한 HSC의 생체내 변형이 당업자의 영역 내에 있다.

[1404] 생체외 편집 요법: 조혈 세포의 정제, 배양, 및 이식에 대한 장기간의 임상 전문 지식은 혈액 시스템에 영향을 미치는 질환 예컨대 SCID, 판코니 빈혈, 위스코트-알드리치 증후군 및 겸상 적혈구 빈혈을 생체외 편집 요법의 주목을 받게 하였다. 조혈 세포에 주목하는 다른 이유는 혈액 장애에 대한 유전자 요법을 설계하려는 이전 노력 덕분에, 비교적 고효율의 전달 시스템이 이미 존재한다는 것이다. 이러한 장점으로, 이러한 요법 방식은 편집된 세포가 피트니스 장점을 보유하는 질환에 적용될 수 있어서, 적은 수의 생착되고, 편집된 세포를 확장시킬 수 있고 질환을 치료할 수 있다. 이러한 질환 중 하나가 HIV이며, 이 감염은 CD4+ T 세포에 피트니스 단점을 야기시킨다.

[1405] 생체외 편집 요법은 현재 유전자 교정 전략을 포함시키기 위해 확대되었다. 생체외 HDR에 대한 장벽은 Genovese와 그 동료들의 최근 논문에서 극복되었는데, 그들은 SCID-X1를 앓는 환자로부터 수득된 조혈 줄기 세포 (HSC)에서 돌연변이된 IL2RG 유전자의 유전자 교정을 달성하였다[Genovese, P., et al. Nature 510, 235-240 (2014)]. Genovese 등은 다중모드 전략을 사용해 HSC에서 유전자 교정을 수행하였다. 먼저, HSC는 IL2RG에 대한 치료적 cDNA를 코딩하는 HDR 주형을 함유하는 통합-결핍 렌티바이러스를 사용해 형질도입되었다.

형질도입 이후에, 세포는 HDR 기반 유전자 교정을 자극하기 위해 IL2RG 내 돌연변이 하스폿을 표적으로 하는 ZFN을 코딩하는 mRNA로 전기천공시켰다. HDR 비율을 증가시키기 위해서, 배양 조건은 HSC 분열을 북돋기 위해 소형 분자로 최적화시켰다. 최적 배양 조건, 뉴클레아제 및 HDR 주형을 사용해 SCID-X1 환자로부터 유전자 교정된 HSC가 치료적으로 관련된 비율로 배양으로 수득되었다. 동일한 유전자 교정 절차를 겪은 비질환 개체로부터의 HSC는 HSC 기능에 대한 금본위로서, 마우스에서 장기간 조혈작용을 지속시킬 수 있었다. HSC는 모든 조혈 세포 유형을 발생시킬 수 있고 자가 이식될 수 있어서, 그들을 모든 조혈 유전 장애에 대한 매우 가치있는 세포 개체군으로 만들었다 [Weissman, I.L. & Shizuru, J.A. Blood 112, 3543-3553 (2008)]. 유전자 교정된 HSC는 원칙적으로 광범위한 유전성 혈액 장애를 치료하는데 사용될 수 있어서, 이 연구를 치료적 게놈 편집을 위한 흥미로운 돌파구로 만들었다.

[1406] 생체내 편집 요법: 생체내 편집은 유리하게 본 개시 및 당분야의 지식으로부터 사용될 수 있다. 전달이 효율적인 장기 시스템의 경우에, 수많은 흥미로운 전임상 치료 성공이 이미 존재하였다. 성공적인 생체내 편집 요법의 첫번째 예는 혈우병 B의 마우스 모델에서 입증되었다 [Li, H., et al. Nature 475, 217-221 (2011)]. 앞서 언급한 바와 같이, 혈우병 B는 응고 캐스캐이드의 핵심 성분인, 인자 IX를 코딩하는 유전자 내 기능 상실 돌연변이로 인해 초래되는 X-연관 열성 장애이다. 중증 발병 개체에서 이의 수준의 1% 이상까지 인자 IX 활성의 회복은 질환을 상당히 더 경증의 형태로 전환시킬 수 있는데, 이러한 수준을 달성하기 위해 어린 나이부터 예방적으로 이러한 환자에게 재조합 인자 Ix의 주입이 대체로 임상 합병증을 완화시키기 때문이다 [Lofqvist, T., et al. Journal of internal medicine 241, 395-400 (1997)]. 따라서, 오직 낮은 수준의 HDR 유전자 교정이 환자에 대한 임상 결과를 변 화시키는데 필수적이다. 또한, 인자 Ix는 편집 시스템을 코딩하는 바이러스 벡터에 의해 효율적으로 형질도입된 장기인 간에서 합성되고 분비된다.

[1407] ZNF 및 교정 HDR 주형을 코딩하는 간친화성 아데노-연관 바이러스 (AAV) 혈청형을 사용하여, 쥐과 간에서 돌연변이된, 인간화 인자 IX 유전자의 최대 7%의 유전자 교정이 달성되었다 [Li, H., et al. Nature 475, 217-221 (2011)]. 이것은 응고 캐스캐이드의 기능 척도로서, 응괴 형성 운동학의 개선을 야기하여서, 최초로 생체내 편집 요법이 실현가능하고 또한 효율적임을 입증하였다. 본 명세서에서 논의된 바와 같이, 당업자는 기능 상실 돌연변이를 반전시키기 위해서 X-연관 열성 장애의 돌연변이를 표적으로 하는 HDR 주형 및 CRISPR-Cas (C2c1) 시스템을 함유하는 입자를 사용해, 혈우병 B를 해결하기 위해, 당분야의 지식, 예를 들어, Li 및 본 명세서에 교시를 이용한다.

[1408] 이 연구를 기반으로, 다른 그룹들은 최근에 CRISPR-Cas에 의한 간의 생체네 게놈 편집을 사용하여 유전성 티로신혈증의 마우스 모델을 치료하였고 심혈관 질환에 대한 보호성을 제공하는 돌연변이를 생성시켰다. 이들 구별되는 2개 적용은 간 이상기능을 포함한 장애에 대한 이 접근법의 다재다능성을 입증한다 [Yin, H., et al. Nature biotechnology 32, 551-553 (2014); Ding, Q., et al. Circulation research 115, 488-492 (2014)]. 다른 장기 시스템에 생체내 편집의 적용은 이 전략이 광범위하게 적용가능하다는 것을 입증하는 것이 필요하다.

최근에, 이러한 방식의 요법으로 치료할 수 있는 장애의 범위를 확장시키기 위해 바이러스 및 비바이러스 벡터를 최적화하려는 노력이 진행 중이다 [Kotterman, M.A. & Schaffer, D.V. Nature reviews. Genetics 15, 445-451 (2014); Yin, H., et al. Nature reviews. Genetics 15, 541-555 (2014)]. 본 명세서에서 논의된 바와 같이, 당업자는 돌연변이를 표적으로 하는 HDR 주형 및 CRISPR-Cas (C2c1) 시스템을 함유하는 입자를 사용해, 유전성 티로신혈증을 해결하기 위해, 당분야의 지식, 예를 들어, Yin 및 본 명세서에 교시를 이용한다.

[1409] 표적화된 결실, 치료적 적용: 유전자의 표적화된 결실이 바람직할 수 있다. 그러므로, 면역결핍 장애, 혈액

병태, 또는 유전성 리소솜 저장 질환, 예를 들어 혈우병 B, SCID, SCID-X1, ADA-SCID, 유전성 티로신혈증, β-지중해빈혈, X-연관 CGD, 위스코트-알드리치 증후군, 판코니 빈혈, 부신백질이영양증 (ALD), 이염성 백질이영양증 (MLD), HIV/AIDS, 다른 대사 장애에 관여되는 유전자, 질환에 관여되는 미스폴딩된 단백질을 코딩하는 유전자, 질환에 관여된 기능 상실을 초래하는 유전자가; 일반적으로 유리하다고 간주되는 입자 시스템과, 임의의 본 명세서에 논의된 전달 시스템을 사용하여, HSC에서 표적화할 수 있는 돌연변이가 바람직하다.

[1410] 본 발명에서, 특히 CRISPR 효소의 면역원성은 에리트로포이에틴에 관해 문헌[Tangri et al]에서 처음 제시된 접근에 따라 감소될 수 있고, 후속적으로 발행될 수 있다. 따라서, 방향 진화(directed evolution) 또는 합리적 설계는 숙주 종(인간 또는 다른 종)에서 CRISPR 효소(예를 들어, C2c1)의 면역원성을 감소시키기 위해 사용될 수 있다.

[1411] 게놈 편집: 본 발명의 CRISPR/Cas (C2c1) 시스템은 본 명세서에서 논의된 거을 포함하여, TALEN 및 ZFN 및 렌티바이러스를 사용하여 제한된 성공을 지니는 이전에 시도한 유전자 돌연변이를 고정하기 위해 사용될 수 있다; 참조: WO2013163628. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다.

일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1412] **뇌, 중추 신경 및 면역 체계 질병의 치료**

[1413] 본 발명은 또한 CRISPR-Cas 시스템의 뇌 또는 뉴런으로의 전달을 고려한다. 일부 실시형태에서, CRISPR-Cas 시스템은 C2c1 단백질을 포함한다. 일부 구현예예서, CRISPR-C21 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 바람직한 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ, 바람직하게 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다.

일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

예를 들어, RNA 간섭 (RNAi)는 헌팅톤병의 질환-야기 유전자인, HTT (참조: 예를 들어 McBride et al., Molecular Therapy vol. 19 no. 12 Dec. 2011, pp. 2152-2162)의 발현을 감소시켜 이 장애에 대한 치료 잠재성을 제공하며, 그러므로 출원인은 CRISPR-Cas 시스템에 사용하고/하거나 조정할 수 있다고 상정한다. CRISPR-Cas 시스템은 알고리즘을 이용하여 생성되어 안티센스 서열의 표적외 부위를 표적화하는 능을 감소시킬 수 있다. CRISPR-Cas 서열은 마우스, 붉은털 원숭이(rhesus) 또는 인간 헌팅턴의 엑손 52에서의 서열 중 하나를 표적화하고, AAV와 같은 바이러스 벡터 내에서 발현될 수 있다. 인간을 포함하여, 동물은 반구 당 약 3 회의 미세주입을 이용하여 주사될 수 있다 (총 6회 주사):첫번째는 두측에서 (rostral) 전교련 (anterior commissure)에 대해 1 mm (12 ㎕)이고, 남은 두 주사 (각각, 12 ㎕ 및 10 ㎕)는 첫번째 주사에 대해 미측으로 3 및 6 mm 간격으로, 약 1 ㎕/분의 속도로 AAV 1e12 vg/ml를 이용하며, 바늘은 추가의 5분 동안 제자리에 두어 주입물이 바늘 끝으로부터 분산되는 것을 허용하였다.

[1414] DiFiglia 등 (PNAS, October 23, 2007, vol. 104, no. 43, 17204-17209) 은 Htt를 표적화하는 siRNA의 성인 선조체 내로의 단일 투여는 돌연변이체 Htt를 사일런스시키고, 뉴런성 병리학을 약화시키고, HD의 신속한-발병, 바이러스 유전자이식 마우스 모델에서 관찰된 비정상의 거동 표현형을 지연시킬 수 있음을 관찰하였다. DiFiglia는 10μM로, 2 ㎕의 Cy3-표지된 cc-siRNA-Htt 또는 컨쥬게이트되지 않은 siRNA-Htt를 마우스의 선조체 내로 주사하였다. Htt로 표적된 CRISPR Cas의 유사한 용량이 본 발명에서 인간의 경우 고려될 수 있으며, 예를 들어 Htt로 표적된 10 μM CRISPR Cas의 약 5-10 ml가 선조체 내로 주사될 수 있다.

[1415] 다른 예에서, Boudreau 등 (Molecular Therapy vol. 17 no. 6 June 2009])은 htt-특이적 RNAi 바이러스를 발현하는 재조합 AAV 혈청형 2/1 벡터 5 ㎕ (4×1012의 바이러스 게놈/ml)를 선조체 내로 주사한다. Htt로 표적된 CRISPR Cas의 유사한 용량이 본 발명에서 인간의 경우에 고려될 수 있으며, 예를 들어 약 10-20 ml의 (4×

1012의 바이러스 게놈/ml) Htt로 표적된 CRISPR Cas가 선조체 내로 주사될 수 있다.

[1416] 또 다른 예에서, HTT로 표적된 CRISPR Cas는 연속 투여될 수 있다 (예를 들어, 문헌 [Yu et al., Cell 150, 895908, August 31, 2012] 참조). Yu 등은 삼투압 펌프를 이용하여 ss-siRNA 또는 인산염 완충 식염수 (PBS) (Sigma Aldrich)를 28일 동안 300 mg/일 전달하기 위하여 0.25 ml/시간 (Model 2004)를 전달하고, 0.5 ㎕/시간 (Model 2002)을 전달하도록 설계된 펌프를 사용하여 14 일 동안 75 mg/일의 양성 대조구 MOE ASO를 전달하는데 사용하였다. 펌프 (Durect Corporation)는 멸균 PBS 내에서 희석된 ss-siRNA 또는 MOE로 충전되었으며, 이후 임플랜테이션(implantation) 전 37℃에서 24 시간 또는 48 시간 (Model 2004) 인큐베이션하였다. 마우스들을 2.5% 이소플루오란으로 마취시키고, 두개골의 기부에 중간선 절단을 하였다. 정위법 가이드를 이용하여, 삽입관을 우측 뇌실 내로 임플랜트하고, Loctite 접착제로 고정시켰다. Alzet 삼투압 미니 펌프에 부착된 카테테르를 삽입관에 부착시키고, 펌프를 중간견갑골 영역에 피하에 위치시켰다. 절단은 5.0 나일론 봉합사를 이용하여 밀폐되었다. 유사한 용량의 Htt로 표적된 CRISPR Cas가 본 발명에서 인간의 경우에 대해 고려될 수 있으며, 예를 들어, 약 500 내지 1000 g/일의 Htt로 표적된 CRISPR Cas가 투여될 수 있다.

[1417] 연속 주입의 또 다른 예에서, Stiles 등 (문헌 [Experimental Neurology 233 (2012) 463471])은 티타늄 바늘 끝을 이용하여 뇌실질내 카테테르를 우측 피곡(putamen) 내로 임플랜트하였다. 카테터는 복부에 피하 이식된 SynchroMed® II 펌프(Medtronic Neurological, Minneapolis, MN)에 연결되었다. 6 μL/일의 인산염 완충 식염수 주입 7일 후, 펌프를 시험 물품으로 다시 충전하고 7일 동안 연속 전달을 위해 프로그램하였다. 약 2.3 내지 11.52 mg/d의 siRNA를 약 0.1에서 0.5 μL/분의 가변 주입 속도로 주입시켰다. 유사한 용량의 Htt로 표적된 CRISPR Cas가 본 발명에서 인간의 경우에 대해 고려될 수 있으며, 예를 들어, 약 20 내지 200 mg/일의, Htt로 표적된 CRISPR Cas가 투여될 수 있다. 또 다른 예에서, Sangamo에게 수여된 미국 특허 공개 20130253040은 또한 헌팅턴병을 치료하기 위해 TALES로부터 본 발명의 핵산-표적화 시스템에 대하여 적합화될 수 있다.

[1418] 다른 예에서, Sangamo에 양도된 미국 특허 공개 제20130253040의 방법이 또한 헌팅턴병의 치료를 위해서 TALES로부터 CRISPR Cas 시스템에 맞게 개조될 수 있다.

[1419] 참조로 본 명세서에 개시된, The Broad Institute 등의 출원인의 WO2015089354 A1는 헌팅톤병 (Hp)에 대한 표적을 기술한다. 헌팅톤병에 관한 CRISPR 복합체의 가능한 표적 유전자는 PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; 및 TGM2이다. 따라서, PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2 중 하나 이상은 본 발명의 일부 구현예에서 헌팅톤병에 대한 표적으로서 선택될 수 있다.

[1420] 다른 트리뉴클레오티드 반복 질환. 이들은 임의의 하기를 포함할 수 있다: 범주 I은 헌팅톤병 (HD) 및 척수소뇌 운동실조증을 포함하고; 확장 범주 I는 일잔적으로 작은 규모로 이종성 확장이 있는 표현형적으로 다양하며, 또한 유전자의 엑손에서도 존재하고; 범주 III은 취약 X 증후군, 근긴장성 이영양증, 척수소뇌 운동실조증의 2, 소아 근대 간질, 및 프리드리히 실조증을 포함한다.

[1421] 본 발명의 추가의 양태는 라포라 질병(Lafora disease)과 관련이 있는 것으로 확인된 EMP2A 및 EMP2B 유전자의 결함을 교정하기 위하여 CRISPR-Cas 시스템을 사용하는 것에 관한 것이다. 라포라병은 청소년기에 간질성 발작으로서 시작할 수 있는 진행성 마이오클로누스 간질을 특징으로 하는 상염색체 열성 질환이다. 질병의 몇몇 경우에서 아직 확인되지 않은 유전자의 돌연변이에 의해 초래될 수 있다. 질병은 발작, 근경련, 걷기 어려움, 치매, 및 결국 사망을 초래한다. 질병의 진행에 대해 효과적으로 주어지는 치료법은 현재 없다. 간질과 연관된 다른 유전 이상은 또한 CRISPR-Cas 시스템에 의해 표적화될 수 있으며 근본적인 유전학은 문헌 [Genetics of Epilepsy and Genetic Epilepsies, edited by Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009]에 더욱 기술되어 있다.

[1422] T 세포 수용체(TCR) 유전자를 불활성화시키는 데 관련된, Sangamo BioSciences, Inc.로 양도된 미국 특허 공개 제20110158957호의 방법은 또한 본 발명의 CRISPR Cas 시스템으로 변형될 수 있다. 다른 예에서, Sangamo BioSciences, Inc.로 양도된 미국 특허 공개 제20100311124호 및 Cellectis로 양도된 미국 특허 공개 제20110225664호의 방법은 둘 다 글루타민 합성효소 유전자 발현 유전자를 불활성화시키는 데 관련된 것으로서, 또한 본 발명의 CRISPR Cas 시스템으로 변형될 수 있다.

[1423] 뇌에 대한 전달 선택사항은 리포솜 내로 DNA 또는 RNA 중 하나 형태로 CRISPR 효소 및 가이드 RNA의 캡슐화 및 트랜스-혈액뇌장벽(BBB) 전달을 위한 분자 트로이목마에 대한 컨쥬게이팅을 포함한다. 분자 트로이목마는 비인간 영장류의 뇌 내로 B-gal 발현 벡터의 전달에 효과적인 것으로 나타났다. 동일한 접근이 CRISPR 효소

및 가이드 RNA를 함유하는 벡터를 전달하는데 사용될 수 있다. 예를 들어, Xia CF 및 Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human 인슐린 수용체 using avidin-biotin technology." Mol Pharm. doi:10.1021/mp800194)은 배양물에서 및 생체내에서 세포에 대한 짧은 간섭 RNA(siRNA)의 전달 방법이 수용체-특이적 단클론성 항체(mAb) 및 아비딘-바이오틴 기술의 조합된 사용에 의해 가능하다는 것을 기재한다. 저자는 또한 표적화 mAb와 siRNA 간의 결합이 아비딘-바이오틴 기술에 의해 안정하기 때문에, 뇌와 같은 먼 부위에서 RNAi 효과가 표적화된 siRNA의 정맥내 투여 후 생체내에서 관찰되었다는 것을 보고한다.

[1424] Zhang 등 (Mol Ther.2003 Jan;7(1):118.))은 루시퍼라제와 같은 리포터를 코딩하는 발현 플라스미드가 인간 인슐린 수용체(HIR)에 대해 단클론성 항체(MAb)를 지니는 생체내 레서스 원숭이 뇌에 대해 표적화된, 85 nm 페길화 면역리포솜을 포함하는 " 인공 바이러스" 의 내부에서 캡슐화되는 방법을 기재한다. HIRMAb는 리포솜이 혈액-뇌 장벽을 가로지르는 통과세포배출 및 정맥내 주사 후 신경 혈장막을 가로지르는 내포작용을 겪는 외생성 유전자를 운반하는 것을 가능하게 한다. 뇌에서 루시퍼라제 유전자 발현의 수준은 래트와 비교하여 레서스 원숭이에서 50 배 더 높았다. 영장류 뇌에서 베타-갈락토시다제 유전자의 널리 퍼진 신경 발현은 조직화학과 공초점 현미경 둘 다에 의해 입증되었다. 저자는 이 접근이 24 시간에 가역적 성체 유전자이식을 실현가능하게 만든다는 것을 나타낸다. 따라서, 면역리포솜의 사용이 바람직하다. 이들은 특이적 표적 또는 세포 표면 단백질을 표적화하는 항체와 함께 사용될 수 있다.

[1425] **알츠하이머병**

[1426] 미국 특허 공개 제20110023153호는 알츠하이머병과 연관된 세포, 동물 및 단백질을 유전적으로 변형시키기 위해 아연 핑거 뉴클레아제를 사용하는 것을 기재한다. 일단 변형된 세포 및 동물은 AD의 연구에 통상 사용되는 척도를 사용하여 AD의 발생 및/또는 진행에 대한 표적화된 돌연변이의 영향을 연구하기 위한 공지된 방법을 사용하여 더 시험될 수 있으며, 이들은 제한은 아니지만 예컨대 학습 및 기억, 불안, 우울감, 중독, 및 감각 운동 기능뿐만 아니라 행동, 기능, 병리, 대사 및 생화학적 기능을 측정하는 분석 등이다.

[1427] 본 개시는 AD 와 관련된 단백질을 코딩하는 어떤 염색체 서열의 편집을 포함한다.

[1428] 일부 구현예에서, 본 발명에 개시된 시스템은 C2c1 - CRISPR 시스템을 포함할 수 있다. 일부 구현예에서, CRISPR-C21 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 AD 관련 유전자에 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 AD 관련 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다. AD-관련 단백질은 전형적으로 AD-관련 단백질과 AD 장애의 실험적 연관성에 기초하여 선택된다. 예를 들어, AD-관련 단백질의 혈중 농도나 생성 비율은 AD 장애를 결여한 집단에 비해 AD 장애를 가진 집단에서 상승되거나 억제될 수 있다. 단백질 수준에서의 차이는 웨스턴 블롯, 면역 조직화학적 염색, 효소 결합 면역 흡착 분석 (ELISA), 및 질량 분석을 포함하는 프로테오믹스 기술을 사용하여 평가될 수 있으나, 이에 제한되지 않는다. 또는 달리, AD-관련 단백질은, 제한은 아니지만, DNA 마이크로어레이 분석, 연속 유전자 발현 분석 (SAGE), 및 정량적 실시간 폴리머라제 연쇄 반응 (Q-PCR)을 포함하는 게놈 기술을 사용하여 단백질을 코딩하는 유전자의 유전자 발현 프로파일을 얻음으로써 확인될 수 있다.

[1429] 알츠하이머병 관련 단백질의 예는 예를 들어, VLDLR 유전자에 의해 코딩되는 극저밀도 리포단백질 수용체 단백질(VLDLR), UBA1 유전자에 의해 코딩되는 유비퀴틴-유사 변형 활성화 효소(UBA1) 또는 UBA3 유전자에 의해 코딩되는 NEDD8-활성화 효소 E1 촉매 서브유닛 단백질(UBE1C)을 포함할 수 있다.

[1430] 비제한적 예로서, AD와 관련된 단백질은 제한은 아니지만 다음에 열거된 단백질들을 포함한다:염색체 서열 암호화된 단백질 ALAS2 델타-아미노레불리네이트 신타제 2(ALAS2) ABCA1 ATP-결합 카세트 수송인자(ABCA1) ACE 안지오텐신 I-전환 효소 (ACE) APOE 아포지방단백질 E 전구체 (APOE) APP 아밀로이드 전구체 단백질(APP) AQP1 아쿠아포린 1 단백질(AQP1) BIN1 Myc 박스-의존성-상호작용 단백질 1 또는 브릿지 통합인자 1 단백질(BIN1) BDNF 뇌-유래 신경영양 인자(BDNF) BTNL8 부티로필린-유사 단백질 8 (BTNL8) C1ORF49 염색체 1 오픈 리딩 프레임 49 CDH4 캐드헤린-4 CHRNB2 뉴런 아세틸콜린 수용체 서브유닛 베타-2 CKLFSF2 CKLF-유사 MARVEL 경막 도메인-함유 단백질 2(CKLFSF2) CLEC4E C-타입 렉틴 도메인 패밀리 4, 멤버 (CLEC4E) CLU 클러스테린 단백질(아포지방단백질

J로도 알려져 있다) CR1 적혈구 보체 수용체 1(CR1, CD35, C3b/C4b 수용체 및 면역 부착 수용체라고도 한다) CR1L 적혈구 보체 수용체 1(CR1L) CSF3R 과립구 콜로니-자극 인자 3 수용체 (CSF3R) CST3 시스타틴 C 또는 시스타틴 3 CYP2C 시토크롬 P450 2C DAPK1 사망-관련 단백질 키나제 1 (DAPK1) ESR1 에스트로겐 수용체 1 IgA의 FCAR Fc 단편 수용체 (FCAR, CD89라고도 한다) IgG의 FCGR3B Fc 단편, 저 친화성 IIIb, 수용체 (FCGR3B 또는 CD16b) FFA2 자유 지방산 수용체 2(FFA2) FGA 피브리노겐 (인자 I) GAB2 GRB2-관련-결합 단백질 2(GAB2) GAB2 GRB2-관련-결합 단백질 2(GAB2) GALP 갈라닌-유사 펩티드 GAPDHS 글리세르알데하이드-3-포스페이트 디하이드로게나제, 정자형성 (GAPDHS) GMPB GMBP HP 햅토글로빈 (HP) HTR7 5-하이드록시 트립타민 (세로토닌) 수용체 7 (아데닐레이트 시클라제-결합) IDE 인슐린 분해 효소 IF127 IF127 IFI6 인터페론, 알파-유도성 단백질 6 (IFI6) IFIT2 인터페론-유도 단백질 테트라트리코펩티드 반복부 2(IFIT2) IL1RN 인터루킨-1 수용체 길항제(IL-1RA) IL8RA 인터루킨 8 수용체, 알파 (IL8RA 또는 CD181) IL8RB 인터루킨 8 수용체, 베타 (IL8RB) JAG1 재그드 1 (JAG1) KCNJ15 칼륨 내부-정류 채널, 서브패밀리 J, 멤버 15 (KCNJ15) LRP6 저밀도 지방단백질 수용체-관련 단백질 6 (LRP6) MAPT 미세소관-관련 단백질 tau (MAPT) MARK4 MAP/미세소관 친화성-조절 키나제 4 (MARK4) MPHOSPH1 M-상 포스포단백질 1 MTHFR 5,10-메틸렌테트라하이드로폴레이트 환원효소 MX2 인터페론-유도 GTP-결합 단백질 Mx2 NBN 니브린, NBN NCSTN 니카스트린이라고도 함 NIACR2 니아신 수용체 2(NIACR2, GPR109B라고도 함) NMNAT3 니코틴아미드 뉴클레오티드 아데닐일트랜스페라제 3 NTM 뉴로트리민 (또는 HNT) ORM1 오로스뮤코이드 1 (ORM1) 또는 알파-1-산 글리코단백질 1 P2RY13 P2Y 퓨리노셉터 13 (P2RY13) PBEF1 니코틴아미드 포스포리보실트랜스페라제(NAmPRTase 또는 Nampt) 프레-B-세포 콜로니-증진 인자 1 (PBEF1) 또는 비스파틴 PCK1 포스포엔올피루베이트 카복시 키나제라고도 함 PICALM 포스파티딜이노시톨 결합 클라스린 조립체 단백질(PICALM) PLAU Uro 키나제-타입 플라스미노겐 활성인자(PLAU) PLXNC1 플렉신 C1 (PLXNC1) PRNP 프리온 단백질 PSEN1 프레세닐린 1 단백질(PSEN1) PSEN2 프레세닐린 2 단백질(PSEN2) PTPRA 단백질 티로신 포스파타제 수용체 타입 A 단백질 (PTPRA) RALGPS2 Ral GEF PH 도메인 및 SH3 결합 모티프 2(RALGPS2) RGSL2 조절인자 G-단백질 신호화 유사 2(RGSL2) SELENBP1 셀레늄 결합 단백질 1 (SELNBP1) SLC25A37 미토페린-1 SORL1 소틸린-관련 수용체 L(DLR 부류) A 반복부-함유 단백질(SORL1) TF 트랜스페린 TFAM 미토콘드리아 전사 인자 A TNF 종양 괴사 인자 TNFRSF10C 종양 괴사 인자 수용체 수퍼패밀리 멤버 10C (TNFRSF10C) TNFSF10 종양 괴사 인자 수용체 수퍼패밀리, (TRAIL) 멤버 10a (TNFSF10) UBA1 유비퀴틴-유사 변형인자 활성화 효소 1 (UBA1) UBA3 NEDD8-활성화 효소 E1 촉매 서브유닛 단백질(UBE1C) UBB 유비퀴틴 B 단백질(UBB) UBQLN1 유비퀼린-1 UCHL1 유비퀴틴 카복실-말단 에스테라제 L1 단백질(UCHL1) UCHL3 유비퀴틴 카복실-말단 하이드롤라제 이소엔자임 L3 단백질(UCHL3) VLDLR 초저밀도 지방단백질 수용체 단백질(VLDLR).

[1431] 예시적인 구현예에서, 염색체 서열이 편집된 AD와 관련된 단백질은 VLDLR 유전자에 의해서 암호화된 초저밀도 지방단백질 수용체 단백질(VLDLR), UBA1 유전자에 의해서 암호화된 유비퀴틴-유사 변형인자 활성화 효소 1(UBA1), UBA3 유전자에 의해서 암호화된 NEDD8-활성화 효소 E1 촉매 서브유닛 단백질(UBE1C), AQP1 유전자에 의해서 암호화된 아쿠아포린 1 단백질(AQP1), UCHL1 유전자에 의해서 암호화된 유비퀴틴 카복실-말단 에스테라제 L1 단백질(UCHL1), UCHL 3 유전자에 의해서 암호화된 유비퀴틴 카복실-말단 하이드롤라제 이소엔자임 L3 단백질(UCHL3), UBB 유전자에 의해서 암호화된 유비퀴틴 B 단백질(UBB), MAPT 유전자에 의해서 암호화된 미세소관-관련 단백질 tau (MAPT), PTPRA 유전자에 의해서 암호화된 단백질 티로신 포스파타제 수용체 타입 A 단백질 (PTPRA), PICALM 유전자에 의해서 암호화된 포스파티딜이노시톨 결합 클라스린 조립체 단백질(PICALM), CLU 유전자에 의해서 암호화된 클러스테린 단백질(아포지방단백질 J로도 알려져 있다), PSEN1 유전자에 의해서 암호화된 프레세닐린 1 단백질, PSEN2 유전자에 의해서 암호화된 프레세닐린 2 단백질, SORL1 유전자에 의해서 암호화된 소틸린-관련 수용체 L (DLR 부류) A 반복부-함유 단백질(SORL1) 단백질, APP 유전자에 의해서 암호화된 아밀로이드 전구체 단백질(APP), APOE 유전자에 의해서 암호화된 아포지방단백질 E 전구체 (APOE), 또는 BDNF 유전자에 의해서 암호화된 뇌-유래 신경영양 인자(BDNF)일 수 있다. 예시적인 구현예에서, 유전자 변형된 동물은 래트이며, AD와 관련된 단백질을 암호화하는 편집된 염색체 서열은 다음과 같다: APP 아밀로이드 전구체 단백질(APP) NM_019288 AQP1 아쿠아포린 1 단백질(AQP1) NM_012778 BDNF 뇌-유래 신경영양 인자 NM_012513 CLU 클러스터린 단백질(NM_053021 아포지방단백질 J로도 알려져 있다) MAPT 미세소관-관련 단백질 NM_017212 tau (MAPT) PICALM 포스파티딜이노시톨 결합 NM_053554 클라스린 조립체 단백질(PICALM) PSEN1 프레세닐린 1 단백질(PSEN1) NM_019163 PSEN2 프레세닐린 2 단백질(PSEN2) NM_031087 PTPRA 단백질 티로신 포스파타제 NM_012763 수용체 타입 A 단백질(PTPRA) SORL1 소틸린-관련 수용체 L(DLR NM_053519, 부류) A 반복부-함유 XM_001065506, 단백질(SORL1) XM_217115 UBA1 유비퀴틴-유사 변형인자 활성화 NM_001014080 효소 1 (UBA1) UBA3 NEDD8-활성화 효소 E1 NM_057205 촉매 서브유닛단백질(UBE1C) UBB 유비퀴틴 B 단백질(UBB) NM_138895 UCHL1 유비퀴틴 카복실-말단 NM_017237 에스테라제 L1 단백질(UCHL1) UCHL3 유비퀴틴 카복실-말단 NM_001110165 하이드롤라제 이소엔

자임 L3 단백질(UCHL3) VLDLR 초저밀도 지방단백질 NM_013155 수용체 단백질(VLDLR).

[1432] 동물 또는 세포는 AD와 관련된 단백질을 코딩하는 1, 2, 3, 4, 5, 6, 7, 8, 9,10, 11, 12, 13, 14, 15 또는 그 이상의 파괴된 염색체 서열 및 AD와 관련된 파괴된 단백질을 코딩하는 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 또는 그 이상의 염색체 통합 서열을 포함할 수 있다.

[1433] 편집된 또는 통합된 염색체 서열은 AD와 관련된 변경된 단백질을 암호화하도록 변형될 수 있다. AD-관련 염색체 서열에서 다수의 돌연변이가 AD와 관련되었다. 예를 들어, APP에서 V7171(즉, 위치 717에서 발린이 이소류신으로 변화됨) 미스센스 돌연변이는 가족성 AD를 야기한다. 프레세닐린-1 단백질에서의 다중 돌연변이, 예컨대 H163R (즉, 위치 163에서 히스티딘이 아르기닌으로 변화됨), A246E (즉, 위치 246에서 알라닌이 글루타메이트로 변화됨), L286V(즉, 위치 286에서 류신이 발린으로 변화됨) 및 C410Y (즉, 위치 410에서 시스테인이 티로신으로 변화됨)는 가족성 알츠하이머병 타입 3을 야기한다. 프레세닐린-2 단백질에서의 돌연변이, 예컨대 N141 I(즉, 위치 141에서 아스파라긴이 이소류신으로 변화됨), M239V (즉, 위치 239에서 메티오닌이 발린으로 변화됨), 및 D439A (즉, 위치 439에서 아스파르테이트가 알라닌으로 변화됨)는 가족성 알츠하이머 타입 4를 야기한다. AD-관련 유전자 및 질환에서 유전자 변이체의 다른 관련성도 본 분야에 잘 알려져 있다. 예를 들어, 문헌[Waring et al. (2008) Arch. Neurol. 65:329334]를 참조하며, 이것의 개시는 그 전체가 여기 참고로 포함된다.

[1434] **세크레타제 장애**

[1435] 미국 특허 공개 제20110023146호는 세크레타제-연관 장애와 연관된 세포, 동물 및 단백질을 유전적으로 변형시키기 위해 아연 핑거 뉴클레아제를 사용하는 것을 기재한다. 세크레타제는 프리-단백질을 그들 생물학적 활성 형태로 프로세싱하는데 필수적이다. 당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 미국 특허 공개 제20110023146호의 것과 유사한 시스템에서 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1436] 세크레타제 경로의 다양한 구성성분에서 결함은 많은 장애들, 특히 특징적 아밀로이드형성 또는 아밀로이드반을 가진 것들, 예컨대 알츠하이머병 (Ad)의 원인이 된다.

[1437] 세크레타제 장애 및 이들 장애와 관련된 단백질은 수많은 장애에 대한 감수성, 장애의 존재, 장애의 중증도, 또는 이들의 어떤 조합에 영향을 미치는 다양한 세트의 단백질이다. 본 개시는 세크레타제 장애와 관련된 단백질을 코딩하는 어떤 염색체 서열의 편집을 포함한다. 세크레타제 장애와 관련된 단백질은 전형적으로 세크레타제-관련 단백질과 세크레타제 장애의 발생의 실험적 연관성에 기초하여 선택된다. 예를 들어, 세크레타제 장애와 연관된 단백질의 생산 속도 또는 순환 농도는 세크레타제 장애가 없는 개체군에 비해서 세크레타제 장애를 갖는 개체군에서 상승 또는 억제될 수 있다. 단백질 수준에서의 차이는 웨스턴 블롯, 면역 조직화학적 염색, 효소 결합 면역 흡착 분석 (ELISA), 및 질량 분석을 포함하는 프로테오믹스 기술을 사용하여 평가될 수 있으나, 이에 제한되지 않는다. 대안적으로, 세크레타제 장애와 연관된 단백질은 제한없이 DNA 마이크로어레이 분석, 유전자 발현의 연속 분석 (SAGE), 및 정량적 실시간 중합효소 연쇄 반응 (Q-PCR)을 포함한 게놈 기술을 사용해 단백질을 코딩하는 유전자의 유전자 발현 프로파일을 수득하여 확인할 수 있다.

[1438] 비제한적인 예로서, 세크레타제 장애와 연관된 단백질은 다음을 포함한다: PSENEN (프레세닐린 인핸서 2 상동체 (예쁜 꼬마 선충 (C. elegans)), CTSB (카텝신 B), PSEN1 (프레세닐린 1), APP (아밀로이드 베타 (A4) 전구체 단백질), APH1B (앞인두 결핍 1 상동체 B (예쁜 꼬마 서충), PSEN2 (프레세닐린 2 (알츠하이머병 4)), BACE1 (베타_부위 APP-절단 효소 1), ITM2B (통합 막 단백질 2B), CTSD (카텝신 D), NOTCH1 (Notch 상동체 1, 전좌-연관 (초파리 (Drosophila)), TNF (종양 괴사 인자 (TNF 수퍼패밀리, 멤버 2)), INS (인슐린), DYT10 (근긴장 이상 10), ADAM17 (ADAM 메탈로펩티다제 도메인 17), APOE (아포리포단백질 E), ACE (안지오텐신 I 전환 효소 (펩티딜-디펩티다제 A) 1), STN (스타틴), TP53 (종양 단백질 p53), IL6 (인터루킨 6 (인터페론, 베타 2)), NGFR (신경 성장 인자 수용체 (TNFR 수퍼패밀리, 멤버 16)), IL1B (인터루킨 1, 베타), ACHE (아세틸콜린에스터

라제 (Yt 혈액군)), CTNNB1 (카테닌 (카데린-연관 단백질), 베타 1, 88kDa), IGF1 (인슐린-유사 성장 인자 1 (소마토메딘 C)), IFNG (인터페론, 감마), NRG1 (뉴레귤린 1), CASP3 (캐스파제 3, 아폽토시스-관련 시스테인 펩티다제), MAPK1 (미토겐-활성화된 단백질 키나제 1), CDH1 (카데린 1, 1형, E-카데린 (상피)), APBB1 (아밀로이드 베타 (A4) 전구체 단백질-결합, 패밀리 B, 멤버 1 (Fe65)), HMGCR (3-히드록시-3-메틸글루타릴-보조효소 A 리덕타제), CREB1 (cAMP 반응성 엘리먼트 결합 단백질 1), PTGS2 (프로스타글란딘-엔도퍼옥시다제 신타제 2 (프로스타글란딘 G/H 신타제 및 시클로옥시게나제)), HES1 (스플릿 1의 모발 및 인헨서, (초파리)), CAT (카탈라제), TGFB1 (형질전환 성장 인자, 베타 1), ENO2 (에놀라제 2 (감마, 신경원)), ERBB4 (v-erb-a 적아세포 백혈병 바이러스 종양유전자 상동체 4 (조류)), TRAPPC10 (수송 단백질 입자 복합체 10), MAOB (모노아민 옥시다제 B), NGF (신경 성장 인자 (베타 폴리펩티드)), MMP12 (매트릭스 메탈로펩티다제 12 (마크로파지 엘라스타제)), JAG1 (jagged 1 (알라질 증후군)), CD40LG (CD40 리간드), PPARG (퍼옥시솜 증식인자-활성화된 수용체 감마), FGF2 (섬유아세포 성장 인자 2 (basic)), IL3 (인터루킨 3 (콜로니-자극 인자, 다수)), LRP1 (저밀도 지단백질 수용체-관련 단백질 1), NOTCH4 (Notch 상동체 4 (초파리)), MAPK8 (미토겐-활성화된 단백질 키나제 8), PREP (프롤릴 엔도펩티다제), NOTCH3 (Notch 상동체 3 (초파리)), PRNP (프라이온 단백질), CTSG (카텝신 G), EGF (상피 성장 인자 (베타-우로가스트론)), REN (레닌), CD44 (CD44 분자 (인도 혈액군)), SELP (셀렉틴 P (과립 막 단백질 140 kDa, 항원 CD62)), GHR (성장 호르몬 수용체), ADCYAP1 (아데닐레이트 시클라제 활성화 폴리펩티드 1 (뇌하수체)), INSR (인슐린 수용체), GFAP (아교세포 섬유 산성 단백질), MMP3 (매트릭스 메탈로펩티다제 3 (스트로멜리신 1, 프로젤라티나제)), MAPK10 (미토겐-활성화된 단백질 키나제 10), SP1 (Sp1 전사 인자), MYC (v-myc 골수구종증 바이러스 종양 유전자 상동체 (조류)), CTSE (카텝신 E), PPARA (퍼옥시솜 증식 인자-활성화된 수용체 알파), JUN (jun 종양유전자), TIMP1 (TIMP 메탈로펩티다제 억제제 1), IL5 (인터루킨 5 (콜로니-자극 인자, 호산구)), IL1A (인터루킨 1, 알파), MMP9 (매트릭스 메탈로펩티다제 9 (젤라티나제 B, 92 kDa 젤라티나제, 92 kDa IV형 콜라게나제)), HTR4 (5-히드록시트립타민 (세로토닌) 수용체 4), HSPG2 (헤파란 술페이트 프로테오글리칸 2), KRAS (v-Ki-ras2 Kirsten 래트 육종 바이러스 종양유전자 상동체), CYCS (시토크롬 c, 체세포), SMG1 (SMG1 상동체, 포스파티딜이노시톨 3-키나제-관련 키나제 (예쁜 꼬마 선충)), IL1R1 (인터루킨 1 수용체, I형), PROK1 (프로키네티신 1), MAPK3 (미토겐-활성화된 단백질 키나제 3), NTRK1 (신경영양성 티로신 키나제, 수용체, 1형), IL13 (인터루킨 13), MME (막 메탈로-엔도펩티다제), TKT (트랜스케톨라제), CXCR2 (케모카인 (C-X-C 모티프) 수용체 2), IGF1R (인슐린-유사 성장 인자 1 수용체), RARA (레티노산 수용체, 알파), CREBBP (CREB 결합 단백질), PTGS1 (프로스타글란딘-엔도퍼옥시드 신타제 1 (프로스타글란딘 G/H 신타제 및 시클로옥시게나제)), GALT (갈락토스-1-포스페이트 우리딜릴트랜스퍼라제), CHRM1 (콜린성 수용체, 무스카린성 1), ATXN1 (아탁신 1), PAWR (PRKC, 아폽토시스, WT1, 조절인자), NOTCH2 (Notch 상동체 2 (초파리)), M6PR (만노스-6-포스페이트 수용체 (양이온 의존적)), CYP46A1 (시토크롬 P450, 패밀리 46, 서브패밀리 A, 폴리펩티드 1), CSNK1 D (카세인 키나제 1, 델타), MAPK14 (미토겐-활성화된 단백질 키나제 14), PRG2 (프로테오글리칸 2, 골수 (자연 살해 세포 활성인자, 호산구 과립 주요 염기 단백질)), PRKCA (단백질 키나제 C, 알파), L1 CAM (L1 세포 부착 분자), CD40 (CD40 분자, TNF 수용체 수퍼패밀리 멤버 5), NR1I2 (핵 수용체 서브패밀리 1, 그룹 I, 멤버 2), JAG2 (jagged 2), CTNND1 (카테닌 (카데린_연관 단백질), 델타 1), CDH2 (카데린 2, 1형, N-카데린 (신경원)), CMA1 (키마제 1, 비만 세포1), SORT1 (솔틸린 1), DLK1 (델타-유사 1 상동체 (초파리)), THEM4 (티오에스터라제 수퍼패밀리 멤버 4), JUP (접합 플라코글로빈), CD46 (CD46 분자, 보체 조절 단백질), CCL11 (케모카인 (C-C 모티프) 리간드 11), CAV3 (카베올린 3), RNASE3 (리보뉴클레아제, RNase A 패밀리, 3 (호산구 양이온 단백질)), HSPA8 (열충격 70kDa 단백질 8), CASP9 (캐스파제 9, 아폽토시스-관련 시스테인 펩티다제), CYP3A4 (시토크롬 P450, 패밀리 3, 서브패밀리 A, 폴리펩티드 4), CCR3 (케모카인 (C-C 모티프) 수용체 3), TFAP2A (전사 인자 AP-2 알파 (활성화 인핸서 결합 단백질 2 알파)), SCP2 (스테롤 캐리어 단백질 2), CDK4 (시클린-독립 키나제 4), HIF1A (저산소증 유도성 인자 1, 알파 서브유닛 (염기성 헬릭스-루프-헬릭스 전사 인자)), TCF7L2 (전사 인자 7-유사 2 (T-세포 특이적, HMG-박스)), IL1R2 (인터루킨 1 수용체, II형), B3GALTL (베타 1,3-갈락토실트랜스퍼라제-유사), MDM2 (Mdm2 p53 결합 단백질 상동체 (마우스)), RELA (v-rel 세망내피증 바이러스 종양 유전자 상동체 A (조류)), CASP7 (캐스파제 7, 아폽토시스-관련 시스테인 펩티다제), IDE (인슐린-분해 효소), FABP4 (지방산 결합 단백질 4, 지방세포), CASK (칼슘/칼모듈린-의존적 세린 단백질 키나제 (MAGUK 패밀리)), ADCYAP1R1 (아데틸레이트 시클라제 활성화 폴리펩티드 1 (뇌하수체) 수용체 I 형), ATF4 (활성화 전사 인자 4 (tax-반응성 인핸서 엘리먼트 B67)), PDGFA (혈소판-유래 성장 인자 알파 폴리펩티드), C21 또는 f33 (염색체 21 오픈 리딩 프레임 33), SCG5 (세크레토그라닌 V (7B2 단백질)), RNF123 (링 핑거 단백질 123), NFKB1 (B-세포의 카파 경쇄 폴리펩티드 유전자 인핸서의 핵인자 1), ERBB2 (v-erb-b2 적아세포 백혈병 바이러스 종양유전자 상동체 2, 신경/교아세포종 유래 종양유전자 상동체 (조류)), CAV1 (카베올린 1, 카베올라 단백질, 22 kDa), MMP7 (매트릭스 메탈로펩티다제 7 (마트릴신, 자궁)), TGFA (형질전환 성장

인자, 알파), RXRA (레티노이드 X 수용체, 알파), STX1A (신탁신 1A (뇌)), PSMC4 (프로테아솜 (프로솜, 마크로파인) 26S 서브유닛, ATPase, 4), P2RY2 (푸린성 수용체 P2Y, G-단백질 커플링, 2), TNFRSF21 (종양 괴사 인자 수용체 수퍼패밀리, 멤버 21), DLG1 (discs, 거대 상동체 1 (초파리)), NUMBL (numb 상동체 (초파리)-유사), SPN (시알로포린), PLSCR1 (인지질 스크램블라제 1), UBQLN2 (유비퀼린 2), UBQLN1 (유비퀼린 1), PCSK7 (프로단백질 컨버타제 서브틸리신/켁신 7형), SPON1 (스폰딘 1, 세포외 매트릭스 단백질), SILV (은 상동체 (마우스)), QPCT (글루타미닐-펩티드 시클로트랜스퍼라제), HESS (스플릿 5의 모발 및 인핸서 (초파리)), GCC1 (GRIP 및 코일드-코일 도메인 함유 1), 및 이의 임의 조합.

[1439] 유전자 변형된 동물 또는 세포는 세크레타제 장애와 관련된 단백질을 코딩하는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 파괴된 염색체 서열 및 세크레타제 장애와 관련된 파괴된 단백질을 코딩하는 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 염색체 통합 서열을 포함할 수 있다.

[1440] **ALS**

[1441] 예를 들어, 미국 특허 공개 제20110023144호는 근위축성 측삭경화증 (ALS) 질환과 관련된 세포, 동물 및 단백질을 유전자 변형시키기 위한 징크 핑거 뉴클레아제의 사용을 설명한다. ALS는 수의 운동에 수반되는 뇌피질, 뇌 줄기, 및 척수에서 특정 신경 세포의 점진적인 꾸준한 변성을 특징으로 한다. 당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 미국 특허 공개 제20110023144호의 것과 유사한 시스템에서 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 바람직한 구현예에서, 스태거드 DSB는 HR 독립적 기전, 예컨대 NHEJ를 통해 복구된다. 일부 구현예에서, 표적 세포는 비분열 세포이다. 특정한 실시형태에서, 표적 항원은 운동 뉴런이다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1442] 운동 뉴런 장애 및 이들 장애와 관련된 단백질은 운동 뉴런 장애 발생에 대한 감수성, 운동 뉴런 장애의 존재, 운동 뉴런 장애의 중증도, 또는 이들의 어떤 조합에 영향을 미치는 다양한 세트의 단백질이다. 본 개시는 특정한 운동 뉴런 장애인 ALS 질환과 관련된 단백질을 코딩하는 어떤 염색체 서열의 편집을 포함한다. ALS와 관련된 단백질은 전형적으로 ALS-관련 단백질과 ALS의 실험적 연관성에 기초하여 선택된다. 예를 들어, ALS 와 관련된 단백질의 혈중 농도나 생성 비율이 ALS가 없는 집단에 비해 ALS를 가진 집단에서 상승되거나 억제될 수 있다. 단백질 수준에서의 차이는 웨스턴 블롯, 면역 조직화학적 염색, 효소 결합 면역 흡착 분석 (ELISA), 및 질량 분석을 포함하는 프로테오믹스 기술을 사용하여 평가될 수 있으나, 이에 제한되지 않는다.
대안적으로, ALS 와 관련된 단백질은, 제한은 아니지만, DNA 마이크로어레이 분석, 연속 유전자 발현 분석 (SAGE), 및 정량적 실시간 중합효소 연쇄 반응 (Q-PCR)을 포함하는 게놈 기술을 사용하여 단백질을 코딩하는 유전자의 유전자 발현 프로파일을 얻음으로써 확인될 수 있다.

[1443] 비제한적인 예로서, ALS와 연관된 단백질은 제한없이 다음을 포함한다: SOD1 수퍼옥시드 디스뮤타제 1, ALS3 근위축성 축삭 가용 경화증 3 SETX 세나탁틴 ALS5 근위축성 축삭 경화증 5 육종 ALS7 근위축성 축삭 경화증 7에 융합된 FUS ALS2 근위축성 축삭 DPP6 디펩티딜-펩티다제 6 경화증 2 NEFH 뉴로필라멘트, 중 PTGS1 프로스타글란틴- 폴리펩티드 엔도퍼옥시다제 신타제 1 SLC1A2 용질 운반체 패밀리 1 TNFRSF10B 종양 괴사 인자 (신경교 고친화성 수용체 수퍼패밀리, 글루타메이트 수송체), 멤버 10b 멤버 2 PRPH 페리페린 HSP90AA1 열충격 단백질 90 kDa 알파 (시토졸), 클래스 A 멤버 1 GRIA2 글루타메이트 수용체, IFNG 인터페론, 감마 이온성, AMPA 2 S100B S100 칼슘 결합 FGF2 섬유아세포 성장 인자 2 단백질 B AOX1 알데히드 옥시다제 1 CS 시트레이트 신타제 TARDBP TAR DNA 결합 단백질 TXN 티오레독신 RAPH1 Ras 연합 MAP3K5 미토겐-활성화 단백질 (RaIGDS/AF-6) 및 키나제 5 플레크스트린 상동성 도메인 1 NBEAL1 뉴로베아킨-유사 1 GPX1 글루타티온 퍼옥시다제 1 ICA1L 섬 세포 자기항원 RAC1 ras-관련 C3 보툴리늄 1.69 kDa-유사 독소 기질 1 MAPT 마이크로튜불-연관 ITPR2 이노시톨 1,4,5- 단백질 tau 트리포스페이트 수용체, 2형 ALS2CR4 근위축성 축삭 GLS 글루타미나제 경화증 2 (소아) 염색체 영역, 후보 4 ALS2CR8 근위축성 축삭 CNTFR 섬모 향신경성 인자 경화증 2 (소아) 수용체 염색체 영역, 후보 8 ALS2CR11 근위축성 축삭 FOLH1 폴레이트 히드롤라제 1 경화증 2 (소아) 염색체 영역, 서열 P4HB 프롤릴 4-히드롤라제를 갖는 후보 11 FAM117B 패밀리, 유사성 117, 멤버 B 베타 폴리펩티드 CNTF 섬모 향신경성 인자 SQSTM1

시퀘스토솜 1 STRADB STE20-관련 키나제 NAIP NLR 패밀리, 아폽토시스 어댑터 베타 억제성 단백질 YWHAQ 티로신 3- SLC33A1 용질 운반체 패밀리 33 모노옥시게나제/트립토프 (아세틸-CoA 수송체), 5-모노옥시게나제 멤버 1 활성화단백질, 쎄타 폴리펩티드 TRAK2 수송 단백질, 상동체, SAC1 키네신 결합 2 지질 포스파타제 도메인 함유 NIF3L1 NIF3 NGG1 상호작용 INA 인터넥신 신경 인자3-유사 1 중간 필라멘트 단백질, 알파 PARD3B par-3 분획화 COX8A 시토크롬 c 옥시다제 결핍 3 상동체 B 서브유닛 VIIIA CDK15 사이클린-의존적 키나제 HECW1 HECT, C2 및 WW 15 도메인 함유 E3 유비퀴틴 단백질 리가제 1 NOS1 산화질소 신타제 1 MET met 프로토-종양 유전자 SOD2 수퍼옥시드 디스뮤타제 2, HSPB1 열충격 27 kDa 미토콘드리아 단백질 1 NEFL 뉴로필라멘트, 경 CTSB 카텝신 B 폴리펩티드 ANG 안지오게닌, HSPA8 열충격 70 kDa 리보뉴클레아제, RNase A 단백질 8 패밀리, 5 VAPB VAMP (소포- ESR1 에스트로겐 수용체 1 연관 막단백질)-연관 단백질 B 및 C SNCA 시누클레인, 알파 HGF 간세포 성장 인자 CAT 카탈라제 ACTB 액틴, 베타 NEFM 뉴로필라멘트, 중간 TH 티로신 히드롤라제 폴리펩티드 BCL2 B-세포 CLL/림프종 2 FAS Fas (TNF 수용체 수퍼패밀리, 멤버 6) CASP3 캐스파제 3, 아폽토시스- CLU 클러스테린 관련 시스테인 펩티다제 SMN1 모터 뉴런의 생존 G6PD 글루코스-6-포스페이트 1, 텔로머 디히드로게나제 BAX BCL2-연관 X HSF1 열충격 전사 단백질 인자 1 RNF19A 링 핑거 단백질 19A JUN jun 종양 유전자 ALS2CR12 근위축성 축삭 HSPA5 열충격 70 kDa 경화증 2 (소아) 단백질 5 염색체 영역, 후보 12 MAPK14 미토겐-활성화 단백질 IL10 인터루킨 10 키나제 14 APEX1 APEX 뉴클레아제 TXNRD1 티오레독신 리덕타제 1 (다기능 DNA 복구 효소) 1 NOS2 산화질소 신타제 2, TIMP1 TIMP 메탈로펩티다제 유도성 억제제 1 CASP9 캐스파제 9, 아폽토시스- XIAP 관련 시스테인 아폽토시스 펩티다제의 X-연관 억제제 GLG1 골지 당단백질 1 EPO 에리쓰로포이어틴 VEGFA 혈관 내피 ELN 엘라스틴 성장 인자 A GDNF 신경교 세포 유래 NFE2L2 핵 인자 (적혈구-신경영양 인자 유래 2)-유사 2 SLC6A3 용질 운반체 패밀리 6 HSPA4 열충격 70 kDa (신경전달인자 단백질 4 수송체, 도파민), 멤버 3 APOE 아포지단백질 E PSMB8 프로테오솜 (프로솜, 마크로파인) 서브유닛, 베타 유형, 8 DCTN1 디낙틴 1 TIMP3 TIMP 메탈로펩티다제 억제제 3 KIFAP3 키네신-연관 SLC1A1 용질 운반체 패밀리 1 단백질 3 (신경원/상피 고친화성 글루타메이트 수송체, 시스템 Xag), 멤버 1 SMN2 모터 뉴런의 생존 CCNC 사이클린 C 2, 동심체 MPP4 막단백질, STUB1 STIP1 상동성 및 U- 팔미토일화4 박스 함유 단백질 1 ALS2 아밀로이드 베타 (A4) PRDX6 퍼옥시리독신 6 전구체 단백질 SYP 시냅토피신 CABIN1 칼시뉴린 결합 단백질 1 CASP1 캐스파제 1, 아폽토시스- GART 포스포리보실글리신아미드 관련 시스테인 포르밀트랜스퍼라제, 펩티다제 포스포리보실글리신아미드 신써타제, 포스포리보실아미노이미다졸 신써타제 CDK5 사이클린-의존적 키나제 5 ATXN3 아탁신 3 RTN4 레티큘론 4 C1QB 보체 성분 1, q 서브성분, B 사슬 VEGFC 신경 성장 인자 HTT 헌팅틴 수용체 PARK7 파킨슨병 7 XDH 잔틴 디히드로게나제 GFAP 신경교 섬유질 산성 MAP2 마이크로튜불-연관 단백질 단백질 2 CYCS 시토크롬 c, IgG의 체세포 FCGR3B Fc 단편, 저친화성 IIIb, UBL5 유비퀴틴-유사 5 수퍼옥시드 디스뮤타제용 CCS 구리 샤페론 MMP9 매트릭스 메탈로펩티다제 SLC18A3 용질 운반체 패밀리 18 9 ((소포 아세틸콜린), 멤버 3 TRPM7 일시적 수용체 HSPB2 열충격 27 kDa 잠재적 양이온 채널, 단백질 2 서브패밀리 M, 멤버 7 AKT1 v-akt 쥐과 흉선종 DERL1 Der1-유사 도메인 패밀리, 바이러스 종양유전자 상동체 1 멤버 1 CCL2 케모카인 (C--C 모티프) NGRN 뉴그린, 뉴라이트 리간드 2 과성장 연관 GSR 글루타티온 리덕타제 TPPP3 튜불린 중합-촉진 단백질 패밀리 멤버 3 APAF1 아폽토시스 펩티다제 BTBD10 BTB (POZ) 도메인 활성화 인자 1 함유 10 GLUD1 글루타메이트 CXCR4 케모카인 (C--X--C 모티프) 디히드로게나제 1 수용체 4 SLC1A3 용질 운반체 패밀리 1 FLT1 fms-관련 티로신 (신경교 고친화성 글루타메이트 수송체), 멤버 3 키나제 1 PON1 파라옥소나제 1 AR 안드로겐 수용체 LIF 백혈병 억제성 인자 ERBB3 v-erb-b2 적아세포 백혈병 바이러스 종양유전자 상동체 3 LGALS1 렉틴, 갈락토시드- CD44 CD44 분자 결합, 가용성, 1 TP53 종양 단백질 p53 TLR3 toll-유사 수용체 3 GRIA1 글루타메이트 수용체, GAPDH 글리세르알데히드-3- 이온성, AMPA 1 포스페이트 디히드로게나제 GRIK1 글루타메이트 수용체, DES 데스민 이온성, 카이네이트 1 CHAT 콜린 아세틸트랜스퍼라제 FLT4 fms-관련 티로신 키나제 4 CHMP2B 크로마틴 변형 BAG1 BCL2-연관 단백질 2B 아타노겐 MT3 메탈로티오테인 3 CHRNA4 콜린성 수용체, 니코틴성, 알파 4 GSS 글루타티온 신써타제 BAK1 BCL2-길항제/킬러 1 KDR 키나제 삽입 도메인 GSTP1 글루타티온 S-트랜스퍼라제 수용체 (III형 pi 1 수용체 티로신 키나제) OGG1 8-옥소구아닌 DNA IL6 인터루킨 6 (인터페론, 글리코실라제 베타 2).

[1444] 동물 또는 세포는 ALS 와 관련된 단백질을 코딩하는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 파괴된 염색체 서열 및 ALS 와 관련된 파괴된 단백질을 코딩하는 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 염색체 통합 서열을 포함할 수 있다. ALS와 관련된 바람직한 단백질은 SOD1 (수퍼옥사이드 디스뮤타제 1), ALS2(근위축성 측삭경화증 2), FUS (육종 융합형), TARDBP (TAR DNA 결합 단백질), VAGFA(혈관 내피 성장인자 A), VAGFB(혈관 내피 성장인자 B), 및 VAGFC(혈관 내피 성장인자 C), 및 이들의 어떤 조합을 포함한다.

[1445] **자폐증**

[1446] 미국 특허 공개 제20110023145호는 자폐 범주성 장애 (ASD)와 연관된 세포, 동물 및 단백질을 유전적으로 변형

시키기 위해 아연 핑거 뉴클레아제를 사용하는 것을 기재한다. 자폐 범주성 장애(ASD)는 사회적 상호작용 및 소통의 정량적 손상, 및 제한된 반복적이며 정형화된 행동, 관심 및 활동 패턴을 특징으로 하는 일군의 장애이다. 세 가지 장애, 즉 자폐, 아스퍼거 증후군(AS) 및 비전형 전반적 발달 장애(PDD-NOS)가 중증도의 정도, 관련된 지적 기능 및 의학적 상태가 다른 연속된 동일한 장애이다. ASD는 대개 유전적으로 결정되는 장애이며, 유전성은 약 90%이다.

[1447] US 특허 공개 제 20110023145호는 본 발명의 CRISPR Cas 시스템에 적용될 수 있는 ASD와 관련된 단백질을 코딩하는 어떤 염색체 서열의 편집을 포함한다. ASD와 관련된 단백질은 전형적으로 ASD와 관련관 단백질과 ASD의 유발 또는 징후의 실험적 연관성에 기초하여 선택된다. 예를 들어, ASD와 관련된 단백질의 혈중 농도나 생성 비율은 ASD를 결여한 집단에 비해 ASD를 가진 집단에서 상승되거나 억제될 수 있다. 단백질 수준에서의 차이는 웨스턴 블롯, 면역 조직화학적 염색, 효소 결합 면역 흡착 분석 (ELISA), 및 질량 분석을 포함하는 프로테오믹스 기술을 사용하여 평가될 수 있으나, 이에 제한되지 않는다. 대안적으로, ASD 와 관련된 단백질은, 제한은 아니지만, DNA 마이크로어레이 분석, 연속 유전자 발현 분석 (SAGE), 및 정량적 실시간 중합효소 연쇄 반응 (Q-PCR)을 포함하는 게놈 기술을 사용하여 단백질을 코딩하는 유전자의 유전자 발현 프로파일을 얻음으로써 확인될 수 있다. 당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 미국 특허 공개 제20110023145호의 것과 유사한 시스템에서 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1448] ASD와 관련된 단백질과 관련될 수 있는 질환 상태 또는 장애의 비제한적 예들은 자폐, 아스퍼거 증후군(AS), 비전형 전반적 발달 장애(PDD-NOS), 레트 증후군, 결절성 경화증, 페닐케톤뇨증, 스미스-렘리-오피츠 증후군 및 취약 X 증후군을 포함한다. 비제한적 예로서, ASD와 관련된 단백질은 제한은 아니지만 다음의 단백질들을 포함한다:ATP10C 아미노인지질-MET MET 수용체 수송 ATP 분해효소 티로신 키나제(ATP10C) BZRAP1 MGLUR5(GRM5) 대사성 글루타메이트 수용체 5(MGLUR5) CDH10 캐드헤린-10 MGLUR6(GRM6) 대사성 글루타메이트 수용체 6(MGLUR6) CDH9 캐드헤린-9 NLGN1 뉴로리긴-1 CNTN4 컨택틴-4 NLGN2 뉴로리긴-2 CNTNAP2 컨택틴-관련 SEMA5A 뉴로리긴-3 단백질-유사 2(CNTNAP2) DHCR7 7-디하이드로콜레스테롤 NLGN4X 뉴로리긴-4 X- 환원효소(DHCR7) 결합 DOC2A 이중 C2-유사 도메인-NLGN4Y 뉴로리긴-4 Y-함유 단백질 알파 결합 DPP6 디펩티딜 NLGN5 뉴로리긴-5 아미노펩티다제-유사 단백질 6 EN2 분절형 2(EN2) NRCAM 뉴런 세포 부착 분자(NRCAM) MDGA2 취약 X 지적 장애 NRXN1 뉴렉신-11(MDGA2) FMR2(AFF2) AF4/FMR2 패밀리 멤버 2 OR4M2 올팩토리 수용체(AFF2) 4M2 FOXP2 포크헤드 박스 단백질 P2 OR4N4 올팩토리 수용체(FOXP2) 4N4 FXR1 취약 X 정신 OXTR 옥시토신 수용체 발달지연, 상염색체(OXTR) 상동체 1 (FXR1) FXR2 취약 X 정신 PAH 페닐알라닌 발달지연, 상염색체 하이드록실라제(PAH) 상동체 2(FXR2) GABRA1 감마-아미노부티르산 PTEN 포스파타제 및 수용체 서브유닛 알파-1 텐신 상동성 (GABRA1)(PTEN) GABRA5 GABAA(.감마.-아미노부티르 PTPRZ1 수용체-타입 산) 수용체 알파 5 티로신-단백질 서브유닛 (GABRA5) 포스파타제 제타(PTPRZ1) GABRB1 감마-아미노부티르산 RELN 렐린 수용체 서브유닛 베타-1(GABRB1) GABRB3 GABAA(.감마.-아미노부티르 RPL10 60S 리보솜 산) 수용체 .베타.3 서브유닛 단백질 L10(GABRB3)GABRG1 감마-아미노부티르산 SEMA5A 세마포린-5A 수용체 서브유닛 감마-1(SEMA5A) (GABRG1) HIRIP3 HIRA-상호작용 단백질 3 SEZ6L2 발작 관련 6 상동체(마우스)-유사 2 HOXA1 호메오박스 단백질 Hox-A1 SHANK3 SH3 및 다중 (HOXA1) 안키린 반복부 도메인 3(SHANK3) IL6 인터루킨-6 SHBZRAP1 SH3 및 다중 안키린 반복부 도메인 3(SHBZRAP1) LAMB1 라미닌 서브유닛 베타-1 SLC6A4 세로토닌 (LAMB1) 수송인자(SERT) MAPK3 미토겐-활성화 단백질 TAS2R1 테이스트 수용체 키나제 3 타입 2 멤버 1 TAS2R1 MAZ Myc-관련 징크 핑거 TSC1 결절성 경화증 단백질 단백질 1 MDGA2 MAM 도메인 함유 TSC2 결절성 경화증 글리코실포스파티딜이노시톨 단백질 2 앵커 2(MDGA2) MECP2 메틸 CpG 결합 UBE3A 유비퀴틴 단백질 단백질 2(MECP2) 리가아제 E3A (UBE3A) MECP2 메틸 CpG 결합 WNT2 무익기형-타입 단백질 2(MECP2) MMTV 통합 부위 패밀리, 멤버 2(WNT2).

[1449] 염색체 서열이 편집된 ASD와 관련된 단백질의 동일성은 다양할 수 있으며 다양할 것이다. 바람직한 구현예에서, 염색체 서열이 편집된 ASD와 관련된 단백질은 BZRAP 1 유전자에 의해서 암호화된 벤조디아자핀 수용체(말초) 관련 단백질 1 (BZRAP1), AFF2 유전자에 의해서 암호화된 AF4/FMR2 패밀리 멤버 2 단백질(AFF2) (MFR2라고

도 한다), FXR1 유전자에 의해서 암호화된 취약 X 지적 장애 상염색체 상동체 1 단백질(FXR1), FXR2 유전자에 의해서 암호화된 취약 X 지적 장애 상염색체 상동체 2 단백질(FXR2), MDGA2 유전자에 의해서 암호화된 MAM 도메인 함유 글리코실포스파티딜이노시톨 앵커 2 단백질(MDGA2), MECP2 유전자에 의해서 암호화된 메틸 CpG 결합 단백질 2(MECP2), MGLUR5-1 유전자에 의해서 암호화된 대사성 글루타메이트 수용체 5 (MGLUR5) (GRM5라고도 한다), NRXN1 유전자에 의해서 암호화된 뉴렉신 1 단백질, 또는 SEMA5A 유전자에 의해서 암호화된 세마포린-5A 단백질(SEMA5A)일 수 있다. 예시적인 구현예에서, 유전자 변형된 동물은 래트이며, ASD와 관련된 단백질을 암호화하는 편집된 염색체 서열은 아래 열거된 것과 같다:BZRAP1 벤조디아자핀 수용체 XM_002727789, (말초) 관련 XM_213427, 단백질 1 (BZRAP1) XM_002724533, XM_001081125 AFF2(FMR2) AF4/FMR2 패밀리 멤버 2 XM_219832, (AFF2) XM_001054673 FXR1 취약 X 정신 NM_001012179 발달지연, 상염색체 상동체 1 (FXR1) FXR2 취약 X 정신 NM_001100647 발달지연, 상염색체 상동체 2(FXR2) MDGA2 MAM 도메인 함유 NM_199269 글리코실포스파티딜이노시톨 앵커 2(MDGA2) MECP2 메틸 CpG 결합 NM_022673 단백질 2(MECP2) MGLUR5 대사성 글루타메이트 NM_017012(GRM5) 수용체 5 (MGLUR5) NRXN1 뉴렉신-1 NM_021767 SEMA5A 세마포린-5A (SEMA5A) NM_001107659

[1450] **트리뉴클레오티드 반복 확장 장애**

[1451] 미국 특허 공개 제20110016540호는 트리뉴클레오티드 반복 확장 장애와 연관된 세포, 동물 및 단백질을 유전적으로 변형시키기 위해 아연 핑거 뉴클레아제를 사용하는 것을 기재한다. 트리뉴클레오티드 반복 확장 장애는 발생 신경생물학과 관련되고 종종 인지뿐만 아니라 감각 운동 기능에 영향을 미치는 복합, 진행성 장애이다.

당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 미국 특허 공개 제20110016540호의 것과 유사한 시스템에서 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3'이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1452] 트리뉴클레오티드 반복 확장 단백질은 트리뉴클레오티드 반복 확장 장애가 발생할 감수성, 트리뉴클레오티드 반복 확장 장애의 존재, 트리뉴클레오티드 반복 확장 장애의 중증도 또는 이들의 임의의 조합과 연관된 다양한 세트의 단백질이다. 트리뉴클레오티드 반복 확장 장애는 반복 유형에 의해 결정되는 2 가지 범주로 나누어진다. 가장 통상적인 반복은, 유전자의 암호 영역에서 존재할 때 아미노산 글루타민(Q)에 대해 암호화나는 트리플렛 CAG이다. 그러므로, 이들 장애는 폴리글루타민 (폴리Q) 장애라고 하고 하기 질환을 포함한다: 헌팅톤병 (HD); 척수구근 근위축증 (SBMA); 척수소뇌 운동실조증 (SCA 유형 1, 2, 3, 6, 7, 및 17); 및 DRPLA (Dentatorubro-Pallidoluysian Atrophy). 남아있는 트리뉴클레오티드 반복 확장 장애는 CAG 트리플렛을 수반하지 않거나 또는 CAG 트리플렛은 유전자의 암호 영역에 있지 않고, 따라서 비 폴리글루타민 장애로서 지칭된다.

비-폴리글루타민 장애는 취약 X 증후군 (FRAXA); 취약 XE 정신 지체증 (FRAXE); 프리드리히 운동실조증 (FRDA); 근긴장성 이영양증 (DM); 및 척수소뇌 운동실조증 (SCA 유형 8, 및 12)을 포함한다.

[1453] 트리뉴클레오티드 반복 확장 장애와 관련된 단백질은 트리뉴클레오티드 반복 확장 장애에 대한 트리뉴클레오티드 반복 확장 장애와 연관된 단백질의 실험적 연관에 기반하여 전형적으로 선택된다. 예를 들어, 트리뉴클레오티드 반복 확장 장애와 연관된 단백질의 생성 속도 또는 순환 농도는 트리뉴클레오티드 반복 확장 장애가 없는 집단에 비해 트리뉴클레오티드 반복 확장 장애를 갖는 집단에서 상승 또는 억제될 수 있다. 단백질 수준에서의 차이는 웨스턴 블롯, 면역 조직화학적 염색, 효소 결합 면역 흡착 분석 (ELISA), 및 질량 분석을 포함하는 프로테오믹스 기술을 사용하여 평가될 수 있으나, 이에 제한되지 않는다. 대안적으로, 트리뉴클레오티드 반복 확장 장애와 연관된 단백질은 DNA 마이크로어레이 분석, 일련 분석 유전자 발현(SAGE), 및 정량적 실시간 중합효소 연쇄반응(Q-PCR)을 포함하지만, 이들로 제한되지 않는 게놈 기법을 사용하여 단백질을 코딩하는 유전자의 유전자 발현 프로파일을 얻음으로써 확인될 수 있다.

[1454] 트리뉴클레오티드 반복부 확장 장애와 연관된 단백질의 비제한적인 예는 다음을 포함한다: AR (안드로겐 수용체), FMR1 (취약 X 정신 지체 1), HTT (헌팅틴), DMPK (근긴장 이상증-단백질 키나제), FXN (프라탁신), ATXN2 (아탁신 2), ATN1 (아트로핀 1), FEN1 (플랍 구조-특이적 엔도뉴클레아제 1), TNRC6A (트리뉴클레오티드 반복 함유 6A), PABPN1 (폴리(A) 결합 단백질, 핵 1), JPH3 (정토필린 3), MED15 (매개인자 복합체 서브유닛

15), ATXN1 (아탁신 1), ATXN3 (아탁신 3), TBP (TATA 박스 결합 단백질), CACNA1A (칼슘 채널, 전압-의존적, P/Q 유형, 알파 1A 서브유닛), ATXN80S (ATXN8 반대 가닥 (비-단백질 코딩)), PPP2R2B (단백질 포스파타제 2, 조절 서브유닛 B, 베타), ATXN7 (아탁신 7), TNRC6B (트리뉴클레오티드 반복 함유 6B), TNRC6C (트리뉴클레오티드 반복 함유 6C), CELF3 (CUGBP, Elav-유사 패밀리 멤버 3), MAB21L1 (mab-21-유사 1 (예쁜 꼬마 선충)), MSH2 (mutS 상동체 2, 결장암, 비용종증 1형 (이. 콜라이)), TMEM185A (경막 단백질 185A), SIX5 (SIX 호메오박스 5), CNPY3 (카노피 3 상동체 (제브라피쉬)), FRAXE (취약 부위, 폴산 유형, 희귀, fra(X)(q28) E), GNB2 (구아닌 뉴클레오티드 결합 단백질 (G 단백질), 베타 폴리펩티드 2), RPL14 (리보솜 단백질 L14), ATXN8 (아탁신 8), INSR (인슐린 수용체), TTR (트랜스타이레틴), EP400 (E1A 결합 단백질 p400), GIGYF2 (GRB10 상호작용 GYF 단백질 2), OGG1 (8-옥소구아닌 DNA 글리코실라제), STC1 (스탄니오칼신 1), CNDP1 (카르노신 디펩티다제 1 (메탈로펩티다제 M20 패밀리)), C10orf2 (염색체 10 오픈 리딩 프레임 2), MAML3 마스터민드-유사 3 (초파리), DKC1 (선천성 이각화증 1, 디스케린), PAXIP1 (PAX 상호작용 (전사-활성화도메인) 단백질 1), CASK (칼슘/칼모듈린-의존적 세린 단백질 키나제 (MAGUK 패밀리)), MAPT (마이크로튜불-연관 단백질 tau), SP1 (Sp1 전사 인자), POLG (중합효소 (DNA 지정), 감마), AFF2 (AF4/FMR2 패밀리, 멤버 2), THBS1 (트롬보스폰딘 1), TP53 (종양 단백질 p53), ESR1 (에스트로겐 수용체 1), CGGBP1 (CGG 트리플릿 반복 결합 단백질 1), ABT1 (기본 전사의 활성인자 1), KLK3 (칼리크레인-관련 펩티다제 3), PRNP (프라이온 단백질), JUN (jun 종양 유전자), KCNN3 (포타슘 중간체/소형 컨덕턴스 칼슘-활성화된 채널, 서브패밀리 N, 멤버 3), BAX (BCL2-연관 X 단백질), FRAXA (취약 부위, 폴산 유형, 희귀, fra(X)(q27.3) A (거대고환, 정신 지체)), KBTBD10 (kelch 반복 및 BTB (POZ) 도메인 함유 10), MBNL1 (머슬블라인드-유사 (초파리)), RAD51 (RAD51 상동체 (RecA 상동체, 이. 콜라이) (에스. 세레비지아)), NCOA3 (핵 수용체 공활성화제 3), ERDA1 (확장 반복 도메인, CAG/CTG 1), TSC1 (결절성 경화증 1), COMP (연골 올리고머 매트릭스 단백질), GCLC (글루타메이트-시스테인 리가제, 촉매 서브유닛), RRAD (Ras-관련 당뇨병 연관), MSH3 (mutS 상동체 3 (이. 콜라이)), DRD2 (도파민 수용체 D2), CD44 (CD44 분자 (인도 혈액군)), CTCF (CCCTC-결합 인자 (아연 핑거 단백질)), CCND1 (사이클린 D1), CLSPN (클라스핀 상동체 (제노푸스 라에비스 (Xenopus laevis)), MEF2A (근세포 인핸서 인자 2A), PTPRU (단백질 티로신 포스파타제, 수용체 유형, U), GAPDH (글리세르알데히드-3-포스페이트 디히드로게나제), TRIM22 (3원 모티프-함유 22), WT1 (빌름스 종양 1), AHR (아릴 탄화수소 수용체), GPX1 (글루타티온 퍼옥시다제 1), TPMT (티오푸린 S-메틸트랜스퍼라제), NDP (노리에병 (가신경교종)), ARX (아리스탈레스 (aristaless) 관련 호메오박스), MUS81 (MUS81 엔도뉴클레아제 상동체 (에스. 세레비지아에)), TYR (티로시나제 (눈피부 백색증 IA)), EGR1 (초기 성장 반응 1), UNG (우라실-DNA 글리코실라제), NUMBL (numb 상동체 (초파리)-유사), FABP2 (지방산 결합 단백질 2, 장), EN2 (엔그레일드 (engrailed) 호메오박스 2), CRYGC (크리스탈린, 감마 C), SRP14 (신호 인식 입자 14 kDa (상동성 Alu RNA 결합 단백질)), CRYGB (크리스탈린, 감마 B), PDCD1 (프로그램된 세포 사멸1), HOXA1 (호메오박스 A1), ATXN2L (아탁신 2-유사), PMS2 (PMS2 감수분열후 분리 증가 2 (에스. 세레비지아에)), GLA (갈락토시다제, 알파), CBL (Cas-Br-M (쥐과) 동종숙주 레트로바이러스 형질전환 서열), FTH1 (페리틴, 헤비 폴리펩티드 1), IL12RB2 (인터루킨 12 수용체, 베타 2), OTX2 (오르토덴티클 호메오박스 2), HOXA5 (호메오박스 A5), POLG2 (중합효소 (DNA 지정), 감마 2, 보조 서브유닛), DLX2 (디스탈-리스 (distal-less) 호메오박스 2), SIRPA (신호-조절 단백질 알파), OTX1 (오르토덴티클 호메오박스 1), AHRR (아릴-탄화수소 수용체 억제인자), MANF (중쇄 성상세포-유래 신경영양 인자), TMEM158 (경막 단백질 158 (유전자/가유전자)), 및 ENSG00000078687.

[1455] 트리뉴클레오티드 반복 확장 장애와 연관된 바람직한 단백질은 HTT(헌팅틴), AR(안드로겐 수용체), FXN(프라탁신), Atxn3(아탁신), Atxn1(아탁신), Atxn2(아탁신), Atxn7(아탁신), Atxn10(아탁신), DMPK(근긴장성 이영양증-단백질 키나제), Atn1(아트로핀 1), CBP(creb 결합 단백질), VLDLR(매우 저밀도 지단백질 수용체), 및 이들의 임의의 조합을 포함한다.

[1456] **청각 질병의 치료**

[1457] 본 발명은 또한 하나 또는 두 귀로 CRISPR-Cas 시스템을 전달하는 것을 고려한다.

[1458] 연구자는 유전자 치료가 현재 난청 치료-즉, 달팽이관 이식을 원조하기 위해 사용될 수 있는지를 조사하였다. 난청은 종종 청각 뉴런으로 신호를 전달할 수 없는 손실되거나 손상된 모세포에 의해 발생된다. 이러한 경우에서, 달팽이관 이식은 신경 세포로 소리에 대한 반응 및 전기 신호 전달을 위해 사용될 수 있다. 그러나 더 적은 성장 인자가 손상된 모세포에 의해 방출되므로 달팽이관으로부터 이들 뉴런이 종종 악화되고 철회된다.

[1459] 미국 특허 출원 제20120328580호는 예를 들어, 주사기, 예를 들어, 단일-용량 주사기를 사용하여 약학적 조성물

을 귀(예를 들어, 귀 투여), 예컨대 달팽이관의 관내강(luminae)으로(예를 들어, 중앙계, Sc 정전계(vestibulae), 및 고실계) 주입을 기재한다. 예를 들어, 본원에 기재된 하나 이상의 화합물은 고막내 주입(예를 들어, 중이 내로), 및/또는 외이, 중이 및/또는 내이 내로 주입에 의해 투여될 수 있다. 이러한 방법은 예를 들어, 인간 귀로 스테로이드 및 항체의 투여를 위해 당 분야에서 관례적으로 사용된다. 주입은 예를 들어, 귀의 내창 또는 달팽이관 캡슐을 통해 실시될 수 있다. 다른 내이 투여 방법이 당 분야에서 공지되었다 (예를 들어, 문헌 [Salt and Plontke, Drug Discovery Today, 10:1299-1306, 2005]).

[1460] 다른 모드의 투여에서, 약학적 조성물은 카테터 또는 펌프를 통해 원 위치 투여될 수 있다. 카테터 또는 펌프는, 예를 들어, 달팽이관 관내강 또는 귀의 내창 및/또는 결장의 루멘으로 약학적 조성물을 향하게 한다. 본원에서 기재된 하나 이상의 화합물을 귀, 예를 들어 인간 귀로 투여하기 위해 적합한 예시적인 약물 전달 장치 및 방법이 McKenna 등의 문헌(미국 공개 번호 제2006/0030837호) 및 Jacobsen 등의 문헌(미국 특허 제7,206,639호)에 기재되어 있다. 몇몇 구현예에서, 카테터 또는 펌프는 수술 과정 동안, 예를 들어, 환자의 귀 (예를 들어, 외이, 중이, 및/또는 내이)에 위치될 수 있다. 몇몇 구현예에서, 카테터 또는 펌프는 수술 과정 필요 없이, 예를 들어, 환자의 귀 (예를 들어, 외이, 중이, 및/또는 내이)에 위치될 수 있다.

[1461] 대안적으로 또는 더하여, 본원에 기재된 하나 이상의 화합물이 달팽이관 이식 또는 외이에 착용하는 보청기와 같은 기계 장치와의 조합으로 투여될 수 있다. 본 발명과 함께 사용하기에 적합한 예시적인 달팽이관 이식은 Edge 등의 문헌에 의해 설명된다(미국 공개 번호 제2007/0093878호).

[1462] 몇몇 구현예에서, 상기 기재된 투여 모드는 임의의 순서로 조합될 수 있으며 동시에 또는 배치될 수 있다.

[1463] 대안적으로 또는 더하여, 본 발명은 예를 들어, CDER Data Standards Manual, 버전 넘버 004(fda.give/cder/dsm/DRG/drg00301.htm에서 이용 가능함)에 기재된 바와 같은, 임의의 미국 식품의약국(Food and Drug Administration) 승인 방법에 따라 투여될 수 있다.

[1464] 일반적으로, 미국 특허 출원 제20120328580호에 기재된 세포 치료 방법이 실험관내 내이의 성숙 세포 유형(예를 들어, 모세포)에 대해 또는 이를 향해 세포의 완전한 부분적 분화를 촉진하기 위해 사용될 수 있다. 그런 다음 이러한 방법으로부터 야기된 세포는 이러한 치료가 필요한 환자에게 이식되거나 주입될 수 있다. 적합한 세포 유형을 확인하고 스크리닝하기 위한 방법을 포함하는 이들 방법을 실시하기 위해 요구되는 세포 배양 방법, 선택된 세포의 완전한 또는 부분적 분화를 촉진하는 방법, 완전히 또는 부분적으로 분화된 세포 유형을 확인하기 위한 방법, 및 완전히 또는 부분적으로 분화된 세포를 이식하는 방법이 하기에 기재된다.

[1465] 본 발명에서 사용하기에 적합한 세포는, 본원에 기재된 하나 이상의 화합물과 예를 들어 실험관내 접촉했을 때, 내이의 성숙 세포, 예를 들어, 모세포(예를 들어, 내부 및/또는 외부 모세포) 내로 완전하게 또는 부분적으로 분화될 수 있는 세포를 포함하나, 이로 제한되지 않는다. 모세포로 분화될 수 있는 예시적인 세포는 줄기 세포(예를 들어, 내이 줄기 세포, 성인 줄기 세포, 골수 유래 줄기 세포, 배아 줄기 세포, 간엽 줄기 세포, 피부 줄기 세포, iPS 세포, 및 지방 유래 줄기 세포), 선조 세포(예를 들어, 내이 선조 세포), 지지 세포(예를 들어, 다이테르스 세포, 주상 세포, 내부 지골 세포, 덮개(tectal) 세포 및 헨젠 세포), 및/또는 생식 세포를 포함하나, 이로 제한되지 않는다. 내이 감각 세포의 대체를 위한 줄기 세포의 사용은 Li 등(미국 특허 공개 제2005/0287127호) 및 Li 등(미국 특허 출원 제11/953,797호)에 기재되어 있다. 내이 감각 세포의 치환을 위한 골수 유래 줄기 세포의 용도는 [Edge et al., PCT/US2007/084654]에 기술된다. iPS 세포는 예를 들어, 하기 문헌에 기술된다: Takahashi et al., Cell, Volume 131, Issue 5, Pages 861-872 (2007); Takahashi and Yamanaka, Cell 126, 663-76 (2006); Okita et al., Nature 448, 260-262 (2007); Yu, J. et al., Science 318(5858):1917-1920 (2007); Nakagawa et al., Nat. Biotechnol. 26:101-106 (2008); 및 Zaehres and Scholer, Cell 131(5):834-835 (2007). 이러한 적합한 세포는 하나 이상의 조직 특이 유전자의 존재를 분석(예를 들어, 정성적 또는 정량적)하여 확인될 수 있다. 예를 들어, 유전자 발현은 하나 이상의 조직-특이적 유전자의 단백질 산물을 탐지하여 탐지될 수 있다. 단백질 탐지 기술은 적절한 항원에 대한 항체를 사용하는 염색 단백질(예를 들어, 세포 추출물 또는 전체 세포를 사용함)을 수반한다. 이러한 경우에서, 적절한 항원은 조직-특이적 유전자 발현의 단백질 산물이다. 원칙적으로 제1 항체(즉, 항원에 결합하는 항체)가 표지될 수 있지만, 제1 항체(예를 들어, 항-IgG)로 향하는 제2 항체를 사용하는 것이 좀더 일반적(및 시각화를 개선함)이다. 이러한 제2 항체는 형광 색소 또는 비색 반응을 위한 적절한 효소 또는 (전자 현미경을 위한) 금 비드 또는 비오틴-아비딘 시스템과 컨쥬게이트되어, 1차 항체의 위치가 인식될 수 있으므로 항원이 인식될 수 있다.

[1466] 본 발명의 CRISPR Cas 분자는 미국 공개 출원, 제20110142917호로부터 변형된 조성물을 가지고, 외이로 약학적 조성물을 직접 적용하여 귀로 전달될 수 있다. 몇몇 구현예에서 약학적 조성물은 이도(ear canal)에 적용된다. 귀로의 전달은 또한 청각 또는 귀 전달으로서 언급될 수 있다.

[1467] 당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 상기 논의된 특허 공개의 것과 유사한 시스템에서 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1468] 일부 구현예에서, 본 발명의 RNA 분자는 리포솜 또는 리포펙틴 제형 등으로 전달되고, 당업자에게 잘 공지된 방법에 의해 제조될 수 있다. 이러한 방법은 예를 들어, 참조로 본 명세서에 편입되는, 예를 들어, 미국 특허 출원 번호 5,593,972, 5,589,466, 및 5,580,859에 기술되어 있다.

[1469] 포유동물 세포로 siRNA의 증강되고 개선된 전달을 특별히 목적으로 하는 전달 시스템이 개발되었고 (참조: 예를 들어, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-10; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen. 2002, 32: 107-108 및 Simeoni et al., NAR 2003, 31, 11: 2717-2724), 본 발명에 적용할 수 있다. siRNA는 최근에 영장류에서의 유전자 발현의 억제를 위해 성공적으로 사용되었다 (참조: 예를 들어, Tolentino et al., Retina 24(4):660 (본 발명에 또한 적용할 수 있음).

[1470] Qi 등은 본 발명의 핵산 표적화 시스템에 적용될 수 있는 신규한 단백질 전달 기술에 의해 온전한 내창을 통해 내이로 효율적인 siRNA 트랜스펙션을 위한 방법을 기재했다(예를 들어, 문헌[Qi et al., Gene Therapy (2013), 19]을 참조). 특히, 온전한 내창 투과를 통해 내부 및 외부 모세포, 팽대부릉(crista ampullaris), 난형낭반 및 구형낭반을 포함하는 내이의 세포 내로 Cy3-표지 siRNA를 트랜스펙션시킬 수 있는, TAT 이중 가닥 RNA-결합 도메인(TAT-DRBDs)이 다양한 내이 질병의 치료 및 청각 기능의 보존을 위한 생체내 이중 가닥 siRNA 전달에 성공적이었다. 10 mM RNA의 약 40㎕가 귀에 투여하기 위한 용량으로서 고려될 수 있다.

[1471] Rejali 등 (Hear Res. 2007 Jun;228(1-2):180-7)에 따르면, 달팽이관 이식 기능은 이식물에 의한 전기적 자극의 표적인 나선청신경절 뉴런의 양호한 보정에 의해 개선될 수 있으며, 뇌 유래 신경성장 인자(brain derived neurotrophic factor:BDNF)는 실험적으로 청각을 잃은 귀에서 나선청신경절 생존을 향상시키는 것으로 이전에 나타났다. Rejali 등은 BDNF 유전자 삽입을 갖는 바이러스 벡터에 의해 형질도입된 섬유아세포의 코팅을 포함하는 달팽이관 이식 전극의변형된 설계를 시험했다. 이 유형의 생체외 유전자 이송을 완수하기 위해, Rejali 등은 BDNF 유전자 카세트 삽입을 갖는 아데노바이러스로 기니아피그 섬유아세포를 형질도입시키고, 이 세포들이 BDNF를 분비하는 것을 밝힌 후 BDNF-분비 세포를 아가로스 겔을 통해 달팽이관 이식 전극에 부착시키고 고실계에 전극을 이식했다. Rejali 등은 BDNF 발현 전극이 대조 전극과 비교했을 때 이식 48일 후에 달팽이관의 기저 회전에서 상당히 많은 나선 신경절 뉴런을 보존할 수 있었음을 밝혔으며 나선 신경절 뉴런 생존을 증강시키기 위해 생체외 유전자 이송과 달팽이관 이식 치료법을 조합하는 가능성을 밝혔다. 이러한 시스템은 귀로의 전달을 위한 본 발명의 핵산 표적화 시스템에 적용될 수 있다.

[1472] Mukherjea 등(Antitoxidants & Redox Signaling, Volume 13, Number 5, 2010)은 손상으로부터 OHC의 보호에 의해 입증되는 바와 같이 짧은 간섭 (si) RNA를 사용하는 NOX3의 녹다운이 시스플라틴 이독성(ototoxicity)을 없애며 청성뇌간반응(auditory brainstem response:ABR)에서 역치 이동을 감소시켰다는 것을 기록하였다. 상이한 용량의 siNOX3 (0.3, 0.6, 및 0.9 ㎍)이 래트에 투여되고 NOX3 발현이 실시간 RT-PCR로 평가됐다. 사용된 NOX3 siRNA의 가장 낮은 용량(0.3 ㎍)은 스크램블드 siRNA의 고막투과 투여 또는 비처리된 달팽이관과 비교했을 때 NOX3 mRNA의 임의의 저해를 보이지 않았다. 그러나, 고용량의 NOX3 siRNA (0.6 및 0.9 ㎍)의 투여는 대조 스크램블드 siRNA에 비하여 NOX3 발현을 감소시켰다. 이러한 시스템은 인간으로의 투여를 위해 약 2 mg 내지 약 4 mg 용량의 CRISPR Cas를 고막투과 투여하기 위한 본 발명의 CRISPR Cas 시스템에 적용될 수 있다.

[1473] Jung 등은 iRNA의 적용 후 소낭에서의 Hes5 수준이 감소하고 이러한 소낭에서의 다수의 모세포가 대조 치료 후보다 상당히 더 큰 것을 입증했다(문헌 [Molecular Therapy, vol. 21 no. 4, 834841 apr. 2013]). 데이터는 siRNA 기술이 내이에서의 복구 및 재생을 유도하는데 유용할 수 있고 Notch 신호화 경로가 특이 유전자 발현 저해에 대한 잠재적으로 유용한 표적임을 제시한다. Jung 등은 8 ㎍의 Hes5 siRNA를 2 ㎕ 부피로 주입했으며, 이는 귀의 전정 상피에 대한 동결건조된 siRNA에 멸균된 일반 식염수를 첨가하여 제조되었다. 이러한 시스템은 인간으로의 투여를 위한 약 1 내지 약 30 mg 용량으로 CRISPR Cas을 귀의 전정 상피에 투여하기 위해 본 발명의 핵산 표적화 시스템에 적용될 수 있다. 당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 상기 논의된 특허 공개의 것과 유사한 시스템에서 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1474] **비분열 세포 (뉴런 및 근육)에서 유전자 표적화**

[1475] 비분열 (비분열, 완전 분화) 세포 유형은 예를 들어 상동성 재조합 (Hr)이 일반적으로 G1 세포-주기 시기에서 억제되기 때문에, 유전자 표적화 또는 게놈 조작에 문제를 제기한다. 그러나, 세포가 정상 DNA 복구를 조절하는 기전을 연구하면서, Durocher는 비분열 세포에서 HR을 " 끔" 로으로 유지시키는 이전에 알려지지 않은 스위치를 발견하였고 이러한 스위치를 다시 켜는 전략을 고안하였다. Orthwein 등 (Daniel Durocher’s lab at the Mount Sinai Hospital in Ottawa, Canada)의 최근 보고 (Nature 16142, published online 9 Dec 2015)는 HR의 억제가 해제될 수 있고 유전자 표적화가 신장 (293T) 및 골육종 (U2OS) 세포에서 성공적이라고 결론내렸다. 종양 억제인자, BRCA1, PALB2 및 BRAC2는 Hr에 의한 DNA DSB 복구를 촉진하는 것으로 알려져 있다. 그들은 PALB2 -BRAC2와 BRCA1의 복합체의 형성이 PALB 상의 유비퀴틴 부위에 의해 지배되어서, E3 유비퀴틴 리가제에 의한 부위상의 작용을 확인하였다. 이러한 E3 유비퀴틴 리가제는 쿨린-3 (CUL3)-RBX1과의 복합체에 KEAP1 (PALB2 -상호작용 단백질)로 구성된다 PALB2 유비퀴틴화는 BRCA1과 이의 상호작용을 억제하고 그 자체로 세포 주기 제어 하에 있는, 디유비퀴틸라제 USP11에 의해 대항된다. DNA-말단 절제의 활성화되 조합된 BRCA1-PALB2 상호작용의 복원은 (pX459 벡터로부터 발현된) USP11 또는 KEAP1에서 지정된 CRISPR-Cas9-기반 유전자-표적화 어세이를 포함한 수많은 방법으로 측정하여, G1에서 상동성 재조합을 유도하기에 충분하다. 그러나, BRCA1-PALB2 상호작용이 KEAP1 고갈 또는 PALB2-KR 돌연변이체의 발현을 사용한 절제-적격 G1 세포에서 복원되었을 때, 유전자-표적화 사건에서 강력한 증가가 검출되었다.

[1476] 따라서, 세포, 특히 비분열, 완전 분화 세포 유형에서 HR의 재활성화가 일부 구현예에서, 바람직하다. 일부 구현예에서, BRCA1-PALB2 상호작용이 일부 구현예에서, 바람직하다. 일부 구현예에서, 표적 세포는 비분열 세포이다. 일부 구현예에서, 표적 세포는 뉴런 또는 근육 세포이다. 일부 구현예에서, 표적 세포는 생체 내에서 표적화된다. 일부 구현예에서, 세포는 G1 이고 HR은 억제된다. 일부 구현예에서, KEAP1 고갈의 사용, 예를 들어, KEAP1 활성의 발현 억제가 바람직하다. KEAP1 고갈은 예를 들어 Orthwein 등이 확인한 바와 같이, siRNA를 통해 달성될 수 있다. 대안적으로, PALB2-KR 돌연변이체 (BRCA1-상호작용 도메인 내 모든 8개 Lys의 결여)의 발현이 KEAP1 고갈과 조합하여 또는 단독으로 바람직하다. PALB2-KR은 세포 주기 위치와 무관하게, BRCA1과 상호작용한다. 따라서, 특히 G1 세포에서, BRCA1-PALB2 상호작용의 촉진 또는 복원은 일부 경우에, 특히 표적 세포가 비분열하는 경우, 또는 제거 및 복귀 (생체외 유전자 표적화)가 문제가 되는 경우, 예를 들어 뉴런 또는 근육 세포에서 바람직하다. KEAP1 siRNA는 ThermoFischer로부터 입수가능하다.
일부 구현예에서, BRCA1-PALB2 복합체가 G1 세포에 전달될 수 있다. 일부 구현예에서, PALB2 탈유비퀴틴화는 예를 들어 디유비퀴틸라제 USP11의 증가된 발현에 의해 촉진될 수 있어서, 구성체가 디유비퀴틸라제 USP11의 발현 또는 활성을 촉진시키거나 또는 상향조절시키는데 제공될 수 있다는 것을 고려한다.

[1477] **안질환의 치료**

[1478] 본 발명은 또한 하나 또는 두 눈으로 CRISPR-Cas 시스템을 전달하는 것을 고려한다.

[1479] 본 발명의 또 다른 양태에서, CRISPR-Cas 시스템은 문헌 [Genetic Diseases of the Eye, Second Edition,

edited by Elias I. Traboulsi, Oxford University Press, 2012]에 추가 기재된 여러 유전적 돌연변이로부터 발생한 안구 결함을 편집하기 위해 사용될 수 있다.

[1480] 특정 구현예에서, 치료 또는 표적화되는 병태는 눈 장애이다. 일부 구현예에서, 눈 장애는 녹내장을 포함할 수 있다. . 일부 구현예에서, 눈 장애는 망막 변성 질환을 포함한다. 일부 구현예에서, 망막 변성 질환은 스타르가르트병, 바르뎃-비들 증후군, 베스트병, 파란색 원추 단색형 색각, 맥락막결손, 원뿔-막대 세포 이영양증, 선천성 고정형 야맹증, 증대 S-추체 증후군, 소아 X-연관 망막분리증, 레버 선천성 흑내장, 말라티아 레벤티네세 (Malattia Leventinesse), 노리에병 또는 X-연관 가족삼출유리체망막병증, 패턴 이영양증, 소르스비 이영양증, 어셔 증후군, 망막색소변성증, 완전색맹 또는 황반 이영양증 또는 변성, 망막색소변성증, 완전색맹, 및 나이 관련 황반 변성으로부터 선택된다. 일부 구현예에서, 망막 퇴행성 질환은 레버 선천성 흑내장 (LCA) 또는 망막 색소변성증이다. 일부 구현예에서, CRISPR 시스템은 임의로 유리체내 주사 또는 망막하 주사를 통해 눈에 전달된다.

[1481] 눈으로의 투여에서, 렌티바이러스 벡터, 상세하게는 말 전염성 빈혈증 바이러스(EIAV)가 특히 바람직하다.

[1482] 다른 실시형태에서, 말 전염성 빈혈 바이러스 (EIAV/ jgm.845)에 기반한 최소 비영장류 렌티바이러스 벡터가 또한 특히 안구 유전자 요법에 대해 고려된다 (참조: 예를 들어, Balagaan, J Gene Med 2006; 8: 275 -285). 벡터는 표적 유전자의 발현을 구동하는 사이토메갈로 바이러스 (CMV) 프로모터를 갖는 것이 고려된다. 전방내, 망막하, 안구내, 및 유리체내 주사가 모두 고려된다 (참조: 예를 들어, Balagaan, J Gene Med 2006; 8: 275 -285, Published online 21 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845). 안내 주입은 수술 현미경의 도움으로 실행될 수 있다. 망막하 및 유리체내 주입에서, 눈은 온화한 손가락 압력에 의해 탈출될 수 있으며 기반부는 현미경용 유리 슬라이드 커버슬립으로 덮인 각막 상에 커플링 배지 용액을 떨어뜨리는 것으로 구성된 콘텍트 렌즈 시스템을 사용하여 시각화되었다. 망막하 주입에서, 5-㎕ 해밀턴 주사기에 고정된 10-mm 34-게이지 바늘의 끝부분이, 망막하 공간에서 바늘의 구멍이 보일 때까지, 우세한 적도 공막(equatorial sclera)을 통해 접선으로 후두극을 향해 직접 시각화 하에서 진전될 수 있다. 그런 다음, 2 ㎕의 벡터 상청액을 주입하여 우세한 수포성 망막 분리를 생성할 수 있으며, 따라서 이는 망막하 벡터 투여를 확인하는 것이다. 이러한 접근법은 RPE에 의해 흡수될 때까지, 일반적으로 48시간의 과정 내에 벡터 상청액이 망막하 공간에서 보유되도록 하는 자가-밀봉 공막절단을 창출한다. 이 과정은 하위 망막 분리를 생성하기 위해 하위 반구에서 반복될 수 있다. 이 기술은 벡터 현탁액으로 대략 70% 의 감각신경 망막 및 RPE의 노출을 야기한다. 유리체내 주입에서, 바늘 끝은 공막을 통해 각공막 경계 1 mm 뒤로 전진될 수 있으며 2 ㎕의 벡터 현탁액은 유리체강으로 주입된다. 전방내 주입에서, 바늘 끝은 각공막 경계 천자를 통해, 중앙 각막을 향해 진전될 수 있으며, 2 ㎕의 벡터 현탁액이 주입될 수 있다. 전방내 주입에서, 바늘 끝은 각공막 경계 천자를 통해, 중앙 각막을 향해 진전될 수 있으며, 2 ㎕의 벡터 현탁액이 주입될 수 있다. 이들 벡터는 1.01.4×1010 또는 1.01.4×109 형질도입 유닛(TU)/ml의 역가로 주입될 수 있다.

[1483] 다른 구현예에서, 망 형태의 노인성 황반변성의 치료를 위한 망막하 주입을 통해 전달되는 혈관형성 억제 단백질 엔도스타틴 및 안지오스타틴을 발현하는 단백질 말 전염성 빈혈증 바이러스-기반 렌티바이러스 유전자 치료 벡터인 RetinoStat®이 또한 고려된다 (예를 들어, 문헌[Binley et al., HUMAN GENE THERAPY 23:980991 (September 2012)] 참조). 이러한 벡터는 본 발명의 CRISPR-Cas 시스템을 위해 변형될 수 있다. 각각의 눈은 총 100 ㎕의 부피로 눈 당 1.1 x 105 형질도입 유닛(TU/eye)의 용량의 RetinoStat®으로 치료될 수 있다.

[1484] 다른 구현예에서, E1-제거, 부분적 E3-제거, E4-제거 아데노바이러스 벡터가 눈으로의 전달을 위해 고려될 수 있다. 진행된 신생혈관 노인성 황반변성(AMD)을 지니는 28 명의 환자에게 E1-, 부분적 E3-, E4-결실 아데노바이러스 벡터 발현 인간 색소 상피-유래 인자(AdPEDF.11)의 단일 유리체강내 주사를 제공하였다(예를 들어, 문헌[Campochiaro et al., Human Gene Therapy 17:167176 (February 2006)] 참조). 106 내지 109.59.5 입자 유닛(PU) 범위의 용량이 조사되었으며 AdPEDF.11에 관련된 심각한 부작용 및 용량-제한 독성이 없었다(예를 들어, 문헌[Campochiaro et al., Human Gene Therapy 17:167176 (February 2006)]을 참조하라). 아데노바이러스 벡터-매개 안구 유전자 이송이 안구 장애의 치료를 위해 실행가능한 접근법을 나타낼 수 있으며 CRISPR Cas 시스템에 적용될 수 있다.

[1485] 다른 구현예에서, RXi Pharmaceuticals의 sd-rxRNA® 시스템이 눈으로의 CRISPR Cas의 전달을 위해 사용되고 그리고/또는 조정될 수 있다. 이 시스템에서, 3 ㎍의 sd-rxRNA의 단일 유리체내 투여는 14일 동안 PPIB mRNA 수준의 서열-특이적 감소를 야기한다. sd-rxRNA® 시스템이 본 발명의 핵산-표적화 시스템에 적용될 수 있으며, 인간에게 투여된 약 3 내지 20 mg 용량의 CRISPR을 고려하는 것이다.

[1486] Millington-Ward 등 (Molecular Therapy, vol. 19 no. 4, 642649 apr. 2011)은 RNAi 표적 부위 상의 변성 위치에서 뉴클레오티드 변경에 기인한 억제에 저항하는 RNA 간섭 (RNAi)-기반 로돕신 억제제 및 코돈-변형 로돕신 대체 유전자를 전달하기 위해 아데노-연관 바이러스 (AAV) 벡터를 기재한다. 6.0 x 108 vp 또는 1.8 x 1010 vp의 AAV의 주입이 Millington-Ward 등에 의해 눈으로 망막하 주입된다. Millington-Ward 등의 AAV 벡터는 본 발명의 CRISPR Cas 시스템에 적용될 수 있으며, 인간에게 약 2 x 1011내지 약 6 x 1013 vp 용량으로 투여되는 것을 고려한 것이다.

[1487] Dalkara 등(Sci Transl Med 5, 189ra76 (2013))은 또한 눈의 유리체액 내로 손상없는 주사 후에 망막을 통해 결함있는 유전자의 야생형 형태를 전달하는 AAV 벡터를 생성하는 생체내 관련 진화에 관한 것이다. Dalkara는 AAV1, 2, 4, 5, 6, 8, 및 9로부터 cap 유전자의 DNA 셔플링에 의해 작제된 칠량체 펩티드 디스플레이 라이브러리 및 AAV 라이브러리를 기재한다. CAG 또는 Rho 프로모터 하에서 GFP를 발현하는 rcAAV 라이브러리 및 rAAV 벡터는 패키징되었으며 디옥시리보뉴클레아제-저항 게놈 역가는 정량 PCR을 통해 수득된다. 라이브러리가 풀링되었으며, 두 라운드의 진화가 실행되었으며, 각각 초기 라이브러리 다양화 및 그 이후의 3 개의 생체내 스크리닝 단계로 구성된다. 각각의 이러한 단계에서, P30 rho-GFP 마우스에게 약 1×1012 vg/ml의 게놈 역가를 갖는 2 ml의 이오딕사놀-정제된, 인산염-완충 식염수(PBS)-투석 라이브러리로 유리체내로 주입되었다.

Dalkara 등의 AAV 벡터는 본 발명의 핵산-표적화 시스템에 적용될 수 있으며, 약 1 × 1015 내지 약 1 × 1016 vg/ml의 용량으로 인간에게 투여하는 것을 고려하는 것이다.

[1488] 또 다른 구현예에서, 로돕신 유전자는 색소성 망막염 (RP)의 치료를 위해 표적화될 수 있으며, Sangamo BioSciences, Inc.로 양도된 미국 특허 공개 제20120204282호의 시스템은 본 발명의 CRISPR Cas 시스템에 따라 변형될 수 있다. 다른 구현예에서, 인간 로돕신 유전자로부터 표적 서열을 절단하는 방법에 대한, Cellectis로 양도된 미국 특허 공개 제20130183282호의 방법은 또한 본 발명의 핵산-표적화 시스템으로 변형될 수 있다. 다른 구현예에서, 레버 선천성 흑내장 10 (Lca10)을 치료하기 위한 CRISPR-Cas 관련 방법 및 조성물에 관한, Editas Medicine에 양도된 US 공개 번호 20150252358의 방법은 또한 본 발명을 위한 핵산-표적화 시스템에 대해 변형될 수 있다.

[1489] 다른 구현예에서, 어셔 증후군 및 망막색소변성증을 치료하기 위한 CRISPR-Cas 관련 방법 및 조성물에 관한, Editas Medicine에 양도된 US 공개 번호 20170073674의 방법은 또한 본 발명을 위한 핵산-표적화 시스템에 대해 변형될 수 있다.

[1490] 일부 구현예에서, CRISPR 단백질은 C2c1이고, 시스템은 I. CRISPR-Cas 시스템 RNA 폴리뉴클레오티드 서열로서, (a) tracr RNA 폴리뉴클레오티드 및 표적 서열과 하이브리드화할 수 있는 가이드 RNA 폴리뉴클레오티드, 및 (b) 직접 반복부 RNA 폴리뉴클레오티드를 포함하는 것인 폴리뉴클레오티드 서열, 및 II. 임의로 적어도 하나 이상의 핵 국재화 서열을 포함하는, C2c1을 코딩하는 폴리뉴클레오티드 서열을 포함하고, 여기서 직접 반복부 서열은 가이드 서열과 하이브리드화하여 표적 서열로 CRISPR 복합체의 서열-특이적 결합을 유도시키고, CRISPR 복합체는 (1) 표적 서열과 하이브리드화하거나 또는 하이브리드화가능한 가이드 서열, 및 (2) 직접 반복부 서열과 복합체를 형성하는 CRISPR 단백질을 포함하고, CRISPR 단백질을 코딩하는 폴리뉴클레오티드 서열은 DNA 또는 RNA이다.

[1491] 일부 구현예에서, C2c1 이펙터 단백질은 T-풍부 PAM을 인식한다. 특정 구현예에서, PAM은 5'-TTN-3' 또는 5'-ATTN-3' 이다. 일정 구현예에서, MPS I와 관련된 관심 유전자좌는 5' 오버행을 갖는 스태거드 절단부를 생성시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 일부 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, 스태거드 절단은 NHEJ 또는 HDR에 후속한다. 일정 구현예에서, 관심 유전자좌는 주형 DNA 서열을 삽입, 또는 " 녹-인" 시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 특정 구현예에서, DNA 삽입은 적절한 배향으로 게놈에 통합되도록 설계된다. Maresca 등 (Genome Res. 2013 Mar; 23(3): 539-546)은 징크 핑거 뉴클레아제 (ZFN) 및 Tale 뉴클레아제 (TALEN)를 사용해 적용가능한 부위 지정된, 정밀 삽입 방법을 기술하는데, 여기서 5' 오버행을 갖는 짧은, 이중-가닥 DNA가 상보성 말단에 결찰되어, 인간 세포주에서 한정된 유전자좌에서 15-kb 외생성 발현 카세트의 정밀한 삽입을 가능하게 하였다. He 등 (Nucleic Acids Res. 2016 May 19; 44(9))은 4.6 kb 무프로모터 ires-eGFP 단편을 GAPDH 유전자좌에 CRISPR/Cas9-유도 부위-특이적 녹-인으로 체세포 LO2 세포에서 최대 20% GFP+ 세포를 산출하였고, NHEJ 경로에 의해 매개되는 인간 배아 줄기 세포에서 1.70% GFP+ 세포를 산출하였다고 기술하였으며, 또한 NHEJ-기반 녹-인이 조사된 모든 인간 세포 유형에서 HDR-매개 유전자 표적화에 비해 더 효율적이라고 보고하였다. C2c1이 5' 오버행을 갖는 스태거드 절단을 생성시키기 때문에, 당업자는 본 명세서에 개시된 CRISPR-C2c1 시스템을 사용해 관심 유전자좌에서 외생성 DNA 삽입을

발생시키기 위해서 Meresca 등 및 He 등에 기술된 거소가 유사한 방법을 사용할 수 있다.

[1492] 일정 구현예에서, 관심 유전자좌는 PAM 서열의 원위 말단에서 CRISPR-C2c1 시스템에 의해 먼저 변형되고, PAM 서열 근처에서 CRISPR-C2c1 시스템에 의해 더욱 변형되며 HDR을 통해 복구된다. 일정 구현예에서, 관심 유전자좌는 HDR을 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ를 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 바람직한 구현예에서, 외생성 DNA는 3' 및 5' 말단 둘 모두 상에서 단일 가이드 DNA(sgDNA)-PAM 서열이 측접된다. 바람직한 구현예에서, 외생성 DNA는 CRISPR-C2c1 절단 이후에 방출된다.

[1493] Wu(Cell Stem Cell,13:659-62, 2013)는 DNA 절단을 유발한 경우 마우스에서 백내장을 야기하는 단일 염기쌍 돌연변이에 Cas9를 안내하는 가이드 RNA를 설계하였다. 그런 다음, 다른 야생형 대립 형질 또는 올리고를 사용하여 돌연변이체 마우스에서 파손된 대립 형질의 서열을 고치고 백내장-발생 유전 결함을 고치는 접합체 복구 메커니즘이 주어졌다.

[1494] 미국 특허 공개 제20120159653호는 황반변성(MD)과 연관된 세포, 동물 및 단백질을 유전적으로 변형시키기 위해 아연 핑거 뉴클레아제를 사용하는 것을 기재한다. 황반변성(MD)은 노인의 시각 장애의 일차 원인이나, 또한 유아기와 같이 어린 발명 연령을 갖는 스타르가르트 질병, 소르스비 기저부(Sorsby fundus), 및 치명적인 소아 신경병성 질병과 같은 소아 질병의 특정 증상이기도 하다. 황반변성은 망막으로의 손상 때문에 시계 중심의 시야 손실(망막황반)을 야기한다. 현재 존재하는 동물 모델은 인간에서 관찰된 바와 같은 질병의 주요 특징을 나타내지 않는다. MD와 연관된 단백질을 코딩하는 돌연변이 유전자를 포함하는 가능한 동물 모델은 또한 매우 다양한 표현형을 생산하며, 이는 인간 질병 및 치료법 개발로 번역되어 문제가 생기게 한다.

[1495] 미국 특허 공개 제20120159653호의 일 양태는 본 발명의 핵산-표적화 시스템에 적용될 수 있는 MD와 연관된 단백질을 코딩하는 임의의 염색체 서열의 편집에 관한 것이다. MD와 연관된 단백질은 통상적으로 MD 장애에 대한 MD와 연관된 단백질의 실험적 연관성에 근거하여 선택된다. 예를 들어, MD와 연관된 단백질의 생산 비율 또는 순환 농도가 MD 장애가 결여된 집단에 비하여 MD 장애를 갖는 집단에서 상승하거나 저해될 수 있다. 단백질 수준에서의 차이는 웨스턴 블롯, 면역 조직화학적 염색, 효소 결합 면역 흡착 분석 (ELISA), 및 질량 분석을 포함하는 프로테오믹스 기술을 사용하여 평가될 수 있으나, 이에 제한되지 않는다. 대안적으로, MD와 관련된 단백질은, 제한은 아니지만, DNA 마이크로어레이 분석, 연속 유전자 발현 분석 (SAGE), 및 정량적 실시간 중합효소 연쇄 반응 (Q-PCR)을 포함하는 게놈 기술을 사용하여 단백질을 코딩하는 유전자의 유전자 발현 프로파일을 얻음으로써 확인될 수 있다.

[1496] 비-제한적인 예로서, MD와 연관된 단백질은 하기 단백질을 포함하나, 이로 제한되지 않는다:(ABCA4) ATP-결합 카세트, 서브-패밀리 A (ABC1), 멤버 4 ACHM1 색맹 (간상체 적색맹) 1 ApoE 아포지단백질 E (ApoE) C1QTNF5 (CTRP5) C1q 및 종양 괴사 인자 관련 단백질 5 (C1QTNF5) C2 보체 성분 2 (C2) C3 보체 성분 (C3) CCL2 케모카인 (C-C 모티프) 리간드 2 (CCL2) CCR2 케모카인 (C-C 모티프) 수용체 2 (CCR2) CD36 분화 클러스터 36 CFB 보체 인자 B CFH 보체 인자 CFH H CFHR1 보체 인자 H-관련 1 CFHR3 보체 인자 H-관련3 CNGB3 환형 뉴클레오티드 게이트 채널 베타 3 CP 세룰로플라스민 (CP) CRP C 반응성 단백질 (CRP) CST3 시스타틴 C 또는 시스타틴 3 (CST3) CTSD 카텝신 D (CTSD) CX3CR1 케모카인 (C-X3-C 모티프) 수용체 1 ELOVL4 (매우 긴 지방산 4의 신장) ERCC6 (절제 복구 교차-상보) 설치류 복구 결핍, FBLN5 피불린-5 FBLN5 피불린 5 FBLN6 피불린 6 FSCN2 파신 (fascin) (FSCN2) HMCN1 헤미센트린 1 HMCN1 헤미센틴 1 HTRA1 HtrA 세린 펩티다아제 1 (HTRA1) HTRA1 HtrA 세린 펩티다아제 1 IL-6 인터루킨 6 IL-8 인터루킨 8 LOC387715 가설 단백질 PLEKHA1 플레크스트린 상동성 도메인-함유 패밀리 A 멤버 1 (PLEKHA1) PROM1 프로미닌 1(PROM1 또는 CD133) PRPH2 페리페린-2 RPGR 색소성 망막염 GTPase 조절자 세르핀G1 세르핀 펩티다아제 저해제, 클레이드 G, 멤버 1 (C1-저해제) TCOF1 트리클 TIMP3 메탈로프로테이나아제 저해제 3 (TIMP3) TLR3 톨-유사 수용체 3.

[1497] 염색체 서열이 편집된 ASD와 관련된 단백질의 동일성은 다양할 수 있으며 다양할 것이다. 바람직한 구현예에서, 염색체 서열이 편집된 MD와 연관된 단백질은 ATP-결합 카세트, ABCR 유전자에 의해 코딩되는 서브-패밀리 A (ABC1) 멤버 4 단백질 (ABCA4), APOE 유전자에 의해 코딩되는 아포지단백질 E 단백질 (APOE), CCL2 유전자에 의해 코딩되는 케모카인 (C-C 모티프) 리간드 2 단백질 (CCL2), CCR2 케모카인 (C-C 모티프) 수용체 2 단백질 (CCR2), CP 유전자에 의해 코딩되는 세룰로플라스민 단백질 (CP), CTSD 유전자에 의해 코딩되는 카텝신 D 단백질 (CTSD), 또는 TIMP3 유전자에 의해 코딩되는 메탈로프로테이나아제 억제제 3 단백질 (TIMP3)일 수 있다. 예시적인 구현예에서, 유전적으로 변형된 동물은 래트이며, MD 와 연관된 단백질을 인코딩하는 편집된 염색체

서열은 하기 것일 수 있다:(ABCA4) ATP 결합 카세트, NM_000350 서브-패밀리 A (ABC1), 멤버 4 APOE 아포지단백질 E NM_138828 (APOE) CCL2 케모카인 (C-C NM_031530 모티프) 리간드 2 (CCL2) CCR2 케모카인 (C-C NM_021866 모티프) 수용체 2 (CCR2) CP 세룰로플라스민 (CP) NM_012532 CTSD 카텝신 D (CTSD) NM_134334 TIMP3 메탈로프로테이나아제 NM_012886 저해제 3 (TIMP3). 동물 또는 세포는 MD와 연관된 단백질을 인코딩하는 1, 2, 3, 4, 5, 6, 7개 또는 그 이상의 방해된 염색체 서열 및 MD와 연관된 방해된 단백질을 인코딩하는 0, 1, 2, 3, 4, 5, 6, 7개 또는 그 이상의 염색체 통합된 서열을 포함할 수 있다.

[1498] 편집되거나 통합된 염색체 서열은 MD와 관련된 변경된 단백질을 암호화하도록 변형될 수 있다. MD-관련 염색체 서열에서의 여러 돌연변이는 MD와 연관되었다. MD와 연관된 염색체 서열 내 돌연변이의 비제한적인 예는 ABCR 단백질에서, E471K (즉, 위치 471의 글루타메이트가 리신으로 변화), R1129L (즉, 위치 1129의 아르기닌이 류신으로 변화), T1428M (즉, 위치 1428의 트레오닌이 메티오닌으로 변화), R1517S (즉, 위치 1517의 아르기닌이 세린으로 변화), I1562T (즉, 위치 1562의 이소류신이 트레오닌으로 변화), 및 G1578R (즉, 위치 1578의 글리신이 아르기닌으로 변화); CCR2 단백질에서, V64I (즉, 위치 192의 발린이 이소류신으로 변화); CP 단백질에서, G969B (즉, 위치 969의 글리신이 아스파라긴 또는 아스파테이트로 변화); TIMP3 단백질에서, S156C (즉, 위치 156의 세린이 시스테인으로 변화), G166C (즉, 위치 166의 글리신이 시스테인으로 변화), G167C (즉, 위치 167의 글리신이 시스테인으로 변화), Y168C (즉, 위치 168의 티로신이 시스테인으로 변화), S170C (즉, 위치 170의 세린이 시스테인으로 변화), Y172C (즉, 위치 172의 티로신이 시스테인으로 변화) 및 S181C (즉, 위치 181의 세린이 시스테인으로 변화)를 포함하는, MD를 초래할 수 있는 것들을 포함한다. MD-관련 유전자 및 질환에서 유전자 변이체의 다른 관련성도 본 분야에 잘 알려져 있다.

[1499] CRISPR 시스템은 상염색체 열성 유전자로 인한 질환을 교정하는데 유용하다. 예를 들어, CRISPR/Cas9는 눈에서 수용체 상실을 초래하는 상염색체 열성을 제거하는데 사용되었다. Bakondi, B. et al., In Vivo CRISPR/Cas9 Gene Editing Corrects 망막 Dystrophy in the S334ter-3 Rat Model of Autosomal Dominant Retinitis Pigmentosa. Molecular Therapy, 2015; DOI: 10.1038/mt.2015.220.

[1500] 당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 상기 논의된 특허 공개의 것과 유사한 시스템에서 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1501] **순환 및 근육 질병의 치료**

[1502] 본 발명은 본 명세서에 설명된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질 시스템을 심장에 전달하는 것을 또한 고려한다. 심장에서, 선호되는 유전자 이송을 보이는 심근 열대성 아데나-연관 바이러스(AAVM), 특히 AAVM41이 바람직하다(예를 들어, Lin-Yanga et al., PNAS, March 10, 2009, vol. 106, no. 10). 투여는 전신성 또는 국소적일 수 있다. 약 1-10 x 1014 용량의 벡터 게놈이 전신성 투여를 위해 고려된다. 참조: Eulalio et al. (2012) Nature 492: 376 및 Somasuntharam et al. (2013) Biomaterials 34: 7790.

[1503] 예를 들어, 미국 특허 공개 제20110023139호는 세포, 동물 및 심혈관 질병에 연관된 단백질을 유전적으로 변형시키기 위한 아연 핑거 뉴클레아제의 사용을 기재한다. 심혈관 질환은 일반적으로 고혈압, 심장마비, 심부전, 및 뇌졸중 및 TIA를 포함한다. 심혈관 질병에 관련된 임의의 염색체 서열 또는 심혈관 질병에 관련된 임의의 염색체 서열에 의해 코딩된 단백질은 본 기재에서 기재된 방법에서 사용될 수 있다. 심혈관-관련 단백질은 통상적으로 심혈관 질병의 발병에 대한 심혈관-관련 단백질의 실험 연관성에 근거하여 스크리닝된다. 예를 들어, 심혈관-관련 단백질의 생산 비율 또는 순환 농도는 심혈관 장애가 결여된 집단에 비하여 심혈관 장애를 갖는 집단에서 상승되거나 감소될 수 있다. 단백질 수준에서의 차이는 웨스턴 블롯, 면역 조직화학적 염색, 효소 결합 면역 흡착 분석 (ELISA), 및 질량 분석을 포함하는 프로테오믹스 기술을 사용하여 평가될 수 있으나, 이에 제한되지 않는다. 대안적으로, 심혈관-관련 단백질은 DNA 마이크로어레이 분석, 유전자 발현의 연속 분석(SAGE), 및 정량 실시간 폴리머라제 캐스케이드(Q-PCR)을 포함하는 게놈 기술을 사용하여 단백질

을 코딩하는 유전자의 유전자 발현 프로파일을 수득함으로써 식별될 수 있으나, 이에 제한되지 않는다.

[1504] 예로서, 염색체 서열은 제한없이, 다음을 포함할 수 있다: IL1B (인터루킨 1, 베타), XDH (잔틴 디히드로게나제), TP53 (종양 단백질 p53), PTGIS (프로스타글란틴 12 (프로스타사이클린) 신타제), MB (미오글로빈), IL4 (인터루킨 4), ANGPT1 (안지오포이어틴 1), ABCG8 (ATP-결합 카세트, 서브-패밀리 G (WHITE), 멤버 8), CTSK (카텝신 K), PTGIR (프로스타글란틴 12 (프로스타사이클린) 수용체 (IP)), KCNJ11 포타슘 내향성 채널, 서브패밀리 J, 멤버 11), INS (인슐린), CRP (C-반응성 단백질, 펜트락신-관련), PDGFRB (혈소판-유래 성장 인자 수용체, 베타 폴리펩티드), CCNA2 (사이클린 A2), PDGFB (혈소판-유래 성장 인자 베타 폴리펩티드 (원숭이 육종 바이러스 (v-sis) 종양 유전자 상동체)), KCNJ5 포타슘 내향성 채널, 서브패밀리 J, 멤버 5), KCNN3 (포타슘 중간체/소형 컨덕턴스 칼슘-활성화된 채널, 서브패밀리 N, 멤버 3), CAPN10 (칼파인 10), PTGES (프로스타글란틴 E 신타제), ADRA2B (아드레날린성, 알파-2B-, 수용체), ABCG5 (ATP-결합 카세트, 서브-패밀리 G (WHITE), 멤버 5), PRDX2 (퍼옥시리독신 2), CAPN5 (칼파인 5), PARP14 (폴리 (ADP-리보스) 중합효소 패밀리, 멤버 14), MEX3C (mex-3 상동체 C (예쁜 꼬마 선충)), ACE 안지오텐신 I 전환 효소 (펩티딜-디펩티다제 A) 1), TNF (종양 괴사 인자 (TNF 수퍼패밀리, 멤버 2)), IL6 (인터루킨 6 (인터페론, 베타 2)), STN (스타틴), SERPINE1 (세르핀 펩티다제 억제제, 클레이드 E (넥신, 플라스미노겐 활성인자 억제제 1형), 멤버 1), ALB (알부민), ADIPOQ (아디포넥틴, C1Q 및 콜라겐 도메인 함유), APOB (아포지단백질 B (Ag(x) 항원 포함)), APOE (아포지단백질 E), LEP (렙틴), MTHFR (5,10-메틸렌테트라히드로폴레이트 리덕타제 (NADPH)), APOA1 (아포지단백질 A-I), EDN1 (엔도텔린 1), NPPB (나트륨뇨설 펩티드 전구체 B), NOS3 (산화질소 신타제 3 (내피 세포)), PPARG (퍼옥시솜 증식인자-활성화 수용체 감마), PLAT (플라스미노겐 활성인자, 조직), PTGS2 (프로스타글란틴-엔도퍼옥시다제 신타제 2 (프로스타글란틴 G/H 신타제 및 시클로옥시게나제)), CETP (콜레스테릴 에스테르 전달 단백질, 혈장), AGTR1 (안지오텐신 II 수용체, 1형), HMGCR (3-히드록시-3-메틸글루타릴-조효소 A 리덕타제), IGF1 (인슐린-유사 성장 인자 1 (소마토메딘 C)), SELE (셀렉틴 E), REN (레닌), PPARA (퍼옥시솜 증식인자-활성화 수용체 알파), PON1 (파라옥소나제 1), KNG1 (키니노겐 1), CCL2 (케모카인 (C-C 모티프) 리간드 2), LPL (지단백질 리파제), VWF (폰 빌레브란트 인자), F2 (응고 인자 II (트롬빈)), ICAM1 (세포내 부착 분자 1), TGFB1 (형질 전환 성장 인자, 베타 1), NPPA (나트륨뇨설 펩티드 전구체 A), IL10 (인터루킨 10), EPO (에리쓰로포이어틴), SOD1 (수퍼옥시드 디스뮤타제 1, 가용성), VCAM1 (혈관 세포 부착 분자 1), IFNG (인터페론, 감마), LPA (지단백질, Lp(a)), MPO (미엘로퍼옥시다제), ESR1 (에스트로겐 수용체 1), MAPK1 (미토겐-활성화 단백질 키나제 1), HP (하프토글로빈), F3 (응고 인자 III (트롬보플라스틴, 조직 인자)), CST3 (시스타틴 C), COG2 (올리고머 골지 복합체의 성분 2), MMP9 (매트릭스 메탈로펩티다제 9 (젤라티나제 B, 92 kDa 젤라티나제, 92 kDa IV형 콜라게나제)), SERPINC1 (세르핀 펩티다제 억제제, 클레이드 C (항트롬빈), 멤버 1), F8 (응고 인자 VIII, 응혈촉진 성분), HMOX1 (헴 옥시게나제 (디사이클링) 1), APOC3 (아포지단백질 C-III), IL8 (인터루킨 8), PROK1 (프로키네티신 1), CBS (시스타티오닌-베타-신타제), NOS2 (산화질소 신타제 2, 유도성), TLR4 (toll-유사 수용체 4), SELP (셀렉틴 P (과립 막단백질 140 kDa, 항원 CD62)), ABCA1 (ATP-결합 카세트, 서브-패밀리 A (ABC1), 멤버 1), AGT (안지오텐시노겐 (세르핀 펩티다제 억제제, 클레이드 A, 멤버 8)), LDLR (저밀도 지단백질 수용체), GPT (글루탐산-피루베이트 트랜사미나제 (알라닌 아미노트랜스퍼라제)), VEGFA (혈관 내피 성장 인자 A), NR3C2 (핵 수용체 서브패밀리 3, 그룹 C, 멤버 2), IL18 (인터루킨 18 (인터페론-감마-유도 인자)), NOS1 (산화질소 신타제 1 (신경원)), NR3C1 (핵 수용체 서브패밀리 3, 그룹 C, 멤버 1 (글루코코르티코이드 수용체)), FGB (피브리노겐 베타 사슬), HGF (간세포 성장 인자 (헤파포이어틴 A; 산란 인자)), IL1A (인터루킨 1, 알파), RETN (레시스틴), AKT1 (v-akt 쥐과 흉선종 바이러스 종양유전자 상동체 1), LIPC (리파제, 간), HSPD1 (열충격 60 kDa 단백질 1 (샤페로닌)), MAPK14 (미토겐-활성화 단백질 키나제 14), SPP1 (분비형 인단백질 1), ITGB3 (인테그린, 베타 3 (혈소판 당단백질 111a, 항원 CD61)), CAT (카탈라제), UTS2 (우로텐신 2), THBD (트롬보모듈린), F10 (응고 인자 X), CP (세룰로플라스민 (페록시다제)), TNFRSF11B (종양 괴사 인자 수용체 수퍼패밀리, 멤버 11b), EDNRA (엔도텔린 수용체 A형), EGFR (상피 성장 인자 수용체 (적아세포 백혈병 바이러스 (v-erb-b) 종양 유전자 상동체, 조류)), MMP2 (매트릭스 메탈로펩티다제 2 (젤라티나제 A, 72 kDa 젤라티나제, 72 kDa IV형 콜라게나제)), PLG (플라스미노겐), NPY (뉴로펩티드 Y), RHOD (ras 상동체 유전자 패밀리, 멤버 D), MAPK8 (미토겐-활성화 단백질 키나제 8), MYC (v-myc 골수구종증 바이러스 종양유전자 상동체 (조류)), FN1 (피브로넥틴 1), CMA1 (카이마제 1, 비만 세포), PLAU (플라스미노겐 활성인자, 우로키나제), GNB3 (구아닌 뉴클레오티드 결합 단백질 (G 단백질), 베타 폴리펩티드 3), ADRB2 (아드레날린성, 베타-2-, 수용체, 표면), APOA5 (아포지단백질 A-V), SOD2 (수퍼옥시드 디스뮤타제 2, 미토콘드리아), F5 (응고 인자 V (프로악세렐린, 불안정 인자)), VDR (비타민 D (1,25-디히드록시비타민 D3) 수용체), ALOX5 (아라키도네이트 5-리폭시게나제), HLA-DRB1 (주요 조직적합성 복합체, 클래스 II, DR 베타 1), PARP1 (폴리 (ADP-리보스) 중합효소 1), CD40LG

(CD40 리간드), PON2 (파라옥소나제 2), AGER (진행성 글리코실화 최종 산물-특이적 수용체), IRS1 (인슐린 수용체 기질 1), PTGS1 (프로스타글란틴-엔도퍼옥시다제 신타제 1 (프로스타글란틴 G/H 신타제 및 시클로옥시게나제)), ECE1 (엔도텔린 전환 효소 1), F7 (응고 인자 VII (혈청 프로트롬비니 전환 가속인자)), URN (인터루킨 1 수용체 길항제), EPHX2 (에폭시드 히드롤라제 2, 세포질), IGFBP1 (인슐린-유사 성장 인자 결합 단백질 1), MAPK10 (미토겐-활성화 단백질 키나제 10), FAS (Fas (TNF 수용체 수퍼패밀리, 멤버 6)), ABCB1 (ATP-결합 카세트, 서브-패밀리 B (MDR/TAP), 멤버 1), JUN (jun 종양 유전자), IGFBP3 (인슐린-유사 성장 인자 결합 단백질 3), CD14 (CD14 분자), PDE5A (포스포디에스터라제 5A, cGMP-특이적), AGTR2 (안지오텐신 II 수용체, 2형), CD40 (CD40 분자, TNF 수용체 수퍼패밀리 멤버 5), LCAT (레시틴-콜레스테롤 아실트랜스퍼라제), CCR5 (케모카인 (C-C 모티프) 수용체 5), MMP1 (매트릭스 메탈로펩티다제 1 (간질 콜라게나제)), TIMP1 (TIMP 메탈로펩티다제 억제제 1), ADM (아드레노메둘린), DYT10 (근긴장이상 10), STAT3 (신호 전달인자 및 전사 활성인자 3 (급성기 반응 인자)), MMP3 (매트릭스 메탈로펩티다제 3 (스트로멜리신 1, 프로젤라티나제)), ELN (엘라스틴), USF1 (상류 전사 인자 1), CFH (보체 인자 H), HSPA4 (열충격 70 kDa 단백질 4), MMP12 (매트릭스 메탈로펩티다제 12 (마크로파지 엘라스타제)), MME (막 메탈로엔도펩티다제), F2R (응고 인자 II (트롬빈) 수용체), SELL (셀렉틴 L), CTSB (카텝신 B), ANXA5 (아넥신 A5), ADRB1 (아드레날린성, 베타-1-, 수용체), CYBA (시토크롬 b-245, 알파 폴리펩티드), FGA (피브리노겐 알파 사슬), GGT1 (감마-글루타미트랜스퍼라제 1), LIPG (리파제, 상피), HIF1A (저산소증 유도성 인자 1, 알파 서브유닛 (염기성 헬릭스-루프-헬릭스 전사 인자)), CXCR4 (케모카인 (C-X-C 모티프) 수용체 4), PROC (단백질 C (응고 인자 Va 및 VIIIa의 불활성인자)), SCARB1 (스캐빈저 수용체 클래스 B, 멤버 1), CD79A (CD79a 분자, 면역글로불린-연관 알파), PLTP (인지질 전달 단백질), ADD1 (아두신 1 (알파)), FGG (피브리노겐 감마 사슬), SAA1 (혈청 아밀로이드 A1), KCNH2 (포타슘 전압-게이팅 채널, 서브패밀리 H (eag-관련), 멤버 2), DPP4 (디펩티딜-펩티다제 4), G6PD (글루코스-6-포스페이트 디히드로게나제), NPR1 (나트륨뇨설 펩티드 수용체 A/구아닐레이트 시클라제 A (심방나트륨이뇨 펩티드 수용체 A)), VTN (비트로넥틴), KIAA0101 (KIAA0101), FOS (FBJ 쥐과 골육종 바이러스 종양유전자 상동체), TLR2 (toll-유사 수용체 2), PPIG (펩티딜프롤릴 이소머라제 G (사이클로필린 G)), IL1R1 (인터루킨 1 수용체, I형), AR (안드로겐 수용체), CYP1A1 (시토크롬 P450, 패밀리 1, 서브패밀리 A, 폴리펩티드 1), SERPINA1 (세르핀 펩티다제 억제제, 클레이드 A (알파-1 안티프로테이나제, 안티트립신), 멤버 1), MTR (5-메틸테트라히드로폴레이트-호모시스테인 메틸트랜스퍼라제), RBP4 (레티놀 결합 단백질 4, 혈장), APOA4 (아포지단백질 A-IV), CDKN2A (사이클린-의존적 키나제 억제제 2A (골수종, p16, 억제 CDK4)), FGF2 (섬유아세포 성장 인자 2 (염기성)), EDNRB (엔도텔린 수용체 B형), ITGA2 (인테그린, 알파 2 (CD49B, 알파 2 서브유닛 of VLA-2 수용체)), CABIN1 (칼시뉴린 결합 단백질 1), SHBG (성 호르몬-결합 globulin), HMGB1 (고-이동성 그룹 박스 1), HSP90B2P (열충격 단백질 90 kDa 베타 (Grp94), 멤버 2 (가유전자)), CYP3A4 (시토크롬 P450, 패밀리 3, 서브패밀리 A, 폴리펩티드 4), GJA1 (간극연접 단백질, 알파 1, 43 kDa), CAV1 (카베올린 1, 카베올라 단백질, 22 kDa), ESR2 (에스트로겐 수용체 2 (ER 베타)), LTA (림포톡신 알파 (TNF 수퍼패밀리, 멤버 1)), GDF15 (성장 분화 인자 15), BDNF (뇌-유래 신경영양인자), CYP2D6 (시토크롬 P450, 패밀리 2, 서브패밀리 D, 폴리펩티드 6), NGF (신경 성장 인자 (베타 폴리펩티드)), SP1 (Sp1 전사 인자), TGIF1 (TGFB-유도 인자 호메오박스 1), SRC (v-src 육종 (슈미트-루핀 A-2) 바이러스 종양유전자 상동체 (조류)), EGF (상피 성장 인자 (베타-우로가스트론)), PIK3CG (포스포이노시티드-3-키나제, 촉매성, 감마 폴리펩티드), HLA-A (주요 조직적합성 복합체, 클래스 I, A), KCNQ1 (포타슘 전압-게이팅 채널, KQT-유사 서브패밀리, 멤버 1), CNR1 (칸나비노이드 수용체 1 (뇌)), FBN1 (피브릴린 1), CHKA (콜린 키나제 알파), BEST1 (베스트로핀 1), APP (아밀로이드 베타 (A4) 전구체 단백질), CTNNB1 (카테닌 (카데린-연관 단백질), 베타 1, 88 kDa), IL2 (인터루킨 2), CD36 (CD36 분자 (트롬보스폰딘 수용체)), PRKAB1 (단백질 키나제, AMP-활성화, 베타 1 비촉매 서브유닛), TPO (갑상선 퍼옥시다제), ALDH7A1 (알데히드 디히드로게나제 7 패밀리, 멤버 A1), CX3CR1 (케모카인 (C-X3-C 모티프) 수용체 1), TH (티로신 히드롤라제), F9 (응고 인자 IX), GH1 (성장 호르몬 1), TF (트랜스페린), HFE (혈색소증), IL17A (인터루킨 17A), PTEN (포스파타제 및 텐신 상동체), GSTM1 (글루타티온 S-트랜스퍼라제 mu 1), DMD (디스트로핀), GATA4 (GATA 결합 단백질 4), F13A1 (응고 인자 XIII, A1 폴리펩티드), TTR (트랜스타이레틴), FABP4 (지방산 결합 단백질 4, 지방세포), PON3 (파라옥소나제 3), APOC1 (아포지단백질 C-I), INSR (인슐린 수용체), TNFRSF1B (종양 괴사 인자 수용체 수퍼패밀리, 멤버 1B), HTR2A (5-히드록시트립타민 (세로토닌) 수용체 2A), CSF3 (콜로니 자극 인자 3 (과립구)), CYP2C9 (시토크롬 P450, 패밀리 2, 서브패밀리 C, 폴리펩티드 9), TXN (티오레독신), CYP11B2 (시토크롬 P450, 패밀리 11, 서브패밀리 B, 폴리펩티드 2), PTH (부갑상선 호르몬), CSF2 (콜로니 자극 인자 2 (과립구-마크로파지)), KDR (키나제 삽입 도메인 수용체 (III형 수용체 티로신 키나제)), PLA2G2A (포스포리파제 A2, 그룹 IIA (혈소판, 활액)), B2M (베타-2-마이크로글로불린), THBS1 (트롬보스폰딘 1), GCG (글루카곤), RHOA (ras 상동체 유전자 패밀리, 멤버 A), ALDH2 (알데히드 디히드로게나제 2 패밀리 (미토콘드리아)), TCF7L2 (전사 인자 7-

유사 2 (T-세포 특이적, HMG-박스)), BDKRB2 (브라디키닌 수용체 B2), NFE2L2 (핵 인자 (적혈구-유래 2)-유사 2), NOTCH1 (Notch 상동체 1, 전좌연관 (초파리)), UGT1A1 (UDP 글루쿠로노실트랜스퍼라제 1 패밀리, 폴리펩티드 A1), IFNA1 (인터페론, 알파 1), PPARD (퍼옥시솜 증식인자-활성화 수용체 델타), SIRT1 (실투인 (침묵 교잡 유형 정보 조절 2 상동체) 1 (에스. 세레비지아에)), GNRH1 (고나도트로핀-방출 호르몬 1 (황체-방출 호르몬)), PAPPA (임신-연관 혈장 단백질 A, 파파리신 1), ARR3 (어레스틴 3, 망막 (X-어레스틴)), NPPC (나트륨뇨설 펩티드 전구체 C), AHSP (알파 헤모글로빈 안정화 단백질), PTK2 (PTK2 단백질 티로신 키나제 2), IL13 (인터루킨 13), MTOR (라파마이신의 기계론적 표적 (세린/트레오닌 키나제)), ITGB2 (인테그린, 베타 2 (보체 성분 3 수용체 3 및 4 서브유닛)), GSTT1 (글루타티온 S-트랜스퍼라제 쎄타 1), IL6ST (인터루킨 6 신호 전달인자 (gp130, 온코스타틴 M 수용체)), CPB2 (카르복시 펩티다제 B2 (혈장)), CYP1A2 (시토크롬 P450, 패밀리 1, 서브패밀리 A, 폴리펩티드 2), HNF4A (간세포 핵 인자 4, 알파), SLC6A4 (용질 운반체 패밀리 6 (신경전달인자 수송체, 세로토닌), 멤버 4), PLA2G6 (포스포리파제 A2, 그룹 VI (시토졸, 칼슘-독립적)), TNFSF11 (종양 괴사 인자 (리간드) 수퍼패밀리, 멤버 11), SLC8A1 (용질 운반체 패밀리 8 (소듐/칼슘 교환체), 멤버 1), F2RL1 (응고 인자 II (트롬빈) 수용체-유사 1), AKR1A1 (알도-케토 리덕타제 패밀리 1, 멤버 A1 (알데히드 리덕타제)), ALDH9A1 (알데히드 디히드로게나제 9 패밀리, 멤버 A1), BGLAP (뼈 감마-카르복시글루타메이트 (gla) 단백질), MTTP (마이크로솜 트리글리세리드 전달 단백질), MTRR (5-메틸테트라히드로폴레이트-호모시스테인 메틸트랜스퍼라제 리덕타제), SULT1A3 (슬포트랜스퍼라제 패밀리, 시토졸, 1A, 페놀-선호, 멤버 3), RAGE (신장 종양 항원), C4B (보체 성분 4B (키도 혈액군), P2RY12 (푸린성 수용체 P2Y, G-단백질 커플링, 12), RNLS (레날라제, FAD-의존적 아민 옥시다제), CREB1 (cAMP 반응성 엘리먼트 결합 단백질 1), POMC (프로오피오멜라노코르틴), RAC1 (ras-관련 C3 보툴리늄 독소 기질 1 (rho 패밀리, 소형 GTP 결합 단백질 Rac1)), LMNA (라민 NC), CD59 (CD59 분자, 보체 조절 단백질), SCN5A (소듐 채널, 전압-게이팅, V형, 알파 서브유닛), CYP1B1 (시토크롬 P450, 패밀리 1, 서브패밀리 B, 폴리펩티드 1), MIF (마크로파지 이동 억제성 인자 (글리코실화-억제 인자)), MMP13 (매트릭스 메탈로펩티다제 13 (콜라게나제 3)), TIMP2 (TIMP 메탈로펩티다제 억제제 2), CYP19A1 (시토크롬 P450, 패밀리 19, 서브패밀리 A, 폴리펩티드 1), CYP21A2 (시토크롬 P450, 패밀리 21, 서브패밀리 A, 폴리펩티드 2), PTPN22 (단백질 티로신 포스파타제, 비수용체 유형 22 (림프구)), MYH14 (미오신, 중쇄 14, 비근육), MBL2 (만노스-결합 렉틴 (단백질 C) 2, 가용성 (옵소닉 결함)), SELPLG (셀렉틴 P 리간드), AOC3 (아민 옥시다제, 구리 함유 3 (혈관 부착 단백질 1)), CTSL1 (카텝신 L1), PCNA (증식 세포 핵 항원), IGF2 (인슐린-유사 성장 인자 2 (소마토메딘 A)), ITGB1 (인테그린, 베타 1 (피브로넥틴 수용체, 베타 폴리펩티드, 항원 CD29는 MDF2, MSK12 포함)), CAST (칼파스타틴), CXCL12 (케모카인 (C-X-C 모티프) 리간드 12 (기질 세포-유래 인자 1)), IGHE (면역글로불린 중쇄 불변 엡실론), KCNE1 (포타슘 전압-게이팅 채널, Isk-관련 패밀리, 멤버 1), TFRC (트랜스페린 수용체 (p90, CD71)), COL1A1 (콜라겐, I형, 알파 1), COL1A2 (콜라겐, I형, 알파 2), IL2RB (인터루킨 2 수용체, 베타), PLA2G10 (포스포리파제 A2, 그룹 X), ANGPT2 (안지오포이어틴 2), PROCR (단백질 C 수용체, 상피 (EPCR)), NOX4 (NADPH 옥시다제 4), HAMP (헵시디니 항미생물성 펩티드), PTPN11 (단백질 티로신 포스파타제, 비수용체 유형 11), SLC2A1 (용질 운반체 패밀리 2 (촉진성 글루코스 수송체), 멤버 1), IL2RA (인터루킨 2 수용체, 알파), CCL5 (케모카인 (C-C 모티프) 리간드 5), IRF1 (인터페론 조절 인자 1), CFLAR (CASP8 및 FADD-유사 아폽토시스 조절제), CALCA (칼시토닌-관련 폴리펩티드 알파), EIF4E (진핵생물 번역 개시 인자 4E), GSTP1 (글루타티온 S-트랜스퍼라제 pi 1), JAK2 (Janus 키나제 2), CYP3A5 (시토크롬 P450, 패밀리 3, 서브패밀리 A, 폴리펩티드 5), HSPG2 (헤파란 술페이트 프로테오글리칸 2), CCL3 (케모카인 (C-C 모티프) 리간드 3), MYD88 (골수 분화 1차 반응 유전자 (88)), VIP (혈관작용성 장 펩티드), SOAT1 (스테롤 O-아실트랜스퍼라제 1), ADRBK1 (아드레날린성, 베타, 수용체 키나제 1), NR4A2 (핵 수용체 서브패밀리 4, 그룹 A, 멤버 2), MMP8 (매트릭스 메탈로펩티다제 8 (호중구 콜라게나제)), NPR2 (나트륨뇨설 펩티드 수용체 B/구아닐레이트 시클라제 B (심방나트륨이뇨 펩티드 수용체 B)), GCH1 (GTP 시클로히드롤라제 1), EPRS (글루타밀-프롤릴-tRNA 신써타제), PPARGC1A (퍼옥시솜 증식인자-활성화 수용체 감마, 공활성인자 1 알파), F12 (응고 인자 XII (하게만 인자)), PECAM1 (혈소판/내피 세포 부착 분자), CCL4 (케모카인 (C-C 모티프) 리간드 4), SERPINA3 (세르핀 펩티다제 억제제, 클레이드 A (알파-1 안티프로테이나제, 안티트립신), 멤버 3), CASR (칼슘-감지 수용체), GJA5 (간극연접 단백질, 알파 5, 40 kDa), FABP2 (지방산 결합 단백질 2, 장), TTF2 (전사 종결 인자, RNA 중합효소 II), PROS1 (단백질 S (알파)), CTF1 (카디오트로핀 1), SGCB (사르코글리칸, 베타 (43 kDa 디스트로핀-연관 당단백질)), YME1L1 (YME1-유사 1 (에스. 세레비지아에)), CAMP (카텔리시딘 항미생물성 펩티드), ZC3H12A (아연 핑거 CCCH-유형 함유 12A), AKR1B1 (알도-케토 리덕타제 패밀리 1, 멤버 B1 (알도스 리덕타제)), DES (데스민), MMP7 (매트릭스 메탈로펩티다제 7 (마트릴리신, 자궁)), AHR (아릴 탄화수소 수용체), CSF1 (콜로니 자극 인자 1 (마크로파지)), HDAC9 (히스톤 디아세틸라제 9), CTGF (결합 조직 성장 인자), KCNMA1 (포타슘 거대 컨덕턴스 칼슘-활성화 채널, 서브패밀리 M, 알파 멤버 1), UGT1A (UDP 글루쿠로노실트랜스퍼라제 1 패밀리, 폴리펩티드 A

복합 유전자좌), PRKCA (단백질 키나제 C, 알파), COMT (카테콜-.베타.-메틸트랜스퍼라제), S100B (S100 칼슘 결합 단백질 B), EGR1 (초기 성장 반응 1), PRL (프로락틴), IL15 (인터루킨 15), DRD4 (도파민 수용체 D4), CAMK2G (칼슘/칼모듈린-의존적 단백질 키나제 II 감마), SLC22A2 (용질 운반체 패밀리 22 (유기 양이온 수송체), 멤버 2), CCL11 (케모카인 (C-C 모티프) 리간드 11), PGF (B321 태반 성장 인자), THPO (트롬보포이어틴), GP6 (당단백질 VI (혈소판)), TACR1 (타카이키닌 수용체 1), NTS (뉴로텐신), HNF1A (HNF1 호메오박스 A), SST (소마토스타틴), KCND1 (포타슘 전압-게이팅 채널, Shal-관련 서브패밀리, 멤버 1), LOC646627 (포스포리파제 억제제), TBXAS1 (트롬복산 A 신타제 1 (혈소판)), CYP2J2 (시토크롬 P450, 패밀리 2, 서브패밀리 J, 폴리펩티드 2), TBXA2R (트롬복산 A2 수용체), ADH1C (알콜 디히드로게나제 1C (클래스 I), 감마 폴리펩티드), ALOX12 (아라키도네이트 12-리폭시게나제), AHSG (알파-2-HS-당단백질), BHMT (베타인-호모시스테인 메틸트랜스퍼라제), GJA4 (간극 연접 단백질, 알파 4, 37 kDa), SLC25A4 (용질 운반체 패밀리 25 (미토콘드리아 캐리어; 아데닌 뉴클레오티드 운반체), 멤버 4), ACLY (ATP 시트레이트 리아제), ALOX5AP (아라키도네이트 5-리폭시게나제-활성화 단백질), NUMA1 (핵 유사분열 장비 단백질 1), CYP27B1 (시토크롬 P450, 패밀리 27, 서브패밀리 B, 폴리펩티드 1), CYSLTR2 (시스테이닐 류코트리엔 수용체 2), SOD3 (수퍼옥시드 디스뮤타제 3, 세포외), LTC4S (류코트리엔 C4 신타제), UCN (우로코르틴), GHRL (그렐린/오베스타틴 프리프로펩티드), APOC2 (아포지단백질 C-II), CLEC4A (C-type 렉틴 도메인 패밀리 4, 멤버 A), KBTBD10 (kelch 반복 및 BTB (POZ) 도메인 함유 10), TNC (테나신 C), TYMS (티미딜레이트 신써타제), SHC1 (SHC (Src 상동성 2 도메인 함유) 형질전환 단백질 1), LRP1 (저밀도 지단백질 수용체-관련 단백질 1), SOCS3 (사이토카인 신호전달의 억제인자 3), ADH1B (알콜 디히드로게나제 1B (클래스 I), 베타 폴리펩티드), KLK3 (칼리크레인-관련 펩티다제 3), HSD11B1 (히드록시스테로이드 (11-베타) 디히드로게나제 1), VKORC1 (비타민 K 에폭시드 리덕타제 복합체, 서브유닛 1), SERPINB2 (세르핀 펩티다제 억제제, 클레이드 B (오발부민), 멤버 2), TNS1 (텐신 1), RNF19A (링 핑거 단백질 19A), EPOR (에리쓰로포이어틴 수용체), ITGAM (인테그린, 알파 M (보체 성분 3 수용체 3 서브유닛)), PITX2 (쌍형성-유사 호메오도메인 2), MAPK7 (미토겐-활성화 단백질 키나제 7), FCGR3A (IgG의 Fc 단편, 저친화성 111a, 수용체 (CD16a)), LEPR (렙틴 수용체), ENG (엔도글린), GPX1 (글루타티온 퍼옥시다제 1), GOT2 (글루타믹-옥살로아세틱 트랜스아미나제 2, 미토콘드리아 (아스파테이트 아미노트랜스퍼라제 2)), HRH1 (히스타민 수용체 H1), NR112 (핵 수용체 서브패밀리 1, 그룹 I, 멤버 2), CRH (코르티코트로핀 방출 호르몬), HTR1A (5-히드록시트립타민 (세로토닌) 수용체 1A), VDAC1 (전압-의존적 음이온 채널 1), HPSE (헤파라나제), SFTPD (계면활성제 단백질 D), TAP2 (수송체 2, ATP-결합 카세트, 서브-패밀리 B (MDR/TAP)), RNF123 (링 핑거 단백질 123), PTK2B (PTK2B 단백질 티로신 키나제 2 베타), NTRK2 (신경영양성 티로신 키나제, 수용체, 2형), IL6R (인터루킨 6 수용체), ACHE (아세틸콜린스테라제 (Yt 혈액군)), GLP1R (글루카곤-유사 펩티드 1 수용체), GHR (성장 호르몬 수용체), GSR (글루타티온 리덕타제), NQO1 (NAD(P)H 디히드로게나제, 퀴논 1), NR5A1 (핵 수용체 서브패밀리 5, 그룹 A, 멤버 1), GJB2 (간극 연접 단백질, 베타 2, 26 kDa), SLC9A1 (용질 운반체 패밀리 9 (소듐/수소 교환체), 멤버 1), MAOA (모노아민 옥시다제 A), PCSK9 (프로단백질 컨버타제 서브틸리신/켁신 9형), FCGR2A (IgG의 Fc 단편, 저친화성 IIa, 수용체 (CD32)), SERPINF1 (세르핀 펩티다제 억제제, 클레이드 F (알파-2 안티플라스민, 피그먼트 상피 유래 인자), 멤버 1), EDN3 (엔도텔린 3), DHFR (디히드로폴레이트 리덕타제), GAS6 (성장 정지-특이적 6), SMPD1 (스핀고미엘린 포스포디에스터라제 1, 산 리소솜), UCP2 (언커플링 단백질 2 (미토콘드리아, 양자 캐리어)), TFAP2A (전사 인자 AP-2 알파 (활성화 인핸서 결합 단백질 2 알파)), C4BPA (보체 성분 4 결합 단백질, 알파), SERPINF2 (세르핀 펩티다제 억제제, 클레이드 F (알파-2 안티플라스민, 피그먼트 상피 유래 인자), 멤버 2), TYMP (티미딘 포스포릴라제), ALPP (알칼리 포스파타제, 태반 (Regan 이소자임)), CXCR2 (케모카인 (C-X-C 모티프) 수용체 2), SLC39A3 (용질 운반체 패밀리 39 (아연 수송체), 멤버 3), ABCG2 (ATP-결합 카세트, 서브-패밀리 G (WHITE), 멤버 2), ADA (아데노신 디아미나제), JAK3 (Janus 키나제 3), HSPA1A (열충격 70 kDa 단백질 1A), FASN (지방산 신타제), FGF1 (섬유아세포 성장 인자 1 (산성)), F11 (응고 인자 XI), ATP7A (ATPase, Cu++ 수송, 알파 폴리펩티드), CR1 (보체 성분 (3b/4b) 수용체 1 (Knops 혈액군)), GFAP (신경교 섬유질 산성 단백질), ROCK1 (Rho-연관, 코일드-코일 함유 단백질 키나제 1), MECP2 (메틸 CpG 결합 단백질 2 (레트 증후군)), MYLK (미오신 경쇄 키나제), BCHE (부티릴콜린스테라제), LIPE (리파제, 호르몬-감응성), PRDX5 (퍼옥시리독신 5), ADORA1 (아데노신 A1 수용체), WRN (베르너 증후군, RecQ 헬리카제-유사), CXCR3 (케모카인 (C-X-C 모티프) 수용체 3), CD81 (CD81 분자), SMAD7 (SMAD 패밀리 멤버 7), LAMC2 (라미닌, 감마 2), MAP3K5 (미토겐-활성화 단백질 키나제 키나제 키나제 5), CHGA (크로모그라닌 A (부갑상선 분비 단백질 1)), IAPP (섬 아밀로이드 폴리펩티드), RHO (로돕신), ENPP1 (엑토뉴클레오티드 파이로포스파타제/포스포디에스터라제 1), PTHLH (부갑상선 호르몬-유사 호르몬), NRG1 (뉴레굴린 1), VEGFC (혈관 내피 성장 인자 C), ENPEP (글루타밀 아미노펩티다제 (아미노펩티다제 A)), CEBPB (CCAAT/인핸서 결합 단백질 (C/EBP), 베타), NAGLU (N-아세틸글루코사미니다제, 알파-), F2RL3 (응고 인자 II (트롬빈) 수용체-유사 3), CX3CL1 (케모

카인 (C-X3-C 모티프) 리간드 1), BDKRB1 (브라디키닌 수용체 B1), ADAMTS13 (트롬보스폰딘 1형 모티프 존재의 ADAM 메탈로펩티다제, 13), ELANE (엘라스타제, 호중구 발현), ENPP2 (엑토뉴클레오티드 파이로포스파타제/포스포디에스터라제 2), CISH (사이토카인 유도성 SH2-함유 단백질), GAST (가스트린), MYOC (미오실린, 섬유주대 유도성 글루코코르티코이드 반응), ATP1A2 (ATPase, Na+/K+ 수송, 알파 2 폴리펩티드), NF1 (뉴로피브로민 1), GJB1 (간극 연접 단백질, 베타 1, 32 kDa), MEF2A (근세포 인핸서 인자 2A), VCL (빈큘린), BMPR2 (뼈 형태형성성 단백질 수용체, II형 (세린/트레오닌 키나제)), TUBB (튜불린, 베타), CDC42 (세포 분열 주기 42 (GTP 결합 단백질, 25 kDa)), KRT18 (케라틴 18), HSF1 (열충격 전사 인자 1), MYB (v-myb 골수아구증 바이러스 종양 유전자 상동체 (조류)), PRKAA2 (단백질 키나제, AMP-활성화, 알파 2 촉매 서브유닛), ROCK2 (Rho-연관, 코일드-코일 함유 단백질 키나제 2), TFPI (조직 인자 경로 억제제 (지단백질-연관 응고 억제제)), PRKG1 (단백질 키나제, cGMP-의존성, I형), BMP2 (뼈 형태형성성 단백질 2), CTNND1 (카테닌 (카데린-연관 단백질), 델타 1), CTH (시스타티오나제 (시스타티오닌 감마-리아제)), CTSS (카텝신 S), VAV2 (vav 2 구아닌 뉴클레오티드 교환 인자), NPY2R (뉴로펩티드 Y 수용체 Y2), IGFBP2 (인슐린-유사 성장 인자 결합 단백질 2, 36 kDa), CD28 (CD28 분자), GSTA1 (글루타티온 S-트랜스퍼라제 알파 1), PPIA (펩티딜프롤릴 이소머라제 A (사이클로필린 A)), APOH (아포지단백질 H (베타-2-당단백질 I)), S100A8 (S100 칼슘 결합 단백질 A8), IL11 (인터루킨 11), ALOX15 (아라키도네이트 15-리폭시게나제), FBLN1 (피불린 1), NR1H3 (핵 수용체 서브패밀리 1, 그룹 H, 멤버 3), SCD (스테아로일-CoA 데사투라제 (델타-9-데사투라제)), GIP (위 억제성 폴리펩티드), CHGB (크로모그라닌 B (세크레토그라닌 1)), PRKCB (단백질 키나제 C, 베타), SRD5A1 (스테로이드-5-알파-리덕타제, 알파 폴리펩티드 1 (3-옥소-5 알파-스테로이드 델타 4-디히드로게나제 알파 1)), HSD11B2 (히드록시스테로이드 (11-베타) 디히드로게나제 2), CALCRL (칼시토닌 수용체-유사), GALNT2 (UDP-N-아세틸-알파-D-갈락토사민:폴리펩티드 N-아세틸갈락토사미닐트랜스퍼라제 2 (GalNAc-T2)), ANGPTL4 (안지오포이어틴-유사 4), KCNN4 (포타슘 중간체/소형 컨덕턴스 칼슘-활성화된 채널, 서브패밀리 N, 멤버 4), PIK3C2A (포스포이노시티드-3-키나제, 클래스 2, 알파 폴리펩티드), HBEGF (헤파린-결합 EGF-유사 성장 인자), CYP7A1 (시토크롬 P450, 패밀리 7, 서브패밀리 A, 폴리펩티드 1), HLA-DRB5 (주요 조직적합성 복합체, 클래스 II, DR 베타 5), BNIP3 (BCL2/아데노바이러스 E1B 19 kDa 상호작용 단백질 3), GCKR (글루코키나제 (헥소키나제 4) 조절제), S100A12 (S100 칼슘 결합 단백질 A12), PADI4 (펩티딜 아르기닌 디이미나제, IV형), HSPA14 (열충격 70 kDa 단백질 14), CXCR1 (케모카인 (C-X-C 모티프) 수용체 1), H19 (H19, 각인된 모성 발현 전사물 (비-단백질 코딩)), KRTAP19-3 (케라틴 연관 단백질 19-3), IDDM2 (인슐린-의존성 진성 당뇨병 2), RAC2 (ras-관련 C3 보툴리늄 독소 기질 2 (rho 패밀리, 소형 GTP 결합 단백질 Rac2)), RYR1 (리아노딘 수용체 1 (골격)), 시계 (시계 상동체 (마우스)), NGFR (신경 성장 인자 수용체 (TNFR 수퍼패밀리, 멤버 16)), DBH (도파민 베타-히드롤라제 (도파민 베타-모노옥시게나제)), CHRNA4 (콜린성 수용체, 니코틴성, 알파 4), CACNA1C (칼슘 채널, 전압-의존적, L형, 알파 1C 서브유닛), PRKAG2 (단백질 키나제, AMP-활성화, 감마 2 비촉매 서브유닛), CHAT (콜린 아세틸트랜스퍼라제), PTGDS (프로스타글란딘 D2 신타제 21 kDa (뇌)), NR1H2 (핵 수용체 서브패밀리 1, 그룹 H, 멤버 2), TEK (TEK 티로신 키나제, 상피), VEGFB (혈관 내피 성장 인자 B), MEF2C (근세포 인핸서 인자 2C), MAPKAPK2 (미토겐-활성화 단백질 키나제-활성화 단백질 키나제 2), TNFRSF11A (종양 괴사 인자 수용체 수퍼패밀리, 멤버 11a, NFKB 활성인자), HSPA9 (열충격 70 kDa 단백질 9 (몰탈린)), CYSLTR1 (시스테이닐 류코트리엔 수용체 1), MAT1A (메티오닌 아데노실트랜스퍼라제 I, 알파), OPRL1 (오피에이트 수용체-유사 1), IMPA1 (이노시톨(myo)-1(or 4)-모노포스파타제 1), CLCN2 (클로라이드 채널 2), DLD (디히드로리포아미드 디히드로게나제), PSMA6 (프로테오솜 (프로솜, 마크로파인) 서브유닛, 알파형, 6), PSMB8 (프로테오솜 (프로솜, 마크로파인) 서브유닛, 베타 유형, 8 (거대 다기능 펩티다제 7)), CHI3L1 (키티나제 3-유사 1 (연골 당단백질-39)), ALDH1B1 (알데히드 디히드로게나제 1 패밀리, 멤버 B1), PARP2 (폴리 (ADP-리보스) 중합효소 2), STAR (스테로이드성 급성 조절 단백질), LBP (지다당류 결합 단백질), ABCC6 (ATP-결합 카세트, 서브-패밀리 C(CFTR/MRP), 멤버 6), RGS2 (G-단백질 신호전달 조절제 2, 24 kDa), EFNB2 (에프린-B2), GJB6 (간극 연접 단백질, 베타 6, 30 kDa), APOA2 (아포지단백질 A-II), AMPD1 (아데노신 모노포스페이트 디아미나제 1), DYSF (디스펠린, 지대형 근이영양증 2B (상염색체 열성)), FDFT1 (파르네실-디포스페이트 파르네실트랜스퍼라제 1), EDN2 (엔도텔린 2), CCR6 (케모카인 (C-C 모티프) 수용체 6), GJB3 (간극 연접 단백질, 베타 3, 31 kDa), IL1RL1 (인터루킨 1 수용체-유사 1), ENTPD1 (엑토뉴클레오시드 트리포스페이트 디포스포히드롤라제 1), BBS4 (프라뎃-비들 증후군 4), CELSR2 (카데린, EGF LAG 7-경로 G-유형 수용체 2 (프라민고 상동체, 초파리)), F11R (F11 수용체), RAPGEF3 (Rap 구아닌 뉴클레오티드 교환 인자 (GEF) 3), HYAL1 (히아루로노글루코사미니다제 1), ZNF259 (아연 핑거 단백질 259), ATOX1 (ATX1 항산화제 단백질 1 상동체 (효모)), ATF6 (활성화 전사 인자 6), KHK (케토헥소키나제 (프룩토키나제)), SAT1 (스퍼미딘/스퍼민 N1-아세틸트랜스퍼라제 1), GGH (감마-글루타밀 히드롤라제 (콘쥬가제, 폴릴폴리감마글루타밀 히드롤라제)), TIMP4 (TIMP 메탈로펩티다제 억제제 4), SLC4A4 (용질 운반체 패밀리 4, 소듐 바이카보네이트 공수송체,

멤버 4), PDE2A (포스포디에스터라제 2A, cGMP-자극), PDE3B (포스포디에스터라제 3B, cGMP-억제), FADS1 (지방산 데사투라제 1), FADS2 (지방산 데사투라제 2), TMSB4X (티모신 베타 4, X-linked), TXNIP (티오레독신 상호작용 단백질), LIMS1 (LIM 및 노화 세포 항원-유사 도메인 1), RHOB (ras 상동체 유전자 패밀리, 멤버 B), LY96 (림프구 항원 96), FOXO1 (포크헤드 박스 O1), PNPLA2 (파타틴-유사 포스포리파제 도메인 함유 2), TRH (트로트로핀-방출 호르몬), GJC1 (간극 연접 단백질, 감마 1, 45 kDa), SLC17A5 (용질 운반체 패밀리 17 (음이온/당 수송체), 멤버 5), FTO (지방량 및 비만 연관), GJD2 (간극 연접 단백질, 델타 2, 36 kDa), PSRC1 (프롤린/세린-풍부 코일드-코일 1), CASP12 (캐스파제 12 (유전자-가유전자)), GPBAR1 (G 단백질-커플링 담즙산 수용체 1), PXK (PX 도메인 함유 세린/트레오닌 키나제), IL33 (인터루킨 33), TRIB1 (tribbles 상동체 1 (초파리)), PBX4 (프리-B-세포 백혈병 호메오박스 4), NUPR1 (핵 단백질, 전사 조절제, 1), 15-Sep(15 kDa 셀레노단백질), CILP2 (연골 중간층 단백질 2), TERC (텔로머라제 RNA 성분), GGT2 (감마-글루타미트랜스퍼라제 2), MT-CO1 (미토콘드리아 코딩 시토크롬 c 옥시다제 I), 및 UOX (우레이트 옥시다제, 가유전자). 임의의 이들 서열은 예를 들어 돌연변이를 처리하기 위해서, CRISPR-Cas 시스템에 대한 표적일 수 있다.

[1505] 추가 구현예에서, 염색체 서열은 Pon1(파라옥소나제 1), LDLR (LDL 수용체), ApoE(아포지방단백질 E), Apo B-100(아포지방단백질 B-100), ApoA (아포지방단백질(a)), ApoA1(아포지방단백질 A1), CBS(시스타티온 B-신타제), 글리코단백질 IIb/IIb, MTHRF(5,10-메틸렌테트라하이드로폴레이트 환원효소(NADPH), 및 이들의 조합으로부터 추가로 선택될 수 있다. 일 반복에서, 심혈관 질환에 연루된 염색체 서열, 및 염색체 서열에 의해 암호화된 단백질은 CRISPR-Cas 시스템에 대한 표적으로서 Cacna1C, Sod1, Pten, Ppar(알파), Apo E, 렙틴, 및 이들의 조합으로부터 선택될 수 있다.

[1506] 당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 상기 기술된 바와 같은 방법과 유사한 시스템에서 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1507] **간 및 신장 질병의 치료**

[1508] 본 발명은 본 명세서에서 설명된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질 시스템을 간 및/또는 신장으로의 전달을 또한 고려한다. 치료적 핵산의 세포 취득을 유도하기 전달 전략은 바이러스-기반, 지질-기반 또는 복합체-기반 전달과 같은 벡터 시스템 또는 물리력, 또는 나노캐리어를 포함한다. 적은 가능성의 임상 관련성을 갖는 초기 적용으로부터, 유체역학적 고압 주사를 이용하여 핵산이 체계적으로 신세포에 어드레스된 경우, 넓은 범위의 유전자 치료적 바이러스 및 비-바이러스 캐리어가 생체 내에서 상이한 동물 신장 질병 모델에서 전사 후 이벤트를 표적하도록 이미 적용된 바 있다 (Csaba Revesz and Peter Hamar (2011). Delivery Methods to Target RNAs in the Kidney, Gene Therapy Applications, Prof. Chunsheng Kang (Ed.), ISBN: 978-953-307-541-9, InTech, Available from: www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-inthe-kidney). 신장으로의 전달 방법은 아라키돈산 대사의 12/15-리폭시게나아제 (12/15-LO) 경로를 표적화하는 작은 간섭 RNA (siRNA)의 생체 내 전달이 제1형 당뇨병의 스트렙토조토신 주사된 당뇨성 신경병증 (DN) 및 신손상을 개선할 수 있는지의 여부를 연구한, 문헌 ([Yuan et al.(Am J Physiol Renal Physiol 295:F605-F617, 2008])에 포함된 것들을 포함할 수 있다. 신장에서 더욱 큰 생체 내 접근 및 siRNA 발현을 달성하기 위하여, Yuan 등은 콜레스테롤과 컨쥬게이트된 이중가닥 12/15-LO siRNA 올리고뉴클레오티드를 사용하였다. 약 400 ㎍의 siRNA를 마우스 내로 피하 주사하였다. Yuang 등의 방법이, 콜레스테롤과 컨쥬게이트된 CRISPR Cas의 1-2g을 신장으로의 전달을 위해 인간으로의 피하 주사를 고려하여 본 발명의 CRISPR Cas 시스템에 적용될 수 있다.

[1509] Molitoris 등 (문헌 [J Am Soc Nephrol 20:1754-1764, 2009])은 신장 내에서 올리고뉴클레오티드 재흡수 부위로서, 근위요세관 세포 (PTC)를 이용하여 세포사멸 경로에서 중추적 단백질인, p53으로 표적된 신손상을 방지하기 위한 siRNA의 효능을 시험하였다. 허혈 손상 후 4시간에 정맥주사된 p53으로의 네이키드 합성 siRNA는 PTC 및 신장 작용을 최대로 보호하였다. Molitoris 등의 데이터는 siRNA의 근위요세관 세포로의 신속한 전

달이 정맥 투여에 뒤따름을 나타낸다. 용량-반응 분석의 경우, 래트는 동일한 4회 시점에 제공된, siP53, 0.33; 1, 3, 또는 5mg/kg의 용량으로 주사되었고, 각각 1.32; 4, 12, 및 20 mg/kg의 누적 용량을 야기하였다. 시험된 모든 siRNA 용량은 더욱 높은 용량을 이용하여 1일에 SCr 감소 효과를 생산하였으며, PBS-처리된 허혈성 대조구 래트에 비하여 대략 5 일에 걸쳐 유효하였다. 12 및 20 mg/kg 누적 용량은 최고의 보호 효과를 제공하였다. Molitoris 등의 방법은 본 발명의 핵산-표적화 시스템에 적용될 수 있으며, 이는 신장으로의 전달을 위해 인간에게 12 및 20 mg/kg의 누적 용량을 고려한다.

[1510] Thompson 등 (문헌 [Nucleic Acid Therapeutics, Volume 22, Number 4, 2012])은 설치류 및 비인간 영장류에서 정맥 투여에 뒤이은 합성의 작은 간섭 RNA I5NP의 독성학적 및 약물역학적 특성을 보고한다. I5NP는 RNA 간섭 (RNAi) 경로를 통해 작용하도록 설계되어 세포사멸 전구 단백질인 p53의 발현을 일시적으로 억제시키고, 주요 심장 외과 수술 동안 발생할 수 있는 급성 신장 손상과 같은 급성 허혈/재관류 손상으로부터 세포를 보호하도록 개발되고 있으며, 신장 이식 후에 발생할 수 있는 이식편 작용을 지연시킬 수 있다. 설치류에서 800mg/kg I5NP, 및 비인간 영장류에서 1,000 mg/kg I5NP의 용량은 이상 반응을 유발하는데 필요로 되었는데, 원숭이에서 상보체의 무증상 활성화 및 약간 증가된 응고 시간을 포함하는 혈액에 대한 효과를 지시하도록 단리되었다. 래트에서, I5NP의 래트 유사체를 이용시 관찰된 추가의 부작용은 없었으며, 이는 그 효과가 I5NP의 의도된 약물학적 활성에 관련된 독성이라기보다는 합성 RNA 이중나선의 클래스 효과를 나타내기 쉬움을 표시한다. 이들을 합하여, 이들 데이터는 급성 허혈/재관류 손상 후에 신장 작용의 보존을 위해 I5NP의 정맥 내 투여의 임상 시험을 지지한다. 원숭이에서 이상 반응이 관찰되지 않은 수준(no observed adverse effect level:NOAEL)은 500 mg/kg이었다. 심혈관, 호흡 및 신경학적 파라미터에 대한 효과는, 25 mg/kg 이하의 용량 수준으로 정맥 내 투여 후에 원숭이에서 관찰되지 않았다. 따라서, 인간 신장으로의 CRISPR Cas의 정맥 내 투여에 대해 유사한 용량이 고려될 수 있다.

[1511] Shimizu 등 (문헌 [J Am Soc Nephrol 21:622-633, 2010])은 폴리(에틸렌 글리콜)-폴리(L-리신)-기반 비히클을 통해 siRNA의 신사구체로의 전달을 표적화하는 시스템을 개발하였다. siRNA/나노캐리어 복합체는, 유창 내피를 지나 이동하여 메산지움(mesangium)으로의 그의 접근을 가능하게 할 수 있는 크기인, 대략적으로 직경 10 내지 20 nm였다. 형광-표지된 siRNA/나노캐리어 복합체의 복강내 주사 후, Shimizu 등은 연장된 기간 동안 혈액 순환 내 siRNA를 검출하였다. 미토겐-활성화된 단백질 키나아제 1 (MAPK1) siRNA/나노캐리어 복합체의 반복된 복강내 투여는 사구체 신염의 마우스 모델에서 사구체 MAPK1 mRNA 및 단백질 발현을 저해하였다. siRNA 축적의 연구를 위해, PIC 나노캐리어와 복합체화된 Cy5-표지된 siRNA (0.5 ml, 5 nmol의 siRNA 함량), 네이키드 Cy5-표지된 siRNA (0.5 ml, 5 nmol), 또는 HVJ-E 내에 캡슐화된 Cy5-표지된 siRNA (0.5 ml, 5 nmol의 siRNA 함량)를 BALBc 마우스들에게 투여하였다. Shimizu 등의 방법은 본 발명의 핵산-표적화 시스템에 적용될 수 있으며, 이는 인간으로의 복강내 투여 및 신장으로의 전달을 위해 약 1-2 리터 내에서 나노캐리어와 복합체화된 CRISPR Cas의 약 10-20 μmol 도스를 고려한다.

[1512] 당업자는 본 명세서에 개시된 바와 같은 C2c1-CRISPR 시스템으로 Shimizu 등, Thompson 등, 및 Molitoris 등에서 기술된 바와 같은 방법과 함께 본 명세서에 개시된 방법을 사용할 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1513] **신장에 대한 전달 방법은 하기와 같이 요약된다:**

전달 방법	담체	표적 RNA	질환	모델	기능 어세이	저자
유체역학 / 지질	TransIT 생체내 유전자 전달 시스템, DOTAP	p85α	급성 신장 손상	허혈-재관류	흡수, 생체분포	Larson et al., Surgery, (Aug 2007), Vol. 142, No. 2, pp. (262-269)
유체역학 / 지질	리포펙타민 2000	Fas	급성 신장 손상	허혈-재관류	혈액 우레아 질소, Fas 면역조직화학, 아폽토시스, 조직학적 채점	Hamar et al., Proc Natl Acad Sci, (Oct 2004), Vol. 101, No. 41, pp. (14883-14888)
유체역학	n.a.	아폽토시스 캐스케이드 구성요소	급성 신장 손상	허혈-재관류	n.a.	Zheng et al., Am J Pathol, (Oct 2008), Vol. 173, No. 4, pp. (973–980)
유체역학	n.a.	핵 인자 카파-b (NFkB)	급성 신장 손상	허혈-재관류	n.a.	Feng et al., Transplantation, (May 2009), Vol. 87, No. 9, pp. (1283–1289)

[1514]

유체역학 /바이러스	리포펙타민 2000	아폽토시스 길항 전사 인자(AATF)	급성 신장 손상	허혈-재관류	아폽토시스, 산화 스트레스, 캐스파제 활성화, 막 지질 과산화	Xie & Guo, Am Soc Nephrol, (Dec 2006), Vol. 17, No. 12, pp. (3336–3346)
유체역학	pBAsi mU6 Neo/ TransIT-EE 유체역학 전달 시스템	그렘린	당뇨병성 신장병증	스트렙토조토신-유도성 당뇨병	단백뇨, 혈청 크레아티닌, 사구체 및 세뇨관 직경, IV 형 콜라겐/BMP7 발현	Q. Zhang et al., PloS ONE, (Jul 2010), Vol. 5, No. 7, e11709, pp. (1-13)
바이러스/ 지질	pSUPER 벡터/리포펙타민	II 형 TGF-β 수용체	신장 간질 섬유증	일측 요로 폐쇄	α-SMA 발현, 콜라겐 함량,	Kushibikia et al., J Controlled Release, (Jul 2005), Vol. 105, No. 3, pp. (318-331)
바이러스	아데노-연관 바이러스-2	미네랄 코르티코이드 수용체	신손상으로 유발된 고혈압	한랭-유도성 고혈압	혈압, 혈청 알부민, 혈청 우레아 질소, 혈청 크레아티닌, 신장 중량, 소변 나트륨	Wang et al., Gene Therapy, (Jul 2006), Vol. 13, No. 14, pp. (1097-1103)
유체역학 /바이러스	pU6 벡터	루시퍼라제	n.a.	n.a.	흡수	Kobayashi et al., Journal of Pharmacology and Experimental Therapeutics, (Feb 2004), Vol. 308, No. 2, pp. (688-693)

[1515]

지질	지단백질, 알부민	apoB1, apoM	n.a.	n.a.	흡수, 지단백질 및 알부민과의 결합 친화성	Wolfrum et al., Nature Biotechnology, (Sep 2007), Vol. 25, No. 10, pp. (1149-1157)
지질	리포펙타민 2000	p53	급성 신장 손상	허혈성 및 시스플라틴-유도성 급성 손상	조직학적 채점, 아폽토시스	Molitoris et al., J Am Soc Nephrol, (Aug 2009), Vol. 20, No. 8, pp. (1754-1764)
지질	DOTAP/DOPE, DOTAP/DOPE/DOPE-PEG2000	COX-2	유방 선암종	MDA-MB-231 유방암 이종이식 편-보유 마우스	세포 생존능, 흡수	Mikhaylova et al., Cancer Gene Therapy, (Mar 2011), Vol. 16, No. 3, pp. (217-226)
지질	콜레스테롤	12/15-리폭시게나제	당뇨병성 신장병증	스트렙토조토신-유도성 당뇨병	알부민뇨증, 소변 크레아티닌, 조직학, I 형 및 IV 형 콜라겐, TGF-β, 피브로넥틴, 플라스미노겐 활성인자 억제제 1	Yuan et al., Am J Physiol Renal Physiol, (Jun 2008), Vol. 295, pp. (F605-F617)
지질	리포펙타민 2000	미토콘드리아 막 44 (TIM44)	당뇨병성 신장병증	스트렙토조토신-유도성 당뇨병	세포 증식 및 아폽토시스, 조직학, ROS, Mn-SOD 및 글루타티온 퍼옥시다제의 미토콘드리아 이입, 세포막 분극화	Y. Zhang et al., J Am Soc Nephrol, (Apr 2006), Vol. 17, No. 4, pp. (1090–1101)

[1516]

유체역학 / 지질	프로테오리포솜	RLIP76	신장 암종	Caki-2 신장암 이종이식 편-보유 마우스	흡수	Singhal et al., Cancer Res, (May 2009), Vol. 69, No. 10, pp. (4244-4251)
중합체	PEG 화 PEI	루시퍼라제 pGL3	n.a.	n.a.	흡수, 생체분포, 적혈구 응집	Malek et al., Toxicology and Applied Pharmacology, (Apr 2009), Vol. 236, No. 1, pp. (97-108)
중합체	PEG 화 폴리-L-리신	MAPK1	루푸스 사구체신염	사구체신염	단백뇨, 사구체경화증, TGF- β, 피브로넥틴, 플라스미노겐 활성인자 억제제 1	Shimizu et al., J Am Soc Nephrology, (Apr 2010), Vol. 21, No. 4, pp. (622-633)
중합체/나노 입자	히알루론산/퀀텀 도트/PEI	VEGF	신장암/흑색종	B16F1 흑색종 종양-보유 마우스	생체분포, 세포독성, 종양 부피, 엔도시토시스	Jiang et al., Molecular Pharmaceutics, (May-Jun 2009), Vol. 6, No. 3, pp. (727-737)
중합체/나노 입자	PEG 화 폴리카프로락톤 나노섬유	GAPDH	n.a.	n.a.	세포 생존능, 흡수	Cao et al, J Controlled Release, (Jun 2010), Vol. 144, No. 2, pp. (203-212)
압타머	스피에겔머 mNOX-E36	CC 케모카인 리간드 2	사구체경화증	편측신절제 마우스	소변 알부민, 소변 크레아티닌, 조직병리학, 사구체 여과율, 마크로파지 계측수, 혈청 Ccl2, Mac- 2+, Ki-67+	Ninichuk et al., Am J Pathol, (Mar 2008), Vol. 172, No. 3, pp. (628-637)

[1517]

압타머	압타머 NOX-F37	바소프레신 (AVP)	울혈성 심부전	n.a.	D-AVP 와의 결합 친화성, AVP 신호전달의 억제, 소변 삼투압 및 나트륨 농도	Purschke et al., Proc Natl Acad Sci, (Mar 2006), Vol. 103, No. 13, pp. (5173-5178)

[1518]

[1519] **간 또는 간 세포 표적화**

[1520] 간 세포의 표적화를 제공한다. 이것은 시험관내일 수 있거나 또는 생체내일 수 있다. 간세포가 바람직하다. 본 명세서의 CRISPR 단백질, 예컨대 C2c1의 전달은 바이러스 벡터, 특히 AAV (특히 AAV2/6) 벡터를 통할 수 있다. 이들은 정맥내 주사에 의해 투여될 수 있다.

[1521] 시험관이거나 또는 생체내이건 무관하게, 간에 대한 바람직한 표적은 알부민 유전자이다. 이것은 알부민이 매우 높은 수준으로 발현되므로 소위 '세이프 하버 (safe harbor)' 라고 하며 그래서 성공적인 유전자 편집 후 알부민 생산의 감소는 견딜만 하다. 간세포의 오직 소분획이 편집되더라도 알부민 프로모터/인핸서로부터 확인되는 높은 수준의 발현은 유용한 수준의 교정 또는 이식유전자 생산 (삽입된 도너 주형으로부터)이 획득되게 하므로 또한 바람직하다.

[1522] 알부민의 인트론 1은 적합한 표적 부위인 것으로 Wechsler 등 (reported at the 57th Annual Meeting and

Exposition of the American Society of Hematology - abstract available online at ash.confex.com/ash/2015/webprogram/Paper86495.html and presented on 6th December 2015)에 의해 확인되었다. 그들 작업은 이러한 표적 부위에서 DNA를 절단하기 위해 Zn 핑거를 사용하였고, 적합한 가이드 서열은 CRISPR 단백질에 의해 동일 부위에서 가이드 절단으로 생성될 수 있다.

[1523] 알부민같은 고도로 발현된 유전자 (고도로 활성인 인핸서/프로모터를 갖는 유전자) 내 표적의 사용은 또는 Wechsler 등이 보고한 바와 같이, 프로모터 무함유 도너 주형을 사용할 수 있게 하고, 또한 이것은 간 표적화 밖에서 광범위하게 적용가능하다. 고도로 발현된 유전자의 다른 예가 공지되어 있다.

[1524] **간의 다른 질환**

[1525] 특정 구현예에서, 본 발명의 CRISPR 단백질은 간 장애 예컨대 트랜스타이레틴 아밀로이드증 (ATTR), 알파-1-안티트립신 결핍증, 및 다른 간-기반 선천성 대사 오류의 치료에서 사용된다. FAP는 트랜스타이레틴 (TTR)을 코딩하는 유전자의 돌연변이로 인해 초래된다. 상염색체 우성 질환이지만, 모든 캐리어가 질환을 발생시키지 않는다. TTR 유전자 내 100개가 넘는 돌연변이가 질환과 연관된 것으로 알려져 있다. 공통 돌연변이의 예는 V30M을 포함한다. 유전자 침묵화를 기반으로 하는 TTR의 치료 원리는 iRNA를 사용한 연구로 입증되었다 (Ueda et al. 2014 Transl Neurogener. 3:19). 윌슨병 (WD)은 오로지 간세포에서 발견되는 ATP7B를 코딩하는 유전자 내 돌연변이로 인해 초래된다. WD와 연관된 500개가 넘는 돌연변이가 존재하고, 동아시아같은 특정 지역에서 유병률이 높다. 다른 예는 A1ATD (SERPINA1 유전자 내 돌연변이로 인해 초래되는 상염색체 열성 질환) 및 PKU (페닐알라닌 히드록실라제 (PAH) 유전자 내 돌연변이로 인해 초래되는 상염색체 열성 질환)이다.

[1526] 일 양상에서, 본 발명은 간 장애를 치료하는 방법을 제공하고, 세포에, tracrRNA와 복합체를 형성하는 C2c1-CRISPR, 가이드 서열을 포함하는 가이드 RNA 및 직접 반복부를 포함하는 C2c1-CRISPR 시스템을 전달하는 단계를 포함하고, 가이드 서열은 간 질환에 관여하는 유전자의 표적 서열과 하이브리드화하고, C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1527] **간-연관 혈액 장애, 특히 혈우병 및 특히 혈우병 B**

[1528] 간세포의 성공적인 유전자 편집은 마우스 (시험관내 및 생체내 모두) 및 인간외 영장류 (생체내)에서 획득되었는데 간세포에서 유전자 편집/게놈 조작을 통한 혈액 장애의 치료가 실현가능함을 보여준다. 특히, 간세포에서 인간 F9 (hF9)의 발현은 인간에서 혈우병 B을 위한 치료를 의미하는 인간외 영장류에서 확인하였다. 일 양상에서, 본 발명은 간-연관 혈액 장애를 치료하는 방법을 제공하고, 세포에, tracrRNA와 복합체를 형성하는 C2c1-CRISPR, 가이드 서열을 포함하는 가이드 RNA 및 직접 반복부를 포함하는 C2c1-CRISPR 시스템을 전달하는 단계를 포함하고, 가이드 서열은 간 질환에 관여하는 유전자의 표적 서열과 하이브리드화하고, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1529] Wechsler 등은 그들이 생체내 유전자 편집을 통해서 인간회 영장류에서 간세포로부터 인간 F9 (hF9)를 성공적으로 발현시켰다는 것을 [57th Annual Meeting and Exposition of the American Society of Hematology (abstract presented 6th December 2015 and available online at ash.confex.com/ash/2015/webprogram/Paper86495.html)]에서 보고하였다. 이것은 1) 알부민 유전자좌의 인

트론 1을 표적화하는 2개 아연 핑거 뉴클레아제 (ZFN), 및 2) 인간 F9 도너 주형 구성체를 사용해 획득되었다. ZFN 및 도너 주형은 정맥내로 주사되어, 일정 비율의 간 세포에서 알부민 유전자좌로 hF9 유전자의 교정되니 카피의 표적화된 삽입을 일으키는, 별개 간친화성 아데노-연관 바이러스 혈청형 2/6 (AAV2/6) 상에서 코딩되었다.

[1530] 알부민 유전자좌는 이러한 가장 풍부한 혈장 단백질의 생산이 10 g/일을 넘기 때문에 "세이프 하버" 로서 선택되었고, 이들 수준의 중등도 감소는 충분히 견딘다. 게놈 편집된 간세포는 고도의 활성 알부민 인핸서/프로모터에 의해 구동된, 알부민보다는 치료적 분량으로 정상 hFIX (hF9)를 생산하였다. 알부민 유전자좌에서 hF9 이식유전자의 표적화된 통합 및 이 유전자의 알부민 전사물로의 스플라이싱이 확인되었다.

[1531] 마우스 연구: C57BL/6 마우스는 꼬리 정맥 주사를 통해서 1.0 x1013 벡터 게놈 (vg)/kg으로 마우스 대리 시약을 코딩하는 AAV2/6 벡터 (n=25) 또는 비히클 (n=20)이 투여되었다. 처치된 마우스에서 혈장 hFIX의 ELISA 분석은 6-개월 연구 지속기간 동안 지속되는 50-1053 ng/mL의 피크 수준을 보였다. 마우스 혈장으로부터의 FIX 활성의 분석은 발현도와 일관된 생활성을 확인하였다.

[1532] 비인간 영장류 (NHP) 연구: 1.2x1013 vg/kg (n=5/그룹)으로 NHP 표적화된 알부민-특이적 ZFN 및 인간 F9 도너를 코딩하는 AAV2/6 벡터의 단일 정맥내 공-주입은 이러한 대형 동물 모델에서 >50 ng/mL (정상의 >1%)을 야기시켰다. 더 높은 AAV2/6 용량 (최대 1.5x1014 vg/kg)의 사용은 연구 기간 (3개월) 동안, 몇몇 동물에서 최대 1000 ng/ml (또는 정상의 20%), 그리고 단일 동물에서 최대 2000 ng/ml (또는 정상의 50%)로 hFIX 수준을 산출하였다.

[1533] 치료는 마우스 및 NHP에서 충분히 내성이었고 치료 용량에서 양쪽 종에서 AAV2/6+ 도너 치료와 관련하여 유의한 독성학적 발견이 없었다. Sangamo (CA, USA)가 FDA에 신청하였고, 생체내 게놈 편집 적용을 위한 세계 최초의 인간 임상 시험을 수행할 허가를 부여받았다. 이것은 지단백질 리파제 결핍증의 Glybera 유전자 요법 치료의 EMEA의 승인 이후에 후속된다.

[1534] 따라서, 일부 구현예에서, 하기 중 어느 하나 또는 전부가 사용되는 것이 바람직하다: 바람직하게 정맥내 주사로 투여되는, AAV (특히 AAV2/6); 특히 알부민의 인트론 1에서, 이식유전지/주형의 유전자 편집/삽입을 위한 표적으로서 알부민; 인간 F9 도너 주형; 및/또는 무프로모터 도너 주형.

[1535] **혈우병 B**

[1536] 따라서, 일부 구현예에서, 본 발명은 혈우병 B를 치료하는데 사용되는 것이 바람직하다. 이와 같이 F9 (인자 IX)가 적합한 가이드 RNA의 제공을 통해 표적화되는 것이 바람직하다. 효소 및 가이드는 F9가 생산되는 간에서 이상적으로 표적화하지만, 그들인 함께 또는 별개로 전달될 수 있다. 일부 구현예에서, 주형이 제공되고, 이것인 인간 F9 유전자이다. hF9 주형은 hF9의 wt 또는 '교정' 형태를 포함하여서 치료가 효과적이라는 것을 이해할 것이다. 일부 구현예에서, 2-벡터 시스템이 사용될 수 있는데 하나는 C2c1용 벡터이고 다른 한 벡터는 복구 주형(들)용이다. 복구 주형은 둘 이상의 복구 주형, 예를 들어 상이한 포유동물 종으로부터의 2개의 F9 서열을 포함할 수 있다. 일부 구현예에서, 마우스 및 인간 F9 서열 둘 모두가 제공된다. 이것인 마우스에게 전달될 수 있다. 제58회 미국 혈액학 학회 (Annual American Society of Hematology Meeting) (Nov 2016), Yang Yang, John White, McMenamin Deirdre, 및 Peter Bell, PhD는 이것이 역가 및 정확도를 증가시킨다는 보고서를 발표하였다. 제2 벡터는 인자 IX의 인간 서열을 마우스 게놈에 삽입하였다. 일부 구현예에서, 표적화된 삽입은 키메라 초활성 인자 IX 단백질의 발현을 야기시켰다. 일부 구현예에서, 이것은 천연 마우스 인자 IX 프로모터의 제어 하에 있다. 증가 용량으로 신생 및 성체 " 녹아웃" 마우스에게 2-성분 시스템 (벡터 1 및 벡터 2)의 주입은 4개월 동안 정상 (또는 그 이상)의 수준으로 안정한 인자 IX 활성의 발현 및 활성을 야기시켰다. 인간 치료 사례에서, 천연 인간 F9 프로모터가 대신 사용될 수 있다. 일부 구현예에서, wt 표현형이 복원된다.

[1537] 대안적인 구현예에서, F9의 혈우병 B 형태는 모델 유기체, 세포, 또는 세포주 (예를 들어, 쥐과 또는 비인간 영장류 모델 유기체, 세포, 또는 세포주)를 생성시키기 위해 전달될 수 있으며, 모델 유기체, 세포, 또는 세포주는 혈우병 B 표현형, 즉 wt F9 생산 불능을 가지거나 또는 보유한다.

[1538] **혈우병 A**

[1539] 일부 구현예에서, F9 (인자 IX) 유전자는 상기 설명된 F8 (인자 VIII) 유전자에 의해 치환될 수 있어서, 혈우병 A의 치료 (교정 F8 유전자의 제공을 통해) 및/또는 혈우병 A 모델 유기체, 세포 또는 세포주의 창출 (F8 유전자

의 비교정, 혈우병 A 형태의 제공을 통해)을 이끈다.

[1540] **혈우병 C**

[1541] 일부 구현예에서, F9 (인자 IX) 유전자는 상기 설명된 F11 (인자 XI)에 의해 치환될 수 있어서, 혈우병 C의 치료 (교정 F11 유전자의 제공을 통해) 및/또는 혈우병 C 모델 유기체, 세포 또는 세포주의 창출 (F11 유전자의 비교정, 혈우병 C 형태의 제공을 통해)을 이끈다.

[1542] **트랜스타이레틴 아밀로이드증**

[1543] 트랜스타이레틴은 타이록신 호르몬 및 레티놀 (비타민 A)에 결합된 레티놀 결합 단백질을 운반하는 혈청 CSF에 존재하는, 간에서 주로 생산되는 단백질이다. 120개가 넘는 상이한 돌연변이가 유전적 유전자 장애인 트랜스타이레틴 아밀로이드증 (ATTR)을 초래하며, 이 단백질의 돌연변이체 형태는 조직, 특히 말초 신경계에서 응집되어, 다발신경병증을 초래한다. 가족 아밀로이드 다발신경병증 (FAP)가 가장 일반적인 TTR 장애이고, 2014년에 유럽에서 100,000명 당 47명이 발병된 것으로 여겨졌다. Val30Met의 TTR 유전자 내 돌연변이가 가장 일반적인 돌연변이로 여겨지며, FAP 사례의 대략 50%를 초래한다. 지금까지 유일하게 알려진 치유법인, 간 이식 부재 하에서, 이 질환은 일반적으로 진단 후 10년 내에 사망한다. 대부분의 사례가 단성유전자이다.

[1544] ATTR의 마우스 모델에서, TTR 유전자는 CRISPR/Cas9의 전달에 의해 용량 의존적 방식으로 편집될 수 있다. 일부 구현예에서, C2c1은 mRNa로서 제공된다. 일부 구현예에서, C2c1 mRNA 및 가이드 RNA는 LNP에 패키징된다. LNP에 패키징된 C2c1 mRNA 및 가이드 RNA를 포함하는 시스템은 간에서 최대 60% 편집 효율을 달성하였으며, 혈청 TTR 수준은 최대 80% 까지 감소되었다. 그러므로, 일부 구현예에서, 트랜스타이레틴은 특히 Val30Met 돌연변이를 교정하기 위해 표적화된다. 그러므로, 일부 실시형태에서, ATTR이 치료된다.

[1545] **알파-1 항트립신 결핍증**

[1546] 알파-1 항트립신 (A1AT)는 폐에서, 결합 조직을 분해하는 효소인, 호중구 엘라스타제의 활성을 감소시키는 기능을 주로 하는 간에서 생산되는 단백질이다. 알피-1 항트립신 결핍증 (ATTD)는 A1AT를 코딩하는, SERPINA1 유전자의 돌연변이에 의해 초래되는 질환이다. A1AT의 손상된 생산은 폐의 결합 조직의 점진적 분해를 초래해 결과적으로 기종 유사 증상을 일으킨다.

[1547] 몇몇 돌연변이가 ATTD를 초래할 수 있지만, 가장 일반적인 돌연변이는 Glu342Lys (Z 대립유전자라고함, 야생형은 M이라고 함) 또는 Glu264Val (S 대립유전자라고함)이고, 각각의 대립유전자는 동등하게 질환 상태에 기여하며, 2개의 영향받는 대립유전자는 보다 현저한 병태생리를 일으킨다. 이들 결과는 민감한 장기, 예컨대 폐의 결합 조직의 분해를 일으킬뿐만 아니라, 간 내 돌연변이체의 축적은 단백질 독성을 일으킬 수 있다. 현재 치료는 기증 인간 혈장으로부터 회수된 단백질의 주사에 의한 A1AT의 대체에 집중한다. 중증 사례에서 폐 및/또는 간 이식이 고려될 수 있다.

[1548] 이 질환의 일반적인 변이체는 역시 단성 유전자이다. 일부 구현예에서, SERPINA1 유전자가 표적화된다. 일부 구현예에서, Glu342Lys 돌연변이 (Z 대립유전자라고 함, 야생형은 M이라고 함), 또는 Glu264Val 돌연변이 (S 대립유전자라고 함)가 교정된다. 그러므로, 일부 구현예에서, 부실 유전자는 야생형 기능성 유전자로의 대체를 요구한다. 일부 구현예에서, 녹아웃 및 복구 접근법이 필요하여서, 복구 주형이 제공된다. 이중-대립유전자 돌연변이의 경우에, 일부 구현예에서, 오직 하나의 가이드 RNA가 동형접합 돌연변이에 필요하지만, 이형접합 돌연변이 경우에 2개 가이드 RNA가 필요할 수 있다. 일부 구현에예서, 폐 또는 간으로 전달된다.

[1549] **선천성 대사 이상**

[1550] 선천성 대사 이상 (IEM)은 대사 과정에 영향을 미치는 질환의 우산 그룹이다. 일부 구현예서, IEM이 치료된다. 대부분의 이들 질환은 본질적으로 단성유전자 (예를 들어, 페닐케톤뇨증)이고 병태생리는 선천적으로 독성인 물질의 비정상적 축적, 또는 필수 물질의 합성 불능을 일으키는 돌연변이로 인한다. IEM의 성질에 따라서, CRISPR/C2c1은 단독으로, 또는 복구 주형을 통한 부실 유전자의 대체와 조합하여, 녹아웃을 촉진하는데 사용될 수 있다. CRISPR/C2c1 기술로 이득을 얻을 수 있는 예시적인 질환을 일부 구현예에서, 원발성 과옥살산뇨증 1형 (PH1), 아르기노숙신산 리아제 결핍증, 오르니틴 트랜스카르바밀라제 결핍증, 페닐케톤뇨증, 또는 PKU, 및 메이플 시럽 소변 질환이다.

[1551] **상피 및 폐 질병의 치료**

[1552] 본 발명은 또한 본 명세서에 설명된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질 시스템을 하나 또는 둘 모두의 폐에 전달하는 것을 고려한다.

[1553] AAV-2-기반 벡터가 CF 기도로 CFTR 전달을 위해 원래 제안되었지만, AAV-1, AAV-5, AAV-6, 및 AAV-9와 같은 다른 혈청형이 다양한 폐 상피 모델에서 개선된 유전자 이송 효율을 나타낸다(예를 들어, Li et al., Molecular Therapy, vol. 17 no. 12, 2067-2077 Dec 2009). 생체내 AAV-1 형질도입된 쥐과 기관 기도가 AAV-5와 동일한 효율을 갖지만, 내피 AAV-1 실험관내 형질도입 인간 기도 상피 세포에서 AAV-2 및 AAV-5보다 100배까지 더 효율적일 수 있음이 입증되었다. 다른 연구는 실험관내 인간 기도 상피(HAE)로의 유전자 전달에서 AAV-5가 AAV-2보다 50배 더 효율적이고 생체내 마우스 폐 기도 상피에서는 상당히 더 효율적임을 보였다. AAV-6은 또한 실험관내 인간 기도 상피 세포 및 생체내 쥐과 기도에서 AAV-2보다 더 효율적임을 보였다. 더 최근의 분리체인 AAV-9는 9개월에 걸쳐 탐지된 유전자 발현을 갖는 생체내 쥐과 코 및 치조 상피에서 AAV-5보더 더 우수한 유전자 이송 효율을 나타냄을 보였으며, 이는 AAV가 CFTR 유전자 전달 벡터를 위해 요망되는 특성인 생체내 장기간 유전자 발현이 가능하게 할 수 있음을 제안한다. 더욱이, AAV-9는 CFTR 발현의 손실 및 최소 면역 결과를 갖는 쥐과 폐로 재투여될 수 있음을 입증하였다. CF 및 비-CF HAE 배양은 몇 시간 동안 100 μl의 AAV 벡터가 정점 표면 상에 접종될 수 있다 (예를 들어, 문헌[Li et al., Molecular Therapy, vol. 17 no. 12, 20672077 Dec 2009]을 참조하라). MOI는 바이러스 농도 및 실험 목적에 의존하여 1 x 103 내지 4 x 105 벡터 게놈/세포로 변한다. 상기 인용된 벡터는 본 발명의 전달 및/또는 투여를 위해 고려된다.

[1554] Zamora 등은 RNA 간섭 치료제의 인간 감염 질병의 치료 및 또한 호흡기 세포융합 바이러스(RSV)-감염 폐 이식 수령인에서 항바이러스 약물의 무작위 시험으로의 적용의 예를 보고했다(Am J Respir Crit Care Med Vol 183. pp 531538, 2011). Zamora 등은 RSV 호흡관 감염을 갖는 LTX 수령자에서 임의, 이중-맹검, 플라시보조절 시험을 실시했다. 환자는 RSV에 대해 표준 관리를 받았다. 에어로졸화 ALN-RSV01 (0.6 mg/kg) 또는 위약이 3일 동안 매일 투여됐다. 이 연구는 RSV를 표적화하는 RNAi 치료제가 RSV 감염을 갖는 LTX 수령자에 안전하게 투여될 수 있음을 입증한다. ALN-RSV01의 3회 일일 용량은 호흡관 증상 또는 폐 기능의 장애의 임의의 악화를 야기하지 않았으며 사이토카인 또는 CRP의 유도와 같은 임의의 전신성 염증전 효과를 나타내지 않았다.
약동학은 흡인 후 낮은 일시적인 전신성 노출만을 보였으며, 이는 정맥내 또는 흡입으로 투여된 ALN-RSV01이 엑소뉴클레아제 매개 소화 및 신장 배출을 통해 순환으로부터 신속하게 청소됨을 보이는 임상전 동물 데이터와 일치하는 것이다. Zamora 등의 방법은 본 발명의 핵산-표적화 시스템에 적용될 수 있으며 에어로졸화 CRISPR Cas는 예를 들어, 0.6 mg/kg의 용량으로, 본 발명에 대해 고려될 수 있다.

[1555] 폐 질환에 대해 치료된 대상체는, 예를 들어 자발적으로 호흡하면서 폐에 대해 기관지내삽관으로 전달되는 약학적 유효량의 에어로졸화된 AAV 벡터 시스템을 받을 수 있다. 이와 같이, 일반적으로 AAV 전달에 에어로졸화된 전달이 바람직하다. 전달을 위해 아데노바이러스 또는 AAV 입자가 사용될 수 있다. 각각 하나 이상의 조절 서열에 작동적으로 연결된 적합한 유전자 구성체는 전달 벡터에 클로닝될 수 있다. 이러한 에에서, 하기 구성체가 예로서 제공된다: Cas (C2c1)에 대해 Cbh 또는 EF1a 프로모터, 가이드 RNA)에 대해 U6 또는 H1 프로모터,: 바람직한 배열은 CFTR델타508 표적화 가이드, 델타F508 돌연변이에 대한 복구 주형 및 코돈 최적화된 C2c1 효소와, 임의로 하나 이상의 핵 국재화 신호 또는 서열 (들) (NLS(들)), 예를 들어, 2개 (2) NLS를 사용하는 것이다. NLS가 없는 구성체가 또한 고려된다.

[1556] **근육계의 질병 치료**

[1557] 본 발명은 또한 근육(들)에 본 명세서에 기재된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질 시스템을 전달하는 것을 고려한다.

[1558] Bortolanza 등(Molecular Therapy vol. 19 no. 11, 2055-2064 Nov. 2011) 은 안면견갑상완근 이영양증 (FSHD)의 발병 후 FRG1 마우스에서의 RNA 간섭 발현 카세트의 전신 전달이 독성의 징조 없이 용량-의존성 장기간 FRG1 녹다운을 이끌어냄을 보였다. Bortolanza 등은 5 x 1012 vg의 rAAV6-sh1fRG1의 단일 정맥내 주입이 FRG1 마우스의 근육 조직병리학 및 근육 기능을 구제함을 발견했다. 상세하게는, 200 μl 생리학적 용액 중 2 x 1012 또는 5 x 1012 vg의 벡터가 25-게이지 Terumo 주사기를 사용하여 꼬리 정맥으로 주입되었다. Bortolanza 등의 방법은 AAV 발현 CRISPR Cas에 적용될 수 있으며, 약 2 × 1015 또는 2 × 1016 vg의 벡터의 용량으로 인간에게 주사될 수 있다.

[1559] Dumonceaux 등은 미오스타틴 수용체 AcvRIIb mRNA(sh-AcvRIIb)에 대한 RNA 간섭의 기술을 사용하여 미오스타틴 경로를 저해한다(Molecular Therapy vol. 18 no. 5, 881887 May 2010). 유사-디스트로핀의 복구는 벡터화 U7 엑손-스키핑 기술(U7-DYS)에 의해 매개되었다. sh-AcvrIIb 구성체 단독, U7-DYS 구성체 단독, 또는 두 구성

체의 조합을 운반하는 아데노-관련 벡터는 이영양증 mdx 마우스의 전경골근(TA) 근육으로 주입되었다. 주입은 1011 AAV 바이러스 게놈으로 실행되었다. Dumonceaux 등의 방법은 AAV 발현 CRISPR Cas에 적용될 수 있으며, 약 1014 또는 약 1015 vg의 벡터의 용량으로 인간에게 주사될 수 있다.

[1560] Kinouchi 등은 아텔로콜라겐(ATCOL)을 갖는 화학적으로 비변형된 siRNA의 입자 형성을 통해 정상 또는 질병에 걸린 마우스의 골격으로의 생체내 siRNA 전달의 효능을 보고한다(Gene Therapy (2008) 15, 11261130). 마우스 골격근 또는 정맥내로 골격근 성장의 음성 조절자인 미오스타틴을 표적화하는 siRNA의 ATCOL-매개 국소 적용은 적용 몇 주 후 내에 근육량의 증가를 초래했다. 이들 결과는 siRNA의 ATCOL-매개 적용이 근위축증을 포함하는 질병에 대한 미래의 치료적 용도를 위한 강력한 도구임을 시사한다. MstsiRNA(최종 농도, 10 mM)는 제조업자의 지시에 따라 ATCOL(국소 투여를 위한 최종 농도, 0.5%)(AteloGene, Kohken, Tokyo, Japan)와 혼합되었다. 넴부탈(25mg/kg, i.p.)로 마우스(20-주령 수컷 C57BL/6)를 마취한 후, Mst-siRNA/ATCOL 복합체가 교근 및 대퇴이두근 근육에 주사되었다. Kinouchi 등의 방법은 CRISPR Cas에 적용될 수 있으며 인간에게, 예를 들어, 40 μM 용액의 약 500 내지 1000 ml의 용량으로 근육에 주입된다. Hagstrom 등은 포유류의 다리 근육을 통해 근육 세포(근섬유)로 핵산의 유효하고 반복가능한 전달이 가능한 혈관내, 비바이러스 방법을 기재한다(Molecular Therapy Vol. 10, No. 2, August 2004). 과정은 지혈대 또는 혈압 측정용 커프스에 의해 일시적으로 분리된 다리의 말단 정맥으로 네이키드 플라스미드 DNA 또는 siRNA의 주입을 포함한다. 근섬유로의 핵산 전달은 근육 조직으로 핵산 용액이 유출되기에 충분한 부피로 이의 신속한 주입을 가능하게 한다. 골격근에서의 높은 수준의 트랜스유전자 발현은 최소 독성을 갖는 작은 동물 및 큰 동물 모두에서 달성된다. 다리 근육으로의 siRNA 전달 증거 또한 수득되었다. 레서스 원숭이에 플라스미드 DNA 정맥내 주사를 위해서, 3방향 스톱콕이 각각이 단일 시린지에 로딩된, 2 개의 주사기 펌프(Model PHD 2000; Harvard Instruments)에 연결되었다. 파파베린 주사 후 5 분 뒤에 pDNA (40-100ml 식염수 중 15.5 내지 25.7 mg)가 1.7 또는 2.0 ml/s의 속도로 주사되었다. 이는 본 발명의 CRISPR Cas를 발현하는 플라스미드 DNA를 위해 인간에 대해 800 내지 2000 ml 식염수 약 300 내지 500 mg의 주입으로 규모가 커질 수 있다. 래트로의 아데노 바이러스 벡터 주입에서, 2 x 109 전염성 입자는 3 ml의 생리 식염수(NSS)로 주입되었다. 이는 본 발명의 CRISPR Cas를 발현하는 플라스미드 DNA를 위해 인간에 대해 10 리터의 NSS로 주입되는 약 1 x 1013 전염성 입자의 주입으로 규모가 커질 수 있다. siRNA에서, 래트는 12.5 μg의 siRNA로 대복재 정맥에 주입되었으며 영장류는 대복재 정맥에 750 μg의 siRNA가 주입되었다. 이는 본 발명의 CRISPR Cas를 위해 예를 들어, 인간의 대복재 정맥으로 약 15 내지 약 50 mg의 주입으로 규모가 커질 수 있다.

[1561] 또한, 예를 들어, 듀크대의 공개된 출원 WO2013163628 A2호 (돌연변이된 유전자의 유전적 편집(Genetic Correction of Mutated Genes))를 참조하며, 이는 예를 들어 디스트로핀 유전자에서의 돌연변이로 인한 근육 변성을 결과로서 일으키는 열성의 치명적인 X-관련 장애인, 뒤시엔느 근 이영양증 ("DMD")의 원인이 되는 것들과 같은, 뉴클레아제 매개된 비-상동성 말단 접합을 통해 편집될 수 있는 절단된 유전자 생성물 및 때이른 정지 코돈을 유발하는 예를 들어 프레임시프트 돌연변이를 편집하고자 하는 노력을 설명한다. DMD를 유발하는 디스트로핀 돌연변이의 대다수는 리딩 프레임을 방해하고, 디스트로핀 유전자에서 때이른 번역 종료를 유발하는 엑손의 결실이다. 디스트로핀은 근세포 온전성 및 작용을 조절하는데 원인이 되는 세포막의 디스트로글리칸 복합체에 구조 안정성을 제공하는 세포질성 단백질이다. 본 명세서에서 상호가능하게 사용된 바와 같은 디스트로핀 유전자 또는 "DMD 유전자" 는 유전자좌 Xp21에서 2.2 메가염기이다. 일차 전사는 약 14 kb인 성숙 mRNA를 이용하여 약 2,400 kb로 측정된다. 79 엑손은 3500 개가 넘는 아미노산인 단백질을 코딩한다. 엑손 51은 DMD 환자에서 종종 프레임-방해 결실에 인접하며, 올리고뉴클레오티드-기반 엑손 스키핑(skipping)에 대한 임상 시험에서 표적된다. 엑손 51 스키핑 화합물 에테플러센(eteplirsen)에 대한 임상 시험은 최근에 48주에 걸쳐 상당한 기능적 이점을 보고하였고, 기준에 비해 평균 47% 디스트로핀 양성 섬유를 지녔다. 엑손 51에서 돌연변이는 NHEJ-기반 게놈 편집에 의해 영구 편집에 이상적으로 적합화된다.

[1562] Cellectis에게 수여된 미국 특허 제20130145487호의 방법은, 인간 디스트로핀 유전자 (DMD)로부터의 표적 서열을 절단하는 메가뉴클레아제 변이체에 관한 것으로, 이는 또한 본 발명의 핵산-표적화 시스템에 대해 변형될 수 있다. 바람직한 구현예에서, 핵산 표적화 시스템은 CRISPR-C2c1 시스템을 포함한다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1

시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1563] **피부병의 치료**

[1564] 본 발명은 또한 피부로의 본 명세서에 기재된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질 시스템의 전달을 고려한다.

[1565] Hickerson 등(Molecular Therapy-Nucleic Acids (2013) 2, e129)은 인간 및 쥐과 피부에 자체-전달 (sd)-siRNA를 전달하기 위한 동력화된 마이크로니들 어레이 피부 전달 장치에 관한 것이다. siRNA-기반 피부 치료제가 임상으로 바뀌기 위한 주된 도전은 효과적인 전달 시스템의 개발이다. 다양한 피부 전달 기술에서의 상당한 노력이 쏟아졌으나 제한된 성공만을 거두었다. 피부가 siRNA로 치료되는 임상 연구에서, 피하 주사 주입과 연관된 강렬한 통증이 추가 환자의 등록을 불가능하게 하였으며, 이는 개선되며 좀더 "환자-친화적인" (즉, 적은 통증 또는 무통증) 전달 접근법에 대한 요구를 강조하는 것이다. 마이크로니들은 초기 장벽인 각질층을 거쳐 siRNA를 포함하는 대전하 카고를 전달하는 효과적인 방식을 대표하며, 일반적으로 통상적인 피하 주사보다 통증이 덜한 것으로 간주된다. Hickerson 등에 의해 사용된 동력화 마이크로니들 어레이(MMNA) 장치를 포함하는, 동력화 "스탬프 유형" 마이크로니들 장치는 무모 마우스 연구에서 안전함을 보였으며 (i) 미용 산업에서의 광범위한 사용 및 (ii) 모든 지원자들이 독감 예방주사보다 훨씬 적은 통증으로 장치를 사용함을 발견한 제한된 시험에 의해 입증된 바와 같이 적은 통증 또는 무통증을 초래하며, 이는 이 장치를 사용하는 siRNA 전달이 피하 주사 주입을 사용한 이전 임상 시험에 경험했던 것보다 훨씬 더 적은 통증을 야기할 수 있음을 제안하는 것이다. MMNA 장치 (Triple-M 또는 Tri-M으로 Bomtech Electronic Co, Seoul, South Korea가 시판)는 마우스 및 인간 피부에 siRNA의 전달을 위해 조정되었다. sd-siRNA 용액 (최대 300 μl의 0.1 mg/ml RNA) 이 0.1 mm 깊이로 설정된 1회용 Tri-M 바늘 카트리지의 챔버 (Bomtech)에 도입되었다. 인간 피부를 치료하기 위해서, (수술 직후 수득된) 비식별된 피부를 손으로 펼쳐서 처치 전에 코크 플랫폼에 고정시켰다. 모든 피내 주입은 28-게이지 0.5-인치 바늘을 갖는 인슐린 주사기를 사용해 실행되었다. Hickerson 등의 MMNA 장치 및 방법은 본 발명의 CRISPR Cas, 예를 들어, 피부에 300 μl까지의 용량으로 0.1 mg/ml CRISPR Cas를 전달하기 위해 사용되고 그리고/또는 적응될 수 있다.

[1566] Leachman 등의 문헌 (Molecular Therapy, vol. 18 no. 2, 442446 Feb. 2010)은 피부에 제1 짧은 간섭 RNA(siRNA)-기반 치료제를 사용하여 장애성 발바닥 각피증을 포함하는 상염색체 우성 증후군인 흔하지 않은 피부 장애 선천성 경조증(PC)의 치료를 위한 Ib상 임상 시험에 관한 것이다. TD101로 명명된 이 siRNA는 야생형 K6a mRNA 없이 케라틴 6a(K6a) N171K 돌연변이체 mRNA를 특이적으로 그리고 강력하게 표적화한다.

[1567] Zheng 등 (PNAS, July 24, 2012, vol. 109, no. 30, 1197511980)은 구형의 핵산 나노입자 접합체(SNA-NCs)인 고도로 지향적인 공유적으로 고정화된 siRNA의 밀집된 쉘로 싸인 금 코어는 적용 후 몇 시간 내에 실험관내 각질세포, 마우스 피부 및 인간 표피를 거의 100% 자유롭게 투과함을 보였다. Zheng 등은 60 h 동안 25 nM 표피성장인자 수용체 (EGFR) SNA-NC의 단일 적용이 인간 피부에서 효과적인 유전자 녹아웃을 나타낸다는 것을 증명했다. 유사한 용량이 피부로의 투여를 위해 SNA-NC에 고정된 CRISPR Cas에 대해 고려될 수 있다. Zheng 등, Leachman 등, 및 Hickerson 등의 방법은 또한 본 발명의 핵산-표적화 시스템을 위해 변형될 수 있다.

바람직한 구현예에서, 핵산 표적화 시스템은 CRISPR- C2c1 시스템을 포함한다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1568] **암**

[1569] 일부 구현예에서, 암의 치료, 예방 또는 진단이 제공된다. 표적은 FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC 또는 TRBC 유전자 중 하나 이상이다. 암은 림프종, 만성 림프구성 백혈병 (CLL), B 세포 급성 림프구성 백혈병 (B-ALL), 급성 림프아구성 백혈병, 급성 골수성 백혈병, 비호지킨 림프구 (NHL), 미만성 거대 세포 림프종 (DLCL), 다발성 골수종, 신장 세포 암종 (RCC), 신경아세포종, 직결장암, 거세 저항성 전립선 암, 전이

성 신장 세포 암종, 전이성 비소세포 폐암, 유방암, 방광암, 난소암, 흑색종, 육종, 전립선암, 폐암, 식도암, 간세포 암종, 췌장암, 성상세포종, 중피종, 두경부암, 및 수모세포종 중 하나 이상일 수 있다. 이것은 조작된 키메라 항원 수용체 (CAR) T 세포로 구현될 수 있다. 이것은 WO2015161276에 기술되어 있고, 이의 개시는 참조로 본 명세서에 편입되고, 하기 본 명세서에 기술된다.

[1570] PD-1 녹아웃 T 세포를 생성시켜 CRISPR-Cas9 시스템을 사용한 식도암, 침윤성 방광암, 호르몬 불응성 전립선암, 전이성 신장 세포 암종, 전이성 비소세포 폐암, IV기 위 암종, IV 기 비인두 암종, IV 기 t-세포 암종, 및 엡스테인-바 바이러스 연관 악성종을 포함하느 다수 암의 치료가 제안되고 기술되었다. 참조: Niu et al. Cell 2014, 156(4): 836-43; Rosenberg et al, Science 2015, 348 (6230): 62-8; Sharma et al, Cell 2015, 161(2): 205-14; Bidnur et al, Bladder Cancer, 2016, 2(1): 15-25; Kim et al, Investig Clin Urol. 2016, 57 Suppl 1: S98-S105; Argon-Ching et al, Future Oncol., 2016, 12(17): 2049-58; Festino et al, Drugs 2016, 76(9): 925-45; Zibelman et al, Future Oncl., 2016, 12(19): 2227-42; Doni et al., J. Urol., 2017 197(1): 14-22; Yi et al, Biochim Biophys Acta., 2016, 1866(2): 197-207; Taube et al, Oncoimmunology, 2014, 3(11)L e963413; Yatsuda et al, Nihon Rinsho, 2014, 72(12): 2174-8; Modena et al. Oncol Rev. 2016, 10(1): 293; Bishop et al. Oncotarget, 2015, 6(1): 234-42; Gandini et al, Crit Rev Oncol Hematol. 2016, 100:88-98; Koshikin et al, Expert Opin Pharmacother. 2016 17(9):1225-32; Hofmann et al., Eur J Cancer. 2016, 60:190-209; Gunturi et al, Curr Treat Options Oncol. 2014, 15(1):137-46; Bockorny et al., Expert Opin Biol Ther. 2013, 13(6):911-25; Garon et al, N Engl J Med 2015, 372(21)L 2018-28; Brahmer et al, N Eng J Med, 2015 373(2): 123-35; Borghaei et al, N Engl JH Med 2015, 373(17): 1627-39; Kim et al, Gastroenterology, 2015 148(1): 137-147; Quan et al, PloS One, 2015, 10(9): 30136476; Louis et al, J Immunother., 2010, 33(9): 983-90; Lloyd et al., Frot Immunol., 2013, 4:221; Su et al, Sci Rep. 2016, 6: 20070. 말초 혈액 림프구를 채취하였고 프로그램된 세포 사멸 단백질 1 (PDCD1) 유전자를 실험실에서 CRISPR Cas9를 통해 녹아웃시킬 것이다 (PD-1 녹아웃 T-세포). 림프구를 선택하고 생체외에서 확장시키고 환자에게 다시 주입시킨다. 총 2 x 10^7/kg의 PD-1 녹아웃 T 세포가 1회 사이클로 주입된다. 각 사이클은 3회 투여로 나뉘며, 제1 투여는 20% 주입되고, 제2 투여에서 30%, 나머지 50%는 제3 투여이다. 후기 식도암 및 침윤성 방광암의 경우, 20 mg/kg의 사이클로포스파미드의 단일 용량이 세포 주입 전 3일에 i.v로 투여된다. 5일 후에, 720000 국제 단위 (IU)/kg/일 (내성이면)의 인터루킨-2 (Il-2)가 투여된다. 환자는 총 2, 3, 4, 사이클의 치료를 받는다.

[1571] 암의 치료 또는 예방에 적합한 표적 유전자는 일부 구현에에서, 그 개시가 참조로 본 명세서에 편입되는 Wo2015048577에 기술된 것들을 포함할 수 있다. WO2015161276 및 WO2015048577의 방법이 또한 본 발명의 핵산-표적화 시스템이 변형시키는데 사용될 수 있다.

[1572] 본 명세서에 개시된 CRISPR-C2c1 시스템은 암 치료에서 상기 기술된 방법과 적용될 수 있다. 바람직한 구현예에서, 핵산 표적화 시스템은 CRISPR- C2c1 시스템을 포함한다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다. C2c1은 Cas9에 의해 생성되는 PAM의 근위 말단에서의 절단과 대조적으로, PAM의 원위 말단에서 이중 가닥 파손을 생성시킨다 (Jinek et al., 2012; Cong et al., 2013). C2c1 돌연변이된 표적 서열은 단일 gRNA에 의한 반복된 절단에 감수성일 수 있어서, HDR 매개 게놈 편집에서 C2c1의 적용을 촉진한다고 제안된다 (Front Plant Sci. 2016 Nov 14;7:1683). 일정 구현에에서, 관심 유전자좌는 상동성 지정 복구 (HR 또는 HDR)를 통해 CRISPR-C2c1 복합체에 의해 변형된다. 일정 구현예에서, 관심 유전자좌는 HR 독립적인 CRISPR-C2c1 복합체에 의해 변형된다. 일정 구현에에서, 관심 유전자좌는 비상동성 말단 결합 (NHEJ)을 통해 CRISPR-C2c1 복합체에 의해 변형된다.

[1573] C2c1는 Cas9에 의해 생성되는 블런트 말단부와 대조적으로, 5' 오버행을 갖는 스태거드 절단부를 생성시킨다 (Garneau et al., Nature. 2010;468:67-71; Gasiunas et al., Proc Natl Acad Sci U S A. 2012;109:E2579-2586). 절단 산물의 이러한 구조는 포유동물 게놈으로 비상동성 말단 결합 (NHEJ)-기반 유전자 삽입을 촉진하기 위해 특이 유리할 수 있다 (Maresca et al., Genome research. 2013;23:539-546). 일부 구현예에서,

CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 외생성 DNA 서열을 도입시킨다. 일정 구현예에서, 관심 유전자좌는 주형 DNA 서열을 삽입, 또는 "녹-인" 시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 특정 구현예에서, DNA 삽입은 적절한 배향으로 게놈에 통합되도록 설계된다. 바람직한 구현예에서, 관심 유전자좌는 비분열 세포에서 CRISPR-C2c1 시스템에 의해 변형되고, 여기서 상동성 지정 복구 (HDR) 기전을 통한 게놈 편집이 특히 도전적이다 (Chan et al., Nucleic acids research. 2011;39:5955-5966). Maresca 등 (Genome Res. 2013 Mar; 23(3): 539-546)은 징크 핑거 뉴클레아제 (ZFN) 및 Tale 뉴클레아제 (TALEN)를 사용해 적용가능한 부위 지정된, 정밀 삽입 방법을 기술하는데, 여기서 5' 오버행을 갖는 짧은, 이중-가닥 DNA가 상보성 말단에 결찰되어, 인간 세포주에서 한정된 유전자좌에서 15-kb 외생성 발현 카세트의 정밀한 삽입을 가능하게 하였다. He 등 (Nucleic Acids Res. 2016 May 19; 44(9))은 4.6 kb 무프로모터 ires-eGFP 단편을 GAPDH 유전자좌에 CRISPR/Cas9-유도 부위-특이적 녹-인으로 체세포 LO2 세포에서 최대 20% GFP+ 세포를 산출하였고, NHEJ 경로에 의해 매개되는 인간 배아 줄기 세포에서 1.70% GFP+ 세포를 산출하였다고 기술하였으며, 또한 NHEJ-기반 녹-인이 조사된 모든 인간 세포 유형에서 HDR-매개 유전자 표적화에 비해 더 효율적이라고 보고하였다. C2c1이 5' 오버행을 갖는 스태거드 절단을 생성시키기 때문에, 당업자는 본 명세서에 개시된 CRISPR-C2c1 시스템을 사용해 관심 유전자좌에서 외생성 DNA 삽입을 발생시키기 위해서 Meresca 등 및 He 등에 기술된 거소가 유사한 방법을 사용할 수 있다.

[1574] 일정 구현예에서, 관심 유전자좌는 PAM 서열의 원위 말단에서 CRISPR-C2c1 시스템에 의해 먼저 변형되고, PAM 서열 근처에서 CRISPR-C2c1 시스템에 의해 더욱 변형되며 HDR을 통해 복구된다. 일정 구현예에서, 관심 유전자좌는 HDR을 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ를 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 바람직한 구현예에서, 외생성 DNA는 3' 및 5' 말단 둘 모두 상에서 단일 가이드 DNA(sgDNA)-PAM 서열이 측접된다. 바람직한 구현예에서, 외생성 DNA는 CRISPR-C2c1 절단 이후에 방출된다. 참조: Zhang et al., Genome Biology201718:35; He et al., Nucleic Acids Research, 44: 9, 2016.

[1575] **어셔 증후군 및 망막 색소변성증-39**

[1576] 일부 구현예에서, 어셔 증후군 또는 망막 색소변성증-39의 치료, 예방 또는 진단이 제공된다. 표적은 바람직하게 USH2A 유전자이다. 일부 구현예에서, 위치 2299에서 G 결실 (2299delG)의 교정이 제공된다. 이것은 WO2015134812A1에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1577] **자가면역 및 염증성 장애**

[1578] 일부 구현예에서, 자가면역 및 염증성 장애가 치료된다. 이들은 예를 들어 다발성 경화증 (MS) 또는 류마티스성 관절염 (RA)을 포함한다.

[1579] **낭성 섬유증 (CF)**

[1580] 일부 구현예에서, 낭성 섬유증의 치료, 예방 또는 진단이 제공된다. 표적은 바람직하게 SCNN1A 또는 CFTR 유전자이다. 이것은 WO2015157070에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다.

[1581] Schwank 등 (문헌 [Cell Stem Cell, 13:65358, 2013])은 인간 줄기 세포에서 낭포성 섬유증과 연관된 결함을 편집하기 위해 CRISPR-Cas9를 사용하였다. 이 팀의 표적은 이온 채널을 위한 유전자인, 낭포성 섬유증 막투과 전도체 수용체(CFTR). CFTR에서의 결실은 낭포성 섬유증 환자에서 단백질의 미스폴딩을 발생시킨다. 낭포성 섬유증을 갖는 2명의 아동으로부터의 세포 시료로부터 발생된 배양된 장 줄기 세포를 사용하여, Schwank 등은 삽입될 복구 서열을 함유하는 공여 플라스미드와 함께 CRISPR을 사용하여 편집할 수 있었다. 그런 다음, 연구자들은 세포를 장 "오르가노이드," 또는 소형 소화관으로 성장시켰으며, 이들이 정상적으로 작용함을

보였다. 이러한 경우에서, 클론성 오르가노이드의 대략 절반이 적절한 유전 편집을 겪었다.

[1582] 일부 구현예에서, 예를 들어, 낭성 섬유증이 치료된다. 그러므로 폐로의 전달이 바람직하다. F508 돌연변이 (델타-F508, 명칭 CFTRΔF508 또는 F508del-CFTR) 가 바람직하게 교정된다. 일부 구현예에서, 표적은 ABCC7, CF 또는 MRP7일 수 있다.

[1583] 다른 구현에에서, 낭성 섬유증의 치료를 위한 Crispr-Cas 관련 방법 및 조성물에 관한, Editas medicine에게 양도된 특허 공개 US20170022507의 방법은 본 발명에 개시된 CRISPR-Cas 시스템에 대해 변형될 수 있다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1584] **뒤센 근이영양증**

[1585] 뒤센 근이영양증 (DMD)은 출생시 남성 5000명 중 대략 1명에서 발병되는 열성, 성-연관 근 소모 질환이다. 디스트로핀 유전자의 돌연변이가 골격근에서 디스트로핀의 부재를 초래하는데, 이것은 정상적으로는 근육 섬유의 세포골격을 기저층에 연결시키는 기능을 한다. 이들 돌연변이로인한 디스트로핀의 부재는 체세포로 과도한 칼슘 유입을 초래하여 미토콘드리아를 파열시켜, 세포를 파괴하게 된다. 현재 치료는 DMD의 증상을 완화시키는데 집중하며 평균 기대 수명은 대략 26세이다.

[1586] 일정 유형의 DMD에 대한 치료로서 CRISPR/Cas9 효능은 마우스 모델에서 입증되었다. 이러한 한 연구에서, 근이영양증 표현형은 돌연변이체 엑손을 녹아웃시켜 마우스에서 부분적으로 교정되어 기능성 단백질이 생성되었다 (참조: Nelson et al. (2016) Science, Long et al. (2016) Science, and Tabebordbar et al. (2016) Science).

[1587] 일부 구현예에서, DMD를 치료하는 Crispr 관련 방법과 관련된, Editas Medine에 양도된 특허 공개 WO2016161380의 방법이 본 발명의 CRISPR-Cas 시스템의 적용을 위해 변형될 수 있다. 일부 실시형태에서, DMD가 치료된다. 일부 구현예에서, 주사에 의해 근육으로 전달된다. 일부 구현예에서, CRISPR 단백질은 C2c1이고, 시스템은 I. CRISPR-Cas 시스템 RNA 폴리뉴클레오티드 서열로서, (a) tracr RNA 폴리뉴클레오티드 및 표적 서열과 하이브리드화할 수 있는 가이드 RNA 폴리뉴클레오티드, 및 (b) 직접 반복부 RNA 폴리뉴클레오티드를 포함하는 것인 폴리뉴클레오티드 서열, 및 II. 임의로 적어도 하나 이상의 핵 국재화 서열을 포함하는, C2c1을 코딩하는 폴리뉴클레오티드 서열을 포함하고, 여기서 직접 반복부 서열은 가이드 서열과 하이브리드화하여 표적 서열로 CRISPR 복합체의 서열-특이적 결합을 유도시키고, CRISPR 복합체는 (1) 표적 서열과 하이브리드화하거나 또는 하이브리드화가능한 가이드 서열, 및 (2) 직접 반복부 서열과 복합체를 형성하는 CRISPR 단백질을 포함하고, CRISPR 단백질을 코딩하는 폴리뉴클레오티드 서열은 DNA 또는 RNA이다.

[1588] 일부 구현예에서, CRISPR- C2c1 단백질은 T-풍부 PAM을 인식한다. 특정 구현예에서, PAM은 5'-TTN-3' 또는 5'-ATTN-3' 이다. 일정 구현예에서, 관심 유전자좌는 주형 DNA 서열을 삽입, 또는 " 녹-인" 시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 특정 구현예에서, DNA 삽입은 적절한 배향으로 게놈에 통합되도록 설계된다.

Maresca 등 (Genome Res. 2013 Mar; 23(3): 539-546)은 징크 핑거 뉴클레아제 (ZFN) 및 Tale 뉴클레아제 (TALEN)를 사용해 적용가능한 부위 지정된, 정밀 삽입 방법을 기술하는데, 여기서 5' 오버행을 갖는 짧은, 이중-가닥 DNA가 상보성 말단에 결찰되어, 인간 세포주에서 한정된 유전자좌에서 15-kb 외생성 발현 카세트의 정밀한 삽입을 가능하게 하였다. He 등 (Nucleic Acids Res. 2016 May 19; 44(9))은 4.6 kb 무프로모터 ires-eGFP 단편을 GAPDH 유전자좌에 CRISPR/Cas9-유도 부위-특이적 녹-인으로 체세포 LO2 세포에서 최대 20% GFP+ 세포를 산출하였고, NHEJ 경로에 의해 매개되는 인간 배아 줄기 세포에서 1.70% GFP+ 세포를 산출하였다고 기술하였으며, 또한 NHEJ-기반 녹-인이 조사된 모든 인간 세포 유형에서 HDR-매개 유전자 표적화에 비해 더 효율적이라고 보고하였다. C2c1이 5' 오버행을 갖는 스태거드 절단을 생성시키기 때문에, 당업자는 본 명세서에 개시된 CRISPR-C2c1 시스템을 사용해 관심 유전자좌에서 외생성 DNA 삽입을 발생시키기 위해서 Meresca 등 및 He 등에 기술된 것과 유사한 방법을 사용할 수 있다.

[1589] 일정 구현예에서, 관심 유전자좌는 PAM 서열의 원위 말단에서 CRISPR-C2c1 시스템에 의해 먼저 변형되고, PAM 서열 근처에서 CRISPR-C2c1 시스템에 의해 더욱 변형되며 HDR을 통해 복구된다. 일정 구현예에서, 관심 유전자좌는 HDR을 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ를 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 바람직한 구현예에서, 외생성 DNA는 3' 및 5' 말단 둘 모두 상에서 단일 가이드 DNA(sgDNA)-PAM 서열이 측접된다. 바람직한 구현예에서, 외생성 DNA는 CRISPR-C2c1 절단 이후에 방출된다.

[1590] **1a 포함, 당원축적병**

[1591] 1a 당원축적병은 글루코스-6-포스파타제 효소의 결핍으로 인한 유전 질환이다. 이러한 결핍은 글리코겐 및 글루코스 신생 합성으로 유리 글루코스를 생성시키는 간의 능력을 손상시킨다. 일부 구현예에서, 글루코스-6-포스파타제를 코딩하는 유전자가 표적이 된다. 일부 구현예에서, 당원 축적병이 치료된다. 일부 구현예에서, 지질 입자, 예컨대, LNP에 C2c1 (단백질 또는 mRNA 형태)의 캡슐화에 의해 간에 전달된다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1592] 일부 구현예에서, 1a를 포함한, 당원축적병이 표적이 되고, 바람직하게, 예를 들어 병태/질환/감염과 연관된 폴리뉴클레오티드를 표적으로 하여 치료된다. 연관된 폴리뉴클레오티드는 유전자를 포함할 수 있는, DNA를 포함한다 (여기서 유전자는 임의의 코딩 서열 및 조절 엘리먼트 예컨대 인핸서 또는 프로모터를 포함). 일부 구현예에서, 연관된 폴리뉴클레오티드는 SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, 또는 PFKM 유전자를 포함할 수 있다.

[1593] **후를러 증후군**

[1594] 후를러 증후군은 또한 I형 점액다당류증 (MPS I)이라고도 알려져 있고, 후를러 질환은 리소솜 내 점액다당류의 분해를 담당하는 효소인, 알파-L-이두로니다제의 결핍으로 인한 글리코사미노글리칸 (이전에 뮤코폴리사카라이드라고 알려짐)의 축적을 야기시키는 유전 장애이다. 후를러 증후군은 종종 리소솜 저장 질환으로 분류되고 헌터 증후군과 임상적으로 관련된다. 헌터 증후군은 X-연관된 반면 후를러 증후군은 상염색체 열성이다. MPS I은 증상의 중증도를 기반으로 3개 아형으로 분류된다. 모든 3개 유형은 효소 α-L-이두로니다제의 부재, 또는 불충분한 수준에 의해 초래된다. MPS I H 또는 후를러 증후군은 MPS I 아형 중 가장 중증이다. 다른 2개 유형은 MPS I S 또는 샤이에 증후군 및 MPS I H-S 또는 후를러-샤이에 증후군이다. MPS I 부모에게 태어난 아이는 염색체 4의 4p16.3 부위로 맵핑된 결함성 IDUA 유전자를 보유한다. 이의 이두로니다제 효소 단백질 산물때문에 이 유전자는 IDUA라고 명명된다. 2001년 현재, IDUA 유전자 내 52개의 상이한 돌연변이가 후를러 증후군을 야기하는 것으로 확인되었다. 레트로바이러스, 렌티바이러스, AAV, 및 심지어 비바이러스 벡터를 통한 이두로니다제 유전자의 전달에 의한 MPS I의 마우스, 개, 및 고양이 모델의 성공적 치료.

[1595] 일부 구현예에서, α-L-이두로니다제 유전자가 표적이 되고 복구 주형이 바람직하게 제공된다. 일부 구현예에서, CRISPR 단백질은 C2c1이고, 시스템은 I. CRISPR-Cas 시스템 RNA 폴리뉴클레오티드 서열로서, (a) 표적 서열과 하이브리드화할 수 있는 가이드 RNA 폴리뉴클레오티드, 및 (b) 직접 반복부 RNA 폴리뉴클레오티드를 포함하는 것인 폴리뉴클레오티드 서열, 및 II. 임의로 적어도 하나 이상의 핵 국재화 서열을 포함하는, C2c1을 코딩하는 폴리뉴클레오티드 서열을 포함하고, 여기서 직접 반복부 서열은 가이드 서열과 하이브리드화하여 표적 서열로 CRISPR 복합체의 서열-특이적 결합을 유도시키고, CRISPR 복합체는 (1) 표적 서열과 하이브리드화하거나 또는 하이브리드화가능한 가이드 서열, 및 (2) 직접 반복부 서열과 복합체를 형성하는 CRISPR 단백질을 포함하고, CRISPR 단백질을 코딩하는 폴리뉴클레오티드 서열은 DNA 또는 RNA이다. 일부 구현예에서, C2c1 이펙터 단백질은 T-풍부 PAM을 인식한다. 특정 구현예에서, PAM은 5'-TTN-3' 또는 5'-ATTN-3' 이다. 일정 구현예에서, MPS I와 관련된 관심 유전자좌는 5' 오버행을 갖는 스태거드 절단부를 생성시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 일부 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, 스태거드 절단은

NHEJ 또는 HDR에 후속한다. 일정 구현예에서, 관심 유전자좌는 주형 DNA 서열을 삽입, 또는 " 녹-인" 시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 특정 구현예에서, DNA 삽입은 적절한 배향으로 게놈에 통합되도록 설계된다. Maresca 등 (Genome Res. 2013 Mar; 23(3): 539-546)은 징크 핑거 뉴클레아제 (ZFN) 및 Tale 뉴클레아제 (TALEN)를 사용해 적용가능한 부위 지정된, 정밀 삽입 방법을 기술하는데, 여기서 5' 오버행을 갖는 짧은, 이중-가닥 DNA가 상보성 말단에 결찰되어, 인간 세포주에서 한정된 유전자좌에서 15-kb 외생성 발현 카세트의 정밀한 삽입을 가능하게 하였다. He 등 (Nucleic Acids Res. 2016 May 19; 44(9))은 4.6 kb 무프로모터 ires-eGFP 단편을 GAPDH 유전자좌에 CRISPR/Cas9-유도 부위-특이적 녹-인으로 체세포 LO2 세포에서 최대 20% GFP+ 세포를 산출하였고, NHEJ 경로에 의해 매개되는 인간 배아 줄기 세포에서 1.70% GFP+ 세포를 산출하였다고 기술하였으며, 또한 NHEJ-기반 녹-인이 조사된 모든 인간 세포 유형에서 HDR-매개 유전자 표적화에 비해 더 효율적이라고 보고하였다. C2c1이 5' 오버행을 갖는 스태거드 절단을 생성시키기 때문에, 당업자는 본 명세서에 개시된 CRISPR-C2c1 시스템을 사용해 관심 유전자좌에서 외생성 DNA 삽입을 발생시키기 위해서 Meresca 등 및 He 등에 기술된 것과 유사한 방법을 사용할 수 있다.

[1596] 일정 구현예에서, 관심 유전자좌는 PAM 서열의 원위 말단에서 CRISPR-C2c1 시스템에 의해 먼저 변형되고, PAM 서열 근처에서 CRISPR-C2c1 시스템에 의해 더욱 변형되며 HDR을 통해 복구된다. 일정 구현예에서, 관심 유전자좌는 HDR을 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ를 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 바람직한 구현예에서, 외생성 DNA는 3' 및 5' 말단 둘 모두상에서 단일 가이드 DNA(sgDNA)-PAM 서열이 측접된다. 바람직한 구현예에서, 외생성 DNA는 CRISPR-C2c1 절단 이후에 방출된다.

[1597] **HIV 및 AIDS**

[1598] 일부 구현예에서, HIV 및 AIDS의 치료, 예방 또는 진단이 제공된다. 표적은 바람직하게 HIV 내 CCR5 유전자이다. 이것은 WO2015148670A1에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1599] **베타 지중해 빈혈 및 겸상 적혈구 질환 (SCD)**

[1600] 일부 구현예에서, 베타 지중해 빈혈의 치료, 예방 또는 진단을 제공한다. 표적은 바람직하게 BCL11A 유전자이다. 이것은 WO2015148860에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1601] 일부 구현예에서, 겸상 적혈구 질환 (SCD)의 치료, 예방 또는 진단이 제공된다. 표적은 바람직하게 HBB 또는 BCL11A 유전자이다. 이것은 WO2015148863에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유

전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다. 일정 구현예에서, 관심 유전자좌는 주형 DNA 서열을 삽입, 또는 " 녹-인" 시켜서 CRISPR-C2c1 복합체에 의해 변형된다. 특정 구현예에서, DNA 삽입은 적절한 배향으로 게놈에 통합되도록 설계된다. 바람직한 구현예에서, 관심 유전자좌는 비분열 세포에서 CRISPR-C2c1 시스템에 의해 변형되고, 여기서 상동성 지정 복구 (HDR) 기전을 통한 게놈 편집이 특히 도전적이다 (Chan et al., Nucleic acids research. 2011;39:5955-5966). Maresca 등 (Genome Res. 2013 Mar; 23(3): 539-546)은 징크 핑거 뉴클레아제 (ZFN) 및 Tale 뉴클레아제 (TALEN)를 사용해 적용가능한 부위 지정된, 정밀 삽입 방법을 기술하는데, 여기서 5' 오버행을 갖는 짧은, 이중-가닥 DNA가 상보성 말단에 결찰되어, 인간 세포주에서 한정된 유전자좌에서 15-kb 외생성 발현 카세트의 정밀한 삽입을 가능하게 하였다. He 등 (Nucleic Acids Res. 2016 May 19; 44(9))은 4.6 kb 무프로모터 ires-eGFP 단편을 GAPDH 유전자좌에 CRISPR/Cas9-유도 부위-특이적 녹-인으로 체세포 LO2 세포에서 최대 20% GFP+ 세포를 산출하였고, NHEJ 경로에 의해 매개되는 인간 배아 줄기 세포에서 1.70% GFP+ 세포를 산출하였다고 기술하였으며, 또한 NHEJ-기반 녹-인이 조사된 모든 인간 세포 유형에서 HDR-매개 유전자 표적화에 비해 더 효율적이라고 보고하였다. C2c1이 5' 오버행을 갖는 스태거드 절단을 생성시키기 때문에, 당업자는 본 명세서에 개시된 CRISPR-C2c1 시스템을 사용해 관심 유전자좌에서 외생성 DNA 삽입을 발생시키기 위해서 Meresca 등 및 He 등에 기술된 것과 유사한 방법을 사용할 수 있다.

[1602] 일정 구현예에서, 관심 유전자좌는 PAM 서열의 원위 말단에서 CRISPR-C2c1 시스템에 의해 먼저 변형되고, PAM 서열 근처에서 CRISPR-C2c1 시스템에 의해 더욱 변형되며 HDR을 통해 복구된다. 일정 구현예에서, 관심 유전자좌는 HDR을 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 일부 구현예에서, 관심 유전자좌는 NHEJ를 통한 외생성 DNA 서열의 삽입, 결실, 또는 돌연변이를 도입시켜 CRISPR-C2c1 시스템에 의해 변형된다. 바람직한 구현예에서, 외생성 DNA는 3' 및 5' 말단 둘 모두상에서 단일 가이드 DNA(sgDNA)-PAM 서열이 측접된다. 바람직한 구현예에서, 외생성 DNA는 CRISPR-C2c1 절단 이후에 방출된다.

[1603] **헤르페스 심플렉스 바이러스 1 및 2**

[1604] 헤르페스바이러스과는 75-200개 유전자를 갖는 선형 이중 가닥 DNA 게놈으로 구성된 바이러스 패밀리이다. 유전자 편집 목적을 위해서, 가장 일반적으로 연구되는 패밀리는 헤르페스 심플렉서 바이러스 -1 (HSV-1)로서, 다른 바이러스 벡터에 비해 구별되는 다수의 장점을 갖는 바이러스이다 (고찰: Vannuci et al. (2003)). 따라서, 일부 구현예에서, 바이러스 벡터는 HSV 바이러스 벡터이다. 일부 구현예에서, HSV 바이러스 벡터는 HSV-1이다.

[1605] HSV-1은 대략 152 kb의 이중 가닥 DNA의 거대 게놈을 갖는다. 이러한 게놈은 80개 초과의 유전자로 구성되고, 이의 많은 것들은 치환 또는 제거될 수 있어서, 30-150 kb의 유전자 삽입을 허용한다. HSV-1로부터 유래된 바이러스 벡터는 일반적으로 3개 그룹으로 분리된다: 복제-적격 약독화 벡터, 복제-부적격 재조합 벡터, 및 앰플리콘으로 알려진 결함성 헬퍼-의존적 벡터. 벡터로서 HSV-1을 사용하는 유전자 전달은 이전에, 예를 들어, 신경병증성 동통 (참조: 예를 들어, Wolfe et al. (2009) Gene Ther) 및 류마티스성 관절염 (참조: 예를 들어, Burton et al. (2001) Stem Cells)의 경우에 입증되었다.

[1606] 따라서, 일부 구현예에서, 바이러스 벡터는 HSV 바이러스 벡터이다. 일부 구현예에서, HSV 바이러스 벡터는 HSV-1이다. 일부 구현예에서, 벡터는 하나 이상의 CRISPR 성분의 전달을 위해 사용된다. C2c1 및 하나 이상의 가이드 RNA, 예를 들어, 2개 이상, 3개 이상, 또는 4개 이상의 가이드 RNA의 전달에 특히 유용할 수 있다. 일부 구현예에서, 벡터는 다중복합 시스템에서 유용하다. 일부 구현예에서, 전달은 신경병증성 동통 또는 류마티스성 관절염의 치료를 위한 것이다.

[1607] 일부 구현예에서, HSV-1 (헤르페스 심플렉스 바이러스 1)의 치료, 예방, 또는 진단이 제공된다. 표적은 바람직하게 HSV-1의 UL19, UL30, UL48 또는 UL50 유전자이다. 이것은 WO2015153789에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다.

[1608] 다른 구현예에서, HSV-2 (헤르페스 심플렉스 바이러스 2)의 치료, 예방, 또는 진단이 제공된다. 표적은 바람직하게 HSV-2의 UL19, UL30, UL48 또는 UL50 유전자이다. 이것은 WO2015153791에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다.

[1609] 일부 구현예에서, 원발성 개방각 녹내장 (POAG)의 치료, 예방 또는 진단이 제공된다. 표적은 바람직하게

MYOC 유전자이다. 이것은 WO2015153780에 기술되어 있고, 이 개시는 참조로 본 명세서에 편입된다. 본 발명은 상기 기술된 바와 같은 방법과 함께 적용될 수 있다. C2c1 단백질과 관련하여 CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1610] **양자 세포 요법**

[1611] 본 발명은 또한 양자 요법을 위해 세포를 변형시키기 위해서, 본 명세서에 설명된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질 시스템의 사용을 고려한다.

[1612] 본 명세서에서 사용되는, " ACT" , " 양자 세포 요법" 및 " 양자 세포 전달" 은 상호교환적으로 사용될 수 있다. 일정 구현예에서, 양자 세포 요법 (ACT)은 세포의 생착에 의해 새로운 숙주에게 기능성 및 특징을 전달하려는 목적으로 환자에게 세포의 전달을 의미한다 (참조: Mettananda et al., Editing an α-globin enhancer in primary human hematopoietic stem cells as a treatment for β-thalassemia, Nat Commun. 2017 Sep 4;8(1):424). 본 명세서에서 사용되는 용어 "생착 (engraft, 또는 engraftment)"은 조직의 존재하는 세포와 접촉을 통해 생체내에서 관심 조직으로 세포 도입의 과정을 의미한다. 양자 세포 요법 (ACT)는 세포, 가장 일반적으로 면역-유래 세포를 동일한 환자에게 다시 역으로 또는 새로운 숙주에게 면역학적 기능성 및 특징을 전달하려는 목적으로 새로운 수령체 숙주에게 전달하는 것을 의미한다. 가능하다면, 자기유래 세포의 사용이 GVHD 문제를 최소화하여 수령체를 돕는다. 자기유래 종양 침윤성 림프구 (TIL) (Besser et al., (2010) Clin. Cancer Res 16 (9) 2646-55; Dudley et al., (2002) Science 298 (5594): 850-4; and Dudley et al., (2005) Journal of Clinical Oncology 23 (10): 2346-57.) 또는 유전적으로 재지정된 말초 혈액 단핵 세포 (Johnson et al., (2009) Blood 114 (3): 535-46; and Morgan et al., (2006) Science 314(5796) 126-9)의 양자 전달이 흑색종 및 직결장 암종을 포함한 후기 고형 종양 환자를 비롯하여, CD19-발현 혈액학적 악성종 환자 (Kalos et al., (2011) Science Translational Medicine 3 (95): 95ra73)를 성공적으로 치료하는데 사용되었다. 일정 구현예에서, 동종 세포 면역 세포가 전달된다 (참조: 예를 들어, Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266). 본 명세서에서 더욱 기술된 바와 같이, 동종 세포는 동종이식편 반응을 감소시키고 이식편 대 숙주 질환을 예방하기 위해 편집될 수 있다. 따라서, 동종 세포의 사용은 건강한 도너로부터 세포를 수득할 수 있게 하고 진단 후 환자로부터 자기유래 세포를 제조하는 것이 아니라 환자에서 사용을 위해 준비하게 한다.

[1613] 일부 구현예에서, 본 명세서에 기술된 방법은 양자 면역요법의 방법에 관한 것으로서, T 세포가 CRISPR에 의해 생체외에서 편집되어 적어도 하나의 유전자를 조절하고 이후에 이를 필요로 하는 환자에게 투여된다. 일부 구현예에서, CRISPR 편집은 편집된 T 세포의 적어도 하나의 표적 유전자의 발현의 녹아웃 또는 녹다운을 포함한다. 일부 구현예에서, 표적 유전자의 조절 이외에도, T 세포는 또한 (1) 키메라 항원 수용체 (CAR) 또는 T-세포 수용체 (TCR)을 코딩하는 외생성 유전자의 녹인, (2) 면역 체크포인트 수용체의 발현의 녹아웃 또는 녹다운, (3) 내생성 TCR의 발현의 녹아웃 또는 녹다운, (4) 인간 백혈구 항원 클래스 I (HLA-I) 단백질의 발현의 녹아웃 또는 녹다운, 및/또는 (5) 외생성 CAR 또는 TCR에 의해 표적화된 항원을 코딩하는 내생성 유전자의 발현의 녹아웃 또는 녹다운을 위해서 CRISPR에 의해 생체외에서 편집된다.

[1614] 일부 구현예에서, T 세포는 CRISPR 이펙터 단백질, 및 표적 서열에 하이브리드화하는 가이드 서열, tracr 메이트 서열 및 tracr 메이트 서열과 하이브리드화할 수 있는 tracr 서열을 포함하는 가이드 분자를 코딩하는 아데노-연관 바이러스 (AAV)와 생체외에서 접촉한다. 일부 구현예에서, T 세포는 가이드 분자와 복합체 형성하는 CRISPR 이펙터 단백질을 포함하는 리보핵단백질 (RNP)와 (예를 들어, 전기천공을 통해) 생체외에서 접촉하고, 가이드 분자는 표적 서열과 하이브리드화할 수 있는 가이드 서열, tracr 메이트 서열, 및 tracr 메이트 서열과 하이브리드화할 수 있는 tracr 서열을 포함한다. 참조: Rupp et al., Scientific Reports 7:737 (2017); Liu et al., Cell Research 27:154-157 (2017). 일부 구현예에서, T 세포는 CRISPR 이펙터 단백질, 및 표적 서열에 하이브리드화하는 가이드 서열, tracr 메이트 서열 및 tracr 메이트 서열과 하이브리드화할 수 있는 Tracr 서열을 포함하는 가이드 분자를 코딩하는 mRNA와 (예를 들어, 전기천공을 통해) 생체외에서

접촉한다. 참조: Eyquem et al., Nature 543:113-117 (2017). 일부 구현예에서, T 세포는 렌티바이러스 또는 레트로바이러스 벡터와 생체외에서 접촉하지 않는다.

[1615] 일부 구현예에서, 방법은 CAR을 코딩하는 외생성 유전자를 녹인시키기 위해 CRISPR에 의해 생체외에서 T 세포를 편집하여서, 편집된 T 세포가 세포 표면 상에 위치된 특이적 단백질의 발현을 기반으로 암 세포를 인식할 수 있게 하는 단계를 포함한다. 일부 구현예에서, T 세포는 TCR을 코딩하는 외생성 유전자를 녹인시키기 위해 CRISPR에 의해 생체외에서 편집되어서, 편집된 T 세포가 암 세포의 표면 또는 내부로부터 유래된 단백질을 인식할 수 있게 하는 단계를 포함한다. 일부 구현예에서, 방법은 CRISPR 가이드 서열에 의해 표적화되는 게놈 유전자좌로 상동성-지정 복구 (HDR)을 통해 통합될 수 있는, 도너 서열로서 외생성 CAR-코딩 또는 TCR-코딩 서열을 제공하는 단계를 포함한다. 일부 구현예에서, 내생성 TCR α 불변 (TRAC) 유전자좌로 외생성 CAR 또는 TCR의 표적화는 토닉 CAR 신호전달을 감소시킬 수 있고 항원에 단일 또는 반복 노출 후에 CAR의 효과적인 내재화 및 재발현을 촉진하여 이펙터 T-세포 분화 및 고갈을 지연시킨다. 참조: Eyquem et al., Nature 543:113-117 (2017).

[1616] 일부 구현예에서, 방법은 암 세포에 의한 면역억제를 감소시키기 위해 하나 이상의 면역 체크포인트 수용체를 차단하도록 CRISPR에 의해 생체외에서 T 세포를 편집하는 단계를 포함한다. 일부 구현예에서, T 세포는 프로그램된 사멸-1 (PD-1) 신호전달 경로, 예컨대 PD-1 및 PD-L1에 관여되는 내생성 유전자의 녹아웃 또는 녹다운을 위해 CRISPR에 의해 생체외에서 편집된다. 일부 구현예에서, T 세포는 Pdcd1 유전자좌 또는 CD274 유전자좌를 돌연변이시키기 위해 CRISPR에 의해 생체외에서 편집된다. 일부 구현예에서, T 세포는 PD-1의 제1 엑손을 표적화하는 하나 이상의 가이드 서열을 사용하는 CRISPR에 의해 생체외에서 편집된다. 참조: Rupp et al., Scientific Reports 7:737 (2017); Liu et al., Cell Research 27:154-157 (2017).

[1617] 일부 구현예에서, 방법은 동종이계 양자 전달을 허용하도록 잠재적 동종이식편 반응성 TCR을 제거하기 위해 CRISPR에 의해 생체외에서 T 세포를 편집하는 단계를 포함한다. 일부 구현예에서, T 세포는 이식편 대 숙주 질환 (GVHD)를 피하기 위해서 TCR (예를 들어, αβ TCR)을 코딩하는 내생성 유전자를 녹아웃 또는 녹다운시키기 위해 CRISPR에 의해 생체외에서 편집된다. 일부 구현예에서, T 세포는 TRAC를 돌연변이시키기 위해 CRISPR에 의해 생체외에서 편집된다. 일부 구현예에서, T 세포는 TRAC 의 제1 엑손을 표적화하는 하나 이상의 가이드 서열을 사용하는 CRISPR에 의해 생체외에서 편집된다. 참조: Liu et al., Cell Research 27:154-157 (2017). 일부 구현예에서, 방법은 내생성 TCR (예를 들어, CAR cDNA 이후 자가-절단 P2A 펩티드를 코딩하는 도너 서열로)을 동시에 녹아웃시키면서, TRAC 유전자좌로 CAR 또는 TCR을 코딩하는 외생성 유전자를 녹인시키기 위한 CRISPR의 사용을 포함한다. 참조: Eyquem et al., Nature 543:113-117 (2017). 일부 구현예에서, 외생성 유전자는 내생성 TCR 프로모터의 하류에서 작동적으로 삽입된 프로모터-무함유 CAR-코딩 또는 TCR-코딩 서열을 포함한다.

[1618] 일부 구현예에서, 방법은 편집된 T 세포의 면역원성을 최소화시키기 위해서 HLA-I 단백질을 코딩하는 내생성 유전자를 녹아웃 또는 녹다운시키기 위해 CRISPR에 의해 생체외에서 T 세포를 편집하는 단계를 포함한다. 일부 구현예에서, T 세포는 베타-2 마크로글로불린 (B2M) 유전자좌를 돌연변이시키기 위해 CRISPR에 의해 생체외에서 편집된다. 일부 구현예에서, T 세포는 B2M의 제1 엑손을 표적화하는 하나 이상의 가이드 서열을 사용하는 CRISPR에 의해 생체외에서 편집된다. 참조: Liu et al., Cell Research 27:154-157 (2017). 일부 구현예에서, 방법은 내생성 B2M (예를 들어, CAR cDNA 이후 자가-절단 P2A 펩티드를 코딩하는 도너 서열로)을 동시에 녹아웃시키면서, B2M 유전자좌로 CAR 또는 TCR을 코딩하는 외생성 유전자를 녹인시키기 위한 CRISPR의 사용을 포함한다. 참조: Eyquem et al., Nature 543:113-117 (2017). 일부 구현예에서, 외생성 유전자는 내생성 B2M 프로모터의 하류에서 작동적으로 삽입된 프로모터-무함유 CAR-코딩 또는 TCR-코딩 서열을 포함한다.

[1619] 일부 구현예에서, 방법은 외생성 CAR 또는 TCR에 의해 표적화된 항원을 코딩하는 내생성 유전자를 녹아웃 또는 녹다운시키기 위해 CRISPR에 의해 생체외에서 T 세포를 편집하는 단계를 포함한다. 일부 구현예에서, T 세포는 인간 텔로머라제 역전사효소 (hTERT), 서비빈, 마우스 이중 미니트 2 상동체 (MDM2), 시토크롬 P450 1B 1 (CYP1B), HER2/neu, 빌름스 종양 유전자 1 (WT1), 리빈, 알파펙토단백질 (AFP), 암배 상원 (CEA), 뮤신 16 (MUC16), MUC1, 전립선-특이적 막 항원 (PSMA), p53 또는 사이클린 (DI)으로부터 선택된 종양 항원의 발현을 녹아웃 또는 녹다운시키기 위해 CRISPR에 의해 생체외에서 편집된다 (참조: WO2016/011210). 일부 구현예에서, T 세포는 B 세포 성숙화 항원 (BCMA), 경막 활성인자 및 CAML 인터액터 (TACI), 또는 B-세포 활성인자 수용체 (BAFF-R), CD38, CD138, CS-1, CD33, CD26, CD30, CD53, CD92, CD100, CD148, CD150, CD200, CD261, CD262, 또는 Cd362로부터 선택된 항원의 발현을 녹아웃 또는 녹다운시키기 위해 CRISPR에 의해 생체외에

서 편집된다 (참조: WO2017/011804).

[1620] 따라서 본 발명의 양상은 선택된 항원, 예컨대, 종양 연관 항원에 특이적인 면역 시스템 세포, 예컨대, T 세포의 양자 전달을 수반한다 (하기의 문헌들 참조: Maus et al., 2014, Adoptive Immunotherapy for Cancer 또는 Viruses, Annual Review of Immunology, Vol. 32: 189-225; Rosenberg and Restifo, 2015, Adoptive cell transfer as personalized immunotherapy for human cancer, Science Vol. 348 no. 6230 pp. 62- 68; 및, Restifo et al., 2015, Adoptive immunotherapy for cancer: harnessing the T cell Response. Nat. Rev. Immunol. 12(4): 269-281; 및 Jenson and Riddell, 2014, Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells. Immunol Rev. 257(1): 127- 144). 예를 들어, T 세포 수용체 (TCR)의 특이성을 변경시킴으로써, 예를 들어 선택된 펩티드 특이성을 갖는 새로운 TCR α 및 β 쇄를 도입함으로써 T 세포를 유전자 변형시키는 다양한 전략이 사용될 수 있다 (미국 특허 제8,697,854호; 국제 특허 출원 공개: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; 미국 특허 제8,088,379호 참조).

[1621] TCR 변형에 대한 대안으로서 또는 추가로, 기술되었던 매우 다양한 수용체 키메라 구성체를 사용하여, 악성 세포와 같은, 선택 표적에 특이적인 면역반응성 세포, 예컨대, T 세포를 생성하기 위해 키메라 항원 수용체 (CAR)가 사용될 수 있다 (미국 특허 제5,843,728호; 제5,851,828호; 5,912,170호; 제6,004,811호; 제6,284,240호; 제6,392,013호; 제6,410,014호; 제6,753,162호; 제8,211,422호; 및 국제 특허 출원 공개 WO9215322). B 세포 상의 CD19 같은 백혈병 항원에 대한 키메라 항원 수용체 (CAR)을 발현시키도록 조작된 자기유래 T 세포는 재발성 또는 난치성 B-세포 악성종의 치료를 위해 유망한 결과를 보였다. 그러나, 암 환자의 서브세트, 특히 심하게 전처리된 암 환자는 확장 실패때문에 이러한 고도의 능동 요법을 수용할 수 없을 수 있다. 게다가, 그들의 작은 혈액 부피로 인해 소아 암 환자에 대해 효과적인 치료 산물을 제작하는 것이 여전히 도전이 된다.

다른 한편으로, 개인맞춤 자기유래 T 세포 제작 및 광범위하게 "분산된" 접근법을 포함하여 자기유래 CAR-T 세포 요법의 고유한 특징은 자기유래 CAR-T 세포 요법의 산업화에 어려움을 초래한다. 하나 이상의 건강한 미관련 도너로부터 유래되지만, 이식편 대 숙주 질환 (GVHD)을 피할 수 있고 그들 면역원성을 최소화시킨, 보편적 Cd19-특이적 CAR-T 세포 (UCART019)는 의심할 바 없이 상기 언급된 문제를 해결하기 위한 대안적 옵션이다.

대안적인 CAR 구성체는 연속 세대에 속하는 것으로 특징규명될 수 있다. 제1-세대 CAR은 전형적으로 항원에 특이적인 항체의 단일쇄 가변 단편으로 이루어지는데, 예를 들어 가요성 링커에 의해, 예를 들어, CD8α 힌지 도메인 및 CD8α 막관통 도메인에 의해, CD3ζ 또는 FcRγ 중 하나의 막관통 및 세포내 신호전달 도메인에 연결되는 특정 항체의 VH에 연결된 VL을 포함한다 (scFv-CD3ζ 또는 scFv-FcRγ; 미국 특허 제7,741,465호; 미국 특허 제5,912,172호; 미국 특허 제5,906,936호 참조). 제2-세대 CAR은 엔도도메인 내에 하나 이상의 공자극 분자, 예컨대, CD28, OX40 (CD134), 또는 4-1BB(CD137)의 세포내 도메인을 혼입시킨다 (예를 들어 scFv-CD28/OX40/ 4--CD3ζ; 미국 특허 제8,911,993호; 제8,916,381호; 제8,975,071호; 제9,101,584호; 제9,102,760호; 제9,102,761호 참조). 제3-세대 CAR은 CD3ζ-사슬, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, 또는 CD28 신호전달 도메인 (예를 들어, scFv-CD28-4-1BB-CD3ζ 또는 scFv-CD28-OX40-CD3ζ; 참조: U.S 특허 제8,906,682호; U.S. 특허 제8,399,645호; U.S. 특허 제5,686,281호; PCT 공개 번호 WO2014134165; PCT 공개 번호 WO2012079000)같은 공자극성 엔도도메인의 조합을 포함한다. 대안적으로, 공자극은 그들의 천연 αβTCR의 맞물림 후, 예를 들어 전문적 항원-제시 세포 상에서 항원에 의해, 수반하는 공자극으로 활성화되고 확장될 수 있도록 선택되는 항원-특이적 T 세포에서 CAR을 발현시킴으로써 조직될 수 있다. 추가로, 추가의 조작된 수용체가 예를 들어 T-세포 공격의 표적화를 개선시키기 위해 그리고/또는 부작용을 최소화하기 위해 면역반응성 세포 상에 제공될 수 있다. Han 등 (clinicaltrials, A Study Evaluating UCART019 in Patients with Relapsed or Refactory CD19+ Lukemia and Lymphoma)은 내생성 TCR 및 B2M 유전자를 동시에 파괴하기 위해 CAR의 렌티바이러스 전달 및 CRISPR RNA 전기천공을 조합하여 유전자-파괴된 동종이계 CD19-지정 BBζ CAR-T 세포 (UCART019라고 함)를 생성시켰고 숙주-매개 면역을 피하고 GVHD 없이 항백혈병 효과를 전달할 수 있는지 여부가 시험될 것이다.

[1622] 표적 면역반응성 세포를 형질전환시키기 위해 대안의 기법, 예컨대, 원형질체 융합, 리포펙션, 형질감염 또는 전기천공법이 사용될 수 있다. 매우 다양한 벡터, 예컨대, 레트로바가러스 벡터, 렌티바이러스 벡터, 아데노바이러스 벡터, 아데노-연합 바이러스 벡터, 플라스미드 또는 트랜스포존, 예컨대, 슬리핑 뷰티 (Sleeping Beauty) 트랜스포존이 사용될 수 있고 (미국 특허 제6,489,458호; 제7,148,203호; 제7,160,682호; 제7,985,739호; 제8,227,432호 참조), 예를 들어 CD3ζ 및 CD28 또는 CD137 중 하나를 통해 2세대 항원-특이적 CAR 신호전달을 이용하여 CAR을 도입시키는데 사용될 수 있다. 바이러스 벡터는, 예를 들어, HIV, SV40, EBV, HSV 또

는 BPV에 기반한 벡터를 포함할 수 있다.

[1623] 형질전환을 위해 표적화된 세포는, 예를 들어 T 세포, 자연살해 (NK) 세포, 세포독성 T 림프구(CTL), 조절 T 세포, 인간 배아 줄기 세포, 종양-침윤성 림프구(TIL) 또는 림구 세포가 분화될 수 있는 다능성 줄기 세포를 포함할 수 있다. 목적하는 CAR을 발현시키는 T 세포는, 예를 들어, 암 항원 및 공자극 분자를 공동발현시키는 γ-조사된 활성화 및 증식 세포 (AaPC)와 함께 공배양을 통해 선택될 수 있다. 조작된 CAR T-세포는 가용성 인자, 예컨대, IL-2 및 IL-21의 존재 하에 AaPC에 대한 공동 배양에 의해, 확장될 수 있다. 이러한 확장은, 예를 들어, 기억 CAR+ T 세포를 제공하기 위해 수행될 수 있다 (예를 들어, 비효소적 디지털 어레이 및/또는 다중-패널 유세포 분석기에 의해 평가될 수 있음). 이러한 방식으로, 항원-보유 종양에 대해 (선택적으로 목적하는 케모카인, 예컨대, 인터페론-γ의 생성과 함께) 특이적 세포독성 활성을 갖는 CAR T 세포가 제공될 수 있다. 이런 종류의 CAR T 세포는, 예를 들어, 동물 모델에서, 예를 들어 종양 이종이식을 위협하기 위해 사용될 수 있다.

[1624] 일반적으로, CAR은 세포외 도메인, 경막 도메인, 및 세포내 도메인으로 구성되고, 세포외 도메인은 사전결정된 표적에 특이적인 항원-결합 도메인을 포함한다. CAR의 항원-결합 도메인이 종종 항체 또는 항체 단편 (예를 들어, 단쇄 가변 단편, scFv)이지만, 결합 도메인은 이것이 표적의 특이적 인식을 야기시키는 한 특별히 제한되지 않는다. 예를 들어, 일부 구현예에서, 항원-결합 도메인은 수용체를 포함할 수 있어서, CAR은 수용체의 리간드에 결합할 수 있다. 대안적으로, 항원-결합 도메인은 리간드를 포함할 수 있어서, CAR은 그 리간드의 내생성 수용체에 결합할 수 있다.

[1625] CAR의 항원-결합 도메인을 일반적으로 힌지 또는 스페이서에 의해 경막 도메인으로부터 이격된다. 스페이서는 또한 특별히 제한되지 않으며, CAR에 가요성을 부여하도록 설계된다. 예를 들어, 스페이서 도메인은 CH3 도메인의 일부분을 포함하여, 인간 Fc 도메인의 일부, 또는 임의의 면역글로불린, 예컨대 IgA, IgD, IgE, IgG, 또는 IgM, 또는 이의 변이체의 힌지 영역을 포함할 수 있다. 더 나아가서, 힌지 영역은 FcR 또는 다른 잠재적 간섭 객체에 의한 오프-표적 결합을 방지하도록 변형될 수 있다. 예를 들어, 힌지는 FcR에 결합을 감소시키기 위해서 S228P, L235E, 및/또는 N297Q 돌연변이 (카밧 번호매김에 따름)가 존재하거나 또는 없는 IgG4 Fc를 포함할 수 있다. 추가 스페이서/힌지는 제한없이 CD4, CD8 및 CD28 힌지 영역을 포함한다.

[1626] CAR의 경막 도메인은 천연 또는 합성 공급원으로부터 유래할 수 있다. 공급원이 천인 경우, 도메인은 임의의 막 결합 또는 경막 단백질로부터 유래될 수 있다. 본 개시에서 특별한 용도의 경막 영역은 CD8, CD28, CD3, CD45, CD4, CD5, CDS, CD9, CD 16, CD22, CD33, CD37, CD64, CD80, CD86, CD 134, CD137, CD 154, TCR로부터 유래될 수 있다. 대안적으로, 경막 도메인은 합성일 수 있고, 이러한 경우에 류신 및 발린 같은 소수성 잔기를 주로 포함하게 될 것이다. 바람직하게, 페닐알라닌, 트립토판, 및 발린의 트리플릿이 합성 경막 도메인의 각 말단에서 발견될 것이다. 임의로, 짧은 올리고펩티드 또는 폴리펩티드 링커, 바람직하게 2 내지 10개 아미노산 길이인 것이 CAR의 세포질 신호전달 도메인 및 경막 도메인 사이에 연결을 형성할 수 있다. 글리신-세린 더블릿이 특히 적합한 링커를 제공한다.

[1627] 대안적인 CAR 구성체는 연속 세대에 속하는 것으로 특징규명될 수 있다. 제1-세대 CAR은 전형적으로 항원에 특이적인 항체의 단일쇄 가변 단편으로 이루어지는데, 예를 들어 가요성 링커에 의해, 예를 들어, CD8α 힌지 도메인 및 CD8α 막관통 도메인에 의해, CD3ζ 또는 FcRγ 중 하나의 막관통 및 세포내 신호전달 도메인에 연결되는 특정 항체의 VH에 연결된 VL을 포함한다 (scFv-CD3ζ 또는 scFv-FcRγ; 미국 특허 제7,741,465호; 미국 특허 제5,912,172호; 미국 특허 제5,906,936호 참조). 제2-세대 CAR은 엔도도메인 내에 하나 이상의 공자극 분자, 예컨대, CD28, OX40 (CD134), 또는 4-1BB(CD137)의 세포내 도메인을 혼입시킨다 (예를 들어 scFv-CD28/OX40/ 4-1BB-CD3ζ; 미국 특허 제8,911,993호; 제8,916,381호; 제8,975,071호; 제9,101,584호; 제9,102,760호; 제9,102,761호 참조). 제3-세대 CAR은 Cd3ζ-사슬, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, CD2, CD7, LIGHT, LFA-1, NKG2C, B7-H3, CD30, CD40, PD-1, 또는 CD28 신호전달 도메인 (예를 들어, scFv-CD28-4-1BB-CD3ζ 또는 scFv-CD28-OX40-CD3ζ; 참조: U.S. 특허 번호 8,906,682; U.S. 특허 번호 8,399,645; U.S. 특허 번호 5,686,281; PCT 공개 번호 WO2014134165; PCT 공개 번호 WO2012079000) 같은 공자극성 엔도도메인의 조합을 포함한다. 일정 구현예에서, 주요 신호전달 도메인은 CD3 제타, CD3 감마, CD3 델타, CD3 엡실론, 공통 FcR 감마 (FCERIG), FcR 베타 (Fc 엡실론 R1b), CD79a, CD79b, Fc 감마 RIIa, DAP10, 및 DAP12로 이루어진 군으로부터 선택된 단백질의 기능성 신호전달 도메인을 포함한다. 일정 바람직한 구현예에서, 주요 신호전달 도메인은 CD3ζ 또는 FcRγ의 기능성 신호전달 도메인을 포함한다. 일부 구현예에서, 하나 이상의 공자극성 도메인은 CD27, CD28, 4-1BB (CD137), OX40, CD30, CD40, PD-1, ICOS, 림프구 기능-연관 항원-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, CD83에 특이적으로 결합하는 리간드, CDS, ICAM-

1, GITR, BAFFR, HVEM (LIGHTR), SLAMF7, NKp80 (KLRF1), CD160, CD19, CD4, CD8 알파, CD8 베타, IL2R 베타, IL2R 감마, IL7R 알파, ITGA4, VLA1, CD49a, ITGA4, IA4, CD49D, ITGA6, VLA-6, CD49f, ITGAD, CD11d, ITGAE, CD103, ITGAL, CD11a, LFA-1, ITGAM, CD11b, ITGAX, CD11c, ITGB1, CD29, ITGB2, CD18, ITGB7, TNFR2, TRANCE/RANKL, DNAM1 (CD226), SLAMF4 (CD244, 2B4), CD84, CD96 (Tactile), CEACAM1, CRTAM, Ly9 (CD229), CD160 (BY55), PSGL1, CD100 (SEMA4D), CD69, SLAMF6 (NTB-A, Ly108), SLAM (SLAMF1, CD150, IPO-3), BLAME (SLAMF8), SELPLG (CD162), LTBR, LAT, GADS, SLP-76, PAG/Cbp, NKp44, NKp30, NKp46, 및 NKG2D로 이루어진 군으로부터, 각각 독립적으로 선택된 단백질의 기능성 신호전달 도메인을 포함한다. 일정 구현예에서, 하나 이상의 공자극성 신호전달 도메인은 4-1BB, CD27 및 CD28로 이루어진 군으로부터 서로 독립적으로 선택된 단백질의 기능성 신호전달 도메인을 포함한다. 일정 구현예에서, 키메라 항원 수용체는 U.S. 특허 번호 7,446,190에 기술된 바와 같이, CD3ζ 사슬의 세포내 도메인 (예컨대 US 7,446,190의 SEQ ID NO: 14로 표시된, 인간 CD3 제타 사슬의 아미노산 잔기 52-163), CD28 유래 신호전달 영역 및 항원-결합 엘리먼트 (또는 부분 또는 도메인; 예컨대 scFv)을 포함하는, 디자인을 가질 수 있다. CD28 부분은 제타 사슬 부분 및 항원-결합 엘리먼트 사이일 때, CD28의 경막 및 신호전달 도메인을 적합하게 포함할 수 있다 (예컨대 SEQ ID NO: 10의 아미노산 잔기 114-220, US 7,446,190의 SEQ ID NO: 6로 표시된 전체 서열; 이들은 Genbank 식별자 NM_006139 (서열 버전 1, 2 또는 3로 기재된 CD28의 하기 부분을 포함할 수 있음): IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDF AAYRS) (SEQ ID NO:478). 대안적으로, CD28 서열 및 항원-결합 엘리먼트 사이에 제타 서열이 놓일 때, CD28의 세포내 도메인이 단독으로 사용될 수 있다 (예컨대 US 7,446,190의 SEQ ID NO: 9로 표시된 아미노산 서열). 그런 이유로, 일정 구현예는 (a) 인간 CD3ζ 사슬의 세포내 도메인을 포함하는 제타 사슬 부분, (b) 공자극성 신호전달 영역, 및 (c) 항원-결합 엘리먼트 (또는 부분 또는 도메인)를 포함하는 CAR를 적용하고, 공자극성 신호전달 영역은 US 7,446,190의 SEQ ID NO: 6에 의해 코딩되는 아미노산을 포함한다.

[1628] 대안적으로, 공자극은 그들의 천연 αβTCR의 맞물림 후, 예를 들어 전문적 항원-제시 세포 상에서 항원에 의해, 수반하는 공자극으로 활성화되고 확장될 수 있도록 선택되는 항원-특이적 T 세포에서 CAR을 발현시킴으로써 조직될 수 있다. 추가로, 추가의 조작된 수용체가 예를 들어 T-세포 공격의 표적화를 개선시키기 위해 그리고/또는 부작용을 최소화하기 위해 면역반응성 세포 상에 제공될 수 있다.

[1629] 예로서 제한없이, [Kochenderfer et al., (2009) J Immunother. 32 (7): 689-702]은 항-CD19 키메라 항원 수용체 (CAR)를 기술한다. FMC63-28Z CAR은 FMC63 마우스 하이브리도마로부터 유래된 CD19를 인식하는 단쇄 가변 영역 모이어티 (scFv) (Nicholson et al., (1997) Molecular Immunology 34: 1157-1165), 인간 CD28 분자의 일부, 및 인간 TCR-ζ 분자의 세포내 성분을 함유하였다. FMC63-CD828BBZ CAR은 FMC63 scFv, CD8 분자의 힌지 및 경막 영역, CD28 및 4-1Bb의 세포질 도메인, 및 TCR-ζ 분자의 세포질 성분을 함유하였다. FMC63-28Z CAR에 포함된 CD28 분자의 정확한 서열은 Genbank 식별자 NM_006139에 상응하였고; 서열은 아미노산 서열 IEVMYPPPY로 출발한 모든 아미노산을 포함하였고 단백질의 카르복시-말단까지 모든 방식으로 계속된다. 벡터의 항-CD19 scFv 성분을 코딩하기 위해서 저자는 이전에 공기된 CAR의 일부를 기반으로 DNA 서열을 설계하였다 (Cooper et al., (2003) Blood 101: 1637-1644). 이 서열은 5' 말단에서 3' 말단으로 인프레임으로 하기 성분들을 코딩하였다: XhoI 부위, 인간 과립구-마크로파지 콜로니-자극 인자 (GM-CSF) 수용체 α-사슬 신호 서열, FMC63 경쇄 가변 영역 (Nicholson et al., supra), 링커 펩티드 (Cooper et al., supra), FMC63 중쇄 가변 영역 (Nicholson et al., supra), 및 NotI 부위. 이러한 서열을 코딩하는 플라스미드를 XhoI 및 NotI로 분해하였다. MSGV-FMC63-28Z 레트로바이러스 벡터를 형성하기 위해서, FMC63 scFv를 코딩하는 XhoI 및 NotI-분해 단편은 인간 CD28의 세포외 부분의 일부, 인간 CD28의 전체 경막 및 세포질 부분, 및 인간 TCR-ζ 분자의 세포질 부분을 비롯하여, MSGV 레트로바이러스 골격 (Hughes et al., (2005) Human Gene Therapy 16: 457-472)을 코딩하는 제2의 XhoI 및 NotI-분해 단편으로 결찰시켰다 (Maher et al., 2002) Nature Biotechnology 20: 70-75). FMC63-28Z CAR은 특히 재발성/난치성 공격적 B-세포 비호지킨 림프종 (NHL) 환자의 치료를 위해서, Kite Pharma, Inc.에서 개발한 KTE-C19 (axicabtagene ciloleucel) 항-CD19 CAR-T 요법 제품에 포함된다. 그리하여, 일정 구현예에서, 앙자 세포 요법에 의도되는 세포, 보다 특히 면역반응성 세포 예컨대 T 세포는 항원에 특이적으로 결합하는 세포외 항원-결합 엘리먼트 (또는 부분 또는 도메인; 예컨대 scFv), CD3ζ 사슬의 세포내 도메인을 포함하는 세포내 신호전달 도메인, 및 CD28의 신호전달 도메인을 포함하는 공자극성 신호전달 영역을 포함하는 CAR을 포함할 수 있다. 바람직하게, CD28 아미노산 서열은 Genbank 식별자 NM_006139 (서열 버전 1, 2 또는 3)로 기재된 바와 같고 아미노산 서열 IEVMYPPPY에서 출발하여 모든 방식으로 단백질의 카르복시-말단까지 계속된다. 서열은 본 명세서에서 복제한다: IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWVLVVVGGVLACYSLLVTVAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDF

AAYRS (SEQ ID NO:479).

[1630] 추가 항-CD19 CAR은 Wo2015187528에 더욱 기술된다. 보다 특히, 본 명세서에 참조로 편입되는 WO2015187528의 실시예 1 및 표 1은 완전한 인간 항-CD19 단일클론 항체 (47G4, US20100104509에 기술됨) 및 쥐과 항-CD19 단일클론 항체 (Nicholson 등이 기술하고, 상기 설명됨)를 기반으로 항-CD19 CAR의 생성을 입증하였다. 신호 서열 (인간 CD8-알파 또는 GM-CSF 수용체), 세포외 및 경막 영역 (인간 CD8-알파) 및 세포내 T-세포 신호전달 도메인 (CD28-CD3ζ; 4-1BB-CD3ζ; CD27-CD3ζ; CD28-CD27-CD3ζ, 4-1BB-CD27-CD3ζ; CD27-4-1BB-CD3ζ; CD28-CD27-FceRI 감마 사슬; 또는 CD28-FceRI 감마 사슬)의 다양한 조합이 개시되었다. 그리하여, 일정 구현예에서, 양자 세포 요법에 의도되는 세포, 보다 특히 면역반응성 세포 예컨대 T 세포는 항원에 특이적으로 결합하는 세포외-항원 결합 엘리먼트, WO2015187528의 표 1에 기재된 바와 같은 세포외 및 경막 영역, 및 WO201518752의 표 1에 기재된 바와 같은 세포내 T-세포 신호전달 도메인을 포함하는 CAR을 포함할 수 있다. 바람직하게, 항원은 CD19이고, 보다 바람직하게 항원-결합 엘리먼트는 항-CD19 scFv, 보다 더 바람직하게는 WO2015187528의 실시예 1에 기술된 바와 같은 마우스 또는 인간 항-CD19 scFv이다. 일정 구현예에서, CAR은 WO2015187528의 표 1에 기재된 바와 같은 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, 또는 SEQ ID NO: 13의 아미노산 서열을 포함하거나, 그로 본질적으로 이루어지거나, 또는 이루어진다.

[1631] 예로서 제한없이 CD70 항원을 인식하는 키메라 항원 수용체는 WO2012058460A2에 기술되어 있다 (참조: Park et al., CD70 as a target for chimeric antigen receptor T cells in head and neck squamous cell carcinoma, Oral Oncol. 2018 Mar;78:145-150; 및 Jin et al., CD70, a novel target of CAR T-cell therapy for gliomas, Neuro Oncol. 2018 Jan 10;20(1):55-65). CD70은 미만성 거대 B-세포 및 여포성 림프종 및 또한 호지킨 림프종, 발덴스트롬 거대글로불린혈증 및 다발성 골수종, 및 HTLV-1- 및 EBV-연관 악성종의 악성 세포에 의해 발현된다 (Agathanggelou et al. Am.J.Pathol. 1995;147: 1152-1160; Hunter et al., Blood 2004; 104:4881. 26; Lens et al., J Immunol. 2005;174:6212-6219; Baba et al., J Virol. 2008;82:3843-3852.). 또한, CD70은 신장 세포 암종 및 교모세포종 같은 비혈액학적 악성종에 의해 발현된다 (Junker et al., J Urol. 2005;173:2150-2153; Chahlavi et al., Cancer Res 2005;65:5428-5438). 생리적으로, CD70 발현은 일시적이고 고도로 활성화된 T, B, 및 수지상 세포의 서브세트에 제한된다.

[1632] 예로서 제한없이, BCMA를 인식하는 키메라 항원 수용체가 기술되었다 (참조: 예를 들어 US20160046724A1; WO2016014789A2; WO2017211900A1; WO2015158671A1; US20180085444A1; WO2018028647A1; US20170283504A1; 및 WO2013154760A1).

[1633] 본 명세서에 개시된 CRISPR 시스템은 양자 세포 요법에서 표적화하려는 항원을 표적화하기 위해 사용될 수 있다. 일정 구현예에서, 질환 (예컨대 특히 종양 또는 암)의 양자 세포 요법 (예컨대 TIL, CAR, 또는 TCR T-세포 요법)에서 표적화하려는 항원 (예컨대 종양 항원)은 하기로 이루어진 군으로부터 선택될 수 있다: B 세포 성숙화 항원 (BCMA) (참조: Friedman et al., Effective Targeting of Multiple BCMA-Expressing Hematological Malignancies by Anti-BCMA CAR T Cells, Hum Gene Ther. 2018 Mar 8; Berdeja JG, et al. Durable clinical Responses in heavily pretreated patients with relapsed/refractory multiple myeloma: updated results from a multicenter study of bb2121 anti-Bcma CAR T cell therapy. Blood. 2017;130:740; and Mouhieddine and Ghobrial, Immunotherapy in Multiple Myeloma: The Era of CAR T Cell Therapy, Hematologist, May-June 2018, Volume 15, issue 3); PSA (전립선-특이적 항원); 전립선-특이적 막 항원 (PSMA); PSCA (전립선 줄기 세포 항원); 티로신-단백질 키나제 경막 수용체 ROR1; 섬유아세포 활성화단백질 (FAP); 종양-연관 당단백질 72 (TAG72); 암배 항원 (CEA); 상피 세포 부착 분자 (EPCAM); 메소테린; 인간 상피 성장 인자 수용체 2 (ERBB2 (Her2/neu)); 프로스타제; 프로스타트산 포스파타제 (PAP); 연장 인자 2 돌연변이체 (ELF2M); 인슐린-유사 성장 인자 1 수용체 (IGF-1R); gp100; BCR-ABL (브레이크포인트 클러스터 영역-아벨슨); 티로시나제; 뉴욕 식도 편평 세포 암종 1 (NY-ESO-1); κ-경쇄, LAGE (L 항원); MAGE (골수종 항원); 골수종-연관 항원 1 (MAGE-A1); MAGE A3; MAGE A6; 레구마인; 인간 파필로마마이러스 (HPV) E6; HPV E7; 프로스테린; 서비빈; PCTA1 (갈렉틴 8); 멜란-A/MART-1; Ras 돌연변이체; TRP-1 (티로시나제 관련 단백질 1, 또는 gp75); 티로시나제-관련 단백질 2 (TRP2); TRP-2/INT2 (TRP-2/인트론 2); RAGE (신장 항원); 진행성 당화 최종 산물에 대한 수용체 1 (RAGE1); 신장 편재성 1, 2 (RU1, RU2); 장 카르복실 에스터라제 (iCE); 열충격 단백질 70-2 (HSP70-2) 돌연변이체; 갑상선 자극 호르몬 수용체 (TSHR); CD123; CD171; CD19; CD20; CD22; CD26; CD30; CD33; CD44v7/8 (분화 클러스터 44, 엑손 7/8); CD53; CD92; CD100; CD148; CD150; CD200; CD261; CD262;

CD362; CS-1 (CD2 서브세트 1, CRACC, SLAMF7, CD319, 및 19A24); C-유형 렉틴-유사 분자-1 (CLL-1); 강글리오시드 GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); Tn 항원 (Tn Ag); Fms-유사 티로신 키나제 3 (FLT3); CD38; CD138; CD44v6; B7H3 (CD276); KIT (CD117); 인터루킨-13 수용체 서브유닛 알파-2 (IL-13Ra2); 인터루킨 11 수용체 알파 (IL-11Ra); 전립선 줄기 세포 항원 (PSCA); 프로테아제 세린 21 (PRSS21); 혈관 내피 성장 인자 수용체 2 (VEGFR2); 루이스(Y) 항원; CD24; 혈소판-유래 성장 인자 수용체 베타 (PDGFR-베타); 단계-특이적 배아 항원-4 (SSEA-4); 뮤신 1, 세포 표면 연관 (MUC1); 뮤신 16 (MUC16); 상피 성장 인자 수용체 (EGFR); 상피 성장 인자 수용체 변이체 III (EGFRvIII); 신경 세포 부착 분자 (NCAM); 카본산 언히드라제 IX (CAIX); 프로테오솜 (프로솜, 마크로파인) 서브유닛, 베타 유형, 9 (LMP2); 에프린 A형 수용체 2 (EphA2); 에프린 B2; 푸코실 GM1; 시알릴 루이스 부착 분자 (sLe); 강글리오시드 GM3 (aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); TGS5; 고분자량-골수종-연관 항원 (HMWMAA); o-아세틸-GD2 강글리오시드 (OAcGD2); 폴레이트 수용체 알파; 폴레이트 수용체 베타; 종양 상피 마커 1 (TEM1/CD248); 종양 상피 마커 7-관련 (TEM7R); 클라우딘 6 (CLDN6); G 단백질-커플링 수용체 클래스 C 그룹 5, 멤버 D (GPRC5D); 염색체 X 오픈 리딩 프레임 61 (CXORF61); CD97; CD179a; 역형성 림프종 키나제 (ALK); 폴리시알산; 태반-특이적 1 (PLAC1); globoH 글리코세라미드의 헥사사카라이드 부분 (GloboH); 유선 분화 항원 (NY-BR-1); 우로플라킨 2 (UPK2); A형 간염 바이러스 세포 수용체 1 (HAVCR1); 아드레노셉터 베타 3 (ADRB3); 판넥신 3 (PANX3); G 단백질-커플링 수용체 20 (GPR20); 림프구 항원 6 복합체, 유전자좌 K 9 (LY6K); 후각 수용체 51E2 (OR51E2); TCR 감마 교대식 리딩 프레임 단백질 (TARP); 빌름스 종양 단백질 (WT1); ETS 전좌 변이체 유전자 6, located on 염색체 12p 상에 위치 (ETV6-AML); 정자 단백질 17 (SPA17); X 항원 패밀리, 멤버 1A (XAGE1); 안지오포이어틴-결합 세포 표면 수용체 2 (Tie 2); CT (암/고환 (항원)); 골수종 암 고환 항원-1 (MAD-CT-1); 골수종 암 고환 항원-2 (MAD-CT-2); Fos-관련 항원 1; p53; p53 돌연변이체; 역 텔로머라제 역전사효소 (hTERT); 육종 전좌 브레이크포인트; 아폽토시스의 골수종 억제제 (ML-IAP); ERG (경막 프로테아제, 세린 2 (TMPRSS2) ETS 융합 유전자); N-아세틸 글루코사미닐-트랜스퍼라제 V (NA17); 쌍형성 박스 단백질 Pax-3 (PAX3); 안드로겐 수용체; 사이클린 B1; 사이클린 D1; v-myc 조류 골수구종증 바이러스 종양 유전자 신경아세포종 유래 상동체 (MYCN); Ras 상동체 패밀리 멤버 C (RhoC); 시토크롬 P450 1B1 (CYP1B1); CCCTC-결합 인자 (아연 핑거 단백질)-유사 (BORIS); T 세포-1 또는 3에 의해 인식된 편평 세포 암종 항원 (SART1, SART3); 쌍형성 박스 단백질 Pax-5 (PAX5); 프로아크로신 결합 단백질 sp32 (OY-TES1); 림프구-특이적 단백질 티로신 키나제 (LCK); A 키나제 앵커 단백질 4 (AKAP-4); 활액육종, X 브레이크포인트-1, -2, -3 또는 -4 (SSX1, SSX2, SSX3, SSX4); CD79a; CD79b; CD72; 백혈구-연관 면역글로불린-유사 수용체 1 (LAIR1); IgA 수용체의 Fc 단편 (FCAR); 백혈구 면역글로불린-유사 수용체 서브패밀리 A 멤버 2 (LILRA2); CD300 분자-유사 패밀리 멤버 f (CD300LF); C-type 렉틴 도메인 패밀리 12 멤버 A (CLEC12A); 골수 기질 세포 항원 2 (BST2); EGF-유사 모듈-함유 뮤신-유사 호르몬 수용체-유사 2 (EMR2); 림프구 항원 75 (LY75); 글리피칸-3 (GPC3); Fc 수용체-유사 5 (FCRL5); 마우스 이중 미니트 2 상동체 (MDM2); 리빈; 알파페토단백질 (AFP); 경막 활성인자 및 CAML 인터랙터 (TACI); B-세포 활성화 인자 수용체 (BAFF-R); V-Ki-ras2 커스텐 래트 육종 바이러스 종양유전자 상동체 (KRAS); 면역글로불린 람다-유사 폴리펩티드 1 (IGLL1); 707-AP (707 알라닌 프롤린); ART-4 (T4 세포에 의해 인식된 선암종 항원); BAGE (B 항원; b-카테닌/m, b-카테닌/돌연변이됨); CAMEL (골수종 상의 CTL-인식된 항원); CAP1 (암배 항원 펩티드 1); CASP-8 (캐스파제-8); CDC27m (세포-분열 주기 27 돌연변이됨); CDK4/m (사이클린-의존성 키나제 4 돌연변이됨); Cyp-B (사이클로필린 B); DAM (분화 항원 골수종); EGP-2 (상피 당단백질 2); EGP-40 (상피 당단백질 40); Erbb2, 3, 4 (적아세포 백혈병 바이러스 종양유전자 상동체-2, -3, 4); FBP (폴레이트 결합 단백질); fAchR (태아 아세틸콜린 수용체); G250 (당단백질 250); GAGE (G 항원); GnT-V (N-아세틸글루코사미닐트랜스퍼라제 V); HAGE (헬리코스 항원); ULA-A (인간 백혈구 항원-A); HST2 (인간 시그넷 링 종양 2); KIAA0205; KDR (키나제 삽입 도메인 수용체); LDLR/FUT (저밀도 지질 수용체/GDP L-푸코스: b-D-갈락토시다제 2-a-L 푸코실트랜스퍼라제); L1CAM (L1 세포 부착 분자); MC1R (멜라노코르틴 1 수용체); 미오신/m (미오신 돌연변이됨); MUM-1, -2, -3 (골수종 편재성 돌연변이 1, 2, 3); NA88-A (환자 M88의 NA cDNA 클론); KG2D (자연 살해 그룹 2, 멤버 D) 리간드; 종양태아 항원 (h5T4); p190 마이너 bcr-abl (190KD bcr-abl의 단백질); Pml/RARa (전골수구성 백혈병/레티노산 수용체 a); PRAME (골수종의 우선 발현 항원); SAGE (육종 항원); TEL/AML1 (전좌 Ets-패밀리 백혈병/급성 골수성 백혈병 1); TPI/m (트리오스포스페이트 이소머라제 돌연변이됨); CD70; 세포영양막 당단백질 (TPBG); αvβo 인테그린, B7-H3; B7-H6; CD20; CD44; 콘드로이틴 술페이트 프로테오글리칸 4 (CSPG4), bDGalpNAc(1-4)[aNeu5Ac(2-8)aNeu5Ac(2-3)]bDGalp(1-4)bDGlcp(1-1)Cer (GD2), aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer (GD3); 인간 백혈구 항원 A1 MAGE 패밀리 멤버 A1 (HLA-A1^{+}MAGEA1); 인간 백혈구 항원 A2 MAGE 패밀리 멤버 A1 (HLA-A2^{+}MAGEA1); 인간 백혈구 항원 A3 MAGE 패밀리 멤버 A1 (HLA-A3^{+}MAGEA1); MAGEA1; 인

간 백혈구 항원 A1 뉴욕 식도 편평 세포 암종 1 (FILA-A1^{+}NY-ESO-1); 인간 백혈구 항원 A2 뉴욕 식도 편평 세포 암종 1 (HLA-A2+NY-ESO-1), 람다 경쇄, 카파 경쇄, 종양 상피 마커 5 (TEM5), 종양 상피 마커 7 (TEM7), 종양 상피 마커 8 (TEM8), TEM5, TEM7, TEM8, IFN-유도성 p78, 멜라노트랜스페린 (p97), 인간 칼리크레인 (huK2), Axl, ROR2, FKBP11, KAMP3, ITGA8, FCRL5, LAGA-1, CD133, cD34, EBV 핵 항원-1 (EBNA1), 잠복 막단백질 1 (LMP1) 및 LMP2A, CD75, gp100, MICA, MICB, MART1, 암배 항원, CA-125, MAGEC2, CTAG2, CTAG1, pd-12, CLA, CD142, CD73, CD49c, CD66c, CD104, CD318, TSPAN8, CLEC14, 인간 면역결핍 바이러스 1 (HIV-1) 역전사효소 (RT), Cd16, BLTA, IL-2, IL-7, IL-15, IL-21,IL-12, CCR4, CCR2b, 헤파라나제, CD137L, LEM, 및 Bcl-2, Msln, Cd8, IL-15, 4-1BBL, OX40L, 4- IBB, cd95, cd27, HVENM, CXCR4; 및 이의 조합. 일부 예에서, 표적화하려는 항원은 CXCR일 수 있다. 일부 예에서, 표적화하려는 항원은 PD-1일 수 있다.

[1634] 일정 구현예예서, 질환 (예컨대 특히 종양 또는 암)의 양자 세포 요법 (예컨대 특히 CAR 또는 TCR T-세포 요법)에서 표적화하려는 항원은 종양-특이적 항원 (TSA)이다.

[1635] 일정 구현예예서, 질환 (예컨대 특히 종양 또는 암)의 양자 세포 요법 (예컨대 특히 CAR 또는 TCR T-세포 요법)에서 표적화하려는 항원은 네오항원이다.

[1636] 일정 구현예예서, 질환 (예컨대 특히 종양 또는 암)의 양자 세포 요법 (예컨대 특히 CAR 또는 TCR T-세포 요법)에서 표적화하려는 항원은 종양-연관 항원 (TAA)이다.

[1637] 일정 구현예예서, 질환 (예컨대 특히 종양 또는 암)의 양자 세포 요법 (예컨대 특히 CAR 또는 TCR T-세포 요법)에서 표적화하려는 항원은 유니버설 종양 항원이다. 일부 바람직한 구현예에서, 유니버설 종양 항원은 인간 텔로머라제 역전사효소 (hTERT), 서비빈, 마우스 이중 미니트 2 상동체 (MDM2), 시토크롬 P450 1B 1 (CYP1B), HER2/neu, 빌름스 종양 유전자 1 (WT1), 리빈, 알파펙토단백질 (AFP), 암배 상원 (CEA), 뮤신 16 (MUC16), MUC1, 전립선-특이적 막 항원 (PSMA), p53 또는 사이클린 (Di), 및 이의 임의 조합으로 이루어진 군으로부터 선택된다.

[1638] 일정 구현예에서, 질환 (예컨대 특히 종양 또는 암)의 양자 세포 요법 (예컨대 특히 CAR 또는 TCR T-세포 요법)에서 표적화하려는 항원 (예컨대 종양 항원)은 CD19, BCMA, CD70, CLL-1, MAGE A3, MAGE A6, HPV E6, HPV E7, WT1, CD22, CD171, ROR1, MUC16, 및 SSX2로 이루어진 군으로부터 선택될 수 있다. 일정한 바람직한 구현예에서, 항원은 CD19일 수 있다. 예를 들어, CD19는 혈액학적 악성종, 예컨대 림프종, 보다 특히 B-세포 림프종, 예컨대 제한없이 미만성 거대 B-세포 림프종, 원발성 종격동 B-세포 림프종, 형질전환 여포성 림프종, 변연부 림프종, 맨틀 세포 림프종, 성인 및 소아 ALL을 포함한 급성 림프아구성 백혈병, 비호지킨 림프종, 무통성 비호지킨 림프종, 또는 만성 림프구성 백혈병에서 표적화될 수 있다. 예를 들어, BCMA는 다발성 골수종 또는 형질 세포 백혈병에서 표적화될 수 있다 (참조: 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic Chimeric Antigen Receptor T Cells Targeting B Cell Maturation 항원).

예를 들어, CLL1은 급성 골수성 백혈병에서 표적화될 수 있다. 예를 들어, MAGE A3, MAGE A6, SSX2, 및/또는 KRAS는 고형 종양에서 표적화될 수 있다. 예를 들어, HPV E6 및/또는 HPV E7은 자궁경부암 또는 두경부암에서 표적화될 수 있다. 예를 들어, WT1은 급성 골수성 백혈병 (AML), 골수이형성 증후군 (MDS), 만성 골수성 백혈병 (CML), 비소세포 폐암, 유방, 췌장, 난소 또는 직결장 암, 또는 중피종에서 표적화될 수 있다.

예를 들어, CD22는 비호지킨 림프종, 미만성 거대 B-세포 림프종, 또는 급성 림프아구성 백혈병을 포함한, B 세포 악성종에서 표적화될 수 있다. 예를 들어, CD171은 신경모세포종, 교모세포종, 또는 폐, 췌장, 또는 난소 암에서 표적화될 수 있다. 예를 들어, ROR1은 비소세포 폐암, 삼중 음성 유방암, 췌장암, 전립선암, ALL, 만성 림프구성 백혈병, 또는 맨틀 세포 림프종을 포함한, ROR1+ 악성종에서 표적화될 수 있다. 예를 들어, MUC16은 MUC16ecto+ 상피 난소, 나팔관 또는 원발성 복막 암에서 표적화될 수 있다. 예를 들어, CD70은 혈액학적 악성종뿐만 아니라 고형암 예컨대 신장 세포 암종 (RCC), 신경교종 (예를 들어, GBM), 및 두경부암 (HNSCC)에서 표적화될 수 있다. CD70은 혈액학적 종양과 고형암 둘 모두에서 발현되지만, 정상 조직에서 이의 발현은 림프구 세포 유형의 서브세트에 제한된다 (참조: 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic CRISPR Engineered Anti-CD70 CAR-T Cells Demonstrate Potent Preclinical Activity Against Both Solid and Hematological Cancer Cells).

[1639] 일부 구현예에서, 표적 항원은 바이러스 항원이다. 많은 바이러스 항원 표적은 HIV, HTLV 및 다른 바이러스의 바이러스 게놈으로부터 유래된 펩티드를 포함한 것들이 동정되고 공지되어 있다 (참조: Addo et al. (2007) PLoS ONE, 2, e321; Tsomides et al. (1994) J Exp Med, 180, 1283-93; Utz et al. (1996) J Virol, 70, 843-51). 예시적인 바이러스 항원은 제한없이 A형 간염 바이러스, B형 간염 바이러스 (예를 들어, HBV 코어 및

표면 항원 (HBVc, HBVs)), C형 간염 바이러스 (HCV), 엡스테인-바 바이러스 (예를 들어, EBVA), 인간 파필로마 마이러스 (HPV; 예를 들어 E6 및 E7), 인간 면역결핍 1형 바이러스 (HIV1), 카포시 육종 헤르페스 바이러스 (KSHV), 인간 파필로마 바이러스 (HPV), 인플루엔자 바이러스, 라싸 바이러스, HTLN-i, HIN-1, HIN-IL CMN, EBN 또는 HPN을 포함한다. 일부 구현예에서, 표적 단백질은 박테리아 항원 또는 다른 병원성 항원, 예컨대 마이코박테리움 튜버큘로시스 (Mycobacterium tuberculosis) (MT) 항원, 트리파노솜 (trypanosome), 예를 들어, 티판소마 크루지 (Tiypansoma cruzi) (T. cruzi), 항원예컨대 표면 항원 (TSA), 또는 말라리아 항원이다. 특이적 바이러스 항원 또는 에피토프 또는 다른 병원성 항원 또는 펩티드 에피토프는 공지되어 있다 (참조: Addo et al. (2007) PLoS ONE, 2, e321; Anikeeva et al. (2009) Clin Immunol, 130, 98-109). 일부 구현예에서, 항원은 암과 연관된 바이러스, 예컨대 종양원성 바이러스로부터 유래된 항원이다. 예를 들어, 종양원성 바이러스는 일정 바이로스로부터의 감염이 상이한 유형의 암의 발생을 초래하는 것, 예를 들어, A형 간염 바이러스, B형 간염 바이러스 (HBV), C형 간염 바이러스 (HCV), 인간 파필로마 바이러스 (HPV), 간염 바이러스 감염, 엡스타인-바 바이러스 (EBV), 인간 헤르페스 바이러스 8 (HHV-8), 인간 T-세포 백혈병 바이러스 -1 (HTLV- 1), 인간 T-세포 백혈병 바이러스-2 (HTLV-2), 또는 사이토메갈로바이러스 (CMV) 항원이다. 일부 구현예에서, 바이러스 항원은 일부 경우에 자궁경부 및/또는 두경부 암을 발생시킬 더 큰 위험성을 초래하는 HPV 항원이다. 일부 구현예에서, 항원은 HPV-16 항원, 및 HPV-18 항원, 및 HPV-31 항원, HPV-33 항원 또는 HPV-35 항원이다. 일부 구현예에서, 바이러스 항원은 HPV-16 항원 (예를 들어, HPV-16 의 E1, E2, E6 또는 E7 단백질의 혈청반응성 영역, 예를 들어, U.S. Pat. No. 6,531, 127 참고) 또는 HPV-18 항원 (예를 들어, U.S. Pat. No. 5,840,306 에 기재된 바와 같은, HPV-18 의 LI 및/또는 L2 단백질의 혈청반응성 영역) 이다.

[1640] 일부 구현예에서, 바이러스 항원은 일부 경우에서, HBV 또는 HCV 음성 대상체에 비해서 간암이 발생될 더 큰 위험성을 초래할 수 있는 HBV 또는 HCV 항원이다. 예를 들어, 일부 구현예에서, 이종성 항원은 HBV 항원, 예컨대 B형 간염 바이러스 코어 항원 또는 B형 간염 바이러스 엔벨로프 항원이다 (US2012/0308580).

[1641] 일부 구현예에서, 바이러스 항원은 일부 경우에서 EBV 음성 대상체에 비해서 버킷 림프종, 비인두 암종 및 비호지킨 질환을 발생시킬 더 큰 위험성을 초래할 수 있는 EBV 항원이다. 예를 들어, EBV는 인간 헤르페스 바이러스로서, 일부 경우에 다양한 조직 기원의 수많은 인간 종양과 연관된 것으로 확인된다. 주로 무증상성 감염으로 발견되지만, EBV-양성 종양은 바이러스 유전자 산물, 예컨대 EBNA-1, LMP-1 및 LMP-2A의 활발한 발현을 특징으로 할 수 있다. 일부 구현예에서, 이종성 항원은 엡스타인-바 핵 항원 (EBNA)-1, EBNA-2, EBNA-3A, EBNA-3B, EBNA-3C, EBNA-리더 단백질 (EBNA- LP), 잠복 막 단백질 LMP- 1, LMP-2A 및 LMP-2B, EBV-EA, EBV-MA 또는 EBV- VCA를 포함할 수 있는 EBV 항원이다. 일부 구현예에서, 바이러스 항원은 HTLV-1 또는 HTLV-2 음성 대조군에 비해서 T-세포 백혈병을 발생시킬 더 큰 위험성을 초래할 수 있는, HTLV-1 또는 HTLV-2 항원이다. 예를 들어, 일부 구현예에서, 이종성 항원은 HTLV- 항원, 예컨대 TAX이다.

[1642] 일부 구현예에서, 바이러스 항원은 HHV-8 음성 대상체에 비해서 카포시 육종을 발생시킬 더 큰 위험성을 초래할 수 있는, HHV-8 항원이다. 일부 구현예에서, 이종성 항원은 CMV 항원, 예컨대 pp65 또는 pp64 (참조: U.S. 특허 번호 8361473).

[1643] 일부 구현예에서, 바이러스 항원은 바이러스-특이적 표면 항원 예컨대 HIV-특이적 항원 (예컨대 HIV gp120); EBV-특이적 항원, CMV-특이적 항원, HPV-특이적 항원, 라싸 바이러스-특이적 항원, 인플루엔자 바이러스-특이적 항원을 비롯하여 이들 표면 마커의 임의 유도체 또는 변이체이다.

[1644] 일 양상에서, 본 발명은 특히 1형 신경섬유종증 (NF1) 신경유전 병태에 의해 유도되는 중추신경계의 종양의 치료를 제공한다. NF1을 갖는 개체는 NF1 유전자에 배선 돌연변이를 갖고 태어나지만, 자폐증 및 주의력 결핍부터 뇌 및 말초 신경초 종양의 범위에 걸쳐서, 수많은 개별 신경학적 문제가 발생될 수 있다. 본 발명은 환자-특이적 질환 모델을 개발하고 동질유전자 배경에서 유도 만능 줄기 세포 (iPSC)-유래 질환 관련 세포를 연구하기 위해 사용될 수 있다. 유도 만능 줄기 세포 또는 iPSC로도 알려진, 배아 줄기 세포 (ESC)-유사 세포는 성인 환자의 피부 또는 혈액 세포로부터 생성될 수 있다. 최근의 연구 노력은 NF1 환자에 영향을 미치는 중추 및 말초 신경계 (CNS 및 PNS)에서 다양한 세포 유형으로 iPSC를 분화시키는 배양 프로토콜을 개발하기 위해 시작되었다. 본 발명의 CRISPR C2c1 시스템은 현존 돌연변이체 유전자를 복구하거나 또는 새로운 돌연변이를 창출하여 특이적 질환 유전자를 유전자 편집하는데 사용될 수 있다. NF1 연구의 최전선에 자리잡기 위해서, 국립 아동 의료 센터 (Children's National Medical Center)의 길버트 가족 신경섬유종증 연구소 (Gilbert Family Neurofibromatosis Institute) (GFNI)는 이들 최근의 흥미로운 연구 개발을 조사하고, 환자-특이적 인간 NF1 질환 모델을 체계적으로 개발하고, 개별 NF 환자에 대한 약물 스크리닝 및 평가를 위한 도구를

제공하는 것이 중요할 것이다.

[1645] 앞서 언급한 것과 같은 접근은 선택된 항원에 결합하는 수용체를 인식하는 항원을 포함하는 유효량의 면역반응성 세포를 투여함으로써, 질환, 예컨대, 신생물을 갖는 대상체를 치료하고/하거나 생존을 증가시키는 방법을 제공하는 데 적합할 수 있으며, 여기서 결합은 면역반응성 세포를 활성화시켜서, 질환 (예컨대, 신생물, 병원균 감염, 자가면역 장애 또는 동종이계 이식 반응)을 치료하거나 또는 예방한다. CAR T 세포 요법에서 용량은, 예를 들어, 림프구고갈 과정의 존재 또는 부재로, 예를 들어 시클로포스파미드 존재에서, 106 내지 109개의 세포/kg의 투여를 수반할 수 있다.

[1646] 당업자는 상기 기술된 바와 유사한 시스템에서 본 발명에 개시된 CRISPR-C2c1 시스템일 수 있다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부 서열인 PAM 서열을 인식할 수 있다. 일부 구현예에서, PAM 서열은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손 (DSB)을 표적 유전자에 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 주형 DNA 서열을 도입시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 유전자를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 표적 유전자의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1647] 일 구현예에서, 치료는 면역억제 치료를 겪는 환자에게 투여될 수 있다. 세포 또는 세포 개체군은 이러한 면역억제제에 대한 수용체를 코딩하는 유전자의 불활성화에 기인하여 적어도 하나의 면역억제제에 대해 내성될 수 있다. 이론에 국한하지 않지만, 면역억제 치료는 환자 내에서 본 발명에 따른 면역억제 또는 T 세포의 선택 및 확장에 도움을 주어야 한다.

[1648] 본 발명에 따른 세포 또는 세포 집단의 투여는 에어로졸 흡입, 주사, 섭취, 수혈, 이식 주입 또는 이식을 비롯한 임의의 편리한 방법으로 수행될 수 있다. 세포 또는 세포 개체군은 환자에게 피하로, 피내로, 종양내로, 낭내로, 척수내로, 근육내로, 정맥내로 또는 림프내 주사로 또는 복강내로 투여될 수 있다. 일 구현예에서, 본 발명의 세포 조성물은 바람직하게는 정맥내 주사에 의해 투여된다.

[1649] 세포 또는 세포 개체군의 투여는 104 내지 109 세포/kg 체중, 바람직하게는 105 내지 106 세포/kg 체중 (해당 범위 내에서 세포의 모든 정수 값을 포함함)의 투여로 이루어질 수 있다. CAR T 세포 요법에서 용량은, 예를 들어, 림프구고갈 과정의 존재 또는 부재로, 예를 들어 시클로포스파미드 존재에서, 106 내지 109개의 세포/kg의 투여를 수반할 수 있다. 세포 또는 세포 개체군은 1회 이상의 용량으로 투여될 수 있다. 다른 구현예에서, 유효량의 세포는 단일 용량으로서 투여된다. 다른 구현예에서, 유효량의 세포는 일정 기간에 걸쳐 1회 초과로서 투여된다. 투여 시기는 담당 의사의 판단 내이고, 환자의 임상적 병태에 의존한다. 세포 또는 세포 개체군은 임의의 공급원, 예컨대, 혈액 은행 또는 공혈자로부터 얻을 수 있다. 각각의 요구가 다르지만, 특정 질환 또는 병태에 대해 유효량의 주어진 세포 유형의 최적 범위의 결정은 당업자의 기술 내이다. 유효량은 치료적 또는 예방적 이득을 제공하는 양을 의미한다. 투여되는 용량은 연령, 수용자의 건강상태 및 체중, 있다면 병용 치료의 종류, 치료 빈도 및 목적하는 효과의 특성에 따를 것이다.

[1650] 다른 구현예에서, 해당 세포를 포함하는 유효량의 세포 또는 조성물은 비경구로 투여된다. 투여는 정맥내 투여일 수 있다. 투여는 종양 내에서 주사에 의해 직접적으로 행해질 수 있다.

[1651] 가능한 유해 반응에 대해 보호하기 위해, 조작된 면역반응성 세포는 특정 신호에 대한 노출에 취약한 세포를 제공하는 이식유전자의 형태로 유전자이식 안전성 스위치를 구비할 수 있다. 예를 들어, 단순 포진 바이러스 티미딘 키나제(TK) 유전자가 이 방법에서, 줄기 세포 이식 후 공여자 림프구 주입으로서 사용되는 동종이계 T 림프구 내로 도입에 의해 사용될 수 있다 (Greco, et al., Improving the safety of cell therapy with the TK-suicide gene. Front. Pharmacol. 2015; 6: 95). 이러한 세포에서, 뉴클레오시드 프로드러그, 예컨대, 간시클로비어 또는 아시클로비어의 투여는 세포 사멸을 야기한다. 대안의 안전성 스위치 구성체는, 예를 들어 활성 효소를 형성하는 2개의 비기능성 icasp9 분자와 함께 소형 분자 이량체의 투여에 의해 촉발되는 유도성 카파제 9를 포함한다. 세포 증식 제어를 구현하기 위한 광범위하게 다양한 대안이 기술되었다 (참조: U.S. 특허 출원 공개 번호 20130071414; PCT 특허 공개 WO2011146862; PCT 특허 공개 WO2014011987; PCT 특허 공개 WO2013040371; Zhou et al. BLOOD, 2014, 123/25:3895 - 3905; Di Stasi et al., The New England Journal of Medicine 2011; 365:1673-1683; Sadelain M, The New England Journal of Medicine 2011; 365:1735-173; Ramos et al., Stem Cells 28(6):1107-15 (2010)).

[1652] 양자 요법의 추가적인 개선에서, 본 명세서에 기재된 바와 같은 기능화된 CRISPR-Cas 시스템에 의한 게놈 편집은, 예를 들어 편집 CAR T 세포를 제공하는 대안적인 실행을 위해 면역반응성 세포를 재단하는데 사용될 수 있다 (Poirot et al., 2015, Multiplex genome edited T-cell manufacturing platform for " off-the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853). 예를 들어, 면역반응성 세포는 HLA II형 및/또는 I형 분자 부류의 일부 또는 모두의 발현을 결실시키도록 또는 목적하는 면역 반응, 예컨대, PD1 유전자를 억제할 수 있는 선택 유전자를 녹아웃시키도록 편집될 수 있다.

[1653] 세포는 본 명세서에 기술된 바와 같은 임의의 CRISPR 시스템 및 이의 사용 방법을 사용해 편집될 수 있다. CRISPR 시스템은 임의의 본 명세서에 기재된 방법에 의해 면역 세포에 전달될 수 있다. 바람직한 구현예에서, 세포는 생체외에서 편집되고, 이를 필요로 하는 대상체에게 전달된다. 면역 반응성 세포 세포, CAR-T 세포 또는 양자 세포 전달을 위해 사용되는 임의의 세포는 편집될 수 있다. 편집은 잠재적인 동종반응성 T-세포 수용체 (TCR)를 제거하기 위해, 화학치료제의 표적을 붕괴시키기 위해, 면역 체크포인트를 차단하기 위해, T 세포를 활성화시키기 위해 및/또는 기능적으로 감손되거나 또는 기능장애 CD8+ T-세포의 분화 및/또는 증식을 증가시키기 위해 수행될 수 있다 (국제 특허 출원 공개: WO2013176915, WO2014059173, WO2014172606, WO2014184744 및 WO2014191128). 편집은 유전자의 불활성화를 초래할 수 있다.

[1654] 유전자를 불활성화시켜서, 관심 유전자가 기능성 단백질 형태로 발현되지 않게 의도한다. 특정 구현예에서, CRISPR 시스템은 특히 하나의 표적화된 유전자에서 절단을 촉매하여 상기 표적화된 유전자를 불활성화시킨다. 초래된 핵산 가닥 파손은 일반적으로 상동성 재조합 또는 비상동성 말단 연결 (NHEJ)의 별개 기전을 통해 복구된다. 그러나, NHEJ는 종종 절단 부위에서 DNA 서열에 변화를 초래하는 불완전한 복구 과정이다. 비상동성 말단 연결 (NHEJ)를 통한 복구가 종종 작은 삽입 또는 결실 (Indel)을 야기시키고 특이적 유전자 녹아웃의 생성에 사용될 수 있다. 절단 유도된 돌연변이유발 사건이 발생된 세포는 당분야에 충분히 공지된 방법으로 확인 및/또는 선택할 수 있다.

[1655] T 세포 수용체(TCR)는 항원 제시에 반응하는 T 세포의 활성화에 참여하는 세포 표면 수용체이다. TCR은 일반적으로, 이형이량체를 형성하도록 조립되고 세포 표면 상에 제시되는 T 세포 수용체 복합체를 형성하기 위해 CD3-형질도입 서브유닛과 회합하는 2개의 쇄, 즉, α 및 β로 이루어진다. TCR의 각각의 α 및 β 쇄는 면역글로불린-유사 N-말단의 가변 (V) 및 불변 (C) 영역, 소수성 막관통 도메인 및 짧은 세포질 영역으로 이루어진다. 면역글로불린 분자에 대해서와 같이, α 및 β 쇄의 가변 영역은 V(D)J 재조합에 의해 생성되어, T 세포 개체군 내에서 매우 다양한 항원 특이성을 생성시킨다. 그러나, 온전한 항원을 인식하는 면역글로불린과 대조적으로, T 세포는 MHC 분자와 회합되는 가공된 펩티드 단편에 의해 활성화되어, MHC 제한으로서 알려진 T 세포에 의한 항원 인식에 대해 여분의 차원을 도입한다. T 세포 수용체를 통한 도너와 수용자 사이의 MHC의 인식 차이는 T 세포 증식 및 이식편 대 숙주 반응 (GVHD)의 잠재적 발생을 야기한다. TCR α 또는 TCR β의 불활성화는 T 세포 표면으로부터 TCR의 제거를 초래하여 동종항원의 인식을 방지하고 그에 따라 GVHD를 방지할 수 있다. 그러나, TCR 붕괴는 일반적으로 CD3 신호전달 성분의 제거를 초래하고 추가적인 T 세포 확장 수단을 변경시킨다.

[1656] 동종이계 세포는 숙주 면역계에 의해 빠르게 거부된다. 비-방사선조사 혈액 생성물에 존재하는 동종이계 백혈구는 5 내지 6일 이하 동안 지속된다는 것이 입증되었다 (Boni, Muranski et al. 2008 Blood 1;112(12):4746- 54). 따라서, 동종이계 세포의 거부를 방지하기 위해, 숙주의 면역계는 일반적으로 어느 정도까지는 억제되어야 한다. 그러나, 양자 세포 전달의 경우에, 면역억제 약물의 사용이 또한 도입된 치료적 T 세포에 대해 해로운 효과를 갖는다. 그러므로, 이들 조건에서 양자 면역요법 접근을 효과적으로 사용하기 위해, 도입된 세포는 면역억제 치료에 내성일 필요가 있을 것이다. 따라서, 특정 구현예에서, 본 발명은, 바람직하게는 면역억제제에 대한 표적을 코딩하는 적어도 하나의 유전자의 불활성화에 의해, T 세포를 면역억제제에 내성이 되도록 변형시키는 단계를 추가로 포함한다. 면역억제제는 몇몇 작용 메커니즘 중 하나에 의해 면역 기능을 억제하는 작용제이다. 면역억제제는 칼시뉴린 억제제, 라파마이신의 표적, 인터루킨-2 수용체 α-쇄 차단제, 이노신 일인산 디히드로게나제의 억제제, 다이하이드로엽산 리덕타제의 억제제, 코티코스테로이드 또는 면역억제성 항대사물질일 수 있지만, 이들로 제한되지 않는다. 본 발명은 T 세포 내 면역억제제의 표적을 불활성화시킴으로써 면역요법을 위해 T 세포에 면역억제 내성을 부여하도록 허용한다. 비제한적 예로서, 면역억제제에 대한 표적은 면역억제제, 예컨대: CD52, 글루코코티코이드 수용체(GR), FKBP 패밀리 유전자 멤버 및 시클로필린 패밀리 유전자 멤버에 대한 수용체일 수 있다.

[1657] 면역 체크포인트는 면역반응을 늦추거나 중단시키고, 면역 세포의 제어되지 않은 활성으로부터의 과도한 조직 손상을 방지하는 억제 경로이다. 일정 구현예에서, 표적화된 면역 체크포인트는 프로그램된 사멸-1 (PD-1

또는 CD279) 유전자 (PDCD1)이다. 다른 구현예에서, 표적화된 면역 체크포인트는 세포독성 T-림프구-연관 항원 (CTLA-4)이다. 추가의 구현예에서, 표적화된 면역 체크포인트는 CD28 및 CTLA4 Ig 수퍼패밀리의 다른 멤버 예컨대 BTLA, LAG3, ICOS, PDL1 또는 KIR이다. 더욱 추가의 구현예에서, 표적화된 면역 체크포인트는 TNFR 수퍼패밀리의 멤버 예컨대 CD40, OX40, CD137, GITR, CD27 또는 TIM-3 이다.

[1658] 추가적인 면역 체크포인트는 Src 상동성 2 도메인-함유 단백질 티로신 포스파타제 1 (SHP-1)을 포함한다 (Watson HA, et al., SHP-1: the next checkpoint target for cancer immunotherapy? Biochem Soc Trans. 2016 Apr 15;44(2):356- 62). SHP-1은 널리 발현되는 억제성 단백질 티로신 포스파타제 (PTP)이다. T-세포에서, 이는 항원-의존적 활성화 및 증식의 음성 조절자이다. 이는 시토졸 단백질이고, 따라서 항체-매개 요법을 받을 수 없지만, 활성화 및 증식에서 이의 역할은 양자 전달 전략, 예컨대, 키메라 항원 수용체 (CAR) T 세포에서 유전자 조작을 위한 매력적인 표적이 되게 한다. 면역관문은 또한 Ig 및 ITIM 도메인 (TIGIT/Vstm3/WUCAM/VSIG9) 및 VISTA를 갖는 T 세포 면역수용체를 포함할 수 있다 (Le Mercier I, et al., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. Front. Immunol. 6:418).

[1659] WO2014172606은 고갈된 CD8+ T-세포의 증식 및/또는 활성을 증가시키기 위해서, 그리고 CD8+ T-세포 고갈을 감소 (예를 들어, 기능적으로 고갈되거나 또는 비반응성인 CD8+ 면역 세포를 감소)시키기 위해 MT1 및/또는 MT1 억제제의 사용에 관한 것이다. 일정 구현예에서, 메탈로티오네인은 양자 전달된 T 세포 내 유전자 편집에 의해 표적화된다.

[1660] 일정 구현예에서, 유전자 편집의 표적은 면역 체크포인트 단백질의 발현에 관여되는 적어도 하나의 표적화된 유전자좌일 수 있다. 이러한 표적은 CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBRII, TGFRBRI, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1, FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1, TIM-3, CEACAM-1, CEACAM-3, 또는 CEACAM-5을 포함할 수 있지만, 이들로 제한되지 않는다. 바람직한 구현예에서, PD-1 또는 CTLA-4 유전자의 발현에 관여되는 유전자의 유전자좌가 표적화된다. 다른 바람직한 구현예에서, 유전자의 조합은, 예컨대, PD-1 및 TIGIT로 표적화되지만, 이들로 제한되지 않는다.

[1661] 다른 구현예에서, 적어도 2개의 유전자가 편집된다. 유전자의 쌍은 PD1과 TCRα, PD1과 TCRβ, CTLA-4와 TCRα, CTLA-4와 TCRβ, LAG3과 TCRα, LAG3과 TCRβ, Tim3과 TCRα, Tim3과 TCRβ, BTLA와 TCRα, BTLA와 TCRβ, BY55와 TCRα, BY55와 TCRβ, TIGIT와 TCRα, TIGIT와 TCRβ, B7H5와 TCRα, B7H5와 TCRβ, LAIR1과 TCRα, LAIR1과 TCRβ, SIGLEC10과 TCRα, SIGLEC10과 TCRβ, 2B4와 TCRα, 2B4와 TCRβ를 포함할 수 있지만, 이들로 제한되지 않는다.

[1662] T 세포의 유전자 변형 이전이건 또는 이후이건 간에, T 세포는 일반적으로 예를 들어, 미국 특허 제6,352,694호; 제6,534,055호; 제6,905,680호; 제5,858,358호; 제6,887,466호; 제6,905,681호; 제7,144,575호; 제7,232,566호; 제7,175,843; 5,883,223호; 제6,905,874호; 제6,797,514호; 제6,867,041호; 및 제7,572,631호에 기재된 바와 같은 방법을 이용하여 활성화 및 확장될 수 있다. T 세포는 시험관내 또는 생체내에서 확장될 수 있다.

[1663] 본 발명의 실시는 달리 나타내지 않는 한, 당업계의 기술 내에 있는 면역학, 생화학, 화학, 분자 생물학, 미생물학, 세포 생물학, 유전체학 및 재조합 DNA의 통상의 기술을 사용한다. 참조: MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989) (Sambrook, Fritsch and Maniatis); MOLECULAR CLONING: A LABORATORY MANUAL, 4th edition (2012) (Green and Sambrook); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1987) (F. M. Ausubel, et al. eds.); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (1995) (M.J. MacPherson, B.D. Hames and G.R. Taylor eds.); ANTIBODIES, A LABORATORY MANUAL (1988) (Harlow and Lane, eds.); ANTIBODIES A LABORATORY MANUAL, 2nd edition (2013) (E.A. Greenfield ed.); 및 ANIMAL CELL CULTURE (1987) (R.I. Freshney, ed.).

[1664] 본 발명의 실시는 달리 표시하지 않으면 유전자 변형 마우스의 생성을 위해 통상의 기술을 적용한다. 참조: Marten H. Hofker and Jan van Deursen, TRANSGENIC MOUSE METHODS AND PROTOCOLS, 2nd edition (2011).

[1665] **CRISPR 시스템을 이용하는 스크리닝/진단/치료**

[1666] **암**

[1667] 본 발명의 방법 및 조성물은 세포의 약물 내성 및 지속성과 연관된 세포 상태, 성분 및 메커니즘을 확인하는 데 사용될 수 있다. Terai 등 (Cancer Research, 19-Dec-2017, doi: 10.1158/0008- 5472.CAN-17-1904)]은 에를로티닙/THZ1 상승효과가 향상된 다중 유전자뿐만 아니라 상승효과를 억제하는 성분 및 경로를 확인하기 위해 에를로티닙 + THZ1(CDK7/12 저해제) 병용 요법으로 처리된 EGFR-의존적 폐암 PC9 세포에서의 게놈-와이드 CRISPR/Cas9 인핸서/억제자 선별을 보고하였다. Wang 등 (Cell Rep. 2017 Feb 7;18(6):1543- 1557. doi: 10.1016/j.celrep.2017.01.031.; Krall et al., Elife. 2017 Feb 1;6. pii: e18970. doi: 10.7554/eLife.18970)]은 MAPK 저해제에 대한 내성의 매개체를 확인하기 위한 게놈-와이드 CRISPR 기능 상실 선별의 사용을 보고하였다. Donovan 등 (PLoS One. 2017 Jan 24;12(1):e0170445. doi: 10.1371/journal.pone.0170445. eCollection 2017)]은 MAPK 신호전달 경로 유전자의 신규한 기능획득 및 약물 내성 대립유전자를 동정하기 위해 돌연변이유발에 대한 CRISPR-매개 접근을 사용하였다. Wang 등 (Cell. 2017 Feb 23;168(5):890-903.e15. doi: 10.1016/j.cell.2017.01.013. Epub 2017 Feb 2)은 종양유전자 Ras와 합성의 치명적 상호작용 및 유전자 네트워크를 확인하기 위해 게놈-와이드 CRISPR 스크린을 사용하였다. Chow 등 (Nat Neurosci. 2017 Oct;20(10):1329- 1341. doi: 10.1038/nn.4620. Epub 2017 Aug 14)]은 교아세포종에서 기능성 억제자를 확인하기 위해 교아세포종에서 아데노-연관 바이러스-매개, 자생종 유전자 CRISPR 선별을 개발하였다. Xue 등 (Nature. 2014 Oct 16;514(7522):380-4. doi: 10.1038/nature13589. Epub 2014 Aug 6)은 마우스 간에서 암의 유전자의 CRISPR-매개 직접 돌연변이를 적용하였다.

[1668] Chen 등 (J Clin Invest. 2017 Dec 4. pii: 90793. doi: 10.1172/JCI90793. [Epub ahead of print])은 EZH2에 대한 MYCN-증폭된 신경아세포종 의존성을 확인하기 위해 CRISPR-기반 스크린을 사용하였다. MYCN-증식 신경아세포종을 갖는 환자에서 EZH2 억제제의 시험을 뒷받침한다.

[1669] Vijai 등 (Cancer Discov. 2016 Nov;6(11):1267- 1275. Epub 2016 Sep 21)]은 유방암의 위험을 평가하기 위해 유방 상피 세포주에서 이형접합적 돌연변이를 생성하는 CRISPR의 사용을 보고하였다.

[1670] Chakraborty 등 (Sci Transl Med. 2017 Jul 12;9(398). pii: eaal5272. doi: 10.1126/scitranslmed.aal5272)]은 투명 세포 신세포 암종을 치료하기 위한 잠재적 표적으로서 EZH1을 동정하기 위해 CRISPR-기반 선별을 사용하였다.

[1671] **대사 질환**

[1672] 본 발명의 방법 및 조성물은 가족성 과콜레스테롤혈증, 혈우병, 오르니틴 트랜스 카비미라제 결핍증, 유전성 티로신혈증 1형, 및 알파-1 항트립신 결핍증을 포함하지만, 이들로 제한되지 않는 간의 유전성 대사 질환 치료에서 통상적인 유전자 요법 이상의 이점을 제공한다. 참조: Bryson et al., Yale J. Biol. Med. 90(4):553-566, 19-Dec-2017.

[1673] Bompada 등 (Int J Biochem Cell Biol. 2016 Dec;81(Pt A):82- 91. doi: 10.1016/j.biocel.2016.10.022. Epub 2016 Oct 29)]은 히스톤 아세틸화가 TXNIP 유전자 발현에서 글루코스-유도 증가 및 이에 의한 당독성-유도 아폽토시스의 중요한 조절자로서 작용한다는 것을 입증하기 위해 췌장 베타 세포에서 히스톤 아세틸트랜스퍼라제를 녹아웃시키기 위한 CRISPR의 사용을 기재하였다.

[1674] **눈**

[1675] 본 발명은 망막의 유전성 및 후천성 안질환의 효율적인 치료를 제공한다. Holmgaard 등 (Mol. Ther. Nucleic Acids 9:89- 99, 15-Dec-2017 doi:10.1016/j.omtn.2017.08.016. Epub 2017 Sep 21)]은 Vegfa에 표적화된 SpCas9를 암호화하는 렌티바이러스 벡터(LV)에 의해 SpCas9가 전달될 때 고빈도의 삽입결실과, 형질도입 세포에서 VEGFA의 상당한 감소가 있었다는 것을 보고하였다. Duan 등 (J Biol Chem. 2016 Jul 29;291(31):16339-47. doi: 10.1074/jbc.M116.729467. Epub 2016 May 31)은 인간 초대 망막 색소 상피 세포에서 MDM2 게놈 유전자좌를 표적으로 하기 위한 CRISPR의 용도를 기술한다.

[1676] 본 발명의 방법 및 조성물은 노인성 황반 변성을 포함하는 안질환의 치료에 유사하게 적용 가능하다.

[1677] Huang 등 (Nat Commun. 2017 Jul 24;8(1):112. doi:10.1038/s41467-017-00140-3]은 혈관신생-연관 질환을 치료하기 위해 VEGFR2를 편집하는 데 CRISPR를 사용하였다.

[1678] **청력**

[1679] Gao 등 (Nature. 2017 Dec 20. doi:10.1038/nature25164. [Epub ahead of print])]은 마우스에서 Tmc1 유전자를 표적화하고 진행성 청력상실 및 귀먹음을 감소시키기 위해 CRISPR-Cas9를 이용하는 게놈 편집을 보고하였다.

[1680] **근육**

[1681] Provenzano 등 (Mol Ther Nucleic Acids. 9:337- 348. 15-Dec-2017;. doi: 10.1016/j.omtn.2017.10.006. Epub 2017 Oct 14)은 근긴장성 이영양증 1 환자로부터의 근원성 세포에서 CTG 확장 및 정상 표현형으로의 영구한 전환의 CRISPR/Cas9-매개 결실을 보고하였다. 본 발명의 방법 및 조성물은 CTG 확장으로 제한되지 않는 뉴클레오티드 반복 장애에 유사하게 적용 가능하다. Tabebordbar 등 (2016 Jan 22;351(6271):407- 411. doi: 10.1126/science.aad5177. Epub 2015 Dec 31)]은 DMD에서 파괴적 돌연변이를 보정하기 위해 Dmd 엑손 23 좌위를 편집하는 CRISPR의 사용을 보고한다. Tabebordbar는 프로그램 가능한 CRISPR 복합체가 신생 및 성체 마우스에서 말단으로 분화된 골격근 섬유 및 심장 근육 세포뿐만 아니라 근위성 세포에 국소로 그리고 전신으로 전달될 수 있으며, 그들이 표적화된 유전자 변형을 매개하는 경우에, 디스트로핀 발현을 회복시키고, 이영양성 근육의 기능성 결핍증을 부분적으로 회복시킬 수 있다는 것을 나타낸다. 참조: Nelson et al., (Science. 2016 Jan 22;351(6271):403-7. doi: 10.1126/science.aad5143. Epub 2015 Dec 31).

[1682] **감염성 질환**

[1683] Sidik 등 (Cell. 2016 Sep 8;166(6):1423-1435.e12. doi: 10.1016/j.cell.2016.08.019. Epub 2016 Sep 2) 및 Patel 등 (Nature. 2017 Aug 31;548(7669):537-542. doi: 10.1038/nature23477. Epub 2017 Aug 7)은 톡소플라스마에서 CRISPR 스크린 및 항기생충 중재술의 확장을 기술한다.

[1684] 숙주-병원균 상호작용의 근간을 이루는 성분 및 과정을 확인하기 위한 게놈-와이드 CRISPR 스크리닝의 몇몇 보고가 있다. 예에는 다음의 문헌들이 포함된다: Blondel et al. (Cell Host Microbe. 2016 Aug 10;20(2):226-37. doi: 10.1016/j.chom.2016.06.010. Epub 2016 Jul 21), Shapiro et al. (Nat Microbiol. 2018 Jan;3(1):73-82. doi: 10.1038/s41564-017-0043-0. Epub 2017 Oct 23) 및 Park et al. (Nat Genet. 2017 Feb;49(2):193-203. doi: 10.1038/ng.3741. Epub 2016 Dec 19).

[1685] Ma 등 (Cell Host Microbe. 2017 May 10;21(5):580- 591.e7. doi: 10.1016/j.chom.2017.04.005)]은 치료적 개입을 위한 바이러스 형질전환-유도 합성 치사 표적을 확인하기 위해 게놈-와이드 CRISPR 기능 상실 선별을 사용하였다.

[1686] **심혈관 질환**

[1687] CRISPR 시스템은 혈관 질환과 연관된 유전자 또는 유전자 변이체를 동정하기 위한 도구로서 사용될 수 있다. 이는 잠재적인 치료 또는 예방 표적을 동정하는 데 유용하다. Xu 등 (Atherosclerosis, 2017 Sep. 21 pii: S0021-9150(17)31265-0. doi: 10.1016/j.atherosclerosis.2017.08.031. [Epub ahead of print])은 LDL-C의 혈장 수준 조절에서 ANGPTL3의 역할을 확인하기 위해서 ANGPT13을 녹아웃시키기 위한 CRISPR의 용도를 보고한다. Gupta 등 (Cell. 2017 Jul 27;170(3):522-533.e15. doi: 10.1016/j.cell.2017.06.049)은 혈관 질환과 연관된 유전자 변이체를 동정하기 위해 줄기 세포-유래 내피 세포를 편집하기 위한 CRISPR의 용도를 보고한다. Beaudoin 등 (Arterioscler Thromb Vasc Biol. 2015 Jun;35(6):1472-1479. doi: 10.1161/ATVBAHA.115.305534. Epub 2015 Apr 2)은 유전자좌에서 전사 인자 MEF2의 결합을 파괴하기 위한 CRISPR 게놈 편집의 용도를 보고한다. 혈관내피에서 PHACTR1 기능이 관상동맥 질환이 영향을 미치는 방법을 연구하는 단계를 설정한다. Pashos 등 (Cell Stem Cell. 2017 Apr 6;20(4):558- 570.e10. doi: 10.1016/j.stem.2017.03.017.)]은 기능성 변이체 및 지질 기능성 유전자를 동정하기 위해 다능성 줄기 세포 및 간세포-유사 세포를 표적화하는 CRISPR 기술을 이용하는 것을 보고한다.

[1688] 표적을 동정하기 위한 도구로서 사용되는 것에 추가로, CRISPR 시스템은 공지된 표적에 대해 심혈관 질환을 치료하거나 또는 예방하기 위해 직접적으로 사용될 수 있다. Khera 등 (Nat Rev Genet. 2017 Jun;18(6):331-344. doi:10.1038/nrg.2016.160. Epub 2017 Mar 13)]은 새로운 치료의 생물학 연구의 기저를 이루는 원인 위험 인자의 더 양호한 이해를 용이하게 하기 위해 사용되는 관상동맥 위험에 대략 60개의 유전자 좌위를 연결하는 통상적인 변이체 연관 연구를 기재하였다. Khera는, 예를 들어 PCSK9에서의 비활성화 돌연변이가 순환 LDL 콜레스테롤 수준을 감소시키고, PCSK9 억제제 개발에서 강력한 관심을 야기하는 CAD의 위험을 감소시켰다는 것을 설명한다. 추가로, APOC3 또는 LPA에서 보호적 돌연변이를 모방하도록 설계된 안티센스 올리고뉴클레오티드는 트라이글리세라이드 수준의 대략 70% 감소 및 순환 지질단백질(a) 수준의 80% 감소를 각각 입증하였다.

또한, Wang 등 (Arterioscler Thromb Vasc Biol. 2016 May;36(5):783-6. doi: 10.1161/ATVBAHA.116.307227. Epub 2016 Mar 3) and Ding et al. (Circ Res. 2014 Aug 15;115(5):488-92. doi: 10.1161/CIRCRESAHA.115.304351. Epub 2014 Jun 10.)은 심혈관 질환의 예방을 위해 유전자 Pcsk9를 표적화하기 위한 CRISPR의 용도를 보고한다.

[1689] 본 발명은 신경학적 질환 및 장애를 연구하고 치료하기 위한 방법 및 조성물을 제공한다. Nakayama 등 (Am J Hum Genet. 2015 May 7;96(5):709- 19. doi:10.1016/j.ajhg.2015.03.003. Epub 2015 Apr 9)]은 인간 CNS 발생에서 PYCR2의 역할을 연구하기 위한 그리고 소두증 및 수초형성부전증에 대한 잠재적 표적을 확인하기 위한 CRISPR의 사용을 보고한다. Swiech 등 (Nat Biotechnol. 2015 Jan;33(1):102- 6. doi:10.1038/nbt.3055. Epub 2014 Oct 19)]은 생체내에서 성체 마우스 뇌에서의 단일(Mecp2)뿐만 아니라 다중 유전자(Dnmt1, Dnmt3a 및 Dnmt3b)를 표적화하기 위한 CRISPR의 사용을 보고한다. Shin 등 (Hum Mol Genet. 2016 Oct 15;25(20):4566- 4576. doi:10.1093/hmg/ddw286)]은 헌팅톤병 돌연변이를 비활성화시키기 위한 CRISPR의 사용을 보고한다. Platt 등 (Cell Rep. 2017 Apr 11;19(2):335- 350. doi: 10.1016/j.celrep.2017.03.052)]은 자폐 스펙트럼 장애에서 Chd8의 역할을 확인하기 위한 CRISPR 녹인 마우스의 사용을 보고한다. Seo 등 (J Neurosci. 2017 Oct 11;37(41):9917-9924. doi: 10.1523/JNEUROSCI.0621-17.2017. Epub 2017 Sep 14)은 신경퇴행성 장애의 모델을 생성시키기 위한 CRISPR의 용도를 기 술한다. Petersen 등 (Neuron. 2017 Dec 6;96(5):1003- 1012.e7. doi: 10.1016/j.neuron.2017.10.008. Epub 2017 Nov 2)]은 재수초화되지 않는 질환에 대한 잠재적 표적을 동정하기 위해 희소돌기교세포 전구 세포에서의 액티빈 A 수용체 I형의 CRISPR 녹아웃을 입증한다. 본 발명의 방법 및 조성물은 유사하게 적용 가능하다.

[1690] **CRISPR 기술의 다른 적용분야**

[1691] Renneville 등 (Blood. 2015 Oct 15;126(16):1930-9. doi: 10.1182/blood-2015-06-649087. Epub 2015 Aug 28)]은 태아 헤모글로빈 발현에서 EHMT1 및 EMHT2의 역할을 연구하기 위한 그리고 SCD에 대한 신규한 치료적 표적을 동정하기 위한 CRISPR의 사용을 보고한다.

[1692] Tothova 등 (Cell Stem Cell. 2017 Oct 5;21(4):547-555.e8. doi: 10.1016/j.stem.2017.07.015)은 인간 골수성 질환의 모델을 생성시키기 위한 조혈 줄기 및 전구체 세포에서 CRISPR의 용도를 보고한다.

[1693] Giani 등 (Cell Stem Cell. 2016 Jan 7;18(1):73- 78. doi: 10.1016/j.stem.2015.09.015. Epub 2015 Oct 22)]은 인간 다능성 줄기 세포에서 CRISPR/Cas9 게놈 편집에 의한 SH2B3의 비활성화는 보존된 분화에 의한 향상된 적혈구 세포 확장을 허용하였다는 것을 보고한다.

[1694] Wakabayashi 등 (Proc Natl Acad Sci U S A. 2016 Apr 19;113(16):4434-9. doi: 10.1073/pnas.1521754113. Epub 2016 Apr 4)은 인간 적혈구 장애에서 넌코딩 변이체의 병원성을 조사하고 GATA1 전사 활성에 관한 통찰력을 획득하기 위해 CRISPR을 적용하였다.

[1695] Mandal 등 (Cell Stem Cell. 2014 Nov 6;15(5):643- 52. doi: 10.1016/j.stem.2014.10.004. Epub 2014 Nov 6)]은 1차 인간 CD4+ T 세포 및 CD34+ 조혈 줄기 및 전구 세포(HSPC)에서 2가지의 임상적으로 적절한 유전자, 즉, B2M 및 CCR5의 CRISPR/Cas9 표적화를 기재한다.

[1696] Polfus 등 (Am J Hum Genet. 2016 Sep 1;99(3):785. doi: 10.1016/j.ajhg.2016.08.002. Epub 2016 Sep 1)]은 조혈 세포주를 편집하고, 후속적으로 1차 인간 조혈 줄기 및 전구 세포에서 표적화된 녹다운 실험을 하고, 인간 조혈에서 GFI1B 변이체의 역할을 연구하기 위해 CRISPR을 사용하였다.

[1697] Najm 등 (Nat Biotechnol. 2017 Dec 18. doi: 10.1038/nbt.4048. [Epub ahead of print])]은 MAPK 경로 유전자 및 세포자멸사 유전자를 포함하는 다중 세포유형에 걸쳐 합성 치사 및 완충 유전자 쌍을 확인하기 위해 고-복잡성 풀링 이중-녹아웃 라이브러리를 생성하는 이중 표적화를 달성하기 위한 SaCas9와 SpCas9 쌍을 갖는 CRISPR 복합체의 사용을 보고한다.

[1698] Manguso 등 (Nature. 2017 Jul 27;547(7664):413-418. doi: 10.1038/nature23270. Epub 2017 Jul 19.)은 신규한 면역요법 표적의 동정 및/또는 확증을 위한 CRISPR 스크린의 용도를 보고한다. 참조: Roland et al. (Proc Natl Acad Sci USA. 2017 Jun 20;114(25):6581-6586. doi: 10.1073/pnas.1701263114. Epub 2017 Jun 12.); Erb et al. (Nature. 2017 Mar 9;543(7644):270-274. doi: 10.1038/nature21688. Epub 2017 Mar 1.); Hong et al., (Nat Commun. 2016 Jun 22;7:11987. doi: 10.1038/ncomms11987); Fei et al., (Proc Natl Acad Sci USA. 2017 Jun 27;114(26):E5207-E5215. doi: 10.1073/pnas.1617467114. Epub 2017 Jun 13.); Zhang et

al., (Cancer Discov. 2017 Sep 29. doi: 10.1158/2159-8290.CD-17-0532. [Epub ahead of print]).

[1699] Joung 등 (Nature. 2017 Aug 17;548(7667):343-346. doi: 10.1038/nature23451. Epub 2017 Aug 9.)은 긴 넌코딩 RNA (lncRNA)를 분석하기 위해 게놈-와이드 스크리닝의 용도를 보고한다; 참조: Zhu et al., (Nat Biotechnol. 2016 Dec;34(12):1279-1286. doi: 10.1038/nbt.3715. Epub 2016 Oct 31); Sanjana et al., (Science. 2016 Sep 30;353(6307):1545-1549).

[1700] Barrow 등 (Mol Cell. 2016 Oct 6;64(1):163-175. doi: 10.1016/j.molcel.2016.08.023. Epub 2016 Sep 22.)은 미토콘드리아 질환에 대한 치료 표적을 검색하기 위한 게놈-와이드 CRISPR 스크린의 용도를 보고한다. 참조: Vafai et al., (PLoS One. 2016 Sep 13;11(9):e0162686. doi: 10.1371/journal.pone.0162686. eCollection 2016).

[1701] Guo 등 (Elife. 2017 Dec 5;6. pii: e29329. doi: 10.7554/eLife.29329)]은 인간 성장을 위한 생물학적 메커니즘을 설명하기 위해 인간 연골세포를 표적화하는 CRISPR의 사용을 보고한다.

[1702] Ramanan 등 (Sci Rep. 2015 Jun 2;5:10833. doi: 10.1038/srep10833)]은 HBV 게놈에서 보존된 영역을 표적화하고 절단하는 CRISPR의 사용을 보고한다.

[1703] **유전자 구동**

[1704] 본 발명은 예를 들어, PCT 출원 공개 WO 2015/105928에 기술된 유전자 구동과 유사한 시스템에서, RNA-가이드된 유전자 구동을 제공하기 위해서, 본 명세서에 기술된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질 시스템의 용도를 포괄한다. 이러한 종류의 시스템은 예를 들어 RNA-가이드된 DNA 뉴클레아제 및 하나 이상의 가이드 RNA를 코딩하는 핵산 서열을 배선 세포로 도입시켜서, 진핵생물 배선 세포를 변경시키기 위한 방법을 제공할 수 있다. 가이드 RNA는 배선 세포의 게놈 DNA 상의 하나 이상의 표적 위치에 상보성이도록 설계될 수 있다.

RNA 가이드된 DNA 뉴클레아제를 코딩하는 핵산 서열 및 가이드 RNA를 코딩하는 핵산 서열은 측접된 서열 사이에 구성체 상에 제공될 수 있고, 프로모터는 배선 세포가 또한 측접 서열 사이에 위치된 임의의 바람직한 카고-코딩 서열과 함께, RNA 가이드된 DNA 뉴클레아제 및 가이드 RNA를 발현할 수 있도록 배열된다. 측접 서열은 전형적으로 선택된 표적 염색체 상의 상응하는 서열과 동일한 서열을 포함하게 될 것이고, 그래서 측접 서열은 배선 세포를 외래 핵산 서열에 동형접합이게 만들기 위해서, 상동성 재조합같은 기전에 의해 표적 절단 부위에서 게놈 DNA로 외래 핵산 구성체 서열의 삽입을 촉진하도록 구성체에 의해 코딩되는 성분과 함께 작용한다.

이러한 방식으로, 유전자-구동 시스템은 번식 개체군 전체에서 바람직한 카고 유전자를 이입시킬 수 있다 (Gantz et al., 2015, Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi, PNAS 2015, published ahead of print November 23, 2015, doi:10.1073/pnas.1521077112; Esvelt et al., 2014, Concerning RNA-guided gene drives for the alteration of wild populations eLife 2014;3:e03401). 일부 구현예에서, 본 발명은 곤충의 번식 개체군 전체에서 바람직한 카고 융전지를 이입시키는 유전자-구동 시스템에 의해, 말라리아, 지카 바이러스, 웨스트나일 바이러스, 일본 뇌염 바이러스, 및 뎅기 바이러스를 포함한 곤충 매개 질환을 통제하기 위한 방법을 제공한다. 일부 구현예에서, 유전자-구동 시스템은 CRISPR- C2c1 시스템이다. 특정 구현예에서, 곤충은 모기이다. 선택 구현예예서, 표적 서열은 게놈 내 소수의 잠재적 오프-표적 부위를 갖는 것이 선택될 수 있다. 다중 가이드 RNA를 사용하여, 표적 유전자좌 내 다중 부위의 표적화는 절단 빈도를 증가시킬 수 있고, 구동 내성 대립유전자의 진화를 방해할 수 있다. 절두된 가이드 RNA는 오프-표적 절단을 감소시킬 수 있다. 쌍형성 닉카제는 특이성을 더 증가시키기 위해, 단일 뉴클레아제 대신에 사용될 수 있다. 유전자 구동 구성체는 예를 들어 상동성 재조합 유전자를 활성화시키고/시키거나 비상동성 말단-연결을 억제하기 위해, 전사 조절인자를 코딩하는 카고 서열을 포함할 수 있다. 표적 부위는 필수 유전자 내에서 선택되어서, 비상동성 말단-연결 사건이 구동-내성 대립유전자를 생성하는 대신에 치사성을 초래할 수 있다. 유전자 구동 구성체는 다양한 온도 범위에서 다양한 숙주에서 기능하도록 조작될 수 있다 (Cho et al. 2013, Rapid and Tunable Control of Protein Stability in Caenorhabditis elegans Using a Small Molecule, PLoS ONE 8(8): e72393. doi:10.1371/journal.pone.0072393). 본 명세서에 개시된 CRISPR-C2c1 시스템은 Ganz 등 및 Cho 등이 기술한 바와 유사한 구동 구성체 및 시스템에 적용될 수 있고, 일부 구현예에서, CRISPR-C2c1 시스템은 번식 조절에 관여되는 유전자를 변형시킨다. 일부 구현예에서, CRISPR-C2c1 시스템은 질환 관련 유전자를 변형시킨다.

일정 구현예에서, CRISPR-C2c1 시스템은 가축 생물량 관련 유전자를 변형시킨다. 일정 구현예에서, CRISPR-C2c1 시스템은 가축 형질 관련 유전자를 변형시킨다. 일부 구현예에서, 형질 관련 유전자는 해충 및 진균 감염 감수성에 관여된다. 특정 구현예에서, CRISPR-C2c1 시스템은 곤충 세포에 전달된다. 특정 구

현예에서, 곤충 세포는 꿀벌 세포이다. 일부 구현에예서, CRISPR-C2c1 시스템은 동물 세포에 전달된다. 일부 구현에예서, CRISPR-C2c1 시스템은 비동물 포유동물 세포에 전달된다. 특정 구현예에서, 형질 관련 유전자는 지방 과다의 조절에 관여된다. C2c1 단백질과 관련하여, CRISPR-C2c1 시스템은 T-풍부의 PAM 서열을 인식한다. 일부 구현예에서, PAM은 5' TTN 3' 또는 5' ATTN 3' 이며, 여기서 N 은 임의의 뉴클레오티드이다. 일부 구현예에서, CRISPR-C2c1 시스템은 5' 오버행을 갖는 하나 이상의 스태거드 이중 가닥 파손(DSB)을 도입시킨다. 특정 구현예에서, 5' 오버행은 7 nt 이다. 일부 구현예에서, CRISPR-C2c1 시스템은 HR 또는 NHEJ를 통해서 스태거드 DSB에서 외생성 주형 DNA 서열을 도입시킨다. 일부 구현예에서, C2c1 이펙터 단백질은 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, C2c1 이펙터 단백질은 닉카제이다. 일부 구현예에서, CRISPR-C2c1 시스템은 표적 관심 유전자좌를 변형시키는 기능성 도메인과 연관된 촉매적으로 불활성화된 C21 단백질을 포함한다. 특정 구현예에서, CRISPR-C2c1 시스템은 단일 돌연변이를 도입시킨다. 다른 특정 구현예에서, CRISPR-C2c1 시스템은 가축의 게놈을 변형시키지 않고 관심 표적 유전자좌의 전사물에 단일 뉴클레오티드 변형을 도입시킨다.

[1705] **이종이식**

[1706] 본 발명은 또한 이식을 위해 변형된 조직을 제공하는데 사용하도록 적합화된 RNA-가이드된 DNA 뉴클레아제를 제공하기 위해서, 본 명세서에 기술된 CRISPR-Cas 시스템, 예를 들어 C2c1 이펙터 단백질 시스템의 용도를 고려한다. 예를 들어, RNA-가이드된 DNA 뉴클레아제는 예를 들어 인간 면역계에 의해 인식되는 에피토프를 코딩하는 유전자, 즉 이종항원 유전자의 발현을 파괴하여, 유전자이식 돼지 (예컨대 인간 헴 옥시게나제-1 유전자이식 돼지 품종) 같은 동물에서 선택된 유전자를 녹아웃, 녹다운 또는 파괴시키는데 사용될 수 있다. 파괴를 위한 후보 돼지 유전자는 예를 들어 α(1,3)-갈락토실트랜스퍼라제 및 시티딘 모노포스페이트-N-아세틸뉴라민산 히드록실라제 유전자를 포함할 수 있다 (참조; PCT 공개 출원 WO 2014/066505). 또한, 내생성 레트로바이러스를 코딩하는 유전자는 예를 들어, 모든 돼지 내생성 레트로바이러스를 코딩하는 유전자를 파괴할 수 있다 (참조: Yang et al., 2015, Genome-wide inactivation of porcine endogenous retroviruses (PERVs), Science 27 November 2015: Vol. 350 no. 6264 pp. 1101-1104). 또한, RNA-가이드된 DNA 뉴클레아제는 초급성 거부에 대한 보호성을 제공하기 위해 인간 CD55 유전자같은, 이종이식 도너 동물에서 추가 유전자의 통합을 위한 부위를 표적화하는데 사용될 수 있다.

[1707] **일반적 유전자 치료 고려사항**

[1708] 질병-연관 유전자 및 폴리뉴클레오티드 및 질병 특이 정보의 예는 월드 와이드 웹 상에서 입수 가능한 존스홉킨스대학교(Johns Hopkins University)(미국 메릴랜드주 볼티모어 소재)의 맥쿠식-네이선스 유전의학 연구소(McKusick-Nathans Institute of Genetic Medicine) 및 국립 생명 기술 정보 센터(National Center for Biotechnology Information), 국립 의학 도서관(National Library of Medicine)(미국 메릴랜드주 베데스다 소재)로부터 입수할 수 있다.

[1709] 이들 유전자 및 경로에서의 돌연변이는 작용에 영향을 미치는 부적절한 양의 단백질 또는 부적절한 단백질의 생산을 초래할 수 있다. 유전자, 질병 및 단백질의 추가 예는 2012년 12월 12일 출원된 US 가출원 61/736,527로부터 참고로 본 명세서에 포함한다. 이러한 유전자, 단백질 및 경로는 본 발명의 CRISPR 복합체의 표적 폴리뉴클레오티드일 수 있다. 질환-연관 유전자 및 폴리뉴클레오티드의 예는 표 7 및 8에 열거되어 있다. 신호전달 생화학 경로-관련 유전자 및 폴리뉴클레오티드의 예는 표 9에 열거되어 있다.

표 7	
질환/장애	유전자(들)
신생물	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR 알파; PPAR 감마; WT1 (빌름스 종양); FGF 수용체 패밀리 구성원 (5 구성원: 1, 2, 3, 4, 5); CDKN2a; APC; RB (망막모세포종); MEN1; VHL; BRCA1; BRCA2; AR (안드로겐 수용체); TSG101; IGF; IGF 수용체; Igf1 (4 변이체); Igf2 (3 변이체); Igf 1 수용체; Igf 2 수용체; Bax; Bcl2; 캐스파제 패밀리 (9 구성원: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
나이-관련 황반 변성증	Abcr; Ccl2; Cc2; cp (세룰로플라스민); Timp3; 카텝신 D; Vldlr; Ccr2
조현병 장애	뉴레귤린 1 (Nrg1); Erb4 (뉴레귤린에 대한 수용체); 콤플렉신 1 (Cplx1); Tph1 트립토판 히드록실라제; Tph2 트립토판 히드록실라제 2; 뉴렉신 1; GSK3; GSK3a; GSK3b 5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA; DTNBP1; Dao (Dao1)
트리뉴클레오티드 반복 장애	HTT (헌팅톤 Dx); SBMA/SMAX1/AR (케네디 Dx); FXN/X25 (프리드리히 운동 실조); ATX3 (마카도-조셉 Dx); ATXN1 및 ATXN2 (척추소뇌성 운동실조증); DMPK (근긴장성 이영양증); 아트로핀-1 및 Atn1 (DRPLA Dx); CBP (Creb-BP - 전신 불안정증); VLDLR (알츠하이머); Atxn7; Atxn10
취약 X 증후군	FMR2; FXR1; FXR2; mGLUR5
세크리타제 관련 장애	APH-1 (알파 및 베타); 프레세닐린 (Psen1); 니카스트린 (Ncstn); PEN-2
기타	Nos1; Parp1; Nat1; Nat2
프라이온-관련 장애	Prp

[1710]

ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c)
약물 중독	Prkce (알콜); Drd2; Drd4; ABAT (알콜); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (알콜)
자폐증	Mecp2; BZRAP1; MDGA2; Sema5A; 뉴렉신 1; 취약 X (FMR2 (AFF2); FXR1; FXR2; Mglur5)
알츠하이머 질환	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; 클러스테린; PS1; SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1, 아쿠아포린 1); Uchl1; Uchl3; APP
염증	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL- 17b; IL-17c; IL-17d; IL-17f); II-23; Cx3cr1; ptpn22; TNFa; IBD 에 대한 NOD2/CARD15; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3cl1
파킨슨병	x-시누클레인; DJ-1; LRRK2; 파킨; PINK1

[1711]

[1712]

표 8	
혈액 및 응고 질환 및 장애	빈혈증 (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); 노출 림프구 증후군 (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), 출혈 장애 (TBXA2R, P2RX1, P2X1); 인자 H 및 인자 H-유사 1 (HF1, CFH, HUS); 인자 V 및 인자 VIII (MCFD2); 인자 VII 결핍증 (F7); 인자 X 결핍증 (F10); 인자 XI 결핍증 (F11); 인자 XII 결핍증 (F12, HAF); 인자 XIIIA 결핍증 (F13A1, F13A); 인자 XIIIB 결핍증 (F13B); 판코니 빈혈증 (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); 혈구탐식성 림프구조직구증가증 장애 (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); 혈우병 A (F8, F8C, HEMA); 혈우병 B (F9, HEMB), 출혈성 장애 (PI, ATT, F5); 백혈구 결핍증 및 장애 (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); 겸상 적혈구 빈혈증 (HBB); 탈라세미아 (HBA2, HBB, HBD, LCRB, HBA1).
세포 이상조절 및 종양학 질환 및 장애	B-세포 비호지킨 림프종 (BCL7A, BCL7); 백혈병 (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS,

	AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
염증 및 면역 관련 질환 및 장애	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); 자가면역 림프증식 증후군 (TNFRSF6, APT1, FAS, CD95, ALPS1A); 복합 면역결핍증, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), HIV 감수성 또는 감염 (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5)); 면역결핍증 (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); 염증 (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), II-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); 중증 복합 면역결핍증 (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
물질대사, 간, 신장, 및 단백질 질환 및 장애	아밀로이드 신경병증 (TTR, PALB); 아밀로이드증 (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); 간경변증 (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); 낭성 섬유증 (CFTR, ABCC7, CF, MRP7); 당원 축적병 (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); 간 선종, 142330 (TCF1, HNF1A, MODY3), 간부전증, 조기 발병, 및 신경학적 장애 (SCOD1, SCO1), 간 리파제 결핍증 (LIPC), 간모세포종, 암 및 암종 (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; 수질 낭성 신장 질환 (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); 페닐케톤뇨증 (PAH, PKU1, QDPR, DHPR, PTS); 다낭성 신장 및 간 질환(FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
근육/골격 질환 및 장애	베커형 근이영양증 (DMD, BMD, MYF6), 듀센형 근이영양증 (DMD, BMD); 에머리-드라이푸스 근이영양증 (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); 안면견갑상완형 근이영양증 (FSHMD1A, FSHD1A);

[1713]

	근이영양증 (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); 골화석증 (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); 근위축증 (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
신경학적 및 신경원성 질환 및 장애	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); 알츠하이머병 (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); 자폐증 (Mecp2, BZRAP1, MDGA2, Sema5A, 뉴렉신 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); 취약 X 증후군 (FMR2, FXR1, FXR2, mGLUR5); 헌팅톤병 및 질환 유사 장애 (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); 파킨슨병 (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); 레트 증후군 (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-시누클레인, DJ-1); 조현병 (뉴레귤린 1 (Nrg1), Erb4 (뉴레귤린에 대한 수용체), 콤플렉신 1 (Cplx1), Tph1 트립토판 히드록실라제, Tph2, 트립토판 히드록실라제 2, 뉴렉신 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); 세크리타제 관련 장애 (APH-1 (알파 및 베타), 프레세닐린 (Psen1), 니카스트린, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); 트리뉴클레오티드 반복 장애 (HTT (헌팅톤 Dx), SBMA/SMAX1/AR (케네디 Dx), FXN/X25 (프리드리히 운동 실조), ATX3 (마카도-조셉 Dx), ATXN1 및 ATXN2 (척추소뇌성 운동실조증), DMPK (근긴장성 이영양증), 아트로핀-1 및 Atn1 (DRPLA Dx), CBP (Creb-BP - 전신 불안정증), VLDLR (알츠하이머), Atxn7, Atxn10).
안구 질환 및 장애	나이-관련 황반 변성증 (Abcr, Ccl2, Cc2, cp (세룰로플라스민), Timp3, 카텝신 D, Vldlr, Ccr2); 백내장 (CRYAA, CRYA1, CRYBB2,

[1714]

	CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); 각막 혼탁 및 이영양증 (APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); 선천성 편평 각막증(KERA, CNA2); 녹내장 (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); 레베르 선천성 흑암시 (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); 황반 이영양증 (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

[1715]

표 9	
세포 기능	유전자
PI3K/AKT 신호전달	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2; PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1; AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2; PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2; ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3; PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7; YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A; CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1; PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2; TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK; HSP90AA1; RPS6KB1
ERK/MAPK 신호전달	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1; PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A; PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN; EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ; PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1; PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1; CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
글루코코르티코이드 수용체 신호전달	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1; MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I; PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2; MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1; MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13; RELA;

[1716]

	STAT5A; MAPK9; NOS2A; PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3; MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP; CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1; STAT1; IL6; HSP90AA1
축삭 유도 신호전달	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12; IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2; ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2; CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11; PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1; GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA
에프린 수용체 신호전달	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14; CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4; AKT3; SGK
액틴 세포골격 신호전달	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1; PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6; ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8; F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7; PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1; MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3; ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRAF; VAV3; SGK
헌팅톤병 신호전달	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2; MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2; PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST; GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2; HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A; HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1; PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX; ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
아폽토시스 신호전달	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1; BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8; BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF; RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2; CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2; BIRC2; TTK; CSNK1A1;

[1717]

	BRAF; BAX; PRKCA; SGK; CASP3; BIRC3; PARP1
B 세포 수용체 신호전달	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11; AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A; MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1; MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9; EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN; GSK3B; ATF4; AKT3; VAV3; RPS6KB1
백혈구 혈관외유출 신호전달	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA; RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11; MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12; PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB; MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK; MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2; CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK; CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
인테그린 신호전달	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A; TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2; CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8; CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA; SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1; TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL; BRAF; GSK3B; AKT3
급성기 반응 신호전달	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11; AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14; PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1; TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN; AKT3; IL1R1; IL6
PTEN 신호전달	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11; MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA; CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1; MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR; RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2; AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1; NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2; GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
p53 신호전달	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A; BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2; PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1; PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9; CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A; HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1; SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN; SNAI2; GSK3B; BAX; AKT3
아릴 탄화수소 수용체 신호전달	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1; NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1; SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1; MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1; SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF; CDKN1A;

[1718]

	NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1; CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1; HSP90AA1
생체이물 물질대사 신호전달	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1; NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A; PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1; ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD; GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL; NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1; CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1; NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1; HSP90AA1
SAPK/JNK 신호전달	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1; GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA; FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1; GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2; PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1; CRKL; BRAF; SGK
PPAr/RXR 신호전달	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN; RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2; ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8; IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A; NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1; TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1; ADIPOQ
NF-KB 신호전달	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6; TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2; KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1
뉴레귤린 신호전달	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1; MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI; CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS; PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3; EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA; HSP90AA1; RPS6KB1
Wnt & 베타 카테닌 신호전달	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO; AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A; WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK; LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1; PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOX2
인슐린 수용체 신호전달	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3; MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1; GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK; RPS6KB1

[1719]

IL-6 신호전달	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3; MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3; MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
간 담즙정체	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA; RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8; PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1; TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8; CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA; IL6
IGF-1 신호전달	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2; PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8; IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A; YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3; FOXO1; SRF; CTGF; RPS6KB1
NRF2-매개 산화 스트레스 반응	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1; NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL; NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP; MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1; GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
간 섬유증/간 성상 세포 활성화	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF; SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9; IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8; PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX; IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR 신호전달	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB; NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3; NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2; PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA; MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Fc 엡실론 RI 신호전달	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD; MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3; VAV3; PRKCA
G-단백질 결합 수용체 신호전달	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB; PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB; PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK; PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3; PRKCA
이노시톨 포스페이트 물질대사	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; DYRK1A; MAP2K2;

[1720]

	PIP5K1A; PIK3R1; MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF 신호전달	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC; PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2; PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC; JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF 신호전달	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;
	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3; BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN; RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN; VEGFA; AKT3; FOXO1; PRKCA
자연 살해 세포 신호전달	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11; KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6; PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1; PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
세포 주기: G1/S 체크포인트 조절	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC; ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11; HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1; E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1; GSK3B; RBL1; HDAC6
T 세포 수용체 신호전달	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS; NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN; MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10; JUN; VAV3
사멸 수용체 신호전달	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD; FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8; DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB; CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3; BIRC3
FGF 신호전달	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11; AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8; MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1; AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4; AKT3; PRKCA; HGF
GM-CSF 신호전달	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A; STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3; ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3; STAT1
근위축성 측삭 경화증 신호전달	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2; PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1; PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1; APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat 신호전달	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A; PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1;

[1721]

	AKT3; STAT1
니코티네이트 및 니코틴아미드 물질대사	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1; PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1; PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2; MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
케모카인 신호전달	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13; RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1; MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 신호전달	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS; STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2; JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
시냅스 장기 저하	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS; PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3; KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
에스트로겐 수용체 신호전달	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2; SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1; HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP; MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
단백질 유비퀴틴화 경로	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4; CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8; USP1; VHL; HSP90AA1; BIRC3
IL-10 신호전달	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1; JUN; IL1R1; IL6
VDR/RXR 활성화	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1; NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD; RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1; LRP5; CEBPB; FOXO1; PRKCA
TGF-베타 신호전달	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2; MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll-유사 수용체 신호전달	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK; NFKB1; TLR2; JUN
p38 MAPK 신호전달	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS; CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2; MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1; SRF; STAT1
뉴로트로핀/TRK 신호전달	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A; RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; CDC42; JUN; ATF4

[1722]

FXR/RXR 활성화	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8; APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A; TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
시냅스 장기 강화	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1; PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS; PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1; ATF4; PRKCA
칼슘 신호전달	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1; CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11; HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4; HDAC6
EGF 신호전달	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1; STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
심혈관계의 저산소증 신호전달	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT; HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM; VEGFA; JUN; ATF4; VHL; HSP90AA1
RXR 기능의 LPS/IL-1 매개 억제	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1; MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2; TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR 활성화	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA; NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1; SREBF1; IL1R1; CCL2; IL6; MMP9
아밀로이드 프로세싱	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2; CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1; PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 신호전달	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1; PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1; FRAP1; AKT3; RPS6KB1
세포 주기: G2/M DNA 손상 체크포인트 조절	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC; CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A; PRKDC; ATM; SFN; CDKN2A
심혈관계의 산화질소 신호전달	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3; CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1; VEGFA; AKT3; HSP90AA1
푸린 물질대사	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4; PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C; NT5E; POLD1; NME1
cAMP-매개 신호전달	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3; SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
미토콘드리아	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9; PARK7; PSEN1;

[1723]

기능이상	PARK2; APP; CASP3
Notch 신호전달	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2; PSEN1; NOTCH3; NOTCH1; DLL4
소포체 스트레스 경로	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4; EIF2AK3; CASP3
피리미딘 물질대사	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B; NT5E; POLD1; NME1
파킨슨 신호전달	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7; PARK2; CASP3
심장 & 베타 아드레날린성 신호전달	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC; PPP2R5C
해당과정/포도당 신생합성	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
인터페론 신호전달	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
소닉 헤지호그 신호전달	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
글리세로인지질 물질대사	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
인지질 분해	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
트립토판 물질대사	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
리신 분해	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
뉴클레오티드 절단 복구 경로	ERCC5; ERCC4; XPA; XPC; ERCC1
전분 및 수크로스 물질대사	UCHL1; HK2; GCK; GPI; HK1
아미노당 물질대사	NQO1; HK2; GCK; HK1
아라키돈산 물질대사	PRDX6; GRN; YWHAZ; CYP1B1
일주기 리듬 신호전달	CSNK1E; CREB1; ATF4; NR1D1

[1724]

응고 시스템	BDKRB1; F2R; SERPINE1; F3
도파민 수용체 신호전달	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
글루타티온 물질대사	IDH2; GSTP1; ANPEP; IDH1
글리세로지질 물질대사	ALDH1A1; GPAM; SPHK1; SPHK2
리놀레산 물질대사	PRDX6; GRN; YWHAZ; CYP1B1
메티오닌 물질대사	DNMT1; DNMT3B; AHCY; DNMT3A
피루베이트 물질대사	GLO1; ALDH1A1; PKM2; LDHA
아르기닌 및 프롤린 물질대사	ALDH1A1; NOS3; NOS2A
에이코사노이드 신호전달	PRDX6; GRN; YWHAZ
프룩토스 및 만노스 물질대사	HK2; GCK; HK1
갈락토스 물질대사	HK2; GCK; HK1
스틸벤, 쿠마린 및 리그닌 생합성	PRDX6; PRDX1; TYR
항원 제시 경로	CALR; B2M
스테로이드의 생합성	NQO1; DHCR7
부타노에이트 물질대사	ALDH1A1; NLGN1
시트레이트 회로	IDH2; IDH1
지방산 물질대사	ALDH1A1; CYP1B1
글리세로인지질 물질대사	PRDX6; CHKA
히스티딘	PRMT5; ALDH1A1

[1725]

물질대사	
이노시톨 물질대사	ERO1L; APEX1
시토크롬 p450 에 의한 생체이물의 물질대사	GSTP1; CYP1B1
메탄 물질대사	PRDX6; PRDX1
페닐알라닌 물질대사	PRDX6; PRDX1
프로파노에이트 물질대사	ALDH1A1; LDHA
셀레노아미노산 물질대사	PRMT5; AHCY
스핀고지질 물질대사	SPHK1; SPHK2
아미노포스포네이트 물질대사	PRMT5
안드로겐 및 에스트로겐 물질대사	PRMT5
아스코르베이트 및 알다레이트 물질대사	ALDH1A1
담즙산 생합성	ALDH1A1
시스테인 물질대사	LDHA
지방산 생합성	FASN
글루타메이트 수용체 신호전달	GNB2L1
NRF2-매개 산화 스트레스 반응	PRDX1
펜토스 포스페이트 경로	GPI
펜토스 및	UCHL1

[1726]

글루쿠로네이트 상호전환	
레티놀 물질대사	ALDH1A1
리보플라빈 물질대사	TYR
티로신 물질대사	PRMT5, TYR
유비퀴논 생합성	PRMT5
발린, 류신, 및 이소류신 분해	ALDH1A1
글리신, 세린, 및 트레오닌 물질대사	CHKA
리신 분해	ALDH1A1
통증/미각	TRPM5; TRPA1
통증	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2; Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a
미토콘드리아 기능	AIF; CytC; SMAC (디아블로); Aifm-1; Aifm-2
발달 신경학	BMP-4; 코르딘 (Chrd); 노긴 (Nog); WNT (Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); 베타-카테닌; Dkk-1; 프리즐 관련 단백질; Otx-2; Gbx2; FGF-8; 릴린; Dab1; unc-86 (Pou4f1 or Brn3a); Numb; Reln

[1727]

[1728] 또한, 본 발명의 구현예는 유전자의 녹아웃, 유전자의 증폭 및 DNA 반복부 불안정성 및 신경계 장애와 관련된 특정 돌연변이의 복구에 관한 방법 및 조성물에 관한 것이다(문헌[Robert D.Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct 13, 2011 - Medical). 탠덤 반복 서열의 특이적 양태는 20 개 초과의 인간 질병에 원인임이 발견되었다(문헌[New insights into repeat instability:role of RNA·DNA hybrids.McIvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep-Oct;7(5):551-8). 본 발명의 이펙터 단백질 시스템은 게놈 불안정성의 이들 결함을 편집하기 위해 활용될 수 있다.

[1729] 본 발명의 여러 추가 양태는 국립 보건원의 웹사이트에서 토픽 서브섹션 유전 장애(Genetic Disorders)(웹사이트 health.nih.gov/topic/GeneticDisorders) 하에 추가 기재된 광범위한 유전 질병과 연관된 결함을 편집하는 것에 관한 것이다. 유전 뇌 질병은 부신백질 이영양증, 뇌량의 무발육, 에카르디 증후군, 알퍼스 병, 알츠하이머병, 바쓰 증후군, 바텐병, CADASIL, 소뇌 변성, 파브리병, 게르스트만-슈트로이슬러-샤인커병, 헌팅턴병 및 기타 다른 트리플릿 반복 장애, 라이병, 레시-니한 증후군, 멘케스병, 미토콘드리아 근병증 및 NINDS 거대후두각을 포함할 수 있으나, 이로 제한되지 않는다. 이들 질병은 국립보건원의 웹사이트에서 서브섹션 유전 뇌 장애 하에 추가 기재되어 있다.

[1730] 본 개시 전반에서 CRISPR 또는 CRISPR-Cas 복합체 또는 시스템을 언급하였다. CRISPR 시스템 또는 복합체는 핵산 분자를 표적화할 수 있고, 예를 들어 CRISPR-C2c1 복합체는 표적 DNA 분자를 표적화하고 절단하거나 또는 닉킹하거나 또는 단순히 자리잡을 수 있다 (C2c1이 닉카제 또는 "데드" 가 되게하는 돌연변이를 갖는지에 따라서). 이러한 시스템 또는 복합체는 후보 질환 유전자의 조직-특이적 및 일시적으로 제어된 표적화된 결실을 달성할 수 있다. 예들은 콜레스테롤 및 지방산 대사, 아밀로이드 질환, 우성 음성 질환, 잠복 바이러스 감염에 수반되는 유전자를 포함하지만, 이에 제한되지 않는다. 따라서, 이러한 시스템 또는 복합체에 대한 표적 서열이 후보 질환 유전자에 있을 수 있다, eg.:

표 10 – 질환 및 표적					
질환	유전자	스페이서	PAM	기전	참조문헌
고콜레스테롤혈증	HMG-CR	GCCAAATTGGACGACCCTCG (SEQ ID NO:480)	CGG	녹아웃	Fluvastatin: a review of its pharmacology and use in the management of Hypercholesterolemia.(Plosker GL et al. Drugs 1996, 51(3):433-459)
고콜레스테롤혈증	SQLE	CGAGGAGACCCCCGTTTCGG (SEQ ID NO:481)	TGG	녹아웃	Potential role of nonstatin cholesterol lowering agents (Trapani et al. IUBMB Life, Volume 63, Issue 11, pages 964–971, November 2011)
고지혈증	DGAT1	CCCGCCGCCGCCGTGGCTCG (SEQ ID NO:482)	AGG	녹아웃	DGAT1 inhibitors as anti-obesity and anti-diabetic agents. (Birch AM et al. Current Opinion in Drug Discovery & Development [2010, 13(4):489-496)
백혈병	BCR-ABL	TGAGCTCTACGAGATCCACA (SEQ ID NO:483)	AGG	녹아웃	Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi).(Fuchs et al. Oncogene 2002, 21(37):5716-5724)

[1731]

[1732] 따라서, 본 발명은 CRISPR 또는 CRISPR-Cas 복합체와 관련하여, 조혈 장애의 교정을 고려한다. 예를 들어, 중증 복합형 면역 결핍증 (SCID)은 림프구 B에서의 작용성 결함과 항상 관련되는 림프구 T 성숙에서의 결함으로부터 야기한다 (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). SCID 형태 중 하나인 아데노신 디아미나제(Deaminase:ADA) 결핍증의 경우에, 환자는 재조합 아데노신 디아미나제 효소의 주사에 의해 치료될 수 있다. ADA 유전자가 SCID 환자에서 돌연변이된 것으로 보여진 이후로 (Giblett et al., Lancet, 1972, 2, 1067-1069), SCID와 관련된 여러 다른 유전자가 확인되었다 (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). SCID에 대한 4 개의 주요 원인이 있다:(i) SCID의 가장 빈번한 형태인, SCID-X1 (X-관련 SCID 또는 X-SCID)은 IL2RG 유전자에서의 돌연변이에 의해 발생하며, 이는 성숙 T 림프구 및 NK 세포의 부재를 야기한다. IL2RG는 적어도 5 개 인터루킨 수용체 복합체의 공통 성분인 감마 C 단백질을 코딩한다 (문헌 [Noguchi, et al., Cell, 1993, 73, 147157]). 이들 수용체는 JAK3 키나제를 통해 몇몇 표적을 활성화시키는데 (Macchi et al., Nature, 1995, 377, 65-68), 이의 불활성화는 감마 C 불활성화와 동일한 증후군을 초래하고; (ii) ADA 유전자의 돌연변이는 림프종 전구체에 치명적이어서, 이후 B, T, 및 NK 세포의 유사 부재를 야기시키는 푸린 물질대사의 결함을 야기하고; (iii) V(D)J 재조합은 면역글로불린 및 T 림프구 수용체 (TCR)의 성숙화에서 필수 단계이다. 이러한 과정에 관여되는 3개 유전자, 재조합 활성화 유전자 1 및 2 (RAG1 및 RAG2) 및 Artemis 내 돌연변이는 성숙한 T 및 B 림프구의 부재를 야기하고; (iv) T 세포 특이적 신호전달에 관여하는, CD45 같은 다른 유전자 내 돌연변이가 또하나 보고되었지만, 그들은 소수 사례를 대표한다 (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109).

본 발명의 양상에서, CRISPR 또는 CRISPR-Cas 복합체 고려 시스템과 관련하여, 본 발명은 문헌 [Genetic Diseases of the Eye, Second Edition, edited by Elias I. Traboulsi, Oxford University Press, 2012]에 추가 기재된 여러 유전적 돌연변이로부터 발생한 안구 결함을 교정할 수 있는데 사용할 수 있다고 고려한다. 교정하려는 안구 결함의 비제한적인 예는 황반 변성 (MD), 망막색소변성증 (RP)을 포함한다. 안구 결함과 연관된 유전자 및 단백질의 비제한적 예로서, MD와 관련된 단백질은 다음의 단백질을 포함하지만, 이들로 제한되지 않는다:(ABCA4) ATP-결합 카세트, 서브 패밀리 A(ABC1), 멤버 4 ACHM1 완전색맹(간상체 전색맹) 1 ApoE 아포지방단백질 E(ApoE) C1QTNF5(CTRP5) C1q 및 종양 괴사 인자 관련 단백질 5(C1QTNF5) C2 상보체 구성성분 2(C2) C3 상보체 구성성분(C3) CCL2 케모카인(C-C 모티프) 리간드 2(CCL2) CCR2 케모카인(C-C 모티프) 수용체 2(CCR2) 분화 36의 CD36 클러스터 CFB 상보체 인자 B CFH 상보체 인자 CFH H CFHR1 상보체 인자 H-관련 1 CFHR3 상보체 인자 H-관련 3 CNGB3 사이클릭 뉴클레오티드화된 채널 베타 3 CP 셀룰로플라스민(CP) CRP C 반응 단백질(CRP) CST3 시스타틴 C 또는 시스타틴 3(CST3) CTSD 카텝신 D(CTSD) CX3CR1 케모카인(C-X3-C 모티프) 수

용체 1 ELOVL4 매우 장쇄 지방산 4 의 신장 ERCC6 절단 수선 교차 상보성 설치류 수선 결함, 상보체화 그룹 6 FBLN5 피불린-5 FBLN5 피불린 5 FBLN6 피불린 6 FSCN2 파신(FSCN2) HMCN1 헤미센트린 1 HMCN1 헤미센트린 1 HTRA1 HtrA 세린 펩티다제 1(HTRA1) HTRA1 HtrA 세린 펩티다제 1 IL-6 인터루킨 6 IL-8 인터루킨 8 LOC387715 가설 단백질 PLEKHA1 플렉스트린 상동성 도메인-함유 패밀리 A 멤버 1 (PLEKHA1) PROM1 프로미닌 1(PROM1 또는 CD133) PRPH2 페리페린-2 RPGR 망막색소변성증 GTPase 조절자 세르핀G1 세르핀 펩티다제 저해제, 클레이트 G, 멤버 1(C1- 저해제) TCOF1 트리클 TIMP3 메탈로프로테이나제 저해제 3 (TIMP3) TLR3 Toll-유사 수용체 3 본 발명은 CRISPR 또는 CRISPR-Cas 복합체와 관련하여, 심장으로 전달을 고려한다. 심장에서, 선호되는 유전자 이송을 보이는 심근 열대성 아데나-연관 바이러스(AAVM), 특히 AAVM41이 바람직하다 (참조: Lin-Yanga et al., PNAS, March 10, 2009, vol. 106, no. 10). 예를 들어, 미국 특허 공개 제20110023139호는 세포, 동물 및 심혈관 질병에 연관된 단백질을 유전적으로 변형시키기 위한 아연 핑거 뉴클레아제의 사용을 기재한다. 심혈관 질환은 일반적으로 고혈압, 심장마비, 심부전, 및 뇌졸중 및 TIA를 포함한다. 예로서, 염색체 서열은 제한 없이 다음을 포함할 수 있다: IL1B (인터루킨 1, 베타), XDH (잔틴 디히드로게나제), TP53 (종양 단백질 p53), PTGIS (프로스타글란틴 12 (프로스타사이클린) 신타제), MB (미오글로빈), IL4 (인터루킨 4), ANGPT1 (안지오포이어틴 1), ABCG8 (ATP-결합 카세트, 서브-패밀리 G (WHITE), 멤버 8), CTSK (카텝신 K), PTGIR (프로스타글란틴 12 (프로스타사이클린) 수용체 (IP)), KCNJ11 (포타슘 내향성 채널, 서브패밀리 J, 멤버 11), INS (인슐린), CRP (C-반응성 단백질, 펜트락신-관련), PDGFRB (혈소판-유래 성장 인자 수용체, 베타 폴리펩티드), CCNA2 (사이클린 A2), PDGFB (혈소판-유래 성장 인자 베타 폴리펩티드 (원숭이 육종 바이러스 (v-sis) 종양 유전자 상동체)), KCNJ5 포타슘 내향성 채널, 서브패밀리 J, 멤버 5), KCNN3 (포타슘 중간체/소형 컨덕턴스 칼슘-활성화된 채널, 서브패밀리 N, 멤버 3), CAPN10 (칼파인 10), PTGES (프로스타글란틴 E 신타제), ADRA2B (아드레날린성, 알파-2B-, 수용체), ABCG5 (ATP-결합 카세트, 서브-패밀리 G (WHITE), 멤버 5), PRDX2 (퍼옥시리독신 2), CAPN5 (칼파인 5), PARP14 (폴리 (ADP-리보스) 중합효소 패밀리, 멤버 14), MEX3C (mex-3 상동체 C (예쁜 꼬마 선충)), ACE 안지오텐신 I 전환 효소 (펩티딜-디펩티다제 A) 1), TNF (종양 괴사 인자 (TNF 수퍼패밀리, 멤버 2)), IL6 (인터루킨 6 (인터페론, 베타 2)), STN (스타틴), SERPINE1 (세르핀 펩티다제 억제제, 클레이드 E (넥신, 플라스미노겐 활성인자 억제제 1형), 멤버 1), ALB (알부민), ADIPOQ (아디포넥틴, C1Q 및 콜라겐 도메인 함유), APOB (아포지단백질 B (Ag(x) 항원 포함)), APOE (아포지단백질 E), LEP (렙틴), MTHFR (5,10-메틸렌테트라히드로폴레이트 리덕타제 (NADPH)), APOA1 (아포지단백질 A-I), EDN1 (엔도텔린 1), NPPB (나트륨뇨설 펩티드 전구체 B), NOS3 (산화질소 신타제 3 (내피 세포)), PPARG (퍼옥시솜 증식인자-활성화 수용체 감마), PLAT (플라스미노겐 활성인자, 조직), PTGS2 (프로스타글란틴-엔도퍼옥시다제 신타제 2 (프로스타글란틴 G/H 신타제 및 시클로옥시게나제)), CETP (콜레스테릴 에스테르 전달 단백질, 혈장), AGTR1 (안지오텐신 II 수용체, 1형), HMGCR (3-히드록시-3-메틸글루타릴-조효소 A 리덕타제), IGF1 (인슐린-유사 성장 인자 1 (소마토메딘 C)), SELE (셀렉틴 E), REN (레닌), PPARA (퍼옥시솜 증식인자-활성화 수용체 알파), PON1 (파라옥소나제 1), KNG1 (키니노겐 1), CCL2 (케모카인 (C-C 모티프) 리간드 2), LPL (지단백질 리파제), VWF (폰 빌레브란트 인자), F2 (응고 인자 II (트롬빈)), ICAM1 (세포내 부착 분자 1), TGFB1 (형질전환 성장 인자, 베타 1), NPPA (나트륨뇨설 펩티드 전구체 A), IL10 (인터루킨 10), EPO (에리쓰로포이어틴), SOD1 (수퍼옥시드 디스뮤타제 1, 가용성), VCAM1 (혈관 세포 부착 분자 1), IFNG (인터페론, 감마), LPA (지단백질, Lp(a)), MPO (미엘로퍼옥시다제), ESR1 (에스트로겐 수용체 1), MAPK1 (미토겐-활성화 단백질 키나제 1), HP (하프토글로빈), F3 (응고 인자 III (트롬보플라스틴, 조직 인자)), CST3 (시스타틴 C), COG2 (올리고머 골지 복합체의 성분 2), MMP9 (매트릭스 메탈로펩티다제 9 (젤라티나제 B, 92 kDa 젤라티나제, 92 kDa IV형 콜라게나제)), SERPINC1 (세르핀 펩티다제 억제제, 클레이드 C (항트롬빈), 멤버 1), F8 (응고 인자 VIII, 응혈촉진 성분), HMOX1 (헴 옥시게나제 (디사이클링) 1), APOC3 (아포지단백질 C-III), IL8 (인터루킨 8), PROK1 (프로키네티신 1), CBS (시스타티오닌-베타-신타제), NOS2 (산화질소 신타제 2, 유도성), TLR4 (toll-유사 수용체 4), SELP (셀렉틴 P (과립 막단백질 140 kDa, 항원 CD62)), ABCA1 (ATP-결합 카세트, 서브-패밀리 A (ABC1), 멤버 1), AGT (안지오텐시노겐 (세르핀 펩티다제 억제제, 클레이드 A, 멤버 8)), LDLR (저밀도 지단백질 수용체), GPT (글루탐산-피루베이트 트랜사미나제 (알라닌 아미노트랜스퍼라제)), VEGFA (혈관 내피 성장 인자 A), NR3C2 (핵 수용체 서브패밀리 3, 그룹 C, 멤버 2), IL18 (인터루킨 18 (인터페론-감마-유도 인자)), NOS1 (산화질소 신타제 1 (신경원)), NR3C1 (핵 수용체 서브패밀리 3, 그룹 C, 멤버 1 (글루코코르티코이드 수용체)), FGB (피브리노겐 베타 사슬), HGF (간세포 성장 인자 (헤파포이어틴 A; 산란 인자)), IL1A (인터루킨 1, 알파), RETN (레시스틴), AKT1 (v-akt 쥐과 흉선종 바이러스 종양유전자 상동체 1), LIPC (리파제, 간), HSPD1 (열충격 60 kDa 단백질 1 (샤페로닌)), MAPK14 (미토겐-활성화 단백질 키나제 14), SPP1 (분비형 인단백질 1), ITGB3 (인테그린, 베타 3 (혈소판 당단백질 111a, 항원 CD61)), CAT (카탈라제), UTS2 (우로텐신 2), THBD (트롬보모듈린), F10 (응고 인자 X), CP (세룰로플라스민 (페록시다제)), TNFRSF11B (종양 괴사 인자 수용체 수퍼패밀리,

멤버 11b), EDNRA (엔도텔린 수용체 A형), EGFR (상피 성장 인자 수용체 (적아세포 백혈병 바이러스 (v-erb-b) 종양 유전자 상동체, 조류)), MMP2 (매트릭스 메탈로펩티다제 2 (젤라티나제 A, 72 kDa 젤라티나제, 72 kDa IV 형 콜라게나제)), PLG (플라스미노겐), NPY (뉴로펩티드 Y), RHOD (ras 상동체 유전자 패밀리, 멤버 D), MAPK8 (미토겐-활성화 단백질 키나제 8), MYC (v-myc 골수구종증 바이러스 종양유전자 상동체 (조류)), FN1 (피브로넥틴 1), CMA1 (카이마제 1, 비만 세포), PLAU (플라스미노겐 활성인자, 우로키나제), GNB3 (구아닌 뉴클레오티드 결합 단백질 (G 단백질), 베타 폴리펩티드 3), ADRB2 (아드레날린성, 베타-2-, 수용체, 표면), APOA5 (아포지단백질 A-V), SOD2 (수퍼옥시드 디스뮤타제 2, 미토콘드리아), F5 (응고 인자 V (프로악세렐린, 불안정 인자)), VDR (비타민 D (1,25-디히드록시비타민 D3) 수용체), ALOX5 (아라키도네이트 5-리폭시게나제), HLA-DRB1 (주요 조직적합성 복합체, 클래스 II, DR 베타 1), PARP1 (폴리 (ADP-리보스) 중합효소 1), CD40LG (CD40 리간드), PON2 (파라옥소나제 2), AGER (진행성 글리코실화 최종 산물-특이적 수용체), IRS1 (인슐린 수용체 기질 1), PTGS1 (프로스타글란틴-엔도퍼옥시다제 신타제 1 (프로스타글란틴 G/H 신타제 및 시클로옥시게나제)), ECE1 (엔도텔린 전환 효소 1), F7 (응고 인자 VII (혈청 프로트롬비니 전환 가속인자)), URN (인터루킨 1 수용체 길항제), EPHX2 (에폭시드 히드롤라제 2, 세포질), IGFBP1 (인슐린-유사 성장 인자 결합 단백질 1), MAPK10 (미토겐-활성화 단백질 키나제 10), FAS (Fas (TNF 수용체 수퍼패밀리, 멤버 6)), ABCB1 (ATP-결합 카세트, 서브-패밀리 B (MDR/TAP), 멤버 1), JUN (jun 종양 유전자), IGFBP3 (인슐린-유사 성장 인자 결합 단백질 3), CD14 (CD14 분자), PDE5A (포스포디에스터라제 5A, cGMP-특이적), AGTR2 (안지오텐신 II 수용체, 2형), CD40 (CD40 분자, TNF 수용체 수퍼패밀리 멤버 5), LCAT (레시틴-콜레스테롤 아실트랜스퍼라제), CCR5 (케모카인 (C-C 모티프) 수용체 5), MMP1 (매트릭스 메탈로펩티다제 1 (간질 콜라게나제)), TIMP1 (TIMP 메탈로펩티다제 억제제 1), ADM (아드레노메둘린), DYT10 (근긴장이상 10), STAT3 (신호 전달인자 및 전사의 활성인자 3 (급성기 반응 인자)), MMP3 (매트릭스 메탈로펩티다제 3 (스트로멜리신 1, 프로젤라티나제)), ELN (엘라스틴), USF1 (상류 전사 인자 1), CFH (보체 인자 H), HSPA4 (열충격 70 kDa 단백질 4), MMP12 (매트릭스 메탈로펩티다제 12 (마크로파지 엘라스타제)), MME (막 메탈로엔도펩티다제), F2R (응고 인자 II (트롬빈) 수용체), SELL (셀렉틴 L), CTSB (카텝신 B), ANXA5 (아넥신 A5), ADRB1 (아드레날린성, 베타-1-, 수용체), CYBA (시토크롬 b-245, 알파 폴리펩티드), FGA (피브리노겐 알파 사슬), GGT1 (감마-글루타미트랜스퍼라제 1), LIPG (리파제, 상피), HIF1A (저산소증 유도성 인자 1, 알파 서브유닛 (염기성 헬릭스-루프-헬릭스 전사 인자)), CXCR4 (케모카인 (C-X-C 모티프) 수용체 4), PROC (단백질 C (응고 인자 Va 및 VIIIa의 불활성인자)), SCARB1 (스캐빈저 수용체 클래스 B, 멤버 1), CD79A (CD79a 분자, 면역글로불린-연관 알파), PLTP (인지질 전달 단백질), ADD1 (아두신 1 (알파)), FGG (피브리노겐 감마 사슬), SAA1 (혈청 아밀로이드 A1), KCNH2 (포타슘 전압-게이팅 채널, 서브패밀리 H (eag-관련), 멤버 2), DPP4 (디펩티딜-펩티다제 4), G6PD (글루코스-6-포스페이트 디히드로게나제), NPR1 (나트륨뇨설 펩티드 수용체 A/구아닐레이트 시클라제 A (심방나트륨이뇨 펩티드 수용체 A)), VTN (비트로넥틴), KIAA0101 (KIAA0101), FOS (FBJ 쥐과 골육종 바이러스 종양유전자 상동체), TLR2 (toll-유사 수용체 2), PPIG (펩티딜프롤릴 이소머라제 G (사이클로필린 G)), IL1R1 (인터루킨 1 수용체, I형), AR (안드로겐 수용체), CYP1A1 (시토크롬 P450, 패밀리 1, 서브패밀리 A, 폴리펩티드 1), SERPINA1 (세르핀 펩티다제 억제제, 클레이드 A (알파-1 안티프로테이나제, 안티트립신), 멤버 1), MTR (5-메틸테트라히드로폴레이트-호모시스테인 메틸트랜스퍼라제), RBP4 (레티놀 결합 단백질 4, 혈장), APOA4 (아포지단백질 A-IV), CDKN2A (사이클린-의존적 키나제 억제제 2A (골수종, p16, CDK4 억제)), FGF2 (섬유아세포 성장 인자 2 (염기성)), EDNRB (엔도텔린 수용체 B형), ITGA2 (인테그린, 알파 2 (CD49B, VLA-2 수용체의 알파 2 서브유닛)), CABIN1 (칼시뉴린 결합 단백질 1), SHBG (성 호르몬-결합 글로불린), HMGB1 (고-이동성 그룹 박스 1), HSP90B2P (열충격 단백질 90 kDa 베타 (Grp94), 멤버 2 (가유전자)), CYP3A4 (시토크롬 P450, 패밀리 3, 서브패밀리 A, 폴리펩티드 4), GJA1 (간극 연접 단백질, 알파 1, 43 kDa), CAV1 (카베올린 1, 카베올라 단백질, 22 kDa), ESR2 (에스트로겐 수용체 2 (ER 베타)), LTA (림포톡신 알파 (TNF 수퍼패밀리, 멤버 1)), GDF15 (성장 분화 인자 15), BDNF (뇌-유래 신경영양 인자), CYP2D6 (시토크롬 P450, 패밀리 2, 서브패밀리 D, 폴리펩티드 6), NGF (신경 성장 인자 (베타 폴리펩티드)), SP1 (Sp1 전사 인자), TGIF1 (TGFB-유도 인자 호메오박스 1), SRC (v-src 육종 (슈미트-루핀 A-2) 바이러스 종양유전자 상동체 (조류)), EGF (상피 성장 인자 (베타-우로가스트론)), PIK3CG (포스포이노시티드-3-키나제, 촉매성, 감마 폴리펩티드), HLA-A (주요 조직적합성 복합체, 클래스 I, A), KCNQ1 (포타슘 전압-게이팅 채널, KQT-유사 서브패밀리, 멤버 1), CNR1 (칸나비노이드 수용체 1 (뇌)), FBN1 (피브릴린 1), CHKA (콜린 키나제 알파), BEST1 (베스트로핀 1), APP (아밀로이드 베타 (A4) 전구체 단백질), CTNNB1 (카테닌 (카데린-연관 단백질), 베타 1, 88 kDa), IL2 (인터루킨 2), CD36 (CD36 분자 (트롬보스폰딘 수용체)), PRKAB1 (단백질 키나제, AMP-활성화, 베타 1 비촉매 서브유닛), TPO (갑상선 퍼옥시다제), ALDH7A1 (알데히드 디히드로게나제 7 패밀리, 멤버 A1), CX3CR1 (케모카인 (C-X3-C 모티프) 수용체 1), TH (티로신 히드롤라제), F9 (응고 인자 IX), GH1 (성장 호르몬 1), TF (트랜스페린), HFE (혈색소증), IL17A (인터루킨 17A), PTEN (포스파타제 및 텐신 상

동체), GSTM1 (글루타티온 S-트랜스퍼라제 mu 1), DMD (디스트로핀), GATA4 (GATA 결합 단백질 4), F13A1 (응고 인자 XIII, A1 폴리펩티드), TTR (트랜스타이레틴), FABP4 (지방산 결합 단백질 4, 지방세포), PON3 (파라옥소나제 3), APOC1 (아포지단백질 C-I), INSR (인슐린 수용체), TNFRSF1B (종양 괴사 인자 수용체 수퍼패밀리, 멤버 1B), HTR2A (5-히드록시트립타민 (세로토닌) 수용체 2A), CSF3 (콜로니 자극 인자 3 (과립구)), CYP2C9 (시토크롬 P450, 패밀리 2, 서브패밀리 C, 폴리펩티드 9), TXN (티오레독신), CYP11B2 (시토크롬 P450, 패밀리 11, 서브패밀리 B, 폴리펩티드 2), PTH (부갑상선 호르몬), CSF2 (콜로니 자극 인자 2 (과립구-마크로파지)), KDR (키나제 삽입 도메인 수용체 (a type III 수용체 티로신 키나제)), PLA2G2A (포스포리파제 A2, 그룹 IIA (혈소판, 활액)), B2M (베타-2-마이크로글로불린), THBS1 (트롬보스폰딘 1), GCG (글루카곤), RHOA (ras 상동체 유전자 패밀리, 멤버 A), ALDH2 (알데히드 디히드로게나제 2 패밀리 (미토콘드리아)), TCF7L2 (전사 인자 7-유사 2 (T-세포 특이적, HMG-box)), BDKRB2 (브라디키닌 수용체 B2), NFE2L2 (핵 인자 (적혈구-유래 2)-유사 2), NOTCH1 (Notch 상동체 1, 전좌연관 (초파리)), UGT1A1 (UDP 글루쿠로노실트랜스퍼라제 1 패밀리, 폴리펩티드 A1), IFNA1 (인터페론, 알파 1), PPARD (퍼옥시솜 증식인자-활성화 수용체 델타), SIRT1 (실투인 (침묵 교잡 유형 정보 조절 2 상동체) 1 (에스. 세레비지아에)), GNRH1 (고나도트로핀-방출 호르몬 1 (황체-방출 호르몬)), PAPPA (임신-연관 혈장 단백질 A, 파파리신 1), ARR3 (어레스틴 3, 망막 (X-어레스틴)), NPPC (나트륨뇨설 펩티드 전구체 C), AHSP (알파 헤모글로빈 안정화 단백질), PTK2 (PTK2 단백질 티로신 키나제 2), IL13 (인터루킨 13), MTOR (라파마이신의 기계론적 표적 (세린/트레오닌 키나제)), ITGB2 (인테그린, 베타 2 (보체 성분 3 수용체 3 및 4 서브유닛)), GSTT1 (글루타티온 S-트랜스퍼라제 쎄타 1), IL6ST (인터루킨 6 신호 전달인자 (gp130, 온코스타틴 M 수용체)), CPB2 (카르복시 펩티다제 B2 (혈장)), CYP1A2 (시토크롬 P450, 패밀리 1, 서브패밀리 A, 폴리펩티드 2), HNF4A (간세포 핵 인자 4, 알파), SLC6A4 (용질 운반체 패밀리 6 (신경전달인자 수송체, 세로토닌), 멤버 4), PLA2G6 (포스포리파제 A2, 그룹 VI (시토졸, 칼슘-독립적)), TNFSF11 (종양 괴사 인자 (리간드) 수퍼패밀리, 멤버 11), SLC8A1 (용질 운반체 패밀리 8 (소듐/칼슘 교환체), 멤버 1), F2RL1 (응고 인자 II (트롬빈) 수용체-유사 1), AKR1A1 (알도-케토 리덕타제 패밀리 1, 멤버 A1 (알데히드 리덕타제)), ALDH9A1 (알데히드 디히드로게나제 9 패밀리, 멤버 A1), BGLAP (뼈 감마-카르복시글루타메이트 (gla) 단백질), MTTP (마이크로솜 트리글리세리드 전달 단백질), MTRR (5-메틸테트라히드로폴레이트-호모시스테인 메틸트랜스퍼라제 리덕타제), SULT1A3 (슬포트랜스퍼라제 패밀리, 시토졸, 1A, 페놀-선호, 멤버 3), RAGE (신장 종양 항원), C4B (보체 성분 4B (키도 혈액군), P2RY12 (푸린성 수용체 P2Y, G-단백질 커플링, 12), RNLS (레날라제, FAD-의존적 아민 옥시다제), CREB1 (cAMP 반응성 엘리먼트 결합 단백질 1), POMC (프로오피오멜라노코르틴), RAC1 (ras-관련 C3 보툴리늄 독소 기질 1 (rho 패밀리, 소형 GTP 결합 단백질 Rac1)), LMNA (라민 NC), CD59 (CD59 molecule, 보체 조절 단백질), SCN5A (소듐 채널, 전압-게이팅, V형, 알파 서브유닛), CYP1B1 (시토크롬 P450, 패밀리 1, 서브패밀리 B, 폴리펩티드 1), MIF (마크로파지 이동 억제성 인자 (글리코실화-억제 인자)), MMP13 (매트릭스 메탈로펩티다제 13 (콜라게나제 3)), TIMP2 (TIMP 메탈로펩티다제 억제제 2), CYP19A1 (시토크롬 P450, 패밀리 19, 서브패밀리 A, 폴리펩티드 1), CYP21A2 (시토크롬 P450, 패밀리 21, 서브패밀리 A, 폴리펩티드 2), PTPN22 (단백질 티로신 포스파타제, 비수용체 유형 22 (림프구)), MYH14 (미오신, 중쇄 14, 비근육), MBL2 (만노스-결합 렉틴 (단백질 C) 2, 가용성 (옵소닉 결함)), SELPLG (셀렉틴 P 리간드), AOC3 (아민 옥시다제, 구리 함유 3 (혈관 부착 단백질 1)), CTSL1 (카텝신 L1), PCNA (증식 세포 핵 항원), IGF2 (인슐린-유사 성장 인자 2 (소마토메딘 A)), ITGB1 (인테그린, 베타 1 (피브로넥틴 수용체, 베타 폴리펩티드, 항원 CD29은 MDF2, MSK12 포함)), CAST (칼파스타틴), CXCL12 (케모카인 (C-X-C 모티프) 리간드 12 (기질 세포-유래 인자 1)), IGHE (면역글로불린 중쇄 불변 엡실론), KCNE1 (포타슘 전압-게이팅 채널, Isk-관련 패밀리, 멤버 1), TFRC (트랜스페린 수용체 (p90, CD71)), COL1A1 (콜라겐, I형, 알파 1), COL1A2 (콜라겐, I형, 알파 2), IL2RB (인터루킨 2 수용체, 베타), PLA2G10 (포스포리파제 A2, 그룹 X), ANGPT2 (안지오포이어틴 2), PROCR (단백질 C 수용체, 상피 (EPCR)), NOX4 (NADPH 옥시다제 4), HAMP (헵시디니 항미생물성 펩티드), PTPN11 (단백질 티로신 포스파타제, 비수용체 유형 11), SLC2A1 (용질 운반체 패밀리 2 (촉진성 글루코스 수송체), 멤버 1), IL2RA (인터루킨 2 수용체, 알파), CCL5 (케모카인 (C-C 모티프) 리간드 5), IRF1 (인터페론 조절 인자 1), CFLAR (CASP8 및 FADD-유사 아폽토시스 조절제), CALCA (칼시토닌-관련 폴리펩티드 알파), EIF4E (진핵생물 번역 개시 인자 4E), GSTP1 (글루타티온 S-트랜스퍼라제 pi 1), JAK2 (Janus 키나제 2), CYP3A5 (시토크롬 P450, 패밀리 3, 서브패밀리 A, 폴리펩티드 5), HSPG2 (헤파란 술페이트 프로테오글리칸 2), CCL3 (케모카인 (C-C 모티프) 리간드 3), MYD88 (골수 분화 1차 반응 유전자 (88)), VIP (혈관작용성 장 펩티드), SOAT1 (스테롤 O-아실트랜스퍼라제 1), ADRBK1 (아드레날린성, 베타, 수용체 키나제 1), NR4A2 (핵 수용체 서브패밀리 4, 그룹 A, 멤버 2), MMP8 (매트릭스 메탈로펩티다제 8 (호중구 콜라게나제)), NPR2 (나트륨뇨설 펩티드 수용체 B/구아닐레이트 시클라제 B (심방나트륨이뇨 펩티드 수용체 B)), GCH1 (GTP 시클로히드롤라제 1), EPRS (글루타밀-프롤릴-tRNA 신써타제), PPARGC1A (퍼옥시솜 증식인자-활성화 수용체 감마, 공활성인자 1 알파), F12 (응고 인자 XII (하게만 인

자)), PECAM1 (혈소판/내피 세포 부착 분자), CCL4 (케모카인 (C-C 모티프) 리간드 4), SERPINA3 (세르핀 펩티다제 억제제, 클레이드 A (알파-1 안티프로테이나제, 안티트립신), 멤버 3), CASR (칼슘-감지 수용체), GJA5 (간극 연접 단백질, 알파 5, 40 kDa), FABP2 (지방산 결합 단백질 2, 장), TTF2 (전사 종결 인자, RNA 중합효소 II), PROS1 (단백질 S (알파)), CTF1 (카디오트로핀 1), SGCB (사르코글리칸, 베타 (43 kDa 디스트로핀-연관 당단백질)), YME1L1 (YME1-유사 1 (에스. 세레비지아에)), CAMP (카텔리시딘 항미생물성 펩티드), ZC3H12A (아연 핑거 CCCH-유형 함유 12A), AKR1B1 (알도-케토 리덕타제 패밀리 1, 멤버 B1 (알도스 리덕타제)), DES (데스민), MMP7 (매트릭스 메탈로펩티다제 7 (마트릴리신, 자궁)), AHR (아릴 탄화수소 수용체), CSF1 (콜로니 자극 인자 1 (마크로파지)), HDAC9 (히스톤 디아세틸라제 9), CTGF (결합 조직 성장 인자), KCNMA1 (포타슘 거대 컨덕턴스 칼슘-활성화 채널, 서브패밀리 M, 알파 멤버 1), UGT1A (UDP 글루쿠로노실트랜스퍼라제 1 패밀리, 폴리펩티드 A 복합 유전자좌), PRKCA (단백질 키나제 C, 알파), COMT (카테콜-.베타.-메틸트랜스퍼라제), S100B (S100 칼슘 결합 단백질 B), EGR1 (초기 성장 반응 1), PRL (프로락틴), IL15 (인터루킨 15), DRD4 (도파민 수용체 D4), CAMK2G (칼슘/칼모듈린-의존적 단백질 키나제 II 감마), SLC22A2 (용질 운반체 패밀리 22 (유기 양이온 수송체), 멤버 2), CCL11 (케모카인 (C-C 모티프) 리간드 11), PGF (B321 태반 성장 인자), THPO (트롬보포이어틴), GP6 (당단백질 VI (혈소판)), TACR1 (타카이키닌 수용체 1), NTS (뉴로텐신), HNF1A (HNF1 호메오박스 A), SST (소마토스타틴), KCND1 (포타슘 전압-게이팅 채널, Shal-관련 서브패밀리, 멤버 1), LOC646627 (포스포리파제 억제제), TBXAS1 (트롬복산 A 신타제 1 (혈소판)), CYP2J2 (시토크롬 P450, 패밀리 2, 서브패밀리 J, 폴리펩티드 2), TBXA2R (트롬복산 A2 수용체), ADH1C (알콜 디히드로게나제 1C (클래스 I), 감마 폴리펩티드), ALOX12 (아라키도네이트 12-리폭시게나제), AHSG (알파-2-HS-당단백질), BHMT (베타인-호모시스테인 메틸트랜스퍼라제), GJA4 (간극 연접 단백질, 알파 4, 37 kDa), SLC25A4 (용질 운반체 패밀리 25 (미토콘드리아 캐리어; 아데닌 뉴클레오티드 운반체), 멤버 4), ACLY (ATP 시트레이트 리아제), ALOX5AP (아라키도네이트 5-리폭시게나제-활성화 단백질), NUMA1 (핵 유사분열 장비 단백질 1), CYP27B1 (시토크롬 P450, 패밀리 27, 서브패밀리 B, 폴리펩티드 1), CYSLTR2 (시스테이닐 류코트리엔 수용체 2), SOD3 (수퍼옥시드 디스뮤타제 3, 세포외), LTC4S (류코트리엔 C4 신타제), UCN (우로코르틴), GHRL (그렐린/오베스타틴 프리프로펩티드), APOC2 (아포지단백질 C-II), CLEC4A (C-type 렉틴 도메인 패밀리 4, 멤버 A), KBTBD10 (kelch 반복 및BTB (POZ) 도메인 함유 10), TNC (테나신 C), TYMS (티미딜레이트 신써타제), SHC1 (SHC (Src 상동성 2 도메인 함유) 형질전환 단백질 1), LRP1 (저밀도 지단백질 수용체-관련 단백질 1), SOCS3 (사이토카인 신호전달의 억제인자 3), ADH1B (알콜 디히드로게나제 1B (클래스 I), 베타 폴리펩티드), KLK3 (칼리크레인-관련 펩티다제 3), HSD11B1 (히드록시스테로이드 (11-베타) 디히드로게나제 1), VKORC1 (비타민 K 에폭시드 리덕타제 복합체, 서브유닛 1), SERPINB2 (세르핀 펩티다제 억제제, 클레이드 B (오발부민), 멤버 2), TNS1 (텐신 1), RNF19A (링 핑거 단백질 19A), EPOR (에리쓰로포이어틴 수용체), ITGAM (인테그린, 알파 M (보체 성분 3 수용체 3 서브유닛)), PITX2 (쌍형성-유사 호메오도메인 2), MAPK7 (미토겐-활성화 단백질 키나제 7), FCGR3A (IgG의 Fc 단편, 저친화성 111a, 수용체 (CD16a)), LEPR (렙틴 수용체), ENG (엔도글린), GPX1 (글루타티온 퍼옥시다제 1), GOT2 (글루타믹-옥살로아세틱 트랜스아미나제 2, 미토콘드리아 (아스파테이트 아미노트랜스퍼라제 2)), HRH1 (히스타민 수용체 H1), NR112 (핵 수용체 서브패밀리 1, 그룹 I, 멤버 2), CRH (코르티코트로핀 방출 호르몬), HTR1A (5-히드록시트립타민 (세로토닌) 수용체 1A), VDAC1 (전압-의존적 음이온 채널 1), HPSE (헤파라나제), SFTPD (계면활성제 단백질 D), TAP2 (수송체 2, ATP-결합 카세트, 서브-패밀리 B (MDR/TAP)), RNF123 (링 핑거 단백질 123), PTK2B (PTK2B 단백질 티로신 키나제 2 베타), NTRK2 (신경영양성 티로신 키나제, 수용체, 2형), IL6R (인터루킨 6 수용체), ACHE (아세틸콜린스테라제 (Yt 혈액군)), GLP1R (글루카곤-유사 펩티드 1 수용체), GHR (성장 호르몬 수용체), GSR (글루타티온 리덕타제), NQO1 (NAD(P)H 디히드로게나제, 퀴논 1), NR5A1 (핵 수용체 서브패밀리 5, 그룹 A, 멤버 1), GJB2 (간극 연접 단백질, 베타 2, 26 kDa), SLC9A1 (용질 운반체 패밀리 9 (소듐/수소 교환체), 멤버 1), MAOA (모노아민 옥시다제 A), PCSK9 (프로단백질 컨버타제 서브틸리신/켁신 9형), FCGR2A (IgG의 Fc 단편, 저친화성 IIa, 수용체 (CD32)), SERPINF1 (세르핀 펩티다제 억제제, 클레이드 F (알파-2 안티플라스민, 피그먼트 상피 유래 인자), 멤버 1), EDN3 (엔도텔린 3), DHFR (디히드로폴레이트 리덕타제), GAS6 (성장 정지-특이적 6), SMPD1 (스핀고미엘린 포스포디에스터라제 1, 산 리소솜), UCP2 (언커플링 단백질 2 (미토콘드리아, 양자 캐리어)), TFAP2A (전사 인자 AP-2 알파 (활성화 인핸서 결합 단백질 2 알파)), C4BPA (보체 성분 4 결합 단백질, 알파), SERPINF2 (세르핀 펩티다제 억제제, 클레이드 F (알파-2 안티플라스민, 피그먼트 상피 유래 인자), 멤버 2), TYMP (티미딘 포스포릴라제), ALPP (알칼리 포스파타제, 태반 (Regan 이소자임)), CXCR2 (케모카인 (C-X-C 모티프) 수용체 2), SLC39A3 (용질 운반체 패밀리 39 (아연 수송체), 멤버 3), ABCG2 (ATP-결합 카세트, 서브-패밀리 G (WHITE), 멤버 2), ADA (아데노신 디아미나제), JAK3 (Janus 키나제 3), HSPA1A (열충격 70 kDa 단백질 1A), FASN (지방산 신타제), FGF1 (섬유아세포 성장 인자 1 (산성)), F11 (응고 인자 XI), ATP7A (ATPase, Cu++ 수송, 알파 폴리펩티드), CR1 (보체 성분 (3b/4b) 수용체 1 (Knops 혈액군)),

GFAP (신경교 섬유질 산성 단백질), ROCK1 (Rho-연관, 코일드-코일 함유 단백질 키나제 1), MECP2 (메틸 CpG 결합 단백질 2 (레트 증후군)), MYLK (미오신 경쇄 키나제), BCHE (부티릴콜린스테라제), LIPE (리파제, 호르몬-감응성), PRDX5 (퍼옥시리독신 5), ADORA1 (아데노신 A1 수용체), WRN (베르너 증후군, RecQ 헬리카제-유사), CXCR3 (케모카인 (C-X-C 모티프) 수용체 3), CD81 (CD81 molecule), SMAD7 (SMAD 패밀리 멤버 7), LAMC2 (라미닌, 감마 2), MAP3K5 (미토겐-활성화 단백질 키나제 키나제 키나제 5), CHGA (크로모그라닌 A (부갑상선 분비 단백질 1)), IAPP (섬 아밀로이드 폴리펩티드), RHO (로돕신), ENPP1 (엑토뉴클레오티드 파이로포스파타제/포스포디에스터라제 1), PTHLH (부갑상선 호르몬-유사 호르몬), NRG1 (뉴레귤린 1), VEGFC (혈관 내피 성장 인자 C), ENPEP (글루타밀 아미노펩티다제 (아미노펩티다제 A)), CEBPB (CCAAT/인핸서 결합 단백질 (C/EBP), 베타), NAGLU (N-아세틸글루코사미니다제, 알파-), F2RL3 (응고 인자 II (트롬빈) 수용체-유사 3), CX3CL1 (케모카인 (C-X3-C 모티프) 리간드 1), BDKRB1 (브라디키닌 수용체 B1), ADAMTS13 (트롬보스폰딘 1형 모티프 존재의 ADAM 메탈로펩티다제, 13), ELANE (엘라스타제, 호중구 발현), ENPP2 (엑토뉴클레오티드 파이로포스파타제/포스포디에스터라제 2), CISH (사이토카인 유도성 SH2-함유 단백질), GAST (가스트린), MYOC (미오실린, 섬유주대 유도성 글루코코르티코이드 반응), ATP1A2 (ATPase, Na+/K+ 수송, 알파 2 폴리펩티드), NF1 (뉴로피브로민 1), GJB1 (간극 연접 단백질, 베타 1, 32 kDa), MEF2A (근세포 인핸서 인자 2A), VCL (빈큘린), BMPR2 (뼈 형태형성성 단백질 수용체, II형 (세린/트레오닌 키나제)), TUBB (튜불린, 베타), CDC42 (세포 분열 주기 42 (GTP 결합 단백질, 25 kDa)), KRT18 (케라틴 18), HSF1 (열충격 전사 인자 1), MYB (v-myb 골수아구증 바이러스 종양 유전자 상동체 (조류)), PRKAA2 (단백질 키나제, AMP-활성화, 알파 2 촉매 서브유닛), ROCK2 (Rho-연관, 코일드-코일 함유 단백질 키나제 2), TFPI (조직 인자 경로 억제제 (지단백질-연관 응고 억제제)), PRKG1 (단백질 키나제, cGMP-의존성, I형), BMP2 (뼈 형태형성성 단백질 2), CTNND1 (카테닌 (카데린-연관 단백질), 델타 1), CTH (시스타티오나제 (시스타티오닌 감마-리아제)), CTSS (카텝신 S), VAV2 (vav 2 구아닌 뉴클레오티드 교환 인자), NPY2R (뉴로펩티드 Y 수용체 Y2), IGFBP2 (인슐린-유사 성장 인자 결합 단백질 2, 36 kDa), CD28 (CD28 분자), GSTA1 (글루타티온 S-트랜스퍼라제 알파 1), PPIA (펩티딜프롤릴 이소머라제 A (사이클로필린 A)), APOH (아포지단백질 H (베타-2-당단백질 I)), S100A8 (S100 칼슘 결합 단백질 A8), IL11 (인터루킨 11), ALOX15 (아라키도네이트 15-리폭시게나제), FBLN1 (피불린 1), NR1H3 (핵 수용체 서브패밀리 1, 그룹 H, 멤버 3), SCD (스테아로일-CoA 데사투라제 (델타-9-데사투라제)), GIP (위 억제성 폴리펩티드), CHGB (크로모그라닌 B (세크레토그라닌 1)), PRKCB (단백질 키나제 C, 베타), SRD5A1 (스테로이드-5-알파-리덕타제, 알파 폴리펩티드 1 (3-옥소-5 알파-스테로이드 델타 4-디히드로게나제 알파 1)), HSD11B2 (히드록시스테로이드 (11-베타) 디히드로게나제 2), CALCRL (칼시토닌 수용체-유사), GALNT2 (UDP-N-아세틸-알파-D-갈락토사민:폴리펩티드 N-아세틸갈락토사미닐트랜스퍼라제 2 (GalNAc-T2)), ANGPTL4 (안지오포이어틴-유사 4), KCNN4 (포타슘 중간체/소형 컨덕턴스 칼슘-활성화된 채널, 서브패밀리 N, 멤버 4), PIK3C2A (포스포이노시티드-3-키나제, 클래스 2, 알파 폴리펩티드), HBEGF (헤파린-결합 EGF-유사 성장 인자), CYP7A1 (시토크롬 P450, 패밀리 7, 서브패밀리 A, 폴리펩티드 1), HLA-DRB5 (주요 조직적합성 복합체, 클래스 II, DR 베타 5), BNIP3 (BCL2/아데노바이러스 E1B 19 kDa 상호작용 단백질 3), GCKR (글루코키나제 (헥소키나제 4) 조절제), S100A12 (S100 칼슘 결합 단백질 A12), PADI4 (펩티딜 아르기닌 디이미나제, IV형), HSPA14 (열충격 70 kDa 단백질 14), CXCR1 (케모카인 (C-X-C 모티프) 수용체 1), H19 (H19, 각인된 모성 발현 전사물 (비-단백질 코딩)), KRTAP19-3 (케라틴 연관 단백질 19-3), IDDM2 (인슐린-의존성 진성 당뇨병 2), RAC2 (ras-관련 C3 보툴리늄 독소 기질 2 (rho 패밀리, 소형 GTP 결합 단백질 Rac2)), RYR1 (리아노딘 수용체 1 (골격)), 시계 (시계 상동체 (마우스)), NGFR (신경 성장 인자 수용체 (TNFR 수퍼패밀리, 멤버 16)), DBH (도파민 베타-히드롤라제 (도파민 베타-모노옥시게나제)), CHRNA4 (콜린성 수용체, 니코틴성, 알파 4), CACNA1C (칼슘 채널, 전압-의존적, L형, 알파 1C 서브유닛), PRKAG2 (단백질 키나제, AMP-활성화, 감마 2 비촉매 서브유닛), CHAT (콜린 아세틸트랜스퍼라제), PTGDS (프로스타글란딘 D2 신타제 21 kDa (뇌)), NR1H2 (핵 수용체 서브패밀리 1, 그룹 H, 멤버 2), TEK (TEK 티로신 키나제, 상피), VEGFB (혈관 내피 성장 인자 B), MEF2C (근세포 인핸서 인자 2C), MAPKAPK2 (미토겐-활성화 단백질 키나제-활성화 단백질 키나제 2), TNFRSF11A (종양 괴사 인자 수용체 수퍼패밀리, 멤버 11a, NFKB 활성인자), HSPA9 (열충격 70 kDa 단백질 9 (몰탈린)), CYSLTR1 (시스테이닐 류코트리엔 수용체 1), MAT1A (메티오닌 아데노실트랜스퍼라제 I, 알파), OPRL1 (오피에이트 수용체-유사 1), IMPA1 (이노시톨(myo)-1(or 4)-모노포스파타제 1), CLCN2 (클로라이드 채널 2), DLD (디히드로리포아미드 디히드로게나제), PSMA6 (프로테오솜 (프로솜, 마크로파인) 서브유닛, 알파형, 6), PSMB8 (프로테오솜 (프로솜, 마크로파인) 서브유닛, 베타 유형, 8 (거대 다기능 펩티다제 7)), CHI3L1 (키티나제 3-유사 1 (연골 당단백질-39)), ALDH1B1 (알데히드 디히드로게나제 1 패밀리, 멤버 B1), PARP2 (폴리 (ADP-리보스) 중합효소 2), STAR (스테로이드성 급성 조절 단백질), LBP (지다당류 결합 단백질), ABCC6 (ATP-결합 카세트, 서브-패밀리 C(CFTR/MRP), 멤버 6), RGS2 (G-단백질 신호전달 조절제 2, 24 kDa), EFNB2 (에프린-B2), GJB6 (간극 연접 단백질, 베타 6, 30 kDa), APOA2 (아포지단백질 A-II), AMPD1 (아

데노신 모노포스페이트 디아미나제 1), DYSF (디스펠린, 지대형 근이영양증 2B (상염색체 열성)), FDFT1 (파르네실-디포스페이트 파르네실트랜스퍼라제 1), EDN2 (엔도텔린 2), CCR6 (케모카인 (C-C 모티프) 수용체 6), GJB3 (간극 연접 단백질, 베타 3, 31 kDa), IL1RL1 (인터루킨 1 수용체-유사 1), ENTPD1 (엑토뉴클레오시드 트리포스페이트 디포스포히드롤라제 1), BBS4 (프라뎃-비들 증후군 4), CELSR2 (카데린, EGF LAG 7-경로 G-유형 수용체 2 (프라민고 상동체, 초파리)), F11R (F11 수용체), RAPGEF3 (Rap 구아닌 뉴클레오티드 교환 인자 (GEF) 3), HYAL1 (히아루로노글루코사미니다제 1), ZNF259 (아연 핑거 단백질 259), ATOX1 (ATX1 항산화제 단백질 1 상동체 (효모)), ATF6 (활성화 전사 인자 6), KHK (케토헥소키나제 (프룩토키나제)), SAT1 (스퍼미딘/스퍼민 N1-아세틸트랜스퍼라제 1), GGH (감마-글루타밀 히드롤라제 (콘쥬가제, 폴릴폴리감마글루타밀 히드롤라제)), TIMP4 (TIMP 메탈로펩티다제 억제제 4), SLC4A4 (용질 운반체 패밀리 4, 소듐 바이카보네이트 공수송체, 멤버 4), PDE2A (포스포디에스터라제 2A, cGMP-자극), PDE3B (포스포디에스터라제 3B, cGMP-억제), FADS1 (지방산 데사투라제 1), FADS2 (지방산 데사투라제 2), TMSB4X (티모신 베타 4, X-연관), TXNIP (티오레독신 상호작용 단백질), LIMS1 (LIM 및 노화 세포 항원-유사 도메인 1), RHOB (ras 상동체 유전자 패밀리, 멤버 B), LY96 (림프구 항원 96), FOXO1 (포크헤드 박스 O1), PNPLA2 (파타틴-유사 포스포리파제 도메인 함유 2), TRH (트로트로핀-방출 호르몬), GJC1 (간극 연접 단백질, 감마 1, 45 kDa), SLC17A5 (용질 운반체 패밀리 17 (음이온/당 수송체), 멤버 5), FTO (지방량 및 비만 연관), GJD2 (간극 연접 단백질, 델타 2, 36 kDa), PSRC1 (프롤린/세린-풍부 코일드-코일 1), CASP12 (캐스파제 12 (유전자-가유전자)), GPBAR1 (G 단백질-커플링 담즙산 수용체 1), PXK (PX 도메인 함유 세린/트레오닌 키나제), IL33 (인터루킨 33), TRIB1 (tribbles 상동체 1 (초파리)), PBX4 (pre-B-세포 백혈병 호메오박스 4), NUPR1 (핵 단백질, 전사 조절제, 1), 15-Sep(15 kDa 셀레노단백질), CILP2 (연골 중간층 단백질 2), TERC (텔로머라제 RNA 성분), GGT2 (감마-글루타미트랜스퍼라제 2), MT-CO1 (미토콘드리아 코딩 시토크롬 c 옥시다제 I), and UOX (우레이트 옥시다제, 가유전자). 추가 구현예에서, 염색체 서열은 Pon1(파라옥소나제 1), LDLR (LDL 수용체), ApoE(아포지방단백질 E), Apo B-100(아포지방단백질 B-100), ApoA (아포지방단백질(a)), ApoA1(아포지방단백질 A1), CBS(시스타티온 B-신타제), 글리코단백질 IIb/IIb, MTHRF(5,10-메틸렌테트라하이드로폴레이트 환원효소(NADPH), 및 이들의 조합으로부터 추가로 선택될 수 있다. 일 반복에서, 심혈관 질환에 연루된 염색체 서열, 및 염색체 서열에 의해 암호화된 단백질은 Cacna1C, Sod1, Pten, Ppar(알파), Apo E, 렙틴, 및 이들의 조합으로부터 선택될 수 있다. 따라서 본 명세서의 내용은 CRISPR 또는 CRISPR-Cas 시스템 또는 복합체에 관한 예시적인 표적을 제공한다.

[1733] **면역 직교 오솔로그**

[1734] 일부 구현예에서, CRISPR 효소가 대상체에서 발현되거나 또는 투여되는 것이 필요할 때, CRISPR 효소의 면역원성은 대상체에게 CRISPR 효소의 면역 직교 오솔로그를 순차적으로 발현 또는 투여하여 감소시킬 수 있다. 본 명세서에서 사용되는 용어 "면역 직교 오솔로그" 는 유사하거나 또는 실질적으로 동일한 기능 또는 활성을 갖지만, 서로에 의해 생성된 면역 반응과 교차-반응성은 거의 없거나 또는 전혀 없는 오솔로그 단백질을 의미한다. 이러한 오솔로그의 순차적 발현 또는 투여는 강력하거나 또는 임의의 부수적 면역 반응을 유발시키지 않을 것이다. 면역 직교 오솔로그는 현존 항체에 의한 중화를 피할 수 있다. 오솔로그를 발현하는 세포는 숙주의 면역계 (예를 들어, 활성화된 CTL)에 의한 제거를 피할 수 있다. 일부 예에서, 상이한 종으로부터의 CRISPR 효소 오솔로그가 면역 직교 오솔로그일 수 있다.

[1735] 면역 직고 오솔로그는 후보 오솔로그 세트의 서열, 구조, 및 면역원성을 분석하여 동정될 수 있다. 예시적 방법에서, 면역 직교 오솔로그 세트는 a) 서열 유사성이 낮거나 또는 전혀 없는 후보물의 서브세트를 확인하기 위해서 후보 오솔로그 세트 (예를 들어, 상이한 종 유래 오솔로그)의 서열을 비교하는 단계; b) 면역 중복이 없거나 또는 낮은 후보물을 확인하기 위해 후보물의 서브세트의 멤버 중에 면역 중복을 평가하는 단계에 의해 동정될 수 있다. 일부 경우에서, 후보물 간 면역 중복은 후보 오솔로그 및 MHC (예를 들어, MHC I형 및/또는 MHC II) 간 결합을 결정하여 평가할 수 있다. 대안적으로 또는 추가로, 후보물간 면역 중복은 후보 오솔로그의 B-세포 에피토프를 결정하여 평가될 수 있다. 일 예에서, 면역 직교 오솔로그는 [Moreno AM et al., BioRxiv](2018년 1월 10일 온라인 공개, doi: doi.org/10.1101/245985)에 기술된 방법을 사용해 동정될 수 있다.

[1736] 본 출원은 또한 하기 번호 표시된 문장에 기재된 바와 같은 양상 및 구현예를 제공한다:

[1737] 1. i) 표 1 또는 2로 부터의 Cas12b 이펙터 단백질, ii) 표적 서열에 하이브리드화할 수 있는 가이드 서열을 포함하는 가이드를 포함하는, 비천연 발생 또는 조작된 시스템.

[1738] 2. 문장 1의 시스템에 있어서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스 (Alicyclobacillus

kakegawensis), 바실러스 (Bacillus) sp. V3-13, 바실러스 히사시이 (Bacillus hisashii), 렌티스파에리아 박테리움 (Lentisphaeria bacterium), 및 라세이엘라 세디미니스 (Laceyella sediminis)로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다.

[1739] 3. 문장 1 또는 2의 시스템에 있어서, tracr RNA는 직접 반복부 서열의 5' 말단에서 crRNA에 융합된다.

[1740] 4. 전술한 문장 중 어느 하나의 시스템에 있어서, 2종의 상이한 표적 서열 또는 동일 표적 서열의 상이한 영역에 하이브리드화할 수 있는 둘 이상의 가이드 서열을 포함한다.

[1741] 5. 전술한 문장 중 어느 하나의 시스템에 있어서, 가이드 서열은 원핵생물 세포 내 하나 이상의 표적 서열에 하이브리드화한다.

[1742] 6. 전술한 문장 중 어느 하나의 시스템에 있어서, 가이드 서열은 진핵생물 세포 내 하나 이상의 표적 서열에 하이브리드화한다.

[1743] 7. 전술한 문장 중 어느 하나의 시스템에 있어서, Cas12b 이펙터 단백질은 하나 이상의 핵 국재화 신호 (NLS)를 포함한다.

[1744] 8. 전술한 문장 중 어느 하나의 시스템에 있어서, Cas12b 이펙터 단백질은 촉매적으로 불활성이다.

[1745] 9. 전술한 문장 중 어느 하나의 시스템에 있어서, Cas12b 이펙터 단백질은 하나 이상의 기능성 도메인과 회합된다.

[1746] 10. 문장 9의 시스템에 있어서, 하나 이상의 기능성 도메인은 하나 이상의 표적 서열을 절단한다.

[1747] 11. 문장 10의 시스템에 있어서, 기능성 도메인은 하나 이상의 표적 서열의 전사 또는 번역을 변형시킨다.

[1748] 12. 전술한 문장 중 어느 하나의 시스템에 있어서, Cas12b 이펙터 단백질은 하나 이상의 기능성 도메인과 회합되고; Cas12b 이펙터 단백질은 RuvC 및/또는 Nuc 도메인 내에 하나 이상의 돌연변이를 함유하여서, 형성된 CRISPR 복합체는 표적 서열에서 또는 그에 인접하여 후성적 변형자 또는 전사 또는 번역 활성화 또는 억제 신호를 전달할 수 있다.

[1749] 13. 전술한 문장 중 어느 하나의 시스템에 있어서, Cas12b 이펙터 단백질은 아데노신 디아미나제 또는 시티딘 디아미나제와 회합된다.

[1750] 14. 전술한 문장 중 어느 하나의 시스템에 있어서, 재조합 주형을 더 포함한다.

[1751] 15. 제14항의 시스템에 있어서, 재조합 주형은 상동성-지정 복구 (HDR)에 의해 삽입된다.

[1752] 16. 표 1 또는 2로부터의 Cas12b 이펙터 단백질을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제1 조절 엘리먼트, 및 i) a) 가이드 서열을 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트, 및 b) tracr RNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제3 조절 엘리먼트, 또는 ii) 가이드 서열 및 tracr RNA를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 제2 조절 엘리먼트를 포함하는 하나 이상의 벡터를 포함하는, Cas12b 벡터 시스템.

[1753] 17. 문장 16의 벡터 시스템에 있어서, Cas12b 이펙터 단백질을 코딩하는 뉴클레오티드 서열은 진핵생물 세포에서 발현을 위해 코돈 최적화된다.

[1754] 18. 문장 16 또는 17의 벡터 시스템에 있어서, 단일 벡터에 포함된다.

[1755] 19. 문장 17 내지 18 중 어느 하나의 벡터 시스템에 있어서, 하나 이상의 벡터는 바이러스 벡터를 포함한다.

[1756] 20. 문장 17 내지 19 중 어느 하나의 벡터 시스템에 있어서, 하나 이상의 벡터는 하나 이상의 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관 또는 헤르페스 심플렉스 바이러스 벡터를 포함한다.

[1757] 21. i) 표 1 또는 2로부터 선택된 Cas12b 이펙터 단백질, ii) 하나 이상의 표적 서열에 하이브리드화할 수 있는 가이드 서열, 및 iii) tracr RNA를 포함하는 비천연 발생 또는 조작된 조성물의 하나 이상의 핵산 성분 및 Cas12b 이펙터 단백질을 전달하도록 구성된 전달 시스템.

[1758] 22. 문장 21의 전달 시스템에 있어서, 하나 이상의 벡터, 또는 하나 이상의 폴리뉴클레오티드 분자를 포함하고, 하나 이상의 벡터 또는 폴리뉴클레오티드 분자는 비천연 발생 또는 조작된 조성물의 하나 이상의 핵산 성분 및 Cas12b 이펙터 단백질을 코딩하는 하나 이상의 폴리뉴클레오티드 분자를 포함한다.

[1759] 23. 문장 21 또는 22의 전달 시스템에 있어서, 리포솜(들), 입자(들), 엑소솜(들), 미세소포(들), 유전자총, 또는 바이러스 벡터(들)를 포함하는 전달 비히클을 포함한다.

[1760] 24. 치료적 치료 방법에서 사용을 위한 문장 1 내지 15 중 어느 하나의 비천연 발생 또는 조작된 시스템, 문장 16 내지 20 중 어느 하나의 벡터 시스템, 또는 문장 21 내지 23 중 어느 하나의 전달 시스템.

[1761] 25. 하나 이상의 관심 표적 서열을 변형시키는 방법으로서, 방법은 하나 이상의 표적 서열을 i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질, ii) 하나 이상의 표적 서열에 하이브리드화할 수 있는 가이드 서열, 및 iii) tracr RNA를 포함하는 하나 이상의 비천연 발생 또는 조작된 조성물과 접촉시켜서, crRNA 및 tracr RNA와 복합체를 형성하는 Cas12b 이펙터 단백질을 포함하는 CRISPR 복합체가 형성되는 것인 단계를 포함하고, 가이드 서열은 세포에서 하나 이상의 표적 서열과의 서열-특이적 결합을 유도하여서, 하나 이상의 표적 서열의 발현이 변형된다.

[1762] 26. 문장 25의 방법에 있어서, 하나 이상의 표적 서열의 변형은 표적 DNA의 절단을 포함한다.

[1763] 27. 문장 25 또는 26의 방법에 있어서, 하나 이상의 표적 서열의 변형은 하나 이상의 표적 서열의 발현 증가 또는 감소를 포함한다.

[1764] 28. 문장 25 내지 27 중 어느 하나의 방법에 있어서, 조성물은 재조합 주형을 더 포함하고, 하나 이상의 표적 서열의 변형은 재조합 주형 또는 이의 부분의 삽입을 포함한다.

[1765] 29. 문장 25 내지 28 중 어느 하나의 방법에 있어서, 표적 유전자는 원핵생물 세포에 존재한다.

[1766] 30. 문장 25 내지 29 중 어느 하나의 방법에 있어서, 하나 이상의 표적 서열은 진핵생물 세포에 존재한다.

[1767] 31. 변형된 관심 표적을 포함하는 세포 또는 이의 자손으로서, 하나 이상의 표적 서열은 문장 25 내지 30 중 어느 하나에 따른 방법으로 변형되었고, 임의로 치료적 T 세포 또는 항체-생성 B-세포이거나 또는 상기 세포는 식물 세포이다.

[1768] 32. 문장 31의 세포에 있어서, 세포는 원핵생물 세포이다.

[1769] 33. 문장 31의 세포에 있어, 세포는 진핵생물 세포이다.

[1770] 34. 문장 31 내지 33 중 어느 하나에 따른 세포에 있어서, 하나 이상의 표적 서열의 변형은 적어도 하나의 유전자 산물의 변경된 발현을 포함하는 세포; 적어도 하나의 유전자 산물의 변경된 발현을 포함하고, 적어도 하나의 유전자 산물의 발현이 증가된 것인 세포; 또는 적어도 하나의 유전자 산물의 변경된 발현을 포함하고, 적어도 하나의 유전자 산물의 발현이 감소된 것인 세포; 또는 내생성 또는 비내생성 생물학적 산물 또는 화학적 화합물을 생산 및/또는 분비하는 세포 또는 개체군을 초래한다.

[1771] 35. 문장 31 또는 34에 따른 진핵생물 세포에 있어서, 세포는 포유동물 세포 또는 인간 세포이다.

[1772] 36. 문장 31 내지 35 중 어느 하나에 따른 세포를 포함하는 세포주, 또는 이의 자손.

[1773] 37. 문장 31 내지 35 중 어느 하나에 따른 하나 이상의 세포를 포함하는 다세포 유기체.

[1774] 38. 문장 31 내지 35 중 어느 하나에 따른 하나 이상의 세포를 포함하는 식물 또는 동물 모델.

[1775] 39. 문장 31 내지 35 중 어느 하나의 세포로부터의 유전자 산물 또는 문장 36의 세포주 또는 문장 37의 유기체 또는 문장 38의 식물 또는 동물 모델.

[1776] 40. 문장 39의 유전자 산물로서, 발현된 유전자 산물의 양은 변경된 발현을 갖지 않는 세포로부터의 유전자 산물의 양을 초과하거나 또는 그 미만이다.

[1777] 41. 표 1 또는 2로부터의 단리된 Cas12b 이펙터 단백질.

[1778] 42. 문장 41의 Cas12b 이펙터 단백질을 코딩하는 단리된 핵산.

[1779] 43. 문장 42에 따른 단리된 핵산으로서, DNA이고 crRNA 및 tracr RNA를 코딩하는 서열을 더 포함한다.

[1780] 44. 문장 42 또는 43에 따른 핵산 또는 문장 41의 Cas12b를 포함하는 단리된 진핵생물 세포.

[1781] 45. i) 표 1 또는 2로부터의 Cas12b 이펙터 단백질을 코딩하는 mRNA, ii) 가이드 서열, 및 iii) tracr RNA를 포함하는 비천연 발생 또는 조작된 시스템.

[1782] 46. 문장 45에 따른 비천연 발생 또는 조작된 시스템으로서, tracr RNA는 직접 반복부의 5' 말단에서 crRNA에 융합된다.

[1783] 47. 표적화 도메인 및 아데노신 디아미나제, 시티딘 디아미나제, 또는 이의 촉매 도메인을 포함하는 부위 지정 염기 편집을 위한 조작된 조성물로서, 표적화 도메인은 Cas12b 이펙터 단백질, 또는 올리고뉴클레오티드-결합 활성을 보유한 이의 단편, 및 가이드 분자를 포함한다.

[1784] 48. 문장 47의 조성물로서, Cas12b 이펙터 단백질은 촉매적으로 불활성이다.

[1785] 49. 문장 47 또는 48의 조성물로서, Cas12b 이펙터 단백질은 표 1 또는 2로부터 선택된다.

[1786] 50. 문장 47-49 중 어느 하나의 조성물로서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp. V3-13, 바실러스 히사시이, 렌티스파에리아 박테리움, 및 라세이엘라 세디미니스로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다.

[1787] 51. 하나 이상의 관심 표적 올리고뉴클레오티드의 아데노신 또는 시티딘을 변형시키는 방법으로서, 상기 하나 이상의 표적 올리고뉴클레오티드에 문장 47 내지 50 중 어느 하나에 따른 조성물을 전달하는 단계를 포함한다.

[1788] 52. 문장 51의 방법에 있어서, 병원성 T→C 또는 A→G 점 돌연변이를 함유하는 전사물에 의해 초개된 질환의 치료 또는 예방에서 사용을 위한 것이다.

[1789] 53. 문장 51 또는 52 중 어느 하나의 방법으로 수득되고/되거나 문장 47-50 중 어느 하나의 조성물을 포함하는 단리된 세포.

[1790] 54. 문장 53의 세포 또는 이의 자손에 있어서, 상기 진핵생물 세포는, 바람직하게 인간 또는 비인간 동물 세포, 임의로 치료적 T 세포 또는 항체-생성 B-세포이거나 또는 상기 세포는 식물 세포이다.

[1791] 55. 문장 53 또는 54의 상기 변형된 세포 또는 이의 자손을 포함하는 비인간 동물.

[1792] 56. 문장 54의 상기 변형된 세포를 포함하는 식물.

[1793] 57. 요법, 바람직하게 세포 요법에서 사용을 위한 문장 53 또는 54에 따른 변형된 세포.

[1794] 58. 표적 올리고뉴클레오티드의 아데닌 또는 시토신을 변형시키는 방법으로서, 상기 표적 올리고뉴클레오티드에: 촉매적 불활성 Cas12b 단백질; 직접 반복부에 연결된 가이드 서열을 포함하는 가이드 분자; 및 아데노신 또는 시티딘 디아미나제 단백질 또는 이의 촉매 도메인을 전달하는 단계를 포함하고; 상기 아데노신 또는 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 상기 촉매적 불활성 Cas12b 단백질에 공유적으로 또는 비공유적으로 연결되거나 또는 상기 가이드 분자는 전달 후에 이에 연결되도록 적합화되고; 상기 가이드 분자는 상기 촉매적 불활성 Cas12b와 복합체를 형성하고 상기 복합체가 상기 표적 올리고뉴클레오티드에 결합하도록 유도하며, 상기 가이드 서열은 상기 표적 올리고뉴클레오티드 내 표적 서열과 하이브리드할 수 있어서 올리고뉴클레오티드 듀플렉스를 형성한다.

[1795] 59. 문장 58의 방법에 있어서, (A) 상기 시토신은 상기 올리고뉴클레오티드 듀플렉스를 형성하는 상기 표적 서열 밖에 존재하고, 상기 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 상기 올리고뉴클레오티드 듀플렉스 밖의 상기 시토신을 탈아미노화시키거나, 또는 (B) 상기 시토신은 상기 올리고뉴클레오티드 듀플렉스를 형성하는 상기 표적 서열 내에 존재하고, 상기 가이드 서열은 상기 시토신에 상응하는 위치에서 비쌍형성 아데닌 또는 우라실을 포함하여 상기 RNA 듀플렉스에 C-A 또는 C-U 미스매치를 생성시키고, 시티딘 디아미나제 단백질 또는 이의 촉매 도메인은 비쌍형성 아데닌 또는 우라실 반대쪽 올리고뉴클레오티드 듀플렉스 내 시토신을 탈아미노화시킨다.

[1796] 60. 문장 58 또는 59의 방법에 있어서, 상기 아데노신 디아미나제 단백질 또는 이의 촉매 도메인은 상기 올리고뉴클레오티드 듀플렉스 내 상기 아데닌 또는 시토신을 탈아미노화시킨다.

[1797] 61. 문장 58-60 중 어느 하나의 방법에 있어서, Cas12b 단백질은 표 1 또는 2로부터 선택된다.

[1798] 62. 문장 61의 방법에 있어서, Cas12b 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp. V3-13, 바실러스 히사시이, 렌티스파에리아 박테리움, 및 라세이엘라 세디미니스로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다.

[1799] 63. 하나 이상의 시험관내 샘플에서 핵산 올리고뉴클레오티드 표적 서열의 존재를 검출하기 위한 시스템으로서,

Cas12b 단백질; 하나 이상의 표적 서열과 일정 정도의 상보성을 갖도로 설계된 가이드 서열을 포함하고, Cas12b 단백질과 복합체를 형성하도록 설계된 적어도 하나의 가이드 폴리뉴클레오티드; 및 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체를 포함하고; Cas12b 단백질은 부차적 뉴클레아제 활성을 나타내고 하나 이상의 표적 서열에 의해 활성화되면 올리고뉴클레오티드 기반 차폐성 구성체의 비표적 서열을 절단한다.

[1800] 64. 하나 이상의 시험관내 샘플에서 표적 폴리펩티드의 존재를 검출하기 위한 시스템으로서, Cas12b 단백질; 하나 이상의 검출 압타머로서, 각각 하나 이상의 표적 폴리펩티드 중 하나에 결합하도록 설계되고, 각각의 검출 압타머는 차폐된 프로모터 결합 부위 또는 차폐된 프라이머 결합 부위 및 기폭제 서열 주형을 포함하는 것인 하나 이상의 압타머; 및 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체를 포함한다.

[1801] 65. 문장 64 또는 65의 시스템에 있어서, 표적 서열 또는 기폭제 서열을 증폭시키기 위한 핵산 증폭 시약을 더 포함한다.

[1802] 66. 문장 652의 시스템에 있어서, 핵산 증폭 시약은 등온 증폭 시약이다.

[1803] 67. 문장 63 내지 66 중 어느 하나의 시스템에 있어서, Cas12b 단백질은 표 1 또는 2로부터 선택된다.

[1804] 68. 문장 67의 시스템에 있어서, Cas12b 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp. V3-13, 바실러스 히사시이, 렌티스파에리아 박테리움, 및 라세이엘라 세디미니스로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다.

[1805] 69. 하나 이상의 시험관내 샘플에서 하나 이상의 서열을 검출하기 위한 방법에 있어서, 하나 이상의 샘플을 i) Cas12b 이펙터 단백질, ii) 하나 이상의 표적 서열과 일정 정도의 상보성을 갖도록 설계된 가이드 서열을 포함하고, Cas12b 이펙터 단백질과 복합체를 형성하도록 설계된 적어도 하나의 가이드 폴리뉴클레오티드; 및 iii) 비표적 서열을 포함하는 올리고뉴클레오티드-기반 차폐성 구성체와 접촉시키는 단계를 포함하고, 상기 Cas12 이펙터 단백질은 부차적 뉴클레아제 활성을 나타내고 올리고뉴클레오티드-기반 차폐성 구성체의 비표적 서열을 절단한다.

[1806] 70. 문장 69의 방법에 있어서, Cas12b 이펙터 단백질은 표 1 또는 2로부터 선택된다.

[1807] 71. 문장 70의 방법에 있어서, Cas12b 이펙터 단백질은 알리시클로바실러스 카케가웬시스, 바실러스 sp. V3-13, 바실러스 히사시이, 렌티스파에리아 박테리움, 및 라세이엘라 세디미니스로 이루어진 군으로부터 선택되는 박테리아로부터 기원한다.

[1808] 72. 효소 또는 리포터 모이어티의 불활성 제1 부분에 연결된 Cas12b 단백질을 포함하는 비천연 발생 또는 조작된 조성물로서, 효소 또는 리포터 모이어티는 효소 또는 리포터 모이어티의 상보성 부분과 접촉될 때 재구성된다.

[1809] 73. 문장 72의 조성물에 있어서, 효소 또는 리포터 모이어티는 단백질가수분해 효소를 포함한다.

[1810] 74. 문장 72 또는 73의 조성물에 있어서, Cas12 단백질은 제1 Cas12b 단백질 및 효소 또는 리포터 모이어티의 상보성 부분에 연결된 제2 Cas12b 단백질을 포함한다.

[1811] 75. 문장 72-74 중 어느 하나의 조성물에 있어서, 제1 Cas12b 단백질과 복합체를 형성할 수 있고 표적 핵산의 제1 표적 서열에 하이브리드화할 수 있는 제1 가이드; 및 ii) 제2 Cas12b 단백질과 복합체를 형성할 수 있고, 표적 핵산의 제2 표적 서열에 하이브리드화할 수 있는 제2 가이드를 더 포함한다.

[1812] 76. 문장 72-75 중 어느 하나의 조성물에 있어서, 효소는 캐스파제를 포함한다.

[1813] 77. 문장 72-75 중 어느 하나의 조성물에 있어서, 효소는 담배 식각 바이러스 (TEV)를 포함한다.

[1814] 78. 표적 올리고뉴클레오티드를 함유하는 세포에서 단백질가수분해 활성을 제공하는 방법으로서, a) 세포 또는 세포 개체군을, i) 단백질가수분해 효소의 불활성 부분에 연결된 제1 Cas12b 이펙터 단백질; ii) 단백질가수분해 효소의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 단백질가수분해 효소의 단백질가수분해 활성은 단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 이펙터 단백질; iii) 제1 Cas12b 이펙터 단백질에 결합하고 표적 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드; 및 iv) 제2 Cas12b 이펙터 단백질에 결합하고 표적 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드와 접촉시켜서, 단백질 가수분해 효소의 제1 부분 및 상보성 부분이 접촉하고 단백질가수분해 효소의 단백질가수분해 활성이 재구성되는 것인 단계를 포함한다.

[1815] 79. 문장 78의 방법에 있어서, 단백질가수분해 효소는 캐스파제이다.

[1816] 80. 문장 79의 방법에 있어서, 단백질가수분해 효소는 TEV 프로테아제이고, TEV 프로테아제의 단백질가수분해 활성은 재구성되어서, TEV 기질이 절단되고 활성화된다.

[1817] 81. 문장 80의 방법에 있어서, TEV 기질은 TEV 표적 서열을 함유하도록 조작된 프로캐스파제여서, TEV 프로테아제에 의한 절단은 프로캐스파제를 활성화시킨다.

[1818] 82. 관심 올리고뉴클레오티드를 함유하는 세포를 확인하는 방법으로서, 방법은 세포 내 올리고뉴클레오티드를, i) 단백질가수분해 효소의 불활성 제1 부분에 연결된 제1 Cas12b 이펙터 단백질; ii) 단백질가수분해 효소의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 단백질가수분해 효소의 활성은 단백질가수분해 효소의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 이펙터 단백질; iii) 제1 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드; iv) 제2 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드; 및 v) 검출가능하게 절단되는 리포터를 포함하는 조성물과 접촉시키는 단계를 포함하고, 단백질가수분해 효소의 제1 부분 및 상보성 부분은 관심 올리고뉴클레오티드가 세포에 존재할 때 접촉하여서, 단백질가수분해 효소의 활성이 재구성되고 리포터를 검출가능하게 절단한다.

[1819] 83. 관심 올리고뉴클레오티드를 함유하는 세포를 확인하는 방법으로서, 바아법은 세포 내 올리고뉴클레오티드를, i) 리포터의 불활성 제1 부분에 연결된 제1 Cas12b 이펙터 단백질; ii) 리포터의 상보성 부분에 연결된 제2 Cas12b 이펙터 단백질로서, 리포터의 활성은 리포터의 제1 부분 및 상보성 부분이 접촉될 때 재구성되는 것인 이펙터 단백질; iii) 제1 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제1 표적 서열에 하이브리드화하는 제1 가이드; iv) 제2 Cas12b 이펙터 단백질에 결합하고 올리고뉴클레오티드의 제2 표적 서열에 하이브리드화하는 제2 가이드; 및 v) 리포터를 포함하는 조성물과 접촉시키는 단계를 포함하고, 리포터의 제1 부분 및 상보성 부분은 관심 올리고뉴클레오티드가 세포에 존재할 때 접촉하여서, 리포터의 활성이 재구성된다.

[1820] 84. 문장 82 또는 83에 있어서, 리포터는 형광 단백질 또는 발광 단백질이다.

[1821] 본 발명은 청구항에 기술된 본 발명의 범주를 제한하지 않는, 하기 실시예에서 더욱 설명된다.

[1822] **실시예:**

[1823] **실시예 1:**

[1824] 표 11은 예시적인 C2c1 오솔로그의 아미노산 서열을 표시한다.

[1825]

표 11 - C2c1 오솔로그

숙주	아미노산 서열
알리시클로바실러스 마크로스포란 지이더스 균주 DSM 17980 (SEQ ID NO:484)	MVAVKSIKVKLMLGHLPEIREGLWHLHEAVNLGVRYYTEWLAL LRQGNLYRRGKDGAQECYMTAEQCRQELLVRLRDRQKRNGHT GDPGTDEELLGVARRLYELLVPQSVGKKGQAQMLASGFLSPLAD PKSEGGKGTSKSGRKPAWMGMKEAGDSRWVEAKARYEANKAK DPTKQVIASLEMYGLRPLFDVFTETYKTIRWMPLGKHQGVRAWD RDMFQQSLERLMSWESWNERVGAEFARLVDRRDRFREKHFTGQ EHLVALAQRLEQEMKEASPGFESKSSQAHRITKRALRGADGIIDD WLKLSEGEPVDRFDEILRKRQAQNPRRFGSHDLFLKLAEPVFQPL WREDPSFLSRWASYNEVLNKLEDAKQFATFTLPSPCSNPVWARF ENAEGTNIFKYDFLFDHFGKGRHGVRFQRMIVMRDGVPTEVEGI

	VVPIAPSRQLDALAPNDAASPIDVFVGDPAAPGAFRGQFGGAKIQ YRRSALVRKGRREEKAYLCGFRLPSQRRTGTPADDAGEVFLNLS LRVESQSEQAGRRNPPYAAVFHISDQTRRVIVRYGEIERYLAEHP DTGIPGSRGLTSGLRVMSVDLGLRTSAAISVFRVAHRDELTPDAH GRQPFFFPIHGMDHLVALHERSHLIRLPGETESKKVRSIREQRLDR LNRLRSQMASLRLLVRTGVLDEQKRDRNWERLQSSMERGGERM PSDWWDLFQAQVRYLAQHRDASGEAWGRMVQAAVRTLWRQL AKQVRDWRKEVRRNADKVKIRGIARDVPGGHSLAQLDYLERQY RFLRSWSAFSVQAGQVVRAERDSRFAVALREHIDNGKKDRLKKL ADRILMEALGYVYVTDGRRAGQWQAVYPPCQLVLLEELSEYRFS NDRPPSENSQLMVWSHRGVLEELIHQAQVHDVLVGTIPAAFSSRF DARTGAPGIRCRRVPSIPLKDAPSIPIWLSHYLKQTERDAAALRPG ELIPTGDGEFLVTPAGRGASGVRVVHADINAAHNLQRRLWENFD LSDIRVRCDRREGKDGTVVLIPRLTNQRVKERYSGVIFTSEDGVSF TVGDAKTRRRSSASQGEGDDLSDEEQELLAEADDARERSVVLFR DPSGFVNGGRWTAQRAFWGMVHNRIETLLAERFSVSGAAEKVR G
바실러스 히사시이 균주 C4 (SEQ ID NO:485)	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAI YEHHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKD EVFNILRELYEELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKG TASSGRKPRWYNLKIAGDPSWEEEKKKWEEDKKKDPLAKILGKL AEYGLIPLFIPYTDSNEPIVKEIKWMEKSRNQSVRRLDKDMFIQAL ERFLSWESWNLKVKEEYEKVEKEYKTLEERIKEDIQALKALEQY EKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPS EKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPY LYATFCEIDKKKKDAKQQATFTLADPINHPLWVRFEERSGSNLNK YRILTEQLHTEKLKKKLTVQLDRLIYPTESGGWEEKGKVDIVLLP SRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFDRD HLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPK VVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAA AASIFEVVDQKPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVK SREVLRKAREDNLKLMNQKLNFLRNVLHFQQFEDITEREKRVTK WISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLKQLHKRLE VEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLR PTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALG YCYDVRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMK WSRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSV VTKEKLQDNRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFIS LSKDRKCVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQVDG QTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWVNAGKLKIKKGS SKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKW MAAGVFFGKLERILISKLTNQYSISTIEDDSSKQSM
칸디다투스 린도우박테리아 박테리움	MPRDDLDLLTNLNSTAKGIRERGKTKEGTDKKKSGRKSSWPMD KAAWETAKTSDSSAHFLEKLKQHPDLKDAFGNLSSGGSKKLEYY KKLAGSAPWKESQSVILEKAARWKEAKQEREEKEQDSSEHGSKA

[1826]

RIFCSPLOWO2 (SEQ ID NO:486)	AYRRLFDAGCLPMPEFAKYIDENQIEFGDLKLSDCGAEWKRGM WNQAGQRVRSHMGWQRRREKENAVYSLRKELFEKGGAIRRKKS EELTPEDILPGKAAPDQNDWQERPAYGNQMWFIGLRSYEENEMA KYAEEAGMGSRSAPRIRRGTIKGWSKLRERWLQILKRNPQATRD DLIGELNALRSQDPRAYGDARLFDWLSKTDQRFLWDGFDADGKI LCGRDDRDCVSAFVAYNEEFADEPSSITLTETDERLHPVWPFFGE SSAVPYEIEYDLETACPTAIRLPLLVGKENGGYAERQGTRLPLAE YADLASSFQLPTPVRLDVLVEIREVTRAGRKVTCPFSYFKQNGV WYVREGEIPSGESIQIKQTDRKIENGKIFISSKLRMAYRDDLMVSP ATGDFGSIKILWERIELASHVDQKKLPETAPARSRVFVSFSCNVVE RAPRKQLTRKPDAVVVTIPSGVDQGLVVVSTDVRTGKSKSSSAPP LPPGSRLWPADAVHGDPPLRILSVDLGHRHSAYAVWELGLQQKS WRAGVLKGSTQTPVYADCTGTGLLCLPGDGEDTPAEEESLRLRS RQIRRRLNLQNSILRVSRLLSLDKFEKTIFEQSDVRDRPNKKGLRI RRRCRTEKTPLSEAEVRKNCDKAAEILIRWADTDAMAKSLAATG NADISFWKYMAVKNPPLSAVVDVAPSTIVPDDGPDRETLKKKRQ EEEEKFASSIYENRVKLAGALCSGYDADHRRPATGGLWHDLDRT LIREISYGDRGQKGNPRKLNNEGILRLLRRPPRARPDWREFHRTL NDANRIPKGRTLRGGLSMGRLNFLKEVGDFVKKWSCRPRWPGD RRHIPPGQLFDRQDAEHLEHLRDDRIKRLAHLIVAQALGFEPDIRR GLWKYVDGSTGEILWQHPETRRFFAEGAAGELREVSRPAEIDDD AAARPHTVSAPAHIVVFENLIRYRFQSDRPKTENAGLMQWAHRQ IVHFTKQVASLYGLKVAMVYAAFSSKFCSRCGSPGARVSRFDPA WRNQEWFKRRTSNPRSKVDHSLKRASEDPTADETRPWVLIEGGK EFVCANAKCSAHDEPLNADENAAANIGLRFLRGVEDFRTKVNPA GALKGKLRFETGIHSFRPPVSGSPFWSPMAEPAQKKKIGAAAPGA DVDEAGDADESGVVVLFRDPSGAFRNKQYWYEGKIFWSNVMM AVEAKIAGASVGAKPVAASWGQAQPQSGPGLAKPGGD
엘루시미크로비아 박테리움 RIFOXYA12 (SEQ ID NO:487)	MNRIYQGRVTKVEVPDGKDEKGNIKWKKLENWSDILWQHHMLF QDAVNYYTLALAAISGSAVGSDEKSIILREWAVQVQNIWEKAKK KATVFEGPQKRLTSILGLEQNASFDIAAKHILRTSEAKPEQRASAL IRLLEEIDKKNHNVVCGERLPFFCPRNIQSKRSPTSKAVSSVQEQK RQEEVRRFHNMQPEEVVKNAVTLDISLFKSSPKIVFLEDPKKARA ELLKQFDNACKKHKELVGIKKAFTESIDKHGSSLKVPAPGSKPSG LYPSAIVFKYFPVDITKTVFLKATEKLAMGKDREVTNDPIADARV NDKPHFDYFTNIALIREKEKNRAAWFEFDLAAFIEAIMSPHRFYQ DTQKRKEAARKLEEKIKAIEGKGGQFKESDSEDDDVDSLPGFEGD TRIDLLRKLVTDTLGWLGESETPDNNEGKKTEYSISERTLRIFPDIQ KQWSELAEKGETTEGKLLEVLKHEQTEHQSDFGSATLYQHLAKP EFHPIWLKSGTEEWHAENPLKAWLNYKELQYELTDKKRPIHFTP AHPVYSPRYFDFPKKSETEEKEVSKNTHSLTTSLASEHIKNSLQFT AGLIRKTNVGKKAIKARFSYSAPRLRRDCLRSENNENLYKAPWL QPMMRALGIDEEKADRQNFANTRITLMAKGLDDIQLGFPVEANS QELQKEVSNGISWKGQFNWGGIASLSALRWPHEKKPKNPPEQPW WGIDSFSCLAVDLGQRYAGAFARLDVSTIEKKGKSRFIGEACDKK WYAKVSRMGLLRLPGEDVKVWRDASKIDKENGFAFRKELFGEK

[1827]

	GRSATPLEAEETAELIKLFGANEKDVMPDNWSKELSFPEQNDKLL IVARRAQAAVSRLHRWAWFFDEAKRSDDAIREILESDDTDLKQK VNKNEIEKVKETIISLLKVKQELLPTLLTRLANRVLPLRGRSWEW KKHHQKNDGFILDQTGKAMPNVLIRGQRGLSMDRIEQITELRKRF QALNQSLRRQIGKKAPAKRDDSIPDCCPDLLEKLDHMKEQRVNQ TAHMILAEALGLKLAEPPKDKKELNETCDMHGAYAKVDNPVSFI VIEDLSRYRSSQGRSPRENSRLMKWCHRAVRDKLKEMCEVFFPL CERRKAGSAWVSLPPLLETPAAYSSRFCSRSGVAGFRAVEVIPGF ELKYPWSWLKDKKDKAGNLAKEALNIRTVSEQLKAFNQDKPEK PRTLLVPIAGGPIFVPISEVGLSSFGLKPQVVQADINAAINLGLRAIS DPRIWEIHPRLRTEKRDGRLFAREKRKYGEEKVEVQPSKNEKAK KVKDDRKPNYFADFSGKVDWGFGNIKNESGLTLVSGKALWWTI NQLQWERCFDINKRHIEDWSNKQKQ
옴니트로피카 WOR_2 박테리움 RIFCSPHIGHO 2 (SEQ ID NO:488)	MNRIYQGRVTKVEKLKNGKSPDDREELKDWQTALWRHHELFQD AVSYYTLALAAMAEGLPDKHPINVLRKRMEEAWEEFPRKTVTPA KNLRDSVRPWLGLSESASFGDALKKILPPAPENKEVRALAVALLA EKARTLKPQKTSASYWGRFCDDLKKKPNWDYSEEELARKTGSG DWVAGLWSEDALNKIDELAKSLKLSSLVKCVPDGQINPEGARNL VKEALDHLEGVSNGTKKEKNDPGPAKKTNNWLRQHASDVRNFI HKNKNQFSSLPNGRLITERARGGGININKTYAGVLFKAFPCPFTFD YVRAAVPEPKVKKVDQEKKSEQSATWTELEKRILRIGDDPIELAR KNNKPIFKAFTALEKWSDQNSKSCWSDFDKCAFEEALKTLNQFN QKTEEREKRRSEAEAELKYMMDENPEWKPKKETEGDDVREVPIL KGDPRYEKLVKLFGDLDEEGSEHATGKIYGPSRASLRGFGKLRNE WVDLFTKANDNPREQDLQKAVTGFQREHKLDMGYTAFFLKLCE RDYWDIWRDDTEVEVKKIREKRWVKSVVYAAADTRELAEELER LQEPVRYTPAEPQFSRRLFMFSDIKGKQGAKHIREGLVEVSLAVK DQSGKYGTCRVRLHYSAPRLIRDHLSDGSSSMWLQPMMAALGL SSDARGCFTRDSKGNVKEPAVALMSDFVGRKRELRMLLNFPVDL DISKLEENIGKKARWEKQMNTAYEKNKLKQRFHLIWPGMELKET QEPGQFWWDNPTIQKEGMYCLAIDLSQRRAADYALLHAGVNRD SKTFVELGQAGGQSWFTKLCAAGSLRLPGEDTEVIREGKRQIELS GKKGRNATQSEYDQAIALAKQLLHNENSAELESAARDWLGDNA KRFSFPEQNDKLIDLYYGALSRYKTWLRWSWRLTEQHKELWDK TLDEIRKVPYFASWGELAGNGTNEATVQQLQKLIADAAVDLRNF LEKALLHIAYRALPLRENTWRWIENGKDGKGKPLHLLVSDGQSP AEIPWLRGQRGLSIARIEQLENFRRAVLSLNRLLRHEIGTKPEFGSS TCGESLPDPCPDLTDKIVRLKEERVNQTAHLIIAQSLGVRLKGHSL FTEEREKADMHGEHEVIPGRSPVDFVVLEDLSRYTTDKSRSRSEN SRLMKWCHRKINEKVKLLAEPFGIPVIEVFASYSSKFDARTGAPG FRAVEVTSEDRPFWRKTIEKQSVAREVFDCLDNLVGKGLNGIHL VLPQNGGPLFIAAVKEDQPLPAIRQADINAAVNIGLRAIAGPSCYH AHPKVRLIKGESGTDKGKWLPRKGKEANKRENAQFGNVDLDLE VKFNRLDIDSDVLKGDNTNLFHDPLNIACYGFATIQNLQHPFLAH ASAVFSRQKGAVARLQWEVCRAINSRRLEAWQKKAEKAAVKR

[1828]

피시스파에라에 박테리움 ST-NAGAB-D1 (SEQ ID NO:489)	MATKSYRARILTDSRLAAALDRTHVVFVESLKQMINTYLRMQNG KFGPDHKKLAQIMLSRSNTFAHGVMDQITRDQPTSTLDEEWTDL ARRIHKTTGPLFLQAERFATVKNRAIHTKSRGKVIPSPETLAVPAK FWHQVCDSASAYIRSNRELMQQWRKDRAAWLKDKNEWQQKHP EFMQFYNGPYQNFLKLCDDDRITSQLAAEQQPTASKNNRPRKTG KRFARWHLWYKWLSENPEIIEWRNKASASDFKTVTDDVRKQIIT KYPQQNKYITRLLDWLEDNNPELKTLENLRRTYVKKFDSFKRPPT LTLPSPYRHPYWFTMELDQFYKKADFENGTIQLLLIDEDDDGNW FFNWMPASLKPDPRLVPSWRAETFETEGRFPPYLGGKIGKKLSRP APTDAERKAGIAGAKLMIKNNRSELLFTVFEQDCPPRVKWAKTK NRKCPADNAFSSDGKTRKPLRILSIDLGIRHIGAFALTQGTRNDSA WQTESLKKGIINSPSIPPLRQVRRHDYDLKRKRRRHGKPVKGQRS NANLQAHRTNMAQDRFKKGASAIVSLAREHSADLILFENLHSLK FSAFDERWMNRQLRDMNRRHIVELVSEQAPEFGITVKDDINPWM TSRICSNCNLPGFRFSMKKKNPYREKLPREKCTDFGYPVWEPGGH LFRCPHCDHRVNADINAAANLANKFFGLGYWNNGLKYDAETKT FTVHTDKKTPPLIFKPRPQFDLWADSVKTRKQLGPDPF
플란크토마이세테스 박테리움 RBG_13_46_10 (SEQ ID NO:490)	MSVRSFQARVECDKQTMEHLWRTHKVFNERLPEIIKILFKMKRG ECGQNDKQKSLYKSISQSILEANAQNADYLLNSVSIKGWKPGTA KKYRNASFTWADDAAKLSSQGIHVYDKKQVLGDLPGMMSQMV CRQSVEAISGHIELTKKWEKEHNEWLKEKEKWESEDEHKKYLDL REKFEQFEQSIGGKITKRRGRWHLYLKWLSDNPDFAAWRGNKA VINPLSEKAQIRINKAKPNKKNSVERDEFFKANPEMKALDNLHGY YERNFVRRRKTKKNPDGFDHKPTFTLPHPTIHPRWFVFNKPKTNP EGYRKLILPKKAGDLGSLEMRLLTGEKNKGNYPDDWISVKFKAD PRLSLIRPVKGRRVVRKGKEQGQTKETDSYEFFDKHLKKWRPAK LSGVKLIFPDKTPKAAYLYFTCDIPDEPLTETAKKIQWLETGDVTK KGKKRKKKVLPHGLVSCAVDLSMRRGTTGFATLCRYENGKIHIL RSRNLWVGYKEGKGCHPYRWTEGPDLGHIAKHKREIRILRSKRG KPVKGEESHIDLQKHIDYMGEDRFKKAARTIVNFALNTENAASK NGFYPRADVLLLENLEGLIPDAEKERGINRALAGWNRRHLVERVI EMAKDAGFKRRVFEIPPYGTSQVCSKCGALGRRYSIIRENNRREIR FGYVEKLFACPNCGYCANADHNASVNLNRRFLIEDSFKSYYDWK RLSEKKQKEEIETIESKLMDKLCAMHKISRGSISK
스피로카에테스 박테리움 GWB1_27_13 (SEQ ID NO:491)	MSFTISYPFKLIIKNKDEAKALLDTHQYMNEGVKYYLEKLLMFR QEKIFIGEDETGKRIYIEETEYKKQIEEFYLIKKTELGRNLTLTLDEF KTLMRELYICLVSSSMENKKGFPNAQQASLNIFSPLFDAESKGYIL KEENNNISLIHKDYGKILLKRLRDNNLIPIFTKFTDIKKITAKLSPTA LDRMIFAQAIEKLLSYESWCKLMIKERFDKEVKIKELENKCENKQ ERDKIFEILEKYEEERQKTFEQDSGFAKKGKFYITGRMLKGFDEIK EKWLKEKDRSEQNLINILNKYQTDNSKLVGDRNLFEFIIKLENQC LWNGDIDYLKIKRDINKNQIWLDRPEMPRFTMPDFKKHPLWYRY EDPSNSNFRNYKIEVVKDENYITIPLITERNNEYFEENYTFNLAKL KKLSENITFIPKSKNKEFEFIDSNDEEEDKKDQKKSKQYIKYCDTA KNTSYGKSGGIRLYFNRNELENYKDGKKMDSYTVFTLSIRDYKS

[1829]

	LFAKEKLQPQIFNTVDNKITSLKIQKKFGNEEQTNFLSYFTQNQIT KKDWMDEKTFQNVKELNEGIRVLSVDLGQRFFAAVSCFEIMSEI DNNKLFFNLNDQNHKIIRINDKNYYAKHIYSKTIKLSGEDDDLYK ERKINKNYKLSYQERKNKIGIFTRQINKLNQLLKIIRNDEIDKEKFK ELIETTKRYVKNTYNDGIIDWNNVDNKILSYENKEDVINLHKELD KKLEIDFKEFIRECRKPIFRSGGLSMQRIDFLEKLNKLKRKWVART QKSAESIVLTPKFGYKLKEHINELKDNRVKQGVNYILMTALGYIK DNEIKNDSKKKQKEDWVKKNRACQIILMEKLTEYTFAEDRPREE NSKLRMWSHRQIFNFLQQKASLWGILVGDVFAPYTSKCLSDNNA PGIRCHQVTKKDLIDNSWFLKIVVKDDAFCDLIEINKENVKNKSIK INDILPLRGGELFASIKDGKLHIVQADINASRNIAKRFLSQINPFRV VLKKDKDETFHLKNEPNYLKNYYSILNFVPTNEELTFFKVEENKD IKPTKRIKMDKHEKESTDEGDDYSKNQIALFRDDSGIFFDKSLWV DGKIFWSVVKNKMTKLLRERNNKKNGSK
베루코미크로비아세아에 박테리움 UBA2429 (SEQ ID NO:492)	MPLSRIYQGRTNSLIILTPTPQEPWDHKALARFDSPLWRHHALFQ DAVNYYQLCLVALASSDGTRPLSKLHEQMKASWDEAKTDTEDS WRVRLARRLGIPAASLFEAALAKVLEGNEAPERARELAGELLLD KIEGDIQQAGRGYWPRFCDPKANPTYDYSATARASASGLTKLAA VIHAENVTEEALKQVAAEMDLSWTVKLQPDKNFVGAEARARLL EAAHHFIKVAESPPTKLAEVLARFPDGLALWQALPEKIAALPEET QVPRNRKASPDLTFATLLFQHFPSLFTAAVLGLSVGKPKSVKAPK VVEKVSARRKANAVTQAVVIEEPEIDFAELGDDPIKLARGERGFV FPAFTSLSFWAVPGPHVPVWKEFDIAAFKEALKTVNQFKLKTSER NALLAEAQRRLDYMDEKTHDWKTGDSDEPGHIPPRLKSDPNFTL IQALTQDEGVSNKATGDQHIPKGVYTGGLRGFYAIKKDWCELWE RKADKSQGTPTEEELISIVTDYQRDHVYDVGDVGLFRALCEPRF WPLWQPLTDEQEAERIKAGRAKDMISAYRVWLELQEDVVRLAQ PIRFTPAHAENSRRLFMFSDISGSHGAEFGSDGKSLEVSIAYDVDG KLQPVRAKLEFSAPRAARDELEGLSGGSESMRWFQPMMKALDC PEVEMPALEKCAVSLMPDVVKKGGGKWVRLLLNFPATLEPEGLI RHIGKQAMWYKQFNGTYKPRTQQLDTGLHLYWPGLEKAPEAED AAAWWNREEIRAKGFSVLSVDLGQRDAGAWALLESRSDKAFSR NRQPFIELGEAGGKLWSTALLGLGMLRLPGEDARTGALDDQGKR AVEFHGKAGRNALEAEWQEAREMALLFGGEEAKSRLGPGFDHL SHSKQNEELLRILSRAQSRLARFHRWSCRIHEKPEATGDDVIDYG QVDELLTKTAEAMLENLKALYTNAGGILDSKSKQPLTLVGLRKK LEAQKVEPEKIAAVLKPHAEIIFQRLGTLIPELKQHLRVSLERLAN RELPLRHREWVVWNEAFEKLEQGNFKKEENPKWIRGQRGLSMARI EQIENLRKRFMSLRRQMSLIPGEQVKQGVEDKGQRQPEPCEDILN KLDRMKQQRVNQTAHLILAQALGLRLRPHLANDAEREEKDIHGE YELIPGRKPVDFIVMEDLSRYLSSQGRAPSENGRLMKWCHRAVL AKLKQMCEPFGIPVLEVPAAYSSRFCALTGVPGFRAVEVHDGNA EDFRWKRLIKKAEKDKSSKDAEAAAMLFDQLHDLNIEAREARKQ DKKLPLRTLFAPVAGGPLFIPMVGGGPRQADMNAAINLGLRAIAS PTCLRARPKIRAELKDGKHQAMLGNKLEKAAALTLEPPKEPTKE LAAQKRTNFFLDEKFVGKFDTAHVTTSGKKLRLSGGMSLWKAIK

[1830]

	DGAWQRVKKINDARIAKWKNNPPPEPDPDDEIQF
알리시클로바실러스 카케가웬시스 (SEQ ID NO:493)	MAVKSIKVKLRLSECPDILAGMWQLHRATNAGVRYYTEWVSLM RQEILYSRGPDGGQQCYMTAEDCQRELLRRLRNRQLHNGRQDQP GTDADLLAISRRLYEILVLQSIGKRGDAQQIASSFLSPLVDPNSKG GRGEAKSGRKPAWQKMRDQGDPRWVAAREKYEQRKAVDPSKE ILNSLDALGLRPLFAVFTETYRSGVDWKPLGKSQGVRTWDRDMF QQALERLMSWESWNRRVGEEYARLFQQKMKFEQEHFAEQSHLV KLARALEADMRAASQGFEAKRGTAHQITRRALRGADRVFEIWKS IPEEALFSQYDEVIRQVQAEKRRDFGSHDLFAKLAEPKYQPLWRA DETFLTRYALYNGVLRDLEKARQFATFTLPDACVNPIWTRFESSQ GSNLHKYEFLFDHLGPGRHAVRFQRLLVVESEGAKERDSVVVPV APSGQLDKLVLREEEKSSVALHLHDTARPDGFMAEWAGAKLQY ERSTLARKARRDKQGMRSWRRQPSMLMSAAQMLEDAKQAGDV YLNISVRVKSPSEVRGQRRPPYAALFRIDDKQRRVTVNYNKLSAY LEEHPDKQIPGAPGLLSGLRVMSVDLGLRTSASISVFRVAKKEEV EALGDGRPPHYYPIHGTDDLVAVHERSHLIQMPGETETKQLRKLR EERQAVLRPLFAQLALLRLLVRCGAADERIRTRSWQRLTKQGRE FTKRLTPSWREALELELTRLEAYCGRVPDDEWSRIVDRTVIALWR RMGKQVRDWRKQVKSGAKVKVKGYQLDVVGGNSLAQIDYLEQ QYKFLRRWSFFARASGLVVRADRESHFAVALRQHIENAKRDRLK KLADRILMEALGYVYEASGPREGQWTAQHPPCQLIILEELSAYRF SDDRPPSENSKLMAWGHRGILEELVNQAQVHDVLVGTVYAAFSS RFDARTGAPGVRCRRVPARFVGATVDDSLPLWLTEFLDKHRLDK NLLRPDDVIPTGEGEFLVSPCGEEAARVRQVHADINAAQNLQRRL WQNFDITELRLRCDVKMGGEGTVLVPRVNNARAKQLFGKKVLV SQDGVTFFERSQTGGKPHSEKQTDLTDKELELIAEADEARAKSVV LFRDPSGHIGKGHWIRQREFWSLVKQRIESHTAERIRVRGVGSSL D
바실러스 sp._V3-13 (SEQ ID NO:494)	MAIRSIKLKMKTNSGTDSIYLRKALWRTHQLINEGIAYYMNLLTL YRQEAIGDKTKEAYQAELINIIRNQQRNNGSSEEHGSDQEILALLR QLYELIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNAGRKP RWKRLKEEGNPDWELEKKKDEERKAKDPTVKIFDNLNKYGLLP LFPLFTNIQKDIEWLPLGKRQSVRKWDKDMFIQAIERLLSWESWN RRVADEYKQLKEKTESYYKEHLTGGEEWIEKIRKFEKERNMELE KNAFAPNDGYFITSRQIRGWDRVYEKWSKLPESASPEELWKVVA EQQNKMSEGFGDPKVFSFLANRENRDIWRGHSERIYHIAAYNGL QKKLSRTKEQATFTLPDAIEHPLWIRYESPGGTNLNLFKLEEKQK KNYYVTLSKIIWPSEEKWIEKENIEIPLAPSIQFNRQIKLKQHVKGK QEISFSDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFF NLVVDVAPLQETRNGRLQSPIGKALKVISSDFSKVIDYKPKELMD WMNTGSASNSFGVASLLEGMRVMSIDMGQRTSASVSIFEVVKEL PKDQEQKLFYSINDTELFAIHKRSFLLNLPGEVVTKNNKQQRQER RKKRQFVRSQIRMLANVLRLETKKTPDERKKAIHKLMEIVQSYDS WTASQKEVWEKELNLLTNMAAFNDEIWKESLVELHHRIEPYVG QIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRLLISWSKRSRTP GEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIMTALGFKYD

[1831]

	KEEKDRYKRWKETYPACQIILFENLNRYLFNLDRSRRENSRLMK WAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCH ALTEEDLKAGSNTLKRLIEDGFINESELAYLKKGDIIPSQGGELFV TLSKRYKKDSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPC QLARMGEDKLYIPKSQTETIKKYFGKGSFVKNNTEQEVYKWEKS EKMKIKTDTTFDLQDLDGFEDISKTIELAQEQQKKYLTMFRDPSG YFFNNETWRPQKEYWSIVNNIIKSCLKKKILSNKVEL
데술파틸하브듐 부티라티보란스 (SEQ ID NO:495)	MPLSNNPPVTQRAYTLRLRGADPSDLSWREALWHTHEAVNKGA KVFGDWLLTLRGGLDHTLADTKVKGGKGKPDRDPTPEERKARRI LLALSWLSVESKLGAPSSYIVASGDEPAKDRNDNVVSALEEILQS RKVAKSEIDDWKRDCSASLSAAIRDDAVWVNRSKVFDEAVKSV GSSLTREEAWDMLERFFGSRDAYLTPMKDPEDKSSETEQEDKAK DLVQKAGQWLSSRYGTSEGADFCRMSDIYGKIAAWADNASQGG SSTVDDLVSELRQHFDTKESKATNGLDWIIGLSSYTGHTPNPVHE LLRQNTSLNKSHLDDLKKKANTRAESCKSKIGSKGQRPYSDAILN DVESVCGFTYRVDKDGQPVSVADYSKYDVDYKWGTARHYIFAV MLDHAARRISLAHKWIKRAEAERHKFEEDAKRIANVPARAREWL DSFCKERSVTSGAVEPYRIRRRAVDGWKEVVAAWSKSDCKSTED RIAAARALQDDSEIDKFGDIQLFEALAEDDALCVWHKDGEATNE PDFQPLIDYSLAIEAEFKKRQFKVPAYRHPDELLHPVFCDFGKSR WKINYDVHKNVQAPFYRGLCLTLWTGSEIKPVPLCWQSKRLTRD LALGNNHRNDAASAVTRADRLGRAASNVTKSDMVNITGLFEQA DWNGRLQAPRQQLEAIAVVRDNPRLSEQERNLRMCGMIEHIRWL VTFSVKLQPQGPWCAYAEQHGLNTNPQYWPHADTNRDRKVHA RLILPRLPGLRVLSVDLGHRYAAACAVWEAVNTETVKEACQNV GRDMPKEHDLYLHIKVKKQGIGKQTEVDKTTIYRRIGADTLPDG RPHPAPWARLDRQFLIKLQGEEKDAREASNEEIWALHQMECKLD RTKPLIDRLIASGWGLLKRQMARLDALKELGWIPAPDSSENLSRE DGEAKDYRESLAVDDLMFSAVRTLRLALQRHGNRARIAYYLISE VKIRPGGIQEKLDENGRIDLLQDALALWHELFSSPGWRDEAAKQ LWDSRIATLAGYKAPEENGDNVSDVAYRKKQQVYREQLRNVAK TLSGDVITCKELSDAWKERWEDEDQRWKKLLRWFKDWVLPSGT QANNATIRNVGGLSLSRLATITEFRRKVQVGFFTRLRPDGTRHEIG EQFGQKTLDALELLREQRVKQLASRIAEAALGIGSEGGKGWDGG KRPRQRINDSRFAPCHAVVIENLANYRPDETRTRLENRRLMTWS ASKVHKYLSEACQLNGLYLCTVSAWYTSRQDSRTGAPGIRCQDV SVREFMQSPFWRKQVKQAEAKHDENKGDARERFLCELNKTWKA KTPAEWKKAGFVRIPLRGGEIFVSADSKSPSAKGIHADLNAAANI GLRALTDPDWPGKWWYVPCDPVSFESKMDYVKGCAAVKVGQP LRQPAQTNADGAASKIRKGKKNRTAGTSKEKVYLWRDISAFPLE SNEIGEWKETSAYQNDVQYRVIRMLKEHIKSLDNRTGDNVEG
데술포나트로늄 티오디스뮤탄스 (SEQ ID	MVLGRKDDTAELRRALWTTHEHVNLAVAEVERVLLRCRGRSY WTLDRRGDPVHVPESQVAEDALAMAREAQRRNGWPVVGEDEEI LLALRYLYEQIVPSCLLDDLGKPLKGDAQKIGTNYAGPLFDSDTC RRDEGKDVACCGPFHEVAGKYLGALPEWATPISKQEFDGKDASH

[1832]

NO:496)	LRFKATGGDDAFFRVSIEKANAWYEDPANQDALKNKAYNKDD WKKEKDKGISSWAVKYIQKQLQLGQDPRTEVRRKLWLELGLLPL FIPVFDKTMVGNLWNRLAVRLALAHLLSWESWNHRAVQDQALA RAKRDELAALFLGMEDGFAGLREYELRRNESIKQHAFEPVDRPY VVSGRALRSWTRVREEWLRHGDTQESRKNICNRLQDRLRGKFG DPDVFHWLAEDGQEALWKERDCVTSFSLLNDADGLLEKRKGYA LMTFADARLHPRWAMYEAPGGSNLRTYQIRKTENGLWADVVLL SPRNESAAVEEKTFNVRLAPSGQLSNVSFDQIQKGSKMVGRCRY QSANQQFEGLLGGAEILFDRKRIANEQHGATDLASKPGHVWFKL TLDVRPQAPQGWLDGKGRPALPPEAKHFKTALSNKSKFADQVRP GLRVLSVDLGVRSFAACSVFELVRGGPDQGTYFPAADGRTVDDP EKLWAKHERSFKITLPGENPSRKEEIARRAAMEELRSLNGDIRRL KAILRLSVLQEDDPRTEHLRLFMEAIVDDPAKSALNAELFKGFGD DRFRSTPDLWKQHCHFFHDKAEKVVAERFSRWRTETRPKSSSWQ DWRERRGYAGGKSYWAVTYLEAVRGLILRWNMRGRTYGEVNR QDKKQFGTVASALLHHINQLKEDRIKTGADMIIQAARGFVPRKN GAGWVQVHEPCRLILFEDLARYRFRTDRSRRENSRLMRWSHREI VNEVGMQGELYGLHVDTTEAGFSSRYLASSGAPGVRCRHLVEED FHDGLPGMHLVGELDWLLPKDKDRTANEARRLLGGMVRPGML VPWDGGELFATLNAASQLHVIHADINAAQNLQRRFWGRCGEAIR IVCNQLSVDGSTRYEMAKAPKARLLGALQQLKNGDAPFHLTSIP NSQKPENSYVMTPTNAGKKYRAGPGEKSSGEEDELALDIVEQAE ELAQGRKTFFRDPSGVFFAPDRWLPSEIYWSRIRRRIWQVTLERN SSGRQERAEMDEMPY
렌티스파에리아 박테리움 (SEQ ID NO:497)	MAVELNRIYQGRVNHVYIFDENQNQVSVDNGDDLLFVHHELYQ DAINYYLVALAAMALDSKDSLFGKFKMQIRAVWNDFYRNGQLR PGLKHSLIRSLGHAAELNTSNGADIAMNLILEDGGIPSEILNAALE HLAEKCTGDVSQLGKTFFPRFCDTAYHGNWDVDAKSFSEKKGR QRLVDALYSLHPVQAVQELAPEIEIGWGGVKTQTGKFFTGDEAK ASLKKAISYFLQDTGKNSPELQEYFSVAGKQPLEQYLGKIDTFPEI SFGRISSHQNINISNAMWILKFFPDQYSVDLIKNLIPNKKYEIGIAP QWGDDPVKLSRGKRGYTFRAFTDLAMWEKNWKVFDRAAFSDA LKTINQFRNKTQERNDQLKRYCAALNWMDGESSDKKPPVEPAD ADAVDEAATSVLPILAGDKRWNALLQLQKELGICNDFTENELMD YGLSLRTIRGYQKLRSMMLEKEEKMRAKTADDEEISQALQEIIIKF QSSHRDTIGSVSLFLKLAEPKYFCVWHDADKNQNFASVDMVAD AVRYYSYQEEKARLEEPIQITPADARYSRRVSDLYALVYKNAKE CKTGYGLRPDGNFVFEIAQKNAKGYAPAKVVLAFSAPRLKRDGL IDKEFSAYYPPVLQAFLREEEAPKQSFKTTAVILMPDWDKNGKRR ILLNFPIKLDVSAIHQKTDHRFENQFYFANNTNTCLLWPSYQYKK PVTWYQGKKPFDVVAVDLGQRSAGAVSRITVSTEKREHSVAIGE AGGTQWYAYRKFSGLLRLPGEDATVIRDGQRTEELSGNAGRLST EEETVQACVLCKMLIGDATLLGGSDEKTIRSFPKQNDKLLIAFRR ATGRMKQLQRWLWMLNENGLCDKAKTEISNSDWLVNKNIDNV LKEEKQHREMLPAILLQIADRVLPLRGRKWDWVLNPQSNSFVLQ QTAHGSGDPHKKICGQRGLSFARIEQLESLRMRCQALNRILMRKT

[1833]

	GEKPATLAEMRNNPIPDCCPDILMRLDAMKEQRINQTANLILAQA LGLRHCLHSESATKRKENGMHGEYEKIPGVEPAAFVVLEDLSRY RFSQDRSSYENSRLMKWSHRKILEKLALLCEVFNVPILQVGAAYS SKFSANAIPGFRAEECSIDQLSFYPWRELKDSREKALVEQIRKIGH RLLTFDAKATIIMPRNGGPVFIPFVPSDSKDTLIQADINASFNIGLR GVADATNLLCNNRVSCDRKKDCWQVKRSSNFSKMVYPEKLSLS FDPIKKQEGAGGNFFVLGCSERILTGTSEKSPVFTSSEMAKKYPNL MFGSALWRNEILKLERCCKINQSRLDKFIAKKEVQNEL
라세이엘라 세디미니스 (SEQ ID NO:498)	MSIRSFKLKIKTKSGVNAEELRRGLWRTHQLINDGIAYYMNWLV LLRQEDLFIRNEETNEIEKRSKEEIQGELLERVHKQQQRNQWSGE VDDQTLLQTLRHLYEEIVPSVIGKSGNASLKARFFLGPLVDPNNK TTKDVSKSGPTPKWKKMKDAGDPNWVQEYEKYMAERQTLVRL EEMGLIPLFPMYTDEVGDIHWLPQASGYTRTWDRDMFQQAIERL LSWESWNRRVRERRAQFEKKTHDFASRFSESDVQWMNKLREYE AQQEKSLEENAFAPNEPYALTKKALRGWERVYHSWMRLDSAAS EEAYWQEVATCQTAMRGEFGDPAIYQFLAQKENHDIWRGYPER VIDFAELNHLQRELRRAKEDATFTLPDSVDHPLWVRYEAPGGTNI HGYDLVQDTKRNLTLILDKFILPDENGSWHEVKKVPFSLAKSKQF HRQVWLQEEQKQKKREVVFYDYSTNLPHLGTLAGAKLQWDRN FLNKRTQQQIEETGEIGKVFFNISVDVRPAVEVKNGRLQNGLGKA LTVLTHPDGTKIVTGWKAEQLEKWVGESGRVSSLGLDSLSEGLR VMSIDLGQRTSATVSVFEITKEAPDNPYKFFYQLEGTELFAVHQR SFLLALPGENPPQKIKQMREIRWKERNRIKQQVDQLSAILRLHKK VNEDERIQAIDKLLQKVASWQLNEEIATAWNQALSQLYSKAKEN DLQWNQAIKNAHHQLEPVVGKQISLWRKDLSTGRQGIAGLSLWS IEELEATKKLLTRWSKRSREPGVVKRIERFETFAKQIQHHINQVKE NRLKQLANLIVMTALGYKYDQEQKKWIEVYPACQVVLFENLRS YRFSYERSRRENKKLMEWSHRSIPKLVQMQGELFGLQVADVYA AYSSRYHGRTGAPGIRCHALTEADLRNETNIIHELIEAGFIKEEHRP YLQQGDLVPWSGGELFATLQKPYDNPRILTLHADINAAQNIQKRF WHPSMWFRVNCESVMEGEIVTYVPKNKTVHKKQGKTFRFVKVE GSDVYEWAKWSKNRNKNTFSSITERKPPSSMILFRDPSGTFFKEQ EWVEQKTFWGKVQSMIQAYMKKTIVQRMEE
메틸로박테리움 노둘란스 (장형) (SEQ ID NO:499)	MYEAIVLADDANAQLANAFLGPLTDPNSAGFLEAFNKVDRPAPS WLDQVPASDPIDPAVLAEANAWLDTDAGRAWLVDTGAPPRWRS LAAKQDPIWPREFARKLGELRKEAASGTSAIIKALKRDFGVLPLF QPSLAPRILGSRSSLTPWDRLAFRLAVGHLLSWESWCTRARDEHT ARVQRLEQFSSAHLKGDLATKVSTLREYERARKEQIAQLGLPMG ERDFLITVRMTRGWDDLREKWRRSGDKGQEALHAIIATEQTRKR GRFGDPDLFRWLARPENHHVWADGHADAVGVLARVNAMERLV ERSRDTALMTLPDPVAHPRSAQWEAEGGSNLRNYQLEAVGGEL QITLPLLKAADDGRCIDTPLSFSLAPSDQLQGVVLTKQDKQQKIT YCTNMNEVFEAKLGSADLLLNWDHLRGRIRDRVDAGDIGSAFLK LALDVAHVLPDGVDDQLARAAFHFQSAKGAKSKHADSVQAGLR VLSIDLGVRSFATCSVFELKDTAPTTGVAFPLAEFRLWAVHERSF TLELPGENVGAAGQQWRAQADAELRQLRGGLNRHRQLLRAATV

[1834]

	QKGERDAYLTDLREAWSAKELWPFEASLLSELERCSTVADPLWQ DTCKRAARLYRTEFGAVVSEWRSRTRSREDRKYAGKSMWSVQH LTDVRRFLQSWSLAGRASGDIRRLDRERGGVFAKDLLDHIDALK DDRLKTGADLIVQAARGFQRNEFGYWVQKHAPCHVILFEDLSRY RMRTDRPRRENSQLMQWAHRGVPDMVGMQGEIYGIQDRRDPDS ARKHARQPLAAFCLDTPAAFSSRYHASTMTPGIRCHPLRKREFED QGFLELLKRENEGLDLNGYKPGDLVPLPGGEVFVCLNANGLSRIH ADINAAQNLQRRFWTQHGDAFRLPCGKSAVQGQIRWAPLSMGK RQAGALGGFGYLEPTGHDSGSCQWRKTTEAEWRRLSGAQKDRD EAAAAEDEELQGLEEELLERSGERVVFFRDPSGVVLPTDLWFPSA AFWSIVRAKTVGRLRSHLDAQAEASYAVAAGL
오피투타세아에 박테리움 (SEQ ID NO:500)	MSLNRIYQGRVAAVETGTALAKGNVEWMPAAGGDEVLWQHHE LFQAAINYYLVALLALADKNNPVLGPLISQMDNPQSPYHVWGSF RRQGRQRTGLSQAVAPYITPGNNAPTLDEVFRSILAGNPTDRATL DAALMQLLKACDGAGAIQQEGRSYWPKFCDPDSTANFAGDPAM LRREQHRLLLPQVLHDPAITHDSPALGSFDTYSIATPDTRTPQLTG PKARARLEQAITLWRVRLPESAADFDRLASSLKKIPDDDSRLNLQ GYVGSSAKGEVQARLFALLLFRHLERSSFTLGLLRSATPPPKNAE TPPPAGVPLPAASAADPVRIARGKRSFVFRAFTSLPCWHGGDNIH PTWKSFDIAAFKYALTVINQIEEKTKERQKECAELETDFDYMHGR LAKIPVKYTTGEAEPPPILANDLRIPLLRELLQNIKVDTALTDGEA VSYGLQRRTIRGFRELRRIWRGHAPAGTVFSSELKEKLAGELRQF QTDNSTTIGSVQLFNELIQNPKYWPIWQAPDVETARQWADAGFA DDPLAALVQEAELQEDIDALKAPVKLTPADPEYSRRQYDFNAVS KFGAGSRSANRHEPGQTERGHNTFTTEIAARNAADGNRWRATH VRIHYSAPRLLRDGLRRPDTDGNEALEAVPWLQPMMEALAPLPT LPQDLTGMPVFLMPDVTLSGERRILLNLPVTLEPAALVEQLGNAG RWQNQFFGSREDPFALRWPADGAVKTAKGKTHIPWHQDRDHFT VLGVDLGTRDAGALALLNVTAQKPAKPVHRIIGEADGRTWYASL ADARMIRLPGEDARLFVRGKLVQEPYGERGRNASLLEWEDARNI ILRLGQNPDELLGADPRRHSYPEINDKLLVALRRAQARLARLQNR SWRLRDLAESDKALDEIHAERAGEKPSPLPPLARDDAIKSTDEAL LSQRDIIRRSFVQIANLILPLRGRRWEWRPHVEVPDCHILAQSDPG TDDTKRLVAGQRGISHERIEQIEELRRRCQSLNRALRHKPGERPVL GRPAKGEEIADPCPALLEKINRLRDQRVDQTAHAILAAALGVRLR APSKDRAERRHRDIHGEYERFRAPADFVVIENLSRYLSSQDRARS ENTRLMQWCHRQIVQKLRQLCETYGIPVLAVPAAYSSRFSSRDGS AGFRAVHLTPDHRHRMPWSRILARLKAHEEDGKRLEKTVLDEAR AVRGLFDRLDRFNAGHVPGKPWRTLLAPLPGGPVFVPLGDATPM QADLNAAINIALRGIAAPDRHDIHHRLRAENKKRILSLRLGTQRE KARWPGGAPAVTLSTPNNGASPEDSDALPERVSNLFVDIAGVAN FERVTIEGVSQKFATGRGLWASVKQRAWNRVARLNETVTDNNR NEEEDDIPM

[1835]

표 12. Tracr 및 직접 반복부

피시스파에라에 박테리움 ST-NAGAB-D1	
tracer 영역	CAACATGCTCGCTTTGCGAAGGCTGACGGCCCGCTCTCATTTGGCATTGCCGGGAGCCGGAGTTTTCGGAAGAGAGTGTCGACGACTGCTGATCTCCGCATCCGCGTCCTGTTCGCCAGGCCGGGTCGGGTGTACGGATCATGCTGGCAGCAGTCTACGCCG (SEQ ID NO:501)
추정 성숙 tracr 1	CGGAAGAGAGUGUCGACGACUGCUGAUCUCCGCAUCCGCGUCCUGUUCGCCAGGCCGGGU (SEQ ID NO:502)
추정 성숙 tracr 2	CCAACAUGCUCGCUUUGCGAAGGCUGACGGCCCGCUCUCAUUUGGCAUUGCCGGGAGCCGGA (SEQ ID NO:503)
추정 성숙 tracr 3	UCGCUUUGCGAAGGCUGACGGCCCGCUCUCAUUUGGCAUUGCCGGGAGCCGGAGUUUUCGGAAGAGAGUGUCGACGACUGCUGAUCUCCGCAUCCGCGUCCUGUUCGCCAGGCCGGGU (SEQ ID NO:504)
추정 성숙 tracr 4	AUGCUCGCUUUGCGAAGGCUGACGGCCCGCUCUCAUUUGGCAUUGCCGGGAGCCGGAGUUUUCGGAAGAGAGUGUCGACGACUGCUG (SEQ ID NO:505)
추정 성숙 tracr 5	AGUUUUCGGAAGAGAGUGUCGACGACUGCUGAUCUCCGCAUCCGCGUCCUGUUCGCCAGGCCGGGUCGGGU (SEQ ID NO:506)
추정 성숙 tracr 6	CGCCUAUCAGCCAACAUGCUCGCUUUGCGAAGGCUGACGGCCCGCUCUCAUUUGGCAUUGC (SEQ ID NO:507)
추정 성숙 tracr 7	CGCCUAUCAGCCAACAUGCUCGCUUUGCGAAGGCUGACGGCCCGCUCUCAUUUGGCAUUGCCGGGAGCCGGAGUUUUCGGAAGAGAG (SEQ ID NO:508)
추정 성숙 tracr 8	UGAUCUCCGCAUCCGCGUCCUGUUCGCCAGGCCGGGUCGGGUGUACGGAUCAUGCUGGCAGCAGUCUACGCCGAGAACAUUCGC (SEQ ID NO:509)
추정 성숙 tracr 9	UGAUCUCCGCAUCCGCGUCCUGUUCGCCAGGCCGGGUCGGGUGUACGGAUCAUGCUGGCAGCAGUCUACGCCGAGAACAUUCGC (SEQ ID NO:510)
추정 성숙 tracr 10	UCCGCAUCCGCGUCCUGUUCGCCAGGCCGGGUCGGGUGUACGGAUCAUGCUGGCAGCAGUCUACGCCGAGAACAUUCGCUUUU (SEQ ID NO:511)
직접 반복부	GGCGCAACCCGCACACAACCGCGAATGGACAC (SEQ ID NO:512)
성숙 직접 반복부	CCGCGAATGGACAC (SEQ ID NO:513)

[1836]

[1837] 표 13은 추가의 예시적인 Cas12b 오솔로그를 표시한다.

표 13 - C2c1 오솔로그	
숙주	GenBank 식별자 (GI) 번호
알리시클로바실러스 아시도테레스트리스	544884152
알리시클로바실러스 콘타미난스	652589596
데술포비브리오 이노피나투스	652932497
데술포나트로뉴 티오디스뮤탄스	667765471
오피투타세아에 박테리움 TAV5	497199019
투베리바실러스 칼리두스	654153037
바실러스 씨모아밀로보란스	754485389
브레비바실러스 sp. CF112	495056180
바실러스 sp. NSP2-1	651512544
데술파틸하브듐 부티라티보란스	654874074
알리시클로바실러스 허바리우스	652569729
알리시클로바실러스 콘타미난스	652589403
시트로박터 프룬디이 ATCC 8090	411770298
시트로박터 프룬디이	696372964
브레비바실러스 아그리	492410745
브레비바실러스 아그리	492410748

[1838]

브레비바실러스 sp. CF112	495062547
메틸로박테리움 노둘란스	506407588
메틸로박테리움 노둘란스 ORS 2060	219945206
메틸로박테리움 노둘란스	760065057

[1839]

[1840] **실시예 2 - 아데노신 디아미나제의 선택 및 디자인:**

[1841] 다수의 AD를 사용하고, 각각은 다양한 수준의 활성을 가질 것이다. 이들 AD는 하기를 포함한다:

[1842] 1. 인간 ADAR(hADAR1, hADAR2, hADAR3)

[1843] 2. 오징어 롤리고 페아레이 (Loligo pealeii) ADAR(sqADAR2a, sqADAR2b)

[1844] 3. ADAT (인간 ADAT, 초파리 ADAT)

[1845] DNA-RNA 헤테로듀플렉스에 대한 ADAR 반응 활성을 증가시키기 위해 돌연변이를 또한 사용할 수 있다. 예를 들어, 인간 ADAR 유전자에 대해, hADAR1d(E1008Q) 또는 hADAR2d(E488Q) 돌연변이를 사용하여 DNA-RNA 표적에 대한 그들의 활성을 증가시킨다.

[1846] 각각의 ADAR은 다양한 수준의 서열 내용 요건을 가진다. 예를 들어, hADAR1d(E1008Q)에 대해, tAg 및 aAg 부위는 효율적으로 탈아미노화되는 반면, aAt 및 cAc는 덜 효율적으로 편집되고, gAa 및 gAc는 훨씬 더 적게 편

집된다. 그러나, ADAR이 다르면, 내용 요건은 다를 것이다.

[1847] 시스템의 한 형태를 나타내는 개략도를 도 1에 제공한다. 예시적인 Cpf1-AD 융합 단백질의 아미노산 서열을 도 2 내지 도 4에 제공한다.

[1848] **실시예 3 - 피시스파에라에 박테리움 ST-NAGAB-D1로부터의 C2c1 (Cas12b)의 특징규명**

[1849] 이. 콜라이 (stbl3)을 피시스파에라에 박테리움 CRISPR-C2c1 유전자좌의 내생성 게놈 서열의 일부를 함유하는 저카피수 플라스미드 (pACYC184)로 형질전환시켰다. 전체 RNA는 14시간 동안 배양된 세포로부터 추출하였고 RNA를 준비하고 소형 RNA 시퀀싱으로 분석하였다. 방법은 [Zetsche et al. 2015]가 기술된 대로 하였다.

[1850] 소형 RNAseq는 tracer RNA의 위치 및 성숙한 crRNA의 아키텍처를 밝혀주었다. 성숙한 crRNA는 아마도 14 nt의 직접 반복부와 이어서 20-24 nt의 가이드 서열인 듯 하다. 높은 판독수를 갖는 잠재적 tracr 서열은 도 2에 도시되어 있고 서열은 표 12에 표시되어 있다. RNA 폴딩 예측을 기반으로 직접 반복부 (DR)를 갖는 tracrRNA 듀플렉스의 구조는 도 2에 도시된다.

[1851] PAM 스크리닝은 [Zetsche et al., 2015] 대로 수행하였다. 특히, Stbl3 이 콜라이는 인식가능한 프로토스페이서의 5'에 위치된 상이한 PAM 서열을 코딩하는 10 ng 플라스미드 DNA로 형질전환시켰고, 콜로니 계측을 수행하였다. 감소된 콜로니 형성이 TTH PAM (H = A, T, C)의 경우에 확인되었다.

[1852] **실시예 4 - 비색 검출**

[1853] DNA 사중체가 생물분자 피분석물 검출에 사용될 수 있다 (도 6). 한 경우에, OTA-압타머 (파란색)는 OTA를 인식하여, 입체형태 변화를 초래해 사중체 (빨간색)를 노출시켜 헤민에 결합되게 한다. 헤민-사중체 복합체는 퍼옥시다제 활성을 가져서, TMB 기질을 유색 형태 (일반적으로 용액 중에서 파란색)로 산화시킬 수 있다. 따라서, 사중체는 본 명세서에 기술된 CRISPR 부차적 활성에 의해 분해될 수 있다. 출원인은 또한 본 명세서에 기술된 CRISPR 부차적 활성의 일부로서 분해할 수 있는 이들 사중체의 RNA 형태를 생성시켰다. 분해는 RNA 압타머의 손실을 초래하여서, 핵산 표적 존재 하에서 색상 신호의 손실을 초래한다. 예시적인 2개 디자인을 하기에 예시한다.

[1854] 1) rUrGrGrGrUrUrGrGrGrUrUrGrGrGrUrUrGrGrGrA (SEQ ID NO:514)

[1855] 2) rUrGrGrGrUrUrUrGrGrGrUrUrUrGrGrGrUrUrUrGrGrGrA (SEQ ID NO:515)

[1856] 구아닌은 사중체 구조를 생성하는 핵심 염기쌍을 형성하고 이것이 헤민 분자에 결합한다. 출원인은 디-뉴클레오티드 데이터가 구아닌이 불충분하게 분해된다는 것을 보여주므로 사중체의 분해를 허용하도록 우리딘 (굵은 체로 표시)과 구아닌의 세트를 배치하였다.

[1857] 비색 어세이는 본 명세서에 기술된 바와 같은 진단 어세이에서 사용을 위해 적용가능하다. 일 구현예에서, 적절한 사중체를 시험 샘플 및 Cas12 시스템과 인큐베이션시킨다. 다른 구현예에서, 적절한 사중체를 시험 샘플 및 Cas13 시스템과 인큐베이션시킨다. 예를 들어, 표적 서열의 Cas13 식별을 허용하고 부차적 활성에 의한 압타머의 분해를 위한 인큐베이션 기간 이후에, 기질을 첨가할 수 있다. 그러고 나서 흡광도를 측정할 수 있다. 다른 구현예에서, 기질을 사중체 및 CRISPR Cas9, Cas12, 또는 Cas13 시스템과 함께 어세이에 포함시킨다.

[1858] **실시예 5**

[1859] 도 13 은 상이한 sgRNA를 도시한다. 도 14 는 상이한 표적 부위에 대한, 플라스미드 형질감염 이후, 도 13의 상이한 sgRNA로 수득된 indel 백분율을 도시한다. 사용된 Cas12b는 바실러스 히사시이 균주 C4로부터의 것이었다.

[1860] **실시예 6**

[1861] 표 14는 Cas12b의 예시적인 오솔로그를 표시한다.

표 14

명칭	수탁 번호	유기체	서열
Aac (알리시클로바실러스 아시도테레스트리스) (SEQ ID NO:516)	WP_021296342	*알리시클로바실러스 아시도테레스트리스*	MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPD

[1862]

			DGKLGSEGLLSGLRVMSVDLGLRTSASISV FRVARKDELKPNSKGRVPFFFPIKGNDNLV AVHERSQLLKLPGETESKDLRAIREERQRT LRQLRTQLAYLRLLVRCGSEDVGRRERSW AKLIEQPVDAANHMTPDWREAFENELQKL KSLHGICSDKEWMDAVYESVRRVWRHMG KQVRDWRKDVRSGERPKIRGYAKDVVGG NSIEQIEYLERQYKFLKSWSFFGKVSGQVIR AEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDERGKGKWVAKYPPCQLI LLEELSEYQFNNDRPPSENNQLMQWSHRG VFQELINQAQVHDLLVGTMYAAFSSRFDA RTGAPGIRCRRVPARCTQEHNPEPFPWWL NKFVVEHTLDACPLRADDLIPTGEGEIFVSP FSAEEGDFHQIHADLNAAQNLQQRLWSDF DISQIRLRCDWGEVDGELVLIPRLTGKRTA DSYSNKVFYTNTGVTYYERERGKKRRKVF AQEKLSEEEAELLVEADEAREKSVVLMRD PSGIINRGNWTRQKEFWSMVNQRIEGYLV KQIRSRVPLQDSACENTGDI
Ak (알리시클로바실러스 카케가웬시스) (SEQ ID NO:517)	WP_067936067	*알리시클로바실러스 카케가웬시스*	MAVKSIKVKLRLSECPDILAGMWQLHRAT NAGVRYYTEWVSLMRQEILYSRGPDGGQQ CYMTAEDCQRELLRRLRNRQLHNGRQDQP GTDADLLAISRRLYEILVLQSIGKRGDAQQI ASSFLSPLVDPNSKGGRGEAKSGRKPAWQ KMRDQGDPRWVAAREKYEQRKAVDPSKE ILNSLDALGLRPLFAVFTETYRSGVDWKPL GKSQGVRTWDRDMFQQALERLMSWESW NRRVGEEYARLFQQKMKFEQEHFAEQSHL VKLARALEADMRAASQGFEAKRGTAHQIT RRALRGADRVFEIWKSIPEEALFSQYDEVIR

[1863]

			QVQAEKRRDFGSHDLFAKLAEPKYQPLWR ADETFLTRYALYNGVLRDLEKARQFATFT LPDACVNPIWTRFESSQGSNLHKYEFLFDH LGPGRHAVRFQRLLVVESEGAKERDSVVV PVAPSGQLDKLVLREEEKSSVALHLHDTAR PDGFMAEWAGAKLQYERSTLARKARRDK QGMRSWRRQPSMLMSAAQMLEDAKQAG DVYLNISVRVKSPSEVRGQRRPPYAALFRI DDKQRRVTVNYNKLSAYLEEHPDKQIPGA PGLLSGLRVMSVDLGLRTSASISVFRVAKK EEVEALGDGRPPHYYPIHGTDDLVAVHERS HLIQMPGETETKQLRKLREERQAVLRPLFA QLALLRLLVRCGAADERIRTRSWQRLTKQ GREFTKRLTPSWREALELELTRLEAYCGRV PDDEWSRIVDRTVIALWRRMGKQVRDWR KQVKSGAKVKVKGYQLDVVGGNSLAQID YLEQQYKFLRRWSFFARASGLVVRADRES HFAVALRQHIENAKRDRLKKLADRILMEA LGYVYEASGPREGQWTAQHPPCQLIILEEL SAYRFSDDRPPSENSKLMAWGHRGILEELV NQAQVHDVLVGTVYAAFSSRFDARTGAPG VRCRRVPARFVGATVDDSLPLWLTEFLDK HRLDKNLLRPDDVIPTGEGEFLVSPCGEEA ARVRQVHADINAAQNLQRRLWQNFDITEL RLRCDVKMGGEGTVLVPRVNNARAKQLF GKKVLVSQDGVTFFERSQTGGKPHSEKQT DLTDKELELIAEADEARAKSVVLFRDPSGH IGKGHWIRQREFWSLVKQRIESHTAERIRV RGVGSSLD
Am (알리시클	WP_0749 48407	*알리시클로바*	MNVAVKSIKVKLMLGHLPEIREGLWHLHE AVNLGVRYYTEWLALLRQGNLYRRGKDG

[1864]

로바실러스 마크로스포란지이더스) (SEQ ID NO:518)		*실러스 마크로스포란 지이더스*	AQECYMTAEQCRQELLVRLRDRQKRNGH TGDPGTDEELLGVARRLYELLVPQSVGKK GQAQMLASGFLSPLADPKSEGGKGTSKSG RKPAWMGMKEAGDSRWVEAKARYEANK AKDPTKQVIASLEMYGLRPLFDVFTETYKT IRWMPLGKHQGVRAWDRDMFQQSLERLM SWESWNERVGAEFARLVDRRDRFREKHFT GQEHLVALAQRLEQEMKEASPGFESKSSQ AHRITKRALRGADGIIDDWLKLSEGEPVDR FDEILRKRQAQNPRRFGSHDLFLKLAEPVF QPLWREDPSFLSRWASYNEVLNKLEDAKQ FATFTLPSPCSNPVWARFENAEGTNIFKYD FLFDHFGKGRHGVRFQRMIVMRDGVPTEV EGIVVPIAPSRQLDALAPNDAASPIDVFVGD PAAPGAFRGQFGGAKIQYRRSALVRKGRR EEKAYLCGFRLPSQRRTGTPADDAGEVFLN LSLRVESQSEQAGRRNPPYAAVFHISDQTR RVIVRYGEIERYLAEHPDTGIPGSRGLTSGL RVMSVDLGLRTSAAISVFRVAHRDELTPDA HGRQPFFFPIHGMDHLVALHERSHLIRLPG ETESKKVRSIREQRLDRLNRLRSQMASLRL LVRTGVLDEQKRDRNWERLQSSMERGGE RMPSDWWDLFQAQVRYLAQHRDASGEA WGRMVQAAVRTLWRQLAKQVRDWRKEV RRNADKVKIRGIARDVPGGHSLAQLDYLE RQYRFLRSWSAFSVQAGQVVRAERDSRFA VALREHIDNGKKDRLKKLADRILMEALGY VYVTDGRRAGQWQAVYPPCQLVLLEELSE YRFSNDRPPSENSQLMVWSHRGVLEELIHQ AQVHDVLVGTIPAAFSSRFDARTGAPGIRC RRVPSIPLKDAPSIPIWLSHYLKQTERDAAA

[1865]

			LRPGELIPTGDGEFLVTPAGRGASGVRVVH ADINAAHNLQRRLWENFDLSDIRVRCDRR EGKDGTVVLIPRLTNQRVKERYSGVIFTSE DGVSFTVGDAKTRRRSSASQGEGDDLSDE EQELLAEADDARERSVVLFRDPSGFVNGG RWTAQRAFWGMVHNRIETLLAERFSVSGA AEKVRG
Bh (바실러스 히사시이) (SEQ ID NO:519)	WP_095142515	*바실러스 히사시이*	MATRSFILKIEPNEEVKKGLWKTHEVLNHG IAYYMNILKLIRQEAIYEHHEQDPKNPKKV SKAEIQAELWDFVLKMQKCNSFTHEVDKD EVFNILRELYEELVPSSVEKKGEANQLSNK FLYPLVDPNSQSGKGTASSGRKPRWYNLKI AGDPSWEEEKKKWEEDKKKDPLAKILGKL AEYGLIPLFIPYTDSNEPIVKEIKWMEKSRN QSVRRLDKDMFIQALERFLSWESWNLKVK EEYEKVEKEYKTLEERIKEDIQALKALEQY EKERQEQLLRDTLNTNEYRLSKRGLRGWR EIIQKWLKMDENEPSEKYLEVFKDYQRKH PREAGDYSVYEFLSKKENHFIWRNHPEYPY LYATFCEIDKKKKDAKQQATFTLADPINHP LWVRFEERSGSNLNKYRILTEQLHTEKLKK KLTVQLDRLIYPTESGGWEEKGKVDIVLLP SRQFYNQIFLDIEEKGKHAFTYKDESIKFPL KGTLGGARVQFDRDHLRRYPHKVESGNV GRIYFNMTVNIEPTESPVSKSLKIHRDDFPK VVNFKPKELTEWIKDSKGKKLKSGIESLEIG LRVMSIDLGQRQAAAASIFEVVDQKPDIEG KLFFPIKGTELYAVHRASFNIKLPGETLVKS REVLRKAREDNLKLMNQKLNFLRNVLHFQ QFEDITEREKRVTKWISRQENSDVPLVYQD ELIQIRELMYKPYKDWVAFLKQLHKRLEV

[1866]

			EIGKEVKHWRKSLSDGRKGLYGISLKNIDE IDRTRKFLLRWSLRPTEPGEVRRLEPGQRF AIDQLNHLNALKEDRLKKMANTIIMHALG YCYDVRKKKWQAKNPACQIILFEDLSNYN PYEERSRFENSKLMKWSRREIPRQVALQGE IYGLQVGEVGAQFSSRFHAKTGSPGIRCSV VTKEKLQDNRFFKNLQREGRLTLDKIAVL KEGDLYPDKGGEKFISLSKDRKCVTTHADI NAAQNLQKRFWTRTHGFYKVYCKAYQVD GQTVYIPESKDQKQKIIEEFGEGYFILKDGV YEWVNAGKLKIKKGSSKQSSSELVDSDILK DSFDLASELKGEKLMLYRDPSGNVFPSDK WMAAGVFFGKLERILISKLTNQYSISTIEDD SSKQSM
Bs (바실러스 sp. NSP2) (SEQ ID NO:520)	WP_0265 57978	*바실러스 sp.* *NSP2*	MAIRSIKLKLKTHTGPEAQNLRKGIWRTHR LLNEGVAYYMKMLLLFRQESTGERPKEEL QEELICHIREQQQRNQADKNTQALPLDKAL EALRQLYELLVPSSVGQSGDAQIISRKFLSP LVDPNSEGGKGTSKAGAKPTWQKKKEAN DPTWEQDYEKWKKRREEDPTASVITTLEE YGIRPIFPLYTNTVTDIAWLPLQSNQFVRT WDRDMLQQAIERLLSWESWNKRVQEEYA KLKEKMAQLNEQLEGGQEWISLLEQYEEN RERELRENMTAANDKYRITKRQMKGWNE LYELWSTFPASASHEQYKEALKRVQQRLR GRFGDAHFFQYLMEEKNRLIWKGNPQRIH YFVARNELTKRLEEAKQSATMTLPNARKH PLWVRFDARGGNLQDYYLTAEADKPRSRR FVTFSQLIWPSESGWMEKKDVEVELALSR QFYQQVKLLKNDKGKQKIEFKDKGSGSTF NGHLGGAKLQLERGDLEKEEKNFEDGEIG

[1867]

			SVYLNVVIDFEPLQEVKNGRVQAPYGQVL QLIRRPNEFPKVTTYKSEQLVEWIKASPQH SAGVESLASGFRVMSIDLGLRAAAATSIFS VEESSDKNAADFSYWIEGTPLVAVHQRSY MLRLPGEQVEKQVMEKRDERFQLHQRVK FQIRVLAQIMRMANKQYGDRWDELDSLK QAVEQKKSPLDQTDRTFWEGIVCDLTKVL PRNEADWEQAVVQIHRKAEEYVGKAVQA WRKRFAADERKGIAGLSMWNIEELEGLRK LLISWSRRTRNPQEVNRFERGHTSHQRLLT HIQNVKEDRLKQLSHAIVMTALGYVYDER KQEWCAEYPACQVILFENLSQYRSNLDRST KENSTLMKWAHRSIPKYVHMQAEPYGIQI GDVRAEYSSRFYAKTGTPGIRCKKVRGQD LQGRRFENLQKRLVNEQFLTEEQVKQLRP GDIVPDDSGELFMTLTDGSGSKEVVFLQAD INAAHNLQKRFWQRYNELFKVSCRVIVRD EEEYLVPKTKSVQAKLGKGLFVKKSDTAW KDVYVWDSQAKLKGKTTFTEESESPEQLE DFQEIIEEAEEAKGTYRTLFRDPSGVFFPES VWYPQKDFWGEVKRKLYGKLRERFLTKA R
Bt (바실러스 써모아밀로보란스) (SEQ ID NO:521)	WP_041902512	*바실러스 써모아밀로보란스*	MATRSFILKIEPNEEVKKGLWKTHEVLNHG IAYYMNILKLIRQEAIYEHHEQDPKNPKKV SKAEIQAELWDFVLKMQKCNSFTHEVDKD VVFNILRELYEELVPSSVEKKGEANQLSNK FLYPLVDPNSQSGKGTASSGRKPRWYNLKI AGDPSWEEEKKKWEEDKKKDPLAKILGKL AEYGLIPLFIPFTDSNEPIVKEIKWMEKSRN QSVRRLDKDMFIQALERFLSWESWNLKVK EEYEKVEKEHKTLEERIKEDIQAFKSLEQY

[1868]

			EKERQEQLLRDTLNTNEYRLSKRGLRGWR EIIQKWLKMDENEPSEKYLEVFKDYQRKH PREAGDYSVYEFLSKKENHFIWRNHPEYPY LYATFCEIDKKKKDAKQQATFTLADPINHP LWVRFEERSGSNLNKYRILTEQLHTEKLKK KLTVQLDRLIYPTESGGWEEKGKVDIVLLP SRQFYNQIFLDIEEKGKHAFTYKDESIKFPL KGTLGGARVQFDRDHLRRYPHKVESGNV GRIYFNMTVNIEPTESPVSKSLKIHRDDFPK FVNFKPKELTEWIKDSKGKKLKSGIESLEIG LRVMSIDLGQRQAAAASIFEVVDQKPDIEG KLFFPIKGTELYAVHRASFNIKLPGETLVKS REVLRKAREDNLKLMNQKLNFLRNVLHFQ QFEDITEREKRVTKWISRQENSDVPLVYQD ELIQIRELMYKPYKDWVAFLKQLHKRLEV EIGKEVKHWRKSLSDGRKGLYGISLKNIDE IDRTRKFLLRWSLRPTEPGEVRRLEPGQRF AIDQLNHLNALKEDRLKKMANTIIMHALG YCYDVRKKKWQAKNPACQIILFEDLSNYN PYEERSRFENSKLMKWSRREIPRQVALQGE IYGLQVGEVGAQFSSRFHAKTGSPGIRCSV VTKEKLQDNRFFKNLQREGRLTLDKIAVL KEGDLYPDKGGEKFISLSKDRKLVTTHADI NAAQNLQKRFWTRTHGFYKVYCKAYQVD GQTVYIPESKDQKQKIIEEFGEGYFILKDGV YEWGNAGKLKIKKGSSKQSSSELVDSDILK DSFDLASELKGEKLMLYRDPSGNVFPSDK WMAAGVFFGKLERILISKLTNQYSISTIEDD SSKQSM
Bv (바실러스	WP_1016 61451	*바실러스 sp.* *V3-13*	MAIRSIKLKMKTNSGTDSIYLRKALWRTHQ LINEGIAYYMNLLTLYRQEAIGDKTKEAYQ

[1869]

[1870]

sp. V3-13) (SEQ ID NO:522)			AELINIIRNQQRNNGSSEEHGSDQEILALLR QLYELIIPSSIGESGDANQLGNKFLYPLVDP NSQSGKGTSNAGRKPRWKRLKEEGNPDW ELEKKKDEERKAKDPTVKIFDNLNKYGLLP LFPLFTNIQKDIEWLPLGKRQSVRKWDKD MFIQAIERLLSWESWNRRVADEYKQLKEK TESYYKEHLTGGEEWIEKIRKFEKERNMEL EKNAFAPNDGYFITSRQIRGWDRVYEKWS KLPESASPEELWKVVAEQQNKMSEGFGDP KVFSFLANRENRDIWRGHSERIYHIAAYNG LQKKLSRTKEQATFTLPDAIEHPLWIRYESP GGTNLNLFKLEEKQKKNYYVTLSKIIWPSE EKWIEKENIEIPLAPSIQFNRQIKLKQHVKG KQEISFSDYSSRISLDGVLGGSRIQFNRKYI KNHKELLGEGDIGPVFFNLVVDVAPLQETR NGRLQSPIGKALKVISSDFSKVIDYKPKELM DWMNTGSASNSFGVASLLEGMRVMSIDM GQRTSASVSIFEVVKELPKDQEQKLFYSIND TELFAIHKRSFLLNLPGEVVTKNNKQQRQE RRKKRQFVRSQIRMLANVLRLETKKTPDE RKKAIHKLMEIVQSYDSWTASQKEVWEKE LNLLTNMAAFNDEIWKESLVELHHRIEPYV GQIVSKWRKGLSEGRKNLAGISMWNIDEL EDTRRLLISWSKRSRTPGEANRIETDEPFGS SLLQHIQNVKDDRLKQMANLIIMTALGFK YDKEEKDRYKRWKETYPACQIILFENLNR YLFNLDRSRRENSRLMKWAHRSIPRTVSM QGEMFGLQVGDVRSEYSSRFHAKTGAPGI RCHALTEEDLKAGSNTLKRLIEDGFINESEL AYLKKGDIIPSQGGELFVTLSKRYKKDSDN NELTVIHADINAAQNLQKRFWQQNSEVYR

			VPCQLARMGEDKLYIPKSQTETIKKYFGKG SFVKNNTEQEVYKWEKSEKMKIKTDTTFD LQDLDGFEDISKTIELAQEQQKKYLTMFRD PSGYFFNNETWRPQKEYWSIVNNIIKSCLK KKILSNKVEL
CLb (칸디다투 스 린도우박테 리아) (SEQ ID NO:523)	OGH5599 4	*칸디다투스 린도우박테리 아*	MPRDDLDLLTNLNSTAKGIRERGKTKEGT DKKKSGRKSSWPMDKAAWETAKTSDSSA HFLEKLKQHPDLKDAFGNLSSGGSKKLEY YKKLAGSAPWKESQSVILEKAARWKEAKQ EREEKEQDSSEHGSKAAYRRLFDAGCLPM PEFAKYIDENQIEFGDLKLSDCGAEWKRG MWNQAGQRVRSHMGWQRRREKENAVYS LRKELFEKGGAIRRKKSEELTPEDILPGKAA PDQNDWQERPAYGNQMWFIGLRSYEENE MAKYAEEAGMGSRSAPRIRRGTIKGWSKL RERWLQILKRNPQATRDDLIGELNALRSQD PRAYGDARLFDWLSKTDQRFLWDGFDAD GKILCGRDDRDCVSAFVAYNEEFADEPSSI TLTETDERLHPVWPFFGESSAVPYEIEYDLE TACPTAIRLPLLVGKENGGYAERQGTRLPL AEYADLASSFQLPTPVRLDVLVEIREVTRA GRKVTCPFSYFKQNGVWYVREGEIPSGESI QIKQTDRKIENGKIFISSKLRMAYRDDLMV SPATGDFGSIKILWERIELASHVDQKKLPET APARSRVFVSFSCNVVERAPRKQLTRKPDA VVVTIPSGVDQGLVVVSTDVRTGKSKSSSA PPLPPGSRLWPADAVHGDPPLRILSVDLGH RHSAYAVWELGLQQKSWRAGVLKGSTQT PVYADCTGTGLLCLPGDGEDTPAEEESLRL RSRQIRRRLNLQNSILRVSRLLSLDKFEKTIF EQSDVRDRPNKKGLRIRRRCRTEKTPLSEA

[1871]

			EVRKNCDKAAEILIRWADTDAMAKSLAAT GNADISFWKYMAVKNPPLSAVVDVAPSTI VPDDGPDRETLKKKRQEEEEKFASSIYENR VKLAGALCSGYDADHRRPATGGLWHDLD RTLIREISYGDRGQKGNPRKLNNEGILRLLR RPPRARPDWREFHRTLNDANRIPKGRTLRG GLSMGRLNFLKEVGDFVKKWSCRPRWPG DRRHIPPGQLFDRQDAEHLEHLRDDRIKRL AHLIVAQALGFEPDIRRGLWKYVDGSTGEI LWQHPETRRFFAEGAAGELREVSRPAEIDD DAAARPHTVSAPAHIVVFENLIRYRFQSDR PKTENAGLMQWAHRQIVHFTKQVASLYGL KVAMVYAAFSSKFCSRCGSPGARVSRFDP AWRNQEWFKRRTSNPRSKVDHSLKRASED PTADETRPWVLIEGGKEFVCANAKCSAHD EPLNADENAAANIGLRFLRGVEDFRTKVNP AGALKGKLRFETGIHSFRPPVSGSPFWSPM AEPAQKKKIGAAAPGADVDEAGDADESGV VVLFRDPSGAFRNKQYWYEGKIFWSNVM MAVEAKIAGASVGAKPVAASWGQAQPQS GPGLAKPGGD
Db (데술파틸 하브듐 부티라티보 란스) (SEQ ID NO:524)	WP_0283 26052	*데술파틸하브듐 부티라티보란스*	MPLSNNPPVTQRAYTLRLRGADPSDLSWR EALWHTHEAVNKGAKVFGDWLLTLRGGL DHTLADTKVKGGKGKPDRDPTPEERKARR ILLALSWLSVESKLGAPSSYIVASGDEPAKD RNDNVVSALEEILQSRKVAKSEIDDWKRD CSASLSAAIRDDAVWVNRSKVFDEAVKSV GSSLTREEAWDMLERFFGSRDAYLTPMKD PEDKSSETEQEDKAKDLVQKAGQWLSSRY GTSEGADFCRMSDIYGKIAAWADNASQGG SSTVDDLVSELRQHFDTKESKATNGLDWII

[1872]

			GLSSYTGHTPNPVHELLRQNTSLNKSHLDD LKKKANTRAESCKSKIGSKGQRPYSDAILN DVESVCGFTYRVDKDGQPVSVADYSKYD VDYKWGTARHYIFAVMLDHAARRISLAHK WIKRAEAERHKFEEDAKRIANVPARAREW LDSFCKERSVTSGAVEPYRIRRRAVDGWK EVVAAWSKSDCKSTEDRIAAARALQDDSEI DKFGDIQLFEALAEDDALCVWHKDGEATN EPDFQPLIDYSLAIEAEFKKRQFKVPAYRHP DELLHPVFCDFGKSRWKINYDVHKNVQAP FYRGLCLTLWTGSEIKPVPLCWQSKRLTRD LALGNNHRNDAASAVTRADRLGRAASNV TKSDMVNITGLFEQADWNGRLQAPRQQLE AIAVVRDNPRLSEQERNLRMCGMIEHIRWL VTFSVKLQPQGPWCAYAEQHGLNTNPQY WPHADTNRDRKVHARLILPRLPGLRVLSV DLGHRYAAACAVWEAVNTETVKEACQNV GRDMPKEHDLYLHIKVKKQGIGKQTEVDK TTIYRRIGADTLPDGRPHPAPWARLDRQFLI KLQGEEKDAREASNEEIWALHQMECKLDR TKPLIDRLIASGWGLLKRQMARLDALKEL GWIPAPDSSENLSREDGEAKDYRESLAVDD LMFSAVRTLRLALQRHGNRARIAYYLISEV KIRPGGIQEKLDENGRIDLLQDALALWHEL FSSPGWRDEAAKQLWDSRIATLAGYKAPE ENGDNVSDVAYRKKQQVYREQLRNVAKT LSGDVITCKELSDAWKERWEDEDQRWKK LLRWFKDWVLPSGTQANNATIRNVGGLSL SRLATITEFRRKVQVGFFTRLRPDGTRHEIG EQFGQKTLDALELLREQRVKQLASRIAEAA LGIGSEGGKGWDGGKRPRQRINDSRFAPC

[1873]

			HAVVIENLANYRPDETRTRLENRRLMTWS ASKVHKYLSEACQLNGLYLCTVSAWYTSR QDSRTGAPGIRCQDVSVREFMQSPFWRKQ VKQAEAKHDENKGDARERFLCELNKTWK AKTPAEWKKAGFVRIPLRGGEIFVSADSKS PSAKGIHADLNAAANIGLRALTDPDWPGK WWYVPCDPVSFESKMDYVKGCAAVKVG QPLRQPAQTNADGAASKIRKGKKNRTAGT SKEKVYLWRDISAFPLESNEIGEWKETSAY QNDVQYRVIRMLKEHIKSLDNRTGDNVEG
Dt (데술포나트로늅 티오디스뮤탄스) (SEQ ID NO:525)	WP_031386437	*데술포나트로늅 티오디스뮤탄스*	MVLGRKDDTAELRRALWTTHEHVNLAVA EVERVLLRCRGRSYWTLDRRGDPVHVPES QVAEDALAMAREAQRRNGWPVVGEDEEI LLALRYLYEQIVPSCLLDDLGKPLKGDAQK IGTNYAGPLFDSDTCRRDEGKDVACCGPFH EVAGKYLGALPEWATPISKQEFDGKDASH LRFKATGGDDAFFRVSIEKANAWYEDPAN QDALKNKAYNKDDWKKEKDKGISSWAVK YIQKQLQLGQDPRTEVRRKLWLELGLLPLF IPVFDKTMVGNLWNRLAVRLALAHLLSWE SWNHRAVQDQALARAKRDELAALFLGME DGFAGLREYELRRNESIKQHAFEPVDRPYV VSGRALRSWTRVREEWLRHGDTQESRKNI CNRLQDRLRGKFGDPDVFHWLAEDGQEA LWKERDCVTSFSLLNDADGLLEKRKGYAL MTFADARLHPRWAMYEAPGGSNLRTYQIR KTENGLWADVVLLSPRNESAAVEEKTFNV RLAPSGQLSNVSFDQIQKGSKMVGRCRYQ SANQQFEGLLGGAEILFDRKRIANEQHGAT DLASKPGHVWFKLTLDVRPQAPQGWLDG KGRPALPPEAKHFKTALSNKSKFADQVRP

[1874]

			GLRVLSVDLGVRSFAACSVFELVRGGPDQ GTYFPAADGRTVDDPEKLWAKHERSFKIT LPGENPSRKEEIARRAAMEELRSLNGDIRR LKAILRLSVLQEDDPRTEHLRLFMEAIVDD PAKSALNAELFKGFGDDRFRSTPDLWKQH CHFFHDKAEKVVAERFSRWRTETRPKSSS WQDWRERRGYAGGKSYWAVTYLEAVRG LILRWNMRGRTYGEVNRQDKKQFGTVAS ALLHHINQLKEDRIKTGADMIIQAARGFVP RKNGAGWVQVHEPCRLILFEDLARYRFRT DRSRRENSRLMRWSHREIVNEVGMQGELY GLHVDTTEAGFSSRYLASSGAPGVRCRHL VEEDFHDGLPGMHLVGELDWLLPKDKDR TANEARRLLGGMVRPGMLVPWDGGELFA TLNAASQLHVIHADINAAQNLQRRFWGRC GEAIRIVCNQLSVDGSTRYEMAKAPKARLL GALQQLKNGDAPFHLTSIPNSQKPENSYVM TPTNAGKKYRAGPGEKSSGEEDELALDIVE QAEELAQGRKTFFRDPSGVFFAPDRWLPSE IYWSRIRRRIWQVTLERNSSGRQERAEMDE MPY
Eb (엘루시미 크로비아 박테리움) (SEQ ID NO:526)	OGS0232 6	*엘루시미크로 비아 박테리움*	MNRIYQGRVTKVEVPDGKDEKGNIKWKK LENWSDILWQHHMLFQDAVNYYTLALAAI SGSAVGSDEKSIILREWAVQVQNIWEKAK KKATVFEGPQKRLTSILGLEQNASFDIAAK HILRTSEAKPEQRASALIRLLEEIDKKNHNV VCGERLPFFCPRNIQSKRSPTSKAVSSVQEQ KRQEEVRRFHNMQPEEVVKNAVTLDISLF KSSPKIVFLEDPKKARAELLKQFDNACKKH KELVGIKKAFTESIDKHGSSLKVPAPGSKPS GLYPSAIVFKYFPVDITKTVFLKATEKLAM

[1875]

			GKDREVTNDPIADARVNDKPHFDYFTNIAL IREKEKNRAAWFEFDLAAFIEAIMSPHRFY QDTQKRKEAARKLEEKIKAIEGKGGQFKES DSEDDDVDSLPGFEGDTRIDLLRKLVTDTL GWLGESETPDNNEGKKTEYSISERTLRIFPD IQKQWSELAEKGETTEGKLLEVLKHEQTE HQSDFGSATLYQHLAKPEFHPIWLKSGTEE WHAENPLKAWLNYKELQYELTDKKRPIHF TPAHPVYSPRYFDFPKKSETEEKEVSKNTH SLTTSLASEHIKNSLQFTAGLIRKTNVGKKA IKARFSYSAPRLRRDCLRSENNENLYKAPW LQPMMRALGIDEEKADRQNFANTRITLMA KGLDDIQLGFPVEANSQELQKEVSNGISWK GQFNWGGIASLSALRWPHEKKPKNPPEQP WWGIDSFSCLAVDLGQRYAGAFARLDVST IEKKGKSRFIGEACDKKWYAKVSRMGLLR LPGEDVKVWRDASKIDKENGFAFRKELFG EKGRSATPLEAEETAELIKLFGANEKDVMP DNWSKELSFPEQNDKLLIVARRAQAAVSR LHRWAWFFDEAKRSDDAIREILESDDTDLK QKVNKNEIEKVKETIISLLKVKQELLPTLLT RLANRVLPLRGRSWEWKKHHQKNDGFILD QTGKAMPNVLIRGQRGLSMDRIEQITELRK RFQALNQSLRRQIGKKAPAKRDDSIPDCCP DLLEKLDHMKEQRVNQTAHMILAEALGLK LAEPPKDKKELNETCDMHGAYAKVDNPVS FIVIEDLSRYRSSQGRSPRENSRLMKWCHR AVRDKLKEMCEVFFPLCERRKAGSAWVSL PPLLETPAAYSSRFCSRSGVAGFRAVEVIPG FELKYPWSWLKDKKDKAGNLAKEALNIRT VSEQLKAFNQDKPEKPRTLLVPIAGGPIFVP

[1876]

			ISEVGLSSFGLKPQVVQADINAAINLGLRAI SDPRIWEIHPRLRTEKRDGRLFAREKRKYG EEKVEVQPSKNEKAKKVKDDRKPNYFADF SGKVDWGFGNIKNESGLTLVSGKALWWTI NQLQWERCFDINKRHIEDWSNKQKQ
Lb (렌티스파 에리아 박테리움) (SEQ ID NO:527)		*렌티스파에리아 박테리움*	MAVELNRIYQGRVNHVYIFDENQNQVSVD NGDDLLFVHHELYQDAINYYLVALAAMA LDSKDSLFGKFKMQIRAVWNDFYRNGQLR PGLKHSLIRSLGHAAELNTSNGADIAMNLI LEDGGIPSEILNAALEHLAEKCTGDVSQLG KTFFPRFCDTAYHGNWDVDAKSFSEKKGR QRLVDALYSLHPVQAVQELAPEIEIGWGG VKTQTGKFFTGDEAKASLKKAISYFLQDTG KNSPELQEYFSVAGKQPLEQYLGKIDTFPEI SFGRISSHQNINISNAMWILKFFPDQYSVDL IKNLIPNKKYEIGIAPQWGDDPVKLSRGKR GYTFRAFTDLAMWEKNWKVFDRAAFSDA LKTINQFRNKTQERNDQLKRYCAALNWM DGESSDKKPPVEPADADAVDEAATSVLPIL AGDKRWNALLQLQKELGICNDFTENELMD YGLSLRTIRGYQKLRSMMLEKEEKMRAKT ADDEEISQALQEIIIKFQSSHRDTIGSVSLFL KLAEPKYFCVWHDADKNQNFASVDMVAD AVRYYSYQEEKARLEEPIQITPADARYSRR VSDLYALVYKNAKECKTGYGLRPDGNFVF EIAQKNAKGYAPAKVVLAFSAPRLKRDGLI DKEFSAYYPPVLQAFLREEEAPKQSFKTTA VILMPDWDKNGKRRILLNFPIKLDVSAIHQ KTDHRFENQFYFANNTNTCLLWPSYQYKK PVTWYQGKKPFDVVAVDLGQRSAGAVSRI TVSTEKREHSVAIGEAGGTQWYAYRKFSG

[1877]

			LLRLPGEDATVIRDGQRTEELSGNAGRLST EEETVQACVLCKMLIGDATLLGGSDEKTIR SFPKQNDKLLIAFRRATGRMKQLQRWLW MLNENGLCDKAKTEISNSDWLVNKNIDNV LKEEKQHREMLPAILLQIADRVLPLRGRKW DWVLNPQSNSFVLQQTAHGSGDPHKKICG QRGLSFARIEQLESLRMRCQALNRILMRKT GEKPATLAEMRNNPIPDCCPDILMRLDAM KEQRINQTANLILAQALGLRHCLHSESATK RKENGMHGEYEKIPGVEPAAFVVLEDLSR YRFSQDRSSYENSRLMKWSHRKILEKLALL CEVFNVPILQVGAAYSSKFSANAIPGFRAE ECSIDQLSFYPWRELKDSREKALVEQIRKIG HRLLTFDAKATIIMPRNGGPVFIPFVPSDSK DTLIQADINASFNIGLRGVADATNLLCNNR VSCDRKKDCWQVKRSSNFSKMVYPEKLSL SFDPIKKQEGAGGNFFVLGCSERILTGTSEK SPVFTSSEMAKKYPNLMFGSALWRNEILKL ERCCKINQSRLDKFIAKKEVQNEL
Ls (라세이엘라 세디미니스) (SEQ ID NO:528)	WP_1063 41859	*라세이엘라 세디미니스*	MSIRSFKLKIKTKSGVNAEELRRGLWRTHQ LINDGIAYYMNWLVLLRQEDLFIRNEETNE IEKRSKEEIQGELLERVHKQQQRNQWSGEV DDQTLLQTLRHLYEEIVPSVIGKSGNASLK ARFFLGPLVDPNNKTTKDVSKSGPTPKWK KMKDAGDPNWVQEYEKYMAERQTLVRLE EMGLIPLFPMYTDEVGDIHWLPQASGYTRT WDRDMFQQAIERLLSWESWNRRVRERRA QFEKKTHDFASRFSESDVQWMNKLREYEA QQEKSLEENAFAPNEPYALTKKALRGWER VYHSWMRLDSAASEEAYWQEVATCQTAM RGEFGDPAIYQFLAQKENHDIWRGYPERVI

[1878]

			DFAELNHLQRELRRAKEDATFTLPDSVDHP LWVRYEAPGGTNIHGYDLVQDTKRNLTLI LDKFILPDENGSWHEVKKVPFSLAKSKQFH RQVWLQEEQKQKKREVVFYDYSTNLPHL GTLAGAKLQWDRNFLNKRTQQQIEETGEI GKVFFNISVDVRPAVEVKNGRLQNGLGKA LTVLTHPDGTKIVTGWKAEQLEKWVGESG RVSSLGLDSLSEGLRVMSIDLGQRTSATVS VFEITKEAPDNPYKFFYQLEGTELFAVHQR SFLLALPGENPPQKIKQMREIRWKERNRIK QQVDQLSAILRLHKKVNEDERIQAIDKLLQ KVASWQLNEEIATAWNQALSQLYSKAKEN DLQWNQAIKNAHHQLEPVVGKQISLWRK DLSTGRQGIAGLSLWSIEELEATKKLLTRW SKRSREPGVVKRIERFETFAKQIQHHINQVK ENRLKQLANLIVMTALGYKYDQEQKKWIE VYPACQVVLFENLRSYRFSYERSRRENKKL MEWSHRSIPKLVQMQGELFGLQVADVYA AYSSRYHGRTGAPGIRCHALTEADLRNETN IIHELIEAGFIKEEHRPYLQQGDLVPWSGGE LFATLQKPYDNPRILTLHADINAAQNIQKR FWHPSMWFRVNCESVMEGEIVTYVPKNKT VHKKQGKTFRFVKVEGSDVYEWAKWSKN RNKNTFSSITERKPPSSMILFRDPSGTFFKEQ EWVEQKTFWGKVQSMIQAYMKKTIVQRM EE
Mn (메틸로박 테리움 노둘란스)	WP_0437 47912	*메틸로박테리 움 노둘란스*	MYEAIVLADDANAQLANAFLGPLTDPNSA GFLEAFNKVDRPAPSWLDQVPASDPIDPAV LAEANAWLDTDAGRAWLVDTGAPPRWRS LAAKQDPIWPREFARKLGELRKEAASGTSA IIKALKRDFGVLPLFQPSLAPRILGSRSSLTP

[1879]

(SEQ ID NO:529)			WDRLAFRLAVGHLLSWESWCTRARDEHT ARVQRLEQFSSAHLKGDLATKVSTLREYE RARKEQIAQLGLPMGERDFLITVRMTRGW DDLREKWRRSGDKGQEALHAIIATEQTRK RGRFGDPDLFRWLARPENHHVWADGHAD AVGVLARVNAMERLVERSRDTALMTLPDP VAHPRSAQWEAEGGSNLRNYQLEAVGGE LQITLPLLKAADDGRCIDTPLSFSLAPSDQL QGVVLTKQDKQQKITYCTNMNEVFEAKLG SADLLLNWDHLRGRIRDRVDAGDIGSAFL KLALDVAHVLPDGVDDQLARAAFHFQSA KGAKSKHADSVQAGLRVLSIDLGVRSFAT CSVFELKDTAPTTGVAFPLAEFRLWAVHER SFTLELPGENVGAAGQQWRAQADAELRQL RGGLNRHRQLLRAATVQKGERDAYLTDLR EAWSAKELWPFEASLLSELERCSTVADPL WQDTCKRAARLYRTEFGAVVSEWRSRTRS REDRKYAGKSMWSVQHLTDVRRFLQSWS LAGRASGDIRRLDRERGGVFAKDLLDHIDA LKDDRLKTGADLIVQAARGFQRNEFGYWV QKHAPCHVILFEDLSRYRMRTDRPRRENSQ LMQWAHRGVPDMVGMQGEIYGIQDRRDP DSARKHARQPLAAFCLDTPAAFSSRYHAST MTPGIRCHPLRKREFEDQGFLELLKRENEG LDLNGYKPGDLVPLPGGEVFVCLNANGLS RIHADINAAQNLQRRFWTQHGDAFRLPCG KSAVQGQIRWAPLSMGKRQAGALGGFGY LEPTGHDSGSCQWRKTTEAEWRRLSGAQK DRDEAAAAEDEELQGLEEELLERSGERVVF FRDPSGVVLPTDLWFPSAAFWSIVRAKTVG RLRSHLDAQAEASYAVAAGL

[1880]

Ob (옴니트로피카 박테리움) (SEQ ID NO:530)	WP_009513281	*옴니트로피카 WOR_2*	MNRIYQGRVTKVEKLKNGKSPDDREELKD WQTALWRHHELFQDAVSYYTLALAAMAE GLPDKHPINVLRKRMEEAWEEFPRKTVTP AKNLRDSVRPWLGLSESASFGDALKKILPP APENKEVRALAVALLAEKARTLKPQKTSA SYWGRFCDDLKKKPNWDYSEEELARKTGS GDWVAGLWSEDALNKIDELAKSLKLSSLV KCVPDGQINPEGARNLVKEALDHLEGVSN GTKKEKNDPGPAKKTNNWLRQHASDVRN FIHKNKNQFSSLPNGRLITERARGGGININK TYAGVLFKAFPCPFTFDYVRAAVPEPKVK KVDQEKKSEQSATWTELEKRILRIGDDPIEL ARKNNKPIFKAFTALEKWSDQNSKSCWSD FDKCAFEEALKTLNQFNQKTEEREKRRSEA EAELKYMMDENPEWKPKKETEGDDVREV PILKGDPRYEKLVKLFGDLDEEGSEHATGK IYGPSRASLRGFGKLRNEWVDLFTKANDN PREQDLQKAVTGFQREHKLDMGYTAFFLK LCERDYWDIWRDDTEVEVKKIREKRWVKS VVYAAADTRELAEELERLQEPVRYTPAEP QFSRRLFMFSDIKGKQGAKHIREGLVEVSL AVKDQSGKYGTCRVRLHYSAPRLIRDHLS DGSSSMWLQPMMAALGLSSDARGCFTRD SKGNVKEPAVALMSDFVGRKRELRMLLNF PVDLDISKLEENIGKKARWEKQMNTAYEK NKLKQRFHLIWPGMELKETQEPGQFWWD NPTIQKEGMYCLAIDLSQRRAADYALLHA GVNRDSKTFVELGQAGGQSWFTKLCAAGS LRLPGEDTEVIREGKRQIELSGKKGRNATQ SEYDQAIALAKQLLHNENSAELESAARDW LGDNAKRFSFPEQNDKLIDLYYGALSRYKT

[1881]

			WLRWSWRLTEQHKELWDKTLDEIRKVPY FASWGELAGNGTNEATVQQLQKLIADAAV DLRNFLEKALLHIAYRALPLRENTWRWIEN GKDGKGKPLHLLVSDGQSPAEIPWLRGQR GLSIARIEQLENFRRAVLSLNRLLRHEIGTK PEFGSSTCGESLPDPCPDLTDKIVRLKEERV NQTAHLIIAQSLGVRLKGHSLFTEEREKAD MHGEHEVIPGRSPVDFVVLEDLSRYTTDKS RSRSENSRLMKWCHRKINEKVKLLAEPFGI PVIEVFASYSSKFDARTGAPGFRAVEVTSE DRPFWRKTIEKQSVAREVFDCLDNLVGKG LNGIHLVLPQNGGPLFIAAVKEDQPLPAIRQ ADINAAVNIGLRAIAGPSCYHAHPKVRLIK GESGTDKGKWLPRKGKEANKRENAQFGN VDLDLEVKFNRLDIDSDVLKGDNTNLFHD PLNIACYGFATIQNLQHPFLAHASAVFSRQ KGAVARLQWEVCRAINSRRLEAWQKKAE KAAVKR
Opb (오피투타세아에 박테리움) (SEQ ID NO:531)	WP_009513281	*오피투타세아에 박테리움*	MSLNRIYQGRVAAVETGTALAKGNVEWM PAAGGDEVLWQHHELFQAAINYYLVALLA LADKNNPVLGPLISQMDNPQSPYHVWGSF RRQGRQRTGLSQAVAPYITPGNNAPTLDEV FRSILAGNPTDRATLDAALMQLLKACDGA GAIQQEGRSYWPKFCDPDSTANFAGDPAM LRREQHRLLLPQVLHDPAITHDSPALGSFD TYSIATPDTRTPQLTGPKARARLEQAITLW RVRLPESAADFDRLASSLKKIPDDDSRLNL QGYVGSSAKGEVQARLFALLLFRHLERSSF TLGLLRSATPPPKNAETPPPAGVPLPAASA ADPVRIARGKRSFVFRAFTSLPCWHGGDNI HPTWKSFDIAAFKYALTVINQIEEKTKERQ

[1882]

			KECAELETDFDYMHGRLAKIPVKYTTGEA EPPPILANDLRIPLLRELLQNIKVDTALTDG EAVSYGLQRRTIRGFRELRRIWRGHAPAGT VFSSELKEKLAGELRQFQTDNSTTIGSVQLF NELIQNPKYWPIWQAPDVETARQWADAGF ADDPLAALVQEAELQEDIDALKAPVKLTP ADPEYSRRQYDFNAVSKFGAGSRSANRHE PGQTERGHNTFTTEIAARNAADGNRWRAT HVRIHYSAPRLLRDGLRRPDTDGNEALEAV PWLQPMMEALAPLPTLPQDLTGMPVFLMP DVTLSGERRILLNLPVTLEPAALVEQLGNA GRWQNQFFGSREDPFALRWPADGAVKTA KGKTHIPWHQDRDHFTVLGVDLGTRDAG ALALLNVTAQKPAKPVHRIIGEADGRTWY ASLADARMIRLPGEDARLFVRGKLVQEPY GERGRNASLLEWEDARNIILRLGQNPDELL GADPRRHSYPEINDKLLVALRRAQARLAR LQNRSWRLRDLAESDKALDEIHAERAGEK PSPLPPLARDDAIKSTDEALLSQRDIIRRSFV QIANLILPLRGRRWEWRPHVEVPDCHILAQ SDPGTDDTKRLVAGQRGISHERIEQIEELRR RCQSLNRALRHKPGERPVLGRPAKGEEIAD PCPALLEKINRLRDQRVDQTAHAILAAALG VRLRAPSKDRAERRHRDIHGEYERFRAPA DFVVIENLSRYLSSQDRARSENTRLMQWC HRQIVQKLRQLCETYGIPVLAVPAAYSSRF SSRDGSAGFRAVHLTPDHRHRMPWSRILA RLKAHEEDGKRLEKTVLDEARAVRGLFDR LDRFNAGHVPGKPWRTLLAPLPGGPVFVP LGDATPMQADLNAAINIALRGIAAPDRHDI HHRLRAENKKRILSLRLGTQREKARWPGG

[1883]

			APAVTLSTPNNGASPEDSDALPERVSNLFV DIAGVANFERVTIEGVSQKFATGRGLWAS VKQRAWNRVARLNETVTDNNRNEEEDDIP M
Phyci (피시스파 에라에 박테리움) (SEQ ID NO:532)	AQT6968 5	*피시스파에라 에 박테리움*	MATKSYRARILTDSRLAAALDRTHVVFVE SLKQMINTYLRMQNGKFGPDHKKLAQIML SRSNTFAHGVMDQITRDQPTSTLDEEWTDL ARRIHKTTGPLFLQAERFATVKNRAIHTKS RGKVIPSPETLAVPAKFWHQVCDSASAYIR SNRELMQQWRKDRAAWLKDKNEWQQKH PEFMQFYNGPYQNFLKLCDDDRITSQLAAE QQPTASKNNRPRKTGKRFARWHLWYKWL SENPEIIEWRNKASASDFKTVTDDVRKQIIT KYPQQNKYITRLLDWLEDNNPELKTLENL RRTYVKKFDSFKRPPTLTLPSPYRHPYWFT MELDQFYKKADFENGTIQLLLIDEDDDGN WFFNWMPASLKPDPRLVPSWRAETFETEG RFPPYLGGKIGKKLSRPAPTDAERKAGIAG AKLMIKNNRSELLFTVFEQDCPPRVKWAK TKNRKCPADNAFSSDGKTRKPLRILSIDLGI RHIGAFALTQGTRNDSAWQTESLKKGIINS PSIPPLRQVRRHDYDLKRKRRRHGKPVKG QRSNANLQAHRTNMAQDRFKKGASAIVSL AREHSADLILFENLHSLKFSAFDERWMNRQ LRDMNRRHIVELVSEQAPEFGITVKDDINP WMTSRICSNCNLPGFRFSMKKKNPYREKL PREKCTDFGYPVWEPGGHLFRCPHCDHRV NADINAAANLANKFFGLGYWNNGLKYDA ETKTFTVHTDKKTPPLIFKPRPQFDLWADS VKTRKQLGPDPF
Planc	OHB6217	*플란크토마이*	MSVRSFQARVECDKQTMEHLWRTHKVFN

[1884]

(플란크토마이세테스 박테리움) (SEQ ID NO:533)	5	*세테스 박테리움*	ERLPEIIKILFKMKRGECGQNDKQKSLYKSISQSILEANAQNADYLLNSVSIKGWKPGTAKKYRNASFTWADDAAKLSSQGIHVYDKKQVLGDLPGMMSQMVCRQSVEAISGHIELTKKWEKEHNEWLKEKEKWESEDEHKKYLDLREKFEQFEQSIGGKITKRRGRWHLYLKWLSDNPDFAAWRGNKAVINPLSEKAQIRINKAKPNKKNSVERDEFFKANPEMKALDNLHGYYERNFVRRRKTKKNPDGFDHKPTFTLPHPTIHPRWFVFNKPKTNPEGYRKLILPKKAGDLGSLEMRLLTGEKNKGNYPDDWISVKFKADPRLSLIRPVKGRRVVRKGKEQGQTKETDSYEFFDKHLKKWRPAKLSGVKLIFPDKTPKAAYLYFTCDIPDEPLTETAKKIQWLETGDVTKKGKKRKKKVLPHGLVSCAVDLSMRRGTTGFATLCRYENGKIHILRSRNLWVGYKEGKGCHPYRWTEGPDLGHIAKHKREIRILRSKRGKPVKGEESHIDLQKHIDYMGEDRFKKAARTIVNFALNTENAASKNGFYPRADVLLLENLEGLIPDAEKERGINRALAGWNRRHLVERVIEMAKDAGFKRRVFEIPPYGTSQVCSKCGALGRRYSIIRENNRREIRFGYVEKLFACPNCGYCANADHNASVNLNRRFLIEDSFKSYYDWKRLSEKKQKEEIETIESKLMDKLCAMHKISRGSISK
Sb (스피로카에테스 박테리움) (SEQ ID	OHD16008	*스피로카에테스 박테리움*	MSFTISYPFKLIIKNKDEAKALLDTHQYMNEGVKYYLEKLLMFRQEKIFIGEDETGKRIYIEETEYKKQIEEFYLIKKTELGRNLTLTLDEFKTLMRELYICLVSSSMENKKGFPNAQQASLNIFSPLFDAESKGYILKEENNNISLIHKDYGKILLKRLRDNNLIPIFTKFTDIKKITAKLSPT

[1885]

NO:534)			ALDRMIFAQAIEKLLSYESWCKLMIKERFD KEVKIKELENKCENKQERDKIFEILEKYEEE RQKTFEQDSGFAKKGKFYITGRMLKGFDEI KEKWLKEKDRSEQNLINILNKYQTDNSKL VGDRNLFEFIIKLENQCLWNGDIDYLKIKR DINKNQIWLDRPEMPRFTMPDFKKHPLWY RYEDPSNSNFRNYKIEVVKDENYITIPLITE RNNEYFEENYTFNLAKLKKLSENITFIPKSK NKEFEFIDSNDEEEDKKDQKKSKQYIKYCD TAKNTSYGKSGGIRLYFNRNELENYKDGK KMDSYTVFTLSIRDYKSLFAKEKLQPQIFN TVDNKITSLKIQKKFGNEEQTNFLSYFTQN QITKKDWMDEKTFQNVKELNEGIRVLSVD LGQRFFAAVSCFEIMSEIDNNKLFFNLNDQ NHKIIRINDKNYYAKHIYSKTIKLSGEDDDL YKERKINKNYKLSYQERKNKIGIFTRQINK LNQLLKIIRNDEIDKEKFKELIETTKRYVKN TYNDGIIDWNNVDNKILSYENKEDVINLHK ELDKKLEIDFKEFIRECRKPIFRSGGLSMQRI DFLEKLNKLKRKWVARTQKSAESIVLTPKF GYKLKEHINELKDNRVKQGVNYILMTALG YIKDNEIKNDSKKKQKEDWVKKNRACQIIL MEKLTEYTFAEDRPREENSKLRMWSHRQI FNFLQQKASLWGILVGDVFAPYTSKCLSDN NAPGIRCHQVTKKDLIDNSWFLKIVVKDD AFCDLIEINKENVKNKSIKINDILPLRGGELF ASIKDGKLHIVQADINASRNIAKRFLSQINP FRVVLKKDKDETFHLKNEPNYLKNYYSIL NFVPTNEELTFFKVEENKDIKPTKRIKMDK HEKESTDEGDDYSKNQIALFRDDSGIFFDK SLWVDGKIFWSVVKNKMTKLLRERNNKK

[1886]

			NGSK
Vb (베루코미 크로비아세 아에 박테리움) (SEQ ID NO:535)		*베루코미크로* *비아세아에* *박테리움*	MPLSRIYQGRTNSLIILTPTPQEPWDHKALA RFDSPLWRHHALFQDAVNYYQLCLVALAS SDGTRPLSKLHEQMKASWDEAKTDTEDS WRVRLARRLGIPAASLFEAALAKVLEGNE APERARELAGELLLDKIEGDIQQAGRGYWP RFCDPKANPTYDYSATARASASGLTKLAA VIHAENVTEEALKQVAAEMDLSWTVKLQP DKNFVGAEARARLLEAAHHFIKVAESPPTK LAEVLARFPDGLALWQALPEKIAALPEETQ VPRNRKASPDLTFATLLFQHFPSLFTAAVL GLSVGKPKSVKAPKVVEKVSARRKANAVT QAVVIEEPEIDFAELGDDPIKLARGERGFVF PAFTSLSFWAVPGPHVPVWKEFDIAAFKEA LKTVNQFKLKTSERNALLAEAQRRLDYMD EKTHDWKTGDSDEPGHIPPRLKSDPNFTLI QALTQDEGVSNKATGDQHIPKGVYTGGLR GFYAIKKDWCELWERKADKSQGTPTEEELI SIVTDYQRDHVYDVGDVGLFRALCEPRFW PLWQPLTDEQEAERIKAGRAKDMISAYRV WLELQEDVVRLAQPIRFTPAHAENSRRLFM FSDISGSHGAEFGSDGKSLEVSIAYDVDGK LQPVRAKLEFSAPRAARDELEGLSGGSESM RWFQPMMKALDCPEVEMPALEKCAVSLM PDVVKKGGGKWVRLLLNFPATLEPEGLIR HIGKQAMWYKQFNGTYKPRTQQLDTGLH LYWPGLEKAPEAEDAAAWWNREEIRAKG FSVLSVDLGQRDAGAWALLESRSDKAFSR NRQPFIELGEAGGKLWSTALLGLGMLRLP GEDARTGALDDQGKRAVEFHGKAGRNAL EAEWQEAREMALLFGGEEAKSRLGPGFDH

[1887]

			LSHSKQNEELLRILSRAQSRLARFHRWSCRI HEKPEATGDDVIDYGQVDELLTKTAEAML ENLKALYTNAGGILDSKSKQPLTLVGLRKK LEAQKVEPEKIAAVLKPHAEIIFQRLGTLIPE LKQHLRVSLERLANRELPLRHREWVVWNEA FEKLEQGNFKKEENPKWIRGQRGLSMARIE QIENLRKRFMSLRRQMSLIPGEQVKQGVED KGQRQPEPCEDILNKLDRMKQQRVNQTAH LILAQALGLRLRPHLANDAEREEKDIHGEY ELIPGRKPVDFIVMEDLSRYLSSQGRAPSEN GRLMKWCHRAVLAKLKQMCEPFGIPVLEV PAAYSSRFCALTGVPGFRAVEVHDGNAED FRWKRLIKKAEKDKSSKDAEAAAMLFDQL HDLNIEAREARKQDKKLPLRTLFAPVAGGP LFIPMVGGGPRQADMNAAINLGLRAIASPT CLRARPKIRAELKDGKHQAMLGNKLEKAA ALTLEPPKEPTKELAAQKRTNFFLDEKFVG KFDTAHVTTSGKKLRLSGGMSLWKAIKDG AWQRVKKINDARIAKWKNNPPPEPDPDDE IQF

[1888]

[1889] 표 15는 표 14에 표시된 Ls, Ak, Bv, Phyci, 및 Planc의 Cas12b 오솔로그에 대한 crRNa, tracrRNA 및 sgRNA의 예시적인 서열을 표시한다. 도15A-15C 는 PAM 발굴, 각각 Ls, Ak 및 Bv에서 Cas12b 오솔로그를 사용한 정제된 단백질 및 RNA에 의한 시험관내 절단을 도시한다. 도 15D-15E 는 각각 Phyci 및 Planc의 Cas12B 오솔로그를 사용한 정제된 단백질 및 RNA에 의한 시험관내 절단을 도시한다.

[1890] 표 15

명칭	crRNA	tracrRNA	sgRNA
Ls	UAGAUGAAUUAAAUGUGAUUAGCAC (SEQ ID NO:536)	UUUGCCUAAAGGGCAAAGAAUACUGUGCGUGUGCUAAGGAUGGAAAAAUCCAUUCAACCACAGGAUUACAUUAUUUAUC (SEQ ID NO:537)	UUUGCCUAAAGGGCAAAGAAUACUGUGCGUGUGCUAAGGAUGGAAAAAAUCCAUUCAACCACAGGAUUACAUUAUUUAUCAAAAGAUGAAUAAAUGUGAUUAGCAC (SEQ ID NO:538)
Ak	AGCGAGCGGUCUGAGAAGUGGCACU (SEQ ID NO:539)	GUCGUCUAUAGGACGGCGAGGACAACGGGAAGUGCCAAUGUGCUCUUUCCAAGAGCAAACACCCCGUUGGCUUCAAGAUGACCGCUCGC (SEQ ID NO:540)	GUCGUCUAUAGGACGGCGAGGACAACGGGAAGUGCCAAUGUGCUCUUUCCAAGAGCAAACACCCCGUUGGCUUCUCAGACCGCUCGAAAACGAGCGGUCUGAGAAGUGGCACU (SEQ ID NO:541)
Bv	GCAGAAAUAAUGAUGAUUGGCAC (SEQ ID NO:542)	GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGGAUUAUCUUAUUUCUGC (SEQ ID NO:543)	GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUAAGUAAUUAAAAAUUACCCACCACAGGAUUAUCUUAUUUCUGCAAAAGCAGAAAUAAGAUGAUUGGCAC (SEQ ID

[1891]

			NO:544)
Phyci			UCAGCCAACAUGCU CGCUUUGCGAAGGC UGACGGCCCGCUCU CAUUUGGCAUUGCC GGGAGCCGGAGUUU UCGGAAGAGAGUGU CGACGACUGCUGAU CUCCGCAUCCGCGU CCUGUUCGCCAGGC CGGGUCGGGUGUAC GGAUCAUGCUGGCA GCAGUCUACGCCGA GAACAUUCGCUACC GCGAAUGGACAC (SEQ ID NO:545)
Planc			GUGGAGUAAGGUCG GAGUAACGACCGAA CGUUUAGCGUGCUA UAGGCCGCUGAAUG CCACACAGCGAUGU GUUUUGAGUGUCAA UAGCUGCUGACCCA AAGGCCAAAAGCCG AGUAGGGCUUGACU GAUGCGGUUUAUAU CGCACAUAGGCGGC AGUAACACAUAUCG CGUCAAUCAAAUUU AUUGAUGGACAC (SEQ ID NO:546)

[1892]

[1893] **실시예 7**

[1894] 알리시클로바실러스 마크로스포란지이더스 Cas12b에 대한 예시적인 직접 반복부 서열, crRNA 서열, tracrRNA 서열 및 sgRNA가 하기 표 16에 표시된다.

[1895] **표 16**

Am Cas12b 직접 반복부 서열	
21nt	GGGCGAUCUGAGAAGUGGCAC (SEQ ID NO:547)
29nt	CGUUGAGCGGGCGAUCUGAGAAGUGGCAC (SEQ ID NO:548)
전체-길이	GUCGGAUCGUUGAGCGGGCGAUCUGAGAAGUGGCAC (SEQ ID NO:549)
crRNA	
nC1523 (21nt DR + 23nt 가이드)	CGATCTGAGAAGTGGCACGAGAAGTCATTTAATAAGGCCAC (SEQ ID NO:550)
tracr RNA	
nC1518	CCGAUCUAUAGGACGGCAGAUUCAACGGGAUG (SEQ ID NO:551)
nC1519	CCGAUCUAUAGGACGGCAGAUUCAACGGGAUGUGCCAA (SEQ ID NO:552)
nC1520	CCGAUCUAUAGGACGGCAGAUUCAACGGGAUGUGCCAAUGCACUCUUUCCAGGAGUGAACA (SEQ ID NO:553)
nC1521	CCGAUCUAUAGGACGGCAGAUUCAACGGGAUGUGCCAAUGCACUCUUUCCAGGAGUGAACACCCCGUUGGCUUCAACAUGAUCGCCCGCUCAACGGU (SEQ ID NO:554)
nC1522	UGCCAAUGCACUCUUUCCAGGAGUGAACACCCCGUUGGCUUCAACAUGAUCGCCCGCUC (SEQ ID NO:555)
sgRNA	
sgRNA1 full tracr-CTA loop-full DR	GCCGAUCUAUAGGACGGCAGAUUCAACGGGAUGUGCCAAUGCACUCUUUCCAGGAGUGAACACCCCGUUGGCUUCAACAUGAUCGCCCGCUCAACGGUCCCUAGUCGGAUCGUUGAGCGGGCGAUCUGAGAAGUGGCAC (SEQ ID NO:556)
sgRNA2 96nt tracr-CTA loop- 29nt DR	GCCGAUCUAUAGGACGGCAGAUUCAACGGGAUGUGCCAAUGCACUCUUUCCAGGAGUGAACACCCCGUUGGCUUCAACAUGAUCGCCCGCUCAACGCUACGUUGAGCGGGCGAUCUGAGAAGUGGCAC (SEQ ID NO:557)
sgRNA3 88nt tracr-CTA loop-21nt DR	GCCGAUCUAUAGGACGGCAGAUUCAACGGGAUGUGCCAAUGCACUCUUUCCAGGAGUGAACACCCCGUUGGCUUCAACAUGAUCGCCCCUAGGGCGAUCUGAGAAGUGGCAC (SEQ ID NO:558)

[1896]

[1897] 도 16 은 정제된 AmCas12b (AmC2C1) 단백질 및 소형 RNAseq로부터의 상이한 예상 tracr RNA에 의한 시험관내 절단 어세이를 도시한다. 우리는 TTA가 이 시점에 C2c1에 대한 공통 PAM이므로 TTTA PAM을 사용하였다.

[1898] 다양한 sgRNA 디자인이 도 17A-17E에 도시된다. 도 17은 tracr RNA (붉은색)에 어닐링된 전장 AmC2C1 직접 반복부 서열 (녹색)을 도시한다. Tracr은 소형 RANseq로 예측되었고 시험관내에서 확인되었다. 파란색 원형 = 5' 말단; 붉은색 원형 = 3' 말단. 도 17B는 tracr RNA (붉은색)에 어닐링된 21 nt의 AmC2C1 직접 반복부 서열 (녹색)을 도시한다. Tracr은 소형 RANseq로 예측되었고 시험관내에서 확인되었다. 파란색 원형 = 5' 말단; 붉은색 원형 = 3' 말단. 도 17C는 전장 직접 반복부 및 tracr과 CTA 루프의 융합을 도시한다. 도 17D는 29nt의 직접 반복부 및 tracr과 CTA 루프를 도시한다. 도 17E는 21nt의 직접 반복부 및 tracr과 CTA 루프를 도시한다.

[1899] 도 18 은 sgRNA 효율의 비교를 위한 AmC2C1에 의한 시험관내 절단을 도시한다.

[1900] AmC2C1 RuvC의 돌연변이체를 생성시켰고 그들 활성은 HEK 세포 용해물을 사용해 시험하였다 (도 19).

[1901] Cas12b 오솔로그에 대한 PAM은 시험관내 PAM 스크린으로 결정하였다. 간략하게, Cas12b 단백질 및 sgRNa는 PAM 라이브러리 플라스미드오 인큐베이션시켰다. 결과를 도 20 에 나타낸다.

[1902] **실시예 8**

[1903] 바실러스 히사시이 Cas12b (BhC2C1)를 정제하였고 이의 활성을 상이한 온도에서 시험하였다. 도 21A-21D는 소형 RNAseq tracr 예측, 생체내 스크린으로부터의 BhC2C1 (바실러스 히사시이 Cas12b) PAM, BhC2C1 단백질 정제, 각각 37℃ 및 48℃에서 BhC2c1 단백질 및 예측 tracrRNA에 의한 시험관내 절단을 도시한다.

[1904] 도 2A-22D는 BhC2C1 sgRNA 디자인을 도시한다. 예를 들어, 도 22A는 20 nt 직접 반복부 (녹색) 및 예측 tracr RNA (붉은색)를 도시한다.

[1905] BhC2C1의 직접 반복부 서열, tracr RNA 서열, 및 sgRNA 서열은 하기 표 17에 표시된다.

[1906] **표 17**

BhC2C1 직접 반복부 서열	
20 nt	AUGAUACGAGGCAUUAGCAC (SEQ ID NO:559)
전체 길이	GUCCAAGAAAAAAGAAAUGAUACGAGGCAUUAGCAC (SEQ ID NO:560)
tracr RNA	
JS1568	ACGAGGUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUA AGUGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGU CUCUUUU (SEQ ID NO:561)
sgRNA	
sgRNA1	GAGGUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUAAG UGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGUCU CUACGAGGCAUUAGCAC (SEQ ID NO:562)
sgRNA2	ACGAGGUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUA AGUGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGU CUCUAUGAUACGAGGCAUUAGCAC (SEQ ID NO:563)
sgRNA3	ACGAGGUUCUGUCUUUUGGUCAGGACAACCGUCUAGCUAUA AGUGCUGCAGGGUGUGAGAAACUCCUAUUGCUGGACGAUGU CUCUUACGAGGCAUUAGCAC (SEQ ID NO:564)

[1907]

[1908] BhC2c1을 플라스미드에 클로닝하였다. 플라스미드의 맵은 도 23에 도시된다. 스캐폴드는 다음과 같았다:

[1909] GTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGTGTGAGAAACTCCTATTGCTGGACGACGCCTCTTACGAGGCGTTAGCACn23_spacer (SEQ ID NO:565).

[1910] **실시예 9**

[1911] 도 24 는 상이한 표적 부위에 대한, 플라스미드 형질감염 이후, 상이한 sgRNA로 수득된 indel 백분율을 도시한다. 사용된 Cas12b는 바실러스 V3-13 (Wp_101661451)로부터의 것이었다. 단백질 서열, sgRNA 서열, 및 표적화 부위는 하기 표 18에 표시된다.

[1912] **표 18**

BvCas12b 단백질 서열	
NLS-BvCas12b-NLS-3xHA (BvCas12b 밑줄)) (SEQ ID NO:566)	MAPKKKRKVGIHGVPAAAIRSIKLKMKTNSGTDSIYL RKALWRTHQLINEGIAYYMNLLTLYRQEAIGDKTKE AYQAELINIIRNQQRNNGSSEEHGSDQEILALLRQLYE LIIPSSIGESGDANQLGNKFLYPLVDPNSQSGKGTSNA GRKPRWKRLKEEGNPDWELEKKKDEERKAKDPTVK IFDNLNKYGLLPLFPLFTNIQKDIEWLPLGKRQSVRK WDKDMFIQAIERLLSWESWNRRVADEYKQLKEKTE SYYKEHLTGGEEWIEKIRKFEKERNMELEKNAFAPN

[1913]

	DGYFITSRQIRGWDRVYEKWSKLPESASPEELWKVV AEQQNKMSEGFGDPKVFSFLANRENRDIWRGHSERI YHIAAYNGLQKKLSRTKEQATFTLPDAIEHPLWIRYE SPGGTNLNLFKLEEKQKKNYYVTLSKIIWPSEEKWIE KENIEIPLAPSIQFNRQIKLKQHVKGKQEISFSDYSSRI SLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFFNL VVDVAPLQETRNGRLQSPIGKALKVISSDFSKVIDYK PKELMDWMNTGSASNSFGVASLLEGMRVMSIDMGQ RTSASVSIFEVVKELPKDQEQKLFYSINDTELFAIHKR SFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRML ANVLRLETKKTPDERKKAIHKLMEIVQSYDSWTASQ KEVWEKELNLLTNMAAFNDEIWKESLVELHHRIEPY VGQIVSKWRKGLSEGRKNLAGISMWNIDELEDTRRL LISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDR LKQMANLIIMTALGFKYDKEEKDRYKRWKETYPAC QIILFENLNRYLFNLDRSRRENSRLMKWAHRSIPRTV SMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCHAL TEEDLKAGSNTLKRLIEDGFINESELAYLKKGDIIPSQ GGELFVTLSKRYKKDSDNNELTVIHADINAAQNLQK RFWQQNSEVYRVPCQLARMGEDKLYIPKSQTETIKK YFGKGSFVKNNTEQEVYKWEKSEKMKIKTDTTFDLQ DLDGFEDISKTIELAQEQQKKYLTMFRDPSGYFFNNE TWRPQKEYWSIVNNIIKSCLKKKILSNKVELKRPAAT KKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYD VPDYA*
sgRNA 서열	
sgRNA.1 (SEQ ID NO:567)	GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCA UAAGUAAUUAAAAAUUACCCACCACAGGAUUAUC UUAUUUCUGCAAAAGCAGAAAUAAGAUGAUUGG CAC-nnnnnnnnnnnnnnnnnnnnnnn (23 nt 스페이서)
sgRNA.2 (SEQ ID NO:568)	GGUGACCUAUAGGGUCAAUGAAUCUGUGCGUGU GCCAUAAGUAAUUAAAAAUUACCCACCACAGGAU UAUCUUAUUUCUGCAAAAGCAGAAAUAAGAUGA UUGGCAC-nnnnnnnnnnnnnnnnnnnnnnn (23 nt 스페이서)
sgRNA.3 (SEQ ID NO:569)	CUAUAGGGUCAAUGAAUCUGUGCGUGUGCCAUA AGUAAUUAAAAAUUACCCACCACAGGAUUAUCUU AUUUCUGCAAAAGCAGAAAUAAGAUGAUUGGCA C-nnnnnnnnnnnnnnnnnnnnnnn (23 nt 스페이서)
sgRNA.4 (SEQ ID NO:570)	GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCA UAAGUAAUUAAAAAUUACCCACCACAGGAUCAUC UUAAAAAAGAUGAUUGGCAC- nnnnnnnnnnnnnnnnnnnnnnn (23 nt 스페이서)

[1914]

sgRNA.5 (SEQ ID NO:571)	GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCA UAAGUAAUUAAAAAUUACCCACCACAGGAUCAUC UUAUUUCAAAAGAAAUAAGAUGAUUGGCAC-nnnnnnnnnnnnnnnnnnnnnnn (23 nt 스페이서)
sgRNA.6 (SEQ ID NO:572)	GACCUAUAGGGUCAAUGAAUCUGUGCGUGUGCCA UAAGUAAUUAAAAAUUACCCACCACAGGAGCACC UGAAAACAGGUGCUUGGCAC-nnnnnnnnnnnnnnnnnnnnnnn (23 nt 스페이서)
표적 부위	
DNMT1-1: (SEQ ID NO:573)	CCCTTCAGCTAAAATAAAGGAGG
DNMT1-2: (SEQ ID NO:574)	GGCTCAGCAGGCACCTGCCTCAG
DNMT1-3: (SEQ ID NO:575)	ACGTACTGATGTTAACAGCTGAC
VEGFA-1: (SEQ ID NO:576)	GGGACTGGAGTTGCTTCATGTAC
VEGFA-2: (SEQ ID NO:577)	CTGACCTCCCAAACAGCTACATA
EMX1-1: (SEQ ID NO:578)	TTCATGGAGAAAATATTCAGAAT
EMX1-2: (SEQ ID NO:579)	TCTCCATGAAAAATACTGGGGTC

[1915]

[1916] **실시예 10**

[1917] BvCas12b (바실러스 sp. V3-13 Cas12b)는 플라스미드 (pcDNA3-BvCas12b)에 클로닝하였다. 플라스미드의 맵은 도 25에 도시된다. 클로닝된 구성체의 서열은 하기 표 19 에 제시되어 있다.

[1918] **표 19**

pcDNA3-BvCas12b 의 서열
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT

[1919]

CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT
CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG
TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCG
ACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTA
GTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGA
AGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATA
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG
CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCA

[1920]

[1921]

ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG
GCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCA
CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTA
TCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAA
GCTTGCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAG
CAGCCGGATCCgccgtgaagtccatcaaagtgaagctgcggctgagcgagtgccccgatattctggctggaatgtggcagctgc
acagagccacaaatgccggcgtgcggtactacacagaatgggtgtccctgatgcggcaagagatcctgtacagcagaggccctgatggc
ggccagcagtgttatatgaccgccgaggattgccagagagagctgctgcggagactgcggaatagacagctgcataacggccggcagg
atcagcctggaacagatgctgatctgctggccatcagcagacggctgtacgagattctggtgctgcagagcatcggcaaaagaggcgacg
cccagcagattgccagcagctttctgagccctctggtggaccccaacagcaaaggtggaagaggcgaggccaagagcggaagaaaacc
tgcctggcagaagatgcgcgaccagggcgatcctagatgggttgccgctagagagaagtacgagcagcggaaggccgtggatcccagc
aaagagattctgaacagcctggacgccctgggcctcagacctctgtttgccgtgttcaccgagacatacagatccggcgtggactggaagc
ctctgggcaaatctcagggcgtcagaacctgggacagagacatgtttcagcaggccctggaacggctgatgagctgggagagctggaat
cggagagtgggcgaagagtacgccagactgttccagcagaaaatgaagttcgagcaagagcacttcgccgagcagagccacctggtca
aactggctagagccctggaagccgatatgagagccgcctctcagggcttcgaggccaaaagaggaacagcccaccagatcaccagaag
ggcactgagaggggccgacagagtgttcgagatctggaagtctatccccgaggaagccctgttcagccagtacgacgaagtgatcagac
aggtgcaggccgagaagcggagagatttcggcagccatgacctgttcgccaagctggccgagcctaagtatcagccccttggagagcc
gacgagacattcctgaccagatacgccctgtacaacggcgtgctgcgcgatctggaaaaggccagacagttcgccaccttcacactgcct
gatgcctgcgtgaaccccatctggaccagattcgagtctagccagggcagcaacctgcacaaatacgagtttctgttcgaccacctcggac
ctggcagacacgccgtcagatttcagagactgctggtggtggaaagcgagggcgccaaagaaagggatagcgtggtggtgcctgtggct
ccttctggccaactggataagctggtgctgagggaagaagagaagtccagcgtcgccctgcatctgcacgataccgctagacccgatggc
ttcatggctgaatgggctggcgccaaactgcagtacgagagaagcaccctggccagaaaagccagacgggacaagcagggcatgaga
agctggcggagacagccctccatgctgatgtctgccgctcagatgctggaagatgccaaacaggctggcgacgtgtacctgaacatcagc
gtgcgcgtgaagtctcccagtgaagtgcgaggacagaggcggcctccttacgccgctctgtttagaatcgacgacaagcagcggagagt
gaccgtgaactacaacaagctgagcgcctacctggaagaacaccccgataagcagatccctggcgctcctggactgctgtctggactgag
agtgatgtccgtggacctgggcctgagaacaagcgccagcatctccgtgttcagagtggccaagaaagaagaggtggaagccctcggag
atggccggcctcctcactactatcctatccacggcaccgatgacctggtggccgtgcacgaaagatcccacctgattcagatgcccggcga
aaccgagacaaagcagctgcggaagctgagagaagaacggcaggccgtgctgaggccactgtttgctcaactggcactgctgagactgc

[1922]

tcgtcagatgtggcgccgctgacgagagaatcagaaccagatcctggcagcggctgaccaagcagggaagagagttcaccaagagact
gacccctagctggcgcgaggctctggaactggaactgacaagactcgaggcctactgcggcagagtgcccgatgatgagtggtccagaa
tcgtggacagaaccgtgattgccctgtggcggagaatgggcaagcaagtgcgcgattggcggaagcaagtgaagtccggggccaaagt
gaaagtgaagggctaccagctggatgtcgtcggcggaaattctctggcccagatcgactatctggaacagcagtacaagttcctgcggcgt
tggagcttcttcgccagagcttctggcctggtcgtgcgggccgatagagaaagccattttgccgtggctctgagacagcacatcgagaacg
ccaagcgggacagactgaagaaactggccgaccggatcctgatggaagcactgggctatgtgtacgaggccagcggacctagagaagg
ccagtggacagctcagcaccctccttgccagctgatcattctcgaggaactgtccgcctaccggttcagcgacgatagacctcctagcgag
aacagcaaactgatggcctggggccacagaggcatcctcgaagaactggtcaaccaggctcaggtgcacgatgtgctcgtgggcacagt
gtacgccgccttcagcagcagattcgacgctagaacaggtgctcccggcgtcagatgcagaagagtgcctgccagatttgtgggcgccac
cgtggatgattctctgccactgtggctgaccgagttcctggacaagcaccggctggataagaacctgctgcggcccgacgatgtgatccca
acaggcgaaggcgaattcctggtgtccccttgtggcgaagaggctgccagagttagacaggttcacgccgacatcaacgctgcccagaac
ctgcagagaaggctgtggcagaacttcgacatcaccgagctgaggctgagatgcgacgtgaagatgggcggagagggaacagtgctgg
tgcccagagtgaacaacgccagagccaagcagctgttcggcaagaaggtgctggtttcccaggacggcgtgaccttcttcgagagatctc
agacaggcggcaagccccacagcgagaagcagaccgatctgaccgacaaagaactcgagctgatcgccgaggccgatgaggccaga
gctaaaagcgtggtgctgttcagggatcctagcggccacattggcaaaggccactggatccggcagcgcgagttttggagtctggtcaagc
agaggatcgagagccacaccgccgagcggattagagttagaggcgtgggaagctccctggacGGATCCAAAAGGCCGG
CGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGggatctTACCCATACGATGT
TCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCC
GACTATGCCtaaGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAG
GGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTG
TTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCT
TTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCT
GGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG
CATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGG
CTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGG
TGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCG
CTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACT
TGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCC
TTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAAC
ACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCC

TATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGA
ATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATG
CAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCC
AGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGC
CCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCA
TGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTA
TTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCG
GGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGC
ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCT
ATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCG
GCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCT
GAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTC
CTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGG
GCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTAT
CCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCAT
TCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGT
CTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACT
GTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATG
GCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCG
ACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTG
ATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTA
TCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTG
AGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACG
AGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCG
GGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCA
CCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAA
TTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATC
AATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCAT
GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACA
TTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTG

[1923]

CATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCA
GCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA
(SEQ ID NO:580)

[1924]

[1925] **실시예 11**

[1926] BhCas12b (바실러스 히사시이 Cas12b)는 플라스미드 (pcDNA3-BhCas12b)에 클로닝하였다. 플라스미드의 맵은 도 26에 도시된다. 클로닝된 구성체의 서열은 하기 표 20 에 제시되어 있다.

[1927] 표 20

pcDNA3-BhCas12b 의 서열
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT
CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT
CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG

[1928]

[1929]

TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCG
ACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTA
GTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGA
AGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATA
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG
CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCA
ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG
GCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCA
CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTA
TCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAA
GCTTGCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAG
CAGCCGGATCCGCCACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTG
AAGAAAGGCCTCTGGAAAACCCACGAGGTGCTGAACCACGGAATCGCCTACTACAT
GAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACC
CCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTC
GTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGT

GTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGA
AGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAAC
AGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACC
TGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGA
TAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGA
TCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGT
GGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATT
CAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGA
ATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGAC
ATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCT
GCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCT
GGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAA
GTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATT
ACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCAC
CCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGA
CGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGT
CCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGC
AGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATC
TACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCT
GCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGC
ACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGC
GGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGA
AAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAG
AGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCA
ACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACT
GAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGG
GACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGAC
ATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGA
GCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCT
GCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGC
GGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTC

[1930]

[1931]

ACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGA
GCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCT
GAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGG
CGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACAT
CGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCG
AACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTG
AATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCAT
CATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGA
ACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGAGG
AAAGGTCCCGCTTCGAGAACAGCAAGCTCATGAAGTGGTCCAGACGCGAGATCCCC
AGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGC
TCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTAGCG
TCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAG
GGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGA
CAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACAC
ACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCAC
GGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACAT
CCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACT
TCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAG
AAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAG
ACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGAC
CCCAGCGGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGA
AAGCTGGAACGCATCCTGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCAT
CGAGGACGACAGCAGCAAGCAGTCTATGGGATCCAAAAGGCCGGCGGCCACGAAA
AAGGCCGGCCAGGCAAAAAAGAAAAAGggatctTACCCATACGATGTTCCAGATTACG
CTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCtaa
GAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAA
ACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCT
CCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGT
GGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAT

GCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTA
TCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTC
CTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTA
GGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGAT
GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAG
TCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCT
CGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAA
TGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTA
GGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCT
CAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTA
TGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTT
TTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTG
AGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATC
CATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGA
TGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTG
GGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGG
GGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGG
ACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTG
CTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGG
GCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGA
TGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGC
GAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGG
ATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTC
AAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTT
GCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCT
GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAG
AGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCG
ATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTG
GGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTC

[1932]

CACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTG
GATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTT
TATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAT
CATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTG
TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGC
ATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATC
GGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTC
ACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAG
GCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA (SEQ ID NO:581)

[1933]

[1934] **실시예 12**

[1935] EbCas12b (엘루시미크로비아 박테리움 Cas12b)는 플라스미드 (pcDNA3-EbCas12b)에 클로닝하였다. 플라스미드의 맵은 도 27에 도시된다. 클로닝된 구성체의 서열은 하기 표 21 에 제시되어 있다.

[1936] 표 21

pcDNA3-EbCas12b 의 서열
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT
CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT
CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG
TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG

[1937]

CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCG
ACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTA
GTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGA
AGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATA
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG
CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCA
ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG
GCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCA
CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTA
TCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAA
GCTTGCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAG
CAGCCGGATCCAACCGGATCTACCAGGGCAGAGTGACCAAGGTGGAAGTGCCCGAT
GGCAAGGACGAGAAGGGCAACATCAAGTGGAAGAAGCTGGAAAATTGGAGCGACA
TCCTGTGGCAGCACCACATGCTGTTCCAGGACGCCGTGAACTACTACACACTGGCCC
TGGCCGCCATCTCTGGATCTGCTGTTGGCAGCGACGAGAAGTCCATCATCCTGAGAG
AATGGGCCGTGCAGGTCCAGAACATCTGGGAGAAAGCCAAGAAAAAGGCCACCGT
GTTCGAGGGCCCACAGAAGAGACTGACCAGCATCCTGGGCCTTGAGCAGAACGCCA
GCTTCGACATTGCCGCCAAGCACATCCTGAGGACCTCTGAGGCCAAGCCTGAGCAG
AGAGCTAGCGCCCTGATCAGACTGCTGGAAGAGATCGACAAGAAAAACCACAACGT
CGTGTGCGGCGAGCGGCTGCCTTTTTTCTGCCCTCGGAACATCCAGAGCAAGCGGAG
CCCTACAAGCAAGGCCGTGTCTAGCGTGCAAGAGCAGAAACGGCAAGAGGAAGTG
CGGCGGTTCCACAACATGCAGCCTGAGGAAGTGGTCAAGAACGCCGTGACACTGGA
CATCAGCCTGTTCAAGAGCAGCCCCAAGATCGTGTTCCTGGAAGATCCCAAGAAGG
CCAGAGCCGAGCTGCTGAAGCAGTTCGACAACGCCTGCAAGAAACACAAAGAACTC
GTGGGCATCAAGAAAGCCTTCACCGAGTCCATCGACAAGCACGGCTCTAGCCTGAA
GGTGCCAGCTCCTGGCTCTAAGCCTAGCGGCCTGTATCCTAGCGCCATCGTGTTCAA
GTACTTCCCCGTGGATATTACCAAGACCGTGTTTCTGAAGGCCACAGAGAAGCTGGC
CATGGGCAAAGACCGGGAAGTGACCAACGATCCTATCGCCGACGCCAGAGTGAACG
ACAAGCCCCACTTCGACTACTTCACCAACATTGCCCTGATCCGCGAGAAAGAGAAG
AACAGAGCCGCTTGGTTTGAGTTCGATCTGGCCGCCTTTATCGAGGCCATCATGAGC
CCTCACAGATTCTACCAGGACACCCAGAAGCGGAAAGAGGCCGCCAGAAAGCTGGA
AGAAAAGATCAAGGCCATCGAAGGCAAAGGCGGGCAGTTCAAAGAGAGCGACAGC
GAGGACGACGACGTGGACTCTCTGCCTGGATTTGAGGGCGACACCAGAATCGACCT

[1938]

[1939]

GCTGCGGAAGCTGGTCACCGATACACTTGGATGGCTGGGCGAGAGCGAGACACCCG
ATAACAACGAGGGCAAAAAGACCGAGTACAGCATCAGCGAGCGGACCCTGAGAAT
CTTCCCCGACATCCAGAAGCAGTGGAGCGAGCTGGCCGAGAAAGGCGAGACAACA
GAGGGAAAGCTGCTCGAAGTGCTGAAACACGAGCAGACCGAGCACCAGAGCGATTT
CGGAAGCGCCACACTGTATCAGCACCTGGCCAAGCCAGAGTTTCACCCCATCTGGCT
GAAGTCCGGCACCGAGGAATGGCACGCCGAGAATCCTCTGAAAGCCTGGCTGAACT
ACAAAGAGCTGCAGTACGAGCTGACCGACAAGAAGCGGCCCATCCACTTTACCCCT
GCTCACCCTGTGTACAGCCCCAGATACTTCGACTTCCCCAAGAAGTCCGAAACCGAA
GAGAAAGAGGTGTCCAAGAACACCCACAGCCTGACCACAAGCCTGGCCAGCGAGC
ACATCAAGAACTCCCTGCAGTTTACAGCCGGGCTGATCAGAAAGACCAACGTGGGC
AAGAAGGCTATCAAGGCCCGGTTCAGCTACAGCGCCCCTAGACTGAGAAGGGACTG
CCTGAGAAGCGAGAACAACGAGAACCTGTACAAGGCCCCTTGGCTCCAGCCTATGA
TGAGAGCCCTGGGCATCGACGAGGAAAAGGCCGACAGACAGAACTTCGCCAACACC
AGGATCACCCTGATGGCCAAAGGCCTGGACGACATTCAGCTGGGCTTTCCCGTGGA
AGCCAACAGCCAAGAACTGCAGAAAGAAGTGTCTAACGGCATCAGCTGGAAGGGC
CAGTTCAACTGGGGAGGAATCGCCTCTCTGTCTGCCCTGAGATGGCCCCACGAGAA
GAAGCCCAAGAATCCTCCTGAGCAGCCTTGGTGGGGCATCGATAGCTTTAGCTGCCT
GGCCGTGGATCTGGGCCAGAGATATGCTGGCGCCTTCGCCAGACTGGACGTGTCCA
CCATTGAGAAAAAGGGCAAGAGCCGGTTCATCGGCGAGGCCTGCGACAAAAAGTGG
TACGCCAAGGTGTCCCGGATGGGCCTGTTGAGACTTCCTGGCGAGGACGTGAAAGT
GTGGCGGGATGCCAGCAAGATTGACAAAGAGAACGGCTTCGCCTTCCGGAAAGAGC
TGTTCGGCGAGAAGGGAAGATCCGCCACACCTCTGGAAGCCGAGGAAACCGCCGAG
CTGATCAAGCTGTTTGGAGCCAACGAGAAGGACGTGATGCCCGACAACTGGTCTAA
AGAGCTGAGCTTCCCCGAGCAGAATGACAAGCTGCTGATCGTGGCTCGGAGAGCCC
AGGCTGCTGTTAGCAGACTGCATAGATGGGCATGGTTCTTCGACGAGGCCAAGAGA
TCCGACGACGCCATCAGAGAGATTCTGGAAAGCGACGACACCGACCTGAAGCAGAA
AGTGAACAAGAACGAGATCGAGAAAGTCAAAGAGACAATCATCTCCCTGCTGAAAG
TCAAGCAAGAGCTGCTGCCCACACTGCTGACCAGACTGGCCAATAGAGTGCTGCCC
CTGAGAGGCAGATCCTGGGAGTGGAAAAAGCACCACCAGAAGAACGACGGCTTCAT
CCTGGACCAGACCGGCAAGGCCATGCCTAACGTGCTGATTAGAGGACAGCGGGGCC
TGAGCATGGACCGGATCGAGCAGATTACCGAGCTGAGAAAGCGGTTTCAGGCCCTG
AACCAGAGCCTGCGGAGACAGATCGGAAAGAAGGCCCCTGCCAAGCGGGACGACT
CTATCCCTGATTGCTGCCCCGATCTGCTGGAAAAACTGGACCACATGAAGGAACAG
CGCGTGAACCAGACAGCCCACATGATTCTGGCCGAGGCACTGGGACTGAAACTGGC
CGAGCCTCCTAAGGACAAGAAAGAACTGAACGAGACATGCGACATGCACGGCGCCT
ACGCCAAAGTGGACAACCCCGTGTCCTTCATCGTGATCGAGGACCTGAGCCGGTAC
AGAAGCAGCCAAGGCAGAAGCCCCAGAGAAAACAGCCGACTGATGAAGTGGTGCC
ACAGGGCCGTCAGAGACAAGCTGAAAGAAATGTGCGAGGTGTTCTTCCCACTGTGC
GAGAGAAGAAAGGCCGGCTCTGCTTGGGTTTCCCTGCCTCCTCTGCTTGAAACACCA
GCCGCCTACAGCAGCAGATTCTGCAGCAGATCTGGCGTGGCCGGCTTCAGAGCCGT
GGAAGTGATTCCTGGCTTCGAGCTGAAGTACCCCTGGTCTTGGCTGAAGGATAAGAA
GGACAAGGCCGGCAATCTGGCCAAAGAAGCCCTGAACATCCGGACCGTGTCTGAGC
AGCTGAAGGCCTTTAACCAGGACAAGCCCGAGAAGCCCAGGACACTGCTGGTGCCT
ATTGCCGGCGGACCTATCTTCGTGCCCATCTCTGAAGTGGGCCTGTCCAGCTTCGGA
CTGAAGCCTCAGGTTGTGCAGGCCGACATCAACGCCGCCATCAATCTGGGACTCAG
AGCCATCAGCGACCCTCGGATTTGGGAGATTCACCCCAGACTGCGGACCGAGAAGA

GAGATGGCAGACTGTTCGCCAGAGAGAAACGGAAGTACGGCGAAGAGAAGGTCGA
GGTGCAGCCCAGCAAGAATGAGAAGGCCAAAAAAGTGAAGGACGACCGGAAGCCT
AACTACTTCGCCGATTTCAGCGGCAAGGTGGACTGGGGCTTTGGCAACATTAAGAAC
GAGTCCGGCCTGACACTGGTGTCTGGCAAAGCACTGTGGTGGACCATCAACCAGCT
GCAGTGGGAGAGATGCTTTGACATCAACAAGCGGCACATCGAGGACTGGTCCAACA
AGCAGAAGCAAGGATCCAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAA
AAAGAAAAAGggatctTACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGC
CTGATTATGCATACCCATATGATGTCCCCGACTATGCCtaaGAATTCTGCAGATATCCA
GCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTC
GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTG
ACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCG
CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA
GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG
CTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGT
AGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACT
TGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTC
GCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGT
GCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGG
CCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATA
GTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGA
TTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACA
AAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCC
CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAAC
CAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATC
TCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTC
CGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGA
GGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGG
AGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGA
TCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAG
GTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACA
ATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTT
TTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCG
GCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCAC
TGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGT
CATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGC
TGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCG
AGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAA
GAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCC
CGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGT
GGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCG
CTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATG
GGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGC
CTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACC
GACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTAT
GAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGC
GGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATG

[1940]

GTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGC
ATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTC
GACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTG
TTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCT
GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTT
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGG
AGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCT
CGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTAT
CCACAGAATCAGGGGATAACGCAGGAA (SEQ ID NO:582)

[1941]

[1942] **실시예 13**

[1943] AkCas12b (알리시클로바실러스 카케가웬시스 Cas12b)는 플라스미드 (pcDNA3-AkCas12b)에 클로닝하였다. 플라스미드의 맵은 도 28에 도시된다. 클로닝된 구성체의 서열은 하기 표 22 에 제시되어 있다.

[1944] 표 22

pcDNA3-AkCas12b 의 서열
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT
CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT
CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG
TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCG

[1945]

[1946]

ACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTA
GTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGA
AGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATA
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG
CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCA
ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG
GCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCA
CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTA
TCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAA
GCTTGCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAG
CAGCCGGATCCgccgtgaagtccatcaaagtgaagctgcggctgagcgagtgccccgatattctggctggaatgtggcagctgc
acagagccacaaatgccggcgtgcggtactacacagaatgggtgtccctgatgcggcaagagatcctgtacagcagaggccctgatggc
ggccagcagtgttatatgaccgccgaggattgccagagagagctgctgcggagactgcggaatagacagctgcataacggccggcagg
atcagcctggaacagatgctgatctgctggccatcagcagacggctgtacgagattctggtgctgcagagcatcggcaaaagaggcgacg
cccagcagattgccagcagctttctgagccctctggtggaccccaacagcaaaggtggaagaggcgaggccaagagcggaagaaaacc
tgcctggcagaagatgcgcgaccagggcgatcctagatgggttgccgctagagagaagtacgagcagcggaaggccgtggatcccagc
aaagagattctgaacagcctggacgccctgggcctcagacctctgtttgccgtgttcaccgagacatacagatccggcgtggactggaagc
ctctgggcaaatctcagggcgtcagaacctgggacagagacatgtttcagcaggccctggaacggctgatgagctgggagagctggaat
cggagagtgggcgaagagtacgccagactgttccagcagaaaatgaagttcgagcaagagcacttcgccgagcagagccacctggtca
aactggctagagccctggaagccgatatgagagccgcctctcagggcttcgaggccaaaagaggaacagcccaccagatcaccagaag
ggcactgagaggggccgacagagtgttcgagatctggaagtctatccccgaggaagccctgttcagccagtacgacgaagtgatcagac
aggtgcaggccgagaagcggagagatttcggcagccatgacctgttcgccaagctggccgagcctaagtatcagccccttggagagcc
gacgagacattcctgaccagatacgccctgtacaacggcgtgctgcgcgatctggaaaaggccagacagttcgccaccttcacactgcct
gatgcctgcgtgaaccccatctggaccagattcgagtctagccagggcagcaacctgcacaaatacgagtttctgttcgaccacctcggac
ctggcagacacgccgtcagatttcagagactgctggtggtggaaagcgagggcgccaaagaaagggatagcgtggtggtgcctgtggct
ccttctggccaactggataagctggtgctgagggaagaagagaagtccagcgtcgccctgcatctgcacgataccgctagacccgatggc
ttcatggctgaatgggctggcgccaaactgcagtacgagagaagcaccctggccagaaaagccagacgggacaagcagggcatgaga
agctggcggagacagccctccatgctgatgtctgccgctcagatgctggaagatgccaaacaggctggcgacgtgtacctgaacatcagc
gtgcgcgtgaagtctcccagtgaagtgcgaggacagaggcggcctccttacgccgctctgtttagaatcgacgacaagcagcggagagt
gaccgtgaactacaacaagctgagcgcctacctggaagaacaccccgataagcagatccctggcgctcctggactgctgtctggactgag
agtgatgtccgtggacctgggcctgagaacaagcgccagcatctccgtgttcagagtggccaagaaagaagaggtggaagccctcggag
atggccggcctcctcactactatcctatccacggcaccgatgacctggtggccgtgcacgaaagatcccacctgattcagatgcccggcga
aaccgagacaaagcagctgcggaagctgagagaagaacggcaggccgtgctgaggccactgtttgctcaactggcactgctgagactgc
tcgtcagatgtggcgccgctgacgagagaatcagaaccagatcctggcagcggctgaccaagcagggaagagagttcaccaagagact
gaccctagctggcgcgaggctctggaactggaactgacaagactcgaggcctactgcggcagagtgcccgatgatgagtggtccagaa
tcgtggacagaaccgtgattgccctgtggcggagaatgggcaagcaagtgcgcgattggcggaagcaagtgaagtccgggccaaagt
gaaagtgaagggctaccagctggatgtcgtcggcggaaattctctggcccagatcgactatctggaacagcagtacaagttcctgcggcgt
tggagcttcttcgccagagcttctggcctggtcgtgcgggccgatagagaaagccattttgccgtggctctgagacagcacatcgagaacg
ccaagcgggacagactgaagaaactggccgaccggatcctgatggaagcactgggctatgtgtacgaggccagcggacctagagaagg

ccagtggacagctcagcaccctccttgccagctgatcattctcgaggaactgtccgcctaccggttcagcgacgatagacctcctagcgag
aacagcaaactgatggcctggggccacagaggcatcctcgaagaactggtcaaccaggctcaggtgcacgatgtgctcgtgggcacagt
gtacgccgccttcagcagcagattcgacgctagaacaggtgctcccggcgtcagatgcagaagagtgcctgccagatttgtgggcgccac
cgtggatgattctctgccactgtggctgaccgagttcctggacaagcaccggctggataagaacctgctgcggcccgacgatgtgatccca
acaggcgaaggcgaattcctggtgtccccttgtggcgaagaggctgccagagttagacaggttcacgccgacatcaacgctgcccagaac
ctgcagagaaggctgtggcagaacttcgacatcaccgagctgaggctgagatgcgacgtgaagatgggcggagagggaacagtgctgg
tgcccagagtgaacaacgccagagccaagcagctgttcggcaagaaggtgctggtttcccaggacggcgtgaccttcttcgagagatctc
agacaggcggcaagccccacagcgagaagcagaccgatctgaccgacaaagaactcgagctgatcgccgaggccgatgaggccaga
gctaaaagcgtggtgctgttcagggatcctagcggccacattggcaaaggccactggatccggcagcgcgagttttggagtctggtcaagc
agaggatcgagagccacaccgccgagcggattagagttagaggcgtgggaagctccctggacGGATCCAAAAGGCCGG
CGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGggatctTACCCATACGATGT
TCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCC
GACTATGCCtaaGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAG
GGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTG
TTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCT
TTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCT
GGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG
CATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGG
CTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGG
TGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCG
CTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACT
TGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCC
TTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAAC
ACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCC
TATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGA
ATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATG
CAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCC
AGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGC
CCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCA
TGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTA
TTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCG
GGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGC
ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCT
ATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCG
GCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCT
GAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTC
CTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGG
GCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTAT
CCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCAT
TCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGT
CTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACT
GTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATG
GCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCG
ACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTG
ATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTA

[1947]

TCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTG
AGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACG
AGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCG
GGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCA
CCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAA
TTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATC
AATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCAT
GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACA
TTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTG
CATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCA
GCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA
(SEQ ID NO:583)

[1948]

[1949] **실시예 14**

[1950] PhyciCas12b (피시스파에라에 박테리움 Cas12b)는 플라스미드 (pcDNA3-PhyciCas12b)에 클로닝하였다. 플라스미드의 맵은 도 29에 도시된다. 클로닝된 구성체의 서열은 하기 표 23 에 제시되어 있다.

[1951] 표 23

pcDNA3-PhyciCas12b 의 서열
gacggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttg tgtgttggaggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttaggg ttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcatt agttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgt caataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagt acatcaagtgtatcatatgccaagtacgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatg ggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtt tgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac aactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccactg cttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcgtttaaacttaagcttgccaccATGGCCCCA AAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGCCACCAAGAGCT ACAGAGCCAGAATCCTGACCGACAGCAGACTGGCCGCTGCTCTGGATAGAACCCAC GTGGTGTTTGTGGAAAGCCTGAAGCAGATGATCAACACCTACCTGCGGATGCAGAA CGGCAAGTTCGGCCCCGACCACAAGAAACTGGCCCAGATCATGCTGAGCCGGTCCA ACACATTTGCCCACGGCGTGATGGACCAGATCACCAGAGATCAGCCCACCAGCACA CTGGACGAGGAATGGACCGACCTGGCCAGAAGAATCCACAAGACAACCGGACCTCT GTTCCTGCAAGCCGAGAGATTCGCCACCGTGAAGAACAGAGCCATCCACACCAAGT CCAGAGGCAAAGTGATCCCATCTCCTGAGACACTGGCCGTGCCTGCCAAGTTCTGGC ACCAAGTGTGCGATAGCGCCAGCGCCTACATCAGATCCAACCGCGAACTGATGCAG CAGTGGCGGAAAGATAGAGCCGCCTGGCTGAAGGACAAGAACGAGTGGCAGCAGA AACACCCCGAGTTCATGCAGTTCTACAACGGCCCCTACCAGAACTTCCTGAAGCTGT GCGACGACGACAGAATCACCTCTCAGCTGGCTGCCGAGCAGCAGCCTACAGCCAGC AAGAACAACAGACCCAGAAAGACCGGCAAGCGCTTCGCCAGATGGCACCTGTGGTA CAAGTGGCTGAGCGAGAACCCCGAGATCATCGAATGGCGGAACAAGGCCTCCGCCA GCGACTTTAAGACCGTGACCGATGACGTGCGGAAGCAGATCATTACCAAGTATCCC CAGCAGAACAAGTACATCACCCGGCTGCTGGACTGGCTGGAAGATAACAACCCCGA GCTGAAAACCCTGGAAAACCTGCGGCGGACCTACGTGAAGAAGTTCGACAGCTTCA

[1952]

AGCGGCCTCCTACACTGACCCTGCCATCTCCATACAGACACCCCTACTGGTTCACCA
TGGAACTGGACCAGTTTTACAAGAAGGCCGACTTCGAGAACGGCACCATCCAGCTG
CTGCTGATCGACGAGGACGACGACGGCAACTGGTTCTTCAACTGGATGCCCGCCTCT
CTGAAGCCCGATCCTAGACTGGTGCCTTCTTGGAGAGCCGAAACCTTCGAGACAGA
GGGCAGATTCCCTCCTTACCTCGGCGGCAAGATCGGCAAGAAGCTGAGCAGACCTG
CTCCTACCGACGCCGAGAGAAAGGCTGGAATTGCCGGCGCTAAGCTGATGATTAAG
AACAATCGGAGCGAGCTGCTGTTCACCGTGTTCGAGCAGGACTGCCCTCCTAGAGTG
AAGTGGGCCAAGACCAAGAACCGGAAGTGCCCTGCCGACAACGCCTTTAGCTCCGA
CGGCAAGACCAGAAAGCCCCTGAGAATCCTGTCCATCGACCTGGGCATCAGACACA
TCGGCGCCTTCGCTCTGACACAGGGCACCAGAAATGATAGCGCCTGGCAGACCGAG
AGCCTGAAGAAGGGCATCATCAACAGCCCTAGCATCCCTCCACTGCGGCAAGTGCG
GAGACACGACTACGACCTGAAGCGGAAAAGACGGCGGCACGGCAAGCCTGTGAAG
GGCCAGAGAAGCAACGCCAATCTGCAGGCCCACAGGACCAACATGGCCCAGGACA
GATTCAAGAAGGGCGCCTCTGCCATCGTGTCACTGGCCAGAGAGCATAGCGCCGAC
CTGATCCTGTTCGAGAACCTGCACAGCCTGAAGTTCAGCGCCTTCGACGAGCGGTGG
ATGAACAGACAGCTGCGGGACATGAACCGGCGGCACATCGTGGAACTGGTGTCTGA
ACAGGCCCCTGAGTTCGGCATCACAGTGAAGGACGACATCAACCCCTGGATGACCA
GCCGGATCTGCAGCAACTGTAACCTGCCTGGCTTCAGGTTCAGCATGAAGAAGAAG
AACCCCTACCGCGAGAAGCTGCCCAGAGAGAAGTGCACCGATTTCGGCTACCCTGT
GTGGGAACCTGGCGGCCACCTGTTTAGATGCCCTCACTGCGACCACAGAGTGAACG
CCGACATTAACGCCGCTGCCAACCTGGCCAACAAGTTCTTTGGCCTCGGCTACTGGA
ACAACGGCCTGAAGTACGATGCCGAGACAAAGACCTTCACCGTGCACACCGACAAG
AAAACCCCACCTCTGATCTTCAAGCCCAGACCTCAGTTCGATCTGTGGGCCGACAGC
GTGAAAACACGGAAGCAGCTTGGCCCCGATCCTTTCAAAAGGCCGGCGGCCACGAA
AAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGATT
ACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATG
CCTAAGaattctgcagatatccagcacagtggcggccgctcgagtctagagggcccgtttaaacccgctgatcagcctcgactgtgcc
ttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgagga
aattgcatcgcattgtctgagtaggtgtcattctattctgggggtggggtggggcaggacagcaagggggaggattgggaagacaatagc
aggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtag
cggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcc
cttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgacc
ccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaata
gtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaa
aatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggca
gaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgc
atctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggct
gactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggctttttggaggcctaggctttt
gcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgc
acgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccg
gctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatc
gtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgcc
ggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccgg
ctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctgga
cgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccat
ggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatca

[1953]

ggacatagcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccg
attcgcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacct
gccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcg
cggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaa
ataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttgg
cgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctg
gggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatga
atcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcg
gcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggcc
agcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgct
caagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccct
gccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcg
ttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggta
agacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggt
ggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatc
cggcaaacaaaccaccgctggtagcggttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttt
tctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaa
ttaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatc
tgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgat
accgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactt
tatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacag
gcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaa
aaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctc
ttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctctt
gcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctca
aggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtga
gcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattga
agcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaa
agtgccacctgacgtc (SEQ ID NO:584)

[1954]

[1955] **실시예 15**

[1956] PlancCas12b (플란크토마이세테스 박테리움 Cas12b)는 플라스미드 (pcDNA3-PlancCas12b)에 클로닝하였다. 플라스미드의 맵은 도 30에 도시된다. 클로닝된 구성체의 서열은 하기 표 24 에 제시되어 있다.

[1957] **표 24**

pcDNA3-PlancCas12b 의 서열
gacggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttg tgtgttggaggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacaattgcatgaagaatctgcttaggg ttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcatt agttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgt caataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagt acatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatg ggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtt tgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaac aactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccactg cttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcgtttaaacttaagcttgccaccATGGCCCCA AAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAGCGTGCGGAGCTT TCAGGCCAGAGTGGAATGCGACAAGCAGACCATGGAACACCTGTGGCGGACCCACA

[1958]

AGGTGTTCAACGAGAGACTGCCCGAGATCATCAAGATCCTGTTCAAGATGAAGCGG
GGCGAGTGCGGCCAGAACGATAAGCAGAAGTCCCTGTACAAGAGCATCAGCCAGA
GCATCCTGGAAGCCAACGCTCAGAACGCCGACTACCTGCTGAACAGCGTGTCCATC
AAAGGCTGGAAGCCTGGCACCGCCAAGAAGTACAGAAACGCCAGCTTCACCTGGGC
CGACGATGCCGCTAAACTGTCTAGCCAGGGCATCCACGTGTACGACAAGAAACAGG
TGCTGGGCGACCTGCCTGGCATGATGTCTCAGATGGTCTGCAGGCAGAGCGTGGAA
GCCATCTCTGGACACATCGAGCTGACCAAGAAGTGGGAGAAAGAACACAACGAGTG
GCTGAAAGAAAAAGAGAAATGGGAGTCCGAGGACGAGCACAAGAAGTATCTGGAC
CTGCGCGAGAAGTTCGAGCAGTTTGAGCAGAGCATCGGCGGCAAGATCACCAAGAG
AAGAGGCCGGTGGCACCTGTACCTGAAGTGGCTGAGCGACAACCCTGATTTTGCCG
CCTGGCGGGGAAACAAGGCCGTGATCAATCCTCTGAGCGAGAAGGCCCAGATCAGG
ATCAACAAGGCCAAGCCGAACAAGAAGAACAGCGTCGAGCGGGACGAGTTCTTCA
AGGCCAATCCTGAGATGAAGGCCCTGGACAACCTGCACGGCTACTACGAGCGGAAT
TTCGTGCGGCGGAGAAAGACAAAGAAGAACCCCGACGGCTTCGACCACAAGCCTAC
CTTCACACTGCCCCATCCTACCATCCATCCTCGTTGGTTCGTGTTCAACAAGCCTAAG
ACAAACCCCGAGGGCTACCGCAAGCTGATCCTGCCTAAAAAGGCCGGCGATCTGGG
CAGCCTGGAAATGAGACTGCTGACCGGCGAGAAGAACAAGGGCAACTACCCCGAC
GACTGGATCAGCGTGAAGTTTAAGGCCGATCCTCGGCTGAGCCTGATCAGACCCGTG
AAAGGCAGACGGGTTGTGCGGAAGGGCAAAGAGCAGGGCCAGACCAAAGAGACAG
ACAGCTACGAGTTTTTCGACAAGCACCTGAAGAAGTGGCGGCCTGCCAAACTGTCTG
GCGTGAAGCTGATCTTCCCCGACAAGACACCTAAGGCCGCCTACCTGTACTTCACCT
GTGACATCCCCGACGAGCCCCTGACCGAGACAGCCAAGAAAATCCAGTGGCTGGAA
ACCGGCGACGTGACCAAAAAGGGCAAGAAACGCAAAAAGAAGGTGCTGCCCCACG
GCCTGGTGTCCTGTGCTGTTGATCTGAGCATGCGGAGAGGCACCACCGGCTTTGCCA
CACTGTGCAGATACGAGAATGGCAAGATCCACATCCTGCGGAGCCGGAACCTGTGG
GTCGGATACAAAGAAGGCAAGGGCTGTCACCCCTACAGATGGACAGAGGGACCCG
ACCTGGGACACATTGCCAAGCACAAGAGAGAGATCAGAATCCTGCGGTCCAAGCGG
GGCAAGCCTGTGAAGGGCGAAGAGAGCCACATCGACCTGCAGAAACACATCGACTA
CATGGGCGAAGATCGGTTCAAGAAGGCCGCCAGAACCATCGTGAACTTCGCCCTGA
ACACCGAGAACGCCGCCAGCAAGAATGGCTTCTACCCCAGAGCTGACGTGCTGCTG
CTGGAAAACCTGGAAGGACTGATCCCCGATGCCGAGAAAGAGCGGGGCATCAATAG
AGCCCTGGCCGGCTGGAATAGACGGCACCTGGTTGAGCGCGTGATCGAGATGGCCA
AGGATGCCGGCTTCAAGCGGCGGGTGTTCGAGATCCCACCTTACGGCACAAGCCAA
GTGTGCAGCAAATGTGGCGCCCTGGGCAGAAGATACAGCATCATCAGAGAGAACAA
CCGGCGCGAGATCAGATTCGGCTACGTGGAAAAGCTGTTCGCCTGTCCTAACTGCGG
CTACTGCGCCAACGCCGATCACAATGCCAGCGTGAACCTGAACCGGCGGTTCCTGAT
CGAGGACAGCTTCAAGTCCTACTACGACTGGAAGCGGCTGTCCGAGAAGAAGCAGA
AAGAGGAAATCGAGACAATCGAGTCCAAGCTGATGGATAAGCTGTGCGCCATGCAC
AAGATCAGCCGGGGCAGCATCAGCAAGAAAAGGCCGGCGGCCACGAAAAAGGCCG
GCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGATTACGCTTAT
CCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGaatt
ctgcagatatccagcacagtggcggccgctcgagtctagagggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagc
catctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcatt
gtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggg
gatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagc
gcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgcca

[1959]

cgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttga
ttagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttc
caaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattta
acaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaag
catgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtca
gcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttat
ttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctccc
gggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccg
gccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcag
gggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacg
acgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctc
ctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcg
accaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcag
gggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgctt
gccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttgg
ctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcg
ccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagattt
cgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgc
tggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttca
ctgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcata
gctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagt
gagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgc
ggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcag
ctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccag
gaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtgg
cgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggata
cctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgg
gctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcg
ccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggct
acactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaacca
ccgctggtagcggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgac
gctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagtttt
aaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcat
ccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccca
cgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatcc
agtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtca
cgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagct
ccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccat
ccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaat
acgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgct
gttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaa
ggcaaaatgccgcaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggt
tattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgt
c (SEQ ID NO:585)

[1960]

[1961] **실시예 16**

[1962] BvCas12b를 함유하는 플라스미드 pZ143-pcDNA3-BvCas12b를 생성시켰다. 플라스미드의 맵은 도 31에 표시되고 클로닝된 구성체의 서열은 하기 표 25에 제시된다.

[1963] 표 25

pZ143-pcDNA3-BvCas12b 의 서열
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT
CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT
CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG
TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCG
ACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTA
GTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGA
AGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATA
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG
CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCA
ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG
GCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCA
CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA

[1964]

ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTA
TCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAA
GCTTGCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAG
CAGCCGGATCCGCCATCCGGTCCATCAAGCTGAAGATGAAGACCAACAGCGGCACC
GACAGCATCTACCTGAGAAAAGCCCTGTGGCGGACCCACCAGCTGATCAATGAGGG
AATCGCCTACTACATGAACCTGCTGACCCTGTACCGGCAAGAGGCCATCGGCGACA
AGACCAAAGAAGCCTATCAGGCCGAGCTGATTAACATCATCCGGAACCAGCAGCGG
AACAACGGCAGCTCTGAGGAACACGGCTCCGACCAAGAAATTCTGGCCCTGCTGAG
ACAGCTGTACGAGCTGATCATCCCCAGCAGCATCGGCGAATCTGGCGACGCTAATC
AGCTGGGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAACAGCCAGTCTGGCAAG
GGCACATCTAACGCCGGCAGAAAGCCCAGATGGAAGCGGCTGAAAGAGGAAGGCA
ACCCCGACTGGGAACTCGAGAAGAAGAAGGACGAGGAACGCAAGGCCAAGGATCC
CACCGTGAAGATCTTTGACAACCTGAACAAATACGGCCTGCTGCCTCTGTTCCCACT
GTTCACCAACATCCAGAAAGACATCGAGTGGCTGCCCCTGGGCAAGAGACAGTCTG
TGCGGAAGTGGGACAAAGACATGTTCATCCAGGCCATCGAGAGACTGCTGAGCTGG
GAGAGCTGGAACAGAAGAGTGGCCGACGAGTACAAACAGCTGAAAGAAAAGACCG
AGAGCTACTACAAAGAGCACCTGACAGGCGGCGAGGAATGGATCGAGAAGATCCG
GAAGTTCGAGAAAGAACGGAACATGGAACTGGAAAAGAACGCCTTCGCTCCCAACG
ACGGCTACTTCATCACCAGCAGACAGATCAGAGGCTGGGACAGAGTGTACGAGAAG
TGGTCCAAGCTGCCCGAGTCTGCTAGCCCTGAGGAACTGTGGAAAGTGGTGGCCGA
GCAGCAGAACAAGATGTCCGAAGGCTTCGGCGACCCCAAGGTGTTCAGCTTCCTGG
CCAACAGAGAGAACCGGGACATTTGGAGAGGCCACAGCGAGCGGATCTACCACATT
GCCGCCTACAACGGCCTGCAGAAGAAGCTGAGCCGGACCAAAGAGCAGGCCACCTT
CACACTGCCTGACGCCATTGAACACCCTCTGTGGATCAGATACGAGAGCCCTGGCG
GCACCAACCTGAATCTGTTCAAGCTGGAAGAGAAACAGAAAAAGAACTACTACGTG
ACCCTGAGCAAGATCATCTGGCCCAGCGAGGAAAAGTGGATTGAGAAAGAGAACAT
CGAGATCCCTCTGGCTCCCAGCATCCAGTTCAACCGGCAGATTAAGCTGAAGCAGC
ACGTGAAGGGCAAGCAAGAGATCAGCTTCAGCGACTACAGCAGCCGGATCAGCCTG
GATGGTGTTCTCGGCGGCAGCAGAATCCAGTTTAATCGGAAGTACATCAAGAACCA
CAAAGAGCTGCTCGGAGAGGGCGACATCGGCCCCGTGTTCTTTAACCTGGTGGTGG
ATGTGGCCCCTCTGCAAGAAACCAGAAACGGCAGACTGCAGAGCCCCATCGGCAAG
GCCCTGAAAGTGATCAGCAGCGACTTCTCCAAAGTGATCGACTACAAGCCGAAAGA
ACTCATGGATTGGATGAATACCGGCAGCGCCAGCAACAGCTTTGGAGTGGCTTCTCT
GCTGGAAGGCATGAGAGTGATGAGCATCGACATGGGCCAGAGAACCAGCGCCTCCG
TGTCCATCTTCGAGGTCGTGAAAGAACTGCCCAAGGATCAAGAGCAGAAGCTGTTCT
ACAGCATCAACGACACCGAGCTGTTCGCCATCCACAAGCGGAGCTTTCTGCTGAACC
TGCCTGGCGAGGTGGTCACCAAGAACAACAAGCAGCAGCGGCAAGAGCGGCGGAA
AAAGCGGCAGTTTGTGCGGAGCCAGATCAGAATGCTGGCCAACGTGCTGCGGCTGG
AAACAAAGAAAACCCCTGACGAGCGGAAGAAGGCCATTCACAAGCTGATGGAAAT
CGTGCAGAGCTACGACAGCTGGACCGCCAGCCAGAAAGAAGTGTGGGAGAAAGAG
CTGAATCTCCTGACCAACATGGCCGCCTTCAATGACGAGATCTGGAAAGAAAGCCT
GGTGGAACTGCACCACCGGATCGAGCCTTACGTGGGACAGATCGTGTCCAAGTGGC
GGAAGGGCCTGTCTGAGGGCAGAAAGAATCTGGCCGGCATCAGCATGTGGAACATC
GACGAACTGGAAGATACCAGGCGGCTGCTGATTTCCTGGTCCAAGAGAAGCAGAAC
CCCAGGCGAGGCCAACAGGATCGAAACCGATGAGCCTTTCGGCAGCAGCCTGCTCC

[1965]

AGCACATTCAGAACGTGAAGGACGACAGACTGAAGCAGATGGCCAACCTGATCATC
ATGACAGCCCTGGGCTTTAAGTACGACAAAGAGGAAAAGGACCGGTACAAGCGGTG
GAAAGAGACATACCCCGCCTGCCAGATCATCCTGTTCGAGAACCTGAACCGCTACCT
GTTCAACCTCGACCGGTCCAGACGCGAGAACAGCAGACTGATGAAGTGGGCCCATC
GGAGCATCCCCAGAACCGTGTCTATGCAGGGCGAGATGTTCGGCCTGCAAGTGGGC
GACGTTCGGAGCGAGTACAGCTCCAGATTCCACGCCAAAACAGGCGCCCCTGGCAT
CAGATGTCACGCCCTGACTGAAGAGGATCTGAAGGCCGGCAGCAACACCCTGAAGA
GACTGATCGAGGACGGCTTCATCAATGAGAGCGAGCTGGCCTACCTGAAGAAGGGC
GATATCATCCCTAGCCAAGGCGGCGAACTGTTCGTGACACTGTCCAAGCGGTACAA
GAAGGACAGCGACAACAACGAGCTGACCGTGATCCACGCCGACATCAACGCCGCTC
AGAATCTGCAGAAGCGGTTTTGGCAGCAAAACAGCGAGGTGTACAGAGTGCCCTGT
CAGCTGGCCAGAATGGGCGAAGATAAGCTGTACATCCCCAAGAGCCAGACCGAGAC
AATCAAGAAGTATTTCGGCAAGGGCTCCTTCGTGAAGAACAATACCGAACAAGAGG
TCTACAAGTGGGAGAAGTCCGAGAAAATGAAGATCAAGACGGACACCACCTTCGAC
CTGCAAGACCTGGATGGCTTCGAGGACATCAGCAAGACCATTGAGCTGGCACAAGA
GCAGCAAAAGAAATACCTGACCATGTTCAGGGACCCCAGCGGCTACTTTTTCAACA
ATGAGACATGGCGGCCTCAAAAAGAATACTGGTCCATCGTGAACAACATCATCAAG
AGCTGCCTCAAGAAGAAGATCCTGAGCAACAAGGTCGAGCTGGGATCCAAAAGGCC
GGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGggatctTACCCATACGAT
GTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCC
CCGACTATGCCtaaGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAG
AGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATC
TGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTC
CTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTC
TGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG
GCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGG
GCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTG
GTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC
GCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATC
GGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAAC
TTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC
CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAA
CACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGC
CTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGG
AATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTAT
GCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCC
CAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCG
CCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCC
ATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCT
ATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCC
GGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCG
CATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGC
TATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCC
GGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCC
TGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTT
CCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTG

[1966]

GGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTA
TCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCA
TTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGG
TCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAC
TGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCAT
GGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATC
GACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGT
GATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGT
ATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCT
GAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAC
GAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCC
GGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCC
ACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAA
ATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCAT
CAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCA
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACA
TTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTG
CATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCA
GCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA
(SEQ ID NO:586)

[1967]

[1968] BvCas12b sgRNA 스캐폴드를 함유하는 플라스미드 pZ147-BvCas12b-sgRNA-scaffold를 생성시켰다. 플라스미

드의 맵은 도 32에 표시되고 클로닝된 구성체의 서열은 하기 표 26에 제시된다.

[1969] **표 26**

pZ147-BvCas12b-sgRNA-scaffold 의 서열
accttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaa gcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgc gccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatat catggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtg atattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatc gccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattcca ccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttct tcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattct agttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcc tgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaact cacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagag gcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaa aggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaa aaaggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccc gacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgc ctttctcccttcgggaagcgtggcgctttctatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgc acgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggca gcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaa gaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggta

[1970]

gcggttttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtgg aacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatct aaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgc ctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccg gctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaa ttgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgt ttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcct ccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatg cttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataat accgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatcc agttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatg ccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcat gagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtcgacggatc gggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttgga ggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacTTAATTAAgagggcctatttcccatgatt ccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgt agaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttgg ctttatatatcttgtggaaaggacgaaacaccgGACCTATAGGGTCAATGAATCTGTGCGTGTGCCATAA GTAATTAAAAATTACCCACCACAGGAGCACCTGAAAACAGGTGCTTGGCACggagacga tatatcgtctctttttttcaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagatatacgcgttgacatt gattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggccc gcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtc aatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaa atggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgat gcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgtttt ggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtct atataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagctggct agcgtttaaacttaAGCTTGCCGGTgccaccatgGTGAGCAAGGGCGAGGAGGATAACATGGCCA TCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCAC GAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCG CCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCC CTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACT ACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGG ACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATC TACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAA GAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCC TGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGC TGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACA ACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAA CAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAA GtacccatacgatgttccagattacgctTAAGaattctgcagatatccagcacagtggcggccgctcgagtctagagggcccgtttaaa cccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactccc actgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaag ggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccagctggggctctag ggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccc tagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttcc

[1971]

gatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccc
tttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggat
tttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaa
agtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccag
caggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccag
ttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgag
gaggctttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgttt
cgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcg
gctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttgtcaagaccgacctgtccggtgccctgaatgaactg
caggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggga
ctggctgctattgggcgaagtgccggggcaggatctcctgtcatctc (SEQ ID NO:587)

[1972]

[1973] **실시예 17**

[1974] BhCas12b sgRNA 스캐폴드를 함유하는 플라스미드 pZ148-BhCas12b-sgRNA-scaffold를 생성시켰다. 플라스미드의 맵은 도 33에 표시되고 클로닝된 구성체의 서열은 하기 표 27에 제시된다.

[1975] **표 27**

pZ148-BhCas12b-sgRNA-scaffold 의 서열
accttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaa gcgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgc gccagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatat catggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtg atattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatc gccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattcca ccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttct tcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattct agttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcc tgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaact cacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagag gcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaa aggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaa aaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccc gacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgc ctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgc acgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggca gcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaa gaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggta gcggttttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtgg aacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatct aaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgc ctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccg gctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaa ttgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgt ttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcct ccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatg cttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataat accgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatcc

[1976]

agttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatg
ccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttttcaatattattgaagcatttatcagggttattgtctcat
gagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtcgacggatc
gggagatctcccgatcccctatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttgga
ggtcgctgagtagtgcgcgagcaaaatttaagctacaacaaggcaaggcttgaccgacTTAATTAAgagggcctatttcccatgatt
ccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgt
agaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttgg
ctttatatatcttgtggaaaggacgaaacaccgGTTCTGTCTTTTGGTCAGGACAACCGTCTAGCTATAA
GTGCTGCAGGGTGTGAGAAACTCCTATTGCTGGACGATGTCTCTTACGAGGCATTAG
CACggagacgatatatcgtctctttttttcaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagat
atacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataactt
acggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggga
ctttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgac
gtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgc
tattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtc
aatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgta
cggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggag
acccaagctggctagcgtttaaacttaAGCTTGCCGGTgccaccatgGTGAGCAAGGGCGAGGAGGATAA
CATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGA
ACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCAC
CCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACA
TCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACA
TCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGA
ACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGC
GAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTA
ATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGA
CGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCAC
TACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGG
CGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCA
TCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAG
CTGTACAAGtacccatacgatgttccagattacgctTAAGaattctgcagatatccagcacagtggcggccgctcgagtctaga
gggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctgga
aggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtgggg
caggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaacca
gctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctaca
cttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggct
ccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatag
acggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttt
gatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtca
gttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtc
cccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgccccttaactccgcccatcccgccc
ctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattc
cagaagtagtgaggaggctttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacag
gatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggca
caacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgc
cctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactga
agcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctc (SEQ ID NO:588)

[1977]

[1978] S893, K846, 및 E836에 돌연변이를 갖는 BhCas12b를 함유하는 플라스미드 pZ149-BhCas12b-S893R-K846R-E836G를 생성시켰다. 플라스미드의 맵은 도 34에 표시되고 클로닝된 구성체의 서열은 하기 표 28에 제시된다.

[1979] 표 28

pcDNA3-PlancCas12b 의 서열
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT
CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT
CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG
TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCG
ACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTA
GTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGA
AGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATA
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG

[1980]

[1981]

TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG
CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCA
ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG
GCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCA
CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTA
TCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAA
GCTTGCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAG
CAGCCGGATCCGCCACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTG
AAGAAAGGCCTCTGGAAAACCCACGAGGTGCTGAACCACGGAATCGCCTACTACAT
GAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACC
CCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTC
GTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGT
GTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGA
AGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAAC
AGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACC
TGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGA
TAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGA
TCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGT
GGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATT
CAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGA
ATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGAC
ATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCT
GCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCT
GGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAA
GTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATT
ACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCAC
CCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGA
CGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGT
CCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGC
AGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATC
TACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCT
GCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGC
ACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGC
GGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGA
AAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAG
AGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCA
ACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACT
GAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGG
GACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGAC
ATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGA
GCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCT
GCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGC
GGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTC

[1982]

ACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGA
GCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCT
GAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGG
CGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACAT
CGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCG
AACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTG
AATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCAT
CATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGA
ACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACGGAG
AAAGGTCCCGCTTCGAGAACAGCCGGCTCATGAAGTGGTCCAGACGCGAGATCCCC
AGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGC
TCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTCGGG
TCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAG
GGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGA
CAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACAC
ACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCAC
GGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACAT
CCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACT
TCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAG
AAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAG
ACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGAC
CCCAGCGGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGA
AAGCTGGAACGCATCCTGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCAT
CGAGGACGACAGCAGCAAGCAGTCTATGGGATCCAAAAGGCCGGCGGCCACGAAA
AAGGCCGGCCAGGCAAAAAAGAAAAAGggatctTACCCATACGATGTTCCAGATTACG
CTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCtaa
GAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAA
ACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCT
CCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGT
GGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAT
GCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTA
TCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTC
CTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTA
GGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGAT
GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAG
TCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCT
CGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAA
TGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTA
GGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCT
CAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTA
TGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTT
TTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTG
AGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATC

CATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGA
TGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTG
GGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGG
GGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGG
ACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTG
CTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGG
GCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGA
TGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGC
GAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGG
ATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTC
AAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTT
GCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCT
GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAG
AGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCG
ATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTG
GGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTC
CACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTG
GATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTT
TATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAT
CATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTG
TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGC
ATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATC
GGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTC
ACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAG
GCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA (SEQ ID NO:589)

[1983]

[1984] S893, K846, 및 E836에 돌연변이를 갖는 BhCas12b를 함유하는 플라스미드 pZ150- pCDNA3-BhCas12b-S893R-K846R-E836K를 생성시켰다. 플라스미드의 맵은 도 35에 표시되고 클로닝된 구성체의 서열은 하기 표 29에 제시된다.

[1985] **표 29**

pZ150-pCDNA3-BhCas12b-S893R-K846R-E836K 의 서열
AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG
TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT
CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA

[1986]

GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA
TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT
GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT
CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT
CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG
TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA
AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCG
ACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTA
GTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGA
AGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATA
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTG
GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG
CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCA
ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTC
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG
GCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCA
CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTA
TCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAA
GCTTGCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAG
CAGCCGGATCCGCCACCAGATCCTTCATCCTGAAGATCGAGCCCAACGAGGAAGTG
AAGAAAGGCCTCTGGAAAACCCACGAGGTGCTGAACCACGGAATCGCCTACTACAT
GAATATCCTGAAGCTGATCCGGCAAGAGGCCATCTACGAGCACCACGAGCAGGACC
CCAAGAATCCCAAGAAGGTGTCCAAGGCCGAGATCCAGGCCGAGCTGTGGGATTTC
GTGCTGAAGATGCAGAAGTGCAACAGCTTCACACACGAGGTGGACAAGGACGAGGT
GTTCAACATCCTGAGAGAGCTGTACGAGGAACTGGTGCCCAGCAGCGTGGAAAAGA
AGGGCGAAGCCAACCAGCTGAGCAACAAGTTTCTGTACCCTCTGGTGGACCCCAAC
AGCCAGTCTGGAAAGGGAACAGCCAGCAGCGGCAGAAAGCCCAGATGGTACAACC
TGAAGATTGCCGGCGATCCCTCCTGGGAAGAAGAGAAGAAGAAGTGGGAAGAAGA
TAAGAAAAAGGACCCGCTGGCCAAGATCCTGGGCAAGCTGGCTGAGTACGGACTGA
TCCCTCTGTTCATCCCCTACACCGACAGCAACGAGCCCATCGTGAAAGAAATCAAGT

[1987]

GGATGGAAAAGTCCCGGAACCAGAGCGTGCGGCGGCTGGATAAGGACATGTTCATT
CAGGCCCTGGAACGGTTCCTGAGCTGGGAGAGCTGGAACCTGAAAGTGAAAGAGGA
ATACGAGAAGGTCGAGAAAGAGTACAAGACCCTGGAAGAGAGGATCAAAGAGGAC
ATCCAGGCTCTGAAGGCTCTGGAACAGTATGAGAAAGAGCGGCAAGAACAGCTGCT
GCGGGACACCCTGAACACCAACGAGTACCGGCTGAGCAAGAGAGGCCTTAGAGGCT
GGCGGGAAATCATCCAGAAATGGCTGAAAATGGACGAGAACGAGCCCTCCGAGAA
GTACCTGGAAGTGTTCAAGGACTACCAGCGGAAGCACCCTAGAGAGGCCGGCGATT
ACAGCGTGTACGAGTTCCTGTCCAAGAAAGAGAACCACTTCATCTGGCGGAATCAC
CCTGAGTACCCCTACCTGTACGCCACCTTCTGCGAGATCGACAAGAAAAAGAAGGA
CGCCAAGCAGCAGGCCACCTTCACACTGGCCGATCCTATCAATCACCCTCTGTGGGT
CCGATTCGAGGAAAGAAGCGGCAGCAACCTGAACAAGTACAGAATCCTGACCGAGC
AGCTGCACACCGAGAAGCTGAAGAAAAAGCTGACAGTGCAGCTGGACCGGCTGATC
TACCCTACAGAATCTGGCGGCTGGGAAGAGAAGGGCAAAGTGGACATTGTGCTGCT
GCCCAGCCGGCAGTTCTACAACCAGATCTTCCTGGACATCGAGGAAAAGGGCAAGC
ACGCCTTCACCTACAAGGATGAGAGCATCAAGTTCCCTCTGAAGGGCACACTCGGC
GGAGCCAGAGTGCAGTTCGACAGAGATCACCTGAGAAGATACCCTCACAAGGTGGA
AAGCGGCAACGTGGGCAGAATCTACTTCAACATGACCGTGAACATCGAGCCTACAG
AGTCCCCAGTGTCCAAGTCTCTGAAGATCCACCGGGACGACTTCCCCAAGGTGGTCA
ACTTCAAGCCCAAAGAACTGACCGAGTGGATCAAGGACAGCAAGGGCAAGAAACT
GAAGTCCGGCATCGAGTCCCTGGAAATCGGCCTGAGAGTGATGAGCATCGACCTGG
GACAGAGACAGGCCGCTGCCGCCTCTATTTTCGAGGTGGTGGATCAGAAGCCCGAC
ATCGAAGGCAAGCTGTTTTTCCCAATCAAGGGCACCGAGCTGTATGCCGTGCACAGA
GCCAGCTTCAACATCAAGCTGCCCGGCGAGACACTGGTCAAGAGCAGAGAAGTGCT
GCGGAAGGCCAGAGAGGACAATCTGAAACTGATGAACCAGAAGCTCAACTTCCTGC
GGAACGTGCTGCACTTCCAGCAGTTCGAGGACATCACCGAGAGAGAGAAGCGGGTC
ACCAAGTGGATCAGCAGACAAGAGAACAGCGACGTGCCCCTGGTGTACCAGGATGA
GCTGATCCAGATCCGCGAGCTGATGTACAAGCCTTACAAGGACTGGGTCGCCTTCCT
GAAGCAGCTCCACAAGAGACTGGAAGTCGAGATCGGCAAAGAAGTGAAGCACTGG
CGGAAGTCCCTGAGCGACGGAAGAAAGGGCCTGTACGGCATCTCCCTGAAGAACAT
CGACGAGATCGATCGGACCCGGAAGTTCCTGCTGAGATGGTCCCTGAGGCCTACCG
AACCTGGCGAAGTGCGTAGACTGGAACCCGGCCAGAGATTCGCCATCGACCAGCTG
AATCACCTGAACGCCCTGAAAGAAGATCGGCTGAAGAAGATGGCCAACACCATCAT
CATGCACGCCCTGGGCTACTGCTACGACGTGCGGAAGAAGAAATGGCAGGCTAAGA
ACCCCGCCTGCCAGATCATCCTGTTCGAGGATCTGAGCAACTACAACCCCTACAAGG
AAAGGTCCCGCTTCGAGAACAGCCGGCTCATGAAGTGGTCCAGACGCGAGATCCCC
AGACAGGTTGCACTGCAGGGCGAGATCTATGGCCTGCAAGTGGGAGAAGTGGGCGC
TCAGTTCAGCAGCAGATTCCACGCCAAGACAGGCAGCCCTGGCATCAGATGTCGGG
TCGTGACCAAAGAGAAGCTGCAGGACAATCGGTTCTTCAAGAATCTGCAGAGAGAG
GGCAGACTGACCCTGGACAAAATCGCCGTGCTGAAAGAGGGCGATCTGTACCCAGA
CAAAGGCGGCGAGAAGTTCATCAGCCTGAGCAAGGATCGGAAGTGCGTGACCACAC
ACGCCGACATCAACGCCGCTCAGAACCTGCAGAAGCGGTTCTGGACAAGAACCCAC
GGCTTCTACAAGGTGTACTGCAAGGCCTACCAGGTGGACGGCCAGACCGTGTACAT
CCCTGAGAGCAAGGACCAGAAGCAGAAGATCATCGAAGAGTTCGGCGAGGGCTACT
TCATTCTGAAGGACGGGGTGTACGAATGGGTCAACGCCGGCAAGCTGAAAATCAAG
AAGGGCAGCTCCAAGCAGAGCAGCAGCGAGCTGGTGGATAGCGACATCCTGAAAG
ACAGCTTCGACCTGGCCTCCGAGCTGAAAGGCGAAAAGCTGATGCTGTACAGGGAC

[1988]

CCCAGCGGCAATGTGTTCCCCAGCGACAAATGGATGGCCGCTGGCGTGTTCTTCGGA
AAGCTGGAACGCATCCTGATCAGCAAGCTGACCAACCAGTACTCCATCAGCACCAT
CGAGGACGACAGCAGCAAGCAGTCTATGGGATCCAAAAGGCCGGCGGCCACGAAA
AAGGCCGGCCAGGCAAAAAAGAAAAAGggatctTACCCATACGATGTTCCAGATTACG
CTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCtaa
GAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAA
ACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCT
CCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGT
GGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAT
GCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTA
TCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTC
CTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTA
GGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGAT
GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAG
TCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCT
CGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAA
TGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTA
GGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCT
CAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTA
TGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTT
TTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTG
AGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATC
CATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGA
TGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTG
GGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGG
GGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGG
ACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTG
CTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGG
GCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGA
TGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGC
GAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGG
ATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTC
AAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTT
GCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCT
GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAG
AGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCG
ATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTG
GGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTC
CACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTG
GATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTT
TATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAT
CATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTG

[1989]

TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGC
ATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATC
GGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTC
ACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAG
GCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA (SEQ ID NO:590)

[1990]

[1991] **실시예 18**

[1992] BhCas12b에 대한 이. 콜라이 PAM은 다양한 조건 하에서 시험관내 PAM 스크린으로 결정하였다. 결과를 도 36에 제시된다.

[1993] **실시예 19**

[1994] BvCas12b에 대한 이. 콜라이 PAM은 다양한 조건 하에서 시험관내 PAM 스크린으로 결정하였다. 결과를 도 37에 제시된다.

[1995] **실시예 20**

[1996] BhCas12b의 변이체를 생성시켰다. 돌연변이는 아래의 표 30 에 도시된다.

[1997] 표 30

BhCas12b 변이체	변이체 내 돌연변이
BhCas12b 변이체 1	S893R
BhCas12b 변이체 2	S893R/K846R
BhCas12b 변이체 3	K846R/E837G
BhCas12b 변이체 4	S893R/K846R/E837G

[1998]

[1999] 변이체의 활성은 상이한 결합 부위에서 indel 백분율을 시험하여 평가하였다. 시험의 결과가 표 38 에 제시되어 있다.

[2000] 실시예 21

[2001] BhCas12b의 추가 변이체를 생성시켰고 그들 활성을 시험하였으며 실시예 20에서 생성된 변이체와 비교하였다.

[2002] 추가 변이체는 돌연변이 S893R 및 K846R을 함유하였고, 돌연변이 E837H, E837K, E837N, E837L, E837I, D533G, N644K, D680P, L741Q, L792Q, F881L, V895A, V980E, T984A, K1022E, 또는 M1073I를 더 함유하였다. 변이체의 활성은 상이한 결합 부위에서 indel 백분율을 시험하여 평가하였다. 시험의 결과가 표 39 에 제시되어 있다.

[2003] 실시예 22

[2004] BhCas12b (실시예 20의 변이체 4) 및 야생형 BvCas12b에 의한 절단에 의한 HDR은 상이한 부위에서 시험하였다. DNMT1-1에서 HDR의 결과는 도 40A에 도시되고 VEGFA-2에서의 HDR은 도 40B에 도시된다.

[2005] 실시예 23

[2006] 이 실시예는 ssODN 도너로 수행된 실험을 비롯하여 Cas12b 오솔로그의 활성 접근 상이한 PAM을 시험하기 위해 293T 세포에서 수행된 실험을 도시한다.

[2007] 도 41A 는 TTTV PAM에서 AsCas12a 및 BhCas12b 변이체 4 및 BvCas12b ATTN PAMS의 indel 백분율의 비교를 도시한다. 도 41B 는 상이한 PAM 서열에서 BhCas12b 변이체 4 및 BvCas12b 활성의 붕괴를 도시한다.

[2008] 도 42A 는 ssDNA 도너와 함께 도입시키려는 바람직한 변화를 포함한 VEGFA 표적의 개략도를 도시한다. 도 42B 는 VEGFA 표적 부위에서 각 뉴클레아제의 indel 활성을 도시한다. 도 42C 는 VEGFA 부위에서 바람직한 편집 (2개 뉴클레오티드 치환)을 함유하는 세포의 백분율을 도시한다. 도 42D 는 ssDNA 도너와 함께 도입시키려는 바람직한 변화를 포함한 DNMT1 표적의 개략도를 도시한다. 도 42E 는 DNMT1 표적 부위에서 각 뉴클레아제의 indel 활성을 도시한다. 도 42E 는 DNMT1 부위에서 바람직한 편집 (2개 뉴클레오티드 치환)을 함유하는 세포의 백분율을 도시한다. 도 42C 및 42E의 경우, 완벽한 편집은 파란색으로 도시되었고 붉은색 막대는 바람직한 2개 뉴클레오티드 치환, 및 추가 돌연변이없는 양쪽 PAM의 돌연변이를 포함하는 개략도에서 표시된 바와 같이 완벽하게 교정된 유전자좌를 함유하는 세포의 백분율을 표시한다.

[2009] 실시예 24

[2010] CXCR4 유전자를 표적화하는 sgRNA와 BhCas12b (v4) 및 BvCas12b 리보핵단백질 (RNP) 복합체를 시험관내에서 조립하였고 Lonza 4D-Nucleofector를 사용해 인간 CD4+ T 세포를 전기천공시켰다. 인간 CD4+ T 세포는 2명의 다른 도너로부터 획득하였다. RNP는 3 μM 최종 농도로 $3x10^5$ 세포에 전달되었다. 전기천공된 세포는 48시간 후에 회수하였고 indel 돌연변이는 표적화 심층 시퀀싱으로 판독하였다. 도 43의 좌측 패널은 CXCR4의 표적화된 엑손 및 각각 BhCas12b (v4) 및 BvCas12b에 의해 표적화된 CXCR4 서열을 도시한다. 도 43의 우측 패널은 2명 도너로부터의 T 세포에서 CXCR4에 대한 BhCas12b (v4) 및 BvCas12b의 효과를 도시한 indel 백분율을 도시한다.

[2011] 실시예 25 - CRISPR-Cas12b를 사용한 게놈 편집

[2012] V형 CRISPR 이펙터 Cas12b (C2c1로도 알려짐)는 적어도 부분적으로 특징규명된 패밀리 멤버의 고온 요구성으로 인해, 인간 세포에서 게놈 편집을 위해 개발하는 것이 도전이었다. 여기서 출원인은 다양한 Cas12b 패밀리

를 조사하였고 바실러스 히사시이로부터 인간 유전자 편집을 위한 예시적인 유망한 후보물 BhCas12b를 확인하였다. 37℃에서, 야생형 BhCas12b는 이중 가닥 파손을 형성하는 대신에 비표적 DNA 가닥을 우선적으로 닉킹하여서, 낮은 편집 효율을 초래한다. 접근법의 조합들을 사용하여, 출원인은 이러한 한계를 극복한 BhCas12b에 대한 기능 획득 돌연변이를 동정하였다. 돌연변이체 BhCas12b는 인간 세포주 및 생체외 초대 인간 T 세포에서 강력한 게놈 편집을 촉진하였고, 에스. 피오게네스 Cas9와 비교해 더 큰 특이성을 나타냈다. 이러한 작업은 인간 세포에서 게놈 편집을 위해, Cas9 및 Cpf1/Cas12a 이외에도, 제3의 RNA-가이드된 뉴클레아제 플랫폼을 확립한다.

[2013] 여기서 출원인은 중온성 Cas12b 효소를 검색하였고 37℃에서 비표적 DNA 가닥을 우선적으로 닉킹시키는, 바실러스 히사시이로부터의 유망한 후보물, BhCas12b를 동정하였다. 접근법의 조합을 사용하여, 출원인은 이러한 한계를 극복하고 37℃에서 양쪽 DNA 가닥을 절단하는 BhCas12b 변이체를 조작하였다. 출원인은 또한, 조작된 BhCas12b 변이체를 자연적으로 모방한, 바이킹 우주선이 조립된2 청정실에서 단리된 샘플로부터 시퀀싱된 제2 바실러스 sp. 오솔로그, BvCas12b를 동정하였다. 양쪽 특징규명된 Cas12b 뉴클레아제는 인간 세포에서 효율적인 게놈-편집을 촉진하였고 Cas9와 비교하여 더 높은 특이성을 나타냈다. 따라서, BhCas12b 및 BvCas12b의 특징규명 및 조작은 인간 세포에서 고도로 특이적인 게놈 편집을 위한 새로운 도구를 제공하여, CRISPR-Cas 시스템의 이러한 신규 클래스의 잠재성을 열어준다.

[2014] 게놈 편집 도구는 재프로그램가능하고 고도로 특이적인 것이 바람직할 수 있고, 원핵생물 CRISPR 및 CRISPR-연관 단백질 (CRISPR-Cas) 시스템은 자연적으로 이들 성질을 부여한다[3,4]. 현행 게놈 편집 기술은 게놈 절단을 위한 단일-단백질 이펙터 뉴클레아제를 함유하는, 클래스 2 CRISPR-Cas 시스템에 집중하고 있지만, 클래스 2 뉴클레아제의 오직 하기 2개 패밀리가 지금까지 인간 세포에서 게놈 편집을 위해 이용되었다: tracrRNA7 와 기능할 수 있고 HNH 및 RuvC 뉴클레아제 도메인[8,9]을 함유하는, Cas9[5,6] 및 짧은 crRNA를 사용하고 단일 RuvC 도메인을 함유하는 Cas12a[10]. 여기서 출원인은 단일 RuvC 도메인을 함유하고 tracrRNA11 를 요구하는 클래스 2 엔도뉴클레아제의 제3 패밀리, Cas12b에 집중하였다 (도 44a). Cas12b 단백질이 종종 Cas9 및 Cas12a에 비해 작고 게놈 편집에 잠재적으로 유망하게 보이지만, 알리시클로바실러스 악시도테레스트리스 (Alicyclobacillus acidoterrestris) 유래의 최고로 특징규명된 Cas12b 뉴클레아제 (AacCas12b)는 48℃[1]에서 최적 DNA 절단 활성을 나타낸다. 충분히 특징규명된 패밀리[10,12] 내에서 Cas 이펙터의 다양한 성질을 고려하면, 출원인은 저온에서 활성이 있는 Cas12b 패밀리 멤버을 동정하고자 하였고 따라서 인간 게놈 편집을 위해 적합화시킬 수 있었다.

[2015] 질문으로서 이전에 검출된 Cas12b 단백질을 사용한 최신 서열 데이터베이스의 BLAST 검색은 V-B형 유전자좌 내에서 코딩되는 약 25개 멤버의 Cas12b 패밀리를 동정하였다. V-B형 시스템은 박테리아 간에 광범위하게 산재되어 있고, Cas12b의 계통수 형태 (도 48a)는 일반적인 박테리아 분류법을 따르지 않아서, 광대한 수평 이동을 시사한다. 그러나, 특히, 계통수에서 강력하게 뒷받침되는 분기군을 형성하는 V-B형 유전자좌의 대략 절반은 바실라레스 (Bacillales) 박테리아 목의 멤버에서 확인된다. 출원인은 실험 연구를 위해 다양한 박테리아로부터 14개의 비규명된 Cas12b 유전자를 선택하여, 고온균으로 인식된 것들과 이전에 기술된 멤버은 피하였다. 모든 공지된 클래스 2 DNA-표적화 CRISPR-Cas 뉴클레아제는 DNA 절단을 위해 프로토스페이서-인접 모티프 (PAM)[8,10] 를 필요로 하고, Cas12b 패밀리의 초기 특징규명은 표적 부위[1]의 5' 면 상의 PAM을 밝혀주었다.

확인된 유전자좌가 기능성 CRISPR-Cas 시스템이라는 것을 확증하고 그들의 PAM을 확인하기 위해서, 14개 후보 각각에 대해서, 출원인은 이. 콜라이에서 그들의 천연 측접 서열과 인간 코된 최적화된 Cas12b를 발현시켰고 무작위화된 5' PAM 라이브러리로 형질전환된 세포를 챌린지시킨 후에 심층 시퀀싱을 수행하였다 (도 48b 및 48c). 출원인은 14개의 시험된 Cas12b 시스템 (AkCas12b, BhCas12b, EbCas12b, 및 LsCas12b) 중 4개에서 결실을 검출하여 이종성 숙주에서 기능성 DNA 간섭을 시사하였다. 고갈된 PAM은 스페이서 상류의 1-4 bp 위치에서 T-풍부하여, 이전에 연구된 Cas12b 멤버[11]에서 관찰된 선호도와 일관되었다. 출원인은 필요한 RNA 성분을 확인하기 위해 이. 콜라이 용해물 상에서 소형 RNA-Seq을 수행하였고 Cas12b 및 CRISPR 어레이 사이 영역에 대해 맵핑된 추정 tracrRNA를 확인하였다 (도 49a-49d).

[2016] Cas12b를 생화학적으로 특징규명하기 위해서, 출원인은 확인된 PAM을 함유하는 표적 상에서, 정제된 Cas12b 단백질 및 예상 tracrRNA 및 crRNA로 시험관내 활성에 대해 시험하였다 (도 44b, 도 49e). 출원인은 EbCas12b 및 LsCas12b로 오직 최소의 활성을 관찰하였지만, AkCas12b 및 BhCas12b 둘 모두는 37℃에서 강력한 절단을 나타내서, 인간 세포에서 추가 연구를 타당하게 한다. 세포에서 게놈 편집이 단일 가이드 RNA

(sgRNA)[13]에 의해 더 효율적이라는 것을 고려하여, 출원인은 AkCas12b 및 BhCas12b에 대한 sgRNA를 설계하였고 시험관내에서 그들 활성을 검증하였다 (도 49f). 출원인은 U6 프로모터에 의해 구동되는 NLS-태그화 Cas12b 및 sgRNA를 발현하는 플라스미드로 293T 세포를 형질감염시켰고 표적화 심층 시퀀싱을 통해 삽입 또는 결실 (indel) 돌연변이의 형성을 통해 뉴클레아제 활성을 모니터링하였다. 양쪽 Cas1b 단백질에 대해 관찰된 indel 비율은 검출가능하였지만 1% 이하였다 (도 44c 및 44d). 효율을 증가시키기 위해서, 출원인은 tracrRNA 및 crRNA를 변경시키고, 헤어핀 미스매치를 제거하고, 5' 출발 부위 및 스페이서 길이를 변형시켜서 sgRNA 스캐폴드 내 변화 효과를 시험하였다 (도 44c-44e, 도 50). AkCas12b sgRNA 내 변경이 적은 효과를 갖지만, BhCas12b sgRNA의 5-nt 5' 절두는 다수 표적에 걸쳐서 실질적으로 활성을 개선시켰다.

[2017] 출원인은 빈번하게 시험관내 절단 반응의 겔 전기영동 동안, 가장 특히 AkCas12b에서, 더 느린 이동을 관찰하였으며, Cas12b가 이중 가닥 DNA (dsDNA) 기질을 닉킹할 수 있다는 것을 의미한다 (도 44b). 차등적으로 표지된 DNA 가닥과의 반응은 AkCas12b 및 BhCas12b가 비표적 가닥을 우선적으로 절단하고, 이러한 거동은 저온에서 보다 현저하다는 것을 밝혔다 (도 45a). 표적-가닥을 절단하는 불능성은 게놈 편집 도구로서 BhCas12b의 가능성을 감소시키므로, 출원인은 단백질 조작을 통해 이러한 한계를 해결하고자 하였다.

[2018] 표적-가닥은 BhCas12b의 RuvC 활성 부위에 충분치 않게 접근가능할 수 있다. 출원인은 BhCas12b 내 이러한 포켓의 성질이 표적-가닥 접근성 및 DNA 절단을 개선시킬 수 있는지 여부를 시험하였다. 출원인은 AacCas12b와의 정렬을 통해서 확인된 12 BhCas12b 잔기를 돌연변이시켰고, 이 잔기는 또한 바실러스 써모아밀로보란스 (Bacillus thermoamylovorans) 유래의 거의 동일한 Cas12b (BthCas12b)의 구조에서 관찰되었다 (BthCas12b는 또한 세포에서 활성을 나타냈지만, BhCas12b 만큼 효율적이지는 않음 도 51a)[15]. 출원인은 총 268 BhCas12b 단일 돌연변이체로 2개 표적 부위에서 indel 활성을 측정하였고 K846R 및 S893R을 포함한 몇몇 돌연변이에서 증가된 활성을 확인하였으며, 이중 돌연변이체로서 가산적 효과를 나타냈다 (도 45b 및 45c, 도 51b 및 51c). 양으로 하전된 아르기닌 측쇄는 종종 핵산의 골격과 상호작용하므로[16], 돌연변이체의 증가된 DNA-결합 친화성은 표적-가닥을 RuvC 활성 부위를 향해 당겨서 DNA 절단을 촉진하는 것을 돕는게 가능하다.

[2019] 직교적 접근법으로서, 출원인은 표적-가닥 절단의 온도-의존성을 해결하고자 하였다. 출원인은 66 표면 노출 잔기에서 글리신 치환을 생성시키고 다시 2개 표적 부위에서 indel 활성을 시험하였다. 현저하게, 출원인은 E837G 변이체를 갖는 야생형에 비해서 2배 이상 개선을 관찰하였고, 이 위치는 가이드 RNA:DNA 듀플렉스 및 RuvC 활성 부위 사이에 위치된다 (도 45d 및 45e). 돌연변이의 조합 시험은 점진적으로 활성 변이체를 야기시켰는데, K846R/S893R/E837G를 함유하는 최종 BhCas12b v4 변이체가 다수 표적에 걸쳐 최고 활성을 나타냈다 (도 45f 및 45g). 인간 세포에서 이들 결과와 일관되게, 정제된 BhCas12b v4 단백질은 37℃에서 증가된 dsDNA 절단 활성 및 닉킹된 dsDNA의 분명한 감소를 나타냈다 (도 45h, 도 51g-51j).

[2020] 출원인의 Cas12b 효소의 초기 선택은 스크리닝된 변이체의 다양성을 증가시키기 위해 동일 종으로부터의 오솔로그는 피하였다. 그러나, BhCas12b에 의한 긍정적인 게놈 편집 결과를 고려하여, 출원인은 바실러스 sp.멤버을 재논의하였고, 바이킹 우주선이 조립된 청정실에서 단리된, 바실러스 sp. V3-13 로부터 최근에 기탁된 게놈에서 코딩되는 Cas12b 오솔로그 (BhCas12b와 41% 서열 동일성)를 찾았다[2]. 출원인은 이 단백질을 특징규명하였고, 본 명세서에서 BvCas12b라고 하며, BvCas12b가 37℃에서 ATTN PAM을 갖는 표적 DNA를 효율적으로 절단한다는 것을 발견하였다 (도 52). 흥미롭게도, BhCas12b v4 돌연변이 K846R 및 S893R은 각각 BvCas12b의 R849 및 H896에 상응하고 (도 53a), BvCas12b가 자연적으로 최적 dsDNA 절단 활성을 진화시켰다는 것을 시사한다. 이러한 아이디어와 일관되게, 출원인은 시험관내에서 BvCas12b로 임의의 닉킹 산물을 검출하지 않았다 (도 53b). 또한, 글리신 치환이 BhCas12b E837G에 상응하므로, BvCas12b의 표적 가닥 포켓 내 표적화 돌연변이는 모두 활성을 감소시켰다 (도 53c-53e).

[2021] 강력한 게놈 편집 도구는 표적 범위에 걸쳐 효과적이고 특이적인 것이 바람직할 수 있다. 출원인은 이전에 연구된 Cas 뉴클레아제와 비교하여 보다 철저하게 Cas12b를 조사하였다. 출원인은 293T 세포에서 5개 유전자에 걸쳐 56개 표적에서 BhCas12b v4 및 BvCas12b를 시험하였고 양성 대조군으로서 ATTN PAM에서 AsCas12a를 사용해 ATTN PAM에서 강력한 절단을 발견하였다 (도 46a). Cas12b에 의해 형성된 indel 패턴의 분석은 우세한 5-15bp 결실을 밝혀주었다 (도 46b). 출원인은 TTTN 및 GTTN PAM의 서브세트에서 높은 Cas12b 활성을 관찰하였지만, 이러한 활성은 덜 강력하였다 (도 54a). 출원인은 매치되는 부위에서 BhCas12b v4 및 BvCas12b의 활성 사이에 약한 상관성만을 관찰하였고 (R^2=0.48), 수많은 표적이 2개 뉴클레아제 중 하나에 의해 보다 효율적으로 절단되었다 (도 54b). 이들 발견은 다수 오솔로그의 이득을 강조하고 계속적으로 Cas 뉴클

레아제의 표적화 규칙을 철저하게 조사할 필요가 있었다. 인간 게놈에서 ATTN 출현율의 분석은 Cas12b 효소에 대한 유사한 표적가능성을 밝혀주었다 (도 54c). SpCas9 및 AsCas12a와 대조적으로, BhCas12b에 의해 형성된 indel 패턴의 분석은 5-15 bp의 두드러지게 더 큰 결실을 밝혀주었다 (도 46f). 단일-가닥 올리고뉴클레오티드 (ssODN) 도너와 Cas12b 뉴클레아제의 공동 형질감염은 TTTC PAM 표적에서 SpCas9 및 AsCas12a와 비슷한 편집 효율을 야기시켰고 (도 46c-46e), ATTC PAM 표적에서 더 높은 편집 효율을 야기시켰다 (도 54d-54f).

인간 세포에서 BhCas12b v4의 효율을 더욱 평가하기 위해서, 출원인은 초대 인간 T 세포를 편집하는 Cas12b 리보핵단백질 (RNP)의 능력을 시험하였다. 출원인은 BhCas12b v4-sgRNA 복합체를 생성시켰고 그들을 전기천공을 통해서 인간 CD4+ T 세포로 전달하였다. BhCas12b v4 RNP는 3개 시험된 표적에 걸쳐서 32-49%의 indel 비율을 나타냈다 (도 46g). 함께, 이들 데이터는 BhCas12 v4 및 BvCas12b는 치료적으로 관련된 인간 세포 유형을 포함하여, 다양한 게놈 편집 상황에서 기능성의 프로그램가능한 뉴클레아제로서 이용될 수 있다는 것을 시사한다.

[2022] 다음으로 출원인은 세포에서 Cas12b 특이성을 결정하고자 하였다. 출원인은 상이한 Cas 뉴클레아제 간에 비슷한 indel 활성을 갖는 9개 표적 부위를 선택하였고 (도 47a), 이들 표적으로 Guide-Seq 분석을 수행하였다[19].

출원인은 Cas12b 뉴클레아제 및 AsCas12a 둘 모두에 대해 임의의 오프-표적 부위를 검출하지 않았는데 반해서, SpCas9는 9개 시험된 가이드 중 6개에서 현저한 오프-표적 절단을 초래하여 (도 47b, 도 55), 이의 기지의 혼잡성과 일관된다[13,20]. 예를 들어, 표적 3의 경우에, 출원인은 SpCas9로 101개 삽입 부위를 검출하였고, 판독치의 오직 10%만이 표적 부위로 맵핑되었으며, 2개 Cas12b 효소 중 어느 것에서도 오프-표적 부위는 없었다. 비매치된 부위에서 추가의 가이드-Seq 실험은 BhCas12b v4의 경우에 14개 부위 중 오직 2개에서, 그리고 BvCas12b의 경우에 15개 부위 중 1개에서 유의한 오프-표적 절단이 검출되었다 (도 56a, 57). 이들 발견과 일관되게, 출원인은 위치 1-20에서 가이드 RNA 및 표적 DNA 간에 이중 미스매치로 제한된 indel 활성을 관찰하였고 심지어 단일 미스매치에 대해 낮은 내성이 관찰되었다 (도 56b 및 56c). 이들 결과는 시험관내에서 AacCas12b의 보고된 특이성과 일관되고[21], 세포에서 관찰된 낮은 오프-표적 활성에 대해 분자 기전을 제공한다.

[2023] 여기서 출원인은 인간 세포에서 게놈 편집에 적합한 V형 CRISPR Cas12b 패밀리의 처음 2개 멤버를 기술한다. 많은 Cas12b 뉴클레아제가 고온에서 강력한 선호도를 보이지만, 우리의 광범위한 스크리닝은 37℃에서 고도로 활성인 이러한 패밀리의 멤버의 동정을 야기시켰다. 더 나아가서, BhCas12b의 우리의 조작은 dsDNA 절단의 효율에서 실질적인 증가를 야기시켰고 게놈 편집 도구로서 다른 Cas12b 뉴클레아제의 가능성을 열어주는 프레임워크를 제공한다. BhCas12b 및 BvCas12b 둘 모두는 비교적 조밀한 단백질 (각각 약 1100개 아미노산)이므로, 아데노-연관 바이러스 (AAV)로 효율적인 패키징에 적합하다. 그들의 높은 표적 특이성과 조합하여, 이들 Cas12b 효소는 생체내 게놈 편집을 위한 유망한 신규 도구이다.

[2024] **보충 정보**

[2025] **Cas12b 패밀리 단백질의 다중 정렬**

[2026] 서열은 수탁 번호로 표시된다. 바실러스 sp V3-13 (WP 101661451.1) 및 바실러스 히사시이 (WP_095142515.1)의 서열은 붉은색으로 강조된다. 이 작업에서 돌연변이된 12개 잔기는 비. 히사시이 (WP_095142515.1) 서열에서 붉게 강조하여 표시된다. 치환이 BhCas12 v4 돌연변이체에서 DNA 절단 효율에 실질적으로 영향을 미친 잔기는 붉은색 강조되고, 노란색으로 표시된다.

[2027] **재료 및 방법**

[2028] Cas12b 서열 정렬 및 계통수 재구성

[2029] 정렬은 MUSCLE 프로그램 (v 3.7)[23]을 사용해 구성되었다. 정렬은 하기 아미노산 그룹에 대한 100% 일치에 따라서 www.bioinformatics.org/sms2/color_align_cons.htm 서버를 사용해 착색하였다: GAVLI, FYW, CM, ST, KRH, DENQ, P. 50% 초과의 갭을 갖는 위치는 계통수 재구성을 위해 사용된 정렬로부터 폐기하였다. PHYML 프로그램 (v. 20120412)[24]을 사용하여 최대-가능성 뿌리없는 계통수를 생성시켰다. 동일한 프로그램을 통한 사용하여, 선택된 분지에 대해 표시된, 부트스트랩 값을 산출하였다.

[2030] **Cas12b 발현 플라스미드의 생성**

[2031] Cas12b 유전자좌를 합성하였고 이. 콜라이에서 발현을 위해 pACYC184 (Genewiz)에 클로닝하였다. Cas12b 오픈 리딩 프레임 (ORF)는 인간 발현을 위해 코돈 최적화하였디만 ORF 측접된 상류 및 하류 서열은 변화시키지

않고 남겨두었다. CRISPR 어레이는 3 직접 반복부로 단축시켰고 제1 내생성 스페이서는 FnCpf1 프로토스페이서 1 (FnPSP1) 서열 (GAGAAGTCATTTAATAAGGCCACTGTTAAAA) (SEQ ID No:591)로 치환하였다.

[2032] **PAM 발굴**

[2033] PAM의 동정은 이전에 기술된 대로 수행하였다[10]. 간략하게, pACYC184-Cas12b 시스템을 발현하는 이. 콜라이 세포는 Z-competent kit (Zymo Research)로 적격하게 만들었다. pACYC184-Cas12b 또는 빈 pACYC184를 발현하는 세포는 FnPSP1 표적 부위의 5' 측면 상에 무작위 8N 서열을 갖는 PAM 라이브러리로 형질전환시켰고, 16시간 동안 밤새 성장시켰다. 플라스미드 DNA를 단리하였고, 라이브러리는 75-cycle NextSeq kit (Illumina)를 사용해 시퀀싱하였다. 라이브러리 내 PAM 표시는 고객 Python 스크립트를 사용해 결정하였고 2개 독립 복제물로 대조군과 Cas12b를 비교하였다. 서열 모티프는 Weblogo tool (weblogo.berkeley.edu)을 사용해 생성시켰다. PAM 휠은 Krona plots (github.com/marbl/Krona/wiki)[22]을 사용해 생성시켰다.

[2034] **박테리아 RNA 시퀀싱**

[2035] 소형 RNA-Seq는 이전에 기술된 대로 수행하였다[1,10]. 간략하게, RNA는 TRIzol을 사용해 이. 콜라이 용해물로부터 제조한 후에 BeadBeater (BioSpec Products)를 사용해 균질화시켰다. rRNA는 Ribo-Zero kit (Illumina)를 사용해 제거하였고 라이브러리는 NEBNext Small RNA Library Kit for Illumina (NEB)를 사용해 제조하였다. 라이브러리는 2x150 쌍형성-말단 MiSeq run (Illumina)으로 시퀀싱하였고 판독치를 정렬하고 Geneious R9 (Biomatters)로 분석하였다.

[2036] **Cas12b 단백질의 정제**

[2037] Cas12b 유전자는 박테리아 발현 플라스미드 (T7-TwinStrep-SUMO-NLS-Cas12b-NLS-3xHA)에 클로닝하였고 BL21(DE3) 세포 (Novagen #70956의 pLysS-tRNA 플라스미드 함유 NEB #C2527H)에서 발현시켰다. 세포는 중간 대수기까지 Terrific 액체배지에서 성장시켰고 온도를 20℃로 낮추었다. 발현은 수확전에 0.6 OD에서 0.25 mM IPTG를 사용해 16-20시간 동안 유도시켰고 -80℃에 세포를 냉동시켰다. 세포 페이스트를 EDTA-무함유 완전 프로테아제 억제제 (Roche)가 보충된 용해 완충액 (50 mM TRIS pH 8, 500 mM NaCl, 5% 글리세롤, 1 mM DTT)에 재현탁시켰다. 세포는 LM20 미세유동기 장치 (Microfluidics)를 사용해 용해시켰고 Strep-Tactin Superflow Plus 레진 (Qiagen)에 결합된 용해물을 제거하였다. 레진은 용해 완충액으로 세척하였고 Cas12b 단백질은 5 mM 데스티오바이오틴이 보충된 용해 완충액으로 용리하였다. TwinStrep-SUMO 태그는 홈메이드 SUMO 프로테아제 Ulp1를 사용해 1:100의 프로테아제 대 Cas12b의 중량비로 4℃에서 밤새 분해하여 제거하였다. 절단된 Cas12b는 200 mM NaCl 까지 희석하였고 200 mM-1 M NaCl 농도구배로 AKTA Pure 25 L (GE Healthcare Life Sciences) 상에서 HiTrap Heparin HP 컬럼을 사용해 정제하였다. Cas12b를 사용하는 분획을 모았고 농축하였으며 25 mM TRIS pH 8, 500 mM NaCl, 5% 글리세롤, 1 mM DTT의 최종 저장 완충액을 사용해 Superdex 200 Increase 컬럼 (GE Healthcare Life Sciences) 상에 로딩하였다. 정제된 Cas12b 단백질을 5 uM 또는 73 uM 스톡으로 농축하였고, -80℃에 저장 전에 액체 질소에서 급속 냉동시켰다.

[2038] **시험관내 RNA 합성**

[2039] 모든 RNA는 짧은 T7 올리고뉴클레오티드와 바람직한 RNA의 역상보체를 함유하는 DNA 올리고뉴클레오티드를 어닐링시켜 생성시켰다. 시험관내 전사는 HiScribe T7 High Yield RNA 합성 키트 (NEB)를 사용해 37℃에서 8-12시간 동안 수행하였고 RNA는 Agencourt AMPure RNA Clean beads (Beckman Coulter)를 사용해 정제하였다.

[2040] **시험관내 절단 반응**

[2041] DNA 기질은 FnPSP1 표적 부위를 함유하는 pUC19 플라스미드의 PCR 증폭으로 생성시켰다. 전형적인 반응물은 100 ng의 DNA 기질, 250 nM의 Cas12b 단백질, 500 nM의 RNA 및 20 mM TRIS pH 6.5, 6 mM MgCl2의 최종 1x 반응 완충액을 함유하였다. 반응은 20 mM EDTA로 켄칭하였고, RNA는 5 ug RNAse A (Qiagen)로 37℃에서 5분 동안 분해시켰으며, DNA 산물은 PCR cleanup kit (Qiagen)를 사용해 정제하였다. 반응물을 Novex 10% TBE PAGE 겔 상에서 1x TBE 완충액 (Thermo Fisher Scientific) 중에 러닝하였고 SYBR Gold (Thermo Fisher Scientific)로 염색하였다. 표지된 DNA 기질은 IR700 및 IR800 접합된 DNA 올리고뉴클레오티드 (IDT)로 생성시켰다. 변성 겔의 경우, DNA는 동일 부피의 100% 포름아미드와 혼합하고 나서 95℃에서 5분 동안 열변성시켰다. 산물은 60℃로 예열시킨 1x TBE 완충액 중에 Novex Urea-PAGE 겔 (Thermo Fisher Scientific) 상에서 분리하였고 Odyssey CLx 장치 (LI-COR) 상에서 이미지화시켰다. 적용가능한 경우에, DNA 절단 또는 닉킹의 정량은 식, 100 x (1 - sqrt(1 - (b+c)/(a+b+c)))으로 결정하였고, 여기서 a는 미분해 산물의 통합 강도

이고, b 및 c는 각 절단 또는 닉킹 산물의 통합 강도이다.

[2042] **포유동물 발현 구성체 및 돌연변이유발**

[2043] Cas12b 유전자는 그들의 상응하는 pACYC184 플라스미드로부터 증폭시켰고 N- 및 C-말단 NLS 태그 및 C-말단 3xHa 태그를 함유하는 pCDNA3.1에 클로닝시켰다. 가이드 발현 플라스미드는 U6 프로모터 뒤에 2개의 인버티드 BsmBI IIS형 제한효소 부위를 함유하는 sgRNA 스캐폴드를 클로닝하여 생성시켰다. 가이드는 2개의 어닐링된 상보성 올리고뉴클레오티드와 골든 게이트 조립을 통해 스캐폴드에 클로닝하였다. 표시하지 않으면, 모든 가이드는 23-nt 길이였다. 바람직한 Cas12b 돌연변이는 Gibson Assembly Master Mix (NEB)를 사용하여 조립시킨 2개의 중복 Cas12b PCR 산물을 생성시키도록 올리고뉴클레오티드 상에 정렬시켰다. 사용된 가이드 서열은 하기 표 31에 표시된다.

[2044] **표 31**

[2045]

유전자	표적	5' PAM	서열
DNMT1	1	GTTCT	AGACCCAGAGGCTCAAGTGAGCA (SEQ ID NO:592)
DNMT1	2	ATTTT	AGCTGAAGGGAAATAAAAGGAAA (SEQ ID NO:593)
VEGFA	3	ATTCT	TCTCCCCTGGGAAGCATCCCTGG (SEQ ID NO:594)
EMX1	4	ATTTT	TCATGGAGAAAATATTCAGAATC (SEQ ID NO:595)
DNMT1	5	TTTC	CCTCACTCCTGCTCGGTGAATTT (SEQ ID NO:596)
DNMT1	6	TTTG	AGGAGTGTTCAGTCTCCGTGAAC (SEQ ID NO:597)
VEGFA	7	TTTG	GGAGGTCAGAAATAGGGGGTCCA (SEQ ID NO:598)
VEGFA	8	TTTC	CAAAGCCCATTCCCTCTTTAGCC (SEQ ID NO:599)
DNMT1	9	ATTT	CCCTTCAGCTAAAATAAAGGAGG (SEQ ID NO:600)
VEGFA	10	ATTC	TTCTCCCCTGGGAAGCATCCCTG (SEQ ID NO:601)
GRIN2B	11	ATTC	TGCAGAGCAAATACCAGAGATAA (SEQ ID NO:602)
PDCD1	12	ATTG	CGCCGGGCCCTGACCACGCTCAT (SEQ ID NO:603)
CXCR4	13	ATTC	CCGACTTCATCTTTGCCAACGTC (SEQ ID NO:604)
EMX1	14	ATTT	TAGAGCACTGGCATGGGGATGGG (SEQ ID NO:605)
EMX1	15	ATTC	TTGCTCCAGAGGCCCCCCTTGGG (SEQ ID NO:606)
DNMT1	16	ATTC	CTGGTGCCAGAAACAGGGGTGAC (SEQ ID NO:607)
CXCR4	17	ATTC	TGGGCTTCAAGCAACTTGTAGTG (SEQ ID NO:608)
CXCR4	18	ATTT	TGTAATTGGTTCTACCAAAGAAG (SEQ ID NO:609)
CXCR4	19	ATTT	AGAGGCGGAGGGCGGCGTGCCTG (SEQ ID NO:610)
GRIN2B	20	TTTC	CTTCAGCCCAAGAACAGTACAAG (SEQ ID NO:611)
CXCR4	21	TTTC	TCTGTGAGTCGAGGAGAAACGAC (SEQ ID NO:612)

HPRT1	22	TTTC	CTTGGGTGTGTTAAAAGTGACCA (SEQ ID NO:613)
DNMT1	23	-	TCACTCCTGCTCGGTGAATT (SEQ ID NO:614)
EMX1	24	-	GCTACAGGCAGAGACAAAGG (SEQ ID NO:615)
VEGFA	25	-	AGGTCAGAAATAGGGGGTCC (SEQ ID NO:616)
VEGFA	26	-	CAGGCTGTGAACCTTGGTGG (SEQ ID NO:617)
VEGFA	27	-	GACCCCCTCCACCCCGCCTC (SEQ ID NO:618)
GRIN2B	28	-	GTATCTAGCCTCTTCTAAGAC (SEQ ID NO:619)
VEGFA	29	-	TCTCCCCTGGGAAGCATCCC (SEQ ID NO:620)
EMX1	30	-	GAGTCCGAGCAGAAGAAGAA (SEQ ID NO:621)
TUBB	31	-	TTTTGGGAGTAAGAAAAGGT (SEQ ID NO:622)
VEGFA	32	-	AGTGTCCAGGGATGCTTCCC (SEQ ID NO:623)
DNMT1	33	TTTC	CCTCACTCCTGCTCGGTGAATTT (SEQ ID NO:624)
VEGFA	34	TTTG	GGAGGTCAGAAATAGGGGGTCCA (SEQ ID NO:625)
EMX1	35	TTTG	GATGGCGACTTCAGGCACAGGAT (SEQ ID NO:626)
EMX1	36	TTTG	GGAAGTGTCCAGGGATGCTTCCC (SEQ ID NO:627)
DNMT1	37	ATTT	CCCTTCAGCTAAAATAAAGGAGG (SEQ ID NO:628)
DNMT1	38	ATTT	GGCTCAGCAGGCACCTGCCTCAG (SEQ ID NO:629)
VEGFA	39	ATTT	GGGACTGGAGTTGCTTCATGTAC (SEQ ID NO:630)
EMX1	40	ATTT	TCTCCATGAAAAATACTGGGGTC (SEQ ID NO:631)
EMX1	41	ATTT	TTCATGGAGAAAATATTCAGAAT (SEQ ID NO:632)
GRIN2B	42	ATTG	GCAGCTACAGGCAGAGACAAAGG (SEQ ID NO:633)
EMX1	43	ATTT	CCTGGAAACCATCCAGGCCTTGT (SEQ ID NO:634)
DNMT1	44	ATTG	GGTCAGCTGTTAACATCAGTACG (SEQ ID NO:635)
CXCR4	45	ATTT	TCTTCACGGAAACAGGGTTCCTT (SEQ ID NO:636)
EMX1	46	TTTG	TGGTTGCCCACCCTAGTCATTGG (SEQ ID NO:637)
EMX1	47	TTTG	GATGGCGACTTCAGGCACAGGAT (SEQ ID NO:638)
DNMT1	49	TTTC	CCTCACTCCTGCTCGGTGAATTT (SEQ ID NO:639)
VEGFA	50	TTTG	GGAAGTGTCCAGGGATGCTTCCC (SEQ ID NO:640)
DNMT1	51	ATTT	GGCTCAGCAGGCACCTGCCTCAG (SEQ ID NO:641)
EMX1	52	ATTT	TTCATGGAGAAAATATTCAGAAT (SEQ ID NO:642)
VEGFA	53	ATTT	CTGACCTCCCAAACAGCTACATA (SEQ ID NO:643)

[2046]

[2047] **세포 배양 및 형질감염**

[2048] HEK293T 세포 (ATCC)는 고 글루코스, 소듐 피루베이트, 및 GlutaMAX (Thermo Fisher Scientific), 1x 페니실린-스트렙토마이신 (Thermo Fisher Scientific), 및 10% 태아 소 혈청 (Seradigm)이 존재하는 둘베코의 변형 이글 배지에서 배양하였다. 세포는 90% 이하의 융합점으로 유지시켰고 MycoAlert detection kit (Lonza)를 사용해 마이코플라스마 음성을 시험하였다. Indel 분석을 위해서, 96웰 플레이트는 형질감염 시 대략 75%의 융합점을 위해 형질감염 전에 대략 16시간 동안 17,500 세포/웰이 파종되었다. 각각의 96-웰은 20 uL의 Opti-MEM (Thermo Fisher Scientific)과 0.6 uL의 TransIt-LT1 형질감염 시약 (Mirus) 중 100 ng의 가이드 플라스미드 및 100 ng의 뉴클레아제 발현 플라스미드로 형질감염시켰다. 세포는 QuickExtract DNA 추출 용액 (Lucigen)으로 형질감염 후 72시간에 수확하였다.

[2049] HDR 실험을 위해서, 100 ng의 뉴클레아제, 100 ng의 가이드 및 100 ng의 ssODN가 96웰 당 0.9 uL의 TransIt-LT1 형질감염 시약 (Mirus)으로 형질감염되었다. ssODN는 Ultramer DNA 올리고뉴클레오티드 (IDT)로서 주문하였고 각 말단에 3 포스포로티오에이트 변형을 함유하였다.

[2050] **Indel 돌연변이의 심층 시퀀싱**

[2051] 표적화된 indel 분석은 Illumina P5 어댑터 및 고유한 샘플-특이적 바코드를 첨가하기 위해 2-라운드 PCR 전략을 사용해 NEBNext High-Fidelity 2x PCR Master Mix (NEB)로 관심 게놈 영역을 증폭시켜 수행하였다. 라이브러리는 1x200 cycle MiSeq runs (Illumina)로 시퀀싱하였다. Indel 비율은 Outknocker 2[25]를 사용해 측정하였다.

[2052] (www.outknocker.org/outknocker2.htm).

[2053] **오프-표적 분석**

[2054] 오프-표적 절단 부위는 변형 라이브러리 제조로 가이드-Seq를 사용해 확인하였다. 간략하게, 세포는 0.5 uL GeneJuice 형질감염 시약 (Millipore) 존재의 50 uL Opti-MEM 중 75 ng 뉴클레아제 플라스미드, 25 ng 가이드 플라스미드, 및 100 ng 어닐링된 dsDNA 올리고를 사용해 96웰 플레이트에서 형질감염시켰다.

[2055] F: /5phos/G*T*TGTGAGCAAGGGCGAGGAGGATAACGCCTCTCTCCCAGCGACT*A*T (SEQ ID NO:644)

[2056] R: /5phos/A*T*AGTCGCTGGGAGAGAGGCGTTATCCTCCTCGCCCTTGCTCACA*A*C (SEQ ID NO:645)

[2057] 세포는 72시간 후에 회수하였고 각 실험에 대해 10 웰을 모았다. 1E6 세포를 용해시켰고 게놈 DNA를 Tn5로 태그화시킨 후에 플라스미드 미니-프렙 컬럼 (Qiagen)을 사용해 정제하였다. 라이브러리는 Tn5 어댑터-특이적 프라이머 및 DNA 도너 내 네스티드 프라이머를 사용해 KOD Hot Start DNA 중합효소 (Millipore)로 2 라운드의 PCR 증폭을 사용해 제조하였다. 라이브러리는 75-cycle NextSeq 키트 (Illumina)로 시퀀싱하였다. 판독치는 BrowserGenome.org[26]를 사용해 인간 게놈 (hg38)에 대해 맵핑하였다.

[2058] **T 세포 배양**

[2059] 인간 CD4+ T 세포 (STEMCELL Technologies)는 5mM HEPES, pH 8.0 (Gibco), 50 ug/mL 페니실린/스트렙토마이신 (Gibco), 50 uM 2-머캅토에탄올 (Sigma-Aldrich), 5 mM MEM 불필수 아미노산 (Gibco), 5 mM 소듐 피루베이트 (Gibco) 및 10% FBS (Seradigm)가 보충된 RMPI 1640 (Glutamax Supplement, Gibco)에서 배양하였다. 세포는 10 ug/mL의 항-CD3 (UCHT1, eBioscience, Invitrogen) 및 항-CD28 (CD28.2, eBioscience, Invitrogen) 단일클론 항체가 코팅된 디쉬 상에서 2일 마다 플레이팅하여 해동 후 5-7일 동안 활성화시켰다.

[2060] RNP 복합체화 및 전달

[2061] BhCas12b sgRNA는 3' 말단에 3개 2'O-메틸 변형을 갖게 합성하였다 (Integrated DNA Technologies). RNP는 10 mg/mL 단백질과 50 uM의 어닐링된 RNA를 1:1 몰비율로, 37℃에서 15분 동안 인큐베이션시켜 형성시켰다. RNP는 전기영동까지 얼음 상에 저장하였다.

[2062] 세포는 Amaxa P3 Primary Cell 4D-Nucleofector X Kit (Lonza)를 사용해 전기천공시켰다. 반응 당, 3 x 10^5 자극된 CD4+ T 세포를 펠렛화시켰고 20uL의 P3 완충액에 재현탁시켰다. crRNA 및 tracrRNA와 사전복합체 형성된, 4.5 uM Cas9 또는 Cas12b 단백질을 첨가하였고 혼합물을 전기천공 큐벳으로 옮겼다. 세포는 프로그램 Eh-115를 사용하여 Amaxa 4D-Nucleofector (Lonza) 상에서 전기천공하였다. 80 uL의 예열된 완전 배지를 즉시 펄스 후 세포에 첨가하였고 세포를 큐벳에서 회수하기 위해서 30분 동안 37 C에 인큐베이션시켰다.

회수 후, 50 uL의 세포 현탁액을 40 IU/mL Il-2의 최종 농도로 80 IU/mL IL-2 (STEMCELL Technologies)가 더해진 50 uL의 완전 배지에 첨가하였다. 세포는 CD3/CD28 사전코팅된 96웰 플레이트 상에 플레이팅하였다.

세포는 indel 분석을 위해 48시간 후에 회수하였다.

[2063] 참조문헌

- 1 Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **60**, 385-397, doi:10.1016/j.molcel.2015.10.008 (2015).
- 2 Seuylemezian, A., Cooper, K., Schubert, W. & Vaishampayan, P. Draft Genome Sequences of 12 Dry-Heat-Resistant Bacillus Strains Isolated from the Cleanrooms Where the Viking Spacecraft Were Assembled. *Genome announcements* **6**, doi:10.1128/genomeA.00094-18 (2018).
- 3 Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866-869, doi:10.1126/science.aat5011 (2018).
- 4 Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262-1278, doi:10.1016/j.cell.2014.05.010 (2014).
- 5 Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
- 6 Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).
- 7 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607, doi:10.1038/nature09886 (2011).
- 8 Bolotin, A., Ouinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiol-Sgm* **151**, 2551-2561, doi:10.1099/mic.0.28048-0 (2005).
- 9 Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**, doi:Artn 710.1186/1745-6150-1-7 (2006).
- 10 Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771, doi:10.1016/j.cell.2015.09.038 (2015).

[2064]

- 11 Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182, doi:10.1038/nrmicro.2016.184 (2017).
- 12 Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019-1027, doi:10.1126/science.aaq0180 (2017).
- 13 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-+, doi:10.1038/nbt.2647 (2013).
- 14 Yang, H., Gao, P., Rajashankar, K. R. & Patel, D. J. PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* **167**, 1814-1828 e1812, doi:10.1016/j.cell.2016.11.053 (2016).
- 15 Wu, D., Guan, X., Zhu, Y., Ren, K. & Huang, Z. Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA. *Cell Res* **27**, 705-708, doi:10.1038/cr.2017.46 (2017).
- 16 Suzuki, M. A framework for the DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure (London, England : 1993)* **2**, 317-326 (1994).
- 17 Mavromatis, K., Tsigos, I., Tzanodaskalaki, M., Kokkinidis, M. & Bouriotis, V. Exploring the role of a glycine cluster in cold adaptation of an alkaline phosphatase. *Eur J Biochem* **269**, 2330-2335 (2002).
- 18 Saavedra, H. G., Wrabl, J. O., Anderson, J. A., Li, J. & Hilser, V. J. Dynamic allostery can drive cold adaptation in enzymes. *Nature* **558**, 324-+, doi:10.1038/s41586-018-0183-2 (2018).
- 19 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197, doi:10.1038/nbt.3117 (2015).
- 20 Fu, Y. F. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-+, doi:10.1038/nbt.2623 (2013).
- 21 Liu, L. *et al.* C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. *Mol Cell* **65**, 310-322, doi:10.1016/j.molcel.2016.11.040 (2017).
- 22 Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Molecular Cell* **62**, 137-147, doi:10.1016/j.molcel.2016.02.031 (2016).
- 23 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

[2065]

- 24 Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).
- 25 Schmid-Burgk, J. L. *et al.* OutKnocker: a web tool for rapid and simple genotyping of designer nuclease edited cell lines. *Genome Res* **24**, 1719-1723, doi:10.1101/gr.176701.114 (2014).
- 26 Schmid-Burgk, J. L. & Hornung, V. BrowserGenome.org: web-based RNA-seq data analysis and visualization. *Nat Methods* **12**, 1001, doi:10.1038/nmeth.3615 (2015).
- Jinek, M. et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821 (2012).
- Teng, F. et al. Repurposing CRISPR-Cas12b for mammalian genome engineering. *Cell Discov* **4**, 63 (2018).

[2066]

[2067] **실시예 26**

[2068] 도 58은 Cas12b (C2c1) 구조를 도시한다 (PDB 구조 5U30 기반). 도면은 구조적으로 예측된 ssDNA 경로, 및 추

가 ssDNA에 접근하기 위해, 부분적으로 또는 전체로, 제거될 수 있는 도메인을 도시한다.

[2069] **실시예 27**

[2070] ADAR 활성에 영향을 미치는 ADAR 내 돌연변이는 효모 스크리닝을 사용하여 스크리닝하였다. 스크린은 다수 라운드로 수행되었다. 각 라운드의 스크리닝은 한 세트의 후보 돌연변이를 산출하였다. 그 다음으로 후보 돌연변이는 포유동물 세포에서 검증하였다. 상위-수행 돌연변이를 마지막 형태의 돌연변이에 첨가하였고 재스크리닝하였다. 10라운드에서 스크리닝된 돌연변이는 하기 표 32에 표시되어 있다. n 라운드에서 동정된 돌연변이체는 " RESCUE vn-1" 로 표시하였다. 본 명세서에서 논의되는 바와 같이 RESCUE는 아데노신 디아미나제 활성을 시티딘 디아미나제 활성으로 전환시킨 돌연변이를 의미한다.

[2071] **표 32**

RESCUE 라운드	ADAR 돌연변이	플라스미드
RESCUEv0	E488Q	pAB0048
RESCUEv1	E488Q, V351G	pAB0359
RESCUEv2	E488Q, V351G, S486A	pAB1188

[2072]

RESCUEv3	E488Q, V351G, S486A, T375S	pAB0642
RESCUEv4	E488Q, V351G, S486A, T375S, S370C,	pAB1072
RESCUEv5	E488Q, V351G, S486A, T375S, S370C, P462A	pAB1135
RESCUEv6	E488Q, V351G, S486A, T375S, S370C, P462A, N597I	pAB1146
RESCUEv7	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I	pAB1194
RESCUEv8	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V	pAB1220
RESCUEv9	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I	pAB1327

[2073]

[2074] RESCUE 돌연변이체의 용량 반응은 T 모티프 (도 59) 및 C 및 G 모티프 (도 60) 상에서 시험되었다. RESCUE v3, v6, v7, 및 v8로 내생성 표적화를 시험하였다 (도 61 및 62).

[2075] RESCUE v9에 대한 돌연변이에 대한 스크리닝을 수행하였다 (도 63). RESCUEv9에 대한 잠재적 돌연변이를 확인하였다 (도 64). 염기 플립 및 모티프 시험을 수행하였다 (도 65). RESCUEv9의 효과는 상이한 모티프 플립에서 수행되었다 (도 66). 데이터는 v9가 C-플립 가이드에서 더 양호하게 작동한다는 것을 시사한다. RESCUE v1 및 v8에 의한 B6 및 B12 간 비교는 50 bp 가이드 (도 67) 및 30 bp 가이드 (도 68)로 수행하였다.

[2076] **실시예 28**

[2077] 이 실시예는 RESCUE 라운드 1-12의 결과를 요약한다 (도 69-80 참조). 시험된 추가 표현형은 PCSK9, Stat3, IRS1, 및 TFEB를 포함하였다. PCSK9는 클로닝이 프로모터를 개선시켰음을 보여주었다. Stat3은 부위 상에서 ~10% 편집을 보였다. 신호전달의 억제는 루시퍼라제 리포터로 시험하였다. IRS1의 경우, 합성 부위의 표적화는 전-지방세포 이동 전에 시험될 것이다. TFEB의 경우, 표적화는 전사 인자 -> 오토파지의 전좌를 초래하게 설계될 수 있다. 또한, 12개 내생성 포스포사이트 표적 및 48개 합성 표적의 패널이 시험될 것이다. 효모에서 스크리닝은 S22P 존재의 V11 배경에서 계속될 것이다. 상위 히트는 V13에 대해 V12에서 스크리닝되었고 새로운 라운드의 효모 히트를 평가하였다. 루피퍼라제에 대한 수백개 추가 스크린 히트를 평가할 것이고 Ade2 편집은 특이성 스크리닝을 위해 검증될 것이다. 유전자 셔플링이 또한 라이브러리 복잡성 및 상이한 효모 리포터에 대해 시험될 것이다.

[2078] **실시예 29**

[2079] 이 실시예는 Cas12b 및 이의 변이체 및 디아미나제를 사용한 염기 편집을 위한 예시적인 접근법을 입증한다.

[2080] 도 81, 83-86은 C에서 T 염기 편집 능력을 갖는 Cas12b Bhv4 절두체를 도시한다. 촉매적 불활성 Bhv4의 C-

말단 142 아미노산 (dBhv4Δ142- 불활성화 돌연변이 D574A, 새로운 크기 966 아미노산)의 제거 및 링커 및 래트 Apobec 도메인과 C-말단부의 융합 후에, C에서 T 염기 편집은 비표적 가닥 상의 가이드 염기쌍 위치 14에서 최대 10.95%의 빈도로 관찰되었다. 6.97% 편집 효율이 가이드 위치 15에서 검출되었다. 이 활성은 가이드 의존적이었다. 존재하는 구성체와의 융합 또는 자유 발현을 통해서, 우라실-DNA 글리코실라제 억제제 (UGI) 도메인의 첨가는 이러한 C에서 T로의 전환을 증가시킨다. 열거된 가이드 서열 (대문자)은 HEK 293T 세포에서 GRIN2B 내부의 영역을 표적으로 한다.

[2081] 도 87 은 전체 길이 BhCas12b 를 사용한 예시적인 염기 편집 접근법을 도시한다. 제2 NLS 서열은 도메인들이 서로 거리를 두도록 N-말단 rApobec을 첨가하였다.

[2082] **실시예 30**

[2083] 도 88A는 BhCas12b v4 및 다른 오솔로그 AaCas12b의 indel 활성 간 비교를 도시한다 ([Teng F. et al., Repurposing CRISPR-Cas12b for mammalian genome engineering in HEK293T cells]에 기술된 바와 같음). 도 88B 및 88C는 BhCas12b v4 또는 BhCas12b를 발현하는 AAV1/2로 뉴런의 형질도입을 입증한다. 이러한 디자인은 inde 활성으로 측정시 더 높은 활성을 나타내었다. 폴리A 서열은 최적화된 서열에서 길어졌고 U6 프로모터 및 sgRNA 스캐폴드는 반대 가닥으로 이동하였다.

[2084] 이 연구에서의 서열은 하기 표 33 에 나타냈다. px602-bh-optimize-AAV의 맵은 도 89A에 도시되어 있고, px602-bv-optimize-AAV의 맵은 도 89B에 도시되어 있다.

[2085] **표 33**

서열 명칭	서열
BhCas12b Mecp2 표적 (SEQ ID NO:646):	TACTTTAGAGCGAAAGGCTTTTC
BhCas12b Map2 표적 (SEQ ID NO:647):	AGATACCAAAGAGAACGGGATCA
BvCas12b Mecp2 표적 (SEQ ID NO:648):	TGACTTCACTGTAACTGGGAGAG
BvCas12b Map2 표적 (SEQ ID NO:649):	TCTCCCAGGAAGGGCAGCAGGCT
BvCas12b NeuN 표적 (SEQ ID NO:650):	TAGACGTGGAGATCATTTTTAAC
오리지널 폴리 A (SEQ ID NO:651):	Aataaaatatctttatttttcattacatctgtgtgttggttttttgtgtg
bGH 폴리 A (최적화된 구성체) (SEQ ID NO:652):	Ctgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagagaatagcaggcatgctgggga
NLS-AaCas12b-NLS-3xHA	MAPKKKRKVGIHGVPAAAVKSMKVKLRLDNMPEIRA

[2086]

(SEQ ID NO:653)	GLWKLHTEVNAGVRYYTEWLSLLRQENLYRRSPNGD GEQECYKTAEECKAELLERLRARQVENGHCGPAGSD DELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLA DKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKA KAEARKSTDRTADVLRALADFGLKPLMRVYTDSDMS SVQWKPLRKGQAVRTWDRDMFQQAIERMMSWESW NQRVGEAYAKLVEQKSRFEQKNFVGQEHLVQLVNQL QQDMKEASHGLESKEQTAHYLTGRALRGSDKVFEK WEKLDPDAPFDLYDTEIKNVQRRNTRRFGSHDLFAKL AEPKYQALWREDASFLTRYAVYNSIVRKLNHAKMFA TFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGEGR HAIRFQKLLTVEDGVAKEVDDVTVPISMSAQLDDLLP RDPHELVALYFQDYGAEQHLAGEFGGAKIQYRRDQL NHLHARRGARDVYLNLSVRVQSQSEARGERRPPYAA VFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGL LSGLRVMSVDLGLRTSASISVFRVARKDELKPNSEGR VPFCFPIEGNENLVAVHERSQLLKLPGETESKDLRAIR EERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWA KLIEQPMDANQMTPDWREAFEDELQKLKSLYGICGD REWTEAVYESVRRVWRHMGKQVRDWRKDVRSGER PKIRGYQKDVVGGNSIEQIEYLERQYKFLKSWSFFGK VSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRII MEALGYVYALDDERGKGKWVAKYPPCQLILLEELSE YQFNNDRPPSENNQLMQWSHRGVFQELLNQAQVHD LLVGTMYAAFSSRFDARTGAPGIRCRRVPARCAREQN PEPFPWWLNKFVAEHKLDGCPLRADDLIPTGEGEFFV SPFSAEEGDFHQIHADLNAAQNLQRRLWSDFDISQIRL RCDWGEVDGEPVLIPRTTGKRTADSYGNKVFYTKTG VTYYERERGKKRRKVFAQEELSEEEAELLVEADEARE KSVVLMRDPSGIINRGDWTRQKEFWSMVNQRIEGYL VKQIRSRVRLQESACENTGDIKRPAATKKAGQAKKK

[2087]

	KGSYPYDVPDYAYPYDVPDYAYPYDVPDYA*
px602-bh-optimize-AAV (SEQ ID NO:654)	cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcga cctttggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaa ctccatcactaggggttcctgcggcctctagactcgaggggctggaagctacctttga catcatttcctctgcgaatgcatgtataatttctacagaacctattagaaaggatcaccca gcctctgcttttgtacaactttcccttaaaaaactgccaattccactgctgtttggcccaat agtgagaacttttttcctgctgcctcttggtgcttttgcctatggcccctattctgcctgctg aagacactcttgccagcatggacttaaaccccctccagctctgacaatcctctttctctttt gttttacatgaagggtctggcagccaaagcaatcactcaaagttcaaaccttatcatttttt tgctttgttcctcttggccttggttttgtacatcagctttgaaaataccatcccagggttaat gctggggttaatttataactaagagtgctctagttttgcaatacaggacatgctataaaa atggaaagataccggtgccacAAGAGGTAAGGGTTTAAGGGAT GGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGG AGCACCTGCCTGAAATCACTTTTTTTCAGGTTGGcatg gccccaaagaagaagcggaaggtcggtatccacggagtcccagcagccgccacc agatccttcatcctgaagatcgagcccaacgaggaagtgaagaaaggcctctggaa aacccacgaggtgctgaaccacggaatcgcctactacatgaatatcctgaagctgat ccggcaagaggccatctacgagcaccacgagcaggaccccaagaatcccaagaa ggtgtccaaggccgagatccaggccgagctgtgggatttcgtgctgaagatgcaga agtgcaacagcttcacacacgaggtggacaaggacgaggtgttcaacatcctgaga gagctgtacgaggaactggtgcccagcagcgtggaaaagaagggcgaagccaac cagctgagcaacaagtttctgtaccctctggtggaccccaacagccagtctggaaag ggaacagccagcagcggcagaaagcccagatggtacaacctgaagattgccggc gatccctcctgggaagaagagaagaagaagtgggaagaagataagaaaaaggac ccgctggccaagatcctgggcaagctggctgagtacggactgatccctctgttcatcc cctacaccgacagcaacgagcccatcgtgaaagaaatcaagtggatggaaaagtcc cggaaccagagcgtgcggcggctggataaggacatgttcattcaggccctggaacg gttcctgagctgggagagctggaacctgaaagtgaaagaggaatacgagaaggtcg agaaagagtacaagaccctggaagagaggatcaaagaggacatccaggctctgaa ggctctggaacagtatgagaaagagcggcaagaacagctgctgcgggacaccctg aacaccaacgagtaccggctgagcaagagaggccttagaggctggcgggaaatca tccagaaatggctgaaaatggacgagaacgagccctccgagaagtacctggaagtg ttcaaggactaccagcggaagcaccctagagaggccggcgattacagcgtgtacga gttcctgtccaagaaagagaaccacttcatctggcggaatcaccctgagtaccctac ctgtacgccaccttctgcgagatcgacaagaaaaagaaggacgccaagcagcagg ccaccttcacactggccgatcctatcaatcaccctctgtgggtccgattcgaggaaag aagcggcagcaacctgaacaagtacagaatcctgaccgagcagctgcacaccgag aagctgaagaaaaagctgacagtgcagctggaccggctgatctaccctacagaatct ggcggctgggaagagaagggcaaagtggacattgtgctgctgcccagccggcagt tctacaaccagatcttcctggacatcgaggaaaagggcaagcacgccttcacctaca aggatgagagcatcaagttccctctgaagggcacactcggcggagccagagtgca gttcgacagagatcacctgagaagataccctcacaaggtggaaagcggcaacgtgg gcagaatctacttcaacatgaccgtgaacatcgagcctacagagtccccagtgtccaa gtctctgaagatccaccgggacgacttccccaaggtggtcaacttcaagcccaaaga actgaccgagtggatcaaggacagcaagggcaagaaactgaagtccggcatcgag tccctggaaatcggcctgagagtgatgagcatcgacctgggacagagacaggccg

[2088]

	ctgccgcctctatttttcgaggtggtggatcagaagcccgacatcgaaggcaagctgtt tttcccaatcaagggcaccgagctgtatgccgtgcacagagccagcttcaacatcaa gctgcccggcgagacactggtcaagagcagagaagtgctgcggaaggccagaga ggacaatctgaaactgatgaaccagaagctcaacttcctgcggaacgtgctgcacttc cagcagttcgaggacatcaccgagagagagaagcgggtcaccaagtggatcagca gacaagagaacagcgacgtgcccctggtgtaccaggatgagctgatccagatccgc gagctgatgtacaagccttacaaggactgggtcgccttcctgaagcagctccacaag agactggaagtcgagatcggcaaagaagtgaagcactggcggaagtccctgagcg acggaagaaagggcctgtacggcatctccctgaagaacatcgacgagatcgatcgg acccggaagttcctgctgagatggtccctgaggcctaccgaacctggcgaagtgcgt agactggaacccggccagagattcgccatcgaccagctgaatcacctgaacgccct gaaagaagatcggctgaagaagatggccaacaccatcatcatgcacgccctgggct actgctacgacgtgcggaagaagaaatggcaggctaagaaccccgcctgccagat catcctgttcgaggatctgagcaactacaaccctacgaggaaaggtcccgcttcga gaacagcaagctcatgaagtggtccagacgcgagatccccagacaggttgcactgc agggcgagatctatggcctgcaagtgggagaagtgggcgctcagttcagcagcag attccacgccaagacaggcagccctggcatcagatgtagcgtcgtgaccaaagaga agctgcaggacaatcggttcttcaagaatctgcagagagagggcagactgaccctg gacaaaatcgccgtgctgaaagagggcgatctgtacccagacaaaggcggcgaga agttcatcagcctgagcaaggatcggaagtgcgtgaccacacacgccgacatcaac gccgctcagaacctgcagaagcggttctggacaagaacccacggcttctacaaggt gtactgcaaggcctaccaggtggacggccagaccgtgtacatccctgagagcaagg accagaagcagaagatcatcgaagagttcggcgagggctacttcattctgaaggac ggggtgtacgaatgggtcaacgccggcaagctgaaaatcaagaagggcagctcca agcagagcagcagcgagctggtggatagcgacatcctgaaagacagcttcgacctg gcctccgagctgaaaggcgaaaagctgatgctgtacagggaccccagcggcaatgt gttccccagcgacaaatggatggccgctggcgtgttcttcggaaagctggaacgcat cctgatcagcaagctgaccaaccagtactccatcagcaccatcgaggacgacagca gcaagcagtctatgaaaaggccggcggccacgaaaaaggccggccaggcaaaaa agaaaaagggatcctacccatacgatgttccagattacgcttatccctacgacgtgcct gattatgcatacccatatgatgtccccgactatgcctaagaattcctagagctcgctgat cagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgcctt ccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgca tcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagca agggggaggattgggaagagaatagcaggcatgctggggacggccaaaaaaaga gaccatatatggtctccgtgctaatgcctcgtaagagacatcgtccagcaataggagtt tctcacaccctgcagcacttatagctagacggttgtcctgaccaaaagacagaaccg gtgtttcgtcctttccacaagatatataaagccaagaaatcgaaatactttcaagttacg gtaagcatatgatagtccattttaaaacataatttttaaaactgcaaactacccaagaaatt attactttctacgtcacgtattttgtactaatatctttgtgtttacagtcaaattaattccaatt atctctctaacagccttgtatcgtatatgcaaatatgaaggaatcatgggaaataggcc ctcgcggccgcaggaacccctagtgatggagttggccactccctctctgcgcgctcg ctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggctttgcccgg gcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctgatgcgg tattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagta cgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtga

[2089]

	ccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgc cacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgat ttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtg ggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaata gtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgattt ataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaattt aacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctg ctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgc cctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgg gagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagg gcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtca ggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattc aaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaag gaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgcct tcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgg gtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagtttt cgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggt attatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcag aatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacag taagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttactt ctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggat catgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgac gagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaact ggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggata aagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaat ctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggt aagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaa cgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcag accaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatcta ggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccact gagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgc gtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccgg atcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatacc aaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcac cgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataag tcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcg ggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccg aactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggaga aaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgag ggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctct gacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacg ccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgt
px602-bv-optimize-AAV (SEQ ID NO:655)	cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcga cctttggtcgcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaa ctccatcactaggggttcctgcggcctctagactcgaggggctggaagctacctttga

[2090]

	catcatttcctctgcgaatgcatgtataatttctacagaacctattagaaaggatcaccca gcctctgcttttgtacaactttcccttaaaaaactgccaattccactgctgtttggcccaat agtgagaacttttttcctgctgcctcttggtgcttttgcctatggcccctattctgcctgctg aagacactcttgccagcatggacttaaacccctccagctctgacaatcctctttctctttt gttttacatgaagggtctggcagccaaagcaatcactcaaagttcaaaccttatcatttttt tgctttgttcctcttggccttggttttgtacatcagctttgaaaataccatcccagggttaat gctggggttaatttataactaagagtgctctagttttgcaatacaggacatgctataaaaa atggaaagataccggtgccaccatggccccaaagaagaagcggaaggtcggtatc cacggagtcccagcagccgccatccggtccatcaagctgaagatgaagaccaaca gcggcaccgacagcatctacctgagaaaagccctgtggcggacccaccagctgatc aatgagggaatcgcctactacatgaacctgctgaccctgtaccggcaagaggccatc ggcgacaagaccaaagaagcctatcaggccgagctgattaacatcatccggaacca gcagcggaacaacggcagctctgaggaacacggctccgaccaagaaattctggcc ctgctgagacagctgtacgagctgatcatccccagcagcatcggcgaatctggcgac gctaatcagctgggcaacaagtttctgtaccctctggtggaccccaacagccagtctg gcaagggcacatctaacgccggcagaaagcccagatggaagcggctgaaagagg aaggcaaccccgactgggaactcgagaagaagaaggacgaggaacgcaaggcc aaggatcccaccgtgaagatctttgacaacctgaacaaatacggcctgctgcctctgtt cccactgttcaccaacatccagaaagacatcgagtggctgcccctgggcaagagac agtctgtgcggaagtgggacaaagacatgttcatccaggccatcgagagactgctga gctgggagagctggaacagaagagtggccgacgagtacaaacagctgaaagaaa agaccgagagctactacaaagagcacctgacaggcggcgaggaatggatcgaga agatccggaagttcgagaaagaacggaacatggaactggaaaagaacgccttcgct cccaacgacggctacttcatcaccagcagacagatcagaggctgggacagagtgta cgagaagtggtccaagctgcccgagtctgctagccctgaggaactgtggaaagtgg tggccgagcagcagaacaagatgtccgaaggcttcggcgaccccaaggtgttcagc ttcctggccaacagagagaaccgggacatttggagaggccacagcgagcggatcta ccacattgccgcctacaacggcctgcagaagaagctgagccggaccaaagagcag gccaccttcacactgcctgacgccattgaacaccctctgtggatcagatacgagagc cctggcggcaccaacctgaatctgttcaagctggaagagaaacagaaaaagaacta ctacgtgaccctgagcaagatcatctggcccagcgaggaaaagtggattgagaaag agaacatcgagatccctctggctcccagcatccagttcaaccggcagattaagctga agcagcacgtgaagggcaagcaagagatcagcttcagcgactacagcagccggat cagcctggatggtgttctcggcggcagcagaatccagtttaatcggaagtacatcaag aaccacaaagagctgctcggagagggcgacatcggccccgtgttctttaacctggtg gtggatgtggcccctctgcaagaaaccagaaacggcagactgcagagccccatcg gcaaggccctgaaagtgatcagcagcgacttctccaaagtgatcgactacaagccg aaagaactcatggattggatgaataccggcagcgccagcaacagctttggagtggct tctctgctggaaggcatgagagtgatgagcatcgacatgggccagagaaccagcgc ctccgtgtccatcttcgaggtcgtgaaagaactgcccaaggatcaagagcagaagct gttctacagcatcaacgacaccgagctgttcgccatccacaagcggagctttctgctg aacctgcctggcgaggtggtcaccaagaacaacaagcagcagcggcaagagcgg cggaaaaagcggcagtttgtgcggagccagatcagaatgctggccaacgtgctgcg gctggaaacaaagaaaaccccctgacgagcggaagaaggccattcacaagctgatg gaaatcgtgcagagctacgacagctggaccgccagccagaaagaagtgtgggaga aagagctgaatctcctgaccaacatggccgccttcaatgacgagatctggaaagaaa

[2091]

	gcctggtggaactgcaccaccggatcgagccttacgtgggacagatcgtgtccaagt ggcggaagggcctgtctgagggcagaaagaatctggccggcatcagcatgtggaa catcgacgaactggaagataccaggcggctgctgatttcctggtccaagagaagca gaaccccaggcgaggccaacaggatcgaaaccgatgagcctttcggcagcagcct gctccagcacattcagaacgtgaaggacgacagactgaagcagatggccaacctga tcatcatgacagccctgggctttaagtacgacaaagaggaaaaggaccggtacaag cggtggaaagagacataccccgcctgccagatcatcctgttcgagaacctgaaccg ctacctgttcaacctcgaccggtccagacgcgagaacagcagactgatgaagtggg cccatcggagcatccccagaaccgtgtctatgcagggcgagatgttcggcctgcaa gtgggcgacgttcggagcgagtacagctccagattccacgccaaaacaggcgccc ctggcatcagatgtcacgccctgactgaagaggatctgaaggccggcagcaacacc ctgaagagactgatcgaggacggcttcatcaatgagagcgagctggcctacctgaa gaagggcgatatcatccctagccaaggcggcgaactgttcgtgacactgtccaagc ggtacaagaaggacagcgacaacaacgagctgaccgtgatccacgccgacatcaa cgccgctcagaatctgcagaagcggttttggcagcaaaacagcgaggtgtacagag tgccctgtcagctggccagaatgggcgaagataagctgtacatccccaagagccag accgagacaatcaagaagtatttcggcaagggctccttcgtgaagaacaataccgaa caagaggtctacaagtgggagaagtccgagaaaatgaagatcaagacggacacca ccttcgacctgcaagacctggatggcttcgaggacatcagcaagaccattgagctgg cacaagagcagcaaaagaaatacctgaccatgttcagggaccccagcggctacttttt tcaacaatgagacatggcggcctcaaaaagaatactggtccatcgtgaacaacatcat caagagctgcctcaagaagaagatcctgagcaacaaggtcgagctgaaaaggccg gcggccacgaaaaaggccggccaggcaaaaaagaaaaagggatcctacccatac gatgttccagattacgcttatccctacgacgtgcctgattatgcatacccatatgatgtcc ccgactatgcctaagaattcctagagctcgctgatcagcctcgactgtgccttctagttg ccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactc ccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattc tattctgggggggtggggtggggcaggacagcaagggggaggattgggaagagaat agcaggcatgctggggacggccaaaaaaagagaccatatatggtctccgtgccaag cacctgttttcaggtgctcctgtggtgggtaatttttaattacttatggcacacgcacaga ttcattgaccctataggtccggtgtttcgtcctttccacaagatatataaagccaagaaat cgaaatactttcaagttacggtaagcatatgatagtccattttaaaacataatttaaaact gcaaactacccaagaaattattactttctacgtcacgtattttgtactaatatctttgtgttta cagtcaaattaattccaattatctctctaacagccttgtatcgtatatgcaaatatgaagg aatcatgggaaataggccctcgcggccgcaggaacccctagtgatggagttggcca ctccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccga cgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgc aggggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatac gtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcgggtgtgg tggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcg ctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggg gctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattt gggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgtt ggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctat ctcgggctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatg agctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatggt

[2092]

	gcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgcc aacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagaca agctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaa cgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataat aatggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctattt gtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatg cttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattc ccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaa gatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagc ggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaa gttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtc gccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcat cttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgata acactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctt ttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaat gaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaa cgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatag actggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctg gctggtttattgctgataaatctggagccggtgagcgtggaagccgcggtatcattgc agcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggag tcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgatt aagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttca tttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttа acgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttctt gagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccag cggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttca gcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccactt caagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctg ctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccgga taaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggag cgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccac gcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaac aggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgt cgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcgga gcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggcct tttgctcacatgt

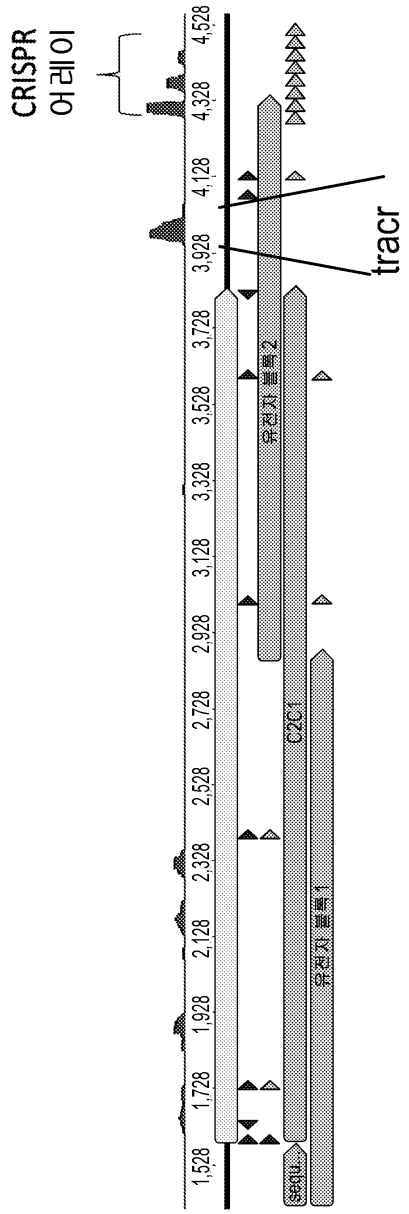
[2093]

[2094] * * *

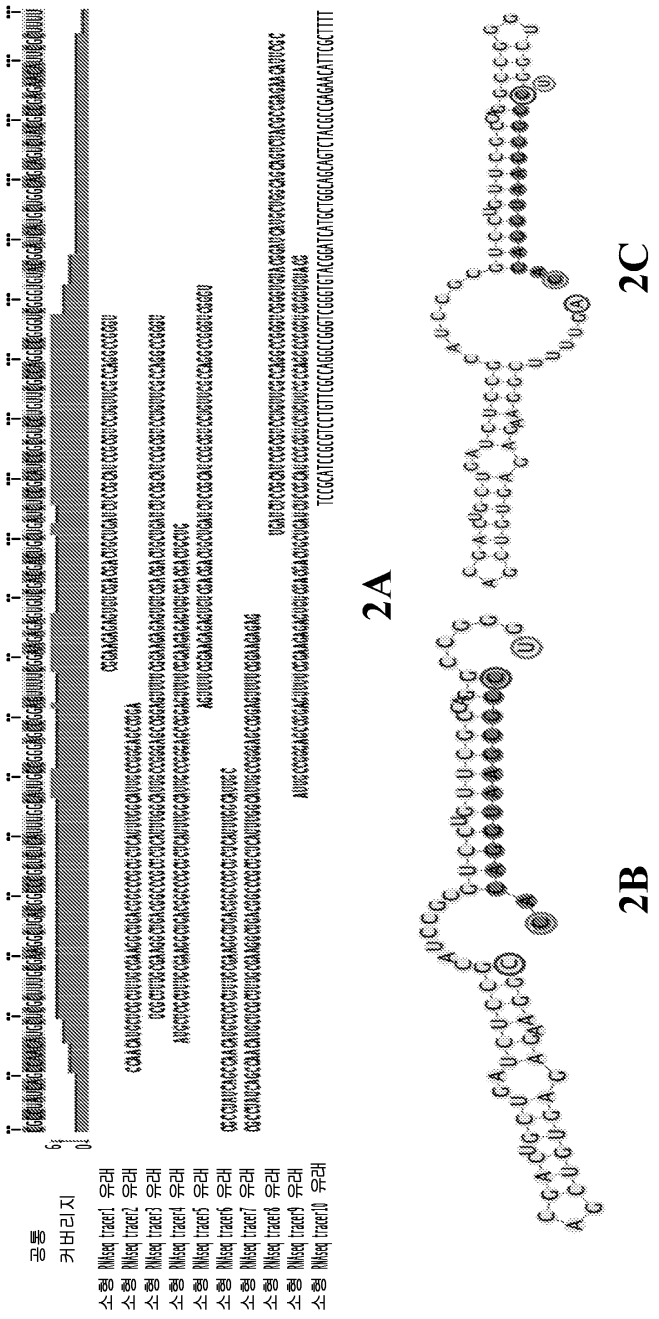
[2095] 본 발명의 기재된 방법, 약학 조성물 및 키트의 다양한 변형 및 이형은 본 발명의 범주 및 취지를 벗어나지 않고 당업자에게 분명해질 것이다. 본 발명이 특별한 구현예와 함께 기재되어 있지만, 더욱 변형될 수 있고 청구되는 본 발명이 이러한 특별한 구현예에 과도하게 제한되어서는 안된다는 것을 이해하게 될 것이다. 실제로, 당업자에게 자명한 본 발명을 수행하기 위해 기재된 방식의 다양한 변형은 본 발명의 범주 내에 포함되는 것으로 의도된다. 본 출원은 일반적으로 본 발명의 원리에 따른 본 발명의 임의의 변형, 용도, 또는 개조를 포괄하는 것으로 의도되고 본 개시물로부터의 이러한 이탈의 포함은 본 발명이 속하는 분야 내에서 공지의 통상적인 관례 내이며 이전에 기재된 본원의 본질적인 특성에 적용될 수 있다.

도면

도면1


CRISPR
어레이
4,528
4,328
4,128
3,928
3,728
3,528
3,328
3,128
2,928
2,728
2,528
2,328
2,128
1,928
1,728
1,528
tracr
유전자 블록 2
C2C1
유전자 블록 1
tracr

도면2

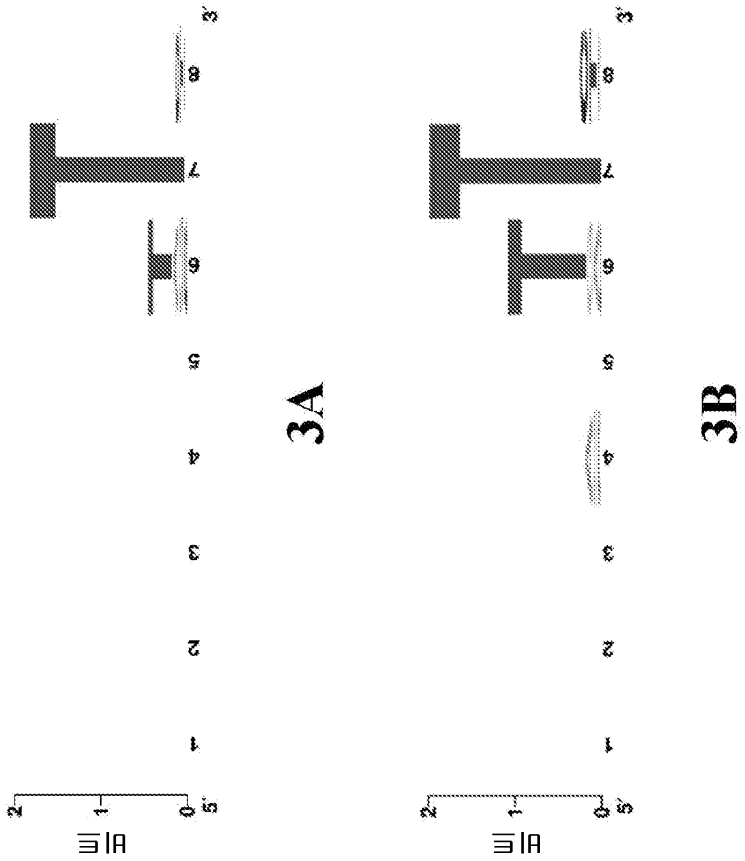

공통
커버리지
소형 NMseq trace1 유래
소형 NMseq trace2 유래
소형 NMseq trace3 유래
소형 NMseq trace4 유래
소형 NMseq trace5 유래
소형 NMseq trace6 유래
소형 NMseq trace7 유래
소형 NMseq trace8 유래
소형 NMseq trace9 유래
소형 NMseq trace10 유래

2A

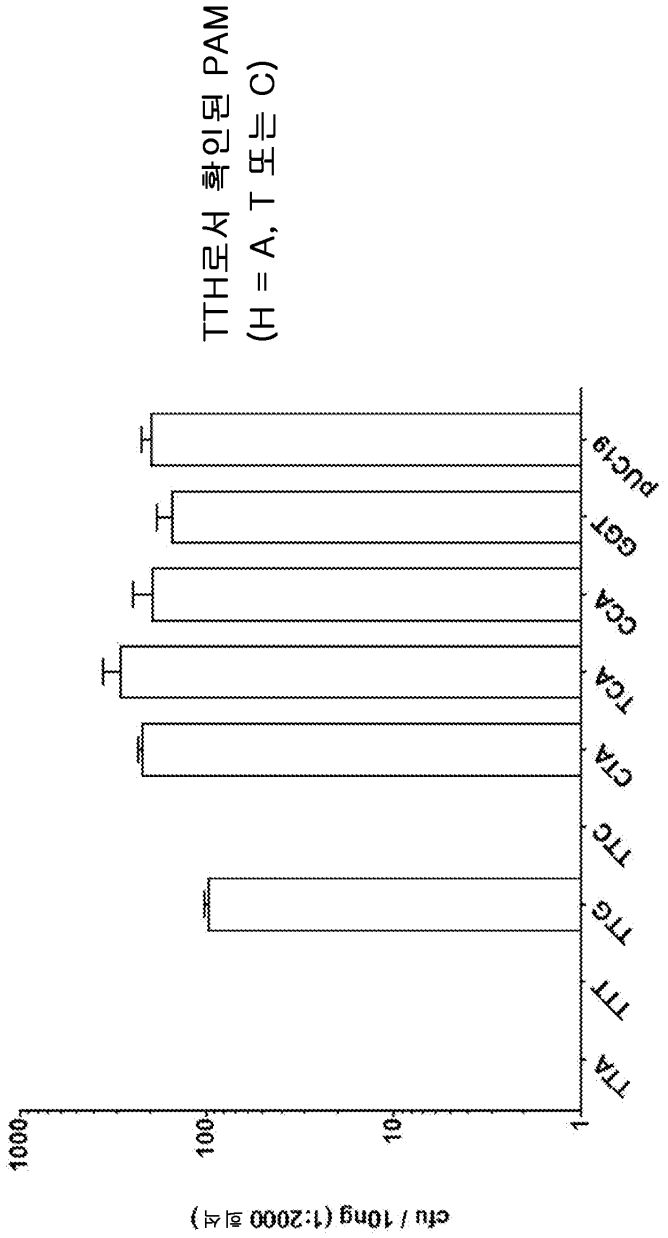
2B

2C

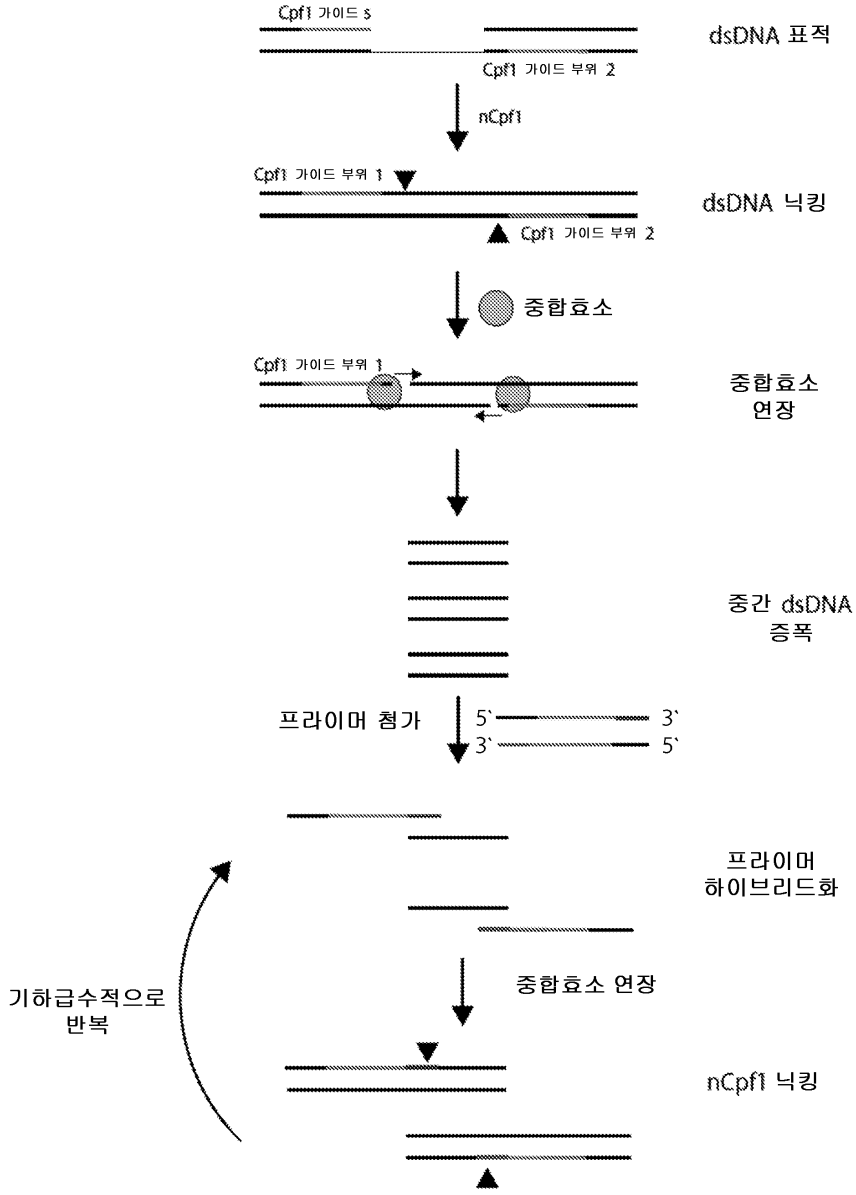
도면3



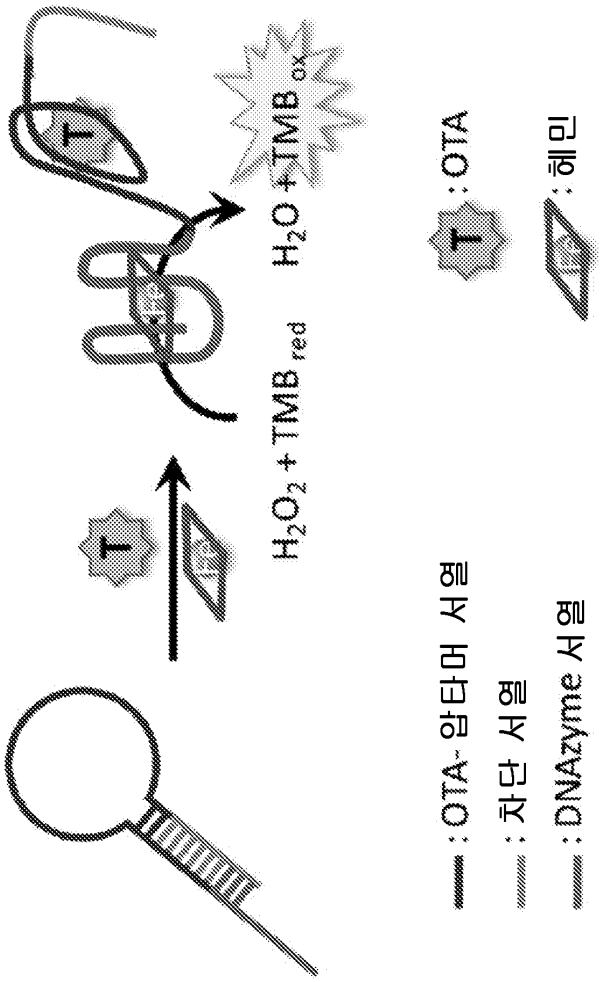
도면4


TTH로서 확인된 PAM
(H = A, T 또는 C)
cfu / 10ng (1:2000 희석)
1000
100
10
1
TTA
TTT
TTG
TTC
CTA
TCA
CCA
GGT
pUC19

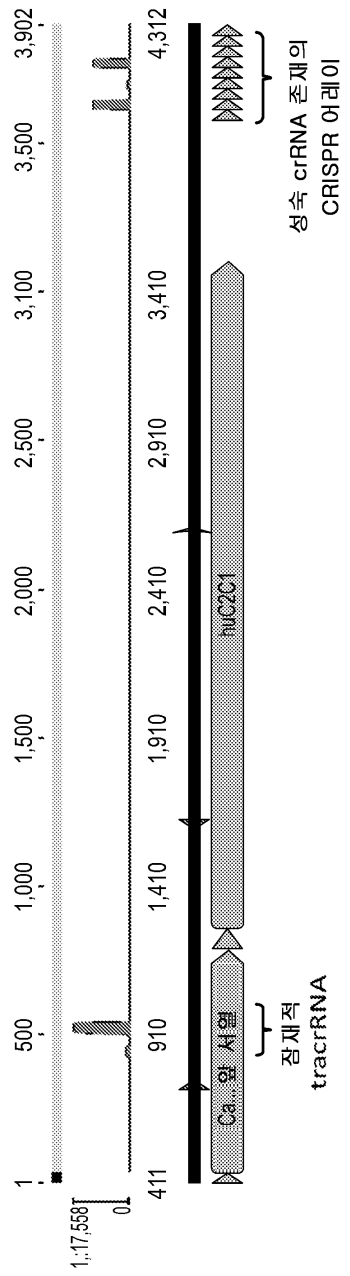
도면5


Cpf1 가이드 s
Cpf1 가이드 부위 2
dsDNA 표적
nCpf1
Cpf1 가이드 부위 1
dsDNA 닉킹
Cpf1 가이드 부위 2
중합효소
Cpf1 가이드 부위 1
중합효소
연장
중간 dsDNA
증폭
프라이머 첨가
5`
3`
3`
5`
프라이머
하이브리드화
중합효소 연장
nCpf1 닉킹
기하급수적으로
반복

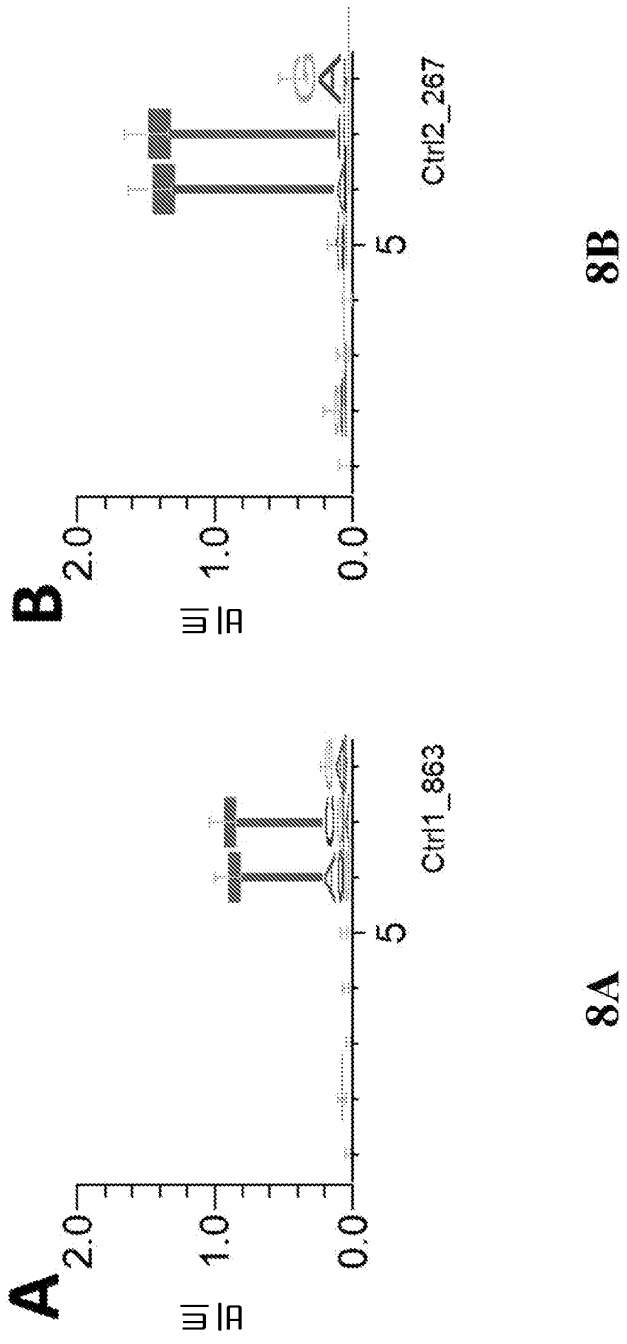
도면6



도면7


1
500
1,000
1,500
2,000
2,500
3,100
3,500
3,902
1:17,558
0
411
910
1,410
1,910
2,410
2,910
3,410
4,312
Ca... 핵 서열
huC2C1
잠재적
tracrRNA
성숙 crRNA 존재의
CRISPR 어레이

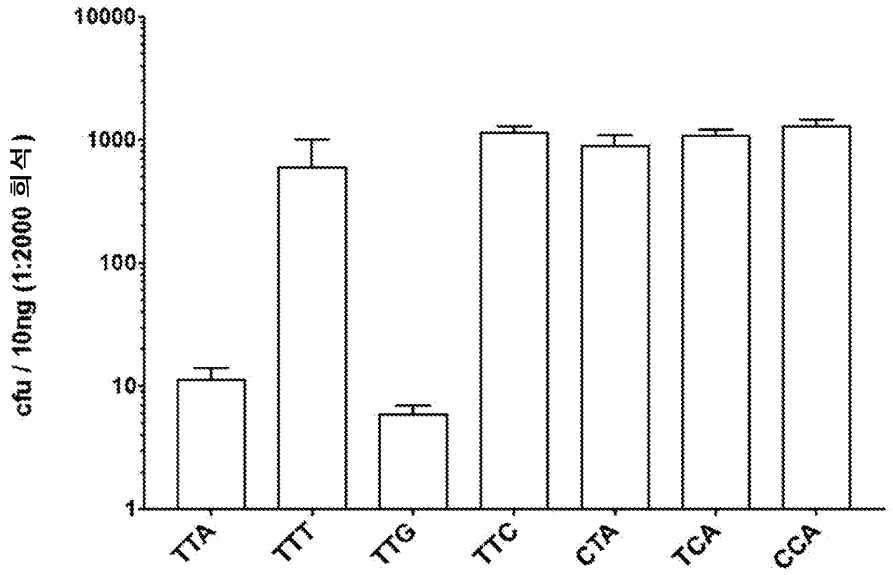
도면8



8A

8B

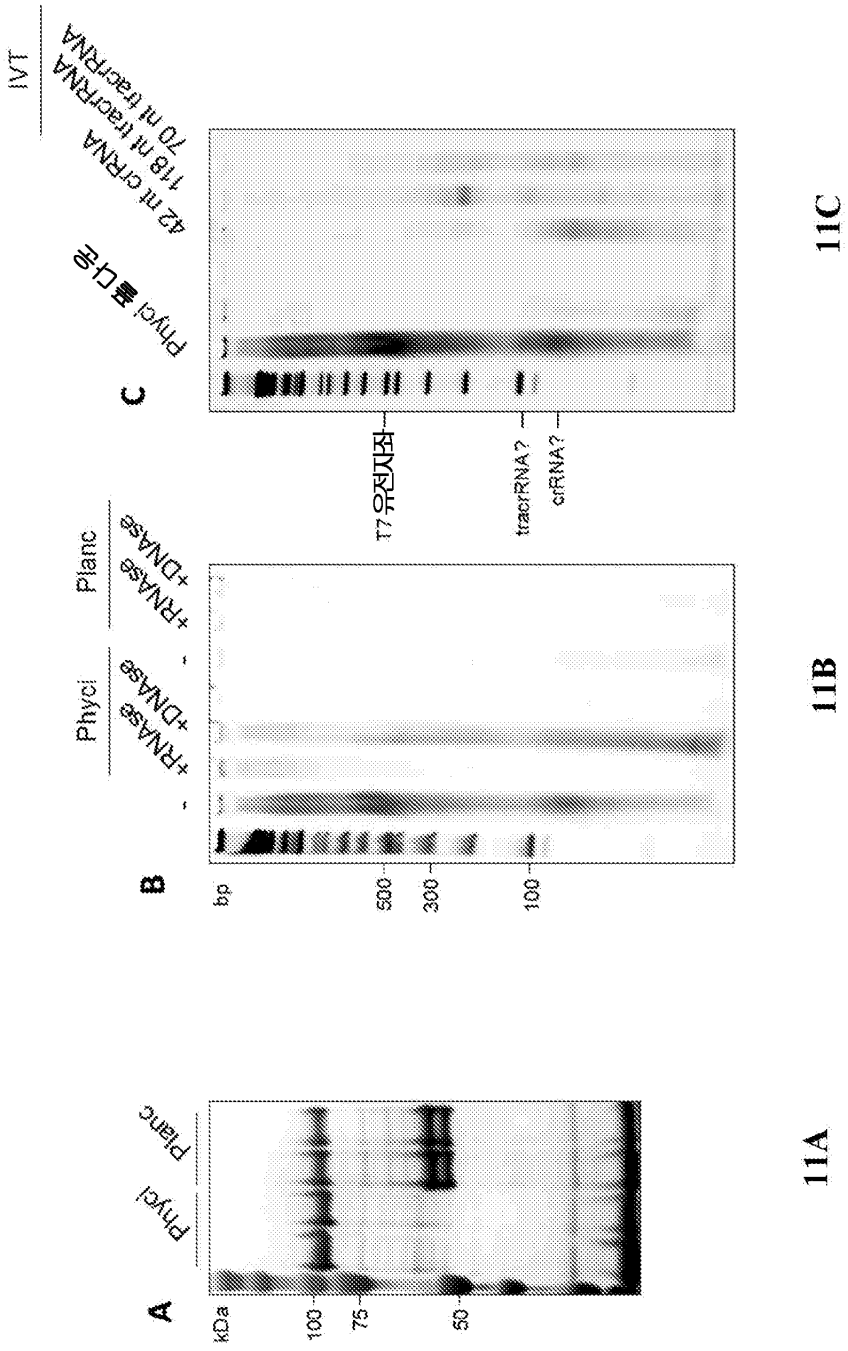
도면9


10000
1000
100
10
1
cfu / 10ng (1:2000 희석)
TTA
TTT
TTG
TTC
CTA
TCA
CCA

도면10

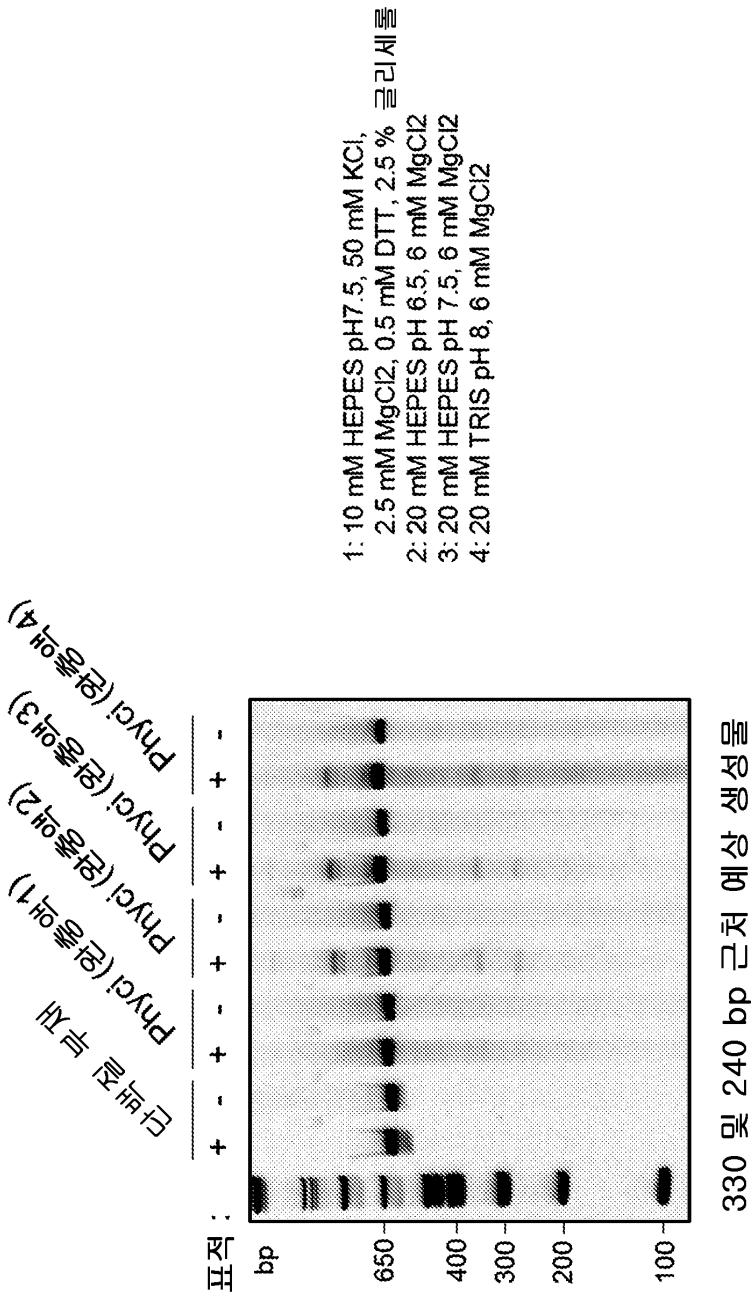

T7
SUMO
PhyciC2C1
T7
tracr 영역 & CRISPR 어레이

도면11

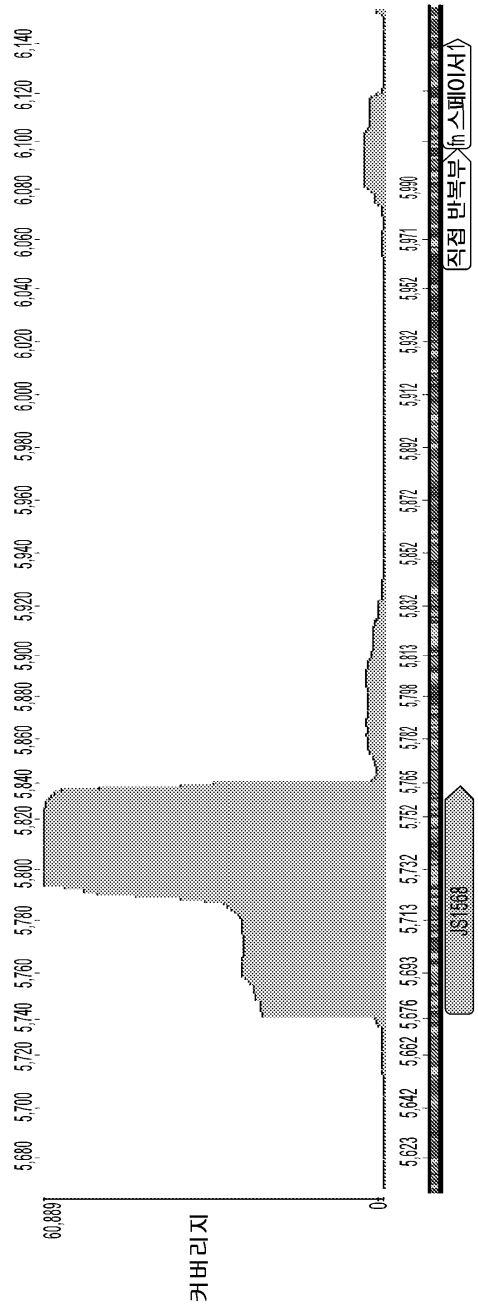


11A 11B 11C

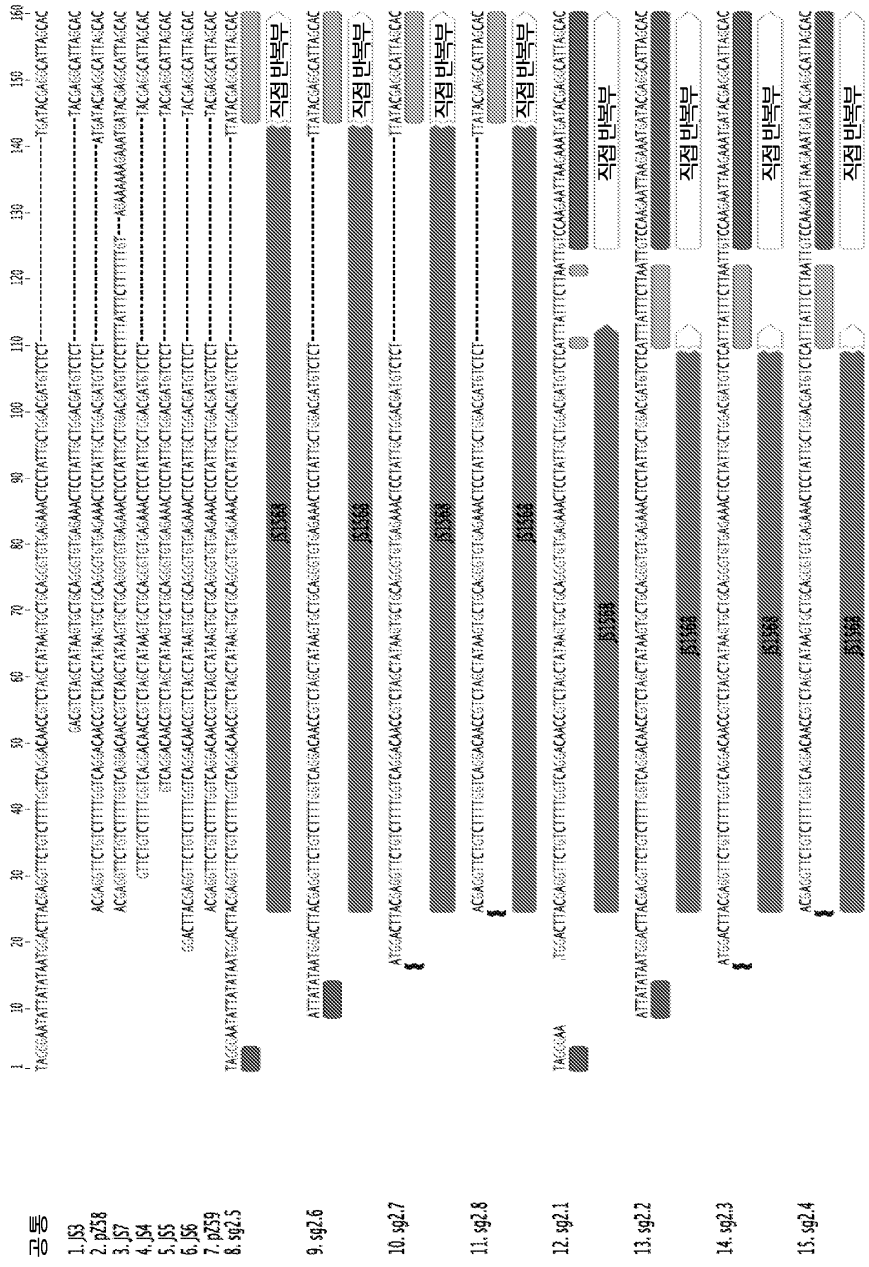
도면12


1: 10 mM HEPES pH7.5, 50 mM KCl,
2.5 mM MgCl2, 0.5 mM DTT, 2.5 % 글리세롤
2: 20 mM HEPES pH 6.5, 6 mM MgCl2
3: 20 mM HEPES pH 7.5, 6 mM MgCl2
4: 20 mM TRIS pH 8, 6 mM MgCl2
330 및 240 bp 근처 예상 생성물
bp
650
400
300
200
100
표적 :
단백질 부재
PhyCi (융통액1)
PhyCi (융통액2)
PhyCi (융통액3)
PhyCi (융통액4)

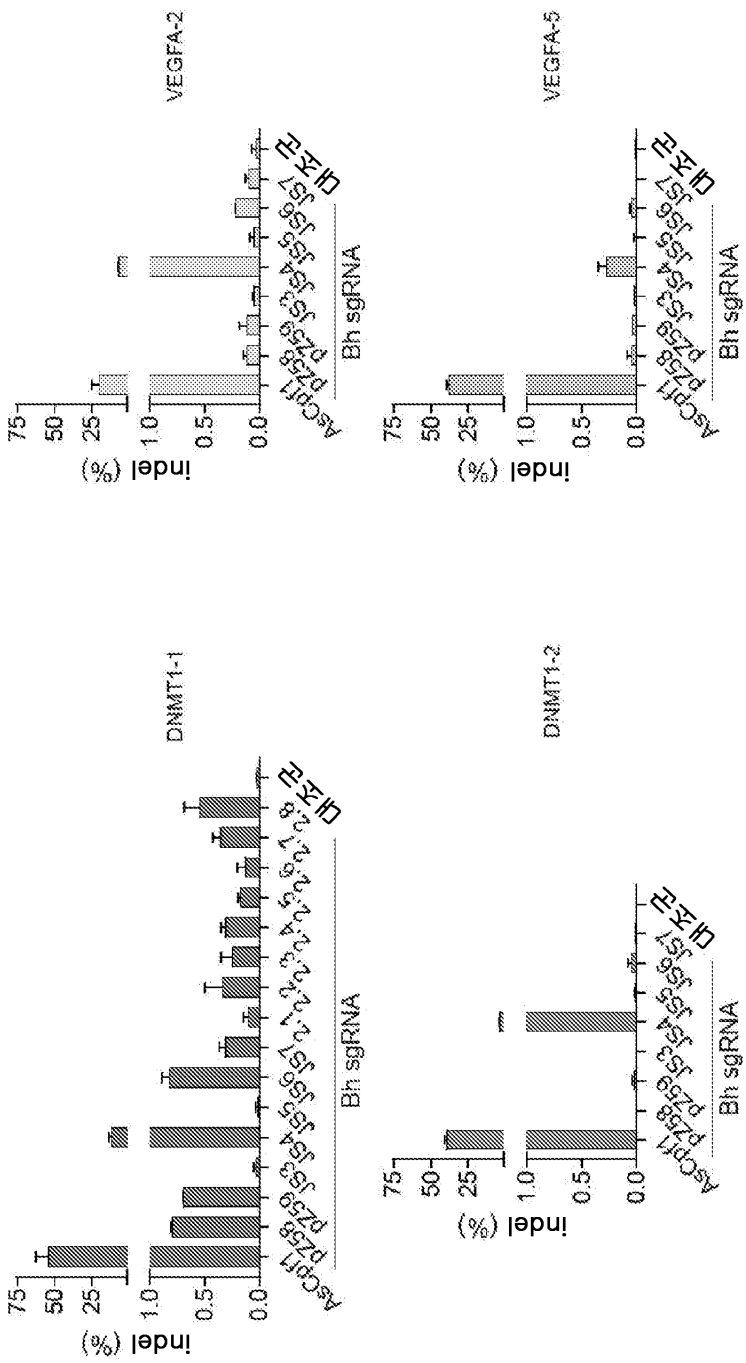
도면13a


커버리지
60,889
0
5,623
5,642
5,662
5,676
5,693
5,713
5,732
5,752
5,765
5,782
5,798
5,813
5,832
5,852
5,872
5,892
5,912
5,932
5,952
5,971
5,990
6,000
6,020
6,040
6,060
6,080
6,100
6,120
6,140
5,680
5,700
5,720
5,740
5,760
5,780
5,800
5,820
5,840
5,860
5,880
5,900
5,920
5,940
5,960
5,980
JS1658
직접 반복부
In 스페이서

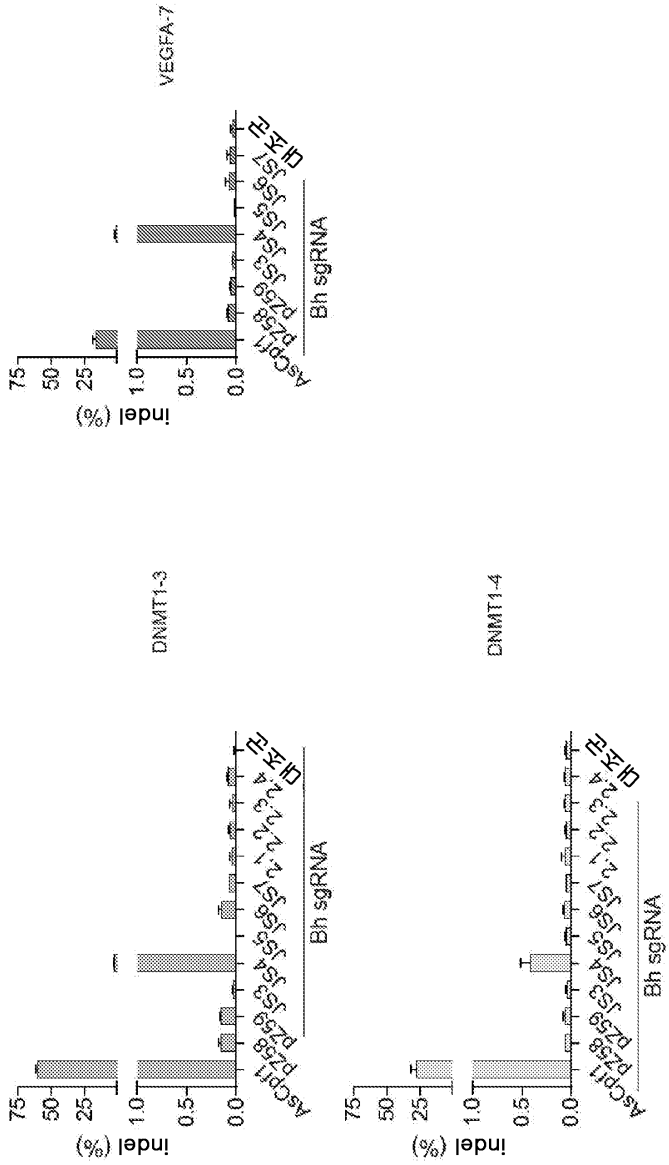
도면13b


공통
1. J53
2. pZ58
3. J57
4. J54
5. J55
6. J56
7. pZ59
8. sg2.5
9. sg2.6
10. sg2.7
11. sg2.8
12. sg2.1
13. sg2.2
14. sg2.3
15. sg2.4
직접 반복부

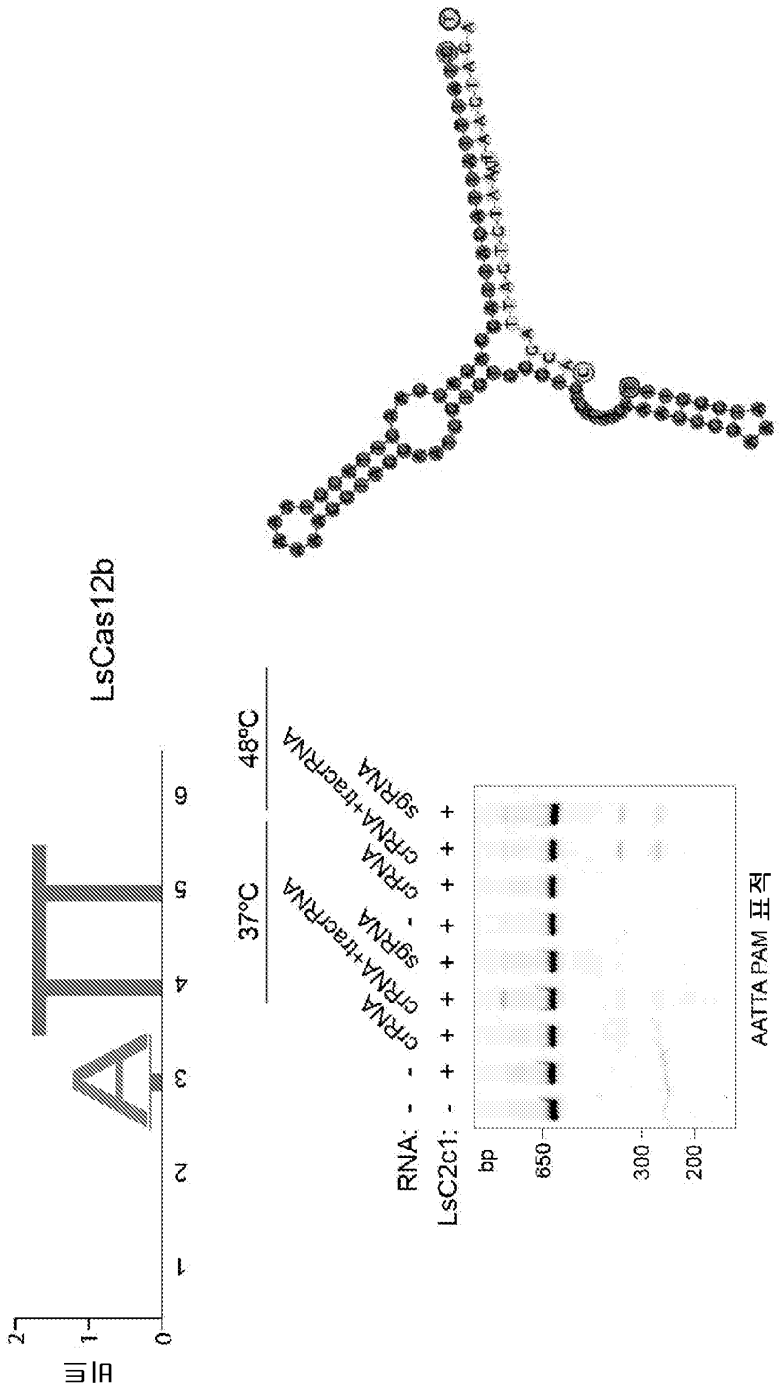
도면14a


VEGFA-2
VEGFA-5
DNMT1-1
DNMT1-2
Indel (%)
Bh sgRNA
AsCpf1
pD259
pD259 pS263
pD259 pS263 pS354
pD259 pS263 pS354 pS356
pD259 pS263 pS354 pS356 pS357
없음

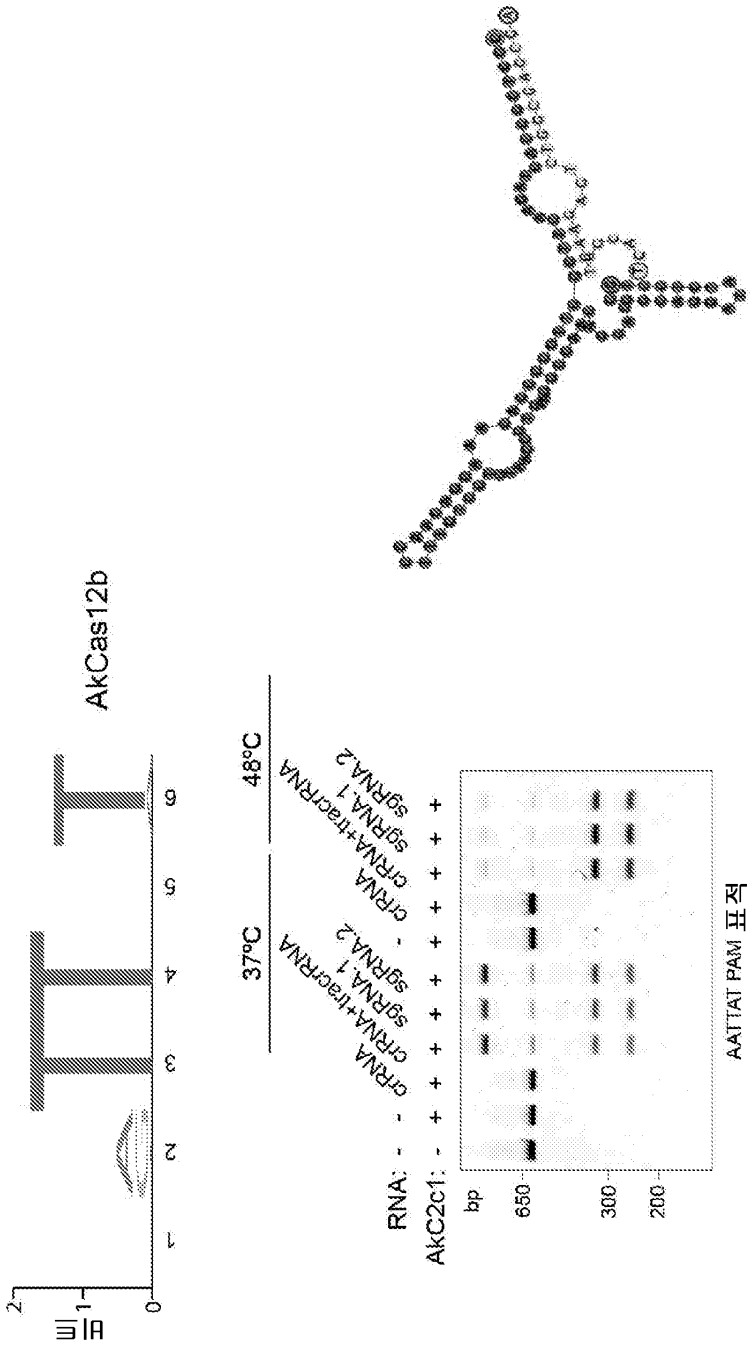
도면14b



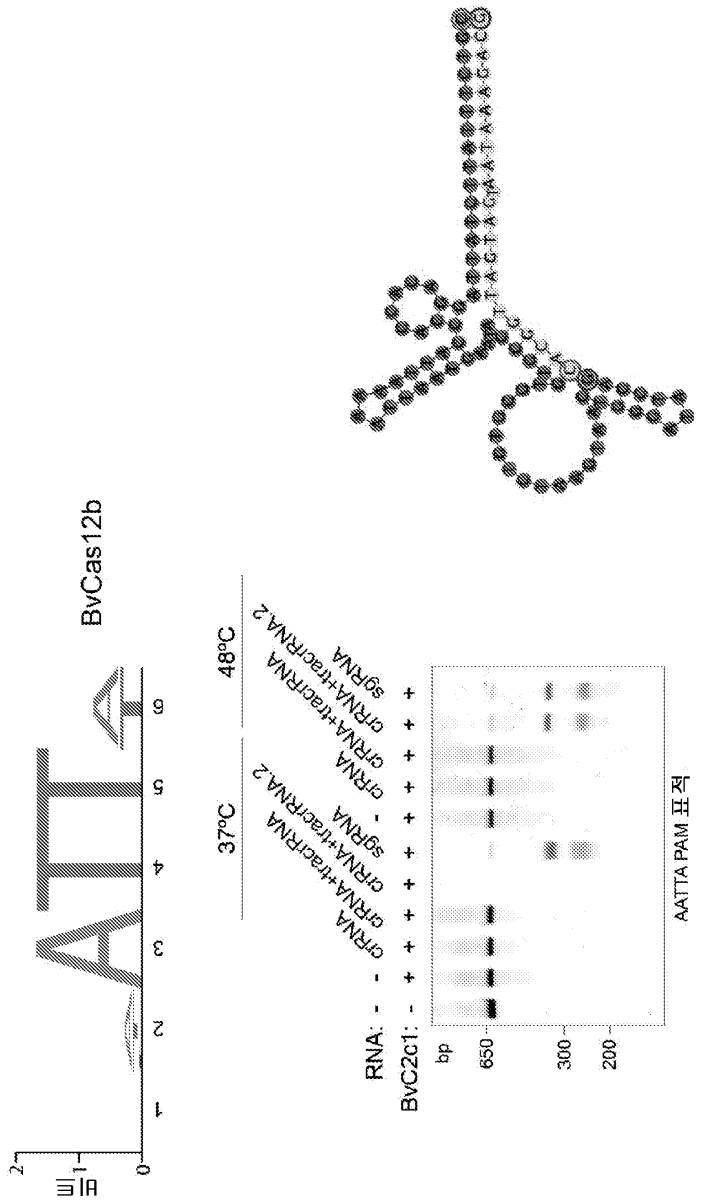
도면15a


LsCas12b
비율
0
1
2
3
4
5
6
37°C
48°C
RNA: - - crRNA crRNA+tracrRNA sgRNA crRNA crRNA+tracrRNA sgRNA
LsC2c1: - + + + + + + + +
bp
650
300
200
AATTA PAM 표적

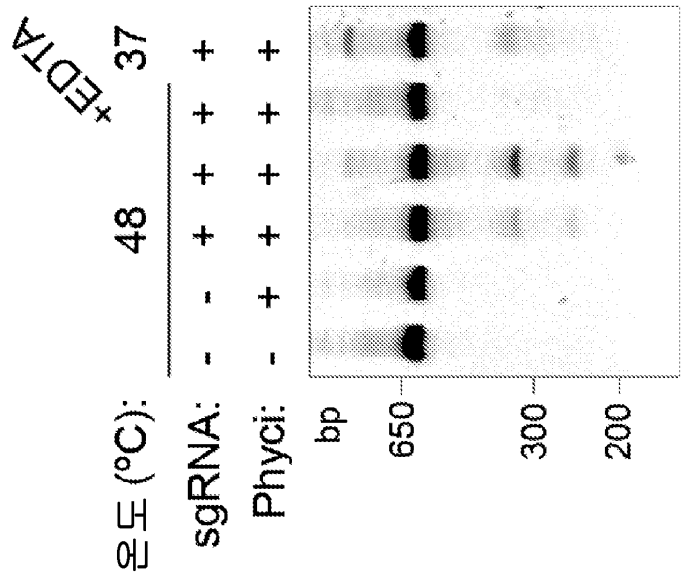
도면15b


AkCas12b
비율
0
1
2
1
2
3
4
5
6
AkC2c1: - + + + + + + + + +
RNA: - - crRNA crRNA+tracrRNA sgRNA 1 sgRNA 2 - crRNA crRNA+tracrRNA sgRNA 1 sgRNA 2
37°C
48°C
bp
650
300
200
AATTAT PAM 표적

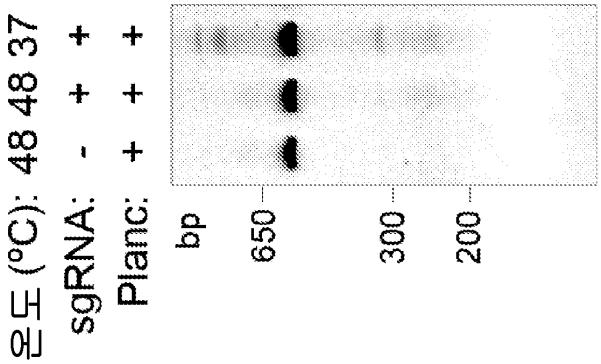
도면15c


BvCas12b
비율
0
1
2
1
2
3
4
5
6
AATTA PAM 표적
bp
650
300
200
BvC2c1: - + + + + + + + +
RNA: - - + + + + + + +
crRNA
crRNA+tracrRNA
crRNA+uacrRNA.2
sgRNA
37°C
crRNA
crRNA+tracrRNA
crRNA+tracrRNA.2
sgRNA
48°C

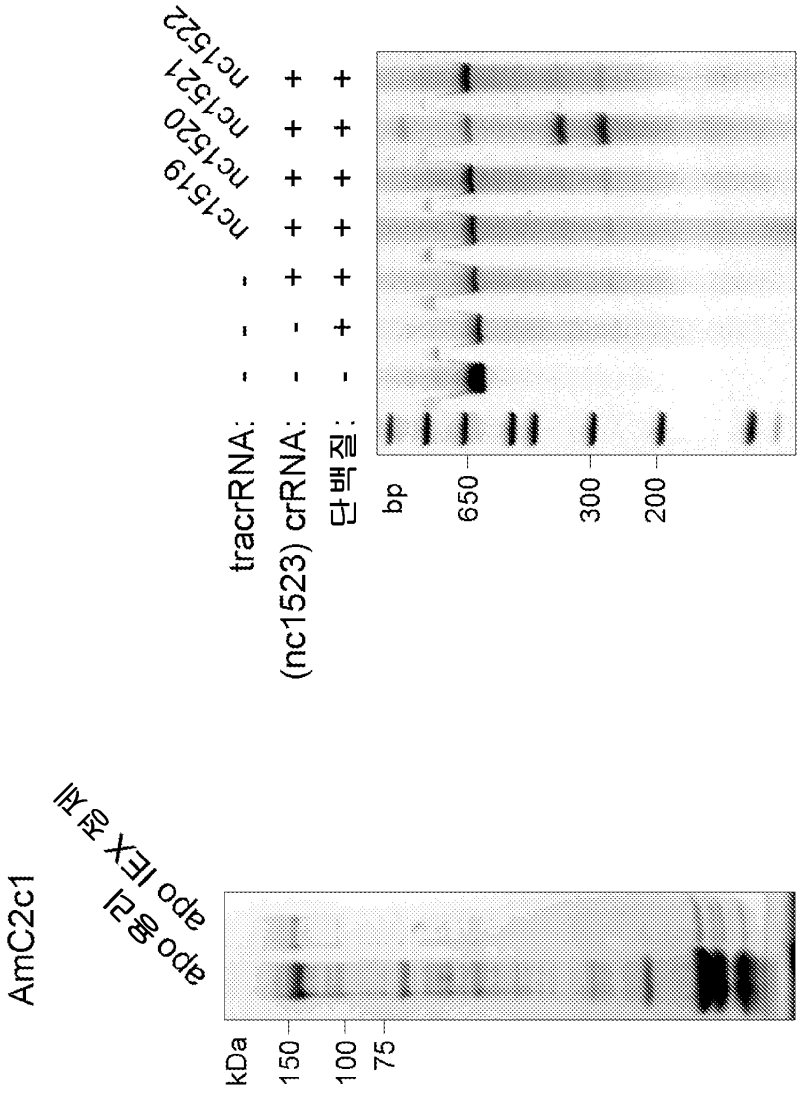
도면15d


온도 (°C):
sgRNA:
Phyci:
48
37
+EDTA
bp
650
300
200

도면15e


온도 (°C):
sgRNA:
Planc:
48
48
37
bp
650
300
200

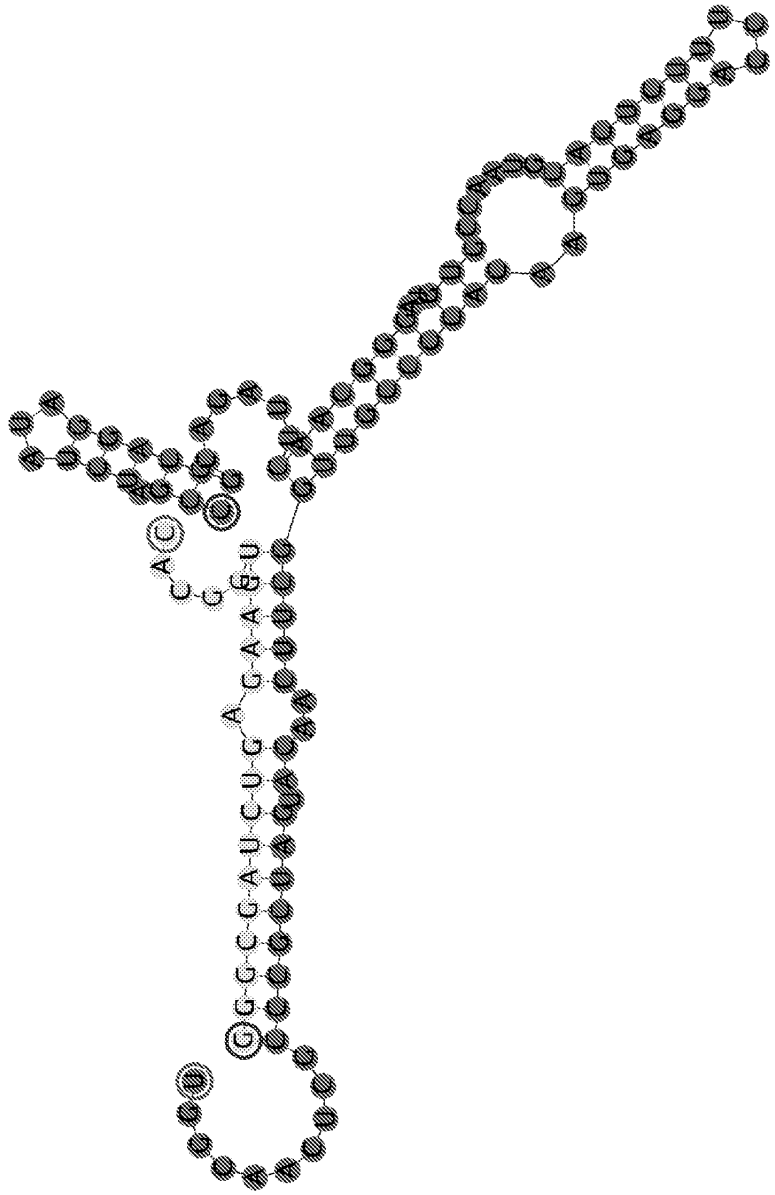
도면16


AmC2c1
apo 정제 IEX
apo 농축
kDa
150
100
75
tracrRNA: - - - + + + + +
(nc1523) crRNA: - - + + + + + +
단백질: - + + + + + + +
nc1519
nc1520
nc1521
nc1522
bp
650
300
200

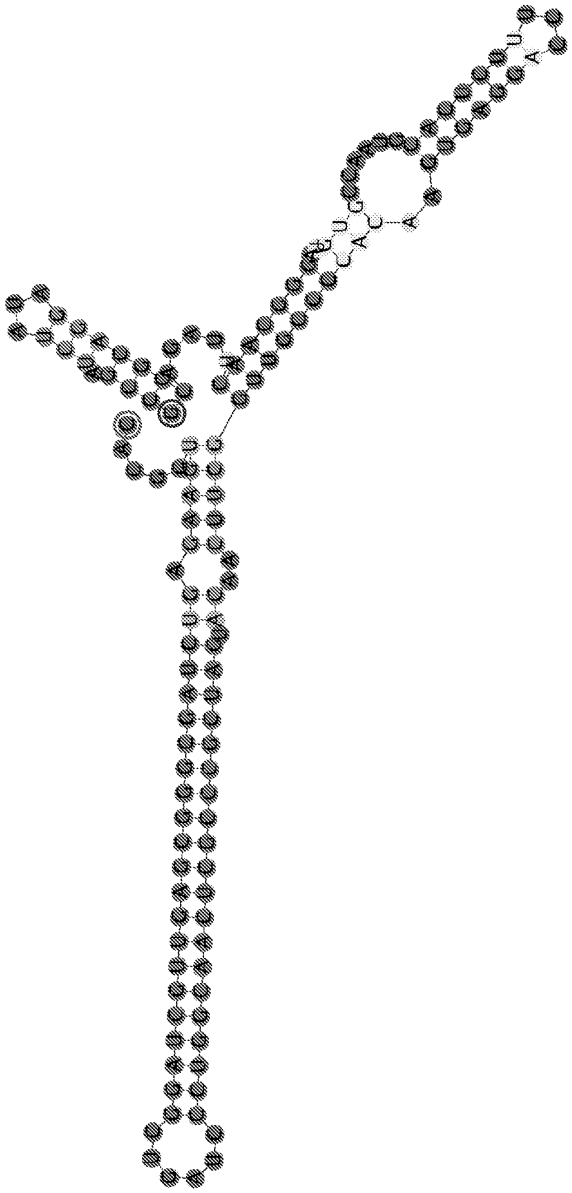
도면17a



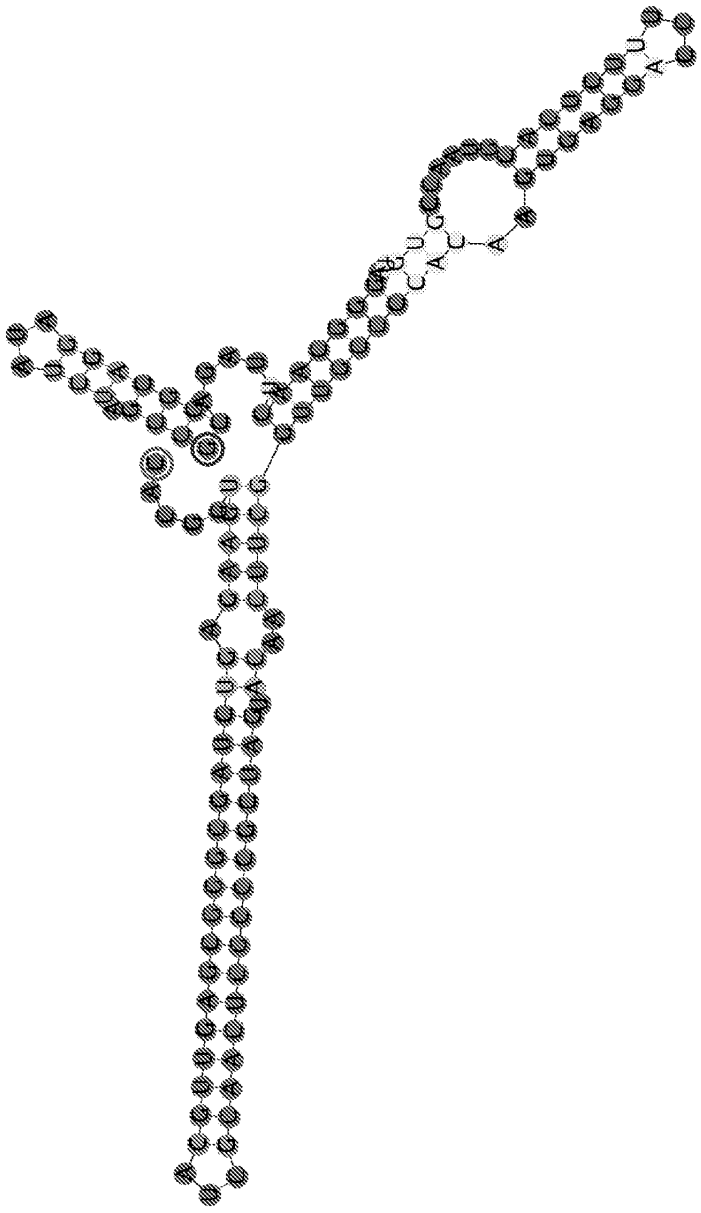
도면17b



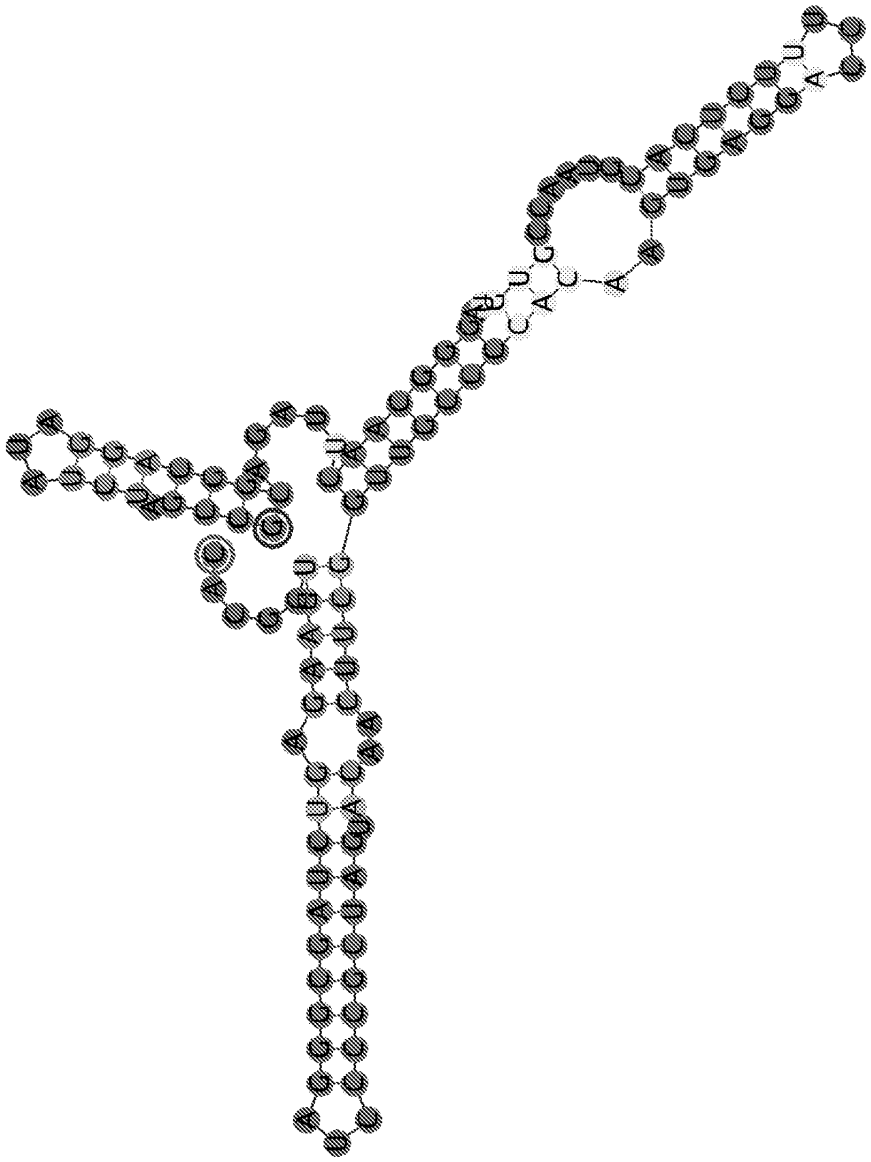
도면17c



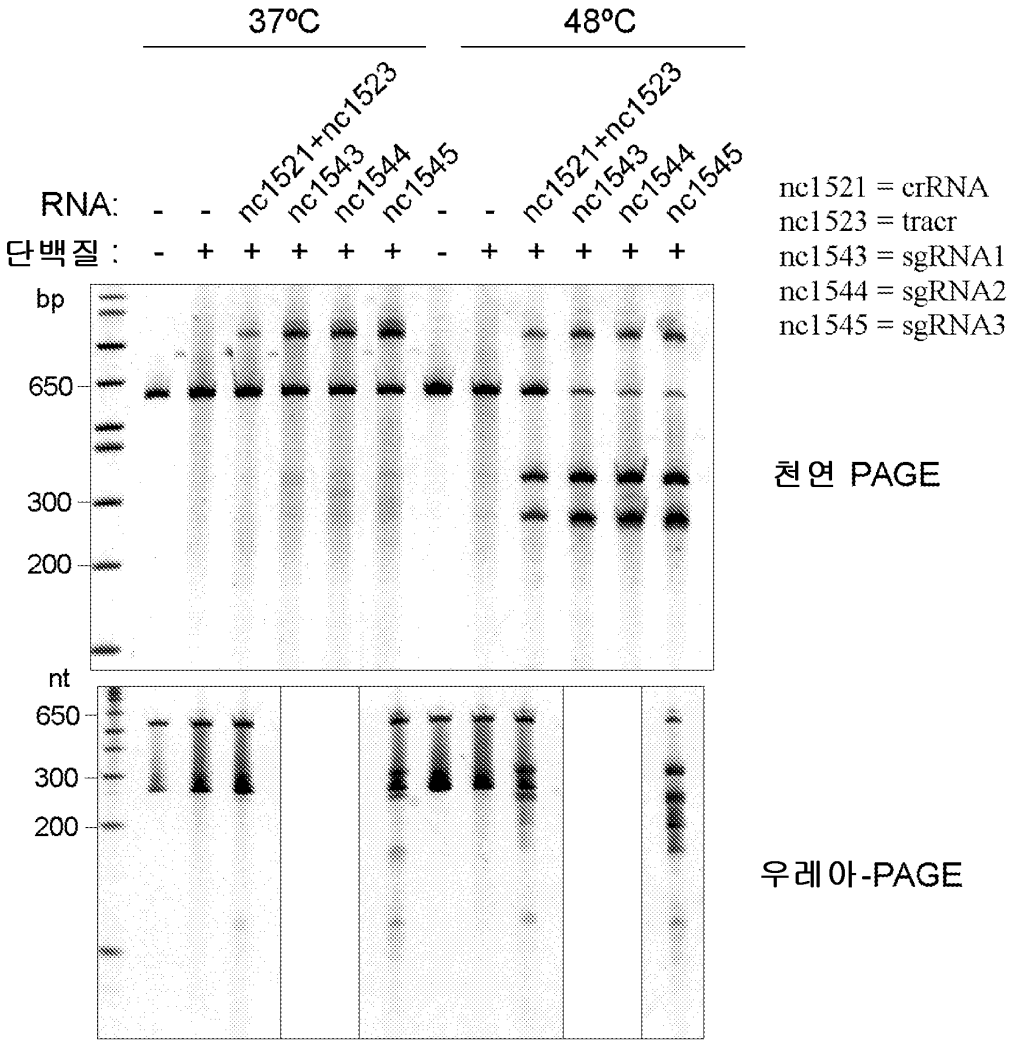
도면17d



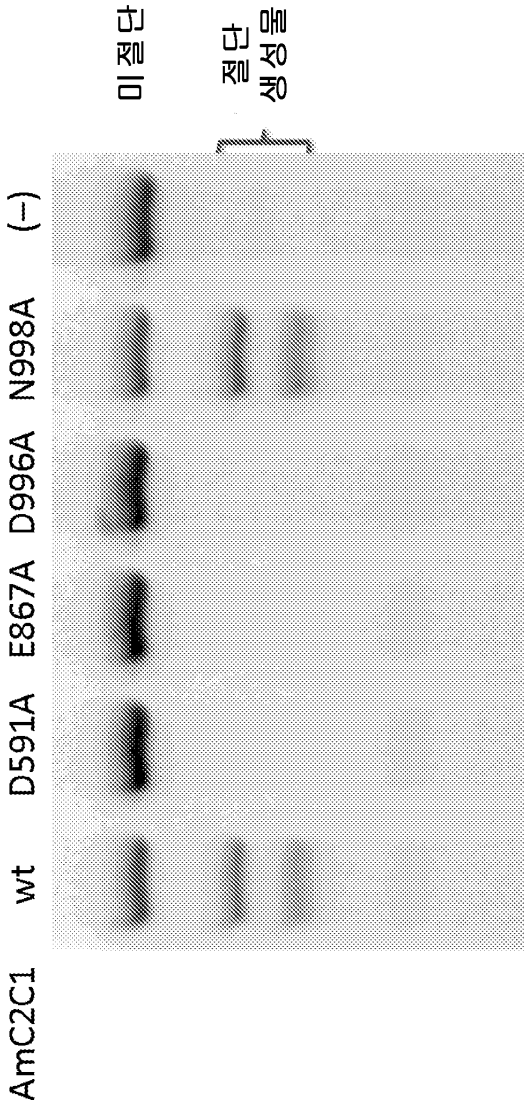
도면17e



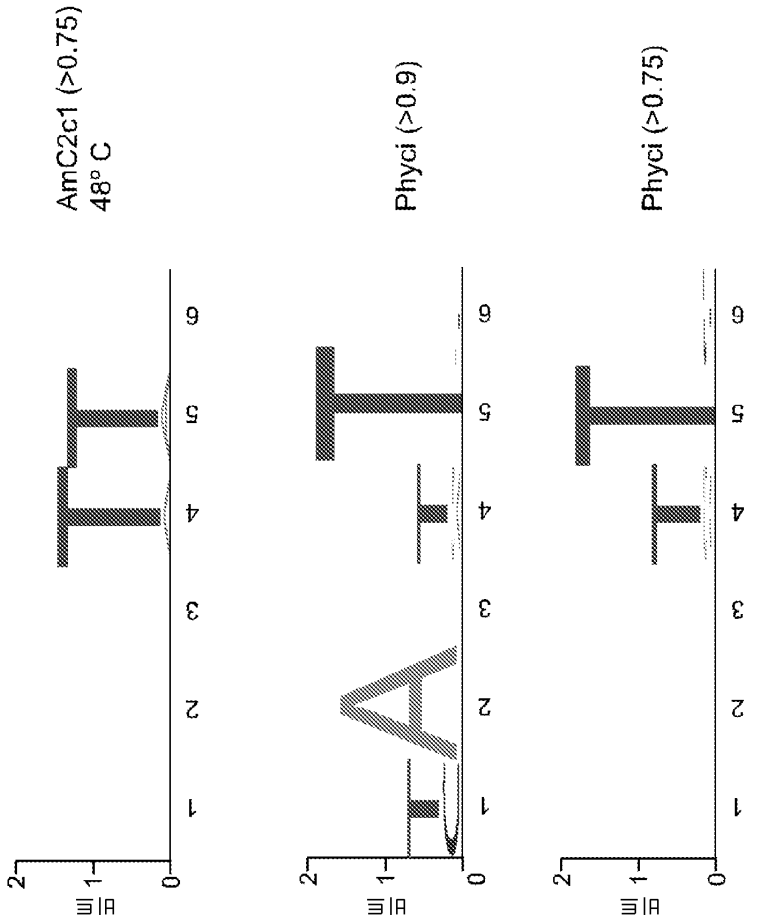
도면18


37℃
48℃
RNA:
nc1521+nc1523
nc1543
nc1544
nc1545
단백질 :
bp
650
300
200
nt
650
300
200
nc1521 = crRNA
nc1523 = tracr
nc1543 = sgRNA1
nc1544 = sgRNA2
nc1545 = sgRNA3
천연 PAGE
우레아-PAGE

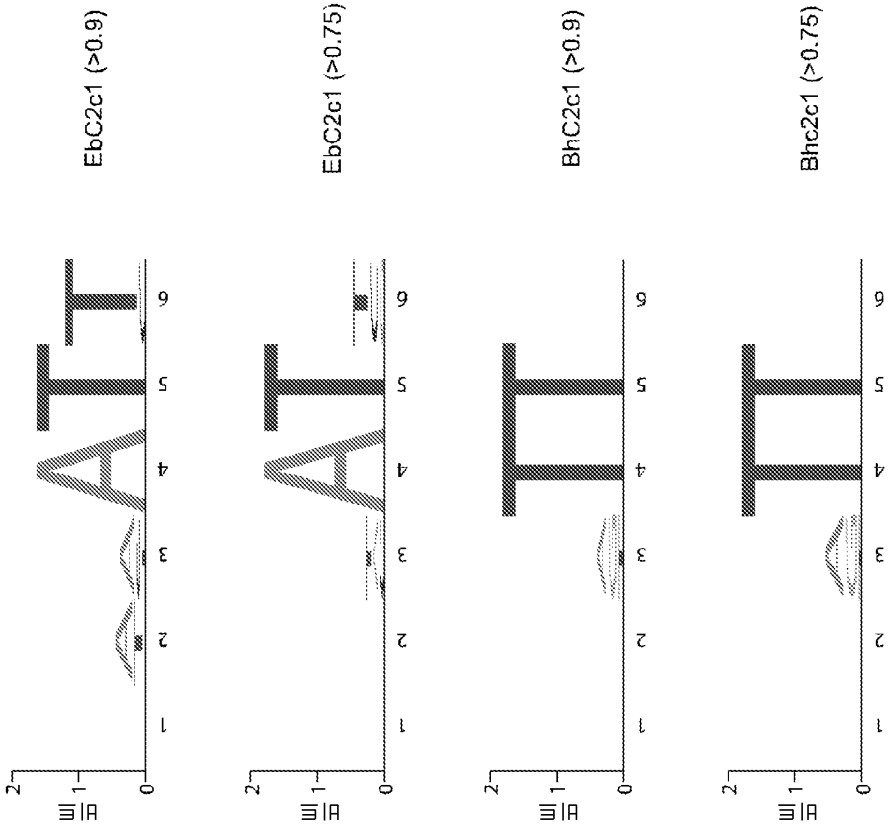
도면19



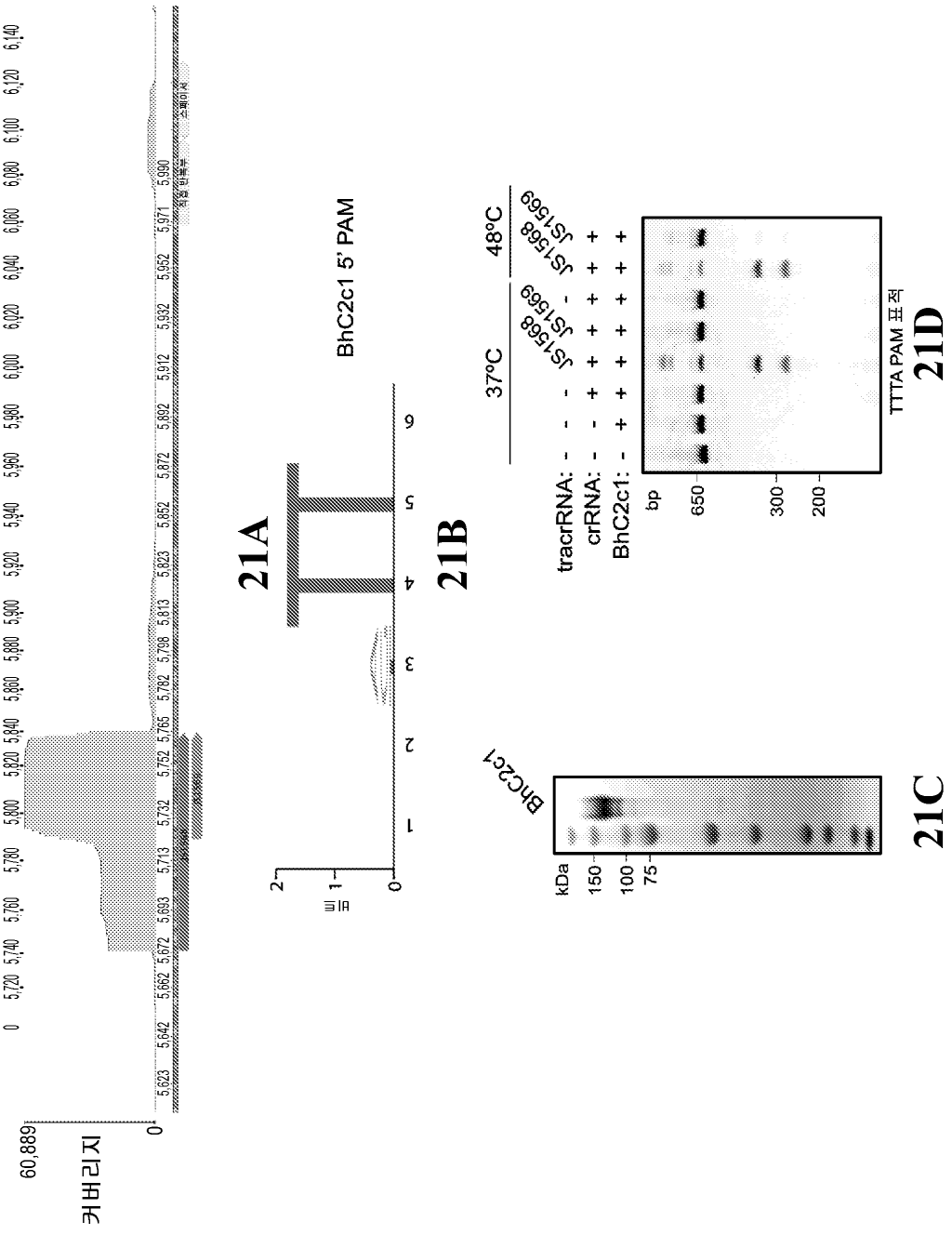
도면20a



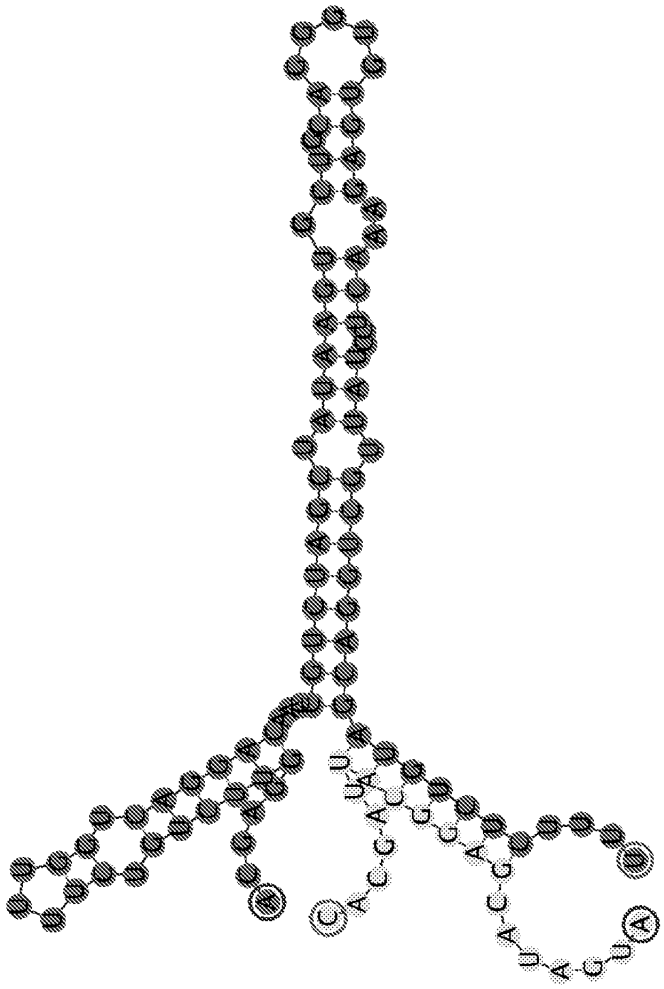
도면20b



도면21


21A
커버리지
60,889
0
BhC2c1 5' PAM
빈도
21B
37°C
48°C
tracrRNA:
crRNA:
BhC2c1:
bp
650
300
200
TTTA PAM 표적
21D
BhC2c1
kDa
150
100
75
21C

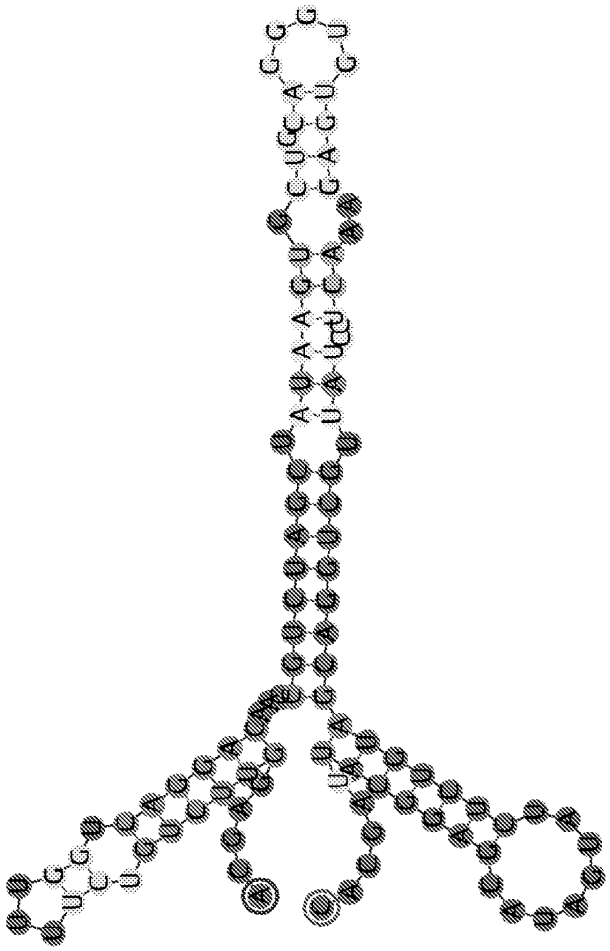
도면22a



도면22b



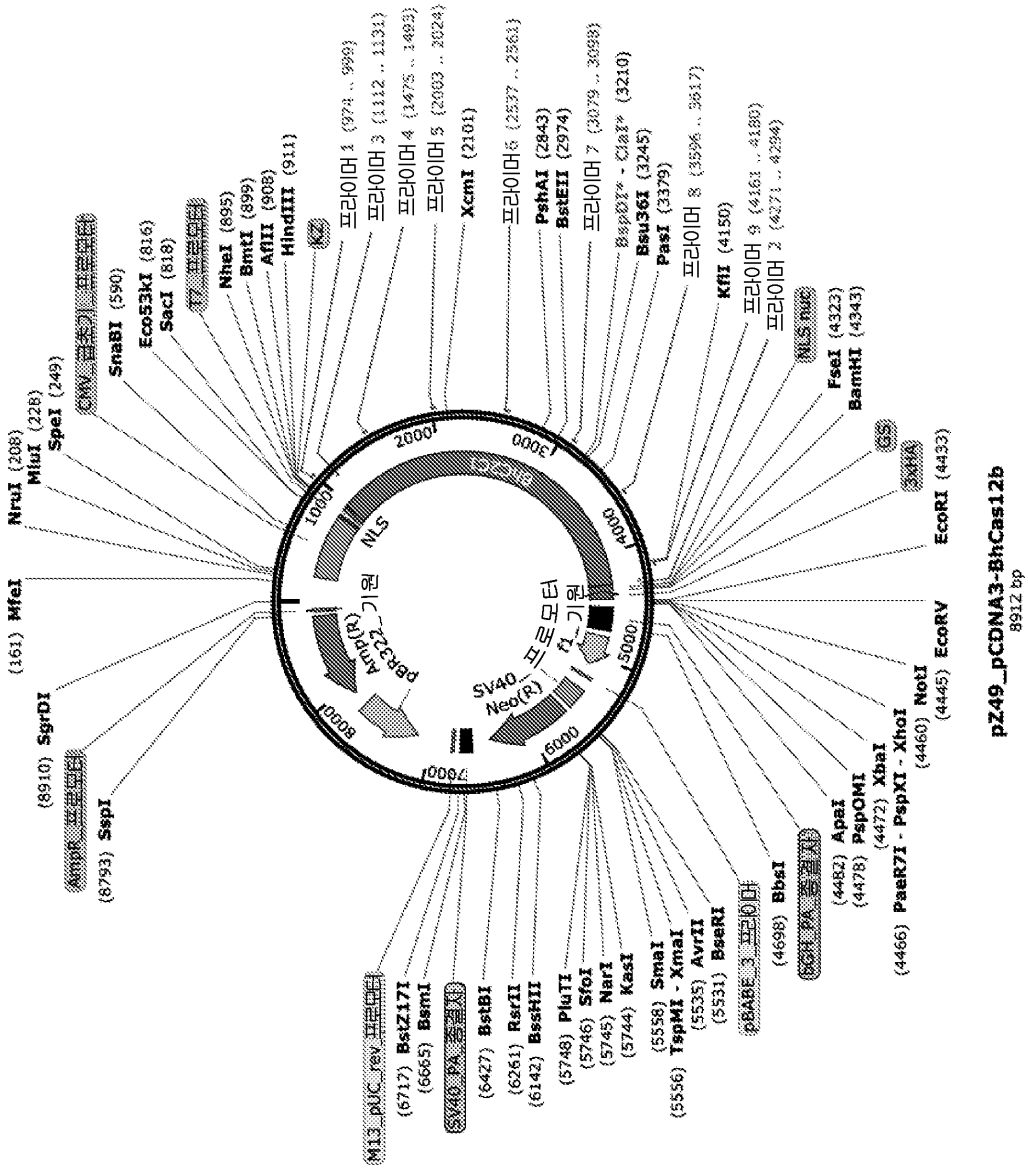
도면22c



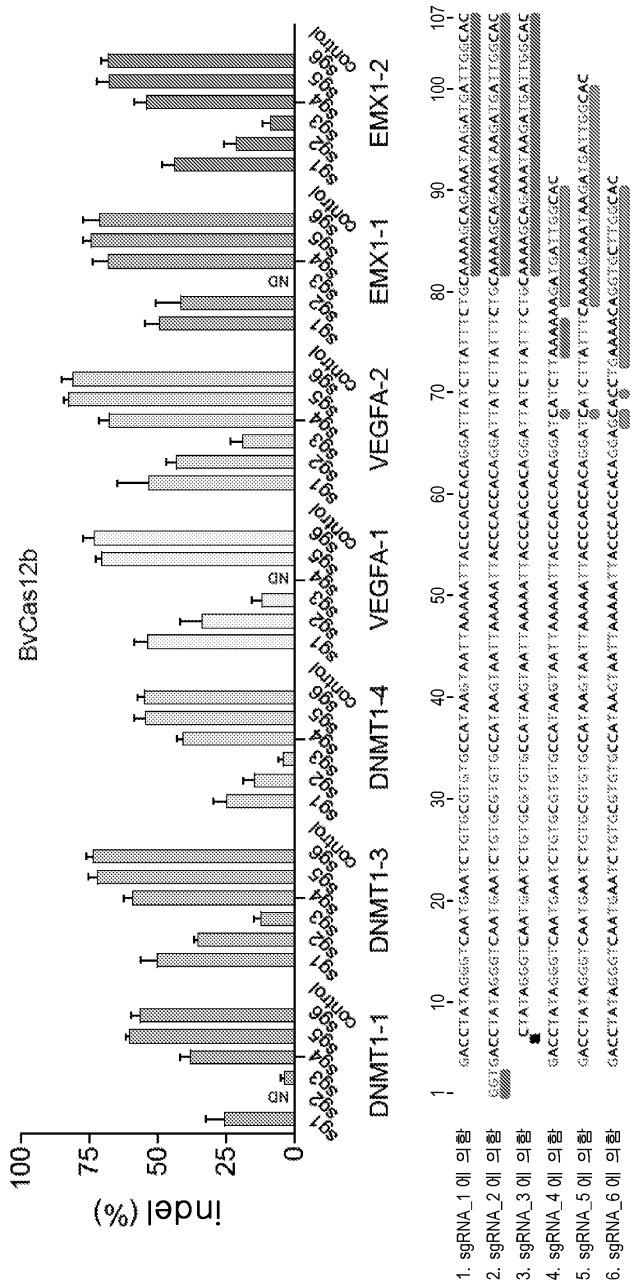
도면22d



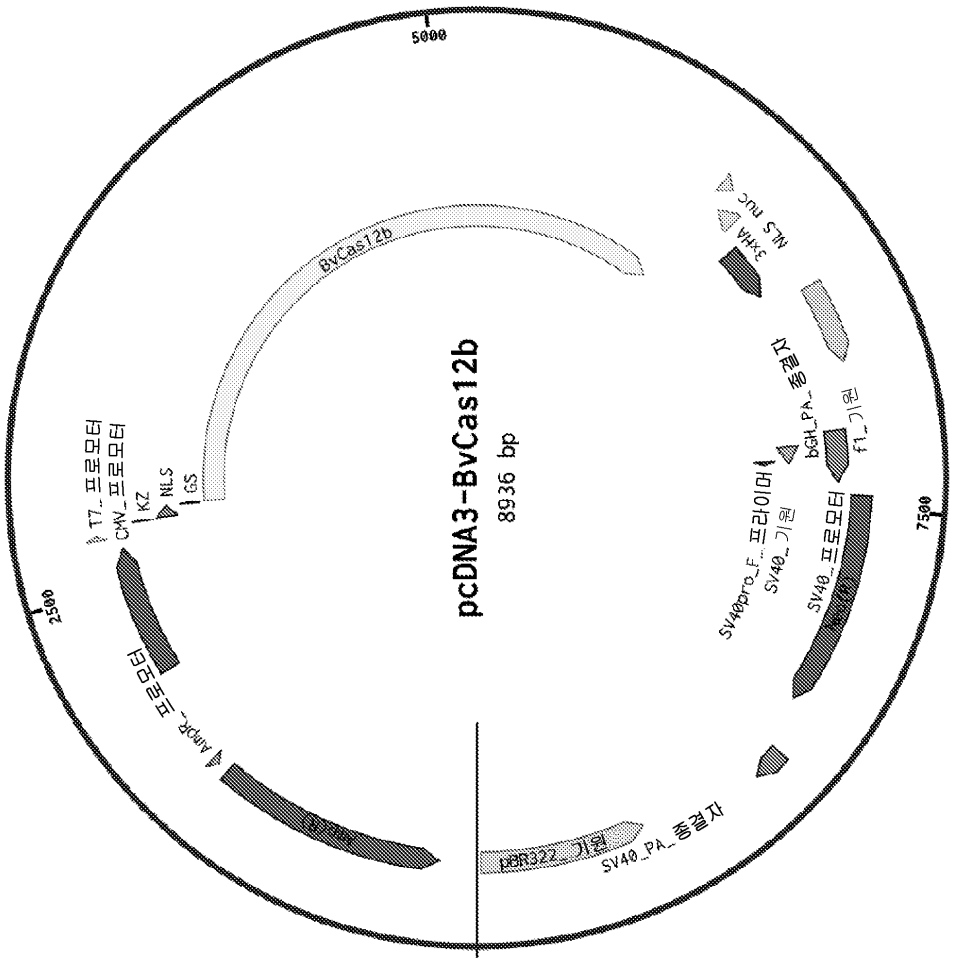
도면23


pZ49_pCDNA3-BhCas12b
8912 bp
CMV_프로모터
T7_프로모터
프라이머 1 (974 .. 999)
프라이머 3 (1112 .. 1131)
프라이머 4 (1475 .. 1493)
프라이머 5 (2003 .. 2024)
프라이머 6 (2537 .. 2561)
프라이머 7 (3079 .. 3098)
프라이머 8 (3596 .. 3617)
프라이머 9 (4161 .. 4180)
프라이머 2 (4271 .. 4294)
NLS
BhCas12b
pBR322_ori
AmpR(R)
SV40_프로모터
Neo(R)
f1_ori
NLS_nuc
GS
3xHA
SV40_PA_종결자
pGH_PA_종결자
M13_puc_rev_프로모터
6BABE_3_프라이머
Amp_프로모터
MfeI (161)
SgrDI (8910)
SspI (8793)
NruI (208)
MluI (228)
SpeI (249)
SnaBI (590)
Eco53kI (816)
SacI (818)
NheI (895)
BmtI (899)
AflII (908)
HindIII (911)
KpnI
EcoRI (4433)
EcoRV (4445)
NotI (4460)
XhoI
XbaI (4472)
PspOMI (4478)
ApaI (4482)
BbsI (4698)
BseRI (5531)
AvrII (5535)
XmaI
SmaI (5558)
KasI (5744)
NarI (5745)
SfoI (5746)
PluTI (5748)
BssHII (6142)
RsrII (6261)
BstBI (6427)
BsmI (6665)
BstZ17I (6717)
PaeR7I - PspXI
SspDB
FseI (4323)
BamHI (4343)
BsaI (3210)
BsiWI - ClaI
PasI (3379)
BstEII (2974)
PshAI (2843)
XcmI (2101)
KflI (4150)

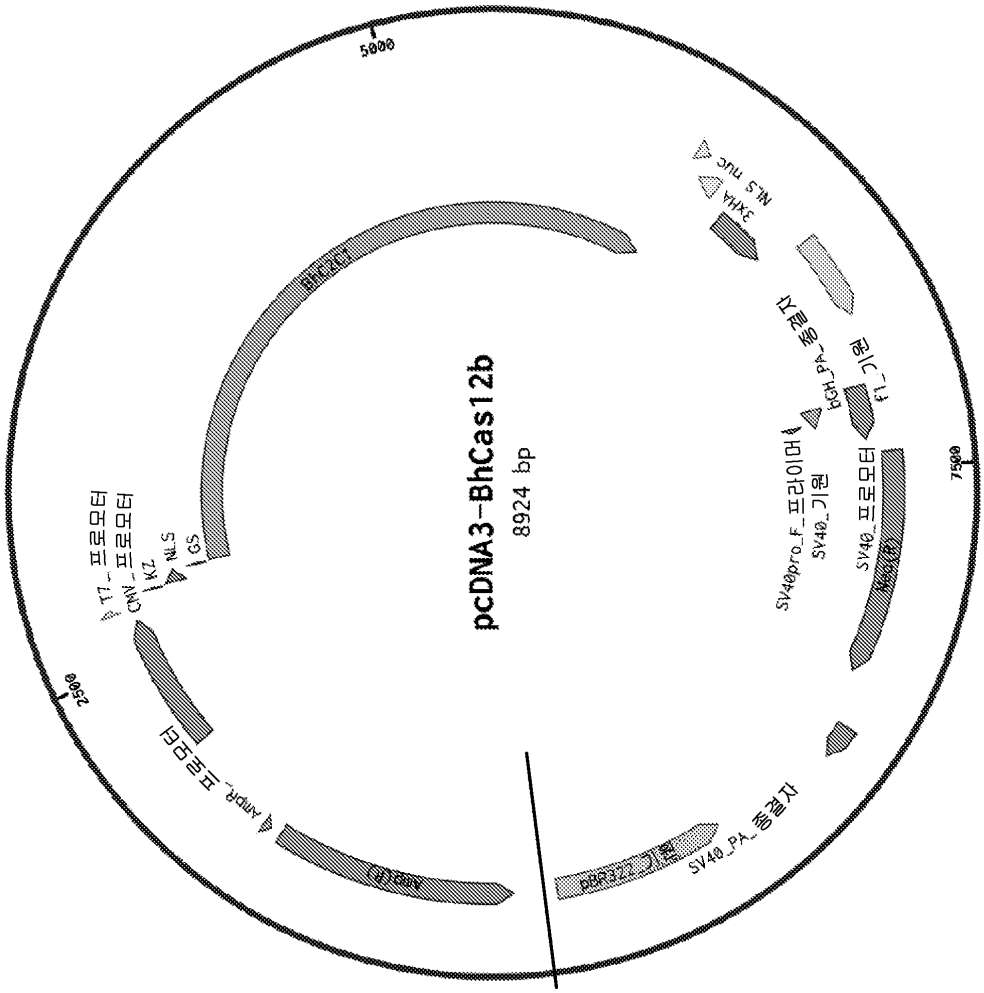
도면24


BvCas12b
Indel (%)
100
75
50
25
0
DNMT1-1
DNMT1-3
DNMT1-4
VEGFA-1
VEGFA-2
EMX1-1
EMX1-2
control
sg1
sg2
sg3
sg4
sg5
sg6
ND
1. sgRNA_1 에 의함
2. sgRNA_2 에 의함
3. sgRNA_3 에 의함
4. sgRNA_4 에 의함
5. sgRNA_5 에 의함
6. sgRNA_6 에 의함

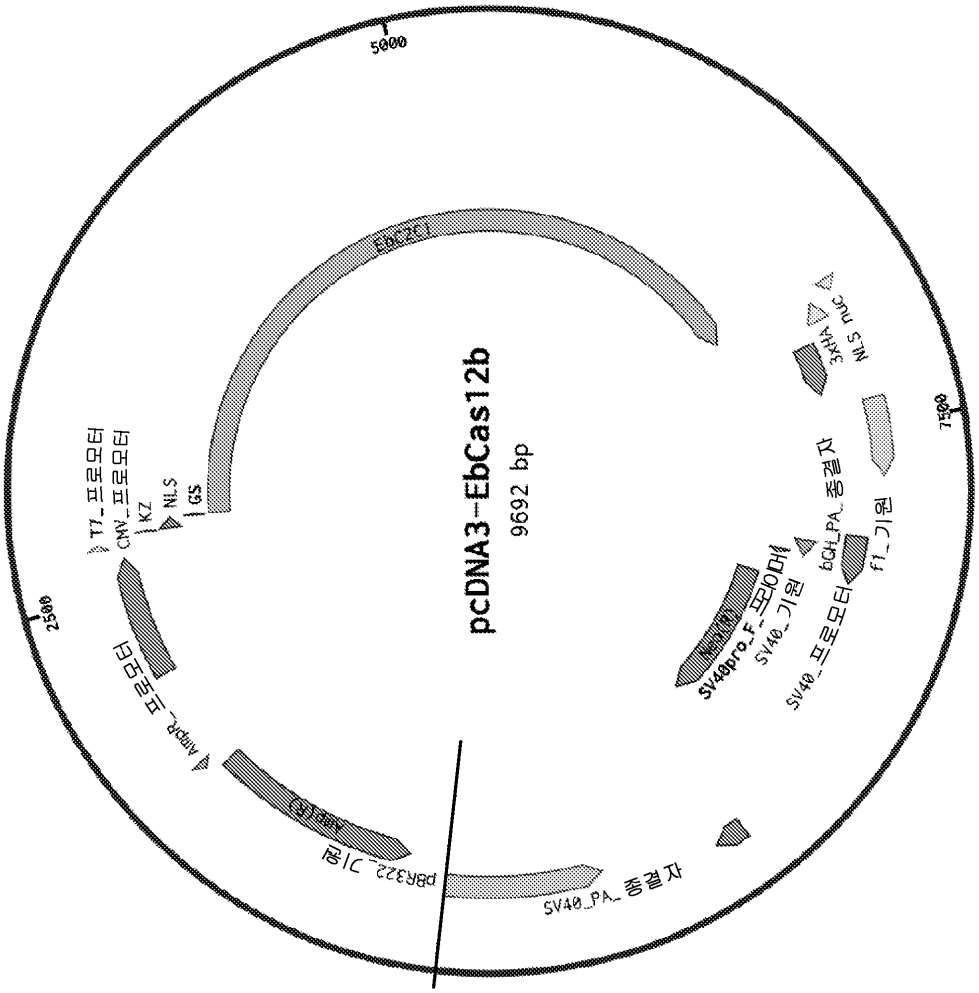
도면25


pcDNA3-BvCas12b
8936 bp
5000
7500
2500
BvCas12b
T7_프로모터
CMV_프로모터
KZ
NLS
GS
3'HA
NLS
SV40_프로모터
SV40_기원
SV40_PA_종결자
f1_기원
pBR322_기원
AmpR_프로모터

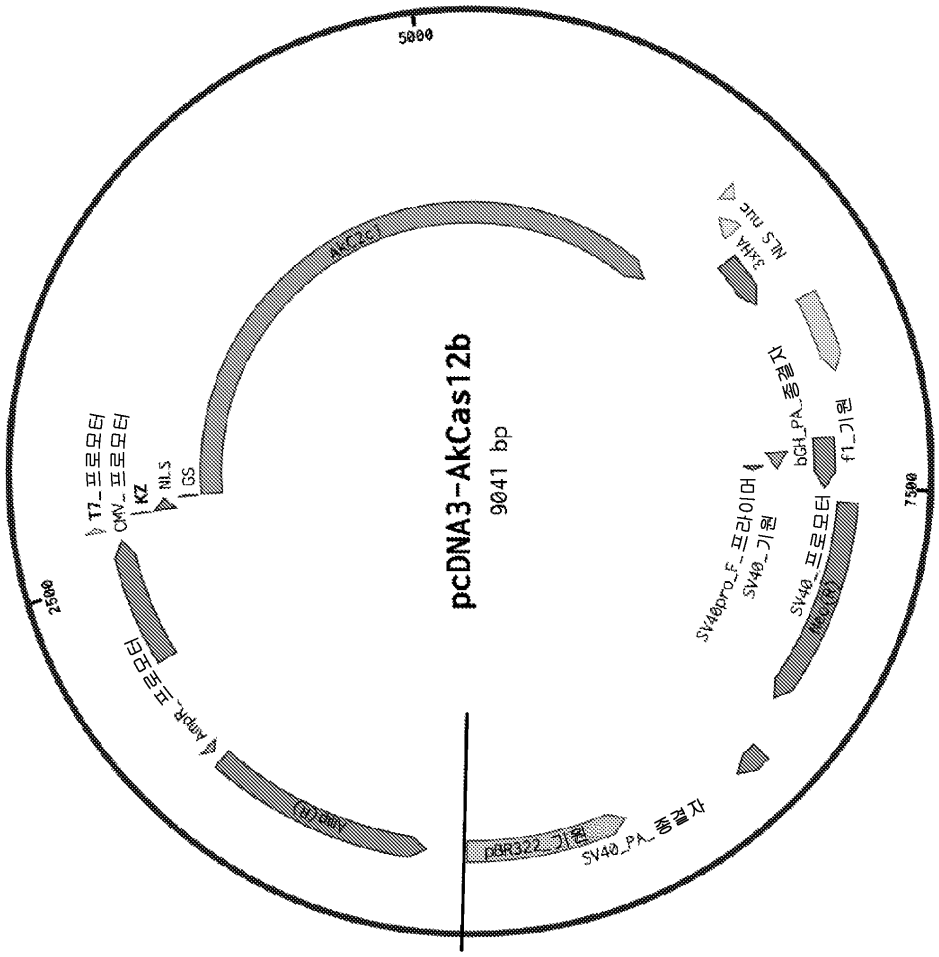
도면26


pcDNA3-BhCas12b
8924 bp
5000
7500
2500
1
BhCas12b
3xHA
NLS nuc
SV40pro_F_프라이머
SV40_기원
bGH_PA_종결자
f1_기원
SV40_프로모터
NEO(R)
SV40_PA_종결자
pBR322_기원
AMP(R)
AmpR_프로모터
T7_프로모터
CMV_프로모터
KZ
NLS
GS

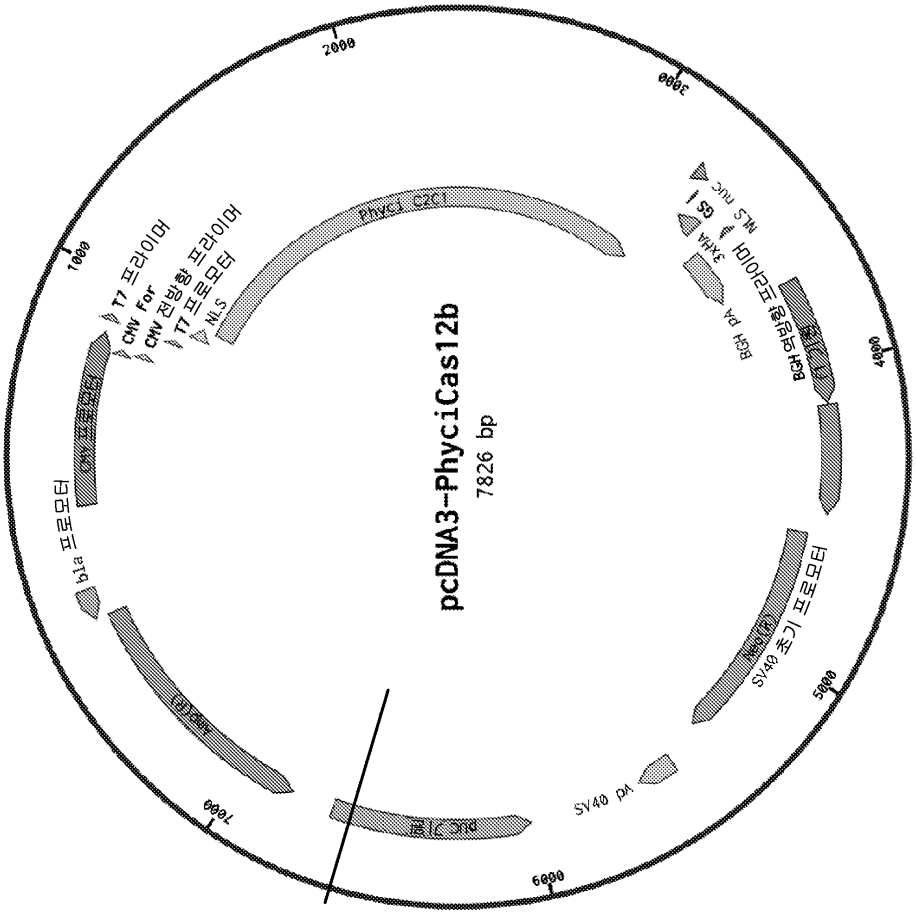
도면27


pcDNA3-EbCas12b
9692 bp
5000
7500
2500
EbCas12
NLS
GS
KZ
T7_프로모터
CMV_프로모터
AmpR_프로모터
pBR322_기원
SV40_PA_종결자
SV40pro_F_프로모터
SV40_기원
bGH_PA_종결자
SV40_프로모터
f1_기원
NLS
3xHA
Nuc

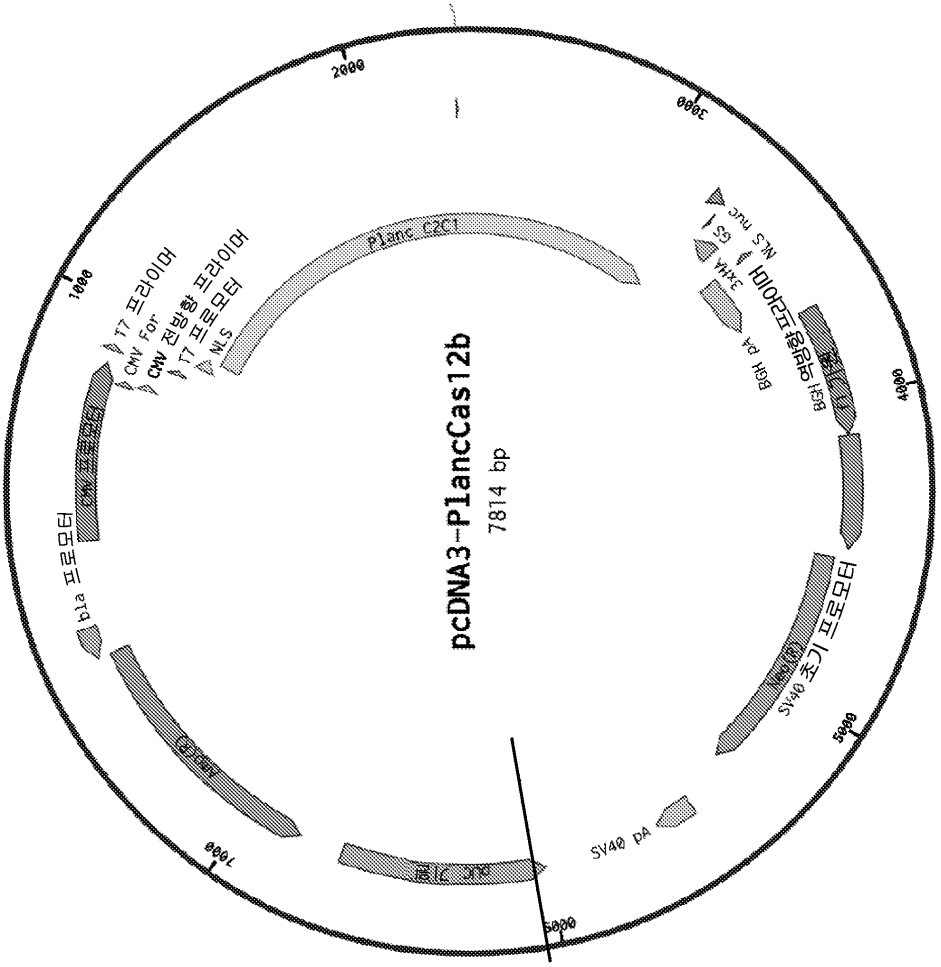
도면28


pcDNA3-AkCas12b
9041 bp
5000
7500
2500
AkC12b
T7_프로모터
CMV_프로모터
KZ
NLS
GS
3xHA
NLS
nuc
bGH_PA_종결자
f1_기원
SV40pro_F_프라이머
SV40_기원
SV40_프로모터
NeoR/KanR
SV40_PA_종결자
pBR322_기원
AmpR_프로모터
Ampicillin

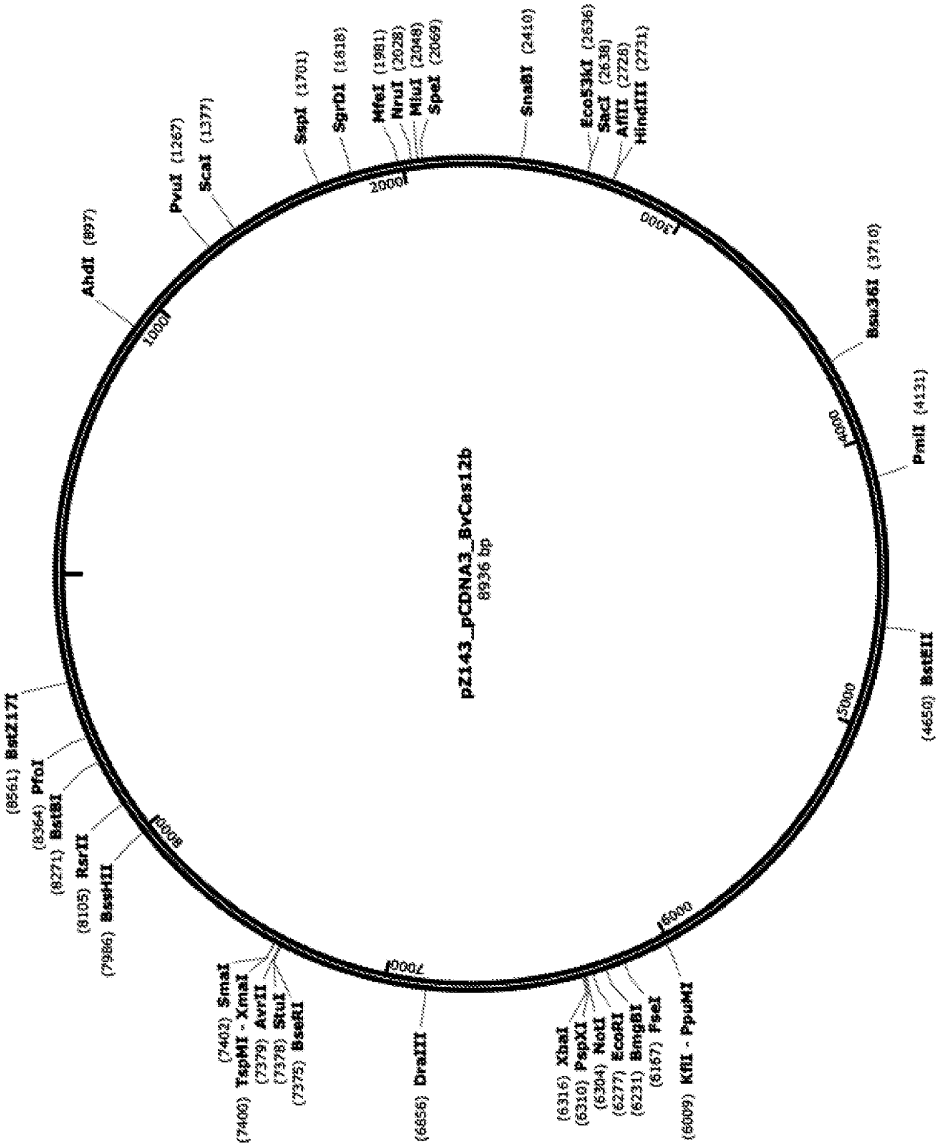
도면29


pcDNA3-PhyciCas12b
7826 bp
PhyCi C2C1
T7 프라이머
CMV For
CMV 전방향 프라이머
T7 프로모터
N-S
CMV 프로모터
bla 프로모터
bla
ori
pUC ori
SV40 pA
SV40초기 프로모터
NLS
3'HA
BGH pA
BGH 역방향 프라이머
1000
2000
3000
4000
5000
6000
7000

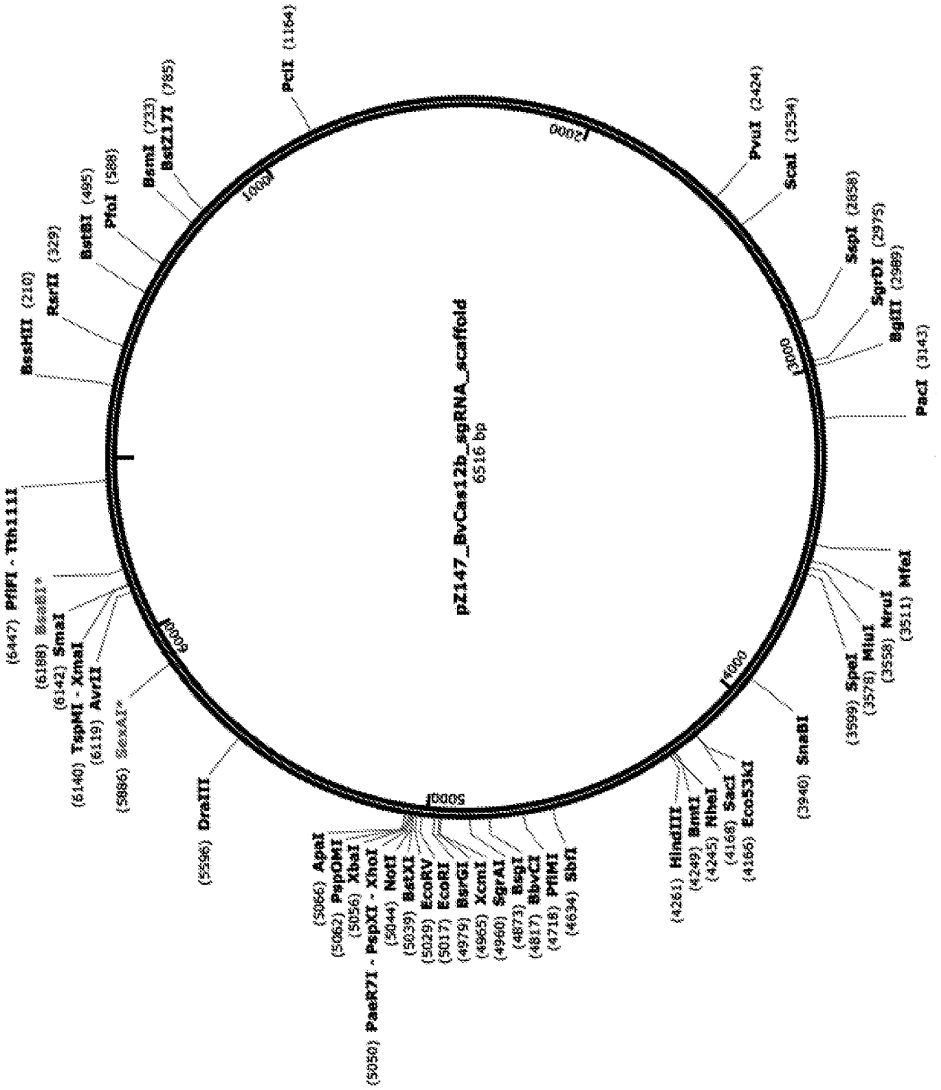
도면30


pcDNA3-PlancCas12b
7814 bp
PLanc F2C1
T7 프라이머
CMV For
CMV 전방향 프라이머
T7 프로모터
W.S.
CMV 프로모터
bla 프로모터
ampR
ori
SV40 pA
SV40 초기 프로모터
NeoR
BGH pA
3XHA
NLS
GS
NLS nuc
1000
2000
3000
4000
5000
6000
7000

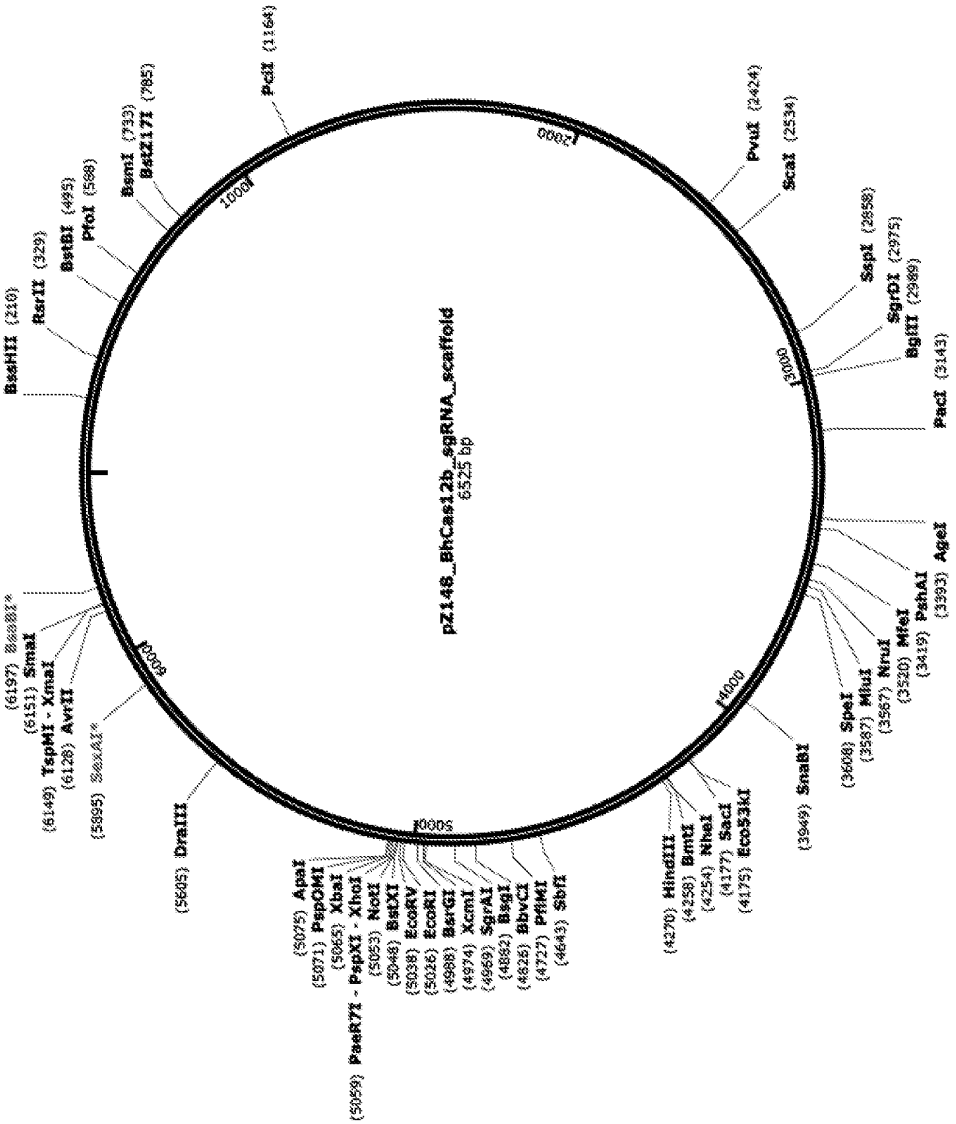
도면31


pZ143_pCDNA3_ByCas12b
8936 bp
1000
2000
3000
4000
5000
6000
7000
8000
AhdI (897)
PvuI (1267)
ScaI (1377)
SspI (1701)
SgrDI (1818)
MfeI (1981)
NruI (2028)
MluI (2049)
SpeI (2069)
SnaBI (2410)
Eco53kI (2636)
SacI (2638)
AflII (2728)
Hind3III (2731)
Bsu36I (3710)
PmlI (4131)
BstEII (4650)
KflI - PpuMI (6009)
FseI (6167)
BmgBI (6231)
EcoRI (6277)
NotI (6304)
PspXI (6310)
XhoI (6316)
DraIII (6856)
StuI (7378)
AvrII (7379)
TspMI - XmaI (7400)
SmaI (7402)
BseRI (7375)
BsmBI (7986)
RsrII (8105)
BstBI (8271)
PfoI (8364)
Bst2I7I (8561)

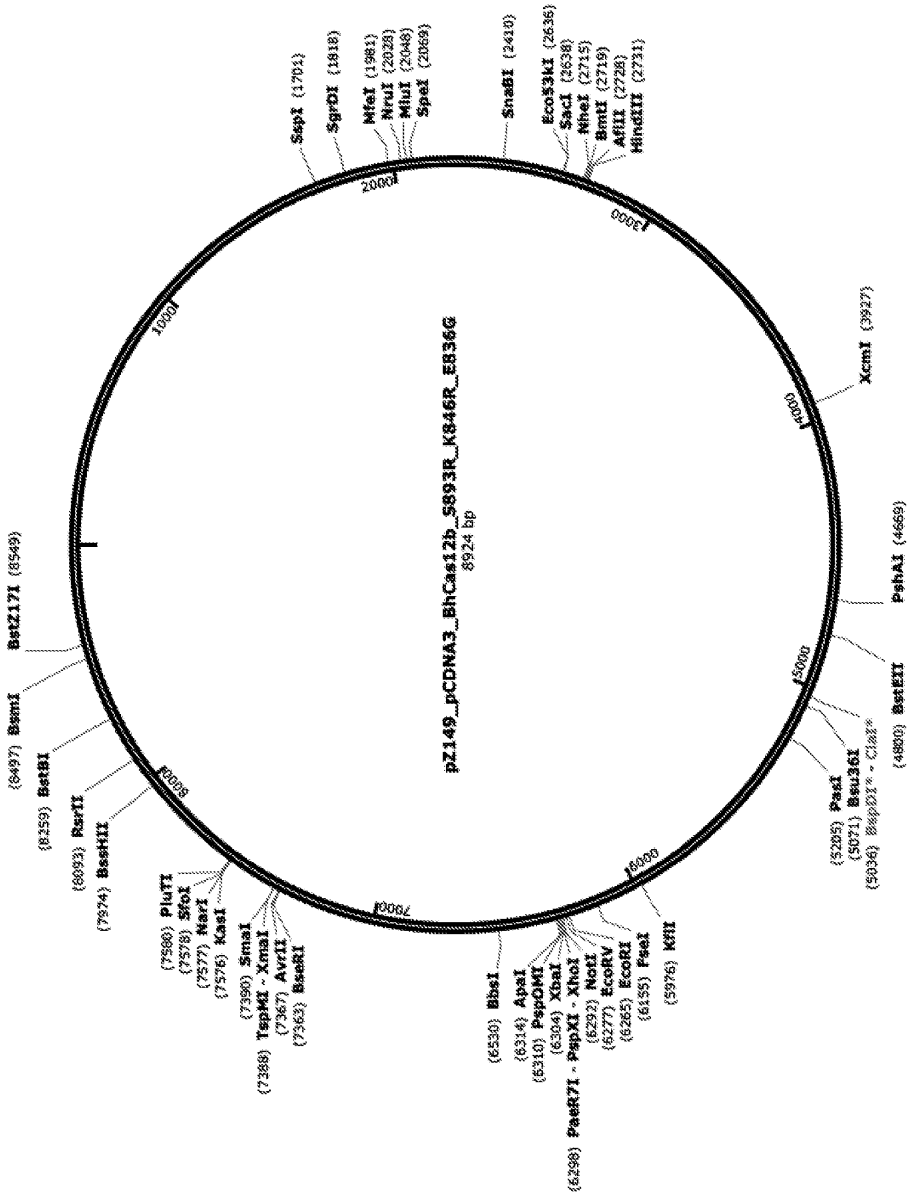
도면32


pZ147_BvCas12b_sgRNA_scaffold
6516 bp
1000
2000
3000
4000
5000
6000
PciI (1164)
BamI (733)
Bst2117I (785)
PtoI (588)
BstBI (495)
RsrII (329)
BssHII (210)
PflFI - Tth111I (6447)
BsaBI* (6188)
SmaI (6142)
TspMI - XmaI (6140)
AvrII (6119)
BsaAI* (5886)
DraIII (5596)
ApaI (5086)
PspOMI (5062)
XbaI (5056)
PspXI - XhoI (5050)
NotI (5044)
BstXI (5039)
EcoRV (5029)
EcoRI (5017)
BsrGI (4979)
XcmI (4965)
SgrAI (4960)
BsgI (4873)
BbvCI (4817)
PflMI (4718)
SbfI (4634)
HindIII (4261)
BmtI (4249)
NheI (4245)
SacI (4168)
Eco53kI (4166)
SnaBI (3940)
SpeI (3599)
MluI (3578)
NruI (3558)
MfeI (3511)
PacI (3143)
SgrDI (2989)
BglII (2975)
SapI (2858)
ScaI (2534)
PvuI (2424)

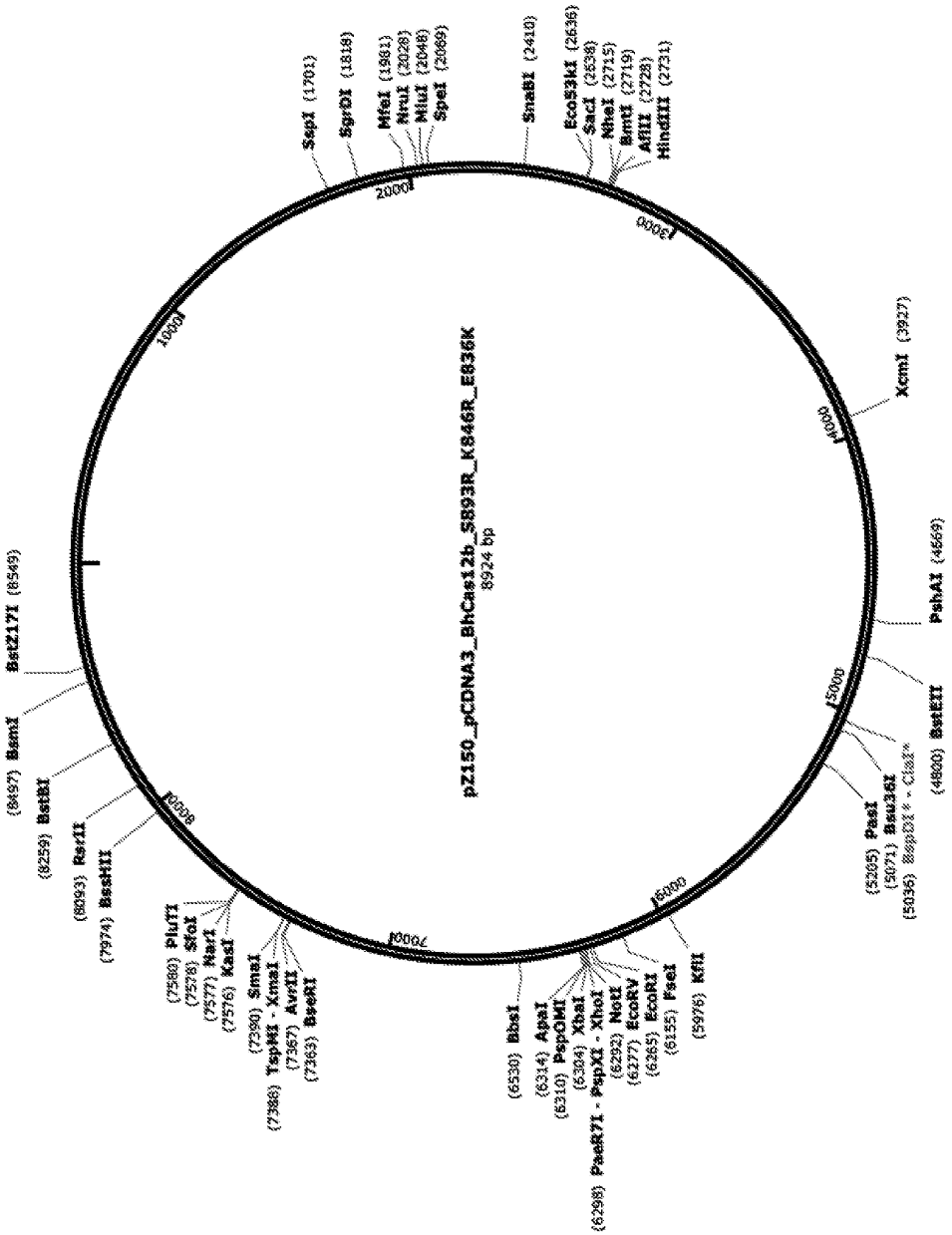
도면33


pZ148_BhCas12b_sgRNA_scaffold
6525 bp
1000
2000
3000
4000
5000
6000
PciI (1164)
BamI (733)
Bst2171 (785)
PtoI (588)
BstBI (495)
RsrII (329)
BssHII (210)
PvuI (2424)
ScaI (2534)
SapI (2658)
SgrDI (2975)
BglII (2989)
PacI (3143)
AgeI (3393)
PshAI (3419)
NruI (3520)
MfeI (3567)
MluI (3587)
SpeI (3608)
SnaBI (3949)
HindIII (4270)
BmtI (4258)
NheI (4254)
SacI (4177)
Eco53kI (4175)
SbfI (4643)
PflMI (4727)
BbvCI (4826)
BsgI (4882)
SgrAI (4969)
XcmI (4974)
BsrGI (4988)
EcoRI (5036)
EcoRV (5038)
BstXI (5048)
NotI (5053)
PspXI - XhoI (5059)
XbaI (5065)
PspOMI (5071)
ApaI (5075)
DraIII (5605)
SexAI* (5895)
TspMI - XmaI (6128)
AvrII (6149)
SmaI (6151)
BsaBI* (6197)

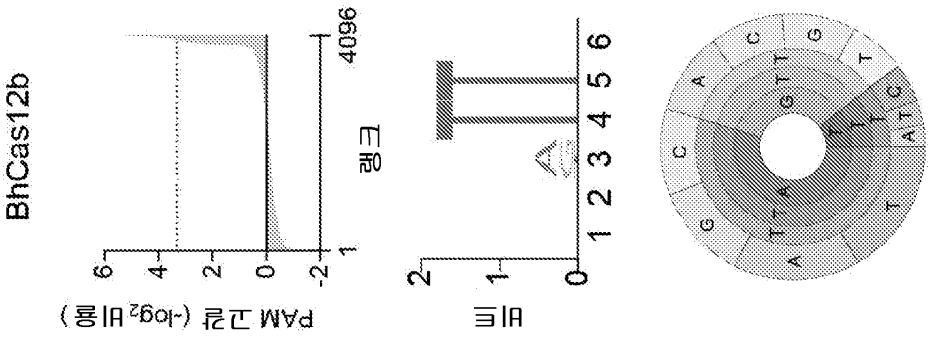
도면34


pZ149_pCDNA3_BhCas12b_S893R_K846R_E836G
8924 bp
1000
2000
3000
4000
5000
6000
7000
8000
SapI (1701)
SgrDI (1818)
NheI (1981)
NruI (2028)
MluI (2048)
SpeI (2069)
SnaBI (2410)
Eco53kI (2636)
SacI (2638)
NheI (2715)
BmtI (2719)
AflII (2728)
HindIII (2731)
XcmI (3927)
PshAI (4669)
BstEII (4830)
Pst1 (5265)
Bsu36I (5071)
BspDI* - ClaI* (5036)
KflI (5976)
PaeI (6155)
EcoRI (6265)
EcoRV (6277)
NotI (6292)
XhoI - PspXI (6304)
XbaI (6334)
PspOMI (6310)
ApaI (6314)
BbsI (6530)
AvrII (7367)
TspMI - XmaI (7369)
SmaI (7390)
KasI (7576)
NarI (7577)
SfoI (7578)
PluTI (7580)
BseRI (7363)
BssHII (7974)
RsrII (8093)
BstBI (8359)
BsmI (8497)
BstZ17I (8549)

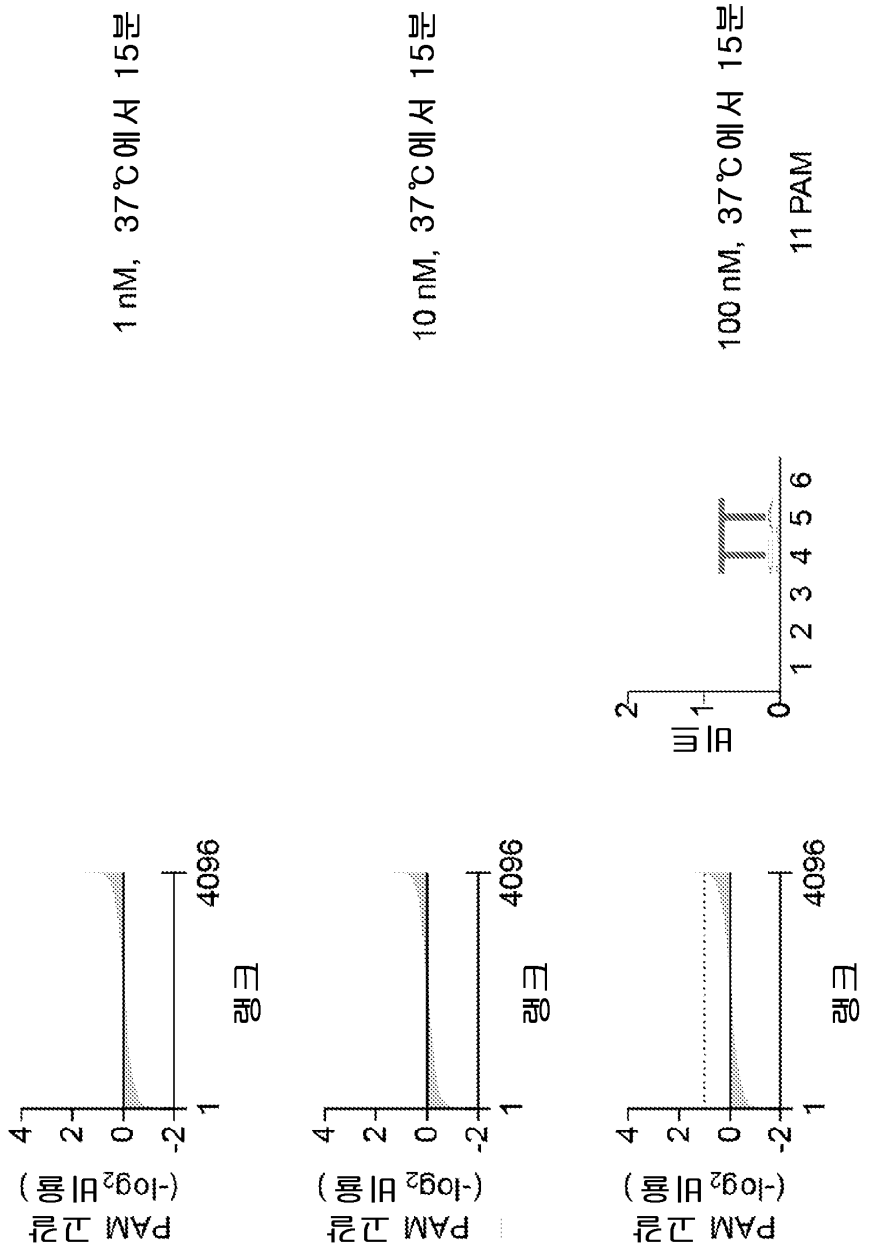
도면35


pZ150_pCDNA3_BhCas12b_S893R_K846R_E836K
8924 bp
SepI (1701)
SgrDI (1818)
MfeI (1961)
NruI (2028)
MluI (2048)
SpeI (2069)
SnaBI (2410)
Eco53kI (2636)
SacI (2638)
NheI (2715)
BmtI (2719)
AflII (2728)
HindIII (2731)
XcmI (3927)
PshAI (4569)
BstEII (4800)
PasI (5205)
Bsu36I (5071)
BspDI* - ClaI*
(5036)
KflI (5976)
PseI (6155)
EcoRI (6265)
EcoRV (6277)
NotI (6292)
XhoI (6304)
PspOMI (6310)
ApaI (6314)
BbsI (6530)
PaeR7I - PspXI
(6298)
TspMI - XmaI (7367)
AvrII (7363)
BseRI (7386)
SmaI (7390)
KasI (7576)
NarI (7577)
SfoI (7578)
PluTI (7580)
BsaHII (7974)
RsrII (8093)
BstBI (8259)
BsmI (8497)
Bst217I (8549)

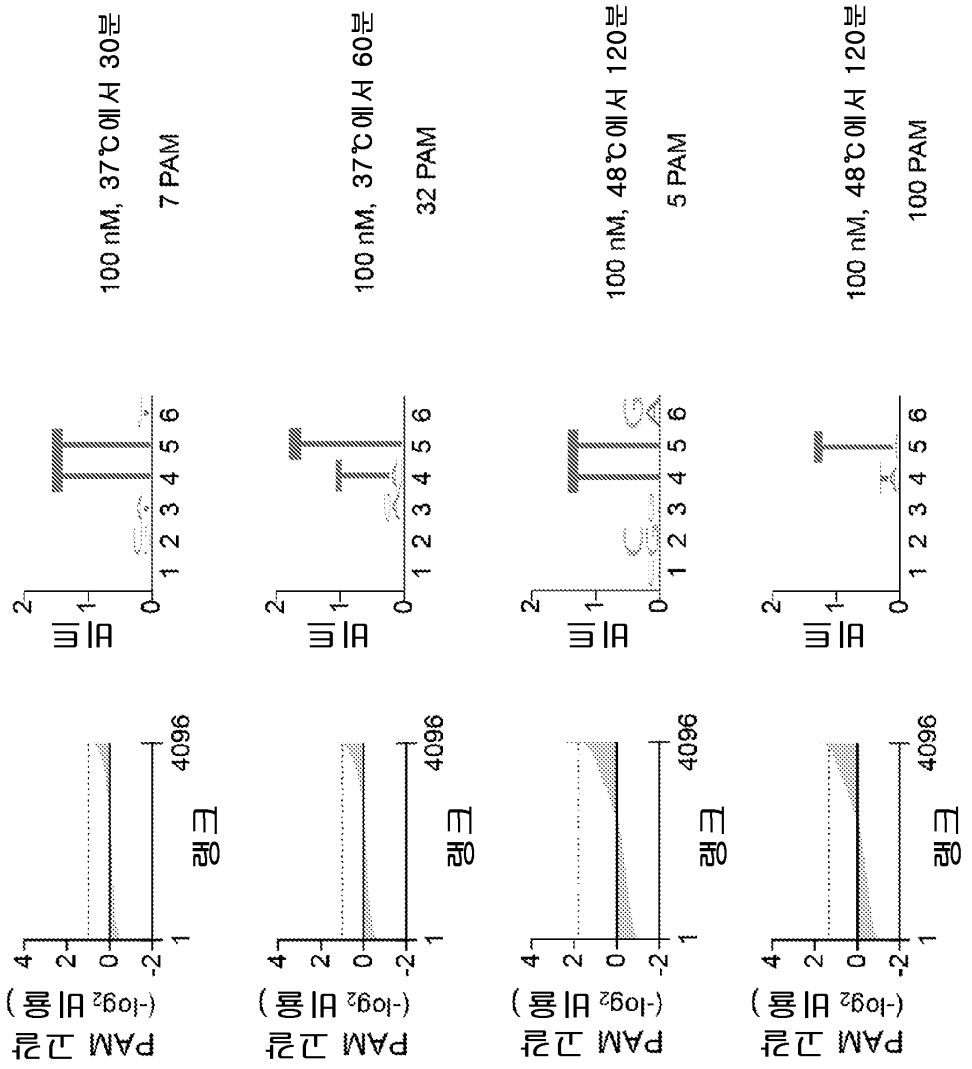
도면36a


BhCas12b
PAM 고갈 (-log2 비율)
6
4
2
0
-2
1
4096
크기
비트
2
1
0
1 2 3 4 5 6

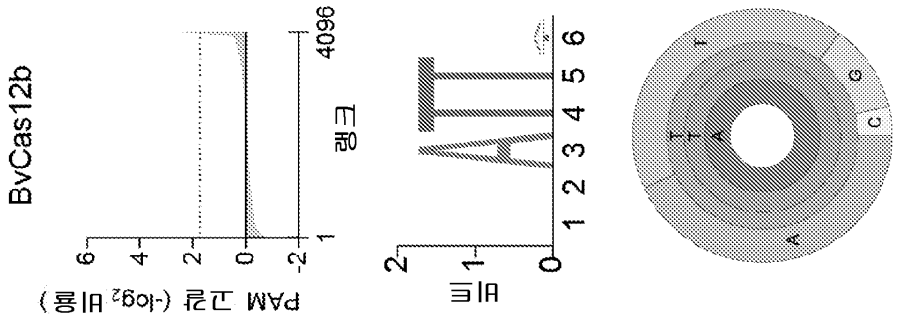
도면36b


1 nM, 37℃에서 15분
10 nM, 37℃에서 15분
100 nM, 37℃에서 15분
11 PAM
비트
랭크
PAM 고갈 (-log2 비율)
4
2
0
-2
1
4096
1 2 3 4 5 6

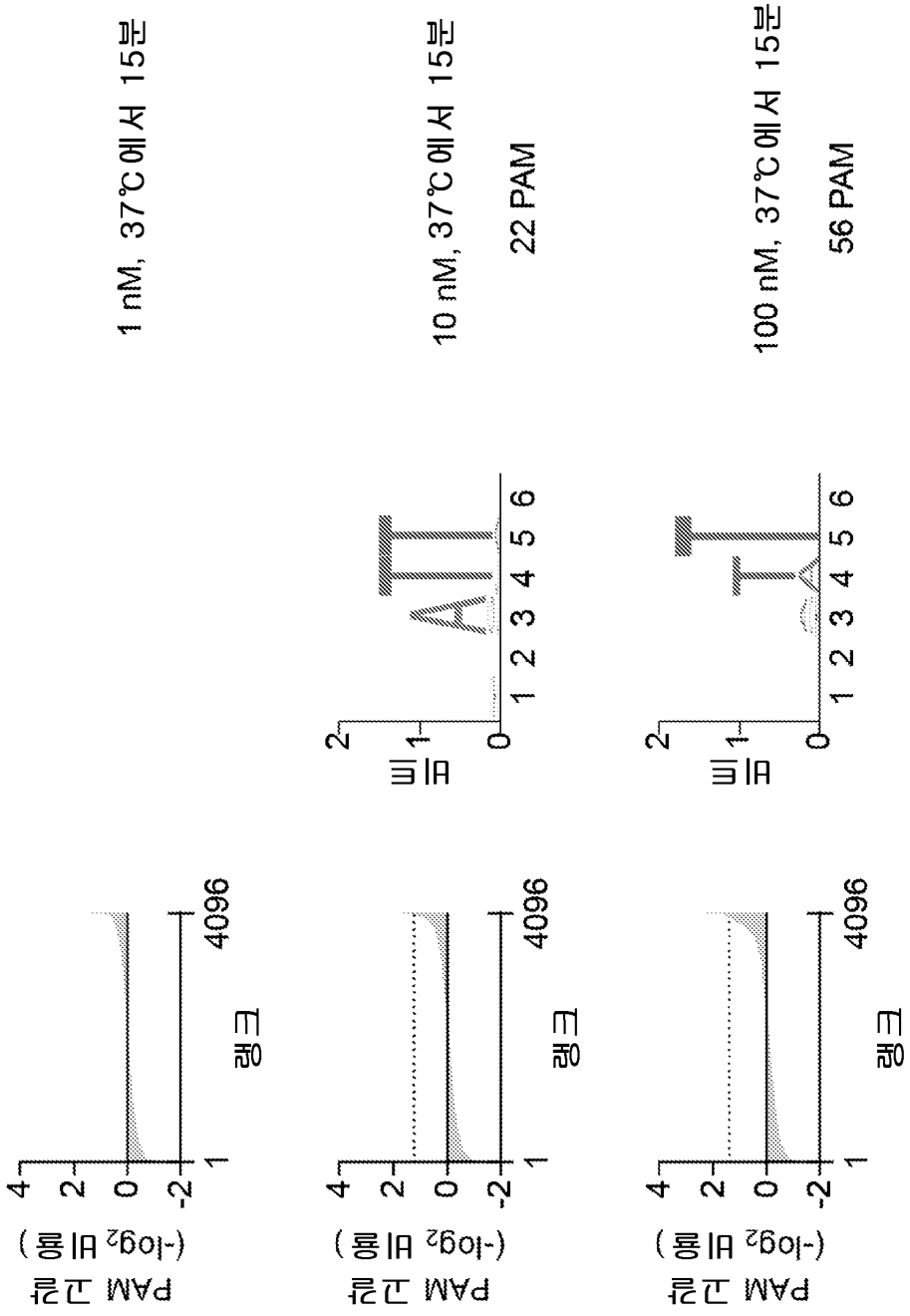
도면36c


100 nM, 37℃에서 30분
7 PAM
100 nM, 37℃에서 60분
32 PAM
100 nM, 48℃에서 120분
5 PAM
100 nM, 48℃에서 120분
100 PAM
비율
1 2 3 4 5 6
PAM 고갈 (-log2 비율)
랜크
4096

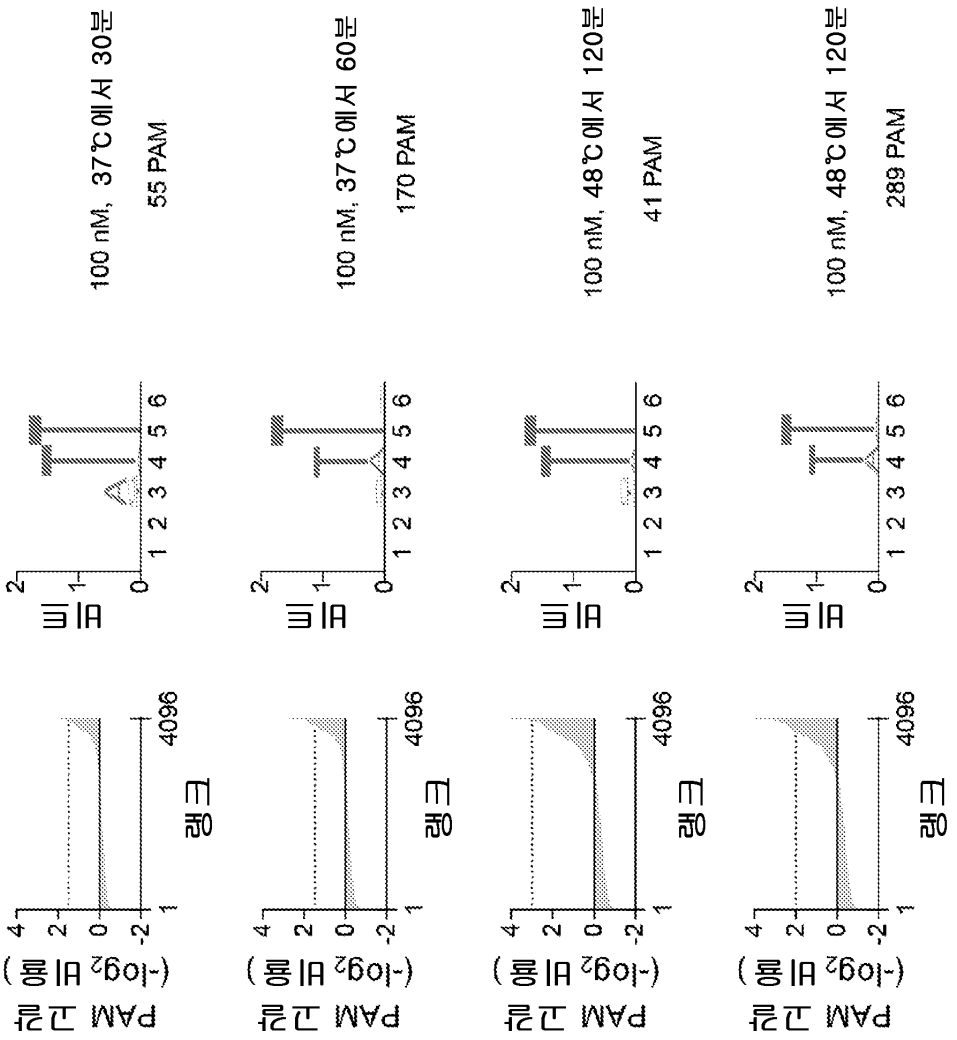
도면37a


BvCas12b
PAM 고갈 (-log2 비율)
랜크
4096
비율
1 2 3 4 5 6
A
C
G
T

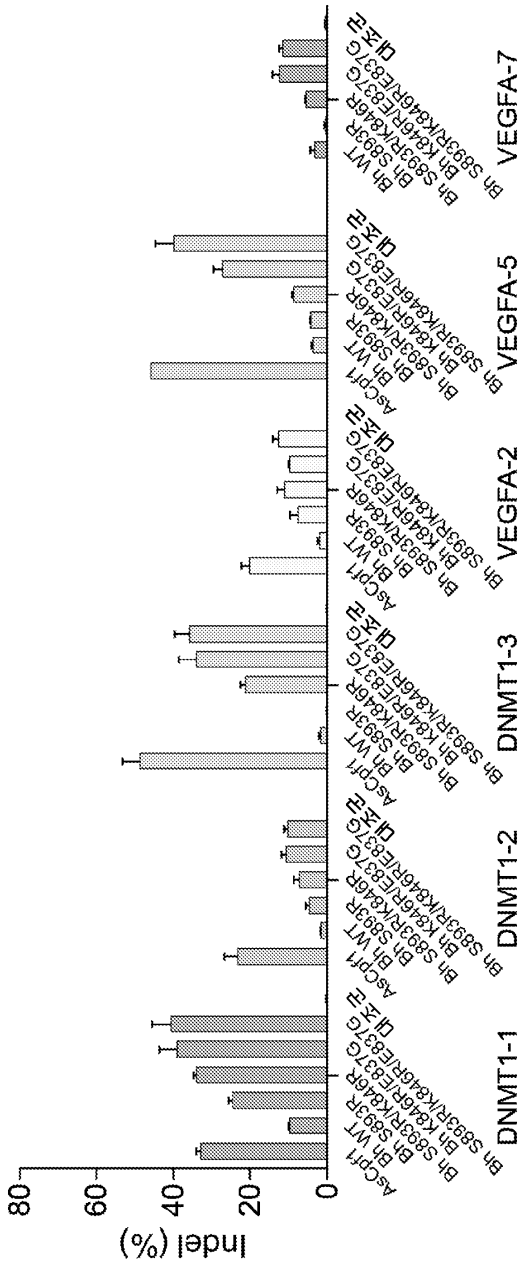
도면37b


1 nM, 37℃에서 15분
10 nM, 37℃에서 15분
22 PAM
100 nM, 37℃에서 15분
56 PAM
비율
1 2 3 4 5 6
0
1
2
랭크
1
4096
PAM 고갈
(-log2 비율)
4
2
0
-2

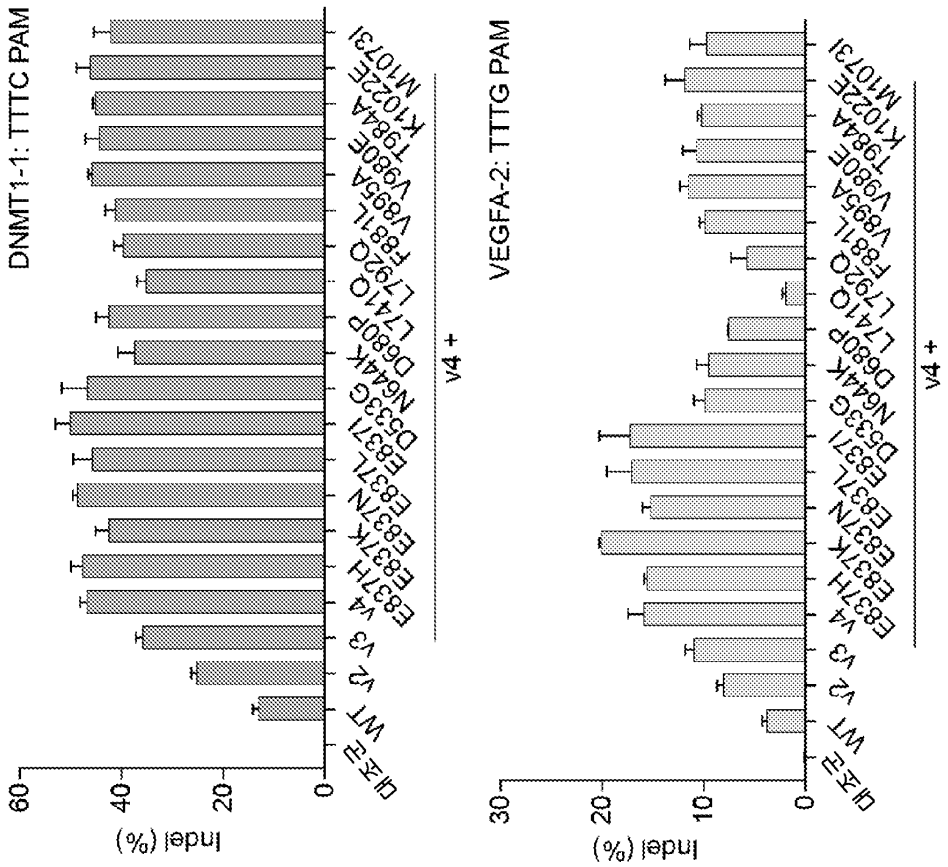
도면37c


100 nM, 37℃에서 30분
55 PAM
100 nM, 37℃에서 60분
170 PAM
100 nM, 48℃에서 120분
41 PAM
100 nM, 48℃에서 120분
289 PAM
비드
랜드
PAM 고갈
(-log2 비율)

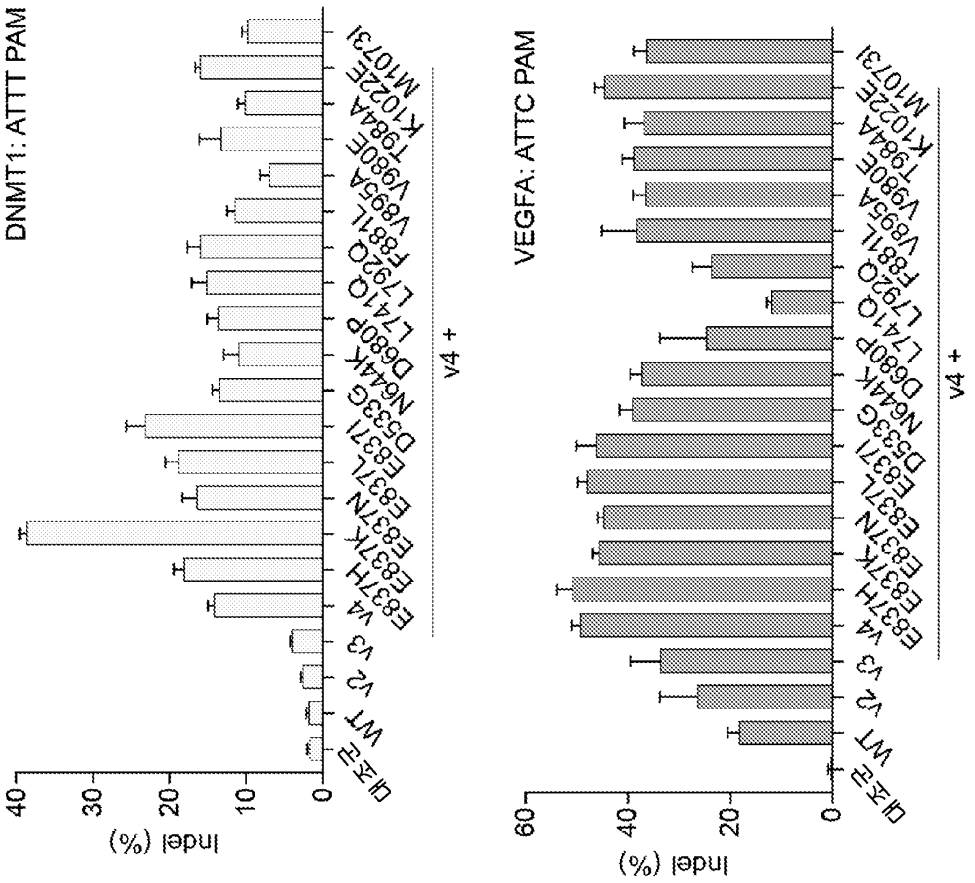
도면38


Indel (%)
0
20
40
60
80
DNMT1-1
DNMT1-2
DNMT1-3
VEGFA-2
VEGFA-5
VEGFA-7
AsCpf1
Bh WT
Bh S893R
Bh K846R/E837G
BH S893R/K846R/E837G
UI 조건

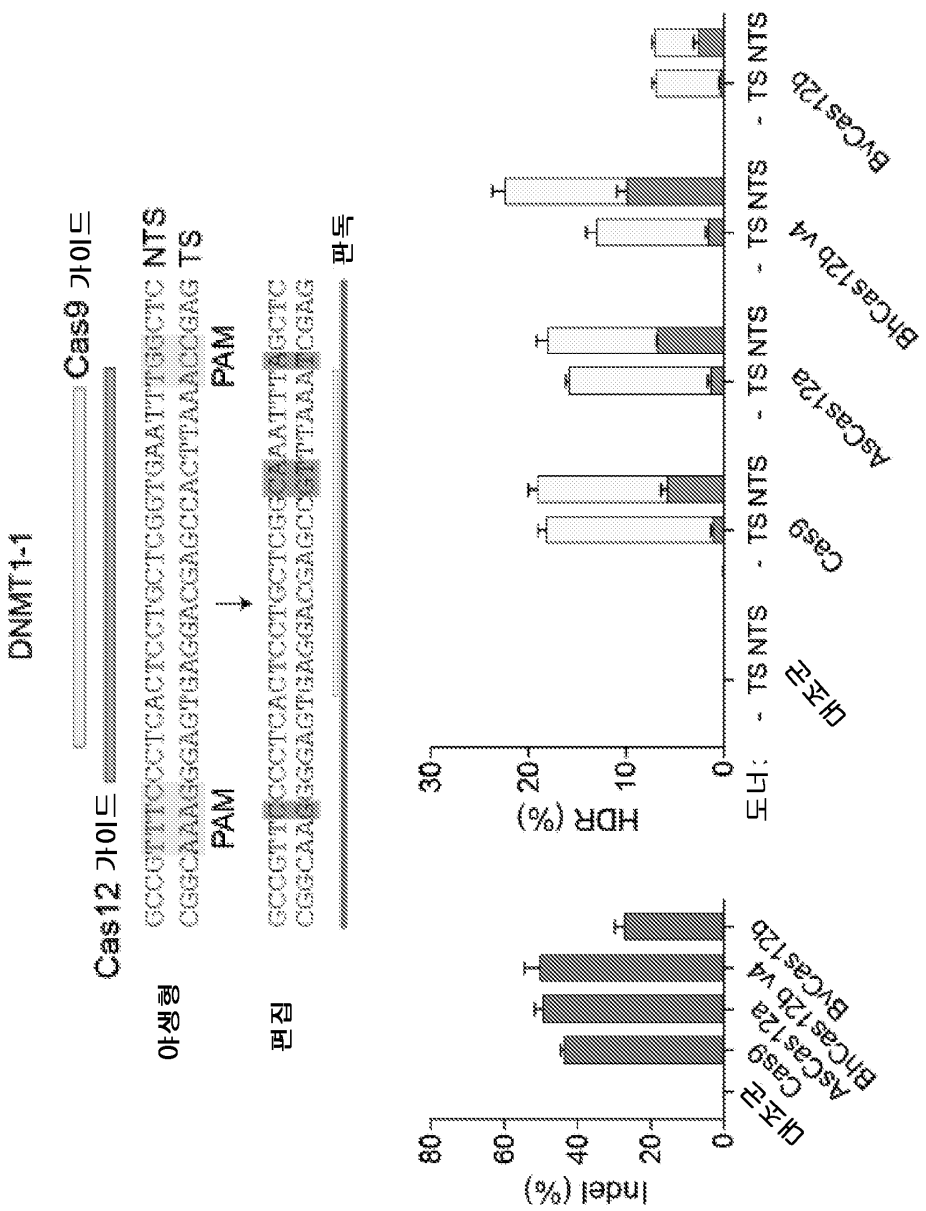
도면39a


DNMT1-1: TTTC PAM
Indel (%)
0
20
40
60
대조군
WT
v2
v3
v4
E837H
E837K
E837N
E837L
E837I
D533G
N644K
D680P
L741Q
L792Q
F881L
V895A
V980E
T984A
K1022E
M1073I
v4 +
VEGFA-2: TTTG PAM
Indel (%)
0
10
20
30
대조군
WT
v2
v3
v4
E837H
E837K
E837N
E837L
E837I
D533G
N644K
D680P
L741Q
L792Q
F881L
V895A
V980E
T984A
K1022E
M1073I
v4 +

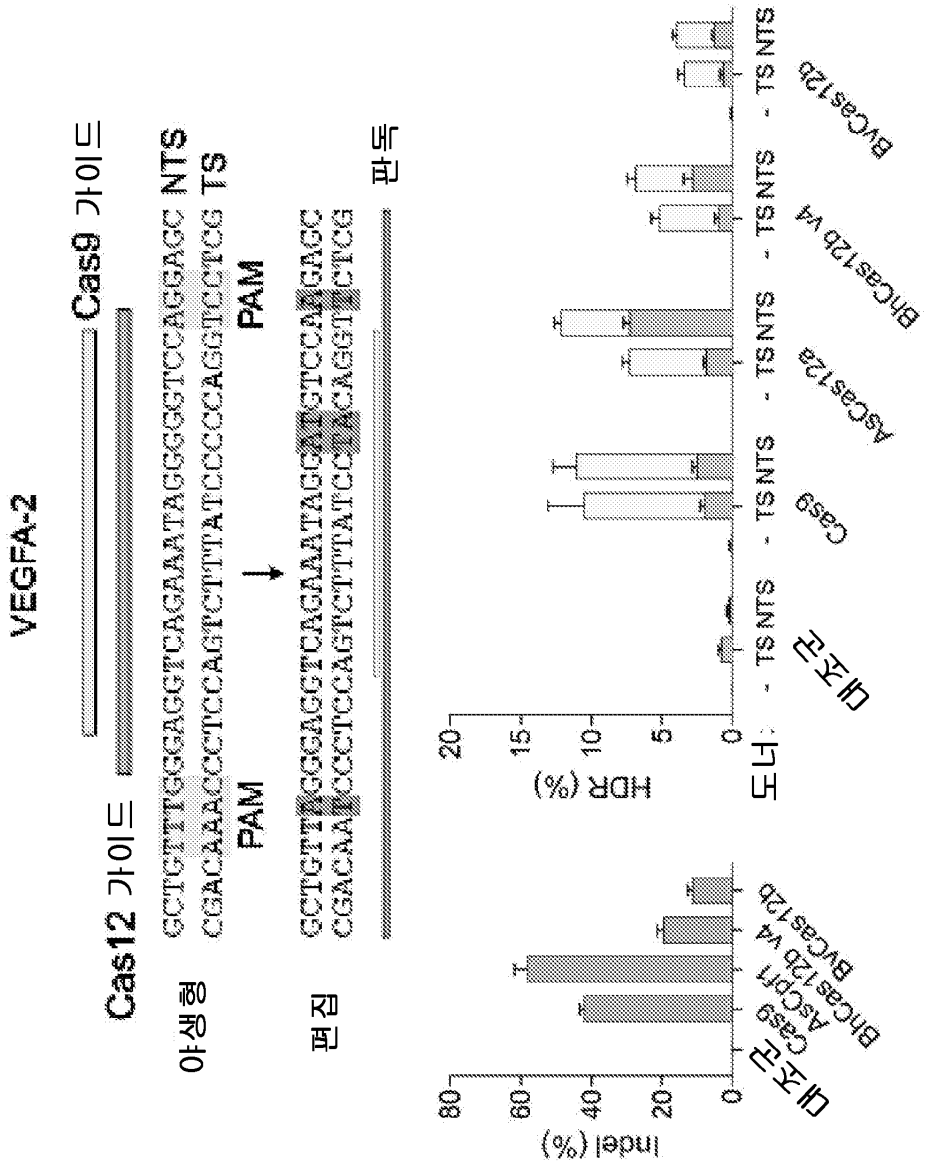
도면39b


DNMT1: ATTT PAM
Indel (%)
0
10
20
30
40
대조군
WT
v2
v3
v4
E837H
E837K
E837N
E837L
E837I
D533G
N64AK
D680P
L741Q
L792Q
F881L
V895A
V980E
T984A
K1022E
M1073I
v4 +
VEGFA: ATTC PAM
Indel (%)
0
20
40
60
대조군
WT
v2
v3
v4
E837H
E837K
E837N
E837L
E837I
D533G
N64AK
D680P
L741Q
L792Q
F881L
V895A
V980E
T984A
K1022E
M1073I
v4 +

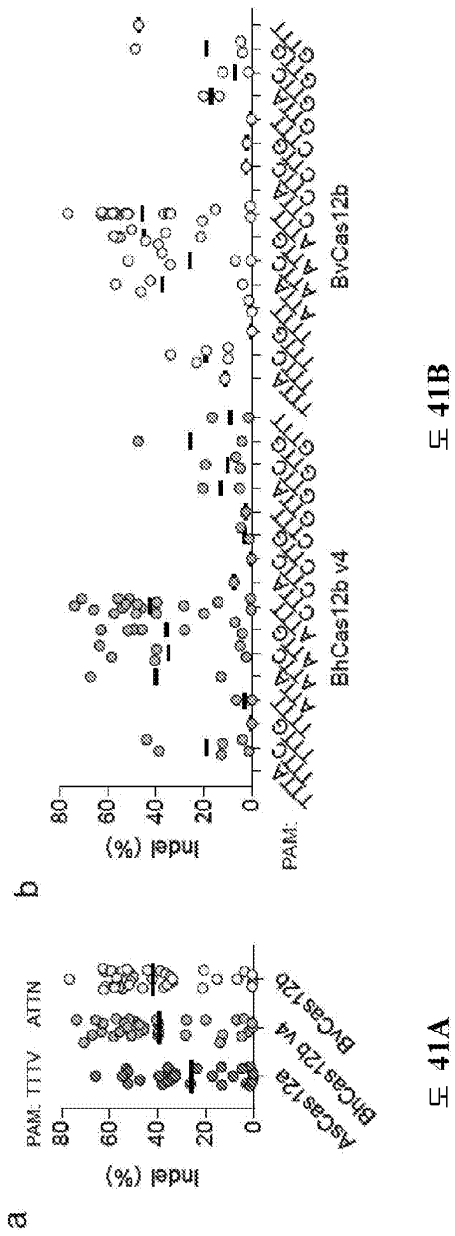
도면40a


DNMT1-1
Cas9 가이드
Cas12 가이드
야생형
편집
PAM
NTS
TS
편독
Indel (%)
HDR (%)
대조군
Cas9
AsCas12a
BhCas12b v4
BvCas12b
도너: - TS NTS - TS NTS - TS NTS - TS NTS

도면40b


VEGFA-2
Cas9 가이드
Cas12 가이드
야생형
NTS
TS
PAM
편집
편집
대조군
Cas9
AsCpf1
BnCas12b
BvCas12b
Indel (%)
HDR (%)
도너:
- TS NTS
Cas9
AsCas12a
BnCas12b V4
BvCas12b

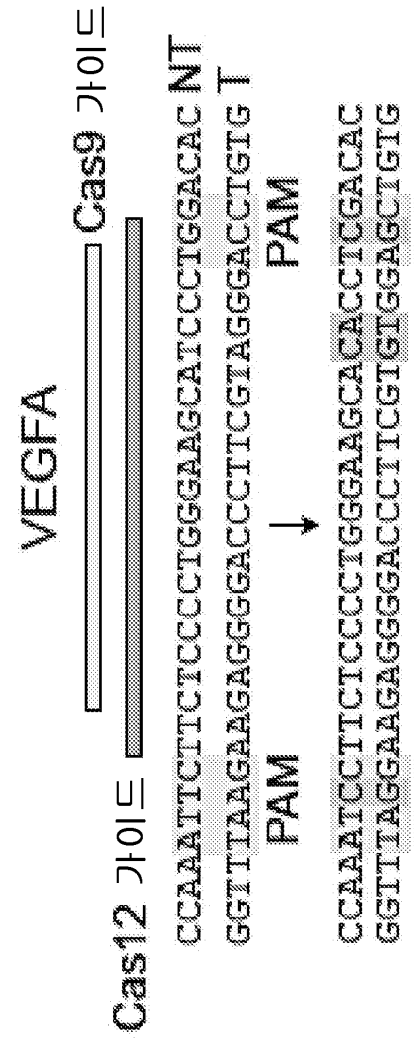
도면41



도 41A

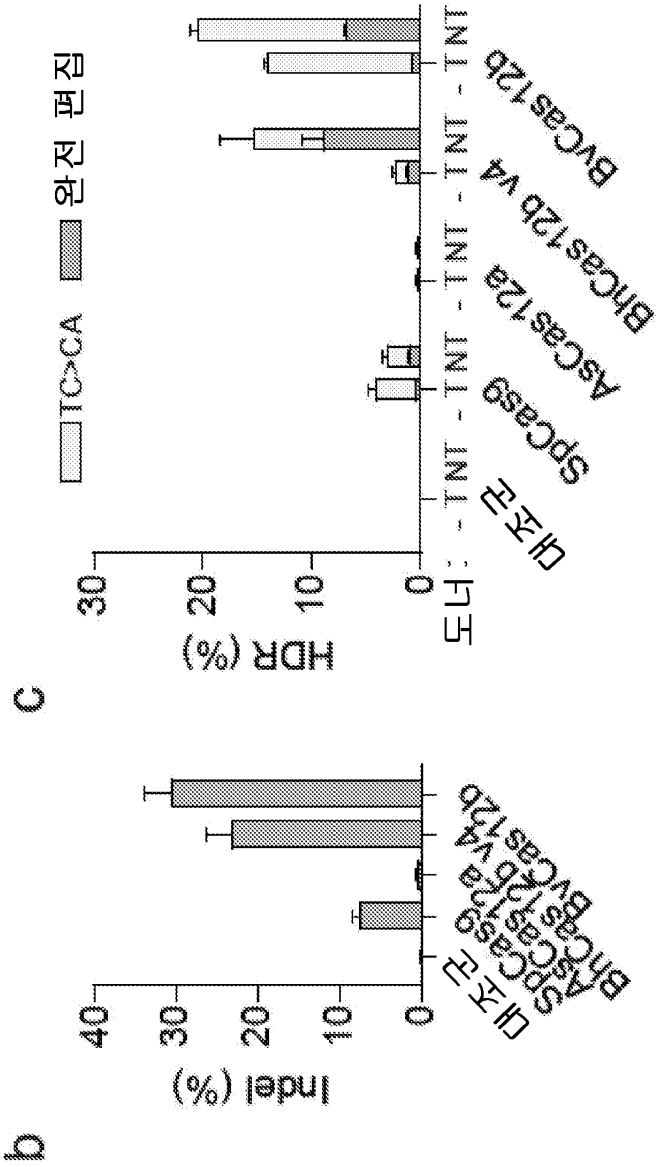
도 41B

도면42a


VEGFA
Cas9 가이드
Cas12 가이드
CCAATTCTTCTCCCCTGGGAAGCATCCCTGGACAC NT
GGTTTAAGAGAGGGGACCCTTCGTAGGGACCTGTG T
PAM
PAM
CCAATCCTTCTCCCCTGGGAAGCACACCTCGACAC
GGTTTAGGAAGAGGGGACCCTTCGTGTGGAGCTGTG

도 42 A

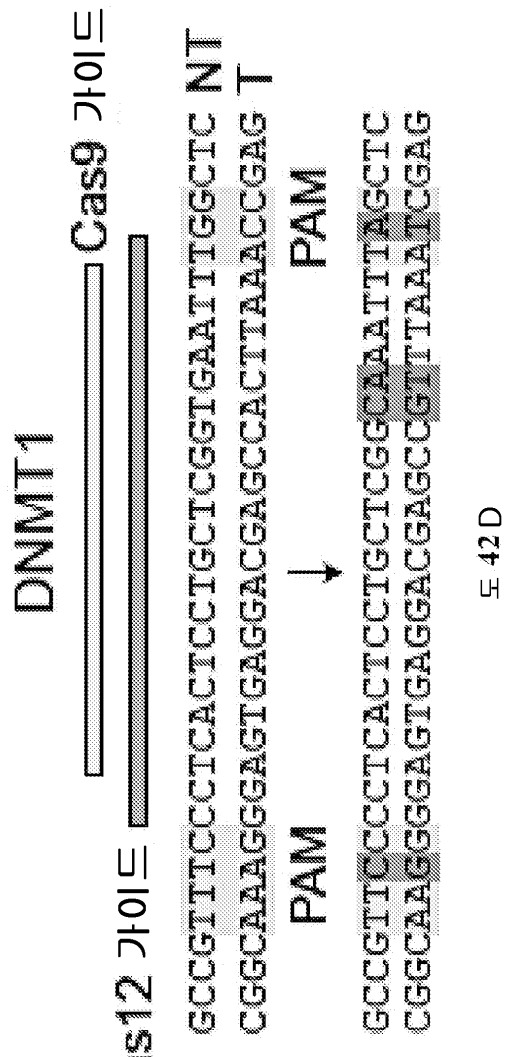
도면42b



도 42B

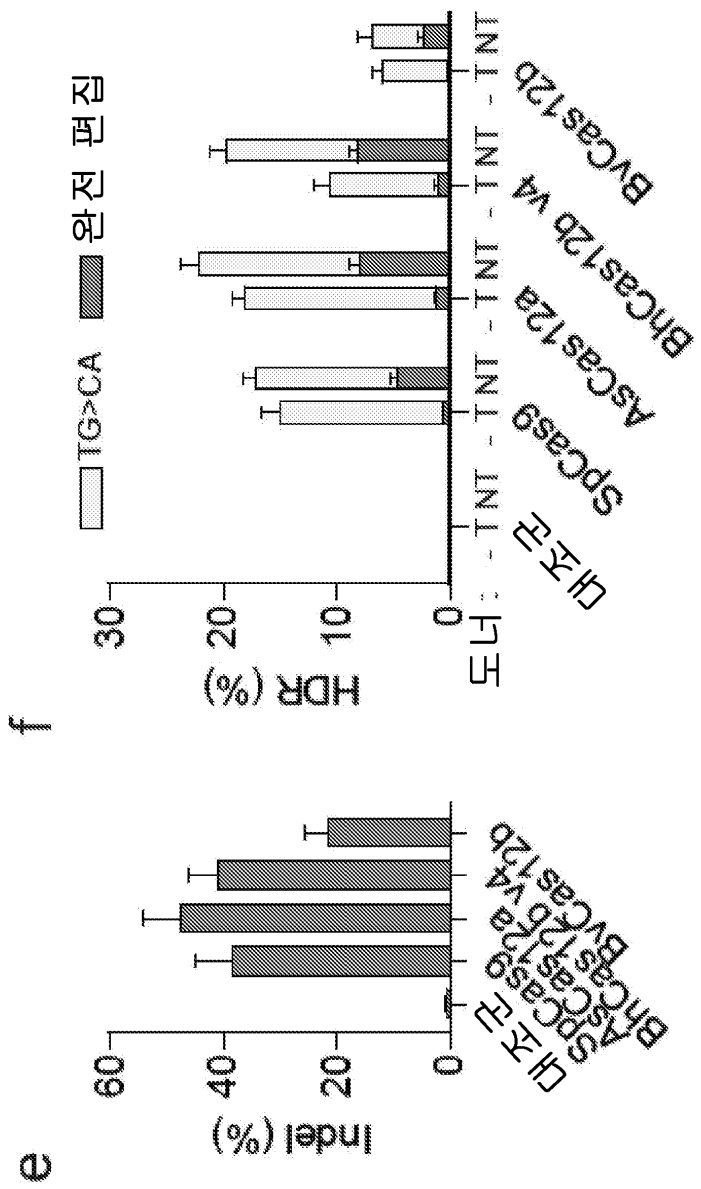
도 42C

도면42c


DNMT1
Cas9 가이드
Cas12 가이드
GCCGTTCCCTCACTCCTGCTCGGTGAATTTGGCTC NT
CGGCAAAGGGAGTGAGGACGAGCCACTTAAACCGAG T
PAM
PAM
GCCGTTCCCTCACTCCTGCTCGGCAAATTTAGCTC
CGGCAAGGGAGTGAGGACGAGCCGTTTAAATCGAG

도 42D

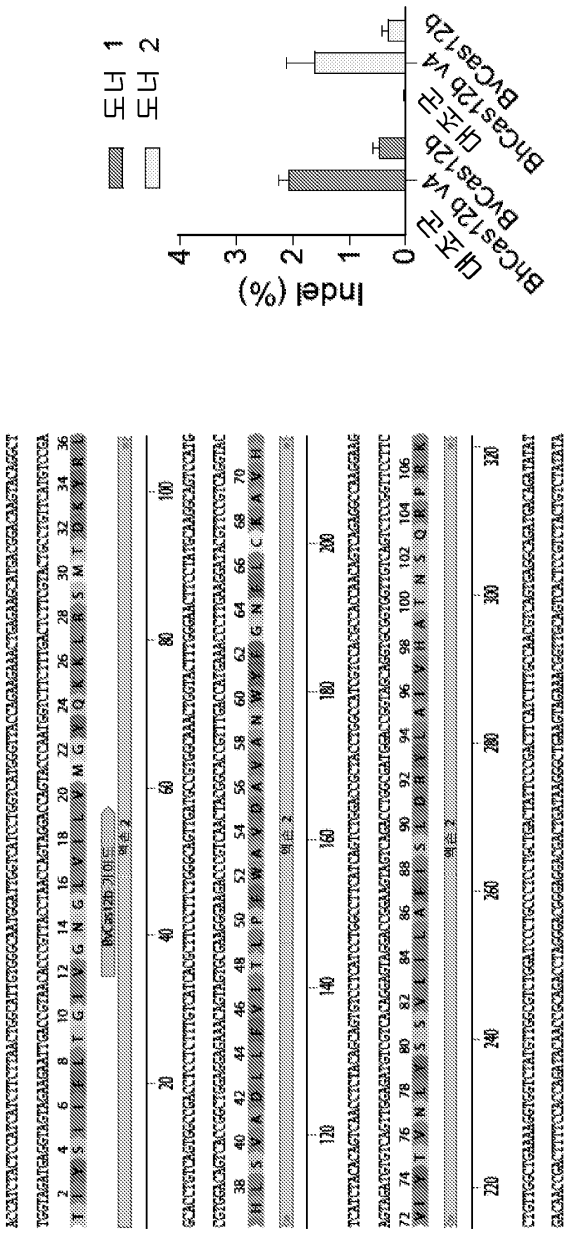
도면42d



도 42F

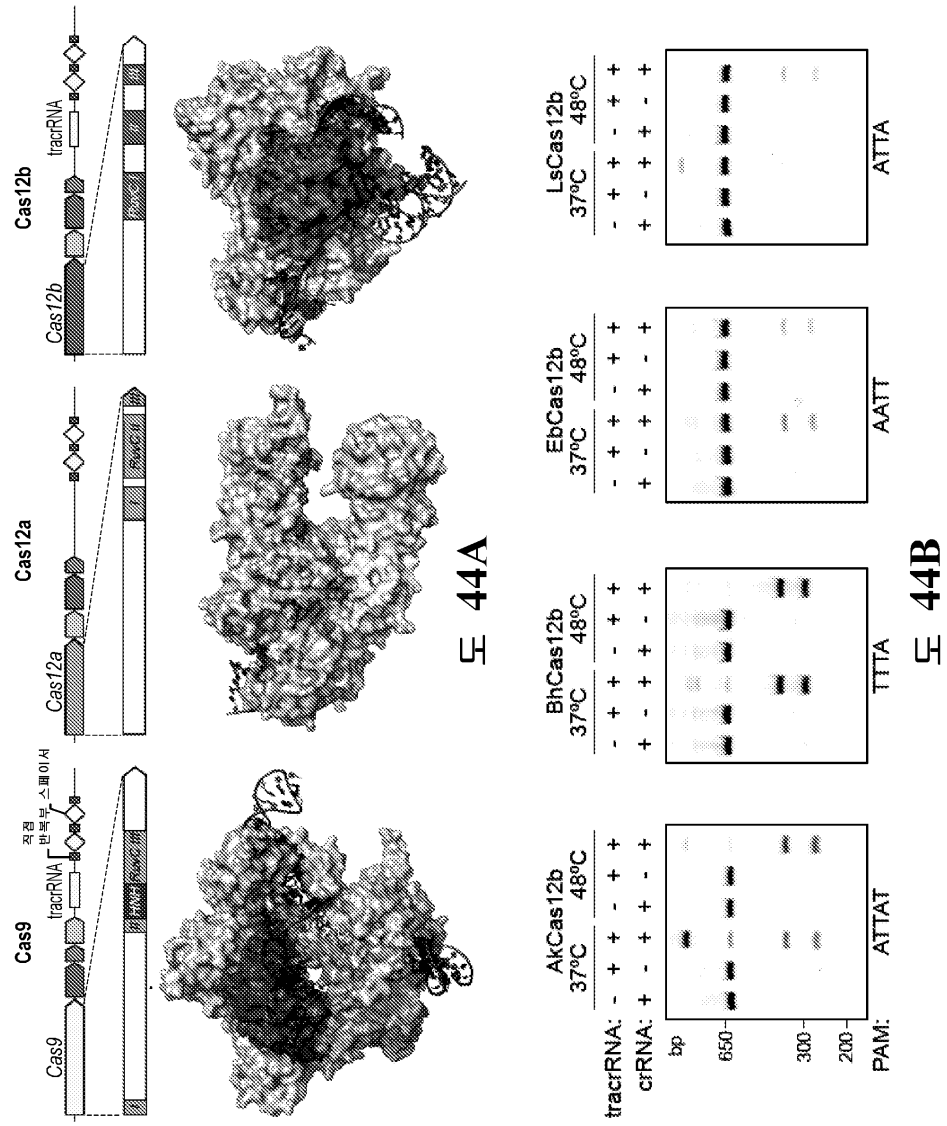
도 42E

도면43


CXCR4 엑손 2
BhCas12b 가이드
엑손 2
BvCas12b 가이드
BhCas12b 표적 : CCGACTTCATCTTTGCCAACGTC
BvCas12b 표적 : TGGGCAATGGATTGGTCATCCTG

도너 1
도너 2
Indel (%)
0
1
2
3
4
대조군
BhCas12b Vs
BvCas12b
대조군
BhCas12b Vs
BvCas12b

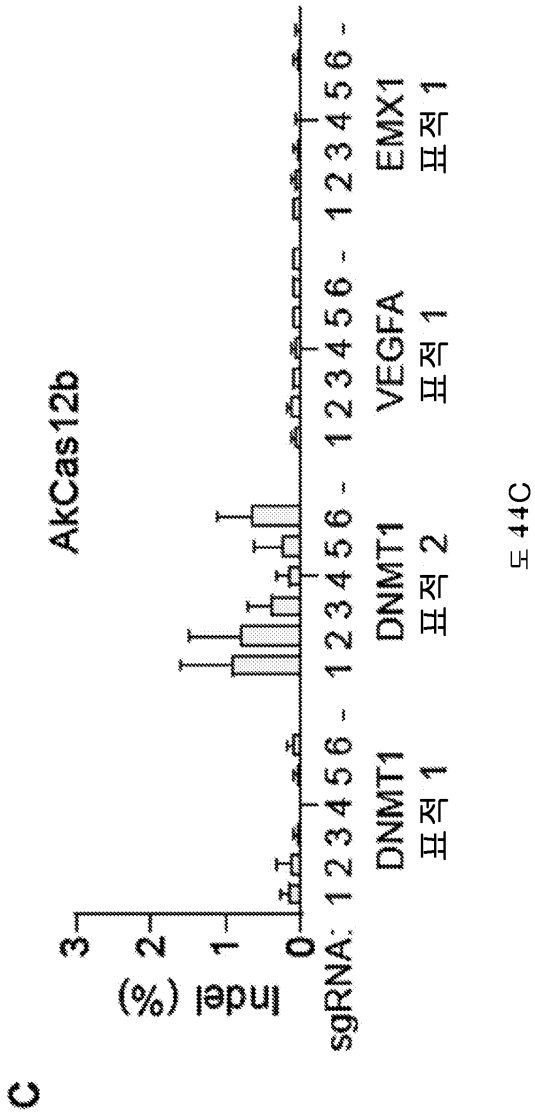
도면44a



도 44A

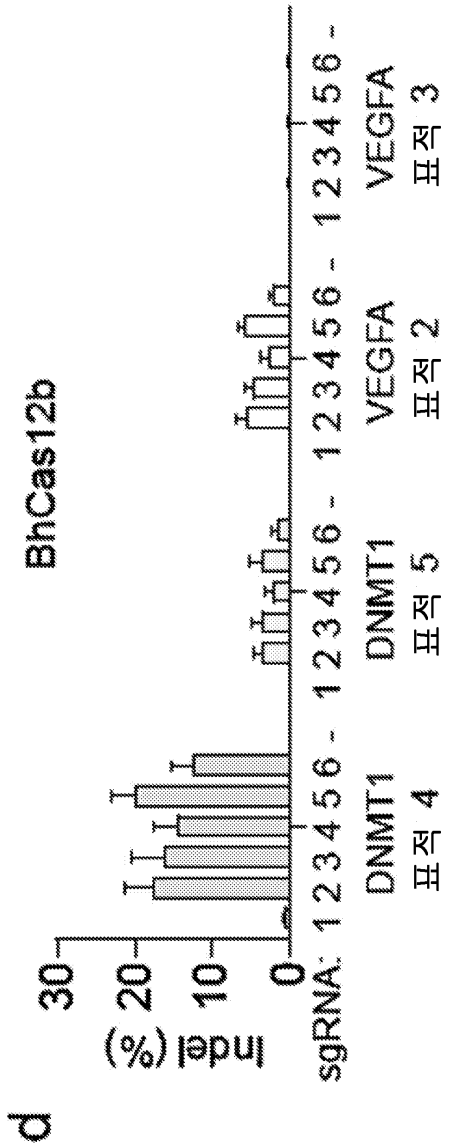
도 44B

도면44b


C
AkCas12b
Indel (%)
0
1
2
3
sgRNA: 1 2 3 4 5 6 - 1 2 3 4 5 6 - 1 2 3 4 5 6 - 1 2 3 4 5 6 -
DNMT1 표적 1
DNMT1 표적 2
VEGFA 표적 1
EMX1 표적 1

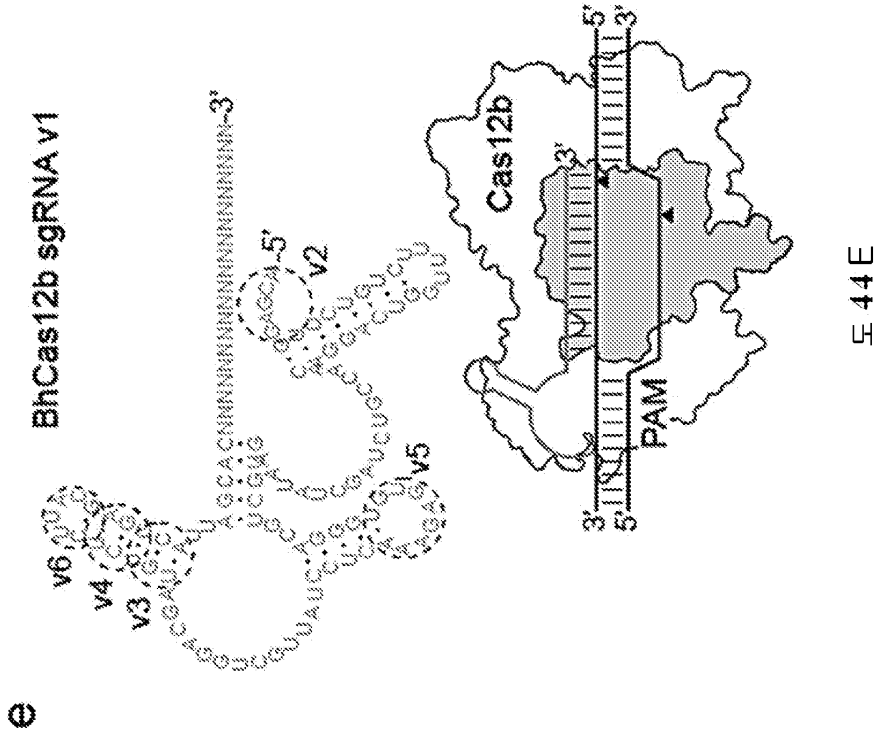
도 44C

도면44c


d
BhCas12b
Indel (%)
30
20
10
0
sgRNA: 1 2 3 4 5 6 - 1 2 3 4 5 6 - 1 2 3 4 5 6 - 1 2 3 4 5 6 -
DNMT1 표적 4
DNMT1 표적 5
VEGFA 표적 2
VEGFA 표적 3

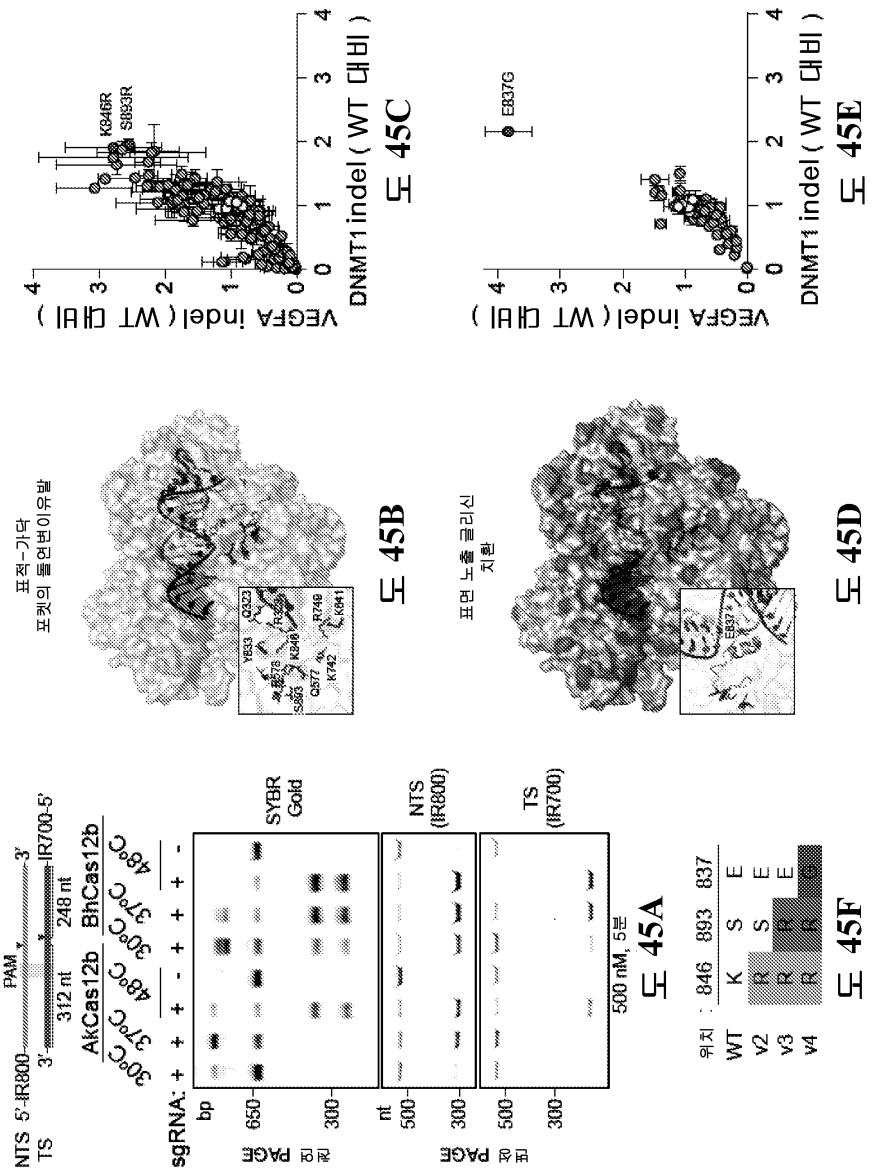
도 44D

도면44d

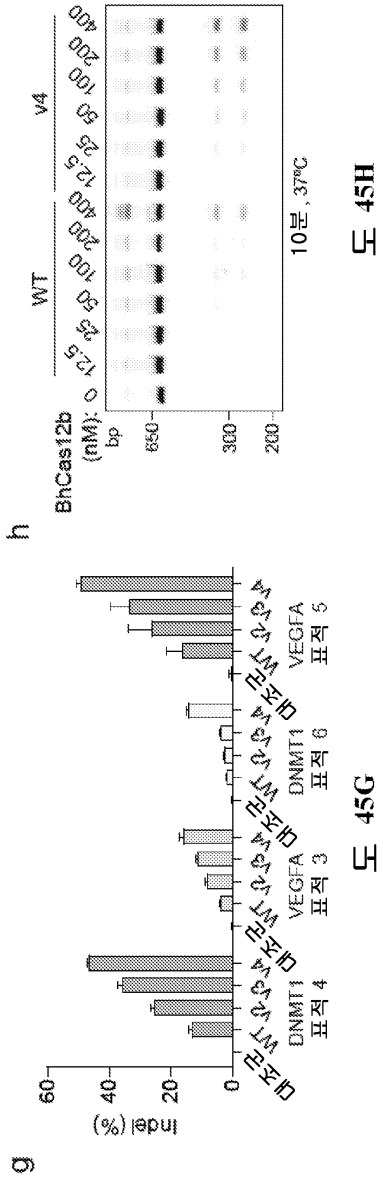


도 44E

도면45a


NTS 5'-IR800
TS 3'
PAM
3'
IR700-5'
312 nt
248 nt
AkCas12b
BhCas12b
30°C
37°C
48°C
sgRNA: + + + - + + + -
SYBR Gold
NTS (IR800)
TS (IR700)
bp
nt
650
300
500
500 nM, 5분
위치 : 846 893 837
WT K S E
v2 R S E
v3 R R E
v4 R R G
표적-가닥
포켓의 돌연변이유발
Y833
Q323
R778
K346
S893
Q377
K742
R749
K641
표면 노출 글리신
치환
E837G
VEGFA indel (WT 대비)
DNMT1 indel (WT 대비)
K846R
S893R
E837G
도 45A
도 45B
도 45C
도 45D
도 45E
도 45F

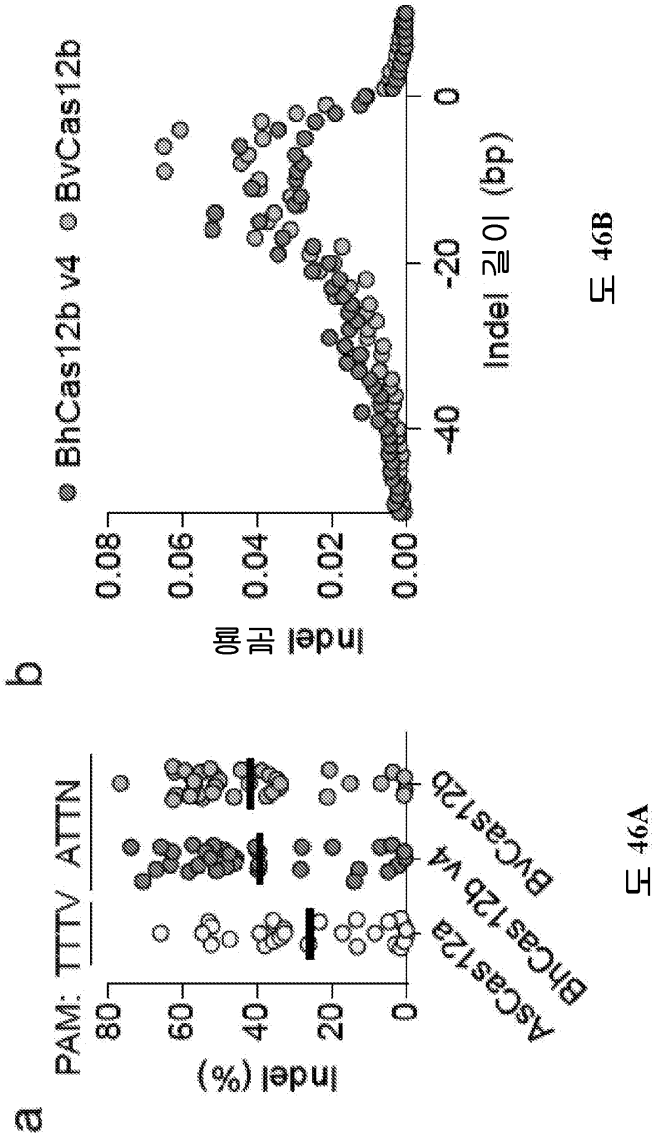
도면45b



도 45H

도 45G

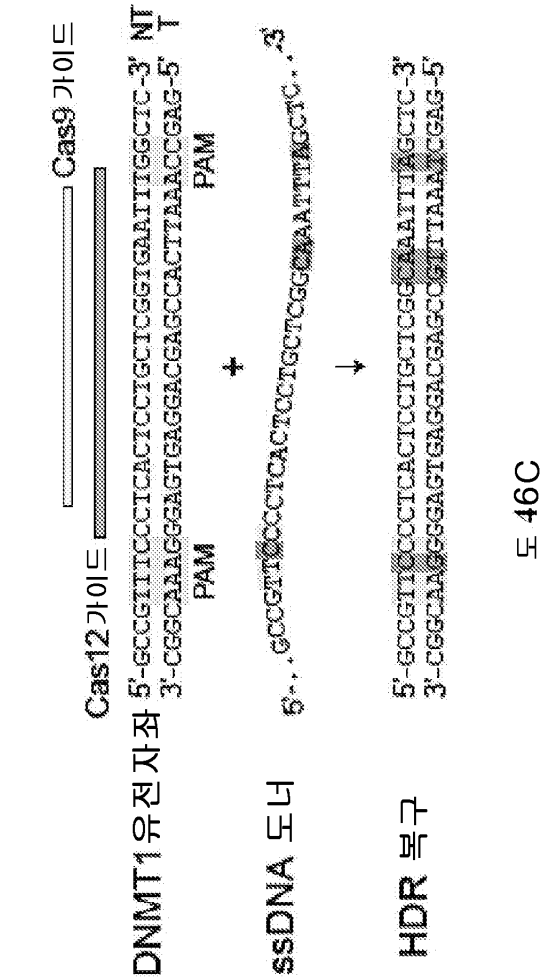
도면46a



도 46A

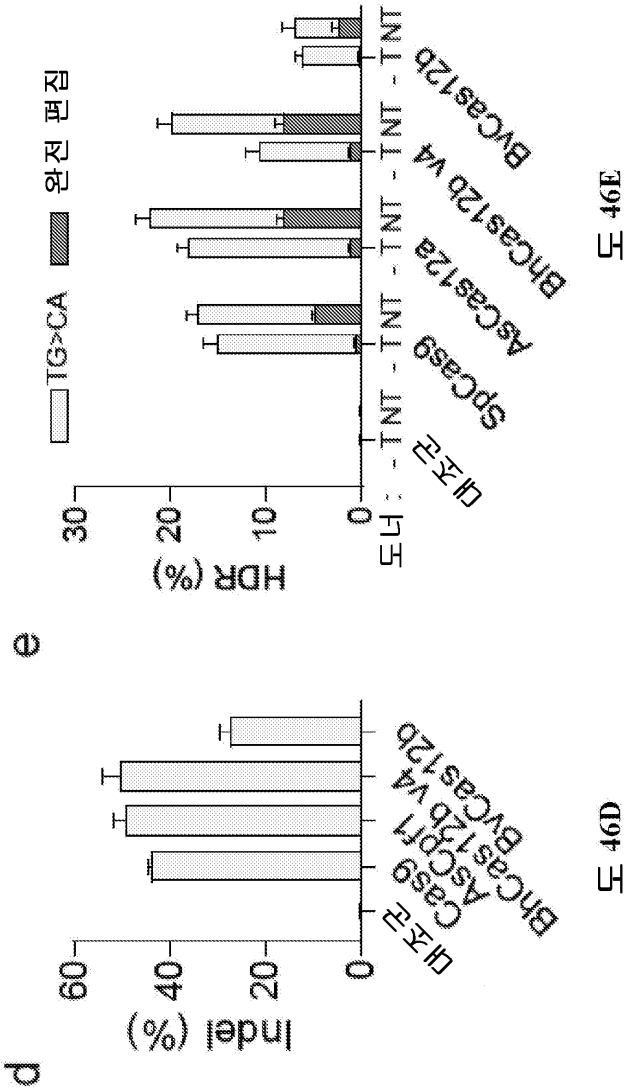
도 46B

도면46b


C
Cas9 가이드
Cas12 가이드
DNMT1 유전자좌
5'-GCCGTTCCCTCACTCCTGCTCGGTGAATTTGGCTC-3' NT
3'-CGGCAAGGGAGTGAGGACGAGCCACTTAAACCGAG-5' T
PAM
PAM
+
ssDNA 도너
5'-...GCCGTTCCCTCACTCCTGCTCGGTGAATTTGGCTC...-3'
HDR 복구
5'-GCCGTTCCCTCACTCCTGCTCGGTGAATTTGGCTC-3'
3'-CGGCAAGGGAGTGAGGACGAGCCACTTAAACCGAG-5'

도 46C

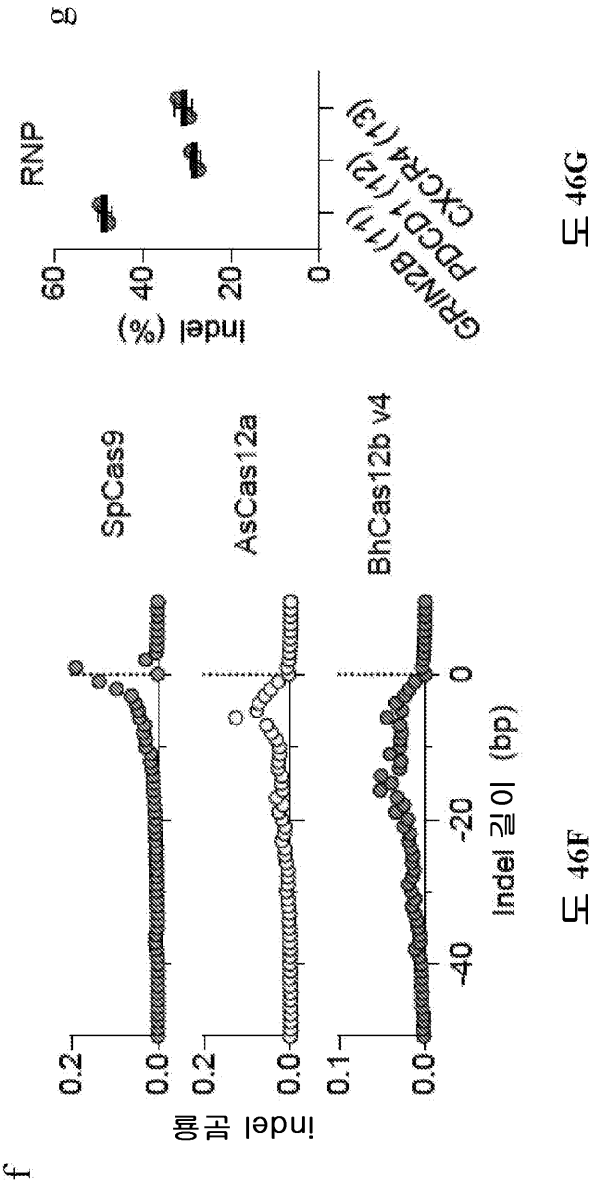
도면46c



도 46E

도 46D

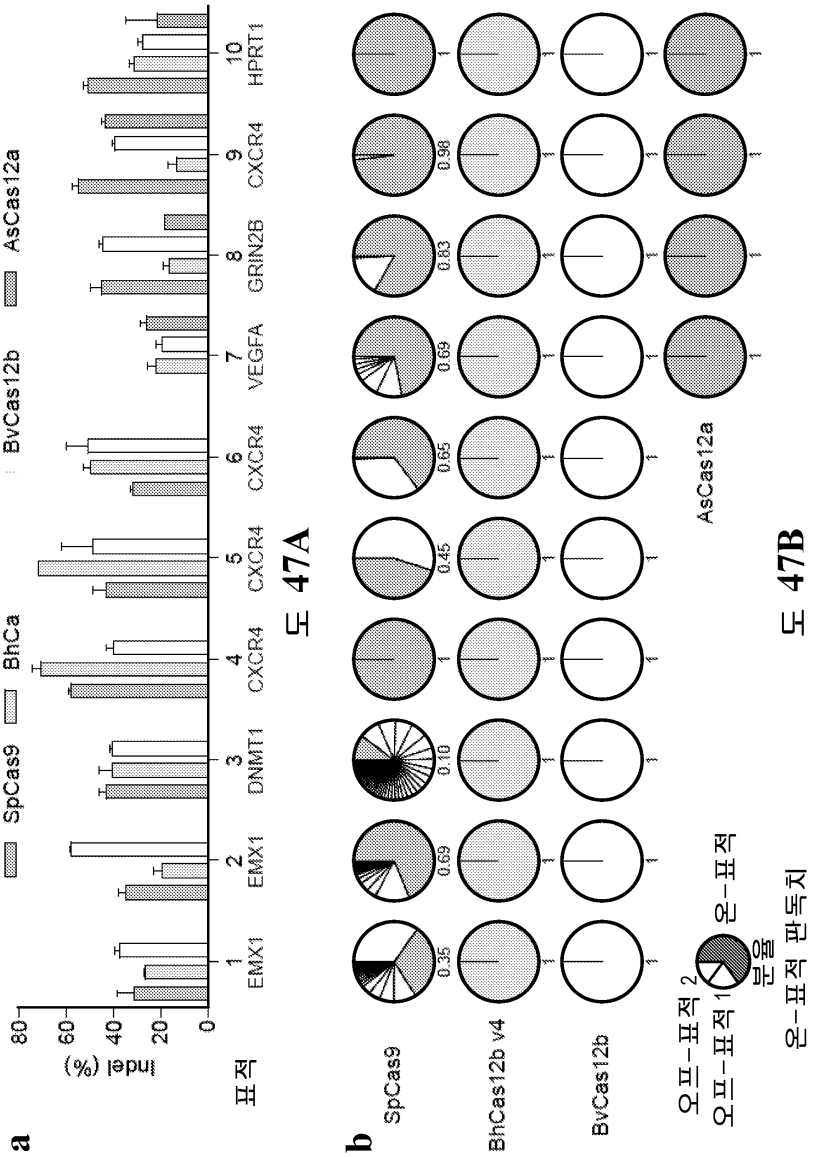
도면46d



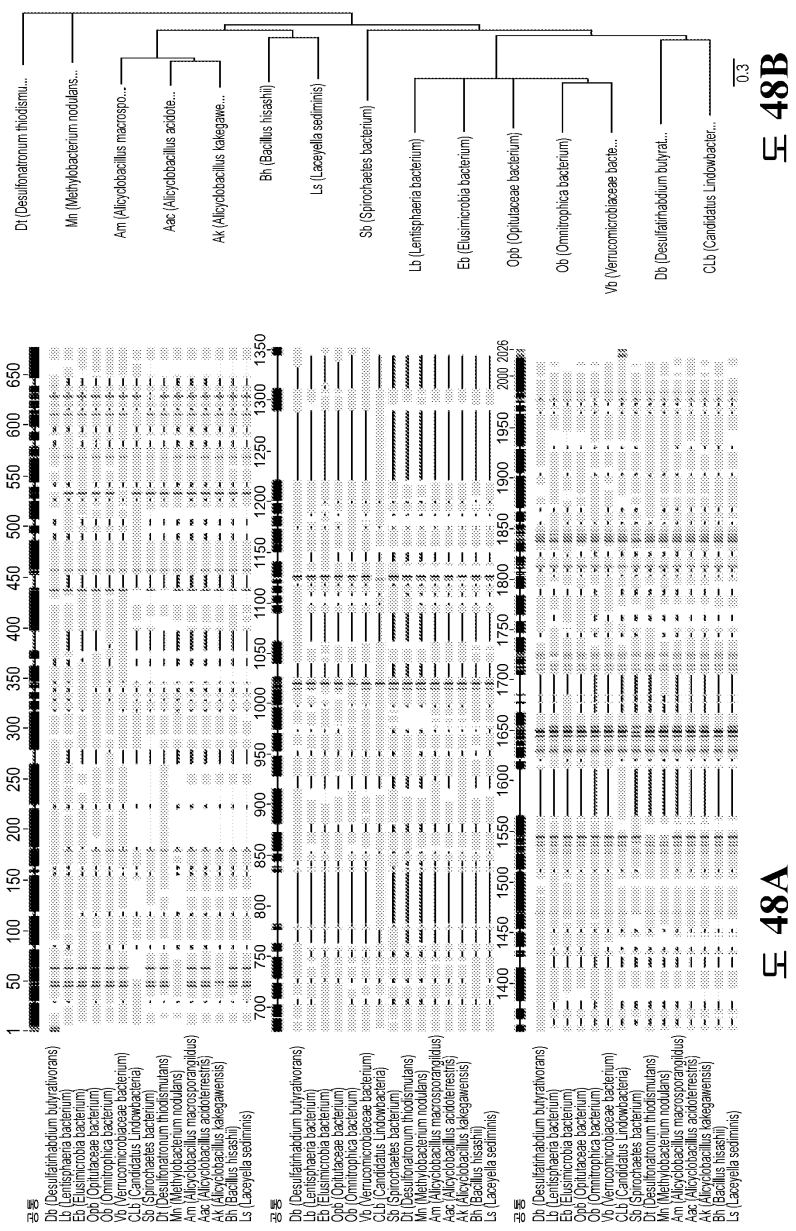
도 46F

도 46G

도면47


a
SpCas9
BhCa
BvCas12b
AsCas12a
indel (%)
80
60
40
20
0
표적
1 EMX1
2 EMX1
3 DNMT1
4 CXCR4
5 CXCR4
6 CXCR4
7 VEGFA
8 GRIN2B
9 CXCR4
10 HPRT1
도 47A
b
SpCas9
BhCas12b v4
BvCas12b
AsCas12a
0.35
0.69
0.10
1
0.45
0.65
0.69
0.83
0.98
오프-표적 2
오프-표적 1
온-표적
분율
온-표적 판독치
도 47B

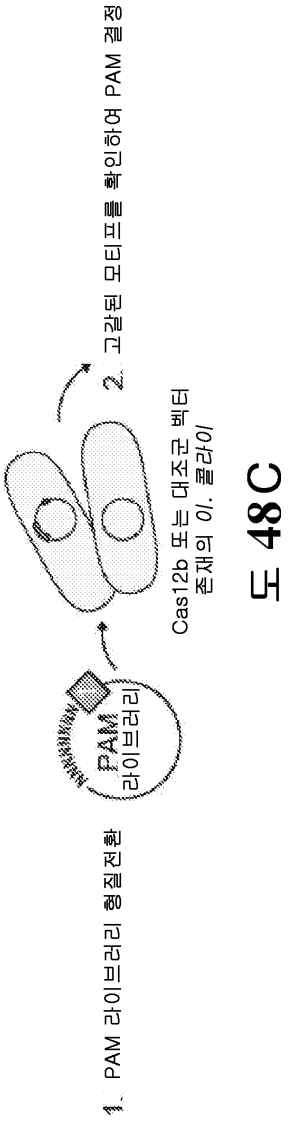
도면48a



도 48A

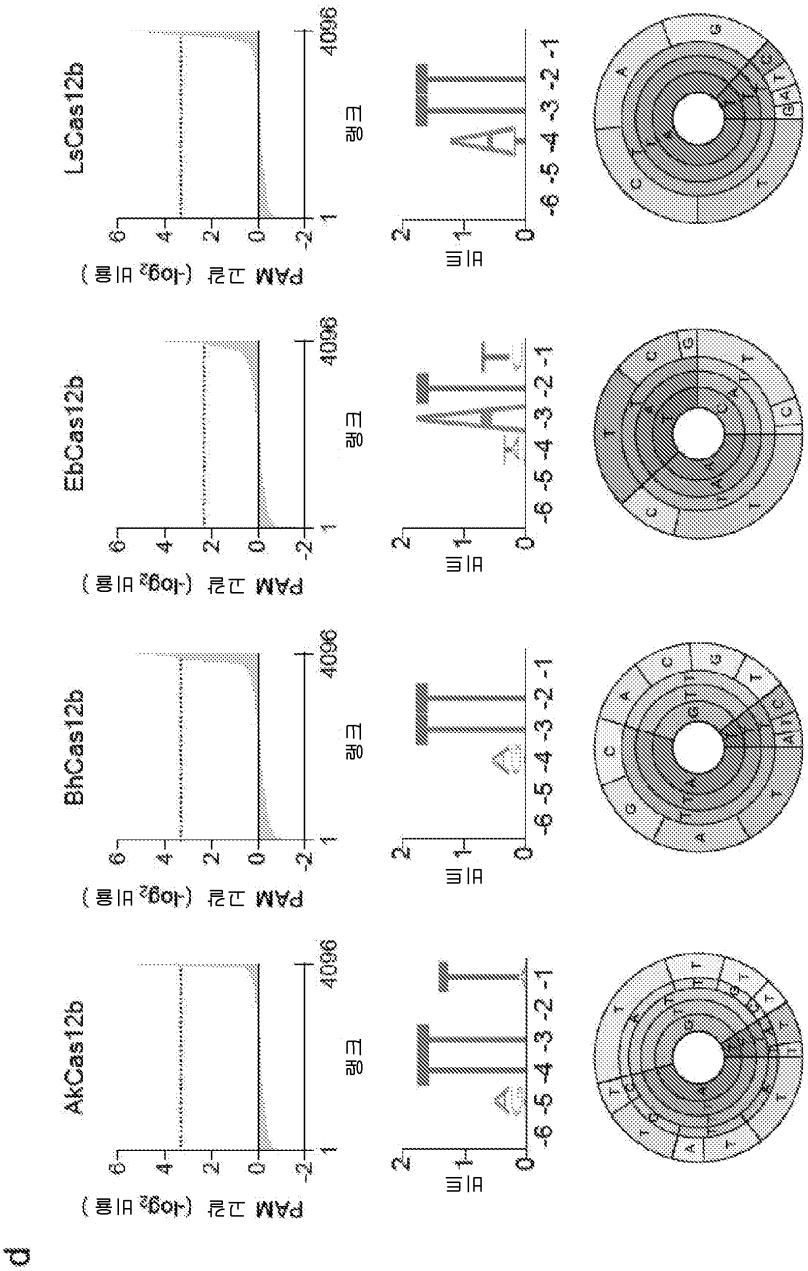
도 48B

도면48b


1. PAM 라이브러리 형질전환
PAM 라이브러리
Cas12b 또는 대조군 벡터
존재의 01. 클라이
2. 고갈된 모티프를 확인하여 PAM 결정

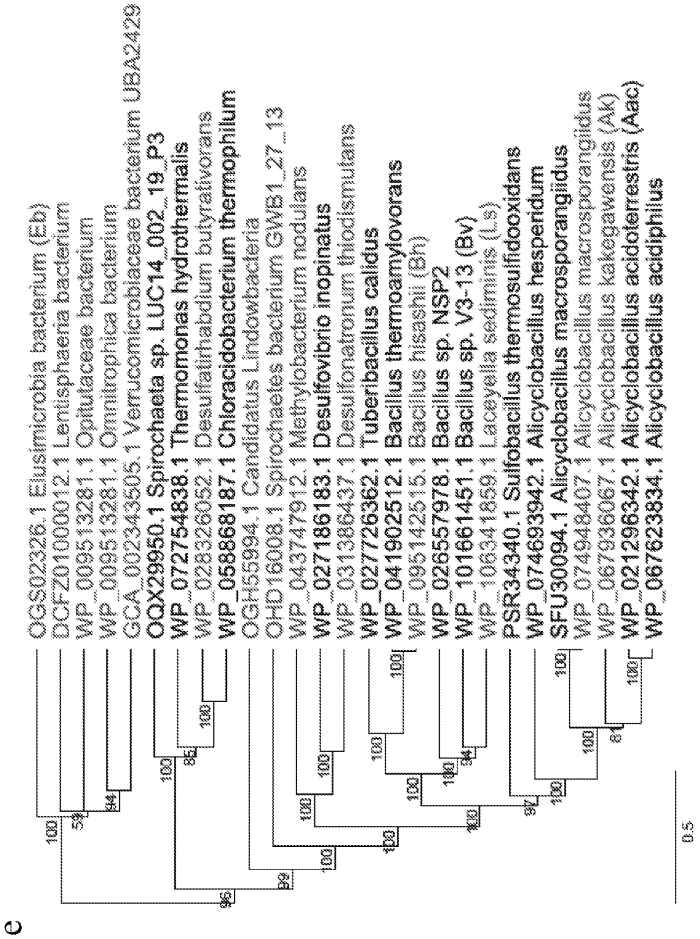
도 48C

도면48c


d
AkCas12b
BhCas12b
EbCas12b
LsCas12b
PAM 고갈 (-log2 비율)
랭크
비트
4096
1
-2
0
2
4
6
-6 -5 -4 -3 -2 -1

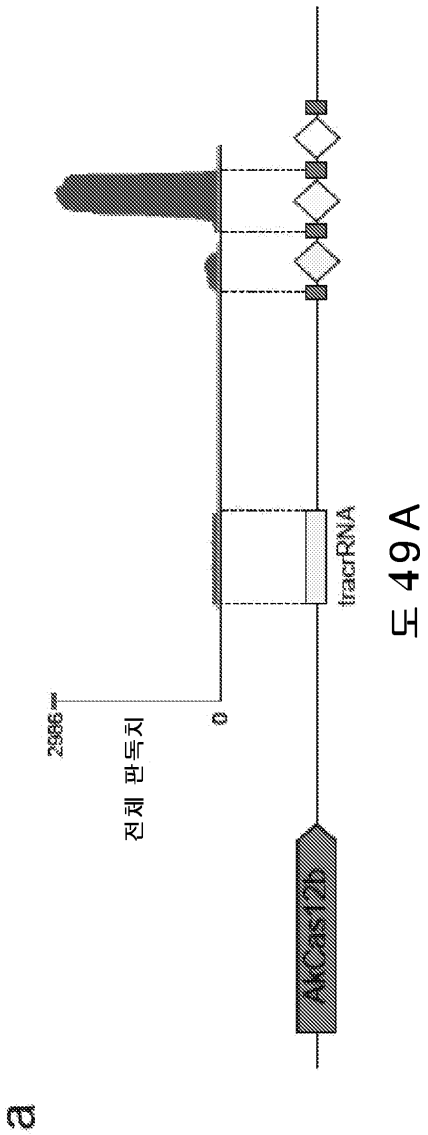
도 48D

도면48d


e
OGS02326.1 Elusimicrobia bacterium (Eb)
DCFZ01000012.1 Lentisphaera bacterium
WP_009513281.1 Opitutaceae bacterium
WP_009513281.1 Omnitrophica bacterium
GCA_002343505.1 Verrucomicrobiaceae bacterium UBA2429
OQX29950.1 Spirochaeta sp. LUC14_002_19_P3
WP_072754838.1 Thermomonas hydrothermalis
WP_028326052.1 Desulfatirhabdium butyrativorans
WP_058868187.1 Chloracidobacterium thermophilum
OGH55994.1 Candidatus Lindowbacteria
OHD16008.1 Spirochaetes bacterium GWB1_27_13
WP_043747912.1 Methylobacterium nodulans
WP_027186183.1 Desulfovibrio inopinatus
WP_031386437.1 Desulfonatronum thiodismutans
WP_027726362.1 Tuberibacillus calidus
WP_041902512.1 Bacillus thermoamylovorans
WP_095142515.1 Bacillus hisashii (Bh)
WP_026557978.1 Bacillus sp. NSP2
WP_101661451.1 Bacillus sp. V3-13 (Bv)
WP_106341859.1 Laceyella sediminis (LS)
PSR34340.1 Sulfobacillus thermosulfidooxidans
WP_074693942.1 Alicyclobacillus hesperidum
SFU30094.1 Alicyclobacillus macrosporangidus
WP_074948407.1 Alicyclobacillus macrosporangidus
WP_067936067.1 Alicyclobacillus kakegawensis (Ak)
WP_021296342.1 Alicyclobacillus acidoterrestris (Aac)
WP_067623834.1 Alicyclobacillus acidiphilus
0.5

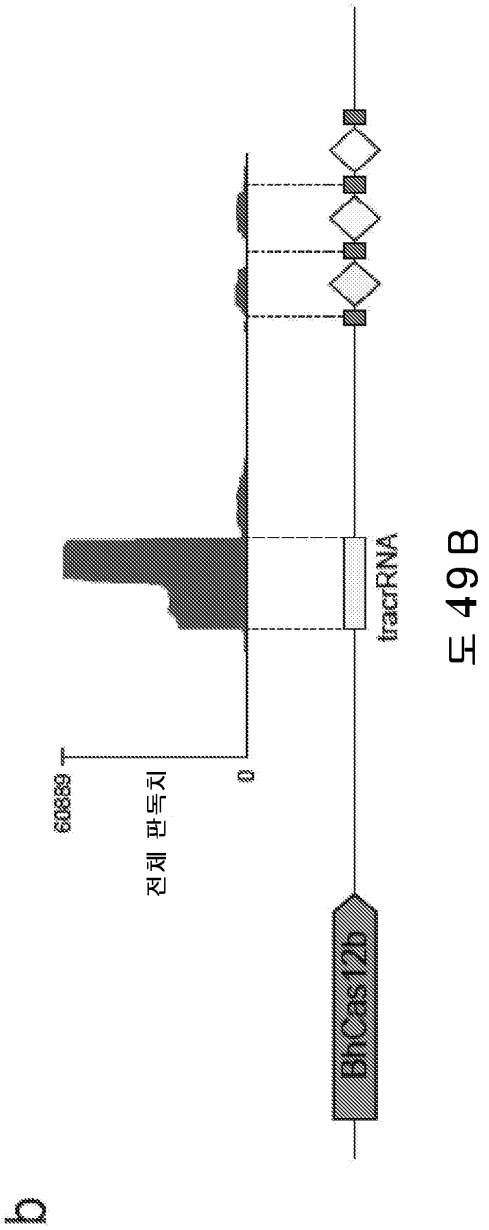
도 48E

도면49a



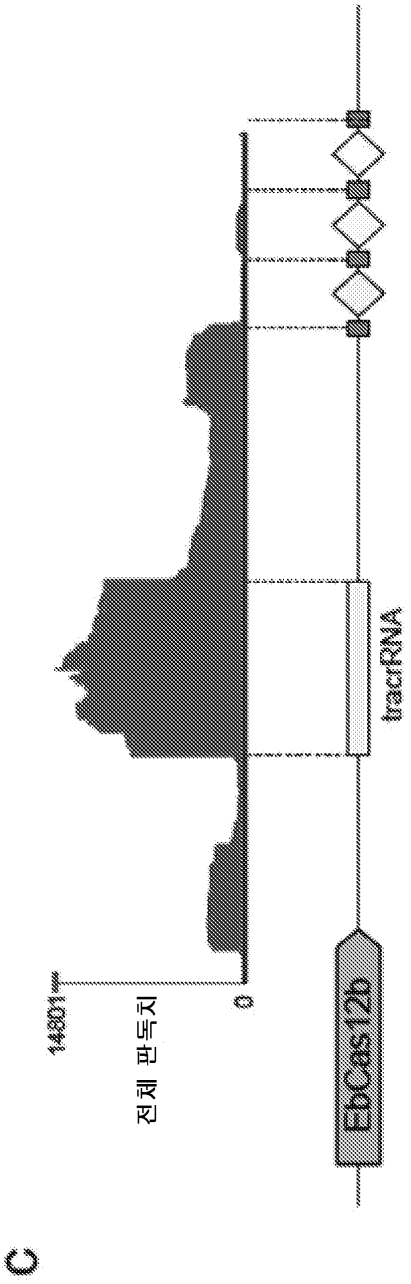
도 49A

도면49b



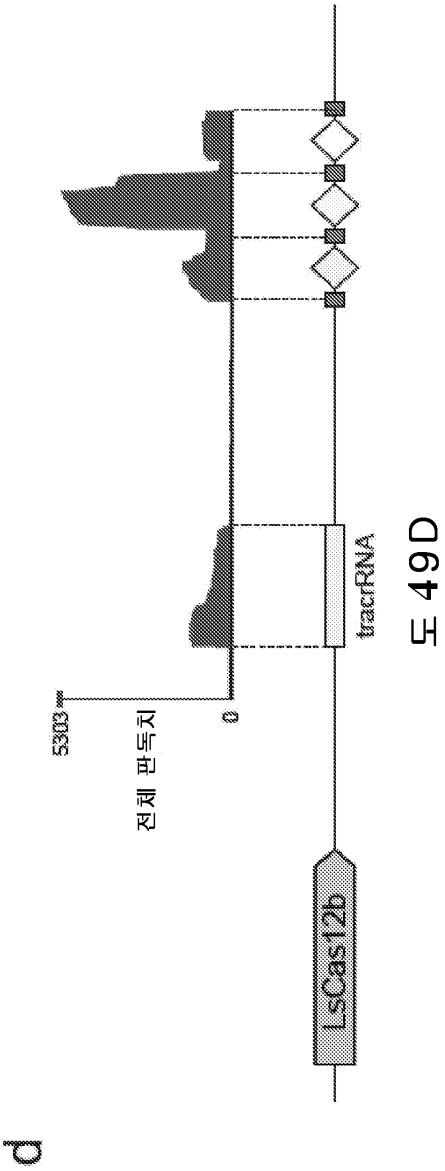
도 49B

도면49c



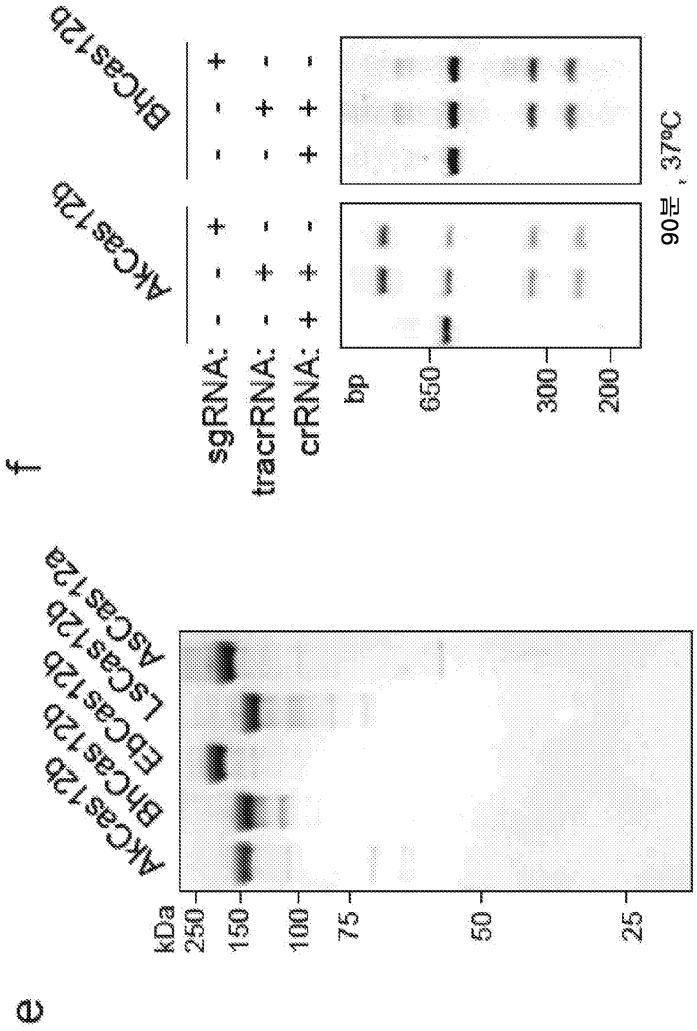
도 49C

도면49d



도 49D

도면49e



도 49E

도 49F

도면50a

도 50A

도 50B

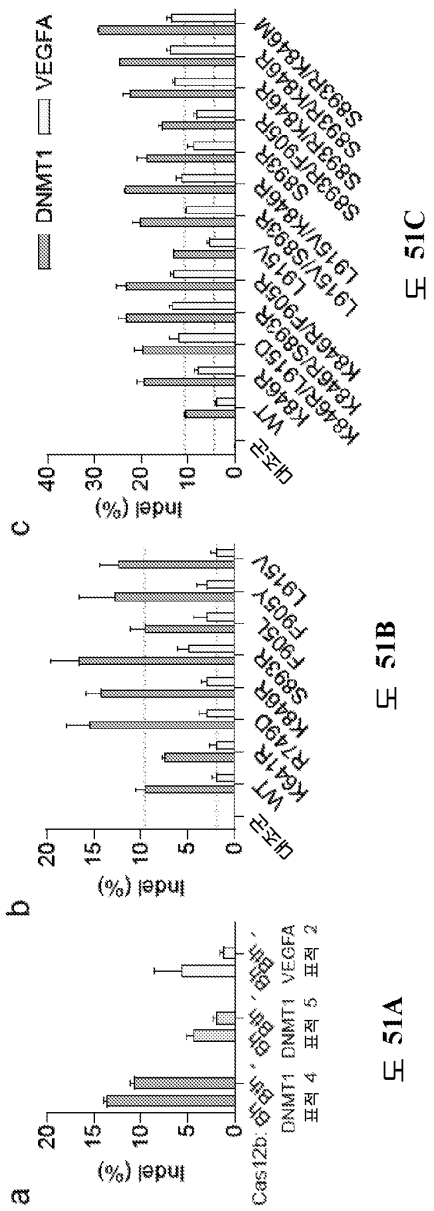
도 50C

도면50b

도 50D

도 50E

도면51a

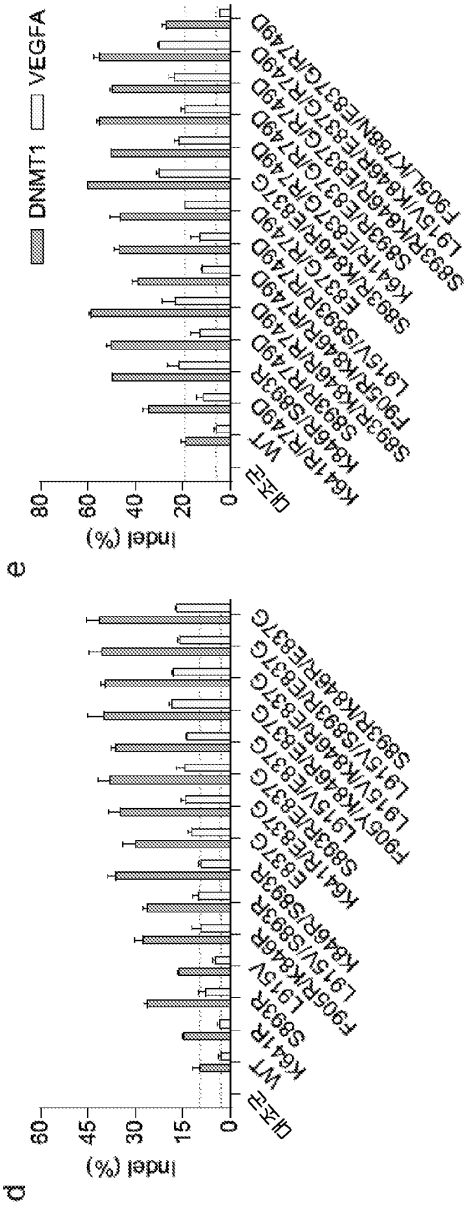


도 51A

도 51B

도 51C

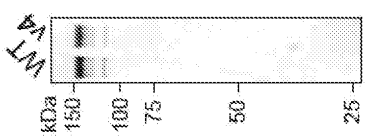
도면51b



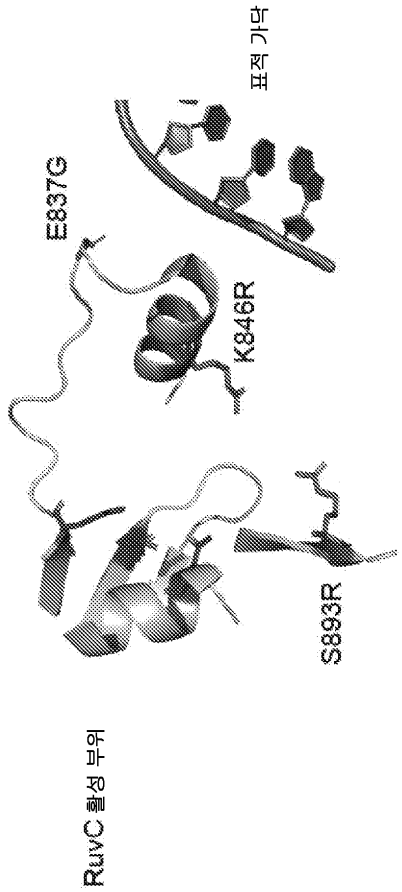
도 51D

도 51E

도면51c

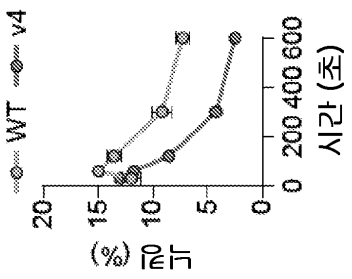


도 51G

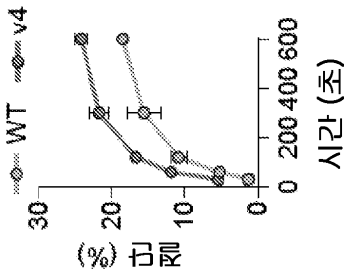


도 51F

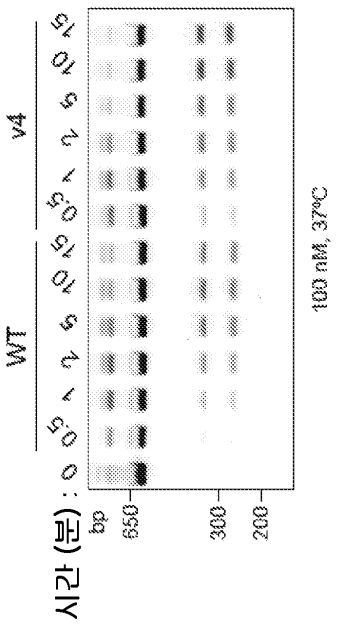
도면51d



도 51J

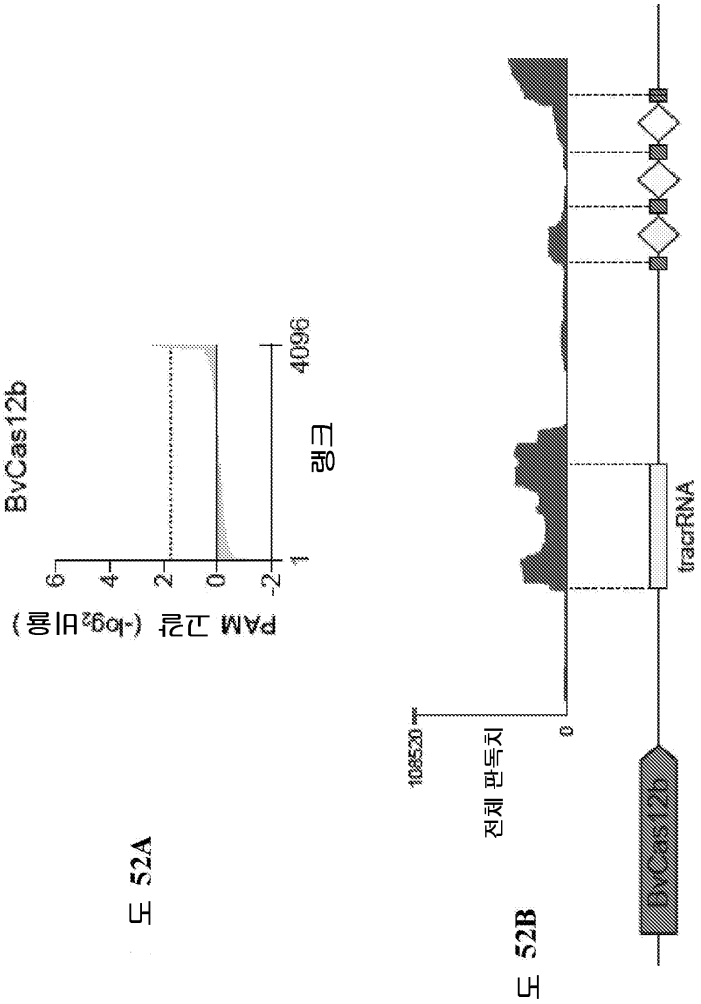


도 51I

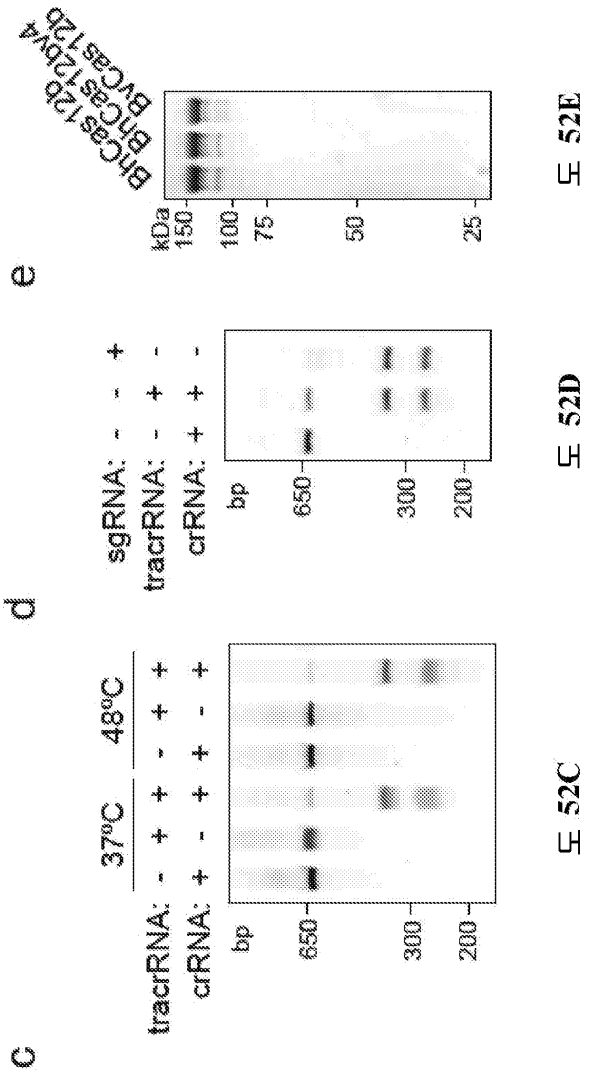


도 51H

도면52a


BvCas12b
PAM 고갈 (-log2배율)
6
4
2
0
-2
1
4096
랭크
도 52A
188520
0
전체 판독치
BvCas12b
tracrRNA
도 52B

도면52b



도 52C 도 52D 도 52E

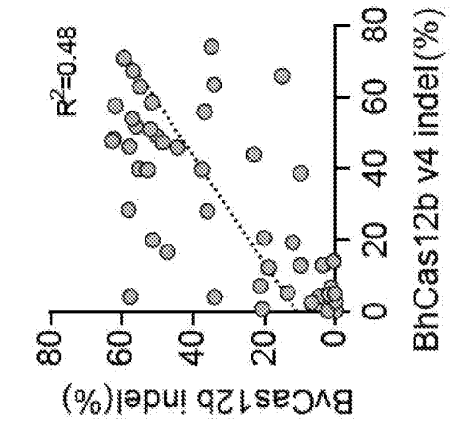
도면52c

도 52F

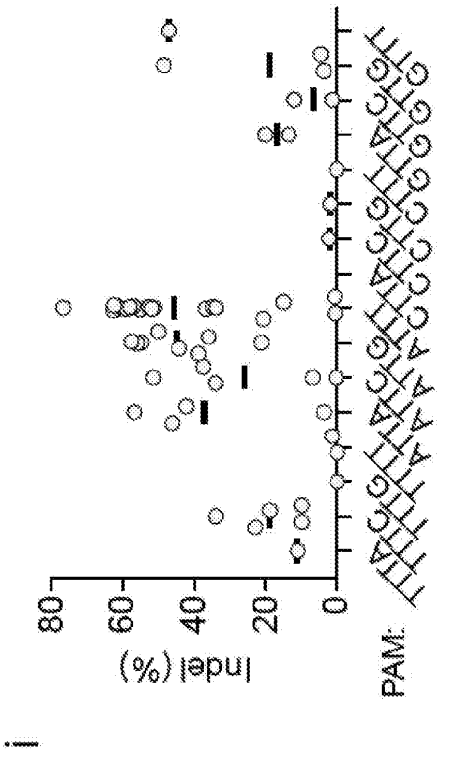
도 52G

도 52H

도면52d

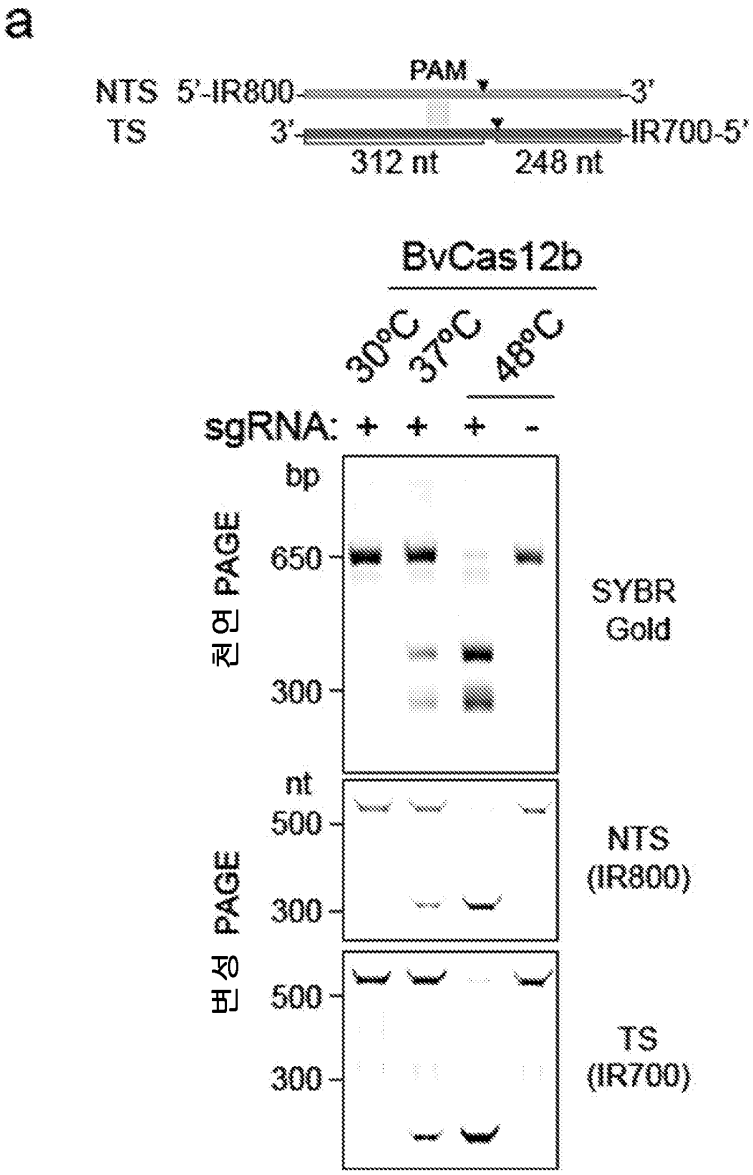


도 52J



도 52I

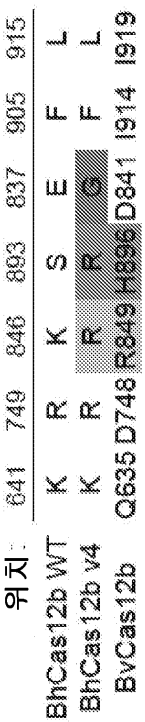
도면53a


a
NTS 5'-IR800
PAM
3'
TS
3'
IR700-5'
312 nt
248 nt
BvCas12b
30°C
37°C
48°C
sgRNA: + + + -
bp
650
300
천연 PAGE
SYBR Gold
nt
500
300
NTS (IR800)
변성 PAGE
500
300
TS (IR700)

도 53A

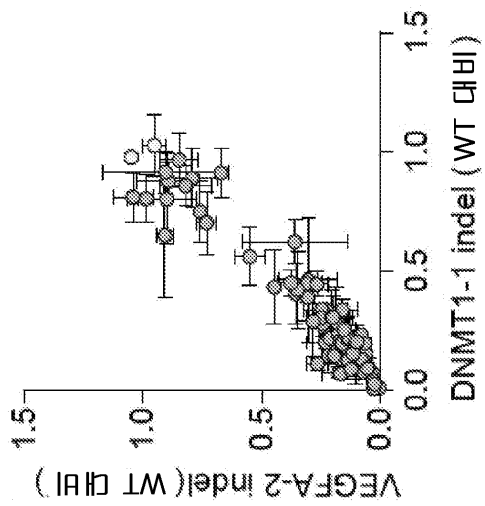
도면53b

b



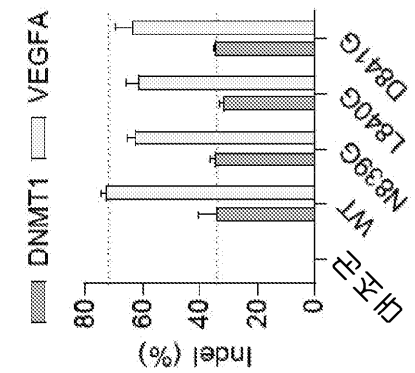
도 53B

c

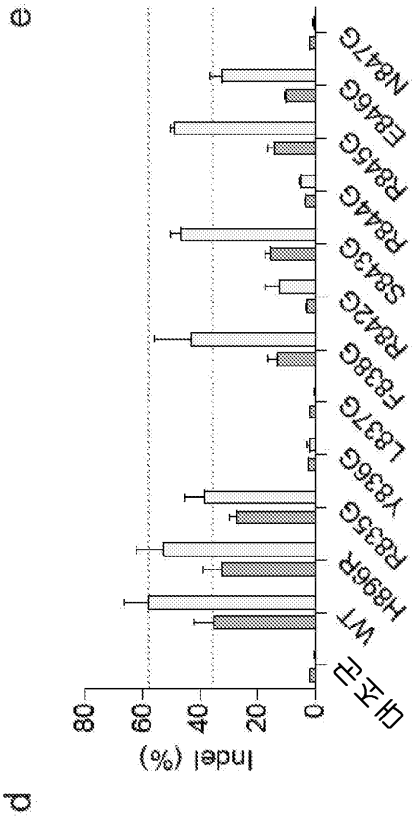


도 53C

도면53c

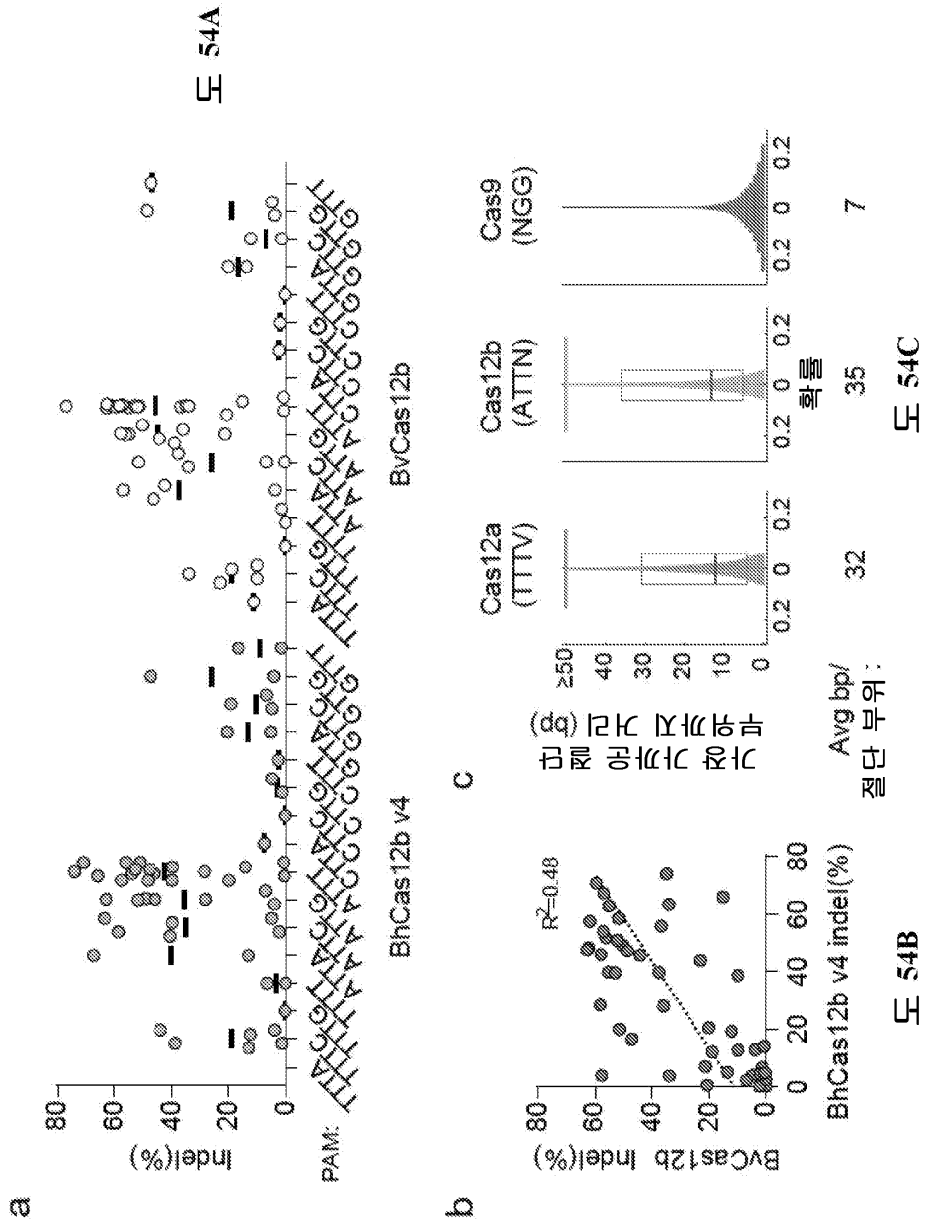


도 53E

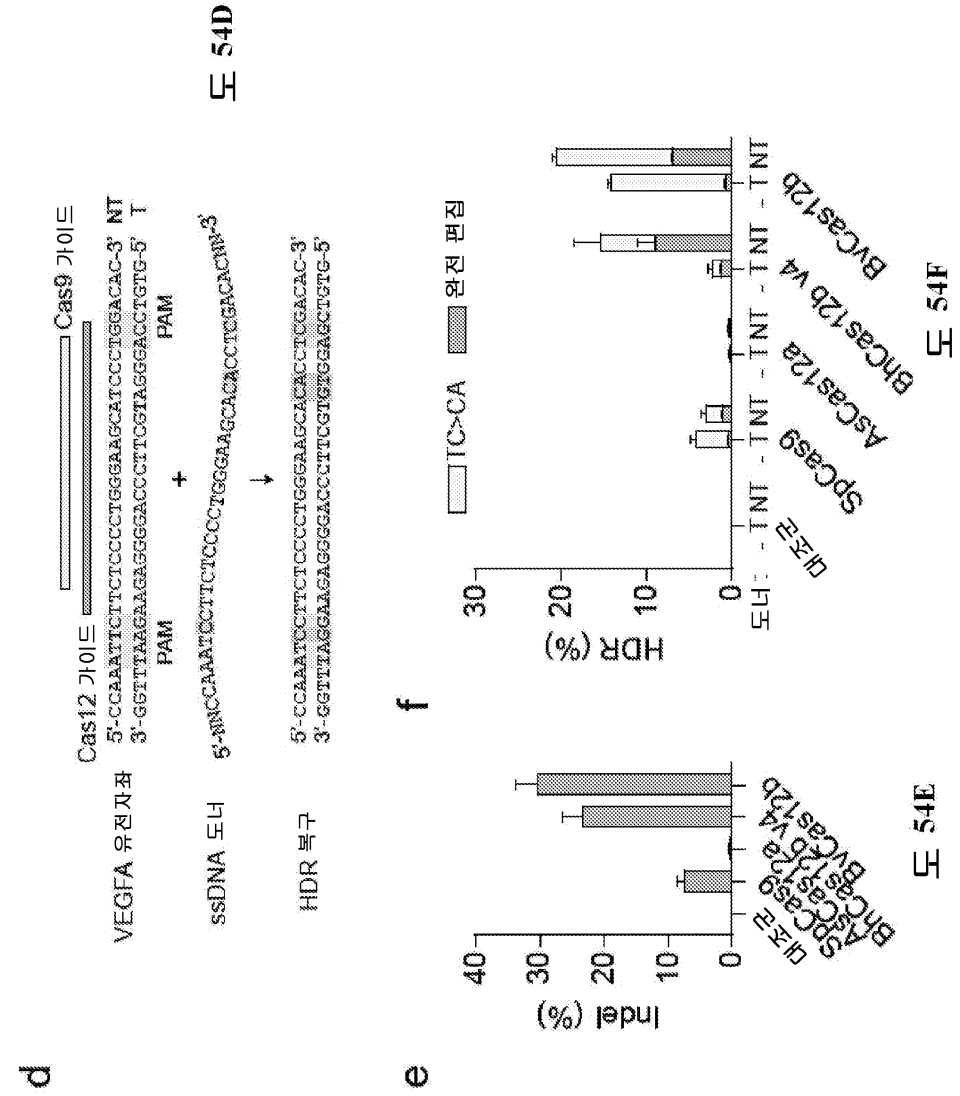


도 53D

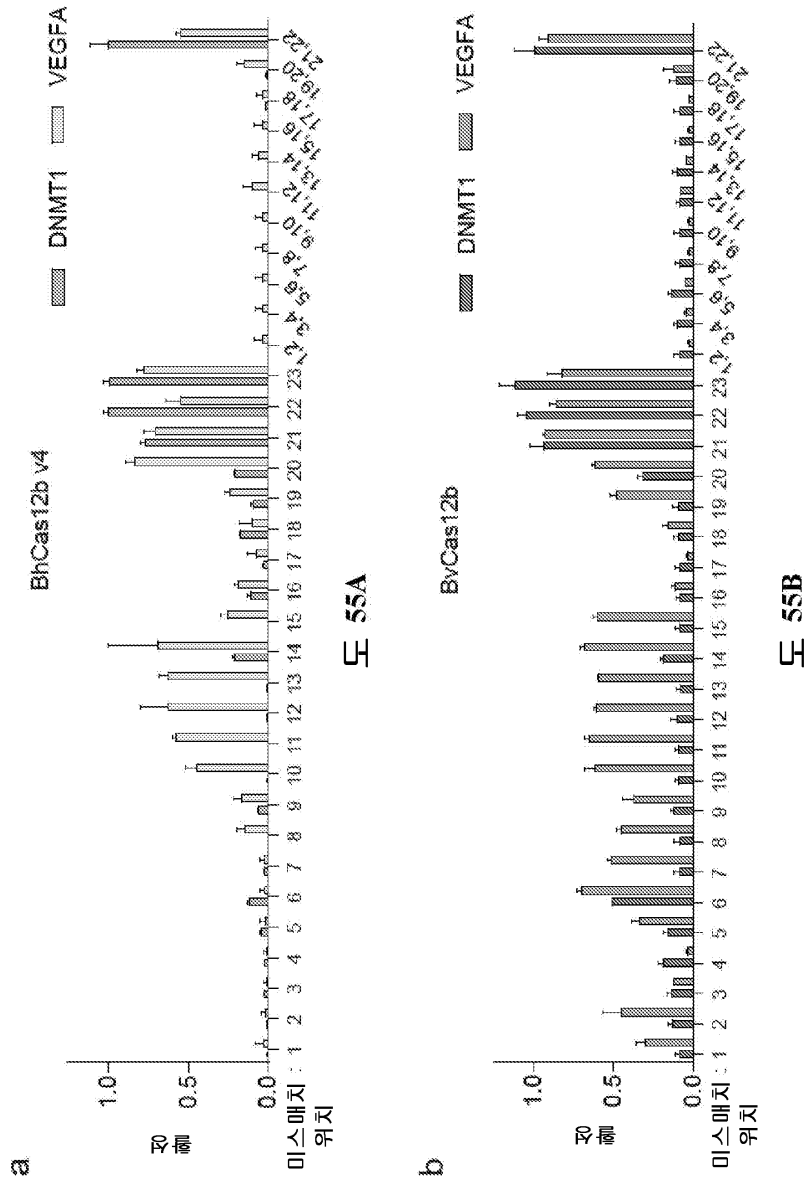
도면54a


a
Indel(%)
PAM:
BhCas12b v4
BvCas12b
도 54A
b
BvCas12b Indel(%)
BhCas12b v4 indel(%)
R²=0.48
도 54B
c
가장 가까운 절단 부위까지 거리 (bp)
Cas12a (TTTV)
Cas12b (ATTN)
Cas9 (NGG)
확률
Avg bp/ 절단 부위 :
32
35
7
도 54C

도면54b


d
Cas9 가이드
Cas12 가이드
VEGFA 유전자좌
5'-CCAAATTCTTCTCCCCTGGGAAGCATCCCTGGACAC-3' NT
3'-GGTTTAAGAAGAGGGGACCCTTCGTAGGGACCTGTG-5' T
PAM
PAM
+
ssDNA 도너
5'-nnCCAAATTCTTCTCCCCTGGGAAGCACACCTCGACACnnn-3'
→
HDR 복구
5'-CCAAATTCTTCTCCCCTGGGAAGCACACCTCGACAC-3'
3'-GGTTTAAGAAGAGGGGACCCTTCGTGTGGAGCTGTG-5'
도 54D
e
Indel (%)
0
10
20
30
40
도너:
-
T
NT
-
T
NT
-
T
NT
-
T
NT
없음
SpCas9
AsCas12a
BhCas12b v4
BvCas12b
도 54E
f
TC>CA
완전 편집
HDR (%)
0
10
20
30
도너:
-
T
NT
-
T
NT
-
T
NT
-
T
NT
없음
SpCas9
AsCas12a
BhCas12b v4
BvCas12b
도 54F

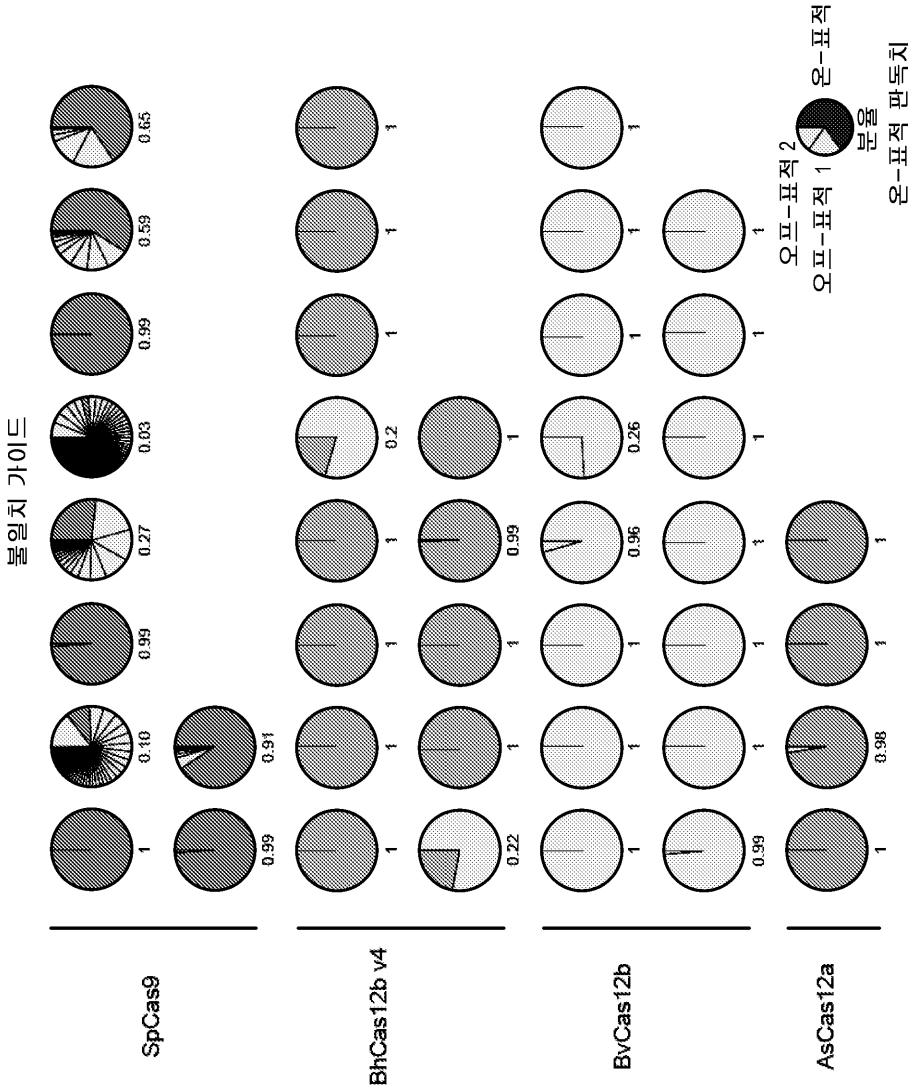
도면55a



도 55A

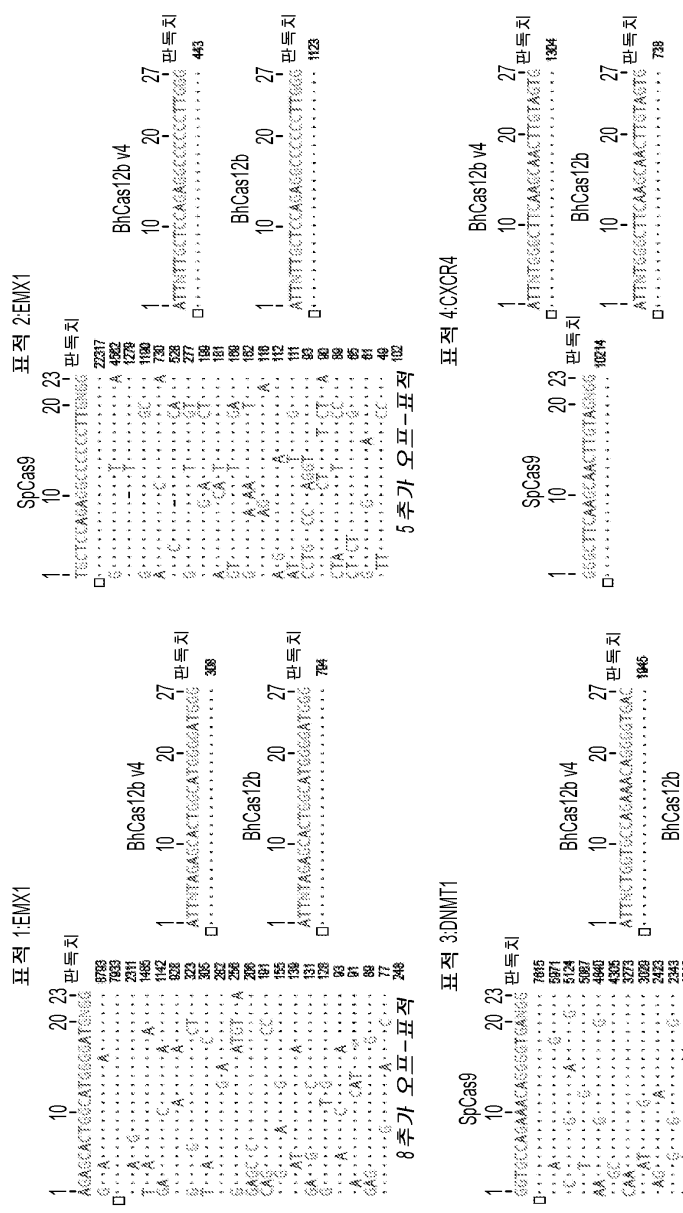
도 55B

도면55b

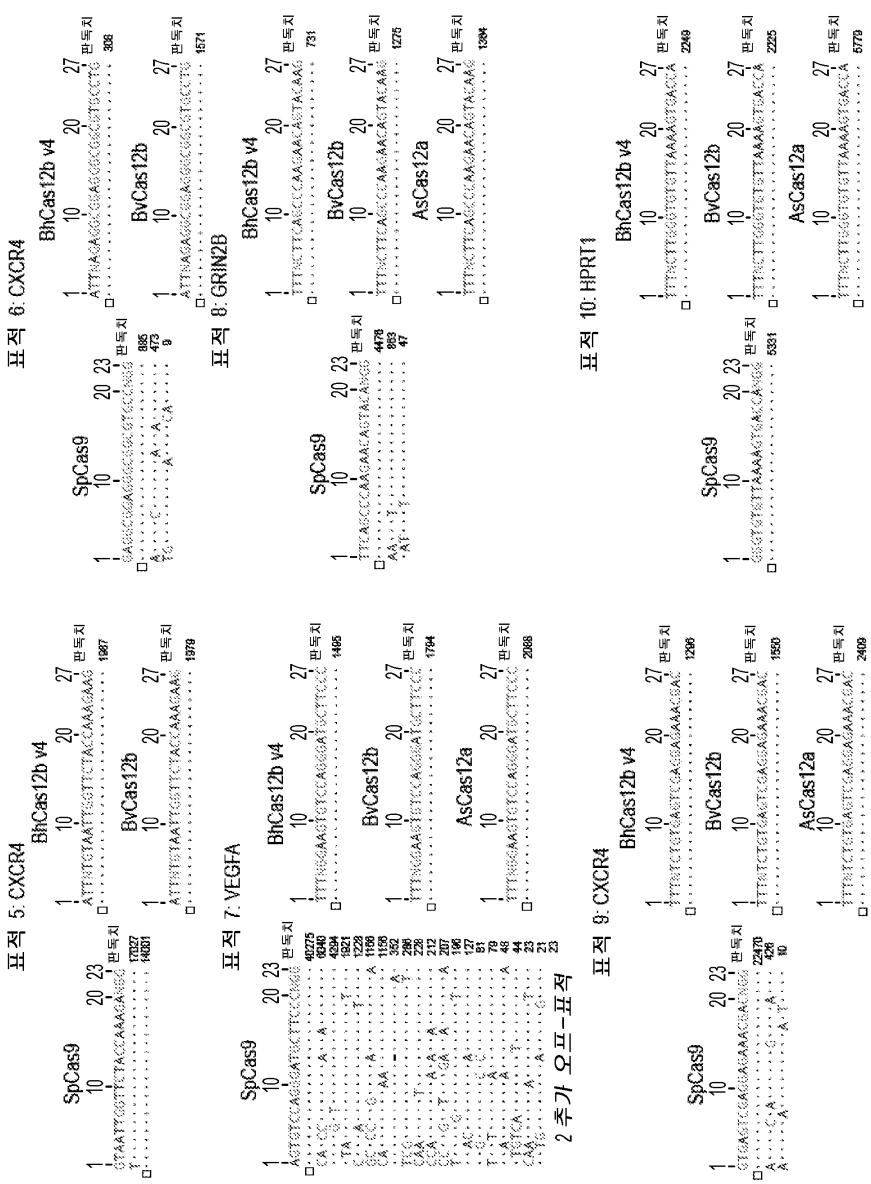

불일치 가이드
SpCas9
BhCas12b v4
BvCas12b
AsCas12a
오프-표적 2
오프-표적 1
온-표적
분율
온-표적 판독치

도 55C

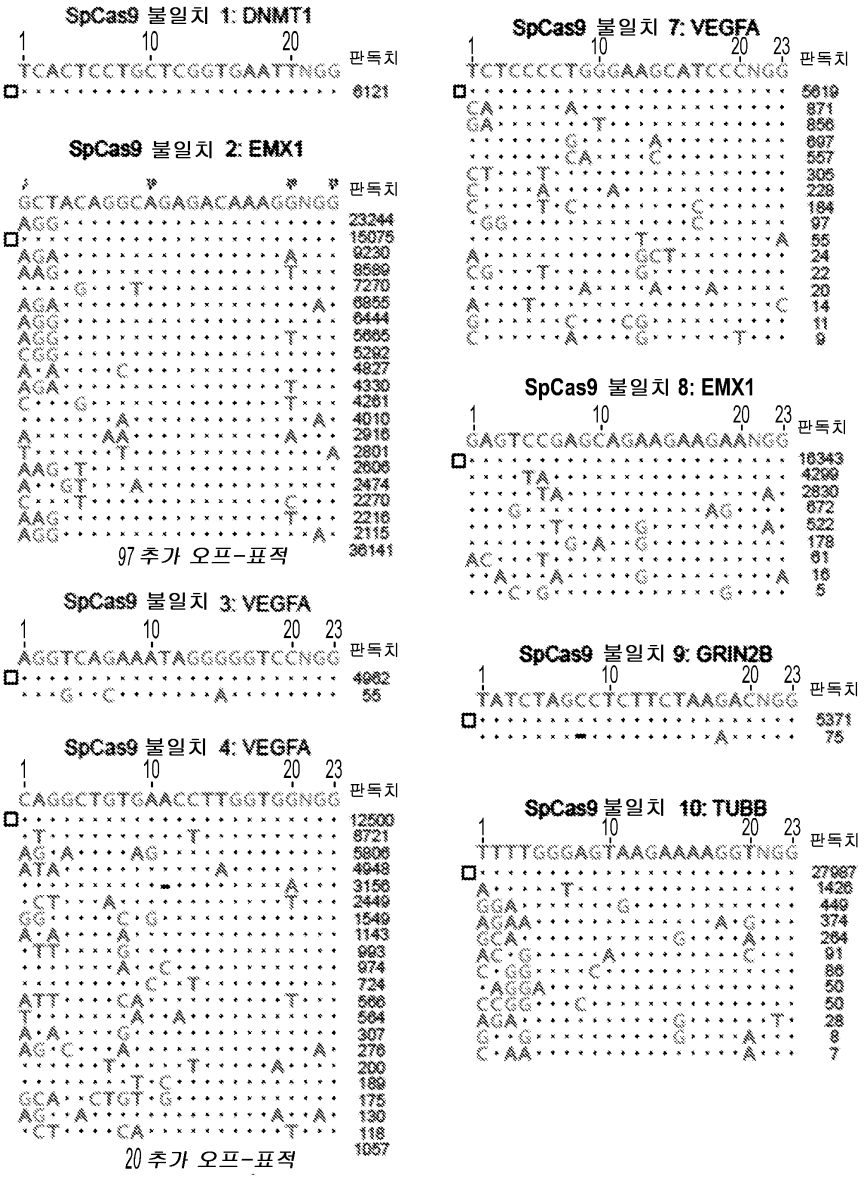
도면56a


표적 1:EMX1
SpCas9
BhCas12b v4
BhCas12b
만득치
8 추가 오프-표적
표적 2:EMX1
5 추가 오프-표적
표적 3:DNMT1
81 추가 오프-표적
표적 4:CXCR4

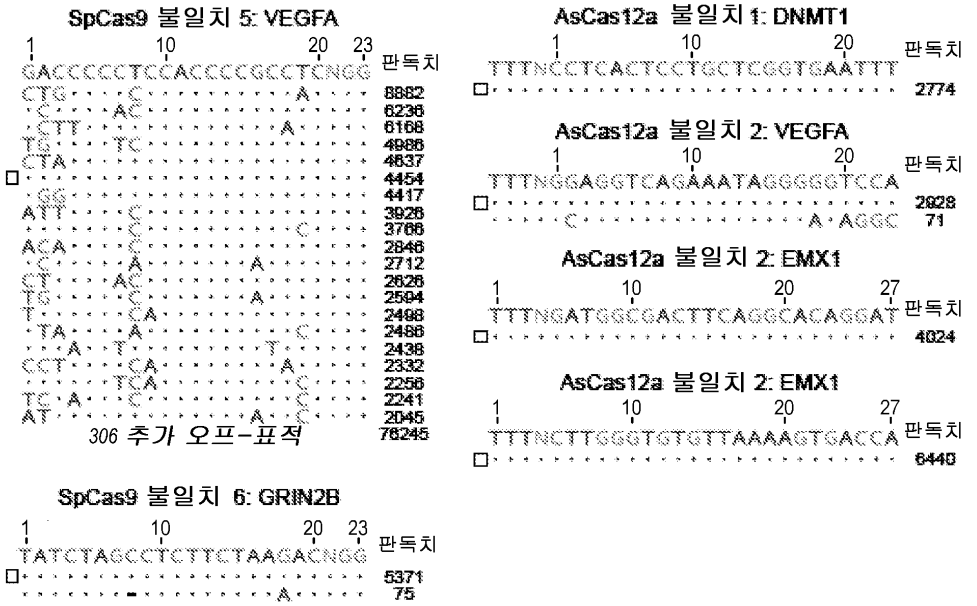
도면56b


표적 5: CXCR4
표적 6: CXCR4
표적 7: VEGFA
표적 8: GRIN2B
표적 9: CXCR4
표적 10: HPRT1
BhCas12b v4
SpCas9
BvCas12b
AsCas12a
편독치
2 추가 오프-표적

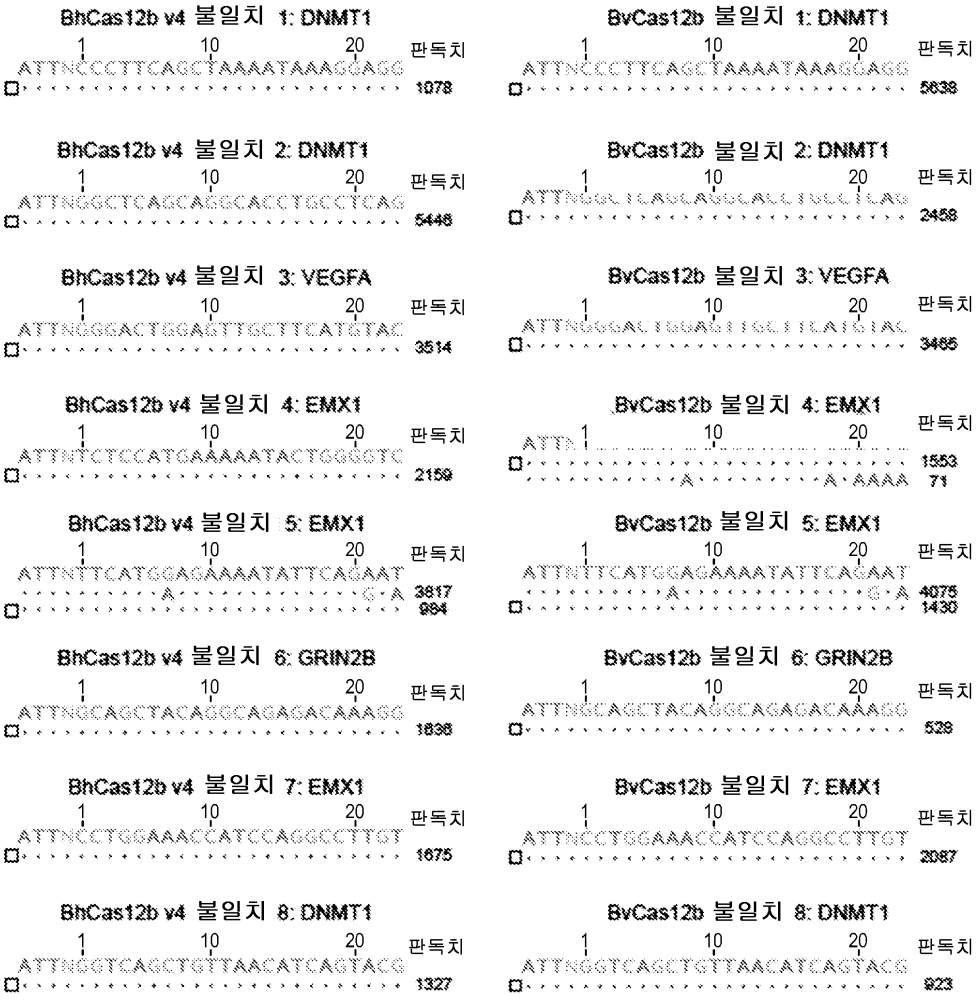
도면57a


SpCas9 불일치 1: DNMT1
판독치
SpCas9 불일치 2: EMX1
판독치
97 추가 오프-표적
SpCas9 불일치 3: VEGFA
판독치
SpCas9 불일치 4: VEGFA
판독치
20 추가 오프-표적
SpCas9 불일치 7: VEGFA
판독치
SpCas9 불일치 8: EMX1
판독치
SpCas9 불일치 9: GRIN2B
판독치
SpCas9 불일치 10: TUBB
판독치

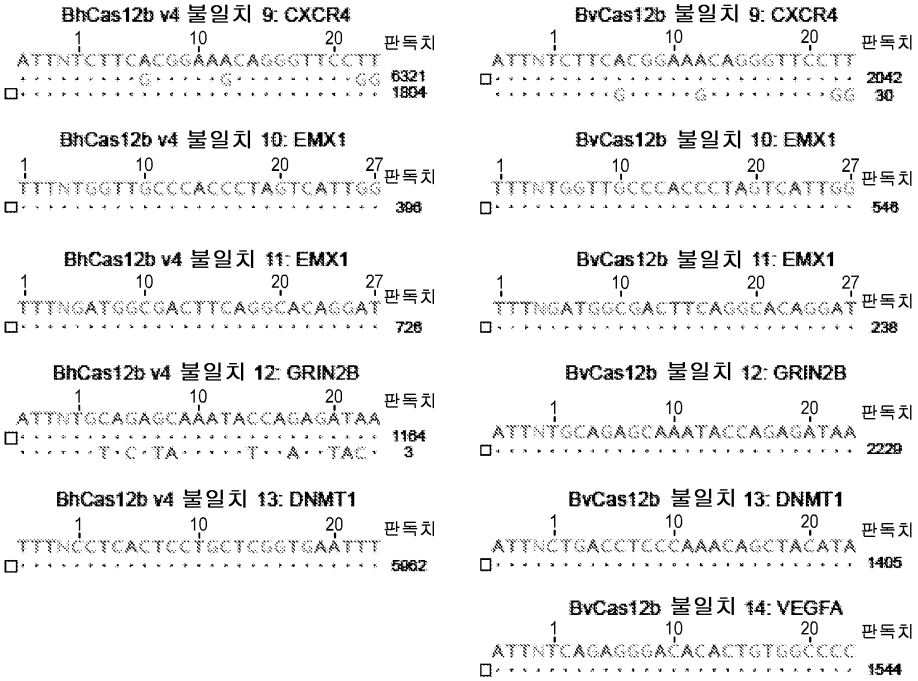
도면57b


SpCas9 불일치 5: VEGFA
판독치
306 추가 오프-표적
AsCas12a 불일치 1: DNMT1
AsCas12a 불일치 2: VEGFA
AsCas12a 불일치 2: EMX1
AsCas12a 불일치 2: EMX1
SpCas9 불일치 6: GRIN2B

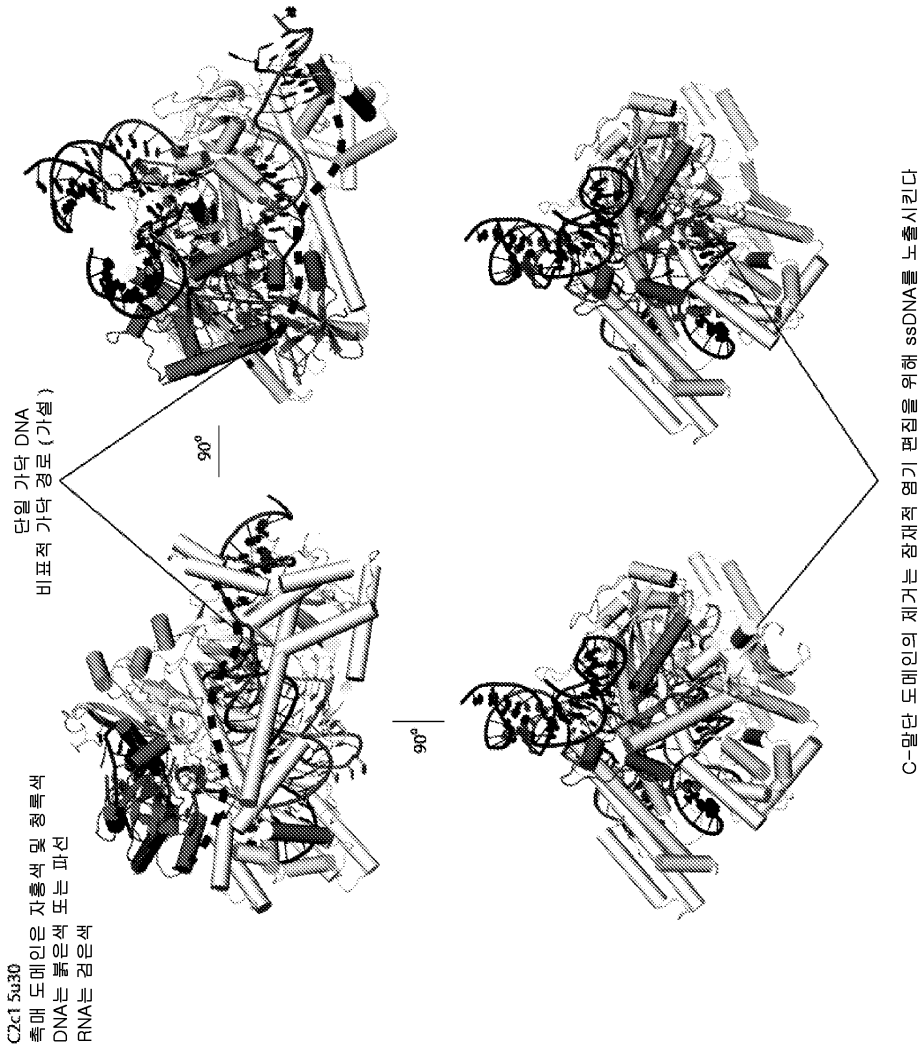
도면57c


BhCas12b v4 불일치 1: DNMT1
BvCas12b 불일치 1: DNMT1
BhCas12b v4 불일치 2: DNMT1
BvCas12b 불일치 2: DNMT1
BhCas12b v4 불일치 3: VEGFA
BvCas12b 불일치 3: VEGFA
BhCas12b v4 불일치 4: EMX1
BvCas12b 불일치 4: EMX1
BhCas12b v4 불일치 5: EMX1
BvCas12b 불일치 5: EMX1
BhCas12b v4 불일치 6: GRIN2B
BvCas12b 불일치 6: GRIN2B
BhCas12b v4 불일치 7: EMX1
BvCas12b 불일치 7: EMX1
BhCas12b v4 불일치 8: DNMT1
BvCas12b 불일치 8: DNMT1
판독치

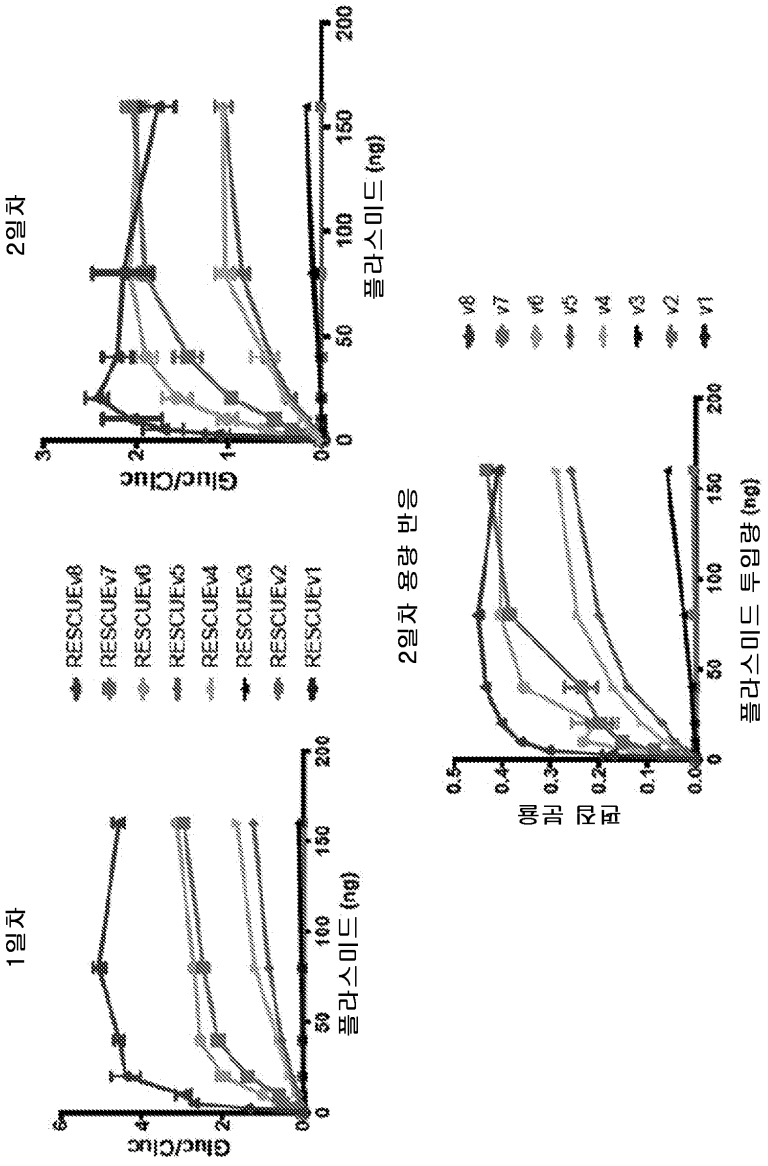
도면57d


BhCas12b v4 불일치 9: CXCR4
판독치
6321
1804
BvCas12b 불일치 9: CXCR4
판독치
2042
30
BhCas12b v4 불일치 10: EMX1
판독치
396
BvCas12b 불일치 10: EMX1
판독치
546
BhCas12b v4 불일치 11: EMX1
판독치
726
BvCas12b 불일치 11: EMX1
판독치
238
BhCas12b v4 불일치 12: GRIN2B
판독치
1184
3
BvCas12b 불일치 12: GRIN2B
판독치
2229
BhCas12b v4 불일치 13: DNMT1
판독치
5982
BvCas12b 불일치 13: DNMT1
판독치
1405
BvCas12b 불일치 14: VEGFA
판독치
1544

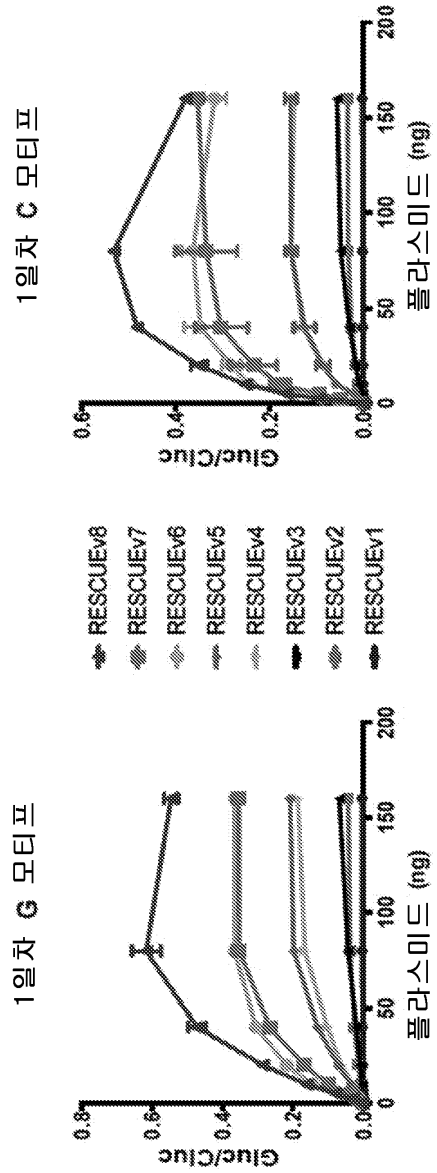
도면58


C2c1 5u30
촉매 도메인은 자홍색 및 청록색
DNA는 붉은색 또는 파신
RNA는 검은색
단일 가닥 DNA
비표적 가닥 경로 (가설)
90°
90°
C-말단 도메인의 제거는 잠재적 엽기 편집을 위해 ssDNA를 노출시킨다

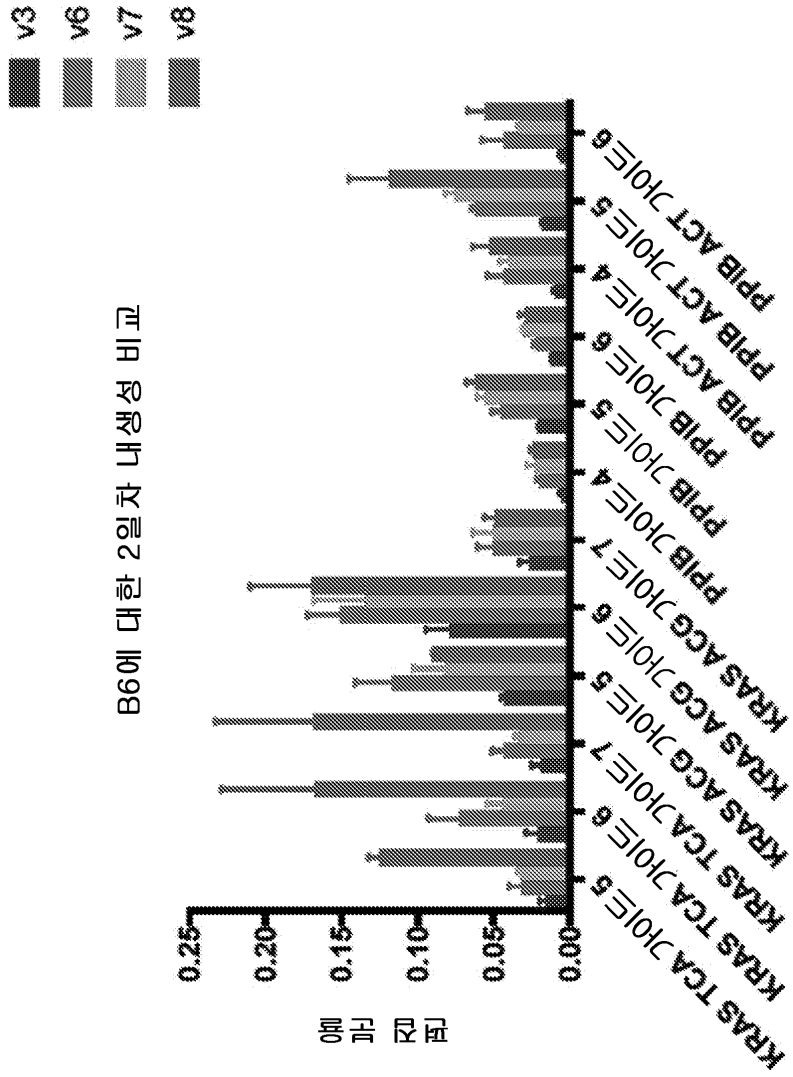
도면59


1일차
2일차
Gluc/Cluc
플라스미드 (ng)
RESCUEv8
RESCUEv7
RESCUEv6
RESCUEv5
RESCUEv4
RESCUEv3
RESCUEv2
RESCUEv1
2일차 용량 반응
편집 분율
플라스미드 투입량 (ng)
v8
v7
v6
v5
v4
v3
v2
v1

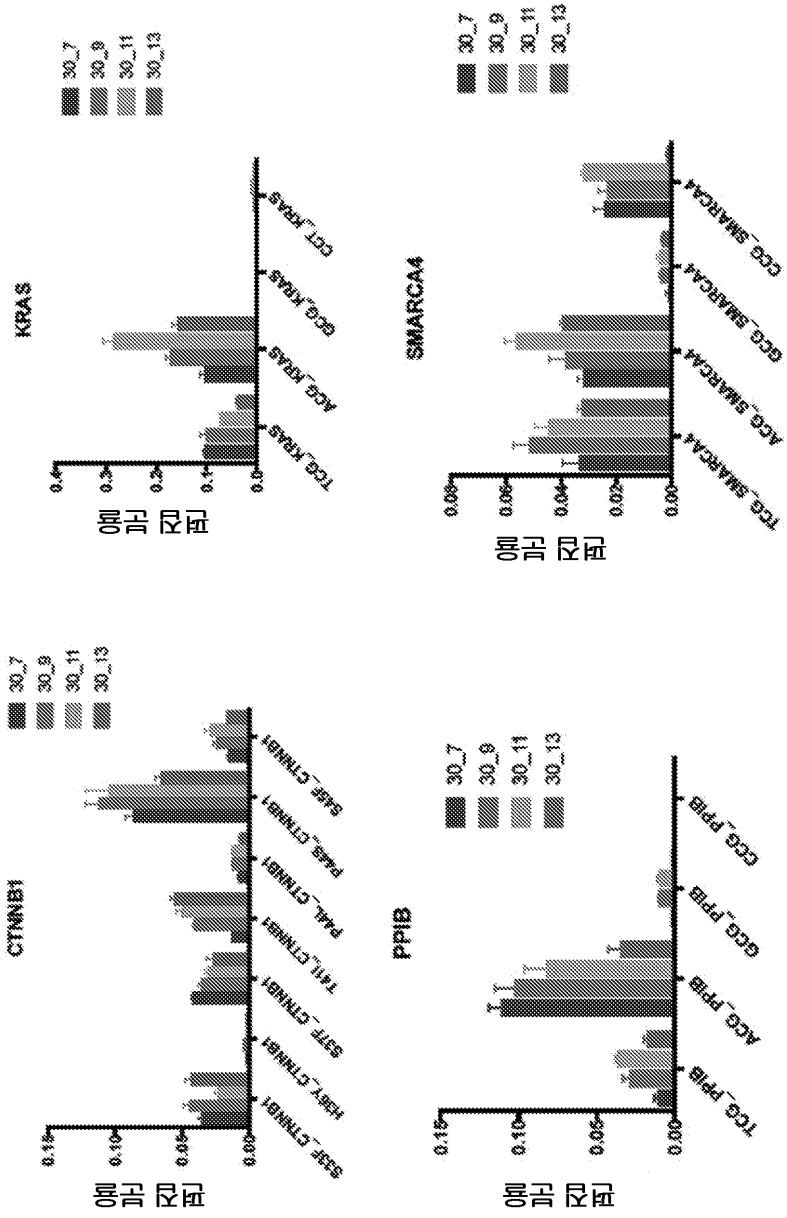
도면60


1일차 C 모티프
Gluc/Cluc
0.0
0.2
0.4
0.6
0
50
100
150
200
플라스미드 (ng)
RESCUEv8
RESCUEv7
RESCUEv6
RESCUEv5
RESCUEv4
RESCUEv3
RESCUEv2
RESCUEv1
1일차 G 모티프
Gluc/Cluc
0.0
0.2
0.4
0.6
0.8
0
50
100
150
200
플라스미드 (ng)

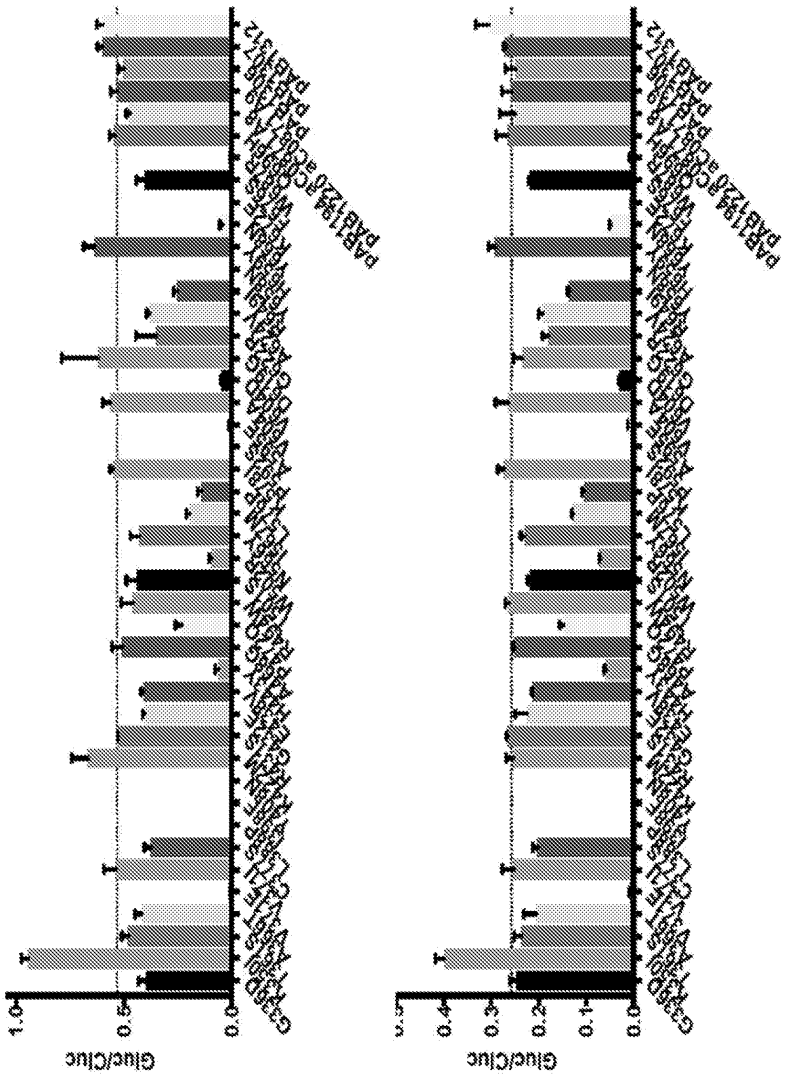
도면61


v3
v6
v7
v8
B6에 대한 2일차 내생성 비교
편집 빈율
0.25
0.20
0.15
0.10
0.05
0.00
KRAS TCA 가이드 5
KRAS TCA 가이드 6
KRAS TCA 가이드 7
KRAS ACG 가이드 5
KRAS ACG 가이드 6
KRAS ACG 가이드 7
PPIB 가이드 4
PPIB 가이드 5
PPIB 가이드 6
PPIB ACT 가이드 4
PPIB ACT 가이드 5
PPIB ACT 가이드 6

도면62


KRAS
30_7
30_9
30_11
30_13
TCG_KRAS
ACG_KRAS
GCG_KRAS
CCT_KRAS
편집 분율
SMARCA4
TCG_SMARCA4
ACG_SMARCA4
GCG_SMARCA4
CCG_SMARCA4
편집 분율
CTNNB1
GGF_CTNNB1
HGF_CTNNB1
GTT_CTNNB1
TAT_CTNNB1
PAA_CTNNB1
PAS_CTNNB1
GAF_CTNNB1
편집 분율
PPIB
TCG_PPIB
ACG_PPIB
GCG_PPIB
CCG_PPIB
편집 분율

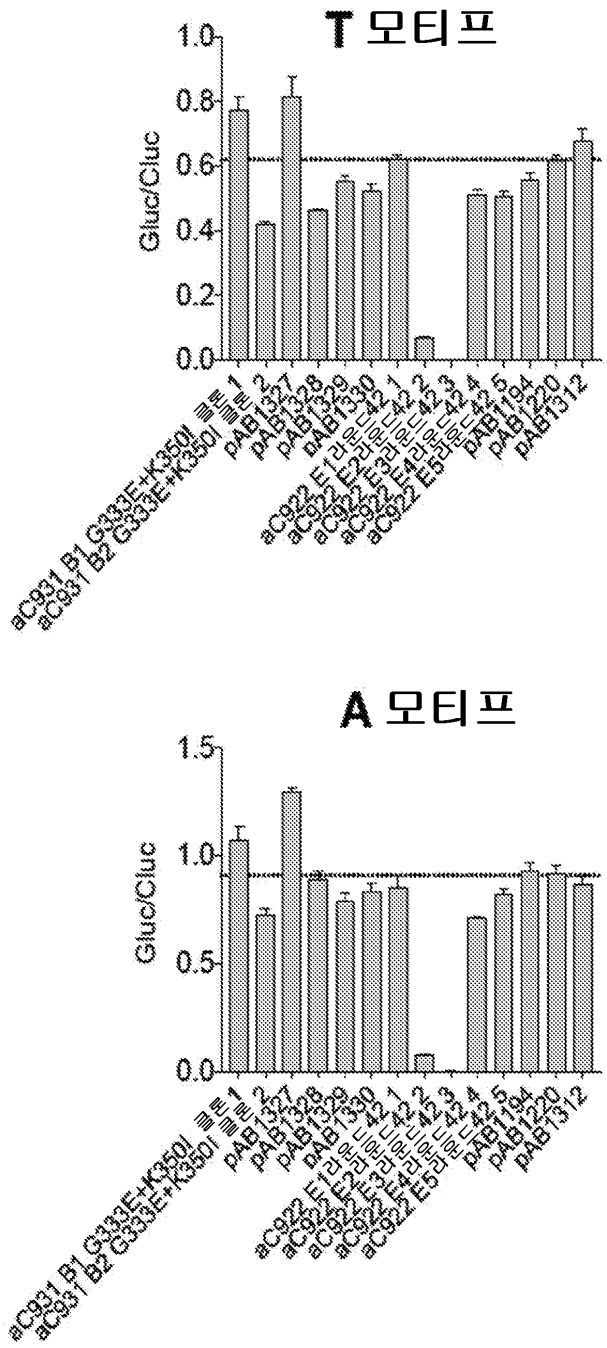
도면63


Gluc/Cluc
1.0
0.5
0.0
pAB194-C
pAB120-C
Gluc/Cluc
0.5
0.4
0.3
0.2
0.1
0.0
pAB194-C
pAB120-C

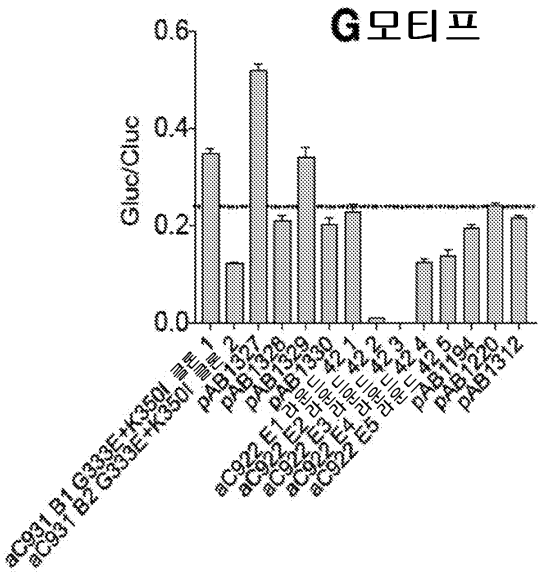
1일차

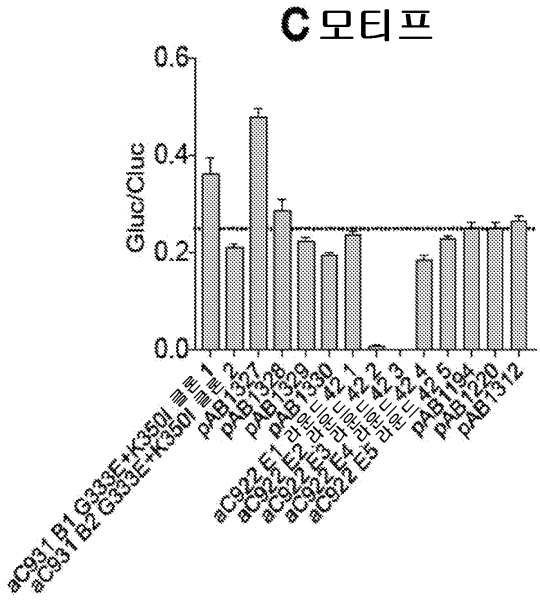
2일차

도면64a

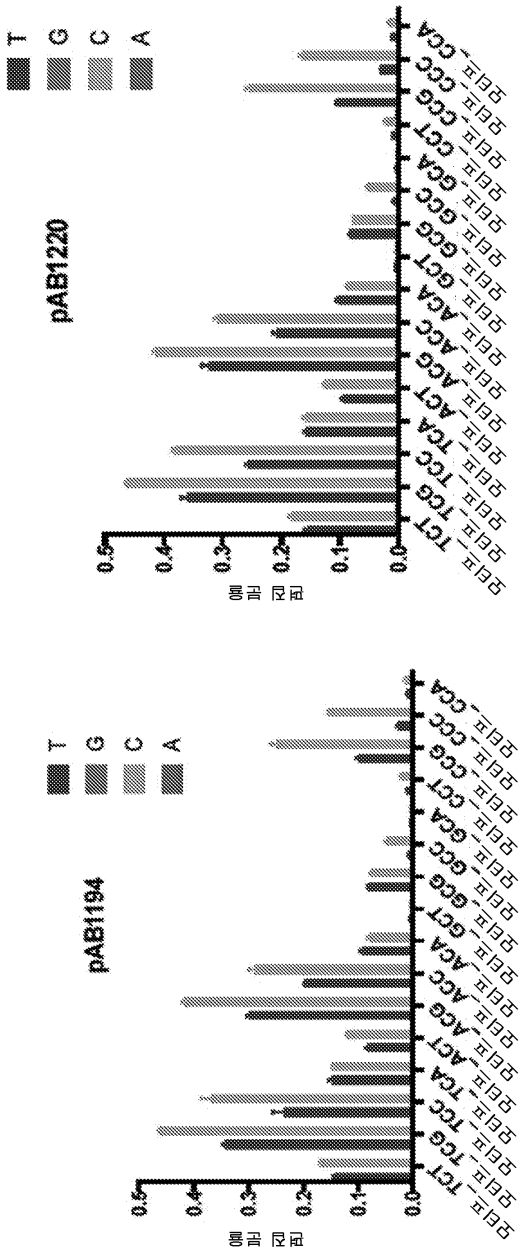

T 모티프
Gluc/Cluc
1.0
0.8
0.6
0.4
0.2
0.0
클론 1
클론 2
pAB1327
pAB1328
pAB1329
pAB1330
E1 리니어드-42 1
E2 리니어드-42 2
E3 리니어드-42 3
E4 리니어드-42 4
E5 리니어드-42 5
pAB1194
pAB1220
pAB1312
αC931 B1 G333E+K350I
αC931 B2 G333E+K350I
αC922 E1 리니어드-42 1
αC922 E2 리니어드-42 2
αC922 E3 리니어드-42 3
αC922 E4 리니어드-42 4
αC922 E5 리니어드-42 5
A 모티프
Gluc/Cluc
1.5
1.0
0.5
0.0
클론 1
클론 2
pAB1327
pAB1328
pAB1329
pAB1330
E1 리니어드-42 1
E2 리니어드-42 2
E3 리니어드-42 3
E4 리니어드-42 4
E5 리니어드-42 5
pAB1194
pAB1220
pAB1312
αC931 B1 G333E+K350I
αC931 B2 G333E+K350I
αC922 E1 리니어드-42 1
αC922 E2 리니어드-42 2
αC922 E3 리니어드-42 3
αC922 E4 리니어드-42 4
αC922 E5 리니어드-42 5

도면64b

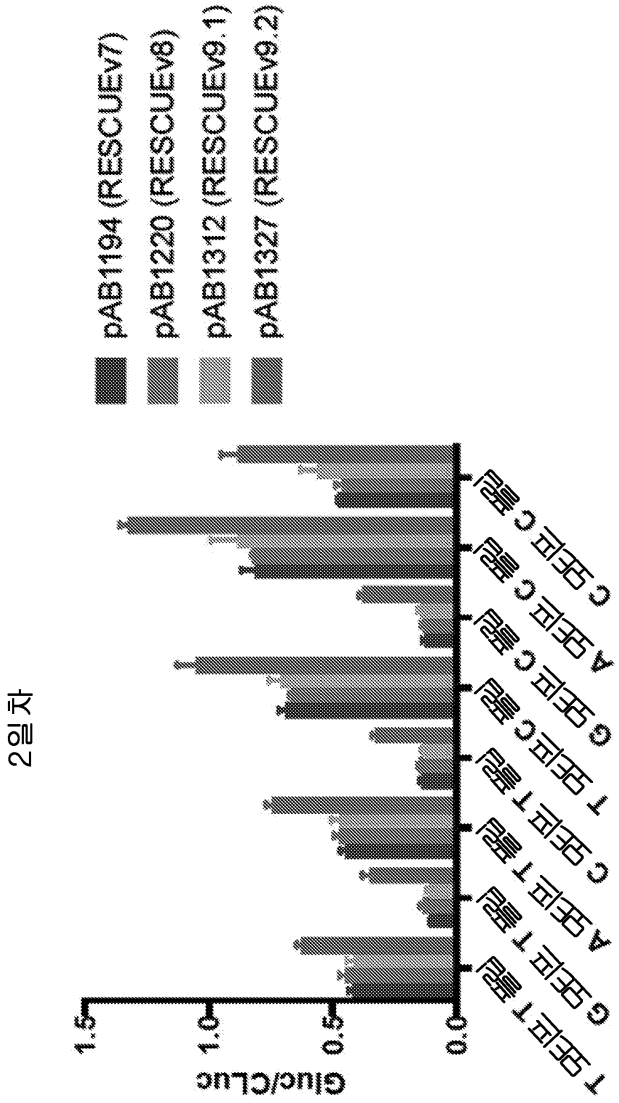

G모티프
Gluc/Cluc
0.6
0.4
0.2
0.0
aC931 B1 G333E+K350I
aC931 B2 G333E+K350I
클론 1
클론 2
pAB1327
pAB1328
pAB1329
pAB1330
라이브러리 42 1
라이브러리 42 2
라이브러리 42 3
라이브러리 42 4
라이브러리 42 5
pAB1194
pAB1220
pAB1312
aC922 E1
aC922 E2
aC922 E3
aC922 E4
aC922 E5


C모티프
Gluc/Cluc
0.6
0.4
0.2
0.0
aC931 B1 G333E+K350I
aC931 B2 G333E+K350I
클론 1
클론 2
pAB1327
pAB1328
pAB1329
pAB1330
라이브러리 42 1
라이브러리 42 2
라이브러리 42 3
라이브러리 42 4
라이브러리 42 5
pAB1194
pAB1220
pAB1312
aC922 E1
aC922 E2
aC922 E3
aC922 E4
aC922 E5

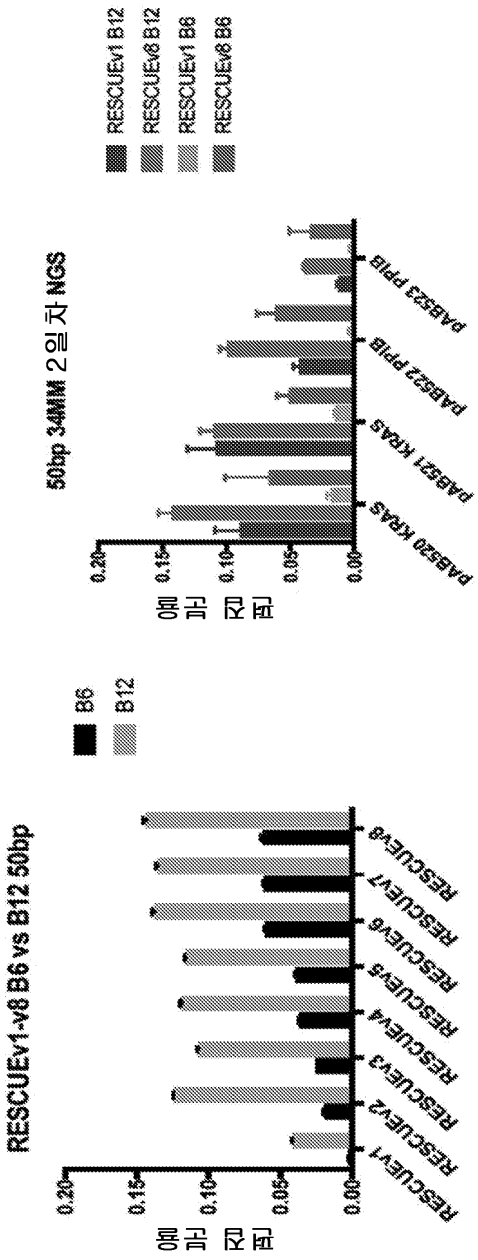
도면65


T
G
C
A
pAB1220
0.5
0.4
0.3
0.2
0.1
0.0
TCT
TCG
TCC
TCA
ACT
ACG
ACC
ACA
GCT
GCG
GCC
GCA
CCT
CCG
CCC
CCA
pAB1194

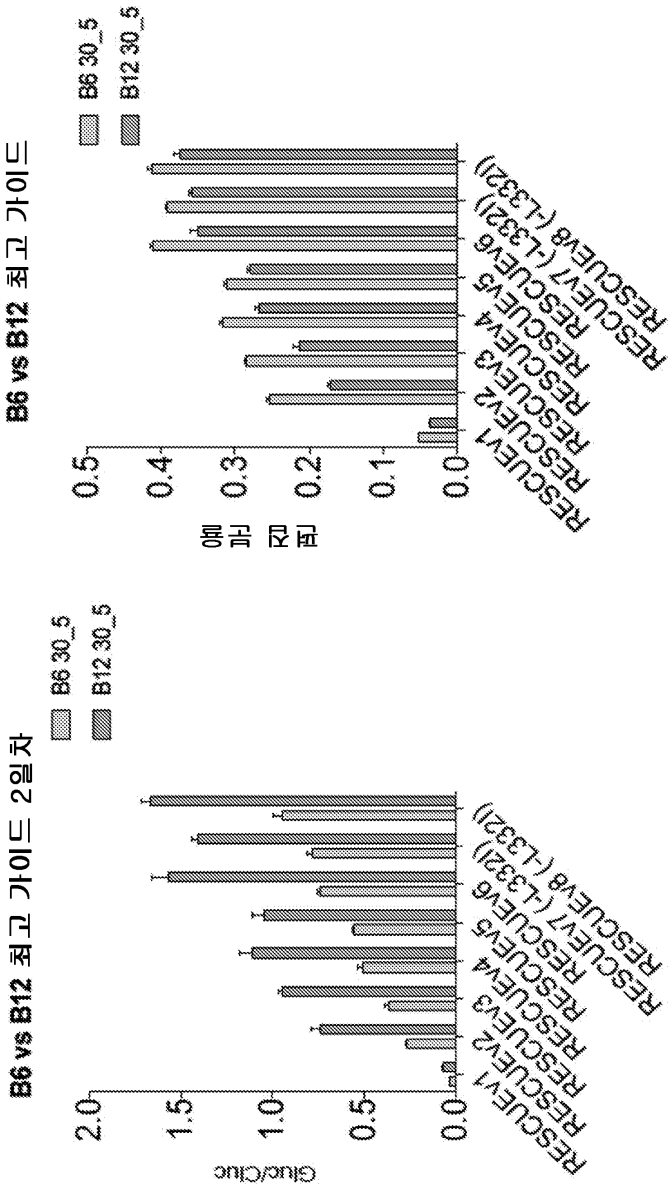
도면66


pAB1194 (RESCUEv7)
pAB1220 (RESCUEv8)
pAB1312 (RESCUEv9.1)
pAB1327 (RESCUEv9.2)
2일 차
Gluc/cLuc
1.5
1.0
0.5
0.0
T 오프타겟 gRNA
G 오프타겟 gRNA
A 오프타겟 gRNA
C 오프타겟 gRNA
T 오프타겟 gRNA
G 오프타겟 gRNA
A 오프타겟 gRNA
C 오프타겟 gRNA

도면67


RESCUEv1 B12
RESCUEv8 B12
RESCUEv1 B6
RESCUEv8 B6
50bp 34MM 2일 차 NGS
편집 빈율
0.20
0.15
0.10
0.05
0.00
pAS520 KRAS
pAS521 KRAS
pAS522 PPIB
pAS523 PPIB
B6
B12
RESCUEv1-v8 B6 vs B12 50bp
편집 빈율
RESCUEv1
RESCUEv2
RESCUEv3
RESCUEv4
RESCUEv5
RESCUEv6
RESCUEv7
RESCUEv8

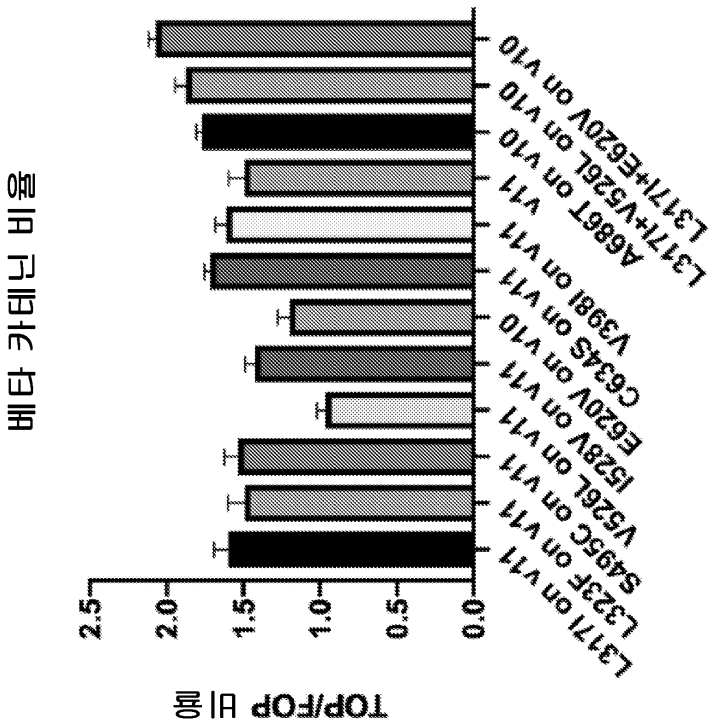
도면68


B6 vs B12 최고 가이드
B6 30_5
B12 30_5
형광 분율
0.0
0.1
0.2
0.3
0.4
0.5
RESCUEV1
RESCUEV2
RESCUEV3
RESCUEV4
RESCUEV5
RESCUEV6
RESCUEV7 (L332I)
RESCUEV8 (L332I)
B6 vs B12 최고 가이드 2일차
B6 30_5
B12 30_5
Gluc/Gluc
0.0
0.5
1.0
1.5
2.0
RESCUEV1
RESCUEV2
RESCUEV3
RESCUEV4
RESCUEV5
RESCUEV6
RESCUEV7 (L332I)
RESCUEV8 (L332I)

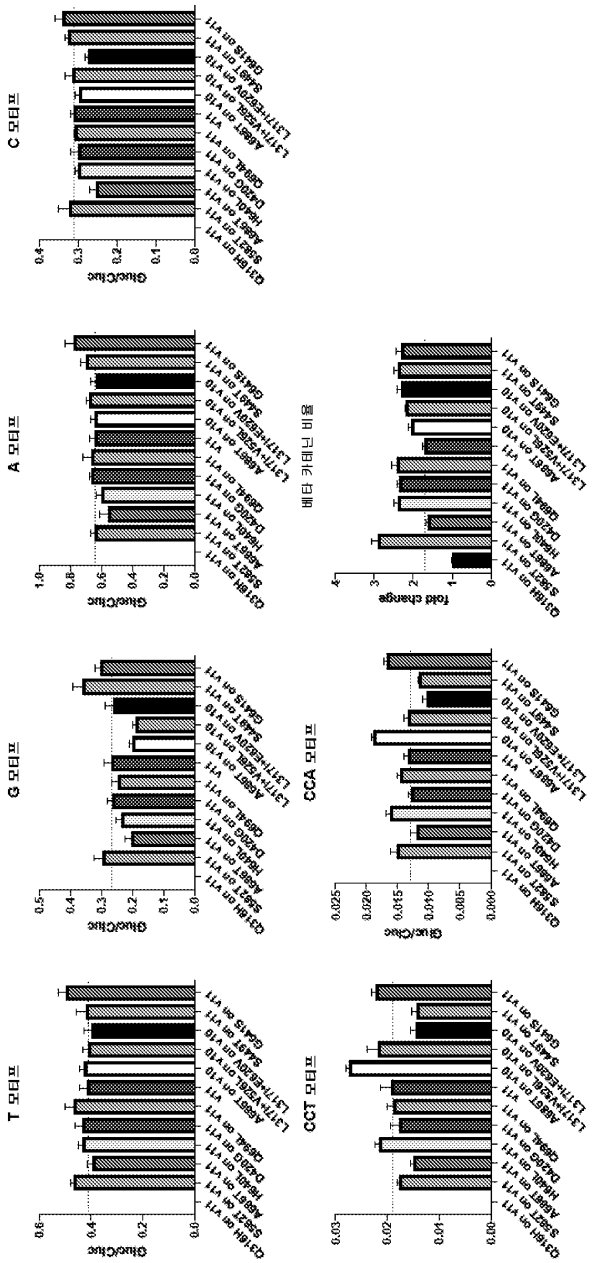
도면69

RESCUE 라운드	ADAR 돌연변이	플라스미드
RESCUEv0	E488Q	pAB0048
RESCUEv1	E488Q, V351G	pAB0359
RESCUEv2	E488Q, V351G, S486A	pAB1188
RESCUEv3	E488Q, V351G, S486A, T375S	pAB0642
RESCUEv4	E488Q, V351G, S486A, T375S, S370C,	pAB1072
RESCUEv5	E488Q, V351G, S486A, T375S, S370C, P462A	pAB1135
RESCUEv6	F488Q, V351G, S486A, T375S, S370C, P462A, N597I	pAB1146
RESCUEv7	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, I332I	pAB1184
RESCUEv8	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V	pAB1220
RESCUEv9	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I	pAB1327
RESCUEv10	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L	pAB1411
RESCUEv11	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G	pAB1481
RESCUEv12	E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T	pAB1553

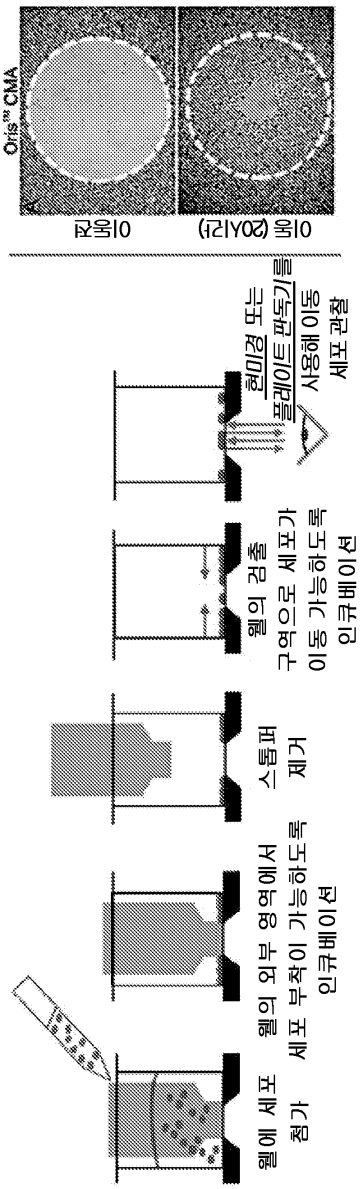
도면70


베타 카테닌 비율
TOP/FOP 비율
2.5
2.0
1.5
1.0
0.5
0.0
L371I on V11
L323F on V11
S495C on V11
V526L on V11
I528V on V11
E620V on V11
C634S on V10
V398I on V11
V11
A686T on V10
L371I+V526L on V10
L371I+E620V on V10

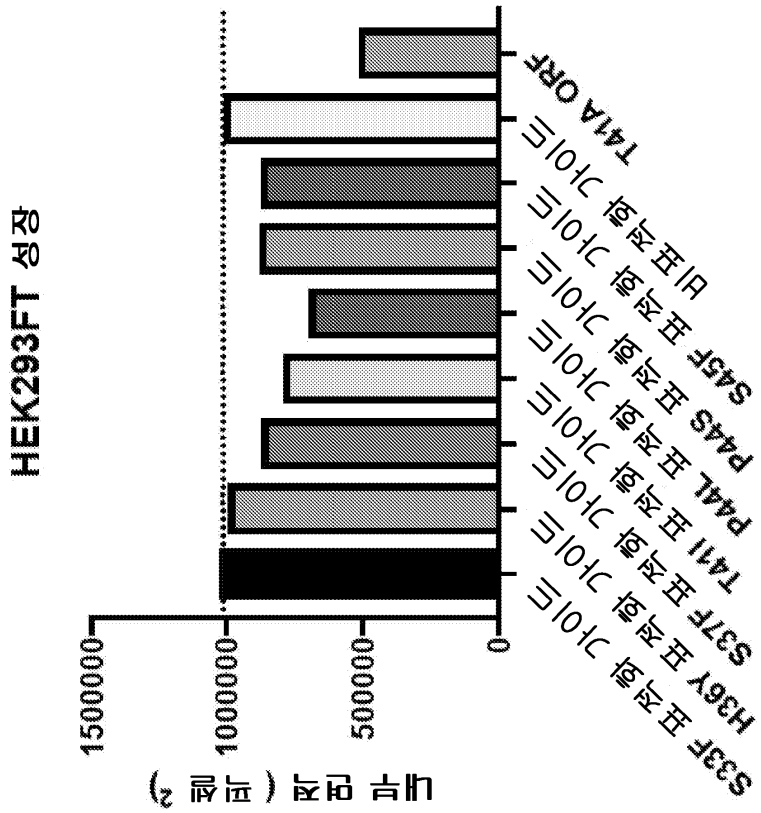
도면71



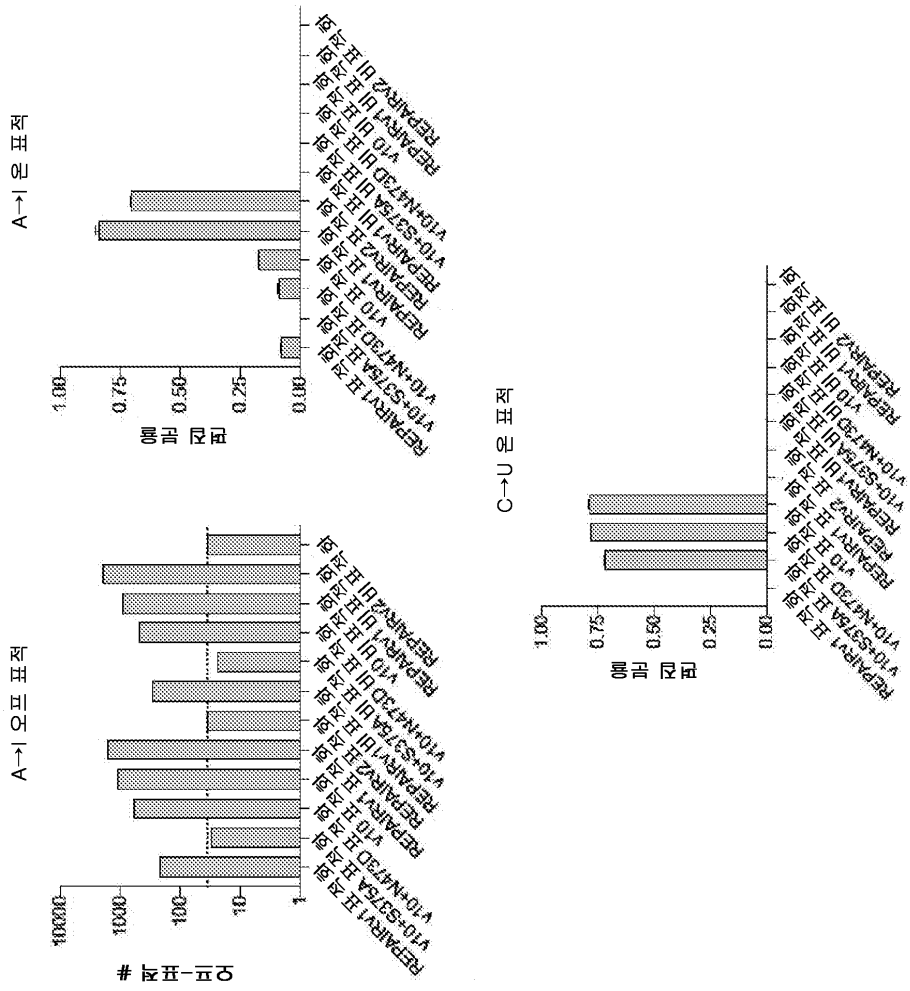
도면72


웰에 세포 첨가
웰의 외부 영역에서 세포 부착이 가능하도록 인큐베이션
스톱퍼 제거
웰의 검출 구역으로 세포가 이동 가능하도록 인큐베이션
현미경 또는 플레이트 판독기를 사용해 이동 세포 관찰
Oris™ CMA
이동 (20시간)
이동전

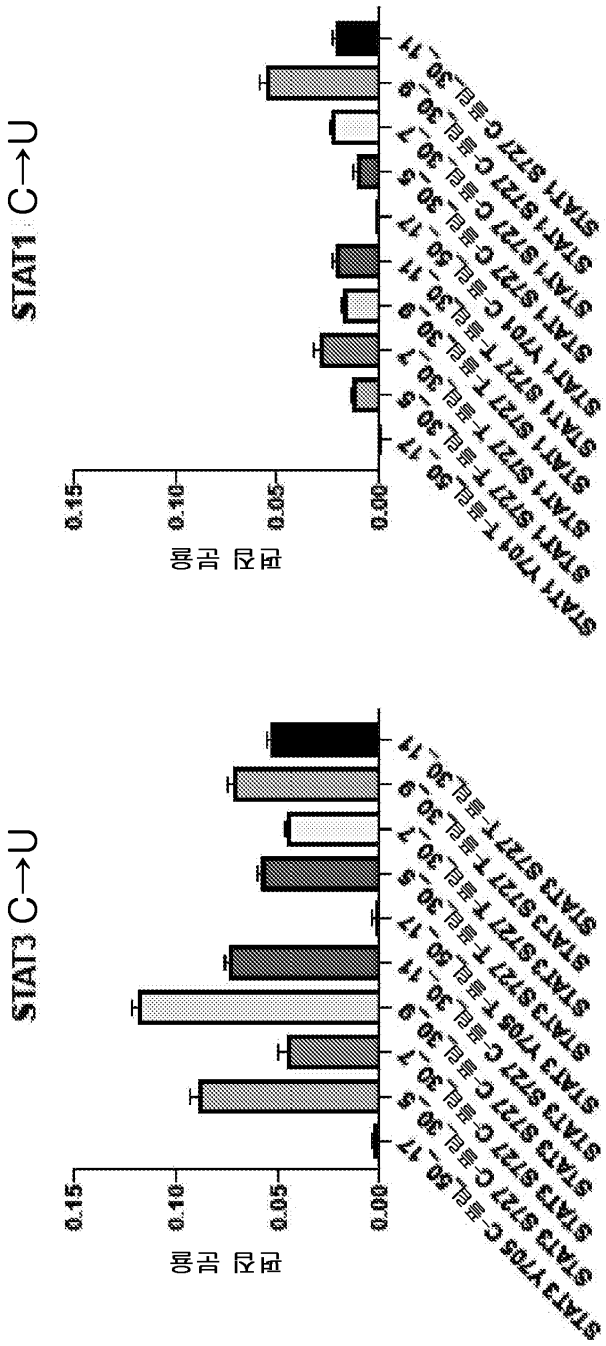
도면73


HEK293FT 성장
내부 면적 (픽셀 2)
1500000
1000000
500000
0
S33F 표적화 가이드
H36Y 표적화 가이드
S37F 표적화 가이드
T41I 표적화 가이드
P44L 표적화 가이드
P44S 표적화 가이드
S45F 표적화 가이드
비표적화 가이드
T41A ORF

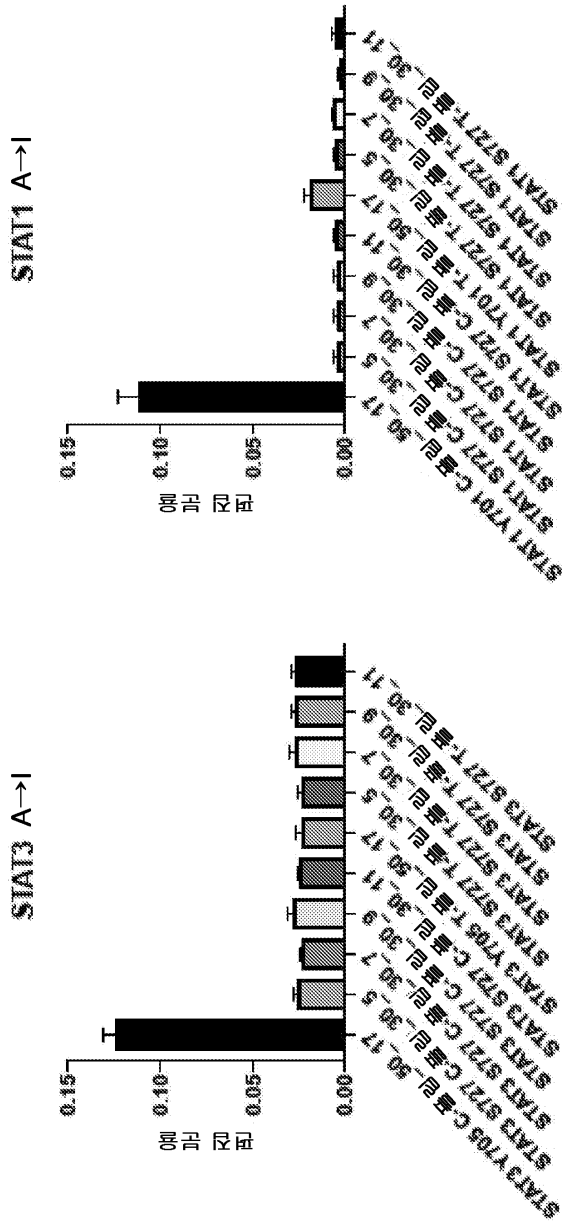
도면74


A→I 온 표적
편집 분율
1.00
0.75
0.50
0.25
0.00
A→I 오프 표적
오프-표적 #
10000
1000
100
10
1
C→U 온 표적
편집 분율
1.00
0.75
0.50
0.25
0.00

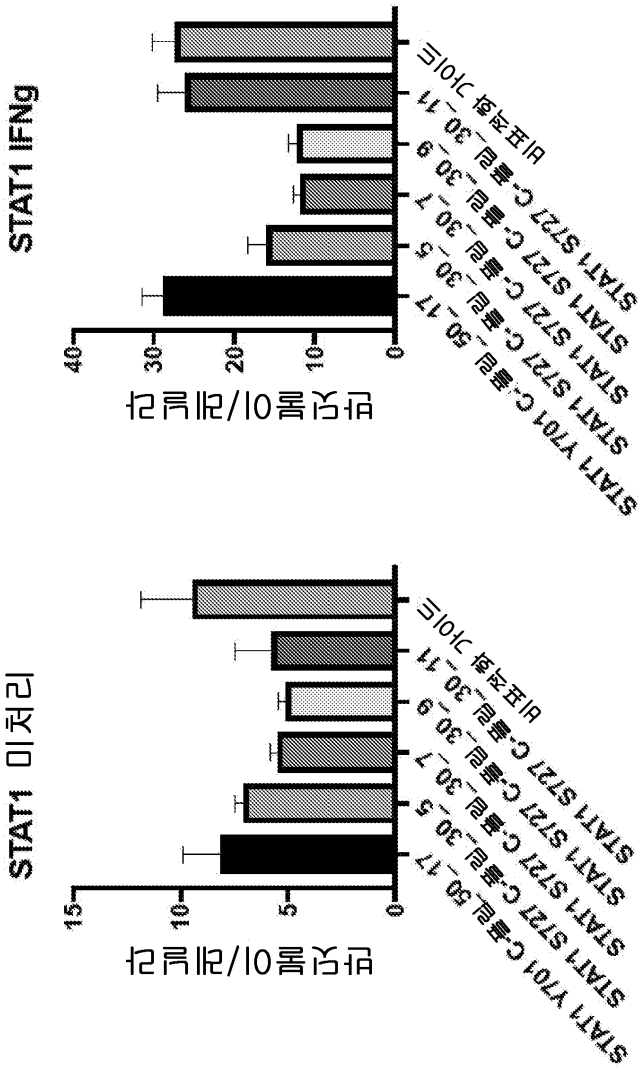
도면75a


STAT1 C→U
편집 빈율
0.15
0.10
0.05
0.00
STAT1 YT01 T-뉴클_50_17
STAT1 ST21 T-뉴클_30_5
STAT1 ST21 T-뉴클_30_7
STAT1 YT01 C-뉴클_30_9
STAT1 ST21 C-뉴클_50_11
STAT1 ST21 C-뉴클_30_5
STAT1 ST21 C-뉴클_30_7
STAT1 ST21 C-뉴클_30_9
STAT1 ST21 C-뉴클_30_11
STAT3 C→U
편집 빈율
0.15
0.10
0.05
0.00
STAT3 YT05 C-뉴클_50_17
STAT3 ST21 C-뉴클_30_5
STAT3 ST21 C-뉴클_30_7
STAT3 YT05 T-뉴클_30_9
STAT3 ST21 T-뉴클_50_11
STAT3 ST21 T-뉴클_30_5
STAT3 ST21 T-뉴클_30_7
STAT3 ST21 T-뉴클_30_9
STAT3 ST21 T-뉴클_30_11

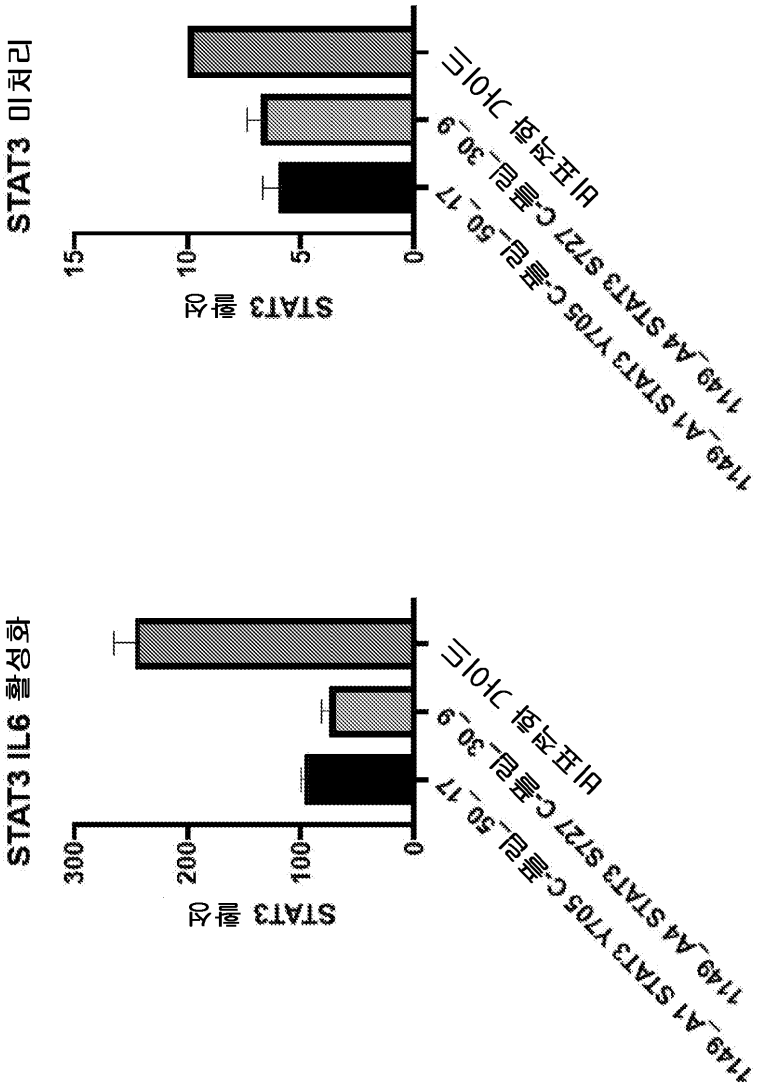
도면75b


STAT1 A→I
편집 빈율
0.15
0.10
0.05
0.00
STAT1 VT01 C:홀딥_90_17
STAT1 ST27 C:홀딥_30_5
STAT1 ST27 C:홀딥_30_7
STAT1 ST27 C:홀딥_30_9
STAT1 VT01 T:홀딥_30_17
STAT1 ST27 T:홀딥_90_17
STAT1 ST27 T:홀딥_30_5
STAT1 ST27 T:홀딥_30_7
STAT1 ST27 T:홀딥_30_9
STAT1 ST27 T:홀딥_30_11
STAT3 A→I
편집 빈율
0.15
0.10
0.05
0.00
STAT3 VT05 C:홀딥_90_17
STAT3 ST27 C:홀딥_30_5
STAT3 ST27 C:홀딥_30_7
STAT3 ST27 C:홀딥_30_9
STAT3 VT05 T:홀딥_30_11
STAT3 ST27 T:홀딥_90_17
STAT3 ST27 T:홀딥_30_5
STAT3 ST27 T:홀딥_30_7
STAT3 ST27 T:홀딥_30_9
STAT3 ST27 T:홀딥_30_11

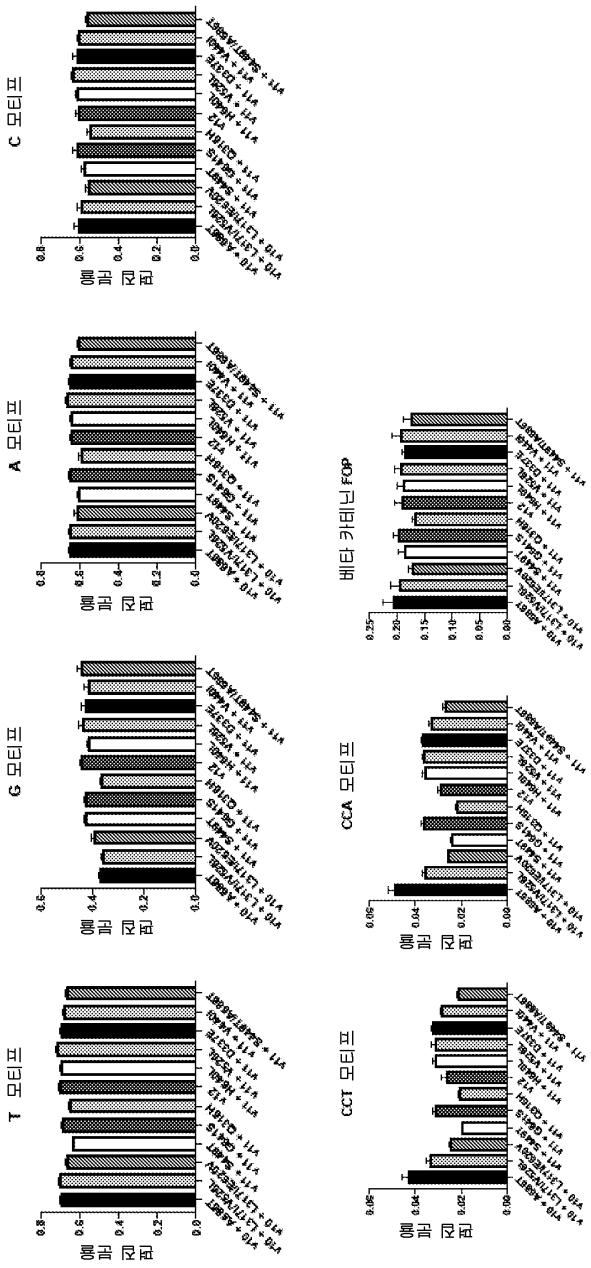
도면76


STAT1 IFNg
반딧불이/레닐라
0
10
20
30
40
STAT1 Y701 C-플립_50_17
STAT1 S727 C-플립_30_5
STAT1 S727 C-플립_30_7
STAT1 S727 C-플립_30_9
STAT1 S727 C-플립_30_11
비표적화 가이드
STAT1 미처리
반딧불이/레닐라
0
5
10
15
STAT1 Y701 C-플립_50_17
STAT1 S727 C-플립_30_5
STAT1 S727 C-플립_30_7
STAT1 S727 C-플립_30_9
STAT1 S727 C-플립_30_11
비표적화 가이드

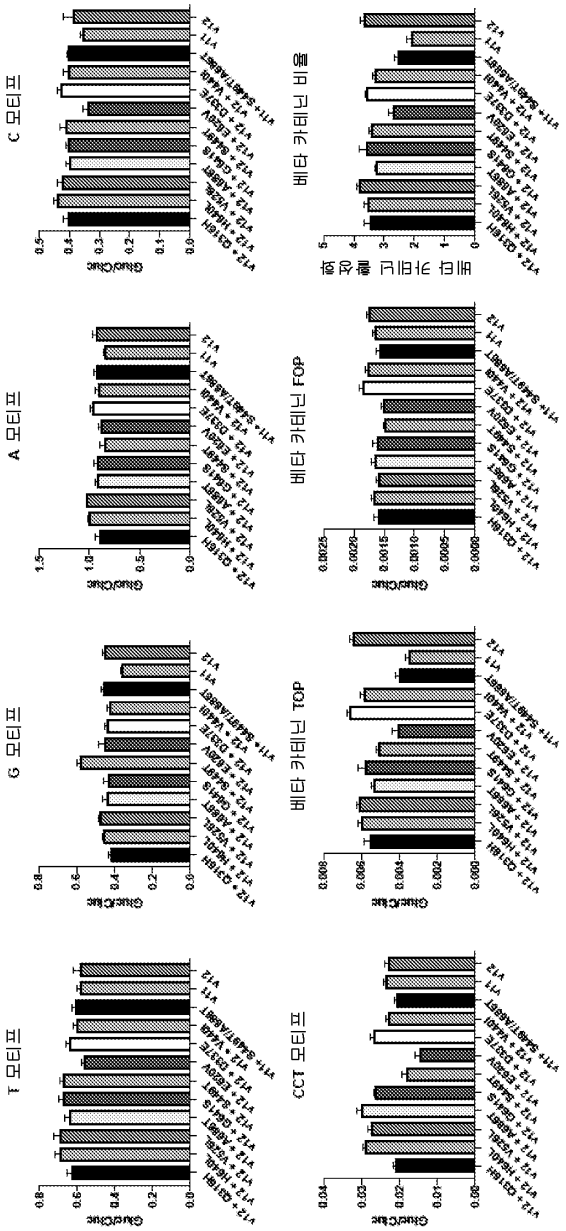
도면77


STAT3 미처리
STAT3 활성
0
5
10
15
1149_A1 STAT3 YT05 C-폴딩_50_17
1149_A4 STAT3 S727 C-폴딩_30_9
비표적화 가이드
STAT3 IL6 활성화
STAT3 활성
0
100
200
300
1149_A1 STAT3 YT05 C-폴딩_50_17
1149_A4 STAT3 S727 C-폴딩_30_9
비표적화 가이드

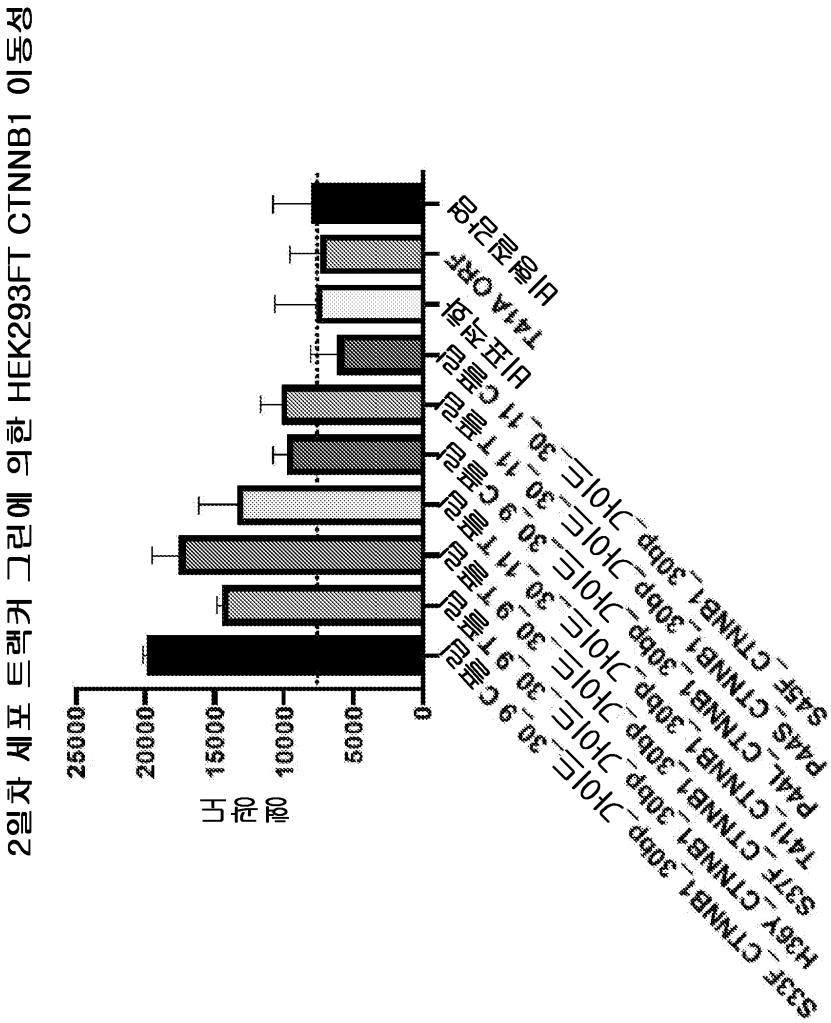
도면78



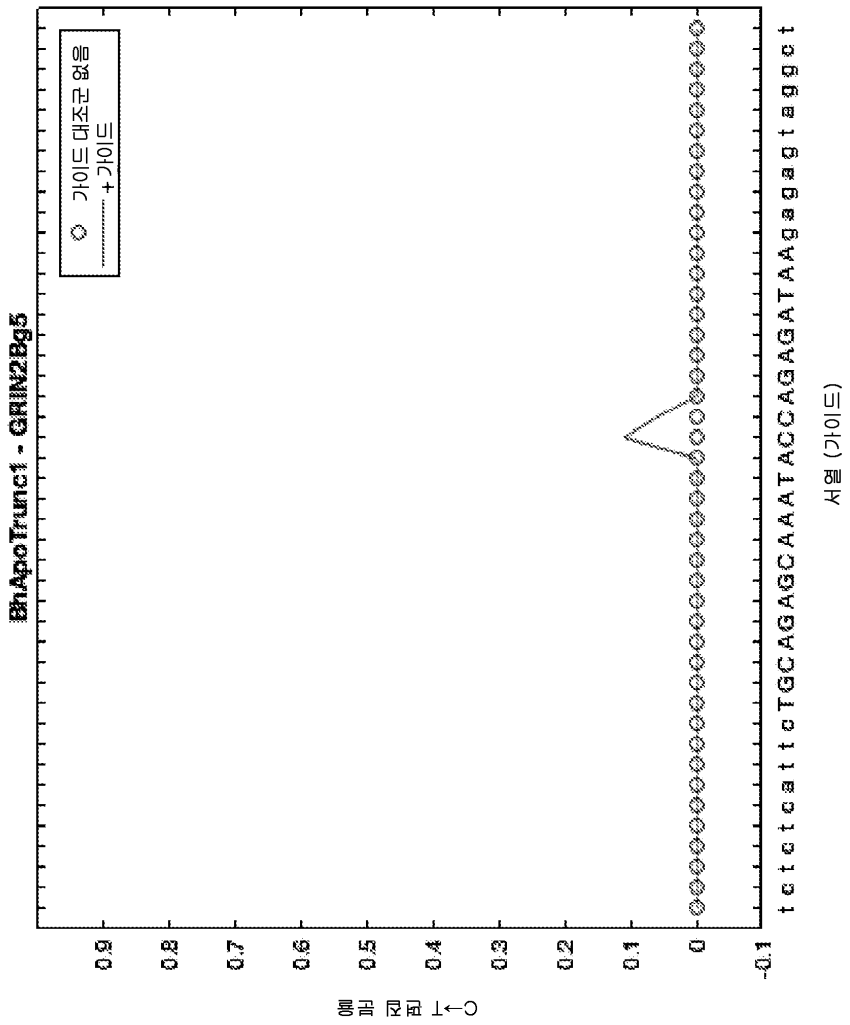
도면79



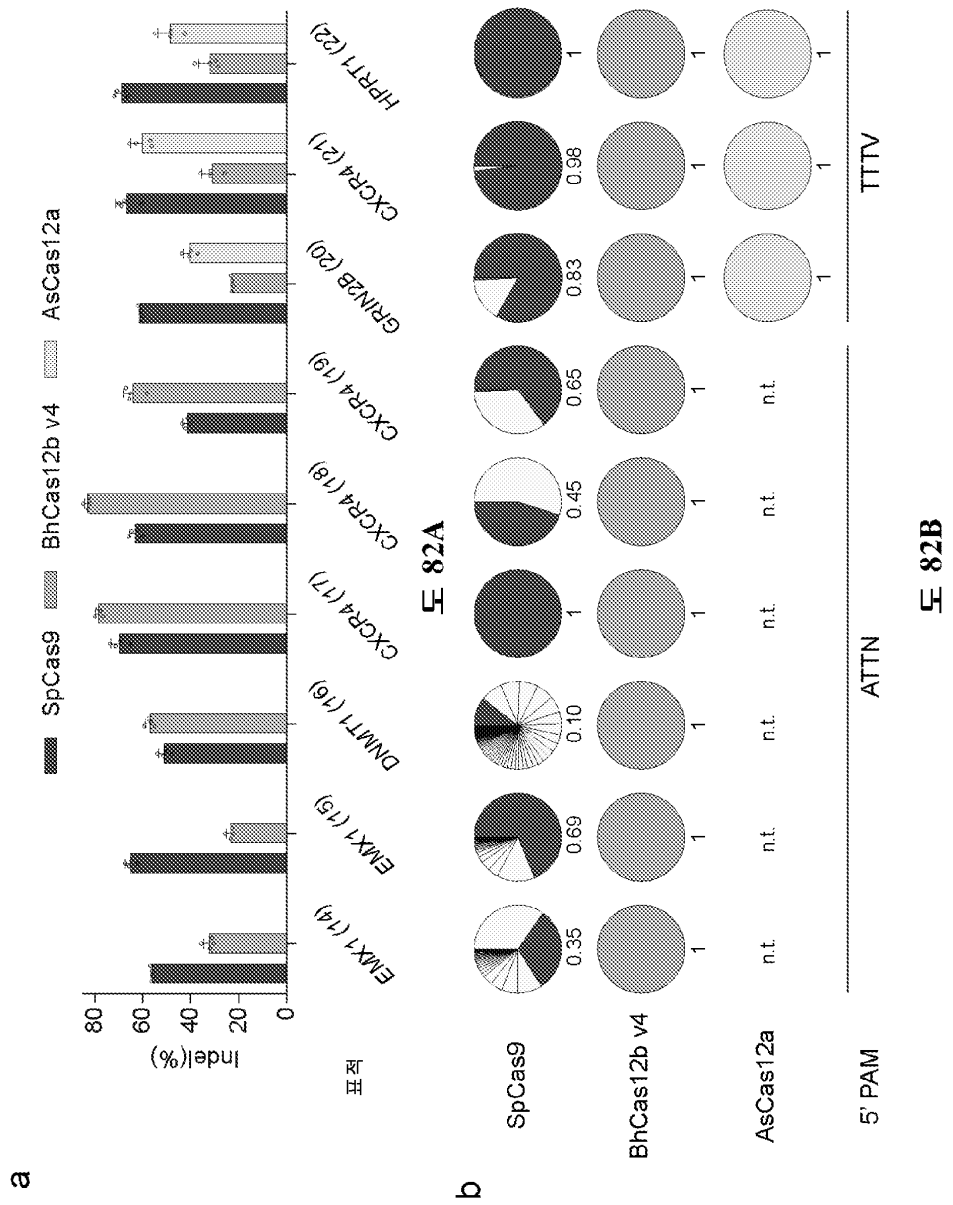
도면80


2일차 세포 트랙커 그린에 의한 HEK293FT CTNNB1 이동성
형광도
25000
20000
15000
10000
5000
0
30_9 Cells
30_9 트랜펙션
30_11 트랜펙션
30_9 Cells
30_11 Cells
비표적화
T41A ORF
비형질감염
S33F_CTNNB1_30bp_가이드
H36Y_CTNNB1_30bp_가이드
S37F_CTNNB1_30bp_가이드
T41I_CTNNB1_30bp_가이드
P44L_CTNNB1_30bp_가이드
P44S_CTNNB1_30bp_가이드
S45F_CTNNB1_30bp_가이드

도면81


EhApoTrunc1 - GHIN2Bg5
가이드 대조군 없음
+가이드
C→T 편집 분율
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
서열 (가이드)

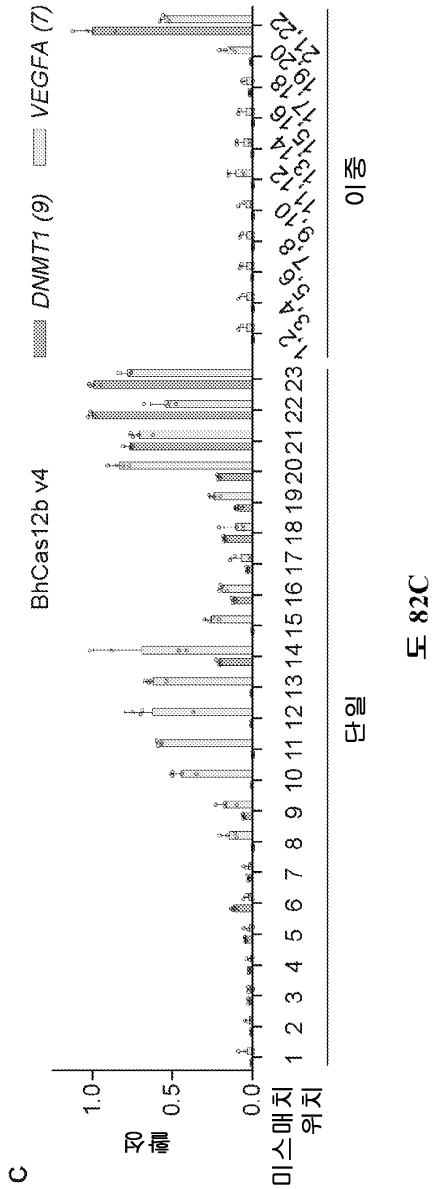
도면82a



도 82A

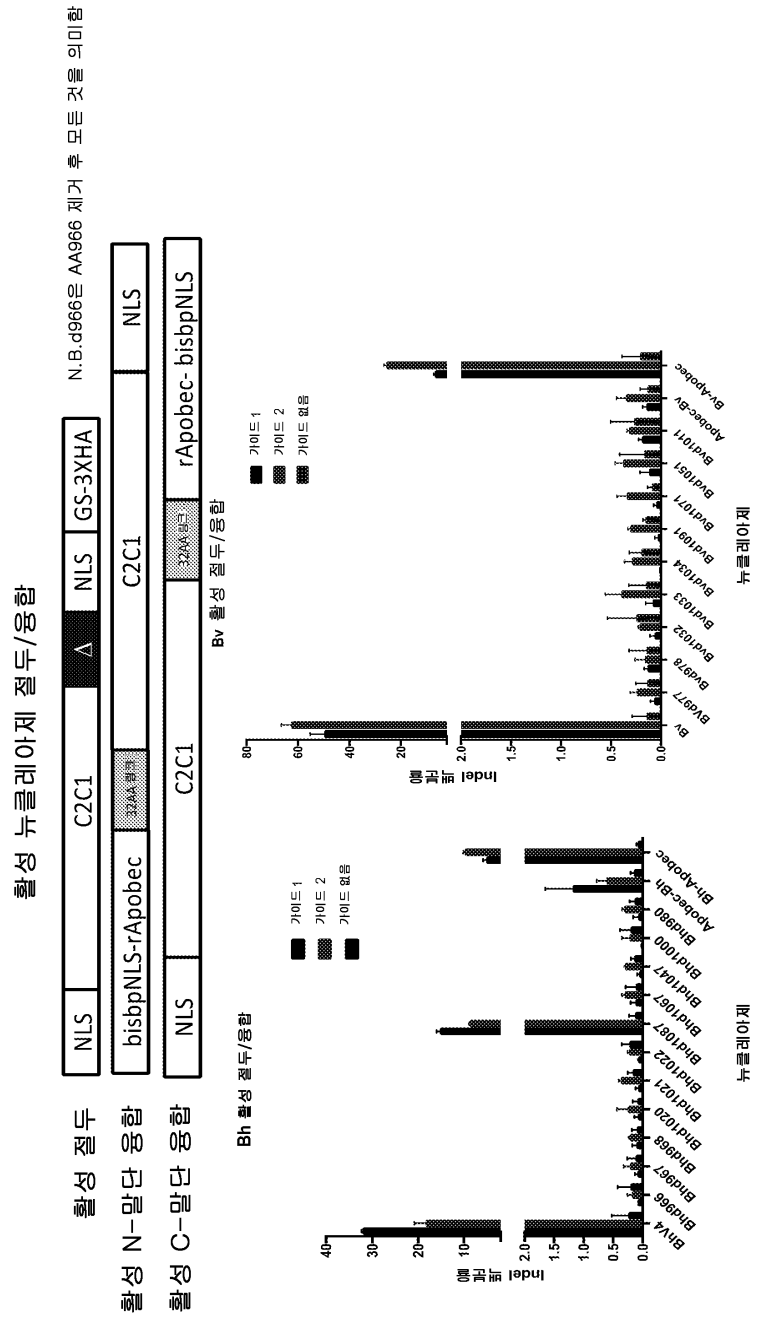
도 82B

도면82b

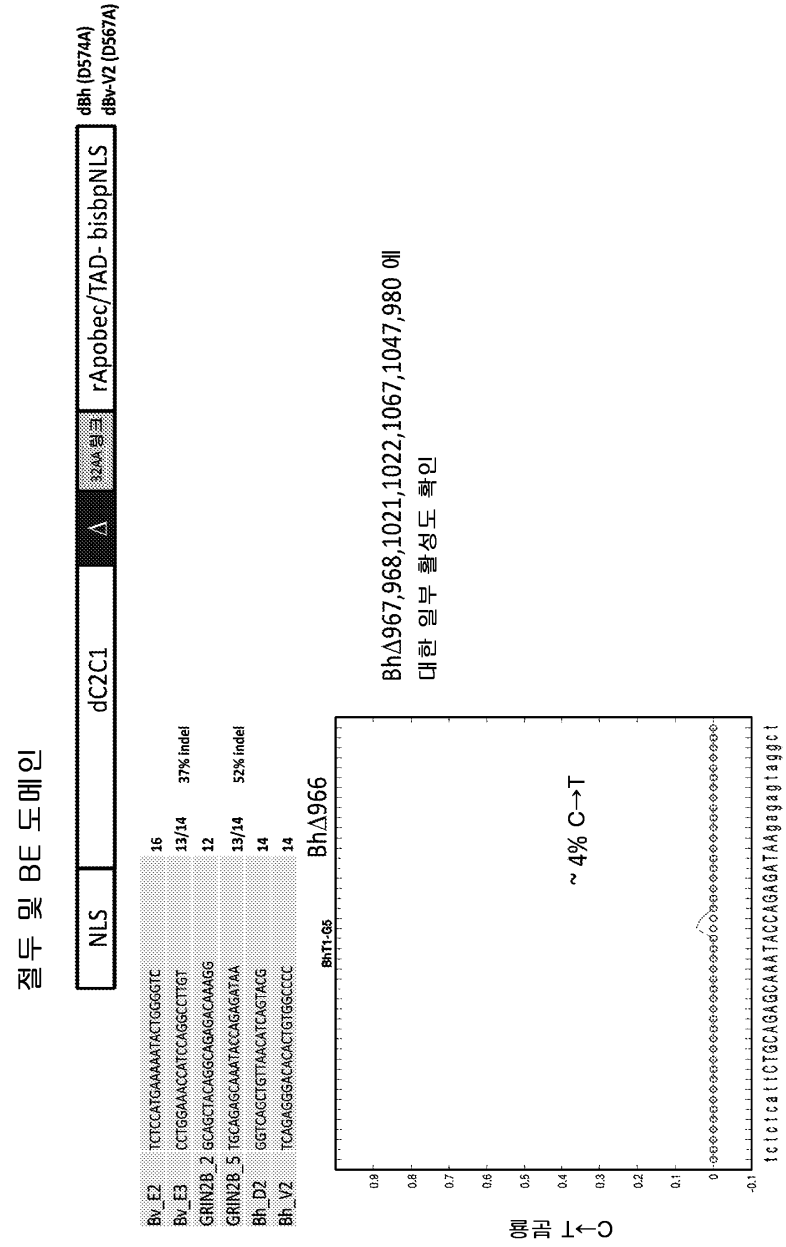

C
BhCas12b v4
DMMT1 (9)
VEGFA (7)
활성
1.0
0.5
0.0
미스매치 위치
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
단일
이중

도 82C

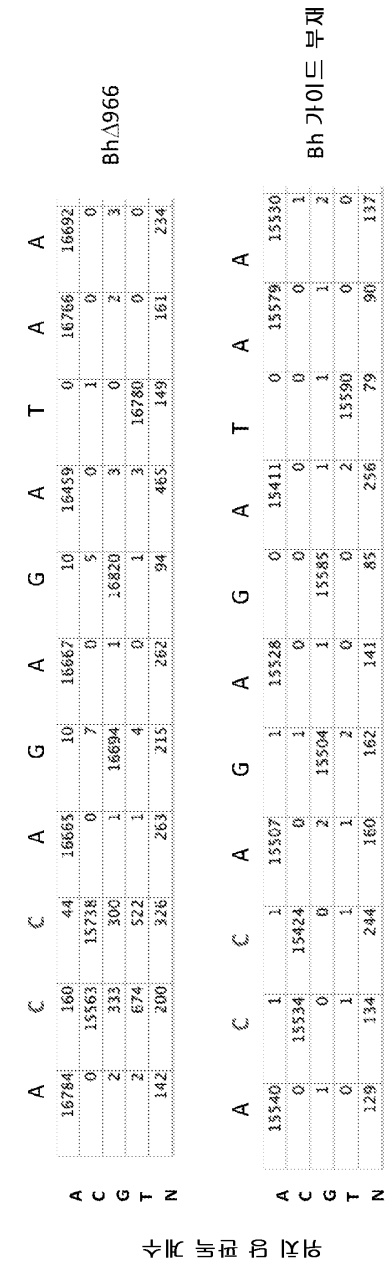
도면83


활성 뉴클레아제 절두/융합
활성 절두
활성 N-말단 융합
활성 C-말단 융합
NLS
C2C1
Δ
NLS
GS-3XHA
N.B.d966은 AA966 제거 후 모든 것을 의미함
bisbpNLS-rApobec
22AA 링커
C2C1
NLS
NLS
C2C1
32AA 링커
rApobec- bisbpNLS
Bh 활성 절두/융합
가이드 1
가이드 2
가이드 없음
Indel 빈도율
뉴클레아제
Bv 활성 절두/융합
가이드 1
가이드 2
가이드 없음
Indel 빈도율
뉴클레아제

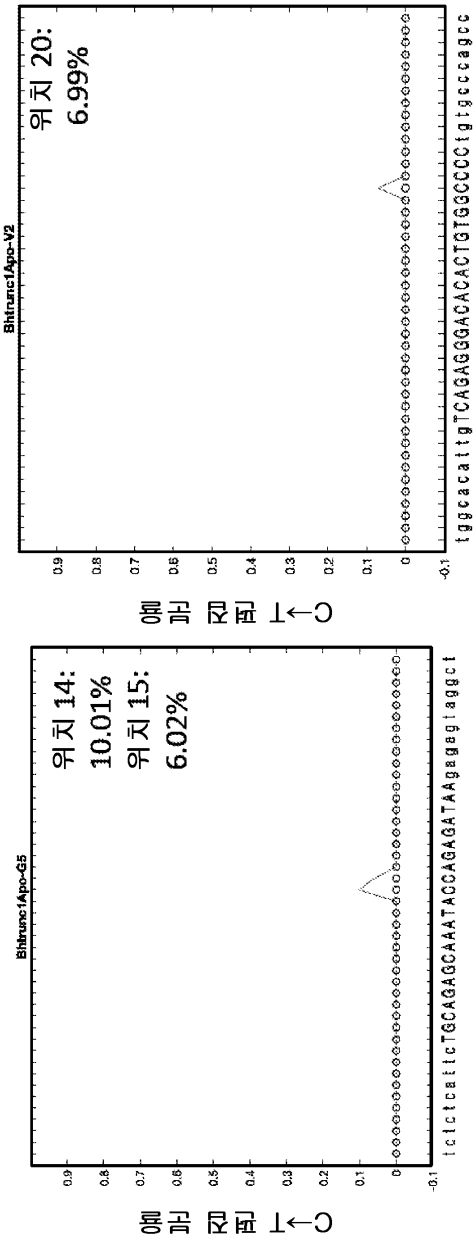
도면84


절두 및 BE 도메인
NLS
dC2C1
Δ
32AA링크
rApobec/TAD- bisbpNLS
dBh (D574A)
dBv-V2 (D567A)
Bv_E2 TCTCCATGAAAAATACTGGGGTC 16
Bv_E3 CCTGGAAACCATCCAGGCCTTGT 13/14 37% indel
GRIN2B_2 GCAGCTACAGGCAGAGACAAAGG 12
GRIN2B_5 TGCAGAGCAAATACCAGAGATAA 13/14 52% indel
Bh_D2 GGTCAGCTGTTAACATCAGTACG 14
Bh_V2 TCAGAGGGACACACTGTGGCCCC 14
BhT1-65
BhΔ966
~ 4% C→T
C→T 빈율
BhΔ967,968,1021,1022,1067,1047,980 에
대한 일부 활성도 확인

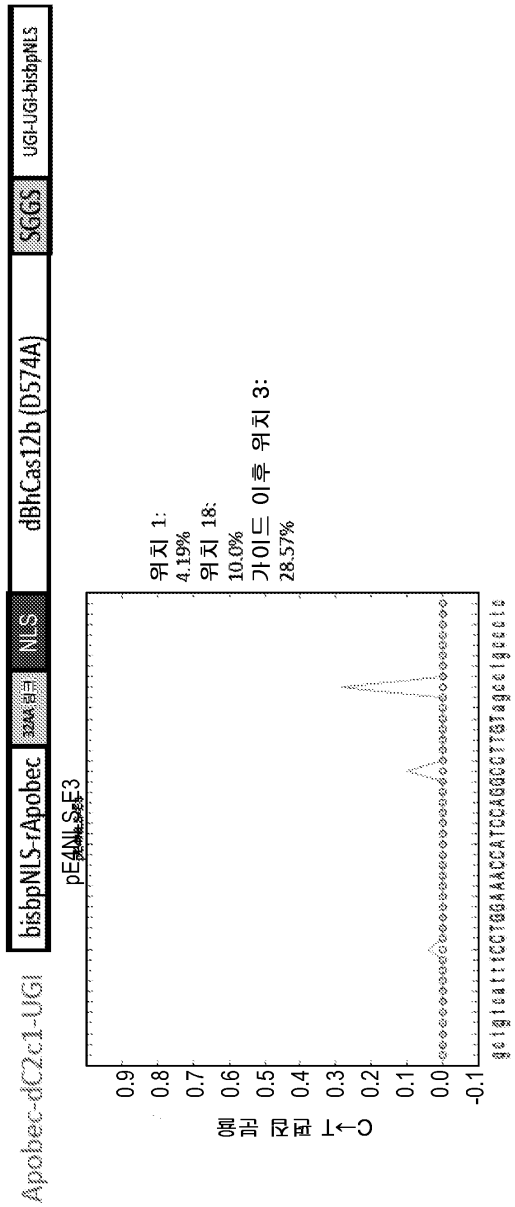
도면85


Bh 가이드 부재
BhΔ966
위치 당 판독 계수
A C G T N
A C G T N
A C G T N

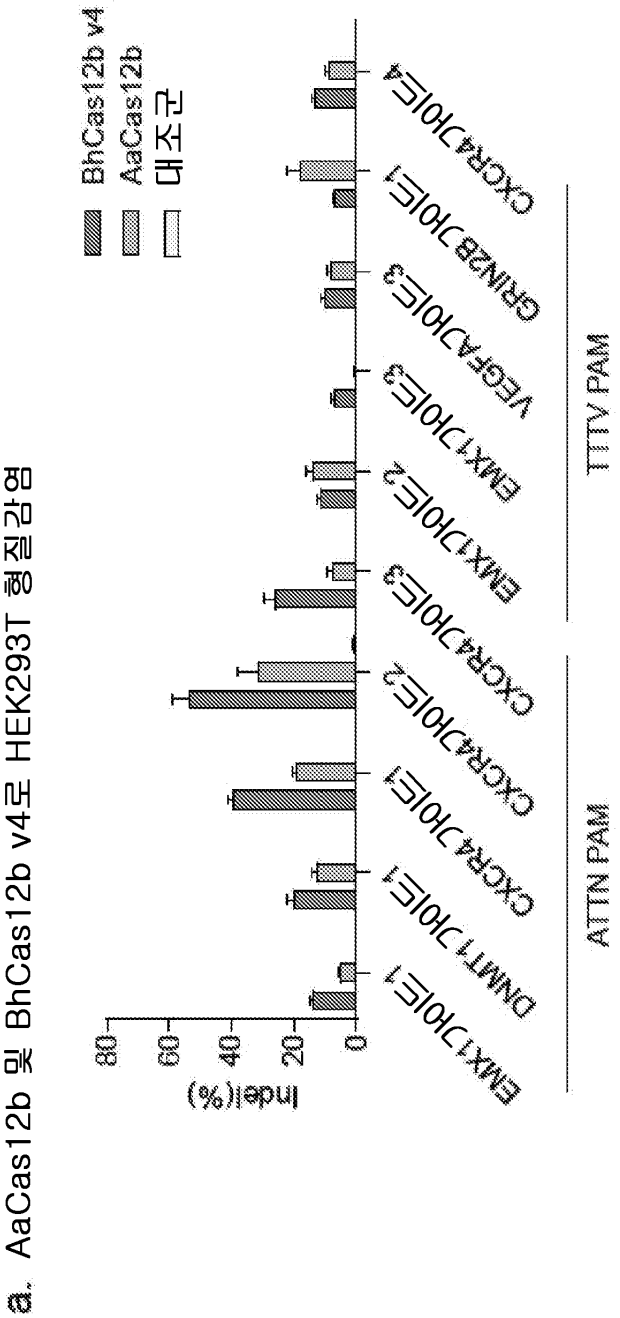
도면86


Bhtrunc1Apo-V2
위치 20:
6.99%
tggcacattgTCAGAGGGACACACTGTGGCCCCtgtgcccagcc
C→T 편집 빈율
VEGFA-가이드2
Bhtrunc1Apo-G5
위치 14:
10.01%
위치 15:
6.02%
tctctcatttcTGCAGAGCAAATACCAGAGATAAggaggtaggct
C→T 편집 빈율
GRIN2b-가이드5

도면87

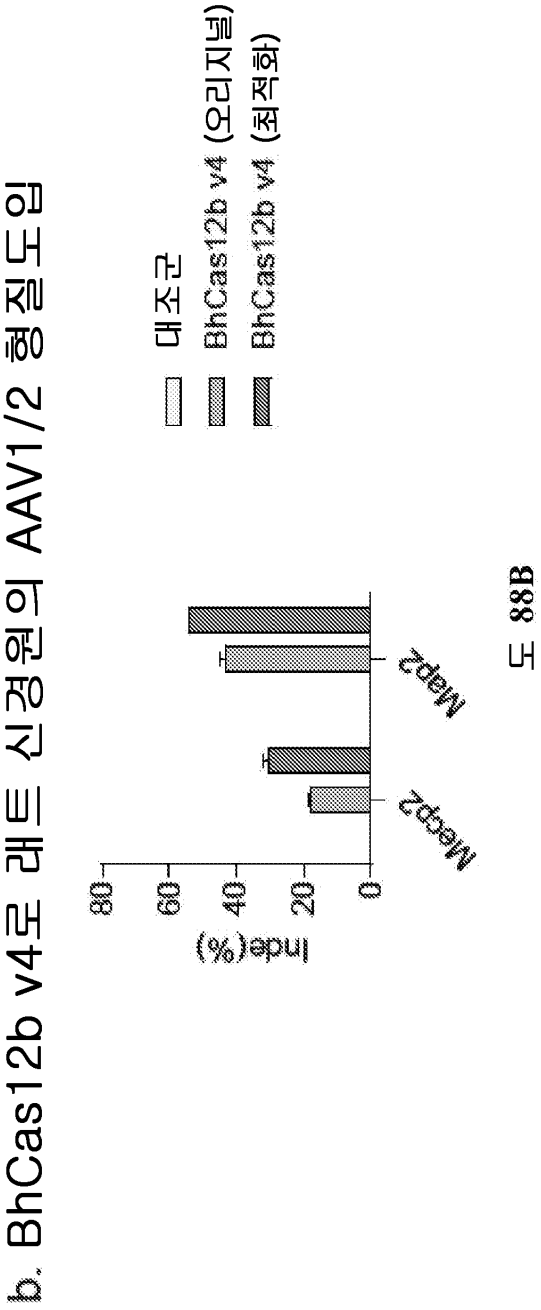

Apobec-dC2c1-UGI
bisbpNLS-rApobec
22A 링커
NLS
dBhCas12b (D574A)
SGGS
UGI-UGI-bisbpNLS
pE4NL-sgE3
C→T 편집 분율
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
위치 1:
4.19%
위치 18:
10.0%
가이드 이후 위치 3:
28.57%

도면88a


a. AaCas12b 및 BhCas12b v4로 HEK293T 형질감염
BhCas12b v4
AaCas12b
대조군
Indel(%)
0
20
40
60
80
EMX1 가이드 1
DNMT1 가이드 1
CXCR4 가이드 1
CXCR4 가이드 2
CXCR4 가이드 3
EMX1 가이드 2
EMX1 가이드 3
VEGFA 가이드 3
GRIN2B 가이드 1
CXCR4 가이드 4
ATTN PAM
TTTV PAM

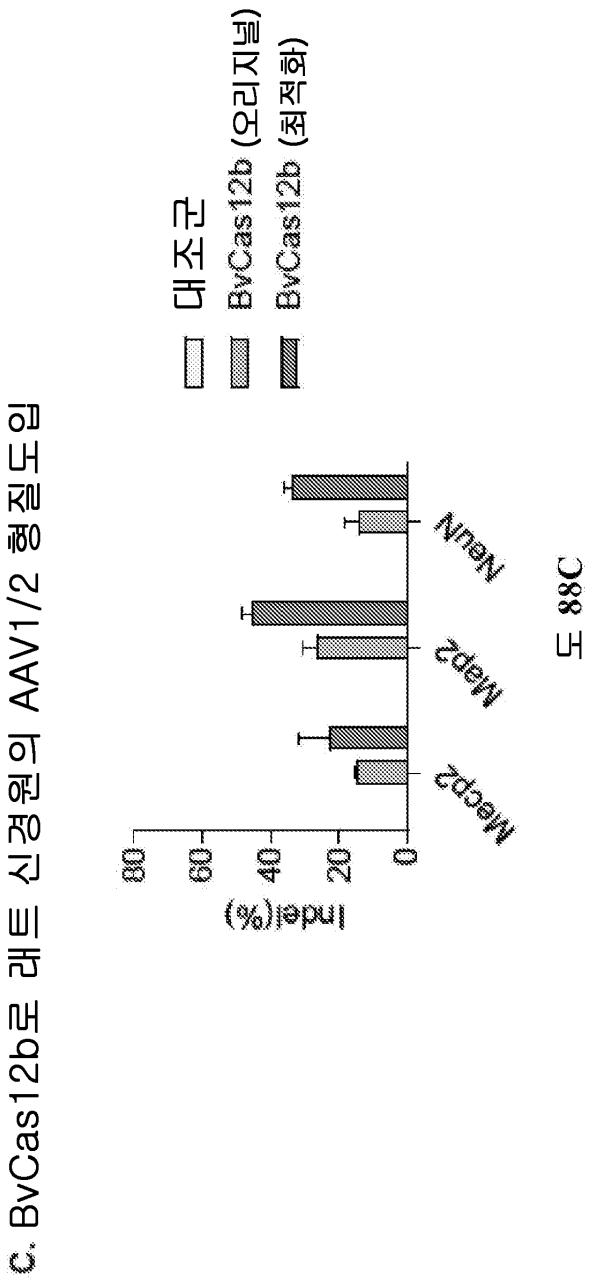
도 88A

도면88b


b. BhCas12b v4로 래트 신경원의 AAV1/2 형질도입
대조군
BhCas12b v4 (오리지널)
BhCas12b v4 (최적화)
Indel(%)
0
20
40
60
80
Mecp2
Map2

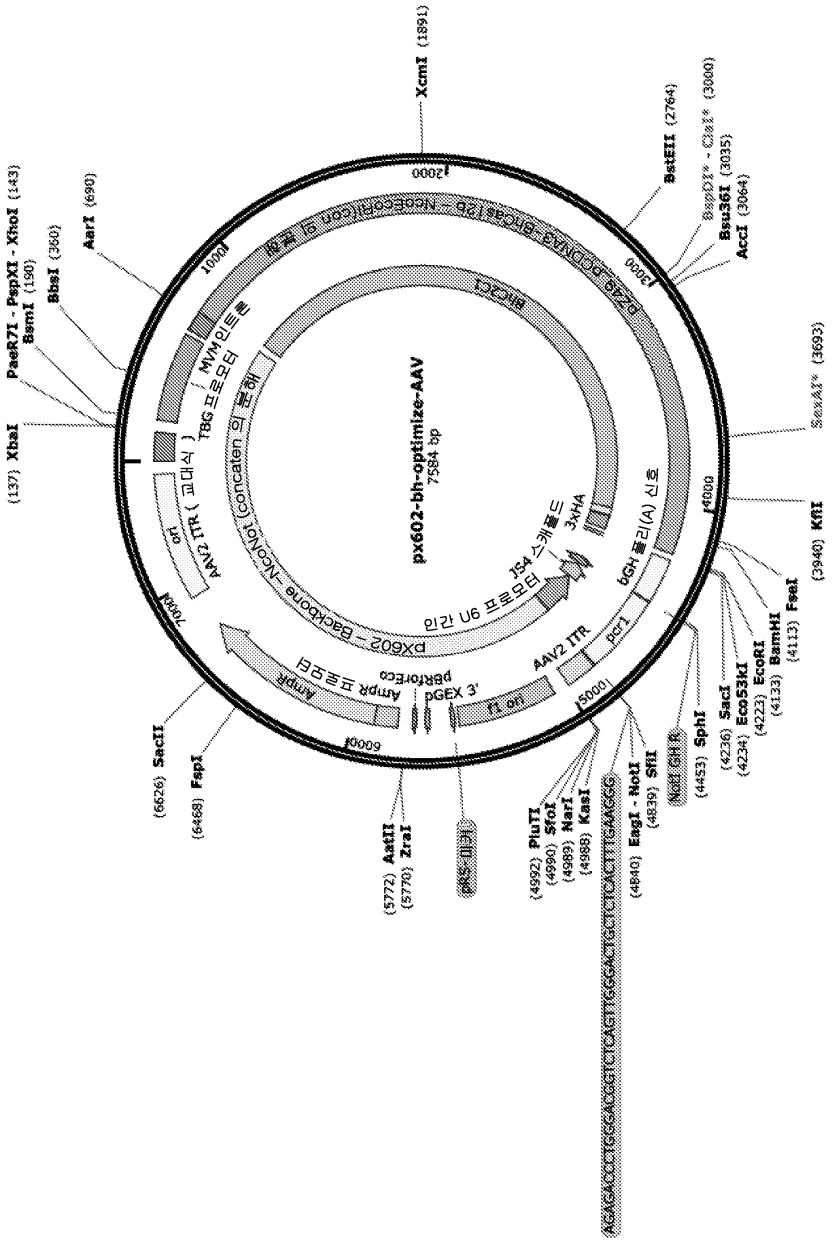
도 88B

도면88c


C. BvCas12b로 래트 신경원의 AAV1/2 형질도입
대조군
BvCas12b (오리지널)
BvCas12b (최적화)
Indel(%)
0
20
40
60
80
Mecp2
Map2
Neun

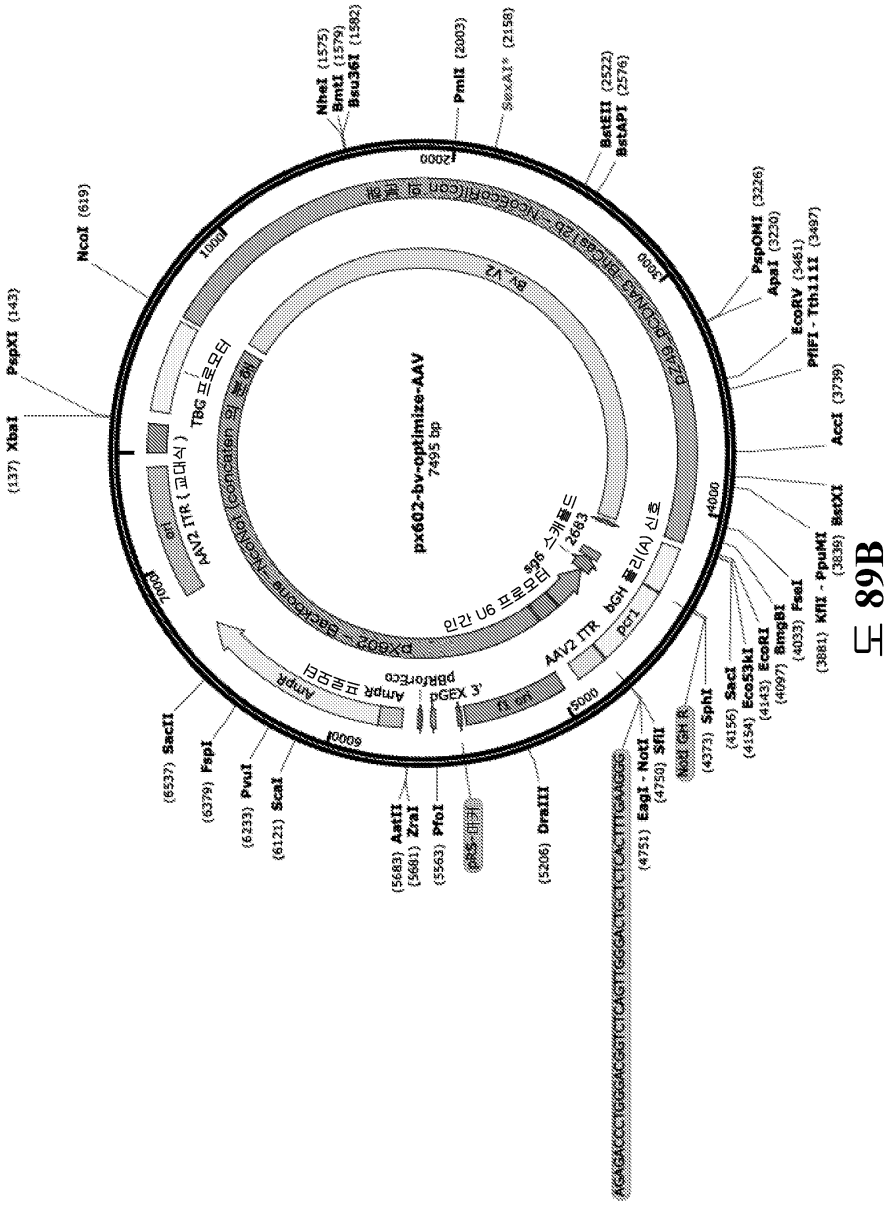
도 88C

도면89a


px602-bh-optimize-AAV
7584 bp
pBR322 ori
AAV2 ITR (교대식)
T86 프로모터
NVM 인트론
Concaten 의 분해
인간 U6 프로모터
3xHA
bGH 폴리(A) 신호
AAV2 ITR
px602 - Backbone -NcoNotOpt
AmpR 프로모터
pGEX 3'
f1 ori
pcr1
AmpR
pBRforEco
pRS-SRH
XcmI (1891)
BstEII (2764)
Bsu36I (3035)
AccI (3064)
SexAI* (3693)
KflI (3940)
FseI (4113)
BamHI (4133)
EcoRI (4223)
Eco53kI (4234)
SacI (4236)
SphI (4453)
NotI GHR (4839)
SfiI (4840)
EagI - NotI (4839)
MluI (4988)
NarI (4989)
SfoI (4990)
PluTI (4992)
ZraI (5770)
AatII (5772)
FspI (6468)
SacII (6626)
ACAGACCCTGCGCAAGGTCTCAGTTGGGACTGCTCTCACTTTGAAGCGC

도 89A

도면89b



도 89B

서 열 목 록

SEQUENCE LISTING

<110> The Broad Institute, Inc.

Massachusetts Institute of Technology

Zhang, Feng

Strecker, Jonathan

Slaymaker, Ian

Jones, Sara

<120> Novel CAS12B Enzymes and Systems

<130> BROD-2670US

<150> 62/715,640

<151> 2018-08-07

<150> 62/744,080

<151> 2018-10-10

<150> 62/751,196

<151> 2018-10-26

<150> 62/794,929

<151> 2019-01-21

<150> 62/831,028

<151> 2019-04-08

<150> PCT/US2019/045582

<151> 2019-08-07

<160

> 663

<170> PatentIn version 3.5

<210> 1

<211> 188

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 1

cgccuaucag ccaacaugcu cgcuuugcga aggcugacgg cccgcucuca uuuggcauug 60

ccgggagccg gaguuuucgg aagagagugu cgacgacugc ugaucuccgc auccgcgucc 120

uguucgccag gccgggucgg guguacggau caugcuggca gcagucuacg ccgagaacau 180

ucgcuuuu 188

<210> 2

<211> 60

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 2

cggaagagag ugucgacgac ugcugaucuc cgcauccgcg uccuguucgc caggccgggu 60

<210> 3

<211> 62

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 3

ccaacaugcu cgcuuugcga aggcugacgg cccgcucuca uuuggcauug ccgggagccg 60

ga 62

<210> 4

<211> 118

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 4

ucgcuuugcg aaggcugacg gcccgcucuc auuuggcauu gccgggagcc ggaguuuucg 60

gaagagagug ucgacgacug cugaucuccg cauccgcguc cuguucgcca ggccgggu 118

<210> 5

<211> 87

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 5

augcucgcuu ugcgaaggcu gacggcccgc ucucauuugg cauugccggg agccggaguu 60

uucggaagag agugucgacg acugcug 87

<210> 6

<211> 71

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 6

aguuuucgga agagaguguc gacgacugcu gaucuccgca uccgcguccu guucgccagg 60

ccgggucggg u 71

<210> 7

<211> 61

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 7

cgccuaucag ccaacaugcu cgcuuugcga aggcugacgg cccgcucuca uuuggcauug 60

c 61

<210> 8

<211> 87

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 8

cgccuaucag ccaacaugcu cgcuuugcga aggcugacgg cccgcucuca uuuggcauug 60

ccgggagccg gaguuuucgg aagagag 87

<210> 9

<211> 84

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 9

ugaucuccgc auccgcgucc uguucgccag gccgggucgg guguacggau caugcuggca 60

gcagucuacg ccgagaacau ucgc 84

<210> 10

<211> 91

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 10

auugccggga gccggaguuu ucggaagaga gugucgacga cugcugaucu ccgcauccgc 60

guccuguucg ccaggccggg ucgggguguac g 91

<210> 11

<211> 83

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 11

tccgcatccg cgtcctgttc gccaggccgg gtcgggtgta cggatcatgc tggcagcagt 60

ctacgccgag aacattcgct ttt 83

<210> 12

<211> 60

<212> RNA

<213> Artificial Sequence

<220><223

> Synthetic Oligonucleotide

<400> 12

cggaagagag ugucgacgac ugcugaucuc cgcauccgcg uccuguucgc caggccgggu 60

<210> 13

<211> 85

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 13

aguuuucgga agagaguguc gacgacugcu gaucuccgca uccgcguccu guucgccagg 60

ccgggucggg uccgcgaaug gacac 85

<210> 14

<211> 129

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 14

tagggaatat tatataatgg acttacgagg ttctgtcttt tggtcaggac aaccgtctag 60

ctataagtgc tgcagggtgt gagaaactcc tattgctgga cgatgtctct tgatacgagg 120

cattagcac 129

<210> 15

<211> 75

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 15

gacgtctagc tataagtgct gcagggtgtg agaaactcct attgctggac gatgtctctt 60

acgaggcatt agcac 75

<210> 16

<211> 106

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 16

acgaggttct gtcttttggt caggacaacc gtctagctat aagtgctgca gggtgtgaga 60

aactcctatt gctggacgat gtctctatga tacgaggcat tagcac 106

<210> 17

<211> 133

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 17

acgaggttct gtcttttggt caggacaacc gtctagctat aagtgctgca gggtgtgaga 60

aactcctatt gctggacgat gtctctttta tttctttttt gtagaaaaaa gaaatgatac 120

gaggcattag cac 133

<210> 18

<211> 97

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 18

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgatatctc ttacgaggca ttagcac 97

<210> 19

<211> 84

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 19

gtcaggacaa ccgtctagct ataagtgctg cagggtgtga gaaactccta ttgctggacg 60

atgtctctta cgaggcatta gcac 84

<210> 20

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 20

ggacttacga ggttctgtct tttggtcagg acaaccgtct agctataagt gctgcagggt 60

gtgagaaact cctattgctg gacgatgtct cttacgaggc attagcac 108

<210> 21

<211> 102

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 21

acgaggttct gtcttttggt caggacaacc gtctagctat aagtgctgca gggtgtgaga 60

aactcctatt gctggacgat gtctcttacg aggcattagc ac 102

<210> 22

<211> 129

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 22

tagggaatat tatataatgg acttacgagg ttctgtcttt tggtcaggac aaccgtctag 60

ctataagtgc tgcagggtgt gagaaactcc tattgctgga cgatgtctct ttatacgagg 120

cattagcac 129

<210> 23

<211> 121

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 23

attatatatt ggacttacga ggttctgtct tttggtcagg acaaccgtct agctataagt 60

gctgcagggt gtgagaaact cctattgctg gacgatgtct ctttatacga ggcattagca 120

c 121

<210> 24

<211> 113

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 24

atggacttac gaggttctgt cttttggtca ggacaaccgt ctagctataa gtgctgcagg 60

gtgtgagaaa ctcctattgc tggacgatgt ctctttatac gaggcattag cac 113

<210> 25

<211> 105

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 25

acgaggttct gtcttttggt caggacaacc gtctagctat aagtgctgca gggtgtgaga 60

aactcctatt gctggacgat gtctctttat acgaggcatt agcac 105

<210> 26

<211> 160

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 26

tagggaatat tatataatgg acttacgagg ttctgtcttt tggtcaggac aaccgtctag 60

ctataagtgc tgcagggtgt gagaaactcc tattgctgga cgatgtctca tttatttctt 120

aattgtccaa gaattaagaa atgatacgag gcattagcac 160

<210> 27

<211> 152

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 27

attatataat ggacttacga ggttctgtct tttggtcagg acaaccgtct agctataagt 60

gctgcagggt gtgagaaact cctattgctg gacgatgtct catttatttc ttaattgtcc 120

aagaattaag aaatgatacg aggcattagc ac 152

<210> 28

<211> 144

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 28

atggacttac gaggttctgt cttttggtca ggacaaccgt ctagctataa gtgctgcagg 60

gtgtgagaaa ctcctattgc tggacgatgt ctcatttatt tcttaattgt ccaagaatta 120

agaaatgata cgaggcatta gcac 144

<210> 29

<211> 136

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 29

acgaggttct gtcttttggt caggacaacc gtctagctat aagtgctgca gggtgtgaga 60

aactcctatt gctggacgat gtctcattta tttcttaatt gtccaagaat taagaaatga 120

tacgaggcat tagcac 136

<210> 30

<211> 80

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 30

tttgcctaaa gggcaaagaa tactgtgcgt gtgctaagga tggaaaaaat ccattcaacc 60

acaggattac attatttatc 80

<210> 31

<211> 89

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 31

gtcgtctata ggacggcgag gacaacggga actgccaatg tgctctttcc aagagcaaac 60

accccgttgg cttcaagatg accgctcgc 89

<210> 32

<211> 78

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 32

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggattatctt atttctgc 78

<210> 33

<211> 97

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 33

uggcaacucg cccgcuagua caacuucggu ugccccacaa gugaggaccu uucucacgua 60

accguguagg gcaacuuaga cggcaggaua ucuagcc 97

<210> 34

<211> 97

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 34

uggcaacucg cccgcuagua caacuucggu ugccccacaa gugaggaccu uucucacgua 60

accguguagg gcaacuuaga cgccaggaua ucuagcc 97

<210> 35

<211> 138

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 35

gccgaucuau aggacggcag auucaacggg augugccaau gcacucuuuc caggagugaa 60

caccccguug gcuucaacau gaucgcccgc ucaacggucc cuagucggau cguugagcgg 120

gcgaucugag aagggcac 138

<210> 36

<211> 128

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 36

cacggugaag agucuagcgg gcgaguugca ucgcaacucg cccgcuagua caacuucggu 60

ugccccacaa gugaggaccu uucucacgua accguguagg gcaacuuaga ccgcaggaua 120

ucuagccg 128

<210> 37

<211> 112

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 37

cacggugaag agucuagcgg gaucccgcu aguacaacuu cgguugcccc acaagugagg 60

accuuucuca cguaaccgug uagggcaacu uagacggcag gauaucuagc cg 112

<210> 38

<211> 89

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 38

acgagguucu gucuuuuggu caggacaacc gucuagcuau aagugcugca gggugugaga 60

aacuccuauu gcuggacgau gucucuuuu 89

<210> 39

<211> 98

<212> RNA

<213> Artificial Sequence

<220><223

> Synthetic Oligonucleotide

<400> 39

gagguucugu cuuuggucag gacaaccguc uagcuauaag ugcugcaggg ugugagaaac 60

uccuauugcu ggacgauguc ucuacgaggc auuagcac 98

<210> 40

<211> 106

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 40

acgagguucu gucuuuuggu caggacaacc gucuagcuau aagugcugca gggugugaga 60

aacuccuauu gcuggacgau gucucuauga uacgaggcau uagcac 106

<210> 41

<211> 102

<212> RNA

<213>

Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 41

acgagguucu gucuuuuggu caggacaacc gucuagcuau aagugcugca gggugugaga 60

aacuccuauu gcuggacgau gucucuuacg aggcauuagc ac 102

<210> 42

<211> 104

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 42

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggattatctt atttctgcaa aagcagaaat aagatgattg gcac 104

<210> 43

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 43

ggtgacctat agggtcaatg aatctgtgcg tgtgccataa gtaattaaaa attacccacc 60

acaggattat cttatttctg caaaagcaga aataagatga ttggcac 107

<210> 44

<211> 101

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 44

ctatagggtc aatgaatctg tgcgtgtgcc ataagtaatt aaaaattacc caccacagga 60

ttatcttatt tctgcaaaag cagaaataag atgattggca c 101

<210> 45

<211> 88

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 45

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggatcatctt aaaaaagatg attggcac 88

<210> 46

<211> 98

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 46

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggatcatctt atttcaaaag aaataagatg attggcac 98

<210> 47

<211> 88

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 47

gacctatagg gtcaatgaat ctgtgcgtgt gccataagta attaaaaatt acccaccaca 60

ggagcacctg aaaacaggtg cttggcac 88

<210> 48

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 48

gccgtttccc tcactcctgc tcggtgaatt tggctc 36

<210> 49

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 49

gagccaaatt caccgagcag gagtgaggga aacggc 36

<210> 50

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 50

gccgttcccc tcactcctgc tcggcaaatt tagctc 36

<210> 51

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 51

gagctaaatt tgccgagcag gagtgagggg aacggc 36

<210> 52

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 52

gctgtttggg aggtcagaaa taggggtcc aggagc 36

<210> 53

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 53

gctcctggac cccctatttc tgacctccca aacagc 36

<210> 54

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 54

gctgttaggg aggtcagaaa taggatgtcc aagagc 36

<210> 55

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 55

gctcttggac atcctatttc tgacctccct aacagc 36

<210> 56

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 56

ccaaattctt ctcccctggg aagcatccct ggacac 36

<210> 57

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 57

gtgtcgaggt gtgcttccca ggggagaagg atttgg 36

<210> 58

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 58

ccaaatcctt ctcccctggg aagcacacct cgacac 36

<210> 59

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 59

ggtttaggaa gaggggaccc ttcgtgtgga gctgtg 36

<210> 60

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 60

gccgtttccc tcactcctgc tcggtgaatt tggctc 36

<210> 61

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 61

gagccaaatt caccgagcaa gagtgaggga aacggc 36

<210> 62

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 62

gccgttcccc tcactcctgc tcggcaaatt tagctc 36

<210> 63

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 63

gagctaaatt tgccgagcag gagtgagggg aacggc 36

<210> 64

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 64

accatctact ccatcatctt cttaactggc attgtgggca atggattggt catcctggtc 60

atgggttacc agaagaaact gagaagcatg acggacaagt acaggct 107

<210> 65

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 65

agcctgtact tgtccgtcat gcttctcagt ttcttctggt aacccatgac caggatgacc 60

aatccattgc ccacaatgcc agttaagaag atgatggagt agatggt 107

<210> 66

<211> 36

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 66

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

1 5 10 15

Val Ile Leu Val Met Gly Tyr Gln Lys Lys Leu Arg Ser Met Thr Asp

20 25 30

Lys Tyr Arg Leu

35

<210> 67

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 67

gcacctgtca gtggccgacc tcctctttgt catcacgctt cccttctggg cagttgatgc 60

cgtggcaaac tggtactttg ggaacttcct atgcaaggca gtccatg 107

<210> 68

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 68

catggactgc cttgcatagg aagttcccaa agtaccagtt tgccacggca tcaactgccc 60

agaagggaag cgtgatgaca aagaggaggt cggccactga caggtgc 107

<210> 69

<211> 35

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic Peptide

<400> 69

His Leu Ser Val Ala Asp Leu Leu Phe Val Ile Thr Leu Pro Phe Trp

1 5 10 15

Ala Val Asp Ala Val Ala Asn Trp Tyr Phe Gly Asn Phe Leu Cys Lys

20 25 30

Ala Val His

35

<210> 70

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 70

tcatctacac agtcaacctc tacagcagtg tcctcatcct ggccttcatc agtctggacc 60

gctacctggc catcgtccac gccaccaaca gtcagaggcc aaggaag 107

<210> 71

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 71

cttccttggc ctctgactgt tggtggcgtg gacgatggcc aggtagcggt ccagactgat 60

gaaggccagg atgaggacac tgctgtagag gttgactgtg tagatga 107

<210> 72

<211> 36

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 72

Val Ile Tyr Thr Val Asn Leu Tyr Ser Ser Val Leu Ile Leu Ala Phe

1 5 10 15

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

20 25 30

Arg Pro Arg Lys

35

<210> 73

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 73

ctgttggctg aaaaggtggt ctatgttggc gtctggatcc ctgccctcct gctgactatt 60

cccgacttca tctttgccaa cgtcagtgag gcagatgaca gatatat 107

<210

> 74

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 74

atatatctgt catctgcctc actgacgttg gcaaagatga agtcgggaat agtcagcagg 60

agggcaggga tccagacgcc aacatagacc acctttcag ccaacag 107

<210> 75

<211> 36

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 75

Leu Leu Ala Glu Lys Val Val Tyr Val Gly Val Trp Ile Pro Ala Leu

1 5 10 15

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

20 25 30

Asp Arg Tyr Ile

35

<210> 76

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 76

ccgacttcat ctttgccaac gtc 23

<210> 77

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 77

tgggcaatgg attggtcatc ctg 23

<210> 78

<211> 122

<212> RNA

<213> Bacillus hisashii

<220><221> misc_feature

<222> (103)..(122)

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

<400> 78

acgagguucu gucuuuuggu caggacaacc gucuagcuau aagugcugca gggugugaga 60

aacuccuauu gcuggacgau gucucuuacg aggcauuagc acnnnnnnnn nnnnnnnnnn 120

nn 122

<210> 79

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 79

gccgtttccc tcactcctgc tcggtgaatt tggctc 36

<210> 80

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 80

gagccaaatt caccgagcag gagtgaggga aacggc 36

<210> 81

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 81

gccgttcccc tcactcctgc tcggcaaatt tagctc 36

<210> 82

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 82

gccgttcccc tcactcctgc tcggcaaatt tagctc 36

<210> 83

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 83

gagctaaatt tgccgagcag gagtgagggg aacggc 36

<210> 84

<211> 115

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 84

gucgucuaua ggacggcgag gacaacggga agugccaaug ugcucuuucc aagagcaaac 60

accccguugg cuucaagaug accgcucgaa aacgagcggu cugagaagug gcacu 115

<210> 85

<211> 113

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 85

gucgucuaua ggacggcgag gacaacggga agugccaaug ugcucuuucc aagagcaaac 60

accccguugg cuucucagac cgcucgaaaa cgagcggucu gagaaguggc acu 113

<210> 86

<211> 118

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 86

acagucgucu auaggacggc gaggacaacg ggaagugcca augugcucuu uccaagagca 60

aacaccccgu uggcuucaag augaccgcuc gaaaacgagc ggucugagaa guggcacu 118

<210> 87

<211> 112

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 87

gucuauagga cggcgaggac aacgggaagu gccaaugugc ucuuuccaag agcaaacacc 60

ccguuggcuu caagaugacc gcucgaaaac gagcggucug agaaguggca cu 112

<210> 88

<211> 111

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 88

gucgucuaua ggacggcgag gacaacgggg ugccaaugug cucuuuccaa gagcaaacac 60

cccguuggcu ucucagaccg cucgaaaacg agcggucuga gaaguggcac u 111

<210>

89

<211> 105

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 89

gucgucuaua ggacggcgag gacaacggga agugccaaug ugcucuuucc aagagcaaac 60

accccguugg cuucucagac cgaaaacggu cugagaagug gcacu 105

<210> 90

<211> 102

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 90

acgagguucu gucuuuuggu caggacaacc gucuagcuau aagugcugca gggugugaga 60

aacuccuauu gcuggacgau gucucuuacg aggcauuagc ac 102

<210> 91

<211> 97

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 91

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcac 97

<210> 92

<211> 97

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 92

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgacgccuc uuacgaggcg uuagcac 97

<210> 93

<211> 95

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 93

guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgacgccuu acgaggcguu agcac 95

<210> 94

<211> 92

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 94

guucugucuu uuggucagga caaccgucca gcuauaagug cugcagggug cagaaacucc 60

uauugcugga cgacgccuua cggcguuagc ac 92

<210> 95

<211> 115

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 95

guucugucuu uuggucagga caaccgucca gcuauaagug cugcagggug cagaaacucc 60

uauugcugga cgacgccuca uuuauuucaa aagaaauguu acgaggcguu agcac 115

<210> 96

<211> 135

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (116)..(135)

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

<400> 96

gucgucuaua ggacggcgag gacaacggga agugccaaug ugcucuuucc aagagcaaac 60

accccguugg cuucaagaug accgcucgaa aacgagcggu cugagaagug gcacunnnnn 120

nnnnnnnnnn nnnnn 135

<210> 97

<211> 104

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 97

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggauuaucuu auuucugcaa aagcagaaau aagaugauug gcac 104

<210> 98

<211> 107

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 98

ggugaccuau agggucaaug aaucugugcg ugugccauaa guaauuaaaa auuacccaac 60

acaggauuau cuuauuucug caaaagcaga aauaagauga uuggcac 107

<210> 99

<211> 101

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 99

cuauaggguc aaugaaucug ugcgugugcc auaaguaauu aaaaauuacc caccacagga 60

uuaucuuauu ucugcaaaag cagaaauaag augauuggca c 101

<210> 100

<211> 88

<212> RNA

<213> Artificial Sequence

<220><223

> Synthetic Oligonucleotide

<400> 100

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggaucaucuu aaaaaagaug auuggcac 88

<210> 101

<211> 98

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 101

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggaucaucuu auuucaaaag aaauaagaug auuggcac 98

<210> 102

<211> 88

<212> RNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 102

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggagcaccug aaaacaggug cuuggcac 88

<210> 103

<211> 124

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (105)..(124)

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

<400> 103

gaccuauagg gucaaaguau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggauuaucuu auuucugcaa aagcagaaau aagaugauug gcacnnnnnn nnnnnnnnnn 120

nnnn 124

<210> 104

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 104

ccaaattctt ctcccctggg aagcatccct ggacac 36

<210> 105

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 105

gtgtccaggg atgcttccca ggggagaaga atttgg 36

<210> 106

<211> 40

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (1)..(2)

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

<220><221> misc_feature

<222> (39)..(40)

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

<400> 106

nnccaaatcc ttctcccctg ggaagcacac ctcgacacnn 40

<210> 107

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 107

ccaaatcctt ctcccctggg aagcacacct cgacac 36

<210> 108

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 108

gtgtcgaggt gtgcttccca ggggagaagg atttgg 36

<210> 109

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 109

agagcactgg catggggatg ngg 23

<210> 110

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 110

ggaacactgg catggagatg ngg 23

<210> 111

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 111

agaacagtgg catggggatg ngg 23

<210> 112

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 112

tgaacactgg catggggaag ngg 23

<210> 113

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 113

gaagcactgc catgggaatg ngg 23

<210> 114

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 114

agagcactgg aatgggaatg ngg 23

<210> 115

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 115

ggagcgctgg catggggact ngg 23

<210> 116

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 116

tgaacactgg catggggctg ngg 23

<210> 117

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 117

agagcactgg caggaggatg ngg 23

<210> 118

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 118

ggagcactgg catgggatgt nga 23

<210> 119

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 119

gagcccctgg catggggatg ngg 23

<210> 120

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 120

caggcactgg catggggacc ngg 23

<210> 121

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 121

aggatacagg caggggaatg ngg 23

<210> 122

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 122

agaatactgg catgggaatg ngg 23

<210> 123

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 123

gaaggactgg cacggggatg ngg 23

<210> 124

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 124

ggagcactgg taggggatg ngg 23

<210> 125

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 125

agagaactgc catgggaatg ngg 23

<210> 126

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 126

aaagcactgg ccatgggatg ngg 23

<210> 127

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 127

gaggcactgg catggggtg ngg 23

<210> 128

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 128

agagcacggg catgaggatc ngg 23

<210> 129

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 129

attntagagc actggcatgg ggatggg 27

<210> 130

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 130

attntagagc actggcatgg ggatggg 27

<210> 131

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 131

tgctccagag gccccccttg ngg 23

<210> 132

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 132

ggctccagag gctccccttg nga 23

<210> 133

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 133

tgctccagag gctccccttg ngg 23

<210> 134

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 134

ggctccagag gccccccctgc ngg 23

<210> 135

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 135

agctccagag ccccccttg nga 23

<210> 136

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 136

tgccccagag gccccctca ngg 23

<210> 137

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 137

ggctccagag gctcccctgt ngg 23

<210> 138

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 138

tgctccaggg acccccctct ngg 23

<210> 139

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 139

agctccagac actccccttg ngg 23

<210> 140

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 140

gtctccagag gctcccctga ngg 23

<210> 141

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 141

ggctccaaaa acccccctttngg 23

<210> 142

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 142

tgctccaagg gccccccttgnag 23

<210> 143

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 143

aggtccagag gccacccttgngg 23

<210> 144

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 144

atctccagag gcctccctggngg 23

<210> 145

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 145

cctgccccag aggtcccttg ngg 23

<210> 146

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 146

tgctccagag ctccccttct nga 23

<210> 147

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 147

ctatccagag gctcccctcc ngg 23

<210> 148

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 148

ctcctcagag gccccctgg ngg 23

<210> 149

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 149

ggctccaggg gccccacttg ngg 23

<210> 150

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 150

ttttccagag gccccctcc ngg 23

<210> 151

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 151

attnttgctc cagaggcccc ccttggg 27

<210> 152

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 152

attnttgctc cagaggcccc ccttggg 27

<210> 153

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 153

ggtgccagaa acaggggtga ngg 23

<210> 154

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 154

ggtaccagaa acaggggggа ngg 23

<210> 155

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 155

gctgccagga acagaggtga ggg 23

<210> 156

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 156

ggttccagaa agaggggtga ngg 23

<210> 157

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 157

aatgccagga acaggggtgg ngg 23

<210> 158

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (20)..(20)

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

<400> 158

gggccagaaa caggggtgan gg 22

<210> 159

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 159

caagccagaa acaggggtga ngg 23

<210> 160

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 160

ggtatcagaa gcaggggtga ngg 23

<210> 161

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 161

gaggccagaa aaaggggtga ngg 23

<210> 162

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 162

ggtggcagga acaggggtgg ngg 23

<210> 163

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 163

agtgacagaa acaggaatga ngg 23

<210> 164

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 164

gaagccagaa gcaggggtga ngg 23

<210> 165

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 165

ggagccagaa gcaggggaga ngg 23

<210> 166

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 166

ggagccagaa acagggatgg ngg 23

<210> 167

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 167

agtaccagga acaggggagt ngg 23

<210> 168

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 168

agtgccagga gcaggggaga ngg 23

<210> 169

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 169

ggtaccaaaa aaaggggaga ngg 23

<210> 170

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 170

gatgccagag gcaggggaga ngg 23

<210> 171

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 171

agggccagaa acaggaggga ngg 23

<210> 172

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 172

ggtgccagga acgggggtga ngg 23

<210> 173

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 173

attnctggtg ccagaaacag gggtgac 27

<210> 174

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 174

attnctggtg ccagaaacag gggtgac 27

<210> 175

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 175

gggcttcaag caacttgtag cgg 23

<210> 176

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 176

attntgggct tcaagcaact tgtagtg 27

<210> 177

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 177

attntgggct tcaagcaact tgtagtg 27

<210> 178

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 178

gtaattggtt ctaccaaaga ngg 23

<210> 179

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 179

ttaattggtt ctaccaaaga ngg 23

<210> 180

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 180

attntgtaat tggttctacc aaagaag 27

<210> 181

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 181

attntgtaat tggttctacc aaagaag 27

<210> 182

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 182

gaggcggagg gcggcgtgcc ngg 23

<210> 183

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 183

aaggccgagg gcagcatgcc ngg 23

<210> 184

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 184

tgggcggagg gaggcgcacc ngg 23

<210> 185

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 185

attnagaggc ggagggcggc gtgcctg 27

<210> 186

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 186

attnagaggc ggagggcggc gtgcctg 27

<210> 187

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 187

agtgtccagg gatgcttccc ngg 23

<210> 188

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 188

catccccagg gaagcatccc ngg 23

<210> 189

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 189

agtggctagg gatgcttccc ngg 23

<210> 190

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 190

atagtccagg gatgcttcct ngg 23

<210> 191

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 191

cgtgaccagg gatgcttctc ngg 23

<210> 192

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 192

gctccccggg gaagcttccc nga 23

<210> 193

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 193

catgtccaga aatgcttccc ngg 23

<210> 194

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 194

agtgtccagg gatgcttccc nag 23

<210> 195

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 195

tcggtccagg gatgcttccc ntg 23

<210> 196

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 196

caagtccatg gatgcttccc ngg 23

<210> 197

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 197

ccagtccagg aaagcatccc ngg 23

<210> 198

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 198

cctggcctgg ggagcatccc nga 23

<210> 199

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 199

tgtgtgcagg gatgcttcct ngg 23

<210> 200

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 200

agactccagg gaagcttccc ngg 23

<210> 201

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 201

agggtccagg cacgcttccc ngg 23

<210> 202

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 202

tgtttccagg aatgcttccc ngg 23

<210> 203

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 203

agagtccagg aatgcttccc nga 23

<210> 204

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 204

atgtcacagg gattcttccc ngg 23

<210> 205

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 205

caagtccaga gatgcttcct ngg 23

<210> 206

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 206

agtgtccagg gaagcttcgc ngg 23

<210> 207

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 207

tttnggaagt gtccagggat gcttccc 27

<210> 208

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 208

tttnggaagt gtccagggat gcttccc 27

<210> 209

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 209

tttnggaagt gtccagggat gcttccc 27

<210> 210

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 210

ttcagcccaa gaacagtaca ngg 23

<210> 211

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 211

aacagtccaa gaacagtaca ngg 23

<210> 212

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 212

tatagctcaa gaacagtaca ngg 23

<210> 213

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 213

tttncttcag cccaagaaca gtacaag 27

<210> 214

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 214

tttncttcag cccaagaaca gtacaag 27

<210> 215

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 215

tttncttcag cccaagaaca gtacaag 27

<210> 216

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 216

gtgagtcgag gagaaacgac ngg 23

<210> 217

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 217

atgagccaag gagagacgaa ngg 23

<210> 218

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 218

atgagtagag gagaaaagac ngg 23

<210> 219

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 219

tttntctgtg agtcgaggag aaacgac 27

<210> 220

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 220

tttntctgtg agtcgaggag aaacgac 27

<210> 221

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 221

tttntctgtg agtcgaggag aaacgac 27

<210> 222

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 222

gggtgtgtta aaagtgacca ngg 23

<210> 223

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 223

tttncttggg tgtgttaaaa gtgacca 27

<210> 224

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 224

tttncttggg tgtgttaaaa gtgacca 27

<210> 225

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 225

tttncttggg tgtgttaaaa gtgacca 27

<210> 226

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 226

tcactcctgc tcggtgaatt ngg 23

<210> 227

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 227

gctacaggca gagacaaagg cgg 23

<210> 228

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 228

aggacaggca gagacaaagg cgg 23

<210> 229

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 229

agaacaggca gagacaaaga cgg 23

<210> 230

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 230

aagacaggca gagacaaagt cgg 23

<210> 231

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 231

gctagaggta gagacaaagg cgg 23

<210> 232

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 232

agaacaggca gagacaaagg cag 23

<210> 233

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 233

aggacaggca gagacaaagg cgg 23

<210> 234

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 234

aggacaggca gagacaaagt cgg 23

<210> 235

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 235

cggacaggca gagacaaagg cgg 23

<210> 236

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 236

acaacagcca gagacaaagg cgg 23

<210> 237

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 237

agaacaggca gagacaaagt cgg 23

<210> 238

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 238

cctagaggca gagacaaagt cgg 23

<210> 239

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 239

gctacagaca gagacaaagg cag 23

<210> 240

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 240

gctacaggca gagacaaagg cgg 23

<210> 241

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 241

actacaaaca gagacaaaga cgg 23

<210> 242

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 242

tctacagtca gagacaaagg cga 23

<210> 243

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 243

aagataggca gagacaaagg cgg 23

<210> 244

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 244

cctataggca gagacaaagc cgg 23

<210> 245

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 245

aagacaggca gagacaaagt cgg 23

<210> 246

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 246

aggacaggca gagacaaagg cag 23

<210> 247

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 247

aggtcagaaa taggggtcc ngg 23

<210> 248

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 248

agggcacaaa taggaggtcc ngg 23

<210> 249

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 249

caggctgtga accttggtgg cgg 23

<210> 250

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 250

ctggctgtga actttggtgg cgg 23

<210> 251

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 251

aggactgtag accttggtgg cgg 23

<210> 252

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 252

atagctgtga acctaggtgg cgg 23

<210> 253

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 253

caggctgtga accttggtga cgg 23

<210> 254

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 254

cctgctatga accttggtgt cgg 23

<210> 255

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 255

ggggctgcgg accttggtgg cgg 23

<210> 256

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 256

aaagctgaga accttggtgg cgg 23

<210> 257

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 257

cttgctggga accttggtgg cgg 23

<210> 258

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 258

caggctgaga cccttggtgg cgg 23

<210> 259

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 259

caggctgtgc actttggtgg cgg 23

<210> 260

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 260

attgctgcaa accttggtgt cgg 23

<210> 261

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 261

taggctgtaa aacttggtgg cgg 23

<210> 262

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 262

aaagctggga accttggtgg cgg 23

<210> 263

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 263

aggcctgaga accttggtgg cag 23

<210> 264

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 264

caggctttga actttggtag cgg 23

<210> 265

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 265

caggctgtta cccttggtgg cgg 23

<210> 266

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 266

gcagcctgta gccttggtgg cgg 23

<210> 267

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 267

agggatgtga accttggtag cag 23

<210> 268

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 268

cctgctgcaa accttggtgt cgg 23

<210> 269

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 269

gacccctcc accccgcctc ngg 23

<210> 270

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 270

ctgccccccc accccgccac ngg 23

<210> 271

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 271

gcccccaccc accccgcctc ngg 23

<210> 272

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 272

gcttccctcc accccgcatc ngg 23

<210> 273

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 273

tgcccctccc accccgcctc ngg 23

<210> 274

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 274

ctacccctcc accccgcctc ngg 23

<210> 275

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 275

gggcccctcc accccgcctc ngg 23

<210> 276

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 276

attcccccc accccgcctc ngg 23

<210> 277

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 277

gacccccccc accccgcccc ngg 23

<210> 278

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 278

acacccccccc accccgcctc ngg 23

<210> 279

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 279

gccccccacc accccacctc ngg 23

<210> 280

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 280

ctccccaccc accccgcctc ngg 23

<210> 281

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 281

tgccccccccc accccacctc ngg 23

<210> 282

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 282

taccccccac accccgcctc ngg 23

<210> 283

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 283

gtaccccacc accccgcccc ngg 23

<210> 284

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 284

gacaccttcc accccgtctc ngg 23

<210> 285

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 285

cctcccccac accccgcatc ngg 23

<210> 286

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 286

gacccctcac accccgcccc ngg 23

<210> 287

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 287

tccacccccc accccgcccc ngg 23

<210> 288

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 288

atccccctcc accccacccc ngg 23

<210> 289

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 289

tatctagcct cttctaagac ngg 23

<210> 290

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 290

tatctagcct cttctaaaac ngg 23

<210> 291

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 291

tttncctcac tcctgctcgg tgaattt 27

<210> 292

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 292

tttnggaggt cagaaatagg gggtcca 27

<210> 293

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 293

tttngcaggt cagaaatagg gagaggc 27

<210> 294

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 294

tttngatggc gacttcaggc acaggat 27

<210> 295

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 295

tttncttggg tgtgttaaaa gtgacca 27

<210> 296

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 296

tctcccctgg gaagcatccc ngg 23

<210> 297

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 297

catccccagg gaagcatccc ngg 23

<210> 298

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 298

gatcccctgt gaagcatccc ngg 23

<210> 299

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 299

tctccccggg gaaacatccc ngg 23

<210> 300

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 300

tctcccccag gaaccatccc ngg 23

<210> 301

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 301

cttcctctgg gaagcatccc ngg 23

<210> 302

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 302

cctccactgg aaagcatccc ngg 23

<210> 303

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 303

cctcctccgg gaagcacccc ngg 23

<210> 304

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 304

tggcccctgg gaagcacccc ngg 23

<210> 305

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 305

tctcccctgg gatgcatccc nga 23

<210> 306

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 306

actcccctgg gagctatccc ngg 23

<210> 307

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 307

cgtcctctgg gaggcatccc ngg 23

<210> 308

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 308

tctcccctag gaaacatacc ngg 23

<210> 309

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 309

actctcctgg gaagcatccc ngc 23

<210> 310

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 310

gctcccccgg gcggcatccc ngg 23

<210> 311

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 311

cctccccagg gaggcatcct ngg 23

<210> 312

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 312

tctcccctgg gaagcatccc ngg 23

<210> 313

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 313

gagttagagc agaagaagaa ngg 23

<210> 314

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 314

gagtctaagc agaagaagaa nag 23

<210> 315

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 315

gaggccgagc agaagaaaga ngg 23

<210> 316

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 316

gagtcctagc aggagaagaa nag 23

<210> 317

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 317

gagtccggga aggagaagaa ngg 23

<210> 318

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 318

acgtctgagc agaagaagaa ngg 23

<210> 319

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 319

gaatccaagc aggagaagaa nga 23

<210> 320

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 320

gagccggagc agaagaagga ngg 23

<210> 321

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 321

tatctagcct cttctaagac ngg 23

<210> 322

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 322

tatctagcct cttctaaaac ngg 23

<210> 323

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 323

ttttgggagt aagaaaaggt ngg 23

<210> 324

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 324

atttggtagt aagaaaaggt ngg 23

<210> 325

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 325

ggatgggagt gagaaaaggt ngg 23

<210> 326

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 326

agaagggagt aagaaaaagg ngg 23

<210> 327

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 327

gcatgggagt aagagaagga ngg 23

<210> 328

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 328

actggggaga aagaaaaggc ngg 23

<210> 329

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 329

ctgggggact aagaaaaggt ngg 23

<210> 330

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 330

taggaggagt aagaaaaggt ngg 23

<210> 331

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 331

ccgggggcgt aagaaaaggt ngg 23

<210> 332

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 332

agatgggagt aagagaaggt ntg 23

<210> 333

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 333

gttggggagt aagagaagga ngg 23

<210> 334

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (21)..(21)

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

<400> 334

ctaagggagt aagaaaagga ngg 23

<210> 335

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 335

attncccttc agctaaaata aaggagg 27

<210> 336

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 336

attnggctca gcaggcacct gcctcag 27

<210> 337

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 337

attngggact ggagttgctt catgtac 27

<210> 338

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 338

attntctcca tgaaaaatac tggggtc 27

<210> 339

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 339

attnttcatg gagaaaatat tcagaat 27

<210> 340

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 340

attnttcatg aagaaaatat tcaggaa 27

<210> 341

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 341

attngcagct acaggcagag acaaagg 27

<210> 342

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 342

attncctgga aaccatccag gccttgt 27

<210> 343

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 343

attnggtcag ctgttaacat cagtacg 27

<210> 344

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 344

attnccccttc agctaaaata aaggagg 27

<210> 345

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 345

attnggctca gcaggcacct gcctcag 27

<210> 346

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 346

attngggact ggagttgctt catgtac 27

<210> 347

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 347

attntctcca tgaaaaatac tggggtc 27

<210> 348

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 348

attntctcca taaaaaatac tagaaaa 27

<210> 349

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 349

attnttcatg gagaaaatat tcagaat 27

<210> 350

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 350

attnttcatg aagaaaatat tcagaaa 27

<210> 351

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 351

attngcagct acaggcagag acaaagg 27

<210> 352

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 352

attncctgga aaccatccag gccttgt 27

<210> 353

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 353

attnggtcag ctgttaacat cagtacg 27

<210> 354

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 354

attntcttca cggaaacagg gttcctt 27

<210> 355

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 355

attntcttcg cggaagcagg gttccgg 27

<210> 356

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 356

tttntggttg cccaccctag tcattgg 27

<210> 357

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 357

tttngatggc gacttcaggc acaggat 27

<210> 358

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 358

attntgcaga gcaaatacca gagataa 27

<210> 359

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 359

attntgtaca taaaatatca aagtaca 27

<210> 360

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 360

tttncctcac tcctgctcgg tgaattt 27

<210> 361

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 361

attntcttca cggaaacagg gttcctt 27

<210> 362

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 362

attntcttcg cggaagcagg gttccgg 27

<210> 363

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 363

tttntggttg cccaccctag tcattgg 27

<210> 364

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 364

tttngatggc gacttcaggc acaggat 27

<210> 365

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 365

attntgcaga gcaaatacca gagataa 27

<210> 366

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 366

attnctgacc tcccaaacag ctacata 27

<210> 367

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (4)..(4)

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

<400> 367

attntcagag ggacacactg tggcccc 27

<210> 368

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 368

tctctcattc tgcagagcaa ataccagaga taagagagta ggct 44

<210> 369

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 369

tctccatgaa aaatactggg gtc 23

<210> 370

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 370

cctggaaacc atccaggcct tgt 23

<210> 371

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 371

gcagctacag gcagagacaa agg 23

<210> 372

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 372

tgcagagcaa ataccagaga taa 23

<210> 373

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 373

ggtcagctgt taacatcagt acg 23

<210> 374

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 374

tcagagggac acactgtggc ccc 23

<210> 375

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 375

tctctcattc tgcagagcaa ataccagaga taagagagta ggct 44

<210> 376

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 376

tctctcattc tgcagagcaa ataccagaga taagagagta ggct 44

<210> 377

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 377

tggcacattg tcagagggac acactgtggc ccctgtgccc agcc 44

<210> 378

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 378

gctgtcattt cctggaaacc atccaggcct tgtagcctgc cctc 44

<210> 379

<211> 1147

<212> PRT

<213> Alicyclobacillus kakegawensis

<400> 379

Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Ser Glu Cys Pro
1 5 10 15

Asp Ile Leu Ala Gly Met Trp Gln Leu His Arg Ala Thr Asn Ala Gly
20 25 30

Val Arg Tyr Tyr Thr Glu Trp Val Ser Leu Met Arg Gln Glu Ile Leu
35 40 45

Tyr Ser Arg Gly Pro Asp Gly Gly Gln Gln Cys Tyr Met Thr Ala Glu
50 55 60

Asp Cys Gln Arg Glu Leu Leu Arg Arg Leu Arg Asn Arg Gln Leu His
65 70 75 80

Asn Gly Arg Gln Asp Gln Pro Gly Thr Asp Ala Asp Leu Leu Ala Ile
85 90 95

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

100 105 110

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

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

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

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

Glu Thr Phe Leu Thr Arg Tyr Ala Leu Tyr Asn Gly Val Leu Arg Asp
355 360 365

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

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

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

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

595 600 605

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

610 615 620

His Tyr Tyr Pro Ile His Gly Thr Asp Asp Leu Val Ala Val His Glu

625 630 635 640

Arg Ser His Leu Ile Gln Met Pro Gly Glu Thr Glu Thr Lys Gln Leu

645 650 655

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

660 665 670

Gln Leu Ala Leu Leu Arg Leu Leu Val Arg Cys Gly Ala Ala Asp Glu

675 680 685

Arg Ile Arg Thr Arg Ser Trp Gln Arg Leu Thr Lys Gln Gly Arg Glu

690 695 700

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

705 710 715 720

Leu Thr Arg Leu Glu Ala Tyr Cys Gly Arg Val Pro Asp Asp Glu Trp

725 730 735

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

740 745 750

Lys Gln Val Arg Asp Trp Arg Lys Gln Val Lys Ser Gly Ala Lys Val

755 760 765

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

770 775 780

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

785 790 795 800

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

805 810 815

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

820 825 830

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

835 840 845

Val Tyr Glu Ala Ser Gly Pro Arg Glu Gly Gln Trp Thr Ala Gln His
850 855 860
Pro Pro Cys Gln Leu Ile Ile Leu Glu Glu Leu Ser Ala Tyr Arg Phe
865 870 875 880

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

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

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

Thr Gly Gly Lys Pro His Ser Glu Lys Gln Thr Asp Leu Thr Asp
1070 1075 1080
Lys Glu Leu Glu Leu Ile Ala Glu Ala Asp Glu Ala Arg Ala Lys

1085 1090 1095

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

1100 1105 1110

His Trp Ile Arg Gln Arg Glu Phe Trp Ser Leu Val Lys Gln Arg

1115 1120 1125

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

1130 1135 1140

Ser Ser Leu Asp

1145

<210> 380

<211> 1112

<212> PRT

<213> Bacillus species

<400> 380

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

1 5 10 15

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

20 25 30

Asn Glu Gly Ile Ala Tyr Tyr Met Asn Leu Leu Thr Leu Tyr Arg Gln

35 40 45

Glu Ala Ile Gly Asp Lys Thr Lys Glu Ala Tyr Gln Ala Glu Leu Ile

50 55 60

Asn Ile Ile Arg Asn Gln Gln Arg Asn Asn Gly Ser Ser Glu Glu His

65 70 75 80

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

85 90 95

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

100 105 110

Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn Ser Gln Ser Gly Lys

115 120 125

Gly Thr Ser Asn Ala Gly Arg Lys Pro Arg Trp Lys Arg Leu Lys Glu

130 135 140

Glu Gly Asn Pro Asp Trp Glu Leu Glu Lys Lys Lys Asp Glu Glu Arg
145 150 155 160

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

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

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

Ala Tyr Asn Gly Leu Gln Lys Lys Leu Ser Arg Thr Lys Glu Gln Ala
355 360 365
Thr Phe Thr Leu Pro Asp Ala Ile Glu His Pro Leu Trp Ile Arg Tyr
370 375 380
Glu Ser Pro Gly Gly Thr Asn Leu Asn Leu Phe Lys Leu Glu Glu Lys

385 390 395 400

Gln Lys Lys Asn Tyr Tyr Val Thr Leu Ser Lys Ile Ile Trp Pro Ser

405 410 415

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

420 425 430

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

435 440 445

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

450 455 460

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

465 470 475 480

Asn His Lys Glu Leu Leu Gly Glu Gly Asp Ile Gly Pro Val Phe Phe

485 490 495

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

500 505 510

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

515 520 525

Ser Lys Val Ile Asp Tyr Lys Pro Lys Glu Leu Met Asp Trp Met Asn

530 535 540

Thr Gly Ser Ala Ser Asn Ser Phe Gly Val Ala Ser Leu Leu Glu Gly

545 550 555 560

Met Arg Val Met Ser Ile Asp Met Gly Gln Arg Thr Ser Ala Ser Val

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

Gln Gln Arg Gln Glu Arg Arg Lys Lys Arg Gln Phe Val Arg Ser Gln

625 630 635 640

Ile Arg Met Leu Ala Asn Val Leu Arg Leu Glu Thr Lys Lys Thr Pro
645 650 655
Asp Glu Arg Lys Lys Ala Ile His Lys Leu Met Glu Ile Val Gln Ser
660 665 670

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

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

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

Gln Gly Glu Met Phe Gly Leu Gln Val Gly Asp Val Arg Ser Glu Tyr
865 870 875 880
Ser Ser Arg Phe His Ala Lys Thr Gly Ala Pro Gly Ile Arg Cys His

885 890 895

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

900 905 910

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

915 920 925

Lys Gly Asp Ile Ile Pro Ser Gln Gly Gly Glu Leu Phe Val Thr Leu

930 935 940

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

945 950 955 960

His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Gln

965 970 975

Gln Asn Ser Glu Val Tyr Arg Val Pro Cys Gln Leu Ala Arg Met Gly

980 985 990

Glu Asp Lys Leu Tyr Ile Pro Lys Ser Gln Thr Glu Thr Ile Lys Lys

995 1000 1005

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

1010 1015 1020

Val Tyr Lys Trp Glu Lys Ser Glu Lys Met Lys Ile Lys Thr Asp

1025 1030 1035

Thr Thr Phe Asp Leu Gln Asp Leu Asp Gly Phe Glu Asp Ile Ser

1040 1045 1050

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

1055 1060 1065

Met Phe Arg Asp Pro Ser Gly Tyr Phe Phe Asn Asn Glu Thr Trp

1070 1075 1080

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

1085 1090 1095

Ser Cys Leu Lys Lys Lys Ile Leu Ser Asn Lys Val Glu Leu

1100 1105 1110

<210> 381

<211> 1489

<212> PRT

<213> Desulfatirhabdium butyrativorans

<400> 381

Met Pro Leu Ser Asn Asn Pro Pro Val Thr Gln Arg Ala Tyr Thr Leu
1 5 10 15

Arg Leu Arg Gly Ala Asp Pro Ser Asp Leu Ser Trp Arg Glu Ala Leu
20 25 30

Trp His Thr His Glu Ala Val Asn Lys Gly Ala Lys Val Phe Gly Asp
35 40 45

Trp Leu Leu Thr Leu Arg Gly Gly Leu Asp His Thr Leu Ala Asp Thr

50 55 60

Lys Val Lys Gly Gly Lys Gly Lys Pro Asp Arg Asp Pro Thr Pro Glu
65 70 75 80

Glu Arg Lys Ala Arg Arg Ile Leu Leu Ala Leu Ser Trp Leu Ser Val
85 90 95

Glu Ser Lys Leu Gly Ala Pro Ser Ser Tyr Ile Val Ala Ser Gly Asp
100 105 110

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

115 120 125

Ile Leu Gln Ser Arg Lys Val Ala Lys Ser Glu Ile Asp Asp Trp Lys
130 135 140

Arg Asp Cys Ser Ala Ser Leu Ser Ala Ala Ile Arg Asp Asp Ala Val
145 150 155 160

Trp Val Asn Arg Ser Lys Val Phe Asp Glu Ala Val Lys Ser Val Gly
165 170 175

Ser Ser Leu Thr Arg Glu Glu Ala Trp Asp Met Leu Glu Arg Phe Phe

180 185 190

Gly Ser Arg Asp Ala Tyr Leu Thr Pro Met Lys Asp Pro Glu Asp Lys
195 200 205

Ser Ser Glu Thr Glu Gln Glu Asp Lys Ala Lys Asp Leu Val Gln Lys
210 215 220

Ala Gly Gln Trp Leu Ser Ser Arg Tyr Gly Thr Ser Glu Gly Ala Asp

225 230 235 240

Phe Cys Arg Met Ser Asp Ile Tyr Gly Lys Ile Ala Ala Trp Ala Asp

245 250 255

Asn Ala Ser Gln Gly Gly Ser Ser Thr Val Asp Asp Leu Val Ser Glu

260 265 270

Leu Arg Gln His Phe Asp Thr Lys Glu Ser Lys Ala Thr Asn Gly Leu

275 280 285

Asp Trp Ile Ile Gly Leu Ser Ser Tyr Thr Gly His Thr Pro Asn Pro

290 295 300

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

305 310 315 320

Asp Asp Leu Lys Lys Lys Ala Asn Thr Arg Ala Glu Ser Cys Lys Ser

325 330 335

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

340 345 350

Asp Val Glu Ser Val Cys Gly Phe Thr Tyr Arg Val Asp Lys Asp Gly

355 360 365

Gln Pro Val Ser Val Ala Asp Tyr Ser Lys Tyr Asp Val Asp Tyr Lys

370 375 380

Trp Gly Thr Ala Arg His Tyr Ile Phe Ala Val Met Leu Asp His Ala

385 390 395 400

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

405 410 415

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

420 425 430

Ala Arg Ala Arg Glu Trp Leu Asp Ser Phe Cys Lys Glu Arg Ser Val

435 440 445

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

450 455 460

Gly Trp Lys Glu Val Val Ala Ala Trp Ser Lys Ser Asp Cys Lys Ser

465 470 475 480

Thr Glu Asp Arg Ile Ala Ala Ala Arg Ala Leu Gln Asp Asp Ser Glu
485 490 495
Ile Asp Lys Phe Gly Asp Ile Gln Leu Phe Glu Ala Leu Ala Glu Asp

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

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

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

690 695 700
Thr Phe Ser Val Lys Leu Gln Pro Gln Gly Pro Trp Cys Ala Tyr Ala
705 710 715 720
Glu Gln His Gly Leu Asn Thr Asn Pro Gln Tyr Trp Pro His Ala Asp

725 730 735

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

740 745 750

Pro Gly Leu Arg Val Leu Ser Val Asp Leu Gly His Arg Tyr Ala Ala

755 760 765

Ala Cys Ala Val Trp Glu Ala Val Asn Thr Glu Thr Val Lys Glu Ala

770 775 780

Cys Gln Asn Val Gly Arg Asp Met Pro Lys Glu His Asp Leu Tyr Leu

785 790 795 800

His Ile Lys Val Lys Lys Gln Gly Ile Gly Lys Gln Thr Glu Val Asp

805 810 815

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

820 825 830

Arg Pro His Pro Ala Pro Trp Ala Arg Leu Asp Arg Gln Phe Leu Ile

835 840 845

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

850 855 860

Ile Trp Ala Leu His Gln Met Glu Cys Lys Leu Asp Arg Thr Lys Pro

865 870 875 880

Leu Ile Asp Arg Leu Ile Ala Ser Gly Trp Gly Leu Leu Lys Arg Gln

885 890 895

Met Ala Arg Leu Asp Ala Leu Lys Glu Leu Gly Trp Ile Pro Ala Pro

900 905 910

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

915 920 925

Arg Glu Ser Leu Ala Val Asp Asp Leu Met Phe Ser Ala Val Arg Thr

930 935 940

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

945 950 955 960

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

965 970 975

Leu Asp Glu Asn Gly Arg Ile Asp Leu Leu Gln Asp Ala Leu Ala Leu

980 985 990

Trp His Glu Leu Phe Ser Ser Pro Gly Trp Arg Asp Glu Ala Ala Lys

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

Lys Glu Arg Trp Glu Asp Glu Asp Gln Arg Trp Lys Lys Leu Leu

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

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

1115 1120 1125

Thr Arg Leu Arg Pro Asp Gly Thr Arg His Glu Ile Gly Glu Gln

1130 1135 1140

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

1145 1150 1155

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

1160 1165 1170

Ile Gly Ser Glu Gly Gly Lys Gly Trp Asp Gly Gly Lys Arg Pro

1175 1180 1185

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

1190 1195 1200

Val Ile Glu Asn Leu Ala Asn Tyr Arg Pro Asp Glu Thr Arg Thr

1205 1210 1215

Arg Leu Glu Asn Arg Arg Leu Met Thr Trp Ser Ala Ser Lys Val

1220 1225 1230

His Lys Tyr Leu Ser Glu Ala Cys Gln Leu Asn Gly Leu Tyr Leu

1235 1240 1245

Cys Thr Val Ser Ala Trp Tyr Thr Ser Arg Gln Asp Ser Arg Thr

1250 1255 1260

Gly Ala Pro Gly Ile Arg Cys Gln Asp Val Ser Val Arg Glu Phe

1265 1270 1275

Met Gln Ser Pro Phe Trp Arg Lys Gln Val Lys Gln Ala Glu Ala

1280 1285 1290

Lys His Asp Glu Asn Lys Gly Asp Ala Arg Glu Arg Phe Leu Cys

1295 1300 1305

Glu Leu Asn Lys Thr Trp Lys Ala Lys Thr Pro Ala Glu Trp Lys

1310 1315 1320

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

1325 1330 1335

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

1340 1345 1350

Asp Leu Asn Ala Ala Ala Asn Ile Gly Leu Arg Ala Leu Thr Asp

1355 1360 1365

Pro Asp Trp Pro Gly Lys Trp Trp Tyr Val Pro Cys Asp Pro Val

1370 1375 1380

Ser Phe Glu Ser Lys Met Asp Tyr Val Lys Gly Cys Ala Ala Val

1385 1390 1395

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

1400 1405 1410

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

1415 1420 1425

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

1430 1435 1440

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

1445 1450 1455

Ala Tyr Gln Asn Asp Val Gln Tyr Arg Val Ile Arg Met Leu Lys

1460 1465 1470

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

1475 1480 1485

Gly

<210> 382

<211> 1194

<212> PRT

<213> Desulfonatronum thiodismutans

<400> 382

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

1 5 10 15

Trp Thr Thr His Glu His Val Asn Leu Ala Val Ala Glu Val Glu Arg

20 25 30

Val Leu Leu Arg Cys Arg Gly Arg Ser Tyr Trp Thr Leu Asp Arg Arg

35 40 45

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

50 55 60

Ala Met Ala Arg Glu Ala Gln Arg Arg Asn Gly Trp Pro Val Val Gly

65 70 75 80

Glu Asp Glu Glu Ile Leu Leu Ala Leu Arg Tyr Leu Tyr Glu Gln Ile

85 90 95

Val Pro Ser Cys Leu Leu Asp Asp Leu Gly Lys Pro Leu Lys Gly Asp

100 105 110

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

115 120 125

Thr Cys Arg Arg Asp Glu Gly Lys Asp Val Ala Cys Cys Gly Pro Phe

130 135 140

His Glu Val Ala Gly Lys Tyr Leu Gly Ala Leu Pro Glu Trp Ala Thr

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

Asn Lys Ala Tyr Asn Lys Asp Asp Trp Lys Lys Glu Lys Asp Lys Gly

210 215 220

Ile Ser Ser Trp Ala Val Lys Tyr Ile Gln Lys Gln Leu Gln Leu Gly

225 230 235 240

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

245 250 255

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

260 265 270

Leu Trp Asn Arg Leu Ala Val Arg Leu Ala Leu Ala His Leu Leu Ser

275 280 285

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

290 295 300

Ala Lys Arg Asp Glu Leu Ala Ala Leu Phe Leu Gly Met Glu Asp Gly

305 310 315 320

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

325 330 335

Gln His Ala Phe Glu Pro Val Asp Arg Pro Tyr Val Val Ser Gly Arg

340 345 350

Ala Leu Arg Ser Trp Thr Arg Val Arg Glu Glu Trp Leu Arg His Gly

355 360 365

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

370 375 380

Leu Arg Gly Lys Phe Gly Asp Pro Asp Val Phe His Trp Leu Ala Glu

385 390 395 400

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

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

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

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

625 630 635 640
Phe Pro Ala Ala Asp Gly Arg Thr Val Asp Asp Pro Glu Lys Leu Trp

645 650 655

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

660 665 670

Ser Arg Lys Glu Glu Ile Ala Arg Arg Ala Ala Met Glu Glu Leu Arg

675 680 685

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

690 695 700

Val Leu Gln Glu Asp Asp Pro Arg Thr Glu His Leu Arg Leu Phe Met

705 710 715 720

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

725 730 735

Phe Lys Gly Phe Gly Asp Asp Arg Phe Arg Ser Thr Pro Asp Leu Trp

740 745 750

Lys Gln His Cys His Phe Phe His Asp Lys Ala Glu Lys Val Val Ala

755 760 765

Glu Arg Phe Ser Arg Trp Arg Thr Glu Thr Arg Pro Lys Ser Ser Ser

770 775 780

Trp Gln Asp Trp Arg Glu Arg Arg Gly Tyr Ala Gly Gly Lys Ser Tyr

785 790 795 800

Trp Ala Val Thr Tyr Leu Glu Ala Val Arg Gly Leu Ile Leu Arg Trp

805 810 815

Asn Met Arg Gly Arg Thr Tyr Gly Glu Val Asn Arg Gln Asp Lys Lys

820 825 830

Gln Phe Gly Thr Val Ala Ser Ala Leu Leu His His Ile Asn Gln Leu

835 840 845

Lys Glu Asp Arg Ile Lys Thr Gly Ala Asp Met Ile Ile Gln Ala Ala

850 855 860

Arg Gly Phe Val Pro Arg Lys Asn Gly Ala Gly Trp Val Gln Val His

865 870 875 880

Glu Pro Cys Arg Leu Ile Leu Phe Glu Asp Leu Ala Arg Tyr Arg Phe

885 890 895

Arg Thr Asp Arg Ser Arg Arg Glu Asn Ser Arg Leu Met Arg Trp Ser

900 905 910

His Arg Glu Ile Val Asn Glu Val Gly Met Gln Gly Glu Leu Tyr Gly

915 920 925

Leu His Val Asp Thr Thr Glu Ala Gly Phe Ser Ser Arg Tyr Leu Ala

930 935 940

Ser Ser Gly Ala Pro Gly Val Arg Cys Arg His Leu Val Glu Glu Asp

945 950 955 960

Phe His Asp Gly Leu Pro Gly Met His Leu Val Gly Glu Leu Asp Trp

965 970 975

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

980 985 990

Leu Gly Gly Met Val Arg Pro Gly Met Leu Val Pro Trp Asp Gly Gly

995 1000 1005

Glu Leu Phe Ala Thr Leu Asn Ala Ala Ser Gln Leu His Val Ile

1010 1015 1020

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

1025 1030 1035

Gly Arg Cys Gly Glu Ala Ile Arg Ile Val Cys Asn Gln Leu Ser

1040 1045 1050

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

1055 1060 1065

Arg Leu Leu Gly Ala Leu Gln Gln Leu Lys Asn Gly Asp Ala Pro

1070 1075 1080

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

1085 1090 1095

Tyr Val Met Thr Pro Thr Asn Ala Gly Lys Lys Tyr Arg Ala Gly

1100 1105 1110

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

1115 1120 1125

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

1130 1135 1140

Phe Arg Asp Pro Ser Gly Val Phe Phe Ala Pro Asp Arg Trp Leu

1145 1150 1155

Pro Ser Glu Ile Tyr Trp Ser Arg Ile Arg Arg Arg Ile Trp Gln

1160 1165 1170

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

1175 1180 1185

Met Asp Glu Met Pro Tyr

1190

<210> 383

<211> 1326

<212> PRT

<213> Unknown

<220><223> Lentisphaeria bacterium

<400> 383

Met Ala Val Glu Leu Asn Arg Ile Tyr Gln Gly Arg Val Asn His Val

1 5 10 15

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

20 25 30

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

35 40 45

Tyr Leu Val Ala Leu Ala Ala Met Ala Leu Asp Ser Lys Asp Ser Leu

50 55 60

Phe Gly Lys Phe Lys Met Gln Ile Arg Ala Val Trp Asn Asp Phe Tyr

65 70 75 80

Arg Asn Gly Gln Leu Arg Pro Gly Leu Lys His Ser Leu Ile Arg Ser

85 90 95

Leu Gly His Ala Ala Glu Leu Asn Thr Ser Asn Gly Ala Asp Ile Ala

100 105 110

Met Asn Leu Ile Leu Glu Asp Gly Gly Ile Pro Ser Glu Ile Leu Asn

115 120 125

Ala Ala Leu Glu His Leu Ala Glu Lys Cys Thr Gly Asp Val Ser Gln

130 135 140

Leu Gly Lys Thr Phe Phe Pro Arg Phe Cys Asp Thr Ala Tyr His Gly

145 150 155 160

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

165 170 175

Arg Leu Val Asp Ala Leu Tyr Ser Leu His Pro Val Gln Ala Val Gln

180 185 190

Glu Leu Ala Pro Glu Ile Glu Ile Gly Trp Gly Gly Val Lys Thr Gln

195 200 205

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

210 215 220

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

225 230 235 240

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

245 250 255

Gly Lys Ile Asp Thr Phe Pro Glu Ile Ser Phe Gly Arg Ile Ser Ser

260 265 270

His Gln Asn Ile Asn Ile Ser Asn Ala Met Trp Ile Leu Lys Phe Phe

275 280 285

Pro Asp Gln Tyr Ser Val Asp Leu Ile Lys Asn Leu Ile Pro Asn Lys

290 295 300

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

305 310 315 320

Leu Ser Arg Gly Lys Arg Gly Tyr Thr Phe Arg Ala Phe Thr Asp Leu

325 330 335

Ala Met Trp Glu Lys Asn Trp Lys Val Phe Asp Arg Ala Ala Phe Ser

340 345 350

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

355 360 365

Asn Asp Gln Leu Lys Arg Tyr Cys Ala Ala Leu Asn Trp Met Asp Gly

370 375 380

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

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

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

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

625 630 635 640

Pro Lys Gln Ser Phe Lys Thr Thr Ala Val Ile Leu Met Pro Asp Trp

645 650 655

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

660 665 670

Asp Val Ser Ala Ile His Gln Lys Thr Asp His Arg Phe Glu Asn Gln

675 680 685

Phe Tyr Phe Ala Asn Asn Thr Asn Thr Cys Leu Leu Trp Pro Ser Tyr

690 695 700

Gln Tyr Lys Lys Pro Val Thr Trp Tyr Gln Gly Lys Lys Pro Phe Asp

705 710 715 720

Val Val Ala Val Asp Leu Gly Gln Arg Ser Ala Gly Ala Val Ser Arg

725 730 735

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

740 745 750

Ala Gly Gly Thr Gln Trp Tyr Ala Tyr Arg Lys Phe Ser Gly Leu Leu

755 760 765

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

770 775 780

Glu Glu Leu Ser Gly Asn Ala Gly Arg Leu Ser Thr Glu Glu Glu Thr

785 790 795 800

Val Gln Ala Cys Val Leu Cys Lys Met Leu Ile Gly Asp Ala Thr Leu

805 810 815

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

820 825 830

Asp Lys Leu Leu Ile Ala Phe Arg Arg Ala Thr Gly Arg Met Lys Gln

835 840 845

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

850 855 860

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

865 870 875 880

Asp Asn Val Leu Lys Glu Glu Lys Gln His Arg Glu Met Leu Pro Ala

885 890 895

Ile Leu Leu Gln Ile Ala Asp Arg Val Leu Pro Leu Arg Gly Arg Lys

900 905 910

Trp Asp Trp Val Leu Asn Pro Gln Ser Asn Ser Phe Val Leu Gln Gln

915 920 925

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

930 935 940

Gly Leu Ser Phe Ala Arg Ile Glu Gln Leu Glu Ser Leu Arg Met Arg

945 950 955 960

Cys Gln Ala Leu Asn Arg Ile Leu Met Arg Lys Thr Gly Glu Lys Pro

965 970 975

Ala Thr Leu Ala Glu Met Arg Asn Asn Pro Ile Pro Asp Cys Cys Pro

980 985 990

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

995 1000 1005

Thr Ala Asn Leu Ile Leu Ala Gln Ala Leu Gly Leu Arg His Cys

1010 1015 1020

Leu His Ser Glu Ser Ala Thr Lys Arg Lys Glu Asn Gly Met His

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

Ser Tyr Glu Asn Ser Arg Leu Met Lys Trp Ser His Arg Lys Ile

1070 1075 1080

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

1085 1090 1095

Leu Gln Val Gly Ala Ala Tyr Ser Ser Lys Phe Ser Ala Asn Ala

1100 1105 1110

Ile Pro Gly Phe Arg Ala Glu Glu Cys Ser Ile Asp Gln Leu Ser

1115 1120 1125

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

1130 1135 1140

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

1145 1150 1155

Ala Lys Ala Thr Ile Ile Met Pro Arg Asn Gly Gly Pro Val Phe

1160 1165 1170

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

1175 1180 1185

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

1190 1195 1200

Ala Thr Asn Leu Leu Cys Asn Asn Arg Val Ser Cys Asp Arg Lys

1205 1210 1215

Lys Asp Cys Trp Gln Val Lys Arg Ser Ser Asn Phe Ser Lys Met

1220 1225 1230

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

1235 1240 1245

Gln Glu Gly Ala Gly Gly Asn Phe Phe Val Leu Gly Cys Ser Glu

1250 1255 1260

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

1265 1270 1275

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

1280 1285 1290

Leu Trp Arg Asn Glu Ile Leu Lys Leu Glu Arg Cys Cys Lys Ile

1295 1300 1305

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

1310 1315 1320

Asn Glu Leu

1325

<210> 384

<211> 1090

<212> PRT

<213> Laceyella sediminis

<400> 384

Met Ser Ile Arg Ser Phe Lys Leu Lys Ile Lys Thr Lys Ser Gly Val
1 5 10 15

Asn Ala Glu Glu Leu Arg Arg Gly Leu Trp Arg Thr His Gln Leu Ile
20 25 30

Asn Asp Gly Ile Ala Tyr Tyr Met Asn Trp Leu Val Leu Leu Arg Gln
35 40 45

Glu Asp Leu Phe Ile Arg Asn Glu Glu Thr Asn Glu Ile Glu Lys Arg
50 55 60

Ser Lys Glu Glu Ile Gln Gly Glu Leu Leu Glu Arg Val His Lys Gln
65 70 75 80

Gln Gln Arg Asn Gln Trp Ser Gly Glu Val Asp Asp Gln Thr Leu Leu
85 90 95

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

Lys Ser Gly Asn Ala Ser Leu Lys Ala Arg Phe Phe Leu Gly Pro Leu
115 120 125

Val Asp Pro Asn Asn Lys Thr Thr Lys Asp Val Ser Lys Ser Gly Pro
130 135 140

Thr Pro Lys Trp Lys Lys Met Lys Asp Ala Gly Asp Pro Asn Trp Val
145 150 155 160

Gln Glu Tyr Glu Lys Tyr Met Ala Glu Arg Gln Thr Leu Val Arg Leu
165 170 175

Glu Glu Met Gly Leu Ile Pro Leu Phe Pro Met Tyr Thr Asp Glu Val
180 185 190

Gly Asp Ile His Trp Leu Pro Gln Ala Ser Gly Tyr Thr Arg Thr Trp
195 200 205

Asp Arg Asp Met Phe Gln Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu
210 215 220

Ser Trp Asn Arg Arg Val Arg Glu Arg Arg Ala Gln Phe Glu Lys Lys

225 230 235 240

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

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

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

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

Asn Phe Leu Asn Lys Arg Thr Gln Gln Gln Ile Glu Glu Thr Gly Glu
485 490 495

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

Glu Val Lys Asn Gly Arg Leu Gln Asn Gly Leu Gly Lys Ala Leu Thr
515 520 525

Val Leu Thr His Pro Asp Gly Thr Lys Ile Val Thr Gly Trp Lys Ala
530 535 540

Glu Gln Leu Glu Lys Trp Val Gly Glu Ser Gly Arg Val Ser Ser Leu
545 550 555 560

Gly Leu Asp Ser Leu Ser Glu Gly Leu Arg Val Met Ser Ile Asp Leu
565 570 575

Gly Gln Arg Thr Ser Ala Thr Val Ser Val Phe Glu Ile Thr Lys Glu
580 585 590

Ala Pro Asp Asn Pro Tyr Lys Phe Phe Tyr Gln Leu Glu Gly Thr Glu
595 600 605

Leu Phe Ala Val His Gln Arg Ser Phe Leu Leu Ala Leu Pro Gly Glu
610 615 620

Asn Pro Pro Gln Lys Ile Lys Gln Met Arg Glu Ile Arg Trp Lys Glu
625 630 635 640

Arg Asn Arg Ile Lys Gln Gln Val Asp Gln Leu Ser Ala Ile Leu Arg
645 650 655

Leu His Lys Lys Val Asn Glu Asp Glu Arg Ile Gln Ala Ile Asp Lys
660 665 670

Leu Leu Gln Lys Val Ala Ser Trp Gln Leu Asn Glu Glu Ile Ala Thr
675 680 685

Ala Trp Asn Gln Ala Leu Ser Gln Leu Tyr Ser Lys Ala Lys Glu Asn
690 695 700

Asp Leu Gln Trp Asn Gln Ala Ile Lys Asn Ala His His Gln Leu Glu
705 710 715 720

Pro Val Val Gly Lys Gln Ile Ser Leu Trp Arg Lys Asp Leu Ser Thr

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

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

785 790 795 800

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

805 810 815

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

820 825 830

Glu Asn Leu Arg Ser Tyr Arg Phe Ser Tyr Glu Arg Ser Arg Arg Glu

835 840 845

Asn Lys Lys Leu Met Glu Trp Ser His Arg Ser Ile Pro Lys Leu Val

850 855 860

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

865 870 875 880

Ala Tyr Ser Ser Arg Tyr His Gly Arg Thr Gly Ala Pro Gly Ile Arg

885 890 895

Cys His Ala Leu Thr Glu Ala Asp Leu Arg Asn Glu Thr Asn Ile Ile

900 905 910

His Glu Leu Ile Glu Ala Gly Phe Ile Lys Glu Glu His Arg Pro Tyr

915 920 925

Leu Gln Gln Gly Asp Leu Val Pro Trp Ser Gly Gly Glu Leu Phe Ala

930 935 940

Thr Leu Gln Lys Pro Tyr Asp Asn Pro Arg Ile Leu Thr Leu His Ala

945 950 955 960

Asp Ile Asn Ala Ala Gln Asn Ile Gln Lys Arg Phe Trp His Pro Ser

965 970 975

Met Trp Phe Arg Val Asn Cys Glu Ser Val Met Glu Gly Glu Ile Val
980 985 990
Thr Tyr Val Pro Lys Asn Lys Thr Val His Lys Lys Gln Gly Lys Thr
995 1000 1005

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

Trp Gly Lys Val Gln Ser Met Ile Gln Ala Tyr Met Lys Lys Thr
1070 1075 1080
Ile Val Gln Arg Met Glu Glu
1085 1090

<210> 385
<211> 1041
<212> PRT
<213> Methylobacterium nodulans
<400> 385
Met Tyr Glu Ala Ile Val Leu Ala Asp Asp Ala Asn Ala Gln Leu Ala
1 5 10 15
Asn Ala Phe Leu Gly Pro Leu Thr Asp Pro Asn Ser Ala Gly Phe Leu
20 25 30

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

85 90 95

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

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

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

Gly His Ala Asp Ala Val Gly Val Leu Ala Arg Val Asn Ala Met Glu
290 295 300
Arg Leu Val Glu Arg Ser Arg Asp Thr Ala Leu Met Thr Leu Pro Asp
305 310 315 320
Pro Val Ala His Pro Arg Ser Ala Gln Trp Glu Ala Glu Gly Gly Ser
325 330 335

Asn Leu Arg Asn Tyr Gln Leu Glu Ala Val Gly Gly Glu Leu Gln Ile
340 345 350

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

Pro Leu Ser Phe Ser Leu Ala Pro Ser Asp Gln Leu Gln Gly Val Val
370 375 380

Leu Thr Lys Gln Asp Lys Gln Gln Lys Ile Thr Tyr Cys Thr Asn Met
385 390 395 400

Asn Glu Val Phe Glu Ala Lys Leu Gly Ser Ala Asp Leu Leu Leu Asn
405 410 415

Trp Asp His Leu Arg Gly Arg Ile Arg Asp Arg Val Asp Ala Gly Asp
420 425 430

Ile Gly Ser Ala Phe Leu Lys Leu Ala Leu Asp Val Ala His Val Leu
435 440 445

Pro Asp Gly Val Asp Asp Gln Leu Ala Arg Ala Ala Phe His Phe Gln
450 455 460

Ser Ala Lys Gly Ala Lys Ser Lys His Ala Asp Ser Val Gln Ala Gly
465 470 475 480

Leu Arg Val Leu Ser Ile Asp Leu Gly Val Arg Ser Phe Ala Thr Cys
485 490 495

Ser Val Phe Glu Leu Lys Asp Thr Ala Pro Thr Thr Gly Val Ala Phe
500 505 510

Pro Leu Ala Glu Phe Arg Leu Trp Ala Val His Glu Arg Ser Phe Thr
515 520 525

Leu Glu Leu Pro Gly Glu Asn Val Gly Ala Ala Gly Gln Gln Trp Arg
530 535 540

Ala Gln Ala Asp Ala Glu Leu Arg Gln Leu Arg Gly Gly Leu Asn Arg
545 550 555 560

His Arg Gln Leu Leu Arg Ala Ala Thr Val Gln Lys Gly Glu Arg Asp
565 570 575

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

580 585 590

Pro Phe Glu Ala Ser Leu Leu Ser Glu Leu Glu Arg Cys Ser Thr Val

595 600 605

Ala Asp Pro Leu Trp Gln Asp Thr Cys Lys Arg Ala Ala Arg Leu Tyr

610 615 620

Arg Thr Glu Phe Gly Ala Val Val Ser Glu Trp Arg Ser Arg Thr Arg

625 630 635 640

Ser Arg Glu Asp Arg Lys Tyr Ala Gly Lys Ser Met Trp Ser Val Gln

645 650 655

His Leu Thr Asp Val Arg Arg Phe Leu Gln Ser Trp Ser Leu Ala Gly

660 665 670

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

675 680 685

Phe Ala Lys Asp Leu Leu Asp His Ile Asp Ala Leu Lys Asp Asp Arg

690 695 700

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

705 710 715 720

Arg Asn Glu Phe Gly Tyr Trp Val Gln Lys His Ala Pro Cys His Val

725 730 735

Ile Leu Phe Glu Asp Leu Ser Arg Tyr Arg Met Arg Thr Asp Arg Pro

740 745 750

Arg Arg Glu Asn Ser Gln Leu Met Gln Trp Ala His Arg Gly Val Pro

755 760 765

Asp Met Val Gly Met Gln Gly Glu Ile Tyr Gly Ile Gln Asp Arg Arg

770 775 780

Asp Pro Asp Ser Ala Arg Lys His Ala Arg Gln Pro Leu Ala Ala Phe

785 790 795 800

Cys Leu Asp Thr Pro Ala Ala Phe Ser Ser Arg Tyr His Ala Ser Thr

805 810 815

Met Thr Pro Gly Ile Arg Cys His Pro Leu Arg Lys Arg Glu Phe Glu

820 825 830

Asp Gln Gly Phe Leu Glu Leu Leu Lys Arg Glu Asn Glu Gly Leu Asp
835 840 845

Leu Asn Gly Tyr Lys Pro Gly Asp Leu Val Pro Leu Pro Gly Gly Glu
850 855 860

Val Phe Val Cys Leu Asn Ala Asn Gly Leu Ser Arg Ile His Ala Asp
865 870 875 880

Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg Phe Trp Thr Gln His Gly
885 890 895

Asp Ala Phe Arg Leu Pro Cys Gly Lys Ser Ala Val Gln Gly Gln Ile
900 905 910

Arg Trp Ala Pro Leu Ser Met Gly Lys Arg Gln Ala Gly Ala Leu Gly
915 920 925

Gly Phe Gly Tyr Leu Glu Pro Thr Gly His Asp Ser Gly Ser Cys Gln
930 935 940

Trp Arg Lys Thr Thr Glu Ala Glu Trp Arg Arg Leu Ser Gly Ala Gln
945 950 955 960

Lys Asp Arg Asp Glu Ala Ala Ala Ala Glu Asp Glu Glu Leu Gln Gly
965 970 975

Leu Glu Glu Glu Leu Leu Glu Arg Ser Gly Glu Arg Val Val Phe Phe
980 985 990

Arg Asp Pro Ser Gly Val Val Leu Pro Thr Asp Leu Trp Phe Pro Ser
995 1000 1005

Ala Ala Phe Trp Ser Ile Val Arg Ala Lys Thr Val Gly Arg Leu
1010 1015 1020

Arg Ser His Leu Asp Ala Gln Ala Glu Ala Ser Tyr Ala Val Ala
1025 1030 1035

Ala Gly Leu
1040

<210> 386

<211> 1388

<212> PRT

<213> Unknown

<220><223> Opitutaceae bacterium

<400> 386

Met Ser Leu Asn Arg Ile Tyr Gln Gly Arg Val Ala Ala Val Glu Thr
1 5 10 15
Gly Thr Ala Leu Ala Lys Gly Asn Val Glu Trp Met Pro Ala Ala Gly
20 25 30
Gly Asp Glu Val Leu Trp Gln His His Glu Leu Phe Gln Ala Ala Ile
35 40 45
Asn Tyr Tyr Leu Val Ala Leu Leu Ala Leu Ala Asp Lys Asn Asn Pro
50 55 60

Val Leu Gly Pro Leu Ile Ser Gln Met Asp Asn Pro Gln Ser Pro Tyr
65 70 75 80
His Val Trp Gly Ser Phe Arg Arg Gln Gly Arg Gln Arg Thr Gly Leu
85 90 95
Ser Gln Ala Val Ala Pro Tyr Ile Thr Pro Gly Asn Asn Ala Pro Thr
100 105 110
Leu Asp Glu Val Phe Arg Ser Ile Leu Ala Gly Asn Pro Thr Asp Arg
115 120 125

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

Ile Thr His Asp Ser Pro Ala Leu Gly Ser Phe Asp Thr Tyr Ser Ile
195 200 205
Ala Thr Pro Asp Thr Arg Thr Pro Gln Leu Thr Gly Pro Lys Ala Arg
210 215 220
Ala Arg Leu Glu Gln Ala Ile Thr Leu Trp Arg Val Arg Leu Pro Glu

225 230 235 240

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

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Pro Pro Lys Asn Ala Glu Thr Pro Pro Pro Ala Gly Val Pro Leu Pro

305 310 315 320

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

325 330 335

Phe Val Phe Arg Ala Phe Thr Ser Leu Pro Cys Trp His Gly Gly Asp

340 345 350

Asn Ile His Pro Thr Trp Lys Ser Phe Asp Ile Ala Ala Phe Lys Tyr

355 360 365

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

370 375 380

Lys Glu Cys Ala Glu Leu Glu Thr Asp Phe Asp Tyr Met His Gly Arg

385 390 395 400

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

405 410 415

Pro Ile Leu Ala Asn Asp Leu Arg Ile Pro Leu Leu Arg Glu Leu Leu

420 425 430

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

435 440 445

Tyr Gly Leu Gln Arg Arg Thr Ile Arg Gly Phe Arg Glu Leu Arg Arg

450 455 460

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

465 470 475 480

Lys Glu Lys Leu Ala Gly Glu Leu Arg Gln Phe Gln Thr Asp Asn Ser
485 490 495
Thr Thr Ile Gly Ser Val Gln Leu Phe Asn Glu Leu Ile Gln Asn Pro
500 505 510

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

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

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

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

725 730 735

Thr Ala Lys Gly Lys Thr His Ile Pro Trp His Gln Asp Arg Asp His

740 745 750

Phe Thr Val Leu Gly Val Asp Leu Gly Thr Arg Asp Ala Gly Ala Leu

755 760 765

Ala Leu Leu Asn Val Thr Ala Gln Lys Pro Ala Lys Pro Val His Arg

770 775 780

Ile Ile Gly Glu Ala Asp Gly Arg Thr Trp Tyr Ala Ser Leu Ala Asp

785 790 795 800

Ala Arg Met Ile Arg Leu Pro Gly Glu Asp Ala Arg Leu Phe Val Arg

805 810 815

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

820 825 830

Leu Leu Glu Trp Glu Asp Ala Arg Asn Ile Ile Leu Arg Leu Gly Gln

835 840 845

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

850 855 860

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

865 870 875 880

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

885 890 895

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

900 905 910

Pro Ser Pro Leu Pro Pro Leu Ala Arg Asp Asp Ala Ile Lys Ser Thr

915 920 925

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

930 935 940

Gln Ile Ala Asn Leu Ile Leu Pro Leu Arg Gly Arg Arg Trp Glu Trp

945 950 955 960

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

965 970 975

Pro Gly Thr Asp Asp Thr Lys Arg Leu Val Ala Gly Gln Arg Gly Ile
980 985 990

Ser His Glu Arg Ile Glu Gln Ile Glu Glu Leu Arg Arg Arg Cys Gln
995 1000 1005

Ser Leu Asn Arg Ala Leu Arg His Lys Pro Gly Glu Arg Pro Val
1010 1015 1020

Leu Gly Arg Pro Ala Lys Gly Glu Glu Ile Ala Asp Pro Cys Pro
1025 1030 1035

Ala Leu Leu Glu Lys Ile Asn Arg Leu Arg Asp Gln Arg Val Asp
1040 1045 1050

Gln Thr Ala His Ala Ile Leu Ala Ala Ala Leu Gly Val Arg Leu
1055 1060 1065

Arg Ala Pro Ser Lys Asp Arg Ala Glu Arg Arg His Arg Asp Ile
1070 1075 1080

His Gly Glu Tyr Glu Arg Phe Arg Ala Pro Ala Asp Phe Val Val
1085 1090 1095

Ile Glu Asn Leu Ser Arg Tyr Leu Ser Ser Gln Asp Arg Ala Arg
1100 1105 1110

Ser Glu Asn Thr Arg Leu Met Gln Trp Cys His Arg Gln Ile Val
1115 1120 1125

Gln Lys Leu Arg Gln Leu Cys Glu Thr Tyr Gly Ile Pro Val Leu
1130 1135 1140

Ala Val Pro Ala Ala Tyr Ser Ser Arg Phe Ser Ser Arg Asp Gly
1145 1150 1155

Ser Ala Gly Phe Arg Ala Val His Leu Thr Pro Asp His Arg His
1160 1165 1170

Arg Met Pro Trp Ser Arg Ile Leu Ala Arg Leu Lys Ala His Glu
1175 1180 1185

Glu Asp Gly Lys Arg Leu Glu Lys Thr Val Leu Asp Glu Ala Arg
1190 1195 1200

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

1205 1210 1215

His Val Pro Gly Lys Pro Trp Arg Thr Leu Leu Ala Pro Leu Pro

1220 1225 1230

Gly Gly Pro Val Phe Val Pro Leu Gly Asp Ala Thr Pro Met Gln

1235 1240 1245

Ala Asp Leu Asn Ala Ala Ile Asn Ile Ala Leu Arg Gly Ile Ala

1250 1255 1260

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

1265 1270 1275

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

1280 1285 1290

Ala Arg Trp Pro Gly Gly Ala Pro Ala Val Thr Leu Ser Thr Pro

1295 1300 1305

Asn Asn Gly Ala Ser Pro Glu Asp Ser Asp Ala Leu Pro Glu Arg

1310 1315 1320

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

1325 1330 1335

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

1340 1345 1350

Gly Leu Trp Ala Ser Val Lys Gln Arg Ala Trp Asn Arg Val Ala

1355 1360 1365

Arg Leu Asn Glu Thr Val Thr Asp Asn Asn Arg Asn Glu Glu Glu

1370 1375 1380

Asp Asp Ile Pro Met

1385

<210> 387

<211> 1108

<212> PRT

<213> Bacillus species

<400> 387

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

1 5 10 15

Glu Ala Gln Asn Leu Arg Lys Gly Ile Trp Arg Thr His Arg Leu Leu
20 25 30
Asn Glu Gly Val Ala Tyr Tyr Met Lys Met Leu Leu Leu Phe Arg Gln
35 40 45

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

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

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

Met Ala Gln Leu Asn Glu Gln Leu Glu Gly Gly Gln Glu Trp Ile Ser
245 250 255
Leu Leu Glu Gln Tyr Glu Glu Asn Arg Glu Arg Glu Leu Arg Glu Asn

260 265 270

Met Thr Ala Ala Asn Asp Lys Tyr Arg Ile Thr Lys Arg Gln Met Lys

275 280 285

Gly Trp Asn Glu Leu Tyr Glu Leu Trp Ser Thr Phe Pro Ala Ser Ala

290 295 300

Ser His Glu Gln Tyr Lys Glu Ala Leu Lys Arg Val Gln Gln Arg Leu

305 310 315 320

Arg Gly Arg Phe Gly Asp Ala His Phe Phe Gln Tyr Leu Met Glu Glu

325 330 335

Lys Asn Arg Leu Ile Trp Lys Gly Asn Pro Gln Arg Ile His Tyr Phe

340 345 350

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

355 360 365

Ala Thr Met Thr Leu Pro Asn Ala Arg Lys His Pro Leu Trp Val Arg

370 375 380

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

385 390 395 400

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

405 410 415

Trp Pro Ser Glu Ser Gly Trp Met Glu Lys Lys Asp Val Glu Val Glu

420 425 430

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

435 440 445

Asp Lys Gly Lys Gln Lys Ile Glu Phe Lys Asp Lys Gly Ser Gly Ser

450 455 460

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

465 470 475 480

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

485 490 495

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

500 505 510

Asn Gly Arg Val Gln Ala Pro Tyr Gly Gln Val Leu Gln Leu Ile Arg
515 520 525

Arg Pro Asn Glu Phe Pro Lys Val Thr Thr Tyr Lys Ser Glu Gln Leu
530 535 540

Val Glu Trp Ile Lys Ala Ser Pro Gln His Ser Ala Gly Val Glu Ser
545 550 555 560

Leu Ala Ser Gly Phe Arg Val Met Ser Ile Asp Leu Gly Leu Arg Ala
565 570 575

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

Ala Ala Asp Phe Ser Tyr Trp Ile Glu Gly Thr Pro Leu Val Ala Val
595 600 605

His Gln Arg Ser Tyr Met Leu Arg Leu Pro Gly Glu Gln Val Glu Lys
610 615 620

Gln Val Met Glu Lys Arg Asp Glu Arg Phe Gln Leu His Gln Arg Val
625 630 635 640

Lys Phe Gln Ile Arg Val Leu Ala Gln Ile Met Arg Met Ala Asn Lys
645 650 655

Gln Tyr Gly Asp Arg Trp Asp Glu Leu Asp Ser Leu Lys Gln Ala Val
660 665 670

Glu Gln Lys Lys Ser Pro Leu Asp Gln Thr Asp Arg Thr Phe Trp Glu
675 680 685

Gly Ile Val Cys Asp Leu Thr Lys Val Leu Pro Arg Asn Glu Ala Asp
690 695 700

Trp Glu Gln Ala Val Val Gln Ile His Arg Lys Ala Glu Glu Tyr Val
705 710 715 720

Gly Lys Ala Val Gln Ala Trp Arg Lys Arg Phe Ala Ala Asp Glu Arg
725 730 735

Lys Gly Ile Ala Gly Leu Ser Met Trp Asn Ile Glu Glu Leu Glu Gly
740 745 750

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

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

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

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

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

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

Arg Asp Pro Ser Gly Val Phe Phe Pro Glu Ser Val Trp Tyr Pro
1070 1075 1080
Gln Lys Asp Phe Trp Gly Glu Val Lys Arg Lys Leu Tyr Gly Lys
1085 1090 1095
Leu Arg Glu Arg Phe Leu Thr Lys Ala Arg
1100 1105

<210> 388
<211> 658
<212> PRT
<213> Methylobacterium nodulans
<400> 388

Met Leu Thr Lys Gln Asp Lys Gln Gln Lys Ile Thr Tyr Cys Thr Asn
1 5 10 15

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

Gln Ser Ala Lys Gly Ala Lys Ser Lys His Ala Asp Ser Val Gln Ala
85 90 95
Gly Leu Arg Val Leu Ser Ile Asp Leu Gly Val Arg Ser Phe Ala Thr

100 105 110

Cys Ser Val Phe Glu Leu Lys Asp Thr Ala Pro Thr Thr Gly Val Ala

115 120 125

Phe Pro Leu Ala Glu Phe Arg Leu Trp Ala Val His Glu Arg Ser Phe

130 135 140

Thr Leu Glu Leu Pro Gly Glu Asn Val Gly Ala Ala Gly Gln Gln Trp

145 150 155 160

Arg Ala Gln Ala Asp Ala Glu Leu Arg Gln Leu Arg Gly Gly Leu Asn

165 170 175

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

180 185 190

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

195 200 205

Trp Pro Phe Glu Ala Ser Leu Leu Ser Glu Leu Glu Arg Cys Ser Thr

210 215 220

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

225 230 235 240

Tyr Arg Thr Glu Phe Gly Ala Val Val Ser Glu Trp Arg Ser Arg Thr

245 250 255

Arg Ser Arg Glu Asp Arg Lys Tyr Ala Gly Lys Ser Met Trp Ser Val

260 265 270

Gln His Leu Thr Asp Val Arg Arg Phe Leu Gln Ser Trp Ser Leu Ala

275 280 285

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

290 295 300

Val Phe Ala Lys Asp Leu Leu Asp His Ile Asp Ala Leu Lys Asp Asp

305 310 315 320

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

325 330 335

Gln Arg Asn Glu Phe Gly Tyr Trp Val Gln Lys His Ala Pro Cys His

340 345 350

Val Ile Leu Phe Glu Asp Leu Ser Arg Tyr Arg Met Arg Thr Asp Arg
355 360 365

Pro Arg Arg Glu Asn Ser Gln Leu Met Gln Trp Ala His Arg Gly Val
370 375 380

Pro Asp Met Val Gly Met Gln Gly Glu Ile Tyr Gly Ile Gln Asp Arg
385 390 395 400

Arg Asp Pro Asp Ser Ala Arg Lys His Ala Arg Gln Pro Leu Ala Ala
405 410 415

Phe Cys Leu Asp Thr Pro Ala Ala Phe Ser Ser Arg Tyr His Ala Ser
420 425 430

Thr Met Thr Pro Gly Ile Arg Cys His Pro Leu Arg Lys Arg Glu Phe
435 440 445

Glu Asp Gln Gly Phe Leu Glu Leu Leu Lys Arg Glu Asn Glu Gly Leu
450 455 460

Asp Leu Asn Gly Tyr Lys Pro Gly Asp Leu Val Pro Leu Pro Gly Gly
465 470 475 480

Glu Val Phe Val Cys Leu Asn Ala Asn Gly Leu Ser Arg Ile His Ala
485 490 495

Asp Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg Phe Trp Thr Gln His
500 505 510

Gly Asp Ala Phe Arg Leu Pro Cys Gly Lys Ser Ala Val Gln Gly Gln
515 520 525

Ile Arg Trp Ala Pro Leu Ser Met Gly Lys Arg Gln Ala Gly Ala Leu
530 535 540

Gly Gly Phe Gly Tyr Leu Glu Pro Thr Gly His Asp Ser Gly Ser Cys
545 550 555 560

Gln Trp Arg Lys Thr Thr Glu Ala Glu Trp Arg Arg Leu Ser Gly Ala
565 570 575

Gln Lys Asp Arg Asp Glu Ala Ala Ala Ala Glu Asp Glu Glu Leu Gln
580 585 590

Gly Leu Glu Glu Glu Leu Leu Glu Arg Ser Gly Glu Arg Val Val Phe

595 600 605

Phe Arg Asp Pro Ser Gly Val Val Leu Pro Thr Asp Leu Trp Phe Pro

610 615 620

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

625 630 635 640

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

645 650 655

Gly Leu

<210> 389

<211> 1147

<212> PRT

<213> Alicyclobacillus kakegawensis

<400> 389

Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Ser Glu Cys Pro

1 5 10 15

Asp Ile Leu Ala Gly Met Trp Gln Leu His Arg Ala Thr Asn Ala Gly

20 25 30

Val Arg Tyr Tyr Thr Glu Trp Val Ser Leu Met Arg Gln Glu Ile Leu

35 40 45

Tyr Ser Arg Gly Pro Asp Gly Gly Gln Gln Cys Tyr Met Thr Ala Glu

50 55 60

Asp Cys Gln Arg Glu Leu Leu Arg Arg Leu Arg Asn Arg Gln Leu His

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

Pro Asn Ser Lys Gly Gly Arg Gly Glu Ala Lys Ser Gly Arg Lys Pro

130 135 140

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

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

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

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

370 375 380
Val Asn Pro Ile Trp Thr Arg Phe Glu Ser Ser Gln Gly Ser Asn Leu

385 390 395 400

His Lys Tyr Glu Phe Leu Phe Asp His Leu Gly Pro Gly Arg His Ala

405 410 415

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

420 425 430

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

435 440 445

Leu Val Leu Arg Glu Glu Glu Lys Ser Ser Val Ala Leu His Leu His

450 455 460

Asp Thr Ala Arg Pro Asp Gly Phe Met Ala Glu Trp Ala Gly Ala Lys

465 470 475 480

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

485 490 495

Gln Gly Met Arg Ser Trp Arg Arg Gln Pro Ser Met Leu Met Ser Ala

500 505 510

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

515 520 525

Ile Ser Val Arg Val Lys Ser Pro Ser Glu Val Arg Gly Gln Arg Arg

530 535 540

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

545 550 555 560

Thr Val Asn Tyr Asn Lys Leu Ser Ala Tyr Leu Glu Glu His Pro Asp

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

His Tyr Tyr Pro Ile His Gly Thr Asp Asp Leu Val Ala Val His Glu

625 630 635 640

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

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

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

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

885 890 895

His Arg Gly Ile Leu Glu Glu Leu Val Asn Gln Ala Gln Val His Asp

900 905 910

Val Leu Val Gly Thr Val Tyr Ala Ala Phe Ser Ser Arg Phe Asp Ala

915 920 925

Arg Thr Gly Ala Pro Gly Val Arg Cys Arg Arg Val Pro Ala Arg Phe

930 935 940

Val Gly Ala Thr Val Asp Asp Ser Leu Pro Leu Trp Leu Thr Glu Phe

945 950 955 960

Leu Asp Lys His Arg Leu Asp Lys Asn Leu Leu Arg Pro Asp Asp Val

965 970 975

Ile Pro Thr Gly Glu Gly Glu Phe Leu Val Ser Pro Cys Gly Glu Glu

980 985 990

Ala Ala Arg Val Arg Gln Val His Ala Asp Ile Asn Ala Ala Gln Asn

995 1000 1005

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

1010 1015 1020

Leu Arg Cys Asp Val Lys Met Gly Gly Glu Gly Thr Val Leu Val

1025 1030 1035

Pro Arg Val Asn Asn Ala Arg Ala Lys Gln Leu Phe Gly Lys Lys

1040 1045 1050

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

1055 1060 1065

Thr Gly Gly Lys Pro His Ser Glu Lys Gln Thr Asp Leu Thr Asp

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

His Trp Ile Arg Gln Arg Glu Phe Trp Ser Leu Val Lys Gln Arg

1115 1120 1125

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

1130 1135 1140

Ser Ser Leu Asp

1145

<210> 390

<211> 1112

<212> PRT

<213> Bacillus species

<400> 390

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

1 5 10 15

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

20 25 30

Asn Glu Gly Ile Ala Tyr Tyr Met Asn Leu Leu Thr Leu Tyr Arg Gln

35 40 45

Glu Ala Ile Gly Asp Lys Thr Lys Glu Ala Tyr Gln Ala Glu Leu Ile

50 55 60

Asn Ile Ile Arg Asn Gln Gln Arg Asn Asn Gly Ser Ser Glu Glu His

65 70 75 80

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

85 90 95

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

100 105 110

Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn Ser Gln Ser Gly Lys

115 120 125

Gly Thr Ser Asn Ala Gly Arg Lys Pro Arg Trp Lys Arg Leu Lys Glu

130 135 140

Glu Gly Asn Pro Asp Trp Glu Leu Glu Lys Lys Lys Asp Glu Glu Arg

145 150 155 160

Lys Ala Lys Asp Pro Thr Val Lys Ile Phe Asp Asn Leu Asn Lys Tyr

165 170 175

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

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

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

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

420 425 430

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

435 440 445

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

450 455 460

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

465 470 475 480

Asn His Lys Glu Leu Leu Gly Glu Gly Asp Ile Gly Pro Val Phe Phe

485 490 495

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

500 505 510

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

515 520 525

Ser Lys Val Ile Asp Tyr Lys Pro Lys Glu Leu Met Asp Trp Met Asn

530 535 540

Thr Gly Ser Ala Ser Asn Ser Phe Gly Val Ala Ser Leu Leu Glu Gly

545 550 555 560

Met Arg Val Met Ser Ile Asp Met Gly Gln Arg Thr Ser Ala Ser Val

565 570 575

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

580 585 590

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

595 600 605

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

610 615 620

Gln Gln Arg Gln Glu Arg Arg Lys Lys Arg Gln Phe Val Arg Ser Gln

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

Asn Leu Leu Thr Asn Met Ala Ala Phe Asn Asp Glu Ile Trp Lys Glu

690 695 700

Ser Leu Val Glu Leu His His Arg Ile Glu Pro Tyr Val Gly Gln Ile

705 710 715 720

Val Ser Lys Trp Arg Lys Gly Leu Ser Glu Gly Arg Lys Asn Leu Ala

725 730 735

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

740 745 750

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

755 760 765

Ile Glu Thr Asp Glu Pro Phe Gly Ser Ser Leu Leu Gln His Ile Gln

770 775 780

Asn Val Lys Asp Asp Arg Leu Lys Gln Met Ala Asn Leu Ile Ile Met

785 790 795 800

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

805 810 815

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

820 825 830

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

835 840 845

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

850 855 860

Gln Gly Glu Met Phe Gly Leu Gln Val Gly Asp Val Arg Ser Glu Tyr

865 870 875 880

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

885 890 895

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

900 905 910

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

915 920 925

Lys Gly Asp Ile Ile Pro Ser Gln Gly Gly Glu Leu Phe Val Thr Leu

930 935 940

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

945 950 955 960

His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Gln

965 970 975

Gln Asn Ser Glu Val Tyr Arg Val Pro Cys Gln Leu Ala Arg Met Gly

980 985 990

Glu Asp Lys Leu Tyr Ile Pro Lys Ser Gln Thr Glu Thr Ile Lys Lys

995 1000 1005

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

1010 1015 1020

Val Tyr Lys Trp Glu Lys Ser Glu Lys Met Lys Ile Lys Thr Asp

1025 1030 1035

Thr Thr Phe Asp Leu Gln Asp Leu Asp Gly Phe Glu Asp Ile Ser

1040 1045 1050

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

1055 1060 1065

Met Phe Arg Asp Pro Ser Gly Tyr Phe Phe Asn Asn Glu Thr Trp

1070 1075 1080

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

1085 1090 1095

Ser Cys Leu Lys Lys Lys Ile Leu Ser Asn Lys Val Glu Leu

1100 1105 1110

<210> 391

<211> 1326

<212> PRT

<213> Unknown

<220>

<223> Lentisphaeria bacterium

<400> 391

Met Ala Val Glu Leu Asn Arg Ile Tyr Gln Gly Arg Val Asn His Val
1 5 10 15
Tyr Ile Phe Asp Glu Asn Gln Asn Gln Val Ser Val Asp Asn Gly Asp
20 25 30
Asp Leu Leu Phe Val His His Glu Leu Tyr Gln Asp Ala Ile Asn Tyr
35 40 45
Tyr Leu Val Ala Leu Ala Ala Met Ala Leu Asp Ser Lys Asp Ser Leu
50 55 60

Phe Gly Lys Phe Lys Met Gln Ile Arg Ala Val Trp Asn Asp Phe Tyr
65 70 75 80
Arg Asn Gly Gln Leu Arg Pro Gly Leu Lys His Ser Leu Ile Arg Ser
85 90 95
Leu Gly His Ala Ala Glu Leu Asn Thr Ser Asn Gly Ala Asp Ile Ala
100 105 110
Met Asn Leu Ile Leu Glu Asp Gly Gly Ile Pro Ser Glu Ile Leu Asn
115 120 125

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

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

245 250 255

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

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

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

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

Ser Val Ser Leu Phe Leu Lys Leu Ala Glu Pro Lys Tyr Phe Cys Val
500 505 510

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

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

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

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

740 745 750

Ala Gly Gly Thr Gln Trp Tyr Ala Tyr Arg Lys Phe Ser Gly Leu Leu

755 760 765

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

770 775 780

Glu Glu Leu Ser Gly Asn Ala Gly Arg Leu Ser Thr Glu Glu Glu Thr

785 790 795 800

Val Gln Ala Cys Val Leu Cys Lys Met Leu Ile Gly Asp Ala Thr Leu

805 810 815

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

820 825 830

Asp Lys Leu Leu Ile Ala Phe Arg Arg Ala Thr Gly Arg Met Lys Gln

835 840 845

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

850 855 860

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

865 870 875 880

Asp Asn Val Leu Lys Glu Glu Lys Gln His Arg Glu Met Leu Pro Ala

885 890 895

Ile Leu Leu Gln Ile Ala Asp Arg Val Leu Pro Leu Arg Gly Arg Lys

900 905 910

Trp Asp Trp Val Leu Asn Pro Gln Ser Asn Ser Phe Val Leu Gln Gln

915 920 925

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

930 935 940

Gly Leu Ser Phe Ala Arg Ile Glu Gln Leu Glu Ser Leu Arg Met Arg

945 950 955 960

Cys Gln Ala Leu Asn Arg Ile Leu Met Arg Lys Thr Gly Glu Lys Pro

965 970 975

Ala Thr Leu Ala Glu Met Arg Asn Asn Pro Ile Pro Asp Cys Cys Pro

980 985 990

```
Asp Ile Leu Met Arg Leu Asp Ala  Met Lys Glu Gln Arg  Ile Asn Gln
        995                 1000                1005
Thr Ala  Asn Leu Ile Leu Ala  Gln Ala Leu Gly Leu  Arg His Cys
    1010                1015                1020

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

Leu Glu  Lys Leu Ala Leu Leu  Cys Glu Val Phe Asn  Val Pro Ile
    1085                1090                1095
Leu Gln  Val Gly Ala Ala Tyr  Ser Ser Lys Phe Ser  Ala Asn Ala
    1100                1105                1110
Ile Pro  Gly Phe Arg Ala Glu  Glu Cys Ser Ile Asp  Gln Leu Ser
    1115                1120                1125
Phe Tyr  Pro Trp Arg Glu Leu  Lys Asp Ser Arg Glu  Lys Ala Leu
    1130                1135                1140

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

Ala Thr  Asn Leu Leu Cys Asn  Asn Arg Val Ser Cys  Asp Arg Lys
    1205                1210                1215
Lys Asp  Cys Trp Gln Val Lys  Arg Ser Ser Asn Phe  Ser Lys Met
```

1220 1225 1230

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

1235 1240 1245

Gln Glu Gly Ala Gly Gly Asn Phe Phe Val Leu Gly Cys Ser Glu

1250 1255 1260

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

1265 1270 1275

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

1280 1285 1290

Leu Trp Arg Asn Glu Ile Leu Lys Leu Glu Arg Cys Cys Lys Ile

1295 1300 1305

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

1310 1315 1320

Asn Glu Leu

1325

<210> 392

<211> 1090

<212> PRT

<213> Laceyella sediminis

<400> 392

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

1 5 10 15

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

20 25 30

Asn Asp Gly Ile Ala Tyr Tyr Met Asn Trp Leu Val Leu Leu Arg Gln

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

Gln Thr Leu Arg His Leu Tyr Glu Glu Ile Val Pro Ser Val Ile Gly
100 105 110
Lys Ser Gly Asn Ala Ser Leu Lys Ala Arg Phe Phe Leu Gly Pro Leu

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

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

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

305 310 315 320
Met Arg Gly Glu Phe Gly Asp Pro Ala Ile Tyr Gln Phe Leu Ala Gln
325 330 335
Lys Glu Asn His Asp Ile Trp Arg Gly Tyr Pro Glu Arg Val Ile Asp

340 345 350

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

355 360 365

Asp Ala Thr Phe Thr Leu Pro Asp Ser Val Asp His Pro Leu Trp Val

370 375 380

Arg Tyr Glu Ala Pro Gly Gly Thr Asn Ile His Gly Tyr Asp Leu Val

385 390 395 400

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

405 410 415

Pro Asp Glu Asn Gly Ser Trp His Glu Val Lys Lys Val Pro Phe Ser

420 425 430

Leu Ala Lys Ser Lys Gln Phe His Arg Gln Val Trp Leu Gln Glu Glu

435 440 445

Gln Lys Gln Lys Lys Arg Glu Val Val Phe Tyr Asp Tyr Ser Thr Asn

450 455 460

Leu Pro His Leu Gly Thr Leu Ala Gly Ala Lys Leu Gln Trp Asp Arg

465 470 475 480

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

485 490 495

Ile Gly Lys Val Phe Phe Asn Ile Ser Val Asp Val Arg Pro Ala Val

500 505 510

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

515 520 525

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

530 535 540

Glu Gln Leu Glu Lys Trp Val Gly Glu Ser Gly Arg Val Ser Ser Leu

545 550 555 560

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

565 570 575

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

580 585 590

Ala Pro Asp Asn Pro Tyr Lys Phe Phe Tyr Gln Leu Glu Gly Thr Glu
595 600 605
Leu Phe Ala Val His Gln Arg Ser Phe Leu Leu Ala Leu Pro Gly Glu
610 615 620
Asn Pro Pro Gln Lys Ile Lys Gln Met Arg Glu Ile Arg Trp Lys Glu

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

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

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

820 825 830
Glu Asn Leu Arg Ser Tyr Arg Phe Ser Tyr Glu Arg Ser Arg Arg Glu

835 840 845

Asn Lys Lys Leu Met Glu Trp Ser His Arg Ser Ile Pro Lys Leu Val

850 855 860

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

865 870 875 880

Ala Tyr Ser Ser Arg Tyr His Gly Arg Thr Gly Ala Pro Gly Ile Arg

885 890 895

Cys His Ala Leu Thr Glu Ala Asp Leu Arg Asn Glu Thr Asn Ile Ile

900 905 910

His Glu Leu Ile Glu Ala Gly Phe Ile Lys Glu Glu His Arg Pro Tyr

915 920 925

Leu Gln Gln Gly Asp Leu Val Pro Trp Ser Gly Gly Glu Leu Phe Ala

930 935 940

Thr Leu Gln Lys Pro Tyr Asp Asn Pro Arg Ile Leu Thr Leu His Ala

945 950 955 960

Asp Ile Asn Ala Ala Gln Asn Ile Gln Lys Arg Phe Trp His Pro Ser

965 970 975

Met Trp Phe Arg Val Asn Cys Glu Ser Val Met Glu Gly Glu Ile Val

980 985 990

Thr Tyr Val Pro Lys Asn Lys Thr Val His Lys Lys Gln Gly Lys Thr

995 1000 1005

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

1010 1015 1020

Lys Trp Ser Lys Asn Arg Asn Lys Asn Thr Phe Ser Ser Ile Thr

1025 1030 1035

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

1040 1045 1050

Gly Thr Phe Phe Lys Glu Gln Glu Trp Val Glu Gln Lys Thr Phe

1055 1060 1065

Trp Gly Lys Val Gln Ser Met Ile Gln Ala Tyr Met Lys Lys Thr

1070 1075 1080

Ile Val Gln Arg Met Glu Glu

1085 1090

<210> 393

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 393

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

1 5 10 15

<210> 394

<211> 30

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 394

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

20 25 30

<210> 395

<211> 45

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 395

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly

20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

35 40 45

<210> 396

<211> 60

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 396

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly

20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly

35 40 45

Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

50 55 60

<210> 397

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 397

acttgtttaa gt 12

<210> 398

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (1)..(4)

<223> n = any nucleotide that is complementary to nnnn at positions 9-12

<220><221> misc_feature

<222> (9)..(12)

<223> n = any nucleotide that is complementary to nnnn at positions 1-4

<400> 398

nnnngtttnn nn 12

<210> 399

<211> 16

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 399

ggcaccgagt cggtgc 16

<210> 400

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (1)..(7)

<223> n = any nucleotide that is complementary to nnnnnnn at positions 11 to 17

<220><221> misc_feature

<222> (11)..(17)

<223> n = any nucleotide that is complementary to nnnnnnn at positions 1 to 7

<400> 400

nnnnnnnagt nnnnnnn 17

<210> 401

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (2)..(5)

<223> y = t/u or c

<220><221> misc_feature

<222> (8)..(20)

<223> n = any nucleotide

<220><221> misc_feature

<222> (8)..(8)

<223> n= bulge in a duplex

<220><221> misc_feature

<222> (21)..(23)

<223> n= bulge in a duplex

<220><221> misc_feature

<222> (23)..(23)

<223> n = any nucleotide

<220><221> misc_feature

<222> (26)..(29)

<223> r = g or a

<400> 401

gyyyyagnnn nnnnnnnnnn aanuurrrru 30

<210> 402

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 402

Gly Gly Gly Ser

1

<210> 403

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 403

Gly Gly Gly Gly Ser

1 5

<210> 404

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 404

Gly Gly Ser Gly Gly Ser Gly Gly Ser

1 5

<210> 405

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 405

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser

20 25

<210> 406

<211> 35

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 406

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly

20 25 30

Gly Gly Ser

35

<210> 407

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 407

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

1 5 10

<210> 408

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 408

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser

20

<210> 409

<211> 40

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 409

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly

20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser

35 40

<210> 410

<211> 50

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 410

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly

20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly

35 40 45

Gly Ser

50

<210> 411

<211> 55

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 411

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly

1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly

20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly

35 40 45

Gly Ser Gly Gly Gly Gly Ser

50 55

<210> 412

<211> 32

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 412

Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser

1 5 10 15

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

20 25 30

<210> 413

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic Peptide

<400> 413

Pro Lys Lys Lys Arg Lys Val Glu Ala Ser Ser Pro Lys Lys Arg Lys

1 5 10 15

Val Glu Ala Ser

20

<210> 414

<211> 9

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 414

ggtggtagt 9

<210> 415

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400

> 415

ggtggtagtg gagggagcgg cggttca 27

<210> 416

<211> 72

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 416

ggtggaggag gctctggtgg aggcggtagc ggaggcggag ggtcgggtgg tagtggaggg 60

agcggcggtt ca 72

<210> 417

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 417

tcgggatctg agacgcctgg gacctcggaa tcggctacgc ccgaaagt 48

<210> 418

<211> 192

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 418

gtggataaca aatttaacaa agaaatgtgg gcggcgtggg aagaaattcg taacctgccg 60

aacctgaacg gctggcagat gaccgcgttt attgcgagcc tggtggatga tccgagccag 120

agcgcgaacc tgctggcgga agcgaaaaaa ctgaacgatg cgcaggcgcc gaaaaccggc 180

ggtggttctg gt 192

<210> 419

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 419

ggtggttctg ccggtggctc cggttctggc tccagcggtg gcagctctgg tgcgtccggc 60

acgggtactg cgggtggcac tggcagcggt tccggtactg gctctggc 108

<210> 420

<211> 284

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 420

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

1 5 10 15

Leu Ala Tyr Ile Leu Ser Met Glu Pro Cys Gly His Cys Leu Ile Ile

20 25 30

Asn Asn Val Asn Phe Cys Arg Glu Ser Gly Leu Arg Thr Arg Thr Gly

35 40 45

Ser Asn Ile Asp Cys Glu Lys Leu Arg Arg Arg Phe Ser Ser Leu His
50 55 60
Phe Met Val Glu Val Lys Gly Asp Leu Thr Ala Lys Lys Met Val Leu
65 70 75 80
Ala Leu Leu Glu Leu Ala Arg Gln Asp His Gly Ala Leu Asp Cys Cys
85 90 95

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

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

Ser Gly Ser Trp Tyr Val Glu Thr Leu Asp Asp Ile Phe Glu Gln Trp
225 230 235 240
Ala His Ser Glu Asp Leu Gln Ser Leu Leu Leu Arg Val Ala Asn Ala
245 250 255
Val Ser Val Lys Gly Ile Tyr Lys Gln Met Pro Gly Cys Phe Asn Phe
260 265 270
Leu Arg Lys Lys Leu Phe Phe Lys Thr Ser Val Asp
275 280
<210> 421
<

211> 263

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 421

Ser Glu Ser Gln Thr Leu Asp Lys Val Tyr Gln Met Lys Ser Lys Pro
1 5 10 15

Arg Gly Tyr Cys Leu Ile Ile Asn Asn His Asn Phe Ala Lys Ala Arg
20 25 30

Glu Lys Val Pro Lys Leu His Ser Ile Arg Asp Arg Asn Gly Thr His
35 40 45

Leu Asp Ala Gly Ala Leu Thr Thr Thr Phe Glu Glu Leu His Phe Glu

50 55 60

Ile Lys Pro His Asp Asp Cys Thr Val Glu Gln Ile Tyr Glu Ile Leu
65 70 75 80

Lys Ile Tyr Gln Leu Met Asp His Ser Asn Met Asp Cys Phe Ile Cys
85 90 95

Cys Ile Leu Ser His Gly Asp Lys Gly Ile Ile Tyr Gly Thr Asp Gly
100 105 110

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

115 120 125

Cys Pro Ser Leu Ala Gly Lys Pro Lys Val Phe Phe Ile Gln Ala Cys
130 135 140

Gln Gly Asp Asn Tyr Gln Lys Gly Ile Pro Val Glu Thr Asp Ser Glu
145 150 155 160

Glu Gln Pro Tyr Leu Glu Met Asp Leu Ser Ser Pro Gln Thr Arg Tyr
165 170 175

Ile Pro Asp Glu Ala Asp Phe Leu Leu Gly Met Ala Thr Val Asn Asn

180 185 190

Cys Val Ser Tyr Arg Asn Pro Ala Glu Gly Thr Trp Tyr Ile Gln Ser
195 200 205

Leu Cys Gln Ser Leu Arg Glu Arg Cys Pro Arg Gly Asp Asp Ile Leu

210 215 220

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

225 230 235 240

Lys Asn Met Gly Lys Gln Met Pro Gln Pro Thr Phe Thr Leu Arg Lys

245 250 255

Lys Leu Val Phe Pro Ser Asp

260

<210> 422

<211> 147

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 422

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

1 5 10 15

Gly Leu Cys Ile Ile Ile Asn Asn Lys Asn Phe His Lys Ser Thr Gly

20 25 30

Met Thr Ser Arg Ser Gly Thr Asp Val Asp Ala Ala Asn Leu Arg Glu

35 40 45

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

50 55 60

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

65 70 75 80

Ser Lys Arg Ser Ser Phe Val Cys Val Leu Leu Ser His Gly Glu Glu

85 90 95

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

100 105 110

Asn Phe Phe Arg Gly Asp Arg Cys Arg Ser Leu Thr Gly Lys Pro Lys

115 120 125

Leu Phe Ile Ile Gln Ala Cys Arg Gly Thr Glu Leu Asp Cys Gly Ile

130 135 140

Glu Thr Asp

145

<210> 423

<211> 118

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 423

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

1 5 10 15

Thr Ile Cys His Leu Thr Asn Glu Ser Asp Gly His Thr Thr Ser Leu

20 25 30

Tyr Gly Ile Gly Phe Gly Pro Phe Ile Ile Thr Asn Lys His Leu Phe

35 40 45

Arg Arg Asn Asn Gly Thr Leu Leu Val Gln Ser Leu His Gly Val Phe

50 55 60

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

65 70 75 80

Asp Met Ile Ile Ile Arg Met Pro Lys Asp Phe Pro Pro Phe Pro Gln

85 90 95

Lys Leu Lys Phe Arg Glu Pro Gln Arg Glu Glu Arg Ile Cys Leu Val

100 105 110

Thr Thr Asn Phe Gln Thr

115

<210> 424

<211> 101

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 424

Lys Ser Met Ser Ser Met Val Ser Asp Thr Ser Cys Thr Phe Pro Ser

1 5 10 15

Ser Asp Gly Ile Phe Trp Lys His Trp Ile Gln Thr Lys Asp Gly Gln

20 25 30

Cys Gly Ser Pro Leu Val Ser Thr Arg Asp Gly Phe Ile Val Gly Ile

35 40 45

His Ser Ala Ser Asn Phe Thr Asn Thr Asn Asn Tyr Phe Thr Ser Val

50 55 60

Pro Lys Asn Phe Met Glu Leu Leu Thr Asn Gln Glu Ala Gln Gln Trp

65 70 75 80

Val Ser Gly Trp Arg Leu Asn Ala Asp Ser Val Leu Trp Gly Gly His

85 90 95

Lys Val Phe Met Val

100

<210> 425

<211> 939

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 425

Ala Thr Gly Gly Cys Ala Gly Ala Thr Gly Ala Thr Cys Ala Gly Gly

1 5 10 15

Gly Cys Thr Gly Thr Ala Thr Thr Gly Ala Ala Gly Ala Gly Cys Ala

20 25 30

Gly Gly Gly Gly Gly Thr Thr Gly Ala Gly Gly Ala Thr Thr Cys Ala

35 40 45

Gly Cys Ala Ala Ala Thr Gly Ala Ala Gly Ala Thr Thr Cys Ala Gly

50 55 60

Thr Gly Gly Ala Ala Ala Ala Thr Cys Thr Cys Thr Ala Cys Thr Thr

65 70 75 80

Cys Cys Ala Gly Gly Cys Thr Ala Ala Gly Cys Cys Ala Gly Ala Cys

85 90 95

Cys Gly Gly Thr Cys Cys Thr Cys Gly Thr Thr Thr Gly Thr Ala Cys

100 105 110

Cys Gly Thr Cys Cys Cys Thr Cys Thr Thr Cys Ala Gly Thr Ala Ala

115 120 125

Gly Ala Ala Gly Ala Ala Gly Ala Ala Ala Ala Ala Thr Gly Thr Cys

130 135 140

Ala Cys Cys Ala Thr Gly Cys Gly Ala Thr Cys Cys Ala Thr Cys Ala

145 150 155 160

Ala Gly Ala Cys Cys Ala Cys Cys Cys Gly Gly Gly Ala Cys Cys Gly

165 170 175

Ala Gly Thr Gly Cys Cys Thr Ala Cys Ala Thr Ala Thr Cys Ala Gly

180 185 190

Thr Ala Cys Ala Ala Cys Ala Thr Gly Ala Ala Thr Thr Thr Thr Gly

195 200 205

Ala Ala Ala Ala Gly Cys Thr Gly Gly Gly Cys Ala Ala Ala Thr Gly

210 215 220

Cys Ala Thr Cys Ala Thr Ala Ala Thr Ala Ala Ala Cys Ala Ala Cys

225 230 235 240

Ala Ala Gly Ala Ala Cys Thr Thr Thr Gly Ala Thr Ala Ala Ala Gly

245 250 255

Thr Gly Ala Cys Ala Gly Gly Thr Ala Thr Gly Gly Gly Cys Gly Thr

260 265 270

Thr Cys Gly Ala Ala Ala Cys Gly Gly Ala Ala Cys Ala Gly Ala Cys

275 280 285

Ala Ala Ala Gly Ala Thr Gly Cys Cys Gly Ala Gly Gly Cys Gly Cys

290 295 300

Thr Cys Thr Thr Cys Ala Ala Gly Thr Gly Cys Thr Thr Cys Cys Gly

305 310 315 320

Ala Ala Gly Cys Cys Thr Gly Gly Gly Thr Thr Thr Thr Gly Ala Cys

325 330 335

Gly Thr Gly Ala Thr Thr Gly Thr Cys Thr Ala Thr Ala Ala Thr Gly

340 345 350

Ala Cys Thr Gly Cys Thr Cys Thr Thr Gly Thr Gly Cys Cys Ala Ala

355 360 365

Gly Ala Thr Gly Cys Ala Ala Gly Ala Thr Cys Thr Gly Cys Thr Thr

370 375 380

Ala Ala Ala Ala Ala Ala Gly Cys Thr Thr Cys Thr Gly Ala Ala Gly

385 390 395 400

Ala Gly Gly Ala Cys Cys Ala Thr Ala Cys Ala Ala Ala Thr Gly Cys

405 410 415

Cys Gly Cys Cys Thr Gly Cys Thr Thr Cys Gly Cys Cys Thr Gly Cys

420 425 430

Ala Thr Cys Cys Thr Cys Thr Thr Ala Ala Gly Cys Cys Ala Thr Gly

435 440 445

Gly Ala Gly Ala Ala Gly Ala Ala Ala Ala Thr Gly Thr Ala Ala Thr

450 455 460

Thr Thr Ala Thr Gly Gly Gly Ala Ala Ala Gly Ala Thr Gly Gly Thr

465 470 475 480

Gly Thr Cys Ala Cys Ala Cys Cys Ala Ala Thr Ala Ala Ala Gly Gly

485 490 495

Ala Thr Thr Thr Gly Ala Cys Ala Gly Cys Cys Cys Ala Cys Thr Thr

500 505 510

Thr Ala Gly Gly Gly Gly Gly Gly Ala Thr Ala Gly Ala Thr Gly Cys

515 520 525

Ala Ala Ala Ala Cys Cys Cys Thr Thr Thr Thr Ala Gly Ala Gly Ala

530 535 540

Ala Ala Cys Cys Cys Ala Ala Ala Cys Thr Cys Thr Thr Cys Thr Thr

545 550 555 560

Cys Ala Thr Thr Cys Ala Gly Gly Cys Thr Thr Gly Cys Cys Gly Ala

565 570 575

Gly Gly Gly Ala Cys Cys Gly Ala Gly Cys Thr Thr Gly Ala Thr Gly

580 585 590

Ala Thr Gly Gly Cys Ala Thr Cys Cys Ala Gly Gly Cys Cys Gly Ala

595 600 605

Ala Ala Ala Thr Cys Thr Cys Thr Ala Cys Thr Thr Cys Cys Ala Gly

610 615 620

Thr Cys Gly Gly Gly Gly Cys Cys Cys Ala Thr Cys Ala Ala Thr Gly

625 630 635 640

Ala Cys Ala Cys Ala Gly Ala Thr Gly Cys Thr Ala Ala Thr Cys Cys

645 650 655

Thr Cys Gly Ala Thr Ala Cys Ala Ala Gly Ala Thr Cys Cys Cys Ala

660 665 670

Gly Thr Gly Gly Ala Ala Gly Cys Thr Gly Ala Cys Thr Thr Cys Cys

675 680 685

Thr Cys Thr Thr Cys Gly Cys Cys Thr Ala Thr Thr Cys Cys Ala Cys

690 695 700

Gly Gly Thr Thr Cys Cys Ala Gly Gly Cys Thr Ala Thr Thr Ala Cys

705 710 715 720

Thr Cys Gly Thr Gly Gly Ala Gly Gly Ala Gly Cys Cys Cys Ala Gly

725 730 735

Gly Ala Ala Gly Ala Gly Gly Cys Thr Cys Cys Thr Gly Gly Thr Thr

740 745 750

Thr Gly Thr Gly Cys Ala Ala Gly Cys Cys Cys Thr Cys Thr Gly Cys

755 760 765

Thr Cys Cys Ala Thr Cys Cys Thr Gly Gly Ala Gly Gly Ala Gly Cys

770 775 780

Ala Cys Gly Gly Ala Ala Ala Ala Gly Ala Cys Cys Thr Gly Gly Ala

785 790 795 800

Ala Ala Thr Cys Ala Thr Gly Cys Ala Gly Ala Thr Cys Cys Thr Cys

805 810 815

Ala Cys Cys Ala Gly Gly Gly Thr Gly Ala Ala Thr Gly Ala Cys Ala

820 825 830

Gly Ala Gly Thr Thr Gly Cys Cys Ala Gly Gly Cys Ala Cys Thr Thr

835 840 845

Thr Gly Ala Gly Thr Cys Thr Cys Ala Gly Thr Cys Thr Gly Ala Thr

850 855 860

Gly Ala Cys Cys Cys Ala Cys Ala Cys Thr Thr Cys Cys Ala Thr Gly
865 870 875 880
Ala Gly Ala Ala Gly Ala Ala Gly Cys Ala Gly Ala Thr Cys Cys Cys
885 890 895
Cys Thr Gly Thr Gly Thr Gly Gly Thr Cys Thr Cys Cys Ala Thr Gly
900 905 910
Cys Thr Cys Ala Cys Cys Ala Ala Gly Gly Ala Ala Cys Thr Cys Thr

915 920 925
Ala Cys Thr Thr Cys Ala Gly Thr Cys Ala Ala
930 935

<210> 426
<211> 861
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Peptide
<400> 426
Ala Thr Gly Gly Ala Gly Ala Ala Cys Ala Cys Thr Gly Ala Ala Ala
1 5 10 15
Ala Cys Thr Cys Ala Gly Thr Gly Gly Ala Thr Thr Cys Ala Ala Ala
20 25 30

Ala Thr Cys Cys Ala Thr Thr Ala Ala Ala Ala Ala Thr Thr Thr Gly
35 40 45
Gly Ala Ala Cys Cys Ala Ala Ala Gly Ala Thr Cys Ala Thr Ala Cys
50 55 60
Ala Thr Gly Gly Ala Ala Gly Cys Gly Ala Ala Thr Cys Ala Ala Thr
65 70 75 80
Gly Gly Ala Ala Ala Ala Thr Cys Thr Cys Thr Ala Cys Thr Thr Cys
85 90 95

Cys Ala Gly Thr Cys Thr Gly Gly Ala Ala Thr Ala Thr Cys Cys Cys
100 105 110
Thr Gly Gly Ala Cys Ala Ala Cys Ala Gly Thr Thr Ala Thr Ala Ala
115 120 125

Ala Ala Thr Gly Gly Ala Thr Thr Ala Thr Cys Cys Thr Gly Ala Gly
130 135 140
Ala Thr Gly Gly Gly Thr Thr Thr Ala Thr Gly Thr Ala Thr Ala Ala
145 150 155 160

Thr Ala Ala Thr Thr Ala Ala Thr Ala Ala Thr Ala Ala Gly Ala Ala
165 170 175
Thr Thr Thr Thr Cys Ala Thr Ala Ala Ala Ala Gly Cys Ala Cys Thr
180 185 190
Gly Gly Ala Ala Thr Gly Ala Cys Ala Thr Cys Thr Cys Gly Gly Thr
195 200 205
Cys Thr Gly Gly Thr Ala Cys Ala Gly Ala Thr Gly Thr Cys Gly Ala
210 215 220

Thr Gly Cys Ala Gly Cys Ala Ala Ala Cys Cys Thr Cys Ala Gly Gly
225 230 235 240
Gly Ala Ala Ala Cys Ala Thr Thr Cys Ala Gly Ala Ala Ala Cys Thr
245 250 255
Thr Gly Ala Ala Ala Thr Ala Thr Gly Ala Ala Gly Thr Cys Ala Gly
260 265 270
Gly Ala Ala Thr Ala Ala Ala Ala Ala Thr Gly Ala Thr Cys Thr Thr
275 280 285

Ala Cys Ala Cys Gly Thr Gly Ala Ala Gly Ala Ala Ala Thr Thr Gly
290 295 300
Thr Gly Gly Ala Ala Thr Thr Gly Ala Thr Gly Cys Gly Thr Gly Ala
305 310 315 320
Thr Gly Thr Thr Thr Cys Thr Ala Ala Ala Gly Ala Ala Gly Ala Thr
325 330 335
Cys Ala Cys Ala Gly Cys Ala Ala Ala Ala Gly Gly Ala Gly Cys Ala
340 345 350

Gly Thr Thr Thr Thr Gly Thr Thr Thr Gly Thr Gly Thr Gly Cys Thr
355 360 365
Thr Cys Thr Gly Ala Gly Cys Cys Ala Thr Gly Gly Thr Gly Ala Ala

370 375 380

Gly Ala Ala Gly Gly Ala Ala Thr Ala Ala Thr Thr Thr Thr Thr Gly

385 390 395 400

Gly Ala Ala Cys Ala Ala Ala Thr Gly Gly Ala Cys Cys Thr Gly Thr

405 410 415

Thr Gly Ala Cys Cys Thr Gly Ala Ala Ala Ala Ala Ala Ala Thr Ala

420 425 430

Ala Cys Ala Ala Ala Cys Thr Thr Thr Thr Thr Cys Ala Gly Ala Gly

435 440 445

Gly Gly Gly Ala Thr Cys Gly Thr Thr Gly Thr Ala Gly Ala Ala Gly

450 455 460

Thr Cys Thr Ala Ala Cys Thr Gly Gly Ala Ala Ala Ala Cys Cys Cys

465 470 475 480

Ala Ala Ala Cys Thr Thr Thr Thr Cys Ala Thr Thr Ala Thr Thr Cys

485 490 495

Ala Gly Gly Cys Cys Thr Gly Cys Cys Gly Thr Gly Gly Thr Ala Cys

500 505 510

Ala Gly Ala Ala Cys Thr Gly Gly Ala Cys Thr Gly Thr Gly Gly Cys

515 520 525

Ala Thr Thr Gly Ala Gly Ala Cys Ala Gly Ala Ala Ala Ala Thr Cys

530 535 540

Thr Cys Thr Ala Cys Thr Thr Cys Cys Ala Gly Ala Gly Thr Gly Gly

545 550 555 560

Thr Gly Thr Thr Gly Ala Thr Gly Ala Thr Gly Ala Cys Ala Thr Gly

565 570 575

Gly Cys Gly Thr Gly Thr Cys Ala Thr Ala Ala Ala Ala Thr Ala Cys

580 585 590

Cys Ala Gly Thr Gly Gly Ala Gly Gly Cys Cys Gly Ala Cys Thr Thr

595 600 605

Cys Thr Thr Gly Thr Ala Thr Gly Cys Ala Thr Ala Cys Thr Cys Cys

610 615 620

Ala Cys Ala Gly Cys Ala Cys Cys Thr Gly Gly Thr Thr Ala Thr Thr
625 630 635 640
Ala Thr Thr Cys Thr Thr Gly Gly Cys Gly Ala Ala Ala Thr Thr Cys
645 650 655
Ala Ala Ala Gly Gly Ala Thr Gly Gly Cys Thr Cys Cys Thr Gly Gly
660 665 670

Thr Thr Cys Ala Thr Cys Cys Ala Gly Thr Cys Gly Cys Thr Thr Thr
675 680 685
Gly Thr Gly Cys Cys Ala Thr Gly Cys Thr Gly Ala Ala Ala Cys Ala
690 695 700
Gly Thr Ala Thr Gly Cys Cys Gly Ala Cys Ala Ala Gly Cys Thr Thr
705 710 715 720
Gly Ala Ala Thr Thr Thr Ala Thr Gly Cys Ala Cys Ala Thr Thr Cys
725 730 735

Thr Thr Ala Cys Cys Cys Gly Gly Gly Thr Thr Ala Ala Cys Cys Gly
740 745 750
Ala Ala Ala Gly Gly Thr Gly Gly Cys Ala Ala Cys Ala Gly Ala Ala
755 760 765
Thr Thr Thr Gly Ala Gly Thr Cys Cys Thr Thr Thr Thr Cys Cys Thr
770 775 780
Thr Thr Gly Ala Cys Gly Cys Thr Ala Cys Thr Thr Thr Thr Cys Ala
785 790 795 800

Thr Gly Cys Ala Ala Ala Gly Ala Ala Ala Cys Ala Gly Ala Thr Thr
805 810 815
Cys Cys Ala Thr Gly Thr Ala Thr Thr Gly Thr Thr Thr Cys Cys Ala
820 825 830
Thr Gly Cys Thr Cys Ala Cys Ala Ala Ala Ala Gly Ala Ala Cys Thr
835 840 845
Cys Thr Ala Thr Thr Thr Thr Thr Ala Thr Cys Ala Cys
850 855 860

<210>

427

<211> 164

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 427

Met Ala Glu Gly Ser Val Ala Arg Gln Pro Asp Leu Leu Thr Cys Asp
1 5 10 15

Asp Glu Pro Ile His Ile Pro Gly Ala Ile Gln Pro His Gly Leu Leu
20 25 30

Leu Ala Leu Ala Ala Asp Met Thr Ile Val Ala Gly Ser Asp Asn Leu
35 40 45

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

50 55 60

Ala Asp Val Phe Asp Ser Glu Thr His Asn Arg Leu Thr Ile Ala Leu
65 70 75 80

Ala Glu Pro Gly Ala Ala Val Gly Ala Pro Ile Thr Val Gly Phe Thr
85 90 95

Met Arg Lys Asp Ala Gly Phe Ile Gly Ser Trp His Arg His Asp Gln
100 105 110

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

115 120 125

Ser Ala Leu Glu Lys Glu Val Ser Ala Leu Glu Lys Glu Val Ser Ala
130 135 140

Leu Glu Lys Glu Val Ser Ala Leu Glu Lys Glu Val Ser Ala Leu Glu
145 150 155 160

Lys Gly Gly Ser

<210> 428

<211> 239

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 428

Met Gly Gly Ser Lys Val Ser Ala Leu Lys Glu Lys Val Ser Ala Leu

1 5 10 15

Lys Glu Lys Val Ser Ala Leu Lys Glu Lys Val Ser Ala Leu Lys Glu

20 25 30

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

35 40 45

Ala Glu Pro Gln Ala Phe Phe Arg Arg Thr Asn Ser Ala Ile Arg Arg

50 55 60

Leu Gln Ala Ala Glu Thr Leu Glu Ser Ala Cys Ala Ala Ala Ala Gln

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

Gln Ala Arg Arg Leu Tyr Thr Ile Asn Pro Val Arg Ile Ile Pro Asp

130 135 140

Ile Asn Tyr Arg Pro Val Pro Val Thr Pro Tyr Leu Asn Pro Val Thr

145 150 155 160

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

165 170 175

Val His Leu Glu Phe Met Arg Asn Ile Gly Met His Gly Thr Met Ser

180 185 190

Ile Ser Ile Leu Arg Gly Glu Arg Leu Trp Gly Leu Ile Val Cys His

195 200 205

His Arg Thr Pro Tyr Tyr Val Asp Leu Asp Gly Arg Gln Ala Cys Glu

210 215 220

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

225 230 235

<210> 429

<211> 154

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 429

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
1 5 10 15

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Leu Ile
35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50 55 60

Leu Gly Tyr Gly Leu Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys
65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
100 105 110

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

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Thr
145 150

<210> 430

<211> 85

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 430

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

1 5 10 15

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

20 25 30

Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu

35 40 45

Ser Tyr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His

50 55 60

Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met

65 70 75 80

Asp Glu Leu Tyr Lys

85

<210> 431

<211> 156

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 431

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

1 5 10 15

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

20 25 30

Lys Gly Asp Ala Thr Arg Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr

35 40 45

Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr

50 55 60

Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Lys His Met Lys Arg His

65 70 75 80

Asp Phe Phe Lys Ser Ala Met Pro Lys Gly Tyr Val Gln Glu Arg Thr

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

Asn Ser His Lys Val Tyr Ile Thr Ala Asp Lys Arg

145 150 155

<210> 432

<211> 81

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 432

Lys Asn Gly Ile Lys Ala Lys Phe Lys Ile Arg His Asn Val Lys Asp

1 5 10 15

Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly

20 25 30

Arg Gly Pro Val Leu Leu Pro Arg Asn His Tyr Leu Ser Thr Arg Ser

35 40 45

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

50 55 60

Glu Phe Val Thr Ala Ala Gly Ile Lys His Gly Arg Asp Glu Arg Tyr

65 70 75 80

Lys

<210> 433

<211> 229

<212> PRT

<213> Rattus rattus

<400> 433

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

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

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

195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys
225
<210> 434
<211> 236
<212> PRT
<213> Homo sapiens
<400> 434
Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg

1 5 10 15

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

20 25 30

Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg

35 40 45

Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val

50 55 60

Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met

65 70 75 80

Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys

85 90 95

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

100 105 110

Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro

145 150 155 160

Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met

165 170 175

Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys

180 185 190

Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu

195 200 205

His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu

210 215 220

Ala Thr Gly Leu Ile His Pro Ser Val Ala Trp Arg

225 230 235

<210> 435

<211> 325

<212> PRT

<213> Homo sapiens

<400> 435

Met Glu Leu Lys Tyr His Pro Glu Met Arg Phe Phe His Trp Phe Ser
1 5 10 15

Lys Trp Arg Lys Leu His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr
20 25 30

Ile Ser Trp Ser Pro Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe
35 40 45

Leu Ala Glu Asp Pro Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu
50 55 60

Tyr Tyr Phe Trp Asp Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys
65 70 75 80

Gln Lys Arg Asp Gly Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp
85 90 95

Glu Phe Gln His Cys Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu
100 105 110

Phe Glu Pro Trp Asn Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile
115 120 125

Met Leu Gly Glu Ile Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr
130 135 140

Phe Asn Phe Asn Asn Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr
145 150 155 160

Leu Cys Tyr Glu Val Glu Arg Met His Asn Asp Thr Trp Val Leu Leu
165 170 175

Asn Gln Arg Arg Gly Phe Leu Cys Asn Gln Ala Pro His Lys His Gly
180 185 190

Phe Leu Glu Gly Arg His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro
195 200 205

Phe Trp Lys Leu Asp Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr
210 215 220

Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile

225 230 235 240

Ser Lys Asn Lys His Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr

245 250 255

Asp Asp Gln Gly Arg Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala

260 265 270

Gly Ala Lys Ile Ser Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp

275 280 285

Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly

290 295 300

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

305 310 315 320

Gln Asn Gln Glu Asn

325

<210> 436

<211> 208

<212> PRT

<213> Petromyzon marinus

<400> 436

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

1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg

20 25 30

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

35 40 45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly

50 55 60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg

65 70 75 80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro

85 90 95

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

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

<210> 437
<211> 200
<212> PRT
<213> Homo sapiens

<400> 437
Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45
Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95
Phe Leu Arg Gly Asn Pro Tyr Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

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

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu Leu Asp

195 200

<210> 438

<211> 181

<212> PRT

<213> Homo sapiens

<400> 438

Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val

20 25 30

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

35 40 45

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

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

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

100 105 110

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

Arg Arg Ile Leu Leu
180

<210> 439
<211> 25
<212> RNA
<213> Artificial Sequence
<220><223> Synthetic Oligonucleotide
<400> 439
gggaacaaag cugaaguacu uaccc 25
<210> 440
<211> 18
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Oligonucleotide
<400> 440
gggtagggcg ggttggga 18
<210> 441
<211> 25
<212> DNA
<213> Artificial Sequence
<220><223
> Synthetic Oligonucleotide
<220><221> misc_feature
<222> (25)..(25)

<223> 3' Thiol modification

<400> 441

ttataactat tcctaaaaaa aaaaa 25

<210> 442

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (1)..(1)

<223> 5' Thiol modification

<400> 442

aaaaaaaaaa ctcccctaat aacaat 26

<210> 443

<211> 45

<212> RNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 443

ggguaggaau aguuauaauu ucccuuuccc auuguuauua gggag 45

<210> 444

<211> 12

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (1)..(1)

<223> 5? biotin modification

<220><221> misc_feature

<222> (12)..(12)

<223> 3? quencher

<400> 444

ucucguacgu uc 12

<210> 445

<211

> 24

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (1)..(1)

<223> 5? biotin modification

<220><221> misc_feature

<222> (24)..(24)

<223> 3? quencher

<400> 445

ucucguacgu ucucucguac guuc 24

<210> 446

<211> 66

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 446

tgtggttggt gtggttggtt catggtcata ttggtttttt ttttttttc caaccacagt 60

ctctgt 66

<210> 447

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 447

ggttggtagt ctcgaattgc tctctttcac tggcc 35

<210> 448

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 448

gaaattaata cgactcacta taggggttg gttcatggtc atattggt 48

<210> 449

<211> 57

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 449

gaaattaata cgactcacta taggggttg gtgtggttgg ttcatggtca tattggt 57

<210> 450

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 450

ggccagtgaa agagagcaat tcgagactac c 31

<210> 451

<211> 64

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 451

gauuuagacu accccaaaaa cgaaggggac uaaaacccag ugaaagagag caauucgaga 60

cuac 64

<210> 452

<211> 64

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 452

gauuuagacu accccaaaaa cgaaggggac uaaaacaaag agagcaauuc gagacuacca 60

acca 64

<210> 453

<211> 64

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 453

gauuuagacu accccaaaaa cgaaggggac uaaaacagac uaccaaccac agagacugug 60

guug 64

<210> 454

<211> 106

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 454

gttagatcgc aagcatatca ttgcgcttgc gatctaactg ctgcgccgcc gggaaaatac 60

tgtacggtta gatcgcatag tctcgaattg ctctctttca ctggcc 106

<210> 455

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 455

gttagatcgc aagcatatca ttgcgcttgc gatctaactg ctgcgccgcc gggaaaatac 60

tgtacggtta g 71

<210> 456

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 456

atcgcatagt ctcgaattgc tctctttcac tggcc 35

<210> 457

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 457

gaaattaata cgactcacta tagggatcgc aagcatatca ttgcgcttgc 50

<210> 458

<211> 31

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 458

ggccagtgaa agagagcaat tcgagactat g 31

<210> 459

<211> 64

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 459

gauuuagacu accccaaaaa cgaaggggac uaaaacccag ugaaagagag caauucgaga 60

cuau 64

<210> 460

<211> 64

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 460

gauuuagacu accccaaaaa cgaaggggac uaaaacagag caauucgaga cuaugcgauc 60

uaac 64

<210> 461

<211> 64

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 461

gauuuagacu accccaaaaa cgaaggggac uaaaacacua ugcgaucuaa ccguacagua 60

uuuu 64

<210> 462

<211> 7

<212> PRT

<213> Artificial Sequence

<220><

223> Synthetic Peptide

<400> 462

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 463

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 463

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 464

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 464

Pro Ala Ala Lys Arg Val Lys Leu Asp

1 5

<210> 465

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 465

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro

1 5 10

<210> 466

<211> 38

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 466

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> 467

<211> 42

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 467

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> 468

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 468

Val Ser Arg Lys Arg Pro Arg Pro

1 5

<210> 469

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 469

Pro Pro Lys Lys Ala Arg Glu Asp

1 5

<210> 470

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 470

Pro Gln Pro Lys Lys Lys Pro Leu

1 5

<210> 471

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 471

Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro

1 5 10

<210> 472

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 472

Asp Arg Leu Arg Arg

1 5

<210> 473

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223

> Synthetic Peptide

<400> 473

Pro Lys Gln Lys Lys Arg Lys

1 5

<210> 474

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 474

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu

1 5 10

<210> 475

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 475

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg

1 5 10

<210> 476

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 476

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> 477

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 477

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15

Lys

<210> 478

<211> 107

<212> PRT

<213> Homo sapiens

<400> 478

Ile Glu Val Met Tyr Pro Pro Pro Tyr Leu Asp Asn Glu Lys Ser Asn
1 5 10 15

Gly Thr Ile Ile His Val Lys Gly Lys His Leu Cys Pro Ser Pro Leu
20 25 30

Phe Pro Gly Pro Ser Lys Pro Phe Trp Val Leu Val Val Val Gly Gly
35 40 45

Val Leu Ala Cys Tyr Ser Leu Leu Val Thr Val Ala Phe Ile Ile Phe

50 55 60

Trp Val Arg Ser Lys Arg Ser Arg Leu Leu His Ser Asp Tyr Met Asn
65 70 75 80

Met Thr Pro Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln Pro Tyr
85 90 95

Ala Pro Pro Arg Asp Phe Ala Ala Tyr Arg Ser
100 105

<210> 479

<211> 107

<212> PRT

<213> Homo sapiens

<400> 479

Ile Glu Val Met Tyr Pro Pro Pro Tyr Leu Asp Asn Glu Lys Ser Asn

1 5 10 15

Gly Thr Ile Ile His Val Lys Gly Lys His Leu Cys Pro Ser Pro Leu

20 25 30

Phe Pro Gly Pro Ser Lys Pro Phe Trp Val Leu Val Val Val Gly Gly

35 40 45

Val Leu Ala Cys Tyr Ser Leu Leu Val Thr Val Ala Phe Ile Ile Phe

50 55 60

Trp Val Arg Ser Lys Arg Ser Arg Leu Leu His Ser Asp Tyr Met Asn

65 70 75 80

Met Thr Pro Arg Arg Pro Gly Pro Thr Arg Lys His Tyr Gln Pro Tyr

85 90 95

Ala Pro Pro Arg Asp Phe Ala Ala Tyr Arg Ser

100 105

<210> 480

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 480

gccaaattgg acgaccctcg 20

<210> 481

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 481

cgaggagacc cccgtttcgg 20

<210> 482

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 482

cccgccgccg ccgtggctcg 20

<210> 483

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 483

tgagctctac gagatccaca 20

<210> 484

<211> 1145

<212> PRT

<213> Alicyclobacillus macrosporangiidus

<400> 484

Met Val Ala Val Lys Ser Ile Lys Val Lys Leu Met Leu Gly His Leu
1 5 10 15

Pro Glu Ile Arg Glu Gly Leu Trp His Leu His Glu Ala Val Asn Leu
20 25 30

Gly Val Arg Tyr Tyr Thr Glu Trp Leu Ala Leu Leu Arg Gln Gly Asn
35 40 45

Leu Tyr Arg Arg Gly Lys Asp Gly Ala Gln Glu Cys Tyr Met Thr Ala
50 55 60

Glu Gln Cys Arg Gln Glu Leu Leu Val Arg Leu Arg Asp Arg Gln Lys
65 70 75 80

Arg Asn Gly His Thr Gly Asp Pro Gly Thr Asp Glu Glu Leu Leu Gly
85 90 95

Val Ala Arg Arg Leu Tyr Glu Leu Leu Val Pro Gln Ser Val Gly Lys
100 105 110

Lys Gly Gln Ala Gln Met Leu Ala Ser Gly Phe Leu Ser Pro Leu Ala
115 120 125

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

130 135 140

Pro Ala Trp Met Gly Met Lys Glu Ala Gly Asp Ser Arg Trp Val Glu

145 150 155 160

Ala Lys Ala Arg Tyr Glu Ala Asn Lys Ala Lys Asp Pro Thr Lys Gln

165 170 175

Val Ile Ala Ser Leu Glu Met Tyr Gly Leu Arg Pro Leu Phe Asp Val

180 185 190

Phe Thr Glu Thr Tyr Lys Thr Ile Arg Trp Met Pro Leu Gly Lys His

195 200 205

Gln Gly Val Arg Ala Trp Asp Arg Asp Met Phe Gln Gln Ser Leu Glu

210 215 220

Arg Leu Met Ser Trp Glu Ser Trp Asn Glu Arg Val Gly Ala Glu Phe

225 230 235 240

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

245 250 255

Gly Gln Glu His Leu Val Ala Leu Ala Gln Arg Leu Glu Gln Glu Met

260 265 270

Lys Glu Ala Ser Pro Gly Phe Glu Ser Lys Ser Ser Gln Ala His Arg

275 280 285

Ile Thr Lys Arg Ala Leu Arg Gly Ala Asp Gly Ile Ile Asp Asp Trp

290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

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

355 360 365

Leu Glu Asp Ala Lys Gln Phe Ala Thr Phe Thr Leu Pro Ser Pro Cys

370 375 380

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

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

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

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

625 630 635 640

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

645 650 655

Gln Arg Leu Asp Arg Leu Asn Arg Leu Arg Ser Gln Met Ala Ser Leu

660 665 670

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

675 680 685

Asn Trp Glu Arg Leu Gln Ser Ser Met Glu Arg Gly Gly Glu Arg Met

690 695 700

Pro Ser Asp Trp Trp Asp Leu Phe Gln Ala Gln Val Arg Tyr Leu Ala

705 710 715 720

Gln His Arg Asp Ala Ser Gly Glu Ala Trp Gly Arg Met Val Gln Ala

725 730 735

Ala Val Arg Thr Leu Trp Arg Gln Leu Ala Lys Gln Val Arg Asp Trp

740 745 750

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

755 760 765

Ala Arg Asp Val Pro Gly Gly His Ser Leu Ala Gln Leu Asp Tyr Leu

770 775 780

Glu Arg Gln Tyr Arg Phe Leu Arg Ser Trp Ser Ala Phe Ser Val Gln

785 790 795 800

Ala Gly Gln Val Val Arg Ala Glu Arg Asp Ser Arg Phe Ala Val Ala

805 810 815

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

820 825 830

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

835 840 845

Gly Arg Arg Ala Gly Gln Trp Gln Ala Val Tyr Pro Pro Cys Gln Leu

850 855 860

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

865 870 875 880

Pro Ser Glu Asn Ser Gln Leu Met Val Trp Ser His Arg Gly Val Leu

885 890 895

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

900 905 910

Ile Pro Ala Ala Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro

915 920 925

Gly Ile Arg Cys Arg Arg Val Pro Ser Ile Pro Leu Lys Asp Ala Pro

930 935 940

Ser Ile Pro Ile Trp Leu Ser His Tyr Leu Lys Gln Thr Glu Arg Asp

945 950 955 960

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

965 970 975

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

980 985 990

His Ala Asp Ile Asn Ala Ala His Asn Leu Gln Arg Arg Leu Trp Glu

995 1000 1005

Asn Phe Asp Leu Ser Asp Ile Arg Val Arg Cys Asp Arg Arg Glu

1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

Gly Val Ser Phe Thr Val Gly Asp Ala Lys Thr Arg Arg Arg Ser

1055 1060 1065

Ser Ala Ser Gln Gly Glu Gly Asp Asp Leu Ser Asp Glu Glu Gln

1070 1075 1080

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

1085 1090 1095

Leu Phe Arg Asp Pro Ser Gly Phe Val Asn Gly Gly Arg Trp Thr

1100 1105 1110

Ala Gln Arg Ala Phe Trp Gly Met Val His Asn Arg Ile Glu Thr

1115 1120 1125

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

1130 1135 1140

Arg Gly

1145

<210> 485

<211> 1108

<212> PRT

<213> Bacillus hisashii

<400> 485

Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
1 5 10 15

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
20 25 30

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
35 40 45

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
50 55 60

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
65 70 75 80

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
85 90 95

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
100 105 110

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
115 120 125

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
130 135 140

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
145 150 155 160

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly

165 170 175

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

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

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

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

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
420 425 430

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

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

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

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
625 630 635 640
Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
645 650 655
Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

660 665 670

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

675 680 685

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

690 695 700

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

705 710 715 720

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

725 730 735

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

740 745 750

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

755 760 765

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

770 775 780

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

785 790 795 800

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln

805 810 815

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn

820 825 830

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

835 840 845

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

850 855 860

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

865 870 875 880

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

885 890 895

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

900 905 910

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

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

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

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

<210> 486
<211> 1364
<212> PRT
<213> Unknown
<220><223> Uncultured organism
<400> 486

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

1 5 10 15

Lys Gly Ile Arg Glu Arg Gly Lys Thr Lys Glu Gly Thr Asp Lys Lys

20 25 30

Lys Ser Gly Arg Lys Ser Ser Trp Pro Met Asp Lys Ala Ala Trp Glu

35 40 45

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

50 55 60

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

65 70 75 80

Ser Lys Lys Leu Glu Tyr Tyr Lys Lys Leu Ala Gly Ser Ala Pro Trp

85 90 95

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

100 105 110

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

115 120 125

Ser Lys Ala Ala Tyr Arg Arg Leu Phe Asp Ala Gly Cys Leu Pro Met

130 135 140

Pro Glu Phe Ala Lys Tyr Ile Asp Glu Asn Gln Ile Glu Phe Gly Asp

145 150 155 160

Leu Lys Leu Ser Asp Cys Gly Ala Glu Trp Lys Arg Gly Met Trp Asn

165 170 175

Gln Ala Gly Gln Arg Val Arg Ser His Met Gly Trp Gln Arg Arg Arg

180 185 190

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

195 200 205

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

210 215 220

Ile Leu Pro Gly Lys Ala Ala Pro Asp Gln Asn Asp Trp Gln Glu Arg

225 230 235 240

Pro Ala Tyr Gly Asn Gln Met Trp Phe Ile Gly Leu Arg Ser Tyr Glu

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

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

305 310 315 320

Ala Tyr Gly Asp Ala Arg Leu Phe Asp Trp Leu Ser Lys Thr Asp Gln

325 330 335

Arg Phe Leu Trp Asp Gly Phe Asp Ala Asp Gly Lys Ile Leu Cys Gly

340 345 350

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

355 360 365

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

370 375 380

Leu His Pro Val Trp Pro Phe Phe Gly Glu Ser Ser Ala Val Pro Tyr

385 390 395 400

Glu Ile Glu Tyr Asp Leu Glu Thr Ala Cys Pro Thr Ala Ile Arg Leu

405 410 415

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

420 425 430

Thr Arg Leu Pro Leu Ala Glu Tyr Ala Asp Leu Ala Ser Ser Phe Gln

435 440 445

Leu Pro Thr Pro Val Arg Leu Asp Val Leu Val Glu Ile Arg Glu Val

450 455 460

Thr Arg Ala Gly Arg Lys Val Thr Cys Pro Phe Ser Tyr Phe Lys Gln

465 470 475 480

Asn Gly Val Trp Tyr Val Arg Glu Gly Glu Ile Pro Ser Gly Glu Ser

485 490 495

Ile Gln Ile Lys Gln Thr Asp Arg Lys Ile Glu Asn Gly Lys Ile Phe
500 505 510
Ile Ser Ser Lys Leu Arg Met Ala Tyr Arg Asp Asp Leu Met Val Ser

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

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

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

705 710 715 720
Ile Leu Arg Val Ser Arg Leu Leu Ser Leu Asp Lys Phe Glu Lys Thr
725 730 735
Ile Phe Glu Gln Ser Asp Val Arg Asp Arg Pro Asn Lys Lys Gly Leu

740 745 750

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

755 760 765

Glu Val Arg Lys Asn Cys Asp Lys Ala Ala Glu Ile Leu Ile Arg Trp

770 775 780

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

785 790 795 800

Asp Ile Ser Phe Trp Lys Tyr Met Ala Val Lys Asn Pro Pro Leu Ser

805 810 815

Ala Val Val Asp Val Ala Pro Ser Thr Ile Val Pro Asp Asp Gly Pro

820 825 830

Asp Arg Glu Thr Leu Lys Lys Lys Arg Gln Glu Glu Glu Glu Lys Phe

835 840 845

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

850 855 860

Ser Gly Tyr Asp Ala Asp His Arg Arg Pro Ala Thr Gly Gly Leu Trp

865 870 875 880

His Asp Leu Asp Arg Thr Leu Ile Arg Glu Ile Ser Tyr Gly Asp Arg

885 890 895

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

900 905 910

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

915 920 925

Arg Thr Leu Asn Asp Ala Asn Arg Ile Pro Lys Gly Arg Thr Leu Arg

930 935 940

Gly Gly Leu Ser Met Gly Arg Leu Asn Phe Leu Lys Glu Val Gly Asp

945 950 955 960

Phe Val Lys Lys Trp Ser Cys Arg Pro Arg Trp Pro Gly Asp Arg Arg

965 970 975

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

980 985 990

Glu His Leu Arg Asp Asp Arg Ile Lys Arg Leu Ala His Leu Ile Val
995 1000 1005

Ala Gln Ala Leu Gly Phe Glu Pro Asp Ile Arg Arg Gly Leu Trp
1010 1015 1020

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

1025 1030 1035

Glu Thr Arg Arg Phe Phe Ala Glu Gly Ala Ala Gly Glu Leu Arg
1040 1045 1050

Glu Val Ser Arg Pro Ala Glu Ile Asp Asp Asp Ala Ala Ala Arg
1055 1060 1065

Pro His Thr Val Ser Ala Pro Ala His Ile Val Val Phe Glu Asn
1070 1075 1080

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

1085 1090 1095

Ala Gly Leu Met Gln Trp Ala His Arg Gln Ile Val His Phe Thr
1100 1105 1110

Lys Gln Val Ala Ser Leu Tyr Gly Leu Lys Val Ala Met Val Tyr
1115 1120 1125

Ala Ala Phe Ser Ser Lys Phe Cys Ser Arg Cys Gly Ser Pro Gly
1130 1135 1140

Ala Arg Val Ser Arg Phe Asp Pro Ala Trp Arg Asn Gln Glu Trp

1145 1150 1155

Phe Lys Arg Arg Thr Ser Asn Pro Arg Ser Lys Val Asp His Ser
1160 1165 1170

Leu Lys Arg Ala Ser Glu Asp Pro Thr Ala Asp Glu Thr Arg Pro
1175 1180 1185

Trp Val Leu Ile Glu Gly Gly Lys Glu Phe Val Cys Ala Asn Ala
1190 1195 1200

Lys Cys Ser Ala His Asp Glu Pro Leu Asn Ala Asp Glu Asn Ala

1205 1210 1215

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

1220 1225 1230

Arg Thr Lys Val Asn Pro Ala Gly Ala Leu Lys Gly Lys Leu Arg

1235 1240 1245

Phe Glu Thr Gly Ile His Ser Phe Arg Pro Pro Val Ser Gly Ser

1250 1255 1260

Pro Phe Trp Ser Pro Met Ala Glu Pro Ala Gln Lys Lys Lys Ile

1265 1270 1275

Gly Ala Ala Ala Pro Gly Ala Asp Val Asp Glu Ala Gly Asp Ala

1280 1285 1290

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

1295 1300 1305

Phe Arg Asn Lys Gln Tyr Trp Tyr Glu Gly Lys Ile Phe Trp Ser

1310 1315 1320

Asn Val Met Met Ala Val Glu Ala Lys Ile Ala Gly Ala Ser Val

1325 1330 1335

Gly Ala Lys Pro Val Ala Ala Ser Trp Gly Gln Ala Gln Pro Gln

1340 1345 1350

Ser Gly Pro Gly Leu Ala Lys Pro Gly Gly Asp

1355 1360

<210> 487

<211> 1364

<212> PRT

<213> Unknown

<220><223> Uncultured organism

<400> 487

Met Asn Arg Ile Tyr Gln Gly Arg Val Thr Lys Val Glu Val Pro Asp

1 5 10 15

Gly Lys Asp Glu Lys Gly Asn Ile Lys Trp Lys Lys Leu Glu Asn Trp

20 25 30

Ser Asp Ile Leu Trp Gln His His Met Leu Phe Gln Asp Ala Val Asn

35 40 45

Tyr Tyr Thr Leu Ala Leu Ala Ala Ile Ser Gly Ser Ala Val Gly Ser

50 55 60

Asp Glu Lys Ser Ile Ile Leu Arg Glu Trp Ala Val Gln Val Gln Asn
65 70 75 80

Ile Trp Glu Lys Ala Lys Lys Lys Ala Thr Val Phe Glu Gly Pro Gln
85 90 95

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

Ile Ala Ala Lys His Ile Leu Arg Thr Ser Glu Ala Lys Pro Glu Gln
115 120 125

Arg Ala Ser Ala Leu Ile Arg Leu Leu Glu Glu Ile Asp Lys Lys Asn
130 135 140

His Asn Val Val Cys Gly Glu Arg Leu Pro Phe Phe Cys Pro Arg Asn
145 150 155 160

Ile Gln Ser Lys Arg Ser Pro Thr Ser Lys Ala Val Ser Ser Val Gln
165 170 175

Glu Gln Lys Arg Gln Glu Glu Val Arg Arg Phe His Asn Met Gln Pro
180 185 190

Glu Glu Val Val Lys Asn Ala Val Thr Leu Asp Ile Ser Leu Phe Lys
195 200 205

Ser Ser Pro Lys Ile Val Phe Leu Glu Asp Pro Lys Lys Ala Arg Ala
210 215 220

Glu Leu Leu Lys Gln Phe Asp Asn Ala Cys Lys Lys His Lys Glu Leu
225 230 235 240

Val Gly Ile Lys Lys Ala Phe Thr Glu Ser Ile Asp Lys His Gly Ser
245 250 255

Ser Leu Lys Val Pro Ala Pro Gly Ser Lys Pro Ser Gly Leu Tyr Pro
260 265 270

Ser Ala Ile Val Phe Lys Tyr Phe Pro Val Asp Ile Thr Lys Thr Val
275 280 285

Phe Leu Lys Ala Thr Glu Lys Leu Ala Met Gly Lys Asp Arg Glu Val
290 295 300

Thr Asn Asp Pro Ile Ala Asp Ala Arg Val Asn Asp Lys Pro His Phe
305 310 315 320
Asp Tyr Phe Thr Asn Ile Ala Leu Ile Arg Glu Lys Glu Lys Asn Arg
325 330 335

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

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

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

Thr Asp Lys Lys Arg Pro Ile His Phe Thr Pro Ala His Pro Val Tyr
530 535 540
Ser Pro Arg Tyr Phe Asp Phe Pro Lys Lys Ser Glu Thr Glu Glu Lys

545 550 555 560

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

565 570 575

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

580 585 590

Asn Val Gly Lys Lys Ala Ile Lys Ala Arg Phe Ser Tyr Ser Ala Pro

595 600 605

Arg Leu Arg Arg Asp Cys Leu Arg Ser Glu Asn Asn Glu Asn Leu Tyr

610 615 620

Lys Ala Pro Trp Leu Gln Pro Met Met Arg Ala Leu Gly Ile Asp Glu

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

Gln Phe Asn Trp Gly Gly Ile Ala Ser Leu Ser Ala Leu Arg Trp Pro

690 695 700

His Glu Lys Lys Pro Lys Asn Pro Pro Glu Gln Pro Trp Trp Gly Ile

705 710 715 720

Asp Ser Phe Ser Cys Leu Ala Val Asp Leu Gly Gln Arg Tyr Ala Gly

725 730 735

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

740 745 750

Arg Phe Ile Gly Glu Ala Cys Asp Lys Lys Trp Tyr Ala Lys Val Ser

755 760 765

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

770 775 780

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

785 790 795 800

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

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

Val Lys Gln Glu Leu Leu Pro Thr Leu Leu Thr Arg Leu Ala Asn Arg
915 920 925
Val Leu Pro Leu Arg Gly Arg Ser Trp Glu Trp Lys Lys His His Gln
930 935 940
Lys Asn Asp Gly Phe Ile Leu Asp Gln Thr Gly Lys Ala Met Pro Asn
945 950 955 960
Val Leu Ile Arg Gly Gln Arg Gly Leu Ser Met Asp Arg Ile Glu Gln
965 970 975

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

Gly Leu Lys Leu Ala Glu Pro Pro Lys Asp Lys Lys Glu Leu Asn

1040 1045 1050

Glu Thr Cys Asp Met His Gly Ala Tyr Ala Lys Val Asp Asn Pro

1055 1060 1065

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

1070 1075 1080

Gln Gly Arg Ser Pro Arg Glu Asn Ser Arg Leu Met Lys Trp Cys

1085 1090 1095

His Arg Ala Val Arg Asp Lys Leu Lys Glu Met Cys Glu Val Phe

1100 1105 1110

Phe Pro Leu Cys Glu Arg Arg Lys Ala Gly Ser Ala Trp Val Ser

1115 1120 1125

Leu Pro Pro Leu Leu Glu Thr Pro Ala Ala Tyr Ser Ser Arg Phe

1130 1135 1140

Cys Ser Arg Ser Gly Val Ala Gly Phe Arg Ala Val Glu Val Ile

1145 1150 1155

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

1160 1165 1170

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

1175 1180 1185

Thr Val Ser Glu Gln Leu Lys Ala Phe Asn Gln Asp Lys Pro Glu

1190 1195 1200

Lys Pro Arg Thr Leu Leu Val Pro Ile Ala Gly Gly Pro Ile Phe

1205 1210 1215

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

1220 1225 1230

Gln Val Val Gln Ala Asp Ile Asn Ala Ala Ile Asn Leu Gly Leu

1235 1240 1245

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

1250 1255 1260

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

1265 1270 1275

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

Ile Asn Gln Leu Gln Trp Glu Arg Cys Phe Asp Ile Asn Lys Arg
1340 1345 1350
His Ile Glu Asp Trp Ser Asn Lys Gln Lys Gln
1355 1360

<210> 488
<211> 1413
<212> PRT
<213> Unknown
<220><223> Uncultured organism
<400> 488
Met Asn Arg Ile Tyr Gln Gly Arg Val Thr Lys Val Glu Lys Leu Lys
1 5 10 15
Asn Gly Lys Ser Pro Asp Asp Arg Glu Glu Leu Lys Asp Trp Gln Thr

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

85 90 95
Trp Leu Gly Leu Ser Glu Ser Ala Ser Phe Gly Asp Ala Leu Lys Lys
100 105 110

Ile Leu Pro Pro Ala Pro Glu Asn Lys Glu Val Arg Ala Leu Ala Val
115 120 125
Ala Leu Leu Ala Glu Lys Ala Arg Thr Leu Lys Pro Gln Lys Thr Ser
130 135 140
Ala Ser Tyr Trp Gly Arg Phe Cys Asp Asp Leu Lys Lys Lys Pro Asn

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

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

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

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

355 360 365

Trp Ser Asp Gln Asn Ser Lys Ser Cys Trp Ser Asp Phe Asp Lys Cys

370 375 380

Ala Phe Glu Glu Ala Leu Lys Thr Leu Asn Gln Phe Asn Gln Lys Thr

385 390 395 400

Glu Glu Arg Glu Lys Arg Arg Ser Glu Ala Glu Ala Glu Leu Lys Tyr

405 410 415

Met Met Asp Glu Asn Pro Glu Trp Lys Pro Lys Lys Glu Thr Glu Gly

420 425 430

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

435 440 445

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

450 455 460

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

465 470 475 480

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

485 490 495

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

500 505 510

Glu His Lys Leu Asp Met Gly Tyr Thr Ala Phe Phe Leu Lys Leu Cys

515 520 525

Glu Arg Asp Tyr Trp Asp Ile Trp Arg Asp Asp Thr Glu Val Glu Val

530 535 540

Lys Lys Ile Arg Glu Lys Arg Trp Val Lys Ser Val Val Tyr Ala Ala

545 550 555 560

Ala Asp Thr Arg Glu Leu Ala Glu Glu Leu Glu Arg Leu Gln Glu Pro

565 570 575

Val Arg Tyr Thr Pro Ala Glu Pro Gln Phe Ser Arg Arg Leu Phe Met

580 585 590

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

595 600 605

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

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

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

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

850 855 860

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

865 870 875 880

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

885 890 895

Gly Ala Leu Ser Arg Tyr Lys Thr Trp Leu Arg Trp Ser Trp Arg Leu

900 905 910

Thr Glu Gln His Lys Glu Leu Trp Asp Lys Thr Leu Asp Glu Ile Arg

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

Arg Ala Leu Pro Leu Arg Glu Asn Thr Trp Arg Trp Ile Glu Asn Gly

980 985 990

Lys Asp Gly Lys Gly Lys Pro Leu His Leu Leu Val Ser Asp Gly Gln

995 1000 1005

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

1010 1015 1020

Ile Ala Arg Ile Glu Gln Leu Glu Asn Phe Arg Arg Ala Val Leu

1025 1030 1035

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

1040 1045 1050

Phe Gly Ser Ser Thr Cys Gly Glu Ser Leu Pro Asp Pro Cys Pro

1055 1060 1065

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

1070 1075 1080

Gln Thr Ala His Leu Ile Ile Ala Gln Ser Leu Gly Val Arg Leu

1085 1090 1095

Lys Gly His Ser Leu Phe Thr Glu Glu Arg Glu Lys Ala Asp Met

1100 1105 1110

His Gly Glu His Glu Val Ile Pro Gly Arg Ser Pro Val Asp Phe

1115 1120 1125

Val Val Leu Glu Asp Leu Ser Arg Tyr Thr Thr Asp Lys Ser Arg

1130 1135 1140

Ser Arg Ser Glu Asn Ser Arg Leu Met Lys Trp Cys His Arg Lys

1145 1150 1155

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

1160 1165 1170

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

1175 1180 1185

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

1190 1195 1200

Arg Pro Phe Trp Arg Lys Thr Ile Glu Lys Gln Ser Val Ala Arg

1205 1210 1215

Glu Val Phe Asp Cys Leu Asp Asn Leu Val Gly Lys Gly Leu Asn

1220 1225 1230

Gly Ile His Leu Val Leu Pro Gln Asn Gly Gly Pro Leu Phe Ile

1235 1240 1245

Ala Ala Val Lys Glu Asp Gln Pro Leu Pro Ala Ile Arg Gln Ala

1250 1255 1260

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

1265 1270 1275

Pro Ser Cys Tyr His Ala His Pro Lys Val Arg Leu Ile Lys Gly

1280 1285 1290

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

1295 1300 1305

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

1310 1315 1320

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

1325 1330 1335

Leu Lys Gly Asp Asn Thr Asn Leu Phe His Asp Pro Leu Asn Ile

1340 1345 1350

Ala Cys Tyr Gly Phe Ala Thr Ile Gln Asn Leu Gln His Pro Phe

1355 1360 1365

Leu Ala His Ala Ser Ala Val Phe Ser Arg Gln Lys Gly Ala Val

1370 1375 1380

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

1385 1390 1395

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

1400 1405 1410

<210> 489

<211> 747

<212> PRT

<213> Unknown

<220><223> Bacterium in the Phycisphaerae family

<400> 489

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

1 5 10 15

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

20 25 30

Gln Met Ile Asn Thr Tyr Leu Arg Met Gln Asn Gly Lys Phe Gly Pro

35 40 45

Asp His Lys Lys Leu Ala Gln Ile Met Leu Ser Arg Ser Asn Thr Phe

50 55 60

Ala His Gly Val Met Asp Gln Ile Thr Arg Asp Gln Pro Thr Ser Thr

65 70 75 80

Leu Asp Glu Glu Trp Thr Asp Leu Ala Arg Arg Ile His Lys Thr Thr

85 90 95

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

100 105 110

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

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

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

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

355 360 365

Val Pro Ser Trp Arg Ala Glu Thr Phe Glu Thr Glu Gly Arg Phe Pro

370 375 380

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

385 390 395 400

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

405 410 415

Lys Asn Asn Arg Ser Glu Leu Leu Phe Thr Val Phe Glu Gln Asp Cys

420 425 430

Pro Pro Arg Val Lys Trp Ala Lys Thr Lys Asn Arg Lys Cys Pro Ala

435 440 445

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

450 455 460

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

465 470 475 480

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

485 490 495

Gly Ile Ile Asn Ser Pro Ser Ile Pro Pro Leu Arg Gln Val Arg Arg

500 505 510

His Asp Tyr Asp Leu Lys Arg Lys Arg Arg Arg His Gly Lys Pro Val

515 520 525

Lys Gly Gln Arg Ser Asn Ala Asn Leu Gln Ala His Arg Thr Asn Met

530 535 540

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

545 550 555 560

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

565 570 575

Lys Phe Ser Ala Phe Asp Glu Arg Trp Met Asn Arg Gln Leu Arg Asp

580 585 590

Met Asn Arg Arg His Ile Val Glu Leu Val Ser Glu Gln Ala Pro Glu

595 600 605

Phe Gly Ile Thr Val Lys Asp Asp Ile Asn Pro Trp Met Thr Ser Arg

610 615 620

Ile Cys Ser Asn Cys Asn Leu Pro Gly Phe Arg Phe Ser Met Lys Lys
625 630 635 640

Lys Asn Pro Tyr Arg Glu Lys Leu Pro Arg Glu Lys Cys Thr Asp Phe
645 650 655

Gly Tyr Pro Val Trp Glu Pro Gly Gly His Leu Phe Arg Cys Pro His
660 665 670

Cys Asp His Arg Val Asn Ala Asp Ile Asn Ala Ala Ala Asn Leu Ala

675 680 685

Asn Lys Phe Phe Gly Leu Gly Tyr Trp Asn Asn Gly Leu Lys Tyr Asp
690 695 700

Ala Glu Thr Lys Thr Phe Thr Val His Thr Asp Lys Lys Thr Pro Pro
705 710 715 720

Leu Ile Phe Lys Pro Arg Pro Gln Phe Asp Leu Trp Ala Asp Ser Val
725 730 735

Lys Thr Arg Lys Gln Leu Gly Pro Asp Pro Phe
740 745

<210> 490
<211> 743
<212> PRT
<213> Unknown
<220><223> Member of the Planctomycetes phylum
<400> 490
Met Ser Val Arg Ser Phe Gln Ala Arg Val Glu Cys Asp Lys Gln Thr
1 5 10 15

Met Glu His Leu Trp Arg Thr His Lys Val Phe Asn Glu Arg Leu Pro
20 25 30

Glu Ile Ile Lys Ile Leu Phe Lys Met Lys Arg Gly Glu Cys Gly Gln
35 40 45

Asn Asp Lys Gln Lys Ser Leu Tyr Lys Ser Ile Ser Gln Ser Ile Leu

50 55 60

Glu Ala Asn Ala Gln Asn Ala Asp Tyr Leu Leu Asn Ser Val Ser Ile
65 70 75 80

Lys Gly Trp Lys Pro Gly Thr Ala Lys Lys Tyr Arg Asn Ala Ser Phe
85 90 95

Thr Trp Ala Asp Asp Ala Ala Lys Leu Ser Ser Gln Gly Ile His Val
100 105 110

Tyr Asp Lys Lys Gln Val Leu Gly Asp Leu Pro Gly Met Met Ser Gln
115 120 125

Met Val Cys Arg Gln Ser Val Glu Ala Ile Ser Gly His Ile Glu Leu
130 135 140

Thr Lys Lys Trp Glu Lys Glu His Asn Glu Trp Leu Lys Glu Lys Glu
145 150 155 160

Lys Trp Glu Ser Glu Asp Glu His Lys Lys Tyr Leu Asp Leu Arg Glu
165 170 175

Lys Phe Glu Gln Phe Glu Gln Ser Ile Gly Gly Lys Ile Thr Lys Arg
180 185 190

Arg Gly Arg Trp His Leu Tyr Leu Lys Trp Leu Ser Asp Asn Pro Asp
195 200 205

Phe Ala Ala Trp Arg Gly Asn Lys Ala Val Ile Asn Pro Leu Ser Glu
210 215 220

Lys Ala Gln Ile Arg Ile Asn Lys Ala Lys Pro Asn Lys Lys Asn Ser
225 230 235 240

Val Glu Arg Asp Glu Phe Phe Lys Ala Asn Pro Glu Met Lys Ala Leu
245 250 255

Asp Asn Leu His Gly Tyr Tyr Glu Arg Asn Phe Val Arg Arg Arg Lys
260 265 270

Thr Lys Lys Asn Pro Asp Gly Phe Asp His Lys Pro Thr Phe Thr Leu
275 280 285

Pro His Pro Thr Ile His Pro Arg Trp Phe Val Phe Asn Lys Pro Lys
290 295 300

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

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

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

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

545 550 555 560

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

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

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

<210> 491
<211> 1119
<212> PRT
<213> Unknown

<220><223> Member of the Spirochaetes phylum
<400> 491

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

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

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

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

245 250 255

Phe Ala Lys Lys Gly Lys Phe Tyr Ile Thr Gly Arg Met Leu Lys Gly

260 265 270

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

275 280 285

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

290 295 300

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

305 310 315 320

Gln Cys Leu Trp Asn Gly Asp Ile Asp Tyr Leu Lys Ile Lys Arg Asp

325 330 335

Ile Asn Lys Asn Gln Ile Trp Leu Asp Arg Pro Glu Met Pro Arg Phe

340 345 350

Thr Met Pro Asp Phe Lys Lys His Pro Leu Trp Tyr Arg Tyr Glu Asp

355 360 365

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

370 375 380

Glu Asn Tyr Ile Thr Ile Pro Leu Ile Thr Glu Arg Asn Asn Glu Tyr

385 390 395 400

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

405 410 415

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

420 425 430

Ile Asp Ser Asn Asp Glu Glu Glu Asp Lys Lys Asp Gln Lys Lys Ser

435 440 445

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

450 455 460

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

465 470 475 480

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

485 490 495

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

500 505 510

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

515 520 525

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

530 535 540

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

545 550 555 560

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

565 570 575

Gln Arg Phe Phe Ala Ala Val Ser Cys Phe Glu Ile Met Ser Glu Ile

580 585 590

Asp Asn Asn Lys Leu Phe Phe Asn Leu Asn Asp Gln Asn His Lys Ile

595 600 605

Ile Arg Ile Asn Asp Lys Asn Tyr Tyr Ala Lys His Ile Tyr Ser Lys

610 615 620

Thr Ile Lys Leu Ser Gly Glu Asp Asp Asp Leu Tyr Lys Glu Arg Lys

625 630 635 640

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

645 650 655

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

660 665 670

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

675 680 685

Thr Thr Lys Arg Tyr Val Lys Asn Thr Tyr Asn Asp Gly Ile Ile Asp

690 695 700

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

705 710 715 720

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

725 730 735

Lys Glu Phe Ile Arg Glu Cys Arg Lys Pro Ile Phe Arg Ser Gly Gly

740 745 750

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

755 760 765

Arg Lys Trp Val Ala Arg Thr Gln Lys Ser Ala Glu Ser Ile Val Leu

770 775 780

Thr Pro Lys Phe Gly Tyr Lys Leu Lys Glu His Ile Asn Glu Leu Lys

785 790 795 800

Asp Asn Arg Val Lys Gln Gly Val Asn Tyr Ile Leu Met Thr Ala Leu

805 810 815

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

820 825 830

Lys Glu Asp Trp Val Lys Lys Asn Arg Ala Cys Gln Ile Ile Leu Met

835 840 845

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

850 855 860

Asn Ser Lys Leu Arg Met Trp Ser His Arg Gln Ile Phe Asn Phe Leu

865 870 875 880

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

885 890 895

Pro Tyr Thr Ser Lys Cys Leu Ser Asp Asn Asn Ala Pro Gly Ile Arg

900 905 910

Cys His Gln Val Thr Lys Lys Asp Leu Ile Asp Asn Ser Trp Phe Leu

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Glu Thr Phe His Leu Lys Asn Glu Pro Asn Tyr Leu Lys Asn Tyr

1010 1015 1020

Tyr Ser Ile Leu Asn Phe Val Pro Thr Asn Glu Glu Leu Thr Phe

1025 1030 1035

Phe Lys Val Glu Glu Asn Lys Asp Ile Lys Pro Thr Lys Arg Ile

1040 1045 1050

Lys Met Asp Lys His Glu Lys Glu Ser Thr Asp Glu Gly Asp Asp

1055 1060 1065

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

1070 1075 1080

Phe Phe Asp Lys Ser Leu Trp Val Asp Gly Lys Ile Phe Trp Ser

1085 1090 1095

Val Val Lys Asn Lys Met Thr Lys Leu Leu Arg Glu Arg Asn Asn

1100 1105 1110

Lys Lys Asn Gly Ser Lys

1115

<210> 492

<211> 1447

<212> PRT

<213> Unknown

<220><223> Member of the phylum Verrucomicrobiaceae

<400> 492

Met Pro Leu Ser Arg Ile Tyr Gln Gly Arg Thr Asn Ser Leu Ile Ile

1 5 10 15

Leu Thr Pro Thr Pro Gln Glu Pro Trp Asp His Lys Ala Leu Ala Arg

20 25 30

Phe Asp Ser Pro Leu Trp Arg His His Ala Leu Phe Gln Asp Ala Val

35 40 45

Asn Tyr Tyr Gln Leu Cys Leu Val Ala Leu Ala Ser Ser Asp Gly Thr

50 55 60

Arg Pro Leu Ser Lys Leu His Glu Gln Met Lys Ala Ser Trp Asp Glu

65 70 75 80

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

85 90 95

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

100 105 110

Leu Glu Gly Asn Glu Ala Pro Glu Arg Ala Arg Glu Leu Ala Gly Glu

115 120 125

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

130 135 140

Tyr Trp Pro Arg Phe Cys Asp Pro Lys Ala Asn Pro Thr Tyr Asp Tyr

145 150 155 160

Ser Ala Thr Ala Arg Ala Ser Ala Ser Gly Leu Thr Lys Leu Ala Ala

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

Ala Arg Phe Pro Asp Gly Leu Ala Leu Trp Gln Ala Leu Pro Glu Lys

245 250 255

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

260 265 270

Ser Pro Asp Leu Thr Phe Ala Thr Leu Leu Phe Gln His Phe Pro Ser

275 280 285

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

290 295 300

Val Lys Ala Pro Lys Val Val Glu Lys Val Ser Ala Arg Arg Lys Ala

305 310 315 320

Asn Ala Val Thr Gln Ala Val Val Ile Glu Glu Pro Glu Ile Asp Phe

325 330 335

Ala Glu Leu Gly Asp Asp Pro Ile Lys Leu Ala Arg Gly Glu Arg Gly

340 345 350

Phe Val Phe Pro Ala Phe Thr Ser Leu Ser Phe Trp Ala Val Pro Gly

355 360 365

Pro His Val Pro Val Trp Lys Glu Phe Asp Ile Ala Ala Phe Lys Glu

370 375 380

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

385 390 395 400

Ala Leu Leu Ala Glu Ala Gln Arg Arg Leu Asp Tyr Met Asp Glu Lys

405 410 415

Thr His Asp Trp Lys Thr Gly Asp Ser Asp Glu Pro Gly His Ile Pro

420 425 430

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

435 440 445

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

450 455 460

Lys Gly Val Tyr Thr Gly Gly Leu Arg Gly Phe Tyr Ala Ile Lys Lys

465 470 475 480

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

485 490 495

Pro Thr Glu Glu Glu Leu Ile Ser Ile Val Thr Asp Tyr Gln Arg Asp

500 505 510

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

515 520 525

Pro Arg Phe Trp Pro Leu Trp Gln Pro Leu Thr Asp Glu Gln Glu Ala

530 535 540

Glu Arg Ile Lys Ala Gly Arg Ala Lys Asp Met Ile Ser Ala Tyr Arg

545 550 555 560

Val Trp Leu Glu Leu Gln Glu Asp Val Val Arg Leu Ala Gln Pro Ile
565 570 575
Arg Phe Thr Pro Ala His Ala Glu Asn Ser Arg Arg Leu Phe Met Phe

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

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

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

770 775 780
Ser Arg Ser Asp Lys Ala Phe Ser Arg Asn Arg Gln Pro Phe Ile Glu
785 790 795 800
Leu Gly Glu Ala Gly Gly Lys Leu Trp Ser Thr Ala Leu Leu Gly Leu

805 810 815

Gly Met Leu Arg Leu Pro Gly Glu Asp Ala Arg Thr Gly Ala Leu Asp

820 825 830

Asp Gln Gly Lys Arg Ala Val Glu Phe His Gly Lys Ala Gly Arg Asn

835 840 845

Ala Leu Glu Ala Glu Trp Gln Glu Ala Arg Glu Met Ala Leu Leu Phe

850 855 860

Gly Gly Glu Glu Ala Lys Ser Arg Leu Gly Pro Gly Phe Asp His Leu

865 870 875 880

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

885 890 895

Gln Ser Arg Leu Ala Arg Phe His Arg Trp Ser Cys Arg Ile His Glu

900 905 910

Lys Pro Glu Ala Thr Gly Asp Asp Val Ile Asp Tyr Gly Gln Val Asp

915 920 925

Glu Leu Leu Thr Lys Thr Ala Glu Ala Met Leu Glu Asn Leu Lys Ala

930 935 940

Leu Tyr Thr Asn Ala Gly Gly Ile Leu Asp Ser Lys Ser Lys Gln Pro

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Ser Leu Glu Arg Leu Ala Asn Arg Glu Leu Pro Leu Arg His Arg

1010 1015 1020

Glu Trp Val Trp Asn Glu Ala Phe Glu Lys Leu Glu Gln Gly Asn

1025 1030 1035

Phe Lys Lys Glu Glu Asn Pro Lys Trp Ile Arg Gly Gln Arg Gly

1040 1045 1050

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

1055 1060 1065

Phe Met Ser Leu Arg Arg Gln Met Ser Leu Ile Pro Gly Glu Gln

1070 1075 1080

Val Lys Gln Gly Val Glu Asp Lys Gly Gln Arg Gln Pro Glu Pro

1085 1090 1095

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

1100 1105 1110

Val Asn Gln Thr Ala His Leu Ile Leu Ala Gln Ala Leu Gly Leu

1115 1120 1125

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

1130 1135 1140

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

1145 1150 1155

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

1160 1165 1170

Gly Arg Ala Pro Ser Glu Asn Gly Arg Leu Met Lys Trp Cys His

1175 1180 1185

Arg Ala Val Leu Ala Lys Leu Lys Gln Met Cys Glu Pro Phe Gly

1190 1195 1200

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

1205 1210 1215

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

1220 1225 1230

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

1235 1240 1245

Glu Lys Asp Lys Ser Ser Lys Asp Ala Glu Ala Ala Ala Met Leu

1250 1255 1260

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

1265 1270 1275

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

1280 1285 1290

Ala Gly Gly Pro Leu Phe Ile Pro Met Val Gly Gly Gly Pro Arg

1295 1300 1305

Gln Ala Asp Met Asn Ala Ala Ile Asn Leu Gly Leu Arg Ala Ile

1310 1315 1320

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

1325 1330 1335

Leu Lys Asp Gly Lys His Gln Ala Met Leu Gly Asn Lys Leu Glu

1340 1345 1350

Lys Ala Ala Ala Leu Thr Leu Glu Pro Pro Lys Glu Pro Thr Lys

1355 1360 1365

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

1370 1375 1380

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

1385 1390 1395

Lys Leu Arg Leu Ser Gly Gly Met Ser Leu Trp Lys Ala Ile Lys

1400 1405 1410

Asp Gly Ala Trp Gln Arg Val Lys Lys Ile Asn Asp Ala Arg Ile

1415 1420 1425

Ala Lys Trp Lys Asn Asn Pro Pro Pro Glu Pro Asp Pro Asp Asp

1430 1435 1440

Glu Ile Gln Phe

1445

<210> 493

<211

> 1147

<212> PRT

<213> Alicyclobacillus kakegawensis

<400> 493

Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Ser Glu Cys Pro

1 5 10 15

Asp Ile Leu Ala Gly Met Trp Gln Leu His Arg Ala Thr Asn Ala Gly

20 25 30

Val Arg Tyr Tyr Thr Glu Trp Val Ser Leu Met Arg Gln Glu Ile Leu
35 40 45
Tyr Ser Arg Gly Pro Asp Gly Gly Gln Gln Cys Tyr Met Thr Ala Glu

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

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

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

245 250 255
Glu Gln Ser His Leu Val Lys Leu Ala Arg Ala Leu Glu Ala Asp Met
260 265 270
Arg Ala Ala Ser Gln Gly Phe Glu Ala Lys Arg Gly Thr Ala His Gln

275 280 285

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

290 295 300

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

305 310 315 320

Arg Gln Val Gln Ala Glu Lys Arg Arg Asp Phe Gly Ser His Asp Leu

325 330 335

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

340 345 350

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

355 360 365

Leu Glu Lys Ala Arg Gln Phe Ala Thr Phe Thr Leu Pro Asp Ala Cys

370 375 380

Val Asn Pro Ile Trp Thr Arg Phe Glu Ser Ser Gln Gly Ser Asn Leu

385 390 395 400

His Lys Tyr Glu Phe Leu Phe Asp His Leu Gly Pro Gly Arg His Ala

405 410 415

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

420 425 430

Arg Asp Ser Val Val Val Pro Val Ala Pro Ser Gly Gln Leu Asp Lys

435 440 445

Leu Val Leu Arg Glu Glu Glu Lys Ser Ser Val Ala Leu His Leu His

450 455 460

Asp Thr Ala Arg Pro Asp Gly Phe Met Ala Glu Trp Ala Gly Ala Lys

465 470 475 480

Leu Gln Tyr Glu Arg Ser Thr Leu Ala Arg Lys Ala Arg Arg Asp Lys

485 490 495

Gln Gly Met Arg Ser Trp Arg Arg Gln Pro Ser Met Leu Met Ser Ala

500 505 510

Ala Gln Met Leu Glu Asp Ala Lys Gln Ala Gly Asp Val Tyr Leu Asn

515 520 525

Ile Ser Val Arg Val Lys Ser Pro Ser Glu Val Arg Gly Gln Arg Arg
530 535 540
Pro Pro Tyr Ala Ala Leu Phe Arg Ile Asp Asp Lys Gln Arg Arg Val
545 550 555 560
Thr Val Asn Tyr Asn Lys Leu Ser Ala Tyr Leu Glu Glu His Pro Asp

565 570 575
Lys Gln Ile Pro Gly Ala Pro Gly Leu Leu Ser Gly Leu Arg Val Met
580 585 590
Ser Val Asp Leu Gly Leu Arg Thr Ser Ala Ser Ile Ser Val Phe Arg
595 600 605
Val Ala Lys Lys Glu Glu Val Glu Ala Leu Gly Asp Gly Arg Pro Pro
610 615 620
His Tyr Tyr Pro Ile His Gly Thr Asp Asp Leu Val Ala Val His Glu

625 630 635 640
Arg Ser His Leu Ile Gln Met Pro Gly Glu Thr Glu Thr Lys Gln Leu
645 650 655
Arg Lys Leu Arg Glu Glu Arg Gln Ala Val Leu Arg Pro Leu Phe Ala
660 665 670
Gln Leu Ala Leu Leu Arg Leu Leu Val Arg Cys Gly Ala Ala Asp Glu
675 680 685
Arg Ile Arg Thr Arg Ser Trp Gln Arg Leu Thr Lys Gln Gly Arg Glu

690 695 700
Phe Thr Lys Arg Leu Thr Pro Ser Trp Arg Glu Ala Leu Glu Leu Glu
705 710 715 720
Leu Thr Arg Leu Glu Ala Tyr Cys Gly Arg Val Pro Asp Asp Glu Trp
725 730 735
Ser Arg Ile Val Asp Arg Thr Val Ile Ala Leu Trp Arg Arg Met Gly
740 745 750
Lys Gln Val Arg Asp Trp Arg Lys Gln Val Lys Ser Gly Ala Lys Val

755 760 765
Lys Val Lys Gly Tyr Gln Leu Asp Val Val Gly Gly Asn Ser Leu Ala

770 775 780

Gln Ile Asp Tyr Leu Glu Gln Gln Tyr Lys Phe Leu Arg Arg Trp Ser

785 790 795 800

Phe Phe Ala Arg Ala Ser Gly Leu Val Val Arg Ala Asp Arg Glu Ser

805 810 815

His Phe Ala Val Ala Leu Arg Gln His Ile Glu Asn Ala Lys Arg Asp

820 825 830

Arg Leu Lys Lys Leu Ala Asp Arg Ile Leu Met Glu Ala Leu Gly Tyr

835 840 845

Val Tyr Glu Ala Ser Gly Pro Arg Glu Gly Gln Trp Thr Ala Gln His

850 855 860

Pro Pro Cys Gln Leu Ile Ile Leu Glu Glu Leu Ser Ala Tyr Arg Phe

865 870 875 880

Ser Asp Asp Arg Pro Pro Ser Glu Asn Ser Lys Leu Met Ala Trp Gly

885 890 895

His Arg Gly Ile Leu Glu Glu Leu Val Asn Gln Ala Gln Val His Asp

900 905 910

Val Leu Val Gly Thr Val Tyr Ala Ala Phe Ser Ser Arg Phe Asp Ala

915 920 925

Arg Thr Gly Ala Pro Gly Val Arg Cys Arg Arg Val Pro Ala Arg Phe

930 935 940

Val Gly Ala Thr Val Asp Asp Ser Leu Pro Leu Trp Leu Thr Glu Phe

945 950 955 960

Leu Asp Lys His Arg Leu Asp Lys Asn Leu Leu Arg Pro Asp Asp Val

965 970 975

Ile Pro Thr Gly Glu Gly Glu Phe Leu Val Ser Pro Cys Gly Glu Glu

980 985 990

Ala Ala Arg Val Arg Gln Val His Ala Asp Ile Asn Ala Ala Gln Asn

995 1000 1005

Leu Gln Arg Arg Leu Trp Gln Asn Phe Asp Ile Thr Glu Leu Arg

1010 1015 1020

Leu Arg Cys Asp Val Lys Met Gly Gly Glu Gly Thr Val Leu Val
1025 1030 1035
Pro Arg Val Asn Asn Ala Arg Ala Lys Gln Leu Phe Gly Lys Lys
1040 1045 1050
Val Leu Val Ser Gln Asp Gly Val Thr Phe Phe Glu Arg Ser Gln
1055 1060 1065
Thr Gly Gly Lys Pro His Ser Glu Lys Gln Thr Asp Leu Thr Asp

1070 1075 1080
Lys Glu Leu Glu Leu Ile Ala Glu Ala Asp Glu Ala Arg Ala Lys
1085 1090 1095
Ser Val Val Leu Phe Arg Asp Pro Ser Gly His Ile Gly Lys Gly
1100 1105 1110
His Trp Ile Arg Gln Arg Glu Phe Trp Ser Leu Val Lys Gln Arg
1115 1120 1125
Ile Glu Ser His Thr Ala Glu Arg Ile Arg Val Arg Gly Val Gly

1130 1135 1140
Ser Ser Leu Asp
1145

<210> 494
<211> 1112
<212> PRT
<213> Bacillus species
<400> 494
Met Ala Ile Arg Ser Ile Lys Leu Lys Met Lys Thr Asn Ser Gly Thr
1 5 10 15
Asp Ser Ile Tyr Leu Arg Lys Ala Leu Trp Arg Thr His Gln Leu Ile
20 25 30
Asn Glu Gly Ile Ala Tyr Tyr Met Asn Leu Leu Thr Leu Tyr Arg Gln

35 40 45
Glu Ala Ile Gly Asp Lys Thr Lys Glu Ala Tyr Gln Ala Glu Leu Ile
50 55 60
Asn Ile Ile Arg Asn Gln Gln Arg Asn Asn Gly Ser Ser Glu Glu His

65 70 75 80

Gly Ser Asp Gln Glu Ile Leu Ala Leu Leu Arg Gln Leu Tyr Glu Leu

85 90 95

Ile Ile Pro Ser Ser Ile Gly Glu Ser Gly Asp Ala Asn Gln Leu Gly

100 105 110

Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn Ser Gln Ser Gly Lys

115 120 125

Gly Thr Ser Asn Ala Gly Arg Lys Pro Arg Trp Lys Arg Leu Lys Glu

130 135 140

Glu Gly Asn Pro Asp Trp Glu Leu Glu Lys Lys Lys Asp Glu Glu Arg

145 150 155 160

Lys Ala Lys Asp Pro Thr Val Lys Ile Phe Asp Asn Leu Asn Lys Tyr

165 170 175

Gly Leu Leu Pro Leu Phe Pro Leu Phe Thr Asn Ile Gln Lys Asp Ile

180 185 190

Glu Trp Leu Pro Leu Gly Lys Arg Gln Ser Val Arg Lys Trp Asp Lys

195 200 205

Asp Met Phe Ile Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu Ser Trp

210 215 220

Asn Arg Arg Val Ala Asp Glu Tyr Lys Gln Leu Lys Glu Lys Thr Glu

225 230 235 240

Ser Tyr Tyr Lys Glu His Leu Thr Gly Gly Glu Glu Trp Ile Glu Lys

245 250 255

Ile Arg Lys Phe Glu Lys Glu Arg Asn Met Glu Leu Glu Lys Asn Ala

260 265 270

Phe Ala Pro Asn Asp Gly Tyr Phe Ile Thr Ser Arg Gln Ile Arg Gly

275 280 285

Trp Asp Arg Val Tyr Glu Lys Trp Ser Lys Leu Pro Glu Ser Ala Ser

290 295 300

Pro Glu Glu Leu Trp Lys Val Val Ala Glu Gln Gln Asn Lys Met Ser

305 310 315 320

Glu Gly Phe Gly Asp Pro Lys Val Phe Ser Phe Leu Ala Asn Arg Glu
325 330 335
Asn Arg Asp Ile Trp Arg Gly His Ser Glu Arg Ile Tyr His Ile Ala
340 345 350
Ala Tyr Asn Gly Leu Gln Lys Lys Leu Ser Arg Thr Lys Glu Gln Ala

355 360 365
Thr Phe Thr Leu Pro Asp Ala Ile Glu His Pro Leu Trp Ile Arg Tyr
370 375 380
Glu Ser Pro Gly Gly Thr Asn Leu Asn Leu Phe Lys Leu Glu Glu Lys
385 390 395 400
Gln Lys Lys Asn Tyr Tyr Val Thr Leu Ser Lys Ile Ile Trp Pro Ser
405 410 415
Glu Glu Lys Trp Ile Glu Lys Glu Asn Ile Glu Ile Pro Leu Ala Pro

420 425 430
Ser Ile Gln Phe Asn Arg Gln Ile Lys Leu Lys Gln His Val Lys Gly
435 440 445
Lys Gln Glu Ile Ser Phe Ser Asp Tyr Ser Ser Arg Ile Ser Leu Asp
450 455 460
Gly Val Leu Gly Gly Ser Arg Ile Gln Phe Asn Arg Lys Tyr Ile Lys
465 470 475 480
Asn His Lys Glu Leu Leu Gly Glu Gly Asp Ile Gly Pro Val Phe Phe

485 490 495
Asn Leu Val Val Asp Val Ala Pro Leu Gln Glu Thr Arg Asn Gly Arg
500 505 510
Leu Gln Ser Pro Ile Gly Lys Ala Leu Lys Val Ile Ser Ser Asp Phe
515 520 525
Ser Lys Val Ile Asp Tyr Lys Pro Lys Glu Leu Met Asp Trp Met Asn
530 535 540
Thr Gly Ser Ala Ser Asn Ser Phe Gly Val Ala Ser Leu Leu Glu Gly

545 550 555 560
Met Arg Val Met Ser Ile Asp Met Gly Gln Arg Thr Ser Ala Ser Val

565 570 575

Ser Ile Phe Glu Val Val Lys Glu Leu Pro Lys Asp Gln Glu Gln Lys

580 585 590

Leu Phe Tyr Ser Ile Asn Asp Thr Glu Leu Phe Ala Ile His Lys Arg

595 600 605

Ser Phe Leu Leu Asn Leu Pro Gly Glu Val Val Thr Lys Asn Asn Lys

610 615 620

Gln Gln Arg Gln Glu Arg Arg Lys Lys Arg Gln Phe Val Arg Ser Gln

625 630 635 640

Ile Arg Met Leu Ala Asn Val Leu Arg Leu Glu Thr Lys Lys Thr Pro

645 650 655

Asp Glu Arg Lys Lys Ala Ile His Lys Leu Met Glu Ile Val Gln Ser

660 665 670

Tyr Asp Ser Trp Thr Ala Ser Gln Lys Glu Val Trp Glu Lys Glu Leu

675 680 685

Asn Leu Leu Thr Asn Met Ala Ala Phe Asn Asp Glu Ile Trp Lys Glu

690 695 700

Ser Leu Val Glu Leu His His Arg Ile Glu Pro Tyr Val Gly Gln Ile

705 710 715 720

Val Ser Lys Trp Arg Lys Gly Leu Ser Glu Gly Arg Lys Asn Leu Ala

725 730 735

Gly Ile Ser Met Trp Asn Ile Asp Glu Leu Glu Asp Thr Arg Arg Leu

740 745 750

Leu Ile Ser Trp Ser Lys Arg Ser Arg Thr Pro Gly Glu Ala Asn Arg

755 760 765

Ile Glu Thr Asp Glu Pro Phe Gly Ser Ser Leu Leu Gln His Ile Gln

770 775 780

Asn Val Lys Asp Asp Arg Leu Lys Gln Met Ala Asn Leu Ile Ile Met

785 790 795 800

Thr Ala Leu Gly Phe Lys Tyr Asp Lys Glu Glu Lys Asp Arg Tyr Lys

805 810 815

Arg Trp Lys Glu Thr Tyr Pro Ala Cys Gln Ile Ile Leu Phe Glu Asn
820 825 830
Leu Asn Arg Tyr Leu Phe Asn Leu Asp Arg Ser Arg Arg Glu Asn Ser
835 840 845
Arg Leu Met Lys Trp Ala His Arg Ser Ile Pro Arg Thr Val Ser Met
850 855 860
Gln Gly Glu Met Phe Gly Leu Gln Val Gly Asp Val Arg Ser Glu Tyr

865 870 875 880
Ser Ser Arg Phe His Ala Lys Thr Gly Ala Pro Gly Ile Arg Cys His
885 890 895
Ala Leu Thr Glu Glu Asp Leu Lys Ala Gly Ser Asn Thr Leu Lys Arg
900 905 910
Leu Ile Glu Asp Gly Phe Ile Asn Glu Ser Glu Leu Ala Tyr Leu Lys
915 920 925
Lys Gly Asp Ile Ile Pro Ser Gln Gly Gly Glu Leu Phe Val Thr Leu

930 935 940
Ser Lys Arg Tyr Lys Lys Asp Ser Asp Asn Asn Glu Leu Thr Val Ile
945 950 955 960
His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Gln
965 970 975
Gln Asn Ser Glu Val Tyr Arg Val Pro Cys Gln Leu Ala Arg Met Gly
980 985 990
Glu Asp Lys Leu Tyr Ile Pro Lys Ser Gln Thr Glu Thr Ile Lys Lys

995 1000 1005
Tyr Phe Gly Lys Gly Ser Phe Val Lys Asn Asn Thr Glu Gln Glu
1010 1015 1020
Val Tyr Lys Trp Glu Lys Ser Glu Lys Met Lys Ile Lys Thr Asp
1025 1030 1035
Thr Thr Phe Asp Leu Gln Asp Leu Asp Gly Phe Glu Asp Ile Ser
1040 1045 1050
Lys Thr Ile Glu Leu Ala Gln Glu Gln Gln Lys Lys Tyr Leu Thr

1055 1060 1065

Met Phe Arg Asp Pro Ser Gly Tyr Phe Phe Asn Asn Glu Thr Trp

1070 1075 1080

Arg Pro Gln Lys Glu Tyr Trp Ser Ile Val Asn Asn Ile Ile Lys

1085 1090 1095

Ser Cys Leu Lys Lys Lys Ile Leu Ser Asn Lys Val Glu Leu

1100 1105 1110

<210> 495

<211> 1489

<212> PRT

<213> Desulfatirhabdium butyrativorans

<400> 495

Met Pro Leu Ser Asn Asn Pro Pro Val Thr Gln Arg Ala Tyr Thr Leu

1 5 10 15

Arg Leu Arg Gly Ala Asp Pro Ser Asp Leu Ser Trp Arg Glu Ala Leu

20 25 30

Trp His Thr His Glu Ala Val Asn Lys Gly Ala Lys Val Phe Gly Asp

35 40 45

Trp Leu Leu Thr Leu Arg Gly Gly Leu Asp His Thr Leu Ala Asp Thr

50 55 60

Lys Val Lys Gly Gly Lys Gly Lys Pro Asp Arg Asp Pro Thr Pro Glu

65 70 75 80

Glu Arg Lys Ala Arg Arg Ile Leu Leu Ala Leu Ser Trp Leu Ser Val

85 90 95

Glu Ser Lys Leu Gly Ala Pro Ser Ser Tyr Ile Val Ala Ser Gly Asp

100 105 110

Glu Pro Ala Lys Asp Arg Asn Asp Asn Val Val Ser Ala Leu Glu Glu

115 120 125

Ile Leu Gln Ser Arg Lys Val Ala Lys Ser Glu Ile Asp Asp Trp Lys

130 135 140

Arg Asp Cys Ser Ala Ser Leu Ser Ala Ala Ile Arg Asp Asp Ala Val

145 150 155 160

Trp Val Asn Arg Ser Lys Val Phe Asp Glu Ala Val Lys Ser Val Gly
165 170 175
Ser Ser Leu Thr Arg Glu Glu Ala Trp Asp Met Leu Glu Arg Phe Phe
180 185 190

Gly Ser Arg Asp Ala Tyr Leu Thr Pro Met Lys Asp Pro Glu Asp Lys
195 200 205
Ser Ser Glu Thr Glu Gln Glu Asp Lys Ala Lys Asp Leu Val Gln Lys
210 215 220
Ala Gly Gln Trp Leu Ser Ser Arg Tyr Gly Thr Ser Glu Gly Ala Asp
225 230 235 240
Phe Cys Arg Met Ser Asp Ile Tyr Gly Lys Ile Ala Ala Trp Ala Asp
245 250 255

Asn Ala Ser Gln Gly Gly Ser Ser Thr Val Asp Asp Leu Val Ser Glu
260 265 270
Leu Arg Gln His Phe Asp Thr Lys Glu Ser Lys Ala Thr Asn Gly Leu
275 280 285
Asp Trp Ile Ile Gly Leu Ser Ser Tyr Thr Gly His Thr Pro Asn Pro
290 295 300
Val His Glu Leu Leu Arg Gln Asn Thr Ser Leu Asn Lys Ser His Leu
305 310 315 320

Asp Asp Leu Lys Lys Lys Ala Asn Thr Arg Ala Glu Ser Cys Lys Ser
325 330 335
Lys Ile Gly Ser Lys Gly Gln Arg Pro Tyr Ser Asp Ala Ile Leu Asn
340 345 350
Asp Val Glu Ser Val Cys Gly Phe Thr Tyr Arg Val Asp Lys Asp Gly
355 360 365
Gln Pro Val Ser Val Ala Asp Tyr Ser Lys Tyr Asp Val Asp Tyr Lys
370 375 380

Trp Gly Thr Ala Arg His Tyr Ile Phe Ala Val Met Leu Asp His Ala
385 390 395 400
Ala Arg Arg Ile Ser Leu Ala His Lys Trp Ile Lys Arg Ala Glu Ala

405 410 415

Glu Arg His Lys Phe Glu Glu Asp Ala Lys Arg Ile Ala Asn Val Pro

420 425 430

Ala Arg Ala Arg Glu Trp Leu Asp Ser Phe Cys Lys Glu Arg Ser Val

435 440 445

Thr Ser Gly Ala Val Glu Pro Tyr Arg Ile Arg Arg Arg Ala Val Asp

450 455 460

Gly Trp Lys Glu Val Val Ala Ala Trp Ser Lys Ser Asp Cys Lys Ser

465 470 475 480

Thr Glu Asp Arg Ile Ala Ala Ala Arg Ala Leu Gln Asp Asp Ser Glu

485 490 495

Ile Asp Lys Phe Gly Asp Ile Gln Leu Phe Glu Ala Leu Ala Glu Asp

500 505 510

Asp Ala Leu Cys Val Trp His Lys Asp Gly Glu Ala Thr Asn Glu Pro

515 520 525

Asp Phe Gln Pro Leu Ile Asp Tyr Ser Leu Ala Ile Glu Ala Glu Phe

530 535 540

Lys Lys Arg Gln Phe Lys Val Pro Ala Tyr Arg His Pro Asp Glu Leu

545 550 555 560

Leu His Pro Val Phe Cys Asp Phe Gly Lys Ser Arg Trp Lys Ile Asn

565 570 575

Tyr Asp Val His Lys Asn Val Gln Ala Pro Phe Tyr Arg Gly Leu Cys

580 585 590

Leu Thr Leu Trp Thr Gly Ser Glu Ile Lys Pro Val Pro Leu Cys Trp

595 600 605

Gln Ser Lys Arg Leu Thr Arg Asp Leu Ala Leu Gly Asn Asn His Arg

610 615 620

Asn Asp Ala Ala Ser Ala Val Thr Arg Ala Asp Arg Leu Gly Arg Ala

625 630 635 640

Ala Ser Asn Val Thr Lys Ser Asp Met Val Asn Ile Thr Gly Leu Phe

645 650 655

Glu Gln Ala Asp Trp Asn Gly Arg Leu Gln Ala Pro Arg Gln Gln Leu
660 665 670
Glu Ala Ile Ala Val Val Arg Asp Asn Pro Arg Leu Ser Glu Gln Glu
675 680 685
Arg Asn Leu Arg Met Cys Gly Met Ile Glu His Ile Arg Trp Leu Val
690 695 700

Thr Phe Ser Val Lys Leu Gln Pro Gln Gly Pro Trp Cys Ala Tyr Ala
705 710 715 720
Glu Gln His Gly Leu Asn Thr Asn Pro Gln Tyr Trp Pro His Ala Asp
725 730 735
Thr Asn Arg Asp Arg Lys Val His Ala Arg Leu Ile Leu Pro Arg Leu
740 745 750
Pro Gly Leu Arg Val Leu Ser Val Asp Leu Gly His Arg Tyr Ala Ala
755 760 765

Ala Cys Ala Val Trp Glu Ala Val Asn Thr Glu Thr Val Lys Glu Ala
770 775 780
Cys Gln Asn Val Gly Arg Asp Met Pro Lys Glu His Asp Leu Tyr Leu
785 790 795 800
His Ile Lys Val Lys Lys Gln Gly Ile Gly Lys Gln Thr Glu Val Asp
805 810 815
Lys Thr Thr Ile Tyr Arg Arg Ile Gly Ala Asp Thr Leu Pro Asp Gly
820 825 830

Arg Pro His Pro Ala Pro Trp Ala Arg Leu Asp Arg Gln Phe Leu Ile
835 840 845
Lys Leu Gln Gly Glu Glu Lys Asp Ala Arg Glu Ala Ser Asn Glu Glu
850 855 860
Ile Trp Ala Leu His Gln Met Glu Cys Lys Leu Asp Arg Thr Lys Pro
865 870 875 880
Leu Ile Asp Arg Leu Ile Ala Ser Gly Trp Gly Leu Leu Lys Arg Gln
885 890 895

Met Ala Arg Leu Asp Ala Leu Lys Glu Leu Gly Trp Ile Pro Ala Pro

900 905 910

Asp Ser Ser Glu Asn Leu Ser Arg Glu Asp Gly Glu Ala Lys Asp Tyr

915 920 925

Arg Glu Ser Leu Ala Val Asp Asp Leu Met Phe Ser Ala Val Arg Thr

930 935 940

Leu Arg Leu Ala Leu Gln Arg His Gly Asn Arg Ala Arg Ile Ala Tyr

945 950 955 960

Tyr Leu Ile Ser Glu Val Lys Ile Arg Pro Gly Gly Ile Gln Glu Lys

965 970 975

Leu Asp Glu Asn Gly Arg Ile Asp Leu Leu Gln Asp Ala Leu Ala Leu

980 985 990

Trp His Glu Leu Phe Ser Ser Pro Gly Trp Arg Asp Glu Ala Ala Lys

995 1000 1005

Gln Leu Trp Asp Ser Arg Ile Ala Thr Leu Ala Gly Tyr Lys Ala

1010 1015 1020

Pro Glu Glu Asn Gly Asp Asn Val Ser Asp Val Ala Tyr Arg Lys

1025 1030 1035

Lys Gln Gln Val Tyr Arg Glu Gln Leu Arg Asn Val Ala Lys Thr

1040 1045 1050

Leu Ser Gly Asp Val Ile Thr Cys Lys Glu Leu Ser Asp Ala Trp

1055 1060 1065

Lys Glu Arg Trp Glu Asp Glu Asp Gln Arg Trp Lys Lys Leu Leu

1070 1075 1080

Arg Trp Phe Lys Asp Trp Val Leu Pro Ser Gly Thr Gln Ala Asn

1085 1090 1095

Asn Ala Thr Ile Arg Asn Val Gly Gly Leu Ser Leu Ser Arg Leu

1100 1105 1110

Ala Thr Ile Thr Glu Phe Arg Arg Lys Val Gln Val Gly Phe Phe

1115 1120 1125

Thr Arg Leu Arg Pro Asp Gly Thr Arg His Glu Ile Gly Glu Gln

1130 1135 1140

Phe Gly Gln Lys Thr Leu Asp Ala Leu Glu Leu Leu Arg Glu Gln
1145 1150 1155
Arg Val Lys Gln Leu Ala Ser Arg Ile Ala Glu Ala Ala Leu Gly
1160 1165 1170
Ile Gly Ser Glu Gly Gly Lys Gly Trp Asp Gly Gly Lys Arg Pro
1175 1180 1185
Arg Gln Arg Ile Asn Asp Ser Arg Phe Ala Pro Cys His Ala Val
1190 1195 1200

Val Ile Glu Asn Leu Ala Asn Tyr Arg Pro Asp Glu Thr Arg Thr
1205 1210 1215
Arg Leu Glu Asn Arg Arg Leu Met Thr Trp Ser Ala Ser Lys Val
1220 1225 1230
His Lys Tyr Leu Ser Glu Ala Cys Gln Leu Asn Gly Leu Tyr Leu
1235 1240 1245
Cys Thr Val Ser Ala Trp Tyr Thr Ser Arg Gln Asp Ser Arg Thr
1250 1255 1260

Gly Ala Pro Gly Ile Arg Cys Gln Asp Val Ser Val Arg Glu Phe
1265 1270 1275
Met Gln Ser Pro Phe Trp Arg Lys Gln Val Lys Gln Ala Glu Ala
1280 1285 1290
Lys His Asp Glu Asn Lys Gly Asp Ala Arg Glu Arg Phe Leu Cys
1295 1300 1305
Glu Leu Asn Lys Thr Trp Lys Ala Lys Thr Pro Ala Glu Trp Lys
1310 1315 1320

Lys Ala Gly Phe Val Arg Ile Pro Leu Arg Gly Gly Glu Ile Phe
1325 1330 1335
Val Ser Ala Asp Ser Lys Ser Pro Ser Ala Lys Gly Ile His Ala
1340 1345 1350
Asp Leu Asn Ala Ala Ala Asn Ile Gly Leu Arg Ala Leu Thr Asp
1355 1360 1365
Pro Asp Trp Pro Gly Lys Trp Trp Tyr Val Pro Cys Asp Pro Val

1370 1375 1380

Ser Phe Glu Ser Lys Met Asp Tyr Val Lys Gly Cys Ala Ala Val
1385 1390 1395
Lys Val Gly Gln Pro Leu Arg Gln Pro Ala Gln Thr Asn Ala Asp
1400 1405 1410
Gly Ala Ala Ser Lys Ile Arg Lys Gly Lys Lys Asn Arg Thr Ala
1415 1420 1425
Gly Thr Ser Lys Glu Lys Val Tyr Leu Trp Arg Asp Ile Ser Ala
1430 1435 1440

Phe Pro Leu Glu Ser Asn Glu Ile Gly Glu Trp Lys Glu Thr Ser
1445 1450 1455
Ala Tyr Gln Asn Asp Val Gln Tyr Arg Val Ile Arg Met Leu Lys
1460 1465 1470
Glu His Ile Lys Ser Leu Asp Asn Arg Thr Gly Asp Asn Val Glu
1475 1480 1485
Gly

<210> 496
<211> 1194
<212> PRT
<213> Desulfonatronum thiodismutans
<400> 496
Met Val Leu Gly Arg Lys Asp Asp Thr Ala Glu Leu Arg Arg Ala Leu

1 5 10 15
Trp Thr Thr His Glu His Val Asn Leu Ala Val Ala Glu Val Glu Arg
20 25 30
Val Leu Leu Arg Cys Arg Gly Arg Ser Tyr Trp Thr Leu Asp Arg Arg
35 40 45
Gly Asp Pro Val His Val Pro Glu Ser Gln Val Ala Glu Asp Ala Leu
50 55 60
Ala Met Ala Arg Glu Ala Gln Arg Arg Asn Gly Trp Pro Val Val Gly

65 70 75 80

Glu Asp Glu Glu Ile Leu Leu Ala Leu Arg Tyr Leu Tyr Glu Gln Ile

85 90 95

Val Pro Ser Cys Leu Leu Asp Asp Leu Gly Lys Pro Leu Lys Gly Asp

100 105 110

Ala Gln Lys Ile Gly Thr Asn Tyr Ala Gly Pro Leu Phe Asp Ser Asp

115 120 125

Thr Cys Arg Arg Asp Glu Gly Lys Asp Val Ala Cys Cys Gly Pro Phe

130 135 140

His Glu Val Ala Gly Lys Tyr Leu Gly Ala Leu Pro Glu Trp Ala Thr

145 150 155 160

Pro Ile Ser Lys Gln Glu Phe Asp Gly Lys Asp Ala Ser His Leu Arg

165 170 175

Phe Lys Ala Thr Gly Gly Asp Asp Ala Phe Phe Arg Val Ser Ile Glu

180 185 190

Lys Ala Asn Ala Trp Tyr Glu Asp Pro Ala Asn Gln Asp Ala Leu Lys

195 200 205

Asn Lys Ala Tyr Asn Lys Asp Asp Trp Lys Lys Glu Lys Asp Lys Gly

210 215 220

Ile Ser Ser Trp Ala Val Lys Tyr Ile Gln Lys Gln Leu Gln Leu Gly

225 230 235 240

Gln Asp Pro Arg Thr Glu Val Arg Arg Lys Leu Trp Leu Glu Leu Gly

245 250 255

Leu Leu Pro Leu Phe Ile Pro Val Phe Asp Lys Thr Met Val Gly Asn

260 265 270

Leu Trp Asn Arg Leu Ala Val Arg Leu Ala Leu Ala His Leu Leu Ser

275 280 285

Trp Glu Ser Trp Asn His Arg Ala Val Gln Asp Gln Ala Leu Ala Arg

290 295 300

Ala Lys Arg Asp Glu Leu Ala Ala Leu Phe Leu Gly Met Glu Asp Gly

305 310 315 320

Phe Ala Gly Leu Arg Glu Tyr Glu Leu Arg Arg Asn Glu Ser Ile Lys

325 330 335

Gln His Ala Phe Glu Pro Val Asp Arg Pro Tyr Val Val Ser Gly Arg

340 345 350

Ala Leu Arg Ser Trp Thr Arg Val Arg Glu Glu Trp Leu Arg His Gly

355 360 365

Asp Thr Gln Glu Ser Arg Lys Asn Ile Cys Asn Arg Leu Gln Asp Arg

370 375 380

Leu Arg Gly Lys Phe Gly Asp Pro Asp Val Phe His Trp Leu Ala Glu

385 390 395 400

Asp Gly Gln Glu Ala Leu Trp Lys Glu Arg Asp Cys Val Thr Ser Phe

405 410 415

Ser Leu Leu Asn Asp Ala Asp Gly Leu Leu Glu Lys Arg Lys Gly Tyr

420 425 430

Ala Leu Met Thr Phe Ala Asp Ala Arg Leu His Pro Arg Trp Ala Met

435 440 445

Tyr Glu Ala Pro Gly Gly Ser Asn Leu Arg Thr Tyr Gln Ile Arg Lys

450 455 460

Thr Glu Asn Gly Leu Trp Ala Asp Val Val Leu Leu Ser Pro Arg Asn

465 470 475 480

Glu Ser Ala Ala Val Glu Glu Lys Thr Phe Asn Val Arg Leu Ala Pro

485 490 495

Ser Gly Gln Leu Ser Asn Val Ser Phe Asp Gln Ile Gln Lys Gly Ser

500 505 510

Lys Met Val Gly Arg Cys Arg Tyr Gln Ser Ala Asn Gln Gln Phe Glu

515 520 525

Gly Leu Leu Gly Gly Ala Glu Ile Leu Phe Asp Arg Lys Arg Ile Ala

530 535 540

Asn Glu Gln His Gly Ala Thr Asp Leu Ala Ser Lys Pro Gly His Val

545 550 555 560

Trp Phe Lys Leu Thr Leu Asp Val Arg Pro Gln Ala Pro Gln Gly Trp

565 570 575

Leu Asp Gly Lys Gly Arg Pro Ala Leu Pro Pro Glu Ala Lys His Phe

580 585 590

Lys Thr Ala Leu Ser Asn Lys Ser Lys Phe Ala Asp Gln Val Arg Pro

595 600 605

Gly Leu Arg Val Leu Ser Val Asp Leu Gly Val Arg Ser Phe Ala Ala

610 615 620

Cys Ser Val Phe Glu Leu Val Arg Gly Gly Pro Asp Gln Gly Thr Tyr

625 630 635 640

Phe Pro Ala Ala Asp Gly Arg Thr Val Asp Asp Pro Glu Lys Leu Trp

645 650 655

Ala Lys His Glu Arg Ser Phe Lys Ile Thr Leu Pro Gly Glu Asn Pro

660 665 670

Ser Arg Lys Glu Glu Ile Ala Arg Arg Ala Ala Met Glu Glu Leu Arg

675 680 685

Ser Leu Asn Gly Asp Ile Arg Arg Leu Lys Ala Ile Leu Arg Leu Ser

690 695 700

Val Leu Gln Glu Asp Asp Pro Arg Thr Glu His Leu Arg Leu Phe Met

705 710 715 720

Glu Ala Ile Val Asp Asp Pro Ala Lys Ser Ala Leu Asn Ala Glu Leu

725 730 735

Phe Lys Gly Phe Gly Asp Asp Arg Phe Arg Ser Thr Pro Asp Leu Trp

740 745 750

Lys Gln His Cys His Phe Phe His Asp Lys Ala Glu Lys Val Val Ala

755 760 765

Glu Arg Phe Ser Arg Trp Arg Thr Glu Thr Arg Pro Lys Ser Ser Ser

770 775 780

Trp Gln Asp Trp Arg Glu Arg Arg Gly Tyr Ala Gly Gly Lys Ser Tyr

785 790 795 800

Trp Ala Val Thr Tyr Leu Glu Ala Val Arg Gly Leu Ile Leu Arg Trp

805 810 815

Asn Met Arg Gly Arg Thr Tyr Gly Glu Val Asn Arg Gln Asp Lys Lys
820 825 830
Gln Phe Gly Thr Val Ala Ser Ala Leu Leu His His Ile Asn Gln Leu

835 840 845
Lys Glu Asp Arg Ile Lys Thr Gly Ala Asp Met Ile Ile Gln Ala Ala
850 855 860
Arg Gly Phe Val Pro Arg Lys Asn Gly Ala Gly Trp Val Gln Val His
865 870 875 880
Glu Pro Cys Arg Leu Ile Leu Phe Glu Asp Leu Ala Arg Tyr Arg Phe
885 890 895
Arg Thr Asp Arg Ser Arg Arg Glu Asn Ser Arg Leu Met Arg Trp Ser

900 905 910
His Arg Glu Ile Val Asn Glu Val Gly Met Gln Gly Glu Leu Tyr Gly
915 920 925
Leu His Val Asp Thr Thr Glu Ala Gly Phe Ser Ser Arg Tyr Leu Ala
930 935 940
Ser Ser Gly Ala Pro Gly Val Arg Cys Arg His Leu Val Glu Glu Asp
945 950 955 960
Phe His Asp Gly Leu Pro Gly Met His Leu Val Gly Glu Leu Asp Trp

965 970 975
Leu Leu Pro Lys Asp Lys Asp Arg Thr Ala Asn Glu Ala Arg Arg Leu
980 985 990
Leu Gly Gly Met Val Arg Pro Gly Met Leu Val Pro Trp Asp Gly Gly
995 1000 1005
Glu Leu Phe Ala Thr Leu Asn Ala Ala Ser Gln Leu His Val Ile
1010 1015 1020
His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg Phe Trp

1025 1030 1035
Gly Arg Cys Gly Glu Ala Ile Arg Ile Val Cys Asn Gln Leu Ser
1040 1045 1050
Val Asp Gly Ser Thr Arg Tyr Glu Met Ala Lys Ala Pro Lys Ala

1055 1060 1065

Arg Leu Leu Gly Ala Leu Gln Gln Leu Lys Asn Gly Asp Ala Pro

1070 1075 1080

Phe His Leu Thr Ser Ile Pro Asn Ser Gln Lys Pro Glu Asn Ser

1085 1090 1095

Tyr Val Met Thr Pro Thr Asn Ala Gly Lys Lys Tyr Arg Ala Gly

1100 1105 1110

Pro Gly Glu Lys Ser Ser Gly Glu Glu Asp Glu Leu Ala Leu Asp

1115 1120 1125

Ile Val Glu Gln Ala Glu Glu Leu Ala Gln Gly Arg Lys Thr Phe

1130 1135 1140

Phe Arg Asp Pro Ser Gly Val Phe Phe Ala Pro Asp Arg Trp Leu

1145 1150 1155

Pro Ser Glu Ile Tyr Trp Ser Arg Ile Arg Arg Arg Ile Trp Gln

1160 1165 1170

Val Thr Leu Glu Arg Asn Ser Ser Gly Arg Gln Glu Arg Ala Glu

1175 1180 1185

Met Asp Glu Met Pro Tyr

1190

<210> 497

<211> 1326

<212> PRT

<213> Unknown

<220><223> Lentisphaeria bacterium

<400> 497

Met Ala Val Glu Leu Asn Arg Ile Tyr Gln Gly Arg Val Asn His Val

1 5 10 15

Tyr Ile Phe Asp Glu Asn Gln Asn Gln Val Ser Val Asp Asn Gly Asp

20 25 30

Asp Leu Leu Phe Val His His Glu Leu Tyr Gln Asp Ala Ile Asn Tyr

35 40 45

Tyr Leu Val Ala Leu Ala Ala Met Ala Leu Asp Ser Lys Asp Ser Leu

50 55 60

Phe Gly Lys Phe Lys Met Gln Ile Arg Ala Val Trp Asn Asp Phe Tyr

65 70 75 80

Arg Asn Gly Gln Leu Arg Pro Gly Leu Lys His Ser Leu Ile Arg Ser

85 90 95

Leu Gly His Ala Ala Glu Leu Asn Thr Ser Asn Gly Ala Asp Ile Ala

100 105 110

Met Asn Leu Ile Leu Glu Asp Gly Gly Ile Pro Ser Glu Ile Leu Asn

115 120 125

Ala Ala Leu Glu His Leu Ala Glu Lys Cys Thr Gly Asp Val Ser Gln

130 135 140

Leu Gly Lys Thr Phe Phe Pro Arg Phe Cys Asp Thr Ala Tyr His Gly

145 150 155 160

Asn Trp Asp Val Asp Ala Lys Ser Phe Ser Glu Lys Lys Gly Arg Gln

165 170 175

Arg Leu Val Asp Ala Leu Tyr Ser Leu His Pro Val Gln Ala Val Gln

180 185 190

Glu Leu Ala Pro Glu Ile Glu Ile Gly Trp Gly Gly Val Lys Thr Gln

195 200 205

Thr Gly Lys Phe Phe Thr Gly Asp Glu Ala Lys Ala Ser Leu Lys Lys

210 215 220

Ala Ile Ser Tyr Phe Leu Gln Asp Thr Gly Lys Asn Ser Pro Glu Leu

225 230 235 240

Gln Glu Tyr Phe Ser Val Ala Gly Lys Gln Pro Leu Glu Gln Tyr Leu

245 250 255

Gly Lys Ile Asp Thr Phe Pro Glu Ile Ser Phe Gly Arg Ile Ser Ser

260 265 270

His Gln Asn Ile Asn Ile Ser Asn Ala Met Trp Ile Leu Lys Phe Phe

275 280 285

Pro Asp Gln Tyr Ser Val Asp Leu Ile Lys Asn Leu Ile Pro Asn Lys

290 295 300

Lys Tyr Glu Ile Gly Ile Ala Pro Gln Trp Gly Asp Asp Pro Val Lys
305 310 315 320
Leu Ser Arg Gly Lys Arg Gly Tyr Thr Phe Arg Ala Phe Thr Asp Leu

325 330 335
Ala Met Trp Glu Lys Asn Trp Lys Val Phe Asp Arg Ala Ala Phe Ser
340 345 350
Asp Ala Leu Lys Thr Ile Asn Gln Phe Arg Asn Lys Thr Gln Glu Arg
355 360 365
Asn Asp Gln Leu Lys Arg Tyr Cys Ala Ala Leu Asn Trp Met Asp Gly
370 375 380
Glu Ser Ser Asp Lys Lys Pro Pro Val Glu Pro Ala Asp Ala Asp Ala

385 390 395 400
Val Asp Glu Ala Ala Thr Ser Val Leu Pro Ile Leu Ala Gly Asp Lys
405 410 415
Arg Trp Asn Ala Leu Leu Gln Leu Gln Lys Glu Leu Gly Ile Cys Asn
420 425 430
Asp Phe Thr Glu Asn Glu Leu Met Asp Tyr Gly Leu Ser Leu Arg Thr
435 440 445
Ile Arg Gly Tyr Gln Lys Leu Arg Ser Met Met Leu Glu Lys Glu Glu

450 455 460
Lys Met Arg Ala Lys Thr Ala Asp Asp Glu Glu Ile Ser Gln Ala Leu
465 470 475 480
Gln Glu Ile Ile Ile Lys Phe Gln Ser Ser His Arg Asp Thr Ile Gly
485 490 495
Ser Val Ser Leu Phe Leu Lys Leu Ala Glu Pro Lys Tyr Phe Cys Val
500 505 510
Trp His Asp Ala Asp Lys Asn Gln Asn Phe Ala Ser Val Asp Met Val

515 520 525
Ala Asp Ala Val Arg Tyr Tyr Ser Tyr Gln Glu Glu Lys Ala Arg Leu
530 535 540
Glu Glu Pro Ile Gln Ile Thr Pro Ala Asp Ala Arg Tyr Ser Arg Arg

545 550 555 560

Val Ser Asp Leu Tyr Ala Leu Val Tyr Lys Asn Ala Lys Glu Cys Lys

565 570 575

Thr Gly Tyr Gly Leu Arg Pro Asp Gly Asn Phe Val Phe Glu Ile Ala

580 585 590

Gln Lys Asn Ala Lys Gly Tyr Ala Pro Ala Lys Val Val Leu Ala Phe

595 600 605

Ser Ala Pro Arg Leu Lys Arg Asp Gly Leu Ile Asp Lys Glu Phe Ser

610 615 620

Ala Tyr Tyr Pro Pro Val Leu Gln Ala Phe Leu Arg Glu Glu Glu Ala

625 630 635 640

Pro Lys Gln Ser Phe Lys Thr Thr Ala Val Ile Leu Met Pro Asp Trp

645 650 655

Asp Lys Asn Gly Lys Arg Arg Ile Leu Leu Asn Phe Pro Ile Lys Leu

660 665 670

Asp Val Ser Ala Ile His Gln Lys Thr Asp His Arg Phe Glu Asn Gln

675 680 685

Phe Tyr Phe Ala Asn Asn Thr Asn Thr Cys Leu Leu Trp Pro Ser Tyr

690 695 700

Gln Tyr Lys Lys Pro Val Thr Trp Tyr Gln Gly Lys Lys Pro Phe Asp

705 710 715 720

Val Val Ala Val Asp Leu Gly Gln Arg Ser Ala Gly Ala Val Ser Arg

725 730 735

Ile Thr Val Ser Thr Glu Lys Arg Glu His Ser Val Ala Ile Gly Glu

740 745 750

Ala Gly Gly Thr Gln Trp Tyr Ala Tyr Arg Lys Phe Ser Gly Leu Leu

755 760 765

Arg Leu Pro Gly Glu Asp Ala Thr Val Ile Arg Asp Gly Gln Arg Thr

770 775 780

Glu Glu Leu Ser Gly Asn Ala Gly Arg Leu Ser Thr Glu Glu Glu Thr

785 790 795 800

Val Gln Ala Cys Val Leu Cys Lys Met Leu Ile Gly Asp Ala Thr Leu
805 810 815
Leu Gly Gly Ser Asp Glu Lys Thr Ile Arg Ser Phe Pro Lys Gln Asn
820 825 830
Asp Lys Leu Leu Ile Ala Phe Arg Arg Ala Thr Gly Arg Met Lys Gln

835 840 845
Leu Gln Arg Trp Leu Trp Met Leu Asn Glu Asn Gly Leu Cys Asp Lys
850 855 860
Ala Lys Thr Glu Ile Ser Asn Ser Asp Trp Leu Val Asn Lys Asn Ile
865 870 875 880
Asp Asn Val Leu Lys Glu Glu Lys Gln His Arg Glu Met Leu Pro Ala
885 890 895
Ile Leu Leu Gln Ile Ala Asp Arg Val Leu Pro Leu Arg Gly Arg Lys

900 905 910
Trp Asp Trp Val Leu Asn Pro Gln Ser Asn Ser Phe Val Leu Gln Gln
915 920 925
Thr Ala His Gly Ser Gly Asp Pro His Lys Lys Ile Cys Gly Gln Arg
930 935 940
Gly Leu Ser Phe Ala Arg Ile Glu Gln Leu Glu Ser Leu Arg Met Arg
945 950 955 960
Cys Gln Ala Leu Asn Arg Ile Leu Met Arg Lys Thr Gly Glu Lys Pro

965 970 975
Ala Thr Leu Ala Glu Met Arg Asn Asn Pro Ile Pro Asp Cys Cys Pro
980 985 990
Asp Ile Leu Met Arg Leu Asp Ala Met Lys Glu Gln Arg Ile Asn Gln
995 1000 1005
Thr Ala Asn Leu Ile Leu Ala Gln Ala Leu Gly Leu Arg His Cys
1010 1015 1020
Leu His Ser Glu Ser Ala Thr Lys Arg Lys Glu Asn Gly Met His

1025 1030 1035
Gly Glu Tyr Glu Lys Ile Pro Gly Val Glu Pro Ala Ala Phe Val

1040 1045 1050

Val Leu Glu Asp Leu Ser Arg Tyr Arg Phe Ser Gln Asp Arg Ser

1055 1060 1065

Ser Tyr Glu Asn Ser Arg Leu Met Lys Trp Ser His Arg Lys Ile

1070 1075 1080

Leu Glu Lys Leu Ala Leu Leu Cys Glu Val Phe Asn Val Pro Ile

1085 1090 1095

Leu Gln Val Gly Ala Ala Tyr Ser Ser Lys Phe Ser Ala Asn Ala

1100 1105 1110

Ile Pro Gly Phe Arg Ala Glu Glu Cys Ser Ile Asp Gln Leu Ser

1115 1120 1125

Phe Tyr Pro Trp Arg Glu Leu Lys Asp Ser Arg Glu Lys Ala Leu

1130 1135 1140

Val Glu Gln Ile Arg Lys Ile Gly His Arg Leu Leu Thr Phe Asp

1145 1150 1155

Ala Lys Ala Thr Ile Ile Met Pro Arg Asn Gly Gly Pro Val Phe

1160 1165 1170

Ile Pro Phe Val Pro Ser Asp Ser Lys Asp Thr Leu Ile Gln Ala

1175 1180 1185

Asp Ile Asn Ala Ser Phe Asn Ile Gly Leu Arg Gly Val Ala Asp

1190 1195 1200

Ala Thr Asn Leu Leu Cys Asn Asn Arg Val Ser Cys Asp Arg Lys

1205 1210 1215

Lys Asp Cys Trp Gln Val Lys Arg Ser Ser Asn Phe Ser Lys Met

1220 1225 1230

Val Tyr Pro Glu Lys Leu Ser Leu Ser Phe Asp Pro Ile Lys Lys

1235 1240 1245

Gln Glu Gly Ala Gly Gly Asn Phe Phe Val Leu Gly Cys Ser Glu

1250 1255 1260

Arg Ile Leu Thr Gly Thr Ser Glu Lys Ser Pro Val Phe Thr Ser

1265 1270 1275

Ser Glu Met Ala Lys Lys Tyr Pro Asn Leu Met Phe Gly Ser Ala
1280 1285 1290
Leu Trp Arg Asn Glu Ile Leu Lys Leu Glu Arg Cys Cys Lys Ile
1295 1300 1305
Asn Gln Ser Arg Leu Asp Lys Phe Ile Ala Lys Lys Glu Val Gln
1310 1315 1320
Asn Glu Leu
1325

<210> 498
<211> 1090

<212> PRT
<213> Laceyella sediminis
<400> 498

Met Ser Ile Arg Ser Phe Lys Leu Lys Ile Lys Thr Lys Ser Gly Val
1 5 10 15
Asn Ala Glu Glu Leu Arg Arg Gly Leu Trp Arg Thr His Gln Leu Ile
20 25 30
Asn Asp Gly Ile Ala Tyr Tyr Met Asn Trp Leu Val Leu Leu Arg Gln
35 40 45
Glu Asp Leu Phe Ile Arg Asn Glu Glu Thr Asn Glu Ile Glu Lys Arg
50 55 60
Ser Lys Glu Glu Ile Gln Gly Glu Leu Leu Glu Arg Val His Lys Gln
65 70 75 80
Gln Gln Arg Asn Gln Trp Ser Gly Glu Val Asp Asp Gln Thr Leu Leu
85 90 95
Gln Thr Leu Arg His Leu Tyr Glu Glu Ile Val Pro Ser Val Ile Gly
100 105 110
Lys Ser Gly Asn Ala Ser Leu Lys Ala Arg Phe Phe Leu Gly Pro Leu
115 120 125
Val Asp Pro Asn Asn Lys Thr Thr Lys Asp Val Ser Lys Ser Gly Pro
130 135 140
Thr Pro Lys Trp Lys Lys Met Lys Asp Ala Gly Asp Pro Asn Trp Val

145 150 155 160

Gln Glu Tyr Glu Lys Tyr Met Ala Glu Arg Gln Thr Leu Val Arg Leu

165 170 175

Glu Glu Met Gly Leu Ile Pro Leu Phe Pro Met Tyr Thr Asp Glu Val

180 185 190

Gly Asp Ile His Trp Leu Pro Gln Ala Ser Gly Tyr Thr Arg Thr Trp

195 200 205

Asp Arg Asp Met Phe Gln Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu

210 215 220

Ser Trp Asn Arg Arg Val Arg Glu Arg Arg Ala Gln Phe Glu Lys Lys

225 230 235 240

Thr His Asp Phe Ala Ser Arg Phe Ser Glu Ser Asp Val Gln Trp Met

245 250 255

Asn Lys Leu Arg Glu Tyr Glu Ala Gln Gln Glu Lys Ser Leu Glu Glu

260 265 270

Asn Ala Phe Ala Pro Asn Glu Pro Tyr Ala Leu Thr Lys Lys Ala Leu

275 280 285

Arg Gly Trp Glu Arg Val Tyr His Ser Trp Met Arg Leu Asp Ser Ala

290 295 300

Ala Ser Glu Glu Ala Tyr Trp Gln Glu Val Ala Thr Cys Gln Thr Ala

305 310 315 320

Met Arg Gly Glu Phe Gly Asp Pro Ala Ile Tyr Gln Phe Leu Ala Gln

325 330 335

Lys Glu Asn His Asp Ile Trp Arg Gly Tyr Pro Glu Arg Val Ile Asp

340 345 350

Phe Ala Glu Leu Asn His Leu Gln Arg Glu Leu Arg Arg Ala Lys Glu

355 360 365

Asp Ala Thr Phe Thr Leu Pro Asp Ser Val Asp His Pro Leu Trp Val

370 375 380

Arg Tyr Glu Ala Pro Gly Gly Thr Asn Ile His Gly Tyr Asp Leu Val

385 390 395 400

Gln Asp Thr Lys Arg Asn Leu Thr Leu Ile Leu Asp Lys Phe Ile Leu
405 410 415
Pro Asp Glu Asn Gly Ser Trp His Glu Val Lys Lys Val Pro Phe Ser
420 425 430
Leu Ala Lys Ser Lys Gln Phe His Arg Gln Val Trp Leu Gln Glu Glu

435 440 445
Gln Lys Gln Lys Lys Arg Glu Val Val Phe Tyr Asp Tyr Ser Thr Asn
450 455 460
Leu Pro His Leu Gly Thr Leu Ala Gly Ala Lys Leu Gln Trp Asp Arg
465 470 475 480
Asn Phe Leu Asn Lys Arg Thr Gln Gln Gln Ile Glu Glu Thr Gly Glu
485 490 495
Ile Gly Lys Val Phe Phe Asn Ile Ser Val Asp Val Arg Pro Ala Val

500 505 510
Glu Val Lys Asn Gly Arg Leu Gln Asn Gly Leu Gly Lys Ala Leu Thr
515 520 525
Val Leu Thr His Pro Asp Gly Thr Lys Ile Val Thr Gly Trp Lys Ala
530 535 540
Glu Gln Leu Glu Lys Trp Val Gly Glu Ser Gly Arg Val Ser Ser Leu
545 550 555 560
Gly Leu Asp Ser Leu Ser Glu Gly Leu Arg Val Met Ser Ile Asp Leu

565 570 575
Gly Gln Arg Thr Ser Ala Thr Val Ser Val Phe Glu Ile Thr Lys Glu
580 585 590
Ala Pro Asp Asn Pro Tyr Lys Phe Phe Tyr Gln Leu Glu Gly Thr Glu
595 600 605
Leu Phe Ala Val His Gln Arg Ser Phe Leu Leu Ala Leu Pro Gly Glu
610 615 620
Asn Pro Pro Gln Lys Ile Lys Gln Met Arg Glu Ile Arg Trp Lys Glu

625 630 635 640
Arg Asn Arg Ile Lys Gln Gln Val Asp Gln Leu Ser Ala Ile Leu Arg

645 650 655

Leu His Lys Lys Val Asn Glu Asp Glu Arg Ile Gln Ala Ile Asp Lys

660 665 670

Leu Leu Gln Lys Val Ala Ser Trp Gln Leu Asn Glu Glu Ile Ala Thr

675 680 685

Ala Trp Asn Gln Ala Leu Ser Gln Leu Tyr Ser Lys Ala Lys Glu Asn

690 695 700

Asp Leu Gln Trp Asn Gln Ala Ile Lys Asn Ala His His Gln Leu Glu

705 710 715 720

Pro Val Val Gly Lys Gln Ile Ser Leu Trp Arg Lys Asp Leu Ser Thr

725 730 735

Gly Arg Gln Gly Ile Ala Gly Leu Ser Leu Trp Ser Ile Glu Glu Leu

740 745 750

Glu Ala Thr Lys Lys Leu Leu Thr Arg Trp Ser Lys Arg Ser Arg Glu

755 760 765

Pro Gly Val Val Lys Arg Ile Glu Arg Phe Glu Thr Phe Ala Lys Gln

770 775 780

Ile Gln His His Ile Asn Gln Val Lys Glu Asn Arg Leu Lys Gln Leu

785 790 795 800

Ala Asn Leu Ile Val Met Thr Ala Leu Gly Tyr Lys Tyr Asp Gln Glu

805 810 815

Gln Lys Lys Trp Ile Glu Val Tyr Pro Ala Cys Gln Val Val Leu Phe

820 825 830

Glu Asn Leu Arg Ser Tyr Arg Phe Ser Tyr Glu Arg Ser Arg Arg Glu

835 840 845

Asn Lys Lys Leu Met Glu Trp Ser His Arg Ser Ile Pro Lys Leu Val

850 855 860

Gln Met Gln Gly Glu Leu Phe Gly Leu Gln Val Ala Asp Val Tyr Ala

865 870 875 880

Ala Tyr Ser Ser Arg Tyr His Gly Arg Thr Gly Ala Pro Gly Ile Arg

885 890 895

Cys His Ala Leu Thr Glu Ala Asp Leu Arg Asn Glu Thr Asn Ile Ile
900 905 910
His Glu Leu Ile Glu Ala Gly Phe Ile Lys Glu Glu His Arg Pro Tyr
915 920 925
Leu Gln Gln Gly Asp Leu Val Pro Trp Ser Gly Gly Glu Leu Phe Ala
930 935 940
Thr Leu Gln Lys Pro Tyr Asp Asn Pro Arg Ile Leu Thr Leu His Ala

945 950 955 960
Asp Ile Asn Ala Ala Gln Asn Ile Gln Lys Arg Phe Trp His Pro Ser
965 970 975
Met Trp Phe Arg Val Asn Cys Glu Ser Val Met Glu Gly Glu Ile Val
980 985 990
Thr Tyr Val Pro Lys Asn Lys Thr Val His Lys Lys Gln Gly Lys Thr
995 1000 1005
Phe Arg Phe Val Lys Val Glu Gly Ser Asp Val Tyr Glu Trp Ala

1010 1015 1020
Lys Trp Ser Lys Asn Arg Asn Lys Asn Thr Phe Ser Ser Ile Thr
1025 1030 1035
Glu Arg Lys Pro Pro Ser Ser Met Ile Leu Phe Arg Asp Pro Ser
1040 1045 1050
Gly Thr Phe Phe Lys Glu Gln Glu Trp Val Glu Gln Lys Thr Phe
1055 1060 1065
Trp Gly Lys Val Gln Ser Met Ile Gln Ala Tyr Met Lys Lys Thr

1070 1075 1080
Ile Val Gln Arg Met Glu Glu
1085 1090

<210> 499
<211> 1041
<212> PRT
<213> Methylobacterium nodulans
<400> 499
Met Tyr Glu Ala Ile Val Leu Ala Asp Asp Ala Asn Ala Gln Leu Ala

1 5 10 15

Asn Ala Phe Leu Gly Pro Leu Thr Asp Pro Asn Ser Ala Gly Phe Leu

20 25 30

Glu Ala Phe Asn Lys Val Asp Arg Pro Ala Pro Ser Trp Leu Asp Gln

35 40 45

Val Pro Ala Ser Asp Pro Ile Asp Pro Ala Val Leu Ala Glu Ala Asn

50 55 60

Ala Trp Leu Asp Thr Asp Ala Gly Arg Ala Trp Leu Val Asp Thr Gly

65 70 75 80

Ala Pro Pro Arg Trp Arg Ser Leu Ala Ala Lys Gln Asp Pro Ile Trp

85 90 95

Pro Arg Glu Phe Ala Arg Lys Leu Gly Glu Leu Arg Lys Glu Ala Ala

100 105 110

Ser Gly Thr Ser Ala Ile Ile Lys Ala Leu Lys Arg Asp Phe Gly Val

115 120 125

Leu Pro Leu Phe Gln Pro Ser Leu Ala Pro Arg Ile Leu Gly Ser Arg

130 135 140

Ser Ser Leu Thr Pro Trp Asp Arg Leu Ala Phe Arg Leu Ala Val Gly

145 150 155 160

His Leu Leu Ser Trp Glu Ser Trp Cys Thr Arg Ala Arg Asp Glu His

165 170 175

Thr Ala Arg Val Gln Arg Leu Glu Gln Phe Ser Ser Ala His Leu Lys

180 185 190

Gly Asp Leu Ala Thr Lys Val Ser Thr Leu Arg Glu Tyr Glu Arg Ala

195 200 205

Arg Lys Glu Gln Ile Ala Gln Leu Gly Leu Pro Met Gly Glu Arg Asp

210 215 220

Phe Leu Ile Thr Val Arg Met Thr Arg Gly Trp Asp Asp Leu Arg Glu

225 230 235 240

Lys Trp Arg Arg Ser Gly Asp Lys Gly Gln Glu Ala Leu His Ala Ile

245 250 255

Ile Ala Thr Glu Gln Thr Arg Lys Arg Gly Arg Phe Gly Asp Pro Asp
260 265 270
Leu Phe Arg Trp Leu Ala Arg Pro Glu Asn His His Val Trp Ala Asp
275 280 285
Gly His Ala Asp Ala Val Gly Val Leu Ala Arg Val Asn Ala Met Glu

290 295 300
Arg Leu Val Glu Arg Ser Arg Asp Thr Ala Leu Met Thr Leu Pro Asp
305 310 315 320
Pro Val Ala His Pro Arg Ser Ala Gln Trp Glu Ala Glu Gly Gly Ser
325 330 335
Asn Leu Arg Asn Tyr Gln Leu Glu Ala Val Gly Gly Glu Leu Gln Ile
340 345 350
Thr Leu Pro Leu Leu Lys Ala Ala Asp Asp Gly Arg Cys Ile Asp Thr

355 360 365
Pro Leu Ser Phe Ser Leu Ala Pro Ser Asp Gln Leu Gln Gly Val Val
370 375 380
Leu Thr Lys Gln Asp Lys Gln Gln Lys Ile Thr Tyr Cys Thr Asn Met
385 390 395 400
Asn Glu Val Phe Glu Ala Lys Leu Gly Ser Ala Asp Leu Leu Leu Asn
405 410 415
Trp Asp His Leu Arg Gly Arg Ile Arg Asp Arg Val Asp Ala Gly Asp

420 425 430
Ile Gly Ser Ala Phe Leu Lys Leu Ala Leu Asp Val Ala His Val Leu
435 440 445
Pro Asp Gly Val Asp Asp Gln Leu Ala Arg Ala Ala Phe His Phe Gln
450 455 460
Ser Ala Lys Gly Ala Lys Ser Lys His Ala Asp Ser Val Gln Ala Gly
465 470 475 480
Leu Arg Val Leu Ser Ile Asp Leu Gly Val Arg Ser Phe Ala Thr Cys

485 490 495
Ser Val Phe Glu Leu Lys Asp Thr Ala Pro Thr Thr Gly Val Ala Phe

500 505 510

Pro Leu Ala Glu Phe Arg Leu Trp Ala Val His Glu Arg Ser Phe Thr

515 520 525

Leu Glu Leu Pro Gly Glu Asn Val Gly Ala Ala Gly Gln Gln Trp Arg

530 535 540

Ala Gln Ala Asp Ala Glu Leu Arg Gln Leu Arg Gly Gly Leu Asn Arg

545 550 555 560

His Arg Gln Leu Leu Arg Ala Ala Thr Val Gln Lys Gly Glu Arg Asp

565 570 575

Ala Tyr Leu Thr Asp Leu Arg Glu Ala Trp Ser Ala Lys Glu Leu Trp

580 585 590

Pro Phe Glu Ala Ser Leu Leu Ser Glu Leu Glu Arg Cys Ser Thr Val

595 600 605

Ala Asp Pro Leu Trp Gln Asp Thr Cys Lys Arg Ala Ala Arg Leu Tyr

610 615 620

Arg Thr Glu Phe Gly Ala Val Val Ser Glu Trp Arg Ser Arg Thr Arg

625 630 635 640

Ser Arg Glu Asp Arg Lys Tyr Ala Gly Lys Ser Met Trp Ser Val Gln

645 650 655

His Leu Thr Asp Val Arg Arg Phe Leu Gln Ser Trp Ser Leu Ala Gly

660 665 670

Arg Ala Ser Gly Asp Ile Arg Arg Leu Asp Arg Glu Arg Gly Gly Val

675 680 685

Phe Ala Lys Asp Leu Leu Asp His Ile Asp Ala Leu Lys Asp Asp Arg

690 695 700

Leu Lys Thr Gly Ala Asp Leu Ile Val Gln Ala Ala Arg Gly Phe Gln

705 710 715 720

Arg Asn Glu Phe Gly Tyr Trp Val Gln Lys His Ala Pro Cys His Val

725 730 735

Ile Leu Phe Glu Asp Leu Ser Arg Tyr Arg Met Arg Thr Asp Arg Pro

740 745 750

Arg Arg Glu Asn Ser Gln Leu Met Gln Trp Ala His Arg Gly Val Pro
755 760 765
Asp Met Val Gly Met Gln Gly Glu Ile Tyr Gly Ile Gln Asp Arg Arg
770 775 780
Asp Pro Asp Ser Ala Arg Lys His Ala Arg Gln Pro Leu Ala Ala Phe
785 790 795 800
Cys Leu Asp Thr Pro Ala Ala Phe Ser Ser Arg Tyr His Ala Ser Thr

805 810 815
Met Thr Pro Gly Ile Arg Cys His Pro Leu Arg Lys Arg Glu Phe Glu
820 825 830
Asp Gln Gly Phe Leu Glu Leu Leu Lys Arg Glu Asn Glu Gly Leu Asp
835 840 845
Leu Asn Gly Tyr Lys Pro Gly Asp Leu Val Pro Leu Pro Gly Gly Glu
850 855 860
Val Phe Val Cys Leu Asn Ala Asn Gly Leu Ser Arg Ile His Ala Asp

865 870 875 880
Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg Phe Trp Thr Gln His Gly
885 890 895
Asp Ala Phe Arg Leu Pro Cys Gly Lys Ser Ala Val Gln Gly Gln Ile
900 905 910
Arg Trp Ala Pro Leu Ser Met Gly Lys Arg Gln Ala Gly Ala Leu Gly
915 920 925
Gly Phe Gly Tyr Leu Glu Pro Thr Gly His Asp Ser Gly Ser Cys Gln

930 935 940
Trp Arg Lys Thr Thr Glu Ala Glu Trp Arg Arg Leu Ser Gly Ala Gln
945 950 955 960
Lys Asp Arg Asp Glu Ala Ala Ala Ala Glu Asp Glu Glu Leu Gln Gly
965 970 975
Leu Glu Glu Glu Leu Leu Glu Arg Ser Gly Glu Arg Val Val Phe Phe
980 985 990
Arg Asp Pro Ser Gly Val Val Leu Pro Thr Asp Leu Trp Phe Pro Ser

995 1000 1005

Ala Ala Phe Trp Ser Ile Val Arg Ala Lys Thr Val Gly Arg Leu

1010 1015 1020

Arg Ser His Leu Asp Ala Gln Ala Glu Ala Ser Tyr Ala Val Ala

1025 1030 1035

Ala Gly Leu

1040

<210> 500

<211> 1388

<212> PRT

<213> Unknown

<220><223> Member of the phylum Opitutaceae

<400> 500

Met Ser Leu Asn Arg Ile Tyr Gln Gly Arg Val Ala Ala Val Glu Thr

1 5 10 15

Gly Thr Ala Leu Ala Lys Gly Asn Val Glu Trp Met Pro Ala Ala Gly

20 25 30

Gly Asp Glu Val Leu Trp Gln His His Glu Leu Phe Gln Ala Ala Ile

35 40 45

Asn Tyr Tyr Leu Val Ala Leu Leu Ala Leu Ala Asp Lys Asn Asn Pro

50 55 60

Val Leu Gly Pro Leu Ile Ser Gln Met Asp Asn Pro Gln Ser Pro Tyr

65 70 75 80

His Val Trp Gly Ser Phe Arg Arg Gln Gly Arg Gln Arg Thr Gly Leu

85 90 95

Ser Gln Ala Val Ala Pro Tyr Ile Thr Pro Gly Asn Asn Ala Pro Thr

100 105 110

Leu Asp Glu Val Phe Arg Ser Ile Leu Ala Gly Asn Pro Thr Asp Arg

115 120 125

Ala Thr Leu Asp Ala Ala Leu Met Gln Leu Leu Lys Ala Cys Asp Gly

130 135 140

Ala Gly Ala Ile Gln Gln Glu Gly Arg Ser Tyr Trp Pro Lys Phe Cys

145 150 155 160

Asp Pro Asp Ser Thr Ala Asn Phe Ala Gly Asp Pro Ala Met Leu Arg

165 170 175

Arg Glu Gln His Arg Leu Leu Leu Pro Gln Val Leu His Asp Pro Ala

180 185 190

Ile Thr His Asp Ser Pro Ala Leu Gly Ser Phe Asp Thr Tyr Ser Ile

195 200 205

Ala Thr Pro Asp Thr Arg Thr Pro Gln Leu Thr Gly Pro Lys Ala Arg

210 215 220

Ala Arg Leu Glu Gln Ala Ile Thr Leu Trp Arg Val Arg Leu Pro Glu

225 230 235 240

Ser Ala Ala Asp Phe Asp Arg Leu Ala Ser Ser Leu Lys Lys Ile Pro

245 250 255

Asp Asp Asp Ser Arg Leu Asn Leu Gln Gly Tyr Val Gly Ser Ser Ala

260 265 270

Lys Gly Glu Val Gln Ala Arg Leu Phe Ala Leu Leu Leu Phe Arg His

275 280 285

Leu Glu Arg Ser Ser Phe Thr Leu Gly Leu Leu Arg Ser Ala Thr Pro

290 295 300

Pro Pro Lys Asn Ala Glu Thr Pro Pro Pro Ala Gly Val Pro Leu Pro

305 310 315 320

Ala Ala Ser Ala Ala Asp Pro Val Arg Ile Ala Arg Gly Lys Arg Ser

325 330 335

Phe Val Phe Arg Ala Phe Thr Ser Leu Pro Cys Trp His Gly Gly Asp

340 345 350

Asn Ile His Pro Thr Trp Lys Ser Phe Asp Ile Ala Ala Phe Lys Tyr

355 360 365

Ala Leu Thr Val Ile Asn Gln Ile Glu Glu Lys Thr Lys Glu Arg Gln

370 375 380

Lys Glu Cys Ala Glu Leu Glu Thr Asp Phe Asp Tyr Met His Gly Arg

385 390 395 400

Leu Ala Lys Ile Pro Val Lys Tyr Thr Thr Gly Glu Ala Glu Pro Pro
405 410 415
Pro Ile Leu Ala Asn Asp Leu Arg Ile Pro Leu Leu Arg Glu Leu Leu
420 425 430
Gln Asn Ile Lys Val Asp Thr Ala Leu Thr Asp Gly Glu Ala Val Ser
435 440 445
Tyr Gly Leu Gln Arg Arg Thr Ile Arg Gly Phe Arg Glu Leu Arg Arg

450 455 460
Ile Trp Arg Gly His Ala Pro Ala Gly Thr Val Phe Ser Ser Glu Leu
465 470 475 480
Lys Glu Lys Leu Ala Gly Glu Leu Arg Gln Phe Gln Thr Asp Asn Ser
485 490 495
Thr Thr Ile Gly Ser Val Gln Leu Phe Asn Glu Leu Ile Gln Asn Pro
500 505 510
Lys Tyr Trp Pro Ile Trp Gln Ala Pro Asp Val Glu Thr Ala Arg Gln

515 520 525
Trp Ala Asp Ala Gly Phe Ala Asp Asp Pro Leu Ala Ala Leu Val Gln
530 535 540
Glu Ala Glu Leu Gln Glu Asp Ile Asp Ala Leu Lys Ala Pro Val Lys
545 550 555 560
Leu Thr Pro Ala Asp Pro Glu Tyr Ser Arg Arg Gln Tyr Asp Phe Asn
565 570 575
Ala Val Ser Lys Phe Gly Ala Gly Ser Arg Ser Ala Asn Arg His Glu

580 585 590
Pro Gly Gln Thr Glu Arg Gly His Asn Thr Phe Thr Thr Glu Ile Ala
595 600 605
Ala Arg Asn Ala Ala Asp Gly Asn Arg Trp Arg Ala Thr His Val Arg
610 615 620
Ile His Tyr Ser Ala Pro Arg Leu Leu Arg Asp Gly Leu Arg Arg Pro
625 630 635 640
Asp Thr Asp Gly Asn Glu Ala Leu Glu Ala Val Pro Trp Leu Gln Pro

645 650 655

Met Met Glu Ala Leu Ala Pro Leu Pro Thr Leu Pro Gln Asp Leu Thr

660 665 670

Gly Met Pro Val Phe Leu Met Pro Asp Val Thr Leu Ser Gly Glu Arg

675 680 685

Arg Ile Leu Leu Asn Leu Pro Val Thr Leu Glu Pro Ala Ala Leu Val

690 695 700

Glu Gln Leu Gly Asn Ala Gly Arg Trp Gln Asn Gln Phe Phe Gly Ser

705 710 715 720

Arg Glu Asp Pro Phe Ala Leu Arg Trp Pro Ala Asp Gly Ala Val Lys

725 730 735

Thr Ala Lys Gly Lys Thr His Ile Pro Trp His Gln Asp Arg Asp His

740 745 750

Phe Thr Val Leu Gly Val Asp Leu Gly Thr Arg Asp Ala Gly Ala Leu

755 760 765

Ala Leu Leu Asn Val Thr Ala Gln Lys Pro Ala Lys Pro Val His Arg

770 775 780

Ile Ile Gly Glu Ala Asp Gly Arg Thr Trp Tyr Ala Ser Leu Ala Asp

785 790 795 800

Ala Arg Met Ile Arg Leu Pro Gly Glu Asp Ala Arg Leu Phe Val Arg

805 810 815

Gly Lys Leu Val Gln Glu Pro Tyr Gly Glu Arg Gly Arg Asn Ala Ser

820 825 830

Leu Leu Glu Trp Glu Asp Ala Arg Asn Ile Ile Leu Arg Leu Gly Gln

835 840 845

Asn Pro Asp Glu Leu Leu Gly Ala Asp Pro Arg Arg His Ser Tyr Pro

850 855 860

Glu Ile Asn Asp Lys Leu Leu Val Ala Leu Arg Arg Ala Gln Ala Arg

865 870 875 880

Leu Ala Arg Leu Gln Asn Arg Ser Trp Arg Leu Arg Asp Leu Ala Glu

885 890 895

Ser Asp Lys Ala Leu Asp Glu Ile His Ala Glu Arg Ala Gly Glu Lys

900 905 910

Pro Ser Pro Leu Pro Pro Leu Ala Arg Asp Asp Ala Ile Lys Ser Thr

915 920 925

Asp Glu Ala Leu Leu Ser Gln Arg Asp Ile Ile Arg Arg Ser Phe Val

930 935 940

Gln Ile Ala Asn Leu Ile Leu Pro Leu Arg Gly Arg Arg Trp Glu Trp

945 950 955 960

Arg Pro His Val Glu Val Pro Asp Cys His Ile Leu Ala Gln Ser Asp

965 970 975

Pro Gly Thr Asp Asp Thr Lys Arg Leu Val Ala Gly Gln Arg Gly Ile

980 985 990

Ser His Glu Arg Ile Glu Gln Ile Glu Glu Leu Arg Arg Arg Cys Gln

995 1000 1005

Ser Leu Asn Arg Ala Leu Arg His Lys Pro Gly Glu Arg Pro Val

1010 1015 1020

Leu Gly Arg Pro Ala Lys Gly Glu Glu Ile Ala Asp Pro Cys Pro

1025 1030 1035

Ala Leu Leu Glu Lys Ile Asn Arg Leu Arg Asp Gln Arg Val Asp

1040 1045 1050

Gln Thr Ala His Ala Ile Leu Ala Ala Ala Leu Gly Val Arg Leu

1055 1060 1065

Arg Ala Pro Ser Lys Asp Arg Ala Glu Arg Arg His Arg Asp Ile

1070 1075 1080

His Gly Glu Tyr Glu Arg Phe Arg Ala Pro Ala Asp Phe Val Val

1085 1090 1095

Ile Glu Asn Leu Ser Arg Tyr Leu Ser Ser Gln Asp Arg Ala Arg

1100 1105 1110

Ser Glu Asn Thr Arg Leu Met Gln Trp Cys His Arg Gln Ile Val

1115 1120 1125

Gln Lys Leu Arg Gln Leu Cys Glu Thr Tyr Gly Ile Pro Val Leu

1130 1135 1140

Ala Val Pro Ala Ala Tyr Ser Ser Arg Phe Ser Ser Arg Asp Gly

1145 1150 1155

Ser Ala Gly Phe Arg Ala Val His Leu Thr Pro Asp His Arg His

1160 1165 1170

Arg Met Pro Trp Ser Arg Ile Leu Ala Arg Leu Lys Ala His Glu

1175 1180 1185

Glu Asp Gly Lys Arg Leu Glu Lys Thr Val Leu Asp Glu Ala Arg

1190 1195 1200

Ala Val Arg Gly Leu Phe Asp Arg Leu Asp Arg Phe Asn Ala Gly

1205 1210 1215

His Val Pro Gly Lys Pro Trp Arg Thr Leu Leu Ala Pro Leu Pro

1220 1225 1230

Gly Gly Pro Val Phe Val Pro Leu Gly Asp Ala Thr Pro Met Gln

1235 1240 1245

Ala Asp Leu Asn Ala Ala Ile Asn Ile Ala Leu Arg Gly Ile Ala

1250 1255 1260

Ala Pro Asp Arg His Asp Ile His His Arg Leu Arg Ala Glu Asn

1265 1270 1275

Lys Lys Arg Ile Leu Ser Leu Arg Leu Gly Thr Gln Arg Glu Lys

1280 1285 1290

Ala Arg Trp Pro Gly Gly Ala Pro Ala Val Thr Leu Ser Thr Pro

1295 1300 1305

Asn Asn Gly Ala Ser Pro Glu Asp Ser Asp Ala Leu Pro Glu Arg

1310 1315 1320

Val Ser Asn Leu Phe Val Asp Ile Ala Gly Val Ala Asn Phe Glu

1325 1330 1335

Arg Val Thr Ile Glu Gly Val Ser Gln Lys Phe Ala Thr Gly Arg

1340 1345 1350

Gly Leu Trp Ala Ser Val Lys Gln Arg Ala Trp Asn Arg Val Ala

1355 1360 1365

Arg Leu Asn Glu Thr Val Thr Asp Asn Asn Arg Asn Glu Glu Glu

1370 1375 1380

Asp Asp Ile Pro Met

1385

<210> 501

<211> 162

<212> DNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 501

caacatgctc gctttgcgaa ggctgacggc ccgctctcat ttggcattgc cgggagccgg 60

agttttcgga agagagtgtc gacgactgct gatctccgca tccgcgtcct gttcgccagg 120

ccgggtcggg tgtacggatc atgctggcag cagtctacgc cg 162

<210> 502

<211> 60

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 502

cggaagagag ugucgacgac ugcugaucuc cgcauccgcg uccuguucgc caggccgggu 60

<210> 503

<211> 62

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 503

ccaacaugcu cgcuuugcga aggcugacgg cccgcucuca uuuggcauug ccgggagccg 60

ga 62

<210> 504

<211> 118

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 504

ucgcuuugcg aaggcugacg gcccgcucuc auuuggcauu gccgggagcc ggaguuuucg 60

gaagagagug ucgacgacug cugaucuccg cauccgcguc cuguucgcca ggccgggu 118

<210> 505

<211> 87

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 505

augcucgcuu ugcgaaggcu gacggcccgc ucucauuugg cauugccggg agccggaguu 60

uucggaagag agugucgacg acugcug 87

<210> 506

<211> 71

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 506

aguuuucgga agagaguguc gacgacugcu gaucuccgca uccgcguccu guucgccagg 60

ccgggucggg u 71

<210> 507

<211> 61

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 507

cgccuaucag ccaacaugcu cgcuuugcga aggcugacgg cccgcucuca uuuggcauug 60

c 61

<210> 508

<211> 87

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 508

cgccuaucag ccaacaugcu cgcuuugcga aggcugacgg cccgcucuca uuuggcauug 60

ccgggagccg gaguuuucgg aagagag 87

<210> 509

<211> 84

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 509

ugaucuccgc auccgcgucc uguucgccag gccgggucgg guguacggau caugcuggca 60

gcagucuacg ccgagaacau ucgc 84

<210> 510

<211> 84

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 510

ugaucuccgc auccgcgucc uguucgccag gccgggucgg guguacggau caugcuggca 60

gcagucuacg ccgagaacau ucgc 84

<210> 511

<211> 83

<212> RNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 511

uccgcauccg cguccuguuc gccaggccgg gucgggugua cggaucaugc uggcagcagu 60

cuacgccgag aacauucgcu uuu 83

<210> 512

<211> 32

<212> DNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 512

ggcgcaaccc gcacacaacc gcgaatggac ac 32

<210> 513

<211> 14

<212> DNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 513

ccgcgaatgg acac 14

<210> 514

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 514

uggguuugggu uggguuggga 20

<210> 515

<211> 23

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 515

uggguuuggg uuuggguuug gga 23

<210> 516

<211> 1129

<212> PRT

<213> Alicyclobacillus acidoterrestris

<400> 516

Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Asp Asp Met Pro
1 5 10 15

Glu Ile Arg Ala Gly Leu Trp Lys Leu His Lys Glu Val Asn Ala Gly
20 25 30

Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu
35 40 45

Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Asp Lys Thr

50 55 60

Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln
65 70 75 80
Val Glu Asn Gly His Arg Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu
85 90 95
Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly
100 105 110
Ala Lys Gly Asp Ala Gln Gln Ile Ala Arg Lys Phe Leu Ser Pro Leu

115 120 125

Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn
130 135 140
Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu
145 150 155 160
Glu Glu Lys Glu Lys Ala Glu Thr Arg Lys Ser Ala Asp Arg Thr Ala
165 170 175
Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg

180 185 190

Val Tyr Thr Asp Ser Glu Met Ser Ser Val Glu Trp Lys Pro Leu Arg
195 200 205
Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala
210 215 220
Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Gln
225 230 235 240
Glu Tyr Ala Lys Leu Val Glu Gln Lys Asn Arg Phe Glu Gln Lys Asn

245 250 255

Phe Val Gly Gln Glu His Leu Val His Leu Val Asn Gln Leu Gln Gln
260 265 270
Asp Met Lys Glu Ala Ser Pro Gly Leu Glu Ser Lys Glu Gln Thr Ala
275 280 285
His Tyr Val Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu

290 295 300

Lys Trp Gly Lys Leu Ala Pro Asp Ala Pro Phe Asp Leu Tyr Asp Ala

305 310 315 320

Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His

325 330 335

Asp Leu Phe Ala Lys Leu Ala Glu Pro Glu Tyr Gln Ala Leu Trp Arg

340 345 350

Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Leu

355 360 365

Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp

370 375 380

Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn

385 390 395 400

Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Arg Arg His

405 410 415

Ala Ile Arg Phe His Lys Leu Leu Lys Val Glu Asn Gly Val Ala Arg

420 425 430

Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Glu Gln Leu Asp

435 440 445

Asn Leu Leu Pro Arg Asp Pro Asn Glu Pro Ile Ala Leu Tyr Phe Arg

450 455 460

Asp Tyr Gly Ala Glu Gln His Phe Thr Gly Glu Phe Gly Gly Ala Lys

465 470 475 480

Ile Gln Cys Arg Arg Asp Gln Leu Ala His Met His Arg Arg Arg Gly

485 490 495

Ala Arg Asp Val Tyr Leu Asn Val Ser Val Arg Val Gln Ser Gln Ser

500 505 510

Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu

515 520 525

Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp

530 535 540

Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu
545 550 555 560
Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser

565 570 575
Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro
580 585 590
Asn Ser Lys Gly Arg Val Pro Phe Phe Phe Pro Ile Lys Gly Asn Asp
595 600 605
Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly
610 615 620
Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg

625 630 635 640
Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val
645 650 655
Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys
660 665 670
Leu Ile Glu Gln Pro Val Asp Ala Ala Asn His Met Thr Pro Asp Trp
675 680 685
Arg Glu Ala Phe Glu Asn Glu Leu Gln Lys Leu Lys Ser Leu His Gly

690 695 700
Ile Cys Ser Asp Lys Glu Trp Met Asp Ala Val Tyr Glu Ser Val Arg
705 710 715 720
Arg Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp
725 730 735
Val Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Ala Lys Asp Val
740 745 750
Val Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr

755 760 765
Lys Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val
770 775 780
Ile Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His

785 790 795 800

Ile Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile

805 810 815

Ile Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Glu Arg Gly Lys

820 825 830

Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu

835 840 845

Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn

850 855 860

Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Ile

865 870 875 880

Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala

885 890 895

Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys

900 905 910

Arg Arg Val Pro Ala Arg Cys Thr Gln Glu His Asn Pro Glu Pro Phe

915 920 925

Pro Trp Trp Leu Asn Lys Phe Val Val Glu His Thr Leu Asp Ala Cys

930 935 940

Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Ile Phe

945 950 955 960

Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala

965 970 975

Asp Leu Asn Ala Ala Gln Asn Leu Gln Gln Arg Leu Trp Ser Asp Phe

980 985 990

Asp Ile Ser Gln Ile Arg Leu Arg Cys Asp Trp Gly Glu Val Asp Gly

995 1000 1005

Glu Leu Val Leu Ile Pro Arg Leu Thr Gly Lys Arg Thr Ala Asp

1010 1015 1020

Ser Tyr Ser Asn Lys Val Phe Tyr Thr Asn Thr Gly Val Thr Tyr

1025 1030 1035

Tyr Glu Arg Glu Arg Gly Lys Lys Arg Arg Lys Val Phe Ala Gln
1040 1045 1050
Glu Lys Leu Ser Glu Glu Glu Ala Glu Leu Leu Val Glu Ala Asp
1055 1060 1065
Glu Ala Arg Glu Lys Ser Val Val Leu Met Arg Asp Pro Ser Gly

1070 1075 1080
Ile Ile Asn Arg Gly Asn Trp Thr Arg Gln Lys Glu Phe Trp Ser
1085 1090 1095
Met Val Asn Gln Arg Ile Glu Gly Tyr Leu Val Lys Gln Ile Arg
1100 1105 1110
Ser Arg Val Pro Leu Gln Asp Ser Ala Cys Glu Asn Thr Gly Asp
1115 1120 1125
Ile

<210> 517
<211> 1147
<212> PRT
<
213> Alicyclobacillus kakegawensis
<400> 517
Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Ser Glu Cys Pro
1 5 10 15
Asp Ile Leu Ala Gly Met Trp Gln Leu His Arg Ala Thr Asn Ala Gly
20 25 30
Val Arg Tyr Tyr Thr Glu Trp Val Ser Leu Met Arg Gln Glu Ile Leu
35 40 45
Tyr Ser Arg Gly Pro Asp Gly Gly Gln Gln Cys Tyr Met Thr Ala Glu

50 55 60
Asp Cys Gln Arg Glu Leu Leu Arg Arg Leu Arg Asn Arg Gln Leu His
65 70 75 80
Asn Gly Arg Gln Asp Gln Pro Gly Thr Asp Ala Asp Leu Leu Ala Ile
85 90 95
Ser Arg Arg Leu Tyr Glu Ile Leu Val Leu Gln Ser Ile Gly Lys Arg

100 105 110

Gly Asp Ala Gln Gln Ile Ala Ser Ser Phe Leu Ser Pro Leu Val Asp

115 120 125

Pro Asn Ser Lys Gly Gly Arg Gly Glu Ala Lys Ser Gly Arg Lys Pro

130 135 140

Ala Trp Gln Lys Met Arg Asp Gln Gly Asp Pro Arg Trp Val Ala Ala

145 150 155 160

Arg Glu Lys Tyr Glu Gln Arg Lys Ala Val Asp Pro Ser Lys Glu Ile

165 170 175

Leu Asn Ser Leu Asp Ala Leu Gly Leu Arg Pro Leu Phe Ala Val Phe

180 185 190

Thr Glu Thr Tyr Arg Ser Gly Val Asp Trp Lys Pro Leu Gly Lys Ser

195 200 205

Gln Gly Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala Leu Glu

210 215 220

Arg Leu Met Ser Trp Glu Ser Trp Asn Arg Arg Val Gly Glu Glu Tyr

225 230 235 240

Ala Arg Leu Phe Gln Gln Lys Met Lys Phe Glu Gln Glu His Phe Ala

245 250 255

Glu Gln Ser His Leu Val Lys Leu Ala Arg Ala Leu Glu Ala Asp Met

260 265 270

Arg Ala Ala Ser Gln Gly Phe Glu Ala Lys Arg Gly Thr Ala His Gln

275 280 285

Ile Thr Arg Arg Ala Leu Arg Gly Ala Asp Arg Val Phe Glu Ile Trp

290 295 300

Lys Ser Ile Pro Glu Glu Ala Leu Phe Ser Gln Tyr Asp Glu Val Ile

305 310 315 320

Arg Gln Val Gln Ala Glu Lys Arg Arg Asp Phe Gly Ser His Asp Leu

325 330 335

Phe Ala Lys Leu Ala Glu Pro Lys Tyr Gln Pro Leu Trp Arg Ala Asp

340 345 350

Glu Thr Phe Leu Thr Arg Tyr Ala Leu Tyr Asn Gly Val Leu Arg Asp
355 360 365
Leu Glu Lys Ala Arg Gln Phe Ala Thr Phe Thr Leu Pro Asp Ala Cys

370 375 380
Val Asn Pro Ile Trp Thr Arg Phe Glu Ser Ser Gln Gly Ser Asn Leu
385 390 395 400
His Lys Tyr Glu Phe Leu Phe Asp His Leu Gly Pro Gly Arg His Ala
405 410 415
Val Arg Phe Gln Arg Leu Leu Val Val Glu Ser Glu Gly Ala Lys Glu
420 425 430
Arg Asp Ser Val Val Val Pro Val Ala Pro Ser Gly Gln Leu Asp Lys

435 440 445
Leu Val Leu Arg Glu Glu Glu Lys Ser Ser Val Ala Leu His Leu His
450 455 460
Asp Thr Ala Arg Pro Asp Gly Phe Met Ala Glu Trp Ala Gly Ala Lys
465 470 475 480
Leu Gln Tyr Glu Arg Ser Thr Leu Ala Arg Lys Ala Arg Arg Asp Lys
485 490 495
Gln Gly Met Arg Ser Trp Arg Arg Gln Pro Ser Met Leu Met Ser Ala

500 505 510
Ala Gln Met Leu Glu Asp Ala Lys Gln Ala Gly Asp Val Tyr Leu Asn
515 520 525
Ile Ser Val Arg Val Lys Ser Pro Ser Glu Val Arg Gly Gln Arg Arg
530 535 540
Pro Pro Tyr Ala Ala Leu Phe Arg Ile Asp Asp Lys Gln Arg Arg Val
545 550 555 560
Thr Val Asn Tyr Asn Lys Leu Ser Ala Tyr Leu Glu Glu His Pro Asp

565 570 575
Lys Gln Ile Pro Gly Ala Pro Gly Leu Leu Ser Gly Leu Arg Val Met
580 585 590
Ser Val Asp Leu Gly Leu Arg Thr Ser Ala Ser Ile Ser Val Phe Arg

595 600 605

Val Ala Lys Lys Glu Glu Val Glu Ala Leu Gly Asp Gly Arg Pro Pro

610 615 620

His Tyr Tyr Pro Ile His Gly Thr Asp Asp Leu Val Ala Val His Glu

625 630 635 640

Arg Ser His Leu Ile Gln Met Pro Gly Glu Thr Glu Thr Lys Gln Leu

645 650 655

Arg Lys Leu Arg Glu Glu Arg Gln Ala Val Leu Arg Pro Leu Phe Ala

660 665 670

Gln Leu Ala Leu Leu Arg Leu Leu Val Arg Cys Gly Ala Ala Asp Glu

675 680 685

Arg Ile Arg Thr Arg Ser Trp Gln Arg Leu Thr Lys Gln Gly Arg Glu

690 695 700

Phe Thr Lys Arg Leu Thr Pro Ser Trp Arg Glu Ala Leu Glu Leu Glu

705 710 715 720

Leu Thr Arg Leu Glu Ala Tyr Cys Gly Arg Val Pro Asp Asp Glu Trp

725 730 735

Ser Arg Ile Val Asp Arg Thr Val Ile Ala Leu Trp Arg Arg Met Gly

740 745 750

Lys Gln Val Arg Asp Trp Arg Lys Gln Val Lys Ser Gly Ala Lys Val

755 760 765

Lys Val Lys Gly Tyr Gln Leu Asp Val Val Gly Gly Asn Ser Leu Ala

770 775 780

Gln Ile Asp Tyr Leu Glu Gln Gln Tyr Lys Phe Leu Arg Arg Trp Ser

785 790 795 800

Phe Phe Ala Arg Ala Ser Gly Leu Val Val Arg Ala Asp Arg Glu Ser

805 810 815

His Phe Ala Val Ala Leu Arg Gln His Ile Glu Asn Ala Lys Arg Asp

820 825 830

Arg Leu Lys Lys Leu Ala Asp Arg Ile Leu Met Glu Ala Leu Gly Tyr

835 840 845

Val Tyr Glu Ala Ser Gly Pro Arg Glu Gly Gln Trp Thr Ala Gln His
850 855 860
Pro Pro Cys Gln Leu Ile Ile Leu Glu Glu Leu Ser Ala Tyr Arg Phe
865 870 875 880
Ser Asp Asp Arg Pro Pro Ser Glu Asn Ser Lys Leu Met Ala Trp Gly

885 890 895
His Arg Gly Ile Leu Glu Glu Leu Val Asn Gln Ala Gln Val His Asp
900 905 910
Val Leu Val Gly Thr Val Tyr Ala Ala Phe Ser Ser Arg Phe Asp Ala
915 920 925
Arg Thr Gly Ala Pro Gly Val Arg Cys Arg Arg Val Pro Ala Arg Phe
930 935 940
Val Gly Ala Thr Val Asp Asp Ser Leu Pro Leu Trp Leu Thr Glu Phe

945 950 955 960
Leu Asp Lys His Arg Leu Asp Lys Asn Leu Leu Arg Pro Asp Asp Val
965 970 975
Ile Pro Thr Gly Glu Gly Glu Phe Leu Val Ser Pro Cys Gly Glu Glu
980 985 990
Ala Ala Arg Val Arg Gln Val His Ala Asp Ile Asn Ala Ala Gln Asn
995 1000 1005
Leu Gln Arg Arg Leu Trp Gln Asn Phe Asp Ile Thr Glu Leu Arg

1010 1015 1020
Leu Arg Cys Asp Val Lys Met Gly Gly Glu Gly Thr Val Leu Val
1025 1030 1035
Pro Arg Val Asn Asn Ala Arg Ala Lys Gln Leu Phe Gly Lys Lys
1040 1045 1050
Val Leu Val Ser Gln Asp Gly Val Thr Phe Phe Glu Arg Ser Gln
1055 1060 1065
Thr Gly Gly Lys Pro His Ser Glu Lys Gln Thr Asp Leu Thr Asp

1070 1075 1080
Lys Glu Leu Glu Leu Ile Ala Glu Ala Asp Glu Ala Arg Ala Lys

1085 1090 1095

Ser Val Val Leu Phe Arg Asp Pro Ser Gly His Ile Gly Lys Gly

1100 1105 1110

His Trp Ile Arg Gln Arg Glu Phe Trp Ser Leu Val Lys Gln Arg

1115 1120 1125

Ile Glu Ser His Thr Ala Glu Arg Ile Arg Val Arg Gly Val Gly

1130 1135 1140

Ser Ser Leu Asp

1145

<210> 518

<211> 1146

<212> PRT

<213> Alicyclobacillus macrosporangiidus

<400> 518

Met Asn Val Ala Val Lys Ser Ile Lys Val Lys Leu Met Leu Gly His

1 5 10 15

Leu Pro Glu Ile Arg Glu Gly Leu Trp His Leu His Glu Ala Val Asn

20 25 30

Leu Gly Val Arg Tyr Tyr Thr Glu Trp Leu Ala Leu Leu Arg Gln Gly

35 40 45

Asn Leu Tyr Arg Arg Gly Lys Asp Gly Ala Gln Glu Cys Tyr Met Thr

50 55 60

Ala Glu Gln Cys Arg Gln Glu Leu Leu Val Arg Leu Arg Asp Arg Gln

65 70 75 80

Lys Arg Asn Gly His Thr Gly Asp Pro Gly Thr Asp Glu Glu Leu Leu

85 90 95

Gly Val Ala Arg Arg Leu Tyr Glu Leu Leu Val Pro Gln Ser Val Gly

100 105 110

Lys Lys Gly Gln Ala Gln Met Leu Ala Ser Gly Phe Leu Ser Pro Leu

115 120 125

Ala Asp Pro Lys Ser Glu Gly Gly Lys Gly Thr Ser Lys Ser Gly Arg

130 135 140

Lys Pro Ala Trp Met Gly Met Lys Glu Ala Gly Asp Ser Arg Trp Val
145 150 155 160
Glu Ala Lys Ala Arg Tyr Glu Ala Asn Lys Ala Lys Asp Pro Thr Lys

165 170 175
Gln Val Ile Ala Ser Leu Glu Met Tyr Gly Leu Arg Pro Leu Phe Asp
180 185 190
Val Phe Thr Glu Thr Tyr Lys Thr Ile Arg Trp Met Pro Leu Gly Lys
195 200 205
His Gln Gly Val Arg Ala Trp Asp Arg Asp Met Phe Gln Gln Ser Leu
210 215 220
Glu Arg Leu Met Ser Trp Glu Ser Trp Asn Glu Arg Val Gly Ala Glu

225 230 235 240
Phe Ala Arg Leu Val Asp Arg Arg Asp Arg Phe Arg Glu Lys His Phe
245 250 255
Thr Gly Gln Glu His Leu Val Ala Leu Ala Gln Arg Leu Glu Gln Glu
260 265 270
Met Lys Glu Ala Ser Pro Gly Phe Glu Ser Lys Ser Ser Gln Ala His
275 280 285
Arg Ile Thr Lys Arg Ala Leu Arg Gly Ala Asp Gly Ile Ile Asp Asp

290 295 300
Trp Leu Lys Leu Ser Glu Gly Glu Pro Val Asp Arg Phe Asp Glu Ile
305 310 315 320
Leu Arg Lys Arg Gln Ala Gln Asn Pro Arg Arg Phe Gly Ser His Asp
325 330 335
Leu Phe Leu Lys Leu Ala Glu Pro Val Phe Gln Pro Leu Trp Arg Glu
340 345 350
Asp Pro Ser Phe Leu Ser Arg Trp Ala Ser Tyr Asn Glu Val Leu Asn

355 360 365
Lys Leu Glu Asp Ala Lys Gln Phe Ala Thr Phe Thr Leu Pro Ser Pro
370 375 380
Cys Ser Asn Pro Val Trp Ala Arg Phe Glu Asn Ala Glu Gly Thr Asn

385 390 395 400

Ile Phe Lys Tyr Asp Phe Leu Phe Asp His Phe Gly Lys Gly Arg His

405 410 415

Gly Val Arg Phe Gln Arg Met Ile Val Met Arg Asp Gly Val Pro Thr

420 425 430

Glu Val Glu Gly Ile Val Val Pro Ile Ala Pro Ser Arg Gln Leu Asp

435 440 445

Ala Leu Ala Pro Asn Asp Ala Ala Ser Pro Ile Asp Val Phe Val Gly

450 455 460

Asp Pro Ala Ala Pro Gly Ala Phe Arg Gly Gln Phe Gly Gly Ala Lys

465 470 475 480

Ile Gln Tyr Arg Arg Ser Ala Leu Val Arg Lys Gly Arg Arg Glu Glu

485 490 495

Lys Ala Tyr Leu Cys Gly Phe Arg Leu Pro Ser Gln Arg Arg Thr Gly

500 505 510

Thr Pro Ala Asp Asp Ala Gly Glu Val Phe Leu Asn Leu Ser Leu Arg

515 520 525

Val Glu Ser Gln Ser Glu Gln Ala Gly Arg Arg Asn Pro Pro Tyr Ala

530 535 540

Ala Val Phe His Ile Ser Asp Gln Thr Arg Arg Val Ile Val Arg Tyr

545 550 555 560

Gly Glu Ile Glu Arg Tyr Leu Ala Glu His Pro Asp Thr Gly Ile Pro

565 570 575

Gly Ser Arg Gly Leu Thr Ser Gly Leu Arg Val Met Ser Val Asp Leu

580 585 590

Gly Leu Arg Thr Ser Ala Ala Ile Ser Val Phe Arg Val Ala His Arg

595 600 605

Asp Glu Leu Thr Pro Asp Ala His Gly Arg Gln Pro Phe Phe Phe Pro

610 615 620

Ile His Gly Met Asp His Leu Val Ala Leu His Glu Arg Ser His Leu

625 630 635 640

Ile Arg Leu Pro Gly Glu Thr Glu Ser Lys Lys Val Arg Ser Ile Arg
645 650 655
Glu Gln Arg Leu Asp Arg Leu Asn Arg Leu Arg Ser Gln Met Ala Ser
660 665 670
Leu Arg Leu Leu Val Arg Thr Gly Val Leu Asp Glu Gln Lys Arg Asp

675 680 685
Arg Asn Trp Glu Arg Leu Gln Ser Ser Met Glu Arg Gly Gly Glu Arg
690 695 700
Met Pro Ser Asp Trp Trp Asp Leu Phe Gln Ala Gln Val Arg Tyr Leu
705 710 715 720
Ala Gln His Arg Asp Ala Ser Gly Glu Ala Trp Gly Arg Met Val Gln
725 730 735
Ala Ala Val Arg Thr Leu Trp Arg Gln Leu Ala Lys Gln Val Arg Asp

740 745 750
Trp Arg Lys Glu Val Arg Arg Asn Ala Asp Lys Val Lys Ile Arg Gly
755 760 765
Ile Ala Arg Asp Val Pro Gly Gly His Ser Leu Ala Gln Leu Asp Tyr
770 775 780
Leu Glu Arg Gln Tyr Arg Phe Leu Arg Ser Trp Ser Ala Phe Ser Val
785 790 795 800
Gln Ala Gly Gln Val Val Arg Ala Glu Arg Asp Ser Arg Phe Ala Val

805 810 815
Ala Leu Arg Glu His Ile Asp Asn Gly Lys Lys Asp Arg Leu Lys Lys
820 825 830
Leu Ala Asp Arg Ile Leu Met Glu Ala Leu Gly Tyr Val Tyr Val Thr
835 840 845
Asp Gly Arg Arg Ala Gly Gln Trp Gln Ala Val Tyr Pro Pro Cys Gln
850 855 860
Leu Val Leu Leu Glu Glu Leu Ser Glu Tyr Arg Phe Ser Asn Asp Arg

865 870 875 880
Pro Pro Ser Glu Asn Ser Gln Leu Met Val Trp Ser His Arg Gly Val

885 890 895

Leu Glu Glu Leu Ile His Gln Ala Gln Val His Asp Val Leu Val Gly

900 905 910

Thr Ile Pro Ala Ala Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala

915 920 925

Pro Gly Ile Arg Cys Arg Arg Val Pro Ser Ile Pro Leu Lys Asp Ala

930 935 940

Pro Ser Ile Pro Ile Trp Leu Ser His Tyr Leu Lys Gln Thr Glu Arg

945 950 955 960

Asp Ala Ala Ala Leu Arg Pro Gly Glu Leu Ile Pro Thr Gly Asp Gly

965 970 975

Glu Phe Leu Val Thr Pro Ala Gly Arg Gly Ala Ser Gly Val Arg Val

980 985 990

Val His Ala Asp Ile Asn Ala Ala His Asn Leu Gln Arg Arg Leu Trp

995 1000 1005

Glu Asn Phe Asp Leu Ser Asp Ile Arg Val Arg Cys Asp Arg Arg

1010 1015 1020

Glu Gly Lys Asp Gly Thr Val Val Leu Ile Pro Arg Leu Thr Asn

1025 1030 1035

Gln Arg Val Lys Glu Arg Tyr Ser Gly Val Ile Phe Thr Ser Glu

1040 1045 1050

Asp Gly Val Ser Phe Thr Val Gly Asp Ala Lys Thr Arg Arg Arg

1055 1060 1065

Ser Ser Ala Ser Gln Gly Glu Gly Asp Asp Leu Ser Asp Glu Glu

1070 1075 1080

Gln Glu Leu Leu Ala Glu Ala Asp Asp Ala Arg Glu Arg Ser Val

1085 1090 1095

Val Leu Phe Arg Asp Pro Ser Gly Phe Val Asn Gly Gly Arg Trp

1100 1105 1110

Thr Ala Gln Arg Ala Phe Trp Gly Met Val His Asn Arg Ile Glu

1115 1120 1125

Thr Leu Leu Ala Glu Arg Phe Ser Val Ser Gly Ala Ala Glu Lys
1130 1135 1140

Val Arg Gly
1145

<210> 519

<211> 1108

<212> PRT

<213> Bacillus hisashii

<400> 519

Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
1 5 10 15

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
20 25 30

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
35 40 45

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
50 55 60

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
65 70 75 80

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
85 90 95

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
100 105 110

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
115 120 125

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
130 135 140

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
145 150 155 160

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
165 170 175

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp

180 185 190

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg

195 200 205

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

210 215 220

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu

225 230 235 240

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys

245 250 255

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg

260 265 270

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu

275 280 285

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu

290 295 300

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys

305 310 315 320

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr

325 330 335

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro

340 345 350

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

355 360 365

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn

370 375 380

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn

385 390 395 400

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys

405 410 415

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

420 425 430

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
435 440 445

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
450 455 460

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
465 470 475 480

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
485 490 495

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
500 505 510

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
515 520 525

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
530 535 540

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
545 550 555 560

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
565 570 575

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
580 585 590

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
595 600 605

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
610 615 620

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
625 630 635 640

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
645 650 655

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
660 665 670

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu

675 680 685

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val

690 695 700

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys

705 710 715 720

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

725 730 735

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys

740 745 750

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg

755 760 765

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu

770 775 780

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

785 790 795 800

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln

805 810 815

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn

820 825 830

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met

835 840 845

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu

850 855 860

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg

865 870 875 880

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr

885 890 895

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu

900 905 910

Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu

915 920 925

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
930 935 940

Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
945 950 955 960

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
965 970 975

Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp
980 985 990

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu
995 1000 1005

Lys Asp Gly Val Tyr Glu Trp Val Asn Ala Gly Lys Leu Lys Ile
1010 1015 1020

Lys Lys Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser
1025 1030 1035

Asp Ile Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly
1040 1045 1050

Glu Lys Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro
1055 1060 1065

Ser Asp Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu
1070 1075 1080

Arg Ile Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr
1085 1090 1095

Ile Glu Asp Asp Ser Ser Lys Gln Ser Met
1100 1105

<210> 520

<211> 1108

<212

> PRT

<213> Bacillus species

<400> 520

Met Ala Ile Arg Ser Ile Lys Leu Lys Leu Lys Thr His Thr Gly Pro
1 5 10 15

Glu Ala Gln Asn Leu Arg Lys Gly Ile Trp Arg Thr His Arg Leu Leu

20 25 30

Asn Glu Gly Val Ala Tyr Tyr Met Lys Met Leu Leu Leu Phe Arg Gln

35 40 45

Glu Ser Thr Gly Glu Arg Pro Lys Glu Glu Leu Gln Glu Glu Leu Ile

50 55 60

Cys His Ile Arg Glu Gln Gln Gln Arg Asn Gln Ala Asp Lys Asn Thr

65 70 75 80

Gln Ala Leu Pro Leu Asp Lys Ala Leu Glu Ala Leu Arg Gln Leu Tyr

85 90 95

Glu Leu Leu Val Pro Ser Ser Val Gly Gln Ser Gly Asp Ala Gln Ile

100 105 110

Ile Ser Arg Lys Phe Leu Ser Pro Leu Val Asp Pro Asn Ser Glu Gly

115 120 125

Gly Lys Gly Thr Ser Lys Ala Gly Ala Lys Pro Thr Trp Gln Lys Lys

130 135 140

Lys Glu Ala Asn Asp Pro Thr Trp Glu Gln Asp Tyr Glu Lys Trp Lys

145 150 155 160

Lys Arg Arg Glu Glu Asp Pro Thr Ala Ser Val Ile Thr Thr Leu Glu

165 170 175

Glu Tyr Gly Ile Arg Pro Ile Phe Pro Leu Tyr Thr Asn Thr Val Thr

180 185 190

Asp Ile Ala Trp Leu Pro Leu Gln Ser Asn Gln Phe Val Arg Thr Trp

195 200 205

Asp Arg Asp Met Leu Gln Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu

210 215 220

Ser Trp Asn Lys Arg Val Gln Glu Glu Tyr Ala Lys Leu Lys Glu Lys

225 230 235 240

Met Ala Gln Leu Asn Glu Gln Leu Glu Gly Gly Gln Glu Trp Ile Ser

245 250 255

Leu Leu Glu Gln Tyr Glu Glu Asn Arg Glu Arg Glu Leu Arg Glu Asn

260 265 270

Met Thr Ala Ala Asn Asp Lys Tyr Arg Ile Thr Lys Arg Gln Met Lys
275 280 285
Gly Trp Asn Glu Leu Tyr Glu Leu Trp Ser Thr Phe Pro Ala Ser Ala
290 295 300
Ser His Glu Gln Tyr Lys Glu Ala Leu Lys Arg Val Gln Gln Arg Leu

305 310 315 320
Arg Gly Arg Phe Gly Asp Ala His Phe Phe Gln Tyr Leu Met Glu Glu
325 330 335
Lys Asn Arg Leu Ile Trp Lys Gly Asn Pro Gln Arg Ile His Tyr Phe
340 345 350
Val Ala Arg Asn Glu Leu Thr Lys Arg Leu Glu Glu Ala Lys Gln Ser
355 360 365
Ala Thr Met Thr Leu Pro Asn Ala Arg Lys His Pro Leu Trp Val Arg

370 375 380
Phe Asp Ala Arg Gly Gly Asn Leu Gln Asp Tyr Tyr Leu Thr Ala Glu
385 390 395 400
Ala Asp Lys Pro Arg Ser Arg Arg Phe Val Thr Phe Ser Gln Leu Ile
405 410 415
Trp Pro Ser Glu Ser Gly Trp Met Glu Lys Lys Asp Val Glu Val Glu
420 425 430
Leu Ala Leu Ser Arg Gln Phe Tyr Gln Gln Val Lys Leu Leu Lys Asn

435 440 445
Asp Lys Gly Lys Gln Lys Ile Glu Phe Lys Asp Lys Gly Ser Gly Ser
450 455 460
Thr Phe Asn Gly His Leu Gly Gly Ala Lys Leu Gln Leu Glu Arg Gly
465 470 475 480
Asp Leu Glu Lys Glu Glu Lys Asn Phe Glu Asp Gly Glu Ile Gly Ser
485 490 495
Val Tyr Leu Asn Val Val Ile Asp Phe Glu Pro Leu Gln Glu Val Lys

500 505 510
Asn Gly Arg Val Gln Ala Pro Tyr Gly Gln Val Leu Gln Leu Ile Arg

515 520 525

Arg Pro Asn Glu Phe Pro Lys Val Thr Thr Tyr Lys Ser Glu Gln Leu

530 535 540

Val Glu Trp Ile Lys Ala Ser Pro Gln His Ser Ala Gly Val Glu Ser

545 550 555 560

Leu Ala Ser Gly Phe Arg Val Met Ser Ile Asp Leu Gly Leu Arg Ala

565 570 575

Ala Ala Ala Thr Ser Ile Phe Ser Val Glu Glu Ser Ser Asp Lys Asn

580 585 590

Ala Ala Asp Phe Ser Tyr Trp Ile Glu Gly Thr Pro Leu Val Ala Val

595 600 605

His Gln Arg Ser Tyr Met Leu Arg Leu Pro Gly Glu Gln Val Glu Lys

610 615 620

Gln Val Met Glu Lys Arg Asp Glu Arg Phe Gln Leu His Gln Arg Val

625 630 635 640

Lys Phe Gln Ile Arg Val Leu Ala Gln Ile Met Arg Met Ala Asn Lys

645 650 655

Gln Tyr Gly Asp Arg Trp Asp Glu Leu Asp Ser Leu Lys Gln Ala Val

660 665 670

Glu Gln Lys Lys Ser Pro Leu Asp Gln Thr Asp Arg Thr Phe Trp Glu

675 680 685

Gly Ile Val Cys Asp Leu Thr Lys Val Leu Pro Arg Asn Glu Ala Asp

690 695 700

Trp Glu Gln Ala Val Val Gln Ile His Arg Lys Ala Glu Glu Tyr Val

705 710 715 720

Gly Lys Ala Val Gln Ala Trp Arg Lys Arg Phe Ala Ala Asp Glu Arg

725 730 735

Lys Gly Ile Ala Gly Leu Ser Met Trp Asn Ile Glu Glu Leu Glu Gly

740 745 750

Leu Arg Lys Leu Leu Ile Ser Trp Ser Arg Arg Thr Arg Asn Pro Gln

755 760 765

Glu Val Asn Arg Phe Glu Arg Gly His Thr Ser His Gln Arg Leu Leu
770 775 780
Thr His Ile Gln Asn Val Lys Glu Asp Arg Leu Lys Gln Leu Ser His
785 790 795 800
Ala Ile Val Met Thr Ala Leu Gly Tyr Val Tyr Asp Glu Arg Lys Gln
805 810 815
Glu Trp Cys Ala Glu Tyr Pro Ala Cys Gln Val Ile Leu Phe Glu Asn

820 825 830
Leu Ser Gln Tyr Arg Ser Asn Leu Asp Arg Ser Thr Lys Glu Asn Ser
835 840 845
Thr Leu Met Lys Trp Ala His Arg Ser Ile Pro Lys Tyr Val His Met
850 855 860
Gln Ala Glu Pro Tyr Gly Ile Gln Ile Gly Asp Val Arg Ala Glu Tyr
865 870 875 880
Ser Ser Arg Phe Tyr Ala Lys Thr Gly Thr Pro Gly Ile Arg Cys Lys

885 890 895
Lys Val Arg Gly Gln Asp Leu Gln Gly Arg Arg Phe Glu Asn Leu Gln
900 905 910
Lys Arg Leu Val Asn Glu Gln Phe Leu Thr Glu Glu Gln Val Lys Gln
915 920 925
Leu Arg Pro Gly Asp Ile Val Pro Asp Asp Ser Gly Glu Leu Phe Met
930 935 940
Thr Leu Thr Asp Gly Ser Gly Ser Lys Glu Val Val Phe Leu Gln Ala

945 950 955 960
Asp Ile Asn Ala Ala His Asn Leu Gln Lys Arg Phe Trp Gln Arg Tyr
965 970 975
Asn Glu Leu Phe Lys Val Ser Cys Arg Val Ile Val Arg Asp Glu Glu
980 985 990
Glu Tyr Leu Val Pro Lys Thr Lys Ser Val Gln Ala Lys Leu Gly Lys
995 1000 1005
Gly Leu Phe Val Lys Lys Ser Asp Thr Ala Trp Lys Asp Val Tyr

1010 1015 1020

Val Trp Asp Ser Gln Ala Lys Leu Lys Gly Lys Thr Thr Phe Thr

1025 1030 1035

Glu Glu Ser Glu Ser Pro Glu Gln Leu Glu Asp Phe Gln Glu Ile

1040 1045 1050

Ile Glu Glu Ala Glu Glu Ala Lys Gly Thr Tyr Arg Thr Leu Phe

1055 1060 1065

Arg Asp Pro Ser Gly Val Phe Phe Pro Glu Ser Val Trp Tyr Pro

1070 1075 1080

Gln Lys Asp Phe Trp Gly Glu Val Lys Arg Lys Leu Tyr Gly Lys

1085 1090 1095

Leu Arg Glu Arg Phe Leu Thr Lys Ala Arg

1100 1105

<210> 521

<211> 1108

<212> PRT

<213> Bacillus thermoamylovorans

<400> 521

Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val

1 5 10 15

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala

20 25 30

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu

35 40 45

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu

50 55 60

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn

65 70 75 80

Ser Phe Thr His Glu Val Asp Lys Asp Val Val Phe Asn Ile Leu Arg

85 90 95

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu

100 105 110

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
115 120 125
Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
130 135 140
Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys

145 150 155 160
Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
165 170 175
Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Phe Thr Asp
180 185 190
Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
195 200 205
Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu

210 215 220
Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
225 230 235 240
Tyr Glu Lys Val Glu Lys Glu His Lys Thr Leu Glu Glu Arg Ile Lys
245 250 255
Glu Asp Ile Gln Ala Phe Lys Ser Leu Glu Gln Tyr Glu Lys Glu Arg
260 265 270
Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu

275 280 285
Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
290 295 300
Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
305 310 315 320
Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
325 330 335
Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro

340 345 350
Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys

355 360 365

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn

370 375 380

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn

385 390 395 400

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys

405 410 415

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly

420 425 430

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg

435 440 445

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His

450 455 460

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr

465 470 475 480

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr

485 490 495

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met

500 505 510

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys

515 520 525

Ile His Arg Asp Asp Phe Pro Lys Phe Val Asn Phe Lys Pro Lys Glu

530 535 540

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly

545 550 555 560

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly

565 570 575

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys

580 585 590

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu

595 600 605

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
610 615 620
Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
625 630 635 640
Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
645 650 655
Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp

660 665 670
Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
675 680 685
Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
690 695 700
Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
705 710 715 720
Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu

725 730 735
Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
740 745 750
Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
755 760 765
Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
770 775 780
Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile

785 790 795 800
Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
805 810 815
Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
820 825 830
Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
835 840 845
Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu

850 855 860

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg

865 870 875 880

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr

885 890 895

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu

900 905 910

Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu

915 920 925

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg

930 935 940

Lys Leu Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln

945 950 955 960

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys

965 970 975

Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp

980 985 990

Gln Lys Gln Lys Ile Ile Glu Glu Phe Gly Glu Gly Tyr Phe Ile Leu

995 1000 1005

Lys Asp Gly Val Tyr Glu Trp Gly Asn Ala Gly Lys Leu Lys Ile

1010 1015 1020

Lys Lys Gly Ser Ser Lys Gln Ser Ser Ser Glu Leu Val Asp Ser

1025 1030 1035

Asp Ile Leu Lys Asp Ser Phe Asp Leu Ala Ser Glu Leu Lys Gly

1040 1045 1050

Glu Lys Leu Met Leu Tyr Arg Asp Pro Ser Gly Asn Val Phe Pro

1055 1060 1065

Ser Asp Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu

1070 1075 1080

Arg Ile Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr

1085 1090 1095

Ile Glu Asp Asp Ser Ser Lys Gln Ser Met

1100 1105

<210> 522

<211> 1112

<212> PRT

<213> Bacillus species

<400> 522

Met Ala Ile Arg Ser Ile Lys Leu Lys Met Lys Thr Asn Ser Gly Thr
1 5 10 15

Asp Ser Ile Tyr Leu Arg Lys Ala Leu Trp Arg Thr His Gln Leu Ile
20 25 30

Asn Glu Gly Ile Ala Tyr Tyr Met Asn Leu Leu Thr Leu Tyr Arg Gln
35 40 45

Glu Ala Ile Gly Asp Lys Thr Lys Glu Ala Tyr Gln Ala Glu Leu Ile

50 55 60

Asn Ile Ile Arg Asn Gln Gln Arg Asn Asn Gly Ser Ser Glu Glu His
65 70 75 80

Gly Ser Asp Gln Glu Ile Leu Ala Leu Leu Arg Gln Leu Tyr Glu Leu
85 90 95

Ile Ile Pro Ser Ser Ile Gly Glu Ser Gly Asp Ala Asn Gln Leu Gly
100 105 110

Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn Ser Gln Ser Gly Lys

115 120 125

Gly Thr Ser Asn Ala Gly Arg Lys Pro Arg Trp Lys Arg Leu Lys Glu
130 135 140

Glu Gly Asn Pro Asp Trp Glu Leu Glu Lys Lys Lys Asp Glu Glu Arg
145 150 155 160

Lys Ala Lys Asp Pro Thr Val Lys Ile Phe Asp Asn Leu Asn Lys Tyr
165 170 175

Gly Leu Leu Pro Leu Phe Pro Leu Phe Thr Asn Ile Gln Lys Asp Ile

180 185 190

Glu Trp Leu Pro Leu Gly Lys Arg Gln Ser Val Arg Lys Trp Asp Lys

195 200 205

Asp Met Phe Ile Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu Ser Trp

210 215 220

Asn Arg Arg Val Ala Asp Glu Tyr Lys Gln Leu Lys Glu Lys Thr Glu

225 230 235 240

Ser Tyr Tyr Lys Glu His Leu Thr Gly Gly Glu Glu Trp Ile Glu Lys

245 250 255

Ile Arg Lys Phe Glu Lys Glu Arg Asn Met Glu Leu Glu Lys Asn Ala

260 265 270

Phe Ala Pro Asn Asp Gly Tyr Phe Ile Thr Ser Arg Gln Ile Arg Gly

275 280 285

Trp Asp Arg Val Tyr Glu Lys Trp Ser Lys Leu Pro Glu Ser Ala Ser

290 295 300

Pro Glu Glu Leu Trp Lys Val Val Ala Glu Gln Gln Asn Lys Met Ser

305 310 315 320

Glu Gly Phe Gly Asp Pro Lys Val Phe Ser Phe Leu Ala Asn Arg Glu

325 330 335

Asn Arg Asp Ile Trp Arg Gly His Ser Glu Arg Ile Tyr His Ile Ala

340 345 350

Ala Tyr Asn Gly Leu Gln Lys Lys Leu Ser Arg Thr Lys Glu Gln Ala

355 360 365

Thr Phe Thr Leu Pro Asp Ala Ile Glu His Pro Leu Trp Ile Arg Tyr

370 375 380

Glu Ser Pro Gly Gly Thr Asn Leu Asn Leu Phe Lys Leu Glu Glu Lys

385 390 395 400

Gln Lys Lys Asn Tyr Tyr Val Thr Leu Ser Lys Ile Ile Trp Pro Ser

405 410 415

Glu Glu Lys Trp Ile Glu Lys Glu Asn Ile Glu Ile Pro Leu Ala Pro

420 425 430

Ser Ile Gln Phe Asn Arg Gln Ile Lys Leu Lys Gln His Val Lys Gly

435 440 445

Lys Gln Glu Ile Ser Phe Ser Asp Tyr Ser Ser Arg Ile Ser Leu Asp
450 455 460
Gly Val Leu Gly Gly Ser Arg Ile Gln Phe Asn Arg Lys Tyr Ile Lys
465 470 475 480
Asn His Lys Glu Leu Leu Gly Glu Gly Asp Ile Gly Pro Val Phe Phe
485 490 495
Asn Leu Val Val Asp Val Ala Pro Leu Gln Glu Thr Arg Asn Gly Arg

500 505 510
Leu Gln Ser Pro Ile Gly Lys Ala Leu Lys Val Ile Ser Ser Asp Phe
515 520 525
Ser Lys Val Ile Asp Tyr Lys Pro Lys Glu Leu Met Asp Trp Met Asn
530 535 540
Thr Gly Ser Ala Ser Asn Ser Phe Gly Val Ala Ser Leu Leu Glu Gly
545 550 555 560
Met Arg Val Met Ser Ile Asp Met Gly Gln Arg Thr Ser Ala Ser Val

565 570 575
Ser Ile Phe Glu Val Val Lys Glu Leu Pro Lys Asp Gln Glu Gln Lys
580 585 590
Leu Phe Tyr Ser Ile Asn Asp Thr Glu Leu Phe Ala Ile His Lys Arg
595 600 605
Ser Phe Leu Leu Asn Leu Pro Gly Glu Val Val Thr Lys Asn Asn Lys
610 615 620
Gln Gln Arg Gln Glu Arg Arg Lys Lys Arg Gln Phe Val Arg Ser Gln

625 630 635 640
Ile Arg Met Leu Ala Asn Val Leu Arg Leu Glu Thr Lys Lys Thr Pro
645 650 655
Asp Glu Arg Lys Lys Ala Ile His Lys Leu Met Glu Ile Val Gln Ser
660 665 670
Tyr Asp Ser Trp Thr Ala Ser Gln Lys Glu Val Trp Glu Lys Glu Leu
675 680 685
Asn Leu Leu Thr Asn Met Ala Ala Phe Asn Asp Glu Ile Trp Lys Glu

690 695 700

Ser Leu Val Glu Leu His His Arg Ile Glu Pro Tyr Val Gly Gln Ile

705 710 715 720

Val Ser Lys Trp Arg Lys Gly Leu Ser Glu Gly Arg Lys Asn Leu Ala

725 730 735

Gly Ile Ser Met Trp Asn Ile Asp Glu Leu Glu Asp Thr Arg Arg Leu

740 745 750

Leu Ile Ser Trp Ser Lys Arg Ser Arg Thr Pro Gly Glu Ala Asn Arg

755 760 765

Ile Glu Thr Asp Glu Pro Phe Gly Ser Ser Leu Leu Gln His Ile Gln

770 775 780

Asn Val Lys Asp Asp Arg Leu Lys Gln Met Ala Asn Leu Ile Ile Met

785 790 795 800

Thr Ala Leu Gly Phe Lys Tyr Asp Lys Glu Glu Lys Asp Arg Tyr Lys

805 810 815

Arg Trp Lys Glu Thr Tyr Pro Ala Cys Gln Ile Ile Leu Phe Glu Asn

820 825 830

Leu Asn Arg Tyr Leu Phe Asn Leu Asp Arg Ser Arg Arg Glu Asn Ser

835 840 845

Arg Leu Met Lys Trp Ala His Arg Ser Ile Pro Arg Thr Val Ser Met

850 855 860

Gln Gly Glu Met Phe Gly Leu Gln Val Gly Asp Val Arg Ser Glu Tyr

865 870 875 880

Ser Ser Arg Phe His Ala Lys Thr Gly Ala Pro Gly Ile Arg Cys His

885 890 895

Ala Leu Thr Glu Glu Asp Leu Lys Ala Gly Ser Asn Thr Leu Lys Arg

900 905 910

Leu Ile Glu Asp Gly Phe Ile Asn Glu Ser Glu Leu Ala Tyr Leu Lys

915 920 925

Lys Gly Asp Ile Ile Pro Ser Gln Gly Gly Glu Leu Phe Val Thr Leu

930 935 940

Ser Lys Arg Tyr Lys Lys Asp Ser Asp Asn Asn Glu Leu Thr Val Ile

945 950 955 960

His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Gln

965 970 975

Gln Asn Ser Glu Val Tyr Arg Val Pro Cys Gln Leu Ala Arg Met Gly

980 985 990

Glu Asp Lys Leu Tyr Ile Pro Lys Ser Gln Thr Glu Thr Ile Lys Lys

995 1000 1005

Tyr Phe Gly Lys Gly Ser Phe Val Lys Asn Asn Thr Glu Gln Glu

1010 1015 1020

Val Tyr Lys Trp Glu Lys Ser Glu Lys Met Lys Ile Lys Thr Asp

1025 1030 1035

Thr Thr Phe Asp Leu Gln Asp Leu Asp Gly Phe Glu Asp Ile Ser

1040 1045 1050

Lys Thr Ile Glu Leu Ala Gln Glu Gln Gln Lys Lys Tyr Leu Thr

1055 1060 1065

Met Phe Arg Asp Pro Ser Gly Tyr Phe Phe Asn Asn Glu Thr Trp

1070 1075 1080

Arg Pro Gln Lys Glu Tyr Trp Ser Ile Val Asn Asn Ile Ile Lys

1085 1090 1095

Ser Cys Leu Lys Lys Lys Ile Leu Ser Asn Lys Val Glu Leu

1100 1105 1110

<210> 523

<211> 1364

<212> PRT

<213> Unknown

<220><223> Member of the candidate phylum Lindowbacteriae

<400> 523

Met Pro Arg Asp Asp Leu Asp Leu Leu Thr Asn Leu Asn Ser Thr Ala

1 5 10 15

Lys Gly Ile Arg Glu Arg Gly Lys Thr Lys Glu Gly Thr Asp Lys Lys

```
            20                  25                  30
Lys Ser Gly Arg Lys Ser Ser Trp Pro Met Asp Lys Ala Ala Trp Glu
        35                  40                  45
Thr Ala Lys Thr Ser Asp Ser Ser Ala His Phe Leu Glu Lys Leu Lys
    50                  55                  60
Gln His Pro Asp Leu Lys Asp Ala Phe Gly Asn Leu Ser Ser Gly Gly

65                  70                  75                  80
Ser Lys Lys Leu Glu Tyr Tyr Lys Lys Leu Ala Gly Ser Ala Pro Trp
                85                  90                  95
Lys Glu Ser Gln Ser Val Ile Leu Glu Lys Ala Ala Arg Trp Lys Glu
            100                 105                 110
Ala Lys Gln Glu Arg Glu Glu Lys Glu Gln Asp Ser Ser Glu His Gly
        115                 120                 125
Ser Lys Ala Ala Tyr Arg Arg Leu Phe Asp Ala Gly Cys Leu Pro Met

    130                 135                 140
Pro Glu Phe Ala Lys Tyr Ile Asp Glu Asn Gln Ile Glu Phe Gly Asp
145                 150                 155                 160
Leu Lys Leu Ser Asp Cys Gly Ala Glu Trp Lys Arg Gly Met Trp Asn
                165                 170                 175
Gln Ala Gly Gln Arg Val Arg Ser His Met Gly Trp Gln Arg Arg Arg
            180                 185                 190
Glu Lys Glu Asn Ala Val Tyr Ser Leu Arg Lys Glu Leu Phe Glu Lys

        195                 200                 205
Gly Gly Ala Ile Arg Arg Lys Lys Ser Glu Glu Leu Thr Pro Glu Asp
    210                 215                 220
Ile Leu Pro Gly Lys Ala Ala Pro Asp Gln Asn Asp Trp Gln Glu Arg
225                 230                 235                 240
Pro Ala Tyr Gly Asn Gln Met Trp Phe Ile Gly Leu Arg Ser Tyr Glu
                245                 250                 255
Glu Asn Glu Met Ala Lys Tyr Ala Glu Glu Ala Gly Met Gly Ser Arg

            260                 265                 270
```

Ser Ala Pro Arg Ile Arg Arg Gly Thr Ile Lys Gly Trp Ser Lys Leu
275 280 285
Arg Glu Arg Trp Leu Gln Ile Leu Lys Arg Asn Pro Gln Ala Thr Arg
290 295 300
Asp Asp Leu Ile Gly Glu Leu Asn Ala Leu Arg Ser Gln Asp Pro Arg
305 310 315 320
Ala Tyr Gly Asp Ala Arg Leu Phe Asp Trp Leu Ser Lys Thr Asp Gln

325 330 335
Arg Phe Leu Trp Asp Gly Phe Asp Ala Asp Gly Lys Ile Leu Cys Gly
340 345 350
Arg Asp Asp Arg Asp Cys Val Ser Ala Phe Val Ala Tyr Asn Glu Glu
355 360 365
Phe Ala Asp Glu Pro Ser Ser Ile Thr Leu Thr Glu Thr Asp Glu Arg
370 375 380
Leu His Pro Val Trp Pro Phe Phe Gly Glu Ser Ser Ala Val Pro Tyr

385 390 395 400
Glu Ile Glu Tyr Asp Leu Glu Thr Ala Cys Pro Thr Ala Ile Arg Leu
405 410 415
Pro Leu Leu Val Gly Lys Glu Asn Gly Gly Tyr Ala Glu Arg Gln Gly
420 425 430
Thr Arg Leu Pro Leu Ala Glu Tyr Ala Asp Leu Ala Ser Ser Phe Gln
435 440 445
Leu Pro Thr Pro Val Arg Leu Asp Val Leu Val Glu Ile Arg Glu Val

450 455 460
Thr Arg Ala Gly Arg Lys Val Thr Cys Pro Phe Ser Tyr Phe Lys Gln
465 470 475 480
Asn Gly Val Trp Tyr Val Arg Glu Gly Glu Ile Pro Ser Gly Glu Ser
485 490 495
Ile Gln Ile Lys Gln Thr Asp Arg Lys Ile Glu Asn Gly Lys Ile Phe
500 505 510
Ile Ser Ser Lys Leu Arg Met Ala Tyr Arg Asp Asp Leu Met Val Ser

515 520 525

Pro Ala Thr Gly Asp Phe Gly Ser Ile Lys Ile Leu Trp Glu Arg Ile

530 535 540

Glu Leu Ala Ser His Val Asp Gln Lys Lys Leu Pro Glu Thr Ala Pro

545 550 555 560

Ala Arg Ser Arg Val Phe Val Ser Phe Ser Cys Asn Val Val Glu Arg

565 570 575

Ala Pro Arg Lys Gln Leu Thr Arg Lys Pro Asp Ala Val Val Val Thr

580 585 590

Ile Pro Ser Gly Val Asp Gln Gly Leu Val Val Val Ser Thr Asp Val

595 600 605

Arg Thr Gly Lys Ser Lys Ser Ser Ser Ala Pro Pro Leu Pro Pro Gly

610 615 620

Ser Arg Leu Trp Pro Ala Asp Ala Val His Gly Asp Pro Pro Leu Arg

625 630 635 640

Ile Leu Ser Val Asp Leu Gly His Arg His Ser Ala Tyr Ala Val Trp

645 650 655

Glu Leu Gly Leu Gln Gln Lys Ser Trp Arg Ala Gly Val Leu Lys Gly

660 665 670

Ser Thr Gln Thr Pro Val Tyr Ala Asp Cys Thr Gly Thr Gly Leu Leu

675 680 685

Cys Leu Pro Gly Asp Gly Glu Asp Thr Pro Ala Glu Glu Glu Ser Leu

690 695 700

Arg Leu Arg Ser Arg Gln Ile Arg Arg Arg Leu Asn Leu Gln Asn Ser

705 710 715 720

Ile Leu Arg Val Ser Arg Leu Leu Ser Leu Asp Lys Phe Glu Lys Thr

725 730 735

Ile Phe Glu Gln Ser Asp Val Arg Asp Arg Pro Asn Lys Lys Gly Leu

740 745 750

Arg Ile Arg Arg Arg Cys Arg Thr Glu Lys Thr Pro Leu Ser Glu Ala

755 760 765

Glu Val Arg Lys Asn Cys Asp Lys Ala Ala Glu Ile Leu Ile Arg Trp

770 775 780

Ala Asp Thr Asp Ala Met Ala Lys Ser Leu Ala Ala Thr Gly Asn Ala

785 790 795 800

Asp Ile Ser Phe Trp Lys Tyr Met Ala Val Lys Asn Pro Pro Leu Ser

805 810 815

Ala Val Val Asp Val Ala Pro Ser Thr Ile Val Pro Asp Asp Gly Pro

820 825 830

Asp Arg Glu Thr Leu Lys Lys Lys Arg Gln Glu Glu Glu Glu Lys Phe

835 840 845

Ala Ser Ser Ile Tyr Glu Asn Arg Val Lys Leu Ala Gly Ala Leu Cys

850 855 860

Ser Gly Tyr Asp Ala Asp His Arg Arg Pro Ala Thr Gly Gly Leu Trp

865 870 875 880

His Asp Leu Asp Arg Thr Leu Ile Arg Glu Ile Ser Tyr Gly Asp Arg

885 890 895

Gly Gln Lys Gly Asn Pro Arg Lys Leu Asn Asn Glu Gly Ile Leu Arg

900 905 910

Leu Leu Arg Arg Pro Pro Arg Ala Arg Pro Asp Trp Arg Glu Phe His

915 920 925

Arg Thr Leu Asn Asp Ala Asn Arg Ile Pro Lys Gly Arg Thr Leu Arg

930 935 940

Gly Gly Leu Ser Met Gly Arg Leu Asn Phe Leu Lys Glu Val Gly Asp

945 950 955 960

Phe Val Lys Lys Trp Ser Cys Arg Pro Arg Trp Pro Gly Asp Arg Arg

965 970 975

His Ile Pro Pro Gly Gln Leu Phe Asp Arg Gln Asp Ala Glu His Leu

980 985 990

Glu His Leu Arg Asp Asp Arg Ile Lys Arg Leu Ala His Leu Ile Val

995 1000 1005

Ala Gln Ala Leu Gly Phe Glu Pro Asp Ile Arg Arg Gly Leu Trp

1010 1015 1020

Lys Tyr Val Asp Gly Ser Thr Gly Glu Ile Leu Trp Gln His Pro

1025 1030 1035

Glu Thr Arg Arg Phe Phe Ala Glu Gly Ala Ala Gly Glu Leu Arg

1040 1045 1050

Glu Val Ser Arg Pro Ala Glu Ile Asp Asp Asp Ala Ala Ala Arg

1055 1060 1065

Pro His Thr Val Ser Ala Pro Ala His Ile Val Val Phe Glu Asn

1070 1075 1080

Leu Ile Arg Tyr Arg Phe Gln Ser Asp Arg Pro Lys Thr Glu Asn

1085 1090 1095

Ala Gly Leu Met Gln Trp Ala His Arg Gln Ile Val His Phe Thr

1100 1105 1110

Lys Gln Val Ala Ser Leu Tyr Gly Leu Lys Val Ala Met Val Tyr

1115 1120 1125

Ala Ala Phe Ser Ser Lys Phe Cys Ser Arg Cys Gly Ser Pro Gly

1130 1135 1140

Ala Arg Val Ser Arg Phe Asp Pro Ala Trp Arg Asn Gln Glu Trp

1145 1150 1155

Phe Lys Arg Arg Thr Ser Asn Pro Arg Ser Lys Val Asp His Ser

1160 1165 1170

Leu Lys Arg Ala Ser Glu Asp Pro Thr Ala Asp Glu Thr Arg Pro

1175 1180 1185

Trp Val Leu Ile Glu Gly Gly Lys Glu Phe Val Cys Ala Asn Ala

1190 1195 1200

Lys Cys Ser Ala His Asp Glu Pro Leu Asn Ala Asp Glu Asn Ala

1205 1210 1215

Ala Ala Asn Ile Gly Leu Arg Phe Leu Arg Gly Val Glu Asp Phe

1220 1225 1230

Arg Thr Lys Val Asn Pro Ala Gly Ala Leu Lys Gly Lys Leu Arg

1235 1240 1245

Phe Glu Thr Gly Ile His Ser Phe Arg Pro Pro Val Ser Gly Ser

1250 1255 1260

Pro Phe Trp Ser Pro Met Ala Glu Pro Ala Gln Lys Lys Lys Ile

1265 1270 1275

Gly Ala Ala Ala Pro Gly Ala Asp Val Asp Glu Ala Gly Asp Ala

1280 1285 1290

Asp Glu Ser Gly Val Val Val Leu Phe Arg Asp Pro Ser Gly Ala

1295 1300 1305

Phe Arg Asn Lys Gln Tyr Trp Tyr Glu Gly Lys Ile Phe Trp Ser

1310 1315 1320

Asn Val Met Met Ala Val Glu Ala Lys Ile Ala Gly Ala Ser Val

1325 1330 1335

Gly Ala Lys Pro Val Ala Ala Ser Trp Gly Gln Ala Gln Pro Gln

1340 1345 1350

Ser Gly Pro Gly Leu Ala Lys Pro Gly Gly Asp

1355 1360

<210> 524

<211> 1489

<212> PRT

<213> Desulfatirhabdium butyrativorans

<400> 524

Met Pro Leu Ser Asn Asn Pro Pro Val Thr Gln Arg Ala Tyr Thr Leu

1 5 10 15

Arg Leu Arg Gly Ala Asp Pro Ser Asp Leu Ser Trp Arg Glu Ala Leu

20 25 30

Trp His Thr His Glu Ala Val Asn Lys Gly Ala Lys Val Phe Gly Asp

35 40 45

Trp Leu Leu Thr Leu Arg Gly Gly Leu Asp His Thr Leu Ala Asp Thr

50 55 60

Lys Val Lys Gly Gly Lys Gly Lys Pro Asp Arg Asp Pro Thr Pro Glu

65 70 75 80

Glu Arg Lys Ala Arg Arg Ile Leu Leu Ala Leu Ser Trp Leu Ser Val
85 90 95
Glu Ser Lys Leu Gly Ala Pro Ser Ser Tyr Ile Val Ala Ser Gly Asp
100 105 110
Glu Pro Ala Lys Asp Arg Asn Asp Asn Val Val Ser Ala Leu Glu Glu
115 120 125
Ile Leu Gln Ser Arg Lys Val Ala Lys Ser Glu Ile Asp Asp Trp Lys
130 135 140

Arg Asp Cys Ser Ala Ser Leu Ser Ala Ala Ile Arg Asp Asp Ala Val
145 150 155 160
Trp Val Asn Arg Ser Lys Val Phe Asp Glu Ala Val Lys Ser Val Gly
165 170 175
Ser Ser Leu Thr Arg Glu Glu Ala Trp Asp Met Leu Glu Arg Phe Phe
180 185 190
Gly Ser Arg Asp Ala Tyr Leu Thr Pro Met Lys Asp Pro Glu Asp Lys
195 200 205

Ser Ser Glu Thr Glu Gln Glu Asp Lys Ala Lys Asp Leu Val Gln Lys
210 215 220
Ala Gly Gln Trp Leu Ser Ser Arg Tyr Gly Thr Ser Glu Gly Ala Asp
225 230 235 240
Phe Cys Arg Met Ser Asp Ile Tyr Gly Lys Ile Ala Ala Trp Ala Asp
245 250 255
Asn Ala Ser Gln Gly Gly Ser Ser Thr Val Asp Asp Leu Val Ser Glu
260 265 270

Leu Arg Gln His Phe Asp Thr Lys Glu Ser Lys Ala Thr Asn Gly Leu
275 280 285
Asp Trp Ile Ile Gly Leu Ser Ser Tyr Thr Gly His Thr Pro Asn Pro
290 295 300
Val His Glu Leu Leu Arg Gln Asn Thr Ser Leu Asn Lys Ser His Leu
305 310 315 320
Asp Asp Leu Lys Lys Lys Ala Asn Thr Arg Ala Glu Ser Cys Lys Ser

325 330 335

Lys Ile Gly Ser Lys Gly Gln Arg Pro Tyr Ser Asp Ala Ile Leu Asn
340 345 350
Asp Val Glu Ser Val Cys Gly Phe Thr Tyr Arg Val Asp Lys Asp Gly
355 360 365
Gln Pro Val Ser Val Ala Asp Tyr Ser Lys Tyr Asp Val Asp Tyr Lys
370 375 380
Trp Gly Thr Ala Arg His Tyr Ile Phe Ala Val Met Leu Asp His Ala
385 390 395 400

Ala Arg Arg Ile Ser Leu Ala His Lys Trp Ile Lys Arg Ala Glu Ala
405 410 415
Glu Arg His Lys Phe Glu Glu Asp Ala Lys Arg Ile Ala Asn Val Pro
420 425 430
Ala Arg Ala Arg Glu Trp Leu Asp Ser Phe Cys Lys Glu Arg Ser Val
435 440 445
Thr Ser Gly Ala Val Glu Pro Tyr Arg Ile Arg Arg Arg Ala Val Asp
450 455 460

Gly Trp Lys Glu Val Val Ala Ala Trp Ser Lys Ser Asp Cys Lys Ser
465 470 475 480
Thr Glu Asp Arg Ile Ala Ala Ala Arg Ala Leu Gln Asp Asp Ser Glu
485 490 495
Ile Asp Lys Phe Gly Asp Ile Gln Leu Phe Glu Ala Leu Ala Glu Asp
500 505 510
Asp Ala Leu Cys Val Trp His Lys Asp Gly Glu Ala Thr Asn Glu Pro
515 520 525

Asp Phe Gln Pro Leu Ile Asp Tyr Ser Leu Ala Ile Glu Ala Glu Phe
530 535 540
Lys Lys Arg Gln Phe Lys Val Pro Ala Tyr Arg His Pro Asp Glu Leu
545 550 555 560
Leu His Pro Val Phe Cys Asp Phe Gly Lys Ser Arg Trp Lys Ile Asn
565 570 575

Tyr Asp Val His Lys Asn Val Gln Ala Pro Phe Tyr Arg Gly Leu Cys
580 585 590

Leu Thr Leu Trp Thr Gly Ser Glu Ile Lys Pro Val Pro Leu Cys Trp
595 600 605
Gln Ser Lys Arg Leu Thr Arg Asp Leu Ala Leu Gly Asn Asn His Arg
610 615 620
Asn Asp Ala Ala Ser Ala Val Thr Arg Ala Asp Arg Leu Gly Arg Ala
625 630 635 640
Ala Ser Asn Val Thr Lys Ser Asp Met Val Asn Ile Thr Gly Leu Phe
645 650 655

Glu Gln Ala Asp Trp Asn Gly Arg Leu Gln Ala Pro Arg Gln Gln Leu
660 665 670
Glu Ala Ile Ala Val Val Arg Asp Asn Pro Arg Leu Ser Glu Gln Glu
675 680 685
Arg Asn Leu Arg Met Cys Gly Met Ile Glu His Ile Arg Trp Leu Val
690 695 700
Thr Phe Ser Val Lys Leu Gln Pro Gln Gly Pro Trp Cys Ala Tyr Ala
705 710 715 720

Glu Gln His Gly Leu Asn Thr Asn Pro Gln Tyr Trp Pro His Ala Asp
725 730 735
Thr Asn Arg Asp Arg Lys Val His Ala Arg Leu Ile Leu Pro Arg Leu
740 745 750
Pro Gly Leu Arg Val Leu Ser Val Asp Leu Gly His Arg Tyr Ala Ala
755 760 765
Ala Cys Ala Val Trp Glu Ala Val Asn Thr Glu Thr Val Lys Glu Ala
770 775 780

Cys Gln Asn Val Gly Arg Asp Met Pro Lys Glu His Asp Leu Tyr Leu
785 790 795 800
His Ile Lys Val Lys Lys Gln Gly Ile Gly Lys Gln Thr Glu Val Asp
805 810 815
Lys Thr Thr Ile Tyr Arg Arg Ile Gly Ala Asp Thr Leu Pro Asp Gly

820 825 830

Arg Pro His Pro Ala Pro Trp Ala Arg Leu Asp Arg Gln Phe Leu Ile

835 840 845

Lys Leu Gln Gly Glu Glu Lys Asp Ala Arg Glu Ala Ser Asn Glu Glu

850 855 860

Ile Trp Ala Leu His Gln Met Glu Cys Lys Leu Asp Arg Thr Lys Pro

865 870 875 880

Leu Ile Asp Arg Leu Ile Ala Ser Gly Trp Gly Leu Leu Lys Arg Gln

885 890 895

Met Ala Arg Leu Asp Ala Leu Lys Glu Leu Gly Trp Ile Pro Ala Pro

900 905 910

Asp Ser Ser Glu Asn Leu Ser Arg Glu Asp Gly Glu Ala Lys Asp Tyr

915 920 925

Arg Glu Ser Leu Ala Val Asp Asp Leu Met Phe Ser Ala Val Arg Thr

930 935 940

Leu Arg Leu Ala Leu Gln Arg His Gly Asn Arg Ala Arg Ile Ala Tyr

945 950 955 960

Tyr Leu Ile Ser Glu Val Lys Ile Arg Pro Gly Gly Ile Gln Glu Lys

965 970 975

Leu Asp Glu Asn Gly Arg Ile Asp Leu Leu Gln Asp Ala Leu Ala Leu

980 985 990

Trp His Glu Leu Phe Ser Ser Pro Gly Trp Arg Asp Glu Ala Ala Lys

995 1000 1005

Gln Leu Trp Asp Ser Arg Ile Ala Thr Leu Ala Gly Tyr Lys Ala

1010 1015 1020

Pro Glu Glu Asn Gly Asp Asn Val Ser Asp Val Ala Tyr Arg Lys

1025 1030 1035

Lys Gln Gln Val Tyr Arg Glu Gln Leu Arg Asn Val Ala Lys Thr

1040 1045 1050

Leu Ser Gly Asp Val Ile Thr Cys Lys Glu Leu Ser Asp Ala Trp

1055 1060 1065

Lys Glu Arg Trp Glu Asp Glu Asp Gln Arg Trp Lys Lys Leu Leu
1070 1075 1080
Arg Trp Phe Lys Asp Trp Val Leu Pro Ser Gly Thr Gln Ala Asn
1085 1090 1095

Asn Ala Thr Ile Arg Asn Val Gly Gly Leu Ser Leu Ser Arg Leu
1100 1105 1110
Ala Thr Ile Thr Glu Phe Arg Arg Lys Val Gln Val Gly Phe Phe
1115 1120 1125
Thr Arg Leu Arg Pro Asp Gly Thr Arg His Glu Ile Gly Glu Gln
1130 1135 1140
Phe Gly Gln Lys Thr Leu Asp Ala Leu Glu Leu Leu Arg Glu Gln
1145 1150 1155

Arg Val Lys Gln Leu Ala Ser Arg Ile Ala Glu Ala Ala Leu Gly
1160 1165 1170
Ile Gly Ser Glu Gly Gly Lys Gly Trp Asp Gly Gly Lys Arg Pro
1175 1180 1185
Arg Gln Arg Ile Asn Asp Ser Arg Phe Ala Pro Cys His Ala Val
1190 1195 1200
Val Ile Glu Asn Leu Ala Asn Tyr Arg Pro Asp Glu Thr Arg Thr
1205 1210 1215

Arg Leu Glu Asn Arg Arg Leu Met Thr Trp Ser Ala Ser Lys Val
1220 1225 1230
His Lys Tyr Leu Ser Glu Ala Cys Gln Leu Asn Gly Leu Tyr Leu
1235 1240 1245
Cys Thr Val Ser Ala Trp Tyr Thr Ser Arg Gln Asp Ser Arg Thr
1250 1255 1260
Gly Ala Pro Gly Ile Arg Cys Gln Asp Val Ser Val Arg Glu Phe
1265 1270 1275

Met Gln Ser Pro Phe Trp Arg Lys Gln Val Lys Gln Ala Glu Ala
1280 1285 1290
Lys His Asp Glu Asn Lys Gly Asp Ala Arg Glu Arg Phe Leu Cys

1295 1300 1305

Glu Leu Asn Lys Thr Trp Lys Ala Lys Thr Pro Ala Glu Trp Lys

1310 1315 1320

Lys Ala Gly Phe Val Arg Ile Pro Leu Arg Gly Gly Glu Ile Phe

1325 1330 1335

Val Ser Ala Asp Ser Lys Ser Pro Ser Ala Lys Gly Ile His Ala

1340 1345 1350

Asp Leu Asn Ala Ala Ala Asn Ile Gly Leu Arg Ala Leu Thr Asp

1355 1360 1365

Pro Asp Trp Pro Gly Lys Trp Trp Tyr Val Pro Cys Asp Pro Val

1370 1375 1380

Ser Phe Glu Ser Lys Met Asp Tyr Val Lys Gly Cys Ala Ala Val

1385 1390 1395

Lys Val Gly Gln Pro Leu Arg Gln Pro Ala Gln Thr Asn Ala Asp

1400 1405 1410

Gly Ala Ala Ser Lys Ile Arg Lys Gly Lys Lys Asn Arg Thr Ala

1415 1420 1425

Gly Thr Ser Lys Glu Lys Val Tyr Leu Trp Arg Asp Ile Ser Ala

1430 1435 1440

Phe Pro Leu Glu Ser Asn Glu Ile Gly Glu Trp Lys Glu Thr Ser

1445 1450 1455

Ala Tyr Gln Asn Asp Val Gln Tyr Arg Val Ile Arg Met Leu Lys

1460 1465 1470

Glu His Ile Lys Ser Leu Asp Asn Arg Thr Gly Asp Asn Val Glu

1475 1480 1485

Gly

<210> 525

<211> 1194

<212> PRT

<213> Desulfonatronum thiodismutan

<400> 525

Met Val Leu Gly Arg Lys Asp Asp Thr Ala Glu Leu Arg Arg Ala Leu
1 5 10 15
Trp Thr Thr His Glu His Val Asn Leu Ala Val Ala Glu Val Glu Arg

20 25 30
Val Leu Leu Arg Cys Arg Gly Arg Ser Tyr Trp Thr Leu Asp Arg Arg
35 40 45
Gly Asp Pro Val His Val Pro Glu Ser Gln Val Ala Glu Asp Ala Leu
50 55 60
Ala Met Ala Arg Glu Ala Gln Arg Arg Asn Gly Trp Pro Val Val Gly
65 70 75 80
Glu Asp Glu Glu Ile Leu Leu Ala Leu Arg Tyr Leu Tyr Glu Gln Ile

85 90 95
Val Pro Ser Cys Leu Leu Asp Asp Leu Gly Lys Pro Leu Lys Gly Asp
100 105 110
Ala Gln Lys Ile Gly Thr Asn Tyr Ala Gly Pro Leu Phe Asp Ser Asp
115 120 125
Thr Cys Arg Arg Asp Glu Gly Lys Asp Val Ala Cys Cys Gly Pro Phe
130 135 140
His Glu Val Ala Gly Lys Tyr Leu Gly Ala Leu Pro Glu Trp Ala Thr

145 150 155 160
Pro Ile Ser Lys Gln Glu Phe Asp Gly Lys Asp Ala Ser His Leu Arg
165 170 175
Phe Lys Ala Thr Gly Gly Asp Asp Ala Phe Phe Arg Val Ser Ile Glu
180 185 190
Lys Ala Asn Ala Trp Tyr Glu Asp Pro Ala Asn Gln Asp Ala Leu Lys
195 200 205
Asn Lys Ala Tyr Asn Lys Asp Asp Trp Lys Lys Glu Lys Asp Lys Gly

210 215 220
Ile Ser Ser Trp Ala Val Lys Tyr Ile Gln Lys Gln Leu Gln Leu Gly
225 230 235 240
Gln Asp Pro Arg Thr Glu Val Arg Arg Lys Leu Trp Leu Glu Leu Gly

245 250 255

Leu Leu Pro Leu Phe Ile Pro Val Phe Asp Lys Thr Met Val Gly Asn

260 265 270

Leu Trp Asn Arg Leu Ala Val Arg Leu Ala Leu Ala His Leu Leu Ser

275 280 285

Trp Glu Ser Trp Asn His Arg Ala Val Gln Asp Gln Ala Leu Ala Arg

290 295 300

Ala Lys Arg Asp Glu Leu Ala Ala Leu Phe Leu Gly Met Glu Asp Gly

305 310 315 320

Phe Ala Gly Leu Arg Glu Tyr Glu Leu Arg Arg Asn Glu Ser Ile Lys

325 330 335

Gln His Ala Phe Glu Pro Val Asp Arg Pro Tyr Val Val Ser Gly Arg

340 345 350

Ala Leu Arg Ser Trp Thr Arg Val Arg Glu Glu Trp Leu Arg His Gly

355 360 365

Asp Thr Gln Glu Ser Arg Lys Asn Ile Cys Asn Arg Leu Gln Asp Arg

370 375 380

Leu Arg Gly Lys Phe Gly Asp Pro Asp Val Phe His Trp Leu Ala Glu

385 390 395 400

Asp Gly Gln Glu Ala Leu Trp Lys Glu Arg Asp Cys Val Thr Ser Phe

405 410 415

Ser Leu Leu Asn Asp Ala Asp Gly Leu Leu Glu Lys Arg Lys Gly Tyr

420 425 430

Ala Leu Met Thr Phe Ala Asp Ala Arg Leu His Pro Arg Trp Ala Met

435 440 445

Tyr Glu Ala Pro Gly Gly Ser Asn Leu Arg Thr Tyr Gln Ile Arg Lys

450 455 460

Thr Glu Asn Gly Leu Trp Ala Asp Val Val Leu Leu Ser Pro Arg Asn

465 470 475 480

Glu Ser Ala Ala Val Glu Glu Lys Thr Phe Asn Val Arg Leu Ala Pro

485 490 495

Ser Gly Gln Leu Ser Asn Val Ser Phe Asp Gln Ile Gln Lys Gly Ser
500 505 510
Lys Met Val Gly Arg Cys Arg Tyr Gln Ser Ala Asn Gln Gln Phe Glu
515 520 525
Gly Leu Leu Gly Gly Ala Glu Ile Leu Phe Asp Arg Lys Arg Ile Ala

530 535 540
Asn Glu Gln His Gly Ala Thr Asp Leu Ala Ser Lys Pro Gly His Val
545 550 555 560
Trp Phe Lys Leu Thr Leu Asp Val Arg Pro Gln Ala Pro Gln Gly Trp
565 570 575
Leu Asp Gly Lys Gly Arg Pro Ala Leu Pro Pro Glu Ala Lys His Phe
580 585 590
Lys Thr Ala Leu Ser Asn Lys Ser Lys Phe Ala Asp Gln Val Arg Pro

595 600 605
Gly Leu Arg Val Leu Ser Val Asp Leu Gly Val Arg Ser Phe Ala Ala
610 615 620
Cys Ser Val Phe Glu Leu Val Arg Gly Gly Pro Asp Gln Gly Thr Tyr
625 630 635 640
Phe Pro Ala Ala Asp Gly Arg Thr Val Asp Asp Pro Glu Lys Leu Trp
645 650 655
Ala Lys His Glu Arg Ser Phe Lys Ile Thr Leu Pro Gly Glu Asn Pro

660 665 670
Ser Arg Lys Glu Glu Ile Ala Arg Arg Ala Ala Met Glu Glu Leu Arg
675 680 685
Ser Leu Asn Gly Asp Ile Arg Arg Leu Lys Ala Ile Leu Arg Leu Ser
690 695 700
Val Leu Gln Glu Asp Asp Pro Arg Thr Glu His Leu Arg Leu Phe Met
705 710 715 720
Glu Ala Ile Val Asp Asp Pro Ala Lys Ser Ala Leu Asn Ala Glu Leu

725 730 735
Phe Lys Gly Phe Gly Asp Asp Arg Phe Arg Ser Thr Pro Asp Leu Trp

740 745 750

Lys Gln His Cys His Phe Phe His Asp Lys Ala Glu Lys Val Val Ala

755 760 765

Glu Arg Phe Ser Arg Trp Arg Thr Glu Thr Arg Pro Lys Ser Ser Ser

770 775 780

Trp Gln Asp Trp Arg Glu Arg Arg Gly Tyr Ala Gly Gly Lys Ser Tyr

785 790 795 800

Trp Ala Val Thr Tyr Leu Glu Ala Val Arg Gly Leu Ile Leu Arg Trp

805 810 815

Asn Met Arg Gly Arg Thr Tyr Gly Glu Val Asn Arg Gln Asp Lys Lys

820 825 830

Gln Phe Gly Thr Val Ala Ser Ala Leu Leu His His Ile Asn Gln Leu

835 840 845

Lys Glu Asp Arg Ile Lys Thr Gly Ala Asp Met Ile Ile Gln Ala Ala

850 855 860

Arg Gly Phe Val Pro Arg Lys Asn Gly Ala Gly Trp Val Gln Val His

865 870 875 880

Glu Pro Cys Arg Leu Ile Leu Phe Glu Asp Leu Ala Arg Tyr Arg Phe

885 890 895

Arg Thr Asp Arg Ser Arg Arg Glu Asn Ser Arg Leu Met Arg Trp Ser

900 905 910

His Arg Glu Ile Val Asn Glu Val Gly Met Gln Gly Glu Leu Tyr Gly

915 920 925

Leu His Val Asp Thr Thr Glu Ala Gly Phe Ser Ser Arg Tyr Leu Ala

930 935 940

Ser Ser Gly Ala Pro Gly Val Arg Cys Arg His Leu Val Glu Glu Asp

945 950 955 960

Phe His Asp Gly Leu Pro Gly Met His Leu Val Gly Glu Leu Asp Trp

965 970 975

Leu Leu Pro Lys Asp Lys Asp Arg Thr Ala Asn Glu Ala Arg Arg Leu

980 985 990

Leu Gly Gly Met Val Arg Pro Gly Met Leu Val Pro Trp Asp Gly Gly
995 1000 1005

Glu Leu Phe Ala Thr Leu Asn Ala Ala Ser Gln Leu His Val Ile
1010 1015 1020

His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg Phe Trp
1025 1030 1035

Gly Arg Cys Gly Glu Ala Ile Arg Ile Val Cys Asn Gln Leu Ser

1040 1045 1050

Val Asp Gly Ser Thr Arg Tyr Glu Met Ala Lys Ala Pro Lys Ala
1055 1060 1065

Arg Leu Leu Gly Ala Leu Gln Gln Leu Lys Asn Gly Asp Ala Pro
1070 1075 1080

Phe His Leu Thr Ser Ile Pro Asn Ser Gln Lys Pro Glu Asn Ser
1085 1090 1095

Tyr Val Met Thr Pro Thr Asn Ala Gly Lys Lys Tyr Arg Ala Gly

1100 1105 1110

Pro Gly Glu Lys Ser Ser Gly Glu Glu Asp Glu Leu Ala Leu Asp
1115 1120 1125

Ile Val Glu Gln Ala Glu Glu Leu Ala Gln Gly Arg Lys Thr Phe
1130 1135 1140

Phe Arg Asp Pro Ser Gly Val Phe Phe Ala Pro Asp Arg Trp Leu
1145 1150 1155

Pro Ser Glu Ile Tyr Trp Ser Arg Ile Arg Arg Arg Ile Trp Gln

1160 1165 1170

Val Thr Leu Glu Arg Asn Ser Ser Gly Arg Gln Glu Arg Ala Glu
1175 1180 1185

Met Asp Glu Met Pro Tyr
1190

<210> 526

<211> 1364

<212> PRT

<213> Unknown

<220><223> Member of the phylum Elusimicrobia

<400> 526

Met Asn Arg Ile Tyr Gln Gly Arg Val Thr Lys Val Glu Val Pro Asp
1 5 10 15

Gly Lys Asp Glu Lys Gly Asn Ile Lys Trp Lys Lys Leu Glu Asn Trp
20 25 30

Ser Asp Ile Leu Trp Gln His His Met Leu Phe Gln Asp Ala Val Asn
35 40 45

Tyr Tyr Thr Leu Ala Leu Ala Ala Ile Ser Gly Ser Ala Val Gly Ser
50 55 60

Asp Glu Lys Ser Ile Ile Leu Arg Glu Trp Ala Val Gln Val Gln Asn
65 70 75 80

Ile Trp Glu Lys Ala Lys Lys Lys Ala Thr Val Phe Glu Gly Pro Gln
85 90 95

Lys Arg Leu Thr Ser Ile Leu Gly Leu Glu Gln Asn Ala Ser Phe Asp
100 105 110

Ile Ala Ala Lys His Ile Leu Arg Thr Ser Glu Ala Lys Pro Glu Gln
115 120 125

Arg Ala Ser Ala Leu Ile Arg Leu Leu Glu Glu Ile Asp Lys Lys Asn
130 135 140

His Asn Val Val Cys Gly Glu Arg Leu Pro Phe Phe Cys Pro Arg Asn
145 150 155 160

Ile Gln Ser Lys Arg Ser Pro Thr Ser Lys Ala Val Ser Ser Val Gln
165 170 175

Glu Gln Lys Arg Gln Glu Glu Val Arg Arg Phe His Asn Met Gln Pro
180 185 190

Glu Glu Val Val Lys Asn Ala Val Thr Leu Asp Ile Ser Leu Phe Lys
195 200 205

Ser Ser Pro Lys Ile Val Phe Leu Glu Asp Pro Lys Lys Ala Arg Ala
210 215 220

Glu Leu Leu Lys Gln Phe Asp Asn Ala Cys Lys Lys His Lys Glu Leu

225 230 235 240

Val Gly Ile Lys Lys Ala Phe Thr Glu Ser Ile Asp Lys His Gly Ser

245 250 255

Ser Leu Lys Val Pro Ala Pro Gly Ser Lys Pro Ser Gly Leu Tyr Pro

260 265 270

Ser Ala Ile Val Phe Lys Tyr Phe Pro Val Asp Ile Thr Lys Thr Val

275 280 285

Phe Leu Lys Ala Thr Glu Lys Leu Ala Met Gly Lys Asp Arg Glu Val

290 295 300

Thr Asn Asp Pro Ile Ala Asp Ala Arg Val Asn Asp Lys Pro His Phe

305 310 315 320

Asp Tyr Phe Thr Asn Ile Ala Leu Ile Arg Glu Lys Glu Lys Asn Arg

325 330 335

Ala Ala Trp Phe Glu Phe Asp Leu Ala Ala Phe Ile Glu Ala Ile Met

340 345 350

Ser Pro His Arg Phe Tyr Gln Asp Thr Gln Lys Arg Lys Glu Ala Ala

355 360 365

Arg Lys Leu Glu Glu Lys Ile Lys Ala Ile Glu Gly Lys Gly Gly Gln

370 375 380

Phe Lys Glu Ser Asp Ser Glu Asp Asp Asp Val Asp Ser Leu Pro Gly

385 390 395 400

Phe Glu Gly Asp Thr Arg Ile Asp Leu Leu Arg Lys Leu Val Thr Asp

405 410 415

Thr Leu Gly Trp Leu Gly Glu Ser Glu Thr Pro Asp Asn Asn Glu Gly

420 425 430

Lys Lys Thr Glu Tyr Ser Ile Ser Glu Arg Thr Leu Arg Ile Phe Pro

435 440 445

Asp Ile Gln Lys Gln Trp Ser Glu Leu Ala Glu Lys Gly Glu Thr Thr

450 455 460

Glu Gly Lys Leu Leu Glu Val Leu Lys His Glu Gln Thr Glu His Gln

465 470 475 480

Ser Asp Phe Gly Ser Ala Thr Leu Tyr Gln His Leu Ala Lys Pro Glu
485 490 495
Phe His Pro Ile Trp Leu Lys Ser Gly Thr Glu Glu Trp His Ala Glu
500 505 510
Asn Pro Leu Lys Ala Trp Leu Asn Tyr Lys Glu Leu Gln Tyr Glu Leu
515 520 525
Thr Asp Lys Lys Arg Pro Ile His Phe Thr Pro Ala His Pro Val Tyr

530 535 540
Ser Pro Arg Tyr Phe Asp Phe Pro Lys Lys Ser Glu Thr Glu Glu Lys
545 550 555 560
Glu Val Ser Lys Asn Thr His Ser Leu Thr Thr Ser Leu Ala Ser Glu
565 570 575
His Ile Lys Asn Ser Leu Gln Phe Thr Ala Gly Leu Ile Arg Lys Thr
580 585 590
Asn Val Gly Lys Lys Ala Ile Lys Ala Arg Phe Ser Tyr Ser Ala Pro

595 600 605
Arg Leu Arg Arg Asp Cys Leu Arg Ser Glu Asn Asn Glu Asn Leu Tyr
610 615 620
Lys Ala Pro Trp Leu Gln Pro Met Met Arg Ala Leu Gly Ile Asp Glu
625 630 635 640
Glu Lys Ala Asp Arg Gln Asn Phe Ala Asn Thr Arg Ile Thr Leu Met
645 650 655
Ala Lys Gly Leu Asp Asp Ile Gln Leu Gly Phe Pro Val Glu Ala Asn

660 665 670
Ser Gln Glu Leu Gln Lys Glu Val Ser Asn Gly Ile Ser Trp Lys Gly
675 680 685
Gln Phe Asn Trp Gly Gly Ile Ala Ser Leu Ser Ala Leu Arg Trp Pro
690 695 700
His Glu Lys Lys Pro Lys Asn Pro Pro Glu Gln Pro Trp Trp Gly Ile
705 710 715 720
Asp Ser Phe Ser Cys Leu Ala Val Asp Leu Gly Gln Arg Tyr Ala Gly

725 730 735

Ala Phe Ala Arg Leu Asp Val Ser Thr Ile Glu Lys Lys Gly Lys Ser

740 745 750

Arg Phe Ile Gly Glu Ala Cys Asp Lys Lys Trp Tyr Ala Lys Val Ser

755 760 765

Arg Met Gly Leu Leu Arg Leu Pro Gly Glu Asp Val Lys Val Trp Arg

770 775 780

Asp Ala Ser Lys Ile Asp Lys Glu Asn Gly Phe Ala Phe Arg Lys Glu

785 790 795 800

Leu Phe Gly Glu Lys Gly Arg Ser Ala Thr Pro Leu Glu Ala Glu Glu

805 810 815

Thr Ala Glu Leu Ile Lys Leu Phe Gly Ala Asn Glu Lys Asp Val Met

820 825 830

Pro Asp Asn Trp Ser Lys Glu Leu Ser Phe Pro Glu Gln Asn Asp Lys

835 840 845

Leu Leu Ile Val Ala Arg Arg Ala Gln Ala Ala Val Ser Arg Leu His

850 855 860

Arg Trp Ala Trp Phe Phe Asp Glu Ala Lys Arg Ser Asp Asp Ala Ile

865 870 875 880

Arg Glu Ile Leu Glu Ser Asp Asp Thr Asp Leu Lys Gln Lys Val Asn

885 890 895

Lys Asn Glu Ile Glu Lys Val Lys Glu Thr Ile Ile Ser Leu Leu Lys

900 905 910

Val Lys Gln Glu Leu Leu Pro Thr Leu Leu Thr Arg Leu Ala Asn Arg

915 920 925

Val Leu Pro Leu Arg Gly Arg Ser Trp Glu Trp Lys Lys His His Gln

930 935 940

Lys Asn Asp Gly Phe Ile Leu Asp Gln Thr Gly Lys Ala Met Pro Asn

945 950 955 960

Val Leu Ile Arg Gly Gln Arg Gly Leu Ser Met Asp Arg Ile Glu Gln

965 970 975

Ile Thr Glu Leu Arg Lys Arg Phe Gln Ala Leu Asn Gln Ser Leu Arg

980 985 990

Arg Gln Ile Gly Lys Lys Ala Pro Ala Lys Arg Asp Asp Ser Ile Pro

995 1000 1005

Asp Cys Cys Pro Asp Leu Leu Glu Lys Leu Asp His Met Lys Glu

1010 1015 1020

Gln Arg Val Asn Gln Thr Ala His Met Ile Leu Ala Glu Ala Leu

1025 1030 1035

Gly Leu Lys Leu Ala Glu Pro Pro Lys Asp Lys Lys Glu Leu Asn

1040 1045 1050

Glu Thr Cys Asp Met His Gly Ala Tyr Ala Lys Val Asp Asn Pro

1055 1060 1065

Val Ser Phe Ile Val Ile Glu Asp Leu Ser Arg Tyr Arg Ser Ser

1070 1075 1080

Gln Gly Arg Ser Pro Arg Glu Asn Ser Arg Leu Met Lys Trp Cys

1085 1090 1095

His Arg Ala Val Arg Asp Lys Leu Lys Glu Met Cys Glu Val Phe

1100 1105 1110

Phe Pro Leu Cys Glu Arg Arg Lys Ala Gly Ser Ala Trp Val Ser

1115 1120 1125

Leu Pro Pro Leu Leu Glu Thr Pro Ala Ala Tyr Ser Ser Arg Phe

1130 1135 1140

Cys Ser Arg Ser Gly Val Ala Gly Phe Arg Ala Val Glu Val Ile

1145 1150 1155

Pro Gly Phe Glu Leu Lys Tyr Pro Trp Ser Trp Leu Lys Asp Lys

1160 1165 1170

Lys Asp Lys Ala Gly Asn Leu Ala Lys Glu Ala Leu Asn Ile Arg

1175 1180 1185

Thr Val Ser Glu Gln Leu Lys Ala Phe Asn Gln Asp Lys Pro Glu

1190 1195 1200

Lys Pro Arg Thr Leu Leu Val Pro Ile Ala Gly Gly Pro Ile Phe

1205 1210 1215

Val Pro Ile Ser Glu Val Gly Leu Ser Ser Phe Gly Leu Lys Pro

1220 1225 1230

Gln Val Val Gln Ala Asp Ile Asn Ala Ala Ile Asn Leu Gly Leu

1235 1240 1245

Arg Ala Ile Ser Asp Pro Arg Ile Trp Glu Ile His Pro Arg Leu

1250 1255 1260

Arg Thr Glu Lys Arg Asp Gly Arg Leu Phe Ala Arg Glu Lys Arg

1265 1270 1275

Lys Tyr Gly Glu Glu Lys Val Glu Val Gln Pro Ser Lys Asn Glu

1280 1285 1290

Lys Ala Lys Lys Val Lys Asp Asp Arg Lys Pro Asn Tyr Phe Ala

1295 1300 1305

Asp Phe Ser Gly Lys Val Asp Trp Gly Phe Gly Asn Ile Lys Asn

1310 1315 1320

Glu Ser Gly Leu Thr Leu Val Ser Gly Lys Ala Leu Trp Trp Thr

1325 1330 1335

Ile Asn Gln Leu Gln Trp Glu Arg Cys Phe Asp Ile Asn Lys Arg

1340 1345 1350

His Ile Glu Asp Trp Ser Asn Lys Gln Lys Gln

1355 1360

<210> 527

<211> 1326

<212> PRT

<213> Unknown

<220><223> Member of the phylum Lentisphaeria

<400> 527

Met Ala Val Glu Leu Asn Arg Ile Tyr Gln Gly Arg Val Asn His Val

1 5 10 15

Tyr Ile Phe Asp Glu Asn Gln Asn Gln Val Ser Val Asp Asn Gly Asp

20 25 30

Asp Leu Leu Phe Val His His Glu Leu Tyr Gln Asp Ala Ile Asn Tyr
35 40 45
Tyr Leu Val Ala Leu Ala Ala Met Ala Leu Asp Ser Lys Asp Ser Leu
50 55 60
Phe Gly Lys Phe Lys Met Gln Ile Arg Ala Val Trp Asn Asp Phe Tyr
65 70 75 80
Arg Asn Gly Gln Leu Arg Pro Gly Leu Lys His Ser Leu Ile Arg Ser
85 90 95

Leu Gly His Ala Ala Glu Leu Asn Thr Ser Asn Gly Ala Asp Ile Ala
100 105 110
Met Asn Leu Ile Leu Glu Asp Gly Gly Ile Pro Ser Glu Ile Leu Asn
115 120 125
Ala Ala Leu Glu His Leu Ala Glu Lys Cys Thr Gly Asp Val Ser Gln
130 135 140
Leu Gly Lys Thr Phe Phe Pro Arg Phe Cys Asp Thr Ala Tyr His Gly
145 150 155 160

Asn Trp Asp Val Asp Ala Lys Ser Phe Ser Glu Lys Lys Gly Arg Gln
165 170 175
Arg Leu Val Asp Ala Leu Tyr Ser Leu His Pro Val Gln Ala Val Gln
180 185 190
Glu Leu Ala Pro Glu Ile Glu Ile Gly Trp Gly Gly Val Lys Thr Gln
195 200 205
Thr Gly Lys Phe Phe Thr Gly Asp Glu Ala Lys Ala Ser Leu Lys Lys
210 215 220

Ala Ile Ser Tyr Phe Leu Gln Asp Thr Gly Lys Asn Ser Pro Glu Leu
225 230 235 240
Gln Glu Tyr Phe Ser Val Ala Gly Lys Gln Pro Leu Glu Gln Tyr Leu
245 250 255
Gly Lys Ile Asp Thr Phe Pro Glu Ile Ser Phe Gly Arg Ile Ser Ser
260 265 270
His Gln Asn Ile Asn Ile Ser Asn Ala Met Trp Ile Leu Lys Phe Phe

275 280 285

Pro Asp Gln Tyr Ser Val Asp Leu Ile Lys Asn Leu Ile Pro Asn Lys
290 295 300
Lys Tyr Glu Ile Gly Ile Ala Pro Gln Trp Gly Asp Asp Pro Val Lys
305 310 315 320
Leu Ser Arg Gly Lys Arg Gly Tyr Thr Phe Arg Ala Phe Thr Asp Leu
325 330 335
Ala Met Trp Glu Lys Asn Trp Lys Val Phe Asp Arg Ala Ala Phe Ser
340 345 350

Asp Ala Leu Lys Thr Ile Asn Gln Phe Arg Asn Lys Thr Gln Glu Arg
355 360 365
Asn Asp Gln Leu Lys Arg Tyr Cys Ala Ala Leu Asn Trp Met Asp Gly
370 375 380
Glu Ser Ser Asp Lys Lys Pro Pro Val Glu Pro Ala Asp Ala Asp Ala
385 390 395 400
Val Asp Glu Ala Ala Thr Ser Val Leu Pro Ile Leu Ala Gly Asp Lys
405 410 415

Arg Trp Asn Ala Leu Leu Gln Leu Gln Lys Glu Leu Gly Ile Cys Asn
420 425 430
Asp Phe Thr Glu Asn Glu Leu Met Asp Tyr Gly Leu Ser Leu Arg Thr
435 440 445
Ile Arg Gly Tyr Gln Lys Leu Arg Ser Met Met Leu Glu Lys Glu Glu
450 455 460
Lys Met Arg Ala Lys Thr Ala Asp Asp Glu Glu Ile Ser Gln Ala Leu
465 470 475 480

Gln Glu Ile Ile Ile Lys Phe Gln Ser Ser His Arg Asp Thr Ile Gly
485 490 495
Ser Val Ser Leu Phe Leu Lys Leu Ala Glu Pro Lys Tyr Phe Cys Val
500 505 510
Trp His Asp Ala Asp Lys Asn Gln Asn Phe Ala Ser Val Asp Met Val
515 520 525

Ala Asp Ala Val Arg Tyr Tyr Ser Tyr Gln Glu Glu Lys Ala Arg Leu
530 535 540

Glu Glu Pro Ile Gln Ile Thr Pro Ala Asp Ala Arg Tyr Ser Arg Arg
545 550 555 560
Val Ser Asp Leu Tyr Ala Leu Val Tyr Lys Asn Ala Lys Glu Cys Lys
565 570 575
Thr Gly Tyr Gly Leu Arg Pro Asp Gly Asn Phe Val Phe Glu Ile Ala
580 585 590
Gln Lys Asn Ala Lys Gly Tyr Ala Pro Ala Lys Val Val Leu Ala Phe
595 600 605

Ser Ala Pro Arg Leu Lys Arg Asp Gly Leu Ile Asp Lys Glu Phe Ser
610 615 620
Ala Tyr Tyr Pro Pro Val Leu Gln Ala Phe Leu Arg Glu Glu Glu Ala
625 630 635 640
Pro Lys Gln Ser Phe Lys Thr Thr Ala Val Ile Leu Met Pro Asp Trp
645 650 655
Asp Lys Asn Gly Lys Arg Arg Ile Leu Leu Asn Phe Pro Ile Lys Leu
660 665 670

Asp Val Ser Ala Ile His Gln Lys Thr Asp His Arg Phe Glu Asn Gln
675 680 685
Phe Tyr Phe Ala Asn Asn Thr Asn Thr Cys Leu Leu Trp Pro Ser Tyr
690 695 700
Gln Tyr Lys Lys Pro Val Thr Trp Tyr Gln Gly Lys Lys Pro Phe Asp
705 710 715 720
Val Val Ala Val Asp Leu Gly Gln Arg Ser Ala Gly Ala Val Ser Arg
725 730 735

Ile Thr Val Ser Thr Glu Lys Arg Glu His Ser Val Ala Ile Gly Glu
740 745 750
Ala Gly Gly Thr Gln Trp Tyr Ala Tyr Arg Lys Phe Ser Gly Leu Leu
755 760 765
Arg Leu Pro Gly Glu Asp Ala Thr Val Ile Arg Asp Gly Gln Arg Thr

770 775 780

Glu Glu Leu Ser Gly Asn Ala Gly Arg Leu Ser Thr Glu Glu Glu Thr

785 790 795 800

Val Gln Ala Cys Val Leu Cys Lys Met Leu Ile Gly Asp Ala Thr Leu

805 810 815

Leu Gly Gly Ser Asp Glu Lys Thr Ile Arg Ser Phe Pro Lys Gln Asn

820 825 830

Asp Lys Leu Leu Ile Ala Phe Arg Arg Ala Thr Gly Arg Met Lys Gln

835 840 845

Leu Gln Arg Trp Leu Trp Met Leu Asn Glu Asn Gly Leu Cys Asp Lys

850 855 860

Ala Lys Thr Glu Ile Ser Asn Ser Asp Trp Leu Val Asn Lys Asn Ile

865 870 875 880

Asp Asn Val Leu Lys Glu Glu Lys Gln His Arg Glu Met Leu Pro Ala

885 890 895

Ile Leu Leu Gln Ile Ala Asp Arg Val Leu Pro Leu Arg Gly Arg Lys

900 905 910

Trp Asp Trp Val Leu Asn Pro Gln Ser Asn Ser Phe Val Leu Gln Gln

915 920 925

Thr Ala His Gly Ser Gly Asp Pro His Lys Lys Ile Cys Gly Gln Arg

930 935 940

Gly Leu Ser Phe Ala Arg Ile Glu Gln Leu Glu Ser Leu Arg Met Arg

945 950 955 960

Cys Gln Ala Leu Asn Arg Ile Leu Met Arg Lys Thr Gly Glu Lys Pro

965 970 975

Ala Thr Leu Ala Glu Met Arg Asn Asn Pro Ile Pro Asp Cys Cys Pro

980 985 990

Asp Ile Leu Met Arg Leu Asp Ala Met Lys Glu Gln Arg Ile Asn Gln

995 1000 1005

Thr Ala Asn Leu Ile Leu Ala Gln Ala Leu Gly Leu Arg His Cys

1010 1015 1020

Leu His Ser Glu Ser Ala Thr Lys Arg Lys Glu Asn Gly Met His
1025 1030 1035
Gly Glu Tyr Glu Lys Ile Pro Gly Val Glu Pro Ala Ala Phe Val
1040 1045 1050

Val Leu Glu Asp Leu Ser Arg Tyr Arg Phe Ser Gln Asp Arg Ser
1055 1060 1065
Ser Tyr Glu Asn Ser Arg Leu Met Lys Trp Ser His Arg Lys Ile
1070 1075 1080
Leu Glu Lys Leu Ala Leu Leu Cys Glu Val Phe Asn Val Pro Ile
1085 1090 1095
Leu Gln Val Gly Ala Ala Tyr Ser Ser Lys Phe Ser Ala Asn Ala
1100 1105 1110

Ile Pro Gly Phe Arg Ala Glu Glu Cys Ser Ile Asp Gln Leu Ser
1115 1120 1125
Phe Tyr Pro Trp Arg Glu Leu Lys Asp Ser Arg Glu Lys Ala Leu
1130 1135 1140
Val Glu Gln Ile Arg Lys Ile Gly His Arg Leu Leu Thr Phe Asp
1145 1150 1155
Ala Lys Ala Thr Ile Ile Met Pro Arg Asn Gly Gly Pro Val Phe
1160 1165 1170

Ile Pro Phe Val Pro Ser Asp Ser Lys Asp Thr Leu Ile Gln Ala
1175 1180 1185
Asp Ile Asn Ala Ser Phe Asn Ile Gly Leu Arg Gly Val Ala Asp
1190 1195 1200
Ala Thr Asn Leu Leu Cys Asn Asn Arg Val Ser Cys Asp Arg Lys
1205 1210 1215
Lys Asp Cys Trp Gln Val Lys Arg Ser Ser Asn Phe Ser Lys Met
1220 1225 1230

Val Tyr Pro Glu Lys Leu Ser Leu Ser Phe Asp Pro Ile Lys Lys
1235 1240 1245
Gln Glu Gly Ala Gly Gly Asn Phe Phe Val Leu Gly Cys Ser Glu

1250 1255 1260

Arg Ile Leu Thr Gly Thr Ser Glu Lys Ser Pro Val Phe Thr Ser

1265 1270 1275

Ser Glu Met Ala Lys Lys Tyr Pro Asn Leu Met Phe Gly Ser Ala

1280 1285 1290

Leu Trp Arg Asn Glu Ile Leu Lys Leu Glu Arg Cys Cys Lys Ile

1295 1300 1305

Asn Gln Ser Arg Leu Asp Lys Phe Ile Ala Lys Lys Glu Val Gln

1310 1315 1320

Asn Glu Leu

1325

<210> 528

<211> 1090

<212> PRT

<213> Laceyella sediminis

<400> 528

Met Ser Ile Arg Ser Phe Lys Leu Lys Ile Lys Thr Lys Ser Gly Val

1 5 10 15

Asn Ala Glu Glu Leu Arg Arg Gly Leu Trp Arg Thr His Gln Leu Ile

20 25 30

Asn Asp Gly Ile Ala Tyr Tyr Met Asn Trp Leu Val Leu Leu Arg Gln

35 40 45

Glu Asp Leu Phe Ile Arg Asn Glu Glu Thr Asn Glu Ile Glu Lys Arg

50 55 60

Ser Lys Glu Glu Ile Gln Gly Glu Leu Leu Glu Arg Val His Lys Gln

65 70 75 80

Gln Gln Arg Asn Gln Trp Ser Gly Glu Val Asp Asp Gln Thr Leu Leu

85 90 95

Gln Thr Leu Arg His Leu Tyr Glu Glu Ile Val Pro Ser Val Ile Gly

100 105 110

Lys Ser Gly Asn Ala Ser Leu Lys Ala Arg Phe Phe Leu Gly Pro Leu

115 120 125

```
Val Asp Pro Asn Asn Lys Thr Thr Lys Asp Val Ser Lys Ser Gly Pro
    130                 135                 140
Thr Pro Lys Trp Lys Lys Met Lys Asp Ala Gly Asp Pro Asn Trp Val

145                 150                 155                 160
Gln Glu Tyr Glu Lys Tyr Met Ala Glu Arg Gln Thr Leu Val Arg Leu
                165                 170                 175
Glu Glu Met Gly Leu Ile Pro Leu Phe Pro Met Tyr Thr Asp Glu Val
            180                 185                 190
Gly Asp Ile His Trp Leu Pro Gln Ala Ser Gly Tyr Thr Arg Thr Trp
        195                 200                 205
Asp Arg Asp Met Phe Gln Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu

    210                 215                 220
Ser Trp Asn Arg Arg Val Arg Glu Arg Arg Ala Gln Phe Glu Lys Lys
225                 230                 235                 240
Thr His Asp Phe Ala Ser Arg Phe Ser Glu Ser Asp Val Gln Trp Met
                245                 250                 255
Asn Lys Leu Arg Glu Tyr Glu Ala Gln Gln Glu Lys Ser Leu Glu Glu
            260                 265                 270
Asn Ala Phe Ala Pro Asn Glu Pro Tyr Ala Leu Thr Lys Lys Ala Leu

        275                 280                 285
Arg Gly Trp Glu Arg Val Tyr His Ser Trp Met Arg Leu Asp Ser Ala
    290                 295                 300
Ala Ser Glu Glu Ala Tyr Trp Gln Glu Val Ala Thr Cys Gln Thr Ala
305                 310                 315                 320
Met Arg Gly Glu Phe Gly Asp Pro Ala Ile Tyr Gln Phe Leu Ala Gln
                325                 330                 335
Lys Glu Asn His Asp Ile Trp Arg Gly Tyr Pro Glu Arg Val Ile Asp

            340                 345                 350
Phe Ala Glu Leu Asn His Leu Gln Arg Glu Leu Arg Arg Ala Lys Glu
        355                 360                 365
Asp Ala Thr Phe Thr Leu Pro Asp Ser Val Asp His Pro Leu Trp Val
```

370 375 380

Arg Tyr Glu Ala Pro Gly Gly Thr Asn Ile His Gly Tyr Asp Leu Val

385 390 395 400

Gln Asp Thr Lys Arg Asn Leu Thr Leu Ile Leu Asp Lys Phe Ile Leu

405 410 415

Pro Asp Glu Asn Gly Ser Trp His Glu Val Lys Lys Val Pro Phe Ser

420 425 430

Leu Ala Lys Ser Lys Gln Phe His Arg Gln Val Trp Leu Gln Glu Glu

435 440 445

Gln Lys Gln Lys Lys Arg Glu Val Val Phe Tyr Asp Tyr Ser Thr Asn

450 455 460

Leu Pro His Leu Gly Thr Leu Ala Gly Ala Lys Leu Gln Trp Asp Arg

465 470 475 480

Asn Phe Leu Asn Lys Arg Thr Gln Gln Gln Ile Glu Glu Thr Gly Glu

485 490 495

Ile Gly Lys Val Phe Phe Asn Ile Ser Val Asp Val Arg Pro Ala Val

500 505 510

Glu Val Lys Asn Gly Arg Leu Gln Asn Gly Leu Gly Lys Ala Leu Thr

515 520 525

Val Leu Thr His Pro Asp Gly Thr Lys Ile Val Thr Gly Trp Lys Ala

530 535 540

Glu Gln Leu Glu Lys Trp Val Gly Glu Ser Gly Arg Val Ser Ser Leu

545 550 555 560

Gly Leu Asp Ser Leu Ser Glu Gly Leu Arg Val Met Ser Ile Asp Leu

565 570 575

Gly Gln Arg Thr Ser Ala Thr Val Ser Val Phe Glu Ile Thr Lys Glu

580 585 590

Ala Pro Asp Asn Pro Tyr Lys Phe Phe Tyr Gln Leu Glu Gly Thr Glu

595 600 605

Leu Phe Ala Val His Gln Arg Ser Phe Leu Leu Ala Leu Pro Gly Glu

610 615 620

Asn Pro Pro Gln Lys Ile Lys Gln Met Arg Glu Ile Arg Trp Lys Glu
625 630 635 640
Arg Asn Arg Ile Lys Gln Gln Val Asp Gln Leu Ser Ala Ile Leu Arg
645 650 655
Leu His Lys Lys Val Asn Glu Asp Glu Arg Ile Gln Ala Ile Asp Lys

660 665 670
Leu Leu Gln Lys Val Ala Ser Trp Gln Leu Asn Glu Glu Ile Ala Thr
675 680 685
Ala Trp Asn Gln Ala Leu Ser Gln Leu Tyr Ser Lys Ala Lys Glu Asn
690 695 700
Asp Leu Gln Trp Asn Gln Ala Ile Lys Asn Ala His His Gln Leu Glu
705 710 715 720
Pro Val Val Gly Lys Gln Ile Ser Leu Trp Arg Lys Asp Leu Ser Thr

725 730 735
Gly Arg Gln Gly Ile Ala Gly Leu Ser Leu Trp Ser Ile Glu Glu Leu
740 745 750
Glu Ala Thr Lys Lys Leu Leu Thr Arg Trp Ser Lys Arg Ser Arg Glu
755 760 765
Pro Gly Val Val Lys Arg Ile Glu Arg Phe Glu Thr Phe Ala Lys Gln
770 775 780
Ile Gln His His Ile Asn Gln Val Lys Glu Asn Arg Leu Lys Gln Leu

785 790 795 800
Ala Asn Leu Ile Val Met Thr Ala Leu Gly Tyr Lys Tyr Asp Gln Glu
805 810 815
Gln Lys Lys Trp Ile Glu Val Tyr Pro Ala Cys Gln Val Val Leu Phe
820 825 830
Glu Asn Leu Arg Ser Tyr Arg Phe Ser Tyr Glu Arg Ser Arg Arg Glu
835 840 845
Asn Lys Lys Leu Met Glu Trp Ser His Arg Ser Ile Pro Lys Leu Val

850 855 860
Gln Met Gln Gly Glu Leu Phe Gly Leu Gln Val Ala Asp Val Tyr Ala

865 870 875 880

Ala Tyr Ser Ser Arg Tyr His Gly Arg Thr Gly Ala Pro Gly Ile Arg

885 890 895

Cys His Ala Leu Thr Glu Ala Asp Leu Arg Asn Glu Thr Asn Ile Ile

900 905 910

His Glu Leu Ile Glu Ala Gly Phe Ile Lys Glu Glu His Arg Pro Tyr

915 920 925

Leu Gln Gln Gly Asp Leu Val Pro Trp Ser Gly Gly Glu Leu Phe Ala

930 935 940

Thr Leu Gln Lys Pro Tyr Asp Asn Pro Arg Ile Leu Thr Leu His Ala

945 950 955 960

Asp Ile Asn Ala Ala Gln Asn Ile Gln Lys Arg Phe Trp His Pro Ser

965 970 975

Met Trp Phe Arg Val Asn Cys Glu Ser Val Met Glu Gly Glu Ile Val

980 985 990

Thr Tyr Val Pro Lys Asn Lys Thr Val His Lys Lys Gln Gly Lys Thr

995 1000 1005

Phe Arg Phe Val Lys Val Glu Gly Ser Asp Val Tyr Glu Trp Ala

1010 1015 1020

Lys Trp Ser Lys Asn Arg Asn Lys Asn Thr Phe Ser Ser Ile Thr

1025 1030 1035

Glu Arg Lys Pro Pro Ser Ser Met Ile Leu Phe Arg Asp Pro Ser

1040 1045 1050

Gly Thr Phe Phe Lys Glu Gln Glu Trp Val Glu Gln Lys Thr Phe

1055 1060 1065

Trp Gly Lys Val Gln Ser Met Ile Gln Ala Tyr Met Lys Lys Thr

1070 1075 1080

Ile Val Gln Arg Met Glu Glu

1085 1090

<210> 529

<211> 1041

<212> PRT

<213> Methylobacterium nodulans

<400> 529

Met Tyr Glu Ala Ile Val Leu Ala Asp Asp Ala Asn Ala Gln Leu Ala

1 5 10 15

Asn Ala Phe Leu Gly Pro Leu Thr Asp Pro Asn Ser Ala Gly Phe Leu

20 25 30

Glu Ala Phe Asn Lys Val Asp Arg Pro Ala Pro Ser Trp Leu Asp Gln

35 40 45

Val Pro Ala Ser Asp Pro Ile Asp Pro Ala Val Leu Ala Glu Ala Asn

50 55 60

Ala Trp Leu Asp Thr Asp Ala Gly Arg Ala Trp Leu Val Asp Thr Gly

65 70 75 80

Ala Pro Pro Arg Trp Arg Ser Leu Ala Ala Lys Gln Asp Pro Ile Trp

85 90 95

Pro Arg Glu Phe Ala Arg Lys Leu Gly Glu Leu Arg Lys Glu Ala Ala

100 105 110

Ser Gly Thr Ser Ala Ile Ile Lys Ala Leu Lys Arg Asp Phe Gly Val

115 120 125

Leu Pro Leu Phe Gln Pro Ser Leu Ala Pro Arg Ile Leu Gly Ser Arg

130 135 140

Ser Ser Leu Thr Pro Trp Asp Arg Leu Ala Phe Arg Leu Ala Val Gly

145 150 155 160

His Leu Leu Ser Trp Glu Ser Trp Cys Thr Arg Ala Arg Asp Glu His

165 170 175

Thr Ala Arg Val Gln Arg Leu Glu Gln Phe Ser Ser Ala His Leu Lys

180 185 190

Gly Asp Leu Ala Thr Lys Val Ser Thr Leu Arg Glu Tyr Glu Arg Ala

195 200 205

Arg Lys Glu Gln Ile Ala Gln Leu Gly Leu Pro Met Gly Glu Arg Asp

210 215 220

Phe Leu Ile Thr Val Arg Met Thr Arg Gly Trp Asp Asp Leu Arg Glu

225 230 235 240

Lys Trp Arg Arg Ser Gly Asp Lys Gly Gln Glu Ala Leu His Ala Ile

245 250 255

Ile Ala Thr Glu Gln Thr Arg Lys Arg Gly Arg Phe Gly Asp Pro Asp

260 265 270

Leu Phe Arg Trp Leu Ala Arg Pro Glu Asn His His Val Trp Ala Asp

275 280 285

Gly His Ala Asp Ala Val Gly Val Leu Ala Arg Val Asn Ala Met Glu

290 295 300

Arg Leu Val Glu Arg Ser Arg Asp Thr Ala Leu Met Thr Leu Pro Asp

305 310 315 320

Pro Val Ala His Pro Arg Ser Ala Gln Trp Glu Ala Glu Gly Gly Ser

325 330 335

Asn Leu Arg Asn Tyr Gln Leu Glu Ala Val Gly Gly Glu Leu Gln Ile

340 345 350

Thr Leu Pro Leu Leu Lys Ala Ala Asp Asp Gly Arg Cys Ile Asp Thr

355 360 365

Pro Leu Ser Phe Ser Leu Ala Pro Ser Asp Gln Leu Gln Gly Val Val

370 375 380

Leu Thr Lys Gln Asp Lys Gln Gln Lys Ile Thr Tyr Cys Thr Asn Met

385 390 395 400

Asn Glu Val Phe Glu Ala Lys Leu Gly Ser Ala Asp Leu Leu Leu Asn

405 410 415

Trp Asp His Leu Arg Gly Arg Ile Arg Asp Arg Val Asp Ala Gly Asp

420 425 430

Ile Gly Ser Ala Phe Leu Lys Leu Ala Leu Asp Val Ala His Val Leu

435 440 445

Pro Asp Gly Val Asp Asp Gln Leu Ala Arg Ala Ala Phe His Phe Gln

450 455 460

Ser Ala Lys Gly Ala Lys Ser Lys His Ala Asp Ser Val Gln Ala Gly

465 470 475 480

Leu Arg Val Leu Ser Ile Asp Leu Gly Val Arg Ser Phe Ala Thr Cys
485 490 495
Ser Val Phe Glu Leu Lys Asp Thr Ala Pro Thr Thr Gly Val Ala Phe
500 505 510
Pro Leu Ala Glu Phe Arg Leu Trp Ala Val His Glu Arg Ser Phe Thr

515 520 525
Leu Glu Leu Pro Gly Glu Asn Val Gly Ala Ala Gly Gln Gln Trp Arg
530 535 540
Ala Gln Ala Asp Ala Glu Leu Arg Gln Leu Arg Gly Gly Leu Asn Arg
545 550 555 560
His Arg Gln Leu Leu Arg Ala Ala Thr Val Gln Lys Gly Glu Arg Asp
565 570 575
Ala Tyr Leu Thr Asp Leu Arg Glu Ala Trp Ser Ala Lys Glu Leu Trp

580 585 590
Pro Phe Glu Ala Ser Leu Leu Ser Glu Leu Glu Arg Cys Ser Thr Val
595 600 605
Ala Asp Pro Leu Trp Gln Asp Thr Cys Lys Arg Ala Ala Arg Leu Tyr
610 615 620
Arg Thr Glu Phe Gly Ala Val Val Ser Glu Trp Arg Ser Arg Thr Arg
625 630 635 640
Ser Arg Glu Asp Arg Lys Tyr Ala Gly Lys Ser Met Trp Ser Val Gln

645 650 655
His Leu Thr Asp Val Arg Arg Phe Leu Gln Ser Trp Ser Leu Ala Gly
660 665 670
Arg Ala Ser Gly Asp Ile Arg Arg Leu Asp Arg Glu Arg Gly Gly Val
675 680 685
Phe Ala Lys Asp Leu Leu Asp His Ile Asp Ala Leu Lys Asp Asp Arg
690 695 700
Leu Lys Thr Gly Ala Asp Leu Ile Val Gln Ala Ala Arg Gly Phe Gln

705 710 715 720
Arg Asn Glu Phe Gly Tyr Trp Val Gln Lys His Ala Pro Cys His Val

725 730 735

Ile Leu Phe Glu Asp Leu Ser Arg Tyr Arg Met Arg Thr Asp Arg Pro

740 745 750

Arg Arg Glu Asn Ser Gln Leu Met Gln Trp Ala His Arg Gly Val Pro

755 760 765

Asp Met Val Gly Met Gln Gly Glu Ile Tyr Gly Ile Gln Asp Arg Arg

770 775 780

Asp Pro Asp Ser Ala Arg Lys His Ala Arg Gln Pro Leu Ala Ala Phe

785 790 795 800

Cys Leu Asp Thr Pro Ala Ala Phe Ser Ser Arg Tyr His Ala Ser Thr

805 810 815

Met Thr Pro Gly Ile Arg Cys His Pro Leu Arg Lys Arg Glu Phe Glu

820 825 830

Asp Gln Gly Phe Leu Glu Leu Leu Lys Arg Glu Asn Glu Gly Leu Asp

835 840 845

Leu Asn Gly Tyr Lys Pro Gly Asp Leu Val Pro Leu Pro Gly Gly Glu

850 855 860

Val Phe Val Cys Leu Asn Ala Asn Gly Leu Ser Arg Ile His Ala Asp

865 870 875 880

Ile Asn Ala Ala Gln Asn Leu Gln Arg Arg Phe Trp Thr Gln His Gly

885 890 895

Asp Ala Phe Arg Leu Pro Cys Gly Lys Ser Ala Val Gln Gly Gln Ile

900 905 910

Arg Trp Ala Pro Leu Ser Met Gly Lys Arg Gln Ala Gly Ala Leu Gly

915 920 925

Gly Phe Gly Tyr Leu Glu Pro Thr Gly His Asp Ser Gly Ser Cys Gln

930 935 940

Trp Arg Lys Thr Thr Glu Ala Glu Trp Arg Arg Leu Ser Gly Ala Gln

945 950 955 960

Lys Asp Arg Asp Glu Ala Ala Ala Ala Glu Asp Glu Glu Leu Gln Gly

965 970 975

Leu Glu Glu Glu Leu Leu Glu Arg Ser Gly Glu Arg Val Val Phe Phe
980 985 990
Arg Asp Pro Ser Gly Val Val Leu Pro Thr Asp Leu Trp Phe Pro Ser
995 1000 1005
Ala Ala Phe Trp Ser Ile Val Arg Ala Lys Thr Val Gly Arg Leu
1010 1015 1020
Arg Ser His Leu Asp Ala Gln Ala Glu Ala Ser Tyr Ala Val Ala

1025 1030 1035
Ala Gly Leu
1040

<210> 530
<211> 1413
<212> PRT
<213> Unknown
<220><223> Member of the candidate phylum Omnitrophica
<400> 530
Met Asn Arg Ile Tyr Gln Gly Arg Val Thr Lys Val Glu Lys Leu Lys
1 5 10 15
Asn Gly Lys Ser Pro Asp Asp Arg Glu Glu Leu Lys Asp Trp Gln Thr
20 25 30
Ala Leu Trp Arg His His Glu Leu Phe Gln Asp Ala Val Ser Tyr Tyr

35 40 45
Thr Leu Ala Leu Ala Ala Met Ala Glu Gly Leu Pro Asp Lys His Pro
50 55 60
Ile Asn Val Leu Arg Lys Arg Met Glu Glu Ala Trp Glu Glu Phe Pro
65 70 75 80
Arg Lys Thr Val Thr Pro Ala Lys Asn Leu Arg Asp Ser Val Arg Pro
85 90 95
Trp Leu Gly Leu Ser Glu Ser Ala Ser Phe Gly Asp Ala Leu Lys Lys

100 105 110
Ile Leu Pro Pro Ala Pro Glu Asn Lys Glu Val Arg Ala Leu Ala Val
115 120 125

Ala Leu Leu Ala Glu Lys Ala Arg Thr Leu Lys Pro Gln Lys Thr Ser
130 135 140
Ala Ser Tyr Trp Gly Arg Phe Cys Asp Asp Leu Lys Lys Lys Pro Asn
145 150 155 160
Trp Asp Tyr Ser Glu Glu Glu Leu Ala Arg Lys Thr Gly Ser Gly Asp

165 170 175
Trp Val Ala Gly Leu Trp Ser Glu Asp Ala Leu Asn Lys Ile Asp Glu
180 185 190
Leu Ala Lys Ser Leu Lys Leu Ser Ser Leu Val Lys Cys Val Pro Asp
195 200 205
Gly Gln Ile Asn Pro Glu Gly Ala Arg Asn Leu Val Lys Glu Ala Leu
210 215 220
Asp His Leu Glu Gly Val Ser Asn Gly Thr Lys Lys Glu Lys Asn Asp

225 230 235 240
Pro Gly Pro Ala Lys Lys Thr Asn Asn Trp Leu Arg Gln His Ala Ser
245 250 255
Asp Val Arg Asn Phe Ile His Lys Asn Lys Asn Gln Phe Ser Ser Leu
260 265 270
Pro Asn Gly Arg Leu Ile Thr Glu Arg Ala Arg Gly Gly Gly Ile Asn
275 280 285
Ile Asn Lys Thr Tyr Ala Gly Val Leu Phe Lys Ala Phe Pro Cys Pro

290 295 300
Phe Thr Phe Asp Tyr Val Arg Ala Ala Val Pro Glu Pro Lys Val Lys
305 310 315 320
Lys Val Asp Gln Glu Lys Lys Ser Glu Gln Ser Ala Thr Trp Thr Glu
325 330 335
Leu Glu Lys Arg Ile Leu Arg Ile Gly Asp Asp Pro Ile Glu Leu Ala
340 345 350
Arg Lys Asn Asn Lys Pro Ile Phe Lys Ala Phe Thr Ala Leu Glu Lys

355 360 365
Trp Ser Asp Gln Asn Ser Lys Ser Cys Trp Ser Asp Phe Asp Lys Cys

370 375 380

Ala Phe Glu Glu Ala Leu Lys Thr Leu Asn Gln Phe Asn Gln Lys Thr

385 390 395 400

Glu Glu Arg Glu Lys Arg Arg Ser Glu Ala Glu Ala Glu Leu Lys Tyr

405 410 415

Met Met Asp Glu Asn Pro Glu Trp Lys Pro Lys Lys Glu Thr Glu Gly

420 425 430

Asp Asp Val Arg Glu Val Pro Ile Leu Lys Gly Asp Pro Arg Tyr Glu

435 440 445

Lys Leu Val Lys Leu Phe Gly Asp Leu Asp Glu Glu Gly Ser Glu His

450 455 460

Ala Thr Gly Lys Ile Tyr Gly Pro Ser Arg Ala Ser Leu Arg Gly Phe

465 470 475 480

Gly Lys Leu Arg Asn Glu Trp Val Asp Leu Phe Thr Lys Ala Asn Asp

485 490 495

Asn Pro Arg Glu Gln Asp Leu Gln Lys Ala Val Thr Gly Phe Gln Arg

500 505 510

Glu His Lys Leu Asp Met Gly Tyr Thr Ala Phe Phe Leu Lys Leu Cys

515 520 525

Glu Arg Asp Tyr Trp Asp Ile Trp Arg Asp Asp Thr Glu Val Glu Val

530 535 540

Lys Lys Ile Arg Glu Lys Arg Trp Val Lys Ser Val Val Tyr Ala Ala

545 550 555 560

Ala Asp Thr Arg Glu Leu Ala Glu Glu Leu Glu Arg Leu Gln Glu Pro

565 570 575

Val Arg Tyr Thr Pro Ala Glu Pro Gln Phe Ser Arg Arg Leu Phe Met

580 585 590

Phe Ser Asp Ile Lys Gly Lys Gln Gly Ala Lys His Ile Arg Glu Gly

595 600 605

Leu Val Glu Val Ser Leu Ala Val Lys Asp Gln Ser Gly Lys Tyr Gly

610 615 620

Thr Cys Arg Val Arg Leu His Tyr Ser Ala Pro Arg Leu Ile Arg Asp
625 630 635 640
His Leu Ser Asp Gly Ser Ser Ser Met Trp Leu Gln Pro Met Met Ala
645 650 655
Ala Leu Gly Leu Ser Ser Asp Ala Arg Gly Cys Phe Thr Arg Asp Ser
660 665 670
Lys Gly Asn Val Lys Glu Pro Ala Val Ala Leu Met Ser Asp Phe Val

675 680 685
Gly Arg Lys Arg Glu Leu Arg Met Leu Leu Asn Phe Pro Val Asp Leu
690 695 700
Asp Ile Ser Lys Leu Glu Glu Asn Ile Gly Lys Lys Ala Arg Trp Glu
705 710 715 720
Lys Gln Met Asn Thr Ala Tyr Glu Lys Asn Lys Leu Lys Gln Arg Phe
725 730 735
His Leu Ile Trp Pro Gly Met Glu Leu Lys Glu Thr Gln Glu Pro Gly

740 745 750
Gln Phe Trp Trp Asp Asn Pro Thr Ile Gln Lys Glu Gly Met Tyr Cys
755 760 765
Leu Ala Ile Asp Leu Ser Gln Arg Arg Ala Ala Asp Tyr Ala Leu Leu
770 775 780
His Ala Gly Val Asn Arg Asp Ser Lys Thr Phe Val Glu Leu Gly Gln
785 790 795 800
Ala Gly Gly Gln Ser Trp Phe Thr Lys Leu Cys Ala Ala Gly Ser Leu

805 810 815
Arg Leu Pro Gly Glu Asp Thr Glu Val Ile Arg Glu Gly Lys Arg Gln
820 825 830
Ile Glu Leu Ser Gly Lys Lys Gly Arg Asn Ala Thr Gln Ser Glu Tyr
835 840 845
Asp Gln Ala Ile Ala Leu Ala Lys Gln Leu Leu His Asn Glu Asn Ser
850 855 860
Ala Glu Leu Glu Ser Ala Ala Arg Asp Trp Leu Gly Asp Asn Ala Lys

865 870 875 880

Arg Phe Ser Phe Pro Glu Gln Asn Asp Lys Leu Ile Asp Leu Tyr Tyr

885 890 895

Gly Ala Leu Ser Arg Tyr Lys Thr Trp Leu Arg Trp Ser Trp Arg Leu

900 905 910

Thr Glu Gln His Lys Glu Leu Trp Asp Lys Thr Leu Asp Glu Ile Arg

915 920 925

Lys Val Pro Tyr Phe Ala Ser Trp Gly Glu Leu Ala Gly Asn Gly Thr

930 935 940

Asn Glu Ala Thr Val Gln Gln Leu Gln Lys Leu Ile Ala Asp Ala Ala

945 950 955 960

Val Asp Leu Arg Asn Phe Leu Glu Lys Ala Leu Leu His Ile Ala Tyr

965 970 975

Arg Ala Leu Pro Leu Arg Glu Asn Thr Trp Arg Trp Ile Glu Asn Gly

980 985 990

Lys Asp Gly Lys Gly Lys Pro Leu His Leu Leu Val Ser Asp Gly Gln

995 1000 1005

Ser Pro Ala Glu Ile Pro Trp Leu Arg Gly Gln Arg Gly Leu Ser

1010 1015 1020

Ile Ala Arg Ile Glu Gln Leu Glu Asn Phe Arg Arg Ala Val Leu

1025 1030 1035

Ser Leu Asn Arg Leu Leu Arg His Glu Ile Gly Thr Lys Pro Glu

1040 1045 1050

Phe Gly Ser Ser Thr Cys Gly Glu Ser Leu Pro Asp Pro Cys Pro

1055 1060 1065

Asp Leu Thr Asp Lys Ile Val Arg Leu Lys Glu Glu Arg Val Asn

1070 1075 1080

Gln Thr Ala His Leu Ile Ile Ala Gln Ser Leu Gly Val Arg Leu

1085 1090 1095

Lys Gly His Ser Leu Phe Thr Glu Glu Arg Glu Lys Ala Asp Met

1100 1105 1110

His Gly Glu His Glu Val Ile Pro Gly Arg Ser Pro Val Asp Phe

1115 1120 1125

Val Val Leu Glu Asp Leu Ser Arg Tyr Thr Thr Asp Lys Ser Arg

1130 1135 1140

Ser Arg Ser Glu Asn Ser Arg Leu Met Lys Trp Cys His Arg Lys

1145 1150 1155

Ile Asn Glu Lys Val Lys Leu Leu Ala Glu Pro Phe Gly Ile Pro

1160 1165 1170

Val Ile Glu Val Phe Ala Ser Tyr Ser Ser Lys Phe Asp Ala Arg

1175 1180 1185

Thr Gly Ala Pro Gly Phe Arg Ala Val Glu Val Thr Ser Glu Asp

1190 1195 1200

Arg Pro Phe Trp Arg Lys Thr Ile Glu Lys Gln Ser Val Ala Arg

1205 1210 1215

Glu Val Phe Asp Cys Leu Asp Asn Leu Val Gly Lys Gly Leu Asn

1220 1225 1230

Gly Ile His Leu Val Leu Pro Gln Asn Gly Gly Pro Leu Phe Ile

1235 1240 1245

Ala Ala Val Lys Glu Asp Gln Pro Leu Pro Ala Ile Arg Gln Ala

1250 1255 1260

Asp Ile Asn Ala Ala Val Asn Ile Gly Leu Arg Ala Ile Ala Gly

1265 1270 1275

Pro Ser Cys Tyr His Ala His Pro Lys Val Arg Leu Ile Lys Gly

1280 1285 1290

Glu Ser Gly Thr Asp Lys Gly Lys Trp Leu Pro Arg Lys Gly Lys

1295 1300 1305

Glu Ala Asn Lys Arg Glu Asn Ala Gln Phe Gly Asn Val Asp Leu

1310 1315 1320

Asp Leu Glu Val Lys Phe Asn Arg Leu Asp Ile Asp Ser Asp Val

1325 1330 1335

Leu Lys Gly Asp Asn Thr Asn Leu Phe His Asp Pro Leu Asn Ile

1340 1345 1350

Ala Cys Tyr Gly Phe Ala Thr Ile Gln Asn Leu Gln His Pro Phe

1355 1360 1365

Leu Ala His Ala Ser Ala Val Phe Ser Arg Gln Lys Gly Ala Val

1370 1375 1380

Ala Arg Leu Gln Trp Glu Val Cys Arg Ala Ile Asn Ser Arg Arg

1385 1390 1395

Leu Glu Ala Trp Gln Lys Lys Ala Glu Lys Ala Ala Val Lys Arg

1400 1405 1410

<210> 531

<211> 1388

<212> PRT

<213> Unknown

<220><223> Member of the family Opitutaceae

<400> 531

Met Ser Leu Asn Arg Ile Tyr Gln Gly Arg Val Ala Ala Val Glu Thr

1 5 10 15

Gly Thr Ala Leu Ala Lys Gly Asn Val Glu Trp Met Pro Ala Ala Gly

20 25 30

Gly Asp Glu Val Leu Trp Gln His His Glu Leu Phe Gln Ala Ala Ile

35 40 45

Asn Tyr Tyr Leu Val Ala Leu Leu Ala Leu Ala Asp Lys Asn Asn Pro

50 55 60

Val Leu Gly Pro Leu Ile Ser Gln Met Asp Asn Pro Gln Ser Pro Tyr

65 70 75 80

His Val Trp Gly Ser Phe Arg Arg Gln Gly Arg Gln Arg Thr Gly Leu

85 90 95

Ser Gln Ala Val Ala Pro Tyr Ile Thr Pro Gly Asn Asn Ala Pro Thr

100 105 110

Leu Asp Glu Val Phe Arg Ser Ile Leu Ala Gly Asn Pro Thr Asp Arg

115 120 125

Ala Thr Leu Asp Ala Ala Leu Met Gln Leu Leu Lys Ala Cys Asp Gly
130 135 140
Ala Gly Ala Ile Gln Gln Glu Gly Arg Ser Tyr Trp Pro Lys Phe Cys
145 150 155 160
Asp Pro Asp Ser Thr Ala Asn Phe Ala Gly Asp Pro Ala Met Leu Arg
165 170 175
Arg Glu Gln His Arg Leu Leu Leu Pro Gln Val Leu His Asp Pro Ala

180 185 190
Ile Thr His Asp Ser Pro Ala Leu Gly Ser Phe Asp Thr Tyr Ser Ile
195 200 205
Ala Thr Pro Asp Thr Arg Thr Pro Gln Leu Thr Gly Pro Lys Ala Arg
210 215 220
Ala Arg Leu Glu Gln Ala Ile Thr Leu Trp Arg Val Arg Leu Pro Glu
225 230 235 240
Ser Ala Ala Asp Phe Asp Arg Leu Ala Ser Ser Leu Lys Lys Ile Pro

245 250 255
Asp Asp Asp Ser Arg Leu Asn Leu Gln Gly Tyr Val Gly Ser Ser Ala
260 265 270
Lys Gly Glu Val Gln Ala Arg Leu Phe Ala Leu Leu Leu Phe Arg His
275 280 285
Leu Glu Arg Ser Ser Phe Thr Leu Gly Leu Leu Arg Ser Ala Thr Pro
290 295 300
Pro Pro Lys Asn Ala Glu Thr Pro Pro Pro Ala Gly Val Pro Leu Pro

305 310 315 320
Ala Ala Ser Ala Ala Asp Pro Val Arg Ile Ala Arg Gly Lys Arg Ser
325 330 335
Phe Val Phe Arg Ala Phe Thr Ser Leu Pro Cys Trp His Gly Gly Asp
340 345 350
Asn Ile His Pro Thr Trp Lys Ser Phe Asp Ile Ala Ala Phe Lys Tyr
355 360 365
Ala Leu Thr Val Ile Asn Gln Ile Glu Glu Lys Thr Lys Glu Arg Gln

370 375 380

Lys Glu Cys Ala Glu Leu Glu Thr Asp Phe Asp Tyr Met His Gly Arg

385 390 395 400

Leu Ala Lys Ile Pro Val Lys Tyr Thr Thr Gly Glu Ala Glu Pro Pro

405 410 415

Pro Ile Leu Ala Asn Asp Leu Arg Ile Pro Leu Leu Arg Glu Leu Leu

420 425 430

Gln Asn Ile Lys Val Asp Thr Ala Leu Thr Asp Gly Glu Ala Val Ser

435 440 445

Tyr Gly Leu Gln Arg Arg Thr Ile Arg Gly Phe Arg Glu Leu Arg Arg

450 455 460

Ile Trp Arg Gly His Ala Pro Ala Gly Thr Val Phe Ser Ser Glu Leu

465 470 475 480

Lys Glu Lys Leu Ala Gly Glu Leu Arg Gln Phe Gln Thr Asp Asn Ser

485 490 495

Thr Thr Ile Gly Ser Val Gln Leu Phe Asn Glu Leu Ile Gln Asn Pro

500 505 510

Lys Tyr Trp Pro Ile Trp Gln Ala Pro Asp Val Glu Thr Ala Arg Gln

515 520 525

Trp Ala Asp Ala Gly Phe Ala Asp Asp Pro Leu Ala Ala Leu Val Gln

530 535 540

Glu Ala Glu Leu Gln Glu Asp Ile Asp Ala Leu Lys Ala Pro Val Lys

545 550 555 560

Leu Thr Pro Ala Asp Pro Glu Tyr Ser Arg Arg Gln Tyr Asp Phe Asn

565 570 575

Ala Val Ser Lys Phe Gly Ala Gly Ser Arg Ser Ala Asn Arg His Glu

580 585 590

Pro Gly Gln Thr Glu Arg Gly His Asn Thr Phe Thr Thr Glu Ile Ala

595 600 605

Ala Arg Asn Ala Ala Asp Gly Asn Arg Trp Arg Ala Thr His Val Arg

610 615 620

Ile His Tyr Ser Ala Pro Arg Leu Leu Arg Asp Gly Leu Arg Arg Pro

625 630 635 640

Asp Thr Asp Gly Asn Glu Ala Leu Glu Ala Val Pro Trp Leu Gln Pro

645 650 655

Met Met Glu Ala Leu Ala Pro Leu Pro Thr Leu Pro Gln Asp Leu Thr

660 665 670

Gly Met Pro Val Phe Leu Met Pro Asp Val Thr Leu Ser Gly Glu Arg

675 680 685

Arg Ile Leu Leu Asn Leu Pro Val Thr Leu Glu Pro Ala Ala Leu Val

690 695 700

Glu Gln Leu Gly Asn Ala Gly Arg Trp Gln Asn Gln Phe Phe Gly Ser

705 710 715 720

Arg Glu Asp Pro Phe Ala Leu Arg Trp Pro Ala Asp Gly Ala Val Lys

725 730 735

Thr Ala Lys Gly Lys Thr His Ile Pro Trp His Gln Asp Arg Asp His

740 745 750

Phe Thr Val Leu Gly Val Asp Leu Gly Thr Arg Asp Ala Gly Ala Leu

755 760 765

Ala Leu Leu Asn Val Thr Ala Gln Lys Pro Ala Lys Pro Val His Arg

770 775 780

Ile Ile Gly Glu Ala Asp Gly Arg Thr Trp Tyr Ala Ser Leu Ala Asp

785 790 795 800

Ala Arg Met Ile Arg Leu Pro Gly Glu Asp Ala Arg Leu Phe Val Arg

805 810 815

Gly Lys Leu Val Gln Glu Pro Tyr Gly Glu Arg Gly Arg Asn Ala Ser

820 825 830

Leu Leu Glu Trp Glu Asp Ala Arg Asn Ile Ile Leu Arg Leu Gly Gln

835 840 845

Asn Pro Asp Glu Leu Leu Gly Ala Asp Pro Arg Arg His Ser Tyr Pro

850 855 860

Glu Ile Asn Asp Lys Leu Leu Val Ala Leu Arg Arg Ala Gln Ala Arg

865 870 875 880

Leu Ala Arg Leu Gln Asn Arg Ser Trp Arg Leu Arg Asp Leu Ala Glu

885 890 895

Ser Asp Lys Ala Leu Asp Glu Ile His Ala Glu Arg Ala Gly Glu Lys

900 905 910

Pro Ser Pro Leu Pro Pro Leu Ala Arg Asp Asp Ala Ile Lys Ser Thr

915 920 925

Asp Glu Ala Leu Leu Ser Gln Arg Asp Ile Ile Arg Arg Ser Phe Val

930 935 940

Gln Ile Ala Asn Leu Ile Leu Pro Leu Arg Gly Arg Arg Trp Glu Trp

945 950 955 960

Arg Pro His Val Glu Val Pro Asp Cys His Ile Leu Ala Gln Ser Asp

965 970 975

Pro Gly Thr Asp Asp Thr Lys Arg Leu Val Ala Gly Gln Arg Gly Ile

980 985 990

Ser His Glu Arg Ile Glu Gln Ile Glu Glu Leu Arg Arg Arg Cys Gln

995 1000 1005

Ser Leu Asn Arg Ala Leu Arg His Lys Pro Gly Glu Arg Pro Val

1010 1015 1020

Leu Gly Arg Pro Ala Lys Gly Glu Glu Ile Ala Asp Pro Cys Pro

1025 1030 1035

Ala Leu Leu Glu Lys Ile Asn Arg Leu Arg Asp Gln Arg Val Asp

1040 1045 1050

Gln Thr Ala His Ala Ile Leu Ala Ala Ala Leu Gly Val Arg Leu

1055 1060 1065

Arg Ala Pro Ser Lys Asp Arg Ala Glu Arg Arg His Arg Asp Ile

1070 1075 1080

His Gly Glu Tyr Glu Arg Phe Arg Ala Pro Ala Asp Phe Val Val

1085 1090 1095

Ile Glu Asn Leu Ser Arg Tyr Leu Ser Ser Gln Asp Arg Ala Arg

1100 1105 1110

Ser Glu Asn Thr Arg Leu Met Gln Trp Cys His Arg Gln Ile Val

1115 1120 1125

Gln Lys Leu Arg Gln Leu Cys Glu Thr Tyr Gly Ile Pro Val Leu

1130 1135 1140

Ala Val Pro Ala Ala Tyr Ser Ser Arg Phe Ser Ser Arg Asp Gly

1145 1150 1155

Ser Ala Gly Phe Arg Ala Val His Leu Thr Pro Asp His Arg His

1160 1165 1170

Arg Met Pro Trp Ser Arg Ile Leu Ala Arg Leu Lys Ala His Glu

1175 1180 1185

Glu Asp Gly Lys Arg Leu Glu Lys Thr Val Leu Asp Glu Ala Arg

1190 1195 1200

Ala Val Arg Gly Leu Phe Asp Arg Leu Asp Arg Phe Asn Ala Gly

1205 1210 1215

His Val Pro Gly Lys Pro Trp Arg Thr Leu Leu Ala Pro Leu Pro

1220 1225 1230

Gly Gly Pro Val Phe Val Pro Leu Gly Asp Ala Thr Pro Met Gln

1235 1240 1245

Ala Asp Leu Asn Ala Ala Ile Asn Ile Ala Leu Arg Gly Ile Ala

1250 1255 1260

Ala Pro Asp Arg His Asp Ile His His Arg Leu Arg Ala Glu Asn

1265 1270 1275

Lys Lys Arg Ile Leu Ser Leu Arg Leu Gly Thr Gln Arg Glu Lys

1280 1285 1290

Ala Arg Trp Pro Gly Gly Ala Pro Ala Val Thr Leu Ser Thr Pro

1295 1300 1305

Asn Asn Gly Ala Ser Pro Glu Asp Ser Asp Ala Leu Pro Glu Arg

1310 1315 1320

Val Ser Asn Leu Phe Val Asp Ile Ala Gly Val Ala Asn Phe Glu

1325 1330 1335

Arg Val Thr Ile Glu Gly Val Ser Gln Lys Phe Ala Thr Gly Arg

1340 1345 1350

Gly Leu Trp Ala Ser Val Lys Gln Arg Ala Trp Asn Arg Val Ala

1355 1360 1365

Arg Leu Asn Glu Thr Val Thr Asp Asn Asn Arg Asn Glu Glu Glu

1370 1375 1380

Asp Asp Ile Pro Met

1385

<210> 532

<211> 747

<212> PRT

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 532

Met Ala Thr Lys Ser Tyr Arg Ala Arg Ile Leu Thr Asp Ser Arg Leu

1 5 10 15

Ala Ala Ala Leu Asp Arg Thr His Val Val Phe Val Glu Ser Leu Lys

20 25 30

Gln Met Ile Asn Thr Tyr Leu Arg Met Gln Asn Gly Lys Phe Gly Pro

35 40 45

Asp His Lys Lys Leu Ala Gln Ile Met Leu Ser Arg Ser Asn Thr Phe

50 55 60

Ala His Gly Val Met Asp Gln Ile Thr Arg Asp Gln Pro Thr Ser Thr

65 70 75 80

Leu Asp Glu Glu Trp Thr Asp Leu Ala Arg Arg Ile His Lys Thr Thr

85 90 95

Gly Pro Leu Phe Leu Gln Ala Glu Arg Phe Ala Thr Val Lys Asn Arg

100 105 110

Ala Ile His Thr Lys Ser Arg Gly Lys Val Ile Pro Ser Pro Glu Thr

115 120 125

Leu Ala Val Pro Ala Lys Phe Trp His Gln Val Cys Asp Ser Ala Ser

130 135 140

Ala Tyr Ile Arg Ser Asn Arg Glu Leu Met Gln Gln Trp Arg Lys Asp

145 150 155 160

Arg Ala Ala Trp Leu Lys Asp Lys Asn Glu Trp Gln Gln Lys His Pro

165 170 175

Glu Phe Met Gln Phe Tyr Asn Gly Pro Tyr Gln Asn Phe Leu Lys Leu

180 185 190

Cys Asp Asp Asp Arg Ile Thr Ser Gln Leu Ala Ala Glu Gln Gln Pro

195 200 205

Thr Ala Ser Lys Asn Asn Arg Pro Arg Lys Thr Gly Lys Arg Phe Ala

210 215 220

Arg Trp His Leu Trp Tyr Lys Trp Leu Ser Glu Asn Pro Glu Ile Ile

225 230 235 240

Glu Trp Arg Asn Lys Ala Ser Ala Ser Asp Phe Lys Thr Val Thr Asp

245 250 255

Asp Val Arg Lys Gln Ile Ile Thr Lys Tyr Pro Gln Gln Asn Lys Tyr

260 265 270

Ile Thr Arg Leu Leu Asp Trp Leu Glu Asp Asn Asn Pro Glu Leu Lys

275 280 285

Thr Leu Glu Asn Leu Arg Arg Thr Tyr Val Lys Lys Phe Asp Ser Phe

290 295 300

Lys Arg Pro Pro Thr Leu Thr Leu Pro Ser Pro Tyr Arg His Pro Tyr

305 310 315 320

Trp Phe Thr Met Glu Leu Asp Gln Phe Tyr Lys Lys Ala Asp Phe Glu

325 330 335

Asn Gly Thr Ile Gln Leu Leu Leu Ile Asp Glu Asp Asp Asp Gly Asn

340 345 350

Trp Phe Phe Asn Trp Met Pro Ala Ser Leu Lys Pro Asp Pro Arg Leu

355 360 365

Val Pro Ser Trp Arg Ala Glu Thr Phe Glu Thr Glu Gly Arg Phe Pro

370 375 380

Pro Tyr Leu Gly Gly Lys Ile Gly Lys Lys Leu Ser Arg Pro Ala Pro

385 390 395 400

Thr Asp Ala Glu Arg Lys Ala Gly Ile Ala Gly Ala Lys Leu Met Ile
405 410 415
Lys Asn Asn Arg Ser Glu Leu Leu Phe Thr Val Phe Glu Gln Asp Cys

420 425 430
Pro Pro Arg Val Lys Trp Ala Lys Thr Lys Asn Arg Lys Cys Pro Ala
435 440 445
Asp Asn Ala Phe Ser Ser Asp Gly Lys Thr Arg Lys Pro Leu Arg Ile
450 455 460
Leu Ser Ile Asp Leu Gly Ile Arg His Ile Gly Ala Phe Ala Leu Thr
465 470 475 480
Gln Gly Thr Arg Asn Asp Ser Ala Trp Gln Thr Glu Ser Leu Lys Lys

485 490 495
Gly Ile Ile Asn Ser Pro Ser Ile Pro Pro Leu Arg Gln Val Arg Arg
500 505 510
His Asp Tyr Asp Leu Lys Arg Lys Arg Arg Arg His Gly Lys Pro Val
515 520 525
Lys Gly Gln Arg Ser Asn Ala Asn Leu Gln Ala His Arg Thr Asn Met
530 535 540
Ala Gln Asp Arg Phe Lys Lys Gly Ala Ser Ala Ile Val Ser Leu Ala

545 550 555 560
Arg Glu His Ser Ala Asp Leu Ile Leu Phe Glu Asn Leu His Ser Leu
565 570 575
Lys Phe Ser Ala Phe Asp Glu Arg Trp Met Asn Arg Gln Leu Arg Asp
580 585 590
Met Asn Arg Arg His Ile Val Glu Leu Val Ser Glu Gln Ala Pro Glu
595 600 605
Phe Gly Ile Thr Val Lys Asp Asp Ile Asn Pro Trp Met Thr Ser Arg

610 615 620
Ile Cys Ser Asn Cys Asn Leu Pro Gly Phe Arg Phe Ser Met Lys Lys
625 630 635 640
Lys Asn Pro Tyr Arg Glu Lys Leu Pro Arg Glu Lys Cys Thr Asp Phe

645 650 655

Gly Tyr Pro Val Trp Glu Pro Gly Gly His Leu Phe Arg Cys Pro His

660 665 670

Cys Asp His Arg Val Asn Ala Asp Ile Asn Ala Ala Ala Asn Leu Ala

675 680 685

Asn Lys Phe Phe Gly Leu Gly Tyr Trp Asn Asn Gly Leu Lys Tyr Asp

690 695 700

Ala Glu Thr Lys Thr Phe Thr Val His Thr Asp Lys Lys Thr Pro Pro

705 710 715 720

Leu Ile Phe Lys Pro Arg Pro Gln Phe Asp Leu Trp Ala Asp Ser Val

725 730 735

Lys Thr Arg Lys Gln Leu Gly Pro Asp Pro Phe

740 745

<210> 533

<211> 743

<212> PRT

<213> Unknown

<220><223> Member of the phylum Planktomycetes

<400> 533

Met Ser Val Arg Ser Phe Gln Ala Arg Val Glu Cys Asp Lys Gln Thr

1 5 10 15

Met Glu His Leu Trp Arg Thr His Lys Val Phe Asn Glu Arg Leu Pro

20 25 30

Glu Ile Ile Lys Ile Leu Phe Lys Met Lys Arg Gly Glu Cys Gly Gln

35 40 45

Asn Asp Lys Gln Lys Ser Leu Tyr Lys Ser Ile Ser Gln Ser Ile Leu

50 55 60

Glu Ala Asn Ala Gln Asn Ala Asp Tyr Leu Leu Asn Ser Val Ser Ile

65 70 75 80

Lys Gly Trp Lys Pro Gly Thr Ala Lys Lys Tyr Arg Asn Ala Ser Phe

85 90 95

Thr Trp Ala Asp Asp Ala Ala Lys Leu Ser Ser Gln Gly Ile His Val

100 105 110

Tyr Asp Lys Lys Gln Val Leu Gly Asp Leu Pro Gly Met Met Ser Gln
115 120 125
Met Val Cys Arg Gln Ser Val Glu Ala Ile Ser Gly His Ile Glu Leu
130 135 140
Thr Lys Lys Trp Glu Lys Glu His Asn Glu Trp Leu Lys Glu Lys Glu
145 150 155 160
Lys Trp Glu Ser Glu Asp Glu His Lys Lys Tyr Leu Asp Leu Arg Glu
165 170 175

Lys Phe Glu Gln Phe Glu Gln Ser Ile Gly Gly Lys Ile Thr Lys Arg
180 185 190
Arg Gly Arg Trp His Leu Tyr Leu Lys Trp Leu Ser Asp Asn Pro Asp
195 200 205
Phe Ala Ala Trp Arg Gly Asn Lys Ala Val Ile Asn Pro Leu Ser Glu
210 215 220
Lys Ala Gln Ile Arg Ile Asn Lys Ala Lys Pro Asn Lys Lys Asn Ser
225 230 235 240

Val Glu Arg Asp Glu Phe Phe Lys Ala Asn Pro Glu Met Lys Ala Leu
245 250 255
Asp Asn Leu His Gly Tyr Tyr Glu Arg Asn Phe Val Arg Arg Arg Lys
260 265 270
Thr Lys Lys Asn Pro Asp Gly Phe Asp His Lys Pro Thr Phe Thr Leu
275 280 285
Pro His Pro Thr Ile His Pro Arg Trp Phe Val Phe Asn Lys Pro Lys
290 295 300

Thr Asn Pro Glu Gly Tyr Arg Lys Leu Ile Leu Pro Lys Lys Ala Gly
305 310 315 320
Asp Leu Gly Ser Leu Glu Met Arg Leu Leu Thr Gly Glu Lys Asn Lys
325 330 335
Gly Asn Tyr Pro Asp Asp Trp Ile Ser Val Lys Phe Lys Ala Asp Pro
340 345 350

Arg Leu Ser Leu Ile Arg Pro Val Lys Gly Arg Arg Val Val Arg Lys
355 360 365

Gly Lys Glu Gln Gly Gln Thr Lys Glu Thr Asp Ser Tyr Glu Phe Phe
370 375 380
Asp Lys His Leu Lys Lys Trp Arg Pro Ala Lys Leu Ser Gly Val Lys
385 390 395 400
Leu Ile Phe Pro Asp Lys Thr Pro Lys Ala Ala Tyr Leu Tyr Phe Thr
405 410 415
Cys Asp Ile Pro Asp Glu Pro Leu Thr Glu Thr Ala Lys Lys Ile Gln
420 425 430

Trp Leu Glu Thr Gly Asp Val Thr Lys Lys Gly Lys Lys Arg Lys Lys
435 440 445
Lys Val Leu Pro His Gly Leu Val Ser Cys Ala Val Asp Leu Ser Met
450 455 460
Arg Arg Gly Thr Thr Gly Phe Ala Thr Leu Cys Arg Tyr Glu Asn Gly
465 470 475 480
Lys Ile His Ile Leu Arg Ser Arg Asn Leu Trp Val Gly Tyr Lys Glu
485 490 495

Gly Lys Gly Cys His Pro Tyr Arg Trp Thr Glu Gly Pro Asp Leu Gly
500 505 510
His Ile Ala Lys His Lys Arg Glu Ile Arg Ile Leu Arg Ser Lys Arg
515 520 525
Gly Lys Pro Val Lys Gly Glu Glu Ser His Ile Asp Leu Gln Lys His
530 535 540
Ile Asp Tyr Met Gly Glu Asp Arg Phe Lys Lys Ala Ala Arg Thr Ile
545 550 555 560

Val Asn Phe Ala Leu Asn Thr Glu Asn Ala Ala Ser Lys Asn Gly Phe
565 570 575
Tyr Pro Arg Ala Asp Val Leu Leu Leu Glu Asn Leu Glu Gly Leu Ile
580 585 590
Pro Asp Ala Glu Lys Glu Arg Gly Ile Asn Arg Ala Leu Ala Gly Trp

595 600 605

Asn Arg Arg His Leu Val Glu Arg Val Ile Glu Met Ala Lys Asp Ala
610 615 620

Gly Phe Lys Arg Arg Val Phe Glu Ile Pro Pro Tyr Gly Thr Ser Gln
625 630 635 640

Val Cys Ser Lys Cys Gly Ala Leu Gly Arg Arg Tyr Ser Ile Ile Arg
645 650 655

Glu Asn Asn Arg Arg Glu Ile Arg Phe Gly Tyr Val Glu Lys Leu Phe
660 665 670

Ala Cys Pro Asn Cys Gly Tyr Cys Ala Asn Ala Asp His Asn Ala Ser
675 680 685

Val Asn Leu Asn Arg Arg Phe Leu Ile Glu Asp Ser Phe Lys Ser Tyr
690 695 700

Tyr Asp Trp Lys Arg Leu Ser Glu Lys Lys Gln Lys Glu Glu Ile Glu
705 710 715 720

Thr Ile Glu Ser Lys Leu Met Asp Lys Leu Cys Ala Met His Lys Ile
725 730 735

Ser Arg Gly Ser Ile Ser Lys
740

<210> 534

<211> 1119

<212> PRT

<213> Unknown

<220><223> Member of the Spirochaetes

<400> 534

Met Ser Phe Thr Ile Ser Tyr Pro Phe Lys Leu Ile Ile Lys Asn Lys
1 5 10 15

Asp Glu Ala Lys Ala Leu Leu Asp Thr His Gln Tyr Met Asn Glu Gly
20 25 30

Val Lys Tyr Tyr Leu Glu Lys Leu Leu Met Phe Arg Gln Glu Lys Ile
35 40 45

Phe Ile Gly Glu Asp Glu Thr Gly Lys Arg Ile Tyr Ile Glu Glu Thr

50 55 60

Glu Tyr Lys Lys Gln Ile Glu Glu Phe Tyr Leu Ile Lys Lys Thr Glu

65 70 75 80

Leu Gly Arg Asn Leu Thr Leu Thr Leu Asp Glu Phe Lys Thr Leu Met

85 90 95

Arg Glu Leu Tyr Ile Cys Leu Val Ser Ser Ser Met Glu Asn Lys Lys

100 105 110

Gly Phe Pro Asn Ala Gln Gln Ala Ser Leu Asn Ile Phe Ser Pro Leu

115 120 125

Phe Asp Ala Glu Ser Lys Gly Tyr Ile Leu Lys Glu Glu Asn Asn Asn

130 135 140

Ile Ser Leu Ile His Lys Asp Tyr Gly Lys Ile Leu Leu Lys Arg Leu

145 150 155 160

Arg Asp Asn Asn Leu Ile Pro Ile Phe Thr Lys Phe Thr Asp Ile Lys

165 170 175

Lys Ile Thr Ala Lys Leu Ser Pro Thr Ala Leu Asp Arg Met Ile Phe

180 185 190

Ala Gln Ala Ile Glu Lys Leu Leu Ser Tyr Glu Ser Trp Cys Lys Leu

195 200 205

Met Ile Lys Glu Arg Phe Asp Lys Glu Val Lys Ile Lys Glu Leu Glu

210 215 220

Asn Lys Cys Glu Asn Lys Gln Glu Arg Asp Lys Ile Phe Glu Ile Leu

225 230 235 240

Glu Lys Tyr Glu Glu Glu Arg Gln Lys Thr Phe Glu Gln Asp Ser Gly

245 250 255

Phe Ala Lys Lys Gly Lys Phe Tyr Ile Thr Gly Arg Met Leu Lys Gly

260 265 270

Phe Asp Glu Ile Lys Glu Lys Trp Leu Lys Glu Lys Asp Arg Ser Glu

275 280 285

Gln Asn Leu Ile Asn Ile Leu Asn Lys Tyr Gln Thr Asp Asn Ser Lys

290 295 300

Leu Val Gly Asp Arg Asn Leu Phe Glu Phe Ile Ile Lys Leu Glu Asn

305 310 315 320

Gln Cys Leu Trp Asn Gly Asp Ile Asp Tyr Leu Lys Ile Lys Arg Asp

325 330 335

Ile Asn Lys Asn Gln Ile Trp Leu Asp Arg Pro Glu Met Pro Arg Phe

340 345 350

Thr Met Pro Asp Phe Lys Lys His Pro Leu Trp Tyr Arg Tyr Glu Asp

355 360 365

Pro Ser Asn Ser Asn Phe Arg Asn Tyr Lys Ile Glu Val Val Lys Asp

370 375 380

Glu Asn Tyr Ile Thr Ile Pro Leu Ile Thr Glu Arg Asn Asn Glu Tyr

385 390 395 400

Phe Glu Glu Asn Tyr Thr Phe Asn Leu Ala Lys Leu Lys Lys Leu Ser

405 410 415

Glu Asn Ile Thr Phe Ile Pro Lys Ser Lys Asn Lys Glu Phe Glu Phe

420 425 430

Ile Asp Ser Asn Asp Glu Glu Glu Asp Lys Lys Asp Gln Lys Lys Ser

435 440 445

Lys Gln Tyr Ile Lys Tyr Cys Asp Thr Ala Lys Asn Thr Ser Tyr Gly

450 455 460

Lys Ser Gly Gly Ile Arg Leu Tyr Phe Asn Arg Asn Glu Leu Glu Asn

465 470 475 480

Tyr Lys Asp Gly Lys Lys Met Asp Ser Tyr Thr Val Phe Thr Leu Ser

485 490 495

Ile Arg Asp Tyr Lys Ser Leu Phe Ala Lys Glu Lys Leu Gln Pro Gln

500 505 510

Ile Phe Asn Thr Val Asp Asn Lys Ile Thr Ser Leu Lys Ile Gln Lys

515 520 525

Lys Phe Gly Asn Glu Glu Gln Thr Asn Phe Leu Ser Tyr Phe Thr Gln

530 535 540

Asn Gln Ile Thr Lys Lys Asp Trp Met Asp Glu Lys Thr Phe Gln Asn
545 550 555 560
Val Lys Glu Leu Asn Glu Gly Ile Arg Val Leu Ser Val Asp Leu Gly

565 570 575
Gln Arg Phe Phe Ala Ala Val Ser Cys Phe Glu Ile Met Ser Glu Ile
580 585 590
Asp Asn Asn Lys Leu Phe Phe Asn Leu Asn Asp Gln Asn His Lys Ile
595 600 605
Ile Arg Ile Asn Asp Lys Asn Tyr Tyr Ala Lys His Ile Tyr Ser Lys
610 615 620
Thr Ile Lys Leu Ser Gly Glu Asp Asp Asp Leu Tyr Lys Glu Arg Lys

625 630 635 640
Ile Asn Lys Asn Tyr Lys Leu Ser Tyr Gln Glu Arg Lys Asn Lys Ile
645 650 655
Gly Ile Phe Thr Arg Gln Ile Asn Lys Leu Asn Gln Leu Leu Lys Ile
660 665 670
Ile Arg Asn Asp Glu Ile Asp Lys Glu Lys Phe Lys Glu Leu Ile Glu
675 680 685
Thr Thr Lys Arg Tyr Val Lys Asn Thr Tyr Asn Asp Gly Ile Ile Asp

690 695 700
Trp Asn Asn Val Asp Asn Lys Ile Leu Ser Tyr Glu Asn Lys Glu Asp
705 710 715 720
Val Ile Asn Leu His Lys Glu Leu Asp Lys Lys Leu Glu Ile Asp Phe
725 730 735
Lys Glu Phe Ile Arg Glu Cys Arg Lys Pro Ile Phe Arg Ser Gly Gly
740 745 750
Leu Ser Met Gln Arg Ile Asp Phe Leu Glu Lys Leu Asn Lys Leu Lys

755 760 765
Arg Lys Trp Val Ala Arg Thr Gln Lys Ser Ala Glu Ser Ile Val Leu
770 775 780
Thr Pro Lys Phe Gly Tyr Lys Leu Lys Glu His Ile Asn Glu Leu Lys

785 790 795 800

Asp Asn Arg Val Lys Gln Gly Val Asn Tyr Ile Leu Met Thr Ala Leu

805 810 815

Gly Tyr Ile Lys Asp Asn Glu Ile Lys Asn Asp Ser Lys Lys Lys Gln

820 825 830

Lys Glu Asp Trp Val Lys Lys Asn Arg Ala Cys Gln Ile Ile Leu Met

835 840 845

Glu Lys Leu Thr Glu Tyr Thr Phe Ala Glu Asp Arg Pro Arg Glu Glu

850 855 860

Asn Ser Lys Leu Arg Met Trp Ser His Arg Gln Ile Phe Asn Phe Leu

865 870 875 880

Gln Gln Lys Ala Ser Leu Trp Gly Ile Leu Val Gly Asp Val Phe Ala

885 890 895

Pro Tyr Thr Ser Lys Cys Leu Ser Asp Asn Asn Ala Pro Gly Ile Arg

900 905 910

Cys His Gln Val Thr Lys Lys Asp Leu Ile Asp Asn Ser Trp Phe Leu

915 920 925

Lys Ile Val Val Lys Asp Asp Ala Phe Cys Asp Leu Ile Glu Ile Asn

930 935 940

Lys Glu Asn Val Lys Asn Lys Ser Ile Lys Ile Asn Asp Ile Leu Pro

945 950 955 960

Leu Arg Gly Gly Glu Leu Phe Ala Ser Ile Lys Asp Gly Lys Leu His

965 970 975

Ile Val Gln Ala Asp Ile Asn Ala Ser Arg Asn Ile Ala Lys Arg Phe

980 985 990

Leu Ser Gln Ile Asn Pro Phe Arg Val Val Leu Lys Lys Asp Lys Asp

995 1000 1005

Glu Thr Phe His Leu Lys Asn Glu Pro Asn Tyr Leu Lys Asn Tyr

1010 1015 1020

Tyr Ser Ile Leu Asn Phe Val Pro Thr Asn Glu Glu Leu Thr Phe

1025 1030 1035

Phe Lys Val Glu Glu Asn Lys Asp Ile Lys Pro Thr Lys Arg Ile

1040 1045 1050

Lys Met Asp Lys His Glu Lys Glu Ser Thr Asp Glu Gly Asp Asp

1055 1060 1065

Tyr Ser Lys Asn Gln Ile Ala Leu Phe Arg Asp Asp Ser Gly Ile

1070 1075 1080

Phe Phe Asp Lys Ser Leu Trp Val Asp Gly Lys Ile Phe Trp Ser

1085 1090 1095

Val Val Lys Asn Lys Met Thr Lys Leu Leu Arg Glu Arg Asn Asn

1100 1105 1110

Lys Lys Asn Gly Ser Lys

1115

<210> 535

<211> 1447

<212> PRT

<213> Unknown

<220><223> Member of the phylum Verrucomicrobiaceae

<400> 535

Met Pro Leu Ser Arg Ile Tyr Gln Gly Arg Thr Asn Ser Leu Ile Ile

1 5 10 15

Leu Thr Pro Thr Pro Gln Glu Pro Trp Asp His Lys Ala Leu Ala Arg

20 25 30

Phe Asp Ser Pro Leu Trp Arg His His Ala Leu Phe Gln Asp Ala Val

35 40 45

Asn Tyr Tyr Gln Leu Cys Leu Val Ala Leu Ala Ser Ser Asp Gly Thr

50 55 60

Arg Pro Leu Ser Lys Leu His Glu Gln Met Lys Ala Ser Trp Asp Glu

65 70 75 80

Ala Lys Thr Asp Thr Glu Asp Ser Trp Arg Val Arg Leu Ala Arg Arg

85 90 95

Leu Gly Ile Pro Ala Ala Ser Leu Phe Glu Ala Ala Leu Ala Lys Val

100 105 110

Leu Glu Gly Asn Glu Ala Pro Glu Arg Ala Arg Glu Leu Ala Gly Glu
115 120 125
Leu Leu Leu Asp Lys Ile Glu Gly Asp Ile Gln Gln Ala Gly Arg Gly

130 135 140
Tyr Trp Pro Arg Phe Cys Asp Pro Lys Ala Asn Pro Thr Tyr Asp Tyr
145 150 155 160
Ser Ala Thr Ala Arg Ala Ser Ala Ser Gly Leu Thr Lys Leu Ala Ala
165 170 175
Val Ile His Ala Glu Asn Val Thr Glu Glu Ala Leu Lys Gln Val Ala
180 185 190
Ala Glu Met Asp Leu Ser Trp Thr Val Lys Leu Gln Pro Asp Lys Asn

195 200 205
Phe Val Gly Ala Glu Ala Arg Ala Arg Leu Leu Glu Ala Ala His His
210 215 220
Phe Ile Lys Val Ala Glu Ser Pro Pro Thr Lys Leu Ala Glu Val Leu
225 230 235 240
Ala Arg Phe Pro Asp Gly Leu Ala Leu Trp Gln Ala Leu Pro Glu Lys
245 250 255
Ile Ala Ala Leu Pro Glu Glu Thr Gln Val Pro Arg Asn Arg Lys Ala

260 265 270
Ser Pro Asp Leu Thr Phe Ala Thr Leu Leu Phe Gln His Phe Pro Ser
275 280 285
Leu Phe Thr Ala Ala Val Leu Gly Leu Ser Val Gly Lys Pro Lys Ser
290 295 300
Val Lys Ala Pro Lys Val Val Glu Lys Val Ser Ala Arg Arg Lys Ala
305 310 315 320
Asn Ala Val Thr Gln Ala Val Val Ile Glu Glu Pro Glu Ile Asp Phe

325 330 335
Ala Glu Leu Gly Asp Asp Pro Ile Lys Leu Ala Arg Gly Glu Arg Gly
340 345 350
Phe Val Phe Pro Ala Phe Thr Ser Leu Ser Phe Trp Ala Val Pro Gly

355 360 365

Pro His Val Pro Val Trp Lys Glu Phe Asp Ile Ala Ala Phe Lys Glu

370 375 380

Ala Leu Lys Thr Val Asn Gln Phe Lys Leu Lys Thr Ser Glu Arg Asn

385 390 395 400

Ala Leu Leu Ala Glu Ala Gln Arg Arg Leu Asp Tyr Met Asp Glu Lys

405 410 415

Thr His Asp Trp Lys Thr Gly Asp Ser Asp Glu Pro Gly His Ile Pro

420 425 430

Pro Arg Leu Lys Ser Asp Pro Asn Phe Thr Leu Ile Gln Ala Leu Thr

435 440 445

Gln Asp Glu Gly Val Ser Asn Lys Ala Thr Gly Asp Gln His Ile Pro

450 455 460

Lys Gly Val Tyr Thr Gly Gly Leu Arg Gly Phe Tyr Ala Ile Lys Lys

465 470 475 480

Asp Trp Cys Glu Leu Trp Glu Arg Lys Ala Asp Lys Ser Gln Gly Thr

485 490 495

Pro Thr Glu Glu Glu Leu Ile Ser Ile Val Thr Asp Tyr Gln Arg Asp

500 505 510

His Val Tyr Asp Val Gly Asp Val Gly Leu Phe Arg Ala Leu Cys Glu

515 520 525

Pro Arg Phe Trp Pro Leu Trp Gln Pro Leu Thr Asp Glu Gln Glu Ala

530 535 540

Glu Arg Ile Lys Ala Gly Arg Ala Lys Asp Met Ile Ser Ala Tyr Arg

545 550 555 560

Val Trp Leu Glu Leu Gln Glu Asp Val Val Arg Leu Ala Gln Pro Ile

565 570 575

Arg Phe Thr Pro Ala His Ala Glu Asn Ser Arg Arg Leu Phe Met Phe

580 585 590

Ser Asp Ile Ser Gly Ser His Gly Ala Glu Phe Gly Ser Asp Gly Lys

595 600 605

Ser Leu Glu Val Ser Ile Ala Tyr Asp Val Asp Gly Lys Leu Gln Pro
610 615 620
Val Arg Ala Lys Leu Glu Phe Ser Ala Pro Arg Ala Ala Arg Asp Glu
625 630 635 640
Leu Glu Gly Leu Ser Gly Gly Ser Glu Ser Met Arg Trp Phe Gln Pro

645 650 655
Met Met Lys Ala Leu Asp Cys Pro Glu Val Glu Met Pro Ala Leu Glu
660 665 670
Lys Cys Ala Val Ser Leu Met Pro Asp Val Val Lys Lys Gly Gly Gly
675 680 685
Lys Trp Val Arg Leu Leu Leu Asn Phe Pro Ala Thr Leu Glu Pro Glu
690 695 700
Gly Leu Ile Arg His Ile Gly Lys Gln Ala Met Trp Tyr Lys Gln Phe

705 710 715 720
Asn Gly Thr Tyr Lys Pro Arg Thr Gln Gln Leu Asp Thr Gly Leu His
725 730 735
Leu Tyr Trp Pro Gly Leu Glu Lys Ala Pro Glu Ala Glu Asp Ala Ala
740 745 750
Ala Trp Trp Asn Arg Glu Glu Ile Arg Ala Lys Gly Phe Ser Val Leu
755 760 765
Ser Val Asp Leu Gly Gln Arg Asp Ala Gly Ala Trp Ala Leu Leu Glu

770 775 780
Ser Arg Ser Asp Lys Ala Phe Ser Arg Asn Arg Gln Pro Phe Ile Glu
785 790 795 800
Leu Gly Glu Ala Gly Gly Lys Leu Trp Ser Thr Ala Leu Leu Gly Leu
805 810 815
Gly Met Leu Arg Leu Pro Gly Glu Asp Ala Arg Thr Gly Ala Leu Asp
820 825 830
Asp Gln Gly Lys Arg Ala Val Glu Phe His Gly Lys Ala Gly Arg Asn

835 840 845
Ala Leu Glu Ala Glu Trp Gln Glu Ala Arg Glu Met Ala Leu Leu Phe

850 855 860

Gly Gly Glu Glu Ala Lys Ser Arg Leu Gly Pro Gly Phe Asp His Leu

865 870 875 880

Ser His Ser Lys Gln Asn Glu Glu Leu Leu Arg Ile Leu Ser Arg Ala

885 890 895

Gln Ser Arg Leu Ala Arg Phe His Arg Trp Ser Cys Arg Ile His Glu

900 905 910

Lys Pro Glu Ala Thr Gly Asp Asp Val Ile Asp Tyr Gly Gln Val Asp

915 920 925

Glu Leu Leu Thr Lys Thr Ala Glu Ala Met Leu Glu Asn Leu Lys Ala

930 935 940

Leu Tyr Thr Asn Ala Gly Gly Ile Leu Asp Ser Lys Ser Lys Gln Pro

945 950 955 960

Leu Thr Leu Val Gly Leu Arg Lys Lys Leu Glu Ala Gln Lys Val Glu

965 970 975

Pro Glu Lys Ile Ala Ala Val Leu Lys Pro His Ala Glu Ile Ile Phe

980 985 990

Gln Arg Leu Gly Thr Leu Ile Pro Glu Leu Lys Gln His Leu Arg Val

995 1000 1005

Ser Leu Glu Arg Leu Ala Asn Arg Glu Leu Pro Leu Arg His Arg

1010 1015 1020

Glu Trp Val Trp Asn Glu Ala Phe Glu Lys Leu Glu Gln Gly Asn

1025 1030 1035

Phe Lys Lys Glu Glu Asn Pro Lys Trp Ile Arg Gly Gln Arg Gly

1040 1045 1050

Leu Ser Met Ala Arg Ile Glu Gln Ile Glu Asn Leu Arg Lys Arg

1055 1060 1065

Phe Met Ser Leu Arg Arg Gln Met Ser Leu Ile Pro Gly Glu Gln

1070 1075 1080

Val Lys Gln Gly Val Glu Asp Lys Gly Gln Arg Gln Pro Glu Pro

1085 1090 1095

Cys Glu Asp Ile Leu Asn Lys Leu Asp Arg Met Lys Gln Gln Arg
1100 1105 1110
Val Asn Gln Thr Ala His Leu Ile Leu Ala Gln Ala Leu Gly Leu
1115 1120 1125
Arg Leu Arg Pro His Leu Ala Asn Asp Ala Glu Arg Glu Glu Lys
1130 1135 1140
Asp Ile His Gly Glu Tyr Glu Leu Ile Pro Gly Arg Lys Pro Val

1145 1150 1155
Asp Phe Ile Val Met Glu Asp Leu Ser Arg Tyr Leu Ser Ser Gln
1160 1165 1170
Gly Arg Ala Pro Ser Glu Asn Gly Arg Leu Met Lys Trp Cys His
1175 1180 1185
Arg Ala Val Leu Ala Lys Leu Lys Gln Met Cys Glu Pro Phe Gly
1190 1195 1200
Ile Pro Val Leu Glu Val Pro Ala Ala Tyr Ser Ser Arg Phe Cys

1205 1210 1215
Ala Leu Thr Gly Val Pro Gly Phe Arg Ala Val Glu Val His Asp
1220 1225 1230
Gly Asn Ala Glu Asp Phe Arg Trp Lys Arg Leu Ile Lys Lys Ala
1235 1240 1245
Glu Lys Asp Lys Ser Ser Lys Asp Ala Glu Ala Ala Ala Met Leu
1250 1255 1260
Phe Asp Gln Leu His Asp Leu Asn Ile Glu Ala Arg Glu Ala Arg

1265 1270 1275
Lys Gln Asp Lys Lys Leu Pro Leu Arg Thr Leu Phe Ala Pro Val
1280 1285 1290
Ala Gly Gly Pro Leu Phe Ile Pro Met Val Gly Gly Gly Pro Arg
1295 1300 1305
Gln Ala Asp Met Asn Ala Ala Ile Asn Leu Gly Leu Arg Ala Ile
1310 1315 1320
Ala Ser Pro Thr Cys Leu Arg Ala Arg Pro Lys Ile Arg Ala Glu

1325 1330 1335

Leu Lys Asp Gly Lys His Gln Ala Met Leu Gly Asn Lys Leu Glu

1340 1345 1350

Lys Ala Ala Ala Leu Thr Leu Glu Pro Pro Lys Glu Pro Thr Lys

1355 1360 1365

Glu Leu Ala Ala Gln Lys Arg Thr Asn Phe Phe Leu Asp Glu Lys

1370 1375 1380

Phe Val Gly Lys Phe Asp Thr Ala His Val Thr Thr Ser Gly Lys

1385 1390 1395

Lys Leu Arg Leu Ser Gly Gly Met Ser Leu Trp Lys Ala Ile Lys

1400 1405 1410

Asp Gly Ala Trp Gln Arg Val Lys Lys Ile Asn Asp Ala Arg Ile

1415 1420 1425

Ala Lys Trp Lys Asn Asn Pro Pro Pro Glu Pro Asp Pro Asp Asp

1430 1435 1440

Glu Ile Gln Phe

1445

<210> 536

<211

> 25

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 536

uagaugaauu aaaugugauu agcac 25

<210> 537

<211> 80

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 537

uuugccuaaa gggcaaagaa uacugugcgu gugcuaagga uggaaaaaau ccauucaacc 60

acaggauuac auuauuuauc 80

<210> 538

<211> 105

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 538

uuugccuaaa gggcaaagaa uacugugcgu gugcuaagga uggaaaaaau ccauucaacc 60

acaggauuac auuauuuauc aaaagaugaa uaaugugauu agcac 105

<210> 539

<211> 25

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 539

agcgagcggu cugagaagug gcacu 25

<210> 540

<211> 89

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400

> 540

gucgucuaua ggacggcgag gacaacggga agugccaaug ugcucuuucc aagagcaaac 60

accccguugg cuucaagaug accgcucgc 89

<210> 541

<211> 113

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 541

gucgucuaua ggacggcgag gacaacggga agugccaaug ugcucuuucc aagagcaaac 60

accccguugg cuucucagac cgcucgaaaa cgagcggucu gagaaguggc acu 113

<210> 542

<211> 23

<212> RNA

<213> Artificial Sequence

<220><223

> Synthetic Oligonucleotide

<400> 542

gcagaaauaa ugaugauugg cac 23

<210> 543

<211> 78

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 543

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggauuaucuu auuucugc 78

<210> 544

<211> 104

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 544

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggauuaucuu auuucugcaa aagcagaaau aagaugauug gcac 104

<210> 545

<211> 194

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 545

ucagccaaca ugcucgcuuu gcgaaggcug acggcccgcu cucauuuggc auugccggga 60

gccggaguuu ucggaagaga gugucgacga cugcugaucu ccgcauccgc guccuguucg 120

ccaggccggg ucggguguac ggaucaugcu ggcagcaguc uacgccgaga acauucgcua 180

ccgcgaaugg acac 194

<210> 546

<211> 194

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 546

guggaguaag gucggaguaa cgaccgaacg uuuagcgugc uauaggccgc ugaaugccac 60

acagcgaugu guuuugagug ucaauagcug cugacccaaa ggccaaaagc cgaguagggc 120

uugacugaug cgguuuauau cgcacauagg cggcaguaac acauaucgcg ucaaucaaau 180

uuauugaugg acac 194

<210> 547

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 547

gggcgaucug agaaguggca c 21

<210> 548

<211> 29

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 548

cguugagcgg gcgaucugag aaguggcac 29

<210> 549

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 549

gucggaucgu ugagcgggcg aucugagaag uggcac 36

<210> 550

<211> 41

<212

> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 550

cgatctgaga agtggcacga gaagtcattt aataaggcca c 41

<210> 551

<211> 32

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 551

ccgaucuaua ggacggcaga uucaacggga ug 32

<210> 552

<211> 38

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 552

ccgaucuaua ggacggcaga uucaacggga ugugccaa 38

<210> 553

<211> 61

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 553

ccgaucuaua ggacggcaga uucaacggga ugugccaaug cacucuuucc aggagugaac 60

a 61

<210> 554

<211> 97

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 554

ccgaucuaua ggacggcaga uucaacggga ugugccaaug cacucuuucc aggagugaac 60

accccguugg cuucaacaug aucgcccgcu caacggu 97

<210> 555

<211> 59

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 555

ugccaaugca cucuuuccag gagugaacac cccguuggcu ucaacaugau cgcccgcuc 59

<210> 556

<211> 139

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 556

gccgaucuau aggacggcag auucaacggg augugccaau gcacucuuuc caggagugaa 60

caccccguug gcuucaacau gaucgcccgc ucaacggucc cuagucggau cguugagcgg 120

gcgaucugag aaguggcac 139

<210> 557

<211> 128

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 557

gccgaucuau aggacggcag auucaacggg augugccaau gcacucuuuc caggagugaa 60

caccccguug gcuucaacau gaucgcccgc ucaacgcuac guugagcggg cgaucugaga 120

aguggcac 128

<210> 558

<211> 112

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 558

gccgaucuau aggacggcag auucaacggg augugccaau gcacucuuuc caggagugaa 60

caccccguug gcuucaacau gaucgccccu agggcgaucu gagaaguggc ac 112

<210> 559

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 559

augauacgag gcauuagcac 20

<210> 560

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 560

guccaagaaa aaagaaauga uacgaggcau uagcac 36

<210> 561

<211> 89

<212> RNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 561

acgagguucu gucuuuuggu caggacaacc gucuagcuau aagugcugca gggugugaga 60

aacuccuauu gcuggacgau gucucuuuu 89

<210> 562

<211> 99

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 562

gagguucugu cuuuugguca ggacaaccgu cuagcuauaa gugcugcagg gugugagaaa 60

cuccuauugc uggacgaugu cucuacgagg cauuagcac 99

<210> 563

<211> 106

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 563

acgagguucu gucuuuuggu caggacaacc gucuagcuau aagugcugca gggugugaga 60

aacuccuauu gcuggacgau gucucuauga uacgaggcau uagcac 106

<210> 564

<211> 102

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 564

acgagguucu gucuuuuggu caggacaacc gucuagcuau aagugcugca gggugugaga 60

aacuccuauu gcuggacgau gucucuuacg aggcauuagc ac 102

<210> 565

<211> 120

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (98)..(120)

<223> n = any nucleotide

<400> 565

gttctgtctt ttggtcagga caaccgtcta gctataagtg ctgcagggtg tgagaaactc 60

ctattgctgg acgacgcctc ttacgaggcg ttagcacnnn nnnnnnnnnn nnnnnnnnnn 120

<210> 566

<211> 1173

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 566

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala

1 5 10 15

Ala Ala Ile Arg Ser Ile Lys Leu Lys Met Lys Thr Asn Ser Gly Thr

20 25 30

Asp Ser Ile Tyr Leu Arg Lys Ala Leu Trp Arg Thr His Gln Leu Ile

35 40 45

Asn Glu Gly Ile Ala Tyr Tyr Met Asn Leu Leu Thr Leu Tyr Arg Gln

50 55 60

Glu Ala Ile Gly Asp Lys Thr Lys Glu Ala Tyr Gln Ala Glu Leu Ile

65 70 75 80

Asn Ile Ile Arg Asn Gln Gln Arg Asn Asn Gly Ser Ser Glu Glu His

85 90 95

Gly Ser Asp Gln Glu Ile Leu Ala Leu Leu Arg Gln Leu Tyr Glu Leu

100 105 110

Ile Ile Pro Ser Ser Ile Gly Glu Ser Gly Asp Ala Asn Gln Leu Gly

115 120 125

Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn Ser Gln Ser Gly Lys

130 135 140

Gly Thr Ser Asn Ala Gly Arg Lys Pro Arg Trp Lys Arg Leu Lys Glu

145 150 155 160

Glu Gly Asn Pro Asp Trp Glu Leu Glu Lys Lys Lys Asp Glu Glu Arg

165 170 175

Lys Ala Lys Asp Pro Thr Val Lys Ile Phe Asp Asn Leu Asn Lys Tyr

180 185 190

Gly Leu Leu Pro Leu Phe Pro Leu Phe Thr Asn Ile Gln Lys Asp Ile

195 200 205

Glu Trp Leu Pro Leu Gly Lys Arg Gln Ser Val Arg Lys Trp Asp Lys

210 215 220

Asp Met Phe Ile Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu Ser Trp

225 230 235 240

Asn Arg Arg Val Ala Asp Glu Tyr Lys Gln Leu Lys Glu Lys Thr Glu

245 250 255

Ser Tyr Tyr Lys Glu His Leu Thr Gly Gly Glu Glu Trp Ile Glu Lys

260 265 270

Ile Arg Lys Phe Glu Lys Glu Arg Asn Met Glu Leu Glu Lys Asn Ala
275 280 285
Phe Ala Pro Asn Asp Gly Tyr Phe Ile Thr Ser Arg Gln Ile Arg Gly
290 295 300
Trp Asp Arg Val Tyr Glu Lys Trp Ser Lys Leu Pro Glu Ser Ala Ser
305 310 315 320
Pro Glu Glu Leu Trp Lys Val Val Ala Glu Gln Gln Asn Lys Met Ser

325 330 335
Glu Gly Phe Gly Asp Pro Lys Val Phe Ser Phe Leu Ala Asn Arg Glu
340 345 350
Asn Arg Asp Ile Trp Arg Gly His Ser Glu Arg Ile Tyr His Ile Ala
355 360 365
Ala Tyr Asn Gly Leu Gln Lys Lys Leu Ser Arg Thr Lys Glu Gln Ala
370 375 380
Thr Phe Thr Leu Pro Asp Ala Ile Glu His Pro Leu Trp Ile Arg Tyr

385 390 395 400
Glu Ser Pro Gly Gly Thr Asn Leu Asn Leu Phe Lys Leu Glu Glu Lys
405 410 415
Gln Lys Lys Asn Tyr Tyr Val Thr Leu Ser Lys Ile Ile Trp Pro Ser
420 425 430
Glu Glu Lys Trp Ile Glu Lys Glu Asn Ile Glu Ile Pro Leu Ala Pro
435 440 445
Ser Ile Gln Phe Asn Arg Gln Ile Lys Leu Lys Gln His Val Lys Gly

450 455 460
Lys Gln Glu Ile Ser Phe Ser Asp Tyr Ser Ser Arg Ile Ser Leu Asp
465 470 475 480
Gly Val Leu Gly Gly Ser Arg Ile Gln Phe Asn Arg Lys Tyr Ile Lys
485 490 495
Asn His Lys Glu Leu Leu Gly Glu Gly Asp Ile Gly Pro Val Phe Phe
500 505 510
Asn Leu Val Val Asp Val Ala Pro Leu Gln Glu Thr Arg Asn Gly Arg

515 520 525

Leu Gln Ser Pro Ile Gly Lys Ala Leu Lys Val Ile Ser Ser Asp Phe

530 535 540

Ser Lys Val Ile Asp Tyr Lys Pro Lys Glu Leu Met Asp Trp Met Asn

545 550 555 560

Thr Gly Ser Ala Ser Asn Ser Phe Gly Val Ala Ser Leu Leu Glu Gly

565 570 575

Met Arg Val Met Ser Ile Asp Met Gly Gln Arg Thr Ser Ala Ser Val

580 585 590

Ser Ile Phe Glu Val Val Lys Glu Leu Pro Lys Asp Gln Glu Gln Lys

595 600 605

Leu Phe Tyr Ser Ile Asn Asp Thr Glu Leu Phe Ala Ile His Lys Arg

610 615 620

Ser Phe Leu Leu Asn Leu Pro Gly Glu Val Val Thr Lys Asn Asn Lys

625 630 635 640

Gln Gln Arg Gln Glu Arg Arg Lys Lys Arg Gln Phe Val Arg Ser Gln

645 650 655

Ile Arg Met Leu Ala Asn Val Leu Arg Leu Glu Thr Lys Lys Thr Pro

660 665 670

Asp Glu Arg Lys Lys Ala Ile His Lys Leu Met Glu Ile Val Gln Ser

675 680 685

Tyr Asp Ser Trp Thr Ala Ser Gln Lys Glu Val Trp Glu Lys Glu Leu

690 695 700

Asn Leu Leu Thr Asn Met Ala Ala Phe Asn Asp Glu Ile Trp Lys Glu

705 710 715 720

Ser Leu Val Glu Leu His His Arg Ile Glu Pro Tyr Val Gly Gln Ile

725 730 735

Val Ser Lys Trp Arg Lys Gly Leu Ser Glu Gly Arg Lys Asn Leu Ala

740 745 750

Gly Ile Ser Met Trp Asn Ile Asp Glu Leu Glu Asp Thr Arg Arg Leu

755 760 765

Leu Ile Ser Trp Ser Lys Arg Ser Arg Thr Pro Gly Glu Ala Asn Arg

770 775 780

Ile Glu Thr Asp Glu Pro Phe Gly Ser Ser Leu Leu Gln His Ile Gln

785 790 795 800

Asn Val Lys Asp Asp Arg Leu Lys Gln Met Ala Asn Leu Ile Ile Met

805 810 815

Thr Ala Leu Gly Phe Lys Tyr Asp Lys Glu Glu Lys Asp Arg Tyr Lys

820 825 830

Arg Trp Lys Glu Thr Tyr Pro Ala Cys Gln Ile Ile Leu Phe Glu Asn

835 840 845

Leu Asn Arg Tyr Leu Phe Asn Leu Asp Arg Ser Arg Arg Glu Asn Ser

850 855 860

Arg Leu Met Lys Trp Ala His Arg Ser Ile Pro Arg Thr Val Ser Met

865 870 875 880

Gln Gly Glu Met Phe Gly Leu Gln Val Gly Asp Val Arg Ser Glu Tyr

885 890 895

Ser Ser Arg Phe His Ala Lys Thr Gly Ala Pro Gly Ile Arg Cys His

900 905 910

Ala Leu Thr Glu Glu Asp Leu Lys Ala Gly Ser Asn Thr Leu Lys Arg

915 920 925

Leu Ile Glu Asp Gly Phe Ile Asn Glu Ser Glu Leu Ala Tyr Leu Lys

930 935 940

Lys Gly Asp Ile Ile Pro Ser Gln Gly Gly Glu Leu Phe Val Thr Leu

945 950 955 960

Ser Lys Arg Tyr Lys Lys Asp Ser Asp Asn Asn Glu Leu Thr Val Ile

965 970 975

His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Gln

980 985 990

Gln Asn Ser Glu Val Tyr Arg Val Pro Cys Gln Leu Ala Arg Met Gly

995 1000 1005

Glu Asp Lys Leu Tyr Ile Pro Lys Ser Gln Thr Glu Thr Ile Lys

1010 1015 1020

Lys Tyr Phe Gly Lys Gly Ser Phe Val Lys Asn Asn Thr Glu Gln

1025 1030 1035

Glu Val Tyr Lys Trp Glu Lys Ser Glu Lys Met Lys Ile Lys Thr

1040 1045 1050

Asp Thr Thr Phe Asp Leu Gln Asp Leu Asp Gly Phe Glu Asp Ile

1055 1060 1065

Ser Lys Thr Ile Glu Leu Ala Gln Glu Gln Gln Lys Lys Tyr Leu

1070 1075 1080

Thr Met Phe Arg Asp Pro Ser Gly Tyr Phe Phe Asn Asn Glu Thr

1085 1090 1095

Trp Arg Pro Gln Lys Glu Tyr Trp Ser Ile Val Asn Asn Ile Ile

1100 1105 1110

Lys Ser Cys Leu Lys Lys Lys Ile Leu Ser Asn Lys Val Glu Leu

1115 1120 1125

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys

1130 1135 1140

Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr

1145 1150 1155

Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1160 1165 1170

<210> 567

<211> 127

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (105)..(127)

<223> n = any nucleotide

<400> 567

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggauuaucuu auuucugcaa aagcagaaau aagaugauug gcacnnnnnn nnnnnnnnnn 120

nnnnnnn 127

<210> 568

<211> 130

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (108)..(130)

<223> n = any nucleotide

<400> 568

ggugaccuau agggucaaug aaucugugcg ugugccauaa guaauuaaaa auuacccacc 60

acaggauuau cuuauuucug caaaagcaga aauaagauga uuggcacnnn nnnnnnnnnn 120

nnnnnnnnnn 130

<210> 569

<211> 124

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (102)..(124)

<223> n = any nucleotide

<400> 569

cuauaggguc aaugaaucug ugcgugugcc auaaguaauu aaaaauuacc caccacagga 60

uuaucuuauu ucugcaaaag cagaaauaag augauuggca cnnnnnnnnn nnnnnnnnnn 120

nnnn 124

<210> 570

<211> 111

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (89)..(111)

<223> n = any nucleotide

<400> 570

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggaucaucuu aaaaaagaug auuggcacnn nnnnnnnnnn nnnnnnnnnn n 111

<210> 571

<211> 121

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (99)..(121)

<223> n = any nucleotide

<400> 571

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggaucaucuu auuucaaaag aaauaagaug auuggcacnn nnnnnnnnnn nnnnnnnnnn 120

n 121

<210> 572

<211> 111

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (89)..(111)

<223> n = any nucleotide

<400> 572

gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggagcaccug aaaacaggug cuuggcacnn nnnnnnnnnn nnnnnnnnnn n 111

<210> 573

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 573

cccttcagct aaaataaagg agg 23

<210> 574

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 574

ggctcagcag gcacctgcct cag 23

<210> 575

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 575

acgtactgat gttaacagct gac 23

<210> 576

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 576

gggactggag ttgcttcatg tac 23

<210> 577

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 577

ctgacctccc aaacagctac ata 23

<210> 578

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 578

ttcatggaga aaatattcag aat 23

<210> 579

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 579

tctccatgaa aaatactggg gtc 23

<210> 580

<211> 9041

<212> DNA

<213> Bacillus species

<400> 580

agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 60

cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga 120

ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 180

tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 240

gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc 300

gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 360

gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 420

ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt 480

ggcctaacta cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag 540

ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 600

gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 660

ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 720

tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 780

ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 840

gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg 900

tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac 960

cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1020

ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc 1080
gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta 1140
caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1200
gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc 1260
ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac 1320
tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 1380
caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 1440

tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 1500
cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca 1560
ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa 1620
aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 1680
tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 1740
gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc 1800
gaaaagtgcc acctgacgtc gacggatcgg gagatctccc gatcccctat ggtgcactct 1860

cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt 1920
ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag gcttgaccga 1980
caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc 2040
cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa ttacggggtc 2100
attagttcat agcccatata tggagttccg cgttacataa cttacggtaa atggcccgcc 2160
tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt 2220
aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt aaactgccca 2280

cttggcagta catcaagtgt atcatatgcc aagtacgccc cctattgacg tcaatgacgg 2340
taaatggccc gcctggcatt atgcccagta catgacctta tgggactttc ctacttggca 2400
gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc agtacatcaa 2460
tgggcgtgga tagcggtttg actcacgggg atttccaagt ctccacccca ttgacgtcaa 2520
tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta acaactccgc 2580
cccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa gcagagctct 2640
ctggctaact agagaaccca ctgcttactg gcttatcgaa attaatacga ctcactatag 2700

ggagacccaa gctggctagc gtttaaactt aagcttgcca ccatggcccc aaagaagaag 2760

cggaaggtcg gtatccacgg agtcccagca gccggatccg ccgtgaagtc catcaaagtg 2820

aagctgcggc tgagcgagtg ccccgatatt ctggctggaa tgtggcagct gcacagagcc 2880

acaaatgccg gcgtgcggta ctacacagaa tgggtgtccc tgatgcggca agagatcctg 2940

tacagcagag gccctgatgg cggccagcag tgttatatga ccgccgagga ttgccagaga 3000

gagctgctgc ggagactgcg gaatagacag ctgcataacg gccggcagga tcagcctgga 3060

acagatgctg atctgctggc catcagcaga cggctgtacg agattctggt gctgcagagc 3120

atcggcaaaa gaggcgacgc ccagcagatt gccagcagct ttctgagccc tctggtggac 3180

cccaacagca aaggtggaag aggcgaggcc aagagcggaa gaaaacctgc ctggcagaag 3240

atgcgcgacc agggcgatcc tagatgggtt gccgctagag agaagtacga gcagcggaag 3300

gccgtggatc ccagcaaaga gattctgaac agcctggacg ccctgggcct cagacctctg 3360

tttgccgtgt tcaccgagac atacagatcc ggcgtggact ggaagcctct gggcaaatct 3420

cagggcgtca gaacctggga cagagacatg tttcagcagg ccctggaacg gctgatgagc 3480

tgggagagct ggaatcggag agtgggcgaa gagtacgcca gactgttcca gcagaaaatg 3540

aagttcgagc aagagcactt cgccgagcag agccacctgg tcaaactggc tagagccctg 3600

gaagccgata tgagagccgc ctctcagggc ttcgaggcca aaagaggaac agcccaccag 3660

atcaccagaa gggcactgag aggggccgac agagtgttcg agatctggaa gtctatcccc 3720

gaggaagccc tgttcagcca gtacgacgaa gtgatcagac aggtgcaggc cgagaagcgg 3780

agagatttcg gcagccatga cctgttcgcc aagctggccg agcctaagta tcagcccctt 3840

tggagagccg acgagacatt cctgaccaga tacgccctgt acaacggcgt gctgcgcgat 3900

ctggaaaagg ccagacagtt cgccaccttc acactgcctg atgcctgcgt gaaccccatc 3960

tggaccagat tcgagtctag ccagggcagc aacctgcaca aatacgagtt tctgttcgac 4020

cacctcggac ctggcagaca cgccgtcaga tttcagagac tgctggtggt ggaaagcgag 4080

ggcgccaaag aaagggatag cgtggtggtg cctgtggctc cttctggcca actggataag 4140

ctggtgctga gggaagaaga gaagtccagc gtcgccctgc atctgcacga taccgctaga 4200

cccgatggct tcatggctga atgggctggc gccaaactgc agtacgagag aagcaccctg 4260

gccagaaaag ccagacggga caagcagggc atgagaagct ggcggagaca gccctccatg 4320

ctgatgtctg ccgctcagat gctggaagat gccaaacagg ctggcgacgt gtacctgaac 4380

atcagcgtgc gcgtgaagtc tcccagtgaa gtgcgaggac agaggcggcc tccttacgcc 4440

gctctgttta gaatcgacga caagcagcgg agagtgaccg tgaactacaa caagctgagc 4500

gcctacctgg aagaacaccc cgataagcag atccctggcg ctcctggact gctgtctgga 4560

ctgagagtga tgtccgtgga cctgggcctg agaacaagcg ccagcatctc cgtgttcaga 4620

gtggccaaga aagaagaggt ggaagccctc ggagatggcc ggcctcctca ctactatcct 4680

atccacggca ccgatgacct ggtggccgtg cacgaaagat cccacctgat tcagatgccc 4740

ggcgaaaccg agacaaagca gctgcggaag ctgagagaag aacggcaggc cgtgctgagg 4800

ccactgtttg ctcaactggc actgctgaga ctgctcgtca gatgtggcgc cgctgacgag 4860

agaatcagaa ccagatcctg gcagcggctg accaagcagg gaagagagtt caccaagaga 4920

ctgaccccta gctggcgcga ggctctggaa ctggaactga caagactcga ggcctactgc 4980

ggcagagtgc ccgatgatga gtggtccaga atcgtggaca gaaccgtgat tgccctgtgg 5040

cggagaatgg gcaagcaagt gcgcgattgg cggaagcaag tgaagtccgg ggccaaagtg 5100

aaagtgaagg gctaccagct ggatgtcgtc ggcggaaatt ctctggccca gatcgactat 5160

ctggaacagc agtacaagtt cctgcggcgt tggagcttct tcgccagagc ttctggcctg 5220

gtcgtgcggg ccgatagaga aagccatttt gccgtggctc tgagacagca catcgagaac 5280

gccaagcggg acagactgaa gaaactggcc gaccggatcc tgatggaagc actgggctat 5340

gtgtacgagg ccagcggacc tagagaaggc cagtggacag ctcagcaccc tccttgccag 5400

ctgatcattc tcgaggaact gtccgcctac cggttcagcg acgatagacc tcctagcgag 5460

aacagcaaac tgatggcctg gggccacaga ggcatcctcg aagaactggt caaccaggct 5520

caggtgcacg atgtgctcgt gggcacagtg tacgccgcct tcagcagcag attcgacgct 5580

agaacaggtg ctcccggcgt cagatgcaga agagtgcctg ccagatttgt gggcgccacc 5640

gtggatgatt ctctgccact gtggctgacc gagttcctgg acaagcaccg gctggataag 5700

aacctgctgc ggcccgacga tgtgatccca acaggcgaag gcgaattcct ggtgtcccct 5760

tgtggcgaag aggctgccag agttagacag gttcacgccg acatcaacgc tgcccagaac 5820

ctgcagagaa ggctgtggca gaacttcgac atcaccgagc tgaggctgag atgcgacgtg 5880

aagatgggcg gagagggaac agtgctggtg cccagagtga acaacgccag agccaagcag 5940

ctgttcggca agaaggtgct ggtttcccag gacggcgtga ccttcttcga gagatctcag 6000

acaggcggca agccccacag cgagaagcag accgatctga ccgacaaaga actcgagctg 6060

atcgccgagg ccgatgaggc cagagctaaa agcgtggtgc tgttcaggga tcctagcggc 6120

cacattggca aaggccactg gatccggcag cgcgagtttt ggagtctggt caagcagagg 6180

atcgagagcc acaccgccga gcggattaga gttagaggcg tgggaagctc cctggacgga 6240

tccaaaaggc cggcggccac gaaaaaggcc ggccaggcaa aaaagaaaaa gggatcttac 6300

ccatacgatg ttccagatta cgcttatccc tacgacgtgc ctgattatgc atacccatat 6360

gatgtccccg actatgccta agaattctgc agatatccag cacagtggcg gccgctcgag 6420

tctagagggc ccgtttaaac ccgctgatca gcctcgactg tgccttctag ttgccagcca 6480

tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac tcccactgtc 6540
ctttcctaat aaaatgagga aattgcatcg cattgtctga gtaggtgtca ttctattctg 6600
gggggtgggg tggggcagga cagcaagggg gaggattggg aagacaatag caggcatgct 6660
ggggatgcgg tgggctctat ggcttctgag gcggaaagaa ccagctgggg ctctaggggg 6720
tatccccacg cgccctgtag cggcgcatta agcgcggcgg gtgtggtggt tacgcgcagc 6780
gtgaccgcta cacttgccag cgccctagcg cccgctcctt tcgctttctt cccttccttt 6840
ctcgccacgt tcgccggctt tccccgtcaa gctctaaatc gggggctccc tttagggttc 6900

cgatttagtg ctttacggca cctcgacccc aaaaaacttg attagggtga tggttcacgt 6960
agtgggccat cgccctgata gacggttttt cgcccctttga cgttggagtc cacgttcttt 7020
aatagtggac tcttgttcca aactggaaca acactcaacc ctatctcggt ctattctttt 7080
gatttataag ggattttgcc gatttcggcc tattggttaa aaaatgagct gatttaacaa 7140
aaatttaacg cgaattaatt ctgtggaatg tgtgtcagtt agggtgtgga aagtccccag 7200
gctccccagc aggcagaagt atgcaaagca tgcatctcaa ttagtcagca accaggtgtg 7260
gaaagtcccc aggctcccca gcaggcagaa gtatgcaaag catgcatctc aattagtcag 7320

caaccatagt cccgccccta actccgccca tcccgcccct aactccgccc agttccgccc 7380
attctccgcc ccatggctga ctaatttttt ttatttatgc agaggccgag gccgcctctg 7440
cctctgagct attccagaag tagtgaggag gctttttttgg aggcctaggc ttttgcaaaa 7500
agctcccggg agcttgtata tccattttcg gatctgatca agagacagga tgaggatcgt 7560
ttcgcatgat tgaacaagat ggattgcacg caggttctcc ggccgcttgg gtggagaggc 7620
tattcggcta tgactgggca caacagacaa tcggctgctc tgatgccgcc gtgttccggc 7680
tgtcagcgca ggggcgcccg gttctttttg tcaagaccga cctgtccggt gccctgaatg 7740

aactgcagga cgaggcagcg cggctatcgt ggctggccac gacgggcgtt ccttgcgcag 7800
ctgtgctcga cgttgtcact gaagcgggaa gggactggct gctattgggc gaagtgccgg 7860
ggcaggatct cctgtcatct caccttgctc ctgccgagaa agtatccatc atggctgatg 7920
caatgcggcg gctgcatacg cttgatccgg ctacctgccc attcgaccac caagcgaaac 7980
atcgcatcga gcgagcacgt actcggatgg aagccggtct tgtcgatcag gatgatctgg 8040
acgaagagca tcaggggctc gcgccagccg aactgttcgc caggctcaag gcgcgcatgc 8100
ccgacggcga ggatctcgtc gtgacccatg gcgatgcctg cttgccgaat atcatggtgg 8160

aaaatggccg cttttctgga ttcatcgact gtggccggct gggtgtggcg gaccgctatc 8220

aggacatagc gttggctacc cgtgatattg ctgaagagct tggcggcgaa tgggctgacc 8280

gcttcctcgt gctttacggt atcgccgctc ccgattcgca gcgcatcgcc ttctatcgcc 8340

ttcttgacga gttcttctga gcgggactct ggggttcgaa atgaccgacc aagcgacgcc 8400

caacctgcca tcacgagatt tcgattccac cgccgccttc tatgaaaggt tgggcttcgg 8460

aatcgttttc cgggacgccg gctggatgat cctccagcgc ggggatctca tgctggagtt 8520

cttcgcccac cccaacttgt ttattgcagc ttataatggt tacaaataaa gcaatagcat 8580

cacaaatttc acaaataaag catttttttc actgcattct agttgtggtt tgtccaaact 8640

catcaatgta tcttatcatg tctgtatacc gtcgacctct agctagagct tggcgtaatc 8700

atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg 8760

agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac tcacattaat 8820

tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg 8880

aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg cttcctcgct 8940

cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc 9000

ggtaatacgg ttatccacag aatcagggga taacgcagga a 9041

<210> 581

<211> 8924

<212> DNA

<213> Bacillus hisashii

<400> 581

agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 60

cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga 120

ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 180

tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 240

gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc 300

gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 360

gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 420

ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt 480

ggcctaacta cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag 540

ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 600

gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 660

ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 720

tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 780

ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 840

gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg 900

tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac 960

cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1020

ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc 1080

gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta 1140

caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1200

gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc 1260

ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac 1320

tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 1380

caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 1440

tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 1500

cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca 1560

ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa 1620

aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 1680

tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 1740

gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc 1800

gaaaagtgcc acctgacgtc gacggatcgg gagatctccc gatcccctat ggtgcactct 1860

cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt 1920

ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag gcttgaccga 1980

caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc 2040

cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa ttacggggtc 2100

attagttcat agcccatata tggagttccg cgttacataa cttacggtaa atggcccgcc 2160

tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt 2220

aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt aaactgccca 2280

cttggcagta catcaagtgt atcatatgcc aagtacgccc cctattgacg tcaatgacgg 2340

taaatggccc gcctggcatt atgcccagta catgacctta tgggactttc ctacttggca 2400

gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc agtacatcaa 2460

tgggcgtgga tagcggtttg actcacgggg atttccaagt ctccacccca ttgacgtcaa 2520

tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta acaactccgc 2580

cccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa gcagagctct 2640

ctggctaact agagaaccca ctgcttactg gcttatcgaa attaatacga ctcactatag 2700

ggagacccaa gctggctagc gtttaaactt aagcttgcca ccatggcccc aaagaagaag 2760

cggaaggtcg gtatccacgg agtcccagca gccggatccg ccaccagatc cttcatcctg 2820

aagatcgagc ccaacgagga agtgaagaaa ggcctctgga aaacccacga ggtgctgaac 2880

cacggaatcg cctactacat gaatatcctg aagctgatcc ggcaagaggc catctacgag 2940

caccacgagc aggaccccaa gaatcccaag aaggtgtcca aggccgagat ccaggccgag 3000

ctgtgggatt tcgtgctgaa gatgcagaag tgcaacagct tcacacacga ggtggacaag 3060

gacgaggtgt tcaacatcct gagagagctg tacgaggaac tggtgcccag cagcgtggaa 3120

aagaagggcg aagccaacca gctgagcaac aagtttctgt accctctggt ggaccccaac 3180

agccagtctg gaaagggaac agccagcagc ggcagaaagc ccagatggta caacctgaag 3240

attgccggcg atccctcctg ggaagaagag aagaagaagt gggaagaaga taagaaaaag 3300

gacccgctgg ccaagatcct gggcaagctg gctgagtacg gactgatccc tctgttcatc 3360

ccctacaccg acagcaacga gcccatcgtg aaagaaatca agtggatgga aaagtcccgg 3420

aaccagagcg tgcggcggct ggataaggac atgttcattc aggccctgga acggttcctg 3480

agctgggaga gctggaacct gaaagtgaaa gaggaatacg agaaggtcga gaaagagtac 3540

aagaccctgg aagagaggat caaagaggac atccaggctc tgaaggctct ggaacagtat 3600

gagaaagagc ggcaagaaca gctgctgcgg gacaccctga acaccaacga gtaccggctg 3660

agcaagagag gccttagagg ctggcgggaa atcatccaga aatggctgaa aatggacgag 3720

aacgagccct ccgagaagta cctggaagtg ttcaaggact accagcggaa gcaccctaga 3780

gaggccggcg attacagcgt gtacgagttc ctgtccaaga aagagaacca cttcatctgg 3840

cggaatcacc ctgagtaccc ctacctgtac gccaccttct gcgagatcga caagaaaaag 3900

aaggacgcca agcagcaggc caccttcaca ctggccgatc ctatcaatca ccctctgtgg 3960

gtccgattcg aggaaagaag cggcagcaac ctgaacaagt acagaatcct gaccgagcag 4020

ctgcacaccg agaagctgaa gaaaaagctg acagtgcagc tggaccggct gatctaccct 4080

acagaatctg gcggctggga agagaagggc aaagtggaca ttgtgctgct gcccagccgg 4140

cagttctaca accagatctt cctggacatc gaggaaaagg gcaagcacgc cttcacctac 4200

aaggatgaga gcatcaagtt ccctctgaag ggcacactcg gcggagccag agtgcagttc 4260

gacagagatc acctgagaag ataccctcac aaggtggaaa gcggcaacgt gggcagaatc 4320

tacttcaaca tgaccgtgaa catcgagcct acagagtccc cagtgtccaa gtctctgaag 4380

atccaccggg acgacttccc caaggtggtc aacttcaagc ccaaagaact gaccgagtgg 4440

atcaaggaca gcaagggcaa gaaactgaag tccggcatcg agtccctgga aatcggcctg 4500

agagtgatga gcatcgacct gggacagaga caggccgctg ccgcctctat tttcgaggtg 4560

gtggatcaga agcccgacat cgaaggcaag ctgtttttcc caatcaaggg caccgagctg 4620

tatgccgtgc acagagccag cttcaacatc aagctgcccg gcgagacact ggtcaagagc 4680

agagaagtgc tgcggaaggc cagagaggac aatctgaaac tgatgaacca gaagctcaac 4740

ttcctgcgga acgtgctgca cttccagcag ttcgaggaca tcaccgagag agagaagcgg 4800

gtcaccaagt ggatcagcag acaagagaac agcgacgtgc ccctggtgta ccaggatgag 4860

ctgatccaga tccgcgagct gatgtacaag ccttacaagg actgggtcgc cttcctgaag 4920

cagctccaca agagactgga agtcgagatc ggcaaagaag tgaagcactg gcggaagtcc 4980

ctgagcgacg gaagaaaggg cctgtacggc atctccctga agaacatcga cgagatcgat 5040

cggacccgga agttcctgct gagatggtcc ctgaggccta ccgaacctgg cgaagtgcgt 5100

agactggaac ccggccagag attcgccatc gaccagctga atcacctgaa cgccctgaaa 5160

gaagatcggc tgaagaagat ggccaacacc atcatcatgc acgccctggg ctactgctac 5220

gacgtgcgga agaagaaatg gcaggctaag aaccccgcct gccagatcat cctgttcgag 5280

gatctgagca actacaaccc ctacgaggaa aggtcccgct tcgagaacag caagctcatg 5340

aagtggtcca gacgcgagat ccccagacag gttgcactgc agggcgagat ctatggcctg 5400

caagtgggag aagtgggcgc tcagttcagc agcagattcc acgccaagac aggcagccct 5460

ggcatcagat gtagcgtcgt gaccaaagag aagctgcagg acaatcggtt cttcaagaat 5520

ctgcagagag agggcagact gaccctggac aaaatcgccg tgctgaaaga gggcgatctg 5580

tacccagaca aaggcggcga gaagttcatc agcctgagca aggatcggaa gtgcgtgacc 5640

acacacgccg acatcaacgc cgctcagaac ctgcagaagc ggttctggac aagaacccac 5700

ggcttctaca aggtgtactg caaggcctac caggtggacg gccagaccgt gtacatccct 5760

gagagcaagg accagaagca gaagatcatc gaagagttcg gcgagggcta cttcattctg 5820

aaggacgggg tgtacgaatg ggtcaacgcc ggcaagctga aaatcaagaa gggcagctcc 5880

aagcagagca gcagcgagct ggtggatagc gacatcctga aagacagctt cgacctggcc 5940

tccgagctga aaggcgaaaa gctgatgctg tacagggacc ccagcggcaa tgtgttcccc 6000

agcgacaaat ggatggccgc tggcgtgttc ttcggaaagc tggaacgcat cctgatcagc 6060

aagctgacca accagtactc catcagcacc atcgaggacg acagcagcaa gcagtctatg 6120

ggatccaaaa ggccggcggc cacgaaaaag gccggccagg caaaaaagaa aaagggatct 6180

tacccatacg atgttccaga ttacgcttat ccctacgacg tgcctgatta tgcataccca 6240
tatgatgtcc ccgactatgc ctaagaattc tgcagatatc cagcacagtg gcggccgctc 6300
gagtctagag ggcccgttta aacccgctga tcagcctcga ctgtgccttc tagttgccag 6360
ccatctgttg tttgcccctc ccccgtgcct tccttgaccc tggaaggtgc cactcccact 6420
gtcctttcct aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt 6480
ctgggggggtg gggtggggca ggacagcaag ggggaggatt gggaagacaa tagcaggcat 6540
gctggggatg cggtgggctc tatggcttct gaggcggaaa gaaccagctg gggctctagg 6600

gggtatcccc acgcgccctg tagcggcgca ttaagcgcgg cgggtgtggt ggttacgcgc 6660
agcgtgaccg ctacacttgc cagcgcccta gcgcccgctc ctttcgcttt cttcccttcc 6720
tttctcgcca cgttcgccgg ctttccccgt caagctctaa atcgggggct ccctttaggg 6780
ttccgattta gtgctttacg gcacctcgac cccaaaaaac ttgattaggg tgatggttca 6840
cgtagtgggc catcgccctg atagacggtt tttcgccctt tgacgttgga gtccacgttc 6900
tttaatagtg gactcttgtt ccaaactgga acaacactca accctatctc ggtctattct 6960
tttgatttat aagggatttt gccgatttcg gcctattggt taaaaaatga gctgatttaa 7020

caaaaattta acgcgaatta attctgtgga atgtgtgtca gttagggtgt ggaaagtccc 7080
caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccaggt 7140
gtggaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 7200
cagcaaccat agtcccgccc ctaactccgc ccatcccgcc cctaactccg cccagttccg 7260
cccattctcc gccccatggc tgactaattt tttttattta tgcagaggcc gaggccgcct 7320
ctgcctctga gctattccag aagtagtgag gaggcttttt tggaggccta ggcttttgca 7380
aaaagctccc gggagcttgt atatccattt tcggatctga tcaagagaca ggatgaggat 7440

cgtttcgcat gattgaacaa gatggattgc acgcaggttc tccggccgct tgggtggaga 7500
ggctattcgg ctatgactgg gcacaacaga caatcggctg ctctgatgcc gccgtgttcc 7560
ggctgtcagc gcaggggcgc ccggttcttt ttgtcaagac cgacctgtcc ggtgccctga 7620
atgaactgca ggacgaggca gcgcggctat cgtggctggc cacgacgggc gttccttgcg 7680
cagctgtgct cgacgttgtc actgaagcgg gaagggactg gctgctattg ggcgaagtgc 7740
cggggcagga tctcctgtca tctcaccttg ctcctgccga gaaagtatcc atcatggctg 7800
atgcaatgcg gcggctgcat acgcttgatc cggctacctg cccattcgac caccaagcga 7860

aacatcgcat cgagcgagca cgtactcgga tggaagccgg tcttgtcgat caggatgatc 7920

tggacgaaga gcatcagggg ctcgcgccag ccgaactgtt cgccaggctc aaggcgcgca 7980

tgcccgacgg cgaggatctc gtcgtgaccc atggcgatgc ctgcttgccg aatatcatgg 8040

tggaaaatgg ccgcttttct ggattcatcg actgtggccg gctgggtgtg gcggaccgct 8100

atcaggacat agcgttggct acccgtgata ttgctgaaga gcttggcggc gaatgggctg 8160

accgcttcct cgtgctttac ggtatcgccg ctcccgattc gcagcgcatc gccttctatc 8220

gccttcttga cgagttcttc tgagcgggac tctggggttc gaaatgaccg accaagcgac 8280

gcccaacctg ccatcacgag atttcgattc caccgccgcc ttctatgaaa ggttgggctt 8340

cggaatcgtt ttccgggacg ccggctggat gatcctccag cgcggggatc tcatgctgga 8400

gttcttcgcc cacccccaact tgtttattgc agcttataat ggttacaaat aaagcaatag 8460

catcacaaat ttcacaaata aagcattttt ttcactgcat tctagttgtg gtttgtccaa 8520

actcatcaat gtatcttatc atgtctgtat accgtcgacc tctagctaga gcttggcgta 8580

atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat 8640

acgagccgga agcataaagt gtaaagcctg gggtgcctaa tgagtgagct aactcacatt 8700

aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta 8760

atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc 8820

gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa 8880

ggcggtaata cggttatcca cagaatcagg ggataacgca ggaa 8924

<210> 582

<211> 9692

<212> DNA

<213> Unknown

<220><223> Member of the phylum Elusimicrobia

<400> 582

agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 60

cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga 120

ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 180

tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 240

gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc 300

gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 360

gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 420

ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt 480

ggcctaacta cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag 540

ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 600
gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 660
ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 720
tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 780
ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 840
gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg 900
tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac 960

cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1020
ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc 1080
gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta 1140
caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1200
gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc 1260
ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac 1320
tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 1380

caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 1440
tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 1500
cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca 1560
ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa 1620
aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 1680
tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 1740
gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc 1800

gaaaagtgcc acctgacgtc gacggatcgg gagatctccc gatcccctat ggtgcactct 1860
cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt 1920
ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag gcttgaccga 1980
caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc 2040
cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa ttacggggtc 2100
attagttcat agcccatata tggagttccg cgttacataa cttacggtaa atggcccgcc 2160
tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt 2220

aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt aaactgccca 2280

cttggcagta catcaagtgt atcatatgcc aagtacgccc cctattgacg tcaatgacgg 2340

taaatggccc gcctggcatt atgcccagta catgacctta tgggactttc ctacttggca 2400

gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc agtacatcaa 2460

tgggcgtgga tagcggtttg actcacgggg atttccaagt ctccacccca ttgacgtcaa 2520

tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta acaactccgc 2580

cccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa gcagagctct 2640

ctggctaact agagaaccca ctgcttactg gcttatcgaa attaatacga ctcactatag 2700

ggagacccaa gctggctagc gtttaaactt aagcttgcca ccatggcccc aaagaagaag 2760

cggaaggtcg gtatccacgg agtcccagca gccggatcca accggatcta ccagggcaga 2820

gtgaccaagg tggaagtgcc cgatggcaag gacgagaagg gcaacatcaa gtggaagaag 2880

ctggaaaatt ggagcgacat cctgtggcag caccacatgc tgttccagga cgccgtgaac 2940

tactacacac tggccctggc cgccatctct ggatctgctg ttggcagcga cgagaagtcc 3000

atcatcctga gagaatgggc cgtgcaggtc cagaacatct gggagaaagc caagaaaaag 3060

gccaccgtgt tcgagggccc acagaagaga ctgaccagca tcctgggcct tgagcagaac 3120

gccagcttcg acattgccgc caagcacatc ctgaggacct ctgaggccaa gcctgagcag 3180

agagctagcg ccctgatcag actgctggaa gagatcgaca agaaaaacca caacgtcgtg 3240

tgcggcgagc ggctgccttt tttctgccct cggaacatcc agagcaagcg gagccctaca 3300

agcaaggccg tgtctagcgt gcaagagcag aaacggcaag aggaagtgcg gcggttccac 3360

aacatgcagc ctgaggaagt ggtcaagaac gccgtgacac tggacatcag cctgttcaag 3420

agcagcccca agatcgtgtt cctggaagat cccaagaagg ccagagccga gctgctgaag 3480

cagttcgaca acgcctgcaa gaaacacaaa gaactcgtgg gcatcaagaa agccttcacc 3540

gagtccatcg acaagcacgg ctctagcctg aaggtgccag ctcctggctc taagcctagc 3600

ggcctgtatc ctagcgccat cgtgttcaag tacttccccg tggatattac caagaccgtg 3660

tttctgaagg ccacagagaa gctggccatg ggcaaagacc gggaagtgac caacgatcct 3720

atcgccgacg ccagagtgaa cgacaagccc cacttcgact acttcaccaa cattgccctg 3780

atccgcgaga aagagaagaa cagagccgct tggtttgagt tcgatctggc cgcctttatc 3840

gaggccatca tgagccctca cagattctac caggacaccc agaagcggaa agaggccgcc 3900

agaaagctgg aagaaaagat caaggccatc gaaggcaaag gcgggcagtt caaagagagc 3960

gacagcgagg acgacgacgt ggactctctg cctggatttg agggcgacac cagaatcgac 4020

ctgctgcgga agctggtcac cgatacactt ggatggctgg gcgagagcga gacacccgat 4080

aacaacgagg gcaaaaagac cgagtacagc atcagcgagc ggaccctgag aatcttcccc 4140

gacatccaga agcagtggag cgagctggcc gagaaaggcg agacaacaga gggaaagctg 4200

ctcgaagtgc tgaaacacga gcagaccgag caccagagcg atttcggaag cgccacactg 4260

tatcagcacc tggccaagcc agagtttcac cccatctggc tgaagtccgg caccgaggaa 4320

tggcacgccg agaatcctct gaaagcctgg ctgaactaca aagagctgca gtacgagctg 4380

accgacaaga agcggcccat ccactttacc cctgctcacc ctgtgtacag ccccagatac 4440

ttcgacttcc ccaagaagtc cgaaaccgaa gagaaagagg tgtccaagaa cacccacagc 4500

ctgaccacaa gcctggccag cgagcacatc aagaactccc tgcagtttac agccgggctg 4560

atcagaaaga ccaacgtggg caagaaggct atcaaggccc ggttcagcta cagcgccccc 4620

agactgagaa gggactgcct gagaagcgag aacaacgaga acctgtacaa ggccccttgg 4680

ctccagccta tgatgagagc cctgggcatc gacgaggaaa aggccgacag acagaacttc 4740

gccaacacca ggatcaccct gatggccaaa ggcctggacg acattcagct gggctttccc 4800

gtggaagcca acagccaaga actgcagaaa gaagtgtcta acggcatcag ctggaagggc 4860

cagttcaact ggggaggaat cgcctctctg tctgccctga gatggcccca cgagaagaag 4920

cccaagaatc ctcctgagca gccttggtgg ggcatcgata gctttagctg cctggccgtg 4980

gatctgggcc agagatatgc tggcgccttc gccagactgg acgtgtccac cattgagaaa 5040

aagggcaaga gccggttcat cggcgaggcc tgcgacaaaa agtggtacgc caaggtgtcc 5100

cggatgggcc tgttgagact tcctggcgag gacgtgaaag tgtggcggga tgccagcaag 5160

attgacaaag agaacggctt cgccttccgg aaagagctgt tcggcgagaa gggaagatcc 5220

gccacacctc tggaagccga ggaaaccgcc gagctgatca agctgtttgg agccaacgag 5280

aaggacgtga tgcccgacaa ctggtctaaa gagctgagct tccccgagca gaatgacaag 5340

ctgctgatcg tggctcggag agcccaggct gctgttagca gactgcatag atgggcatgg 5400

ttcttcgacg aggccaagag atccgacgac gccatcagag agattctgga aagcgacgac 5460

accgacctga agcagaaagt gaacaagaac gagatcgaga aagtcaaaga gacaatcatc 5520

tccctgctga aagtcaagca agagctgctg cccacactgc tgaccagact ggccaataga 5580

gtgctgcccc tgagaggcag atcctgggag tggaaaaagc accaccagaa gaacgacggc 5640

ttcatcctgg accagaccgg caaggccatg cctaacgtgc tgattagagg acagcggggc 5700

ctgagcatgg accggatcga gcagattacc gagctgagaa agcggtttca ggccctgaac 5760

cagagcctgc ggagacagat cggaaagaag gcccctgcca agcgggacga ctctatccct 5820

gattgctgcc ccgatctgct ggaaaaactg gaccacatga aggaacagcg cgtgaaccag 5880

acagcccaca tgattctggc cgaggcactg ggactgaaac tggccgagcc tcctaaggac 5940

aagaaagaac tgaacgagac atgcgacatg cacggcgcct acgccaaagt ggacaacccc 6000

gtgtccttca tcgtgatcga ggacctgagc cggtacagaa gcagccaagg cagaagcccc 6060
agagaaaaca gccgactgat gaagtggtgc cacagggccg tcagagacaa gctgaaagaa 6120
atgtgcgagg tgttcttccc actgtgcgag agaagaaagg ccggctctgc ttgggtttcc 6180
ctgcctcctc tgcttgaaac accagccgcc tacagcagca gattctgcag cagatctggc 6240
gtggccggct tcagagccgt ggaagtgatt cctggcttcg agctgaagta cccctggtct 6300
tggctgaagg ataagaagga caaggccggc aatctggcca aagaagccct gaacatccgg 6360
accgtgtctg agcagctgaa ggcctttaac caggacaagc ccgagaagcc caggacactg 6420

ctggtgccta ttgccggcgg acctatcttc gtgcccatct ctgaagtggg cctgtccagc 6480
ttcggactga agcctcaggt tgtgcaggcc gacatcaacg ccgccatcaa tctgggactc 6540
agagccatca gcgaccctcg gatttgggag attcacccca gactgcggac cgagaagaga 6600
gatggcagac tgttcgccag agagaaacgg aagtacggcg aagagaaggt cgaggtgcag 6660
cccagcaaga atgagaaggc caaaaaagtg aaggacgacc ggaagcctaa ctacttcgcc 6720
gatttcagcg gcaaggtgga ctggggcttt ggcaacatta agaacgagtc cggcctgaca 6780
ctggtgtctg gcaaagcact gtggtggacc atcaaccagc tgcagtggga gagatgcttt 6840

gacatcaaca agcggcacat cgaggactgg tccaacaagc agaagcaagg atccaaaagg 6900
ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatctta cccatacgat 6960
gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 7020
gactatgcct aagaattctg cagatatcca gcacagtggc ggccgctcga gtctagaggg 7080
cccgtttaaa cccgctgatc agcctcgact gtgccttcta gttgccagcc atctgttgtt 7140
tgcccctccc ccgtgccttc cttgaccctg gaaggtgcca ctcccactgt cctttcctaa 7200
taaaatgagg aaattgcatc gcattgtctg agtaggtgtc attctattct gggggggtggg 7260

gtggggcagg acagcaaggg ggaggattgg gaagacaata gcaggcatgc tggggatgcg 7320
gtgggctcta tggcttctga ggcggaaaga accagctggg gctctagggg gtatccccac 7380
gcgccctgta gcggcgcatt aagcgcggcg ggtgtggtgg ttacgcgcag cgtgaccgct 7440
acacttgcca gcgccctagc gcccgctcct ttcgctttct tcccttcctt tctcgccacg 7500
ttcgccggct ttccccgtca agctctaaat cgggggctcc ctttagggtt ccgatttagt 7560
gctttacggc acctcgaccc caaaaaactt gattagggtg atggttcacg tagtgggcca 7620
tcgccctgat agacggtttt tcgccctttg acgttggagt ccacgttctt taatagtgga 7680

ctcttgttcc aaactggaac aacactcaac cctatctcgg tctattcttt tgatttataa 7740

gggattttgc cgatttcggc ctattggtta aaaaatgagc tgatttaaca aaaatttaac 7800

gcgaattaat tctgtggaat gtgtgtcagt tagggtgtgg aaagtcccca ggctccccag 7860

caggcagaag tatgcaaagc atgcatctca attagtcagc aaccaggtgt ggaaagtccc 7920

caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccatag 7980

tcccgcccct aactccgccc atcccgcccc taactccgcc cagttccgcc cattctccgc 8040

cccatggctg actaattttt tttatttatg cagaggccga ggccgcctct gcctctgagc 8100

tattccagaa gtagtgagga ggctttttg gaggcctagg cttttgcaaa aagctcccgg 8160

gagcttgtat atccattttc ggatctgatc aagagacagg atgaggatcg tttcgcatga 8220

ttgaacaaga tggattgcac gcaggttctc cggccgcttg ggtggagagg ctattcggct 8280

atgactgggc acaacagaca atcggctgct ctgatgccgc cgtgttccgg ctgtcagcgc 8340

aggggcgccc ggttcttttt gtcaagaccg acctgtccgg tgccctgaat gaactgcagg 8400

acgaggcagc gcggctatcg tggctggcca cgacgggcgt tccttgcgca gctgtgctcg 8460

acgttgtcac tgaagcggga agggactggc tgctattggg cgaagtgccg gggcaggatc 8520

tcctgtcatc tcaccttgct cctgccgaga aagtatccat catggctgat gcaatgcggc 8580

ggctgcatac gcttgatccg gctacctgcc cattcgacca ccaagcgaaa catcgcatcg 8640

agcgagcacg tactcggatg gaagccggtc ttgtcgatca ggatgatctg gacgaagagc 8700

atcaggggct cgcgccagcc gaactgttcg ccaggctcaa ggcgcgcatg cccgacggcg 8760

aggatctcgt cgtgacccat ggcgatgcct gcttgccgaa tatcatggtg gaaaatggcc 8820

gcttttctgg attcatcgac tgtggccggc tgggtgtggc ggaccgctat caggacatag 8880

cgttggctac ccgtgatatt gctgaagagc ttggcggcga atgggctgac cgcttcctcg 8940

tgctttacgg tatcgccgct cccgattcgc agcgcatcgc cttctatcgc cttcttgacg 9000

agttcttctg agcgggactc tggggttcga aatgaccgac caagcgacgc ccaacctgcc 9060

atcacgagat ttcgattcca ccgccgcctt ctatgaaagg ttgggcttcg gaatcgtttt 9120

ccgggacgcc ggctggatga tcctccagcg cggggatctc atgctggagt tcttcgccca 9180

ccccaacttg tttattgcag cttataatgg ttacaaataa agcaatagca tcacaaattt 9240

cacaaataaa gcatttttt cactgcattc tagttgtggt ttgtccaaac tcatcaatgt 9300

atcttatcat gtctgtatac cgtcgacctc tagctagagc ttggcgtaat catggtcata 9360

gctgtttcct gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag 9420

cataaagtgt aaagcctggg gtgcctaatg agtgagctaa ctcacattaa ttgcgttgcg 9480

ctcactgccc gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca 9540

acgcgcgggg agaggcggtt tgcgtattgg gcgctcttcc gcttcctcgc tcactgactc 9600

gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg 9660

gttatccaca gaatcagggg ataacgcagg aa 9692

<210> 583

<211> 9041

<212> DNA

<213> Alicyclobacillus kakegawensis

<400> 583

agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 60

cgtttttcca taggctccgc cccccctgacg agcatcacaa aaatcgacgc tcaagtcaga 120

ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 180

tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 240

gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc 300

gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 360

gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 420

ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt 480

ggcctaacta cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag 540

ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 600

gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 660

ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 720

tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 780

ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 840

gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg 900

tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac 960

cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1020

ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc 1080

gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta 1140

caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1200

gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc 1260

ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac 1320

tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 1380

caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 1440

tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 1500

cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca 1560

ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa 1620

aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 1680

tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 1740

gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc 1800

gaaaagtgcc acctgacgtc gacggatcgg gagatctccc gatcccctat ggtgcactct 1860

cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt 1920

ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag gcttgaccga 1980

caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc 2040

cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa ttacggggtc 2100

attagttcat agcccatata tggagttccg cgttacataa cttacggtaa atggcccgcc 2160

tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt 2220

aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt aaactgccca 2280

cttggcagta catcaagtgt atcatatgcc aagtacgccc cctattgacg tcaatgacgg 2340

taaatggccc gcctggcatt atgcccagta catgacctta tgggactttc ctacttggca 2400

gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc agtacatcaa 2460

tgggcgtgga tagcggtttg actcacgggg atttccaagt ctccacccca ttgacgtcaa 2520

tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta acaactccgc 2580

cccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa gcagagctct 2640

ctggctaact agagaaccca ctgcttactg gcttatcgaa attaatacga ctcactatag 2700

ggagacccaa gctggctagc gtttaaactt aagcttgcca ccatggcccc aaagaagaag 2760

cggaaggtcg gtatccacgg agtcccagca gccggatccg ccgtgaagtc catcaaagtg 2820

aagctgcggc tgagcgagtg ccccgatatt ctggctggaa tgtggcagct gcacagagcc 2880

acaaatgccg gcgtgcggta ctacacagaa tgggtgtccc tgatgcggca agagatcctg 2940

tacagcagag gccctgatgg cggccagcag tgttatatga ccgccgagga ttgccagaga 3000

gagctgctgc ggagactgcg gaatagacag ctgcataacg gccggcagga tcagcctgga 3060

acagatgctg atctgctggc catcagcaga cggctgtacg agattctggt gctgcagagc 3120

atcggcaaaa gaggcgacgc ccagcagatt gccagcagct ttctgagccc tctggtggac 3180

cccaacagca aaggtggaag aggcgaggcc aagagcggaa gaaaacctgc ctggcagaag 3240

atgcgcgacc agggcgatcc tagatgggtt gccgctagag agaagtacga gcagcggaag 3300

gccgtggatc ccagcaaaga gattctgaac agcctggacg ccctgggcct cagacctctg 3360

tttgccgtgt tcaccgagac atacagatcc ggcgtggact ggaagcctct gggcaaatct 3420

cagggcgtca gaacctggga cagagacatg tttcagcagg ccctggaacg gctgatgagc 3480

tgggagagct ggaatcggag agtgggcgaa gagtacgcca gactgttcca gcagaaaatg 3540

aagttcgagc aagagcactt cgccgagcag agccacctgg tcaaactggc tagagccctg 3600

gaagccgata tgagagccgc ctctcagggc ttcgaggcca aaagaggaac agcccaccag 3660

atcaccagaa gggcactgag aggggccgac agagtgttcg agatctggaa gtctatcccc 3720

gaggaagccc tgttcagcca gtacgacgaa gtgatcagac aggtgcaggc cgagaagcgg 3780

agagatttcg gcagccatga cctgttcgcc aagctggccg agcctaagta tcagcccctt 3840

tggagagccg acgagacatt cctgaccaga tacgccctgt acaacggcgt gctgcgcgat 3900

ctggaaaagg ccagacagtt cgccaccttc acactgcctg atgcctgcgt gaaccccatc 3960

tggaccagat tcgagtctag ccagggcagc aacctgcaca aatacgagtt tctgttcgac 4020

cacctcggac ctggcagaca cgccgtcaga tttcagagac tgctggtggt ggaaagcgag 4080

ggcgccaaag aaagggatag cgtggtggtg cctgtggctc cttctggcca actggataag 4140

ctggtgctga gggaagaaga gaagtccagc gtcgccctgc atctgcacga taccgctaga 4200

cccgatggct tcatggctga atgggctggc gccaaactgc agtacgagag aagcaccctg 4260

gccagaaaag ccagacggga caagcagggc atgagaagct ggcggagaca gccctccatg 4320

ctgatgtctg ccgctcagat gctggaagat gccaaacagg ctggcgacgt gtacctgaac 4380

atcagcgtgc gcgtgaagtc tcccagtgaa gtgcgaggac agaggcggcc tccttacgcc 4440

gctctgttta gaatcgacga caagcagcgg agagtgaccg tgaactacaa caagctgagc 4500

gcctacctgg aagaacaccc cgataagcag atccctggcg ctcctggact gctgtctgga 4560

ctgagagtga tgtccgtgga cctgggcctg agaacaagcg ccagcatctc cgtgttcaga 4620

gtggccaaga aagaagaggt ggaagccctc ggagatggcc ggcctcctca ctactatcct 4680

atccacggca ccgatgacct ggtggccgtg cacgaaagat cccacctgat tcagatgccc 4740

ggcgaaaccg agacaaagca gctgcggaag ctgagagaag aacggcaggc cgtgctgagg 4800

ccactgtttg ctcaactggc actgctgaga ctgctcgtca gatgtggcgc cgctgacgag 4860

agaatcagaa ccagatcctg gcagcggctg accaagcagg gaagagagtt caccaagaga 4920

ctgaccccta gctggcgcga ggctctggaa ctggaactga caagactcga ggcctactgc 4980

ggcagagtgc ccgatgatga gtggtccaga atcgtggaca gaaccgtgat tgccctgtgg 5040

cggagaatgg gcaagcaagt gcgcgattgg cggaagcaag tgaagtccgg ggccaaagtg 5100
aaagtgaagg gctaccagct ggatgtcgtc ggcggaaatt ctctggccca gatcgactat 5160
ctggaacagc agtacaagtt cctgcggcgt tggagcttct tcgccagagc ttctggcctg 5220
gtcgtgcggg ccgatagaga aagccatttt gccgtggctc tgagacagca catcgagaac 5280
gccaagcggg acagactgaa gaaactggcc gaccggatcc tgatggaagc actgggctat 5340
gtgtacgagg ccagcggacc tagagaaggc cagtggacag ctcagcaccc tccttgccag 5400
ctgatcattc tcgaggaact gtccgcctac cggttcagcg acgatagacc tcctagcgag 5460

aacagcaaac tgatggcctg gggccacaga ggcatcctcg aagaactggt caaccaggct 5520
caggtgcacg atgtgctcgt gggcacagtg tacgccgcct tcagcagcag attcgacgct 5580
agaacaggtg ctcccggcgt cagatgcaga agagtgcctg ccagatttgt gggcgccacc 5640
gtggatgatt ctctgccact gtggctgacc gagttcctgg acaagcaccg gctggataag 5700
aacctgctgc ggcccgacga tgtgatccca acaggcgaag gcgaattcct ggtgtcccct 5760
tgtggcgaag aggctgccag agttagacag gttcacgccg acatcaacgc tgcccagaac 5820
ctgcagagaa ggctgtggca gaacttcgac atcaccgagc tgaggctgag atgcgacgtg 5880

aagatgggcg gagagggaac agtgctggtg cccagagtga acaacgccag agccaagcag 5940
ctgttcggca agaaggtgct ggtttcccag gacggcgtga ccttcttcga gagatctcag 6000
acaggcggca agccccacag cgagaagcag accgatctga ccgacaaaga actcgagctg 6060
atcgccgagg ccgatgaggc cagagctaaa agcgtggtgc tgttcaggga tcctagcggc 6120
cacattggca aaggccactg gatccggcag cgcgagtttt ggagtctggt caagcagagg 6180
atcgagagcc acaccgccga gcggattaga gttagaggcg tgggaagctc cctggacgga 6240
tccaaaaggc cggcggccac gaaaaaggcc ggccaggcaa aaaagaaaaa gggatcttac 6300

ccatacgatg ttccagatta cgcttatccc tacgacgtgc ctgattatgc atacccatat 6360
gatgtccccg actatgccta agaattctgc agatatccag cacagtggcg gccgctcgag 6420
tctagagggc ccgtttaaac ccgctgatca gcctcgactg tgccttctag ttgccagcca 6480
tctgttgttt gcccctcccc cgtgccttcc ttgaccctgg aaggtgccac tcccactgtc 6540
ctttcctaat aaaatgagga aattgcatcg cattgtctga gtaggtgtca ttctattctg 6600
gggggtgggg tggggcagga cagcaagggg gaggattggg aagacaatag caggcatgct 6660
ggggatgcgg tgggctctat ggcttctgag gcggaaagaa ccagctgggg ctctaggggg 6720

tatccccacg cgccctgtag cggcgcatta agcgcggcgg gtgtggtggt tacgcgcagc 6780

gtgaccgcta cacttgccag cgccctagcg cccgctcctt tcgctttctt cccttccttt 6840

ctcgccacgt tcgccggctt tccccgtcaa gctctaaatc gggggctccc tttagggttc 6900

cgatttagtg ctttacggca cctcgacccc aaaaaacttg attagggtga tggttcacgt 6960

agtgggccat cgccctgata gacggttttt cgccctttga cgttggagtc cacgttcttt 7020

aatagtggac tcttgttcca aactggaaca acactcaacc ctatctcggt ctattctttt 7080

gatttataag ggattttgcc gatttcggcc tattggttaa aaaatgagct gatttaacaa 7140

aaatttaacg cgaattaatt ctgtggaatg tgtgtcagtt agggtgtgga aagtccccag 7200

gctccccagc aggcagaagt atgcaaagca tgcatctcaa ttagtcagca accaggtgtg 7260

gaaagtcccc aggctcccca gcaggcagaa gtatgcaaag catgcatctc aattagtcag 7320

caaccatagt cccgccccta actccgccca tcccgcccct aactccgccc agttccgccc 7380

attctccgcc ccatggctga ctaatttttt ttatttatgc agaggccgag gccgcctctg 7440

cctctgagct attccagaag tagtgaggag gcttttttgg aggcctaggc ttttgcaaaa 7500

agctcccggg agcttgtata tccattttcg gatctgatca agagacagga tgaggatcgt 7560

ttcgcatgat tgaacaagat ggattgcacg caggttctcc ggccgcttgg gtggagaggc 7620

tattcggcta tgactgggca caacagacaa tcggctgctc tgatgccgcc gtgttccggc 7680

tgtcagcgca ggggcgcccg gttctttttg tcaagaccga cctgtccggt gccctgaatg 7740

aactgcagga cgaggcagcg cggctatcgt ggctggccac gacgggcgtt ccttgcgcag 7800

ctgtgctcga cgttgtcact gaagcgggaa gggactggct gctattgggc gaagtgccgg 7860

ggcaggatct cctgtcatct caccttgctc ctgccgagaa agtatccatc atggctgatg 7920

caatgcggcg gctgcatacg cttgatccgg ctacctgccc attcgaccac caagcgaaac 7980

atcgcatcga gcgagcacgt actcggatgg aagccggtct tgtcgatcag gatgatctgg 8040

acgaagagca tcaggggctc gcgccagccg aactgttcgc caggctcaag gcgcgcatgc 8100

ccgacggcga ggatctcgtc gtgacccatg gcgatgcctg cttgccgaat atcatggtgg 8160

aaaatggccg cttttctgga ttcatcgact gtggccggct gggtgtggcg gaccgctatc 8220

aggacatagc gttggctacc cgtgatattg ctgaagagct tggcggcgaa tgggctgacc 8280

gcttcctcgt gctttacggt atcgccgctc ccgattcgca gcgcatcgcc ttctatcgcc 8340

ttcttgacga gttcttctga gcgggactct ggggttcgaa atgaccgacc aagcgacgcc 8400

caacctgcca tcacgagatt tcgattccac cgccgccttc tatgaaaggt tgggcttcgg 8460

aatcgttttc cgggacgccg gctggatgat cctccagcgc ggggatctca tgctggagtt 8520

cttcgcccac cccaacttgt ttattgcagc ttataatggt tacaaataaa gcaatagcat 8580

cacaaatttc acaaataaag catttttttc actgcattct agttgtggtt tgtccaaact 8640

catcaatgta tcttatcatg tctgtatacc gtcgacctct agctagagct tggcgtaatc 8700

atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg 8760

agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac tcacattaat 8820

tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg 8880

aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg cttcctcgct 8940

cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc 9000

ggtaatacgg ttatccacag aatcagggga taacgcagga a 9041

<210> 584

<211> 7826

<212> DNA

<213> Unknown

<220><223> Member of the class Phycisphaerae

<400> 584

gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg 60

ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120

cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180

ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt 240

gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata 300

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 480

atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540

atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca 600

tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg 660

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 720

aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg 780

gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca 840

ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc 900

gtttaaactt aagcttgcca ccatggcccc aaagaagaag cggaaggtcg gtatccacgg 960

agtcccagca gccgccacca agagctacag agccagaatc ctgaccgaca gcagactggc 1020

cgctgctctg gatagaaccc acgtggtgtt tgtggaaagc ctgaagcaga tgatcaacac 1080

ctacctgcgg atgcagaacg gcaagttcgg ccccgaccac aagaaactgg cccagatcat 1140

gctgagccgg tccaacacat ttgcccacgg cgtgatggac cagatcacca gagatcagcc 1200

caccagcaca ctggacgagg aatggaccga cctggccaga agaatccaca agacaaccgg 1260

acctctgttc ctgcaagccg agagattcgc caccgtgaag aacagagcca tccacaccaa 1320

gtccagaggc aaagtgatcc catctcctga gacactggcc gtgcctgcca agttctggca 1380

ccaagtgtgc gatagcgcca gcgcctacat cagatccaac cgcgaactga tgcagcagtg 1440

gcggaaagat agagccgcct ggctgaagga caagaacgag tggcagcaga aacaccccga 1500

gttcatgcag ttctacaacg gcccctacca gaacttcctg aagctgtgcg acgacgacag 1560

aatcacctct cagctggctg ccgagcagca gcctacagcc agcaagaaca acagacccag 1620

aaagaccggc aagcgcttcg ccagatggca cctgtggtac aagtggctga gcgagaaccc 1680

cgagatcatc gaatggcgga acaaggcctc cgccagcgac tttaagaccg tgaccgatga 1740

cgtgcggaag cagatcatta ccaagtatcc ccagcagaac aagtacatca cccggctgct 1800

ggactggctg gaagataaca accccgagct gaaaaccctg gaaaacctgc ggcggaccta 1860

cgtgaagaag ttcgacagct tcaagcggcc tcctacactg accctgccat ctccatacag 1920

acacccctac tggttcacca tggaactgga ccagttttac aagaaggccg acttcgagaa 1980

cggcaccatc cagctgctgc tgatcgacga ggacgacgac ggcaactggt tcttcaactg 2040

gatgcccgcc tctctgaagc ccgatcctag actggtgcct tcttggagag ccgaaacctt 2100

cgagacagag ggcagattcc ctccttacct cggcggcaag atcggcaaga agctgagcag 2160

acctgctcct accgacgccg agagaaaggc tggaattgcc ggcgctaagc tgatgattaa 2220

gaacaatcgg agcgagctgc tgttcaccgt gttcgagcag gactgccctc ctagagtgaa 2280

gtgggccaag accaagaacc ggaagtgccc tgccgacaac gcctttagct ccgacggcaa 2340

gaccagaaag cccctgagaa tcctgtccat cgacctgggc atcagacaca tcggcgcctt 2400

cgctctgaca cagggcacca gaaatgatag cgcctggcag accgagagcc tgaagaaggg 2460

catcatcaac agccctagca tccctccact gcggcaagtg cggagacacg actacgacct 2520

gaagcggaaa agacggcggc acggcaagcc tgtgaagggc cagagaagca acgccaatct 2580

gcaggcccac aggaccaaca tggcccagga cagattcaag aagggcgcct ctgccatcgt 2640

gtcactggcc agagagcata gcgccgacct gatcctgttc gagaacctgc acagcctgaa 2700

gttcagcgcc ttcgacgagc ggtggatgaa cagacagctg cgggacatga accggcggca 2760

catcgtggaa ctggtgtctg aacaggcccc tgagttcggc atcacagtga aggacgacat 2820

caacccctgg atgaccagcc ggatctgcag caactgtaac ctgcctggct tcaggttcag 2880

catgaagaag aagaacccct accgcgagaa gctgcccaga gagaagtgca ccgatttcgg 2940

ctaccctgtg tgggaacctg gcggccacct gtttagatgc cctcactgcg accacagagt 3000

gaacgccgac attaacgccg ctgccaacct ggccaacaag ttctttggcc tcggctactg 3060

gaacaacggc ctgaagtacg atgccgagac aaagaccttc accgtgcaca ccgacaagaa 3120

aaccccacct ctgatcttca agcccagacc tcagttcgat ctgtgggccg acagcgtgaa 3180

aacacggaag cagcttggcc ccgatccttt caaaaggccg gcggccacga aaaaggccgg 3240

ccaggcaaaa aagaaaaagg gatcctaccc atacgatgtt ccagattacg cttatcccta 3300

cgacgtgcct gattatgcat acccatatga tgtccccgac tatgcctaag aattctgcag 3360

atatccagca cagtggcggc cgctcgagtc tagagggccc gtttaaaccc gctgatcagc 3420

ctcgactgtg ccttctagtt gccagccatc tgttgtttgc ccctcccccg tgccttcctt 3480

gaccctggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca 3540

ttgtctgagt aggtgtcatt ctattctggg gggtggggtg gggcaggaca gcaaggggga 3600

ggattgggaa gacaatagca ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc 3660

ggaaagaacc agctggggct ctaggggGta tccccacgcg ccctgtagcg gcgcattaag 3720

cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccctagcgcc 3780

cgctcctttc gctttcttcc cttcctttct cgccacgttc gccggctttc cccgtcaagc 3840

tctaaatcgg gggctccctt tagggttccg atttagtgct ttacggcacc tcgaccccaa 3900

aaaacttgat tagggtgatg gttcacgtag tgggccatcg ccctgataga cggtttttcg 3960

ccctttgacg ttggagtcca cgttctttaa tagtggactc ttgttccaaa ctggaacaac 4020

actcaaccct atctcggtct attcttttga tttataaggg attttgccga tttcggccta 4080

ttggttaaaa aatgagctga tttaacaaaa atttaacgcg aattaattct gtggaatgtg 4140

tgtcagttag ggtgtggaaa gtccccaggc tccccagcag gcagaagtat gcaaagcatg 4200

catctcaatt agtcagcaac caggtgtgga aagtccccag gctccccagc aggcagaagt 4260

atgcaaagca tgcatctcaa ttagtcagca accatagtcc cgcccctaac tccgcccatc 4320

ccgcccctaa ctccgcccag ttccgcccat tctccgcccc atggctgact aatttttttt 4380

atttatgcag aggccgaggc cgcctctgcc tctgagctat tccagaagta gtgaggaggc 4440

ttttttggag gcctaggctt ttgcaaaaag ctcccgggag cttgtatatc cattttcgga 4500

tctgatcaag agacaggatg aggatcgttt cgcatgattg aacaagatgg attgcacgca 4560

ggttctccgg ccgcttgggt ggagaggcta ttcggctatg actgggcaca acagacaatc 4620

ggctgctctg atgccgccgt gttccggctg tcagcgcagg ggcgcccggt tcttttttgtc 4680

aagaccgacc tgtccggtgc cctgaatgaa ctgcaggacg aggcagcgcg gctatcgtgg 4740

ctggccacga cgggcgttcc ttgcgcagct gtgctcgacg ttgtcactga agcgggaagg 4800
gactggctgc tattgggcga agtgccgggg caggatctcc tgtcatctca ccttgctcct 4860
gccgagaaag tatccatcat ggctgatgca atgcggcggc tgcatacgct tgatccggct 4920
acctgcccat tcgaccacca agcgaaacat cgcatcgagc gagcacgtac tcggatggaa 4980
gccggtcttg tcgatcagga tgatctggac gaagagcatc aggggctcgc gccagccgaa 5040
ctgttcgcca ggctcaaggc gcgcatgccc gacggcgagg atctcgtcgt gacccatggc 5100
gatgcctgct tgccgaatat catggtggaa aatggccgct tttctggatt catcgactgt 5160

ggccggctgg gtgtggcgga ccgctatcag gacatagcgt tggctacccg tgatattgct 5220
gaagagcttg gcggcgaatg ggctgaccgc ttcctcgtgc tttacggtat cgccgctccc 5280
gattcgcagc gcatcgcctt ctatcgcctt cttgacgagt tcttctgagc gggactctgg 5340
ggttcgaaat gaccgaccaa gcgacgccca acctgccatc acgagatttc gattccaccg 5400
ccgccttcta tgaaaggttg ggcttcggaa tcgttttccg ggacgccggc tggatgatcc 5460
tccagcgcgg ggatctcatg ctggagttct tcgcccaccc caacttgttt attgcagctt 5520
ataatggtta caaataaagc aatagcatca caaatttcac aaataaagca ttttttttcac 5580

tgcattctag ttgtggtttg tccaaactca tcaatgtatc ttatcatgtc tgtataccgt 5640
cgacctctag ctagagcttg gcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt 5700
atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg 5760
cctaatgagt gagctaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg 5820
gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc 5880
gtattgggcg ctcttccgct tcctcgctca ctgactcgct gcgctcggtc gttcggctgc 5940
ggcgagcggt atcagctcac tcaaaggcgg taatacggtt atccacagaa tcaggggata 6000

acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg 6060
cgttgctggc gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct 6120
caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcgttt cccccctggaa 6180
gctccctcgt gcgctctcct gttccgaccc tgccgcttac cggatacctg tccgcctttc 6240
tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcggtgt 6300
aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg 6360
ccttatccgg taactatcgt cttgagtcca acccggtaag acacgactta tcgccactgg 6420

cagcagccac tggtaacagg attagcagag cgaggtatgt aggcggtgct acagagttct 6480

tgaagtggtg gcctaactac ggctacacta gaagaacagt atttggtatc tgcgctctgc 6540

tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg 6600

ctggtagcgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 6660

aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 6720

ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 6780

gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 6840

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 6900

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 6960

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 7020

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 7080

gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 7140

ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 7200

cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 7260

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 7320

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 7380

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 7440

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 7500

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 7560

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 7620

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 7680

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 7740

tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggggtt ccgcgcacat 7800

ttccccgaaa agtgccacct gacgtc 7826

<210> 585

<211> 7814

<212> DNA

<213> Unknown

<220><223> Member of the phylum Planctomycetes

<400> 585

gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg 60

ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120

cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180

ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt 240

gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata 300

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 480

atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540

atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca 600

tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg 660

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 720

aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg 780

gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca 840

ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc 900

gtttaaactt aagcttgcca ccatggcccc aaagaagaag cggaaggtcg gtatccacgg 960

agtcccagca gccagcgtgc ggagctttca ggccagagtg gaatgcgaca agcagaccat 1020

ggaacacctg tggcggaccc acaaggtgtt caacgagaga ctgcccgaga tcatcaagat 1080

cctgttcaag atgaagcggg gcgagtgcgg ccagaacgat aagcagaagt ccctgtacaa 1140

gagcatcagc cagagcatcc tggaagccaa cgctcagaac gccgactacc tgctgaacag 1200

cgtgtccatc aaaggctgga agcctggcac cgccaagaag tacagaaacg ccagcttcac 1260

ctgggccgac gatgccgcta aactgtctag ccagggcatc cacgtgtacg acaagaaaca 1320

ggtgctgggc gacctgcctg gcatgatgtc tcagatggtc tgcaggcaga gcgtggaagc 1380

catctctgga cacatcgagc tgaccaagaa gtgggagaaa gaacacaacg agtggctgaa 1440

agaaaaagag aaatgggagt ccgaggacga gcacaagaag tatctggacc tgcgcgagaa 1500

gttcgagcag tttgagcaga gcatcggcgg caagatcacc aagagaagag gccggtggca 1560

cctgtacctg aagtggctga gcgacaaccc tgattttgcc gcctggcggg gaaacaaggc 1620

cgtgatcaat cctctgagcg agaaggccca gatcaggatc aacaaggcca agccgaacaa 1680

gaagaacagc gtcgagcggg acgagttctt caaggccaat cctgagatga aggccctgga 1740

caacctgcac ggctactacg agcggaattt cgtgcggcgg agaaagacaa agaagaaccc 1800

cgacggcttc gaccacaagc ctaccttcac actgccccat cctaccatcc atcctcgttg 1860

gttcgtgttc aacaagccta agacaaaccc cgagggctac cgcaagctga tcctgcctaa 1920

aaaggccggc gatctgggca gcctggaaat gagactgctg accggcgaga agaacaaggg 1980

caactacccc gacgactgga tcagcgtgaa gtttaaggcc gatcctcggc tgagcctgat 2040

cagacccgtg aaaggcagac gggttgtgcg gaagggcaaa gagcagggcc agaccaaaga 2100

gacagacagc tacgagtttt tcgacaagca cctgaagaag tggcggcctg ccaaactgtc 2160

tggcgtgaag ctgatcttcc ccgacaagac acctaaggcc gcctacctgt acttcacctg 2220

tgacatcccc gacgagcccc tgaccgagac agccaagaaa atccagtggc tggaaaccgg 2280

cgacgtgacc aaaaagggca agaaacgcaa aaagaaggtg ctgccccacg gcctggtgtc 2340

ctgtgctgtt gatctgagca tgcggagagg caccaccggc tttgccacac tgtgcagata 2400

cgagaatggc aagatccaca tcctgcggag ccggaacctg tgggtcggat acaaagaagg 2460

caagggctgt caccctaca gatggacaga gggacccgac ctgggacaca ttgccaagca 2520

caagagagag atcagaatcc tgcggtccaa gcggggcaag cctgtgaagg gcgaagagag 2580

ccacatcgac ctgcagaaac acatcgacta catgggcgaa gatcggttca agaaggccgc 2640

cagaaccatc gtgaacttcg ccctgaacac cgagaacgcc gccagcaaga atggcttcta 2700

ccccagagct gacgtgctgc tgctggaaaa cctggaagga ctgatccccg atgccgagaa 2760

agagcggggc atcaatagag ccctggccgg ctggaataga cggcacctgg ttgagcgcgt 2820

gatcgagatg gccaaggatg ccggcttcaa gcggcgggtg ttcgagatcc caccttacgg 2880

cacaagccaa gtgtgcagca aatgtggcgc cctgggcaga agatacagca tcatcagaga 2940

gaacaaccgg cgcgagatca gattcggcta cgtggaaaag ctgttcgcct gtcctaactg 3000

cggctactgc gccaacgccg atcacaatgc cagcgtgaac ctgaaccggc ggttcctgat 3060

cgaggacagc ttcaagtcct actacgactg gaagcggctg tccgagaaga agcagaaaga 3120

ggaaatcgag acaatcgagt ccaagctgat ggataagctg tgcgccatgc acaagatcag 3180

ccggggcagc atcagcaaga aaaggccggc ggccacgaaa aaggccggcc aggcaaaaaa 3240

gaaaaaggga tcctacccat acgatgttcc agattacgct tatccctacg acgtgcctga 3300

ttatgcatac ccatatgatg tccccgacta tgcctaagaa ttctgcagat atccagcaca 3360

gtggcggccg ctcgagtcta gagggcccgt ttaaacccgc tgatcagcct cgactgtgcc 3420

ttctagttgc cagccatctg ttgtttgccc ctcccccgtg ccttccttga ccctggaagg 3480

tgccactccc actgtccttt cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag 3540

gtgtcattct attctggggg gtggggtggg gcaggacagc aaggggggagg attgggaaga 3600

caatagcagg catgctgggg atgcggtggg ctctatggct tctgaggcgg aaagaaccag 3660

ctggggctct agggggtatc cccacgcgcc ctgtagcggc gcattaagcg cggcgggtgt 3720

ggtggttacg cgcagcgtga ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc 3780

tttcttccct tcctttctcg ccacgttcgc cggctttccc cgtcaagctc taaatcgggg 3840

gctcccttta gggttccgat ttagtgcttt acggcacctc gaccccaaaa aacttgatta 3900

gggtgatggt tcacgtagtg ggccatcgcc ctgatagacg gtttttcgcc ctttgacgtt 3960

ggagtccacg ttctttaata gtggactctt gttccaaact ggaacaacac tcaaccctat 4020

ctcggtctat tcttttgatt tataagggat tttgccgatt tcggcctatt ggttaaaaaa 4080

tgagctgatt taacaaaaat ttaacgcgaa ttaattctgt ggaatgtgtg tcagttaggg 4140

tgtggaaagt ccccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag 4200

tcagcaacca ggtgtggaaa gtccccaggc tccccagcag gcagaagtat gcaaagcatg 4260

catctcaatt agtcagcaac catagtcccg cccctaactc cgcccatccc gcccctaact 4320

ccgcccagtt ccgcccattc tccgccccat ggctgactaa ttttttttat ttatgcagag 4380

gccgaggccg cctctgcctc tgagctattc cagaagtagt gaggaggctt ttttggaggc 4440

ctaggctttt gcaaaaagct cccgggagct tgtatatcca ttttcggatc tgatcaagag 4500

acaggatgag gatcgtttcg catgattgaa caagatggat tgcacgcagg ttctccggcc 4560

gcttgggtgg agaggctatt cggctatgac tgggcacaac agacaatcgg ctgctctgat 4620

gccgccgtgt tccggctgtc agcgcagggg cgcccggttc tttttgtcaa gaccgacctg 4680

tccggtgccc tgaatgaact gcaggacgag gcagcgcggc tatcgtggct ggccacgacg 4740

ggcgttcctt gcgcagctgt gctcgacgtt gtcactgaag cgggaaggga ctggctgcta 4800

ttgggcgaag tgccggggca ggatctcctg tcatctcacc ttgctcctgc cgagaaagta 4860

tccatcatgg ctgatgcaat gcggcggctg catacgcttg atccggctac ctgcccattc 4920

gaccaccaag cgaaacatcg catcgagcga gcacgtactc ggatggaagc cggtcttgtc 4980

gatcaggatg atctggacga agagcatcag gggctcgcgc cagccgaact gttcgccagg 5040

ctcaaggcgc gcatgcccga cggcgaggat ctcgtcgtga cccatggcga tgcctgcttg 5100

ccgaatatca tggtggaaaa tggccgcttt tctggattca tcgactgtgg ccggctgggt 5160

gtggcggacc gctatcagga catagcgttg gctacccgtg atattgctga agagcttggc 5220

ggcgaatggg ctgaccgctt cctcgtgctt tacggtatcg ccgctcccga ttcgcagcgc 5280

atcgccttct atcgccttct tgacgagttc ttctgagcgg gactctgggg ttcgaaatga 5340

ccgaccaagc gacgcccaac ctgccatcac gagatttcga ttccaccgcc gccttctatg 5400

aaaggttggg cttcggaatc gttttccggg acgccggctg gatgatcctc cagcgcgggg 5460

atctcatgct ggagttcttc gcccacccca acttgtttat tgcagcttat aatggttaca 5520

aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt 5580

gtggtttgtc caaactcatc aatgtatctt atcatgtctg tataccgtcg acctctagct 5640

agagcttggc gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa 5700
ttccacacaa catacgagcc ggaagcataa agtgtaaagc ctggggtgcc taatgagtga 5760
gctaactcac attaattgcg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt 5820
gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgct 5880
cttccgcttc ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat 5940
cagctcactc aaaggcggta atacggttat ccacagaatc aggggataac gcaggaaaga 6000
acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt 6060

ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt 6120
ggcgaaaccc gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc 6180
gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa 6240
gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct 6300
ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga ccgctgcgcc ttatccggta 6360
actatcgtct tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg 6420
gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc 6480

ctaactacgg ctacactaga agaacagtat ttggtatctg cgctctgctg aagccagtta 6540
ccttcggaaa aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtt 6600
tttttgtttg caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga 6660
tcttttctac ggggtctgac gctcagtgga acgaaaactc acgttaaggg attttggtca 6720
tgagattatc aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga agttttaaat 6780
caatctaaag tatatatgag taaacttggt ctgacagtta ccaatgctta atcagtgagg 6840
cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc cccgtcgtgt 6900

agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 6960
acccacgctc accggctcca gatttatcag caataaacca gccagccgga agggccgagc 7020
gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 7080
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt gctacaggca 7140
tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa 7200
ggcgagttac atgatccccc atgttgtgca aaaaagcggt tagctccttc ggtcctccga 7260
tcgttgtcag aagtaagttg gccgcagtgt tatcactcat ggttatggca gcactgcata 7320

attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 7380

agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 7440

ataataccgc gccacatagc agaactttaa aagtgctcat cattggaaaa cgttcttcgg 7500

ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 7560

cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag 7620

gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 7680

tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca 7740

tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 7800

tgccacctga cgtc 7814

<210> 586

<211> 8936

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 586

agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 60

cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga 120

ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 180

tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 240

gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc 300

gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 360

gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 420

ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt 480

ggcctaacta cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag 540

ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 600

gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 660

ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 720

tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 780

ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 840

gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg 900

tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac 960

cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1020

ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc 1080

gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta 1140

caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1200

gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc 1260

ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac 1320

tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 1380

caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 1440

tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 1500

cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca 1560

ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa 1620

aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 1680

tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 1740

gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc 1800

gaaaagtgcc acctgacgtc gacggatcgg gagatctccc gatcccctat ggtgcactct 1860

cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt 1920

ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag gcttgaccga 1980

caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc 2040

cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa ttacggggtc 2100

attagttcat agcccatata tggagttccg cgttacataa cttacggtaa atggcccgcc 2160

tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt 2220

aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt aaactgccca 2280

cttggcagta catcaagtgt atcatatgcc aagtacgccc cctattgacg tcaatgacgg 2340

taaatggccc gcctggcatt atgcccagta catgacctta tgggactttc ctacttggca 2400

gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc agtacatcaa 2460

tgggcgtgga tagcggtttg actcacgggg atttccaagt ctccacccca ttgacgtcaa 2520

tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta acaactccgc 2580

cccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa gcagagctct 2640

ctggctaact agagaaccca ctgcttactg gcttatcgaa attaatacga ctcactatag 2700

ggagacccaa gctggctagc gtttaaactt aagcttgcca ccatggcccc aaagaagaag 2760

cggaaggtcg gtatccacgg agtcccagca gccggatccg ccatccggtc catcaagctg 2820

aagatgaaga ccaacagcgg caccgacagc atctacctga gaaaagccct gtggcggacc 2880

caccagctga tcaatgaggg aatcgcctac tacatgaacc tgctgaccct gtaccggcaa 2940

gaggccatcg gcgacaagac caaagaagcc tatcaggccg agctgattaa catcatccgg 3000

aaccagcagc ggaacaacgg cagctctgag gaacacggct ccgaccaaga aattctggcc 3060

ctgctgagac agctgtacga gctgatcatc cccagcagca tcggcgaatc tggcgacgct 3120

aatcagctgg gcaacaagtt tctgtaccct ctggtggacc ccaacagcca gtctggcaag 3180

ggcacatcta acgccggcag aaagcccaga tggaagcggc tgaaagagga aggcaacccc 3240

gactgggaac tcgagaagaa gaaggacgag gaacgcaagg ccaaggatcc caccgtgaag 3300

atctttgaca acctgaacaa atacggcctg ctgcctctgt tcccactgtt caccaacatc 3360

cagaaagaca tcgagtggct gcccctgggc aagagacagt ctgtgcggaa gtgggacaaa 3420

gacatgttca tccaggccat cgagagactg ctgagctggg agagctggaa cagaagagtg 3480

gccgacgagt acaaacagct gaaagaaaag accgagagct actacaaaga gcacctgaca 3540

ggcggcgagg aatggatcga gaagatccgg aagttcgaga aagaacggaa catggaactg 3600

gaaaagaacg ccttcgctcc caacgacggc tacttcatca ccagcagaca gatcagaggc 3660

tgggacagag tgtacgagaa gtggtccaag ctgcccgagt ctgctagccc tgaggaactg 3720

tggaaagtgg tggccgagca gcagaacaag atgtccgaag gcttcggcga ccccaaggtg 3780

ttcagcttcc tggccaacag agagaaccgg gacatttgga gaggccacag cgagcggatc 3840

taccacattg ccgcctacaa cggcctgcag aagaagctga gccggaccaa agagcaggcc 3900

accttcacac tgcctgacgc cattgaacac cctctgtgga tcagatacga gagccctggc 3960

ggcaccaacc tgaatctgtt caagctggaa gagaaacaga aaaagaacta ctacgtgacc 4020

ctgagcaaga tcatctggcc cagcgaggaa aagtggattg agaaagagaa catcgagatc 4080

cctctggctc ccagcatcca gttcaaccgg cagattaagc tgaagcagca cgtgaagggc 4140

aagcaagaga tcagcttcag cgactacagc agccggatca gcctggatgg tgttctcggc 4200

ggcagcagaa tccagtttaa tcggaagtac atcaagaacc acaaagagct gctcggagag 4260

ggcgacatcg gccccgtgtt ctttaacctg gtggtggatg tggcccctct gcaagaaacc 4320

agaaacggca gactgcagag ccccatcggc aaggccctga aagtgatcag cagcgacttc 4380

tccaaagtga tcgactacaa gccgaaagaa ctcatggatt ggatgaatac cggcagcgcc 4440

agcaacagct ttggagtggc ttctctgctg gaaggcatga gagtgatgag catcgacatg 4500

ggccagagaa ccagcgcctc cgtgtccatc ttcgaggtcg tgaaagaact gcccaaggat 4560

caagagcaga agctgttcta cagcatcaac gacaccgagc tgttcgccat ccacaagcgg 4620

agctttctgc tgaacctgcc tggcgaggtg gtcaccaaga acaacaagca gcagcggcaa 4680

gagcggcgga aaaagcggca gtttgtgcgg agccagatca gaatgctggc caacgtgctg 4740

cggctggaaa caaagaaaac ccctgacgag cggaagaagg ccattcacaa gctgatggaa 4800

atcgtgcaga gctacgacag ctggaccgcc agccagaaag aagtgtggga gaaagagctg 4860

aatctcctga ccaacatggc cgccttcaat gacgagatct ggaaagaaag cctggtggaa 4920

ctgcaccacc ggatcgagcc ttacgtggga cagatcgtgt ccaagtggcg gaagggcctg 4980

tctgagggca gaaagaatct ggccggcatc agcatgtgga acatcgacga actggaagat 5040

accaggcggc tgctgatttc ctggtccaag agaagcagaa ccccaggcga ggccaacagg 5100

atcgaaaccg atgagccttt cggcagcagc ctgctccagc acattcagaa cgtgaaggac 5160

gacagactga agcagatggc caacctgatc atcatgacag ccctgggctt taagtacgac 5220

aaagaggaaa aggaccggta caagcggtgg aaagagacat accccgcctg ccagatcatc 5280

ctgttcgaga acctgaaccg ctacctgttc aacctcgacc ggtccagacg cgagaacagc 5340

agactgatga agtgggccca tcggagcatc cccagaaccg tgtctatgca gggcgagatg 5400

ttcggcctgc aagtgggcga cgttcggagc gagtacagct ccagattcca cgccaaaaca 5460

ggcgcccctg gcatcagatg tcacgccctg actgaagagg atctgaaggc cggcagcaac 5520

accctgaaga gactgatcga ggacggcttc atcaatgaga gcgagctggc ctacctgaag 5580

aagggcgata tcatccctag ccaaggcggc gaactgttcg tgacactgtc caagcggtac 5640

aagaaggaca gcgacaacaa cgagctgacc gtgatccacg ccgacatcaa cgccgctcag 5700

aatctgcaga agcggttttg gcagcaaaac agcgaggtgt acagagtgcc ctgtcagctg 5760

gccagaatgg gcgaagataa gctgtacatc cccaagagcc agaccgagac aatcaagaag 5820

tatttcggca agggctcctt cgtgaagaac aataccgaac aagaggtcta caagtgggag 5880

aagtccgaga aaatgaagat caagacggac accaccttcg acctgcaaga cctggatggc 5940

ttcgaggaca tcagcaagac cattgagctg gcacaagagc agcaaaagaa atacctgacc 6000

atgttcaggg accccagcgg ctacttttttc aacaatgaga catggcggcc tcaaaaagaa 6060

tactggtcca tcgtgaacaa catcatcaag agctgcctca agaagaagat cctgagcaac 6120

aaggtcgagc tgggatccaa aaggccggcg gccacgaaaa aggccggcca ggcaaaaaag 6180

aaaaagggat cttacccata cgatgttcca gattacgctt atccctacga cgtgcctgat 6240

tatgcatacc catatgatgt ccccgactat gcctaagaat tctgcagata tccagcacag 6300

tggcggccgc tcgagtctag agggcccgtt taaacccgct gatcagcctc gactgtgcct 6360

tctagttgcc agccatctgt tgtttgcccc tcccccgtgc cttccttgac cctggaaggt 6420

gccactccca ctgtcctttc ctaataaaat gaggaaattg catcgcattg tctgagtagg 6480

tgtcattcta ttctgggggg tggggtgggg caggacagca agggggagga ttgggaagac 6540

aatagcaggc atgctgggga tgcggtgggc tctatggctt ctgaggcgga aagaaccagc 6600
tggggctcta gggggtatcc ccacgcgccc tgtagcggcg cattaagcgc ggcgggtgtg 6660
gtggttacgc gcagcgtgac cgctacactt gccagcgccc tagcgcccgc tcctttcgct 6720
ttcttccctt cctttctcgc cacgttcgcc ggctttcccc gtcaagctct aaatcggggg 6780
ctccctttag ggttccgatt tagtgcttta cggcacctcg accccaaaaa acttgattag 6840
ggtgatggtt cacgtagtgg gccatcgccc tgatagacgg tttttcgccc tttgacgttg 6900
gagtccacgt tctttaatag tggactcttg ttccaaactg gaacaacact caaccctatc 6960

tcggtctatt cttttgattt ataagggatt ttgccgattt cggcctattg gttaaaaaat 7020
gagctgattt aacaaaaatt taacgcgaat taattctgtg gaatgtgtgt cagttagggt 7080
gtggaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 7140
cagcaaccag gtgtggaaag tccccaggct ccccagcagg cagaagtatg caaagcatgc 7200
atctcaatta gtcagcaacc atagtcccgc ccctaactcc gcccatcccg cccctaactc 7260
cgcccagttc cgcccattct ccgccccatg gctgactaat ttttttatt tatgcagagg 7320
ccgaggccgc ctctgcctct gagctattcc agaagtagtg aggaggcttt tttggaggcc 7380

taggcttttg caaaaagctc ccgggagctt gtatatccat tttcggatct gatcaagaga 7440
caggatgagg atcgtttcgc atgattgaac aagatggatt gcacgcaggt tctccggccg 7500
cttgggtgga gaggctattc ggctatgact gggcacaaca gacaatcggc tgctctgatg 7560
ccgccgtgtt ccggctgtca gcgcaggggc gcccggttct ttttgtcaag accgacctgt 7620
ccggtgccct gaatgaactg caggacgagg cagcgcggct atcgtggctg gccacgacgg 7680
gcgttccttg cgcagctgtg ctcgacgttg tcactgaagc gggaagggac tggctgctat 7740
tgggcgaagt gccggggcag gatctcctgt catctcacct tgctcctgcc gagaaagtat 7800

ccatcatggc tgatgcaatg cggcggctgc atacgcttga tccggctacc tgcccattcg 7860
accaccaagc gaaacatcgc atcgagcgag cacgtactcg gatggaagcc ggtcttgtcg 7920
atcaggatga tctggacgaa gagcatcagg ggctcgcgcc agccgaactg ttcgccaggc 7980
tcaaggcgcg catgcccgac ggcgaggatc tcgtcgtgac ccatggcgat gcctgcttgc 8040
cgaatatcat ggtggaaaat ggccgctttt ctggattcat cgactgtggc cggctgggtg 8100
tggcggaccg ctatcaggac atagcgttgg ctacccgtga tattgctgaa gagcttggcg 8160
gcgaatgggc tgaccgcttc ctcgtgcttt acggtatcgc cgctcccgat tcgcagcgca 8220

tcgccttcta tcgccttctt gacgagttct tctgagcggg actctggggt tcgaaatgac 8280

cgaccaagcg acgcccaacc tgccatcacg agatttcgat tccaccgccg ccttctatga 8340

aaggttgggc ttcggaatcg ttttccggga cgccggctgg atgatcctcc agcgcgggga 8400

tctcatgctg gagttcttcg cccaccccaa cttgtttatt gcagcttata atggttacaa 8460

ataaagcaat agcatcacaa atttcacaaa taaagcattt ttttcactgc attctagttg 8520

tggtttgtcc aaactcatca atgtatctta tcatgtctgt ataccgtcga cctctagcta 8580

gagcttggcg taatcatggt catagctgtt tcctgtgtga aattgttatc cgctcacaat 8640

tccacacaac atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagtgag 8700

ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa acctgtcgtg 8760

ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc 8820

ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc gagcggtatc 8880

agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg caggaa 8936

<210> 587

<211> 6516

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 587

accttgctcc tgccgagaaa gtatccatca tggctgatgc aatgcggcgg ctgcatacgc 60

ttgatccggc tacctgccca ttcgaccacc aagcgaaaca tcgcatcgag cgagcacgta 120

ctcggatgga agccggtctt gtcgatcagg atgatctgga cgaagagcat caggggctcg 180

cgccagccga actgttcgcc aggctcaagg cgcgcatgcc cgacggcgag gatctcgtcg 240

tgacccatgg cgatgcctgc ttgccgaata tcatggtgga aaatggccgc ttttctggat 300

tcatcgactg tggccggctg ggtgtggcgg accgctatca ggacatagcg ttggctaccc 360

gtgatattgc tgaagagctt ggcggcgaat gggctgaccg cttcctcgtg ctttacggta 420

tcgccgctcc cgattcgcag cgcatcgcct tctatcgcct tcttgacgag ttcttctgag 480

cgggactctg gggttcgaaa tgaccgacca agcgacgccc aacctgccat cacgagattt 540

cgattccacc gccgccttct atgaaaggtt gggcttcgga atcgttttcc gggacgccgg 600

ctggatgatc ctccagcgcg gggatctcat gctggagttc ttcgcccacc ccaacttgtt 660

tattgcagct tataatggtt acaaataaag caatagcatc acaaatttca caaataaagc 720

atttttttca ctgcattcta gttgtggttt gtccaaactc atcaatgtat cttatcatgt 780

ctgtataccg tcgacctcta gctagagctt ggcgtaatca tggtcatagc tgtttcctgt 840

gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca taaagtgtaa 900

agcctggggt gcctaatgag tgagctaact cacattaatt gcgttgcgct cactgcccgc 960

tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac gcgcggggag 1020

aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc tgcgctcggt 1080

cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt tatccacaga 1140

atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg 1200

taaaaaggcc gcgttgctgg cgtttttcca taggctccgc cccctgacg agcatcacaa 1260

aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt 1320

tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct 1380

gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct gtaggtatct 1440

cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc 1500

cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa gacacgactt 1560

atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc 1620

tacagagttc ttgaagtggt ggcctaacta cggctacact agaagaacag tatttggtat 1680

ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa 1740

acaaaccacc gctggtagcg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa 1800

aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa 1860

ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt 1920

aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag 1980

ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc gttcatccat 2040

agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac catctggccc 2100

cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat cagcaataaa 2160

ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg cctccatcca 2220

gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa 2280

cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta tggcttcatt 2340

cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc 2400

ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact 2460

catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc 2520

tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg 2580

ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt taaaagtgct 2640

catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc tgttgagatc 2700

cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatctttta ctttcaccag 2760

cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa taagggcgac 2820

acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca tttatcaggg 2880

ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt 2940

tccgcgcaca tttccccgaa aagtgccacc tgacgtcgac ggatcgggag atctcccgat 3000

cccctatggt gcactctcag tacaatctgc tctgatgccg catagttaag ccagtatctg 3060

ctccctgctt gtgtgttgga ggtcgctgag tagtgcgcga gcaaaattta agctacaaca 3120

aggcaaggct tgaccgactt aattaagagg gcctatttcc catgattcct tcatatttgc 3180

atatacgata caaggctgtt agagagataa ttggaattaa tttgactgta aacacaaaga 3240

tattagtaca aaatacgtga cgtagaaagt aataatttct tgggtagttt gcagttttaa 3300

aattatgttt taaaatggac tatcatatgc ttaccgtaac ttgaaagtat ttcgatttct 3360

tggctttata tatcttgtgg aaaggacgaa acaccggacc tatagggtca atgaatctgt 3420

gcgtgtgcca taagtaatta aaaattaccc accacaggag cacctgaaaa caggtgcttg 3480

gcacggagac gatatatcgt ctcttttttt caattgcatg aagaatctgc ttagggttag 3540

gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt gattattgac 3600

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 3660

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 3720

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 3780

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 3840

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 3900

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 3960

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 4020

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 4080

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 4140

acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca ctgcttactg 4200

gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc gtttaaactt 4260

aagcttgccg gtgccaccat ggtgagcaag ggcgaggagg ataacatggc catcatcaag 4320

gagttcatgc gcttcaaggt gcacatggag ggctccgtga acggccacga gttcgagatc 4380

gagggcgagg gcgagggccg cccctacgag ggcacccaga ccgccaagct gaaggtgacc 4440

aagggtggcc ccctgccctt cgcctgggac atcctgtccc ctcagttcat gtacggctcc 4500

aaggcctacg tgaagcaccc cgccgacatc cccgactact tgaagctgtc cttccccgag 4560

ggcttcaagt gggagcgcgt gatgaacttc gaggacggcg gcgtggtgac cgtgacccag 4620

gactcctccc tgcaggacgg cgagttcatc tacaaggtga agctgcgcgg caccaacttc 4680

ccctccgacg gccccgtaat gcagaagaag accatgggct gggaggcctc ctccgagcgg 4740

atgtaccccg aggacggcgc cctgaagggc gagatcaagc agaggctgaa gctgaaggac 4800

ggcggccact acgacgctga ggtcaagacc acctacaagg ccaagaagcc cgtgcagctg 4860

cccggcgcct acaacgtcaa catcaagttg gacatcacct cccacaacga ggactacacc 4920

atcgtggaac agtacgaacg cgccgagggc cgccactcca ccggcggcat ggacgagctg 4980

tacaagtacc catacgatgt tccagattac gcttaagaat tctgcagata tccagcacag 5040

tggcggccgc tcgagtctag agggcccgtt taaacccgct gatcagcctc gactgtgcct 5100

tctagttgcc agccatctgt tgtttgcccc tcccccgtgc cttccttgac cctggaaggt 5160

gccactccca ctgtcctttc ctaataaaat gaggaaattg catcgcattg tctgagtagg 5220

tgtcattcta ttctgggggg tggggtgggg caggacagca agggggagga ttgggaagac 5280

aatagcaggc atgctgggga tgcggtgggc tctatggctt ctgaggcgga aagaaccagc 5340

tggggctcta gggggtatcc ccacgcgccc tgtagcggcg cattaagcgc ggcgggtgtg 5400

gtggttacgc gcagcgtgac cgctacactt gccagcgccc tagcgcccgc tcctttcgct 5460

ttcttccctt cctttctcgc cacgttcgcc ggctttcccc gtcaagctct aaatcggggg 5520

ctccctttag ggttccgatt tagtgcttta cggcacctcg accccaaaaa acttgattag 5580

ggtgatggtt cacgtagtgg gccatcgccc tgatagacgg tttttcgccc tttgacgttg 5640

gagtccacgt tctttaatag tggactcttg ttccaaactg gaacaacact caaccctatc 5700

tcggtctatt cttttgattt ataagggatt ttgccgattt cggcctattg gttaaaaaat 5760

gagctgattt aacaaaaatt taacgcgaat taattctgtg gaatgtgtgt cagttagggt 5820

gtggaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 5880

cagcaaccag gtgtggaaag tccccaggct ccccagcagg cagaagtatg caaagcatgc 5940

atctcaatta gtcagcaacc atagtcccgc ccctaactcc gcccatcccg cccctaactc 6000

cgcccagttc cgcccattct ccgccccatg gctgactaat ttttttttatt tatgcagagg 6060

ccgaggccgc ctctgcctct gagctattcc agaagtagtg aggaggcttt tttggaggcc 6120

taggcttttg caaaaagctc ccgggagctt gtatatccat tttcggatct gatcaagaga 6180

caggatgagg atcgtttcgc atgattgaac aagatggatt gcacgcaggt tctccggccg 6240

cttgggtgga gaggctattc ggctatgact gggcacaaca gacaatcggc tgctctgatg 6300

ccgccgtgtt ccggctgtca gcgcaggggc gcccggttct ttttgtcaag accgacctgt 6360

ccggtgccct gaatgaactg caggacgagg cagcgcggct atcgtggctg gccacgacgg 6420

gcgttccttg cgcagctgtg ctcgacgttg tcactgaagc gggaagggac tggctgctat 6480

tgggcgaagt gccggggcag gatctcctgt catctc 6516

<210> 588

<211> 6525

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 588

accttgctcc tgccgagaaa gtatccatca tggctgatgc aatgcggcgg ctgcatacgc 60

ttgatccggc tacctgccca ttcgaccacc aagcgaaaca tcgcatcgag cgagcacgta 120

ctcggatgga agccggtctt gtcgatcagg atgatctgga cgaagagcat caggggctcg 180

cgccagccga actgttcgcc aggctcaagg cgcgcatgcc cgacggcgag gatctcgtcg 240

tgacccatgg cgatgcctgc ttgccgaata tcatggtgga aaatggccgc ttttctggat 300

tcatcgactg tggccggctg ggtgtggcgg accgctatca ggacatagcg ttggctaccc 360

gtgatattgc tgaagagctt ggcggcgaat gggctgaccg cttcctcgtg ctttacggta 420

tcgccgctcc cgattcgcag cgcatcgcct tctatcgcct tcttgacgag ttcttctgag 480

cgggactctg gggttcgaaa tgaccgacca agcgacgccc aacctgccat cacgagattt 540

cgattccacc gccgccttct atgaaaggtt gggcttcgga atcgttttcc gggacgccgg 600

ctggatgatc ctccagcgcg gggatctcat gctggagttc ttcgcccacc ccaacttgtt 660

tattgcagct tataatggtt acaaataaag caatagcatc acaaatttca caaataaagc 720

atttttttca ctgcattcta gttgtggttt gtccaaactc atcaatgtat cttatcatgt 780

ctgtataccg tcgacctcta gctagagctt ggcgtaatca tggtcatagc tgtttcctgt 840

gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca taaagtgtaa 900

agcctggggt gcctaatgag tgagctaact cacattaatt gcgttgcgct cactgcccgc 960

tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac gcgcggggag 1020

aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc tgcgctcggt 1080

cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt tatccacaga 1140

atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg 1200

taaaaaggcc gcgttgctgg cgtttttcca taggctccgc ccccctgacg agcatcacaa 1260

aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt 1320

tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct 1380

gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct gtaggtatct 1440

cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc 1500

cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa gacacgactt 1560

atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc 1620

tacagagttc ttgaagtggt ggcctaacta cggctacact agaagaacag tatttggtat 1680

ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa 1740

acaaaccacc gctggtagcg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa 1800

aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa 1860

ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt 1920

aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag 1980

ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc gttcatccat 2040

agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac catctggccc 2100

cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat cagcaataaa 2160

ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg cctccatcca 2220

gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa 2280

cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta tggcttcatt 2340

cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc 2400

ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact 2460

catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc 2520

tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg 2580

ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt taaaagtgct 2640

catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc tgttgagatc 2700

cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatctttta ctttcaccag 2760

cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa taagggcgac 2820

acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca tttatcaggg 2880

ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt 2940

tccgcgcaca tttccccgaa aagtgccacc tgacgtcgac ggatcgggag atctcccgat 3000

cccctatggt gcactctcag tacaatctgc tctgatgccg catagttaag ccagtatctg 3060

ctccctgctt gtgtgttgga ggtcgctgag tagtgcgcga gcaaaattta agctacaaca 3120

aggcaaggct tgaccgactt aattaagagg gcctatttcc catgattcct tcatatttgc 3180
atatacgata caaggctgtt agagagataa ttggaattaa tttgactgta aacacaaaga 3240
tattagtaca aaatacgtga cgtagaaagt aataatttct tgggtagttt gcagttttaa 3300
aattatgttt taaaatggac tatcatatgc ttaccgtaac ttgaaagtat ttcgatttct 3360
tggctttata tatcttgtgg aaaggacgaa acaccggttc tgtcttttgg tcaggacaac 3420
cgtctagcta taagtgctgc agggtgtgag aaactcctat tgctggacga tgtctcttac 3480
gaggcattag cacggagacg atatatcgtc tctttttttc aattgcatga agaatctgct 3540

tagggttagg cgttttgcgc tgcttcgcga tgtacgggcc agatatacgc gttgacattg 3600
attattgact agttattaat agtaatcaat tacggggtca ttagttcata gcccatatat 3660
ggagttccgc gttacataac ttacggtaaa tggcccgcct ggctgaccgc ccaacgaccc 3720
ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag ggactttcca 3780
ttgacgtcaa tgggtggagt atttacggta aactgcccac ttggcagtac atcaagtgta 3840
tcatatgcca agtacgcccc ctattgacgt caatgacggt aaatggcccg cctggcatta 3900
tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg tattagtcat 3960

cgctattacc atggtgatgc ggttttggca gtacatcaat gggcgtggat agcggtttga 4020
ctcacgggga tttccaagtc tccaccccat tgacgtcaat gggagtttgt tttggcacca 4080
aaatcaacgg gactttccaa aatgtcgtaa caactccgcc ccattgacgc aaatgggcgg 4140
taggcgtgta cggtgggagg tctatataag cagagctctc tggctaacta gagaacccac 4200
tgcttactgg cttatcgaaa ttaatacgac tcactatagg gagacccaag ctggctagcg 4260
tttaaactta agcttgccgg tgccaccatg gtgagcaagg gcgaggagga taacatggcc 4320
atcatcaagg agttcatgcg cttcaaggtg cacatggagg gctccgtgaa cggccacgag 4380

ttcgagatcg agggcgaggg cgagggccgc ccctacgagg gcacccagac cgccaagctg 4440
aaggtgacca agggtggccc cctgcccttc gcctgggaca tcctgtcccc tcagttcatg 4500
tacggctcca aggcctacgt gaagcacccc gccgacatcc ccgactactt gaagctgtcc 4560
ttccccgagg gcttcaagtg ggagcgcgtg atgaacttcg aggacggcgg cgtggtgacc 4620
gtgacccagg actcctccct gcaggacggc gagttcatct acaaggtgaa gctgcgcggc 4680
accaacttcc cctccgacgg ccccgtaatg cagaagaaga ccatgggctg ggaggcctcc 4740
tccgagcgga tgtaccccga ggacggcgcc ctgaagggcg agatcaagca gaggctgaag 4800

ctgaaggacg gcggccacta cgacgctgag gtcaagacca cctacaaggc caagaagccc 4860

gtgcagctgc ccggcgccta caacgtcaac atcaagttgg acatcacctc ccacaacgag 4920

gactacacca tcgtggaaca gtacgaacgc gccgagggcc gccactccac cggcggcatg 4980

gacgagctgt acaagtaccc atacgatgtt ccagattacg cttaagaatt ctgcagatat 5040

ccagcacagt ggcggccgct cgagtctaga gggcccgttt aaacccgctg atcagcctcg 5100

actgtgcctt ctagttgcca gccatctgtt gtttgcccct cccccgtgcc ttccttgacc 5160

ctggaaggtg ccactcccac tgtcctttcc taataaaatg aggaaattgc atcgcattgt 5220

ctgagtaggt gtcattctat tctgggggggt ggggtggggc aggacagcaa gggggaggat 5280

tgggaagaca atagcaggca tgctggggat gcggtgggct ctatggcttc tgaggcggaa 5340

agaaccagct ggggctctag ggggtatccc cacgcgccct gtagcggcgc attaagcgcg 5400

gcgggtgtgg tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct 5460

cctttcgctt tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta 5520

aatcgggggc tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa 5580

cttgattagg gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct 5640

ttgacgttgg agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc 5700

aaccctatct cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg 5760

ttaaaaaatg agctgattta acaaaaattt aacgcgaatt aattctgtgg aatgtgtgtc 5820

agttagggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcatc 5880

tcaattagtc agcaaccagg tgtggaaagt ccccaggctc cccagcaggc agaagtatgc 5940

aaagcatgca tctcaattag tcagcaacca tagtcccgcc cctaactccg cccatcccgc 6000

ccctaactcc gcccagttcc gcccattctc cgccccatgg ctgactaatt ttttttattt 6060

atgcagaggc cgaggccgcc tctgcctctg agctattcca gaagtagtga ggaggctttt 6120

ttggaggcct aggcttttgc aaaaagctcc cgggagcttg tatatccatt ttcggatctg 6180

atcaagagac aggatgagga tcgtttcgca tgattgaaca agatggattg cacgcaggtt 6240

ctccggccgc ttgggtggag aggctattcg gctatgactg ggcacaacag acaatcggct 6300

gctctgatgc cgccgtgttc cggctgtcag cgcaggggcg cccggttctt tttgtcaaga 6360

ccgacctgtc cggtgccctg aatgaactgc aggacgaggc agcgcggcta tcgtggctgg 6420

ccacgacggg cgttccttgc gcagctgtgc tcgacgttgt cactgaagcg ggaagggact 6480

ggctgctatt gggcgaagtg ccggggcagg atctcctgtc atctc 6525

<210> 589

<211> 8924

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 589

agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 60

cgtttttcca taggctccgc cccccctgacg agcatcacaa aaatcgacgc tcaagtcaga 120

ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 180

tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 240

gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc 300

gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 360

gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 420

ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt 480

ggcctaacta cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag 540

ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 600

gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 660

ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 720

tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 780

ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 840

gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg 900

tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac 960

cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1020

ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc 1080

gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta 1140

caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1200

gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc 1260

ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac 1320

tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 1380

caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 1440

tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 1500

cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca 1560

ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa 1620

aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 1680

tcatactctt ccttttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 1740

gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc 1800

gaaaagtgcc acctgacgtc gacggatcgg gagatctccc gatcccctat ggtgcactct 1860

cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt 1920

ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag gcttgaccga 1980

caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc 2040

cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa ttacggggtc 2100

attagttcat agcccatata tggagttccg cgttacataa cttacggtaa atggcccgcc 2160

tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt 2220

aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt aaactgccca 2280

cttggcagta catcaagtgt atcatatgcc aagtacgccc cctattgacg tcaatgacgg 2340

taaatggccc gcctggcatt atgcccagta catgacctta tgggactttc ctacttggca 2400

gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc agtacatcaa 2460

tgggcgtgga tagcggtttg actcacgggg atttccaagt ctccacccca ttgacgtcaa 2520

tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta acaactccgc 2580

cccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa gcagagctct 2640

ctggctaact agagaaccca ctgcttactg gcttatcgaa attaatacga ctcactatag 2700

ggagacccaa gctggctagc gtttaaactt aagcttgcca ccatggcccc aaagaagaag 2760

cggaaggtcg gtatccacgg agtcccagca gccggatccg ccaccagatc cttcatcctg 2820

aagatcgagc ccaacgagga agtgaagaaa ggcctctgga aaacccacga ggtgctgaac 2880

cacggaatcg cctactacat gaatatcctg aagctgatcc ggcaagaggc catctacgag 2940

caccacgagc aggaccccaa gaatcccaag aaggtgtcca aggccgagat ccaggccgag 3000

ctgtgggatt tcgtgctgaa gatgcagaag tgcaacagct tcacacacga ggtggacaag 3060

gacgaggtgt tcaacatcct gagagagctg tacgaggaac tggtgcccag cagcgtggaa 3120

aagaagggcg aagccaacca gctgagcaac aagtttctgt accctctggt ggaccccaac 3180

agccagtctg gaaagggaac agccagcagc ggcagaaagc ccagatggta caacctgaag 3240

attgccggcg atccctcctg ggaagaagag aagaagaagt gggaagaaga taagaaaaag 3300

gacccgctgg ccaagatcct gggcaagctg gctgagtacg gactgatccc tctgttcatc 3360

ccctacaccg acagcaacga gcccatcgtg aaagaaatca agtggatgga aaagtcccgg 3420

aaccagagcg tgcggcggct ggataaggac atgttcattc aggccctgga acggttcctg 3480

agctgggaga gctggaacct gaaagtgaaa gaggaatacg agaaggtcga gaaagagtac 3540

aagaccctgg aagagaggat caaagaggac atccaggctc tgaaggctct ggaacagtat 3600

gagaaagagc ggcaagaaca gctgctgcgg gacaccctga acaccaacga gtaccggctg 3660

agcaagagag gccttagagg ctggcgggaa atcatccaga aatggctgaa aatggacgag 3720

aacgagccct ccgagaagta cctggaagtg ttcaaggact accagcggaa gcaccctaga 3780

gaggccggcg attacagcgt gtacgagttc ctgtccaaga aagagaacca cttcatctgg 3840

cggaatcacc ctgagtaccc ctacctgtac gccaccttct gcgagatcga caagaaaaag 3900

aaggacgcca agcagcaggc caccttcaca ctggccgatc ctatcaatca ccctctgtgg 3960

gtccgattcg aggaaagaag cggcagcaac ctgaacaagt acagaatcct gaccgagcag 4020

ctgcacaccg agaagctgaa gaaaaagctg acagtgcagc tggaccggct gatctaccct 4080

acagaatctg gcggctggga agagaagggc aaagtggaca ttgtgctgct gcccagccgg 4140

cagttctaca accagatctt cctggacatc gaggaaaagg gcaagcacgc cttcacctac 4200

aaggatgaga gcatcaagtt ccctctgaag ggcacactcg gcggagccag agtgcagttc 4260

gacagagatc acctgagaag ataccctcac aaggtggaaa gcggcaacgt gggcagaatc 4320

tacttcaaca tgaccgtgaa catcgagcct acagagtccc cagtgtccaa gtctctgaag 4380

atccaccggg acgacttccc caaggtggtc aacttcaagc ccaaagaact gaccgagtgg 4440

atcaaggaca gcaagggcaa gaaactgaag tccggcatcg agtccctgga aatcggcctg 4500

agagtgatga gcatcgacct gggacagaga caggccgctg ccgcctctat tttcgaggtg 4560

gtggatcaga agcccgacat cgaaggcaag ctgtttttcc caatcaaggg caccgagctg 4620

tatgccgtgc acagagccag cttcaacatc aagctgcccg gcgagacact ggtcaagagc 4680

agagaagtgc tgcggaaggc cagagaggac aatctgaaac tgatgaacca gaagctcaac 4740

ttcctgcgga acgtgctgca cttccagcag ttcgaggaca tcaccgagag agagaagcgg 4800

gtcaccaagt ggatcagcag acaagagaac agcgacgtgc ccctggtgta ccaggatgag 4860

ctgatccaga tccgcgagct gatgtacaag ccttacaagg actgggtcgc cttcctgaag 4920

cagctccaca agagactgga agtcgagatc ggcaaagaag tgaagcactg gcggaagtcc 4980

ctgagcgacg gaagaaaggg cctgtacggc atctccctga agaacatcga cgagatcgat 5040

cggacccgga agttcctgct gagatggtcc ctgaggccta ccgaacctgg cgaagtgcgt 5100

agactggaac ccggccagag attcgccatc gaccagctga atcacctgaa cgccctgaaa 5160

gaagatcggc tgaagaagat ggccaacacc atcatcatgc acgccctggg ctactgctac 5220

gacgtgcgga agaagaaatg gcaggctaag aaccccgcct gccagatcat cctgttcgag 5280

gatctgagca actacaaccc ctacggagaa aggtcccgct tcgagaacag ccggctcatg 5340

aagtggtcca gacgcgagat ccccagacag gttgcactgc agggcgagat ctatggcctg 5400
caagtgggag aagtgggcgc tcagttcagc agcagattcc acgccaagac aggcagccct 5460
ggcatcagat gtcgggtcgt gaccaaagag aagctgcagg acaatcggtt cttcaagaat 5520
ctgcagagag agggcagact gaccctggac aaaatcgccg tgctgaaaga gggcgatctg 5580
tacccagaca aaggcggcga gaagttcatc agcctgagca aggatcggaa gtgcgtgacc 5640
acacacgccg acatcaacgc cgctcagaac ctgcagaagc ggttctggac aagaacccac 5700
ggcttctaca aggtgtactg caaggcctac caggtggacg gccagaccgt gtacatccct 5760

gagagcaagg accagaagca gaagatcatc gaagagttcg gcgagggcta cttcattctg 5820
aaggacgggg tgtacgaatg ggtcaacgcc ggcaagctga aaatcaagaa gggcagctcc 5880
aagcagagca gcagcgagct ggtggatagc gacatcctga aagacagctt cgacctggcc 5940
tccgagctga aaggcgaaaa gctgatgctg tacagggacc ccagcggcaa tgtgttcccc 6000
agcgacaaat ggatggccgc tggcgtgttc ttcggaaagc tggaacgcat cctgatcagc 6060
aagctgacca accagtactc catcagcacc atcgaggacg acagcagcaa gcagtctatg 6120
ggatccaaaa ggccggcggc cacgaaaaag gccggccagg caaaaaagaa aaagggatct 6180

tacccatacg atgttccaga ttacgcttat ccctacgacg tgcctgatta tgcataccca 6240
tatgatgtcc ccgactatgc ctaagaattc tgcagatatc cagcacagtg gcggccgctc 6300
gagtctagag ggcccgttta aacccgctga tcagcctcga ctgtgccttc tagttgccag 6360
ccatctgttg tttgcccctc ccccgtgcct tccttgaccc tggaaggtgc cactcccact 6420
gtcctttcct aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt 6480
ctgggggggtg gggtggggca ggacagcaag ggggaggatt gggaagacaa tagcaggcat 6540
gctggggatg cggtgggctc tatggcttct gaggcggaaa gaaccagctg gggctctagg 6600

gggtatcccc acgcgccctg tagcggcgca ttaagcgcgg cgggtgtggt ggttacgcgc 6660
agcgtgaccg ctacacttgc cagcgcccta gcgcccgctc ctttcgcttt cttcccttcc 6720
tttctcgcca cgttcgccgg ctttccccgt caagctctaa atcgggggct ccctttaggg 6780
ttccgattta gtgctttacg gcacctcgac cccaaaaaac ttgattaggg tgatggttca 6840
cgtagtgggc catcgccctg atagacggtt tttcgccctt tgacgttgga gtccacgttc 6900
tttaatagtg gactcttgtt ccaaactgga acaacactca accctatctc ggtctattct 6960
tttgatttat aagggatttt gccgatttcg gcctattggt taaaaaatga gctgatttaa 7020

caaaaattta acgcgaatta attctgtgga atgtgtgtca gttagggtgt ggaaagtccc 7080

caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccaggt 7140

gtggaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 7200

cagcaaccat agtcccgccc ctaactccgc ccatcccgcc cctaactccg cccagttccg 7260

cccattctcc gccccatggc tgactaattt tttttattta tgcagaggcc gaggccgcct 7320

ctgcctctga gctattccag aagtagtgag gaggcttttt tggaggccta ggcttttgca 7380

aaaagctccc gggagcttgt atatccattt tcggatctga tcaagagaca ggatgaggat 7440

cgtttcgcat gattgaacaa gatggattgc acgcaggttc tccggccgct tgggtggaga 7500

ggctattcgg ctatgactgg gcacaacaga caatcggctg ctctgatgcc gccgtgttcc 7560

ggctgtcagc gcaggggcgc ccggttcttt ttgtcaagac cgacctgtcc ggtgccctga 7620

atgaactgca ggacgaggca gcgcggctat cgtggctggc cacgacgggc gttccttgcg 7680

cagctgtgct cgacgttgtc actgaagcgg gaagggactg gctgctattg ggcgaagtgc 7740

cggggcagga tctcctgtca tctcaccttg ctcctgccga gaaagtatcc atcatggctg 7800

atgcaatgcg gcggctgcat acgcttgatc cggctacctg cccattcgac caccaagcga 7860

aacatcgcat cgagcgagca cgtactcgga tggaagccgg tcttgtcgat caggatgatc 7920

tggacgaaga gcatcagggg ctcgcgccag ccgaactgtt cgccaggctc aaggcgcgca 7980

tgcccgacgg cgaggatctc gtcgtgaccc atggcgatgc ctgcttgccg aatatcatgg 8040

tggaaaatgg ccgcttttct ggattcatcg actgtggccg gctgggtgtg gcggaccgct 8100

atcaggacat agcgttggct acccgtgata ttgctgaaga gcttggcggc gaatgggctg 8160

accgcttcct cgtgctttac ggtatcgccg ctcccgattc gcagcgcatc gccttctatc 8220

gccttcttga cgagttcttc tgagcgggac tctggggttc gaaatgaccg accaagcgac 8280

gcccaacctg ccatcacgag atttcgattc caccgccgcc ttctatgaaa ggttgggctt 8340

cggaatcgtt ttccgggacg ccggctggat gatcctccag cgcggggatc tcatgctgga 8400

gttcttcgcc caccccaact tgtttattgc agcttataat ggttacaaat aaagcaatag 8460

catcacaaat ttcacaaata aagcattttt ttcactgcat tctagttgtg gtttgtccaa 8520

actcatcaat gtatcttatc atgtctgtat accgtcgacc tctagctaga gcttggcgta 8580

atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat 8640

acgagccgga agcataaagt gtaaagcctg gggtgcctaa tgagtgagct aactcacatt 8700

aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta 8760

atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc 8820

gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa 8880

ggcggtaata cggttatcca cagaatcagg ggataacgca ggaa 8924

<210> 590

<211> 8924

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 590

agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 60

cgtttttcca taggctccgc cccccctgacg agcatcacaa aaatcgacgc tcaagtcaga 120

ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 180

tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 240

gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc 300

gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 360

gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 420

ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggt 480

ggcctaacta cggctacact agaagaacag tatttggtat ctgcgctctg ctgaagccag 540

ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 600

gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 660

ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 720

tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 780

ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 840

gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg 900

tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct gcaatgatac 960

cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1020

ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt aattgttgcc 1080

gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt gccattgcta 1140

caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1200

gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc 1260

ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt atggcagcac 1320

tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 1380

caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 1440

tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 1500

cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaaccca 1560

ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa 1620

aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 1680

tcatactctt ccttttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 1740

gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc 1800

gaaaagtgcc acctgacgtc gacggatcgg gagatctccc gatcccctat ggtgcactct 1860

cagtacaatc tgctctgatg ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt 1920

ggaggtcgct gagtagtgcg cgagcaaaat ttaagctaca acaaggcaag gcttgaccga 1980

caattgcatg aagaatctgc ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc 2040

cagatatacg cgttgacatt gattattgac tagttattaa tagtaatcaa ttacggggtc 2100

attagttcat agcccatata tggagttccg cgttacataa cttacggtaa atggcccgcc 2160

tggctgaccg cccaacgacc cccgcccatt gacgtcaata atgacgtatg ttcccatagt 2220

aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt aaactgccca 2280

cttggcagta catcaagtgt atcatatgcc aagtacgccc cctattgacg tcaatgacgg 2340

taaatggccc gcctggcatt atgcccagta catgacctta tgggactttc ctacttggca 2400

gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc agtacatcaa 2460

tgggcgtgga tagcggtttg actcacgggg atttccaagt ctccacccca ttgacgtcaa 2520

tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta acaactccgc 2580

cccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa gcagagctct 2640

ctggctaact agagaaccca ctgcttactg gcttatcgaa attaatacga ctcactatag 2700

ggagacccaa gctggctagc gtttaaactt aagcttgcca ccatggcccc aaagaagaag 2760

cggaaggtcg gtatccacgg agtcccagca gccggatccg ccaccagatc cttcatcctg 2820

aagatcgagc ccaacgagga agtgaagaaa ggcctctgga aaacccacga ggtgctgaac 2880

cacggaatcg cctactacat gaatatcctg aagctgatcc ggcaagaggc catctacgag 2940

caccacgagc aggaccccaa gaatcccaag aaggtgtcca aggccgagat ccaggccgag 3000

ctgtgggatt tcgtgctgaa gatgcagaag tgcaacagct tcacacacga ggtggacaag 3060

gacgaggtgt tcaacatcct gagagagctg tacgaggaac tggtgcccag cagcgtggaa 3120

aagaagggcg aagccaacca gctgagcaac aagtttctgt accctctggt ggaccccaac 3180

agccagtctg gaaagggaac agccagcagc ggcagaaagc ccagatggta caacctgaag 3240

attgccggcg atccctcctg ggaagaagag aagaagaagt gggaagaaga taagaaaaag 3300

gacccgctgg ccaagatcct gggcaagctg gctgagtacg gactgatccc tctgttcatc 3360

ccctacaccg acagcaacga gcccatcgtg aaagaaatca agtggatgga aaagtcccgg 3420

aaccagagcg tgcggcggct ggataaggac atgttcattc aggccctgga acggttcctg 3480

agctgggaga gctggaacct gaaagtgaaa gaggaatacg agaaggtcga gaaagagtac 3540

aagaccctgg aagagaggat caaagaggac atccaggctc tgaaggctct ggaacagtat 3600

gagaaagagc ggcaagaaca gctgctgcgg gacaccctga acaccaacga gtaccggctg 3660

agcaagagag gccttagagg ctggcgggaa atcatccaga aatggctgaa aatggacgag 3720

aacgagccct ccgagaagta cctggaagtg ttcaaggact accagcggaa gcaccctaga 3780

gaggccggcg attacagcgt gtacgagttc ctgtccaaga aagagaacca cttcatctgg 3840

cggaatcacc ctgagtaccc ctacctgtac gccaccttct gcgagatcga caagaaaaag 3900

aaggacgcca agcagcaggc caccttcaca ctggccgatc ctatcaatca ccctctgtgg 3960

gtccgattcg aggaaagaag cggcagcaac ctgaacaagt acagaatcct gaccgagcag 4020

ctgcacaccg agaagctgaa gaaaaagctg acagtgcagc tggaccggct gatctaccct 4080

acagaatctg gcggctggga agagaagggc aaagtggaca ttgtgctgct gcccagccgg 4140

cagttctaca accagatctt cctggacatc gaggaaaagg gcaagcacgc cttcacctac 4200

aaggatgaga gcatcaagtt ccctctgaag ggcacactcg gcggagccag agtgcagttc 4260

gacagagatc acctgagaag ataccctcac aaggtggaaa gcggcaacgt gggcagaatc 4320

tacttcaaca tgaccgtgaa catcgagcct acagagtccc cagtgtccaa gtctctgaag 4380

atccaccggg acgacttccc caaggtggtc aacttcaagc ccaaagaact gaccgagtgg 4440

atcaaggaca gcaagggcaa gaaactgaag tccggcatcg agtccctgga aatcggcctg 4500

agagtgatga gcatcgacct gggacagaga caggccgctg ccgcctctat tttcgaggtg 4560

gtggatcaga agcccgacat cgaaggcaag ctgtttttcc caatcaaggg caccgagctg 4620

tatgccgtgc acagagccag cttcaacatc aagctgcccg gcgagacact ggtcaagagc 4680

agagaagtgc tgcggaaggc cagagaggac aatctgaaac tgatgaacca gaagctcaac 4740

ttcctgcgga acgtgctgca cttccagcag ttcgaggaca tcaccgagag agagaagcgg 4800

gtcaccaagt ggatcagcag acaagagaac agcgacgtgc ccctggtgta ccaggatgag 4860

ctgatccaga tccgcgagct gatgtacaag ccttacaagg actgggtcgc cttcctgaag 4920

cagctccaca agagactgga agtcgagatc ggcaaagaag tgaagcactg gcggaagtcc 4980

ctgagcgacg gaagaaaggg cctgtacggc atctccctga agaacatcga cgagatcgat 5040

cggacccgga agttcctgct gagatggtcc ctgaggccta ccgaacctgg cgaagtgcgt 5100

agactggaac ccggccagag attcgccatc gaccagctga atcacctgaa cgccctgaaa 5160

gaagatcggc tgaagaagat ggccaacacc atcatcatgc acgccctggg ctactgctac 5220
gacgtgcgga agaagaaatg gcaggctaag aaccccgcct gccagatcat cctgttcgag 5280
gatctgagca actacaaccc ctacaaggaa aggtcccgct tcgagaacag ccggctcatg 5340
aagtggtcca gacgcgagat ccccagacag gttgcactgc agggcgagat ctatggcctg 5400
caagtgggag aagtgggcgc tcagttcagc agcagattcc acgccaagac aggcagccct 5460
ggcatcagat gtcgggtcgt gaccaaagag aagctgcagg acaatcggtt cttcaagaat 5520
ctgcagagag agggcagact gaccctggac aaaatcgccg tgctgaaaga gggcgatctg 5580

tacccagaca aaggcggcga gaagttcatc agcctgagca aggatcggaa gtgcgtgacc 5640
acacacgccg acatcaacgc cgctcagaac ctgcagaagc ggttctggac aagaacccac 5700
ggcttctaca aggtgtactg caaggcctac caggtggacg gccagaccgt gtacatccct 5760
gagagcaagg accagaagca gaagatcatc gaagagttcg gcgagggcta cttcattctg 5820
aaggacgggg tgtacgaatg ggtcaacgcc ggcaagctga aaatcaagaa gggcagctcc 5880
aagcagagca gcagcgagct ggtggatagc gacatcctga aagacagctt cgacctggcc 5940
tccgagctga aaggcgaaaa gctgatgctg tacagggacc ccagcggcaa tgtgttcccc 6000

agcgacaaat ggatggccgc tggcgtgttc ttcggaaagc tggaacgcat cctgatcagc 6060
aagctgacca accagtactc catcagcacc atcgaggacg acagcagcaa gcagtctatg 6120
ggatccaaaa ggccggcggc cacgaaaaag gccggccagg caaaaaagaa aaagggatct 6180
tacccatacg atgttccaga ttacgcttat ccctacgacg tgcctgatta tgcataccca 6240
tatgatgtcc ccgactatgc ctaagaattc tgcagatatc cagcacagtg gcggccgctc 6300
gagtctagag ggcccgttta aacccgctga tcagcctcga ctgtgccttc tagttgccag 6360
ccatctgttg tttgcccctc ccccgtgcct tccttgaccc tggaaggtgc cactcccact 6420

gtcctttcct aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt 6480
ctgggggggtg gggtggggca ggacagcaag ggggaggatt gggaagacaa tagcaggcat 6540
gctggggatg cggtgggctc tatggcttct gaggcggaaa gaaccagctg gggctctagg 6600
gggtatcccc acgcgccctg tagcggcgca ttaagcgcgg cgggtgtggt ggttacgcgc 6660
agcgtgaccg ctacacttgc cagcgcccta gcgcccgctc ctttcgcttt cttcccttcc 6720
tttctcgcca cgttcgccgg ctttccccgt caagctctaa atcgggggct ccctttaggg 6780
ttccgattta gtgctttacg gcacctcgac cccaaaaaac ttgattaggg tgatggttca 6840

cgtagtgggc catcgccctg atagacggtt tttcgccctt tgacgttgga gtccacgttc 6900

tttaatagtg gactcttgtt ccaaactgga acaacactca accctatctc ggtctattct 6960

tttgatttat aagggatttt gccgatttcg gcctattggt taaaaaatga gctgatttaa 7020

caaaaattta acgcgaatta attctgtgga atgtgtgtca gttagggtgt ggaaagtccc 7080

caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccaggt 7140

gtggaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 7200

cagcaaccat agtcccgccc ctaactccgc ccatcccgcc cctaactccg cccagttccg 7260

cccattctcc gccccatggc tgactaattt tttttattta tgcagaggcc gaggccgcct 7320

ctgcctctga gctattccag aagtagtgag gaggcttttt tggaggccta ggcttttgca 7380

aaaagctccc gggagcttgt atatccattt tcggatctga tcaagagaca ggatgaggat 7440

cgtttcgcat gattgaacaa gatggattgc acgcaggttc tccggccgct tgggtggaga 7500

ggctattcgg ctatgactgg gcacaacaga caatcggctg ctctgatgcc gccgtgttcc 7560

ggctgtcagc gcaggggcgc ccggttcttt ttgtcaagac cgacctgtcc ggtgccctga 7620

atgaactgca ggacgaggca gcgcggctat cgtggctggc cacgacgggc gttccttgcg 7680

cagctgtgct cgacgttgtc actgaagcgg gaagggactg gctgctattg ggcgaagtgc 7740

cggggcagga tctcctgtca tctcaccttg ctcctgccga gaaagtatcc atcatggctg 7800

atgcaatgcg gcggctgcat acgcttgatc cggctacctg cccattcgac caccaagcga 7860

aacatcgcat cgagcgagca cgtactcgga tggaagccgg tcttgtcgat caggatgatc 7920

tggacgaaga gcatcagggg ctcgcgccag ccgaactgtt cgccaggctc aaggcgcgca 7980

tgcccgacgg cgaggatctc gtcgtgaccc atggcgatgc ctgcttgccg aatatcatgg 8040

tggaaaatgg ccgcttttct ggattcatcg actgtggccg gctgggtgtg gcggaccgct 8100

atcaggacat agcgttggct acccgtgata ttgctgaaga gcttggcggc gaatgggctg 8160

accgcttcct cgtgctttac ggtatcgccg ctcccgattc gcagcgcatc gccttctatc 8220

gccttcttga cgagttcttc tgagcgggac tctggggttc gaaatgaccg accaagcgac 8280

gcccaacctg ccatcacgag atttcgattc caccgccgcc ttctatgaaa ggttgggctt 8340

cggaatcgtt ttccgggacg ccggctggat gatcctccag cgcggggatc tcatgctgga 8400

gttcttcgcc caccccaact tgtttattgc agcttataat ggttacaaat aaagcaatag 8460

catcacaaat ttcacaaata aagcattttt ttcactgcat tctagttgtg gtttgtccaa 8520

actcatcaat gtatcttatc atgtctgtat accgtcgacc tctagctaga gcttggcgta 8580

atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat 8640

acgagccgga agcataaagt gtaaagcctg gggtgcctaa tgagtgagct aactcacatt 8700

aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta 8760

atgaatcggc caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc 8820

gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa 8880

ggcggtaata cggttatcca cagaatcagg ggataacgca ggaa 8924

<210> 591

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 591

gagaagtcat ttaataaggc cactgttaaa a 31

<210> 592

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 592

agacccagag gctcaagtga gca 23

<210> 593

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 593

agctgaaggg aaataaaagg aaa 23

<210> 594

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 594

tctcccctgg gaagcatccc tgg 23

<210> 595

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 595

tcatggagaa aatattcaga atc 23

<210> 596

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 596

cctcactcct gctcggtgaa ttt 23

<210> 597

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 597

aggagtgttc agtctccgtg aac 23

<210> 598

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 598

ggaggtcaga aatagggggt cca 23

<210> 599

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 599

caaagcccat tccctcttta gcc 23

<210> 600

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 600

cccttcagct aaaataaagg agg 23

<210> 601

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 601

ttctcccctg ggaagcatcc ctg 23

<210> 602

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 602

tgcagagcaa ataccagaga taa 23

<210> 603

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 603

cgccgggccc tgaccacgct cat 23

<210> 604

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 604

ccgacttcat ctttgccaac gtc 23

<210> 605

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 605

tagagcactg gcatggggat ggg 23

<210> 606

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 606

ttgctccaga ggcccccctt ggg 23

<210> 607

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 607

ctggtgccag aaacaggggt gac 23

<210> 608

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 608

tgggcttcaa gcaacttgta gtg 23

<210> 609

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 609

tgtaattggt tctaccaaag aag 23

<210> 610

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 610

agaggcggag ggcggcgtgc ctg 23

<210> 611

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 611

cttcagccca agaacagtac aag 23

<210> 612

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 612

tctgtgagtc gaggagaaac gac 23

<210> 613

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 613

cttgggtgtg ttaaaagtga cca 23

<210> 614

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 614

tcactcctgc tcggtgaatt 20

<210> 615

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 615

gctacaggca gagacaaagg 20

<210> 616

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 616

aggtcagaaa taggggtcc 20

<210> 617

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 617

caggctgtga accttggtgg 20

<210> 618

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 618

gacccctcc accccgcctc 20

<210> 619

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 619

gtatctagcc tcttctaaga c 21

<210> 620

<211> 20

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 620

tctcccctgg gaagcatccc 20

<210> 621

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 621

gagtccgagc agaagaagaa 20

<210> 622

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 622

ttttgggagt aagaaaaggt 20

<210> 623

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 623

agtgtccagg gatgcttccc 20

<210> 624

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 624

cctcactcct gctcggtgaa ttt 23

<210> 625

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 625

ggaggtcaga aatagggggt cca 23

<210> 626

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 626

gatggcgact tcaggcacag gat 23

<210> 627

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 627

ggaagtgtcc agggatgctt ccc 23

<210> 628

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 628

cccttcagct aaaataaagg agg 23

<210> 629

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 629

ggctcagcag gcacctgcct cag 23

<210> 630

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 630

gggactggag ttgcttcatg tac 23

<210> 631

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 631

tctccatgaa aaatactggg gtc 23

<210> 632

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 632

ttcatggaga aaatattcag aat 23

<210> 633

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 633

gcagctacag gcagagacaa agg 23

<210> 634

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 634

cctggaaacc atccaggcct tgt 23

<210> 635

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 635

ggtcagctgt taacatcagt acg 23

<210> 636

<211> 23

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 636

tcttcacgga aacagggttc ctt 23

<210> 637

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 637

tggttgccca ccctagtcat tgg 23

<210> 638

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 638

gatggcgact tcaggcacag gat 23

<210> 639

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 639

cctcactcct gctcggtgaa ttt 23

<210> 640

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 640

ggaagtgtcc agggatgctt ccc 23

<210> 641

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 641

ggctcagcag gcacctgcct cag 23

<210> 642

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 642

ttcatggaga aaatattcag aat 23

<210> 643

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 643

ctgacctccc aaacagctac ata 23

<210> 644

<211> 47

<212> DNA

<213

> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (1)..(1)

<223> 5' posphorylated

<400> 644

gttgtgagca agggcgagga ggataacgcc tctctcccag cgactat 47

<210> 645

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<220><221> misc_feature

<222> (1)..(1)

<223> 5' phosphorylate

<400> 645

atagtcgctg ggagagaggc gttatcctcc tcgcccttgc tcacaac 47

<210> 646

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 646

tactttagag cgaaaggctt ttc 23

<210> 647

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 647

agataccaaa gagaacggga tca 23

<210> 648

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 648

tgacttcact gtaactggga gag 23

<210> 649

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 649

tctcccagga agggcagcag gct 23

<210> 650

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 650

tagacgtgga gatcattttt aac 23

<210> 651

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 651

aataaaatat ctttattttc attacatctg tgtgttggtt ttttgtgtg 49

<210> 652

<211> 208

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 652

ctgtgccttc tagttgccag ccatctgttg tttgcccctc ccccgtgcct tccttgaccc 60

tggaaggtgc cactcccact gtcctttcct aataaaatga ggaaattgca tcgcattgtc 120

tgagtaggtg tcattctatt ctgggggggtg gggtggggca ggacagcaag ggggaggatt 180

gggaagagaa tagcaggcat gctgggga 208

<210> 653

<211> 1190

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Peptide

<400> 653

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Val Lys Ser Met Lys Val Lys Leu Arg Leu Asp Asn Met Pro
20 25 30

Glu Ile Arg Ala Gly Leu Trp Lys Leu His Thr Glu Val Asn Ala Gly
35 40 45

Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu

50 55 60

Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Tyr Lys Thr
65 70 75 80

Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln

85 90 95

Val Glu Asn Gly His Cys Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu

100 105 110

Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly

115 120 125

Ala Lys Gly Asp Ala Gln Gln Ile Ala Arg Lys Phe Leu Ser Pro Leu

130 135 140

Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn

145 150 155 160

Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu

165 170 175

Glu Glu Lys Ala Lys Ala Glu Ala Arg Lys Ser Thr Asp Arg Thr Ala

180 185 190

Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg

195 200 205

Val Tyr Thr Asp Ser Asp Met Ser Ser Val Gln Trp Lys Pro Leu Arg

210 215 220

Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala

225 230 235 240

Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Glu

245 250 255

Ala Tyr Ala Lys Leu Val Glu Gln Lys Ser Arg Phe Glu Gln Lys Asn

260 265 270

Phe Val Gly Gln Glu His Leu Val Gln Leu Val Asn Gln Leu Gln Gln

275 280 285

Asp Met Lys Glu Ala Ser His Gly Leu Glu Ser Lys Glu Gln Thr Ala

290 295 300

His Tyr Leu Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu

305 310 315 320

Lys Trp Glu Lys Leu Asp Pro Asp Ala Pro Phe Asp Leu Tyr Asp Thr

325 330 335

Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His
340 345 350
Asp Leu Phe Ala Lys Leu Ala Glu Pro Lys Tyr Gln Ala Leu Trp Arg
355 360 365
Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Val

370 375 380
Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp
385 390 395 400
Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn
405 410 415
Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Gly Arg His
420 425 430
Ala Ile Arg Phe Gln Lys Leu Leu Thr Val Glu Asp Gly Val Ala Lys

435 440 445
Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Ala Gln Leu Asp
450 455 460
Asp Leu Leu Pro Arg Asp Pro His Glu Leu Val Ala Leu Tyr Phe Gln
465 470 475 480
Asp Tyr Gly Ala Glu Gln His Leu Ala Gly Glu Phe Gly Gly Ala Lys
485 490 495
Ile Gln Tyr Arg Arg Asp Gln Leu Asn His Leu His Ala Arg Arg Gly

500 505 510
Ala Arg Asp Val Tyr Leu Asn Leu Ser Val Arg Val Gln Ser Gln Ser
515 520 525
Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu
530 535 540
Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp
545 550 555 560
Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu

565 570 575
Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser

580 585 590

Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro

595 600 605

Asn Ser Glu Gly Arg Val Pro Phe Cys Phe Pro Ile Glu Gly Asn Glu

610 615 620

Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly

625 630 635 640

Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg

645 650 655

Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val

660 665 670

Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys

675 680 685

Leu Ile Glu Gln Pro Met Asp Ala Asn Gln Met Thr Pro Asp Trp Arg

690 695 700

Glu Ala Phe Glu Asp Glu Leu Gln Lys Leu Lys Ser Leu Tyr Gly Ile

705 710 715 720

Cys Gly Asp Arg Glu Trp Thr Glu Ala Val Tyr Glu Ser Val Arg Arg

725 730 735

Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp Val

740 745 750

Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Gln Lys Asp Val Val

755 760 765

Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr Lys

770 775 780

Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val Ile

785 790 795 800

Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His Ile

805 810 815

Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile Ile

820 825 830

Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Asp Glu Arg Gly Lys
835 840 845
Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu
850 855 860
Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn
865 870 875 880
Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Leu

885 890 895
Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala
900 905 910
Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys
915 920 925
Arg Arg Val Pro Ala Arg Cys Ala Arg Glu Gln Asn Pro Glu Pro Phe
930 935 940
Pro Trp Trp Leu Asn Lys Phe Val Ala Glu His Lys Leu Asp Gly Cys

945 950 955 960
Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Phe Phe
965 970 975
Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala
980 985 990
Asp Leu Asn Ala Ala Gln Asn Leu Gln Arg Arg Leu Trp Ser Asp Phe
995 1000 1005
Asp Ile Ser Gln Ile Arg Leu Arg Cys Asp Trp Gly Glu Val Asp

1010 1015 1020
Gly Glu Pro Val Leu Ile Pro Arg Thr Thr Gly Lys Arg Thr Ala
1025 1030 1035
Asp Ser Tyr Gly Asn Lys Val Phe Tyr Thr Lys Thr Gly Val Thr
1040 1045 1050
Tyr Tyr Glu Arg Glu Arg Gly Lys Lys Arg Arg Lys Val Phe Ala
1055 1060 1065
Gln Glu Glu Leu Ser Glu Glu Glu Ala Glu Leu Leu Val Glu Ala

1070 1075 1080

Asp Glu Ala Arg Glu Lys Ser Val Val Leu Met Arg Asp Pro Ser

1085 1090 1095

Gly Ile Ile Asn Arg Gly Asp Trp Thr Arg Gln Lys Glu Phe Trp

1100 1105 1110

Ser Met Val Asn Gln Arg Ile Glu Gly Tyr Leu Val Lys Gln Ile

1115 1120 1125

Arg Ser Arg Val Arg Leu Gln Glu Ser Ala Cys Glu Asn Thr Gly

1130 1135 1140

Asp Ile Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys

1145 1150 1155

Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr

1160 1165 1170

Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp

1175 1180 1185

Tyr Ala

1190

<210> 654

<211> 7584

<212

> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 654

cctgcaggca gctgcgcgct cgctcgctca ctgaggccgc ccgggcgtcg ggcgacctttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct gcggcctcta gactcgaggg gctggaagct acctttgaca tcatttcctc 180

tgcgaatgca tgtataattt ctacagaacc tattagaaag gatcacccag cctctgcttt 240

tgtacaactt tcccttaaaa aactgccaat tccactgctg tttggcccaa tagtgagaac 300

tttttcctgc tgcctcttgg tgcttttgcc tatggcccct attctgcctg ctgaagacac 360

tcttgccagc atggacttaa acccctccag ctctgacaat cctctttctc ttttgtttta 420

catgaagggt ctggcagcca aagcaatcac tcaaagttca aaccttatca ttttttgctt 480

tgttcctctt ggccttggtt ttgtacatca gctttgaaaa taccatccca gggttaatgc 540

tggggttaat ttataactaa gagtgctcta gttttgcaat acaggacatg ctataaaaat 600

ggaaagatac cggtgccaca agaggtaagg gtttaaggga tggttggttg gtggggtatt 660

aatgtttaat tacctggagc acctgcctga aatcactttt tttcaggttg gcatggcccc 720

aaagaagaag cggaaggtcg gtatccacgg agtcccagca gccgccacca gatccttcat 780

cctgaagatc gagcccaacg aggaagtgaa gaaaggcctc tggaaaaccc acgaggtgct 840

gaaccacgga atcgcctact acatgaatat cctgaagctg atccggcaag aggccatcta 900

cgagcaccac gagcaggacc ccaagaatcc caagaaggtg tccaaggccg agatccaggc 960

cgagctgtgg gatttcgtgc tgaagatgca gaagtgcaac agcttcacac acgaggtgga 1020

caaggacgag gtgttcaaca tcctgagaga gctgtacgag gaactggtgc ccagcagcgt 1080

ggaaaagaag ggcgaagcca accagctgag caacaagttt ctgtaccctc tggtggaccc 1140

caacagccag tctggaaagg gaacagccag cagcggcaga aagcccagat ggtacaacct 1200

gaagattgcc ggcgatccct cctgggaaga agagaagaag aagtgggaag aagataagaa 1260

aaaggacccg ctggccaaga tcctgggcaa gctggctgag tacggactga tccctctgtt 1320

catcccctac accgacagca acgagcccat cgtgaaagaa atcaagtgga tggaaaagtc 1380

ccggaaccag agcgtgcggc ggctggataa ggacatgttc attcaggccc tggaacggtt 1440

cctgagctgg gagagctgga acctgaaagt gaaagaggaa tacgagaagg tcgagaaaga 1500

gtacaagacc ctggaagaga ggatcaaaga ggacatccag gctctgaagg ctctggaaca 1560

gtatgagaaa gagcggcaag aacagctgct gcgggacacc ctgaacacca acgagtaccg 1620

gctgagcaag agaggcctta gaggctggcg ggaaatcatc cagaaatggc tgaaaatgga 1680

cgagaacgag ccctccgaga agtacctgga agtgttcaag gactaccagc ggaagcaccc 1740

tagagaggcc ggcgattaca gcgtgtacga gttcctgtcc aagaaagaga accacttcat 1800

ctggcggaat caccctgagt acccctacct gtacgccacc ttctgcgaga tcgacaagaa 1860

aaagaaggac gccaagcagc aggccacctt cacactggcc gatcctatca atcaccctct 1920

gtgggtccga ttcgaggaaa gaagcggcag caacctgaac aagtacagaa tcctgaccga 1980

gcagctgcac accgagaagc tgaagaaaaa gctgacagtg cagctggacc ggctgatcta 2040

ccctacagaa tctggcggct gggaagagaa gggcaaagtg gacattgtgc tgctgcccag 2100

ccggcagttc tacaaccaga tcttcctgga catcgaggaa aagggcaagc acgccttcac 2160

ctacaaggat gagagcatca agttccctct gaagggcaca ctcggcggag ccagagtgca 2220

gttcgacaga gatcacctga gaagataccc tcacaaggtg gaaagcggca acgtgggcag 2280

aatctacttc aacatgaccg tgaacatcga gcctacagag tccccagtgt ccaagtctct 2340

gaagatccac cgggacgact tccccaaggt ggtcaacttc aagcccaaag aactgaccga 2400

gtggatcaag gacagcaagg gcaagaaact gaagtccggc atcgagtccc tggaaatcgg 2460

cctgagagtg atgagcatcg acctgggaca gagacaggcc gctgccgcct ctattttcga 2520
ggtggtggat cagaagcccg acatcgaagg caagctgttt ttcccaatca agggcaccga 2580
gctgtatgcc gtgcacagag ccagcttcaa catcaagctg cccggcgaga cactggtcaa 2640
gagcagagaa gtgctgcgga aggccagaga ggacaatctg aaactgatga accagaagct 2700
caacttcctg cggaacgtgc tgcacttcca gcagttcgag gacatcaccg agagagagaa 2760
gcgggtcacc aagtggatca gcagacaaga gaacagcgac gtgcccctgg tgtaccagga 2820
tgagctgatc cagatccgcg agctgatgta caagccttac aaggactggg tcgccttcct 2880

gaagcagctc cacaagagac tggaagtcga gatcggcaaa gaagtgaagc actggcggaa 2940
gtccctgagc gacggaagaa agggcctgta cggcatctcc ctgaagaaca tcgacgagat 3000
cgatcggacc cggaagttcc tgctgagatg gtccctgagg cctaccgaac ctggcgaagt 3060
gcgtagactg gaacccggcc agagattcgc catcgaccag ctgaatcacc tgaacgccct 3120
gaaagaagat cggctgaaga agatggccaa caccatcatc atgcacgccc tgggctactg 3180
ctacgacgtg cggaagaaga aatggcaggc taagaacccc gcctgccaga tcatcctgtt 3240
cgaggatctg agcaactaca acccctacga ggaaaggtcc cgcttcgaga acagcaagct 3300

catgaagtgg tccagacgcg agatccccag acaggttgca ctgcagggcg agatctatgg 3360
cctgcaagtg ggagaagtgg gcgctcagtt cagcagcaga ttccacgcca agacaggcag 3420
ccctggcatc agatgtagcg tcgtgaccaa agagaagctg caggacaatc ggttcttcaa 3480
gaatctgcag agagagggca gactgaccct ggacaaaatc gccgtgctga aagagggcga 3540
tctgtaccca gacaaaggcg gcgagaagtt catcagcctg agcaaggatc ggaagtgcgt 3600
gaccacacac gccgacatca acgccgctca gaacctgcag aagcggttct ggacaagaac 3660
ccacggcttc tacaaggtgt actgcaaggc ctaccaggtg gacggccaga ccgtgtacat 3720

ccctgagagc aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat 3780
tctgaaggac ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag 3840
ctccaagcag agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct 3900
ggcctccgag ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt 3960
ccccagcgac aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat 4020
cagcaagctg accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc 4080
tatgaaaagg ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta 4140

cccatacgat gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata 4200

tgatgtcccc gactatgcct aagaattcct agagctcgct gatcagcctc gactgtgcct 4260

tctagttgcc agccatctgt tgtttgcccc tcccccgtgc cttccttgac cctggaaggt 4320

gccactccca ctgtcctttc ctaataaaat gaggaaattg catcgcattg tctgagtagg 4380

tgtcattcta ttctgggggg tggggtgggg caggacagca agggggagga ttgggaagag 4440

aatagcaggc atgctgggga cggccaaaaa aagagaccat atatggtctc cgtgctaatg 4500

cctcgtaaga gacatcgtcc agcaatagga gtttctcaca ccctgcagca cttatagcta 4560

gacggttgtc ctgaccaaaa gacagaaccg gtgtttcgtc ctttccacaa gatatataaa 4620

gccaagaaat cgaaatactt tcaagttacg gtaagcatat gatagtccat tttaaaacat 4680

aattttaaaa ctgcaaacta cccaagaaat tattactttc tacgtcacgt attttgtact 4740

aatatctttg tgtttacagt caaattaatt ccaattatct ctctaacagc cttgtatcgt 4800

atatgcaaat atgaaggaat catgggaaat aggccctcgc ggccgcagga acccctagtg 4860

atggagttgg ccactccctc tctgcgcgct cgctcgctca ctgaggccgg gcgaccaaag 4920

gtcgcccgac gcccgggctt tgcccgggcg gcctcagtga gcgagcgagc gcgcagctgc 4980

ctgcaggggc gcctgatgcg gtattttctc cttacgcatc tgtgcggtat ttcacaccgc 5040

atacgtcaaa gcaaccatag tacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 5100

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 5160

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 5220

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgatttgg 5280

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 5340

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 5400

cgggctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 5460

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgttt acaattttat 5520

ggtgcactct cagtacaatc tgctctgatg ccgcatagtt aagccagccc cgacacccgc 5580

caacacccgc tgacgcgccc tgacgggctt gtctgctccc ggcatccgct tacagacaag 5640

ctgtgaccgt ctccgggagc tgcatgtgtc agaggttttc accgtcatca ccgaaacgcg 5700

cgagacgaaa gggcctcgtg atacgcctat ttttataggt taatgtcatg ataataatgg 5760

tttcttagac gtcaggtggc acttttcggg gaaatgtgcg cggaacccct atttgtttat 5820

ttttctaaat acattcaaat atgtatccgc tcatgagaca ataaccctga taaatgcttc 5880

aataatattg aaaaaggaag agtatgagta ttcaacattt ccgtgtcgcc cttattccct 5940

tttttgcggc attttgcctt cctgtttttg ctcacccaga aacgctggtg aaagtaaaag 6000

atgctgaaga tcagttgggt gcacgagtgg gttacatcga actggatctc aacagcggta 6060

agatccttga gagttttcgc cccgaagaac gttttccaat gatgagcact tttaaagttc 6120

tgctatgtgg cgcggtatta tcccgtattg acgccgggca agagcaactc ggtcgccgca 6180

tacactattc tcagaatgac ttggttgagt actcaccagt cacagaaaag catcttacgg 6240

atggcatgac agtaagagaa ttatgcagtg ctgccataac catgagtgat aacactgcgg 6300

ccaacttact tctgacaacg atcggaggac cgaaggagct aaccgctttt ttgcacaaca 6360

tgggggatca tgtaactcgc cttgatcgtt gggaaccgga gctgaatgaa gccataccaa 6420

acgacgagcg tgacaccacg atgcctgtag caatggcaac aacgttgcgc aaactattaa 6480

ctggcgaact acttactcta gcttcccggc aacaattaat agactggatg gaggcggata 6540

aagttgcagg accacttctg cgctcggccc ttccggctgg ctggtttatt gctgataaat 6600

ctggagccgg tgagcgtgga agccgcggta tcattgcagc actggggcca gatggtaagc 6660

cctcccgtat cgtagttatc tacacgacgg ggagtcaggc aactatggat gaacgaaata 6720

gacagatcgc tgagataggt gcctcactga ttaagcattg gtaactgtca gaccaagttt 6780

actcatatat actttagatt gatttaaaac ttcattttta atttaaaagg atctaggtga 6840

agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagttttcg ttccactgag 6900

cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tccttttttt ctgcgcgtaa 6960

tctgctgctt gcaaacaaaa aaaccaccgc taccagcggt ggtttgtttg ccggatcaag 7020

agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata ccaaatactg 7080

tccttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat 7140

acctcgctct gctaatcctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta 7200

ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg 7260

gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc 7320

gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggcggacagg tatccggtaa 7380

gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac gcctggtatc 7440

tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt 7500

caggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg ttcctggcct 7560

tttgctggcc ttttgctcac atgt 7584

<210> 655

<211> 7495

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 655

cctgcaggca gctgcgcgct cgctcgctca ctgaggccgc ccgggcgtcg ggcgacctttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct gcggcctcta gactcgaggg gctggaagct acctttgaca tcatttcctc 180

tgcgaatgca tgtataattt ctacagaacc tattagaaag gatcacccag cctctgcttt 240

tgtacaactt tcccttaaaa aactgccaat tccactgctg tttggcccaa tagtgagaac 300

tttttcctgc tgcctcttgg tgcttttgcc tatggcccct attctgcctg ctgaagacac 360

tcttgccagc atggacttaa acccctccag ctctgacaat cctctttctc ttttgtttta 420

catgaagggt ctggcagcca aagcaatcac tcaaagttca aaccttatca ttttttgctt 480

tgttcctctt ggccttggtt ttgtacatca gctttgaaaa taccatccca gggttaatgc 540

tggggttaat ttataactaa gagtgctcta gttttgcaat acaggacatg ctataaaaat 600

ggaaagatac cggtgccacc atggccccaa agaagaagcg gaaggtcggt atccacggag 660

tcccagcagc cgccatccgg tccatcaagc tgaagatgaa gaccaacagc ggcaccgaca 720

gcatctacct gagaaaagcc ctgtggcgga cccaccagct gatcaatgag ggaatcgcct 780

actacatgaa cctgctgacc ctgtaccggc aagaggccat cggcgacaag accaaagaag 840

cctatcaggc cgagctgatt aacatcatcc ggaaccagca gcggaacaac ggcagctctg 900

aggaacacgg ctccgaccaa gaaattctgg ccctgctgag acagctgtac gagctgatca 960

tccccagcag catcggcgaa tctggcgacg ctaatcagct gggcaacaag tttctgtacc 1020

ctctggtgga ccccaacagc cagtctggca agggcacatc taacgccggc agaaagccca 1080

gatggaagcg gctgaaagag gaaggcaacc ccgactggga actcgagaag aagaaggacg 1140

aggaacgcaa ggccaaggat cccaccgtga agatctttga caacctgaac aaatacggcc 1200

tgctgcctct gttcccactg ttcaccaaca tccagaaaga catcgagtgg ctgcccctgg 1260

gcaagagaca gtctgtgcgg aagtgggaca aagacatgtt catccaggcc atcgagagac 1320

tgctgagctg ggagagctgg aacagaagag tggccgacga gtacaaacag ctgaaagaaa 1380

agaccgagag ctactacaaa gagcacctga caggcggcga ggaatggatc gagaagatcc 1440

ggaagttcga gaaagaacgg aacatggaac tggaaaagaa cgccttcgct cccaacgacg 1500

gctacttcat caccagcaga cagatcagag gctgggacag agtgtacgag aagtggtcca 1560

agctgcccga gtctgctagc cctgaggaac tgtggaaagt ggtggccgag cagcagaaca 1620

agatgtccga aggcttcggc gaccccaagg tgttcagctt cctggccaac agagagaacc 1680

gggacatttg gagaggccac agcgagcgga tctaccacat tgccgcctac aacggcctgc 1740

agaagaagct gagccggacc aaagagcagg ccaccttcac actgcctgac gccattgaac 1800

accctctgtg gatcagatac gagagccctg gcggcaccaa cctgaatctg ttcaagctgg 1860

aagagaaaca gaaaaagaac tactacgtga ccctgagcaa gatcatctgg cccagcgagg 1920

aaaagtggat tgagaaagag aacatcgaga tccctctggc tcccagcatc cagttcaacc 1980

ggcagattaa gctgaagcag cacgtgaagg gcaagcaaga gatcagcttc agcgactaca 2040

gcagccggat cagcctggat ggtgttctcg gcggcagcag aatccagttt aatcggaagt 2100

acatcaagaa ccacaaagag ctgctcggag agggcgacat cggccccgtg ttctttaacc 2160

tggtggtgga tgtggcccct ctgcaagaaa ccagaaacgg cagactgcag agccccatcg 2220

gcaaggccct gaaagtgatc agcagcgact tctccaaagt gatcgactac aagccgaaag 2280

aactcatgga ttggatgaat accggcagcg ccagcaacag ctttggagtg gcttctctgc 2340

tggaaggcat gagagtgatg agcatcgaca tgggccagag aaccagcgcc tccgtgtcca 2400

tcttcgaggt cgtgaaagaa ctgcccaagg atcaagagca gaagctgttc tacagcatca 2460

acgacaccga gctgttcgcc atccacaagc ggagctttct gctgaacctg cctggcgagg 2520

tggtcaccaa gaacaacaag cagcagcggc aagagcggcg gaaaaagcgg cagtttgtgc 2580

ggagccagat cagaatgctg gccaacgtgc tgcggctgga aacaaagaaa acccctgacg 2640

agcggaagaa ggccattcac aagctgatgg aaatcgtgca gagctacgac agctggaccg 2700

ccagccagaa agaagtgtgg gagaaagagc tgaatctcct gaccaacatg gccgccttca 2760

atgacgagat ctggaaagaa agcctggtgg aactgcacca ccggatcgag ccttacgtgg 2820

gacagatcgt gtccaagtgg cggaagggcc tgtctgaggg cagaaagaat ctggccggca 2880

tcagcatgtg gaacatcgac gaactggaag ataccaggcg gctgctgatt tcctggtcca 2940

agagaagcag aaccccaggc gaggccaaca ggatcgaaac cgatgagcct ttcggcagca 3000

gcctgctcca gcacattcag aacgtgaagg acgacagact gaagcagatg gccaacctga 3060

tcatcatgac agccctgggc tttaagtacg acaaagagga aaaggaccgg tacaagcggt 3120

ggaaagagac ataccccgcc tgccagatca tcctgttcga gaacctgaac cgctacctgt 3180

tcaacctcga ccggtccaga cgcgagaaca gcagactgat gaagtgggcc catcggagca 3240

tccccagaac cgtgtctatg cagggcgaga tgttcggcct gcaagtgggc gacgttcgga 3300

gcgagtacag ctccagattc cacgccaaaa caggcgcccc tggcatcaga tgtcacgccc 3360

tgactgaaga ggatctgaag gccggcagca acaccctgaa gagactgatc gaggacggct 3420

tcatcaatga gagcgagctg gcctacctga agaagggcga tatcatccct agccaaggcg 3480

gcgaactgtt cgtgacactg tccaagcggt acaagaagga cagcgacaac aacgagctga 3540

ccgtgatcca cgccgacatc aacgccgctc agaatctgca gaagcggttt tggcagcaaa 3600

acagcgaggt gtacagagtg ccctgtcagc tggccagaat gggcgaagat aagctgtaca 3660
tccccaagag ccagaccgag acaatcaaga agtatttcgg caagggctcc ttcgtgaaga 3720
acaataccga acaagaggtc tacaagtggg agaagtccga gaaaatgaag atcaagacgg 3780
acaccacctt cgacctgcaa gacctggatg gcttcgagga catcagcaag accattgagc 3840
tggcacaaga gcagcaaaag aaatacctga ccatgttcag ggaccccagc ggctactttt 3900
tcaacaatga gacatggcgg cctcaaaaag aatactggtc catcgtgaac aacatcatca 3960
agagctgcct caagaagaag atcctgagca acaaggtcga gctgaaaagg ccggcggcca 4020

cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat gttccagatt 4080
acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc gactatgcct 4140
aagaattcct agagctcgct gatcagcctc gactgtgcct tctagttgcc agccatctgt 4200
tgtttgcccc tcccccgtgc cttccttgac cctggaaggt gccactccca ctgtcctttc 4260
ctaataaaat gaggaaattg catcgcattg tctgagtagg tgtcattcta ttctgggggg 4320
tggggtgggg caggacagca agggggagga ttgggaagag aatagcaggc atgctgggga 4380
cggccaaaaa aagagaccat atatggtctc cgtgccaagc acctgttttc aggtgctcct 4440

gtggtgggta atttttaatt acttatggca cacgcacaga ttcattgacc ctataggtcc 4500
ggtgtttcgt cctttccaca agatatataa agccaagaaa tcgaaatact ttcaagttac 4560
ggtaagcata tgatagtcca ttttaaaaca taattttaaa actgcaaact acccaagaaa 4620
ttattacttt ctacgtcacg tattttgtac taatatcttt gtgtttacag tcaaattaat 4680
tccaattatc tctctaacag ccttgtatcg tatatgcaaa tatgaaggaa tcatgggaaa 4740
taggccctcg cggccgcagg aacccctagt gatggagttg gccactccct ctctgcgcgc 4800
tcgctcgctc actgaggccg ggcgaccaaa ggtcgcccga cgcccgggct ttgcccgggc 4860

ggcctcagtg agcgagcgag cgcgcagctg cctgcagggg cgcctgatgc ggtattttct 4920
ccttacgcat ctgtgcggta tttcacaccg catacgtcaa agcaaccata gtacgcgccc 4980
tgtagcggcg cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt 5040
gccagcgccc tagcgcccgc tcctttcgct ttcttccctt cctttctcgc cacgttcgcc 5100
ggctttcccc gtcaagctct aaatcggggg ctccctttag ggttccgatt tagtgcttta 5160
cggcacctcg accccaaaaa acttgatttg ggtgatggtt cacgtagtgg gccatcgccc 5220
tgatagacgg tttttcgccc tttgacgttg gagtccacgt tctttaatag tggactcttg 5280

ttccaaactg gaacaacact caaccctatc tcgggctatt cttttgattt ataagggatt 5340

ttgccgattt cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat 5400

tttaacaaaa tattaacgtt tacaatttta tggtgcactc tcagtacaat ctgctctgat 5460

gccgcatagt taagccagcc ccgacacccg ccaacacccg ctgacgcgcc ctgacgggct 5520

tgtctgctcc cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt 5580

cagaggtttt caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgccta 5640

tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg 5700

ggaaatgtgc gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg 5760

ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt 5820

attcaacatt tccgtgtcgc ccttattccc ttttttgcgg cattttgcct tcctgttttt 5880

gctcacccag aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg 5940

ggttacatcg aactggatct caacagcggt aagatccttg agagttttcg ccccgaagaa 6000

cgttttccaa tgatgagcac ttttaaagtt ctgctatgtg gcgcggtatt atcccgtatt 6060

gacgccgggc aagagcaact cggtcgccgc atacactatt ctcagaatga cttggttgag 6120

tactcaccag tcacagaaaa gcatcttacg gatggcatga cagtaagaga attatgcagt 6180

gctgccataa ccatgagtga taacactgcg gccaacttac ttctgacaac gatcggagga 6240

ccgaaggagc taaccgcttt tttgcacaac atgggggatc atgtaactcg ccttgatcgt 6300

tgggaaccgg agctgaatga agccatacca aacgacgagc gtgacaccac gatgcctgta 6360

gcaatggcaa caacgttgcg caaactatta actggcgaac tacttactct agcttcccgg 6420

caacaattaa tagactggat ggaggcggat aaagttgcag gaccacttct gcgctcggcc 6480

cttccggctg gctggtttat tgctgataaa tctggagccg gtgagcgtgg aagccgcggt 6540

atcattgcag cactggggcc agatggtaag ccctcccgta tcgtagttat ctacacgacg 6600

gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg tgcctcactg 6660

attaagcatt ggtaactgtc agaccaagtt tactcatata tactttagat tgatttaaaa 6720

cttcattttt aatttaaaag gatctaggtg aagatccttt ttgataatct catgaccaaa 6780

atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa gatcaaagga 6840

tcttcttgag atccttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg 6900

ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctctttttcc gaaggtaact 6960

ggcttcagca gagcgcagat accaaatact gtccttctag tgtagccgta gttaggccac 7020

cacttcaaga actctgtagc accgcctaca tacctcgctc tgctaatcct gttaccagtg 7080

gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg 7140

gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga 7200

acgacctaca ccgaactgag atacctacag cgtgagctat gagaaagcgc cacgcttccc 7260

gaagggagaa aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg 7320

agggagcttc cagggggaaa cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc 7380

tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc 7440

agcaacgcgg cctttttacg gttcctggcc ttttgctggc ctttgctca catgt 7495

<210> 656

<211> 14

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 656

ccgcgaaugg acac 14

<210> 657

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 657

tagatgaatt aaatgtgatt agcac 25

<210> 658

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 658

agcgagcggt ctgagaagtg gcact 25

<210> 659

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 659

gcagaaataa tgatgattgg cac 23

<210> 660

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 660

gucggaucgu ugagcgggcg aucugagaag uggcac 36

<210> 661

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 661

gggcgaucug agaaguggca c 21

<210> 662

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic Oligonucleotide

<400> 662

augauacgag gcauuagcac 20

<210> 663

<211> 102

<212> PRT

<213

> Artificial Sequence

<220><223> Synthetic Peptide

<400> 663

Ser Gly Val Asp Asp Asp Met Ala Cys His Lys Ile Pro Val Glu Ala

1 5 10 15

Asp Phe Leu Tyr Ala Tyr Ser Thr Ala Pro Gly Tyr Tyr Ser Trp Arg

20 25 30

Asn Ser Lys Asp Gly Ser Trp Phe Ile Gln Ser Leu Cys Ala Met Leu

35 40 45

Lys Gln Tyr Ala Asp Lys Leu Glu Phe Met His Ile Leu Thr Arg Val

50 55 60

Asn Arg Lys Val Ala Thr Glu Phe Glu Ser Phe Ser Phe Asp Ala Thr

65 70 75 80

Phe His Ala Lys Lys Gln Ile Pro Cys Ile Val Ser Met Leu Thr Lys

85 90 95

Glu Leu Tyr Phe Tyr His

100