



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

(11) 공개번호 10-2021-0070305  
(43) 공개일자 2021년06월14일

- (51) 국제특허분류(Int. Cl.)  
A61K 48/00 (2006.01) C07K 14/005 (2006.01)  
C07K 14/46 (2006.01) C12N 15/86 (2006.01)
- (52) CPC특허분류  
A61K 48/0008 (2013.01)  
C07K 14/005 (2013.01)
- (21) 출원번호 10-2021-7011429
- (22) 출원일자(국제) 2019년09월18일  
심사청구일자 없음
- (85) 번역문제출일자 2021년04월16일
- (86) 국제출원번호 PCT/US2019/051786
- (87) 국제공개번호 WO 2020/061229  
국제공개일자 2020년03월26일
- (30) 우선권주장  
62/733,015 2018년09월18일 미국(US)

- (71) 출원인  
브이엔बी이 뉴코 인크.  
미국 10010 뉴욕주 뉴욕 920 브로드웨이 11 플로어
- (72) 발명자  
말론 콜린  
미국 11211 뉴욕주 브루클린 홈볼트 스트리트 345 아파트먼트 2  
페이콘 이안  
미국 11714 뉴욕주 베스페이지 나인스 스트리트 265  
(뒷면에 계속)
- (74) 대리인  
김진희, 김태홍

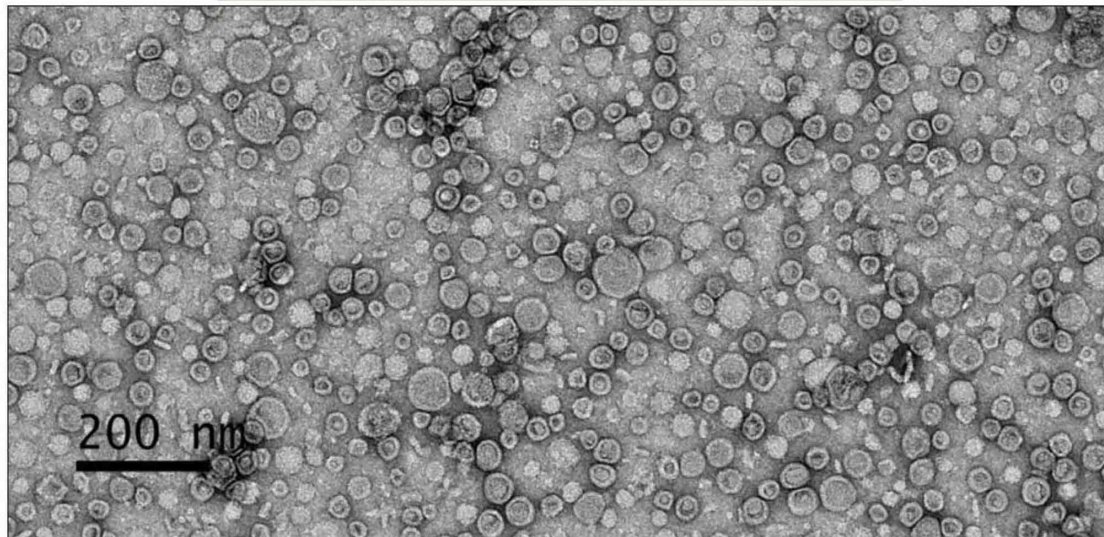
전체 청구항 수 : 총 24 항

(54) 발명의 명칭 ARC 기반 캡시드 및 이의 용도

**(57) 요약**

특정 실시양태에서, 제조함 Arc 및 내인성 Gag 폴리펩티드, 및 제조함 Arc 및 내인성 Gag 폴리펩티드를 사용하는 방법이 본원에 개시된다.

**대표도** - 도5



(52) CPC특허분류

*C07K 14/46* (2013.01)

*C12N 15/86* (2013.01)

(72) 발명자

**길버트 잭**

미국 10010 뉴욕주 뉴욕 브로드웨이 920 일레븐스  
플로어

**피스아레브 안드레이**

미국 10010 뉴욕주 뉴욕 브로드웨이 920 일레븐스  
플로어

**프라이데스 아담**

미국 10010 뉴욕주 뉴욕 브로드웨이 920 일레븐스  
플로어

**크리스프 제시카**

미국 10010 뉴욕주 뉴욕 브로드웨이 920 일레븐스  
플로어

## 명세서

### 청구범위

#### 청구항 1

재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드 및 치료제를 포함하는 캡시드.

#### 청구항 2

제1항에 있어서, 치료제가 핵산인 캡시드.

#### 청구항 3

제2항에 있어서, 핵산이 RNA인 캡시드.

#### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 재조합 Arc 폴리펩티드가 서열번호 1인 아미노산 서열 또는 서열번호 1과 적어도 90% 동일한 아미노산 서열을 포함하는 인간 Arc 폴리펩티드인 캡시드.

#### 청구항 5

제1항 내지 제3항 중 어느 한 항에 있어서, 재조합 Arc 폴리펩티드가

- a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열;
  - b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열;
  - c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열;
  - d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열;
  - e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열;
  - f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열;
  - g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; 또는
  - h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; 또는
  - i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는
  - j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는
  - k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는
  - l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는
  - m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는
  - n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열
- 을 포함하는 Arc 폴리펩티드인 캡시드.

#### 청구항 6

제1항 내지 제3항 중 어느 한 항에 있어서, 재조합 내인성 Gag 폴리펩티드가 인간 내인성 Gag 폴리펩티드인 캡시드.

#### 청구항 7

제6항에 있어서, 재조합 인간 내인성 Gag 폴리펩티드가

- a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열;
  - b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열;
  - c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열;
  - d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열;
  - e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열;
  - f) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열;
  - g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열;
  - h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열;
  - i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열;
  - j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는
  - k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는
  - l) 서열번호 27인 아미노산 서열 또는 서열번호 27과 적어도 90% 동일한 아미노산 서열; 또는
  - m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열
- 을 포함하는 내인성 Gag 폴리펩티드인 캡시드.

**청구항 8**

제조합 Arc 폴리펩티드 또는 제조합 내인성 Gag 폴리펩티드를 포함하고, 제조합 Arc 폴리펩티드는 래트 Arc 폴리펩티드 또는 인간 Arc 폴리펩티드가 아닌, 캡시드.

**청구항 9**

제8항에 있어서, 카고(cargo)를 추가로 포함하는 캡시드.

**청구항 10**

제9항에 있어서, 카고가 핵산인 캡시드.

**청구항 11**

제10항에 있어서, 카고가 RNA인 캡시드.

**청구항 12**

제9항에 있어서, 카고가 치료제인 캡시드.

**청구항 13**

제8항 내지 제12항 중 어느 한 항에 있어서, 제조합 Arc 폴리펩티드가

- a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열;
- b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열;
- c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열;
- d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열;
- e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열;
- f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열;
- g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열;
- h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열;

- i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는
  - j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는
  - k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는
  - l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는
  - m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는
  - n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열
- 을 포함하는 Arc 폴리펩티드인 캡시드.

**청구항 14**

제8항 내지 제12항 중 어느 한 항에 있어서, 재조합 내인성 Gag 폴리펩티드가

- a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열;
  - b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열;
  - c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열;
  - d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열;
  - e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열;
  - f) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열; 또는
  - g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열; 또는
  - h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열; 또는
  - i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열; 또는
  - j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는
  - k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는
  - l) 서열번호 27인 아미노산 서열 또는 서열번호 27과 적어도 90% 동일한 아미노산 서열; 또는
  - m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열
- 을 포함하는 내인성 Gag 폴리펩티드인 캡시드.

**청구항 15**

제1항 내지 제14항 중 어느 한 항의 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드를 코딩하는 DNA를 포함하는 벡터.

**청구항 16**

제1항 내지 제14항 중 어느 한 항의 캡시드를 세포에 투여하는 단계를 포함하는, 카고를 세포에 전달하는 방법.

**청구항 17**

제16항에 있어서, 세포가 진핵 세포인 방법.

**청구항 18**

제16항에 있어서, 세포가 척추동물 세포인 방법.

**청구항 19**

제16항에 있어서, 세포가 포유동물 세포인 방법.

**청구항 20**

제16항에 있어서, 세포가 인간 세포인 방법.

**청구항 21**

제16항 내지 제20항 중 어느 한 항에 있어서, 카고가 핵산인 방법.

**청구항 22**

제21항에 있어서, 세포가 핵산에 의해 코딩되는 유전자를 발현하는 것인 방법.

**청구항 23**

제16항 내지 제20항 중 어느 한 항에 있어서, 카고가 치료제인 방법.

**청구항 24**

제1항 내지 제14항 중 어느 한 항의 캡시드를 세포에 투여하는 단계를 포함하는, 핵산을 세포로 형질감염시키는 방법.

**발명의 설명**

**발명의 내용**

[0001]

**상호 참조**

[0002]

본 출원은 2018년 9월 18일에 출원된 미국 가특허 출원 번호 62/733,015의 이익을 주장하며, 이는 그 전체가 본원에 참조로 포함된다.

[0003]

**개시내용의 요약**

[0004]

특정 실시양태에서, 재조합 및 조작된 Arc 폴리펩티드 및 재조합 및 조작된 내인성 Gag(endo-Gag) 폴리펩티드가 본원에 개시된다. 일부 실시양태에서, 또한 로딩된 또는 비어있는 Arc 기반 캡시드 및 endo-Gag 기반 캡시드, 및 캡시드를 제조하는 방법이 포함된다. 또한, Arc 기반 캡시드 및 endo-Gag 기반 캡시드를 관심 부위에 전달하는 방법이 포함된다.

[0005]

특정 실시양태에서, 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드 및 치료제를 포함하는 캡시드가 본원에 개시된다. 일부 실시양태에서, 치료제는 핵산이다. 일부 실시양태에서, 핵산은 RNA이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 서열번호 1인 아미노산 서열 또는 서열번호 1과 적어도 90% 동일한 아미노산 서열을 포함하는 인간 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열; b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열; c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열; d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열; e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열; f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열; g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는 m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는 n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열을 포함하는 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 인간 내인성 Gag 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열; b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열; c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열; d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열; e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열; f) 서열번호 21인 아미노산 서열 또는 서열번호

호 21과 적어도 90% 동일한 아미노산 서열; 또는 g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열; 또는 h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열; 또는 i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 27인 아미노산 서열 또는 서열번호 27과 적어도 90% 동일한 아미노산 서열 또는 m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열을 포함하는 내인성 Gag 폴리펩티드이다.

[0006]

특정 실시양태에서, 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드를 포함하고, 여기서 재조합 Arc 폴리펩티드는 래트 Arc 폴리펩티드 또는 인간 Arc 폴리펩티드가 아닌, 캡시드가 본원에 개시된다. 일부 실시양태에서, 캡시드는 카고(cargo)를 추가로 포함한다. 일부 실시양태에서, 카고는 핵산이다. 일부 실시양태에서, 카고는 RNA이다. 일부 실시양태에서, 카고는 치료제이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열; b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열; c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열; d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열; e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열; f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열; g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는 m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는 n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열을 포함하는 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열; b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열; c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열; d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열; e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열; f) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열; 또는 g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열; 또는 h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열; 또는 i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 27인 아미노산 서열 또는 서열번호 27과 적어도 90% 동일한 아미노산 서열 또는 m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열을 포함하는 내인성 Gag 폴리펩티드이다.

[0007]

특정 실시양태에서, 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드를 코딩하는 DNA를 포함하는 벡터가 본원에 개시된다. 일부 실시양태에서, 벡터는 치료제를 추가로 코딩한다. 일부 실시양태에서, 치료제는 핵산이다. 일부 실시양태에서, 핵산은 RNA이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 서열번호 1인 아미노산 서열 또는 서열번호 1과 적어도 90% 동일한 아미노산 서열을 포함하는 인간 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열; b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열; c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열; d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열; e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열; f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열; g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는 m) 서열번호 14인 아미노산 서열 또는 서열번호 14과 적어도 90% 동일한 아미노산 서열; 또는 n) 서열번호 15인

아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열을 포함하는 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 인간 내인성 Gag 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열; b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열; c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열; d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열; e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열; f) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열; 또는 g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열; 또는 h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열; 또는 i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 27인 아미노산 서열 또는 서열번호 27과 적어도 90% 동일한 아미노산 서열 또는 m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열을 포함하는 내인성 Gag 폴리펩티드이다.

[0008]

특정 실시양태에서, 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드를 코딩하는 DNA를 포함하고, 여기서 재조합 Arc 폴리펩티드는 래트 Arc 폴리펩티드 또는 인간 Arc 폴리펩티드가 아닌, 벡터가 본원에 개시된다. 일부 실시양태에서, 벡터는 카고를 추가로 코딩한다. 일부 실시양태에서, 카고는 핵산이다. 일부 실시양태에서, 카고는 RNA이다. 일부 실시양태에서, 카고는 치료제이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열; b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열; c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열; d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열; e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열; f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열; g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는 m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는 n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열을 포함하는 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열; b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열; c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열; d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열; e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열; f) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열; 또는 g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열; 또는 h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열; 또는 i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 27인 아미노산 서열 또는 서열번호 27과 적어도 90% 동일한 아미노산 서열 또는 m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열을 포함하는 내인성 Gag 폴리펩티드이다.

[0009]

특정 실시양태에서, 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드 및 치료제를 포함하는 캡시드를 세포에 투여하는 단계를 포함하는, 카고를 세포에 전달하는 방법이 본원에 개시된다. 일부 실시양태에서, 치료제는 핵산이다. 일부 실시양태에서, 핵산은 RNA이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 서열번호 1인 아미노산 서열 또는 서열번호 1과 적어도 90% 동일한 아미노산 서열을 포함하는 인간 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열; b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열; c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열; d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열; e) 서열번호 6인 아미노산 서열 또는 서열번호 6과

적어도 90% 동일한 아미노산 서열; f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열; g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는 m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는 n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열을 포함하는 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 인간 내인성 Gag 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열; b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열; c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열; d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열; e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열; f) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열; 또는 g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열; 또는 h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열; 또는 i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 27인 아미노산 서열 또는 서열번호 27과 적어도 90% 동일한 아미노산 서열 또는 m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열을 포함하는 내인성 Gag 폴리펩티드이다. 일부 실시양태에서, 세포는 진핵 세포이다. 일부 실시양태에서, 세포는 척추동물 세포이다. 일부 실시양태에서, 세포는 포유동물 세포이다. 일부 실시양태에서, 세포는 인간 세포이다. 일부 실시양태에서, 카고는 핵산이다. 일부 실시양태에서, 세포는 핵산에 의해 코딩되는 유전자를 발현한다. 일부 실시양태에서, 카고는 치료제이다.

[0010]

특정 실시양태에서, 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드를 포함하는 캡시드를 세포에 투여하는 단계를 포함하고, 여기서 재조합 Arc 폴리펩티드는 래트 Arc 폴리펩티드 또는 인간 Arc 폴리펩티드가 아닌, 카고를 세포에 전달하는 방법이 본원에 개시된다. 일부 실시양태에서, 캡시드는 카고를 추가로 포함한다. 일부 실시양태에서, 카고는 핵산이다. 일부 실시양태에서, 카고는 RNA이다. 일부 실시양태에서, 카고는 치료제이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열; b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열; c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열; d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열; e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열; f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열; g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는 m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는 n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열을 포함하는 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열; b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열; c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열; d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열; e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열; f) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열; 또는 g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열; 또는 h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열; 또는 i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 27인 아미노산 서열 또는

서열번호 27과 적어도 90% 동일한 아미노산 서열 또는 m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열을 포함하는 내인성 Gag 폴리펩티드이다. 일부 실시양태에서, 세포는 진핵 세포이다. 일부 실시양태에서, 세포는 척추동물 세포이다. 일부 실시양태에서, 세포는 포유동물 세포이다. 일부 실시양태에서, 세포는 인간 세포이다. 일부 실시양태에서, 카고는 핵산이다. 일부 실시양태에서, 세포는 핵산에 의해 코딩되는 유전자를 발현한다. 일부 실시양태에서, 카고는 치료제이다.

[0011] 특정 실시양태에서, 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드 및 치료제를 포함하는 캡시드를 세포에 투여하는 단계를 포함하는, 핵산을 세포로 형질감염시키는 방법이 본원에 개시된다. 일부 실시양태에서, 치료제는 핵산이다. 일부 실시양태에서, 핵산은 RNA이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 서열번호 1인 아미노산 서열 또는 서열번호 1과 적어도 90% 동일한 아미노산 서열을 포함하는 인간 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열; b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열; c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열; d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열; e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열; f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열; g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는 m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는 n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열을 포함하는 Arc 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 인간 내인성 Gag 폴리펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 a) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열; b) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열; c) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열; d) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열; e) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열; f) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열; 또는 g) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열; 또는 h) 서열번호 23인 아미노산 서열 또는 서열번호 23과 적어도 90% 동일한 아미노산 서열; 또는 i) 서열번호 24인 아미노산 서열 또는 서열번호 24와 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 25인 아미노산 서열 또는 서열번호 25와 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 26인 아미노산 서열 또는 서열번호 26과 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 27인 아미노산 서열 또는 서열번호 27과 적어도 90% 동일한 아미노산 서열 또는 m) 서열번호 28인 아미노산 서열 또는 서열번호 28과 적어도 90% 동일한 아미노산 서열을 포함하는 내인성 Gag 폴리펩티드이다.

[0012] 특정 실시양태에서, 재조합 Arc 폴리펩티드 또는 재조합 내인성 Gag 폴리펩티드를 포함하는 캡시드를 세포에 투여하는 단계를 포함하고, 여기서 재조합 Arc 폴리펩티드는 래트 Arc 폴리펩티드 또는 인간 Arc 폴리펩티드가 아닌, 핵산을 세포로 형질감염시키는 방법이 본원에 개시된다. 일부 실시양태에서, 캡시드는 카고를 추가로 포함한다. 일부 실시양태에서, 카고는 핵산이다. 일부 실시양태에서, 카고는 RNA이다. 일부 실시양태에서, 카고는 치료제이다. 일부 실시양태에서, 재조합 Arc 폴리펩티드는 a) 서열번호 2인 아미노산 서열 또는 서열번호 2와 적어도 90% 동일한 아미노산 서열; b) 서열번호 3인 아미노산 서열 또는 서열번호 3과 적어도 90% 동일한 아미노산 서열; c) 서열번호 4인 아미노산 서열 또는 서열번호 4와 적어도 90% 동일한 아미노산 서열; d) 서열번호 5인 아미노산 서열 또는 서열번호 5와 적어도 90% 동일한 아미노산 서열; e) 서열번호 6인 아미노산 서열 또는 서열번호 6과 적어도 90% 동일한 아미노산 서열; f) 서열번호 7인 아미노산 서열 또는 서열번호 7과 적어도 90% 동일한 아미노산 서열; g) 서열번호 8인 아미노산 서열 또는 서열번호 8과 적어도 90% 동일한 아미노산 서열; h) 서열번호 9인 아미노산 서열 또는 서열번호 9와 적어도 90% 동일한 아미노산 서열; i) 서열번호 10인 아미노산 서열 또는 서열번호 10과 적어도 90% 동일한 아미노산 서열; 또는 j) 서열번호 11인 아미노산 서열 또는 서열번호 11과 적어도 90% 동일한 아미노산 서열; 또는 k) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; 또는 l) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; 또는 m) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; 또는 n) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열을 포함하는 Arc 폴리

펩티드이다. 일부 실시양태에서, 재조합 내인성 Gag 폴리펩티드는 a) 서열번호 12인 아미노산 서열 또는 서열번호 12와 적어도 90% 동일한 아미노산 서열; b) 서열번호 13인 아미노산 서열 또는 서열번호 13과 적어도 90% 동일한 아미노산 서열; c) 서열번호 14인 아미노산 서열 또는 서열번호 14와 적어도 90% 동일한 아미노산 서열; d) 서열번호 15인 아미노산 서열 또는 서열번호 15와 적어도 90% 동일한 아미노산 서열; e) 서열번호 16인 아미노산 서열 또는 서열번호 16과 적어도 90% 동일한 아미노산 서열; f) 서열번호 17인 아미노산 서열 또는 서열번호 17과 적어도 90% 동일한 아미노산 서열; g) 서열번호 18인 아미노산 서열 또는 서열번호 18과 적어도 90% 동일한 아미노산 서열; g) 서열번호 19인 아미노산 서열 또는 서열번호 19와 적어도 90% 동일한 아미노산 서열; g) 서열번호 20인 아미노산 서열 또는 서열번호 20과 적어도 90% 동일한 아미노산 서열; g) 서열번호 21인 아미노산 서열 또는 서열번호 21과 적어도 90% 동일한 아미노산 서열; 또는 h) 서열번호 22인 아미노산 서열 또는 서열번호 22와 적어도 90% 동일한 아미노산 서열을 포함하는 내인성 Gag 폴리펩티드이다.

[0013] 특정 실시양태에서, 카고 결합 도메인 및 Arc 또는 endo-Gag 폴리펩티드로부터의 적어도 하나의 캡시드 형성 서브유닛을 포함하는 조작된 Arc 또는 endo-Gag 폴리펩티드가 본원에 개시된다. 일부 실시양태에서, 카고 결합 도메인은 핵산 결합 도메인을 포함한다. 일부 실시양태에서, 카고 결합 도메인은 소분자에 결합하는 폴리펩티드를 포함한다. 일부 실시양태에서, 카고 결합 도메인은 단백질, 펩티드, 또는 항체 또는 이의 결합 단편에 결합하는 폴리펩티드를 포함한다. 일부 실시양태에서, 카고 결합 도메인은 펩티드모방체 또는 뉴클레오티드모방체에 결합하는 폴리펩티드를 포함한다. 일부 실시양태에서, 적어도 하나의 캡시드 형성 서브유닛은 서열번호 1의 CA N-로브(lobe) 및/또는 CA C-로브에 상응하는 폴리펩티드를 포함한다. 일부 실시양태에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 상이한 종의 Arc 또는 endo-Gag 폴리펩티드로부터의 제2 캡시드 형성 서브유닛을 추가 포함한다. 일부 실시양태에서, 제2 캡시드 형성 서브유닛은 서열번호 1의 N-로브 및/또는 C-로브에 상응하는 폴리펩티드를 포함한다. 일부 실시양태에서, 적어도 하나의 캡시드 형성 서브유닛 및 제2 캡시드 형성 서브유닛은 포유동물, 설치류, 조류, 파충류, 어류, 곤충, 진균 또는 식물로부터 선택되는 Arc 또는 endo-Gag의 종으로부터 각각 독립적으로 선택된다. 일부 실시양태에서, 적어도 하나의 캡시드 형성 서브유닛 및 제2 캡시드 형성 서브유닛은 2개의 상이한 종으로부터의 것이다. 일부 실시양태에서, 카고 결합 도메인은 직접적으로 또는 링커를 통해 적어도 하나의 캡시드 형성 서브유닛의 C-말단에 융합된다. 일부 실시양태에서, 카고 결합 도메인은 직접적으로 또는 링커를 통해 적어도 하나의 캡시드 형성 서브유닛의 N-말단에 융합된다. 일부 실시양태에서, 제2 캡시드 형성 서브유닛은 직접적으로 또는 링커를 통해 적어도 하나의 캡시드 형성 서브유닛의 C-말단에 융합된다. 일부 실시양태에서, 제2 캡시드 형성 서브유닛은 직접적으로 또는 링커를 통해 적어도 하나의 캡시드 형성 서브유닛의 N-말단에 융합되고 제2 캡시드 형성 서브유닛은 직접적으로 또는 링커를 통해 적어도 하나의 캡시드 형성 서브유닛의 C-말단에 융합된다. 실시양태에서, 카고 결합 도메인은 직접적으로 또는 링커를 통해 적어도 하나의 캡시드 형성 서브유닛의 C-말단에 융합되고 제2 캡시드 형성 서브유닛은 직접적으로 또는 링커를 통해 적어도 하나의 캡시드 형성 서브유닛의 N-말단에 융합된다. 일부 실시양태에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 제2 폴리펩티드를 추가로 포함한다. 일부 실시양태에서, 제2 폴리펩티드는 직접적으로 또는 링커를 통해 적어도 하나의 캡시드 형성 서브유닛에 융합된다. 일부 실시양태에서, 제2 폴리펩티드는 직접적으로 또는 링커를 통해 카고 결합 도메인에 융합된다. 일부 실시양태에서, 제2 폴리펩티드는 단백질 또는 항체 또는 이의 결합 단편이다. 일부 실시양태에서, 단백질은 인간 단백질 또는 바이러스 단백질이다. 일부 실시양태에서, 단백질은 인간 Gag-유사 단백질이다. 일부 실시양태에서, 단백질은 관심 표적 수용체에 결합하도록 설계된 드 노보 조작된 단백질이다. 일부 실시양태에서, 제2 폴리펩티드는 조작된 Arc 또는 endo-Gag 폴리펩티드에 의해 형성된 캡시드의 관심 표적 부위로의 전달을 안내한다.

[0014] 특정 실시양태에서, 캡시드 형성, 핵산 결합 또는 전달과 관련되지 않은 부분이 제거된 말단절단된 Arc 또는 endo-Gag 폴리펩티드가 본원에 개시된다. 일부 실시양태에서, 부분은 뉴클레오티드 결합 도메인이 인간 Arc RNA 결합 도메인이 아니라면 매트릭스(MA) 도메인, 역전사효소(RT) 도메인, 뉴클레오티드 결합 도메인, 또는 이들의 조합을 포함한다. 일부 실시양태에서, 부분은 CA C-로브 도메인을 포함한다. 일부 실시양태에서, 부분은 N-말단 결실, C-말단 결실, 또는 이들의 조합을 포함한다. 일부 실시양태에서, N-말단 결실은 최대 10개 아미노산, 20개 아미노산, 30개 아미노산 또는 50개 아미노산의 결실을 포함한다. 일부 실시양태에서, C-말단 결실은 최대 10개 아미노산, 20개 아미노산, 30개 아미노산 또는 50개 아미노산의 결실을 포함한다.

[0015] 특정 실시양태에서, 말단절단된 Arc 또는 endo-Gag 폴리펩티드일 수 있는 조작된 Arc 또는 endo-Gag 폴리펩티드 및 조작된 Arc 또는 endo-Gag에 의해 형성되는 캡시드에 의해 캡슐화되는 카고를 포함하는 Arc 또는 endo-Gag 기반 캡시드가 본원에 개시된다. 일부 실시양태에서, 카고는 핵산 분자이다. 일부 실시양태에서, 핵산 분자는 DNA, RNA, 또는 DNA와 RNA의 혼합물이다. 일부 실시양태에서, DNA 및 RNA는 각각 독립적으로 단일 가닥, 이중

가닥, 또는 단일 및 이중 가닥의 혼합물이다. 일부 실시양태에서, 카고는 소분자이다. 일부 실시양태에서, 카고는 단백질이다. 일부 실시양태에서, 카고는 펩티드이다. 일부 실시양태에서, 카고는 항체 또는 이의 결합 단편이다. 일부 실시양태에서, 카고는 펩티드모방체 또는 뉴클레오티드모방체이다. 일부 실시양태에서, Arc 또는 endo-Gag 기반 캡시드는 조작된 Arc 또는 endo-Gag 폴리펩티드와 상이한 Arc 또는 endo-Gag 단백질의 하나 이상의 종으로부터의 하나 이상의 추가의 캡시드 서브유닛을 포함한다. 일부 실시양태에서, Arc 기반 또는 endo-Gag 기반 캡시드는 비-Arc 단백질로부터의 하나 이상의 추가의 캡시드 서브유닛을 포함한다. 일부 실시양태에서, 하나 이상의 추가의 캡시드 서브유닛은 코피아(Copia) 단백질, ASPRV1 단백질, SCAN 도메인 패밀리로부터의 단백질, 부신생물성 Ma 항원 패밀리에 의해 코딩되는 단백질(예컨대, PNMA5, PNMA6, PNMA6A 및 PNMA6B), 레트로트랜스포존 Gag-유사 패밀리로부터의 단백질(예컨대, RTL3, RTL6, RTL8A, RTL8B), 또는 이들의 조합을 포함한다. 일부 실시양태에서, 하나 이상의 추가의 캡시드 서브유닛은 BOP, LDOC1, MOAP1, PEG10, PNMA3, PNMA5, PNMA6A, PNMA6B, RTL3, RTL6, RTL8A, RTL8B, 및 ZNF18을 포함한다. 일부 실시양태에서, 캡시드는 적어도 1 nm, 2 nm, 3 nm, 4 nm, 5 nm, 10nm, 15 nm, 20 nm, 25 nm, 30 nm, 50 nm, 80 nm, 100 nm, 120 nm, 150 nm, 200 nm, 250 nm, 300 nm, 500 nm, 600 nm, 또는 그 초과와 직경을 갖는다. 일부 실시양태에서, 캡시드는 약 1 nm 내지 약 600 nm, 약 1 nm 내지 약 500 nm, 약 1 nm 내지 약 200 nm, 약 1 nm 내지 약 100 nm, 약 1 nm 내지 약 50 nm, 또는 약 1 nm 내지 약 30 nm의 직경을 갖는다. 일부 실시양태에서, 캡시드는 감소된 표적외(off-target) 효과를 갖는다. 일부 실시양태에서, 캡시드는 표적외 효과를 갖지 않는다. 일부 실시양태에서, 캡시드는 생체외에서 형성된다. 일부 실시양태에서, 캡시드는 시험관내에서 형성된다.

[0016] 특정 실시양태에서, 본원에 기재된 재조합 또는 조작된 Arc 폴리펩티드 또는 재조합 또는 조작된 내인성 Gag 폴리펩티드를 코딩하는 핵산 중합체가 본원에 개시된다.

[0017] 특정 실시양태에서, 본원에 기재된 재조합 또는 조작된 Arc 폴리펩티드 또는 재조합 또는 조작된 내인성 Gag 폴리펩티드를 코딩하는 핵산 중합체를 포함하는 벡터가 본원에 개시된다.

[0018] 특정 실시양태에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 또는 복수의 재조합 또는 조작된 endo-Gag 폴리펩티드를 로딩된 캡시드를 생성하기에 충분한 시간 동안 용액에서 카고와 함께 인큐베이션하는 단계를 포함하는, 로딩된 Arc 기반 또는 endo-Gag 기반 캡시드를 제조하는 방법이 본원에 개시된다. 일부 실시양태에서, 방법은 카고와 함께 인큐베이션하기 전에 복수의 조작된 Arc 또는 endo-Gag 폴리펩티드를 포함하는 용액을 복수의 비-Arc 또는 비-endo-Gag 캡시드 형성 서브유닛과 혼합하는 단계를 추가로 포함한다. 일부 실시양태에서, 복수의 비-Arc 또는 비-endo-Gag 캡시드 형성 서브유닛은 복수의 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드와 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 또는 10:1의 비로 혼합된다. 일부 실시양태에서, 복수의 비-Arc 또는 비-endo-Gag 캡시드 형성 서브유닛은 복수의 조작된 Arc 또는 endo-Gag 폴리펩티드와 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 또는 1:10의 비로 혼합된다. 일부 실시양태에서, 방법은 카고와 함께 인큐베이션하기 전에 복수의 말단절단된 Arc 또는 endo-Gag 폴리펩티드를 포함하는 용액을 복수의 비-Arc 또는 endo-Gag 캡시드 형성 서브유닛과 혼합하는 단계를 추가로 포함한다. 일부 실시양태에서, 복수의 비-Arc 또는 endo-Gag 캡시드 형성 서브유닛은 복수의 말단절단된 Arc 또는 endo-Gag 폴리펩티드와 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 또는 10:1의 비로 혼합된다. 일부 실시양태에서, 복수의 비-Arc 또는 비-endo-Gag 캡시드 형성 서브유닛은 복수의 말단절단된 Arc 또는 endo-Gag 폴리펩티드와 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 또는 1:10의 비로 혼합된다. 일부 실시양태에서, 복수의 조작된 Arc 또는 endo-Gag 폴리펩티드는 박테리아 세포 시스템, 곤충 세포 시스템, 또는 포유동물 세포 시스템으로부터 획득된다. 일부 실시양태에서, 복수의 조작된 Arc 또는 endo-Gag 폴리펩티드는 무세포 시스템으로부터 획득된다. 일부 실시양태에서, 복수의 말단절단된 Arc 또는 endo-Gag 폴리펩티드는 박테리아 세포 시스템, 곤충 세포 시스템, 또는 포유동물 세포 시스템으로부터 획득된다. 일부 실시양태에서, 복수의 말단절단된 Arc 또는 endo-Gag 폴리펩티드는 무세포 시스템으로부터 획득된다. 일부 실시양태에서, 로딩된 Arc 기반 또는 endo-Gag 캡시드는 전신 투여용으로 제제화된다. 일부 실시양태에서, 로딩된 Arc 또는 endo-Gag 기반 캡시드는 국소 투여용으로 제제화된다. 일부 실시양태에서, 로딩된 Arc 또는 endo-Gag 기반 캡시드는 비경구 투여용으로 제제화된다. 일부 실시양태에서, 로딩된 Arc 또는 endo-Gag 기반 캡시드는 경구 투여용으로 제제화된다. 일부 실시양태에서, 로딩된 Arc 또는 endo-Gag 기반 캡시드는 국부 투여용으로 제제화된다. 일부 실시양태에서, 로딩된 Arc 또는 endo-Gag 기반 캡시드는 설하 또는 에어로졸 투여용으로 제제화된다.

[0019] 특정 실시양태에서, 관심 부위의 세포를 Arc 기반 또는 Arc 기반 캡시드의 세포 흡수를 촉진하기에 충분한 시간 동안 Arc 기반 또는 Arc 기반 캡시드와 접촉시키는 단계를 포함하는, 카고를 관심 부위에 전달하기 위한 조작된 또는 재조합 Arc 기반 또는 endo-Gag 기반 캡시드의 용도가 본원에 개시된다. 일부 실시양태에서, 세포는 종양

세포이다. 일부 실시양태에서, 종양 세포는 고휘 종양 세포이다. 일부 실시양태에서, 고휘 종양 세포는 방광암, 유방암, 뇌암, 결장직장암, 신장암, 간암, 폐암, 췌장암, 전립선암, 피부암, 위암 또는 갑상선암으로부터의 세포이다. 일부 실시양태에서, 종양 세포는 혈액 악성종양으로부터의 것이다. 일부 실시양태에서, 혈액 악성종양은 B 세포 악성종양, 또는 T 세포 악성종양이다. 일부 실시양태에서, 혈액 악성종양은 만성 림프구성 백혈병 (CLL), 소 림프구성 림프종(SLL), 미만성 거대 B 세포 림프종(DLBCL), 소포 림프종, 맨틀 세포 림프종, 버킷 림프종, 피부 T 세포 림프종, 또는 말초 T 세포 림프종이다. 일부 실시양태에서, 세포는 체세포이다. 일부 실시양태에서, 세포는 줄기 세포 또는 전구 세포이다. 일부 실시양태에서, 세포는 중간엽 줄기 또는 전구 세포이다. 일부 실시양태에서, 세포는 조혈 줄기 또는 전구 세포이다. 일부 실시양태에서, 세포는 근육 세포, 피부 세포, 혈액 세포, 또는 면역 세포이다. 일부 실시양태에서, 표적 단백질은 과발현되거나 세포에서 고갈된다. 일부 실시양태에서, 세포의 표적 유전자는 하나 이상의 돌연변이를 갖는다. 일부 실시양태에서, 세포는 손상된 스플라이싱 메커니즘을 포함한다. 일부 실시양태에서, 용도는 생체내 용도이다. 일부 실시양태에서, Arc 기반 캡시드는 대상체에게 전신 투여된다. 일부 실시양태에서, Arc 기반 또는 endo-Gag 기반 캡시드는 대상체에게 국소 투여를 통해 투여된다. 일부 실시양태에서, Arc 기반 또는 endo-Gag 기반 캡시드는 대상체에게 비경구 투여된다. 일부 실시양태에서, Arc 기반 캡시드는 대상체에게 경구적으로 투여된다. 일부 실시양태에서, Arc 기반 또는 endo-Gag 기반 캡시드는 대상체에게 국부로 투여된다. 일부 실시양태에서, Arc 기반 또는 endo-Gag 기반 캡시드는 대상체에게 설하 또는 에어로졸 투여를 통해 투여된다. 일부 실시양태에서, 용도는 시험관내 또는 생체의 용도이다.

[0020] 특정 실시양태에서, 조작된 Arc 또는 endo-Gag 폴리펩티드, 말단절단된 Arc 또는 endo-Gag 폴리펩티드, 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드를 코딩하는 벡터, 또는 Arc 기반 또는 endo-Gag 기반 캡시드를 포함하는 키트가 본원에 개시된다.

**도면의 간단한 설명**

[0021] 본 개시내용의 다양한 측면은 첨부된 청구범위에서 구체적으로 설명된다. 본 개시내용의 특징 및 이점에 대한 더 나은 이해는 본 개시내용의 원리가 활용되는 예시적인 실시양태를 설명하는 하기의 상세한 설명 및 하기 첨부 도면을 참조함으로써 얻어질 것이다.

도 1은 예시적인 Arc 폴리펩티드의 표현이다.

도 2는 예시적인 조작된 Arc 폴리펩티드의 표현이다.

도 3은 (A) 특정 카고(예컨대, RNA 페이로드)을 운반하거나 (B) 기능의(off-function) 효과를 제거하기 위해 Arc 폴리펩티드를 조작하는 예시적인 방법을 예시한다.

도 4a는 HisTrap 컬럼으로부터 이미다졸 구배로의 용출에 의한 6xHis 태그부착된 인간 Arc의 단리를 도시한다.

도 4b는 NaCl 구배로 용출된 모노 Q 컬럼 상의 잔류 핵산으로부터 6xHis 태그부착된 인간 Arc의 분리를 도시한다.

도 5는 네거티브 염색된 인간 Arc 캡시드의 투과 전자 현미경 이미지를 도시한다.

도 6은 재조합적으로 발현된 Arc 동원체(ortholog)로부터 형성된 네거티브 염색된 캡시드의 투과 전자 현미경 이미지를 도시한다.

도 7은 재조합적으로 발현된 endo-Gag 단백질로부터 형성된 네거티브 염색된 캡시드의 투과 전자 현미경 이미지를 도시한다.

도 8은 HeLa 세포에 의한 Alexa594 표지된 Arc 캡시드의 선택적 내재화를 도시한다.

도 9는 Arc 캡시드에 의한 Cre RNA의 HeLa 세포로의 전달을 도시한다.

도 10은 이중 RNA 페이로드를 전달하는 능력에 대해 Arc 및 endo-Gag 유전자 후보를 스크리닝하는 방법을 예시한다.

**발명을 실시하기 위한 구체적인 내용**

[0022] 진단 또는 치료제를 관심 부위에 정밀하게 투여하는 것은 지속적인 도전을 제시하였다. 핵산을 세포에 전달하는 이용 가능한 방법은 무수한 한계를 갖는다. 예컨대, 유전자 요법에 자주 사용되는 AAV 바이러스 벡터는 면역원

성이고, < 3kb의 제한된 페이로드 용량을 갖고, 열악한 생체 분포를 겪고, 직접 주사로만 투여할 수 있으며, 통합에 의해 숙주 유전자를 파괴할 위험이 있다. 비-바이러스 방법은 상이한 한계를 갖는다. 리포솜은 주로 간에 전달된다. 세포의 소포는 < 1 kb의 제한된 페이로드 용량, 제한된 확장성 및 정제 어려움을 갖는다. 따라서, 치료 페이로드를 전달하는 새로운 방법에 대한 필요성이 인식된다.

[0023] 대부분의 분자는 신체에 고유한 친화력을 가지고 있지 않다. 다른 경우에서, 투여된 작용제는 청소를 위해 간 및 신장에 축적되거나 의도하지 않은 조직 또는 세포 유형에 축적된다. 전달을 개선하는 방법은 소수성 화합물 또는 중합체로 선택한 작용제를 코팅하는 것을 포함한다. 이러한 접근법은 순환에서 상기 작용제의 지속기간을 증가시키고 세포 흡수를 위한 소수성을 증대시킨다. 반면에, 이 접근법은 카고를 전달을 위한 관심 부위로 적극적으로 보내지 않는다.

[0024] 요법이 필요한 부위를 명확하게 표적으로 하기 위해, 치료 화합물은 임의적으로 표적 세포의 표면에 디스플레이 되는 수용체를 인식하고 결합하는 리간드, 항체 및 압타머와 같은 모이어티에 융합된다. 관심 세포에 도달하면, 치료 화합물은 임의적으로 세포내 표적에 추가로 전달된다. 예컨대, 치료 RNA는 세포의 세포질의 리포솜과 접촉하면 단백질로 번역될 수 있다.

[0025] Arc(활성 조절되는 세포골격 관련된 단백질)는 α-아미노-3-히드록시-5-메틸이속사졸-4-프로피온산(AMPA) 유형 글루타메이트 수용체의 세포내이입 트래피킹을 조절한다. Arc 활성화는 시냅스 강도 및 뉴런 가소성과 관련이 있다. 실험 무린 모델에서 Arc 손실의 표현형은 결손이 있는 장기 기억 형성 및 감소된 뉴런 활성화 및 가소성을 포함하였다.

[0026] Arc는 레트로바이러스 Gag 단백질과 유사한 분자 특성을 나타낸다. Arc 유전자는 Ty3/*gypsy* 레트로트랜스포존으로부터 유래되었을 수 있다. 내인성 Gag(endo-Gag) 단백질은 Arc를 포함하여 진행 유기체에 내인성인 임의의 단백질이며, 바이러스 Gag 단백질에 대해 예측된 주석이 달린 유사성을 갖는다. 예시적인 endo-Gag 단백질은 문헌 (Campillos M, Doerks T, Shah PK, and Bork P, Computational characterization of multiple Gag-like human proteins, Trends Genet. 2006 Nov;22(11):585-9)에 게시되어 있다. endo-Gag 단백질은 임의적으로 원핵 또는 진행 세포, 또는 박테리아, 효모, 곤충, 척추동물, 포유동물 또는 인간 세포를 포함하는 임의의 숙주 세포에 의해 재조합적으로 발현된다. 본원에서 기재되는 바와 같이, 일부 실시양태에서 endo-Gag 단백질은 endo-Gag 캡시드로 조립된다.

[0027] 특정 실시양태에서, 관심 카고의 전달을 위해 캡시드로 조립되는 Arc 및 endo-Gag 폴리펩티드가 본원에 게시된다. 일부 실시양태에서, 또한 관심 카고의 전달을 위해 캡시드로 조립되는 조작된 Arc 및 endo-Gag 폴리펩티드가 본원에 기재된다. 추가의 실시양태에서, 관심 카고의 전달을 위한 캡시드, 예컨대 Arc 기반 또는 endo-Gag 기반 캡시드가 본원에 기재된다.

[0028] **Arc 폴리펩티드 및 내인성 Gag 폴리펩티드**

[0029] 특정 실시양태에서, Arc 폴리펩티드가 본원에 게시된다. 특정 실시양태에서, endo-Gag 폴리펩티드가 본원에 게시된다. endo-Gag 서열은 이 섹션에 기재된 임의의 유형의 조작된 Arc 폴리펩티드를 형성하기 위한 Arc 서열에 대한 임의적 대체물이라는 것을 이해해야 한다.

[0030] 일부 예에서, Arc는 비-인간 Arc 폴리펩티드이다. 일부 예에서, Arc 폴리펩티드는 전장 Arc 폴리펩티드(예컨대, 전장 비-인간 Arc 폴리펩티드)를 포함한다. 다른 예에서, Arc 폴리펩티드는 캡시드의 형성에 참여하는 말단절단된 Arc 폴리펩티드와 같은 비-인간 Arc의 단편을 포함한다. 추가의 예에서, Arc 폴리펩티드는 비-인간 Arc 폴리펩티드의 하나 이상의 도메인을 포함하고, 여기서 도메인 중 적어도 하나는 캡시드의 형성에 참여한다. 추가의 예에서, Arc 폴리펩티드는 재조합 Arc 폴리펩티드이다.

[0031] 일부 예에서, endo-Gag는 비-인간 endo-Gag 폴리펩티드이다. 일부 예에서, endo-Gag 폴리펩티드는 전장 endo-Gag 폴리펩티드(예컨대, 전장 비-인간 endo-Gag 폴리펩티드)를 포함한다. 다른 예에서, endo-Gag 폴리펩티드는 캡시드의 형성에 참여하는 말단절단된 endo-Gag 폴리펩티드와 같은 비-인간 endo-Gag의 단편을 포함한다. 추가의 예에서, endo-Gag 폴리펩티드는 비-인간 endo-Gag 폴리펩티드의 하나 이상의 도메인을 포함하고, 여기서 도메인 중 적어도 하나는 캡시드의 형성에 참여한다. 추가의 예에서, endo-Gag 폴리펩티드는 재조합 endo-Gag 폴리펩티드이다.

[0032] 일부 실시양태에서, Arc는 인간 Arc에 고유하지 않은 카고에 결합하도록 변형된 적어도 그의 RNA 결합 도메인을 갖는 인간 Arc 폴리펩티드이다. 일부 예에서, Arc 폴리펩티드는 인간 Arc 단백질에 고유하지 않은 카고에 결합하도록 변형된 적어도 그의 RNA 결합 도메인을 갖는 전장 인간 Arc 폴리펩티드를 포함한다. 다른 예에서, Arc

폴리펩티드는 적어도 그의 RNA 결합 도메인에 변형(들)을 포함하는 인간 Arc 단편을 포함한다. 추가의 예에서, Arc 폴리펩티드는 인간 Arc 폴리펩티드의 하나 이상의 도메인을 포함하고, 여기서 도메인 중 적어도 하나는 캡시드의 형성에 참여하고, 여기서 RNA 결합 도메인은 천연 인간 Arc 단백질은 결합하지 않는 카고에 결합하도록 변형된다. 추가의 예에서, Arc 폴리펩티드는 재조합 인간 Arc 폴리펩티드이며, 적어도 RNA 결합 도메인은 인간 Arc 단백질에 고유하지 않은 카고를 로딩할 수 있도록 변형된다.

[0033] 일부 실시양태에서, Endo-Gag는 인간 endo-Gag에 고유하지 않은 카고에 결합하도록 변형된 적어도 그의 RNA 결합 도메인을 갖는 인간 Endo-Gag 폴리펩티드이다. 일부 예에서, endo-Gag 폴리펩티드는 인간 endo-Gag 단백질에 고유하지 않은 카고에 결합하도록 변형된 적어도 그의 RNA 결합 도메인을 갖는 전장 인간 endo-Gag 폴리펩티드를 포함한다. 다른 예에서, endo-Gag 폴리펩티드는 천연 인간 endo-Gag 단백질이 결합하지 않는 카고에 결합하도록 적어도 그의 RNA 결합 도메인에 변형(들)을 포함하는 인간 endo-Gag 단편을 포함한다. 추가의 예에서, endo-Gag 폴리펩티드는 인간 endo-Gag 폴리펩티드의 하나 이상의 도메인을 포함하고, 여기서 도메인 중 적어도 하나는 캡시드의 형성에 참여하고 여기서 RNA 결합 도메인은 인간 endo-Gag 단백질에 고유하지 않은 카고에 결합하도록 변형된다. 추가의 예에서, endo-Gag 폴리펩티드는 재조합 인간 endo-Gag 폴리펩티드이고, 적어도 RNA 결합 도메인은 인간 endo-Gag 단백질에 고유하지 않은 카고를 로딩할 수 있도록 변형된다.

[0034] 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 조작된 Arc 또는 endo-Gag 폴리펩티드이다. 본원에서 사용되는 바와 같이, 조작된 폴리펩티드는 전장 야생형 폴리펩티드와 서열이 동일하지 않은 재조합 폴리펩티드이다. 일부 예에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 제1 종으로부터의 Arc 또는 endo-Gag 폴리펩티드의 단편 및 제2 종의 Arc 또는 endo-Gag 폴리펩티드로부터의 적어도 추가의 단편을 포함한다. 일부 경우에서, 제1 Arc 또는 endo-Gag 폴리펩티드는 동물계, 식물계, 진균 또는 원생생물계의 계의 구성원으로부터 선택된다. 일부 경우에서, 제1 종은 포유동물, 설치류, 조류, 파충류, 어류, 척추동물, 진핵생물, 곤충, 진균, 또는 식물로부터 선택된다. 일부 경우에서, 제2 Arc 폴리펩티드는 제1 Arc 또는 endo-Gag 폴리펩티드와 동일하거나 상이한 동물계, 식물계, 진균 또는 원생생물계의 계의 구성원으로부터 선택된다. 일부 경우에서, 제2 종은 제1 종과 상이한 포유동물, 설치류, 조류, 파충류, 어류, 척추동물, 진핵생물, 곤충, 진균, 또는 식물로부터 선택된다.

[0035] 일부 실시양태에서, 재조합 또는 조작된 Arc 폴리펩티드로서의 발현하기 위한 예시적인 포유동물 Arc 또는 endo-Gag 단백질은 호모 사피엔스(*homo sapiens*) 종으로부터 것이다. 재조합 또는 조작된 Arc 폴리펩티드로서의 발현을 위한 영장류 Arc 또는 endo-Gag 단백질의 추가의 예시적인 종은 고릴라(*gorilla*), 판고 아벨리(*pongo abelii*), 판 파니쿠스(*pan paniscus*), 마카카 네메스트리나(*macaca nemestrina*), 클로로세부스 사바에우스(*chlorocebus sabaeus*), 파피오 아누비스(*papio anubis*), 리노피테쿠스 록셀라나(*rhinopithecus roxellana*), 마카카 파시쿨라리스(*macaca fascicularis*), 노마스쿠스 류코게니스(*nomascus leucogenys*), 칼리트릭스 자쿠스(*callithrix jacchus*), 아오투스 난시마에(*aotus nancymae*), 세부스 카푸시누스 이미타토르(*cebus capucinus imitator*), 사이미리 볼리비엔시스 볼리비엔시스(*saimiri boliviensis boliviensis*), 오톨레무르 가르네티이(*otolemur garnettii*), 마카카 몰라타(*macaca mulatta*), 및 마카카 파시쿨라리스(*macaca fascicularis*)를 포함한다.

[0036] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 설치류 Arc 또는 endo-Gag 단백질의 예시적인 종 목록은 푸코미스 다마렌시스(*fukomys damarensis*), 마이크로세버스 무리누스(*microcebus murinus*), 헤테로세팔루스 글라베르(*heterocephalus glaber*), 프로피테쿠스 코퀘렐리(*propithecus coquereli*), 마르모타 마르모타(*marmota marmota marmota*), 갈레오프테루스 바리에가투스(*galeopterus variegatus*), 카비아 포르셀루스(*cavia porcellus*), 디포도미스 오르디이(*dipodomys ordii*), 옥토돈 데구스(*octodon degus*), 카스토르 카나덴시스 나노스팔락스 갈릴리(*castor canadensis nannospalax galili*), 카를리토 시리차(*carlito syrichta*), 친칠라 라니게라(*chinchilla lanigera*), 무스 무스쿨루스(*mus musculus*), 익티도미스 트리데셀리네아투스(*ictidomys tridecemlineatus*), 라투스 노르베기쿠스(*rattus norvegicus*), 마이크로투스 오크로가스테르(*microtus ochrogaster*), 오톨레무르 가르네티이(*otolemur garnettii*), 메리오네스 운구이쿨라투스(*meriones unguiculatus*), 크리세툴루스 그리세우스(*crisetulus griseus*), 라투스 노르베기쿠스(*rattus norvegicus*), 네오토마 레피다(*neotoma lepida*), 자쿨루스 자쿨루스(*jaculus jaculus*), 무스텔라 푸토리우스 푸로(*mustela putorius furo*), 메소크리세투스 아우라투스(*mesocricetus auratus*), 투파이아 친넨시스(*tupaia chinensis*), 크리세툴루스 그리세우스(*crisetulus griseus*), 크리소클로리스 아시아티카(*chrysochloris asiatica*), 엘레판툴루스 에드워드디(*elephantulus edwardii*), 에리나세우스 에우로파에우스(*erinaceus europaeus*), 오초토나 프린셉스(*ochotona princeps*), 소렉스 아라네우스(*sorex araneus*), 모노델피스 도메스티카(*monodelphis domestica*), 에키노프스 텔파이리(*echinops telfairi*), 및 콘딜루라 크리스타타(*condylura cristata*)를 포함한다.

다.

- [0037] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 Arc 또는 endo-Gag 단백질의 예시적인 종 목록은 볼페스 볼페스(*vulpes vulpes*), 카니스 루푸스 딥고(*canis lupus dingo*), 펠리스 카투스(*felis catus*), 판테라 파르두스(*panthera pardus*), 칼로리누스 우르시누스(*callorhinus ursinus*), 오도베누스 로스마루스 디베르겐스(*odobenus rosmarus divergens*), 에쿠우 아시누스(*equus asinus*), 수스 스크로파(*sus scrofa*), 마니스 자바이카(*manis javanica*), 세라토테리움 시뮴 시뮴(*ceratotherium simum simum*), 렙토니코테스 웨델리이(*leptonychotes weddellii*), 엔히드라 루트리스 케노니(*enhydra lutris kenyonii*), 리포테스 복실리페르(*lipotes vexillifer*), 보스 그룬니엔스(*bos grunniens*), 부발루스 부발리스(*bubalus bubalis*), 카멜루스 드로메다리우스(*camelus dromedarius*), 비쿠그나 파코스(*vicugna pacos*), 오르시누스 오르카(*orcinus orca*), 네오모나쿠스 샤우인슬란드(*neomonachus schauinslandi*), 투르시오프스 트룬카투스(*tursiops truncatus*), 보스 타우루스(*bos taurus*), 카프라 히르쿠스(*capra hircus*), 델피나프테루스 레우카스(*delphinapterus leucas*), 오비스 아리에스 무시몬(*ovis aries musimon*), 발라에노프테라 아쿠토로스트라타 스캄모니(*balaenoptera acutorostrata scammoni*), 네오포카에나 아시아에오리엔탈리스 아시아에오리엔탈리스(*neophocaena asiaeorientalis asiaeorientalis*), 미니오프테루스 나탈렌시스(*miniopterus natalensis*), 프테로푸스 알렉토(*pteropus alecto*), 피세테르 카토돈(*physeter catodon*), 록소돈타 아프리카나(*loxodonta africana*), 오리테로푸스 아페르 아페르(*orycteropus afer afer*), 보스 무투스(*bos mutus*), 데스모두스 로툰두스(*desmodus rotundus*), 히포시데로스 아르미게르(*hipposideros armiger*), 아일루로포다 멜라노레우카(*ailuropoda melanoleuca*), 트리케쿠스 마나투스 라티로스트리스(*trichechus manatus latirostris*), 로우세투스 라티로스트리스(*rousettus latirostris*), 로우세투스 아에집티아쿠스(*rousettus aegyptiacus*), 에프테시쿠스 푸스쿠스(*eptesicus fuscus*), 리놀로푸스 시니쿠스(*rhinolophus sinicus*), 세르부스 엘라푸스 히펠라푸스(*cervus elaphus hippelaphus*), 오도코일레우스 비르기니아누스 텍사누스(*odocoileus virginianus texanus*), 판톨로프스 호드그소니아(*pantholops hodgsonii*), 카멜루스 박트리아누스(*camelus bactrianus*), 사르코필루스 하리시이(*sarcophilus harrisii*), 파스콜락토스 시네레우스(*phascolarctos cinereus*), 및 오르니토린쿠스 아나티누스(*ornithorhynchus anatinus*)를 포함한다.
- [0038] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 조류 Arc 또는 endo-Gag 단백질의 예시적인 종 목록은 갈루스 갈루스(*gallus gallus*), 코르부스 코르닉스(*corvus cornix*), 코르닉스(*cornix*), 파루스 메이저(*parus major*), 코르부스 브라키린코스(*corvus brachyrhynchos*), 드로마이우스 노바에홀란드리아에(*dromaius novaehollandiae*), 및 아프테릭스 로위(*apteryx rowi*)를 포함한다.
- [0039] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 파충류 Arc 단백질의 예시적인 종 목록은 피톤 비비타투스(*python bivittatus*), 포고나 비티셉스(*pogona vitticeps*), 아놀리스 카롤리넨시스(*anolis carolinensis*), 프로토보트로프스 무크로스쿠아마투스(*protobothrops mucrosquamatus*), 엘리게이터 시넨시스(*alligator sinensis*), 크로코딜루스 포로수스(*crocodylus porosus*), 가비알리스 간게티쿠스(*gavialis gangeticus*), 엘리게이터 미시시피엔시스(*alligator mississippiensis*), 펠로디스쿠스 시넨시스(*pelodiscus sinensis*), 테라페네 멕시코나 트리운구이스(*terrapene mexicana triunguis*), 크리세미스 픽타 벨리이(*chrysemys picta bellii*), 첼로니아 미다스(*chelonina mydas*), 나노라나 파케리(*nanorana parkeri*), 제노푸스 트로피칼리스(*xenopus tropicalis*), 제노푸스 라에비스(*xenopus laevis*), 및 라티메리아 칼룸나에(*latimeria chalumnae*)를 포함한다.
- [0040] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 어류 Arc 단백질의 예시적인 종 목록은 온코린쿠스 미키스(*oncorhynchus mykiss*), 아칸토크로미스 폴리아칸투스(*acanthochromis polyacanthus*), 온코린쿠스 키수치(*oncorhynchus kisutch*), 카라시우스 아우라투스(*carassius auratus*), 및 아우스트로퓌нду루스 림나에우스(*austrofundulus limnaeus*)를 포함한다.
- [0041] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 곤충 Arc 또는 endo-Gag 단백질의 예시적인 종 목록은 드로소필라 세라타(*drosophila serrata*), 드로소필라 비펙티나타(*drosophila bipectinata*), 솔레놉시스 인빅타(*solenopsis invicta*), 템노토락스 쿠르비스피노수스(*temnothorax curvispinosus*), 드로소필라 멜라노가스테르(*drosophila melanogaster*), 아그릴루스 플라니페니스(*agrilus planipennis*), 캄포토투스 플로리다누스(*camponotus floridanus*), 포고노미르멕스 바르바투스(*pogonomyrmex barbatus*), 닐라파르바타 루겐스(*nilaparvata lugens*), 봄빅스 모리(*bombyx mori*), 트리볼리움 카스타네움(*tribolium castaneum*), 및 렙티노타르사 데셈리네아타(*leptinotarsa decemlineata*)를 포함한다.

- [0042] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 식물 Arc 또는 endo-Gag 단백질의 예시적인 종 목록은 스피나시아 올레라세아(*spinacia oleracea*) 및 에리트란테 구타타(*erythranthe guttata*)를 포함한다.
- [0043] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 진균 단백질의 예시적인 종 목록은 사카로미세스 세레비시아에(*Saccharomyces cerevisiae*), 리조푸스 델레마르(*rhizopus delemar*), 푸사리움 옥시스포룸(*fusarium oxysporum*), 크립토크쿠스 네오포르만스(*cryptococcus neoformans*), 리조파구스 일레굴라리스(*rhizophagus irregularis*), 푸사리움 푸지쿠로이(*fusarium fujikuroi*), 칸디다 알비칸스(*candida albicans*), 트리코피톤 루브룸(*trichophyton rubrum*), 피레노포라 트리티시-레펜티스(*pyrenophora tritici-repentis*), 리조푸스 미크로스포루스(*rhizopus microsporus*), 리족토니아 솔라니(*rhizoctonia solani*), 아스페르길루스 플라부스(*aspergillus flavus*), 베르티실리움 달리아에(*verticillium dahliae*), 푸사리움 베르티실리오이데스(*fusarium verticillioides*), 아스페르길루스 니게르(*aspergillus niger*), 푸사리움 그라미네아룸(*fusarium graminearum*), 아스페르길루스 푸미가투스(*aspergillus fumigatus*), 지모셉토리아 트리티시(*zymoseptoria tritici*), 및 트리코데르마 하르시아눔(*trichoderma harzianum*)을 포함한다.
- [0044] 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드로서의 발현을 위한 원생생물 Arc 또는 endo-Gag 단백질의 예시적인 종 목록은 엔타모에바 히스톨리티카(*entamoeba histolytica*), 파울리넬라 미크로포라(*paulinella micropora*), 구일라르디아 테타(*guillardia theta*), 옥시르히스 마리나(*oxyrrhis marina*), 세미나비스 로부스타(*seminavis robusta*), 유글레나 롱가(*euglena longa*), 나에글레리아 그루베리(*naegleria gruberi*), 및 트리코모나스 바기날리스(*trichomonas vaginalis*)를 포함한다.
- [0045] 일부 예에서, Arc 또는 endo-Gag는 캡시드 조립/형성(CA) 도메인, 카고 결합 도메인(예컨대, RNA 결합 도메인) 및 임의적으로 매트릭스(MA) 도메인, 역전사효소(RT) 도메인, 또는 이들의 조합을 포함한다. 일부 경우에서, CA 도메인은 N-로브 도메인 및 C-로브 도메인으로 더 나뉜다. 일부 경우에서, 카고 결합 도메인은 RNA 결합 도메인, DNA 결합 도메인, 단백질 결합 도메인, 펩티드 결합 도메인, 항체 결합 도메인, 소분자 결합 도메인, 또는 펩티드모방체/뉴클레오티드모방체 결합 도메인을 포함한다. 예시적인 카고 결합 도메인은 GPCR, 항체 또는 이의 결합 단편, 지단백질, 인테그린, 티로신 키나제, DNA 결합 단백질, RNA 결합 단백질, 뉴클레아제, 리가제, 프로테아제, 인테그라제, 이소머라제, 포스파타제, GTPase, 아로마타제, 에스테라제, 어댑터 단백질, G-단백질, GEF, 사이토카인, 인터류킨, 인터류킨 수용체, 인터페론, 인터페론 수용체, 카스파제, 전사 인자, 신경영양 인자 및 그들의 수용체, 성장 인자 및 그들의 수용체, 신호 인식 입자 및 수용체 성분, 세포외 매트릭스 단백질, 막의 통합 성분, 리보솜 단백질, 번역 신장 인자, 번역 개시 인자, GPI 고정된 단백질, 조직 인자, 디스트로핀, 우트로핀, 디스트로브레틴, 이들의 임의의 융합, 조합, 서브유닛, 유도체 또는 도메인으로부터의 도메인을 포함하지만 이에 제한되지 않는다.
- [0046] 일부 실시양태에서, 캡시드 형성 또는 핵산 결합에 관여하지 않는 하나 이상의 비-필수 영역은 Arc 또는 endo-Gag 폴리펩티드를 생성하기 위해 Arc 또는 endo-Gag 단백질로부터 제거된다. 이러한 경우에서, 하나 이상의 비-필수 영역, 예컨대 N-말단 영역(예컨대, 최대 10개 아미노산, 최대 20개 아미노산, 최대 30개 아미노산, 또는 최대 50개 아미노산), C-말단 영역(예컨대, 최대 10개 아미노산, 최대 20개 아미노산, 최대 30개 아미노산, 또는 최대 50개 아미노산), RT 도메인, MA 도메인, 또는 이들의 조합이 Arc 또는 endo-Gag 폴리펩티드를 생성하기 위해 Arc 또는 endo-Gag 단백질로부터 삭제된다. 일부 경우에서, 캡시드 조립/형성 및 카고 결합과 관련된 필수 영역만 Arc 또는 endo-Gag 폴리펩티드에 잔류한다. 추가의 경우에서, 캡시드 조립/형성에 관여하는 필수 영역(예컨대, N-로브 및/또는 C-로브)만 Arc 폴리펩티드에 잔류한다.
- [0047] 특정 실시양태에서, RT 도메인, MA 도메인, 및/또는 내인성 RNA 결합 도메인은 다른 카고 결합 도메인으로 대체된다: 예컨대, DNA 결합 도메인, 단백질 결합 도메인, 펩티드 결합 도메인, 항체 결합 도메인, 소분자 결합 도메인, 펩티드모방체 결합 도메인, 또는 뉴클레오티드모방체 결합 도메인으로 대체된다. 일부 실시양태에서, Arc 또는 endo-Gag 폴리펩티드는 캡시드 형성, 핵산 결합, 또는 전달에 관여하는 도메인의 말단절단 또는 변형을 포함한다.
- [0048] 일부 실시양태에서, Arc 또는 endo-Gag 폴리펩티드는 MA 도메인, CA N-로브, CA C-로브, 카고 결합 도메인, 및 RT 도메인을 포함한다. 일부 예에서, Arc 폴리펩티드는 N-말단에서 C-말단으로 MA 도메인, CA N-로브, CA C-로브, RT 도메인 및 카고 결합 도메인을 포함한다. 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 MA 도메인, RT 도메인, 카고 결합 도메인, CA N-로브 및 CA C-로브를 포함한다. 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 카고 결합 도메인, MA 도메인, RT 도메인, CA N-로브 및 CA

C-로브를 포함한다. 일부 예에서, 도메인은 캡시드 조립 및 카고 결합을 방해하지 않는 순서로 배열된다. 일부 예에서, 각각의 도메인은 각각의 2개의 플랭킹 도메인에 직접 또는 간접적으로 융합된다.

[0049] 일부 실시양태에서, Arc 또는 endo-Gag 폴리펩티드는 MA 도메인, CA N-로브, CA C-로브, 및 카고 결합 도메인을 포함한다. 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 MA 도메인, CA N-로브, CA C-로브, 및 카고 결합 도메인을 포함한다. 일부 경우에서, Arc 폴리펩티드는 N-말단에서 C-말단으로 MA 도메인, 카고 결합 도메인, CA N-로브 및 CA C-로브 도메인을 포함한다. 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 카고 결합 도메인, MA 도메인, CA N-로브, 및 CA C-로브를 포함한다. 일부 예에서, 도메인은 캡시드 조립 및 카고 결합을 방해하지 않는 순서로 배열된다. 일부 예에서, 각각의 도메인은 각각의 2개의 플랭킹 도메인에 직접 또는 간접적으로 융합된다.

[0050] 일부 실시양태에서, Arc 또는 endo-Gag 폴리펩티드는 CA N-로브, CA C-로브 및 카고 결합 도메인을 포함한다. 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 CA N-로브, CA C-로브 및 카고 결합 도메인을 포함한다. 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 카고 결합 도메인, CA N-로브 및 CA C-로브를 포함한다. 일부 예에서, 도메인은 캡시드 조립 및 카고 결합을 방해하지 않는 순서로 배열된다. 일부 예에서, 각각의 도메인은 각각의 2개의 플랭킹 도메인에 직접 또는 간접적으로 융합된다.

[0051] 일부 실시양태에서, Arc 또는 endo-Gag 폴리펩티드는 CA N-로브 및 카고 결합 도메인을 포함한다. 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 CA N-로브 및 카고 결합 도메인을 포함한다. 일부 예에서, Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 카고 결합 도메인 및 CA N-로브를 포함한다. 일부 예에서, 도메인은 캡시드 조립 및 카고 결합을 방해하지 않는 순서로 배열된다. 일부 경우에서, 2개의 도메인은 직접 또는 간접적으로 서로 융합된다.

[0052] 일부 실시양태에서, Arc 또는 endo-Gag 폴리펩티드는 조작된 Arc 폴리펩티드를 생성하기 위해 하나 이상의 추가의 종으로부터의 카고 결합 도메인, CA 도메인, MA 도메인, 또는 RT 도메인을 포함하도록 조작된다. 예컨대, 조작된 Arc 또는 endo-Gag 폴리펩티드는 제1 종으로부터의 카고 결합 도메인, CA 도메인, MA 도메인, 또는 RT 도메인 및 제2 종으로부터의 카고 결합 도메인, CA 도메인, MA 도메인, 또는 RT 도메인을 포함한다. 일부 경우에서, 제1 종은 진핵생물, 척추동물, 인간, 포유동물, 설치류, 조류, 파충류, 어류, 곤충, 진균, 또는 식물로부터 선택된다. 일부 경우에서, 제2 종은 제1 종과 상이한 진핵생물, 척추동물, 인간, 포유동물, 설치류, 조류, 파충류, 어류, 곤충, 진균, 또는 식물로부터 선택된다.

[0053] 일부 예에서, 조작된 또는 endo-Gag Arc 폴리펩티드는 제1 종으로부터의 카고 결합 도메인 및 제2 종으로부터의 CA 도메인(예컨대, CA N-로브 및 임의적으로 CA C-로브)을 포함한다. 조작된 Arc 또는 endo-Gag 폴리펩티드는 임의적으로 제1 종 또는 제2 종으로부터의 MA 도메인 및 RT 도메인을 포함한다. 일부 경우에서, 제1 종은 진핵생물, 척추동물, 인간, 포유동물, 설치류, 조류, 파충류, 어류, 곤충, 진균, 또는 식물로부터 선택된다. 일부 경우에서, 제2 종은 제1 종과 상이한 진핵생물, 척추동물, 인간, 포유동물, 설치류, 조류, 파충류, 어류, 곤충, 진균, 또는 식물로부터 선택된다.

[0054] 일부 예에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 카고 결합 도메인, 제1 CA 도메인, 제2 CA 도메인, 및 임의적으로 MA 도메인 및/또는 RT 도메인을 포함한다. 일부 경우에서, 카고 결합 도메인, 제1 CA 도메인, 및 임의적으로 MA 도메인 및/또는 RT 도메인은 제1 종으로부터의 것이고 제2 CA 도메인은 제2 종으로부터의 것이다. 일부 경우에서, 제1 CA 도메인은 제1 종으로부터의 것이고 카고 결합 도메인, 제2 CA 도메인, 및 임의적으로 MA 도메인 및/또는 RT 도메인은 제2 종으로부터의 것이다. 일부 예에서, 도메인은 캡시드 조립 및 카고 결합을 방해하지 않는 순서로 배열된다. 일부 예에서, 각각의 도메인은 각각의 2개의 인접한 도메인에 직접 또는 간접적으로 융합된다.

[0055] 일부 예에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 카고 결합 도메인, 제1 CA 도메인 및 제2 CA 도메인을 포함한다. 일부 경우에서, 카고 결합 도메인 및 제1 CA 도메인은 제1 종으로부터의 것이고 제2 CA 도메인은 제2 종으로부터의 것이다. 일부 경우에서, 제1 CA 도메인은 제1 종으로부터의 것이고 카고 결합 도메인 및 제2 CA 도메인은 제2 종으로부터의 것이다. 이러한 경우에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 카고 결합 도메인, 제1 CA 도메인 및 제2 CA 도메인을 포함한다. 이러한 경우에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 N-말단에서 C-말단으로 제1 CA 도메인, 제2 CA 도메인, 및 카고 결합 도메인을 포함한다. 일부 예에서, 도메인은 캡시드 조립 및 카고 결합을 방해하지 않는 순서로 배열된다. 일부 예에서, 각각의 도메인은 각각의 2개의 플랭킹 도메인에 직접 또는 간접적으로 융합된다.

합된다.

- [0056] 일부 예에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 제2 폴리펩티드를 추가로 포함한다. 일부 예에서, 제2 폴리펩티드는 링커를 통해 카고 결합 도메인, 제1 CA 도메인, 제2 CA 도메인, 존재하는 경우 MA 도메인, 또는 존재하는 경우 RT 도메인 중 하나 이상에 직접 또는 간접적으로 융합된다. 일부 경우에서, 제2 폴리펩티드는 단백질(예컨대, 인간 단백질), 항체 또는 이의 결합 단편, 바이러스 단백질, Gag-유사 단백질(예컨대, 인간 Gag-유사 단백질), 또는 관심 표적 수용체에 결합하도록 설계된 드 노보 조작된 단백질을 포함한다. 일부 예에서, 항체 또는 이의 결합 단편은 인간화 항체 또는 이의 결합 단편, 무린 항체 또는 이의 결합 단편, 키메라 항체 또는 이의 결합 단편, 모노클로날 항체 또는 이의 결합 단편, 다중특이적 항체 또는 이의 단편 결합, 이중특이적 항체 또는 이의 결합 단편, 1가 Fab', 2가 Fab<sub>2</sub>, F(ab')<sub>3</sub> 단편, 단일 쇠 가변 단편(scFv), 비스-scFv, (scFv)<sub>2</sub>, 디아바디, 미니바디, 나노바디, 트리아바디, 테트라바디, 디스파이드 안정화된 Fv 단백질(dsFv), 단일 도메인 항체(sdAb), Ig NAR, 낙타류 항체 또는 이의 결합 단편, 또는 이의 화학적으로 변형된 유도체를 포함한다. 일부 예에서, 제2 폴리펩티드는 조작된 Arc 폴리펩티드에 의해 형성된 캡시드를 관심 표적 부위에 전달하는 것을 안내한다.
- [0057] 일부 실시양태에서, (예컨대, Arc 폴리펩티드 또는 endo-Gag 폴리펩티드를 코딩하는) 본 개시 내용의 핵산 서열 또는 아미노산 서열은 본원에 제공된 아미노산 서열과 적어도 70% 상동성, 적어도 71% 상동성, 적어도 72% 상동성, 적어도 73% 상동성, 적어도 74% 상동성, 적어도 75% 상동성, 적어도 76% 상동성, 적어도 77% 상동성, 적어도 78% 상동성, 적어도 79% 상동성, 적어도 80% 상동성, 적어도 81% 상동성, 적어도 82% 상동성, 적어도 83% 상동성, 적어도 84% 상동성, 적어도 85% 상동성, 적어도 86% 상동성, 적어도 87% 상동성, 적어도 88% 상동성, 적어도 89% 상동성, 적어도 90% 상동성, 적어도 91% 상동성, 적어도 92% 상동성, 적어도 93% 상동성, 적어도 94% 상동성, 적어도 95% 상동성, 적어도 96% 상동성, 적어도 97% 상동성, 적어도 98% 상동성, 적어도 99% 상동성, 적어도 99.1% 상동성, 적어도 99.2% 상동성, 적어도 99.3% 상동성, 적어도 99.4% 상동성, 적어도 99.5% 상동성, 적어도 99.6% 상동성, 적어도 99.7% 상동성, 적어도 99.8% 상동성, 적어도 99.9% 또는 적어도 99.99% 상동성을 갖는다. NCBI BLAST, 클루스탈(Clustal) W, MAFFT, 클루스탈 오메가(Clustal Omega), AlignMe, 프랄린 (Praline) 또는 또 다른 적절한 방법 또는 알고리즘과 같은 2개 이상의 서열 간의 상동성을 결정하기 위해 다양한 방법 및 소프트웨어 프로그램이 사용된다.
- [0058] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 1의 아미노산 서열 또는 서열번호 1과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 폴리펩티드이다.
- [0059] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 2의 아미노산 서열 또는 서열번호 2와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 범고래 폴리펩티드이다.
- [0060] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 3의 아미노산 서열 또는 서열번호 3과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 흰꼬리 사슴 폴리펩티드이다.
- [0061] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 4의 아미노산 서열 또는 서열번호 4와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 오리너구리 폴리펩티드이다.
- [0062] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 5의 아미노산 서열 또는 서열번호 5와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 거위 폴리펩티드이다.
- [0063] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 6의 아미노산 서열 또는 서열번호 6과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 달마티안 펠리칸(Dalmatian pelican) 폴리펩티드이다.
- [0064] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 7의 아미노산 서열 또는 서열번호 7과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 흰꼬리 수리 폴리펩티드이다.
- [0065] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 8의 아미노산 서열 또는 서열번호 8과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 킹 코브라 폴리펩티드이다.
- [0066] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 9의 아미노산 서열 또는 서열번호 9와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 조기어류(Ray finned fish) 폴리펩티드이다.

- [0067] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 10의 아미노산 서열 또는 서열번호 10과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 향유 고래 폴리펩티드이다.
- [0068] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 11의 아미노산 서열 또는 서열번호 11과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 칠면조 폴리펩티드이다.
- [0069] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 12의 아미노산 서열 또는 서열번호 12와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 중앙 턱수염 도마뱀(central bearded dragon) 폴리펩티드이다.
- [0070] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 13의 아미노산 서열 또는 서열번호 13과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 양쯔강 악어(chinese alligator) 폴리펩티드이다.
- [0071] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 14의 아미노산 서열 또는 서열번호 14와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 미국 악어(American alligator) 폴리펩티드이다.
- [0072] 특정 실시양태에서, Arc 폴리펩티드는 서열번호 15의 아미노산 서열 또는 서열번호 15와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 일본 게코(Japanese gekko) 폴리펩티드이다.
- [0073] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 16의 아미노산 서열 또는 서열번호 16과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 PNMA3 폴리펩티드이다.
- [0074] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 17의 아미노산 서열 또는 서열번호 17과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 PNMA5 폴리펩티드이다.
- [0075] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 18의 아미노산 서열 또는 서열번호 18과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 PNMA6A 폴리펩티드이다.
- [0076] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 19의 아미노산 서열 또는 서열번호 19와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 PNMA6B 폴리펩티드이다.
- [0077] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 20의 아미노산 서열 또는 서열번호 20과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 RTL3 폴리펩티드이다.
- [0078] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 21의 아미노산 서열 또는 서열번호 21과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 RTL6 폴리펩티드이다.
- [0079] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 22의 아미노산 서열 또는 서열번호 22와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 RTL8A 폴리펩티드이다.
- [0080] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 23의 아미노산 서열 또는 서열번호 23과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 RTL8B 폴리펩티드이다.
- [0081] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 24의 아미노산 서열 또는 서열번호 24와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 BOP 폴리펩티드이다.
- [0082] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 25의 아미노산 서열 또는 서열번호 25와 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 LDOC1 폴리펩티드이다.
- [0083] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 26의 아미노산 서열 또는 서열번호 26과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 ZNF18 폴리펩티드이다.
- [0084] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 27의 아미노산 서열 또는 서열번호 27과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 MOAP1 폴리펩티드이다.
- [0085] 특정 실시양태에서, endo-Gag 폴리펩티드는 서열번호 28의 아미노산 서열 또는 서열번호 28과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 인간 PEG10 폴리펩티드이다.

- [0086] 일부 경우에서, 재조합 Arc 또는 endo-Gag 폴리펩티드는 도 1에 예시되는 Arc 폴리펩티드이다.
- [0087] 일부 경우에서, 조작된 Arc 또는 endo-Gag 폴리펩티드는 도 2에 예시되는 조작된 Arc 폴리펩티드이다.
- [0088] 링커
- [0089] 특정 실시양태에서, 본 개시 내용의 폴리펩티드는 링커를 포함한다. 일부 실시양태에서, 링커는 펩티드 링커이다. 일부 예에서, 링커는 경성 링커이다. 다른 예에서, 링커는 연성 링커이다. 일부 경우에서, 링커는 비-절단 가능한 링커이다. 다른 경우에서, 링커는 절단 가능한 링커이다. 추가의 경우에서, 링커는 선형 구조 또는 비-선형 구조(예컨대, 사이클릭 구조)를 포함한다.
- [0090] 특정 실시양태에서, 비-절단 가능한 링커는 다양한 길이의 짧은 펩티드를 포함한다. 예시적인 비-절단 가능한 링커는 (EAAAK)<sub>n</sub>(서열번호 70) 또는 (EAAAR)<sub>n</sub>(서열번호 71)을 포함하며, 여기서 n은 1 내지 5이고, 글루탐산-프롤린 잔기 또는 리신-프롤린 반복의 최대 30개 잔기이다. 일부 실시양태에서, 비-절단 가능한 링커는 (GGGS)<sub>n</sub>(서열번호 72) 또는 (GGG)<sub>n</sub>(서열번호 73)을 포함하며, 여기서 n은 1 내지 10이고; KESGSVSSEQLAQFRSLD(서열번호 74); 또는 EGKSSGSGSESKST(서열번호 75)이다. 일부 실시양태에서, 비-절단 가능한 링커는 폴리-Gly/Ala 중합체를 포함한다.
- [0091] 특정 실시양태에서, 링커는 절단 가능한 링커, 예컨대 세포의 절단 가능한 링커 또는 세포내 절단 가능한 링커이다. 일부 예에서, 링커는 특정 조건의 존재 또는 특정 환경(예컨대, 생리적 조건 하)에서의 절단을 위해 설계된다. 예컨대, 특정 효소에 의하는 것과 같은 특정 조건에 의한 절단을 위한 링커의 설계는 특정 위치에 대한 세포 흡수의 표적화를 가능하게 한다.
- [0092] 일부 실시양태에서, 링커는 pH 민감성 링커이다. 한 예에서, 링커는 염기성 pH 조건 하에서 절단된다. 다른 예에서, 링커는 산성 pH 조건 하에서 절단된다.
- [0093] 일부 실시양태에서, 링커는 트롬빈, 메탈로프로테아제, 푸린, 카텡신 B, 피사 효소(예컨대, 칼파인) 등을 포함하지만 이에 제한되지 않는 세린 프로테아제와 같은 내인성 효소(예컨대, 프로테아제)에 의해 생체내에서 절단된다. 예시적인 절단 가능한 링커는 GGAANLVRGG(서열번호 76); SGRIGFLRTA(서열번호 77); SGRSA(서열번호 78); GFLG(서열번호 79); ALAL(서열번호 80); FK; PIC(Et)F-F(서열번호 81)(여기서 C(Et)는 S-에틸시스테인을 나타냄); PR(S/T)(L/I)(S/T)(서열번호 82); DEVD(서열번호 83); GWEHDG(서열번호 84); RPLALWRS(서열번호 85); 또는 이들의 조합을 포함하지만 이에 제한되지 않는다.
- [0094] 캡시드
- [0095] 일부 실시양태에서, 캡시드가 본원에 개시된다. 일부 예에서, 캡시드는 Arc 폴리펩티드 및/또는 endo-Gag 폴리펩티드, 예컨대 코피아 단백질, ASPRV1 단백질, SCAN 도메인 패밀리로부터의 단백질, 부신생물성 Ma 항원 패밀리에 의해 코딩되는 단백질, 레트로트랜스포존 Gag-유사 패밀리로부터 선택되는 단백질 또는 단백질의 조합, 또는 이들의 조합을 포함한다. 예시적인 endo-Gag 폴리펩티드는 BOP, LDOC1, MOAP1, PEG10, PNMA3, PNMA5, PNMA6A, PNMA6B, RTL3, RTL6, RTL8A, RTL8B, 및 ZNF18이다. 일부 예에서, Arc 폴리펩티드, 코피아 단백질, ASPRV1 단백질, SCAN 도메인 패밀리로부터의 단백질, 부신생물성 Ma 항원 패밀리에 의해 코딩되는 단백질, 및 레트로트랜스포존 Gag-유사 패밀리로부터 선택되는 단백질 또는 단백질의 조합은 각각 독립적으로 전장 폴리펩티드이다. 다른 예에서, Arc 폴리펩티드, 코피아 단백질, ASPRV1 단백질, SCAN 도메인 패밀리로부터의 단백질, 부신생물성 Ma 항원 패밀리에 의해 코딩되는 단백질, 및 레트로트랜스포존 Gag-유사 패밀리로부터 선택되는 단백질 또는 단백질의 조합은 각각 독립적으로, 예컨대 캡시드의 서브유닛을 형성할 수 있는, 이의 기능적 단편이다.
- [0096] Arc 기반 캡시드 및 endo-Gag 기반 캡시드
- [0097] 일부 실시양태에서, 캡시드는 Arc 기반 캡시드를 포함한다. 일부 실시양태에서, 캡시드는 endo-Gag 기반 캡시드를 포함한다. 일부 예에서, Arc 기반 및/또는 endo-Gag 캡시드는 상술된 복수의 재조합 Arc 폴리펩티드 및/또는 endo-Gag 폴리펩티드, 상술된 복수의 조작된 Arc 폴리펩티드 및/또는 endo-Gag 폴리펩티드, 또는 이들의 조합을 포함한다. 일부 경우에서, Arc 기반 캡시드는 복수의 재조합 Arc 폴리펩티드를 포함한다. 다른 경우에서, Arc 기반 캡시드는 복수의 조작된 Arc 폴리펩티드를 포함한다. 일부 경우에서, endo-Gag 기반 캡시드는 복수의 재조합 endo-Gag 폴리펩티드를 포함한다. 다른 경우에서, endo-Gag 기반 캡시드는 복수의 조작된 endo-Gag 폴리펩티드를 포함한다.
- [0098] 일부 실시양태에서, Arc 기반 또는 endo-Gag 기반 캡시드는 제1 종으로부터의 제1 복수의 Arc 및/또는 endo-Gag

폴리펩티드 및 적어도 제2 종으로부터의 제2 복수의 Arc 및/또는 endo-Gag 폴리펩티드를 포함한다. 일부 경우에서, 제1 종은 진핵생물, 척추동물, 인간, 포유동물, 설치류, 조류, 파충류, 어류, 곤충, 진균, 또는 식물로부터 선택된다. 일부 경우에서, 제2 종은 제1 종과 상이한 진핵생물, 척추동물, 인간, 포유동물, 설치류, 조류, 파충류, 어류, 곤충, 진균, 또는 식물로부터 선택된다.

[0099] 일부 예에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 20:1, 50:1, 또는 100:1이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1:1이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 2:1이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 4:1이다. 일부 경우에서, 복수의 제1 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 5:1이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 8:1이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 복수의 제2 Arc 폴리펩티드의 비는 10:1이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 20:1이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 50:1이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 100:1이다. 일부 예에서, 비는 몰 농도의 비교이다. 일부 예에서, 비는 캡시드 형성 서브유닛의 수의 비교이다(예컨대, 각각의 또는 조작된 Arc 폴리펩티드는 캡시드 서브유닛을 형성함).

[0100] 일부 예에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1:2, 1:3, 1:4, 1 : 5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:20, 또는 1:50이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1:2이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1 : 5이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1:8이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1:10이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1:20이다. 일부 경우에서, 제1 복수의 Arc 또는 endo-Gag 폴리펩티드 대 제2 복수의 Arc 또는 endo-Gag 폴리펩티드의 비는 1:50이다. 일부 예에서, 비는 몰 농도의 비교이다. 일부 예에서, 비는 캡시드 형성 서브유닛의 수의 비교이다(예컨대, 각각의 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드는 캡시드 서브유닛을 형성함).

[0101] 일부 실시양태에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 복수의 재조합 또는 조작된 Arc 폴리펩티드 및 복수의 비-Arc 단백질을 포함한다. 비-Arc 단백질의 예시적인 종은 코피아, ASPRV1, SCAN 도메인 패밀리로부터 선택되는 단백질 또는 단백질의 조합, 부신생물성 Ma 항원 패밀리로부터 선택되는 단백질 또는 단백질의 조합, 및 레트로트랜스포존 Gag-유사 패밀리로부터 선택되는 단백질 또는 단백질의 조합을 포함하지만 이에 제한되지 않는다. 비-Arc 단백질의 예시적인 종은 BOP, LDOC1, MOAP1, PEG10, PNMA3, PNMA5, PNMA6A, PNMA6B, RTL3, RTL6, RTL8A, RTL8B, 및 ZNF18을 포함한다.

[0102] 일부 예에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 1:1, 2:1, 3:1, 4:1, 5:1, 6:1이다. 7:1, 8:1, 9:1, 10:1, 20:1, 50:1 또는 100:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 1:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 2:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 4:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 5:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 8:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 10:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 20:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 50:1이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 100:1이다. 일부 예에서, 비는 몰 농도의 비교이다. 일부 예에서, 비는 캡시드 형성 서브유닛의 수의 비교이다(예컨대, 각각의 재조합 또는 조작된 Arc 폴리펩티드는 캡시드 서브유닛을 형성함).

[0103] 일부 예에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:20, 또는 1:50이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드

대 복수의 비-Arc 단백질의 비는 1:2이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 1:5이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 1:8이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 1:10이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 1:20이다. 일부 경우에서, 복수의 재조합 또는 조작된 Arc 폴리펩티드 대 복수의 비-Arc 단백질의 비는 1:50이다. 일부 예에서, 비는 몰 농도의 비교이다. 일부 예에서, 비는 캡시드 형성 서브유닛의 수의 비교이다(예컨대, 각각의 재조합 또는 조작된 Arc 폴리펩티드는 캡시드 서브유닛을 형성함).

[0104] 일부 실시양태에서, 캡시드는 적어도 1 nm의 직경을 갖는다. 일부 예에서, 캡시드는 적어도 2 nm, 3 nm, 4 nm, 5 nm, 10 nm, 15 nm, 20 nm, 25 nm, 30 nm, 40 nm, 50 nm, 60 nm, 70 nm, 80 nm, 90 nm, 100 nm, 150 nm, 200 nm, 300 nm, 400 nm, 500 nm, 600 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 5 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 10 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 20 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 30 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 40 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 50 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 80 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 100 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 200 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 300 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 400 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 500 nm, 또는 그 초과 직경을 갖는다. 일부 예에서, 캡시드는 적어도 600 nm, 또는 그 초과 직경을 갖는다.

[0105] 일부 실시양태에서, 캡시드는 최대 1 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 2 nm, 3 nm, 4 nm, 5 nm, 10 nm, 15 nm, 20 nm, 25 nm, 30 nm, 40 nm, 50 nm, 60 nm, 70 nm, 80 nm, 90 nm, 100 nm, 150 nm, 200 nm, 300 nm, 400 nm, 500 nm, 600 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 5 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 10 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 20 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 30 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 40 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 50 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 80 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 100 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 200 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 400 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 500 nm, 또는 그 미만의 직경을 갖는다. 일부 예에서, 캡시드는 최대 600 nm, 또는 그 미만의 직경을 갖는다.

[0106] 일부 실시양태에서, 캡시드는 약 1 nm, 2 nm, 3 nm, 4 nm, 5 nm, 10 nm, 15 nm, 20 nm, 25 nm, 30 nm, 40 nm, 50 nm, 60 nm, 70 nm, 80 nm, 90 nm, 100 nm, 150 nm, 200 nm, 300 nm, 400 nm, 500 nm, 또는 600 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 5 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 10 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 20 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 30 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 40 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 50 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 80 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 100 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 200 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 300 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 400 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 500 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 600 nm의 직경을 갖는다.

[0107] 일부 실시양태에서, 캡시드는 약 1 nm 내지 약 600 nm의 직경을 갖는다. 일부 예에서, 캡시드는 약 2 nm 내지 약 500 nm, 약 2 nm 내지 약 400 nm, 약 2 nm 내지 약 300 nm, 약 2 nm 내지 약 200 nm, 약 2 nm 내지 약 100 nm, 약 2 nm 내지 약 50 nm, 약 2 nm 내지 약 30 nm, 약 20 nm 내지 약 400 nm, 약 20 nm 내지 약 300 nm, 약 20 nm 내지 약 200 nm, 약 20 nm 내지 약 100 nm, 약 20 nm 내지 약 50 nm, 약 20 nm 내지 약 30 nm, 약 30 nm 내지 약 500 nm, 약 30 nm 내지 약 400 nm, 약 30 nm 내지 약 300 nm, 약 30 nm 내지 약 200 nm, 약 30 nm 내지 약 100 nm, 약 30 nm 내지 약 50 nm, 약 50 nm 내지 약 300 nm, 약 50 nm 내지 약 200 nm, 약 50 nm 내지 약 100 nm, 약 2 nm 내지 약 25 nm, 약 2 nm 내지 약 20 nm, 약 2 nm 내지 약 10 nm, 약 5 nm 내지 약 25 nm, 약 5 nm 내지 약 20 nm, 약 5 nm 내지 약 10 nm, 약 10 nm 내지 약 25 nm, 또는 약 10 nm 내지 약 20 nm의 직경을 갖는다.

[0108] 일부 실시양태에서, 캡시드는 감소된 표적의 효과를 갖는다. 일부 경우에서, 표적의 효과는 10%, 5%, 4%, 3%,

2%, 1% 또는 0.5% 미만이다. 일부 경우에서, 표적의 효과는 10%, 5%, 4%, 3%, 2%, 1% 또는 0.5% 이하이다.

- [0109] 일부 경우에 캡시드는 표적의 효과를 갖지 않는다.
- [0110] 특정 실시양태에서, Arc 및/또는 endo-Gag 기반 캡시드의 형성은 생체의 또는 시험관내에서 발생한다.
- [0111] 일부 예에서, Arc 및/또는 endo-Gag 기반 캡시드는 생체내에서 조립된다.
- [0112] 일부 예에서, Arc 및/또는 endo-Gag 기반 캡시드는 실온에서 안정하다. 일부 경우에서, Arc 및/또는 endo-Gag 기반 캡시드는 비어 있다. 다른 경우에서, Arc 및/또는 endo-Gag 기반 캡시드는 로딩된다(예컨대, 카고 및/또는 치료제, 예컨대 DNA 또는 RNA가 로딩됨).
- [0113] 일부 예에서, Arc 및/또는 endo-Gag 기반 캡시드는 약 2°C 내지 약 37°C의 온도에서 안정하다. 일부 예에서, Arc 및/또는 endo-Gag 기반 캡시드는 약 2°C 내지 약 8°C, 약 2°C 내지 약 4°C, 약 20°C 내지 약 37°C, 약 25°C 내지 약 37°C, 약 20°C 내지 약 30°C, 약 25°C 내지 약 30°C, 또는 약 30°C 내지 약 37°C의 온도에서 안정하다. 일부 경우에서, Arc 및/또는 endo-Gag 기반 캡시드는 비어 있다. 다른 경우에서, Arc 및/또는 endo-Gag 기반 캡시드는 로딩된다(예컨대, 카고 및/또는 치료제, 예컨대 DNA 또는 RNA가 로딩됨).
- [0114] 일부 예에서, Arc 및/또는 endo-Gag 기반 캡시드는 적어도 약 1일, 2일, 4일, 5일, 7일, 14일, 28일, 30일, 60일, 2개월, 3개월, 4개월, 5개월, 6개월, 12개월, 18개월, 24개월, 3년, 5년, 또는 그 초과 동안 안정하다. 일부 경우에서, Arc 및/또는 endo-Gag 기반 캡시드는 최소 분해, 예컨대 분해되는 Arc 및/또는 endo-Gag 기반 캡시드의 총 집단을 기준으로 하여 10%, 5%, 4%, 3%, 2%, 1%, 0.5% 미만을 갖는다. 일부 경우에서, Arc 및/또는 endo-Gag 기반 캡시드는 비어 있다. 다른 경우에서, Arc 및/또는 endo-Gag 기반 캡시드는 로딩된다(예컨대, 치료제, 예컨대 DNA 또는 RNA가 로딩됨).
- [0115] **추가적 캡시드**
- [0116] 일부 실시양태에서, 캡시드는 코피아 단백질을 포함한다. 일부 예에서, 코피아 단백질은 드로소필라 멜라노가스테르(*Drosophila melanogaster*)(UniProtKB-P04146), 세라티티스 카피타테(*Ceratitis capitata*)(UniProtKB-W8BHY5) 또는 드로소필라 시물란스(*Drosophila simulans*)(UniProtKB-Q08461)로부터의 것이다.
- [0117] 일부 실시양태에서, 캡시드는 단백질 ASPRV1을 포함한다. ASPRV1 단백질은 피부 장벽의 발달 및 유지에 참여하는 구조 단백질이다. 일부 예에서, 단백질 ASPRV1은 호모 사피엔스(UniProtKB-Q53RT3)로부터의 것이다.
- [0118] 일부 실시양태에서, 캡시드는 SCAN 도메인 패밀리로부터의 단백질을 포함한다. SCAN 도메인은 징크 핑거 전사 인자의 슈퍼패밀리이다. SCAN 도메인은 루이신 풍부 영역(LeR)으로도 알려져 있으며 자가-회합 또는 다른 단백질과의 선택적 회합을 매개하는 단백질 상호 작용 도메인으로 기능한다.
- [0119] 일부 실시양태에서, 캡시드는 부신생물성 Ma 항원 패밀리로부터의 단백질을 포함한다. 부신생물성 Ma 항원 패밀리는 neurO- 및 교환-특이적 단백질의 약 14개의 구성원을 포함한다.
- [0120] 일부 실시양태에서, 캡시드는 레트로트랜스포존 Gag-유사 유전자에 의해 코딩되는 단백질을 포함한다.
- [0121] 일부 실시양태에서, 캡시드는 BOP, LDOC1, MOAP1, PEG10, PNMA3, PNMA5, PNMA6A, PNMA6B, RTL3, RTL6, RTL8A, RTL8B, 및/또는 ZNF18을 포함한다.
- [0122] **카고**
- [0123] 일부 실시양태에서, 본 개시 내용의 조성물(예컨대, 캡시드)은 카고를 포함한다. 일부 실시양태에서, 카고는 치료제이다. 일부 실시양태에서, 카고는 핵산 분자, 소분자, 단백질, 펩티드, 항체 또는 이의 결합 단편, 펩티드 모방체, 또는 뉴클레오티드모방체이다. 일부 예에서, 카고는, 예컨대 하나 이상의 약물을 포함하는 치료 카고이다. 일부 예에서, 카고는 프로파일링을 위한 진단 도구, 예컨대 하나 이상의 마커(예컨대, 하나 이상의 질환 표현형과 관련된 마커)를 포함한다. 추가의 예에서, 카고는 이미징 도구를 포함한다.
- [0124] 일부 예에서, 카고는 핵산 분자이다. 예시적인 핵산 분자는 DNA, RNA, 또는 DNA와 RNA의 혼합물을 포함한다. 일부 예에서, 핵산 분자는 DNA 중합체이다. 일부 경우에서, DNA는 단일 가닥 DNA 중합체이다. 다른 경우에서, DNA는 이중 가닥 DNA 중합체이다. 추가의 경우에서, DNA는 단일 및 이중 가닥 DNA 중합체의 하이브리드이다.
- [0125] 일부 실시양태에서, 핵산 분자는 RNA 중합체, 예컨대 단일 가닥 RNA 중합체, 이중 가닥 RNA 중합체, 또는 단일 및 이중 가닥 RNA 중합체의 하이브리드이다. 일부 예에서, RNA는 안티센스 올리고리보뉴클레오티드, siRNA,

mRNA, tRNA, rRNA, snRNA, shRNA, 마이크로RNA 또는 비-코딩 RNA를 포함하고/하거나 코딩한다.

- [0126] 일부 실시양태에서, 핵산 분자는 DNA와 RNA의 하이브리드를 포함한다.
- [0127] 일부 실시양태에서, 핵산 분자는 임의적으로 DNA, RNA, 또는 DNA와 RNA의 하이브리드를 포함하는 안티센스 올리고뉴클레오티드이다.
- [0128] 일부 예에서, 핵산 분자는 mRNA 분자를 포함하고/하거나 코딩한다.
- [0129] 일부 실시양태에서, 핵산 분자는 RNAi 분자를 포함하고/하거나 코딩한다. 일부 경우에서, RNAi 분자는 마이크로RNA(miRNA) 분자이다. 다른 경우에서, RNAi 분자가 siRNA 분자이다. miRNA 및/또는 siRNA는 임의적으로 이중가닥 또는 헤어핀으로서 또한 임의적으로 전구체 분자로 캡슐화된다.
- [0130] 일부 실시양태에서, 핵산 분자는 핵산 기반 요법에 사용하기 위한 것이다. 일부 예에서, 핵산 분자는 유전자 발현 조절(예컨대, mRNA 번역 또는 분해 조절), RNA 스플라이싱 조절 또는 RNA 간섭을 위한 것이다. 일부 경우에서, 핵산 분자는 유전자 발현 조절, RNA 스플라이싱 조절, 또는 RNA 간섭에 사용하기 위한 안티센스 올리고뉴클레오티드, 마이크로RNA 분자, siRNA 분자, mRNA 분자를 포함하고/하거나 코딩한다.
- [0131] 일부 예에서, 핵산 분자는 유전자 편집에 사용하기 위한 것이다. 예시적인 유전자 편집 시스템은 CRISPR-Cas 시스템, 징크 핑거 뉴클레아제(ZFN) 시스템, 및 전사 활성화제-유사 이펙터 뉴클레아제(TALEN) 시스템을 포함하지만 이에 제한되지 않는다. 일부 경우에서, 핵산 분자는 CRISPR-Cas 시스템, ZFN 시스템, 또는 TALEN 시스템에 관여하는 성분을 포함하고/하거나 코딩한다.
- [0132] 일부 경우에서, 핵산 분자는 치료 및/또는 예방 백신 생성을 위한 항원 생성에 사용하기 위한 것이다. 예컨대, 핵산 분자는 발현되고 바람직한 면역 반응(예컨대, 전염증 면역 반응, 항염증 면역 반응, B 세포 반응, 항체 반응, T 세포 반응, CD4+ T 세포 반응, CD8+ T 세포 반응, Th1 면역 반응, Th2 면역 반응, Th17 면역 반응, Treg 면역 반응, 또는 이들의 조합)을 유도하는 항원을 코딩한다.
- [0133] 일부 경우에서, 핵산 분자는 핵산 효소를 포함한다. 핵산 효소는 촉매 활성을 갖는 RNA 분자(예컨대, 리보자임) 또는 DNA 분자(예컨대, 데옥시리보자임)이다. 일부 예에서, 핵산 분자는 리보자임이다. 다른 예에서, 핵산 분자는 데옥시리보자임이다. 일부 경우에서, 핵산 분자는 바이오센서 및/또는 분자 스위치로 기능하는 MNAzyme이다 (예컨대, 문헌(Mokany, et al., "MNAzymes, a versatile new class of nucleic acid enzymes that can function as biosensors and molecular switches," JACS 132(2): 1051-1059 (2010)) 참조).
- [0134] 일부 예에서, 핵산 분자의 예시적인 표적은 *UL123*(인간 시토크갈로바이러스), *APOB*, *AR*(안드로젠 수용체) 유전자, *KRAS*, *PCSK9*, *CFTR*, 및 *SMN*(예컨대, *SMN2*)을 포함하지만 이에 제한되지 않는다.
- [0135] 일부 실시양태에서, 핵산 분자는 길이가 적어도 5개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, 1000, 1500, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 10개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 12개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 15개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 18개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 19개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 20개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 21개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 22개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 23개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 24개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 25개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 26개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 27개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 28개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 29개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 30개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 40개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 50개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 100개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 200개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 300개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어도 500개 뉴클레오티드 또는 그 초과이다. 일부 예에서, 핵산 분자는 길이가 적어



약 5 내지 약 1000개 뉴클레오티드 길이, 약 5 내지 약 500개 뉴클레오티드 길이, 약 5 내지 약 100개 뉴클레오티드 길이, 약 5 내지 약 50개 뉴클레오티드 길이, 약 5 내지 약 40개 뉴클레오티드 길이, 약 5 내지 약 30개 뉴클레오티드 길이, 약 5 내지 약 25개 뉴클레오티드 길이, 약 5 내지 약 20개 뉴클레오티드 길이, 약 10 내지 약 8000개 뉴클레오티드 길이, 약 10 내지 약 7000개 뉴클레오티드 길이, 약 10 내지 약 6000개 뉴클레오티드 길이, 약 10 내지 약 5000개 뉴클레오티드 길이, 약 10 내지 약 4000개 뉴클레오티드 길이, 약 10 내지 약 3000개 뉴클레오티드 길이, 약 10 내지 약 2000개 뉴클레오티드 길이, 약 10 내지 약 1000개 뉴클레오티드 길이, 약 10 내지 약 500개 뉴클레오티드 길이, 약 10 내지 약 100개 뉴클레오티드 길이, 약 10 내지 약 50개 뉴클레오티드 길이, 약 10 내지 약 40개 뉴클레오티드 길이, 약 10 내지 약 30개 뉴클레오티드 길이, 약 10 내지 약 25개 뉴클레오티드 길이, 약 10 내지 약 20개 뉴클레오티드 길이, 약 18 내지 약 8000개 뉴클레오티드 길이, 약 18 내지 약 7000개 뉴클레오티드 길이, 약 18 내지 약 6000개 뉴클레오티드 길이, 약 18 내지 약 5000개 뉴클레오티드 길이, 약 18 내지 약 4000개 뉴클레오티드 길이, 약 18 내지 약 3000개 뉴클레오티드 길이, 약 18 내지 약 2000개 뉴클레오티드 길이, 약 18 내지 약 1000개 뉴클레오티드 길이, 약 18 내지 약 500개 뉴클레오티드 길이, 약 18 내지 약 100개 뉴클레오티드 길이, 약 18 내지 약 50개 뉴클레오티드 길이, 약 18 내지 약 40개 뉴클레오티드 길이, 약 18 내지 약 30개 뉴클레오티드 길이, 약 12 내지 약 50개 뉴클레오티드 길이, 약 20 내지 약 40개 뉴클레오티드 길이, 약 20 내지 약 30개 뉴클레오티드 길이, 또는 약 25 내지 약 30개 뉴클레오티드 길이이다.

- [0139] 일부 실시양태에서, 핵산 분자는 천연, 합성, 또는 인공 뉴클레오티드 유사체 또는 염기를 포함한다. 일부 경우에서, 핵산 분자는 DNA, RNA 및/또는 뉴클레오티드 유사체의 조합을 포함한다. 일부 경우에서, 합성 또는 인공 뉴클레오티드 유사체 또는 염기는 리보스 모이어티, 포스페이트 모이어티, 뉴클레오시드 모이어티, 또는 이들의 조합 중 하나 이상에서 변형을 포함한다.
- [0140] 일부 실시양태에서, 상술된 뉴클레오티드 유사체 또는 인공 뉴클레오티드 염기는 리보스 모이어티의 2' 히드록실기에서 변형을 갖는 핵산을 포함한다. 일부 예에서, 변형은 H, OR, R, 할로, SH, SR, NH<sub>2</sub>, NHR, NR<sub>2</sub> 또는 CN을 포함하며, 여기서 R은 알킬 모이어티이다. 예시적인 알킬 모이어티는 할로젠, 황, 티올, 티오에테르, 티오에스테르, 아민(1차, 2차 또는 3차), 아마이드, 에테르, 에스테르, 알콜 및 산소를 포함하지만 이에 제한되지 않는다. 일부 예에서, 알킬 모이어티는 변형을 추가로 포함한다. 일부 예에서, 변형은 아조기, 케토기, 알데히드기, 카르복실기, 니트로기, 니트로소기, 니트릴기, 헤테로사이클(예컨대, 이미다졸, 히드라지노 또는 히드록실아미노)기, 및 이소시아네이트 또는 시아네이트기, 또는 황 함유 기(예컨대, 술폭사이드, 술폰, 술파이드 또는 디술파이드)를 포함하지만 이에 제한되지 않는다. 일부 경우에서, 알킬 모이어티는 헤테로 치환을 추가로 포함한다. 일부 예에서, 헤테로사이클릭기의 탄소는 질소, 산소 또는 황으로 치환된다. 일부 예에서, 헤테로사이클릭 치환은 모르폴리노, 이미다졸, 및 피롤리디노를 포함하지만 이에 제한되지 않는다.
- [0141] 일부 예에서, 2' 히드록실기에서의 변형은 2'-O-메틸 변형 또는 2'-O-메톡시 에틸(2'-O-MOE) 변형이다. 일부 경우에서, 2'-O-메틸 변형은 리보스 모이어티의 2' 히드록실기에 메틸기를 추가하는 반면, 2'-O-메톡시에틸 변형은 리보스 모이어티의 2' 히드록실기에 메톡시에틸기를 추가한다.
- [0142] 일부 예에서, 2' 히드록실기에서의 변형은 프로필 링커를 포함하는 연장된 아민기가 아민기를 2' 산소에 결합하는 2'-O-아미노프로필 변형이다. 일부 예에서, 이 변형은 슈가당 아민기로부터 하나의 양전하를 도입함으로써 올리고뉴클레오티드 분자의 포스페이트 유래된 전체 음전하를 중화시키고 그에 따라 양쪽성 이온 특성으로 인해 세포 흡수 특성을 개선한다.
- [0143] 일부 예에서, 2' 히드록실기에서의 변형은 2' 탄소에 결합된 산소 분자가 메틸렌기에 의해 4' 탄소에 연결됨으로써 2'-C,4'-C-옥시-메틸렌 연결된 비사이클릭 리보뉴클레오티드 단량체를 형성하는 잠금 또는 브리징된 리보스 변형(예컨대, 잠금 핵산 또는 LNA)이다.
- [0144] 일부 실시양태에서, 2' 히드록실기에서의 추가의 변형은 2'-데옥시, T-데옥시-2'-플루오로, 2'-O-아미노프로필(2'-O-AP), 2'-O-디메틸아미노에틸(2'-O-DMAOE), 2'-O-디메틸아미노프로필(2'-O-DMAP), T-O-디메틸아미노에틸 옥시에틸(2'-O-DMAEOE), 또는 2'-ON-메틸아세트아미도(2'-O-NMA)를 포함한다.
- [0145] 일부 실시양태에서, 뉴클레오티드 유사체는 변형된 염기, 예컨대 5-프로피닐우리딘, 5-프로피닐시티딘, 6-메틸아데닌, 6-메틸구아닌, N,N,-디메틸아데닌, 2-프로필아데닌, 2 프로필구아닌, 2-아미노아데닌, 1-메틸이노신, 3-메틸우리딘, 5-메틸시티딘, 5-메틸우리딘 및 5 위치에 변형을 갖는 다른 뉴클레오티드, 5-(2-아미노) 프로필우리딘, 5-할로시티딘, 5-할로우리딘, 4-아세틸시티딘, 1-메틸아데노신, 2-메틸아데노신, 3-메틸시티딘, 6-메틸우리딘, 2-메틸구아노신, 7-메틸구아노신, 2,2-디메틸구아노신, 5-메틸아미노에틸우리딘, 5-메틸옥시우리딘, 데

아자뉴클레오티드(예컨대, 7-테아자-아데노신, 6-아조우리딘, 6-아조시티딘, 또는 6-아조티미딘), 5-메틸-2-티오우리딘, 다른 티오 염기(예컨대, 2-티오우리딘, 4-티오우리딘, 및 2-티오시티딘), 디히드로우리딘, 슈도우리딘, 큐오신, 아르케오신, 나프틸 및 치환된 나프틸기, 임의의 O- 및 N-알킬화 퓨린 및 피리미딘(예컨대, N6-메틸아데노신, 5-메틸카르보닐메틸우리딘, 우리딘 5-옥시아세트산, 피리딘-4-온, 또는 피리딘-2-온), 페닐 및 변형된 페닐기, 예컨대 아미노페놀 또는 2,4,6-트리메톡시 벤젠, G-클램프 뉴클레오티드로 작용하는 변형된 시토신, 8-치환된 아데닌 및 구아닌, 5-치환된 우라실 및 티민, 아자피리미딘, 카르복시히드록시알킬 뉴클레오티드, 카르복시알킬아미노알킬 뉴클레오티드, 및 알킬카르보닐알킬화 뉴클레오티드를 포함하지만 이에 제한되지 않는다. 변형된 뉴클레오티드는 또한 슈가 모이어티와 관련하여 변형된 뉴클레오티드 뿐만 아니라 리보실이 아닌 슈가 또는 이의 유사체를 갖는 뉴클레오티드를 포함한다. 예컨대, 슈가 모이어티는 일부 경우에서 만노스, 아라비노스, 글루코피라노스, 갈락토피라노스, 4'-티오리보스, 및 다른 슈가, 헤테로사이클, 또는 카르보사이클이거나 이에 기초한다. 용어 뉴클레오티드는 또한 범용 염기를 포함한다. 예로서, 범용 염기는 3-니트로피롤, 5-니트로인돌, 또는 네볼라린을 포함하지만 이에 제한되지 않는다.

[0146] 일부 실시양태에서, 뉴클레오티드 유사체는 모르폴리노, 펩티드 핵산(PNA), 메틸포스포네이트 뉴클레오티드, 티올포스포네이트 뉴클레오티드, 2'-플루오로 N3-P5'-포스포르아미다이트, 또는 1',5'-안히드로핵시톨 핵산(HNA)을 추가로 포함한다. 모르폴리노 또는 포스포로디아미데이트 모르폴리노 올리고(PMO)는 구조가 천연 핵산 구조를 모방하지만 정상적인 슈가 및 포스페이트 구조와는 다른 합성 분자를 포함하였다. 일부 예에서, 5원 리보스 고리는 4개의 탄소, 1개의 질소 및 1개의 산소를 함유하는 6원 모르폴리노 고리로 치환된다. 일부 경우에서, 리보스 단량체는 포스페이트기 대신 포스포로디아미데이트기에 의해 연결된다. 이러한 경우에서, 골격 변경은 모든 양전하 및 음전하를 제거하여 하전된 올리고뉴클레오티드에 사용되는 것과 같은 세포 전달체의 도움 없이 세포막을 통과할 수 있는 모르폴리노 중성 분자를 만든다.

[0147] 일부 실시양태에서, 펩티드 핵산(PNA)은 슈가 고리 또는 포스페이트 결합을 함유하지 않으며 염기는 올리고글리신-유사 분자에 의해 부착되고 적절하게 이격되어 골격 전하를 제거한다.

[0148] 일부 실시양태에서, 하나 이상의 변형은 임의적으로 뉴클레오티드간 결합에서 발생한다. 일부 예에서, 변형된 뉴클레오티드간 결합은 포스포로티오에이트; 포스포로디티오에이트; 메틸포스포네이트; 5'-알킬렌포스포네이트; 5'-메틸포스포네이트; 3'-알킬렌 포스포네이트; 보론트리플루오리데이트; 3'-5' 결합 또는 2'-5' 결합의 보라노 포스페이트 에스테르 및 셀레포스페이트; 포스포트리에스테르; 티오노알킬포스포트리에스테르; 수소 포스포네이트 결합; 알킬 포스포네이트; 알킬포 포노티오에이트; 아릴포스포노티오에이트; 포스포로셀레노에이트; 포스포로디셀레노에이트; 포스포네이트; 포스포르아미데이트; 3'-알킬포스포르아미데이트; 아미노알킬포스포르아미데이트; 티오노포스포르아미데이트; 포스포로피페라지데이트; 포스포로아닐로 티오에이트; 포스포로아닐리데이트; 케톤; 술폰; 술폰아미드; 카르보네이트; 카르바메이트; 메틸렌히드라조; 메틸렌디메틸히드라조; 포름아세탈; 티오포름아세탈; 옥심; 메틸렌이미노; 메틸렌메틸이미노; 티오아미데이트; 리보아세틸기와 결합; 아미노에틸 글리신; 실릴 또는 실록산 결합; 예컨대, 포화 또는 불포화 및/또는 치환 및/또는 헤테로 원자를 함유하는 1 내지 10개의 탄소의 헤테로 원자가 있거나 없는 알킬 또는 시클로알킬 결합; 모르폴리노 구조, 아미드 또는 폴리아미드와의 결합(여기서 염기는 골격의 아자 질소에 직접 또는 간접적으로 부착됨); 및 이들의 조합을 포함하지만 이에 제한되지 않는다.

[0149] 일부 실시양태에서, 하나 이상의 변형은 변형이 중성 또는 하전되지 않은 골격을 생성하는 변형된 포스페이트 골격을 포함한다. 일부 예에서, 포스페이트 골격은 알킬화에 의해 변형되어 하전되지 않은 또는 중성 포스페이트 골격을 생성한다. 본원에서 사용되는 바와 같이, 알킬화는 메틸화, 에틸화 및 프로필화를 포함한다. 일부 경우에서, 알킬화와 관련하여 본원에서 사용되는 알킬기는 1 내지 6개의 탄소 원자를 함유하는 선형 또는 분지형 포화 탄화수소기를 지칭한다. 일부 예에서, 예시적인 알킬기는 메틸, 에틸, N-프로필, 이소프로필, N-부틸, 이소부틸, sec-부틸, tert-부틸, N-펜틸, 이소펜틸, 네오펜틸, 헥실, 이소헥실, 1,1-디메틸부틸, 2,2-디메틸부틸, 3,3-디메틸부틸, 및 2-에틸부틸기를 포함하지만 이에 제한되지 않는다. 일부 경우에서, 변형된 포스페이트는 미국 특허 번호 9481905에 기재된 포스페이트기이다.

[0150] 일부 실시양태에서, 추가의 변형된 포스페이트 골격은 메틸 포스포네이트, 에틸 포스포네이트, 메틸티오포스포네이트, 또는 메톡시 포스포네이트를 포함한다. 일부 경우에서, 변형된 포스페이트는 메틸포스포네이트이다. 일부 경우에서, 변형된 포스페이트는 에틸포스포네이트이다. 일부 경우에서, 변형된 포스페이트는 메틸 티오포스포네이트이다. 일부 경우에서, 변형된 포스페이트는 메톡시포스포네이트이다.

[0151] 일부 실시양태에서, 하나 이상의 변형은 임의적으로 리보스 모이어티, 포스페이트 골격 및 뉴클레오시드의

변형, 또는 3' 또는 5' 말단에서 뉴클레오티드 유사체의 변형을 추가로 포함한다. 예컨대, 3' 말단은 임의적으로 3' 양이온 기를 포함하거나 3'-3' 연결로 3'-말단에서 뉴클레오시드를 반전시킨다. 또 다른 대안에서, 3'-말단은 임의적으로 아미노알킬기, 예컨대 3'C 5-아미노 알킬 dT와 접합된다. 추가의 대안에서, 3'-말단은 임의적으로 비염기 부위, 예컨대 비퓨린 또는 비피리미딘 부위와 접합된다. 일부 예에서, 5'-말단은 아미노알킬기, 예컨대 5'-0-알킬아미노 치환체와 접합된다. 일부 경우에서, 5'-말단은 비염기 부위, 예컨대 비퓨린 또는 비피리미딘 부위와 접합된다.

[0152] 일부 실시양태에서, 예시적인 핵산 카고는 포미비르센, 미포메르센, AZD5312(AstraZeneca), 누시네르센 및 SB010(Sterna Biologicals)을 포함하지만 이에 제한되지 않는다.

[0153] 소분자

[0154] 일부 실시양태에서, 카고는 소분자이다. 일부 예에서, 소분자는 억제제(예컨대, 팬(pan) 억제제 또는 선택적 억제제)이다. 다른 예에서, 소분자는 활성화제이다. 추가의 경우에서, 소분자는 효능제, 길항제, 부분 효능제, 혼합 효능제/길항제 또는 경쟁적 길항제이다.

[0155] 일부 실시양태에서, 소분자는 진통제, 항불안제, 항부정맥제, 항박테리아제, 항생제, 항응고제 및 혈전용해제, 항경련제, 항우울제, 지사제, 항구토제, 항진균제, 항히스타민제, 항고혈압제, 항염증제, 항신생물제, 항정신병제, 해열제, 항바이러스제, 바르비투레이트, 베타 차단제, 기관지 확장제, 감기 치료제, 코르티코스테로이드, 기침 억제제, 세포독성제, 충혈제거제, 이노제, 거담제, 호르몬, 혈당강하제, 면역억제제, 완하제, 근육이완제, 성 호르몬, 수면제, 또는 신경안정제의 부류에 속하는 약물이다.

[0156] 일부 실시양태에서, 소분자는 억제제, 예컨대 티로신 키나제 경로 또는 세린/트레오닌 키나제 경로와 같은 키나제 경로의 억제제이다. 일부 경우에서, 소분자는 이중 단백질 키나제 억제제이다. 일부 경우에서, 소분자는 지질 키나제 억제제이다.

[0157] 일부 경우에서, 소분자는 뉴라미니다제 억제제이다.

[0158] 일부 경우에서, 소분자는 탄산 안히드라제 억제제이다.

[0159] 일부 실시양태에서, 소분자의 예시적인 표적은 혈관 내피 성장 인자 수용체 1(VEGFR1), 혈관 내피 성장 인자 수용체 2(VEGFR2), 혈관 내피 성장 인자 수용체 3(VEGFR3), 섬유모세포 성장 인자 수용체 1(FGFR1), 섬유모세포 성장 인자 수용체 2(FGFR2), 섬유모세포 성장 인자 수용체 3(FGFR3), 섬유모세포 성장 인자 수용체 4(FGFR4), 사이클린 의존적 키나제 4(CDK4), 사이클린 의존적 키나제 6(CDK6), 수용체 티로신 키나제, 포스포이노시티드 3-키나제(PI3K) 이소형(예컨대, PI3K  $\delta$ , p110  $\delta$  로도 알려짐), 야누스 키나제 1(JAK1), 야누스 키나제 3(JAK3), 혈소판 유래된 성장 인자 수용체(PDFG-R) 패밀리로부터의 수용체, 및 탄산 안히드라제(예컨대, 탄산 안히드라제 I)를 포함하지만 이에 제한되지 않는다.

[0160] 일부 실시양태에서, 소분자는 바이러스 단백질, 예컨대 바이러스 외피 단백질을 표적으로 한다. 일부 실시양태에서, 소분자는 숙주 세포에 대한 바이러스 흡착을 감소시킨다. 일부 실시양태에서, 소분자는 숙주 세포로의 바이러스 진입을 감소시킨다. 일부 실시양태에서, 소분자는 숙주 또는 숙주 세포에서 바이러스 복제를 감소시킨다. 일부 실시양태에서, 소분자는 바이러스 조립을 감소시킨다.

[0161] 일부 실시양태에서, 예시적인 소분자 카고는 렌바티닙, 팔보시클립, 레고라페닙, 이텔랄리십, 토파시티닙, 닌테다닙, 자나미비르, 에톡졸아미드, 및 아르테미시닌을 포함하지만 이에 제한되지 않는다.

[0162] 단백질

[0163] 일부 실시양태에서, 카고는 단백질이다. 일부 예에서, 단백질이 전장 단백질이다. 다른 예에서, 단백질은 단편, 예컨대 기능적 단편이다. 일부 경우에서, 단백질이 자연 발생 단백질이다. 추가의 경우에서, 단백질은 드 노보 조작된 단백질이다. 추가의 경우에서, 단백질은 융합 단백질이다. 추가의 경우에서, 단백질은 재조합 단백질이다. 예시적인 단백질은 Fc 융합 단백질, 항응고제, 혈액 인자, 골 형태형성 단백질, 효소, 성장 인자, 호르몬, 인터페론, 인터류킨, 및 혈전용해제를 포함하지만 이에 제한되지 않는다.

[0164] 일부 예에서, 단백질은 효소 대체 요법에 사용하기 위한 것이다.

[0165] 일부 경우에서, 단백질은 치료 및/또는 예방 백신 생성을 위한 항원 생성에 사용하기 위한 것이다. 예컨대, 단백질은 바람직한 면역 반응(예컨대, 전염증 면역 반응, 항염증 면역 반응, B 세포 반응, 항체 반응, T 세포 반응, CD4+ T 세포 반응, CD8+ T 세포 반응, Th1 면역 반응, Th2 면역 반응, Th17 면역 반응, Treg 면역 반응,

또는 이들의 조합)을 유도하는 항원을 포함한다.

- [0166] 일부 예에서, 예시적인 단백질 카고는 로미플로스티뮴, 리라글루타이드, 인간 성장 호르몬(rHGH), 인간 인슐린(BHI), 난포 자극 호르몬(FSH), 인자 VIII, 에리트로포이에틴(EPO), 과립구 콜로니 자극 인자(G-CSF), 알파-갈락토시다제 A, 알파-L-이두로니다제, N-아세틸갈락토스아민-4-수퍼타제, 도르나제 알파, 조직 플라스미노겐 활성화제(TPA), 글루코세레브로시다제, 인터페론-베타-1a, 인슐린-유사 성장 인자 1(IGF-1) 또는 라스부리카제를 포함하지만 이에 제한되지 않는다.
- [0167] 펩티드
- [0168] 일부 실시양태에서, 카고는 펩티드이다. 일부 경우에서, 펩티드는 자연 발생 펩티드이다. 다른 예에서, 펩티드는 인공적으로 조작된 펩티드 또는 재조합 펩티드이다. 일부 경우에서, 펩티드는 G-단백질 커플링된 수용체, 이온 채널, 미생물, 항 미생물 표적, 촉매 또는 다른 Ig 패밀리 수용체, 세포내 표적, 막 고정된 표적 또는 세포외 표적을 표적으로 한다.
- [0169] 일부 경우에서, 펩티드는 적어도 2개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 10개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 15개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 20개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 30개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 40개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 50개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 60개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 70개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 80개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 90개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 적어도 100개의 아미노산을 포함한다.
- [0170] 일부 경우에서, 펩티드는 최대 3개의 아미노산을 포함한다. 일부 경우에서, 펩티드는 최대 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 10개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 15개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 20개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 30개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 40개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 50개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 60개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 70개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 80개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 90개의 아미노산을 포함한다. 일부 경우에서, 펩티드가 최대 100개의 아미노산을 포함한다.
- [0171] 일부 경우에서, 펩티드는 약 1 내지 약 10 kDa를 포함한다. 일부 경우에서, 펩티드는 약 1 내지 약 9 kDa, 약 1 내지 약 6 kDa, 약 1 내지 약 5 kDa, 약 1 내지 약 4 kDa, 약 1 내지 약 3 kDa, 약 2 내지 약 8 kDa, 약 2 내지 약 6 kDa, 약 2 내지 약 4 kDa, 약 1.2 내지 약 2.8 kDa, 약 1.5 내지 약 2.5 kDa, 또는 약 1.5 내지 약 2 kDa를 포함한다.
- [0172] 일부 실시양태에서, 펩티드는 사이클릭 펩티드이다. 일부 예에서, 사이클릭 펩티드는 매크로사이클릭 펩티드이다. 다른 예에서, 사이클릭 펩티드는 제한된 펩티드이다. 사이클릭 펩티드는, 예컨대 헤드-대-테일, 헤드-대-측쇄, 측쇄-대-테일 및 측쇄-대-측쇄 결합과 같은 다양한 결합으로 조립된다. 일부 예에서, 사이클릭 펩티드(예컨대, 매크로사이클릭 또는 제한된 펩티드)는 분자량이 약 500 달톤 내지 약 2000 달톤이다. 다른 예에서, 사이클릭 펩티드(예컨대, 매크로사이클릭 또는 제한된 펩티드)는 약 10개 아미노산 내지 약 100개 아미노산, 약 10개 아미노산 내지 약 70개 아미노산, 또는 약 10개 아미노산 내지 약 50개 아미노산 범위이다.
- [0173] 일부 경우에서, 펩티드는 치료 및/또는 예방 백신 생성을 위한 항원 생성에 사용하기 위한 것이다. 예컨대, 펩티드는 바람직한 면역 반응(예컨대, 전염증 면역 반응, 항염증 면역 반응, B 세포 반응, 항체 반응, T 세포 반응, CD4+ T 세포 반응, CD8+ T 세포 반응, Th1 면역 반응, Th2 면역 반응, Th17 면역 반응, Treg 면역 반응, 또는 이들의 조합)을 유도하는 항원을 포함한다.
- [0174] 일부 실시양태에서, 펩티드는 천연 아미노산, 비천연 아미노산, 또는 이들의 조합을 포함한다. 일부 예에서, 아미노산 잔기는 아미노기 및 카르복실기를 모두 함유하는 분자를 지칭한다. 적합한 아미노산은 제한 없이 자연 발생 아미노산의 D- 및 L-이성체 뿐만 아니라 유기 합성 또는 다른 대사 경로에 의해 제조되는 비-자연 발생 아미노산을 포함한다. 본원에서 사용되는 바와 같이 용어 아미노산은 제한 없이 α-아미노산, 천연 아미노산, 비-천연 아미노산, 및 아미노산 유사체를 포함한다.

- [0175] 일부 예에서,  $\alpha$ -아미노산은  $\alpha$ -탄소로 지정된 탄소에 결합된 아미노기 및 카르복실기를 모두 함유하는 분자를 지칭한다.
- [0176] 일부 예에서,  $\beta$ -아미노산은  $\beta$  배치의 아미노기 및 카르복실기를 모두 함유하는 분자를 지칭한다.
- [0177] 일부 실시양태에서, 아미노산 유사체는 라세미 혼합물이다. 일부 예에서, 아미노산 유사체의 D 이성체가 사용된다. 일부 경우에서, 아미노산 유사체의 L 이성체가 사용된다. 일부 예에서, 아미노산 유사체는 R 또는 S 배치의 키랄 중심을 포함한다.
- [0178] 일부 실시양태에서, 예시적인 펩티드 카고는 페기네사티드, 인슐린, 부신피질자극 호르몬(ACTH), 칼시토닌, 옥시토신, 바소프레신, 옥트레올리드 및 류프로렐린을 포함하지만 이에 제한되지 않는다.
- [0179] 일부 실시양태에서, 예시적인 펩티드 카고는 텔라반신, 달바반신, 오리타반신, 아니둘라폰진, 란레오티드, 파시레오티드, 로미넵신, 리나클로티드 및 페기네사티드를 포함하지만 이에 제한되지 않는다.
- [0180] 항체
- [0181] 일부 실시양태에서, 카고는 항체 또는 이의 결합 단편이다. 일부 경우에서, 항체 또는 이의 결합 단편은 인간화 항체 또는 이의 결합 단편, 무린 항체 또는 이의 결합 단편, 키메라 항체 또는 이의 결합 단편, 모노클로날 항체 또는 이의 결합 단편, 이중특이적 항체 또는 이의 결합 단편, 1가 Fab', 2가 Fab<sub>2</sub>, F(ab)'<sub>3</sub> 단편, 단일 쇄가변 단편(scFv), 비스-scFv, (scFv)<sub>2</sub>, 디아바디, 미니바디, 나노바디, 트리아바디, 테트라바디, 디슈파이드 안정화된 Fv 단백질(dsFv), 단일 도메인 항체(sdAb), Ig NAR, 낙타류 항체 또는 이의 결합 단편, 또는 이의 화학적으로 변형된 유도체를 포함한다.
- [0182] 일부 예에서, 항체 또는 결합 단편은 세포 표면 단백질을 인식한다. 일부 예에서, 세포 표면 단백질은 암 세포에 의해 발현되는 항원이다. 일부 예에서, 세포 표면 단백질은 네오에피토프이다. 일부 예에서, 세포 표면 단백질은 야생형 단백질과 비교하여 하나 이상의 돌연변이를 포함한다. 예시적인 암 항원은 알파 태아단백질, ASLG659, B7-H3, BAFF-R, 브레비칸, CA125(MUC16), CA15-3, CA19-9, 암배아 항원(CEA), CA242, CRIPTO(CR, CR1, CRGF, CRIPTO, TDGF1, 기형암종 유래된 성장 인자), CTLA-4, CXCR5, E16(LAT1, SLC7A5), FcRH2(IFGP4, IRTA4, SPAP1A(SH2 도메인 함유 포스파타제 앵커 단백질 1a), SPAP1B, SPAP1C)), 표피 성장 인자, ETBR, Fc 수용체-유사 단백질 1(FCRH1), GEDA, HLA-DOB(MHC 부류 II 분자의 베타 서브유닛(Ia 항원), 인간 옴모 형성자극 호르몬, ICOS, IL-2 수용체, IL20R  $\alpha$ , 면역 글로블린 슈퍼패밀리 수용체 전위 관련된 2(IRTA2), L6, 루이스 Y, 루이스 X, MAGE-1, MAGE-2, MAGE-3, MAGE 4, MART1, 메소텔린, MDP, MPF(SMR, MSLN), MCP1(CCL2), 대식구 억제 인자(MIF), MPG, MSG783, 뮤신, MUC1-KLH, Napi3b(SLC34A2), 넥틴-4, Neu 온코진 산물, NCA, 태반 알칼리성 포스파타제, 전립선 특이적 막 항원(PMSA), 전립선 산 포스파타제, PSCA h1g, 항-트랜스페린 수용체, p97, 퓨린성 수용체 P2X 리간드 게이팅된 이온 채널 5(P2X5), LY64(림프구 항원 64(RP105), gp100, P21, 전립선의 6개의 막형단 상피 항원(STEAP1), STEAP2, Sema 5b, 종양 관련된 당단백질 72(TAG-72), Trpm4(BR22450, FLJ20041, TRPM4, TRPM4B, 일시적 수용체 전위 양이온 채널, 서브패밀리 M, 멤버 4) 등을 포함하지만 이에 제한되지 않는다.
- [0183] 일부 예에서, 세포 표면 단백질은 분화 클러스터(CD) 세포 표면 마커를 포함한다. 예시적인 CD 세포 표면 마커는 CD1, CD2, CD3, CD4, CD5, CD6, CD7, CD8, CD9, CD10, CD11a, CD11b, CD11c, CD11d, CDw12, CD13, CD14, CD15, CD15s, CD16, CDw17, CD18, CD19, CD20, CD21, CD22, CD23, CD24, CD25, CD26, CD27, CD28, CD29, CD30, CD31, CD32, CD33, CD34, CD35, CD36, CD37, CD38, CD39, CD40, CD41, CD42, CD43, CD44, CD45, CD45RO, CD45RA, CD45RB, CD46, CD47, CD48, CD49a, CD49b, CD49c, CD49d, CD49e, CD49f, CD50, CD51, CD52, CD53, CD54, CD55, CD56, CD57, CD58, CD59, CDw60, CD61, CD62E, CD62L(L-셀렉틴), CD62P, CD63, CD64, CD65, CD66a, CD66b, CD66c, CD66d, CD66e, CD71, CD79(예컨대, CD79a, CD79b), CD90, CD95(Fas), CD103, CD104, CD125(IL5RA), CD134(OX40), CD137(4-1BB), CD152(CTLA-4), CD221, CD274, CD279(PD-1), CD319(SLAMF7), CD326(EpCAM) 등을 포함하지만 이에 제한되지 않는다.
- [0184] 일부 실시양태에서, 예시적인 항체 또는 이의 결합 단편은 잘루투무맵(HuMax-EFGr, Genmab), 아바고보맵(Menarini), 아비투주맵(Merck), 아테카투무맵(MT201), 알라시주맵 페골, 알레투주맵(Campath®), 맵캄파쓰, 또는 캄파쓰-1H; 류코시트), 알로무네(BioTransplant), 아마투스맵(Morphotek, Inc.), 항-VEGF(Genentech), 아나투모맵 마페나투스, 아폴리주맵(hu1D10), 아스크린바쿠맵(Pfizer Inc.), 아테졸리주맵(MPDL3280A; Genentech/Roche), B43.13(오바렉스, AltaRex Corporation), 바실릭시맵(Simulect®, Novartis), 벨리무맵

(Benlysta®, GlaxoSmithKline), 베바시주맵(Avastin®, Genentech), 블리나투모맵(블린시토, AMG103; Amgen), BEC2(ImClone Systems Inc.), 카를루맵(Janssen Biotech), 카투막소맵(Removab, Trion Pharma), CEAcide(Immunomedics), 세톡시맵(Erbix®), ImClone), 시타투주맵 보가톡스(VB6-845), 시톡투무맵(IMC-A12, ImClone Systems Inc.), 코나투무맵(AMG 655, Amgen), 다세투주맵(SGN-40, huS2C6; Seattle Genetics, Inc.), 다라투무맵(Darzalex®, Janssen Biotech), 데투모맵, 드로지투맵(Genentech), 두르발루맵(MedImmune), 두시기투맵(MedImmune), 에드레콜로맵(MAb17-1A, 과노렉스, Glaxo Wellcome), 엘로투주맵(Empliciti™, Bristol-Myers Squibb), 에미베투주맵(Eli Lilly), 예나바투주맵(Facet Biotech Corp.), 엔포르투맵 보도틴(Seattle Genetics, Inc.), 예노블리투주맵(MGA271, MacroGenics, Inc.), 엔시톡수맵(Neogenix Oncology, Inc.), 예프라투주맵(림포시드, Immunomedics, Inc.), 예르투막소맵(Rexmun®, Trion Pharma), 예타라시주맵(아베그린, MedImmune), 파르레투주맵(MORAb-003, Morphotek, Inc), FBTA05(림포문, Trion Pharma), 피클라투주맵(AVEO Pharmaceuticals), 피기투무맵(CP-751871, Pfizer), 플란보투맵(ImClone Systems), 프레솔리무맵(GC1008, Aanofi-Aventis), 푸톡시맵, 글락시맵, 가니투맵(Amgen), 기렌톡시맵(Rencarex®, Willex AG), IMAB362(클라우딕시맵, Ganymed Pharmaceuticals AG), 이말루맵(Baxalta), IMC-1C11(ImClone Systems), IMC-C225(Imclone Systems Inc.), 임가투주맵(Genentech/Roche), 인테투무맵(Centocor, Inc.), 이펠리무맵(Yervoy®, Bristol-Myers Squibb), 이라투무맵(Medarex, Inc.), 이사톡시맵(SAR650984, Sanofi-Aventis), 라베투주맵(CEA-CIDE, Immunomedics), 렉사투무맵(ETR2-ST01, Cambridge Antibody Technology), 린투주맵(SGN-33, Seattle Genetics), 루카투무맵(Novartis), 루밀릭시맵, 마파투무맵(HGS-ETR1, Human Genome Sciences), 마투주맵(EMD 72000, Merck), 밀라투주맵(hLL1, Immunomedics, Inc.), 미투모맵(BEC-2, ImClone Systems), 나르나투맵(ImClone Systems), 네시투무맵(Portrazza™, Eli Lilly), 네스바쿠맵(Regeneron Pharmaceuticals), 니모투주맵(h-R3, BIOMAb EGFR, TheraCIM, 테랄록, 또는 CIMAher; Biotech Pharmaceutical Co.), 니볼루맵(Opdivo®, Bristol-Myers Squibb), 오비누투주맵(가지바 또는 가지바로; Hoffmann-La Roche), 오카라투주맵(AME-133v, LY2469298; Mentrik Biotech, LLC), 오파투무맵(Arzerra®, Genmab), 오나르투주맵(Genentech), 온톡시주맵(Morphotek, Inc.), 오레고보맵(오바렉스®, AltaRex Corp.), 오틀레르투주맵(Emergent BioSolutions), 과니투무맵(ABX-EGF, Amgen), 판코맵(Glycotope GmbH), 파르사투주맵(Genentech), 파트리투맵, 펌브롤리주맵(Keytruda®, Merck), 펌투모맵(테라긴, Antisoma), 페르투주맵(페르제타, Genentech), 피딜리주맵(CT-011, Medivation), 폴라투주맵 보도틴(Genentech/Roche), 프리투무맵, 라코투모맵(Vaxira®, Recombio), 라무시루맵(Cyramza®, ImClone Systems Inc.), 리톡시맵(Rituxan®, Genentech), 로바투무맵(Schering-Plough), 세리반투맵(Sanofi/Merrimack Pharmaceuticals, Inc.), 시브로투주맵, 실톡시맵(Sylvant™, Janssen Biotech), 스마트 MI95(Protein Design Labs, Inc.), 스마트 ID10(Protein Design Labs, Inc.), 탈발루맵(LY2127399, Eli Lilly), 타플리투모맵 팜톡스, 테나투모맵, 테프로투무맵(Roche), 테톨로맵, TGN1412(CD28-SuperMAB 또는 TAB08), 티가투주맵(CD-1008, Daiichi Sankyo), 토시투모맵, 트라스투주맵(Herceptin®), 트레멜리무맵(CP-672,206; Pfizer), 투코투주맵 셀몰류킨(EMD Pharmaceuticals), 우블리톡시맵, 우렐루맵(BMS-663513, Bristol-Myers Squibb), 볼로식시맵(M200, Biogen Idec), 및 자톡시맵을 포함하지만 이에 제한되지 않는다.

[0185] 일부 예에서, 항체 또는 이의 결합 단편은 항체-약물 접합체(ADC)이다. 일부 경우에서, ADC의 페이로드는, 예컨대 아우리스타틴 유도체, 메이탄신, 메이탄시노이드, 탁산, 칼리케아미신, 세마도틴, 듀오카르마이신, 피롤로벤조디아제핀(PDB) 또는 투블리신을 포함하지만 이에 제한되지 않는다. 일부 예에서, 페이로드 는 모노 메틸 아우리스타틴 E(MMAE) 또는 모노메틸 아우리스타틴 F(MMAF)를 포함한다. 일부 예에서, 페이로드는 DM2(메르탄신) 또는 DM4를 포함한다. 일부 예에서, 페이로드는 피롤로벤조디아제핀 이량체를 포함한다.

[0186] *추가 카고*

[0187] 일부 실시양태에서, 카고는 펩티드모방체이다. 펩티드모방체는 펩티드를 모방하도록 설계된 작은 단백질-유사 중합체이다. 일부 예에서, 펩티드모방체에는 D- 펩티드를 포함한다. 다른 예에서, 펩티드모방체는 L-펩티드를 포함한다. 예시적인 펩티드모방체는 펩티드 및 β-펩티드를 포함한다.

[0188] 일부 실시양태에서, 카고는 뉴클레오티드모방체이다.

[0189] **벡터 및 발현 시스템**

[0190] 특정 실시양태에서, 상술된 Arc 폴리펩티드, endo-Gag 폴리펩티드, 조작된 Arc 및 조작된 endo-Gag 폴리펩티드는 플라즈미드 벡터에 의해 코딩된다. 일부 실시양태에서, 벡터는 진핵 또는 원핵 공급원으로부터 유래된 임의의 적합한 벡터를 포함한다. 일부 경우에서, 벡터는 박테리아(예컨대, 이. 콜라이(*E. coli*)), 곤충, 효모(예컨대

대, 피키아 파스토리스(*Pichia pastoris*)), 조류, 또는 포유동물 공급원으로부터 취득된다.

- [0191] 예시적인 박테리아 벡터는 pACYC177, pASK75, pBAD 벡터 시리즈, pBADM 벡터 시리즈, pET 벡터 시리즈, pETM 벡터 시리즈, pGEX 벡터 시리즈, pHAT, pHAT2, pMal-c2, pMal-p2, pQE 벡터 시리즈, pRSET A, pRSET B, pRSET C, pTrcHis2 series, pZA31-Luc, pZE21-MCS-1, pFLAG ATS, pFLAG CTS, pFLAG MAC, pFLAG Shift-12c, pTAC-MAT-1, pFLAG CTC, 또는 pTAC-MAT-2를 포함한다.
- [0192] 예시적인 곤충 벡터는 pFastBac1, pFastBac DUAL, pFastBac ET, pFastBac HTa, pFastBac HTb, pFastBac HTc, pFastBac M30a, pFastBac M30b, pFastBac, M30c, pVL1392, pVL1393, pVL1393 M10, pVL1393 M11, pVL1393 M12, FLAG 벡터, 예컨대 pPolh-FLAG1 또는 pPolh-MAT 2, 또는 MAT 벡터, 예컨대 pPolh-MAT1, 또는 pPolh-MAT2를 포함한다.
- [0193] 일부 경우에서, 효모 벡터는 Gateway® pDEST™ 14 벡터, Gateway® pDEST™ 15 벡터, Gateway® pDEST™ 17 벡터, Gateway® pDEST™ 24 벡터, Gateway® pYES-DEST52 벡터, pBAD-DEST49 Gateway® 목표 벡터, pA0815 피키아 벡터, pFLD1 피키 파스토리스(*Pichi pastoris*) 벡터, pGAPZA,B, & C 피키아 파스토리스 벡터, pPIC3.5K 피키아 벡터, pPIC6 A, B, & C 피키아 벡터, pPIC9K 피키아 벡터, pTEF1/Zeo, pYES2 효모 벡터, pYES2/CT 효모 벡터, pYES2/NT A, B, & C 효모 벡터, 또는 pYES3/CT 효모 벡터를 포함한다.
- [0194] 예시적인 조류 벡터는 pChlamy-4 벡터 또는 MCS 벡터를 포함한다.
- [0195] 포유동물 벡터의 예는 일시적 발현 벡터 또는 안정적인 발현 벡터를 포함한다. 포유동물 일시적 발현 벡터는 p3xFLAG-CMV 8, pFLAG-Myc-CMV 19, pFLAG-Myc-CMV 23, pFLAG-CMV 2, pFLAG-CMV 6a,b,c, pFLAG-CMV 5.1, pFLAG-CMV 5a,b,c, p3xFLAG-CMV 7.1, pFLAG-CMV 20, p3xFLAG-Myc-CMV 24, pCMV-FLAG-MAT1, pCMV-FLAG-MAT2, pBICEP-CMV 3, 또는 pBICEP-CMV 4를 포함한다. 포유동물 안정적인 발현 벡터는 pFLAG-CMV 3, p3xFLAG-CMV 9, p3xFLAG-CMV 13, pFLAG-Myc-CMV 21, p3xFLAG-Myc-CMV 25, pFLAG-CMV 4, p3xFLAG-CMV 10, p3xFLAG-CMV 14, pFLAG-Myc-CMV 22, p3xFLAG-Myc-CMV 26, pBICEP-CMV 1, 또는 pBICEP-CMV 2를 포함한다.
- [0196] 일부 예에서, 무세포 시스템은 세포로부터의 세포질 및/또는 핵 성분의 혼합물이며 시험관내 핵산 합성에 사용된다. 일부 경우에서, 무세포 시스템은 원핵 세포 성분 또는 진핵 세포 성분을 사용한다. 때때로, 핵산 합성은, 예컨대 드로소필라 세포, 제노푸스 난자, 또는 HeLa 세포(ATCC® CCL-2™)를 기반으로 하는 무세포 시스템에서 취득된다. 예시적인 무세포 시스템은 이. 콜라이 S30 추출물 시스템, 이. 콜라이 T7 S30 시스템, 또는 PURExpress®를 포함하지만 이에 제한되지 않는다.
- [0197] 숙주 세포
- [0198] 일부 실시양태에서, 숙주 세포는 천연 유래 세포 또는 유전자조작으로 변형된 세포와 같은 임의의 적합한 세포를 포함한다. 일부 예에서, 숙주 세포는 생식 숙주 세포이다. 일부 예에서, 숙주 세포는 진핵 세포이다. 다른 예에서 숙주 세포는 원핵 세포이다. 일부 경우에서, 진핵 세포는 진균(예컨대, 효모 세포), 동물 세포, 또는 식물 세포를 포함한다. 일부 경우에서, 원핵 세포가 박테리아 세포이다. 박테리아 세포의 예는 그람 양성 박테리아 또는 그람 음성 박테리아를 포함한다. 일부 실시양태에서 그람 음성 박테리아는 혐기성, 막대형, 또는 둘 다이다.
- [0199] 일부 경우에서, 그람 양성 박테리아는 악티노박테리아(*Actinobacteria*), 피르미쿠테스(*Firmicutes*), 또는 테네리쿠테스(*Tenericutes*)를 포함한다. 일부 경우에서, 그람 음성 박테리아는 아퀴피카에(*Aquificae*), 데이노코쿠스-페르무스(*Deinococcus-Thermus*), 피브로박테레스-클로로비/박테로이테테스(*Fibrobacteres-Chlorobi/Bacteroidetes*)(FCB군), 푸소박테리아(*Fusobacteria*), 겐마티모나데테스(*Gemmatimonadetes*), 니트로스피라에(*Nitrospirae*), 플란크토미세테스-베루코미크로비아/클라미디아에(*Planctomycetes-Verrucomicrobia/Chlamydiae*)(PVC군), 프로테오박테리아(*Proteobacteria*), 스피로카에테스(*Spirochaetes*) 또는 시네르گی스테테스(*Synergistetes*). 일부 실시양태에서, 박테리아는 아시도박테리아(*Acidobacteria*), 클로로플렉시(*Chloroflexi*), 크리시오게네테스(*Chrysiogenetes*), 시아노박테리아(*Cyanobacteria*), 데페리박테레스(*Deferribacteres*), 디티오글로미(*Dictyoglomi*), 테르모데술포박테리아(*Thermodesulfobacteria*) 또는 테르모토가에(*Thermotogae*)를 포함한다. 일부 실시양태에서, 박테리아 세포는 에스케리키아 콜라이(*Escherichia coli*), 클로스트리디움 보툴리눔(*Clostridium botulinum*) 또는 콜라이 바실리(*Coli bacilli*)이다.
- [0200] 예시적인 원핵 숙주 세포는 BL21, Mach1™, DH10B™, TOP10, DH5 α, DH10Bac™, OmniMax™, MegaX™, DH12S™,

INV110, TOP10F', INV αF, TOP10/P3, ccdB 서바이벌(Survival), PIR1, PIR2, Stb12<sup>TM</sup>, Stb13<sup>TM</sup>, 또는 Stb14<sup>TM</sup>을 포함하지만 이에 제한되지 않는다.

[0201] 일부 예에서, 동물 세포는 척추동물 또는 무척추동물로부터의 세포를 포함한다. 일부 경우에서, 동물 세포는 해양 무척추동물, 어류, 곤충, 양서류, 파충류, 포유동물, 또는 인간으로부터의 세포를 포함한다. 일부 경우에서, 진균 세포는 맥주 효모, 빵 효모 또는 와인 효모와 같은 효모 세포를 포함한다.

[0202] 진균은 효모, 몰드(mold), 사상균, 담자균 또는 접합균과 같은 자낭균을 포함한다. 일부 예에서, 효모는 아스코미코타(*Ascomycota*) 또는 바시디오미코타(*Basidiomycota*)를 포함한다. 일부 경우에서, 아스코미코타는 사카로미코티나(*Saccharomycotina*)(진정 효모, 예컨대 사카로미세스 세레비시아에(빵 효모)) 또는 타프린노미코티나(*Taphrinomycotina*)(예컨대, 쉬조사카로미세테스(*Schizosaccharomycetes*)(분열 효모))를 포함한다. 일부 경우에서, 바시디오미코타(*Basidiomycota*)는 아가리코미코티나(*Agaricomycotina*)(예컨대, 트레멜로미세테스(*Tremellomycetes*)) 또는 푸시니오미코티나(*Pucciniomycotina*)(예컨대, 마이크로보트리요미세테스(*Microbotryomycetes*))를 포함한다.

[0203] 예시적 효모 또는 사상균은, 예컨대 사카로미세스, 쉬조사카로미세스, 칸디다, 피키아, 한세놀라(*Hansenula*), 클루이베로미세스(*Kluyveromyces*), 지고사카로미세스(*Zygosaccharomyces*), 야로위아(*Yarrowia*), 트리코스포론(*Trichosporon*), 로도스포리디(*Rhodospiridi*), 아스페르길루스(*Aspergillus*), 푸사리움(*Fusarium*), 또는 트리코데르마(*Trichoderma*) 속을 포함한다. 예시적 효모 또는 사상균은, 예컨대 사카로미세스 세레비시아에, 쉬조사카로미세스 폼베(*Schizosaccharomyces pombe*), 칸디다 우틸리스(*Candida utilis*), 칸디다 보이디니(*Candida boidini*), 칸디다 알비카스(*Candida albicans*), 칸디다 트로피칼리스(*Candida tropicalis*), 칸디다 스텔라토이데아(*Candida stellatoidea*), 칸디다 글라블라타(*Candida glabrata*), 칸디다 크루세이(*Candida krusei*), 칸디다 파라프실로시스(*Candida parapsilosis*), 칸디다 구일리에르몽디이(*Candida guilliermondii*), 칸디다 비스와나티이(*Candida viswanathii*), 칸디다 루시타니아에(*Candida lusitaniae*), 로도토롤라 무실라기노사(*Rhodotorula mucilaginosa*), 피키아 메타놀리카(*Pichia metanolica*), 피키아 안구스타(*Pichia angusta*), 피키아 파스토리스(*Pichia pastoris*), 피키아 아노말라(*Pichia anomala*), 한세놀라 폴리모르파(*Hansenula polymorpha*), 클루이베로미세스 락티스(*Kluyveromyces lactis*), 지고사카로미세스 로옥시이(*Zygosaccharomyces rouxii*), 야로위아 리폴리티카(*Yarrowia lipolytica*), 트리코스포론 풀루란스(*Trichosporon pullulans*), 로도스포리디움 토루-아스페르길루스 니거(*Rhodospiridium toru-Aspergillus niger*), 아스페르길루스 니둘란스(*Aspergillus nidulans*), 아스페르길루스 아와모리(*Aspergillus awamori*), 아스페르길루스 오리자에(*Aspergillus oryzae*), 트리코데르마 레세이(*Trichoderma reesei*), 야로위아 리폴리티카(*Yarrowia lipolytica*), 브레타노미세스 브룩셀렌시스(*Brettanomyces bruxellensis*), 칸디다 스텔라타(*Candida stellata*), 쉬조사카로미세스 폼베(*Schizosaccharomyces pombe*), 토룰라스포라 델브루엑키이(*Torulaspora delbrueckii*), 지고사카로미세스 바일리이(*Zygosaccharomyces bailii*), 크립토코쿠스 네오포르만스(*Cryptococcus neoformans*), 크립토코쿠스 가티이(*Cryptococcus gattii*), 또는 사카로미세스 보올라르디이(*Saccharomyces boulardii*) 종을 포함한다.

[0204] 예시적인 효모 숙주 세포는 GS115, KM71H, SMD1168, SMD1168H 및 X-33과 같은 피키아 파스토리스 효모 균주; 및 INVSc1과 같은 사카로미세스 세레비시아에 효모 균주를 포함하지만 이에 제한되지 않는다.

[0205] 일부 예에서, 추가의 동물 세포는 연체동물, 절지동물, 환형동물 또는 해면동물로부터 수득되는 세포를 포함한다. 일부 경우에서, 추가의 동물 세포는 포유동물 세포, 예컨대 인간, 영장류, 유인원, 말, 소, 돼지, 개, 고양이 또는 설치류로부터의 세포이다. 일부 경우에서, 설치류는 마우스, 래트, 햄스터, 게르빌, 햄스터, 친칠라, 팬시 래트 또는 기니피그를 포함한다.

[0206] 예시적인 포유동물 숙주 세포는 293A 세포주, 293FT 세포주, 293F 세포, 293 H 세포, CHO DG44 세포, CHO-S 세포, CHO-K1 세포, Expi293F<sup>TM</sup> 세포, Flp-In<sup>TM</sup> T-REx<sup>TM</sup> 293 세포주, Flp-In<sup>TM</sup>-293 세포주, Flp-In<sup>TM</sup>-3T3 세포주, Flp-In<sup>TM</sup>-BHK 세포주, Flp-In<sup>TM</sup>-CHO 세포주, Flp-In<sup>TM</sup>-CV-1 세포주, Flp-In<sup>TM</sup>-Jurkat 세포주, FreeStyle<sup>TM</sup> 293-F 세포, FreeStyle<sup>TM</sup> CHO-S 세포, GripTite<sup>TM</sup> 293 MSR 세포주, GS-CHO 세포주, HepaRG<sup>TM</sup> 세포, T-REx<sup>TM</sup> Jurkat 세포주, Per.C6 세포, T-REx<sup>TM</sup>-293 세포주, T-REx<sup>TM</sup>-CHO 세포주, 및 T-REx<sup>TM</sup>-HeLa 세포주를 포함하지만 이에 제한되지 않는다.

[0207] 일부 예에서, 포유동물 숙주 세포는 1차 세포이다. 일부 예에서, 포유동물 숙주 세포는 안정한 세포주, 또는 관

심 유전 물질을 자체 게놈에 통합하고 여러 세대의 세포 분열 후에 유전 물질의 산물을 발현할 수 있는 능력을 가진 세포주이다. 일부 경우에서, 포유동물 숙주 세포는 일시적 세포주, 또는 관심 유전 물질을 자체 게놈에 통합하지 않고 여러 세대의 세포 분열 후에 유전 물질의 산물을 발현할 능력이 없는 세포주이다.

- [0208] 예시적인 곤충 숙주 세포는 드로스필라 S2 세포, Sf9 세포, Sf21 세포, High Five™ 세포 및 expresSF+® 세포를 포함하지만 이에 제한되지 않는다.
- [0209] 일부 예에서, 식물 세포는 조류로부터의 세포를 포함한다. 예시적인 곤충 세포주는 클라미도모나스 레이하르트티(*Chlamydomonas reinhardtii*), 또는 시네코코쿠스 엘롱가투스 PPC 7942(*Synechococcus elongatus PPC 7942*)로부터의 스트레인을 포함한다.
- [0210] **사용 방법**
- [0211] 특정 실시양태에서, 카고를 캡슐화하는 캡시드를 제조하는 방법이 본원에 개시된다. 일부 실시양태에서, 방법은 복수의 Arc 또는 endo-Gag 폴리펩티드, 조작된 Arc 또는 endo-Gag 폴리펩티드, 및/또는 재조합 Arc 또는 endo-Gag 폴리펩티드를 로딩된 Arc 기반 캡시드 또는 endo-Gag 기반 캡시드를 생성하기에 충분한 시간 동안 용액에서 카고와 함께 인큐베이션하는 단계를 포함한다.
- [0212] 일부 예에서, 방법은 카고와 함께 인큐베이션하기 전에 복수의 조작된 및/또는 재조합 Arc 폴리펩티드를 포함하는 용액을 복수의 비-Arc 캡시드 형성 서브유닛과 혼합하는 단계를 포함한다. 일부 경우에서, 복수의 비-Arc 캡시드 형성 서브유닛은 복수의 조작된 및/또는 재조합 Arc 폴리펩티드와 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 또는 10:1의 비로 혼합된다. 다른 경우에서, 복수의 비-Arc 캡시드 형성 서브유닛은 복수의 조작된 및/또는 재조합 Arc 폴리펩티드와 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 또는 1:10의 비로 혼합된다.
- [0213] 일부 경우에서, 로딩된 Arc 기반 캡시드 또는 endo-Gag 기반 캡시드를 생성하기에 충분한 시간은 적어도 약 5분, 적어도 약 10분, 적어도 약 20분, 적어도 약 30분, 적어도 약 1시간, 적어도 약 2시간, 적어도 약 4시간, 적어도 약 6시간, 적어도 약 10시간, 적어도 약 12시간, 적어도 약 24시간, 또는 그 초과이다.
- [0214] 일부 경우에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 약 2°C 내지 약 37°C의 온도에서 제조된다. 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 약 2°C 내지 약 8°C, 약 2°C 내지 약 4°C, 약 20°C 내지 약 37°C, 약 25°C 내지 약 37°C, 약 20°C 내지 약 30°C, 약 25°C 내지 약 30°C, 또는 약 30°C 내지 약 37°C의 온도에서 제조된다.
- [0215] 일부 경우에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 실온에서 제조된다.
- [0216] 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 전신 투여용으로 추가로 제제화된다.
- [0217] 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 국소 투여용으로 추가로 제제화된다.
- [0218] 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 비경구(예컨대, 동맥내, 관절내, 피내, 병변내, 근육내, 안내, 골내 주입, 복강내, 척추강내, 정맥내, 유리체내, 또는 피하) 투여용으로 추가로 제제화된다.
- [0219] 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 국부 투여용으로 추가로 제제화된다.
- [0220] 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 경구 투여용으로 추가로 제제화된다.
- [0221] 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 설하 투여용으로 추가로 제제화된다.
- [0222] 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 에어로졸 투여용으로 추가로 제제화된다.
- [0223] 특정 실시양태에서, 또한 관심 부위로 카고를 전달하기 위한 Arc 기반 캡시드 또는 endo-Gag 기반 캡시드의 용도가 본원에 기재된다. 일부 예에서, 방법은 관심 부위의 세포를 캡시드의 세포 흡수를 촉진하기에 충분한 시간 동안 Arc 기반 캡시드 또는 endo-Gag 기반 캡시드와 접촉시키는 단계를 포함한다.
- [0224] 일부 경우에서, 세포는 근육 세포, 피부 세포, 혈액 세포, 또는 면역 세포(예컨대, T 세포 또는 B 세포)이다.
- [0225] 일부 예에서, 세포는 종양 세포, 예컨대 고형 종양 세포 또는 혈액 악성종양으로부터의 세포이다. 일부 경우에서, 고형 종양 세포는 방광암, 유방암, 뇌암, 결장직장암, 신장암, 간암, 폐암, 췌장암, 전립선암, 피부암, 위암, 또는 갑상선암으로부터의 세포이다. 일부 경우에서, 혈액 악성종양으로부터의 세포는 B 세포 악성종양 또는 T 세포 악성종양으로부터의 것이다. 일부 경우에서, 세포는 백혈병, 림프종, 골수종, 만성 림프구성 백혈병 (CLL), 소 림프구성 림프종(SLL), 미만성 거대 B 세포 림프종(DLBCL), 소포 림프종, 맨틀 세포 림프종, 버킷 림

프종, 피부 T 세포 림프종, 말초 T 세포 림프종, 다발성 골수종, 형질세포종, 급성 림프아구성 백혈병(ALL), 급성 골수성 백혈병(AML) 또는 만성 골수성 백혈병(CML)으로부터의 것이다.

- [0226] 일부 실시양태에서, 세포는 체세포이다. 일부 예에서, 세포는 혈액 세포, 피부 세포, 결합 조직 세포, 뼈 세포, 근육 세포, 또는 기관으로부터의 세포이다.
- [0227] 일부 실시양태에서, 세포는 상피 세포, 결합 조직 세포, 근육 세포, 또는 뉴런이다.
- [0228] 일부 예에서, 세포는 내배엽 세포, 중배엽 세포, 또는 외배엽이다. 일부 예에서, 내배엽은 호흡기, 장, 간, 담낭, 췌장, 랑게르한스섬, 갑상선 또는 후장의 세포를 포함한다. 일부 경우에서, 중배엽은 골연골전구세포, 근육 세포, 소화계로부터의 세포, 신장 줄기 세포, 생식계로부터의 세포, 순환계로부터의 세포(예컨대, 내피 세포)를 포함한다. 외배엽으로부터의 예시적인 세포는 상피 세포, 뇌하수체 전엽 세포, 말초 신경계 세포, 신경내분비계 세포, 눈 세포, 중추 신경계 세포, 뇌실막 세포, 또는 송과체 세포를 포함한다. 일부 경우에서, 중추 및 말초 신경계로부터 유래되는 세포는 뉴런, 슈반 세포, 위성 아교 세포, 희돌기아교세포 또는 성상 세포를 포함한다. 일부 경우에서, 뉴런은 인터뉴런, 피라미드 뉴런, 가바성(gabaergic) 뉴런, 도파민성 뉴런, 세로토닌성 뉴런, 글루타메이트성 뉴런, 척수로부터의 운동 뉴런, 또는 억제성 척수 뉴런을 추가로 포함한다.
- [0229] 일부 실시양태에서, 세포는 줄기 세포 또는 전구 세포이다. 일부 경우에서, 세포가 중간엽 줄기 또는 전구 세포이다. 다른 경우에서, 세포는 조혈 줄기 또는 전구 세포이다.
- [0230] 일부 경우에서, 표적 단백질은 세포에서 과발현되거나 세포에서 고갈된다. 일부 경우에서, 표적 단백질은 세포에서 과발현된다. 추가의 경우에서, 표적 단백질은 세포에서 고갈된다.
- [0231] 일부 경우에서, 세포의 표적 유전자는 하나 이상의 돌연변이를 갖는다.
- [0232] 일부 경우에서, 세포는 손상된 스플라이싱 메커니즘을 포함한다.
- [0233] 일부 예에서, Arc 기반 캡시드는 이를 필요로 하는 대상체에게 전신 투여된다.
- [0234] 다른 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 이를 필요로 하는 대상체에게 국소적으로 투여된다.
- [0235] 일부 실시양태에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 이를 필요로 하는 대상체에게 비경구적으로, 경구적으로, 국부적으로, 설하를 통해, 또는 에어로졸에 의해 투여된다. 일부 경우에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 이를 필요로 하는 대상체에게 비경구적으로 투여된다. 다른 경우에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 이를 필요로 하는 대상체에게 경구적으로 투여된다. 추가의 경우에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 이를 필요로 하는 대상체에게 국부적으로, 설하를 통해, 또는 에어로졸에 의해 투여된다.
- [0236] 일부 실시양태에서, 전달 성분은 관심 부위의 표적 전달을 위해 Arc 기반 캡시드 또는 endo-Gag 기반 캡시드와 조합된다. 일부 예에서, 전달 성분은 담체, 예컨대 미셀, 리포솜, 또는 미세소포와 같은 세포외 소포; 또는 바이러스 외피를 포함한다.
- [0237] 일부 예에서, 전달 성분은 자체 전달 성분을 포함하지 않는(예컨대, 제2 폴리펩티드가 존재하지 않는) Arc 기반 캡시드 또는 endo-Gag 기반 캡시드에 대한 1차 전달 비히클의 역할을 한다. 이러한 경우에서, 전달 성분은 Arc 기반 캡시드 또는 endo-Gag 기반 캡시드를 관심 표적 부위로 유도하고 임의적으로 세포내 흡수를 촉진한다.
- [0238] 다른 예에서, 전달 성분은 Arc 기반 캡시드의 제2 폴리펩티드의 표적 특이성 및/또는 민감도를 향상시킨다. 이러한 경우에서, 전달 성분은 표적 부위에 대한 Arc 기반 캡시드 또는 endo-Gag 기반 캡시드의 특이성 및/또는 친화성을 향상시킨다. 추가의 경우에서, 전달 성분은 특이성 및/또는 친화성을 약 2배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 20배, 30배, 50배, 100배, 200배, 500배, 또는 그 초과로 향상시킨다. 추가의 경우에서, 전달 성분은 특이성 및/또는 친화성을 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 500%, 또는 그 초과로 향상시킨다. 또한, 전달 성분은 임의적으로 표적외 효과를 약 2배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 20배, 30배, 50배, 100배, 200배, 500배, 또는 그 초과로 감소시킨다. 또한, 전달 성분은 임의적으로 표적외 효과를 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 500%, 또는 그 초과로 감소시킨다.
- [0239] 추가의 예에서, 전달 성분은 Arc 기반 캡시드를 일반적인 표적 영역(예컨대, 중앙 미세환경)으로 수송하는 제1 비히클의 역할을 하고 Arc 기반 또는 endo-Gag 기반 캡시드의 제2 폴리펩티드는 Arc 기반 캡시드 또는 endo-Gag

기반 캡시드를 특정 표적 부위로 유도하고 임의적으로 세포내 흡수를 촉진하는 제2 전달 분자의 역할을 한다. 이러한 경우에서, 전달 성분은 표적외 효과를 약 2배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 20배, 30배, 50배, 100배, 200배, 500배, 또는 그 초과로 감소시킨다. 이러한 경우에서, 전달 성분은 표적외 효과를 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 500%, 또는 그 초과로 감소시킨다.

- [0240] 추가의 예에서, 전달 성분은 Arc 기반 캡시드를 관심 표적 부위로 수송하는 제1 비히클의 역할을 하고, Arc 기반 또는 endo-Gag 기반 캡시드의 제2 폴리펩티드는 세포내 흡수를 촉진하는 제2 전달 분자의 역할을 한다.
- [0241] 일부 실시양태에서, 전달 성분은 세포의 소포를 포함한다. 일부 예에서, 세포의 소포는 미세소포, 리포솜 또는 미셀을 포함한다. 일부 예에서, 세포의 소포는 약 10 nm 내지 약 2000 nm, 약 10 nm 내지 약 1000 nm, 약 10 nm 내지 약 800 nm, 약 20 nm 내지 약 600 nm, 약 30 nm 내지 약 500 nm, 약 50 nm 내지 약 200 nm, 또는 약 80 nm 내지 약 100 nm의 직경을 갖는다.
- [0242] 일부 실시양태에서, 전달 성분은 미세소포를 포함한다. 순환하는 미세소포 또는 미세입자로도 알려진 미세소포는 인지질을 포함하는 막 결합된 소포이다. 일부 예에서, 미세소포는 약 50 nm 내지 약 1000 nm, 약 100 nm 내지 약 800 nm, 약 200 nm 내지 약 500 nm, 또는 약 50 nm 내지 약 400 nm의 직경을 갖는다.
- [0243] 일부 예에서, 미세소포는 세포막 역전, 세포외배출, 흘림(shedding), 블레빙(blebbing), 또는 출아(budding)로부터 유래된다. 일부 경우에서, 미세소포는 분화된 세포로부터 생성된다. 다른 예에서, 미세소포는 미분화된 세포로부터, 예컨대 모세포, 전구 세포 또는 줄기 세포에 의해 생성된다.
- [0244] 일부 실시양태에서, 전달 성분은 리포솜을 포함한다. 일부 예에서, 리포솜은 관심 부위 또는 영역으로의 표적 전달을 위해 리포솜의 표면에 존재하는 복수의 리포펩티드를 포함한다. 일부 경우에서, 리포솜이 표적 세포와 융합되어 리포솜의 내용물이 표적 세포로 비워진다. 일부 경우에서, 리포솜이 식세포성 세포에 의해 세포내이입된다. 세포내이입은 리소좀 지질의 리소좀내 분해 및 캡슐화된 작용제의 방출로 이어진다.
- [0245] 통합에 적합한 예시적인 리포솜은 다층 소포(MLV), 올리고층 소포(OLV), 단층 소포(UV), 작은 단층 소포(SUV), 중간 크기 단층 소포(MUV), 큰(large) 단층 소포(LUV), 거대(giant) 단층 소포(GUV), 다낭(multivesicular) 소포(MVV), 역상 증발법으로 제조된 단일 또는 올리고층 소포(REV), 역상 증발법으로 제조된 다층 소포(MLV-REV), 안정한 다수층(plurilamellar) 소포(SPLV), 냉동 및 해동된 MLV(FATMLV), 압출 방법으로 제조된 소포(VET), 프렌치 프레스에 의해 제조된 소포(FPV), 융합에 의해 제조된 소포(FUV), 탈수-재수화 소포(DRV), 및 버블솜(BSV)을 포함하지만 이에 제한되지 않는다. 일부 예에서, 리포솜은 알피폴(Amphipol)(A8-35)을 포함한다. 리포솜 제조 기술은, 예컨대 문헌(COLLOIDAL DRUG DELIVERY SYSTEMS, vol. 66 (J. Kreuter ed., Marcel Dekker, Inc. (1994)))에 기재되어 있다.
- [0246] 제조 방법에 따라, 리포솜은 단층 또는 다층이고, 직경이 약 20 nm부터 약 1000 nm 초과로 크기가 다양하다.
- [0247] 일부 예에서, 본원에 제공된 리포솜은 또한 담체 지질을 포함한다. 일부 실시양태에서, 담체 지질은 인지질이다. 리포솜을 형성할 수 있는 담체 지질은 디팔미토일포스파티딜콜린(DPPC), 포스파티딜콜린(PC; 레시틴), 포스파티드산(PA), 포스파티딜글리세롤(PG), 포스파티딜에탄올아민(PE), 또는 포스파티딜세린(PS)을 포함하지만 이에 제한되지 않는다. 다른 적합한 인지질은 추가로 디스테아로일포스파티딜콜린(DSPC), 디미리스토일포스파티딜콜린(DMPC), 디팔미토일포스파티딜글리세롤(DPPG), 디스테아로일포스파티딜글리세롤(DSPG), 디미리스토일포스파티딜글리세롤(DMPCG); 디미리스토일포스파티드산(DMPA), 디스테아로일포스파티드산(DSPA), 디팔미토일포스파티딜세린(DPSP), 디미리스토일포스파티딜세린(ASD), 디스테아로일포스파티딜세린(DSPS), 디팔미토일포스파티딜에탄올아민(DPPE), 디미리스토일포스파티딜에탄올아민(DMPE), 디스테아로일포스파티딜에탄올아민(DSPE) 등을 포함하지만 이에 제한되지 않는다. 일부 실시양태에서, 리포솜은 리포솜 형성을 조정하는 스테롤(예컨대, 콜레스테롤)을 추가로 포함한다. 담체 지질은 임의적으로 비-포스페이트 극성 지질이다.
- [0248] 일부 실시양태에서, 전달 성분은 미셀을 포함한다. 일부 예에서, 미셀은 약 2 nm 내지 약 250 nm, 약 20 nm 내지 약 200 nm, 약 20 nm 내지 약 100 nm, 또는 약 50 내지 약 100 nm의 직경을 갖는다.
- [0249] 일부 경우에서, 미셀은 소수성 코어가 친수성 셸로 둘러싸여진 코어 셸 구조를 특징으로 하는 중합체 미셀이다. 일부 경우에서, 친수성 셸은 친수성 중합체 또는 공중합체 및 pH 민감성 성분을 추가로 포함한다.
- [0250] 예시적인 친수성 중합체 또는 공중합체는 폴리(N-치환된 아크릴아미드), 폴리(N-아크릴로일 피롤리딘), 폴리(N-아크릴로일 피페리딘), 폴리(N-아크릴-L-아미노산 아미드), 폴리(에틸 옥사졸린), 메틸셀룰로스, 히드록시프로필 아크릴레이트, 히드록시알킬 셀룰로스 유도체 및 폴리(비닐 알콜), 폴리(N-이소프로필 아크릴아미드), 폴리

(N-비닐-2-피롤리돈), 폴리에틸렌글리콜 유도체, 및 이들의 조합을 포함하지만 이에 제한되지 않는다.

- [0251] pH 민감성 모이어티는 알킬아크릴산, 예컨대 메타크릴산, 에틸아크릴산, 프로필 아크릴산 및 부틸 아크릴산과 같은 알킬 아크릴산, 또는 아미노산, 예컨대 글루탐산을 포함하지만 이에 제한되지 않는다.
- [0252] 일부 경우에서, 소수성 모이어티는 미셀의 코어를 구성하고, 예컨대 옥타 데실 아크릴레이트와 같은 단일 알킬 쇠 또는 포스파티딜에탄올아민 또는 디옥타데실아민과 같은 이중 쇠 알킬 화합물을 포함한다. 일부 경우에서, 소수성 모이어티는 임의적으로 폴리(락트산) 또는 폴리(e-카프롤락톤)와 같은 수불용성 중합체이다.
- [0253] pH 민감성을 나타내는 중합체성 미셀도 고려되며, 예컨대 메타크릴산, 메타크릴산 에스테르 및 아크릴산 에스테르, 폴리비닐 아세테이트 프탈레이트, 히드록시프로필 메틸 셀룰로스 프탈레이트, 셀룰로스 아세테이트 프탈레이트, 또는 셀룰로스 아세테이트 트리멜리테이트로부터의 공중합체를 포함하지만 이에 제한되지 않는 pH-민감성 중합체를 사용하여 형성된다.
- [0254] 일부 실시양태에서, 전달 성분은 바이러스 외피를 포함한다. 바이러스 외피는 당단백질, 인지질, 및 숙주로부터 수득되는 추가의 단백질을 포함한다. 일부 예에서, 바이러스 외피는 광범위한 표적 세포에 대해 허용된다. 다른 경우에서, 바이러스 외피는 허용되지 않으며 관심 표적 세포에 특이적이다. 일부 경우에서, 바이러스 외피는 세포-특이적 결합 단백질 및 임의적으로 표적 세포로의 카고의 용해를 돕는 용해형성 분자를 포함한다. 일부 경우에서, 바이러스 외피는 내인성 바이러스 외피를 포함한다. 다른 경우에서, 바이러스 외피는 하나 이상의 외래 단백질을 포함하는 변형된 외피이다.
- [0255] 일부 예에서, 바이러스 외피는 DNA 바이러스로부터 유래된다. 예시적인 외피 DNA 바이러스는 헤르페스비리다에 (*Herpesviridae*), 폭스비리다에(*Poxviridae*), 및 헤파드나비리다에(*Hepadnaviridae*) 과로부터의 바이러스를 포함한다.
- [0256] 다른 경우에서, 바이러스 외피는 RNA 바이러스로부터 유래된다. 예시적인 외피 RNA 바이러스는 부냐비리다에 (*Bunyaviridae*), 코로나비리다에(*Coronaviridae*), 필로비리다에(*Filoviridae*), 플라비비리다에(*Flaviviridae*), 오르토믹소비리다에(*Orthomyxoviridae*), 파라믹소비리다에(*Paramyxoviridae*), 라브도비리다에(*Rhabdoviridae*) 및 토가비리다에(*Togaviridae*) 과로부터의 바이러스를 포함한다.
- [0257] 추가의 예에서, 바이러스 외피는 레트로비리다에(*Retroviridae*) 과로부터의 바이러스로부터 유래된다.
- [0258] 일부 실시양태에서, 바이러스 외피는 종양용해성 바이러스, 예컨대 헤르페스비리다에(*Herpesviridae*)(예컨대, HSV1) 또는 폭스비리다에(*Poxviridae*)(예컨대, 박시니아 바이러스 및 믹소마 바이러스) 과로부터의 종양용해성 DNA 바이러스; 또는 라브도비리다에(*Rhabdoviridae*)(예컨대, VSV) 또는 파라믹소비리다에(*Paramyxoviridae*)(예컨대, MV 및 NDV) 과로부터의 종양용해성 RNA 바이러스로부터의 것이다.
- [0259] 일부 예에서, 바이러스 외피는 항원 또는 세포 표면 분자에 결합하는 외래 또는 조작된 단백질을 추가로 포함한다. 표적화를 위한 예시적인 항원 및 세포 표면 분자는 P-당단백질, Her2/Neu, 에리트로포이에틴(EPO), 표피 성장 인자 수용체(EGFR), 혈관 내피 성장 인자 수용체(VEGF-R), 카드헤린, 암배아 항원(CEA), CD4, CD8, CD19, CD20, CD33, CD34, CD45, CD117(c-kit), CD133, HLA-A, HLA-B, HLA-C, 케모카인 수용체 5(CCR5), 줄기 세포 마커 ABCG2 수송체, 난소암 항원 CA125, 면역글로불린, 인테그린, 전립선 특이적 항원(PSA), 전립선 줄기 세포 항원(PSCA), 수지상 세포 특이적 세포간 부착 분자 3-그래빙 비-인테그린(DC-SIGN), 티로글로불린, 파립구-대식구 콜로니 자극 인자(GM-CSF), 근원성 분화 촉진 인자-1(MyD-1), Leu-7(CD57), LeuM-1, 모노클로날 항체 Ki-67(Ki-67)에 의해 정의되는 세포 증식 관련된 인간 핵 항원, 바이러스 외피 단백질, HIV gp120 또는 트랜스페린 수용체를 포함하지만 이에 제한되지 않는다.
- [0260] 일부 실시양태에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 시험관 내 사용을 위한 것이다.
- [0261] 일부 예에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 생체외 사용을 위한 것이다.
- [0262] 일부 경우에서, Arc 기반 캡시드 또는 endo-Gag 기반 캡시드는 생체내 사용을 위한 것이다.
- [0263] **키트/제조 물품**
- [0264] 특정 실시양태에서, 본원에 기재된 하나 이상의 방법과 함께 사용하기 위한 키트 및 제조 물품이 본원에 개시된다. 이러한 키트는 바이알, 튜브 등과 같은 하나 이상의 용기를 수용하도록 구획화된 캐리어, 패키지 또는 용기를 포함하며, 각각의 용기(들)는 본원에 기재된 방법에 사용되는 개별 요소 중 하나를 포함한다. 적합한 용기는, 예컨대 병, 바이알, 주사기 및 시험관을 포함한다. 일 실시양태에서, 용기는 유리 또는 플라스틱과 같

은 다양한 물질로부터 형성된다.

[0265] 예컨대, 용기(들)는 상술된 재조합 또는 조작된 Arc 또는 endo-Gag 폴리펩티드를 포함한다. 이러한 키트는 임의적으로 본원에 기재된 방법에서의 사용과 관련된 식별 설명 또는 라벨 또는 지침을 포함한다. 예컨대, 키트는 전형적으로 내용물이 나열된 라벨 및/또는 사용 지침, 및 사용 지침을 갖는 패키지 삽입물을 포함한다. 전형적으로 지침 세트도 포함될 것이다.

[0266] **특정 용어**

[0267] 달리 정의되지 않는 한, 본원 사용된 모든 기술 및 과학 용어는 일반적으로 이해되는 것과 동일한 의미를 갖는다. 상세한 설명은 단지 예시적이고 설명적인 것이며 청구된 임의의 주제를 제한하지 않는다는 것을 이해해야 한다. 본 출원에서, 단수의 사용은 특별히 달리 언급되지 않는 한 복수를 포함한다. 명세서에서 사용되는 바와 같이, 단수 형태는 문맥이 달리 명확하게 지시하지 않는 한 복수 지시 대상을 포함한다는 점에 유의해야 한다. 본 출원에서 "또는"의 사용은 달리 명시되지 않는 한 "및/또는"을 의미한다. 또한, 용어 "포함하는" 뿐만 아니라 "포함한다" 및 "포함되는"과 같은 다른 형태의 사용은 제한적이지 않다.

[0268] 본 발명의 다양한 특징이 단일 실시양태의 맥락에서 설명될 수 있지만, 특징은 또한 개별적으로 또는 임의의 적절한 조합으로 제공될 수 있다. 반대로, 본 발명은 명확성을 위해 별도의 실시양태의 맥락에서 본원에서 설명될 수 있지만, 본 발명은 또한 단일 실시양태로 구현될 수 있다.

[0269] 본 명세서에서 "일부 실시양태", "일 실시양태", "한 실시양태" 또는 "다른 실시양태"에 대한 언급은, 실시양태와 관련하여 설명된 특정 특징, 구조 또는 특성이 본 발명의 적어도 일부 실시양태에 포함되며 반드시 모든 실시양태에 필요한 것은 아니라는 것을 의미한다.

[0270] 본원에서 사용되는 바와 같이, 범위 및 양은 "약" 특정 값 또는 범위로 표현될 수 있다. 약은 또한 정확한 양을 포함한다. 따라서, "약 5  $\mu$ L"는 "약 5  $\mu$ L" 및 "5  $\mu$ L"을 의미한다. 일반적으로, 용어 "약"은 실험 오차 내에 있을 것으로 예상되는 양을 포함한다.

[0271] 본원에 사용된 섹션 제목은 조직화 목적으로만 사용되며 기재된 주제를 제한하는 것으로 해석되어서는 안된다.

[0272] 본원에서 사용되는 바와 같이, 본원에 기재된 CA N-로브의 서열은 인간 CA N-로브에 상응한다. 일부 예에서, 인간 CA N-로브는 서열번호 1의 잔기 207-278을 포함한다. 일부 예에서, 본원에 기재된 CA N-로브는 서열번호 1의 잔기 207-278과 약 30%, 40%, 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% 98% 또는 99% 서열 동일성을 포함한다. 일부 경우에서, 본원에 기재된 CA N-로브는 인간 CA N-로브와 구조적 유사성을 공유한다. 예컨대, 본원에 기재된 CA N-로브는 인간 CA N-로브와 약 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% 98% 또는 99% 구조적 유사성을 공유한다. 일부 경우에서, CA N-로브는 높은 구조적 유사성(예컨대, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% 98% 또는 99% 구조적 유사성)을 공유하지만 높은 서열 동일성은 공유하지 않는다(예컨대, 서열 동일성은 80% 미만, 70% 미만, 60% 미만, 50% 미만, 40% 미만 또는 30% 미만임). 일부 경우에서, CA N-로브는 서열번호 1의 잔기 207-278을 포함한다.

[0273] 본원에서 사용되는 바와 같이, 본원에 기재된 CA C-로브의 서열은 인간 CA C-로브에 상응한다. 일부 예에서, 인간 CA C-로브는 서열번호 1의 잔기 278-370을 포함한다. 일부 예에서, 본원에 기재된 CA C-로브는 서열번호 1의 잔기 278-370과 약 30%, 40%, 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% 98% 또는 99% 서열 동일성을 포함한다. 일부 경우에서, 본원에 기재된 CA C-로브는 인간 CA C-로브와 구조적 유사성을 공유한다. 예컨대, 본원에 기재된 CA C-로브는 인간 CA C-로브와 약 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% 98% 또는 99% 구조적 유사성을 공유한다. 일부 경우에서, CA C-로브는 높은 구조적 유사성(예컨대, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% 98% 또는 99% 구조적 유사성)을 공유하지만 높은 서열 동일성은 공유하지 않는다(예컨대, 서열 동일성은 80% 미만, 70% 미만, 60% 미만, 50% 미만, 40% 미만 또는 30% 미만임). 일부 경우에서, CA C-로브는 서열번호 1의 잔기 278-370을 포함한다.

[0274] 본원에 사용된 용어 "개체(들)", "대상체(들)" 및 "환자(들)"는 임의의 포유동물을 의미한다. 일부 실시양태에서, 포유동물은 인간이다. 일부 실시양태에서, 포유동물은 비-인간이다. 어떠한 용어도 의료 종사자(예컨대, 의사, 등록 간호사, 임상 간호사, 의사 보조원, 잡역부 또는 호스피스 작업자)의 감독(예컨대, 지속적 또는 간헐적)에 의해 특징지어지는 상황에 요구되거나 이에 제한되지 않는다.

[0275] **실시예**

- [0276] 이들 실시예는 예시 목적으로만 제공되며 본원에 제공된 청구범위의 범위를 제한하지 않는다.
- [0277] **실시예 1 - 재조합 Arc 단백질 및 조작된 Arc 단백질을 코딩하는 DNA 벡터의 구축**
- [0278] Arc 발현을 위한 재조합 DNA 벡터를 구축하기 위해, 개시 메티오닌을 제외한 전장 cDNA 오픈 리딩 프레임을 클로닝 벡터에 삽입한 후 표준 방법에 따라 발현 벡터에 전달한다. endo-Gag 단백질을 발현하기 위한 재조합 DNA 벡터를 구축하는 데에도 동일한 접근법이 이용된다. 인간 Arc cDNA는 주석이 달린 매트릭스 도메인(MA) 및 캡시드 도메인을 포함한다. 캡시드 도메인은 N-말단 로브(NTD) 및 C-말단 로브(CTD)를 갖는다. 도 1은 인간 Arc 단백질의 구조 및 비단백, 오리너구리 및 오르카(Orca)의 예측되는 Arc 구조를 예시한다.
- [0279] 조작된 Arc 단백질을 코딩하는 cDNA는 임의적으로 상이한 종으로부터의 Arc 서열을 재조합함으로써(도 2), 다른 단백질로부터의 기능 도메인을 Arc 단백질에 삽입함으로써(도 3a), Arc 단백질의 서열을 변형함으로써(도 3b), 및/또는 도 2-3에 예시된 접근법의 임의의 조합에 의해 생성된다. 조작된 endo-Gag 단백질을 코딩하는 cDNA도 마찬가지로 상이한 종으로부터의 endo-Gag 서열을 재조합함으로써, 다른 단백질로부터의 기능 도메인을 endo-Gag 단백질에 삽입함으로써, endo-Gag 단백질의 서열을 변형함으로써, 및/또는 이들 접근법의 임의의 조합에 의해 생성된다. 또한, 조작된 endo-Gag 단백질은 임의적으로 Arc 서열을 함유하고 조작된 Arc 단백질은 임의적으로 endo-Gag 서열을 함유한다. 조작된 Arc 및 endo-Gag 단백질 단량체는 캡시드로 조립된다.
- [0280] 표 1의 Arc 및 endo-Gag 단백질을 코딩하는 cDNA를 pET-41 a(+)(EMD Millipore (Novagen) Cat# 70566)로부터 유래되는 발현 벡터에 삽입하였다. pET-41 a (+)의 전체 클로닝 부위를 제거하고, 서열번호 58의 아미노산 서열을 갖는 대안적 N-말단 태그를 코딩하는 서열번호 57의 뉴클레오티드 서열을 갖고 6xHis 태그(서열번호 59), 6개 아미노산 스페이서(서열번호 60), 및 AcTEV™ 절단 부위(서열번호 61)를 포함하는 DNA로 대체하였다. 시작 메티오닌 코돈이 없는 Arc 및 endo-Gag 오픈 리딩 프레임을 깃슨(Gibson) 조립에 의해 AcTEV™ 절단 부위 뒤에 삽입하였다(Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases". Nature Methods. 6 (5): 343-345). 발현 및 AcTEV™ 절단 후, 생성된 Arc 또는 endo-Gag 단백질의 N-말단은 AcTEV™ 절단 부위로부터 단일 잔기 글리신을 갖는다.
- [0281] 서열번호 57
- ATGCATCACCATCACCATCACGGCTCAGGGTCTGGTAGCGAAAATCTGT
- [0282] ACTTCCAGGGG
- [0283] 서열번호 58 MHHHHHHGSGSGSENLYFQG
- [0284] 서열번호 59 HHHHHHH
- [0285] 서열번호 60 GSGSGS
- [0286] 서열번호 61 ENLYFQG

표 1

Arc 및 endo-Gag 폴리펩티드 및 뉴클레오티드의 서열

유전자명	종		서열 ID	서열번호	
	일반명	고유명		아미노산	DNA
Arc	인간	호모 사피엔스( <i>Homo sapiens</i> )	NP_056008.1	1	29
Arc	범고래	오르시누스 오르카( <i>Orcinus orca</i> )	XP_004265337.1	2	30
Arc	흰꼬리 사슴	오도코일레우스 비르기니아누스 텍사누스( <i>Odocoileus virginianus texanus</i> )	XP_020755692.1	3	31
Arc	오리너구리	오르니토르힌쿠스 아나티누스( <i>Ornithorhynchus anatinus</i> )	XP_001512750.1	4	32
Arc	거위	안세르 시그노이데스 도메스티쿠스( <i>Anser cygnoides domesticus</i> )	XP_013046406.1	5	33
Arc	달마티온 펠리칸	펠레카누스 크리스푸스( <i>Pelecanus crispus</i> )	KFQ60200.1	6	34
Arc	흰꼬리 수리	할리아에투스 알비실라( <i>Haliaeetus albicilla</i> )	KFQ04633.1	7	35
Arc	킹 코브라	오피오파구스 한나( <i>Ophiophagus hannah</i> )	ETE60609.1	8	36
Arc	조기어	아우스트로폰돌루스 림나에우스( <i>Austrofundulus limnaeus</i> )	XP_013881732.1	9	37
Arc	향유 고래	피세테르 카토돈( <i>Physeter catodon</i> )	XP_007119193.2	10	38
Arc	칠면조	멜레아그리스 갈로파보( <i>Meleagris gallopavo</i> )	XP_010707654.1	11	39
Arc	중앙 턱수염 도마뱀	포고나 비티셉스( <i>Pogona vitticeps</i> )	XP_020633722.1	12	40
Arc	양쯔강 악어	엘리게이터 시넨시스( <i>Alligator sinensis</i> )	XP_006027442.1	13	41
Arc	미국 악어	엘리게이터 미시시피엔시스( <i>Alligator mississippiensis</i> )	XP_019337372.1	14	42

[0287]

Arc	일본 게코	게코 자포니쿠스( <i>Gekko japonicus</i> )	XP_015273745.1	15	43
PNM A3	인간	호모 사피엔스	NP_001269464.1	16	44
PNM A5	인간	호모 사피엔스	NP_001096620.1	17	45
PNM A6A	인간	호모 사피엔스	NP_116271.3	18	46
PNM A6B	인간	호모 사피엔스	SP_P0C5W0.1	19	47
RTL3	인간	호모 사피엔스	NP_689907.1	20	48
RTL6	인간	호모 사피엔스	NP_115663.2	21	49
RTL8 A	인간	호모 사피엔스	NP_001071640.1	22	50
RTL8 B	인간	호모 사피엔스	NP_001071641.1	23	51
BOP	인간	호모 사피엔스	NP_078903.3	24	52
LDO C1	인간	호모 사피엔스	NP_036449.1	25	53
ZNF18	인간	호모 사피엔스	NP_001290210.1	26	54
MOA P1	인간	호모 사피엔스	AAG31786.1	27	55
PEG10	인간	호모 사피엔스	NP_055883.2	28	56

[0288]

[0289]

실시예 2 - Arc 및 endo-Gag 단백질의 발현 및 정제

- [0290] Arc 및 endo-Gag 오픈 리딩 프레임을 포함하는 발현 벡터 구축물을 Rosetta 2 (DE3)pLysS 이. 콜라이 균주 (Millipore Sigma, Cat# 71403)로 형질전환시켰다. Arc 또는 endo-Gag 발현을 0.1mM IPTG로 유도된 후 16°C에서 16시간 인큐베이션하였다. 세포 펠렛을 20mM 인산나트륨 pH 7.4, 0.1M NaCl, 40mM 이미다졸, 1mM DTT, 및 10% 글리세롤에서 초음파 처리에 의해 용해시켰다. 용해물을 과량의 TURBO DNase(Thermo Fisher Scientific, Cat# AM2238), RNase 콕테일(Thermo Fisher Scientific, Cat# AM2286) 및 벤조나제 뉴클레아제(Millipore Sigma, Cat# 71205)로 처리하여 핵산을 제거하였다. NaCl 농도를 0.5M로 조정하기 위해 NaCl을 용해물에 첨가한 다음 원심분리 및 여과하여 세포 과편을 제거하였다. 6xHis 태그부착된 제조합 단백질을 HisTrap HP 컬럼(GE Healthcare, Cat# 17-5247-01)에 로딩하고, 버퍼 A(20mM 인산나트륨 pH 7.4, 0.5M NaCl, 40mM 이미다졸 및 10% 글리세롤)로 세척하고, 버퍼 B의 선형 구배(20mM 인산나트륨 pH 7.4, 0.5M NaCl, 500mM 이미다졸, 및 10% 글리세롤)로 용출하였다. 수집 튜브에 용출물 1ml당 10 µl의 0.5M EDTA pH 8.0을 미리 보충하였다. 생성된 Arc 또는 endo-Gag 단백질은 SDS-PAGE 분석에 의해 나타나는 바와 같이 일반적으로 95% 초과로 순수하며, 1L의 박테리아 배양물당 최대 50mg의 수율을 갖는다(도 4a).
- [0291] 모노 Q 5/50 GL 컬럼(GE Healthcare, Cat# 17516601)에서 음이온 교환 크로마토그래피로 잔류 핵산을 제거하였다. 컬럼에 로딩하기 전에, 제조합 단백질을 제조업자의 프로토콜에 따라 "피어스(Pierce) 단백질 농축기 PES, 10K MWCO, 5-20 ml"(Thermo Scientific, Cat#88528)를 사용하여 버퍼 C(20 mM Tris-HCl pH 8.0, 100 mM NaCl, 및 10% 글리세롤)로 버퍼 교환하였다. 로딩 후, 모노 Q 수지를 2ml의 버퍼 C로 세척하였다. Arc 및 endo-Gag 단백질을 버퍼 D의 선형 구배(20mM Tris-HCl pH 8.0, 500mM NaCl 및 10% 글리세롤)를 사용하여 용출하였다. RNA는 Arc로부터 효율적으로 분리되고 600mM NaCl에서 용출되었다(도 4b).
- [0292] N-말단 6xHis 태그 및 스페이서를 10kDa MWCO PES 농축기를 사용하여 모노 Q 정제된 Arc의 농축 피크 분획으로부터 제거한 다음 10% v/v의 AcTEV™ 프로테아제(Invitrogen™ #12575023)로 처리하였다. 절단 효율은 SDS-PAGE 검정에 의해 나타나는 바와 같이 99% 초과이다. 이어서, 단백질을 HisTrap 버퍼 A로 희석하고 HisTrap HP 수지로 세척한다. 생성된 정제된 Arc는 N-말단 글리신 잔기를 가지며 개시 메티오닌을 함유하지 않는다.
- [0293] **실시예 3 - 캡시드 조립**
- [0294] 절단된 Arc 단백질(1mg/mL)을 20kDa MWCO 투석 카세트에 로딩하고 실온에서 1M 인산나트륨(pH 7.5)에서 밤새 투석하였다. 다음날, 용액을 카세트로부터 꺼내고 마이크로원심분리기 튜브로 옮기고 탁상용 원심분리기에서 5분 동안 최대 속도로 회전시켰다. 상청액을 100kDa MWCO 재생 셀룰로스 아미콘 한외여과 원심 농축기로 옮겼다. 버퍼를 PBS pH 7.5로 교환하고 부피를 20배 감소시켰다.
- [0295] 캡시드 조립을 투과 전자 현미경으로 검정하였다. EM 그리드(탄소 지지 필름, 사각 그리드, 400 메쉬 5-6 nm, 구리, CF400-Cu-UL)는 글로 방전에 의해 제조되었다. 정제된 Arc의 5 µL 샘플을 20초 동안 그리드에 적용한 다음 여과지를 사용하여 빨아들였다. 이어서, 그리드를 MilliQ H<sub>2</sub>O로 세척하고 H<sub>2</sub>O 중의 1% 우라닐 아세테이트 5 µL로 30초 동안 염색한 다음 1분 동안 공기 건조시켰다. Arc 캡시드의 이미지를 가탄(Gatan) 4k x 4k 원부(OneView) 카메라가 장착된 FEI Talos L120C TEM을 사용하여 획득하였다. 도 5는 농축된 인간 Arc 캡시드를 도시한다. 도 6은 다른 척추동물 종으로부터 제조합적으로 발현된 Arc 동원체로부터 형성된 캡시드를 도시한다. 도 7은 다른 척추동물 종으로부터 제조합적으로 발현된 endo-Gag 유전자로부터 형성된 캡시드를 도시한다.
- [0296] **실시예 4 - Arc 캡시드의 선택적 세포 내재화**
- [0297] 단리된 제조합 인간 Arc 단백질(0.5mg/ml)로부터 조립된 캡시드를 PBS(pH 8.5) 중의 NHS 에스테르 Alexa Fluor™ 594-NHS 염료(Invitrogen™ #A20004)(DMSO에 용해됨) 50몰 과량과 반응시켜 형광 표지하였다. 어둠 속에서 2시간 동안 반응을 진행시켰다. Alexa594 표지된 캡시드를 PBS(pH 7.5)로 밤새 어둠 속에서 실온에서 적어도 2회의 버퍼 교환과 함께 투석하여 임의의 표지되지 않은 염료를 제거하였다.
- [0298] HeLa 세포(ATCC® CCL-2™)를 실험 24시간 전에 96 웰 플레이트에 시딩하여 처리를 위해 ~80% 합류에 도달하였다. 이어서, 표지된 캡시드를 0.05mg/ml의 최종 캡시드 농도로 완전한 조직 배양 배지에 스파이킹하였다. 처리를 37°C에서 4시간 동안 진행된 다음, 이미징 전에 10µg/ml 핵스트(Hoechst) 핵 염료를 함유하는 이미징 배지(DMEM, 페놀 레드 없음, 10% FBS 및 20mM HEPES 포함)로 세포를 3회 세척한다. 형광 현미경 검사에서 점상 염색 패턴을 나타내었으며, 이는 Arc 캡시드가 HeLa 세포에 의해 내재화되었음을 시사한다(도 8). 동일한 조건 하에 Alexa Fluor™ 594 표지된 소 혈청 알부민(BSA)(최종 농도 0.05 mg/ml) 또는 45.6 µM Alexa Fluor™ 594를 투여한 후 세포내 염색이 거의 또는 전혀 관찰되지 않았다.

[0299] 실시예 5 - Arc 캡시드에 의한 이중 RNA 전달

[0300] (1M 인산나트륨으로의 투석에 의한) 캡시드 형성 동안 과량의 RNA에서 스파이킹함으로써 인간 Arc 캡시드에 Cre RNA를 로딩하였다. Cre RNA 로딩된 캡시드를 37°C에서 4시간 동안 0.05 mg/ml의 최종 캡시드 농도로 생물학적 삼중으로 HeLa 세포에 투여하였다. 이어서, RNA 추출 전에 세포를 얼음 냉각된 1xPBS로 3회 세척하였다 (Invitrogen™ TRIzol™ 시약 # 15596026). 정제된 세포 관련 RNA를 qPCR에 의해 기술적 삼중으로 정량화하고, 값을 세포 GAPDH 수준으로 정규화하고, 단백질 정제로부터 이어질 수 있는 에세리키아 콜라이 rrsA mRNA 및 Arc RNA와 비교하였다. 표 2는 PCR 반응에 사용된 프라이머를 나타낸다. 검출된 세포 관련 Cre RNA의 양은 Arc 캡시드에 Cre RNA를 로딩한 경우 Cre RNA가 로딩되지 않은 대조군 캡시드와 비교하여 >27배 더 높았다(도 9).

표 2

Arc 캡시드에 의해 HeLa 세포에 전달된 RNA 의 qPCR 정량화를 위한 프라이머

유전자- 프라이머	서열	서열번호
GAPDH-F	AAGCTCATTTCTGGTATGACAACGA	62
GAPDH-R	AGGGTCTCTCTTCTCTTGTGCT	63
rrsA-F	GCTCAACCTGGGAACATGATCTGAT	64
rrsA-R	TAATCCTGTTTGCTCCCCACGCTTT	65
Arc CDS-F	GGCCCCTCAGCTCCAGTGATTC	66
Arc CDS-R	CCTGTTGTCACTCTCCTGGCTCTGA	67
Cre CDS-F	GCCAAGACATAAGAAACCTCGCCT	68
Cre CDS-R	GTGAATCAACATCCTCCCTCCGTC	69

[0301]

[0302] 도 10은 Arc 또는 endo-Gag 캡시드에 의한 이중 RNA의 전달을 입증하는 대안적인 방법을 예시한다. 6xHis-태그 부착된 Arc 또는 endo-Gag 유전자를 숙주 세포에서 발현시킨다. 생성된 Arc 단백체를 캡시드 형성 조건 하에서 번역 가능한 Cre mRNA와 혼합하여 Cre mRNA 로딩된 캡시드를 형성한다. Cre 로딩된 캡시드를 LoxP-루시페라제 리포터 마우스에 투여한다. Cre mRNA가 마우스 세포에 성공적으로 전달되고 Cre 리컴비나제 단백질이 후속 번역 되면 리포터의 LoxP 부위가 재조합되어 임의적으로 루시페린 투여 시 생물발광 이미징에 의해 검출되는 루시페라제 발현이 유도된다. 이 방법은 후보 Arc 및 endo-Gag 유전자의 전달 가능성을 시험하는 데 이용된다. 양성 루시페라제 신호는 후보 Arc 또는 endo-Gag 유전자가 이중 카고를 통합하고 그 카고를 표적 세포에 전달하는 캡시드로 조립할 수 있는 Arc 또는 endo-Gag 단백질을 코딩함을 나타낸다.

[0303] 본 발명의 바람직한 실시양태가 본원에 제시되고 설명되었지만, 이러한 실시양태는 단지 예로서 제공된다는 것이 당업자에게 명백할 것이다. 이제 본 발명을 벗어나지 않고 다양한 변형, 변화 및 대체가 당업자에게 떠오를 것이다. 본원에 기재된 본 발명의 실시양태에 대한 다양한 대안이 본 발명을 실시하는 데 사용될 수 있음을 이해해야 한다. 하기의 청구범위는 본 발명의 범위를 정의하고, 이들 청구범위 내의 방법 및 구조 및 그 균등물이 이에 의해 커버되는 것으로 의도된다.

표 3

Arc 및 endo-Gag 아미노산 및 뉴클레오티드 서열

서열 번호 1

GELDHRTSGGLHAYPGPRGGQVAKPNVILQIGKCAEMLEHVRRTHRHLLAEVSKQVERELKGLHRSVKGLESN
LDGYVPTSDSQRWKKSIKACLCRCQETIANLERWVKREMHVWREVFYRLERWADRLESTGGKYPVGSSESARHT
VSVGVGGPESYCHEADGYDYTVSPYAITPPPAAGELPGQEPAEAQQYQPWVPGEDGQPSPGVDTQIFEDPREF
LSHLEEYLRQVGGSEEWLSQIQNHMNGPAKKWWEFKQGSVKNWVEFKKEFLQYSEGTLREAIQRELDLPQ
KQGEPLDQFLWRKRDLYQTLVDADEEEIIQYVVGTLQPKLKRFLRHPLPKTLEQLIQRGMEVQDDLEQAAEPA
GPHLPVEDEAETLTPAPNSESVASDRTQPE

서열 번호 2

GELDQRTTGGLHAYPAPRGGPVAKPNVILQIGKCAEMLEHVRRTHRHLLTEVSKQVERELKGLHRSVKGLESN
LDGYVPTGDSQRWRKSIKACLCRCQETIANLERWVKREMHVWREVFYRLERWADRLESMGGKYPVGSNPSR
HTTSVGVGGPESYGHEADTYDYTVSPYAITPPPAAGELPGQEAVEAQYPPWGLGEDGQPSPGVDTQIFEDPR
EFLSHLEEYLRQVGGSEEWLSQIQNHMNGPAKKWWEYKQGSVKNWVEFKKEFLQYSEGALSREAVQRELDL
PQKQGEPLDQFLWRKRDLYQTLVDADEEEIIQYVVGTLQPKLKRFLRPPLPKTLEQLIQKGMVEVDGLEQVAEP
ASPHLPTEEESEALTPALTSSESVASDRTQPE

서열 번호 3

GELDHRTTGGLHAYPAPRGGPAAKPNVILQIGKCAEMLEHVRRTHRHLLAEVSKQVERELKGLHRSVKGLESN
LDGYVPTGDSQRWKKSIKACLSRCQETIANLERWVKREMHVWREVFYRLERWADRLESGGGKYPVGSPPARH
TVSVGVGGPESYQDADNYDYTVSPYAITPPPAAGQLPGQEEVEAQYPPWAPGEDGQLSPGVDTQVDFEDPR
EFLRHLEDYLRQVGGSEEWLSQIQNHMNGPAKKWWEYKQGSVKNWVEFKKEFLQYSEGTLREAIQRELDL
PQKQGEPLDQFLWRKRDLYQTLVDADEEEIIQYVVGTLQPKLKRFLRPPLPKTLEQLIQKGMVEVDGLEQAAEP
AAEEAEALTPALTNESVASDRTQPE

[0304]

서열번호 4

GELDRLNPSGLHPSSGLHPYPGLRGGATAKPNVILQIGKCRAEMLEHVRKTRHLLTEVSRQVERELKGLHKS  
GKLESNLDGYVPSDSQRWKKSIKACLARCQETIAHLERWVKREMNWVREVFYRLERWADRLEAMGGKYPAG  
EQARRTVSVGVGGPETCCPGDESYDCPISPYAVPPSTGESPELDQGDQHYQQWFALPEESPVSPGVDTQIFED  
PREFLRHLEKYLKQVGGTEEDWLSQIQNHMNGPAKKWWEYKQGSVKNWLEFKKEFLQYSEGTLTRDALKREL  
DLPQKQGEPLDQFLWRKRDLYQTLVYDADEEEIIQYVVGTLQPKLKRFLHHPKTLLEQLIQRGQEVQNGLEPTD  
DPAGQRTQSEDNDESLTPAVTNESTASEGTLPE

서열번호 5

GQLDNVTNAGIHSFQGHARGVANKPNVILQIGKCRAEMLEHVRRTHRHLLSEVSKQVERELKGLQKQSVGKLENN  
LEDHVPTDNQRWKKSIKACLARCQETIAHLERWVKREMNWKEVFFRLEKWADRLESMGGKYCPGEHGKQT  
VSVGVGGPEIRPSEGEIYDIALDMSQMYALTPPPGEVPSIPQAHDSYQWVSVSEDAPASPVETQVFEDPREFLS  
HLEEYLKQVGGTEEYWLSQIQNHMNGPAKKWWEYKQDSVKNWVFEKKEFLQYSEGTLTRDAIKRELDLPQKE  
GEPLDQFLWRKRDLYQTLVYDADEEEIIQYVVGTLQPKLKRFLSYPLPKTLEQLIQRGKEVQGNMDHSEDESPQR  
TPEIQSGDSVESMPPSTTASPVPSTASPVPSNGTQPEPPSPATVI

서열번호 6

GQLDNVTNAGIHSFQGHARGVANKPNVILQIGKCRAEMLEHVRRTHRHLLSEVSKQVERELKGLQKQSVGKLENN  
LEDHVPTDNQRWKKSIKACLARCQETIAHLERWVKREMNWKEVFFRLEKWADRLESMGGKYCPGEHGKQT  
VSVGVGGPEIRPSEGEIYDIALDMSQMYALTPPPGEVPSIPQAHDSYQWVSVSEDAPASPVETQVFEDPREFLS  
HLEEYLKQVGGTEEYWLSQIQNHMNGPAKKWWEYKQDSVKNWVFEKKEFLQYSEGTLTRDAIKRELDLPQKE  
GEPLDQFLWRKRDLYQTLVYDADEEEIIQYVVGTLQPKLKRFLSYPLPKTLEQLIQRGKEVQGNMDHSEDESPQR  
TPEIQSGDSVDSVPPSTTASPVPSTASPVPSNGTQPE

서열번호 7

GQLDNVTNAGIHSFQGHARGVANKPNVILQIGKCRAEMLEHVRRTHRHLLSEVSKQVERELKGLQKQSVGKLENN  
LEDHVPTDNQRWKKSIKACLARCQETIAHLERWVKREMNWKEVFFRLEKWADRLESMGGKYCPGDHGKQT  
VSVGVGGPEIRPSEGEIYDIALDMSQMYALTPPPGEVPSIPQAHDSYQWVSTSEDAPASPVETQVFEDPREFLS  
HLEEYLKQVGGTEEYWLSQIQNHMNGPAKKWWEYKQDSVKNWVFEKKEFLQYSEGTLTRDAIKRELDLPQKE  
GEPLDQFLWRKRDLYQTLVYDADEEEIIQYVVGTLQPKLKRFLSYPLPKTLEQLIQRGKEVQGNMDHSEDESPQR  
TPEIQSGDSVDSVPPSTTASPVPSTASPVPSNGTQPE

서열번호 8

GSWGLQRHVADERRGLATPTYGAVCSIREKKASQLSGQSCLEKELLGWKCTEAVEMMQVDNFHGNLHSCQ  
GHRGMANHKPNVILQIGKCRAEMLDHVRRTHRHLLTEVSKQVERELKSLQKQSVGKLENNLEDHVPSAAENQR  
WKKSIKACLARCQETIAHLERWVKREINWKEVFFRLEKWADRLESGGGKYGPGDQSRQTVSVGVGAPEIQPR  
KEEYDYALDMSQMYALTPPPMPGEPNVPQSHDSYQWITISDDSPSPVETQIFEDPREFLTHLEDYLKQVGGT  
EEYWLSQIQNHMNGPAKKWWEYKQDSVKNWLEFKKEFLQYSEGTLTRDAIKQELDLPQKDGEPDQFLWRK  
RDLYQTLYDAEEEEVIQYVVGTLQPKLKRFLSHYPKTLLEQLIQRGKEVEGNLDNSEEPSPQRSPKHQLGGSVESL  
PPSSTASPVASDETHPDVSAPPVTVI

[0305]

서열번호 9

GDGETQAENPSTSLNNTDEDILEQLKIVMDQQHLYQKELKASFELSRKMFQMEQMNSKQTDLLLEHQKQ  
TVKHVDKRVEYLRAQFDASLGRWLKEQHADITTKIPEIITVKEDISLCLSTLCSIAEDIQTSRATTVTGHAAVQTH  
PVDLLGEHHLGTTGHPRLQSTRVQKDDVPESVSLFMQGEARSRIVGKSPIKQFPTFGKANDSSDPLQYLERC  
EDFLALNPLTDEELMATLRNVLHGTSRDWWVVARHKIQTWREFNKHFRAAFSEDEYELAERVRNRRIQKEDE  
SIRDFAYMYQSLCKRWNPACIEGDVVKLILKNINQLPSQLRSRVTVDLVLRLGQLEKDRQNQLQYELRKSSG  
KIIQKSSSCTSALPNTKSTPNQQNPATSNRPPQVYCWRCCKGHHAPASCPQWKADKHRAQPSRSSGPQTLTNL  
QAQDI

서열번호 10

GELDQRAAGGLRAYPAPRGGPVAKPSVILQIGKRAEMLEHVRRTHRHLLTEVSKQVERELKGLHRSVKGLEGN  
LDGYVPTGDSQRWKKSIKACLCRCQETIANLERWVKREMHVWREVFYRLERWADRLESMGGKYPVGTNPSR  
HTVSVGGVGGPEGYSHADTYDYTVSPYAITPPAAGELPGQEAQYPPWGLGEDGQPGQVDTQIFEDP  
REFLSHLEEYLQVGGSEYWLSQIQNHMNGPAKKWWEFKQGSVKNWVEFKKEFLQYSEGTLRDAIKRELDL  
PQKQGEPLDQFLWRKRDLQTLVDAEEEEIIQYVVGTLQPKLKRFLRPLPKTLEQLIQKMEVQDGLQAAEP  
ASPRLPPEESEALPALTSESVASDRTQPE

서열번호 11

GQLDNVTNAGHSFQGHGRGVANKPNVILQIGKRAEMLEHVRRTHRHLLSEVSKQVERELKGLQKSVGKLENN  
LEDHVPTDNQRWKKSIKACLCRCQETIAHLERWVKREMNWKEVFFRLEKWADRLESMGGKYCPGEHGKQT  
VSVGVGGPEIRPSEGEIYDIALDMSQMYALTTPGPEVPSIPQAHDSYQWVSSEDAPASPVETQIFEDPHEFLS  
HLEEYLKQVGGTEEYWLSQIQNHMNGPAKKWWEYKQDSVKNWVEFKKEFLQYSEGTLRDAIKRELDLPQKE  
GEPLDQFLWRKRDLQTLVDAEEEEIIQYVVGTLQPKLKRFLSYPLPKTLEQLIQRGKEVQGNMDHSEEPSQR  
TPEIQSGDSVESMPPSTTASPVPSNGTQPEPPSPATVI

서열번호 12

GQLENINQGSLSHAFQGHGRGVHNNKPNVILQIGKRAEMLEHVRRTHRHLLTEVSKQVERELKGLQKSVGKLE  
NNLEDHVPSAAENQRWKKSIKACLCRCQETIANLERWVKREMNWKEVFFRLEKWADRLESMGGKYCHADQ  
GRQTVSVGVGGPEVRPSEGEIYDIALDMSQMYALTTPPMGDVPVIPPQPHDSYQWVTDPEEAPPSPVETQIFE  
DPREFLTHLEDYKQVGGTEEYWLSQIQNHMNGPAKKWWEYKQDSVKNWVEFKKEFLQYSEGTLRDAIKQE  
LDLPQKEGEPLDQFLWRKRDLQTLVDAEEEEIIQYVVGTLQPKLKRFLSHYPKLEQLIQRGKEVEGNLDNSE  
EPPSPQRTPEHQLGDSVESLPPSTTASPAQSDKTQPEISLPPTVI

서열번호 13

GQLDSVTNAGVHTYQGHRSVANKPNVILQIGKCRTEMLEHVRRTHRHLLTEVSKQVERELKGLQKSVGKLENN  
LEDHVPTDNQRWKKSIKACLCRCQETIAHLERWVKREMNWKEVFFRLEKWADRLESMGGKYCPTDSARQT  
VSVGVGGPEIRPSEGEIYDIALDMSQMYALTTPSGELPSVPQPHDSYQWVTPEDAPASPVETQVFEDPREFLC  
HLEEYLKQVGGTEEYWLSQIQNHMNGPAKKWWEYKQDQTVKNWVEFKKEFLQYSEGTLRDAIKRELDLPQKD  
GEPLDQFLWRKRDLQTLVIDAEEEQIIQYVVGTLQPKLKRFLSYPLPKTLEQLIQKKEVQGSGLDHSEEPSQRA  
SEARTGDSVETLPPSTTSSPNTSSGTQPEAPSPATVI

서열번호 14

GQLDSVTNAGVHTYQGHGRGVANKPNVILQIGKCRTEMLEHVRRTHRHLLTEVSKQVERELKGLQKSVGKLENN  
LEDHVPTDNQRWKKSIKACLCRCQETIAHLERWVKREMNWKEVFFRLEKWADRLESMGGKYCPTDSARQT

[0306]

VSVGVGGEIRPSEGEIYDYALDMSQMYALTSPGELPSIQPHDSYQWVTSPEDAPASPVETQVFEDPREFLC  
HLEEYKQVGGTEEYWLSQIQNHMNGPAKKWWEYKQDTVKNWVEFKKEFLQYSEGTLTRDAIKRELDLPQKD  
GEPLDQFLWRKRDLYQTLYDADEEQIQYVVGTLPKPKRFLSYPLKTLQLELIQKGEVQGSLDHSEEPSQRA  
SEARTGDSVESLPPSTTTSPNASSGTQPEAPSPPATVI

서열번호 15

GQLENVNHGNLHSFQGHRRGGVANKPNVILQIGKCAEMLDHVRRTHRHLLTEVSKQVERELKGLQKSVGKLE  
NNLEDHVPASVENQRWKKSIKACLSRCQETIAHLERWVKREMNWVKEVFFRRLERWADRLES GGGKYCHGDN  
HRQTVSVGVGGPEVRPSEGEIYDYALDMSQMYALTSPGDPVVPVVSQPHDSYQWVTVPEDTPPSPVETQIFED  
PREFLTHLEDYKQVGGTEEYWLSQIQNHMNGPAKKWWEYKQDSVKNWLEFKKEFLQYSEGTLTRDAIKEELD  
LPQKDGPELDQFLWRKRDLYQTLYEADEEEVIQYVVGTLPKPKRFLSHYPKTLQLELIQKGEVEGNLDNSEE  
PTPQRTPEHQLCGSVESLPPSSTVSPVASDGTQPEATSPLPATVI

서열번호 16

GPLTLLQDWCRGEHLNTRCMLILGIPEDCGEDEFEEETLQEAACRHLGRYRVIGRMFRREENAQAALLELAQDIDY  
ALLPREIPGKGGPWEVIVKPRNSDGEFLNRLNRFLEEERRTVSDMNRVLGSDTNC SAPRVTISPEFWTWAQTLG  
AAVQPLLEQMLYRELRFVSGNTISIPGALAFDAWLEHTTEMLQMWQVPEGEKRRRLMECLRGPALQVVSGLR  
ASNASITVEECLAALQVFGPVESHKIAQVKLCKAQEQAGEKVSFVLRLEPLLQRAVENNVVSRNNVNQTRKLR  
VLSGATLPDKLRDKLMLKQRRKPPGFALVLLREEEWEATLGPDRSLEGLVAPRPPARITGVGAVPLPAS  
GNSFDARPSQYRRRRGQHRRGVVARAGSRGSRKRKRHTFCYSCGEDGHIRVQCINPSNLLAKETKEILEG  
GEREAQTNSR

서열번호 17

GALTLEDWCKGMDMDPRKALLIVGIPMECSEVEIQDQTVKAGLQPLCAYRVLGRMFRREDNAKAVFIELADTV  
NYTTLPSHIPGKGGSEWVVKPRNPDEFLSRLNYFLKDEGRSMTDVARALGCCSLPAESLDAEVMPPQVRSPL  
EPPKESMWWYRKLKVFSGTASPSPGEETEFEDWLEQVTEIMPIWQVSEVEKRRRLESRLGPAISIMRVLQANNS  
ITVEQLDALKQIFGDKEDFRASQFRFLQTSPIKIGEVSTFLLRLEPLLQKAVHKSPLSVRSTDMIRLKHLLARVAM  
TPALRGKLELLDQRCPPNFLEMLKIRDEEWEENTEAVMKNEKPSGRGRGASGRQARAEASVSAPQATVQA  
RSFSDSSPQTIQGLPPLVRRRLLGSESTRGEDHGQATYPKAENQTPGREGPQAAGEELGNEAGAGAMSHPK  
PWET

서열번호 18

GAVTMLQDWCRWGMVGNARRGILLGIPEDCDDAEFQESLEAALRPMGHFTVLGKAFREEDNATAALVELDRE  
VNYALVPREIPGTGGPWNVVFPVPRCSGEEFLGLGRVHFPEQEGQMVESVAGALGVGLRRVCWLRSIGQAVQ  
PWVEAVRCQSLGVFSGRDQPAPGEESFEVWLDHTTEMLHVWQGVSERERRRRLLEGLRGTAQLVHALLAEN  
PARTAQDCLAALAVFGDNESQATIRVKCLTAQQQSGERLSAFVLRLEVLQKAMEKEALARASADRVRRLRQM  
LTRAHLTEPLDEALRKLRMAGRSPSFLEMLGLVRESEAWASLARSVRAQTQEGAGARAGAAVARASTKVEA  
VPGGPGREPEGLLQAGGQEAEEELLQEGKPKVLEECDN

서열번호 19

GAVTMLQDWCRWGMVGNARRGILLGIPEDCDDAEFQESLEAALRPMGHFTVLGKVFREEDNATAALVELDRE  
VNYALVPREIPGTGGPWNVVFPVPRCSGEEFLGLGRVHFPEQEGQMVESVAGALGVGLRRVCWLRSIGQAVQ  
PWVEAVRYQSLGVFSGRDQPAPGEESFEVWLDHTTEMLHVWQGVSERERRRRLLEGLRGTAQLVHALLAEN  
PARTAQDCLAALAVFGDNESQATIRVKCLTAQQQSGERLSAFVLRLEVLQKAMEKEALARASADRVRRLRQM

[0307]

LTRAHLTEPLDEALRKLRMAGRSPSFLMGLVRESEAWEASLARSVRAQTQEGAGARAGAQAVARASTKVEA  
VPGGPGREPEGLRQAGGQEAELLQEGLKPVLEECDN

서열번호 20

GVEDLAASYIVLKLLENEIRQAQVQWLMEENAALQAQIPELQKSQAKEYDLLRKSSEAKEPQKLEPHMNPPAA  
WEAQKTPEFKEPKQPPPEQDLLPWEPPAAWELQEAPAPESLAPPATRESQKPPMAHEIPTVLEGQGPANTQ  
DATIAQEPKNSEPDPPNIEKQEAPEYQETAQAQLEFLELPPQEPEPSNAQEFLELSAAQESLEGLIVVETSAA  
EFPQAPIGLEATDFLQYTLTFSGDSQKLPFLVQLYSYMRVRGHLYPTEAALVSFVGNCFSGRAGWWFQLLLDI  
QSPLEQCESFIPVLQDTFDNPNENMKDANQCIHQLCQEGEHVATHFHLLIAQELNWDSTLWIQFQEGGLASSIQ  
DELSHTSPATNLSDLITQCISLEEKDPNPLGKSSSAEGDGPESPPAENQPMQAAINCPHISEAEWVRWHKGR  
CLYCGYPGHFARDPCVKPHQALQAGNIQACQ

서열번호 21

GVQPQTSKAESPALAASPNAMDDVIDTLTSLRLNSALRREASTLRAEKANLTNMLESVMAELTLRTRARIPG  
ALQITPPISSITNGTRPMTTPPTSLPEPFGDPGRLAGFLMQMDRFMIFQASRFPGEARVAFLVSRLTGEAEK  
WAIPHMQPDSPLRNNYQGLAELRRYKSPRLRHARRAQIRKTSASNRAVRERQMLCRQLASAGTGPCPVHPAS  
NGTSPAPALPARARNL

서열번호 22

GDGRVQLMKALLAGPLRPAARRWRNPIPFPEFDGTDRLPEFIVQTSSYMFVDENTFSNDALKVTLITRLTGP  
ALQWVPIYIRKESPLLNDYRGFLAEMKRIVFGWEEDEDF

서열번호 23

GEGRVQLMKALLARPLRPAARRWRNPIPFPEFDGTDRLPEFIVQTSSYMFVDENTFSNDALKVTLITRLTGP  
ALQWVPIYIKESPLSDYRGFLAEMKRIVFGWEEDEDF

서열번호 24

GPRGRCRQQGPRIWAAANYANAHPWQQMDKASPGVAYTPLVDPWIERPCCGDTVCVTRTMEQKSTASG  
TCGGKPAERGPLAGHMPSSRPHRVDFCWPVPGSDPGTDFGSPWLLDRFLAQLGDYMSFHFHYQDNISRVCEI  
LRRLTGRAQAWAAPYLDGDLPLPDDYELFCQDLKEVVQDPNSFAEYHAVVTCPPLASSQLPVAPQLPVVRQYL  
ARFLEGLALDMGTAPRSLPAAMATPAVSGSNSVRSALFEQQLKESTPGPKPEPPVLPSSSTCSSKPGPVEPASSQ  
PEEAAPTVPRLSESANPPAQRDPAPHPGGPKPKTEEEVLETEGDQEVSLGTPQEVVEAPETPGEPPLSPGF

서열번호 25

GVDELVLLLHALLMRHRALSIENTQLMEQLRLLVCERASLLRQVRPPSPVFPETFNCESSRLEPFIVQTASYML  
VNENRFCNDAMKVAFLISLLTGEAEVWVPIEMDSPILGDYRAFLDEMKGCFGWDDDEDDDEEEEDDY

서열번호 26

GPVDLGQALGLLPLAKAEDSQFSEDAALQEELSSPETARQLFRQFRYQVMGPHETLQKRLKCFWLQPEV  
HTKEQILEILMLEQFLTILPGEIQMWVRKQCPGSGEEAVTLVESLKGDPQRLWQWISIQVLGQDILSEKME  
SPSCQVGEVEPHLEVVPQELGLENSSSGPGELLSHIVKEESDTEAELALASQPARLEERLIRDQDLGASLL  
PAAPQEQWRQLDSTQKEQYWDLMEYTGKMSVSGAGISHPKSDLTNSIEFGEELAGIYLHVNEKIPRPTCIGDRQ  
ENDKENL

[0308]

NLENHRDQELLHASQASGEVPSQASLRGFFTEDEPGCFGEENLPEALQNIQDEGTGEQLSPQERISEKQLGQ  
HLPNPHSGEMSTMWLEEKRETSQKQPRAPMAQKLPTRCFCGKTFYRNSQLIFHQRTHTGETYFQCTICKKAF  
LRSSDFVKHQRTHTGEKPKCDYCGKGFSDFSGLRHHKEIHTGEKPYKCPICEKSFQRSNFRHQRVHTGEKPY  
KCSHCGKSFWSSSLDKHQRSHLGGKPFQ

서열번호 27

GTLRLLLEDWCRGMDMNPRAKALLIAGISQSCSVAEIEEALQAGLAPLGEYRLLGRMFRRDENRKVALVGLTAETS  
HALVPKEIPGKGGIWRVIFKPPDPDNFTLSRLNEFLAGEGNTVGELSRALGHENGLDPEQGMIPEMWAPMLA  
QALEALQPALQCLKYKLRVFSGRESPEPEEEEEFRWWMFHTTQMIKAWQVPDVEKRRRRLLESLRGPALDVIRVL  
KINNPLITVDECLQALEEVFVGTNDNPRELQVKYLTYYHKDEEKLSAYVLRLEPLLQKLVQRGAIERDAVNQARLDQ  
VIAGAVHKTRRELNLPEDGPAFGFLQVLVLIKDYEAEEEEALLQALEGNFT

서열번호 28

GTERRRDELSEIINNREKVMKQSEENNNLQSQVQKLTENTTLREQVEPTPEDEDDIELRGAAAAAPPPPIE  
EECPEDLPEKFDGNPDM LAPFMAQCQIFMEKSTRDFSVDVRVCFVTSMMTGRAARWASAKLERSHYLMHN  
YPAFMMEMKHFEDPQRREVAKRIRRLRQGMGSVIDYSNAFQMI AQDLWNEPALIDQYHEGLSDHIQEEL  
SHLEVAKLSALIGQCIHERRLAAAAARKPRSPRALVLP HIASHHQVDPTPEVGGARMRLTQEEKERRRKLNL  
CLYCGTGGHYADNCPAKASKSSPAGKLP GPAVEGPSATGPEIIRSPQDDASSPHLQVMLQIHLPRHTLFRAM  
IDSGAGNFIDHEYVAQNGIPLRIKDWPIVLEAIDGRPIASGPVVHETHDLIVDLGDHREVLSDFTQSPFFPVV  
GVRWLSTHDPNITWSTRSIVFDSEYCRYHCRMYSPIPPSLPPPAPQPLYYPVDGYRVYQPVRYVYVQNVYTPV  
DEHYVPDHRVLDPHIEMIPGAHSIPSGHVYSLEPEMAALRDFVARNVKDLITPTIAPNGAQLVQVKGWKL  
QVSYDCRAPNFTIQNQYPRLSIPNLEDAQHLATYTFVQPQIPGYQTYPTYAAYPTYPVGFAWYVPVGRDQGRS  
LYVPVMITWNPWHYRQPPVPQYPPQPPPPPPPPPPSYSTL

서열번호 29

GGGGAGCTGGACCACGGACCAGCGGGGGCTCCACGCCTACCCGGGCGCGGGGGCGGCGAGGTGGCC  
AAGCCCAACGTGATCCTGCAGATCGGGAAGTGCCGGGCGGAGATGCTGGAGCACGTGCGGGCGGACGCAC  
CGGCACCTGCTGGCCGAGGTGTCCAAGCAGGTGGAGCGCGAGCTGAAGGGGCTGCACCGGTGCGTGGG  
AAGCTGGAGAGCAACCTGGACGGCTACGTGCCACGAGCGACTCGCAGCGCTGGAAGAAGTCCATCAAG  
GCCTGCCTGTGCCGTGCCAGGAGACCATCGCCAACCTGGAGCGCTGGGTCAAGCGCGAGATGCACGTGT  
GGCGCGAGGTGTTCTACCGCTGGAGCGCTGGGCCGACCGCTGGAGTCCACGGGCGGCAAGTACCCGGT  
GGGACGCGAGTACGCCGCCACACCGTTTCCGTGGGCGTGGGGGGTCCCGAGAGCTACTGCCACGAGGC  
AGACGGCTACGACTACACCGTACGCCCCACGCCATCACCCCGCCCCAGCCGCTGGCGAGCTGCCCGGGC  
AGGAGCCCGCGAGGCCAGCAGTACCAGCCGTGGTCCCCGGCGAGGACGGGAGCCAGCCAGCCCGGCG  
TGGACACGCAGATCTTCGAGGACCCTCGAGAGTTCCTGAGCCACCTAGAGGAGTACTTGCGGACAGGTGGG  
CGGCTCTGAGGAGTACTGGCTGTCCAGATCCAGAATCACATGAACGGGCGGCAAGAAGTGGTGGGA  
GTTCAAGCAGGGCTCCGTGAAGAACTGGGTGGAGTCAAGAAGGAGTTCCTGCAGTACAGCGAGGGCAC  
GCTGTCCCGAGAGGCCATCCAGCGCGAGCTGGACTGCCGAGAAGCAGGGCGAGCCGCTGGACCAATTC  
CTGTGGCGCAAGCGGGACCTGTACCAGACGCTCTACGTGGACGCGGACGAGGAGGAGATCATCCAGTAC  
GTGGTGGGACCCCTGCAGCCCAAGCTCAAGCGTTTCTGCGCCACCCCTGCCAAGACCCCTGGAGCAGCT  
CATCCAGAGGGGATGGAGGTGAGGATGACCTGGAGCAGCGGCGGAGCCGCGCCGCCCCACCTCC  
GGTGGAGGATGAGGCGGAGACCCTACGCCCCCCCCAACAGCGAGTCCGTGGCCAGTGACCGGACCCA  
GCCCAG

[0309]

서열번호 30

GGGGAATTGGATCAACGTAACCTACCGGTGGCCCTTACGCATACCTGCACCACGCGGGGCCCTGTGCGGA
AGCCAAATGTCATCCTGCAGATTGGGAAGTGCCGGGCTGAGATGCTGGAGCACGTCCGTCGGACGCATCG
TCATCTTCTACTGAGGTGTCAAACAGGTGGAGCGTGAACCTCAAAGCTTGCACCGCAGCGTTGGGAAAC
TTGAAAGCAACTTAGATGGCTATGTGCCGACTGGCGACAGCCAGCGTTGGCGTAAGTCCATCAAAGCATGT
TTGTGTCGTTGCCAGGAAACGATTGCAAACCTGGAGCGTTGGGTCAAACGGGAGATGCATGCTGGCGTG
AAGTATTTTATCGTTTAGAGCGTTGGGCCGATCGTTAGAGAGCATGGGTGGTAAGTACCCTGTGGGGAGC
AACCTTCTCGGCATACGACGTGAGTCGGTGTGGCGGGCCGGAGTCTACGGTCATGAAGCGGACACCTA
CGACTATACCGTAAGCCCTTATGCTATTACCCACCACCTGCGGCCGGCAATTACCTGGCCAGGAAGCCG
TTGAGGCTCAACAATACCCTCCTTGGGGGCTGGGCGAGGATGGTCAACCTAGCCAGGGGTAGACACGCA
AATCTTTGAGGACCCACGGGAGTTTCTTCCACCTGGAAGAACTGCGTCAGGTTGGTGGGAGCGAAG
AATACTGGCTGTCAAAATCAAACCATATGAATGGTCTGCAAAAAATGGTGGGAATATAAACAGGGT
TCCGTGAAAAACTGGGTTGAGTTAAAAAGGAGTTTCTCAATATTCCGAGGGCGCCCTCAGTCGGGAGGC
GGTCCAACGCGAGTTGGACTTGCCACAGAAACAGGGGAAACCACTCGATCAATTCCTTTGGCGGAAACGT
GACCTTTACCAGACATTGACGTGGATGCAGATGAGGAAGAAATTATCCAATATGTTGTGGGGACCCTGCA
GCCGAAACTGAAACGTTTCTTCCGCCCGCGCTGCCTAAAACGTTGGAACAACCTATTAGAAAGGTATGG
AGGTCGAGGATGGCTTAGAACAAAGTCGCAGAGCCGCCCTCGCCACACTTGCCTACAGAGGAGGAATCGGA
GGCGTGACCCAGCACTTACATCAGAGTCAGTGGCATCAGACCGGACACAACCCAGAG

서열번호 31

GGGGAGTTAGATCACCGTACAACGGGGGGGTTGCACGCATACCTGCTCCAGTGCCGGGCCGGCAGCTA
AGCCAAACGTAATCCTGCAGATTGGGAAGTGCCGGGCGAGAGATGTTGGAGCACGTCCGGCGGACCCACCG
GCACCTCCTGGCTGAAGTGTCTAAACAAGTAGAACGGGAACTCAAAGGTCTTCATCGTAGCGTCGGGAAAT
TGGAATCGAATTTGGACGGGTATGTTCTACAGGCGACTCACAGCGGTGAAAAAGAGCATCAAGGCCTG
CCTGAGTCGCTGCCAGGAGACGATTGTAACTCGAACGCTGGGTTAAGCGGGAGATGCACGTTTGGCGC
GAAGTCTTCTACCGGCTGGAGCGTTGGGCTGATCGGCTCGAATCTGGTGGGGGTAAGTATCCAGTTGGGT
CCGACCTGCTGCCACACAGTCTCAGTTGGCGTAGTGGGCCGGAGTCGTATTGCCAAGATGCGGACAA
CTATGATTATACAGTTTCCCATACGCGATCACACCACCGCCGGCAGCAGGGCAGCTGCCAGGTGAGGAAG
AGGTTGAGGCCAGCAGTATCCACCATGGGCCCCAGGGGAAGACGGCCAGCTTCTCCTGGGGTGGACAC
TCAAGTTTTGAAGATCCGCGTGAATTTCTGCGGCATTTAGAAGATTATCTCCGCCAGGTCCGGGGGCTG
AAGAGTATTGGTTAAGCCAAATCAAACCATATGAACGGCCCGCCAAGAAAGTGGTGGGAGTACAAGCA
AGGGTCTGTGAAAAATTGGGTGGAGTTAAGAAAGAATTCTTGAATATTCTGAGGGCACTCTTTCGCGTG
AAGCCATCCAACGCGAACTCGACTTACCGCAGAAACAAGGGGAACCTCTCGACCAATTTCTGTGGCGCAAA
CGCGACCTGTACCAGACTCTTACGTGATGCTGAGGAGGAAGAAATTATCAATACGTAGTTGGCACACT
GCAGCCTAAGCTTAAACGGTTTTTACGTCCACCATTGCCGAAGACGCTTGAACAACCTCATCCAGAAGGGTA
TGGAGTTCAAGATGGTCTGGAACAGGCAGCGGAACAGCGCGGAGGAGGCAGAAGCCCTGACACCTG
CGTTAACTAACGAGTCTGTCCGAGCGACCCGACCCAGCCGGAA

서열번호 32

GGGGAATTAGACCCCTGAACCAAGCTCAGGCCTGCATCCATCCTCTGTTTTGCATCCATACCCAGGTCTC
CGGGGCGGGCAACCCGGAAGCCTAATGTATTTGCAAATGGCAAATGCCGTGCGGAAATGCTTGAAC
ACGTCCGCAAAACTCACCGTCATCTCCTCACAGAAATATCGCGCAAGTAGAACGCGAGCTCAAAGGCCTT
CACAAAAGTGTGGCAAGTTGGAATCAAATCTTGATGGGTACGTACCGTCAAGCGACTCCCAACGCTGGAA
GAAAAGCATTAAAGCGTGCTTATCCCGTTGCCAAGAGACGATTGCGCATTTAGAACGCTGGGTTAAACGTG
AAATGAATGATGGCGTGAAGTGTCTACCGTTTGGAACTGGGGCGACCGTCTGGAGGCTATGGGCGG
TAAGTATCCTGCCGGTGAGCAGGCCCGGCTACAGTTTTCAGTGGGCGTTGGGGGCCCTGAGACATGTTGT

[0310]

CCAGGGGATGAAAGTTATGATTGTCCGATTCTCCGTATGCAGTTCACCTTCCACCGGGGAGTCTCCGGAA  
TCCTTAGACCAAGGGGATCAGCACTATCAGCAGTGGTTTGCCTCCCGGAGGAGTCCCCTGTTAGCCCTGG  
GGTTGATACCCAGATCTTTGAAGATCCTCGCGAGTTTTTACGTCATCTGGAGAAGTACCTGAAACAAGTCG  
GCGGGACAGAGGAAGACTGGCTTCTCAAATCCAGAATCACATGAATGGGCCGGCGAAGAAGTGGTGGG  
AGTACAAGCAAGGGAGTGTAAAGAATTGGCTTGAATTTAAGAAGGAATTTTTACAGTATTCGGAGGGCAC  
ACTGACGCGGGACGCGTTGAAACGTGAACTGGATCTCCACAGAAAACAAGGCCAACCCTTGATCAATTTT  
TATGGCGGAAGCGCGACTTATATCAGACACTTACGTTGACGCCGATGAAGAGGAAATCATTAGTACGTC  
GTGGGCACTTTCAGCCGAAATTAACGCTTCTCCATCACCCACTCCCTAAGACGCTTGAGCAGCTTATC  
CAACGGGGCCAAGAAGTTCAGAATGGTCTGGAGCCTACCGACGATCTGCAGGCCAACGCACTCAATCGG  
AGGACAACGACGAAAGCCTTACCCCTGCCGTACCAATGAGAGTACTGCAAGCGAGGGCACCTGCCAGA  
G

서열번호 33

GGGCAGCTTGATAACGTTACAAACGCGGGCATCCACTCCTTCCAGGGGCATCGTGGCGTAGCGAATAAGC  
CAAATGTCTTCTGCAAATTGGTAAATGTCTGTGCGGAAATGCTGGAGCAGTTCGCGCACCCACCGCCAT  
TTATTATCTGAAGTATCTAAGCAGGTAGAACGTGAGCTGAAAGGGCTGAAAAGTCCGTGGGCAAGCTCG  
AGAATAACTTGGAGGATCATGTCCCTACAGATAACCAACGCTGGAAGAAGTCCATTAAGCGTCTTGGCT  
CGTTGTCAAGAGACTATCGCGCATTTAGAGCGTTGGGTGAAACGCGAAATGAACGCTCTGGAAGGAGGTGT  
TTTTCCGGCTGGAAAAGTGGGCAGACCGGCTGGAGTCAATGGGTGGCAAGTACTGCCCGGGCGAACACG  
GGAAACAACCCGTCAGTGTAGGCGTGGGGGTCTGAAATCCGGCCTTCGGAGGGGGAAATTTATGATTA  
TGCTCTGGATATGAGCCAGATGTATGCACTCACCCACCTCCAGGCGAAATGCCATCAATCCCAAGCCCA  
TGACAGCTATCAGTGGGTTAGTGTCTCAGAAGATGCCCCGGGAGCCCTGTCGAAACCCAGGTATTTGAGG  
ACCTCGGGAAATTCCTGTCTCACCTGGAGGAATACCTGAAGCAGGTAGGCGGCACGAGGAGTATTGGTT  
GTCCAGATCCAGAATCACATGAATGGTCCGGCAAAAAATGGTGGGAATATAAACAGGACTCCGTAAA  
AACTGGGTTGAGTTAAAAAGGAATTTCTGCAATACTCTGAAGTACTTTAACTCGGATGCTATTAAGCGT  
GAACTCGACTTCCCGCAAAAGGAAGGTGAACCTCTTGACCAATTCCTTTGGCGGAAGCGGGACCTCTATCA  
GACACTTTACGTGGACGCGGATGAGGAGGAGATCATTAGTATGTGGTCCGTCACCTGCAGCCGAAGCTC  
AAGCGTTTCTGAGCTATCTCTCCAAAGACTTTAGAACAGCTCATCCAGCGCGTAAAGAAGTGCAGGG  
TAACATGGATCACTCCGATGAGCCTTCGCCGAGCGTACACCTGAAATTCATCAGGTGACTCCGTAGAAT  
CTATGCCACCTTCAACAACGGCATCTCCGGTTCATCTAATGGTACCCAACCTGAGCCGCGGAGCCGCCAG  
CCACCGTTATC

서열번호 34

GGGCAACTTGACAACGTAACAAACGCTGGGATTCACTCCTTTCAGGGCCACCGCGGTGTCGCCAACAAAGCC  
AAACGTAATCTTGCAAATGGCAAATGCCGTGCGGAGATGTTGGAACACGTTCTGCTACACATCGTCACT  
TGCTGTGCGGAAGTCTCTAAACAAGTAGAACGTGAACTTAAAGGGCTTCAAAGTCAAGTCGGCAAATTGGAA  
AACAACTTGAAGACCATGTACCAACCGACAATCAGCGTTGGAAAAAGTCTATCAAAGCTTGCTGGCCCG  
TTGTCAAGAGACGATTGCTCACCTGGAGCGGTGGGTAAAGCGGAGATGAATGTGTGGAAAGAGGTCTTC  
TTCCGCTTGGAAAAATGGGCCGACCGTTTTGGAGTCCATGGGCGGTAATATTGTCCGGGTGAACATGGTA  
AGCAAACAGTCTCTGTGGCGTGGTGGGCCGGAGATTCCGCCTTCTGAAGGCGAGATTTACGATTATGC  
GCTCGACATGTCCAGATGTATGCGCTTACACCACCACCGGGCGAGGTACCAAGCATTCTCAAGCGCATG  
ACAGTTATCAGTGGGTTAGCGTATCCGAAGACGCTCCTGCCTCGCGGTAGAGACCCAGTTTTTGAAGAT  
CCTCGTGAATTTTAAAGCCACTTGGAGGAGTATTTGAAGCAGGTAGGGGGGACAGAGGAATATTGGCTGT  
CTCAGATCCAGAACCACATGAATGGCCCCGCTAAAAAGTGGTGGGAATACAACAAGATTCGGTAAAGAA  
TTGGGTAGAATTTAAAAAGGATTTTTACAGTACTCAGAGGGGACTCTACGCGTGATGCGATCAAACGCG  
AGTTGGATCTTCTCAAAGAGGGGGAGCCACTCGATCAGTTCCTCTGGCGCAAGCGGGATCTTACCAA

[0311]

ACACTCTACGTAGACGCAGACGAAGAAGAGATCATCCAGTACGTGGTGGGTACGCTCCAGCCGAAACTCA  
AACGTTTCTCAGCTACCCACTTCTAAGACTCTGGAACAACCTGATTCAGCGGGGCAAAGAGGTCCAGGGT  
AACATGGACCATTAGAGGAACCTAGTCCGCAACGTACACCTGAGATCCAATCTGGGGATTCTGTGATT  
GGTCCACCTTCTACAACAGCGTCTCCGTGCCGTCAAATGGGACCCAACACAGAG

서열번호 35

GGGCAGCTTGATAATGTAACCAATGCAGGTATCCACTTTCCAGGGTCACCGCGGTGTGGCAAACAAGCC  
AAATGTTATTCTGCAAATTTGGTAAGTGTGCGCGTGAAGTGTAGAACACGTCCGGCGCACGCATCGGCATC  
TCCTGTGAGAGTTTCAAAGCAGGTAGAGCTGAATTAAGGGCTCCAGAAGTCCGTAGGTAAACTCGA  
AAATAATCTTGAAGACCGTTCCTACCGATAATCAACGGTGGAAAAAGTCAATCAAGCGGTGTTAGCAC  
GGTGTGAGGAAACGATCGCGCACCTCGAACGTTGGGTGAAGCGGAAATGAATGTCTGAAAGAAGTGT  
CTCCGGCTTGAGAAAGTGGGCTGATCGGCTCGAATCCATGGGTGGCAAATATTGTCCAGGTGATCATGGCA  
AGCAAACGGTCTCCGTGGTGTGGTGGTCCGGAATCCGGCCGAGCGAGGGTGAATCTATGACTACGC  
TCTTGATATGTCCAGATGTATGCACTCACTCCTCCCGGGTGAAGTCCCGTGCATCCCGCAGGCGCATGA  
CTCATACCAATGGGTGTGACTAGCGAAGACGCACCGCCTCCCTGTTGAAACTCAAGTATTGAGGAC  
CGCGTGAGTTCCTGAGCCATTTAGAGGAGTACCTTAAGCAGGTTGGTGGTACCGAGGAATACTGGTTGAG  
CCAGATTGAGAAATCATGAAACGGGCGGCTAAGAAATGGTGGGAATACAAGCAGGATTCAGTCAAGAAAT  
TGGGTGCAATTAAGAAGGAGTTTTTGCAGTACAGTGAAGGGACGCTCACACGCGACGCTATCAAACGGG  
AGCTGGACCTGCCACAAAAGGAGGGTGAACCGCTTATCAGTTTCTTTGGCGCAAGCGTGTATCTGATCAA  
ACCTGTATGTGGACGCTGACGAAGAAGAGATCATTCACTACGTGGTGGGACTCTGCAACCAAAGCTGA  
AGCGTTTTCTTTCTTCTCCTCCTAAGACACTGGAACAGTTAATCCAACGTGGCAAGGAGGTCCAGGGTA  
ATATGGACCACTCTGAGGAACCGAGCCGCAACGTAATCTGAAATTCAGAGCGGGGATAGTGTGACTC  
AGTTCCTCAAGTACGACCGCATCCCGGTCCCAAGTAACGGTACCCAACACAGAG

서열번호 36

GGGTCTTGGGGCTTGAACGTCACGTGGCTGATGAACGTCGTGGCCTCGCTACGCCTACCTACGGCGGGT  
TTGTTCCATTGGGAGAAAAAGCCTCCCAACTGAGCGGCCAGAGCTGTTGGAGAAAGAGTTGCTTGGT  
GGAAATGTACGGAGGCAATCGTGGAAATGATGCAAGTCGATAACTTTAACACCGTAACCTACATAGCTGC  
CAAGGCCATCGGGGATGGCAAATCACAACCGAACGTAATCCTTCAAATCGGGAATGTGCGCGAGAAA  
TGTTAGACCAGTGGTGGCACCACCGCATCTTTGACGGAGTTTGAAGCAGGTAGAACCGGAATTG  
AAGTCTCTCAAAGTGGTGGCAAGCTCGAGAATAATCTGGAAGACCGTGCATCGGCGAGCGGAGA  
ACCAACGTTGGAAGAAATCAATTAAGCCTGCCTGGCCCGGTGCAAGAAACAATTGCTCACTCGAACGC  
TGGGTTAAACGCGAAATCAACGTCGAAAGAAGTATTTCTGCTGAGAGAGTGGGCGGACCGCCTTG  
AGTCGGGTGGGGCAAGTATGGGCCTGGTACCAAGTCTGCAACTGTAAGTGTGGTGTGGGGCCCC  
AGAAATCCAACCGGAAAGAAAGAAATCTATGACTACGCTCTGACATGTGCGAGATGTATGCCTTAACAC  
CACCGCCGATGGGTGAAGACCCAACCTACCTCAATCCACGATAGCTACCAAGTGGATTACCATCTCAGAC  
GATTCACCTCCGTCGCCAGTGGAAACTCAATTTTCGAGGATCCACGCAATTCCTTACCCATCTCGAGGAT  
TATCTTAAGCAAGTGGGCGGACTGAAGAATTTGGTTGAGTCAGATTCAAAATCATATGAACGGTCCGGC  
CAAGAAATGGTGGGAGTACAAACAAGATTCGTTGAAAACTGGTTGGAATTCAGAAGGAATTCCTTCAA  
TACTCTGAGGGTACTTTGACAGTGAACGCAATTAACAAGAAGTACTTACCGCAGAAGGACGGCGAGCC  
ATTGGATCAATTTCTTTGGCGGAAGCGGACCTGTATCAGACGCTCTATATTGATGAGAGGAGGAAGAA  
GTAATCCAATACGTTGTTGGCACACTCCAACCGAAATTAACGTTTCTTTCCACCCGTATCCGAAAATT  
TGGAACAGTTAATCCAACGTGGGAAAGAGGTTGGAAGCAACCTCGATAACTCTGAGGAGCCTAGCCCGCA  
ACGGAGTCAAAGCACCAATTTGGTGGTACGTCGAGAGCCTCCACCTTCGTCGACCGCAAGTCTGTTG  
CGTCAGACGAGACTACCCAGAGCTGAGCGCACCTCCGGTAACGGTGATT

[0312]

서열번호 37

GGGGACGGCGAGACTCAAGCTGAGAAATCCATCTACCAGCTTGAACAACACTGACGAAGATATCTTGAAC
AGCTCAAGAAAATTGTCATGGATCAACAACACCTGTATCAGAAAGAATTAAGGCATCTTTTGAACAAC
AGTCGCAAAATGTTTTCCAGATGGAACAAATGAATAGCAAGCAAACGGATCTGCTTTTAGAACATCAAAA
ACAGACTGTCAAACATGTAGACAAGCGCGTGGAGTATTTGCGGGCGCAATTCGATGCATCGTTAGGCTGG
CGGTTGAAAGAGCAACACGCGGATATTACGACCAAATCATTCTGAGATCATCAAACGGTGAAGGAAG
ATATTAGCCTGTGCTTTCTACGCTCTGCAATATCGCTGAAGATATCCAGACATCACGGGCTACCACTGTCA
CAGGGCATGCTGCCGTACAACCCATCCTGTGGATCTTTGGGTGAACACCAATTTAGGGACACGGGGCAC
CCACGTTACAGTCGACCCGTGTAGGGAAACCAGACGACGTACCTGAGTCGCCGGTAAGCCTGTTTATGCA
AGGTGAGGCGGTTCCCGGATCGTTGGCAAGAGTCCGATTAACCTGCAATTTCCGACGTTCCGGCAAAGCA
AACGATTCTCCGACCCACTCCAATATCTGGAGCGGTGTGAGGACTTTCTTGTCTTAAACCTTTAACTGATG
AGGAACTTATGGCTACTTTGCGGAATGTGTACATGGCACCTCTCGGGATTGGTGGGATGTCGCACGTCAT
AAAATCCAAACTTGGCGTGAGTTAATAAACACTTCCGGGGCGGCTTCTCAGCGAGGATTATGAAGATGA
GTTGGCTGAGCGCGTCCGTAACCGCATCAAAAAGAAGATGAGTCTATCCGCGATTTGCTTATATGTATC
AGTCTTGTGCAAGCGGTGGAACCTGCTATCTGCGAAGGTGATGTAGTAAAGTCATCCTGAAGAACATC
AATCCACAACCTGCCGTCTCAGTTACGCTCCCGGTACGACCGTGGATGAGCTTGTGCTTGGGCCAGCA
GCTTGAAAAGATCGTCAGAATCAGCTCCAATATGAGCTTCGGAAGAGTTCGGGCAAAATTATCAAAAAT
CTAGTTGTCGCGAAACTTACGCGTCCCGAACGAAAGAGTACACCTAATCAACAAAACCTGCTACCAAGT
AACCCTCTCCACAGGTGATTGCTGGCGGTGTAAGGGTCAACATGCCCTGCCTTGTCCGCAATGGAA
AGCTGATAAGCACCGTGCACACCTTCGCGGAGTTCTGGGCCAACAACCTGACTAATCTCCAAGCTCAAG
ACATC

서열번호 38

GGGGAATTGGATCAACGTGCGGCAGGGGGCTTGC GCGCGTACCCGGCGCCGCTGGTGGTCCAGTTGCC
AAACCGAGCGTAATTTTCCAGATTGGTAAGTCCGCGCTGAGATGCTGGAACACGTCGCCCGCACGCATCG
CCATCTTCTGACGGAGGTAAGTAAACAAGTGGAGCGCGAACTCAAGGGGTTACATCGGTCTGTCGGTAAG
TTGGAGGGCAATTTAGACGGCTATGTGCTACCGGTGATTCCCAACGCTGGAAAAAAGTATCAAGGCGT
GTCTCTGCCGGTGTCAAGAAACAATGCAAATCTCGAGCGTTGGGTGAAACGTGAGATGCATGTTGGCGT
GAGGTATTCTATCGTTTGAACCGTGGGCAGACCGTTTGGAGTCTATGGGGGGCAAGTATCCGGTGGGCA
CTAACCCGTCGCGGCACACAGTAAGTGTGCGGGTAGGGGGCCCGAAGGCTATTCTCATGAAGCGGATAC
TTATGACTACCGGTGTCTCCGTATGCTATCACGCCACCGCTGCCGCGGTGAGTTGCCTGGTCAAGAGG
CTGTCGAGGCACAACAGTACCCTCCATGGGGTCTGGGGGAGGACGGGCAACAGGTCCGGGCGTGGACA
CGCAGATTTTTGAGGACCTCGCGAATTTTTGAGCCACTTAGAGGAGTACCTGCGGCAAGTAGGGGGGAG
TGAAGAGTACTGGTTATCGCAAATTCAAAATCATATGAATGGCCCTGCGAAGAAATGGTGGGAGTTCAAAC
AGGGGTCAGTCAAGAAATGGGTGAGTTAAGAAAGAATTTTTGCAATACAGTGAGGGTACGTTGAGTCCG
CGAGGCCATCCAACGTGAACTGGACCTCCCTCAGAAGCAGGGGGAGCCGTTAGATCAATTTTTATGGCGG
AAACGTGACTTATACCAAACCTCTACGTTGACGCTGAGGAAGAAGAAATTTCAATATGTTGTCGGTAC
GCTGCAGCCAAAAGCTGAAGCGGTTCTCCGCTCCTCACTCCCTAAAACCTTAGAACAAATTAATCCAAAAGG
CATGGAAGTTACAGGACGGTTAGAACAAGCGGCCAACCAGGCTCTCCGCTGTCGCCCGGAAGAGGA
GAGTGAGGCTCTACGCTGCGCTACGAGCGAATCAGTAGCCTCCGATCGGACACAGCCAGAG

서열번호 39

GGGCAGCTTGACAAATGTGACGAACGCGGGGATTCACAGCTTTCAAGGGCACCGCGCGTCCCAACAAAC
CGAATGTACTTCTGCAAATCGGTAATGTCGTGCTGAAATGCTTGAACAGTTCGTCGTACCCATCGTCACT
TGCTTTCTGAAGTATCAAAAAGTGGAGCGGAACTCAAAGGCTGCAAAAGTCAAGTGGGTAAATTTGGA
GAATAACCTCGAAGACCATGTACTACAGACAACCAGCGGTGGAAAAAATCTATCAAGGCATGCCTCGCTC

[0313]

GTTGCCAGGAGACTATTGCCCATCTTGAGCGGTGGGTGAAACGTGAAATGAACGTATGGAAGGAAGTATT  
TTTTCGCTTAGAGAAGTGGGCTGATCGTCTTGAATCGATGGGCGCAAGTACTGCTGGGGAACACGGC  
AAACAACTGTATCTGTCGGCGTGGGGGGCCCCGAGATCCGGCCATCGGAAGGGGAAATTTATGATTATG  
CTCTCGACATGTCCCAAATGTATGCTCTCACACCAGGGCCAGGGGAAAGTACCGTCAATTCGCAAGCACAC  
GACAGCTACCAATGGGTATCTGTGAGCGAGGACGCGCTGCCTCCTCGGTTGAGACGCAAATCTTTGAGG  
ACCCACATGAATTTTTGTCTCATCTTGAAGAATATCTCAACAGGTTGGCGGCACAGAAGAATACTGGTTAT  
CTCAGATCCAGAATCACATGAACGGCCCGGCTAAAAAGTGGTGGGAGTATAAGCAAGATTCGTAAGAA  
CTGGGTGCAATCAAGAAAGAGTTTCTTCAATACTCTGAGGGTACTCTGACGCGCGATGCAATTAAGCGGG  
AGTTAGACCTTCCACAAAAGAGGGGGAGCCTCTTGACCAGTTCTGTGGCGTAAGCGCGACCTCTATCAG  
ACACTTACGTGACGCTGATGAAGAAGAGATTATTCAATATGTTGTGGGTACCTGCAGCCAAAGCTTAA  
GCGTTTCTTAGCTACCCACTCCGAAAACCTGAGCAGCTCATTCAACGCGGTAAGGAAGTGCAGGGCA  
ACATGGACCACTCTGAAGAGCCTAGCCCGCAGCGCACTCTGAAATCCAATCAGGTGACAGTGTGGAGTCA  
ATGCCGCGCTCAACCACCGCTTCTCCGGTACCTAGCAACGGGACGCAACCAGAGCTCCAAGCCACCGGC  
TACAGTCATC

서열번호 40

GGGCAACTTGAGAAATTAACCAAGGTTCCCTGCACGCGTTTCAGGGTCATCGCGCGTGGTCCATAACAA  
CAAGCCTAACGTTATTCTCCAGATCGGGAAAGTGCCGCGCCGAAATGCTGAGCATGTGCGGCGCACCCATC  
GCCATTTGCTCACTGAAGTATCAAAACAGGTGGAGCGTGAGTTGAAGGGGTTGCAGAAAAGTGTAGGCAA  
ACTTGAAAATAATTTAGAAGACCACGTACCAAGTGCGGCTGAGAACCAACGCTGGAAGAAGTCGATTAAA  
GCCTGCTTAGCGCTTGTGAGGAGACCATTCGAACTTGGAAACGCTGGGTTAAACGTGAGATGAATGTTT  
GAAGGAGGTCTTTTTCCGCTTAGAGCGCTGGGCAGATCGCCTCGAATCCGGGGTGGCAAGTACTGCCAT  
GCAGACCAGGGTCGCCAACTGTACGCGTAGGTGTTGGTGGTCTGAAGTGCCTCCGCTGAAAGGTGAAA  
TTTACGATTACGCGTTGGATATGAGCCAAATGTACGCTTACTCCGCCCTATGGGTGATGTTCCAGTAA  
TTCTCAGCCGATGACAGTTATCAGTGGGTGACAGATCCGGAAGAAGCGCCACCAAGTCCGGTTGAGAC  
ACAAATTTTCGAGGACCCTCGGGAGTTTCTGACCATCTTGAGGATTTTAAAAACAAGTCGGCGGGACAG  
AGGAATATTGGCTCTCAGATCCAAAATCATATGAATGGGCCAGGAAAAGTGGTGGGAATATAACA  
GGATAGTGTGAAGAACTGGCTTGAGTTCAAAAAGAATTCTTGCACTCAGAAGGCACGTTAACGCGG  
GACGCTATTAACAGGAACCTGACCTTCCACAAAAGAAGGGGAACCGCTGGATCAATTCCTGCGCGCAA  
ACGGATTTGTACCAAACCTCTACGTCGAGGCAGAAGAAGAGGAGTCAATCAATATGAGTTGGCACAC  
TGCAACCAAAACTGAAGCGGTTTCTTCTCATCCGTACCCTAAAACCTGGAGCAACTCATCCAGCGGGGA  
AGGAAGTTGAGGGGAATTTGGACAATAGTGAAGAACCCTCCACAGCGGACCCGAGAACATCAGCTGGG  
GGACAGTGTGGAATCTTTCGCGCTAGTACTACGGCTTCGCTGCCGTTCCGGATAAAAACGCAACCTGAGA  
TTAGCTTACCTCAACTACAGTCATT

서열번호 41

GGGCAATTAGATTCCGTTAACCAATGCGGGCGTCCACACCTACCAGGGCCATCGGAGCGTCGCCAATAAAC  
CTAACGTCATTTCAAATCGGGAAATGTCGGACTGAGATGCTGGAGCATGTCCGTGGACTCATCGCCAC  
CTGCTCAGAGAAGTGTCAAAGCAAGTGAACGTGAACCTCAAGGGCTTACAGAAGAGCGTGGGCAAATGG  
AAAACAATCTTGAAGACCATGTCCCAACTGACAATCAGCGGTGGAAGAAGTAATCAAGGCATGTCTCGCG  
CGTTGCCAAGAGACCATTGCTCACCTTGAGCGGTGGGTGAAACGTGAAATGAACGTGTGGAAGGAGGTGT  
TCTTCGGTTAGAACGCTGGGCCGACCGCTTGAATCAATGGGTGGTAAATCTGCCGACGGACTCTGCA  
CGTCAGACAGTTAGCGTTGGGGTGGGGGGCCCCGAAATTCGGCCTAGTGAAGGCGAAATCTATGACTACG  
CGCTCGATATGAGCCAAATGTACGCTTACGCCGTACCCGGCGAATTGCCGTCCTCCCTCAACCGCATG  
ATTCATACCAAGTGGTCACTAGTCCGGAAGACGCTCCGGCGTACCAGTTGAAACGCAAGTATTTCGAGGAT  
CCTCGGGAGTTCTTGTGTCATTTGGAAGAGTACCTGAAGCAGGTTGGCGGTACAGAGGAATATTGGCTGA

[0314]

GCCAGATTCAGAATCATATGAATGGTCCTGCAAAAAAGTGGTGGGAATATAACAAGACACGGTTAAGAA  
TTGGGTGGAATTCAAGAAGGAGTTCCTACAATACAGTGAGGGTACACTTACCCGTGATGCGATTAAGCGG  
GAATTAGACCTCCCGCAAAGGACGGTGAAGCTCTGGATCAATTTTTATGGCGTAAGCGTGACCTCTATCA  
GACATTATACATTGATGCCGATGAAGAACAGATCATTCACTAGCTCGTGGGGACATTGCAACCTAAACTCA  
AGCGGTTCTTGCTATCCACTTCCAAAACTCTTGAACAATTAATCCAGAAAGGGAAGGAGGTGCAGGGT  
TCACCTTGACCACAGCGAGGACCGAGTCTCAACGTGCGAGCGAGGCTCGGACGGGCGATAGTGTGGAA  
ACCTTGCCGCTTCTACCACTACATCACCAAAATACGTATCTGGTACACAGCCAGAGGCACCATCGCTCCA  
GCGACGGTAATC

서열번호 42

GGGCAGTTAGACAGTGTGACTAACGCCGGGGTGCATACGTACCAGGGGCACCGGGGTGCGCAATAAG  
CCAAATGTAATTCTCCAGATTGGGAAGTGTCTGACAGAGATGTTGGAACATGTCCGTGCGACTCATCGCCA  
CTTGCTCACCAGGTTCTCAAACAAGTAGAACCGGAACCTCAAGGGGCTCCAGAAGAGTGTGGGAAGTTG  
GAGAATAACCTCGAAGACCAGTTCGACAGATAACCAACGGTGGAAAAAGTCTATTAAGCCTGTCTCGC  
CCGTTGTCAAGAGACAATCGCACACTTGGAACGCTGGGTCAAACGGGAGATGAATGTGTGGAAGGAAGTC  
TTCTCCGTCTCGAGCGGTGGCGGATCGTTAGAAAAGTATGGCGGTAAATATTGCCAACTGACTCGGC  
TCGTCAAACGGTGTGCGTTGGCGTAGGCGGCCCGAAATTCGCCCTAGCGAGGGTGAGATCTATGACTAT  
GCACTTGACATGAGTCAGATGTATGCGTTAACTCCGTCCGACGGGGAGCTTCCAAGTATCCACAGCCTCA  
CGATAGTTATCAATGGGTAACCTTCTCCTGAAGACGCCCCAGCATCCCCAGTTGAGACACAAGTATTCGAGG  
ACCTCGTGAGTTTCTGTACCTCGAGGAGTACCTTAAACAGGTAGGCGGGACCGAAGAGTACTGGTTA  
TCGCAAATCCAAAACCATATGAATGGTCTGCCAAAAAGTGGTGGGAGTATAACAAGATACTGTGAAGA  
ATTGGGTAGAGTTCAAGAAAGAGTTCTTACAGTACTCTGAGGGGACGTTAACTCGTATGCGATCAAGCGC  
GAATTGGATTTACCTCAGAAGGACGGCGAGCCACTCGACCAGTCTTATGGCGCAAGCGTGACTTGTATCA  
AACCCCTTATATCGATGCTGACGAGGAACAATTTACAGTACGTAGTCGGTACGTTGCAACCAAACTTAA  
ACGCTTCTGAGCTACCCATTACCTAAAACGTTGGAGCAACTGATCCAGAAAGGTAAGAGGTGCAAGGG  
AGCCTGGATCATAGTGAAGAACCGAGCCCTCAGCGGGCTTCTGAAGCTCGGACCGGTGATAGCGTGAAT  
CTTACCACCTAGTACCACACACCGCCGAATGCGTCATCTGGTACCCAACCTGAAGCGCTTCCCACCTG  
CTACAGTCATT

서열번호 43

GGGCAGCTCGAGAATGTCAACCATGGGAACCTCCATTCTTTCAAGGTCATCGCGGCGGCTGCGCAACAA  
GCCAAACGTTATCTTGACAGATCGGTAATGTCTGTCAGAGATGCTGGACCACGTCCGGCGGACCCACCGG  
CATTTACTGACAGAGGTATCGAAACAGGTTGAACGTGAGTTGAAGGGTTACAGAAATCAGTAGGGAAT  
TAGAAAATAACTTAGAAGACCATGTCCTTCAGCCGTTGAAAACAGCGTTGGAAAAATCGATCAAGGCC  
TGCCTTTCCCGCTGCCAAGAGACCATTGCCACCTTGAGCGTTGGGTGAAGCGCGAGATGAACGTATGAA  
AGAGGTTTTCTCCGCTTAGAGCGGTGGGAGATCGGTTGGAATCTGGGGCGGGAAATATTGTCACGGT  
GATAATCATCGTCAAACAGTATCAGTCGGTGTGGCGGCCCTGAGGTACGTCCATCTGAAGGCGAAATTTA  
CGATTACGCTCTCGACATGTCGCAAATGTACGCTTAAACACCGCTAGCCAGGGGATGTGCTGTAGTTA  
GCCAGCCGACGACAGCTATCAGTGGGTTACGTTCCGGAGGATACCCCTCATCCCCGGTGGAGACGCA  
AATCTTCGAGGACCCACGGGAGTTCTTGACCCACTTAGAGGATTACTTAAAGCAAGTGGGGGTACAGAG  
GAATATTGGTTATCTCAGATCCAGAATCACATGAACGGGCCAGCCAAGAAGTGGTGGGAGTATAAGCAAG  
ACTCAGTAAAAAATTGGCTCGAGTTAAGAAGGAATTCCTCAGTATCCGAGGGGACACTTACGCGCGAC  
GCTATCAAGGAAGAAGTACCTCCCGCAAAGGACGGGGAACCTCTTGATCAGTTCTGTGGCGCAAGC  
GCGACTTGTACCAGACCTGTACGTGGAGGCGGATGAGGAGGAGGTGATCCAGTATGTTGTGGGGACTTT  
ACAACCTAAATTAAGCGTTTTCTCTCACACCCTTACCCGAAAACGTTAGAGCAACTTATCCAACGGGGCAA  
AGAGGTGGAAGGGAACCTCGACAATTAGAGGAACCAACCTCAGCGTACTCCAGAACCAACTGTGT

[0315]

GGTTCTGTAGAATCGCTGCCTCTCTACCGTCAGTCCAGTGGCTAGCGATGGTACTCAACCTGAGACT  
TCGCCATTGCCAGCGACTGTTATT

서열번호 44

GGGCCATTGACGTTGTTACAAGACTGGTGTCTGGTGAACATTTAAACACCCCGGTCATGTTGATCCT  
CGGTATCCCAGAAGATTGCGGCGAGGATGAGTTCGAAGAGACACTTCAGGAGGCGTGTCCGCAATTTAGGG  
CGGTACCGCGTGATCGGCCGATGTTCCGTCGTGAGGAAAATGCCAAGCGATCCTCTTGAATTGGCGCA  
GGATATTGACTATGCCTTACTCCCTCGGGAAATCCCTGGGAAAGCGGGCCTTGGGAGGTAATTGTGAAG  
CCGCGTAATCCGACGGCGAATCTTAAATCGGCTTAATCGCTTTCTTGAAGAGGAGCGCCGTACGGTCTCC  
GATATGAACCGTGTGTTGGGCTCGGATACTAACTGTTACGTCCTCGTGTACACATTAGTCTGAATTTG  
ACTTGGGCACAGACGCTGGGCGCAGCTGTCCAACATTGCTCGAACAGATGCTTACCGGGAGTTACGGG  
TCTTCAGTGGCAATACGATTTCCATCCAGGTGCTCTCGTTTTGACGCGTGGCTGGAGCATACCACGGAAA  
TGCTTCAAATGTGGCAGGTGCCTGAAGGGGAGAAAACGGCGCGCTTGATGGAGTGTTCGCGGGGCCAG  
CCCTGCAAGTCGTTAGTGGGTTACGTGCATCGAATGCCAGTACTGTGCAAGAGTGTCTTGTGCACTG  
CAGCAGGTATTCCGTCAGTGGAAAAGTCATAAGATTGCCAAGTAAAGTTATGCAAAGCTTACCAGGAGG  
CTGGGGAAAAAGTAAGCAGCTTCGTTTTGCGTTTTGAGCCACTGCTTCAGCGTGTGTAGAAAAACAAGTG  
GTCAGTCGCGCAATGTCAACCAACACGCTTAAAGCGTGTCTGTGGGCGCCACCTTCTGACAAGCTG  
CGTGATAAATTGAAGTTAATGAAACAGCGCCGTAAACCGCGGGTTTTCTTGGCGTTGGTTAAACTGTTACG  
TGAAAGAGGAGGAGTGGGAGGCCACTTAGGGCCAGACCGCGAGTCATTGGAGGGGTTAGAAGTGGCACC  
GCGCCCGCCAGCAGGATTACGGGTGTTGGCGCAGTACCTTCCGGCATCCGGGAATTCATTGATGCC  
GTCTTCGCAAGGGTACCGGCGCGTCCGGGTCGTGGTTCAGCACCCTCGGGGCGGCGTTGCTCGTGCAAG  
CTCTCGTGGCTCTCGTAAGCGGAAACGGCACACCTTCTGCTATTCTGTGGTGAAGGATGGCCATATTCGTG  
CCAATGCATTAACCTTAGCAATCTCTGTGGCTAAGGAGACCAAAGAGATTTTGAAGGGGGAGAACGT  
GAAGCGCAAACGAATTCACGT

서열번호 45

GGGGCTTACGCTCTTAGAAGACTGGTGTAAAGGTATGGACATGGACCCGCGAAGGCTCTCTGATTGT  
AGGTATCCGATGGAATGCAGTGAGGTGGAAATCCAGGATACAGTTAAAGCTGGTCTTCAACCTCTGTGCG  
CTTATCGTGTACTCGGCCGATGTTCCGGCGGGAGGATAATGCGAAGGCTGTTTTATTGAGCTGGCAGAC  
ACCGTGAATTACACCACGTTACCGTCTCACATTCGGGTAAAGGGGTTCTGGGAAGTCGTTGTTAAACC  
TCGGAACCTGACGACGAGTTCCTTTCTCGGCTTAACTACTTCTTGAAGATGAGGGCCGCTCGATGACGG  
ATGTCGCCCCGGCCTGGGGTGTGTAGCTTACCTGCGGAATCACTGGACGCGGAAGTAATGCCACAGGT  
CCGCTCCCCACCATTAGAACCTCAAAGAGAGATGTGGTACCGTAAGTAAAAAGTGTGTTAGTGGTACCG  
CGTCGCCTTCCGCGGGGAGGAGACATTTGAGGACTGGTTAGAGCAAGTACCAGATCATGCCTATCTG  
GCAAGTATCTGAAGTTGAAAAGCGCCGTCGGTTACTGGAGTCACTCCGGGGCCCGGCACTCTCAATTATGC  
GCGTGTTACAAGCCAATAACGATAGCATTACCGTTGAACAGTGTGTTGGATGCATTAAGCAGATCTTTGGC  
GACAAGGAAGACTTCCGTGCCTCTCAATTTGTTTTCTTCAAACGTCCTTAAAATGGGGAGAAGGTGAGT  
ACGTTCTGTGCTGCTTTAGAGCCACTCTTGCAAAAGGCGGTTACAAAGAGCCACTTTCGGTACGTAGTACT  
GATATGATTCGGTTAAAGCACCTGTTGGCACGCGTAGCCATGACCCCGCACTGCGTGGTAAACTCGAATT  
ACTCGACCAACGCGGGTGCCACCTAATTTCTTGGAGCTGATGAAGCTGATCCGGGATGAGGAAGAGTGG  
GAGAATACTGAAGCTGTGATGAAAAATAAAGAGAAACCTTCAAGTCTGTCGCCGCGTGCATCAGGCCGTC  
AAGTCTCGCGCCGAGGCCAGTGAAGTGTCCGCAAGCAACAGTCCAAGCACGTAGCTTCTGATTCTAGC  
CCGACAGCAGATTACGGGGGCTTACCACCTCTGTCAAGCGTCGGCGCCTTTGGGTTCCGGAGAGCACAG  
TGGGGAAAGACCAGGGCAAGCTACTTATCCGAAAGCAGAGAATCAGACTCCAGGGCGTGAGGGCCCGCA  
GGCGGCTGGGGAGGAATTGTAATGAGGCCGGGGCCGGCGCATGTCCACCCGAAACCGTGGGAAAC  
C

[0316]

서열번호 46

GGGGCTGTGACAATGCTCCAGGACTGGTGCCGTTGGATGGGCGTGAACGCTCGGCGGGGGCTGTTAATCT  
TAGGTATCCCTGAAGACTGTGACGATGCAGAGTTCCAAGAGTCGTTAGAAGCTGCACTCCGTCCTATGGGT  
CACTTTACTGTACTCGGTAAGGCCTTCCGCGAGGAAGACAACGCTACCGCTGCGCTGGTGAATTAGATCG  
CGAGGTTAATTACGCACTTGTTCACGCGAAATCCGGGCACCGCGGGCCTTGAACGTCGTGTTGTTTC  
CTCGGTGCTCCGCGGAGGAATCCTGGGGTTAGGCCGCTGTTCCACTTTCCTGAACAGGAGGGCCAAATG  
GTAGAATCGGTTGCGGGGCACTGGGGGTAGGTTGCGCCGCTGTGTTGTTACGCTCGATCGGGCAA  
GCTGTACAACCATGGGTAGAAGCTGTTGCTGCCAAAGCTTAGGGGTATTTAGTGGTCGTGATCAACCTGC  
ACCTGGTGAAGAAAAGCTTCGAGGTCTGGTTGGATCATACGACCGAGATGTTGCATGTGTGGCAAGGCGTG  
TCGGAACGGGAACGCGCCGTCGTTGCTGGAAGGGTGCCTGGCACAGCCTTACAACCTGTACATGCCTT  
ACTGGCAGAAAATCCGGCACGGACAGCACAAGATTGCTTGGCTGCATTAGCCCAAGTTTTTGGTGATAACG  
AAAGCCAGGCAACGATTGCTGTAAATGTTTGACAGCCCAACAGCAGAGTGGCGAACGCCTCTCTGCGTTC  
GTTCTCCGCTTAGAAGTACTTCTGCAAAAGGCTATGGAGAAGGAAGCATTGGCGCGCGCTCAGCGGATC  
GGGTGCGTCTTCGTAGATGCTGACACGCGCACATCTCACAGAGCCGTTGGATGAAGCCTTACGAAATTG  
CGTATGGCAGGGCGTTCTCCGTTCTTTTTGGAAATGCTCGGCTTAGTACGCGAGTCAGAGGCTGGGAGGC  
AAGTCTGGCTCGTCCGTCGGGGCGCAAACCCAGGAGGTTGACGGGGCCCGGGGGGGCCCAAGCAGT  
TGCGCGTCCAGCCTAAGGTTGAAGCTGTACCTGGTGGCCCTGGCCGGGAGCCAGAAGGTCTCTCCAA  
GCCGGGGCCAAAGACGCGGAAGAACTTCTCAAGAGGGCTTAAAGCCGGTTTTAGAGGAATGTGACAAAT

서열번호 47

GGGGCGGTCCACATGTTGCAAGACTGGTGTGCGGTGGATGGGCGTGAATGCTCGGCGGGGTTTATTGATCT  
TGGGTATCCAGAAAGACTGTGACGACGCCGAGTTTCAGGAGTCGCTCGAGGCCGCCCTTCGTCCAATGGG  
GCATTTTACGGTTCTGGGCAAGGTGTTCCGTGAAGAGGATAACGCTACAGCAGCTTGTGGAGCTTGACC  
GTGAGGTGAATTATGCGTTAGTACCTCGCGAGATTCCAGGTACCGGTGGGCCATGGAACGTAGTCTTCGTC  
CCACGTTGCTCGGGGGAGGAATTTCTGGGGCTTGGGCGCTATTCCACTTTCAGAACAGGAAGGGCAGA  
TGGTGAAAGCGTAGCAGGCGCTCTGGCGTTGGTCTCCGGCGCGTGTGCTGGTTACGCTCCATCGGCCAA  
GCAGTCCAACCATGGGTTGAAGCCGTACGCTATCAATCTTAGGTGCTTCTCAGGCCGTGACCAAGCCGGC  
GCCTGGTGAGGAATCCTTCAAGTCTGGCTCGATCATACAACTGAGATGCTGCATGTATGGCAAGGTGTCT  
CAGAGCGGGAACGCGCGGGCGGTTATTAGAGGGGCTCCGTGGGACTGCGCTCCAATTAAGTACATGCGCT  
TTTGGCCGAAAATCCAGCCCGTACTGCCCAAGATTGCTGGCAGCACTCGCCCAAGTATTGCGCGACAACG  
AATCGCAGGCAACAATCCGCGTAAAGTGTCTTACAGCACAGCAGCAGTCAGGGGAACGCTTAGTGCGTTC  
GTTCTGCGGCTGGAAGTGTACTCCAGAAAGCCATGGAAGGAGGCATTGGCTCGCGCAGCGCTGACC  
GTGTACGCTGCGGCAATGCTTACTCGCGCACATCTACCGAGCCTCTCGATGAAGCACTGCGGAAACTG  
CGATGGCAGGCGCCAGCCCGTCTTCTGAAATGTTAGGCTTAGTCCGGGAGTCCGAAGCCTGGGAGG  
CCAGTCTGGCACGGTCAAGTGGGGCACAAACGCAAGAGGGTGCAGGGGCACGGGCGGGTGCACAAGCA  
GTTGCACGTGCTTCACTAAAGTTGAGGCAAGTGGCGGGTGGGCCAGGCCGTGAACCGGAGGGTTTGGCC  
AAGCCGGCGGGCAGGAAGCCGAAGAATTACTCAAGAAGGTTTTAAACCGGTTTTGGAGGAATGCGATAA  
C

서열번호 48

GGGGTGAAGATTTGGCGGCATTTACATCGTATTAAGCTTGAGAACGAAATCCGGCAGGCGCAGGTCC  
AATGGTTAATGGAGGAAAACGCCCGCTGACAGGCCAGATCCCTGAACTTCAAAAGTCGCAAGCCGCGAA  
GGAGTATGATCTTCTGCGTAAATCTTGGAGGGCAAGGAGCCGCAAAAACCTGCCAGAACATATGAATCCAC  
CGGCCGCTTGGGAAGCACAAAAGACTCCAGAGTTTAAGGAACACAGAAACCTCCTGAACCACAGGATTT  
GCTTCCTTGGGAGCCGCTGCTGCTGGGAGTTGCAAGAAGCACCGGCTGCCCTGAGTCACTGGCTCCGC  
CTGCAACCCGTGAGTCTCAGAAACACCTATGGCGCATGAAATCCCTACTGTATTGGAGGGCAAGGGCCT

[0317]

GCCAACACACAAGACGCTACGATTGCTCAAGAACCAAGAATAGCGAGCCGCAAGACCCCTCCAAATATCG  
AGAAACCTCAGGAAGCTCCGGAATATCAAGAAACAGCGGCACAGTTGGAGTTTTAGAACTTCTCCACCT  
CAGGAGCCACTCGAACCGAGCAATGCGCAAGAATTTCTCGAGTTGTCGGCTGCCAGGATCCTTAGAAG  
GCCTATTGTAGTTGAAACGTCGCGGCTTCGGAGTTCCACAGGCTCTATCGGGCTTGAAGCCACCGAC  
TTTCCGCTGCAGTACACGCTTACCTTCTGCGCAGCCAGAAGTTGCCAGAATTTTTGGTCCAACCTAC  
AGTTATATGCGGGTACGTGGGCATTATACCTACCGAGGCGGGCTTAGTGTCTGTTGTAGGCAATTGTTT  
CTCAGGGCGCGGGCTGGTGGTTTTCAAGTTGTTTTGGATATCCAGTCGCCTCTGTTAGAACAGTGTGAAA  
GTTTTATCCCGGTTCTCAAGACACATTTGACAATCCGGAACATGAAGGACGCAAACCAATGCATCCACC  
AGCTTTGTCAGGGGAGGGTCAATGTGGCCACACTTCCACCTCATTGCACAAGAGCTTAATTGGGATGAA  
AGCACGCTGTGGATCCAGTTCCAGGAAGGCTGGCCTCATCCATCCAGGATGAACCTTCCCATACATCGCT  
GCTACCAACCTGAGTGATCTGATTACTCAATGCATCTCATTAGAGGAAAAGCCTGACCCAAACCGTTAGG  
GAAGTCCTCCTCGGCGGAGGGGATGGCCGGAAGTCCGCCAGCAGAAAACCAACCTATGCAAGCTGCG  
ATCAATTGTCTCACATTTCCGAAGCAGAGTGGGTTCTGTTGGCACAAGGCGGGCTTTGTCTATTGCGGC  
TATCCGGTCACTTCGCACGTGATTGCCAGTGAAGCCACACAGGCGTTACAGGCAGGGAACATTACAGGC  
TTGCCAA

서열번호 49

GGGGTGCAGCCGACAGTAGCAAAGCTGAATCGCCGGCTCTCGTCCCTACCGAACGCACAAATGGATG  
ACGTTATTGATACATTAACCTCCCTGCGTCTGACGAATTCGGCTCTGCGGCGGAGGCTAGCACTTTCGG  
GCCGAGAAAGCAAATTAACATAATGCTCGAGTCAAGTATGCGCGAGTTAACGCTGTTACGGACCCGTGC  
GCGGATTCGGGGGCCCTGCAGATTACGCCACCAATTTGTTCTATTACTAGCAACGGTACTCGCCGATGA  
CGACTCTCCAACAGTTTACCTGAACGTTTTCTGGCGATCCTGGCCGTTAGCTGGTTTTCTTATGCAGAT  
GGACCGTTTTATGATCTTCAAGCTAGCCGTTTTCCAGGGGAGGCAGAGCGTGTTCGTTCTGGTGTCCG  
GCTTAACCTGGCGAAGCAGAAAATGGGCCATTCTCACATGCAACCAGACTCTCTTTGCGTAACAACATC  
AAGGCTTCTTAGCAGAGTTACGGCGGACCTATAAGAGCCGTTGCGTCACGCCCGGCGGGCGCAAATCCG  
GAAGACATCGGCCTCGAACGGGCAGTCCGTGAACGCCAAATGCTTTGCCGGCAACTTGCATCAGCAGGT  
ACAGGCCATGCCCGGTACACCTGTAGTAACGGGACTTCCCGGACCGGCATTACCAGCACGGGCGC  
GTAACCTA

서열번호 50

GGGGACGGTCGGGTACAGTTGATGAAGGCTTTATTGGCTGGCCCTTACGTCGGGCGGCACGCCGTTGGC  
GGAATCCTATTCCATTTCCAGAGACTTTTGTGGGATACTGATCGCCTCCCGGAGTTTATCGTCCAACTT  
CGTCCTACATGTTGTTGACGAAAATACTTTCTAACGACGCTCTGAAAGTGACATTTCTATTACCCGGCT  
GACAGGTCCAGCCTTGAATGGGTATTCCGTACATTCGTAAGAAAGCCCGCTTCTTAACGACTATCGGG  
GTTTCTGGCCGAGATGAAGCGGGTTTTGGGTGGGAAGAGGACGAGGACTTT

서열번호 51

GGGGAAGGTCGGGTGCAACTTATGAAAGCGTTGCTTCCCCGCCGCTTCGTCAGCAGCACGTGCTGGC  
GGAATCCAATTCCTTTCCCGGAGACTTTTACGGGGACACCGATCCGCTCCAGAGTTTATTGTGCAGACG  
TCAAGCTATATGTTGTTGGATGAGAACCGTTCTTAACGACGCGTTGAAAGTGACTTTCTTAATTACGCGT  
TTGACTGGCCGGCTTTACAATGGGTGATTCCATACATTAAGAAAGAGTACCAGCTTCTCAGTGATTATCGC  
GTTTTTTAGCCGAGATGAAGCGGGTCTTCGGGTGGGAAGAAGACGAAGACTTT

서열번호 52

[0318]

GGGCCGCGTGGGCGTTGCCGTCAACAAGGTCCTCGGATTCCGATTGGGCGAGCGCCAACTATGCCAACG  
CCCACCCGTGGCAACAAATGGATAAGGCTTCGCCAGGCGTTGCTTACACACCTTTGGTTGATCCTGGATTG  
AGCGGCCTTGTGCGGTGACACGGTTTGTGTGCGCACCACAATGGAACAGAAGACACAGCGTCAGGCAC  
TTGTGTGGTAAGCCTGCTGAGCGTGTCTCTCGCGGGGCATATGCCGAGCTCACGCCACATCGGGTTG  
ATTTCTGTTGGGTTCTGGTAGCGACCCAGGCACATTCGACGGCAGTCCATGGCTCTTAGATCGCTTTTGG  
CGCAACTTGGTGATTACATGAGTTTTACTTTGAACACTACCAGGACAATATCAGCCGTGTCTGCGAGATT  
TTCGTCGGTTAACGGCCGCGCTCAGGCATGGGCTGCTCCTTACCTGGACGGGGACCTTCCACTGCCAGAC  
GACTACGAATTGTTTTGCAAGACCTTAAGGAGGTAGTACAGGACCCTAACAGTTTCGCCGAGTATCACGC  
CGTGGTGACTTGTCCACTCCCTCTTGCTTCGTCCTCAACTTCTGTAGCTCCTCAGCTTCCGGTGTACGCCAA  
TACCTTGCAGCTTCTTGGAGGGCCTTGTCTTGGATATGGGTACGGCGCCTCGGTCACTCCCGGCCGTAT  
GGCCACACCGGCAGTCTCCGGCTCGAACTCCGTTTCTGTTCTGCCTATTTGAACAACAACCTCACAAGGA  
ATCCACTCCAGGCCCGAAAGAGCCACCTGTTCTCCCTAGCTCGACTTGTCTAGCAAAACCGGGTCTGTCTGA  
ACCAGCCAGTTCACAACCTGAAGAGGCTGCTCTACCCCGGTGCCGCTTTGTCAGAGTCGGCTAACCCAC  
CGGCTCAGCTCCAGACCCTGCTCACCTGGTGGTCTAAACCACAAAAAACCGAAGAGGAAGTTTTAGAA  
ACTGAGGGGGACAGGAAGTTAGCTGGGGACGCCGACGAGGTCGTAGAAGCGCCGAAACACCAGG  
TGAACCACCGCTCAGCCCTGGGTTT

서열번호 53

GGGGTTGATGAATTGGTGCTCTTGTGCACGCGCTGTTAATGCGCCATCGGGCGCTTCCATTGAAAATCT  
CAGTTGATGGAGCACTTCGCTTGTGGTCTGCGAACGGGCGAGCCTTCTTCGTCAAGTACGTCGCCGAG  
CTGTCAGTGCATTTCTGAGACTTTAACGGGGAGTACACGGTTACCTGAGTTTATCGTCCAAACCGC  
AAGCTATATGTTAGTTAATGAAAATCGCTTTGCAATGACGCAATGAAAAGTCGCTTTTTGATTAGCTTCTT  
ACTGGTGAAGCAGAAGAATGGGTCGCTCCATACATTGAGATGGATTACCAATTCTGGGGACTACCGTGC  
GTTCTTGGATGAGATGAAGCAGTGTGGTGGGACGATGATGAAGATGACGACGATGAGGAAGAGGA  
GGATGACTAT

서열번호 54

GGGCTGTGGATTTAGTACAGGCTTTGGGTTGTTGCCATCCCTCGTAAGGCCGAAGATCCCAATTTAG  
CGAAAGCGATGCAGCTTTACAGGAGGAATTGTCTTCCGGAAACCGCACGGCAACTTTTTCGTCAATTTCC  
CTATCAAGTCATGTCGGGGCCTCATGAAACACTGAAACAGTTACGGAAAGTTATGTTTTAGTGGCTGCAAC  
CTGAAGTCCATACAAGGAACAATCTCGAAATCTGATGCTGGAACAGTTCTTGACCATTCTGCCTGGTG  
AAATTCAGATGTGGGTCCGCAAGCAGTGCCCTGGTAGTGGGGAGGAGGCGGTTACGTTAGTAGAATCCCT  
GAAAGGTGATCCACAACGGCTCTGGCAATGGATCTCCATCCAAGTCTGGGTGAGGATATCCTGTCTGAGA  
AAATGGAGTCACTTCTTCCAGGTGGGCGAAGTGAGCCACACTGGAAGTTGTACCTCAGGAACTGGG  
GTTAGAGAATTCATCTCAGGGCCGGGGAACTTCTTTCGCACATCGTGAAGAGGAGTCTGACACTGAAG  
CAGAGTTGGCGTTAGCGGCATCCAGCCAGCTCGTTTGAAGAACGGCTGATTCGGGATCAGGACCTTGG  
GGCGTCCCTCCTCCGGCAGCACCCGAGGAGCAATGGCGTCAATTAGACAGCACTCAAAAAGAAACAATATT  
GGGACCTGATGCTGGAGACCTACGGCAAAATGGTATCCGGCGGGGTATCTCACACCCGAAGTCCGATTT  
AACGAACTCAATTGAGTTCGGTGAAGAGTTGGCAGGTATTTATTTACATGTAACGAAAAGATTCGCGGC  
CTACCTGCATTGGTGACCCCAAGAAAACGACAAAAGAAAACCTTAATTTGGAAAACCATCGTGACCAGGAA  
TTATTACATGCCAGCTGCCAGGCTCGGGCGAAGTGCCATCCAGGCATCGTTACGTGGCTTCTTTACCGAG  
GACGAACCTGTTGCTTCGGCGAAGGGGAGAACCTTCTGAGGCACTTCAGAATATCCAGGATGAGGGGA  
CTGGCGAACAGCTGAGCCGCAAGAACGCAATGTAAGAAAACGTTGGGTCAACATTTGCCAAATCCGCAC  
TCGGGGGAGATGTCAGCATGTGGCTTGAAGAAAACGGGAGACCAGCAGAAAGGCCAACCCAGTGCA  
CCAATGGCGCAGAAATTGCCAACGTGCCGCAATGTGGCAAAACGTTTTATCGCAATAGTCAACTTATCTTT

[0319]

CACCAACGCACACACACCCGGTGAGACATATTTCAATGCACCATCTGCAAAAAGGCGTTTCTCCGGTCATCT  
GATTTTCGTGAAACATCAGCGGACTACTACTGGCGAAAAACCTTGAAATGTGACTATTGTGGCAAGGGCTT  
TAGTGATTTTAGCGGGCTTCGGCATCACGAGAAGATCCATACCGGCGAGAAGCCATACAAGTGCCAATCT  
GTGAGAAATCTTTATCCAGCGCAGTAATTTTAAACCGCCACCAACGGGTTTACACCGGTGAAAAGCCTTATA  
AATGCTCGCATTGTGGCAAGAGCTTCAGCTGGAGCTCCTCGCTCGATAAGCATCAACGTTTACATCTGGGG  
AAGAAGCCGTTCCAA

서열번호 55

GGGACTCTCCGCTTACTTGAGGATTGGTGTGCGGGGATGGACATGAACCCACGTAAGGCCCTTCTTATCGC  
CGGGATTTCCAGTCATGTTTCAGTCGCCGAGATTGAAGAGGCGCTCCAAGCCGGGCTTGTCTTTAGGCG  
AGTATCGTCTCCTGGGCGGATGTTTCGCCGCGATGAAAATCGCAAAGTAGCGTTGGTTGGTCTCACAGCT  
GAAACTAGCCATGCGCTTGTACTAAAGAAATCCTGGTAAAGCGGGATCTGGCGGTTATTTTAAACC  
ACCGGACCCGGACAATACGTTTCTTCTCGTTTGAATGAGTTCCTCGCGGGCGAGGGGATGACGGTGGGG  
GAACTTAGTCGTCTTGGTACGAAAAATGGGTCAATTAGACCCTGAACAGGGTATGATTCCGGAAATGTG  
GGCGCCGATGCTGGCACAGGCTCTGGAGGCTTCCAACCGCTTACAGTGCCTTAAGTACAAGAAGCTGC  
GCGTTTTTTCAGGGCGGAGTCTCCAGAGCCGGGTGAGGAGGAATTCGGCCGTTGGATGTTCCATACCACC  
CAGATGATCAAAGCGTGGCAGGTGCCGGATGTCGAGAAAACCGCCGGCTGTTGGAATCACTCCGCGGGC  
CGGCACTTGACGTTATTGCGGTTCTGAAAATTAACAACCCGTTAATTACGCTAGATGAATGTTTCAAGCAC  
TTGAAGAGGCTTTGGGGTACTGACAATCCTCGGGAATGCAAGTAAATACTTAAACGACCTACCATAAG  
GACGAGGAGAAATATCAGCCTACGTAAGTCTGCGGCTGGAACCGCTGCTGCAGAAGCTGTCAGCGGGGG  
CTATTGAACGGGACGCTGTTAATCAGGCTCGCCTGGATCAGGTAATCGCTGGGGCGGTACATAAAACTATC  
CGCCGTGAGCTGAACCTGCCTGAAGACGGGCCGCGCCAGGCTTTCTTCACTCCTGTTTTGATTAAGGA  
TTACGAGGCAGCTGAAGAGGAGGAAGCATTACTTCAGGCCATTCTGAAGGGAACCTTACT

서열번호 56

GGGACAGAACCGGCTCGCGACGAATTAAGTGAAGAAATTAATAATCTTCTGTAAGGTTATGAAACAGA  
GTGAGGAAAACAACAATCTTCAATCCCAAGTCCAGAAACTACTGAGGAGAATACTACTCCGTGAGCAA  
GTTGAACCTACACCTGAAGATGAAGATGACGACATTGAGTTGCGGGGCGCAGCAGCCGACGCGCGCCTC  
CGCCGCGATCGAGGAGGAATGCCCGGAGGATTTACCGGAAAATTTGATGGTAATCCGGACATGTTAGC  
GCCATTCATGCCCCAGTGCCAAATTTTTATGAAAAGTCTACGCGCATTTTGTAGTATCGCGTACGTGT  
ATGTTTTGTGACGAGCATGATGACTGTCGCGCAGCCGTTGGGCGTCAAGAAATTTGAGCGGTGCGAC  
TACCTGATGCATAATTACCCGGCTTCATGATGGAGATGAAACACGTTGTTGAAGACCCGACGCGGGGG  
AGGTGGCCAAACGCAAGATCCGCGGTTGCGGCAGGGCATGGGCAGCGTAATTGATTATAGTAATGCGTT  
TCAAATGATTGCGCAGGATCTGGATTGGAATGAACCTGCTCTCATTGATCAATATCATGAAGGGCTTAGTG  
ACCATATCAAGAGGAACCTCTCACCTGGAAGTGGCTAAATCTCTCTCCGCCCTTATTGGCCAATGCATTC  
ATATTGAGCGCGCTTTGACGTCGCTGCTCGGAAACCGCTAGTCCACCACGGCTTTAGTGCTC  
CCACATATCGCGTACACCATCAAGTAGATCCTACTGAGCCAGTGGGGGTCACGCATGCGCTTAACCCA  
AGAAGAAAAGGAACGTCGTCGTAAGCTGAATTTATGCTGTACTGCGGCACTGGTGCCATTATGCCGATA  
ACTGTCTGCCAAAAGCCAGTAAGTCAAGCCCGGCTGGGAAACTTCCAGGTCCTGCCGTCGAGGGCCCTTCT  
GCTACCGGCCAGAGATTATCCGCTCCCCGCAAGACGATGCGTCGTCGCTCATCTCCAGGTAATGCTCCAA  
ATCCACCTCCTGGCCGCGACACTCTTTGTCGGGCGATGATTGACTCTGGGGCGTCTGGTAATTTTATT  
GATCACGAGTATGTTGCTCAAAATGATATCCCTCTCCGGATCAAAGACTGGCCTATTCTGTTGAAGCCATC  
GATGGCCGTCGATCGCGAGCGGTCTGTGGTTTCAATGAAACGCATGACCTCATCGTTGATCTGGGTGACCA  
CCGTGAAGTATTATCCTTTGATGACTCAGTACCGTTTTTTCCAGTTGTTTTGGGGTCCGTTGGCTTTTCG  
ACTCACGATCCTAACATCAGTGGTGCACACGGTCGATTGTCTCGAATCGGAATATTGCTGTTATCATTGC  
CGCATGTATTACCAATTCGCCGCTCTCTCCGCGCCTGCGCCGCAACCTCTGTATTACCCGGTGGAC

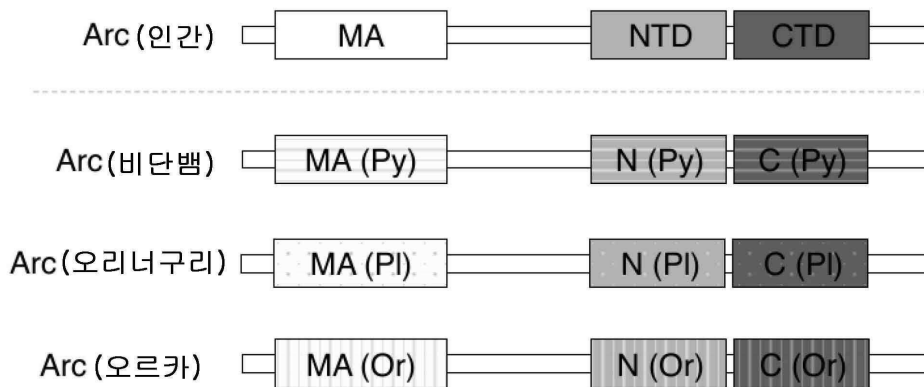
[0320]

GGTTACCGTGTACCAGCCAGTTGCTACTACTACTACGTACAAAACGTGTACACGCCTGTTGATGAACACGTG  
TACCCAGATCACCGCTGGTCGACCCTCATATTGAGATGATCCCGGGTGCACACTCGATCCCATCGGGCCAT  
GTTTATTCTTGTCTGAGCCAGAAATGGCCGCTTACGGGATTTTGTGGCCGGAATGTCAAAGACGGCCT  
GATTACCCGACAATTGCACAAACGGTGTCTCAGGTGTTGACAGTGAAGCGGGGCTGGAAGTTGCAAGTC  
AGCTATGATTGTCGTGCGCAACAACCTTCACTATTGAGAACCAATATCCACGTCCTCAGCATCCCTAATCTCG  
AGGACCAGGCACATCTTGAACATATACTGAATTTGTACCTCAGATTCCTGGCTATCAGACTTATCCTACGT  
ATGCTGCTACCCAACATACCCGGTAGGTTTTCGATGTTACCCAGTAGGCCGGGACGGGCAGGGCCGCTCT  
TTATATGTTCTGTGATGATTACATGAAACCCGATTGGTACCAGCCGCTCCGGTCCCACAGTACCACCTC  
CTCAACCTCCACCACCTCCGCCGCTCTCCACCCGACCTTCTTACTCGACATTA

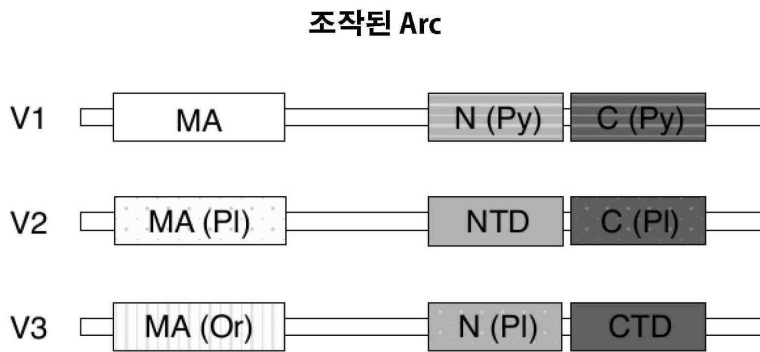
[0321]

도면

도면1

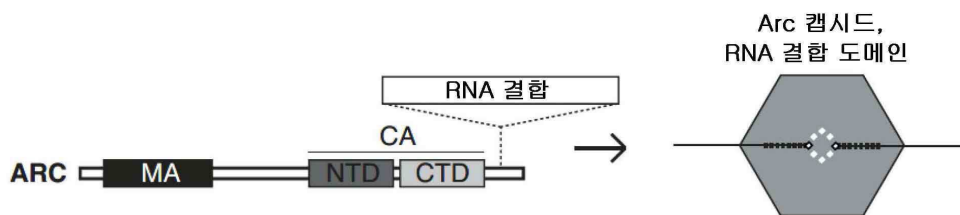


도면2



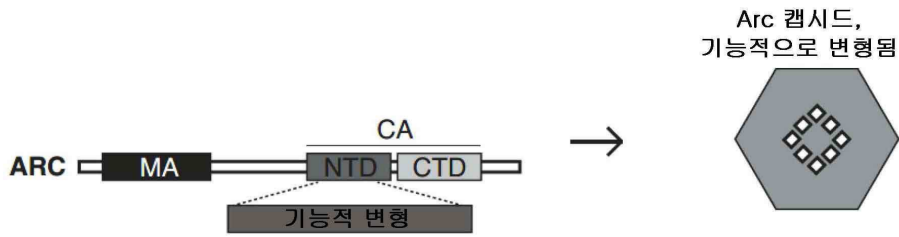
도면3a

**특정 RNA 페이로드를 운반하기 위한 Arc 재조작**

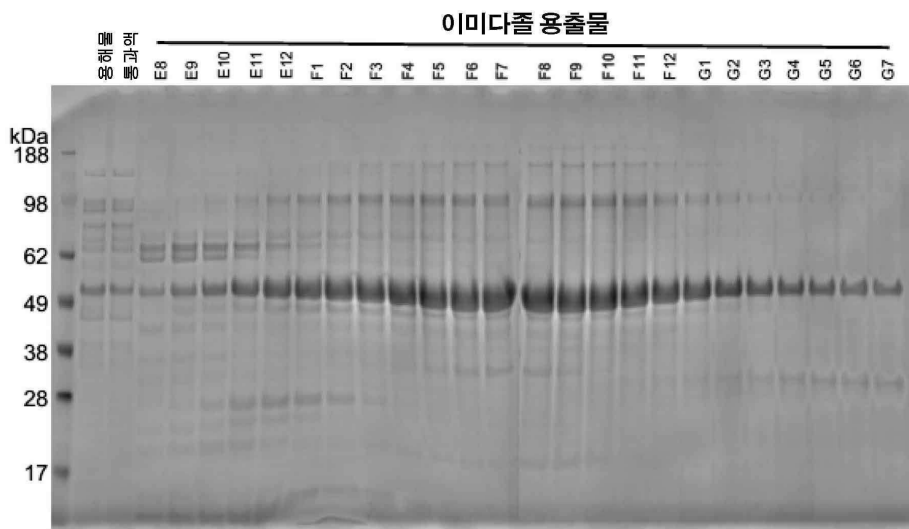


도면3b

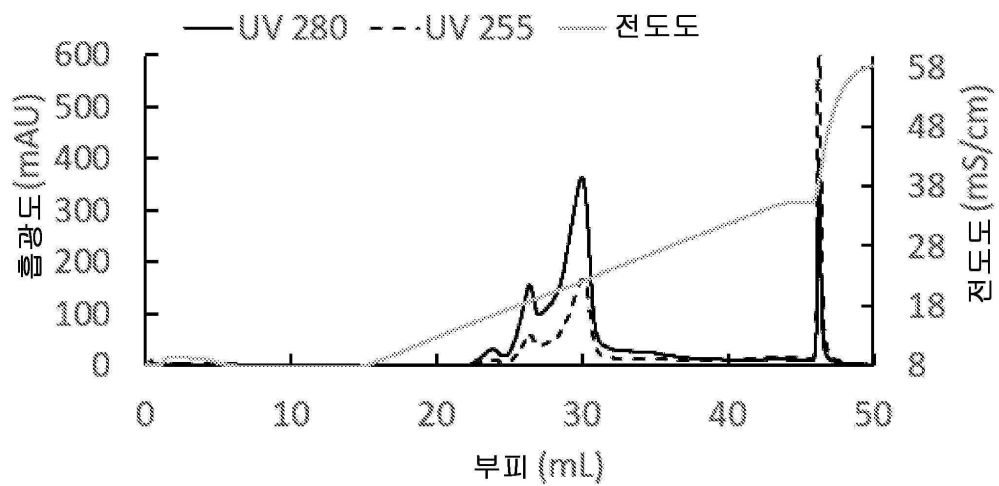
기능외 효과를 제거하기 위한 Arc 재조작



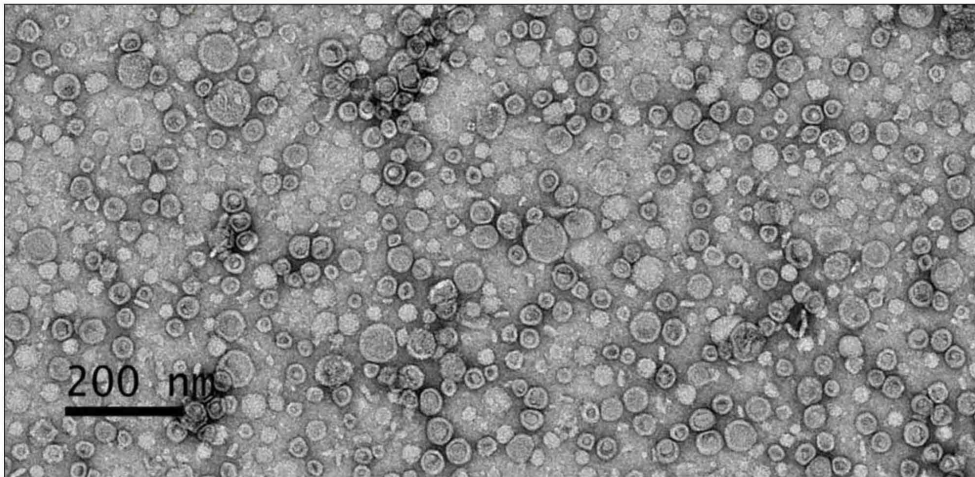
도면4a



도면4b

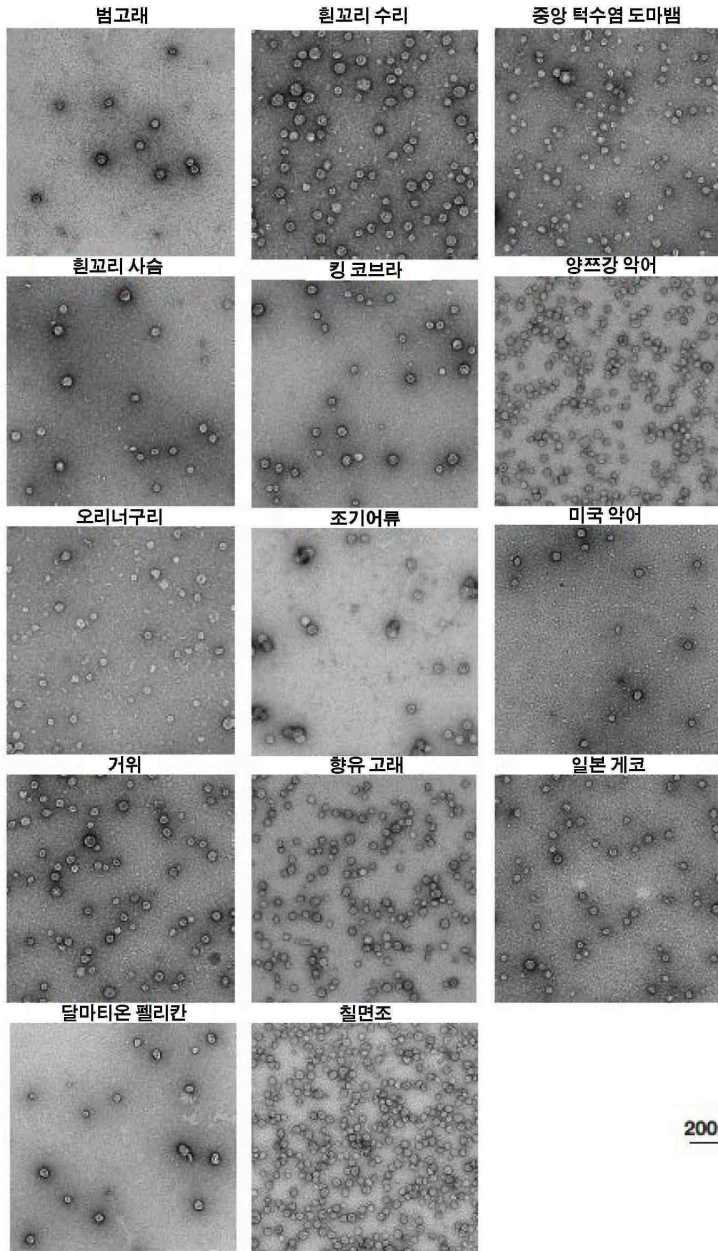


도면5



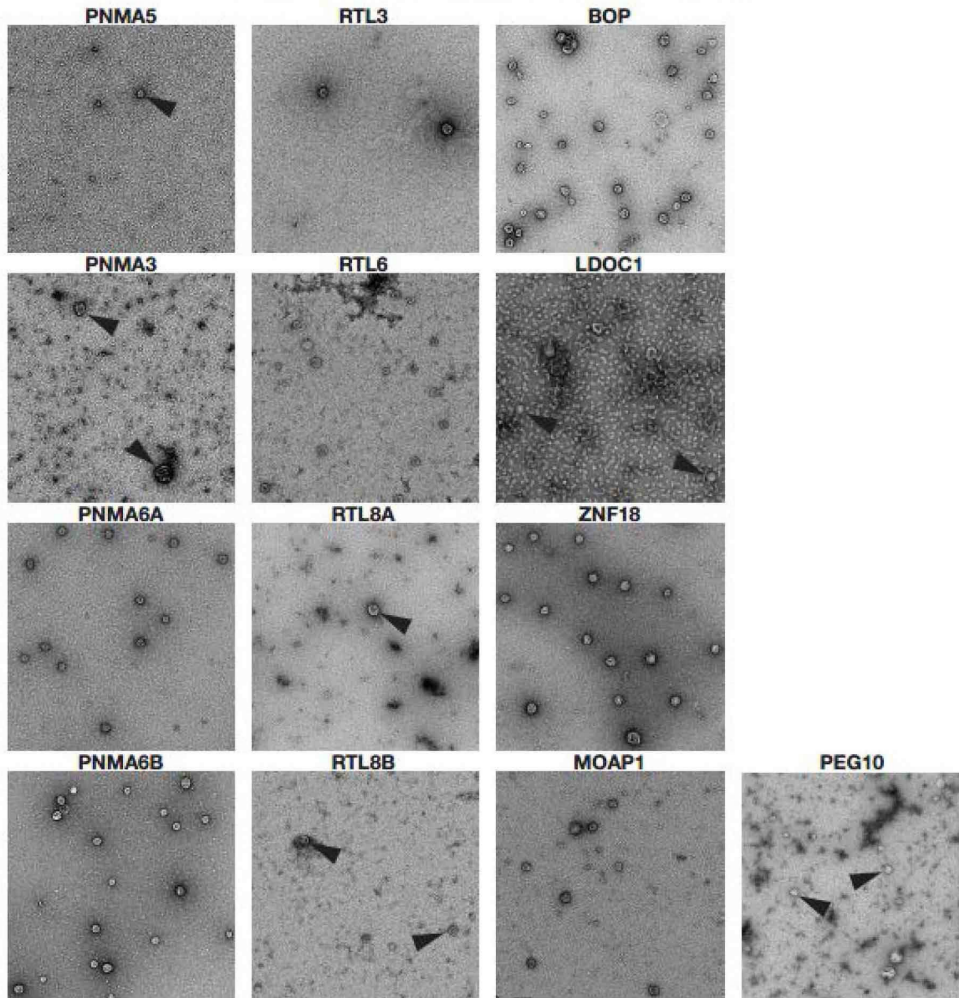
도면6

Arc 동원체



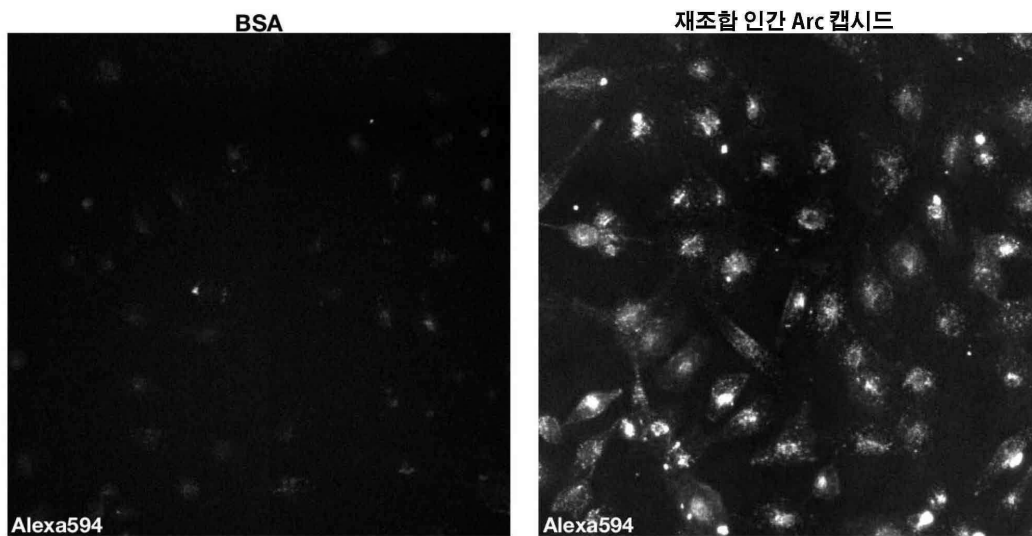
도면7

내인성 인간 Gag 유사 유전자

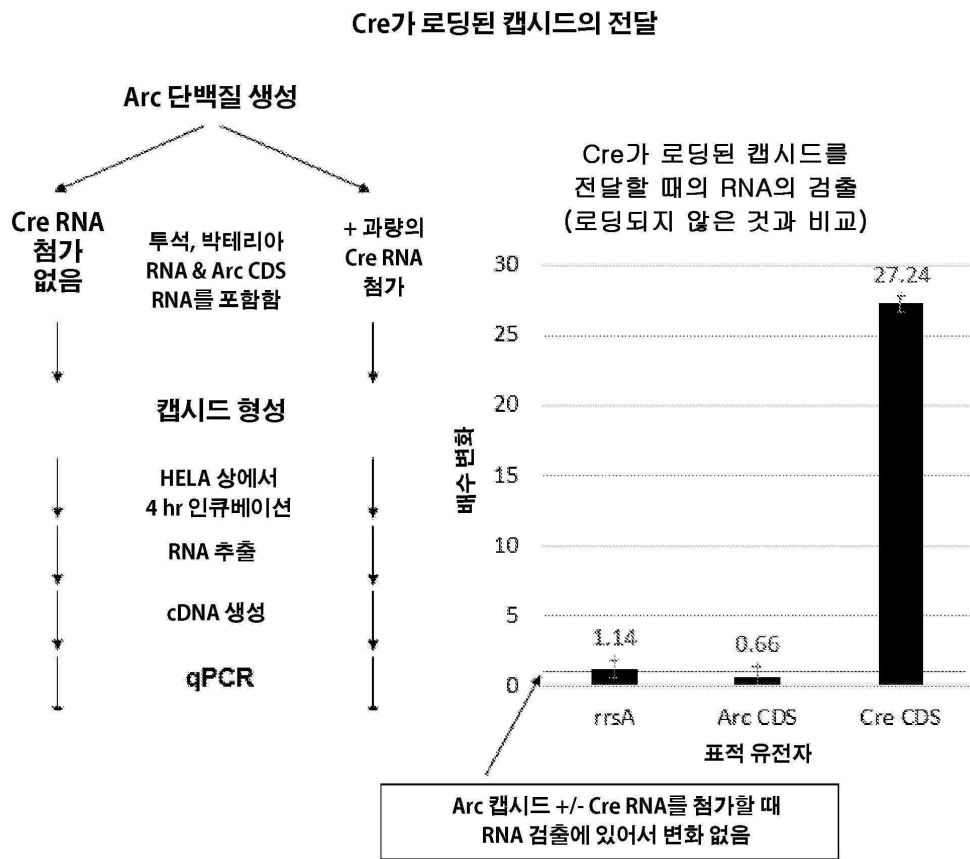


200nm

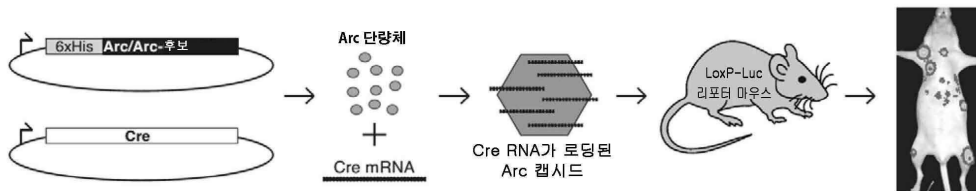
도면8



도면9



도면10



**서열 목록**

SEQUENCE LISTING

<110> VNV NEWCO INC.

<120> ARC-BASED CAPSIDS AND USES THEREOF

<130> 54838-702.601

<140> PCT/US2019/051786

<141> 2019-09-18

<150> 62/733,015

<151> 2018-09-18

<160> 85

<170> PatentIn version 3.5

<210> 1

<211> 396

<212> PRT

<213> Homo sapiens

<400> 1

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

1                    5                    10                    15

Pro Arg Gly Gly Gln Val Ala Lys Pro Asn Val Ile Leu Gln Ile Gly

                  20                    25                    30

Lys Cys Arg Ala Glu Met Leu Glu His Val Arg Arg Thr His Arg His

                  35                    40                    45

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

                  50                    55                    60

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

65                    70                    75                    80

Thr Ser Asp Ser Gln Arg Trp Lys Lys Ser Ile Lys Ala Cys Leu Cys

                  85                    90                    95

Arg Cys Gln Glu Thr Ile Ala Asn Leu Glu Arg Trp Val Lys Arg Glu

                  100                    105                    110

Met His Val Trp Arg Glu Val Phe Tyr Arg Leu Glu Arg Trp Ala Asp

                  115                    120                    125

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

                  130                    135                    140

Arg His Thr Val Ser Val Gly Val Gly Gly Pro Glu Ser Tyr Cys His

145                    150                    155                    160

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

                  165                    170                    175

Pro Pro Ala Ala Gly Glu Leu Pro Gly Gln Glu Pro Ala Glu Ala Gln

                  180                    185                    190

Gln Tyr Gln Pro Trp Val Pro Gly Glu Asp Gly Gln Pro Ser Pro Gly

                  195                    200                    205



Pro Arg Gly Gly Pro Val Ala Lys Pro Asn Val Ile Leu Gln Ile Gly  
 20 25 30  
 Lys Cys Arg Ala Glu Met Leu Glu His Val Arg Arg Thr His Arg His  
 35 40 45  
 Leu Leu Thr Glu Val Ser Lys Gln Val Glu Arg Glu Leu Lys Gly Leu  
 50 55 60  
 His Arg Ser Val Gly Lys Leu Glu Ser Asn Leu Asp Gly Tyr Val Pro  
 65 70 75 80  
  
 Thr Gly Asp Ser Gln Arg Trp Arg Lys Ser Ile Lys Ala Cys Leu Cys  
 85 90 95  
 Arg Cys Gln Glu Thr Ile Ala Asn Leu Glu Arg Trp Val Lys Arg Glu  
 100 105 110  
 Met His Val Trp Arg Glu Val Phe Tyr Arg Leu Glu Arg Trp Ala Asp  
 115 120 125  
 Arg Leu Glu Ser Met Gly Gly Lys Tyr Pro Val Gly Ser Asn Pro Ser  
 130 135 140  
  
 Arg His Thr Thr Ser Val Gly Val Gly Gly Pro Glu Ser Tyr Gly His  
 145 150 155 160  
 Glu Ala Asp Thr Tyr Asp Tyr Thr Val Ser Pro Tyr Ala Ile Thr Pro  
 165 170 175  
 Pro Pro Ala Ala Gly Glu Leu Pro Gly Gln Glu Ala Val Glu Ala Gln  
 180 185 190  
 Gln Tyr Pro Pro Trp Gly Leu Gly Glu Asp Gly Gln Pro Ser Pro Gly  
 195 200 205  
  
 Val Asp Thr Gln Ile Phe Glu Asp Pro Arg Glu Phe Leu Ser His Leu  
 210 215 220  
 Glu Glu Tyr Leu Arg Gln Val Gly Gly Ser Glu Glu Tyr Trp Leu Ser  
 225 230 235 240  
 Gln Ile Gln Asn His Met Asn Gly Pro Ala Lys Lys Trp Trp Glu Tyr  
 245 250 255  
 Lys Gln Gly Ser Val Lys Asn Trp Val Glu Phe Lys Lys Glu Phe Leu





Ile Ile Gln Tyr Val Val Gly Thr Leu Gln Pro Lys Leu Lys Arg Phe  
 325 330 335

Leu Arg Pro Pro Leu Pro Lys Thr Leu Glu Gln Leu Ile Gln Lys Gly  
 340 345 350

Met Glu Val Gln Asp Gly Leu Glu Gln Ala Ala Glu Pro Ala Ala Glu  
 355 360 365

Glu Ala Glu Ala Leu Thr Pro Ala Leu Thr Asn Glu Ser Val Ala Ser  
 370 375 380

Asp Arg Thr Gln Pro Glu  
 385 390

<210> 4  
 <211> 401  
 <212> PRT  
 <213> *Ornithorhynchus anatinus*  
 <400> 4

Gly Glu Leu Asp Arg Leu Asn Pro Ser Ser Gly Leu His Pro Ser Ser  
 1 5 10 15

Gly Leu His Pro Tyr Pro Gly Leu Arg Gly Gly Ala Thr Ala Lys Pro  
 20 25 30

Asn Val Ile Leu Gln Ile Gly Lys Cys Arg Ala Glu Met Leu Glu His  
 35 40 45

Val Arg Lys Thr His Arg His Leu Leu Thr Glu Val Ser Arg Gln Val  
 50 55 60

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

Asn Leu Asp Gly Tyr Val Pro Ser Ser Asp Ser Gln Arg Trp Lys Lys  
 85 90 95

Ser Ile Lys Ala Cys Leu Ser Arg Cys Gln Glu Thr Ile Ala His Leu  
 100 105 110

Glu Arg Trp Val Lys Arg Glu Met Asn Val Trp Arg Glu Val Phe Tyr  
 115 120 125

Arg Leu Glu Arg Trp Ala Asp Arg Leu Glu Ala Met Gly Gly Lys Tyr

130                      135                      140  
 Pro Ala Gly Glu Gln Ala Arg Arg Thr Val Ser Val Gly Val Gly Gly  
 145                      150                      155                      160  
 Pro Glu Thr Cys Cys Pro Gly Asp Glu Ser Tyr Asp Cys Pro Ile Ser  
  
 165                      170                      175  
 Pro Tyr Ala Val Pro Pro Ser Thr Gly Glu Ser Pro Glu Ser Leu Asp  
 180                      185                      190  
 Gln Gly Asp Gln His Tyr Gln Gln Trp Phe Ala Leu Pro Glu Glu Ser  
 195                      200                      205  
 Pro Val Ser Pro Gly Val Asp Thr Gln Ile Phe Glu Asp Pro Arg Glu  
 210                      215                      220  
 Phe Leu Arg His Leu Glu Lys Tyr Leu Lys Gln Val Gly Gly Thr Glu  
  
 225                      230                      235                      240  
 Glu Asp Trp Leu Ser Gln Ile Gln Asn His Met Asn Gly Pro Ala Lys  
 245                      250                      255  
 Lys Trp Trp Glu Tyr Lys Gln Gly Ser Val Lys Asn Trp Leu Glu Phe  
 260                      265                      270  
 Lys Lys Glu Phe Leu Gln Tyr Ser Glu Gly Thr Leu Thr Arg Asp Ala  
 275                      280                      285  
 Leu Lys Arg Glu Leu Asp Leu Pro Gln Lys Gln Gly Glu Pro Leu Asp  
  
 290                      295                      300  
 Gln Phe Leu Trp Arg Lys Arg Asp Leu Tyr Gln Thr Leu Tyr Val Asp  
 305                      310                      315                      320  
 Ala Asp Glu Glu Glu Ile Ile Gln Tyr Val Val Gly Thr Leu Gln Pro  
 325                      330                      335  
 Lys Leu Lys Arg Phe Leu His His Pro Leu Pro Lys Thr Leu Glu Gln  
 340                      345                      350  
 Leu Ile Gln Arg Gly Gln Glu Val Gln Asn Gly Leu Glu Pro Thr Asp  
  
 355                      360                      365  
 Asp Pro Ala Gly Gln Arg Thr Gln Ser Glu Asp Asn Asp Glu Ser Leu  
 370                      375                      380

Thr Pro Ala Val Thr Asn Glu Ser Thr Ala Ser Glu Gly Thr Leu Pro  
 385                      390                      395                      400  
 Glu

<210> 5

<211> 404

<212> PRT

<213> Anser cygnoides domesticus

<400> 5

Gly Gln Leu Asp Asn Val Thr Asn Ala Gly Ile His Ser Phe Gln Gly  
 1                      5                      10                      15

His Arg Gly Val Ala Asn Lys Pro Asn Val Ile Leu Gln Ile Gly Lys  
                     20                      25                      30

Cys Arg Ala Glu Met Leu Glu His Val Arg Arg Thr His Arg His Leu  
                     35                      40                      45

Leu Ser Glu Val Ser Lys Gln Val Glu Arg Glu Leu Lys Gly Leu Gln  
                     50                      55                      60

Lys Ser Val Gly Lys Leu Glu Asn Asn Leu Glu Asp His Val Pro Thr  
 65                      70                      75                      80

Asp Asn Gln Arg Trp Lys Lys Ser Ile Lys Ala Cys Leu Ala Arg Cys  
                     85                      90                      95

Gln Glu Thr Ile Ala His Leu Glu Arg Trp Val Lys Arg Glu Met Asn  
                     100                      105                      110

Val Trp Lys Glu Val Phe Phe Arg Leu Glu Lys Trp Ala Asp Arg Leu  
                     115                      120                      125

Glu Ser Met Gly Gly Lys Tyr Cys Pro Gly Glu His Gly Lys Gln Thr  
                     130                      135                      140

Val Ser Val Gly Val Gly Gly Pro Glu Ile Arg Pro Ser Glu Gly Glu  
 145                      150                      155                      160

Ile Tyr Asp Tyr Ala Leu Asp Met Ser Gln Met Tyr Ala Leu Thr Pro  
                     165                      170                      175

Pro Pro Gly Glu Met Pro Ser Ile Pro Gln Ala His Asp Ser Tyr Gln

180 185 190  
 Trp Val Ser Val Ser Glu Asp Ala Pro Ala Ser Pro Val Glu Thr Gln  
 195 200 205

Val Phe Glu Asp Pro Arg Glu Phe Leu Ser His Leu Glu Glu Tyr Leu  
 210 215 220

Lys Gln Val Gly Gly Thr Glu Glu Tyr Trp Leu Ser Gln Ile Gln Asn  
 225 230 235 240

His Met Asn Gly Pro Ala Lys Lys Trp Trp Glu Tyr Lys Gln Asp Ser  
 245 250 255

Val Lys Asn Trp Val Glu Phe Lys Lys Glu Phe Leu Gln Tyr Ser Glu  
 260 265 270

Gly Thr Leu Thr Arg Asp Ala Ile Lys Arg Glu Leu Asp Leu Pro Gln  
 275 280 285

Lys Glu Gly Glu Pro Leu Asp Gln Phe Leu Trp Arg Lys Arg Asp Leu  
 290 295 300

Tyr Gln Thr Leu Tyr Val Asp Ala Asp Glu Glu Glu Ile Ile Gln Tyr  
 305 310 315 320

Val Val Gly Thr Leu Gln Pro Lys Leu Lys Arg Phe Leu Ser Tyr Pro  
 325 330 335

Leu Pro Lys Thr Leu Glu Gln Leu Ile Gln Arg Gly Lys Glu Val Gln  
 340 345 350

Gly Asn Met Asp His Ser Asp Glu Pro Ser Pro Gln Arg Thr Pro Glu  
 355 360 365

Ile Gln Ser Gly Asp Ser Val Glu Ser Met Pro Pro Ser Thr Thr Ala  
 370 375 380

Ser Pro Val Pro Ser Asn Gly Thr Gln Pro Glu Pro Pro Ser Pro Pro  
 385 390 395 400

Ala Thr Val Ile

<210> 6

<211> 395

<212> PRT

<213> *Pelecanus crispus*

<400> 6

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

225                    230                    235                    240  
 His Met Asn Gly Pro Ala Lys Lys Trp Trp Glu Tyr Lys Gln Asp Ser

                         245                    250                    255  
 Val Lys Asn Trp Val Glu Phe Lys Lys Glu Phe Leu Gln Tyr Ser Glu

                         260                    265                    270  
 Gly Thr Leu Thr Arg Asp Ala Ile Lys Arg Glu Leu Asp Leu Pro Gln

                         275                    280                    285  
 Lys Glu Gly Glu Pro Leu Asp Gln Phe Leu Trp Arg Lys Arg Asp Leu

                         290                    295                    300  
 Tyr Gln Thr Leu Tyr Val Asp Ala Asp Glu Glu Glu Ile Ile Gln Tyr

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

                         325                    330                    335  
 Leu Pro Lys Thr Leu Glu Gln Leu Ile Gln Arg Gly Lys Glu Val Gln

                         340                    345                    350  
 Gly Asn Met Asp His Ser Glu Glu Pro Ser Pro Gln Arg Thr Pro Glu

                         355                    360                    365  
 Ile Gln Ser Gly Asp Ser Val Asp Ser Val Pro Pro Ser Thr Thr Ala

                         370                    375                    380  
 Ser Pro Val Pro Ser Asn Gly Thr Gln Pro Glu

385                    390                    395

<210> 7

<211> 395

<212> PRT

<213> *Haliaeetus albicilla*

<400> 7

Gly Gln Leu Asp Asn Val Thr Asn Ala Gly Ile His Ser Phe Gln Gly  
 1                    5                    10                    15

His Arg Gly Val Ala Asn Lys Pro Asn Val Ile Leu Gln Ile Gly Lys  
                          20                    25                    30

Cys Arg Ala Glu Met Leu Glu His Val Arg Arg Thr His Arg His Leu



Lys Glu Gly Glu Pro Leu Asp Gln Phe Leu Trp Arg Lys Arg Asp Leu

290 295 300

Tyr Gln Thr Leu Tyr Val Asp Ala Asp Glu Glu Glu Ile Ile Gln Tyr

305 310 315 320

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

325 330 335

Leu Pro Lys Thr Leu Glu Gln Leu Ile Gln Arg Gly Lys Glu Val Gln

340 345 350

Gly Asn Met Asp His Ser Glu Glu Pro Ser Pro Gln Arg Thr Pro Glu

355 360 365

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

370 375 380

Ser Pro Val Pro Ser Asn Gly Thr Gln Pro Glu

385 390 395

<210> 8

<211> 465

<212> PRT

<213> Ophiophagus hannah

<400> 8

Gly Ser Trp Gly Leu Gln Arg His Val Ala Asp Glu Arg Arg Gly Leu

1 5 10 15

Ala Thr Pro Thr Tyr Gly Ala Val Cys Ser Ile Arg Glu Lys Lys Ala

20 25 30

Ser Gln Leu Ser Gly Gln Ser Cys Leu Glu Lys Glu Leu Leu Gly Trp

35 40 45

Lys Cys Thr Glu Ala Ile Val Glu Met Met Gln Val Asp Asn Phe Asn

50 55 60

His Gly Asn Leu His Ser Cys Gln Gly His Arg Gly Met Ala Asn His

65 70 75 80

Lys Pro Asn Val Ile Leu Gln Ile Gly Lys Cys Arg Ala Glu Met Leu

85 90 95

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



Val Glu Tyr Leu Arg Ala Gln Phe Asp Ala Ser Leu Gly Trp Arg Leu  
 85 90 95

Lys Glu Gln His Ala Asp Ile Thr Thr Lys Ile Ile Pro Glu Ile Ile  
 100 105 110

Gln Thr Val Lys Glu Asp Ile Ser Leu Cys Leu Ser Thr Leu Cys Ser  
 115 120 125

Ile Ala Glu Asp Ile Gln Thr Ser Arg Ala Thr Thr Val Thr Gly His  
 130 135 140

Ala Ala Val Gln Thr His Pro Val Asp Leu Leu Gly Glu His His Leu  
 145 150 155 160

Gly Thr Thr Gly His Pro Arg Leu Gln Ser Thr Arg Val Gly Lys Pro  
 165 170 175

Asp Asp Val Pro Glu Ser Pro Val Ser Leu Phe Met Gln Gly Glu Ala  
 180 185 190

Arg Ser Arg Ile Val Gly Lys Ser Pro Ile Lys Leu Gln Phe Pro Thr  
 195 200 205

Phe Gly Lys Ala Asn Asp Ser Ser Asp Pro Leu Gln Tyr Leu Glu Arg  
 210 215 220

Cys Glu Asp Phe Leu Ala Leu Asn Pro Leu Thr Asp Glu Glu Leu Met  
 225 230 235 240

Ala Thr Leu Arg Asn Val Leu His Gly Thr Ser Arg Asp Trp Trp Asp  
 245 250 255

Val Ala Arg His Lys Ile Gln Thr Trp Arg Glu Phe Asn Lys His Phe  
 260 265 270

Arg Ala Ala Phe Leu Ser Glu Asp Tyr Glu Asp Glu Leu Ala Glu Arg  
 275 280 285

Val Arg Asn Arg Ile Gln Lys Glu Asp Glu Ser Ile Arg Asp Phe Ala  
 290 295 300

Tyr Met Tyr Gln Ser Leu Cys Lys Arg Trp Asn Pro Ala Ile Cys Glu  
 305 310 315 320

Gly Asp Val Val Lys Leu Ile Leu Lys Asn Ile Asn Pro Gln Leu Pro



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



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

385                      390                      395                      400

Ala Thr Val Ile

<210> 12

<211> 409

<212> PRT

<213> *Pogona vitticeps*

<400> 12

Gly Gln Leu Glu Asn Ile Asn Gln Gly Ser Leu His Ala Phe Gln Gly

1                      5                      10                      15

His Arg Gly Val Val His Asn Asn Lys Pro Asn Val Ile Leu Gln Ile

                    20                      25                      30

Gly Lys Cys Arg Ala Glu Met Leu Glu His Val Arg Arg Thr His Arg

                    35                      40                      45

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

                    50                      55                      60

Leu Gln Lys Ser Val Gly Lys Leu Glu Asn Asn Leu Glu Asp His Val

65                      70                      75                      80

Pro Ser Ala Ala Glu Asn Gln Arg Trp Lys Lys Ser Ile Lys Ala Cys

                    85                      90                      95

Leu Ala Arg Cys Gln Glu Thr Ile Ala Asn Leu Glu Arg Trp Val Lys

                    100                      105                      110

Arg Glu Met Asn Val Trp Lys Glu Val Phe Phe Arg Leu Glu Arg Trp

                    115                      120                      125

Ala Asp Arg Leu Glu Ser Gly Gly Gly Lys Tyr Cys His Ala Asp Gln

                    130                      135                      140

Gly Arg Gln Thr Val Ser Val Gly Val Gly Gly Pro Glu Val Arg Pro

145                      150                      155                      160

Ser Glu Gly Glu Ile Tyr Asp Tyr Ala Leu Asp Met Ser Gln Met Tyr

                    165                      170                      175

Ala Leu Thr Pro Pro Pro Met Gly Asp Val Pro Val Ile Pro Gln Pro

                    180                      185                      190

His Asp Ser Tyr Gln Trp Val Thr Asp Pro Glu Glu Ala Pro Pro Ser  
 195 200 205

Pro Val Glu Thr Gln Ile Phe Glu Asp Pro Arg Glu Phe Leu Thr His  
 210 215 220

Leu Glu Asp Tyr Leu Lys Gln Val Gly Gly Thr Glu Glu Tyr Trp Leu  
 225 230 235 240

Ser Gln Ile Gln Asn His Met Asn Gly Pro Ala Lys Lys Trp Trp Glu  
 245 250 255

Tyr Lys Gln Asp Ser Val Lys Asn Trp Leu Glu Phe Lys Lys Glu Phe  
 260 265 270

Leu Gln Tyr Ser Glu Gly Thr Leu Thr Arg Asp Ala Ile Lys Gln Glu  
 275 280 285

Leu Asp Leu Pro Gln Lys Glu Gly Glu Pro Leu Asp Gln Phe Leu Trp  
 290 295 300

Arg Lys Arg Asp Leu Tyr Gln Thr Leu Tyr Val Glu Ala Glu Glu Glu  
 305 310 315 320

Glu Val Ile Gln Tyr Val Val Gly Thr Leu Gln Pro Lys Leu Lys Arg  
 325 330 335

Phe Leu Ser His Pro Tyr Pro Lys Thr Leu Glu Gln Leu Ile Gln Arg  
 340 345 350

Gly Lys Glu Val Glu Gly Asn Leu Asp Asn Ser Glu Glu Pro Ser Pro  
 355 360 365

Gln Arg Thr Pro Glu His Gln Leu Gly Asp Ser Val Glu Ser Leu Pro  
 370 375 380

Pro Ser Thr Thr Ala Ser Pro Ala Gly Ser Asp Lys Thr Gln Pro Glu  
 385 390 395 400

Ile Ser Leu Pro Pro Thr Thr Val Ile  
 405

<210> 13

<211> 404

<212> PRT

<213> Alligator sinensis

<400> 13

Gly Gln Leu Asp Ser Val Thr Asn Ala Gly Val His Thr Tyr Gln Gly

1                    5                    10                    15

His Arg Ser Val Ala Asn Lys Pro Asn Val Ile Leu Gln Ile Gly Lys

                  20                    25                    30

Cys Arg Thr Glu Met Leu Glu His Val Arg Arg Thr His Arg His Leu

                  35                    40                    45

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

                  50                    55                    60

Lys Ser Val Gly Lys Leu Glu Asn Asn Leu Glu Asp His Val Pro Thr

65                    70                    75                    80

Asp Asn Gln Arg Trp Lys Lys Ser Ile Lys Ala Cys Leu Ala Arg Cys

                  85                    90                    95

Gln Glu Thr Ile Ala His Leu Glu Arg Trp Val Lys Arg Glu Met Asn

                  100                    105                    110

Val Trp Lys Glu Val Phe Phe Arg Leu Glu Arg Trp Ala Asp Arg Leu

                  115                    120                    125

Glu Ser Met Gly Gly Lys Tyr Cys Pro Thr Asp Ser Ala Arg Gln Thr

                  130                    135                    140

Val Ser Val Gly Val Gly Gly Pro Glu Ile Arg Pro Ser Glu Gly Glu

145                    150                    155                    160

Ile Tyr Asp Tyr Ala Leu Asp Met Ser Gln Met Tyr Ala Leu Thr Pro

                  165                    170                    175

Ser Pro Gly Glu Leu Pro Ser Val Pro Gln Pro His Asp Ser Tyr Gln

                  180                    185                    190

Trp Val Thr Ser Pro Glu Asp Ala Pro Ala Ser Pro Val Glu Thr Gln

                  195                    200                    205

Val Phe Glu Asp Pro Arg Glu Phe Leu Cys His Leu Glu Glu Tyr Leu

                  210                    215                    220

Lys Gln Val Gly Gly Thr Glu Glu Tyr Trp Leu Ser Gln Ile Gln Asn



20 25 30  
 Cys Arg Thr Glu Met Leu Glu His Val Arg Arg Thr His Arg His Leu  
 35 40 45  
 Leu Thr Glu Val Ser Lys Gln Val Glu Arg Glu Leu Lys Gly Leu Gln  
 50 55 60  
 Lys Ser Val Gly Lys Leu Glu Asn Asn Leu Glu Asp His Val Pro Thr  
 65 70 75 80  
  
 Asp Asn Gln Arg Trp Lys Lys Ser Ile Lys Ala Cys Leu Ala Arg Cys  
 85 90 95  
 Gln Glu Thr Ile Ala His Leu Glu Arg Trp Val Lys Arg Glu Met Asn  
 100 105 110  
 Val Trp Lys Glu Val Phe Phe Arg Leu Glu Arg Trp Ala Asp Arg Leu  
 115 120 125  
 Glu Ser Met Gly Gly Lys Tyr Cys Pro Thr Asp Ser Ala Arg Gln Thr  
 130 135 140  
  
 Val Ser Val Gly Val Gly Gly Pro Glu Ile Arg Pro Ser Glu Gly Glu  
 145 150 155 160  
 Ile Tyr Asp Tyr Ala Leu Asp Met Ser Gln Met Tyr Ala Leu Thr Pro  
 165 170 175  
 Ser Pro Gly Glu Leu Pro Ser Ile Pro Gln Pro His Asp Ser Tyr Gln  
 180 185 190  
 Trp Val Thr Ser Pro Glu Asp Ala Pro Ala Ser Pro Val Glu Thr Gln  
 195 200 205  
  
 Val Phe Glu Asp Pro Arg Glu Phe Leu Cys His Leu Glu Glu Tyr Leu  
 210 215 220  
 Lys Gln Val Gly Gly Thr Glu Glu Tyr Trp Leu Ser Gln Ile Gln Asn  
 225 230 235 240  
 His Met Asn Gly Pro Ala Lys Lys Trp Trp Glu Tyr Lys Gln Asp Thr  
 245 250 255  
 Val Lys Asn Trp Val Glu Phe Lys Lys Glu Phe Leu Gln Tyr Ser Glu  
 260 265 270

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

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

Ala Thr Val Ile

<210> 15

<211> 408

<212> PRT

<213> Gekko japonicus

<400> 15

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



Val Ile Gln Tyr Val Val Gly Thr Leu Gln Pro Lys Leu Lys Arg Phe  
 325 330 335

Leu Ser His Pro Tyr Pro Lys Thr Leu Glu Gln Leu Ile Gln Arg Gly  
 340 345 350

Lys Glu Val Glu Gly Asn Leu Asp Asn Ser Glu Glu Pro Thr Pro Gln  
 355 360 365

Arg Thr Pro Glu His Gln Leu Cys Gly Ser Val Glu Ser Leu Pro Pro  
 370 375 380

Ser Ser Thr Val Ser Pro Val Ala Ser Asp Gly Thr Gln Pro Glu Thr  
 385 390 395 400

Ser Pro Leu Pro Ala Thr Val Ile  
 405

<210> 16  
 <211> 455  
 <212> PRT  
 <213> Homo sapiens  
 <400> 16

Gly Pro Leu Thr Leu Leu Gln Asp Trp Cys Arg Gly Glu His Leu Asn  
 1 5 10 15

Thr Arg Arg Cys Met Leu Ile Leu Gly Ile Pro Glu Asp Cys Gly Glu  
 20 25 30

Asp Glu Phe Glu Glu Thr Leu Gln Glu Ala Cys Arg His Leu Gly Arg  
 35 40 45

Tyr Arg Val Ile Gly Arg Met Phe Arg Arg Glu Glu Asn Ala Gln Ala  
 50 55 60

Ile Leu Leu Glu Leu Ala Gln Asp Ile Asp Tyr Ala Leu Leu Pro Arg  
 65 70 75 80

Glu Ile Pro Gly Lys Gly Gly Pro Trp Glu Val Ile Val Lys Pro Arg  
 85 90 95

Asn Ser Asp Gly Glu Phe Leu Asn Arg Leu Asn Arg Phe Leu Glu Glu  
 100 105 110

Glu Arg Arg Thr Val Ser Asp Met Asn Arg Val Leu Gly Ser Asp Thr



Pro Ala Ser Gly Asn Ser Phe Asp Ala Arg Pro Ser Gln Gly Tyr Arg  
 370 375 380  
 Arg Arg Arg Gly Arg Gly Gln His Arg Arg Gly Gly Val Ala Arg Ala  
 385 390 395 400  
 Gly Ser Arg Gly Ser Arg Lys Arg Lys Arg His Thr Phe Cys Tyr Ser  
 405 410 415  
 Cys Gly Glu Asp Gly His Ile Arg Val Gln Cys Ile Asn Pro Ser Asn  
 420 425 430  
 Leu Leu Leu Ala Lys Glu Thr Lys Glu Ile Leu Glu Gly Gly Glu Arg  
 435 440 445  
 Glu Ala Gln Thr Asn Ser Arg  
 450 455  
 <210> 17  
 <211> 448  
 <212> PRT  
 <213> Homo sapiens  
 <400> 17  
 Gly Ala Leu Thr Leu Leu Glu Asp Trp Cys Lys Gly Met Asp Met Asp  
 1 5 10 15  
 Pro Arg Lys Ala Leu Leu Ile Val Gly Ile Pro Met Glu Cys Ser Glu  
 20 25 30  
 Val Glu Ile Gln Asp Thr Val Lys Ala Gly Leu Gln Pro Leu Cys Ala  
 35 40 45  
 Tyr Arg Val Leu Gly Arg Met Phe Arg Arg Glu Asp Asn Ala Lys Ala  
 50 55 60  
 Val Phe Ile Glu Leu Ala Asp Thr Val Asn Tyr Thr Thr Leu Pro Ser  
 65 70 75 80  
 His Ile Pro Gly Lys Gly Gly Ser Trp Glu Val Val Val Lys Pro Arg  
 85 90 95  
 Asn Pro Asp Asp Glu Phe Leu Ser Arg Leu Asn Tyr Phe Leu Lys Asp  
 100 105 110  
 Glu Gly Arg Ser Met Thr Asp Val Ala Arg Ala Leu Gly Cys Cys Ser



Val Gln Ala Arg Ser Phe Ser Asp Ser Ser Pro Gln Thr Ile Gln Gly  
 370 375 380

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

385 390 395 400

Thr Arg Gly Glu Asp His Gly Gln Ala Thr Tyr Pro Lys Ala Glu Asn  
 405 410 415

Gln Thr Pro Gly Arg Glu Gly Pro Gln Ala Ala Gly Glu Glu Leu Gly  
 420 425 430

Asn Glu Ala Gly Ala Gly Ala Met Ser His Pro Lys Pro Trp Glu Thr  
 435 440 445

<210> 18

<211> 399

<212> PRT

<213> Homo sapiens

<400> 18

Gly Ala Val Thr Met Leu Gln Asp Trp Cys Arg Trp Met Gly Val Asn  
 1 5 10 15

Ala Arg Arg Gly Leu Leu Ile Leu Gly Ile Pro Glu Asp Cys Asp Asp  
 20 25 30

Ala Glu Phe Gln Glu Ser Leu Glu Ala Ala Leu Arg Pro Met Gly His  
 35 40 45

Phe Thr Val Leu Gly Lys Ala Phe Arg Glu Glu Asp Asn Ala Thr Ala  
 50 55 60

Ala Leu Val Glu Leu Asp Arg Glu Val Asn Tyr Ala Leu Val Pro Arg  
 65 70 75 80

Glu Ile Pro Gly Thr Gly Gly Pro Trp Asn Val Val Phe Val Pro Arg  
 85 90 95

Cys Ser Gly Glu Glu Phe Leu Gly Leu Gly Arg Val Phe His Phe Pro  
 100 105 110

Glu Gln Glu Gly Gln Met Val Glu Ser Val Ala Gly Ala Leu Gly Val  
 115 120 125

Gly Leu Arg Arg Val Cys Trp Leu Arg Ser Ile Gly Gln Ala Val Gln  
 130 135 140

Pro Trp Val Glu Ala Val Arg Cys Gln Ser Leu Gly Val Phe Ser Gly  
 145 150 155 160

Arg Asp Gln Pro Ala Pro Gly Glu Glu Ser Phe Glu Val Trp Leu Asp  
 165 170 175

His Thr Thr Glu Met Leu His Val Trp Gln Gly Val Ser Glu Arg Glu  
 180 185 190

Arg Arg Arg Arg Leu Leu Glu Gly Leu Arg Gly Thr Ala Leu Gln Leu  
 195 200 205

Val His Ala Leu Leu Ala Glu Asn Pro Ala Arg Thr Ala Gln Asp Cys  
 210 215 220

Leu Ala Ala Leu Ala Gln Val Phe Gly Asp Asn Glu Ser Gln Ala Thr  
 225 230 235 240

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

Ser Ala Phe Val Leu Arg Leu Glu Val Leu Leu Gln Lys Ala Met Glu  
 260 265 270

Lys Glu Ala Leu Ala Arg Ala Ser Ala Asp Arg Val Arg Leu Arg Gln  
 275 280 285

Met Leu Thr Arg Ala His Leu Thr Glu Pro Leu Asp Glu Ala Leu Arg  
 290 295 300

Lys Leu Arg Met Ala Gly Arg Ser Pro Ser Phe Leu Glu Met Leu Gly  
 305 310 315 320

Leu Val Arg Glu Ser Glu Ala Trp Glu Ala Ser Leu Ala Arg Ser Val  
 325 330 335

Arg Ala Gln Thr Gln Glu Gly Ala Gly Ala Arg Ala Gly Ala Gln Ala  
 340 345 350

Val Ala Arg Ala Ser Thr Lys Val Glu Ala Val Pro Gly Gly Pro Gly  
 355 360 365

Arg Glu Pro Glu Gly Leu Leu Gln Ala Gly Gly Gln Glu Ala Glu Glu

370 375 380

Leu Leu Gln Glu Gly Leu Lys Pro Val Leu Glu Glu Cys Asp Asn

385 390 395

<210> 19

<211> 399

<212> PRT

<213> Homo sapiens

<400> 19

Gly Ala Val Thr Met Leu Gln Asp Trp Cys Arg Trp Met Gly Val Asn

1 5 10 15

Ala Arg Arg Gly Leu Leu Ile Leu Gly Ile Pro Glu Asp Cys Asp Asp

20 25 30

Ala Glu Phe Gln Glu Ser Leu Glu Ala Ala Leu Arg Pro Met Gly His

35 40 45

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

50 55 60

Ala Leu Val Glu Leu Asp Arg Glu Val Asn Tyr Ala Leu Val Pro Arg

65 70 75 80

Glu Ile Pro Gly Thr Gly Gly Pro Trp Asn Val Val Phe Val Pro Arg

85 90 95

Cys Ser Gly Glu Glu Phe Leu Gly Leu Gly Arg Val Phe His Phe Pro

100 105 110

Glu Gln Glu Gly Gln Met Val Glu Ser Val Ala Gly Ala Leu Gly Val

115 120 125

Gly Leu Arg Arg Val Cys Trp Leu Arg Ser Ile Gly Gln Ala Val Gln

130 135 140

Pro Trp Val Glu Ala Val Arg Tyr Gln Ser Leu Gly Val Phe Ser Gly

145 150 155 160

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

165 170 175

His Thr Thr Glu Met Leu His Val Trp Gln Gly Val Ser Glu Arg Glu



Gly Val Glu Asp Leu Ala Ala Ser Tyr Ile Val Leu Lys Leu Glu Asn  
 1                    5                    10                    15  
  
 Glu Ile Arg Gln Ala Gln Val Gln Trp Leu Met Glu Glu Asn Ala Ala  
                   20                    25                    30  
 Leu Gln Ala Gln Ile Pro Glu Leu Gln Lys Ser Gln Ala Ala Lys Glu  
                   35                    40                    45  
 Tyr Asp Leu Leu Arg Lys Ser Ser Glu Ala Lys Glu Pro Gln Lys Leu  
                   50                    55                    60  
 Pro Glu His Met Asn Pro Pro Ala Ala Trp Glu Ala Gln Lys Thr Pro  
 65                    70                    75                    80  
  
 Glu Phe Lys Glu Pro Gln Lys Pro Pro Glu Pro Gln Asp Leu Leu Pro  
                   85                    90                    95  
 Trp Glu Pro Pro Ala Ala Trp Glu Leu Gln Glu Ala Pro Ala Ala Pro  
                   100                    105                    110  
 Glu Ser Leu Ala Pro Pro Ala Thr Arg Glu Ser Gln Lys Pro Pro Met  
                   115                    120                    125  
 Ala His Glu Ile Pro Thr Val Leu Glu Gly Gln Gly Pro Ala Asn Thr  
                   130                    135                    140  
  
 Gln Asp Ala Thr Ile Ala Gln Glu Pro Lys Asn Ser Glu Pro Gln Asp  
 145                    150                    155                    160  
 Pro Pro Asn Ile Glu Lys Pro Gln Glu Ala Pro Glu Tyr Gln Glu Thr  
                   165                    170                    175  
 Ala Ala Gln Leu Glu Phe Leu Glu Leu Pro Pro Pro Gln Glu Pro Leu  
                   180                    185                    190  
 Glu Pro Ser Asn Ala Gln Glu Phe Leu Glu Leu Ser Ala Ala Gln Glu  
                   195                    200                    205  
  
 Ser Leu Glu Gly Leu Ile Val Val Glu Thr Ser Ala Ala Ser Glu Phe  
                   210                    215                    220  
 Pro Gln Ala Pro Ile Gly Leu Glu Ala Thr Asp Phe Pro Leu Gln Tyr  
 225                    230                    235                    240  
 Thr Leu Thr Phe Ser Gly Asp Ser Gln Lys Leu Pro Glu Phe Leu Val

245 250 255  
 Gln Leu Tyr Ser Tyr Met Arg Val Arg Gly His Leu Tyr Pro Thr Glu  
 260 265 270

Ala Ala Leu Val Ser Phe Val Gly Asn Cys Phe Ser Gly Arg Ala Gly  
 275 280 285

Trp Trp Phe Gln Leu Leu Leu Asp Ile Gln Ser Pro Leu Leu Glu Gln  
 290 295 300

Cys Glu Ser Phe Ile Pro Val Leu Gln Asp Thr Phe Asp Asn Pro Glu  
 305 310 315 320

Asn Met Lys Asp Ala Asn Gln Cys Ile His Gln Leu Cys Gln Gly Glu  
 325 330 335

Gly His Val Ala Thr His Phe His Leu Ile Ala Gln Glu Leu Asn Trp  
 340 345 350

Asp Glu Ser Thr Leu Trp Ile Gln Phe Gln Glu Gly Leu Ala Ser Ser  
 355 360 365

Ile Gln Asp Glu Leu Ser His Thr Ser Pro Ala Thr Asn Leu Ser Asp  
 370 375 380

Leu Ile Thr Gln Cys Ile Ser Leu Glu Glu Lys Pro Asp Pro Asn Pro  
 385 390 395 400

Leu Gly Lys Ser Ser Ser Ala Glu Gly Asp Gly Pro Glu Ser Pro Pro  
 405 410 415

Ala Glu Asn Gln Pro Met Gln Ala Ala Ile Asn Cys Pro His Ile Ser  
 420 425 430

Glu Ala Glu Trp Val Arg Trp His Lys Gly Arg Leu Cys Leu Tyr Cys  
 435 440 445

Gly Tyr Pro Gly His Phe Ala Arg Asp Cys Pro Val Lys Pro His Gln  
 450 455 460

Ala Leu Gln Ala Gly Asn Ile Gln Ala Cys Gln  
 465 470 475

<210> 21  
 <211> 239

<212> PRT

<213> Homo sapiens

<400> 21

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

1                    5                    10                    15

Ser Pro Asn Ala Gln Met Asp Asp Val Ile Asp Thr Leu Thr Ser Leu

                  20                    25                    30

Arg Leu Thr Asn Ser Ala Leu Arg Arg Glu Ala Ser Thr Leu Arg Ala

                  35                    40                    45

Glu Lys Ala Asn Leu Thr Asn Met Leu Glu Ser Val Met Ala Glu Leu

                  50                    55                    60

Thr Leu Leu Arg Thr Arg Ala Arg Ile Pro Gly Ala Leu Gln Ile Thr

65                    70                    75                    80

Pro Pro Ile Ser Ser Ile Thr Ser Asn Gly Thr Arg Pro Met Thr Thr

                  85                    90                    95

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

                  100                    105                    110

Ala Gly Phe Leu Met Gln Met Asp Arg Phe Met Ile Phe Gln Ala Ser

                  115                    120                    125

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

                  130                    135                    140

Thr Gly Glu Ala Glu Lys Trp Ala Ile Pro His Met Gln Pro Asp Ser

145                    150                    155                    160

Pro Leu Arg Asn Asn Tyr Gln Gly Phe Leu Ala Glu Leu Arg Arg Thr

                  165                    170                    175

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

                  180                    185                    190

Ser Ala Ser Asn Arg Ala Val Arg Glu Arg Gln Met Leu Cys Arg Gln

                  195                    200                    205

Leu Ala Ser Ala Gly Thr Gly Pro Cys Pro Val His Pro Ala Ser Asn

                  210                    215                    220

Gly Thr Ser Pro Ala Pro Ala Leu Pro Ala Arg Ala Arg Asn Leu

225

230

235

<210> 22

<211> 113

<212> PRT

<213> Homo sapiens

<400> 22

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

1                    5                    10                    15

Arg Pro Ala Ala Arg Arg Trp Arg Asn Pro Ile Pro Phe Pro Glu Thr

                  20                    25                    30

Phe Asp Gly Asp Thr Asp Arg Leu Pro Glu Phe Ile Val Gln Thr Ser

                  35                    40                    45

Ser Tyr Met Phe Val Asp Glu Asn Thr Phe Ser Asn Asp Ala Leu Lys

50

55

60

Val Thr Phe Leu Ile Thr Arg Leu Thr Gly Pro Ala Leu Gln Trp Val

65                    70                    75                    80

Ile Pro Tyr Ile Arg Lys Glu Ser Pro Leu Leu Asn Asp Tyr Arg Gly

                  85                    90                    95

Phe Leu Ala Glu Met Lys Arg Val Phe Gly Trp Glu Glu Asp Glu Asp

                  100                    105                    110

Phe

<210> 23

<211> 113

<212> PRT

<213> Homo sapiens

<400> 23

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

1                    5                    10                    15

Arg Pro Ala Ala Arg Arg Trp Arg Asn Pro Ile Pro Phe Pro Glu Thr

                  20                    25                    30

Phe Asp Gly Asp Thr Asp Arg Leu Pro Glu Phe Ile Val Gln Thr Ser

35 40 45  
 Ser Tyr Met Phe Val Asp Glu Asn Thr Phe Ser Asn Asp Ala Leu Lys  
 50 55 60

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

<210> 24

<211> 364

<212> PRT

<213> Homo sapiens

<400> 24

Gly Pro Arg Gly Arg Cys Arg Gln Gln Gly Pro Arg Ile Pro Ile Trp

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

Phe Glu His Tyr Gln Asp Asn Ile Ser Arg Val Cys Glu Ile Leu Arg

130 135 140

Arg Leu Thr Gly Arg Ala Gln Ala Trp Ala Ala Pro Tyr Leu Asp Gly

145 150 155 160

Asp Leu Pro Leu Pro Asp Asp Tyr Glu Leu Phe Cys Gln Asp Leu Lys

165 170 175

Glu Val Val Gln Asp Pro Asn Ser Phe Ala Glu Tyr His Ala Val Val

180 185 190

Thr Cys Pro Leu Pro Leu Ala Ser Ser Gln Leu Pro Val Ala Pro Gln

195 200 205

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

210 215 220

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

225 230 235 240

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

245 250 255

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

260 265 270

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

275 280 285

Ser Ser Gln Pro Glu Glu Ala Ala Pro Thr Pro Val Pro Arg Leu Ser

290 295 300

Glu Ser Ala Asn Pro Pro Ala Gln Arg Pro Asp Pro Ala His Pro Gly

305 310 315 320

Gly Pro Lys Pro Gln Lys Thr Glu Glu Glu Val Leu Glu Thr Glu Gly

325 330 335

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

340 345 350

Glu Thr Pro Gly Glu Pro Pro Leu Ser Pro Gly Phe

355 360

<210> 25

<211> 146

<212> PRT

<213> Homo sapiens

<400> 25

Gly Val Asp Glu Leu Val Leu Leu Leu His Ala Leu Leu Met Arg His

1                    5                    10                    15

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

20                    25                    30

Leu Val Cys Glu Arg Ala Ser Leu Leu Arg Gln Val Arg Pro Pro Ser

35                    40                    45

Cys Pro Val Pro Phe Pro Glu Thr Phe Asn Gly Glu Ser Ser Arg Leu

50                    55                    60

Pro Glu Phe Ile Val Gln Thr Ala Ser Tyr Met Leu Val Asn Glu Asn

65                    70                    75                    80

Arg Phe Cys Asn Asp Ala Met Lys Val Ala Phe Leu Ile Ser Leu Leu

85                    90                    95

Thr Gly Glu Ala Glu Glu Trp Val Val Pro Tyr Ile Glu Met Asp Ser

100                    105                    110

Pro Ile Leu Gly Asp Tyr Arg Ala Phe Leu Asp Glu Met Lys Gln Cys

115                    120                    125

Phe Gly Trp Asp Asp Asp Glu Asp Asp Asp Asp Glu Glu Glu Glu Asp

130                    135                    140

Asp Tyr

145

<210> 26

<211> 549

<212>

> PRT

<213> Homo sapiens

<400> 26

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

1                    5                    10                    15

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



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

515                                      520                                      525  
 Phe Ser Trp Ser Ser Ser Leu Asp Lys His Gln Arg Ser His Leu Gly  
 530                                      535                                      540  
 Lys Lys Pro Phe Gln  
 545  
 <210> 27  
 <211> 351  
 <212> PRT  
 <213> Homo sapiens  
 <400> 27  
 Gly Thr Leu Arg Leu Leu Glu Asp Trp Cys Arg Gly Met Asp Met Asn  
 1                                      5                                      10                                      15  
 Pro Arg Lys Ala Leu Leu Ile Ala Gly Ile Ser Gln Ser Cys Ser Val  
  
 20                                      25                                      30  
 Ala Glu Ile Glu Glu Ala Leu Gln Ala Gly Leu Ala Pro Leu Gly Glu  
 35                                      40                                      45  
 Tyr Arg Leu Leu Gly Arg Met Phe Arg Arg Asp Glu Asn Arg Lys Val  
 50                                      55                                      60  
 Ala Leu Val Gly Leu Thr Ala Glu Thr Ser His Ala Leu Val Pro Lys  
 65                                      70                                      75                                      80  
 Glu Ile Pro Gly Lys Gly Gly Ile Trp Arg Val Ile Phe Lys Pro Pro  
  
 85                                      90                                      95  
 Asp Pro Asp Asn Thr Phe Leu Ser Arg Leu Asn Glu Phe Leu Ala Gly  
 100                                      105                                      110  
 Glu Gly Met Thr Val Gly Glu Leu Ser Arg Ala Leu Gly His Glu Asn  
 115                                      120                                      125  
 Gly Ser Leu Asp Pro Glu Gln Gly Met Ile Pro Glu Met Trp Ala Pro  
 130                                      135                                      140  
 Met Leu Ala Gln Ala Leu Glu Ala Leu Gln Pro Ala Leu Gln Cys Leu  
  
 145                                      150                                      155                                      160  
 Lys Tyr Lys Lys Leu Arg Val Phe Ser Gly Arg Glu Ser Pro Glu Pro  
 165                                      170                                      175

Gly Glu Glu Glu Phe Gly Arg Trp Met Phe His Thr Thr Gln Met Ile  
 180 185 190  
 Lys Ala Trp Gln Val Pro Asp Val Glu Lys Arg Arg Arg Leu Leu Glu  
 195 200 205  
 Ser Leu Arg Gly Pro Ala Leu Asp Val Ile Arg Val Leu Lys Ile Asn  
 210 215 220  
 Asn Pro Leu Ile Thr Val Asp Glu Cys Leu Gln Ala Leu Glu Glu Val  
 225 230 235 240  
 Phe Gly Val Thr Asp Asn Pro Arg Glu Leu Gln Val Lys Tyr Leu Thr  
 245 250 255  
 Thr Tyr His Lys Asp Glu Glu Lys Leu Ser Ala Tyr Val Leu Arg Leu  
 260 265 270  
 Glu Pro Leu Leu Gln Lys Leu Val Gln Arg Gly Ala Ile Glu Arg Asp  
 275 280 285  
 Ala Val Asn Gln Ala Arg Leu Asp Gln Val Ile Ala Gly Ala Val His  
 290 295 300  
 Lys Thr Ile Arg Arg Glu Leu Asn Leu Pro Glu Asp Gly Pro Ala Pro  
 305 310 315 320  
 Gly Phe Leu Gln Leu Leu Val Leu Ile Lys Asp Tyr Glu Ala Ala Glu  
 325 330 335  
 Glu Glu Glu Ala Leu Leu Gln Ala Ile Leu Glu Gly Asn Phe Thr

340 345 350

<210> 28

<211> 708

<212> PRT

<213> Homo sapiens

<400> 28

Gly Thr Glu Arg Arg Arg Asp Glu Leu Ser Glu Glu Ile Asn Asn Leu  
 1 5 10 15  
 Arg Glu Lys Val Met Lys Gln Ser Glu Glu Asn Asn Asn Leu Gln Ser  
 20 25 30  
 Gln Val Gln Lys Leu Thr Glu Glu Asn Thr Thr Leu Arg Glu Gln Val



Arg Arg Lys Leu Asn Leu Cys Leu Tyr Cys Gly Thr Gly Gly His Tyr  
 290 295 300

Ala Asp Asn Cys Pro Ala Lys Ala Ser Lys Ser Ser Pro Ala Gly Lys  
 305 310 315 320

Leu Pro Gly Pro Ala Val Glu Gly Pro Ser Ala Thr Gly Pro Glu Ile  
 325 330 335

Ile Arg Ser Pro Gln Asp Asp Ala Ser Ser Pro His Leu Gln Val Met  
 340 345 350

Leu Gln Ile His Leu Pro Gly Arg His Thr Leu Phe Val Arg Ala Met  
 355 360 365

Ile Asp Ser Gly Ala Ser Gly Asn Phe Ile Asp His Glu Tyr Val Ala  
 370 375 380

Gln Asn Gly Ile Pro Leu Arg Ile Lys Asp Trp Pro Ile Leu Val Glu  
 385 390 395 400

Ala Ile Asp Gly Arg Pro Ile Ala Ser Gly Pro Val Val His Glu Thr  
 405 410 415

His Asp Leu Ile Val Asp Leu Gly Asp His Arg Glu Val Leu Ser Phe  
 420 425 430

Asp Val Thr Gln Ser Pro Phe Phe Pro Val Val Leu Gly Val Arg Trp  
 435 440 445

Leu Ser Thr His Asp Pro Asn Ile Thr Trp Ser Thr Arg Ser Ile Val  
 450 455 460

Phe Asp Ser Glu Tyr Cys Arg Tyr His Cys Arg Met Tyr Ser Pro Ile  
 465 470 475 480

Pro Pro Ser Leu Pro Pro Pro Ala Pro Gln Pro Pro Leu Tyr Tyr Pro  
 485 490 495

Val Asp Gly Tyr Arg Val Tyr Gln Pro Val Arg Tyr Tyr Tyr Val Gln  
 500 505 510

Asn Val Tyr Thr Pro Val Asp Glu His Val Tyr Pro Asp His Arg Leu  
 515 520 525

Val Asp Pro His Ile Glu Met Ile Pro Gly Ala His Ser Ile Pro Ser



ctgaaggggc tgcaccggtc ggtcgggaag ctggagagca acctggacgg ctacgtgcc 240

acgagcgact cgcagcgtg gaagaagtcc atcaaggcct gcctgtgccg ctgccaggag 300

accatcgcca acctggagcg ctgggtcaag cgcgagatgc acgtgtggcg cgaggtgttc 360

taccgctgg agcgtgggc cgaccgcctg ggtccacgg gcggcaagta cccggtgggc 420

agcgagttag cccgccacac cgtttccgtg ggcgtggggg gtcccagag ctactgccac 480

gaggcagacg gctacgacta caccgtcagc ccctacgcca tcacccgcc cccagccgt 540

ggcgagctgc ccgggcagga gcccggcag gccagcagc accagccgtg ggtccccggc 600

gaggacgggc agcccagccc cggcgtggac acgcagatct tcgaggacce tcgagagttc 660

ctgagccacc tagaggagta cttgcggcag gtgggcggct ctgaggagta ctggctgtcc 720

cagatccaga atcacatgaa cgggccggcc aagaagtgtt gggagttaa gcagggtcc 780

gtgaagaact ggggtggagt caagaaggag ttctctcagt acagcgaggg cacgctgtcc 840

cgagaggcca tccagcgcga gctggacctg ccgcagaagc agggcgagcc gctggaccag 900

ttctgtggc gcaagcggga cctgtaccag acgctctacg tggacgcgga cgaggaggag 960

atcatccagt acgtgggtgg caccctgcag cccaagetca agcgtttctt gcgccaccc 1020

ctgccaaga ccctggagca gctcatccag aggggcatgg aggtgcagga tgacctggag 1080

caggcggccg agccggccgg cccccacct cccgtggagg atgaggcgga gaccctcacg 1140

cccgcacca acagcgagtc cgtggccagt gaccggacce agcccag 1188

<210> 30

<211> 1188

<212> DNA

<213> *Orcinus orca*

<400> 30

ggggaattgg atcaacgtac taccggtggc cttcacgcat accctgcacc acgcgggggc 60

cctgtcgcga agccaaatgt catcctgcag attgggaagt gccgggctga gatgctggag 120

cacgtccgtc ggacgcacg tcatctctt actgaggtgt caaacaggt ggagcgtgaa 180

ctcaaaggct tgcaccgag cgttgggaaa cttgaaagca acttagatgg ctatgtgccg 240

actggcgaca gccagcgttg gcgtaagtcc atcaaagcat gtttgtgtcg ttgccaggaa 300

acgattgcaa acctggagcg ttgggtcaaa cgggagatgc atgtctggcg tgaagtattt 360

tatcgttag agcgttgggc cgatcgttta gagagcatgg gtgtaagta ccctgtgggg 420

agcaaccctt ctccgcatc gacgtcagtc ggtgttggcg ggccggagtc ctacggtcat 480

gaagcggaca cctacgacta taccgtaagc cttatgcta ttacccacc acctgcggcc 540

ggcgaattac ctggccagga agccgttgag gctcaacaat accctccttg ggggctgggc 600  
 gaggatggtc aacctagccc aggggtagac acgcaaatct ttgaggaccc acgggagttt 660  
  
 ctttcccacc tggaagaata cctgcgtcag gttgggtggga gcgaagaata ctggctgtca 720  
 caaattcaaa accatatgaa tggctctgca aaaaaatggt gggaatataa acagggttcc 780  
 gtgaaaaact gggttgagtt taaaaaggag tttcttcaat attccgaggg cgcctcagt 840  
 cgggaggcgg tccaacgcga gttggacttg ccacagaaac agggggaacc actcgatcaa 900  
 ttcctttggc ggaaacgtga cctttaccag acattgtacg tggatgcaga tgaggaagaa 960  
 attatccaat atgttgtggg gacctgcag ccgaaactga aacgtttcct tcgcccgcg 1020  
 ctgcctaaaa cgittgaaca acttattcag aaaggtatgg aggtcgagga tggcttagaa 1080  
  
 caagtgcag agccggcctc gccacacttg cctacagagg aggaatcgga ggcgctgacc 1140  
 ccagcactta catcagagtc agtggcatca gaccggacac aaccagag 1188  
  
 <210> 31  
 <211> 1170  
 <212> DNA  
 <213> *Odocoileus virginianus texanus*  
 <400> 31  
  
 ggggagttag atcacgtac aacggggggg ttgcacgcat accctgctcc acgtggcggg 60  
 ccggcagcta agccaaact aatcctgcag attgggaagt gccgggcaga gatgttggag 120  
 cacgtccggc ggaccaccg gcacctctg gctgaagtgt ctaaacaagt agaacgggaa 180  
 ctcaaaggtc ttcacgtag cgtcgggaaa ttggaatcga atttggacgg gtatgttct 240  
  
 acaggcgact cacagcgtg gaaaaagagc atcaaggcct gcctgagtcg ctgccaggag 300  
 acgattgcta acctcgaacg ctgggttaag cgggagatgc acgtttggcg cgaagtcttc 360  
 taccggctgg agcgttgggc tgatcggctc gaatctggtg ggggtaagta tccagttggg 420  
 tccgacctg ctgccacac agtctcagtt ggcgtaggtg ggccggagtc gtattgcaa 480  
 gatgcggaca actatgatta tacagtttcc ccatacgcga tcacaccacc gccggcagca 540  
 gggcagctgc caggtcagga agaggttgag gccagcagt atccacatg ggccccaggg 600  
 gaagacggcc agctttctcc tggggtggac actcaagttt ttgaagatcc gcgtgaattt 660  
  
 ctgcggcatt tagaagatta tctccgccag gtcggggggg ctgaagagta ttggttaagc 720  
 caaattcaaa accatatgaa cggcccggcc aagaagtgtt gggagtaca gcaagggtct 780  
 gtgaaaaatt gggttgagtt taagaaagaa tttcttcaat attctgaggg cactctttcg 840  
 cgtgaagcca tccaacgcga actcgactta ccgagaaac aaggggaacc tctcgaccaa 900

tttctgtggc gcaaacgcga cctgtaccag actctttacg tcatgctga ggaggaagaa 960  
 attattcaat acgtagtgg cacactgcag cctaagctta aacggttttt acgtccacca 1020  
 ttgccgaaga cgcttgaaca actcatccag aagggtatgg aggttcaaga tggcttgaa 1080  
  
 caggcagcgg aaccagcggc ggaggaggca gaagccctga cacctgcgtt aactaacgag 1140  
 tctgtcgcga gcgaccgcac ccagccggaa 1170  
 <210> 32  
 <211> 1203  
 <212> DNA  
 <213> *Ornithorhynchus anatinus*  
 <400> 32  
 ggggaattag accgcctgaa cccaagctca ggccctgcac catcctctgg ttgcatcca 60  
 taccaggtc tccggggcgg ggcaaccgcg aagcctaag tcattttgca aattggcaaa 120  
 tgccgtgcgg aaatgcttga acacgtccgc aaaactcacc gtcctctct cacagaagta 180  
 tcgcgccaag tagaacgcga gctcaaagc cttcacaaaa gtgttgcaa gttggaatca 240  
  
 aatcttgatg ggtacgtacc gtcaagcgac tccaacgct ggaagaaaag cattaaggcg 300  
 tgcttatccc gttgccaaga gacgattgag catttagaac gctgggttaa acgtgaaatg 360  
 aatgtatggc gtgagggtt ctaccgtttg gaacgttggg cggaccgtct ggagctatg 420  
 ggcggtaaat atcctgccgg tgagcaggcc cggcgtacag tttcagtggg cgttgggggc 480  
 cctgagacat gttgtccagg ggatgaaagt tatgattgac cgatttctcc gtatgcagtt 540  
 ccaccttcca ccggcagtc tccggaatcc ttagaccaag gggatcagca ctatcagcag 600  
 tggtttgccc tcccggagga gtcccctgtt agccctgggg ttgatacca gatctttgaa 660  
  
 gatcctcgcg agtttttacg tcatctggag aagtacctga aacaagtcgg cgggacagag 720  
 gaagactggc tttctcaaat ccagaatcac atgaatgggc cggcgaagaa gtggtgggag 780  
 tacaagcaag ggagtgttaa gaattggctt gaatttaaga aggaattttt acagtattcg 840  
 gagggcacac tgacgcggga cgcgttgaac cgtgaactgg atctcccaca gaaacaaggc 900  
 gaaccacttg atcaattttt atggcgggaag cgcgacttat atcagacact ctacgttgac 960  
 gccgatgaag aggaatcat tcatgctgc gtgggcactc ttcagccgaa attaaaacgc 1020  
 tttctccatc acccactccc taagacgctt gacgagctta tccaacgggg ccaagaagtt 1080  
  
 cagaatggte tggagcctac cgacgatcct gcaggccaac gcaactcaatc ggaggacaac 1140  
 gacgaaagcc ttaccctgac cgtcaccat gagagtactg caagcagagg caccctgcca 1200  
 gag 1203

<210> 33

<211> 1212

<212> DNA

<213> *Anser cygnoides domesticus*

<400> 33

gggcagcttg ataacgttac aaacgcgggc atccactcct tccaggggca tctggcgta 60  
 gcgaataagc caaatgtcat tctgcaaatt ggtaaatgtc gtgcggaaat gctggagcac 120  
 gttcgcgcga cccaccgcca tttattatct gaagtatcta agcaggtaga acgtgagctg 180

aaagggctgc aaaagtccgt gggcaagctc gagaataact tggaggatca tgtccctaca 240  
 gataaccaac gctggaagaa gtccattaaa gcgtgcttgg ctctgtgtca agagactatc 300  
 gcgcatttag agcgttgggt gaaacgcgaa atgaacgtct ggaaggaggt gtttttcgg 360  
 ctggaanaagt gggcagaccg gctggagtca atgggtggca agtactgccc gggcgaacac 420  
 gggaaacaaa ccgtcagtggt aggcgtgggg ggtcctgaaa tccggccttc ggagggggaa 480  
 atttatgatt atgctctgga tatgagccag atgtatgcac tcaccccacc tccagcgcaa 540  
 atgcatcaa tcccacaagc ccatgacagc tatcagtggg ttagtgtctc agaagatgcc 600

ccggcgagcc ctgtcgaac ccaggtattt gaggaccctc gggaaattcct gtctcacctg 660  
 gaggaatacc tgaagcaggt aggcggcacg gaggagtatt ggttgtccca gatccagaat 720  
 cacatgaatg gtccggcaaa aaaatggtgg gaatataaac aggactccgt taaaaactgg 780  
 gttgagttta aaaaggaatt cttgcaatac tctgaaggta ctttaactcg ggatgctatt 840  
 aagcgtgaac tgcacttggc gcaaaaggaa ggtgaacctc ttgaccaatt cctttggcgg 900  
 aagcgggacc tctatcagac actttacgtg gacgcgatg aggaggagat cattcagtat 960  
 gtggtcggta cctgcagcc gaagetcaag cgtttcctga gctatcctct cccaaagact 1020

ttagaacagc tcatccagcg cggtaaagaa gtgcaggta acatggatca ctccgatgag 1080  
 ccttcgccgc agcgtacacc tgaattcaa tcaggtgact ccgtagaatc tatgccacct 1140  
 tcaacaacgg catctccggt tccatcta at ggtacceaac ctgagccgcc gagcccgcca 1200  
 gccaccgtta tc 1212

<210> 34

<211> 1185

<212> DNA

<213> *Pelecanus crispus*

<400> 34

gggcaacttg acaacgtaac aaacgtggg attcactcct ttcaggcca ccgcggtgctc 60

gccaacaagc caaacgtaat cttgcaaatt ggcaaatgcc gtgcggagat gttggaacac 120

gttcgtcgta cacatcgta cttgctgtcg gaagtctcta aacaagtaga acgtgaactt 180

aaagggttc aaaagtcagt cggcaaatg gaaaacaacc ttgaagacca tgtaccaacc 240

gacaatcagc gtiggaaaa gtctatcaaa gcttgcctgg cccgttgtca agagacgatt 300

gctcacctgg agcgggtggg aaagcgcgag atgaatgtgt ggaaagaggt cttcttccgc 360

ttggaaaaat gggccgaccg tttggagtcc atgggcggta aatattgtcc gggatgaacat 420

ggtaagcaaa cagtctctgt gggcgttggg gggccggaga ttcggccttc tgaagcgcgag 480

atctacgatt atgcgctcga catgtcccag atgtatgcgc ttacaccacc accgggcgag 540

gtaccaagca ttctcaagc gcatgacagt tatcagtggg ttagcgtatc cgaagacgct 600

cctgcctcgc cggtagagac ccaggTTTT gaagatcctc gtgaatTTTT aagccacttg 660

gaggagtatt tgaagcaggt aggggggaca gaggaatatt ggctgtctca gatccagaac 720

cacatgaatg gcccgctaa aaagtgggtg gaatacaaac aagattcggg aaagaattgg 780

gtagaattta aaaaggagtt tttacagtac tcagagggga ctctcacgcg tgatgcgac 840

aaacgcgagt tggatcttc tcaaaaagag ggggagccac tcgatcagtt cctctggcgc 900

aagcgggatc tctaccaaac actctacgta gacgcagacg aagaagagat catccagtac 960

gtggtgggta cgtccagcc gaaactcaaa cgtttctca gctaccact tcttaagact 1020

ctggaacaac tgattcagcg gggcaaagag gtccagggta acatggacca ttcagaggaa 1080

cctagtcgc aacgtacacc tgagatcaa tctggggatt ctgtcgattc ggttccacct 1140

tctacaacag cgtctccggg gccgtcaaat gggacceaac cagag 1185

<210> 35

<211> 1185

<212> DNA

<213> *Haliaeetus albicilla*

<400> 35

gggcagcttg ataatgtaac caatgcaggt atccactctt tccagggtca cgcggtgtg 60

gcaacaagc caaatgttat tctgcaaatt gtaagtgtc gcgctgagat gttagaacac 120

gtccggcgca cgcacggca tctctgtca gaggtttcaa agcaggtaga gctgaatta 180

aaggcctcc agaagtcctg aggtaaactc gaaaataatc ttgaagacca cgttctacc 240

gataatcaac ggtggaaaa gtcaatcaag gcgtgcttag cacggtgtca ggaaacgatc 300

gcgcacctcg aacgttgggt gaagcgcgaa atgaatgtct ggaaagaagt gttcttccgg 360

cttgagaagt gggctgatcg gctcgaatcc atgggtggca aatattgtcc aggtgatcat 420

ggcaagcaaa cggctctcgt cgggtgttggg ggtccggaaa tccggccgag cgagggtgaa 480  
 atctatgact acgctcttga tatgtcccag atgtatgcac tcactcctcc gccgggtgag 540  
  
 gtccccgtcga tcccgcaggc gcatgactca taccaatggg tgtcgactag cgaagacgca 600  
 ccagcctccc ctgttgaaac tcaagtattc gaggaccgc gtgagttcct gagccattta 660  
 gaggagtacc ttaagcaggt tgggtgtacc gaggaatact ggttgagcca gattcagaat 720  
 cacatgaacg ggccggctaa gaaatgggtg gaatacaagc aggattcagt caagaattgg 780  
 gtcgaattta agaaggagtt tttgcagtac agtgagggga cgctcacacg cgacgctatc 840  
 aaacgggagc tggacctgcc acaaaaggag ggtgaaccgc ttgatcagtt tctttggcgc 900  
 aagcgtgatc tgiatcaaac cctgtatgtg gacgctgacg aagaagagat cattcagtac 960  
  
 gtggttggga ctctgcaacc aaagctgaag cgttttcttt cttatcctct ccctaagaca 1020  
 ctggaacagt taatccaacg tggcaaggag gtccagggta atatggacca ctctgaggaa 1080  
 ccgagcccgc aactactacc tgaaattcag agcggggata gtgtcgactc agttcctcca 1140  
 agtacaccg catccccggt cccaagtaac ggtaccaaac cagag 1185  
  
 <210> 36  
 <211> 1395  
 <212> DNA  
 <213> Ophiophagus hannah  
 <400> 36  
  
 gggctcttggg gcttgaacg tcacgtggct gatgaacgtc gtggcctcgc tacgcctacc 60  
 tacggcgcgg tttgttccat tcgggagaaa aaagcctccc aactgagcgg ccagagctgt 120  
  
 ttggagaaaag agttgcttgg ttggaatgt acggaggcaa tcgtggaat gatgcaagtc 180  
 gataacttta accacggtaa cttacatagc tgccaaggcc atcgggggat ggcaaatcac 240  
 aaaccgaacg taatccttca aatcgggaaa tgtcgcgcag aaatgttaga ccacgtgcgt 300  
 cgcaccacc gccatctctt gacggaggtt tcgaagcagg tagaacgca attgaagtct 360  
 ctccaaaagt cggttggcaa gctcgagaat aatctggaag accacgtgcc atcggcagcg 420  
 gagaaccaac gtggaagaa atcaattaa gcctgcctgg cccggtgcca agaacaatt 480  
 gctcacctcg aacgctgggt taaacgcgaa atcaacgtct ggaaagaagt attctttcgt 540  
  
 ctggagaagt gggcggaccg ccttgagtcg ggtgggggca agtatgggc tggtagacaa 600  
 agtcgtcaaa ctgtaagtgt cgggtttggg gccccagaaa tccaaccgcg gaaagaagaa 660  
 atctatgact acgctctcga catgtcgcag atgtatgcct taacaccacc gccgatgggt 720  
 gaagacccaa acgtacctca atcccagat agctaccagt ggattacat ctgagacgat 780

tcacctccgt cgccagtgga aactcaaatt ttcgaggatc cacgcgaatt ccttacccat 840  
 ctcgaggatt atcttaagca agtgggcggg actgaagaat attggttgag tcagattcaa 900  
 aatcatatga acggtccggc caagaaatgg tgggagtaca aacaagattc cgtgaaaaac 960  
  
 tggttggaat tcaagaagga attccttcaa tactctgagg gtactttgac acgtgacgca 1020  
 attaaacaag aacttgactt accgcagaag gacggcgagc cattggatca atttctttgg 1080  
 cggaagcggg acctgtatca gacgetctat attgatgcag aggaggaaga agtaatccaa 1140  
 tacgtttgtg gcacactcca accgaaatta aacgtttcc tttcccacc gtatccgaaa 1200  
 actttggaac agttaatcca acgtgggaaa gaggtggaag gcaacctcga taactctgag 1260  
 gagcctagcc cgcaacggag tccaaagcac caattgggtg gtagcgtcga gagcctccca 1320  
 ccttcgtcga ccgcaagtcc tgttcgtcga gacgagactc acccagacgt gagcgcacct 1380  
  
 ccggtaacgg tgatt 1395  
 <210> 37  
 <211> 1353  
 <212> DNA  
 <213> Austrofundulus limnaeus  
 <400> 37  
 ggggacggcg agactcaagc tgagaateca tctaccagct tgaacaacac tgacgaagat 60  
 atcttggaa acgtcaagaa aattgtcatg gatcaacaac acctgtatca gaaagaatta 120  
 aaggcatctt ttgaacaact cagtcgcaaa atgttttccc agatggaaca aatgaatagc 180  
 aagcaaacgg atctgctttt agaacatcaa aaacagactg tcaaacatgt agacaagcgc 240  
 gtggagtatt tgcgggcgca attcgatgca tcgttaggct ggcggtttaa agagcaacac 300  
  
 gcggatatta cgacaaaaat cattectgag atcatccaaa cggatgaagga agatattagc 360  
 ctgtgtcttt ctacgctctg cagtatcgct gaagatatcc agacatcacg ggctaccact 420  
 gtcacagggc atgctgccgt acaaacccat cctgtggatc ttttgggtga acaccattta 480  
 gggaccacgg ggcacccacg cttacagtgc acccgtgtag ggaaaccaga cgacgtacct 540  
 gagtcgccgg taagcctggt tatgcaaggt gaggcgcgtt cccggatcgt tggcaagagt 600  
 ccgattaaac tgcaatttcc gacgttcggc aaagcaaacg attcttccga cccactccaa 660  
 tatctggagc ggtgtgagga ctttcttctt cttaacctt taactgatga ggaacttatg 720  
  
 gctactttgc ggaatgtgtt acatggcacc tctcgggatt ggtgggatgt cgcacgtcat 780  
 aaaaatccaa cttggcgtga gtttaataaa cacttccggg cggttttct cagcgaggat 840  
 tatgaagatg agttggctga gcgcgtccgt aaccgcatcc aaaaagaaga tgagtctatc 900

cgcgatttcg cttatatgta tcagtccttg tgcaagcggg ggaaccctgc tatctgcgaa 960  
 ggtgatgtag taaagctcat cctgaagaac atcaatccac aactgccctc tcagttacgc 1020  
 tcccgggtca cgaccgtgga tgagcttggt cgcttgggcc agcagcttga aaaagatcgt 1080  
 cagaatcagc tccaatatga gcttcggaag agttccggca aaattatcca aaaatctagt 1140  
  
 tcgtgcgaaa cttcagcgt cccgaacacg aagagtacac ctaatcaaca aaacctgct 1200  
 accagtaacc gtctccaca ggtgtattgc tggcgggtgta agggtcacca tgccctgcc 1260  
 tcttgccgc aatggaaagc tgataagcac cgtgcgcaac cttcgcggag ttctgggcca 1320  
 caaactctga ctaatctcca agctcaagac atc 1353  
  
 <210> 38  
 <211> 1188  
 <212> DNA  
 <213> *Physeter catodon*  
 <400> 38  
  
 ggggaattgg atcaacgtgc ggcagggggc ttgcgcgctg acccggcgcc gcgtggtggt 60  
 ccagttgcc aaccgagcgt aattcttcag attggttaagt gccgcgctga gatgctggaa 120  
  
 cagtcgcc gcacgcacg ccatcttctg acggaggtaa gtaacaagt ggagcgcgaa 180  
 ctcaaggggt tacatcgctc tgctcgtaag ttggaggcca atttagacgg ctatgtgcct 240  
 accggtgatt cccaacgctg gaaaaaaaagt atcaaggcgt gtctctgccg gtgtcaggaa 300  
 acaattgcaa atctcgagcg ttgggtgaaa cgtgagatgc atgtttggcg tgaggtattc 360  
 tatcgtttgg aacgggtggc agaccgtttg gactctatgg ggggcaagta tccgggtggc 420  
 actaaccctg cgcgccacac agtaagtgtc ggggtagggg gcccggaagg ctattctcat 480  
 gaagcggata cttatgacta cacggtgtct cgtatgcta tcacgccacc gcctgccgcg 540  
  
 ggtgagttgc ctggtcaaga ggctgtcgag gcacaacagt accctccatg gggcttgggg 600  
 gaggacgggc aaccaggtcc gggcgtggac acgcagattt ttgaggacce tcgcaattt 660  
 ttgagccact tagaggagta cctgcggcaa gtagggggga gtgaagagta ctggttatcg 720  
 caaattcaaa atcatatgaa tggccctgcg aagaaatggt gggagttaa acaggggtca 780  
 gtcaagaatt gggctgagtt taagaaagaa tttttgcaat acagtgaggg tacgttgagt 840  
 cgcgaggcca tccaacgtga actggaccic cctcagaagc agggggagcc gttagatcaa 900  
 tttttatggc gaaacgtga cttatacaa accctctacg ttgacgctga ggaagaagaa 960  
  
 attattcaat atgttgtcgg tacgtgcag ccaaagctga agcggttcct ccgtctcca 1020  
 ctccctaaaa cttagaaca attaatcaa aaaggcatgg aagttcagga cgggttagaa 1080

caagcggccg aaccggcctc tccgcgtctg ccgccggaag aggagagtga ggctcttacg 1140  
 cctgcctca cgagcgaatc agtagcctcc gatcggacac agccagag 1188  
 <210> 39  
 <211> 1212  
 <212> DNA  
 <213> Meleagris gallopavo  
 <400> 39  
 gggcagcttg acaatgtgac gaacgcgggg attcacagct ttcaaggca ccgcggcgtc 60  
 gccacaacaac cgaatgtcat tctgcaaatc ggtaaatgtc gtgctgaaat gcttgagcac 120  
 gttcgtcgta cccatcgta cttgctttct gaagtatcaa aacaagtgga gcgggaactc 180  
 aaaggcctgc aaaagtcagt gggtaaatg gagaataacc tcgaagacca tgtacctaca 240  
 gacaaccagc ggtggaaaaa atctatcaag gcatgcctcg ctcgttgcca ggagactatt 300  
 gcccatcttg agcgggtggg gaaacgtgaa atgaacgtat ggaaggaagt atttttcgc 360  
 ttagagaagt gggctgatcg tcttgaatcg atggcgggca agtactgtcc tggggaacac 420  
 ggcaaacaaa ctgtatctgt cggcgtgggg gggccggaga tccggccatc ggaaggggaa 480  
 atttatgatt atgctctcga catgtcccaa atgtatgctc tcacaccagg gccaggggaa 540  
 gtaccgtcaa ttccgcaage acacgacagc taccaatggg tatctgtgag cgaggacgcg 600  
 cctgcctctc cggttgagac gcaaatcttt gaggaccac atgaatttt gtctcatctt 660  
 gaagaatata tcaaacaggt tggcggcaca gaagaatact ggttatctca gatccagaat 720  
 cacatgaacg gcccggttaa aaagtgggtg gagtataagc aagattccgt aaagaactgg 780  
 gtcgaattca agaaagagtt tcttcaatac tctgagggta ctctgacgcg cgatgcaatt 840  
 aagcgggagt tagacctcc acaaaaagag ggggagcctc ttgaccagtt cctgtggcgt 900  
 aagcgcgacc tctatcagac actttacgtc gacgctgatg aagaagagat tattcaatat 960  
 gttgtgggta cctgcagcc aaagcttaag cgtttcctta gctaccact tccgaaaact 1020  
 ctggagcagc tcattcaacg cggtaaggaa gtgcagggca acatggacca ctctgaagag 1080  
 cctagcccgc agcgcactcc tgaaatcaa tcaggtgaca gtgtggagtc aatgcccccg 1140  
 tcaaccaccg ctctccggg acctagcaac gggacgcaac cagacctcc aagcccaccg 1200  
 gctacagtca tc 1212  
 <210> 40  
 <211> 1227  
 <212> DNA  
 <213> Pogona vitticeps

<400> 40

gggcaacttg agaataattaa ccaaggttcc ctgcacgcgt ttcagggtca tcgcggcgtg 60

gtccataaca acaagcetaa cgttattctc cagatcggga agtgccgcgc cgaaatgctg 120

gagcatgtgc ggcgcaccca tcgccatttg ctactgaag tatcaaaaca ggtggagcgt 180

gagttgaagg ggttcagaa aagttaggc aaacttgaat ataatttaga agaccacgta 240

ccaagtgcgg ctgagaacca acgctggaag aagtcgatta aagcctgctt agcgcgttgt 300

caggagacca ttgcgaactt ggaacgctgg gttaaactg agatgaatgt ttggaaggag 360

gtctttttcc gcttagagcg ctgggcagat cgcctcgaat ccgggggttg caagtactgc 420

catgcagacc agggtcgcca aactgtcagc gtaggtgttg gtgtcctga agtgcgtccg 480

tctgaaggtg aaatttacga ttacgcgttg gatatgagcc aaatgtacgc cttgactccg 540

ccgcctatgg gtgatgtcc agtaattcct cagccgatg acagtatca gtgggtgaca 600

gatccggaag aagcgcacc aagtcggtt gagacacaaa ttttcgagga cctcgggag 660

tttctgacce atcttgagga ttatttaaaa caagtcggcg ggacagagga atattggctc 720

tcacagatcc aaaaatcat gaatgggcca gcgaaaaagt ggtgggaata taaacaggat 780

agtgtgaaga actggcttga gttcaaaaa gaattcttgc agtactcaga aggcacgtta 840

acgcgggacg ctattaaaca ggaacttgac ctccacaaa aagaaggga accgctggat 900

caattctctt ggcgcaaacg cgatttgtac caactctct acgtcgagc agaagaagag 960

gaggtcatcc aatatgtagt tggcacactg caacaaaac tgaagcgtt tctttctcat 1020

ccgtacccta aaacctgga gcaactcatc cagcgcggga aggaagtga ggggaatttg 1080

gacaatagt aagaaccgtc tccacagcg acccagaac atcagctggg ggacagtgtg 1140

gaatctttgc cgcctagtac tacggttcg cctgccggtt cggataaaac gcaacctgag 1200

attagettac ctccaactac agtcatt 1227

<210> 41

<211> 1212

<212> DNA

<213> Alligator sinensis

<400> 41

gggcaattag attcggtaac caatcggggc gtccacacct accagggcca tcggagcgtc 60

gccaataaac ctaacgtcat tcttcaaatc gggaaatgct ggactgagat gctggagcat 120

gtccgtcgga ctcatcgcca cctgetcaca gaagtgtcaa agcaagtgga acgtgaactc 180

aagggtctac agaagagcgt gggcaaactg gaaaacaatc ttgaagacca tgtccaact 240

gacaatcagc ggtggaagaa gtcaatcaag gcatgtctcg cgcgttgcca agagaccatt 300  
gctcaccttg agcggtaggt gaaacgtgaa atgaacgtgt ggaaggaggt gttcttccgg 360  
ttagaacgct gggccgaccg ccttgaatca atgggtggta aatactgccc gacggactct 420

gcacgtcaga cagttagcgt tggggtgggg ggcccggaaa ttcgccttag tgaaggcgaa 480  
atctatgact acgcgtcga tatgagccaa atgtacgctc ttacccgctc accgggcgaa 540  
ttgccgtccg tcctcaacc gcatgattca taccagtggg tcaactagtc ggaagacgct 600  
ccggcgtcac cagttgaaac gcaggtattc gaggatcctc gggagtctt gtgtcatttg 660  
gaagagtacc tgaagcaggt tggcgtaca gaggaatatt ggctgagcca gattcagaat 720  
catatgaatg gtctgcaaa aaagtgggtg gaatataaac aagacacggt taagaattgg 780  
gtggaattca agaaggagtt cttacaatac agtgagggtg cacttaccg tgatgcgatt 840

aagcgggaat tagacctccc gcaaaaggac ggtgagcctc tggatcaatt tttatggcgt 900  
aagcgtgacc tctatcagac attatacatt gatgccgatg aagaacagat cattcagtac 960  
gtcgtgggga cattgcaacc taaactcaag cgttcttgt cctatccact tccaaaaact 1020  
cttgaacaat taatccagaa agggaaggag gtgcagggtt cacttgacca cagcgaggag 1080  
ccgagtcctc aacgtgcgag cgaggctcgg acgggcgata gtgtgaaac cttgccgct 1140  
tctaccacta catcaccaaa tacgtcatct ggtacacagc cagaggcacc atcgctcca 1200  
gcgacggtaa tc 1212

<210> 42

<211> 1212

<212> DNA

<213> Alligator mississippiensis

<400> 42

gggcagttag acagtgtgac taacccggg gtgcatacgt accaggggca ccgcggggtc 60  
gccaataage caaatgtaat tctccagatt gggaagtgtc gtacagagat gttggaacat 120  
gtccgtcgca ctcatcgcca cttgetcacc gaggtctcca aacaagtaga acgcaactc 180  
aaggggctcc agaagagtgt tgggaagtgt gagaataacc tcgaagacca cgttccgaca 240  
gataaccaac ggtggaaaaa gtctattaaa gcctgtctcg cccgttgtca agagacaatc 300  
gcacacttgg aacgctgggt caaacgggag atgaatgtgt ggaaggaagt cttcttccgt 360

ctcgagcggg gggcggatcg tttagaaagt atggcggta aatattgccc aactgactcg 420  
gctcgtcaaa cgggtgtcgt tggcgtagc ggcccggaaa ttcgccttag cgagggtgag 480  
atctatgact atgcacttga catgagttag atgtatgctg taactccgct gccaggggag 540

cttccaagta ttccacagcc tcacgatagt tatcaatggg taacttctcc tgaagacgcc 600  
 ccagcatccc cagttgagac acaagtattc gaggaccctc gtgagtttct ctgtcacctc 660  
 gaggagtacc ttaaacaggt aggcgggacc gaagagtact ggttatcgca aatccaaaac 720  
 catatgaatg gtctgcca aaagtgggtg gagtataaac aagatactgt gaagaattgg 780  
  
 gtagagtcca agaaagagtt cttacagtac tctgagggga cgtaactcg tgatgcgatc 840  
 aagcgcgaat tggatttacc tcagaaggac ggcgagccac tcgaccagtt cttatggcgc 900  
 aagcgtgact tgiatcaaac cctttatc gatgctgacg aggaacaaat tatccagtac 960  
 gtagtcgcta cgttgcaacc aaaacttaa cgttttctga gctaccatt acctaaaacg 1020  
 ttggagcaac tgatccagaa aggtaaagag gtgcaaggga gcctggatca tagtgaagaa 1080  
 ccgagcctc agcgggcttc tgaagctcgg accggtgata gcgtcgaatc ttaccacct 1140  
 agtaccacaa ccagcccga tgcgtcatct ggtaccaac ctgaagcgcc ttccccacct 1200  
  
 gctacagtca tt 1212  
 <210> 43  
 <211> 1224  
 <212> DNA  
 <213> Gekko japonicus  
 <400> 43  
  
 gggcagctcg agaatgtcaa ccatgggaac ctccattctt ttcaaggtca tcgcgcggc 60  
 gtgccaaca agccaaactg tatcttgag atcggtaaat gtcgtgcaga gatgctggac 120  
 cacgtccggc ggaccaccg gcatttactg acagaggtat cgaacaggt tgaactgag 180  
 ttgaaggggt tacagaaatc agtaggaaa ttagaaaata acttagaaga ccatgtcct 240  
 tcagccgttg aaaaccagcg ttggaaaaaa tcgatcaagg cctgccttc ccgctgcca 300  
  
 gagaccattg cccacctga gcgttgggtg aagcgcgaga tgaactatg gaaagaggtt 360  
 ttcttccgct tagagcgggtg ggcagatcgg ttggaatctg ggggcgggaa atattgtcac 420  
 ggtgataate atcgtcaaac agtatcagtc ggtggtggcg gccctgaggt acgtccatct 480  
 gaagcgaaa tttacgatta cgctctcgac atgtcgaaa tgtacgcttt aacaccgct 540  
 agcccagggg atgtgcctgt agttagccag ccgcacgaca gctatcagtg ggttacggtt 600  
 ccggaggata cccctccatc cccggtggag acgcaaatct tcgaggacc acgggagttc 660  
 ttgaccact tagaggatta cttaagcaa gtgggggta cagaggaata ttggttatct 720  
  
 cagatccaga atcacatgaa cgggccagcc aagaagtggg gggagtataa gcaagactca 780  
 gtaaaaaatt ggctcgagtt taagaaggaa ttccttcagt attccgaggg gacacttacg 840

cgcgacgcta tcaaggaaga acttgacctc ccgcaaaagg acggggaacc tcttgatcag 900  
 ttctgtggc gcaagcgcga cttgtaccag accctgtacg tggaggcgga tgaggaggag 960  
 gtgatccagt atgttgtggg gactttacaa cctaaattaa agcgttttct ctcacacct 1020  
 taccgaaaa cgttagagca acttatccaa cggggcaaag aggtggaagg gaacctcgac 1080  
 aattcagagg aaccaacacc tcagcgtact ccagaacacc aactgtgtgg ttctgtagaa 1140  
  
 tcgctgccte cttcctctac cgteagteca gtggctagcg atggtactca acctgagact 1200  
 tcgccattgc cagcgactgt tatt 1224  
 <210> 44  
 <211> 1365  
 <212> DNA  
 <213> Homo sapiens  
 <400> 44  
  
 gggccattga cgttgttaca agactggtgt cgtggtgaac atttaaacac ccgccggtgc 60  
 atgttgatcc tcggtatccc agaagattgc ggcgaggatg agttcgaaga gacacttcag 120  
 gaggcgtgtc gccatttagg gcggtaccgc gtgatcggcc gcatgttccg tcgtgaggaa 180  
 aatgcccaag cgatcctctt ggaattggcg caggatattg actatgcctt actccctcgg 240  
  
 gaaatccctg ggaaaggcgg gccttgggag gtaattgtga agcccgtaa ttccgacggc 300  
 gaattcttaa atcgcttaa tcgctttctt gaagaggagc gccgtacggt ctccgatatg 360  
 aaccgtgttt tgggctcgga tactaactgt tcagctcttc gtgtcaccat tagtctgaa 420  
 ttctggactt gggcacagac gctgggcgca gctgtccaac cattgctcga acagatgctc 480  
 taccgggagt tacgggtctt cagtggcaat acgatttcca tcccagggtc tctcgctttt 540  
 gacgcgtggc tggagcatac cacggaaatg cttcaaatgt ggcaggtgcc tgaaggggag 600  
 aaacggcggc gcttgatgga gtgtttgcgg gggccagccc tgcaagtcgt tagtgggtta 660  
  
 cgtgcatcga atgccagtat cactgtcgaag gtagtctctg ctgcaactgca gcaggtattc 720  
 ggtccagtgg aaagtataa gattgcccga gtaaagtat gcaaagctta ccaggaggct 780  
 ggggaaaaag taagcagctt cgttttgcgt ttggagccac tgcttcagcg tgctgtagaa 840  
 aacaactggt tcagtcgccg caatgtcaac caaacacgtc ttaagcgtgt tctgtcgggc 900  
 gccacccttc ctgacaagct gcgtgataaa ttgaagttaa tgaaacagcg ccgtaaacgg 960  
 ccgggtttct tggcgttggg taaactgtta cgtgaagagg aggagtggga ggccacctta 1020  
 gggccagacc gcgagtcatt ggaggggtta gaagtggcac cgcgcccgcc agcacggatt 1080  
  
 acgggtgttg gcgcagtacc tcttccggca tccgggaatt catttgatgc ccgtccttcg 1140

caagggtacc ggcgccgtcg gggtcgtggt cagcacctgc ggggcggcgt tgctcgtgca 1200  
 ggctctcgtg gctctcgtaa gcggaacgg cacaccttct gctattcctg tggtaggat 1260  
 ggccatattc gtgtccaatg cattaacct agcaatctcc tgttgctaa ggagacaaa 1320  
 gagatitttg aagggggaga acgtgaagcg caaacgaatt cacgt 1365

<210> 45  
 <211> 1344  
 <212> DNA  
 <213> Homo sapiens  
 <400> 45

ggggctctta cgctcttaga agactggtgt aagggtatgg acatggacc gcggaaggct 60

ctctgattg taggtattcc gatggaatgc agtgaggtgg aaatccagga tacagttaaa 120  
 gctggcttc aacctctgtg cgcttatcgt gtactcggcc gtatgttccg gcgggaggat 180  
 aatgcgaagg ctgttttcat tgagctggca gacaccgtga attacaccac gttaccgtct 240  
 cacattccgg gtaaaggggg ttctctggaa gtcgttgta aacctcggaa cctgacgac 300  
 gagttccttt ctcgcttaa ctactcttg aaagatgagg gccgctcgat gacggatgtc 360  
 gcccgggcac tggggtgctg tagcttacct gcggaatcac tggacgcgga agtaatgcca 420  
 caggtccgct ccccaccatt agaacctcca aaagagagta tgtggtaccg taagttaaaa 480

gtgttagtg gtaccgctc gccttcgccg ggggaggaga catttgagga ctggttagag 540  
 caagtcaccg agatcatgcc tatctggcaa gtatctgaag ttgaaaagcg ccgtcggtta 600  
 ctggagtac tccggggccc ggcaactctca attatgcgcg tgttacaagc caataacgat 660  
 agcattaccg ttgaacagt tttggatgca ttaaagcaga tctttggcga caaggaagac 720  
 ttccgtgct ctcaatttcg ttttcttcaa acgtccccta aaattgggga gaaggtgagt 780  
 acgttctcgc tgcgtttaga gccactcttg caaaaggccg ttcacaagag cccactttcg 840  
 gtacgtagta ctgatatgat tcggttaaag cacctgttgg cacgcgtagc catgacccc 900

gcactgcgtg gtaaactcga attactcgac caacgcgggt gcccaccta ttttcttgag 960  
 ctgatgaagc tgatccggga tgaggaagag tgggagaata ctgaagctgt gatgaaaaat 1020  
 aaagagaaac cttcaggtcg tggcccggt gcatcaggcc gcaagctcg cgccaggcc 1080  
 agtgaagtg ctccgaagc aacagtcca gcacgtagct tctctgattc tagcccgcag 1140  
 acgattcagg ggggcttacc acctctgtc aagcgtcggc gccttttggg ttcggagagc 1200  
 acacgtgggg aagaccagg gcaagctact tatccgaaag cagagaatca gactccaggg 1260  
 cgtgagggcc cgcaggcggc tggggaggaa cttggtaatg aggccggggc cggcgcgatg 1320

tccccccga aaccgtggga aacc 1344

<210> 46

<211> 1197

<212> DNA

<213> Homo sapiens

<400> 46

ggggctgtga caatgctcca ggactggtgc cgttggatgg gcgtgaacgc tcggcggggg 60

ctgttaatct taggtatccc tgaagactgt gacgatgcag agttccaaga gtcgttagaa 120

gctgcactcc gtcctatggg tcactttact gtactcggta aggcttccg cgaggaagac 180

aacgctaccg ctgcgctggt ggaattagat cgcgaggtta attacgact tgttccacgc 240

gaaattccgg gcaccggcgg gccttggaac gtcgtgttcg ttctcgggtg ctccggcgag 300

gaattcctgg ggttaggccg cgtgttccac tttcctgaac aggagggcca aatggtagaa 360

tcggttgcgg gggcactggg ggtaggtctg cgccgcgtgt gttggttacg ctcgatcggg 420

caagctgtac aaccatgggt agaagctgtt cctgcctaaa gcttaggggt atttagtgtt 480

cgtgatcaac ctgcacctgg tgaagaaagc ttcgaggtct ggttggatca tacgaccgag 540

atgttgcatg tgtggcaagg cgtgtcggaa cgggaacggc gccgtcgtct gctggaaggg 600

ctgcgtggca cagccttaca acttgtacat gccttactgg cagaaaatcc ggcacggaca 660

gcacaagatt gcttggctgc attagcccaa gtttttgggtg ataacgaaag ccaggcaacg 720

attcgtgtta aatgtttgac agcccaacag cagagtggcg aacgcctctc tgcgttcggt 780

ctccgcttag aagtacttct gcaaaaggct atggagaagg aagcattggc gcgcgcgtca 840

gcggatcggg tgcgtcttcg tcagatgctg acacgcgcac atctcacaga gccgttggat 900

gaagccttac gaaattgcg tatggcaggg cgttctccgt cttttttgga aatgctcggc 960

ttagtacgcg agtcagagcg ctgggaggca agtctggctc ggtccgtccg ggcgcaaacc 1020

caggagggtg caggggcccc ggcgggggcc caagcagttg cgcgtgccag cactaaggtt 1080

gaagctgtac ctggtggccc tggccgggag ccagaaggtc tctccaagc cgggggccaa 1140

gaagcgggaag aacttctcca agagggtta aagccggttt tagaggaatg tgacaat 1197

<210> 47

<211> 1197

<212> DNA

<213> Homo sapiens

<400> 47

ggggcgggtca ccatgttgca agactggtgt cgttggatgg gcgtgaatgc tcggcggggg 60

ttattgatct tgggtatccc agaagactgt gacgacgccg agtttcagga gtcgctcgag 120  
 gccgccttc gtccaatggg gcattttacg gttctgggca aggtgttccg tgaagaggat 180  
 aacgctacag cagctcttgt ggagcttgac cgtgaggtga attatgcgtt agtacctcgc 240  
 gagattccag gtaccgggtgg gccatggaac gtagtcttcg tcccacgttg ctcgggggag 300

gaatttctgg ggcttgggcg cgtattccac tttccagaac aggaagggca gatggtcgaa 360  
 agcgtagcag gcgctcttgg cgttggtctc eggcgcgtgt gctggttacg ctccatcggc 420  
 caagcagtcc aaccatgggt tgaagccgta cgctatcaat ctttaggtgt cttctcaggc 480  
 cgtgaccage cggcgcctgg tgaggaatcc ttcgaagtct ggctcgatca tacaactgag 540  
 atgctcatg tatggcaagg tgtctcagag cgggaacggc ggcgcggtt attagagggg 600  
 ctccgtggga ctgcgctcca attagtacat gcgcttttgg ccgaaaatcc agcccgtact 660  
 gcccaagatt gtctggcagc actcgcccaa gtattcggcg acaacgaatc gcaggcaaca 720

atccgcgtaa agtgtcttac agcacagcag cagtcagggg aacgtcttag tgcgttcgtt 780  
 ctgcggctgg aagtgttact ccagaaagcc atggaaaagg aggcattggc tcgcgcgagc 840  
 gctgaccgtg tacgtctcgg gcaaatgctt actcgcgcac atctcaccga gcctctcgat 900  
 gaagcactgc gaaactcgc catggcaggc cgcagcccgt ctttctgga aatgttaggc 960  
 ttagtccggg agtccgaagc ctgggaggcc agtctggcac ggtcagtgcg ggcacaaacg 1020  
 caagaggggtg caggggcacg ggcgggtgca caagcagttg cacgtgcctc cactaaagtt 1080  
 gaggcagtgc cgggtgggcc aggccgtgaa ccggagggtt tgcgccaagc cggcgggcag 1140

gaagccgaag aattactcca agaaggttta aaaccggttt tggaggaatg cgataac 1197

<210> 48  
 <211> 1425  
 <212> DNA  
 <213> Homo sapiens  
 <400> 48

ggggtggaag atttggcggc atcttacatc gtattaaagc ttgagaacga aatccggcag 60  
 gcgcaggtcc aatggttaat ggaggaaaac gccgcctgc aggccagat cctgaactt 120  
 caaaaagtcg aagccgcgaa ggagtatgat ctctcgcgta aatcttcgga ggcgaaggag 180  
 ccgcaaaaac tgcagaaca tatgaatcca ccggccgctt gggaagcaca aaagactcca 240  
 gagttaagga aaccacagaa acctctgaa ccacaggatt tgcttccttg ggagccgctt 300

gctgcctggg agttgcaaga agcaccgget gccctgagt cactggctcc gectgcaacc 360  
 cgtgagtctc agaaaccacc tatggcgcgcat gaaatcccta ctgtattgga ggggcaaggg 420

cctgccaaca cacaagacgc tacgattgct caagaaccaa agaatagcga gccgcaagac 480  
 cctccaata tcgagaaacc tcaggaagct ccggaatata aagaacacgc ggcacagttg 540  
 gagtttttag aacttcctcc acctcaggag ccaactcgaac cgagcaatgc gcaagaattt 600  
 ctcgagttgt cggtcgccca ggagtcctta gaaggcctca ttgtagttga aacgtccgcg 660  
 gcttcggagt tcccacaggc tcctatcggg ctfgaagcca ccgactttcc gctgcagtac 720  
  
 acgcttacct tcctcggcga cagccagaag ttgccagaat ttttggcca actctacagt 780  
 tatatcgggg tacgtgggca cttataccct accgaggcgg cgttagtgtc gttttaggc 840  
 aattgtttct caggcgcgcg gggctggtgg ttccagtgc ttttggatat ccagtgcct 900  
 ctgttagaac agtgtgaaag ttttatcccg gttctccaag acacatttga caatccggaa 960  
 aacatgaagg acgcaaacca atgcatccac cagctttgtc agggcgaggg tcatgtggcc 1020  
 acacacttcc acctcattgc acaagagctt aattgggatg aaagcacgct gtggatccag 1080  
 ttccaggaag gcctggcctc atccatccag gatgaacttt cccatacatc gcctgctacc 1140  
  
 aacctgagtg atctgattac tcaatgcac tcattagagg aaaagcctga cccaaaccg 1200  
 ttagggaagt cctcctcggc ggagggggat ggcccggaaa gtcgccagc agaaaaccaa 1260  
 cctatgcaag ctgcgatcaa ttgtcctcac atttccgaag cagagtgggt tcgttggcac 1320  
 aaaggccgcg tttgtctcta ttgcggctat ccgggtcact tcgcacgtga ttgccagtg 1380  
 aagccacacc aggcgttaca ggcagggaac attcaggctt gcaa 1425  
  
 <210> 49  
 <211> 717  
 <212> DNA  
 <213> Homo sapiens  
 <400> 49  
  
 ggggtgcagc cgcagactag caaagctgaa tcgccgctc tcgctgcctc accgaacgca 60  
  
 caaatggatg acgttattga tacattaacc tcctcgcgc tgacgaattc ggctctgcgg 120  
 cgggaggcta gcaactctcg ggccgagaaa gcaaatata ctaatatgct cgagtcagtg 180  
 atggccgagt taacgtgtt acggaccctg gcgcggattc cgggggcctc gcagattacg 240  
 ccaccaattt cgtctattac tagcaacggt actcgcctga tgacactcc tccaactagt 300  
 ttacctgaac cgtttctgg cgatcctggc cggttagctg gtttccttat gcagatggac 360  
 cgttttatga tcittcaagc tagccggttt ccaggggagg cagagcgtgt tgcgttcctg 420  
 gtgtcgcgct taactggcga agcagaaaaa tgggccattc ctcacatgca accagactct 480  
  
 cctttcgtga acaactatca aggtctctta gcagagttac ggcggaccta taagagcccg 540

ttgctcagc cccggcgggc gcaaatccgg aagacatcgg cctcgaaccg ggcagtcctg 600  
 gaacgcaaaa tgctttgccg gcaacttgca tcagcaggta caggcccatg cccggtacac 660  
 cctgctagta acgggacttc cccggcaccg gcattaccag cacgggcgcg taactta 717

<210> 50

<211> 339

<212> DNA

<213> Homo sapiens

<400> 50

ggggacggtc gggtagactt gatgaagget ttattggctg gccctttacg tccggcggca 60  
 cgccgttggc ggaatcctat tccatttcca gagacttttg atggggatac tgatcgctc 120

ccggagttta tcgtccaac ttcgtcctac atgttcgttg acgaaaatac tttctctaac 180  
 gacgctctga aagtgcattt tctcattacc cgctgacag gtccagcctt gcaatgggtc 240  
 attccgtaca ttcgtaaaga aagcccctt cttaacgact atcggggttt cctggccgag 300  
 atgaagcggg tttttgggtg ggaagaggac gaggacttt 339

<210> 51

<211> 339

<212> DNA

<213> Homo sapiens

<400> 51

ggggaaggtc gggtagactt tatgaaagcg ttgcttggcc gcccgcttcg tccagcagca 60  
 cgtcgtctgc ggaatccaat tcccttcccg gagacttttg acggggacac cgatcgctc 120

ccagagttca ttgtgcagac gtcaagctat atgttcgttg atgagaacac gttctctaac 180  
 gacgcgttga aagtgcattt ctttaattacg cgtttgactg gcccgctttt acaatgggtg 240  
 attccataca ttaagaaaga gtcaccgctt ctcaagtatt atcgcgggtt ttagccgag 300  
 atgaagcggg tcttcgggtg ggaagaagac gaagacttt 339

<210> 52

<211> 1092

<212> DNA

<213> Homo sapiens

<400> 52

gggcccgtg ggcgttgccg tcaacaaggt cctcggattc cgatttgggc agcggccaac 60  
 tatgccaacg cccaccctg gcaacaaatg gataaggctt cgccaggcgt tgcttacaca 120

cctttggttg atccttgat tgagcggcct tgttgcggtg acacggtttg tgtgcgcacc 180  
 acaatggaac agaagagcac agcgtcaggc acttgtggtg gtaagcctgc tgagcgtggt 240  
 cctctcgcgg ggcataatgcc gagctcacgc ccacatcggg ttgatttctg ttgggttctt 300  
 ggtagcgacc caggcacatt cgacggcagt ccatggctct tagatcgctt tttggcgcaa 360  
 cttggtgatt acatgagttt tcactttgaa cactaccagg acaatcag cctgtctgc 420  
 gagattcttc gtcggttaac gggccgcgct caggcatggg ctgctctta cctggacggg 480  
 gaccttcac tgccagacga ctacgaattg ttttgcgaag accttaagga ggtagtacag 540

gaccctaaca gtttcgccga gtatcacgcc gtggtgactt gtccactccc tcttcttctg 600  
 tcccaacttc ctgtagctcc tcagcttccg gtggtacgcc aataccttgc gcgcttcttg 660  
 gagggccttg ctttgatgat gggtagcggc cctcggtcac tcccggccgc tatggccaca 720  
 ccggcagtct ccggctcga aatccgttct cgttctgctt tattgaaca acaactcaca 780  
 aaggaatcca ctccaggccc gaaagagcca cctgttctcc ctgctcgcac ttgctctagc 840  
 aaaccgggtc ctgtcgaacc agccagtcca caacctgaag aggctgctcc taccgggtg 900  
 ccgctttgt cagagtcggc taaccaccg gctcagcgtc cagacctgc tcacctggt 960

ggtcctaaac cacaaaaaac cgaagaggaa gtttagaaa ctgaggggga ccaggaagtt 1020  
 agcctgggga cgccgcagga ggtcgtagaa gcgccgaaa caccaggtga accaccgctc 1080  
 agccctgggt tc 1092

<210> 53

<211> 438

<212> DNA

<213> Homo sapiens

<400> 53

ggggttgatg aattggtgct cttgttgac gcgctgtaa tgcgccatcg ggcgctttcc 60  
 attgaaaatt ctcatgtgat ggagcaactt cgcttgttgg tctgcgaacg ggcgagcctt 120  
 cttctcagg tacgtccgcc gagctgtcca gtccatttc ctgagacttt taacggggag 180

tcatcacggt tacctgagtt catcgtcaa accgcaagct atatgttagt taatgaaat 240  
 cgcttttgca atgacgcaat gaaagtcgct tttttgatta gccttcttac tggatgaagca 300  
 gaagaatggg tcgtcccata cattgagatg gattcaccaa ttcttgggga ctaccgtgcg 360  
 ttcttgatg agatgaagca gtgttttggg tgggacgatg atgaagatga cgacgatgag 420  
 gaagaggagg atgactat 438

<210> 54

<211> 1647

<212> DNA

<213> Homo sapiens

<400> 54

```

gggcctgtgg atttaggtca ggctttgggg ttgttgccat ccctcgctaa ggccgaagat      60

tcccaattta gcgaaagcga tgcagcttta caggaggaat tgtcttctcc gaaaccgca      120
cggaactttt ttcgtcaatt tcgctataca gtcattgtcgg ggccctcatga aacctgaaa      180
cagttacgga agttatgttt tcagtggctg caacctgaag tccatacaaa ggaacaaatc      240
ctcgaaatte tgatgctgga acagttcttg accattctgc ctggtgaaat tcagatgtgg      300
gtccgcaagc agtgccttgg tagtggggag gaggcggtta cgttagtaga atccctgaaa      360
ggtgatccac aacggctctg gcaatggatc tccatccaag tcctgggtca ggatatcctg      420
tctgagaaaa tggagtccac ttcttgccag gtgggcgaag tggagccaca cctggaagtt      480

glacctcagg aactggggtt agagaattca tcttcagggc cgggggaact tctttcgcac      540
atcgtgaaag aggagtctga cactgaagca gagttggcgt tagcggcatc ccagccagct      600
cgtttggag aacggctgat tcgggatcag gaccttgggg cgtccctcct cccggcagca      660
ccgcaggagc aatggcgtca attagacagc actcaaaaag aacaatattg ggacctgatg      720
ctggagacct acggcaaaat ggtatccggc gcgggtatct cacaccgaa gtccgattta      780
acgaactcaa ttgagttcgg tgaagagttg gcaggtatctt attacatgt aaacgaaaag      840
attccgcggc ctacctgcat tggtgaccgc caagaaaacg acaagaaaa ccttaatttg      900

gaaaaccate gtgaccagga attattacat gccagctgcc aggcctcggg cgaagtcca      960
tcccaggcat cgttacgtgg cttctttacc gaggacgaac ctggttgctt cggcgaaggg      1020
gagaaccttc ctgaggcact tcagaatata caggatgagg ggactggcga acagctgagc      1080
ccgcaagaac gcattagtga aaaacagttg ggtcaacatt tgccaaatcc gcaactcggg      1140
gagatgtcga cgatgtggct tgaagaaaaa cgggagacca gccagaaagg ccaaccacgt      1200
gcaccaatgg cgcagaaatt gccaacgtgc cgcgaatgtg gcaaaacgtt ttatcgcaat      1260
agtcaactta tctttacca acgcacacac accggtgaga cataatctca atgcaccate      1320

tgcaaaaagg cgtttctccg gtcatctgat ttcgtgaaac atcagcggac tcatactggc      1380
gaaaaacctt gtaaatgtga ctattgtgga aagggtctta gtgattttag cgggcttcgg      1440
catcacgaga agatccatac cggcgagaag ccatacaagt gtccaatctg tgagaaatct      1500
ttcatccage gcagtaattt taaccgccac caaccgggtc acaccggtga aaagccttat      1560
aatgctcgc attgtggcaa gagcttcagc tggagctcct cgctcgataa gcatcaacgt      1620
tcacatctgg ggaagaagcc gttccaa      1647
    
```

<210> 55

<211> 1053

<212> DNA

<213> Homo sapiens

<400> 55

```

gggactctcc gcttacttga ggattggtgt cgggggatgg acatgaacce acgtaaggcc      60
cttcttatcg ccgggatttc ccagtcatgt tcagtcgccg agattgaaga ggcgctccaa     120
gccgggcttg ctcccttagg cgagtatcgt ctcccttgggc ggatgtttcg ccgcatgaa     180
aatcgcaaag tagcgttggg ttggtctcaca gctgaaacta gccatgcgct tgtacctaaa    240
gaaattcctg gtaaaggcgg gatctggcgg gttatittta aaccaccgga cccggacaat     300
acgtttcttt ctcgtttgaa tgagtctctc gcgggcgagg ggatgacggt gggggaactt    360
agtcgtgctc ttggtcacga aaatgggtca ttagaccctg aacagggtat gattccggaa     420

atgtgggcgc cgatgctggc acaggctctg gaggctctcc aaccggcttt acagtgcctt     480
aagtacaaga agctgcgcgt tttttcaggg cgcgagcttc cagagccggg tgaggaggaa     540
ttcggccggt ggatgttcca taccaccag atgatcaaag cgtggcaggt gccgcatgct     600
gagaaacgcc gccggctggt ggaatcactc cgcgggccgg cacttgacgt tattcgggtt     660
ctgaaaatta acaaccctgt aattacggta gatgaatggt tgcaagcact tgaagaggtc     720
tttgggtgta ctgacaatcc tcggaattg caagtaaat acttaacgac ctaccataag     780
gacgaggaga aattatcagc ctacgtactg cggctggaac cgctgctgca gaagctcgtc     840

cagcgggggg ctattgaacg ggacgctggt aatcaggctc gcctggatca ggtaatcgct     900
ggggcggtag ataaaactat ccgccgtgag ctgaacctgc ctgaagacgg gccggcgcca     960
ggctttcttc aactcctcgt ttgattaag gattacgagg cagctgaaga ggaggaagca    1020
ttacttcagg ccattcttga agggaacttt act                                  1053

```

<210> 56

<211> 2124

<212> DNA

<213> Homo sapiens

<400> 56

```

gggacagAAC ggcgtcgcga cgaattaagt gaagaaatta ataatcttcg tgaaaaggtt      60
atgaaacaga gtgaggaAAA caacaatctt caatcccaag tccagaaact cactgaggag     120

aatactacac tccgtgagca agttgaacct acacctgaag atgaagatga cgacattgag     180

```

ttgcggggcg cagcagccgc agccgcgcct ccgccgccga tgcaggagga atgcccggag 240  
 gatttaccgg aaaaatttga tggtaatccg gacatgttag cgccattcat ggcccagtgc 300  
 caaatTTTTA tggaaaagtc tacgcgcgat tttagtgtag atcgcgtacg tgtatgtttt 360  
 gtgacgagca tgatgactgg tgcgcagcc cgttgggcgt cagcгааatt ggagcggtcg 420  
 cactactga tgcataatta cccgcgcttc atgatggaga tgaaacacgt gtttgaagac 480  
 ccgcagcggc gggagggtggc caaacgcaag atccggcggg tgcggcaggg catgggcagc 540  
  
 gtaattgatt atagtaatgc gtttcaaatg attgvcagcagg atctggattg gaatgaacct 600  
 gctctcattg atcaatatca tgaaggcctt agtgaccata ttcaagagga actctctcac 660  
 ctggaagtgg ctaaactctt ctccgccctt attggccaat gcattcatat tgagcgcctg 720  
 cttgcacgtg ctgctgccgc tgcгааaccg cgtagtcac cacgggcttt agtgctccca 780  
 catatcgcgt cacaccatca agtagatcct actgagccag tggggggtgc acgcatgcgc 840  
 ttaaccaag aagaaaagga acgtcgtcgt aagctgaatt tatgcctgta ctgcggcact 900  
 ggtggccatt atcccataa ctgtcctgcc aaagccagta agtcaagccc ggctgggaaa 960  
  
 ctccaggte ctgcctcga gggccttct gctaccggcc cagagattat ccgctccccg 1020  
 caagacgatg cgtcgtgcc tcatctccag gtaatgctcc aaatccacct cctggccgg 1080  
 cacacactct ttgtccgggc gatgattgac tctggggcgt ctggtaattt tattgatcac 1140  
 gagtatgttg ctcaaaatgg tatecctctc cggatcaaag actggcctat tctggttgaa 1200  
 gccatcgatg gccgtccgat cgcgagcggg cctgtggttc atгааaccga tgacctatc 1260  
 gttgatctgg gtgaccaccg tgaagtatta tcctttgatg tgactcagtc accgtttttt 1320  
 ccagttgttt tggcgtccg ttggctttcg actcacgatc ctaacatcac gtggtcgaca 1380  
  
 cggtcgattg tcttcgattc ggaatattgt cgttatcatt gccgatgta ttcaccaatt 1440  
 ccgccgtctc tcccgcccc tgcgcgcгаа cctcctctgt attacccggg ggacggttac 1500  
 cgtgtttacc agccagttcg ctactactac gtacaaaacg tgtacacgcc tgttgatgaa 1560  
 cacgtgtacc cagatcaccg cctggtcgac cctcatattg agatgatccc ggggtcgcac 1620  
 tcgatcccat cgggccaatg ttattccttg tctgagccag aaatggccgc cttaagggat 1680  
 tttgtggccc ggaatgtcaa agacggcctg attacccga caattgcacc aaacggtgct 1740  
 caggtgttgc aggtgaagcg gggctggaag ttgcaagtca gctatgattg tctgvcgcca 1800  
  
 aacaacttca ctattcagaa ccaatatcca cgtctcagca tccctaatct cgaggaccag 1860  
 gcacatcttg caacatafac tgaatttgta cctcagattc ctggctatca gacttatcct 1920  
 acgtatgctg cctacceaac ataccgggta ggtttcgcac ggtaccagt aggccgggac 1980  
 gggcagggcc gctctttata tgttctgtc atgattacat ggaacccgca ttggtaccgc 2040

cagcctccgg tccacagta cccacctcct caacctccac cacctccgcc gcctcctcca 2100  
 ccgccacctt cttactcgac atta 2124  
 <210> 57  
 <211> 60  
 <212> DNA  
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 oligonucleotide  
 <400> 57  
 atgcatcacc atcaccatca cggctcaggg tctggtagcg aaaatctgta cttccagggg 60  
 <210> 58  
 <211> 20  
 <212> PRT  
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 peptide  
 <400> 58  
 Met His His His His His His Gly Ser Gly Ser Gly Ser Glu Asn Leu  
 1 5 10 15  
 Tyr Phe Gln Gly  
 20  
 <210> 59  
 <211> 6  
 <212>  
 > PRT  
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 6xHis tag  
 <400> 59  
 His His His His His His  
 1 5  
 <210> 60  
 <211> 6  
 <212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 60

Gly Ser Gly Ser Gly Ser

1 5

<210> 61

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic peptide

<400> 61

Glu Asn Leu Tyr Phe Gln Gly

1 5

<210> 62

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic primer

<400> 62

aagctcattt cctggtatga caacga 26

<210> 63

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic primer

<400> 63

agggtctctc tcttctctt gtgct 25

<210> 64

<211> 25

<212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     primer  
 <400> 64  
 gctcaacctg ggaactgcat ctgat 25  
 <210> 65  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     primer  
 <400> 65  
 taatcctggt tgctccccac gcttt 25  
 <210> 66  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 ><223> Description of Artificial Sequence: Synthetic  
     primer  
 <400> 66  
 ggcccctcag ctccagtgat tc 22  
 <210> 67  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Description of Artificial Sequence: Synthetic  
     primer  
 <400> 67  
 cctggtgtca ctctcctggc tctga 25  
 <210> 68  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 primer

<400> 68

gccaagacat aagaacctc gcct 24

<210> 69

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 primer

<400> 69

gtgaatcaac atcctccctc cgtc 24

<210> 70

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 Peptide

<220><221> MISC\_FEATURE

<222> (1)..(25)

<223> This sequence may encompass 1-5 "Glu Ala Ala Ala Lys"

repeating units

<400> 70

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu

1                    5                    10                    15

Ala Ala Ala Lys Glu Ala Ala Ala Lys

20                    25

<210> 71

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
 Peptide

<220><221> MISC\_FEATURE

<222> (1)..(25)

<223> This sequence may encompass 1-5 "Glu Ala Ala Ala Arg"  
repeating units

<

<400> 71

Glu Ala Ala Ala Arg Glu Ala Ala Ala Arg Glu Ala Ala Ala Arg Glu

1                    5                    10                    15

Ala Ala Ala Arg Glu Ala Ala Ala Arg

                  20                    25

<210> 72

<211> 50

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

Polypeptide

<220><221> MISC\_FEATURE

<222> (1)..(50)

<223> This sequence may encompass 1-10 "Gly Gly Gly Gly Ser"  
repeating units

<400> 72

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

1                    5                    10                    15

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

                  20                    25                    30

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

                  35                    40                    45

Gly Ser

50

<210> 73

<211> 40

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

Polypeptide

<220><221> MISC\_FEATURE

<222> (1)..(40)

<223> This sequence may encompass 1-10 "Gly Gly Gly Ser"  
repeating units

<400> 73

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

1                    5                    10                    15

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

                  20                    25                    30

Gly Gly Gly Ser Gly Gly Gly Ser

                  35                    40

<210> 74

<211> 18

<212> PRT

<213> Artificial Sequence

<220>

><223> Description of Artificial Sequence: Synthetic

Peptide

<400> 74

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

1                    5                    10                    15

Leu Asp

<210> 75

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 75

Glu Gly Lys Ser Ser Gly Ser Gly Ser Glu Ser Lys Ser Thr

1                    5                    10

<210> 76

<

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<400> 76

Gly Gly Ala Ala Asn Leu Val Arg Gly Gly

1                    5                    10

<210> 77

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<400> 77

Ser Gly Arg Ile Gly Phe Leu Arg Thr Ala

1                    5                    10

<210> 78

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<400> 78

Ser Gly Arg Ser Ala

1                    5

<210> 79

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<400> 79

Gly Phe Leu Gly

1

<210> 80

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<400> 80

Ala Leu Ala Leu

1

<210> 81

<211>

> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
Peptide

<220><221> MOD\_RES

<222> (3)..(3)

<223> S-ethylcysteine

<400> 81

Pro Ile Cys Phe Phe

1                    5

<210> 82

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<220><221> MOD\_RES

<222> (3)..(3)

<223> Ser or Thr

<220><221> MOD\_RES

<222> (4)..(4)

<223> Leu or Ile

<220><221> MOD\_RES

<222

> (5)..(5)

<223> Ser or Thr

<400> 82

Pro Arg Xaa Xaa Xaa

1                    5

<210> 83

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<400> 83

Asp Glu Val Asp

1

<210> 84

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<400> 84

Gly Trp Glu His Asp Gly

1                    5

<210> 85

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic  
peptide

<400> 85

Arg Pro Leu Ala Leu Trp Arg Ser

1

5