



**(19) 대한민국특허청(KR)**  
**(12) 공개특허공보(A)**

(11) 공개번호 10-2014-0115335  
(43) 공개일자 2014년09월30일

- |  |   |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)<br/>C07K 14/195 (2006.01) C07K 19/00 (2006.01)<br/>C12N 15/67 (2006.01) C12N 9/22 (2006.01)<br/>A61K 38/16 (2006.01)</p> <p>(21) 출원번호 10-2014-7021336</p> <p>(22) 출원일자(국제) 2012년12월21일<br/>심사청구일자 없음</p> <p>(85) 번역문제출일자 2014년07월29일</p> <p>(86) 국제출원번호 PCT/EP2012/076674</p> <p>(87) 국제공개번호 WO 2013/098244<br/>국제공개일자 2013년07월04일</p> <p>(30) 우선권주장<br/>1122458.1 2011년12월30일 영국(GB)</p> | <p>(71) 출원인<br/>바게닝겐 유니버시티<br/>네덜란드, 엔엘-6708 피비 바게닝겐, 드로브달세스티그 4</p> <p>(72) 발명자<br/>브라운스 스텐 요한 조제프<br/>네덜란드 엔엘-6706 쥐알 바게닝겐 콜카케르베그 138</p> <p>반 데르 오스트 존<br/>네덜란드 엔엘-6871 에이치제트 렌콤 브람 슈트 리플란드베그 116</p> <p>(74) 대리인<br/>김성기, 김진희</p> |
|--|---|

전체 청구항 수 : 총 46 항

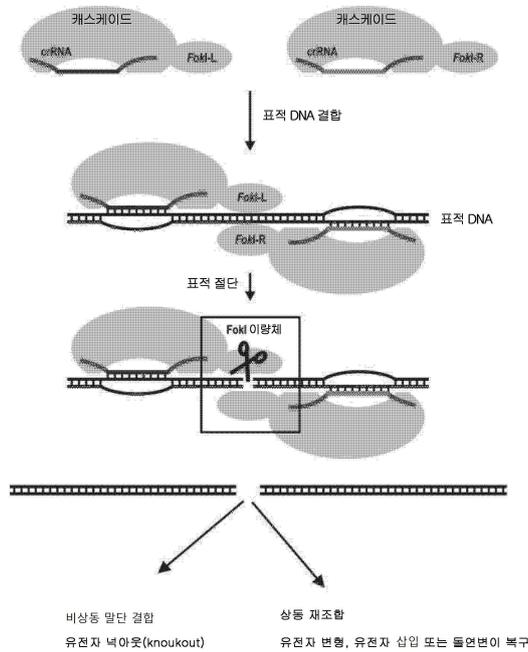
(54) 발명의 명칭 변형 캐스케이드 리보핵단백질 및 이의 용도

**(57) 요약**

본 발명은 적응성 항바이러스적 방어를 위한, CRISPR(clustered regularly interspaced short palindromic repeat)가 연관된 복합체(캐스케이드(Cascade))에 관한 것이며; 캐스케이드 단백질 복합체는 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 보유하는 추가의 아미노산 서열을 갖는 하나 이상의 서

(뒷면에 계속)

**대표도 - 도7**



브유닛을 포함하는 CRISPR 연관된 단백질 서브유닛 Cas7, Cas5 및 Cas6을 적어도 포함한다. 추가의 활성을 갖는 캐스케이드 복합체는 RNA 분자와 조합되어 리보핵단백질 복합체를 생성한다. RNA 분자는 표적 서열에 대하여 실질적인 상보성을 갖도록 선택된다. 표적화된 리보핵단백질은 상동 재조합, 비상동 말단 연결, 유전자 변형, 유전자 삽입, 돌연변이 복구에 있어서 핵산의 정확한 절단을 위한, 또는 상기 핵산의 가시화, 전사 활성화 또는 억제를 위한 유전 공학적 도구로서 이용될 수 있다. FokI 이량체에 융합된 한 쌍의 리보뉴클레오티드를 사용함으로써 DNA에서 이중 가닥 절단을 일으켜 서열 특이적 방식으로 이들 적용을 촉진한다.

---

## 특허청구의 범위

### 청구항 1

CRISPR(clustered regularly interspaced short palindromic repeat) 연관 단백질 서브유닛(subunit):

- 서열 번호 3의 아미노산 서열 또는 이와의 동일성이 18% 이상인 서열을 갖는 Cas7,
- 서열 번호 4의 아미노산 서열 또는 이와의 동일성이 17% 이상인 서열을 갖는 Cas5, 및
- 서열 번호 5의 아미노산 서열 또는 이와의 동일성이 16% 이상인 서열을 갖는 Cas6

을 적어도 포함하며, 여기서, 상기 서브유닛들 중 하나 이상은 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 제공하는 추가의 아미노산 서열을 포함하는 것인, 항바이러스적 방어를 위한 CRISPR(clustered regularly interspaced short palindromic repeat) 연관 복합체(캐스케이드(Cascade)), 캐스케이드 단백질 복합체, 또는 이의 일부.

### 청구항 2

제1항에 있어서, Cas6 서브유닛은 서열 번호 17의 아미노산 서열 또는 이와의 동일성이 16% 이상인 서열을 갖는 Cas6e 서브유닛인 단백질 복합체.

### 청구항 3

제1항 또는 제2항에 있어서, 서열 번호 2의 아미노산 서열 또는 이와의 동일성이 20% 이상인 서열을 갖는 Cse2 서브유닛을 추가로 포함하며, 경우에 따라, 이 서브유닛이 상기 추가의 아미노산 서열을 포함하는 Cse2 서브유닛인 단백질 복합체.

### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 서열 번호 1의 아미노산 서열 또는 이와의 동일성이 9% 이상인 서열을 갖는 Cse1 서브유닛을 추가로 포함하며, 경우에 따라, 이 서브유닛이 상기 추가의 아미노산 서열을 포함하는 Cse1 서브유닛인 단백질 복합체.

### 청구항 5

제4항에 있어서, 타입 I CRISPR-Cas 시스템 단백질 복합체; 바람직하게는 서브타입(subtype) I-E CRISPR-Cas 단백질 복합체인 단백질 복합체.

### 청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 상기 추가의 아미노산 서열은 상기 하나 이상의 서브유닛에 번역에 의해 융합되거나 공유 결합되며; 바람직하게는 상기 추가의 아미노산 서열은 Cse1, Cse2, Cas7, Cas5, Cas6 또는 Cas6e 서브유닛 중 하나 이상의 적어도 N 말단 및/또는 C 말단에 융합되거나 연결된 것인 단백질 복합체.

### 청구항 7

제6항에 있어서, 상기 추가의 아미노산 서열은 Cse1, Cse2 또는 Cas5 서브유닛의 N 말단 또는 C 말단; 바람직하게는 Cse1 서브유닛의 N 말단, Cse2 서브유닛의 N 말단 또는 Cas7 서브유닛의 N 말단에 융합되거나 연결된 것인 단백질 복합체.

### 청구항 8

제7항에 있어서, 상기 추가의 아미노산 서열은 단백질이며; 임의로 헬리카아제, 뉴클레아제, 뉴클레아제-헬리카아제(예를 들어 Cas3), DNA 메틸트랜스퍼라아제(예를 들어 Dam), 또는 DNA 데메틸라아제, 히스톤 메틸트랜스퍼라아제, 히스톤 데메틸라아제, 아세틸라아제, 데아세틸라아제, 포스파타아제, 키나아제, 전사 (공동)활성자, RNA 폴리머라아제 서브유닛, 전사 억제자, DNA 결합 단백질, DNA 구조화 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질(예를 들어, mCherry 또는 중금속 결합 단백질), 신호 펩티드(예를 들

어, Tat-신호 서열), 세포내(subcellular) 국제화 서열(예를 들어, 핵 국제화 서열) 또는 항체 에피토프(epitope)로부터 선택되는 것인 단백질 복합체.

**청구항 9**

제8항에 있어서, 뉴클레아제는 타입 II 제한 엔도뉴클레아제; 바람직하게는 FokI; 더 바람직하게는 변형된 FokI, 예를 들어 KKR 샤키(Sharkey) 또는 ELD 샤키로부터 선택되는 것인 단백질 복합체.

**청구항 10**

제1항 내지 제9항 중 어느 한 항에 있어서, 표적 핵산 서열과의 동일성이 50% 이상인 리보뉴클레오티드 서열을 포함하는 RNA 분자를 추가로 포함하며, 단백질 복합체와 RNA 분자는 리보핵단백질(ribonucleoprotein) 복합체를 형성하는 것인 단백질 복합체.

**청구항 11**

제10항에 있어서, RNA 분자의 일부는 표적 서열과의 동일성이 50% 이상인 리보핵단백질 복합체.

**청구항 12**

제11항에 있어서, RNA 분자의 일부는 그의 길이를 따라 표적 서열에 대하여 적어도 실질적으로 상보적인 리보핵단백질 복합체.

**청구항 13**

제10항 내지 제12항 중 어느 한 항에 있어서, RNA 분자의 길이는 35개 내지 75개 잔기의 범위인 리보핵단백질 복합체.

**청구항 14**

제10항 내지 제13항 중 어느 한 항에 있어서, 소정의 핵산 서열을 표적화하는 데 사용되는 RNA 분자의 일부는 길이가 32개 또는 33개 잔기인 리보핵단백질 복합체.

**청구항 15**

제10항 내지 제14항 중 어느 한 항에 있어서, RNA 분자는, 표적 서열에 대하여 적어도 실질적인 상보성을 갖는 RNA 서열에 대하여 5'에 있는 8개 잔기를 포함하는 것인 리보핵단백질 복합체.

**청구항 16**

제10항 내지 제15항 중 어느 한 항에 있어서, RNA 분자는, 표적 서열에 대하여 적어도 실질적인 상보성을 갖는 RNA 서열에 대하여 3'에 헤어핀(hairpin) 및 테트라뉴클레오티드 루프(loop) 형성 서열을 갖는 것인 리보핵단백질 복합체.

**청구항 17**

- a. 서열 번호 1의 아미노산 서열 또는 이와 동일성이 9% 이상인 서열을 갖는 Cse1 서브유닛;
- b. 서열 번호 2의 아미노산 서열 또는 이와 동일성이 20% 이상인 서열을 갖는 Cse2 서브유닛;
- c. 서열 번호 3의 아미노산 서열 또는 이와 동일성이 18% 이상인 서열을 갖는 Cas7 서브유닛;
- d. 서열 번호 4의 아미노산 서열 또는 이와 동일성이 17% 이상인 서열을 갖는 Cas5 서브유닛;
- e. 서열 번호 5의 아미노산 서열 또는 이와 동일성이 16% 이상인 서열을 갖는 Cas6 서브유닛

으로부터 선택되는 하나 이상의 CRISPR 연관 단백질 서브유닛을 코딩하는 단리된 핵산 분자로서, 여기서, 적어도 a, b, c, d 또는 e는 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 제공하는 추가의 아미노산 서열을 포함하는 것인 단리된 핵산 분자.

**청구항 18**

제17항에 있어서, 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 제공하는 추가의 아

미노산 서열은 CRISPR 연관 단백질 서브유닛에 융합되는 것인 단리된 핵산 분자.

**청구항 19**

제18항에 있어서, 상기 추가의 아미노산 서열은, 헬리카아제, 뉴클레아제, 뉴클레아제-헬리카아제(예를 들어 Cas3), DNA 메틸트랜스퍼라아제(예를 들어 Dam), DNA 데메틸라아제, 히스톤 메틸트랜스퍼라아제, 히스톤 데메틸라아제, 아세틸라아제, 데아세틸라아제, 포스파타아제, 키나아제, 전사 (공동)활성자, RNA 폴리머라아제 서브유닛, 전사 억제자, DNA 결합 단백질, DNA 구조화 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질(예를 들어, mCherry 또는 중금속 결합 단백질), 신호 펩티드(예를 들어, Tat-신호 서열), 세포내 국제화 서열(예를 들어, 핵 국제화 서열) 또는 항체 에피토프로부터 선택되는 것인 단리된 핵산 분자.

**청구항 20**

제17항 내지 제19항 중 어느 한 항의 핵산 분자를 포함하는 발현 벡터.

**청구항 21**

제20항에 있어서, 제10항 내지 제16항 중 어느 한 항에 정의된 RNA 분자를 코딩하는 뉴클레오티드 서열을 추가로 포함하는 것인 발현 벡터.

**청구항 22**

표적 핵산을

- a. 제10항 내지 제16항 중 어느 한 항의 리보핵단백질 복합체; 또는
- b. 제1항 내지 제9항 중 어느 한 항의 단백질 복합체 및 제10항 내지 제16항 중 어느 한 항에 정의된 RNA 분자

와 접촉시키는 단계를 포함하는, 표적 핵산의 변형, 가시화, 전사 활성화 또는 전사 억제의 방법.

**청구항 23**

제22항의 발현 벡터, 및 제10항 내지 제18항 중 어느 한 항에 정의된 RNA 분자를 코딩하는 뉴클레오티드 서열을 포함하는 추가의 발현 벡터로 세포를 형질감염시키거나, 형질전환시키거나 또는 형질도입시키는 단계를 포함하는, 세포에서의 표적 핵산의 변형, 가시화, 전사 활성화 또는 전사 억제의 방법.

**청구항 24**

제22항의 발현 벡터로 세포를 형질감염시키거나, 형질전환시키거나 또는 형질도입시키는 단계, 및 그 후 제10항 내지 제18항 중 어느 한 항에 정의된 RNA 분자를 세포에 또는 세포 내에 투입하는 단계를 포함하는, 세포에서의 표적 핵산의 변형, 가시화, 전사 활성화 또는 전사 억제의 방법.

**청구항 25**

제21항의 발현 벡터로 세포를 형질감염시키거나, 형질전환시키거나 또는 형질도입시키는 단계를 포함하는, 세포에서의 표적 핵산의 변형, 가시화, 전사 활성화 또는 전사 억제의 방법.

**청구항 26**

제22항 내지 제25항 중 어느 한 항에 있어서, 핵산 또는 염색질의 변형 또는 가시화 활성을 갖는 추가의 아미노산 서열은 마커이며, 마커는 표적 핵산 또는 염색질과 연관하고; 바람직하게는 마커는 단백질; 임의로 형광 단백질, 예를 들어 녹색 형광 단백질(green fluorescent protein; GFP) 또는 황색 형광 단백질(yellow fluorescent protein; YFP)인, 표적 핵산의 변형 또는 가시화의 방법.

**청구항 27**

제22항 내지 제26항 중 어느 한 항에 있어서, 표적 핵산은 DNA; 바람직하게는 dsDNA인 방법.

**청구항 28**

제22항 내지 제26항 중 어느 한 항에 있어서, 표적 핵산은 RNA; 바람직하게는 mRNA인 방법.

**청구항 29**

제22항 내지 제26항 중 어느 한 항에 있어서, 핵산은 dsDNA이며, 핵산 또는 염색질의 변형 활성을 갖는 추가의 아미노산 서열은 뉴클레아제 또는 뉴클레아제-헬리카아제이고, 변형은 소정의 유전자좌에서의 단일 가닥 또는 이중 가닥 절단(break)인, 표적 핵산의 변형 방법.

**청구항 30**

dsDNA 분자로부터 뉴클레오티드 서열의 적어도 일부분을 제거하고; 경우에 따라 유전자(들)의 기능을 녹아웃(knockout)시키기 위하여, 소정의 유전자좌에서 세포 중 dsDNA 분자를 비상동 말단 결합시키는 방법으로서, 제29항에 청구된 표적 핵산을 변형시키는 방법을 이용하여 이중 가닥 절단을 일으키는 단계를 포함하는 것인 방법.

**청구항 31**

기존의 뉴클레오티드 서열을 변형시키거나 소정의 뉴클레오티드 서열을 삽입하기 위하여 소정의 유전자좌에서 세포 중 dsDNA 분자 내로 핵산을 상동 재조합시키는 방법으로서,

제29항에 청구된 표적 핵산을 변형시키는 방법을 이용하여 소정의 유전자좌에서 단일 가닥 또는 이중 가닥 절단을 일으키는 단계를 포함하는 것인 방법.

**청구항 32**

제22항 내지 제25항 중 어느 한 항의 방법에서 청구된 표적 핵산 서열을 변형시키는 단계를 포함하는, 유기체에서 유전자 발현을 변형시키거나, 활성화하거나 또는 억제하는 방법으로서,

핵산은 dsDNA이며, 핵산 또는 염색질의 변형, 전사 활성화 또는 억제 활성을 갖는 추가의 아미노산 서열은 DNA 변형 효소(예를 들어 데메틸라아제 또는 데아세틸라아제), 전사 활성화자 또는 전사 억제자로부터 선택되는 것인 방법.

**청구항 33**

제22항 내지 제25항 중 어느 한 항의 방법에서 청구된 표적 핵산 서열을 변형시키는 단계를 포함하는, 유기체에서 유전자 발현을 변형시키거나, 활성화하거나 또는 억제하는 방법으로서,

핵산은 mRNA이며, 핵산 또는 염색질의 변형, 전사 활성화 또는 억제 활성을 갖는 추가의 아미노산 서열은 리보뉴클레아제이며; 임의로 엔도뉴클레아제, 3' 엑소뉴클레아제 또는 5' 엑소뉴클레아제로부터 선택되는 것인 방법.

**청구항 34**

제22항 내지 제33항 중 어느 한 항에 있어서, 세포는 원핵 생물인 방법.

**청구항 35**

제22항 내지 제33항 중 어느 한 항에 있어서, 세포는 진핵 세포, 예를 들어 식물 세포, 효모 세포, 진균류 세포, 곤충 세포, 포유류 세포 또는 인간 세포인 방법.

**청구항 36**

제35항에 있어서, 포유류 또는 인간의 세포는 인간 배아 줄기 세포 이외의 줄기 세포; 바람직하게는 단리된 줄기 세포인 방법.

**청구항 37**

제33항 내지 제35항 중 어느 한 항에 있어서, 세포를 시험관 내에서 형질감염시키는 것인 방법.

**청구항 38**

제22항 내지 제36항 중 어느 한 항에 있어서, 표적 핵산은 3차 구조, 임의로 수퍼코일형(supercoiled)을 가지며, 바람직하게는 표적 핵산은 네거티브(negative) 수퍼코일형인 방법.

**청구항 39**

제1항 내지 제9항 중 어느 한 항의 캐스케이드 단백질 복합체를 포함하는 약학 조성물.

**청구항 40**

제10항 내지 제16항 중 어느 한 항의 리보핵단백질 복합체를 포함하는 약학 조성물.

**청구항 41**

제17항 내지 제19항 중 어느 한 항의 단리된 핵산, 또는 제20항 또는 제21항의 발현 벡터를 포함하는 약학 조성물.

**청구항 42**

제1항 내지 제9항 중 어느 한 항의 캐스케이드 단백질 복합체 및 제10항 내지 제16항 중 어느 한 항에 정의된 RNA 분자를 포함하는 키트.

**청구항 43**

약제로 사용하기 위한 제1항 내지 제9항 중 어느 한 항의 캐스케이드 단백질 복합체.

**청구항 44**

약제로 사용하기 위한 제10항 내지 제16항 중 어느 한 항의 리보핵단백질 복합체.

**청구항 45**

약제로 사용하기 위한 제17항 내지 제19항 중 어느 한 항의 단리된 핵산.

**청구항 46**

약제로 사용하기 위한 제20항 또는 제21항의 발현 벡터.

**명세서**

**기술분야**

- [0001] 본 발명은 유전 공학 분야, 더 구체적으로는 원핵 생물 및 진핵 생물을 비롯한 유기체의 유전자 및/또는 게놈 변형 분야에 관한 것이다. 또한 본 발명은 생체 내이든지 시험관 내이든지 간에, 게놈 분석 및 유전자 변형 방법에서 사용하기 위한 부위 특이적 도구를 제조하는 방법에 관한 것이다. 더 구체적으로, 본 발명은 서열 특이적 방식으로 핵산 서열을 인식하여 이와 연관되는 리보핵단백질 분야에 관한 것이다.
- [0002] 박테리아 및 고세균은 침습성 DNA에 대하여 매우 다양한 방어 기작을 갖는다. 소위 CRISPR/Cas 방어 시스템은 숙주 염색체 상의 CRISPR(clustered regularly interspaced short palindromic repeat)의 유전자좌 내에 플라즈미드 및 바이러스 DNA 단편을 삽입시킴으로써 적응 면역을 제공한다. '스페이서(spacer)'로 공지된 바이러스 또는 플라즈미드 유래된 서열은 숙주 유래된 서열의 반복에 의해 서로로부터 분리되어 있다. 이들 반복 요소는 이러한 면역계의 유전 기억이며, 각각의 CRISPR 유전자좌는 외래 유전자 요소와의 이전의 조우 동안 획득되는 독특한 '스페이서' 서열의 다양한 레퍼토리(repertoire)를 포함한다.
- [0003] 외래 DNA의 획득은 제1 면역화 단계이지만, 보호는 CRISPR가 전사되고 이들 긴 전사체가 짧은 CRISPR 유래된 RNA(CRISPR-derived RNA; crRNA)로 프로세싱되는(processed) 것을 필요로 하는데, 상기 crRNA 각각은 외래 핵산 챌린저(challenger)에 대하여 상보성인 독특한 스페이서 서열을 포함한다.
- [0004] crRNA에 더하여, 몇몇 유기체에서의 유전자 실험에 의하면, crRNA 생물발생에 있어서 그리고 표적화된 간섭에 있어서, CRISPR 연관된(CRISPR-associated; Cas) 단백질의 독특한 세트가 면역 획득 단계에 필요함이 나타났다. 또한, 계통적으로 특유한 CRISPR 시스템으로부터의 Cas 단백질의 하위세트(subset)는 crRNA를 포함하는 큰 복합체로 어셈블링되는(assemble) 것으로 밝혀졌다.

- [0005] CRISPR/Cas 시스템의 다양성의 최근의 재평가에 의하면, *cas* 유전자 함량이 다양하고 CRISPR 방어 경로 전체에 걸쳐 주요한 차이를 나타내는 3가지의 특유한 타입으로 분류되었다(문헌[Makarova K. et al (2011) Nature Reviews Microbiology - AOP 9 May 2011; doi:10.1038/nrmicro2577]). (마카로바(Makarova) 분류법 및 CRISPR 연관 유전자의 명명법이 본 명세서에서 채용된다.) CRISPR 유전자좌의 RNA 전사체(pre-crRNA)는 타입 I 및 타입 III 시스템에서 CRISPR 연관된(Cas) 엔도리보뉴클레아제(endoribonuclease)에 의해 또는 타입 II 시스템에서 RNase III에 의해 반복 서열 내에서 특이적으로 절단되며; 생성된 crRNA는 침습 DNA 또는 RNA 중 어느 하나의 상보성 서열을 검출하기 위한 가이드(guide) RNA로서 Cas 단백질 복합체에 의해 이용된다. 표적 핵산의 절단은 룰러 정착 기작(ruler-anchored mechanism)에서 RNA를 절단하는 피로코쿠스 푸리오수스(*Pyrococcus furiosus*) 타입 III-B 시스템의 경우 시험관 내에서 입증되었으며, 더욱 최근에는 상보성 표적 서열(프로토스페이스어(protospacer))에서 DNA를 절단하는 스트렙토코쿠스 서모필레스(*Streptococcus thermophilus*) 타입 II 시스템의 경우 생체 내에서 입증되었다. 이와는 대조적으로, 타입 I 시스템의 경우, CRISPR 간섭의 기작은 여전히 많이 알려져 있지 않다.
- [0006] 에스케리키아 콜라이(*Escherichia coli*) 주 K12는 CRISPR/Cas 타입 I-E(이전에는 CRISPR 서브타입 E(Cse)으로 공지됨)를 보유한다. 이것은 8가지의 *cas* 유전자(*cas1*, *cas2*, *cas3* 및 *cse1*, *cse2*, *cas7*, *cas5*, *cas6e*) 및 하류 CRISPR(타입 2 반복체)을 함유한다. 에스케리키아 콜라이 K12에서, 상기 8가지의 *cas* 유전자는 CRISPR 유전자좌의 상류에 코딩되어 있다. Cas1 및 Cas2는 표적 간섭에 필요한 것으로 보이지 않지만, 새로운 표적 서열 획득에 참여할 가능성이 있다. 이와는 대조적으로 하기 6가지의 Cas 단백질: Cse1, Cse2, Cas3, Cas7, Cas5 및 Cas6e (이전에는 각각 CasA, CasB, Cas3, CasC/Cse4, CasD 및 CasE/Cse3으로도 공지됨)는 람다 파지 챌린지(challenge)에 대한 보호에 필수적이다. 이들 단백질 중 하기 5가지: Cse1, Cse2, Cas7, Cas5 및 Cas6e(이전에는 각각 CasA, CasB, CasC/Cse4, CasD 및 CasE/Cse3으로 공지됨)는 crRNA와 어셈블링되어 캐스케이드로 지칭되는 다중 서브유닛(multi-subunit) 리보핵단백질(RNP)을 형성한다.
- [0007] 이. 콜라이에 있어서, 캐스케이드는 같지 않은 화학량론적 양의 하기 5가지의 기능적 필수 Cas 단백질: Cse1, Cse2, Cas7, Cas5, Cas6e(즉, 이전의 명명법 CasA<sub>1</sub>B<sub>2</sub>C<sub>3</sub>D<sub>4</sub>E<sub>1</sub> 하) 및 61 nt의 CRISPR 유래된 RNA로 구성된 405 kDa 리보핵단백질 복합체이다. 캐스케이드는 복합체의 어셈블리 및 안정성에 있어서, 그리고 침입 핵산 서열의 동정에 있어서 crRNA에 의존하는 절대(obligate) RNP이다. 캐스케이드는 crRNA의 스페이스어 서열에 상보성인 외래 핵산을 찾아서 상기 외래 핵산과 결합하는 감시(surveillance) 복합체이다.
- [0008] 제목이 "Structural basis for CRISPR RNA-guided DNA recognition by Cascade"인 문헌[Jore et al. (2011) Nature Structural & Molecular Biology 18: 529 - 537]에는 캐스케이드의 Cas6e 서브유닛에 의해 프리-crRNA 전구체가 절단되어서 CRISPR 복합체에 의해 유지된 성숙 61 nt crRNA를 생성하는 방법이 기술되어 있다. crRNA는 crRNA 스페이스어와 상보성 프로토스페이스어 사이의 염기쌍 형성을 통하여 이중 가닥(double stranded; ds) DNA에 캐스케이드가 서열 특이적으로 결합하여 소위 R-루프를 형성하기 위한 가이드 RNA로서의 역할을 한다. 이는 ATP 비의존성 과정인 것으로 공지되어 있다.
- [0009] 제목이 "Small CRISPR RNAs guide antiviral defense in prokaryotes"인 문헌[Brouns S.J.J., et al (2008) Science 321: 960 - 964]에는 crRNA가 로딩된 캐스케이드가 생체 내 파지 내성을 위하여 Cas3을 필요로 함이 교시되어 있다.
- [0010] 제목이 "CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea"인 문헌[Marraffini L. & Sontheimer E. (2010) Nature Reviews Genetics 11: 181 - 190]은 본 분야의 기술에서의 지식의 상태를 요약한 개관 논문이다. 약간의 제안이 CRISPR 기반의 응용 및 기술에 관하여 이루어져 있지만, 이는 주로 낙농업계를 위한 가정화(domesticated) 박테리아의 파지 내성 주를 생성하는 분야에 있다. 피로코쿠스 푸리오수스에서 crRNP 복합체에 의한 시험관 내에서의 RNA 분자의 특이적 절단이 추가의 개발을 기다리고 있는 어떤 것으로서 제안되어 있다. 또한 CRISPR 시스템의 조작은 병원에서 항생제 내성 박테리아 주의 전파를 감소시키는 가능한 방법으로서 제안되어 있다. 상기 저자는 이들 기술 분야에서 상기 기술의 잠재적인 유용성을 탐구하기 위하여 추가의 연구 노력이 필요함을 강조한다.
- [0011] 발명의 명칭이 [Genetic cluster of strains of *Streptococcus thermophilus* having unique rheological properties for dairy fermentation]인 미국 특허 제2011236530 A1호(마노우리(Manoury) 등)에는 밀크(milk)를 발효시켜서 이것이 고도로 점성이고 약하게 끈적끈적한(ropy) 것으로 되도록 하는 특정한 에스. 서모필루스 주가 개시되어 있다. 규정된 서열의 특정 CRISPR 유전자좌가 개시되어 있다.
- [0012] 발명의 명칭이 [Cas6 polypeptides and methods of use]인 미국 특허 제2011217739 A1호(턴즈(Terns) 등)에는

Cas6 엔도리보뉴클레아제(endoribonuclease) 활성을 갖는 폴리펩티드가 개시되어 있다. 상기 폴리펩티드는 Cas6 인식 도메인 및 절단 부위를 갖는 표적 RNA 폴리뉴클레오티드를 절단한다. 절단은 시험관 내에서 또는 생체 내에서 실시될 수 있다. 미생물, 예컨대 이. 콜라이 또는 할로페락스 볼카니이(*Haloferax volcanii*)는 Cas6 엔도리보뉴클레아제 활성이 발현되도록 유전자 변형된다.

- [0013] 발명의 명칭이 [Bifidobacteria CRISPR sequences]인 국제 특허 공개 제W02010054154호(다니스코(Danisco))에는 비피더스균(*Bifidobacteria*)에서 발견되는 다양한 CRISPR 서열 및 파지 내성 특성이 변경된 박테리아의 유전자 변경 주의 제조에 있어서의 그의 용도가 개시되어 있다.
- [0014] 발명의 명칭이 [Prokaryotic RNAi-like system and methods of use]인 미국 특허 제2011189776 A1호(턴즈 등)에는 시험관 내에서 또는 원핵 미생물에서 생체 내에서 표적 폴리뉴클레오티드를 불활성화시키는 방법이 기술되어 있다. 상기 방법에서는 스페이서의 바로 상류의 CRISPR 유전자좌로부터의 반복체로부터 선택된 5 내지 10개 뉴클레오티드의 5' 영역을 갖는 psiRNA가 이용된다. 3' 영역은 표적 폴리뉴클레오티드의 일부에 실질적으로 상보성이다. 또한 기술되어 있는 것은 psiRNA 및 표적 폴리뉴클레오티드의 존재 하에서 엔도뉴클레아제 활성을 갖는 폴리펩티드이다.
- [0015] 발명의 명칭이 [Use of CRISPR associated genes (CAS)]인 유럽 특허 제2341149 A1호(다니스코)에는 1가지 이상의 Cas 유전자가 박테리오파지에 대한 박테리아 세포의 내성의 조정에 사용될 수 있는 방법이 기술되어 있으며, 특히 유제품에서 중균 배양 또는 프로바이오틱 배양을 제공하는 박테리아가 기술되어 있다.
- [0016] 발명의 명칭이 [Compositions and methods for downregulating prokaryotic genes]인 국제 특허 공개 제 W02010075424호(더 리전즈 오브 더 유니버시티 오브 캘리포니아(The Regents of the University of California))에는 단리된 폴리뉴클레오티드를 포함하는 CRISPR 어레이(array)가 개시되어 있다. CRISPR의 하나 이상의 스페이서는 원핵생물의 유전자에 상보성이어서, 이것은 상기 유전자의 발현을 하향 조절할 수 있으며, 이는 특히 상기 유전자가 바이오연료(biofuel) 제조와 결부되어 있을 경우 그러하다.
- [0017] 발명의 명칭이 [Cultures with improved phage resistance]인 국제 특허 공개 제W02008108989호(다니스코)에는 박테리오파지 내성 박테리아 주의 선발과, 또한, 파지 RNA의 소정 영역과의 동일성이 100%인 추가의 스페이서를 갖는 주의 선발이 개시되어 있다. 개선된 주 조합 및 중균 배양 로테이션(rotation)이 유업에서의 사용에 대하여 기술되어 있다. 특정한 파지가 생물방제제로서의 사용에 대하여 기술되어 있다.
- [0018] 발명의 명칭이 [Molecular typing and subtyping of *Salmonella* by identification of the variable nucleotide sequences of the CRISPR loci]인 국제 특허 공개 제W02009115861호(인스티튜트 파스퇴르 (Institut Pasteur))에는 살모넬라(*Salmonella*) 속의 박테리아의 검출 및 동정 방법이 개시되어 있는데, 이는 CRISPR 유전자좌에 함유된 그의 가변 뉴클레오티드 서열의 사용에 의한 것이다.
- [0019] 발명의 명칭이 [Detection and typing of bacterial strains]인 국제 특허 공개 제W02006073445호(다니스코)에는 식품, 식이 보충제 및 환경 샘플 중 박테리아 주의 검출 및 타입 결정(typing)이 개시되어 있다. 락토바실러스(*Lactobacillus*) 주는 특정 CRISPR 뉴클레오티드 서열을 통하여 동정된다.
- [0020] 제목이 [Genome editing with engineered zinc finger nucleases]인 문헌[Urnov F et al. (2010) Nature 11: 636 - 646]은 징크 핑거(zinc finger) 뉴클레아제에 관한 그리고 이것이 일련의 모델 유기체에서 역유전학 분야에서 어떻게 수단이 되었는지에 관한 개관 논문이다. 징크 핑거 뉴클레아제는 정확하게 표적화된 게놈 절단이 가능하고 이어서 후속적인 복구 과정에서 유전자 변형이 가능하도록 개발되었다. 그러나, 징크 핑거 뉴클레아제는 다수의 징크 핑거 DNA 결합 도메인을 DNA 절단 도메인에 융합시킴으로써 생성된다. DNA 서열 특이성은 몇몇 징크 핑거들을 연속하여 커플링시킴으로써 달성되는데, 상기 징크 핑거 각각은 3개 뉴클레오티드 모티프를 인식한다. 상기 기술에서의 상당한 결점은, 절단될 것이 요구되는 각각의 새로운 DNA 유전자좌에 대하여 새로운 징크 핑거가 개발될 필요가 있다는 것이다. 이는 DNA 결합 특이성을 보장하기 위하여 단백질의 엔지니어링 및 광범위한 스크리닝을 요구한다.
- [0021] 유전 공학 및 게놈 연구 분야에서, 서열/부위 특이적 핵산 검출 및/또는 절단용의 개량된 제제에 대한 지속적인 요구가 있다.
- [0022] 본 발명자는 헬리카아제(helicase)-뉴클레아제 활성을 갖는 Cas3을 발현하는 특정한 박테리아가 Cas3을 Cse1과의 융합물로서 발현한다는 놀라운 발견을 하였다. 또한 본 발명자는 예기치 않게도 Cse1과 다른 뉴클레아제 효소와의 인공 융합물을 제조할 수 있었다.
- [0023] 또한 본 발명자는 캐스캐이드에 의한 Cas3 비의존성 표적 DNA 인식이 Cas3에 의한 절단을 위한 DNA를 마킹하

며(mark), 캐스케이드 DNA 결합이 상기 표적 DNA의 토폴로지적(topolygical) 요건에 의해 좌우됨을 발견하였다.

- [0024] 본 발명자는 캐스케이드가 이완 표적 플라스미드에 결합할 수 없지만 놀랍게도 캐스케이드가 네거티브 수퍼코일형(negatively supercoiled; nSC) 토폴로지를 갖는 표적에 대하여 높은 친화도를 나타낸다는 것을 추가로 밝혀냈다.
- [0025] 따라서, 제1 측면에서 본 발명은 항바이러스적 방어를 위한, CRISPR 연관된 복합체(캐스케이드), 캐스케이드 단백질 복합체, 또는 이의 일부를 제공하며, 이는 적어도 하기의 CRISPR 연관된 단백질 서브유닛:
  - [0026] - 서열 번호 3의 아미노산 서열 또는 이와의 동일성이 18% 이상인 서열을 갖는 Cas7 (또는 COG 1857),
  - [0027] - 서열 번호 4의 아미노산 서열 또는 이와의 동일성이 17% 이상인 서열을 갖는 Cas5 (또는 COG1688), 및
  - [0028] - 서열 번호 5의 아미노산 서열 또는 이와의 동일성이 16% 이상인 서열을 갖는 Cas6 (또는 COG 1583)
- [0029] 을 포함하고; 여기서, 상기 서브유닛들 중 하나 이상은 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 제공하는 추가의 아미노산 서열을 포함한다.
- [0030] 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열을 포함하는 서브유닛은 "하나 이상의 기능성 모이어티(moiety)에 연결된 서브유닛"으로 칭해질 수 있는 것의 일례로서, 기능성 모이어티는 상기 추가의 아미노산 서열로 구성된 폴리펩티드 또는 단백질이다. 전사 활성화 활성은 소정의 유전자를 활성화하거나 상향조절하는 것이고; 전사 억제 활성은 소정의 유전자를 억제하거나 하향조절하는 것일 수 있다. 유전자의 선발은 RNA 분자를 포함하는 본 발명의 캐스케이드 복합체의 표적화로 인한 것이며, 이는 하기에 추가로 기술되어 있다.
- [0031] 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열은 바람직하게는 연결 아미노산 잔기로 형성된다. 이들 추가의 아미노산은, 연결하고 관심있는 Cas 또는 Cse 서브유닛(들)의 일부를 형성하는 폴리펩티드 또는 단백질로 볼 수 있다. 바람직하게는 이러한 폴리펩티드 또는 단백질 서열은 보통은 어떠한 Cas 또는 Cse 서브유닛 아미노산 서열의 일부도 아니다. 환언하면, 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열은 Cas 또는 Cse 서브유닛 아미노산 서열 또는 이의 일부 이외의 것일 수 있으며, 즉, Cas3 서브유닛 아미노산 서열 또는 이의 일부 이외의 것일 수 있다.
- [0032] 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열은, 필요할 경우, Cas 또는 Cse 서브유닛(들)과 동일한 유기체, 예를 들어 이. 콜라이로부터 수득되거나 이로부터 유래될 수 있다.
- [0033] 상기에 대하여 부가적으로 및/또는 대안적으로, 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열은 Cas 또는 Cse 서브유닛(들)의 아미노산 서열에 대하여 "이종성"일 수 있다. 따라서, 상기 추가의 아미노산 서열은 Cas 및/또는 Cse 서브유닛(들)이 유래되거나 기원하는 유기체와는 다른 유기체로부터 수득되거나 유래될 수 있다.
- [0034] 전체에 걸쳐, 서열 동일성은 미국 국립 생물 공학 정보 센터(National Center for Biotechnology Information) 웹서버에서 BLAST 및 후속적인 코발트(Cobalt) 다중 서열 정렬로 결정될 수 있으며, 여기서, 당해 서열은 기준 서열(예를 들어 서열 번호 3, 4 또는 5)과 비교된다. 아미노산 서열들은 주어진 기준 서열(예를 들어 서열 번호 3, 4 또는 5)과의 동일성 백분율 또는 BLOSUM62 매트릭스 기반의 서열 유사성 백분율 면에서 정의될 수 있다. 서열 유사성 또는 동일성은 기준을 이용한 보존성 백분율을 계산하기 전에 최상의 정렬을 만드는 초기 단계를 포함하며, 서열들의 진화적 관계의 척도를 반영한다.
- [0035] Cas7은 서열 번호 3과의 서열 유사성이 31% 이상일 수 있으며; Cas5는 서열 번호 4와의 서열 유사성이 26% 이상일 수 있다. Cas6은 서열 번호 5와의 서열 유사성이 27% 이상일 수 있다.
- [0036] Cse1/CasA (502 AA)의 경우:
- [0037] >gi|16130667|ref|NP\_417240.1| 캐스케이드 항바이러스 복합 단백질을 포함하는 CRISP RNA (crRNA) [에스케리키아 콜라이 주 K-12 아주 MG1655]

MNLLIDNWIPVRPRNGGKVQIINLQSLYCSRQDQWRLSLPRDDMELAAALALLVCIGQII  
APAKDDVEFRHRIMNPLTEDEFQQLIAPWIDMFYLNHAEHPFMQTKGVKANDVTPM  
EKLLAGVSGATNCAFVNQPGQGEALCGGCTAIALFNQANQAPGFGGGFKSGLRGGT  
PVTTFVRGIDLRSTVLLNVLTLPRLQKQFPNESHTEHQPTWIKPIKSNEIPASSIGFVR  
GLFWQPAHIELCDPIGIGKSCCGQESNLRYTGFLKEKFTFTVNGLWPHPHSPCLVTV  
KKGEVEEKFLAFTTSAPSWTQISRNVVVDKIIQNGENRVAAVVNQFRNIAPQSPLELI  
MGGYRNNQASILERRHDVLMFNQGWQYGNVINEIVTVGLGYKTALRKALYTFAE  
GFKNKDFKGAGVSVHETAERHFYRQSELLIPDVLANVNFSADEVIADLRDKLHQL  
CEMLFNQSVAPYAHHPKLITLALARATLYKHLRELKPPQGGPSNG [서열 번호 1]

[0038]

Cse2/CasB (160 AA)의 경우:

[0039]

>gi|16130666|ref|NP\_417239.1| 캐스케이드 항바이러스 복합 단백질을 포함하는 CRISP RNA (crRNA) [에스케리  
키아 콜라이 주 K-12 아주 MG1655]

[0040]

MADEIDAMALYRAWQQLDNGSCAQIRRVSEPDELRDIPAFYRLVQPFQWENPRHQQ  
ALLRMVFCLSAGKNVIRHQDKKSEQTTGISLGRALANSGRINERRIFQLIRADRTADM  
VQLRLLTHAEPVLDWPLMARMLTWWGKRERQQLLEDFVLTNNKNA [서열 번호 2]

[0041]

Cas7/CasC/Cse4 (363 AA)의 경우:

[0042]

>gi|16130665|ref|NP\_417238.1| 캐스케이드 항바이러스 복합 단백질을 포함하는 CRISP RNA (crRNA) [에스케리  
키아 콜라이 주 K-12 아주 MG1655]

[0043]

MSNFINIHVLIHSPSCLNRDDMNMQKDAIFGGKRRVRISSQSLKRAMRKSGYYAQN  
IGESSLRTHLAQLRDVLRQKLGFRFDQKIIDKTLALLSGKSVDEAEKISADAVTPWV  
VGEIAWFCEQVAKAEADNLDDKLLKVLKEDIAAIRVNLQQGVDIASGRMATSGM  
MTELKVDGAMSAHAITHQVDSIDWFTA VDDLQEQSAHLGTQEFSSGVFYRY  
ANINLAQLQENLGGASREQALEIATHVVHMLATEVPGAKQRTYAAFNPADMVMVN  
FSDMPLSMANAFEKAVKAKDGFLQPSIQAFNQYWDRVANGYGLNGAAAQFSLSDV  
DPITAQVKQMPTLEQLKSWVRNNGEA [서열 번호 3]

[0044]

Cas5/CasD (224 AA)의 경우:

[0045]

>gi|90111483|ref|NP\_417237.2| 캐스케이드 항바이러스 복합 단백질을 포함하는 CRISP RNA (crRNA) [에스케리  
키아 콜라이 주 K-12 아주 MG1655]

[0046]

MRSYLILRLAGPMQAWGQPTFEGTRPTGRFPTRSGLLGLLGACLGIRDDTSSLQAL  
SESVQFAVRCDELILDDRRVSVTGLRDYHTVLGAREDYRGLKSHETIQTWREYLCD  
ASFTVALWLTPHATMVISELEKAVLKPRYTPYLGRRSCPLTHPLFLGTCQASDPQKA  
LLNYEPVGGDIYSEESVTGHHLKFTARDEPMITLPRQFASREWYVIKGGMDVVSQ  
[서열 번호 4]

[0047]

Cas6e/CasE (199 AA)의 경우:

[0048]

>gi|16130663|ref|NP\_417236.1| CRISPR RNA 전구체 절단 효소; 캐스케이드 항바이러스 복합 단백질을 포함하  
는 CRISP RNA (crRNA) [에스케리키아 콜라이 주 K-12 아주 MG1655]

[0049]

MYLSKVIIARAWSRDLYQLHQGLWHLFPNRPDAARDFLFHVEKRNTPEGCHVLLQS  
AQMPVSTAVATVIKTKQVEFQLQVGVPLYFRLRANPIKTILDNQRKRLDSKGNIKRCR  
VPLIKEAEQIAWLQRKLGNAARVEDVHPISERPQYFSGDGKSGKIQTVCFEGLTIND  
APALIDLVOQQGIGPAKSMGCGLLSLAPL [서열 번호 5]

[0050]

[0051] 본 발명의 범주 내에 있는 서열 변이체의 범위를 규정함에 있어서, 의심의 여지를 피하기 위해, 상기 각각의  
동일성 백분율의 면에서 특정되는 가장 넓은 측면의 범위의 변이체로부터 출발하여 서열 번호 1, 2, 3, 4 또  
는 5의 서열 각각에 대하여 적용되는 변이의 정도에 대한 각각의 선택적 제한이다. 따라서 변이체의 범위는  
하기를 포함할 수 있다: 16% 이상, 또는 17% 이상, 또는 18% 이상, 또는 19% 이상, 또는 20% 이상, 또는 21%  
이상, 또는 22% 이상, 또는 23% 이상, 또는 24% 이상, 또는 25% 이상, 또는 26% 이상, 또는 27% 이상, 또는  
28% 이상, 또는 29% 이상, 또는 30% 이상, 또는 31% 이상, 또는 32% 이상, 또는 33% 이상, 또는 34% 이상, 또

는 35% 이상, 또는 36% 이상, 또는 37% 이상, 또는 38% 이상, 또는 39% 이상, 또는 40% 이상, 또는 41% 이상, 또는 42% 이상, 또는 43% 이상, 44% 이상, 또는 45% 이상, 또는 46% 이상, 또는 47% 이상, 또는 48% 이상, 또는 49% 이상, 또는 50% 이상, 또는 51% 이상, 또는 52% 이상, 또는 53% 이상, 또는 54% 이상, 또는 55% 이상, 또는 56% 이상, 또는 57% 이상, 또는 58% 이상, 또는 59% 이상, 또는 60% 이상, 또는 61% 이상, 또는 62% 이상, 또는 63% 이상, 또는 64% 이상, 또는 65% 이상, 또는 66% 이상, 또는 67% 이상, 또는 68% 이상, 또는 69% 이상, 또는 70% 이상, 또는 71% 이상, 72% 이상, 또는 73% 이상, 또는 74% 이상, 또는 75% 이상, 또는 76% 이상, 또는 77% 이상, 또는 78% 이상, 또는 79% 이상, 또는 80% 이상, 또는 81% 이상, 또는 82% 이상, 또는 83% 이상, 또는 84% 이상, 또는 85% 이상, 또는 86% 이상, 또는 87% 이상, 또는 88% 이상, 또는 89% 이상, 또는 90% 이상, 또는 91% 이상, 또는 92% 이상, 또는 93% 이상, 또는 94% 이상, 또는 95% 이상, 또는 96% 이상, 또는 97% 이상, 또는 98% 이상, 또는 99% 이상, 또는 100%의 아미노산 서열 동일성.

[0052] 전체에 걸쳐, 마카로바(Makarova) 등(2011)의 명명법이 Cas 단백질 서브유닛의 정의에서 사용되고 있다. 마카로바 등의 논문의 제5면의 표 2에는 Cas 유전자와, 이것이 속하는 패밀리(family) 및 슈퍼패밀리(superfamily)의 명칭이 열거되어 있다. 전체에 걸쳐, Cas 단백질 또는 Cse 단백질 서브유닛에 대한 언급은 이들 서브유닛이 부분을 형성하는 패밀리 또는 슈퍼패밀리에 대한 교차 언급을 포함한다.

[0053] 전체에 걸쳐, 본 발명의 Cas 및 Cse 서브유닛의 기준 서열은 당해 아미노산 서열을 코딩하는 뉴클레오티드 서열로서 정의될 수 있다. 예를 들어, Cas7에 대한 서열 번호 3의 아미노산 서열은 그 아미노산 서열을 코딩하는 모든 핵산 서열을 또한 포함한다. 따라서 본 발명의 범주 내에 포함되는 Cas7의 변이체는 기준 핵산 서열과의 적어도 규정된 아미노산 동일성 또는 유사성 백분율의 뉴클레오티드 서열과, 이외에도 그 하한치와 100% 사이의 모든 가능한 동일성 또는 유사성 백분율의 것을 포함한다.

[0054] 본 발명의 캐스케이드 복합체는 1가지 초과와 상이한 박테리아 또는 고세균 원핵생물로부터 유래되거나 변형된 서브유닛으로 구성될 수 있다. 또한, 상이한 Cas 서브타입들로부터의 서브유닛들이 혼합될 수 있다.

[0055] 바람직한 측면에서, Cas6 서브유닛은 하기 서열 번호 17의 Cas6e 서브유닛, 또는 이와 동일성이 16% 이상인 서열이다.

[0056] 바람직한 Cas6e 서브유닛의 서열은 >gi|16130663|ref|NP\_417236.1| CRISPR RNA 전구체 절단 효소; 캐스케이드 항바이러스 복합 단백질을 포함하는 CRISP RNA (crRNA) [에스케리키아 콜라이 주 K-12 아주 MG1655]이다:

MYLSKVIIARAWSRDLYQLHQGLWHLFPNRPDAARDFLFHVEKRNTPEGCHVLLQS  
AQMPVSTAVATVIKTKQVEFQLQVGVPLYFRLRANPIKTILDNQKRLDSKGNIKRCR  
VPLIKEAEQIAWLQRKLGNAARVEDVHPISERPQYFSGDGKSGKIQTVCFEGVLTIND  
APALIDLVQQGIGPAKSMGCGLLSLAPL [서열 번호 17]

[0057] 본 발명의 캐스케이드 복합체, 또는 이의 부분 - 이는 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열을 포함하는 하나 이상의 서브유닛을 포함함 - 은 서열 번호 2의 아미노산 서열을 갖는 Cse2(또는 YgcK-유사) 서브유닛 또는 이와 동일성이 20% 이상인 서열, 또는 이들의 일부를 추가로 포함할 수 있다. 대안적으로, Cse 서브유닛은 서열 번호 2의 서열과의 유사성이 38% 이상인 것으로 정의된다. 경우에 따라, 본 발명의 단백질 복합체 내에서, 이것은 핵산 또는 염색질 변형 활성을 갖는 추가의 아미노산 서열을 포함하는 Cse2 서브유닛이다.

[0059] 부가적으로 또는 대안적으로, 본 발명의 캐스케이드 복합체는 서열 번호 1의 아미노산 서열을 갖는 Cse1(또는 YgcL-유사) 서브유닛 또는 이와 동일성이 9% 이상인 서열, 또는 이들의 일부를 추가로 포함할 수 있다. 경우에 따라 본 발명의 단백질 복합체 내에서, 이것은 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열을 포함하는 Cse1 서브유닛이다.

[0060] 바람직한 실시 양태에서, 본 발명의 캐스케이드 복합체는 타입 I CRISPR-Cas 시스템 단백질 복합체; 더 바람직하게는 서브타입 I-E CRISPR-Cas 단백질 복합체이거나 이것은 타입 I-A 또는 타입 I-B 복합체를 기반으로 할 수 있다. 타입 I-C, D 또는 F 복합체가 가능하다. 이. 콜라이 시스템을 기반으로 하는 특히 바람직한 실시 양태에서, 서브유닛은 하기의 화학량론적 양을 가질 수 있다: Cse<sub>1</sub>Cse<sub>2</sub>Cas<sub>7</sub>Cas<sub>5</sub><sub>1</sub>, Cas<sub>6</sub><sub>1</sub> 또는 Cse<sub>1</sub>Cse<sub>2</sub>Cas<sub>7</sub>Cas<sub>5</sub><sub>1</sub>Cas<sub>6</sub><sub>1</sub>.

[0061] 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열은 상기 하나 이상의 서브유닛에 천연 또는 인공 단백질 발현 시스템에서의 발현을 통하여 번역에 의해 융합되거나, 화학적 합성 단계에 의해 공유 결합될 수 있으며; 바람직하게는 상기 하나 이상의 기능성 모이어티는 Cse1,

Cse2, Cas7, Cas5, Cas6 또는 Cas6e 서브유닛 중 하나 이상의 적어도 N 말단 영역 및/또는 C 말단 영역에 융합되거나 연결된다. 특히 바람직한 실시 양태에서, 핵산 또는 염색질 변형 활성을 갖는 추가의 아미노산 서열은 Cse1, Cse2 또는 Cas5 서브유닛의 N 말단 또는 C 말단에 융합되거나 연결되며; 더 바람직하게는 상기 연결은 Cse1 서브유닛의 N 말단, Cse2 서브유닛의 N 말단, 또는 Cas7 서브유닛의 N 말단의 영역 내에 있다.

[0062] 핵산 또는 염색질의 변형, 활성화, 억제 또는 가시화 활성을 갖는 추가의 아미노산 서열은 단백질일 수 있으며, 이는 임의로 헬리카아제, 뉴클레아제, 뉴클레아제-헬리카아제, DNA 메틸트랜스퍼라아제(예를 들어 Dam), 또는 DNA 데메틸라아제, 히스톤 메틸트랜스퍼라아제, 히스톤 데메틸라아제, 아세틸라아제, 테아세틸라아제, 포스파타아제, 키나아제, 전사 (공동)활성자, RNA 폴리머라아제 서브밋(submit), 전사 억제자, DNA 결합 단백질, DNA 구조화 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질(예를 들어, mCherry 또는 중금속 결합 단백질), 신호 펩티드(예를 들어, Tat-신호 서열), 세포내(subcellular) 국재화 서열(예를 들어, 핵 국재화 서열) 또는 항체 에피토프(epitope)로부터 선택된다.

[0063] 관심있는 단백질은 캐스케이드 단백질 서브유닛이 그 서열 기원을 갖는 박테리아 중 이외의 종으로부터의 이종성 단백질일 수 있다.

[0064] 단백질이 뉴클레아제일 경우, 이것은 타입 II 제한 엔도뉴클레아제, 예컨대 FokI, 또는 이의 돌연변이체 또는 활성 부분으로부터 선택되는 것일 수 있다. 사용될 수 있는 다른 타입 II 제한 엔도뉴클레아제는 EcoRI, EcoRV, BglII, BamHI, BspI 및 BspMI를 포함한다. 바람직하게는 본 발명의 1가지의 단백질 복합체는 FokI의 N 말단 도메인에 융합될 수 있으며, 본 발명의 또 다른 단백질 복합체는 FokI의 C 말단 도메인에 융합될 수 있다. 그 후, 이들 두 단백질 복합체는 핵산에 있어서 유리한 유전자좌 특이적 이중 가닥 절단(break)을 달성하도록 함께 사용될 수 있으며, 이에 의해 유전 물질 내의 절단 위치는 사용자의 설계 및 선택에 있는데, 이는 RNA 성분(하기에 정의되고 기술됨)에 의해 인도되는 바와 같고 표적 핵산 가닥(하기에 또한 더 상세하게 기술됨) 내의 소위 "프로토프에이저 인접 모티프(protospacer adjacent motif; PAM)" 서열의 존재로 인한 것이다.

[0065] 바람직한 실시 양태에서, 본 발명의 단백질 복합체는 변형된 제한 엔도뉴클레아제, 예를 들어 FokI인 추가의 아미노산 서열을 갖는다. 상기 변형은 바람직하게는 촉매 도메인 내에 있다. 바람직한 실시 양태에서, 변형된 FokI은 KKR샤키(Sharkey) 또는 ELD 샤키이며, 이는 단백질 복합체의 Cse1 단백질에 융합된다. 본 발명의 이들 복합체의 바람직한 응용에서, 이들 복합체 중 2가지(KKR샤키 및 ELD 샤키)는 함께 조합된 형태로 존재할 수 있다. 상이하게 변형된 FokI을 이용하는 단백질 복합체들의 헤테로이량체(heterodimer) 쌍은 표적화된 이중 가닥 핵산 절단에 있어서 특별한 이점을 갖는다. 호모이량체(homodimer)가 사용되면, 비특이적 활성으로 인하여 비표적 부위에서 더 많은 절단이 있는 것이 가능하다. 헤테로이량체 접근법은 유리하게는 물질의 샘플에서 절단의 충실도를 증가시킨다.

[0066] 상기에 정의되고 기술된 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열을 갖는 캐스케이드 복합체는 본원에 정의된 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 엔티티(entity) 중 임의의 것을 이용하여, 절단하거나, 태그하거나 또는 달리 일부의 방식으로, 예를 들어 메틸화로 변경시키기를 원하는 정확한 유전자좌를 소정의 물질 중에서 사용자가 선별하는 것을 가능케 하는 것이 유리한 본 발명의 전체 시스템의 성분 부분이다. 이 시스템의 다른 성분 부분은 변형시키거나 절단하거나 태그하고자 하는 DNA 또는 RNA 상의 올바른 유전자좌로 본 발명의 캐스케이드 복합체를 안내하는 가이드로서 작용하는 RNA 분자이다.

[0067] 바람직하게는 본 발명의 캐스케이드 복합체는 또한 소정의 표적 핵산 서열에 대한 동일성이 50% 이상인 리보뉴클레오티드 서열을 포함하는 RNA 분자를 포함하며, 여기서, 본 단백질 복합체와 RNA 분자는 리보핵단백질 복합체를 형성한다. 바람직하게는 리보핵단백질 복합체는 RNA 분자가 그의 의도된 표적 핵산 서열에 혼성화될 때 형성된다. 리보핵단백질 복합체는 캐스케이드-기능성 모이어티 조합 및 RNA 분자 및 핵산(DNA 또는 RNA)의 필요한 성분들이 생체 내이든지 시험관 내이든지 간에 적합한 생리학적 조건에서 함께 존재할 때 형성된다. 임의의 특정 이론에 구애되고자 함이 없이, 본 발명자는 dsDNA, 특히 네거티브 수퍼코일형 DNA의 맥락에서, dsDNA와 연관된 캐스케이드 복합체는 이중 가닥의 부분적인 풀림을 야기하고, 이는 그 후 RNA가 한 가닥과 결합되게 하고; 그 후 상기 RNA 서열의 적어도 일부에 실질적으로 상보성인 표적 서열에 도달할 때까지 전 리보핵단백질 복합체가 DNA 가닥을 따라서 이동하며, 이 시점에서 RNA와 DNA 가닥 사이의 안정한 상호작용이 일어나며, 그 유전자좌에서의 DNA의 변형에 의한 것이든지, 뉴클레아제 절단에 의한 것이든지 태그에 의한 것이든지 간에 기능성 모이어티의 기능은 효과가 나타난다.

[0068] 바람직한 실시 양태에서, RNA 분자의 일부는 표적 핵산 서열에 대한 동일성이 50% 이상이며, 더 바람직하게는

표적 서열에 대한 동일성이 95% 이상이다. 더 바람직한 실시 양태에서, RNA 분자의 일부는 그의 길이를 따라서 표적 DNA 서열에 대하여 실질적으로 상보성이며; 즉, 연결성이거나 비연접성일 수 있는 단지 1개, 2개, 3개, 4개 또는 5개의 미스매치(mismatch)가 있다. RNA 분자 (또는 이의 일부)는 표적 서열에 대한 동일성이 51% 이상, 또는 52% 이상, 또는 53% 이상, 또는 54% 이상, 또는 55% 이상, 또는 56% 이상, 또는 57% 이상, 또는 58% 이상, 또는 59% 이상, 또는 60% 이상, 또는 61% 이상, 또는 62% 이상, 또는 63% 이상, 또는 64% 이상, 또는 65% 이상, 또는 66% 이상, 또는 67% 이상, 또는 68% 이상, 또는 69% 이상, 또는 70% 이상, 또는 71% 이상, 또는 72% 이상, 또는 73% 이상, 또는 74% 이상, 또는 75% 이상, 또는 76% 이상, 또는 77% 이상, 또는 78% 이상, 또는 79% 이상, 또는 80% 이상, 또는 81% 이상, 또는 82% 이상, 또는 83% 이상, 또는 84% 이상, 또는 85% 이상, 또는 86% 이상, 또는 87% 이상, 또는 88% 이상, 또는 89% 이상, 또는 90% 이상, 또는 91% 이상, 또는 92% 이상, 또는 93% 이상, 또는 94% 이상, 또는 95% 이상, 또는 96% 이상, 또는 97% 이상, 또는 98% 이상, 또는 99% 이상, 또는 100%일 수 있다.

- [0069] 표적 핵산은 DNA(ss 또는 ds) 또는 RNA일 수 있다.
- [0070] 다른 바람직한 실시 양태에서, RNA 분자 또는 이의 일부는 표적 핵산과의 동일성이 70% 이상이다. 이러한 수준의 동일성에서, 표적 핵산은 바람직하게는 dsDNA이다.
- [0071] 바람직하게는 RNA 분자는 표적 핵산 서열에 대하여 높은 특이성 및 친화성을 요구한다. 1 pM 내지 1 μM, 바람직하게는 1 ~ 100 nM의 범위의 해리 상수( $K_d$ )가 바람직하며, 이는 바람직하게는 천연 겔 전기영동법, 또는 대안적으로 등은 적정 열량법, 표면 플라즈몬 공명법, 또는 형광 기반의 적정법에 의해 결정되는 바와 같다. 친화도는 전기영동 이동성 시프트 분석법(electrophoretic mobility shift assay; EMSA)을 이용하여 측정될 수 있는데, 이는 겔 지연 분석법으로도 칭해진다(문헌[Semenova E et al. (2011) Proc. Natl. Acad. Sci. USA 108: 10098 - 10103] 참조).
- [0072] 바람직하게는 RNA 분자는 CRISPR RNA(crRNA) 분자로서 원핵생물에서 자연으로부터 공지된 것에서 모델링된다. crRNA 분자의 구조는 이미 확립되어 있으며, 문헌[Jore et al. (2011) Nature Structural & Molecular Biology 18:529 - 537]에 더욱 상세하게 설명되어 있다. 간략하게는, 타입 I-E의 성숙 crRNA는 흔히 길이가 61개 뉴클레오티드이며, 8개 뉴클레오티드의 5' "핸들(handle)", 32개 뉴클레오티드의 "스페이스" 서열, 및 21개 뉴클레오티드의 3' 서열(이는 테트라뉴클레오티드 루프를 갖는 헤어핀(hairpin)을 형성함)로 이루어진다. 그러나, 본 발명에서 사용되는 RNA는, 길이에 있어서든지, 영역에 있어서든지 또는 특정 RNA 서열에 있어서든지 간에 자연 발생 crRNA의 설계에 대하여 엄격하게 설계될 필요는 없다. 그러나 명확한 것은, 본 발명에서 사용하기 위한 RNA 분자는 공개 데이터베이스의 유전자 서열 정보 또는 새롭게 발견된 유전자 서열 정보를 기반으로 하여 설계되고, 그 후, 예를 들어 전체적 또는 부분적으로 화학적 합성에 의해 인공적으로 만들어질 수 있다는 것이다. 본 발명의 RNA 분자는 또한 설계되어서 유전자 변형 세포 또는 무세포 발현 시스템에서의 발현에 의해 생성될 수 있으며, 이 옵션은 그 RNA 서열의 일부 또는 전부의 합성을 포함할 수 있다.
- [0073] crRNA의 구조 및 요건이 또한 문헌[Semenova E et al.(2011) Proc. Natl. Acad. Sci. USA 108: 10098 - 10103]에 기술되어 있다. 스페이스 서열의 5' 말단을 형성하는 소위 "SEED" 부분이 있으며, 이는, 이것에 대하여 5'에서 8개 뉴클레오티드의 5' 핸들이 측면에 위치한다. 문헌[Semenova et al. (2011)]에서는 모든 SEED 서열 잔기가 표적 서열에 대하여 상보성이어야 하지만, 6 위치의 잔기의 경우 미스매치가 용인될 수 있음이 발견되었다. 이와 유사하게, 표적 유전자좌(즉, 서열)에서 인도된 본 발명의 리보핵단백질 복합체의 RNA 성분의 설계 및 제조시에, SEED 서열에 대한 필요한 매치 및 미스매치 규칙이 적용될 수 있다.
- [0074] 따라서 본 발명은 표적 핵산 분자에서 단일 염기 변화를 검출 및/또는 위치화하는 방법을 포함하며, 이는 핵산 샘플을 이상에서 기술된 본 발명의 리보핵단백질 복합체와 접촉시키거나 이상에서 기술된 본 발명의 별도의 RNA 성분 및 캐스케이드 복합체와 접촉시키는 단계를 포함하며, 여기서, RNA 성분의 서열(리보핵단백질 복합체 내에 있을 때를 포함함)는 이것이 8개 뉴클레오티드 잔기의 연결 서열의 6 위치에서의 단일 염기 변화에 정상 대립유전자와 돌연변이 대립유전자를 구별하도록 하는 것이다.
- [0075] 본 발명의 실시 양태에서, RNA 분자는 길이가 35 내지 75개의 범위의 잔기일 수 있다. 바람직한 실시 양태에서, 소정의 핵산 서열에 대하여 상보성이고 상기 소정의 핵산 서열을 표적화하는 데 사용되는 RNA의 부분은 길이가 32개 또는 33개 잔기이다. (자연 발생 crRNA의 맥락에서, 이것은 스페이스 부분에 상응하며; 문헌[Semenova et al. (2011)]의 도 1에 도시된 바와 같다).
- [0076] 본 발명의 리보핵단백질 복합체는 핵산 표적 서열에 대하여 적어도 실질적인 상보성을 갖는 RNA 서열에 대하

여 5'에 8개 잔기를 포함하는 RNA 성분을 부가적으로 가질 수 있다. (핵산 표적 서열에 대하여 적어도 실질적인 상보성을 갖는 RNA 서열은 스페이서 서열로서 crRNA의 맥락에서 상응하는 것으로 이해된다. RNA의 5' 측면 서열은 crRNA의 5' 헐들에 상응하는 것으로 간주된다. 이는 문헌[Semenova et al. (2011)]의 도 1에 나타내어져 있다).

[0077] 본 발명의 리보핵단백질 복합체는 DNA 표적 서열에 대하여 적어도 실질적인 상보성을 갖는 RNA 서열에 대하여 3'인 헤어핀 및 테트라뉴클레오티드 루프 형성 서열을 가질 수 있다. (crRNA의 맥락에서, 이는 문헌[Semenova et al. (2011)]의 도 1에 나타낸 바와 같이, 스페이서 서열 측면에 위치하는 3' 헐들에 상응한다).

[0078] 일부 실시 양태에서, RNA는 CRISPR RNA(crRNA)일 수 있다.

[0079] 본 발명의 캐스케이드 단백질 및 복합체는 표적 핵산(이는 DNA 또는 RNA일 수 있음)의 존재 하에 리보핵단백질 복합체를 형성하는 RNA 인도 성분과의 그의 결합 활성의 면에서 시험관 내에서 특성화될 수 있다. 본 발명의 복합체와 그의 핵산 표적과의 상호작용에 대한 기능 분석법으로서 전기영동 이동성 시프트 분석법(EMSA)이 사용될 수 있다. 기본적으로, 본 발명의 캐스케이드-기능성 모이어티 복합체는 핵산 표적과 혼합되며, 캐스케이드-기능성 모이어티 복합체의 안정한 상호작용은 EMSA에 의해 또는 기능성 모이어티 밖에서의 특정한 리드아웃(readout), 예를 들어 소정의 부위에서의 표적 DNA의 엔도뉴클레아제적 절단(endonucleolytic cleavage)에 의해 모니터링된다. 이는 공지된 특이성 및 표적 DNA 분자 내의 절단 부위를 갖는 구매가능한 효소를 사용하여 추가의 제한효소 단편 길이 분석에 의해 결정될 수 있다.

[0080] 인도 RNA의 존재 하에서 본 발명의 캐스케이드 단백질 또는 복합체의 DNA 또는 RNA에의 결합을 가시화하는 것은 주사/원자력 현미경(SFM/AFM) 영상화를 이용하여 달성될 수 있으며, 이는 본 발명의 기능성 복합체의 존재에 대한 분석을 제공할 수 있다.

[0081] 또한 본 발명은 하기:

[0082] a. 서열 번호 1의 아미노산 서열 또는 이와의 동일성이 9% 이상인 서열을 갖는 Cse1 서브유닛;

[0083] b. 서열 번호 2의 아미노산 서열 또는 이와의 동일성이 20% 이상인 서열을 갖는 Cse2 서브유닛;

[0084] c. 서열 번호 3의 아미노산 서열 또는 이와의 동일성이 18% 이상인 서열을 갖는 Cas7 서브유닛;

[0085] d. 서열 번호 4의 아미노산 서열 또는 이와의 동일성이 17% 이상인 서열을 갖는 Cas5 서브유닛;

[0086] e. 서열 번호 5의 아미노산 서열 또는 이와의 동일성이 16% 이상인 서열을 갖는 Cas6 서브유닛

[0087] 으로부터 선택되는 하나 이상의 CRISPR 연관된 단백질 서브유닛을 코딩하는 핵산 분자를 제공하며, 여기서, 적어도 a, b, c, d 또는 e는 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열을 포함한다.

[0088] 핵산 또는 염색질의 변형, 가시화, 전사 활성화 또는 전사 억제 활성을 갖는 추가의 아미노산 서열은 바람직하게는 CRISPR 연관된 단백질 서브유닛에 융합된다.

[0089] 상기에 정의된 본 발명의 핵산에서, 뉴클레오티드 서열은 각각의 서열 번호 1, 서열 번호 2, 서열 번호 3, 서열 번호 4 또는 서열 번호 5의 서열을 코딩하는 것일 수 있거나, 이에 대한 변이체 서열의 범위를 정의함에 있어서, 이것은 바람직하게는 엄격한 조건 하에서, 더 바람직하게는 매우 높은 엄격도 조건 하에서 그 뉴클레오티드 서열에 서열 혼성화가능한 것일 수 있다. 다양한 엄격한 혼성화 조건은 당업자에게 친숙하다. 핵산 분자의 혼성화는 2개의 상보성 핵산 분자가 소정량의 서로에게의 수소 결합을 겪을 때 일어나는데, 상기 수소 결합은 왓슨-크릭(Watson-Crick) 염기쌍 형성으로 공지된다. 혼성화 엄격도는 핵산을 둘러싸고 있는 환경(즉, 화학적/물리적/생물학적) 조건, 온도, 혼성화 방법의 성질, 및 사용되는 핵산 분자의 조성 및 길이에 따라 달라질 수 있다. 특정한 정도의 엄격도를 획득하는 데 요구되는 혼성화 조건에 관한 계산은 문헌[Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001)]; 및 문헌[Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes Part I, Chapter 2 (Elsevier, New York, 1993)]에 논의되어 있다.  $T_m$ 은 핵산 분자의 주어진 가닥의 50%가 그의 상보성 가닥에 혼성화되는 온도이다. 하기는 예시적인 혼성화 조건 세트이며, 한정되지 않는다:

[0090] 매우 높은 엄격도 (90% 이상의 동일성을 공유하는 서열들이 혼성화되게 함)

- [0091] 혼성화: 5x SSC, 65°C, 16시간
- [0092] 2회 세척: 2x SSC, 실온(RT), 각각 15분
- [0093] 2회 세척: 0.5x SSC, 65°C, 각각 20분
- [0094] 높은 엄격도 (80% 이상의 동일성을 공유하는 서열들이 혼성화되게 함)
- [0095] 혼성화: 5x-6x SSC, 65°C-70°C, 16-20시간
- [0096] 2회 세척: 2x SSC, RT, 각각 5-20분
- [0097] 2회 세척: 1x SSC, 55°C-70°C, 각각 30분
- [0098] 낮은 엄격도 (50% 이상의 동일성을 공유하는 서열들이 혼성화되게 함)
- [0099] 혼성화: 6x SSC, RT 내지 55°C, 16-20시간
- [0100] 2회 이상 세척: 2x-3x SSC, RT 내지 55°C, 각각 20-30분.
- [0101] 핵산 분자는 단리된 핵산 분자일 수 있으며, RNA 또는 DNA 분자일 수 있다.
- [0102] 상기 추가의 아미노산 서열은 헬리카아제, 뉴클레아제, 뉴클레아제-헬리카아제(예를 들어, Cas3), DNA 메틸트랜스퍼라아제(예를 들어 Dam), DNA 데메틸라아제, 히스톤 메틸트랜스퍼라아제, 히스톤 데메틸라아제, 아세틸라아제, 데아세틸라아제, 포스파타아제, 키나아제, 전사 (공동)활성자, RNA 폴리머라아제 서브유닛, 전사 억제자, DNA 결합 단백질, DNA 구조화 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질(예를 들어, mCherry 또는 중금속 결합 단백질), 신호 펩티드(예를 들어, Tat-신호 서열), 세포내(subcellular) 국재화 서열(예를 들어, 핵 국재화 서열) 또는 항체 에피토프로부터 선택될 수 있다. 상기 추가의 아미노산 서열은 관련 캐스케이드 단백질 서브유닛(들)이 유래되는 유기체로부터의 상이한 단백질이거나 또는 상기 단백질로부터 유래될 수 있다.
- [0103] 본 발명은 이상에서 정의된 핵산 분자를 포함하는 발현 벡터를 포함한다. 1가지의 발현 벡터는 단일 캐스케이드 단백질 서브유닛을 코딩하는 뉴클레오티드 서열 및 또한 추가의 아미노산 서열을 코딩하는 뉴클레오티드 서열을 함유할 수 있으며, 이에 의해, 발현시에 상기 서브유닛 및 추가의 아미노산 서열이 융합된다. 다른 발현 벡터는 어떠한 추가의 아미노산 서열에도 융합되지 않은 단지 하나 이상의 캐스케이드 단백질 서브유닛을 코딩하는 뉴클레오티드 서열을 포함할 수 있다.
- [0104] 핵산 또는 염색질 변형 활성을 갖는 추가의 아미노산 서열은 링커 폴리펩티드를 통하여 캐스케이드 서브유닛들 중 임의의 것에 융합될 수 있다. 링커는 최대 약 60개 또는 최대 약 100개의 아미노산 잔기의 임의의 길이의 것일 수 있다. 바람직하게는 링커는 10 내지 60개, 더 바람직하게는 10 내지 20개의 범위의 수의 아미노산을 갖는다. 아미노산은 바람직하게는 극성 아미노산 및/또는 작은 아미노산 및/또는 하전된 아미노산(예를 들어 Gln, Ser, Thr, Pro, Ala, Glu, Asp, Lys, Arg, His, Asn, Cys, Tyr)이다. 링커 펩티드는 바람직하게는 융합 기능성 모이어티 및 상기 모이어티가 융합하여 표적 뉴클레오티드와 적당히 상호작용하게 하는 캐스케이드의 서브유닛의 올바른 이격 및 위치화를 수득하도록 설계된다.
- [0105] 본 발명의 발현 벡터(발현시에 캐스케이드 단백질 서브유닛에 융합되는 아미노산 잔기를 코딩하는 뉴클레오티드를 포함하거나 포함하지 않음)는 이상에서 정의된 RNA 분자를 코딩하는 서열을 추가로 포함할 수 있다. 그 결과, 이러한 발현 벡터를 적절한 숙주 내에서 사용하여 본 발명의 리보핵단백질을 생성할 수 있으며, 본 리보핵단백질은 소정의 뉴클레오티드 서열을 표적화할 수 있다.
- [0106] 따라서, 본 발명은 표적 핵산을 변형시키거나, 가시화하거나, 또는 표적 핵산의 전사를 활성화 또는 억제하는 방법으로서 상기 핵산을 이상에서 정의된 리보핵단백질 복합체와 접촉시키는 단계를 포함하는 방법을 또한 제공한다. 변형은 상기 핵산의 절단에 의한 것이거나 상기 핵산에의 결합에 의한 것일 수 있다.
- [0107] 또한 본 발명은 표적 핵산을 변형시키거나, 가시화하거나, 또는 표적 핵산의 전사를 활성화 또는 억제하는 방법으로서 상기 핵산을 이상에서 정의된 캐스케이드 단백질 복합체 및 이상에서 정의된 RNA 분자와 접촉시키는 단계를 포함하는 방법을 또한 포함한다.
- [0108] 따라서, 상기 방법에 따르면, 표적 핵산을 변형시키거나, 가시화하거나, 또는 표적 핵산의 전사를 활성화 또는 억제하는 것은 시험관 내에서 그리고 무세포 환경에서 실시될 수 있으며, 즉, 본 방법은 용액 중에서 자유롭든지 고상을 포함하든지 간에 생화학 반응으로서 실시된다. 표적 핵산은 예를 들어 고상에 결합될 수 있다.

- [0109] 무세포 환경에서, 표적 핵산, 캐스케이드 단백질 복합체 및 RNA 분자 각각의 첨가 순서는 평균적인 당업자의 옵션에 있다. 상기 3가지 성분은 동시에, 임의의 소정의 순서로 순차적으로, 또는 상이한 시점에 소정의 순서로 별도로 첨가될 수 있다. 따라서, 특정한 방법 단계의 시퀀스(sequence)에서 표적 핵산 및 RNA를 반응 믹스에 동시에 첨가하고 그 후 본 발명의 캐스케이드 단백질 복합체를 별도로 이후에 첨가하는 것이 가능하다.
- [0110] 표적 핵산을 변형시키거나, 가시화하거나, 또는 표적 핵산의 전사를 활성화 또는 억제하는 것은 단리된 세포이든지 다중세포 조직, 기관 또는 유기체의 일부로서든지 간에 세포 내에서 원위지에서 이루어질 수 있다. 따라서, 전 조직 및 기관의 맥락에서, 그리고 유기체의 맥락에서, 본 방법은 생체 내에서 실시될 수 있거나, 이것은 전 조직, 기관 또는 유기체로부터 세포를 단리하고, 그 후, 동일한 유기체 내에서든지 상이한 유기체 내에서든지 간에, 리보핵단백질 복합체로 처리된 세포를 그의 이전의 위치, 또는 상이한 위치로 되돌림으로써 실시될 수 있다. 따라서, 본 방법은 동종이식편, 자가이식편, 동계이식편 및 이종이식편을 포함한다.
- [0111] 이들 실시 양태에서, 본 발명의 리보핵단백질 복합체 또는 캐스케이드 단백질 복합체는 세포 내로의 전달의 적절한 형태를 요구하며, 이는 당업계의 숙련자에게 공지되어 있는데, 이는 세포질 내로든지 핵 내로든지 간에 미세주입을 포함한다.
- [0112] 또한 별도로 존재할 때, RNA 분자는 캐스케이드 단백질 복합체와 동시에 전달되든지, 이와 별도로 전달되든지 순차적으로 전달되든지 간에, 세포 내로의 전달의 적절한 형태를 요구한다. 세포 내로의 RNA의 이러한 도입 형태는 당업계의 숙련자에게 공지되어 있으며, 통상적인 형질감염법을 통한 시험관 내 또는 생체 외(ex vivo) 전달을 포함할 수 있다. 물리적 방법, 예컨대 미세주입법 및 전기천공법과, 이외에도 칼슘 동시침전법 및 구매가능한 양이온성 중합체 및 지질, 및 세포 침투 펩티드, 세포 침투 입자(유전자 건(gene-gun))가 각각 사용될 수 있다. 예를 들어, 세포질로 전달되든지 핵으로 전달되든지 간에 바이러스를 전달 비히클로 사용할 수 있다 - 예를 들어, 본 발명의 캐스케이드 단백질 복합체 또는 본 발명의 리보핵단백질 복합체의 바이러스 입자에의 (가역성) 융합을 통하여 - . 바이러스 전달(예를 들어 아데노바이러스 전달) 또는 아그로박테리움(Agrobacterium) 매개된 전달이 이용될 수 있다.
- [0113] 또한 본 발명은 세포에서 표적 핵산을 변형시키거나, 가시화하거나, 또는 표적 핵산의 전사를 활성화 또는 억제하는 방법으로서 이상에서 기술된 발현 벡터들 중 임의의 것을 이용하여 세포를 형질감염시키거나, 형질전환시키거나 또는 형질도입시키는 단계를 포함하는 방법을 포함한다. 형질감염, 형질전환 또는 형질도입 방법은 당업계의 숙련자에게 공지된 유형의 것이다. 본 발명의 캐스케이드 복합체의 발현을 생성하는 데 사용되는 1가지의 발현 벡터가 있을 경우 그리고 RNA를 세포에 직접적으로 첨가한다면, 동일하거나 상이한 형질도입 방법, 형질전환 방법 또는 형질도입 방법이 이용될 수 있다. 이와 유사하게, 본 발명의 캐스케이드-기능성 융합 복합체의 발현을 생성하는 데 사용되는 1가지의 발현 벡터가 있다면 그리고 발현을 통하여 원위지에서 RNA를 생성하기 위하여 또 다른 발현 벡터가 사용되고 있다면 동일하거나 상이한 형질감염 방법, 형질전환 방법 또는 형질도입 방법이 이용될 수 있다.
- [0114] 다른 실시 양태에서, 본 발명의 캐스케이드 복합체를 코딩하는 mRNA를 세포 내로 도입하여서 캐스케이드 복합체를 그 세포 내에서 발현시킨다. 캐스케이드 복합체를 소정의 표적 서열로 안내하는 RNA도, 상기 mRNA와 동시에 도입하든지, 이와 별도로 도입하든지 순차적으로 도입하든지 간에, 세포 내로 도입하여서, 필요한 리보핵단백질 복합체가 그 세포 내에서 형성되게 한다.
- [0115] 표적 핵산을 변형 또는 가시화하는 전술한 방법에서, 추가의 아미노산 서열은 마커일 수 있으며, 마커는 표적 핵산과 결합되고; 바람직하게는 마커는 단백질; 임의로 형광 단백질, 예를 들어 녹색 형광 단백질(green fluorescent protein; GFP) 또는 황색 형광 단백질(yellow fluorescent protein; YFP) 또는 mCherry이다. 시험관 내이든지, 생체 외이든지 또는 시험관 내이든지 간에, 본 발명의 방법은 핵산 분자에서, 바람직하게는 더욱 고차의 구조체, 예컨대 수퍼코일형 플라스미드 또는 염색체의 형태, 또는 단일 가닥 표적 핵산, 예컨대 mRNA에서 표적 유전자좌를 직접적으로 가시화하는 데 사용될 수 있다. 표적 유전자좌의 직접적 가시화는 전자현미경법 또는 형광 현미경법을 이용할 수 있다.
- [0116] 작은 분자일 수 있는 스핀(spin) 표지체, 방사성 표지체 및 유기 염료 분자를 포함하는 다른 종류의 표지체를 사용하여 표적 핵산을 마킹할 수 있다.
- [0117] 상기에 기술된 본 발명의 방법에서, 표적 핵산은 DNA, 바람직하게는 dsDNA이지만, 표적은 RNA, 바람직하게는 mRNA일 수 있다.
- [0118] dsDNA인 표적 핵산을 변형시키거나, 가시화하거나, 전사 활성화시키거나 전사 억제하는 본 발명의 방법에서,

핵산 또는 염색질 변형 활성을 갖는 추가의 아미노산 서열은 뉴클레아제 또는 헬리카아제-뉴클레아제일 수 있으며, 변형은 바람직하게는 소정의 유전자좌에서 단일 가닥 또는 이중 가닥 절단이다. 이러한 방식으로, DNA의 독특한 서열 특이적 절단은 캐스케이드-기능성 모이어티 복합체를 사용하여 엔지니어링될 수 있다. 최종 리보핵단백질 복합체의 RNA 성분의 선택된 서열은 상기 추가의 아미노산 서열의 작용에 대하여 소정의 서열 특이성을 제공한다.

[0119] 따라서, 본 발명은 dsDNA 분자로부터 뉴클레오티드 서열의 적어도 일부분을 제거하고; 경우에 따라, 유전자(들)의 기능을 너아웃(knockout)시키기 위하여 소정의 유전자좌에서 세포 내에서 dsDNA 분자를 비상동 말단 결합시키는 방법을 또한 제공하며, 여기서, 본 방법은 이상에서 기술된 표적 핵산의 변형 방법들 중 임의의 것을 사용하여 이중 가닥 절단을 일으키는 단계를 포함한다.

[0120] 추가로 본 발명은 기존의 뉴클레오티드 서열을 변형시키거나 소정의 뉴클레오티드 서열을 삽입하기 위하여 소정의 유전자좌에서 세포 내에서 dsDNA 분자 내로 핵산을 상동 재조합시키는 방법을 제공하며, 여기서, 본 방법은 이상에서 기술된 표적 핵산의 변형 방법들 중 임의의 것을 사용하여 소정의 유전자좌에서 이중 또는 단일 가닥 절단을 하는 단계를 포함한다.

[0121] 따라서 본 발명은 이상에서 기술된 방법들 중 임의의 것에 따라 표적 핵산 서열을 변형시키거나, 표적 핵산의 전사를 활성화시키거나 표적 핵산의 전사를 억제하는 단계를 포함하는, 유기체에서 유전자 발현을 변형시키거나, 활성화시키거나 억제하는 방법을 또한 제공하며, 여기서, 핵산은 dsDNA이며, 기능성 모이어티는 DNA 변형 효소(예를 들어, 데메틸라아제 또는 데아세틸라아제), 전사 활성화자 또는 전사 억제자로부터 선택된다.

[0122] 추가로 본 발명은 이상에서 기술된 방법들 중 임의의 것에 따라 표적 핵산 서열을 변형시키거나, 표적 핵산 서열의 전사를 활성화 또는 억제하는 단계를 포함하는, 유기체에서 유전자 발현을 변형시키거나, 활성화시키거나 억제하는 방법을 제공하며, 여기서, 핵산은 mRNA이며, 기능성 모이어티는 리보뉴클레아제로서, 이는 임의로 엔도뉴클레아제, 3' 엑소뉴클레아제 또는 5' 엑소뉴클레아제로부터 선택된다.

[0123] 상기에 기술된 본 발명의 방법들 중 임의의 것에서, 본 방법에 처해지는 세포는 원핵 세포일 수 있다. 이와 유사하게, 세포는 진핵 세포, 예를 들어 식물 세포, 곤충 세포, 효모 세포, 진균류 세포, 포유류 세포 또는 인간 세포일 수 있다. 세포가 포유류 또는 인간의 것이면, 이것은 줄기 세포일 수 있다(그러나, 임의의 인간 배아 줄기 세포가 아닐 수 있다). 본 발명에서 사용하기 위한 이러한 줄기 세포는 바람직하게는 단리된 줄기 세포이다. 경우에 따라, 본 발명의 임의의 방법에 따르면, 세포는 시험관 내에서 형질감염된다.

[0124] 그러나 바람직하게는, 본 발명의 방법들 중 임의의 것에서, 표적 핵산은 특정한 3차 구조를 가지며, 임의로 수퍼코일형이며, 더 바람직하게는 표적 핵산은 네거티브 수퍼코일형이다. 유리하게는, 시험관 내에서 생성되든지 세포 내에서 형성되든지 세포의 발현 기구를 통하여 세포 내에서 형성되든지 간에, 본 발명의 리보핵단백질 복합체는, 특정 서열을 표지하든지 태그하든지, 핵산 구조를 변형시키든지, 유전자 발현을 스위치 온하든지 스위치 오프하든지, 표적 서열 그 자체의 변형이 단일 또는 이중 가닥 절단, 이어서 하나 이상의 뉴클레오티드 잔기 또는 카세트의 삽입을 포함하든지 간에, 소정의 성분의 기능적 활성을 적용하기 위하여 접근이 어려운 유전자좌를 표적화하는 데 사용될 수 있다.

[0125] 본 발명은 이상에서 기술된 본 발명의 캐스케이드 단백질 복합체 또는 리보핵단백질 복합체를 포함하는 약학 조성물을 또한 포함한다.

[0126] 본 발명은 이상에서 기술된 본 발명의 단리된 핵산 또는 발현 벡터를 포함하는 약학 조성물을 추가로 포함한다.

[0127] 또한 제공되는 것은 이상에서 기술된 본 발명의 캐스케이드 단백질 복합체 및 이상에서 기술된 본 발명의 RNA 분자를 포함하는 키트이다.

[0128] 본 발명은 약제로서 사용하기 위한, 이상에서 기술된 캐스케이드 단백질 복합체 또는 리보핵단백질 복합체 또는 핵산 또는 벡터를 포함한다.

[0129] 본 발명은 특정한 게놈 유전자좌에서 원핵 또는 진핵 숙주의 DNA를 물리적으로 변경시키거나, 주어진 유전자좌에서 소정 유전자의 발현 패턴을 변화시키는 다양한 가능성을 허용한다. 숙주 게놈 DNA는 적합한 캐스케이드 서브유닛에 융합된, 각각 뉴클레아제, 메틸라아제, 형광 단백질, 전사 활성화자 또는 억제자와 같은 기능성 도메인에 의해 절단되거나 메틸화에 의해 변형되거나, 형광에 의해 가시화되거나, 전사 활성화되거나 전사 억제될 수 있다. 게다가, 캐스케이드의 RNA 안내된 RNA 결합 능력은 형광 캐스케이드 융합 단백질을 이용한 생 세포에서의 RNA 트래피킹(trafficking)의 모니터링을 가능케 하며, 숙주 mRNA를 격리시키거나 파괴하여 숙주

세포의 유전자 발현 수준의 간섭을 야기하는 방법을 제공한다.

[0130] 본 발명의 방법들 중 임의의 것에서, 표적 핵산은 하기 뉴클레오티드 트리플렛(triplet)들 중 1가지 이상의 존재에 의해 규정될 수 있으며, 바람직하게는 dsDNA의 경우 상기 존재에 의해 그렇게 규정될 수 있다: 5'-CTT-3', 5'-CAT-3', 5'-CCT-3', 또는 5'-CTC-3'(또는 표적이 RNA일 경우 5'-CUU-3', 5'-CAU-3', 5'-CCU-3', 또는 5'-CTC-3). 트리플렛의 위치는 본 발명의 리보핵단백질의 RNA 분자 성분이 혼성화되는 서열에 인접하여 표적 가닥 내에 있다. 상기 트리플렛은 리보핵단백질의 RNA 분자 성분과의 염기쌍 형성이 표적의 5'에서 3'으로(하류) 일어나지 않는(한편, 이것은 본 발명의 리보핵단백질의 RNA 분자 성분의 RNA 서열의 바람직한 길이에 따른 지점으로부터 표적 서열의 상류에서 일어남) 표적 가닥 서열 내의 지점을 마킹한다. 천연 타입 I CRISPR 시스템의 맥락에서, 트리플렛은 "PAM"(프로토스페이서 인접 모티프)으로 공지된 것에 상응한다. ssDNA 또는 ssRNA 표적에 있어서, 상기 트리플렛들 중 하나의 존재는 그렇게 필요하지 않다.

[0131] 이제 본 발명을 상세하게 그리고 특정 실시예 및 도면을 참조로 하여 설명한다.

[0132] 도 1은 겔 시프트(gel-shift) 분석법의 결과를 나타내며, 여기서, 캐스케이드는 네거티브 슈퍼코일형(nSC) 플라스미드 DNA에는 결합하지만 이완 DNA에는 결합하지 않는다. 도 1의 A) 표적화 (J3) crRNA를 함유하는 J3-캐스케이드를 포함하는 nSC 플라스미드 DNA의 겔 시프트. 1:0.5의 몰비에서 최대 1:256의 몰비의, pUC-λ:캐스케이드에 의해, pUC-λ를 2배 증가량의 J3-캐스케이드와 혼합하였다. 제1 레인 및 마지막 레인은 단지 pUC-λ를 포함한다. 도 1의 B) 비표적화 (R44) crRNA를 함유하는 R44-캐스케이드를 이용한, 도 1의 A)와 같은 겔 시프트. 도 1의 C) Nt.BspQI Nick(nicked) pUC-λ를 이용한, 도 1의 A)와 같은 겔 시프트. 도 1의 D) PdmI 선형화 pUC-λ를 이용한, 도 1의 A)와 같은 겔 시프트. 도 1의 E) 자유 J3-캐스케이드의 농도에 대하여 도시된, J3-캐스케이드에 결합된 pUC-λ의 분율의 피트(fit)는 특이적 결합에 대한 해리 상수(Kd)를 제공한다. 도 1의 F) 자유 R44-캐스케이드의 농도에 대하여 도시된, R44-캐스케이드에 결합된 pUC-λ의 분율의 피트는 비특이적 결합에 대한 해리 상수(Kd)를 제공한다. 도 1의 G) 프로토스페이서 서열 내의 독특한 BsmI 제한효소 부위를 사용하여, 제한효소 분석에 의해 모니터링한 프로토스페이서에의 캐스케이드의 특이적 결합. 레인 1 및 5는 단지 pUC-λ를 포함한다. 레인 2 및 6은 pUC-λ가 캐스케이드와 혼합된 것을 포함한다. 레인 3 및 7은 pUC-λ를 캐스케이드와 혼합하고 후속적으로 BsmI를 첨가한 것을 포함한다. 레인 4 및 8은 pUC-λ를 BsmI와 혼합한 것을 포함한다. 도 1의 H) 플라스미드의 한 가닥을 후속적으로 Nt.BspQI로 절단한, 캐스케이드에 결합된 pUC-λ의 겔 시프트. 레인 1 및 6은 단지 pUC-λ를 포함한다. 레인 2 및 7은 pUC-λ가 캐스케이드와 혼합된 것을 포함한다. 레인 3 및 8은 pUC-λ가 캐스케이드와 혼합되고 후속적으로 Nt.BspQI 니킹(nicking)된 것을 포함한다. 레인 4 및 9는 pUC-λ가 캐스케이드와 혼합되고, 이어서 R-루프 내의 대체된 가닥에 대하여 상보성인 ssDNA 프로브를 첨가하고 후속적으로 Nt.BspQI로 니킹시킨 것을 포함한다. 레인 5 및 10은 pUC-λ가 Nt.BspQI로 니킹된 것을 포함한다. 도 1의 H) 플라스미드를 후속적으로 Nt.BspQI로 니킹시킨, 캐스케이드에 결합된 pUC-λ의 겔 시프트. 레인 1 및 6은 단지 pUC-λ를 포함한다. 레인 2 및 7은 pUC-λ가 캐스케이드와 혼합된 것을 포함한다. 레인 3 및 8은 pUC-λ가 캐스케이드와 혼합되고 후속적으로 Nt.BspQI로 절단된 것을 포함한다. 레인 4 및 9는 pUC-λ가 캐스케이드와 혼합되고, 이어서 R-루프 내의 대체된 가닥에 대하여 상보성인 ssDNA 프로브가 첨가되고, 후속적으로 Nt.BspQI로 절단된 것을 포함한다. 레인 5 및 10은 pUC-λ가 Nt.BspQI로 절단된 것을 포함한다. 도 1의 I) 플라스미드의 양 가닥이 후속적으로 EcoRI로 절단된, 캐스케이드에 결합된 pUC-λ의 겔 시프트. 레인 1 및 6은 단지 pUC-λ를 포함한다. 레인 2 및 7은 pUC-λ가 캐스케이드와 혼합된 것을 포함한다. 레인 3 및 8은 pUC-λ가 캐스케이드와 혼합되고 후속적으로 EcoRI로 절단된 것을 포함한다. 레인 4 및 9는 pUC-λ가 캐스케이드와 혼합되고, 이어서 R-루프 내의 대체된 가닥에 대하여 상보성인 ssDNA 프로브가 첨가되고 후속적으로 EcoRI로 절단된 것을 포함한다. 레인 5 및 10은 pUC-λ가 EcoRI로 절단된 것을 포함한다.

[0133] 도 2는 캐스케이드가 프로토스페이서 결합시에 표적 DNA의 굽힘을 유도하는 방법을 입증하는 주사 힘 현미경 사진을 나타낸다. 도 2의 A ~ P) 표적화 (J3) crRNA를 함유하는 J3-캐스케이드를 포함하는 nSC 플라스미드의 주사 힘 현미경법 영상. pUC-λ를 1:7의 pUC-λ:캐스케이드의 비로 J3-캐스케이드와 혼합하였다. 각각의 영상은 500 x 500 nm 표면 영역을 나타낸다. 백색 도트는 캐스케이드에 상응한다.

[0134] 도 3은 표적 인식시에 캐스케이드 및 Cas3이 상호작용함을 이분자 형광 상보(Bimolecular fluorescence complementation; BiFC) 분석법이 보여 주는 방법을 나타낸다. A) 파지 λ 게놈 상의 7개의 프로토스페이서를 표적화하는 CRISPR 7Tm 및 캐스케이드 ΔCse1과, Cse1-N155비너스(Venus) 및 Cas3-C85비너스 융합 단백질을 발현하는 세포의 비너스 형광. B) (A)에서의 세포의 명시야 영상. C) (A)와 (B)의 오버레이(overlay). D) 캐스케이드 ΔCse1 및 CRISPR7Tm과, Cse1-N155비너스 및 Cas3-C85비너스의 융합 단백질을 발현하는 파지 λ 감염

세포의 비너스 형광. **E)** (G)에서의 세포의 명시야 영상. **F)** (G)와 (H)의 오버레이. **G)** 캐스케이드  $\Delta$ Cse1 및 비표적화 CRISPR R44와, N155비너스 및 C85비너스의 단백질을 발현하는 파지  $\lambda$  감염 세포의 비너스 형광. **H)** (J)에서의 세포의 명시야 영상. **I)** (J)와 (K)의 오버레이. **J)** LSM 뷰어(카를 차이스(Carl Zeiss))의 프 로필 도구를 이용하여 결정할 경우, 각각의 주의 4 내지 7개의 개개의 세포의 형광 강도의 평균.

[0135] 도 4는 CRISPR 간섭 동안의 Cas3 뉴클레아제 및 헬리카아제 활성을 나타낸다. 도 4의 A) 캐스케이드, Cas3 돌연변이체 및 CRISPR J3을 발현하는 적격성(competent) BL21-AI 세포를 pUC- $\lambda$ 로 형질전환시켰다. 콜로니 형성 단위(Colony forming unit)/마이크로그람의 pUC- $\lambda$ (cfu/ $\mu$ g의 DNA)를 Cas3 돌연변이체를 발현하는 주들 각각에 대하여 도시한다. 야생형 Cas3 및 CRISPR J3 또는 CRISPR R44를 발현하는 세포는 각각 양성 대조구 및 음성 대조구로서의 역할을 한다. 도 4의 B) 캐스케이드, Cas3 돌연변이체 및 CRISPR 코딩 플라스미드를 지니고, 이 외에도 pUC- $\lambda$ 를 지닌 BL21-AI 세포를 *cas* 유전자 및 CRISPR의 발현을 억제하는 조건 하에서 성장시킨다. t=0 에서 발현을 유도한다. 시간이 지남에 따라 pUC- $\lambda$ 를 상실한 세포의 백분율을 나타내며, 이는 압피실린 민감 성 세포와 압피실린 내성 세포의 비로 결정되는 바와 같다.

[0136] 도 5는 캐스케이드-Cas3 융합 복합체가 생체 내에서 내성을 제공하고 시험관 내에서 뉴클레아제 활성을 갖게 되는 방법을 나타낸다. 도 5의 A) 정제된 캐스케이드 및 캐스케이드-Cas3 융합 복합체의 쿠마시 블루 (Coomassie Blue) 염색된 SDS-PAGE. 도 5의 B) 캐스케이드-Cas3 융합 복합체 및 표적화 (J3) 또는 비표적화 (R44) CRISPR를 발현하는 세포에서의 그리고 캐스케이드 및 Cas3이 별도로 표적화 (J3) CRISPR과 함께인 것을 발현하는 세포에서의 파지  $\lambda$ 의 플라크화 효율. 도 5의 C) J3-캐스케이드-Cas3 융합 복합체를 포함하는 nSC 표적 플라스미드의 겔 시프트(2가 금속 이온의 부재 하에서). 1:0.5의 몰비에서 최대 1:128의 몰비의, pUC- $\lambda$ :J3-캐스케이드-Cas3에 의해, pUC- $\lambda$ 를 2배 증가량의 J3-캐스케이드-Cas3과 혼합하였다. 제1 레인 및 마지막 레인은 단지 pUC- $\lambda$ 를 포함한다. 도 5의 D) J3-캐스케이드-Cas3 융합 복합체를 포함하는 nSC 비표적 플라스미드의 겔 시프트(2가 금속 이온의 부재 하에서). 1:0.5의 몰비에서 최대 1:128의 몰비의, pUC-p7:J3-캐스케이드-Cas3에 의해, pUC-p7을 2배 증가량의 J3-캐스케이드-Cas3과 혼합하였다. 제1 레인 및 마지막 레인은 단지 pUC-p7을 포함한다. 도 5의 E) 10 mM MgCl<sub>2</sub>의 존재 하에서 nSC 표적 플라스미드 (pUC- $\lambda$ , 좌 측) 또는 nSC 비표적 플라스미드 (pUC-p7, 우측)를 J3-캐스케이드-Cas3과 함께 인큐베이션시킴. 레인 1 및 7 은 단지 플라스미드를 포함한다. 도 5의 F) 2 mM ATP의 존재 하에서 도 5의 E)에서와 같이 분석. 도 5g) 돌연 변이 J3-캐스케이드-Cas3K320N 복합체를 이용하여 도 5의 E)에서와 같이 분석. 도 5의 H) 2 mM ATP의 존재 하 에서 도 5g)에서와 같이 분석.

[0137] 도 6은 이. 콜라이에서 CRISPR 간섭 타입 I 경로의 모델을 나타내는 개략도이다.

[0138] 도 7은 비상동성 말단 연결 또는 상동성 재조합 과정의 일부로서 블런트(blunt) 말단을 생성하도록 dsDNA를 절단하는 FokI 이량체를 생성하기 위하여 본 발명의 캐스케이드-FokI 융합물 실시 양태를 사용하는 방법을 나 타내는 개략도이다.

[0139] 도 8은 표적 인식시에 캐스케이드 및 Cas3이 상호작용함을 BiFC 분석법이 보여 주는 방법을 나타낸다. Cse1을 포함하지 않는 캐스케이드, Cse1-N155비너스 및 Cas3-C85비너스와, 파지 람다 게놈 상의 7개의 프로토스페이 서를 표적화하는 CRISPR 7Tm 또는 비표적화 CRISPRR44 중 어느 하나를 발현하는 세포의 명시야 영상 및 비너 스 형광의 오버레이. CRISPR 7Tm을 발현하는 세포는 파지 람다로 감염될 때에만 형광성인 반면, CRISPR R44를 발현하는 세포는 비형광성이다. 고도로 강렬한 형광 도트(세포 외부)는 광 반사 염 결정으로 인한 것이다. 백 색 막대는 10 마이크로미터에 상응한다.

[0140] 도 9는, CRISPR J3, 캐스케이드 및 Cas3(야생형 또는 S483AT485A)을 코딩하는 4개의 클론의 pUC- $\lambda$  서열[서열 번호 39 내지 42]은 이들이 프로토스페이의 (부분) 결실을 지니거나 시드(seed) 영역 내에 단일 점 돌연변 이를 지닌 탈출(escape) 돌연변이체임을 나타낸다는 것을 도시한다.

[0141] 도 10은 타입 I-ECRISPR/Cas 시스템을 함유하는 유기체로부터의 *cas3* 유전자들의 서열 정렬을 나타낸다. 스트 렙토마이세스(*Streptomyces*) 종 SPB78 유래의 *cas3-cse1* 유전자(제1 서열, 등록 번호 : ZP\_07272643.1) [서열 번호 43], 스트렙토마이세스 그리세우스(*Streptomyces griseus*)에서의 상기 유전자(제2 서열, 등록 번호: YP\_001825054) [서열 번호 44], 및 카테놀리스포라 아시디필라(*Catenulispora acidiphila*) DSM 44928에서의 상기 유전자(제3 서열, 등록 번호: YP\_003114638) [서열 번호 45]와, 에스. 그리세우스로부터의 폴리펩티드 링커 서열을 포함하는 인공 이. 콜라이 Cas3-Cse1 융합 단백질 [서열 번호 46]의 정렬.

[0142] 도 11은 단지 헤테로이량체가 KKR 및 ELD 뉴클레아제 도메인으로 이루어지도록 FokI 뉴클레아제 도메인들이 돌연변이된 그리고 반대쪽 결합 부위들 사이의 거리는 캐스케이드 뉴클레아제 쌍 사이의 최적 거리를 결정하

도록 변화될 수 있는 캐스케이드<sup>KKR/ELD</sup> 뉴클레아제 쌍의 설계를 나타낸다.

[0143] 도 12는 캐스케이드-FokI 뉴클레아제 쌍에 의한 게놈 표적화를 나타내는 개략도이다.

[0144] 도 13은 캐스케이드-뉴클레아제 복합체의 SDS PAGE 겔을 나타낸다.

[0145] 도 14는 플라스미드 DNA에서의 캐스케이드<sup>KKR/ELD</sup>의 시험관 내 절단 분석법의 전기영동 겔을 나타낸다.

[0146] 도 15는 캐스케이드<sup>KKR/ELD</sup>의 절단 패턴 및 빈도를 나타낸다 [서열 번호 47].

[0147] **실시예 - 사용한 재료 및 방법**

[0148] **균주, 유전자 클로닝, 플라스미드 및 벡터**

[0149] 이. 콜라이 BL21-AI 및 이. 콜라이 BL21(DE3) 균주를 전체에 걸쳐 사용하였다. 표 1에는 이 연구에서 사용한 모든 플라스미드가 열거되어 있다. 이전에 기술된 pWUR408, pWUR480, pWUR404 및 pWUR547을 스트렙(Strep)-태그 II R44-캐스케이드의 생성용으로 사용하였으며, pWUR408, pWUR514 및 pWUR630을 스트렙-태그 II J3-캐스케이드의 생성용으로 사용하였다(문헌[Jore et al., (2011) Nature Structural & Molecular Biology 18, 529-536]; 문헌[Semenova et al., (2011) Proceedings of the National Academy of Sciences of the United States of America 108, 10098-10103]). pUC-λ (pWUR610) 및 pUC-p7 (pWUR613)은 다른 곳에 기술되었다(문헌 [Jore et al., 2011]; 문헌[Semenova et al., 2011]). C85비너스 단백질은 pWUR647에 의해 코딩되는데, 이는 BamHI 부위와 NotI 부위 사이에 클로닝된 합성 GA1070943 구성물(표 2)(진아트(Geneart))을 포함하는 pET52b (노바겐(Novagen))에 상응한다. N155비너스 단백질은 pWUR648에 의해 코딩되는데, 이는 NotI 부위와 XhoI 부위 사이에 클로닝된 합성 GA1070941 구성물(표 2)(진아트)을 포함하는 pRSF1b(노바겐)에 상응한다. Cas3-C85비너스 융합 단백질은 pWUR649에 의해 코딩되는데, 이는 NcoI 부위와 BamHI 부위 사이에 프라이머 BG3186 및 BG3213(표 3)을 이용한 Cas3 증폭 생성물을 포함하는 pWUR647에 상응한다. CasA-N155비너스 융합 단백질은 pWUR650에 의해 코딩되는데, 이는 NcoI 부위와 BamHI 부위 사이에 프라이머 BG3303 및 BG3212(표 3)를 사용한 CasA 증폭 생성물을 포함하는 pWUR648에 상응한다. CRISPR 7Tm은 pWUR651에 의해 코딩되는데, 이는 NcoI 부위와 KpnI 부위 사이에 클로닝된 합성 GA1068859 구성물(표 2)(진아트)을 포함하는 pACYCDuet-1(노바겐)에 상응한다. 캐스케이드 코딩 pWUR400, 캐스케이드ΔCse1 코딩 WUR401 및 Cas3 코딩 pWUR397은 이전에 기술되었다(문헌[Jore et al., 2011]). Cas3H74A 코딩 pWUR652를 프라이머 BG3093, BG3094(표 3)를 이용하여 pWUR397의 부위 지정 돌연변이 유발을 이용하여 구성하였다.

**표 1**

사용한 플라스미드

[0150]

플라스미드	유전자의 설명 및 순서 (5'-3')	제한효소 부위	프라이머	출처
pWUR397	pRSF-1b 중 <i>cas3</i> , 태그 없음			1
pWUR400	pCDF-1b 중 <i>casA-casB-casC-casD-casE</i> , 태그 없음			1
pWUR401	pCDF-1b 중 <i>casB-casC-casD-casE</i> , 태그 없음			1
pWUR404	pCDF-1b 중 <i>casE</i> , 태그 없음			1
pWUR408	pRSF-1b 중 <i>casA</i> , 태그 없음			1
pWUR480	pET52b 중의, Strep-tag II (N-말단)를 포함하는 <i>casB-casC-casD</i>			1
pWUR514	pET52b 중의, Strep-tag II (N-말단)를 포함하는 <i>casB-casC-casD-CasE</i>			2
pWUR547	pACYCDuet-1 중의, 이. 콜라이 R44 CRISPR, 7x 스페이서 nr. 2			2
pWUR613	pUC-p7; 350 bp 파지 P7 앰플리콘(amplicon) 상의 R44 프로토스페이서를 함유하는 pUC19			2
pWUR630	pACYCDuet-1 중 CRISPR 폴리 J3, 5x 스페이서 J3	NcoI/KpnI		본 연구
pWUR610	pUC-λ; 350 bp 파지 λ 앰플리콘 상의 J3 프로토스페이서를 함유하는 pUC19			3
pWUR647	C85비너스; pET52b 중 GA1070943 (표 S1)	BamHI/NotI		본 연구

pWUR648	<i>N155비너스</i> ; pRSF1b 중 GA1070941 (표 S1)	Not I/Xho I			본 연구	연
pWUR649	<i>cas3-C85비너스</i> ; <i>cas3</i> 앰플리콘을 함유하는 pWUR647	NcoI/Bam HI	BG3186 BG3213	+	본 연구	연
pWUR650	<i>casA-N155비너스</i> ; <i>casA</i> 앰플리콘을 함유하는 pWUR648	NcoI/Not I	BG3303 BG3212	+	본 연구	연
pWUR651	CRISPR 7Tm; pACYCDuet-1 중 GA1068859 (표 S1)	NcoI/Kpn I			본 연구	연
	pCDF-1b 중의, 스트랩트-태그 II (N-말단)를 포함하는 <i>casB-casC-casD-CasE</i>				본 연구	연
	<i>cas3-casA</i> 융합물				본 연구	연
	<i>cas3H74A-CasA</i> 융합물				본 연구	연
	<i>cas3D75A-CasA</i> 융합물				본 연구	연
	<i>cas3K320N-CasA</i> 융합물				본 연구	연
	<i>cas3D452N-CasA</i> 융합물				본 연구	연

[0151] 상기 표에서 출처 1은 문헌[Brouns et al (2008) Science 321, 960-964]이다.

[0152] 상기 표에서 출처 2는 문헌[Jore et al (2011) Nature Structural & Molecular Biology 18:529 - 537]이다.

**표 2**

**합성 구성물**

GA1070943

ACTGGAAGCGGGCAGTGAAAGGAAGGCCATGAGGCCAGTTAATTAAGCGGA  
 TCCTGGCGGGCAGCGGGCGGGCAGCGACAAGCAGAAGAACGGCATCAAGG  
 CGAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACC  
 ACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGCTGCTGCCGACAACC  
 ACTACCTGAGCTACCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC  
 ACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGA  
 GCTGTACAAGTAAGCGGCCGCGGCGCGCTAGGCCCTGACGGCCTTCCTTCAATT  
 CGCCCTATAGTGAG [서열 번호 6]

[0153]

GA1070941

CACTATAGGGCGAATTGGCGGAAGGCCGTCAGGCCGATTTAATTAAGCGGCC  
 GCAGGCGGCGCAGCGGCGGGCAGCATGGTGAGCAAGGGCGAGGAGCTGTT  
 CACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAA  
 GTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT  
 GAAGCTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACC  
 ACCCTCGGCTACGGCCTGCAGTGCTTCGCCCCGCTACCCCGACCACATGAAGCAGC  
 ACGACTTCTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTT  
 CTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGA  
 CACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA  
 CATCTGGGGACAAGCTGGAGTACAACAGCCACAACGTCTATATCAC  
 GGCCTAACTCGAGGGCGCGCCCTGGGCCTCATGGGCCTTCCGCTCACTGCCCGCT  
 TTCCAG [서열 번호 7]

[0154]

GA1068859

CACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGCATGAGCTCCATGGAAA  
CAAAGAATTAGCTGATCTTTAATAATAAGGAAATGTTACATTAAGTTGGTGGGT  
TGTTTTTATGGGAAAAAATGCTTTAAGAACAAATGTATACTTTTAGAGAGTTCCC

[0155]

CGCGCCAGCGGGGATAAACCGGGCCGATTGAAGGTCCGGTGGATGGCTTAAAAG  
AGTTCCCCGCGCCAGCGGGGATAAACCGCCGAGGTACAGCAGGTAGCGCAGAT  
CATCAAGAGTTCCCCGCGCCAGCGGGGATAAACCGACTTCTCTCCGAAAAGTCA  
GGACGCTGTGGCAGAGTTCCCCGCGCCAGCGGGGATAAACCGCCTACGCGCTGA  
ACGCCAGCGGTGTGGTGAATGAGTTCCCCGCGCCAGCGGGGATAAACCGGTGTG  
GCCATGCACGCCTTTAACGGTGAAGTGGAGTTCCCCGCGCCAGCGGGGATAAAC  
CGCACGAACTCAGCCAGAACGACAAAACAAAAGGCGAGTTCCCCGCGCCAGCGG  
GGATAAACCGGCACCAGTACGCGCCCCACGCTGACGGTTTCTGAGTTCCCCGCGC  
CAGCGGGGATAAACCGCAGCTCCCATTTTCAAACCCAGGTACCCTGGGCCTCATG  
GGCCTTCGCTCACTGCCCGCTTTCCAG [서열 번호 8]

[0156]

GA1047360

GAGCTCCCGGGCTGACGGTAATAGAGGCACCTACAGGCTCCGGTAAAACGGAAA  
CAGCGCTGGCCTATGCTTGAAACTTATTGATCAACAAATTGCGGATAGTGTAT  
TTTTGCCCTCCCAACACAAGCTACCGCGAATGCTATGCTTACGAGAATGGAAGCG  
AGCGCGAGCCACTTATTTTCATCCCCAAATCTTATTCTTGCTCATGGCAATTCAG  
GTTTAACCACCTCTTCAATCAATAAAATCACGCGCGATTACTGAACAGGGGCAA  
GAAGAAGCGTGGGTTCAGTGTGTCAGTGGTGTGTCACAAAGCAATAAGAAAAGTG  
TTTCTTGGGCAAATCGGCGTTTGCACGATTGATCAGGTGTTGATTTCCGGTATTGCC  
AGTAAACACCGCTTTATCCGTGGTTTGGGAATTGGTAGATCTGTTTTAATTGTTA  
ATGAAGTTCATGCTTACGACACCTATATGAACGGCTTGCTCGAGGCAGTGCTCAA  
GGCTCAGGCTGATGTGGGAGGGAGTGTATTCTTCTTTCCGCAACCCTACCAATG  
AAACAAAAACAGAAGCTTCTGGATACTTATGGTCTGCATACAGATCCAGTGGAA  
AATAACTCCGCATATCCACTATTAAGTGGCGAGGTGTGAATGGTGCGCAACGTT  
TTGATCTGCTAGCGGATCCGGTACC [서열 번호 9]

[0157]

표 3

프라이머

BG3186	ATAGCGCCATGGAACCTTTTAAATATATATGCCATTA [서열 번호 10]
BG3213	ACAGTGGGATCCGCTTTGGGATTGTCAGGGATGACTCTGGT [서열 번호 11]
BG3303	ATAGCGTCATGAATTTGCTTATTGATAACTGGATTCCCTGTACG [서열 번호 12]

[0158]

BG3212	ACAGTGGCGGCCGCGCCATTTGATGGCCCTCCTTGCGGTTTTAA [서열 번호 13]
BG3076	CGTATATCAAACCTTTCCAATAGCATGAAGAGCAATGAAAAATAAC [서열 번호 14]
BG3449	ATGATACCGCGAGACCCACGCTC [서열 번호 15]
BG3451	CGGATAAAGTTGCAGGACCACTTC [서열 번호 16]

[0159]

[0160] **단백질 생성 및 정제**

[0161] 캐스케이드를 기술된 바와 같이 발현시키고 정제하였다(문헌[Jore et al., 2011]). 정제 전체에 걸쳐, pH 7.5의 20 mM HEPES, 75 mM NaCl, 1 mM DTT, 2 mM EDTA를 함유하는 완충제를 재현탁 및 세척에 사용하였다. 4 mM 테스티오바이오틴을 함유하는 상기 완충제에서 단백질 용출을 수행하였다. 캐스케이드-Cas3 융합 복합체를 동일한 방식으로 발현시키고 정제하였으며, 이때 세척 단계는 pH7.5의 20 mM HEPES, 200 mM NaCl 및 1 mM DTT를 이용하여 수행하고 용출은 pH 7.5의 20 mM HEPES, 75 mM NaCl, 1 mM DTT(4 mM 테스티오바이오틴을 함유함)에서 수행하였다.

[0162] **전기영동 이동성 시프트 분석**

[0163] 정제한 캐스케이드 또는 캐스케이드 부분복합체(subcomplex)를 pH 7.5의 20 mM HEPES, 75 mM NaCl, 1 mM DTT, 2 mM EDTA를 함유하는 완충제에서 pUC-λ와 혼합하고, 37°C에서 15분 동안 인큐베이션하였다. 샘플들을 0.8% TAE 아가로스(Agarose) 겔 상에서 하룻밤 러닝시키고(run), TAE 중에 1:10000으로 희석시킨 SyBR 세이프(safe)(인비트로젠)로 30분 동안 사후 염색시켰다. BsmI(퍼멘타스(Fermentas)) 또는 Nt.BspQI (뉴 잉글랜드 바이오랩스(New England Biolabs))를 이용한 절단을 5 mM MgCl<sub>2</sub>가 보충된 HEPES 반응 완충제에서 수행하였다.

[0164] **주사 힘 현미경법(Scanning Force Microscopy)**

[0165] 정제한 캐스케이드를 pH 7.5의 20 mM HEPES, 75 mM NaCl, 0.2 mM DTT, 0.3 mM EDTA를 함유하는 완충제에서 pUC-λ (7:1의 비, 50 nM 캐스케이드, 35 nM DNA)와 혼합하고, 37°C에서 15분 동안 인큐베이션하였다. 후속적으로, AFM 샘플 제조를 위하여, 상기 인큐베이션 혼합물을 이중 증류수 중에 10배로 희석시키고, MgCl<sub>2</sub>를 1.2 mM의 최종 농도로 첨가하였다. 상기 단백질-DNA 복합체의 침적 및 영상화를 이전에 기술된 바와 같이 실시하였다(문헌[Dame et al., (2000) Nucleic Acids Res. 28: 3504 - 3510]).

[0166] **형광 현미경법**

[0167] cas 유전자 코딩 플라스미드 상에 CRISPR을 지닌 BL21-AI 세포를 암피실린 (100 µg/ml), 카나마이신 (50 µg/ml), 스트렙토마이신 (50 µg/ml) 및 클로람페니콜 (34 µg/ml)을 함유하는 루리아-베르타니 브로쓰(Luria-Bertani broth; LB)에서 37°C에서 하룻밤 성장시켰다. 하룻밤 배양물을 신선한 항생제 함유 LB 중에 1:100으로 희석시키고, 37°C에서 1시간 동안 성장시켰다. cas 유전자 및 CRISPR의 발현은 L-아라비노스를 0.2%의 최종 농도로 그리고 IPTG를 1 mM의 최종 농도로 첨가하여 1시간 동안 유도하였다. 감염을 위하여, 세포를 4의 감염 다중도(Multiplicity of Infection; MOI)로 파지 람다와 혼합시켰다. 세포를 폴리-L-라이신으로 덮인 현미경 슬라이드에 적용하고, 40x 유침 대물렌즈(1.3의 N.A.) 및 여기 광원(514 nm)으로서의 아르곤 레이저 및 530-600 nm에서의 검출을 이용하여, 악시오버트(Axiovert) 도립 현미경을 기반으로 한 차이스 LSM510 공초점 레이저 주사 현미경을 이용하여 분석하였다. 모든 측정에 있어서 핀홀을 203 µm로 설정하였다.

[0168] **pUC-λ 형질전환 연구**

[0169] 카나마이신 (50 µg/ml), 스트렙토마이신 (50 µg/ml) 및 클로람페니콜 (34 µg/ml)을 함유하는 LB에 하룻밤 예비접종물을 접종하고, 이를 0.3의 OD<sub>600</sub>으로 성장시켰다. cas 유전자 및 CRISPR의 발현을 0.2% L-아라비노스 및 1 mM IPTG를 이용하여 45분 동안 유도하였다. 세포를 4°C에서 원심분리함으로써 수집하고, 100 mM RbCl<sub>2</sub>, 50 mM MnCl<sub>2</sub>, 30 mM 아세트산칼륨, 10 mM CaCl<sub>2</sub> 및 15% 글리세롤을 함유하는 빙냉 완충제(pH 5.8) 중에 재현탁 시킴으로써 적격성으로 만들었다. 3시간 인큐베이션한 후, 세포를 수집하여 10 mM MOPS, 10 mM RbCl, 75 mM CaCl<sub>2</sub>, 15% 글리세롤을 함유하는 완충제(pH 6.8) 중에 재현탁시켰다. 80 ng pUC-λ의 첨가, 이어서 42°C에서 1분간의 열충격(heat-shock), 및 얼음 상에서의 5분간의 저온 충격(cold-shock)에 의해 형질전환을 수행하였다. 다음, 세포를 LB에서 37°C에서 45분 동안 성장시킨 후, 0.2% L-아라비노스, 1 mM IPTG, 암피실린 (100 µg/ml), 카나마이신 (50 µg/ml), 스트렙토마이신 (50 µg/ml) 및 클로람페니콜 (34 µg/ml)을 함유하는 LB-한천 플레이트 상에 도말하였다.

[0170] cas 유전자 및 CRISPR 코딩 플라스미드를 함유하는 BL21-AI 세포를 pUC-λ로 형질전환시키는 한편, 상기 세포를 0.2% 글루코스의 존재 하에 성장시켜서 T7-폴리머라아제 유전자의 발현을 억제하여 플라스미드 큐어링(curing)을 분석하였다. cas 유전자 및 CRISPR의 발현은 상기 세포를 수집하고 0.2% 아라비노스 및 1 mM IPTG를 함유하는 LB 중에 재현탁시킴으로써 유도하였다. 세포를 스트렙토마이신, 카나마이신 및 클로람페니콜 (pUC-λ에 대하여 비선택적) 또는 암피실린, 스트렙토마이신, 카나마이신 및 클로람페니콜 (pUC-λ에 대하여

선택적) 중 어느 한 가지를 함유하는 LB-한천 상에 도달하였다. 하룻밤 성장시킨 후, 플라스미드 상실의 백분율을 선택적 플레이트 및 비선택적 플레이트 상에서의 콜로니 형성 단위의 비로부터 계산할 수 있다.

[0171] **파지 람다 감염 연구**

[0172] 파지 감염에 대한 숙주의 민감성을 문헌[Brouns et al (2008) Science 321, 960-964]에서와 같이, 독성 파지 람다( $\lambda_{vir}$ )를 사용하여 테스트하였다. 감염에 대한 숙주의 민감성은 문헌[Brouns et al (2008)]에 기술된 바와 같이 플라크화 효율(비표적화 R44 CRISPR을 함유하는 주의 플라크 카운트에 대한 항- $\lambda$  CRISPR을 함유하는 주의 플라크 카운트의 비)을 계산하였다.

[0173] **실시예 1 - 캐스케이드는 네거티브 수퍼코일형 표적 DNA에 배타적으로 결합한다**

[0174] pUC- $\lambda$ 로 나타낸 3 kb pUC19 유래 플라스미드는 J3-캐스케이드(스페이서 J3를 함유하는 crRNA와 연관된 캐스케이드(문헌[Westra et al (2010) Molecular Microbiology 77, 1380-1393])에 의해 표적화되는 파지  $\lambda$ 의 J 유전자의 일부에 상응하는 350 bp DNA 단편을 함유한다. 전기영동 이동성 시프트 분석은, 캐스케이드가 단지 네거티브 수퍼코일형(nSC) 표적 플라스미드에 대하여 높은 친화도를 가짐을 나타낸다. 6:1의 J3-캐스케이드:pUC- $\lambda$ 의 몰비에서, 모든 nSC 플라스미드에 캐스케이드가 결합하였으며(도 1의 A 참조), 반면 비표적화 crRNA R44를 지닌 캐스케이드(R44-캐스케이드)는 128:1의 몰비에서 비특이적 결합을 나타냈다(도 1의 B 참조). nSC pUC- $\lambda$ 의 해리 상수(Kd)는 J3-캐스케이드의 경우  $13 \pm 1.4$  nM이고(도 1의 E 참조) R44-캐스케이드의 경우  $429 \pm 152$  nM인 것으로 결정되었다(도 1의 F 참조).

[0175] J3-캐스케이드는 이완된 표적 DNA, 예컨대 닉(도 1의 C 참조) 또는 선형 pUC- $\lambda$ (도 1의 D 참조)에는 측정가능한 친화도로 결합할 수 없었으며, 이는 캐스케이드가 nSC 토폴로지를 갖는 더욱 큰 DNA 기질에 대하여 높은 친화도를 가짐을 나타내는 것이었다.

[0176] 비특이적 결합과 특이적 결합을 구별하기 위하여, 프로토스페이서 내에 위치하는 BsmI 제한효소 부위를 사용하였다. BsmI 효소를 pUC- $\lambda$ 에 첨가하면 R44-캐스케이드의 존재 하에서 선형 생성물이 제공되지만 (도 1의 G, 라인 4 참조), pUC- $\lambda$ 는 J3-캐스케이드의 존재 하에서는 BsmI 절단으로부터 보호되는데(도 1의 G, 라인 7 참조), 이는 프로토스페이서에서의 특이적 결합을 나타내는 것이다. 이는 nSC 플라스미드에 있어서 프로토스페이서 서열에의 캐스케이드의 시험관 내에서의 서열 특이적 결합에 Cas3이 요구되지 않음을 나타낸다.

[0177] nSC pUC- $\lambda$ 에의 캐스케이드의 결합에 이어서 Nt.BspQI로 니킹시키면, OC 토폴로지가 생긴다. 캐스케이드는 가닥 니킹 후 플라스미드로부터 방출되며, 이는 이동성 시프트의 부재에 의해 알 수 있는 바와 같다(도 1의 H 참조, 라인 8을 라인 10과 비교). 이와는 대조적으로, 캐스케이드는, 대체된 가닥에 대하여 상보성인 ssDNA 프로브를 반응물에 첨가한 후 Nt.BspQI로 DNA를 절단할 때 그의 DNA 표적에 여전히 결합된 채로 있다(도 1의 H, 라인 9 참조). 프로브는 이완된 표적 DNA 상의 캐스케이드 R-루프를 인공적으로 안정화시킨다. 캐스케이드 결합 후 pUC- $\lambda$ 의 둘 모두의 DNA 가닥이 절단될 때 유사한 관찰이 이루어진다(도 1의 I, 라인 8 및 라인 9 참조).

[0178] **실시예 2- 캐스케이드는 결합된 표적 dna의 굽힘을 유도한다**

[0179] 정제된 캐스케이드와 pUC- $\lambda$  사이에 형성된 복합체를 가시화하였다. 단일 결합 J3-캐스케이드 복합체를 함유하는 특정 복합체들이 형성된 반면, 비특이적 R44-캐스케이드는 동일한 조건 하에서 이 분석법에서 어떠한 DNA 결합 복합체도 생성하지 않는다. 관찰된 81가지의 DNA 분자 중 76%는 J3-캐스케이드가 결합된 것으로 밝혀졌다(도 2의 A~P 참조). 이들 복합체 중 대부분의 경우에 캐스케이드는 루프의 정점(apex)에서 발견되었으며(86%), 반면에, 단지 작은 분율이 비정점 위치에서 발견되었다(14%). 이들 데이터는, 아마도 DNA 이중가닥의 국소적 용융을 돕기 위하여 캐스케이드 결합이 DNA의 굽힘 및 가능하게는 랩핑(wrapping)을 야기함을 나타낸다.

[0180] **실시예 3 - Cas3 및 Cse1: Cas3의 자연 발생 융합물은 프로토스페이서 인식시에 캐스케이드와 상호작용한다**

[0181] 도 S3은, Cas3 및 Cse1이 스트렙토마이세스 종 SPB78(등록 번호: ZP\_07272643.1)에서, 스트렙토마이세스 그리세우스(등록 번호 YP\_001825054)에서, 그리고 카테눌리스포라 애시디필라(*Catenulispora acidiphila*) DSM 44928 (등록 번호 YP\_003114638)에서 융합 단백질로서 나타남을 타입 I-ECRISPR/Cas 시스템을 함유하는 유기체로부터의 cas3 유전자의 서열 분석이 보여 줌을 나타낸다.

[0182] **실시예 4 - 이분자 형광 상보(BiFC)는 캐스케이드의 일부를 형성하는 Cse1 융합 단백질이 계속하여 Cas3과 상호작용하는 방법을 나타낸다**

- [0183] 파지  $\lambda$  감염 전 및 후에 생체 내에서의 Cas3과 캐스케이드 사이의 상호작용을 모니터링하기 위하여 BiFC 실험을 이용하였다. BiFC 실험은 형광 단백질, 예를 들어 황색 형광 단백질(YFP)의 비형광성 절반이, 두 절반이 아주 근접하여 나타날 때 재폴딩하여(refold) 형광 분자를 형성하는 능력에 의존한다. 이와 같이, 이것은 단백질-단백질 상호작용을 보여 주는 도구를 제공하며, 그 이유는 국소 농도가 높을 경우, 예를 들어, 형광 단백질의 두 절반이 상호작용 파트너에 융합될 때 재폴딩 효율이 크게 향상되기 때문이다. Cse1은 YFP의 개선된 버전인 비너스의 N-말단 155개 아미노산과 C-말단에서 융합되었다(Cse1-N155비너스)(문헌[Nagai et al (2002) Nature Biotechnology 20, 87-90]). Cas3은 비너스의 C-말단 85개 아미노산에 C-말단에서 융합되었다(Cas3-C85비너스).
- [0184] BiFC 분석은, 캐스케이드가 침입 DNA의 부재 하에서 Cas3과 상호작용하지 않음을 보여 준다(도 3의 ABC, 도 3의 P 및 도 8). 그러나, 파지  $\lambda$  로 감염시에, 캐스케이드 $\Delta$ Cse1, Cse1-N155비너스 및 Cas3-C85비너스를 발현하는 세포는, 이들이 항- $\lambda$  CRISPR 7Tm을 동시 발현할 경우 형광성이다(도 3의 DEF, 도 3의 P 및 도 8). 이들이 비표적화 CRISPR R44를 동시 발현할 때(도 3의 GHI, 도 3의 P 및 도 8), 세포는 여전히 비형광성인 채로 있다. 이는 캐스케이드 및 Cas3이 프로토스페이서 인식시에 감염 동안 특이적으로 상호작용하고, 캐스케이드-Cas3 이원 이펙터(effector) 복합체에서 Cse1 및 Cas3이 서로와 아주 근접하여 있음을 나타낸다.
- [0185] 또한 이들 결과는, 심지어 Cas3 그 자신이 또한 융합 단백질일 때에도, Cse1과 이중성 단백질의 융합물이 캐스케이드 및 crRNA의 리보핵단백질 형성을 파괴하지도 않고, 이것이 캐스케이드 및 Cas3과 표적 파지 DNA의 상호작용을 파괴하지도 않음을 꽤 명백하게 나타낸다.
- [0186] **실시예 5 - 설계된 Cas3-Cse1 융합물의 제조는 생체 내에서 기능적 활성을 갖는 단백질을 제공한다**
- [0187] Cas3 DNA 절단 활성화에 대한 시험관 내 증거의 제공은 정제된 활성 Cas3을 요구하였다. 다양한 가용화 전략에도 불구하고, 이. 콜라이 BL21에서 과다생성된 Cas3(문헌[Howard et al (2011) Biochem. J. 439, 85-95])은 주로 불활성 응집체 및 봉입체 형태로 존재한다. 따라서, Cas3은 에스. 그리세우스에서 Cas3-Cse1 융합 단백질의 링커와 동일한 링커를 함유하는 Cas3-Cse1 융합 단백질로서 생성되었다(도 10 참조). 캐스케이드 $\Delta$ Cse1 및 CRISPR J3과 동시 발현될 때, 상기 융합물-복합체는 가용성이었으며, 캐스케이드와 동일한 겔보기 화학량으로 고순도로 수득되었다(도 5의 A). 이 복합체의 기능성을 파지  $\lambda$  감염에 대한 내성 제공에 대하여 테스트할 때, 융합물-복합체 J3-캐스케이드-Cas3을 발현하는 세포에서의 플라크화 효율(efficiency of plaquing; eop)은 별도의 단백질들을 발현하는 세포에서와 동일하였다(도 5의 B).
- [0188] J3-캐스케이드-Cas3 융합물-복합물은 생체 내에서 기능성이기 때문에, 이 복합체를 이용하여 시험관 내 DNA 절단 분석을 실시하였다. J3-캐스케이드-Cas3을 2가 금속의 부재 하에서 pUC- $\lambda$ 와 함께 인큐베이션할 때, 캐스케이드에 대하여 관찰된 것과 유사한 몰비에서 플라스미드 결합이 관찰된 반면(도 5의 C), 비표적 플라스미드 (pUC-p7, pUC- $\lambda$ 와 동일한 크기의, 그러나 프로토스페이서가 결여된 pUC19 유래 플라스미드)에의 특이적 결합은 높은 몰비에서만 일어났으며(도 5의 D), 이는 상기 복합체의 특이적인 DNA 결합이 캐스케이드 단독의 것과 또한 유사함을 나타내는 것이었다.
- [0189] 흥미롭게도, J3-캐스케이드-Cas3 융합 복합체는 nSC 표적 플라스미드에서 마그네슘 의존성 엔도뉴클레아제 활성을 나타낸다. 10 mM Mg<sup>2+</sup>의 존재 하에서 J3-캐스케이드-Cas3은 nSC pUC- $\lambda$ 를 니킹시키지만 (도 5의 E, 라인 3-7), 표적 서열을 함유하지 않거나(도 5의 E, 라인 9-13) 이완된 토폴로지를 갖는 기질에 대해서는 절단이 관찰되지 않는다. 생성된 OC 밴드의 시프트는 관찰되지 않으며, 이는 ATP 의존성 Cas3 헬리카아제 활성을 요구하지 않고서, 절단 후에 캐스케이드가 자발적으로 해리된다는 이전의 관찰과 일치한다. 대신, Cas3의 헬리카아제 활성은 엑소뉴클레오라이트(exonucleolytic) 플라스미드 분해에 연루된 것으로 보인다. 마그네슘 및 ATP 둘 모두를 반응물에 첨가할 때, 완전한 플라스미드 분해가 일어났다(도 5의 H).
- [0190] 본 발명자는 캐스케이드 단독이 이완된 DNA 상의 프로토스페이서에 결합할 수 없음을 알아냈다. 이와는 대조적으로, 본 발명자는 캐스케이드가 네거티브 슈퍼코일형 DNA 내의 표적을 효율적으로 찾아내고, 후속적으로 Cse1 서브유닛을 통하여 Cas3을 모집한다는 것을 알아냈다. Cas3 HD-뉴클레아제 도메인에 의한 엔도뉴클레오라이트 절단은 슈퍼코일링의 상실을 통하여 DNA로부터의 캐스케이드의 자발적 방출을 야기하며, 이는 캐스케이드를 재이동시켜 새로운 표적을 찾아내게 한다. 그 후, 표적은 Cas3의 공동 ATP 의존성 헬리카아제 활성 및 HD 뉴클레아제 활성에 의해 점진적으로 풀려져서 절단되며, 이는 완전한 표적 DNA 분해 및 침입자 중화에 이르게 된다.
- [0191] 도 6을 참조하면, 그리고 임의의 특정한 이론에 구애되고자 함이 없이, 이. 콜라이에서의 CRISPR-간접 타입 I

경로의 작동 기작은 하기를 포함할 수 있다: (1) 먼저, crRNA를 지닌 캐스케이드는 인접 PAM에 의해 프로토스페이서에 대하여 nSC 플라스미드 DNA를 스캔한다. 이 단계 동안 가닥 분리가 일어나는지에 관해서는 공지되어 있지 않다. (2) 서열 특이적 프로토스페이서 결합은 R-루프를 형성하는, crRNA와 DNA의 상보성 가닥 사이의 염기쌍 형성을 통하여 달성된다. 결합시에, 캐스케이드는 DNA의 굽힘을 유도한다. (3) 캐스케이드의 Cse1 서브유닛은 DNA 결합시에 Cas3을 모집한다. 이는 핵산 결합시에 일어나는 캐스케이드의 배좌 변화에 의해 달성될 수 있다. (4) Cas3의 HD 도메인(더욱 어두운 부분)은 R-루프의 대체된 가닥의 Mg<sup>2+</sup> 의존성 니킹을 촉매함으로써 표적 플라스미드의 토폴로지를 nSC로부터 이완된 OC로 변경시킨다. (5a 및 5b) 플라스미드 이완은 캐스케이드의 자발적 해리를 야기한다. 한편, Cas3은 연속적인 절단 활성을 위하여 표적 dsDNA 풀림을 위한 헬리카아제 도메인 및 HD-뉴클레아제 도메인을 요구하는, 표적 플라스미드에서의 ATP 의존성 엑소뉴클레아제 활성을 나타낸다. (6) Cas3은 전체 플라스미드를 ATP 의존적인 방식으로 분해시키는데, 그 이유는 이것이 표적 dsDNA를 따라서 점진적으로 이동하고, 표적 dsDNA를 풀어서 절단하기 때문이다.

[0192] **실시예 6 - 인공 Cas-스트랩 태그 융합 단백질의 제조 및 캐스케이드 복합체의 어셈블리**

[0193] 문헌[Jore et al (2011) Nature Structural & Molecular Biology 18: 529 - 537]의 추가 표 3에 열거된 발현 플라스미드를 사용하여, 문헌[Brouns et al (2008) Science 321: 960-4 (2008)]에 기술된 바와 같이 캐스케이드 복합체를 생성하고 정제한다. 캐스케이드는 CasB(또는 CasCDE에서 CasC)에 융합된 N-말단 스트랩-태그 II에 의해 일상적으로 정제된다. 크기 배제 크로마토그래피(수퍼덱스(Superdex) 200 HR 10/30(지이(GE)))를 20 mM 트리스(Tris)-HCl (pH 8.0), 0.1 M NaCl, 1 mM 디티오트레이톨을 사용하여 수행한다. 캐스케이드 제제(대략 0.3 mg)를 2.5 mM MgCl<sub>2</sub>의 존재 하에 37°C에서 15분 동안 DNase I(인비트로젠)과 함께 인큐베이션한 후 크기 배제 분석을 한다. 동시 정제된 핵산들은 동일한 체적의 페놀:클로로포름:이소아밀알코올(25:24:1)(pH 8.0)(플루카(Fluka))를 이용하여 추출함으로써 단리하고, 2.5 mM MgCl<sub>2</sub>가 보충된 DNase I(인비트로젠) 또는 RNase A(퍼멘타스) 중 어느 하나와 함께 37°C에서 10분 동안 인큐베이션한다. 스트랩-태그의 아미노산 서열에 융합된 Cas 서브유닛 단백질이 생성된다.

[0194] 스트랩-태그 캐스케이드 서브유닛의 생물학적 활성을 나타내는 플라크 분석을 박테리오파지 람다를 사용하여 수행하고, 플라크화 효율(EOP)을 문헌[Brouns et al (2008)]에 기술된 바와 같이 계산하였다.

[0195] crRNA의 정제에 있어서, 50 mm x 4.6 mm I. D.의 DNasep 컬럼(미국 캘리포니아주 새너제이의 트랜스게노믹(Transgenomic))을 이용하여 UV<sub>260nm</sub> 검출기를 갖춘 애질런트(Agilent) 1100 HPLC (애질런트)에서 이온쌍 역상 HPLC에 의해 샘플들을 분석한다. 크로마토그래피 분석을 하기 완충제 조건을 이용하여 수행한다: A) 0.1 M 트리에틸암모늄 아세테이트(TEAA)(pH 7.0)(플루카); B) 25% LC MS 등급 아세토니트릴(v/v)(피셔(Fisher))을 포함하는 완충제 A. crRNA는 15% 완충제 B에서 시작하여 12.5분 후에 60% B로 연장되는 선형 구배, 이어서 1.0 ml/min의 유량으로 2분에 걸친 100% B의 선형 연장을 이용하여 75°C에서 온전한 정제 캐스케이드를 주입함으로써 수득한다. 환형 포스페이트 말단의 가수분해는 4°C에서 1시간 동안 0.1 M의 최종 농도의 HCl 중에서 HPLC 정제된 crRNA를 인큐베이션함으로써 수행하였다. 샘플들을 진공 농축기(에펜도르프(Eppendorf))에서 5 내지 10 μl 농축시킨 후 ESI-MS 분석을 한다.

[0196] crRNA의 전기스프레이 이온화 질량 분광 분석법은, 온라인 모세관 액체 크로마토그래피 시스템(얼티밋(Ultimate) 3000, 영국의 디오넥스(Dionex))에 결합된 UHR-TOF 질량 분광계(맥시스(maxis)) 또는 HCT 울트라(Ultra) PTM 디스커버리(Discovery) 기기(둘 모두 브루커 달토닉스(Bruker Daltonics))를 사용하여 네거티브 모드로 수행한다. RNA 분리는 모놀리식(monolithic)(PS-DVB) 모세관 컬럼(200 μm x 50 mm I.D., 영국의 다이오넥스)을 사용하여 수행한다. 크로마토그래피를 하기 완충제 조건을 사용하여 수행한다: C) 트리에틸아민(TEA)로 pH 7.0으로 조정된 0.4 M 1,1,1,3,3,3-헥사플루오로-2-프로판올(HFIP, 시그마-알드리치(Sigma-Aldrich)) 및 0.1 mM TEAA, 및 D) 50% 메탄올(v/v)(피셔)을 포함하는 완충제 C. RNA 분석은 20% 완충제 D, 5분 후에 40% D로의 연장, 이어서 2 μl/min의 유량으로 8분에 걸친 60% D로의 선형 연장을 이용하여 50°C에서 수행한다.

[0197] 캐스케이드 단백질은 5 μM의 단백질 농도에서 0.15 M 아세트산암모늄(pH 8.0)에서 천연 질량 분광법에 의해 분석한다. 단백질 제제는 컷오프(cut-off)가 10 kDa인 원심분리 필터(밀리포어(Millipore))를 사용하여 4°C에서 5회의 연속 농축 및 희석 단계에 의해 수득한다. 단백질을 붕규산염 유리 모세관으로부터 스프레이하고, 고 질량 검출에 있어서의 최적의 성능에 대하여 조정된 LCT 전기스프레이 비행시간형 또는 변형 사중극 비행 시간형 기기[둘 모두 영국의 워터스(Waters)]에서 분석한다(문헌[Tahallah N et al (2001) Rapid Commun Mass

Spectrom **15**: 596-601 (2001)] 및 문헌[van den Heuvel, R.H. et al. Anal Chem **78**: 7473-83 (2006)] 참조). 개개의 Cas 단백질의 정확한 질량 측정은 변성 조건(50% 아세토니트릴, 50% MQ, 0.1% 포름산) 하에서 획득하였다. 용액 중 부분 복합체를, 2-프로판올을 5% (v/v)의 최종 농도로 스프레이 용액에 첨가함으로써 생성하였다. 기기 설정치는 하기와 같았다: 니들 전압: 대략 1.2 kV, 콘(cone) 전압: 대략 175 V, 소스(source) 압력: 9 mbar. 제논을  $1.5 \times 10^{-2}$  mbar의 압력에서의 탠덤(tandem) 질량 분광 분석용 충돌 가스로서 사용하였다. 충돌 전압은 10 V와 200 V 사이에서 변화였다.

[0198] 전기영동 이동성 시프트 분석법(EMSA)을 이용하여 표적 핵산에 대한 캐스케이드 복합체의 기능적 활성을 입증한다. EMSA는 캐스케이드, CasBCDE 또는 CasCDE를 pH 7.5의 50 mM 트리스-Cl, 100 mM NaCl 중 1 nM의 표지된 핵산과 함께 인큐베이션함으로써 수행한다. 연어 정자 DNA(인비트로젠)를 경쟁자로서 사용한다. EMSA 반응물을 37°C에서 20 내지 30분 동안 인큐베이션한 후 5% 폴리아크릴아미드 겔에서 전기영동한다. 겔을 건조시키고, 인광체 보관 스크린 및 PMI 인광체 영상장치(바이오-라드(Bio-Rad))를 사용하여 분석한다. 캐스케이드의 표적 DNA 결합 및 절단 활성을 1-10 mM의 Ca, Mg 또는 Mn 이온의 존재 하에 테스트한다.

[0199] DNA 표적은 겔 정제된 긴 올리고뉴클레오티드(이소젠 라이프 사이언시즈(Isogen Life Sciences) 또는 바이오레지오(Biolegio))이며, 이는 문헌[Jore et al (2011)]의 추가 표 3에 열거되어 있다. 상기 올리고뉴클레오티드를  $\gamma$  <sup>32</sup>P-ATP(퍼킨엘머(PerkinElmer) 및 T4 키나아제(퍼멘타스)를 사용하여 말단 표지한다. 이중 가닥 DNA 표적은 상보성 올리고뉴클레오티드들을 어닐링하고, 남아 있는 ssDNA를 엑소뉴클레아제 I(퍼멘타스)로 절단함으로써 제조한다. 표지된 RNA 표적을,  $\alpha$  <sup>32</sup>P-CTP(퍼킨엘머)를 이용하여 T7 맥시스크립트(Maxiscript) 또는 T7 메가 쇼트스크립트(Mega Shortscript) 키트(암비온(Ambion))를 사용하여 그리고 DNase I(퍼멘타스) 절단에 의해 주형을 제거하여 시험관 내에서 전사시킨다. 이중 가닥 RNA 표적은 상보성 RNA들을 어닐링하고 여분의 ssRNA를 RNase T1(퍼멘타스)으로 절단하고 이어서 페놀 추출함으로써 제조한다.

[0200] 플라스미드 이동성 시프트 분석법은 R44 프로토스페이서를 포함하는 플라스미드 pWUR613을 사용하여 수행한다. 상기 프로토스페이서를 포함하는 단편을 프라이머 BG3297 및 BG 3298을 사용하여 박테리오파지 P7 게놈 DNA로부터 PCR 증폭시킨다(문헌[Jore et al (2011)]의 추가 표 3 참조). 플라스미드(0.4  $\mu$ g) 및 캐스케이드를 5 mM 트리스-HCl (pH 7.5) 및 20 mM NaCl을 함유하는 완충제 중에 1:10의 몰비로 혼합하고, 37°C에서 30분 동안 인큐베이션하였다. 그 후, 캐스케이드 단백질을 프로테이나아제 K 처리(플루카) (0.15 U, 15 min, 37°C), 이어서 페놀/클로로포름 추출에 의해 제거하였다. 그 후, RNA-DNA 복합체를 RNaseH (프로메가(Promega)) (2 U, 1 h, 37°C)로 처리하였다.

[0201] 캐스케이드 단백질 복합체 또는 활성 부분 복합체(RNA 성분을 포함함)(crRNA와 같음)를 형성하는 스트랩-태그-Cas 단백질 서브유닛 융합물은 핵산 표적의 스캐닝과 특이적 부착 및 절단의 예상된 생물학적 및 기능적 활성을 갖는다. 형광 염료의 아미노산 사슬과의 Cas 서브유닛의 융합물은 또한 캐스케이드 복합체 및 부분 복합체(RNA 성분을 포함함)(crRNA와 같음)를 형성하는데, 이는 생물학적 및 기능적 활성을 유지하고 예를 들어 dsDNA에서의 표적 핵산 서열의 위치의 가시화를 허용한다.

[0202] **실시예 7 - 캐스케이드-뉴클레아제 쌍 및 시험관 내에서의 뉴클레아제 활성의 테스트**

[0203] "샤키(Sharkey)"로 표기된 6가지 돌연변이를 랜덤 돌연변이 유발 및 스크리닝에 의해 도입하여 플라보박테리움 오케아노코이트스(*Flavobacterium okeanokoites*) 제한 효소 FokI로부터의 비특이적 뉴클레아제 도메인의 뉴클레아제 활성 및 안정성을 개선시켰다(문헌[Guo, J., et al. (2010) J. Mol. Biol. **400**: 96-107] 참조). 표적에서 멀리 있는 곳(off-target)의 절단 활성을 감소시키는 다른 돌연변이를 도입하였다. 이는 ZFN 쌍의 FokI 이량체 계면에서의 정전기적 상호작용의 엔지니어링에 의해 달성되는데, 이는 양으로 하전된 계면을 갖는 1가지의 FokI 변이체(KKR, E490K, I538K, H537R) 및 음으로 하전된 계면을 갖는 또 다른 것(ELD, Q486E, I499L, N496D)을 생성한다(문헌[Doyon, Y., et al. (2011) Nature Methods **8**: 74-9] 참조). 이들 변이체 각각은 호모이량체로서 촉매적으로 불활성이며, 그것 때문에 표적에서 멀리 있는 곳의 절단의 빈도를 감소시킨다.

[0204] **캐스케이드-뉴클레아제의 설계**

[0205] 본 발명자는 개선된 FokI 뉴클레아제를 Cse1의 N-말단에 번역에 의해 융합시켜서 Cse1의 변이체를 생성하였는데, 이는 각각 FokI<sup>KKR</sup>-Cse1 및 FokI<sup>ELD</sup>-Cse1이었다. 이들 두 변이체는 캐스케이드 서브유닛들(Cse2, Cas7, Cas5 및 Cas6e), 및 균일한 스페이서를 갖는 2가지의 독특한 CRISPR 중 하나와 함께 동시 발현시킨다. 이는 캐스케이드<sup>KKR</sup> 복합체에 균일한 P7-crRNA를 로딩하고, 캐스케이드<sup>ELD</sup> 복합체에 균일한 M13 g8-crRNA를

로딩한다. 이들 복합체를 N-말단 스트랩II-태그 Cse2를 사용하여 정제하며, 이는 문헌[Jore, M.M., et al., (2011) Nat. Struct. Mol. Biol. 18(5): 529-536]에 기술된 바와 같다. 더욱이, 추가의 정제 단계를 N-말단 HIS-태그 FokI을 사용하여 실시하여, 전장의 온전한 캐스케이드-뉴클레아제 융합 복합체를 정제하는 것을 보장할 수 있다.

[0206] 이 실시예에서 사용한 융합 단백질의 뉴클레오티드 및 아미노산 서열은 하기와 같았다:

[0207] >FokI-(샤키-ELD)-Cse1의 뉴클레오티드 서열

ATGGCTCAACTGGTTAAAAGCGAACTGGAAGAGAAAAAAGTGAAGTGCGCCAC  
AAACTGAAATATGTGCCGCATGAATATATCGAGCTGATTGAAATTGCACGTAATC  
CGACCCAGGATCGTATTCTGGAAATGAAAGTGATGGAATTTTTTATGAAAGTGTA

[0208] CGGCTATCGCGGTGAACATCTGGGTGGTAGCCGTAACCGGATGGTGCAATTTAT  
ACCGTTGGTAGCCCGATTGATTATGGTGTATTGTTGATACCAAAGCCTATAGCG  
GTGGTTATAATCTGCCGATTGGTCAGGCAGATGAAATGGAACGTTATGTGGAAG  
AAAATCAGACCCGTGATAAACATCTGAATCCGAATGAATGGTGAAAGTTTATC  
CGAGCAGCGTTACCGAGTTTAAATTCCTGTTTGTAGCGGTCAC TTCAAAGGCAA  
CTATAAAGCACAGCTGACCCGTCTGAATCATATTACCAATTGTAATGGTGCAGTT  
CTGAGCGTTGAAGAACTGCTGATTGGTGGTGAATGATTAAGCAGGCACCCTG  
ACCCTGGAAGAAGTTCGTGCAAAATTTAACAAATGGCGAAATCAACTTTGCGGAT  
CCCACCAACCGCGCGAAAGGCCTGGAAGCGGTGAGCGTGGCGAGC

[0209] atgaatttct  
tattgataactggattcctgtacgcccgcgaacgggggaaagtccaaatcataaactgcaatcgctatactgcagtagagatcagt  
ggcgattaagttgccccgtgacgataggaactggccgctttagcactgctggtttgcaattatcgccccggcaaaagatg  
acgttgaattcgcacatcgcaataatgaatccgctcactgaagatgagttcaacaactcatcgccgctggatagatgttctaccttaat  
cacgcagaacatcccttatgcagacaaaggtgtcaaaagcaaatgatgtactccaatgaaaaactgttgctgggtaagcggcg  
cgacgaattgtgcaattgtcaatcaaccggggcaggggtgaagcattatgtggtggatgcaactgctgattgcttattcaaccaggcgaat  
caggcaccaggtttggtggtggttttaaaagcgtttacgtggaggaacacctgtaacaacgttcgtacgtggatgatcttctgtaaa  
cggtgtactcaatgtctcacattacctgtctcaaaaacaattcctaataatcacatacggaaaaccaacctacctggattaacct  
atcaagtcfaatgagctatactgctctcaatgggtttgtccgtggtctattctggcaaccagcgcatattgaattatgcatccatt  
gggattgtaaatgttctgtgtgacaggaagcaatttgcgttataccggtttcttaagaaaaattacctttacagttaatgggctat  
ggccccatccgcatcccttctgtgtaacagtcgaagaaaggggaggtgaggaaaaattctgtcttaccacctccgcacatcat  
ggacacaaatcagccgagttgtggtagataagattatcaaaatgaaaatggaatcgctggcgcggttgaatcaattcagaaat  
attgcccgcaaaagtccttgaattgattgggggatatcgaataatcaagcatctattctgacggcgctcatgatgttggatgtt  
aatcaggggtggcaacaatcggcaatgtgataaacgaaatagtgactgttgggttggatataaaacagccttacgcaaggcgttata  
taccttgcagaaggttttaaaataaagactcaaaagggccggagtcctctgtcatgagactgcagaaaggcafttctatcgacagag  
tgaattattaattcccgatgactggcgaatgtaattttccagcgtgatgagtaataagctgatttacgagacaaactcatcaattgtg  
gaaatgctatttaataatctgtagctccctatgacatcatcctaaatgaatgaagcaccattagcgttccccgcgccagctatacaaca  
tttacgggagttaaaaccgaaggggccatcaaatggctga **[서열 번호 18]**

[0211] >FokI-(샤키-ELD)-Cse1의 단백질 서열

MAQLVKSELEEKSELRHKLKYPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGY  
RGEHLGGSRPDGAITYVGVSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVEENQTR  
DKHLNPNEWVKVYSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEEL  
LIGGEMIKAGTLTLEEVRKFNNGEINFADPTNRAKGLEAVSVASMNLLIDNWIPVRP

[0212]

RNGGKVQIINLQSLYCSRQWRLSLPRDDMELAAALALLVCIGQIIAPAKDDVEFRHRI  
MNPLTEDEFQQLIAPWIDMFYLNHAEHPFMQTKGVKANDVTPMEKLLAGVSGATN  
CAFVNQPGQGEALCGGCTAIALFNQANQAPGFGGGFKSGLRGGTPVTTFVRGIDLRS  
TVLLNVLTLPRLQKQFPNESHTEENQPTWIKPIKSNESIPASSIGFVRGLFWQPAHIELC  
DPIGIGKSCCCGQESNLRYTGFLKEKFTFTVNGLWPHPHSPCLVTVKKGEVEEKFLAF  
TTSAPSWTQISR VVVDKIIQENENGNRVA AVVNQFRNIAPQSLELIMGGYRNNQASIL  
ERRHDVLMFNQGWQYGNVINEIVTVGLGYKTALRKALYTF AEGFKNKDFKAGV  
SVHETAERHFYRQSELLIPDVLANVNFSADEVIADLRDKLHQLCEMLFNQSVAPYA  
HHPKLISLALARATLYKHLRELKPQGGPSNG\* [서열 번호 19]

[0213]

[0214]

>FokI-(샤키-KKR)-Cse1의 뉴클레오티드 서열

ATGGCTCAACTGGTTAAAAGCGAACTGGAAGAGAAAAAAGTGAAGTGCGCCAC  
AAACTGAAATATGTGCCGCATGAATATATCGAGCTGATTGAAATGCACGTAATC  
CGACCCAGGATCGTATTCTGGAAATGAAAGTGATGGAATTTTTTATGAAAGTGTA  
CGGCTATCGCGGTGAACATCTGGGTGGTAGCCGTAACCCGGATGGTGCAATTTAT  
ACCGTTGGTAGCCCCGATTGATTATGGTGTTATTGTTGATACCAAAGCCTATAGCG  
GTGGTTATAATCTGCCGATTGGTCAGGCAGATGAAATGCAGCGTTATGTGAAAG  
AAAATCAGACCCGCAACAACATATTAACCCGAATGAATGGTGGAAAGTTTATC  
CGAGCAGCGTTACCGAGTTTAAATTCCTGTTTGTAGCGGTCACCTCAAAGGCAA  
CTATAAAGCACAGCTGACCCGTCTGAATCGTAAAACCAATTGTAATGGTGCAGTT  
CTGAGCGTTGAAGAACTGCTGATTGGTGGTGAATGATTAAGCAGGCACCCTG  
ACCCTGGAAGAAGTTCGTCGCAAAATTAACAATGGCGAAATCAACTTTGCGGAT  
CCCACCAACCGCGCGAAAGGCCTGGAAGCGGTGAGCGTGGCGAGC

[0215]

Catgaatttgcctattgataactggattctgtacgcccggaacgggggaaagtccaatcataaatctgcaatcgtatactgcagtagagatcagt  
ggcgattaagtttgcctcgtacgatafggaactggccgcttagcactgctggtttgcattgggcaattatgccccggcaaaagatg  
acgttgaattcgcacatgcaataatgaatccgctcactgaagatgagttcaacaactcatcgccgctggatagatatgttctaccttaat  
cacgcagaacatcccttatgcagacaaaagggtc aaagcaaatgatgtactccaatgaaaaactgttgctgggtaagcggcg  
cgacgaatgtgcaattgtcaatcaaccggggcagggtgaagcattatgtggtggatgcaactgcgattgcgttattcaaccaggcgaat  
caggcaccaggtttgggtggttttaaaagggtttacgtggaggaacacgtgtaacaacgttcgtacgtgggatcgtcttctca  
cgggttactcaatgtcctcacattacctgtcttcaaaaacaattcctaataatgaatcacatacgaaaaccaacctacctgattaaacct  
atcaagtcgaatgagctataacctgtctcaatgggtttgctcgtctattctggcaaccagcgcatattgaattatgcgatccatt  
gggatggtaaatgttctgtgtgacaggaagcaatttgcgtataccggtttctaaaggaaaaattacctttacagttaatgggctat

[0216]

ggccccatccgcatcccttctgtgtaacagtaagaaaggagggtgaggaataattctgtcttaccacctccgaccatcat  
ggacacaaatcagccgagttgtgtagataagattattcaaatgaaatgaaatcgctggcggcgggtgtgaatcaatcagaaat  
attgcgccgaaagtctcttgaattgattfaggggatcgtataatcaagcatctattcttgaacggcgtcatgatgtttgatgttt  
aatcaggggtgcaacaatcggcaatgtgataaacgaaatagtactgttggttgggataaaacagccttacgcaaggcgttata  
tacctttgcagaagggttataaaataaagacttcaaaaggccggagtcctgttcatgagactgcagaaaggcatttctatcgacagag  
tgaattattaattcccagctacggcaatgtaattttccaggctgatgaggaatagctgatttacgagacaaactcatcaattgtgt  
gaaatgctatttaataatctgtagctccctatgcacatcatctaaattaataagcacattagcgttggccgcccacgctatacaaca  
tttacgggagttaaaaccgcaaggaggccatcaaatggctga [서열 번호 20]

[0217]

[0218] >FokI-(샤키-KKR)-Cse1의 단백질 서열

MAQLVKSELEEKSELRHKLKYPHEYIELIEIARNPTQDRILEMKVMEFFMKVYGY  
 RGEHLGGSRKPDAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQT  
 RNKHINPNEWWKVYPSSVTEFKFLFVSGHFKNYKAQLTRLNRKTNCGAVLSVEE  
 LLIGGEMIKAGTLLEEVRRKFNNGEINFADPTNRAKGLEAVSVASMNLLIDNWIPV  
 RPRNGGKVQIINLQSLYCSRQWRLSLPRDDMELAAALALLVCIGQIIAPAKDDVEFR  
 HRIMNPLTEDEFQQLIAPWIDMFYLNHAEHPMQTKGVKANDVTPMEKLLAGVSGA  
 TNCAFVNQPGQGEALCGGCTAIALFNQANQAPGFGGFKSGLRGGTPVTTFVRGIDL  
 RSTVLLNVLTLPRQLKQFPNESHTENQPTWIKPIKSNEIPASSIGFVRGLFWQPAHIEL  
 CDPIGIGKSCCGQESNLRYTGFLKEKFTFTVNGLWPHPHSPCLVTVKKGEVEEKFL  
 AFTTSAPSWTQISR VVVDKIIQENGNRVAAVVNQFRNIAPQSPLELIMGGYRNNQA  
 SILERRHDVLMFNQGWQQYGNVINEIVTVGLGYKTALRKALYTF AEGFKNKDFKGA  
 GVSVHETAERHFYRQSELLIPDVLANVNFSADEVIADLRDKLHLQCEMLFNQSVAP  
 YAHPKLISTLALARATLYKHLRELKPGGPNNG\* [서열 번호 21]

[0219]

[0220] >His6-이중-1부분(monopartite) NLS SV40-FokI-(샤키-KKR)-Cse1의 뉴클레오티드 서열

ATGcatcaccatcatcaccacCCGAAAAAAAAAGCGCAAAGTGGATCCGAAGAAAAACGTAAAG  
 TTGAAGATCCGAAAGACATGGCTCAACTGGTTAAAAGCGAACTGGAAGAGAAAA  
 AAAGTGAAGTGCGCCACAACTGAAATATGTGCCGCATGAATATATCGAGCTGA  
 TTGAAATTGCACGTAATCCGACCCAGGATCGTATTCTGGAAATGAAAGTGATGG  
 AATTTTTTATGAAAGTGTACGGCTATCGCGGTGAACATCTGGGTGGTAGCCGTAA  
 ACCGGATGGTGCAATTTATAACCGTTGGTAGCCCGATTGATTATGGTGTTATTGTT

[0221]

GATACCAAAGCCTATAGCGGTGGTTATAATCTGCCGATTGGTCAGGCAGATGAA  
 ATGCAGCGTTATGTGAAAGAAAATCAGACCCGCAACAAACATATTAACCCGAAT  
 GAATGGTGGAAAGTTTATCCGAGCAGCGTTACCGAGTTTAAATTCCTGTTTGTTA  
 GCGGTCACTTCAAAGGCAACTATAAAGCACAGCTGACCCGTCTGAATCGTAAAA  
 CCAATTGTAATGGTGCAGTTCTGAGCGTTGAAGAAGTCTGATTGGTGGTGAAT  
 GATTAAGCAGGCACCCTGACCCTGGAAGAAGTTCGTCGCAAATTTAACAATGG  
 CGAAATCAACTTTGCGGATCCCACCAACCGCGCGAAAGGCTGGAAGCGGTG  
 AGCGTGGCGAGC

[0222]

atgaatttgcttattgataactggattctgtacgccccgaaacgggggaaagtccaatcataaat

ctgcaatcgctatactgcagtagagatcagtgggcgaattaagttgccccgtgacgatatggaactggccgcttagcactgctggttgc  
atgggcaaatfatcgccccggcaaaagatgacgftgaattcgacatcgcataatgaatccgctcactgaagatgagttcaacaactc  
atcgccgctggatagatagtctacctaatacagcagaacatcccctttatgcagacaaaagggtgcaagcaaatgatgtgactcca  
atggaaaaactgftggctgggtaagcggcgcgacgaattgfcalttgcaatcaaccggggcagggtgaaacattatggtggat  
gcactgcgattgctgattcaaccaggcgaatcaggcaccagggtttggtggtggtttaaaagggttacgtggaggaaacacctgtaa  
caacgttcgtacgftggatcgtcttcgtcaacgggtgtactcaatgctcctacattacctcgtctcaaaaacaattcctaataatcac  
atacggaaaaccaactacctggaftaacctcaagccaatgagctatactgctcgaattgggttgcggtgctattctgg  
caaccagcgcataattgaattatgcatccattgggattggtaaatgtctgtgacagaaagcaattgctgtataccggtttctt  
aaggaaaaattacctttacagtaattgggctatggccccatccgctccctgtgctgtaacagcaagaaaggagggtgaggaa  
aaattctgtcttaccacctccgaccatcatggacacaaatcagccgagttggtgtagataagattatcaaaatgaaatggaaatc  
gcgtggcgcggtgtaatacagaataatgcccgaagtcctctgaattgattatgggggatatcgaataataagcat  
ctattctgaacggcgctgatgtgtgatttaacagggggtggcaacaatacggcaatgtgataaacgaaatagtgactgtggtttg  
ggatataaacagccttacgcaaggcgttatataccttgcagaagggttataaaataaagactcaagggggcggagctctctgtcat  
gagactgcagaaaggcattctatcgacagagtgtaattattaattcccgatgactggcgaatgtaattttccaggctgatgaggtaat  
agctgatttacgagacaaactcatcaattgtgtgaaatgctatttaatactgtagctccctatgcacatcactaaattaataagcaca  
ttagcgttgccccgccacgctatacaaacatttacgggagttaaaaccgcaaggaggccatcaaatggctga [서열 번호 22]

[0223]

[0224]

>His6-이중-1부분 NLS SV40-FokI-(샤키-KKR)-Cse1의 단백질 서열

MHHHHHHPKKKRKVDPKKKRKVEDPKDMAQLVKSELEEKSELRHKLKYVPHEYI  
ELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPdGAIYTVGSPIDYGVIVD  
TKAYSGGYNLPIGQADEMQRVYKVENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGH  
FKGNYKAQLTRLNRKTNCGAVLSVEELLIGGEMIKAGTLTLEEVRKFNNGEINFA  
DPTNRAKGLEAVSVASMNLLIDNWIPVRPRNGGKVQIINLQSLYCSRdQWRLSLPRD

[0225]

DMELAALALLVCIGQIIAPAKDDVEFRHRIMNPLTEDEFQQLIAPWIDMFYLNHAEHP  
FMQTKGVKANDVTPMEKLLAGVSGATNCAFVNQPGQGEALCGGCTAIALFNQANQ  
APGFGGGFKSGLRGGTPVTFVRGIDLRSTVLLNVLTLPRLQKQFPNESHTENQPTWI  
KPIKSNESIPASSIGFVRGLFWQPAHIELCDPIGIGKSCCGQESNLYTGFLKEKFTFT  
VNGLWPHPHSPCLVTYKKGEVEEKFLAFTTSAPSWTQISRVVVDKIIQNENGNRVAA  
VVNQFRNIAPQSPLEIMGGYRNNQASILERRHDVLMFNQGWQQYGNVINEIVTVGL  
GYKTALRKALYFAEGFKNKDFKGAGVSVHETAERHFYRQSELLIPDVLANVNFsq  
ADEVIADLRDKLHQLCEMLFNQSVAPYAHHPKLISTLALARATLYKHLRELKPQGGP

[0226]

SNG\* [서열 번호 23]

[0227] >His6-이중-1부분 NLS SV40-FokI(샤키)-ELD)-Cse1의 뉴클레오티드 서열

ATGcatcaccatcatcaccacCCGAAAAAAAAAGCGCAAAGTGGATCCGAAAGAAAAACGTAAAG  
TTGAAGATCCGAAAGACATGGCTCAACTGGTTAAAAGCGAACTGGAAGAGAAAA  
AAGTGAAGTGCGCCACAACTGAAATATGTGCCGCATGAATATATCGAGCTGAT  
TGAAATTGCACGTAATCCGACCCAGGATCGTATTCTGGAAATGAAAGTGATGGA  
ATTTTTATGAAAGTGTACGGCTATCGCGGTGAACATCTGGGTGGTAGCCGTAAA  
CCGGATGGTGCAATTTATACCGTTGGTAGCCCCGATTGATTATGGTGTATTGTTG  
ATACCAAAGCCTATAGCGGTGGTTATAATCTGCCGATTGGTCAGGCAGATGAAA  
TGGAACGTTATGTGGAAGAAAAATCAGACCCGTGATAAACATCTGAATCCGAATG  
AATGGTGGAAAGTTTATCCGAGCAGCGTTACCGAGTTTAAATTCCTGTTTGTAG  
CGGTCACTTCAAAGGCAACTATAAAGCACAGCTGACCCGTCTGAATCATATTACC  
AATTGTAATGGTGCAGTCTGAGCGTTGAAGAAGTCTGATTGGTGGTGAAATGA  
TTAAAGCAGGCACCCTGACCCTGGAAGAAGTTCGTTCGCAAATTTAACAATGGCG  
AAATCAACTTTGCGGATCCCACCAACCGCGCGAAAGGCCTGGAAGCGGTGAG  
**CGTGGCGAGC**

[0228] Catgaatttgcttattgataactggattcctgtacgcccgcgaaacgggggaaagtccaatcataaactg  
caatcgtatactgcagtagagatcagtgcgatfaagttgccccgtgacgataggaaactggccctttagcactgctgtttgcattg  
ggcaaattatgccccggcaaaagatgacgttgaatttcgacatcgcataatgaatccgctcactgaagatgagttcaacaactcatg  
cgccgtggatagatgttctaccttaatcagcagacaatcctttatgcagaccaaaagggtgcaaaagcaaatgatgtgactccaatgg  
aaaaactgttgctgggtaagcggcgacgcaatgtgcattgtcaatcaaccggggcagggtgaagcattatgtgtggatgcac  
tgcgattgcttattcaaccaggcgaatcaggcaccagggtttgggtgggtttfaaaagcgggttacgtggaggaaacacctgaaacaac  
gttcgtacgtgggatcgtactcgttcaacgggttactcaatgtctccacattacctgcttcaaaaaaatttctaataatcacatacag  
gaaaaccaacctacctggaffaaacctatcaagtccaatgagctatacctgcttcgcaattgggtttgcccgtgtctattctggcaacc

[0229] agcgcataattgaattatgcgatcccattgggattgtaaatgttctgctgtggacaggaaagcaattgcttataccggtttcttaagga  
aaaattacccttacagttaatgggctatgcccaccatccgattcccctgtctgtaaacagtaagaaaggagggttgaggaaaattt  
cttcttaccacctccgaccatcatggacacaaatcagccgagttgtgtagataagattatcaaaatgaaatggaaatcgcgtg  
gcgcggtgtgaatcaattcagaaatattgcccgcgaaagtctctgaattgattgggggatatcgaataatcaagcatctattct  
tgaacggcgtcatgatgttgaatcaggggtgcaacaatacggcaatgtgataaacgaaatagtactgttgggttgggatat  
aaaaacagccttacgcaagggttatataccttgcagaagggttaaaaataaagactcaaaaggccggagctctgttcatgagact  
gcagaaaggcatttctatcgacagatgaattataattcccgatgtactggcgaatgttaattttccaggctgatgaggtaatagctga  
ttacgagacaaactcatcaatgtgtgaaatgctatftaataatctgtagctccctatgcacatcatcctaaatfaataagcacattagcg  
cttccccgcgccacgtatacaaacatttacgggagttaaaaccgcaaggaggccatcaatggctga [서열 번호 24]

[0230]

[0231] >His6-이중-1부분 NLS SV40-FokI-(샤키-ELD)-Cse1의 단백질 서열

MHHHHHHPKRRKVDPKRRKVEDPKDMAQLVKSELEEKSELRHKLKYVPHEYI  
 ELIEIARNPTQDRILEMKVMEFFMKVYGYRGEHLGGSRKPDGAIYTVGSPIDYGVIVD  
 TKAYSGGYNLPYGQADEMERYVEENQTRDKHLNPNEWKVVYSSVTEFKFLVSGH  
 FKGNYKAQLTRLNHNITCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFA  
 DPTNRAKGLEAVSVASMNLLIDNWPVPRNGGKVQIINLQSLYCSRQWRLSLPRD  
 DMELAALALLVCIGQIIAPAKDDVEFRHRIMNPLTEDEFQQLIAPWIDMFYLNHAEHP  
 FMQTKGVKANDVTPMEKLLAGVSGATNCAFVNQPGQGEALCGGCTAIALFNQANQ  
 APGFGGGFKSGLRGGTPVTTFVRGIDLRSTVLLNVLTLPRQLKQFPNESHTEENQPTWI  
 KPIKSNESIPASSIGFVRGLFWQPAHIELCDPIGIGKCSGCGQESNLRVTGFLKEKFTFT  
 VNGLWPHPHSPCLVTVKKGEVEEKFLAFTTSAPSWTQISRVVVDKIIQNENGNRVAA  
 VVNQFRNIAPQSPLEIMGGYRNNQASILERRHDVLMFNQGWQQYGNVINEIVTVGL  
 GYKTALRKALYFAEGFKNKDFKGAGVSVHETAERHFYRQSELLIPDVLANVNSQ  
 ADEVIADLRDKLHQLCEMLFNQSVAPYAHHPKLISTLALARATLYKHLRELKPQGGP  
 SNG\* [서열 번호 25]

[0232]

[0233] DNA 절단 분석

[0234] 복합체들의 특이성 및 활성을 기질로서 인공적으로 구성된 표적 플라스미드를 사용하여 테스트하였다. 이 플라스미드는 반대쪽 가닥들 상에 M13 및 P7 결합 부위를 포함하여서 둘 모두의 FokI 도메인이 서로와 대면하게 된다(도 11 참조). 캐스케이드 결합 부위들 사이의 거리는 25 염기쌍과 50 염기쌍 사이에서 변하며, 이때 증분은 5 bp이다. 캐스케이드의 결합 부위들은 임의의 4가지의 공지된 PAM 서열들(5'-프로토스페이서-CTT/CAT/CTC/CCT-3') 중 임의의 것이 측면에 위치할 필요가 있기 때문에, 이 거리 범위는 거의 모든 주어진 서열에 있어서 이러한 쌍을 설계하기에 충분한 융통성을 제공한다.

[0235] 사용한 표적 플라스미드의 서열은 하기와 같다. 숫자는 M13 표적 부위와 P7 표적 부위 사이의 거리를 나타낸다. 프로토스페이서는 볼드체로 나타내며, PAM은 밑줄이 그어져 있다:

[0236] 표적 플라스미드의 서열. 수는 M13 표적 부위와 P7 표적 부위 사이의 거리를 나타낸다. (프로토스페이서: 볼드체, PAM: 밑줄이 그어져 있음)

[0237] >50 bp

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG  
**TCTTGCTTTTCGAGCGCTAGCTCTAGAACTAGTCCTCAGCCTAGGCCTCGTTCGGA**  
**AGCTGTCTTTCGCTGCTGAGGGTGACGATCCCGCATAGGCGGCCTTTAACTCg**  
 gatcc [서열 번호 26]

[0238]

[0239] >45 bp

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG  
**TCTTTTTCGAGCGCTAGCTCTAGAACTAGTCCTCAGCCTAGGCCTCGTTCAAGCTG**  
**TCTTTCGCTGCTGAGGGTGACGATCCCGCATAGGCGGCCTTTAACTCg**gatcc  
 [서열 번호 27]

[0240]

[0241] >40 bp

gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG  
**TCTTCGAGCGCTAGCTCTAGAACTAGTCCTCAGCCTAGGCCTCGAAGCTGTCTTT**  
**CGCTGCTGAGGGTGACGATCCCGCATAGGCGGCCTTTAACTCg**gatcc [서열 번호 28]

[0242]

[0243] >35 bp  
 gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG  
 TCTTGCCTAGCTCTAGAAGTCTAGCCTAGGCCTAAGCTGTCTTTTCGCT  
 GCTGAGGGTGACGATCCCGCATAGGCGGCCTTTAACTCgcatcc [서열 번호 29]

[0244] >30 bp  
 gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG  
 TCTTGCCTAGCTCTAGAAGTCTAGCCTAGGAAGCTGTCTTTTCGCTGCTGA  
 GGGTGACGATCCCGCATAGGCGGCCTTTAACTCgcatcc [서열 번호 30]

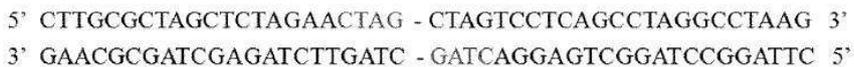
[0247] >25 bp  
 gaattcACAACGGTGAGCAAGTCACTGTTGGCAAGCCAGGATCTGAACAATACCG  
 TCTTCTCTAGAAGTCTAGCCTAGCCTAGGAAGCTGTCTTTTCGCTGCTGAGGGTG  
 ACGATCCCGCATAGGCGGCCTTTAACTCgcatcc [서열 번호 31]

[0248] 표적 플라스미드의 절단을 아가로스 겔에서 분석하였으며, 여기서, 네거티브 수퍼코일형(nSC) 플라스미드는 선형화된 또는 니킹된 플라스미드와 구별될 수 있다. 표적 벡터 중 캐스케이드<sup>KKR/ELD</sup> 쌍의 절단 부위는 아가로스 겔로부터 선형 절단 생성물을 분리하고, 이. 콜라이 DNA 폴리머라아제의 클레나우(Klenow) 단편을 이용한 FokI 절단에 의해 남겨진 리세스(recessed) 3' 말단을 필인(filling in)하여 블런트 말단을 생성함으로써 결정하였다. 선형 벡터를 자가 라이게이션시키고, 형질전환시키고, 증폭시키고, 분리하고, 서열결정하였다. 리세스 3' 말단의 필인 및 재라이게이션은 FokI 절단에 의해 남겨진 오버행(overhang)을 나타내는 서열에 있어서의 가외의 뉴클레오티드에 이르게 된다. 이 서열 관독을 원래의 서열에 대하여 정렬함으로써, 절단 부위들을 클론 수준에서 찾아내어 지도화할 수 있다. 하기에서, FokI 절단에 의해 남겨진 리세스 3' 말단의 필인 후 당해 서열 내에 혼입된 추가의 염기는 밑줄이 그어져 있다:

**FokI 절단**



**3' 필인, 라이게이션**



[0250] 상부로부터 하부로 관독하면, 상기 5'-3' 서열은 각각 서열 번호 32 내지 35이다.  
 [0251]

**[0252] 인간 세포에서의 표적 유전자좌의 절단**

[0253] 인간 CCR5 유전자는 C-C 케모카인 수용체 타입 5 단백질을 코딩하는데, 이는 백혈구 세포의 표면 상의 인간 면역결핍 바이러스(HIV)의 수용체로서의 역할을 한다. 인공 GFP 유전자좌에 더하여 CCR5 유전자를 쌍형의 캐스케이드<sup>KKR/ELD</sup> 뉴클레아제를 사용하여 표적화한다. 적합한 결합 부위 쌍을 CCR5의 코딩 영역 상에서 선택한다. 결합 부위들 각각을 표적화하는 균일한 스페이서를 포함하는 2개의 별도의 CRISPR 어레이를 DNA 합성을 통하여 구성한다(진아트).

[0254] 사용한 CRISPR 설계 및 인간 CCR5 표적 유전자 선택은 하기와 같다:

[0255] >전 ORF(위치 347-1446)를 포함하는 계놈 인간 CCR5 서열의 일부.

GGTGAACAAGATGGATTATCAAGTGTCAAGTCCAATCTATGACATCAATTATTA  
TACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCAAATCGCAGCCCGCTCCT  
GCCTCCGCTCTACTCACTGGTGTTCATCTTTGGTTTTGTGGGCAACATGCTGGTCA  
TCCTCATCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCT  
CAACCTGGCCATCTCTGACCTGTTTTTCTTCTTACTGTCCCCTTCTGGGCTCACT  
ATGCTGCCGCCAGTGGGACTTTGGAAATACAATGTGTCAACTCTTGACAGGGCT  
CTATTTTATAGGCTTCTTCTCTGGAATCTTCTTCATCATCCTCCTGACAATCGATA  
GGTACCTGGCTGTGCTCCATGCTGTGTTGCTTTAAAAGCCAGGACGGTCACCTT  
TGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCCTCTCTCCA

[0256] GGAATCATCTTTACCAGATCTCAAAAAGAAGGTCTTCATTACACCTGCAGCTCTC  
ATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAGATAGT  
CATCTTGGGGCTGGTCTGCGCTGCTTGTTCATGGTCATCTGCTACTCGGGAATC  
CTAAAACTCTGCTTCGGTGTGAAATGAGAAGAAGAGGCACAGGGCTGTGAGG  
CTTATCTTACCATCATGATTGTTTATTTTCTTCTTGGGCTCCCTACAACATTGTC  
CTTCTCTGAACACCTTCCAGGAATTTTGGCCTGAATAATGTCAGTAGCTCTA  
ACAGGTTGGACCAAGCTATGCAGGTGACAGAGACTCTGGGATGACGCACTGCT  
GCATCAACCCCATCATCTATGCCTTTGTCGGGGAGAAGTTCAGAACTACCTCTT  
AGTCTTCTTCCAAAAGCACATTGCCAAACGCTTCTGCAAATGCTGTTCTATTTTCC  
AGCAAGAGGCTCCCAGCGAGCAAGCTCAGTTTACACCCGATCCACTGGGGAGC  
AGGAAATATCTGTGGGCTTGTGACACGGACTCAAGTGGGCTGGTGACCCAGTC

[서열 번호 36]

[0257]

[0258] Red1/2: 선택된 표적 부위(거리: 34 bp, PAM 5'-CTT-3'). "Red 1"은 상기에서 첫 번째로 나타나는 밑줄이 그 어진 서열이다. Red2는 두 번째의 밑줄이 그어진 서열이다.

[0259] >CRISPR 어레이 red1 (이탤릭체 = 스페이서, 볼드체 = 반복체)

ccatggTAATACGACTCACTATAGGGAGAATTAGCTGATCTTTAATAATAAGGAAAT  
GTTACATTAAGGTTGGTGGGTTGTTTTATGGGAAAAAATGCTTTAAGAACAAT  
GTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAAACCGCAAACACAGCA  
TGGACGACAGCCAGGTACCTAGAGTTCCCCGCGCCAGCGGGGATAAAACCGCAA  
CACAGCATGGACGACAGCCAGGTACCTAGAGTTCCCCGCGCCAGCGGGGATAAA  
CCGCAAACACAGCATGGACGACAGCCAGGTACCTAGAGTTCCCCGCGCCAGCGG  
GGATAAAACCGAAAACAAAAGGCTCAGTCGGAAGACTGGGCCTTTTGTTTAACC  
CCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGggtacc [서열 번호 37]

[0260]

[0261] >CRISPR 어레이 red2 (이탤릭체: 스페이서, 볼드체: 반복체)

ccatggTAATACGACTCACTATAGGGAGAATTAGCTGATCTTTAATAATAAGGAAAT  
GTTACATTAAGGTTGGTGGGTTGTTTTATGGGAAAAAATGCTTTAAGAACAAT  
GTATACTTTTAGAGAGTTCCCCGCGCCAGCGGGGATAAAACCGTGTGATCACTTG  
GGTGGTGGCTGTGTTTGCCTGAGTTCCCCGCGCCAGCGGGGATAAAACCGTGTGA  
TCACTTGGGTGGTGGCTGTGTTTGCCTGAGTTCCCCGCGCCAGCGGGGATAAAC  
CGTGTGATCACTTGGGTGGTGGCTGTGTTTGCCTGAGTTCCCCGCGCCAGCGGGG  
ATAAACCGAAAACAAAAGGCTCAGTCGGAAGACTGGGCCTTTTGTTTAACCCC  
TTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGggtacc [서열 번호 38]

[0262]

[0263] **인간 세포의 핵 내로의 Cascade<sup>KKR/ELD</sup>의 전달**

[0264] 캐스케이드는 다중서브유닛 단백질-RNA 복합체로서 매우 안정하며, 이. 콜라이에서 mg의 양으로 용이하게 생성된다. 이. 콜라이로부터 정제된 온전한 형태의 상기 복합체의 형질감염 또는 미세주입이 전달법으로서 사용된다(도 12 참조). 도 12에 나타난 바와 같이, 캐스케이드-FokI 뉴클레아제를 이. 콜라이로부터 정제하며, 단백질 형질감염 소포(vesicle) 중에 봉지한다. 그 후, 이들은 인간 HepG2 세포의 세포막과 융합되어 뉴클레아제를 세포질 내에 방출한다(단계 2). 그 후, NLS 서열들은 임포틴(importin) 단백질에 의해 인식되는데, 이는 핵공 통과를 용이하게 한다(단계 3). 그 후, 캐스케이드<sup>KKR</sup> (채워지지 않은(open) 직사각형) 및 캐스케이드<sup>ELD</sup> (채워진 직사각형)는 그의 표적 부위를 찾아내서 절단하며(단계 4), 이는 DNA 복구 경로를 유도하는데, 상기 경로는 표적 부위를 변경시켜서 소정의 변화에 이르게 된다. 캐스케이드<sup>KKR/ELD</sup> 뉴클레아제는 단지 1회 작용할 필요가 있으며, DNA 상에 코딩되어서 세포 중에 영구적으로 존재하는 것을 요구하지 않는다.

[0265] 캐스케이드를 인간 세포 내로 전달하기 위하여, 피어스(Pierce), 엔이비(NEB), 퍼멘타스 및 클론테크(Clontech)를 포함하는 다양한 공급처로부터의 단백질 형질감염 시약을 사용한다. 이들 시약은 최근에 항체 전달용으로 개발되었으며, 넓은 범위의 인간 세포주의 형질감염에 유용하고, 이때 효율은 90% 이하이다. 인간 HepG2 세포를 형질감염시킨다. 또한, CHO-K1, COS-7, HeLa, 및 비배아 줄기 세포를 포함하는 다른 세포주를 형질감염시킨다.

[0266] 캐스케이드<sup>KKR/ELD</sup> 뉴클레아제 쌍을 핵 내로 임포트(import)하기 위하여, 시미안 바이러스 40(SV 40)의 큰 T-항원으로부터의 텐덤 1부분 핵 국제화 신호(NLS)를 FokI의 N-말단에 융합시킨다. 이는 단지 온전한 캐스케이드<sup>ELD/KKR</sup>의 핵 내로의 임포트를 보장한다(핵공 복합체는 RNA 폴리머라아제(550 kDa) 및 다른 큰 단백질 복합체를 전좌시킨다). 형질전환 이전의 점검으로서, 캐스케이드<sup>KKR/ELD</sup> 뉴클레아제 쌍의 뉴클레아제 활성을 정제된 복합체 및 CCR5 PCR 앰플리콘을 사용하여 시험관 내에서 점검하여 비생산성 캐스케이드<sup>KKR/ELD</sup> 뉴클레아제 쌍의 형질감염을 배제한다.

[0267] **검정 분석**

[0268] 형질감염된 세포를 수 일동안 배양하여 계대한다. 그 후, 생체 내 표적 DNA 절단 효율을 문헌[Guschin, D.Y., et al (2010) Methods Mol. Biol., 649: 247-256]의 검정 분석법을 이용하여 평가한다. 간략하게는, 표적 DNA 유전자좌의 PCR 앰플리콘을 미처리 세포로부터의 PCR 앰플리콘과 1:1로 혼합한다. 이들을 가열하고 어닐링시켜서 NHEJ에 의해 잘못 복구된 표적 부위에서의 미스매치가 생기게 한다. 그 후, 미스매치 뉴클레아제를 사용하여 단지 미스매치된 DNA 분자를 절단하여서, 캐스케이드<sup>KKR/ELD</sup>에 의한 표적 DNA 절단이 완전할 때 최대 50%의 절단을 제공한다. 그 후, 이 절차 후에, 처리된 세포의 표적 DNA 앰플리콘의 서열 결정을 한다. 이 분석법은 전달 절차의 신속한 평가 및 최적화를 허용한다.

[0269] **캐스케이드-뉴클레아제 쌍의 생성**

[0270] 캐스케이드-뉴클레아제 복합체를 상기에 설명한 바와 같이 구성하였다. 스트랩II-태그 Cse2 서브유닛 을 이용한 이. 콜라이로부터의 친화성 정제는 천연 캐스케이드와 비교할 때 예상된 화학량론적 양을 갖는 복합체를 생성한다. 도 13을 참조하면, 이는 단지 스트랩트액틴을 이용하여 정제한지 24시간 후 천연 캐스케이드 (1), P7 CrRNA를 포함하는 캐스케이드<sup>KKR</sup> 및 M13 CrRNA를 포함하는 캐스케이드<sup>ELD</sup>의 화학량론적 양을 나타낸다. 천연 캐스케이드 (1)에서의 밴드는 상부로부터 하부로 Cse1, Cas7, Cas5, Cas6e, Cse2이다. 캐스케이드<sup>KKR/ELD</sup>는 FokI-Cse1 융합 밴드와, 단백질 분해적 분해의 결과로서 FokI의 작은 부분을 포함하는 Cse1을 대표하는 추가의 밴드를 나타낸다.

[0271] 온전한 FokI-Cse1 융합 단백질을 제외하고, 본 발명자는 FokI-Cse1-융합 단백질의 분획이 단백질 분해에 의해 절단되어서, 단지 링커 및 FokI의 작은 부분이 그에 부착된 것을 포함하는 Cse1 단백질을 생성함을 관찰하였다(이는 질량 분광법에 의해 확인된 바와 같으며, 데이터는 예시되어 있지 않다). 대부분의 단백질 단리에서, 분해된 융합 단백질의 분율은 대략 40%이다. 단리된 단백질은 추가의 0.1% 트윈(Tween) 20 및 50% 글리세롤을 포함하는 용출 완충제(pH 7.5의 20 mM HEPES, 75 mM NaCl, 1 mM DTT, 4 mM 테스티오바이오틴)에서 -20°C에서 안정하게 보관한다. 이들 보관 조건 하에서, 복합체의 완전성 및 활성은 3주 이상 동안 안정한 것으로 밝혀졌다(데이터는 예시되어 있지 않음).

[0272] **His6-tag 및 NLS의 캐스케이드-뉴클레아제로의 도입**

[0273] 캐스케이드 뉴클레아제 융합물 설계를 변형시켜 핵인 국재화 신호(Nucleolar Localization Signal; NLS)를 포함시켜서 진핵 세포의 핵 내로의 수송을 가능하게 했다. 이를 위하여, 시미안 바이러스 SV40의 큰 T-항원으로 부터의 탠덤 1부분 NLS(서열: PKKKRKVDPKKKRKV)를, N-말단에서 His<sub>6</sub>-태그가 바로 선행하는 FokI-CseI 융합 단백질의 N-말단에 번역에 의해 융합시켰다. His<sub>6</sub>-태그 (서열: MHHHHHH)는 스트랩II 정제 후 추가의 Ni<sup>2+</sup>-수지 친화성 정제 단계를 허용한다. 이러한 추가의 단계는 단지 전장 캐스케이드-뉴클레아제 융합 복합체의 단리를 가능하게 하며, 비생산적 뉴클레아제 쌍을 형성하는, 온전하지 않은 캐스케이드 복합체의 표적 부위에의 결합을 제거함으로써 절단 효율을 증가시킨다.

[0274] **시험관 내 절단 분석**

[0275] 캐스케이드<sup>KKR/ELD</sup> 활성 및 특이성을 상기에 기술한 바와 같이 시험관 내에서 분석하였다. 도 14의 A는 37°C에서 30분 동안 캐스케이드<sup>KKR/ELD</sup>와 함께 인큐베이션한, 25-50 bp (5 bp의 증분, 레인 1-6)의 프로토스페이서들 사이의 거리를 갖는 플라스미드를 나타낸다. 레인 10은 3가지 가능한 토폴로지의 표적 플라스미드를 포함하며; 가장 낮은 밴드는 처음의 네거티브 수퍼코일형(nSC) 형태의 플라스미드를 나타내며, 중간 밴드는 선형화된 형태(XbaI으로 절단)를 나타내는 한편, 위쪽 밴드는 개방 원형(open circular; OC) 형태(Nt.BbrCI을 이용한 니킹 후)를 나타낸다. 레인 7은 둘 모두의 결합 부위가 제거된 플라스미드의 인큐베이션을 나타낸다(음성 대조구). 따라서, 도 14의 A는 결합 부위들이 5 bp 증분으로 25 내지 50 염기쌍에 의해 분리된 다양한 표적 플라스미드를 사용한 전형적인 절단 분석을 나타낸다(레인 1 내지 6). 25-50 bp의 거리를 갖는 이들 플라스미드를 각각 항 P7 및 M13 crRNA를 지닌 캐스케이드<sup>KKR/ELD</sup>와 함께 인큐베이션하였다. 결합 부위를 포함하지 않는 플라스미드는 대조구로서의 역할을 하였다(레인 7).

[0276] 원래의 플라스미드는 네거티브 수퍼코일형으로 존재하며(nSC, 대조 레인 8), 니킹된 또는 선형화된 생성물은 명백하게 구별가능하다. 인큐베이션시에, 결합 부위가 30, 35 및 40 염기쌍만큼 분리되어 있을 때 선형 절단 생성물이 형성된다(레인 2, 3, 4). 25, 45 및 50 염기쌍의 거리에서(레인 1, 5, 6), 표적 플라스미드는 불완전하게 절단되어 니 형태(OC)에 이르게 되는 것으로 보였다. 이들 결과는 30 내지 40 bp의 거리를 갖는 플라스미드에서의 최상의 절단을 나타내며, 이는 임의의 주어진 유전자좌에 있어서 crRNA 쌍을 설계할 때 충분한 융통성을 제공한다. 더욱 짧은 거리 및 더욱 긴 거리 둘 모두는 DSB를 덜 생성하면서 니킹 활성을 증가시킨다. 두 프로토스페이서가 제거된 플라스미드에서는 매우 적은 활성이 있는데, 이는 표적 특이성을 나타내는 것이다 (레인 7).

[0277] **절단 조건**

[0278] 절단 분석을 위한 최적 완충제 조건을 평가하기 위하여, 그리고 복합체의 활성이 생리학적 조건에서 예상될지를 추정하기 위하여, 하기의 두 완충제를 선택하였다: (1) NEB4 (뉴 잉글랜드 바이오랩스, 50 mM 아세트산칼륨, 20 mM 트리스-아세테이트, 10 mM 아세트산마그네슘, 1 mM 디티오프레이트, pH 7.9) 및 (2) 완충제 0 (퍼멘타스, 50 mM 트리스-HCl, 10 mM MgCl<sub>2</sub>, 100 mM NaCl, 0.1 mg/mL BSA, pH 7.5). 이들 둘 중, 상업용의 온전한 FokI 효소의 최적 활성을 위하여 NEB4가 권고된다. 완충제 0를 급속 스크린에 의해 선택하여 우수한 활성 및 특이성을 제공하였다(데이터는 예시되어 있지 않음). 도 14의 B는 상이한 완충제 및 상이한 인큐베이션 시간들을 이용한 인큐베이션을 나타낸다. 레인 1-4는 퍼멘타스 완충제 0를 이용하여 인큐베이션하였으며(레인 1, 2: 15분, 레인 3, 4: 30분), 레인 5, 6은 NEB4를 이용하여 인큐베이션하였다(30분). 레인 1, 3, 5는 35 bp의 이격치를 갖는 표적 플라스미드를 사용하였으며, 레인 2, 4, 6은 비표적 플라스미드(결합 부위 없음)를 사용하였다. 레인 7, 8은 각각 단지 캐스케이드<sup>KKR</sup> 또는 캐스케이드<sup>ELD</sup>(완충제 0)를 이용하여 인큐베이션하였다. 레인 9는 도 14의 A에서와 같은 토폴로지 마커이다. 레인 10 및 11은 캐스케이드를 첨가하지 않고서 인큐베이션한 표적 및 비표적 플라스미드를 나타낸다. 따라서, 도 14의 B에서, 35 염기쌍의 거리를 갖는 표적 플라스미드(레인 1, 3, 5) 및 비표적 대조 플라스미드(레인 2, 4, 6)에서 활성을 테스트하였다. NEB4에서 많은 양의 비특이적 니킹 및 더욱 적은 절단이 있었으며(레인 5, 6), 반면 완충제 0는 단지 표적 플라스미드에서 활성을 나타내며, 이때 많은 양의 특이적 절단 및 적은 니킹을 있다(레인 1-4). 그 차이는 완충제 0 중의 NaCl 농도에 의해 야기될 가능성이 있으며, 더욱 높은 이온 강도는 단백질-단백질 상호작용을 약화시켜 비특이적 활성이 덜해지게 한다. 15분 또는 30분의 인큐베이션은 표적 플라스미드 및 비표적 플라스미드 둘 모두에서 차이

를 거의 나타내지 않는다(각각 라인 1,2 또는 3,4). 단지 1가지 유형의 캐스케이드(P7<sup>KKR</sup> 또는 M13<sup>ELD</sup>)의 첨가는 예상되는 바와 같이 절단 활성을 생성하지 않는다(라인 7, 8). 이 실험은 NaCl 농도가 세포 내의 생리학적 염수 농도(137 mM NaCl) 근처인 100 mM 이상일 때, 설계된 쌍에 의한 특이적 캐스케이드 뉴클레아제 활성이 나타남을 보여 준다. 캐스케이드 뉴클레아제 쌍은 표적에서 멀리 있는 곳의 절단 활성은 무시가능함을 나타내면서 진행 세포에서는 생체 내에서 완전히 활성일 것으로 예상된다.

[0279] **절단 부위**

[0280] 35 bp의 이격치를 갖는 표적 플라스미드(pTarget35)에서의 절단 부위를 결정하였다. 도 15는 35 염기쌍의 이격치를 갖는 표적 플라스미드에서 서열 결정이 캐스케이드<sup>KKR/ELD</sup>에 의해 상류 및 하류 절단 부위를 보여 주는 방법을 나타낸다. 도 15의 A에는 잠재적인 절단 부위에 주석을 단 pTarget35 내의 표적 영역이 예시되어 있다. 프로토스페이서의 부분은 적색 및 청색으로 나타낸다. 도 15의 B) 막대 차트는 서열 결정된 클론들 내에서의 4가지의 상이한 절단 패턴 및 그의 상대적인 풍부함을 나타낸다. 청색 막대는 생성된 오버행을 나타내는 반면, 각각의 막대의 좌측 및 우측 경계는 좌측 및 우측 절단 부위를 나타낸다(주석에 대해서는 도 15의 B 참조).

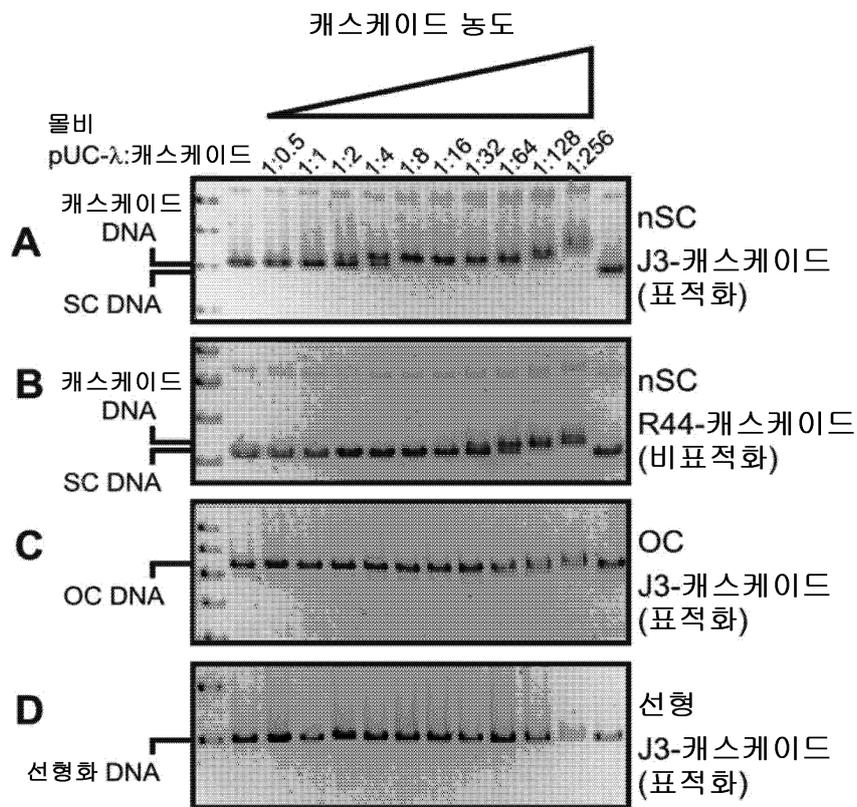
[0281] 도 15의 A는 pTarget35의 원래 서열을 나타내며, 이때 -7로부터 +7까지 절단 부위가 넘버링되며, 여기서, 두 프로토스페이서들 사이의 중간에 0이 있다(적색 및 청색으로 나타냄). 17개의 클론을 서열 결정하였으며, 이들 전부는 대략 위치 0에서 절단을 나타내는데, 이는 3 내지 5 bp의 다양한 오버행을 생성한다(도 15의 B 참조). 4의 오버행이 가장 풍부하며(누적으로 88%), 반면 3 및 5의 오버행은 단지 1회 나타난다(각각 6%). 절단은 예상되는 바와 같이 정확하게 일어났으며, 이때 어떠한 클론도 표적에서 먼 곳에서의 절단을 나타내지 않았다.

[0282] **인간 세포에서의 표적 유전자좌의 절단.**

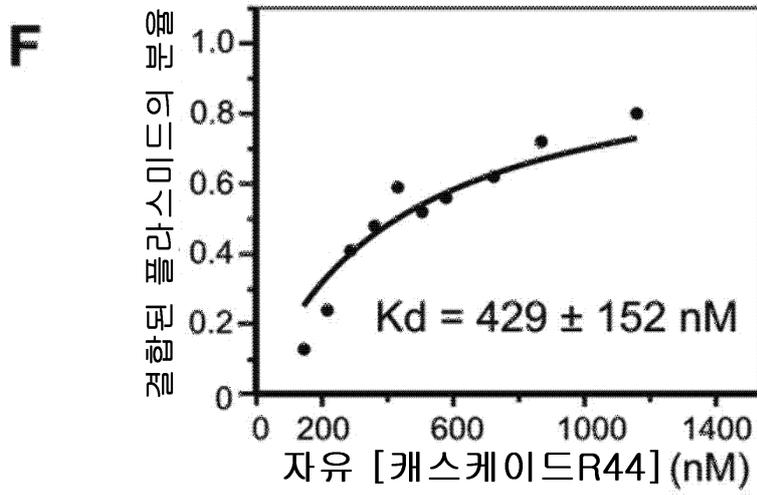
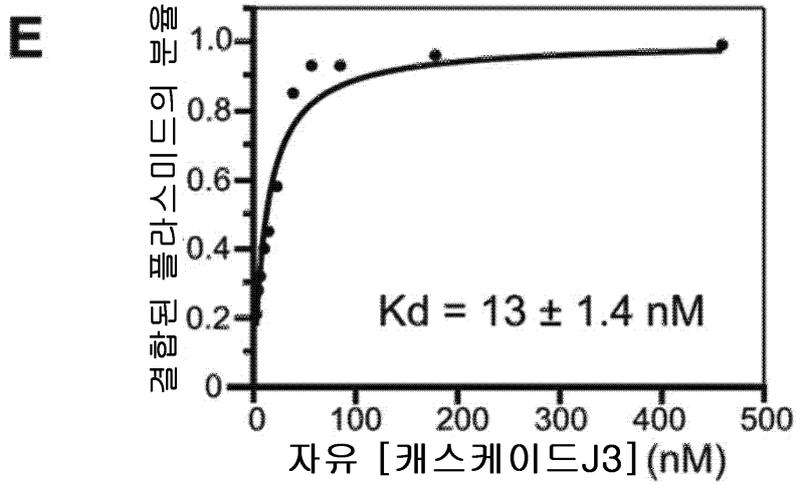
[0283] 캐스케이드<sup>KKR/ELD</sup> 뉴클레아제를 N-말단 His<sub>6</sub>-태그, 이어서 이중 1부분 핵인 국제화 신호를 함유하도록 성공적으로 변형시켰다. 이들 변형된 캐스케이드 뉴클레아제 융합 단백질들을, 각각이 인간 CCR5 유전자 내의 결합 부위를 표적화하는 두 합성 구성 CRISPR 어레이들 중 어느 하나와 함께 동시 발현시켰다. 먼저, 이 새로운 뉴클레아제 쌍의 활성을 CCR5 유전자의 이 영역을 포함하는 플라스미드에서 테스트함으로써 시험관 내에서 상기 활성을 입증한다. 상기 뉴클레아제 쌍으로 인간 세포주, 예를 들어 HeLa 세포주를 형질감염시킨다. 표적 절단 효율을 상기에 기술한 바와 같이 검정 분석법을 이용하여 평가한다.

도면

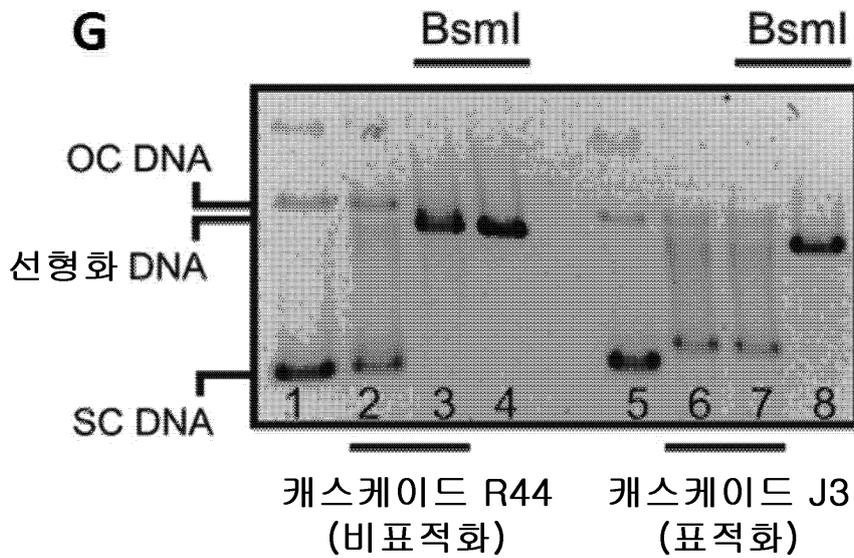
도면1a



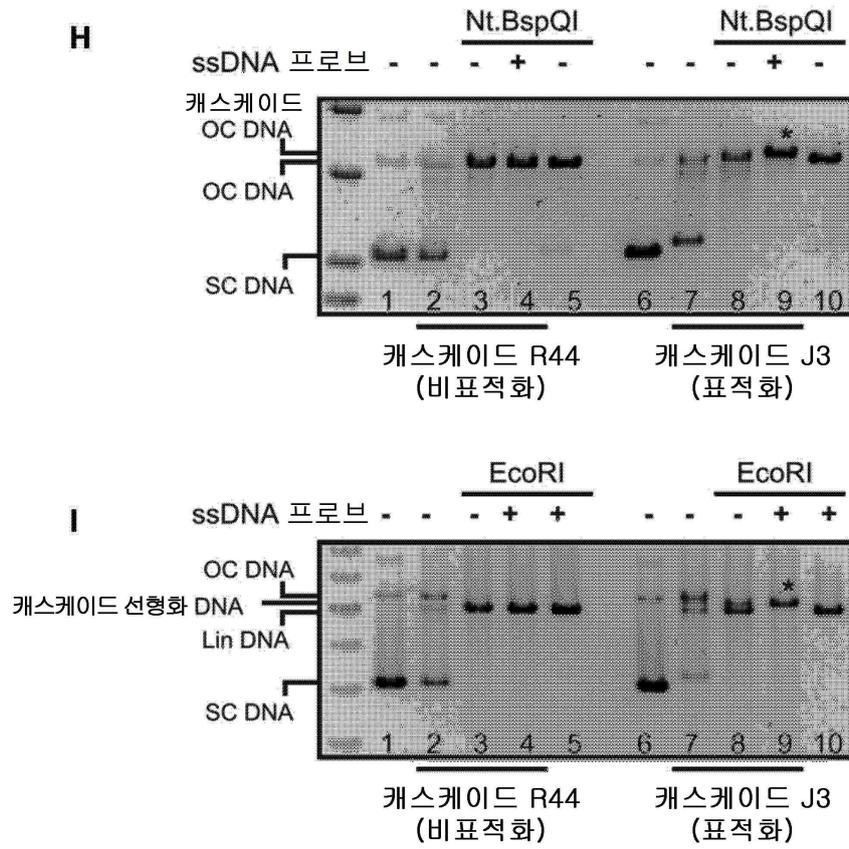
도면1b



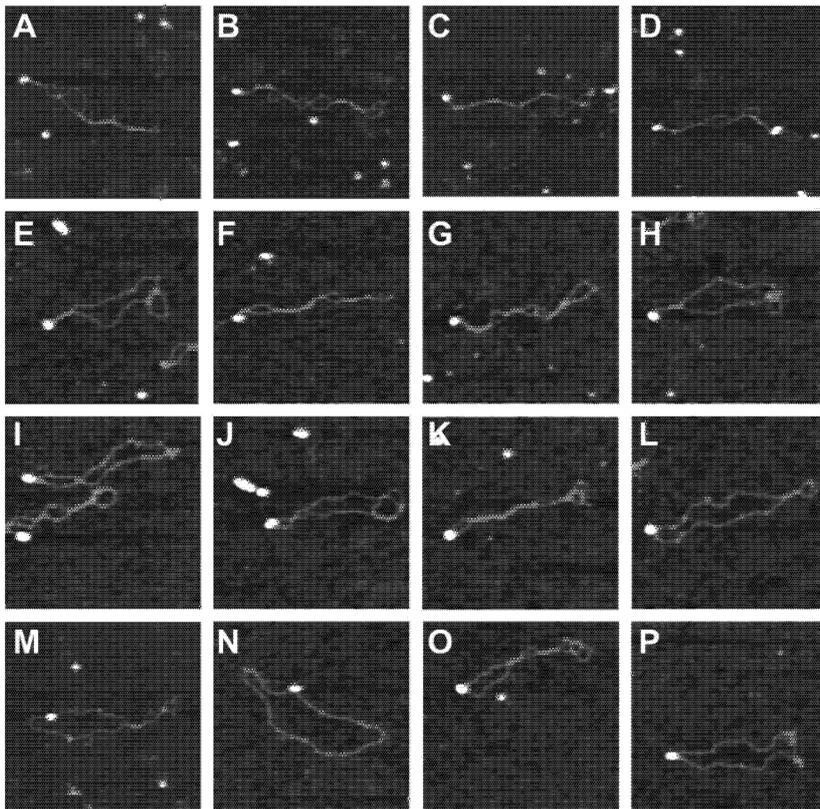
도면1c



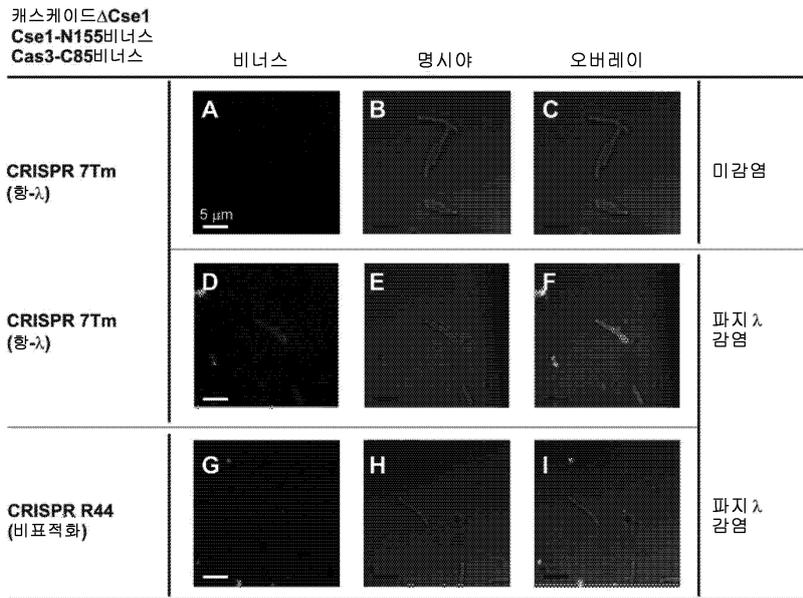
도면1d



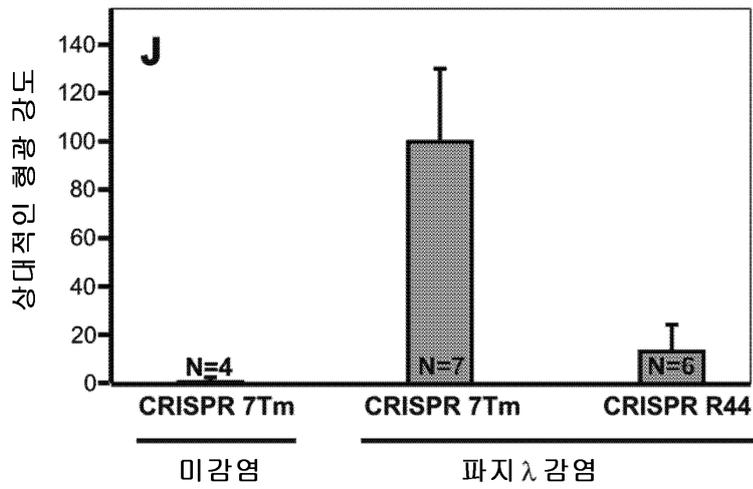
도면2



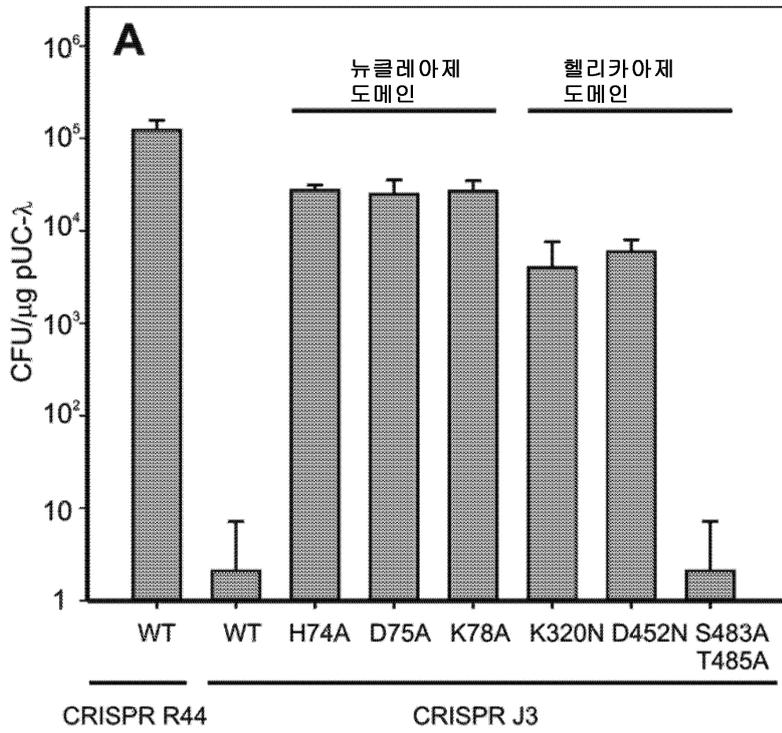
도면3a



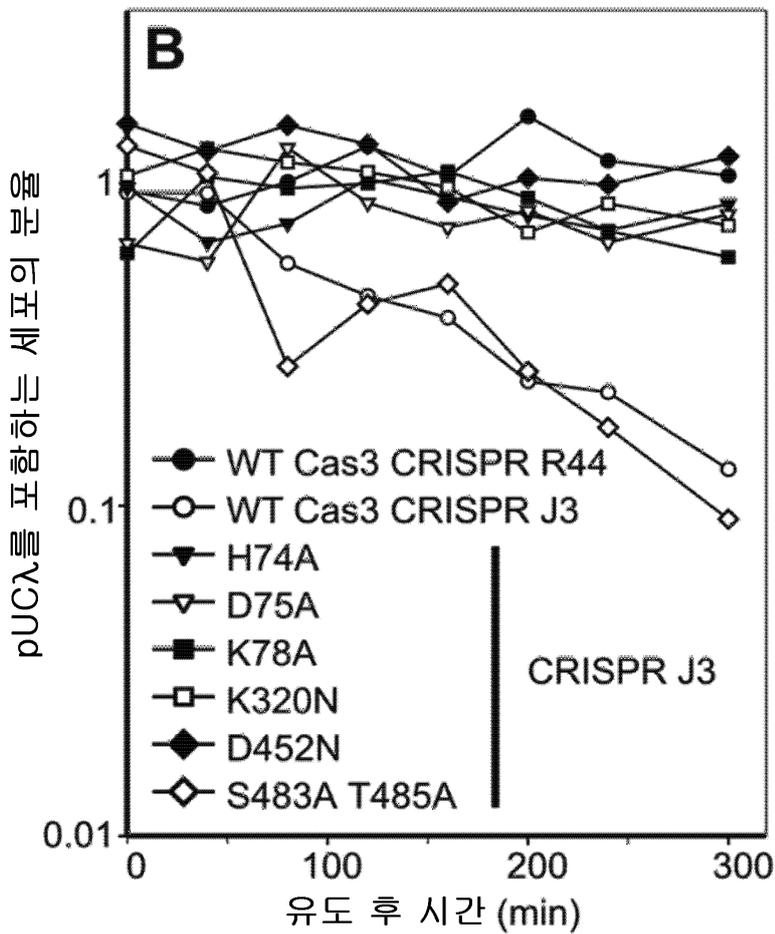
도면3b



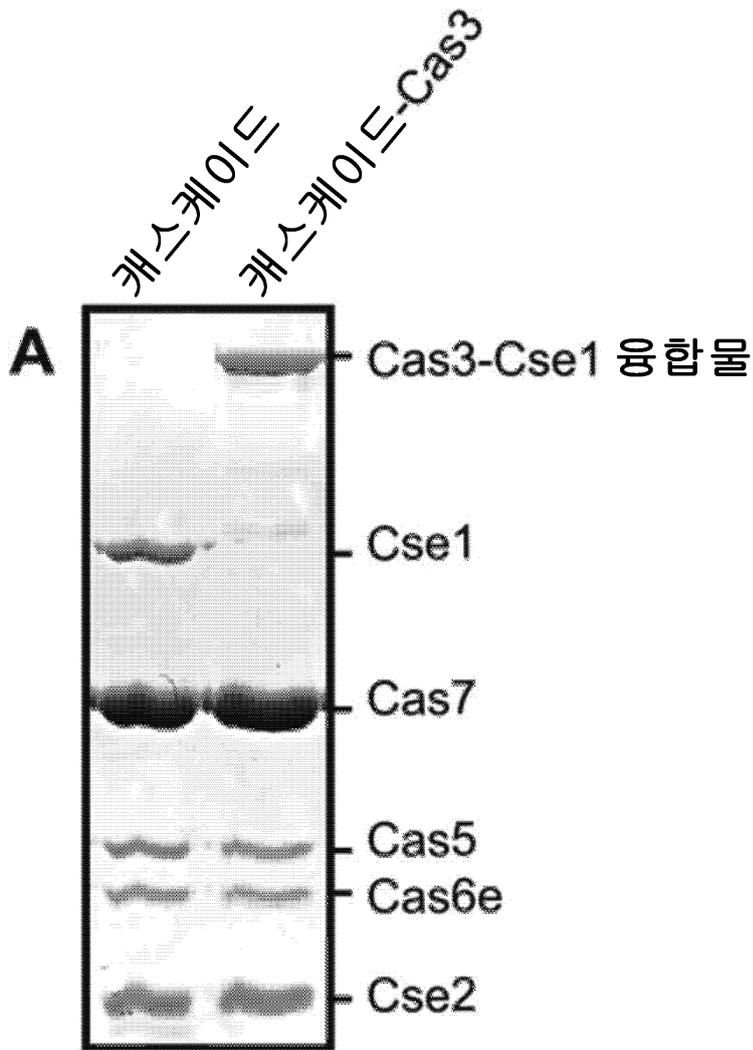
도면4a



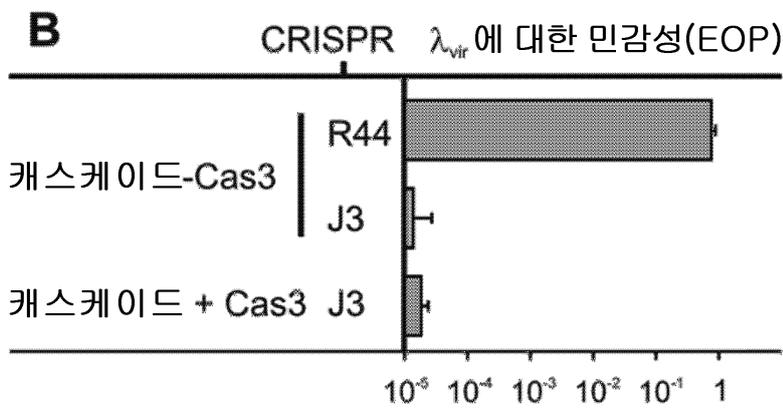
도면4b



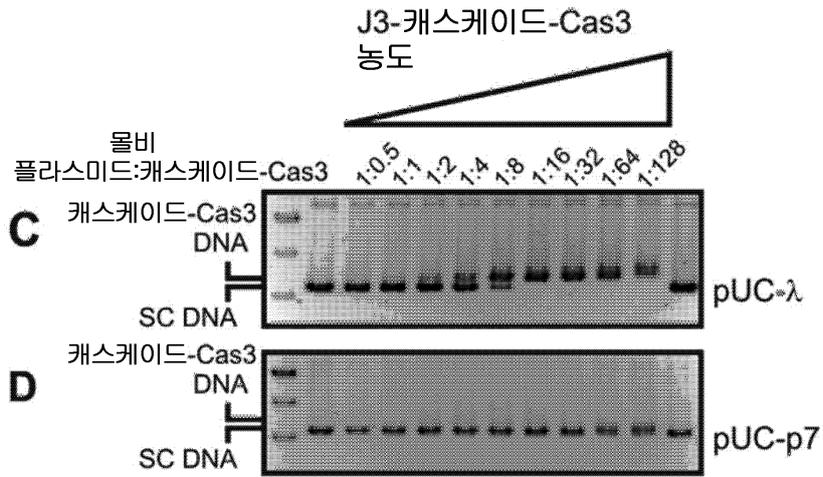
도면5a



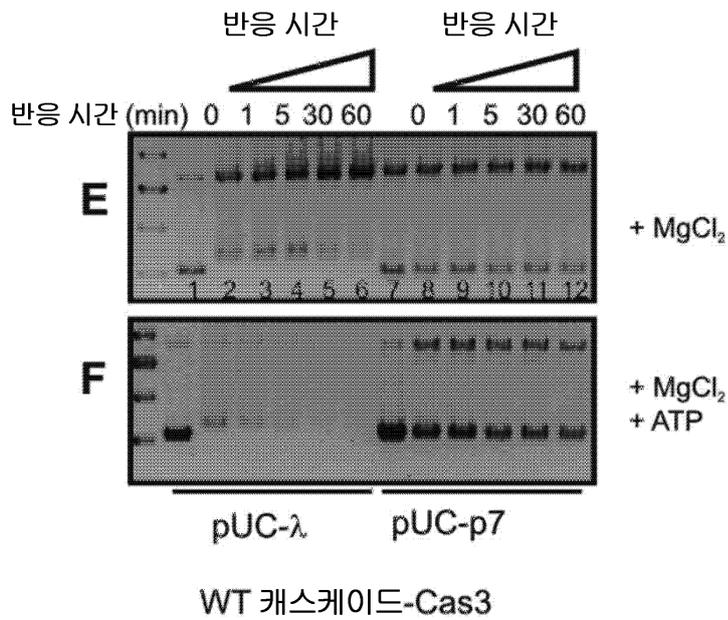
도면5b



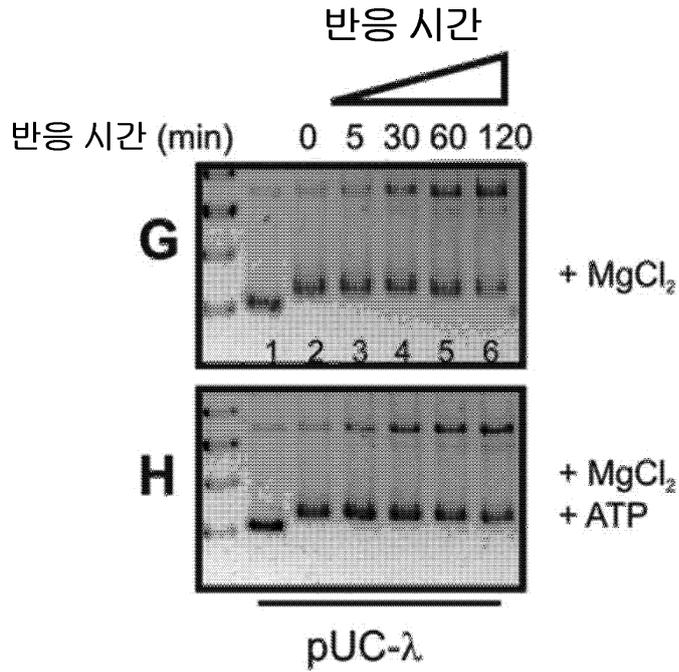
도면5c



도면5d

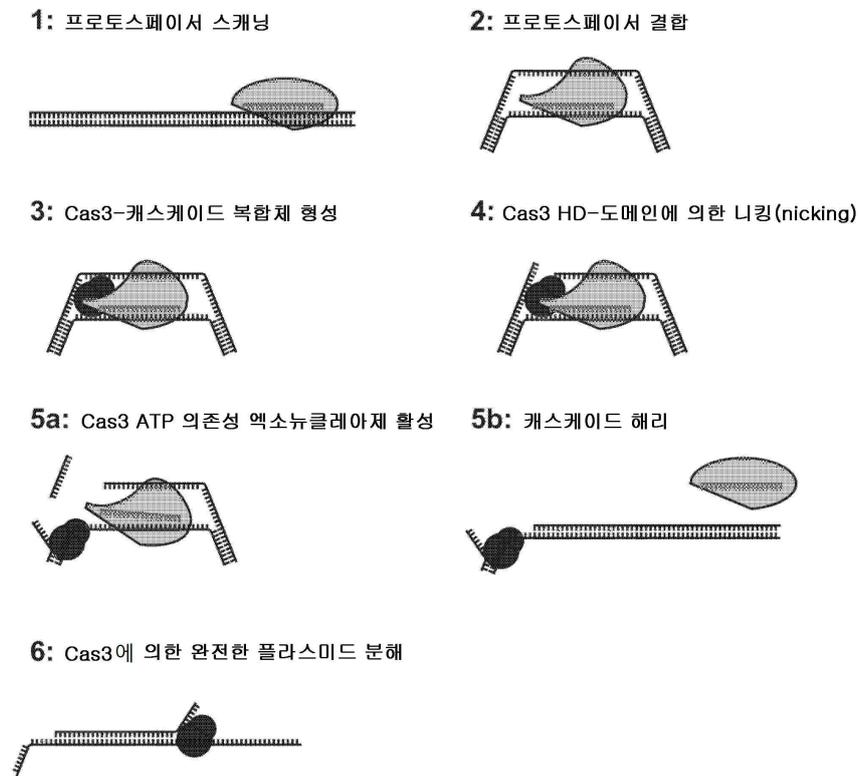


도면5e

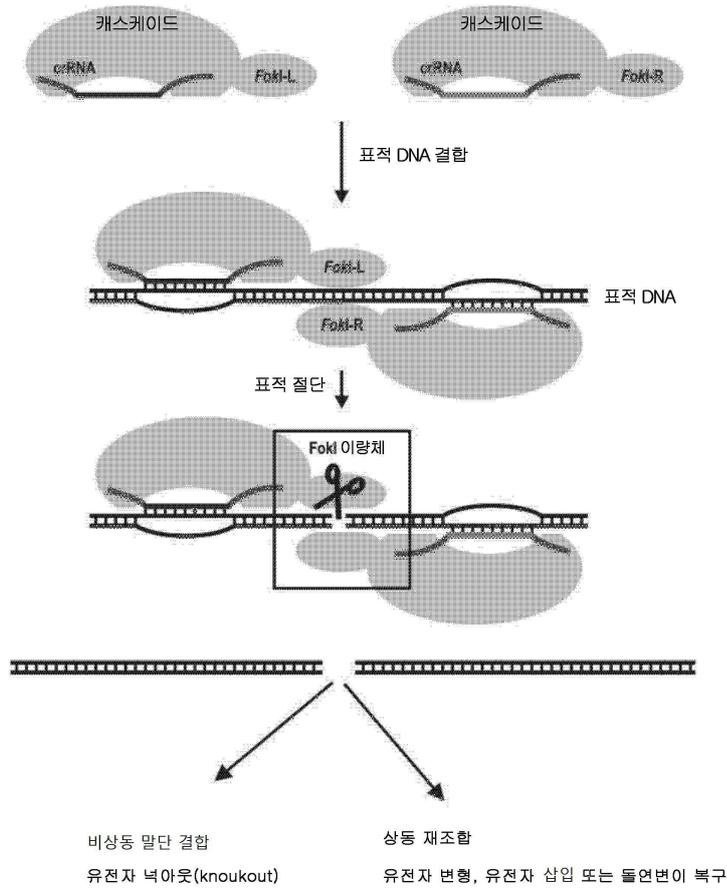


캐스케이드-Cas3 K320N

도면6



도면7



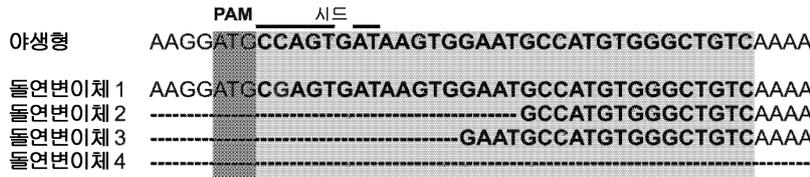
도면8

캐스케이드  $\Delta$ Cse1  
Cse1-N155비너스  
Cas3-C85비너스

오버레이

CRISPR 7Tm (황-λ)		미감염
CRISPR 7Tm (황-λ)		파지 λ 감염
CRISPR R44 (비표적화)		파지 λ 감염

도면9



도면10a

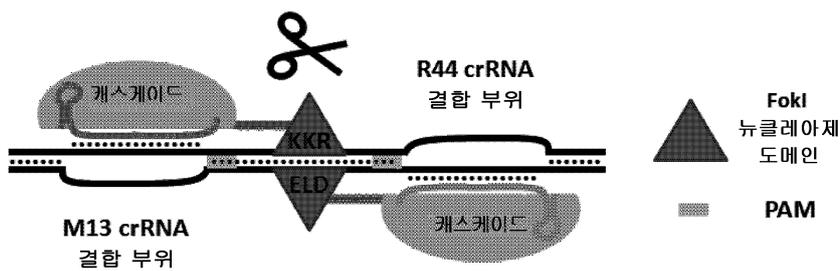
5. sp. S978	----	20	----	40	----	60	----	80	----	100
2. sp. S978	----	20	----	40	----	60	----	80	----	100
3. sp. S978	----	20	----	40	----	60	----	80	----	100
EC_C333-C3	----	20	----	40	----	60	----	80	----	100
5. sp. S978	----	120	----	140	----	160	----	180	----	200
2. sp. S978	----	120	----	140	----	160	----	180	----	200
3. sp. S978	----	120	----	140	----	160	----	180	----	200
EC_C333-C3	----	120	----	140	----	160	----	180	----	200
5. sp. S978	----	220	----	240	----	260	----	280	----	300
2. sp. S978	----	220	----	240	----	260	----	280	----	300
3. sp. S978	----	220	----	240	----	260	----	280	----	300
EC_C333-C3	----	220	----	240	----	260	----	280	----	300
5. sp. S978	----	320	----	340	----	360	----	380	----	400
2. sp. S978	----	320	----	340	----	360	----	380	----	400
3. sp. S978	----	320	----	340	----	360	----	380	----	400
EC_C333-C3	----	320	----	340	----	360	----	380	----	400
5. sp. S978	----	420	----	440	----	460	----	480	----	500
2. sp. S978	----	420	----	440	----	460	----	480	----	500
3. sp. S978	----	420	----	440	----	460	----	480	----	500
EC_C333-C3	----	420	----	440	----	460	----	480	----	500
5. sp. S978	----	520	----	540	----	560	----	580	----	600
2. sp. S978	----	520	----	540	----	560	----	580	----	600
3. sp. S978	----	520	----	540	----	560	----	580	----	600
EC_C333-C3	----	520	----	540	----	560	----	580	----	600
5. sp. S978	----	620	----	640	----	660	----	680	----	700
2. sp. S978	----	620	----	640	----	660	----	680	----	700
3. sp. S978	----	620	----	640	----	660	----	680	----	700
EC_C333-C3	----	620	----	640	----	660	----	680	----	700
5. sp. S978	----	720	----	740	----	760	----	780	----	800
2. sp. S978	----	720	----	740	----	760	----	780	----	800
3. sp. S978	----	720	----	740	----	760	----	780	----	800
EC_C333-C3	----	720	----	740	----	760	----	780	----	800
5. sp. S978	----	820	----	840	----	860	----	880	----	900
2. sp. S978	----	820	----	840	----	860	----	880	----	900
3. sp. S978	----	820	----	840	----	860	----	880	----	900
EC_C333-C3	----	820	----	840	----	860	----	880	----	900
5. sp. S978	----	920	----	940	----	960	----	980	----	1000
2. sp. S978	----	920	----	940	----	960	----	980	----	1000
3. sp. S978	----	920	----	940	----	960	----	980	----	1000
EC_C333-C3	----	920	----	940	----	960	----	980	----	1000

도면10b

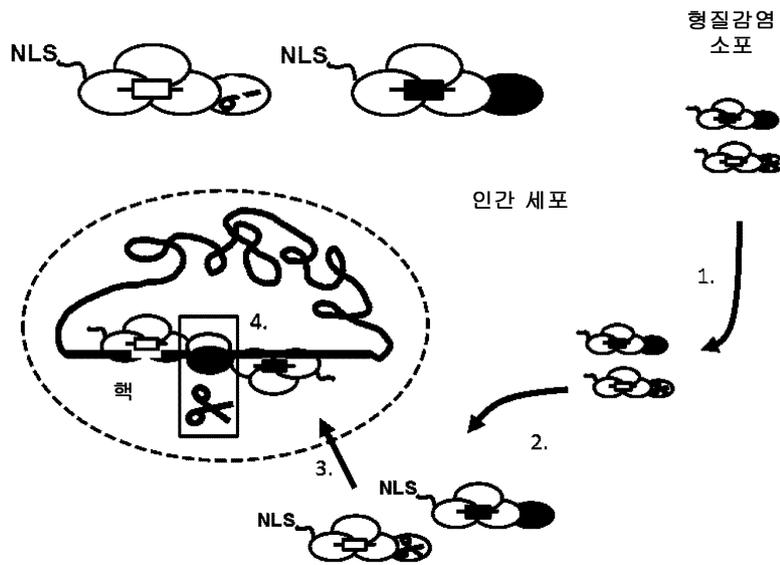
```

5_5q_3p278 : RRRPOLLHPRRSPFQTRGLITANKNEVSLNRLLVAVDPKGGPFSNRRPQNRQLGFKEARIRLVHAKQVDTISGHTTQVGDPRNEACGCPYRGRANKSNLGGRLGEG : 1252
5_6p_13q25 : RRPFLLHPRRFRQWGLHTTKHEVWALMLTAVDVPKGEAFSSRRRQVHRLGAEAKRRLVHTHADVLSGHSISGKCGARVCKKVPYQVQWVGGALGGVFAEGG : 1189
5_8c1d1d6t : RRRPOLLHPRATPRVQVYGLQTKNGEVRLLKTLVADVPDQGLTTRRRPQVQLSFAEAKRRLVHTHADVDTISGHSISGKCGARVCKKVPYQVQWVGGALGGVFAEGG : 1181
EC_Ca33-C3 : IDMPFLMHAERPRRQTKRQVAK--DVTREKRLLAGVSGATKCAFLNPPQDQEGALGGCTALALPKQARQVAFSGGFSGLRG-----GPTVTVRGI : 1084
* : 1308
5_5q_3p278 : MLKTELLMLLADDTQVYHA-AEVDPRRRAE-PSGP--AFAADLGLRVEGLRDLVYRQSSRRTRLLHRSADQVGVAVYQSLRPHRRHQHRRPITSRKRSPTQERKA : 1355
5_6p_13q25 : SAKTELLMLLIPDTEELTSEPKADSRFRPQVVEQDFAKRRAPRDLVYRQSSRRTRLLHTEGDAIVGVYQSLRPAKRRQTEPITQWRRSFQERKLE : 1286
5_8c1d1d6t : TLKTELLMLLIPADTRRQVTSAGDQVRRMAGTRKAK--SVAADLSEKRVVAGLIDLVYRQSSRRTRLLHTEGDAIVGVYQSLRPAKRRQTEPITQWRRSFQERKLE : 1287
EC_Ca33-C3 : DMSTVELVETLPRLDQRPFRMSHTERRPHTIRPK--SRESFRASISGRPRQLRPPRMLLECPPLG--KAKTSCDDESNLHTTFLNHTTWNQSLRPHR : 1186
* : 1400
5_5q_3p278 : QENLVYEP-ENDP5KRAMKQGLLAC-RETSAGDQVSDQRLAK-VOGMAKRLTEGSLRPGVYLIRTRVSGAVGTQDSVDEVDQVLFPAVLLHQRNVG : 1479
5_6p_13q25 : QENLVYEP-ENDP5KRAMKQGLLAC-RETSAGDQVSDQRLAK-VOGMAKRLTEGSLRPGVYLIRTRVSGAVGTQDSVDEVDQVLFPAVLLHQRNVG : 1481
5_8c1d1d6t : QENLVYEP-ENDP5KRAMKQGLLAC-RETSAGDQVSDQRLAK-VOGMAKRLTEGSLRPGVYLIRTRVSGAVGTQDSVDEVDQVLFPAVLLHQRNVG : 1487
EC_Ca33-C3 : SGLVYRKRVEEHTKATTTSMSTLISNVVSNLQKSNHNVAVAVQVRIANQSPKLE-----LRSQKRNQSLRHRNP-----VLRNRSQKQVTV : 1280
* : 1430
5_5q_3p278 : DNAYDVALHDAEKAVGALAKLADLALAV-CTDPRRNTTARDLGGFTLLDTHYRMRRLGQTSDFEERHRNRKQVRLVAELGERLLRQAGAAAEGRVET-GRG : 1584
5_6p_13q25 : DNAYDVALHDAEKAVGALAKLADLALAV-CTDPRRNTTARDLGGFTLLDTHYRMRRLGQTSDFEERHRNRKQVRLVAELGERLLRQAGAAAEGRVET-GRG : 1584
5_8c1d1d6t : DNAYDVALHDAEKAVGALAKLADLALAV-CTDPRRNTTARDLGGFTLLDTHYRMRRLGQTSDFEERHRNRKQVRLVAELGERLLRQAGAAAEGRVET-GRG : 1587
EC_Ca33-C3 : NVNNEIVTYGLQVITLRKALVYTAEGRHKRHPKGGVSHETALRHRVRSQELLTRDVLAWNFSQADVDVADLIDKLNQLCENLFRQSVAFVARRKLLSTLALA : 1387
* : 1520
5_5q_3p278 : TRMLDQAAELRTRRLRERLTAIPRPS-SPRNPVSEEA----- : 1604
5_6p_13q25 : TRMLDQAAELRTRRLRERLTAIPRPS-SPRNPVSEEA----- : 1599
5_8c1d1d6t : ESRMDQSLADLYRKGIAKALSSRLCKRPTDRCFPGKIDRPPA----- : 1407
EC_Ca33-C3 : RALTYMHLRELNKRGDSNG----- : 1407
    
```

도면11

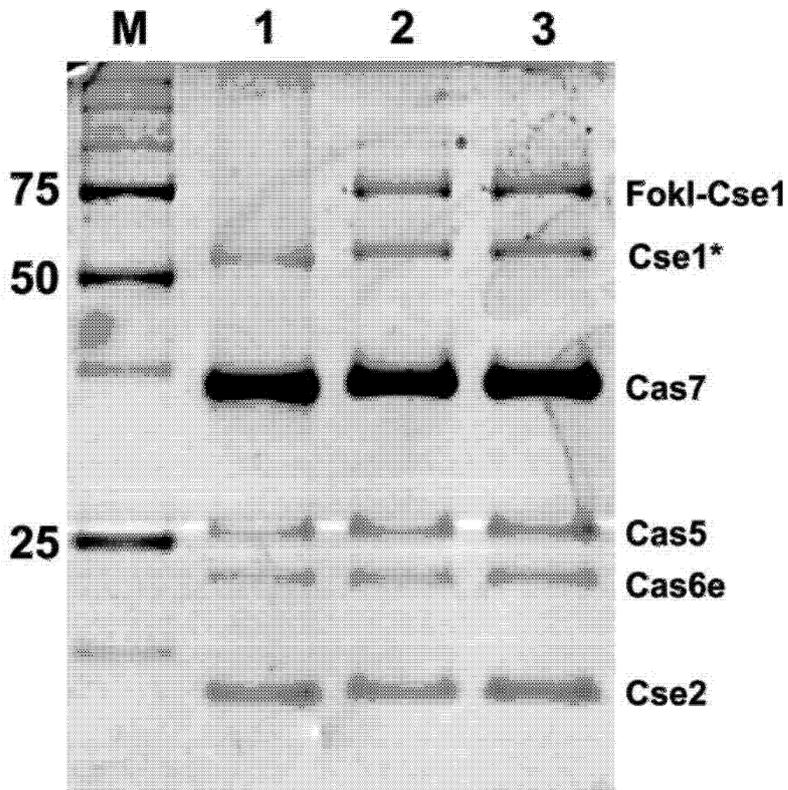


도면12



도면13

<b>FokI 변이체</b>	-	<b>KKR</b>	<b>ELD</b>
<b>로딩된 crRNA</b>	<b>J3</b>	<b>P7</b>	<b>M13</b>





서열 목록

SEQUENCE LISTING

<110> Wageningen Universiteit

<120> MODIFIED CASCADE RIBONUCLEOPROTEINS AND USES THEREOF

<130> P200547WO

<150> GB1122458.1

<151> 2011-12-30

<160> 47

<170> PatentIn version 3.5

<210> 1

<211> 502

<212> PRT

<213> Escherichia coli

<400> 1

Met Asn Leu Leu Ile Asp Asn Trp Ile Pro Val Arg Pro Arg Asn Gly

1                    5                    10                    15

Gly Lys Val Gln Ile Ile Asn Leu Gln Ser Leu Tyr Cys Ser Arg Asp

20                    25                    30

Gln Trp Arg Leu Ser Leu Pro Arg Asp Asp Met Glu Leu Ala Ala Leu

35                    40                    45

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

50                    55                    60

Val Glu Phe Arg His Arg Ile Met Asn Pro Leu Thr Glu Asp Glu Phe

65                    70                    75                    80

Gln Gln Leu Ile Ala Pro Trp Ile Asp Met Phe Tyr Leu Asn His Ala

85                    90                    95

Glu His Pro Phe Met Gln Thr Lys Gly Val Lys Ala Asn Asp Val Thr

100                    105                    110

Pro Met Glu Lys Leu Leu Ala Gly Val Ser Gly Ala Thr Asn Cys Ala

115                    120                    125

Phe Val Asn Gln Pro Gly Gln Gly Glu Ala Leu Cys Gly Gly Cys Thr

130                    135                    140

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

145                    150                    155                    160  
 Gly Phe Lys Ser Gly Leu Arg Gly Gly Thr Pro Val Thr Thr Phe Val  
                                  165                    170                    175  
 Arg Gly Ile Asp Leu Arg Ser Thr Val Leu Leu Asn Val Leu Thr Leu  
                                  180                    185                    190  
 Pro Arg Leu Gln Lys Gln Phe Pro Asn Glu Ser His Thr Glu Asn Gln  
                                  195                    200                    205  
 Pro Thr Trp Ile Lys Pro Ile Lys Ser Asn Glu Ser Ile Pro Ala Ser  
  
                                  210                    215                    220  
 Ser Ile Gly Phe Val Arg Gly Leu Phe Trp Gln Pro Ala His Ile Glu  
 225                    230                    235                    240  
 Leu Cys Asp Pro Ile Gly Ile Gly Lys Cys Ser Cys Cys Gly Gln Glu  
                                  245                    250                    255  
 Ser Asn Leu Arg Tyr Thr Gly Phe Leu Lys Glu Lys Phe Thr Phe Thr  
                                  260                    265                    270  
 Val Asn Gly Leu Trp Pro His Pro His Ser Pro Cys Leu Val Thr Val  
  
                                  275                    280                    285  
 Lys Lys Gly Glu Val Glu Glu Lys Phe Leu Ala Phe Thr Thr Ser Ala  
                                  290                    295                    300  
 Pro Ser Trp Thr Gln Ile Ser Arg Val Val Val Asp Lys Ile Ile Gln  
 305                    310                    315                    320  
 Asn Glu Asn Gly Asn Arg Val Ala Ala Val Val Asn Gln Phe Arg Asn  
                                  325                    330                    335  
 Ile Ala Pro Gln Ser Pro Leu Glu Leu Ile Met Gly Gly Tyr Arg Asn  
  
                                  340                    345                    350  
 Asn Gln Ala Ser Ile Leu Glu Arg Arg His Asp Val Leu Met Phe Asn  
                                  355                    360                    365  
 Gln Gly Trp Gln Gln Tyr Gly Asn Val Ile Asn Glu Ile Val Thr Val  
                                  370                    375                    380  
 Gly Leu Gly Tyr Lys Thr Ala Leu Arg Lys Ala Leu Tyr Thr Phe Ala  
 385                    390                    395                    400

Glu Gly Phe Lys Asn Lys Asp Phe Lys Gly Ala Gly Val Ser Val His

405 410 415

Glu Thr Ala Glu Arg His Phe Tyr Arg Gln Ser Glu Leu Leu Ile Pro

420 425 430

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

435 440 445

Asp Leu Arg Asp Lys Leu His Gln Leu Cys Glu Met Leu Phe Asn Gln

450 455 460

Ser Val Ala Pro Tyr Ala His His Pro Lys Leu Ile Ser Thr Leu Ala

465 470 475 480

Leu Ala Arg Ala Thr Leu Tyr Lys His Leu Arg Glu Leu Lys Pro Gln

485 490 495

Gly Gly Pro Ser Asn Gly

500

<210> 2

<211> 160

<212> PRT

<213> Escherichia coli

<400> 2

Met Ala Asp Glu Ile Asp Ala Met Ala Leu Tyr Arg Ala Trp Gln Gln

1 5 10 15

Leu Asp Asn Gly Ser Cys Ala Gln Ile Arg Arg Val Ser Glu Pro Asp

20 25 30

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

35 40 45

Trp Glu Asn Pro Arg His Gln Gln Ala Leu Leu Arg Met Val Phe Cys

50 55 60

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

65 70 75 80

Gln Thr Thr Gly Ile Ser Leu Gly Arg Ala Leu Ala Asn Ser Gly Arg

85 90 95

Ile Asn Glu Arg Arg Ile Phe Gln Leu Ile Arg Ala Asp Arg Thr Ala  
 100 105 110  
 Asp Met Val Gln Leu Arg Arg Leu Leu Thr His Ala Glu Pro Val Leu  
 115 120 125  
 Asp Trp Pro Leu Met Ala Arg Met Leu Thr Trp Trp Gly Lys Arg Glu  
 130 135 140  
 Arg Gln Gln Leu Leu Glu Asp Phe Val Leu Thr Thr Asn Lys Asn Ala

145 150 155 160

<210> 3

<211> 363

<212> PRT

<213> Escherichia coli

<400> 3

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

Arg Lys Ser Gly Tyr Tyr Ala Gln Asn Ile Gly Glu Ser Ser Leu Arg  
 50 55 60

Thr Ile His Leu Ala Gln Leu Arg Asp Val Leu Arg Gln Lys Leu Gly  
 65 70 75 80

Glu Arg Phe Asp Gln Lys Ile Ile Asp Lys Thr Leu Ala Leu Leu Ser  
 85 90 95

Gly Lys Ser Val Asp Glu Ala Glu Lys Ile Ser Ala Asp Ala Val Thr  
 100 105 110

Pro Trp Val Val Gly Glu Ile Ala Trp Phe Cys Glu Gln Val Ala Lys  
 115 120 125

Ala Glu Ala Asp Asn Leu Asp Asp Lys Lys Leu Leu Lys Val Leu Lys  
 130 135 140

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

145                    150                    155                    160  
 Ala Leu Ser Gly Arg Met Ala Thr Ser Gly Met Met Thr Glu Leu Gly  
                                  165                    170                    175

Lys Val Asp Gly Ala Met Ser Ile Ala His Ala Ile Thr Thr His Gln  
                                  180                    185                    190

Val Asp Ser Asp Ile Asp Trp Phe Thr Ala Val Asp Asp Leu Gln Glu  
                                  195                    200                    205

Gln Gly Ser Ala His Leu Gly Thr Gln Glu Phe Ser Ser Gly Val Phe  
                                  210                    215                    220

Tyr Arg Tyr Ala Asn Ile Asn Leu Ala Gln Leu Gln Glu Asn Leu Gly  
 225                    230                    235                    240

Gly Ala Ser Arg Glu Gln Ala Leu Glu Ile Ala Thr His Val Val His  
                                  245                    250                    255

Met Leu Ala Thr Glu Val Pro Gly Ala Lys Gln Arg Thr Tyr Ala Ala  
                                  260                    265                    270

Phe Asn Pro Ala Asp Met Val Met Val Asn Phe Ser Asp Met Pro Leu  
                                  275                    280                    285

Ser Met Ala Asn Ala Phe Glu Lys Ala Val Lys Ala Lys Asp Gly Phe  
                                  290                    295                    300

Leu Gln Pro Ser Ile Gln Ala Phe Asn Gln Tyr Trp Asp Arg Val Ala  
 305                    310                    315                    320

Asn Gly Tyr Gly Leu Asn Gly Ala Ala Ala Gln Phe Ser Leu Ser Asp  
                                  325                    330                    335

Val Asp Pro Ile Thr Ala Gln Val Lys Gln Met Pro Thr Leu Glu Gln  
                                  340                    345                    350

Leu Lys Ser Trp Val Arg Asn Asn Gly Glu Ala  
                                  355                    360

<210> 4

<211> 224

<212> PRT

<213> Escherichia coli

<400> 4

Met Arg Ser Tyr Leu Ile Leu Arg Leu Ala Gly Pro Met Gln Ala Trp  
 1                    5                    10                    15  
 Gly Gln Pro Thr Phe Glu Gly Thr Arg Pro Thr Gly Arg Phe Pro Thr  
                   20                    25                    30  
 Arg Ser Gly Leu Leu Gly Leu Leu Gly Ala Cys Leu Gly Ile Gln Arg  
                   35                    40                    45  
 Asp Asp Thr Ser Ser Leu Gln Ala Leu Ser Glu Ser Val Gln Phe Ala  
  
                   50                    55                    60  
 Val Arg Cys Asp Glu Leu Ile Leu Asp Asp Arg Arg Val Ser Val Thr  
 65                    70                    75                    80  
 Gly Leu Arg Asp Tyr His Thr Val Leu Gly Ala Arg Glu Asp Tyr Arg  
                   85                    90                    95  
 Gly Leu Lys Ser His Glu Thr Ile Gln Thr Trp Arg Glu Tyr Leu Cys  
                   100                    105                    110  
 Asp Ala Ser Phe Thr Val Ala Leu Trp Leu Thr Pro His Ala Thr Met  
  
                   115                    120                    125  
 Val Ile Ser Glu Leu Glu Lys Ala Val Leu Lys Pro Arg Tyr Thr Pro  
                   130                    135                    140  
 Tyr Leu Gly Arg Arg Ser Cys Pro Leu Thr His Pro Leu Phe Leu Gly  
 145                    150                    155                    160  
 Thr Cys Gln Ala Ser Asp Pro Gln Lys Ala Leu Leu Asn Tyr Glu Pro  
                   165                    170                    175  
 Val Gly Gly Asp Ile Tyr Ser Glu Glu Ser Val Thr Gly His His Leu  
  
                   180                    185                    190  
 Lys Phe Thr Ala Arg Asp Glu Pro Met Ile Thr Leu Pro Arg Gln Phe  
                   195                    200                    205  
 Ala Ser Arg Glu Trp Tyr Val Ile Lys Gly Gly Met Asp Val Ser Gln  
                   210                    215                    220

<210> 5

<211> 199

<212> PRT

<213> Escherichia coli

<400> 5

Met Tyr Leu Ser Lys Val Ile Ile Ala Arg Ala Trp Ser Arg Asp Leu  
 1                    5                    10                    15

Tyr Gln Leu His Gln Gly Leu Trp His Leu Phe Pro Asn Arg Pro Asp  
                   20                    25                    30

Ala Ala Arg Asp Phe Leu Phe His Val Glu Lys Arg Asn Thr Pro Glu  
                   35                    40                    45

Gly Cys His Val Leu Leu Gln Ser Ala Gln Met Pro Val Ser Thr Ala  
                   50                    55                    60

Val Ala Thr Val Ile Lys Thr Lys Gln Val Glu Phe Gln Leu Gln Val  
 65                    70                    75                    80

Gly Val Pro Leu Tyr Phe Arg Leu Arg Ala Asn Pro Ile Lys Thr Ile  
                   85                    90                    95

Leu Asp Asn Gln Lys Arg Leu Asp Ser Lys Gly Asn Ile Lys Arg Cys  
                   100                    105                    110

Arg Val Pro Leu Ile Lys Glu Ala Glu Gln Ile Ala Trp Leu Gln Arg  
                   115                    120                    125

Lys Leu Gly Asn Ala Ala Arg Val Glu Asp Val His Pro Ile Ser Glu  
                   130                    135                    140

Arg Pro Gln Tyr Phe Ser Gly Asp Gly Lys Ser Gly Lys Ile Gln Thr  
 145                    150                    155                    160

Val Cys Phe Glu Gly Val Leu Thr Ile Asn Asp Ala Pro Ala Leu Ile  
                   165                    170                    175

Asp Leu Val Gln Gln Gly Ile Gly Pro Ala Lys Ser Met Gly Cys Gly  
                   180                    185                    190

Leu Leu Ser Leu Ala Pro Leu  
                   195

<210> 6

<211> 392

<212> DNA

<213> Artificial Sequence

<220><223> GA1070943

<400> 6

actggaagc gggcagtga aggaaggccc atgaggccag ttaattaagc ggatcctggc 60  
 ggcggcagcg gcggcggcag cgacaagcag aagaacggca tcaaggcga cttcaagatc 120  
 cgccacaaca tcgaggacgg cggcgtgcag ctgcgccacc actaccagca gaacaccccc 180  
 atcggcgacg gccccgtgct gctgcccacc aaccactacc tgagctacca gtccgcctg 240  
 agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc tggagtctgt gaccgccgcc 300  
 gggatcactc tcggcatgga cgagctgtac aagtaagcgg ccgcggcgcg cctagcctt 360  
 gacggccttc cttcaattcg ccctatagt ag 392

<210> 7

<211> 603

<212> DNA

<213> Artificial Sequence

<220><223> GA1070941

<400> 7

cactataggg cgaattggcg gaagcccgtc aaggcccat ttaattaagc ggccgcaggc 60  
 ggcggcagcg gcggcggcag catggtgagc aagggcgagg agctgttcac cggggtggtg 120  
 cccatcctgg tcgagctgga cggcgacgta aacggccaca agttcagcgt gtccggcgag 180  
 ggcgaggcgc atgccaccta cggcaagctg accctgaagc tcatctgcac caccgcaag 240  
 ctgcccgctg cctggcccac cctcgtgacc accctcggt acggcctgca gtgcttcgcc 300  
 cgctaccccg accacatgaa gcagcacgac ttcttcaagt ccgcatgcc cgaaggctac 360  
  
 gtccaggagc gcaccatctt cttcaaggac gacggcaact acaagaccg cgccgaggtg 420  
 aagttcgagg gcgacacct ggtgaaccgc atcgagctga agggcatcga cttcaaggag 480  
 gacggcaaca tctggggca caagctggag tacaactaca acagccaaa cgtctatata 540  
 acggcctaac tcgagggcgc gccctgggcc tcatgggcct tccgctcact gcccgctttc 600  
 cag 603

<210> 8

<211> 679

<212> DNA

<213> Artificial Sequence

<220><223> GA1068859

<400> 8

cactataggg cgaattggcg gaaggccgtc aaggccgcat gagctccatg gaaacaaaga 60

attagctgat cttaataat aaggaaatgt tacattaagg ttggtgggtt gtttttatgg 120

gaaaaaatgc ttaagaaca aatgtatact ttagagagt tccccgccc agcggggata 180

aaccgggccg attgaagtc cggtaggatg cttaaaagag tccccgccc cagcggggat 240

aaaccgccg aggtacagca ggtagcgcag atcatcaaga gttccccgccc ccagcgggga 300

taaaccgact tcctccgaa aagtcaggac gctgtggcag agttccccg gccagcgggg 360

ataaacgcc tacgcgctga acgccagcgg tgtggtgaat gagttccccg gccagcgggg 420

gataaacggg tgtggccatg cagccttta acggtgaact ggagttcccc gccagcgggg 480

ggataaacg cacgaactca gccagaacga caaacaaaag gcgagttccc cgcgccagcgc 540

gggataaac ggcaccagta cgcgccccac gctgacgggt tctgagttcc ccgccagcgc 600

ggggataaac cgcagctccc atttcaaac ccaggtaccc tgggcctcat gggccttccc 660

ctcactgccc gttttccag 679

<210> 9

<211> 685

<212> DNA

<213> Artificial Sequence

<220><223> GA1047360

<400> 9

gagctcccgg gctgacggta atagaggcac ctacaggctc cggtaaaacg gaaacagcgc 60

tggcctatgc ttggaaactt attgatcaac aaattgcgga tagtgttatt ttgcccctcc 120

caacacaagc taccgcaat gctatgctta cgagaatgga agcagcgcgc agccacttat 180

tttcatcccc aaatcttatt ctgtctcatg gcaattcag gtttaaccac ctctttcaat 240

caataaaatc acgcgcgatt actgaacagg ggcaagaaga agcgtgggtt cagtgttgtc 300

agtggttgtc acaaagcaat aagaaagtgt ttcttgggca aatcggcgtt tgcacgattg 360

atcaggtggt gatttcggta ttgccagtta aacaccgctt tatccgtggt ttgggaattg 420

gtagatctgt ttaattggtt aatgaagttc atgcttacga cacctatatg aacggcttgc 480

tcgaggcagt gctcaaggct caggctgatg tgggaggagg tgttattctt ctttccgcaa 540

ccctaccaat gaaacaaaaa cagaagcttc tggatactta tggctgcat acagatccag 600

tggaaaataa ctccgcatat ccactcatta actggcgagg tgtgaatggt gcgcaacgtt 660

ttgatctgct agcggatccg gtacc 685

<210> 10

<211> 37  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer BG3186  
 <400> 10  
 atagcgccat ggaacctttt aaatataat gccatta 37  
 <210> 11  
 <211> 41  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer BG3213  
 <400>  
 11  
 acagtgggat ccgctttggg atttgcaggg atgactctgg t 41  
 <210> 12  
 <211> 43  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer BG3303  
 <400> 12  
 atagcgtcat gaatttgctt attgataact ggattcctgt acg 43  
 <210> 13  
 <211> 44  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer BG3212  
 <400> 13  
 acagtggcgg ccgcgccatt tgatggcctt ccttgcggtt ttaa 44  
 <210> 14  
 <211> 45  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer BG3076  
 <400> 14  
 cgtatatcaa actttccaat agcatgaaga gcaatgaaaa ataac 45

<210> 15

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Primer BG3449

<400> 15

atgataccgc gagaccacg ctc 23

<210> 16

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Primer BG3451

<400> 16

cggataaagt tgcaggacca cttc 24

<210> 17

<211> 199

<212>

PRT

<213> Escherichia coli

<400> 17

Met Tyr Leu Ser Lys Val Ile Ile Ala Arg Ala Trp Ser Arg Asp Leu

1 5 10 15

Tyr Gln Leu His Gln Gly Leu Trp His Leu Phe Pro Asn Arg Pro Asp

20 25 30

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

35 40 45

Gly Cys His Val Leu Leu Gln Ser Ala Gln Met Pro Val Ser Thr Ala

50 55 60

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

65 70 75 80

Gly Val Pro Leu Tyr Phe Arg Leu Arg Ala Asn Pro Ile Lys Thr Ile

85 90 95

Leu Asp Asn Gln Lys Arg Leu Asp Ser Lys Gly Asn Ile Lys Arg Cys

100 105 110

Arg Val Pro Leu Ile Lys Glu Ala Glu Gln Ile Ala Trp Leu Gln Arg

115	120	125	
Lys Leu Gly Asn Ala Ala Arg Val Glu Asp Val His Pro Ile Ser Glu			
130	135	140	
Arg Pro Gln Tyr Phe Ser Gly Asp Gly Lys Ser Gly Lys Ile Gln Thr			
145	150	155	160
Val Cys Phe Glu Gly Val Leu Thr Ile Asn Asp Ala Pro Ala Leu Ile			
165	170	175	
Asp Leu Val Gln Gln Gly Ile Gly Pro Ala Lys Ser Met Gly Cys Gly			
180	185	190	
Leu Leu Ser Leu Ala Pro Leu			
195			
<210> 18			
<211> 2154			
<212> DNA			
<213> Artificial Sequence			
<220><223> Fusion protein			
<400> 18			
atggctcaac tggtaaag cgaactggaa gagaaaaaaaa gtgaactgcg ccacaaactg			60
aatatgtg cgcgatgaata tatcgagctg attgaaattg cacgtaatcc gaccaggat			120
cgtaattctgg aatgaaagt gatggaattt ttatgaaag tgtacggcta tcgcggtgaa			180
catctgggtg gtaccgtaa accggatggt gcaatttata ccgttgtag cccgattgat			240
tatggtgtta ttgtgatac caaagcctat agcgggtggtt ataactgcc gattggtcag			300
gcagatgaaa tggaaactgta tgtggaagaa aatcagaccc gtgataaaca tctgaatccg			360
aatgaatggt gaaagttaa tccgagcagc gttaccgagt taaattcct gttgttagc			420
ggtcacttca aaggcaacta taaagcacag ctgaccgctc tgaatcatat taccaattgt			480
aatggtgcag ttctgagcgt tgaagaactg ctgattggtg gtgaaatgat taaagcaggc			540
acctgacc tggagaagt tcgtcgcaa ttaacaatg gcgaaatcaa ctttcggat			600
cccccaacc gcgcgaaagg cctggaagcg gtgagcgtgg cgagcatgaa ttgcttatt			660
gataactgga ttctgtacg cccgcgaaac ggggggaaag tccaatcat aaatctgcaa			720
tcgctatact gcagtagaga tcagtggcga ttaagtttgc cccgtgacga tatggaactg			780
gccgctttag cactgctggt ttgcattggg caaattatcg ccccggcaa agatgacggt			840

gaatttcgac atcgcataat gaatccgctc actgaagatg agtttcaaca actcatcgcg 900  
 ccgtggatag atatgttcta ccttaatcac gcagaacatc cctttatgca gaccaaaggt 960  
 gtcaaagcaa atgatgtgac tccaatggaa aaactgttgg ctgggtaag cggcgcgacg 1020  
 aattgtgcat ttgtcaatca accggggcag ggtgaagcat tatgtggtgg atgcactgcg 1080  
 attgcgttat tcaaccaggc gaatcaggca ccaggttttg gtggtggttt taaaagcgg 1140

ttacgtggag gaacacctgt aacaacgttc gtacgtggga tcgatcttcg ttcaacggtg 1200  
 ttactcaatg tctcacatt acctcgtctt caaaaacaat ttctaatga atcacatacg 1260  
 gaaaaccaac ctacctggat taaacctatc aagtccaatg agtctatacc tgcttcgtea 1320  
 attgggtttg tccgtggtct attctggcaa ccagcgcata ttgaattatg cgatcccatt 1380  
 gggattggta aatgttcttg ctgtggacag gaaagcaatt tgcgttatac cggttttctt 1440  
 aaggaanaat ttacctttac agttaatggg ctatggcccc atccgcttc cccttctctg 1500  
 gtaacagtca agaaagggga ggttgaggaa aaatttcttg ctttcaccac ctccgcacca 1560

tcatggacac aaatcagccg agttgtggta gataagatta ttcaaatga aaatggaaat 1620  
 cgcgtggcgg cggttgtgaa tcaattcaga aatattgcgc cgcaaagtcc tcttgaattg 1680  
 attatggggg gatatcgtaa taatcaagca tctattcttg aacggcgtca tgatgtgttg 1740  
 atgtttaatc aggggtggca acaatcggc aatgtgataa acgaaatagt gactgttgg 1800  
 ttgggatata aaacagcctt acgcaaggcg ttatatacct ttgcagaagg gtttaanaat 1860  
 aaagacttca aaggggccgg agtctctgtt catgagactg cagaaaggca tttctatcga 1920  
 cagagtgaat tattaattcc cgatgtactg gcgaatgta attttcca ggctgatgag 1980

gtaatagctg atttacgaga caaacttcat caattgtgtg aaatgctatt taatcaatct 2040  
 gtagctccct atgcacatca tctaaatta ataagcacat tagcgcttgc ccgcgccacg 2100  
 ctatacaaac atttacggga gttaaaaccg caaggagggc catcaaatgg ctga 2154

<210> 19

<211> 717

<212> PRT

<213> Artificial Sequence

<220><223> Fusion protein

<400> 19

Met Ala 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 Pro 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 Glu 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 Glu Arg Tyr Val Glu Glu Asn Gln  
 100 105 110  
 Thr Arg Asp Lys His Leu 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 Ala Asp Pro Thr Asn Arg Ala Lys Gly Leu  
 195 200 205  
 Glu Ala Val Ser Val Ala Ser Met Asn Leu Leu Ile Asp Asn Trp Ile  
  
 210 215 220  
 Pro Val Arg Pro Arg Asn Gly Gly Lys Val Gln Ile Ile Asn Leu Gln  
 225 230 235 240  
 Ser Leu Tyr Cys Ser Arg Asp Gln Trp Arg Leu Ser Leu Pro Arg Asp  
 245 250 255  
 Asp Met Glu Leu Ala Ala Leu Ala Leu Leu Val Cys Ile Gly Gln Ile  
 260 265 270  
 Ile Ala Pro Ala Lys Asp Asp Val Glu Phe Arg His Arg Ile Met Asn



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

530 535 540

Val Val Asn Gln Phe Arg Asn Ile Ala Pro Gln Ser Pro Leu Glu Leu

545 550 555 560

Ile Met Gly Gly Tyr Arg Asn Asn Gln Ala Ser Ile Leu Glu Arg Arg

565 570 575

His Asp Val Leu Met Phe Asn Gln Gly Trp Gln Gln Tyr Gly Asn Val

580 585 590

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

595 600 605

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

610 615 620

Gly Ala Gly Val Ser Val His Glu Thr Ala Glu Arg His Phe Tyr Arg

625 630 635 640

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

645 650 655

Gln Ala Asp Glu Val Ile Ala Asp Leu Arg Asp Lys Leu His Gln Leu

660 665 670

Cys Glu Met Leu Phe Asn Gln Ser Val Ala Pro Tyr Ala His His Pro

675 680 685

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

690 695 700

Leu Arg Glu Leu Lys Pro Gln Gly Gly Pro Ser Asn Gly

705 710 715

<210> 20

<211> 2154

<212> DNA

<213> Artificial Sequence

<220

><223> Fusion protein

<400> 20

atggctcaac tggttaaaag cgaactggaa gagaaaaaaa gtgaactgcg ccacaaactg

aaatatgtgc cgcatgaata tatcgagctg attgaaattg cacgtaatcc gacccaggat 120  
 cgtattctgg aaatgaaagt gatggaattt tttatgaaag tgtacggcta tcgcggtgaa 180  
 catctgggtg gtagccgtaa accggatggt gcaatttata ccgttggtag cccgattgat 240  
 tatggtgtta ttgtgatac caaacctat agcgggtggtt ataatctgcc gattggtcag 300  
 gcagatgaaa tgcagcgtta tgtgaaagaa aatcagacct gcaacaaaca tattaaccg 360  
 aatgaatggt ggaagttta tccgagcagc gttaccgagt ttaaattcct gtttgttagc 420  
  
 ggtcacttca aaggcaacta taaagcacag ctgaccctgc tgaatcgtaa aaccaattgt 480  
 aatggtgcag ttctgagcgt tgaagaactg ctgattggtg gtgaaatgat taaagcaggc 540  
 accctgacct tggaagaagt tcgtcgcaaa ttaacaatg gcgaaatcaa ctttgcggat 600  
 cccaccaacc gcgcgaaagg cctggaagcg gtgagcgtgg cgagcatgaa ttgcttatt 660  
 gataactgga ttctgtacg cccgcgaaac ggggggaaag tccaaatcat aaatctgcaa 720  
 tcgctatact gcagtagaga tcagtggcga ttaagtttgc cccgtgacga tatggaactg 780  
 gccgctttag cactgctggt ttgcattggg caaattatcg ccccggcaaa agatgacgtt 840  
  
 gaatttcgac atcgcataat gaatccgctc actgaagatg agtttcaaca actcatcgcg 900  
 ccgtggatag atatgttcta ccttaatcac gcagaacatc cttttatgca gaccaaaggt 960  
 gtcaaagcaa atgatgtgac tccaatggaa aaactgttgg ctgggtaag cggcgcgacg 1020  
 aattgtgcat ttgtcaatca accggggcag ggtgaagcat tatgtggtgg atgcactgcg 1080  
 attgcgttat tcaaccaggc gaatcaggca ccaggttttg gtggtggttt taaaagcgg 1140  
 ttacgtggag gaacacctgt aacaacgttc gtacgtggga tcgatcttcg ttcaacggtg 1200  
 ttactcaatg tctcaccatt acctcgtctt caaaaacaat ttctaatga atcacatag 1260  
  
 gaaaaccaac ctacctggat taaacctatc aagtccaatg agtctatacc tgcttcgtca 1320  
 attgggtttg tccgtggtct attctggcaa ccagcgcata ttgaattatg cgatccatt 1380  
 gggattggtg aatgttcttg ctgtggacag gaaagcaatt tgcgttatac cggttttctt 1440  
 aagaaaaat ttaccttac agttaatggg ctatggcccc atccgattc cccttgtctg 1500  
 gtaacagtca agaaagggga ggttgaggaa aaatttcttg ctttcaccac ctccgacca 1560  
 tcatggacac aaatcagccg agtttgggia gataagatta ttcaaatga aaatggaaat 1620  
 cgctggcgg cggttgtgaa tcaattcaga aatattgccc cgcaaagtcc tcttgaattg 1680  
  
 attatggggg gatatcgtaa taatcaagca tctattcttg aacggcgtca tgatgtgtt 1740  
 atgtttaate aggggtggca acaatcggc aatgtgataa acgaaatagt gactgttgt 1800  
 ttgggatata aaacagcctt acgcaaggcg ttatatacct ttgcagaagg gtttaaaat 1860  
 aaagacttca aaggggccgg agtctctgtt catgagactg cagaaaggca tttctatcga 1920

cagagtgaat tattaattcc cgatgtactg gcgaatgtta atttttccca ggctgatgag 1980  
 gtaatagctg atttacgaga caaacttcat caattgtgtg aaatgctatt taatcaatct 2040  
 gtagctccct atgcacatca tcctaaatta ataagcacaat tagcgcttgc cgcgcccacg 2100

ctatacaaac atttacggga gttaaaaccg caaggagggc catcaaatgg ctga 2154

<210> 21

<211> 717

<212> PRT

<213> Artificial Sequence

<220><223> Fusion protein

<400> 21

Met Ala 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 Pro 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 Glu 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 Lys 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 Arg Lys Thr Asn Cys

145                    150                    155                    160

Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met





Cys Glu Met Leu Phe Asn Gln Ser Val Ala Pro Tyr Ala His His Pro

675 680 685  
 Lys Leu Ile Ser Thr Leu Ala Leu Ala Arg Ala Thr Leu Tyr Lys His

690 695 700  
 Leu Arg Glu Leu Lys Pro Gln Gly Gly Pro Ser Asn Gly

705 710 715

<210> 22

<211> 2235

<212> DNA

<213> Artificial Sequence

<220><223> Fusion protein

<400> 22

atgcatcacc atcatcacca cccgaaaaaa aagcgcaaag tggatccgaa gaaaaaacgt 60

aaagtgaag atccgaaaga catggctcaa ctggtaaaa gcgaactgga agagaaaaaa 120

agtgaactgc gccacaaact gaaatatgtg ccgcatgaat atatcgagct gattgaaatt 180

gcacgtaatc cgaccacgga tcgtattctg gaaatgaaag tgatggaatt ttttatgaaa 240

gtgtacggct atcgcggtga acatctgggt ggtagccgta aaccggatgg tgcaatttat 300

accgttggta gcccgattga ttatggtggt attgttgata ccaaagccta tagcgggtggt 360

tataatctgc cgattgggtca ggcagatgaa atgcagcgtt atgtgaaaga aatcagacc 420

cgcaacaaac atattaaccc gaatgaatgg tggaaaagttt atccgagcag cgttaccgag 480

tttaaattcc tgtttgttag cggtcacttc aaaggcaact ataaagcaca gctgaccctg 540

ctgaatcgta aaaccaattg taatggtgca gttctgagcg ttgaagaact gctgattggt 600

ggtgaaatga ttaaagcagg caccctgacc ctggaagaag ttcgtcgcaa atttaacaat 660

ggcgaatca actttgcgga tcccaccaac cgcgcgaaag gcctggaagc ggtgagcgtg 720

gcgagcatga atttgcttat tgataactgg attcctgtac gcccgcgaaa cgggggggaaa 780

gtccaaatca taaatctgca atcgtatac tgcagtagag atcagtggcg attaagtttg 840

ccccgtgacg atatggaact ggccgcttta gcaactgctgg tttgcattgg gcaaattatc 900

gccccggcaa aagatgacgt tgaatttga catcgcataa tgaatccgct cactgaagat 960

gagtttcaac aactcatcgc gccgtggata gatatgttct accttaatca cgcagaacat 1020

ccctttatgc agaccaaagg tgtcaaagca aatgatgtga ctccaatgga aaaactgttg 1080

gctggggtaa gcggcgcgac gaattgtgca tttgtcaatc aaccggggca ggtggaagca 1140

ttatgtggtg gatgcactgc gattgcgtta ttcaaccagg cgaatcaggc accaggtttt 1200  
 ggtgggtggtt ttaaaagcgg tttactgga ggaacacctg taacaacgtt cgtacgtggg 1260  
 atcgatcttc gttcaacggt gttactcaat gtccctecat tacctcgtct tcaaaaacaa 1320  
 tttcctaag aatcacatac ggaaaaccaa cctacctgga ttaaacctat caagtccaat 1380

gagtctatac ctgcttcgtc aattgggttt gtcctgggtc tattctggca accagcgcac 1440  
 attgaattat gcgatcccat tgggattggt aaatgttctt gctgtggaca ggaaagcaat 1500  
 ttgcgttata ccggttttct taaggaaaaa ttaccttta cagttaatgg gctatggccc 1560  
 catccgcatt ccccttgtct ggtaacagtc aagaaagggg aggttgagga aaaatttctt 1620  
 gctttcacca cctccgacc atcatggaca caaatcagcc gagtgtggt agataagatt 1680  
 attcaaatg aaaaaggaaa tcgcgtggcg gcggttga atcaattcag aaatattgcg 1740  
 ccgcaaagtc ctcttgaatt gattatgggg ggatatcgta ataacaagc atctattctt 1800

gaacggcgtc atgatgtgtt gatgtttaat caggggtggc aacaatacgg caatgtgata 1860  
 aacgaaatag tgactgttgg tttgggatat aaaacagcct tacgcaaggc gttatatacc 1920  
 tttgcagaag ggtttaaaaa taaagacttc aaagggccg gagtctctgt tcatgagact 1980  
 gcagaaagge atttctatcg acagagtga ttattaattc ccgatgtact ggcaatgtt 2040  
 aatthttccc aggctgatga ggtaatagct gatttacgag acaaacttca tcaattgtgt 2100  
 gaaatgctat ttaatcaatc tgtagctccc tatgcacatc atcctaaatt aataagcaca 2160  
 ttagcgcttg cccgcgccac gctatacaaa catttacggg agttaaacc gcaaggaggg 2220

ccatcaaatg gctga 2235

<210> 23

<211> 744

<212> PRT

<213> Artificial Sequence

<220><223> Fusion protein

<400> 23

Met His His His His His His Pro Lys Lys Lys Arg Lys Val Asp Pro

1 5 10 15

Lys Lys Lys Arg Lys Val Glu Asp Pro Lys Asp Met Ala Gln Leu Val

20 25 30

Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys

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





atgcatcacc atcatcacca cccgaaaaaa aagcgcaaag tggatccgaa gaaaaaacgt 60  
 aaagtgaag atccgaaaga catggctcaa ctggttaaaa gcgaactgga agagaaaaaa 120  
 agtgaactgc gccacaaact gaaatatgtg ccgcatgaat atatcgagct gattgaaatt 180  
 gcacgtaatc cgaccaggga tcgtattctg gaaatgaaag tgatggaatt ttttatgaaa 240  
 gtgtacggct atcgcggtga acatctgggt ggtagccgta aaccggatgg tgcaatttat 300  
 accgttggta gcccgattga ttatggtggt attgttgata ccaaagccta tagcgggtggt 360  
  
 tataatctgc cgattggica ggcagatgaa atggaacgtt atgtggaaga aaatcagacc 420  
 cgtgataaac atctgaatcc gaatgaatgg tggaaagttt atccgagcag cgttaccgag 480  
 tttaaattcc tgtttgtag cggtcacttc aaaggcaact ataaagcaca gctgaccctg 540  
 ctgaatcata ttaccaatg taatggtgca gttctgagcg ttgaagaact gctgattggt 600  
 ggtgaaatga ttaaagcagg caccctgacc ctggaagaag ttcgtcgcaa atttaacaat 660  
 ggcgaaatca actttcgga tcccaccaac cgcgcgaaag gcctggaagc ggtgagcgtg 720  
 gcgagcatga atttgcttat tgataactgg attcctgtac gcccgcgaaa cggggggaaa 780  
  
 gtccaaatca taaatctgca atcgctatac tgcaagtagag atcagtggcg attaagtttg 840  
 ccccgtagc atatggaact ggccgcttta gcaactgctgg tttgcatgga gcaaattatc 900  
 gccccggcaa aagatgacgt tgaatttca catcgcataa tgaatccgct cactgaagat 960  
 gagtttcaac aactcatcgc gccgtggata gatatgttct acctaatca cgcagaacat 1020  
 ccctttatgc agaccaaagg tgtcaaagca aatgatgtga ctccaatgga aaaactgttg 1080  
 gctggggtaa gcggcgcgac gaattgtgca tttgtcaatc aaccggggca gggatgaagca 1140  
 ttatgtggtg gatgactgc gattgcgtta ttcaaccagg cgaatcagc accaggtttt 1200  
  
 ggtggtggtt taaaagcgg tttactgga ggaacacctg taacaacgtt cgtactggg 1260  
 atcgatcttc gttcaacggt gttactcaat gtcctcacat tacctcgtct tcaaaaacia 1320  
 tttcctaatg aatcacatac ggaaaaccaa cctacctgga ttaaacctat caagtccaat 1380  
 gagtctatac ctgcttcgtc aattgggttt gtcctggtc tattctggca accagcgcac 1440  
 attgaattat gcgatccat tgggattggt aatgttctt gctgtggaca ggaaagcaat 1500  
 ttgcgttata ccggttttct taaggaaaaa tttacctta cagttaatgg gctatggccc 1560  
 catccgatt cccttgtct ggtaacagtc aagaaagggg aggttgagga aaaatttctt 1620  
  
 gctttacca cctccgacc atcatggaca caaatcagcc gagttgtggt agataagatt 1680  
 attcaaatg aaaaaggaaa tcgctggcg gcggttgtga atcaattcag aatatgacg 1740  
 ccgcaaagtc ctcttgaatt gattatgggg ggatatcgta ataacaagc atctattctt 1800  
 gaacggcgtc atgatgtgtt gatgtttaat caggggtggc aacaatcgg caatgtgata 1860

aacgaaatag tgactgttgg tttgggatat aaaacagcct tacgcaaggc gttatatacc 1920  
 ttgcagaag ggtttaaaaa taaagacttc aaagggcccg gagtctctgt tcatgagact 1980  
 gcagaaaggc atttctatcg acagagtga tattaattc ccgatgtact ggcgaatggt 2040

aatTTTTccc aggctgatga ggtaatagct gatttacgag acaaacttca tcaattgtgt 2100  
 gaaatgctat ttaatcaatc tgtagctccc tatgcacatc atcctaaatt aataagcaca 2160  
 ttagcgcttg cccgcgccac gctatacaaa catttacggg agttaaacc gcaaggaggg 2220  
 ccatcaaatg gctga 2235

<210> 25

<211> 744

<212> PRT

<213> Artificial Sequence

<220><223> Fusion protein

<400> 25

Met His His His His His His Pro Lys Lys Lys Arg Lys Val Asp Pro

1 5 10 15

Lys Lys Lys Arg Lys Val Glu Asp Pro Lys Asp Met Ala Gln Leu Val

20 25 30

Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys

35 40 45

Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Pro

50 55 60

Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys

65 70 75 80

Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly Ser Arg Lys Pro Asp

85 90 95

Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val

100 105 110

Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala

115 120 125

Asp Glu Met Glu Arg Tyr Val Glu Glu Asn Gln Thr Arg Asp Lys His

130 135 140

Leu Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu  
 145                      150                      155                      160  
 Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala  
                                  165                      170                      175  
 Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu  
                                  180                      185                      190  
 Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr  
                                  195                      200                      205  
  
 Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn  
                                  210                      215                      220  
 Phe Ala Asp Pro Thr Asn Arg Ala Lys Gly Leu Glu Ala Val Ser Val  
 225                      230                      235                      240  
 Ala Ser Met Asn Leu Leu Ile Asp Asn Trp Ile Pro Val Arg Pro Arg  
                                  245                      250                      255  
 Asn Gly Gly Lys Val Gln Ile Ile Asn Leu Gln Ser Leu Tyr Cys Ser  
                                  260                      265                      270  
  
 Arg Asp Gln Trp Arg Leu Ser Leu Pro Arg Asp Asp Met Glu Leu Ala  
                                  275                      280                      285  
 Ala Leu Ala Leu Leu Val Cys Ile Gly Gln Ile Ile Ala Pro Ala Lys  
                                  290                      295                      300  
 Asp Asp Val Glu Phe Arg His Arg Ile Met Asn Pro Leu Thr Glu Asp  
 305                      310                      315                      320  
 Glu Phe Gln Gln Leu Ile Ala Pro Trp Ile Asp Met Phe Tyr Leu Asn  
                                  325                      330                      335  
  
 His Ala Glu His Pro Phe Met Gln Thr Lys Gly Val Lys Ala Asn Asp  
                                  340                      345                      350  
 Val Thr Pro Met Glu Lys Leu Leu Ala Gly Val Ser Gly Ala Thr Asn  
                                  355                      360                      365  
 Cys Ala Phe Val Asn Gln Pro Gly Gln Gly Glu Ala Leu Cys Gly Gly  
                                  370                      375                      380  
 Cys Thr Ala Ile Ala Leu Phe Asn Gln Ala Asn Gln Ala Pro Gly Phe  
 385                      390                      395                      400

Gly Gly Gly Phe Lys Ser Gly Leu Arg Gly Gly Thr Pro Val Thr Thr  
 405 410 415  
 Phe Val Arg Gly Ile Asp Leu Arg Ser Thr Val Leu Leu Asn Val Leu  
 420 425 430  
 Thr Leu Pro Arg Leu Gln Lys Gln Phe Pro Asn Glu Ser His Thr Glu  
 435 440 445  
 Asn Gln Pro Thr Trp Ile Lys Pro Ile Lys Ser Asn Glu Ser Ile Pro  
 450 455 460  
  
 Ala Ser Ser Ile Gly Phe Val Arg Gly Leu Phe Trp Gln Pro Ala His  
 465 470 475 480  
 Ile Glu Leu Cys Asp Pro Ile Gly Ile Gly Lys Cys Ser Cys Cys Gly  
 485 490 495  
 Gln Glu Ser Asn Leu Arg Tyr Thr Gly Phe Leu Lys Glu Lys Phe Thr  
 500 505 510  
 Phe Thr Val Asn Gly Leu Trp Pro His Pro His Ser Pro Cys Leu Val  
 515 520 525  
  
 Thr Val Lys Lys Gly Glu Val Glu Glu Lys Phe Leu Ala Phe Thr Thr  
 530 535 540  
 Ser Ala Pro Ser Trp Thr Gln Ile Ser Arg Val Val Val Asp Lys Ile  
 545 550 555 560  
 Ile Gln Asn Glu Asn Gly Asn Arg Val Ala Ala Val Val Asn Gln Phe  
 565 570 575  
 Arg Asn Ile Ala Pro Gln Ser Pro Leu Glu Leu Ile Met Gly Gly Tyr  
 580 585 590  
  
 Arg Asn Asn Gln Ala Ser Ile Leu Glu Arg Arg His Asp Val Leu Met  
 595 600 605  
 Phe Asn Gln Gly Trp Gln Gln Tyr Gly Asn Val Ile Asn Glu Ile Val  
 610 615 620  
 Thr Val Gly Leu Gly Tyr Lys Thr Ala Leu Arg Lys Ala Leu Tyr Thr  
 625 630 635 640  
 Phe Ala Glu Gly Phe Lys Asn Lys Asp Phe Lys Gly Ala Gly Val Ser

645 650 655

Val His Glu Thr Ala Glu Arg His Phe Tyr Arg Gln Ser Glu Leu Leu

660 665 670

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

675 680 685

Ile Ala Asp Leu Arg Asp Lys Leu His Gln Leu Cys Glu Met Leu Phe

690 695 700

Asn Gln Ser Val Ala Pro Tyr Ala His His Pro Lys Leu Ile Ser Thr

705 710 715 720

Leu Ala Leu Ala Arg Ala Thr Leu Tyr Lys His Leu Arg Glu Leu Lys

725 730 735

Pro Gln Gly Gly Pro Ser Asn Gly

740

<210> 26

<211> 168

<212> DNA

<213> Artificial Sequence

<220><223> Target plasmid

<400> 26

gaattcacaacggtgagcaagtcactgttg gcaagccagg atctgaacaa taccgtcttg 60

ctttcgagcgtagctctag aactagtcct cagcctaggc ctcgttccga agctgtcttt 120

cgctgctgag ggtgacgatc ccgcataggc ggctttaaac tcggatcc 168

<210> 27

<211> 163

<212> DNA

<213> Artificial Sequence

<220><223> Target plasmid

<400> 27

gaattcacaacggtgagcaagtcactgttg gcaagccagg atctgaacaa taccgtcttt 60

tcgagcgcta gctctagaac tagtcctcag cctaggcctc gttcaagctg tctttcgctg 120

ctgagggatga cgatcccga taggcggcct ttaactcgga tcc 163

<210> 28

<211> 158  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Target plasmid  
 <400> 28  
 gaattcaca cggtagca gtcactgtt gcaagccagg atctgaaca taccgtcttc 60  
 gagcgctagc tctagaacta gtcctcagcc taggcctcga agctgtcttt cgctgctgag 120  
 ggtgacgate ccgcatagc ggcctttaac tcggatcc 158  
 <210> 29  
 <211> 153  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Target plasmid  
 <400> 29  
 gaattcaca cggtagca gtcactgtt gcaagccagg atctgaaca taccgtcttg 60  
 cgctagctct agaactagtc ctcagcctag gctaagctg tctttcgctg ctgagggtga 120  
 cgatcccga taggcggcct ttaactcgga tcc 153  
 <210> 30  
 <211> 148  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Target plasmid  
 <400> 30  
 gaattcaca cggtagca gtcactgtt gcaagccagg atctgaaca taccgtcttg 60  
 ctagctctag aactagtctt cagcctagga agctgtcttt cgctgctgag ggtgacgatc 120  
 ccgcatagc ggcctttaac tcggatcc 148  
 <210> 31  
 <211> 143  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Target plasmid  
 <400> 31  
 gaattcaca cggtagca gtcactgtt gcaagccagg atctgaaca taccgtcttc 60  
 tctagaacta gtcctcagcc taggaagctg tctttcgctg ctgagggtga cgatcccga 120

taggcggcct ttaactcgga tcc 143

<210> 32

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Cleavage product

<400> 32

cttgcgctag ctctagaact agtcctcagc ctaggcctaa g 41

<210> 33

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Cleavage product

<400> 33

cttaggccta ggctgaggac tagttctaga gctagcgcaa g 41

<210> 34

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Fill-in ligation

<400> 34

cttgcgctag ctctagaact agctagtctc cagcctaggc ctaag 45

<210> 35

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Fill-in ligation

<400> 35

cttaggccta ggctgaggac tagctagttc tagagctagc gcaag 45

<210> 36

<211> 1100

<212> DNA

<213> Homo sapiens

<400> 36

ggtggaacaa gatggattat caagtgtcaa gtccaatcta tgacatcaat tattatacat 60  
 cggagccctg ccaaaaaatc aatgtgaagc aaatcgcagc ccgcctctctg cctccgctct 120  
  
 actcactggt gttcatcttt ggTTTTgtgg gcaacatgct ggtcaccctc atcctgataa 180  
 actgcaaaaag gctgaagagc atgactgaca tctacctgct caacctggcc atctctgacc 240  
 tgttttttct tcttactgtc ccttctctggg ctcaactatgc tgccgccagc tgggactttg 300  
 gaaatacaat gtgtcaactc ttgacagggc tctatTTTTat aggcttcttc tctggaatct 360  
 tcttcatcat cctcctgaca atcgataggi acctggctgt cgtccatgct gtgtttgctt 420  
 taaaagccag gacggtcacc tttgggggtg tgacaagtgt gatcacttgg gtgggtggctg 480  
 tgtttgcgtc tctcccagga atcatcttta ccagatctca aaaagaaggt cttcattaca 540  
  
 cctgcagctc tcattttcca tacagtcagt atcaattctg gaagaatttc cagacattaa 600  
 agatagtcat ctTggggctg gtcctgccgc tgcttTgcat ggtcatctgc tactcgggaa 660  
 tcctaaaaac tctgcttcgg tgTcgaaatg agaagaagag gcacagggct gtgaggctta 720  
 tcttcacat catgattgtt tattttctct tctgggctcc ctacaacatt gtccttctcc 780  
 tgaacacctt ccaggaattc tttggcctga ataattgcag tagctctaac aggttggacc 840  
 aagctatgca ggtgacagag actctTggga tgacgcaact ctgcatcaac cccatcatct 900  
 atgcctttgt cggggagaag ttcagaaact acctcttagt cttcttccaa aagcacattg 960  
  
 ccaaacgctt ctgcaaatgc tgttctattt tccagcaaga ggctcccagc cgagcaagct 1020  
 cagtttacac ccgatccact ggggagcagg aaatatctgt gggcttTgta cacggactca 1080  
 agtgggctgg tgaccagtc 1100  
  
 <210> 37  
 <211> 424  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> CRISPR array red1  
 <400> 37  
  
 ccatgtaat acgactcact ataggagaa ttagctgac ttaataata aggaaatgtt 60  
 acattaaggt tggTgggttg tttttatggg aaaaaatgct ttaagaaca atgtatactt 120  
 ttagagagtt ccccgcgcca gcggggataa accgcaaca cagcatggac gacagccagg 180  
  
 tacctagagt tccccgcc agcggggata aaccgcaac acagcatgga cgacagccag 240  
 gtacctagag tccccgcgc cagcggggat aaaccgcaa cacagcatgg acgacagcca 300  
 ggtacctaga gttccccgcg ccagcgggga taaaccgaaa acaaaaggct cagtcggaag 360

actgggcctt ttgtttaac cccttggggc ctctaaacgg gtcttgaggg gtttttggg 420  
tacc 424  
<210> 38  
<211> 424  
<212> DNA  
<213> Artificial Sequence  
<220><223> CRISPR array red2  
<400> 38  
ccatgtaat acgactcact atagggagaa ttagctgatc ttaataata aggaaatgtt 60  
  
acattaaggt tggtaggttg tttttatggg aaaaaatgct ttaagaaca atgtatactt 120  
ttagagagtt ccccgcgcca gcggggataa accgtgtgat cacttgggtg gtggctgtgt 180  
ttgcgtgagt tcccgcgcc agcggggata aaccgtgtga tcacttgggt ggtggctgtg 240  
tttgcgtgag ttcccgcgc cagcggggat aaaccgtgtg atcacttggg tggtagctgt 300  
gtttgcgtga gttcccgcg ccagcgggga taaaccgaaa acaaaaggct cagtcggaag 360  
actgggcctt ttgtttaac cccttggggc ctctaaacgg gtcttgaggg gtttttggg 420  
tacc 424  
  
<210> 39  
<211> 43  
<212> DNA  
<213> Artificial Sequence  
<220><223> plasmid  
<400> 39  
aaggatgcca gtgataagtg gaatgcatg tgggctgtca aaa 43  
<210> 40  
<211> 43  
<212> DNA  
<213> Artificial Sequence  
<220><223> plasmid mutant 1  
<400> 40  
aaggatgcca gtgataagtg gaatgcatg tgggctgtca aaa 43  
<210> 41  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220><223> plasmid mutant 2

<400> 41

gccatgtggg ctgtcaaaa

19

<210> 42

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> plasmid mutant 3

<400> 42

gaatgccatg tgggctgtca aaa

23

<210> 43

<211> 1604

<212> PRT

<213> Streptomyces sp. SPB78

<400> 43

Met Pro Asp Gln Leu Asn Ala Pro Thr Pro Leu Gly Asp Arg Leu Thr

1                    5                    10                    15

Gly Ala Val Arg Thr Val Trp Ala Lys His Asp Arg Asp Thr Gly Lys

                  20                    25                    30

Trp Leu Pro Leu Trp Arg His Met Thr Asp Ser Ala Ala Val Ala Gly

                  35                    40                    45

Leu Leu Trp Asp His Trp Leu Pro Arg Asn Ile Lys Asp Leu Ile Ala

                  50                    55                    60

Glu Pro Leu Pro Gly Gly Val Ala Asp Ala Arg Ser Leu Cys Val Trp

65                    70                    75                    80

Leu Ala Gly Thr His Asp Ile Gly Lys Ala Thr Pro Ala Phe Ala Cys

                  85                    90                    95

Gln Val Asp Glu Leu Ala Gly Val Met Thr Ala Ala Gly Leu Asp Met

                  100                    105                    110

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

                  115                    120                    125

Leu Ala Gly Gln Val Leu Leu Gln Glu Trp Leu Glu Glu Arg Arg Gly



Arg Ser Val Leu Leu Ala His Ala Lys Ser Ala Leu Gln Glu Asp Tyr  
 385                    390                    395                    400  
 Ala Thr Leu Met Arg Glu Ser His Arg Thr Ile Ala Ala Val Asp Ala  
                          405                    410                    415  
  
 Tyr Gly Asp Asp Ser Arg Pro Arg Lys Gly Arg Pro Ala Ala Asp Gly  
                          420                    425                    430  
 Val Arg Arg Lys Ala Pro Ala Glu Leu Val Ala His Gln Trp Leu Arg  
                          435                    440                    445  
 Gly Arg Lys Lys Gly Leu Leu Ala Ser Phe Ala Val Gly Thr Ile Asp  
                          450                    455                    460  
 Gln Leu Leu Met Ala Gly Leu Lys Ser Arg His Leu Ala Leu Arg His  
 465                    470                    475                    480  
  
 Leu Ala Met Ala Gly Lys Val Val Val Ile Asp Glu Val His Ala Tyr  
                          485                    490                    495  
 Asp Thr Tyr Met Asn Ala Tyr Leu Asp Arg Val Leu Ala Trp Leu Gly  
                          500                    505                    510  
 Glu Tyr Arg Val Pro Val Val Val Leu Ser Ala Thr Leu Pro Ala Arg  
                          515                    520                    525  
 Arg Arg Gly Glu Leu Ala Ala Ala Tyr Thr Gly Glu Asp Ala Gln Ala  
                          530                    535                    540  
  
 Leu Thr Glu Ala Thr Gly Tyr Pro Leu Leu Thr Ala Val Val Pro Gly  
 545                    550                    555                    560  
 Arg Glu Ala Val Gln Phe Val Ala Ala Ala Ser Gly Arg Gly Ser Asp  
                          565                    570                    575  
 Val Leu Leu Glu Lys Leu Asp Asp Asp Asp Glu Ala Leu Ala Asp Arg  
                          580                    585                    590  
 Leu Asp Thr Asp Leu Ala Asp Gly Gly Cys Ala Leu Val Val Arg Asn  
                          595                    600                    605  
  
 Thr Val Asp Arg Val Met Asp Thr Ala Ser Val Leu Arg Glu Arg Phe  
                          610                    615                    620  
 Gly Ala Asp His Val Thr Val Ala His Ala Arg Phe Val Asp Leu Asp

625                      630                      635                      640  
 Arg Ala Arg Lys Asp Ser Glu Leu Leu Ala Arg Phe Gly Pro Pro Asp  
    645                      650                      655  
 Pro Asp Gly Gly Ser Pro Gln Arg Pro Arg Asn Ala His Ile Val Val  
    660                      665                      670  
  
 Ala Ser Gln Val Ala Glu Gln Ser Leu Asp Val Asp Phe Asp Leu Leu  
    675                      680                      685  
 Val Ser Asp Leu Cys Pro Val Asp Leu Leu Leu Gln Arg Met Gly Arg  
    690                      695                      700  
 Leu His Arg His Pro Arg Gly Arg Asp Gln Glu Arg Arg Pro Ala Arg  
 705                      710                      715                      720  
 Leu Arg Gln Ala Arg Cys Leu Val Thr Gly Val Gly Trp Asp Thr Ser  
    725                      730                      735  
  
 Pro Ala Pro Glu Ala Asp Glu Gly Ser Arg Ala Ile Tyr Gly Ala Tyr  
    740                      745                      750  
 Ser Leu Leu Arg Ser Leu Ala Val Leu Ala Pro His Leu Gly Thr Ala  
    755                      760                      765  
 Gly Ala Ala Gly His Pro Leu Arg Leu Pro Glu Asp Ile Ser Pro Leu  
    770                      775                      780  
 Val Arg Arg Ala Tyr Gly Glu Glu Asp Pro Cys Pro Pro Glu Trp Glu  
 785                      790                      795                      800  
  
 Pro Val Leu Ala Pro Ala Arg Asp Lys Tyr Arg Thr Ala Arg Glu Arg  
    805                      810                      815  
 Gln Ser Gln Lys Ala Glu Val Phe Arg Leu Asp Glu Val Arg Lys Ala  
    820                      825                      830  
 Gly Arg Pro Leu Ile Gly Trp Ile Asp Ala Gly Val Gly Asp Ala Asp  
    835                      840                      845  
 Asp Thr Pro Val Gly Arg Ala Gln Val Arg Asp Thr Lys Glu Gly Leu  
    850                      855                      860  
  
 Glu Val Leu Val Val Arg Arg Arg Ala Asp Gly Ser Leu Cys Thr Leu  
 865                      870                      875                      880

Pro Trp Leu Asp Lys Gly Arg Gly Gly Leu Glu Leu Pro Val Asp Ala  
                   885                  890                  895  
 Val Pro Ser Ala Leu Ala Ala Arg Ala Val Ala Ala Ser Gly Leu Arg  
                   900                  905                  910  
 Leu Pro Tyr His Phe Thr Ser Ser Pro Gln Thr Leu Asp Arg Thr Leu  
                   915                  920                  925  
  
 Ala Glu Leu Glu Glu Leu Tyr Val Pro Ala Trp Gln Glu Lys Glu Ser  
                   930                  935                  940  
 His Trp Ile Ala Gly Glu Leu Ile Leu Ala Leu Asp Glu Glu Gly Arg  
 945                  950                  955                  960  
 Ala Ala Leu Ala Gly Gln Gln Leu Val Tyr Asn Pro Glu Glu Gly Leu  
                   965                  970                  975  
 Leu Val Ala Ser Ala Asp Ala Asn Thr Glu Ala Thr Ser Gly Arg Val  
                   980                  985                  990  
  
 Met Asp Gly Lys Pro Ser Ser Ala Gly Asp Gly Lys Pro Gly His Ala  
                   995                  1000                  1005  
 Ala Asp Gly Asn Arg Ala Arg Thr Thr Val Gly Gln Ser Pro Ala  
                   1010                  1015                  1020  
 Asp Arg Gln Thr His Gln Pro Pro Glu Gly Glu Arg His Pro Val  
                   1025                  1030                  1035  
 Pro Pro Ser Ala Ala Pro Pro Pro Ala Arg Pro Ser Phe Asp Leu  
                   1040                  1045                  1050  
  
 Thr Ser Arg Pro Trp Leu Pro Val Leu Leu Lys Asp Gly Ser Glu  
                   1055                  1060                  1065  
 Arg Glu Leu Ser Leu Pro Glu Val Phe Asp Gln Ala Arg Asp Ile  
                   1070                  1075                  1080  
 Arg Arg Leu Val Gly Asp Leu Pro Thr Gln Asp Phe Ala Leu Thr  
                   1085                  1090                  1095  
 Arg Met Leu Leu Ala Leu Leu Tyr Asp Ala Leu Ser Glu Pro Gly  
                   1100                  1105                  1110  
  
 Gly Asp Met Ala Pro Ala Asp Thr Asp Ala Trp Glu Glu Leu Trp

1115	1120	1125
Leu Ser Gln Ser Ala Tyr Ala	Ala Pro Val Ala Ala	Tyr Leu His
1130	1135	1140
Arg Tyr Arg Glu Arg Phe Asp	Leu Leu His Pro Glu	Ser Pro Phe
1145	1150	1155
Phe Gln Thr Pro Gly Leu Arg	Thr Ala Lys Asn Glu	Val Phe Ser
1160	1165	1170
Leu Asn Arg Leu Val Ala Asp	Val Pro Asn Gly Asp	Pro Phe Phe
1175	1180	1185
Ser Met Arg Arg Pro Gly Val	Asp Arg Leu Gly Phe	Ala Glu Ala
1190	1195	1200
Ala Arg Trp Leu Val His Ala	Gln Ala Tyr Asp Thr	Ser Gly Ile
1205	1210	1215
Lys Thr Gly Ala Val Gly Asp	Pro Arg Val Lys Ala	Gly Lys Gly
1220	1225	1230
Tyr Pro Gln Gly Pro Ala Trp	Ala Gly Asn Leu Gly	Gly Val Leu
1235	1240	1245
Leu Glu Gly Asp Asn Leu His	Glu Thr Leu Leu Leu	Asn Leu Ile
1250	1255	1260
Ala Gly Asp Thr Pro Gly Val	His Ala Ala Glu Val	Asp Arg Pro
1265	1270	1275
Ala Trp Arg Ala Glu Pro Ser	Gly Pro Ala Pro Ala	Pro Asp Leu
1280	1285	1290
Gly Leu Arg Pro Tyr Gly Leu	Arg Asp Leu Tyr Thr	Trp Gln Ser
1295	1300	1305
Arg Arg Ile Arg Leu His His	Asp Ala Asp Gly Val	His Gly Val
1310	1315	1320
Val Leu Ala Tyr Gly Asp Ser	Leu Glu Pro His Asn	Arg His Gly
1325	1330	1335
His Glu Pro Met Thr Ser Trp	Arg Arg Ser Pro Thr	Gln Glu Lys
1340	1345	1350

Lys Arg Gln Glu Asn Leu Val Tyr Leu Pro Arg Glu His Asp Pro			
1355	1360	1365	
Ser Arg Leu Ala Trp Arg Gly Met Asp Gly Leu Leu Ala Gly Arg			
1370	1375	1380	
Glu Thr Gly Ser Ala Gln Gly Pro Asp Gly Ala Asp Arg Leu Ala			
1385	1390	1395	
Pro Lys Val Val Gln Trp Ala Ala Gln Leu Thr Thr Glu Gly Leu			
1400	1405	1410	
Leu Pro Arg Gly Tyr Leu Ile Arg Thr Arg Val Ile Gly Ala Arg			
1415	1420	1425	
Tyr Gly Thr Gln Gln Ser Val Ile Asp Glu Val Val Asp Asp Gly			
1430	1435	1440	
Val Leu Met Pro Ala Val Leu Leu His Glu Ala Asp Arg Arg Tyr			
1445	1450	1455	
Gly Asp Lys Ala Val Asp Ala Leu His Asp Ala Glu Lys Ala Val			
1460	1465	1470	
Gly Ala Leu Ala Gln Leu Ala Ala Asp Leu Ala Leu Ala Val Gly			
1475	1480	1485	
Thr Asp Pro Glu Pro Gly Arg Asn Thr Ala Arg Asp Leu Gly Phe			
1490	1495	1500	
Gly Thr Leu Asp Thr His Tyr Arg Arg Trp Leu Arg Glu Leu Gly			
1505	1510	1515	
Gly Thr Ser Asp Pro Glu Glu His Arg Asp Arg Trp Lys Gln Glu			
1520	1525	1530	
Val Arg Arg Leu Val Ala Glu Leu Gly Glu Arg Leu Leu Asp Gly			
1535	1540	1545	
Ala Gly Pro Ala Ala Trp Glu Gly Arg Leu Val Glu Thr Gly Lys			
1550	1555	1560	
Gly Thr Arg Trp Leu Asn Asp Ala Ala Ala Glu Leu Arg Phe Arg			
1565	1570	1575	
Thr Arg Leu Arg Glu Phe Leu Thr Thr Ala Pro Asp Thr Pro Thr			
1580	1585	1590	

Ser Pro Arg Pro Ala Pro Val Glu Ser Pro Ala

1595 1600

<210> 44

<211> 1559

<212> PRT

<213> Streptomyces griseus

<400> 44

Met Ser Asn Thr Pro Met Ser Arg Asp His Pro Glu Ser Leu Ser Ala

1 5 10 15

Tyr Ala Arg Leu Ser Pro Val Ser Arg Thr Ala Trp Gly Lys His Asp

20 25 30

Arg Gln Thr Glu Gln Trp Leu Pro Leu Trp Arg His Met Ala Asp Ser

35 40 45

Ala Ala Val Ala Glu Arg Leu Trp Asp Gln Trp Val Pro Asp Asn Val

50 55 60

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

65 70 75 80

Arg Val Ala Val Phe Leu Ala Cys Val His Asp Ile Gly Lys Ala Thr

85 90 95

Pro Ala Phe Ala Cys Gln Val Asp Gly Leu Ala Asp Arg Met Arg Ala

100 105 110

Ala Gly Leu Ser Met Pro Tyr Leu Lys Gln Phe Gly Leu Asp Arg Arg

115 120 125

Met Ala Pro His Gly Leu Ala Gly Gln Leu Leu Leu Gln Glu Trp Leu

130 135 140

Ala Glu Arg Phe Gly Trp Ser Glu Arg Ala Ser Gly Gln Phe Ala Val

145 150 155 160

Val Ala Gly Gly His His Gly Thr Pro Pro Asp His Gln His Ile His

165 170 175

Asp Leu Gly Leu Arg Pro His Leu Leu Arg Thr Ala Gly Glu Ser Gln

180 185 190

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



Leu Ala Pro Val Asp Leu Val Leu Gln Arg Met Gly Arg Leu His Arg  
 690 695 700  
 His Pro Arg Thr Arg Pro Pro Arg Leu Ser Arg Ala Arg Cys Leu Ile  
 705 710 715 720  
 Thr Gly Val Glu Asp Trp His Ala Glu Arg Pro Val Pro Val Arg Gly  
 725 730 735  
 Ser Leu Ala Val Tyr Gln Gly Pro His Thr Leu Leu Arg Ala Leu Ala  
 740 745 750  
  
 Val Leu Gly Pro His Leu Asp Gly Val Pro Leu Val Leu Pro Asp His  
 755 760 765  
 Ile Ser Pro Leu Val Gln Ala Ala Tyr Asp Glu Arg Pro Val Gly Pro  
 770 775 780  
 Ala His Trp Ala Pro Val Leu Asp Glu Ala Arg Arg Gln Tyr Leu Thr  
 785 790 795 800  
 Arg Leu Ala Glu Lys Arg Glu Arg Ala Asp Val Phe Arg Leu Gly Pro  
 805 810 815  
  
 Val Arg Arg Pro Gly Arg Pro Leu Phe Gly Trp Leu Asp Gly Asn Ala  
 820 825 830  
 Gly Asp Ala Asp Asp Ser Arg Thr Gly Arg Ala Gln Val Arg Asp Ser  
 835 840 845  
 Glu Glu Ser Leu Glu Val Leu Val Val Gln Arg Arg Ala Asp Gly Arg  
 850 855 860  
 Leu Thr Thr Val Ser Trp Leu Asp Gly Gly Arg Gly Gly Leu Asp Leu  
 865 870 875 880  
  
 Pro Glu His Ala Pro Pro Pro Pro Arg Ala Ala Glu Val Val Ala Ala  
 885 890 895  
 Cys Ala Leu Thr Leu Pro Arg Ser Leu Thr His Pro Gly Val Ile Asp  
 900 905 910  
 Arg Thr Ile Ala Glu Leu Glu Arg Phe Val Val Pro Ala Trp Gln Val  
 915 920 925  
 Lys Glu Cys Pro Trp Leu Ala Gly Glu Leu Leu Leu Val Leu Asp Glu  
 930 935 940

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

1175	1180	1185
Ala Ser Leu Arg Glu Thr Leu Leu Leu Asn Leu Ile Pro Thr Asp		
1190	1195	1200
Glu Asp Ile Leu Thr Ser Glu Pro Lys Ala Asp Leu Pro Val Trp		
1205	1210	1215
Arg Arg Glu Thr Pro Pro Gly Pro Gly Val Val Glu Gly Asp Pro		
1220	1225	1230
Ser Ala Pro Arg Pro Ala Gly Pro Arg Asp Leu Tyr Thr Trp Gln		
1235	1240	1245
Ser Arg Arg Leu Leu Leu His Thr Glu Gly Ser Asp Ala Ile Gly		
1250	1255	1260
Val Val Leu Gly Tyr Gly Asp Pro Leu Ser Pro Ala Asn Arg Gln		
1265	1270	1275
Lys Thr Glu Pro Met Thr Gly Trp Arg Arg Ser Pro Ala Gln Glu		
1280	1285	1290
Lys Lys Leu Gly Arg Pro Leu Val Tyr Leu Pro Arg Gln His Asp		
1295	1300	1305
Pro Gly Arg Ala Ala Trp Arg Gly Leu Ala Ser Leu Leu Tyr Pro		
1310	1315	1320
Gln Gly Glu Asp Gly Asp Thr Thr Gly Arg Gly Thr Asp Arg Ser		
1325	1330	1335
Arg Pro Ala Gly Ile Val Arg Trp Leu Ala Leu Leu Ser Thr Glu		
1340	1345	1350
Gly Val Leu Pro Lys Gly Ser Leu Ile Arg Thr Arg Leu Val Gly		
1355	1360	1365
Ala Val Tyr Gly Thr Gln Gln Ser Val Val Asp Asp Val Val Asp		
1370	1375	1380
Asp Ser Ile Ala Leu Pro Val Val Leu Leu His Gln Asp Arg Arg		
1385	1390	1395
Leu His Gly Ala Val Ala Val Asp Ala Val Ala Asp Ala Glu Arg		
1400	1405	1410

Ala Val Ser Ala Leu Gly His Leu Ala Gly Asn Leu Ala Arg Ala  
 1415 1420 1425

Ser Gly Ser Glu Ala Gly Pro Ala Thr Ala Thr Ala Arg Asp Gln  
 1430 1435 1440

Gly Phe Gly Ala Leu Asp Gly Pro Tyr Arg Arg Trp Leu Val Asp  
 1445 1450 1455

Leu Ala Glu Asp Thr Asp Leu Glu Arg Ala Arg Ala Ala Trp Arg  
 1460 1465 1470

Asp Thr Val Arg Leu Val Val Leu Gly Ile Gly Arg Glu Leu Leu  
 1475 1480 1485

Asp Ala Ala Gly Arg Ala Ala Ala Glu Gly Arg Val Ile Glu Leu  
 1490 1495 1500

Pro Gly Val Gly Lys Arg Trp Ile Asp Ser Ser Arg Ala Asp Leu  
 1505 1510 1515

Trp Phe Arg Thr Arg Ile Asn Arg Val Leu Pro Arg Pro Leu Pro  
 1520 1525 1530

Glu Ala His Ala Pro Thr Ala Asp Ile His Ala Gly His Ala Val  
 1535 1540 1545

Arg Ala Asp Glu Ala Leu Ser Glu Glu Thr Val  
 1550 1555

<210> 45

<211> 1540

<212> PRT

<213> *Catenulispora acidiphila*

<400> 45

Met Phe Asn Val Gly Ser Thr Arg Cys Trp Gly Asp Gly Gly Leu Arg  
 1 5 10 15

Asn Ala Ala Glu Asp Leu Ser Ala Ala Thr Arg Ser Ala Trp Ala Lys  
 20 25 30

Ser Asp Pro Asp Ser Gly Gln Ser Leu Ser Leu Ile Arg His Leu Ala

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



545                    550                    555                    560  
 Gln Ala Val Gly Thr Glu Pro Ala Glu Gly Arg Arg Val Asp Val Asn  
                                  565                    570                    575  
 Val Glu Ala Phe Asp Asp Asp Leu Gly Arg Leu Ala Asp Arg Leu Glu  
                                  580                    585                    590  
 Ala Glu Leu Val Asp Gly Gly Cys Ala Leu Ile Ile Arg Asn Thr Val  
                                  595                    600                    605  
 Gly Arg Val Leu Gln Thr Ala Gln Gln Leu Arg Glu Arg Phe Gly Ala  
  
                                  610                    615                    620  
 Gly Gln Val Thr Val Ala His Ser Arg Phe Ile Asp Leu Asp Arg Ala  
 625                    630                    635                    640  
 Arg Lys Asp Ala Asp Leu Leu Ala Arg Phe Gly His Asp Gly Ala Arg  
                                  645                    650                    655  
 Pro Arg Arg His Ile Val Val Ala Ser Gln Val Ala Glu Gln Ser Leu  
                                  660                    665                    670  
 Asp Ile Asp Phe Asp Leu Leu Val Thr Asp Leu Ala Pro Ile Asp Leu  
  
                                  675                    680                    685  
 Val Leu Gln Arg Met Gly Arg Val His Arg His His Arg Gly Gly Pro  
                                  690                    695                    700  
 Glu Gln Ser Glu Arg Pro Pro Ser Leu Arg Thr Ala Arg Cys Leu Val  
 705                    710                    715                    720  
 Thr Gly Val Asp Trp Ala Gly Ile Pro Ser Ala Pro Ile Ala Gly Ser  
                                  725                    730                    735  
 Val Ala Val Tyr Gly Leu His Pro Leu Leu Arg Ser Leu Ala Val Leu  
  
                                  740                    745                    750  
 Gln Pro Tyr Leu Thr Gly Ser Ala Leu Thr Leu Pro Gly Asp Ile Asn  
                                  755                    760                    765  
 Pro Leu Val Gln Cys Ala Tyr Ala Gln Ser Phe Val Ala Pro Thr Gly  
                                  770                    775                    780  
 Trp Gly Glu Ala Met Asp Ala Ala Gln Ala Glu His Met Ala His Ile  
 785                    790                    795                    800

Val Gln Gln Arg Glu Gly Ala Met Ala Phe Cys Leu Asp Glu Val Arg  
 805 810 815  
 Gly Pro Gly Arg Ser Leu Ile Gly Trp Ile Asp Gly Gly Val Gly Asp  
 820 825 830  
 Ala Asp Asp Thr Arg Ala Gly Arg Ala Gln Val Arg Asp Ser Pro Glu  
 835 840 845  
 Thr Ile Glu Val Leu Val Val Gln Arg Gly Ser Asp Gly Val Leu Arg  
 850 855 860  
 Thr Leu Pro Trp Leu Asp Arg Gly Arg Gly Gly Leu Glu Leu Pro Thr  
 865 870 875 880  
 Glu Ala Val Pro Pro Pro Arg Ala Ala Arg Ala Ala Ala Ala Ser Ala  
 885 890 895  
 Leu Arg Leu Pro Gly Leu Phe Ala Lys Pro Trp Met Phe Asp Arg Val  
 900 905 910  
 Leu Arg Glu Leu Glu Arg Glu Tyr His Glu Ala Trp Gln Ala Lys Glu  
 915 920 925  
 Ser Ser Trp Leu Gln Gly Glu Leu Leu Leu Val Leu Asp Glu Glu Cys  
 930 935 940  
 Arg Thr Val Leu Ala Gly Tyr Glu Leu Ser Tyr Asn Pro Asp Asp Gly  
 945 950 955 960  
 Leu Glu Met Val Met Pro Gly Glu Pro His Ala Ala Val Val Arg Asp  
 965 970 975  
 Lys Glu Ala Ser Asp Asp Lys Thr Ala Ser Phe Asp Leu Thr Ser Ala  
 980 985 990  
 Pro Trp Leu Pro Val Leu Tyr Ala Asp Gly Met Gln Gly Val Leu Ser  
 995 1000 1005  
 Leu Arg Asp Val Phe Ala Gln Ser Asn Leu Ile Arg Arg Leu Val  
 1010 1015 1020  
 Gly Asp Leu Pro Thr Gln Asp Phe Ala Leu Leu Arg Leu Leu Leu  
 1025 1030 1035  
 Ala Val Leu Tyr Asp Ala Val Asp Gly Pro Arg Asp Gly Gln Asp

1040                      1045                      1050  
 Trp Glu Asp Leu Trp Thr Ser Asp Asp Pro Phe Ala Ala Val Pro

1055                      1060                      1065  
 Ala Tyr Leu Asp Ser His Arg Glu Arg Phe Asp Leu Leu His Pro

1070                      1075                      1080  
 Ala Thr Pro Phe Tyr Gln Val Pro Gly Leu Gln Thr Ala Lys Gly

1085                      1090                      1095  
 Glu Val Gly Pro Leu Asn Lys Ile Val Ala Asp Val Pro Asp Gly

1100                      1105                      1110  
 Asp Pro Phe Leu Thr Met Arg Met Pro Gly Val Glu Gln Leu Ser

1115                      1120                      1125  
 Phe Ala Glu Ala Ala Arg Trp Leu Val His Thr Gln Ala Phe Asp

1130                      1135                      1140  
 Thr Ser Gly Ile Lys Ser Gly Val Val Gly Asp Pro Lys Ala Val

1145                      1150                      1155  
 Asn Gly Lys Arg Tyr Pro Gln Gly Val Ala Trp Leu Gly Asn Leu

1160                      1165                      1170  
 Gly Gly Val Phe Ala Glu Gly Asp Thr Leu Arg Gln Thr Leu Leu

1175                      1180                      1185  
 Leu Asn Leu Ile Pro Ala Asp Thr Thr Asn Leu Gln Val Thr Ser

1190                      1195                      1200  
 Ala Gln Asp Val Pro Ala Trp Arg Gly Thr Asn Gly Arg Ala Gly

1205                      1210                      1215  
 Ser Asp His Ala Asp Ala Glu Pro Arg Val Pro Ala Gly Leu Arg

1220                      1225                      1230  
 Asp Leu Tyr Thr Trp Gln Ser Arg Arg Ile Arg Leu Glu Tyr Asp

1235                      1240                      1245  
 Thr Arg Gly Val Thr Gly Ala Val Leu Thr Tyr Gly Asp Glu Leu

1250                      1255                      1260  
 Thr Ala His Asn Lys His Gly Val Glu Pro Met Thr Gly Trp Arg

1265                      1270                      1275

Arg Ser Lys Pro Gln Glu Lys Lys Leu Gly Leu Ser Thr Val Tyr  
 1280 1285 1290

Met Pro Gln Gln His Asp Pro Thr Arg Ala Ala Trp Arg Gly Ile  
 1295 1300 1305

Glu Ser Leu Leu Ala Gly Ser Ala Gly Ser Gly Ser Ser Gln Thr  
 1310 1315 1320

Gly Glu Pro Ala Ser His Tyr Arg Pro Lys Ile Val Asp Trp Leu  
 1325 1330 1335

Gly Glu Leu Ala His His Gly Asn Leu Pro Ser Arg Gly Leu Ile  
 1340 1345 1350

Arg Val Arg Thr Ser Gly Ala Val Tyr Gly Thr Gln Gln Ser Ile  
 1355 1360 1365

Ile Asp Glu Val Val Ser Asp Glu Leu Thr Met Ala Val Val Leu  
 1370 1375 1380

Leu His Glu Asp Asp Pro Arg Phe Gly Lys Ala Ala Val Thr Ala  
 1385 1390 1395

Val Lys Asp Ala Asp Ser Ala Val Ala Ala Leu Gly Asp Leu Ala  
 1400 1405 1410

Ser Asp Leu Ala Arg Ala Ala Gly Leu Asp Pro Glu Pro Glu Arg  
 1415 1420 1425

Val Thr Ala Arg Asp Arg Ala Phe Gly Ala Leu Asp Gly Pro Tyr  
 1430 1435 1440

Arg Arg Trp Leu Leu Asp Leu Gly Asn Ser Thr Asp Pro Ala Ala  
 1445 1450 1455

Met Arg Ala Val Trp Gln Gly Arg Val Tyr Asp Ile Ile Ala Val  
 1460 1465 1470

Gln Gly Gln Met Leu Leu Asp Ser Ala Gly Ser Ala Ala Ala Gln  
 1475 1480 1485

Gly Arg Met Val Lys Thr Thr Arg Gly Glu Arg Trp Met Asp Asp  
 1490 1495 1500

Ser Leu Ala Asp Leu Tyr Phe Lys Gly Arg Ile Ala Lys Ala Leu

1505                      1510                      1515  
 Ser Ser Arg Leu Gly Lys Lys Pro Thr Asp Pro Gly Glu Pro Val  
 1520                      1525                      1530  
 Gly Ile Gln Glu Asp Pro Ala  
 1535                      1540

<210> 46

<211> 1407

<212> PRT

<213> Artificial Sequence

<220><223> Fusion Cas3-Cse1

<400> 46

Met Glu Pro Phe Lys Tyr Ile Cys His Tyr Trp Gly Lys Ser Ser Lys  
 1                      5                      10                      15  
 Ser Leu Thr Lys Gly Asn Asp Ile His Leu Leu Ile Tyr His Cys Leu  
                     20                      25                      30  
 Asp Val Ala Ala Val Ala Asp Cys Trp Trp Asp Gln Ser Val Val Leu  
                     35                      40                      45  
 Gln Asn Thr Phe Cys Arg Asn Glu Met Leu Ser Lys Gln Arg Val Lys  
  
                     50                      55                      60  
 Ala Trp Leu Leu Phe Phe Ile Ala Leu His Asp Ile Gly Lys Phe Asp  
 65                      70                      75                      80  
 Ile Arg Phe Gln Tyr Lys Ser Ala Glu Ser Trp Leu Lys Leu Asn Pro  
                     85                      90                      95  
 Ala Thr Pro Ser Leu Asn Gly Pro Ser Thr Gln Met Cys Arg Lys Phe  
                     100                      105                      110  
 Asn His Gly Ala Ala Gly Leu Tyr Trp Phe Asn Gln Asp Ser Leu Ser  
  
                     115                      120                      125  
 Glu Gln Ser Leu Gly Asp Phe Phe Ser Phe Phe Asp Ala Ala Pro His  
                     130                      135                      140  
 Pro Tyr Glu Ser Trp Phe Pro Trp Val Glu Ala Val Thr Gly His His  
 145                      150                      155                      160  
 Gly Phe Ile Leu His Ser Gln Asp Gln Asp Lys Ser Arg Trp Glu Met



Gly Gln Ile Gly Val Cys Thr Ile Asp Gln Val Leu Ile Ser Val Leu  
 420 425 430

Pro Val Lys His Arg Phe Ile Arg Gly Leu Gly Ile Gly Arg Ser Val  
 435 440 445

Leu Ile Val Asp Glu Val His Ala Tyr Asp Thr Tyr Met Asn Gly Leu  
 450 455 460

Leu Glu Ala Val Leu Lys Ala Gln Ala Asp Val Gly Gly Ser Val Ile  
 465 470 475 480

Leu Leu Ser Ala Thr Leu Pro Met Lys Gln Lys Gln Lys Leu Leu Asp  
 485 490 495

Thr Tyr Gly Leu His Thr Asp Pro Val Glu Asn Asn Ser Ala Tyr Pro  
 500 505 510

Leu Ile Asn Trp Arg Gly Val Asn Gly Ala Gln Arg Phe Asp Leu Leu  
 515 520 525

Ala His Pro Glu Gln Leu Pro Pro Arg Phe Ser Ile Gln Pro Glu Pro  
 530 535 540

Ile Cys Leu Ala Asp Met Leu Pro Asp Leu Thr Met Leu Glu Arg Met  
 545 550 555 560

Ile Ala Ala Ala Asn Ala Gly Ala Gln Val Cys Leu Ile Cys Asn Leu  
 565 570 575

Val Asp Val Ala Gln Val Cys Tyr Gln Arg Leu Lys Glu Leu Asn Asn  
 580 585 590

Thr Gln Val Asp Ile Asp Leu Phe His Ala Arg Phe Thr Leu Asn Asp  
 595 600 605

Arg Arg Glu Lys Glu Asn Arg Val Ile Ser Asn Phe Gly Lys Asn Gly  
 610 615 620

Lys Arg Asn Val Gly Arg Ile Leu Val Ala Thr Gln Val Val Glu Gln  
 625 630 635 640

Ser Leu Asp Val Asp Phe Asp Trp Leu Ile Thr Gln His Cys Pro Ala  
 645 650 655

Asp Leu Leu Phe Gln Arg Leu Gly Arg Leu His Arg His His Arg Lys



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

1145	1150	1155
Gly Gln Glu Ser Asn Leu Arg Tyr Thr Gly Phe Leu Lys Glu Lys		
1160	1165	1170
Phe Thr Phe Thr Val Asn Gly Leu Trp Pro His Pro His Ser Pro		
1175	1180	1185
Cys Leu Val Thr Val Lys Lys Gly Glu Val Glu Glu Lys Phe Leu		
1190	1195	1200
Ala Phe Thr Thr Ser Ala Pro Ser Trp Thr Gln Ile Ser Arg Val		
1205	1210	1215
Val Val Asp Lys Ile Ile Gln Asn Glu Asn Gly Asn Arg Val Ala		
1220	1225	1230
Ala Val Val Asn Gln Phe Arg Asn Ile Ala Pro Gln Ser Pro Leu		
1235	1240	1245
Glu Leu Ile Met Gly Gly Tyr Arg Asn Asn Gln Ala Ser Ile Leu		
1250	1255	1260
Glu Arg Arg His Asp Val Leu Met Phe Asn Gln Gly Trp Gln Gln		
1265	1270	1275
Tyr Gly Asn Val Ile Asn Glu Ile Val Thr Val Gly Leu Gly Tyr		
1280	1285	1290
Lys Thr Ala Leu Arg Lys Ala Leu Tyr Thr Phe Ala Glu Gly Phe		
1295	1300	1305
Lys Asn Lys Asp Phe Lys Gly Ala Gly Val Ser Val His Glu Thr		
1310	1315	1320
Ala Glu Arg His Phe Tyr Arg Gln Ser Glu Leu Leu Ile Pro Asp		
1325	1330	1335
Val Leu Ala Asn Val Asn Phe Ser Gln Ala Asp Glu Val Ile Ala		
1340	1345	1350
Asp Leu Arg Asp Lys Leu His Gln Leu Cys Glu Met Leu Phe Asn		
1355	1360	1365
Gln Ser Val Ala Pro Tyr Ala His His Pro Lys Leu Ile Ser Thr		
1370	1375	1380

Leu Ala Leu Ala Arg Ala Thr Leu Tyr Lys His Leu Arg Glu Leu  
1385 1390 1395

Lys Pro Gln Gly Gly Pro Ser Asn Gly  
1400 1405

<210> 47

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> CASCADE sequence cleavage

<400> 47

ccgtcttgcg ctagctctag aactagtect cagcctagge ctaagctgt