



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2025년06월16일

(11) 등록번호 10-2821006

(24) 등록일자 2025년06월11일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) C12N 15/63 (2006.01)(52) CPC특허분류
C12N 9/22 (2025.05)
C12N 15/63 (2013.01)

(21) 출원번호 10-2023-7009916(분할)

(22) 출원일자(국제) 2014년09월05일

심사청구일자 2023년04월21일

(85) 번역문제출일자 2023년03월22일

(65) 공개번호 10-2023-0044544

(43) 공개일자 2023년04월04일

(62) 원출원 특허 10-2021-7014697

원출원일자(국제) 2014년09월05일

심사청구일자 2021년06월11일

(86) 국제출원번호 PCT/US2014/054291

(87) 국제공개번호 WO 2015/035162

국제공개일자 2015년03월12일

(30) 우선권주장

61/874,609 2013년09월06일 미국(US)

(뒷면에 계속)

(56) 선행기술조사문헌

JP2013534417 A*

US20110123509 A1

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

프레지던트 앤드 펠로우즈 오브 하바드 칼리지

미합중국, 메사추세츠 02138, 캠브리지, 퀸시스트리트17

(72) 발명자

리우, 데이비드, 알.

미국 02420 매사추세츠주 렉싱턴 위트먼 서클 3

길링거, 존, 폴

미국 81432 콜로라도주 리지웨이 베어 커브 드라이브 340

툼슨, 데이비드, 비.

미국 02138 매사추세츠주 캠브리지 #11 폴렌 스트리트 50

(74) 대리인

양영준, 김영

전체 청구항 수 : 총 24 항

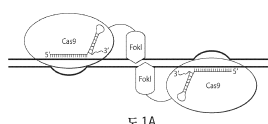
심사관 : 한지혜

(54) 발명의 명칭 Cas9 변이체 및 그의 용도

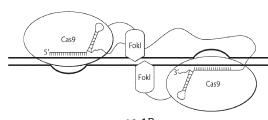
(57) 요약

본 개시내용의 일부 측면은 RNA-프로그램 가능한 엔도뉴클레아제, 예컨대 Cas9의 특이성을 개선시키기 위한 조성물, 방법 및 키트를 제공한다. 또한, 핵산 표적을 절단하기 위한 개선된 특이성을 지니도록 조작된, Cas9의 변이체, 예를 들어 Cas9 이량체 및 융합 단백질이 제공된다. 또한, Cas9 융합 단백질 (예를 들어, 뉴클레아제 촉매 도메인 또는 재조합효소 촉매 도메인과 융합된 뉴클레아제-불활성화 Cas9)을 이용하여 부위-특이적 핵산 변형시키기 위한 조성물, 방법 및 키트가 제공된다. 이러한 Cas9 변이체는 DNA의 부위-특이적 변형, 예를 들어 게놈 변형과 관련되는 임상 및 연구 환경에 유용하다.

대표도



도 1A



도 1B

(52) CPC특허분류
C07K 2319/09 (2013.01)

(30) 우선권주장
61/915,414 2013년12월12일 미국(US)
61/980,315 2014년04월16일 미국(US)
14/320,467 2014년06월30일 미국(US)
14/320,498 2014년06월30일 미국(US)

명세서

청구범위

청구항 1

Cas9 단백질 및 추가의 단백질 도메인을 포함하는 융합 단백질을 포함하는 폴리펩티드이며, Cas9 단백질 및 추가의 단백질 도메인은 SGSETPGTSESATPES (서열 16), SGSETPGTSESA (서열 17) 및 SGSETPGTSESATPEGSGGS (서열 18)로 이루어진 군으로부터 선택되는 링커 서열에 의해 연결되는 것인, 폴리펩티드.

청구항 2

제1항에 있어서, 링커 서열이 SGSETPGTSESATPES (서열 16)인 폴리펩티드.

청구항 3

제1항에 있어서, 링커 서열이 SGSETPGTSESA (서열 17)인 폴리펩티드.

청구항 4

제1항에 있어서, 링커 서열이 SGSETPGTSESATPEGSGGS (서열 18)인 폴리펩티드.

청구항 5

제1항에 있어서, Cas9 단백질이 야생형 Cas9 단백질, 뉴클레아제-불활성화 Cas9 (dCas9) 단백질 또는 Cas9 니카제인 폴리펩티드.

청구항 6

제1항에 있어서, Cas9 단백질이 서열 2, 4, 또는 5와 80% 이상, 85% 이상, 90% 이상, 95% 이상, 98% 이상 99% 이상 또는 100% 동일한 아미노산 서열을 포함하는 것인 폴리펩티드.

청구항 7

제1항에 있어서, Cas9 단백질이 서열 4의 D10, H840 또는 D10 및 H840 둘 다에 상응하는 위치에서의 아미노산 치환을 포함하는 것인 폴리펩티드.

청구항 8

제7항에 있어서, 서열 4의 D10 및 H840에 상응하는 위치에서의 아미노산 치환이 D10A 또는 H840A인 폴리펩티드.

청구항 9

제7항에 있어서, 서열 4의 D10 및 H840에 상응하는 위치에서의 아미노산 치환이 D10A 및 H840A인 폴리펩티드.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서, 융합 단백질이 핵 국재화 신호 (NLS) 도메인을 추가로 포함하는 것인 폴리펩티드.

청구항 11

제1항 내지 제9항 중 어느 한 항에 있어서, Cas9 단백질이 gRNA에 결합하는 것인 폴리펩티드.

청구항 12

제1항 내지 제9항 중 어느 한 항에 있어서, 융합 단백질이 구조 NH₂-[Cas9 단백질]-[추가 단백질 도메인]-COOH 또는 NH₂-[추가 단백질 도메인]-[Cas9 단백질]-COOH를 포함하고, 여기서 상기 "-"는 링커 서열을 나타내는 것인 폴리펩티드.

청구항 13

제1항 내지 제9항 중 어느 한 항의 폴리펩티드 및 gRNA를 포함하는 복합체.

청구항 14

제1항 내지 제9항 중 어느 한 항의 폴리펩티드를 코딩하는 폴리뉴클레오티드.

청구항 15

제14항의 폴리뉴클레오티드를 포함하는 벡터.

청구항 16

제1항 내지 제9항 중 어느 한 항의 폴리펩티드를 발현하기 위한 유전적 구조물을 포함하는 단리된 세포.

청구항 17

제16항에 있어서, 원핵 세포인 단리된 세포.

청구항 18

제16항에 있어서, 진핵 세포인 단리된 세포.

청구항 19

제18항에 있어서, 인간 세포인 단리된 세포.

청구항 20

제1항 내지 제9항 중 어느 한 항의 폴리펩티드 및 제약상 허용되는 부형제; 제1항 내지 제9항 중 어느 한 항의 폴리펩티드를 코딩하는 폴리뉴클레오티드 및 제약상 허용되는 부형제; 제1항 내지 제9항 중 어느 한 항의 폴리펩티드를 코딩하는 폴리뉴클레오티드를 포함하는 벡터 및 제약상 허용되는 부형제; 제1항 내지 제9항 중 어느 한 항의 폴리펩티드를 발현하기 위한 유전적 구조물을 포함하는 단리된 세포 및 제약상 허용되는 부형제; 또는 제1항 내지 제9항 중 어느 한 항의 폴리펩티드 및 gRNA를 포함하는 복합체 및 제약상 허용되는 부형제를 포함하는 제약 조성물.

청구항 21

제1항 내지 제9항 중 어느 한 항의 폴리펩티드, 제1항 내지 제9항 중 어느 한 항의 폴리펩티드를 코딩하는 폴리뉴클레오티드, 제1항 내지 제9항 중 어느 한 항의 폴리펩티드를 코딩하는 폴리뉴클레오티드를 포함하는 벡터, 제1항 내지 제9항 중 어느 한 항의 폴리펩티드를 발현하기 위한 유전적 구조물을 포함하는 세포, 또는 제1항 내지 제9항 중 어느 한 항의 폴리펩티드 및 gRNA를 포함하는 복합체를 포함하는 키트.

청구항 22

제1항 내지 제9항 중 어느 한 항의 폴리펩티드 또는 제1항 내지 제9항 중 어느 한 항의 폴리펩티드 및 gRNA를 포함하는 복합체를 DNA 서열과 접촉시키는 단계를 포함하는, DNA 서열을 편집하기 위한 시험관 내 또는 생체 외 방법.

청구항 23

제20항의 제약 조성물을 DNA 서열과 접촉시키는 단계를 포함하는, DNA 서열을 편집하기 위한 시험관 내 또는 생체 외 방법.

청구항 24

제20항에 있어서, 의약으로서 사용하기 위한 제약 조성물.

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

발명의 설명

기술 분야

관련 출원

본 출원은 35 U.S.C. § 365(c) 하에 2014년 6월 30일에 출원된 미국 출원 번호 14/320,498 및 2014년 6월 30일에 출원된 미국 출원 번호 14/320,467을 우선권 주장하며, 또한 35 U.S.C. § 119(e) 하에 2013년 9월 6일에 출원된 미국 가출원 번호 61/874,609; 및 2013년 12월 12일에 출원된 미국 가출원 번호 61/915,414; 및 2014년 4월 16일에 출원된 미국 가출원 번호 61/980,315를 우선권 주장하며, 이들 각각은 본원에 참조로 포함된다.

배경 기술

부위-특이적 엔도뉴클레아제는 이론상, 게놈 내의 단일 부위의 표적화된 조작을 허용해주고 치료 및 연구 적용을 위한 유전자 표적화의 맥락에서 유용하다. 포유류를 포함한 각종 유기체에서는, 부위-특이적 엔도뉴클레아제가 비-상동 말단 결합 또는 상동 재조합을 자극함으로써 게놈 공학을 위해 사용되어 왔다. 강력한 연구 도구를 제공하는 것 이외에도, 부위-특이적 뉴클레아제는 또한, 유전자 요법 작용제로서의 잠재력을 지니고 있고, 다음 2개의 부위-특이적 엔도뉴클레아제가 최근에 임상 시험에 들어갔다: (1) 항-HIV 치료적 접근 방식의 일부로서 인간 CCR-5 대립 유전자를 표적으로 하는 CCR5-2246 (NCT00842634, NCT01044654, NCT01252641); 및 (2) 항암 치료적 접근 방식의 일부로서 인간 VEGF-A 프로모터를 표적으로 하는 VF24684 (NCT01082926).

표적을 벗어난 활성(off-target activity)을 수반하지 않거나 최소한의 표적을 벗어난 활성만을 수반하는 것으로 의도된 뉴클레아제 표적 부위를 특이적으로 절단하는 것이 부위-특이적 엔도뉴클레아제의 임상 적용을 위한 전제 조건이고, 또한 기본 연구 적용에서 고-효율 게놈 조작을 위한 전제 조건이다. 예를 들어, 조작된 부위-특이적 결합성 도메인의 불완전한 특이성은 의도한 표적 이외의 게놈 유전자 자리의 바람직하지 못한 변경 및 세포성 독성과 관련이 있어 왔다. 그러나, 현재 입수 가능한 대부분의 뉴클레아제는 상당히 표적을 벗어난 활성을 나타내므로, 임상 적용에 적합하지 않을 수 있다. 임상 및 연구 환경에서 사용하기 위하여 최근에 생겨난 뉴클레아제 플랫폼(platform)은 RNA-가이드된 뉴클레아제, 예컨대 Cas9이다. 이들 뉴클레아제가 특이적 표적 부위의 절단을 지시하는 가이드 RNA (gRNA)와 결합할 수 있긴 하지만, 특정의 Cas9:gRNA 복합체에 대해서는 표적을 벗어난 활성이 여전히 관찰된다 (Pattanayak et al., "High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity." *Nat Biotechnol.* 2013; doi: 10.1038/nbt.2673). 따라서, 개선된 특이성을 지닌 뉴클레아제를 조작하기 위한 기술이 필요하다.

표적화된 유전적 조작에 유용한 또 다른 부류의 효소는 부위-특이적 재조합효소 (SSR)이다. 이들 효소는 짧은 DNA 서열을 인식하고 이와 결합함으로써 DNA 절편의 재배열을 수행하는데, 여기에서 이들은 DNA 골격을 절단하고, 관련된 2개의 DNA 나선을 교환하며 DNA 가닥을 재연결시킨다. 이러한 재배열로 인해, DNA 절편의 표적화된 삽입, 역위, 절제 또는 전위가 허용된다. 그러나, 부위-특이적 엔도뉴클레아제와 같이, 자연 발생적 SSR은 전형적으로, 특이적 컨센스 서열을 인식하고 이와 결합하므로, 이 점에 있어서 제한된다. 변경되고/되거나 개선된 특이성을 지닌 재조합효소(recombinase)를 조작하기 위한 기술이 또한 필요하다.

발명의 내용

본 개시내용의 일부 측면은 일부 조작된 부위-특이적 엔도뉴클레아제의 보고된 독성이 표적을 벗어난 DNA 절단에 근거한다는 인식에 기초한다. 따라서 본원에 기재된 특정 측면은 서열 (예를 들어, 목적하는 표적에서 2개 이상의 부위에 뉴클레아제 결합을 갖는다)의 수를 증가시키고/시키거나 2개 이상의 단백질 간의 뉴클레아제의

활성 (예를 들어, 표적 결합성 및 표적 절단성)을 분할시키는 것이 뉴클레아제의 특이성을 증가시킴으로써, 표적을 벗어난 효과가 발생할 가능성이 감소된다는 발견에 관한 것이다. 따라서, 본 개시내용의 일부 측면은 부위-특이적 뉴클레아제, 특히 RNA-프로그램 가능한 엔도뉴클레아제, 예컨대 Cas9 엔도뉴클레아제의 특이성을 개선하기 위한 전략, 조성물, 시스템 및 방법을 제공한다. 본 개시내용의 특정 측면은 개선된 특이성을 갖도록 조작된 Cas9 엔도뉴클레아제의 변이체를 제공한다.

[0007] 본 개시내용의 다른 측면은 현재 입수 가능한 부위-특이적 재조합효소 (SSR)가 전형적으로, 별개의 컨센서스 서열을 인식하고 이와 결합하는 것으로 제한된다는 인식에 기초한다. 따라서 본원에 기재된 특정 측면은 RNA-프로그램 가능한 (뉴클레아제-불활성화) 뉴클레아제 (또는 그의 RNA-결합성 도메인)와 재조합효소 도메인 간의 융합이, 이론상 예를 들어, 기술자에 의해 선택된 어떠한 부위 [예를 들어, 재조합되는 구역의 서열에 따라서 조작되거나 선택되는 가이드 RNA (gRNA)에 의해 지정된 부위]에서도 DNA와 결합하고 재조합할 수 있는 신규 재조합효소를 제공한다는 발견에 관한 것이다. 따라서, 이러한 신규 재조합효소는 특히, 표적화된 삽입, 결실, 역위, 전위 또는 기타 게놈 변형에 유용하다. 따라서, 예를 들어 상기 표적화된 게놈 조작을 위하여, 이들 본 발명이 재조합효소 융합 단백질을 사용하는 방법이 또한 제공된다.

[0008] 따라서, 본 개시내용의 한 실시양태는, 예를 들어 다음 2개의 도메인을 포함하는 융합 단백질 및 그의 이량체를 제공한다: (i) 뉴클레아제-불활성화 Cas9 도메인; 및 (ii) 뉴클레아제 도메인 (예를 들어, FokI DNA 절단 도메인의 단량체) (예를 들어, 도 1A, 6d 참조). 이러한 융합 단백질은 세포의 핵 내로 수송될 융합 단백질에 대하여 신호를 보내는 핵 국재화 신호 (NLS) 도메인을 추가로 포함할 수 있다. 일부 실시양태에서, 융합 단백질의 하나 이상의 도메인은 링커에 의해 분리된다. 특정 실시양태에서, 이러한 링커는 비-펩티드성 링커이다. 특정 실시양태에서, 링커는 펩티드 링커이다. 펩티드 링커의 경우, 이러한 펩티드 링커는 XTEN 링커, 트리-펩티드 GGS의 하나 이상의 반복 서열을 포함하는 아미노산 서열, 또는 도 12a에 제공된 바와 같은 어느 서열을 포함할 수 있다. 일부 실시양태에서, 융합 단백질은 서열 9, 서열 10, 서열 11, 또는 서열 12, 또는 서열 9, 서열 10, 서열 11, 또는 서열 12 중 어느 하나의 변이체 또는 단편으로서 제시된 뉴클레오티드 서열에 의해 코딩된다. 상기 뉴클레아제-불활성화 Cas9 도메인은 가이드 RNA (gRNA)와 결합할 수 있다. 특정 실시양태에서, 각각 표적 핵산의 2개의 별개의 영역과 결합되는 gRNA를 포함하는, 상기 융합 단백질의 이량체를 가짐으로써, 예를 들어 표적 핵산과의 결합을 지시하는 단일 gRNA를 포함하는 단량체성 RNA-가이드된 뉴클레아제와 비교해서 개선된 특이성이 제공된다.

[0009] 본 발명의 또 다른 측면에 따라서, 본 발명의 Cas9 변이체를 이용한 부위-특이적 DNA 절단 방법이 제공된다. 이러한 방법은 전형적으로, (a) DNA를 본 발명의 융합 단백질 (예를 들어, 뉴클레아제-불활성화 Cas9 도메인과 FokI DNA 절단 도메인을 포함하는 융합 단백질)과 접촉시키는 단계 (여기서, 불활성 Cas9 도메인은 DNA의 특정 영역과 혼성화되는 gRNA와 결합된다); 및 (b) 상기 DNA를 제2 융합 단백질 (예를 들어, 뉴클레아제-불활성화 Cas9와 FokI DNA 절단 도메인을 포함하는 융합 단백질)과 접촉시키는 단계 (여기서, 이러한 제2 융합 단백질의 불활성 Cas9 도메인은 DNA의 제2 영역과 혼성화되는 제2의 gRNA와 결합된다)를 포함한다 [여기서, 단계 (a) 및 (b)에서의 융합 단백질의 결합으로 인해, 융합 단백질의 뉴클레아제 도메인이 이량체화되어, DNA가 상기 결합된 융합 단백질들 사이의 영역에서 절단된다]. 일부 실시양태에서, 단계 (a) 및 (b)의 gRNA는 DNA의 동일 가닥과 혼성화되거나, 또는 단계 (a) 및 (b)의 gRNA는 DNA의 반대 가닥과 혼성화된다. 일부 실시양태에서, 단계 (a) 및 (b)의 gRNA는 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어진 DNA의 영역과 혼성화된다. 일부 실시양태에서, DNA는 특정 개체, 예컨대 인간 내에 존재할 수 있는 특정 세포, 예를 들어 진핵 세포 또는 원핵 세포 내에 있다.

[0010] 또 다른 실시양태에 따라서, 본 발명의 융합 단백질의 이량체 (예를 들어, 뉴클레아제-불활성화 Cas9와 FokI DNA 절단 도메인을 포함하는 융합 단백질의 이량체)를 포함하는 복합체가 제공된다. 일부 실시양태에서, 이러한 이량체의 각 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 단일 연장된 gRNA와 결합되어, 이량체의 하나의 융합 단백질이 gRNA의 특정 부분과 결합되고 이량체의 다른 융합 단백질이 gRNA의 또 다른 부분과 결합되도록 한다 (예를 들어, 도 1B 참조). 일부 실시양태에서, 상기 gRNA는 50개 이상, 75개 이상, 100개 이상, 150개 이상, 200개 이상, 250개 이상, 또는 300개 이상 뉴클레오티드 길이이다. 일부 실시양태에서, 표적 핵산과 혼성화되는 상기 연장된 gRNA의 영역은 15 내지 25개, 19 내지 21개, 또는 20개 뉴클레오티드를 포함한다.

[0011] 또 다른 실시양태에서, DNA를, 단일 연장된 gRNA와 결합된 본 발명의 2개 융합 단백질의 복합체와 접촉시키는 것을 포함하는, 부위-특이적 DNA 절단 방법이 제공된다. 일부 실시양태에서, gRNA는 절단될 DNA의 2개의 별도의 영역과 혼성화되는 2개 부분을 함유하고; 복합체는 이러한 gRNA의 부분들이 상기 2개의 영역과 혼성화됨에

따라 DNA와 결합되며; 이러한 복합체의 결합으로 인해, 융합 단백질의 뉴클레아제 도메인이 이량체화되어, 이러한 도메인이 상기 결합된 융합 단백질들 사이의 영역에서 DNA를 절단한다. 일부 실시양태에서, gRNA의 2개 부분은 DNA의 동일 가닥과 혼성화된다. 다른 실시양태에서, gRNA의 2개 부분은 DNA의 반대 가닥과 혼성화된다. 일부 실시양태에서, gRNA의 2개 부분은 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어진 DNA의 영역과 혼성화된다. 일부 실시양태에서, DNA는 특정 개체, 예컨대 인간 내에 존재할 수 있는 특정 세포, 예를 들어 진핵 세포 또는 원핵 세포 내에 있다.

[0012] 본 발명의 또 다른 실시양태에 따라서, Cas9 단백질의 단편을 포함하는 분할(split) Cas9 단백질 (분할 Cas9 단백질을 포함하는 융합 단백질 포함)이 제공된다. 일부 실시양태에서, Cas9의 gRNA 결합성 도메인은 포함하지만, DNA 절단 도메인은 포함하지 않는 단백질이 제공된다. 다른 실시양태에서, Cas9의 DNA 절단 도메인은 포함하지만, gRNA 결합성 도메인은 포함하지 않는 단백질이 제공된다. 일부 실시양태에서, 다음 2개의 도메인을 포함하는 융합 단백질이 제공된다: (i) 뉴클레아제-불활성화 Cas9 도메인, 및 (ii) Cas9의 gRNA 결합성 도메인 (예를 들어, 여기서 도메인 (ii)은 DNA 절단 도메인을 포함하지 않는다) (예를 들어, 도 2B 참조). 일부 실시양태에서, 다음 2개의 도메인을 포함하는 융합 단백질이 제공된다: (i) 뉴클레아제-불활성화 Cas9 도메인, 및 (ii) DNA 절단 도메인 (예를 들어, 여기서 도메인 (ii)은 gRNA 결합성 도메인을 포함하지 않는다) [예를 들어, 도 2C ("B" 도메인을 포함한, 우측 상의 융합 단백질) 참조]. 일부 실시양태에서, 본원에 기재된 단백질 중 어느 것의 단백질 이량체가 제공된다. 예를 들어, 일부 실시양태에서, 이량체는 분할 Cas9 단백질의 2개의 절반, 예를 들어 (i) Cas9의 gRNA 결합성 도메인은 포함하지만, DNA 절단 도메인은 포함하지 않는 단백질, 및 (ii) Cas9의 DNA 절단 도메인은 포함하지만, gRNA 결합성 도메인은 포함하지 않는 단백질을 포함한다 (예를 들어, 도 2A 참조). 일부 실시양태에서, 이량체는 분할 Cas9 단백질의 하나의 절반; 및 분할 Cas9 단백질의 다른 절반을 포함하는 융합 단백질을 포함한다 (예를 들어, 도 2B 참조). 예를 들어, 특정 실시양태에서 이러한 이량체는 (i) Cas9의 gRNA 결합성 도메인은 포함하지만, DNA 절단 도메인은 포함하지 않는 단백질, 및 (ii) 뉴클레아제-불활성화 Cas9와 DNA 절단 도메인을 포함하는 융합 단백질을 포함한다. 다른 실시양태에서, 상기 이량체는 (i) Cas9의 DNA 절단 도메인은 포함하지만, gRNA 결합성 도메인은 포함하지 않는 단백질, 및 (ii) 뉴클레아제-불활성화 Cas9와 Cas9의 gRNA 결합성 도메인을 포함하는 융합 단백질을 포함한다. 일부 실시양태에서, 2개의 융합 단백질 (각각의 융합 단백질은 뉴클레아제-불활성화 Cas9와 분할 Cas9의 하나의 절반을 포함한다)을 포함하는 이량체가 제공된다 (예를 들어, 도 2C 참조). 예를 들어, 특정 실시양태에서, 이러한 이량체는 (i) 뉴클레아제-불활성화 Cas9와 Cas9의 gRNA 결합성 도메인을 포함하는 융합 단백질, 및 (ii) 뉴클레아제-불활성화 Cas9와 DNA 절단 도메인을 포함하는 융합 단백질을 포함한다. 일부 실시양태에서, 이와 같이 제공된 단백질 이량체 중 어느 것이 하나 이상의 gRNA(들)와 연합한다.

[0013] 일부 실시양태에서, 본 발명의 단백질 이량체를 활용하는 부위-특이적 DNA 절단 방법이 제공된다. 예를 들어, 일부 실시양태에서, 이러한 방법은 DNA를, (i) Cas9의 gRNA 결합성 도메인은 포함하지만, DNA 절단 도메인은 포함하지 않는 단백질, 및 (ii) Cas9의 DNA 절단 도메인은 포함하지만, gRNA 결합성 도메인은 포함하지 않는 단백질을 포함하는 단백질 이량체와 접촉시키는 단계 (여기서, 이러한 이량체는 DNA의 특정 영역과 혼성화되는 gRNA와 결합되고, 이러한 DNA의 절단이 일어난다)를 포함한다 (예를 들어, 도 2A 참조). 일부 실시양태에서, 부위-특이적 DNA 절단에 사용된 단백질 이량체는 (i) Cas9의 gRNA 결합성 도메인은 포함하지만, DNA 절단 도메인은 포함하지 않는 단백질, 및 (ii) 뉴클레아제-불활성화 Cas9와 DNA 절단 도메인을 포함하는 융합 단백질을 포함한다 (예를 들어, 도 2B 참조). 일부 실시양태에서, 부위-특이적 DNA 절단에 사용된 이량체는 (i) Cas9의 DNA 절단 도메인은 포함하지만, gRNA 결합성 도메인은 포함하지 않는 단백질, 및 (ii) 뉴클레아제-불활성화 Cas9와 Cas9의 gRNA 결합성 도메인을 포함하는 융합 단백질을 포함한다. 일부 실시양태에서, 이러한 단백질 이량체는 DNA의 2개 영역과 혼성화되는 2개의 gRNA와 결합되고, 이러한 DNA의 절단이 일어난다 (예를 들어, 도 2B 참조). 일부 실시양태에서, 상기 2개의 gRNA는 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어진 DNA의 영역과 혼성화된다. 일부 실시양태에서, 부위-특이적 DNA 절단에 사용된 이량체는 다음 2개의 융합 단백질을 포함한다: (i) 뉴클레아제-불활성화 Cas9와 Cas9의 gRNA 결합성 도메인을 포함하는 융합 단백질, 및 (ii) 뉴클레아제-불활성화 Cas9와 DNA 절단 도메인을 포함하는 융합 단백질. 일부 실시양태에서, 단백질 이량체는 DNA의 3개 영역과 혼성화되는 3개의 gRNA와 결합되고, 이러한 DNA의 절단이 일어난다. 이러한 배열을 가짐으로써, 예를 들어 2개 이상의 gRNA (또는 표적과 혼성화되는 2개 이상의 영역을 포함하는 gRNA)와 연합된 이량체를 이용하여, 예를 들어 표적 핵산의 2개 이상의 영역을 표적화함으로써, 표적 핵산의 단일 영역과 결합되는 뉴클레아제와 비교해서 절단 특이성이 증가된다. 일부 실시양태에서, 3개의 gRNA는 제1 영역과 제2 영역 사이, 및 제2 영역과 제3 영

역 사이가 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어진 DNA의 영역과 혼성화된다. 일부 실시양태에서, DNA는 특정 개체, 예컨대 인간 내에 존재할 수 있는 특정 세포, 예를 들어 진핵 세포 또는 원핵 세포 내에 있다.

[0014] 또 다른 실시양태에 따라서, 최소한의 Cas9 단백질이 제공되는데, 예를 들어 이러한 단백질은 N- 및/또는 C-말단 끝을 잘라낸 것을 포함하고, RNA 결합성과 DNA 절단 활성을 보유하고 있다. 일부 실시양태에서, 상기와 같이 N-말단 끝을 잘라낸 것은 5개 이상, 10개 이상, 15개 이상, 20개 이상, 25개 이상, 40개 이상, 40개 이상, 50개 이상, 75개 이상, 100개 이상, 또는 150개 이상의 아미노산을 제거한다. 일부 실시양태에서, C-말단 끝을 잘라내는 것은 5개 이상, 10개 이상, 15개 이상, 20개 이상, 25개 이상, 40개 이상, 40개 이상, 50개 이상, 75개 이상, 100개 이상, 또는 150개 이상의 아미노산을 제거한다. 일부 실시양태에서, 상기와 같이 최소화된 Cas9 단백질은 추가로, 결합된 gRNA를 포함한다.

[0015] 일부 실시양태에서, DNA를 최소화된 Cas9 단백질:gRNA 복합체와 접촉시키는 단계를 포함하는 부위-특이적 DNA 절단 방법이 제공된다.

[0016] 또 다른 실시양태에 따라서, Cas9 (또는 그의 단편)의 이량체 (여기서, 이러한 이량체는 단일 gRNA를 통하여 조정된다)가 제공된다. 일부 실시양태에서, 단일 gRNA는 (i) 각각 Cas9 단백질과 결합할 수 있고, (ii) 각각 표적 핵산 서열 (예를 들어, DNA 서열)과 혼성화되는 2개 이상의 부분을 포함한다. 일부 실시양태에서, 표적 핵산과 혼성화되는 gRNA의 부분은 각각, 표적 핵산 서열에 대해 상보적인 5개 이하, 10개 이하, 또는 15개 이하의 뉴클레오티드를 포함한다. 일부 실시양태에서, 표적 핵산과 혼성화되는 gRNA의 부분은 링커 서열에 의해 분리된다. 일부 실시양태에서, 이러한 링커 서열은 표적 핵산과 혼성화된다 (예를 들어, 도 4 참조). 일부 실시양태에서, DNA를, 단일 gRNA를 통하여 조정된 Cas9 단백질의 이량체와 접촉시키는 단계를 포함하는 부위-특이적 DNA 절단 방법이 제공된다.

[0017] 또 다른 실시양태에 따라서, 본 개시내용은, 예를 들어 다음 2개의 도메인을 포함하는 융합 단백질, 및 그의 이량체 및 사량체를 제공한다: (i) 뉴클레아제-불활성화 Cas9 도메인; 및 (ii) 재조합효소 촉매 도메인 (예를 들어, 도 5 참조). 이러한 재조합효소 촉매 도메인은 일부 실시양태에서, Hin 재조합효소, Gin 재조합효소, 또는 Tn3 위치 특이성 재조합 촉진 효소(resolvase)의 재조합효소 촉매 도메인으로부터 유래된다. 뉴클레아제-불활성화 Cas9 도메인은, 예를 들어 상기 융합 단백질이 표적 핵산 서열을 표적으로 하도록 gRNA와 결합할 수 있다. 융합 단백질은 세포의 핵 내로 수송될 융합 단백질에 대하여 신호를 보내는 핵 국재화 신호 (NLS) 도메인을 추가로 포함할 수 있다. 일부 실시양태에서, 융합 단백질의 하나 이상의 도메인은 링커에 의해 분리된다. 특정 실시양태에서, 이러한 링커는 비-펩티드성 링커이다. 특정 실시양태에서, 링커는 펩티드 링커이다. 펩티드 링커의 경우, 이러한 펩티드 링커는 XTEN 링커, 트리-펩티드 GGS의 하나 이상의 반복 서열을 포함하는 아미노산 서열, 또는 도 12a에 제공된 바와 같은 어느 서열을 포함할 수 있다.

[0018] 또 다른 실시양태에서, 본원에 기재된 본 발명의 RNA-가이드된 재조합효소 융합 단백질을 활용하는, 부위-특이적 재조합 방법이 제공된다. 일부 실시양태에서, 이러한 방법은 2개의 별도의 DNA 분자를 재조합하는 데 유용하고, 이는 (a) 제1 DNA를 제1 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 뉴클레아제-불활성화 Cas9 도메인은 제1 DNA의 특정 영역과 혼성화되는 제1 gRNA와 결합된다); (b) 제1 DNA를 제2 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제2 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 제1 DNA의 제2 영역과 혼성화되는 제2 gRNA와 결합된다); (c) 제2 DNA를 제3 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제3 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 제2 DNA의 특정 영역과 혼성화되는 제3 gRNA와 결합된다); 및 (d) 제2 DNA를 제4 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제4 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 제2 DNA의 제2 영역과 혼성화되는 제4 gRNA와 결합된다)를 포함한다 [여기서, 단계 (a) 내지 (d)에서의 융합 단백질의 결합으로 인해, 상기 DNA가 재조합되도록 하는 조건 하에 융합 단백질의 재조합효소 촉매 도메인이 사량체화된다]. 일부 실시양태에서, 단일 DNA 분자의 2개 영역 간의 부위-특이적 재조합 방법이 제공된다. 일부 실시양태에서, 이러한 방법은 (a) 특정 DNA를 제1 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 뉴클레아제-불활성화 Cas9 도메인은 상기 DNA의 특정 영역과 혼성화되는 제1 gRNA와 결합된다); (b) 상기 DNA를 제2 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제2 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 상기 DNA의 제2 영역과 혼성화되는 제2 gRNA와 결합된다); (c) 상기 DNA를 제3 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제3 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 상기 DNA의 제3 영역과 혼성화되는 제3 gRNA와 결합된다); 및 (d) 상기 DNA를 제4 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계

(여기서, 제4 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 상기 DNA의 제4 영역과 혼성화되는 제4 gRNA와 결합된다)를 포함한다 [여기서, 단계 (a) 내지 (d)에서의 융합 단백질의 결합으로 인해, 상기 DNA가 재조합되도록 하는 조건 하에 융합 단백질의 재조합효소 촉매 도메인이 사랑체화된다]. 부위-특이적 재조합 방법을 포함한 일부 실시양태에서, 동일한 DNA 분자와 혼성화되는 gRNA는 이러한 DNA 분자의 반대 가닥과 혼성화된다. 예를 들어, 단일 DNA 분자의 부위-특이적 재조합을 포함한 일부 실시양태에서, 2개의 gRNA는 상기 DNA의 하나의 가닥과 혼성화되고, 다른 2개의 gRNA는 반대 가닥과 혼성화된다. 일부 실시양태에서, 이들 gRNA는 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어진 그들 각각의 DNA의 영역 (예를 들어, 동일한 가닥 상에 있음)과 혼성화된다. 일부 실시양태에서, 상기 DNA는 특정 개체, 예컨대 인간 내에 존재하거나 또는 이로부터 획득될 수 있는 특정 세포, 예를 들어 진핵 세포 또는 원핵 세포 내에 있다.

[0019] 또 다른 실시양태에 따라서, 예를 들어 본원에 기재된 Cas9 단백질 (예를 들어, Cas9 변이체, Cas9 이량체, Cas9 융합 단백질, Cas9 단편, 최소화 Cas9 단백질, 절단 도메인을 수반하지 않는 Cas9 변이체, gRNA 도메인을 수반하지 않는 Cas9 변이체, Cas9-재조합효소 융합물 등) 중 어느 것을 코딩하는 폴리뉴클레오티드가 제공된다. 일부 실시양태에서, 본원에 기재된 gRNA 중 어느 것을 코딩하는 폴리뉴클레오티드가 제공된다. 일부 실시양태에서, 본원에 기재된 본 발명의 Cas9 단백질 및 본원에 기재된 바와 같은 gRNA(들)의 어느 조합물을 코딩하는 폴리뉴클레오티드가 제공된다. 일부 실시양태에서, 본원에 기재된 폴리뉴클레오티드를 포함하는 벡터가 제공된다. 일부 실시양태에서, 본원에 기재된 Cas9 단백질 및/또는 gRNA 중 어느 것을 코딩하는 폴리뉴클레오티드를 포함하는 재조합 단백질 발현을 위한 벡터가 제공된다. 일부 실시양태에서, 본원에 기재된 Cas9 단백질 및/또는 gRNA 중 어느 것을 발현하기 위한 유전적 구조물을 포함하는 세포가 제공된다.

[0020] 일부 실시양태에서, 키트가 제공된다. 예를 들어, 본원에 기재된 Cas9 단백질 및/또는 gRNA 중 어느 것을 포함하는 키트가 제공된다. 일부 실시양태에서, 본원에 기재된 폴리뉴클레오티드, 예를 들어 Cas9 단백질 및/또는 gRNA를 코딩하는 것 중 어느 것을 포함하는 키트가 제공된다. 일부 실시양태에서, 재조합 단백질 발현을 위한 벡터 (여기서, 이러한 벡터는 본원에 기재된 Cas9 단백질 및/또는 gRNA 중 어느 것을 코딩하는 폴리뉴클레오티드를 포함한다)를 포함하는 키트가 제공된다. 일부 실시양태에서, 본원에 기재된 Cas9 단백질 및/또는 gRNA 중 어느 것을 발현하기 위한 유전적 구조물을 포함하는 세포를 포함하는 키트가 제공된다.

[0021] 본 발명의 다른 이점, 특징 및 용도는 본 발명의 특징의 비-제한적 실시양태의 상세한 설명; 일정한 비율로 도시되지 않고 도식적인 도면; 및 청구범위로부터 명백할 것이다.

도면의 간단한 설명

[0022] **도 1**은 본 발명의 특정 실시양태를 도식적으로 상세히 열거한 것이다. (A) 이 실시양태에서는, 뉴클레아제-불활성화 Cas9 단백질이 FokI 뉴클레아제 도메인의 단량체와 융합된다. 이중 가닥 DNA-절단은 표적 부위에서 FokI 단량체의 이량체화를 통하여 달성되고, 이는 2개의 별개의 Cas9:gRNA 복합체의 동시 결합에 의존적이다. (B) 이 실시양태에서는, 대체 입체배치가 제공되는데, 여기서는 2개의 별개의 gRNA 모티프를 함유하는 단일 연장된 gRNA의 작용을 통하여 2개의 Cas9-FokI 융합물이 조정된다. 이러한 gRNA 모티프는 별개의 영역 내의 표적과 혼성화되는 영역 뿐만 아니라 각 융합 단백질과 결합되는 영역을 포함한다. 상기 연장된 gRNA는 협조적 결합을 증강시킬 수 있고, 상기 융합물의 특이성 프로파일을 변경시킬 수 있다.

도 2는 본 발명의 특정 실시양태를 도식적으로 상세히 열거한 것이다. (A) 이 실시양태에서는, 이량체성 분할 Cas9가 A) gRNA-결합성 능력을 B) dsDNA 절단으로부터 분리시켜 준다. 단백질의 양 절반을 공동-국재하여 연합시키고 뉴클레아제-활성 상태가 되도록 재폴딩하는 경우에 DNA 절단이 일어난다. (B) 이 실시양태에서는, 뉴클레아제-불활성화 Cas9 돌연변이체가 분할 Cas9 뉴클레아제의 A-절반 (또는 일부 실시양태에서, B-절반)과 융합된다. 표적 부위에서 Cas9-A-절반 (또는 Cas9-B-절반) 융합물과 불활성 gRNA-결합성 Cas9 B-절반 (또는 A-절반) 둘 다가 각각 결합되면, dsDNA는 분할 단백질 재어셈블리를 수행할 수 있게 된다. 이러한 분할 Cas9-쌍 형성은 2개의 별개의 gRNA-결합성 Cas9 단백질을 이용하여, 분할 뉴클레아제-활성 Cas9가 반드시 정확한 표적 서열 상에서만 재어셈블리되도록 할 수 있다. (C) 이 실시양태에서는, 뉴클레아제-불활성화 Cas9 돌연변이체가 분할 Cas9 뉴클레아제의 A-절반과 융합된다. 별도의 뉴클레아제-불활성화 Cas9 돌연변이체는 분할 Cas9 뉴클레아제의 B-절반과 융합된다. 하나의 뉴클레아제-불활성화 Cas9 돌연변이체가 gRNA 표적 부위와 결합되고, 다른 뉴클레아제-불활성화 Cas9 돌연변이체가 제2의 gRNA 표적 부위와 결합되면, 분할 Cas9 절반은 이량체화될 수 있고 제3의 gRNA 표적과 결합되어, dsDNA를 절단할 수 있는 완전한 활성의 Cas9 뉴클레아제가 될 수 있다. 이러한 분할 Cas9-쌍 형성은 3개의 별개의 gRNA-결합성 Cas9 단백질을 이용하여, 분할 뉴클레아제-활성 Cas9가

반드시 정확한 표적 서열 상에서만 제어셈블리되도록 한다. 불활성 Cas9 대신 다른 DNA-결합성 도메인 (아연 핑거, TALE 단백질 등)을 이용하여 분할 Cas9 뉴클레아제의 제어셈블리를 완료할 수 있다.

도 3은 Cas9 활성을 위한 필수 도메인을 포함하는 최소한의 Cas9 단백질을 도식적으로 나타낸 것이다. 완전한 길이의 Cas9는 >150 kDa의 단백질을 초래하는 4.1 kb 유전자이다. 특이적 결실 및/또는 끝을 잘라내는 것은 Cas9의 활성 (예를 들어, gRNA 결합성 및 DNA 절단 활성)에는 영향을 미치지 않으면서 그의 크기를 감소시킨다. 이와 같이 최소화된 Cas9 단백질은, 예를 들어 바이러스 벡터, 예컨대 AAV (수용 서열 약 <4700 bp) 또는 렌티 바이러스 (수용 서열 약 <9 kb)를 이용하여 세포에 전달하는 효능을 증가시키거나, 또는 다중화 gRNA/Cas9 접근 방식을 추구하는 경우에 그 효능을 증가시킨다.

도 4는 2개의 Cas9 단백질이 2개의 별개의 gRNA 모티프를 함유하는 단일 연장된 RNA의 작용을 통하여 어떻게 조정될 수 있는지를 도시한 것이다. 각각의 gRNA 표적화 영역은 단일 Cas9:gRNA 단위가 그 자체에 의해서 효율적으로 결합될 수 없도록 단축된다. 정상적인 20 nt 표적화 서열은 그 일부분 (예를 들어, 5' 초기 10 nt)이 일부 비-특이적 링커 서열, 예컨대 AAAAAAAAAA (서열 13)로 변화되도록 변경시켰는데, gRNA의 10 nt 만이 여전히 표적 결합을 지시하고 있다 (또 다른 한편으로, 상기 5' 10 nt 가 완전히 잘려진다). 이러한 "저-친화성" gRNA 단위는 링커 서열에 의해 분리된, 제2의 별개의 저 친화성 gRNA 단위 하류를 수반한 탠덤(tandem) gRNA 구조물의 일부로서 존재한다. 일부 실시양태에서, 2개 초과와 저 친화성 gRNA 단위 (예를 들어, 3개 이상, 4개 이상, 5개 이상 등)가 존재한다. 일부 실시양태에서, 링커는 표적 핵산 상보성 서열을 포함한다 (예를 들어, DNA 표적과 접촉하는 링커 영역으로써 도시된 바와 같음).

도 5는 Cas9-재조합효소 융합물이 목적하는 서열 (부위)에서 표적 DNA와 결합 및 조합하기 위해 gRNA를 통하여 어떻게 조정될 수 있는지를 도식적으로 나타낸 것이다. (a, b) 뉴클레아제-불활성화 Cas9 (dCas9)단백질이 재조합효소 도메인 (Rec)의 단량체와 융합된다. 부위-특이적 재조합은 표적 부위에서 재조합효소 촉매 도메인 단량체를 이량체화 (a)한 다음, 별도의 Cas9-재조합 부위 상에 어셈블리된 2개의 이량체를 사랑체화 (b)함으로써 달성된다. dCas9:gRNA 복합체와의 융합물은 양옆에 위치하고 있는 표적 부위의 서열 실체를 결정하는 반면, 재조합효소 촉매 도메인은 코어 서열 (2개의 dCas9-결합 부위 사이의 서열)의 실체를 결정한다. (b) 재조합은 dCas9-재조합효소 사랑체 복합체 내에서의 가닥 절단, 교환, 및 재결합을 통하여 진행된다.

도 6은 Cas9와 *FokI*-dCas9 융합 변이체의 구조를 도시한 것이다. (a) 가이드 RNA (gRNA)와 복합체를 형성하는 Cas9 단백질이 표적 DNA와 결합된다. 에스. 피오게네스 (*S. pyogenes*) Cas9 단백질은 dsDNA의 풀림을 개시하고 gRNA:DNA 염기쌍 형성을 개시하는 PAM 서열 NGG를 인식한다. (b) *FokI*-dCas9 융합물 구조를 시험하였다. NLS, *FokI* 뉴클레아제, 및 dCas9의 4가지 별개의 입체배치를 어셈블리하였다. 17개의 단백질 링커 변이체를 또한 시험하였다. (c) gRNA 표적 부위를 GFP 내에서 시험하였다. 7개 gRNA 표적 부위를 선택하여, PAM이 상기 절단된 스페이스 서열로부터 멀리 떨어진 배향 (배향 A)에서 *FokI*-dCas9 활성을 시험하였다. 취합하면, 이들 7개 gRNA는 *FokI*-dCas9 융합 변이체가 5 내지 43 bp 범위의 스페이스 길이 전반에 걸쳐 시험될 수 있게 해주었다. PAM이 스페이스 서열에 인접한 배향 B를 시험하기 위해 사용된 가이드 RNA에 관해서는 도 9를 참조할 수 있다. (d) dCas9와 융합된 *FokI* 뉴클레아제의 단량체는 표적 유전자 자리 내의 별도의 부위와 결합된다. 인접하게 결합된 *FokI*-dCas9 단량체 만이 dsDNA 절단을 촉발시키는 촉매적 활성의 *FokI* 뉴클레아제 이량체를 어셈블리할 수 있다. (c)에 제시된 서열은 다음과 같이 확인된다: "EmGFP (bp 326-415)"는 서열 204에 상응하고; "G1"은 서열 205에 상응하며; "G2"는 서열 206에 상응하고; "G3"은 서열 207에 상응하며; "G4"는 서열 208에 상응하고; "G5"는 서열 209에 상응하며; "G6"은 서열 210에 상응하고; "G7"은 서열 211에 상응한다.

도 7은 fCas9, Cas9 니카제(nickase), 및 야생형 Cas9에 의한 게놈 DNA 변형을 나타낸 것이다. (a)는 gRNA를 수반하지 않거나, 또는 *GFP* 유전자를 배향 A로 표적화하는 가변 스페이스 길이의 gRNA 쌍을 수반한 fCas9, Cas9 니카제, 또는 야생형 Cas9의 GFP 붕괴 활성을 도시한 그래프를 나타낸다. (b)는 fCas9, Cas9 니카제 또는 야생형 Cas9, 및 *GFP* 부위를 표적화하는 14 bp 간격으로 떨어진 2개의 gRNA (gRNA G3 및 G7; 도 6c) (각각의 gRNA를 개별적으로 수반하거나, 또는 gRNA를 전혀 수반하지 않음)로 처리된 세포로부터 증폭된 재생한 표적 부위 DNA의 서베이어(Surveyor) 절단 검정의 PAGE 분석으로부터 삽입-결실 변형 효율(Indel modification efficiency)을 보여주는 겔의 영상이다. 이러한 삽입-결실 변형 비율(%)은 샘플에 대한 각 레인 아래에 제시되는데, 탐지 한계 (약 2%) 초과와 변형을 나타낸다. (c-g)는 (c) *GFP* 부위를 표적화하는 14 또는 25 bp 간격으로 떨어진 gRNA의 2개 쌍, (d) *CLTA* 부위를 표적화하는 19 bp 간격으로 떨어진 gRNA의 1개 쌍, (e) *EMX* 부위를 표적화하는 23 bp 간격으로 떨어진 gRNA의 1개 쌍, (f) *HBB* 부위를 표적화하는 16 bp 간격으로 떨어진 gRNA의 1개 쌍, 및 (g) *VEGF* 부위를 표적화하는 14 또는 16 bp 간격으로 떨어진 gRNA의 2개 쌍에 대한 삽입-결실 변형 효율을 도시한 그래프를 나타낸다. 오차 막대는 상이한 날에 수행된 3개의 생물학적 복제물로부터의 평균의 표

준 오차를 반영한다.

도 8은 fCas9, Cas9 니카제, 및 야생형 Cas9의 DNA 변형 특이성을 나타낸 것이다. (a)는 gRNA 쌍을 배향 A로 이용하여 야생형 Cas9, Cas9 니카제, 및 fCas9에 의한 GFP 유전자 붕괴를 도기한 그래프를 나타낸다. fCas9의 고 활성을 위해서는 약 15 및 25 bp의 스페이서 길이가 필요한데, 이는 대략 1회의 DNA 나선형 회전만큼 떨어져 있다. (b)는 gRNA 쌍을 배향 B로 이용하여 GFP 유전자 붕괴를 도기한 그래프를 나타낸다. Cas9 니카제는 gRNA 쌍의 어느 하나의 배향을 받아들이지만, fCas9는 그렇지 못하다. (c)는 2개의 gRNA의 존재에 의존하는, Cas9 니카제 또는 야생형 Cas9에 의해서가 아니라 fCas9에 의한 GFP 유전자 붕괴를 도기한 그래프를 나타낸다. 4개의 단일 gRNA를 다양한 스페이서 길이의 3개의 gRNA 쌍과 함께 시험하였다. 14 또는 25 bp의 스페이서 길이를 수반한 배향 A에 gRNA 쌍 (각각 gRNA 1+5, 및 gRNA 3+7)이 존재하면, fCas9는 활성이지만, 10-bp 스페이서를 수반한 gRNA 쌍 (gRNA 1+4)이 사용되는 경우에는 그렇지 않다. (a-c)에서, "처리하지 않음"은 세포에게 플라스미드 DNA를 전달하지 않는 것을 지칭한다. (d-f)는 fCas9, Cas9 니카제, 또는 야생형 Cas9와, (d) *CLTA* 부위를 표적화하는 19 bp 간격으로 떨어진 2개의 gRNA (gRNA C1 및 C2), (e) *EMX* 부위를 표적화하는 23 bp 간격으로 떨어진 2개의 gRNA (gRNA E1 및 E2), 또는 (f, g) *VEGF* 부위를 표적화하는 14 bp 간격으로 떨어진 2개의 gRNA (gRNA V1 및 V2)로 처리된 인간 세포로부터 증폭된 게놈의 표적에 적응한 부위와 표적을 벗어난 부위의 고 처리량 DNA 서열 분석으로부터 삽입-결실 돌연변이 빈도를 도기한 그래프를 나타낸다. (g)는 *VEGF* 표적을 벗어난 부위 1에서의 게놈 변형을 측정하기 위한 2가지 심층 시험을 도기한 그래프를 나타낸다. 시험 1은 각 샘플에 대한 $> 8 \times 10^5$ 개 서열 판독치 및 150 ng의 게놈 입력 DNA를 사용하였고; 시험 2는 각 샘플에 대한 $> 23 \times 10^5$ 개 서열 판독치 및 600 ng의 게놈 입력 DNA를 사용하였다. (d-g)에는, 모든 유의적 [P 값 < 0.005 피셔 (Fisher)의 직접 확률 시험] 삽입-결실 빈도가 제시된다. P 값은 표 3에 열거되어 있다. (d-f)의 경우, 표적에 적응한 샘플과 표적을 벗어난 샘플 각각을 1회 서열 분석하였는데, 표적에 적응한 샘플당 $> 10,000$ 개 서열이 분석되었고, 표적을 벗어난 샘플당 평균 76,260개 서열이 분석되었다 (표 3). (c)에 제시된 서열은 상단에서 하단으로, 다음과 같이 확인된다: 도 8c의 상단에서 발견된 서열은 서열 204에 상응하고; "G1"은 서열 205에 상응하며; "G3"은 서열 207에 상응하고; "G5"는 서열 209에 상응하며; "G7"은 서열 211에 상응하고; "G1+4"는 서열 205 및 서열 208에 상응하며; "G1+5"는 서열 205 및 서열 209에 상응하고; "G3+7"은 서열 207 및 서열 211에 상응한다.

도 9는 게놈 GFP 유전자 내의 표적 DNA 서열을 나타낸 것이다. 7개 gRNA 표적 부위를 선택하여, PAM이 상기 절단된 스페이서 서열과 인접한 배향 (배향 B)으로 FokI-dCas9 후보 활성을 시험하였다. 취합하면, 이들 7개 gRNA는 FokI-dCas9 융합 변이체가 4 내지 42 bp 범위의 6개 스페이서 길이 전반에 걸쳐 시험될 수 있게 해주었다. 제시된 서열은 다음과 같이 확인된다: "EmGFP (bp 297-388)"는 서열 212에 상응하고; "G8"은 서열 213에 상응하며; "G9"는 서열 214에 상응하고; "G10"은 서열 215에 상응하며; "G11"은 서열 216에 상응하고; "G12"는 서열 217에 상응하며; "G13"은 서열 218에 상응하고; "G14"는 서열 219에 상응한다.

도 10은 게놈 DNA-변형 활성을 측정하기 위한 GFP 붕괴 검정을 나타낸 것이다. (a)는 후보 FokI-dCas9 융합 구조물의 활성을 시험하기 위해 사용된 게놈적으로 통합된 EmGFP 유전자를 구성적으로 발현하는 HEK293-유래된 세포주를 도식적으로 도기한 것이다. 이들 세포를 적당한 뉴클레아제 및 gRNA 발현 플라스미드로 공동-형질감염시키면, EmGFP 코딩 서열 내에서 dsDNA 절단이 초래되어, 변이성 NHEJ가 자극되고 GFP의 발현을 붕괴시킬 수 있는 삽입-결실이 발생하여, 세포성 형광의 상실이 야기된다. 이어서, GFP 형광의 상실을 디스플레이하는 세포의 분획을 유동 세포계수법에 의해 정량화한다. (b)는 야생형 Cas9 및 gRNA 발현 플라스미드로 공동-형질감염시키기 전 및 후에 EmGFP-HEK293 세포의 200x 배율 하에 전형적인 표면형광(epifluorescence) 현미경 영상을 나타낸다.

도 11은 상이한 배향과 다양한 스페이서 길이의 gRNA 쌍과 조합된 FokI-dCas9 융합물 후보의 활성을 도기한 그래프를 나타낸 것이다. 도 6b에 기재된 융합물 구조를 대상으로 하여, 모든 (a) 배향 A gRNA 스페이서 및 (b) 배향 B gRNA 스페이서 (도 6c 및 도 9) 전반에 걸쳐 GFP 기능 상실 리포터를 이용하여 유동 세포계수법에 의해 기능성에 관하여 시험하였다. 제시된 모든 FokI-dCas9 융합물 데이터는 단일 시험에 따른 결과이다. 야생형 Cas9 및 Cas9 니카제 데이터는 2개 복제물의 평균인 반면, '처리하지 않음' 음성 대조군 데이터는 6개 복제물의 평균이며, 오차 막대는 하나의 표준 편차를 나타낸다. Y-축 전반에 걸친 회색 점선은 동일자로 수행된 '처리하지 않음' 대조군의 평균에 상응한다. "(GGG)x3"으로서 제시된 서열은 서열 14에 상응한다.

도 12는 NLS-FokI-dCas9에서의 단백질 링커의 최적화를 나타낸 것이다. (a)는 시험된 모든 링커 변이체의 표를 나타낸다. 야생형 Cas9 및 Cas9 니카제가 비교를 위해 포함되었다. FokI와 dCas9 사이에 (GGG)₃ (서열 14) 링

커를 수반한 초기 활성 구조물 NLS-FokI-dCas9을 특정 범위의 대체 링커 전반에 걸쳐 시험하였다. fCas9에 대한 링커의 최종 선택이 강조되었다. (b)는 링커 변이체를 수반한 FokI-dCas9 융합물의 활성을 도시한 그래프를 나타낸다. 각 변이체를 gRNA 쌍 배향 A를 이용하여 5 내지 43 bp 범위의 스페이서 길이 전반에 걸쳐 시험하였다. gRNA가 결합된 대조군 ("gRNA 없음")이 각각의 별도의 융합 구조물에 포함되었다. NLS-FokI-dCas9 변이체 L8이 가장 우수한 활성을 나타냈는데, 이는 Cas9 니카제의 활성에 근접하다. 변이체 L4 내지 L9는 14-bp 및 25-bp 스페이서 길이를 수반한 피크 활성을 나타내는데, 이는 2개의 최적 스페이서 길이가 대략 dsDNA의 1회의 나선형 회전만큼 떨어져 있다는 것을 제안하고 있다. (a)에 제시된 서열은 다음과 같이 확인된다: GSGSGSGGS는 서열 14에 상응하고; GSGSGSGSGSGSGSGGS는 서열 15에 상응하며; MKIIEQLPSA는 서열 22에 상응하고; VRHKLKRVGS는 서열 23에 상응하고; VPFLLEPDNINGKTC는 서열 19에 상응하며; GHGTGTSGSGSS는 서열 24에 상응하고; MSRPDPA는 서열 25에 상응하며; GSAGSAAGSGEF는 서열 20에 상응하고; SGSETPGTSESA는 서열 17에 상응하며; SGSETPGTSESATPES는 서열 16에 상응하고; SGSETPGTSESATPEGGSGGS는 서열 18에 상응하며; GGSN은 서열 301에 상응하고; SIVAQLSRPDPA는 서열 21에 상응한다.

도 13은 내인성 인간 EMX, VEGF, CLTA, 및 HBB 유전자 내의 표적 DNA 서열을 나타낸 것이다. 내인성 인간 EMX, VEGF, CLTA, 및 HBB 유전자 내에서 시험된 gRNA 표적 부위가 제시되어 있다. 13개 gRNA 표적 부위를 선택하여, PAM이 상기 절단된 스페이서 서열로부터 멀리 떨어진 배향 (배향 A)으로 최적화된 fCas9 융합물의 활성을 시험하였다. 취합하면, 이들 13개 gRNA는 fCas9 융합 변이체가 5 내지 47 bp 범위의 8개 스페이서 길이 전반에 걸쳐 시험될 수 있게 해주었다. 제시된 서열은 다음과 같이 확인된다: "CLTA-1"은 서열 220에 상응하고; "C1"은 서열 221에 상응하며; "C2"는 서열 222에 상응하고; "C3"은 서열 224에 상응하며; "C4"는 서열 225에 상응하고; "HBC"는 서열 226에 상응하며; "H1"은 서열 227에 상응하고; "H2"는 서열 228에 상응하며; "H3"은 서열 229에 상응하고; "H4"는 서열 230에 상응하며; "H5"는 서열 231에 상응하고; "H6"은 서열 232에 상응하며; "H7"은 서열 233에 상응하고; "EMX"는 서열 234에 상응하며; "E1"은 서열 235에 상응하고; "E2"는 서열 236에 상응하며; "E3"은 서열 237에 상응하고; "VEGF"는 서열 238에 상응하며; "V1"은 서열 239에 상응하고; "V2"는 서열 240에 상응하며; "V3"은 서열 241에 상응하고; "V4"는 서열 242에 상응한다.

도 14는 fCas9, Cas9 니카제, 및 야생형 Cas9에 의한 게놈 DNA 변형의 스페이서 길이 선호도를 도시하는 그래프를 나타낸 것이다. (a) GFP 부위를 표적화하는 gRNA의 쌍, (b) CLTA 부위를 표적화하는 gRNA의 쌍, (c) EMX 부위를 표적화하는 gRNA의 쌍, (d) HBB 부위를 표적화하는 gRNA의 쌍, 및 (e) VEGF 부위를 표적화하는 gRNA의 쌍에 대한 삽입-결실 변형 효율. 오차 막대는 상이한 날에 수행된 3개의 생물학적 복제물로부터의 평균의 표준 오차를 반영한다.

도 15는 다양한 양의 Cas9 및 gRNA 발현 플라스미드를 이용한 fCas9, Cas9 니카제, 및 야생형 Cas9에 의한 게놈 DNA 변형의 효율을 도시한 그래프를 나타낸 것이다. fCas9, Cas9 니카제 또는 야생형 Cas9와 2개의 표적 부위 gRNA로 처리된 세포 집단으로부터 증폭된 재생한 표적-부위 DNA의 서베이어 검정으로부터의 삽입-결실 변형 효율. 700 ng의 Cas9 발현 플라스미드와 250 ng의 gRNA 발현 플라스미드 (총 950 ng), 350 ng의 Cas9 발현 플라스미드와 125 ng의 gRNA 발현 플라스미드 (총 475 ng), 175 ng의 Cas9 발현 플라스미드와 62.5 ng의 gRNA 발현 플라스미드 (총 238 ng) 또는 88 ng의 Cas9 발현 플라스미드와 31 ng의 gRNA 발현 플라스미드 (총 119 ng)를 적당한 양의 불활성 캐리어 플라스미드로 형질감염시켜 모든 처리군 전반에 걸쳐 950 ng의 플라스미드의 균일한 형질감염을 보장하였다. (a) CLTA 부위를 표적화하는 19-bp 간격으로 떨어진 gRNA, (b) EMX 부위를 표적화하는 23-bp 간격으로 떨어진 gRNA, 및 (c) VEGF 부위를 표적화하는 14-bp 간격으로 떨어진 gRNA에 대한 삽입-결실 변형 효율. 오차 막대는 별도의 날에 수행된 3개의 생물학적 복제물로부터의 평균의 표준 오차를 나타낸다.

도 16은 단일 gRNA의 존재 하에 게놈 DNA를 변형시킬 수 있는 fCas9, Cas9 니카제, 및 야생형 Cas9의 능력을 나타낸 것이다. (A)는 Cas9 단백질과 gRNA(들)의 표시된 조합물로 처리된 세포의 DNA로부터의 게놈 GFP 표적의 서베이어 검정을 도시하는 겔의 영상을 나타낸 것이다. 단일 gRNA는 fCas9와 Cas9 니카제 둘 다에 대하여 탐지 가능한 수준 (< 2% 변형)에서 게놈 변형을 유도하지 않는다. 야생형 Cas9는 시험된 모든 단일 gRNA와 쌍을 이룬 gRNA에 대하여 GFP 표적을 효과적으로 변형시킨다. fCas9와 Cas9 니카제 둘 다의 경우에는, 적당하게 쌍을 형성한 gRNA가 야생형 Cas9와 거의 동등한 수준으로 게놈 변형을 유도시킨다. (B)는 야생형 Cas9, Cas9 니카제, 또는 fCas9 중 어느 하나를 발현하는 플라스미드와 단일 gRNA (G1, G3, G5 또는 G7)를 발현하는 단일 플라스미드, 또는 각각 상이한 gRNA (G1+G5, 또는 G3+G7)를 발현하는 2개의 플라스미드로 처리된 인간 세포로부터 단리된 150 ng 게놈 DNA로부터 증폭된 GFP 표적에 적중한 부위를 서열 분석한 결과를 도시한 그래프를 나타낸다. 음성 대조군으로서, 어떠한 gRNA 발현 플라스미드도 이용하지 않는 상기와 같은 형질감염 및 서열 분석을 3중으로 수행하였다. 오차 막대는 s.d.를 나타낸다. GFP 표적 부위에 1개 초과 삽입 또는 결실을 수반한

서열 (G1 결합성 부위에서 시작하여 G7 결합성 부위에서 끝난다)이 삽입-결실로서 간주된다. 삽입-결실 비율 (%)은 삽입-결실의 수를 서열 총 수로 나눔으로써 계산하였다. 야생형 Cas9가 모든 gRNA 처리군 전반에 걸쳐 삽입-결실을 야기시켰지만, fCas9와 Cas9 니카제는 쌍을 이룬 gRNA가 존재하는 경우에만 삽입-결실을 효율적으로 야기시켰다 (> 1%). fCas9와 단일 gRNA에 의해 유도된 삽입-결실은 gRNA 없는 대조군 상에서는 탐지되지 않았지만, Cas9 니카제와 단일 gRNA는 표적 GFP 서열을 평균 0.12%의 비율로 변형시켰다.

도 17은 게놈 표적의 fCas9 삽입-결실 빈도가 gRNA 쌍 스페이서 길이 선호도에 어떻게 영향을 미치는지를 도시하는 그래프를 나타낸 것이다. 이러한 그래프는 야생형 Cas9 뉴클레아제와 공동-발현된 동일한 gRNA의 삽입-결실 변형 효율에 대해 표준화시킨 fCas9의 삽입-결실 변형 효율과 스페이서 길이 (두 gRNA 간의 bp 수) 간의 관계를 보여준다. X-축 아래의 착색 삼각형은 시험되긴 하였지만, 표시된 표적 유전자에 대해 탐지 가능한 삽입-결실을 전혀 산출하지 않았던 스페이서 길이를 의미한다. 이들 결과는 fCas9가 DNA를 효율적으로 절단하기 위해 절반-부위 사이에 약 15 bp 또는 약 25 bp를 필요로 한다는 것을 제안한다.

도 18은 내인성 유전자 자리에서 Cas9 뉴클레아제, Cas9 니카제, 또는 fCas9 뉴클레아제에 의해 유도된 변형을 나타낸 것이다. (a)는 야생형 Cas9 뉴클레아제, Cas9 니카제, 또는 fCas9 뉴클레아제, 및 VEGF 표적에 적중한 부위를 표적화하는 2개의 gRNA (gRNA V1 및 gRNA V2)를 발현하는 단일 플라스미드를 이용하여 VEGF 표적에 적중한 부위에서 변형된 서열의 예를 나타낸다. 제시된 각 예에 대하여, 변형되지 않은 게놈 부위가 첫 번째 서열이고, 그 다음이 결실을 함유하는 상단 8개 서열이다. 각 서열 앞의 숫자는 서열 분석 계수치를 표시한다. gRNA 표적 부위는 볼드체의 대문자로 표기된다. (b)는 VEGF 표적을 벗어난 부위 1VEG_Off1에 대하여 (a)에서와 동일한 분석이다. (c)는 VEGF 표적을 벗어난 부위 1에 대한 2개의 gRNA의 잠재적 결합 방식을 나타낸다. 상단 가닥은 표준 방식으로 결합되는 반면, 하단 가닥은 G:U 염기 쌍을 포함하는 gRNA:DNA 염기 쌍형성을 통하여 제2의 gRNA인 gRNA V2와 결합된다. (a)에 제시된 서열은 상단에서 하단으로, 다음과 같이 확인된다: 서열 243; 서열 244; 서열 245; 서열 246; 서열 247; 서열 248; 서열 249; 서열 250; 서열 251; 서열 252; 서열 253; 서열 254; 서열 255; 서열 256; 서열 257; 서열 258; 서열 259; 서열 260; 서열 261; 서열 262; 서열 263; 서열 264; 서열 265; 서열 266; 서열 267; 서열 268; 및 서열 269. (b)에 제시된 서열은 상단에서 하단으로, 다음과 같이 확인된다: 서열 270; 서열 271; 서열 272; 서열 273; 서열 274; 서열 275; 서열 276; 서열 277; 서열 278; 서열 279; 서열 280; 서열 281; 서열 282; 서열 283; 서열 284; 서열 285; 서열 286; 서열 287; 서열 288; 서열 289; 서열 290; 서열 291; 서열 292; 서열 293; 서열 294; 서열 295; 및 서열 296. (c)에 제시된 서열은 상단에서 하단으로, 다음과 같이 확인된다: 서열 297; 서열 298; 서열 299; 및 서열 300.

도 19는 게놈 CCR5 유전자 내에서의 표적 DNA 서열을 나타낸 것이다. (a) PAM이 상기 절단된 스페이서 서열과 인접한 배향 (배향 A)으로 Cas9 변이체 (예를 들어, FokI-dCas9) 활성을 시험하기 위하여 8개의 gRNA 표적 부위를 확인하였다. (b) PAM이 상기 절단된 스페이서 서열과 인접한 배향 (배향 B)으로 Cas9 변이체 (예를 들어, FokI-dCas9) 활성을 시험하기 위하여 6개의 gRNA 표적 부위를 확인하였다. 취합하면, 이들 14개 gRNA는 Cas9 융합 변이체가 0 내지 74 bp 범위의 스페이서 길이 전반에 걸쳐 시험될 수 있게 해준다. (a)에 제시된 서열은 다음과 같이 확인된다: "CRA"는 서열 302에 상응하고; "CRA-1"은 서열 303에 상응하며; "CRA-2"는 서열 304에 상응하며; "CRA-3"은 서열 305에 상응하고; "CRA-4"는 서열 306에 상응하며; "CRA-5"는 서열 307에 상응하고; "CRA-6"은 서열 308에 상응하며; "CRA-7"은 서열 309에 상응하고; "CRA-8"은 서열 310에 상응한다. (b)에 제시된 서열은 다음과 같이 확인된다: "CRB"는 서열 311에 상응하고; "CB-1"은 서열 312에 상응하며; "CB-2"는 서열 313에 상응하고; "CB-3"은 서열 314에 상응하며; "CB-4"는 서열 315에 상응하고; "CB-5"는 서열 316에 상응하며; "CB-6"은 서열 317에 상응한다.

도 20은 FokI-dCas9 (fCas9) 구조물을 함유하는 예시되는 플라스미드를 상세히 열거하는 벡터 맵을 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0023] 정의

본원 및 청구범위에 사용된 바와 같은, 단수 형태는 문맥상 달리 명백하게 표시하지 않는 한은 단수 및 복수의 언급 대상을 포함한다. 따라서, 예를 들어, "작용제"에 대한 언급은 단일 작용제 및 복수 개의 상기 작용제를 포함한다.

용어 "Cas9" 또는 "Cas9 뉴클레아제"는 Cas9 단백질, 또는 그의 단편 (예를 들어, Cas9의 활성 또는 불활성 DNA 절단 도메인, 및/또는 Cas9의 gRNA 결합성 도메인을 포함하는 단백질)을 포함하는 RNA-가이드된 뉴클레아제를

지칭한다. Cas9 뉴클레아제는 또한, 종종 casn1 뉴클레아제 또는 CRISPR [클러스터링되고 규칙적으로 사이에 공간을 남겨둔 짧은 팔린드롬성(palindromic) 반복 서열]-관련 뉴클레아제로서 지칭된다. CRISPR은 이동성 유전적 요소 (바이러스, 전위 요소 및 접합 플라스미드)에 대항하여 보호를 제공하는 적응 면역 체계이다. CRISPR 클러스터는 스페이서, 선행 이동 요소에 대해 상보적인 서열, 및 표적 침입 핵산을 함유한다. CRISPR 클러스터는 CRISPR RNA (crRNA)로 전사되고 프로세싱된다. 유형 II CRISPR 시스템에서 pre-crRNA를 정확하게 프로세싱하기 위해서는, 트랜스-코딩된 소형 RNA (tracrRNA), 내인성 리보뉴클레아제 3 (rnc) 및 Cas9 단백질이 필요하다. 상기 tracrRNA는 crRNA 전구의 리보뉴클레아제 3-지원 프로세싱에 대한 가이드로서 제공된다. 연속해서, Cas9/crRNA/tracrRNA는 상기 스페이서에 대해 상보적인 선행 또는 환상 dsDNA 표적을 내부핵산분해적으로 절단한다. crRNA에 대해 상보적이지 않은 표적 가닥을 먼저, 내부핵산분해적으로 커팅한 다음, 3'-5' 외부핵산 분해적으로 트리밍한다. 사실상, DNA-결합과 절단을 위해서는, 전형적으로 단백질과 둘 다의 RNA가 필요하다. 그러나, 단일 가이드 RNA ("sgRNA", 또는 간단히 "gRNA")는 crRNA와 tracrRNA 둘 다의 측면이 단일 RNA 종 내로 혼입되도록 조작될 수 있다 (예를 들어, 문헌 [Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012)] 참조; 그의 전문이 본원에 참조로 포함된다). Cas9는 자기 대 비-자기를 구별하는 데 도움을 주기 위해 CRISPR 반복 서열 내의 짧은 모티프 (PAM 또는 프로토스페이서 인접 모티프)를 인식한다. Cas9 뉴클레아제 서열과 구조는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다 (예를 들어, 문헌 ["Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti et al., J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001)]; ["CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011)]; 및 ["A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012)] 참조; 각각의 전문이 본원에 참조로 포함된다). Cas9 오르소로그(ortholog)가 에스. 피오케네스 및 에스. 더모필루스 (*S. thermophilus*)를 포함하지만, 이에 제한되지 않는 각종 종에서 보고되었다. 부가의 적합한 Cas9 뉴클레아제 및 서열이 본 개시내용을 근거로 하여 관련 기술분야의 통상의 기술자에게 명백할 것이며, 이러한 Cas9 뉴클레아제 및 서열은 문헌 [Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; 그의 전문이 본원에 참조로 포함된다]에 개시된 유기체 및 유전자 자리로부터의 Cas9 서열을 포함한다. 일부 실시양태에서, Cas9 뉴클레아제는 불활성 (예를 들어, 불활성화) DNA 절단 도메인을 갖는다. 뉴클레아제-불활성화 Cas9 단백질은 "dCas9" 단백질 (뉴클레아제 "사멸" Cas9에 대함)로서 상호 교환적으로 지칭될 수 있다. 일부 실시양태에서, dCas9는 부분으로든 전체로든, 다음 서열 5로서 제시된 아미노산 세트에 상응하거나 이를 포함한다. 일부 실시양태에서, dCas9의 변이체 (예를 들어, 서열 5의 변이체)가 제공된다. 예를 들어, 일부 실시양태에서, 예를 들어 뉴클레아제 불활성화 Cas9 (dCas9)를 초래하는, D10A 및 H840A 이외의 돌연변이를 갖는 변이체가 제공된다. 이러한 돌연변이는 한 예로서, D10 및 H840에서의 다른 아미노산 치환, 또는 Cas9의 뉴클레아제 도메인 내에서의 다른 치환 (예를 들어, HNH 뉴클레아제 서브도메인 및/또는 RuvC1 서브도메인 내에서의 치환)을 포함한다. 일부 실시양태에서, 서열 5와 약 70% 이상 동일하거나, 약 80% 이상 동일하거나, 약 90% 이상 동일하거나, 약 95% 이상 동일하거나, 약 98% 이상 동일하거나, 약 99% 이상 동일하거나, 약 99.5% 이상 동일하거나, 또는 약 99.9% 이상 동일한 dCas9의 변이체 또는 동족체 (예를 들어, 서열 5의 변이체)가 제공된다. 일부 실시양태에서, 서열 5 보다 약 5개 아미노산, 약 10개 아미노산, 약 15개 아미노산, 약 20개 아미노산, 약 25개 아미노산, 약 30개 아미노산, 약 40개 아미노산, 약 50개 아미노산, 약 75개 아미노산, 약 100개 아미노산 또는 그 초과 정도가 더 짧거나 더 긴 아미노산 서열을 갖는 dCas9의 변이체 (예를 들어, 서열 5의 변이체)가 제공된다.

[0026] dCas9 (D10A 및 H840A):

```
MDKKYSIGLAIGTNSVGVAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKFLIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRLENL
IAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKQLSKDLYDDDLNLLAQIGDQYADFLAAKNLSDAI
LLSDILRVNTEITKAPLSAMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVLKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
IKDKDFLDNEENEDILEDIVLTTLTFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIAKKGILQTVKVVDELVKV
MGRHKPENIVIEARENQTTQKGQKNSRERMKRIEIEGKELGSQLKEHPVENTQLQNEKLYLYLQNGRDMYVD
QELDINRLSDYDVAIVPQSFLKDDSIDNKVLTSDKNRGSNDNPSEEVVKMKMNYRQLLNAKLITQRKFDNL
TKAERGGSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALEIKKPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVAKVEKGSKSKLKSVELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIEQISE
FSKRVLADANLDKVL SAYNKHDKPIREQAENIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (서열 5)
```

[0027]

[0028] 불활성 DNA 절단 도메인을 갖는 Cas9 단백질 (또는 그의 단편)을 생성시키는 방법은 공지되어 있다 (예를 들어, 본 실시예; 및 문헌 [Jinek et al., *Science*. 337:816-821(2012)]; [Qi et al., "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell*. 28;152(5):1173-83] 참조; 각각의 전문이 본원에 참조로 포함된다). 예를 들어, Cas9의 DNA 절단 도메인은 2개의 서브도메인, 즉 HNH 뉴클레아제 서브도메인 및 RuvC1 서브도메인을 포함하는 것으로 공지되어 있다. HNH 서브도메인은 gRNA에 대해 상보적인 가닥을 절단하는 반면, RuvC1 서브도메인은 비-상보적 가닥을 절단한다. 이들 서브도메인 내의 돌연변이는 Cas9의 뉴클레아제 활성을 침묵시킬 수 있다. 예를 들어, 돌연변이 D10A 및 H840A는 에스. 피오게네스 Cas9의 뉴클레아제 활성을 완전히 불활성화시킨다 (예를 들어, 본 실시예; 및 문헌 [Jinek et al., *Science*. 337:816-821(2012)]; [Qi et al., *Cell*. 28;152(5):1173-83 (2013)] 참조). 일부 실시양태에서, Cas9의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 실시양태에서, 단백질은 다음 2개의 Cas9 도메인 중 하나를 포함한다: (1) Cas9의 gRNA 결합성 도메인; 또는 (2) Cas9의 DNA 절단 도메인. 일부 실시양태에서, Cas9 또는 그의 단편을 포함하는 단백질은 "Cas9 변이체"로서 지칭된다. Cas9 변이체는 Cas9, 또는 그의 단편에 대한 상동성을 공유한다. 예를 들어, Cas9 변이체는 야생형 Cas9와 약 70% 이상 동일하거나, 약 80% 이상 동일하거나, 약 90% 이상 동일하거나, 약 95% 이상 동일하거나, 약 98% 이상 동일하거나, 약 99% 이상 동일하거나, 약 99.5% 이상 동일하거나, 또는 약 99.9% 이상 동일하다. 일부 실시양태에서, Cas9 변이체는 Cas9의 단편 (예를 들어, gRNA 결합성 도메인 또는 DNA-절단 도메인)을 포함하므로, 이러한 단편은 야생형 Cas9의 상응하는 단편과 약 70% 이상 동일하거나, 약 80% 이상 동일하거나, 약 90% 이상 동일하거나, 약 95% 이상 동일하거나, 약 98% 이상 동일하거나, 약 99% 이상 동일하거나, 약 99.5% 이상 동일하거나, 또는 약 99.9% 이상 동일하다. 일부 실시양태에서, 야생형 Cas9는 스트렙토코쿠스 피오게네스 (*Streptococcus pyogenes*)로부터의 Cas9 (NCBI 참조 서열: NC_017053.1, 서열 1 (뉴클레오티드); 서열 2 (아미노산))에 상응한다.

```
ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATG
ATTATAAGGTTCCGCTCAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAATCT
TATAGGGGCTCTTTTATTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGA
AGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAG
ATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGGAAGAACAAGAAGCATGAACGTCATCC
TATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAATCTATCATCTGCGAAAA
AAATTGGCAGATTCTACTGATAAAGCGGATTTCGCTTAATCTATTGGCCTTAGCGCATATGATTAAGT
TTCGTGGTCATTTTTGATTGAGGAGATTTAAATCCTGATAATAGTGATGTGGACAACTATTTATCCA
GTTGGTACAAATCTACAATCAATTTTGAAGAAAACCTTATTAACGCAAGTAGAGTAGATGCTAAAGCG
ATTCTTTCGACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGTGAGAAGA
GAAATGGCTTGGTTGGGAATCTCATTGCTTTGTGCTTGGGATTGACCCCTAATTTTAAATCAAAATTTGA
TTTGGCAGAAGATGCTAAATACAGCTTTCAAAGATACCTACGATGATGATTAGATAAATTTATTGGCG
CAAATTGGAGATCAATATGCTGATTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTCA
ATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCTATCAGCTTCAATGATTAAGCGCTACGATGA
ACATCATCAAGACTTGACTCTTTTAAAGCTTTAGTTCGACAACAACCTCCAGAAAAGTATAAAGAAATC
TTTTTGTATCAAAACGAGATATGACAGGTTATATTGATGGGGGAGCTAGCCAAAGAAGAAATTTATA
AATTTATCAACCAATTTTGAAGAAAATGGATGGTACTGAGGAATTATTGGTGAACATAAATCGTGAAGA
TTTGCTGCGCAAGCAACGACCTTTGACAACGGCTCTATTTCCCATCAAAATCACTTGGGTGAGCTGCAT
GCTATTTTGAAGACAGAAGACTTTTATCCATTTTAAAGACAATCGTGAGAAGATTGAAAAAATCT
TGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCTGATGACTCG
GAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCA
```

[0029]

TTTATTGAACGCATGACAACTTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGC
 TTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAAACC
 AGCATTTCCTTTCAGGTGAACAGAGAAAGCCATTGTTGATTTACTCTTCAAACAAATCGAAAAAGTAACC
 GTTAAGCAATTAAGAAGATTATTTCAAAAAATAGAAATGTTTTGATAGTGTGAAATTTTCAGGAGTTG
 AAGATAGATTAAATGCTTCATTAGGCGCCTACCATGATTGCTAAAAATATTAAAGATAAAGATTTTTT
 GGATAATGAAGAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTGGAAGATAGGGGG
 ATGATTGAGGAAAGACTTAAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTC
 GCCGTATACTGGTTGGGACGCTTTGTCTCGAAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAA
 AACAAATATTAGATTTTTTGAATCAGATGGTTTTGCCAATCGCAATTTATGCAGCTGATCCATGATGAT
 AGTTTGACATTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGA
 TTGCTAACTTAGCTGGCAGTCTGCTATTAAAAAGGTATTTTACAGACTGTAAAAATTTGTTGATGAAC
 GGTCAAAGTAATGGGGCATAAGCCAGAAAAATATCGTTATTGAAATGGCACGTGAAAAATCAGACAACTCAA
 AAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGA
 TTCTTAAAGAGCATCCTGTTGAAAAATCTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAA
 TGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAAGTATTGATGATGTCGATCACATT
 GTTCCACAAAGTTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAAATCGTG
 GTAAATCGGATAACGTTCCAAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACCTCTAAA
 CGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGGAGTGAAC
 GTTCAAAAGCTGGTTTTATCAAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAATTT
 TGGATAGTCGCATGAATACTAAATACGATGAAAAATGATAAACTTATTGAGAGGTTAAAGTGATTACCTT
 AAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTCCAAATCTATAAAGTACGTGAGATTAAACAATTAC
 CATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACGCTTTGATTAAAGAAATATCCAAACTTG
 AATCGGAGTTTGTCTATGGTGATATATAAGTTTATGATGTTTCGTAAGTATTGCTAAAGTCTGAGCAAGA
 AATAGGCAAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACCTCTTCAAAACAGAAATTACA
 CTGCAAAATGGAGAGATTTCGCAACGCGCTCTAATCGAAACTAATGGGGAACCTGGAGAAATTTGCTGGG
 ATAAAGGGCGAGATTTTGCCACAGTGCAGCAAGTATTGCCATGCCCAAGTCAATATTGTCAAGAAAC
 AGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAGAAATTCGGACAAGCTTATTGCT
 CGTAAAAAGACTGGGATCCAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAG
 TGGTTGCTAAGGTGGAAGGGAATCGAAGAAGTTAAATCCGTTAAAGAGTTACTAGGGATCACAAAT
 TATGGAAGAGATTTCCCTTTGAAAAAATCCGATTGACTTTTAGAAGCTAAAGGATATAAGGAAGTTAAA
 AAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG
 CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAAATATGTGAATTTTTTATATTT
 AGCTAGTCATTATGAAAAGTTGAAGGTTAGTCCAGAAGATAACGAACAAAAACAATTTGTTGTGGAGCAG
 CATAAGCATTATTTAGATGAGATTATTGAGCAATCAGTGAATTTCTAAGCGTGTATTTTAGCAGATG
 CCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAAGCAGAAAA
 TATTATTCAATTATTACGTTGACGAATCTGGAGCTCCGCTGCTTTTAAATATTTTGATACAACAATT
 GATCGTAAACGATATACGCTTACAAAAAGAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTC
 TTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (서열 1)

MDKKYSIGLDIGTNSVGWAVITDDYKVPSSKFKVLGNTDRHSIKKNLIGALLFGSGETAETRLKRTARR
 RYTRRKNRICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
 KLAADSDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFILQVQIYNQLFEENPINASRVDAKA
 ILSARLSKSRRLLENLIAQLPGEKRNLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDLDNLLA
 QIGDQYADFLFAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
 AILRRQEDFYFPLKDNREKIEKILTFRIPIYVGPLARGNSRFAMWTRKSEETITPWNFEVVVDKGASAQ
 FIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
 VKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRG
 MIEERLKYTAHLFDDKVMQLKRRRYTGWRLSRKLINGIRDKQSGKITLDFLKSDGFANRNFMLIHDD
 SLTFKEDIQKAQVSGQGHSHEQIANLAGSPAIIKKGILQTVKIVDELVKVMGHKPNINIVEMARENQTTQ
 KGQKNSRERMKRIEKGELGSQLKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHI
 VPQSFIKDDSIDKNVLTSDKNRGSNDNVPSEEVVKMKNYWRQLLNKLTITQRKFDNLTKAERGGLSEL
 DKAGFIKRQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNY
 HHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEI GKATAKYFFYSNIMNFKTEIT
 LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKTEVQTGGFSKESILPKRNSDKLIA
 RKKDWDPKKYGGFDSPTVAYSVLVVAKEKGSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVK
 KDLIIKLPKYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLKGPSPEDNEQKQLFVEQ
 HKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
 DRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGD (서열 2)

[0030]

[0031]

일부 실시양태에서, 야생형 Cas9는 서열 3 (뉴클레오티드) 및/또는 서열 4 (아미노산)에 상응하거나 이를 포함한다.

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCCTAATCCGTTGGATGGGCTGTCATAACCGATGAATAC
 AAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGCTATTTCGATTAAAAAGAATCTTATCGGTGCC
 CTCTATTTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTGCG
 AAGAACCAGATATGTTACTTACAAGAAATTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTACCGT
 TTGGAAGACTCTTCCCTGTGCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAG
 GTGGCATAATCATGAAAAGTACCAACGATTATCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
 CTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACCTTCTCATTGAGGGTGATCTAAAT
 CCGGACAACCTCGGATGTGCAAACTGTTTATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
 ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACTG
 ATCGCACAAATTACCCGGAGAGAAGAAAAATGGGTGTTGTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCA
 AATTTTAAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTC
 GACAATCTACTGGCACAATTTGGAGATCAGTATGCGGACTTATTTTGGGTGCCCCAAAACCTTAGCGATGCAATC
 ATCTTATCTGACATCTGAGAGTTAATACTGAGATTACCAAGGCGCCGTATCCGCTTCAATGATCAAAAGGTAC
 GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTGAGCAACTGCCTGAGAAATATAAGGAAATA
 TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTT
 ATCAAAACCATATTAGAGAAGATGGATGGGACGGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGA
 ATTCACCGGACTTTCGACATGAGCATTCCACATCAATCCACTTAGGCGAATTGCTATGCTATACTTAGAAGG
 CAGGAGGATTTTTATCCGTTCTCAAAGACAATCGTGAAGGATTGAGAAAACTCTAACCTTTTCGCATACCTTAC
 TATGTGGGACCCCTGGCCCGAGGGAACCTCTCGGTTGCGATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
 TGGAATTTTGAAGAGTTGTGCGATAAAGGTGCGTCAGCTCAATCGTTTATCGAGAGGATGACCAACTTTGACAAG
 AATTTACCGAAGTCAAGAAAGATGCTTAAGCAGATTACTTTACGAGTATTTACAGTGTAACATGAACACTCAG
 AAAGTTAAGTATGTCAGTGGGCGATGCGTAACCCGCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGAT
 CTGTTATTCAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAGAGGACTACTTTAAGAAATTTGAATGCTTC
 GATTCTGTCGAGATCTCCGGGTAGAAGATCGATTAAATGCGTCACCTTGGTACGTATCATGACCTCTCTAAAGATA
 ATTAAGATAACGAACTTCGTGATAACGAAGATAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTC
 TTTGAAGATCGGGAATGATTGAGGAAGACATAAAACATACGCTCACCTGTTGACGATAAGGTTATGAAACAG
 TTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTGCGGAAACTTATCAACGGGATAAGAGACAAGCAAGT
 GGTAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACCTTATGACGCTGATCCATGATGAC
 CTCTTAACCTTCAAAGAGGATATACAAAGGCACAGGTTTCCGGACAAGGGGACTCATGACGAAACATATTGCG
 AATCTTGCTGTTGCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTC
 ATGGGACGTCACAAACCGGAAAAACATTGTAATCGAGATGGCAGCGAAAATCAAACGACTCAGAAGGGGCAAAAA
 AACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGTTAATAAGAACTGGGCGAGCCAGATCTTAAAGGAGCATCCT
 TGGGAAAAATACCAAGTTCGAGAGGAGTACAAAGGCAAGAACTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
 CAGGAAGTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGAT
 TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCAGGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTC
 GTAAAGAAAAATGAAGAACTATTGGCGGCGAGCTCTTAATGCGAAACTGATAACGCAAGAAAGTTTCGATAACTTA
 ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACCTTGACAAGGCCGAGTTTATTAACGTCAGCTCGTGGAACCCGC
 CAAATCAAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATT
 CGGGAAGTCAAAGTAATCACTTTAAAGTCAAATTTGGTGTGCGACTTCAGAAAGGATTTCAATTCTATAAAGTT
 AGGGAGATAAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAAGAAA
 TACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCTGTAAGATGATCGCGAAAAAGC
 GAACAGGAGATAGGCAAGGCTACAGCCAAATACCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAATC
 ACTCTGGCAAACGGAGAGATACGCAACGACCTTTAATTGAACCAATGGGGAGACAGGTGAAATCGTATGGGAT
 AAGGGCCGGGACTTCGCGACGGTGAGAAAAAGTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTG
 CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGAC
 TGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCATTCTGTCTTAGTAGTGGCAAAAAGTTGAG
 AAGGGAAAAATCCAAGAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAA
 AAGAACCCCATCGACTTCCTTGAAGCGAAAGGTTACAGGAAGTAAAAAGGATCTCATAATTAACTACCAAAG
 TATAGTCTGTTTGAAGTTAGAAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGAGAGCTTCAAAGGGGAACGAA
 CTCGCACTACCGTCTAAATACGTGAATTTCTGTATTATAGCGTCCATTACGAGAAGTTGAAAGGTTTACCTGAA
 GATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTGACGAAATCATAGAGCAAAATTCGGAA
 TTCAGTAAGAGAGTATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAA
 CCAATACGTGAGCAGGCGGAAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGATTCAAG
 TATTTTGACACACATAGATCGCAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTACCCAA
 TCCATCAGGGATTATATGAAACTCGGATAGATTGTGACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGG
 AAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC
 AAGGCTGCAGGA (서열 3)

[0032]

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLQTYNQLFEEENPINASGVDAKILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDYDDDLNLLAQIGDQYADFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
KGTILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPY
YVGPLARGNSRFAMWTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKLI
IKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTIILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKV
MGRHKPENIVEMARENQTQKGQKNSRERMKRIEKGELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVD
QELDINRLSDYDVIDHIVPQSFLKDDSIDNKVLTSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGSELDDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVGTALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTETVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVAVKEKGSKKLKSVKELLGITIMERSSEFNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLKGPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVL SAYNKRDRKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ
SITGLYETRIDLSQLGGD (서열 4)

[0033]

[0034]

일부 실시양태에서, Cas9는 다음으로부터의 Cas9를 지칭한다: 코리네박테륨 울세란스 (*Corynebacterium ulcerans*) (NCBI Refs: NC_015683.1, NC_017317.1); 코리네박테륨 디프테리아 (*Corynebacterium diphtheria*) (NCBI Refs: NC_016782.1, NC_016786.1); 스피로플라스마 시르피디콜라 (*Spiroplasma syrrhpidicola*) (NCBI Ref: NC_021284.1); 프레보텔라 인테르메디아 (*Prevotella intermedia*) (NCBI Ref: NC_017861.1); 스피로플라스마 타이와넨세 (*Spiroplasma taiwanense*) (NCBI Ref: NC_021846.1); 스트렙토코쿠스 이니아에 (*Streptococcus iniae*) (NCBI Ref: NC_021314.1); 벨리엘라 발티카 (*Belliella baltica*) (NCBI Ref: NC_018010.1); 사이크로플렉수스 토르퀴시 (*Psychroflexus torquis*) (NCBI Ref: NC_018721.1); 스트렙토코쿠스 더모필루스 (*Streptococcus thermophilus*) (NCBI Ref: YP_820832.1); 리스테리아 인노쿠아 (*Listeria innocua*) (NCBI Ref: NP_472073.1); 캄필로박터 제주니 (*Campylobacter jejuni*) (NCBI Ref: YP_002344900.1) 또는 네이세리아 메닝기티디스 (*Neisseria meningitidis*) (NCBI Ref: YP_002342100.1).

[0035]

용어 "접합하는", "접합된" 및 "접합"은 두 실체, 예를 들어 두 분자, 예컨대 두 단백질, 두 도메인 (예를 들어, 결합성 도메인과 절단 도메인), 또는 단백질과 작용제, 예를 들어 단백질 결합성 도메인과 소분자의 연합을 지칭한다. 일부 측면에서, 이러한 연합은 단백질 (예를 들어, RNA-프로그램 가능한 뉴클레아제)과 핵산 (예를 들어, 가이드 RNA) 간의 연합이다. 상기 연합은, 예를 들어 직접 또는 간접 (예를 들어, 링커를 포함) 공유 연쇄를 통해서 이루어질 수 있다. 일부 실시양태에서, 상기 연합은 공유적이다. 일부 실시양태에서, 두 분자는 양 분자를 연결하는 링커를 통하여 접합된다. 예를 들어, 두 단백질이 서로 접합하여, 예를 들어 결합성 도메인과 조작된 뉴클레아제의 절단 도메인이 접합되어 단백질 융합물을 형성하는 일부 실시양태에서, 이러한 두 단백질은 폴리펩티드 링커, 예를 들어 하나의 단백질의 C-말단을 다른 단백질의 N-말단과 연결시켜 주는 아미노산 서열을 통하여 접합될 수 있다.

[0036]

핵산 서열의 맥락에서 본원에 사용된 바와 같은 용어 "컨센서스 서열"은 복수 개의 유사한 서열 내의 각 위치에서 발견된 가장 빈번한 뉴클레오티드 잔기를 나타내는 계산된 서열을 지칭한다. 전형적으로, 컨센서스 서열은 유사한 서열을 서로 비교하고 유사한 서열 모티프를 계산하는 서열 정렬에 의해 결정된다. 뉴클레아제 표적 부위 서열의 맥락에서, 뉴클레아제 표적 부위의 컨센서스 서열은 일부 실시양태에서, 소정의 뉴클레아제에 의해 가장 빈번하게 결합되거나 또는 가장 높은 친화도로 결합된 서열일 수 있다. 재조합효소 표적 부위 서열의 맥락에서, 재조합효소 표적 부위의 컨센서스 서열은 일부 실시양태에서, 소정의 재조합효소에 의해 가장 빈번하게 결합되거나 또는 가장 높은 친화도로 결합된 서열일 수 있다.

[0037]

본원에 사용된 바와 같은 용어 "조작된"은 인간에 의해 설계, 생성, 제조, 합성 및/또는 제작된 바 있는 단백질 분자, 핵산, 복합체, 물질 또는 실체를 지칭한다. 따라서, 조작된 생성물은 자연에서는 존재하지 않는 생성물이다.

[0038]

본원에 사용된 바와 같은 용어 "유효량"은 목적하는 생물학적 반응을 유발시키기에 충분한 생물학적 활성체의 양을 지칭한다. 예를 들어, 일부 실시양태에서, 뉴클레아제의 유효량은 이러한 뉴클레아제에 의해 특이적으로 결합되고 절단된 표적 부위의 절단을 유도시키기에 충분한 뉴클레아제의 양을 지칭할 수 있다. 일부 실시양태에서, 재조합효소의 유효량은 이러한 재조합효소에 의해 특이적으로 결합되고 재조합된 표적 부위에서 재조합을 유도시키기에 충분한 재조합효소의 양을 지칭할 수 있다. 통상의 기술자에 의해 인지되는 바와 같이, 특정 작용제, 예를 들어 뉴클레아제, 재조합효소, 혼성체 단백질, 융합 단백질, 단백질 이량체, 단백질 (또는 단백질

이량체)과 폴리뉴클레오티드의 복합체, 또는 폴리뉴클레오티드의 유효량은 각종 인자, 예를 들어 목적하는 생물학적 반응, 특이적 대립 유전자, 계놈, 표적 부위, 표적화되는 세포 또는 조직, 및 사용되는 작용제에 따라 다양할 수 있다.

[0039] 본원에 사용된 바와 같은 용어 "상동성"은 뉴클레오티드 및/또는 아미노산 서열의 수준에서 고도로 관련이 있는 핵산 또는 폴리펩티드를 지칭하는 것으로 관련 기술분야에서 이해되는 용어이다. 서로 상동성인 핵산 또는 폴리펩티드는 "동족체"로 지칭된다. 두 서열 간의 상동성은 관련 기술분야의 통상의 기술자에게 공지된 서열 정렬 방법에 의해 결정될 수 있다. 본 발명에 따라서, 두 서열이 20개 이상, 30개 이상, 40개 이상, 50개 이상, 60개 이상, 70개 이상, 80개 이상, 90개 이상, 100개 이상, 120개 이상, 150개 이상, 또는 200개 이상의 아미노산의 하나 이상의 연장물에 대하여, 약 50 내지 60% 이상 동일하거나, 예를 들어 하나의 서열 또는 다른 서열에 포함된 모든 잔기의 약 50 내지 60% 이상에서 동일한 잔기 (예를 들어, 아미노산 잔기)를 공유하거나, 약 70% 이상 동일하거나, 약 80% 이상 동일하거나, 약 90% 이상 동일하거나, 약 95% 이상 동일하거나, 약 98% 이상 동일하거나, 약 99% 이상 동일하거나, 약 99.5% 이상 동일하거나, 또는 약 99.9% 이상 동일한 경우에, 이들 두 서열은 상동성인 것으로 간주된다.

[0040] 본원에 사용된 바와 같은 용어 "링커"는 2개의 인접한 분자 또는 모이어티(moiety), 예를 들어 뉴클레아제 (예를 들어, FokI)의 절단 도메인과 결합성 도메인 (예를 들어, dCas9)을 연결하는 분자 또는 화학 기를 지칭한다. 일부 실시양태에서, 링커는 핵 국제화 신호 (NLS) 도메인을 또 다른 단백질 (예를 들어, Cas9 단백질 또는 뉴클레아제 또는 제조합효소 또는 그의 융합물)과 연결시켜 준다. 일부 실시양태에서, 링커는 RNA-프로그램 가능한 뉴클레아제의 gRNA 결합성 도메인과 제조합효소의 촉매 도메인을 연결시켜 준다. 일부 실시양태에서, 링커는 dCas9와 제조합효소를 연결시켜 준다. 전형적으로, 링커는 2개 기, 분자 또는 기타 모이어티 사이에 위치되거나 또는 이들이 양옆에 위치하고 있고, 공유 결합을 통하여 각자 연결되므로, 둘을 연결시켜 준다. 일부 실시양태에서, 링커는 아미노산 또는 복수 개의 아미노산 (예를 들어, 펩티드 또는 단백질)이다. 일부 실시양태에서, 링커는 유기 분자, 기, 중합체 또는 화학적 모이어티이다. 일부 실시양태에서, 링커는 펩티드 링커이다. 일부 실시양태에서, 펩티드 링커는 1개 이상, 2개 이상, 3개 이상, 4개 이상, 5개 이상, 6개 이상, 7개 이상, 8개 이상, 9개 이상, 10개 이상, 15개 이상, 20개 이상, 25개 이상, 30개 이상, 40개 이상, 50개 이상의 아미노산을 갖는, 아미노산의 모든 연장물이다. 일부 실시양태에서, 펩티드 링커는 트리-펩티드 Gly-Gly-Ser, 예를 들어 서열 (GGG)_n (여기서, n은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 반복 서열을 나타낸다)을 포함하는 반복 서열을 포함한다. 일부 실시양태에서, 링커는 서열 (GGG)₆ (서열 15)을 포함한다. 일부 실시양태에서, 펩티드 링커는 16개 잔기 "XTEN" 링커, 또는 그의 변이체이다 (예를 들어, 본 실시예; 및 문헌 [Schellenberger et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* 27, 1186-1190 (2009)] 참조). 일부 실시양태에서, XTEN 링커는 서열 SGSETPGTSESATPES (서열 16), SGSETPGTSESA (서열 17), 또는 SGSETPGTSESATPEGSGGS (서열 18)을 포함한다. 일부 실시양태에서, 펩티드 링커는 도 12a에 제공된 바와 같은 모든 링커, 예를 들어 VPFLLEPDNINGKTC (서열 19), GSAGSAAGSGEF (서열 20), SIVAQLSRPDPA (서열 21), MKIIEQLPSA (서열 22), VRHKLKRVGS (서열 23), GHGTGSTGSGSS (서열 24), MSRPDPA (서열 25); 또는 GGSM (서열 301)으로부터 선택된 하나 이상이다.

[0041] 본원에 사용된 바와 같은 용어 "돌연변이"는 특정 서열, 예를 들어 핵산 또는 아미노산 내의 특정 잔기를 또 다른 잔기로 치환시키거나, 또는 특정 서열 내의 하나 이상의 잔기를 결실 또는 삽입하는 것을 지칭한다. 돌연변이는 전형적으로, 본래의 잔기를 확인한 다음, 서열 내에서의 이러한 잔기의 위치를 확인하고 새로이 치환된 잔기의 실체를 확인함으로써 본원에 기재된다. 본원에 제공된 아미노산 치환 (돌연변이)을 만들기 위한 각종 방법이 관련 기술분야에 널리 공지되어 있고, 예를 들어 문헌 [Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 의해 제공된다.

[0042] 본원에 사용된 바와 같은 용어 "뉴클레아제"는 핵산 분자 내의 2개의 뉴클레오티드 잔기를 연결시켜 주는 포스포디에스테르 결합을 절단할 수 있는 작용제, 예를 들어 단백질을 지칭한다. 일부 실시양태에서, "뉴클레아제"는 불활성 DNA 절단 도메인을 갖는 단백질을 지칭하므로, 이러한 뉴클레아제는 포스포디에스테르 결합을 절단할 수 없다. 일부 실시양태에서, 뉴클레아제는 단백질, 예를 들어 핵산 분자와 결합할 수 있고 이러한 핵산 분자 내의 뉴클레오티드 잔기를 연결시켜 주는 포스포디에스테르 결합을 절단할 수 있는 효소이다. 뉴클레아제는 폴리뉴클레오티드 쇠 내의 포스포디에스테르 결합을 절단하는 엔도뉴클레아제, 또는 폴리뉴클레오티드 쇠의 말단에서 포스포디에스테르 결합을 절단하는 엑소뉴클레아제일 수 있다. 일부 실시양태에서, 뉴클레아제는 "인

식 서열", "뉴클레아제 표적 부위" 또는 "표적 부위"로서 본원에 지칭되기도 하는, 특이적 뉴클레오티드 서열 내의 특이적 포스포디에스테르 결합과 결합하고/하거나 이러한 결합을 절단하는 부위-특이적 뉴클레아제이다. 일부 실시양태에서, 뉴클레아제는 표적 부위를 보완하는 서열을 갖는 RNA (예를 들어, 가이드 RNA인 "gRNA")와 연합됨으로써 (예를 들어, 이와 결합됨으로써) 뉴클레아제의 서열 특이성을 제공하는 RNA-가이드된 (즉, RNA-프로그래밍 가능한) 뉴클레아제이다. 일부 실시양태에서, 뉴클레아제는 단일 가닥 표적 부위를 인식하는 반면, 다른 실시양태에서는, 뉴클레아제가 이중 가닥 표적 부위, 예를 들어 이중 가닥 DNA 표적 부위를 인식한다. 많은 자연 발생적 뉴클레아제, 예를 들어 많은 자연 발생적 DNA 제한 뉴클레아제의 표적 부위는 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다. 많은 경우에 있어서, DNA 뉴클레아제, 예컨대 EcoRI, HindIII, 또는 BamHI는 4 내지 10개 염기 쌍 길이의 팔린드롬성 이중 가닥 DNA 표적 부위를 인식하고, 이러한 표적 부위 내의 특이적 위치에서 2개의 DNA 가닥 각각을 커팅한다. 일부 엔도뉴클레아제는 이중 가닥 핵산 표적 부위를 대칭적으로 커팅하는데, 즉 양 가닥을 동일한 위치에서 커팅하여 그 말단이 염기 쌍 형성된 뉴클레오티드를 포함하도록 한다 (이는 본원에서 평행 말단으로서 지칭되기도 함). 기타 엔도뉴클레아제는 이중 가닥 핵산 표적 부위를 비대칭적으로 커팅하는데, 즉 각 가닥을 상이한 위치에서 커팅하여 그 말단이 쌍을 형성하지 않은 뉴클레오티드를 포함하도록 한다. 이중 가닥 DNA 분자의 말단에 쌍을 형성하지 않은 뉴클레오티드는 또한, "오버행"으로서 지칭되는데, 예를 들어 쌍을 형성하지 않은 뉴클레오티드(들)가 각각의 DNA 가닥의 5' 또는 3' 말단을 형성하는 지에 따라서 "5'-오버행"으로서 또는 "3'-오버행"으로서 지칭된다. 쌍을 형성하지 않은 뉴클레오티드(들)로 종결되는 이중 가닥 DNA 분자 말단은 또한, 점착성 말단으로서 지칭되는데, 이는 이들이 쌍을 형성하지 않은 상보적 뉴클레오티드(들)를 포함하는 다른 이중 가닥 DNA 분자 말단에 "점착"될 수 있기 때문이다. 뉴클레아제 단백질은 전형적으로, 단백질과 핵산 기질의 상호 작용을 매개하고, 또한 일부 경우에는, 표적 부위와 특이적으로 결합하는 "결합성 도메인"과, 핵산 골격 내의 포스포디에스테르 결합의 절단을 촉매하는 "절단 도메인"을 포함한다. 일부 실시양태에서 뉴클레아제 단백질은 단량체성 형태의 핵산 분자와 결합하고 이를 절단할 수 있는 반면, 다른 실시양태에서, 뉴클레아제 단백질은 표적 핵산 분자를 절단하기 위하여 이량체화되거나 또는 다량체화되어야 한다. 자연 발생적 뉴클레아제의 결합성 도메인 및 절단 도메인 뿐만 아니라 뉴클레아제 결합 특이성 표적 부위를 창출시키기 위해 융합될 수 있는 모듈(modular) 결합성 도메인 및 절단 도메인은 관련 기술분야의 통상의 기술자에게 널리 공지되어 있다. 예를 들어, RNA-프로그래밍 가능한 뉴클레아제 (예를 들어, Cas9), 또는 불활성 DNA 절단 도메인을 갖는 Cas9 단백질의 결합성 도메인을 결합성 도메인 (예를 들어, 표적 부위에 대한 결합을 지시하는 gRNA와 결합한다)으로서 사용하여, 목적하는 표적 부위와 특이적으로 결합할 수 있고, 절단 도메인, 예를 들어 FokI의 절단 도메인과 융합 또는 접합되어, 표적 부위를 절단하는 조작된 뉴클레아제를 창출시킬 수 있다. 일부 실시양태에서, 본원에 제공된 Cas9 융합 단백질은 FokI의 절단 도메인을 포함하므로, "fCas9" 단백질로서 지칭된다. 일부 실시양태에서, 예를 들어 fCas9 단백질에서의 FokI의 절단 도메인은 부분으로든 전체로든, 다음에 서열 6으로서 제시된 아미노산 서열 (또는 그의 변이체)에 상응하거나 이를 포함한다. 일부 실시양태에서, FokI 절단 도메인의 변이체 또는 동족체는 표적 핵산 내의 표적 부위에서 또 다른 FokI 절단 도메인과 이량체화함으로써 (예를 들어, fCas9 융합 단백질의 일부로서), 표적 핵산의 절단을 초래할 수 있는 임의의 변이체 또는 동족체를 포함한다. 일부 실시양태에서, 서열 6과 약 70% 이상 동일하거나, 약 80% 이상 동일하거나, 약 90% 이상 동일하거나, 약 95% 이상 동일하거나, 약 98% 이상 동일하거나, 약 99% 이상 동일하거나, 약 99.5% 이상 동일하거나, 또는 약 99.9% 이상 동일한 FokI 절단 도메인의 변이체 (예를 들어, 서열 6의 변이체)가 제공된다. 일부 실시양태에서, 서열 6 보다 약 5개 아미노산, 약 10개 아미노산, 약 15개 아미노산, 약 20개 아미노산, 약 25개 아미노산, 약 30개 아미노산, 약 40개 아미노산, 약 50개 아미노산, 약 75개 아미노산, 약 100개 아미노산 또는 그 초과 정도가 더 짧거나 더 긴 아미노산 서열을 갖는 FokI 절단 도메인의 변이체 (예를 들어, 서열 6의 변이체)가 제공된다.

[0043] FokI의 절단 도메인:

```
GSQLVKSELEEKKSRLRHKLKYPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTV
GSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWKVPSSVTEFKFLFVSGHFKGNYKAQ
LTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF (서열 6)
```

[0044]

[0045]

본원에 사용된 바와 같은 용어 "핵산" 및 "핵산 분자"는 뉴클레오타이드 및 산성 모이어티를 포함하는 화합물, 예를 들어 뉴클레오시드, 뉴클레오티드, 또는 뉴클레오티드의 중합체를 지칭한다. 전형적으로, 중합체성 핵산, 예를 들어 3개 이상의 뉴클레오티드를 포함하는 핵산 분자는 선형 분자인데, 여기서는 인접한 뉴클레오티드가 포스포디에스테르 연쇄를 통하여 서로 연결된다. 일부 실시양태에서, "핵산"은 개개의 핵산 잔기 (예를 들어 뉴클레오티드 및/또는 뉴클레오시드)를 지칭한다. 일부 실시양태에서, "핵산"은 3개 이상의 개개의 뉴클레오티드 잔기를 포함하는 올리고뉴클레오티드 세를 지칭한다. 본원에 사용된 바와 같은 용어 "올리고뉴클레오티드"

및 "폴리뉴클레오티드"는 뉴클레오티드의 중합체 (예를 들어, 3개 이상의 일련의 뉴클레오티드)를 지칭하기 위해 상호 교환적으로 사용될 수 있다. 일부 실시양태에서, "핵산"은 RNA 뿐만 아니라 단일 및/또는 이중 가닥 DNA를 포괄한다. 핵산은, 예를 들어 게놈, 전사물, mRNA, tRNA, rRNA, siRNA, snRNA, gRNA, 플라스미드, 코스미드, 염색체, 염색분체, 또는 기타 자연 발생적 핵산 분자의 맥락에서 자연 발생적일 수 있다. 다른 한편으로, 핵산 분자는 비-자연 발생적 분자, 예를 들어 재조합 DNA 또는 RNA, 인공 염색체, 조작된 게놈, 또는 그의 단편, 또는 합성 DNA, RNA, DNA/RNA 혼성체일 수 있거나 또는 비-자연 발생적 뉴클레오티드 또는 뉴클레오타이드를 포함할 수 있다. 더욱이, 용어 "핵산", "DNA", "RNA" 및/또는 유사한 용어는 핵산 유사체, 즉 포스포디에스테르 골격 이외의 것을 갖는 유사체를 포함한다. 핵산은 자연 공급원으로부터 정제될 수 있고, 재조합 발현 시스템을 이용하여 생성될 수 있으며, 임의로 정제되고, 화학적으로 합성될 수 있다. 적당한 경우, 예를 들어, 화학적으로 합성된 분자의 경우에, 핵산은 뉴클레오타이드 유사체, 예컨대 화학적으로 변형된 염기 또는 당을 갖는 유사체, 및 골격 변형을 포함할 수 있다. 핵산 서열은 달리 표시되지 않는 한 5'에서 3' 방향으로 제시된다. 일부 실시양태에서, 핵산은 자연 뉴클레오타이드 (예를 들어, 아데노신, 티미딘, 구아노신, 시티딘, 우리딘, 데옥시아데노신, 데옥시티미딘, 데옥시구아노신, 및 데옥시시티딘); 뉴클레오타이드 유사체 (예를 들어, 2-아미노아데노신, 2-티오티미딘, 이노신, 피콜로-피리미딘, 3-메틸아데노신, 5-메틸시티딘, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-요오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-시티딘, C5-메틸시티딘, 2-아미노아데노신, 7-테아자아데노신, 7-테아자구아노신, 8-옥소아데노신, 8-옥소구아노신, 0(6)-메틸구아닌 및 2-티오시티딘); 화학적으로 변형된 염기; 생물학적 변형된 염기 (예를 들어, 메틸화 염기); 삽입된 염기; 변형된 당 (예를 들어, 2'-플루오로리보스, 리보스, 2'-데옥시리보스, 아라비노스 및 핵소스); 및/또는 변형된 인산염기 (예를 들어, 포스포로티오에이트 및 5'-N-포스포르아미다이트 연쇄)이거나 이를 포함한다.

[0046] 본원에 사용된 바와 같은 용어 "제약 조성물"은 특정 질환 또는 장애의 치료 및/또는 예방의 맥락에서 특정 대상체에게 투여될 수 있는 조성물을 지칭한다. 일부 실시양태에서, 제약 조성물은 활성 성분, 예를 들어 Cas9 단백질과 융합된 뉴클레아제 또는 재조합효소, 또는 그의 단편 (또는 이러한 융합물을 코딩하는 핵산), 및 임의의 제약상 허용되는 부형제를 포함한다. 일부 실시양태에서, 제약 조성물은 본 발명의 Cas9 변이체/융합물 (예를 들어, fCas9) 단백질(들)과, 이러한 Cas9 변이체/융합 단백질(들)이 표적 핵산을 표적화하는 데 적합한 gRNA(들)를 포함한다. 일부 실시양태에서, 표적 핵산은 특정 유전자이다. 일부 실시양태에서, 표적 핵산은 특정 질환과 연관된 대립 유전자이므로, 이러한 대립 유전자는 Cas9 변이체/융합 단백질(들)의 작용에 의해 절단된다. 일부 실시양태에서, 대립 유전자는 *CLTA* 유전자, *EMX* 유전자, *HBB* 유전자, *VEGF* 유전자, 또는 *CCR5* 유전자의 대립 유전자이다 (예를 들어, 본 실시예; 도 7, 8, 13, 14, 15, 17 및 19 참조).

[0047] 본원에 사용된 바와 같은 용어 "증식성 질환"은 특정 세포 또는 세포 집단이 비정상적으로 상승된 증식률을 나타낸다는 점에서 세포 또는 조직 생체 항상성이 교란되는 모든 질환을 지칭한다. 증식성 질환은 과증식성 질환, 예컨대 전-종양 과형성 병태 및 종양 질환을 포함한다. 종양 질환은 세포의 비정상적인 증식을 특징으로 하고, 양성 종양과 악성 종양 둘 다를 포함한다. 악성 종양은 또한, 암으로서 지칭된다. 일부 실시양태에서, 본원에 제공된 조성물 및 방법은 증식성 질환을 치료하는 데 유용하다. 예를 들어, 일부 실시양태에서, 제약 조성물은 Cas9 (예를 들어, fCas9) 단백질(들)과, 이러한 Cas9 단백질(들)이 *VEGF* 대립 유전자를 표적화하는 데 적합한 gRNA(들)를 포함함으로써, 상기 대립 유전자가 Cas9 단백질(들)의 작용에 의해 불활성화된다 (예를 들어, 본 실시예 참조).

[0048] 용어 "단백질", "펩티드" 및 "폴리펩티드"는 본원에서 상호 교환적으로 사용되고, 펩티드 (아미드) 결합에 의해 함께 연결된 아미노산 잔기의 중합체를 지칭한다. 상기 용어는 모든 크기, 구조 또는 기능의 단백질, 펩티드, 또는 폴리펩티드를 지칭한다. 전형적으로, 단백질, 펩티드, 또는 폴리펩티드는 3개 이상의 아미노산 길이일 것이다. 단백질, 펩티드, 또는 폴리펩티드는 개개의 단백질, 또는 단백질의 컬렉션을 지칭할 수 있다. 단백질, 펩티드, 또는 폴리펩티드 내의 하나 이상의 아미노산은, 예를 들어 화학적 실체, 예컨대 탄수화물 기, 히드록실기, 인산염기, 파르네실기, 이소파르네실기, 지방산기; 접합, 기능화 또는 기타 변형을 위한 링커 등을 부가함으로써 변형될 수 있다. 단백질, 펩티드, 또는 폴리펩티드는 또한, 단일 분자일 수 있거나 또는 다중-분자 복합체일 수 있다. 단백질, 펩티드, 또는 폴리펩티드는 단지, 자연 발생적 단백질 또는 펩티드의 단편일 수 있다. 단백질, 펩티드, 또는 폴리펩티드는 자연 발생적, 재조합, 또는 합성, 또는 그의 모든 조합일 수 있다. 본원에 사용된 바와 같은 용어 "융합 단백질"은 2가지 이상의 상이한 단백질로부터의 단백질 도메인을 포함하는 혼성체 폴리펩티드를 지칭한다. 하나의 단백질은 융합 단백질의 아미노-말단 (N-말단) 부분 또는 카르복시-말단 (C-말단) 부분에 위치할 수 있으므로, "아미노-말단 융합 단백질" 또는 "카르복시-말단 융합 단백질"을 각각 형성할 수 있다. 본원에 제공된 단백질 중 어느 것도 관련 기술분야에 공지된 모든 방법에 의해서 생성될 수 있다. 예를 들어, 본원에 제공된 단백질은 재조합 단백질 발현 및 정제를 통하여 생성될 수 있는데, 이는 펩티

드 링커를 포함하는 융합 단백질에 대해서 특히 적합하다. 재조합 단백질 발현 및 정제를 위한 방법은 널리 공지되어 있고, 이는 문헌 [Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)); 그의 전문이 본원에 참조로 포함된다]에 기재된 것을 포함한다.

[0049]

용어 "RNA-프로그램 가능한 뉴클레아제" 및 "RNA-가이드된 뉴클레아제"는 본원에서 상호 교환적으로 사용되고, 절단에 대한 표적이 아닌 하나 이상의 RNA와 복합체를 형성하는 (예를 들어, 이러한 RNA와 결합되거나 연합되는) 뉴클레아제를 지칭한다. 일부 실시양태에서, RNA와 복합체를 형성하는 경우의 RNA-프로그램 가능한 뉴클레아제는 뉴클레아제:RNA 복합체로서 지칭될 수 있다. 전형적으로, 이와 같이 결합된 RNA(들)는 가이드 RNA (gRNA)로서 지칭된다. gRNA는 2개 이상의 RNA의 복합체로서 존재하거나 또는 단일 RNA 분자로서 존재할 수 있다. 단일 RNA 분자로서 존재하는 gRNA는 단일-가이드 RNA (sgRNA)로서 지칭될 수 있지만, "gRNA"는 단일 분자로서 또는 2개 이상의 분자의 복합체로서 존재하는 가이드 RNA를 지칭하기 위해 상호 교환적으로 사용된다. 전형적으로, 단일 RNA 종으로서 존재하는 gRNA는 다음 2개의 도메인을 포함한다: (1) 표적 핵산과의 상동성을 공유하는 (예를 들어, Cas9 복합체가 상기 표적과 결합하는 것을 지시한다) 도메인; 및 (2) Cas9 단백질과 결합하는 도메인. 일부 실시양태에서, 도메인 (2)은 tracrRNA로서 공지된 서열에 상응하고, 줄기-고리 구조를 포함한다. 예를 들어, 일부 실시양태에서, 도메인 (2)은 문헌 [Jinek et al., *Science* 337:816-821(2012); 그의 전문이 본원에 참조로 포함된다]의 도 1E에 도시된 바와 같이 tracrRNA와 상동성이다. gRNA의 다른 예 (예를 들어, 도메인 2를 포함한 예)는 "전환 가능한 Cas9 뉴클레아제 및 그의 용도"란 발명의 명칭으로, 2013년 9월 6일에 출원된 미국 가출원 번호 61/874,682; "기능성 뉴클레아제에 대한 전달 시스템"이란 발명의 명칭으로, 2013년 9월 6일에 출원된 미국 가출원 번호 61/874,746; "RNA-지시된 표적 DNA 변형 및 전사의 RNA-지시된 조절을 위한 방법 및 조성물"이란 발명의 명칭으로, 2013년 3월 15일에 출원된 PCT 출원 WO 2013/176722; 및 "Cas9-crRNA 복합체에 의한 RNA-지시된 DNA 절단"이란 발명의 명칭으로 2013년 3월 20일에 출원된 PCT 출원 WO 2013/142578 (이들 각각의 전문이 본원에 참조로 포함된다)에서 발견될 수 있다. 또한 gRNA 및 gRNA 구조의 다른 예가 본원에 제공된다 (예를 들어, 본 실시예 참조). 일부 실시양태에서, gRNA는 도메인 (1) 및 (2) 중 2개 이상을 포함하고, 이는 "연장된 gRNA"로서 지칭될 수 있다. 예를 들어, 연장된 gRNA는, 본원에 기재된 바와 같이, 예를 들어 2개 이상의 Cas9 단백질과 결합될 것이고, 2개 이상의 별개의 영역에서 표적 핵산과 결합될 것이다. 이러한 gRNA는 특정 표적 부위에 대한 뉴클레아제/RNA 복합체의 결합을 매개하는 표적 부위를 보완하여, 뉴클레아제:RNA 복합체의 서열 특이성을 제공하는 뉴클레오티드 서열을 포함한다. 일부 실시양태에서, RNA-프로그램 가능한 뉴클레아제는 (CRISPR-관련 시스템) Cas9 엔도뉴클레아제, 예를 들어 스트렙토코쿠스 피오케네스로부터의 Cas9 (Csn1)이다 (예를 들어, 문헌 ["Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001)]; ["CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011)]; 및 ["A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012)] 참조; 각각의 전문이 본원에 참조로 포함된다).

[0050]

RNA-프로그램 가능한 뉴클레아제 (예를 들어, Cas9)는 RNA:DNA 혼성화를 이용하여 표적 DNA 절단 부위를 결정하기 때문에, 이들 단백질은 원칙상, 가이드 RNA에 의해 지정된 모든 서열을 절단할 수 있다. 부위-특이적 절단을 위해 (예를 들어, 게놈을 변형시키기 위해) RNA-프로그램 가능한 뉴클레아제, 예컨대 Cas9를 사용하는 방법은 관련 기술분야에 공지되어 있다 (예를 들어, 문헌 [Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013)]; [Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* 339, 823-826 (2013)]; [Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* 31, 227-229 (2013)]; [Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013)]; [Dicarlo, J.E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* (2013)]; [Jiang, W. et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* 31, 233-239 (2013)] 참조; 각각의 전문이 본원에 참조로 포함된다).

[0051]

본원에 사용된 바와 같은 용어 "재조합효소"는 재조합효소 인식 서열들 간의 DNA 단편의 절제, 통합, 역위 또는

교환 (예를 들어, 전위)을 초래하는, 재조합효소 인식 서열들 간의 DNA의 재조합을 매개하는 부위-특이적 효소를 지칭한다. 재조합효소는 2가지 별개의 계열로 분류될 수 있다: 세린 재조합효소 (예를 들어, 위치 특이성 재조합 촉진 효소 및 전효소) 및 티로신 재조합효소 [예를 들어, 인테그라제(integrase)]. 세린 재조합효소의 예는 Hin, Gin, Tn3, β -six, CinH, ParA, γ δ , Bxb1, Φ C31, TP901, TG1, Φ BT1, R4, Φ RV1, Φ FC1, MR11, A118, U153, 및 gp29를 포함하지만, 이에 제한되지 않는다. 티로신 재조합효소의 예는 Cre, FLP, R, 람다, HK101, HK022, 및 pSAM2를 포함하지만, 이에 제한되지 않는다. 세린 및 티로신 재조합효소 명칭은 재조합 효소가 DNA를 공격하기 위해 사용되고 가닥 교환 동안 DNA와 공유적으로 연결되는, 보존된 친핵성 아미노산 잔기에 기인한다. 재조합효소는 유전자 녹아웃/녹인의 창출 및 유전자 요법 적용을 포함한 수많은 적용을 갖는다 (예를 들어, 문헌 [Brown et al., "Serine recombinases as tools for genome engineering." *Methods*. 2011;53(4):372-9]; [Hirano et al., "Site-specific recombinases as tools for heterologous gene integration." *Appl. Microbiol. Biotechnol.* 2011; 92(2):227-39]; [Chavez and Calos, "Therapeutic applications of the Φ C31 integrase system." *Curr. Gene Ther.* 2011;11(5):375-81]; [Turan and Bode, "Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications." *FASEB J.* 2011; 25(12):4088-107]; [Venken and Bellen, "Genome-wide manipulations of *Drosophila melanogaster* with transposons, Flp recombinase, and Φ C31 integrase." *Methods Mol. Biol.* 2012; 859:203-28]; [Murphy, "Phage recombinases and their applications." *Adv. Virus Res.* 2012; 83:367-414]; [Zhang et al., "Conditional gene manipulation: Cre-ating a new biological era." *J. Zhejiang Univ. Sci. B.* 2012; 13(7):511-24]; [Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." *DNA Repair (Amst).* 2012; 1;11(10):781-8] 참조; 각각의 전문이 본원에 참조로 포함된다). 본원에 제공된 재조합효소가 본 발명의 실시양태에 사용될 수 있는 재조합효소의 독점적 예를 의미하지는 않는다. 본 발명의 방법 및 조성물은 새로운 직교 재조합효소에 대한 데이터베이스를 발굴하거나 또는 규정된 DNA 특이성을 지닌 합성 재조합효소를 설계함으로써 확장될 수 있다 (예를 들어, 문헌 [Groth et al., "Phage integrases: biology and applications." *J. Mol. Biol.* 2004; 335, 667-678]; [Gordley et al., "Synthesis of programmable integrases." *Proc. Natl. Acad. Sci. U S A.* 2009; 106, 5053-5058] 참조; 각각의 전문이 본원에 참조로 포함된다). 본원에 기재된 방법 및 조성물에 유용한 재조합효소의 다른 예는 관련 기술분야의 통상의 기술자에게 공지되어 있고, 발견되거나 생성되는 어떠한 새로운 재조합 효소도 본 발명의 상이한 실시양태에 사용될 수 있는 것으로 예상된다. 일부 실시양태에서, 재조합효소의 촉매 도메인이 뉴클레아제-불활성화 RNA-프로그램 가능한 뉴클레아제 (예를 들어, dCas9, 또는 그의 단편)와 융합되므로, 재조합효소 도메인은 핵산 결합성 도메인을 포함하지 않거나, 또는 표적 핵산과 결합할 수 없다 (예를 들어, 이러한 재조합효소 도메인은 특이적 DNA 결합 활성을 지니지 않도록 조작된다). DNA 결합 활성이 결여된 재조합효소 및 이를 조작하는 방법은 공지되어 있고, 이는 문헌 [Klippel et al., "Isolation and characterisation of unusual gin mutants." *EMBO J.* 1988; 7: 3983-3989]; [Burke et al., "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation." *Mol Microbiol.* 2004; 51: 937-948]; [Olorunniji et al., "Synapsis and catalysis by activated Tn3 resolvase mutants." *Nucleic Acids Res.* 2008; 36: 7181-7191]; [Rowland et al., "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." *Mol Microbiol.* 2009; 74: 282-298]; [Akopian et al., "Chimeric recombinases with designed DNA sequence recognition." *Proc Natl Acad Sci USA.* 2003;100: 8688-8691]; [Gordley et al., "Evolution of programmable zinc finger-recombinases with activity in human cells." *J Mol Biol.* 2007; 367: 802-813]; [Gordley et al., "Synthesis of programmable integrases." *Proc Natl Acad Sci USA.* 2009;106: 5053-5058]; [Arnold et al., "Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity." *EMBO J.* 1999;18: 1407-1414]; [Gaj et al., "Structure-guided reprogramming of serine recombinase DNA sequence specificity." *Proc Natl Acad Sci USA.* 2011;108(2):498-503]; 및 [Proudfoot et al., "Zinc finger recombinases with adaptable DNA sequence specificity." *PLoS One.* 2011;6(4):e19537; 각각의 전문이 본원에 참조로 포함된다]에 기재된 것을 포함한다. 예를 들어, 위치 특이성 재조합 촉진 효소-전효소 군, 예를 들어 Tn3 및 γ δ 위치 특이성 재조합 촉진 효소와 Hin 및 Gin 전효소의 세린 재조합효소는 자기 촉매 도메인과 DNA-결합성 도메인을 수반한 모듈 구조를 갖는다 (예를 들어, 문헌 [Grindley et al., "Mechanism of site-specific recombination." *Ann Rev Biochem.* 2006; 75: 567-605] 참조; 그의 전문이 본원에 참조로 포함된다). 따라서, 이들 재조합효소의 촉매 도메인은, 예를 들어 어떠한 보조 인자 (예를 들어, DNA 결합 활성)도 요구하지 않는 '활성화' 재조합효소 돌연변이체의 단리 후, 본원에 기재된 바와 같은 뉴클레아제-불활성화 RNA-프로그램 가능한 뉴클레아제 (예를 들어, dCas9, 또는 그의 단편)와의

재조합을 잘 받아들일 수 있다 (예를 들어, 문헌 [Klippel et al., "Isolation and characterisation of unusual gin mutants." *EMBO J.* 1988; 7: 3983-3989]; [Burke et al., "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation." *Mol Microbiol.* 2004; 51: 937-948]; [Olorunniji et al., "Synapsis and catalysis by activated Tn3 resolvase mutants." *Nucleic Acids Res.* 2008; 36: 7181-7191]; [Rowland et al., "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." *Mol Microbiol.* 2009; 74: 282-298]; [Akopian et al., "Chimeric recombinases with designed DNA sequence recognition." *Proc Natl Acad Sci USA.* 2003;100: 8688-8691] 참조). 부가적으로, N-말단 촉매 도메인과 C-말단 DNA 결합성 도메인을 갖는 많은 다른 자연 세린 재조합효소가 공지되어 있고 (예를 들어, phiC31 인테그라제, TnpX 유전자 전위효소, IS607 유전자 전위효소), 그들의 촉매 도메인은 본원에 기재된 바와 같은 프로그램 가능한 부위-특이적 재조합효소를 조작하도록 공동-선택될 수 있다 (예를 들어, 문헌 [Smith et al., "Diversity in the serine recombinases." *Mol Microbiol.* 2002;44: 299-307] 참조: 그의 전문이 본원에 참조로 포함된다). 유사하게, 티로신 재조합효소 (예를 들어, Cre, λ 인테그라제)의 코어 촉매 도메인은 공지되어 있고, 본원에 기재된 바와 같은 프로그램 가능한 부위-특이적 재조합효소를 조작하도록 유사하게 공동-선택될 수 있다 (예를 들어, 문헌 [Guo et al., "Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse." *Nature.* 1997; 389:40-46]; [Hartung et al., "Cre mutants with altered DNA binding properties." *J Biol Chem* 1998; 273:22884-22891]; [Shaikh et al., "Chimeras of the Flp and Cre recombinases: Tests of the mode of cleavage by Flp and Cre." *J Mol Biol.* 2000;302:27-48]; [Rongrong et al., "Effect of deletion mutation on the recombination activity of Cre recombinase." *Acta Biochim Pol.* 2005; 52:541-544]; [Kilbride et al., "Determinants of product topology in a hybrid Cre-Tn3 resolvase site-specific recombination system." *J Mol Biol.* 2006; 355:185-195]; [Warren et al., "A chimeric cre recombinase with regulated directionality." *Proc Natl Acad Sci USA.* 2008 105:18278-18283]; [Van Duyne, "Teaching Cre to follow directions." *Proc Natl Acad Sci USA.* 2009 Jan 6;106(1):4-5]; [Numrych et al., "A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophage λ ." *Nucleic Acids Res.* 1990; 18:3953-3959]; [Tirumalai et al., "The recognition of core-type DNA sites by λ integrase." *J Mol Biol.* 1998; 279:513-527]; [Aihara et al., "A conformational switch controls the DNA cleavage activity of λ integrase." *Mol Cell.* 2003; 12:187-198]; [Biswas et al., "A structural basis for allosteric control of DNA recombination by λ integrase." *Nature.* 2005; 435:1059-1066]; 및 [Warren et al., "Mutations in the amino-terminal domain of λ -integrase have differential effects on integrative and excisive recombination." *Mol Microbiol.* 2005; 55:1104-1112] 참조: 각각의 전문이 본원에 참조로 포함된다).

[0052] 핵산 변형 (예를 들어, 게놈 변형)의 맥락에서 용어 "재조합하다" 또는 "재조합"은 2개 이상의 핵산 분자, 또는 단일 핵산 분자의 2개 이상의 영역이 재조합효소 단백질 (예를 들어, 본원에 제공된 본 발명의 재조합효소 융합 단백질)의 작용에 의해 변형되는 과정을 지칭하기 위해 사용된다. 재조합으로 인해, 특히 예를 들어, 하나 이상의 핵산 분자 내에 또는 그 사이에 핵산 분자가 삽입, 역위, 절제 또는 전위될 수 있다.

[0053] 본원에 사용된 바와 같은 용어 "대상체"는 개개의 유기체, 예를 들어 개개의 포유류를 지칭한다. 일부 실시양태에서, 대상체는 인간이다. 일부 실시양태에서, 대상체는 비-인간 포유류이다. 일부 실시양태에서, 대상체는 비-인간 영장류이다. 일부 실시양태에서, 대상체는 설치류이다. 일부 실시양태에서, 대상체는 양, 염소, 소, 고양이, 또는 개이다. 일부 실시양태에서, 대상체는 척추동물, 양서류, 파충류, 어류, 곤충, 파리, 또는 선충이다. 일부 실시양태에서, 대상체는 연구용 동물이다. 일부 실시양태에서, 대상체는 유전적으로 조작된, 예를 들어 유전적으로 조작된 비-인간 대상체이다. 대상체는 어느 한 가지 성별일 수 있고 발생의 어떠한 단계일 수도 있다.

[0054] 뉴클레아제의 맥락에서 본원에 사용된 바와 같은 용어 "표적 핵산" 및 "표적 게놈"은 소정의 뉴클레아제의 하나 이상의 표적 부위를 포함하는 핵산 분자 또는 게놈을 각각 지칭한다. (뉴클레아제-불활성화) RNA-프로그램 가능한 뉴클레아제와 재조합효소 도메인을 포함하는 융합물의 맥락에서, "표적 핵산" 및 "표적 게놈"은 하나 이상의 표적 부위를 포함하는 하나 이상의 핵산 분자(들), 또는 게놈을 각각 지칭한다. 일부 실시양태에서, 표적 핵산(들)은 2개 이상, 3개 이상, 또는 4개 이상의 표적 부위를 포함한다. 일부 실시양태에서, 표적 핵산(들)은 4개의 표적 부위를 포함한다.

[0055] 용어 "표적 부위"는 (1) 뉴클레아제 (예를 들어, 본원에 제공된 Cas9 융합 단백질)에 의해 결합되고

절단되거나, 또는 (2) 재조합효소 (예를 들어, 본원에 제공된 dCas9-재조합효소 융합 단백질)에 의해 결합되고 재조합되는 (예를 들어, 표적 부위에서 또는 그 근처에서) 핵산 분자 내의 서열을 지칭한다. 표적 부위는 단일 가닥 또는 이중 가닥일 수 있다. RNA-가이드된 (예를 들어, RNA-프로그램 가능한) 뉴클레아제 (예를 들어, Cas9 gRNA 결합성 도메인과 활성 Cas9 DNA 절단 도메인 또는 다른 뉴클레아제 도메인, 예컨대 FokI를 포함하는 단백질 이량체)의 맥락에서, 표적 부위는 전형적으로, 상기 RNA-프로그램 가능한 뉴클레아제의 gRNA(들)에 대해 상보적인 뉴클레오티드 서열, 및 이러한 gRNA-상보성 서열에 인접한 3' 말단에서의 프로토스페이서 인접 모티프 (PAM)를 포함한다. 일부 실시양태에서, 예컨대 fCas9와 관련해서는, 표적 부위가 fCas9 단량체와 결합되는 특별한 서열, 및/또는 이량체화 FokI 도메인에 의해 절단되는 상기 결합된 단량체들 사이의 개재 서열을 포괄할 수 있다 (예를 들어, 본 실시예; 및 도 1A, 6d 참조). RNA-가이드된 (예를 들어, RNA-프로그램 가능한, 뉴클레아제-불활성화) 뉴클레아제와 재조합효소 (예를 들어, 재조합효소의 촉매 도메인) 간의 융합물의 맥락에서, 표적 부위는 전형적으로, 상기 RNA-프로그램 가능한 뉴클레아제 도메인의 gRNA에 대해 상보적인 뉴클레오티드 서열, 및 이러한 gRNA-상보성 서열에 인접한 3' 말단에서의 프로토스페이서 인접 모티프 (PAM)를 포함한다. 예를 들어, 일부 실시양태에서, 4개의 재조합효소 단량체가 표적 핵산(들)과 재조합되도록 조정되는데, 각 단량체는 gRNA에 의해 가이드된 (뉴클레아제-불활성화) Cas9 단백질과 융합된다. 이러한 예에서, 각 Cas9 도메인은 별개의 gRNA에 의해 가이드되어 표적 핵산(들)과 결합되므로, 이러한 표적 핵산은 4개의 표적 부위를 포함하는데, 각 부위는 별도의 dCas9-재조합효소 융합물에 의해 표적화된다 (이로써 표적 핵산(들)과 재조합되는 4개의 재조합효소 단량체와 조정된다). RNA-가이드된 뉴클레아제 Cas9 (또는 그의 gRNA-결합성 도메인) 및 Cas9의 본 발명의 융합물의 경우에는, 표적 부위가 일부 실시양태에서, 17개 내지 20개 염기 쌍 플러스 3개 염기 쌍 PAM [예를 들어, NNN (여기서, N은 독립적으로, 임의의 뉴클레오티드를 나타낸다)]일 수 있다. 전형적으로, PAM의 제1 뉴클레오티드는 어떠한 뉴클레오티드일 수도 있지만, 2개의 하류 뉴클레오티드는 특이적 RNA-가이드된 뉴클레아제에 따라서 지정된다. RNA-가이드된 뉴클레아제, 예컨대 Cas9에 대해 예시되는 표적 부위 (예컨대, PAM을 포함)는 관련 기술분야의 통상의 기술자에게 공지되어 있고, NNG, NGN, NAG, 및 NGG (여기서, N은 독립적으로, 임의의 뉴클레오티드를 나타낸다)를 포함하지만, 이에 제한되지 않는다. 또한, 상이한 종 (예를 들어, 에스. 피 오게네스 대신 에스. 더모필루스)으로부터의 Cas9 뉴클레아제는 서열 NGGNG를 포함하는 PAM을 인식한다. NNAGAAW 및 NAAR을 포함하지만, 이에 제한되지 않는 부가의 PAM 서열이 공지되어 있다 (예를 들어, 문헌 [Esvelt and Wang, Molecular Systems Biology, 9:641 (2013)] 참조; 그의 전문이 본원에 참조로 포함된다). 일부 측면에서, RNA-가이드된 뉴클레아제, 예컨대, 예를 들어 Cas9의 표적 부위 (예컨대, PAM 포함)는 구조 $[N_z]-[PAM]$ (여기서, 각각의 N은 독립적으로, 임의의 뉴클레오티드이고, z는 1 내지 50의 정수이다)를 포함할 수 있다. 일부 실시양태에서, z는 2 이상, 3 이상, 4 이상, 5 이상, 6 이상, 7 이상, 8 이상, 9 이상, 10 이상, 11 이상, 12 이상, 13 이상, 14 이상, 15 이상, 16 이상, 17 이상, 18 이상, 19 이상, 20 이상, 25 이상, 30 이상, 35 이상, 40 이상, 45 이상, 또는 50 이상이다. 일부 실시양태에서, z는 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 또는 50이다. 일부 실시양태에서, z는 20이다. 일부 실시양태에서, "표적 부위"는 또한, 뉴클레아제와 결합되긴 하지만, 이에 의해 절단되지 않는 핵산 분자 내의 서열을 지칭할 수 있다. 예를 들어, 본원에 기재된 특정 실시양태는 불활성 (또는 불활성화) Cas9 DNA 절단 도메인을 포함하는 단백질을 제공한다. 이러한 단백질 (예를 들어, 또한 Cas9 RNA 결합성 도메인을 포함하는 경우)은 gRNA에 의해 지정된 표적 부위와 결합할 수 있지만; DNA 절단 부위가 불활성화되기 때문에, 표적 부위는 특별한 단백질에 의해 절단되지 않는다. 그러나, 본원에 기재된 바와 같은 상기 단백질은 전형적으로, 핵산 분자의 절단을 매개하는 또 다른 단백질 (예를 들어, 뉴클레아제) 또는 분자와 접합, 융합 또는 결합된다. 다른 실시양태에서, 상기 단백질은 표적 핵산의 재조합을 매개하는 재조합효소 (또는 재조합효소의 촉매 도메인)와 접합, 융합 또는 결합된다. 일부 실시양태에서, 실제로 절단되거나 재조합되는 서열은 핵산 분자의 절단 또는 재조합을 매개하는 단백질 (예를 들어, 뉴클레아제 또는 재조합효소) 또는 분자에 좌우될 것이고, 일부 경우에는 예를 들어, 불활성화 Cas9 단백질(들)이 결합되는 것으로부터의 거리 또는 근접성에 관한 것이다.

[0056] 이량체화 (또는 다량체화)되는 본 발명의 단백질, 예를 들어 뉴클레아제-불활성화 Cas9 (또는 Cas9 RNA 결합성 도메인)와 DNA 절단 도메인 (예를 들어, FokI 절단 도메인 또는 활성 Cas9 절단 도메인)을 포함하는 단백질의 이량체, 또는 뉴클레아제-불활성화 Cas9 (또는 Cas9 gRNA 결합성 도메인)와 재조합효소 (또는 재조합효소의 촉매 도메인) 간의 융합물을 맥락에서, 표적 부위는 전형적으로, 좌측-절반 부위 (하나의 단백질에 의해 결합됨), 우측-절반 부위 (제2의 단백질에 의해 결합됨), 및 커팅 또는 재조합이 이루어지는 상기 절반 부위들 사이의 스페이스 서열을 포함한다. 일부 실시양태에서, 좌측-절반 부위 또는 우측-절반 부위 중 어느 하나 (스페이스 서열은 아니다)가 커팅되거나 재조합된다. 다른 실시양태에서, 스페이스 서열이 커팅되거나 재조합된다. 이러한

구조 ([좌측-절반 부위]-[스페이스 서열]-[우측-절반 부위])는 본원에서 LSR 구조로서 지칭된다. 일부 실시양태에서, 좌측-절반 부위 및/또는 우측-절반 부위는 RNA-가이드된 표적 부위 (예를 들어, Cas9 표적 부위)에 상응한다. 일부 실시양태에서, 상기 절반 부위 중 어느 하나 또는 둘 다는, 예를 들어 Cas9에 의해 표적화된 전형적인 영역 보다 더 짧거나 더 긴데, 예를 들어 20개 뉴클레오티드 보다 더 짧거나 더 길다. 일부 실시양태에서, 좌측 절반 부위와 우측 절반 부위는 상이한 핵산 서열을 포함한다. 본 발명의 뉴클레아제와 관련된 일부 실시양태에서, 표적 부위는 3개의 RNA-가이드된 뉴클레아제 표적 부위 서열, 예를 들어 Cas9 표적 부위 서열에 상응하는 3개의 서열을 포함하는 서열인데 (예를 들어, 도 2C 참조), 여기서 제1 Cas9 표적 부위 서열과 제2 Cas9 표적 부위 서열, 및 제2 Cas9 표적 부위 서열과 제3 Cas9 표적 부위 서열이 스페이스 서열에 의해 분리된다. 일부 실시양태에서, 스페이스 서열은 5 bp 이상, 6 bp 이상, 7 bp 이상, 8 bp 이상, 9 bp 이상, 10 bp 이상, 11 bp 이상, 12 bp 이상, 13 bp 이상, 14 bp 이상, 15 bp 이상, 16 bp 이상, 17 bp 이상, 18 bp 이상, 19 bp 이상, 20 bp 이상, 25 bp 이상, 30 bp 이상, 35 bp 이상, 40 bp 이상, 45 bp 이상, 50 bp 이상, 60 bp 이상, 70 bp 이상, 80 bp 이상, 90 bp 이상, 100 bp 이상, 125 bp 이상, 150 bp 이상, 175 bp 이상, 200 bp 이상, 또는 250 bp 이상 길이이다. 일부 실시양태에서, 스페이스 서열은 대략 15 bp 내지 대략 25 bp 길이이다. 일부 실시양태에서, 스페이스 서열은 대략 15 bp 길이이다. 일부 실시양태에서, 스페이스 서열은 대략 25 bp 길이이다.

[0057] 본원에 사용된 바와 같은 용어 "전사 활성인자-유사 이펙터" (TALE)는 고도로 가변적인 2개-아미노산 모티프 (반복 가변 2잔기, RVD)를 포함하는, 고도로 보존된 33 내지 34개 아미노산 서열을 함유하는 DNA 결합성 도메인을 포함하는 박테리아성 단백질을 지칭한다. 상기 RVD 모티프는 핵산 서열에 대한 결합 특이성을 결정하고, 목적하는 DNA 서열과 특이적으로 결합하는 것으로 관련 기술분야의 통상의 기술자에게 공지된 방법에 따라서 조작할 수 있다 (예를 들어, 문헌 [Miller, Jeffrey; et.al. (February 2011). "A TALE nuclease architecture for efficient genome editing" *Nature Biotechnology* 29 (2): 143-8]; [Zhang, Feng; et.al. (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". *Nature Biotechnology* 29 (2): 149-53]; [Geißler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". *PLoS ONE* 6 (5): e19509]; [Boch, Jens (February 2011). "TALEs of genome targeting". *Nature Biotechnology* 29 (2): 135-6]; [Boch, Jens; et.al. (December 2009). "Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors". *Science* 326 (5959): 1509-12]; 및 [Moscou, Matthew J.; Adam J. Bogdanove (December 2009). "A Simple Cipher Governs DNA Recognition by TAL Effectors". *Science* 326 (5959): 1501] 참조; 각각의 전문이 본원에 참조로 포함된다). 아미노산 서열과 DNA 인식 간의 단순 관계는 적당한 RVD를 함유하는 반복 절편의 조합을 선택함으로써 특이적 DNA 결합성 도메인을 조작할 수 있게 해주었다.

[0058] 본원에 사용된 바와 같은 용어 "전사 활성인자-유사 요소 뉴클레아제" (TALEN)는 DNA 절단 도메인, 예를 들어 FokI 도메인에 대한 전사 활성인자-유사 이펙터 DNA 결합성 도메인을 포함하는 인공 뉴클레아제를 지칭한다. 조작된 TALE 구조물을 생성시키기 위한 수많은 모듈 어셈블리 계획이 보고되었다 (예를 들어, 문헌 [Zhang, Feng; et.al. (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". *Nature Biotechnology* 29 (2): 149-53]; [Geißler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". *PLoS ONE* 6 (5): e19509]; [Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V. et al. (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting". *Nucleic Acids Research*]; [Morbiter, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). "Assembly of custom TALE-type DNA binding domains by modular cloning". *Nucleic Acids Research*]; [Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes". *Nucleic Acids Research*]; [Weber, E.; Gruetznier, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). Bendahmane, Mohammed. ed. "Assembly of Designer TAL Effectors by Golden Gate Cloning". *PLoS ONE* 6 (5): e19722] 참조; 각각의 전문이 본원에 참조로 포함된다).

[0059] 용어 "치료", "치료하다" 및 "치료하는"은 본원에 기재된 바와 같은 질환 또는 장애, 또는 그의 한 가지 이상의 증상을 역전시키거나, 경감시키거나, 이러한 증상의 발병을 지연시키거나, 또는 증상의 진행을 억제하는 것을 목표로 하는 임상적 개입을 지칭한다. 본원에 사용된 바와 같은 용어 "치료", "치료하다" 및 "치료하는"은 본

원에 기재된 바와 같은 질환 또는 장애, 또는 그의 한 가지 이상의 증상을 역전시키거나, 경감시키거나, 이러한 증상의 발병을 지연시키거나, 또는 증상의 진행을 억제하는 것을 목표로 하는 임상적 개입을 지칭한다. 일부 실시양태에서, 치료는 한 가지 이상의 증상이 발생된 후 및/또는 특정 질환이 진단된 후에 투여될 수 있다. 다른 실시양태에서, 치료는 증상의 부재 하에, 예를 들어 증상의 발병을 방지 또는 지연시키거나, 또는 질환의 발병 또는 진행을 억제하기 위해 투여될 수 있다. 예를 들어, 치료는 증상의 발병 이전에 감수성 개체에게 투여될 수 있다 (예를 들어, 증상 병력을 고려하고/하거나 유전적 또는 기타 감수성 요인을 고려한다). 치료는 또한, 증상이 해소된 후에도, 예를 들어 그의 재발을 방지하거나 지연시키기 위해 지속될 수 있다.

[0060]

용어 "벡터"는 본 발명의 하나 이상의 재조합 폴리뉴클레오타이드, 예를 들어 본원에 제공된 Cas9 단백질 (또는 그의 융합물) 및/또는 gRNA를 코딩하는 것을 포함하는 폴리뉴클레오타이드를 재칭한다. 벡터는 플라스미드, 바이러스성 벡터, 코스미드, 인공 염색체 및 파지미드를 포함하지만, 이에 제한되지 않는다. 벡터는 숙주 세포에서 복제할 수 있고, 벡터가 커팅될 수 있고 목적하는 핵산 서열이 삽입될 수 있는 하나 이상의 엔도뉴클레아제 제한 부위를 추가의 특징으로 한다. 벡터는 이러한 벡터를 이용하여 형질전환시켰거나 시키지 않았거나, 또는 이러한 벡터를 이용하여 게놈적으로 변형시켰거나 시키지 않은 세포의 확인 및/또는 선별에 사용하는 데 적합한 하나 이상의 마커 서열을 함유할 수 있다. 마커는, 예를 들어 항생제 (예를 들어, 카나마이신, 앰피실린) 또는 다른 화합물에 대한 내성 또는 감수성을 증가 또는 감소시키는 단백질을 코딩하는 유전자; 그의 활성이 관련 기술분야에 공지된 표준 검정에 의해 탐지 가능한 효소 (예를 들어, β -갈락토시다제, 알칼리성 포스파타제, 또는 루시페라제)를 코딩하는 유전자; 및 형질전환되거나 형질감염된 세포, 숙주, 집락 또는 플라크의 표현형에 가지적으로 영향을 미치는 유전자를 포함한다. 숙주 세포 [예를 들어, 이. 콜라이 (*E. coli*), 포유류 세포, 예컨대 CHO 세포, 곤충 세포 등]의 형질전환에 적합한 어떠한 벡터도 본 발명에 포괄되는데, 예를 들어 pUC 시리즈, pGEM 시리즈, pET 시리즈, pBAD 시리즈, pTET 시리즈, 또는 pGEX 시리즈에 속하는 벡터가 있다. 일부 실시양태에서, 벡터는 재조합 단백질 생성을 위하여 숙주 세포를 형질전환시키는 데 적합하다. 단백질 (예를 들어, 본원에 제공된 것)을 발현하기 위하여 벡터 및 숙주 세포를 선별 및 조작하는 방법, 세포를 형질전환하는 방법, 및/또는 재조합 단백질을 발현/정제하는 방법은 관련 기술분야에 널리 공지되어 있고, 예를 들어 문헌 [Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))]에 제공되어 있다.

[0061]

본원에 사용된 바와 같은 용어 "아연 핑거"는 특정 폴드와, 이러한 폴드를 안정화시키는 하나 이상의 아연 이온의 배위를 특징으로 하는 소형 핵산-결합성 단백질 구조 모티프를 지칭한다. 아연 핑거는 광범위한 각종 상이한 단백질 구조를 포괄한다 (예를 들어, 문헌 [Klug A, Rhodes D (1987). "Zinc fingers: a novel protein fold for nucleic acid recognition". *Cold Spring Harb. Symp. Quant. Biol.* 52: 473-82] 참조; 그의 전문이 본원에 참조로 포함된다). 아연 핑거는 뉴클레오타이드의 특이적 서열과 결합하도록 설계될 수 있고, 일련의 아연 핑거의 융합물을 포함하는 아연 핑거 어레이는 사실상 목적하는 어떠한 표적 서열과도 결합하도록 설계될 수 있다. 이러한 아연 핑거 어레이는, 예를 들어 핵산 절단 도메인과 접합된 경우에, 단백질, 예를 들어 뉴클레아제의 결합성 도메인을 형성할 수 있다. CysHis₂, 개그 너클(Gag knuckle), 트레블 클레프(Treble clef), 아연 리본, Zn₂/Cys₆, 및 TAZ2 도메인-유사 모티프를 포함하지만, 이에 제한되지 않는 상이한 유형의 아연 핑거 모티프가 관련 기술분야의 통상의 기술자에게 공지되어 있다 (예를 들어, 문헌 [Krishna SS, Majumdar I, Grishin NV (January 2003). "Structural classification of zinc fingers: survey and summary". *Nucleic Acids Res.* 31 (2): 532-50] 참조). 전형적으로, 단일 아연 핑거 모티프는 핵산 분자의 3 또는 4개 뉴클레오타이드와 결합된다. 따라서, 예를 들어 2개의 아연 핑거 모티프를 포함하는 아연 핑거 도메인은 6 내지 8개 뉴클레오타이드와 결합될 수 있고, 3개의 아연 핑거 모티프를 포함하는 아연 핑거 도메인은 9 내지 12개 뉴클레오타이드와 결합될 수 있으며, 4개의 아연 핑거 모티프를 포함하는 아연 핑거 도메인은 12 내지 16개 뉴클레오타이드와 결합될 수 있다. 적합한 모든 단백질 공학 기술을 이용하여, 3 내지 30개 뉴클레오타이드 길이로부터 사실상 목적하는 어떠한 표적 서열과도 결합하도록 신규 아연 핑거 융합물을 설계할 수 있고/있거나 아연 핑거의 DNA-결합 특이성을 변경시킬 수 있다 (예를 들어, 문헌 [Pabo CO, Peisach E, Grant RA (2001). "Design and selection of novel cys2His2 Zinc finger proteins". *Annual Review of Biochemistry* 70: 313-340]; [Jamieson AC, Miller JC, Pabo CO (2003). "Drug discovery with engineered zinc-finger proteins". *Nature Reviews Drug Discovery* 2 (5): 361-368]; 및 [Liu Q, Segal DJ, Ghiara JB, Barbas CF (May 1997). "Design of polydactyl zinc-finger proteins for unique addressing within complex genomes". *Proc. Natl. Acad. Sci. U.S.A.* 94 (11)] 참조; 각각의 전문이 본원에 참조로 포함된다). 핵산을 절단하는 단백질 도메인과 조작된 아연 핑거 어레이 간의 융합을 이용하여, "아연 핑거 뉴클레아제"를 생성시킬 수 있다. 아연 핑거 뉴클레아제는 전형적으로, 핵산 분자

내의 특이적 표적 부위와 결합하는 아연 핑거 도메인; 및 결합성 도메인에 의해 결합된 표적 부위 내에 또는 이러한 부위에 근접하여 핵산 분자를 커팅하는 핵산 절단 도메인을 포함한다. 전형적으로 조작된 아연 핑거 뉴클레아제는 3 내지 6개 개개의 아연 핑거 모티프와 9개 염기 쌍 내지 18개 염기 쌍 길이의 범위의 결합성 표적 부위를 갖는 결합성 도메인을 포함한다. 소정의 게놈 내에서 독특한 표적 부위와 결합하고 이를 절단시키는 것이 요망되는 상황 하에서는 보다 긴 표적 부위가 특히 매력적이다.

[0062] 본원에 사용된 바와 같은 용어 "아연 핑거 뉴클레아제"는 아연 핑거 어레이를 포함하는 결합성 도메인과 접합된 핵산 절단 도메인을 포함하는 뉴클레아제를 지칭한다. 일부 실시양태에서, 이러한 절단 도메인은 유형 II 제한 엔도뉴클레아제 FokI의 절단 도메인이다. 아연 핑거 뉴클레아제는 절단을 위해 소정의 핵산 분자 내의 사실상 목적하는 어떠한 서열도 표적으로 하도록 설계될 수 있고, 복합체 게놈의 맥락에서 독특한 부위와 결합하도록 아연 핑거 결합성 도메인을 설계할 가능성은, 예를 들어 치료적 가치의 표적화된 게놈 변형을 달성하기 위하여, 살아있는 세포 내의 단일 게놈 부위의 표적화된 절단을 허용해 준다. 이중 가닥 브레이크를 목적하는 게놈 유전자 자리로 표적화하는 것을 이용하여, 비-상동 DNA 복구 경로의 변이성 성질에 기인하여, 프레임 시프트 돌연변이를 유전자의 코딩 서열 내로 도입할 수 있다. 관련 기술분야의 통상의 기술자에게 널리 공지된 방법에 의해 관심 부위를 표적으로 하는 아연 핑거 뉴클레아제를 생성시킬 수 있다. 예를 들어, 공지된 특이성의 개개의 아연 핑거 모티프를 조합함으로써, 목적하는 특이성을 지닌 아연 핑거 결합성 도메인을 설계할 수 있다. DNA와 결합된 아연 핑거 단백질 Zif268의 구조는 이러한 분야에서의 많은 작업에 영향을 미쳐왔고, 64개의 가능한 염기 쌍 삼중체 각각에 대한 아연 핑거를 수득한 다음, 이들 모듈 아연 핑거를 혼합 및 매칭하여 목적하는 모든 서열 특이성을 지닌 단백질을 설계한다는 개념이 보고되었다 (Pavletich NP, Pabo CO (May 1991). "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å". *Science* 252 (5007): 809-17; 그의 전문이 본원에 참조로 포함된다). 일부 실시양태에서, 각각 3개의 염기 쌍 DNA 서열을 인식하는 별도의 아연 핑거를 조합하여, 9개 염기 쌍 내지 18개 염기 쌍 길이 범위의 표적 부위를 인식하는 3-, 4-, 5-, 또는 6-핑거 어레이를 생성시킨다. 일부 실시양태에서, 보다 긴 어레이가 고려된다. 다른 실시양태에서, 6 내지 8개 뉴클레오타이드를 인식하는 2-핑거 모듈을 조합하여, 4-, 6-, 또는 8-아연 핑거 어레이를 생성시킨다. 일부 실시양태에서, 박테리아성 또는 파지 디스플레이를 이용하여, 목적하는 핵산 서열, 예를 들어 3 내지 30 bp 길이의 목적하는 뉴클레아제 표적 부위를 인식하는 아연 핑거 도메인을 발생시킨다. 아연 핑거 뉴클레아제는 일부 실시양태에서, 링커, 예를 들어 폴리펩티드 링커를 통하여 서로 융합되거나 또는 달리 접합된, 아연 핑거 결합성 도메인과 절단 도메인을 포함한다. 링커의 길이가, 아연 핑거 도메인에 의해 결합된 핵산 서열로부터의 커팅물의 거리를 결정한다. 보다 짧은 링커가 사용된 경우에는, 절단 도메인이, 상기와 같이 결합된 핵산 서열에 보다 근접한 핵산을 커팅하는 반면, 보다 긴 링커는 상기 결합된 핵산 서열과 커팅물 간의 거리를 보다 크게 만들 것이다. 일부 실시양태에서, 아연 핑거 뉴클레아제의 절단 도메인은 결합된 핵산을 커팅하기 위해 이량체화해야 한다. 이러한 일부 실시양태에서, 이량체는 2개 단량체의 이중-이량체인데, 이러한 단량체 각각은 상이한 아연 핑거 결합성 도메인을 포함한다. 예를 들어, 일부 실시양태에서, 이량체는 FokI 절단 도메인과 접합된 아연 핑거 도메인 A를 포함하는 하나의 단량체와, FokI 절단 도메인과 접합된 아연 핑거 도메인 B를 포함하는 하나의 단량체를 포함할 수 있다. 이러한 비제한적 예에서, 아연 핑거 도메인 A는 표적 부위의 한쪽의 핵산 서열과 결합되고, 아연 핑거 도메인 B는 표적 부위의 다른 쪽의 핵산 서열과 결합되며, 이량체와 FokI 도메인은 아연 핑거 도메인 결합 부위 사이에 있는 핵산을 커팅한다.

[0063] 부위-특이적 뉴클레아제 및 부위-특이적 재조합효소는 시험관내 및 생체 내에서 표적화된 게놈 변형을 위한 강력한 도구이다. 살아있는 세포 내에서의 뉴클레아제 절단으로 인해, 예를 들어 상동 재조합을 통하여 절단되고 복구된 게놈 서열의 변형을 종종 초래하는 DNA 복구 기전이 촉발되는 것으로 보고되었다. 따라서, 특정 게놈 내의 독특한 특이적 서열을 표적화 절단시키면, 통상적인 유전자 표적화 방법으로는 조작하기가 어려운 세포, 예컨대 많은 인간 체세포 또는 배아 줄기 세포를 포함한 살아있는 세포 내에서 유전자를 표적화하고 유전자를 변형시키기 위한 새로운 수단을 이용할 수 있게 된다. 또 다른 접근 방식은 지정된 DNA 절편의 효율적이고 정확한 통합, 결실, 역위 또는 전위를 유발시키는 데 요구되는 기능성 모두를 보유하고 있는 부위-특이적 재조합 효소를 활용한다.

[0064] 질환 관련 서열, 예를 들어 HIV/AIDS 환자에게서의 CCR-5 대립 유전자, 또는 중앙 신생혈관증식에 필요한 유전자의 뉴클레아제-매개된 변형이 임상 상황에서 이용될 수 있고, 2개의 부위 특이적 뉴클레아제가 현재 임상 시험 중이다 (Perez, E.E. et al., "Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases." *Nature biotechnology*. 26, 808-816 (2008); ClinicalTrials.gov identifiers: NCT00842634, NCT01044654, NCT01252641, NCT01082926). 따라서, 거의 어떠한 유전 질환도 부위-특이적 뉴클레아제 및/또는 재조합효소를 이용하여 치료할 수 있고, 이러한 질환은 예를 들어, 트리플리트

(triplet) 확장과 연관된 질환 [예를 들어, 헌팅턴병(Huntington's disease), 근긴장성 퇴행위축, 척수소뇌 실조 등], 낭포성 섬유증 (CFTR 유전자를 표적화함으로써), 혈액학적 질환 (예를 들어, 이상 혈색소증), 암, 자가면역 질환, 및 바이러스성 감염을 포함한다. 본원에 제공된 본 발명의 조성물 및/또는 방법을 이용하여 치료될 수 있는 기타 질환은 연골 형성 부전, 색맹, 산성 말타제 결핍증, 아데노신 데아미나제 결핍증, 부신 백색질 형성 장애증, 에카르디(aicardi) 증후군, 알파-1 항트립신 결핍증, 알파-지중해 빈혈, 안드로젠 무감성 증후군, 아페르트(apert) 증후군, 부정맥 형성 우심실 형성이상, 모세혈관확장성 운동실조, 바르트(barth) 증후군, 베타-지중해 빈혈, 청색 고무 수포 모반 증후군(blue rubber bleb nevus syndrome), 카나반병(canavan disease), 만성 육아종성 질환 (CGD), 고양이울음 증후군(cri du chat syndrome), 더컴병(dercum's disease), 외배엽 형성이상, 판코니(fanconi) 빈혈, 진행성 골화성 섬유형성이상, 취약 X 증후군, 갈락토스혈증, 고셔병(Gaucher's disease), 전신성 강글리오시드증 (예를 들어, GM1), 혈색소 침착증, 베타-글로빈의 6번째 코돈에서의 헤모글로빈 C 돌연변이 (HbC), 혈우병, 헐러(Hurler) 증후군, 저포스파타제증, 클라인펠터(Klinefelter) 증후군, 크라베병(Krabbes Disease), 랑거-기디온(Langer-Giedion) 증후군, 백혈구 부착 결핍증 (LAD), 백색질 형성 장애증, QT 연장 증후군, 마르판(Marfan) 증후군, 뫼비우스(Moebius) 증후군, 점액다당류증 (MPS), 과슬개(nail patella) 증후군, 신장성 요붕증, 신경섬유종증, 니만-피크병(Neimann-Pick disease), 불완전 골형성증, 포르피린병(porphyrria), 프래더-윌리(Prader-Willi) 증후군, 유전조로증, 프로테우스(Proteus) 증후군, 망막모세포종, 레트(Rett) 증후군, 루빈스타인-테이비(Rubinstein-Taybi) 증후군, 산필리포(Sanfilippo) 증후군, 중증 복합 면역결핍증 (SCID), 슈와크만(Shwachman) 증후군, 겸상 적혈구병 (겸상 적혈구 빈혈), 스미스-마게니스(Smith-Magenis) 증후군, 스틱클러(Stickler) 증후군, 테이-새크스병(Tay-Sachs disease), 혈소판감소증 결석 반경 (TAR) 증후군, 트리처 콜린스(Treacher Collins) 증후군, 세염색체증, 결절성 경화증, 터너(Turner) 증후군, 요소 회로 장애, 폰히펠-린다우병(von Hippel-Lindau disease), 바르텐부르크(Waardenburg) 증후군, 윌리엄스(Williams) 증후군, 윌슨(Wilson) 질환, 비스코트-올드리히(Wiskott-Aldrich) 증후군, 및 X 연관 림프구증식 증후군 (XLP)을 포함하지만, 이에 제한되지 않는다.

[0065] 부위-특이적 게놈 변형의 한 가지 측면은 표적을 벗어난 뉴클레아제 또는 재조합효소 효과의 가능성인데, 예를 들어 하나 이상의 뉴클레오티드가 의도한 표적 서열과 상이한 게놈 서열의 절단 또는 재조합이다. 표적을 벗어난 절단/재조합의 바람직하지 못한 부작용은 유전자 표적화 현상 동안 불필요한 유전자 자리 내로의 삽입에서부터 임상 시나리오에서 중증의 합병증까지로 다양하다. 특정 대상체에게 투여된 엔도뉴클레아제 또는 재조합효소에 의해 중앙 억제 유전자 또는 필수 유전자 작용기를 코딩하는 서열이 표적을 벗어나서 절단되거나 재조합되면, 이러한 대상체에게 질환이 유발되거나 심지어 사망할 수 있다. 따라서, 표적을 벗어난 효과를 최소화하는 가장 큰 기회를 갖는 뉴클레아제 및 재조합효소를 설계하는 데 있어서 새로운 전략을 이용하는 것이 요망된다.

[0066] 본 개시내용의 방법 및 조성물은 일부 측면에서, 자신의 의도된 표적에 대하여 개선된 특이성을 지니도록 조작된 뉴클레아제 (및 그의 사용 방법) 및 재조합효소 (및 그의 사용 방법)를 제공하는 기존의 방법 및 조성물에 비해 개선을 나타낸다. 예를 들어, 관련 기술분야에 공지된 뉴클레아제 및 재조합효소 (자연 발생적인 것과 조작된 것 둘 다)는 전형적으로, 특별한 서열을 인식하는 표적 (예를 들어, DNA) 결합성 도메인을 갖는다. 부가적으로, 공지된 뉴클레아제 및 재조합효소는 절단 또는 재조합을 유도할 수 있는 단일 단백질 내의 DNA 결합성 도메인 및 촉매 도메인을 포함할 수 있으므로, 표적을 벗어난 효과가 나타날 기회가 증가되는데, 이는 뉴클레아제 또는 재조합효소 각각의 표적을 벗어난 결합시 절단 또는 재조합이 발생하는 것으로 예상되기 때문이다. 본 발명의 측면은 서열 (예를 들어, 목적하는 표적에서 2개 이상의 부위에 뉴클레아제 결합을 갖는다)의 수를 증가시키고/시키거나 2개 이상의 단백질 간의 뉴클레아제의 활성 (예를 들어, 표적 결합성 및 표적 절단성)을 분할시키는 것이 뉴클레아제의 특이성을 증가시킴으로써, 표적을 벗어난 효과가 발생할 가능성이 감소된다는 인식에 관한 것이다. 본 발명의 다른 측면은 재조합효소 (또는 불활성 DNA 결합성 도메인을 갖는 재조합효소)의 촉매 도메인과 뉴클레아제-불활성화 RNA-프로그램 가능한 뉴클레아제 간의 융합으로 인해, 어떠한 위치에서도 DNA의 표적화된 재조합이 허용된다는 인식에 관한 것이다.

[0067] 부위-특이적 뉴클레아제의 맥락에서, 본원에 제공된 전략, 방법, 조성물 및 시스템을 활용하여, 모든 부위-특이적 뉴클레아제, 예를 들어 Cas9 엔도뉴클레아제, 아연 핑거 뉴클레아제 (ZFN) 및 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN)의 변이체의 특이성을 개선시킬 수 있다. 본원에 기재된 바와 같이 변형에 적합한 뉴클레아제는 본 개시내용을 근거로 하여 관련 기술분야의 통상의 기술자에게 명백할 것이다.

[0068] 특정 실시양태에서, 본원에 제공된 전략, 방법, 조성물 및 시스템을 활용하여 RNA-가이드된 (예를 들어, RNA-프로그램 가능한) 엔도뉴클레아제 Cas9의 특이성을 개선시킨다. 전형적인 엔도뉴클레아제는 단일 표적 서열을 인식하고 절단하는 반면, Cas9 엔도뉴클레아제는 RNA:DNA 혼성화를 이용하여 표적 DNA 절단 부위를 결정하므로,

단일 단량체성 단백질은 원칙상, 가이드 RNA (gRNA)에 의해 지정된 모든 서열을 절단할 수 있게 된다. Cas9:가이드 RNA 복합체가 세포 (문헌 [Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 339, 819-823 (2013)]; [Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science*. 339, 823-826 (2013)]; [Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013)])와 유기체 (문헌 [Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*. 31, 227-229 (2013)]) 둘 다를 변형시키기 위해 성공적으로 사용되어 왔지만, 제브라피시(zebrafish) 배아를 변형시키기 위해 Cas9:가이드 RNA 복합체를 이용하는 연구에서는 ZFN 및 TALEN과 유사한 비율로 독성 (예를 들어, 표적을 벗어난 효과)이 관찰되었다 (Hwang, W.Y. et al. *Nature Biotechnology*. 31, 227-229 (2013)). 추가로, 하나의 DNA 가닥만을 절단하는 Cas9의 최근에 조작된 변이체 ("니카제")는 이중 가닥 브레이크를 2개의 별개의 gRNA 서열에 의해 지정될 수 있게 하지만 (문헌 [Cho, S. W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*. 24, 132-141 (2013)]; [Ran, F. A. et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 154, 1380-1389 (2013)]; [Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833-838 (2013)]), 이들 변이체는 DNA와 개별적으로 결합된 경우에 활성을 유지할 수 있는 각 단량체성 니카제의 능력으로부터 비롯되는 (문헌 [Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339, 819-823 (2013)]; [Jinek, M. et al. *Science* 337, 816-821 (2012)]; [Gasiunas, G., et al., Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* 109, E2579-E2586 (2012)]) 표적을 벗어난 절단 활성으로 인해 여전히 고통받고 있다 ([Ran, F. A. et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 154, 1380-1389 (2013)]; [Fu, Y., et al., Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* (2014)]). 따라서, 본 개시내용의 측면은 신규 조작된 Cas9 변이체를 이용하여 Cas9 표적을 벗어난 효과가 나타날 기회를 저하시키는 것을 목적으로 한다. 한 예에서, 예를 들어 야생형 Cas9 보다 >10배, >50배, >100배, >140배, > 200배 이상 더 높은 특이성을 나타내는, Cas9 니카제 또는 야생형 Cas9와 비교해서 개선된 특이성을 지닌 Cas9 변이체 (예를 들어, fCas9)가 제공된다 (예를 들어, 본 실시예 참조).

[0069] 본 개시내용의 다른 측면은 본 발명의 RNA-가이드된 (예를 들어, RNA-프로그램 가능한) Cas9-재조합효소 융합 단백질을 활용하는 전략, 방법, 조성물 및 시스템을 제공한다. 전형적인 재조합효소는 별개의 표적 서열을 인식하고 재조합하는 반면, 본원에 제공된 Cas9-재조합효소 융합물은 RNA:DNA 혼성화를 이용하여 표적 DNA 재조합 부위를 결정하므로, 이러한 융합 단백질은 원칙상, gRNA(들)에 의해 지정된 모든 영역을 재조합할 수 있게 된다.

[0070] DNA 및 DNA-절단성 뉴클레아제 및/또는 재조합효소에 대한 특별한 관련성이 있긴 하지만, 본원에 제공된 본 발명의 개념, 방법, 전략 및 시스템은 이러한 점에서 제한되지 않고, 모든 핵산:뉴클레아제 또는 핵산:재조합효소 시스템에 적용될 수 있다.

[0071] 뉴클레아제

[0072] 본 개시내용의 일부 측면은 본원에 기재된 방법 및 전략을 이용하여 설계되는, 증강된 특이성을 지닌 부위-특이적 뉴클레아제를 제공한다. 본 개시내용의 일부 실시양태는 이러한 뉴클레아제를 코딩하는 핵산을 제공한다. 본 개시내용의 일부 실시양태는 상기 코딩 핵산을 포함하는 발현 구조물을 제공한다 (예를 들어, 도 20 참조). 예를 들어, 일부 실시양태에서, 특정 게놈 내의 목적하는 표적 부위를 절단하도록 조작시킨 단리된 뉴클레아제가 제공된다. 일부 실시양태에서, 이러한 단리된 뉴클레아제는 RNA-프로그램 가능한 뉴클레아제, 예컨대 Cas9 뉴클레아제의 변이체이다.

[0073] 한 실시양태에서, 다음 2개의 도메인을 포함하는 융합 단백질이 제공된다: (ii) 뉴클레아제 도메인과 융합되거나 결합된 (i) RNA-프로그램 가능한 뉴클레아제 (예를 들어, Cas9 단백질, 또는 그의 단편) 도메인. 예를 들어, 일부 측면에서, Cas9 단백질 (예를 들어, 상기 융합 단백질의 Cas9 도메인)은 RNA (gRNA) 결합 활성을 보유하고 있으므로 gRNA에 대해 상보적인 표적 부위와 결합할 수 있는 뉴클레아제-불활성화 Cas9 (예를 들어, DNA 절단 활성이 결여된 Cas9; "dCas9")를 포함한다. 일부 측면에서, 뉴클레아제-불활성화 Cas9 도메인과 융합된 뉴클레아제는 표적 핵산 (예를 들어, DNA)을 절단하기 위하여 이량체화 (예를 들어, 뉴클레아제의 두 단량체가 함께 온다)를 필요로 하는 모든 뉴클레아제이다. 일부 실시양태에서, 뉴클레아제-불활성화 Cas9와 융합된 뉴클레아제는 FokI DNA 절단 도메인의 단량체이므로, 예를 들어 fCas9로서 지칭된 Cas9 변이체가 생성된다. 이러한

FokI DNA 절단 도메인은 공지되어 있고, 일부 측면에서 이는 FokI (NCBI 수탁 번호 J04623)의 아미노산 388-583에 상응한다. 일부 실시양태에서, FokI DNA 절단 도메인은 FokI의 아미노산 300-583, 320-583, 340-583, 또는 360-583에 상응한다 (또한, 문헌 [Wah et al., "Structure of FokI has implications for DNA cleavage" *Proc. Natl. Acad. Sci. USA*. 1998; 1;95(18):10564-9]; [Li et al., "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain" *Nucleic Acids Res.* 2011; 39(1):359-72]; [Kim et al., "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain" *Proc. Natl Acad. Sci. USA*. 1996; 93:1156-1160] 참조; 각각의 전문이 본원에 참조로 포함된다). 일부 실시양태에서, FokI DNA 절단 도메인은 부분으로든 전체로든, 서열 6으로서 제시된 아미노산 서열에 상응하거나 이를 포함한다. 일부 실시양태에서, FokI DNA 절단 도메인은 본원에 기재된 바와 같은, FokI의 변이체 (예를 들어, 서열 6의 변이체)이다.

[0074]

일부 실시양태에서, 융합 단백질의 이량체, 예를 들어 fCas9의 이량체가 제공된다. 예를 들어, 일부 실시양태에서, 융합 단백질은 그 자신과 이량체를 형성하여 표적 핵산의 절단을 매개한다. 일부 실시양태에서, 융합 단백질, 또는 그의 이량체는 하나 이상의 gRNA와 연합된다. 일부 측면에서, 이량체는 2개의 융합 단백질 (각각은 gRNA 결합 활성을 지닌 Cas9 도메인을 갖는다)을 함유하기 때문에, 핵산 표적의 2개의 별개의 영역을 보완하는 2개의 별개의 gRNA 서열을 이용하여 표적 핵산을 표적화한다 (예를 들어, 도 1A, 6d 참조). 따라서, 이러한 예에서, 양 융합 단백질이 표적 핵산과 결합될 때까지 (예를 들어, gRNA:표적 핵산 염기 쌍형성에 의해 지정된 바와 같은) 표적 핵산의 절단이 일어나지 않고, 뉴클레아제 도메인은 이량체화되며 (예를 들어, FokI DNA 절단 도메인; 상기 융합 단백질의 Cas9:gRNA 도메인의 결합성에 근거한 그들의 근접성에 따른 결과로서), 예를 들어 상기 결합된 Cas9 융합 단백질 사이의 영역 ("스페이스 서열") 내의 표적 핵산을 절단한다. 이는 도 1A 및 6d에 나타난 도식에 의해 예시된다. 이러한 접근 방식은 핵산을 절단하기 위해 뉴클레아제 도메인의 이량체화를 요구하지 않는 야생형 Cas9 및 다른 Cas9 변이체, 예컨대 니카제에 비해 현저한 개선을 나타낸다 ([Ran et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* 154, 1380-1389 (2013)]; [Mali et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833-838 (2013)]). 본 실시예에 기재된 바와 같은 이들 니카제 변이체는 표적에 적응한 부위와 표적을 벗어난 부위에서 발생할 수 있는, 핵산에 대한 단일 니카제의 결합시 절단 또는 니킹(nicking)을 유도할 수 있고, 니킹은 돌연변이를 유발시키는 것으로 공지되어 있다. Cas9 니카제를 코딩하는 예시되는 뉴클레오티드 (서열 7) 및 Cas9 니카제의 예시되는 아미노산 서열 (서열 8)이 다음에 제공된다. 본원에 제공된 변이체는 표적 핵산 절단을 유도하기 위하여 2개의 Cas9 변이체가 서로 근접하여 결합될 필요가 있기 때문에, 표적을 벗어난 절단이 유도될 기회가 감소된다 (예를 들어, 본 실시예 참조). 예를 들어, 일부 실시양태에서, 뉴클레아제 도메인과 융합된 Cas9 변이체 (예를 들어, fCas9)는 야생형 Cas9 또는 다른 Cas9 변이체 (예를 들어, 니카제)의 표적에 적응한:표적을 벗어난 변형 비보다 2배 이상, 5배 이상, 10배 이상, 20배 이상, 30배 이상, 40배 이상, 50배 이상, 60배 이상, 70배 이상, 80배 이상, 90배 이상, 100배 이상, 110배 이상, 120배 이상, 130배 이상, 140배 이상, 150배 이상, 175배 이상, 200배 이상, 250배 이상 더 높은, 표적에 적응한:표적을 벗어난 변형 비를 갖는다. 일부 실시양태에서, 뉴클레아제 도메인과 융합된 Cas9 변이체 (예를 들어, fCas9)는 야생형 Cas9 또는 다른 Cas9 변이체의 표적에 적응한:표적을 벗어난 변형 비보다 약 60배 내지 180배, 약 80배 내지 160배, 약 100배 내지 150배, 또는 약 120배 내지 140배 더 높은, 표적에 적응한:표적을 벗어난 변형 비를 갖는다. 표적에 적응한:표적을 벗어난 변형 비를 결정하는 방법은 공지되어 있고, 본 실시예에 기재된 것을 포함한다. 특정 실시양태에서, 표적에 적응한:표적을 벗어난 변형 비는 특정 유전자 내의 공지된 Cas9 표적을 벗어난 부위의 변형 수 또는 양을 측정함으로써 결정된다. 예를 들어, *CLTA*, *EMX*, 및 *VEGF* 유전자의 Cas9 표적을 벗어난 부위는 공지되어 있고, 이들 부위에서의 변형을 측정하고 시험 단백질과 대조군을 비교할 수 있다. 표적 부위 및 그의 상응하는 공지된 표적을 벗어난 부위 (예를 들어, *CLTA*, *EMX*, 및 *VEGF* 표적을 벗어난 부위에 대한 표 5 참조)는 특별한 Cas9 단백질 또는 변이체로 처리된 세포 (예를 들어, HEK293)로부터 단리된 게놈 DNA로부터 증폭된다. 이어서, 이러한 변형을 고 처리량 서열 분석에 의해 분석한다. 잠재적 게놈성의 표적을 벗어난 부위에 2개 이상의 염기 쌍의 삽입 또는 결실을 함유하고 대조군 gRNA-처리된 샘플과 비교해서 표적 gRNA-처리된 샘플에서 유의적으로 더 큰 수가 존재하는 (P 값 < 0.005, 피셔의 직접 확률 시험) 서열이 Cas9 뉴클레아제-유도된 게놈 변형물로 간주된다.

[0075] Cas9 니카제 (뉴클레오타이드 서열):

ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
CCAAAGAAGAAGCGGAAGGTCCGGTATCCACGGAGTCCCAGCAGCCGATAAAAAGTATTCTATTGGTTTAGCTATC
GGCACTAATTCCGTTGGATGGGCTGTCTAACCCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGG
AACACAGACCGTCATTGATTAAAAAGAACTTTATCGGTGCCCTCCTATTTCGATAGTGGCGAAACGGCAGAGGCG
ACTCGCCTGAAACGAACCGCTCGGAGAAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTT
AGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTCCCGTTTGGAAAGAGTCCCTTCTTGTGCAAGAGGACAAG
AAACATGAACGGCACCCTATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTAT
CACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTAAATCTACTTGGCTCTTGCCCATATG
ATAAAGTTCCGTGGGCACTTTCTCATTTAGGGGTGATCTAAATCCGGACAACCTCGGATGTCGACAAACTGTTTCATC
CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATT
CTTAGCGCCCGCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAAATTACCCGGAGAGAAGAAAAATGGG
TTGTTCCGTTAACCTTATAGCGCTCTCACTAGGCCGTGACACCAAAATTTAAGTCGAACTTCGACTTAGCTGAAGAT
GCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTAT
GCGGACTTATTTTGGCTGCCAAAACCTTAGCGATGCAATCCTCCTATCTGACATAGTGAAGTAAATACTGAG
ATTACCAAGGCGCGCTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAGAACTTGACACTTCTCAAG
GCCCTAGTCCGTGAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAAGGT
TATATTGACGGCGGAGCGAGTCAAGAGGAATTTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACG
GAAGAGTTGCTTGTAAAACCTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTGACAAACGGTAGCATTTCCA
CATCAAAATCCACTTAGCGCAATTGATGCTATCTTAGAAGGCAGGAGGATTTTATCCGTTCTCTCAAAGCAAT
CGTGAAAAGATTGAGAAATCCTAACCTTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACCTCTCGG
TTGCGATGGATGACAAAGAAATCCGAAAGAACGATTACTCCATGGAATTTTGGAGAAAGTTGTCGATAAAGGTGCG
TCAGTCAATCGTTTCATCGAGAGGATGACCAACTTTGACAAAGAAATTTACCGAAGCAAAAAAGTATTGCTTAAGCAC
AGTTTACTTTACAGTATTTTACAGTGTACAATGAACCTCACGAAAGTTAAGTATGTCAGTGGGATGCGTAA
CCCGCCTTTCTAAGCGGAGAACGAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAGTGACAGTT
AAGCAATTGAAAGAGGACTACTTTAAGAAAATGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGA
TTTAAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCTGGATAACGAAAG
AATGAAGATATCTTAGAAGTATAGTGTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTA
AAAACATACGCTCAGCTTCTGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGA
TTGTCGCGGAACTTATCAACGGGATAAGAGACAAGCAAGTGGTAAAACCTATTCTCGATTCTTAAAGAGCGAC
GGCTTCGCCAATAGGAACCTTTATGACGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCA
CAGGTTTCGGGACAAGGGGACTCATTGACGCAACATATTGCGAATCTTGCTGGTTTCGCCAGCCATCAAAAAGGCG
ATATCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCTATGGGACGTCACAAACCGGAAACATTTGTAATC
GAGATGGGACGCGAAATCAACAGCATCAGAAGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAAATAGAAGAG
GGTATTAAAGAACTGGGCGACGAGTCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCGAAGCAGAGAACTT
TACCTCTATTACCTACAAAATGGAAAGGGACATGTATGTTGATCAGGAACCTGGACATAAAACCGTTTATCTGATTAC
GACGTCGATCACAATTGTACCCCAATCCTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGAT
AAGAACCGAGGGAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAGAAAATGAAGAACTATTGGCGGCAGCTC
CTAAATGCGAAACTGATAACGCAAGAAAAGTTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACCT
GACAAGGCCGATTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGAT
TCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAA
TTGGTGTGCGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATACTACCACCATGCGCACGAC
GCTTATCTTAATGCCGTGCTAGGGACCGCACTCATTAAGAAAATACCCGAAGCTAGAAAAGTGAGTTTGTGTATGGT
GATTACAAAGTTTATGACGTCGCTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATAC
TTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAATCACTCTGGCAACCGGAGAGATACGCAAAACGACCT
TTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTT
TTGTCCATGCCCAAGTCAACATAGTAAAGAAAACCTGAGGTGCAGACCGGAGGTTTTCAAAGGAATCGATTCTT
CCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGC
CTACAGTTGCTTATTGTGCTAGTAGTGCCAAAAGTTGAGAAGGGAAAAATCCAAGAACTGAAGTCAGTCAAA
GAATTATTGGGGATAACGATTATGGAGCGCTCGCTTTTGAAGAAGAACCCATCGACTTCTTTGAGGCGAAAGGT
TACAAGGAAGTAAAAAGGATCTCATAATTAACTACCAAAAGTATAGTCTGTTTGAAGTAAAGAAATGGCCGAAAA
CGGATGTTGGTACGCGCGGAGAGCTTCAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCTCTG
TATTTAGCGTCCCATACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAAGCACTTTTGTGAGCAG
CACAAACATTATCTCGACGAAATCATAGAGCAAAATTCGGAATTCAGTAAGAGAGTCATCTAGCTGATGCCAAT
CTGGACAAAGTATTAAGCGCATACAACAAGCACAGGATAAAACCTATCGTGAAGCAGGCGGAAAAATATTATCCAT
TTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGATTCAAGTATTTTGACACAACGATAGATCGCAACGATAC
ACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTACCAATCCATCACGGGATTATATGAAACTCGGATAGAT
TTGTCACAGCTTGGGGGTGAC (서열 7)

[0076]

[0077] Cas9 니카제 (D10A)(아미노산 서열):

```
MDKKYSIGLAIGTNSVGAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLNL
IAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLA AKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELVLKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDQKS
GKTILDFLKS DGFANRNFQMLHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDELVKV
MGRHKPENIVIMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLT RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLS ELDKAGFIKRQLVETRQITKHVAQILDSRMTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVTALIKKYPKLESEFVYGDYKVDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERS SF EKNPIDFLEAKGYKEVKKDLI IKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVLADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ
SITGLYETRIDLSQLGGD (서열 8)
```

[0078]

[0079] 일부 실시양태에서, Cas9 변이체 (예를 들어, fCas9)와 결합되는 gRNA는 스페이서 서열을 기준으로 하여, "A" 및 "B" 배향으로 간주된 2가지 방식 중 한 가지로 배향될 수 있다. 배향 A에서는, PAM 모티프와 결합되는 gRNA의 영역이 스페이서 서열로부터 원위이고, gRNA의 5' 말단이 스페이서 서열에 인접한 반면 (도 6c); 배향 B에서는, PAM 모티프와 결합되는 gRNA의 영역이 스페이서 서열에 인접해 있다 (도 9). 일부 실시양태에서, 이러한 gRNA는 A 또는 B 배향으로 표적 핵산과 결합하도록 (예를 들어, Cas9 변이체, 예컨대 fCas9와의 복합체의 일부로서) 조작되거나 선별된다. 일부 실시양태에서, gRNA는 A 배향으로 표적 핵산과 결합하도록 (예를 들어, Cas9 변이체, 예컨대 fCas9와의 복합체의 일부로서) 조작되거나 선별된다. 일부 실시양태에서, gRNA는 B 배향으로 표적 핵산과 결합하도록 (예를 들어, Cas9 변이체, 예컨대 fCas9와의 복합체의 일부로서) 조작되거나 선별된다.

[0080]

일부 실시양태에서, 융합 단백질의 도메인은, 예를 들어 본원에 기재된 바와 같이 링커를 통하여 연결된다. 특정 실시양태에서, 링커는 펩티드 링커이다. 다른 실시양태에서, 링커는 비-펩티드 링커이다. 일부 실시양태에서, 기능성 도메인은 펩티드 링커 (예를 들어, 융합됨) 또는 비-펩티드 링커를 통하여 본 발명의 융합 단백질과 연결된다. 일부 실시양태에서, 이러한 기능성 도메인은 핵 국재화 신호 (NLS) 도메인이다. NLS 도메인은 핵 수송에 의해 세포 핵 내로 유입하기 위하여 특정 단백질에 "태그를 부착"하거나 또는 신호를 보내는 아미노산 서열을 포함한다. 전형적으로, 이러한 신호는 단백질 표면 상에 노출된 양 전하를 띤 리신 또는 아르기닌의 하나 이상의 짧은 서열로 이루어진다. NLS 서열은 관련 기술분야에 널리 공지되어 있고 (예를 들어, 문헌 [Lange et al., "Classical nuclear localization signals: definition, function, and interaction with importin alpha." *J Biol Chem.* 2007 Feb 23;282(8):5101-5] 참조; 그의 전문이 본원에 참조로 포함된다), 예를 들어 본 실시예 섹션에 기재된 것을 포함한다. 일부 실시양태에서, NLS 서열은 부분으로든 전체로든, 아미노산 서열 MAPKKKKRVGIHRGVP (서열 318)을 포함한다. 링커를 통하여 연합된 도메인 [예를 들어, gRNA 결합성 도메인 (dCas9 도메인), 촉매 뉴클레아제 도메인, 및 NLS 도메인 중 2개 이상]는 어떠한 배향이나 순서로든 연결될 수 있다. 예를 들어, 일부 실시양태에서, 모든 도메인은 융합 단백질의 N-말단, C-말단, 또는 N-말단에서의 도메인과 C-말단에서의 도메인 사이에 존재할 수 있다. 일부 실시양태에서, 본 발명의 융합 단백질 내의 도메인의 배향 또는 순서는 도 6b에 제공된 바와 같다. 융합 단백질이 3개의 도메인 [예를 들어, gRNA 결합성 도메인 (예를 들어, dCas9 도메인), 뉴클레아제 도메인 (예를 들어, FokI), 및 NLS 도메인]을 포함하는 일부 실시양태에서, 각 도메인은 본원에 제공된 바와 같은 링커를 통하여 연결된다. 일부 실시양태에서, 상기 도메인은 링커를 통하여 연결되지 않는다. 일부 실시양태에서, 상기 도메인 중 하나 이상이 링커를 통하여 연결된다.

[0081]

일부 실시양태에서, 3개의 도메인 [예를 들어, gRNA 결합성 도메인 (예를 들어, dCas9 도메인), 뉴클레아제 도메인 (예를 들어, FokI), 및 NLS 도메인]을 포함하는 본 발명의 융합 단백질 (예를 들어, fCas9)은 다음에 제시된 바와 같이, 서열 9, 서열 10, 서열 11, 서열 12, 또는 서열 319로서 제시된 뉴클레오티드 서열 (또는 그의 단편 또는 변이체)에 의해 코딩된다.

[0082]

fCas9 (예를 들어, dCas9-NLS-GGS3클레이-FokI)

ATGGATAAAAAGTATTCTATTGGTTTGTAGCTATCGGCACTAATTCGGTTGGATGGGCTGTCATAACCGATGAATAC
AAAGTACCTTCAAAGAAATTAAGGTGTTGGGGAACACAGACCGTCATTTCGATTAAAAAGAAATCTTATCGGTGCC
CTCCTATTTCGATAGTGGCGAAACCGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC
AAGAACCAGAAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTTTCACCGT
TTGGAAGAGTCTTCTTGTGCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAACATAGTAGATGAG
GTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
CTGAGGTAAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACCTTCTCATTGAGGGTGATCTAAAT
CCGGACAACCTCGGATGTCGACAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCTCTCTAAATCCCGACGGCTAGAAAAACCTG
ATCGCACAAATTACCCGGAGAGAAGAAAAATGGGTTGTTTCGGTAACCTTATAGCGCTCTCACTAGGCTGACACCA
AATTTTTAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTC
GACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGGTGCCAAAAACCTTAGCGATGCAATC
CTCCTATCTGACATACTGAGAGTTAATCTGAGATTACCAAGGCGCGTTATCCGCTTCAATGATCAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTGAGCAACTGCCTGAGAAATATAAGGAAATA
TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTT
ATCAAAACCATATAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGA
AAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGG
CAGGAGGATTTTTTCCGTTTCTCAAAGACAATCGTGAAAAGATTGAGAAAAATCCTAACCTTTTCGCATACCTTAC
TATGTGGGACCCCTGGCCCGAGGGAACCTCTCGGTTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
TGGAATTTTGGAGAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTTCATCGAGAGGATGACCACTTTGACAAG
AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTACAGGTGACAAATGAACCTCAG
AAAGTTAAGTATGCTACTGAGGGCATGCGTAAACCCGCTTTCTAAGCGGAGAACAGAAAGCAATAGTAGAT
CTGTTTATCAAGACCAACCGCAAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAAATTGAATGCTTC
GATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACCTTGGTACGTATCATGACCTCTAAAGATA
ATTAAAGATAAGGACTTCTCGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCCTC
TTTGAAGATCGGGAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTTCGACGATAAGGTTATGAAACAG
TTAAAGAGGCGTCGCTATACGGGCTGGGACGATTGTCGCGGAACTTATCAACGGGATAAGAGACAAGCAAGT
GGTAAACCTATTCTCGATTTCCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGACGCTGATCCATGATGAC
TCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCG
AATCTTGTGTTGCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGAGTGAAGTAAAGGTC
ATGGGACGTCACAAACCGGAAAAACATTGTAATCGAGATGGCAGCGGAAAAATCAAACGACTCAGAAGGGGCAAAAA
AACACTCGAGAGCGGATGAAGAGAATAGAAGAGGTTAATAAGAACTGGGACGCCAGATCTTAAAGAGCAATCCT
GTGGAATAATACCAATTCGAGAAGCAAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
CAGGAAGTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCTTTTGAAGGACGAT
TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCAGGGGAAAAAGTGACAATGTCCAAGCGAGGAAGTC
GTAAGAAAAATGAAGAACTATTGGCGCGAGCTCCTAAATGCGAACTGATAACGCAAGAAAGTTTCGATAACTTA

[0083]

ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACCTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAACCCGC
CAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATT
CGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTTGGTGTGCGGACTTCAGAAAGGATTTTCAATTTCTATAAAGTT
AGGGAGATAAATACTACCACATGCGCAGCAGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAA
TACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCTGTAAGATGATCGCGAAAGC
GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATC
ACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGAT
AAGGGCCGGGACTTCGCGACGGTGAGAAAAAGTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAACTGAGGTG
CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGAC
TGGGACCGGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCTAGTAGTGCGCAAAAGTTGAG
AAGGGAATCCTAAGAACTGAAGTCAGTCAAAGAAATTATGGGGATAACGATTATGGAGCGCTCGTCTTTTGAA
AAGAACCCTATCGACTTCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAGGATCTCATAATTAACTACCAAAG
TATAGTCTGTTTGAAGTTAGAAAAATGGCCGAAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAGGGGAACGAA
CTCGCACTACCGTCTAAATACGTGAATTTCTGTATTTAGCGTCCATTACGAGAAGTTGAAAGGTTACCTGAA
GATAACGAACAGAAACCTTTTGTGTGAGCAGCAAAACATTATCTGACGAAATCATAGAGCAAAATTCGGAA
TTCAGTAAGAGATCATCTAGCTGATGCCAATCTGGACAAAGTATTAGCGCATACAACAAGCACAGGGATAAA
CCCATACGTGAGCAGGCGGAAATATATCCATTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCAATCAAG
TATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTACCAA
TCCATCACGGGATTATATGAACCTCGGATAGATTGTCACAGCTTGGGGGTGACGGATCCCCAAGAAGAAAGAGG
AAAGTCTCGAGCGACTACAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC
AAGGCTGACGAGTACAGTGGAGTGGCGGCAGCGGAGGTTCTGGATCCCACTAGTCAAAGTGAAGTGGAGGAG
AAGAAATCTGAATTCGTCTATAAATTGAAATATGTGCTCATGAATATATTGAATTAATTGAAATGGCAGAAAT
TCCATCAGGATAGAATCTTGAATGAAGGTAATGGAATTTTTATGAAAGTTTATGGATATAGAGGTAACAT
TTGGGTGATCAAGGAACCGGACGGAGCAATTTATAGTGTGATCTCCTATTGATTACGGTGTGATCGGAT
ACTAAAGCTTATAGCGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTGGAAGAAAT
CAAACACGAAACAAATATCAACCTAATGAATGGTGGAAAGTCTATCCATCTTGTGAACGGAAATTAAGTTT
TTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAAT
GGAGCTGTTCTTAGTGTAGAAGAGCTTTAATTTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGAGGAA
GTCAGACGGAAATTTAATAACGGCGAGATAAACTTT (서열 9)

[0084]

[0085]

fCas9 (예를 들어, NLS-dCas9-GGS3클릭-FokI):

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCC
 CCCAAGAAGAAGAGGAAGGTGGGCATTCCACCGCGGGGTACCTATGGATAAAAAAGTATTCTATTGGTTTAGCTATC
 GGCCTAATTCGGTTGGATGGGCTGTCTATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGG
 AACACAGACCGTCAATTCGATTAAAAAGAAATCTTATCGGTGCCCTCCTATTTCGATAGTGGCGAAACGGCAGAGGCG
 ACTCGCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTT
 AGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTACCGTTTGAAGAGTCTTCTTGTGCGAAGAGGACAAG
 AAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTAT
 CACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTAAATCTACTTGGCTCTTGCCCATATG
 ATAAAGTTCGCTGGGCACCTTCTCATTGAGGGTGATCTAAATCCGACAACTCGGATGTCGACAAACTGTTTCATC
 CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATT
 CTTAGCGCCCGCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGAGAGAGAAAAATGGG
 TTGTTCCGTAACCTTATAGCGCTCTCACTAGGCTGACACCAAATTTAAGTCGAACCTCGACTTAGCTGAAGAT
 GCCAATTGACGCTTAGTAAGGACAGTACGATGACGATCTCGACAATCTACTGGCACAATTTGGAGATCAGTAT
 GCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATCTGAG
 ATTACCAAGGCGCGCTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAGACTTGACACTTCTCAAG
 GCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGAGGT
 TATATTGACGGCGGAGCGAGTCAAGAGGAATTTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACG
 GAAGACTTTTACGAGTATTTACAGTGTACAATGAACTCAGCAAGTAAAGTATGTCAGTGAAGGCGATCGCTAAA
 CATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCTCAAAGACAAT
 CGTGAAAAGATTGAGAAAATCCTAACCTTTTCGCATACCTTACTATGTGGGACCCCTGGCCGAGGGAACCTCTCGG
 TTCGATGGATGACAAGAAAGTCCGAGAAACGATTACTCCATGGAATTTTGAAGAGTGTGCGATAAAGGTGCG
 TCAGCTCAATCGTTTATCGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCAC
 AGTTTACTTTACGAGTATTTACAGTGTACAATGAACTCAGCAAGTAAAGTATGTCAGTGAAGGCGATCGCTAAA
 CCCGCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTT
 AAGCAATTGAAAGAGGACTACTTTAAGAAAATGAATGCTTCGATTCTGTCGAGATCTCCGGGTAGAAGATCGA
 TTTAATGCGTCACCTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAG
 AATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTCTTTGAAGATCGGGAATGATTGAGGAAAGACTA
 AAAACATACGCTCACCTGTTTCGACGATAAAGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGA

[0086]

TTGTGCGGAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAACTATTCTCGATTTTCTAAAGAGCGAC
 GGCTTCGCCAATAGGAACCTTATGACAGTGTCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCA
 CAGGTTTCGGACAAGGGGACTCATTGACGAAACATATTGCGAATCTTGCTGGTTCCGCCAGCCATCAAAAAGGGC
 ATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCTATGGGACGTCACAAACCGGAAAAACATTGTAATC
 GAGATGGCAGCGGAAAATCAAACGACTCAGAAAGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAAGAG
 GGTATTAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCTGTGGAAAATACCAATTGCAGAACGAGAAAACCT
 TACCCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACCTGGACATAAACCGTTTATCTGATTAC
 GACGTGATGCCATTGTATCCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGAT
 AAGAACCAGGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTC
 CTAAATGCGAAACTGATAACGCAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTGTGCTGAACTT
 GACAAGGCCGATTATTATAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGAT
 TCCCGAATGAATACGAAATACGAGGAGACGATAAGCTGATTTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAA
 TTGGTGTGCGACTTCAGAAAGGATTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGAC
 GCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGT
 GATTACAAAGTTTATGACGTCCGTAAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATAC
 TTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAATCACCTGCGCAACCGGAGAGATACGCAACGACCT
 TTAATTGAACCAATGGGGGACACAGGTGAAATCGTATGGGATAAAGGGCCGGGACTTCGCGACGGTCAAAAGTT
 TTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACCTGAGGTGCAGACCGGAGGGTTTTTCAAAGGAATCGATTCTT
 CAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGC
 CCTACAGTTGCCATTCTGTCTAGTAGTGGCAAAAGTTGAGAAGGGGAAAAATCCAAGAACTGAAGTCAGTCAAA
 GAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAGAAACCCCATCGACTTCTTTGAGGCGAAAGGT
 TACAAGGAAGTAAAAAAGGATCTCATAAATTAACCTACCAAAGTATAGTCTGTTTGAAGTAAAGAAATGGCCGAAA
 CGGATGTTGGCTAGCGCCGAGAGCTTCAAAGGGGAACGAACCTCGCACTACCGTCTAAATACGTGAATTTCTCTG
 TATTTAGCGTCCCATACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAACTTTTGTGAGCAG
 CAAAAACATTATCTCGACGAAATCATAGAGCAAAATTTCCGAATTGAGTAAGAGAGTCATCTAGCTGATGCCAAT
 CTGGACAAGTATTAAGCGCATACAAACAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAAATATTATCCAT
 TTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCAATCAAGTATTTTGACACAACGATAGATCGCAAAACGATAC
 ACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTACCAATCCATCACGGGATTATATGAAACTCGGATAGAT
 TTGTACAGCTTTGGGGTGACTCAGGTGGAAGTGGCGGCGAGCGGAGGTTCTGGATCCCAACTAGTCAAAAGTGAA
 CTGGAGGAGAAGAAATCTGAACCTTCGTCTATAAATGAAATATGTGCCTCATGAATATATTGAATTAATGAAAT
 GCCAATAATTTCCACTCAGGATAGAATTTCTTGAATGAAGGTAATGAAATTTTATGAAAGTTTATGGATATAGA
 GGTAACATTTGGGTGGATCAAGGAAACCGGACGAGCAATTTACTGTGCGATCTCCTATTGATTACGGTGTG
 ATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTTGGCCAAGCAGATGAAATGCAACGATATGTC
 GAAGAAAATCAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTCTGTAACGGAA
 TTTAAGTTTTTATTGTGAGTGGTCACTTTAAAGGAACTACAAAGCTCAGCTTACACGATTAAATCATATCACT
 AATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTAATTGGTGGAGAAATGATTAAAGCCGCGACATTAAAC
 TTAGAGGAAGTCAGACGGAAATTTAATAACGCGGAGATAAACTTT (서열 10)

[0087]

[0088]

fCas9 (예를 들어, FokI-GGS3링커-dCas9-NLS):

ATGGGATCCCACTAGTCAAAAGTGAAGTGGAGGAGAAGAAATCTGAACCTTCGTCATAAAATTGAAATATGTGCCCT
CATGAATATATTGAATTAATTGAAATTGCCAGAAATCCACTCAGGATAGAAATCTTGAAATGAAGGTAATGGAA
TTTTTTATGAAAGTTTATGGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATCT
GTCGGATCTCCTATTGATTACGGTGTGATCGTGGATACATAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGC
CAAGCAGATGAAATGCAACGATATGTCGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGG
AAAGTCTATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAACTACAAAGCT
CAGCTTACACGATTAAATCATATCACTAATTTGAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTTGGTGG
GAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTTGGC
GGTAGTGGGGGATCTGGGGGAAGTATGGATAAAAAGTATCTATTGGTTTTAGCTATCGGCACATAATCCCGTTGGA
TGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCTG
ATTAAGAAAGAAATCTATCGGTGCCCTCTATTTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACC
GCTCGGAGAAGGTATACACGTCGCAAGAACCAGATATGTTACTTACAAGAAATTTTAGCAATGAGATGGCCAAA
GTTGACGATTCTTTCTTTACCCTTTGGAAGAGTCCCTTCTTGTGCAAGAGGACAAGAAACATGAACGGCACCC
ATCTTTCCGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCACGATTTATCACCTCAGAAAAAAGCTA
GTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCAC
TTTCTCATTGAGGTGATCTAAATCCGGACAACCTCGGATGTCGACAACTGTTTCATCCAGTTAGTACAAACCTAT
AATCAGTTGTTGAAGAGAACCTTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCTCTCT
AAATCCCGACCGCTAGAAAACTGATCGCACAAATACCCGGAGAGAAGAAAAATGGGTGTTTCGGTAACCTTATA
CGCTCTCACTAGGCTGACACCAATTTAAGTCAAGCTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGT
AAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAATGGAGATCAGTATCGGCACTTATTTTTGGCT

[0089]

GCCAAAAACCTTAGCGATGCAATCCTCTCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTA
TCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTGAGCAA
CTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCG
AGTCAAGAGGAATTTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTA
CTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGCG
GAATTGATGCTATATACTTAGAAGCGCAGGAGGATTTTATCCGTTCTCAAAGACAATCGTGAAAAAGATTGAGAAA
ATCCTAACCTTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACCTCTCGGTTTCGCATGGATGACAAGA
AAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATC
GAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCTTAAGCACAGTTTACTTTACGAGTAT
TTCACAGTGTACAAATGAATCAGCAAGTTAAGTATGTCTACTGAGGGCATGCGTAAACCCGCTTTCTTAAGCGGA
GAACAGAGAAGAAACATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGAC
TACTTTAAGAAAATTTGAATGCTTCGATTCTGTCGAGATCTCCGGGTGAGAAGATCGATTTAATGCGTCACTTGGT
ACGTATCATGACCTCTTAAAGATAATTAAGATAAGGACTTCTGGATAACGAAGAGAATGAAGATATCTTAGAA
GATATAGTGTGACTCTTACCTCTTTGAAGATCGGGAATGATTGAGGAAAGACTAAAAACATACGCTCACCTG
TTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTGCGGAAACTTATC
AACGGGATAAGAGACAAGCAAAGTGTTAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAAC
TTTATGACGTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGG
GACTCATTGACGAACATATTGCGAATCTTGCTGGTTTCGCCAGCCATCAAAAAGGCGATATCCAGACAGTCAAA
GTAGTGGATGAGCTAGTTAAGTTCATGGGACGTCACAAACCGGAAACATTGTAATCGAGATGGCACGCGAAAA
CAACGACTCAGAAGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGC
AGCCAGATCTTAAAGGAGCATCTGTGGAAAATACCCAATTGCAGAACGAGAACTTTACCTCTATTACCTACAA
AATGGAAGGGACATGTATGTTGATCAGGAACCTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTA
CCCCAATCCTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCAGGGGAAAAAGT
GACAATGTTCCAAGCGAGGAAGTCGTAAGAAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAACTGATA
ACGCAAGAAAGTTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGCTGAACTTGACAGGCGCGATTTATT
AAACGTCAGCTCGTGGAAACCCGCCAAATCACAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAA
TACGACGAGAACGATAAGCTGATTGCGGAAGTCAAAGTAATCACTTTAAAGTCAAATTTGGTGTGCGGACTTCAGA
AAGGATTTTCAATTTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTC
GTAGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAAGTTTGTGTATGGTGATTACAAAGTTTATGAC
GTCCGTAAGATGATCGCAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATT
ATGAATTTCTTTAAGACGGAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGG
GAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTGTCCATGCCCAAGTC
AACATAGTAAAGAAACTGAGGTGACAGCCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGTAT
AAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCTCTATTCT
GTCTAGTAGTGGCAAAAGTTGAGAAGGGGAAATCCAAGAACTGAAGTCAGTCAAAGAATTATTGGGGATAACG
ATTATGGAGCGCTCGTCTTTTGAAGAAACCCCATCGACTTCTTTGAGGCGAAAGGTTACAAGGAAGTAAAAAG
GATCTCATAAATTAACCTACCAAGTATAGTCTGTTTGTAGTTAGAAAATGGCCGAAACCGGATGTTGGCTAGCGCC
GGAGAGCTTCAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCTGTATTTAGCGTCCCATTA
GAGAAGTTGAAGGTTCAACCTGAAGATAACGACAGAAAGCAACTTTTTGTTGAGCAGCACAACCAATTATCTCGAC
GAAATCATAGAGCAAAATTCGGAATTCAGTAAGAGAGTCATCTAGCTGATGCCAATCTGGACAAAGTATTAAAGC
GCATACAACAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAATATTATCCATTTGTTTACTCTTACCAAC
CTCGGCGCTCCAGCCGATTTCAAGTATTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTG
CTAGACGCGCACTGATTACCAATCCATACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGT
GACGGATCCCCCAAGAGAAGAGGAAAGTCTCGAGCACTACAAAGACCATGACGGTGATTATAAAGATCATGAC
ATCGATTACAAGGATGACGATGACAAGGCTGCAGGA (서열 11)

[0090]

[0091]

fCas9 (예를 들어, NLS -FokI-GGS3링크-dCas9):

ATGGACTACAAGACCATGACGGTGATTATAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCC
CCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGGGTACCTGGAGGTCTATGGGATCCCAACTAGTCAAAAGT
GAACTGGAGGAGAAGAAATCTGAACCTTCGTCTATAAATTGAAATATGTGCCCTCATGAATATATTGAATTAATTGAA
ATTGCCAGAAATTCCTACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTATGAAAGTTTATGGATAT
AGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATCTGTCGGATCTCCTATTGATTACGGT
GTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATAT
GTCGAAGAAAATCAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACG
GAATTTAAGTTTTTATTGTGAGTGGTCACCTTTAAAGGAACTACAAAGCTCAGCTTACACGATTAAATCATATC
ACTAATTGTAAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGACACATTA
ACCTTAGAGGAAGTCAGACGGAATTTAATAACGCGCAGATAAACTTTGGCGGTAGTGGGGGATCTGGGGGAAGT
ATGGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCTGGATGGGCTGTCTATAACCGATGAATAC
AAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGCTCATTCGATTAAAAAGAATCTTATCGGTGCC

[0092]

CTCCTATTGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC
AAGAACCGAATATGTTACTTACAAGAAATTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTACCGT
TTGGAAGAGTCCTTCTTTCGCAAGAGGACAAGAAACATGAACGCGACCCCATCTTTGGAAACATAGTAGATGAG
GTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
CTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTTGAGGTGATCTAAAT
CCGGACAACCTCGGATGTGCAACAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCTT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCTCTCTAAATCCCGACGGCTAGAAAACCTG
ATCGCACAAATTACCCGGAGAGAAGAAAAATGGGTGTTTCGGTAACCTTATAGCGCTCTCACTAGGCTGACACCA
AATTTTAAGTCGAACCTTCGACTTAGCTGAAGATGCCAATTCGAGCTTAGTAAGGACACGTACGATGACGATCTC
GCAATCTAGCTGGCAGAAATGGAGATCAGTATGCGGACTTATTTTGGCTGCCAAAAACCTTAGCGATGCAATC
CTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATA
TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGCGGAGCGAGTCAAGAGGAATCTACAAGTTT
ATCAAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAACTCAATCGCGAAGATCTACTGCGA
AAGCAGCGGACTTTCGACAAACGGTAGCATTTCCACATCAAACTCCACTTAGCGGAATTGCGATGCTACTACTAGAAG
CAGGAGGATTTTTATCCGTTCTCTCAAAGACAATCGTGAAAAGATTGAGAAAATCTAACCTTTTCGCATACCTTAC
TATGTGGGACCCCTGGCCGAGGGAACCTCTCGGTTTCGATGAGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
TGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTTCATCGAGAGGATGACCAACTTTGACAAG
AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTACAGTGATACATGAATCAGC
AAAGTTAAGTATGTCTACTGAGGGCATGCGTAACCCGCTTTCTAAGCGGAGAACAGAGAAGAAAGCAATAGTAGAT
CTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTC
GATTCTGTGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATA
ATTAAAGATAAGGACTTCTTGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCTC
TTTGAAGATCGGGAAGATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTTCGACGATAAGGTTATGAAACAG
TTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTGCGGGAACCTTATCAACGGGATAAGAGACAAGCAAGT
GGTAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGAC
TCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCG
AATCTTGCTGGTTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTC
ATGGGACGTCACAAACCGGAAACATTTGTAATCGAGATTGGCACGCGAAATCAAAACGACTCAGAAGGGGCAAAAA
AACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCT
GTGGAAAATACCCAATTGCAGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
CAGGAACCTGGACATAAACGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGAT
TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCAGGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTC
GTAAGAAAAATGAGAATATTGGCGCGAGCTCTTAAATGCGAAAATGATAACGCAAAAGAAAGTTGATAAATTA
ACTAAAGCTGAGAGGGGTGGCTTGCTGAACTTGACAAGGCCGATTTATTAAACGTCAGCTCGTGGAAACCCGC
CAAATCACAAGCATGTTGACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATT
CGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTGCGACTTCAGAAAGGATTTTCAATTCTATAAAGTT
AGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAAGAA
TACCCGAAGCTAGAAAGTGAGTTTGTATGGTGATTACAAAGTTTATGACGTCCTGAAGATGATCGCGAAAAGC
GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAATC
ACTCTGGCAAAACGGAGAGATACGCAACGACCTTTAATGAAACCAATGGGAGACAGGTGAAATCGTATGGGAT
AAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAACCTGAGGTG
CAGACCGGAGGGTTTTCAAAGGAATCGATTTCTCCAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGAC
TGGGACCCGAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCATTCTGTCTAGTAGTGGCAAAAGTTGAG
AAGGGAATTCAGAAACTGAAGTCAGTCAAAGAATTATTGGGATAACGATTATGGAGCGCTCGTCTTTTGAA
AAGAACCCCATCGACTTCTTTGAGGCGAAAGGTTACAAGGAAGTAAAAAGGATCTCATAATTAACTACCAAAG
TATAGTCTGTTTGAAGTAAAGTGGCCGAAAACGGATGTTGGCTAGCGCCGAGAGCTTCAAAGGGGAACGAA
CTCGCTACCGTCTAAATACGTAATTTCCGTGATTATAGCTCCATTACGAGAAGTTGAAAGGTTTACCTGAA
GATAACGAACAGAAGCACTTTTTGTTGAGCAGCACAAACATTATCTGACGAAATCATAGAGCAAAATTTGCGAA
TTCAGTAAGAGAGTCATCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAAACGACAGGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGATTCAAG
TATTTTGACACAACGATAGATCGCAACGATACACTTACCAAGGAGGTGCTAGACGCGACACTGATTACCAA
TCCATCACGGGATTATATGAAACTCGGATAGATTGTTCACAGCTTGGGGGTGAC (서열 12)

[0093]

[0094]

fCas9:

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCC
CCCAAGAAGAAGAGGAAGGTGGGCATTACCCGCGGGGTACCTGGAGGTCTCGGATCCCAACTAGTCAAAAGTGAA
CTGGAGGAGAAGAAATCTGAACCTTCGTGATATAAATGAAATATGTGCCTCATGAATATATTGAATTAATTGAAAT
GCCAGAAATCCACTCAGGATAGAATCTTTGAAATGAAGGTAATGGAATTTTTATGAAAGTTTATGGATATAGA

[0095]

GGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCTGGATCTCCTATTGATTACGGTGTG
ATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTC
GAAGAAATCAAACACGAAACAAACATATCAACCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAA
TTTAAGTTTTTATTGTGAGTGGTCACTTTAAAGGAACTACAAGCTCAGCTTACACGATTAAATCATATCACT
AATTGTAATGGAGCTGTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACCATTAAAC
TTAGAGGAAGTCAGACGGAATTTAATAACGGCGAGATAAACTTTAGCGGCAGCGAGACTCCCGGGACCTCAGAG
TCCGCCACACCCGAAAGTGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATCCGTGGATGGGCTGTC
ATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTTCGATTAAGAA
AATCTTATCGGTGCCCTCCTATTTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGA
AGGTATACAGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTAGCAATGAGATGGCCAAAGTTGACGAT
TCTTTCTTTACCGTGTGGAAGAGTCTTCTTGTGCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGA
AACATAGTAGATGAGGTGGCATATCATGAAAGTACCAACAGATTTATCACCTCAGAAAAAGCTAGTTGACTCA
ACTGATAAAGCGGACCTGAGGTAACTTACTTGGCTCTTGCCCATATGATAAAAGTTCCGTGGGCACCTTCTCATT
GAGGGTGATCTAAATCCGGACAACCTCGGATGTCGACAAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTG
TTTGAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTAGCGCCCGCTCTCTAAATCCCGA
CGGCTAGAAAACCTGATCGCAAACTTACCCGGAGAGAAGAAAAATGGTGTGTCGGTAACTTATAGCGCTCTCA
CTAGCTGACACCAAAATTTTAAGTCGAACCTTCGACTTAGCTGAAAGATGCCAAATTCGAGCTTAGTAAGGACAG
TACGATGACGATCTCGCAAACTTACTTGGCACAATTTGAGATCAGTATGCGGACTTATTTTGGCTGCCAAAAAC
CTTAGCGATGCAATCCTCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCA
ATGATCAAAAGCGGACCTGAGGTAACTTACTTGGCTCTTGCCCATATGATAAAAGTTCCGTGGGCACCTTCTCATT
AAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAG
GAATCTTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACCTCAATCGC
GAAGATCTACTCGGAAACGCGGACTTTTCGACAACGGTAGCATTCACATCCAAATCCACTTAGGCGAATTGCTAT
GCTACTTAGAAGGCAGGAGGATTTTATCCGTTCTCAAAGACAATCGTGAAGAGATTGAGAAAAATCCTAACC
TTTCGCATACCTTACTATGTGGGACCCCTGGCCGAGGGAACCTCTCGGTTTCGATGGATGACAAGAAAGTCCGAA
GAAACGATTACTCCATGGAATTTTGGAGAGTTGTGCGATAAAGGTGCGTCAGCTCAATCGTTTATCGAGAGGATG
ACCAACTTTGACAGAATTTTACCGAACGAAAAAGTATTGCTTAAGCACAGTTTACTTTIACGAGTATTTCACAGTG
TACAATGAACCTCACGAAAGTTAAGTATGTCAGTGGGCGATGCGTAAACCCGCTTTCTAAGCGGAGAACAGAA
AAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAGTACAGTTAAGCAATTTGAAAGAGGACTACTTTAAG
AAAATTTGAATGCTTCGATTCTGTGCGAGATCTCCGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCAT
GACCTCCTAAAGATAATTAAGATAAGGACTTCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTG
TTGACTCTTACCCTCTTGAAGATCGGGAATGATTGAGGAAAGACTAAAACATACGCTCACCTGTTCGACGAT
AAGGTTATGAAACAGTTAAAGAGCGTCTGCTATACGGGCTGGGGACGATTGTGCGGGAACCTTATCAACGGGATA
AGAGCAACGCAAGAGTGTTAAACTATTCTCGATTCTTAAAGAGCGAGGCTTCGCCAATAGGAACCTTATGCGAG
CTGATCCATGATGACTCTTTAACTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTG
CACGAACATATTGCGAATCTTGTCTGTTTCGCCAGCCATCAAAAAGGCGATCTCCAGACAGTCAAAGTAGTGGAT
GAGCTAGTTAAGGTCTAGGAGCTCACAAACCGGAAACATTTGTAATCGAGATGGCACGCGAAAAATCAAACGACT
CAGAAGGGGCAAAAAACAGCTCGAGAGCGGATGAAGAGAATAGAAGAGGTTATTAAGAACTGGGCAGCCAGATC
TTAAAGGAGCATCTGTGGAATAATACCAATTCGAGAACGAGAACTTTACCTCTATTACCTACAAAATGGAAGG
GACATGTATGTTGATCAGGAACCTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATC
TTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCAGGGGAAAGTGACAATGTT
CCAAGCGAGGAAGTCGTAAGAAAAATGAAGAACTATTGGCGCGAGCTCCTAAATGCGAACTGATAACGCAAGAA
AAGTTGATAAATTAACATAAGCTGAGAGGGGTGGCTTGTCTGAACCTTGACAAGGCCGATTTATTAACAGCTCAG
CTCGTGGAAACCCGCAATACAAAGCATGTTGCACAGATACTAGATTCCGCAATGAATACGAAATACGACGAG
AACGATAAGCTGATTCCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTTGGTGTGCGGACTTCAGAAAGGATTTT
CAATCTATAAAGTTAGGGAGATAAATACTACCACTATGCGCACGACGCTTATCTTAATGCCGTCTGAGGGACC
GCATCTAATGAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCTGTAAG
ATGATCGCGGAAAGCGCAAGGAGATAGGCAAGGCTACAGCCAAATACTCTTTTATTCTAATCATTAATGAAATTT
TTTAAGACGGAATCACTCTGGCAACCGGAGAGATACGCAACGACCTTTAATTGAAACCAATGGGGAGACAGGT
GAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTGTCCATGCCCAAGTCAACATAGTA
AAGAAAACGTAGGTGCGAGACGCGGAGGGTTTCAAAGGAATCGATTCTTCCAAAAGGAATAGTGATAAGCTCATC
GCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCGCTATTCTGTCTCTAGTA
GTGGCAAAAGTTGAGAGGGGAAAAATCCAAGAACTGAAGTCAGTCAAAGATTTATGGGGATAACGATTATGGAG
CGCTCGTCTTTGAAAAGAACCCCATCGACTTCTTGAAGCGAAAGGTTACAAGGAAGTAAAAAGGATCTCATA
ATTAACTACCAAAGTATAGTCTGTTTGAAGTTAGAAAAATGGCCGAAACGGATGTTGGCTAGCGCCGAGAGCTT
CAAAAGGGGAAACGAACTCGCACTACCGTCTAAATACGTGAATTTCTGTATTTAGCGTCCCATTACGAGAAGTTG
AAAGGTTACCTGAAGATAACGAACAGAAGCACTTTTGTGAGCAGCACAAACATTATCTCGACGAAATCATA
GAGCAAAATTCGGAATTCAGTAAGAGAGTCATCTAGCTGATGCCAATCTGGACAAAGTATTAAAGCGCATACAAC
AAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAAATATTATCCATTGTTTACTCTTACCAACCTCGGCGCT
CCAGCCGATTCAGATTTTGAACAACGATAGATCGCAACGATACACTTCTACCAAGGAGGTGCTAGACGCG
ACAGTGATTACCAATCCATACGGGATTATATGAACTCGGATAGATTGTCACAGCTTGGGGTGAC (서열

[0096]

319)

[0097]

일부 실시양태에서, 본 발명의 융합 단백질 (예를 들어, fCas9)은 서열 9 내지 12 또는 서열 319 중 어느 하나
의 동족체에 상응하거나 또는 이에 의해 코딩된다.

[0098]

일부 실시양태에서, 본 발명의 융합 단백질 (예를 들어, fCas9)은 부분으로든 전체로든, 본원에 제공되고 다음
에 제시된 바와 같은, 서열 5, 서열 320, 서열 6, 서열 16, 서열 318, 및 서열 321로서 제시된 아미노산 서열
중 하나 이상을 포함한다. 서열 5, 서열 320, 서열 6, 서열 16, 서열 318, 및 서열 321에 상응하는 각종 도메
인은 서로에 대하여 어떠한 순서로도 배열될 수 있다. 예를 들어, 일부 실시양태에서, dCas9 도메인 (예를 들

어, 서열 5 또는 서열 320)은 아미노 또는 카르복시 말단에 있거나, 또는 아미노 말단과 카르복시 말단 사이의 어딘가에 있다. 유사하게, 서열 6, 서열 16, 서열 318, 및 서열 321에 상응하는 다른 도메인 각각은 본 발명의 융합 단백질 (예를 들어, fCas9)의 아미노 또는 카르복시 말단에 있거나, 또는 아미노 말단과 카르복시 말단 사이의 어딘가에 있다. 각종 도메인 배열을 갖는 본 발명의 융합 단백질의 예는 서열 9 내지 12 및 서열 319에 상응하는 본 발명의 융합 단백질을 포함한다. 일부 실시양태에서, 본 발명의 융합 단백질은 본원에 제공된 바와 같은 도메인 중 어느 것에 대하여 부가되거나 치환될 수 있는, 부가 또는 기타 도메인, 예컨대 기타 링커, 기타 NLS 도메인, 기타 뉴클레아제 도메인, 또는 기타 Cas9 도메인을 포함한다.

[0099] FokI 절단 도메인:

GSQLVKSELEKKSELRHKLKYPHEYYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTV
GSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWKVPSSVTEFKFLFVSGHFKGNYKAQ
LTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRKFNNGEINF (서열 6)

[0100]

[0101] dCas9:

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK
NRICYLQEIFSNEMAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDDTKADL
RLIYLAHMIKIFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLI
AQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAANKLSDAIL
LSDILRVNTEITKAPLSAMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI
KPILEKMDGTIELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYY
VGPLARGNSRFAMTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTK
VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII
KDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM
GRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEEGIKELGSQLKEHPVENTQLQNEKLYLYLQNGRDMYVDQ
ELDINRLSDYDVDAIVPQSFLKDDSIDNKVLRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLT
KAERGLSELKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIIVKKEVQTTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKY
SLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAHYEKLKGSPEQNEQKQLFVEQHKHYLDEIIIEQISEF
SKRVILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQ
ITGLYETRIDLSQLGGD (서열 320)

[0102]

[0103] 3xFLAG TAG:

[0104] MDYKDHGDYKDHDIDYKDDDDK (서열 321)

[0105] NLS 도메인:

[0106] MAPKKKRKVGIIHRGVP (서열 318)

[0107] XTEN 링커:

[0108] SGSETPGTSESATPES (서열 16)

[0109] 일부 실시양태에서, 이량체를 형성하는 융합 단백질은 단일 연장된 gRNA (예를 들어, 2개의 별도의 gRNA와는 달리, 각각 융합 단백질 이량체의 단량체와 결합된다)의 작용을 통하여 조정된다. 따라서, 일부 측면에서, 단일 연장된 gRNA는 표적 핵산을 보완하는 (예를 들어, 2개의 별개의 부위에서 표적 핵산과 결합하는), 링커 서열에 의해 분리된 2개 이상의 부분을 함유하고, 이러한 gRNA는 본원에 기재된 바와 같이, 2개 이상의 융합 단백질과 결합할 수 있다. 이는 도 1B에 나타난 도식에 의해 예시된다. 일부 실시양태에서, 상기 연장된 gRNA 중에서의 2개 부분을 분리시키는 링커 서열은 표적 서열과 상보성을 갖는다. 일부 실시양태에서, 상기 연장된 gRNA는 50개 이상, 60개 이상, 70개 이상, 80개 이상, 90개 이상, 100개 이상, 125개 이상, 150개 이상, 175개 이상, 200개 이상, 250개 이상, 300개 이상, 350개 이상, 400개 이상, 500개 이상, 600개 이상, 700개 이상, 800개 이상, 900개 이상, 또는 1000개 이상의 뉴클레오타이드 길이이다. 표적 핵산을 절단할 수 있는 이량체를 형성하기 위하여, 융합 단백질이 별도의 gRNA를 통해서 조정되는 단일 gRNA를 통해서 조정되는지 간에, 이러한 절단의 특이성은 단일 표적 핵산 결합 부위를 갖는 뉴클레아제와 비교해서 증강되는 것으로 (예를 들어, 표적을 벗어난 절단이 감소되거나 이러한 절단이 없는 것으로) 예상된다. 뉴클레아제의 특이성을 결정하는 방법은 공지되어 있다 (예를 들어, 공개된 PCT 출원, WO 2013/066438; 계류중인 가출원 US 61/864,289; 및 문헌 [Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature Methods* 8, 765-770 (2011)] 참조; 각각의 전문이 본원에 참조로 포

함된다).

[0110] 또 다른 실시양태에 따라서, Cas9 단백질의 이량체가 제공된다. 일부 실시양태에서, 이러한 이량체는 표적 핵산을 보완하는 2개 이상의 부분을 포함하는 단일 연장된 gRNA의 작용을 통하여 조정된다. 일부 실시양태에서, 표적 핵산에 대해 상보적인 부분은 표적 핵산을 보완하는 25개 이하, 24개 이하, 23개 이하, 22개 이하, 21개 이하, 20개 이하, 19개 이하, 18개 이하, 17개 이하, 16개 이하, 15개 이하, 14개 이하, 13개 이하, 12개 이하, 11개 이하, 10개 이하, 9개 이하, 8개 이하, 7개 이하, 6개 이하, 또는 5개 이하의 뉴클레오티드를 포함한다. 일부 실시양태에서, 표적 핵산에 대해 상보적인 부분은 5 내지 30개, 5 내지 25개, 또는 5 내지 20개 뉴클레오티드를 포함한다. 일부 실시양태에서, 표적 핵산에 대해 상보적인 부분은 15 내지 25개, 19 내지 21개, 또는 20개 뉴클레오티드를 포함한다. 일부 실시양태에서, 상기 부분은 표적 핵산을 보완하는 동일한 수의 뉴클레오티드를 포함한다. 일부 실시양태에서, 상기 부분은 표적 핵산을 보완하는 상이한 수의 뉴클레오티드를 포함한다. 예를 들어, 일부 실시양태에서, 상기 연장된 gRNA는 표적 핵산을 보완하는 (예를 들어, 그리고 표적 핵산과 혼성화되는) 2개 부분을 포함하는데, 각 부분은 표적 핵산을 보완하는 5 내지 19개, 10 내지 15개, 또는 10개 뉴클레오티드를 포함한다. 특별한 어떠한 이론에도 얽매지는 것은 아니지만, gRNA 특유의 대략 20개 미만의 뉴클레오티드를 포함하는 부분을 갖게 되면 (예를 들어, 대략 5 내지 19개, 10 내지 15개, 또는 10개 상보적 뉴클레오티드를 포함하는 부분을 갖게 되면), 단일 Cas9:gRNA 단위는 그 자체만으로는 효율적으로 결합될 수 없다. 따라서, 이러한 연장된 gRNA에 의해 잘 적응된 Cas9 단백질 간의 협조적 결합은, 의도된 표적 핵산의 특이성과 절단을 개선시켜 준다. 일부 실시양태에서, 연장된 gRNA의 2개 부분을 분리시켜 주는 링커 서열은 표적 서열과 상보성을 지닌다. 예를 들어, 일부 실시양태에서, 링커 서열은 표적 핵산을 보완하는 5개 이상, 6개 이상, 7개 이상, 8개 이상, 9개 이상, 10개 이상, 11개 이상, 12개 이상, 13개 이상, 14개 이상, 15개 이상, 20개 이상, 25개 이상, 30개 이상, 40개 이상, 또는 50개 이상의 뉴클레오티드를 갖는다. 특별한 어떠한 이론에도 얽매지는 것은 아니지만, Cas9 단백질에 의해 결합되는 것 뿐만 아니라 링커 서열 내의 것을 포함한, 다중 결합 부위 (예를 들어, 다중 저 친화성 결합 부위)를 포함하는 연장된 gRNA를 갖게 되면, 협조적 결합성이 증진됨으로써 증가된 특이성이 제공된다. 이러한 실시양태의 특정 측면이 도 4에 도시된다. 일부 실시양태에서, 본원에 기재된 Cas9 단백질 중 어느 것도 단일 연장된 gRNA를 통하여 잘 적응될 수 있다.

[0111] 또 다른 실시양태에서, RNA-프로그램 가능한 뉴클레아제 (예를 들어, Cas9)의 단편을 포함하는 단백질이 제공된다. 예를 들어, 일부 실시양태에서, Cas9의 gRNA 결합성 도메인을 포함하는 단백질이 제공된다. 일부 실시양태에서, Cas9의 gRNA 결합성 도메인을 포함하는 단백질은 DNA 절단 도메인 (본원에서 Cas9의 "A-절반"으로서 지칭됨)을 포함하지 않는다. 다른 실시양태에서, Cas9의 DNA 절단 도메인(들) (예를 들어, HNH, RuvC1 서브도메인)을 포함하는 단백질이 제공된다. 일부 실시양태에서, 이러한 "DNA 절단 도메인"은 집합적으로, 이중 가닥 DNA 절단을 위해 필요한 Cas9의 부분 (예를 들어, HNH, RuvC1 서브도메인)을 지칭한다. 일부 실시양태에서, Cas9의 DNA 절단 도메인을 포함하는 단백질은 gRNA 결합성 도메인 (본원에서 Cas9의 "B-절반"으로서 지칭됨)을 포함하지 않는다. 일부 실시양태에서, (i) Cas9의 gRNA 결합성 도메인 (예를 들어, A-절반)을 포함하는 단백질, 및 (ii) Cas9의 DNA 절단 도메인 (예를 들어, B-절반)을 포함하는 단백질을 포함하는 이량체가 제공된다. 일부 실시양태에서, 이러한 이량체는 gRNA에 의해 결합된다. 예를 들어, 상기 이량체는 완전한 길이의 Cas9 단백질의 결합 및 절단 활성을 반복 발생시키는 것으로 예상된다. 일부 실시양태에서, 상기 이량체는 본원에서 "이량체성 분할 Cas9"로서 지칭된다. 표적 핵산을 절단하기 위해 이량체성 분할 Cas9를 사용하는 것이, 단일의 완전한 길이의 Cas9 단백질과 비교해서 증가된 특이성을 제공하는 것으로 예상되는데, 이는 상기 단백질의 절반 둘 다가 뉴클레아제-활성 상태가 되도록 연합 및 재폴딩되기 위해 공동-국재되어야만 하기 때문이다. 이러한 전략은 도 2A의 도식에 제시된다.

[0112] 일부 실시양태에서, 다음 2개의 도메인을 포함하는 융합 단백질이 제공된다: (ii) RNA-프로그램 가능한 뉴클레아제의 단편 (예를 들어, 본원에 기재된 바와 같은, Cas9의 A- 또는 B-절반)과 융합되거나 연결된 (i) 표적 핵산과 특이적으로 결합할 수 있는 단백질 (예를 들어, 본원에 기재된 바와 같은, 뉴클레아제-불활성화 RNA 프로그램 가능한 뉴클레아제, 예컨대 뉴클레아제-불활성화 Cas9). 일부 실시양태에서, 전술된 융합 단백질의 도메인 (i)은 DNA 결합성 도메인, 예를 들어 아연 핑거 또는 TALE 단백질의 DNA 결합성 도메인을 포함한다. 일부 실시양태에서, 상기 융합 단백질은 (i) 뉴클레아제-불활성화 Cas9, 및 (ii) Cas9의 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반)을 포함한다. 일부 실시양태에서, 상기 융합 단백질의 도메인 (ii)은 DNA 절단 도메인을 포함하지 않는다. 다른 실시양태에서, 상기 융합 단백질은 (i) 뉴클레아제-불활성화 Cas9, 및 (ii) DNA 절단 도메인 (예를 들어, Cas9 B-절반)을 포함한다. 일부 실시양태에서, 상기 융합 단백질의 도메인 (ii)은 gRNA 결합성 도메인을 포함하지 않는다.

[0113] 일부 실시양태에서, 다음 2개의 단백질을 포함하는 이량체가 제공된다: (i) 뉴클레아제-불활성화 Cas9와 Cas9의 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반과 융합된 뉴클레아제-불활성화 Cas9)을 포함하는 융합 단백질, 및 (ii) Cas9의 DNA 절단 도메인 (예를 들어, Cas9 B-절반)을 포함하는 단백질. 다른 실시양태에서, 상기 이량체는 (i) 뉴클레아제-불활성화 Cas9와 Cas9의 DNA 절단 도메인 (예를 들어, Cas9 B-절반과 융합된 뉴클레아제-불활성화 Cas9)을 포함하는 융합 단백질, 및 (ii) Cas9의 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반)을 포함하는 단백질을 포함한다. 일부 실시양태에서, 이러한 단백질 이량체는 하나 이상의 gRNA를 포함한다. 예를 들어, 일부 실시양태에서, 상기 이량체는 2개의 gRNA를 포함하는데, 하나는 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인에 의해 결합되고; 다른 하나는 A-절반 도메인 (예를 들어, 융합 단백질의 A-절반, 또는 융합 단백질의 일부가 아닌 이량체의 A-절반)에 의해 결합된다. 상기 이량체 (예를 들어, 표적 핵산의 별도의 영역과 결합하는 서열을 갖는 2개의 gRNA와 연합됨)는, 예를 들어 단일 gRNA를 갖는 Cas9 단백질과 비교해서 개선된 특이성을 지닌 것으로 예상된다. 이러한 전략은 도 2B에 도시된다.

[0114] 일부 실시양태에서, 다음 2개의 융합 단백질을 포함하는 단백질 이량체가 제공된다: (i) 뉴클레아제-불활성화 Cas9와 Cas9의 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반과 융합된 뉴클레아제-불활성화 Cas9)을 포함하는 융합 단백질, 및 (ii) 뉴클레아제-불활성화 Cas9와 DNA 절단 도메인 (예를 들어, Cas9 B-절반과 융합된 뉴클레아제-불활성화 Cas9)을 포함하는 융합 단백질. 일부 실시양태에서, 상기 이량체는 하나 이상의 별개의 gRNA와 연합 (예를 들어, 결합)된다. 예를 들어, 일부 실시양태에서, 상기 이량체는 2개 또는 3개의 gRNA와 연합된다. 일부 실시양태에서, 상기 이량체는 3개의 gRNA와 연합된다. 예를 들어, 핵산 표적의 특정 영역에 대한 하나의 뉴클레아제-불활성화 Cas9:gRNA의 결합, 및 핵산 표적의 제2 영역에 대한 다른 뉴클레아제-불활성화 Cas9:gRNA의 결합시, 분할 Cas9 절반들 (예를 들어, 융합 단백질의 A-절반과 B-절반)은 이량체화될 수 있고, 핵산 표적의 제3 영역에 대해 상보적인 제3의 gRNA와 결합되어, 완전한 활성의 Cas9 뉴클레아제로 될 수 있는데, 이는 dsDNA를 절단할 수 있다. 이러한 전략은 도 2C에 예시된다.

[0115] 본 발명의 또 다른 측면에 따라서, 최소화된 Cas9 단백질이 제공된다. "최소화된다"는 것은, Cas9 단백질이 야생형 단백질과 비교해서 아미노산 결실 및/또는 끝을 잘라낸 것을 포함하지만, gRNA 결합 활성, DNA 절단 활성, 또는 둘 다는 보유하고 있다는 것을 의미한다. Cas9 단백질 (예를 들어, 분할 Cas9 단백질, Cas9 A-절반, Cas9 B-절반, 뉴클레아제-불활성화 Cas9 융합 단백질 등)을 설명하는 본원의 실시양태 중 어느 것도 최소화 Cas9 단백질을 활용할 수 있다. 일부 실시양태에서, N-말단 결실 및/또는 끝을 잘라낸 것을 포함하는 최소화 Cas9 단백질이 제공된다. 일부 실시양태에서, C-말단 결실 및/또는 끝을 잘라낸 것을 포함하는 최소화 Cas9 단백질이 제공된다. 일부 실시양태에서, N- 및/또는 C-말단 결실 및/또는 끝을 잘라낸 것을 포함하는 최소화 Cas9 단백질이 제공된다. 일부 실시양태에서, 이러한 최소화 Cas9 단백질은 gRNA 결합 활성과 DNA 절단 활성 둘 다를 보유하고 있다. 일부 실시양태에서, 최소화 Cas9 단백질은 5개 이상, 10개 이상, 15개 이상, 20개 이상, 25개 이상, 40개 이상, 40개 이상, 50개 이상, 75개 이상, 100개 이상, 150개 이상, 200개 이상, 250개 이상, 300개 이상, 350개 이상, 400개 이상, 450개 이상, 또는 500개 이상의 아미노산을 제거하는, N-말단 끝을 잘라내는 것을 포함한다. 일부 실시양태에서, 최소화 Cas9 단백질은 5개 이상, 10개 이상, 15개 이상, 20개 이상, 25개 이상, 40개 이상, 40개 이상, 50개 이상, 75개 이상, 100개 이상, 150개 이상, 200개 이상, 250개 이상, 300개 이상, 350개 이상, 400개 이상, 450개 이상, 또는 500개 이상의 아미노산을 제거하는, C-말단 끝을 잘라내는 것을 포함한다. 일부 실시양태에서, 결실은 Cas9 내에서, 예를 들어 gRNA 결합성 및/또는 DNA 절단에 영향을 미치지 않는 영역에서 이루어진다. 일부 실시양태에서, 최소화 Cas9 단백질은 하나 이상의 gRNA와 연합된다. 특정 실시양태에서, 최소화 Cas9 단백질은 하나의 gRNA와 연합된다.

[0116] 재조합효소

[0117] 본 개시내용의 일부 측면은 본원에 기재된 방법 및 전략을 이용하여 설계되는 RNA-가이드된 재조합효소 융합 단백질을 제공한다. 본 개시내용의 일부 실시양태는 상기 재조합효소를 코딩하는 핵산을 제공한다. 본 개시내용의 일부 실시양태는 상기 코딩 핵산을 포함하는 발현 구조물을 제공한다. 예를 들어, 일부 실시양태에서 특정 게놈 내의 목적하는 표적 부위 (예를 들어, 조작된 재조합효소 중 하나 이상과 결합된 하나 이상의 gRNA에 의해 표적화된 부위)를, 예를 들어 상기 게놈 내의 또 다른 부위 또는 외인성 핵산과 재조합하도록 조작시킨 단리된 재조합효소가 제공된다. 일부 실시양태에서, 이러한 단리된 재조합효소는 RNA-프로그램 가능한 뉴클레아제, 예컨대 Cas9 뉴클레아제의 변이체를 포함한다. 일부 실시양태에서, Cas9 변이체는 뉴클레아제-불활성화 Cas9 (예를 들어, dCas9)이다. 일부 실시양태에서, dCas9는 부분으로든 전체로든, 서열 5 또는 서열 320을 포함하는 뉴클레오티드 서열에 의해 코딩된다. 일부 실시양태에서, dCas9는 서열 5 또는 서열 320의 변이체를 포함하는 뉴클레오티드 서열에 의해 코딩된다.

[0118] 한 실시양태에서, RNA-가이드된 재조합효소 융합 단백질이 제공된다. 전형적으로, 이러한 융합 단백질은 2개 이상의 도메인을 포함한다. 일부 실시양태에서, 상기 융합 단백질은 2개의 도메인을 포함한다. 일부 실시양태에서, 2개 이상의 도메인 중 하나는 뉴클레아제-불활성화 Cas9 (또는 그의 단편, 예를 들어 Cas9 A-절반), 예를 들어 본원에 기재된 것 (예를 들어, dCas9)이다. 재조합효소 융합 단백질의 Cas9 도메인은 하나 이상의 gRNA와 결합할 수 있으므로, 재조합효소 융합 단백질(들)이, 예를 들어 본원에 기재된 바와 같이, 표적 핵산을 지시하거나 또는 이를 표적으로 한다. 2개 이상의 도메인 중 또 다른 도메인은 재조합효소, 또는 그의 단편, 예를 들어 재조합효소의 촉매 도메인이다. "재조합효소의 촉매 도메인"이란, 융합 단백질이 특정 재조합효소의 아미노산 서열 (예를 들어, 특정 재조합효소로부터 유래된 아미노산 서열)을 포함하는 도메인을 포함하므로, 이러한 도메인이 표적 핵산과 접촉되는 경우에 (단독으로, 또는 융합 단백질의 일부를 형성할 수 있거나 형성할 수 없는 다른 재조합효소 촉매 도메인을 포함한 부가 인자와 함께) 재조합을 유도시키기에 충분하다는 것을 의미한다. 일부 실시양태에서, 재조합효소의 촉매 도메인은 재조합효소의 DNA 결합성 도메인을 배제한다. 일부 실시양태에서, 재조합효소의 촉매 도메인은 특정 재조합효소의 일부 또는 전부를 포함하는데, 예를 들어 촉매 도메인은 재조합효소 도메인 및 DNA 결합성 도메인, 또는 그의 일부를 포함할 수 있거나, 또는 촉매 도메인은 DNA 결합 활성을 없애기 위하여 돌연변이되거나 끝이 잘려지는 DNA 결합성 도메인 및 재조합효소 도메인을 포함할 수 있다. 재조합효소 및 재조합효소의 촉매 도메인은 관련 기술분야의 통상의 기술자에게 공지되어 있고, 예를 들어 본원에 기재된 것을 포함한다. 일부 실시양태에서, 촉매 도메인은 어떠한 재조합효소로부터도 유래된다. 일부 실시양태에서, 재조합효소 촉매 도메인은 Tn3 위치 특이성 재조합 촉진 효소, Hin 재조합효소, 또는 Gin 재조합효소의 촉매 도메인이다. 일부 실시양태에서, 촉매 도메인은 부분으로든 전체로든, 다음에 제공된 바와 같은 서열 322를 포함하는 뉴클레오티드 서열에 의해 코딩되는 Tn3 위치 특이성 재조합 촉진 효소 [예를 들어, 스타르크(Stark) Tn3 재조합효소]를 포함한다. 일부 실시양태에서, Tn3 촉매 도메인은 서열 322의 변이체에 의해 코딩된다. 일부 실시양태에서, Tn3 촉매 도메인은 서열 325에 상응하는 폴리펩티드를 코딩하는 폴리뉴클레오티드 (또는 그의 변이체)에 의해 코딩된다. 일부 실시양태에서, 촉매 도메인은 부분으로든 전체로든, 다음에 제공된 바와 같은 서열 323을 포함하는 뉴클레오티드 서열에 의해 코딩되는 Hin 재조합효소를 포함한다. 일부 실시양태에서, Hin 촉매 도메인은 서열 323의 변이체에 의해 코딩된다. 일부 실시양태에서, Hin 촉매 도메인은 서열 326에 상응하는 폴리펩티드를 코딩하는 폴리뉴클레오티드 (또는 그의 변이체)에 의해 코딩된다. 일부 실시양태에서, 촉매 도메인은 부분으로든 전체로든, 다음에 제공된 바와 같은 서열 324를 포함하는 뉴클레오티드 서열에 의해 코딩되는 Gin 재조합효소 (예를 들어, Gin 베타 재조합효소)를 포함한다. 일부 실시양태에서, Gin 촉매 도메인은 서열 324의 변이체에 의해 코딩된다. 일부 실시양태에서, Gin 촉매 도메인은 서열 327에 상응하는 폴리펩티드를 코딩하는 폴리뉴클레오티드 (또는 그의 변이체)에 의해 코딩된다.

[0119] 스타르크 Tn3 재조합효소 (뉴클레오티드: 서열 322; 아미노산: 서열 325):

ATGGCCCTGTTTGGCTACGCACGCGTGTCTACCACTCAACAGTCACTCGATTGCAAGTGAGGGCTCTTAAAGAT
GCCGGAGTGAAGGCAACAGAATTTTACTGATAAGGCCAGCGGAAGCAGCACAGACAGAGAGGGGCTGGATCTC
CTGAGAATGAAGGTAAAGGAGGGTGATGTGATCTTTGGTCAAAAATTGGATCGACTGGGGAGAGACACAGCTGAT
ATGCTTCAGCTTATTAAGAGTTTGACGCTCAGGGTGTGCGGTGAGGTTTATCGATGACGGCATCTCAACCGAC
TCCTACATTGGTCTTATGTTTGTGACAATTTGTCCGCTGTGGCTCAGGCTGAGCGGAGAAGGATTCTCGAAAGG
ACGAATGAGGGACGGCAAGCAGCTAAGTTGAAAGGTATCAAATTTGGCAGACGAAGG (서열 322)

MALFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVKEGDVILVKKLDRLGRDAD
MLQLIKEFDAQGVAVRFIDDGISTDSYIGLMFVTILSAVAQAERRRILERTNEGRQAARKLGKIKFGRRR (서열
325)

[0120]

[0121] Hin 재조합효소 (뉴클레오티드: 서열 323; 아미노산: 서열 326):

ATGGCAACCATTGGCTACATAAGGGTGTCTACCATCGACCAAAATATCGACCTGCAGCGCAACGCTCTGACATCC
GCCAAGTGCATCGGATCTTCGAGGATAGGATCAGTGGCAAGATCGCCAACCGGCCCGGTCTGAAGCGGGCTCTG
AAGTACGTGAATAAGGGCGATACTCTGGTTGTGTGGAAGTTGGATCGCTTGGGTAGATCAGTGAAGAATCTCGTA
GCCCTGATAAGCGAGCTGCACGAGAGGGGTGCACATTTCCATTCTTGACCGATTCCATCGATACGTCTAGCGCC
ATGGGCCGATTCTTCTTTTACGTCAATGTCCGCCCTCGCTGAAATGGAGCGCGAACTTATTGTTGAACGGACTTTG
GCTGGACTGGCAGCGGCTAGAGCACAGGGCCGACTTGA (서열 323)

MATIGYIRVSTIDQNIIDLQRNALTSANCDRIFEDRISGKIANRPLKRALKYVNKGDTLVVWKLDRGRSVKNLV
ALISELHERGAHFHSLTDSIDTSSAMGRFFFYVMSALAEEMERELIVERTLAGLAAARAQGRGLG (서열 326)

[0122]

[0123] Gin 베타 재조합효소 (뉴클레오티드: 서열 324; 아미노산: 서열 327):

ATGCTCATTGGCTATGTAAGGGTCAGCACCAATGACCAAAACACAGACTTGCAACGCAATGCTTTGGTTTGCGCC
GGATGTGAACAGATATTTGAAGATAAACTGAGCGGCACTCGGACAGACAGACCTGGGCTTAAGAGAGCACTGAAA
AGACTGCAGAAGGGGGACACCTGGTCGTCTGGAACTGGATCGCCTCGGACGCAGCATGAAACATCTGATTAGC
CTGGTTGGTGAGCTTAGGGAGAGAGGAATCAACTTCAGAAGCCTGACCGACTCCATCGACACCAGTAGCCCCATG
GGACGATTCTTCTCTATGTGATGGGAGCACTTGTCTGAGATGGAAAGAGAGCTTATTATCGAAAGAACTATGGCT
GGTATCGCTGCTGCCCGGAACAAAGGCAGACGGTTCGGCAGACCGCCGAAGAGCGGC (서열 324)

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTRDPLKRALKRLQKGDTLVVWKLDRIGRSMKHLIS
LVGELRERGINFRSLTDSIDTSSPMGRFFFFVYMGALEMERELIIERTMAGIAAARNKGRFRFGRPPKSG (서열
327)

[0124]

[0125] 일부 실시양태에서, 재조합효소 촉매 도메인은 Cas9 단백질 (예를 들어, sCas9)의 N-말단, C-말단, 또는 N-말단과 C-말단 사이의 어딘가와 융합된다. 일부 실시양태에서, 융합 단백질은 핵 국제화 신호 (NLS; 예를 들어, 본원에 제공된 것 중 어느 것)를 추가로 포함한다. 예를 들어, 일부 실시양태에서, 예시되는 RNA-가이드된 재조합효소 융합 단백질 (예를 들어, Cas9-재조합효소 융합물)의 일반적 구조는 다음 구조 중 하나를 포함한다:

[0126] [NH₂]-[Cas9]-[재조합효소]-[COOH],

[0127] [NH₂]-[재조합효소]-[Cas9],

[0128] [NH₂]-[NLS]-[Cas9]-[재조합효소]-[COOH],

[0129] [NH₂]-[NLS]-[재조합효소]-[Cas9]-[COOH],

[0130] [NH₂]-[Cas9]-[NLS]-[재조합효소]-[COOH],

[0131] [NH₂]-[재조합효소]-[NLS]-[Cas9]-[COOH],

[0132] [NH₂]-[Cas9]-[재조합효소]-[NLS]-[COOH], 또는

[0133] [NH₂]-[재조합효소]-[Cas9]-[NLS]-[COOH]

[0134] (여기서, NLS는 핵 국제화 신호이고, NH₂는 융합 단백질의 N-말단이며, COOH는 융합 단백질의 C-말단이다). 일부 실시양태에서, 링커가 Cas9 도메인과 재조합효소 도메인 사이에 삽입되는데, 예를 들어 본원에 제공된 모든 링커가 있다. 부가의 특징, 예컨대 서열 태그 (예를 들어, 본원에 제공된 것 중 어느 것)이 또한 존재할 수 있다.

[0135] 제약 조성물

[0136] 일부 실시양태에서, 본원에 기재된 뉴클레아제 (예를 들어, 뉴클레아제 또는 뉴클레아제 도메인을 포함하는 융합 단백질) 및 재조합효소 (예를 들어, 재조합효소 또는 재조합효소 촉매 도메인을 포함하는 융합 단백질) 중 어느 것도 제약 조성물의 일부로서 제공된다. 예를 들어, 일부 실시양태는 본원에 제공된 바와 같은 뉴클레아제 및/또는 재조합효소, 또는 이러한 뉴클레아제 및/또는 재조합효소를 코딩하는 핵산, 및 제약상 허용되는 부형제를 포함하는 제약 조성물을 제공한다. 제약 조성물은 임의로, 하나 이상의 부가 치료상 활성 물질을 포함할 수 있다.

[0137] 일부 실시양태에서, 본원에 제공된 조성물은 대상체, 예를 들어 인간 대상체에게 투여되어, 이러한 대상체 내에서의 표적화된 게놈 변형을 수행한다. 일부 실시양태에서, 세포를 상기 대상체로부터 수득하고; 이를 생체 외에서 뉴클레아제 및/또는 재조합효소와 접촉시킨다. 일부 실시양태에서, 대상체로부터 제거되고 생체 외에서 본 발명의 뉴클레아제 및/또는 재조합효소와 접촉시킨 세포를, 임의로 목적하는 게놈 변형을 수행하였거나 또는 세포에서 탐지한 후에 상기 대상체에게 재도입한다. 본원에 제공된 제약 조성물의 설명이 주로, 인간에게 투여하는 데 적합한 제약 조성물에 관한 것이긴 하지만, 이러한 조성물이 일반적으로, 모든 종류의 동물에게 투여하는 데 적합하다는 것을 통상의 기술자는 이해할 것이다. 제약 조성물이 각종 동물에게 투여하는 데 적합하도록 만들기 위해, 인간에게 투여하는 데 적합한 제약 조성물을 변형시키는 것이 널리 이해되고 있고, 통상의 전문 기술을 가진 수의 약리학자는, 있다 하더라도 단지 통상적인 실험을 이용해서 상기 변형을 설계하고/하거나 수행할 수 있다. 제약 조성물이 투여되는 것으로 고려되는 대상체는 인간 및/또는 다른 영장류; 포유류, 가축,

애완용 동물, 및 상업적으로 관련된 포유류, 예컨대 소, 돼지, 말, 양, 고양이, 개, 마우스, 및/또는 래트; 및/또는 조류 (상업적으로 관련된 조류, 예컨대 치킨, 오리, 거위 및/또는 칠면조 포함)를 포함하지만, 이에 제한되지 않는다.

[0138] 본원에 기재된 제약 조성물의 제형은 약리학 분야에서 공지되어 있거나 추후 개발될 어떠한 방법에 의해서도 제조할 수 있다. 일반적으로, 이러한 제조 방법은 활성 성분을 부형제와 연합시키는 단계, 및 이어서, 필요하고/하거나 원하는 경우에는 상기 생성물을 목적하는 단일 용량 단위 또는 다수 용량 단위로 성형하고/하거나 패키징하는 단계를 포함한다.

[0139] 제약 제형은 부가적으로, 본원에 사용된 바와 같이, 목적하는 특별한 투여 형태에 적합한 바 대로, 모든 용매, 분산 매질, 희석제, 또는 기타 액상 비히클, 분산 또는 현탁 보조제, 표면 활성제, 등장제, 증점제 또는 유화제, 보존제, 고형 결합제, 윤활제 등을 포함하는 제약상 허용되는 부형제를 포함할 수 있다. 문헌 [Remington's *The Science and Practice of Pharmacy*, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006); 그의 전문이 본원에 참조로 포함된다]에는 제약 조성물을 제형화하는 데 사용된 각종 부형제 및 그의 제조를 위한 공지된 기술이 개시되어 있다. 뉴클레아제를 포함하는 제약 조성물을 생성시키는 데 적합한 부가의 방법, 시약, 부형제 및 용매에 관해서는 또한, PCT 출원 PCT/US2010/055131을 참조할 수 있다 (그의 전문이 본원에 참조로 포함된다). 예컨대 바람직하지 못한 모든 생물학적 효과를 유발시키거나 또는 달리 유해한 방식으로 제약 조성물의 다른 성분(들)과 상호 작용함으로써, 통상적인 모든 부형제 매질이 특정 물질 또는 그의 유도체와 비화합성인 경우를 제외하고는, 상기 통상적인 부형제 매질의 용도가 본 개시 내용의 범위 내에 있는 것으로 고려된다.

[0140] 일부 실시양태에서, 본 발명에 따르는 조성물은 다음 중 하나 이상을 포함하지만, 그에 제한되지 않는, 각종 질환, 장애 및/또는 병태 중 어느 것을 치료하는 데 사용될 수 있다: 자가면역 장애 (예를 들어, 당뇨병, 루푸스, 다발성 경화증, 건선, 류마티스성 관절염); 염증성 장애 (예를 들어, 관절염, 골반 염증성 질환); 감염성 질환 [예를 들어, 바이러스성 감염 (예를 들어, HIV, HCV, RSV), 박테리아성 감염, 진균성 감염, 패혈증]; 신경학적 장애 [예를 들어, 알츠하이머병(Alzheimer's disease), 헌팅톤병; 자폐증; 뒤시엔느(Duchenne) 근위축증]; 심혈관계 장애 (예를 들어, 아테롬성 경화증, 고콜레스테롤혈증, 혈전증, 응고 장애, 혈관형성성 장애, 예컨대 황반 변형); 증식성 장애 (예를 들어, 암, 양성 종양); 호흡기 장애 (예를 들어, 만성 폐쇄성 폐 질환); 소화기 장애 (예를 들어, 염증성 장 질환, 궤양); 근골격계 장애 (예를 들어, 섬유근육통, 관절염); 내분비, 대사 및 영양 장애 (예를 들어, 당뇨병, 골다공증); 비뇨기와 장애 (예를 들어, 신장 질환); 심리적 장애 (예를 들어, 우울증, 정신분열증); 피부 장애 (예를 들어, 상처, 습진); 혈액 및 림프성 장애 (예를 들어, 빈혈, 혈우병) 등.

[0141] 부위-특이적 핵산 절단 방법

[0142] 본 개시내용의 또 다른 실시양태에서, 부위-특이적 핵산 (예를 들어, DNA) 절단 방법이 제공된다. 일부 실시양태에서, 이러한 방법은 DNA를 본원에 기재된 Cas9:gRNA 복합체 중 어느 것과 접촉시키는 단계를 포함한다. 예를 들어, 일부 실시양태에서, 상기 방법은 DNA를, 다음 2개의 도메인을 포함하는 융합 단백질 (예를 들어, fCas9)과 접촉시키는 단계를 포함한다: (i) 뉴클레아제-불활성화 Cas9 (dCas9); 및 (ii) 뉴클레아제 (예를 들어, FokI DNA 절단 도메인) (여기서, 불활성 Cas9 도메인은 상기 DNA의 특정 영역과 혼성화되는 gRNA와 결합된다). 일부 실시양태에서, 상기 방법은 DNA를 본원에 기재된 제2 융합 단백질 (예를 들어, fCas9)과 접촉시키는 단계 (여기서, 제2 융합 단백질의 뉴클레아제-불활성화 Cas9 (dCas9) 도메인은 DNA의 제2 영역과 혼성화되는 제2의 gRNA와 결합된다)를 추가로 포함한다 (여기서, 상기 융합 단백질의 결합으로 인해, 융합 단백질의 뉴클레아제 도메인이 이량체화되어, 상기 DNA가 상기 결합된 융합 단백질 사이의 영역에서 절단된다) (예를 들어, 도 1A, 6d 참조). 일부 실시양태에서, 각각의 융합 단백질과 결합된 gRNA는 DNA의 동일 가닥과 혼성화되거나, 또는 이들은 DNA의 반대 가닥과 혼성화된다. 일부 실시양태에서, 상기 gRNA는 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어진 DNA의 영역과 혼성화된다. 결합된 Cas9:gRNA 복합체 사이의 영역이 "스페이스 서열"로서 지칭될 수 있는데, 이는 전형적으로 표적 핵산이 절단되는 곳이다 (예를 들어, 도 6c-d 참조). 일부 실시양태에서, 스페이스 서열은 5개 이상, 10개 이상, 15개 이상, 20개 이상, 25개 이상, 30개 이상, 35개 이상, 40개 이상, 45개 이상, 50개 이상, 60개 이상, 70개 이상, 80개 이상, 90개 이상, 또는 100개 이상의 염기 쌍 길이이다. 일부 실시양태에서, 스페이스 서열은 약 5개 내지 약 50개 염기 쌍, 약 10개 내지 약 40개 염기 쌍, 또는 약 15개 내지 약 30개 염기 쌍 길이이다. 일부 실시양태에서, 스페이스 서열은 약 15개 내지 약 25개 염기 쌍 길이

이다. 일부 실시양태에서, 스페이서 서열은 약 15개, 약 20개, 또는 약 25개 염기 쌍 길이이다. 일부 실시양태에서, Cas9:gRNA 복합체는 본원에 기재된 바와 같이, A 배향으로 결합된다. 일부 실시양태에서, Cas9:gRNA 복합체는 본원에 기재된 바와 같이, B 배향으로 결합된다. 일부 실시양태에서, 상기 방법은 야생형 Cas9 또는 다른 Cas9 변이체를 활용하는 방법의 표적에 적중한:표적을 벗어난 변형 비 보다 2배 이상, 5배 이상, 10배 이상, 20배 이상, 30배 이상, 40배 이상, 50배 이상, 60배 이상, 70배 이상, 80배 이상, 90배 이상, 100배 이상, 110배 이상, 120배 이상, 130배 이상, 140배 이상, 150배 이상, 175배 이상, 200배 이상, 또는 250배 이상 더 높은 표적에 적중한:표적을 벗어난 변형 비를 갖는다. 일부 실시양태에서, 상기 방법은 야생형 Cas9 또는 다른 Cas9 변이체를 활용하는 방법의 표적에 적중한:표적을 벗어난 변형 비 보다 약 60배 내지 약 180배, 약 80배 내지 약 160배, 약 100배 내지 약 150배, 또는 약 120배 내지 약 140배 더 높은, 표적에 적중한:표적을 벗어난 변형 비를 갖는다. 표적에 적중한:표적을 벗어난 변형 비를 결정하는 방법은 공지되어 있고, 본 실시예에 기재된 것을 포함한다. 일부 실시양태에서, 융합 단백질은, 예를 들어 본원에 기재된 바와 같이, 단일 gRNA를 통하여 조정되거나 연합된다.

[0143] 일부 실시양태에서, 상기 방법은 특정 핵산을, 본원에 기재된 바와 같은 단일 gRNA와 조정된 (예를 들어, 단일 gRNA에 의해 결합된) Cas9 단백질 (또는 그의 단편)의 이량체와 접촉시키는 단계를 포함한다. 일부 실시양태에서, 이러한 단일 gRNA는 핵산과 혼성화되는 2개 이상의 부분을 포함한다. 일부 실시양태에서, 이러한 부분은 5개 이상, 10개 이상, 15개 이상, 또는 19개 이상의 상보적 뉴클레오티드를 포함한다. 일부 실시양태에서, 상기 부분은 20개 미만의 상보적 뉴클레오티드를 포함한다. 일부 실시양태에서, 링커 서열이 이러한 부분을 분리시키는데, 여기서 링커 서열은 또한, 표적 핵산에 대해 상보적인 (예를 들어, Cas9 단백질에 의해 결합되지 않는) 뉴클레오티드를 포함한다. 일부 실시양태에서, 링커 서열은 표적 핵산과 혼성화되지 않는다.

[0144] 일부 실시양태에서, 상기 방법은 DNA를, 본원에 기재된 바와 같은 융합 단백질의 단백질 이량체와 접촉시키는 단계 (여기서, 융합 단백질은 하나 이상의 gRNA에 의해 결합된다)를 포함한다. 예를 들어, 일부 실시양태에서, 상기 이량체의 하나의 융합 단백질은 Cas9의 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반)을 포함하고 [여기서, 상기 단백질은 DNA 절단 도메인 (예를 들어, Cas9 B-절반)을 포함하지 않는다]; 이량체의 다른 융합 단백질은 Cas9의 DNA 절단 도메인 (예를 들어, Cas9 B-절반)을 포함한다 [여기서, 상기 단백질은 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반)을 포함하지 않는다]. 따라서, 일부 실시양태에서, gRNA (예를 들어, 표적 핵산과 혼성화되는 것)가 상기 이량체의 단량체 중 하나 또는 둘 다와 결합하면, 이러한 이량체가 표적 핵산에 공동-국재되어, 이량체가 뉴클레아제-활성 상태가 되도록 재포딩될 수 있고 표적 핵산을 절단할 수 있게 된다.

[0145] 일부 실시양태에서, 상기 방법은 특정 핵산을 다음 2개의 단백질을 포함하는 단백질 이량체와 접촉시키는 단계를 포함한다: (i) 뉴클레아제-불활성화 Cas9와 Cas9의 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반과 융합된 뉴클레아제-불활성화 Cas9)을 포함하는 융합 단백질, 및 (ii) Cas9의 DNA 절단 도메인 (예를 들어, Cas9 B-절반)을 포함하는 단백질. 다른 실시양태에서, 이량체는 (i) 뉴클레아제-불활성화 Cas9와 Cas9의 DNA 절단 도메인 (예를 들어, Cas9 B-절반과 융합된 뉴클레아제-불활성화 Cas9)을 포함하는 융합 단백질, 및 (ii) Cas9의 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반)을 포함하는 단백질을 포함한다. 일부 실시양태에서, 단백질 이량체는 하나 이상의 gRNA와 연합된다. 예를 들어, 일부 실시양태에서, 이량체는 2개의 gRNA와 연합되는데: 하나는 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인에 의해 결합되고; 다른 하나는 A-절반 도메인 (예를 들어, 융합 단백질의 A-절반이거나, 또는 융합 단백질의 일부가 아닌 이량체의 A-절반)에 의해 결합된다. 일부 실시양태에서, 단백질 이량체는 (i) 뉴클레아제-불활성화 Cas9와 Cas9의 gRNA 결합성 도메인 (예를 들어, Cas9 A-절반과 융합된 뉴클레아제-불활성화 Cas9)을 포함하는 융합 단백질, 및 (ii) 뉴클레아제-불활성화 Cas9와 DNA 절단 도메인 (예를 들어, Cas9 B-절반과 융합된 뉴클레아제-불활성화 Cas9)을 포함하는 융합 단백질을 포함한다. 일부 실시양태에서, 상기 이량체는 하나 이상의 별개의 gRNA와 연합된다. 예를 들어, 일부 실시양태에서, 상기 이량체는 2개 또는 3개의 gRNA와 연합된다. 일부 실시양태에서, 상기 이량체는 3개의 gRNA와 연합된다. 예를 들어, 핵산 표적의 특정 영역에 대한 하나의 뉴클레아제-불활성화 Cas9:gRNA의 결합, 및 핵산 표적의 제2 영역에 대한 다른 뉴클레아제-불활성화 Cas9:gRNA의 결합시, 분할 Cas9 절반들 (예를 들어, 융합 단백질의 A-절반과 B-절반)은 이량체화되고, 핵산 표적의 제3 영역에 대해 상보적인 제3의 gRNA와 결합되어, 완전한 활성의 Cas9 뉴클레아제로 되어 표적 DNA를 절단하게 된다.

[0146] 일부 실시양태에서, 핵산을 부위-특이적으로 절단하는 방법은 특정 핵산 (예를 들어, DNA)을, gRNA와 연합된 최소화 Cas9 단백질 (예를 들어, 본원에 기재된 바와 같음)과 접촉시키는 단계를 포함한다.

[0147] 일부 실시양태에서, 본원에 제공된 방법 중 어느 것도 특정 세포, 예를 들어 박테륨, 효모 세포, 또는 포유류 세포 내의 DNA 상에서 수행될 수 있다. 일부 실시양태에서, 본원에 제공된 어느 Cas9 단백질에 의해 접촉된

DNA는 진핵 세포 내에 있다. 일부 실시양태에서, 상기 방법은 시험관내 또는 생체 외에서의 세포 또는 조직 상에서 수행될 수 있다. 일부 실시양태에서, 진핵 세포는 특정 개체, 예컨대 환자 또는 연구용 동물 내에 있다. 일부 실시양태에서, 상기 개체는 인간이다.

[0148] 부위-특이적 재조합 방법

[0149] 본 개시내용의 또 다른 실시양태에서, 부위-특이적 핵산 (예를 들어, DNA) 재조합 방법이 제공된다. 일부 실시양태에서, 이러한 방법은 2개 이상의 핵산 (예를 들어, DNA) 분자의 2개 이상 영역의 재조합 또는 이러한 2개 이상 영역들 간의 재조합을 유도시키는 데 유용하다. 다른 실시양태에서, 상기 방법은 단일 핵산 분자 (예를 들어, DNA) 내의 2개 이상 영역의 재조합 또는 이러한 2개 이상 영역들 간의 재조합을 유도시키는 데 유용하다. 일부 실시양태에서, 하나 이상의 표적 핵산 분자를 재조합하기 위해서는, 표적 부위에서 사랑체성 복합체가 형성되어야 한다. 전형적으로, 이러한 사랑체는 4개의 본 발명의 RNA-가이드된 재조합효소 융합 단백질 (예를 들어, 본원에 제공된 모든 본 발명의 4개 재조합효소 융합 단백질의 복합체)을 포함한다. 일부 실시양태에서, 사랑체의 각각의 재조합효소 융합 단백질은 각각의 재조합효소 융합 단백질과 결합된 별개의 gRNA를 통하여 특별한 DNA 서열을 표적으로 한다 (예를 들어, 도 5 참조).

[0150] 일부 실시양태에서, 2개의 DNA 분자 간의 부위-특이적 재조합 방법은 (a) 제1 DNA를 제1 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 뉴클레아제-불활성화 Cas9 도메인은 제1 DNA의 특정 영역과 혼성화되는 제1 gRNA와 결합된다); (b) 제1 DNA를 제2 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제2 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 제1 DNA의 제2 영역과 혼성화되는 제2 gRNA와 결합된다); (c) 제2 DNA를 제3 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제3 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 제2 DNA의 특정 영역과 혼성화되는 제3 gRNA와 결합된다); 및 (d) 제2 DNA를 제4 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제4 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 제2 DNA의 제2 영역과 혼성화되는 제4 gRNA와 결합된다)를 포함한다. 단계 (a) 내지 (d)에서의 융합 단백질의 결합으로 인해, 융합 단백질의 재조합효소 촉매 도메인이 사랑체화되어 상기 DNA가 재조합된다. 일부 실시양태에서, 단계 (a) 및 (b)의 gRNA는 제1 DNA의 반대 가닥과 혼성화되고, 단계 (c) 및 (d)의 gRNA는 제2 DNA의 반대 가닥과 혼성화된다. 일부 실시양태에서, 단계 (a) 내지 (d)의 gRNA의 표적 부위는 재조합효소 촉매 도메인의 사랑체화를 허용하는 간격으로 있다. 예를 들어, 일부 실시양태에서, 단계 (a) 내지 (d)의 gRNA의 표적 부위는 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어져 있다. 일부 실시양태에서, 재조합되는 2개 DNA 분자의 2개 영역은, 이러한 재조합되는 영역이 80% 이상, 90% 이상, 95% 이상, 98% 이상, 또는 100% 상동성이 되도록 상동성을 공유한다.

[0151] 또 다른 실시양태에서, 단일 DNA 분자의 2개 영역 간의 부위-특이적 재조합 방법이 제공된다. 일부 실시양태에서, 이러한 방법은 (a) 특정 DNA를 제1 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 뉴클레아제-불활성화 Cas9 도메인은 상기 DNA의 특정 영역과 혼성화되는 제1 gRNA와 결합된다); (b) 상기 DNA를 제2 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제2 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 상기 DNA의 제2 영역과 혼성화되는 제2 gRNA와 결합된다); (c) 상기 DNA를 제3 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제3 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 상기 DNA의 제3 영역과 혼성화되는 제3 gRNA와 결합된다); 및 (d) 상기 DNA를 제4 RNA-가이드된 재조합효소 융합 단백질과 접촉시키는 단계 (여기서, 제4 융합 단백질의 뉴클레아제-불활성화 Cas9 도메인은 상기 DNA의 제4 영역과 혼성화되는 제4 gRNA와 결합된다)를 포함한다. 단계 (a) 내지 (d)에서의 융합 단백질의 결합으로 인해, 융합 단백질의 재조합효소 촉매 도메인이 사랑체화되어 상기 DNA가 재조합된다. 일부 실시양태에서, 단계 (a) 내지 (d)의 gRNA 중 2개는 DNA의 동일 가닥과 혼성화되고, 단계 (a) 내지 (d)의 다른 2개의 gRNA는 DNA의 반대 가닥과 혼성화된다. 일부 실시양태에서, 단계 (a) 및 (b)의 gRNA는 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어진 DNA의 영역과 혼성화되고, 단계 (c) 및 (d)의 gRNA는 10개 이하, 15개 이하, 20개 이하, 25개 이하, 30개 이하, 40개 이하, 50개 이하, 60개 이하, 70개 이하, 80개 이하, 90개 이하, 또는 100개 이하 염기 쌍 떨어진 DNA의 영역과 혼성화된다. 일부 실시양태에서, 재조합되는 DNA 분자의 2개 영역은, 이러한 재조합되는 영역이 80% 이상, 90% 이상, 95% 이상, 98% 이상, 또는 100% 상동성이 되도록 상동성을 공유한다.

[0152] 일부 실시양태에서, 본 발명의 부위-특이적 재조합 방법 중 어느 것도 재조합을 유도시킬 수 있으므로, 이러한 재조합으로 인해 절제 (예를 들어, DNA의 절편이 표적 DNA 분자로부터 절제된다), 삽입 (예를 들어, DNA의 절편이 표적 DNA 분자 내로 삽입된다), 역위 (예를 들어, DNA의 절편이 표적 DNA 분자 내에서 역위된다), 또는 전위

(예를 들어, 하나 이상의 표적 DNA 분자(들) 간의 DNA 절편의 교환)가 발생된다. 일부 실시양태에서, 특별한 재조합 현상 (예를 들어, 절제, 삽입, 역위, 전위 등)은 특히, 결합된 RNA-가이드된 재조합효소 융합 단백질(들)의 배향 (예를 들어, 표적 DNA 분자(들)를 기준으로 함)에 좌우된다. 일부 실시양태에서, RNA-가이드된 재조합효소 융합 단백질이 표적 핵산과 결합되는 배향 또는 방향은, 예를 들어 RNA-가이드된 재조합효소 융합 단백질(들)과 결합된 gRNA의 특별한 서열에 의해 제어될 수 있다. 특별한 재조합 현상을 제어 또는 지시하는 방법은 관련 기술분야에 공지되어 있고, 예를 들어 문헌 [Turan and Bode, "Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications." *FASEB J.* 2011; Dec;25(12):4088-107; 그의 전문이 본원에 참조로 포함된다]에 기재된 것을 포함한다.

[0153] 일부 실시양태에서, 부위-특이적 재조합 방법 중 어느 것도 생체내 또는 시험관 내에서 수행될 수 있다. 일부 실시양태에서, 부위-특이적 재조합 방법 중 어느 것도 특정 세포 내에서 수행된다 (예를 들어, 세포 내의 게놈 DNA와 재조합된다). 이러한 세포는 원핵성 또는 진핵성일 수 있다. 세포, 예컨대 진핵 세포는 본원에 기재된 바와 같은 특정 개체, 예컨대 대상체 (예를 들어, 인간 대상체) 내에 있을 수 있다. 본원에 기재된 방법은 시험관내 및 생체 내에서, 예를 들어 트랜스제닉 세포, 세포주 또는 동물의 발생의 맥락에서, 또는 게놈 서열의 변경, 예를 들어 유전적 결합의 교정에서, 특정 대상체 내에 있거나 이러한 대상체로부터 획득된 세포에서, 세포의 유전적 변형에 유용하다. 일부 실시양태에서, 대상체로부터 획득되고 본원에 제공된 방법에 따라서 변형된 세포를 대상체 (예를 들어, 동일 대상체) 내로 재도입하여, 예를 들어 특정 질환을 치료하거나 또는 농업 또는 생물학적 연구에서 유전적으로 변형된 유기체를 생성시킨다.

[0154] 핵산 서열이 표적 핵산 내로 삽입되도록 2개 이상의 핵산을 재조합하는 것이 요망되는 적용에서는, 삽입될 공여자 서열을 포함하는 핵산이 또한, 예를 들어 세포에 제공된다. "공여자 서열"이란, 하나 이상의 RNA-가이드된 재조합효소 융합 단백질(들)에 의해 유도된 표적 부위에 삽입될 핵산 서열을 의미한다. 일부 실시양태에서, 예를 들어 게놈 변형의 맥락에서, 공여자 서열은 표적 부위에서의 게놈 서열과 상동성을 공유할 것인데, 예를 들어 표적 부위 양옆에 위치하고 있는 뉴클레오타이드 서열, 예를 들어 표적 부위의 약 100개 이하 염기 이내, 예를 들어 약 90개 염기 이내, 약 80개 염기 이내, 약 70개 염기 이내, 약 60개 염기 이내, 약 50개 염기 이내, 약 40개 염기 이내, 약 30개 염기 이내, 약 15개 염기 이내, 약 10개 염기 이내, 약 5개 염기 이내, 또는 표적 부위 바로 양옆에 위치하고 있는 뉴클레오타이드 서열과 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 또는 100% 상동성을 공유할 것이다. 일부 실시양태에서, 공여자 서열은 표적 핵산과 어떠한 상동성도 공유하지 않는데, 예를 들어 표적 부위에서의 게놈 서열과 상동성을 공유하지 않는다. 공여자 서열은 어떠한 길이일 수도 있는데, 예를 들어 10개 이상의 뉴클레오타이드, 50개 이상의 뉴클레오타이드, 100개 이상의 뉴클레오타이드, 250개 이상의 뉴클레오타이드, 500개 이상의 뉴클레오타이드, 1,000개 이상의 뉴클레오타이드, 5,000개 이상의 뉴클레오타이드, 10,000개 이상의 뉴클레오타이드, 100,000개 이상의 뉴클레오타이드 등일 수 있다.

[0155] 전형적으로, 공여자 서열은 이것을 대체되거나 또는 이것이 삽입되는 표적 서열과 동일하지 않다. 일부 실시양태에서, 공여자 서열은 표적 서열 (예를 들어, 표적 게놈 서열)을 기준으로 하여 적어도 하나 이상의 단일 염기 변화, 삽입, 결실, 역위 또는 재배열을 함유한다. 일부 실시양태에서, 공여자 서열은 또한, 관심 DNA 영역과 상동성이 아니고 관심 DNA 영역 내로 삽입되는 것으로 의도되지 않은 서열을 함유하는 벡터 골격을 포함한다.

[0156] 공여자 서열은 표적 (예를 들어, 게놈) 서열과 비교해서 특정의 서열 상이성, 예를 들어 제한 부위, 뉴클레오타이드 다형성, 선별성 마커 (예를 들어, 약물 내성 유전자, 형광성 단백질, 효소 등)를 포함할 수 있는데, 이들은 표적 부위에서의 공여자 서열의 성공적인 삽입을 평가하기 위해 사용될 수 있거나, 또는 일부 경우에는, 다른 목적을 위해 (예를 들어, 표적화된 게놈 유전자 자리에서의 발현을 나타내기 위해) 사용될 수 있다. 일부 실시양태에서, 코딩 영역 내에 위치되는 경우에는, 이러한 뉴클레오타이드 서열 상이성이 아미노산 서열을 변화시키지 않을 것이거나, 또는 침묵 아미노산 변화 (예를 들어, 단백질의 구조 또는 기능에 영향을 미치지 않는 변화)를 만들 것이다. 일부 실시양태에서, 이들 서열 상이성은 양옆에 위치하고 있는 재조합 서열, 예컨대 FLP, loxP 서열 등을 포함할 수 있는데, 이들은 예를 들어, 마커 서열을 제거하기 위하여 나중에 활성화될 수 있다.

[0157] 공여자 서열은 단일 가닥 DNA, 단일 가닥 RNA, 이중 가닥 DNA, 또는 이중 가닥 RNA로서 세포에 제공될 수 있다. 이는 선형 또는 환상 형태로 세포 내로 도입될 수 있다. 선형 형태로 도입되는 경우에는, 공여자 서열의 말단이 관련 기술분야의 통상의 기술자에게 공지된 방법에 의해 보호될 수 있다 (예를 들어, 외부핵산분해적 분해로부터 보호될 수 있다). 예를 들어, 하나 이상의 디옥시뉴클레오타이드 잔기를 선형 분자의 3' 말단에 부가하고/하거나 자기 상보적 올리고뉴클레오타이드를 한쪽 말단 또는 양 말단에 결합시킨다 (예를 들어, 문헌 [Chang et al., *Proc. Natl. Acad. Sci. USA.* 1987; 84:4959-4963]; [Nehls et al., *Science.* 1996; 272:886-889] 참조). 일부 실시양태에서, 공여자 서열은 부가 서열, 예컨대, 예를 들어 복제 기점, 프로모터, 및 항생제 내성을 코딩

하는 유전자를 갖는 벡터 분자의 일부로서 세포 내로 도입될 수 있다. 일부 실시양태에서, 공여자 서열은 네이키드(naked) 핵산으로서, 리포솜 또는 폴옥사머와 같은 작용제와 복합체를 형성한 핵산으로서 도입될 수 있거나, 또는 바이러스 (예를 들어, 아데노바이러스, AAV 등)에 의해 전달될 수 있다.

[0158] 폴리뉴클레오티드, 벡터, 세포, 키트

[0159] 본 개시내용의 또 다른 실시양태에서, 하나 이상의 본 발명의 단백질 및/또는 gRNA를 코딩하는 폴리뉴클레오티드가 제공된다. 예를 들어, Cas9 변이체를 포함하는, 단리된 뉴클레아제 및 재조합효소의 재조합 발현 및 정제를 위한, 예를 들어 본원에 기재된 단백질 중 어느 것을 코딩하는 폴리뉴클레오티드가 제공된다. 일부 실시양태에서, 단리된 폴리뉴클레오티드는 Cas9 절반 부위 (예를 들어, A-절반 및/또는 B-절반)를 코딩하는 하나 이상의 서열을 포함한다. 일부 실시양태에서, 단리된 폴리뉴클레오티드는 Cas9 융합 단백질, 예를 들어 본원에 기재된 Cas9 융합 단백질 중 어느 것 (예를 들어, 뉴클레아제-불활성화 Cas9를 포함하는 것)을 코딩하는 하나 이상의 서열을 포함한다. 일부 실시양태에서, 단리된 폴리뉴클레오티드는 gRNA를 코딩하는 하나 이상의 서열만을 포함하거나, 또는 이를 본원에 기재된 단백질 중 어느 것을 코딩하는 서열과 조합하여 포함한다.

[0160] 일부 실시양태에서, 예를 들어 Cas9 단백질, 및/또는 Cas9 단백질을 포함하는 융합물 (예를 들어, 변이체)의 재조합 발현 및 정제를 위한, 본원에 기재된 단백질 중 어느 것을 코딩하는 벡터가 제공된다. 일부 실시양태에서, 이러한 벡터는, 단리된 폴리뉴클레오티드, 예를 들어 본원에 기재된 것을 포함하거나, 또는 이를 포함하도록 조작된다. 일부 실시양태에서, 상기 벡터는 Cas9 단백질 (본원에 기재된 바와 같음), gRNA, 또는 그의 조합물 (본원에 기재된 바와 같음)을 코딩하는 하나 이상의 서열을 포함한다. 전형적으로, 벡터는 융합 단백질이 숙주 세포에서 발현되도록, 프로모터와 작동적으로 연결된 본 발명의 단백질을 코딩하는 서열을 포함한다.

[0161] 일부 실시양태에서, 예를 들어 본원에 제공된 Cas9 단백질 중 어느 것의 재조합 발현 및 정제를 위한 세포가 제공된다. 이러한 세포는 재조합 단백질 발현에 적합한 모든 세포, 예를 들어 본 발명의 단백질을 발현하거나 또는 발현할 수 있는 유전적 구조물을 포함하는 세포 (예를 들어, 본원에 기재된 하나 이상의 벡터로 형질전환시킨 세포, 또는 게놈 변형을 갖는 세포, 예를 들어 세포의 게놈에 혼입시킨 특정 대립 유전자로부터 본원에 제공된 단백질을 발현하는 것)를 포함한다. 세포를 형질전환시키는 방법, 세포를 유전적으로 변형시키는 방법, 및 이러한 세포에서 유전자 및 단백질을 발현하는 방법은 관련 기술분야에 널리 공지되어 있고, 예를 들어 문헌 [Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012))] 및 [Friedman and Rossi, *Gene Transfer: Delivery and Expression of DNA and RNA, A Laboratory Manual* (1st ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2006))]에 제공된 것을 포함한다.

[0162] 본 개시내용의 일부 측면은 본원에 제공된 바와 같은, Cas9 변이체 및/또는 뉴클레아제 및/또는 재조합효소를 포함하는 키트를 제공한다. 일부 실시양태에서, 이러한 키트는, 예를 들어 본원에 제공된 바와 같은, 본 발명의 Cas9 변이체, 뉴클레아제, 및/또는 재조합효소를 코딩하는 폴리뉴클레오티드를 포함한다. 일부 실시양태에서, 키트는 재조합 단백질 발현을 위한 벡터를 포함하는데, 여기서 이러한 벡터는 본원에 제공된 단백질 중 어느 것을 코딩하는 폴리뉴클레오티드를 포함한다. 일부 실시양태에서, 키트는 본원에 제공된 단백질 중 어느 것을 발현하기 위한 유전적 구조물을 포함하는 세포 (예를 들어, Cas9 단백질, 또는 Cas9 단백질을 포함하는 융합물을 발현하는 데 적합한 모든 세포, 예컨대 박테리아, 효모 또는 포유류 세포)를 포함한다. 일부 실시양태에서, 본원에 제공된 키트 중 어느 것도 하나 이상의 gRNA 및/또는 하나 이상의 gRNA를 발현하기 위한 벡터를 추가로 포함한다. 일부 실시양태에서, 상기 키트는 부형제; 및 뉴클레아제 및/또는 재조합효소를 이러한 부형제와 접촉시켜, 핵산을 뉴클레아제 및/또는 재조합효소와 접촉시키는 데 적합한 조성물을 생성시킴으로써, 표적 핵산과의 혼성화 및 표적 핵산의 절단 및/또는 재조합이 일어나도록 하는 것에 관한 설명서를 포함한다. 일부 실시양태에서, 상기 조성물은 Cas9 단백질을 특정 세포에 전달하는 데 적합하다. 일부 실시양태에서, 상기 조성물은 Cas9 단백질을 특정 대상체에게 전달하는 데 적합하다. 일부 실시양태에서, 부형제는 제약상 허용되는 부형제이다.

[0163] 본 발명의 이들 실시양태 및 기타 실시양태의 기능과 이점은 다음 실시예로부터 보다 상세히 이해될 것이다. 다음 실시예는 본 발명의 이득을 예시하고 특별한 실시양태를 설명하고자 한 것이지만, 본 발명의 전체 범위를 예시하고자 한 것은 아니다. 따라서, 다음 실시예는 본 발명의 범위를 제한하지 않는 것으로 이해될 것이다.

[0164] 실시예

[0165] 실시예 1: 불활성화 Cas9를 FokI 뉴클레아제와 융합시키면 게놈 변형 특이성이 개선된다

[0166] 방법:

[0167] 올리고뉴클레오타이드 및 PCR

[0168] 모든 올리고뉴클레오타이드는 인티그레이티드 DNA 테크놀로지(Integrated DNA Technologies; IDT)로부터 구입하였다. 올리고뉴클레오타이드 서열은 표 1에 열거되어 있다. 달리 언급되지 않는 한은, 1x HF 완충제를 수반한 50 μ l 중의 2 U/ μ l 푸션 핫 스타트 플렉스(Phusion Hot Start Flex) DNA 중합효소 (NEB) 0.4 μ l, 0.2 mM dNTP 혼합물 (0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.2 mM dTTP) (NEB), 0.5 μ M의 각 프라이머 및 다음 프로그램을 이용하여 PCR을 수행하였다: 98 $^{\circ}$ C, 1 min; [98 $^{\circ}$ C, 15 s; 65 $^{\circ}$ C, 15 s; 72 $^{\circ}$ C, 30 s]의 35 주기.

[0169] FokI-dCas9, Cas9 니카제 및 gRNA 발현 플라스미드의 구축

[0170] NLS 및 3xFLAG 태그를 수반한 인간 코돈-최적화 스트렙토코쿠스 피오게네스 Cas9 뉴클레아제 [애드젠(Addgene) 플라스미드 43861]²를 야생형 Cas9 발현 플라스미드로서 사용하였다. 다음 표 1에 열거된 Cas9_Exp 프라이머를 이용하여 주형으로서의 야생형 Cas9 발현 플라스미드의 PCR (72 $^{\circ}$ C, 3 min) 생성물을 집슨 어셈블리 클로닝 키트 (Gibson Assembly Cloning Kit) [뉴 잉글랜드 바이오랩스(New England Biolabs)]와 어셈블리하여 Cas9 및 FokI-dCas9 변이체를 구축하였다. 단일 gRNA 구조물 (gRNA G1 내지 G13)을 코딩하는 발현 플라스미드를 기존에 언급된 바와 같이 클로닝하였다. 간략하게 언급하면, 20-bp 프로토스페이서 표적 서열을 함유하는 표 1에 열거된 gRNA 올리고뉴클레오타이드를 어닐링하고, 이로써 생성되는 4-bp 오버행을 BsmBI-분해된 gRNA 발현 플라스미드 내로 결찰시켰다. 단일 플라스미드 상의 별도의 프로모터로부터의 2개의 별도의 gRNA 구조물의 발현을 코딩하는 gRNA 발현 플라스미드를 2-단계 과정으로 클로닝하였다. 첫째, 하나의 gRNA (gRNA E1, V1, C1, C3, H1, G1, G2 또는 G3)를 상기와 같이 클로닝하고, PCR_Pla-fwd 및 PCR_Pla-rev 프라이머를 이용하여 PCR (72 $^{\circ}$ C, 3 min)하기 위한 주형으로서 사용하며, 1 μ l DpnI (NEB)를 가하고, 반응물을 37 $^{\circ}$ C 하에 30분 동안 인큐베이션한 다음, "제1 gRNA + 벡터 DNA"에 대하여 퀴아퀵(QIAquick) PCR 정제 키트 [퀴아젠(Qiagen)]을 수행하였다. 표 1에 열거된 PCR_gRNA-fwd1, PCR_gRNA-rev1, PCR_gRNA-rev2 및 적당한 PCR_gRNA 프라이머를 이용하여 주형으로서의 100 pg의 BsmBI-분해된 gRNA 발현 플라스미드의 PCR (72 $^{\circ}$ C, 3 min) 생성물을 DpnI 처리하고, "제2 gRNA 삽입 DNA"에 대하여 상기와 같이 정제하였다. 약 200 ng의 "제1 gRNA + 벡터 DNA" 및 약 200 ng의 "제2 gRNA 삽입 DNA"를 실온 하에 (약 21 $^{\circ}$ C) 15분 동안 20 μ l의 총 용적 중에서 1x T4 DNA 리가제 완충제, 1 μ l의 T4 DNA 리가제 (400 U/ μ l, NEB)에서 평활 말단 결찰시켰다. 모든 클로닝을 위하여, 1 μ l의 결찰 또는 어셈블리 반응물을 Mach1 화학적으로 적격한 세포 [라이프 테크놀로지(Life Technologies)] 내로 형질전환시켰다.

[0171] <표 1>

[0172] 올리고뉴클레오타이드. '/5Phos/'는 5' 인산화 올리고뉴클레오타이드를 표시한다

dCas9-NLS-FokI 프라이머:	
Cas9_Exp_CNF_FokI+Plas-Fwd	CGGCGAGATAAACTTTTAA TGACCGGTCATCATCACCA (서열 26)
Cas9_Exp_CNF_Cas9coD10-Rev	CCAACGGAATTAGTGCCGATAGCTAAACCAATAGAATACTTTTATC (서열 27)
Cas9_Exp_CNF_Cas9coD10-Fwd	GATAAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGG (서열 28)
Cas9_Exp_CNF_Cas9coH850-Rev	TTCAAAAAGGATTGGGGTACAATGGCATCGACGTCGTAATCAGATAAAC (서열 29)
Cas9_Exp_CNF_Cas9coH850-Fwd	GTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGA A (서열 30)
Cas9_Exp_CNF_(Cas9)NLS+GGS-Fok-Rev	TTGGGATCCAGAACCTCCTCCTGCAGCCTTGTCATCG (서열 31)
Cas9_Exp_CNF_(Cas9)NLS+GGS3-Fok-Rev	TTGGGATCCAGAACCTCC GCTGCCGCCACTTCCACCTGA TCCTGCAGCCTTGTCATCG (서열 32)
Cas9_Exp_CNF_(Cas9)NLS+GGS-Fok-Fwd	CGATGACAAGGCTGCAGGAGGAGGTTCTGGATCCCAA (서열 33)
Cas9_Exp_CNF_(Cas9)NLS+GGS3-Fok-Fwd	CGATGACAAGGCTGCAGGA TCAGGTGGAAGTGGCGGCAGC GGAGGTTCTGGATCCCAA (서열 34)
Cas9_Exp_CNF_FokI+Plas-Rev	TGGTGATGATGACCGGTCA TAAAAGTTATCTCGCCG (서열 35)
NLS-dCas9-FokI 프라이머:	
Cas9_Exp_NCF_FokI+Plas-Fwd	CGGCGAGATAAACTTTTAA TGACCGGTCATCATCACCA (서열 36)
Cas9_Exp_NCF_PlasS+FLAG(NLS-FokI-Rev	TAGGGAGAGCCGCCACCATGGACTACAAAGACCATGACGG (서열 37)
Cas9_Exp_NCF_NLS+Cas9coD10-Rev	TAAACCAATAGAATACTTTTATC CATAGGTACCCCGCGGTGAATG (서열 38)
Cas9_Exp_NCF_Cas9coD10-Fwd	GATAAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGG (서열 39)
Cas9_Exp_NCF_Cas9coH850-Rev	TTCAAAAAGGATTGGGGTACAATGGCATCGACGTCGTAATCAGATAAAC (서열 40)
Cas9_Exp_NCF_Cas9coH850-Fwd	GTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGA A (서열 41)
Cas9_Exp_NCF_Cas9End+GGS-Fok-Rev	TTGGGATCCAGAACCTCCGTCACCCCAAGCTGTG (서열 42)
Cas9_Exp_NCF_Cas9End+GGS3-Fok-Rev	TTGGGATCCAGAACCTCC GCTGCCGCCACTTCCACCTGA GTCACCCCAAGCTGTG (서열 43)
Cas9_Exp_NCF_Cas9End+GGS-Fok-Fwd	CACAGCTTGGGGGTGACGGAGGTTCTGGATCCCAA (서열 44)
Cas9_Exp_NCF_Cas9End+GGS3-Fok-Fwd	CACAGCTTGGGGGTGAC TCAGGTGGAAGTGGCGGCAGC GGAGGTTCTGGATCCCAA (서열 45)
Cas9_Exp_NCF_FokI+Plas-Rev	TGGTGATGATGACCGGTCA TAAAAGTTATCTCGCCG (서열 46)
FokI-dCas9-NLS 프라이머:	

[0173]

Cas9_Exp_FCN_PlacS+Fok-Fwd	TAGGGAGAGCCGCCACCATGGGATCCCAACTAGTCAAAAAG (서열 47)
Cas9_Exp_FCN_Fok1GGS+Cas-Rev	ACCAATAGAATACTTTTTATCCATGCTGCCACCAAAGTTATCTC (서열 48)
Cas9_Exp_FCN_Fok1GGS3+Cas-Rev	ACCAATAGAATACTTTTTATCCATGCTGCCGCCACTTCCACCTG (서열 49)
Cas9_Exp_FCN_Cas9coD10-Fwd	GATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGG (서열 50)
Cas9_Exp_FCN_Cas9coH850-Rev	CCAACGGAATTAGTGCCGATAGCTAAACCAATAGAATACTTTTTATC (서열 51)
Cas9_Exp_FCN_Cas9coH850-Fwd	GTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTGA A (서열 52)
Cas9_Exp_FCN_Cas9End+PlasmidEn-Rev	TGGTGATGATGACCGGTCA GTCACCCCAAGCTGTG (서열 53)
Cas9_Exp_FCN_Cas9End+PlasmidEn-Fwd	CACAGCTTGGGGGTGAC TGACCGGTCATCATCACCA (서열 54)
Cas9_Exp_FCN_PlacS+Fok-Rev	CTTTTGACTAGTTGGGATCCCATGGTGGCGGCTCTCCCTA (서열 55)
gRNA_G1-상단	ACACCCCTCGAACTTCACCTCGGCGG (서열 56)
gRNA_G2-상단	ACACCGTCGCCCTCGAACTTCACCTG (서열 57)
gRNA_G3-상단	ACACCCAGCTCGATGCGGTTACACAG (서열 58)
gRNA_G4-상단	ACACCGGTGAACCGCATCGAGCTGAG (서열 59)
gRNA_G5-상단	ACACCGCTGAAGGGCATCGACTTCAG (서열 60)
gRNA_G6-상단	ACACCGGCATCGACTTCAAGGAGGAG (서열 61)
gRNA_G7-상단	ACACCAAGGAGGACGGCAACATCCG (서열 62)
gRNA_G8-상단	ACACCACCATCTTCTTCAAGGACGAG (서열 63)
gRNA_G9-상단	ACACCAACTACAAGACCCGCGCCGG (서열 64)
gRNA_G10-상단	ACACCCGCGCCGAGGTGAAGTTCGG (서열 65)
gRNA_G11-상단	ACACCGAAGTTCGAGGGCGACACCCG (서열 66)
gRNA_G12-상단	ACACCTTCGAACCTCACCTCGGCGCG (서열 67)
gRNA_G13-상단	ACACCTCAGCTCGATGCGGTTACCCG (서열 68)
gRNA_G14-상단	ACACCCGATGCCCTTCAGCTCGATGG (서열 69)
gRNA_G1-하단	AAAACCGCCGAGGTGAAGTTCGAGGG (서열 70)
gRNA_G2-하단	AAAACAGGTGAAGTTCGAGGGCGACG (서열 71)
gRNA_G3-하단	AAAACCTGGTGAACCGCATCGAGCTGG (서열 72)
gRNA_G4-하단	AAAACCTCAGCTCGATGCGGTTACCCG (서열 73)
gRNA_G5-하단	AAAACCTGAAGTCGATGCCCTTCAGCG (서열 74)
gRNA_G6-하단	AAAACCTCCTCTGAAGTCGATGCCG (서열 75)
gRNA_G7-하단	AAAACGGATGTTGCCGTCCTCCTTGG (서열 76)
gRNA_G8-하단	AAAACCTCGTCTTGAAGAAGATGGTG (서열 77)
gRNA_G9-하단	AAAACCGGCGCGGGTCTTGTAGTTGG (서열 78)
gRNA_G10-하단	AAAACCGAACTTCACCTCGGCGCGGG (서열 79)
gRNA_G11-하단	AAAACGGGTGTCGCCCTCGAACTTCG (서열 80)
gRNA_G12-하단	AAAACGCGCCGAGGTGAAGTTCGAAG (서열 81)
gRNA_G13-하단	AAAACGGTGAACCGCATCGAGCTGAG (서열 82)
gRNA_G14-하단	AAAACCATCGAGCTGAAGGCATCGG (서열 83)
gRNA_C1-상단	ACACCTGGCCTGCTTGCTAGACTTGG (서열 84)

[0174]

gRNA_C3-상단	ACACCGCAGATGTAGTGTTCACAG (서열 85)
gRNA_H1-상단	ACACCCTTGCCCCACAGGGCAGTAAG (서열 86)
gRNA_E1-상단	ACACCGAGTCCGAGCAGAAGAAG (서열 87)
gRNA_V1-상단	ACACCGGTGGGGGAGTTTGCTCCG (서열 88)
gRNA_C1-하단	AAAACCAAGTCTAGCAAGCAGGCCAG (서열 89)
gRNA_C3-하단	AAAACCTGTGAAACACTACATCTGCG (서열 90)
gRNA_H1-하단	AAAACCTCTTCTTCTGCTCGGACTCG (서열 91)
gRNA_E1-하단	AAAACCTACTGCCCTGTGGGGCAAGG (서열 92)
gRNA_V1-하단	AAAACGGAGCAAACCCCCACCCG (서열 93)
PCR_Pla-fwd	AGG AAA GAA CAT GTG AGC AAA AG (서열 94)
PCR_Pla-rev	CAGCGAGTCAGTGAGCGA (서열 95)
PCR_gRNA-fwd1	CTGTACAAAAAGCAGGCTTA (서열 96)
PCR_gRNA-rev1	AACGTAGGTCTCTACCGCTGTACAAAAAGCAGGCTTA (서열 97)
PCR_gRNA-rev2	AAAAAAGCACCGACTCGGTGCCACTTTTCAAGTTGATAACGGACT AGCCTTATTTAACTTGCTATTCTAGCTCTAAAC (서열 98)
PCR_gRNA_G1	TTGCTATTTCTAGCTCTAAACCGCCGAGGTGAAGTTCGAGGGGTGT TTCGTCCTTTCCA (서열 99)
PCR_gRNA_G2	TTGCTATTTCTAGCTCTAAACAGGTGAAGTTCGAGGGCGACGGTGT TTCGTCCTTTCCA (서열 100)
PCR_gRNA_G3	TTGCTATTTCTAGCTCTAAACCTGGTGAACCGCATCGAGCTGGGTGT TTCGTCCTTTCCA (서열 101)
PCR_gRNA_G4	TTGCTATTTCTAGCTCTAAACCTCAGCTCGATGCGGTTCACCGGTGT TCGTCCTTTCCA (서열 102)
PCR_gRNA_G5	TTGCTATTTCTAGCTCTAAACCTGAAGTCGATGCCCTTCAGCGGTGT TCGTCCTTTCCA (서열 103)
PCR_gRNA_G6	TTGCTATTTCTAGCTCTAAACCTCCTCCTGAAGTCGATGCCGTGT TCGTCCTTTCCA (서열 104)
PCR_gRNA_G7	TTGCTATTTCTAGCTCTAAACCGATGTTGCCGTCCTTGGGTGT TCGTCCTTTCCA (서열 105)
PCR_gRNA_C2	TTGCTATTTCTAGCTCTAAACGCTTGAGGGAGATGAGGACTGGTGT TTCGTCCTTTCCA (서열 106)
PCR_gRNA_C4	TTGCTATTTCTAGCTCTAAACATGACTGTGAAGAGCTTCACGGTGT TTCGTCCTTTCCA (서열 107)
PCR_gRNA_E2	TTGCTATTTCTAGCTCTAAACGAGGACAAAGTACAAACGCGGTGT TTCGTCCTTTCCA (서열 108)
PCR_gRNA_E3	TTGCTATTTCTAGCTCTAAACGAACCGGAGGACAAAGTACAGGTGT TTCGTCCTTTCCA (서열 109)
PCR_gRNA_H2	TTGCTATTTCTAGCTCTAAACACCACCAACTTCATCCACGGGTGT TCGTCCTTTCCA (서열 110)
PCR_gRNA_H3	TTGCTATTTCTAGCTCTAAACGGGCCTCACCACCAACTTCAGGTGT TTCGTCCTTTCCA (서열 111)
PCR_gRNA_H4	TTGCTATTTCTAGCTCTAAACGCCAGGGCCTCACCACCAAGGTGT TTCGTCCTTTCCA (서열 112)
PCR_gRNA_H5	TTGCTATTTCTAGCTCTAAACACCTGCCAGGGCCTCACCAGGTGT TTCGTCCTTTCCA (서열 113)
PCR_gRNA_H6	TTGCTATTTCTAGCTCTAAACCTGATACCAACCTGCCAGGGGTGT TTCGTCCTTTCCA (서열 114)
PCR_gRNA_H7	TTGCTATTTCTAGCTCTAAACCTGTCTTGTAACCTTGGTGT TCGTCCTTTCCA (서열 115)
PCR_gRNA_V2	TTGCTATTTCTAGCTCTAAACGCTCTGGCTAAAGAGGGAATGGTGT TTCGTCCTTTCCA (서열 116)

[0175]

PCR_gRNA_V3	TTGCTATTTCTAGCTCTAAAACCGGCTCTGGCTAAAGAGGGAGGTGT TTCGTCCTTTCCA (서열 117)
PCR_gRNA_V4	TTGCTATTTCTAGCTCTAAAACCTGACACCCCCGGCTCTGGGGTGT TCGTCCTTTCCA (서열 118)
서배이_GFP-fwd	TACGGCAAGCTGACCCTGAA (서열 119)
서배이_GFP-rev	GTCCATGCCGAGAGTGATCC (서열 120)
서배이_CLTA-fwd	GCCAGGGGCTGTATCTTGG (서열 121)
서배이_CLTA-rev	ATGCACAGAAGCACAGGTGA (서열 122)
서배이_EMX-fwd	CTGTGTCCTCTTCTGCCCT (서열 123)
서배이_EMX-rev	CTCTCCGAGGAGAAGGCCAA (서열 124)
서배이_HBB-fwd	GGTAGACCACCAGCAGCCTA (서열 125)
서배이_HBB-rev	CAGTGCCAGAAGAGCCAAGG (서열 126)
서배이_VEGF-fwd	CCACACAGCTTCCCCTTCTC (서열 127)
서배이_VEGF-rev	GAGAGCCGTTCCCTCTTTC (서열 128)
HTS_EXM_ON-fwd	CCTCCCCATTGGCCTGCTTC (서열 129)
HTS_EXM_Off1-fwd	TCGTCCTGCTCTCACTTAGAC (서열 130)
HTS_EXM_Off2-fwd	TTTGTGGCTTGGCCCCAGT (서열 131)
HTS_EXM_Off3-fwd	TGCAGTCTCATGACTTGGCCT (서열 132)
HTS_EXM_Off4-fwd	TTCTGAGGGCTGCTACCTGT (서열 133)
HTS_VEGF_ON-fwd	ACATGAAGCAACTCCAGTCCA (서열 134)
HTS_EXM_Off1-fwd	AGCAGACCCACTGAGTCAACTG (서열 135)
HTS_EXM_Off2-fwd	CCCGCCACAGTCGTGTCAT (서열 136)
HTS_EXM_Off3-fwd	CGCCCCGGTACAAGGTGA (서열 137)
HTS_EXM_Off4-fwd	GTACCGTACATTGTAGGATGTTT (서열 138)
HTS_CLTA2_ON-fwd	CCTCATCTCCTCAAGCAGGC (서열 139)
HTS_CLTA2_Off1-fwd	ATTCTGCTCTTGAGGTATTTGT (서열 140)
HTS_CLTA2_Off2-fwd	CACCTCTGCCTCAAGAGCAGAAAA (서열 141)
HTS_CLTA2_Off3-fwd	TGTGTGTGTGTGTGTAGGACT (서열 142)
HTS_EXM_ON-rev	TCATCTGTGCCCCCTCCCTCC (서열 143)
HTS_EXM_Off-rev	CGAGAAGGAGGTGCAGGAG (서열 144)
HTS_EXM_Off-rev	CGGGAGCTGTTCAGAGGCTG (서열 145)
HTS_EXM_Off-rev	CTCACCTGGGCGAGAAAGGT (서열 146)
HTS_EXM_Off-rev	AAAACCTCAAAGAAATGCCCAATCA (서열 147)
HTS_VEGF_ON-rev	AGACGCTGCTCGCTCCATTC (서열 148)
HTS_EXM_Off1-rev	ACAGGCATGAATCACTGCACCT (서열 149)
HTS_EXM_Off2-rev	GCGGCAACTTCAGACAACCGA (서열 150)
HTS_EXM_Off3-rev	GACCCAGGGGCACCAGTT (서열 151)
HTS_EXM_Off4-rev	CTGCCTTCATTGCTTAAAGTGGAT (서열 152)
HTS_CLTA2_ON-rev	ACAGTTGAAGGAAGGAAACATGC (서열 153)
HTS_CLTA2_Off1-rev	GCTGCATTTGCCATTCCA (서열 154)
HTS_CLTA2_Off2-rev	GTTGGGGGAGGAGGAGCTTAT (서열 155)
HTS_CLTA2_Off3-rev	CTAAGAGCTATAAGGGCAAATGACT (서열 156)

[0176]

[0177]

게놈 GFP의 변형

[0178]

HEK293-GFP 안정적 세포 [젠타겟(GenTarget)]를, 게놈 상에 통합된 에머랄드(Emerald) GFP 유전자 (GFP)를 구성적으로 발현하는 세포주로서 사용하였다. 세포를, 10% (vol/vol) 태내 소 혈청 (FBS, 라이프 테크놀로지) 및 페니실린/스트렙토마이신 [1x, 암레스코(Amresco)]을 보충시킨 둘베크(Dulbecco) 변형 이글 배지 (DMEM, 라이프 테크놀로지)에 유지시켰다. 5×10^4 개의 HEK293-GFP 세포를, 48-웰 콜라겐 코팅된 바이오코트 판 [벡톤 디킨슨 (Becton Dickinson)] 상에 도말하였다. 도말한지 1일 후에, 약 75% 전면생장률을 나타내는 세포를 제조업자의 프로토콜에 따라서 리포펙타민(Lipofectamine) 2000 (라이프 테크놀로지)으로 형질감염시켰다. 간략하게 언급하면, 1.5 μ l의 리포펙타민 2000을 사용하여 950 ng의 총 플라스미드 (Cas9 발현 플라스미드 + gRNA 발현 플라스미드)를 형질감염시켰다. 700 ng의 Cas9 발현 플라스미드, 125 ng의 하나의 gRNA 발현 플라스미드 및 125 ng의 쌍을 형성한 gRNA 발현 플라스미드 (도 6d 및 도 9에 열거된 표적화된 gRNA의 쌍을 수반함). 별도의 웰을 형질감염 대조군으로서 1 μ g의 근적외선 iRFP670 (애드젠 플라스미드 45457)³²로 형질감염시켰다. 형질감염시킨 지 3.5일 후, 세포를 트립신 처리하고, 10% FBS를 보충시킨 DMEM에 재현탁시키며, 20 nm 대역 통과를 나타내는 520 nm 필터 및 488 nm 레이저 여기를 수반한 C6 유동 세포계수기 [아쿠리(Accuri)] 상에서 분석하였다. 각 샘플

플에 대하여, 형질감염 및 유동 세포계수 측정을 1회 수행하였다.

[0179] 계놈 변형의 T7 엔도뉴클레아제 I 서베이어 검증

[0180] HEK293-GFP 안정적 세포를 상기 언급된 바와 같이 Cas9 발현 및 gRNA 발현 플라스미드로 형질감염시켰다. 2개의 별도의 gRNA를 코딩하는 단일 플라스미드를 형질감염시켰다. 발현 플라스미드 (Cas9 발현 + gRNA 발현 플라스미드)의 총량을 적정하는 실험에 대해서는, 700/250, 350/125, 175/62.5, 88/31 ng의 Cas9 발현 플라스미드/ng의 gRNA 발현 플라스미드를, 필요에 따라 불활성 캐리어 플라스미드 pUC19 (NEB)와 조합하여 총 950 ng 형질감염된 플라스미드 DNA에 도달시켰다.

[0181] 계놈 DNA 단리용 키트인 DNAdvance 키트 [아젠코트(Agencourt)]를 이용하여 형질감염시킨지 2일 후에 계놈 DNA를 세포로부터 단리하였다. 간략하게 언급하면, 48-웰 판 중의 세포를 37℃ 하에 5분 동안 40 μ l의 트립신과 함께 인큐베이션하였다. 160 μ l의 DNAdvance 용해 용액을 가하고, 이 용액을 55℃ 하에 2시간 동안 인큐베이션한 후, 아젠코트 DNAdvance 키트 프로토콜에서의 후속 단계를 수행하였다. 40 ng의 단리된 계놈 DNA를 주형으로서 사용하여, 표 1에 명시된 양옆에 위치하고 있는 서베이어(Survey) 프라이머 쌍을 이용하여 표적화된 계놈 유전자 자리를 PCR 증폭시켰다. PCR 생성물을 퀴아퀵 PCR 정제 키트 (퀴아젠)으로 정제하고, 퀀트-iT (Quant-iT™) 피코그린(PicoGreen®) dsDNA 키트 (라이프 테크놀로지)로 정량화하였다. 250 ng의 정제된 PCR DNA를 총 용적 19 μ l 중에서 2 μ l의 NE 완충제 2 (NEB)와 합하고, 변성시킨 다음, 95℃에서 5분, 2℃/s 하의 95 내지 85℃; 0.2℃/s 하의 85 내지 20℃에서 열순환시키면서 재어닐링하였다. 이와 같이 재어닐링된 DNA를 37℃ 하에 15분 동안 1 μ l의 T7 엔도뉴클레아제 I (10 U/ μ l, NEB)과 함께 인큐베이션하였다. 10 μ l의 50% 글리세롤을 상기 T7 엔도뉴클레아제 반응물에 가하고, 12 μ l를 150 V 하에 30분 동안 전기영동시킨 5% TBE 18-웰 크리테리온 (Criterion) PAGE 겔 [바이오-래드(Bio-Rad)] 상에서 분석한 다음, 1x SYBR 골드 (라이프 테크놀로지)로 30분 동안 염색하였다. Cas9-유도된 절단 밴드와 절단되지 않은 밴드를 알파이미저(AlphaImager) HP [알파 이노테크 (Alpha Innotech)] 상에서 가시화하고, 이미지제이(ImageJ) 소프트웨어를 이용하여 정량화하였다³³. 상기와 같이 절단된 밴드의 피크 세기를 모든 밴드 (절단되지 않은 밴드 + 절단된 밴드)의 총 세기로 나누어, 기존에 언급된 바와 같이 유전자 변형 수준을 평가하기 위하여 사용되었던 절단된 분획을 결정하였다²⁸. 각 샘플에 대하여, 형질감염 및 후속 변형 측정을 상이한 날에 세번 되풀이하여 수행하였다.

[0182] 계놈 변형의 고 처리량 서열 분석

[0183] HEK293-GFP 안정적 세포를, Cas9 발현 및 gRNA 발현 플라스미드로 형질감염시키고, 700 ng의 Cas9 발현 플라스미드 플러스 한 쌍의 gRNA를 발현하는 250 ng의 단일 플라스미드를 형질감염시키며 (고 수준), 다만 Cas9 뉴클레아제에 대해서는, 88 ng의 Cas9 발현 플라스미드 플러스 한 쌍의 gRNA를 발현하는 31 ng의 단일 플라스미드를 형질감염시켰다 (저 수준). 계놈 DNA를 상기와 같이 단리하고, 3개의 생물학적 복제물로부터 풀링하였다. 150 ng 또는 600 ng의 풀링된 계놈 DNA를 주형으로서 사용하여, 표 1에 명시된 양옆에 위치하고 있는 HTS 프라이머 쌍을 이용하여 표적에 적중한 계놈 부위와 표적을 벗어난 계놈 부위를 PCR함으로써 증폭시켰다. 조 PCR 생성물의 상대적 양을 겔 전기영동에 의해 정량화하고, 상이한 gRNA 쌍 또는 Cas9 뉴클레아제 유형으로 처리된 샘플을, 퀴아퀵 PCR 정제 키트 (퀴아젠)로 정제하기 전에 등분 농도로 별도로 풀링하였다. 약 500 ng의 풀링된 DNA를 200 V 하에 30분 동안 5% TBE 18-웰 크리테리온 PAGE 겔 (바이오-래드) 상에서 수행하고, 약 125 bp 내지 약 300 bp 길이의 DNA를 단리하고, 퀴아퀵 PCR 정제 키트 (퀴아젠)에 의해 정제하였다. 정제된 DNA를 서열 분석용 적응인자를 함유하는 프라이머를 이용하여 PCR 증폭시키고, 정제한 다음, 기존에 언급된 바와 같이¹ MiSeq 고 처리량 DNA 서열 분석기 [일루미나(Illumina)] 상에서 서열 분석하였다.

[0184] 데이터 분석

[0185] 일루미나 서열 분석용 판독치를 필터링하고, 유닉스 배쉬(Unix Bash)로 작성된 스크립트로 분석하였다. 모든 스크립트는 배쉬로 작성되었다.

[0186] 팻매치(Patmatch) 프로그램³⁸을 사용하여, Cas9 결합 부위에 상응하는 패턴 서열 (배향 A의 경우 CCN N²⁰ 스페이스서 N²⁰NGG 및 배향 B의 경우 N²⁰NGG 스페이스서 CCN N²⁰)에 대한 인간 계놈 (GRCh37/hg19 빌드)을 조사하였다. 계놈 부위의 서열 내에서 삽입-결실을 확인하기 위한 단계는 다음과 같이 발견될 수 있다:

[0187] 1) 서열 판독치를 초기에 필터링하여, 50개 미만 염기의 판독치를 제거하고 B-J가 아닌 일루미나 염기 스코어의 10% 초과를 나타내는 판독치를 제거하였다:

[0188] 예 SeqA-제1 판독치:

TTCTGAGGGCTGCTACCTGTACATCTGCACAAGATTGCCTTTACTCC
ATGCCTTTCTTCTTCTGCTCTAACTCTGACAATCTGTCTTGCCATGCCATAAGCCC
CTATTCTTTCTGTAACCCCAAGATGGTATAAAAAGCATCAATGATTGGGC (서열 157)

[0189]

[0190] 예 SeqA-제2 판독치:

AAAAC TCAAAGAAATGCCCAATCATTGATGCTTTTATACCATCTTG
GGGTTACAGAAAGAATAGGGGCTTATGGCATGGCAAGACAGATTGTCAGAGTTA
GAGCAGAAGAAGAAAGGCATGGAGTAAAGGCAATCTTGTGCAGATGTACAGGT
AA (서열 158)

[0191]

[0192] 2) 1개의 미스매치를 허용하는 SeqA-제1 판독치 중의 SeqA-제2 판독치의 역 보체의 시작으로부터 첫 번째 20개 염기 4개 염기를 발견한다:

[0193] SeqA-제2 판독치의 역 보체:

TTACCTGTACATCTGCACAAGATTGCCTTTACTCCATGCCTTTCTT
CTTCTGCTCTAACTCTGACAATCTGTCTTGCCATGCCATAAGCCCCCTATTCTTTCT
GTAACCCCAAGATGGTATAAAAAGCATCAATGATTGGGCATTTCTTTGAGTTTT
(서열 159)

[0194]

[0195] SeqA-제1 판독치 내에서의 위치

TTCTGAGGGCTGCTACCTGTACATCTGCACAAGATTGCCTTTACTCCATGCCTTT
CTTCTTCTGCTCTAACTCTGACAATCTGTCTTGCCATGCCATAAGCCCCCTATTCTT
TCTGTAACCCCAAGATGGTATAAAAAGCATCAATGATTGGGC (서열 160)

[0196]

[0197] 3) 서열을 정렬한 다음 합하여, 간단한 염기 쌍 정렬에서 5% 초과 미스매치를 수반하는 모든 서열을 제거한다:

[0198] SeqA-제1 판독치와 SeqA-제2 판독치의 조합:

TTCTGAGGGCTGCTACCTGTACATCTGCACAAGATTGCCTTTACTC
CATGCCTTTCTTCTTCTGCTCTAACTCTGACAATCTGTCTTGCCATGCCATAAGCC
CCTATTCTTTCTGTAACCCCAAGATGGTATAAAAAGCATCAATGATTGGGCATTTC
TTTGAGTTTT (서열 161)

[0199]

[0200] 4) 표적 부위를 확인하기 위하여, 양옆에 위치하고 있는 게놈 서열을 대상으로 하여, 양옆에 위치하고 있는 게놈 서열 사이에 1 내지 300개 염기의 다양한 양을 허용해 주는 팻매치 프로그램³⁸으로 조사하였다 (표 2):

[0201] <표 2> 팻매치 서열

표적 부위	하류 게놈 서열	상류 게놈 서열
EMX_On	GGCCTGCTTCGTGGCAATGC(서열 162)	ACCTGGGCCAGGGAGGGAGG(서열 163)
EMX_Off1	CTCACTTAGACTTTCTCTCC(서열 164)	CTCGGAGTCTAGCTCCTGCA(서 열 165)
EMX_Off2	TGGCCCCAGTCTCTCTTCTA(서열 166)	CAGCCTCTGAACAGCTCCCG(서열 167)
EMX_Off3	TGACTTGGCCTTTGTAGGAA(서열 168)	GAGGCTACTGAAACATAAGT(서열 169)
EMX_Off4	TGCTACCTGTACATCTGCAC(서열 170)	CATCAATGATTGGGCATTTC(서 열 171)
VEG_On	ACTCCAGTCCCAAATATGTA(서열 172)	ACTAGGGGGCGCTCGGCCAC(서열 173)
VEG_Off1	CTGAGTCAACTGTAAGCATT(서열 174)	GGCCAGGTGCAGTGATTCAT(서열 175)
VEG_Off2	TCGTGTCATCTTGTGTTGTGC(서열 176)	GGCAGAGCCCAGCGGACACT(서열 177)
VEG_Off3	CAAGGTGAGCCTGGGTCTGT(서열 178)	ATCACTGCCCAAGAAGTGCA(서열 179)
VEG_Off4	TTGTAGGATGTTTAGCAGCA(서열 180)	ACTTGCTCTCTTTAGAGAAC(서 열 181)
CLT2_On	CTCAAGCAGGCCCCGCTGGT(서열 182)	TTTTGGACCAAACCTTTTG(서 열 183)
CLT2_Off1	TGAGGTTATTTGTCCATTGT(서열 184)	TAAGGGGAGTATTTACACCA(서열 185)
CLT2_Off2	TCAAGAGCAGAAAATGTGAC(서열 186)	CTTGCAGGGACCTTCTGATT(서 열 187)
CLT2_Off3	TGTGTGTAGGACTAAACTCT(서열 188)	GATAGCAGTATGACCTTGGG(서열 189)

[0202]

[0203]

인간 게놈 (GRCh37/hg19 빌드) 중의 참조 게놈 부위와 동일한 크기에 상응하는 모든 표적 부위 서열을 변형되지 않은 것으로 간주하였고, 참조 크기에 상응하지 않는 모든 서열을 잠재적 삽입 또는 결실물로 간주하였다. 클러스털(Clustal)W³⁹를 이용하여 참조 크기에 상응하지 않는 서열을 참조 게놈 부위와 정렬시켰다. 2개의 절반-부위 서열 내의 또는 그 사이의 DNA 스페이스 서열 내에 2개 이상의 삽입 또는 1개의 결실을 수반하는 것으로 정렬된 서열이 삽입-결실로 간주되었다. 고 처리량 서열 분석으로 인해서, 1개 염기 쌍의 삽입 또는 결실 (미스-단계화)이 낮은 비율이긴 하지만 적절한 비율로 초래될 수 있기 때문에, 2 bp의 삽입-결실은 Cas9 유도된 변형으로부터 발생될 가능성이 더 크다.

[0204]

서열 분석 실험을 위한 샘플 크기를 최대화하여 (실제 실험 고려 사항 내에서), 효과를 탐지하기 위한 가장 강력한 영향력을 보장하였다. 표 3에서 Cas9-변형된 게놈 부위에 대한 통계적 분석을, 본페로니(Bonferroni) 방법을 이용한 다중 비교 교정을 이용하여 기준에 언급된 바와 같이³⁴ 수행하였다.

[0205]

다음의 결과를 지칭한 표 3은 (A) 야생형 Cas9, Cas9 니카제 또는 fCas9를 발현하는 플라스미드, 및 *CLTA* 표적

에 적중한 부위를 표적화하는 2개의 gRNA (gRNA C3과 gRNA C4)를 발현하는 단일 플라스미드로 처리된 인간 세포로부터 단리된 150 ng 게놈 DNA로부터 증폭된, 기존에 보고된 표적을 벗어난 게놈 부위 및 *CLTA* 표적에 적중한 부위를 서열 분석한 것으로부터의 결과를 나타낸다. 음성 대조군으로서, *GFP* 유전자 표적에 적중한 부위를 표적화하는 2개의 gRNA (gRNA G1, G2 또는 G3과 gRNA G4, G5, G6 또는 G7)를 사용하는 것을 제외하고는, 형질감염 및 서열 분석을 상기와 같이 수행하였다. 삽입-결실: 3개의 Cas9 뉴클레아제-유도된 절단 중 어느 것과 일치하는 삽입 또는 결실을 함유하는 것으로 관찰된 서열의 수. 총: 첫 번째 10,000개 서열만을 대상으로 하여 표적에 적중한 부위 서열에 관하여 분석하는 동안의 서열 계수치의 총 수. 변형된: 삽입-결실의 수를 서열의 총 수로 나눈다 (%로서 나타냄). 잠재적 변형의 상한치는 1개 미만의 삽입-결실이 존재한다고 가정한 다음, 이를 상한치 변형 비율 (%)에 도달하는 총 서열 계수치로 나누거나, 또는 이론적 탐지 한계치 (1/49,500)를 취함으로써 (어느 것이든 더 큰 것을 고른다), 관찰되지 않은 삽입-결실을 수반한 부위에 대하여 계산하였다. P-값: 야생형 Cas9 뉴클레아제, Cas9 니카제 또는 fCas9 뉴클레아제에 대해서는, *CLTA* 표적에 적중한 부위를 표적화하는 2개의 gRNA로 처리된 각 샘플과 *GFP* 표적에 적중한 부위를 표적화하는 2개의 gRNA로 처리된 대조군 샘플 간의 양측 피셔의 직접 확률 시험을 이용하여 기존에 보고된 바와 같이¹⁸ P-값을 계산하였다. < 0.0045의 P-값은 유의적인 것으로 간주되었고, 본페로닌 방법을 이용한 보존적 다중 비교 교정에 근거하여 제시되었다. on:off 특이성은 각 부위에 대한 표적에 적중한 게놈 변형 빈도 대 표적을 벗어난 게놈 변형 빈도의 비이다. (B)는 *EMX* 표적에 적중한 부위를 표적화하는 2개의 gRNA (gRNA E1과 gRNA E2)를 발현하는 단일 플라스미드를 이용하여 *EMX* 표적 부위에 적용된 (A)에서와 같은 실험적 및 분석적 방법을 나타낸다. (C)는 *VEGF* 표적에 적중한 부위를 표적화하는 2개의 gRNA (gRNA V1과 gRNA V2)를 발현하는 단일 플라스미드를 이용하여 *VEGF* 표적 부위에 적용된 (A)에서와 같은 실험적 및 분석적 방법을 나타낸다. (D)는 탐지 감도를 1/198,000로 증가시키기 위하여 600 ng 게놈 DNA로부터 증폭된 *VEGF* 표적에 적중한 부위 및 *VEGF* 표적을 벗어난 부위 1에 적용된 (A)에서와 같은 실험적 및 분석적 방법을 나타낸다.

[0206] <표 3>

[0207] 표적에 적중한 게놈 부위 및 표적을 벗어난 게놈 부위에서 야생형 Cas9, Cas9 니카제, 및 fCas9에 의해 유도된 세포성 변형

(A)

뉴클레아제 유형:	wt Cas9	wt Cas9	Cas9 니카제	fCas9	wt Cas9	Cas9 니카제	fCas9
gRNA 쌍 표적:	<i>CLTA</i>	<i>CLTA</i>	<i>CLTA</i>	<i>CLTA</i>	<i>GFP</i>	<i>GFP</i>	<i>GFP</i>
총 발현 플라스미드 (ng):	1000	125	1000	1000	1000	1000	1000
<u><i>CLTA</i> 부위</u>							
<u>CLT2_On</u>							
삽입-결실	3528	1423	3400	575	3	13	5
총	10000	10000	10000	10000	10000	10000	10000
변형된 (%)	35.280	14.230	34.000	5.750	0.030	0.130	0.050
P-값	<1.0E-300	<1.0E-300	<1.0E-300	1.4E-163			
On:off 특이성	1	1		1			
<u>CLT2_Off1</u>							
삽입-결실	316	44	2	2	1	3	3
총	60620	64755	71537	63079	93883	91306	82055
변형된 (%)	0.521	0.068	0.003	0.003	<0.002	0.003	0.004
P-값	1.3E-126	2.1E-16					
On:off 특이성	68	209		>2850			
<u>CLT2_Off2</u>							
삽입-결실	11	5	3	1	1	1	2
총	72596	51093	59632	35541	69114	64412	39978
변형된 (%)	0.015	0.010	0.005	0.003	<0.002	<0.002	0.005
P-값	6.5E-03						
On:off 특이성	2328	1454		>2850			
<u>CLT2_Off3</u>							
삽입-결실	11	10	0	0	1	1	1
총	52382	44212	54072	48668	55670	58707	54341
변형된 (%)	0.021	0.023	<0.002	<0.002	<0.002	<0.002	<0.002
P-값	2.7E-03	3.5E-03					
On:off 특이성	1680	629		>2850			

[0208]

(B)

뉴클레아제 유형:	wt Cas9	wt Cas9	Cas9 니카제	fCas9	wt Cas9	Cas9 니카제	fCas9
gRNA 쌍:	EMX	EMX	EMX	EMX	GFP	GFP	GFP
총 발현 플라스미드 (ng):	1000	125	1000	1000	1000	1000	1000
<u>EMX 부위</u>							
<u>EMX_On</u>							
삽입-결실	5111	2683	2267	522	0	0	2
총	10000	10000	10000	10000	10000	10000	10000
변형된 (%)	51.110	26.830	22.670	5.220	<0.002	<0.002	0.020
P-값	<1.0E-300	<1.0E-300	<1.0E-300	1.0E-154			
On:off 특이성	1	1	1	1			
<u>EMX_Off1</u>							
삽입-결실	386	122	7	1	4	9	7
총	109787	83420	124564	88424	102817	90020	96526
변형된 (%)	0.352	0.146	0.006	<0.002	0.004	0.010	0.007
P-값	1.3E-103	2.8E-37					
On:off 특이성	145	183	>11222	>2584			
<u>EMX_Off2</u>							
삽입-결실	74	58	3	6	3	0	4
총	98568	94108	105747	78871	81717	79469	79193
변형된 (%)	0.075	0.062	0.003	0.008	0.004	<0.002	0.005
P-값	3.2E-16	1.4E-12					
On:off 특이성	681	435	>11222	>2584			
<u>EMX_Off3</u>							
삽입-결실	736	178	20	14	12	11	17
총	72888	65139	82348	59593	74341	73408	75080
변형된 (%)	1.010	0.273	0.024	0.023	0.016	0.015	0.023
P-값	2.5E-202	3.1E-44					
On:off 특이성	51	98	>11222	>2584			
<u>EMX_Off4</u>							
삽입-결실	4149	620	3	3	6	7	5
총	107537	91695	91368	91605	111736	119643	128088
변형된 (%)	3.858	0.676	0.003	0.003	0.005	0.006	0.004
P-값	<1.0E-300	1.9E-202					
On:off 특이성	13	40	>11222	>2584			

(C)

뉴클레아제 유형: wt Cas9 wt Cas9 Cas9 fCas9 wt Cas9 Cas9 fCas9

[0209]

	니카제				니카제		
	gRNA 쌍: 총 발현 플라스미드 (ng):	VEGF	VEGF	VEGF	GFP	GFP	GFP
VEGF 부위	1000	125	1000	1000	1000	1000	1000
VEG_On							
삼입-결실	5253	2454	1230	1041	8	0	1
총	10000	10000	10000	10000	10000	10000	10000
변형된 (%)	52.530	24.540	12.300	10.410	0.080	<0.002	0.010
P-값	<1.0E-300	<1.0E-300	<1.0E-300	6.6E-286			
On:off 특이성	1	1	1	1			
VEG_Off1							
삼입-결실	2950	603	22	0	0	4	1
총	82198	71163	90434	77557	74765	79738	74109
변형된 (%)	3.589	0.847	0.024	<0.002	<0.002	0.005	<0.002
P-값	<1.0E-300	3.2E-188	2.5E-06				
On:off 특이성	15	29	506	>5150			
VEG_Off2							
삼입-결실	863	72	3	3	0	2	1
총	102501	49836	119702	65107	54247	65753	61556
변형된 (%)	0.842	0.144	0.003	0.005	<0.002	0.003	<0.002
P-값	3.5E-159	9.6E-24					
On:off 특이성	62	170	>6090	>5150			
VEG_Off3							
삼입-결실	260	33	3	2	3	1	0
총	91277	83124	90063	84385	62126	68165	69811
변형된 (%)	0.285	0.040	0.003	0.002	0.005	<0.002	<0.002
P-값	6.8E-54	1.0E-05					
On:off 특이성	184	618	>6090	>5150			
VEG_Off4							
삼입-결실	1305	149	3	2	3	2	4
총	59827	41203	65964	57828	60906	61219	62162
변형된 (%)	2.181	0.362	0.005	0.003	0.005	0.003	0.006
P-값	<1.0E-300	2.7E-54					
On:off 특이성	24	68	>6090	>5150			

(D)

뉴클레아제 유형:	Cas9		Cas9	
	니카제	fCas9	니카제	fCas9
gRNA 쌍:	VEGF	VEGF	GFP	GFP

[0210]

총 발현 플라스미드 (ng):	1000	1000	1000	1000
VEGF 부위				
VEG_On				
삽입-결실	2717	2122	10	13
총	10000	10000	10000	10000
변형된 (%)	27.170	21.220	0.100	0.130
P-값	<1.0E-300	<1.0E-300		
On:off 특이성	1	1		
VEG_Off1				
삽입-결실	67	30	3	2
총	302573	233567	204454	190240
변형된 (%)	0.022	0.013		
P-값	5.9E-12	2.5E-06		
On:off 특이성	1227	1652		

[0211]

[0212]

결과

[0213]

1개의 DNA 가닥만을 절단하는 Cas9의 최근에 조작된 변이체 ("니카제")는, 이중 가닥 브레이크가 2개의 별개의 gRNA 서열⁵⁻⁷에 의해 지정될 수 있게 하지만, DNA와 개별적으로 결합되는 경우에⁹⁻¹¹ 각 단량체성 니카제가 여전히 활성을 유지할 수 있는 능력으로부터 비롯되는 표적을 벗어난 절단 활성^{6,8}으로 인해 여전히 고통받을 수 있게 된다. 이와는 달리, DNA를 절단하기 위해 2개의 *FokI*-dCas9 단량체의 동시적인 DNA 결합과 연합을 필요로 하는 촉매적으로 사멸된 Cas9와의 *FokI* 뉴클레아제 융합물을 개발하는 것이 본원에 기재되어 있다. 이와 같이 조작된 *FokI*-dCas9 (fCas9)의 표적을 벗어난 DNA 절단은 약 15개 또는 25개 염기 쌍 떨어진 2개의 gRNA가 양옆에 위치한 부위만을 절단해야 하는 요구 사항 (니카제 보다 훨씬 더 엄격한 간격 요구 사항이다)에 의해 추가로 저해된다. 인간 세포에서는, fCas9가 니카제와 거의 동등한 수준의 효율, 및 야생형 Cas9 보다 >140배 더 높은 특이성으로 표적 DNA 부위를 변형시켰다. fCas9의 기질 요구 사항에 따르는 표적 부위는 인간 게놈 내에 풍부하여, 34 bp마다 평균 1회 존재한다.

[0214]

세포에서, Cas9:gRNA-유도된 이중 가닥 브레이크는 비-상동 말단 결합 (NHEJ)을 통하여 기능적 유전자 녹아웃을 초래할 수 있거나 또는 외인성 DNA 주형을 수반한 상동성-지시된 복구 (HDR)을 통하여 사실상 모든 서열에 대한 표적 유전자 자리의 변경이 초래될 수 있다^{9,15,16}. Cas9는 특히 편리한 게놈 편집 플랫폼인데¹⁷, 이는 각각의 새로운 관심 표적 부위에 대한 게놈 편집 작용제가 상응하는 gRNA를 단순히 생성시킴으로써 접근될 수 있기 때문이다. 이러한 접근 방식은 세포 및 모델 유기체에서 표적화된 녹아웃과 유전자 삽입을 창출시키기 위해 광범위하게 사용되어 왔고, 또한 그의 잠재적 치료 관련성에 대해 인식되어 왔다.

[0215]

Cas9:gRNA 시스템이 전례없는 수준의 프로그램 가능성과 사용 용이성을 제공하긴 하지만, 연구 도구로서 및 잠재적 치료제로서 Cas9의 유용성과 안전성을 제한할 수 있는 의도하지 않은 유전자 자리의 변형을 초래하는, 표적을 벗어난 게놈 부위를 절단할 수 있는 Cas9의 능력에 관한 연구¹⁻⁵가 보고되었다. 2개의 동시적인 인접한 Cas9:DNA 결합 현상이 일어나는 경우에만 DNA를 절단하도록 Cas9 변이체를 조작하는 것이, 게놈 편집 특이성을 실질적으로 개선시킬 수 있는 것으로 가정되었는데, 이는 2개의 인접한 표적을 벗어난 결합 현상이 발생할 가능성이, 표적을 벗어난 단일 결합 현상이 발생할 가능성 보다 훨씬 더 적기 때문이다 (대략 $1/n^2$ 대 $1/n$). 이러한 접근 방식은 dsDNA의 단일 가닥 만을 절단하는 돌연변이체 Cas9 단백질, 예컨대 니카제의 최근 개발과는 완전히 다르다. 니카제를 사용하여 2개의 가까운 표적 부위의 반대 가닥을 니킹하면, 효과적으로 이중 가닥 브레이크인 것을 생성시킬 수 있고, 쌍을 이룬 Cas9 니카제는 실질적으로 표적에 적중한 DNA 변형을 수행할 수 있는데, 표적을 벗어난 변형은 감소된다^{5,6,8}. 성분 Cas9 니카제 각각은 여전히 촉매적으로 활성이고⁹⁻¹¹ 단일 가닥 DNA 절단 현상은 약한 수준의 돌연변이 유발성이기 때문에^{18,19}, 니카제는 단량체로서 작용하는 경우에도 게놈 변형을 유도시킬 수 있다^{5,7,16}. 실제로, Cas9 니카제는 세포에서 표적을 벗어난 변형을 유도시키는 것으로 기존에 보고되었다^{6,8}. 더욱이, 쌍을 이룬 Cas9 니카제는 약 100 bp 이하 떨어져 결합된 경우에도 dsDNA 절단-유래 변

형 현상을 효율적으로 유도시킬 수 있기 때문에⁶, 쌍을 이룬 니카제에 대한 표적을 벗어난 잠재적 부위의 통계적 수는, 보다 공간적으로 제한된 이량체성 Cas9 절단 시스템의 것 보다 더 많다.

[0216] Cas9:gRNA 시스템의 특이성을 추가로 개선시키기 위하여, 절대적인 이량체성 Cas9 시스템이 본원에 제공된다. 본 실시예에서는, *FokI* 제한 엔도뉴클레아제 절단 도메인을 촉매적으로 사멸된 Cas9 (dCas9)와 융합시키면, 2개의 별개의 *FokI*-dCas9:gRNA 복합체가 특별한 간격 제약을 받는 인접한 부위 ("절반-부위")와 결합되는 경우에만 DNA를 절단할 것인 절대적인 이량체성 Cas9가 창출되었다 (도 6d). 단량체에 의한 단일 가닥 DNA 절단이 독립적으로 발생하는 Cas9 니카제와는 달리, *FokI*-dCas9의 DNA 절단을 위해서는 2개의 별개의 *FokI*-dCas9 단량체를 동시에 결합시켜야 하는데, 이는 단량체성 *FokI* 뉴클레아제 도메인이 촉매적으로 적격하지 않기 때문이다²¹. 이러한 접근 방식은 *FokI*-dCas9 이량체의 양 단량체로 인해 지정된 표적 염기의 수를 배가시킴으로써, 야생형 Cas9와 비교해서 DNA 절단의 특이성을 증가시켰고, 단량체성 *FokI*-dCas9:gRNA 복합체의 불활성에 기인하여 니카제와 비교해서 개선된 특이성을 부여하였으며, *FokI*-dCas9 이량체의 어셈블리에 대한 보다 엄격한 공간적 요구 사항을 부여하였다.

[0217] Cas9와 짧은 기능성 펩티드 태그와의 융합이 gRNA-프로그램된 전사 조절을 가능하게 하는 것으로 보고되긴 하였지만²², Cas9와 활성 효소 도메인과의 융합은 기존에 전혀 보고된 바가 없는 것으로 여겨진다. 따라서, *FokI* 뉴클레아제 도메인, 불활성화 돌연변이 D10A 및 H840A를 함유하는 dCas9, 및 핵 국제화 서열 (NLS)의 별개의 입체 배치를 수반한 광범위한 *FokI*-dCas9 융합 단백질들을 구축하였고 그 특징을 명확히 규명하였다. *FokI*를 dCas9의 N-말단 또는 C-말단과 융합시키고, 상기 NLS가 어느 한 말단에 위치하거나 또는 두 도메인 사이에 위치하도록 변화시켰다 (도 6b). 링커 서열의 길이는 *FokI* 도메인과 dCas9 도메인 사이의 Gly-Gly-Ser (GGG)의 1개 또는 3개 반복 서열로서 변화시켰다. 기존에 개발된 이량체성 뉴클레아제 시스템은 절반-부위들 사이의 스페이서 서열의 길이에 민감하기 때문에^{23,24}, 광범위한 스페이서 서열 길이를 시험 표적 유전자인 에머랄드 GFP (본원에서 GFP로서 후술됨) 내의 2개의 gRNA 결합 부위 사이에서 시험하였다 (도 6c 및 도 9). 상이한 배향을 수반한 gRNA 결합-부위 쌍 2개 세트를 GFP 내에서 선택하였다. 1개 세트는 스페이서 서열로부터 원위의 NGG PAM 서열 쌍에 놓아둔 반면 [gRNA의 5' 말단은 스페이서에 인접해 있다 (배향 A)] (도 6c), 다른 세트는 스페이서에 바로 인접한 PAM 서열에 놓아두었다 (배향 B) (도 9). 전체로서, 7개 쌍의 gRNA가 배향 A에 적합하고, 9개 쌍의 gRNA가 배향 B에 적합하였다. gRNA 표적들을 쌍을 이뤄 조합함으로써, 8개 스페이서 길이를 양 이량체 배향에서 시험하였는데, 배향 A에서는 5 내지 43 bp의 범위이고, 배향 B에서는 4 내지 42 bp의 범위이다. 전체로서, 104개 쌍의 *FokI*-dCas9:gRNA 복합체에 상응하는 DNA 구조물을 생성시키고 시험하여, 4개의 융합 구조물, 17개의 단백질 링커 변이체 (다음에 기재됨), 양 gRNA 배향 및 절반 부위들 사이의 13개 스페이서 길이를 검토하였다.

[0218] 이들 후보 *FokI*-dCas9:gRNA 쌍의 활성을 검증하기 위하여, HEK293 세포에서 안정적으로 통합되고 구성적으로 발현된 GFP 유전자의 DNA 절단 및 NHEJ로 인해 세포성 형광이 상실되는, 기존에 보고된 유동 세포계수법-기반 형광 검정^{2,8}을 사용하였다 (도 10). 비교를 위해, *FokI*-dCas9 변이체의 초기 세트를 대상으로 하여, 양 gRNA 스페이서 배향 세트 A 및 B 전반에 걸쳐 동일한 발현 플라스미드 내의 상응하는 Cas9 니카제 및 야생형 Cas9를 이용하여 나란히 검증하였다. 상응하는 Cas9 단백질 발현 플라스미드를 적당한 쌍의 gRNA 발현 플라스미드와 함께 일시적으로 공동-형질감염시킴으로써, Cas9 단백질 변이체 및 gRNA를 세포에서 생성시켰다. *FokI*-dCas9 변이체, 니카제, 및 야생형 Cas9는 모두, 동일한 gRNA를 이용하여 동일한 DNA 부위를 표적화하였다.

[0219] 대부분의 초기 *FokI*-dCas9 융합 변이체는 불활성이거나 또는 매우 약하게 활성이었다 (도 11). 그러나, NLS-*FokI*-dCas9 구조 (N 말단에서 C 말단으로 열거됨)는 배향 A로 사용되는 경우에 (PAM은 스페이서로부터 원위에 있다) 상응하는 gRNA 없는 대조군 위에서 GFP-음성 세포를 10% 증가시켰다 (도 11a). 이와는 달리, NLS-*FokI*-dCas9 활성은, 스페이서에 인접한 PAM을 수반한 gRNA 쌍 상에서 사용되는 경우에 탐지 가능하지 않았다 (도 11b). 최근에 보고된 Cas9 구조물^{25,26}을 조사한 결과, Cas9 N-말단이 RuvCI 도메인으로부터 돌출되어, gRNA:DNA 이중체의 5' 말단과 접촉되는 것으로 밝혀졌다. 어떠한 특별한 이론에 얽매는 것은 아니지만, 이러한 배열은 N-말단적으로 융합된 *FokI*를 PAM으로부터 원위에 배치시켜, 절단된 스페이서로부터 원위의 PAM을 수반한 gRNA 쌍에 대한 선호도가 생기는 것으로 추정된다 (도 6d). 일부 경우에 다른 *FokI*-dCas9 융합 쌍 형성과 다른 gRNA 배향이 보통 수준의 활성을 나타내긴 하였지만 (도 11), 배향 A에서 gRNA를 수반한 NLS-*FokI*-dCas9가 추가 개발을 위해 선택되었다.

[0220] 이어서, NLS-*FokI*-dCas9 구조에서 NLS와 *FokI* 도메인 사이의 단백질 링커, 및 *FokI* 도메인과 dCas9 사이의 단백질 링커를 최적화하였다. 광범위한 아미노산 조성, 예측되는 유연성, 및 9개 내지 21개 잔기로 다양한 길이

를 수반한 17개 링커를 시험하였다 (도 12a). 게놈 GFP 변형의 가장 높은 수준을 뒷받침하는 바와 같이 (도 12b), 개방되고 연장된 입체 형태를 수반한 기존에 보고된 조작된 단백질²⁷에 근거하여, *FokI* 도메인과 dCas9 사이에서 가요성 18개-잔기 링커, (GGG)₆ (서열 15), 및 16개-잔기 "XTEN" 링커 (도 12a에서의 *FokI*-L8)를 확인하였다.

[0221] 화학적 PEG화에 대한 단백질-기반 기능성 유사체로서 작용하는, 그들의 유체역학적 반경을 증가시킴으로써 해독 상 융합된 생물학적 약물의 혈청 반감기를 연장시키도록 XTEN 단백질을 최초로 설계하였다³⁵. 화학적으로 안정하고, 비-양이온성, 및 비-소수성 일차 서열, 및 연장된 입체 형태를 보유함으로써, XTEN의 일부가 융합 단백질에 대한 안정적이고 불활성인 링커 서열로서 기능할 수 있었다고 가정된다. E-XTEN으로부터의 XTEN 단백질 태그의 서열을 분석하였고, 아미노산 서열 내의 반복되는 모티프를 정렬하였다. *FokI*-dCas9 융합 구조물 *FokI*-L8에 사용된 사열 (도 12a)은 공통의 E-XTEN 모티프의 컨센서스 서열로부터 유래되었고, 이러한 모티프 내의 16개 아미노산을 선택하여 *FokI*-dCas9 링커로서 시험하였다.

[0222] 시험된 많은 *FokI*-dCas9 링커 (최적의 XTEN 링커 포함)는 절반 부위들 사이에 약 15 및 약 25 bp의 스페이서 길이에 대한 현저한 선호도를 나타내는 뉴클레아제를 초래하였는데, 20 bp를 포함한 다른 모든 스페이서 길이는 실질적으로 더 낮은 활성을 나타내었다 (도 12b). 이러한 패턴의 링커 선호도는 *FokI*-dCas9 융합물이 DNA 이중 나선의 반대 면과 결합하여 DNA를 절단해야만 하는 모델과 일치하였는데, 최적의 결합은 약 1.5 또는 2.5회의 나선형 회전만큼 떨어져 일어난다. NLS-*FokI* 링커의 변이는, 특히 XTEN *FokI*-dCas9 링커와 조합되는 경우에, 뉴클레아제 성능에 강력한 영향을 미치지 못하였다 (도 12b).

[0223] NLS-*FokI*-dCas9 구조 내에서 *FokI* 도메인과 dCas9 사이의 링커를 검정하는 것 이외에도, N-말단 NLS와 *FokI* 도메인 사이의 4개의 링커 변이체를 또한 시험하였다 (도 12a). NLS-GSAGSAAGSGEF(서열 20)-*FokI*-dCas9 링커가, 단순한 GGS 링커를 *FokI*와 dCas9 도메인 사이에 사용한 경우에 시험된 다른 NLS-*FokI* 링커 보다 거의 2배 더 우수한 GFP 유전자 변형을 나타내긴 하였지만 (도 12b), GSAGSAAGSGEF (서열 20) 링커는 *FokI*와 dCas9 도메인 사이에 XTEN 링커와 조합되는 경우에 실질적으로 더 우수하게 수행되지 못하였다.

[0224] NLS-GGS-*FokI*-XTEN-dCas9 구조물은 일관되게, 시험된 후보들 중에서 가장 높은 활성을 나타내어, Cas9 니카제 및 야생형 Cas9 뉴클레아제 각각에 대한 약 20% 및 약 30%와 비교해서, 세포의 약 15%에서 GFP의 상실을 유도하였다 (도 7a). fCas9로서 후술되는 상기 구조물을 이용하여, 모든 후속 실험을 수행하였다. 게놈 표적 부위를 효율적으로 변형시킬 수 있는 fCas9의 능력을 확장하기 위하여, T7 엔도뉴클레아제 I 서베이어 검정²⁸을 사용하여, fCas9, Cas9 니카제 또는 야생형 Cas9, 및 배향 A에서의 2개의 별개의 gRNA 또는 음성 대조군으로서 gRNA 없이 처리된 HEK293 세포에서 통합된 GFP 유전자 내의 7개 표적 부위 각각에서의 돌연변이의 양을 측정하였다. 유동 세포계수법-기반 연구와 일치하게, fCas9는 약 15 또는 약 25 bp의 최적의 스페이서 길이를 이용하여 약 20%의 비율로 GFP 표적 부위를 변형시킬 수 있었는데, 이는 니카제-유도된 변형 효율과 거의 동등한 수준이었고 야생형 Cas9의 효율의 대략 2/3였다 (도 7a-c).

[0225] 이어서, 서베이어 검정에 의해 4개의 별개의 내인성 게놈 유전자 자리를 변형시킬 수 있는 최적화된 fCas9의 능력을 평가하였다. *CLTA* (2개 부위), *EMX* (2개 부위), *HBB* (6개 부위), *VEGF* (3개 부위)를, 다양한 길이 간격을 둔 배향 A에서의 1개 부위당 2개의 gRNA로 표적화하였다 (도 13). GFP를 표적화하는 실험 결과와 일치하게, 적당하게 간격을 둔 표적 절반-부위에서, fCas9는 4개 모든 유전자의 효율적인 변형을 유도시켰는데, 8% 내지 22% 범위의 표적 염색체 부위 변형을 나타내었다 (도 7d-g 및 도 14). 표적화된 5개 유전자 (GFP 포함) 각각에서 가장 높은 변형을 초래하는 gRNA 스페이서 길이 중에서, fCas9가 평균 15.6% ($\pm$ 6.3% s.d.) 변형을 유도시킨 반면, Cas9 니카제 및 야생형 Cas9는 각각, 각 유전자에 대한 그들의 최적의 gRNA 쌍으로부터 평균 22.1% ($\pm$ 4.9% s.d.) 및 30.4% ($\pm$ 3.1% s.d.) 변형을 유도시켰다. 형질감염 동안 Cas9 발현 플라스미드와 gRNA 발현 플라스미드의 양을 감소시키는 것이 일반적으로, Cas9 니카제 및 fCas9에 대한 게놈 변형 활성을 비례적으로 감소시키지 않았기 때문에 (도 15a-c), 시험된 조건 하에서는 발현이 제한되지 않는 것으로 예상된다.

[0226] fCas9의 gRNA 요구 사항이 잠재적으로, fCas9의 표적을 벗어난 잠재적 기질의 수를 제한하기 때문에, 표적 GFP 서열을 절단할 수 있는 fCas9, Cas9 니카제, 및 야생형 Cas9의 능력에 대한 가이드 RNA 배향의 효과를 비교하였다. 기존의 보고 내용과 일치하게^{5,6,17}, Cas9 니카제는 가이드 RNA가 야생형 Cas9와 유사하게 배향 A 또는 배향 B로 결합되는 경우에 표적을 효율적으로 절단하였다 (도 8a, b). 이와는 달리, fCas9는 가이드 RNA가 배향 A로 정렬된 경우에 GFP 표적만을 절단하였다 (도 8a). 이러한 배향 요구 사항은 바람직하지 못한 표적을 벗어난

DNA 절단에 대한 기회를 추가로 제한한다.

[0227] 중요하게도, 4개의 단일 gRNA 중 어느 것도 fCas9와 함께 개별적으로 발현되는 경우에는 GFP 붕괴 또는 서베이어 검정에 의해서 어떠한 변형도 관찰되지 않았는데, 이는 *FokI* 활성을 위해서는 2개의 동시적인 결합 현상이 필요하기 때문에 예상된 바와 같다 (도 7b 및 도 8c). 이와는 달리, GFP 붕괴는 야생형 Cas9를 동반한 모든 단일 gRNA의 발현 (예상된 바와 같음), 및 2개의 단일 gRNA의 경우에는 Cas9 니카제를 동반한 발현으로부터 비롯되었다 (도 8c). 놀랍게도, 서베이어 검정 결과, GFP가 단일 gRNA를 수반한 야생형 Cas9에 의해 심하게 변형되긴 하였지만, fCas9 또는 Cas9 니카제는 단일 gRNA로 처리된 세포에서 탐지 가능한 변형을 전혀 나타내지 않은 것으로 (< 약 2%) 밝혀졌다 (도 16A). 단일 gRNA 및 fCas9, Cas9 니카제, 또는 야생형 Cas9로 처리된 세포 내의 GFP 표적 부위에서 삽입-결실을 탐지하기 위한 고 처리량 서열 분석 결과, 야생형 Cas9에 의해 예상되는 실질적 수준의 변형 (서열 판독치의 3 내지 7%)이 밝혀졌다. 4개의 단일 gRNA 중 어느 것의 존재 하에서 fCas9에 의한 변형은 배경 위에서 탐지되지 않았는데 (< 약 0.03% 변형), 이는 DNA를 절단하기 위해 2개의 gRNA를 개입시켜야 하는 fCas9의 요구 사항과 일치한다. 이와는 달리, 단일 gRNA의 존재 하에서의 Cas9 니카제는 표적 부위에서 0.05% 내지 0.16% 범위의 변형 수준을 초래하였다 (도 16B). 단일 gRNA로 Cas9 니카제를 처리한 후 표적 부위에서 진짜의 삽입-결실이 탐지되는 것은, 게놈 DNA 니킹의 돌연변이 유발 가능성을 확증시켜 주는데, 이는 기존의 보고 내용과 일치한다^{5,7,18,19}.

[0228] 그러나, 니카제-유도된 DNA 변형의 관찰된 비율은 유동 세포계수법 검정에서의 훨씬 더 높은 GFP 붕괴 신호를 설명하지 못하였다 (도 8c). Cas9 니카제와 함께 GFP 신호 손실을 유도시켰던 gRNA (gRNA G1 및 G3)는 둘 다, GFP 유전자의 비-주형 가닥을 표적으로 하고; 특정 유전자의 코딩 영역에서 dCas9를 이용하여 비-주형 가닥을 표적화하는 것이 효율적인 전사 억제를 매개하는 것으로 공지되어 있기 때문에²⁹, G1 또는 G3 단일 가이드 RNA와 조합된 Cas9 니카제가 저 수준의 게놈 변형 이외에도, 상당한 전사 억제를 유도시킨 것으로 추정된다. 동일한 효과가 fCas9에 대해서는 관찰되지 않았는데, 이는 fCas9가 전사 기구에 의해 DNA로부터 보다 용이하게 대체될 수 있다는 것을 제안하고 있다. 취합해 보면, 이들 결과는 fCas9가, 단일 가이드 RNA와 결합되는 경우에 DNA를 절단할 수 있는 야생형 Cas9 및 Cas9 니카제의 능력과는 달리, 인접한 부위를 표적화하는 2개의 가이드 RNA의 동시 개입을 요구하는 방식으로 게놈 DNA를 효율적으로 변형시킬 수 있다는 것을 표시한다.

[0229] 상기 결과는 집합적으로, Cas9 니카제와 비교해서 fCas9에 대해서는 훨씬 더 엄격한 스페이스, gRNA 배향 및 가이드 RNA 쌍 형성을 요구한다는 것을 나타낸다. fCas9와는 달리 (도 17), Cas9 니카제는 검정된 모든 스페이스 전반에 걸친 부위 (본 작업에서 배향 A에서 5-bp 내지 47-bp, 및 배향 B에서 4 내지 42 bp)를 절단하였다 (도 8a, b). 이들 관찰 결과는 100 bp 이하 떨어진 스페이스 길이를 이용하여 gRNA에 의해 표적화된 부위를 변형시키는 Cas9 니카제의 기존 보고 내용과 일치한다⁶. Cas9 니카제와 비교해서 fCas9의 보다 엄격한 스페이스 및 gRNA 배향 요구 사항은 fCas9의 잠재적 게놈의 표적을 벗어난 부위의 수를 대략 10배 정도 감소시킨다 (표 4). fCas9의 보다 엄격한 스페이스 요구 사항이 또한, 잠재적으로 표적 가능한 부위의 수를 감소시키긴 하지만, fCas9 스페이스 및 이중 PAM 요구 사항에 따르는 서열이 인간 게놈 내에 34 bp 마다 평균 1회 존재한다 (3.1×10^9 bp 중에 9.2×10^7 개 부위) (표 4). 상이한 PAM 특이성을 지닌 Cas9 상동체의 수가 증가하는 것이³⁰ 본원에 기재된 바와 같이 사용하는 것을 잘 받아들이고, fCas9 접근 방식을 이용하여 표적 가능한 부위의 수를 추가로 증가시킬 것으로 또한 예상된다.

[0230] 표 4에서 (A) 칼럼 2는 2개의 gRNA 결합 부위 사이에 -8 bp 내지 25 bp (칼럼 1)의 스페이스 길이를 허용하는 배향 A에서의 쌍을 이룬 gRNA 결합 부위를 수반한 인간 게놈에서의 부위 수를 나타낸다. 배향 A에서의 gRNA 결합 부위는 스페이스 서열로부터 원위의 NGG PAM 서열을 갖는다 (CCNN₂₀-스페이스-N₂₀NGG). 칼럼 3은 2개의 gRNA 결합 부위 사이에 4 bp 내지 25 bp (칼럼 1)의 스페이스 길이를 허용하는 배향 B에서의 쌍을 이룬 gRNA 결합 부위를 수반한 인간 게놈에서의 부위 수를 나타낸다. 배향 B에서의 gRNA 결합 부위는 스페이스 서열에 인접한 NGG PAM 서열을 갖는다 (N₂₀NGG 스페이스 CCNN₂₀). NC는 인간 게놈 내의 부위 수가 계산되지 않았다는 것을 표시한다. 음성 스페이스 길이는 표시된 수의 염기 쌍에 의해 중복되는 표적 gRNA 결합 부위를 지칭한다. 표 4 (B)는 13 내지 19 bp, 또는 22 내지 29 bp의 스페이스 길이 (fCas9의 스페이스 선호도)를 수반한 배향 A에서의 쌍을 이룬 gRNA 결합 부위의 수의 합을 나타낸다 (도 16). 배향 A에서 -8 bp 내지 100 bp의 스페이스 길이, 또는 배향 B에서 4 내지 42 bp의 스페이스 길이 (Cas9 니카제의 스페이스 선호도)를 수반한 쌍을 이룬 gRNA 결합 부위의 수의 합을 나타낸다 (배향 B에서의 4 내지 42 bp는 도 8b, c에 근거한 것이고, 배향 A에서의 -8 bp 내지

100 bp는 기존의 보고 내용^{36,37}에 근거한 것이다).

<표 4>

인간 게놈에서 fCas9 및 Cas9 니카제에 대한 쌍을 이룬 gRNA 표적 부위 존재도

(A)

스페이서 길이(b)	배향 A 에서 쌍을 이룬 gRNA 부위의 수	배향 B 에서 쌍을 이룬 gRNA 부위의 수
-8	6874293	NC
-7	6785996	NC
-6	6984064	NC
-5	7023260	NC
-4	6487302	NC
-3	6401348	NC
-2	6981383	NC
-1	7230098	NC
0	7055143	NC
1	6598582	NC
2	6877046	NC
3	6971447	NC
4	6505614	5542549
5	6098107	5663458
6	6254974	6819289
7	6680118	6061225
8	7687598	5702252
9	6755736	7306646
10	6544849	6387485
11	6918186	6172852
12	6241723	5799496
13	6233385	7092283
14	6298717	7882433
15	6181422	7472725
16	6266909	6294684
17	6647352	6825904
18	6103603	6973590
19	5896092	6349456
20	6000683	5835825
21	5858015	6056352
22	6116108	6531913
23	5991254	6941816
24	6114969	6572849
25	6135119	5671641

(B)

Cas9 변이체	마람직한 스페이서 길이 (bp)	총 부위
fCas9	배향 A 에서 13 내지 19, 또는 22 내지 29	92354891
Cas9 니카제	배향 A 에서 -8 내지 100 배향 B 에서 4 내지 42	953048977

fCas9의 DNA 절단 특이성을 평가하기 위하여, *CLTA*, *EMX*, 및 *VEGF* 게놈 표적 부위의 공지된 Cas9 표적을 벗어난 부위의 변형을 측정하였다^{1,2,6,8}. 표적 부위 및 그의 상응하는 공지된 표적을 벗어난 부위 (표 5)를, fCas9, Cas9 니카제, 또는 야생형 Cas9 및 *CLTA* 부위를 표적화하는 19 bp 간격으로 떨어진 2개의 gRNA, *EMX* 부위를 표적화하는 23 bp 간격으로 떨어진 2개의 gRNA, *VEGF* 부위를 표적화하는 14 bp 간격으로 떨어진 2개의 gRNA, 또는 음성 대조군으로서 무관한 부위 (GFP)를 표적화하는 2개의 gRNA로 처리된 HEK293 세포로부터 단리된 게놈 DNA로

부터 증폭시켰다. 총 11개의 표적을 벗어난 부위를 고 처리량 서열 분석에 의해 분석하였다.

[0238] 계놈의 표적을 벗어난 절단을 탐지하기 위한 고 처리량 서열 분석 방법의 민감도는 각 계놈 표적 부위의 PCR 증폭 내로 유입된 계놈 DNA (gDNA)의 양에 의해 제한된다. 인간 gDNA의 1 ng 샘플은 약 330개 만의 독특한 계놈을 나타내므로, 각 계놈 부위의 약 330개 만의 독특한 카피가 존재한다. 각 계놈 표적에 대한 PCR 증폭을 총 150 ng의 유입 gDNA 상에서 수행하였는데, 이는 기껏해야 50,000개의 독특한 gDNA 카피로부터 유래된 앰플리콘을 제공한다. 따라서, 고 처리량 서열 분석 검정은 50,000개 중의 1개 미만의 빈도, 또는 0.002%로 발생하는 희귀한 계놈 변형 현상을 탐지할 수 없다.

[0239] <표 5>

[0240] EMX, VEGF, 및 CLTA에서 Cas9 표적 부위의 공지된 표적을 벗어난 기질. EMX, VEGF, 및 CLTA의 계놈의 표적에 적중한 부위와 표적을 벗어난 부위의 목록이 제시되어 있는데, 표적에 적중한 것로부터의 돌연변이가 소문자의 볼드체이다. PAM은 대문자의 볼드체로 제시된다.

	계놈 표적 부위
	GAGTCCGAGCAGAAGAAGAAGGG
EMX_On	(서열 190) GAG g CCGAGCAGAAGAA ag ACGG
EMX_Off1	(서열 191) GAGTCC t AGCAG g AGAAGAAGaG
EMX_Off2	(서열 192) GAGTC ta AGCAGAAGAAGAAGaG
EMX_Off3	(서열 193) GAGT ta GAGCAGAAGAAGAAAGG
EMX_Off4	(서열 194)
	GGGTGGGGGAGTTTGCTCCTGG
VEG_On	(서열 195) GGaTGGaGGGAGTTTGCTCCTGG
VEG_Off1	(서열 196) GGGaGGG t GGAGTTTGCTCCTGG
VEG_Off2	(서열 197) c GG g GGaGGGAGTTTGCTCCTGG
VEG_Off3	(서열 198) GGG ga GGGGaAGTTTGCTCCTGG
VEG_Off4	(서열 199)
	GCAGATGTAGTGTTCACACAGGG
CLT2_On	(서열 200) a CAaATGTAGT a TTTCACACAGGG
CLT2_Off1	(서열 201) c CAGATGTAGT a TT c CCACAGGG
CLT2_Off2	(서열 202) ct AGATGaAGTG c TTCCACATGG
CLT2_Off3	(서열 203)

[0241]

[0242] 잠재적 계놈의 표적을 벗어난 부위에서 2개 이상의 염기 쌍의 삽입 또는 결실을 함유하고 대조군 gRNA-처리된 샘플과 비교해서 표적 gRNA-처리된 샘플에서 상당히 더 큰 수로 존재하는 (P 값 < 0.005, 피셔의 직접 확률 시험) 서열이 Cas9 뉴클레아제-유도된 계놈 변형물로 간주되었다. 검정된 11개의 표적을 벗어난 부위 중 10개에 대하여, fCas9는 검정의 민감도 한계치 내에서 어떠한 탐지 가능한 계놈의 표적을 벗어난 변형을 초래하지 못하였는데 (< 0.002%,), 이는 5% 내지 10%의 실질적인 표적에 적중한 변형 효율을 명확하게 보여준다 (도 8d-f 및 표 3). fCas9-변형된 *VEGF* 표적에 적중한 서열을 상세히 검사한 결과 (도 18a), 아연 핑거 또는 TALE DNA-결합성 도메인과 융합된 *FokI* 뉴클레아제 도메인의 효과와 유사한, 2개의 표적 결합 부위 사이의 DNA 스페이스에서 발생하는 절단과 일치하는 2개 내지 수십 개의 염기 쌍 범위의 결실이 우세한 것으로 밝혀졌다³¹.

[0243] 이와는 달리, 계놈의 표적을 벗어난 DNA 절단이, 검정된 11개 모든 부위에서 야생형 Cas9에 대해 관찰되었다. 표적을 벗어난 fCas9 활성화에 대한 상한으로서 상기 검정의 탐지 한계치를 이용하여, fCas9가 야생형 Cas9 뉴클

레아제 보다 훨씬 더 낮은 표적을 벗어난 변형 비율을 갖는 것으로 계산하였다. 야생형 Cas9 뉴클레아제에 의해 변형된 11개의 표적을 벗어난 부위에서, fCas9는 야생형 Cas9 보다 적어도 140배 더 높은 표적에 적중한:표적을 벗어난 변형 비를 산출시켰다 (도 8d-f).

[0244] 기존의 보고 내용과 일치하게^{5,6,8}, Cas9 니카제는 또한, 야생형 Cas9와 비교해서 실질적으로 더 적은 수의 표적을 벗어난 변형 현상을 유도시켰다 (탐지 가능한 비율로 변형된 11개의 표적을 벗어난 부위 중 1개). 초기 고처리량 서열 분석 검정 결과, *VEGF* 표적을 벗어난 부위 1에서 서열의 0.024%에서 Cas9 니카제에 의해 유도된 유의적 (P 값 $< 10^{-3}$, 피셔의 직접 확률 시험) 변형이 밝혀졌다. Cas9 니카제의 경우에는 유사한 *VEGF* 표적에 적중한 변형 효율이 12.3%이고 fCas9의 경우에는 10.4%임에도 불구하고, 상기 게놈의 표적을 벗어난 부위는 fCas9에 의해 변형되지 않았다 (도 8f 및 표 3C). Cas9 니카제-유도된 변형 수준이 탐지 한계치의 한 자릿수 이내였고, fCas9 변형 수준은 탐지하지 못하였기 때문에, 이 실험을 보다 큰 유입 DNA 샘플과 보다 많은 수의 서열 판독치 (초기 실험 및 반복된 실험 각각에 대하여 150 대 600 ng 게놈 DNA 및 $> 8 \times 10^5$ 개 판독치 대 $> 23 \times 10^5$ 개 판독치)를 이용하여 반복하여, Cas9 니카제 또는 fCas9에 의한 상기 부위에서의 표적을 벗어난 절단을 탐지하였다. 이러한 심층 문의로부터, Cas9 니카제와 fCas9 둘 다 *VEGF* 표적을 벗어난 부위 1을 상당히 변형시키는 것으로 관찰되었다 (P 값 $< 10^{-5}$, 피셔의 직접 확률 시험) (도 8g, 표 3D, 도 18b). *VEGF* 표적을 벗어난 부위 1에서의 변형 비율을 문의하는 양 실험의 경우, fCas9는 Cas9 니카제 보다 더 큰 표적에 적중한:표적을 벗어난 DNA 변형 비를 나타내었다 (fCas9의 경우 $> 5,150$ 및 $1,650$ 대 Cas9 니카제의 경우 510 및 $1,230$, 도 8g).

[0245] *VEGF* 표적을 벗어난 부위 1의 어느 한쪽에는, *VEGF* 표적에 적중한 서열의 2개 절반-부위 중 어느 하나로부터 6개 이하의 돌연변이를 수반한 다른 부위가 존재하지 않는다. 하나의 gRNA (V2)의 첫 번째 11개 염기는 *VEGF* 표적을 벗어난 부위 1 내의 표준 Cas9:gRNA 결합으로부터 풀려난 단일 가닥 DNA와 혼성화될 수도 있다 (도 18c). 이러한 gRNA:DNA 혼성화를 통하여, 제2의 Cas9 니카제 또는 fCas9가 동원되어 상기 표적을 벗어난 부위를 극히 낮은 수준이긴 하지만, 탐지 가능한 수준으로 변형시킬 수 있다는 것이 가능하다. 신중한 gRNA 쌍 설계가 이러한 표적을 벗어난 DNA 절단의 잠재적 방식을 없앨 수 있었는데, 이는 *VEGF* 표적을 벗어난 부위 1이 특정 쌍의 제2의 gRNA를 수반한 11개의 연속되는 잠재적 염기 쌍을 형성할 수 있는 그의 능력에 있어서 고도로 특이하기 때문이다. 일반적으로, fCas9는 *FokI* 뉴클레아제 도메인을 이량체화하고 활성화시키는 데 요구되는 어느 인접한 제2 결합 부위의 부재로 인해, 시험된 게놈의 표적을 벗어난 부위를 변형시킬 수 없었다.

[0246] 본 연구에서 개발된 최적화 *FokI*-dCas9 융합 구조는 표적화된 5개의 모든 게놈 유전자 자리를 변형시켰는데, 이는 fCas9를 이용한 변형이 야생형 Cas9를 이용한 경우 보다 다소 덜 유효하긴 하였지만, 인간 세포에서 게놈 변형을 유도시키기 위해 fCas9를 이용하는 일반론을 입증해준다. fCas9를 사용하는 것은 간단한 일인데, 이는 단지, PAM 서열이 적당한 간격과 배향으로 존재해야만 하고, 야생형 Cas9 또는 Cas9 니카제와 동일한 gRNA를 사용한다. 야생형 Cas9 보다 >140 배 더 낮은, 상기와 같이 관찰된 fCas9의 낮은 표적을 벗어난:표적에 적중한 변형 비는 이량체성 *FokI*의 별개의 작용 방식 [여기서, 2개의 DNA 부위가 지정된 거리 (본원에서는, 약 15 bp 또는 약 25 bp 떨어짐)에서 특이적 절반-부위 배향으로 2개의 *FokI* 도메인에 의해 동시에 점유되는 경우에만 DNA 절단이 진행된다]으로부터 비롯되는 것으로 예상된다. 이로써 생성되는, 특이하게 낮은 fCas9의 표적을 벗어난 활성은 극히 고도의 표적 특이성을 필요로 하는 Cas9:gRNA-기반 기술의 적용, 예컨대 인간 세포의 생체외 또는 생체내 치료적 변형을 가능하게 해준다.

[0247] 참고문헌

1. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).

[0248]

2. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
3. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
4. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* **41**, 9584–9592 (2013).
5. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2013).
6. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).
7. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
8. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* (2014).
doi:10.1038/nbt.2808
9. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
10. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
11. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109**, E2579–E2586 (2012).
12. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
13. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84–87 (2013).
14. Perez, E. E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26**, 808–816 (2008).
15. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471–e00471 (2013).
16. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).

[0249]

17. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).
18. Ramirez, C. L. *et al.* Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.* **40**, 5560–5568 (2012).
19. Wang, J. *et al.* Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res.* **22**, 1316–1326 (2012).
20. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
21. Vanamee, É. S., Santagata, S. & Aggarwal, A. K. FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.* **309**, 69–78 (2001).
22. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
23. Pattanayak, V., Ramirez, C. L., Joung, J. K. & Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* **8**, 765–770 (2011).
24. Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat. Methods* (2014).
doi:10.1038/nmeth.2845
25. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* (2014). doi:10.1016/j.cell.2014.02.001
26. Jinek, M. *et al.* Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* (2014). doi:10.1126/science.1247997
27. Schellenberger, V. *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* **27**, 1186–1190 (2009).
28. Guschin, D. Y. *et al.* in *Eng. Zinc Finger Proteins* (Mackay, J. P. & Segal, D. J.) **649**, 247–256 (Humana Press, 2010).
29. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
30. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
31. Kim, Y., Kweon, J. & Kim, J.-S. TALENs and ZFNs are associated with different mutation signatures. *Nat. Methods* **10**, 185–185 (2013).
32. Shcherbakova, D. M. & Verkhusha, V. V. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nat. Methods* **10**, 751–754 (2013).

[0250]

33. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
34. Sander, J. D. *et al.* In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. *Nucleic Acids Res.* **41**, e181–e181 (2013).
35. Schellenberger, V. *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* **27**, 1186–1190 (2009).
36. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
37. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).
38. Yan, T. *et al.* PatMatch: a program for finding patterns in peptide and nucleotide sequences. *Nucleic Acids Res.* **33**, W262–W266 (2005).
39. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

[0251]

[0252] **실시예 2: Cas9 변이체-매개된 불활성화를 위하여 CCR5를 표적화함**

[0253] 강력한 연구 도구를 제공하는 것 이외에도, 부위-특이적 뉴클레아제는 또한, 유전자 요법 작용제로서의 잠재력을 지니고 있고, 부위-특이적 아연 핑거 엔도뉴클레아제는 최근에 임상 시험에 들어갔다: 항-HIV 치료적 접근 방식의 일부로서 인간 CCR-5 대립 유전자를 표적화하는 CCR5-2246.

[0254] 유사한 접근 방식에서, 본 개시내용의 본 발명의 Cas9 변이체를 사용하여, 예를 들어 특정 대상체로부터 수득된 자기 유래의 T 세포에서 CCR5를 불활성화시킬 수 있는데, 일단 Cas9 변이체에 의해 변형되면, 이를 HIV 감염의 치료 또는 예방을 위하여 상기 대상체 내로 재도입한다.

[0255] 본 실시예에서는, CCR5 유전자를 특정 대상체로부터 수득된 T 세포에서 표적화한다. CCR5 단백질은 HIV의 특성의 공통 유형이 T 세포와 결합하여 이러한 세포 내로 유입됨으로써 이들 세포를 감염시키는 데 필요하다. T 세포는 신체가 HIV와 싸우기 위해 사용되는 백혈구 중 하나이다.

[0256] 일부 사람은 그들의 T 세포 상에서 CCR5 발현이 결여되도록 태어나고, 건강한 상태를 유지하다가 HIV으로의 감염에 대해 내성이 있다. 다른 사람들은 그들의 T 세포 상에서의 CCR5의 발현이 낮고, 그들의 HIV 질환은 덜 중증이며 발병 속도도 더 느리다 (AIDS).

[0257] T 세포 상에서 CCR5 단백질을 결실시키기 위해서, 많은 수의 T 세포를 특정 대상체로부터 단리한다. 이어서, CCR5를 불활성화시킬 수 있는 Cas9 변이체 (예를 들어, fCas9) 및 gRNA를, 바이러스성 벡터, 예를 들어 아테노 바이러스성 벡터를 이용하여 상기 단리된 T 세포로 전달한다. 적합한 Cas9 변이체의 예는 본원에 제공된 본 발명의 융합 단백질을 포함한다. CCR5 대립 유전자를 표적화하는 gRNA에 적합한 표적 서열의 예는 도 19에 기재된 것, 예를 들어 서열 303 내지 310 및 312 내지 317을 포함한다. Cas9 변이체 및 gRNA를 발현할 수 있는 바이러스성 벡터(들)를 상기 단리된 T 세포에 가하여 CCR5 단백질을 녹아웃시킨다. 이러한 T 세포를 대상체에게 재투입하게 되면, 최소한의 아테노바이러스 또는 Cas9 변이체 단백질이 존재하게 된다. 그러나, 대상체에게 투여된 T 세포 상에서의 CCR5 단백질의 제거는 영구적이다. 이어서, 상기 세포를 HIV/AIDS의 치료 또는 예방을 위하여 대상체에게 재도입한다.

[0258] **실시예 3: Cas9-재조합효소 융합 단백질**

[0259] 예시되는 Cas9-재조합효소 융합 단백질이 다음에 제공된다:

[0260] *dCas9-NLS-GGS3 링커-Tn3*

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
DLNPDNSDVKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYA
DLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVK
VMGRHKPENIVEMARENQTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQL
QNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSP
QVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV
AKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD MAPKKR KVG
IHRGVPGSGSGSGSMALFGYARVSTSQSLDLQVRALKDAGVKANRIFTDKASGS
STDREGLDLLRMKVKEGDVILVKKLDRLGRDTADMLQLIKEFDAQGVAVRFIDDGIS
TDSYIGLMFVTILSAVAQAERRRILERTNEGRQA AKLKGIKFGRRR (서열 328)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0261]

[0262] *NLS-dCas9-GGS3 링커-Tn3*

MAPKKR KVGIHRGVPMMDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
HRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL
ALAHMIKFRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFEENPINASGVDAKAILS
ARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYD
DDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTR
IPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDI
VLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IIKKGILQTVKVVDELVKVMGRHKPENIVEMARENQTTQKGQKNSRERMKRIIEGK
ELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVAIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
CLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVY
DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVFAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKE
VKKDLIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIRE
QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDL
SQLGGD GSGSGSGSMALFGYARVSTSQSLDLQVRALKDAGVKANRIFTDKASGS
TDREGLDLLRMKVKEGDVILVKKLDRLGRDTADMLQLIKEFDAQGVAVRFIDDGIST
DSYIGLMFVTILSAVAQAERRRILERTNEGRQA AKLKGIKFGRRR (서열 329)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0263]

[0264]

Tn3-GGS3 링커-dCas9-NLS

MALFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVK
EGDVILVKKLDRLGRDTADMLQLIKEFDAQGVAVRFIDDGISTDSYIGLMFVTILSAV
AQAERRRILERTNEGRQAACKLKGIFGRRRGGS~~GGSGGS~~GSMDDKKYSIGLAIGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR
RKNRICYLQEFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQ
TYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDIL
RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI
DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRK
PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT
HDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQK
AQVSGQGDSLHEHIANLAGSPAIKKILQTVKVVDLVKVMGRHKPENIVIAMAREN
QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY
VDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILD
SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKEL
LGITIMERSSEKPNIDFLEAKGYKEVKKDLIILPKYSLFELENGRKRMLASAGELQK
GNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRV
ILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
TKEVLDATLIHQ~~SITGLYETRIDLSQLGGD~~MAPKKRKRKVGIHRGVP (서열 330)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0265]

[0266]

NLS-Tn3-GGS3 링커-dCas9

MAPKKKRKVGIHRGVPMALFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKAS
 GSSTDREGLDLLRMKVKEGDVILVKKLDRLGRDTADMLQLIKEFDAQGVAVRFIDD
 GISTDSYIGLMFVTILSAVAQAERRRILERTNEGRQA AKLKGIFGRRRGSGSGSGGS
 MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
 TAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHE
 RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG
 DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
 GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYA
 DLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
 KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
 TFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
 WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF DKNLPNEKVLPKHSLLYEYFTV
 YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
 KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRN
 FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDDELVK
 VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK
 NRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
 RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKV
 REINNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVYDV RKMIKSEQEIGK
 ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SM
 PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
 AKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
 LENGKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
 HKHYLDEIIEQISEFSKR VILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGA
 PAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGD (서열 331)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0267]

[0268]

dCas9-NLS-GGS3 링커-Hin

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAHMIKFRGHFLIEG
DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLMS
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVLADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDTLIHQSTGLYETRIDLSQLGGDMAPKKKRKVG
IHRGVPGGSGGSGGSMATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRP
GLKRALKYVNKGDTLVVWKLDRLGRSVKNLVALISELHERGAHFHSLTDSIDTSSA
MGRFFFFYVMSALAEMERELIVERTLAGLAAARAQGRLG (서열 332)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0269]

[0270]

NLS-dCas9-GGS3링커-Hin

MAPKKKRKVGIHRGVPMDDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF
HRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIY
ALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS
RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLNDLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPKEYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDI
VLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI
KKGILQTVKVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIK
ELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGSELDAKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVY
DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIRE
QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQ
LGGDGGSGGSGGSMATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPG
LKRALKYVNKGDTLVVWKLDRLGSRVKNLVALISELHERGAHFHSLTDSIDTSSAM
GRFFFYVMSALAEMERELIVERTLAGLAAARAQGRLG (서열 333)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0271]

[0272]

Hin-GGS3 링커-dCas9-NLS

MATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPGLKRALKYVNBKGD
LVVWKLDRGRSVKNLVALISELHERGAHFHSLTDSIDTSSAMGRFFFYVMSALAE
MERELIVERTLAGLAAARAQGRLLGGSGGSGGSMDDKYSIGLAIGTNSVGWAVITD
EYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
RKKLVDDSTDKADLRILIYLAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQL
FEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKS
NFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE
DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKG
ASAQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAFLSG
EQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQK
GQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGITALI
KKYPKLESEFVYGDYKVYDVRKMIKSEKQEGKATKAYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIM
ERSSEKPNIDFLEAKGYKEVKKDLIKLPKYSLFELENGRKRMLASAGELQKGNELA
LPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSIITGLYETRIDLSQLGGDMAPKKKRKVGIHRGVP (서열 334)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0273]

[0274]

NLS-Hin-GGS3 링커-dCas9

MAPKKKRKVGIHRGVPMATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIAN
 RPGLKRALKYVNVKGDTLVVWKLDRGRSVKNLVALISELHERGAHFHSLTDSIDTSS
 AMGRFFFYVMSALAEMERELIVERTLAGLAAARAQGRLGGS~~GGSGGS~~GSMDDKKYSI
 GLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
 KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNI
 VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS
 DVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL
 FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLLAAK
 NLSDAILLSDLIRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD
 QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP
 HQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
 EETITPWNFEFVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTK
 VKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV
 EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHL
 FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIH
 DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKGILQTVKVVDELVKVMGRHK
 PENIVIEMARENQTTQKGQKNSRERMKRIEELGSGILKEHPVENTQLQNEKLYL
 YYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
 VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR
 QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYH
 HAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFF
 YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVK
 KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVKG
 KSKKLKSVKELLGITIMERSSEKPNIDFLEAKGYKEVKKDLIKLPKYSLELENGRK
 RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD
 EIIEQISEFSKRVILADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAFKYF
 DTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD (서열 335)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0275]

[0276]

dCas9-NLS-GGS3링커-Gin

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEG
DLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRLENLIAQLP
GEKKNGLFGNLIASLGLTPNFKS NFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV KLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSSF EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDMAPKKKRKVG
IHRGVPGSGSGSGSMLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTD
RPGLKRALKRLQKGD TLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSP
MGRFFFYVMGALAEMERELIERTMAGIAAARNKGRRFGRPPKSG (서열 336)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0277]

[0278] *NLS-dCas9-GGS3링커-Gin*

MAPKKKRRKVGIHRGVPMDDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDDSFF
HRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRILIY
ALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS
RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLNDLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKNLREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDI
VLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDQKS
GKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI
KKGILQTVKVVDLVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIIEGK
ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVRINNYYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVY
DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKE
VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK
GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIRE
QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDL SQ
LGGDGGSGGSGGSM LIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTRDRP
GLKRALKRLQKGD TLVVWKLDR LGRSMKHLISLVGELRERGINFRSLTDSIDTSSPM
GRFFFYVMGALAEMERELIERTMAGIAAARNKGRRFGRPPKSG (서열 337)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0279]

[0280]

Gin-GGS3 링커-dCas9-NLS

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD
 TLVWVKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFFYVMGALAE
 MERELIERTMAGIAAARNKGRRFGRPPKSGGSGGSGGSMDDKKYSIGLAIGTNSVG
 WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR
 RKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
 PTIYHLRKKLVDDSTDKADLRILIYLAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQ
 TYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLT
 PNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDIL
 RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI
 DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
 LRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
 VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRK
 PAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT
 HDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKYAHLFDDKVMKQLK
 RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQK
 AQVSGQGDSLHEHIANLAGSPAIKKILQTVKVVDELVKVMGRHKPENIVIAMAREN
 QTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMY
 VDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTSDKNRGKSDNVPSEEVVKKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKHVAQILD
 SRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAV
 VGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEI
 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSK
 ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
 LGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIKLPKYSLEFENGRKRMLASAGELQK
 GNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKR
 ILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
 TKEVL DATLIHQ SITGLYETRIDLSQLGGD MAPKKKRKVGIHRGVP (서열 338)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0281]

[0282] *NLS-Gin-GGS3링커-dCas9*

MAPKKKRKVGIHRGVPMLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRT
DRPGLKRALKRLQKGDTLVVWKLDRLGSMKHLISLVGELRERGINFRSLTDSIDTS
SPMGRFFFFYVMGALAEMERELIERTMAGIAAARNKGRRFGRPPKSGGGSGGSGGS
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE
TAEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLEESFLVEEDKKHE
RHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLLIYLALAHMIKFRGHFLIEG
DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPE
KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNNAVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD (서열 339)

(밑줄: 핵 국제화 신호; 볼드체: 링커 서열)

[0283]

[0284] 실시예 4: Cas9-재조합효소 융합 단백질을 이용한 상동 재조합에 의한 마커 유전자의 도입

[0285]

숙주 세포 유전자의 게놈 서열이 양옆에 위치하고 있는 그린 형광성 단백질 (GFP) 마커 유전자를 수반하는 벡터를, 이러한 GFP 마커와 재조합되는 게놈 유전자 자리 및 GFP 마커 유전자를 표적으로 하는 4개의 적당하게 설계된 gRNA 및 dCas9-재조합효소 융합 단백질 (서열 328 내지 339 중 어느 하나)을 코딩하는 발현 구조물과 함께 세포 내로 도입한다. 4개의 dCas9-재조합효소 융합 단백질은, gRNA의 결합을 통하여 GFP 마커 유전자와 함께 상기 게놈 유전자 자리에서 조정된다 (도 5b). 상기 융합 단백질의 4개의 재조합효소 도메인은 사랑채화되고, 이러한 융합 단백질의 재조합효소 도메인의 재조합효소 활성으로 인해, 상기 게놈 유전자 자리와 마커 유전자 간에 재조합이 발생됨으로써, 마커 유전자가 게놈 유전자 자리 내로 도입된다. 상기 마커 유전자의 도입은 GFP 발현 및/또는 PCR에 의해 확증된다.

[0286]

등가 및 범위

[0287]

관련 기술분야의 통상의 기술자는 단지 통상적인 실험을 이용하여, 본원에 기재된 본 발명의 구체적 실시양태에 대한 많은 등가내용을 인식하거나 또는 확인할 수 있을 것이다. 본 발명의 범위는 상기 설명으로 제한되지 않고, 오히려 첨부된 청구범위에 제시된 바와 같고 같다.

[0288]

청구범위에서, 단수 형태는 달리 표시되지 않거나 또는 문맥상 달리 명백하지 않는 한 하나 이상을 의미한다. 특정 군의 하나 이상의 구성원들 간의 "또는"을 포함하는 청구범위 또는 설명은 이러한 군 구성원 중 하나, 2개 이상, 또는 모두가 달리 표시되지 않거나 또는 문맥상 달리 명백하지 않는 한은 소정의 생성물 또는 공정에 존재하거나, 이용되거나 또는 달리 관련되는 경우에 만족되는 것으로 간주된다. 본 발명은 상기 군의 바로 하나의 구성원이 소정의 생성물 또는 공정에 존재하거나, 이용되거나 또는 달리 관련되는 실시양태를 포함한다. 본 발명은 또한, 상기 군 구성원의 2개 이상 또는 모두가 소정의 생성물 또는 공정에 존재하거나, 이용되거나 또는 달리 관련되는 실시양태를 포함한다.

[0289]

더욱이, 본 발명이 청구항들 중 하나 이상으로부터, 또는 상기 설명의 관련 부분으로부터의 하나 이상의 제한, 요소, 절, 기술적 용어 등이 또 다른 청구항 내로 도입되는 모든 변동, 조합 및 순열을 포괄하는 것으로 이해되어야 한다. 예를 들어, 또 다른 청구항에 종속항인 어떠한 청구항도 동일한 기본 청구항에 종속적인 다른 청구

항에서 발견된 하나 이상의 제한을 포함하도록 변형될 수 있다. 더욱이, 청구항이 조성물을 인용하는 경우, 달리 표시되지 않는 한 또는 모순이나 불일치가 발생할 것으로 관련 기술분야의 통상의 기술자에게 명백하지 않는 한은, 본원에 개시된 목적 중 어느 것을 위하여 이러한 조성물을 사용하는 방법이 포함되고, 본원에 개시된 제조 방법 또는 관련 기술분야에 공지된 다른 방법 중 어느 것에 따라서 상기 조성물을 제조하는 방법이 포함되는 것으로 이해되어야 한다.

[0290] 요소들이 목록으로서, 예를 들어 마쿠쉬(Markush) 군 형식으로 제시되는 경우에는, 이러한 요소들의 각 아군이 또한 개시되고, 어떠한 요소(들)도 이러한 군으로부터 제거될 수 있는 것으로 이해되어야 한다. 또한, 용어 "포함하는"은 부가의 요소 또는 단계의 봉입을 허용하고 개방하기 위한 것이란 사실에 주목해야 한다. 일반적으로, 본 발명 또는 본 발명의 측면이 특별한 요소, 특징, 단계 등을 포함하는 것으로 지칭되는 경우, 본 발명 또는 본 발명의 측면의 특정 실시양태는 이러한 요소, 특징, 단계 등으로 이루어지거나 또는 본질적으로 이루어지는 것으로 이해되어야 한다. 편의상, 이들 실시양태는 본원에서 그런 말로 구체적으로 제시되지 않았다. 따라서, 하나 이상의 요소, 특징, 단계 등을 포함하는 본 발명의 각 실시양태에 대하여, 본 발명은 또한, 이러한 요소, 특징, 단계 등으로 이루어지거나 본질적으로 이루어지는 실시양태를 제공한다.

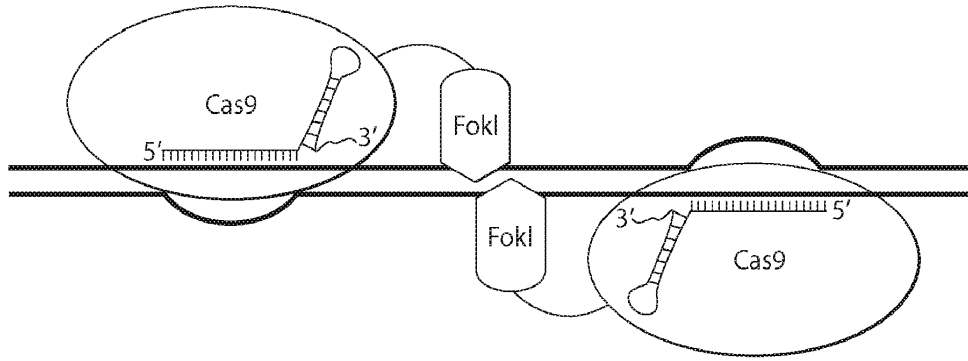
[0291] 범위가 제공되는 경우에는, 종말점이 포함된다. 더욱이, 달리 표시되지 않거나 또는 관련 기술분야의 통상의 기술자의 맥락 및/또는 이해로부터 달리 명백하지 않는 한, 범위로서 표현되는 값은 본 발명의 상이한 실시양태에서 언급된 범위 내의 모든 구체적 값 내지 문맥상 달리 명확하게 지시하지 않는 한은 이러한 범위의 하한치의 단위의 1/10을 추정할 수 있는 것으로 이해되어야 한다. 또한, 달리 표시되지 않거나 또는 관련 기술분야의 통상의 기술자의 맥락 및/또는 이해로부터 달리 명백하지 않는 한, 범위로서 표현되는 값은 소정의 범위 내의 모든 아범위를 추정할 수 있는 것으로 이해되어야 하는데, 여기서 이러한 아범위의 종말점은 상기 범위의 하한치의 단위의 1/10과 동일한 정확도로 표현된다.

[0292] 또한, 본 발명의 특별한 어떠한 실시양태도 어느 하나 이상의 청구항으로부터 명백하게 배제될 수 있는 것으로 이해되어야 한다. 범위가 제공되는 경우, 이러한 범위 내의 어떠한 값도 어느 하나 이상의 청구항으로부터 명백하게 배제될 수 있다. 본 발명의 조성물 및/또는 방법의 모든 실시양태, 요소, 특징, 적용 또는 측면이 어느 하나 이상의 청구항으로부터 배제될 수 있다. 간결성을 위해, 하나 이상의 요소, 특징, 목적 또는 측면이 배제되는 모든 실시양태가 본원에 명백하게 제시되지 않는다.

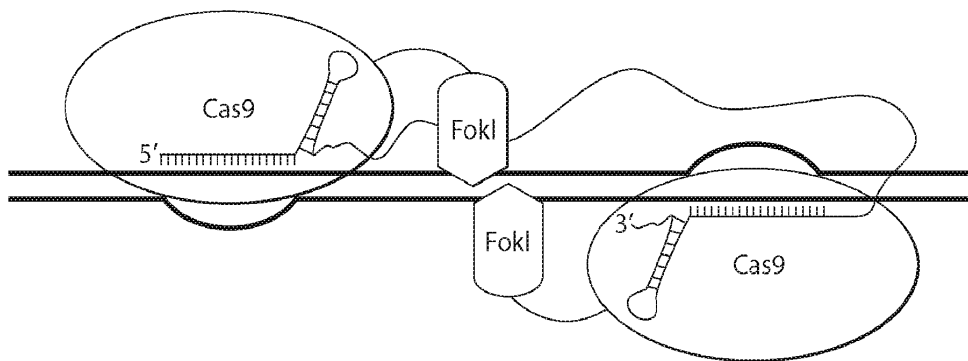
[0293] 상기 열거된 항목을 포함한, 본원에 언급된 모든 공개 공보, 특허 및 서열 데이터베이스 목록은 마치 각각의 개별 공개 공보 또는 특허가 특이적이고도 개별적으로 참조로 포함되는 것으로 표시되었던 것처럼 그들의 전문이 본원에 참조로 포함된다. 분쟁이 있는 경우, 본원의 모든 정의를 포함한 본 출원이 우선할 것이다.

도면

도면1

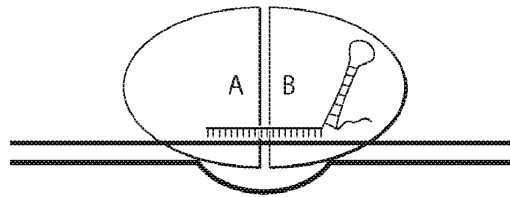


도 1A

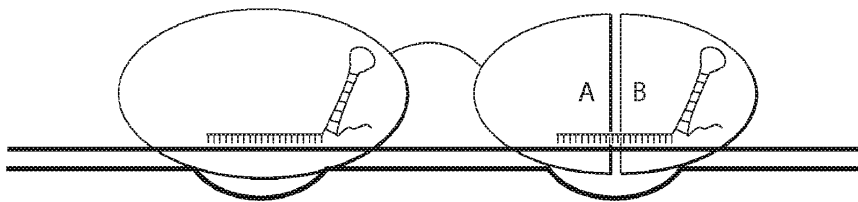


도 1B

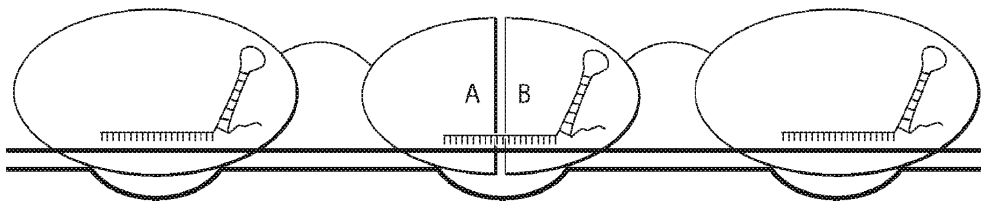
도면2



도 2A

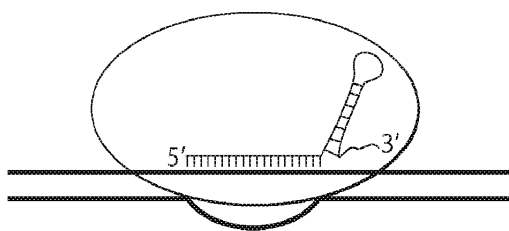


도 2B

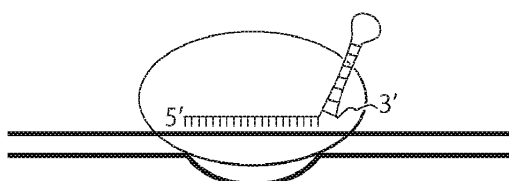


도 2C

도면3

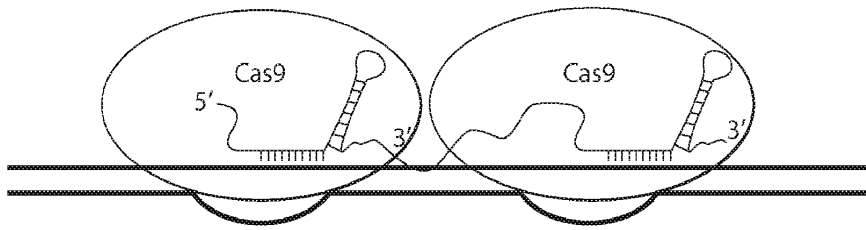


완전한 길이의 Cas9

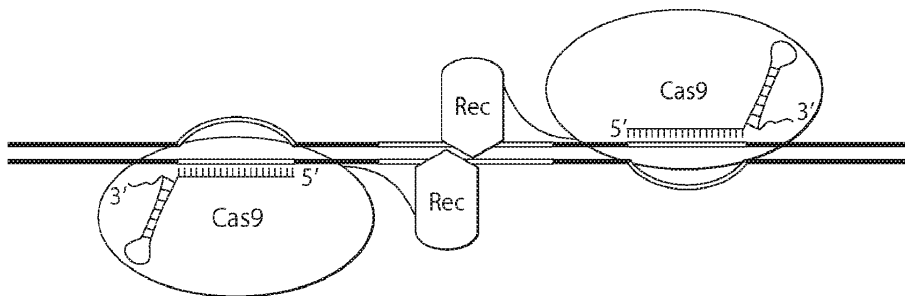


최소화 Cas9

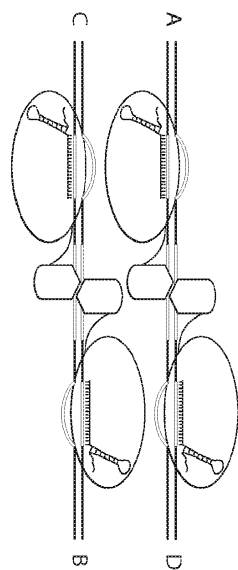
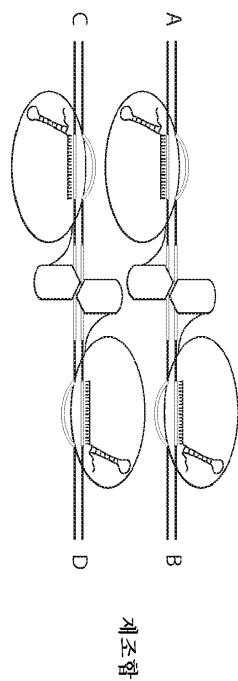
도면4



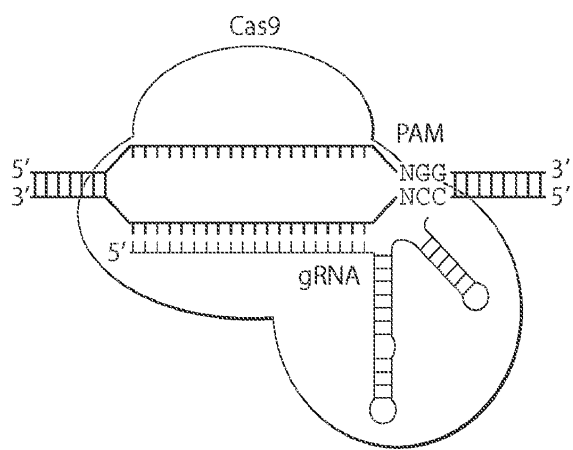
도면5a



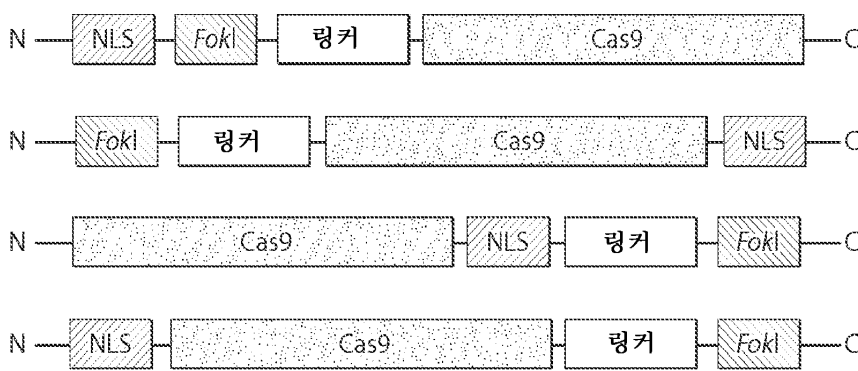
도면5b



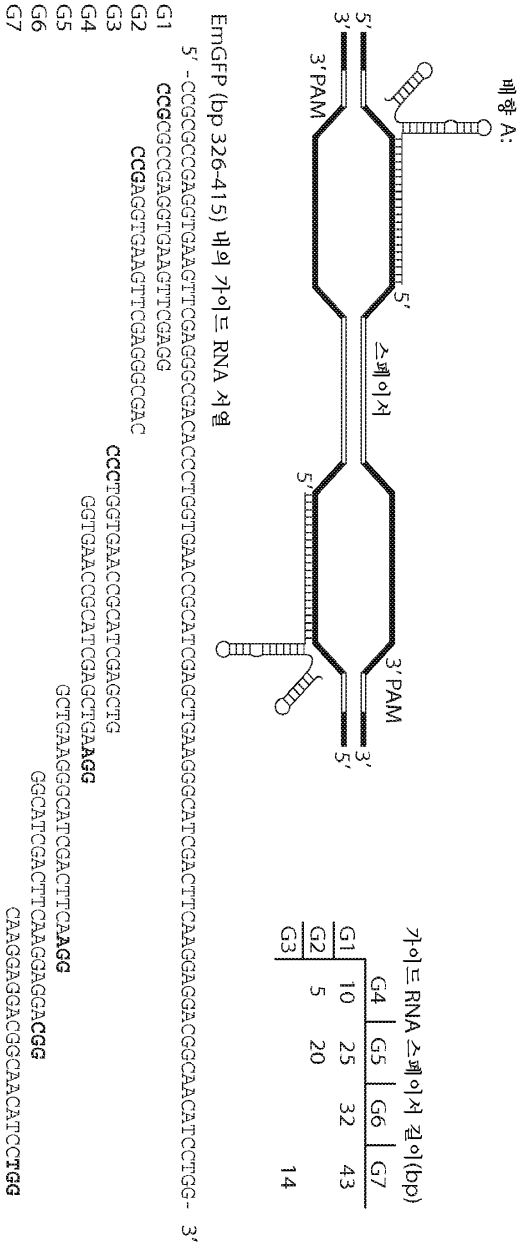
도면6a



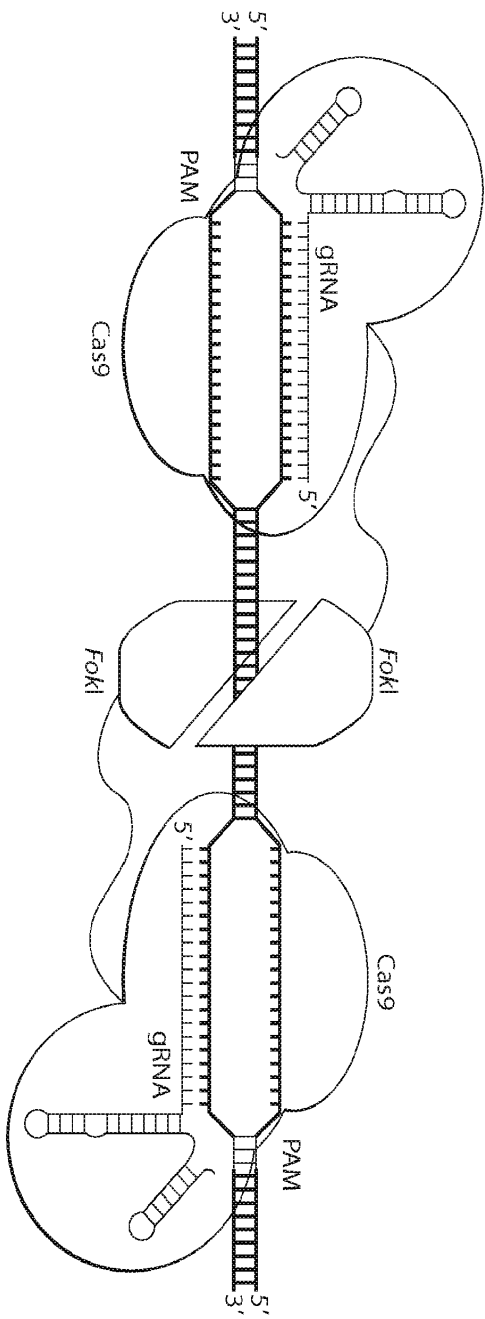
도면6b



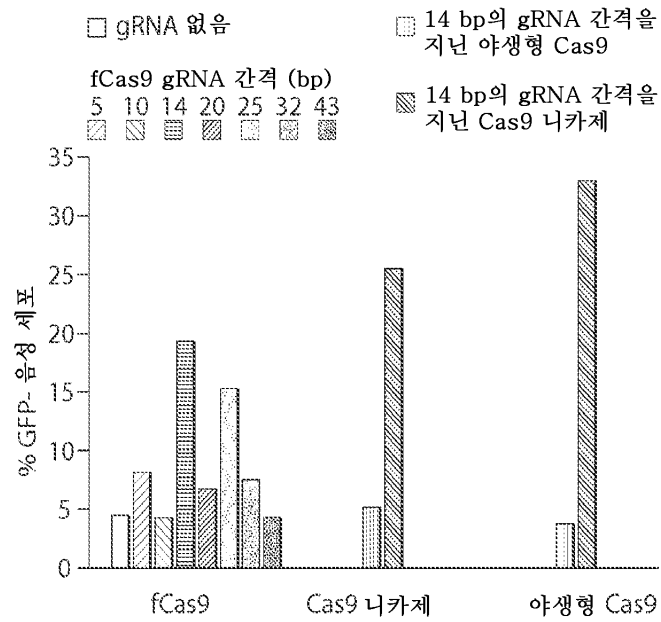
도면6c



도면6d

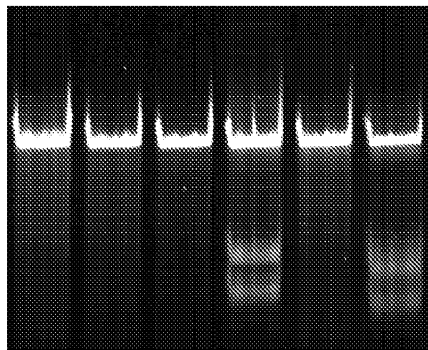


도면 7a



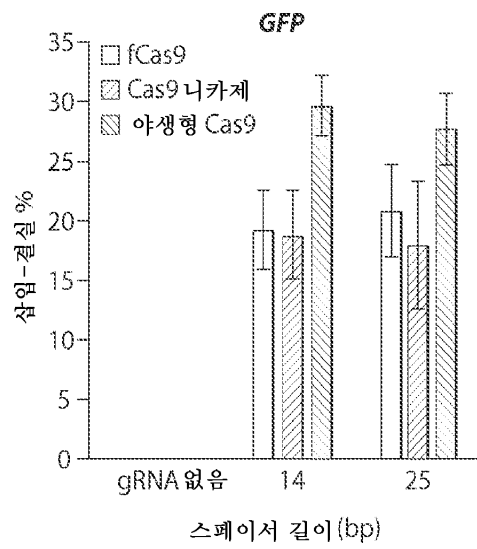
도면 7b

fCas9:	+	-	-	+	-	-
Cas9 니카제 :	-	-	-	-	-	+
gRNA G3:	-	+	-	+	+	+
gRNA G7:	-	-	+	+	+	+

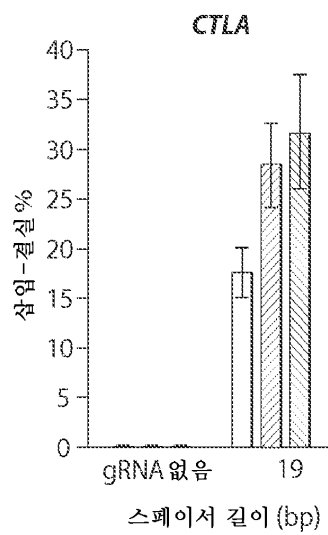


삽입-결실%:	16	21

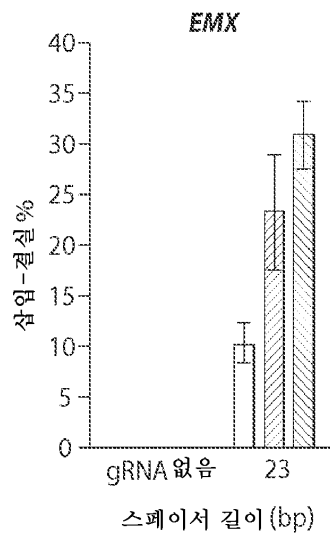
도면7c



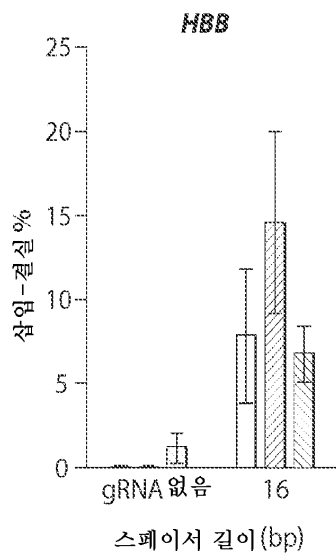
도면7d



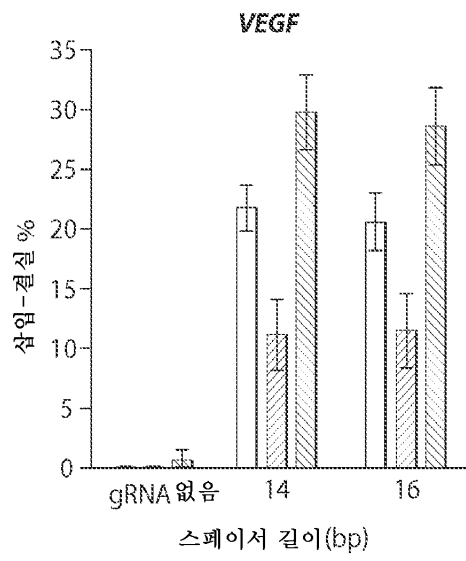
도면7e



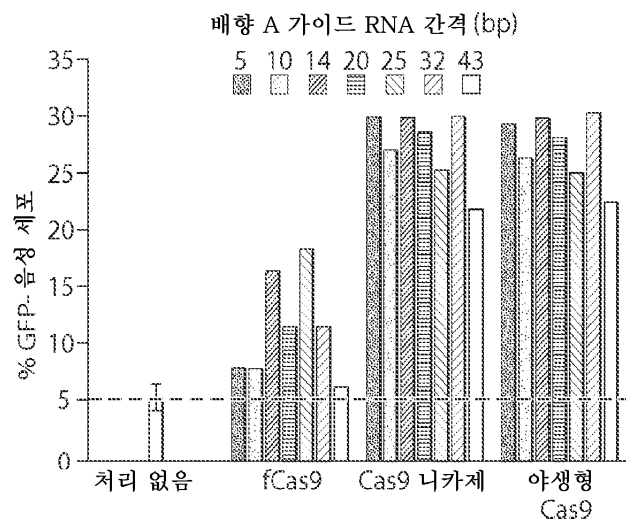
도면7f



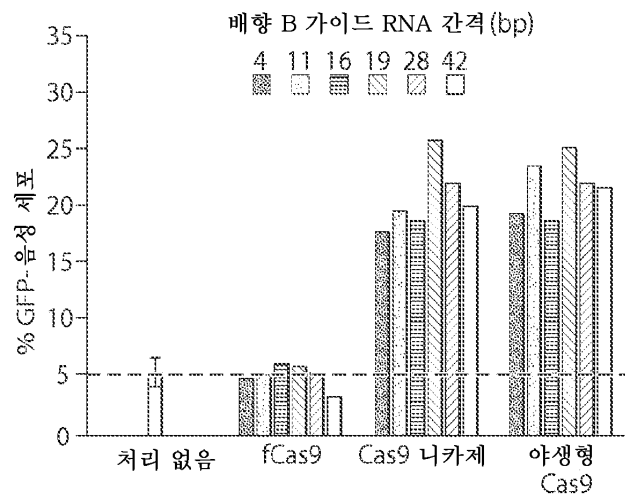
도면7g



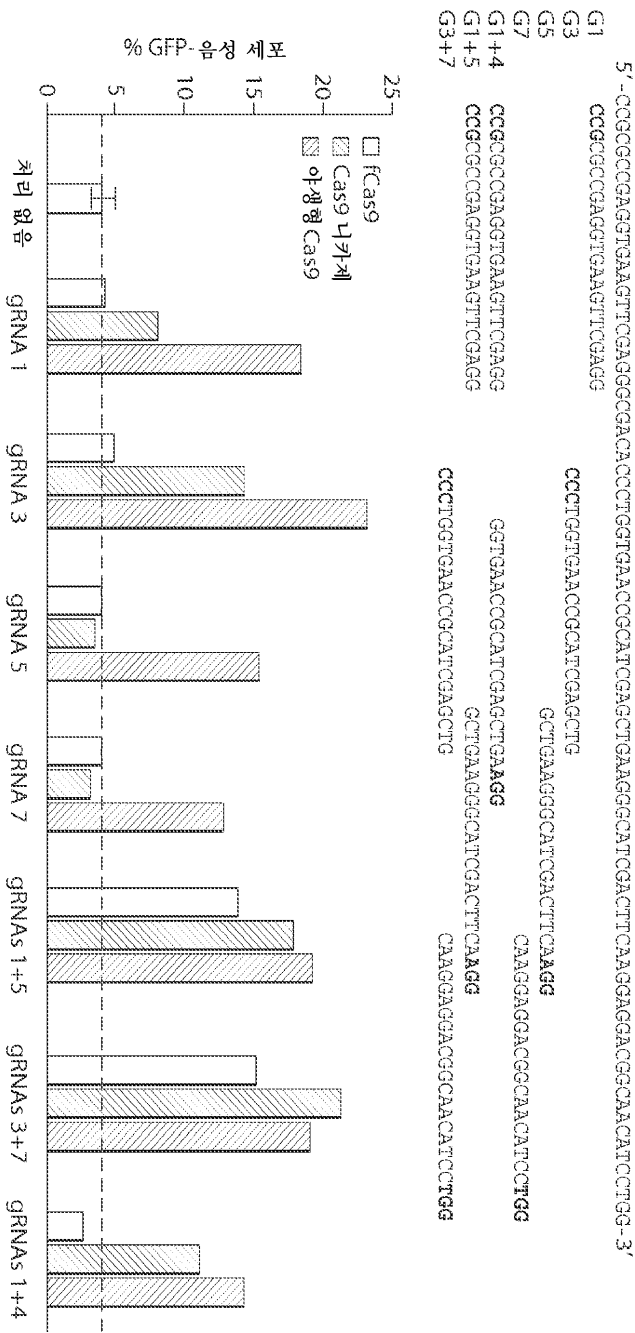
도면8a



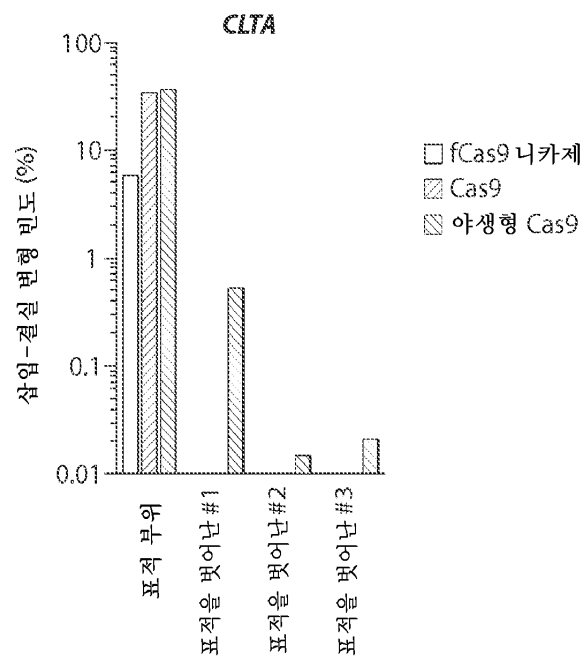
도면 8b



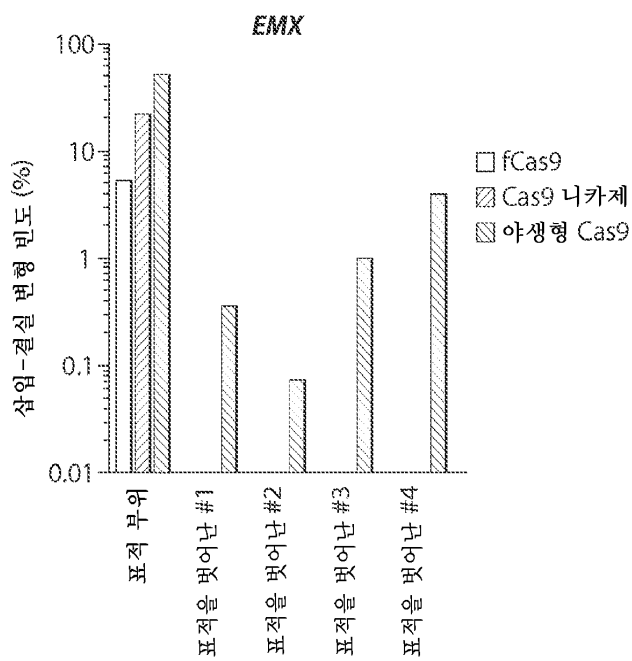
도면8c



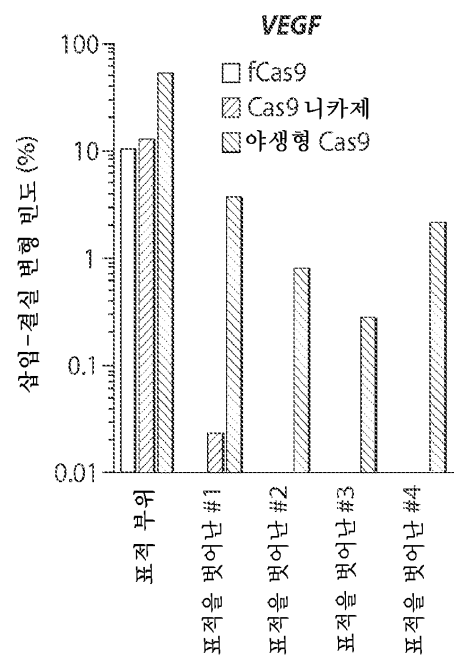
도면8d



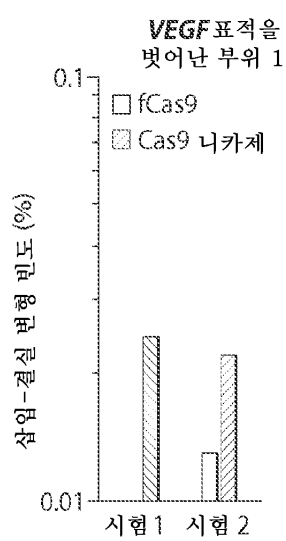
도면8e



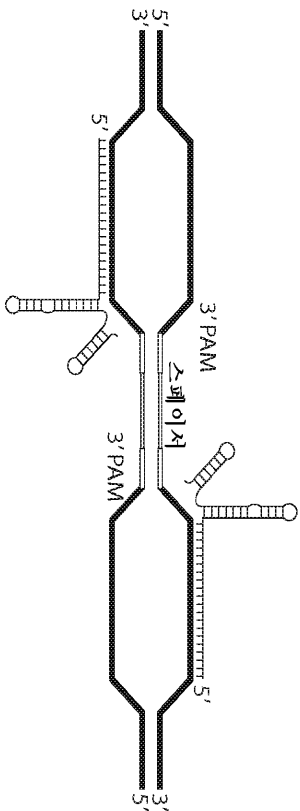
도면8f



도면8g



배향 B:



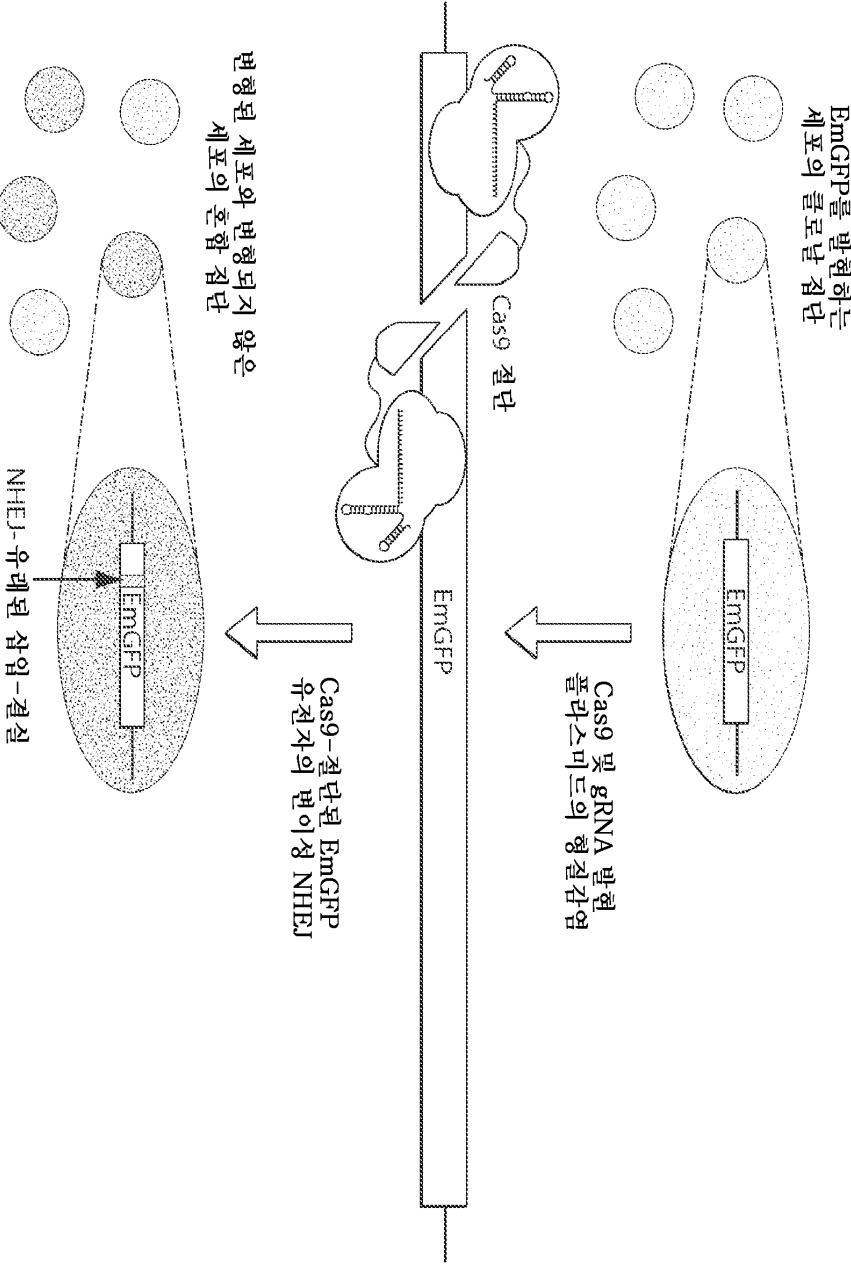
가이드 RNA 스페이서 길이(bp)

	G8	11	42	G12	G13	G14
G8						
G9			19			28
G10						16
G11						4

EmGFP (bp 297-388) 내의 가이드 RNA 서열

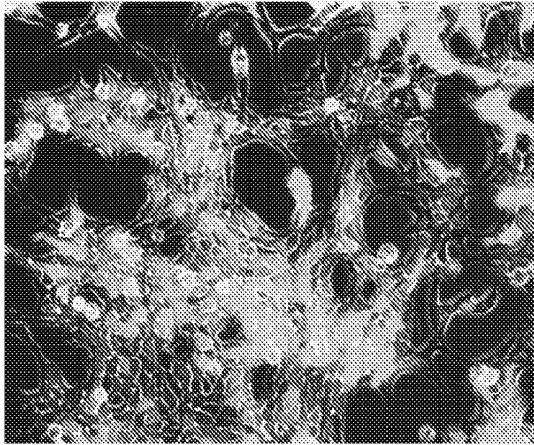
5'-GGAGCGCACCATCTTCTTCAAGGAGCGCACTACAAGACCCGCGCCGAGGTGAAGTTCAGGGCGACACCTTGTTGTAACCGCATCGAGCTGAAGGGCATCG-3'
G8 ACCATCTTCTTCAAGGAGCGG
G9 CAACTACAAGACCCGCGCCGAGG
G10 CCGCGCCGAGGTGAAGTTCAGG
G11 GAAGTTCAGGGCGACACCTTGG
G12 CCGCGCCGAGGTGAAGTTCAG
G13 CCTGGTGAACCGCATCGAGCTCA
G14 CCGCATCGAGCTGAAGGGCATCG

도면10a

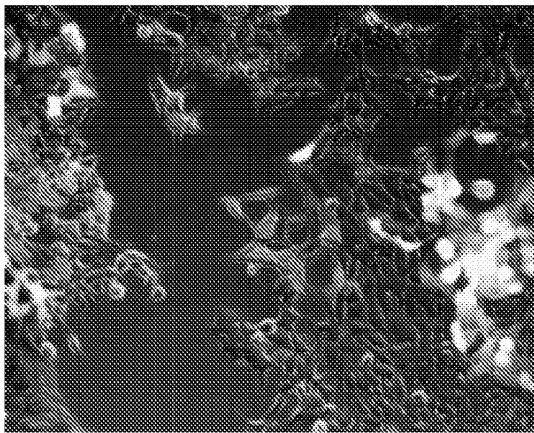


도면10b

처리되지 않은 EmGFP-HEK293 세포

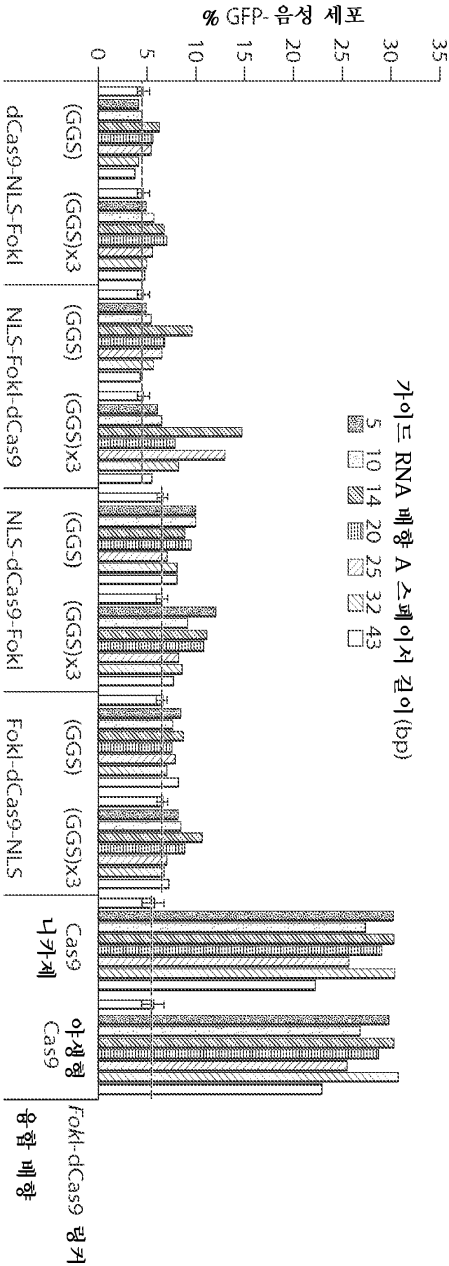


Cas9 및 gRNA 발현
플라스미드 형질감염 ↓

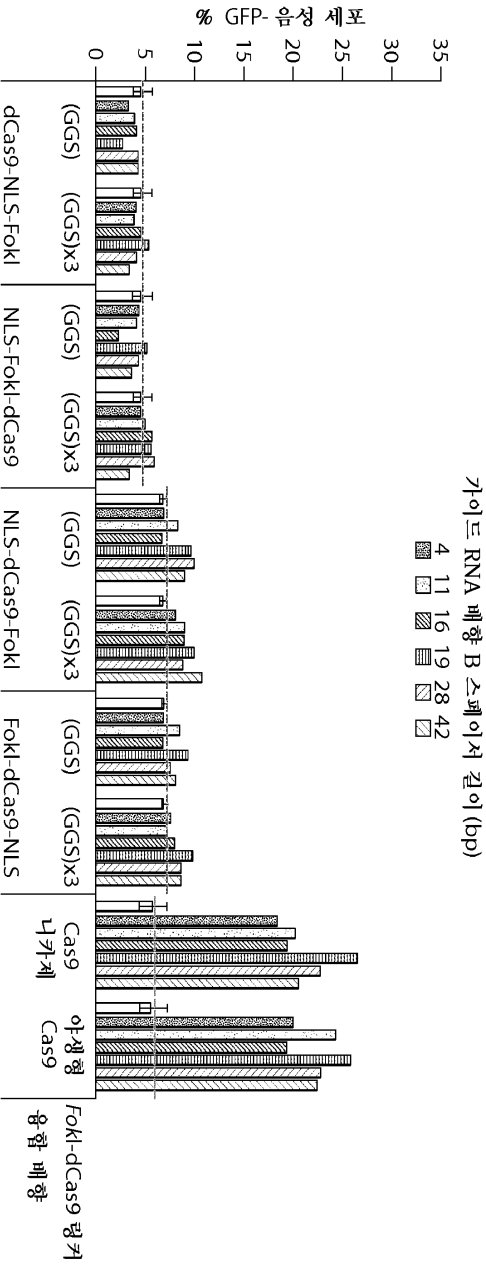


야생형 Cas9 및 gRNA로
형질감염된 EmGFP-HEK293 세포

도면11a



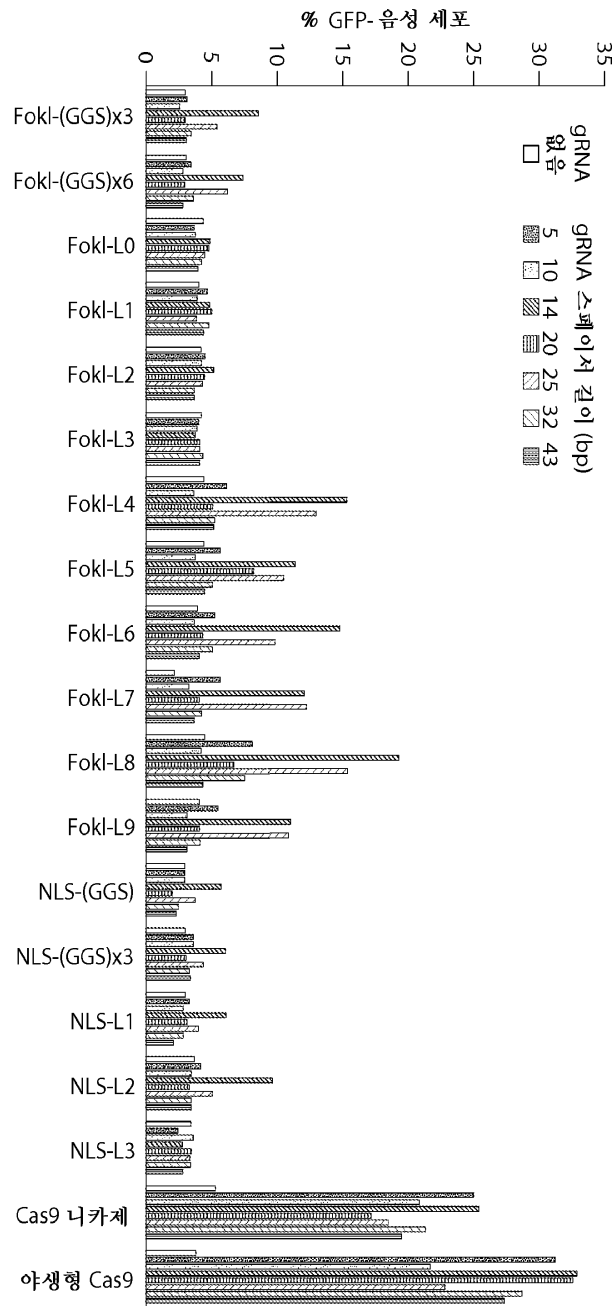
도면11b

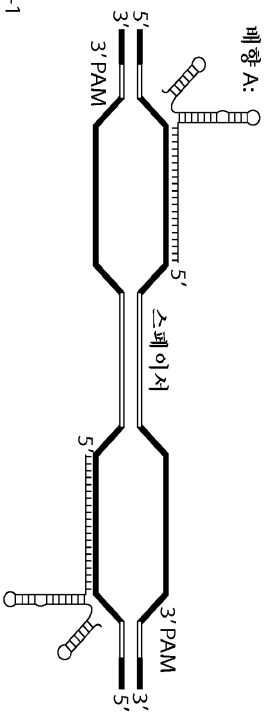


도면 12a

명칭	NLS- 링커 -Fok1	Fok1- 링커 -dCas9
FokI-(GGG)x3	GGG	GGSGGSGGS
FokI-(GGG)x6	GGG	GGSGGSGGSGGSGGSGGS
FokI-L0	GGG	-
FokI-L1	GGG	MKIIQLPSA
FokI-L2	GGG	VRHKLKRVGS
FokI-L3	GGG	VPFLLEPDNINGKTC
FokI-L4	GGG	GHGTGSTGSGSS
FokI-L5	GGG	MSRPDPA
FokI-L6	GGG	GSAGSAAGSGEF
FokI-L7	GGG	SGSETPGTSESA
FokI-L8	GGG	SGSETPGTSESATPES
FokI-L9	GGG	SGSETPGTSESATPEGGSGGS
NLS-(GGG)	GGG	GGSM
NLS-(GGG)x3	GGSGGSGGS	GGSM
NLS-L1	VPFLLEPDNINGKTC	GGSM
NLS-L2	GSAGSAAGSGEF	GGSM
NLS-L3	SIVAQLSRPDPA	GGSM
야생형 Cas9	N/A	N/A
Cas9 니카제	N/A	N/A

도면12b

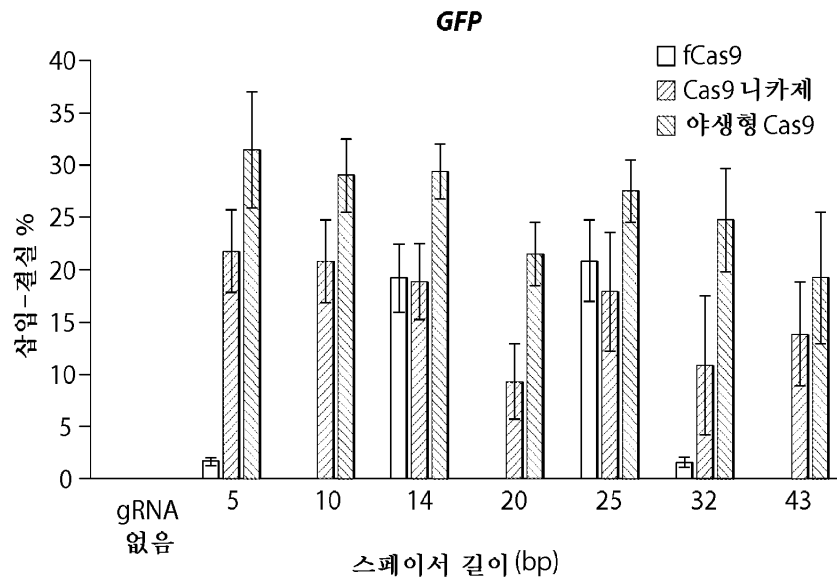




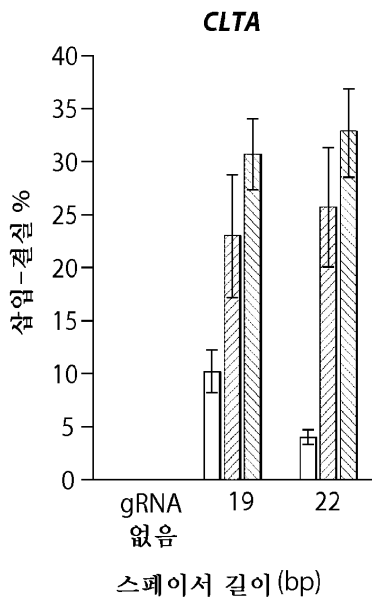
스페이서 (bp)

CLTA-1	CCCCAAGTCTAGCAGACAGAGCCCAAGATGTCTTCCCGCATGCGCTCAGTCCCTCATCTCCCTCAAGCAGG	22
	CCCCAAGTCTAGCAGACAGAGCCA (C1)	
	AGTCTCATCTTCCCTCAAGCAGG (C2)	
CLTA-2	CCCTGTGGAAACACTACAATCTGCATATCTTAATCCTACTCAGTGAAGCTCTTCACAGTCATTGG	19
	CCCTGTGGAAACACTACATCTGC (C3)	
	GTGAAGCTCTTCACAGTCATTGG (C4)	
HBC	CCGTTACTGCCCCCTGTGGGCGAAGGTGAACGTGATGAAGTTGGTGTGAGGCCCTTGGGCGAGGTTGGTATCAAGGTTTACAAGACAGGTTTAAAG	5
	CGTTACTGCCCCCTGTGGGCGAAG (H1)	11
	CGTGGATGAAGTTGGTGGTGGG (H2)	16
	TGAAGTTGGTGGTGAAGGCCCTGG (H3)	20
	TTGGTGGTGAAGGCCCTGGGCAAGG (H4)	28
	TGGTGAAGGCCCTGGGCAAGTTGG (H5)	
	CCCTGGGCGAGGTTGGTATCAAGG (H6)	47
	AAAGTTTACAAGACAGGTTTAAAGG (H7)	
EMX	CCCTTCTTCTTGTGCTCGGACTCAGGCCCTTCTCTCCACAGCTTCTGCGGTTTGAATTGTTCCTCCGCGTTCTGG	23
	CCCTTCTTCTTGTGCTCGGACTC (E1)	29
	GCCGTTTGTACTTTGTCTCCCTCCGG (E2)	
	TGTACTTTGTCTCCCTCCGGTTCTGG (E3)	
VEGF	CCAGGAGCAAACTCCCCCGACCCCTTTCCAAAGCCCATTCCTCTTTAGCCAGAGCCCGGGTGTGCAGACGG	14
	CCAGGAGCAAACTCCCCCACCC (V1)	16
	ATTCCCTCTTTAGCCAGAGCCCGG (V2)	27
	TCCCTCTTTAGCCAGAGCCCGGG (V3)	
	CCAGAGCCGGGTTGTGCAGACCGG (V4)	

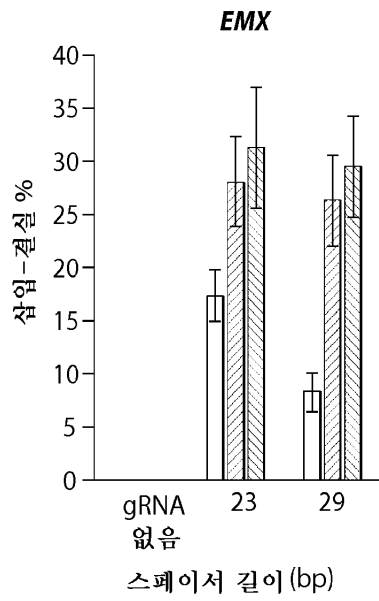
도면14a



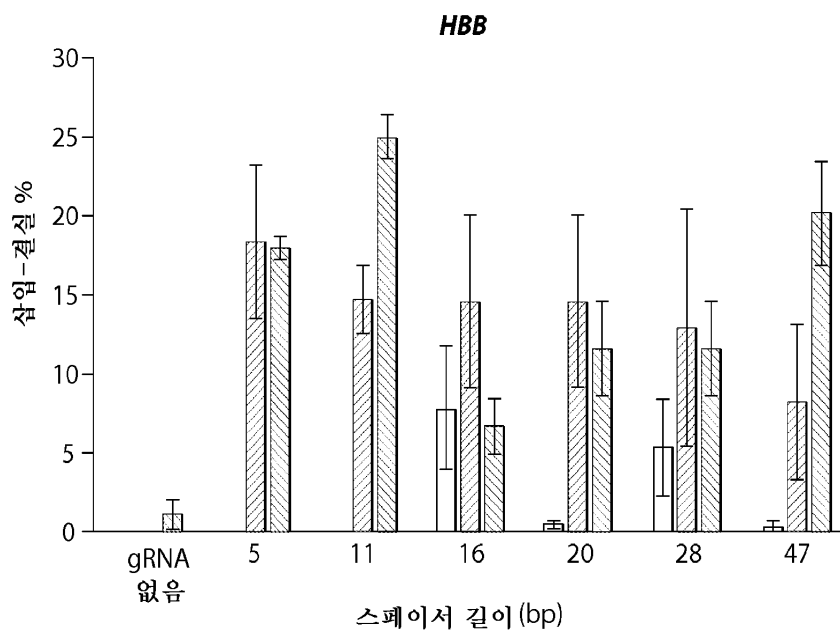
도면14b



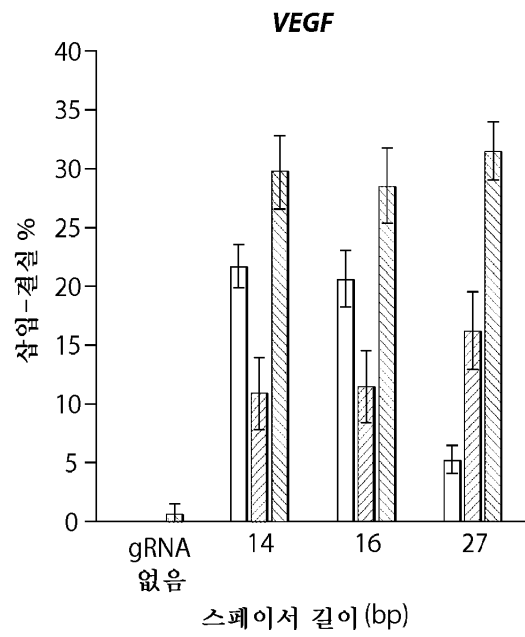
도면14c



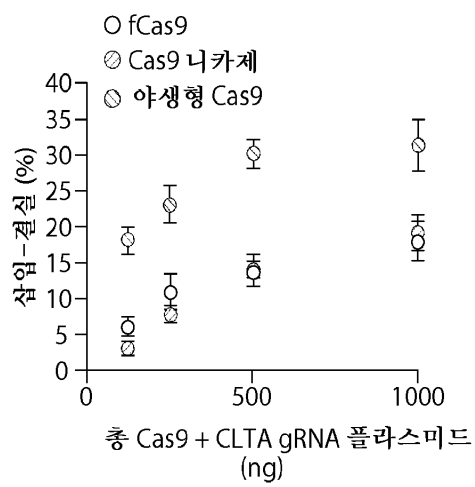
도면14d



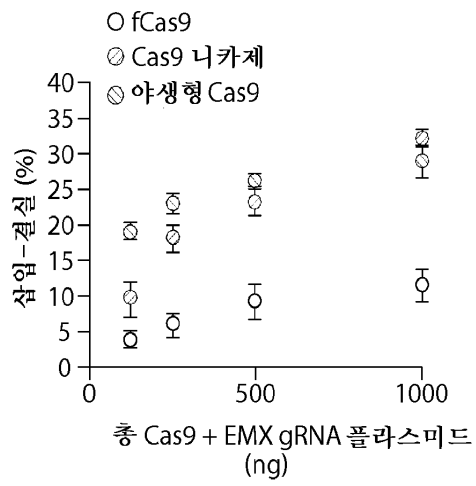
도면14e



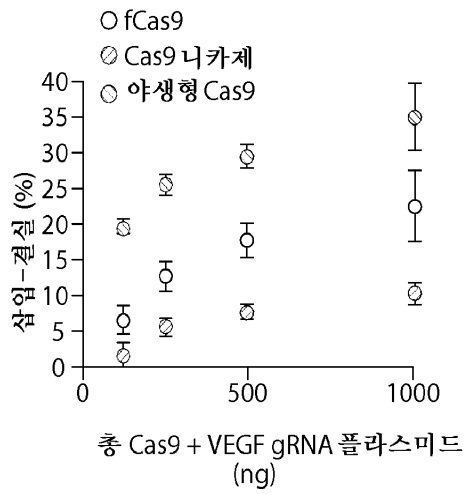
도면15a



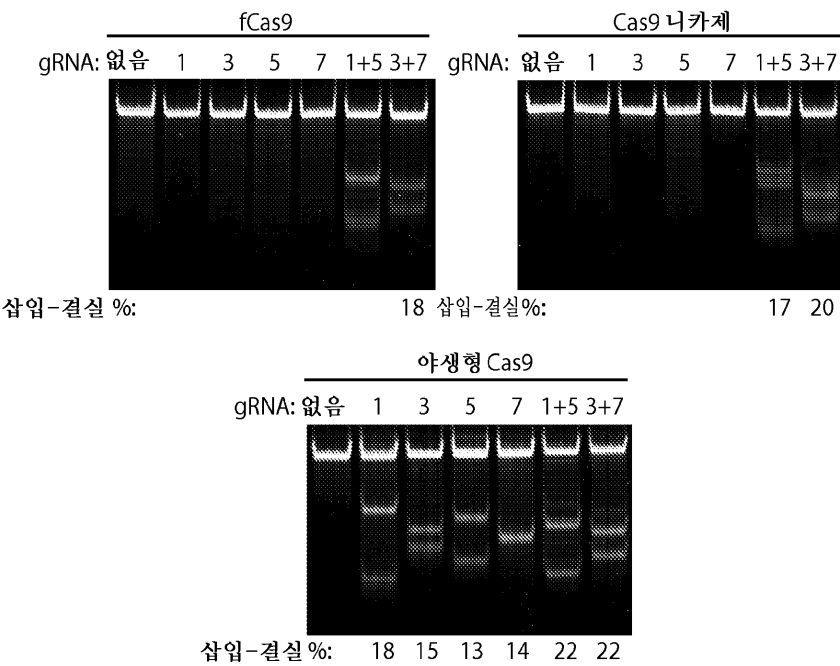
도면15b



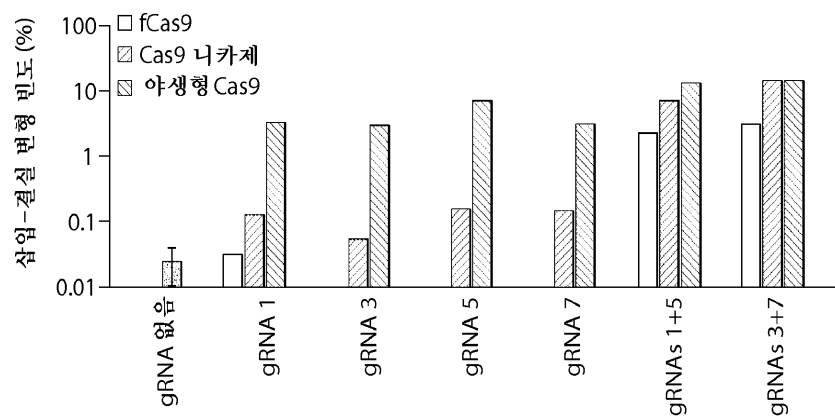
도면15c



도면16

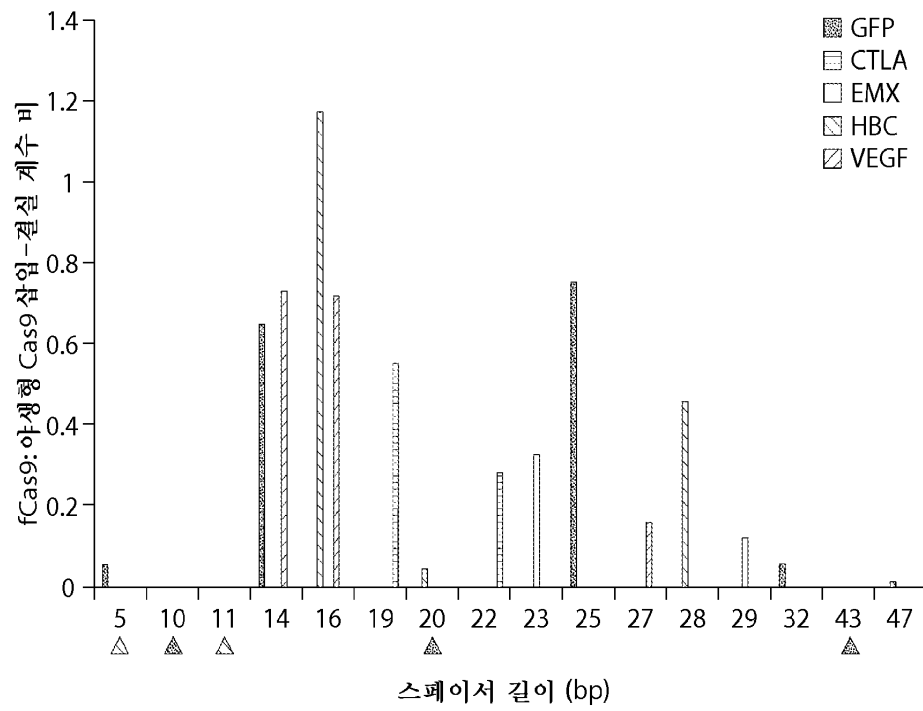


도 16A



도 16B

도면17



VEGF 표적에 작용한 부위의 아생형 Cas9 뉴클레아제 변형:

```
4747 gctgtttgggaagtcagaaaaatagggggtCCAGGAGCAAACTCCCCCAACCCcctttccaagccATTCCCTCTTTAGCGAGAAGCUGggtgtgcagacgycagtc (ref)
4577 gctgtttgggaagtcagaaaaatagggggtCCAGGAGCAAACTCCCCCAACCCcctttccaagccATTCCCTCTTTAGCGAGAAGCUGggtgtgcagacgycagtc
58 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccccaacccctttccaagcccatccctcttaagc-----cggggtgtgcagacgycagtc
54 gctgtttgggaagtcagaaaaatagggggtccaggag-----cccccaacccctttccaagcccatccctcttaagc-----cggggtgtgtcagacgycagtc
43 gctgtttgggaagtcagaaaaatag-----gctgtttgggaagtcagaaaaatag-----cggggtgtgcagacgycagtc
33 gctgtttgggaagtcagaaaaatagggggtccaggagc-----gctgtttgggaagtcagaaaaatag-----cggggtgtgcagacgycagtc
23 gctgtttgggaagtcagaaaaatagggggtccag-----gctgtttgggaagtcagaaaaatag-----cggggtgtgtcagacgycagtc
22 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccccaacccctttccaagcccatccctcttaagccag-----gtgtgtcagacgycagtc
18 gctgtttgggaagtcagaaaaatagggggtccaggag-----t-----agccgggtgtgtcagacgycagtc
```

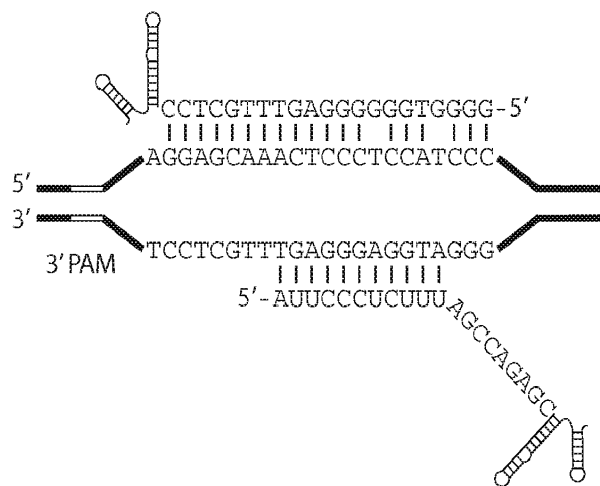
VEGF 표적에 작용한 부위의 Cas9 니카제 변형:

```
8770 gctgtttgggaagtcagaaaaatagggggtCCAGGAGCAAACTCCCCCAACCCcctttccaagccATTCCCTCTTTAGCGAGAAGCUGggtgtgtcagacgycagtc (ref)
78 gctgtttgggaagtcagaaaaatagggggtccag-----gctgtttgggaagtcagaaaaatagggggtccag-----acggcagtc
60 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccccaacccctttccaagcc-----ggtgtgtgcagacgycagtc
58 gctgtttgggaagtcagaaaaatagggggtcca-----aagcccatccctcttaagccaggagccgggtgtgtcagacgycagtc
56 gctgtttgggaagtcagaaaaatagggggt-----gtgcagacgycagtc
49 gctgtttgggaagtcagaaaaatagggggtccag-----cggggtgtgtcagacgycagtc
37 gctgtttgggaagtcagaaaaatagggggtccagg-----gtgtcagacgycagtc
36 gctgtttgggaagtcagaaaaatagggggtccaggagc-----cggggtgtgtcagacgycagtc
27 gctgtttgggaagtcagaaaaatag-----cggggtgtgtcagacgycagtc
```

VEGF 표적에 작용한 부위의 ICas9 뉴클레아제 변형:

```
8959 gctgtttgggaagtcagaaaaatagggggtCCAGGAGCAAACTCCCCCAACCCcctttccaagccATTCCCTCTTTAGCGAGAAGCUGggtgtgtcagacgycagtc (ref)
125 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactccccca-----agccatccctcttaagccaggagccgggtgtgtcagacgycagtc
121 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccccaacccct-----tccctcttaagccaggagccgggtgtgtcagacgycagtc
77 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccccaacccct-----tccctcttaagccaggagccgggtgtgtcagacgycagtc
73 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactccccca-----gccatccctcttaagccaggagccgggtgtgtcagacgycagtc
48 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccccaacccct-----atccctcttaagccaggagccgggtgtgtcagacgycagtc
44 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccccaacccctt-----agccaggagccgggtgtgtcagacgycagtc
24 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccccaacccctt-----aaagccatccctcttaagccaggagccgggtgtgtcagacgycagtc
22 gctgtttgggaagtcagaaaaatagggggtccaggagcaaaactcccc-----aagccatccctcttaagccaggagccgggtgtgtcagacgycagtc
```


도면18c



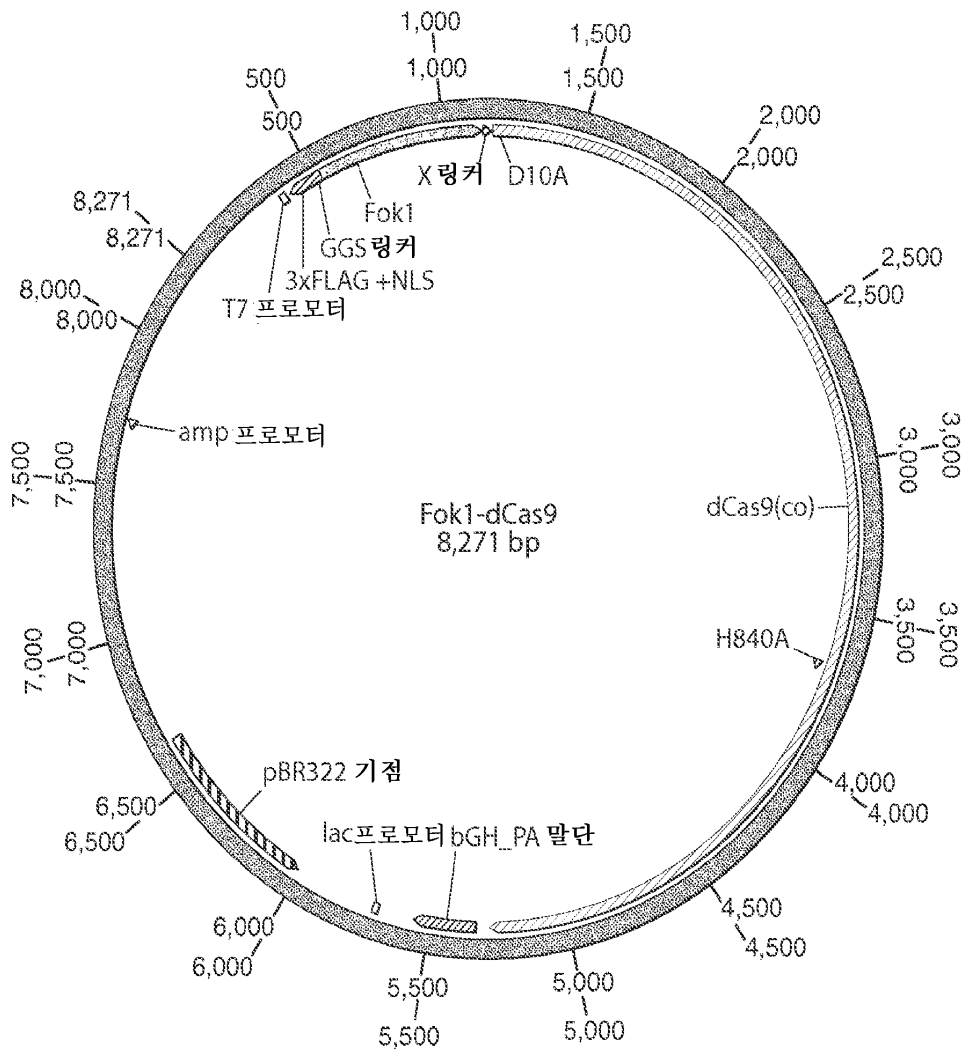
CRA
CCAGCAAGAGGCTCCGAGCGAGCAAGCTCAGTTTACACCCGATCCACTGGGGAGCAGGAATATCTGTGGCTTGTGACACGGACTCAAGTGGGCTGG
CCAGCAAGAGGCTCCGAGCGAG (CRA-1)
CCGAGCGAGCAAGCTCAGTTTA (CRA-2)
CCCGATCCACTGGGGAGCAGGAA (CRA-3)
AGTTTACACCCGATCCACTGGGG (CRA-4)
TGGGGAGCAGGAATATCTGTGG (CRA-5)
ATACTGTGGCTTGTGACACGG (CRA-6)
GCTTGTGACAGGACTCAAGTGG (CRA-8)
TGACACGGACTCAAGTGGGCTGG (CRA-7)

가이드 RNA 스페이서 길이 (bp)

	CRA-4	CRA-5	CRA-6	CRA-8
CRA-1	7	25	38	48
CRA-2		12	25	35
CRA-3			0	10

도면 19a

도면20



서열 목록

SEQUENCE LISTING

<110> President and Fellows of Harvard College

<120> CAS9 VARIANTS AND USES THEREOF

<130> H0824.70146W000

<140> PCT/US2014/054291

<141> 2014-09-05

<150> US 61/980,315

<151> 2014-04-16

<150> US 61/915,414

<151> 2013-12-12

<150> US 61/874,609

<151> 2013-09-06

<160> 339

<170> PatentIn version 3.5

<210> 1

<211> 4104

<212> DNA

<213> Streptococcus pyogenes

<400> 1

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agattttttc aaatgagatg gcgaaagtag atgatatgtt ctttcacga 300

cttgaagagt ctttttttgt ggaagaagac aagaagcatg aacgtcatcc ttttttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcgt 480

atgattaagt ttctgtgtca ttttttgatt gagggagatt taaatcctga taatagtgt 540

gtggacaaac tatttatcca gttgttacia atctacaatc aattatttga agaaaaccct 600

attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttggaat 720

ctcattgctt tgtcattggg attgaccctt aattttaaat caaattttga tttggcagaa 780

gatgctaagt tacagcttcc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttggtt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatagt gaaataacta aggtccctt atcagcttca 960

atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagtctga 1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atgggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttactttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtcgaa gaaacaatta ccccatggaa ttttgaagaa	1440
gttgtcgata aaggtgcttc agctcaatca ttattgaac gcatgacaaa ctttgataaa	1500
aatcttccaa atgaaaaagt actacaaaa catagtgtgc tttatgagta ttttacggtt	1560
tataacgaat tgacaaaggt caaatatgtt actgaggga tgcgaaaacc agcatttctt	1620
tcaggatgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc	1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt	1740
tcaggagtgg aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt	1800
attaaagata aagatTTTTT ggataatgaa gaaaatgaag atatcttaga ggatattgtt	1860
ttacattga cctatttga agataggggg atgattgagg aaagacttaa aacatatgct	1920
cacctctttg atgataaggt gatgaacag cttaaacgtc gccgttatac tggttgggga	1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta	2040
gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat	2100
agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttta	2160
catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggat tttacagact	2220
gtaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcgttatt	2280
gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg	2340
aaacgaatcg aagaaggat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt	2400
gaaaatactc aattgcaaaa tgaaaagctc tatctctatt atctacaaaa tggaagagac	2460
atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt	2520
gtccacaaa gtttcattaa agacgattca atagacaata aggtactaac gcgttctgat	2580
aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac	2640
tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg	2700
aaagctgaac gtggaggttt gagtgaactt gataaagctg gttttatcaa acgccaattg	2760
gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtgc catgaatact	2820
aaatacgtg aaaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa	2880
ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac	2940
catcatgcc atgatgcgt tctaaatgcc gtcgttggaa ctgctttgat taagaaatat	3000
ccaaaacttg aatcggagtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg	3060
attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaat	3120
atcatgaact tcttcaaaa agaaattaca cttgcaaatg gagagattcg caaacgccct	3180
ctaatacgaaa ctaatgggga aactggagaa attgtctggg ataaaggcg agattttgcc	3240

acagtgcgca aagtattgtc catgccccaa gtcaatatgt tcaagaaaac agaagtacag 3300
acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct 3360
cgtaaaaaag actgggatcc aaaaaaatat ggtggttttg atagtccaac ggtagcttat 3420

tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa 3480
gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaaatcc gattgacttt 3540
ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat 3600
agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa 3660
aaaggaaatg agctggctct gccaaagcaa tatgtgaatt ttttatattt agctagtcac 3720
tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag 3780
cataagcatt attagatga gattattgag caaatcagtg aattttctaa gcgtgttatt 3840

ttagcagatg ccaatttaga taaagttctt agtgcataa acaaacatag agacaaacca 3900
atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc 3960
gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc tacaaaagaa 4020
gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat 4080
ttgagtcagc taggagggtga ctga 4104

<210> 2

<211> 1367

<212> PRT

<213> Streptococcus pyogenes

<400> 2

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

1	5	10	15
Gly Trp Ala Val	Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe		
	20	25	30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile			
	35	40	45
Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu			
	50	55	60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys			

65	70	75	80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser			

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

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

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

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

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

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

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp

835 840 845

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

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070	1075	1080
Lys Val Leu Ser Met Pro Gln	Val Asn Ile Val Lys	Lys Thr Glu
1085	1090	1095
Val Gln Thr Gly Gly Phe Ser	Lys Glu Ser Ile Leu	Pro Lys Arg
1100	1105	1110
Asn Ser Asp Lys Leu Ile Ala	Arg Lys Lys Asp Trp	Asp Pro Lys
1115	1120	1125
Lys Tyr Gly Gly Phe Asp Ser	Pro Thr Val Ala Tyr	Ser Val Leu
1130	1135	1140
Val Val Ala Lys Val Glu Lys	Gly Lys Ser Lys Lys	Leu Lys Ser
1145	1150	1155
Val Lys Glu Leu Leu Gly Ile	Thr Ile Met Glu Arg	Ser Ser Phe
1160	1165	1170
Glu Lys Asn Pro Ile Asp Phe	Leu Glu Ala Lys Gly	Tyr Lys Glu
1175	1180	1185
Val Lys Lys Asp Leu Ile Ile	Lys Leu Pro Lys Tyr	Ser Leu Phe
1190	1195	1200
Glu Leu Glu Asn Gly Arg Lys	Arg Met Leu Ala Ser	Ala Gly Glu
1205	1210	1215
Leu Gln Lys Gly Asn Glu Leu	Ala Leu Pro Ser Lys	Tyr Val Asn
1220	1225	1230
Phe Leu Tyr Leu Ala Ser His	Tyr Glu Lys Leu Lys	Gly Ser Pro
1235	1240	1245
Glu Asp Asn Glu Gln Lys Gln	Leu Phe Val Glu Gln	His Lys His
1250	1255	1260
Tyr Leu Asp Glu Ile Ile Glu	Gln Ile Ser Glu Phe	Ser Lys Arg
1265	1270	1275
Val Ile Leu Ala Asp Ala Asn	Leu Asp Lys Val Leu	Ser Ala Tyr
1280	1285	1290
Asn Lys His Arg Asp Lys Pro	Ile Arg Glu Gln Ala	Glu Asn Ile
1295	1300	1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

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

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 3

<211> 4212

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 3

atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc 60

ataaccgatg aatacaaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt 120

cattcgatta aaaagaatct tatcggtgcc ctctatttcg atagtggcga aacggcagag 180

gcgactcgcc tgaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt 240

tacttacaag aaatttttag caatgagatg gccaaagttg acgattcttt ctttcaccgt 300

ttggaagagt ccttccttgt cgaagaggac aagaaacatg aacggcaccc catctttgga 360

aacatagtag atgaggtggc atatcatgaa aagtacccaa cgatttatca cctcagaaaa 420

aagctagtgt actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat 480

atgataaagt tccgtgggca ctttctcatt gagggtgatc taaatccgga caactcggat 540

gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct 600

ataaatgcaa gtggcgtgga tgcgaaggct attcttagcg cccgcctctc taaatcccga 660

cggctagaaa acctgatcgc acaattaccg ggagagaaga aaaatgggtt gttcggtaac 720

cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa 780

gatgccaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca 840

caaattggag atcagtatgc ggacttattt ttggctgcca aaaccttag cgatgcaatc 900

ctctatctg acatactgag agttaatact gagattacca aggcgcggtt atccgcttca 960

atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc ctagtccgt 1020

cagcaactgc ctgagaaata taaggaaata ttttttgatc agtcgaaaaa cgggtacgca	1080
ggttatatcg acggcggagc gagtcaagag gaattctaca agtttatcaa acctatatta	1140
gagaagatgg atgggacgga agagtgtctt gtaaaactca atcgcaaga tctactgca	1200
aagcagcgga ctttcgacaa cggtagcatt cccatcaaa tccacttagg cgaattgcat	1260
gctatactta gaaggcagga ggatttttat cgttctca aagacaatcg tgaaaagatt	1320
gagaaaaacc taacctttcg cataccttac tatgtgggac ccctggcccg agggaaactct	1380
cggttcgcac ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa	1440
gttgtcgata aaggtgcgtc agctcaatcg ttcacgaga ggatgacaa ctttgacaag	1500
aatttaccga acgaaaaagt attgcctaag cacagtttac tttacgagta tttcacagt	1560
tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta	1620
agcggagAAC agaagaaagc aatagtagat ctgttatcca agaccaaccg caaagtgaca	1680
gttaagcaat tgaaagagga ctactttaag aaaattgaat gcttcgattc tgtcgagatc	1740
tccgggtag aagatcgatt taatgcgtca cttggtacgt atcatgacct cctaaagata	1800
attaaagata aggacttctt ggataacgaa gagaatgaag atatcttaga agatatagt	1860
ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgt	1920
cacctgttcg acgataaggt tatgaacag ttaaagaggc gtcgctatac gggctgggga	1980
cgattgtcgc ggaaacttat caacgggata agagacaagc aaagtggtaa aactattctc	2040
gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac	2100
tctttaacct tcaaagagga tatacaaaag gcacaggttt ccggacaagg ggactcattg	2160
cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaaggcat actccagaca	2220
gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaacgga aaacattgta	2280
atcgagatgg cagcgaaaa tcaaacgact cagaaggggc aaaaaacag tcgagagcgg	2340
atgaagagaa tagaagaggg tattaaagaa ctgggcagcc agatcttaa ggagcatcct	2400
gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg	2460
gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatcac	2520
attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgtc tacacgtcgc	2580
gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag	2640
aactattggc ggcagctcct aaatcgcaaa ctgataacgc aaagaaagt cgataactta	2700
actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccgattttat taaacgtcag	2760
ctcgtggaaa cccgccaat cacaagcat gttgcacaga tactagattc ccgaatgaat	2820
acgaaatcgc acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca	2880

aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940
taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000

taccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060
atgatcgca aaagcgaaca ggagataggc aaggctacag ccaataactt cttttattct 3120
aacattatga atttctttaa gacggaatc actctggcaa acggagagat acgcaaacga 3180
cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg cggggacttc 3240
gcgacggtga gaaaagtttt gtccatgcc caagtcaaca tagtaaagaa aactgaggtg 3300
cagaccggag ggttttcaaa ggaatcgatt ctccaaaaa ggaatagtga taagctcatc 3360
gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc 3420

tattctgtcc tagtagtggc aaaagttgag aagggaat ccaagaaact gaagtcagtc 3480
aaagaattat tggggataac gattatggag cgtcgtctt ttgaaaagaa ccccatcgac 3540
ttccttgagg cgaaaggta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600
tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660
caaaagggga acgaactcgc actaccgtct aaatacgtga atttcctgta ttttagctcc 3720
cattacgaga agttgaaagg ttcactgaa gataacgaac agaagcaact tttgttgag 3780
cagcacaac attatctga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840

atcctagctg atgccaatct ggacaaagta ttaagcgcac acaacaagca cagggataaa 3900
cccatacgtg agcaggcgga aaatattatc catttgttta ctcttacaa cctcggcgct 3960
ccagccgcat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020
gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080
gatttgctac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140
tacaagacc atgacggtga ttataaagat catgacatcg attacaagga tgacgatgac 4200
aaggctgcag ga 4212

<210> 4

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 4

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

1 5 10 15

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

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

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

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

 Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255
 Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260	265	270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp		
275	280	285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp		
290	295	300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser		
305	310	315
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys		
325	330	335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe		
340	345	350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser		
355	360	365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp		
370	375	380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg		
385	390	395
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu		
405	410	415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe		
420	425	430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile		
435	440	445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp		
450	455	460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu		
465	470	475
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr		
485	490	495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser		
500	505	510

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

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

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

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

 Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755	760	765	
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile			
770	775	780	
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro			
785	790	795	800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu			
805	810	815	
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg			
820	825	830	
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys			
835	840	845	
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg			
850	855	860	
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys			
865	870	875	880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys			
885	890	895	
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp			
900	905	910	
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr			
915	920	925	
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp			
930	935	940	
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser			
945	950	955	960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg			
965	970	975	
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val			
980	985	990	
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe			
995	1000	1005	

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

1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys	Gln Leu Phe Val Glu	Gln His Lys
1250	1255	1260
His Tyr Leu Asp Glu Ile Ile	Glu Gln Ile Ser Glu	Phe Ser Lys
1265	1270	1275
Arg Val Ile Leu Ala Asp Ala	Asn Leu Asp Lys Val	Leu Ser Ala
1280	1285	1290
Tyr Asn Lys His Arg Asp Lys	Pro Ile Arg Glu Gln	Ala Glu Asn
1295	1300	1305
Ile Ile His Leu Phe Thr Leu	Thr Asn Leu Gly Ala	Pro Ala Ala
1310	1315	1320
Phe Lys Tyr Phe Asp Thr Thr	Ile Asp Arg Lys Arg	Tyr Thr Ser
1325	1330	1335
Thr Lys Glu Val Leu Asp Ala	Thr Leu Ile His Gln	Ser Ile Thr
1340	1345	1350
Gly Leu Tyr Glu Thr Arg Ile	Asp Leu Ser Gln Leu	Gly Gly Asp
1355	1360	1365
<210> 5		
<211> 1368		
<212> PRT		
<213> Artificial Sequence		
<220><223> Synthetic Polypeptide		
<400> 5		
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val		
1	5	10
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe		
	20	30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile		
35	40	45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu		
50	55	60

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

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

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

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

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

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

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

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

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

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

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

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

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

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

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

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

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805

810

815

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

820

825

830

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

835

840

845

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

850

855

860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865

870

875

880

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

885

890

895

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

900

905

910

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

915

920

925

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

930

935

940

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

945

950

955

960

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

965

970

975

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

980

985

990

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

995

1000

1005

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

1010

1015

1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025

1030

1035

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

1040

1045

1050

Asn Gly	Glu Ile Arg Lys Arg	Pro Leu Ile Glu Thr	Asn Gly Glu
1055	1060	1065	
Thr Gly	Glu Ile Val Trp Asp	Lys Gly Arg Asp Phe	Ala Thr Val
1070	1075	1080	
Arg Lys	Val Leu Ser Met Pro	Gln Val Asn Ile Val	Lys Lys Thr
1085	1090	1095	
Glu Val	Gln Thr Gly Gly Phe	Ser Lys Glu Ser Ile	Leu Pro Lys
1100	1105	1110	
Arg Asn	Ser Asp Lys Leu Ile	Ala Arg Lys Lys Asp	Trp Asp Pro
1115	1120	1125	
Lys Lys	Tyr Gly Gly Phe Asp	Ser Pro Thr Val Ala	Tyr Ser Val
1130	1135	1140	
Leu Val	Val Ala Lys Val Glu	Lys Gly Lys Ser Lys	Lys Leu Lys
1145	1150	1155	
Ser Val	Lys Glu Leu Leu Gly	Ile Thr Ile Met Glu	Arg Ser Ser
1160	1165	1170	
Phe Glu	Lys Asn Pro Ile Asp	Phe Leu Glu Ala Lys	Gly Tyr Lys
1175	1180	1185	
Glu Val	Lys Lys Asp Leu Ile	Ile Lys Leu Pro Lys	Tyr Ser Leu
1190	1195	1200	
Phe Glu	Leu Glu Asn Gly Arg	Lys Arg Met Leu Ala	Ser Ala Gly
1205	1210	1215	
Glu Leu	Gln Lys Gly Asn Glu	Leu Ala Leu Pro Ser	Lys Tyr Val
1220	1225	1230	
Asn Phe	Leu Tyr Leu Ala Ser	His Tyr Glu Lys Leu	Lys Gly Ser
1235	1240	1245	
Pro Glu	Asp Asn Glu Gln Lys	Gln Leu Phe Val Glu	Gln His Lys
1250	1255	1260	
His Tyr	Leu Asp Glu Ile Ile	Glu Gln Ile Ser Glu	Phe Ser Lys
1265	1270	1275	
Arg Val	Ile Leu Ala Asp Ala	Asn Leu Asp Lys Val	Leu Ser Ala

1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 6

<211> 198

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 6

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

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

Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro

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

180 185 190
Asn Gly Glu Ile Asn Phe

195

<210> 7

<211> 4221

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 7

atggactata aggaccacga cggagactac aaggatcatg atattgatta caaagacgat	60
gacgataaga tggcccaaaa gaagaagcgg aaggctcggta tccacggagt cccagcagcc	120
gataaaaagt attctattgg ttagctatc ggcactaatt ccgttggatg ggctgtcata	180
accgatgaat acaaagtacc ttcaaagaaa ttaaggtgt tggggaacac agaccgtcat	240
tcgattaaaa agaatttat cgggtgccctc ctattcgata gtggcgaaac ggcagaggcg	300
actcgctga aacgaaccgc tcggagaagg tatacacgtc gcaagaaccg aatatgttac	360
ttacaagaaa ttttagcaa tgagatggcc aaagttgacg attctttctt tcaccgtttg	420
gaagagtcct tccttgtcga agaggacaag aaacatgaac ggcaccccat ctttggaac	480
atagtagatg aggtggcata tcatgaaaag tacccaacga tttatcacct cagaaaaaag	540
ctagttgact caactgataa agcggacctg aggttaatct acttggtctt tgcccatatg	600
ataaagttcc gtgggcactt tctcattgag ggtgatctaa atccggacaa ctcggatgtc	660
gacaaactgt tcattcagtt agtacaaacc tataatcagt tgtttgaaga gaacctata	720
aatgcaagtg gcgtggatgc gaaggctatt cttagcggcc gcctctctaa atcccgcagg	780

ctagaaaacc tgatcgaca attacccgga gagaagaaaa atgggttggtt cggtaacctt	840
atagcgctct cactaggcct gacaccaaatt ttttaagtcga acttcgactt agctgaagat	900
gccaaattgc agcttagtaa ggacacgtac gatgacgac tgcacaatct actggcacia	960
attggagatc agtatgcgga cttatTTTTT gctgccaaaa accttagcga tgcaatcctc	1020
ctatctgaca tactgagagt taatactgag attaccaagg cgccgttatc cgcttcaatg	1080
atcaaaagggt acgatgaaca tcaccaagac ttgacacttc tcaaggccct agtccgtcag	1140
caactgcctg agaaatataa ggaaatattc ttgatcagt cgaaaaacgg gtacgcaggt	1200
tatattgacg gcggagcgag tcaagaggaa ttctacaagt ttatcaaacc catattagag	1260
aagatggatg ggacggaaga gttgcttgta aaactcaatc gcgaagatct actgcgaaag	1320
cagcggactt tgcacaacgg tagcattcca catcaaacc acttaggcga attgcatgct	1380
atacttagaa ggcaggagga tttttatccg ttctcaaag acaatcgtga aaagattgag	1440
aaaatcctaa ctttctgcat accttactat gtgggacccc tggcccagg gaactctcgg	1500
ttcgcattga tgacaagaaa gtccgaagaa acgattactc catggaattt tgaggaagtt	1560
gtcgataaag gtgcgtcagc tcaatcgttc atcgagagga tgaccaactt tgacaagaat	1620
ttaccgaacg aaaaagtatt gcctaagcac agtttacttt acgagtattt cacagtgtac	1680
aatgaactca cgaaagttaa gtatgtcact gagggcatgc gtaaacccgc ctttctaagc	1740
ggagaacaga agaaagcaat agtagatctg ttattcaaga ccaaccgcaa agtgacagtt	1800
aagcaattga aagaggacta ctttaagaaa attgaatgct tcgattctgt cgagatctcc	1860
ggggtagaag atcgatttaa tgcgtcactt ggtacgtatc atgacctcct aaagataatt	1920
aaagataagg acttcctgga taacgaagag aatgaagata tcttagaaga tatagtgttg	1980
actcttacc tctttgaaga tcgggaaatg attgaggaaa gactaaaaac atacgtcac	2040
ctgttcgacg ataaggttat gaaacagtta aagaggcgtc gctatacggg ctggggacga	2100
ttgtcgcgga aacttatcaa cgggataaga gacaagcaaa gtggtlaaac tattctcgat	2160
tttctaaga gcgacggctt gcctaatagg aactttatgc agctgatcca tgatgactct	2220
ttaaccttca aagaggatat acaaaaggca caggtttccg gacaagggga ctcatcgac	2280
gaacatattg cgaatcttgc tggttcgcca gccatcaaaa agggcatact ccagacagtc	2340
aaagtagtgg atgagctagt taaggtcatg ggacgtcaca aaccggaaaa cattgtaatc	2400
gagatggcac gcgaaaatca aacgactcag aaggggcaaa aaaacagtcg agagcggatg	2460
aagagaatag aagagggtat taaagaactg ggcagccaga tcttaaagga gcatcctgtg	2520
gaaaatacc aattgcagaa cgagaaactt tacctctatt acctacaaaa tggaaggagc	2580
atgtatgttg atcaggaact ggacataaac cgtttatctg attacgacgt cgatcacatt	2640

gtacccaat cctttttgaa ggacgattca atcgacaata aagtgccttac acgctcggat 2700
aagaaccgag ggaaaagtga caatgttcca agcgaggaag tcgtaaagaa aatgaagaac 2760

tattggcggc agctcctaaa tgcgaaactg ataacgcaaa gaaagttcga taacttaact 2820
aaagctgaga ggggtggctt gtctgaactt gacaaggccg gatttattaa acgtcagctc 2880
gtggaaaccc gccaaatcac aaagcatgtt gcacagatac tagattcccg aatgaatacg 2940
aaatacgacg agaacgataa gctgattcgg gaagtcaaag taatcacttt aaagtcaaaa 3000
ttggtgtcgg acttcagaaa ggattttcaa ttctataaag ttagggagat aaataactac 3060
caccatgcgc acgacgttta tcttaatgcc gtcgtaggga ccgcactcat taagaaatac 3120
ccgaagctag aaagtgagtt tgtgtatggt gattacaaag tttatgacgt ccgtaagatg 3180

atcgcgaaaa gcgaacagga gataggcaag gctacagcca aatacttctt ttattctaac 3240
attatgaatt tctttaagac ggaaatcact ctggcaaacg gagagatacg caaacgacct 3300
ttaattgaaa ccaatgggga gacaggtgaa atcgatatggg ataagggccg ggacttcgcg 3360
acggtgagaa aagttttgtc catgccccaa gtcaacatag taaagaaaac tgaggtgcag 3420
accggagggt tttaaagga atcgattctt ccaaaaagga atagtataa gctcatcgct 3480
cgtaaaaagg actgggaccc gaaaaagtac ggtggcttcg atagccctac agttgcctat 3540
tctgtcctag tagtggcaaa agttgagaag ggaaaatcca agaaactgaa gtcagtcaaa 3600

gaattattgg ggataacgat tatggagcgc tcgtcttttg aaaagaaccc catcgacttc 3660
cttgaggcga aaggttacaa ggaagtaaaa aaggatctca taattaaact accaaagtat 3720
agtctgtttg agttagaaaa tggccgaaaa cggtatgttg ctacgcccgg agagcttcaa 3780
aagggaacg aactcgact accgtctaaa tacgtgaatt tcctgtatct agcgtcccat 3840
tacgagaagt tgaaaggttc acctgaagat aacgaacaga agcaactttt tgttgagcag 3900
cacaacatt atctcgacga aatcatagag caaatttcgg aattcagtaa gagagtcac 3960
ctagctgatg ccaatctgga caaagtatta agcgcataca acaagcacag ggataaaccc 4020

atacgtgagc aggcggaaaa tattatccat ttgtttactc ttaccaacct cggcgtcca 4080
gccgcattca agtattttga cacaacgata gatcgcaaac gatacacttc taccaaggag 4140
gtgctagacg cgacactgat tcaccaatcc atcacgggat tataatgaaac tcgcatagat 4200
ttgtcacagc ttgggggtga c 4221

<210> 8
<211> 1368
<212> PRT
<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 8

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

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

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

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

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

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

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

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

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

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

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

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

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

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

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

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

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

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

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

1205	1210	1215
Glu Leu Gln Lys Gly Asn Glu	Leu Ala Leu Pro Ser	Lys Tyr Val
1220	1225	1230
Asn Phe Leu Tyr Leu Ala Ser	His Tyr Glu Lys Leu	Lys Gly Ser
1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys	Gln Leu Phe Val Glu	Gln His Lys
1250	1255	1260
His Tyr Leu Asp Glu Ile Ile	Glu Gln Ile Ser Glu	Phe Ser Lys
1265	1270	1275
Arg Val Ile Leu Ala Asp Ala	Asn Leu Asp Lys Val	Leu Ser Ala
1280	1285	1290
Tyr Asn Lys His Arg Asp Lys	Pro Ile Arg Glu Gln	Ala Glu Asn
1295	1300	1305
Ile Ile His Leu Phe Thr Leu	Thr Asn Leu Gly Ala	Pro Ala Ala
1310	1315	1320
Phe Lys Tyr Phe Asp Thr Thr	Ile Asp Arg Lys Arg	Tyr Thr Ser
1325	1330	1335
Thr Lys Glu Val Leu Asp Ala	Thr Leu Ile His Gln	Ser Ile Thr
1340	1345	1350
Gly Leu Tyr Glu Thr Arg Ile	Asp Leu Ser Gln Leu	Gly Gly Asp
1355	1360	1365
<210> 9		
<211> 4836		
<212> DNA		
<213> Artificial Sequence		
<220><223> Synthetic Polynucleotide		
<400> 9		
atggataaaa agtattctat tggtttagct atcggcacta attccgttgg atgggctgtc	60	
ataaccgatg aatacaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt	120	
cattcgatta aaaagaatct tatcggtgcc ctctatttcg atagtggcga aacggcagag	180	
gcgactcgcc tgaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt	240	

tacttacaag aaatTTTTag caatgagatg gccaaagtTg acgattcttt ctttcaccgt	300
ttggaagagt ctttccttTg cgaagaggac aagaaacatg aacggcaccc catctttgga	360
aacatagtag atgaggtggc atatcatgaa aagtacccaa cgatttatca cctcagaaaa	420
aagctagtTg acTcaactga taaagcggac ctgaggTtaa tctactTggc tcttgcccat	480
atgataaagt tccgtgggca ctttctcatt gagggTgac taaatccgga caactcggat	540
gtcgacaaaC tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct	600
ataaatgcaa gtggcgtTga tgcgaaggct attcttagcg cccgcctctc taaatcccga	660
cggctagaaa acctgatcgc acaattaccc ggagagaaga aaaatgggtt gttcggtaac	720
cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa	780
gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca	840
caaatTggag atcagtatgc ggacttattt ttggctgcca aaaaccttag cgatgcaatc	900
ctcctatctg acatactgag agttaatact gagattacca aggcgccgtt atccgcttca	960
atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc cctagtccgt	1020
cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca	1080
ggttatatTg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta	1140
gagaagatgg atgggacgga agagtTgctt gtaaaactca atcgcaaga tctactgcca	1200
aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat	1260
gctatactta gaaggcagga ggatttttat ccgttctca aagacaatcg tgaagagatt	1320
gagaaaaatc taacctttcg cataccttac tatgtgggac ccctggcccg agggaactct	1380
cggttcgcat ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa	1440
gttgtcgata aaggtgcgtc agctcaatcg ttcatcgaga ggatgaccaa ctttgacaag	1500
aatttaccga acgaaaaagt attgcctaag cacagtttac ttacagagta tttcacagtg	1560
tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta	1620
agcggagaac agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca	1680
gttaagcaat tgaagagga ctactttaag aaaattgaat gcttcgattc tgtcgagatc	1740
tccgggtag aagatcgatt taatgcgtca ctTgtacgt atcatgacct cctaaagata	1800
attaaagata aggacttccT ggataacgaa gagaatgaag atatcttaga agatatagtg	1860
ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgct	1920
cacctgttcg acgataaggt tatgaaacag ttaaagaggc gtcgtatatac gggctgggga	1980
cgattgtcgc ggaaacttat caacgggata agagacaagc aaagtggtaa aactattctc	2040
gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac	2100

tctttaacct tcaaagagga tatacaaaaag gcacaggttt ccggacaagg ggactcattg	2160
cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaagggcat actccagaca	2220
gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaacgga aaacattgta	2280
atcgagatgg caccgaaaa tcaaacgact cagaaggggc aaaaaacag tcgagagcgg	2340
atgaagagaa tagaagaggg tattaagaa ctgggcagcc agatcttaaa ggagcatcct	2400
gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg	2460
gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgaatgcc	2520
attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgtc tacacgtcgc	2580
gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag	2640
aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagt cgaataactta	2700
actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccgattttat taaacgtcag	2760
ctcgtggaaa cccgccaat cacaagcat gttgcacaga tactagattc ccgaatgaat	2820
acgaaatacg acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca	2880
aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac	2940
taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa	3000
taccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag	3060
atgatcgca aaagcgaaca ggagataggc aaggctacag ccaataactt cttttattct	3120
aacattatga atttctttaa gacggaatc actctggcaa acggagagat acgcaaacga	3180
cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggc cgggacttc	3240
gcgacggtga gaaaagtttt gtccatgccc caagtcaaca tagtaaagaa aactgaggtg	3300
cagaccggag ggttttcaaa ggaatcgatt ctccaaaaa ggaatagtga taagctcatc	3360
gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc	3420
tattctgtcc tagtagtggc aaaagttgag aagggaat ccaagaaact gaagtcagtc	3480
aaagaattat tggggataac gattatggag cgtcgtctt ttgaaaagaa ccccatcgac	3540
ttccttgagg cgaaaggta caaggaagta aaaaaggatc tcataattaa actaccaaag	3600
tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt	3660
caaaagggga acgaactcgc actaccgtct aaatacgtga atttctgta tttagcgtcc	3720
cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact tttgttgag	3780
cagcacaac attatctcga cgaaatcata gagcaaat cgggaattcag taagagagtc	3840

atcctagctg atgccaatct ggacaaagta ttaagcgcat acaacaagca cagggataaa 3900

cccatacgtg agcaggcgga aaatattatc catttgitta ctcttacaa cctcggcgct 3960

ccagccgcat tcaagtatit tgacacaacg atagatcgca aacgatacac ttctaccaag 4020

gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080

gatttgtcac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140

tacaaagacc atgacgggtg ttataaagat catgacatcg attacaagga tgacgatgac 4200

aaggctgcag gatcagggtg aagtggcggc agcggagggt ctggatccca actagtcaaa 4260

agtgaactgg aggagaagaa atctgaactt cgtcataaat tgaaatatgt gcctcatgaa 4320

tatattgaat taattgaaat tgccagaaat tccactcagg atagaattct tgaaatgaag 4380

gtaatggaat tttttatgaa agtttatgga tatagaggta aacatttggg tggatcaagg 4440

aaaccggacg gagcaattta tactgtcgga tctctattg attacggtgt gatcgtggat 4500

actaaagctt atagcggagg ttataatctg ccaattggcc aagcagatga aatgcaacga 4560

tatgtcgaag aaaatcaaac acgaaacaaa catatcaacc ctaatgaatg gtggaaagtc 4620

tatccatctt ctgtaacgga atttaagttt ttatttgtga gtggtcactt taaaggaaac 4680

tacaaagctc agcttacacg attaaatcat atcactaatt gtaatggagc tgttcttagt 4740

gtagaagagc ttttaattgg tggagaaatg attaaagccg gcacattaac cttagaggaa 4800

gtcagacgga aatttaataa cggcgagata aacttt 4836

<210> 10

<211> 4845

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 10

atggactaca aagaccatga cgggtgattat aaagatcatg acatcgatta caaggatgac 60

gatgacaaga tgcccccaa gaagaagagg aagggtgggca ttcaccgcgg ggtacctatg 120

gataaaaagt attctattgg tttagctatc ggcaactaatt ccgttggatg ggctgtcata 180

accgatgaat acaaagtacc ttcaaagaaa ttttaaggtgt tggggaacac agaccgtcat 240

tcgattaaaa agaattctat cgggtgccctc ctattcgata gtggcgaaac ggcagaggcg 300

actcgctga aacgaaccgc tcggagaagg tatcacgctc gcaagaaccg aatatgttac 360

ttacaagaaa tttttagcaa tgagatggcc aaagttgacg attctttctt tcaccgtttg 420

gaagatcct tccttgtcga agaggacaag aaacatgaac ggcaccccat ctttggaac 480

atagtagatg aggtggcata tcatgaaaag tacccaacga tttatcacct cagaaaaaag 540
 ctagttgact caactgataa agcggacctg aggttaatct acttggctct tgcccatatg 600
 ataaagtcc gtgggcactt tctcattgag ggtgatctaa atccggacaa ctcggatgtc 660

 gacaaactgt tcatccagtt agtacaaacc tataatcagt tgtttgaaga gaaccctata 720
 aatgcaagtg gcgtggatgc gaaggctatt cttagcgcgc gcctctctaa atcccgcagg 780
 ctagaaaacc tgatcgaca attacccgga gagaagaaaa atgggttgtt cggtaacctt 840
 atagcgtct cactaggcct gacaccaaatt ttaagtca acttcgactt agctgaagat 900
 gccaaattgc agcttagtaa ggacacgtac gatgacgac tcgacaatct actggcacia 960
 attggagatc agtatgcgga cttatttttg gctgccaaaa accctagcga tgcaatctc 1020
 ctatctgaca tactgagagt taatactgag attaccaagg cgccgttatt cgcttcaatg 1080

 atcaaaaggt acgatgaaca tcaccaagac ttgacacttc tcaaggccct agtccgtcag 1140
 caactgcctg agaaatataa ggaaatattc tttgatcagt cgaaaaacgg gtacgcagg 1200
 tatattgacg gcggagcgag tcaagaggaa ttctacaagt ttatcaaacc catattagag 1260
 aagatggatg ggacggaaga gttgcttgta aaactcaatc gcgaagatct actgcgaaag 1320
 cagcggactt tcgacaacgg tagcattcca catcaaacc acttaggcga attgcatgct 1380
 atacttagaa ggcaggagga tttttatccg ttctcaaag acaatcgtga aaagattgag 1440
 aaaatcctaa ctttcgcat accctactat gtgggacccc tggcccgagg gaactctcgg 1500

 ttcgcatgga tgacaagaaa gtccgaagaa acgattactc catggaattt tgaggaaagt 1560
 gtcgataaag gtgcgtcagc tcaatcgctt atcgagagga tgaccaactt tgacaagaat 1620
 ttaccgaacg aaaaagtatt gcctaagcac agtttacttt acgagtattt cacagtgtac 1680
 aatgaactca cgaaagttaa gtatgtcact gagggcatgc gtaaacccgc ctttctaage 1740
 ggagaacaga agaaagcaat agtagatctg ttattcaaga ccaaccgcaa agtgacagtt 1800
 aagcaattga aagaggacta ctttaagaaa attgaatgct tcgattctgt cgagatctcc 1860
 ggggtagaag atcgatttaa tgcgtcactt ggtacgtatc atgacctcct aaagataatt 1920

 aaagataagg acttcctgga taacgaagag aatgaagata tcttagaaga tatagtgtt 1980
 actcttacc tctttgaaga tcgggaaatg attgaggaaa gactaaaaac atacgtcac 2040
 ctgttcgacg ataaggttat gaaacagtta aagaggcgtc gctatcggg ctggggacga 2100
 ttgtcgcgga aacttatcaa cgggataaga gacaagcaaa gtggtaaaac tattctcgat 2160
 tttctaaaga gcgacggctt cgccaatagg aactttatgc agctgatcca tgatgactct 2220
 ttaaccttca aagaggatat acaaaaggca caggtttccg gacaagggga ctcatgtcac 2280

gaacatattg cgaatcttgc tggttcgcca gccatcaaaa agggcatact ccagacagtc 2340

aaagtagtgg atgagctagt taaggctatg ggacgtcaca aaccggaaaa cattgtaatc 2400

gagatggcac gcgaaaaatca aacgactcag aaggggcaaaa aaaacagtcg agagcggatg 2460

aagagaatag aagaggggat taaagaactg ggcagccaga tcttaaagga gcatcctgtg 2520

gaaaataccc aattgcagaa cgagaaactt tacctctatt acctacaaaa tggaggggac 2580

atgtatgttg atcaggaact ggacataaac cgtttatctg attacgacgt cgatgccatt 2640

glaccccaat cctttttgaa ggacgattca atcgacaata aagtgccttac acgctcggat 2700

aagaaccgag ggaaaagtga caatgttcca agcgaggaag tcgtaaagaa aatgaagaac 2760

tattggcggc agctcciaaa tgcgaaactg ataacgcaaa gaaagttcga taacttaact 2820

aaagctgaga ggggtggctt gtctgaactt gacaaggccg gatttattaa acgtcagctc 2880

gtggaaaccc gccaaatcac aaagcatgtt gcacagatac tagattcccg aatgaatacg 2940

aaatacgacg agaacgataa gctgattcgg gaagtcaaag taatcacitt aaagcaaaa 3000

ttggtgtcgg acttcagaaa ggattttcaa ttctataaag ttagggagat aaataactac 3060

caccatgcgc acgacgttta tcttaatgcc gtctgtagga ccgcactcat taagaaatac 3120

ccgaagctag aaagtgagtt tgtgtatggt gattacaaag tttatgacgt ccgtaagatg 3180

atcgcaaaaa gcgaacagga gataggcaag gctacagcca aatacttctt ttattctaac 3240

attatgaatt tctttaagac ggaaatcact ctggcaaacg gagagatacg caaacgacct 3300

ttaattgaaa ccaatgggga gacaggtgaa atcgtatggg ataagggccg ggacttcgcg 3360

acggtgagaa aagttttgtc catgccccaa gtcaacatag taaagaaaac tgagggtcag 3420

accggagggt tticaaagga atcgattctt ccaaaaagga atagtgataa gctcatcgct 3480

cgtaaaaagg actgggaccc gaaaaagtac ggtggcttcg atagccctac agttgcctat 3540

tctgtcctag tagtggcaaa agttgagaag ggaaaatcca agaaactgaa gtcagtcaaa 3600

gaattatttg ggataacgat tatggagcgc tcgtcttttg aaaagaaccc catcgacttc 3660

cttgaggcga aaggttacaa ggaagtaaaa aaggatctca taattaaact accaaagtat 3720

agtctgtttg agttagaaaa tggccgaaaa cggatgttgg ctagcgccgg agagcttcaa 3780

aagggaacg aactcgact accgtctaaa tacgtgaatt tcctgtattt agcgtcccat 3840

tacgagaagt tgaaggttc acctgaagat aacgaacaga agcaactttt tgttgagcag 3900

cacaaacatt atctcgacga aatcatagag caaatttcgg aattcagtaa gagagtcac 3960

ctagctgatg ccaatctgga caaagtatta agcgcataca acaagcacag ggataaaccc 4020

atacgtgagc aggcggaaaa tattatccat ttgtttactc ttaccaacct cggcgctcca 4080
gccgcattca agtattttga cacaacgata gatcgcaaac gatacacttc taccaaggag 4140
gtgctagacg cgacactgat tcaccaatcc atcacgggat tatatgaaac tcggatagat 4200
ttgtcacagc ttgggggtga ctcaaggtaga agtggcggca gcggagggtc tggatcccaa 4260
ctagtcaaaa gtgaactgga ggagaagaaa tctgaacttc gtcataaatt gaaatatgtg 4320
cctcatgaat atattgaatt aattgaaatt gccagaaatt ccactcagga tagaattctt 4380
gaaatgaagg taatggaatt ttttatgaaa gtttatggat atagaggtaa acatttgggt 4440

ggatcaagga aaccggacgg agcaatttat actgtcggat ctccatttga ttacggtgtg 4500
atcgtggata ctaaagctta tagcggaggt tataatctgc caattggcca agcagatgaa 4560
atgcaacgat atgtcgaaga aaatcaaaca cgaaacaaac atatcaacc taatgaatgg 4620
tggaaagtct atccatcttc tgtaacggaa ttttaagttt tatttgtgag tggtcacttt 4680
aaaggaaact acaaagctca gcttacacga ttaaatacata tcactaattg taatggagct 4740
gttcttagtg tagaagagct ttttaattgt ggagaaatga ttaaagccgg cacattaacc 4800
ttagaggaag tcagacggaa atttaataac ggcgagataa acttt 4845

<210> 11

<211> 4836

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 11

atgggatccc aactagtcaa aagtgaactg gaggagaaga aatctgaact tcgtcataaa 60
ttgaaatatg tgcctcatga atatatgaa ttaattgaaa ttgccagaaa ttccactcag 120
gatagaattc ttgaaatgaa ggtaatggaa ttttttatga aagtttatgg atatagaggt 180
aaacatttgg gtggatcaag gaaaccggac ggagcaattt atactgtcgg atctcctatt 240
gattacggtg tgatcgtgga tactaaagct tatagcggag gttataatct gccaattggc 300
caagcagatg aaatgcaacg atatgtcgaa gaaaatcaaa cacgaaacaa acatatcaac 360

cctaataaat ggtggaaagt ctatccatct tctgtaacgg aatttaagtt tttatttgtg 420
agtggtcact ttaaagggaa ctacaaagct cagcttacac gattaaatca tactactaat 480
tgtaatggag ctgttcttag tgtagaagag cttttaattg gtggagaaat gattaaagcc 540
ggcacattaa ccttagagga agtcagacgg aaatttaata acggcgagat aaactttggc 600
ggtagtgggg gatctggggg aagtatggat aaaaagtatt ctattggttt agctatcggc 660
actaattccg ttggatgggc tgtcataacc gatgaataca aagtaccttc aaagaaattt 720

aaggtgttgg ggaacacaga ccgtcattcg attaaaaaga atcttatcgg tgccctccta 780

ttcgatagtg gcgaaacggc agaggcgact cgcctgaaac gaaccgctcg gagaaggat 840

acacgtcgca agaaccgaat atgttactta caagaaatit ttagcaatga gatggccaaa 900

gttgacgatt ctttctttca ccgtttggaa gaggcttcc ttgtcgaaga ggacaagaaa 960

catgaacggc accccatctt tggaaacata gtagatgagg tggcatatca tgaaggtac 1020

ccaacgattt atcacctcag aaaaaagcta gttgactcaa ctgataaagc ggacctgagg 1080

ttaatctact tggctcttgc ccatatgata aagtccgtg ggacctttct cattgagggt 1140

gatctaaatc cggacaactc ggatgtcgac aaactgttca tccagttagt acaaacctat 1200

aatcagttgt ttgaagagaa ccctataaat gcaagtggcg tggatgcgaa ggctattctt 1260

agcggccgcc tctctaaatc ccgacggcta gaaaacctga tcgcacaatt acccgagag 1320

aagaaaaatg gggtgttcgg taaccttata gcgctctcac taggcctgac accaaatit 1380

aagtcgaact tcgacttagc tgaagatgcc aaattgcagc ttagtaagga cacgtacgat 1440

gacgatctcg acaatctact ggacacaaat ggagatcagt atgcggactt atttttggct 1500

gccaaaaacc ttagcgatgc aatctctcta tctgacatac tgagagttaa tactgagatt 1560

accaaggcgc cgttatccgc ttcaatgatc aaaaggtagc atgaacatca ccaagacttg 1620

acacttctca aggccctagt ccgtcagcaa ctgcctgaga aatataagga aatattcttt 1680

gatcagtcga aaaacgggta cgcaggttat attgacggcg gagcgagtca agaggaattc 1740

tacaagttta tcaaacccat attagagaag atggatggga cggaagagtt gcttgtaaaa 1800

ctcaatcgcg aagatctact gcgaaagcag cggactttcg acaacggtag cattccacat 1860

caaatccact taggcgaatt gcatgctata cttagaaggc aggaggattt ttatccgttc 1920

ctcaaagaca atcgtgaaaa gattgagaaa atcctaacct ttgcataacc ttactatgtg 1980

ggacccttg cccgagggaa ctctcggttc gcatggatga caagaaagtc cgaagaaacg 2040

attactccat ggaattttga ggaagttgtc gataaagggt cgctcagctca atcgttcac 2100

gagaggatga ccaactttga caagaattta ccgaacgaaa aagtattgcc taagcacagt 2160

ttactttacg agtatttcac agtgtacaat gaactcacga aagttaagta tgtactgag 2220

ggcatgcgta aacccgcctt tctaagcgga gaacagaaga aagcaatagt agatctgtta 2280

ttcaagacca accgcaaagt gacagttaag caattgaaag aggactactt taagaaaatt 2340

gaatgcttcg attctgtcga gatctccggg gtagaagatc gatttaaatgc gtcacttgg 2400

acgtatcatg acctcctaaa gataattaaa gataaggact tcctggataa cgaagagaat 2460

gaagatatct tagaagatat agtgttgact cttaccctct ttgaagatcg ggaaatgatt	2520
gaggaaagac taaaacata cgctcacctg ttcgacgata aggttatgaa acagttaaag	2580
aggcgtcgct atacgggctg gggacgattg tcgcggaac ttatcaacgg gataagagac	2640
aagcaaagtg gtaaaactat tctcgatttt ctaaagagcg acggcttcgc caataggaac	2700
tttatgcagc tgatccatga tgactcttta accttcaaag aggatataca aaaggcacag	2760
gtttccggac aaggggactc attgcacgaa catattgca atcttgctgg ttcgccagcc	2820
atcaaaaagg gcatactcca gacagtcaaa gtagtgatg agctagttaa ggtcatggga	2880
cgtcacaaac cgaaaaacat tgtaatcgag atggcacgcg aaaatcaaac gactcagaag	2940
gggcaaaaaa acagtcgaga gcggatgaag agaatagaag agggatttaa agaactgggc	3000
agccagatct taaaggagca tcctgtggaa aataccaat tgcagaacga gaaactttac	3060
ctctattacc taaaaatgg aaggacatg tatgttgatc aggaactgga cataaacctt	3120
ttatctgatt acgacgtcga tgccattgia cccaatcct ttttgaagga cgattcaatc	3180
gacaataaag tgcttacag ctcggataag aaccgaggga aaagtgaca tgttccaagc	3240
gaggaagtgc taaagaaaaa gaagaactat tggcggcagc tcctaatgc gaaactgata	3300
acgcaaagaa agttcgataa cttaactaaa gctgagaggg gtggcttgtc tgaacttgac	3360
aaggccggat ttattaaacg tcagctcgtg gaaaccgcc aaatcacaaa gcatgttgca	3420
cagatactag attcccgaat gaatacgaac tacgacgaga acgataagct gattcgggaa	3480
gtcaaagtaa tcactttaaa gtcaaaattg gtgtcggact tcagaaagga ttttcaattc	3540
tataaagtta gggagataaa taactaccac catgcgcacg acgcttatct taatgccgtc	3600
gtagggaccg cactcattaa gaaataccg aagctagaaa gtgagtttgt gtatggtgat	3660
tacaaagttt atgacgtccg taagatgatc gcgaaaagcg aacaggagat aggcaaggct	3720
acagccaaat acttctttta ttctaacatt atgaatttct ttaagacgga aatcactctg	3780
gcaaaccggag agatacgcaa acgaccttta attgaaacca atggggagac aggtgaaatc	3840
gtatgggata agggccggga cttcgcgacg gtgagaaaag ttttgtccat gcccgaagtc	3900
aacatagtaa agaaaactga ggtgcagacc ggagggtttt caaaggaatc gattcttcca	3960
aaaaggaata gtgataagct catcgtctgt aaaaaggact gggacccgaa aaagtacggt	4020
ggcttcgata gccctacagt tgcctattct gtccatagtag tggcaaaagt tgagaaggga	4080
aaatccaaga aactgaagtc agtcaaagaa ttattgggga taacgattat ggagcgtcgc	4140
tcttttgaag agaaccat cgacttcctt gaggcgaaag gttacaagga agtaaaaaag	4200
gatctcataa ttaactacc aaagtatagt ctgtttgagt tagaaaatgg ccgaaaacgg	4260
atgttggcta gcgccggaga gcttcaaaag gggaacgaac tcgcactacc gtctaatac	4320

gtgaatttcc tgiatttagc gtcccattac gagaagtga aaggttcacc tgaagataac 4380
gaacagaagc aactttttgt tgagcagcac aaacattatc tcgacgaaat catagagcaa 4440
atttcggaat tcagtaagag agtcatecta gctgatgcca atctggacaa agtattaagc 4500
gcatacaaca agcacaggga taaaccataa cgtgagcagg cggaaaatat tatccatttg 4560

tttactetta ccaacctcgg cgctccagcc gcattcaagt attttgacac aacgatagat 4620
cgcaaacgat acacttttac caaggagggtg ctagacgca cactgattca ccaatccatc 4680
acgggattat atgaaactcg gatagatttg tcacagcttg ggggtgacgg atccccaaag 4740
aagaagagga aagtctcgag cgactacaaa gaccatgacg gtgattataa agatcatgac 4800
atcgattaca aggatgacga tgacaaggct gcagga 4836

<210> 12
<211> 4854
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 12

atggactaca aagaccatga cgggtgattat aaagatcatg acatcgatta caaggatgac 60

gatgacaaga tggcccccaa gaagaagagg aagggtgggca ttcaccgcgg ggtacctgga 120
ggttctatgg gatcccaact agtcaaaagt gaactggagg agaagaaatc tgaacttcgt 180
cataaattga aatatgtgcc tcatgaatat attgaattaa ttgaaattgc cagaaattcc 240
actcaggata gaattcttga aatgaaggta atggaatfff ttatgaaagt ttatggatat 300
agaggtaaac atttgggtgg atcaaggaaa ccggacggag caatttatac tgtcggatct 360
cctattgatt acggtgtgat cgtggatact aaagcttata gcggagggtta taatctgcca 420
attggccaag cagatgaaat gcaacgatat gtcgaagaaa atcaaacacg aaacaacat 480

atcaacccta atgaatgggtg gaaagtctat ccatcttctg taacggaatt taagttttta 540
tttgtgagtg gtcactttta aggaaactac aaagctcagc ttacacgatt aaatcatatc 600
actaattgta atggagctgt tcttagtgta gaagagcttt taattggtgg agaaatgatt 660
aaagccggca cattaacctt agaggaagtc agacggaaat ttaataacgg cgagataaac 720
tttggcggta gtgggggatac tgggggaagt atggataaaa agtatcttat tggtttagct 780
atcggcacta attccgttgg atgggctgtc ataaccgatg aatacaaagt accttcaaag 840
aaatttaagg tgttggggaa cacagaccgt cattcgatta aaaagaatct tatcggtgcc 900

ctcctattcg atagtggcga aacggcagag gcgactcgcc tgaaacgaac cgctcggaga 960

aggtatacac gtcgcaagaa ccgaatatgt tacttacaag aaatttttag caatgagatg 1020
 gccaaagtgt acgattcttt ctttcaccgt ttggaagagt ccttccttgt cgaagaggac 1080
 aagaaacatg aacggcaccc catctttgga aacatagtag atgaggtggc atatcatgaa 1140
 aagtacccaa cgatttatca cctcagaaaa aagctagtgt actcaactga taaagcggac 1200
 ctgaggttaa tctacttggc tcttgcccat atgataaagt tccgtgggca ctttctcatt 1260
 gagggtgatc taaatccgga caactcggat gtcgacaaac tgttcatcca gttagtacaa 1320

 acctataatc agttgtttga agagaaccct ataaatgcaa gtggcgtgga tgcgaaggct 1380
 attcttagcg cccgcctctc taaatcccga cggctagaaa acctgatcgc acaattaccc 1440
 ggagagaaga aaaatgggtt gttcggtaac cttatagcgc tctcactagg cctgacacca 1500
 aattttaagt cgaacttcga cttagctgaa gatgccaaat tgcagcttag taaggacacg 1560
 tacgatgacg atctcgacaa tctactggca caaattggag atcagtatgc ggacttattt 1620
 ttggctgccaa aaaaccttag cgatgcaatc ctccatcttg acatactgag agttaatact 1680
 gagattacca aggcgccgtt atccgcttca atgatcaaaa ggtacgatga acatcaccaa 1740

 gacttgacac ttctcaaggc cctagtccgt cagcaactgc ctgagaaata taaggaaata 1800
 ttctttgatc agtcgaaaaa cgggtacgca gggttatattg acggcggagc gagtcaagag 1860
 gaattctaca agtttatcaa acccatatta gagaagatgg atgggacgga agagttgctt 1920
 gtaaaactca atcggaaga tctactgcga aagcagcgga ctttcgacaa cggtagcatt 1980
 ccacatcaaa tccacttagg cgaattgcat gctatactta gaaggcagga ggatttttat 2040
 ccgttcctca aagacaatcg tgaaaagatt gagaaaatcc taacctttcg cataccttac 2100
 tatgtgggac ccctggcccg aggggaactct cggttcgcat ggatgacaag aaagtccgaa 2160

 gaaacgatta ctccatggaa ttttgaggaa gttgtcgata aaggtagctc agctcaatcg 2220
 ttcatcgaga ggatgaccaa ctttgacaag aatttaccga acgaaaaagt attgcctaag 2280
 cacagtttac ttacgagta ttccacagtg tacaatgaac tcacgaaagt taagtatgtc 2340
 actgagggca tgcgtaaac cgcctttcta agcggagaac agaagaaagc aatagtagat 2400
 ctgttattca agaccaaccg caaagtgaca gttaagcaat tgaaagagga ctactttaag 2460
 aaaattgaat gcttcgattc tgtcgagatc tccggggtag aagatcgatt taatgcgtca 2520
 cttggtagct atcatgacct cctaaagata attaaagata aggacttcct ggataacgaa 2580

 gagaatgaag atatctttaga agatatagtg ttgactctta ccccttttga agatcgggaa 2640
 atgattgagg aaagactaaa aacatacgtt cacctgttcg acgataaggt tatgaaacag 2700
 ttaaagaggc gtcgctatac gggctgggga cgattgtcgc ggaaacttat caacgggata 2760
 agagacaagc aaagtggtaa aactattctc gattttctaa agagcgacgg cttcgccaat 2820

aggaacttta tgcagctgat ccatgatgac tctttaacct tcaaagagga tataaaaaag	2880
gcacaggttt ccggacaagg ggactcattg cacgaacata ttgcgaatct tgctggttcg	2940
ccagccatca aaaaggcat actccagaca gtcaaagtag tggatgagct agttaaggct	3000
atgggacgtc acaaaccgga aaacattgta atcgagatgg cacgcgaaaa tcaaacgact	3060
cagaaggggc aaaaaaacag tcgagagcgg atgaagagaa tagaagaggg tattaagaa	3120
ctgggcagcc agatcttaaa ggagcatcct gtggaaaata cccaattgca gaacgagaaa	3180
ctttacctct attacctaca aaatggaagg gacatgtatg ttgatcagga actggacata	3240
aaccgtttat ctgattacga cgtcgatgcc attgtacccc aatccttttt gaaggacgat	3300
tcaatcgaca ataaagtgt tacacgctcg gataagaacc gagggaaaag tgacaatgtt	3360
ccaagcgagg aagtcgtaaa gaaaatgaag aactattggc ggcagctcct aaatgcgaaa	3420
ctgataacgc aaagaaagt ctgataactta actaaagctg agagggttg cttgtctgaa	3480
cttgacaagg ccgattttat taaacgtcag ctctgtgaaa cccgccaaat cacaaagcat	3540
gttgcacaga tactagattc ccgaatgaat acgaaatagc acgagaacga taagctgatt	3600
cgggaagtca aagtaatcac tttaaagtca aaattggtgt cggacttcag aaaggatttt	3660
caattctata aagttaggga gataaataac taccacatg cgcacgacgc ttatcttaat	3720
gccgtcgtag ggaccgcact cattaagaaa taccgaagc tagaaagtga gtttgtgtat	3780
ggtgattaca aagtttatga cgtccgtaag atgatcgca aaagcgaaca ggagataggc	3840
aaggctacag ccaataactt cttttattct aacattatga atttctttaa gacggaatc	3900
actctggcaa acggagagat acgcaaacga ctttaattg aaaccaatgg ggagacagg	3960
gaaatcgtat gggataaggc ccgggacttc gcgacgtga gaaaagtgtt gtccatgccc	4020
caagtcaaca tagtaagaa aactgagggt cagaccggag ggttttcaa ggaatcgatt	4080
cttccaaaaa ggaatagtga taagctcatc gctcgtaaaa aggactggga cccgaaaaag	4140
tacggtggct tcgatagccc tacagttgcc tattctgtcc tagtagtggc aaaagttgag	4200
aagggaat ccaagaaact gaagtcagtc aaagaattat tggggataac gattatggag	4260
cgctcgtctt ttgaaaagaa ccccatcgac ttccttgagg cgaaagggtt caaggaagta	4320
aaaaaggatc tcataattaa actaccaaag tatagtctgt ttgagttaga aaatggccga	4380
aaacggatgt tggttagcgc cggagagctt caaaagggga acgaactcgc actaccgtct	4440
aaatacgtga atttctgta tttagcgtcc cattacgaga agttgaaagg ttcacctgaa	4500
gataacgaac agaagcaact tttgttgag cagcacaac attatctcga cgaaatcata	4560
gagcaaat ttcggaattcag taagagagtc atcctagctg atgccaatct ggacaaagta	4620

ttaagcgcat acaacaagca cagggataaa cccatacgtg agcaggcgga aaatattatc 4680

catttgttta ctcttaccaa cctcggcgct ccagccgcat tcaagtattt tgacacaacg 4740

atagatcgca aacgatacac ttctaccaag gaggtgctag acgcgacact gattcaccaa 4800

tccatcacgg gattatatga aactcggata gatttgcac agcttggggg tgac 4854

<210> 13

<211> 10

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic Polynucleotide

<400> 13

aaaaaaaaa 10

<210> 14

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 14

Gly Gly Ser Gly Gly Ser Gly Gly Ser

1 5

<210> 15

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 15

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly

1 5 10 15

Gly Ser

<210> 16

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 16

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1 5 10 15

<210> 17

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 17

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala

1 5 10

<210> 18

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 18

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Gly

1 5 10 15

Gly Ser Gly Gly Ser

20

<210> 19

<211> 15

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 19

Val Pro Phe Leu Leu Glu Pro Asp Asn Ile Asn Gly Lys Thr Cys

1 5 10 15

<210> 20

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 20

Gly Ser Ala Gly Ser Ala Ala Gly Ser Gly Glu Phe

1 5 10

<210> 21

<211>

> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 21

Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala

1 5 10

<210> 22

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 22

Met Lys Ile Ile Glu Gln Leu Pro Ser Ala

1 5 10

<210> 23

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 23

Val Arg His Lys Leu Lys Arg Val Gly Ser

1 5 10

<210> 24

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 24

Gly His Gly Thr Gly Ser Thr Gly Ser Gly Ser Ser

1 5 10

<210> 25

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 25

Met Ser Arg Pro Asp Pro Ala

1 5

<210> 26

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 26

cggcgagata aacttttaat gaccggatcat catcacca 38

<210> 27

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 27

ccaacggaat tagtgccgat agctaaacca atagaatact ttttatc 47

<210> 28

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 28

gataaaaagt attctattgg ttttagctatc ggcaactaatt ccgttgg 47

<210> 29
<211> 49
<212> DNA

<213> Artificial Sequence
<220><223> Synthetic Polynucleotide

<400> 29

ttcaaaaagg attggggtac aatggcatcg acgtcgtaat cagataaac 49

<210> 30

<211> 49

<212> DNA

<213> Artificial Sequence
<220><223> Synthetic Polynucleotide

<400> 30

gtttatctga ttacgacgtc gatgccattg taccccaatc ctttttgaa 49

<210> 31

<211> 37

<212> DNA

<213> Artificial Sequence
<220><223> Synthetic Polynucleotide

<400> 31

ttgggatcca gaacctctc ctgcagcctt gtcacg 37

<210> 32

<211> 58

<212> DNA

<213> Artificial Sequence
<220><223> Synthetic Polynucleotide

<400> 32

ttgggatcca gaacctccgc tgccgccaact tccacctgat cctgcagcct tgtcatcg 58

<210> 33

<211> 37

<212> DNA

<213> Artificial Sequence
<220><223> Synthetic Polynucleotide

<400> 33
cgatgacaag gctgcaggag gaggttctgg atcccaa 37

<210> 34
<211> 58
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 34
cgatgacaag gctgcaggat cagggtggaag tggcggcagc ggaggttctg gatcccaa 58

<210> 35
<211> 38
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 35
tggtgatgat gaccggtcat taaaagtta tctcgccg 38

<210> 36
<211> 38
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 36
cggcgagata aacttttaat gaccggtcat catcacca 38

<210> 37
<211> 40
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 37
tagggagagc cgccaccatg gactacaaag accatgacgg 40

<210> 38
<211> 46
<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 38

taaaccaata gaatactttt tatccatagg taccccgcgg tgaatg 46

<210> 39

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 39

gataaaaagt attctattgg tttagctatc ggcactaatt ccgttgg 47

<210> 40

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 40

ttcaaaaagg attggggtac aatggcatcg acgtcgtaat cagataaac 49

<210> 41

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 41

gtttatctga ttacgacgtc gatgccattg tacccaatc cttttgaa 49

<210> 42

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 42

ttgggatcca gaacctccgt caccccaag ctgtg 35

<210> 43

<211> 56

<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 43	
ttgggatcca gaacctccgc tgccgccact tccacctgag tcacccccaa gctgtg	56
<210> 44	
<211> 35	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 44	
cacagcttgg gggtagacga ggttctggat cccaa	35
<210> 45	
<211> 56	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 45	
cacagcttgg gggtagactca ggtggaagtg gcggcagcgg aggttctgga tccaa	56
<210> 46	
<211> 38	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 46	
tggtgatgat gaccggtcat taaaagtta tctcgccg	38
<210> 47	
<211> 40	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 47	
tagggagagc cgccaccatg ggatcccaac tagtcaaaag	40

<210> 48
 <211> 45
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 48
 accaatagaa tactttttat ccatgctgcc accaaagttt atctc 45
 <210> 49
 <211> 44
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 49
 accaatagaa tactttttat ccatgctgcc gccacttcca cctg 44

 <210> 50
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 50
 gataaaaagt attctattgg tttagctatc ggcaactaatt ccgttgg 47
 <210> 51
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 51
 ccaacggaat tagtgccgat agctaaacca atagaatact ttttacc 47
 <210> 52
 <211> 49
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 52

gtttatctga ttacgacgtc gatgccattg tacccaatc cttttgaa 49

<210> 53

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 53

tggtgatgat gaccggtcag tcaccccaaa gctgtg 36

<210> 54

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 54

cacagcttgg ggggtgactga ccggtcatca tcacca 36

<210> 55

<211> 40

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 55

cttttgacta gttgggatcc catggtggcg gctctcccta 40

<210> 56

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 56

acaccctcg aacttcacct cggcgg 26

<210> 57

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 57

acaccgtcgc cctcgaactt cacctg

26

<210> 58

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 58

acaccagct cgatgcggtt caccag

26

<210> 59

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 59

acaccgtga accgcatcga gctgag

26

<210> 60

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 60

acaccgtga agggcatcga cttcag

26

<210> 61

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 61

acaccgcat cgacttcaag gaggag

26

<210> 62

<211> 26

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 62
 acaccaagg aggacggcaa catccg 26
 <210> 63
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 63
 acaccacat cttcttcaag gacgag 26
 <210> 64
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 64
 acaccaact acaagaccg cgccgg 26

 <210> 65
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 65
 acacccgcg ccgagtgaa gttcgg 26
 <210> 66
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 66
 acaccgaagt tcgaggcgga caccg 26
 <210> 67

<211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 67
 acaccttcga acttcacctc ggcgcg 26

<210> 68
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 68
 acacctcagc tcgatgcggt tcaccg 26

<210> 69
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 69
 acaccgatg cccttcagct cgatgg 26

<210> 70
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 70
 aaaaccgccg aggtgaagtt cgaggg 26

<210> 71
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 71

aaaacaggtg aagttcgagg gcgacg	26
<210> 72	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 72	
aaaactggtg aaccgcatcg agctgg	26
<210> 73	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 73	
aaaactcagc tcgatgcggt tcaccg	26
<210> 74	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 74	
aaaactgaag tcgatgccct tcagcg	26
<210> 75	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 75	
aaaactcctc cttgaagtcg atgccg	26
<210> 76	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	

<400> 76	
aaaacggatg ttgccgtcct ccttgg	26
<210> 77	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 77	
aaaactcgtc ctgaagaag atgggtg	26
<210> 78	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 78	
aaaaccggcg cgggtcttgt agttgg	26
<210> 79	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 79	
aaaaccgaac ttcacctcgg cgcggg	26
<210> 80	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 80	
aaaacgggtg tcgccctcga acttcg	26
<210> 81	
<211> 26	
<212> DNA	

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 81

aaaacgcgcc gaggtgaagt tcgaag 26

<210> 82

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 82

aaaacggtga accgcatcga gctgag 26

<210> 83

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 83

aaaaccatcg agctgaaggg catcgg 26

<210> 84

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 84

acacctggcc tgcttgctag acttgg 26

<210> 85

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 85

acaccgcaga tgiagtgttt ccacag 26

<210> 86

<211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 86
 acacccttgc cccacagggc agtaag 26
 <210> 87
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 87
 acaccgagtc cgagcagaag aagaag 26
 <210> 88
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 88
 acaccgggtg gggggagttt gctccg 26

 <210> 89
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 89
 aaaaccaagt ctagcaagca ggccag 26
 <210> 90
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 90
 aaaactgtgg aaacactaca tctgcg 26

<210> 91
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 91
 aaaacttctt ctctgctcg gactcg 26

<210> 92
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 92
 aaaacttact gccctgtggg gcaagg 26

<210> 93
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 93
 aaaacggagc aaactccccc cacccg 26

<210> 94
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 94
 aggaaagaac atgtgagcaa aag 23

<210> 95
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide

<400> 95	
cagcgagtca gtgagcga	18
<210> 96	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 96	
ctgtacaaaa aagcaggctt ta	22
<210> 97	
<211> 39	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 97	
aacgtaggtc tctaccgctg tacaaaaaag caggcttta	39
<210> 98	
<211> 83	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 98	
aaaaaaagca ccgactcggg gccacttttt caagttgata acggactagc cttattttaa	60
cttgctatatt ctagctctaa aac	83
<210> 99	
<211> 60	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 99	
ttgctatttc tagctctaaa accgccgagg tgaagttcga ggggtgtttc gtcctttcca	60
<210> 100	
<211> 60	
<212> DNA	

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 100

ttgctatttc tagctctaaa acaggtgaag ttcgaggcg acggtgttc gtcctttcca 60

<210> 101

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 101

ttgctatttc tagctctaaa actggtgaac cgcacgagc tgggtgttc gtcctttcca 60

<210> 102

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 102

ttgctatttc tagctctaaa actcagctcg atgcggttca ccggtgttc gtcctttcca 60

<210> 103

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 103

ttgctatttc tagctctaaa actgaagtcg atgccttca gcggtgttc gtcctttcca 60

<210> 104

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 104

ttgctatttc tagctctaaa actcctcctt gaagtcgatg ccggtgttc gtcctttcca 60

<210> 105

<211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 105
 ttgctatttc tagctctaaa acggatgttg ccgtcctcct tgggtgtttc gtcctttcca 60

<210> 106
 <211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 106
 ttgctatttc tagctctaaa acgcttgagg gagatgagga ctggtgtttc gtcctttcca 60

<210> 107
 <211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 107
 ttgctatttc tagctctaaa acatgactgt gaagagcttc acggtgtttc gtcctttcca 60

<210> 108
 <211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 108
 ttgctatttc tagctctaaa acgaggacaa agtacaaacg gcggtgtttc gtcctttcca 60

<210> 109
 <211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 109

ttgctatttc tagctctaaa acgaaccgga ggacaaagta cagggtgttc gtcctttcca 60

<210> 110

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 110

ttgctatttc tagctctaaa accaccacca acttcaccca cgggtgttc gtcctttcca 60

<210> 111

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 111

ttgctatttc tagctctaaa acgggcctca ccaccaactt cagggtgttc gtcctttcca 60

<210> 112

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 112

ttgctatttc tagctctaaa acgcccaggg cctcaccacc aagggtgttc gtcctttcca 60

<210> 113

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 113

ttgctatttc tagctctaaa acacctgcc agggcctcac cagggtgttc gtcctttcca 60

<210> 114

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 114
 ttgctatttc tagctctaaa actgatacca acctgcccag ggggtgtttc gtcctttcca 60

<210> 115
 <211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 115
 ttgctatttc tagctctaaa actaaacctg tcttgtaacc ttggtgtttc gtcctttcca 60

<210> 116
 <211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 116
 ttgctatttc tagctctaaa acgctctggc taaagaggga atggtgtttc gtcctttcca 60

<210> 117
 <211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 117
 ttgctatttc tagctctaaa accggtcttg gctaaagagg gaggtgtttc gtcctttcca 60

<210> 118
 <211> 60
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 118
 ttgctatttc tagctctaaa actctgcaca ccccggtctt ggggtgtttc gtcctttcca 60

<210> 119
 <211> 20
 <212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 119

tacggcaagc tgaccctgaa 20

<210> 120

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 120

gtccatgccg agagtgatcc 20

<210> 121

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 121

gccaggggct gttatcttgg 20

<210> 122

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 122

atgcacagaa gcacaggttg a 21

<210> 123

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 123

ctgtgtcctc ttctgcctc 20

<210> 124

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 124
 ctctccgagg agaaggccaa 20
 <210> 125
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 125
 ggtagaccac cagcagccta 20
 <210> 126
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 126
 cagtgccaga agagccaagg 20
 <210> 127
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 127
 ccacacagct tccggttctc 20
 <210> 128
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 128
 gagagccgtt ccctctttgc 20

<210> 129
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 129
 cctcccccatt ggcctgcttc 20

<210> 130
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 130
 tcgtcctgct ctcacttaga c 21

<210> 131
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 131
 ttttgtggct tggccccagt 20

<210> 132
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 132
 tgcagttctca tgacttggcc t 21

<210> 133
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide

<400> 133	
ttctgagggc tgctacctgt	20
<210> 134	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 134	
acatgaagca actccagtcc ca	22
<210> 135	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 135	
agcagaccca ctgagtcaac tg	22
<210> 136	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 136	
cccgccacag tcgtgtcat	19
<210> 137	
<211> 18	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 137	
cgccccgta caaggtga	18
<210> 138	
<211> 23	
<212> DNA	
<213> Artificial Sequence	

<220><223> Synthetic Polynucleotide

<400> 138

gtaccgtaca ttgtaggatg ttt

23

<210> 139

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 139

cctcatctcc ctcaagcagg c

21

<210> 140

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 140

attctgctct tgaggttatt tgt

23

<210> 141

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 141

cacctctgcc tcaagagcag aaaa

24

<210> 142

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 142

tgtgtgtgtg tgtgtgtagg act

23

<210> 143

<211> 20

<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 143	
tcacatgtgc ccctccctcc	20
<210> 144	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 144	
cgagaaggag gtgcaggag	19
<210> 145	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 145	
cgggagctgt tcagaggctg	20
<210> 146	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 146	
ctcacctggg cgagaaaggt	20
<210> 147	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 147	
aaaactcaaa gaaatgccca atca	24

<210> 148	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 148	
agacgctgct cgctccattc	20
<210> 149	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 149	
acaggcatga atcactgcac ct	22
<210> 150	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 150	
gcggcaactt cagacaaccg a	21
<210> 151	
<211> 18	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 151	
gaccaggagg caccagtt	18
<210> 152	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 152	

ctgccttcac tgcttaaaag tggat	25
<210> 153	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 153	
acagttgaag gaaggaaaca tgc	23
<210> 154	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 154	
gctgcatttg cccatttcca	20
<210> 155	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 155	
gttgggggag gaggagctta t	21
<210> 156	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 156	
ctaagagcta taagggaaca tgact	25
<210> 157	
<211> 152	
<212> DNA	
<213> Artificial Sequence	

<220><223> Synthetic Polynucleotide

<400> 157

ttctgagggc tgctacctgt acatctgcac aagattgcct ttactccatg cttttcttct	60
tctgctctaa ctctgacaat ctgtcttgcc atgccataag cccctattct ttctgtaacc	120
ccaagatggt ataaaagcat caatgattgg gc	152

<210> 158

<211> 155

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 158

aaaactcaaa gaaatgccca atcattgatg cttttatacc atcttggggt tacagaaaga	60
ataggggctt atggcatggc aagacagatt gtcagagtta gagcagaaga agaaaggcat	120
ggagtaaagg caatcttgtg cagatgtaca ggtaa	155

<210> 159

<211> 155

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 159

ttacctgtac atctgcacaa gattgccttt actccatgcc tttcttcttc tgctctaact	60
ctgacaatct gtcttgccat gccataagcc cctattcttt ctgtaacccc aagatggtat	120
aaaagcatca atgattgggc atttctttga gtttt	155

<210> 160

<211> 152

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 160

ttctgagggc tgctacctgt acatctgcac aagattgcct ttactccatg cttttcttct	60
tctgctctaa ctctgacaat ctgtcttgcc atgccataag cccctattct ttctgtaacc	120
ccaagatggt ataaaagcat caatgattgg gc	152

<210> 161

<211> 167
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 161
 ttctgagggc tgctacctgt acatctgcac aagattgcct ttactccatg cttttcttct 60

 tctgctctaa ctctgacaat ctgtcttgcc atgccataag cccctattct ttctgtaacc 120
 ccaagatggg ataaaagcat caatgattgg gcatttcttt gagtttt 167
 <210> 162
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 162
 ggccctgcttc gtggcaatgc 20
 <210> 163
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 163
 acctgggccca gggaggaggagg 20

 <210> 164
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 164
 ctcacttaga ctttctctcc 20
 <210> 165
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Synthetic Polynucleotide
 <400> 165
 ctcgaggtct agctcctgca 20
 <210> 166
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 166
 tggccccagt ctctcttcta 20
 <210> 167
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 167
 cagcctctga acagctcccg 20
 <210> 168
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 168
 tgacttgcc ttgtaggaa 20
 <210> 169
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 169
 gaggctactg aaacataagt 20
 <210> 170
 <211> 20

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 170
 tgctacctgt acatctgcac 20
 <210> 171
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 171
 catcaatgat tgggcatttc 20
 <210> 172
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 172
 actccagtcc caaatatgta 20
 <210> 173
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 173
 actagggggc gctcggccac 20
 <210> 174
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 174
 ctgagtcaac tgtaagcatt 20
 <210> 175

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 175
 ggccaggtgc agtgattcat 20

<210> 176
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 176
 tcgtgtcatc ttgtttgtgc 20

<210> 177
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 177
 ggcagagccc agcggacact 20

<210> 178
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 178
 caaggtgagc ctgggtctgt 20

<210> 179
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 179

atcactgccc aagaagtgca	20
<210> 180	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 180	
ttgtaggatg tttagcagca	20
<210> 181	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 181	
acttgctctc tttagagaac	20
<210> 182	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 182	
ctcaagcagg ccccgtggt	20
<210> 183	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 183	
ttttggacca aacctttttg	20
<210> 184	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	

<400> 184	
tgaggttatt tgtccattgt	20
<210> 185	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 185	
taaggggagt atttacacca	20
<210> 186	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 186	
tcaagagcag aaaatgtgac	20
<210> 187	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 187	
cttgcaggga ccttctgatt	20
<210> 188	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 188	
tgtgtgtagg actaaactct	20
<210> 189	
<211> 20	
<212> DNA	

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 189

gatagcagta tgaccttggg 20

<210> 190

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 190

gagtccgagc agaagaagaa ggg 23

<210> 191

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 191

gaggccgagc agaagaaaga cgg 23

<210> 192

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 192

gagtcctagc aggagaagaa gag 23

<210> 193

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 193

gagtctaagc agaagaagaa gag 23

<210> 194

<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 194	
gagtttagagc agaagaagaa agg	23
<210> 195	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 195	
gggtgggggg agtttgctcc tgg	23
<210> 196	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 196	
ggatggaggg agtttgctcc tgg	23
<210> 197	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 197	
gggagggtgg agtttgctcc tgg	23
<210> 198	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 198	
cgggggaggg agtttgctcc tgg	23

<210> 199
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 199
 ggggagggga agtttgctcc tgg 23

<210> 200
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 200
 gcagatgtag tgtttcaca ggg 23

<210> 201
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 201
 acaaagtag tatttcaca ggg 23

<210> 202
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 202
 ccagatgtag tattccaca ggg 23

<210> 203
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide

<400> 203	
ctagatgaag tgcttcaca tgg	23
<210> 204	
<211> 89	
<212> DNA	
<213> Homo sapiens	
<400> 204	
ccgcgccgag gtgaagttcg agggcgacac cctggtgaac cgcatcgagc tgaaggcat	60
cgacttcaag gaggacggca acatcctgg	89
<210> 205	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 205	
ccgcgccgag gtgaagttcg agg	23
<210> 206	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 206	
ccgaggtgaa gttcgagggc gac	23
<210> 207	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 207	
ccctggtgaa ccgcatcgag ctg	23
<210> 208	
<211> 23	
<212> DNA	

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 208

ggtgaaccgc atcgagctga agg 23

<210> 209

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 209

gctgaagggc atcgacttca agg 23

<210> 210

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 210

ggcatcgact tcaaggagga cgg 23

<210> 211

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 211

caaggaggac ggcaacatcc tgg 23

<210> 212

<211> 104

<212> DNA

<213> Homo sapiens

<400> 212

ggagcgacc atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt 60

cgaggcgac accctgggtga accgcatcga gctgaagggc atcg 104

<210> 213

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 213

accatcttct tcaaggacga cgg 23

<210> 214

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 214

caactacaag acccgcgccg agg 23

<210> 215

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 215

ccgcgccgag gtgaagttcg agg 23

<210> 216

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 216

gaagttcgag ggcgacaccc tgg 23

<210> 217

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 217

cccgcccgga ggtgaagttc gag 23

<210> 218
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 218
 cctggtgaac cgcatcgagc tga 23

<210> 219
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 219
 ccgcatcgag ctgaaggcca tcg 23

<210> 220
 <211> 68
 <212> DNA
 <213> Homo sapiens
 <400> 220
 ccccaagtct agcaagcagg ccaaagatgt ctcccgcatg cgctcagtcc tcatctccct 60
 caagcagg 68

<210> 221
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 221
 ccccaagtct agcaagcagg cca 23

<210> 222
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide

<400> 222	
agtcctcadc tccctcaagc agg	23
<210> 223	
<211> 65	
<212> DNA	
<213> Homo sapiens	
<400> 223	
ccctgtggaa acactacadc tgcaatatct taatcctact cagtgaagct cttcacagtc	60
attgg	65
<210> 224	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 224	
ccctgtggaa acactacadc tgc	23
<210> 225	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 225	
gtgaagctct tcacagtcac tgg	23
<210> 226	
<211> 93	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 226	
ccgttactgc cctgtggggc aaggtgaacg tggatgaagt tgggtggtgag gccctgggca	60
ggttggtatc aaggttacaac gacaggttta agg	93
<210> 227	
<211> 23	

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 227
 ccgttactgc cctgtggggc aag 23
 <210> 228
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 228
 cgtggatgaa gttggtggtg gg 22
 <210> 229
 <211> 23
 <212> DNA
 <213> Artificial Sequence

 <220><223> Synthetic Polynucleotide
 <400> 229
 tgaagttggt ggtgaggccc tgg 23
 <210> 230
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 230
 ttggtggtga ggccctgggc agg 23
 <210> 231
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 231
 tggtagggcc ctgggcaggt tgg 23

<210> 232	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 232	
ccctgggcag gttggtatca agg	23
<210> 233	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 233	
aaggttaca gacaggttta agg	23
<210> 234	
<211> 75	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 234	
ccctttcttct tctgctcgga ctcaggccct tcctcctcca gcttctgcg tttgtacttt	60
gtcctccggt tctgg	75
<210> 235	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 235	
ccctttcttct tctgctcgga etc	23
<210> 236	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	

<400> 236	
gccgtttgta cttgtcctc cgg	23
<210> 237	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 237	
tgtactttgt cctccggttc tgg	23
<210> 238	
<211> 73	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 238	
ccaggagcaa actccccca cccctttcc aaagccatt ccctcttag ccagagccgg	60
ggtgtgcaga cgg	73
<210> 239	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400>	
239	
ccaggagcaa actccccca ccc	23
<210> 240	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 240	
attccctctt tagccagagc cgg	23
<210> 241	
<211> 23	

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 241

tccctcttta gccagagccg ggg 23

<210> 242

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 242

ccagagccgg ggtgtgcaga cgg 23

<210> 243

<211> 106

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 243

gctgtttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc ccctttccaa 60

agcccatgcc ctctttagcc agagccgggg tgtgcagacg gcagtc 106

<210> 244

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 244

gctgtttggg aggtcagaaa taggggggtcc aggaagccgg ggtgtgcaga cggcagtc 58

<210> 245

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 245

gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc ccctttccaa 60
agcccatgcc ctcttttagcc ggggtgtgca gacggcagtc 100
<210> 246
<211> 59
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 246
gctgttttggg aggtcagaaa taggggggtcc aggagagccg ggggtgtgcag acggcagtc 59
<210> 247
<211> 45
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 247
gctgttttggg aggtcagaaa tagccgggggt gtgcagacgg cagtc 45
<210> 248
<211> 57
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 248
gctgttttggg aggtcagaaa taggggggtcc aggagccggg gtgtgcagac ggcagtc 57
<210> 249
<211> 54
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 249
gctgttttggg aggtcagaaa taggggggtcc agccgggggtg tgcagacggc agtc 54
<210> 250
<211> 100
<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 250

gctgtttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc ccctttccaa 60

agcccatgcc ctcttttagcc aggggtgtgca gacggcagtc 100

<210> 251

<211> 59

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 251

gctgtttggg aggtcagaaa taggggggtcc aggatagccg ggggtgtgcag acggcagtc 59

<210> 252

<211> 106

<212> DNA

<213>

> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 252

gctgtttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc ccctttccaa 60

agcccatgcc ctcttttagcc agagccggggg tgtgcagacg gcagtc 106

<210> 253

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 253

gctgtttggg aggtcagaaa taggggggtcc agacggcagt c 41

<210> 254

<211> 84

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 254

gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc ccctttccaa 60
agccgggggtg tgcagacggc agtc 84
<210> 255
<211> 77
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 255
gctgttttggg aggtcagaaa taggggggtcc aaagccatt ccctcttttag ccagagccgg 60
ggtggcagac ggcagtc 77
<210> 256
<211> 43
<212> DNA
<213> Artificial Sequence
<220><
223> Synthetic Polynucleotide
<400> 256
gctgttttggg aggtcagaaa taggggggtgt gcagacggca gtc 43
<210> 257
<211> 54
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 257
gctgttttggg aggtcagaaa taggggggtcc agccgggggtg tgcagacggc agtc 54
<210> 258
<211> 50
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 258
gctgttttggg aggtcagaaa taggggggtcc aggggtgtgca gacggcagtc 50
<210> 259
<211> 57

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 259
 gctgttttggg aggtcagaaa gaggggggtcc aggagccggg gtgtgcagac ggcagtc 57
 <210> 260
 <211> 45
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 260
 gctgttttggg aggtcagaaa tagccgggggt gtgcagacgg cagtc 45
 <210> 261
 <211> 106
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 261
 gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc ccctttccaa 60
 agcccatgcc ctctttagcc agagccggggg tgtgcagacg gcagtc 106
 <210> 262
 <211> 94
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 262
 gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccaag cccattccct 60
 cttagccag agccgggggtg tgcagacggc agtc 94
 <210> 263
 <211> 95
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 263

gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc cccttttccc	60
tcttttagcca gagccggggt gtgcagacgg cagtc	95
<210> 264	
<211> 94	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 264	
gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc cccttttcct	60
cttttagccag agccggggtg tgcagacggc agtc	94
<210> 265	
<211> 93	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 265	
gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc ccatttcctc	60
ttagccaga gccggggtgt gcagacggca gtc	93
<210> 266	
<211> 94	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 266	
gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc ccatttcct	60
cttttagccag agccggggtg tgcagacggc agtc	94
<210> 267	
<211> 86	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 267	

gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc cccttttagcc 60

agagccggggg tgtgcagacg gcagtc 86

<210> 268

<211> 103

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 268

gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tccccccacc cccttaaagc 60

ccattccctc tttagccaga gccgggggtgt gcagacggca gtc 103

<210> 269

<211> 93

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 269

gctgttttggg aggtcagaaa taggggggtcc aggagcaaac tcccccaagc ccattccctc 60

tttagccaga gccgggggtgt gcagacggca gtc 93

<210> 270

<211> 132

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 270

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccacagga 60

gcaaactccc tccatccac aaatccgtcc ttagatgtgc acaccaacc tcctaagaaa 120

tagaaggatg at 132

<210> 271

<211> 130

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 271
cattcaacag atacttactg aatgctaata tctcagacag gacattctga caccaccaggc 60
aaactccctc catcccacaa atccgtcctt agatgtgcac acccaacctc ctaagaaata 120
gaaggatgat 130
<210> 272
<211> 123
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 272
cattcaacag atacttactg aatgctaata tctcagacag gacattctga caccaccatcc 60
ctccatccca caaatccgtc cttagatgtg cacaccaaac ctctaagaa atagaaggat 120
gat 123
<210> 273
<211> 129
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 273
cattcaacag atacttactg aatgctaata tctcagacag gacattctga caccaccagca 60
aactccctcc atcccacaaa tccgtcctta gatgtgcaca cccaacctcc taagaaatag 120
aaggatgat 129
<210> 274
<211> 110
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 274
cattcaacag atacttactg aatgctaata tctcagacag gacattctga caccaccacaa 60
atccgtcctt agatgtgcac acccaacctc ctaagaaata gaaggatgat 110
<210> 275
<211> 119
<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 275

cattcaacag atacttactg aatgctaag tctcagacag gacattctga cactccctcc 60

atcccacaaa tccgtcctta gatgtgcaca cccaacctcc taagaaatag aaggatgat 119

<210> 276

<211> 118

<212> DNA

<213> Artificial Sequence

<220>

><223> Synthetic Polynucleotide

<400> 276

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccctcca 60

tcccacaaat ccgtccttag atgtgcacac ccaacctcct aagaaataga aggatgat 118

<210> 277

<211> 64

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 277

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccctcagga 60

tgat 64

<210> 278

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 278

cattcaacag atacttactg aatgctaag tctcagacag gacattctga ctccctccat 60

cccacaaatc cgtccttaga tgtgcacacc caacctccta agaaatagaa ggatgat 117

<210> 279

<211> 132

<212> DNA

<213> Homo sapiens

<400> 279
cattcaacag atacttactg aatgctaata tctcagacag gacattctga caccaccagga 60
gcaaactccc tccatccac aaatccgtcc ttagatgtgc acaccaacc tcctaagaaa 120
tagaaggatg at 132

<210> 280

<211> 111

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 280

cattcaacag atacttactg aatgctaata tctcagacag gacattctga caccaccagga 60
catccgtcct tagatgtgca cacccaacct cctaagaaat agaaggatga t 111

<210> 281

<211> 117

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 281

cattcaacag atacttactg aatgctaata tctcagacag gacattctga ctcctccat 60
cccacaaatc cgtccttaga tgtgcacacc caacctcta agaaatagaa ggatgat 117

<210> 282

<211> 99

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 282

cattcaacag atacttactg aatgctaata tctcagacag gacattctga tccgtcctta 60
gatgtgcaca cccaacctcc taagaaatag aaggatgat 99

<210> 283

<211> 121

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 283

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccacagca 60

tctgatgaca aatccgtcct tagatgtgca cacccaacct cctaagaaat agaaggatga 120

t 121

<210> 284

<211> 130

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 284

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccacagga 60

gcaaactccc tccatccac aaatccgtcc ttagatgtgc acccaacctc ctaagaaata 120

gaaggatgat 130

<210> 285

<211> 130

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 285

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccacagga 60

gcaaactccc tccatccac aaatccgtcc ttagatgtgc acccaacctc ctaagaaata 120

gaaggatgat 130

<210> 286

<211> 132

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 286

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccacagga 60

gcaaactccc tccatccac aaatccgtcc ttagatgtgc acaccaacc tcctaagaaa 120

tagaaggatg at 132

<210> 287

<211> 101
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 287
 cattcaacag atacttactg aatgctaag tctcagacag gacattctga aatccgtcct 60
 tagatgtgca cacccaacct cctaagaaat agaaggatga t 101
 <210> 288
 <211> 132
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 288
 cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccccagga 60
 gcaaactccc tccatccac aaatccgtcc ttagatgtgc acaccaacc tcctaagaaa 120
 tagaaggatg at 132
 <210> 289
 <211> 119
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 289
 cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccccagga 60
 gcaaactccc tccgtcctta gatgtgcaca cccaacctcc taagaaatag aaggatgat 119
 <210> 290
 <211> 121
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 290
 cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccccagga 60
 gcaaactccc tctccgtcct tagatgtgca cacccaacct cctaagaaat agaaggatga 120
 t 121

<210> 291
 <211> 124
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 291
 cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccccagga 60
 gcaaactccc tccattccgt ccttagatgt gcacacccaa cctcctaaga aatagaagga 120

 tgat 124
 <210> 292
 <211> 130
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 292
 cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccccagga 60
 gcaaactccc tccatccac aaatccgtcc ttagatgtgc acccaacctc ctaagaaata 120
 gaaggatgat 130
 <210> 293
 <211> 122
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide

 <400> 293
 cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccccagga 60
 gcaaactccc tctccgtcc ttagatgtgc acaccaacc tcctaagaaa tagaaggatg 120
 at 122
 <210> 294
 <211> 123
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 294

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccacagga 60
gcaaactccc tccatccgtc cttagatgtg cacaccaac ctctaagaa atagaaggat 120

gat 123

<210> 295
<211> 128
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 295

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccacagga 60
gcaaactccc tccatccac aaatccgtcc ttagatgtgc acaccctcct aagaaataga 120
aggatgat 128

<210> 296
<211> 130
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide

<400> 296

cattcaacag atacttactg aatgctaag tctcagacag gacattctga caccacagga 60
gcaaactccc tccatccac aaatccgtcc agatgtgcac acccaacctc ctaagaaata 120
gaaggatgat 130

<210> 297
<211> 21
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic Polynucleotide
<400> 297

ggggtggggg gagtttgctc c 21

<210> 298
<211> 22
<212> DNA
<213> Artificial Sequence
<220><

223> Synthetic Polynucleotide
 <400> 298
 aggagcaaac tccctccatc cc 22
 <210> 299
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 299
 gggatggagg gagtttgctc ct 22
 <210> 300
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 300
 auucccucuu uagccagagc 20
 <210> 301
 <211> 4
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic Polypeptide
 <400> 301
 Gly Gly Ser Met
 1
 <210> 302
 <211> 99
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic Polynucleotide
 <400> 302
 ccagcaagag gctcccgagc gagcaagctc agtttacacc cgatccactg gggagcagga 60
 aatatctgtg ggcttgtgac acggactcaa gtgggctgg 99
 <210> 303

<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 303	
ccagcaagag gctcccgagc gag	23
<210> 304	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 304	
cccgagcgag caagctcagt tta	23
<210> 305	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 305	
cccgatccac tggggagcag gaa	23
<210> 306	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 306	
agtttacacc cgatccactg ggg	23
<210> 307	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 307	

tggggagcag gaaatatctg tggg	24
<210> 308	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 308	
atatctgtgg gcttgtgaca cgg	23
<210> 309	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 309	
gcttgtgaca cggactcaag tgg	23
<210> 310	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 310	
tgacacggac tcaagtgggc tgg	23
<210> 311	
<211> 120	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 311	
cctggccatc tctgacctgt ttttcttct tactgtcccc ttctgggctc actatgctgc	60
cgccagtg gactttggaa atacaatgtg tcaactcttg acagggtct attttatagg	120
<210> 312	
<211> 23	
<212> DNA	

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 312

cctggccatc tctgacctgt ttt 23

<210> 313

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 313

ccccttctgg gctcactatg ctg 23

<210> 314

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 314

ccgcccagtg ggactttgga aat 23

<210> 315

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 315

tcactatgct gccgcccagt ggg 23

<210> 316

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 316

caatgtgtca actcttgaca ggg 23

<210> 317

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 317

ttgacagggc tctatatttat agg 23

<210> 318

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 318

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val Pro

1 5 10 15

<210> 319

<211> 4869

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 319

atggactaca aagaccatga cgggtgattat aaagatcatg acatcgatta caaggatgac 60

gatgacaaga tggcccccaa gaagaagagg aaggtgggca ttcaccgcgg ggtacctgga 120

ggtttctggat cccaactagt caaaagtga ctggaggaga agaaatctga acttcgtcat 180

aaattgaaat atgtgcctca tgaatatatt gaattaattg aaattgccag aaattccact 240

caggatagaa ttcttgaaat gaagtaatg gaatttttta tgaaagtta tggatataga 300

ggtaaacatt tgggtggatc aaggaaaccg gacggagcaa tttatactgt cggatctcct 360

attgattacg gtgtgatcgt ggatactaaa gcttatagcg gaggttataa tctgccaatt 420

ggccaagcag atgaaatgca acgatatgtc gaagaaaatc aaacacgaaa caaacatatc 480

aaccctaatt aatgggtggaa agtctatcca tcttctgtaa cggaatttaa gtttttattt 540

gtgagtggtc actttaaaagg aaactacaaa gtcagctta cagattaaa tcatatcact 600

aattgtaatg gagctgttct tagttagtaa gagcttttaa ttggtggaga aatgattaaa 660

gccggcacat taacctaga ggaagtcaga cggaaattta ataacggcga gataaacttt 720

agcggcagcg agactcccgg gacctcagag tccgccacac ccgaaagtga taaaaagtat 780

tctattgggt tagctatcgg cactaattcc gttggatggg ctgtcataac cgatgaatac 840

aaagtacctt caaagaaatt taagtggttg gggaacacag accgtcattc gattaaaaag 900

aatcttatcg gtgccctcct attcgatagi ggcgaaacgg cagaggcgac tcgacctgaaa 960

cgaaccgctc ggagaaggta tacacgtcgc aagaaccgaa tatgttactt acaagaaatt 1020

tttagcaatg agatggccaa agttgacgat tctttcttcc accgtttgga agagtccttc 1080

cttgtcgaag aggacaagaa acatgaacgg caccatctt ttggaaacat agtagatgag 1140

gtggcatatc atgaaaagta cccaacgatt tatcacctca gaaaaaagct agttgactca 1200

actgataaag cggacctgag gttaatctac ttggctcttg cccatatgat aaagtccgt 1260

gggcactttc tcattgaggg tgatctaaat ccggacaact cggatgtcga caaactgttc 1320

atccagttag tacaaaccta taatcagttg ttgaagaga accctataaa tgcaagtggc 1380

gtggatgcga aggtctatct tagcggccgc ctctctaaat cccgacggct agaaaacctg 1440

atcgacaat taccgggaga gaagaaaaat gggttggtcg gtaaccttat agcgctctca 1500

ctaggcctga caccaaatit taagtcgaac ttgcacttag ctgaagatgc caaattgcag 1560

cttagtaagg acacgtacga tgacgatctc gacaatctac tggcacaat tggagatcag 1620

tatgcggact tatTTTTTgc tgccaaaaac cttagcgatg caatcctcct atctgacata 1680

ctgagagtta atactgagat taccaaggcg ccgttatccg cttcaatgat caaaaggtag 1740

gatgaacatc accaagactt gacacttctc aaggccctag tccgtcagca actgcctgag 1800

aaatataagg aaatattctt tgatcagtcg aaaaacgggt acgcaggtta tattgacggc 1860

ggagcgagtc aagaggaatt ctacaagttt atcaaacca tattagagaa gatggatggg 1920

acggaagagt tgcttgtaaa actcaatcgc gaagatctac tgcgaaagca gcggactttc 1980

gacaacggta gcattccaca tcaaatccac ttaggcgaat tgcatgctat acttagaagg 2040

caggaggatt ttatccgtt cctcaaagac aatcgtgaaa agattgagaa aatcctaacc 2100

tttcgcatac cttactatgt gggaccctg gcccgaggga actctcggtt cgcatggatg 2160

acaagaaagt ccgaagaaac gattactcca tggaattttg aggaagttgt cgataaagg 2220

gcgtcagctc aatcgttcat cgagaggatg accaactttg acaagaattt accgaacgaa 2280

aaagtattgc ctaagcacag ttacttttac gattatttca cagtgtacaa tgaactcacg 2340

aaagttaagt atgtcactga gggcatgcgt aaaccgcct ttctaagcgg agaacagaag 2400

aaagcaatag tagatctgtt attcaagacc aaccgcaaag tgacagttaa gcaattgaaa 2460

gaggactact ttaagaaaat tgaatgcttc gattctgtcg agatctccgg ggtagaagat 2520
cgatttaatg cgtcacttgg tacgtatcat gacctcctaa agataattaa agataaggac 2580
ttcctggata acgaagagaa tgaagataatc ttagaagata tagtggtgac tcttaccctc 2640
tttgaagatc gggaaatgat tgaggaaaga ctaaaaacat acgctcacct gttcgacgat 2700
aaggttatga aacagttaaa gaggcgtcgc tatacgggct ggggacgatt gtcgcggaaa 2760
cttatcaacg ggataagaga caagcaaagt ggtaaaaacta ttctcgattt tctaaagagc 2820
gacggcttcg ccaataggaa ctttatgcag ctgatccatg atgactcttt aaccttcaaa 2880

gaggatatac aaaaggcaca ggtttccgga caaggggact cattgcacga acatattgcg 2940
aatcttgctg gttcgccagc catcaaaaag ggcatactcc agacagtcaa agtagtggat 3000
gagctagtta aggtcatggg acgtcacaaa ccggaaaaca ttgtaatcga gatggcacgc 3060
gaaaatcaaa cgactcagaa ggggcaaaaa aacagtcgag agcggatgaa gagaatagaa 3120
gaggggtatta aagaactggg cagccagatc ttaaaggagc atcctgtgga aaatacccaa 3180
ttgcagaacg agaaacttta cctctattac ctacaaaatg gaaggacat gtatgttgat 3240
caggaactgg acataaacgg tttatctgat tacgacgtcg atgccattgt accccaatcc 3300

tttttgaagg acgattcaat cgacaataaa gtgcttacac gctcggataa gaaccgaggg 3360
aaaagtgaca atgttccaag cgaggaagtc gtaaagaaaa tgaagaacta ttggcggcag 3420
ctcctaaatg cgaaactgat aacgcaaaga aagttcgata acttaactaa agctgagagg 3480
ggtggcttgt ctgaacttga caaggccgga tttattaaac gtcagctcgt ggaaaccgcg 3540
caaatcacia agcatgttgc acagatacta gattcccgaa tgaatacgaa atacgacgag 3600
aacgataagc tgattcggga agtcaaagta atcactttaa agtcaaaatt ggtgtcggac 3660
ttcagaaagg attttcaatt ctataaagtt agggagataa ataactacca ccatgcgcac 3720

gacgcttate ttaatgccgt cgtagggacc gcactcatia agaaataccc gaagctagaa 3780
agtgagtttg tglatggiga ttacaaagtt tatgacgtcc gtaagatgat cgcgaaaagc 3840
gaacaggaga taggcaaggc tacagccaaa tacttctttt attctaakat tatgaatttc 3900
tttaagacgg aaatcactct ggcaaacgga gagatacgca aacgacctt aattgaaacc 3960
aatggggaga caggtgaaat cgtatgggat aagggccggg acttcgcgac ggtgagaaaa 4020
gttttgtcca tgcccaagt caacatagta aagaaaactg aggtgcagac cggagggttt 4080
tcaaaggaat cgattcttcc aaaaaggaat agtgataagc tcatcgctcg taaaaaggac 4140

tgggaccgga aaaagtacgg tggcttcgat agccctacag ttgcctattc tgtcctagta 4200
tgggcaaaag ttgagaaggg aaaatccaag aaactgaagt cagtcaaaga attattgggg 4260
ataacgatta tggagcgctc gtcttttgaa aagaaccca tcgacttcct tgaggcgaaa 4320

ggttacaagg aagtaaaaaa ggatctcata attaaactac caaagtatag tctgtttgag 4380
 ttagaaaatg gccgaaaacg gatgttggct agcgccggag agcttcaaaa ggggaacgaa 4440
 ctgcactac cgtctaaata cgtgaatttc ctgtatttag cgtccatta cgagaagttg 4500
 aaaggttcac ctgaagataa cgaacagaag caactttttg ttgagcagca caaacattat 4560

ctcgacgaaa tcatagagca aatttcggaa ttcagtaaga gattcatcct agctgatgcc 4620
 aatctggaca aagtattaag cgcatacaac aagcacaggg ataaacccat acgtgagcag 4680
 gcggaataa ttatccattt gtttactctt accaacctcg gcgctccagc cgcattcaag 4740
 tattttgaca caacgataga tcgcaaacga tacacttcta ccaaggaggt gctagacgcg 4800
 acactgattc accaatccat cacgggatta tatgaaactc ggatagattt gtcacagctt 4860
 ggggggtgac 4869

<210> 320

<211> 1367

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 320

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115	120	125	
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser			
130	135	140	
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met			
145	150	155	160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp			
165	170	175	
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn			
180	185	190	
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys			
195	200	205	
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu			
210	215	220	
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu			
225	230	235	240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp			
245	250	255	
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp			
260	265	270	
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu			
275	280	285	
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile			
290	295	300	
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met			
305	310	315	320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala			
325	330	335	
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp			
340	345	350	
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln			
355	360	365	

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

370

375

380

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

385

390

395

400

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

405

410

415

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

420

425

430

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

435

440

445

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

450

455

460

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

465

470

475

480

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

485

490

495

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

500

505

510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515

520

525

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

530

535

540

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

545

550

555

560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565

570

575

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

580

585

590

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

595

600

605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620
 Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

 625 630 635 640
 Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

 645 650 655
 Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

 660 665 670
 Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

 675 680 685
 Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

 690 695 700
 Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
 705 710 715 720
 Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

 725 730 735
 Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

 740 745 750
 His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

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

 805 810 815
 Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

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

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
 865 870 875 880
 Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
 885 890 895
 Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
 900 905 910
 Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
 915 920 925
 His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
 930 935 940
 Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
 945 950 955 960
 Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
 965 970 975
 Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
 980 985 990
 Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
 995 1000 1005
 Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
 1010 1015 1020
 Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
 1025 1030 1035
 Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
 1040 1045 1050
 Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
 1055 1060 1065
 Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
 1070 1075 1080
 Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
 1085 1090 1095
 Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100	1105	1110
Asn Ser Asp Lys Leu Ile Ala	Arg Lys Lys Asp Trp	Asp Pro Lys
1115	1120	1125
Lys Tyr Gly Gly Phe Asp Ser	Pro Thr Val Ala Tyr	Ser Val Leu
1130	1135	1140
Val Val Ala Lys Val Glu Lys	Gly Lys Ser Lys Lys	Leu Lys Ser
1145	1150	1155
Val Lys Glu Leu Leu Gly Ile	Thr Ile Met Glu Arg	Ser Ser Phe
1160	1165	1170
Glu Lys Asn Pro Ile Asp Phe	Leu Glu Ala Lys Gly	Tyr Lys Glu
1175	1180	1185
Val Lys Lys Asp Leu Ile Ile	Lys Leu Pro Lys Tyr	Ser Leu Phe
1190	1195	1200
Glu Leu Glu Asn Gly Arg Lys	Arg Met Leu Ala Ser	Ala Gly Glu
1205	1210	1215
Leu Gln Lys Gly Asn Glu Leu	Ala Leu Pro Ser Lys	Tyr Val Asn
1220	1225	1230
Phe Leu Tyr Leu Ala Ser His	Tyr Glu Lys Leu Lys	Gly Ser Pro
1235	1240	1245
Glu Asp Asn Glu Gln Lys Gln	Leu Phe Val Glu Gln	His Lys His
1250	1255	1260
Tyr Leu Asp Glu Ile Ile Glu	Gln Ile Ser Glu Phe	Ser Lys Arg
1265	1270	1275
Val Ile Leu Ala Asp Ala Asn	Leu Asp Lys Val Leu	Ser Ala Tyr
1280	1285	1290
Asn Lys His Arg Asp Lys Pro	Ile Arg Glu Gln Ala	Glu Asn Ile
1295	1300	1305
Ile His Leu Phe Thr Leu Thr	Asn Leu Gly Ala Pro	Ala Ala Phe
1310	1315	1320
Lys Tyr Phe Asp Thr Thr Ile	Asp Arg Lys Arg Tyr	Thr Ser Thr
1325	1330	1335

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

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 321

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 321

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

1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys

20

<210> 322

<211> 432

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 322

atggccctgt ttggctacgc acgcgtgtct accagtcaac agtcactcga ttgcaagtg 60

agggtcttta aagatgccgg agtgaaggca aacagaattt ttactgataa ggccagcgga 120

agcagcacag acagagaggg gctggatctc ctgagaatga aggtaaagga gggatgatgtg 180

atcttgggtca aaaaattgga tcgactgggg agagacacag ctgatatgct tcagcttatt 240

aaagagtttg acgtcaggg tgttgccgtg aggtttatcg atgacggcat ctcaaccgac 300

tcctacattg gtcttatgtt tgtgacaatt ttgtccgtg tggctcaggc tgagcggaga 360

aggattctcg aaaggacgaa tgagggacgg caagcagcta agttgaaagg tatcaaattt 420

ggcagacgaa gg 432

<210> 323

<211> 414

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 323

atggcaacca ttggctacat aagggtgtct accatcgacc aaaatatcga cctgcagcgc	60
aacgctctga catccgccaa ctgcgatcgg atcttcgagg ataggatcag tggcaagatc	120
gccaaaccggc ccggtctgaa gcgggctctg aagtacgtga ataagggcga tactctggtt	180
gtgtggaagt tggatcgctt gggtagatca gtgaagaatc tcgtagccct gataagcgag	240
ctgcacgaga ggggtgcaca ttccattct ctgaccgatt ccatcgatac gtctagcgcc	300
atgggccgat tcttctttta cgtcatgtcc gcctcgtcg aaatggagcg cgaacttatt	360

gttgaacgga ctttggctgg actggcagcg gctagagcac agggccgact tgga	414

<210> 324

<211> 432

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 324

atgctcattg gctatgtaag ggtcagcacc aatgacaaa acacagactt gcaacgcaat	60
gcttttggtt gcgccgatg tgaacagata ttgaagata aactgagcgg cactcggaca	120
gacagacctg ggcttaagag agcactgaaa agactgcaga agggggacac cctggtcgtc	180
tggaaactgg atcgccctcg acgcagcatg aaacatctga ttagcctggt tggtagctt	240
agggagagag gaatcaactt cagaagcctg accgactcca tcgacaccag tagcccatg	300

ggacgattct tcttctatgt gatgggagca cttgctgaga tggaaagaga gcttattatc	360

gaaagaacta tgcttggtat cgctgctgcc cggaacaaag gcagacggtt cggcagaccg	420

ccgaagagcg gc	432

<210> 325

<211> 144

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 325

Met Ala Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu

1	5	10	15

Asp Leu Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg

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

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

<210> 326

<211> 138

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 326

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

85 90 95
 Thr Ser Ser Ala Met Gly Arg Phe Phe Phe Tyr Val Met Ser Ala Leu
 100 105 110
 Ala Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Leu Ala Gly Leu

115 120 125
 Ala Ala Ala Arg Ala Gln Gly Arg Leu Gly

130 135
 <210> 327
 <211> 144
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic Polypeptide

<400> 327
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp

1 5 10 15
 Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30

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

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
 100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Ile Ala
 115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly
 130 135 140

<210> 328

<211> 1537

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 328

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

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

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

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

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

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

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

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

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

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

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

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

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

945	950	955	960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg			
	965	970	975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val			
	980	985	990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe			
	995	1000	1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala			
1010	1015	1020	
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe			
1025	1030	1035	
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala			
1040	1045	1050	
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu			
1055	1060	1065	
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val			
1070	1075	1080	
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr			
1085	1090	1095	
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys			
1100	1105	1110	
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro			
1115	1120	1125	
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val			
1130	1135	1140	
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys			
1145	1150	1155	
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser			
1160	1165	1170	
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys			
1175	1180	1185	

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val
1370 1375 1380

Pro Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Ala Leu Phe Gly
1385 1390 1395

Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu Asp Leu Gln Val
1400 1405 1410

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

1415 1420 1425
 Asp Lys Ala Ser Gly Ser Ser Thr Asp Arg Glu Gly Leu Asp Leu
 1430 1435 1440

 Leu Arg Met Lys Val Lys Glu Gly Asp Val Ile Leu Val Lys Lys
 1445 1450 1455
 Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Leu Gln Leu Ile
 1460 1465 1470
 Lys Glu Phe Asp Ala Gln Gly Val Ala Val Arg Phe Ile Asp Asp
 1475 1480 1485
 Gly Ile Ser Thr Asp Ser Tyr Ile Gly Leu Met Phe Val Thr Ile
 1490 1495 1500

 Leu Ser Ala Val Ala Gln Ala Glu Arg Arg Arg Ile Leu Glu Arg
 1505 1510 1515
 Thr Asn Glu Gly Arg Gln Ala Ala Lys Leu Lys Gly Ile Lys Phe
 1520 1525 1530
 Gly Arg Arg Arg
 1535
 <210> 329
 <211> 1537
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic Polypeptide
 <400> 329
 Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val Pro
 1 5 10 15

 Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
 20 25 30
 Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
 35 40 45
 Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 50 55 60
 Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

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

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
325 330 335

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
340 345 350

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

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
370 375 380

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
385 390 395 400

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
405 410 415

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
420 425 430

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
435 440 445

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
450 455 460

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
465 470 475 480

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
485 490 495

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
500 505 510

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
515 520 525

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
530 535 540

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
545 550 555 560

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

565 570 575
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 580 585 590

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

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

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

 Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 785 790 795 800
 Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 805 810 815

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

820 825 830

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

835 840 845

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

850 855 860

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

865 870 875 880

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035

Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe

1040 1045 1050

Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu

1055	1060	1065
Ala Asn Gly Glu Ile Arg Lys	Arg Pro Leu Ile Glu	Thr Asn Gly
1070	1075	1080
Glu Thr Gly Glu Ile Val Trp	Asp Lys Gly Arg Asp	Phe Ala Thr
1085	1090	1095
Val Arg Lys Val Leu Ser Met	Pro Gln Val Asn Ile	Val Lys Lys
1100	1105	1110
Thr Glu Val Gln Thr Gly Gly	Phe Ser Lys Glu Ser	Ile Leu Pro
1115	1120	1125
Lys Arg Asn Ser Asp Lys Leu	Ile Ala Arg Lys Lys	Asp Trp Asp
1130	1135	1140
Pro Lys Lys Tyr Gly Gly Phe	Asp Ser Pro Thr Val	Ala Tyr Ser
1145	1150	1155
Val Leu Val Val Ala Lys Val	Glu Lys Gly Lys Ser	Lys Lys Leu
1160	1165	1170
Lys Ser Val Lys Glu Leu Leu	Gly Ile Thr Ile Met	Glu Arg Ser
1175	1180	1185
Ser Phe Glu Lys Asn Pro Ile	Asp Phe Leu Glu Ala	Lys Gly Tyr
1190	1195	1200
Lys Glu Val Lys Lys Asp Leu	Ile Ile Lys Leu Pro	Lys Tyr Ser
1205	1210	1215
Leu Phe Glu Leu Glu Asn Gly	Arg Lys Arg Met Leu	Ala Ser Ala
1220	1225	1230
Gly Glu Leu Gln Lys Gly Asn	Glu Leu Ala Leu Pro	Ser Lys Tyr
1235	1240	1245
Val Asn Phe Leu Tyr Leu Ala	Ser His Tyr Glu Lys	Leu Lys Gly
1250	1255	1260
Ser Pro Glu Asp Asn Glu Gln	Lys Gln Leu Phe Val	Glu Gln His
1265	1270	1275
Lys His Tyr Leu Asp Glu Ile	Ile Glu Gln Ile Ser	Glu Phe Ser
1280	1285	1290

Lys Arg Val Ile Leu Ala Asp	Ala Asn Leu Asp Lys	Val Leu Ser
1295	1300	1305
Ala Tyr Asn Lys His Arg Asp	Lys Pro Ile Arg Glu	Gln Ala Glu
1310	1315	1320
Asn Ile Ile His Leu Phe Thr	Leu Thr Asn Leu Gly	Ala Pro Ala
1325	1330	1335
Ala Phe Lys Tyr Phe Asp Thr	Thr Ile Asp Arg Lys	Arg Tyr Thr
1340	1345	1350
Ser Thr Lys Glu Val Leu Asp	Ala Thr Leu Ile His	Gln Ser Ile
1355	1360	1365
Thr Gly Leu Tyr Glu Thr Arg	Ile Asp Leu Ser Gln	Leu Gly Gly
1370	1375	1380
Asp Gly Gly Ser Gly Gly Ser	Gly Gly Ser Met Ala	Leu Phe Gly
1385	1390	1395
Tyr Ala Arg Val Ser Thr Ser	Gln Gln Ser Leu Asp	Leu Gln Val
1400	1405	1410
Arg Ala Leu Lys Asp Ala Gly	Val Lys Ala Asn Arg	Ile Phe Thr
1415	1420	1425
Asp Lys Ala Ser Gly Ser Ser	Thr Asp Arg Glu Gly	Leu Asp Leu
1430	1435	1440
Leu Arg Met Lys Val Lys Glu	Gly Asp Val Ile Leu	Val Lys Lys
1445	1450	1455
Leu Asp Arg Leu Gly Arg Asp	Thr Ala Asp Met Leu	Gln Leu Ile
1460	1465	1470
Lys Glu Phe Asp Ala Gln Gly	Val Ala Val Arg Phe	Ile Asp Asp
1475	1480	1485
Gly Ile Ser Thr Asp Ser Tyr	Ile Gly Leu Met Phe	Val Thr Ile
1490	1495	1500
Leu Ser Ala Val Ala Gln Ala	Glu Arg Arg Arg Ile	Leu Glu Arg
1505	1510	1515
Thr Asn Glu Gly Arg Gln Ala	Ala Lys Leu Lys Gly	Ile Lys Phe

1520 1525 1530

Gly Arg Arg Arg

1535

<210> 330

<211> 1537

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 330

Met Ala Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu

1 5 10 15

Asp Leu Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg

20 25 30

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

35 40 45

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

50 55 60

Lys Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Leu Gln Leu Ile

65 70 75 80

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

85 90 95

Ile Ser Thr Asp Ser Tyr Ile Gly Leu Met Phe Val Thr Ile Leu Ser

100 105 110

Ala Val Ala Gln Ala Glu Arg Arg Arg Ile Leu Glu Arg Thr Asn Glu

115 120 125

Gly Arg Gln Ala Ala Lys Leu Lys Gly Ile Lys Phe Gly Arg Arg Arg

130 135 140

Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Asp Lys Lys Tyr Ser Ile

145 150 155 160

Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp

165 170 175

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

180	185	190	
Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser			
195	200	205	
Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg			
210	215	220	
Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser			
225	230	235	240
Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu			
245	250	255	
Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe			
260	265	270	
Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile			
275	280	285	
Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu			
290	295	300	
Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His			
305	310	315	320
Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys			
325	330	335	
Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn			
340	345	350	
Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg			
355	360	365	
Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly			
370	375	380	
Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly			
385	390	395	400
Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys			
405	410	415	
Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu			
420	425	430	

Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn
435 440 445

Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu
450 455 460

Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu
465 470 475 480

His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu
485 490 495

Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr
500 505 510

Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe
515 520 525

Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val
530 535 540

Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn
545 550 555 560

Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu
565 570 575

Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys
580 585 590

Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu
595 600 605

Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu
610 615 620

Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser
625 630 635 640

Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro
645 650 655

Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr
660 665 670

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

675	680	685	
Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu			
690	695	700	
Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp			
705	710	715	720
Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val			
725	730	735	
Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys			
740	745	750	
Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile			
755	760	765	
Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met			
770	775	780	
Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val			
785	790	795	800
Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser			
805	810	815	
Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile			
820	825	830	
Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln			
835	840	845	
Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala			
850	855	860	
Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu			
865	870	875	880
Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val			
885	890	895	
Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile			
900	905	910	
Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys			
915	920	925	

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

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

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

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

1160	1165	1170
Met Ile Ala Lys Ser Glu Gln	Glu Ile Gly Lys Ala	Thr Ala Lys
1175	1180	1185
Tyr Phe Phe Tyr Ser Asn Ile	Met Asn Phe Phe Lys	Thr Glu Ile
1190	1195	1200
Thr Leu Ala Asn Gly Glu Ile	Arg Lys Arg Pro Leu	Ile Glu Thr
1205	1210	1215
Asn Gly Glu Thr Gly Glu Ile	Val Trp Asp Lys Gly	Arg Asp Phe
1220	1225	1230
Ala Thr Val Arg Lys Val Leu	Ser Met Pro Gln Val	Asn Ile Val
1235	1240	1245
Lys Lys Thr Glu Val Gln Thr	Gly Gly Phe Ser Lys	Glu Ser Ile
1250	1255	1260
Leu Pro Lys Arg Asn Ser Asp	Lys Leu Ile Ala Arg	Lys Lys Asp
1265	1270	1275
Trp Asp Pro Lys Lys Tyr Gly	Gly Phe Asp Ser Pro	Thr Val Ala
1280	1285	1290
Tyr Ser Val Leu Val Val Ala	Lys Val Glu Lys Gly	Lys Ser Lys
1295	1300	1305
Lys Leu Lys Ser Val Lys Glu	Leu Leu Gly Ile Thr	Ile Met Glu
1310	1315	1320
Arg Ser Ser Phe Glu Lys Asn	Pro Ile Asp Phe Leu	Glu Ala Lys
1325	1330	1335
Gly Tyr Lys Glu Val Lys Lys	Asp Leu Ile Ile Lys	Leu Pro Lys
1340	1345	1350
Tyr Ser Leu Phe Glu Leu Glu	Asn Gly Arg Lys Arg	Met Leu Ala
1355	1360	1365
Ser Ala Gly Glu Leu Gln Lys	Gly Asn Glu Leu Ala	Leu Pro Ser
1370	1375	1380
Lys Tyr Val Asn Phe Leu Tyr	Leu Ala Ser His Tyr	Glu Lys Leu
1385	1390	1395

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1400 1405 1410

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1415 1420 1425

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1430 1435 1440

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1445 1450 1455

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1460 1465 1470

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1475 1480 1485

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1490 1495 1500

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1505 1510 1515

Gly Gly Asp Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His
1520 1525 1530

Arg Gly Val Pro
1535

<210> 331

<211> 1537

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 331

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val Pro
1 5 10 15

Met Ala Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu
20 25 30

Asp Leu Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg

35 40 45
 Ile Phe Thr Asp Lys Ala Ser Gly Ser Ser Thr Asp Arg Glu Gly Leu
 50 55 60
 Asp Leu Leu Arg Met Lys Val Lys Glu Gly Asp Val Ile Leu Val Lys
 65 70 75 80
 Lys Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Leu Gln Leu Ile
 85 90 95
 Lys Glu Phe Asp Ala Gln Gly Val Ala Val Arg Phe Ile Asp Asp Gly
 100 105 110

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

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

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

Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile
290 295 300

Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu
305 310 315 320

Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His
325 330 335

Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys
340 345 350

Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn
355 360 365

Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg
370 375 380

Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly
385 390 395 400

Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly
405 410 415

Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys
420 425 430

Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu
435 440 445

Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn
450 455 460

Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu
465 470 475 480

Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu
485 490 495

His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu
500 505 510

Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr
515 520 525

Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe

530 535 540
 Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val
 545 550 555 560

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

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

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

 Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys
 755 760 765
 Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile
 770 775 780

Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met
785 790 795 800
Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val
 805 810 815

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

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

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

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

1025	1030	1035
Val Pro Ser Glu Glu Val Val	Lys Lys Met Lys Asn	Tyr Trp Arg
1040	1045	1050
Gln Leu Leu Asn Ala Lys Leu	Ile Thr Gln Arg Lys	Phe Asp Asn
1055	1060	1065
Leu Thr Lys Ala Glu Arg Gly	Gly Leu Ser Glu Leu	Asp Lys Ala
1070	1075	1080
Gly Phe Ile Lys Arg Gln Leu	Val Glu Thr Arg Gln	Ile Thr Lys
1085	1090	1095
His Val Ala Gln Ile Leu Asp	Ser Arg Met Asn Thr	Lys Tyr Asp
1100	1105	1110
Glu Asn Asp Lys Leu Ile Arg	Glu Val Lys Val Ile	Thr Leu Lys
1115	1120	1125
Ser Lys Leu Val Ser Asp Phe	Arg Lys Asp Phe Gln	Phe Tyr Lys
1130	1135	1140
Val Arg Glu Ile Asn Asn Tyr	His His Ala His Asp	Ala Tyr Leu
1145	1150	1155
Asn Ala Val Val Gly Thr Ala	Leu Ile Lys Lys Tyr	Pro Lys Leu
1160	1165	1170
Glu Ser Glu Phe Val Tyr Gly	Asp Tyr Lys Val Tyr	Asp Val Arg
1175	1180	1185
Lys Met Ile Ala Lys Ser Glu	Gln Glu Ile Gly Lys	Ala Thr Ala
1190	1195	1200
Lys Tyr Phe Phe Tyr Ser Asn	Ile Met Asn Phe Phe	Lys Thr Glu
1205	1210	1215
Ile Thr Leu Ala Asn Gly Glu	Ile Arg Lys Arg Pro	Leu Ile Glu
1220	1225	1230
Thr Asn Gly Glu Thr Gly Glu	Ile Val Trp Asp Lys	Gly Arg Asp
1235	1240	1245
Phe Ala Thr Val Arg Lys Val	Leu Ser Met Pro Gln	Val Asn Ile
1250	1255	1260

Val Lys	Lys Thr Glu Val Gln	Thr Gly Gly Phe Ser	Lys Glu Ser
1265	1270	1275	
Ile Leu	Pro Lys Arg Asn Ser	Asp Lys Leu Ile Ala	Arg Lys Lys
1280	1285	1290	
Asp Trp	Asp Pro Lys Lys Tyr	Gly Gly Phe Asp Ser	Pro Thr Val
1295	1300	1305	
Ala Tyr	Ser Val Leu Val Val	Ala Lys Val Glu Lys	Gly Lys Ser
1310	1315	1320	
Lys Lys	Leu Lys Ser Val Lys	Glu Leu Leu Gly Ile	Thr Ile Met
1325	1330	1335	
Glu Arg	Ser Ser Phe Glu Lys	Asn Pro Ile Asp Phe	Leu Glu Ala
1340	1345	1350	
Lys Gly	Tyr Lys Glu Val Lys	Lys Asp Leu Ile Ile	Lys Leu Pro
1355	1360	1365	
Lys Tyr	Ser Leu Phe Glu Leu	Glu Asn Gly Arg Lys	Arg Met Leu
1370	1375	1380	
Ala Ser	Ala Gly Glu Leu Gln	Lys Gly Asn Glu Leu	Ala Leu Pro
1385	1390	1395	
Ser Lys	Tyr Val Asn Phe Leu	Tyr Leu Ala Ser His	Tyr Glu Lys
1400	1405	1410	
Leu Lys	Gly Ser Pro Glu Asp	Asn Glu Gln Lys Gln	Leu Phe Val
1415	1420	1425	
Glu Gln	His Lys His Tyr Leu	Asp Glu Ile Ile Glu	Gln Ile Ser
1430	1435	1440	
Glu Phe	Ser Lys Arg Val Ile	Leu Ala Asp Ala Asn	Leu Asp Lys
1445	1450	1455	
Val Leu	Ser Ala Tyr Asn Lys	His Arg Asp Lys Pro	Ile Arg Glu
1460	1465	1470	
Gln Ala	Glu Asn Ile Ile His	Leu Phe Thr Leu Thr	Asn Leu Gly
1475	1480	1485	
Ala Pro	Ala Ala Phe Lys Tyr	Phe Asp Thr Thr Ile	Asp Arg Lys

1490 1495 1500
 Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
 1505 1510 1515
 Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
 1520 1525 1530
 Leu Gly Gly Asp
 1535
 <210> 332
 <211> 1531
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic Polypeptide

 <400> 332
 Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
 1 5 10 15
 Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
 20 25 30
 Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45
 Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 50 55 60

 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
 65 70 75 80
 Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
 85 90 95
 Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110
 His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125

 His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140
 Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

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

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

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

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

 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400

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

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

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

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

 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

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

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960

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

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

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

1130	1135	1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys		
1145	1150	1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser		
1160	1165	1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys		
1175	1180	1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu		
1190	1195	1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly		
1205	1210	1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val		
1220	1225	1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser		
1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys		
1250	1255	1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys		
1265	1270	1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala		
1280	1285	1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn		
1295	1300	1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala		
1310	1315	1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser		
1325	1330	1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr		
1340	1345	1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp		
1355	1360	1365

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val
1370 1375 1380

Pro Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Ala Thr Ile Gly
1385 1390 1395

Tyr Ile Arg Val Ser Thr Ile Asp Gln Asn Ile Asp Leu Gln Arg
1400 1405 1410

Asn Ala Leu Thr Ser Ala Asn Cys Asp Arg Ile Phe Glu Asp Arg
1415 1420 1425

Ile Ser Gly Lys Ile Ala Asn Arg Pro Gly Leu Lys Arg Ala Leu
1430 1435 1440

Lys Tyr Val Asn Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
1445 1450 1455

Arg Leu Gly Arg Ser Val Lys Asn Leu Val Ala Leu Ile Ser Glu
1460 1465 1470

Leu His Glu Arg Gly Ala His Phe His Ser Leu Thr Asp Ser Ile
1475 1480 1485

Asp Thr Ser Ser Ala Met Gly Arg Phe Phe Phe Tyr Val Met Ser
1490 1495 1500

Ala Leu Ala Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Leu
1505 1510 1515

Ala Gly Leu Ala Ala Ala Arg Ala Gln Gly Arg Leu Gly
1520 1525 1530

<210> 333

<211> 1531

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 333

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val Pro
1 5 10 15

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

20	25	30	
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe			
35	40	45	
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile			
50	55	60	
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu			
65	70	75	80
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys			
85	90	95	
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser			
100	105	110	
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys			
115	120	125	
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr			
130	135	140	
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp			
145	150	155	160
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His			
165	170	175	
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro			
180	185	190	
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr			
195	200	205	
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala			
210	215	220	
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn			
225	230	235	240
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn			
245	250	255	
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe			
260	265	270	

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

275 280 285

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

290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

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

355 360 365

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

370 375 380

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

385 390 395 400

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

405 410 415

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

420 425 430

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

435 440 445

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

450 455 460

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

465 470 475 480

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

485 490 495

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

500 505 510

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

515 520 525
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

530 535 540
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

545 550 555 560
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

565 570 575
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

580 585 590
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

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

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

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

645 650 655
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

660 665 670
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

675 680 685
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

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

705 710 715 720
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

725 730 735
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

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

755 760 765

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

770 775 780

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

785 790 795 800

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

805 810 815

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

820 825 830

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

835 840 845

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

850 855 860

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

865 870 875 880

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010	1015	1020
Phe Val Tyr Gly Asp Tyr Lys	Val Tyr Asp Val Arg	Lys Met Ile
1025	1030	1035
Ala Lys Ser Glu Gln Glu Ile	Gly Lys Ala Thr Ala	Lys Tyr Phe
1040	1045	1050
Phe Tyr Ser Asn Ile Met Asn	Phe Phe Lys Thr Glu	Ile Thr Leu
1055	1060	1065
Ala Asn Gly Glu Ile Arg Lys	Arg Pro Leu Ile Glu	Thr Asn Gly
1070	1075	1080
Glu Thr Gly Glu Ile Val Trp	Asp Lys Gly Arg Asp	Phe Ala Thr
1085	1090	1095
Val Arg Lys Val Leu Ser Met	Pro Gln Val Asn Ile	Val Lys Lys
1100	1105	1110
Thr Glu Val Gln Thr Gly Gly	Phe Ser Lys Glu Ser	Ile Leu Pro
1115	1120	1125
Lys Arg Asn Ser Asp Lys Leu	Ile Ala Arg Lys Lys	Asp Trp Asp
1130	1135	1140
Pro Lys Lys Tyr Gly Gly Phe	Asp Ser Pro Thr Val	Ala Tyr Ser
1145	1150	1155
Val Leu Val Val Ala Lys Val	Glu Lys Gly Lys Ser	Lys Lys Leu
1160	1165	1170
Lys Ser Val Lys Glu Leu Leu	Gly Ile Thr Ile Met	Glu Arg Ser
1175	1180	1185
Ser Phe Glu Lys Asn Pro Ile	Asp Phe Leu Glu Ala	Lys Gly Tyr
1190	1195	1200
Lys Glu Val Lys Lys Asp Leu	Ile Ile Lys Leu Pro	Lys Tyr Ser
1205	1210	1215
Leu Phe Glu Leu Glu Asn Gly	Arg Lys Arg Met Leu	Ala Ser Ala
1220	1225	1230
Gly Glu Leu Gln Lys Gly Asn	Glu Leu Ala Leu Pro	Ser Lys Tyr
1235	1240	1245

Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly		
1250	1255	1260
Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His		
1265	1270	1275
Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser		
1280	1285	1290
Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser		
1295	1300	1305
Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu		
1310	1315	1320
Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala		
1325	1330	1335
Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr		
1340	1345	1350
Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile		
1355	1360	1365
Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly		
1370	1375	1380
Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Ala Thr Ile Gly		
1385	1390	1395
Tyr Ile Arg Val Ser Thr Ile Asp Gln Asn Ile Asp Leu Gln Arg		
1400	1405	1410
Asn Ala Leu Thr Ser Ala Asn Cys Asp Arg Ile Phe Glu Asp Arg		
1415	1420	1425
Ile Ser Gly Lys Ile Ala Asn Arg Pro Gly Leu Lys Arg Ala Leu		
1430	1435	1440
Lys Tyr Val Asn Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp		
1445	1450	1455
Arg Leu Gly Arg Ser Val Lys Asn Leu Val Ala Leu Ile Ser Glu		
1460	1465	1470
Leu His Glu Arg Gly Ala His Phe His Ser Leu Thr Asp Ser Ile		

1475 1480 1485
Asp Thr Ser Ser Ala Met Gly Arg Phe Phe Phe Tyr Val Met Ser
1490 1495 1500
Ala Leu Ala Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Leu
1505 1510 1515
Ala Gly Leu Ala Ala Ala Arg Ala Gln Gly Arg Leu Gly

1520 1525 1530
<210> 334
<211> 1531
<212> PRT
<213> Artificial Sequence
<220><223> Synthetic Polypeptide
<400> 334
Met Ala Thr Ile Gly Tyr Ile Arg Val Ser Thr Ile Asp Gln Asn Ile
1 5 10 15
Asp Leu Gln Arg Asn Ala Leu Thr Ser Ala Asn Cys Asp Arg Ile Phe
20 25 30
Glu Asp Arg Ile Ser Gly Lys Ile Ala Asn Arg Pro Gly Leu Lys Arg
35 40 45

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

Ala Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Leu Ala Gly Leu
115 120 125
Ala Ala Ala Arg Ala Gln Gly Arg Leu Gly Gly Gly Ser Gly Gly Ser
130 135 140
Gly Gly Ser Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr

145 150 155 160
 Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser
 165 170 175

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

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

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

 Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu
 370 375 380
 Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys
 385 390 395 400

Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp
405 410 415

Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln
420 425 430

Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu
435 440 445

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

Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr
465 470 475 480

Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu
485 490 495

Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly
500 505 510

Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu
515 520 525

Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp
530 535 540

Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln
545 550 555 560

Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe
565 570 575

Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr
580 585 590

Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg
595 600 605

Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn
610 615 620

Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu
625 630 635 640

Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro

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

Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
900 905 910

Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
915 920 925

Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
930 935 940

Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu
945 950 955 960

Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp
965 970 975

Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser
980 985 990

Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
995 1000 1005

Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val
1010 1015 1020

Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu
1025 1030 1035

Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly
1040 1045 1050

Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
1055 1060 1065

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp
1070 1075 1080

Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg
1085 1090 1095

Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe
1100 1105 1110

Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr
1115 1120 1125

His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala

1130	1135	1140
Leu Ile Lys Lys Tyr Pro Lys	Leu Glu Ser Glu Phe	Val Tyr Gly
1145	1150	1155
Asp Tyr Lys Val Tyr Asp Val	Arg Lys Met Ile Ala	Lys Ser Glu
1160	1165	1170
Gln Glu Ile Gly Lys Ala Thr	Ala Lys Tyr Phe Phe	Tyr Ser Asn
1175	1180	1185
Ile Met Asn Phe Phe Lys Thr	Glu Ile Thr Leu Ala	Asn Gly Glu
1190	1195	1200
Ile Arg Lys Arg Pro Leu Ile	Glu Thr Asn Gly Glu	Thr Gly Glu
1205	1210	1215
Ile Val Trp Asp Lys Gly Arg	Asp Phe Ala Thr Val	Arg Lys Val
1220	1225	1230
Leu Ser Met Pro Gln Val Asn	Ile Val Lys Lys Thr	Glu Val Gln
1235	1240	1245
Thr Gly Gly Phe Ser Lys Glu	Ser Ile Leu Pro Lys	Arg Asn Ser
1250	1255	1260
Asp Lys Leu Ile Ala Arg Lys	Lys Asp Trp Asp Pro	Lys Lys Tyr
1265	1270	1275
Gly Gly Phe Asp Ser Pro Thr	Val Ala Tyr Ser Val	Leu Val Val
1280	1285	1290
Ala Lys Val Glu Lys Gly Lys	Ser Lys Lys Leu Lys	Ser Val Lys
1295	1300	1305
Glu Leu Leu Gly Ile Thr Ile	Met Glu Arg Ser Ser	Phe Glu Lys
1310	1315	1320
Asn Pro Ile Asp Phe Leu Glu	Ala Lys Gly Tyr Lys	Glu Val Lys
1325	1330	1335
Lys Asp Leu Ile Ile Lys Leu	Pro Lys Tyr Ser Leu	Phe Glu Leu
1340	1345	1350
Glu Asn Gly Arg Lys Arg Met	Leu Ala Ser Ala Gly	Glu Leu Gln
1355	1360	1365

Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu
 1370 1375 1380
 Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp
 1385 1390 1395
 Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu
 1400 1405 1410
 Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile
 1415 1420 1425

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

Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr
 1490 1495 1500
 Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Met Ala Pro
 1505 1510 1515
 Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val Pro
 1520 1525 1530

<210> 335

<211> 1531

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 335

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val Pro

1 5 10 15
 Met Ala Thr Ile Gly Tyr Ile Arg Val Ser Thr Ile Asp Gln Asn Ile
 20 25 30

Asp Leu Gln Arg Asn Ala Leu Thr Ser Ala Asn Cys Asp Arg Ile Phe
 35 40 45
 Glu Asp Arg Ile Ser Gly Lys Ile Ala Asn Arg Pro Gly Leu Lys Arg
 50 55 60
 Ala Leu Lys Tyr Val Asn Lys Gly Asp Thr Leu Val Val Trp Lys Leu

 65 70 75 80
 Asp Arg Leu Gly Arg Ser Val Lys Asn Leu Val Ala Leu Ile Ser Glu
 85 90 95
 Leu His Glu Arg Gly Ala His Phe His Ser Leu Thr Asp Ser Ile Asp
 100 105 110
 Thr Ser Ser Ala Met Gly Arg Phe Phe Phe Tyr Val Met Ser Ala Leu
 115 120 125
 Ala Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Leu Ala Gly Leu

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

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

 260 265 270
 Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu

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

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

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

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

 515 520 525

Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu
 530 535 540
 Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp
 545 550 555 560
 Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln
 565 570 575
 Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe
 580 585 590
 Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr
 595 600 605
 Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg
 610 615 620
 Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn
 625 630 635 640
 Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu
 645 650 655
 Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro
 660 665 670
 Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr
 675 680 685
 Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser
 690 695 700
 Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg
 705 710 715 720
 Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu
 725 730 735
 Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala
 740 745 750
 Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp
 755 760 765
 Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu

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

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

 900 905 910
 Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
 915 920 925
 Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
 930 935 940
 Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
 945 950 955 960
 Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu

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

Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val

1025

1030

1035

Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys

1040

1045

1050

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

1055

1060

1065

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

1070

1075

1080

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

1085

1090

1095

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

1100

1105

1110

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

1115

1120

1125

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

1130

1135

1140

Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr

1145

1150

1155

Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr

1160

1165

1170

Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser

1175

1180

1185

Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

1190

1195

1200

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly

1205

1210

1215

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

1220

1225

1230

Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys

1235

1240

1245

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

1250 1255 1260
Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn

1265 1270 1275
Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys

1280 1285 1290
Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val

1295 1300 1305
Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val

1310 1315 1320
Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu

1325 1330 1335
Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val

1340 1345 1350
Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu

1355 1360 1365
Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu

1370 1375 1380
Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe

1385 1390 1395
Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu

1400 1405 1410
Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr

1415 1420 1425
Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1430 1435 1440
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn

1445 1450 1455
Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile

1460 1465 1470
His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys

1475 1480 1485

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys

1490 1495 1500

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

1505 1510 1515

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1520 1525 1530

<210> 336

<211> 1537

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 336

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160
 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175
 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 180 185 190
 Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205
 Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220

 Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240
 Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255
 Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270
 Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285

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

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

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415

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

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

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

 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640
 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655
 Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 675 680 685

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

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735

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

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

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 770 775 780

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

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

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

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

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

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

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

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925

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

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

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

 Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
 1115 1120 1125
 Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130	1135	1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys		
1145	1150	1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser		
1160	1165	1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys		
1175	1180	1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu		
1190	1195	1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly		
1205	1210	1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val		
1220	1225	1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser		
1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys		
1250	1255	1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys		
1265	1270	1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala		
1280	1285	1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn		
1295	1300	1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala		
1310	1315	1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser		
1325	1330	1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr		
1340	1345	1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp		
1355	1360	1365

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val
1370 1375 1380

Pro Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Leu Ile Gly Tyr
1385 1390 1395

Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp Leu Gln Arg Asn
1400 1405 1410

Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu Asp Lys Leu
1415 1420 1425

Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala Leu Lys
1430 1435 1440

Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp Arg
1445 1450 1455

Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
1460 1465 1470

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp
1475 1480 1485

Thr Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala
1490 1495 1500

Leu Ala Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala
1505 1510 1515

Gly Ile Ala Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro
1520 1525 1530

Pro Lys Ser Gly
1535

<210> 337

<211> 1537

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 337

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val Pro
1 5 10 15

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
20 25 30

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
35 40 45

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
50 55 60

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
65 70 75 80

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
85 90 95

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
100 105 110

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
115 120 125

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
130 135 140

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
145 150 155 160

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
165 170 175

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
180 185 190

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
195 200 205

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
210 215 220

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
225 230 235 240

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
245 250 255

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

260	265	270
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp		
275	280	285
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp		
290	295	300
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp		
305	310	315
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser		
325	330	335
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys		
340	345	350
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe		
355	360	365
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser		
370	375	380
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp		
385	390	395
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg		
405	410	415
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu		
420	425	430
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe		
435	440	445
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile		
450	455	460
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp		
465	470	475
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu		
485	490	495
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr		
500	505	510

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

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

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

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

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

755	760	765	
Arg His Lys Pro Glu Asn Ile Val	Ile Glu Met Ala Arg Glu Asn Gln		
770	775	780	
Thr Thr Gln Lys Gly Gln Lys Asn Ser	Arg Glu Arg Met Lys Arg Ile		
785	790	795	800
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln	Ile Leu Lys Glu His Pro		
805	810	815	
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys	Leu Tyr Leu Tyr Tyr Leu		
820	825	830	
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln	Glu Leu Asp Ile Asn Arg		
835	840	845	
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val	Pro Gln Ser Phe Leu Lys		
850	855	860	
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr	Arg Ser Asp Lys Asn Arg		
865	870	875	880
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu	Val Val Lys Lys Met Lys		
885	890	895	
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys	Leu Ile Thr Gln Arg Lys		
900	905	910	
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly	Gly Leu Ser Glu Leu Asp		
915	920	925	
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val	Glu Thr Arg Gln Ile Thr		
930	935	940	
Lys His Val Ala Gln Ile Leu Asp Ser Arg	Met Asn Thr Lys Tyr Asp		
945	950	955	960
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys	Val Ile Thr Leu Lys Ser		
965	970	975	
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe	Gln Phe Tyr Lys Val Arg		
980	985	990	
Glu Ile Asn Asn Tyr His His Ala His Asp	Ala Tyr Leu Asn Ala Val		
995	1000	1005	

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

1235	1240	1245
Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly		
1250	1255	1260
Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His		
1265	1270	1275
Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser		
1280	1285	1290
Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser		
1295	1300	1305
Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu		
1310	1315	1320
Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala		
1325	1330	1335
Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr		
1340	1345	1350
Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile		
1355	1360	1365
Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly		
1370	1375	1380
Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Leu Ile Gly Tyr		
1385	1390	1395
Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp Leu Gln Arg Asn		
1400	1405	1410
Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu Asp Lys Leu		
1415	1420	1425
Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala Leu Lys		
1430	1435	1440
Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp Arg		
1445	1450	1455
Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu		
1460	1465	1470

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp
1475 1480 1485

Thr Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala
1490 1495 1500

Leu Ala Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala
1505 1510 1515

Gly Ile Ala Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro
1520 1525 1530

Pro Lys Ser Gly
1535

<210> 338

<211> 1537

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 338

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

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

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

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Ile Ala

115	120	125
Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly		
130	135	140
Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Asp Lys Lys Tyr Ser Ile		
145	150	155
Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp		
165	170	175
Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp		
180	185	190
Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser		
195	200	205
Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg		
210	215	220
Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser		
225	230	235
Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu		
245	250	255
Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe		
260	265	270
Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile		
275	280	285
Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu		
290	295	300
Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His		
305	310	315
Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys		
325	330	335
Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn		
340	345	350
Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg		
355	360	365

Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly
 370 375 380

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

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

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

 Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys
 580 585 590
 Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu
 595 600 605
 Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu

610 615 620
 Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser
 625 630 635 640

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

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

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

 Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln
 835 840 845
 Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala
 850 855 860

Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu
 865 870 875 880
 Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val
 885 890 895

Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile
 900 905 910
 Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys
 915 920 925
 Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu
 930 935 940
 Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln
 945 950 955 960

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

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

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
 1085 1090 1095
 Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

1100	1105	1110
Lys Leu Val Ser Asp Phe Arg	Lys Asp Phe Gln Phe	Tyr Lys Val
1115	1120	1125
Arg Glu Ile Asn Asn Tyr His	His Ala His Asp Ala	Tyr Leu Asn
1130	1135	1140
Ala Val Val Gly Thr Ala Leu	Ile Lys Lys Tyr Pro	Lys Leu Glu
1145	1150	1155
Ser Glu Phe Val Tyr Gly Asp	Tyr Lys Val Tyr Asp	Val Arg Lys
1160	1165	1170
Met Ile Ala Lys Ser Glu Gln	Glu Ile Gly Lys Ala	Thr Ala Lys
1175	1180	1185
Tyr Phe Phe Tyr Ser Asn Ile	Met Asn Phe Phe Lys	Thr Glu Ile
1190	1195	1200
Thr Leu Ala Asn Gly Glu Ile	Arg Lys Arg Pro Leu	Ile Glu Thr
1205	1210	1215
Asn Gly Glu Thr Gly Glu Ile	Val Trp Asp Lys Gly	Arg Asp Phe
1220	1225	1230
Ala Thr Val Arg Lys Val Leu	Ser Met Pro Gln Val	Asn Ile Val
1235	1240	1245
Lys Lys Thr Glu Val Gln Thr	Gly Gly Phe Ser Lys	Glu Ser Ile
1250	1255	1260
Leu Pro Lys Arg Asn Ser Asp	Lys Leu Ile Ala Arg	Lys Lys Asp
1265	1270	1275
Trp Asp Pro Lys Lys Tyr Gly	Gly Phe Asp Ser Pro	Thr Val Ala
1280	1285	1290
Tyr Ser Val Leu Val Val Ala	Lys Val Glu Lys Gly	Lys Ser Lys
1295	1300	1305
Lys Leu Lys Ser Val Lys Glu	Leu Leu Gly Ile Thr	Ile Met Glu
1310	1315	1320
Arg Ser Ser Phe Glu Lys Asn	Pro Ile Asp Phe Leu	Glu Ala Lys
1325	1330	1335

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

1340 1345 1350

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

1355 1360 1365

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

1370 1375 1380

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

1385 1390 1395

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu

1400 1405 1410

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

1415 1420 1425

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

1430 1435 1440

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

1445 1450 1455

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala

1460 1465 1470

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg

1475 1480 1485

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln

1490 1495 1500

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu

1505 1510 1515

Gly Gly Asp Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His

1520 1525 1530

Arg Gly Val Pro

1535

<210> 339

<211> 1537

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 339

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Arg Gly Val Pro
1 5 10 15

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
20 25 30

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
35 40 45

Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
50 55 60

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
65 70 75 80

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
85 90 95

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
100 105 110

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
115 120 125

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Ile Ala
130 135 140

Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly
145 150 155 160

Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Asp Lys Lys Tyr Ser Ile
165 170 175

Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp
180 185 190

Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp
195 200 205

Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser
210 215 220

Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg

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

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

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

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

 Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu
 465 470 475 480

Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu
485 490 495

His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu
500 505 510

Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr
515 520 525

Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe
530 535 540

Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val
545 550 555 560

Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn
565 570 575

Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu
580 585 590

Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys
595 600 605

Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu
610 615 620

Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu
625 630 635 640

Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser
645 650 655

Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro
660 665 670

Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr
675 680 685

Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg
690 695 700

Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu
705 710 715 720

Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp

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

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

1205	1210	1215
Ile Thr Leu Ala Asn Gly Glu	Ile Arg Lys Arg Pro	Leu Ile Glu
1220	1225	1230
Thr Asn Gly Glu Thr Gly Glu	Ile Val Trp Asp Lys	Gly Arg Asp
1235	1240	1245
Phe Ala Thr Val Arg Lys Val	Leu Ser Met Pro Gln	Val Asn Ile
1250	1255	1260
Val Lys Lys Thr Glu Val Gln	Thr Gly Gly Phe Ser	Lys Glu Ser
1265	1270	1275
Ile Leu Pro Lys Arg Asn Ser	Asp Lys Leu Ile Ala	Arg Lys Lys
1280	1285	1290
Asp Trp Asp Pro Lys Lys Tyr	Gly Gly Phe Asp Ser	Pro Thr Val
1295	1300	1305
Ala Tyr Ser Val Leu Val Val	Ala Lys Val Glu Lys	Gly Lys Ser
1310	1315	1320
Lys Lys Leu Lys Ser Val Lys	Glu Leu Leu Gly Ile	Thr Ile Met
1325	1330	1335
Glu Arg Ser Ser Phe Glu Lys	Asn Pro Ile Asp Phe	Leu Glu Ala
1340	1345	1350
Lys Gly Tyr Lys Glu Val Lys	Lys Asp Leu Ile Ile	Lys Leu Pro
1355	1360	1365
Lys Tyr Ser Leu Phe Glu Leu	Glu Asn Gly Arg Lys	Arg Met Leu
1370	1375	1380
Ala Ser Ala Gly Glu Leu Gln	Lys Gly Asn Glu Leu	Ala Leu Pro
1385	1390	1395
Ser Lys Tyr Val Asn Phe Leu	Tyr Leu Ala Ser His	Tyr Glu Lys
1400	1405	1410
Leu Lys Gly Ser Pro Glu Asp	Asn Glu Gln Lys Gln	Leu Phe Val
1415	1420	1425
Glu Gln His Lys His Tyr Leu	Asp Glu Ile Ile Glu	Gln Ile Ser
1430	1435	1440

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
 1445 1450 1455

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
 1460 1465 1470

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
 1475 1480 1485

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
 1490 1495 1500

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
 1505 1510 1515

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
 1520 1525 1530

Leu Gly Gly Asp
 1535